Structure of an Influenza A virus N9 neuraminidase with a tetrabrachion-domain stalk

Victor A. Streltsov,^{a,b,†} Peter M. Schmidt^{a,c,†} and Jennifer L. McKimm-Breschkin^{a,d,*‡}

^aCSIRO Manufacturing, 343 Royal Parade, Parkville, Victoria 3052, Australia, ^bThe Florey Institute of Neuroscience and Mental Health, The University of Melbourne, 30 Royal Parade, Parkville, Victoria 3052, Australia, ^cR&D, CSL Behring GmbH, Emil-von-Behring Strasse 76, 35041 Marburg, Germany, and ^dDepartment of Microbiology and Immunology, The University of Melbourne at the Peter Doherty Institute for Infection and Immunity, 792 Elizabeth Street, Melbourne, Victoria 3000, Australia. *Correspondence e-mail: jmbvirology@gmail.com

Received 16 July 2018

Accepted 18 December 2018

Edited by M. J. van Raaij, Centro Nacional de Biotecnología – CSIC, Spain

† All authors contributed equally.

Keywords: Influenza A virus; neuraminidases; neuraminidase stalk; tetrabrachion tetramerization domain; N9 neuraminidase structure.

PDB references: recombinant N9 with stalk, 6crd; recombinant N9 without stalk, 6mcx; N9 without stalk from egg-grown virus, 6d3b

Supporting information: this article has supporting information at journals.iucr.org/f

The influenza neuraminidase (NA) is a homotetramer with head, stalk, transmembrane and cytoplasmic regions. The structure of the NA head with a stalk has never been determined. The NA head from an N9 subtype influenza A virus, A/tern/Australia/G70C/1975 (H1N9), was expressed with an artificial stalk derived from the tetrabrachion (TB) tetramerization domain from *Staphylothermus marinus*. The NA was successfully crystallized both with and without the TB stalk, and the structures were determined to 2.6 and 2.3 Å resolution, respectively. Comparisons of the two NAs with the native N9 NA structure from egg-grown virus showed that the artificial TB stalk maintained the native NA head structure, supporting previous biological observations.

1. Introduction

Influenza remains a serious disease that continues to circulate globally and, despite the availability of vaccines and antivirals, the most recent estimate from the Centers for Disease Control and Prevention suggests a worldwide annual death toll of between 291 000 and 646 000 (Iuliano *et al.*, 2018). There are two types of influenza, influenza A and influenza B, that cause serious respiratory disease in humans. Influenza A is further subdivided into subtypes based on the two surface glycoproteins hemagglutinin (HA) and neuraminidase (NA). There are 16 HA subtypes and nine NA subtypes that infect humans (Liu *et al.*, 2009). HA binds to terminal sialic acids on the cell membrane to facilitate virus adsorption and entry. NA cleaves sialic acids on the cell surface and on the virus itself, preventing self-aggregation and leading to the release and spread of newly synthesized viruses. NA is a tetramer, with four identical polypeptide chains arranged to give a molecule with a square box-like head (100 × 100 × 60 Å), a narrow central stalk (15 × 100 Å) and a hydrophobic base which serves to anchor the NA in the viral membrane (Ward *et al.*, 1983). It is an N-terminally anchored integral membrane protein with a cytoplasmic tail, which contains six identical polar residues that are found in all influenza A subtypes, a transmembrane region formed by approximately 30 hydrophobic amino acids, the first six of which are conserved, the stalk, which is highly variable in both amino-acid sequence and length, and the head, which contains the catalytic site

(Blok & Air, 1982). Despite large differences in the sequences of the heads of the different NA types and subtypes, all form a structurally similar tetramer. There are nine residues in the active site which interact directly with the substrate and are highly conserved across all influenza A and B NAs. There are a further 10–11 highly conserved residues defined as framework residues, which make no direct contact with the substrate but hold the functional residues in place for binding and catalysis (Varghese *et al.*, 1992; Colman *et al.*, 1993; Burmeister *et al.*, 1992).

Similarly, despite the large variations in the stalk region, all NAs still form a similar tetrameric structure. Although the formation of disulfide bonds between aligned cysteine residues in the stalk of N2 NA monomers has been described, no similarity in the location of the cysteines or in their flanking sequences is found (Blok *et al.*, 1982; Ward *et al.*, 1983). Secondary-structure prediction tools such as *JPred* (Cuff *et al.*, 1998) do not identify any helices or coiled coils, but show about 50% β -sheet content in the full-length stalks, and measurements of full-length and deleted stalks also indicated a mixture of extended and folded structures (Els *et al.*, 1985). Mutations as well as deletions in the stalk are known to affect the activity of NA (Castrucci & Kawaoka, 1993; Dai *et al.*, 2016).

