

**S. D. Streeter, J. E. McGeehan\*  
and G. G. Kneale\***

Biophysics Laboratories, Institute of Biomedical  
and Biomolecular Sciences, School of Biological  
Sciences, University of Portsmouth,  
Portsmouth PO1 2DY, England

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

Received 29 May 2009  
Accepted 20 July 2009

## Overexpression, purification and preliminary X-ray diffraction analysis of the controller protein *C.Csp231I* from *Citrobacter* sp. RFL231

Restriction–modification controller proteins play an essential role in regulating the temporal expression of restriction–modification genes. The controller protein *C.Csp231I* represents a new class of controller proteins. The gene was subcloned to allow overexpression in *Escherichia coli*. The protein was purified to homogeneity and crystallized using the hanging-drop vapour-diffusion method. The crystals diffracted to 2.0 Å resolution and belonged to space group *P2*<sub>1</sub>. An electrophoretic mobility-shift assay provided evidence of strong binding of *C.Csp231I* to a sequence located upstream of the *csp231IC* start codon.

### 1. Introduction

Controller (C) proteins have been identified in many restriction–modification (R–M) systems and play a vital role in the temporal regulation of R–M genes. These helix–turn–helix proteins have been shown to act as regulators of both their own transcription and that of the restriction endonuclease (ENase) encoded within the same operon. In some cases, C proteins also regulate transcription of the methyltransferase (MTase; Tao *et al.*, 1991; Ives *et al.*, 1992; Rimšeliienė *et al.*, 1995; Lubys *et al.*, 1999; Česnavičienė *et al.*, 2003; Semenova *et al.*, 2005; Bogdanova *et al.*, 2008).

Controller proteins have recently been categorized on the basis of ten distinct DNA-recognition motifs (Sorokin *et al.*, 2009). To date, the structures of three C proteins have been reported (McGeehan *et al.*, 2005, 2008; Sawaya *et al.*, 2005); all are highly homologous proteins with similar folds and with similar DNA-recognition sites. Other groups, such as that exemplified by *C.Csp231I* and *C.EcoO109I*, have very different recognition sites and their structures are currently unknown. Kita *et al.* (2002) have previously identified the recognition sequence of *C.EcoO109I* as a 15 bp sequence comprising two palindromic pentanucleotides separated by a nonbinding pentanucleotide sequence, 5'-CTAAG(N<sub>5</sub>)CTTAG-3', located 47 bp upstream of the C gene start codon. This conforms to the sequence motif identified for both *C.Csp231I* and *C.EcoO109I* by bioinformatic analysis (Sorokin *et al.*, 2009).

In the present paper, we report the expression, purification and characterization of *C.Csp231I* together with preliminary crystallization and diffraction analysis.

### 2. Materials and methods

#### 2.1. Cloning and expression

The gene *csp231IC* (Genbank ID AY787793.1) encoding the putative 98-amino-acid controller protein *C.Csp231I* was synthesized and subcloned into the expression vector pET-11a by GenScript (Piscataway, New Jersey, USA). The resultant expression vector was transformed into *Escherichia coli* BL21 (DE3) Gold cells. A single colony was added to 15 ml 2×YT medium containing 100 µg ml<sup>-1</sup> ampicillin and cultured overnight at 310 K while shaking at 225 rev min<sup>-1</sup>. A 10 ml aliquot of the starter culture was used to inoculate 1 l 2×YT medium containing 100 µg ml<sup>-1</sup> ampicillin and

the culture was incubated at 310 K (with shaking at 225 rev min<sup>-1</sup>) until the *A*<sub>600</sub> reached ~0.6, whereupon 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to induce protein expression. The culture was incubated with shaking for a further 3 h prior to cell harvesting by centrifugation.

