

Guogang Xu,<sup>a</sup> Xuejun Li,<sup>a</sup>  
Peter W. Andrew<sup>b</sup> and  
Garry L. Taylor<sup>a\*</sup>

<sup>a</sup>Centre for Biomolecular Sciences, University of  
St Andrews, St Andrews, Fife KY16 9ST,  
Scotland, and <sup>b</sup>Department of Infection,  
Immunity and Inflammation, University of  
Leicester, Leicester LE1 9HN, England

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

Received 16 June 2008

Accepted 29 July 2008

**PDB Reference:** NanA sialidase, 2vvz, r2vvzsf.

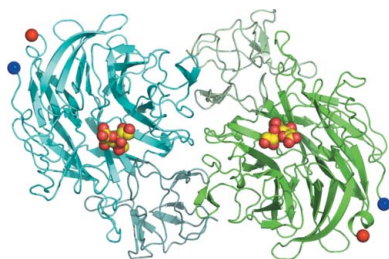
## Structure of the catalytic domain of *Streptococcus pneumoniae* sialidase NanA

*Streptococcus pneumoniae* genomes encode three sialidases, NanA, NanB and NanC, which are key virulence factors that remove sialic acids from various glycoconjugates. The enzymes have potential as drug targets and also as vaccine candidates. The 115 kDa NanA is the largest of the three sialidases and is anchored to the bacterial membrane. Although recombinantly expressed full-length NanA was soluble, it failed to crystallize; therefore, a 56.5 kDa domain that retained full enzyme activity was subcloned. The purified enzyme was crystallized in 0.1 M MES pH 6.5, 30% (w/v) PEG 4000 using the sitting-drop vapour-diffusion method. Data were collected at 100 K to 2.5 Å resolution from a crystal grown in the presence of the inhibitor 2-deoxy-2,3-dehydro-*N*-acetyl neuraminic acid. The crystal belongs to space group  $P2_12_12_1$ , with unit-cell parameters  $a = 49.2$ ,  $b = 95.6$ ,  $c = 226.6$  Å. The structure was solved by molecular replacement and refined to final  $R$  and  $R_{\text{free}}$  factors of 0.246 and 0.298, respectively.

### 1. Introduction

*Streptococcus pneumoniae* is a major human pathogen that is responsible for respiratory-tract infections, septicaemia, otitis media and meningitis. Current broad-spectrum antibiotic treatments for *S. pneumoniae* are increasingly unsuccessful owing to the emergence of drug-resistant strains (Thornsberry *et al.*, 1999). There are multi-valent capsular polysaccharide vaccines available for pneumococcal disease, but their efficacy in certain high-risk groups has been questioned (Siber, 1994). More recently, a conjugate polysaccharide vaccine has been introduced successfully (Prevenar, Wyeth), but there are concerns about how readily it can be introduced globally and about its continuing efficacy. A search is therefore on for new drug and vaccine candidates for pneumococcal diseases. Several virulence factors may contribute to colonization and early infection processes (Jedrzejewski, 2001). Sialidases are one key virulence factor as they remove sialic acid from host cell-surface glycans, probably unmasking certain receptors to facilitate bacterial adherence and colonization (Paton *et al.*, 1993). To date, all *S. pneumoniae* clinical isolates investigated have had prominent sialidase activities. Up to three distinct sialidases, NanA (Camara *et al.*, 1994), NanB (Berry *et al.*, 1996) and NanC, are encoded in *S. pneumoniae* genomes, with a recent study revealing NanA to be present in all clinical strains (Pettigrew *et al.*, 2006). Gene-knockout studies in mouse models have shown that NanA and NanB are essential for *S. pneumoniae* infection (Manco *et al.*, 2006).

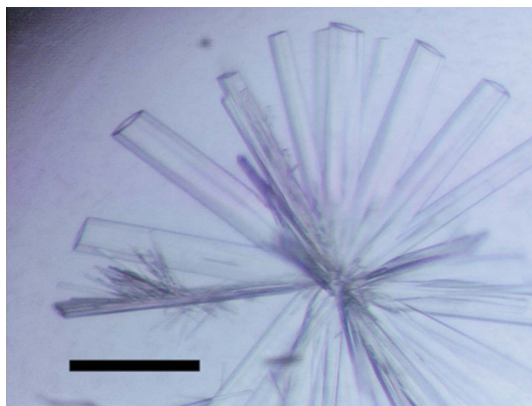
In this paper, we report the 2.5 Å resolution X-ray crystallographic structure of the catalytic domain of *S. pneumoniae* NanA and its complex with the inhibitor 2-deoxy-2,3-dehydro-*N*-acetyl neuraminic acid (Neu5Ac2en). This provides a framework for the structure-based design of specific inhibitors of pneumococcal sialidases as potential therapeutic agents.