Proteolytic digestion of virions releases the enzymatically active NA heads from the membrane-attached stalk region of the enzyme. This enabled the crystallization of NA heads for early structural studies (Ward *et al.*, 1982; McKimm-Breschkin *et al.*, 1991; Laver, 1978; Wright & Laver, 1978). The structure of NA heads with bound sialic acid ligands (Varghese *et al.*, 1992) was critical in the development of zanamivir, the first antiviral that specifically inhibits NA (von Itzstein *et al.*, 1993). Subsequently, structural analysis of drug-resistant mutant NAs has provided a greater understanding of the critical interactions of NA and several NA inhibitors. However, many of the mutant viruses do not grow well, so researchers have turned to the expression of recombinant NA. NA needs to be expressed as a membrane-bound tetramer for optimal activity. While full-length NA with wild-type activity can be expressed in insect cells using recombinant baculoviruses (Mather *et al.*, 1992; Kongkamnerd *et al.*, 2012), the enzyme may be poorly cleaved to generate NA heads for structural studies (Oakley *et al.*, 2010). Furthermore, the mutant NA may be unstable once cleaved. Artificial stalks from known tetramerization domains have therefore been used to generate recombinant, tetrameric, soluble NAs for functional and structural studies. Of the tetramerization domains that have been used for the expression of soluble recombinant influenza NA, which include human vasodilator-stimulated phosphoprotein (VASP; Baranovich *et al.*, 2017; Xu *et al.*, 2008; Gubareva *et al.*, 2017; Yang *et al.*, 2016), yeast GCN-4-pLI and the tetrabrachion (TB) domain from the deep-sea archaeon *Staphylothermus marinus* (Dai *et al.*, 2016; Schmidt *et al.*, 2011), the TB domain has some characteristics that may make it a better choice. Although originating from an archaeon and therefore being unrelated to viral NA, the TB domain has been shown to form extremely stable parallel tetramers (Stetefeld *et al.*, 2000), a

feature that may help to drive the tetramerization of fused NA. In addition, its highly parallel orientation (in contrast to GCN-4-pLI and VASP) seems to be compatible with the parallel orientation of the NA monomers in the tetrameric head of NA.

In agreement with Dai and coworkers, we found that the nature of the artificial tetramerization domain (GCN-4-pLI versus TB) had only a minor effect on the specific activity and K_m values of certain soluble NAs (Schmidt *et al.*, 2011; Dai *et al.*, 2016), whereas for other NAs the TB domain seemed to be superior (Dai *et al.*, 2016). As the latter work included the original NA stalk in the construct, the K_m values that are described in these papers cannot be compared directly. However, it seems that in general the TB domain resulted in a more stable tetrameric NA compared with GCN-4-pLI (Schmidt *et al.*, 2011) or VASP (Dai *et al.*, 2016), making the TB domain the preferred choice to stabilize soluble recombinant NA in its tetrameric state. In order to better characterize the impact of the TB tetramerization domain on the NA structure, we compared the crystal structure of the well characterized N9 NA from egg-grown A/tern/Australia/G70C/1975 influenza virus (G70C; Baker *et al.*, 1987; Smith *et al.*, 2006; Blick *et al.*, 1995; Varghese *et al.*, 1998; Kim *et al.*, 2013; McKimm-Breschkin *et al.*, 2018) with those of recombinant soluble G70C N9 in the presence and absence of a cleavable TB tetramerization domain.

2. Materials and methods

2.1. Protein expression and purification

The N9 NA from the A/tern/Australia/G70C/1975 influenza virus was purified from egg-grown virus as described previously (Blick *et al.*, 1995; Varghese *et al.*, 1997; McKimm-Breschkin *et al.*, 1991). Briefly, after Pronase removal of the NA heads from concentrated virions, the NA was separated from Pronase by size-exclusion chromatography on Superose 12 and concentrated to 5.8 mg ml⁻¹ using a 10 kDa cutoff spin concentrator (Millipore Amicon Ultra-4).

The DNA constructs that were used to express artificial stalk-stabilized influenza N9 NA were designed as shown in Figs. 1(a) and 1(b). Briefly, construct 1 consisted of an N-terminal FLAG-tag followed by a thrombin cleavage site for optional tag removal, the TB tetramerization domain (Peters *et al.*, 1995; Stetefeld *et al.*, 2000) and the N9 NA head domain of G70C (Blick *et al.*, 1995; Fig. 1a). Construct 2 was identical, with the exception that the thrombin cleavage site was positioned between the TB stalk and the NA head, allowing proteolytic removal of the artificial stalk (Fig. 1b). Both constructs were synthesized by GeneArt (Regensburg, Germany) and were cloned into the transfer vector pFastBac downstream of the sequence coding for the melittin signalling peptide (MSP), with the latter enabling efficient secretion of the expressed protein. Constructs were verified by sequencing (Micromon, Victoria, Australia). Bacmids and baculovirus particles were constructed using the Bac-to-Bac system (Life Technologies, Carlsbad, California, USA) according to the

manufacturer's protocols. Recombinant NA constructs were expressed in *Spodoptera frugiperda* (Sf21) insect cells cultured in Sf-900 II serum-free medium (Life Technologies, Carlsbad, California, USA) as described previously (Schmidt *et al.*, 2011). Cells were infected at a density of $\sim 1.7 \times 10^6 \text{ ml}^{-1}$ with a multiplicity of infection of 1. The total expression volumes were 4 l for construct 1 and 3 l for construct 2. Secreted NA activity in the media and cell viability were assessed every day using a fluorescence-based assay (Potier *et al.*, 1979; Schmidt *et al.*, 2011). Five days post-infection, the cells, cell debris and

baculovirus were spun down (40 000g, 1 h, 277 K), and the cleared supernatants were filtered through 1.2 and 0.22 μm filters. Azide and the protease inhibitor E-64 (Johnson & Jiang, 2005) were added to final concentrations of 0.02% and 1 μM , respectively. Affinity purification of FLAG-tagged NAs was performed using an anti-FLAG affinity column (Mini-Leak Low; Kem-En-Tec A/S, Copenhagen, Denmark) at 277 K, and bound FLAG-tagged proteins were eluted with TBA buffer (20 mM Tris pH 6.5, 10 mM CaCl_2 , 0.02% sodium azide) containing 0.25 mg ml^{-1} FLAG peptide. Fractions