2.2. Purification

All purifications were performed at 277 K. The harvested cells were suspended in a buffer consisting of 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM EDTA, 3 mM DTT and disrupted by sonication. Following centrifugation (39 191g, 30 min, 277 K), the supernatant was loaded onto a 5 ml HiTrap heparin HP column (GE Healthcare) and eluted with a 0.1–1 M NaCl gradient. Fractions containing the target protein were pooled and dialysed against 5 l buffer A (50 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA, 1 mM DTT). Following centrifugation (27 216g, 30 min, 277 K), the supernatant was loaded onto a 1 ml HiTrap SP HP column (GE Healthcare) and eluted with a 0.1–1 M NaCl gradient. Fractions containing the target protein were pooled and dialysed against buffer A (as above) containing 500 mM NaCl to reduce interactions between the target protein and contaminants. The dialysate was loaded onto a HiPrep 26/60 Sephacryl S-100 HR (GE Healthcare) column using buffer A with NaCl added to a final concentration of 500 mM. The purified protein was then pooled and dialysed against 5 l buffer A prior to concentration using a 1 ml HiTrap SP HP column (GE Healthcare) with a 0.1–1 M NaCl step gradient. Following a final dialysis step to reduce the NaCl concentration to 100 mM, the protein concentration was determined by UV spectroscopy using an extinction coefficient calculated from the amino-acid sequence of the monomer of *E*<sub>280</sub> = 11 460 M<sup>-1</sup> cm<sup>-1</sup>.

2.3. Dynamic light scattering

Dynamic light scattering (DLS) was performed on a *C.Csp231I* sample at 1.4 mg ml<sup>-1</sup> in 40 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA at 293 K using a Protein Solutions DynaPro temperature-controlled microscopier. The technique provides an estimate of the particle hydrodynamic radius (*R*<sub>h</sub>) and solution molecular weight, as well as the polydispersity of the sample, by analysis of the auto-correlation function. For globular proteins, the value of *R*<sub>h</sub> can be used to estimate the molecular mass *M*<sub>r</sub> using the empirical equation

$$M_r = (1.68 \times R_h)^{2.34}$$

2.4. Electrophoretic mobility-shift assay

Electrophoretic mobility-shift assays (EMSA) were performed using nondenaturing gel electrophoresis. Two complementary DNA strands corresponding to the region upstream of the *C.Csp1396I* gene were purchased (Eurogentec), one of which was labelled with the

Table 1

Crystal parameters and data-collection statistics.

Values in parentheses are for the highest resolution shell.

Crystal parameters	
Space group	<i>P</i> 2 <sub>1</sub>
Unit-cell parameters	<i>a</i> = 49.01, <i>b</i> = 29.53, <i>c</i> = 64.38, α = 90.00, β = 101.91, γ = 90.00
Solvent content (%)	38.7
Data collection	
Temperature (K)	100
Wavelength (Å)	0.9795
Resolution (Å)	50.0–2.0 (2.11–2.00)
No. of measured reflections	41601 (6256)
No. of unique reflections	12439 (1821)
Completeness (%)	99.2 (99.7)
<i>I</i> / <i>σ</i> ( <i>I</i> )	11.7 (4.1)
Multiplicity	3.3 (3.4)
<i>R</i> <sub>merge</sub> †	0.057 (0.276)

† *R*<sub>merge</sub> =  $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where *I*(*hkl*) is the mean intensity of reflection *I*(*hkl*) and *I*<sub>*i*</sub>(*hkl*) is the intensity of an individual measurement of reflection *I*(*hkl*).

fluorescent tag hexachlorofluorescein (hex), and the two strands were annealed to form a duplex. Aliquots of *C.Csp231I* were incubated with 800 nM hex-labelled 96 bp DNA duplex in binding buffer (50 mM Tris-HCl pH 8.0) at 277 K for 30 min. The samples were loaded onto a pre-run 5% native polyacrylamide gel and run at 100 V for 150 min. The gels were then scanned using an FLA-5000 imaging system (FujiFilm).

2.5. Crystallization

Crystallization conditions were screened by the hanging-drop vapour-diffusion method using the PACT screen kit (Molecular Dimensions) at 289 K. Drops were prepared by mixing 2 μl reservoir solution with 2 μl 1.2 mg ml<sup>-1</sup> protein in dialysis buffer and were equilibrated by vapour diffusion against the reservoir solution.

