Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Atomic resolution structure of endoglucanase Cel5A in complex with methyl 4,4^{II},4^{III},4^{IV}-tetrathio-α-cellopentoside highlights the alternative binding modes targeted by substrate mimics

Annabelle Varrot,[a] Martin Schülein,[b] Sandra Fruchard,[c] Hugues Driguez[c] and Gideon J. Davies[a]*

[a]Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York Y010 5DD, England, [b]Novozymes A/S, Smoermosevej 25, DK 2880 Bagsvaerd, Denmark, and [c]Centre de Recherches sur les Macromolécules Végétales, CNRS, BP 53, 38041 Grenoble CEDEX 9, France

Correspondence e-mail: davies@ysbl.york.ac.uk

Received 27 June 2001
Accepted 23 August 2001

PDB Reference: Cel5A–*SDP5* complex, 1h5v.

Many three-dimensional structures of retaining β-D-glycoside hydrolases have been determined, yet oligosaccharide complexes in which the ligand spans the catalytic centre are rare. Those that have been reported so far have revealed two modes of binding: those in which the substrate adopts a distorted skew-boat or envelope conformation in the -1 subsite, reflecting the distortion observed during the catalytic cycle, and those which bypass the true catalytic centre and thus lie in a non-productive manner across the -1 subsite. The three-dimensional structure of a retaining endocellulase, *Bacillus agaradhaerens* Cel5A, in complex with methyl 4,4^{II},4^{III},4^{IV}-tetrathio-α-cellopentoside falls into this latter category. The 1.1 Å structure reveals the binding of five pyranosides, all in the $^4C_1$ chair conformation, occupying the -3, -2, +1 and +2 subsites whilst evading the catalytic machinery located in the true -1 subsite. Such binding is in marked contrast to the structure of another retaining endocellulase, the *Fusarium oxysporum* Cel7B, the identical ligand in which displayed a distorted skew-boat conformation at the active centre. These two binding modes may reflect different steps in the binding and catalytic process.

1. Introduction

Retaining β-glycoside hydrolases perform catalysis *via* a double-displacement mechanism in which a covalent glycosyl-enzyme intermediate is flanked by oxocarbenium-ion-like transition states (most recently reviewed in Rye & Withers, 2000; Zechel & Withers, 2000). Oligosaccharide complexes of these enzymes have rarely revealed binding modes in which the ligand spans the point of cleavage. Indeed, whilst there are many hundreds of deposited PDB files for retaining β-D-glycoside hydrolases available at the present date, only a handful contain a ligand that unambiguously traverses the catalytic centre. These complexes reveal two distinct modes of binding. In one the pyranoside in the -1 subsite is distorted towards the $^1S_3$ (skew-boat) or ^4E (envelope) conformation, reflecting the geometric constraints upon the incipient oxocarbenium ion of the natural catalytic cycle. Structures of this type have been observed for an *F. oxysporum* Cel7B–thiooligosaccharide complex (Sulzenbacher *et al.*, 1996, 1997), the complex of *B. agaradhaerens* Cel5A with 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-cellobioside (Davies, Dauter *et al.*, 1998) and the *Serratia marcesens* chitobiase bound to its natural ligand chitobiose (Tews *et al.*, 1996) (Fig. 1). Alternatively, the oligosaccharide ligand may display no distortion and simply span the active centre, evading the point of cleavage by leaving the anomeric carbon distant from the enzymatic nucleophile. Such complexes have been observed with *Pseudomonas cellulosa* Xyn10A (Harris *et al.*, 1994) and *Trichoderma reesei* Cel7A (Divne *et al.*, 1998).