Figure 1
Design of G70C N9 constructs. (a) Construct 1: G70C N9 with a thrombin-cleavable FLAG-tag. (b) Construct 2: G70C N9 with a cleavable TB stalk. (c) Crystals of construct 1 G70C N9 (N9 + TB stalk): $\sim 200 \mu\text{m}$ long rod and plate. (d) Crystals of thrombin-cleaved construct 2 G70C N9 (N9 - TB stalk): 'cubes' with $\sim 50 \mu\text{m}$ sides.

Table 1
Macromolecule-production information.

	Construct 1: N9 + TB stalk	Construct 2: N9 – TB stalk
Source organism	Influenza virus A/tern/Australia/G70C/1975 (N9)	Influenza virus A/tern/Australia/G70C/1975 (N9)
DNA source	Synthetic gene	Synthetic gene
Cloning vector	pMA (<i>ampR</i>)	pMK-RQ (<i>kanR</i>)
Expression vector	pFastBac1	pFastBac1
Expression host	<i>Spodoptera frugiperda</i> (Sf21)	<i>Spodoptera frugiperda</i> (Sf21)
Complete amino-acid sequence of the construct produced	MKFLVNVALVFMVVYISYIYA+DYKDDDDK+LVPR\$GGGS¶I INETADDIVYRLTVIIDRYESLKNLITLRADRLEMIIND NVSTILA††RDFNNLTGKGLCTINSWHIYGKDNAVRIGEDS DVLVTREPYVSCDPDECRFYALSQGTTRGKHSNGTIHDR SQYRALISWPLSSPPTVYNSRVECIGWSSSTCHDGKTRMS ICISGPNNNASAVIWNRRPVTEINTWARNILRTQSECV CHNGVCPVVFTDGSATGPAETRIYYFKEGKILKWEPLAGT AKHIEECSCYGERAEITCTCRDNWQGSNRPVIRIDPVAMT HTSQYICSPVLTNDNRPNDPTVGKCNDFYPGNNNNGVKGF SYLDGVNTWLGRITISIASRSGYEMLKVPNALTDKSKPTQ GQITVLNTDWSGYSFMDYWAECEYRACFYVELIRGR KEDKVVWTSNSIVSMCSSTEFLLGQWDWPDGAKIEYFL##	DDIVYRLTVIIDRYESLKNLITLRADRLEMIINDNVSTI LA††GGTVLAK¶LVPRGGK\$RDFNNLTGKGLCTINSWHIYG KDNAVRIGEDSDVLVTREPYVSCDPDECRFYALSQGTTR GKHSNGTIHDRSQYRALISWPLSSPPTVYNSRVECIGWSS TSCHDGKTRMSICISGPNNNASAVIWNRRPVTEINTWAR NILRTQSECVCHNGVCPVVFTDGSATGPAETRIYYFKEG KILKWEPLAGTAKHIEECSCYGERAEITCTCRDNWQGSN PVIRIDPVAMTHTSQYICSPVLTNDNRPNDPTVGKCNDFY PGNNNNNGVKGFSYLDGVNTWLGRITISIASRSGYEMLKVP ALTDKSKPTQGGQITVLNTDWSGYSFMDYWAECEYRAC CFYVELIRGRPKEDKVVWTSNSIVSMCSSTEFLLGQWDWPD GAKIEYFL##

† Signal peptide. ‡ FLAG-tag. § Thrombin cleavage site. ¶ Linker. †† TB stalk. ## N9.

showing NA activity were pooled and concentrated to final volumes of 800 and 980 µl, respectively, using a 10 kDa cutoff spin concentrator (Millipore Amicon Ultra-4).

Affinity-purified recombinant NA was subjected to thrombin cleavage to remove the FLAG tag (construct 1) or the entire FLAG-tagged TB stalk (construct 2) for subsequent crystallization trials. Thrombin cleavage of construct 1 was performed by mixing 800 µl NA solution (about 3 mg) with 130 U thrombin (1 U µl⁻¹; Sigma, catalogue No. T6634) and 270 µl phosphate-buffered saline (PBS) followed by incubation at room temperature (RT) for 48 h. Successful cleavage of the FLAG-tag and the functionality of NA heads were confirmed by anti-FLAG Western blotting (Supplementary Fig. S1) and by testing the NA catalytic activity before and after thrombin cleavage. To separate the cleaved tag and the remaining thrombin from the NA, samples were further purified by size-exclusion chromatography on a Superose 12 column (30 × 1 cm; GE Healthcare, Rydalmere, New South Wales, Australia) using a BioLogic DuoFlow System equipped with a QuadTec UV–Vis detector (Bio-Rad, Gladesville, New South Wales, Australia). The NA-containing fractions of construct 1 NA (Supplementary Fig. S2) were pooled and concentrated to 5.0 mg ml⁻¹ for subsequent crystallization.