2.6. X-ray diffraction analysis

Crystals were cryoprotected by transfer to crystallization solution containing 30% (v/v) glycerol prior to cryocooling in liquid nitrogen. Initial indexing suggested that the space group was primitive monoclinic and therefore a 180° data set was collected from a single crystal (of approximate dimensions 100 × 80 × 15 μm) with an oscillation width of 0.5° on beamline I02 of the Diamond Light Source, UK. Data extending to 2.0 Å were collected using an ADSC Q315 CCD detector and processing was performed with *MOSFLM* (Leslie, 1992) and *SCALA* (Collaborative Computational Project, Number 4, 1994). Data-collection statistics are given in Table 1.

2.7. Sequence analysis

Amino-acid sequence alignment was carried out using the *T-Coffee* web server (Armougom *et al.*, 2006) and visualized with the *ESPrpt*

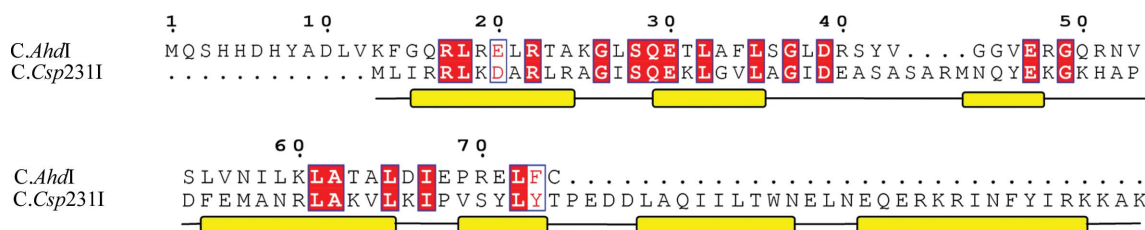
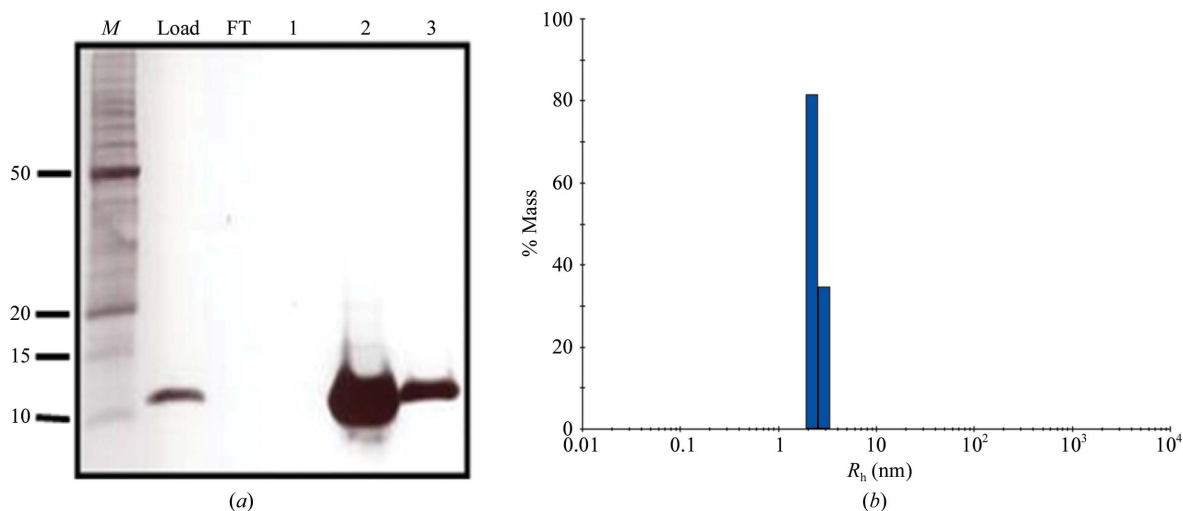


Figure 1

Amino-acid alignment of *C.AhdI* and *C.Csp231I*. Identical amino acids are highlighted in red boxes and similar amino acids in white boxes. Regions predicted to be α-helical are shown as yellow bars.



