Acta Crystallographica Section F
Structural Biology and Crystallization Communications
ISSN 1744-3091

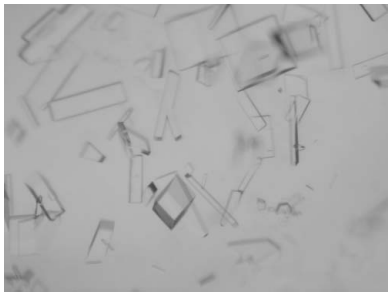
Crystallization of the C-terminal head domain of the avian adenovirus CELO long fibre

Pablo Guardado Calvo,[a] Antonio L. Llamas-Saiz,[b] Patrick Langlois[c] and Mark J. van Raaij[a,b]*

[a]Departamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain, [b]Unidad de Difracción de Rayos X, Laboratorio Integral de Dinámica y Estructura de Biomoléculas José R. Carracido, Edificio CACTUS, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain, and [c]Agence Francaise de Securité Sanitaire des Aliments, Unité Génétique Virale et Biosecurité, Site Les Croix, BP 53, F-22440 Plouſragan, France

Correspondence e-mail: vanraaij@usc.es

Received 7 March 2006
Accepted 3 April 2006

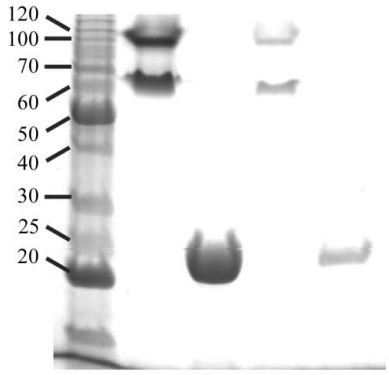


Avian adenovirus CELO contains two different fibres: fibre 1, the long fibre, and fibre 2, the short fibre. The short fibre is responsible for binding to an unknown avian receptor and is essential for infection of birds. The long fibre is not essential, but is known to bind the coxsackievirus and adenovirus receptor protein. Both trimeric fibres are attached to the same penton base, of which each icosahedral virus contains 12 copies. The short fibre extends straight outwards, while the long fibre emerges at an angle. The carboxy-terminal amino acids 579–793 of the avian adenovirus long fibre have been expressed with an amino-terminal hexahistidine tag and the expressed trimeric protein has been purified by nickel-affinity chromatography and crystallized. Crystals were grown at low pH using PEG 10 000 as precipitant and belonged to space group $C2$. The crystals diffracted rotating-anode Cu $K\alpha$ radiation to at least 1.9 Å resolution and a complete data set was collected from a single crystal to 2.2 Å resolution. Unit-cell parameters were $a = 216.5$, $b = 59.2$, $c = 57.5$ Å, $\beta = 101.3^\circ$, suggesting one trimer per asymmetric unit and a solvent content of 46%. The long fibre head does not have significant sequence homology to any other protein of known structure and molecular-replacement attempts with known fibre-head structures were unsuccessful. However, a map calculated using SIRAS phasing shows a clear trimer with a shape similar to known adenovirus fibre-head structures. Structure solution is in progress.

