Mixed-Linkage Cellooligosaccharides: A New Class of Glycoside Hydrolase Inhibitors

Sébastien Fort,[b] Annabelle Varrot,[a] Martin Schülein,[c] Sylvain Cottaz,[b] Hugues Driguez,[b] and Gideon J. Davies*[a]

A new class of inhibitors for β-D-glycoside hydrolases, in which a single α-(1→4)-glycosidic bond is incorporated into an otherwise all-β-(1→4)-linked oligosaccharide, is described. Such mixed β/α-linkage cellooligosaccharides are not transition-state mimics, but instead are capable of utilising binding energy from numerous subsites, spanning either side of the catalytic centre, without the need for substrate distortion. This binding is significant; a mixed α/β-D-tetrasaccharide acts competitively on a number of cellulases, displaying inhibition constants in the range of 40–300 μM. Using the Bacillus agaradhaerens *enzyme Cel5A as a model system, one such mixed β/α-cellooligosaccharide, methyl 4^{II},4^{III}-dithio-α-cellobiosyl-(1→4)-β-cellobioside, displays a K_i value of 100 μM, an inhibition at least 150 times better than is observed with an equivalent all-β-linked compound. The three-dimensional structure of* B. agaradhaerens *Cel5A in complex with methyl 4^{II},4^{III}-dithio-α-cellobiosyl-(1→4)-β-cellobioside has been determined at 1.8 Å resolution. This confirms the expected mode of binding in which the ligand, with all four pyranosides in the $^4C_1$ chair conformation, occupies the −3, −2 and +1 subsites whilst evading the catalytic (−1) subsite. Such "by-pass" compounds offer great scope for the development of a new class of β-D-glycoside hydrolase inhibitors.*

KEYWORDS:

carbohydrates · hydrolases · inhibitors · protein structures · thiooligosaccharides

Introduction

Oligosaccharides offer great potential as therapeutic agents,[1, 2] and much research has been directed towards the synthesis and evaluation of enzyme inhibitors both as mechanistic and structural probes and for clinical assessment. In order to utilise multiple binding sites for oligo- and polysaccharide-degrading enzymes, inhibitors must span the catalytic centre of the enzyme, avoiding enzymatic hydrolysis whilst ideally also accommodating potential substrate distortion in the −1 site (subsite nomenclature reviewed in ref. [3]). For the case of retaining and inverting α-D-glycoside hydrolases, acarbose and related compounds[4] display these features: They are non-hydrolysable whilst the valienamine moiety allows transition-state mimicry, and thus the inhibitor can utilise binding energy from numerous subsites.[5, 6] For the particular case of retaining β-glycoside hydrolases (e→e enzymes in the nomenclature of Sinnott[7, 8]), however, inhibitors which both span the active centre and harness multiple binding sites have been less forthcoming.

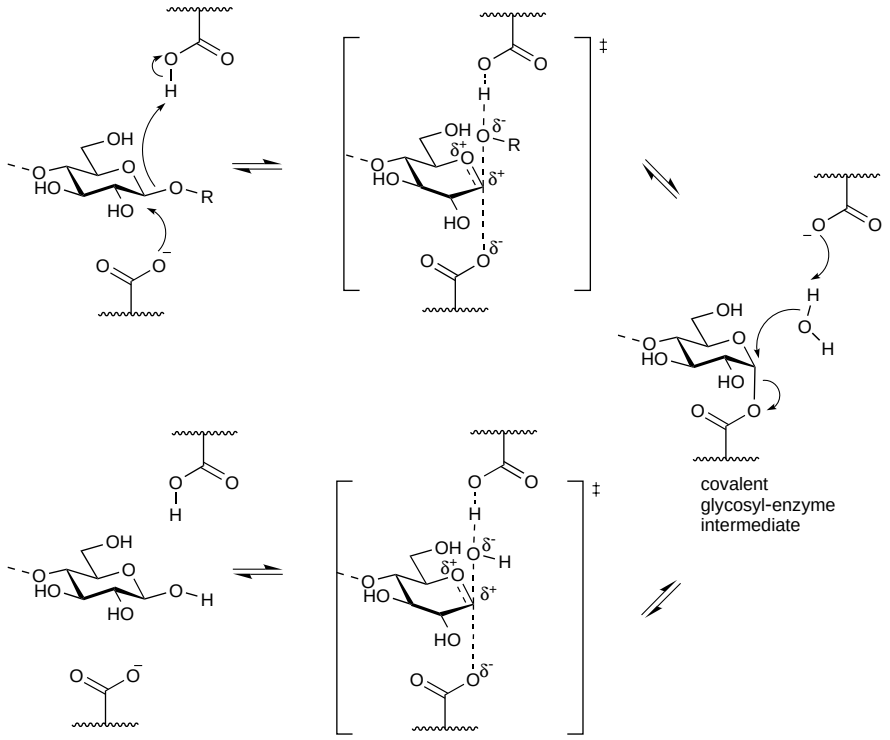
The enzymatic hydrolysis of equatorial glycosidic bonds with net retention of anomeric configuration is performed through the formation and subsequent breakdown of a covalent glycosyl-enzyme intermediate, flanked by oxocarbenium-ion-like transition states, (Scheme 1).[9–12] There is increasing evidence that a number of these enzymes create substantial substrate distortion at the −1 subsite of the active site. This distortion generates numerous catalytic benefits both in allowing "in-line" nucleophilic attack and removing potential repulsive interactions.[13–15] It is also consistent with the dictates of stereoelectronic theory since it places the lone electron pair on the endocyclic oxygen atom antiperiplanar to the scissile bond.[16, 17] Recently, three structures of retaining β-glycoside hydrolases from different sequence-derived families[18–21] have been determined in which unhydrolysed oligosaccharides span the active centre: a family-20 chitobiase in complex with its natural substrate chitobiose (**1**),[22] the endoglucanase Cel7B from *Fusarium oxysporum* in complex with the nonhydrolysable thiooligosaccharide **2**[23, 24] and the endoglucanase Cel5A from *Bacillus agaradhaerens* in complex with unhydrolysed 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-cellobioside (**3**) at low pH[13] (Figure 1).

