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

Crystallization of the C-terminal domain of the bacteriophage T7 fibre protein gp17

Carmela Garcia-Doval[a,b] and Mark J. van Raaij[a]*

[a]Departamento de Estructuras de Macromoleculas, Centro Nacional de Biotecnologia (CNB–CSIC), Calle Darwin 3, 28049 Madrid, Spain, and [b]Departamento de Bioquimica y Biologia Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Vida, 15705 Santiago de Compostela, Spain

Correspondence e-mail: mjvanraaij@cnb.csic.es

Received 10 August 2011
Accepted 28 November 2011

Bacteriophage T7 attaches to its host using the C-terminal domains of its six fibres, which are trimers of the gp17 protein. A C-terminal fragment of gp17 consisting of amino acids 371–553 has been expressed, purified and crystallized. Crystals of two forms were obtained, belonging to space group $P2_12_12_1$ (unit-cell parameters a = 61.2, b = 86.0, c = 118.4 Å) and space group $C222_1$ (unit-cell parameters a = 68.3, b = 145.6, c = 172.1 Å). They diffracted to 1.9 and 2.0 Å resolution, respectively. Both crystals are expected to contain one trimer in the asymmetric unit. Multiwavelength anomalous dispersion phasing with a mercury derivative is in progress.

1. Introduction

Bacteriophages are the most abundant biological entities on earth and are second only to bacteria in mass (Wommack & Colwell, 2000). A large majority of bacteriophages (up to 96%; Ackermann, 2001) belong to the *Caudovirales* order and have a tail at a special vertex. The tail functions as a host-recognition and DNA-injection organelle. Tailed bacteriophages can be subdivided into three families: *Myoviridae*, with long contractile tails, *Siphoviridae*, with long noncontractile tails, and *Podoviridae*, with short, stubby, noncontractile tails (Maniloff & Ackermann, 1998). Bacteriophage T7, an archetypal member of the *Podoviridae* family, has a conical tail 20 nm in length. To this tail, six tail fibres are attached (Steven *et al.*, 1988; Fig. 1*a*). Each tail fibre is a trimer of gene product 17 (gp17) and is responsible for primary host binding, recognizing the *Escherichia coli* lipopolysaccharide (LPS) core heptose region (Picken & Beacham, 1977). Gene 17 codes for a protein containing 553 amino acids (Dunn & Studier, 1983; Fig. 1*b*).

Steven *et al.* (1988) performed electron microscopy on intact bacteriophage T7 tail complexes and on isolated fibres, combined with image averaging, mass measurements on unstained fibres and secondary- and tertiary-structure predictions. They predicted gp17 to contain an N-terminal phage binding domain consisting of residues 1–149 followed by a proximal thin rod-like domain (16 nm in length and 2 nm in width) starting at residue 151 and continuing to about residue 260 (Fig. 1*c*). This rod-like domain may contain a triple α-helical coiled coil. A kink separates the proximal domain from the distal domain, which was predicted to consist of residues 268–553 (the C-terminus). The distal domain consists of a linear succession of four nodules expected to have a high content of β-structure. The four nodules were tentatively assigned as residues 268–365, 366–432, 433–465 and 466–553, respectively, assuming that the polypeptide chain runs collinear with the fibre, with the N-terminus proximal to the phage and the C-terminus most distal (Fig. 1).

Here, we describe the expression, purification and crystallization of residues 371–553 of the bacteriophage T7 fibre protein gp17. The crystallized protein contains the three most C-terminal nodules identified by Steven *et al.* (1988). It is likely to contain the region of the protein involved in LPS binding, as Heineman *et al.* (2008) have identified two mutants of bacteriophage T7 gp17 involved in host

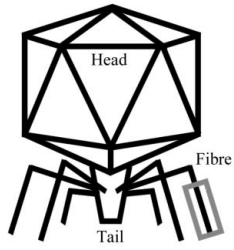
avoidance in this region: a gp17 Gly520 to Glu change in a T7 mutant adapted to avoid *E. coli* B and a Val544 to Ala change in a T7 mutant adapted to avoid *E. coli* K12. Furthermore, Garcia *et al.* (2003) identified a spontaneous host-range change mutant in the highly homologous (86% identical) phage PhiA1122 fibre in this region (Leu523 to Ser, which aligns with Ala518 of bacteriophage T7 gp17). We also report the collection of high-resolution native and multi-wavelength anomalous dispersion data sets from these crystals.

