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

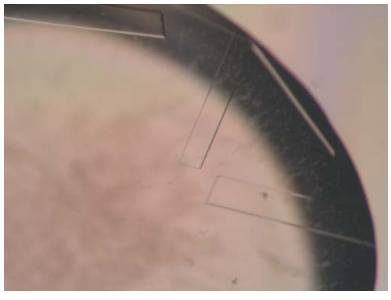
Crystallization of the head and galectin-like domains of porcine adenovirus isolate NADC-1 fibre

Pablo Guardado-Calvo,[a] Antonio L. Llamas-Saiz,[b] Gavin C. Fox,[c] Joel N. Glasgow[d] and Mark J. van Raaij[a,e]*

[a]Departamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Spain, [b]Unidad de Rayos X (RIAIDT), Laboratorio Integral de Dinámica y Estructura de Biomoléculas José R. Carracido, Edificio CACTUS, Universidad de Santiago de Compostela, Spain, [c]Laboratoire des Proteines Membranaires, Institut de Biologie Structurale J. P. Ebel, 41 Rue Jules Horowitz, Grenoble, France, [d]Divisions of Cardiology and Human Gene Therapy, Departments of Medicine, Obstetrics and Gynecology, Pathology, Surgery and the Gene Therapy Center, University of Alabama at Birmingham, Birmingham, Alabama, USA, and [e]Departamento de Biología Estructural, Instituto de Biología Molecular (IBMB-CSIC), Parc Cientific, Baldiri Reixac 10, 08028 Barcelona, Spain

Correspondence e-mail: mark.vanraaij@ibmb.csic.es

Received 21 July 2009
Accepted 21 September 2009



The porcine adenovirus NADC-1 isolate, a strain of porcine adenovirus type 4, has a fibre with an atypical architecture. In addition to a classical virus-attachment region, shaft and head domains, it contains an additional galectin-like domain C-terminal to the head domain and connected to the head domain by a long RGD-containing loop. The galectin-like domain contains two putative carbohydrate-recognition domains. The head and galectin-like domains have been independently crystallized. Diffraction data have been obtained to 3.2 Å resolution from crystals of the head domain and to 1.9 Å resolution from galectin-like domain crystals.

1. Introduction

Adenoviruses may find application as vectors for gene therapy and cancer therapy and as vaccination agents (for recent reviews, see Arnberg, 2009; Bachtarzi *et al.*, 2008; Thacker *et al.*, 2009). The family *Adenoviridae* contains over a hundred known serotypes, including 52 that infect humans. They are non-enveloped viruses with a linear double-stranded DNA genome that can infect all five major vertebrate classes. They have an icosahedral ($T = 25$) capsid consisting of three major proteins: the trimeric hexon, forming the facets of the particle, the pentameric penton base, which forms the vertices, and the trimeric fibre protein, which extends from the penton base at the vertex positions. Proteins IIIa, VI, VIII and IX reinforce the penton–hexon and hexon–hexon interactions (Nemerow *et al.*, 2009). The principal structural determinant of adenovirus tropism is its protruding rod-like fibre protein (reviewed in Noureddini & Curiel, 2005; Nicklin *et al.*, 2005). The distal tip of each fibre is composed of a globular knob domain, which serves as the major viral attachment site for a variety of cellular receptors such as CAR (Roelvink *et al.*, 1998), CD46/80/86 (Short *et al.*, 2006) and sialic acid (Arnberg *et al.*, 2000). An RGD motif localized in the hypervariable loop of the penton base of many adenoviruses binds to integrins (Wickham *et al.*, 1993; Li *et al.*, 2001), which act as secondary receptors, promoting viral internalization. Given the importance of the fibre, structural characterization of different fibre proteins from human and nonhuman adenoviruses has been critical in understanding adenovirus tropism and in developing new vectors with modified tropisms (Nicklin & Baker, 2008). Animal adenoviruses are of particular interest in this regard, as they may be less immunogenic to humans and have novel receptor-binding properties. Recently, structures of canine and fowl adenovirus fibre heads have been published (Seiradake *et al.*, 2006; Guardado-Calvo *et al.*, 2007; El Bakkouri *et al.*, 2008).

