This is a repository copy of *Crystal structures of the GH18 domain of the bifunctional peroxiredoxin-chitinase CotE from Clostridium difficile.*

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/164520/

Version: Published Version

**Article:**
Whittingham, Jean L., Hanai, Shumpei, Brannigan, James A. et al. (6 more authors) (2020) Crystal structures of the GH18 domain of the bifunctional peroxiredoxin-chitinase CotE from Clostridium difficile. Acta crystallographica. Section F, Structural biology communications. pp. 241-249. ISSN 2053-230X

https://doi.org/10.1107/S2053230X20006147

**Reuse**
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
Crystal structures of the GH18 domain of the bifunctional peroxiredoxin–chitinase CotE from *Clostridium difficile*

Jean L. Whittingham, a Shumpei Hanai, b James A. Brannigan, a William T. Ferreira, c Eleanor J. Dodson, a Johan P. Turkenburg, a Jared Cartwright, d Simon M. Cutting c and Anthony J. Wilkinson**

CotE is a coat protein that is present in the spores of *Clostridium difficile*, an obligate anaerobic bacterium and a pathogen that is a leading cause of antibiotic-associated diarrhoea in hospital patients. Spores serve as the agents of disease transmission, and CotE has been implicated in their attachment to the gut epithelium and subsequent colonization of the host. CotE consists of an N-terminal peroxiredoxin domain and a C-terminal chitinase domain. Here, a C-terminal fragment of CotE comprising residues 349–712 has been crystallized and its structure has been determined to reveal a core eight-stranded β-barrel fold with a neighbouring subdomain containing a five-stranded β-sheet. A prominent groove running across the top of the barrel is lined by residues that are conserved in family 18 glycosyl hydrolases and which participate in catalysis. Electron density identified in the groove defines the pentapeptide Gly-Pro-Ala-Met-Lys derived from the N-terminus of the protein following proteolytic cleavage to remove an affinity-purification tag. These observations suggest the possibility of designing peptidomimetics to block *C. difficile* transmission.

1. Introduction

*Clostridium difficile* is an anaerobic, Gram-positive, spore-forming bacterium and an animal pathogen which has emerged as the most frequent cause of antibiotic-associated hospital-acquired diarrhoea (Smits et al., 2016). It is estimated that there are half a million new cases each year of infection with *C. difficile* (CDI), augmented by 75 000–175 000 cases of recurrent CDI. Treatment with antibiotics such as fidaxomicin cures 90% of new cases of CDI; however, recurrent infection is common and thousands of patients are on costly long-term antibiotic treatment, with more than 25 000 deaths per annum in the USA alone (Lessa et al., 2015). The disease follows antibiotic therapy, which disrupts the normal gut microflora, providing an opportunity for *C. difficile* colonization. The pathogen causes a range of clinical conditions in humans which range from mild diarrhoea to life-threatening pseudomembranous colitis, toxic megacolon and colonic perforation.

Spores play an important role in CDI as the agents of disease transmission. As robust and metabolically dormant structures, *C. difficile* spores can survive the dysbiosis of the gut microflora induced by antibiotic treatment. The resulting perturbation of the intestinal microbiota creates an environment in which the spore can germinate and proliferate as
vegetatively growing cells, leading to colonization of the gastrointestinal tract. Disease is primarily associated with the secretion of two inflammatory cytotoxins, TcdA and TcdB, which are responsible for tissue damage (Shen, 2012). These large toxins enter the cells of the colonic epithelium, where they glucosylate small GTPases, disrupting signalling in these cells and giving rise to cytopathic effects and cytotoxicity. Toxin synthesis and spore formation by C. difficile may take place simultaneously upon nutrient limitation, so that while non-sporulating cells in the population produce toxins to generate nutrients, the sporulating cells are designed for survival, dissemination and the initiation of new infections (Daou et al., 2019).

The spore has an elaborate multi-layered structure. Analysis of the proteins released from spores of C. difficile following treatment with SDS–borate–dithiothreitol identified a number of spore-coat enzymes, including superoxide dismutase and catalase (Permpoonpattana et al., 2011). Among the most interesting of the discovered proteins was CotE, an 81 kDa protein comprising an N-terminal domain homologous to 1-Cys peroxiredoxins, a central cysteine-rich interdomain region and a C-terminal domain with homology to family 18 glycosyl hydrolases (GH18). The corresponding enzyme activities were demonstrated for purified recombinant polypeptides comprising these domains (Permpoonpattana et al., 2013). The roles of these enzymatic activities in the spore are intriguing. Insertional mutagenesis showed that cotE is not essential for spore integrity (Permpoonpattana et al., 2013); however, in animal models of CDI, when CotE is absent the capacity of spores to colonize the intestine and induce virulence is markedly reduced (Hong et al., 2017). It was proposed that CotE facilitates host colonization by binding to the mucus layer of the intestine. This in turn implies that CotE may be a target for the prevention of or intervention in CDI. To build a greater understanding of structure–function relationships and to provide a platform for future inhibitor discovery, we have determined crystal structures of the chitinase domain of CotE.

