Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and atomic resolution X-ray diffraction of the catalytic domain of the large sialidase, nanI, from *Clostridium perfringens*

Simon Newstead,[a] Chin-Hsiang Chien,[b] Margaret Taylor[a] and Garry Taylor[a]*

[a]Centre for Biomolecular Sciences, University of St Andrews, St Andrews, Fife KY16 9UA, Scotland, and [b]Institute of Biochemistry, School of Life Science, National Yang-Ming University, Shih-Pai, Taipei, Taiwan

Correspondence e-mail: glt2@st-andrews.ac.uk

Received 8 July 2004
Accepted 5 September 2004

Sialidases catalyse the removal of terminal sialic acids from a range of glycoproteins, glycolipids and oligosaccharides. They have been found in bacteria, viruses and parasites, where they play important roles in pathogenesis and/or microbial nutrition, and in mammalian cells, where they modulate cell-surface glycosylation associated with a range of cellular activities. *Clostridium perfringens*, a causative agent of gas gangrene and peritonitis in humans, possesses three sialidases: nanH, nanI and nanJ, with molecular weights of 42, 77 and 129 kDa, respectively. The two larger enzymes are secreted by the bacterium and are involved in the pathogenesis and nutrition of *Clostridium*. As part of a study to examine the structures of all three enzymes, crystallization of the 77 kDa nanI isoenzyme was attempted. The expressed full-length protein was found to degrade easily; a stable 50 kDa catalytic domain was therefore subcloned. This domain was overexpressed in *Escherichia coli* and produced crystals belonging to space group $P2_12_12_1$, with unit-cell parameters a = 96.98, b = 69.41, c = 72.69 Å and one monomer per asymmetric unit. The crystals diffract to at least 0.92 Å. A molecular-replacement solution was obtained using the catalytic domain of the sialidase from the leech *Macrobdella decora*.

1. Introduction

Sialidases, or neuraminidases, catalyse the removal of terminal sialic acid sugars from a variety of glycoconjugates and play important roles in pathogenesis, bacterial nutrition and cellular interactions (Corfield, 1992). A number of structural studies have been carried out on sialidases from bacteria (Crennell *et al.*, 1993, 1994; Gaskell *et al.*, 1995), viruses (Crennell *et al.*, 2000; Varghese *et al.*, 1983) and eukaryotic sources (Buschiazzo *et al.*, 2000, 2002; Luo *et al.*, 1998). The sialidases from the influenza virus and the paramyxovirus hemagglutinin-neuraminidase have also been used for successful structure-based drug design (Alymova *et al.*, 2004; von Itzstein *et al.*, 1993). The structures of the catalytic domain of sialidases studied to date have a conserved six-bladed β-propeller topology (Taylor, 1996). The non-viral sialidases possess sequence motifs that are the signature of a sialidase: one to five 'Asp-boxes' with sequence S/T-*x*-D-(*x*)-G-*x*-T-W/F and an RIP or RLP motif containing one of the three catalytic arginines (Roggentin *et al.*, 1989). Other than these motifs, the sialidases have very diverse sequences, with many non-viral sialidases having additional carbohydrate-recognizing domains which may serve to increase the catalytic efficiency of the enzymes (Thobhani *et al.*, 2003). The catalytic mechanism of the influenza virus sialidase has been studied (Chong *et al.*, 1992), although catalytic intermediates have not been trapped in the crystal as yet. A recent study of the trans-sialidase

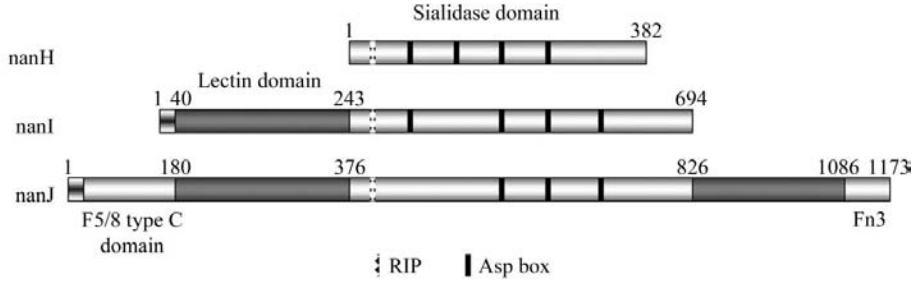

Figure 1
Schematic view of the three *C. perfingens* sialidases. The locations of the RIP and Asp-box motifs are indicated. Both nanH and nanI have a signal peptide at the N-terminus.

from *Trypanosoma cruzi* trapped a covalent intermediate by use of a fluorinated substrate analogue (Amaya *et al.*, 2004; Watts *et al.*, 2003), suggesting that the formation of such a covalent intermediate may be a common step in the mechanism of all sialidases.