**Figure 2** 12% Tris-tricine polyacrylamide gel of purified *C.Csp231I* and dynamic light scattering of sample. (a) Lane *M* contains Benchmark protein ladder with selected molecular masses highlighted (kDa). Lane *FT* is the SP-column flowthrough. Lanes 1, 2 and 3 are selected consecutive fractions. (b) Single monodisperse peak corresponding to an  $R_h$  value of 2.44 nm (13.5% polydispersity).

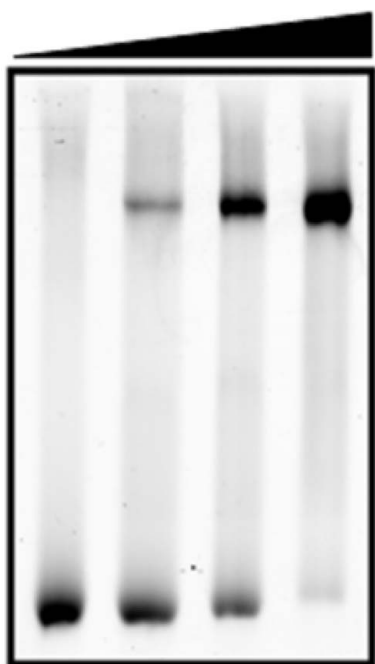
program (Gouet *et al.*, 1999) using the *MultAlin* similarity function with a 0.7 global score value. Secondary-structure prediction was carried out using the *ProteinPredict* server (Rost *et al.*, 2004).

**3. Results and discussion**

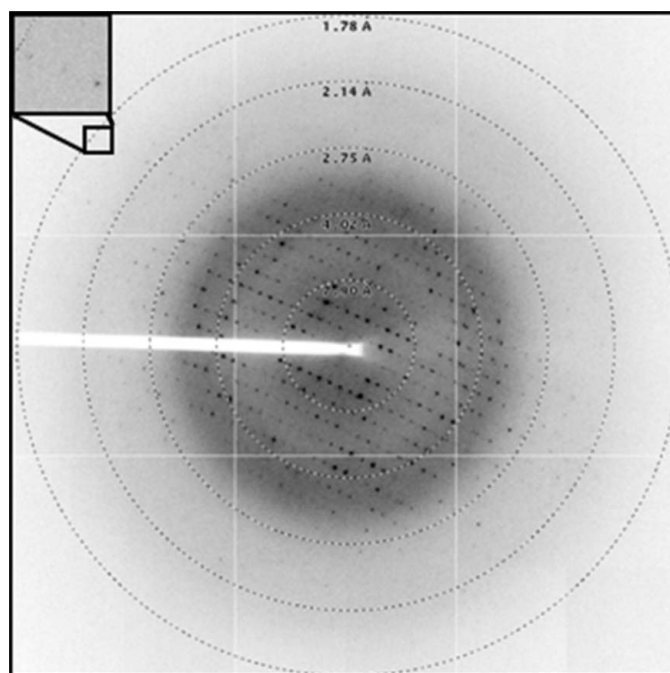
Amino-acid sequence alignment of *C.Csp231I* and *C.AhdI* revealed distinct regions of homology, particularly in the core of the protein (Fig. 1). Overall, the amino-acid sequence identity between these two proteins was 29% over 62 core residues. The main differences in the *C.Csp231I* sequence were a 12-amino-acid truncation at the

N-terminus, a four-amino-acid insertion adjacent to the predicted helix-turn-helix motif and a 32-amino-acid extension of the C-terminal region. The *ProteinPredict* program was used to estimate the secondary structure of *C.Csp231I*. The five characteristic helices conserved among the known C-protein structures were predicted, together with two additional helices located in the extended C-terminal region. Given the significant differences between *C.Csp231I* and the controller protein structures known to date, both in terms of protein structure and DNA-recognition sequence, we decided to embark on a structural analysis of the *C.Csp231I* protein.