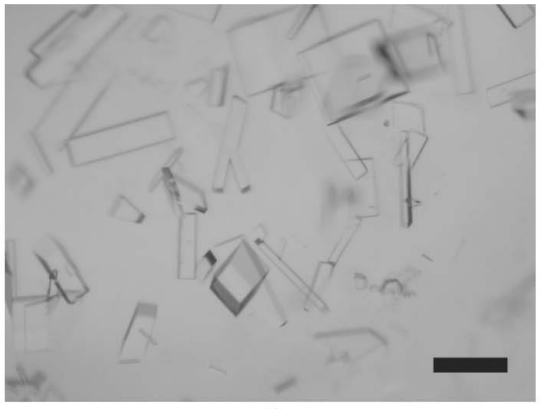
1. Introduction

Avian adenovirus CELO (chick embryo lethal orphan virus) or fowl adenovirus type 1 (FAdV-1; FAV-1) is a non-enveloped virus (Laver *et al.*, 1971; McCracken & Adair, 1993). It infects poultry but is not associated with serious pathologies or economic losses; when experimentally introduced in chickens, it does not cause any evident disease (Cowen *et al.*, 1978). However, this lack of danger has drawn interest in using CELO virus as a gene therapy (Kelleher & Vos, 1994; Stevenson *et al.*, 2006) and anti-cancer vehicle (Logunov *et al.*, 2004; Shashkova *et al.*, 2005) in humans and as a vaccination vehicle for birds (Francois *et al.*, 2004). The virus has a dsDNA genome of 44 kbp encased by an icosahedral capsid (Chiocca *et al.*, 1996). In analogy to the human adenoviruses, the capsid is made up of the hexon and penton proteins, with further minor proteins involved in stabilizing the capsid (Fabry *et al.*, 2005); the hexons form the faces and the pentons the vertices. Each pentameric vertex contains two fibre proteins, the short fibre (fibre 2) emerging straight and the long fibre (fibre 1) at an angle (Hess *et al.*, 1995). Transfection studies have shown that the long fibre binds the coxsackie and adenovirus receptor protein CAR, while the short fibre probably binds an unknown avian receptor (Tan *et al.*, 2001). The short fibre is essential for infection of chicken cells, while the long fibre is not, making it an attractive target for modification.

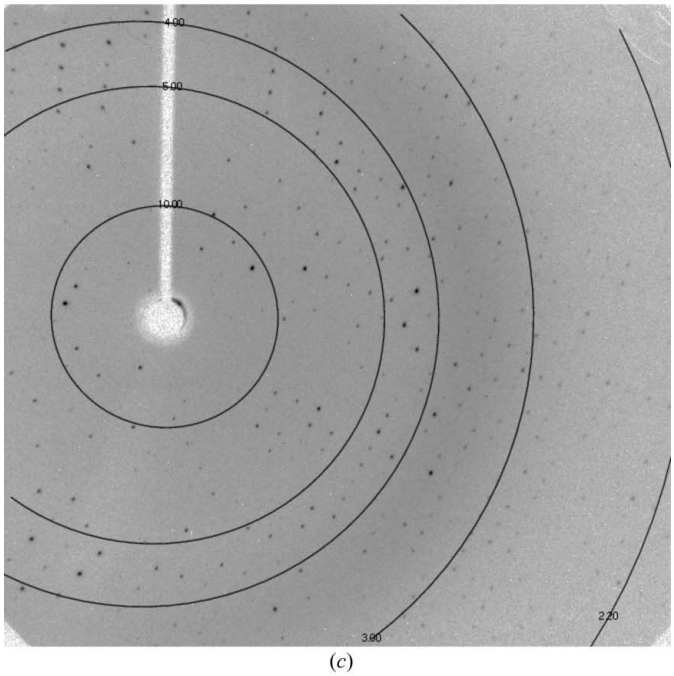
The long-fibre gene encodes 793 amino acids and has two regions consisting of multiple glycine residues (Hess *et al.*, 1995). The 12 consecutive glycines near the amino-terminus (Gly51–Gly62) are presumably important for flexibility; this flexibility may be necessary for accommodating the long fibre as one of two fibres attached to the same penton and may be responsible for the fact that the long fibre emerges from the viral surface at an angle. Crystallographic studies of the human adenovirus penton in the presence of fibre peptides



(*a*)



(*b*)



(*c*)