In contrast to construct 1, construct 2 NA was cleaved poorly by thrombin. 980 µl (approximately 3 mg) of construct 2 NA was mixed with 100 U thrombin (1 U µl⁻¹), 120 µl TBS and incubated for 24 h at RT. Subsequent SDS–PAGE and Western blotting analysis indicated that a major fraction of the NA was not cleaved by thrombin (Supplementary Fig. S3). An additional 50 U thrombin was added and the sample was incubated for a further 24 h at RT, but without improving the outcome. As the remaining noncleaved NA seemed to run at a slightly higher molecular weight on SDS–PAGE compared with the starting material, we suspected that more extensive glycosylation of this NA subpopulation might prevent the protease from accessing its cleavage site owing to steric hindrance. To address this point, the sample was split into two

halves and one half was further incubated with 50 U PNGaseF for an additional 24 h at RT. Subsequent SDS–PAGE analysis showed a slightly sharper NA band (Supplementary Fig. S4), indicating the successful removal of some glyco-structures from the NA; however, the removal of the TB stalk by thrombin was not improved further. The functionality of the NA was monitored by activity assays and the NA was stable during the entire 72 h incubation period. The cleaved material was separated from noncleaved NA, thrombin, PNGaseF and cleaved TB stalk by size-exclusion chromatography on a Superose 12 column as described above (Supplementary Fig. S4). The fractions containing cleaved NA were pooled and the remaining noncleaved NA was removed by incubating the sample with anti-FLAG affinity matrix. The flowthrough was collected and concentrated to 1.3 mg ml⁻¹ for subsequent crystallization. Macromolecule-production information is summarized in Table 1.

2.2. Crystallization

Crystallization of the G70C N9 NA derived from egg-grown virus was achieved under standard conditions, as described previously, in 24-well hanging-drop plates (Blick *et al.*, 1995; Varghese *et al.*, 1998). Crystals of construct 2 NA were obtained by manually mixing 2 µl NA solution (5.8 mg ml⁻¹ in PBS) with 2 µl of crystallant consisting of a 4:1 mixture of 1.4 M KH₂PO₄ and 3.0 M K₂HPO₄ (4 + 1 ml). The well solution (500 µl per well) consisted of 1.4 M KH₂PO₄ and 3.0 M K₂HPO₄ in a 2:1 ratio (2 + 1 ml). The crystallization conditions for thrombin-treated and PNGaseF-treated TB-stalk-less construct 2 NA were identical, except that the hanging drops were seeded with a homogenized egg-grown G70C N9 NA crystal using a whisker.

Crystallization conditions for construct 1 NA were screened using the standard 96-well JCSG sitting-drop crystallization screen at the CSIRO Collaborative Crystallisation Centre. Droplets consisting of 150 nl NA (5.0 mg ml⁻¹) in PBS and

Table 2
Crystallization.

	N9 + TB stalk	N9 – TB stalk	N9, egg-grown virus
Method	Sitting-drop vapour diffusion	Hanging-drop vapour diffusion	Hanging-drop vapour diffusion
Plate type	SD-2 sitting-drop plates (IDEX)	24-well hanging-drop plates (Hampton Research)	24-well hanging-drop plates (Hampton Research)
Temperature (K)	293	293	293
Protein concentration (mg ml ⁻¹)	5.0	1.3	5.8
Buffer composition of protein solution	PBS	PBS	PBS
Composition of reservoir solution	10% PEG 6000, 100 mM HEPES pH 7.0	1.4 M KH ₂ PO ₄ + 3.0 M K ₂ HPO ₄ pH 7.0	1.4 M KH ₂ PO ₄ + 3.0 M K ₂ HPO ₄ pH 7.0
Volume and ratio of drop	300 nl, 1:1	4 µl, 1:1	4 µl, 1:1
Volume of reservoir (µl)	50	500	500

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

	N9 + TB stalk	N9 – TB stalk	N9, egg-grown virus
Diffraction source	MX2, Australian Synchrotron	MX2, Australian Synchrotron	MX1, Australian Synchrotron
Wavelength (Å)	0.95369	0.95369	1.0000
Temperature (K)	100	100	100
Detector	ADSC Quantum 315r	ADSC Quantum 315r	ADSC Quantum 210r
Crystal-to-detector distance (mm)	300	300	130
Rotation range per image (°)	1	1	1
Total rotation range (°)	180	180	180
Exposure time per image (s)	0.5	0.5	1
Space group	<i>P</i> 2 ₁	<i>I</i> 432	<i>I</i> 432
<i>a</i> , <i>b</i> , <i>c</i> (Å)	101.00, 142.33, 163.30	180.72, 180.72, 180.72	180.99, 180.99, 180.99
α , β , γ (°)	90, 91.48, 90	90, 90, 90	90, 90, 90
Mosaicity (°)	1.4	0.6	0.3
Resolution range (Å)	40.00–2.57 (2.63–2.57)	45.20–2.30 (2.35–2.30)	33.07–1.40 (1.44–1.40)
Total No. of reflections	304776	2392459	2560840
No. of unique reflections	108611 (5654)	22761 (1453)	98317 (4847)
Completeness (%)	77.6 (70.7)	98.1 (92.9)	99.0 (99.6)
Multiplicity	2.7 (2.1)	29.2 (10.5)	26.0 (14.3)
$\langle I/\sigma(I) \rangle$	2.9 (1.0)	10.35 (1.3)	29.8 (1.8)
CC _{1/2}	0.22	0.25	0.48
<i>R</i> _{int}	0.387 (0.885)	0.369 (0.950)	0.124 (0.905)
Overall <i>B</i> factor from Wilson plot (Å ²)	27.40	31.70	24.99