Thiooligosaccharides possess many characteristics that should enable the study of active-centre-spanning complexes and have therefore found great use as probes for the study of carbohydrate–enzyme interactions (see, for example, Schmidt *et al.*, 1998; Sulzenbacher *et al.*, 1996; Watson *et al.*, 1999; Zou *et al.*, 1999; for a review, see Driguez, 1995, 2001). They are hydrolytically inert, whilst their conformation is almost identical to the natural substrate. Changes only occur at the glycosidic bond, where the angle is 97° and the bond length is 1.83 Å compared with approximately 117° and 1.41 Å in the natural O-glycosidic bond. This results in a difference of approximately 0.35 Å in the positions of adjacent sugars (Driguez, 2001). Indeed, structures of the maltodextrin phosphorylase from *Escherichia coli* in complex with its natural substrate maltopentaose and a 4-*S*-α-Glc*p*-thiomaltotetraose reveal no appreciable difference in binding of the natural substrate and its thio analogue within the error level of the experiment (Watson *et al.*, 1999). Here, we present the structure of Cel5A from *B. agaradhaerens* at 1.1 Å resolution in complex with a

thiopentasaccharide, methyl 4,4^{II},4^{III},4^{IV}-tetrathio-α-cellopentoside. A comparison of this structure with related complexes highlights the different conformations that even identical substrate mimics may adopt upon binding.

2. Experimental

The catalytic core domain of Cel5A from *B. agaradhaerens* was purified as described previously (Davies, Mackenzie *et al.*, 1998). Crystals grew in the space group $P4_32_12$ from a solution containing 28–30%(*v*/*v*) PEG 400 as precipitant, 100 m*M* triethanolamine (TEA) or HEPES pH 7.0 as buffer and 200 m*M* calcium chloride. The methyl 4,4^{II},4^{III},4^{IV}-tetrathio-α-cellopentoside complex was obtained by cocrystallization after incubation of the enzyme with 1 m*M* of the compound (whose synthesis has been described previously; Schou *et al.*, 1993) for 1 h. Crystals were mounted in rayon-fibre loops directly from the droplets. The presence of 30% PEG 400 in the crystallization conditions meant that no additional cryoprotective agents were necessary. X-ray diffraction data were collected from a single crystal at 100 K at the European Synchrotron Radiation Facility (ESRF, Grenoble), beamline ID14-2 (λ = 0.933), using an ADSC QUANTUM 4 charge-coupled device (CCD) detector. Data were processed and reduced using the *DENZO*/*SCALEPACK* programs (Otwinowski, 1993; Otwinowski & Minor, 1997). All further computing used the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) unless otherwise stated. The native Cel5A coordinates were used as the starting model (PDB code 1qhz; Varrot *et al.*, 2000). Identical cross-validation reflections utilized for the native structure determination were maintained and extended to the full resolution limit of 1.1 Å for refinement with *REFMAC*, thereby maintaining 5% of free reflections in all resolution shells (Murshudov *et al.*, 1997). At no point was a σ cutoff on intensities applied during the refinement. A spherically restrained anisotropic model for the atomic displacement parameters was guided by the behaviour of the cross-validation reflections. H atoms were added in their 'riding' positions and used for geometry and structure-factor calculations (see *REFMAC* documentation for details at http://www.ccp4.ac.uk/).