These structures all reveal pyranoside ring distortions in the −1 subsite to classical $^1S_3$ (skew-boat) or distorted ^4E (envelope) conformations in which the scissile glycosidic bond is pseudo-

[a] *Dr. G. J. Davies, Dr. A. Varrot*
Department of Chemistry
Structural Biology Laboratory
University of York
Heslington, York, Y010 5DD (UK)
Fax: (+44) 1904-410519
E-mail: davies@ysbl.york.ac.uk

[b] *Dr. S. Fort, Dr. S. Cottaz, Dr. H. Driguez*
Centre de Recherches sur les Macromolécules Végétales (CERMAV-CNRS)
Affiliated with Université Joseph Fourier, Grenoble
B.P. 53, 38041 Grenoble cedex 9 (France)

[c] *Dr. M. Schülein*
Novozymes A/S
Smoermosevej 25, 2880 Bagsvaerd (Denmark)



Scheme 1. *Canonical reaction mechanism for a retaining β-glycoside hydrolase (an e →e enzyme in the nomenclature of Sinnott[7, 8]). A covalent glycosyl-enzyme intermediate is formed and subsequently hydrolysed via oxocarbenium-ion-like transition states.*

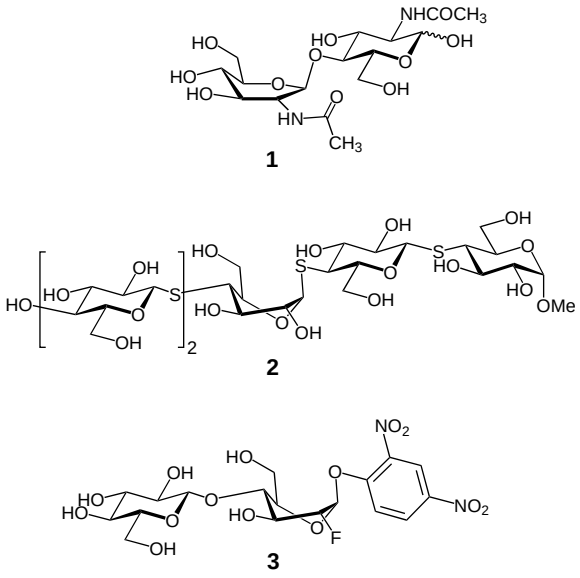


Figure 1. *Substrate distortion observed in retaining (e →e) β-D-glycoside hydrolase structures: a distorted* ^{4}E *conformation for the* N*-acteylglucosamine moiety of chitobiose (**1**) in a complex with chitobiase[22]; $^1S_3$ skew-boat conformation for an endoglucanase – thiooligosaccharide complex[23, 24] (**2**) and a $^1S_3$ skew-boat in an endoglucanase complex with 2,4-dinitrophenyl 2-fluoro-2-deoxy-β-D-cellobioside (**3**) at low pH[13].*

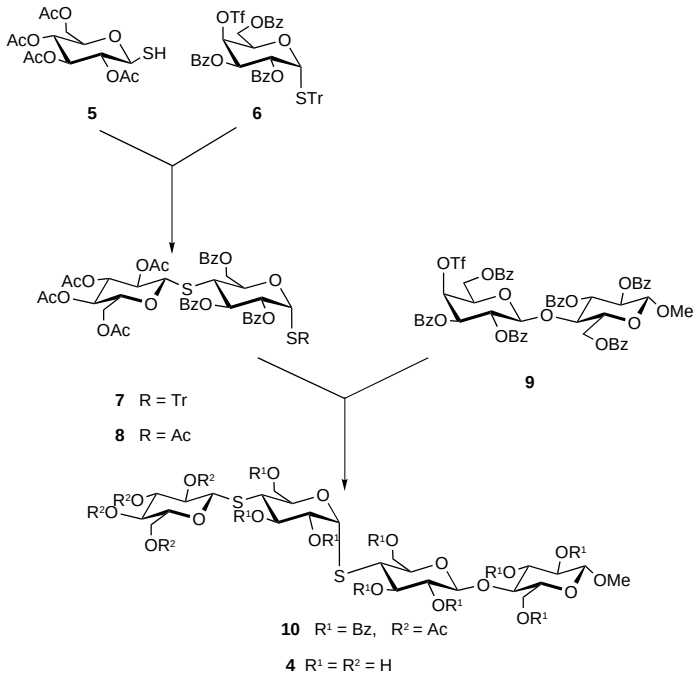
axial and the departing group is displaced 8 – 10 Å "above" the position one might predict were the − 1-subsite sugar modelled in a $^4C_1$ (chair) conformation. Distortion would appear to be driven primarily by interactions of the sugar in the +1 subsite since the product complexes for these, and other, systems show no pyranoside ring distortion.[14, 24] The energetic cost of this distortion is reflected in the negative contribution to binding for this (− 1) subsite.[25] In order to generate oligosaccharide mimics which span the active centre and thus utilise binding energy from subsites on both sides of the scissile bond, inhibitors should accommodate such distortion and be resistant to enzymatic hydrolysis.

Thiooligosaccharides are hydrolytically inert oligosaccharide mimics. They have found great use as probes for enzyme structure and mechanism.[23, 26–29] Here we describe a new class of thiooligosaccharides which are "micromolar" inhibitors of β-glycoside hydrolases. These compounds contain a single α-D-anomeric centre in an otherwise all-β-D-oligosaccharide. We present the synthesis of methyl 4^{II},4^{III}-dithio-α-cellobiosyl-(1 →4)-β-cellobioside (**4**), its behaviour as a competitive inhibitor of different classes of glycoside hydrolases and its "bypass" mode of binding as revealed by X-ray crystallography.

Results and Discussion

Initial X-ray diffraction studies on a variety of *B. agaradhaerens* Cel5A – inhibitor complexes in our laboratory had revealed the serendipitous, preferential binding of trace contaminants of a diastereoisomer containing a single α-1,4-glycosidic linkage in an otherwise all-β-1,4-linked cellooligosaccharide.[30] NMR analysis demonstrated that the purity of all these compounds was greater than 98% so the contaminant species must have displayed significantly tighter binding than the β-linked compound. In order to quantify inhibition by the α/β-oligosaccharides, the tetrasaccharide methyl 4^{I},4^{II}-dithio-α-cellobiosyl-(1 →4)-β-cellobioside (**4**) was specifically synthesised (Scheme 2). This compound not only contains the appropriate α-glycosidic linkage, but also features two *S*-glycosidic linkages to reduce susceptibility to enzymatic hydrolysis.

The tetrasaccharide **4** was prepared as shown in Scheme 2. 2,3,4,6-tetra-*O*-acetyl-β-D-1-thio-glucopyranose (**5**)[31] was coupled with the triflate **6**[32] through cysteamine activation[33] in 70% yield. The resulting disaccharide **7** was converted, in 91% yield, into the SAc derivative **8** by acidic hydrolysis[34] followed by acetylation. In situ de-*S*-acetylation of **8** by using diethylamine and thioglycosylation reaction[35] with the glycosyl acceptor **9**[36] afforded the tetrasaccharide **10** in 62% yield. After de-*O*-acylation, the expected compound **4** was obtained in 99% yield



Scheme 2. *Synthesis of methyl 4^{II},4^{III}-dithio-α-cellobiosyl-(1→4)-β-cellobioside (**4**). Bz = benzoyl, Tf = trifluoromethanesulfonyl, Tr = triphenylmethyl.*

and its structure was confirmed by NMR spectroscopy.[37] Compound **4** was tested as a potential inhibitor for cellulases belonging to family 5 (Cel5A from *B. agaradhaerens* and Cel5 from the fungus *Acremonium* sp.), family 6 (Cel6A and Cel6B from *Humicola insolens*), family 7 (Cel7A and Cel7B from *H. insolens*), family 12 (Cel12 from *Myceliophthora* sp.) and family 45 (Cel45 from *H. insolens*) (Table 1).

