Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray analysis of the controller protein C.AhdI from *Aeromonas hydrophilia*

J. E. McGeehan,[a] S. Streeter,[a] J. B. Cooper,[b] F. Mohammed,[b] G. C. Fox[c] and G. G. Kneale[a]*

[a]Biophysics Laboratories, Institute of Biomedical and Biomolecular Sciences, University of Portsmouth, Portsmouth PO1 2DT, England, [b]School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, England, and [c]BM14 CRG, European Synchrotron Radiation Facility, 6 Rue Jules Horowitz BP 220, 38043 Grenoble CEDEX 9, France

Correspondence e-mail: geoff.kneale@port.ac.uk

Received 2 October 2003
Accepted 12 November 2003

Single crystals of purified homodimeric controller protein from *Aeromonas hydrophilia* (C.AhdI) have been grown under several different conditions using vapour diffusion. X-ray diffraction data have been collected using synchrotron radiation from crystals of both the native and a selenomethionine (SeMet) derivative of the protein. The native crystal form belongs to space group $P2_1$ and data were collected to a resolution of 2.2 Å. Two crystal forms of the SeMet protein have been obtained and were found to belong to space groups $P1$ and $P2_1$; data have been recorded to 2.0 and 1.7 Å resolution, respectively, for the two crystal forms. Three-wavelength MAD data were collected to 1.7 Å for the SeMet derivative crystal, which is isomorphous with the native $P2_1$ crystal.

1. Introduction

Bacterial restriction-modification (RM) systems are gene complexes that allow bacteria to methylate and protect their own DNA whilst cleaving and eliminating unmethylated, *i.e.* foreign, DNA (Wilson & Murray, 1991; Roberts *et al.*, 2003).

C.AhdI is a member of a group of small bacterial proteins (the controller or C proteins) found in some RM systems and believed to temporally regulate the expression of restriction endonucleases (REase) within a bacterial host, allowing methylation of host DNA to proceed prior to REase activity against unmethylated non-self DNA (Tao *et al.*, 1991; Tao & Blumenthal, 1992; Ives *et al.*, 1992; Vijesurier *et al.*, 2000).

Despite some conjecture over their structure (Lunnen *et al.*, 1997), no structural data have been reported on any controller proteins to date. C.AhdI has been shown to form homodimers by analytical ultracentifugation and binds to a region of DNA upstream of the operon containing the *C* and *R* genes of AhdI (Streeter, McGeehan & Kneale, manuscript in preparation). Structural studies are being undertaken and here we report the crystallization and preliminary X-ray diffraction analysis of the protein. The gene encoding C.AhdI, *ahdIC*, was cloned and expressed in *Escherichia coli*. The recombinant product was purified using heparin, cation-exchange and size-exclusion chromatography and concentrated by a step gradient using cation-exchange chromatography. Controller proteins are believed to be active as homodimers (Vijesurier *et al.*, 2000) and in C.AhdI each monomeric unit consists of 74 amino-acid residues (8.37 kDa). Crystals of both the native and the SeMet derivative were obtained that were suitable for X-ray diffraction analysis.

2. Experimental

2.1. Protein expression and purification

The recombinant *A. hydrophilia* controller protein was overexpressed in *E. coli* BL21 (DE3). Production of the selenomethionine (SeMet) derivative of the C protein was undertaken in methionine-auxotrophic *E. coli* B834 (DE3) grown in New Minimal Medium (Budisa *et al.*, 1995). Bacterial cells expressing C.AhdI were lysed by sonication and clarified supernatants were produced by centrifugation. Purification of each form of the protein was performed using heparin, cation-exchange and size-exclusion chromatography. Concentration of the protein involved NaCl step-gradient chromatography on the cation-exchange column followed by dialysis to reduce the NaCl concentration to 100 m*M*. The final protein concentration was approximately 0.5 mg ml^{-1}, which was limited by low solubility. Aggregation was observed at higher protein concentrations.

Purity was checked by tris-tricine PAGE. Additionally, the absence of tryptophan residues in the amino-acid sequence of the protein enables the purity to be determined by fluorescence spectroscopy. The native protein, which had a calculated molecular weight of 8369.6 Da from the amino-acid sequence, corresponds closely to the molecular weight of 8369.5 Da observed by mass spectrometry. SeMet incorporation was evaluated for the derivative protein by mass spectrometry and an increase of 46.9 Da from the native protein to 8416.4 Da was observed, corresponding to

 DOI: 10.1107/S0907444903026143

the single SeMet present in the protein chain. The final protein concentration was estimated from the absorbance at 276 nm using a calculated extinction coefficient of 2900 M^{-1} cm^{-1} for the monomeric form.