Figure 1
(*a*) SDS–PAGE of purified fibre-head protein. Lanes from left to right: molecular-weight markers with their sizes indicated in kDa, high-concentration fibre head without boiling, high-concentration fibre head boiled, low-concentration fibre head without boiling and low-concentration fibre head boiled. (*b*) Crystals of the avian adenovirus long-fibre head. In conditions containing high-molecular-weight PEG and at low pH crystals of diverse shape grow. All crystals analysed belonged to space group *C*2 with nearly identical unit-cell parameters. For more details, see §2. The size bar corresponds to around 0.25 mm. (*c*) Diffraction image. A 1° rotation diffraction image is shown with resolution rings superimposed. Exposure time was 90 s; Cu $K\alpha$ radiation from a rotating-anode source was used.

suggest five fibre-binding sites are available in each penton (Zubieta *et al.*, 2005); therefore, it is possible that one of the amino-termini of the long or short fibres does not bind to the penton. It is possible that the short fibre occupies three binding sites and the long fibre only two. The central domain of the long fibre (residues 72–572) contains repeats compatible with the triple β-spiral fold (Chiocca *et al.*, 1996; van Raaij, Mitraki *et al.*, 1999). These repeats are characterized by the consensus sequence $XX\varphi X\varphi X\varphi X$-1-$X$*$\varphi X\varphi XX$-2, where X is any amino acid, φ is generally hydrophobic and * is generally a proline or glycine, although other small amino acids are also tolerated in this position. At position 1, a central β-turn, insertions of a few amino acids are permitted; at position 2 more extensive insertions are possible as surface loops. The central domain forms the slim central fibre shaft. A second sequence of four consecutive glycines is present near the carboxy-terminal end (amino acids 576–579), possibly forming a flexible region linking the triple β-spiral shaft domain to the head domain.

Here, we report the expression, purification and crystallization of a carboxy-terminal fragment of the avian adenovirus long fibre consisting of amino acids 579–793 containing an amino-terminal hexahistidine and T7 purification tag.

2. Methods

A DNA fragment encoding residues 579–793 of the avian adenovirus long fibre (UniProtKB/TrEMBL accession No. Q64787; Hess *et al.*, 1995) was produced by the polymerase chain reaction and cloned into the expression vector pET28c+ (Novagen, Darmstadt, Germany). The resultant plasmid pET28-CELOlongfib579-793 encodes the mentioned long-fibre fragment fused to an N-terminal purification tag containing a hexahistidine tag and a T7 tag (the N-terminal 11 residues of the bacteriophage T7 gene 10 product). The sequence of the insert was confirmed by DNA-sequence analysis (Sistemas Genómicos, Valencia, Spain).

For expression, *Escherichia coli* strain JM109(DE3) (Promega, Madrid, Spain) was freshly transformed with the plasmid and four 0.7 l cultures were grown aerobically at 310 K to an optical density of 0.6–0.8 measured at 600 nm. Expression was induced by adding 1 m*M* IPTG (isopropyl thio-β-D-galactopyranoside) and was allowed to continue for 3 h. Harvested cells were resuspended in 40 ml cold PBS buffer (4.29 m*M* disodium hydrogen phosphate, 1.47 m*M* potassium dihydrogen phosphate, 2.7 m*M* potassium chloride, 137 m*M* sodium chloride) and frozen at 253 K. Bacteria were lysed by a double pass through an emulsifier (Avestin emulsifier C5, Avestin Europe GmbH, Mannheim, Germany). After removing insoluble material by centrifugation, 3 ml Ni–NTA resin (Qiagen, Valencia, CA, USA) was added. The suspension was incubated for 1 h at 277 K and poured into an empty column. The resin was washed with PBS buffer and elution was performed with a step gradient of imidazole pH 7 in PBS buffer (steps of 0.01, 0.05, 0.125, 0.25, 0.5 and 1.0 *M* imidazole). The protein eluted at imidazole concentrations of 0.125–0.5 *M*. The 0.25 and 0.5 *M* imidazole fractions contained virtually pure protein; the flowthrough and fractions up to 0.125 *M* imidazole were pooled and the procedure was repeated, again keeping the 0.25 and 0.5 *M* imidazole fractions.