Compound **4** successfully inhibits a variety of structurally and mechanistically unrelated cellulases, with inhibition constants ranging from 40 to 265 μM. On four of the enzymes tested the compound shows no inhibition within the concentration range accessible. The tightest binding inhibition is witnessed on retaining glycoside hydrolases with 4 – 5 subsites such as Cel7B from *H. insolens*[38, 39] and Cel5A from *B. agaradhaerens*,[40] as might be expected for a tetrasaccharide inhibitor. Whilst the most powerful inhibition is observed on retaining enzymes, which may reflect the known substrate distortion expected for these systems (discussed below), it is also interesting that an inverting glycoside hydrolase (Cel6A from *H. insolens*) is also inhibited with a K_i value of 200 μM. In the case of Cel5A from *B. agaradhaerens*, it is noteworthy that the equivalent all-β-linked cellooligosaccharide, methyl 4-thio-β-cellobiosyl-(1→4)-β-cellobioside[41] is not an inhibitor for this enzyme, although product insolubility means that we cannot rule out a $K_i > 15$ mM. The α-containing compound therefore binds at least 150 times better than its all-β equivalent, explaining our initial trapping of these diastereosisomers from apparently pure samples. On four of the eight systems tested compound **4** demonstrated no inhibition at detectable levels (implying K_i values > 20 mM). For the inverting endoglucanase Cel45 from *H. insolens* this is not surprising, since its extended substrate-binding cleft requires occupancy of 6 – 7 subsites as a minimum for productive binding.[39, 42, 43] In the case of Cel5A from *Acremonium* and Cel12 from *Myceliophthora*, the reason for non-inhibition is not clear. These enzymes may simply not tolerate an unusual α linkage and misplaced −1 sugar unit within the substrate-binding environment. The spread of observed inhibition constants therefore reflects the number of subsites of the target enzyme and both the tolerance and the degree of adventitious interactions that each individual enzyme makes with the misplaced sugar unit.

Observation of micromolar inhibition, by a predominantly β-linked cellooligosaccharide containing an α-glycosidic bond, on a number of unrelated all-β-glycoside hydrolases is extremely intriguing. The three-dimensional structure of *B. agaradhaerens* Cel5A in complex with **4** was therefore determined by X-ray crystallography at 1.85 Å resolution (Table 2) in order to deter-

Table 1. *Inhibition constants for compound **4** towards enzymes from different glycoside hydrolase families.*[18–21]

Enzyme	Mechanism	Organism	K_i [μM][a]
Cel5A	retaining	*Bacillus agaradhaerens*	100[b]
Cel5A	retaining	*Acremonium* sp.	n.d.[c]
Cel6A	inverting	*Humicola insolens*	200
Cel6B	inverting	*Humicola insolens*	n.d.[c]
Cel7A	retaining	*Humicola insolens*	265[d]
Cel7B	retaining	*Humicola insolens*	40[e]
Cel12	retaining	*Myceliopthora* sp.	n.d.[c]
Cel45	inverting	*Humicola insolens*	n.d.[c, f]

[a] Standard errors < 10%. [b] 4^{II}-thiocellotetraose[41] is not an inhibitor for this enzyme. Cellotriose displays product inhibition with a $K_i > 5000$ μM. [c] No detectable inhibition at concentrations up to 20 mM. [d] This enzyme displays product inhibition by β-D-cellobiose with $K_i \approx 600$ μM. [e] This enzyme is inhibited by an all-β-thiocellooligosaccharide with a $K_i \approx$ 100 μM.[58] [e] This enzyme has seven subsites[42] and cellotetraose is not a substrate.[59]

Table 2. *Refinement and structure quality statistics for the* B. agaradhaerens *Cel5A complex with compound **4**.*

Data quality[a]	
resolution of data [Å]	15 – 1.85 (1.92 – 1.85)
R_{merge}[b]	0.048 (0.265)
mean $I/\sigma(I)$	19.8 (3.7)
completeness [%]	98.7 (96.5)
multiplicity	3.0 (3.0)
Refinement statistics	
no. of protein atoms	2393
no. of ligand atoms	46
no. of water molecules	324
no. of calcium ions	3
resolution used in refinement [Å]	15 – 1.85
R_{cryst}	0.177
R_{free}	0.221
rms deviation 1 – 2 bonds [Å]	0.013
rms deviation 1 – 3 bonds [Å]	0.023
rms deviation chiral volume [Å^3]	0.126
avg main-chain *B* factor [Å^2]	22.3
avg side-chain *B* factor [Å^2]	24.7
avg substrate *B* factor [Å^2]	25.5
avg solvent *B* factor [Å^2]	34.7

[a] Data for the outer shell are given in brackets. [b] $R_{merge} = \Sigma_{hkl}\Sigma_i | I_{hkli} - \langle I_{hkl} \rangle | / \Sigma_{hkl}\Sigma_i I_{hkli}$.

mine the exact mode of binding. Compound **4** binds in the true -3, -2 and $+1$ subsites of the enzyme, whilst "by-passing" the -1 subsite altogether. The electron density is clear for all units and the individual pyranosides, all in the $^4C_1$ chair conformation, refine with temperature factors of 21, 17, 23 and 28 Å^2, respectively (Figure 2). The α-linkage is found between the -2 and $-1_{apparent}$ subsites.

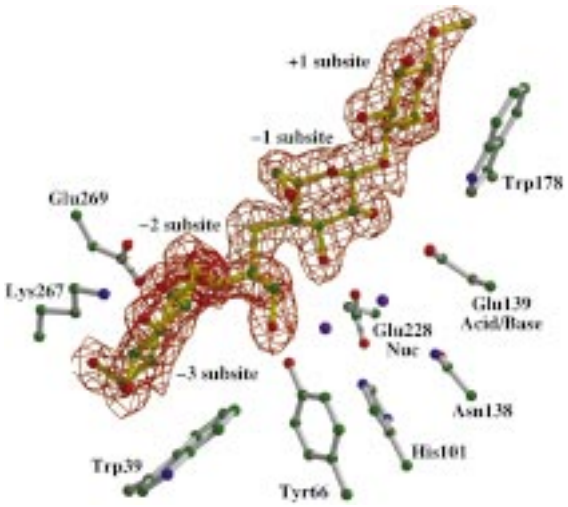


Figure 2. *Observed electron density for methyl 4^{II},4^{III}-dithio-α-cellobiosyl-(1 →4)-β-cellobioside (**4**) bound to Cel5A from* B. agaradhaerens. *The electron density map shown is a maximum likelihood/σ_A-weighted $2F_{obs} - F_{calc}$ map, contoured at 0.6 electrons per Å^2, and is a divergent stereo representation. Nuc = nucleophile.*

Compound **4** therefore makes $-3/-2$ subsite interactions identical to those observed previously in this system, in both product and trapped covalent glycosyl-enzyme complexes.[13, 40, 44] The α-(1 →4)-linkage at the boundary of the -2 and -1 subsites allows the inhibitor to bind in the $+1$ subsite in a similar mode to that which would be expected in a true productive complex, as witnessed both in Cel5A[45] and in *F. oxysporum* Cel7B[23, 24] (Figure 3). The energetic cost of distorting the -1 subsite sugar unit to form a pseudo-axial, but β-configured, linkage is avoided since the sugar evades the correct -1 subsite altogether. The pyranoside in the displaced $-1_{apparent}$ subsite is not in a catalytically viable position. The anomeric C1 carbon atom lies over 6 Å from the nucleophilic oxygen atom of Glu 228 whilst the interglycosidic atom of the potential scissile bond is not in an appropriate position to make a hydrogen bond to the catalytic acid/base Glu 139 (Figure 4).