## 2. Experimental

Full-length NanA could be recombinantly expressed and purified, but failed to crystallize. Limited proteolysis using trypsin, followed by mass spectrometry of the cleavage products, identified a stable sub-domain, which we designate CNanA, that spans residues 319–822 and encompasses a domain that retains the enzyme activity of the full-length NanA. The *S. pneumoniae nanA* gene in a pQE30 vector was used as a template in polymerase chain reaction (PCR) with the following primers: 5'-ACCTCCATGGAAGGAGCGGCTTTAACA-GAGA-3' and 5'-GGGCCTCGAGTTAGACCAATACTTCTGAG-TCG-3' (*NcoI* and *XhoI* restriction sites in bold). The PCR product was then ligated into the pHISTEV vector, containing six histidines and a tobacco etch virus (TEV) cleavage peptide at the N-terminus, and plasmid DNA was extracted using a Mini-Prep Kit (Promega). The plasmid was transformed into *Escherichia coli* BL21 (DE3) expression strain (Novagen) for protein expression. The transformed *E. coli* was inoculated into Luria–Bertani (LB) medium with 100 µg ml<sup>-1</sup> kanamycin at 310 K. 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) was added to induce CNanA expression when the optical density at 600 nm (OD<sub>600</sub>) of the cultures reached 0.6. Cell culture continued at 310 K for 3 h before harvesting by centrifugation at 4500g for 30 min at 277 K. The harvested cell pellets were resuspended in 0.1 M phosphate pH 7.4, 10 mM imidazole and sonicated with 5 × 30 s bursts. Protease-inhibitor cocktail tablets (one tablet per 25 ml extract; Roche Diagnostics) and DNAase (Sigma; final concentration 20 µg ml<sup>-1</sup>) were then added. The crude cell extract was centrifuged at 43 000g for 20 min at 277 K to remove the cell debris and the supernatant was filtered with a syringe-driven filter (0.45 µm) before starting protein purification. Soluble cell extract was loaded onto a 5 ml nickel column (GE Healthcare) and the bound protein was eluted with 300 mM imidazole in 0.1 M phosphate buffer pH 7.4. Protein purity was assessed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI–TOF). Relatively high-purity target protein was pooled for gel filtration using a 120 ml Sephacryl-200 column (GE Healthcare). The purified CNanA was dialysed against 0.1 M Tris–HCl pH 8.0, 150 mM NaCl overnight before concentration and storage.

Purified protein was concentrated to 10.9 mg ml<sup>-1</sup> for crystallization experiments using the sitting-drop vapour-diffusion method at 290 K with the commercial kits Classics (Jena Bioscience), JCSG, Nextal PEGs and Nextal pH Clear (Qiagen). Crystalline materials were observed after 3 d from condition No. 32 of Nextal PEGs [0.1 M MES pH 6.5, 25% (w/v) PEG 4000]. This condition was selected for



**Figure 1**  
Crystals of CNanA. The scale bar represents 0.5 mm.

**Table 1**

Crystallographic summary.

Values in parentheses are for the highest resolution shell.

Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unit-cell parameters (Å)	<i>a</i> = 49.2, <i>b</i> = 95.6, <i>c</i> = 226.6
Maximum resolution (Å)	2.5 (2.54–2.50)
Unique reflections	36773
Completeness	95.2 (71.2)
<i>I</i> /σ( <i>I</i> )	27.3 (4.6)
Mosaicity (°)	0.64
Redundancy	2.9 (2.5)
<i>R</i> <sub>merge</sub> <sup>†</sup>	0.063 (0.248)
<i>V</i> <sub>M</sub> (Å <sup>3</sup> Da <sup>-1</sup> )	2.36
Refinement	
Protein atoms	7442
Other atoms	40 (Neu5Ac2en), 124 waters, 1 Cl <sup>-</sup>
Resolution range (Å)	30–2.5
<i>R</i> <sub>cryst</sub> <sup>‡</sup>	0.246
<i>R</i> <sub>free</sub> <sup>‡</sup>	0.298
Mean temperature factor (Å <sup>2</sup> )	
Protein, monomer <i>A/B</i>	29/45
Neu5Ac2en, monomer <i>A/B</i>	25/45
Waters	25
R.m.s.d. bond lengths (Å)	0.007
R.m.s.d. bond angles (°)	1.145
Ramachandran favoured/outliers (%)	90.5/3.7

