Atomic (0.94 Å) Resolution Structure of an Inverting Glycosidase in Complex with Substrate

Diego M. A. Guérin¹, Marie-Bernard Lascombe¹, Marcelo Costabel¹
Hélène Souchon¹, Victor Lamzin², Pierre Béguin³ and Pedro M. Alzari¹*

¹Unité de Biochimie Structurale, CNRS URA 2185
Institut Pasteur, 25 rue du Dr. Roux, 75724, Paris cédex 15
France
²EMBL, Hamburg Outstation
Notkesstrasse 85, Hamburg
Germany
³Unité de Physiologie Cellulaire, CNRS URA 300
Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris cédex 15
France

*Corresponding author

The crystal structure of Clostridium thermocellum endoglucanase CelA in complex with cellopentaose has been determined at 0.94 Å resolution. The oligosaccharide occupies six D-glucosyl-binding subsites, three on either side of the scissile glycosidic linkage. The substrate and product of the reaction occupy different positions at the reducing end of the cleft, where an extended array of hydrogen-bonding interactions with water molecules fosters the departure of the leaving group. Severe torsional strain upon the bound substrate forces a distorted boat₂,₅ B conformation for the glucosyl residue bound at subsite −1, which facilitates the formation of an oxocarbenium ion intermediate and might favor the breakage of the sugar ring concomitant with catalysis.

Keywords: inverting glycosidase; X-ray crystallography; atomic resolution; protein-carbohydrate interactions; reaction mechanism

Introduction

Glycosidases (EC 3.2.1.X) are ubiquitous enzymes with functions ranging from glycosides degradation to control and mediation of cell-cell interactions. Since the pioneering work of Phillips and co-workers on lysozyme more than 30 years ago,¹ the modes of action of glycosidases have been studied extensively due to their possible use in biomass energy conversion and the potential for specific inhibitors to act as new therapeutic agents in a wide range of infectious diseases. Catalytic domains of glycosidases can be grouped in at least 83 distinct protein families on the basis of amino acid sequence similarities² — ² with enzymes within a same family following a similar catalytic mechanism.³ Several of these families involve enzymes that use a single displacement mechanism, leading to inversion of the anomeric configuration. Extensive biochemical and crystallographic studies of inverting glycosidases revealed a large diversity of structural architectures⁴ — ¹³ and substrate specificities but confirmed, to a large extent, that inverting glycosidases share a common general acid/base mechanism, as originally proposed by Koshland.¹⁴,¹⁵ However, the limited resolution of most crystallographic studies and the paucity of structural information on enzyme-substrate complexes for inverting glycosidases leave important unanswered questions concerning the detailed conformation of bound carbohydrate, the role of substrate distortion on catalysis and the nature of reaction intermediates. We report here structural evidence at atomic resolution of severe carbohydrate distortion for an inverting cellulase, Clostridium thermocellum CelA, that strongly supports an oxocarbenium ion-like transition state and illustrates how substrate-binding energy can be used as a driving force in catalysis.

Inverting endoglucanase CelA from the thermophilic anaerobe C. thermocellum is a component of the cellulosome, an extracellular multi-enzymatic complex that is very efficient in degrading crystalline cellulose.¹⁶ The catalytic domain of CelA, which belongs to family 8 of inverting glycosidases, folds into a (α/α), barrel consisting of six internal, mutually parallel α-helices interconnected by six external helices.¹⁷ This protein folding top-
ology is common among inverting glycosidases, where it is shared by family 9 endoglucanases, family 15 glucoamylases and family 48 cellobiohydrolases, but has not been observed so far for retaining glycosidases. A glutamate residue at the centre of the active-site cleft, Glu95, has been identified as the proton donor in the catalytic reaction of CelA, but the residue acting as a general base could not be identified unambiguously. To fully characterize the reaction mechanism and to analyze the role of substrate conformation on catalysis, we have produced an inactive isosteric mutant of CelA in which the proton donor Glu95 was substituted by Gln (CelA<sub>E95Q</sub>) and determined its crystal structure in complex with substrate. The crystallographic analysis of the complex at atomic resolution (0.94 Å) provides a unique insight into structural aspects of carbohydrate binding and catalysis.