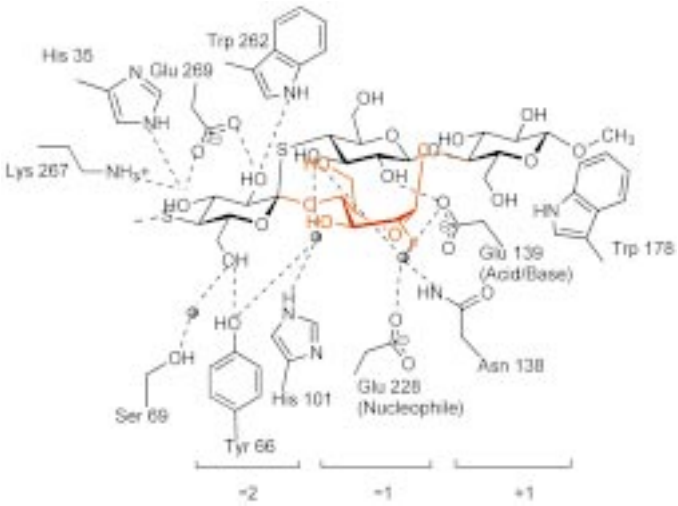


Figure 4. *Interactions of Cel5A from* B. agaradhaerens *with compound **4**. The position of the $^1S_3$ skew-boat ring observed in a "Michaelis" complex of Cel5A[13] is included for reference and is shown in red.*

Published observations of all-β-linked oligosaccharide mimics spanning the active centre of glycoside hydrolases have revealed two modes of binding. They either adopt a distorted ring conformation in the -1 subsite, reflecting the incipient oxocarbenium ion, as observed in three systems: *F. oxysporum* Cel7B,[23] *B. agaradhaerens* Cel5A and *Serratia marcescens* family 20 chitobiase.[22] Alternatively, given sufficient binding energy in the glycone sites, they may simply bind in a nonproductive manner across the active centre as observed in a mutant *Trichoderma reesei* Cel7A complex[46] and a *Pseudomonas fluorescens* Xyn10 – xylopentaose complex.[47] The mixed α/β inhibitors, described here, share features common to both these binding modes. Their intrinsic advantage, yet also their inevitable drawback, is that they "by-pass" the -1 subsite. The -1 subsite is normally associated with transition-state development and an axial glycosidic bond orientation (Figure 5). The power of the mixed β/α approach is that the α linkage in the -2 subsite allows the oligosaccharide to evade the -1 subsite, but to still achieve strong binding to all other subsites on either side of the scissile bond.

These α/β inhibitors span the active centre in a way that imitates the distortion that occurs along the reaction coordinate of retaining β-glycoside hydrolases.[13] The α linkage facilitates binding across the active centre in this manner only because adjacent glycosidic linkages in

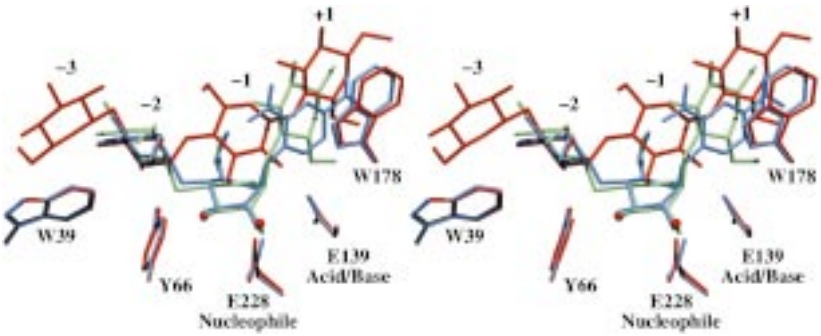


Figure 3. *Active-centre-spanning complexes of retaining cellulases reveal how an α-(1 →4)-glycosidic bond, incorporated into an otherwise all-β-(1 →4)-linked oligosaccharide, imitates substrate distortion at the active centre. The mixed-linkage inhibitor **4** (described here) bound to* B. agaradhaerens *Cel5A is shown in red, 2,4-dinitrophenyl 2-fluoro-2-deoxy-β-D-cellobioside (**3**) bound to* B. agaradhaerens *Cel5A[13] in blue and the thiooligosaccharide **2** bound to* F. oxysporum *Cel7B[23] in pale green.*

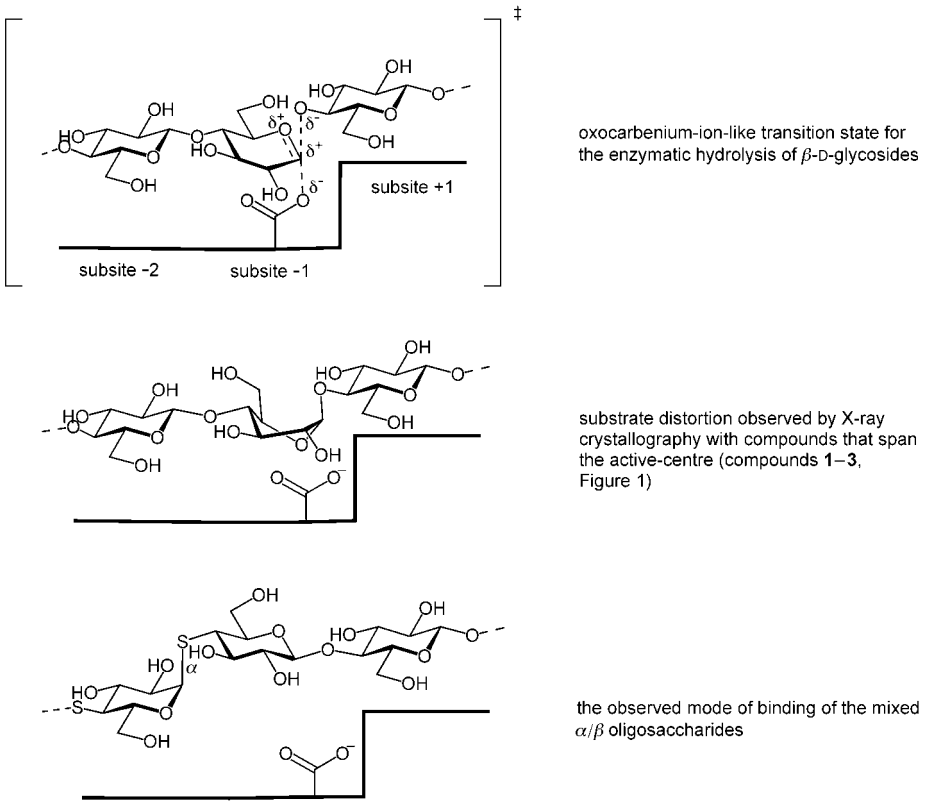


Figure 5. *Schematic representation of the oxocarbenium-ion-like transition state for the enzymatic hydrolysis of glycosides, together with the substrate distortion observed in three systems crystallographically and the mode of binding of the mixed-linkage inhibitor.*