<sup>†</sup>  $R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$ . <sup>‡</sup>  $R_{\text{cryst}}$  and  $R_{\text{free}} = \frac{(\sum |F_o| - |F_c|)}{(\sum |F_o|)}$ ;  $R_{\text{free}}$  was calculated for a 5% set of reflections excluded from the refinement.

crystallization optimization with crystallization drops made up of equal amounts of protein solution and reservoir solution (2 µl each). The best condition for CNanA crystallization was 0.1 M MES pH 6.5, 30% (w/v) PEG 4000. Crystals appeared in about 2 d and reached their maximum size after one week. The crystals have the remarkable habit of a square hollow tube (Fig. 1). Crystals of the Neu5Ac2en complex structure were grown in the presence of 10 mM Neu5Ac2en.

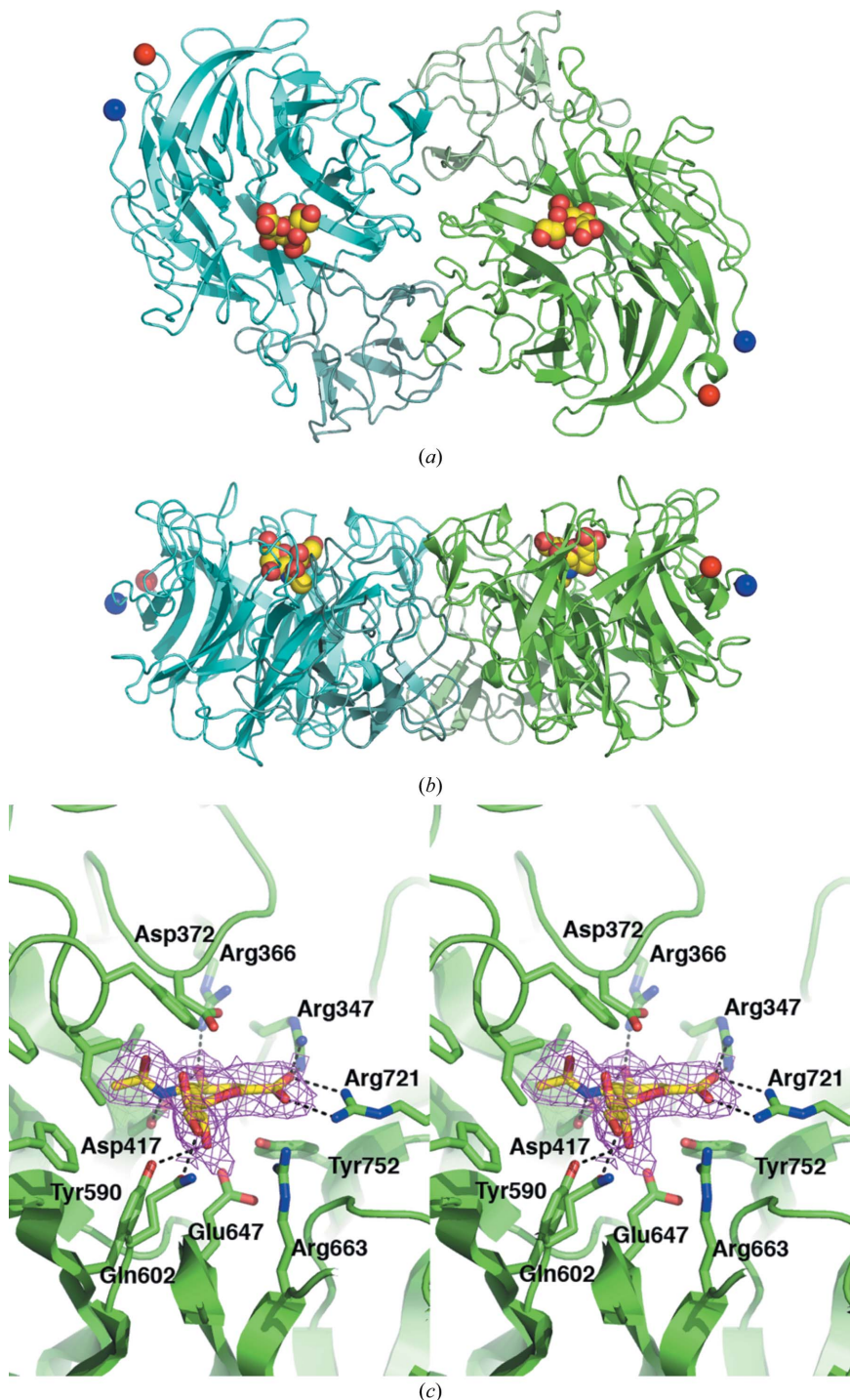
Crystals were cryoprotected by transfer for a few minutes into a solution of the crystallization buffer with 20% (v/v) glycerol before data collection at 100 K. Data were collected in-house (Rigaku-MSX MicroMax-007 HF X-ray generator and Saturn 944+ CCD detector). *HKL-2000* (Otwinowski & Minor, 1997) was used for data processing and scaling and data-collection statistics are shown in Table 1. The catalytic domain of *Clostridium perfringens* sialidase NanI (Newstead *et al.*, 2008; PDB code 2bf6) was used to solve the CNanA structure by molecular replacement using *Phaser* (McCoy *et al.*, 2007) in the CCP4 suite (Collaborative Computational Project, Number 4, 1994). The asymmetric unit contains two CNanA molecules. The automated model-building wizard of the *Phenix* package (Adams *et al.*, 2002) was used to build the initial structure using the 2.5 Å resolution Neu5Ac2en complex data. This procedure built 85% of the residues with an *R* and *R*<sub>free</sub> of 0.29 and 0.35, respectively. *Coot* (Emsley & Cowtan, 2004) and *REFMAC5* (Murshudov *et al.*, 1997) were used to refine and build the final model, which was validated with *MolProbity* (Lovell *et al.*, 2003). Refinement statistics are summarized in Table 1. The first 20 amino acids, including the six-histidine tag and TEV cleavage peptide, and the last 20 amino acids are not visible in the electron-density maps; the final model consists of residues 322–791. Molecule *A* is generally well ordered, whereas molecule *B* shows disorder in its N- and C-terminal regions. Both monomers have Neu5Ac2en bound.