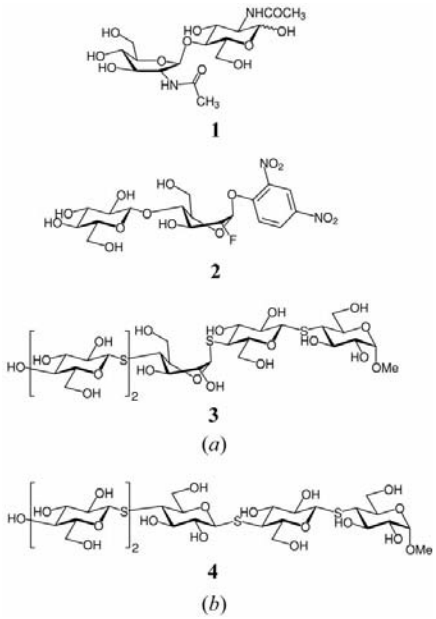

Figure 1
(*a*) Observations of distorted pyranoside rings in the active centre of retaining β-D glycoside hydrolases. (**1**) ^4E conformation of an unhydrolysed chiotobiose in the centre of *S. marcesens* chitobiase (Tews *et al.*, 1996); (**2**) $^1S_3$ skew-boat conformation of 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-D-cellobioside bound to Cel5A (Davies, Dauter *et al.*, 1998); (**3**) $^1S_3$ skew-boat conformation for methyl 4^{II},4^{III},4^{IV}-tetrathio-α-cellopentoside bound to *H. insolens* Cel7B (Sulzenbacher *et al.*, 1996). (*b*) Compound **4** is the same methyl 4^{II},4^{III},4^{IV}-tetrathio-α-cellopentoside observed as a string of $^4C_1$ chair conformed sugars, as observed in Cel5A (this work).

3. Results and discussion

Cel5A–*S*DP5 complex crystals diffract to atomic resolution. Data had an R_{merge} of 0.055 (0.28), a mean $I/\sigma(I)$ of 20.6 (4.2), a completeness of 99.5% (98.7) and a multiplicity of observation of 3.6 (3.2) in the range 20.0–1.10 Å (data in parentheses refer to the outer resolution shell, 1.14–1.10 Å). The Cel5A complex data revealed a 6° rotation of the structure in the crystal lattice relative to the other structures of Cel5A in this tetragonal crystal form. This necessitated ten cycles of rigid-body refinement prior to restrained maximum-likelihood refinement using the native tetragonal coordinates as starting model. The final model structure has an R_{cryst} value of 0.120 and R_{free} = 0.138 for data between 20.0 and 1.1 Å. This model displays deviations from stereochemical target values of 0.017 Å for bonds and 1.8° for angles, and has been deposited with PDB code 1h5v (Fig. 2).

Inspection of the initial electron-density map for Cel5A revealed clear density for all five glucose moieties of the inhibitor spanning all the active site, subsites −3 to +2 (Fig. 3*a*). All glucose moieties are in the $^4C_1$ chair conformation and the −1 subsite sugar is 'displaced' from a catalytically relevant position in the −1 subsite. Conformational change accompanies binding. In previous complex structures, a loop forming one wall of the −1 subsite underwent movement reflecting the occupancy of that subsite. Here, where the true −1 subsite is not occupied but the sugar binds in an alternative position close to the −1 site, the loop presents an intermediate conformation between these 'open' and 'closed' conformations observed previously (Davies, Dauter *et al.*, 1998; Varrot *et al.*, 2000).

The 'displaced' −1 subsite is distinguished by its paucity of direct interactions with the protein. Only the O2 hydroxyl interacts with protein directly, *via* both carbonyl O atoms of the acid/base Glu139 (distances of 3.1 and 2.6 Å, respectively). The O6 hydroxyl interacts only with solvent-exposed water. O3 interacts only with a water molecule which is in turn is hydrogen bonded to both His101 and Tyr66. This alternative −1 binding site is very similar to that occupied by the mixed β/α 'bypass' ligands described previously (Fort *et al.*, 2001) (Fig. 3*b*). The anomeric