cellooligosaccharides are related by an approximate twofold axis.[48] An (axial) α linkage in the -2 subsite mimics, in part, the pseudo-axial, but β-configured glycosidic bond from a catalytically relevant distorted species in the adjacent (-1) subsite. This permits the appropriate placement of the glycon unit in the $+1$ subsite. Whilst most success is observed with retaining glycoside hydrolases, it is noteworthy that moderate inhibition is also observed for an inverting β-glycoside hydrolase, notably one for which experimental evidence for limited substrate distortion is also known.[27]

Whilst this utilisation of the $+1$ subsite sugar unit clearly generates additional binding energy, the disadvantage of the "by-pass" compounds is that they avoid the -1 subsite. This means that, in their present form at least, they are unable to harness any degree of transition-state mimicry. Inhibition constants, in the micromolar range, are therefore relatively modest. We note, however, that a cellobiosyl imidazole, which is a partial transition-state mimic, but which utilises only two (-2, -1) of the enzyme's 5–6 subsites, also displays a K_i value in the 50–70 μM range with *B. agaradhaerens* Cel5A.[49] Furthermore, the success of these compounds is also dependent both upon the tolerance of a misplaced -1 subsite sugar unit and perhaps also its adventitious interaction with the protein. There is no basic reason why the by-passed subsite moiety need actually be a pyranoside, so we imagine that both combinatorial and rational design approaches based upon the β/α framework will, in future, generate much more powerful β-glycoside hydrolase inhibitors harnessing the binding from multiple subsites.

Experimental Section

General: NMR spectra were recorded on Bruker AC 300, Bruker Avance 400 spectrometers. Proton chemical shifts (δ) are reported in ppm downfield from tetramethylsilane (TMS), carbon chemical shifts (δ) are reported in ppm with internal reference of solvent. Complete assignment of the tetrasaccharide **4** was performed by using a combination of COSY and TOCSY 1D, HMQC and HMBC experiments. One-bond ^{13}C—^{1}H correlations were obtained from HMQC data, and the position of the glycosidic linkages was determined by using HMBC data. High-resolution mass spectra (HR-MS) were recorded on a VG ZAB spectrometer and low-resolution spectra (MS) on a Nermag R-1010C spectrometer. Optical rotations were measured with a Perkin – Elmer 341 polarimeter. Microanalyses were performed by the "Laboratoire Central d'Analyses du CNRS" (Vernaison, France). Reaction progress was monitored by analytical thin-layer chromatography using precoated silica gel 60 F254 plates (Merck, Darmstadt, Germany). All reactions in organic media were carried out under argon using freshly distilled solvents. After work-up, organic phases were dried over anhydrous Na_2SO_4.

Triphenylmethyl *S*-(2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl)-(1→4)-2,3,6-tri-*O*-benzoyl-1,4-dithio-α-D-glucopyranoside (7): Distilled pyridine (1.4 mL) and trifluoromethanesulfonic anhydride (480 μL, 2 equiv) were added to an ice-cold solution of triphenylmethyl *S*-2,3,6-tri-*O*-benzoyl-α-D-galactopyranoside[32] (1.1 g, 1.47 mmol) in CH_2Cl_2 (20 mL). The mixture was stirred for 30 min at 0 °C and then for 1 h at room temperature. The solution was diluted with water and extracted with CH_2Cl_2. The organic phase was washed with aq $KHSO_4$ (20%, v/v), sat. aq $NaHCO_3$, dried over anhydrous Na_2SO_4 and concentrated.

2,3,4,6-Tetra-*O*-acetyl-β-D-1-thioglucopyranose (**5**) (535 mg, 1 equiv),[31] 1,4-dithioerythritol (227 mg, 1 equiv) and cysteamine (114 mg, 1 equiv) were successively added to a solution of the crude triflate **6** in hexamethylphosphoramide (HMPA) (7 mL). The mixture was kept at room temperature for 1 h and then precipitated into ice water (100 mL). The solid was collected on Celite, washed with water, then dissolved in CH_2Cl_2. The organic phase was washed with water, dried over Na_2SO_4 and concentrated. Column chromatography (EtOAc/light petroleum, 3:7) afforded the disaccharide **7** (1.12 g, 70% yield). R_f (AcOEt/petroleum ether, 1:1) 0.79; $[\alpha]_D^{25} = +94.1$ ($c = 0.68$, $CHCl_3$); elemental analysis (%): calcd for $C_{60}H_{56}O_{16}S_2$: C 65.68, H 5.14, S 5.85; found: C 65.23, H 5.05, S 5.93; MS (FAB, positive mode): *m/z*: 1119 $[M+Na]^+$; HR-MS (ESI): *m/z* (%): calcd for $C_{60}H_{56}O_{16}S_2$

1119.2907 [*M*Na]$^+$, found 1119.2921; ^{1}H NMR ($CDCl_3$): $\delta = 8.10 - 7.05$ (m, 30 H, H_{arom}), 5.80 (t, 1 H, $J_{2,3} = J_{3,4} = 11$ Hz, H-3^I), 5.39 (dd, 1 H, $J_{1,2} = 5.5$ Hz, H-2^I), 5.25 (d, 1 H, H-1^I), 5.18 – 4.77 (m, 6 H, H-1^{II}, H-2^{II}, H-3^{II}, H-4^{II}, H-5^I, H-6a^I), 4.36 (dd, 1 H, $J_{5,6} = 1.8$ Hz, $J_{a,b} = 12.4$ Hz, H-6b^I), 4.00 (d, 2 H, $J = 2.9$ Hz, H-6a^{II}, H-6b^{II}), 3.63 (m, 1 H, H-5^{II}), 3.26 (t, 1 H, $J_{4,5} = 11$ Hz, H-4^I), 2.00 – 1.62 (m, 12 H, OCOCH_3); ^{13}C NMR ($CDCl_3$): $\delta = 169.9 - 165.4$ (CO), 144.1 ($C_{q,arom}$), 133.4 – 127.0 (CH_{arom}), 82.8, 81.8 (C-1^{I,II}), 75.6, 73.8, 72.3, 71.4, 69.7, 69.5, 68.2, 67.7 (C-2^{I,II}, C-3^{I,II}, C-4^{II}, C-5^{I,II}, *C*Ph_3), 63.6, 61.1 (C-6^{I,II}), 46.0 (C-4^I), 20.4 – 19.9 (OCO*C*H_3).