For crystallization, the pooled fractions were dialysed against ME buffer (10 m*M* MES–NaOH pH 6.5, 1 m*M* EDTA) and applied onto an Uno-S6 cation-exchange column (Biorad, Madrid, Spain) equilibrated in the same buffer. The protein eluted in several peaks between 0.2 and 0.4 *M* NaCl in a linear gradient of 0.0–1.0 *M* NaCl in ME buffer. The fractions all contained pure protein as judged by

Table 1
Crystallographic data statistics.

Values in parentheses are for the highest resolution bin.

Space group	*C*2
Unit-cell parameters (Å, °)	a = 216.45, b = 59.19, c = 57.49, β = 101.3
No. of trimers per ASU†	1
Matthews coefficient† (Å^3 Da^{-1})	2.28
Resolution range (Å)	45.0–2.2 (2.3–2.2)
No. of unique reflections‡	36555 (4586)
Multiplicity	12.2 (5.2)
Completeness (%)	99.6 (96.8)
$\langle I/\sigma(I)\rangle$	27.0 (8.2)
$R_{\rm sym}$§ (%)	7.4 (20.4)

† Estimation based on likely solvent content. ‡ No sigma cutoff or other restrictions were used for inclusion of observed reflections. § $R_{\rm sym} = \sum_h \sum_i |I_{hi} - \langle I_h\rangle| / \sum_h \sum_i |I_{hi}|$, where I_{hi} is the intensity of the ith measurement of the same reflection and $\langle I_h\rangle$ is the mean observed intensity for that reflection.

denaturing gel electrophoresis and were thus pooled. The protein was concentrated to 23 mg ml^{-1} using Centricon concentrators (Millipore, Madrid, Spain), incorporating three washes with TE buffer to eliminate small-molecule impurities.

Crystallization took place by vapour diffusion in sitting-drop CompactClover plates (Jena Biosciences, Jena, Germany) with 0.2 ml reservoirs and drops consisting of 2 µl protein solution mixed with 2 µl reservoir solution. Reservoirs contained 10–12%(*w*/*v*) PEG 10 000, 0.1 *M* HEPES–NaOH pH 4.7–6.1 and 0–10%(*v*/*v*) glycerol; crystallization plates were incubated at 278 K. For data collection, crystals were passed through reservoir solution containing 25%(*v*/*v*) glycerol, flash-frozen to 100 K in a nitrogen stream and maintained at 100 K. Crystallographic data were integrated and scaled using the *HKL*2000 package (Otwinowski & Minor, 1997). Self-rotation functions were calculated using *GLRF* (Tong & Rossmann, 1997).

3. Results and discussion

Structural information about the head domains of several human adenovirus serotypes is available. The first to be solved was the type 5 fibre head (Xia *et al.*, 1994), followed by the type 2 (van Raaij, Louis *et al.*, 1999), type 12 (Bewley *et al.*, 1999), type 3 (Durmort *et al.*, 2001), types 37 and 19p (Burmeister *et al.*, 2004) and the type 41 short-fibre head (Seiradake & Cusack, 2005). They all have identical topology between serotypes and each monomer contains a β-sandwich. The intact heads consist of three β-sandwiches, while the orientation of the monomers in the trimers is also very similar between serotypes.

Both avian short- and long-fibre heads have very limited sequence homology (less than 20% identity) with human adenovirus fibre heads, although it is of course still possible and perhaps likely that they have a similar structure. Furthermore, the long-fibre head probably binds the same coxsackie and adenovirus receptor protein (Tan *et al.*, 2001) as do many human adenovirus fibre heads (groups A and C–F; Roelvink *et al.*, 1998). In order to determine the structure of the head of avian adenovirus long fibre and obtain more information on its function, we have cloned amino acids 579–793, the putative globular head domain, into the bacterial expression vector pET28c+ and obtained soluble protein with an N-terminal His-T7 tag.