Porcine adenovirus was isolated for the first time by Haig *et al.* (1964), is classified within the genus *Mastadenovirus* and has a genome of approximately 34 kb. It is commonly regarded as a low-grade pathogen. Restriction-enzyme mapping and sequence analysis of serotypes 1–5 has shown that serotypes 1–3 are closely related to each other while serotypes 4 and 5 are more divergent. All of the serotypes are distinct from human, murine, canine, bovine and fowl adenoviruses (Hammond & Johnson, 2005). The porcine adenovirus NADC-1 isolate was isolated from a tonsillar swab of an adult sow in 1972 and is thought to be a strain of porcine adenovirus type 4 (Kleiboeker *et al.*, 1993). Its 703-residue fibre protein is unique

among adenovirus strains and contains, from the N-terminus to the C-terminus, a tail domain (residues 1–37) where the interaction sequence with the penton protein is found, a short shaft domain (residues 38–120) with up to six predicted triple β-spiral repeats (van Raaij *et al.*, 1999) and what was predicted to be a larger than average head domain (residues 121–703; Kleiboeker, 1995; Fig. 1). The first 167 amino acids of this domain (121–287) are homologous to other adenoviral head domains; therefore, we propose to designate this as the head domain and differentiate the sequence C-terminal to this as follows. The sequence adjacent to the head domain (288–392) contains an RGD sequence and alanine-rich and glutamate-rich sequences that are likely to be involved in interaction with integrins. Additionally, a domain composed of two carbohydrate-recognition domains (CRDs) repeated in tandem was predicted from sequence analysis (residues 393–681). This galectin-like domain is unique among adenovirus fibres; only one other galectin-like domain has been identified in viruses (in lymphocystis disease virus; Cooper, 2002). The porcine adenovirus galectin-like domain is composed of two carbohydrate-recognition domains linked by a 23-residue sequence rich in prolines. The position of the putative CRDs at the C-terminus of the fibre suggests that the NADC-1 virus may employ interactions with cell-surface galactosides as a method of cell attachment.

Here, we describe the expression, purification, crystallization and crystallographic data collection of two protein-construct crystals containing the head (116–291) and galectin-like domains (393–703) of porcine adenovirus NADC-1.

2. Methods

2.1. Construction of expression vectors

Porcine adenovirus NADC-1 (catalogue No. 030-PDV) was obtained from the National Veterinary Services Laboratories of the USDA Animal and Plant Health Inspection Service (Ames, Iowa). The NADC-1 fibre gene sequence was amplified directly from the viral stock using the polymerase chain reaction (PCR). DNA fragments encoding porcine adenovirus NADC-1 fibre (UniProt accession code Q83467) residues 116–291 (the predicted head domain; Fig. 1) and 393–703 (the predicted galectin-like domain) were produced by PCR from a plasmid (pKan-NMOvPAd4) which encodes the shaft, head and galectin-like domains of the fibre gene and cloned into the expression vector pET28c(+) (Novagen/Merck, Nottingham, England) between the *Bam*HI and *Not*I restriction sites. The start codon of the expression vector pET28c(+) was used; a stop codon was inserted after residue 291 for the head-domain expression clone; the galectin-like domain, being the C-terminal domain, ends at residue 703, the last residue of the porcine adenovirus type 4 fibre. The resultant plasmids pET28c-pAd4Fib116–291 and pET28c-pAd4Fib393–703 encode the head and galectin-like domains, respectively, fused to an N-terminal purification tag containing six consecutive histidine residues and a T7 tag (MGSSH HHHHH SSGLV PRGSH MASMT GGQQG RI; Novagen; the T7 tag is an 11-amino-acid peptide encoded in the leader sequence of T7 bacteriophage gene10 for which specific monoclonal antibodies and antibody-containing chromatography resins are available). The sequences of all inserts were confirmed by DNA-sequence analysis (Sistemas Genómicos, Valencia, Spain).