***S*-Acetyl *S*-(2,3,4,6-tetra-*O*-acetyl-*β*-D-glucopyranosyl)-(1 →4)-2,3,6-tri-*O*-benzoyl-1,4-dithio-*α*-D-glucopyranose (8):** Trifluoroacetic acid (3 mL) was added to a solution of compound **7** (800 mg, 0.73 mmol) and triethylsilane (230 μL, 2 equiv) in CH_2Cl_2 (3 mL). The mixture was kept for 30 min at room temperature, then concentrated and co-evaporated with toluene. Pyridine (3 mL) and acetic anhydride (1 mL) were added to the residue and the mixture was stirred for 1 h at room temperature. The solution was then concentrated and column chromatography (EtOAc/light petroleum, 3:7) afforded compound **8** (596 mg, 91 %). R_f (AcOEt/petroleum ether, 1:1) 0.63; $[\alpha]_D^{25} = +42.5$ ($c = 1.0$, $CHCl_3$); elemental analysis (%): calcd for $C_{43}H_{44}O_{17}S_2$: C 57.58, H 4.94, S 7.15; found: C 57.54, H 4.67, S 7.20; ^{1}H NMR ($CDCl_3$): $\delta = 8.10 - 7.32$ (m, 15 H, H_{arom}), 6.44 (d, 1 H, $J_{1,2} = 5.1$ Hz, H-1^I), 5.68 (t, 1 H, $J_{2,3} = J_{3,4} = 10$ Hz, H-3^I), 5.58 (dd, 1 H, H-2^I), 5.19 – 4.85 (m, 4 H, H-1^{II}, H-2^{II}, H-3^{II}, H-4^{II}), 4.82 (dd, 1 H, $J_{5,6} = 3.7$ Hz, $J_{a,b} = 12.1$ Hz, H-6a^I), 4.70 (dd, 1 H, $J_{5,6} = 1.8$ Hz, H-6b^I), 4.42 (m, 1 H, H-5^I), 4.20 (dd, 1 H, $J_{5,6} = 2.5$ Hz, $J_{a,b} = 12.4$ Hz, H-6a^{II}), 4.07 (dd, 1 H, $J_{5,6} = 5.6$ Hz, H-6b^{II}), 3.77 (ddd, 1 H, H-5^{II}), 3.34 (t, 1 H, $J_{4,5} = 10$ Hz, H-4^I), 2.34 (s, 3 H, SCOCH_3), 2.00 – 1.57 (m, 12 H, OCOCH_3); ^{13}C NMR ($CDCl_3$): $\delta = 190.81$ (*S*CO), 170.03 – 165.8 (O*C*O), 133.4 – 128.3 (CH_{arom}), 81.4, 81.0 (C-1^{I,II}), 75.6, 73.8, 73.6, 71.4, 69.6, 68.5, 68.2 (C-2^{I,II}, C-3^{I,II}, C-4^{II}, C-5^{I,II}), 63.6, 62.1 (C-6^{I,II}), 45.8 (C-4^I), 31.3 (SCO*C*H_3), 20.4 (OCO*C*H_3).

Methyl *S*-(2,3,4,6-tetra-*O*-acetyl-*β*-D-glucopyranosyl)-(1 →4)-*S*-(2,3,6-tri-*O*-benzoyl-4-thio-*α*-D-glucopyranosyl)-(1 →4)-(2,3,6-tri-*O*-benzoyl-4-thio-*β*-D-glucopyranosyl)-(1 →4)-2,3,6-tri-*O*-benzoyl-*β*-D-glucopyranoside (10): Compound **8** (560 mg, 1 equiv) and diethylamine (1 mL) were added to a solution of the triflate **9**[36] (0.62 mmol) in DMF (10 mL). The mixture was stirred for 90 min at room temperature and then concentrated. Column chromatography (EtOAc/light petroleum, 1:1) afforded the expected tetrasaccharide **10** (703 mg, 62 %). R_f (AcOEt/petroleum ether, 1:1) 0.53; $[\alpha]_D^{25} = +88.0$ ($c = 0.67$, $CHCl_3$); elemental analysis (%): calcd for $C_{96}H_{88}O_{32}S_2$: C 63.43, H 4.88, S 3.53; found: C 62.69, H 4.60, S 3.73; MS (FAB, positive mode): *m/z*: 1840 [*M*+Na]$^+$; ^{13}C NMR ($CDCl_3$): $\delta = 171.0 - 164.3$ (CO), 133.4 – 127.8 (CH_{arom}), 101.7, 100.0 (C-1^{I,II}), 84.6, 82.2 (C-1^{III,IV}), 75.7, 75.3, 75.0, 73.6, 73.4, 73.1, 72.9, 72.6, 71.8, 71.3, 71.0, 69.8, 68.5, 67.8 (C-2^{I,II,III,IV}, C-3^{I,II,III,IV}, C-4^{I,IV}, C-5^{I,II,III,IV}), 63.8, 63.5, 62.2, 61.5 (C-6^{I,II,III,IV}), 56.9 (OCH_3), 46.0, 45.7 (C-4^{II,III}), 20.5 – 20.0(OCO*C*H_3).

Methyl *S*-*β*-D-glucopyranosyl-(1 →4)-*S*-(4-thio-*α*-D-glucopyranosyl)-(1 →4)-(4-thio-*β*-D-glucopyranosyl)-(1 →4)-*β*-D-glucopyranoside (4): 1 M sodium methylate (3 mL) was added to a solution of **10** (341 mg, 0.19 mmol) in MeOH (30 mL). After 7 h at room temperature, the mixture was neutralized with Amberlite IR 120 H^+ resin, filtered and concentrated. The residue was dissolved with water, washed with diethyl ether and freeze-dried affording the free tetrasaccharide **4** (132 mg, 99 %). $[\alpha]_D^{25} = +84.9$ ($c = 1.06$, H_2O); HR-MS (FAB): calcd for $C_{25}H_{44}O_{19}S_2$ [*M*+Na]$^+$: 735.1816; found 735.1815; ^{1}H NMR (D_2O): $\delta = 5.60$ (d, 1 H, $J_{1,2} = 5.1$ Hz, H-1^{III}), 4.53 (d, 1 H, $J_{1,2} = 9.9$ Hz, H-1^{IV}), 4.38 (d, 1 H, $J_{1,2} = 8$ Hz, H-1^{II}), 4.28 (d, 1 H, $J_{1,2} = 8$ Hz, H-1^I), 4.15 (d, 1 H, $J_{5,6} = 11$ Hz, H-5^{III}), 3.91 (H-6a^{II}), 3.90 (H-6a^{III}, H-6b^{III}), 3.86 (H-6a^I), 3.77 (H-6a^{IV}), 3.76 (H-2^{III}), 3.73 (H-6b^{II}), 3.68 (H-6b^I), 3.62 (H-6b^{IV}), 3.59 (H-3^{II}), 3.56 (H-5^{II}), 3.52 (H-3^{III}), 3.51 (H-3^I), 3.50 (H-4^I, H-5^I), 3.44 (OCH_3), 3.38 (H-3^{IV}), 3.34 (H-4^{IV}, H-5^{IV}), 3.22 (H-2^{IV}), 3.18 (H-2^{I,II}), 2.76 (t, 1 H, $J_{2,3} = J_{3,4} = 11$ Hz, H-4^{III}), 2.69 (t, 1 H, $J_{2,3} = J_{3,4} = 11$ Hz, H-4^{II}); ^{13}C NMR (D_2O): $\delta = 103.4$ (C-1^I), 102.75 (C-1^{II}), 86.40 (C-1^{III}), 84.07 (C-1^{IV}), 80.12 (C-5^{IV}), 79.10 (C-4^I), 77.45 (C-3^{IV}), 75.96, 75.52 (C-3^{II}, C-5^{II}), 75.10, 74.70 (C-3^I, C-5^I), 74.68, 73.21 (C-2^{I,II}), 73.66 (C-5^{III}), 72.77 (C-2^{III}), 72.70 (C-2^{IV}), 70.30 (C-3^{III}), 69.49 (C-4^{IV}), 61.97 (C-6^{II}), 61.58 (C-6^{III}), 60.97 (C-6^{IV}), 60.39 (C-6^I), 57.56 (OCH_3), 47.52 (C-4^{II}), 47.36 (C-4^{III}).