150 nl crystallant were equilibrated against 50 µl reservoir solution in SD-2 sitting-drop plates (IDEX). The crystallization plates were incubated at 293 K and were automatically inspected at regular intervals. Successful crystallization of construct 1 NA was achieved by mixing construct 1 NA with 100 mM HEPES buffer pH 7.0, 10% PEG 6000 as a precipitant for 48 h. Crystallization information is summarized in Table 2.

2.3. Data collection and processing

X-ray diffraction data sets for single crystals of N9 NA with the TB stalk (N9 + TB stalk), N9 NA without the TB stalk (N9 – TB stalk) and egg-grown virus N9 NA in well solution with ~15% (v/v) glycerol as a cryoprotectant were collected on the MX1/MX2 beamlines at the Australian Synchrotron. The data sets were processed with *HKL-2000* (Otwinowski & Minor, 1997). For the N9 + TB stalk data the Britton plot, *H*-test and maximum-likelihood estimate obtained by *phenix.xtriage* from the *PHENIX* suite (Adams *et al.*, 2010) suggested pseudo-merohedral twin domains ($-h, k, -l$; $h, -k, -l$) with average twin-fraction estimates of 0.865 and

0.135, which were subsequently refined. Further data-collection and processing statistics are given in Table 3.

2.4. Structure solution and refinement

The positions of the N9 head domains and the tetrameric stalk domains were identified in the asymmetric units by molecular replacement with *Phaser* (McCoy *et al.*, 2007) using the structures of N9 (PDB entry 1nnn; Varghese *et al.*, 1997) and TB (PDB entry 1fe6; Stetefeld *et al.*, 2000) without N-linked glycans and waters. Two independent tetrameric N9–TB assemblies were identified in the asymmetric unit of the N9 + TB stalk crystal and one N9 molecule was identified in the asymmetric unit of the N9 – TB stalk and egg-grown virus N9 crystals. The structures with protein molecules alone were refined and N-linked glycans were then added at the predicted sites (Asn75 in TB and Asn86, Asn146 and Asn200 in N9 using N2 numbering) along with water molecules. Iterative refinement and model building were conducted using *REFMAC* (Murshudov *et al.*, 2011) and *Xfit/MIFit* (Rigaku; McRee, 1999). Twin fractions ($-h, k, -l$) of 0.844 and ($h, -k, -l$) of 0.156 were refined using amplitudes from the N9 + TB stalk data set. The progress of the refinement was monitored

Table 4
Structure solution and refinement.

Values in parentheses are for the outer shell.

	N9 + TB stalk	N9 – TB stalk	N9, egg-grown virus
PDB code	6crd	6mcx	6d3b
Resolution range (Å)	40.04–2.57 (2.63–2.57)	45.22–2.30 (2.35–2.30)	33.07–1.40 (1.44–1.40)
Completeness (%)	77.6	98.1	99.0
σ Cutoff	1.0	1.3	1.8
No. of reflections, working set	108611 (2590)	21167 (1453)	92499 (6811)
No. of reflections, test set	5654 (168)	1155 (80)	4866 (348)
Final R_{cryst}	0.201 (0.221)	0.147 (0.278)	0.152 (0.244)
Final R_{free}	0.306 (0.376)	0.210 (0.283)	0.170 (0.259)
No. of non-H atoms			
Protein	27648	3095	3067
Ion	0	1	1
Ligand	1394	129	171
Water	933	365	725
Total	29975	3590	3964
R.m.s. deviations			
Bonds (Å)	0.007	0.009	0.015
Angles (°)	1.78	1.73	1.88
Average B factors (Å ²)			
Protein	25.4	49.9	12.8
Ion	0	53.2	14.1
Ligand	36.5	69.3	26.5
Water	20.3	57.3	30.4
Ramachandran plot			
Most favoured (%)	87.4	95.6	96.1
Allowed (%)	10.6	4.4	3.9

using the R_{free} statistics based on a test set encompassing 5% of the observed diffraction amplitudes (Brünger, 1992). PDB deposition codes and further data-processing details are given in Table 4.