Results and Discussion

Crystallographic data and model quality

Inactive CelA<sub>E95Q</sub> was crystallized in complex with the substrate cellobiose, a complete diffraction data set was collected at 0.94 Å resolution using synchrotron radiation and the structure of the complex was refined to a final crystallographic R-factor of 9.4% (Table 1). The final model includes 363 amino acid residues, 401 fully occupied solvent sites and six enzyme-bound β-glucosyl residues.

The stereochemical quality of the model is comparable to that of other crystal structures refined at atomic resolution. All non-glycine residues display main chain dihedral angles that fall within allowed (90.4%) or additionally allowed (9.6%) regions of the Ramachandran plot as defined by the program PROCHECK. Anisotropic refinement of non-hydrogen atomic temperature factors resulted in a significant quality improvement of the electron density map, in most cases allowing unambiguous determination of the proper orientation of side-chain residues by discriminating between carbon, nitrogen and oxygen atoms.

The average error estimates for the different types of chemical bonds between main-chain atoms, obtained by inversion of the block matrix, lie in the range 0.014-0.018 Å. Although bond distance and angle restraints were relaxed during the final cycles of refinement, these estimates show a remarkable overall agreement with the observed r.m.s. deviations from ideal geometry, which fall in the range 0.014-0.021 Å. However, a few experimental parameters with low error estimates deviate significantly from their target values (data not shown), probably indicating genuine differences from standard protein dictionaries.

### Table 1. Crystallographic data and refinement statistics

<table>
<thead>
<tr>
<th>A. Data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of measured reflections</td>
</tr>
<tr>
<td>Number of unique reflections</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>R-factor</td>
</tr>
<tr>
<td>Free R-factor</td>
</tr>
<tr>
<td>r.m.s. (bond lengths) (Å)</td>
</tr>
<tr>
<td>r.m.s. (angle distances) (Å)</td>
</tr>
<tr>
<td>Protein residues</td>
</tr>
<tr>
<td>Glucosyl residues</td>
</tr>
<tr>
<td>water molecules with full occupancy</td>
</tr>
<tr>
<td>Isotropic average temperature factors (Å²)</td>
</tr>
<tr>
<td>Protein atoms</td>
</tr>
<tr>
<td>sugar atoms</td>
</tr>
<tr>
<td>water molecules</td>
</tr>
</tbody>
</table>

Values in parentheses correspond to the higher-resolution shell.

The substrate-binding region

Carbohydrate binds to a deep acidic cleft of the (α/α) barrel, which runs across the molecular surface at the N-terminal end of the central helices (Figure 1(a)). The enzyme-substrate complex defines six β-glucosyl-binding subsites within this cleft. The sugar chain can be visualized as drawing a V with three subsites on each branch and the scissile glycosidic linkage at the vertex. At the non-reducing end of the substrate, subsites −3 to −1 (using the nomenclature proposed by Davies et al.), the electron density map reveals a single well-defined conformation for the bound β-glucosyl residues, whereas at the opposite end of the cleft there is clear indication of alternate conformations for the sugar residues bound at subsites +1 and +2 (Figure 1(b)). With the exception of a few hydroxyl groups that face the bulk solvent, sugar atoms in these five positions display a moderate anisotropy of their thermal displacements (Figure 2), comparable to that of neighboring protein atoms.

Partial binding of substrate is detected at an additional pocket (subsite +3) within the enzyme cleft. However, the bound glucosyl residue displays higher temperature factors (Figure 2), and no glucosyl binding to this position had been observed in previous studies of wild-type CelA complexed with different cellobiose molecules at lower resolution, possibly due to high mobility or disorder.