The protein could be purified by nickel-affinity chromatography and was shown to be trimeric by gel electrophoresis. Taking advantage of the extraordinary stability of the adenovirus fibre trimers, we compared the electrophoretic mobility of heated and unheated samples avian adenovirus long-fibre head in the presence of sodium dodecylsulfate (Fig. 1*a*). The results showed that the avian adenovirus long-fibre head domain retains its trimeric structure when the sample is not boiled, although two discrete trimeric bands are observed, one probably corresponding to a compact trimer and a slower migrating band corresponding to a partially unfolded 'hydra', similar to the case of native human adenovirus type 2 fibre (Mitraki *et al.*, 1999). Avian reovirus fibre expressed with the same expression and purification tag (van Raaij *et al.*, 2005; Guardado Calvo *et al.*, 2005) as that employed in this work shows the same behaviour (unpublished results).

For crystallization, the protein was further purified by cation-exchange chromatography. Yields of purified concentrated protein were around 3 mg from 2.8 l culture (assuming that the absorbance at 280 nm of a 1 mg ml^{-1} protein solution is 1.3, using the theoretical extinction coefficient from the amino-acid sequence of the expressed protein). Crystals were obtained under acidic conditions in the presence of high-molecular-weight PEG (Fig. 1*b*). Using a Bruker–Nonius FR591 rotating-anode generator equipped with a kappa-CCD2000 detector, we were able to obtain diffraction to 1.9 Å resolution (Fig. 1*c*) and a complete highly redundant data set to 2.2 Å resolution was collected from one crystal (0.28 × 0.13 × 0.05 mm) flash-frozen at 100 K (Table 1). A total of 1473 1° oscillation images were collected in 12 runs. Data could be scaled in the monoclinic space group *C*2; the crystals are most likely to contain one trimer in each asymmetric unit, which would give a solvent content of 46% and a Matthews coefficient of 2.3 Å^3 Da^{-1} (Matthews, 1968).

The self-rotation function is consistent with threefold non-crystallographic symmetry; two clear peaks are present in the κ = 120° section, interrelated by the crystallographic twofold axis. The calculated coordinates of the self-rotation peaks are φ = 77.45, ψ = 56.80° and the corresponding twofold-related peak φ = 102.55, ψ = 123.20°, using polar angles and the *XZK* convention (Fitzgerald, 1988). We were not successful in molecular-replacement attempts using the known structures of human adenovirus head domains, presumably owing to their low sequence identity (less than 20%).

However, a map calculated at 2.3 Å using SIRAS phasing on a mercury derivative shows a clear trimer with a shape similar to known adenovirus fibre-head structures. Structure solution and refinement are in progress; details will be published elsewhere. The structure should reveal the extent of structure homology of the avian adenovirus long-fibre head domain with other adenovirus fibre heads. We also hope to obtain other clues from the structure, such as which regions may be involved in receptor binding.

We thank Guy Schoehn for electron microscopy of the purified protein, Xosé Lois Hermo Parrado and Rebeca Menaya Vargas for technical assistance, Stefan Miller (Profos AG, Regensburg, Germany) for providing crystallization plates and José Martínez-Costas and Javier Benavente for sharing laboratory facilities. This research was sponsored by research grants BMC2005-02974 from the Spanish Ministry of Education and Science and PGIDIT03P-XIC20307PN from the Xunta de Galicia; both the grants and the X-ray apparatus on which the data were collected were co-financed by the European Regional Development Fund programme.

References

Bewley, M. C., Springer, K., Zhang, Y. B., Freimuth, P. & Flanagan, J. M. (1999). *Science*, **286**, 1579–1583.
Burmeister, W. P., Guilligay, D., Cusack, S., Wadell, G. & Arnberg, N. (2004). *J. Virol.* **78**, 7727–7736.
Chiocca, S., Kurzbauer, R., Schaffner, G., Baker, A., Mautner, V. & Cotten, M. (1996). *J. Virol.* **70**, 2939–2949.