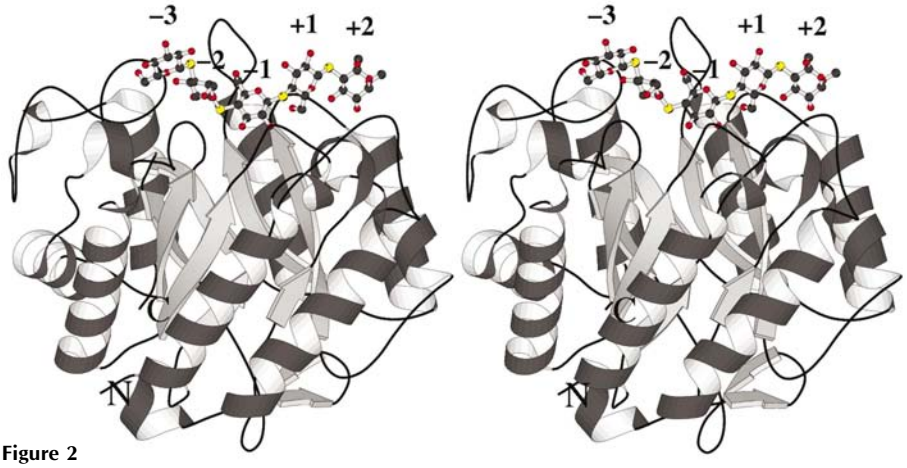

Figure 2
Divergent stereo cartoon of Cel5A in complex with thio-DP5 drawn with the *MOLSCRIPT* program (Kraulis, 1991).

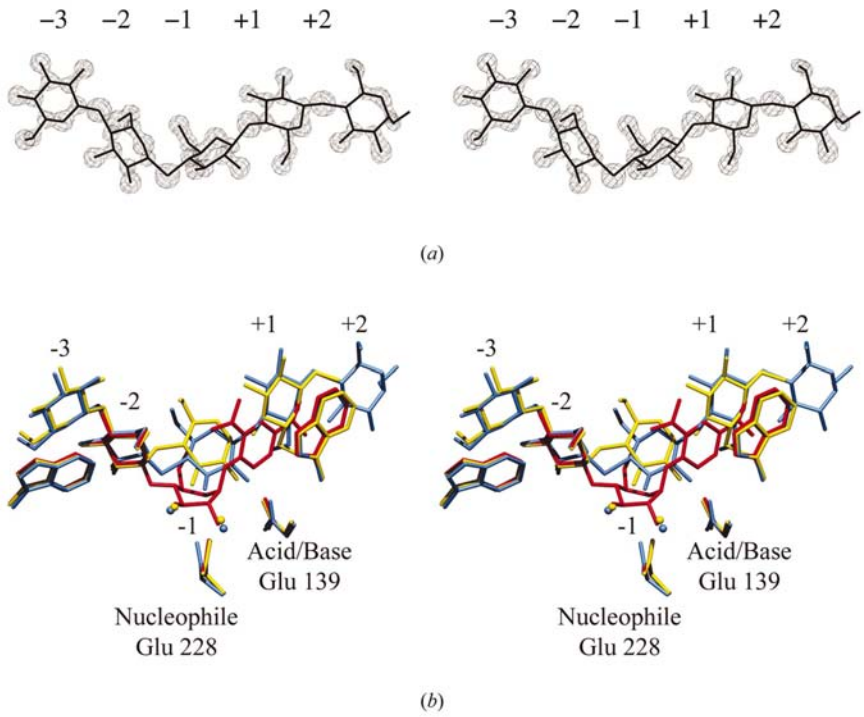

Figure 3
(*a*) Electron density for the Cel5A–*S*DP5 complex. The map shown is a maximum-likelihood weighted $2F_{obs} - F_{calc}$ synthesis contoured at 1.2 e Å^3 (approximately 5σ). The figure is in divergent (wall-eyed) stereo. (*b*) Overlay of different active-centre spanning complexes of *B. agaradhaerens* Cel5A: in red is the hydrolysed 2,4-DNP 2-F-2-deoxy-cellobioside in $^1S_3$ skew-boat conformation in the −1 subsite (Davies, Mackenzie *et al.*, 1998), in yellow the mixed β/α 'bypass' inhibitor complex (Fort *et al.*, 2001) and in blue the *S*DP5 complex (this work). The enzymatic nucleophile (Glu228) and the acid/base (Glu139) are shown for reference.

carbon, C1, lies 6.2 Å from the putative attacking O atom of the nucleophile Glu228 and 4.6 Å from the acid/base Glu139.