The sequencing of the complete *Clostridium perfringens* genome revealed three sialidases, two of which had been previously characterized: the *nanH* gene product with a molecular weight of 43 kDa, which is not secreted, and the *nanJ* gene product, which has a molecular weight of 77 kDa and is secreted (Roggentin *et al.*, 1995). These two sialidases have been studied extensively and shown to exhibit very different kinetic and biochemical properties (Kruse *et al.*, 1996; Traving *et al.*, 1994). The third putative sialidase, nanJ, has a predicted molecular weight of 129 kDa. Fig. 1 shows a schematic view of the three sialidases. The smallest isoenzyme, nanH, contains only a sialidase domain and has only 19% sequence identity with nanI and nanJ. These two latter isoenzymes are more closely related, having 57% sequence identity, and both contain a lectin domain preceding the sialidase domain, as judged by homology with the leech sialidase, the structure of which is known (Luo *et al.*, 1998). The largest sialidase, nanJ, possesses additional domains: a putative F5/8 C-type or discoidin domain (Baumgartner *et al.*, 1998) upstream of the lectin domain and a domain of unknown function following the sialidase domain, with a putative fibronectin type 3 domain at its C-terminus.

The 77 kDa nanI sialidase from *C. perfringens* strain ATCC 10543 consists of 694 amino acids. Sequence alignment with other sialidases shows that it has closest similarity to the intramolecular trans-sialidase of the leech *Macrobdella decora*, with 32% overall sequence identity. The leech sialidase consists of an N-terminal lectin-like domain followed by the canonical β-propeller catalytic domain, into which is inserted an irregular β-stranded domain (Luo *et al.*, 1998). Alignment with the nanI sialidase suggests that it too is composed of two distinct domains.

Here, we report preliminary X-ray diffraction data to very high resolution from the catalytic domain of *C. perfringens* nanI and its solution using the equivalent domain from the leech trans-sialidase. The structure of this sialidase in complex with a variety of substrate and transition-state analogues at atomic resolution should further add to our understanding of this superfamily of enzymes.

2. Materials and methods

2.1. Expression, purification and crystallization of nanI

Full-length nanI from *C. perfringens* ATCC 10543 with a six-histidine tag at the C-terminus was expressed and purified as described previously (Sheu *et al.*, 2002). Crystals were grown at 293 K under oil using a 96-well microbatch system and a Douglas Instruments robot with 2 µl protein solution and 2 µl precipitant solution. Crystals were obtained from Hampton Research Crystal Screen I condition No. 42: 20%(*w*/*v*) PEG 8000 and 0.05 *M* KH_2PO_4. The crystals were extracted from under the oil, cryoprotected in the precipitant with the addition of 20%(*v*/*v*) glycerol for 1 min and flash-frozen at 100 K. Diffraction to 2.5 Å was obtained on an in-house X-ray source and the crystals belonged to space group $P2_12_12_1$, with unit-cell parameters $a = 96.8$, $b = 69.2$, $c = 72.8$ Å. Assuming one 77 kDa monomer per asymmetric unit gives a Matthews coefficient, V_M, of 1.58 Å^3 Da^{-1} and a solvent content of 22%. This unusually tight packing led us to examine the crystal contents. Several crystals were washed extensively and dissolved in water prior to analysis by mass spectrometry and N-terminal sequencing. This revealed a 50.3 kDa protein with an N-terminus beginning VEGA, corresponding to residues 243–694. The full-length protein was prone to degradation (Sheu *et al.*, 2002), which we were unable to prevent and which led to problems in reproducibility of crystals. We therefore decided to subclone the catalytic domain.