2.2. Crystallization

Initial screening for nucleation and crystallization conditions was performed using hanging-drop vapour diffusion (Davies & Segal, 1971) at 295 K. Suitable conditions were identified using the commercially available sparse-matrix screening kits Structure Screen 1 and Structure Screen 2 from Molecular Dimensions Ltd.

Conditions contained within the 'footprint' screen (Stura *et al.*, 1992) were then tested and optimized. In all cases, the drops were prepared on siliconized cover slips and contained 2 µl protein and 2 µl reservoir solution, giving a maximum final concentration of 0.25 mg ml^{-1}. The droplets were equilibrated against a reservoir volume of 0.5 ml in 24-well tissue-culture plates.

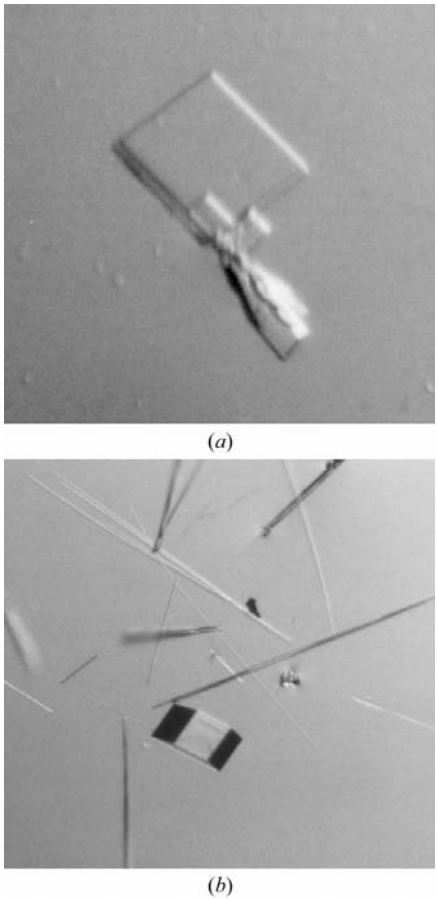

Figure 1
(*a*) Crystal of the native protein, showing the typical plate form. The longest crystal dimension is 0.2 mm. (*b*) Crystals of the SeMet protein at the transition point between the thin needle form and the rectangular form using 20%(*w*/*v*) MPD as an additive. The longest dimension of the rectangular crystal is 0.2 mm.

Table 1
Crystallographic parameters and data-collection statistics.

Values in parentheses refer to the highest resolution shells (2.32–2.20 Å for Native I data, 2.11–2.00 Å for SeMet I data and 1.78–1.69 Å for SeMet II data). Only the peak-wavelength statistics are given for the SeMet II MAD data sets.

	Native I	SeMet I	SeMet II
Beamline	ESRF ID29	ESRF ID29	ESRF BM14
Image-plate system	165 mm MAR CCD	165 mm MAR CCD	133 mm MAR CCD
Crystal-to-detector distance (mm)	200	150	120
Oscillation per image (°)	1.0	1.0	0.7
No. of images	185	360	Inflection, 250; peak 520 + 100; high remote, 250
Temperature (K)	100	100	100
Wavelength (Å)	0.97917	0.93300	See Table 2
Space group	$P2_1$	$P1$	$P2_1$
Unit-cell parameters (Å, °)	a = 24.7, b = 58.0, c = 46.9, α = 90, β = 99.1, γ = 90	a = 24.5, b = 33.2, c = 46.3, α = 90.6, β = 98.5, γ = 94.5	a = 24.4, b = 57.2, c = 46.1, α = 90, β = 98.9, γ = 90
Volume (Å^3)	66343.3	37122.6	63566.2
Monomers per asymmetric unit	2	2	2
Packing density (V_M) (Å^3 Da^{-1})	1.98	2.21	1.89
Solvent content (%)	37.9	44.2	34.9
Resolution range (Å)	20–2.20	20–2.0	20–1.69
Total observations	24004	34332	128403
Unique reflections	6581	9392	14018
Completeness (%)	97.7 (88.4)	96.4 (95.6)	99.0 (93.2)
Multiplicity	3.6 (3.2)	3.7 (3.7)	9.2 (6.8)
$I/\sigma(I)$	3.4 (1.8)	1.7 (1.5)	7.0 (10.2)
$R_{\rm merge}$† (%)	12.0 (35.4)	21.6 (34.8)	4.1 (10.2)

† $R_{\rm merge} = \sum_h \sum_i |(I_{hi} - I_h)| / \sum_h \sum_i (I_{hi})$, where I_h is the mean intensity of the scaled observations I_{hi}.