Bound carbohydrate was modelled as a cellobiose molecule (i.e. the substrate) bound to subsites −3 to +2 and a cellobiose molecule, which mimics the leaving group moiety bound to subsites +1 to +3, as shown in Figure 1(a). This interpretation is supported by several pieces of evidence: (a) a clear disruption of the oligosaccharide chain is observed...
at the reaction center, between subsites $-1$ and $+1$. As a consequence, the sugar ring assigned to the substrate at subsite $+1$ is covalently linked to both adjacent glucosyl residues along the cleft, while the equivalent sugar residue corresponding to the product of the reaction has a free OH-4 hydroxyl group; (b) transglycosylation is not possible because CelA is an inverting enzyme, and (c) steric hindrance due to crystal packing excludes the presence of additional sugar residues beyond subsite $-3$, whereas at the other end of the cleft (beyond subsite $+3$) there is enough space to position two additional sugar residues without interfering with neighboring molecules in the crystal.

The analysis of ligand occupation factors lends further support to this model. When the occupation factors were refined independently for each D-glucosyl residue, the sugar rings bound at subsites $-3$ to $-1$ displayed similar values of occupancy (0.66-0.73), as did those bound to subsites $+1$, $+2$ corresponding to the substrate (0.53-0.54) and the product (0.35-0.37) of the reaction. On the basis of these results, only five independent occupation factors were finally considered for the bound saccharide, as shown in Figure 2. The two extremities of the cellopentaose substrate (subsites $-3$ to $+2$) in the refined model display significantly different values of sugar ring occupancy. This could indicate that a fraction of these molecules

Figure 1. (a) Top view of the CelA$_{E95Q}$-cellopentaose complex. Sugar residues corresponding to the substrate (yellow) and product (violet) of the reaction are shown in stick representation within the enzyme-binding cleft. (b) Electron density map of the glucosyl residues bound at subsites $+1$, $+2$. (c) Stacking interactions at subsite $-2$. (d) Close view of the catalytic center. The glucosyl residue bound at subsite $-1$ displays a distorted $2.5^{B}$ boat conformation. Note the split position for the water nucleophile.
binds to the enzyme-cleft only through their non-reducing terminal residues or, alternatively, that partial substrate cleavage has taken place in the crystal. However, there are some difficulties with both models. The first hypothesis would require an unusual torsional strain of the polysaccharide chain at the catalytic center, whereas the second hypothesis seems unlikely, because the proton donor catalyst Glu95 has been mutated (no activity of CelAE95Q was detected in qualitative assays) and a significant residual activity would be required to account for the observed differences in sugar occupancy.

The inspection of difference Fourier (Fo − Fc) maps reveals residual density in the neighborhood of the reaction center that could not be modelled by a chemically feasible arrangement of solvent molecules. Interestingly, a sugar ring with a broken O5-C1 bond at subsite −1 can fit the density, and the crystallographic refinement of such an open-ring model produced a featureless difference Fourier map (data not shown). It may be tempting to speculate on the possible implications of this model, which could account for the partial binding of substrate to the enzyme cleft and brings to mind the endocyclic mechanism proposed on theoretical grounds for retaining β-glycosidases. However, it should be kept in mind that there is strong evidence against an open-ring mechanism for glycosidases, and that the residual electron density could arise from an overlapping network of partially occupied solvent sites and/or multiple sugar ring conformations at subsite −1. Therefore, further experiments are required to assess the actual significance of the above interpretation.

**Protein-carbohydrate interactions**

The bound oligosaccharide is stabilized by stacking interactions with an array of aromatic side-chain residues lined along the cleft and by an extended network of direct and water-mediated hydrogen bonding interactions with protein atoms (Figure 3). Protein-carbohydrate stacking interactions occur all along the cleft, except for the sugar residue bound at subsite −1 (Figure 3(b)). The β-faces of the glucopyranosyl rings occupying subsites −3 and −2 are stacked on the indole rings of tryptophanyl residues 205 and 132 (Figure 1(c)). At the other end of the cleft, the sugar rings bound at subsites +1 to +3 stack on the phenol groups of three tyrosine residues, at positions 372, 277 and 369, respectively. In particular, the phenolic groups of Tyr372 and Tyr277 are approximately coplanar and provide a gliding surface covering subsites (+1, +2) that could facilitate departure of the leaving group. In the crystal structure, the bound glucosyl residues display two different positions within this plane (Figure 1(b)), which were assigned to the substrate and product of the reaction, respectively, as discussed above. Product release would therefore involve a first step during which the leaving group rotates slightly and shifts away from the reaction center, while preserving the stacking interactions at subsites +1 and +2 (Figure 3(b)).

