Journal of Structural Biology 172 (2010) 353–362

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Putting an N-terminal end to the *Clostridium thermocellum* xylanase Xyn10B story: Crystal structure of the CBM22-1–GH10 modules complexed with xylohexaose

Shabir Najmudin [a,b,*], Benedita A. Pinheiro [a], José A.M. Prates [a], Harry J. Gilbert [c], Maria J. Romão [b], Carlos M.G.A. Fontes [a,**]

[a] CIISA – Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal
[b] REQUIMTE, Departamento de Química, FCT-UNL, 2829-516 Caparica, Portugal
[c] Complex Carbohydrate Research Centre, University of Georgia, Athens, GA 30602-4712, USA

ARTICLE INFO

Article history:
Received 23 April 2010
Received in revised form 26 July 2010
Accepted 29 July 2010
Available online 1 August 2010

Keywords:
Xylanase
Carbohydrate-binding modules
Glycoside hydrolase
Cellulosome
Protein:protein interactions

ABSTRACT

In general, plant cell wall degrading enzymes are modular proteins containing catalytic domains linked to one or more non-catalytic carbohydrate-binding modules (CBMs). Xyn10B from *Clostridium thermocellum* is a typical modular enzyme containing an N-terminal family 22 CBM (CBM22-1), a family 10 glycoside hydrolase catalytic domain (GH10), a second CBM22 (CBM22-2), a dockerin sequence and a C-terminal family 1 carbohydrate esterase (CE1) catalytic domain. The structure of the N-terminal bi-modular CBM22-1–GH10 component of Xyn10B has been determined using a SeMet derivative by SAD to 2.5 Å. The data was extended to 2.0 Å for the non-SeMet mutant complexed with xylohexaose. CBM22-1–GH10 is a 60 kDa protein with an E337A mutation to render the GH10 subunit inactive. Three of the six xylose residues of xylohexaose are shown to be bound in the inactivated GH10 substrate binding cleft, with the other three sugars presumably disordered in the solvent channel. The protein is a dimer in the asymmetric unit with extensive surface contacts between the two GH10 modules and between the CBM22-1 and GH10 modules. Residues from helix H4 of the GH10 module provide the major contacts by fitting into the minor groove of the CBM22-1 module. The orientation of CBM22-1 is such that it would allow the substrate to be loosely bound and subsequently delivered to the active site in a processive manner.

1. Introduction

The complex and intricate structure of plant cell walls restrict the access of hydrolytic enzymes to their target substrates, primarily cellulose and hemicellulose. To overcome their limited accessibility to plant carbohydrates, microbial cellulases and hemicellulases have acquired complex molecular architectures generally comprising catalytic domains that are joined, via linker regions, to non-catalytic carbohydrate-binding modules (CBMs) (Bayer et al., 2004; Gilbert, 2007). The primary role of CBMs is to target the appended catalytic module to its substrate within the cell wall, thereby potentiating catalysis and reducing the accessibility constraints (Boraston et al., 2004). Carbohydrate modifying enzymes and their associated modules, which include CBMs, have been classified into sequence-based families in the CAZy database (Cantarel et al., 2009).

Xylan, the most abundant hemicellulose in plant cell walls, is a complex polysaccharide comprising a backbone of β1,4-xylose resides that are decorated with various sugars and acetyl groups (Waldron and Faulds, 2007). Endo-β1,4-xylanases (henceforth referred to as xylanases) catalyse the random hydrolysis of the xylan backbone. Xyn10B (formerly XynY; EC 3.2.1.8), from *Clostridium thermocellum*, is a multimodular bifunctional enzyme that displays xylanase and esterase activities. It is a component of the cellulosome, a plant cell wall degrading multienzyme complex that is presented on the surface of the bacterium. Xyn10B (1077 amino-acid residues; M_r = 120 kDa) comprises two family 22 CBMs (CBM22-1 and CBM22-2; formerly X6a and X6b, respectively), which flank the glycoside hydrolase family (GH) 10 xylanase module, a type I dockerin and a C-terminal family 1 carbohydrate esterase module (CE1) (Fontes et al., 1995b). CE1 is a ferulate esterase that hydrolyses the linkage between the aromatic acid and arabinoxylan (Tarbouriech et al., 2005; Prates

* Corresponding author at: CIISA – Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal. Tel.: +351 213652876; fax: +351 213652889.
** Corresponding author.
E-mail addresses: shabir@fmv.utl.pt (S. Najmudin), cafontes@fmv.utl.pt (C.M.G.A. Fontes).

