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

Crystallization and preliminary X-ray analysis of phage Mu activator protein C in a complex with promoter DNA

Karthik K. Shanmuganatham,[a] Manimekalai Ravichandran,[b] Martha M. Howe[a] and Hee-Won Park[b,c]*

[a]Department of Molecular Sciences, University of Tennessee Health Science Center, Memphis, TN 38163, USA, [b]Structural Genomics Consortium, Department of Pharmacology, University of Toronto, Toronto, ON, M5G 1L5, Canada, and [c]Department of Structural Biology, St Jude Children's Research Hospital, Memphis, TN 38105, USA

Correspondence e-mail: heewon.park@utoronto.ca

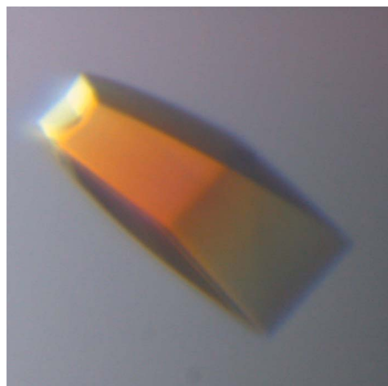
Received 6 March 2007
Accepted 23 May 2007

Bacteriophage Mu C protein is an activator of the four Mu late promoters that drive the expression of genes encoding DNA-modification as well as phage head and tail morphogenesis proteins. This report describes the purification and cocrystallization of wild-type and selenomethionine-substituted C protein with a synthetic late promoter P_{sym}, together with preliminary X-ray diffraction data analysis using SAD phasing. The selenomethionine peak data set was collected from a single crystal which diffracted to 3.1 Å resolution and belonged to space group $P4_1$ or $P4_3$, with unit-cell parameters a = 68.9, c = 187.6 Å and two complexes per asymmetric unit. The structure will reveal the amino acid–DNA interactions and any conformational changes associated with DNA binding.

1. Introduction

Mu is a temperate phage of *Escherichia coli* K-12 and other enteric bacteria (Symonds *et al.*, 1987; Paolozzi & Ghelardini, 2006). Following infection, Mu can form a lysogen or enter a lytic mode of replication for the production of progeny phage particles. The lytic cycle is tightly regulated by a transcriptional cascade, which is divided into early, middle and late phases (Stoddard & Howe, 1989; Marrs & Howe, 1990). The middle operon regulator, Mor protein, is produced from the last gene in the early transcript (Mathee & Howe, 1990). The Mor protein activates the middle promoter P_m and C protein is produced from the last gene in the middle transcript. The C protein activates transcription from the four late promoters P_{lys}, P_I, P_P and P_{mom} (Margolin *et al.*, 1989). Mor and C share a high degree of amino-acid sequence similarity to each other, but not with other known transcription activators, and thus identify a new family of transcription factors (Mathee & Howe, 1990). Both proteins form dimers in solution (Artsimovitch & Howe, 1996; Ramesh & Nagaraja, 1996) and bind an imperfect dyad-symmetry element just upstream and overlapping the −35 region of the promoter (Chiang & Howe, 1993; Artsimovitch & Howe, 1996; Ramesh & Nagaraja, 1996; Sun *et al.*, 1997).

The helix–turn–helix motif (HTH) is a very common DNA-binding motif (Brennan & Matthews, 1989; Wintjens & Rooman, 1996; Luscombe *et al.*, 2000; Aravind *et al.*, 2005). It is a key component of many prokaryotic repressors and activators as well as eukaryotic transcription factors and thereby plays an important role in gene regulation. The simplest form of the HTH motif is a right-handed three-helix bundle in which the first helix serves as a scaffolding helix for positioning the second and third helices (Wintjens & Rooman, 1996; Aravind *et al.*, 2005), which contain the pattern of conserved amino acids typical of this motif (Brennan & Matthews, 1989). Generally, the third helix is the recognition helix, which is responsible for sequence-specific DNA binding by making multiple amino acid–base interactions in the DNA major groove. Based on differences in their structures, the HTH motifs can be classified into distinctive families of closely related structures (Wintjens & Rooman, 1996; Aravind *et al.*, 2005). These families have characteristic features in addition to the three-helix bundle, such as one or more helices just before or after the three-helix bundle (tetrahelical or multihelical), a β-strand hairpin extension (winged HTH) or an unusually long turn



© 2007 International Union of Crystallography
All rights reserved

between the helices (helix–loop–helix; Wintjens & Rooman, 1996; Aravind *et al.*, 2005).

