Crystallization and preliminary X-ray diffraction analysis of *Hypocrea jecorina* Cel7A in two new crystal forms

Annette M. Bodenheimer,^a
Matthew J. Cuneo,^b Paul D.
Swartz,^a Junhong He,^b Hugh M.
O'Neill,^b Dean A. A. Myles,^b
Barbara R. Evans^c and Flora
Meilleur^{a,b*}

^aMolecular and Structural Biochemistry
Department, North Carolina State University,
Raleigh, North Carolina, USA, ^bBiology and Soft
Matter Division, Oak Ridge National Laboratory,
Oak Ridge, Tennessee, USA, and ^cChemical
Science Division, Oak Ridge National
Laboratory, Oak Ridge, Tennessee, USA

Correspondence e-mail: fmeille@ncsu.edu

Received 25 March 2014
Accepted 17 April 2014

Cel7A (previously known as cellobiohydrolase I) from *Hypocrea jecorina* was crystallized in two crystalline forms, neither of which have been previously reported. Both forms co-crystallize under the same crystallization conditions. The first crystal form belonged to space group C2, with unit-cell parameters $a = 152.5$, $b = 44.9$, $c = 57.6$ Å, $\beta = 101.2^\circ$, and diffracted X-rays to 1.5 Å resolution. The second crystal form belonged to space group $P6_322$, with unit-cell parameters $a = b \simeq 155$, $c \simeq 138$ Å, and diffracted X-rays to 2.5 Å resolution. The crystals were obtained using full-length Cel7A, which consists of a large 434-residue N-terminal catalytic domain capable of cleaving cellulose, a 27-residue flexible linker and a small 36-residue C-terminal carbohydrate-binding module (CBM). However, a preliminary analysis of the electron-density maps suggests that the linker and CBM are disordered in both crystal forms. Complete refinement and structure analysis are currently in progress.

1. Introduction

Improving the enzymatic degradation of cellulose remains one of the key challenges in the industrial production of biofuel from ligno-cellulosic biomass. A significant problem is that cellulose is composed of crystalline microfibrils which are insoluble in the native form. In order to digest crystalline cellulose, its surface must be disrupted to allow enzymes to access individual glucan chains. Understanding how cellulases from wood-degrading organisms are able to efficiently break down cellulose could greatly benefit industrial production, both in the design of more effective pre-treatment processes and by guiding the engineering of more efficient enzymes.

The wood-degrading filamentous fungus *Hypocrea jecorina* (an anamorph of *Trichoderma reesei*) secretes a complex mixture of cellulases. Although the secretome of *H. jecorina* is composed of several enzymes including endoglucanases, exoglucanases, polysaccharide monooxygenases and β -glucosidases, the exoglucanase Cel7A comprises about 60–80% of the secreted mixture. Cel7A, previously known as cellobiohydrolase I (CBHI), is a well studied exoglucanase which hydrolyzes cellulose to cellobiose and is one of the few cellulases that are capable of degrading both amorphous and crystalline cellulose. Its ability to bind to and disrupt the surface of crystalline cellulose enables the endoglucanases to access and rapidly hydrolyze the individual glucan chains separated from the cellulose microfibrils by its action, an effect known as synergy. The efficiency with which Cel7A is able to bind, disrupt and hydrolyze crystalline cellulose makes this enzyme a key target for converting biomass into biofuels.

Cel7A consists of a large 434-residue N-terminal catalytic domain (CD) capable of cleaving amorphous cellulose, a flexible linker and a small 36-residue C-terminal carbohydrate-binding module (CBM) that is required for the efficient degradation of crystalline cellulose. Studies suggest that the reducing end of a cellulose chain is threaded through the long catalytic tunnel of the CD, where it is hydrolyzed to cellobiose, which is released at the exit of the tunnel. While the flexibility of the linker region (Beckham *et al.*, 2010) makes the full-length protein structure very challenging to determine using high-resolution techniques, structures of the CD and CBM structures have been determined separately by X-ray crystallography (Divne *et al.*, 1994, 1998; Ståhlberg *et al.*, 1996) and NMR (Kraulis *et al.*, 1989;

Table 1

X-ray data-collection statistics.

Values in parentheses are for the highest resolution shell.