Another salient feature of the complex is the extended network of direct and water-mediated protein-carbohydrate hydrogen bonding interactions. Water molecules involved in these interactions are well defined in density, although in some cases they display highly anisotropic temperature factors or split positions related to alternate sugar conformations. Water-mediated hydrogen bonds predominate at the reducing end of the enzyme cleft (Figure 3(a)), where they could facilitate solvatation and subsequent release of the leaving product moiety. At the other end of the cleft, several protein residues are involved in direct hydrogen bonding interactions with sugar hydroxyl groups. In particular, the hydroxyl groups OH-2 and OH-3 of the glucosyl ring bound...
at subsite $-1$ displace two well-defined water molecules that occupy equivalent positions in the unliganded structure of CelA. These two hydroxyl groups form a bidentate hydrogen bonding interaction with the carboxylic acid group of Asp152 at the bottom of the active center (Figure 1(d)), an interaction that is critical to stabilize the central sugar ring in a strained boat conformation (see below). A similar situation has been described for a family 13 $\alpha$-amylase, which also bind a distorted sugar residue at subsite $-1$. In this structure, it would appear that one of the two hydrogen bonds is particularly strong (with a distance of 2.5 $\text{Å}$), leading the authors to suggest that this putative “low barrier” hydrogen bond could play a role in stabilizing key transition-state interactions for $\alpha$-amylase. In our case, the two hydrogen bonds between the sugar hydroxyl groups and Asp152 have longer distances, 2.66 and 2.77 $\text{Å}$ (Figure 3(a)), which renders less plausible the hypothesis of a low barrier hydrogen bond in CelA.

**The catalytic residues**

Family 8 glycosidases hydrolyze the glycosidic bond via a single displacement mechanism with inversion of the anomeric configuration. The carboxylate group of CelA Glu95 has been identified as the acid catalyst. This functional assignment is
now confirmed by the strong hydrogen bonding interaction of Gln95 with the oxygen atom of the scissile glycosidic linkage (Figure 4(a)). In our previous study, we have identified two aspartic acid residues, Asp152 and Asp278, as possible candidates for the base catalyst, but a definitive assignment could not be done in the absence of bound substrate. The atomic-resolution structure of the reaction center now reveals the nucleophilic water molecule, which is clearly visible in the electron density map (Figure 1(d)), facing the anomeric carbon atom of the glucosyl residue at subsite \(-1\). This water molecule occupies two alternate positions 0.9 Å apart (Figure 4(a)), consistent with a partial sugar occupancy of the enzyme cleft. One of the two solvent sites is sterically hindered by the anomeric carbon atom of the sugar ring (distance 2.8 Å) and superimposes well with an equivalent water molecule observed in the crystal structure of unliganded CelA.\(^{17}\) Besides its interaction with the anomeric carbon at subsite \(-1\), the nucleophilic water molecule is involved in strong hydrogen bonding interactions with the hydroxyl group of Tyr215 and the carboxylate group of Asp278 (Figure 4(a)), suggesting that these two residues play a critical role in catalysis and that Asp278 is the putative base catalyst in CelA. Interestingly, the tyrosine residue is strictly conserved in all family 8 endoglucanases, but the aspartic acid residue is missing from two homologous enzymes from \textit{Bacillus}, where it is replaced by an asparagine residue.\(^{27,28}\) It follows that, at least for some family 8 glycosidases, the enhancement of water nucleophilicity by a carboxylate group might not be required for the reaction to occur. Alternatively, the conserved tyrosine residue could be acting as a base catalyst in these enzymes, although this would require a substantial modification of the tyrosine \(pK_a\) to account for the enzyme activity at slightly acidic pH values.