2. Methods

2.1. Protein expression and purification

Bacteriophage T7 DNA was obtained from Ana Cuervo and José L. Carrascosa of our institute (CNB–CSIC). The gp17 fibre gene was amplified from viral DNA using the polymerase chain reaction (PCR) and cloned into the plasmid pCR2.1-TOPO using topoisomerase I from vaccinia virus (TOPO TA Cloning Kit, Invitrogen, Life Technologies, Alcobendas, Madrid, Spain). A DNA fragment encoding the C-terminal third of the T7 fibre was produced by PCR from the TOPO-gp17 vector and cloned between the *Nco*I and *Hin*dIII restriction sites of the expression vector pET30a+ (Novagen, Merck, Madrid, Spain). We named the resultant plasmid pET30-17(371–553); it encodes residues 371–553 of the T7 fibre gp17 fused to an N-terminal purification tag consisting of residues MHHHH HHSSG LVPRG SGMKE TAAAK FERQH MDSPD LGTDD DDKAM (see also Fig. 1*b*). The sequence of the insert was confirmed by DNA sequence analysis (Sistemas Genomicos, Valencia, Spain).

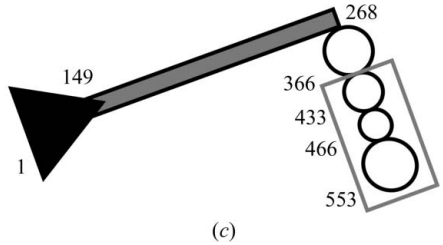
E. coli strain BL21 (DE3) was freshly transformed with pET30-17(371–553) and four 0.9 l cultures were grown aerobically at 310 K to an optical density of between 0.6 and 0.8 measured at 600 nm. Cultures were induced with 1 m*M* isopropyl-β-D-1-thiogalactopyranoside and growth was continued for 16 h at 289 K. Harvested cells were resuspended in 40 ml of a buffer consisting of 50 m*M* Tris–HCl pH 8.0, 4% glycerol, 50 m*M* ammonium chloride, 2 m*M* EDTA, 150 m*M* sodium chloride and frozen at 253 K. Bacteria were lysed by sonication. After removing insoluble material by centrifugation (30 min at 20 000*g*), 2 ml nickel–nitrilotriacetic acid resin (Jena Bioscience, Jena, Germany) was added. The suspension was poured into an empty column and the resin was washed using 50 m*M* Tris–HCl pH 8.5, 0.3 *M* sodium chloride buffer containing 50 m*M* imidazole. Elution was performed with a step gradient of imidazole in the same buffer (0.1, 0.15, 0.2, 0.25, 0.3, 0.4 and 1.0 *M*). The eluted protein was dialysed against 10 m*M* Tris–HCl pH 8.5 and loaded onto an Uno-Q12 quaternary ammonium strong anion-exchange column (Bio-Rad, Madrid, Spain). The protein was eluted with a linear gradient of