2.3. Data collection and processing

Prior to freezing, crystals were cryoprotected by the addition of approximately 10–25%(*v*/*v*) glycerol depending upon the amount of PEG 6000 or 2-methyl-2,4-pentanediol (MPD) present in the reservoirs of each crystal well. The crystals were then mounted in loops, frozen in liquid ethane and stored in liquid nitrogen.

X-ray diffraction data for C.AhdI crystals were measured using synchrotron radiation at 100 K. Native data were collected at station ID29 with a 165 mm MAR CCD and MAD data were collected at station BM14 with a 133 mm MAR CCD, both at the European Synchrotron Research Facility (ESRF), Grenoble, France. Table 1 details the data-collection parameters used for each crystal form. The space groups and unit-cell parameters were determined using the autoindexing and parameter-refinement procedures of *MOSFLM* (Leslie, 1994). Intensity data were sorted and scaled using the *CCP*4 programs *SORTMTZ* and *SCALA* (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

Small crystals were obtained in several buffers contained in the sparse-matrix Structure Screens 1 and 2, with high-molecular-weight PEG the most common component. These preliminary crystals were thin needle clusters and not suitable for X-ray diffraction. Expanding on these results, the 'footprint' screen was tested and yielded larger and more reproducible needle shaped crystals utilizing PEG 6000 as the main precipitant. The pH range was crucial and optimized between 6.0 and 7.0 using 0.2 *M* imidazole–malate. Trials at various temperatures and seeding experiments failed to produce larger crystals.

A series of 20 additives was tested in an attempt to enhance the quality of the crystals. Spermidine and MPD were the most successful additives when used with 25%(*w*/*v*) PEG 6000 at pH 7.0. Titration of spermidine from 10–100 m*M* produced increasingly thicker needles but owing to rapid precipitation at higher concentrations the size could not be optimized further. Titration of MPD, however, produced a transition in the crystal shape from needles to plates between 10 and 15%(*w*/*v*), and then to rectangular crystals at concentrations up to 30%(*w*/*v*). These crystals (Fig. 1*a*) could be reproducibly grown in 7–14 d and, although still small for a Cu $K\alpha$ rotating-anode source, proved to diffract well at the focused synchrotron beamlines. The SeMet derivative of C.AhdI crystallized with the same basic buffers components as the wild-type protein (Fig. 1*b*), but owing to differences in solubility the concentration of PEG 6000 was reduced to 10%(*w*/*v*) for optimal growth.

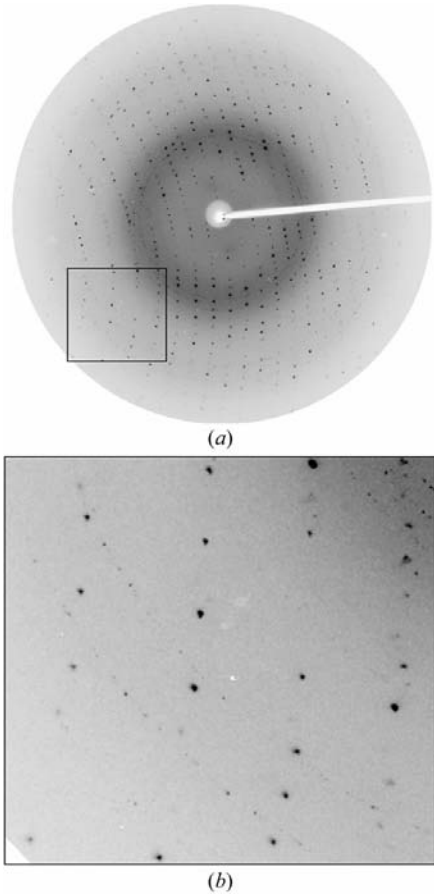

Figure 2
(*a*) X-ray diffraction pattern of a SeMet crystal of C.AhdI. The exposure time was 10 s, the MAR CCD detector distance was 120 mm and the oscillation range per frame was 0.7°. (*b*) The diffraction limit at the edge of the image was 1.7 Å.

Single-wavelength data (Fig. 2) were collected from both a native C.AhdI crystal and a SeMet-derivative crystal, which belong to space groups $P2_1$ and $P1$, respectively, as determined by their systematic absences. Following optimization of the SeMet crystal-growth conditions, three-wavelength MAD data were collected for a SeMet crystal which was isomorphous with the $P2_1$ native crystal. A fluorescence scan of the crystal was performed at the Se *K* edge and the real (f') and imaginary (f'') components of the anomalous scattering correction were calculated using the program *CHOOCH* (Evans & Pettifer, 2001). Table 2 shows the wavelengths used for the MAD data collection with their associated scattering factors. The predicted signal at the peak wavelength for one Se in 74 amino acids is 4.8% for Bijvoet differences and 3.8% for dispersive differences using the calculation of Hendrickson (1990).