## 3. Results and discussion

The structure of CNanA shows the canonical six-bladed β-propeller fold common to all sialidases (Fig. 2). In common with the catalytic

domains of *C. perfringens* NanI (Newstead *et al.*, 2008) and leech sialidase (PDB code 1sli; Luo *et al.*, 1998), CNanA has a small domain, residues 436–535, inserted between the second and third strands of the second sheet. CNanA and NanI share 41% sequence

identity and the r.m.s.d. between 413 C $\alpha$  atoms is 1.26 Å. The two monomers in the asymmetric unit are related by a noncrystallographic twofold axis, superimpose with an r.m.s.d. of 0.39 Å over 470 C $\alpha$  atoms and bury ~1200 Å<sup>2</sup> of surface at their dimer interface. The calibrated gel filtration suggests that CNanA is predominantly a monomer in solution; however, it is possible that full-length NanA may be a dimer on the bacterial surface. The active site contains the usual catalytic residues common to all sialidases (Taylor, 1996): three arginines (Arg347, Arg663, Arg721) that interact with the carboxylate group of sialic acid, a nucleophilic tyrosine (Tyr752) that is proposed to form a covalent intermediate (Watts *et al.*, 2006; Newstead *et al.*, 2008) and its associated glutamic acid (Glu647), an aspartic acid (Asp372) that acts as an acid/base and a hydrophobic pocket that accommodates the acetamidomethyl group of sialic acid. In common with other bacterial sialidases, the hydroxyl group at C4 on Neu5Ac2en interacts with an arginine (Arg366) and an aspartic acid (Asp417). The O8 and O9 hydroxyls of the ligand's glycerol group interact with Tyr590 and Gln602, respectively. The topology of the surface of CNanA around the location of the aglycon is flat and open, in line with the promiscuity shown by NanA towards  $\alpha$ (2,3),  $\alpha$ (2,6) and  $\alpha$ (2,8) linkages (unpublished data) and in contrast to leech sialidase and *Trypanosoma cruzi* *trans*-sialidase, which show specificity for  $\alpha$ (2,3)-linked sialic acids (Amaya *et al.*, 2003). Two inhibitors of the influenza virus neuraminidase, zanamivir and oseltamivir, are currently licensed as treatments for influenza and both were developed based on the framework of Neu5Ac2en. The study reported here lays the groundwork for the potential elaboration of the core of Neu5Ac2en, particularly around the acetamido and glycerol moieties, in order to develop specific inhibitors of the pneumococcal sialidases.