2. Materials and methods

2.1. Macromolecule production

2.1.1. Cloning. The oligonucleotide primers CotE-349F and CotE-712R (Table 1) were used to amplify a 1119 bp fragment from a pET-28b plasmid derivative harbouring a cotE sequence codon-optimized for expression in Escherichia coli. The fragment produced by polymerase chain reaction (PCR) was inserted into the plasmid pETYSBLLIC3C (Fogg & Wilkinson, 2008) by the In-Fusion method (Clontech Laboratories) and the products were used to transform E. coli strain XL1-Blue. Plasmids from kanamycin-resistant colonies were prepared and their ‘cloned’ DNA inserts were sequenced to confirm the presence and authenticity of the expected inserts. The plasmid pETYSBLLIC3C-CotEC was used to transform E. coli BL21 (DE3) cells. It encodes residues 349–712 of CotE fused N-terminally to the sequence MGSSHHHHHHSSGLVLQPSMASVFTGVFSVTQVNLDSVPSQTVNSLYSLLQPGVQIGINDPEVLKEELEGIYGQF.

Table 1

| Source organism | C. difficile |
|-----------------|-------------|
| DNA source      | Synthetic codon-optimized sequence |
| Forward primer  | CotE-349F   |
| Reverse primer  | CotE-712R   |
| Cloning vector  | pETYSBLLIC3C |
| Expression vector | pETYSBLLIC3C |
| Expression host | E. coli BL21(DE3) |

MGSSHHHHHHSSGLVLQPSMASVFTGVFSVTQVNLDSVPSQTVNSLYSLLQPGVQIGINDPEVLKEELEGIYGQF comprising a human rhinovirus (HRV) 3C-cleavable hexahistidine tag (Table 1).

2.1.2. Protein purification. Recombinant protein was produced in E. coli BL21(DE3) cells. The cells were grown with shaking at 310 K in Luria–Bertani broth containing 30 mg ml⁻¹ kanamycin to an OD₆₀₀ of 0.6–0.8. Recombinant protein production was induced by the addition of 1 mM isopropyl β-D-thiogalactopyranoside and incubation at 289 K for 20 h. The cells were harvested by centrifugation, the pellet was resuspended in buffer A (50 mM Tris–HCl pH 7.5, 500 mM NaCl, 10 mM imidazole), to which an EDTA-free protease-inhibitor cocktail tablet (Roche Diagnostics, USA) had been added, and the cells were lysed by sonication on ice. The soluble cell extract was collected following centrifugation and CotE(349–712) was purified in three steps, each of which was carried out at room temperature. Firstly, the soluble cell supernatant was loaded onto a 5 ml nickel-charged HisTrap column (Amersham Pharmacia) equilibrated in buffer A. After washing, the column was developed with a 10–500 mM imidazole gradient in buffer A. Fractions containing CotE(349–712) were identified and pooled. The pooled sample (~2.5 mg ml⁻¹ protein concentration) was treated with a 1:50 ratio of HRV 3C protease (purified in-house) to cleave off the N-terminal tag with simultaneous dialysis overnight against buffer B (20 mM Tris–HCl pH 7.5, 150 mM NaCl). The cleavage products were passed over a second NiNTA agarose column equilibrated in buffer B. In this step, highly purified untagged protein was identified in the flow-through fractions, which were combined, concentrated by centrifugal ultrafiltration (Amicon Ultra) and passed through a Superdex S200 column in buffer B. After gel filtration, the molecular mass of the purified protein was measured by electrospray ionization mass spectrometry to be 17 339 Da, which is within 1 Da of the calculated mass of CotE(349–712) with an N-terminal Gly-Pro-Ala sequence representing a vestige of the cloning and proteolysis procedure.
Composition of reservoir solution 200 m
well diffracting crystals from drops that were formed by these conditions in a hanging-drop format led to the growth of under conditions that contained PEG 3350. Refinement of Clear Strategy Screens I and II (Brzozowski & Walton, 2001) brated against 0.1 ml reservoir solution. Crystals of CotE(349–712), which belonged to space group \( \text{P}1_{n} \) and were equilibrated against 1 ml reservoir solution (Table 2). These crystals belonged to space group \( \text{P}6_{1}2_{1}2_{1} \), and are referred to as the monoclinic crystal form.