	C2	P ₆₃ 22, SERCAT	P ₆₃ 22, GM/CA†
Wavelength (Å)	1.00	1.00	1.85
Resolution range (Å)	35.2–1.5 (1.56–1.50)	37.65–2.70 (2.80–2.70)	47.87–2.50 (2.59–2.50)
Space group	C2	P ₆₃ 22	P ₆₃ 22
Unit-cell parameters			
<i>a</i> (Å)	152.46	155.516	153.68
<i>b</i> (Å)	44.90	155.51	153.68
<i>c</i> (Å)	57.57	137.96	137.57
α (°)	90.00	90	90
β (°)	101.25	90	90
γ (°)	90.00	120	120
<i>R</i> _{merge} ‡ (%)	7.0 (32.8)	12.4 (49.1)	18.8 (144.1)
Unique reflections	59492 (5164)	26987 (2643)	33616 (3293)
Multiplicity	3.3 (2.3)	5.6 (5.2)	11.3 (3.6)
Completeness (%)	97.1 (84.7)	99.3 (99.4)	99.9 (99.8)
$\langle I \rangle / \langle \sigma(I) \rangle$	19.9 (3.5)	18.8 (4.2)	14.7 (1.6)
Wilson <i>B</i> factor (Å ²)	15.7	29.1	42.8

† The percentage of correlation between intensities from random half-data sets, $CC_{1/2}$, is 99.6% (63.4%) (Karplus & Diederichs, 2012). ‡ $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

Creagh *et al.*, 1996), respectively. However, low-resolution solution-scattering information on full-length Cel7A has been obtained using both small-angle X-ray scattering (SAXS; Abuja *et al.*, 1989) and small-angle neutron scattering (SANS; Pingali *et al.*, 2011).

Here, we sought to determine the structure of full-length *H. jecorina* Cel7A (*HjCel7A*) in two new crystal forms in order to probe potential interactions between the different domains.

2. Materials and methods

2.1. *HjCel7A* purification

Full-length *HjCel7A* was purified as previously described (Evans *et al.*, 1994) from a commercial culture filtrate of *T. reesei* ATCC 26921 (Sigma–Aldrich, St Louis, Missouri, USA). Briefly, buffer exchange of the filtrate was performed on a Sephadex G-75 column (GE

Healthcare). Purification was then carried out on a DEAE Sepharose column (GE Healthcare) equilibrated in 50 mM sodium acetate buffer pH 5.0 and eluted with 1 M NaCl. Chromatofocusing on a Mono-P column (Pharmacia) was carried out with a 25 mM *N*-methylpiperazine buffer pH 5.4 using a linear gradient of 1:10 diluted Polybuffer 74 pH 3.0 (GE Healthcare). The buffer was then exchanged to 50 mM sodium acetate pH 5.0. SDS–PAGE was used to check the purity of the full-length protein (Fig. 1).

2.2. Crystal growth

HjCel7A was concentrated to 10 mg ml^{−1}. Crystals were grown by vapor diffusion. Hanging drops consisting of 2 µl protein solution and 0.5 µl reservoir solution were equilibrated against a 1 ml well and incubated at 35°C. Crystals belonging to space groups C2 and P₆₃22 co-crystallized from 25–31% PEG 3350, PEG 4000 or PEG 6000 in 100 mM HEPES pH 6.4 (Fig. 2). Crystal growth was improved by the addition of 0.5 µl Silver Bullet Screen D4 (Hampton Research, Aliso Viejo, California, USA), which consists of 5 mM gadolinium(III) chloride hexahydrate, 5 mM samarium(III) chloride hexahydrate, 50 mM benzamidinium hydrochloride, 0.25% (w/v) salicin, 20 mM HEPES sodium pH 6.8. Ligand-bound crystals were prepared by addition of 10 mM cellohexaose (Megazyme, Bray, Ireland). Crystals were cryoprotected in reservoir solution containing 40% PEG prior to data collection.