3. Results and discussion

The crystal structure of the tetrameric N9 + TB domains shows well characterized folds for both protein domains, as reported previously (Varghese *et al.*, 1997; Stetefeld *et al.*, 2000; Fig. 2). The observable model contains two amino acids (Gly32 and Ser33) from the linker followed by 48 (residues 34–81 in chains *A*, *B*, *E* and *F*) or 47 (residues 35–81 in chains *C*, *D*, *G* and *H*) belonging to the TB domain and 388 belonging to the N9 domain (corresponding to residues 82–468 in N2). The N-linked glycans have the following structures: GlcNAc at residue 75 (TB), GlcNAc₂ at residue 86 (N9), GlcNAc₂-Man₂ at residue 146 (N9) and GlcNAc₂-Man₇ at residue 200 (N9). There are an additional three and one sugar units in the glycan chains attached to the N^δ atoms of Asn146 when compared with the high-resolution egg-grown virus N9 structure at 1.4 Å resolution and the published N9 structure (PDB entry 4gdr; Venkatramani *et al.*, 2012) at 1.55 Å resolution, respectively. N9 – TB stalk has the following PNGaseF-trimmed glycan structures: GlcNAc at residue 86, GlcNAc at residue 146 and

Figure 2
Tetrameric structure of N9 with the TB domain. (a) Front view with the tetrameric helical (TB) domain representing the N9 N-terminal TB stalk domain; (b) top view. This figure was produced using *PyMOL* (v.1.3r1; Schrödinger).

GlcNAc₂-Man₅ at residue 200. It is worth noting that all of the structures have similar high-mannose-type N-glycans (She *et al.*, 2017; Parsons *et al.*, 2017) regardless of the production system.

All eight N9 chains of N9 + TB stalk superposed very well with the N9 – TB stalk structure, with an average r.m.s.d. of 0.359 Å, while chain *A* of the original TB structure (PDB entry 1fe6) superposed with all eight TB domains with an average r.m.s.d. of 0.822 Å. There were no significant

structural differences, including the active site, in the N9 head in the N9 + TB stalk and N9 – TB stalk structures from recombinant proteins (Fig. 3). The N9 – TB stalk structure superposed extremely well with the egg-grown virus N9 head structure in this report (r.m.s.d. of 0.141 Å) as well as with the published structure (PDB entry 7nn9; Varghese *et al.*, 1997; r.m.s.d. of 0.171 Å). All eight N9 chains of N9 + TB stalk superposed with egg-grown virus N9 with an average r.m.s.d. of 0.374 Å. Analysis of the two tetrameric biological units of

Figure 3

Superposition of monomeric N9 heads from recombinant N9 + TB stalk (orange), recombinant N9 – TB stalk (cyan) and egg-grown virus N9 (purple) tetramers. (a) Front view and (b) top view. (c) Superposition of two (cyan and orange) tetrameric heads in the asymmetric unit of N9 + TB stalk shows the flexibility of the TB domain. The thickness of the ribbons is scaled to the *B*-factor values. This figure was produced using *PyMOL*.

the N9 + TB stalk crystal shows that one molecule (chains C, D, G and H) is somewhat distorted and the fourfold rotation deviates from exact fourfold symmetry by up to 0.3°. The superposition of two tetrameric N9 heads resulted in an r.m.s.d. of 0.602 Å and revealed flexibility of the tetrameric TB domains (Fig. 3c). The observed disorder of the N9 heads and the flexibility of the TB domain may explain the crystal twinning and the lower symmetry and resolution of the N9 + TB stalk structure. The superposition was performed with *SUPERPOSE/GESAMT* (Krissinel, 2012) from the *CCP4* package (Winn *et al.*, 2011) using all 388 residues of the N9 domain, and 50 (chains A, B, E and F) and 47 (chains C, D, G and H) of the 52 residues of the TB domain (PDB entry 1fe6; 48 amino-acid residues of TB plus four amino-acid residues of cloning artefacts) resolved in the N9 + TB stalk structure.

All previous structures of influenza NAs have utilized the head proteolytically cleaved from either a natural or artificial stalk for crystallization. Although construct 2 had a thrombin cleavage site, the actual cleavage of the NA head from the TB stalk was very inefficient. This contrasts with construct 1, where thrombin cleavage of the FLAG tag from the base of the TB stalk (Table 1 and Fig. 1a) was very efficient. Structural analysis of the sugar residues provides an insight into the possible reasons for this inefficient cleavage. Both the base of the NA head and the top of the TB stalk are glycosylated at residues Asn75 (TB) and Asn86 (N9) (Fig. 4; the electron density for these glycans is shown in Supplementary Fig. S5). Even though there is a seven-amino-acid linker in construct 2 (Table 1 and Fig. 1b), the sugars from both NA and TB could sterically hinder access of the thrombin to its cleavage site. PNGaseF treatment of the NA did not improve the thrombin cleavage of the TB stalk; this agrees with the obtained

structure, which showed that the glycosylation was still present after incubation with PNGaseF. Many of the other recombinant NA constructs have an additional SPRS sequence between the thrombin cleavage site and the NA head (Xu *et al.*, 2008). This may fortuitously facilitate the access of thrombin to the head–stalk cleavage site.

In summary, our results suggest that the TB domain used to generate stable, soluble, recombinant NA had no significant impact on the structure of the NA head, including the catalytic centre. Given the structural similarities of all influenza NA subtypes, the TB stalk might therefore be a good choice to stabilize some of the more fragile recombinant NAs used in biological screening assays.

Acknowledgements

We acknowledge the use of the Australian Synchrotron protein crystallography MX beamlines and the CSIRO Collaborative Crystallisation Centre (C3), Victoria, Australia.