(*a*)

```
  1-MANVIKTVLT YQLDGSNRDF NIPFEYLARK FVVVTLIGVD RKVLTINTDY- 50
 51-RFATRTTISL TKAWGPADGY TTIELRRVTS TTDRLVDFTD GSILRAYDLN-100
101-VAQIQTMHVA EEARDLTTDT IGVNNDGHLD ARGRRIVNLA NAVDDRDAVP-150
151-FGQLKTMNQN SWQARNEALQ FRNEAETFRN QAEGFKNESS TNATNTKQWR-200
201-DETKGFRDEA KRFKNTAGQY ATSAGNSASA AHQSEVNAEN SATASANSAH-250
251-LAEQQADRAE READKLENYN GLAGAIDKVD GTNVYWKGNI HANGRLYMTT-300
301-NGFDCGQYQQ FFGGVTNRYS VMEWGDENGW LMYVQRREWT TAIGGNIQLV-350
351-VNGQIITQGG AMTGQLKLQN GHVLQLESAS DKAHYILSKD GNRNNWYIGR-400
401-GSDNNNDCTF HSYVHGTTLT LKQDYAVVNK HFHVGQAVVA TDGNIQGTKW-450
451-GGKWLDAYLR DSFVAKSKAW TQVWSGSAGG GVSVTVSQDL RFRNIWIKCA-500
501-NNSWNFFRTG PDGIYFIASD GGWLRFQIHS NGLGFKNIAD SRSVPNAIMV-550
551-ENE                                                  -553
```

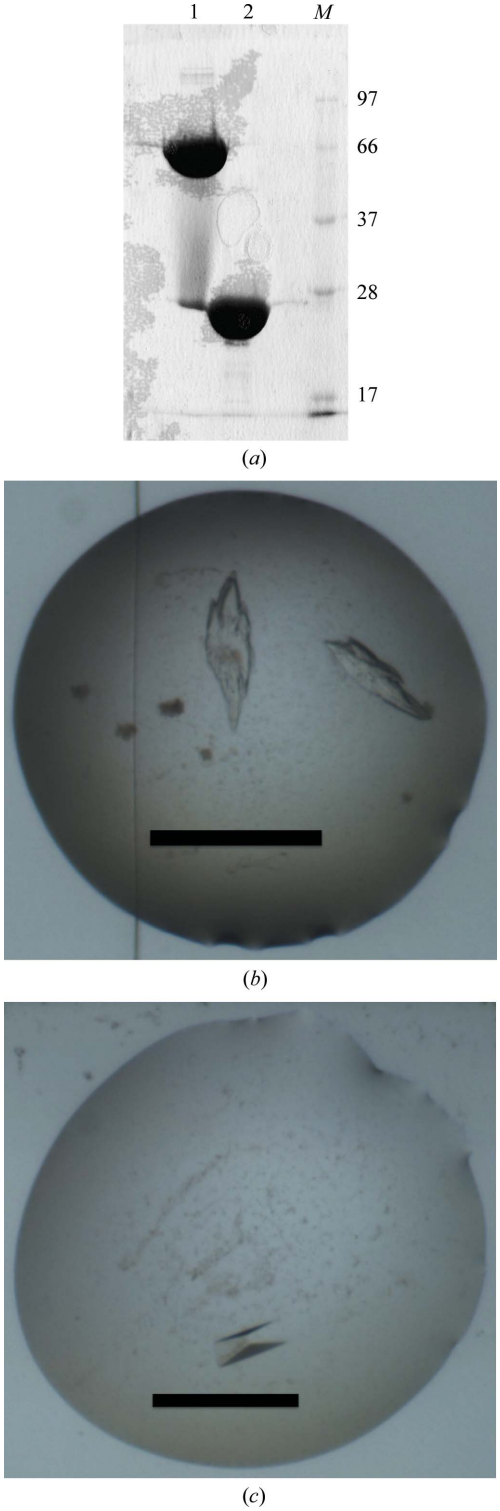
(*b*)



(*c*)

Figure 1
Predicted domain organization of bacteriophage T7 fibre. (*a*) Schematic drawing of bacteriophage T7. (*b*) Amino-acid sequence of gp17. (*c*) Schematic drawing of the fibre. The virus-binding domain (residues 1–149) is shown in black, the rod-like proximal domain (151–260) in grey and the distal domain (268–558) in white. The distal domain is divided into four nodules predicted to contain amino acids 268–365, 366–432, 433–465 and 466–553, respectively (Steven *et al.*, 1988). The C-terminal sequence expressed and crystallized is boxed in grey in parts (*a*) and (*c*) and is underlined in (*b*).