Substrate distortion

Comparison with the ligand-free structure of wild-type CelA indicates that carbohydrate binding induces small structural changes in the protein, mostly slight reorientations of aromatic and polar side-chains in contact with substrate. In contrast, stacking and hydrogen bonding interactions with protein residues impose a continuous torsional strain upon the substrate that modifies the conformation characteristic of \(\beta(1,4)\)-linked glucosyl residues in cellulose (Figure 3(b)) significantly. A kink is observed for the oligosaccharidic chain at the scissile glycosidic linkage, which in turn imposes a distorted \(\beta\)-B conformation to the sugar ring in the \(-1\) subsite (Table 2). Presumably, full substrate occupancy of available subsites (at least \(-3\) to \(+2\)) is energetically required to stabilize the observed oligosaccharide conformation, thus explaining why CelA and other family 8 glycosidases cleave only oligosaccharides with a high degree of polymerization (DP \(>5\)).\(^{29}\) Assuming that the enzyme undergoes no major conformational change, the observed torsional stress on the substrate implies that the initial formation of the protein-sugar com-
plex should be a sequential process, in which one end of the oligosaccharide chain is free to rotate while the other binds to the cleft.

Analysis of the oligosaccharide torsion angles and ring puckering parameters reveals a significant distortion for the sugar ring bound at subsite −1. This ring displays a strained boat \(2,5^B\) conformation, whereas all other glucosyl residues (from both substrate and product) adopt the more favorable \(4C1\) chair conformation (Table 2). The unfavorable \(2,5^B\) conformation is stabilized by the tight anchoring of adjacent glucosyl residues within the enzyme cleft and by a bidentate hydrogen bonding interaction of the carboxylate group of Asp152 with the sugar OH-2 and OH-3 atoms, as discussed above (Figure 1(d)). Previous crystallographic studies of other inverting\(^9,13\) and retaining\(^30,31\) glycosidases in complex with sugar ligands had demonstrated that the pyranose ring at the point of cleavage deviates from the standard \(4C1\) conformation, although in some cases the limited resolution of those studies was not sufficient to distinguish among closely related sugar ring conformations, or indeed from a mixture of them. To our knowledge, the present example marks the first crystallographic observation at atomic resolution of a distorted boat conformation in a glycosidase-substrate complex.

The observed substrate distortion has clear implications for the hydrolytic mechanism. The planarity of the atoms C5, O5, C1, and C2 for the sugar ring in the \(2,5^B\) conformation facilitates the formation of a partial double bond between O5 and C1 during the hydrolytic reaction (Figure 4(c)), which is a requirement of inverting catalytic mechanisms proceeding through a transition state with substantial oxocarbenium ion character.\(^{15,24}\) Interestingly, a similar \(2,5^B\) conformation has been observed for the 2-fluoro-xylose residue bound in the −1 subsite of the glycosyl-enzyme intermediate of a retaining family 11 xylanase,\(^{31}\) and similar arguments were forwarded concerning the planarity of C5, O5, C1, and C2. Therefore, a parallel can be established between the single step of the inverting CelA reaction and the second step (deglycosylation) of the retaining xylanase reaction. Although the two reactions have opposite stereochemical outcomes, the formation and hydrolysis of the transition state is facilitated considerably in both cases by the achievement of a \(2,5^B\) sugar conformation at the initial state (enzyme-substrate complex or glycosyl-enzyme intermediate).

**Table 2. Oligosaccharide conformation**

<table>
<thead>
<tr>
<th>A. Sugar ring</th>
<th>(\Phi^\circ) (deg.)</th>
<th>(\Theta) (deg.)</th>
<th>QT (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>−3</td>
<td>−35</td>
<td>9</td>
<td>0.58</td>
</tr>
<tr>
<td>−2</td>
<td>−6</td>
<td>4</td>
<td>0.59</td>
</tr>
<tr>
<td>−1</td>
<td>127</td>
<td>103</td>
<td>0.54</td>
</tr>
<tr>
<td>+1</td>
<td>−116</td>
<td>11</td>
<td>0.62</td>
</tr>
<tr>
<td>+2</td>
<td>40</td>
<td>6</td>
<td>0.58</td>
</tr>
<tr>
<td>Product</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+1</td>
<td>61</td>
<td>3</td>
<td>0.59</td>
</tr>
<tr>
<td>+2</td>
<td>19</td>
<td>4</td>
<td>0.59</td>
</tr>
<tr>
<td>+3</td>
<td>−87</td>
<td>10</td>
<td>0.66</td>
</tr>
</tbody>
</table>