### Table 2

| Crystal | CotE(349–712), monoclinic | CotE(349–712), hexagonal |
|---------|---------------------------|--------------------------|
| Method  | Vapour diffusion in hanging drops | Vapour diffusion in hanging drops |
| Plate type | 24-well | 24-well |
| Temperature (K) | 291 | 291 |
| Protein concentration (mg ml\(^{-1}\)) | 20 | 20 |
| Buffer composition of protein solution | 20 mM Tris–HCl pH 8.0, 150 mM NaCl | 50 mM Tris–HCl pH 7.5, 500 mM NaCl, 10 mM imidazole |
| Composition of reservoir solution | 200 mM ammonium phosphate, 22.5% PEG 3350 | 1 ml 0.1 M sodium malonate pH 5.5, 13% PEG 3350 and 3 μl C\(_6\)H\(_5\)E\(_6\) |
| Volume and ratio of drop | 3 μl (1 μl protein solution + 2 μl reservoir solution) | 5 μl (2 μl protein solution + 3 μl reservoir solution) |
| Volume of reservoir (ml) | 1 | 1 |

† The mean \( I/\sigma(I) \) falls below 2 at a resolution of 2.2 Å. The data yielded valuable information as shown by CC\(_{1/2}\).

### 2.2. Crystallization

Protein concentrations were determined with an Epoch Microplate Spectrophotometer using an extinction coefficient at 280 nm calculated from the sequence. Crystallization experiments were set up as sitting drops in 96-well plates using Hydra 96 and Mosquito liquid-handling systems to dispense the reservoir and drop solutions, respectively. Drops consisting of 150 μl 20 mg ml\(^{-1}\) protein solution in 20 mM Tris–HCl pH 8.0, 150 mM NaCl and 150 μl reservoir solution were equilibrated against 0.1 ml reservoir solution. Crystals of CotE(349–712) were obtained following screening experiments with Clear Strategy Screens I and II (Brzozowski & Walton, 2001) under conditions that contained PEG 3350. Refinement of these conditions in a hanging-drop format led to the growth of well diffracting crystals from drops that were formed by mixing 1 μl protein solution with 2 μl reservoir solution composed of 200 mM ammonium phosphate, 22.5% PEG 3350 and were equilibrated against 1 ml reservoir solution (Table 2). These crystals belonged to space group \( \text{P}2_{1} \), and are referred to as the monoclinic crystal form.

Following the discovery of electron density in the putative active-site region, we sought to remove the co-purifying ligand. For this purpose, we immobilized the His\(_{6}\)-tagged CotE(349–712) on a Ni\(^{2+}\)-chelation column in buffer A and washed the column extensively with buffer A containing 2 M guanidine hydrochloride followed by a 2–0 M descending gradient of guanidine hydrochloride in buffer A. The polylhistidine tag was subsequently removed from the protein as described above. Crystallization experiments produced diffracting crystals from hanging drops formed by mixing 2 μl 20 mg ml\(^{-1}\) CotE(349–712) with 3 μl well solution consisting of 1 ml 0.1 M sodium malonate pH 5.5, 13% PEG 3350 and 3 μl pentaethylene glycol monooctyl ether (C\(_{6}\)E\(_{5}\)) (Table 2). These crystals belonged to space group \( \text{P}6_{1}2_{1}2_{1} \) and are referred to as the hexagonal crystal form.

### 2.3. Data collection and processing

Single crystals of CotE(349–712) were captured from crystallization drops in fine nylon loops and transferred to a solution of mother liquor containing 15% (v/v) glycerol prior to cryocooling in liquid nitrogen. Crystals were tested on an in-house system and the best diffracting crystals were chosen and sent to Diamond Light Source (DLS) for data collection to high resolution. For the monoclinic and hexagonal crystal forms, data were collected on beamlines I02 and I03, respectivelv. The data were processed using the 3dii option in xia2 (Winter, 2010) and extended to 1.3 and 2.2 Å resolution, respectively, for the monoclinic and hexagonal crystal forms (Table 3). For both crystal forms, there is one protein molecule per asymmetric unit.

### 2.4. Structure solution and refinement

The structures of CotE(349–712) were determined and refined using the CCP4 suite of software as implemented in the CCP4i2 graphical interface (Potterton et al., 2018). The structure of the monoclinic crystal form was solved by molecular replacement in Phaser (McCoy et al., 2007) using the coordinate set for the catalytic domain of chitinase A1 (ChiA1) from Bacillus circulans (PDB entry 1tx; Matsumoto et al., 1999) as the search model. The proteins share 35% sequence identity across 353 and 468 aligned residues of CotE and ChiA1, respectively. The resulting model was refined in REFMACS (Murshudov et al., 2011) and automatic model building was then performed using Buccaneer (Cowtan, 2006). This was followed by iterative rounds of manual model...
building in Coot (Emsley et al., 2010) and refinement in REFMACS. The refinement statistics are summarized in Table 4. The structure of the hexagonal crystal form was solved by molecular replacement using the refined coordinates of the structure of the monoclinic form.