0–1 *M* sodium chloride and eluted at around 0.17 *M* sodium chloride. Fractions containing His-tagged gp17(371–553) were pooled and concentrated to 38 mg ml^{-1} using Amicon Ultra concentrators



(*a*)

(*b*)

(*c*)

Figure 2
Expression, purification and characterization of bacteriophage T7 gp17(371–553). (*a*) SDS–PAGE of purified protein. Lane 1, purified protein without previous heating of the sample; lane 2, protein after incubation at 368 K for 5 min; lane *M*, a mixture of molecular-weight marker proteins (their sizes are indicated in kDa on the right). (*b*, *c*) Crystals belonging to space groups $P2_12_12_1$ (*b*) and $C222_1$ (*c*). The scale bars measure approximately 0.5 mm.

Table 1
Crystallographic data statistics.

Values in parentheses are for the highest resolution bin.

	Native 1	Native 2
Space group	$P2_12_12_1$	$C222_1$
Unit-cell parameters (Å)	a = 61.2, b = 86.0, c = 118.4	a = 68.3, b = 145.6, c = 172.1
Resolution range (Å)	20.0–1.9 (2.00–1.90)	20.0–2.0 (2.11–2.00)
Reflections	50056 (7224)	58236 (8430)
Multiplicity	3.7 (3.7)	3.7 (3.7)
Completeness (%)	99.9 (100.0)	99.9 (100.0)
$\langle I/\sigma(I)\rangle$	6.3 (2.5)	6.0 (2.1)
$R_{\rm merge}$ (%)	9.3 (29.2)	9.7 (34.1)

MAD data set.

	Low-energy remote	Inflection point	Peak	High-energy remote
Wavelength (Å)	1.0121	1.0070	1.0056	0.9919
Unit-cell parameters† (Å)	a = 61.3, b = 86.0, c = 122.4	a = 61.3, b = 85.9, c = 122.4	a = 61.4, b = 86.1, c = 122.5	a = 61.4, b = 86.0, c = 122.4
Resolution range (Å)	20.0–2.7 (2.85–2.70)	20.0–2.7 (2.85–2.70)	20.0–2.7 (2.85–2.70)	20.0–2.7 (2.85–2.70)
Reflections	18382 (2652)	18373 (2651)	18420 (2641)	18392 (2644)
Multiplicity	7.3 (7.4)	7.3 (7.4)	7.3 (7.4)	7.3 (7.4)
Completeness (%)	99.8 (100.0)	99.8 (100.0)	99.8 (100.0)	99.8 (100.0)
$\langle I/\sigma(I)\rangle$	4.5 (1.6)	4.6 (1.7)	4.8 (2.1)	4.7 (2.1)
$R_{\rm merge}$‡ (%)	14.7 (46.2)	14.5 (44.8)	13.7 (35.6)	13.7 (35.9)

† Although the four data sets were collected from the same crystal, unit-cell parameters were refined independently for each data set during data processing. ‡ $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of the same reflection and $\langle I(hkl)\rangle$ is the mean observed intensity for that reflection.

(Millipore Iberica, Madrid, Spain), incorporating three washes with 10 m*M* Tris–HCl pH 8.5 to eliminate salt and impurities. The protein sample was stored frozen at 195 K.

Dynamic light-scattering measurements were performed on protein diluted to 0.5 mg ml^{-1} in 10 m*M* Tris–HCl pH 8.0 using a DynaPro Titan (Wyatt Technology, Santa Barbara, California, USA) following the procedures outlined in the user manual.