B. Glycosidic bond

<table>
<thead>
<tr>
<th>Substrate</th>
<th>(\Phi) (deg.)</th>
<th>(\Psi) (deg.)</th>
<th>Dist. O5(<em>{-})O3(</em>{+1}) (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>−3/−2</td>
<td>−78</td>
<td>126</td>
<td>3.15</td>
</tr>
<tr>
<td>−2/−1</td>
<td>−89</td>
<td>147</td>
<td>4.53</td>
</tr>
<tr>
<td>−1/+1</td>
<td>−44</td>
<td>145</td>
<td>3.91</td>
</tr>
<tr>
<td>+1/+2</td>
<td>−15</td>
<td>120</td>
<td>3.08</td>
</tr>
<tr>
<td>Product</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+1/+2</td>
<td>−74</td>
<td>113</td>
<td>2.76</td>
</tr>
<tr>
<td>+2/+3</td>
<td>−84</td>
<td>138</td>
<td>3.31</td>
</tr>
</tbody>
</table>

The puckering parameters (QT, \(\Theta\), \(\Phi^\circ\)) were calculated as described,\(^{37}\) dihedral angles are defined as: \(\Phi\) (O5\(_{-}\)-C1\(_{-}\)-O4\(_{-}\)+1-C4\(_{+1}\)\(_{-}\)) and \(\Psi\) (C1\(_{-}\)-O4\(_{-}\)+1-C4\(_{+1}\)\(_{-}\)-C3\(_{+1}\)\(_{-}\)).

**Materials and Methods**

**Mutagenesis, crystallisation, data collection and processing**

Overlap extension of mutated PCR fragments was used to generate the 488 bp \(PshI\-BamHI\) segment of pCT128\(^{17}\) bearing the E95Q mutation, which was exchanged against the corresponding wild-type segment. Overexpression, purification, and crystallization of CelA\(_{E95Q}\) were performed as described.\(^{35}\) A complete diffraction data set at 0.94 Å resolution was collected at the BW7 beamline of EMBL/DESY at Hamburg using a 30 cm MarResearch Imaging Plate scanner and monochromatic wavelength (\(\lambda = 0.8815\) Å) from a single frozen crystal at \(T = 100\) K. Data reduction was carried out with the computer programs DENZO and SCALEPACK.\(^{35}\) Data collection statistics are summarized in Table 1.

**Structure refinement**

Crystallographic refinement at atomic resolution was carried out using the restrained conjugate gradient least-squares algorithm as implemented in the program SHELXL97.\(^{36}\) Throughout the whole process, 197,438 reflections (representing 95% of the total data between 10 and 0.94 Å resolution) were used for refinement, and 10,391 reflections were reserved for cross-validation analysis using the free R-factor.\(^{35}\) Manual modelling was carried out using the program O.\(^{36}\) Stereochemical restraints were applied during refinement to bond lengths and angles, chiral volumes, and planar groups. Additional restraints were applied to the temperature factors of bonded atoms and to preclude large anisotropy of solvent molecules. Non-structured solvent was modelled according to Babinet’s principle, as implemented in SHELXL97.\(^{36}\)