3. Results and discussion

3.1. Determination of the structure of the CotE chitinase domain

We initiated structural studies of CotE by generating a series of expression constructs encoding C-terminal fragments of different lengths fused to a cleavable polyhistidine tag. This allowed us to identify CotE(349–712) as a stable, soluble fragment. The fragment encompasses the chitinase domain, which is predicted to span residues 380–685. This protein was purified from overproducing E. coli by immobilized nickel-affinity chromatography, cleavage to remove the affinity tag and gel-filtration chromatography. The protein was first crystallized from polyethylene glycol 3350-containing solutions. The crystals belonged to space group $P_2_1$, with a single protein molecule in the asymmetric unit and a solvent content of 48%. Data extending to 1.3 Å resolution were collected at the Diamond Light Source synchrotron-radiation source and the structure was solved by molecular replacement using the refined coordinates of different lengths fused to a cleavable polyhistidine tag. This allowed us to identify CotE(349–712) as a stable, soluble fragment. The fragment encompasses the chitinase domain, which is predicted to span residues 380–685. This protein was purified from overproducing E. coli by immobilized nickel-affinity chromatography, cleavage to remove the affinity tag and gel-filtration chromatography. The protein was first crystallized from polyethylene glycol 3350-containing solutions. The crystals belonged to space group $P_2_1$, with a single protein molecule in the asymmetric unit and a solvent content of 48%. Data extending to 1.3 Å resolution were collected at the Diamond Light Source synchrotron-radiation source and the structure was solved by molecular replacement using the coordinates of the catalytic domain of chitinase A1 from B. circulans (PDB entry 1lx; Matsumoto et al., 1999) as the search model. The CotE(349–712) structure has been refined to give a crystallographic $R$ value of 10.6% ($R_{\text{free}} = 13.3\%$) for a model comprising residues Ile363–Phe712, 490 waters and a pentapeptide-like entity (currently modelled as Gly-Pro-Ala-Met-Lys) defined by residual electron density located in the active-site region of the structure.

3.2. Structure description

CotE(349–712) has a classical parallel eight-stranded $\alpha\beta$-barrel at its core, on top of which the substrate-binding site resides (Fig. 1). The $(\alpha\beta)_8$-barrel is compact, with the exception of significant insertions following strands $\beta$6 and $\beta$7. Following $\beta$7, residues 605–655 form a distinct subdomain comprising a five-stranded $\beta$-sheet and an $\alpha$-helix which packs across one face of this sheet. The opposite face of the sheet packs against the meandering segment of the polypeptide following strand $\beta$6, which spans residues 550–580. Much of the structure is covalently closed through a disulfide bond linking residues Cys376 and Cys670 (Fig. 1).

A search of the Protein Data Bank for structures similar to the CotE chitinase domain revealed a plethora of coordinate sets with $Q$-scores greater than 0.5, corresponding to r.m.s.d. values in the range 1.3–1.4 Å over 300 or so C$^\alpha$ atoms of matched residues. These included human chitinase/chitotriosidase (PDB entry 1lg2; Fusetti et al., 2002), the mouse lectin YM1 (PDB entry 1v18; Tsai et al., 2004) and bacterial chitinases from Serratia proteamaculans (PDB entry 4lgx; Madhuprakash et al., 2015) and Klebsiella pneumonia (PDB entry 3qok; Midwest Center for Structural Genomics, unpublished work) variously in complexes with monosaccharides, disaccharides and oligosaccharides as well as oligosaccharide and peptide-based inhibitors.

3.3. The substrate-binding and active site

A pronounced groove runs across the top of the molecule when viewed in the orientation shown in Fig. 1(c). In family GH18 enzymes, this forms the substrate-binding and active site. As seen in the structures of other GH18 family members, this groove has a markedly negative electrostatic potential (Figs. 1c and 1d). Located in this groove, we observed a strong positive feature in the electron-density maps that we were able to model as a pentapeptide: Gly-Pro-Ala-Met-Lys (Fig. 2a). There is uncertainty for the side chain of residue 5, but the sequence otherwise corresponds to the first five residues of the expected product of HRV 3C protease cleavage of the recombinant fusion protein. The Gly-Pro-Ala segment is a vestige of the cleavage-recognition sequence, with the Met-Lys segment constituting residues 349 and 350 of CotE. The binding site for the peptide is circumscribed by residues Met550, Tyr485, Glu484, Trp445, Asp553, Ala556, Arg608, Thr614, Thr616, Tyr606, Tyr681 and Trp677, forming a protein–peptide interface of 362 Å$^2$ (Fig. 2b).