2.2. Crystallization and data collection

For high-throughput crystallization (HTX) experiments, protein samples were sent to the EMBL Grenoble Outstation and sitting-drop vapour-diffusion experiments were performed at 293 K using the methods described by Dimasi *et al.* (2007); 576 different conditions were tested from commercially available screens (Crystal Screen, Crystal Screen 2, Crystal Screen Lite, PEG/Ion, MembFac, Natrix, Quick Screen, Grid Screens Ammonium Sulfate, Malonate, PEG 6000, PEG/LiCl and MPD from Hampton Research, Aliso Viejo, California, USA, and JCSG+ and PACT screens from Nextal, Qiagen, Courtabeouf, France). The $P2_12_12_1$ crystals used for data collection were obtained by vapour diffusion in sitting-drop CrysChem plates (Hampton Research) with 0.5 ml reservoirs and drops consisting of 2 µl protein solution mixed with an equal volume of reservoir solution. The reservoir solution consisted of 0.1 *M* Tris–HCl pH 8.5, 0.2 *M* trimethylamine *N*-oxide and between 20 and 25%(*w*/*v*) polyethylene glycol monomethyl ether 2000. Crystals were directly flash-cooled in liquid nitrogen and used for data collection.

For crystallographic phase determination, crystals were soaked with heavy-atom compounds. Compounds were added as powders to the reservoir of a single crystallization trial to provide a saturated solution and were allowed to equilibrate with the drop overnight. The next morning, 1 µl of the saturated reservoir solution was added to the drop and the crystals were flash-cooled in liquid nitrogen after

several minutes, several hours and 1 d. Crystallographic data were collected on beamline BM14 at the European Synchrotron Radiation Facility (ESRF; Grenoble, France) using a MAR Research CCD 225 detector (MAR Research, Norderstedt, Germany). Crystals were maintained at 100 K during data collection. Crystallographic data were integrated with *MOSFLM* (Battye *et al.*, 2011) and scaled with *SCALA* (Evans, 2011). Further processing was performed using programs from *CCP*4 (Winn *et al.*, 2011). Self-rotation functions were calculated with *MOLREP* (Vagin & Teplyakov, 2010).

3. Results and discussion

Expression of full-length T7 fibre with N-terminal purification tags yielded between 3 and 6 mg purified protein per litre of bacterial culture. Full-length gp17 tended to aggregate in solution but, as revealed by electron-microscopy experiments, did show the expected kinked shape consisting of a thinner and a thicker rod connected at an angle. Taking into account secondary-structure predictions, disorder predictions and the predictions of Steven *et al.* (1988) with regards to mapping the structural features onto the amino-acid sequence, we decided to express four different N-terminal deletion mutants containing residues 281–553, 371–553, 431–553 and 461–553 (*i.e.* starting near the borders of the predicted domains and in predicted disordered loop regions). C-terminal deletion mutants were not made because we expect the C-terminal domain to be important for correct trimerization and folding, as observed for other virus fibres (Mitraki *et al.*, 2006). The C-terminal domain is also expected to contain the important receptor-binding region. For gp17(281–553) and gp17(461–553), such small amounts were present in bacterial cell extracts that