Analysis of the amino-acid sequences of Mor and C using the algorithm of Dodd & Egan (1987) predicts that each contains a HTH DNA-binding motif near its C-terminus (Bolker *et al.*, 1989; Mathee & Howe, 1990). The crystal structure of Mor reveals that each monomer has an N-terminal dimerization domain and a C-terminal HTH DNA-binding domain (Kumaraswami *et al.*, 2004). The two N-terminal helices of two monomers intertwine with each other to form a single central dimerization domain. The two flanking HTH domains are proposed to bind to two adjacent major grooves. Since the predicted DNA-binding residues of Mor are too far apart to fit into two adjacent major grooves, DNA binding may be associated with conformational changes in both the Mor dimer and the DNA (Kumaraswami *et al.*, 2004).

C protein is a close homologue of Mor. In a *BLASTP* alignment (Altschul *et al.*, 1997) their sequence identity is 38% and the similarity increases to 55% when chemically similar amino acids are included. They are also proposed to have similar dimerization and HTH domains (Kumaraswami *et al.*, 2004). Since no protein–DNA complex structure has been determined for either protein, we have recently undertaken structural analysis of Mu C protein bound to a synthetic late promoter P_{sym} (Jiang, 1999) in which the binding sites for the HTH domains are symmetrical. The binary complex structure will provide a direct test for the predicted HTH–DNA interactions and the associated protein and DNA conformational changes that have been proposed for DNA-bound Mor and C proteins (Kumaraswami *et al.*, 2004). This report describes the purification, crystallization and preliminary crystallographic analysis of Mu C–DNA complexes formed separately with wild-type and selenomethionine-substituted C protein.

2. Experimental procedures and results

2.1. Expression and purification of Mu C protein

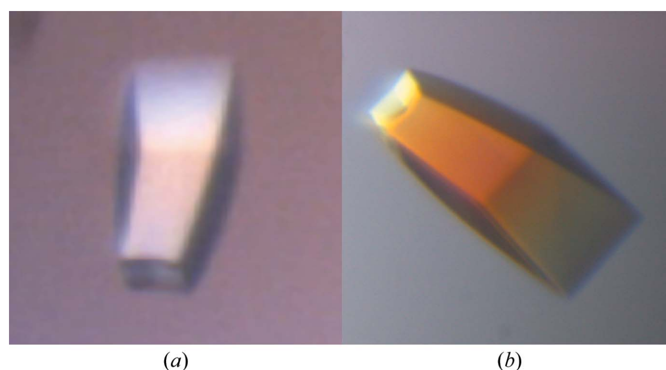
The *C* gene was cloned into an expression vector derived from pACYC184 to create pZZ41, in which the expression of wild-type C protein is under the control of a T7 promoter (Zhao, 1999). Plasmid pZZ41 was transformed into *E. coli* JM109(DE3) (Promega) and a single colony was picked, inoculated into LB containing 34 µg ml^{-1} chloramphenicol and grown at 310 K. When the cells reached an OD_{600} of 0.4, expression of the T7 RNA polymerase was induced by addition of IPTG to a final concentration of 1 m*M*, resulting in high levels of expression of wild-type C protein. After growth of the culture for 3 h at 310 K, the cells were collected by centrifugation at 6000*g* for 10 min at 277 K. The pellet was resuspended in C buffer (25 m*M* HEPES pH 7.0, 50 m*M* NaCl, 1 m*M* EDTA, 5% glycerol, 1 m*M* DTT, 1 m*M* $MgCl_2$) and the cells were lysed using a Microfluidics HC-8000 microfluidizer. The lysed cell suspension was subjected to centrifugation at 20 000*g* for 30 min to remove cell debris.