Funding information

Parts of this study were supported by grant R01A1062721 from the NIAID. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* **D66**, 213–221.
- Baker, A. T., Varghese, J. N., Laver, W. G., Air, G. M. & Colman, P. M. (1987). *Proteins*, **2**, 111–117.
- Baranovich, T., Vongphrachanh, P., Ketmayoon, P., Sisouk, T., Chomlasack, K., Khanthamaly, V., Nguyen, H. T., Mishin, V. P., Marjuki, H., Barnes, J. R., Garten, R. J., Stevens, J., Wentworth, D. E. & Gubareva, L. V. (2017). *Emerg. Infect. Dis.* **23**, 686–690.
- Blick, T. J., Tiong, T., Sahasrabudhe, A., Varghese, J. N., Colman, P. M., Hart, G. J., Bethell, R. C. & McKimm-Breschkin, J. L. (1995). *Virology*, **214**, 475–484.
- Blok, J. & Air, G. M. (1982). *Biochemistry*, **21**, 4001–4007.
- Blok, J., Air, G. M., Laver, W. G., Ward, C. W., Lilley, G. G., Woods, E. F., Roxburgh, C. M. & Inglis, A. S. (1982). *Virology*, **119**, 109–121.
- Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.
- Burmeister, W. P., Ruigrok, R. W. & Cusack, S. (1992). *EMBO J.* **11**, 49–56.
- Castrucci, M. R. & Kawaoka, Y. (1993). *J. Virol.* **67**, 759–764.
- Colman, P. M., Hoyne, P. A. & Lawrence, M. C. (1993). *J. Virol.* **67**, 2972–2980.
- Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. & Barton, G. J. (1998). *Bioinformatics*, **14**, 892–893.
- Dai, M., Guo, H., Dortmans, J. C., Dekkers, J., Nordholm, J., Daniels, R., van Kuppeveld, F. J., de Vries, E. & de Haan, C. A. (2016). *J. Virol.* **90**, 9457–9470.
- Els, M. C., Air, G. M., Murti, K. G., Webster, R. G. & Laver, W. G. (1985). *Virology*, **142**, 241–247.
- Gubareva, L. V., Sleeman, K., Guo, Z., Yang, H., Hodges, E., Davis, C. T., Baranovich, T. & Stevens, J. (2017). *J. Infect. Dis.* **216**, S566–S574.
- Itzstein, M. von, Wu, W.-Y., Kok, G. B., Pegg, M. S., Dyason, J. C., Jin, B., Van Phan, T., Smythe, M. L., White, H. F., Oliver, S. W., Colman,

Figure 4

Construct 1 showing that both the base of the NA head and the top of the TB stalk are glycosylated at residues Asn75 (TB) and Asn86 (N9). Sugars from both ends may hinder the access of thrombin to the cleavage site in construct 2 despite the presence of an additional seven-amino-acid linker. This figure was produced using *PyMOL*.