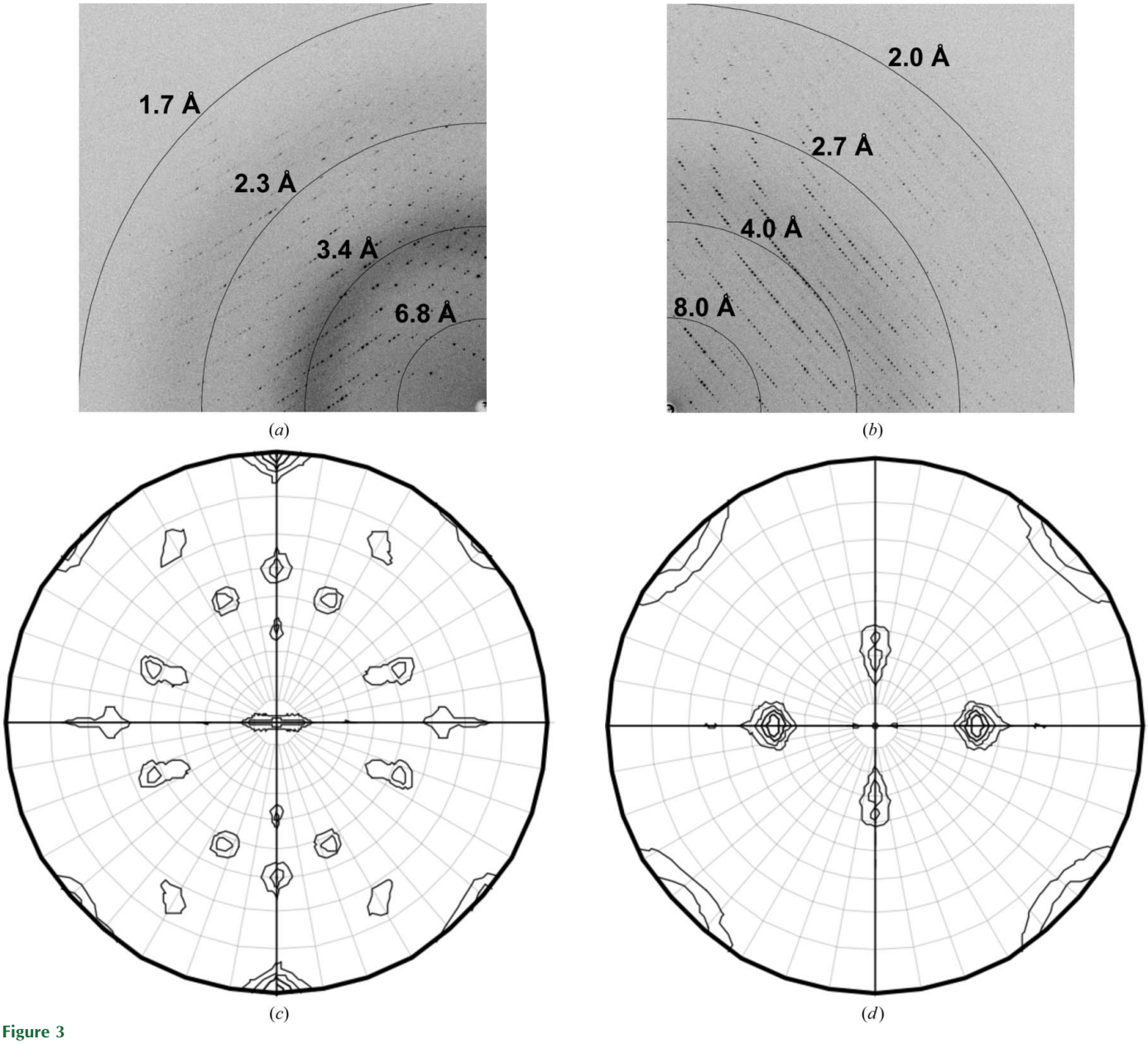


Figure 3
Crystallographic analysis. (*a*, *b*) Diffraction patterns of crystals belonging to space groups $P2_12_12_1$ (*a*) and $C222_1$ (*b*). Only a quarter of each image is shown for clarity. The beam centre is indicated by a cross-hair and resolution rings are shown. (*c*, *d*) Sections of the self-rotation function ($\chi = 120^\circ$) calculated using the $P2_12_12_1$ data (*c*) and the $C222_1$ data (*d*).

purification was not attempted. However, the other two fragments expressed decent amounts of soluble protein and we proceeded with only the largest of these, gp17(371–553).

For purification, we took advantage of the six-histidine tag encoded by the expression vector and used metal-affinity chromatography as described in §2. Histidine-tagged gp17(371–553) eluted at a relatively wide range of high imidazole concentrations (between 0.15 and 0.4 *M*), consistent with the protein being trimeric and thus containing three histidine tags. We have routinely observed similar imidazole ranges for the elution of N-terminally His-tagged trimeric C-terminal fragments of fibre proteins from the same metal resin: 0.2–0.5 *M* for avian reovirus sigmaC (van Raaij *et al.*, 2005), 0.125–0.5 *M* for avian adenovirus short fibre (Guardado-Calvo *et al.*, 2006), 0.05–0.5 *M* for porcine adenovirus fibre (Guardado-Calvo *et al.*, 2009) and 0.2–0.3 *M* for the bacteriophage T4 long tail fibre protein gp37 (Bartual *et al.*, 2010). Metal-affinity chromatography was followed by strong anion-exchange chromatography, in which the protein eluted as a single (albeit somewhat irregularly shaped) peak. After purification, buffer exchange and concentration, the yield was between 10 and 20 mg per litre of bacterial culture. The high solubility of the protein is exemplified by the possibility of concentrating it to at least 50 mg ml^{-1}, although slightly lower concentrations were used for crystallization experiments.