**Figure 2**  
Crystal structure of CNanA. Orthogonal views of the CNanA dimer are shown in (a) and (b), where (b) is related to (a) by a 90° rotation about a horizontal axis. The N- and C-termini are indicated by blue and red spheres, respectively. Molecule A (green) and molecule B (cyan) are drawn with the inserted domains (residues 436–535) drawn in lighter shades. The inhibitor Neu5Ac2en is shown in each monomer and is drawn in space-filling mode. (c) Stereoview of the active site of monomer A showing the hydrogen-bond interactions made between Neu5Ac2en and CNanA, with only key amino acids drawn for clarity. The  $2F_o - F_c$  electron-density map contoured at  $1\sigma$  is only drawn around the inhibitor for clarity. These figures were drawn using PyMOL (DeLano, 2007).

GX was supported by Biocryst Pharmaceuticals Inc, Birmingham, Alabama. Resources of the St Andrews-based Scottish Structural Proteomics Facility, funded by the Scottish Funding Council, the Biotechnology and Biological Sciences Research Council (BBSRC) and the University of St Andrews, were used in this project.

### References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* **D58**, 1948–1954.  
Amaya, M. F., Buschiazzo, A., Nguyen, T. & Alzari, P. M. (2003). *J. Mol. Biol.* **325**, 773–784.  
Berry, A. M., Lock, R. A. & Paton, J. C. (1996). *J. Bacteriol.* **178**, 4854–4860.  
Camara, M., Boulnois, G. J., Andrew, P. W. & Mitchell, T. J. (1994). *Infect. Immun.* **62**, 3688–3695.

- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- DeLano, W. L. (2007). *The PyMOL Molecular Graphics System*. <http://www.pymol.org>.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Jedrzejewski, M. J. (2001). *Microbiol. Mol. Biol. Rev.* **65**, 187–207.
- Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
- Luo, Y., Li, S. C., Chou, M. Y., Li, Y. T. & Luo, M. (1998). *Structure*, **6**, 521–530.
- 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.
- Manco, S., Herson, F., Yesilkaya, H., Paton, J. C., Andrew, P. W. & Kadioglu, A. (2006). *Infect. Immun.* **74**, 4014–4020.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Newstead, S. L., Potter, J. A., Wilson, J. C., Xu, G., Chien, C. H., Watts, A. G., Withers, S. G. & Taylor, G. L. (2008). *J. Biol. Chem.* **283**, 9080–9088.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Paton, J. C., Andrew, P. W., Boulnois, G. J. & Mitchell, T. J. (1993). *Annu. Rev. Microbiol.* **47**, 89–115.
- Pettigrew, M. M., Fennie, K. P., York, M. P., Daniels, J. & Ghaffar, F. (2006). *Infect. Immun.* **74**, 3360–3365.
- Siber, G. R. (1994). *Science*, **265**, 1385–1387.
- Taylor, G. (1996). *Curr. Opin. Struct. Biol.* **6**, 830–837.
- Thornsberry, C., Jones, M. E., Hickey, M. L., Mauriz, Y., Kahn, J. & Sahn, D. F. (1999). *J. Antimicrob. Chemother.* **44**, 749–759.
- Watts, A. G., Oppezzo, P., Withers, S. G., Alzari, P. M. & Buschiazio, A. (2006). *J. Biol. Chem.* **281**, 4149–4155.