Cowen, B., Calnek, B. W., Menendez, N. A. & Ball, R. F. (1978). *Avian Dis.* **22**, 459–470.
Durmort, C., Stehlin, C., Schoehn, G., Mitraki, A., Drouet, E., Cusack, S. & Burmeister, W. P. (2001). *Virology*, **285**, 302–312.
Fabry, C. M., Rosa-Calatrava, M., Conway, J. F., Zubieta, C., Cusack, S., Ruigrok, R. W. & Schoehn, G. (2005). *EMBO J.* **24**, 1645–1654.
Fitzgerald, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.
Francois, A., Chevalier, C., Delmas, B., Eterradossi, N., Toquin, D., Rivallan, G. & Langlois, P. (2004). *Vaccine*, **22**, 2351–2360.
Guardado Calvo, P., Fox, G. C., Hermo Parrado, X. L., Llamas-Saiz, A. L., Costas, C., Martínez-Costas, J., Benavente, J. & van Raaij, M. J. (2005). *J. Mol. Biol.* **354**, 137–149.
Hess, M., Cuzange, A., Ruigrok, R. W., Chroboczek, J. & Jacrot, B. (1995). *J. Mol. Biol.* **252**, 379–385.
Kelleher, Z. T. & Vos, J. M. (1994). *Biotechniques*, **17**, 1110–1117.
Laver, W. G., Younghusband, H. B. & Wrigley, N. G. (1971). *Virology*, **45**, 598–614.
Logunov, D. Y., Ilyinskaya, G. V., Cherenova, L. V., Verhovskaya, L. V., Shmarov, M. M., Chumakov, P. M., Kopnin, B. P. & Naroditsky, B. S. (2004). *Gene Ther.* **11**, 79–84.
McCracken, R. M. & Adair, B. M. (1993). *Viral Infections of Vertebrates*, edited by J. B. McFerran & M. S. McNulty, Vol. 3, pp. 123–144. Amsterdam: Elsevier.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Mitraki, A., Barge, A., Chroboczek, J., Andrieu, J. P., Gagnon, J. & Ruigrok, R. W. (1999). *Eur. J. Biochem.* **264**, 599–606.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Raaij, M. J. van, Louis, N., Chroboczek, J. & Cusack, S. (1999). *Virology*, **262**, 333–343.
Raaij, M. J. van, Mitraki, A., Lavigne, G. & Cusack, S. (1999). *Nature (London)*, **401**, 935–938.
Raaij, M. J. van, Hermo Parrado, X. L., Guardado Calvo, P. G., Fox, G. C., Llamas-Saiz, A. L., Costas, C., Martinez-Costas, J. & Benavente, J. (2005). *Acta Cryst.* F**61**, 651–654.
Roelvink, P. W., Lizonova, A., Lee, J. G., Li, Y., Bergelson, J. M., Finberg, R. W., Brough, D. E., Kovesdi, I. & Wickham, T. J. (1998). *J. Virol.* **72**, 7909–7915.
Seiradake, E. & Cusack, S. (2005). *J. Virol.* **79**, 14088–14094.
Shashkova, E. V., Cherenova, L. V., Kazansky, D. B. & Doronin, K. (2005). *Cancer Gene Ther.* **12**, 617–626.
Stevenson, M., Boos, E., Herbert, C., Hale, A., Green, N., Lyons, M., Chandler, L., Ulbrich, K., van Rooijen, N., Mautner, V., Fisher, K. & Seymour, L. (2006). *Gene Ther.* **13**, 356–368.
Tan, P. K., Michou, A. I., Bergelson, J. M. & Cotten, M. (2001). *J. Gen. Virol.* **82**, 1465–1472.
Tong, L. & Rossmann, M. G. (1997). *Methods Enzymol.* **276**, 594–611.
Xia, D., Henry, L. J., Gerard, R. D. & Deisenhofer, J. (1994). *Structure*, **2**, 1259–1270.
Zubieta, C., Schoehn, G., Chroboczek, J. & Cusack, S. (2005). *Mol. Cell*, **17**, 121–135.