1047-8477/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2010.07.009

et al., 2001). The crystal structures of various modules of Xyn10B have been determined. These include CBM22-2 (Charnock et al., 2000), the dockerin module in complex with a cognate cohesion (Carvalho et al., 2003) and CE1 (Prates et al., 2001). The CE1 esterase (residues 792–1077) displays an α/β-hydrolase fold with a classical Ser-His-Asp catalytic triad. The structure of a hydrolytically inactive mutant, S954A, in complex with ferulic acid demonstrated that the enzyme possesses specificity determinants for both the methoxy and hydroxyl ring substituents of the substrate (Tarbouriech et al., 2005). The dockerin module (residues 730–791), which is responsible for anchoring Xyn10B to the cellulosome, was solved in complex with a cohesin module of the *C. thermocellum* cellulosomal scaffolding protein, termed CipA. Both helices of the dockerin helix–turn–helix motif can interact independently with the cohesin β-sheet, revealing a general dual binding mode of the type I *Clostridium* dockerins to their target cohesins (Carvalho et al., 2003, 2007). CBM22-2 (residues 560–720) exhibits a classic β-jelly-roll fold that interacts strongly with xylans. The ligand-binding cleft contains three aromatic residues and two conserved polar residues that are critical to ligand recognition (Charnock et al., 2000). Trp 53, Tyr 103 and Glu 138 directly participate in xylan recognition, while Tyr 136 and Arg 25, although involved in ligand binding, also maintain the structural integrity of the cleft (Xie et al., 2001). Currently, the role of the N-terminal CBM22, CBM22-1, remains unknown (Dias et al., 2004). Recent work revealed that although CBM22-1 belongs to CBM22 family it does not appear to display carbohydrate-binding activity (Dias et al., 2004).

GH10 xylanases have a typical $(\beta/\alpha)_8$ barrel fold and hydrolyse glycosidic bonds with a net retention of anomeric configuration (Davies and Henrissat, 1995). Consistent with their endo-mode of action GH10 xylanases contain an extended substrate binding cleft that can accommodate between four to seven xylose residues. Each region that binds to a xylose moiety is known as a subsite, which can be given negative or positive numbers depending on whether they recognise sugars that are upstream or downstream of the active site (the −1 subsite)(Davies and Henrissat, 1995). Although GH10 xylanases are often located in modular enzymes, there is a paucity of information on the structural relationship been the modular components of these proteins. While the crystal structure of two GH10 xylanases appended to their respective CBMs have been reported, the structural flexibility of the connecting linker sequences prevented extrapolation of the orientation of the catalytic and non-catalytic modules seen *in crystal* to an *in vivo* setting (Fujimoto et al., 2004; Pell et al., 2004a). While it is possible that the orientation of CBMs and catalytic modules are, indeed, flexible, the biological role of non-catalytic modules that do not bind carbohydrates may require a direct interaction with their cognate catalytic module. Thus, the CBM22-1, which does not appear to recognise carbohydrates, and the GH10 of Xyn10B present a potentially fruitful model for exploring the interactions that may occur in multimodular enzymes.

In this study we probe the structure of the two N-terminal modules of Xyn10B, CBM22-1–GH10. The catalytic apparatus of the xylanase comprises Glu 337 (the catalytic acid/base) and Glu 480 (the nucleophile) (Fontes et al., 1995a). An inactive catalytic acid/base (E337A) mutant was prepared to visualise xylohexaose bound in the substrate binding cleft. Preliminary crystal characterisation of CBM22-1–GH10 E337A in complex with xylohexaose has been reported (Najmudin et al., 2008). In this study SeMet crystals of CBM22-1–GH10 E337A were used to solve the structure by Single- and Multiple-wavelength Anomalous Dispersion (SAD/MAD) X-ray diffraction experiments. The structure of the bi-modular protein provides insights into the structure of modular GH10 enzymes revealing a highly flexible linker region, but extensive contacts between the two protein modules.