In the +1 subsite, there are no direct interactions with the protein except for a hydrophobic stacking platform provided by Trp178, which has rotated approximately 6° around the CG—CD bond compared with the native structure in order to optimize the interaction. The +1 sugar hydrogen bonds only to solvent-accessible water molecules. It is possible that this lack of direct hydrogen bonding may serve to assist product release and prevent tight binding of the sugar in this 'leaving-group' subsite. For the first time in glycoside hydrolase family 5 (nomenclature according to Coutinho & Henrissat, 1999), we observe a glucosyl moiety in the +2 subsite. The O2 hydroxyl is hydrogen bonded to Gln180 OE1, whilst the O6 hydroxyl interacts only with water. The O3 hydroxyl is directly hydrogen bonded to Gln180 NE2 and His206 NE2. A schematic representation of the interactions is shown in Fig. 4. This confirms the role of this residue in substrate binding, predicted through a series of suppressor mutants of a family 5 endocellulase, EGZ, from *Erwinia chrysanthemi* (Bortoli-German *et al.*, 1995). It is possible that in a productive complex the histidine would bind O2 rather than O3. A feature of the −1, +1 and +2 subsites is an unusual water network in which water molecules bridge the O6 of each sugar with either the O3 or O2 hydroxyl of the adjacent sugar in a manner we have not described previously (Fig. 4).

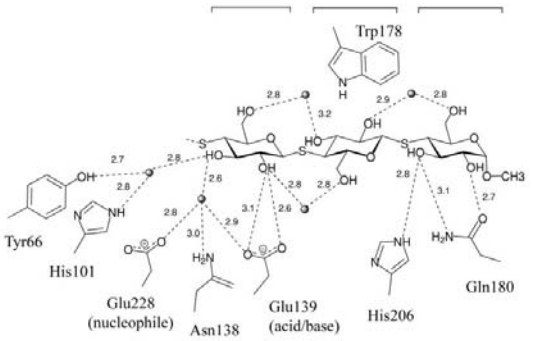

Figure 4
A schematic representation of the interactions of Cel5A with the *S*DP5 inhibitor. Only the displaced −1, +1 and +2 subsites are shown for clarity.

In the previously reported structure of Cel7B, the identical thiopentasaccharide bound in a substantially different manner. In that 2.8 Å study (later confirmed at 1.9 Å; Sulzenbacher, 1999) the sugar in −1 was distorted towards a $^1S_3$ skew-boat conformation. *S*DP5 is a competitive inhibitor for both Cel5A and Cel7B with K_i values of 117 and 30 μ*M*, respectively (Schou *et al.*, 1993). These enzymes are both retaining endoglucanases acting on the same oligosaccharide substrates with 4–5 kinetically significant subsites (Davies, Dauter *et al.*, 1998; Schou *et al.*, 1993), yet the binding modes adopted by the thiopentasaccharides through the −1 subsite are substantially different. Why these enzymes bind the ligand differently is unclear. The binding mode that 'bypasses' the catalytic machinery, exemplified by the Cel5A complex, is necessarily non-productive. It cannot be ruled out, however, that this mode represents an early step in the natural catalytic cycle: an initial binding event with the −1 subsite sugar in the $^4C_1$ chair conformation collapsing to form the productive complex for in-line nucleophilic attack (proposed in Divne *et al.*, 1998). Furthermore, observation of such complexes highlights the potential for overinterpretation of single complex structures since this non-productive binding mode, in the absence of other data, would suggest incorrect roles for many of the residues including the catalytic acid/base and the nucleophile. We have recently utilized the observation that these enzymes can tolerate or even favour an undistorted 'misplaced' sugar in the −1 subsite to harness mixed β/α-linked 'bypass' inhibitors which appear to target this alternate binding mode (Fort *et al.*, 2001). Glycoside hydrolases are important drug targets (Williams & Davies, 2001) and we hope that compounds based upon the three-dimensional structures of glycoside hydrolases in complex with oligosaccharides, such as that described here, will aid the design of novel inhibitors both as mechanistic probes and therapeutic agents in the future.