2.2. Protein expression and purification

Escherichia coli strain BL21 (DE3) was freshly transformed with pET28c-pAd4Fib116–291 or pET28c-pAd4Fib393–703 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. In the case of the head domain, growth was allowed to continue for 3 h at 310 K after induction with 1 m*M* isopropyl β-D-1-thiogalactopyranoside. For expression of the galectin-like domain, cultures were cooled on ice before induction and growth was allowed to continue for 16 h at 289 K. Harvested cells were resuspended in 40 ml of a buffer containing 50 m*M* Tris–HCl pH 8.5, 1 *M* sodium chloride and frozen at 253 K. Bacteria were lysed by a double pass through an emulsifier (Avestin EmulsiFlex C5, Avestin Europe GmbH, Mannheim, Germany). After removing insoluble material by centrifugation (20 min at 20 000*g*), 4 ml Ni–NTA resin (Qiagen, Valencia, California, USA) was added. The suspension was incubated for 1 h at 277 K and poured into an empty column and the resin was washed using the same buffer. For purification of the head domain, elution was performed with a step gradient of imidazole pH 7.0 (10, 20, 50, 100, 200, 400, 600, 800 and 1000 m*M* imidazole). Pure histidine-T7-tagged head domain eluted at high imidazole concentrations (200–800 m*M*), consistent with the proteins being trimeric and thus containing three histidine tags.

For purification of the galectin domain, elution was performed with the following step gradient: 10, 20, 50, 100, 200, 500 and 1000 m*M* imidazole pH 7.0. Histidine-T7-tagged galectin-like domain eluted at imidazole concentrations of between 50 and 500 m*M*. The eluted protein was dialysed against 10 m*M* Tris–HCl pH 8.5, 1 m*M* EDTA (TE buffer pH 8.5) and loaded onto an Uno-Q6 anion-exchange

```
  1 MKRSVPSDFN PVYPYDYQPI SLMPAFYDNY GFHEGPSGVL SLNIANPLGY
 51 TPRKKLCLKL GEGLALDSDG HLRVQIPDMQ AQPPLLYQGH RLSLLFDADA
101 GFHLTEDGAL SLTKTLVYPT LWTGPAPEAN VTFSGENSPS GILRLCLSRT
151 GGTVIGTLSV QGSLTNPSTG QTLGMNLYFD ADGNVLSESN LVRGSWGMKD
201 QDTLVTPIAN GQYLMPNLTA YPRLIQTLTS SYIYTQAHLD HNNSVVDIKI
251 GLNTDLRPTA AYGLSFTMTF TNSPPTSFGT DLVQFGYLGQ DSSPSFLREL
301 PLASEAGYFG KLAAASEEMP APPEAQTQDQ AAEEPPAPAE AEAPAPAEAE
351 AEAEPPRKPP RGDLAALYNR VHSDTRAEDT PTSPELVTTL PDPFVLPLPD
401 GVPTGASIVL EGTLTPSAVF FTLDLVTGPA SLALHFNVRL PLEGEKHIVC
451 NSREGSSNWG EEVRPQEFPF EREKPFVLVI VIQSDTYQIT VNGKPLVDFP
501 QRLQGITRAS LSGDLVFTRL TMYPPGDPRP TTLLPPPAAP LDVIPDAYVL
551 NLPTGLTPRT LLTVTGTPTP LAEFFIVNLV YDLHYDSKNV ALHFNVGFTS
601 DSKGHIACNA RMNGTWGSEI TVSDFPFQRG KPFTLQILTR EADFQVLVDK
651 QPLTQFQYRL KELDQIKYVH MFGHVVQTHL EHQVPDTPVF STAGVSKVYP
701 QIL
```

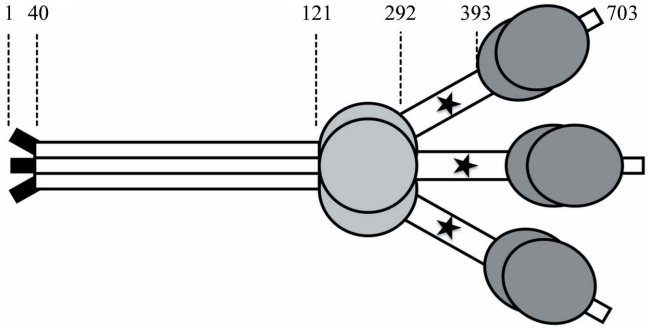


Figure 1
Predicted domain organization of porcine adenovirus type 4 NADC-1 strain fibre. The amino-acid sequence (single-letter code; above) and a schematic drawing (below) are shown. The predicted virus-binding tail (residues 1–37) is shown in black, the shaft domain (40–120) in white, the predicted head domain (121–287) in light grey, the RGD-containing domain (292–392) in white and the predicted galectin-like domain (393–681) in dark grey. The putative integrin-binding RGD sequence is indicated in bold in the sequence and with an asterisk in the drawing. The sequences expressed and crystallized (116–291 and 393–703) are underlined; they correspond to the head and galectin-like domains plus a few N-terminal and C-terminal amino acids. It should be noted that in the absence of electron-microscopy data the relative orientations and dimensions of the domains are unknown.