The $\alpha$-amino group of the peptide makes an ion-pairing interaction with the side-chain carboxylate of Glu484 and a charge–dipole interaction with the phenolic hydroxyl group of Tyr485, as well as polar interactions with surrounding water molecules, one of which forms a bridging polar interaction with the carboxylate of Asp553 (Fig. 2b). The face of the

### Table 4

| Structure refinement. |
|-----------------------|
| Crystal | CotE(349–712), monoclinic | CotE(349–712), hexagonal |
| Resolution range (Å) | 45.09–1.30 (1.33–1.30) | 54.37–2.10 (2.15–2.10) |
| Completeness (%) | 97.89 (96.11) | 99.96 (99.93) |
| $\sigma$ Cutoff | None | None |
| No. of reflections | | |
| Working set | 89793 (6466) | 37274 (2679) |
| Test set | 4557 (328) | 1918 (154) |
| Final $R_{\text{cryst}}$ | 0.106 (0.119) | 0.22 (0.41) |
| Final $R_{\text{free}}$ | 0.133 (0.158) | 0.27 (0.43) |
| Cruickshank DPI | 0.039 | |
| No. of non-H atoms | | |
| Protein | 2843 | 2699 |
| Additives | 53 | 14 |
| Ligand | 33 | — |
| Water | 490 | 138 |
| Total | 3419 | 2851 |
| R.m.s. deviations | | |
| Bonds (Å) | 0.020 | 0.016 |
| Angles (°) | 2.47 | 1.78 |
| Average $B$ factors (Å$^2$) | | |
| Protein | 13.5 | 58.8 |
| Additives | 33.0 | 90.3 |
| Ligand | 28.2 | — |
| Water | 29.6 | 64.0 |
| Ramachandran plot | | |
| Favoured regions (%) | 90.1 | 87.5 |
| Additionally allowed (%) | 9.9 | 11.8 |
| Outliers (%) | 0 | 0.3 |
pyrrolidine ring of Pro2 and the residue 1–2 peptide bond pack closely with the face of the indole ring of Trp445. The carbonyl O atom of Pro2 makes an intramolecular hydrogen bond to the amide >N–H of Lys5, while the amide >N–H of Ala3 makes a charge–dipole interaction with the side-chain carboxylate of Asp553 (Fig. 2b). Further along the ligand backbone, the carbonyl O atom of Ala3 forms another charge–dipole interaction with the guanidino group of Arg608. The backbone >N–H of Met4 forms a hydrogen bond to a well ordered water molecule, with bridging hydrogen bonds to the indole N atom of Trp677 and the phenolic hydroxy group of Tyr681. The side chain of Met4 projects into a pocket lined by the side chains of the protein residues Tyr606 and Tyr681 and the aliphatic faces of Thr614 and Thr616 (Fig. 2b). The electron density becomes more diffuse at Lys5, precluding further model building.

The peptide residing in the binding groove may be an N-terminal degradation product of proteolysis. This would imply high-affinity binding, since the peptide is evidently retained on passage through a gel-filtration column. Alternatively, this pentapeptide may represent the amino-terminus of the intact CotE chitinase-domain polypeptide. If this is the case, then the intervening residues Thr351–Thr362, which are not visible in the electron-density maps, are missing owing to disorder. Lys5 of the peptide and Ile363, the first residue of the polypeptide defined by the electron-density maps, are on opposite faces of the chitinase domain (Fig. 1a) and 45 Å apart, so it is unlikely that they belong to the same molecule. The same lysine and Ile363 in the protein molecule generated by the symmetry operation (−x, y − 1/2, −z) are separated by 24 Å, a distance that could be spanned by the 12 ‘missing’ residues.

3.4. Mechanistic considerations

Chitinases are enzymes that catalyse glycosidic bond cleavage in β-1,4-linked N-acetylglucosamine (GlcNAc)-containing polymers such as chitin. Chitin is widely distributed in nature, but the physiological substrate of CotE is not known. Polymeric structures containing GlcNAc are present in the cell wall of C. difficile, a structure that is known to undergo remodelling during spore formation. The integrity of
C. difficile spores, however, is unaffected by deletion of cotE, suggesting that the substrate may be host-derived. Chitin does not occur in mammals, suggesting that the target of CotE action during infection is a glycoprotein, and evidence has been presented to show that CotE facilitates spore binding to mucin and mucin degradation (Hong et al., 2017).

The family 18 glycosyl hydrolases carry a conserved acidic motif, which occurs in CotE as D⁴⁷⁷GIDIDWEY⁴⁸⁵. As shown in Fig. 2(b), Glu484 and Tyr485 form polar contacts to the α-amino group of the bound peptide in CotE. The reaction mechanisms of these enzymes feature neighbouring-group assistance, in which the carbonyl of the acetyl group on the