2.3. Data collection and refinement

X-ray diffraction data were collected on the SER-CAT and GM/CA beamlines at the Advanced Photon Source (APS), Argonne National Laboratory. Data were collected at −173°C. *HKL*-2000 (Otwinowski & Minor, 1997) was used to index, integrate and scale the data. Data statistics are summarized in Table 1. Structure-factor amplitudes were obtained from intensities using *TRUNCATE* (French & Wilson, 1978) from the *CCP4* suite (Winn *et al.*, 2011). The data from the C2 crystal and from the P₆₃22 crystal grown with cellohexaose were phased using *Phaser* (McCoy *et al.*, 2007) using PDB entry 2v3i (A. Fagerstrom, T. Liljefors, J. Stahlberg, M. Sandgren, U. Berg & R. Isaksson, unpublished work) as the search model. The molecular-replacement solution will be subjected to rounds of refinement using *REFMAC5* (Murshudov *et al.*, 2011; Winn *et al.*, 2011) followed by iterative cycles of manual rebuilding in *Coot*

Figure 1

SDS–PAGE analysis of purified and crystallized *HjCel7A*. The analysis confirms the integrity of full-length Cel7A. Lanes 1 and 2, crystals. Lane 3, molecular-weight marker. Lanes 4 and 5, purified full-length *HjCel7A* from two different preparations.

Figure 2

Crystals of *HjCel7A*. The C2 and P₆₃22 crystal forms cannot be distinguished without collecting X-ray diffraction data.

(Emsley & Cowtan, 2004) and subsequent refinement with *REFMAC5*. The complete models will be reported elsewhere.

Fluorescence analysis of a *P6₃22* crystal on the GM/CA beamline at APS identified a peak at 6720 eV corresponding to the Sm^{3+} absorption edge. A complete single-wavelength anomalous dispersion (SAD) data set was collected at this peak wavelength for experimental phasing. The Sm^{3+} substructure was determined using the *SHELXC/D/E* pipeline (Sheldrick, 2010). Output from the *SHELX* program was then input into *ARP/wARP* (Langer *et al.*, 2008) to build the model. *REFMAC5* will be used for subsequent refinement steps.

3. Results and discussion

Previous crystallographic studies of *HjCel7A* were conducted using the CD prepared from full-length *HjCel7A* submitted to papain cleavage. These CD crystals belonged to space group *I222*. Here, the crystallization of full-length *HjCel7A* produced crystals in two new space groups, *C2* and *P6₃22*, that diffracted X-rays to 1.5 and 2.5 Å resolution, respectively. Preliminary analysis of the electron-density maps shows clear density for the CD in both crystal forms, but

electron density for the linker or CBM is weak or lacking. Crystallographic disorder is expected to some degree as the high flexibility of the linker in *HjCel7A* is known (Tomme *et al.*, 1988; Beckham *et al.*, 2010; Payne *et al.*, 2013). The Matthews coefficient (Matthews, 1968) calculated for the *C2* crystal with one CD alone in the asymmetric unit is $1.98 \text{ Å}^3 \text{ Da}^{-1}$ and the solvent content is 37.89%. The Matthews coefficient calculated for the full-length protein is $1.73 \text{ Å}^3 \text{ Da}^{-1}$ with a solvent content of 28.88% (Winn *et al.*, 2011). For the *P6₃22* crystal, the Matthews coefficient calculated with just the CD is $4.80 \text{ Å}^3 \text{ Da}^{-1}$ and the solvent content is 74.40%, while the Matthews coefficient for the full-length protein was $4.19 \text{ Å}^3 \text{ Da}^{-1}$ with a solvent content of 70.69%. While we cannot exclude proteolysis of the full-length protein during crystallization, SDS–PAGE analysis of melted crystals does not indicate that the protein in the crystals is cleaved (Fig. 1), and the crystal packing observed in both space groups provides ample room for the linker and CBM to be present in the crystal but to be disordered (Fig. 3).

While the *Cel7A* linker and CBM are required for the degradation of crystalline cellulose, the CD alone is equally efficient as the full-length protein on amorphous cellulose (Cruys-Bagger *et al.*, 2013). This suggests that the CD possesses the required machinery not only to hydrolyze a celloextrin chain but also to process along a cellulose chain as cellobiose is being cleaved and released. Molecular-dynamics simulations have suggested that the exo-loop and exit loop that span residues 247–252 and 370–394, respectively, adopt different conformations depending on the binding state of the tunnel (Colussi *et al.*, 2012; Bu *et al.*, 2013). However, experimentally only the exo-loop is observed to move in the available structures of the CD alone (Ossowski *et al.*, 2003; Textor *et al.*, 2012). Analysis of these loop regions in our new *C2* and *P6₃22* crystal forms may shed new light on the processive motion of the protein during its catalytic cycle. Model building and refinement are now in progress.