Table 1
Crystallographic data statistics of porcine adenovirus NADC-1 isolate fibre crystals.

Values in parentheses are for the highest resolution bin (where applicable).

	Head domain	Galectin-like domain
Radiation source	Cu $K\alpha$ (Bruker–Nonius FR591)	ID14-2 (ESRF)
Temperature (K)	295	100
Wavelength (Å)	1.5418	0.9390
Detector	Kappa CCD-2000	ADSC Q4 CCD
Crystal-to-detector distance (mm)	90.0	301.1
Oscillation range (°)	0.6	0.8
No. of images	166	272
Space group	$P2_12_12_1$	$C2$
Unit-cell parameters (Å, °)	a = 125.7, b = 145.4, c = 147.6, $\alpha = \beta = \gamma$ = 90.0	a = 167.6, b = 77.3, c = 94.1, $\alpha = \gamma$ = 90.0, β = 101.5
Mosaicity (°)	0.36	0.68
Unique observations	45035 (4367)	89027 (12129)
Resolution range (Å)	40–3.2 (3.32–3.2)	40–1.9 (2.0–1.9)
Multiplicity	3.8 (3.8)	4.1 (3.8)
Completeness (%)	98.9 (97.5)	96.1 (90.1)
$\langle I/\sigma(I)\rangle$	17.5 (2.9)	10.0 (3.7)
$R_{\rm merge}$† (%)	7.2 (30.7)	7.5 (35.2)
$V_{\rm M}$‡ (Å^3 Da^{-1})	6.0	2.2

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$, where $Ii(hkl)$ is the intensity of the ith measurement of the same reflection and $\langle I(hkl)\rangle$ is the mean observed intensity for that reflection. ‡ According to Matthews (1968) and assuming two domain trimers and four copies of the galectin domain in the asymmetric unit.

column (Bio-Rad, Madrid, Spain). The protein did not bind and eluted in the flowthrough. Fractions containing galectin-domain protein were pooled, concentrated to 1.0 ml using Centricon concentrators (Milipore, Madrid, Spain) and loaded onto a Sephacryl S100 gel-filtration column (GE Healthcare, Alcobendas, Spain) equilibrated with TE buffer pH 8.5 containing 150 m*M* sodium chloride.

2.3. Crystallization and data collection

Crystallization took place by vapour diffusion in sitting-drop CompactClover plates (Jena Biosciences, Jena, Germany), with 0.15 ml reservoirs and drops consisting of 4–5 µl protein solution mixed with equal volumes of reservoir solution. The head-domain protein was concentrated to 40–60 mg ml^{-1} using Centricon concentrators, incorporating three washes with TE buffer pH 8.0 to eliminate small-molecule impurities. 1 *M* lithium sulfate supplemented with polyethylene glycol was used as the reservoir solution for crystallization; equivalent crystals were obtained over the range 1–10%(*w*/*v*) polyethylene glycol 8000. Ligand-free galectin-domain protein was concentrated to 10–20 mg ml^{-1} using Centricon concentrators, incorporating three washes with TE buffer pH 8.0. Galectin-like domain crystals were obtained using 25%(*w*/*v*) polyethylene glycol 3350, 500 m*M* sodium nitrate as reservoir solution and with 2 m*M* sialic acid added to the drop as an additive (although carbohydrate-array experiments showed that the galectin-like domain does not bind sialic acid; results to be published elsewhere). For data collection, galectin-like domain crystals were transferred to reservoir solution containing 20%(*v*/*v*) glycerol, keeping the concentrations of the precipitant components constant. For both head-domain and galectin-like domain protein preparations, the final buffer was TE buffer pH 8, although traces of imidazole or sodium chloride may remain, respectively. We did not check the final pH; owing to the low concentration of buffering agent the pH may have changed upon the addition of precipitant solution and/or cryoprotectant.