The overexpressed wild-type C protein was purified by a four-step chromatography procedure using an ÄKTA FPLC from GE Healthcare, with each purification step being monitored by SDS–PAGE. The supernatant from the cell lysate was loaded onto a Hi-trap heparin affinity column (GE Healthcare) pre-equilibrated with C buffer and the bound proteins were eluted with a 0–500 m*M* NaCl gradient, with C protein eluting between 200 and 300 m*M* NaCl. The peak fractions containing C protein were pooled, diluted to ∼100 m*M* NaCl with NaCl-free C buffer and loaded onto an SP-Sepharose cation-exchange column; bound proteins were eluted with a 100–500 m*M* NaCl gradient, with C protein eluting between 225 and 325 m*M* NaCl. Fractions containing C protein were pooled, adjusted to 1.5 *M* NaCl by adding 5 *M* NaCl and then loaded onto a phenyl-Sepharose hydrophobic column (GE Healthcare); bound proteins were eluted with a 1.5–0 *M* NaCl gradient, with C protein eluting in the 800–100 m*M* range. Finally, pooled C-protein fractions were concentrated to 10–15 ml (approximately eightfold) using YM30 Amicon concentrators (Amicon Bioseparation, Bedford, Massachusetts, USA) and loaded onto a Superdex 75 26/60 size-exclusion column (GE Healthcare) pre-equilibrated with buffer C-DHS (C buffer containing 10 m*M* DTT and 150 m*M* NaCl). The peak C-protein fractions were pooled and concentrated to 30 mg ml^{-1} using Amicon YM30 concentrators and stored in buffer C-DHS at 203 K. Protein purity was estimated to be ∼95% by SDS–PAGE.

The metabolic inhibition method described by Van Duyne *et al.* (1993) was used to incorporate selenomethionine (SeMet) into C protein during expression. A single colony was inoculated into 150 ml LB medium containing 34 µg ml^{-1} chloramphenicol and grown overnight at 310 K. The cells were collected the following day and resuspended in 2 l M9 minimal medium (Symonds *et al.*, 1987) with 34 µg ml^{-1} chloramphenicol and grown for another 12 h or overnight; this culture was used to seed 8 l of the same minimal medium. When the cells reached an OD_{600} of 0.6, the following amino acids (all from Sigma) were added: 800 mg each of L-lysine, L-phenylalanine and L-threonine and 400 mg each of L-isoleucine, L-leucine, L-valine and L-selenomethionine. After the culture had been shaken for 15 min, protein expression was induced by adding IPTG to a final concentration of 1 m*M* and the culture was grown for 12 h at 310 K. Cell lysis and purification of the SeMet C protein were performed as described above for wild-type C protein. Electrospray mass spectrometry (ESI–TOF) was used to confirm the incorporation of two SeMet residues in place of the two naturally occurring methionine residues in C protein, one of which is at the N-terminus and the other of which is in the HTH domain. No other modifications were detected. The wild-type and SeMet C-protein preparations were treated and used under the same conditions unless indicated otherwise.

2.2. Cocrystallization

The oligonucleotides used for crystallization were obtained as separate top and bottom strands from IDT (Integrated DNA Technologies, Coralville, IA, USA) and dissolved in C buffer containing 1 m*M* DTT and 50 m*M* NaCl. Equimolar amounts of the top and bottom strand were mixed and annealed in a thermal cycler by first incubating the oligonucleotides at 5 K above the predicted T_m for 5 min, then at 5 K below the predicted T_m for 5 min and then rapidly



(*a*) (*b*)

Figure 1
Crystals of the complexes produced using (*a*) wild-type C and (*b*) SeMet C. Typical dimensions of both crystals are 0.1 × 0.1 × 0.5 mm.

Table 1
Diffraction data statistics of SeMet C–DNA complex crystals.

Values in parentheses are for the highest resolution shell (3.21–3.10 Å).