Biochemical studies: Kinetics and K_i values were determined in a steady-state linked assay using reduced cellohexaose as the substrate (2 – 400 μM), at pH 7.5 and at 40 °C, with the *Humicola insolens* cellobiodehydrogenase with cytochrome *c*, essentially as described previously.[39, 50] K_i values were determined with the Grafit software (version 4; Erithacus Software Ltd., Horley, UK). *Myceliophotora thermophila* Cel12 was prepared as described in ref. [51] and *Acremonium* Cel5 as described in ref. [52].

X-ray structure analysis: The catalytic core domain of Cel5A from *Bacillus agaradhaerens* was purified as described previously.[40] It was desalted, washed with distilled water and concentrated to 20 mg mL^{-1}. Cel5A was preincubated with 5 mM of compound **4** for 1 h prior to crystallisation and crystals of the tetragonal crystal form of Cel5A were then grown as described.[44] Crystals were mounted in a rayon fibre loop, using the precipitant 30 % (v/v) polyethylene glycol (PEG) 400 as cryoprotectant, and placed in a boiling nitrogen stream at 100 K.

Data collection and processing, model building and refinement: All X-ray diffraction data were collected from a single crystal. A $Cu_{K\alpha}$ rotating anode operating at 50 kV and 100 mA was used as the X-ray source, with a MAR Research image plate detector system using long-focusing mirror optics. Data were processed and reduced by using the DENZO/SCALEPACK programs.[53] All further computing used the CCP4 program suite[54] unless otherwise stated. The structure was determined by the difference Fourier method using the native coordinates as a starting model (PDB code 1QHZ for the tetragonal crystal form). The maximum-likelihood $F_{obs} - F_{calc}$ electron density map revealed unambiguous electron density for all parts of the tetrasaccharide **4**. The structure was refined using the REFMAC program,[55] and the same cross-validation subset of reflections was maintained, as had originally been set aside during the determination of the native enzyme structure.[44] Stereochemical dictionaries for refinement of the substrate moiety were calculated using the program QUANTA (Molecular Simulation, Inc., San Diego, USA). Manual corrections to the model were performed using the X-FIT routines of QUANTA. Water molecules were introduced automatically using the program ARP[56] and verified manually prior to coordinate deposition. Coordinates have been deposited with the Protein Data Bank[57] under PDB code 1E5J. Details of the data and model quality are given in Table 2.

This work was funded, in part, by the European Union (under contract BIO4-CT97-2303), Novo-Nordisk A/S and by the Peter Wall Institute for Advanced Studies and Department of Chemistry of the University of British Columbia. The Structural Biology Laboratory at York is supported by the BBSRC and CERMAV by CNRS. G.J.D is a Royal Society University Research Fellow.

[1] P. Sears, C. H. Wong, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12086 – 12093.
[2] P. Sears, C. H. Wong, *Angew. Chem.* **1999**, *111*, 2446 – 2471; *Angew. Chem. Int. Ed.* **1999**, *38*, 2300 – 2324.
[3] G. J. Davies, K. S. Wilson, B. Henrissat, *Biochem. J.* **1997**, *321*, 557 – 559.
[4] D. D. Schmidt, W. Frommer, B. Junge, L. Müller, W. Wingender, E. Truscheit in *First international symposium on acarbose* (Ed.: W. Creutzfeldt), Excerpta Medica, Amsterdam, **1981**, pp. 5 – 15.