Figure 2
Peptide–CotE interactions. Stereo images are shown with the pentapeptide ligand in ball-and-stick format. (a) The peptide with its associated electron density in maps calculated with coefficients 2mFo – DFo and contoured at 1σ. Residues are labelled using the single-letter code. (b) The pentapeptide ligand displayed as above, with surrounding protein residues in cylinder format (with three-letter code labels). Atoms are coloured by element type (O, red; N, blue; S, yellow; C in green for the ligand or light grey for the protein). Neighbouring water molecules are shown as red spheres. Polar interactions are denoted by dashed lines. (c) Superposition of the methylallosamidin ligand (blue C atoms) from the complex with human chitinase (PDB entry 1hkj; Rao et al., 2003) and the GPAMK ligand (green C atoms) from CotE. Selected conserved residues are shown and numbered as in CotE. The structures were displayed in CCP4mg (McNicholas et al., 2011) following SSM superposition, which gave a r.m.s.d. of 1.4 Å for 292 matching atoms. (d) The argadin (white; PDB entry 1waw; Rao et al., 2005), methylallosamidin (blue; PDB entry 1hkj; Rao et al., 2003) and chitobiose (coloured by atoms with C in grey and O in red; PDB entry 4wkh; Fadel et al., 2015) ligands from human chitinase structures superposed on the GPAMK ligand from CotE. The binding subsites −3, −2 and −1 are labelled.
GlcNAc in the −1 site acts as the nucleophile (Terwisscha van Scheltinga et al., 1995; Tews et al., 1997). An Asp residue (corresponding to Asp482 in CotE) stabilizes the developing positive charge on the −NH of the acetyl group, while a glutamic acid (corresponding to Glu484) promotes breakage of the β-1,4 glycosidic linkage between residues bound in the −1 and +1 subsites by protonating the leaving group. This results in the formation of an oxazolinium intermediate in the −1 site, the positive charge of which is stabilized by neighbouring carboxylates. A water molecule, activated by the Glu residue now serving as a base, attacks at the anomeric C atom with opening of the oxazolinium ring and reformation of the N-acetyl group. As a result, there is an overall retention of configuration at the anomeric C atom (Terwisscha van Scheltinga et al., 1995).

There has been much research into chitinase inhibition, since chitinase inhibitors have potential applications in the treatment of human diseases, including those resulting from bacterial infections (Frederiksen et al., 2013). As a result, many classes of chitinase inhibitor have been discovered, including sugar derivatives such as the natural product allo-samidin and derivatives thereof (Sakuda et al., 1986; MacDonald et al., 2010). The structure of the CotE–peptide complex is compared with that of human macrophage chitinase bound to methylallosamidin (Rao et al., 2003) in Fig. 2(c). Following structure superposition by the SSM procedure as implemented in CCP4mg (McNicholas et al., 2011), the r.m.s.d. for 292 matching atoms is 1.4 Å. Methylallosamidin, which consists of two β-1,4-linked N-acetylglucosamine residues attached to allosamizoline, binds in the −3 to −1 subsites in the chitinase substrate-binding site. Allosamizoline is bicyclic, consisting of a cyclopentitol ring fused to an oxazoline, which mimics the oxazolinium intermediate in the chitinase-catalysed reaction. As seen in Fig. 2(c), the peptide-binding site in CotE partially overlaps with the allosamidin-binding site in human chitinase, particularly in the −3 and −2 subsites.

The cyclic pentapeptides argifin and argadin are of interest in relation to the peptide bound in the substrate-binding site of the chitinase domain of CotE. Argifin and argadin are produced by fungi and have a high potency (nanomolar) towards their insect chitinase targets (Shiomi et al., 2000). Structures have been determined of these inhibitors in complex with bacterial and human chinatines, revealing mimicry of the natural carbohydrate substrate (Rao et al., 2005). Relative to the GPAMK peptide bound to CotE, the cyclic peptide argadin binds more deeply in the substrate-binding cleft. In Fig. 2(d), the argadin (white), methylallosamidin (blue) and chitobiose (grey) ligands from human chitinase complexes are overlaid on the GPAMK peptide (green) from the CotE complex. Argadin occupies subsites −2 and −1, defined by the binding of the chitobiose (GlcNAc₂), and extends towards the +1 site to make interactions with the

---

Figures are not transcribed, but they illustrate the domain swapping in the chitinase domain of CotE. (a) Ribbon rendering of the chitinase domain in the P6_22 crystal structure. The chain is colour-ramped as in Fig. 1(a), and the C\(^\text{\text{α}}\) atoms and side chains of cysteines 376 and 670 are shown as spheres. The C-terminal residues 627–712 extend away from the β-barrel so as to pack onto and complete the β-barrel of a crystallographic symmetry mate, as shown in (c), where the two subunits of the domain-swapped dimer are coloured light blue and green, respectively. The eighth strand of each β-barrel is provided by the partner subunit. (b) A juxtaposition of the C-terminal swapped domains (residues 627–712) in the P_2_1 and P6_22 crystal forms, coloured light green and ice blue, respectively, relative to the β-barrel domain (in white) following the superposition of residues 358–623 is shown.
catalytic residues (Fadel et al., 2015). In its complex, the argadin conformation is stabilized by extensive intramolecular polar interactions. In contrast, there are few polar interactions with the protein; instead, there is quite an extensive buried surface area which will contribute to higher affinity.