Table 2
Anomalous scattering factors for the SeMet derivative crystal at the three wavelengths used in the MAD experiment obtained from *CHOOCH* (Evans & Pettifer, 2001).

Data set	Wavelength (Å)	f'	f''
Inflection	0.979652	−9.8	2.8
Peak	0.979266	−7.6	5.5
High-energy remote	0.904989	−1	3

Table 2 summarizes the crystallographic properties and processing statistics for all three crystals.

The presence of two monomers in the crystallographic asymmetric unit gives a Matthews coefficient for the native crystal of 1.98 Å^3 Da^{-1} and a solvent content of 37.9% (Matthews, 1968). These values are close to that of the $P2_1$ SeMet crystal and the small decrease in packing density is consistent with the additional 50 Da attributed to a single selenium per monomer. Similarly, the $P1$ crystal is calculated to possess two monomers in the asymmetric unit, which is consistent with the biological dimer arrangement of C.AhdI.

The $R_{\rm merge}$ of the native data set was 12.0% compared with 21.6% for the initial SeMet data. Optimization of the SeMet crystal-growth conditions produced larger better quality crystals. A significant improvement in the $R_{\rm merge}$ to 4.1% was achieved at the peak wavelength with the optimized SeMet crystal. Efforts were made to increase the multiplicity of the data and 360° were collected plus a further 180° after a 16° rotation about the κ angle on the goniometer. The latter 180° of data increased the multiplicity from 7.2 to 9.2 overall. Following the peak data, the inflection and remote-wavelength data were collected and were found to have an $R_{\rm merge}$ of 3.4 and 3.3%, respectively. The unit-cell volume increased by only 0.024% and $I/\sigma(I)$ remained stable at 8.8 for the remote data set. This indicates no serious radiation damage over the course of the experiment and work is proceeding to generate phases from this data.

Molecular replacement was attempted using the programs *MOLREP* (Vagin & Teplyakov, 1997), *CNS* (Brünger *et al.*, 1998) and *BEAST* (Read, 2001) using three search models: *Bacillus subtilis* SinR, bacteriophage λ repressor and bacteriophage 434 Cro protein, which have primary sequence identities of 31, 23 and 21%, respectively. No significant solutions were found with the λ repressor and although weak rotation-function solutions were found for the SinR and Cro proteins, the translation function did not yield any significant peaks. This suggests that C.AhdI may have a distinct domain structure and supports efforts to solve the structure by MAD methods.

Determination of the structure of C.AhdI is under way. Co-crystallization experiments with DNA containing the putative recognition site have also begun.

We would like to acknowledge the ESRF, Grenoble, France for provision of synchrotron time and travel support. We are grateful to the BBSRC for funding a PhD studentship. We thank Professor Steve Wood for constructive discussions, Dr Peter Erskine for assistance with data collection and Dr Dave Whitley for computer support.

References

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. & Huber, R. (1995). *Eur. J. Biochem.* **230**, 788–796.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. -S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Davies, D. R. & Segal, D. M. (1971). *Methods Enzymol.* **22**, 266–269.
Evans, G. & Pettifer, R. F. (2001). *J. Appl. Cryst.* **34**, 82–86.
Hendrickson, W. A. (1990). *Science*, **254**, 51–58.
Ives, C. L., Nathan, P. D. & Brooks, J. E. (1992). *J. Bacteriol.* **174**, 7194–7201.
Leslie, A. G. W. (1994). *MOSFLM User Guide.* Cambridge: MRC–LMB.
Lunnen, K., Chang, Z., Morgan, R., Nwankwo, D., Slatko, B. & Wilson, G. (1997). *Proceedings of the Fourth New England Biolabs Workshop on Biological DNA Modification*, p. 84, Igls, Tyrol, Austria.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.
Roberts, R. J. *et al.* (2003). *Nucleic Acids Res.* **31**, 1805–1812.
Stura, E. A., Nemerow, G. R. & Wilson, I. A. (1992). *J. Cryst. Growth*, **122**, 273–285.
Tao, T. & Blumenthal, R. M. (1992). *J. Bacteriol.* **174**, 3395–3398.
Tao, T., Bourne, J. C. & Blumenthal, R. M. (1991). *J. Bacteriol.* **173**, 1367–1375.
Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
Vijesurier, R. M., Carlock, L., Blumenthal, R. M. & Dunbar, J. C. (2000). *J. Bacteriol.* **182**, 477–487.
Wilson, G. G. & Murray, N. E. (1991). *Annu. Rev. Genet.* **25**, 585–627.