2.2. Subcloning, expression and purification

The fragment of the gene encoding the C-terminal catalytic domain (amino acids 243–694) was amplified by PCR from vector DNA encoding the full-length nanI sialidase gene (Sheu *et al.*, 2002). Primers were designed to create an *Nco*I site incorporating the initition codon at the 5′ site and a *Pst*I site following the termination codon. The amplified fragment was digested with *Nco*I and *Pst*I and was cloned into a modified maltose-binding protein fusion vector with a Tev protease-cleavage site upstream of the initiation codon of the nanI protein. The fusion protein has a six-histidine tag at the N-terminus of the maltose-binding protein. The plasmid was transformed into *Escherichia coli* Rosetta (DE3) PlysS (Novagen).

The transformations were cultivated at 312 K in Luria–Bertani medium containing 100 µg ml^{-1} carbenicillin until the optical density at 595 nm reached 0.5–0.6. Isopropyl-β-D-thiogalactoside (IPTG) was added to induce protein expression and cell cultivation was continued for 19 h at 298 K. Cells were harvested by centrifugation, resuspended in buffer *A* (20 m*M* sodium phosphate pH 7.0, 500 m*M* NaCl, 40 m*M* imidazole, 2× complete EDTA-free protease-inhibitor tablets; Roche) and lysed by sonication and rLysozyme treatment (Novagen). After centrifugation, the supernatant was loaded onto a 12 ml Tricorn nickel-chelating column (Amersham Biosciences). The protein was eluted using a step gradient of 500 m*M* imidazole in buffer *A*. Fractions containing the fusion protein were pooled and applied onto a 20 ml amylose column (NEB) and washed using buffer *B* (50 m*M* Tris–HCl pH 7.5, 300 m*M* NaCl, 0.5 m*M* EDTA, 1 m*M* DTT, 20 m*M* imidazole). The protein was eluted using a step gradient of 10 m*M* maltose in buffer *B*. The eluted protein was pooled and digested overnight at 298 K with 20 µg ml^{-1} Tev protease. The cleaved protein was buffer-exchanged using a 10 ml Q-Sepharose column (Amersham Biosciences) by diluting the NaCl to 50 m*M* in the same buffer and eluting with buffer *C* (20 m*M* sodium phosphate pH 7.0, 200 m*M* NaCl). The final purification step used a second nickel-chelating column (5 ml; Amersham Biosciences) to remove any contaminating MBP. The pure protein was concentrated to 20 mg ml^{-1} for crystallization.

2.3. Crystallization

All crystallization experiments were performed using the sitting-drop vapour-diffusion method in Douglas Instruments 96-well crystallization plates at 292 K. The drops were made up of equal amounts (2 µl) of mother liquor and protein solution (20 mg ml^{-1} in buffer *C*). Initial crystals were obtained from PEG/Ion Screen (Hampton Research) condition No. 18 (20%

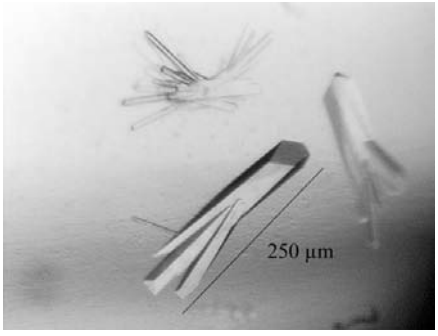

Figure 2
Crystals of the catalytic domain of *C. perfringens* nanI.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shells.

X-ray source	ID14-1 ESRF
Wavelength (Å)	0.934
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	a = 96.77, b = 69.24, c = 72.81
Resolution range (Å)	20–0.97 (1.03–0.97, 0.97–0.92)
Observed reflections	2736520
Unique reflections	290992
Data completeness (%)	87 (87, 54)
Redundancy	3.8 (3.1, 2.5)
$R_{\rm merge}$†	0.095 (0.186, 0.272)
$I/\sigma(I)$	11.8 (5.0, 2.8)

† $R_{\rm merge} = \sum|I(k) - \langle I\rangle|/\sum I(k)$, where $I(k)$ is the value of the kth measurement of the intensity of a reflection, $\langle I\rangle$ is the mean value of the intensity of that reflection and the summation is over all measurements.