X-ray source	APS (beamline 17ID)
Wavelength (Å)	0.979
Detector	CCD
Crystal parameters	
Space group	$P4_1$ or $P4_3$
Unit-cell parameters	
a (Å)	68.9
c (Å)	187.6
Unit-cell volume (Å^3)	890576.6
Data statistics	
Resolution (Å)	3.1
No. of observed reflections	118381
No. of unique reflections	15810
Completeness (%)	99.8 (100.0)
$R_{\rm sym}$†	0.072 (0.282)
Average $I/\sigma(I)$	36.6 (8.0)
Redundancy	7.5
$V_{\rm M}$	2.8
Solvent content	0.56
No. of molecules per ASU	4 protein molecules and 1 dsDNA

† $R_{\rm sym} = \sum_{hkl}[\sum_i |I_{hkl,i} - \langle I_{hkl}\rangle|]/\sum_{hkl,i}\langle I_{hkl}\rangle$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with Miller indices hkl and I_{hkl} is the mean intensity of that reflection.

cooling to room temperature. Protein–DNA complexes were formed by mixing equal volumes of C protein (in buffer C-DHS) and annealed DNA in a 1:1 molar ratio of C dimer to DNA. The complexes were purified using a Superdex 75 26/60 size-exclusion column (GE Healthcare) pre-equilibrated in buffer C-DS (C buffer containing 10 m*M* DTT and 75 m*M* NaCl). The purified complexes were then concentrated to 25–30 mg ml^{-1} using YM30 Amicon concentrators and stored at 203 K.

Crystallization trials with both the wild-type and SeMet complexes were performed by the hanging-drop and sitting-drop vapor-diffusion methods using commercial screens from Hampton Research (Aliso Viejo, CA, USA) and Jena Bioscience (San Diego, CA, USA) as well as custom screens developed at the Structural Genomics Consortium (Toronto, Canada). Crystals obtained from many different conditions were screened and refined by varying the primary precipitant, additives, pH, temperature and the length of the DNA. Good-quality crystals for X-ray analysis were obtained from both protein complexes made with a 22 bp DNA (5′-GTATTATGACTCCA-TAATCCGG-3′ for the top strand and 5′-GGATTATGGAGTCA-TAATACCC-3′ for the bottom strand) within a couple of days at room temperature using the sitting-drop method. Both wild-type and SeMet protein–DNA complexes refined to the same conditions and similar crystals (shown in Fig. 1) were obtained for both by mixing 0.5 µl purified complex solution with 0.5 µl reservoir solution containing 1.4 *M* ammonium sulfate, 50 m*M* trisodium citrate, 21%(*v*/*v*) ethylene glycol and 10 m*M* DTT at pH 5.7.

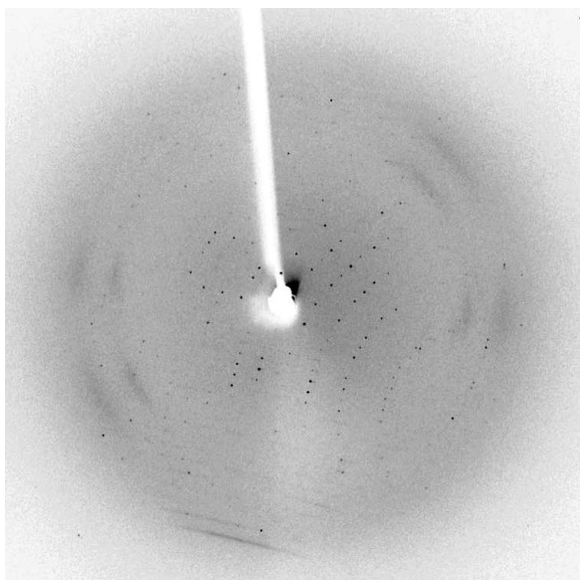


Figure 2
X-ray diffraction pattern for the SeMet C–DNA complex.