3.5. Domain swapping

Before we confirmed the identity of the peptide ligand in the chitinase-domain substrate-binding groove, we sought to remove this ‘ligand’ and crystalize the protein in an unliganded form. To obtain unliganded CotE(349–712), the first Ni–NTA column-chromatography purification procedure was modified so that after the binding and washing steps, the column was washed with buffer C (buffer A containing 2 M guanidine hydrochloride) to partially unfold the immobilized protein and allow the dissociation of endogenous ligand(s). We have used this technique previously to remove ligands from peptide-binding proteins (Hughes et al., 2019). After washing with ten column volumes of buffer C, the concentration of guanidine hydrochloride was decreased to zero in a series of five steps. The protein was then eluted from the column by applying an increasing imidazole concentration gradient. The eluted recombinant protein was subsequently cleaved with HRV 3C protease and further purified as before.

In retrospect this strategy was flawed, since the peptide ligand GPAMK is presumably generated during or after the HRV 3C protease-cleavage step. The partially unfolded/refolded protein was crystallized in a different crystal form (P61,22). Solution of the structure and refinement against data extending to ~2.1 Å resolution revealed that the substrate-binding groove was indeed empty. However, the structure proved to be of a 3D domain-swapped dimer. In this structure (Fig. 3), domain exchange takes place at residue Lys627. As is usual following 3D domain swapping, the respective domains in the two structures can be closely superposed. Thus, 244 Cα atoms in the residue range 368–623 can be superposed with an r.m.s.d. of 0.44 Å, while similarly 78 Cα atoms of residues 627–708 can be superposed with an r.m.s.d. of 0.43 Å.

3D domain swapping is commonly observed in protein crystal structures since it is promoted by high protein concentrations and/or partially denaturing conditions (Schlunegger et al., 1997). Indeed, we have observed this phenomenon in crystals of other sporulation proteins. In Spo0A from B. subtilis (Lewis et al., 2000) and CodY from C. difficile (Daou et al., 2019) α-helices are swapped, leading to dimeric and hexameric assemblies, respectively, while in SpoIE from B. subtilis (Levdikov et al., 2012) three β-strands from a β-sandwich are exchanged in a 3D domain-swapped dimer. In almost all instances 3D domain swapping is a crystallographic artefact and there is no evidence that the 3D domain-swapped dimer of CotE(349–712) is physiologically significant. It is nevertheless structurally interesting since it involves (i) the exchange of a strand, β8, from the β-barrel and (ii) the breakage of the intramolecular disulphide bond between Cys376 and Cys670 and the formation of an intramolecular equivalent.

3D domain swapping in the TIM barrel of a subunit of tryptophan synthase (TrpA) from Streptococcus pneumoniae was recently observed (Michalska et al., 2020). In TrpA, the N-terminal strands 1 and 2 of the TIM barrel are exchanged in dimer formation. In this paper, the authors surveyed 2000 or so TIM-barrel structures in the PDB and found that the integrity of the core barrel was always preserved. They concluded that their TrpA structure is the first example of 3D domain swapping that intrudes into the core of the barrel (Michalska et al., 2020). The 3D domain-swapped CotE(349–712) structure reported here is therefore the second example, with the distinction that it is the C-terminal strand 8 of the barrel that is swapped.

Acknowledgements

We are grateful to Professor Tadayuki Uno for facilitating this work and to Sam Hart for technical assistance. This work was carried out with the support of Diamond Light Source.

Funding information

Funding for this research was provided by: Biotechnology and Biological Sciences Research Council (grant No. BB/J007900/1 to AJW); Cross-Boundary Innovation Program of Osaka University (scholarship to Shumpei Hanai).