PEG 3350, 0.2 *M* potassium nitrate). Good-quality crystals grew in 4–6 d (Fig. 2) and there was no need to refine the conditions. When needed, new crystals could easily be grown by seeding overnight with a crushed crystal.

2.4. X-ray diffraction

The crystals were transferred to a cryoprotectant solution containing 25% PEG 400 in mother liquor before being flash-frozen in a liquid-nitrogen stream at 100 K. Diffraction data were collected to a resolution of 0.92 Å on beamline ID14-1 at ESRF, the resolution limit of this station, using an ADSC Q4R CCD detector system (Fig. 3). Data were collected in 0.5° oscillations with 3 s exposure. A high-resolution pass provided the 0.92 Å data and a subsequent 1.6 Å resolution pass with a 2× attenuator in the beam provided measurements of data overloaded on the first pass. All data were processed, merged and scaled using the programs *MOSFLM* (Leslie, 1992) and *SCALA* from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). The crystals belonged to space group $P2_12_12_1$, with unit-cell parameters a = 96.77, b = 69.24, c = 72.81 Å. Data-collection and processing statistics are summarized in Table 1. The data have an overall completeness of 92% to 0.97 Å and 87% to 0.92 Å. A Matthews coefficient of 2.5 Å^3 Da^{-1}, corresponding to 49% solvent content, suggests the presence of one 50.3 kDa molecule in the asymmetric unit.

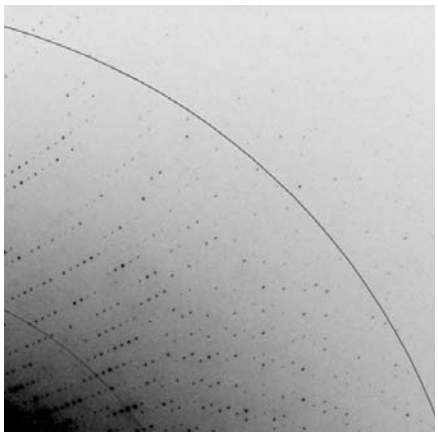

Figure 3
Part of a 0.5° oscillation X-ray diffraction image of a crystal of nanI. The inner circle is at 1.3 Å and the outer circle at 1.0 Å.

2.5. Molecular-replacement solution

A molecular-replacement solution was obtained using *CNS* (Brünger *et al.*, 1998). The structure of the intramolecular trans-sialidase from the leech *M. decora* (PDB code 1sll) was used as a starting model. The search model consisted of just the β-propeller domain, reduced to polyalanine, with the N-terminal lectin and irregular β-strand domains removed. Using data in the range 15–4 Å, a clear cross-rotation solution was found with a height of 0.089 (5.9σ above the mean), the next highest peak being 0.046 (2.3σ above the mean). The translation function, using the same data limits, gave a 'monitor' value of 0.305 (5σ above the mean). No unfavourable molecular contacts were observed for this solution in the crystal packing.

3. Results and discussion

Persistent breakdown of the compete 77 kDa *C. perfringens* nanI sialdiase led us to subclone the stable catalytic domain from residues 243 to 694. Previous studies had subcloned residues 216–694 and found that this domain exhibited a similar specific activity to the full-length protein (Sheu *et al.*, 2002). From its sequence similarity to the leech sialidase, the positions of the conserved RIP and Asp-box motifs and the conservation of other conserved active-site residues (Taylor, 1996), we can assume that all of the catalytic activity resides in the 243–694 domain. The rapid degradation of the full-length protein suggests that the linker between the lectin and sialidase domains is susceptible to protease digestion and that the lectin domain is perhaps flexibly linked. Our main interest is in the catalytic domain, for which we now have very high resolution data. The high order of the crystals, reflected in the Wilson B factor of 5.2 Å^2, suggests that we may be able extend the data to even higher resolution on an appropriate synchrotron beamline. Such unprecedented resolution for a sialidase will enable us to investigate the catalytic mechanism in detail through complexes with substrates and substrate analogues. The sialidase domain of nanI is approximately 70 amino acids longer than nanH and the location of the Asp boxes and conserved active-site residues suggests that, like the leech sialidase, nanI may also possess an additional domain inserted into the β-propeller. The leech sialidase domain is approximately 30 amino acids longer than the nanI sialidase domain, suggesting that any inserted domain in nanI may be different to that observed in the leech enzyme. Structure determination of nanI should also shed light on the role of such an additional domain.