The unliganded structure of CelA refined at 1.65 Å resolution\(^{17}\) was used as the starting model for refinement. The initial model included five glucosyl residues fitted manually to a preliminary (\(F_o - F_c\)) difference electron density map. A few cycles of refinement were carried out assuming atomic isotropic temperature factors, followed by a long stage of several refinement rounds (220 cycles) with anisotropic temperature factors for all non-hydrogen atoms. After each round, \((2F_o - F_c)\) and \((F_o - F_c)\) electron density maps were inspected visually.
to model protein residues with alternate conformations and to add or remove water molecules as required. Hydrogen atoms were included during the first refinement rounds following the introduction of anisotropic temperature factors, but were removed temporarily while modelling alternate side-chain conformations, to avoid numerical instabilities. Towards the end of refinement, stereochemical restraints were softened gradually by duplicating their standard deviations from ideal target values. The final protein model includes 363 amino acid residues, 26 of which were modelled in alternate conformations. Residues in alternate conformations occur mostly at the surface of the protein and, in some cases, split solvent sites could be attached at proper hydrogen-bonding distances to side-chains displaying double conformations.

Addition of solvent molecules was carried out using the automatic water divining procedure as implemented in SHELXL97, selecting from density peaks higher than the automatic water divining procedure as implemented electron density maps around His133.

double conformations. Hydrogen-bonding distances to side-chains displaying cases, split solvent sites could be attached at proper hydrogen-bonding distances to side-chains displaying double conformations.

Addition of solvent molecules was carried out using the automatic water divining procedure as implemented in SHELXL97, selecting from density peaks higher than 0.4 e/Å³. In parallel, water molecules with equivalent isotropic U-factor values greater than 0.8 Å² (B > 63 Å²) were removed systematically from the atomic model, and the remaining solvent structure was checked manually using stereochemical and electron density criteria. A total of 401 solvent sites with full occupancy and 98 additional solvent sites with partial occupancy (corresponding to split positions or related to disordered amino acid side-chains) could be positioned unambiguously in the electron density map. However, the structural comparison of the solvent shell with that modelled for the structure of wild-type CelA at 1.65 Å suggests that only about half of the observed water molecules occupy similar, chemically equivalent, positions in both structures. Solvent occupancies were not refined, except for split sites and water molecules coordinated to a protein region with multiple conformations. The final parameters of the refinement are shown in Table 1.

The electron density is well defined for the whole protein model, except for two short protein regions exposed to the solvent (residues 49-51 at the N-terminal end of the first α helix and residues 394-395 at the C terminus). Furthermore, in well-structured regions of the protein, difference Fourier (Fo − Fe) maps (calculated from a protein model including only heavy atoms) clearly indicate the positions of missing hydrogen atoms, as illustrated in Figure 5 for the hydrogen bonding interaction between Asp91 and His133.

![Figure 5](image-url)

**Figure 5.** View of the 3Fo − 2Fe (blue) and Fo − Fe (orange) electron density maps around His133.

### Protein Data Bank accession code

Atomic coordinates have been deposited with the RCSB Protein Data Bank with accession code 1KWF.

### Acknowledgements

This work has been supported by grants from Institut Pasteur and CNRS. We thank the European Union for support under the TMR/LSF program to the EMBL Hamburg Outstation, reference number ERBFMI- GECT980134. D.M.A.G. and M.C. are UNS-FOMEC fellows.

### References


thiooligosaccharide inhibitor at 2.0 Å resolution. 

EMBO J. 17, 5551-5562.

13. Zou, J., Kleywegt, G. J., Stahlberg, J., Driguez, H., 


Biol. Rev. 28, 416-436.


17. Alzari, P. M., Souchon, H. & Dominguez, R. (1996). The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from Clostridium thermocellum. 

Structure, 4, 265-275.


Bioorg. Chem. 20, 77-88.


Experientia, 75(Suppl.), 111-141.


Chem. Rev. 90, 1171-1202.


Bioorg. Chem. 21, 34-40.


Biochemistry, 36, 10837-10845.


Nucl. Acids Res. 18, 4248.


KSM-330. 


Biochimie, 63, 629-639.


Biochemistry, 35, 15280-15287.


Biochemistry, 38, 5346-5354.


Methods Enzymol. 277, 319-343.


J. Am. Chem. Soc. 97, 1354-1358.

Edited by R. Huber

(Received 28 September 2001; received in revised form 21 December 2001; accepted 27 December 2001)