2. Methods and materials

2.1. Protein expression and purification

Details on the bacterial strains, plasmids and growth condition used for preparation of CBM22-1–GH10 E337A are given in Najmudin et al. (2008). Final concentration of the purified CBM22-1–GH10 E337A was adjusted to 60 mg/ml.

2.2. Crystallisation and data collection

Crystals of seleno-L-methionine-containing (SeMet) and non-SeMet CBM22-1–GH10 E337A were grown by vapour-phase diffusion using the hanging drop method with an equal volume (1 μl) of protein (30 or 60 mg/ml in 5 mM DTT) and reservoir solution at 293 K containing 10 mM Xylohexaose. The crystallisation buffer for the SeMet crystals were: 0.1 M Na citrate tribasic, pH 5.6, 1.0 M $(NH_4)H_2PO_4$; 1 M Na acetate, 0.1 M HEPES pH 7.5 and 0.05 M $CdSO_4$; 0.2 M $(NH_4)_2HPO_4$, 20% PEG 3350, or 1.0 M KH_2PO_4 and that for the non-SeMet crystals were 1 M Na acetate, 0.1 M HEPES pH 7.5 and 0.05 M $CdSO_4$. Crystals grew over a period of 4–6 days for the SeMet and non-SeMet CBM22-1–GH10 E337A. The crystals were cryo-cooled in liquid nitrogen after soaking in fresh mother liquor containing 30% (v/v) glycerol as cryoprotectant for a few seconds. The data for the non-SeMet CBM22-1–GH10 E337A crystals were collected at a wavelength of 0.9340 Å at beamline ID14-EH1 at the ESRF (Grenoble, France) using a Quantum 4 CCD (ADSC). The SAD/MAD data for the SeMet crystals were collected at beamline ID29 at the ESRF (Grenoble, France) using a Quantum 315R CCD (ADSC). The SeMet data were collected at wavelengths of 0.97930 Å (inflection point, $f' = -10.38$, $f'' = 2.91$), 0.97910 Å (peak, $f' = -8.86$, $f'' = 5.24$) and 0.97560 Å (remote, $f' = -5.84$, $f'' = 4.11$). The crystals were delicate and generally suffered radiation damage after each data collection. A complete non-SeMet data set was collected to a nominal resolution of 1.90 Å on the beamline ID14-EH1. All data were processed using the programs MOSFLM (Leslie, 1992) and SCALA (Evans, 2006) from the CCP4 suite (Collaborative-Computational-Project-4, 1994). The CBM22-1–GH10 crystal belong to the trigonal space group ($P3_221$ or $P3_121$) according to the program POINTLESS (Evans, 2006) with cell constants 171–173, 171–173, 131–137 Å, which corresponds to a calculated Matthews coefficient of 4.75 $Å^3\ Da^{-1}$ and a solvent content of 74% consistent with a dimer in the asymmetric unit (Najmudin et al., 2008). Data collection and processing statistics are presented in Table 1.

2.3. Phasing, model building and refinement

At ESRF, the best SeMet crystal (X15) diffracted to a resolution of 2.5 Å on the beamline ID29 and the non-SeMet to 1.90 Å on the beamline ID14-EH1. Location of anomalous scatterers, phasing and density modification were carried out using SHELXD and SHELXE (Sheldrick, 2008). The SeMet crystals suffered severe radiation damage during the first data collection. Thus, only the SeMet peak data was used to determine the positions of 16/22 expected Se atoms in the dimer. The best correlation coefficient output by SHELXD was 55%. The pseudo-free correlation coefficient, contrast and connectivity figures of merit given by SHELXE for the correct heavy-atom enantiomer ($P3_221$) were 67.9%, 1.396 and 0.947, respectively, as opposed to 44.7%, 0.368 and 0.892, respectively, for the wrong hand ($P3_121$). A complete MAD dataset (X7) was eventually collected to a resolution of 2.75 Å on the beamline ID29 in which positions for 15/22 expected Se atoms in the dimer were located. The best correlation coefficient output by SHELXD was 55%. The pseudo-free correlation coefficient, contrast and con-