Comparison of migration properties on sodium dodecyl sulfate polyacrylamide gels with and without previous heat-treatment of the sample confirmed the protein to be trimeric (Fig. 2*a*), as observed for other virus and bacteriophage fibres (Mitraki *et al.*, 1999; Guardado-Calvo *et al.*, 2006; Bartual *et al.*, 2010). After heat treatment (5 min at 368 K) the tagged protein migrates at around its expected size of 25 kDa, while without heat treatment the majority of the protein migrates at an apparent molecular weight of around 65 kDa, which is most consistent with a trimer that stays partially folded under the normally denaturing conditions of electrophoresis in the presence of sodium dodecyl sulfate. To confirm the presence of the trimer in solution, we also performed dynamic light-scattering measurements. Six series of 20 measurements yielded hydrodynamic radii of between 3.5 and 4.7 nm for the main component (comprising between 94 and 99% of the sample according to analysis of the six measuments). The remaining component(s) were much larger aggregates. These hydrodynamic radii correspond to molecular weights between 58 and 125 kDa (average 89 kDa), assuming that the protein is globular. Considering that the size of the monomer is 25.3 kDa and the fact that dynamic light scattering tends to overestimate the size of elongated species, the most reasonable explanation for the observed results is that it forms elongated trimers, which would be consistent with the expected shape observed by Steven *et al.* (1988), *i.e.* three collinear globules (see the boxed region in Fig. 1*c*).

HTX experiments provided the lead condition for obtaining the crystals belonging to space group $P2_12_12_1$: 0.2 *M* trimethylamine-*N*-oxide, 0.1 *M* Tris–HCl pH 8.5, 20%(*w*/*v*) polyethylene glycol monomethyl ether 2000 (Fig. 2*b*). Although they grow in clusters, are not nice-looking and do not show straight edges, single crystals could be separated from the clusters which were found to diffract X-rays well (Fig. 3*a*). HTX experiments also yielded one well diffracting crystal belonging to space group $C222_1$ (Fig. 2*c*); it was grown in 0.2 *M* ammonium citrate pH 7.0, 20%(*w*/*v*) polyethylene glycol 3350 and diffracted X-rays to 2.0 Å resolution (Fig. 3*b*). For crystallographic phase determination, crystals of the $P2_12_12_1$ type were soaked with heavy-atom compounds.

Complete native data sets of good quality were collected from crystals of both crystal forms ($P2_12_12_1$ and $C222_1$), integrated and scaled to 1.9 and 2.0 Å resolution, respectively. Data-processing statistics are summarized in Table 1. Both crystal forms are predicted to contain one trimer in the asymmetric unit, with corresponding solvent contents of 44 and 60%, respectively. A self-rotation function calculated using the $C222_1$ data clearly shows the presence of local threefold symmetry (Fig. 3*d*) consistent with the trimeric nature of the protein; the $\chi = 120^\circ$ section of the self-rotation function of the $P2_12_12_1$ data also shows peaks (Fig. 3*c*), although it is more difficult to interpret.

A complete highly redundant four-wavelength data set to 2.7 Å resolution was collected from a crystal of the $P2_12_12_1$ form soaked for several hours in a solution containing mother liquor that was about 20% saturated in methylmercury(II) chloride (Table 1). Upon soaking, the unit-cell *c* axis showed an increase of more than 3%. As a result, the data were found to not be isomorphous to the native data, but showed significant dispersive and anomalous differences to beyond 3 Å and nine heavy-atom sites could be identified. Phasing, structure solution and refinement will be published elsewhere.