- 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.
- Iuliano, A. D., Roguski, K. M., Chang, H. H., Muscatello, D. J., Palekar, R., Tempia, S., Cohen, C., Gran, J. M., Schanzer, D., Cowling, B. J., Wu, P., Kyncl, J., Ang, L. W., Park, M., Redlberger-Fritz, M., Yu, H., Espenhain, L., Krishnan, A., Emukule, G., van Asten, L., Pereira da Silva, S., Aungkulanon, S., Buchholz, U., Widdowson, M. A., Bresee, J. S., Azziz-Baumgartner, E., Cheng, P., Dawood, F., Foppa, I., Olsen, S., Haber, M., Jeffers, C., MacIntyre, C. R., Newall, A. T., Wood, J. G., Kundi, M., Popow-Kraupp, T., Ahmed, M., Rahman, M., Marinho, F., Sotomayor Proschle, C. V., Vergara Mallegas, N., Luzhao, F., Sa, L., Barbosa-Ramírez, J., Sanchez, D. M., Gomez, L. A., Vargas, X. B., Acosta Herrera, A., Llanés, M. J., Fischer, T. K., Krause, T. G., Mølbak, K., Nielsen, J., Trebbien, R., Bruno, A., Ojeda, J., Ramos, H., an der Heiden, M., Del Carmen Castillo Signor, L., Serrano, C. E., Bhardwaj, R., Chadha, M., Narayan, V., Kosen, S., Bromberg, M., Glatman-Freedman, A., Kaufman, Z., Arima, Y., Oishi, K., Chaves, S., Nyawanda, B., Al-Jarallah, R. A., Kuri-Morales, P. A., Matus, C. R., Corona, M. E. J., Burmaa, A., Darmaa, O., Obtel, M., Cherkaoui, I., van den Wijngaard, C. C., van der Hoek, W., Baker, M., Bandaranayake, D., Bissielo, A., Huang, S., Lopez, L., Newbern, C., Flem, E., Grøneng, G. M., Hauge, S., de Cosío, F. G., de Moltó, Y., Castillo, L. M., Cabello, M. A., von Horoch, M., Medina Osis, J., Machado, A., Nunes, B., Rodrigues, A. P., Rodrigues, E., Calomfirescu, C., Lupulescu, E., Popescu, R., Popovici, O., Bogdanovic, D., Kostic, M., Lazarevic, K., Milosevic, Z., Todorovic, B., Chen, M., Cutter, J., Lee, V., Lin, R., Ma, S., Cohen, A. L., Treurnicht, F., Kim, W. J., Delgado-Sanz, C., de Mateo Ontañón, S., Larrauri, A., León, I. L., Vallejo, F., Born, R., Junker, C., Koch, D., Chuang, J., Huang, W., Kuo, H., Tsai, Y., Bundhamcharoen, K., Chittaganpitch, M., Green, H. K., Pebody, R., Goñi, N., Chiparelli, H., Brammer, L. & Mustaquim, D. (2018). *Lancet*, **391**, 1285–1300.
- Johnson, G. D. & Jiang, W. (2005). *Arch. Biochem. Biophys.* **444**, 7–14.
- Kim, J.-H., Resende, R., Wennekes, T., Chen, H.-M., Bance, N., Buchini, S., Watts, A. G., Pilling, P., Streltsov, V. A., Petric, M., Liggins, R., Barrett, S., McKimm-Breschkin, J. L., Niikura, M. & Withers, S. G. (2013). *Science*, **340**, 71–75.
- Kongkamnerd, J., Milani, A., Cattoli, G., Terregino, C., Capua, I., Beneduce, L., Gallotta, A., Pengo, P., Fassina, G., Miertus, S. & De-Eknamkul, W. (2012). *J. Enzyme Inhib. Med. Chem.* **27**, 5–11.
- Krissinel, E. (2012). *J. Mol. Biochem.* **1**, 76–85.
- Laver, W. G. (1978). *Virology*, **86**, 78–87.
- Liu, S., Ji, K., Chen, J., Tai, D., Jiang, W., Hou, G., Chen, J., Li, J. & Huang, B. (2009). *PLoS One*, **4**, e5022.
- Mather, K. A., White, J. F., Hudson, P. J. & McKimm-Breschkin, J. L. (1992). *Virus Res.* **26**, 127–139.
- 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.
- McKimm-Breschkin, J. L., Barrett, S., Pilling, P. A., Hader, S., Watts, A. G. & Streltsov, V. A. (2018). *J. Med. Chem.* **61**, 1921–1933.
- McKimm-Breschkin, J. L., Caldwell, J. B., Guthrie, R. E. & Kortt, A. A. (1991). *J. Virol. Methods*, **32**, 121–124.
- McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
- 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. D* **67**, 355–367.
- Oakley, A. J., Barrett, S., Peat, T. S., Newman, J., Streltsov, V. A., Waddington, L., Saito, T., Tashiro, M. & McKimm-Breschkin, J. L. (2010). *J. Med. Chem.* **53**, 6421–6431.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Parsons, L. M., An, Y., de Vries, R. P., de Haan, C. A. & Cipollo, J. F. (2017). *J. Proteome Res.* **16**, 398–412.
- Peters, J., Nitsch, M., Köhlhorn, B., Golbik, R., Lupas, A., Kellermann, J., Engelhardt, H., Pfander, J.-P., Müller, S., Goldie, K., Engel, A., Stetter, K.-O. & Baumeister, W. (1995). *J. Mol. Biol.* **245**, 385–401.
- Potier, M., Mameli, L., Bélisle, M., Dallaire, L. & Melançon, S. B. (1979). *Anal. Biochem.* **94**, 287–296.
- Schmidt, P. M., Attwood, R. M., Mohr, P. G., Barrett, S. A. & McKimm-Breschkin, J. L. (2011). *PLoS One*, **6**, e16284.
- She, Y. M., Farnsworth, A., Li, X. & Cyr, T. D. (2017). *Sci. Rep.* **7**, 10232.
- Smith, B. J., Huyton, T., Joosten, R. P., McKimm-Breschkin, J. L., Zhang, J.-G., Luo, C. S., Lou, M.-Z., Labrou, N. E. & Garrett, T. P. J. (2006). *Acta Cryst. D* **62**, 947–952.
- Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Engel, J. & Kammerer, R. A. (2000). *Nature Struct. Biol.* **7**, 772–776.
- Varghese, J. N., Colman, P. M., van Donkelaar, A., Blick, T. J., Sahasrabudhe, A. & McKimm-Breschkin, J. L. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 11808–11812.
- Varghese, J. N., McKimm-Breschkin, J. L., Caldwell, J. B., Kortt, A. A. & Colman, P. M. (1992). *Proteins*, **14**, 327–332.
- Varghese, J. N., Smith, P. W., Sollis, S. L., Blick, T. J., Sahasrabudhe, A., McKimm-Breschkin, J. L. & Colman, P. M. (1998). *Structure*, **6**, 735–746.
- Venkatramani, L., Johnson, E. S., Kolavi, G., Air, G. M., Brouillette, W. J. & Mooers, B. H. (2012). *BMC Struct. Biol.* **12**, 7.
- Ward, C. W., Colman, P. M. & Laver, W. G. (1983). *FEBS Lett.* **153**, 29–33.
- Ward, C. W., Elleman, T. C. & Azad, A. A. (1982). *Biochem. J.* **207**, 91–95.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst. D* **67**, 235–242.
- Wright, C. E. & Laver, W. G. (1978). *J. Mol. Biol.* **120**, 133–136.
- Xu, X., Zhu, X., Dwek, R. A., Stevens, J. & Wilson, I. A. (2008). *J. Virol.* **82**, 10493–10501.
- Yang, H., Carney, P. J., Mishin, V. P., Guo, Z., Chang, J. C., Wentworth, D. E., Gubareva, L. V. & Stevens, J. (2016). *J. Virol.* **90**, 5770–5784.