2.3. Preliminary X-ray analysis

A single crystal containing the wild-type or SeMet complex was harvested from the crystallization drop using a cryoloop (Hampton Research), soaked in its reservoir solution supplemented with additional ethylene glycol to a final concentration of 33%(*v*/*v*), flash-frozen and stored in liquid N_2. For data collection, the flash-frozen crystal was placed directly into a cold nitrogen-gas stream and X-ray diffraction images were collected (Fig. 2). Since the SeMet crystal diffracted to higher resolution, the SeMet peak data set was used for analysis; it was collected using a CCD image-plate detector with synchrotron radiation of wavelength 0.9791 Å at beamline 17ID of the Advanced Photon Source (Argonne National Laboratory, Chicago, USA). The distance between the crystal and detector was 300 mm and a total of 360 oscillation images were recorded with exposure times of 10 s. The diffraction data were indexed, processed and scaled using *DENZO* and *SCALEPACK* from the *HKL* package (Otwinowski & Minor, 1997). Data statistics are given in Table 1.

The crystals belong to space group $P4_1$ or $P4_3$, with unit-cell parameters a = 68.9, c = 187.6 Å, and contain two molecules of the C protein–DNA complex in the asymmetric unit. A total of 118 933 independent reflections were obtained with a redundancy of 7.5 at 3.1 Å. Attempts were made to solve the structure of the protein C–DNA complex using Mor as the model for molecular replacement, since Mor is the only protein of known structure related to C. These attempts were unsuccessful, possibly because C protein has an altered conformation when bound to DNA and thus differs from the search model. We have obtained the phase angles using the SeMet data set; structure determination is in progress.

This work was supported by the American Lebanese–Syrian Associated Charities at St Jude Children's Research Hospital to H-WP, grant MCB-0418108 from the National Science Foundation (USA) to MMH and H-WP and a University of Tennessee Van Vleet Professorship to MMH. The authors appreciate the assistance provided by the IMCA-CAT beamline staff of APS during data collection. KKS is grateful to St Jude Children's Research Hospital and the University of Toronto Structural Genomics Consortium for their hospitality during this project.

References

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. (2005). *FEMS Microbiol. Rev.* **29**, 231–262.
Artsimovitch, I. & Howe, M. M. (1996). *Nucleic Acids Res.* **24**, 450–457.
Bolker, M., Wulczyn, F. G. & Kahmann, R. (1989). *J. Bacteriol.* **171**, 2019–2027.
Brennan, R. G. & Matthews, B. W. (1989). *J. Biol. Chem.* **264**, 1903–1906.
Chiang, L.W. & Howe, M. M. (1993). *Genetics*, **135**, 619–629.
Dodd, I. B. & Egan, J. B. (1987). *J. Mol. Biol.* **194**, 557–564.
Jiang, Y. (1999). PhD dissertation. University of Tennessee, Memphis, TN, USA.

Kumaraswami, M., Howe, M. M. & Park, H.-W. (2004). *J. Biol. Chem.* **279**, 16581–16590.
Luscombe, N. M., Austin, S. E., Berman, H. M. & Thornton, J. M. (2000). *Genome Biol.* **1**, 1–18.
Margolin, W., Rao, G. & Howe, M. M. (1989). *J. Bacteriol.* **171**, 2003–2018.
Marrs, C. F. & Howe, M. M. (1990). *Virology*, **174**, 192–203.
Mathee, K. & Howe, M. M. (1990). *J. Bacteriol.* **172**, 6641–6650.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Paolozzi, L. & Ghelardini, P. (2006). *The Bacteriophages*, edited by R. Calendar, pp. 469–496. Oxford University Press.
Ramesh, V. & Nagaraja, V. (1996). *J. Mol. Biol.* **260**, 22–33.
Stoddard, S. F. & Howe, M. M. (1989). *J. Bacteriol.* **171**, 3440–3448.
Sun, W., Hattman, S. & Kool, E. (1997). *J. Mol. Biol.* **273**, 765–774.
Symonds, N., Toussaint, A., van de Putte, P. & Howe, M. M. (1987). *Phage Mu.* New York: Cold Spring Harbor Laboratory Press.
Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). *J. Mol. Biol.* **229**, 105–124.
Wintjens, R. & Rooman, M. (1996). *J. Mol. Biol.* **262**, 294–313.
Zhao, Z. (1999). PhD dissertation. University of Tennessee, Memphis, TN, USA.