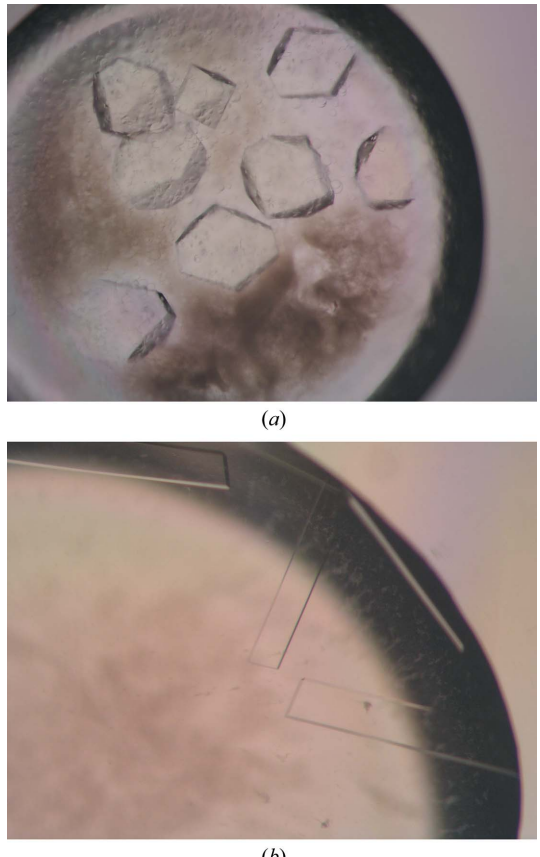
2.4. Computation

Prediction of structural disorder was performed with the programs *FOLDINDEX* (Prilusky *et al.*, 2005) and *RONN* (Yang *et al.*, 2005). Crystallographic data were integrated and scaled using the *HKL*-2000 suite of programs (Otwinowski & Minor, 1997; head-domain crystals) or by a combination of *MOSFLM* (Leslie, 2006) and *SCALA* (Collaborative Computational Project, Number 4, 1994). Data-processing statistics are summarized in Table 1. Self-rotation functions were calculated with the program *XPREP* (Sheldrick & Schneider, 2001).

3. Results and discussion

The fibre of the porcine adenovirus NADC-1 isolate is predicted to be composed of an N-terminal virus-binding peptide that is likely to be disordered in the absence of the rest of the virus, a relatively short shaft region, a head domain and a predicted galectin-like domain (Fig. 1). *In silico* analysis predicts the region between the knob and the galectin-like domains (amino acids 292–392) to be flexible. Consistent with this, truncated versions of the protein containing residues 292–392 either did not form crystals or the crystals obtained did not diffract. The loop of the human adenovirus type 2 penton that contains the RGD sequence is also highly flexible and this flexibility may be related to its biological function (Zubieta *et al.*, 2005). We did obtain diffraction-quality crystals of the head domain and of the galectin-like domain independently.

For the head domain (amino acids 116–291), large crystals were obtained (Fig. 2*a*). A variety of different cryoprotection solutions



(*a*)

(*b*)

Figure 2
Crystals of (*a*) head-domain and (*b*) galectin-domain crystals, measuring approximately 0.5 × 0.5 × 0.5 and 0.4 × 0.1 × 0.02 mm, respectively.

were tried without success (oil-based methods were not tried). However, a complete data set was collected to 3.2 Å resolution at room temperature (Table 1). The self-rotation function (not shown) showed the presence of two independent threefold noncrystallographic axes, suggesting the presence of two trimers in the asymmetric unit and a solvent content of 79.3% (Matthews, 1968). Crystals of the galectin-like domain appeared after 1–2 weeks (Fig. 2*b*). A complete data set was collected to 1.9 Å resolution (Table 1). Four copies of the galectin domain are expected in the asymmetric unit.

Structure solution by molecular replacement has been performed; refinement and analysis are ongoing and will be reported elsewhere. The structures will provide insight into the receptor and carbohydrate ligand-binding properties of the porcine adenovirus NADC-1 isolate fibre. These structural data may also be useful for designing mutants that bind other sugars and chimeric adenoviruses targeting specific cell types (Paul *et al.*, 2008). If suitable cryo-electron microscopy images can be obtained, the high-resolution crystal structures will be fitted into EM maps to produce an overall picture of the relative organization of the domains in the intact fibre.