The authors would like to thank the ESRF, Grenoble for provision of data-collection facilities and for financial assistance. This work was funded in part by the European Union (under contract BIO4-CT97-2303) and Novozymes A/S. The Structural Biology Laboratory at York is supported by the BBSRC and CERMAV by

the CNRS. GJD is a Royal Society University Research Fellow.

References

Bortoli-German, I., Haiech, J., Chippaux, M. & Barras, F. (1995). *J. Mol. Biol.* **246**, 82–94.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Coutinho, P. M. & Henrissat, B. (1999). *Recent Advances in Carbohydrate Engineering*, edited by H. J. Gilbert, G. J. Davies, B. Svensson & B. Henrissat, pp. 3–12. Cambridge: Royal Society of Chemistry.

Davies, G. J., Dauter, M., Brzozowski, A. M., Bjornvad, M. E., Anderson, K. V. & Schülein, M. (1998). *Biochemistry*, **37**, 1926–1932.

Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schülein, M. & Withers, S. G. (1998). *Biochemistry*, **37**, 11707–11713.

Divne, C., Ståhlberg, J., Teeri, T. T. & Jones, T. A. (1998). *J. Mol. Biol.* **275**, 309–325.

Driguez, H. (1995). *Carbohydrate Bioengineering*, edited by S. B. Petersen, B. Svennson & S. Pedersen, pp. 113–124. Amsterdam: Elsevier.

Driguez, H. (2001). *ChemBioChem*, **2**, 311–318.

Fort, S., Varrot, A., Schülein, M., Cottaz, S., Driguez, H. & Davies, G. J. (2001). *ChemBioChem*, **2**, 319–325.

Harris, G. W., Jenkins, J. J., Connerton, I., Cummings, N., Leggio, L. L., Scott, M., Hazlewood, G. P., Laurie, J. I., Gilbert, H. J. & Pickersgill, R. W. (1994). *Structure*, **2**, 1107–1116.

Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Rye, C. S. & Withers, S. G. (2000). *Curr. Opin. Chem. Biol.* **4**, 573–580.

Schmidt, A. K., Cottaz, S., Driguez, H. & Schulz, G. E. (1998). *Biochemistry*, **37**, 5909–5915.

Schou, C., Rasmussen, G., Schülein, M., Henrissat, B. & Driguez, H. (1993). *J. Carbohydr. Chem.* **12**, 743–752.

Sulzenbacher, G. (1999). DPhil thesis, University of York, England.

Sulzenbacher, G., Driguez, H., Henrissat, B., Schülein, M. & Davies, G. J. (1996). *Biochemistry*, **35**, 15280–15287.

Sulzenbacher, G., Schülein, M. & Davies, G. J. (1997). *Biochemistry*, **36**, 5902–5911.

Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S. & Vorgias, C. E. (1996). *Nature Struct. Biol.* **3**, 638–648.

Varrot, A., Schülein, M. & Davies, G. J. (2000). *J. Mol. Biol.* **297**, 819–828.

Watson, K. A., McCleverty, C., Geramia, S., Cottaz, S., Driguez, H. & Johnson, L. N. (1999). *EMBO J.* **18**, 4619–4632.

Williams, S. J. & Davies, G. J. (2001). *Trends Biotechnol.* **19**, 356–362.

Zechel, D. L. & Withers, S. G. (2000). *Acc. Chem. Rev.* **33**, 11–18.

Zou, J.-Y., Kleywegt, G. J., Ståhlberg, J., Driguez, H., Nerinckx, W., Claeyssens, M., Koivula, A., Teeri, T. T. & Jones, T. A. (1999). *Structure*, **7**, 1035–1045.