The authors would like to thank Ruslan Sanishvili (Nukri) and Tim Grüne for assistance with fluorescence analysis and single-wavelength anomalous dispersion during the 2013 APS data-collection workshop and CCP4 school. Funding for this work was provided in part by the NSF (award 1069091). Research at Oak Ridge National Laboratory was supported by the US Department of Energy, Office of Biological and Environmental Research, through the Genomic Science Program, under SFA Biofuels Contract FWP ERKP752, and through support for the Center for Structural Molecular Biology (CSMB) under Contract FWP ERKP291, using facilities supported by the Office of Basic Energy Sciences of the US Department of Energy. Oak Ridge National Laboratory is managed by UT-Battelle LLC for the US Department of Energy under Contract DE-AC05-00OR22725.

References

- Abuja, P. M., Pilz, I., Tomme, P. & Claeysens, M. (1989). *Biochem. Biophys. Res. Commun.* **165**, 615–623.
- Beckham, G. T. *et al.* (2010). *Biophys. J.* **99**, 3773–3781.
- Bu, L., Crowley, M. F., Himmel, M. E. & Beckham, G. T. (2013). *J. Biol. Chem.* **288**, 12175–12186.
- Colussi, F., Garcia, W., Rosseto, F. R., Mello, B. L. S., Oliveira Neto, M. & Polikarpov, I. (2012). *Eur. Biophys. J.* **41**, 89–98.
- Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G. & Haynes, C. A. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 12229–12234.
- Cruys-Bagger, N., Tatsumi, H., Ren, G. R., Borch, K. & Westh, P. (2013). *Biochemistry*, **52**, 8938–8948.
- Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K., Teeri, T. T. & Jones, T. A. (1994). *Science*, **265**, 524–528.

Figure 3
Crystal packing diagram of the *HjCel7A* CD. Top, *C2* crystal packing (*ac* face). Bottom, *P6₃22* crystal packing (*ab* face).

- Divne, C., Ståhlberg, J., Teeri, T. T. & Jones, T. A. (1998). *J. Mol. Biol.* **275**, 309–325.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Evans, B. R., Margalit, R. & Woodward, J. (1994). *Arch. Biochem. Biophys.* **312**, 459–466.
- French, S. & Wilson, K. (1978). *Acta Cryst.* **A34**, 517–525.
- Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
- Kraulis, J., Clore, G. M., Nilges, M., Jones, T. A., Pettersson, G., Knowles, J. & Gronenborn, A. M. (1989). *Biochemistry*, **28**, 7241–7257.
- Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.
- Ossowski, I. von, Ståhlberg, J., Koivula, A., Piens, K., Becker, D., Boer, H., Harle, R., Harris, M., Divne, C., Mahdi, S., Zhao, Y., Driguez, H., Claeysens, M., Sinnott, M. L. & Teeri, T. T. (2003). *J. Mol. Biol.* **333**, 817–829.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Payne, C. M., Resch, M. G., Chen, L., Crowley, M. F., Himmel, M. E., Taylor, L. E. II, Sandgren, M., Ståhlberg, J., Stals, I., Tan, Z. & Beckham, G. T. (2013). *Proc. Natl Acad. Sci. USA*, **110**, 14646–14651.
- Pingali, S. V., O'Neill, H. M., McGaughey, J., Urban, V. S., Rempe, C. S., Petridis, L., Smith, J. C., Evans, B. R. & Heller, W. T. (2011). *J. Biol. Chem.* **286**, 32801–32809.
- Sheldrick, G. M. (2010). *Acta Cryst.* **D66**, 479–485.
- Ståhlberg, J., Divne, C., Koivula, A., Piens, K., Claeysens, M., Teeri, T. T. & Jones, T. A. (1996). *J. Mol. Biol.* **264**, 337–349.
- Textor, L. C., Colussi, F., Silveira, R. L., Serpa, V., de Mello, B. L., Muniz, J. R. C., Squina, F. M., Pereira, N. Jr, Skaf, M. S. & Polikarpov, I. (2012). *FEBS J.* **280**, 56–69.
- Tomme, P., Van Tilbeurgh, H., Pettersson, G., Van Damme, J., Vandekerkhove, J., Knowles, J., Teeri, T. & Claeysens, M. (1988). *Eur. J. Biochem.* **170**, 575–581.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.