We thank Patricia Ferraces-Casais for excellent technical assistance. This research was sponsored by research grants BFU2005-02974, BFU2005-24982-E and BFU2008-01588 and by a predoctoral FPU fellowship to PG-C from the Spanish Ministry of Education and Science. This work was also supported by funds from the European Commission under contracts ERAS-CT-2003-980409 (as part of the European Science Foundation EUROCORES Programme EuroSCOPE) and NMP4-CT-2006-033256 (BeNatural coordinated project) and by the Xunta de Galicia (grant INCITE08E1R203013ES).

References

Arnberg, N. (2009). *Rev. Med. Virol.* **19**, 165–178.
Arnberg, N., Edlund, K., Kidd, A. H. & Wadell, G. (2000). *J. Virol.* **74**, 42–48.
Bachtarzi, H., Stevenson, M. & Fisher, K. (2008). *Expert Opin. Drug Deliv.* **5**, 1231–1240.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cooper, D. N. W. (2002). *Biochim. Biophys. Acta*, **1572**, 209–231.
El Bakkouri, M., Seiradake, E., Cusack, S., Ruigrok, R. W. & Schoehn, G. (2008). *Virology*, **378**, 169–176.
Guardado-Calvo, P., Llamas-Saiz, A. L., Fox, G. C., Langlois, P. & van Raaij, M. J. (2007). *J. Gen. Virol.* **88**, 2407–2416.
Haig, D. A., Clarke, M. C. & Pereira, M. S. (1964). *J. Comp. Pathol.* **74**, 81–84.
Hammond, J. M. & Johnson, M. A. (2005). *Vet. J.* **169**, 17–27.
Kleiboeker, S. B. (1995). *Virus Res.* **39**, 299–309.
Kleiboeker, S. B., Seal, B. S. & Mengeling, W. L. (1993). *Arch. Virol.* **133**, 357–368.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Li, E., Brown, S. L., Stupack, D. G., Puente, X. S., Cheresh, D. A. & Nemerow, G. R. (2001). *J. Virol.* **75**, 5405–5409.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Nemerow, G. R., Pache, L., Reddy, V. & Stewart, P. L. (2009). *Virology*, **384**, 380–388.
Nicklin, S. A. & Baker, A. H. (2008). *Mol. Ther.* **16**, 1904–1905.
Nicklin, S. A., Wu, E., Nemerow, G. R. & Baker, A. H. (2005). *Mol. Ther.* **12**, 384–393.
Noureddini, S. C. & Curiel, D. T. (2005). *Mol. Pharm.* **2**, 341–347.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Paul, C. P., Everts, M., Glasgow, J. N., Dent, P., Fisher, P. B., Ulasov, I. V., Lesniak, M. S., Stoff-Khalili, M. A., Roth, J. C., Preuss, M. A., Dirven, C. M., Lamfers, M. L., Siegal, G. P., Zhu, Z. B. & Curiel, D. T. (2008). *Cancer Biol. Ther.* **7**, 786–793.
Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. & Sussman, J. L. (2005). *Bioinformatics*, **21**, 3435–3438.
Raaij, M. J. van, Mitraki, A., Lavigne, G. & Cusack, S. (1999). *Nature (London)*, **401**, 935–938.
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., Lortat-Jacob, H., Billet, O., Kremer, E. J. & Cusack, S. (2006). *J. Biol. Chem.* **281**, 33704–33716.
Sheldrick, G. M. & Schneider, T. R. (2001). *Methods in Macromolecular Crystallography*, edited by D. Turk & L. Johnson, pp. 72–81. Amsterdam: IOS Press.
Short, J. J., Vasu, C., Holterman, M. J., Curiel, D. T. & Pereboev, A. (2006). *Virus Res.* **122**, 144–153.
Thacker, E. E., Timares, L. & Matthews, Q. L. (2009). *Expert Rev. Vaccines*, **8**, 761–777.
Wickham, T. J., Mathias, P., Cheresh, D. A. & Nemerow, G. R. (1993). *Cell*, **73**, 309–319.
Yang, Z., Thomson, R., McNeil, P. & Esnouf, R. M. (2005). *Bioinformatics*, **21**, 3369–3376.
Zubieta, C., Schoehn, G., Chroboczek, J. & Cusack, S. (2005). *Mol. Cell*, **17**, 121–135.