This work was supported by the University of St Andrews and SN was supported by a BBSRC studentship. The work of expressing the full-length nanI was supported by grant NSC-90-2314-B-010-062 from the National Science Council, Taipei, Taiwan. We thank staff at ESRF for assistance and the EU for funds to access the facility.

References

Alymova, I. V., Taylor, G., Takimoto, T., Lin, T. H., Chand, P., Babu, Y. S., Li, C., Xiong, X. & Portner, A. (2004). *Antimicrob. Agents Chemother.* **48**, 1495–1502.
Amaya, M. F., Watts, A. G., Damager, I., Wehenkel, A., Nguyen, T., Buschiazzo, A., Paris, G., Frasch, A. C., Withers, S. G. & Alzari, P. M. (2004). *Structure*, **12**, 775–784.
Baumgartner, S., Hofmann, K., Chiquet-Ehrismann, R. & Bucher, P. (1998). *Protein Sci.* **7**, 1626–1631.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C. & Alzari, P. M. (2002). *Mol. Cell*, **10**, 757–768.
Buschiazzo, A., Tavares, G. A., Campetella, O., Spinelli, S., Cremona, M. L., Paris, G., Amaya, M. F., Frasch, A. C. & Alzari, P. M. (2000). *EMBO J.* **19**, 16–24.
Chong, A. K. J., Pegg, M. S., Taylor, N. R. & von Itzstein, M. (1992). *Eur. J. Biochem.* **207**, 335–343.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Corfield, T. (1992). *Glycobiology*, **2**, 509–521.
Crennell, S. J., Garman, E. F., Laver, W. G., Vimr, E. R. & Taylor, G. L. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 9852–9856.
Crennell, S., Garman, E., Laver, W., Vimr, E. & Taylor, G. (1994). *Structure*, **2**, 535–544.
Crennell, S., Takimoto, T., Portner, A. & Taylor, G. (2000). *Nature Struct. Biol.* **7**, 1068–1074.
Gaskell, A., Crennell, S. & Taylor, G. (1995). *Structure*, **3**, 1197–1205.
Itzstein, M. von, Wu, W. Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Phan, T. V., Smythe, M. L.,

White, H. F., Oliver, S. W., Colman, P. M., Varghese, J. N., Ryan, D. M., Woods, J. M., Bethell, R. C., Hotham, V. J., Cameron, J. M. & Penn, C. R. (1993). *Nature (London)*, **363**, 418–423.
Kruse, S., Kleineidam, R. G., Roggentin, P. & Schauer, R. (1996). *Protein Expr. Purif.* **7**, 415–422.
Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
Luo, Y., Li, S. C., Chou, M. Y., Li, Y. T. & Luo, M. (1998). *Structure*, **6**, 521–530.
Roggentin, P., Kleineidam, R. G. & Schauer, R. (1995). *Biol. Chem. Hoppe-Seyler*, **376**, 569–575.
Roggentin, P., Rothe, B., Kaper, J. B., Galen, J., Lawrisuk, L., Vimr, E. R. & Schauer, R. (1989). *Glycoconj. J.* **6**, 349–353.
Sheu, S.-Y., Tseng, H.-J., Huang, S.-P. & Chien, C.-H. (2002). *Enzyme Microb. Technol.* **31**, 794–803.
Taylor, G. (1996). *Curr. Opin. Struct. Biol.* **6**, 830–837.
Thobhani, S., Ember, B., Siriwardena, A. & Boons, G. J. (2003). *J. Am. Chem. Soc.* **125**, 7154–7155.
Traving, C., Schauer, R. & Roggentin, P. (1994). *Glycoconj. J.* **11**, 141–151.
Varghese, J. N., Laver, W. G. & Colman, P. M. (1983). *Nature (London)*, **303**, 35–40.
Watts, A. G., Damager, I., Amaya, M. L., Buschiazzo, A., Alzari, P., Frasch, A. C. & Withers, S. G. (2003). *J. Am. Chem. Soc.* **125**, 7532–7533.