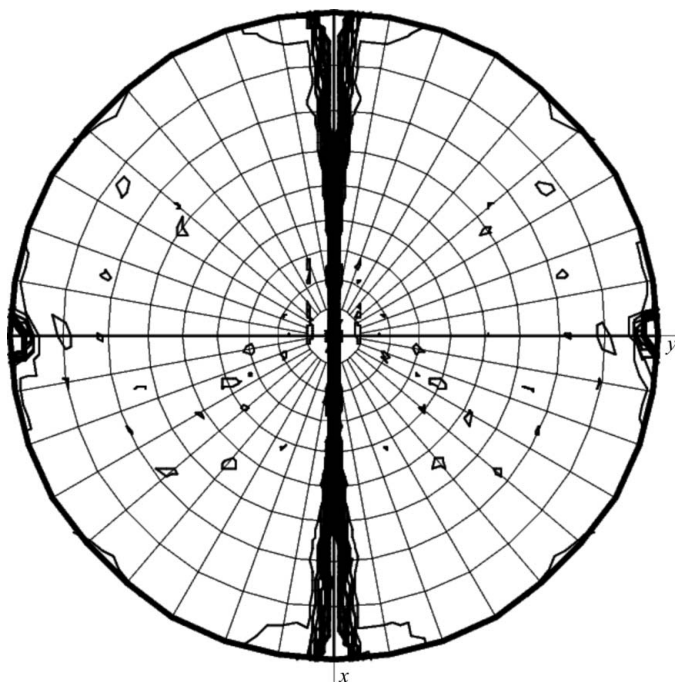
The controller protein *C.Csp231I* was overexpressed in *E. coli* and purified to homogeneity with a final yield of 5 mg l<sup>-1</sup> (Fig. 2a). The molecular mass of the protein was measured as 11 360 Da by



**Figure 3** EMSA of *C.Csp231I* with hex-labelled 96 bp duplex DNA. Increasing concentrations of *C.Csp231I* were incubated with a hex-labelled 96 bp duplex DNA fragment located upstream of the *csp231I*IC start codon at protein:DNA ratios of 0:1, 1:1, 2:1 and 4:1. The DNA concentration was 800 nM throughout.



**Figure 4** Diffraction image from a crystal of *C.Csp231I*. Reflections were observed to a resolution of approximately 1.8 Å (inset).



**Figure 5** Self-rotation function showing NCS. The self-rotation function is shown using a  $\kappa$  angle of  $180^\circ$  and was calculated using data in the resolution range 20.0–4.0 Å. The radius of integration was 35 Å.

electrospray mass spectrometry (University of Leeds, England), which is within 1 Da of that predicted from the amino-acid sequence. The hydrodynamic radius of the protein was measured as 2.4 nm by dynamic light scattering (Fig. 2*b*), from which an estimated molecular mass of 27 kDa was obtained, suggesting that *C.Csp231I* forms homodimers in solution.

In order to confirm the biological activity of the putative transcriptional regulator, its DNA-binding activity was assessed by an electrophoretic gel mobility assay (EMSA), using as substrate a hex-labelled 96 bp oligonucleotide corresponding to a region located directly upstream of the *csp231IC* start codon. Strong binding was observed, with a full shift at a protein:DNA ratio of 4:1 (Fig. 3), suggesting that two protein dimers may interact with this DNA sequence.