As there are no structures in the PDB of proteins with homologous sequences, we expect the structure to reveal one or more new folds. We also hope to obtain other information from the structure, such as determining which regions may be involved in receptor binding. Together with known host-range change mutants, receptor-binding and infection-blocking experiments, this should lead to a more detailed understanding of bacteriophage T7 absorption to *E. coli*. This information will in turn have important consequences for applications of bacteriophages such as phage typing, phage display and phage therapy (Marks & Sharp, 2000).

We thank Ana Cuervo, Jaime Martin-Benito and José L. Carrascosa for providing phage T7 DNA, Patricia Ferraces-Casais for technical assistance, Sergio G. Bartual for help with cloning, Antonio L. Llamas-Saiz, José M. Otero and José-Ignacio Baños-Sanz for diffraction testing of crystals and Silvia Russi and Hassan Belrhali of the European Molecular Biology Laboratory for help with data collection at BM14. We acknowledge the use of the HTX Laboratory at the EMBL Grenoble Outstation (France), which receives funding from the European Community's Seventh Framework Programme under contract No. 227764, and thank Martin Roewer and José A. Márquez for performing the HTX experiments. We thank Armando Albert for the use of the dynamic light-scattering equipment and his help therewith. This research was sponsored by research grant BFU2008-01588 from the Spanish Ministry of Science and Innovation and by a predoctoral FPU fellowship to CG-D from the Spanish Ministry of Education.

References

Ackermann, H. W. (2001). *Arch. Virol.* **146**, 843–857.
Bartual, S. G., Garcia-Doval, C., Alonso, J., Schoehn, G. & van Raaij, M. J. (2010). *Protein Expr. Purif.* **70**, 116–121.
Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* D**67**, 271–281.
Dimasi, N., Flot, D., Dupeux, F. & Márquez, J. A. (2007). *Acta Cryst.* F**63**, 204–208.
Dunn, J. J. & Studier, F. W. (1983). *J. Mol. Biol.* **166**, 477–535.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
Garcia, E., Elliott, J. M., Ramanculov, E., Chain, P. S., Chu, M. C. & Molineux, I. J. (2003). *J. Bacteriol.* **185**, 5248–5262.
Guardado-Calvo, P., Llamas-Saiz, A. L., Fox, G. C., Glasgow, J. N. & van Raaij, M. J. (2009). *Acta Cryst.* F**65**, 1149–1152.
Guardado Calvo, P., Llamas-Saiz, A. L., Langlois, P. & van Raaij, M. J. (2006). *Acta Cryst.* F**62**, 449–452.
Heineman, R. H., Springman, R. & Bull, J. J. (2008). *Am. Nat.* **171**, E149–E157.
Maniloff, J. & Ackermann, H. W. (1998). *Arch. Virol.* **143**, 2051–2063.

Marks, T. & Sharp, R. (2000). *J. Chem. Technol. Biotechnol.* **75**, 6–17.
Mitraki, A., Barge, A., Chroboczek, J., Andrieu, J.-P., Gagnon, J. & Ruigrok, R. W. H. (1999). *Eur. J. Biochem.* **264**, 599–606.
Mitraki, A., Papanikolopoulou, K. & van Raaij, M. J. (2006). *Adv. Protein Chem.* **73**, 97–124.
Picken, R. N. & Beacham, I. R. (1977). *J. Gen. Microbiol.* **102**, 305–318.
Raaij, M. J. van, Hermo Parrado, X. L., Guardado Calvo, P., Fox, G. C., Llamas-Saiz, A. L., Costas, C., Martínez-Costas, J. & Benavente, J. (2005). *Acta Cryst.* F**61**, 651–654.
Steven, A. C., Trus, B. L., Maizel, J. V., Unser, M., Parry, D. A., Wall, J. S., Hainfeld, J. F. & Studier, F. W. (1988). *J. Mol. Biol.* **200**, 351–365.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wommack, K. E. & Colwell, R. R. (2000). *Microbiol. Mol. Biol. Rev.* **64**, 69–114.