[5] R. Mosi, H. Sham, J. C. Uitdehaag, R. Ruiterkamp, B. W. Dijkstra, S. G. Withers, *Biochemistry* **1998**, *37*, 17192 – 17198.
[6] Z. Dauter, M. Dauter, A. M. Brzozowski, A. Pedersen, K. S. Wilson, G. J. Davies, *Biochemistry* **1999**, *38*, 8385 – 8392.
[7] M. L. Sinnott in *Enzyme mechanisms* (Eds.: M. I. Page, A. Williams), Royal Society of Chemistry, London, **1987**, pp. 259 – 297.
[8] M. L. Sinnott, *Chem. Rev.* **1990**, *90*, 1171 – 1202.
[9] D. L. Zechel, S. G. Withers, *Acc. Chem. Rev.* **2000**, *33*, 11 – 18.
[10] G. Davies, B. Henrissat, *Structure* **1995**, *3*, 853 – 859.
[11] G. Davies, M. L. Sinnott, S. G. Withers in *Comprehensive Biological Catalysis*, *Vol. 1* (Ed.: M. L. Sinnott), Academic Press, London, **1997**, pp. 119 – 209.
[12] T. D. Heightman, A. T. Vasella, *Angew. Chem.* **1999**, *111*, 794 – 815; *Angew. Chem. Int. Ed.* **1999**, *38*, 750 – 770.
[13] G. J. Davies, L. Mackenzie, A. Varrot, M. Dauter, A. M. Brzozowski, M. Schülein, S. G. Withers, *Biochemistry* **1998**, *37*, 11707 – 11713.
[14] G. J. Davies in *Carbohydrases from Trichoderma reesei and other microorganisms* (Eds.: M. Claeyssens, K. Piens, W. Nerinckx), Royal Society of Chemistry, London, **1998**, pp. 166 – 173.
[15] T. W. Bentley, P. von R. Schleyer, *Adv. Phys. Org. Chem.* **1977**, *14*, 1 – 67.
[16] P. Deslongchamps, *Stereoelectronic effects in Organic Chemistry*, Pergamon, Oxford, **1983**.
[17] A. J. Kirby, *Acc. Chem. Res.* **1984**, *17*, 305 – 311.
[18] B. Henrissat, A. Bairoch, *Biochem. J.* **1996**, *316*, 695 – 696.
[19] B. Henrissat, *Biochem. J.* **1991**, *280*, 309 – 316.
[20] B. Henrissat, A. Bairoch, *Biochem. J.* **1993**, *293*, 781 – 788.
[21] B. Henrissat, G. J. Davies, *Curr. Opin. Struct. Biol.* **1997**, *7*, 637 – 644.
[22] I. Tews, A. Perrakis, A. Oppenheim, Z. Dauter, K. S. Wilson, C. E. Vorgias, *Nat. Struct. Biol.* **1996**, *3*, 638 – 648.
[23] G. Sulzenbacher, H. Driguez, B. Henrissat, M. Schülein, G. J. Davies, *Biochemistry* **1996**, *35*, 15280 – 15287.
[24] G. Sulzenbacher, M. Schülein, G. J. Davies, *Biochemistry* **1997**, *36*, 5902 – 5911.
[25] D. M. Chipman, V. Grisaro, N. Sharon, *J. Biol. Chem.* **1967**, *242*, 4388 – 4394.
[26] A. K. Schmidt, S. Cottaz, H. Driguez, G. E. Schulz, *Biochemistry* **1998**, *37*, 5909 – 5915.
[27] J.-Y. Zou, G. J. Kleywegt, J. Ståhlberg, H. Driguez, W. Nerinckx, M. Claeyssens, A. Koivula, T. T. Teeri, T. A. Jones, *Structure* **1999**, *7*, 1035 – 1045.
[28] K. A. Watson, C. McCleverty, S. Geramia, S. Cottaz, H. Driguez, L. N. Johnson, *EMBO J.* **1999**, *18*, 4619 – 4632.
[29] H. Driguez in *Carbohydrate Bioengineering* (Eds.: S. B. Petersen, B. Svensson, S. Pedersen), Elsevier, Amsterdam, **1995**, pp. 113 – 124.
[30] A. Varrot, D. Phil. Thesis, University of York (UK), **2000**.
[31] D. Horton, *Methods Carbohydr. Chem.* **1963**, *2*, 433 – 437.
[32] M. Blanc-Muesser, L. Vigne, H. Driguez, J. Lehmann, J. Steck, K. Urbahns, *Carbohydr. Res.* **1992**, *224*, 59 – 71.
[33] M. Blanc-Muesser, H. Driguez, *J. Chem. Soc. Perkin Trans. 1* **1988**, 3345 – 3351.
[34] D. A. Pearson, M. Blanchette, M. L. Baker, C. A. Guindon, *Tetrahedron Lett.* **1989**, *30*, 2739.
[35] S. Bennett, M. von Itzstein, M. J. Kiefel, *Carbohydr. Res.* **1994**, *259*, 293 – 299.
[36] V. Moreau, J. L. Viladot, E. Samain, A. Planas, H. Driguez, *Bioorg. Med. Chem.* **1996**, *4*, 1849 – 1855.
[37] S. Cottaz, H. Driguez, B. Svensson, *Carbohydr. Res.* **1991**, *228*, 299 – 305.
[38] L. F. Mackenzie, G. Sulzenbacher, C. Divne, T. A. Jones, H. Wöldike, M. Schülein, S. G. Withers, G. J. Davies, *Biochem. J.* **1998**, *335*, 409 – 416.
[39] M. Schülein, *J. Biotechnol.* **1997**, *57*, 71 – 81.
[40] G. J. Davies, M. Dauter, A. M. Brzozowski, M. E. Bjornvad, K. V. Anderson, M. Schülein, *Biochemistry* **1998**, *37*, 1926 – 1932.
[41] C. Reverbel-Leroy, G. Parsiegla, V. Moreau, M. Juy, C. Tardif, H. Driguez, J.-P. Bélaich, R. Haser, *Acta Crystallogr. Sect. D* **1998**, *54*, 114 – 118.
[42] G. J. Davies, S. P. Tolley, B. Henrissat, C. Hjort, M. Schülein, *Biochemistry* **1995**, *34*, 16210 – 16220.
[43] M. Schülein, D. F. Tikhomirov, C. Schou in *Proceedings of the second TRICEL symposium on Trichoderma reesei cellulases and other hydrolases*, *Vol. 8* (Eds.: P. Suominen, T. Reinikainen), Foundation for Biotechnical and Industrial Fermentation Research, Espoo, Finland, **1993**, pp. 109 – 116.
[44] A. Varrot, M. Schülein, G. J. Davies, *J. Mol. Biol.* **2000**, *297*, 819 – 828.
[45] G. J. Davies, *Biochem. Soc. Trans.* **1998**, *26*, 167 – 174.
[46] C. Divne, J. Ståhlberg, T. T. Teeri, T. A. Jones, *J. Mol. Biol.* **1997**, *275*, 309 – 325.
[47] G. W. Harris, J. J. Jenkins, I. Connerton, N. Cummings, L. L. Leggio, M. Scott, G. P. Hazlewood, J. I. Laurie, H. J. Gilbert, R. W. Pickersgill, *Structure* **1994**, *2*, 1107 – 1116.
[48] K. Geßler, N. Krauß, T. Steiner, C. Betzel, C. Sandmann, W. Saenger, *Science* **1994**, *266*, 1027 – 1029.
[49] A. Varrot, M. Schülein, M. Pipelier, A. Vasella, G. J. Davies, *J. Am. Chem. Soc.* **1999**, *121*, 2621 – 2622.
[50] C. Schou, M. H. Christensen, M. Schülein, *Biochem. J.* **1998**, *330*, 565 – 571.
[51] N. E. Franks, K. Kida, M. Onishi, H. Sakaguchi, M. Schülein, M. Takahashi, H. Tamagawa, M. Sharyo (Novo Nordisk A/S and Crecia Corporation), WO-A 9727363-A1, **1996**.
[52] L. N. Andersen, M. S. Kauppinen, S. F. Lassen, J. B. Nielsen, K. M. Oxenboli, M. Schülein (Novo Nordisk A/S), WO-A 9611262-A1, **1996**.
[53] "Macromolecular Crystallography": Z. Otwinowski, W. Minor, *Methods Enzymol.* **1997**, *276*, 307 – 326.
[54] "Collaborative Computational Project Number 4": *Acta Crystallogr. Sect. D* **1994**, *50*, 760 – 763.
[55] G. N. Murshudov, A. A. Vagin, E. J. Dodson, *Acta Crystallogr. Sect. D* **1997**, *53*, 240 – 255.
[56] V. S. Lamzin, K. S. Wilson, *Acta Crystallogr. Sect. D* **1993**, *49*, 129 – 147.
[57] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. T. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *J. Mol. Biol.* **1977**, *112*, 535 – 542.
[58] C. Schou, G. Rasmussen, M. Schülein, B. Henrissat, H. Driguez, *J. Carbohydr. Chem.* **1993**, *12*, 743 – 752.
[59] C. Schou, G. Rasmussen, M.-B. Kaltoft, B. Henrissat, M. Schülein, *Eur. J. Biochem.* **1993**, *217*, 947 – 953.

Received: July 26, 2000
Revised version: November 9, 2000 [F 105]