The crystallization conditions for *C.Csp231I* were obtained using the PACT screen (Molecular Dimensions Ltd) and were optimized to produce a number of single crystals suitable for X-ray analysis. Single plate-like crystals of approximately 100  $\mu\text{m}$  in length were observed after one month in reservoir conditions consisting of 0.1 M malate–MES–Tris (MMT) pH 7.0 and 20% (*w/v*) polyethylene glycol (PEG) 1500. The best crystals diffracted to 1.8 Å resolution (Fig. 4), although diffraction was anisotropic and the crystals diffracted less well in other directions. Nevertheless, by probing the crystal to determine the optimum collection volume a complete data set could be collected to 2.0 Å resolution. The data were processed in space group  $P2_1$  and a self-rotation function analysis was performed using *MOLREP* (Vagin

& Teplyakov, 1997). The plot at  $\kappa = 180^\circ$  reveals peaks additional to those resulting from the crystallographic  $2_1$  screw axis, indicating the presence of a noncrystallographic twofold-symmetry axis (Fig. 5). This suggests that *C.Csp231I* forms a dimer in the asymmetric unit of this crystal form, in common with other solved C-protein structures, resulting in a calculated Matthews coefficient of  $2.01 \text{ \AA}^3 \text{ Da}^{-1}$  (Matthews, 1968).

Phase determination by molecular replacement has so far been unsuccessful. This may be an indication of significant differences in the structure of *C.Csp231I* compared with the three available search models *C.AhdI* (McGeehan *et al.*, 2005), *C.Esp1396I* (McGeehan *et al.*, 2008) and *C.BclII* (Sawaya *et al.*, 2005). Indeed, such differences in structure would not be surprising given the presence of the 33-amino-acid extension and 12-amino-acid deletion at the C- and N-termini, respectively. Further attempts to solve the structure by molecular replacement, or if necessary by MAD, are in progress in order to provide detailed structural information on this new class of controller proteins.

We would like to thank Dave Hall and the beamline staff at the Diamond Light Source, UK for their assistance. This research was funded by a BBSRC project grant to G GK (grant reference BB/E000878/1). We would also like to thank RCUK for the provision of an academic fellowship to JEM.

## References

- Armougoum, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V. & Notredame, C. (2006). *Nucleic Acids Res.* **34**, W604–W608.
- Bogdanova, E., Djordjevic, M., Papapanagiotou, I., Heyduk, T. & Severinov, K. (2008). *Nucleic Acids Res.* **36**, 1429–1442.
- Česnavičienė, E., Mitkaitė, G., Stankevičius, K. & Janulaitis, A. (2003). *Nucleic Acids Res.* **31**, 743–749.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Gouet, P., Courcelle, E., Stuart, D. I. & Métoz, F. (1999). *Bioinformatics*, **15**, 305–308.
- Ives, C. L., Nathan, P. D. & Brooks, J. E. (1992). *J. Bacteriol.* **174**, 7194–7201.
- Kita, K., Tsuda, J. & Nakai, S. Y. (2002). *Nucleic Acids Res.* **30**, 3558–3565.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **26**.
- Lubys, A., Jurenaite, S. & Janulaitis, A. (1999). *Nucleic Acids Res.* **27**, 4228–4234.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McGeehan, J. E., Streeter, S. D., Papapanagiotou, I., Fox, G. C. & Kneale, G. G. (2005). *J. Mol. Biol.* **346**, 689–701.
- McGeehan, J. E., Streeter, S. D., Thresh, S. J., Ball, N., Ravelli, R. B. G. & Kneale, G. G. (2008). *Nucleic Acids Res.* **36**, 4778–4787.
- Rimšeliene, R., Vaišvila, R. & Janulaitis, A. (1995). *Gene*, **157**, 217–219.
- Rost, B., Yachdav, G. & Liu, J. (2004). *Nucleic Acids Res.* **32**, W321–W326.
- Sawaya, M. R., Zhu, Z., Mersha, F., Chan, S. H., Dabur, R., Xu, S. Y. & Balendiran, G. K. (2005). *Structure*, **13**, 1837–1847.
- Semenova, E., Minakhin, L., Bogdanova, E., Nagornyykh, M., Vasilov, A., Heyduk, T., Solonin, A., Zakharova, M. & Severinov, K. (2005). *Nucleic Acids Res.* **33**, 6942–6951.
- Sorokin, V., Severinov, K. & Gelfand, M. S. (2009). *Nucleic Acids Res.* **37**, 441–451.
- 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.