References

Brzozowski, A. M. & Walton, J. (2001). J. Appl. Cryst. 34, 97–101.
Cowtan, K. (2006). Acta Cryst. D62, 1002–1011.
Daou, N., Wang, Y., Levdiakov, V. M., Nandanakumar, M., Livny, J., Bouillaut, L., Blagova, E., Zhang, K., Belitsky, B. R., Rhee, K., Wilkinson, A. J., Sun, X. & Sonenshein, A. L. (2019). PLoS One, 14, e0206896.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Acta Cryst. D66, 486–501.
Fadel, F., Zhao, Y., Cachau, R., Cousido-Siah, A., Ruiz, F. X., Harlos, K., Howard, E., Mitschler, A. & Podjarny, A. (2015). Acta Cryst. D71, 1455–1470.
Fogg, M. J. & Wilkinson, A. J. (2008). Biochem. Soc. Trans. 36, 771–775.
Frederiksen, R. F., Paspaalari, D. K., Larsen, T., Storgaard, B. G., Larsen, M. H., Ingmer, H., Palcic, M. M. & Leisner, J. J. (2013). Microbiology, 159, 833–847.
Fusetti, F., von Moeller, H., Houston, D., Rozeboom, H. J., Dijkstra, B. W., Boot, R. G., Aerts, J. M. & van Aalten, D. M. F. (2002). J. Biol. Chem. 277, 25537–25544.
Hong, H. A., Ferreira, W. T., Hosseini, S., Anwar, S., Hitri, K., Wilkinson, A. J., Vahlen, W., Zentek, J., Soloviev, M. & Cutting, S. M. (2017). J. Infect. Dis. 216, 1452–1459.
Hughes, A., Wilson, S., Dodson, E. J., Turkenburg, J. P. & Wilkinson, A. J. (2019). Acta Cryst. F75, 246–253.
Lessa, F. C., Mu, Y., Bamberg, W. M., Beldavs, Z. G., Dumyati, G. K., Dunn, J. R., Farley, M. M., Holzbauer, S. M., Meek, J. I., Phipps, E. C., Wilson, L. E., Winston, L. G., Cohen, J. A., Limbago, B. M., Fridkin, S. K., Gerding, D. N. & McDonald, L. C. (2015). N. Engl. J. Med. 372, 825–834.
Levdikov, V. M., Blagova, E. V., Rawlings, A. E., Jameson, K., Tunaley, J., Hart, D. J., Barak, I. & Wilkinson, A. J. (2012). J. Mol. Biol. 415, 343–358.
Lewis, R. J., Muchová, K., Brannigan, J. A., Barák, I., Leonard, G. & Wilkinson, A. J. (2000). J. Mol. Biol. 297, 757–770.
Macdonald, J. M., Tarling, C. A., Taylor, E. J., Dennis, R. J., Myers, D. S., Knapp, S., Davies, G. J. & Withers, S. G. (2010). Angew. Chem. Int. Ed. 49, 2599–2602.
Madhuprakash, J., Bobbili, K. B., Moerschbacher, B. M., Singh, T. P., Swamy, M. J. & Podile, A. R. (2015). Sci. Rep. 5, 15657.
Matsumoto, T., Nonaka, T., Hashimoto, M., Watanabe, T. & Mitsui, Y. (1999). Proc. Jpn Acad. B Phys. 75, 269–274.
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.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. M. (2011). Acta Cryst. D67, 386–394.
Michalska, K., Kowiel, M., Bigelow, L., Endres, M., Gilski, M., Jaskolski, M. & Joachimiak, A. (2020). Acta Cryst. D76, 166–175.
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.
Permpoonpattana, P., Phetcharaburanin, J., Mikelsone, A., Dembek, M., Tan, S., Brisson, M. C., La Ragione, R., Brisson, A. R., Fairweather, N., Hong, H. A. & Cutting, S. M. (2013). J. Bacteriol. 195, 1492–1503.
Permpoonpattana, P., Tolls, E. H., Nadem, R., Tan, S., Brisson, A. & Cutting, S. M. (2011). J. Bacteriol. 193, 6461–6470.
Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M., Pannu, N. S., Roth, C., Sheldrick, G., Skubak, P., Turkemberg, J., Uski, V., von Delft, F., Waterman, D., Wilson, K., Winn, M. & Wojdyr, M. (2018). Acta Cryst. D74, 68–84.
Rao, F. V., Houston, D. R., Boot, R. G., Aerts, J. M., Hodkinson, M., Adams, D. J., Shiomi, K., Ōmura, S. & van Aalten, D. M. F. (2005). Chem. Biol. 12, 65–76.
Rao, F. V., Houston, D. R., Boot, R. G., Aerts, J. M., Sakuda, S. & Van Aalten, D. M. F. (2003). J. Biol. Chem. 278, 20110–20116.
Sakuda, S., Isogai, A., Matsumoto, S., Suzuki, A. & Koseki, K. (1986). Tetrahedron Lett. 27, 2475–2478.
Schlunegger, M. P., Bennett, M. J. & Eisenberg, D. (1997). Adv. Protein Chem. 50, 61–122.
Shen, A. (2012). J. Innate Immun. 4, 149–158.
Shiomi, K., Arai, N., Iwai, Y., Turberg, A., Kölbl, H. & Ōmura, S. (2000). Tetrahedron Lett. 41, 2141–2143.
Smits, W. K., Lyras, D., Lacy, D. B., Wallace, M. H. & Kuijper, E. J. (2016). Nat. Rev. Dis. Primers, 2, 16020.
Terwisscha van Scheltinga, A. C., Armand, S., Kalk, K. H., Isogai, A., Henrissat, B. & Dijkstra, B. W. (1995). Biochemistry, 34, 15619–15623.
Tews, I., Terwisscha van Scheltinga, A. C., Perrakis, A., Wilson, K. S. & Dijkstra, B. W. (1997). J. Am. Chem. Soc. 119, 7954–7959.
Tsai, M.-L., Liaw, S.-H. & Chang, N.-C. (2004). J. Struct. Biol. 148, 290–296.
Winter, G. (2010). J. Appl. Cryst. 43, 186–190.