Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray analysis of ocr, the product of gene 0.3 of bacteriophage T7

Shane S. Sturrock,[a] David T. F. Dryden,[b] Constandache Atanasiu,[a] Jacqueline Dornan,[a] Sandra Bruce,[a] Andrew Cronshaw,[a] Paul Taylor[a] and Malcolm D. Walkinshaw[a]*

[a]Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh, Scotland, and [b]Department of Chemistry, University of Edinburgh, Edinburgh, Scotland

Correspondence e-mail: m.walkinshaw@ed.ac.uk

Received 22 May 2001
Accepted 10 July 2001

Ocr, the product of gene 0.3 of bacteriophage T7, prevents the action of restriction endonucleases of the host bacteria. The amino-acid sequence of ocr has less than 20% similarity to any protein of known three-dimensional structure. Ocr has been crystallized in a number of different crystal forms and X-ray data for the seleno-L-methionine-substituted form has been collected to a resolution of 1.8 Å. The presence of caesium was found to be required for good crystal growth. Anomalous X-ray data was used to identify possible positions for Se and Cs atoms in the unit cell.

1. Introduction

Bacterial cells have evolved a wide range of defence mechanisms against viral infection. One such mechanism is effected through restriction-modification (R-M) systems such as the type I R-M enzymes (Bickle & Kruger, 1993; Murray, 2000). These enzymes allow the bacteria to recognize their own DNA through the strain-specific methylation of bases by transferring methyl groups from *S*-adenosylmethionine to specific adenine or cytosine bases in the recognition site. Where the host DNA is hemimethylated (only one strand is modified) after DNA replication, the enzyme will methylate the other strand to produce fully methylated DNA. Cleavage of unmethylated DNA (for example, from an invading bacteriophage) by type I R-M systems requires two type I R-M enzymes attached at two distant recognition sites. A complex DNA translocation process driven by ATP hydrolysis occurs prior to DNA cleavage (Studier & Bandyopadhyay, 1988; Ellis *et al.*, 1999; Berge *et al.*, 2000).

To counteract the bacterial R-M systems, bacteriophages have developed anti-restriction systems (Kruger & Schroeder, 1981; Kruger & Bickle, 1983; Bickle & Kruger, 1993). It has been shown that bacteriophage T7 is not subject to restriction or modification by *Escherichia coli* B and *E. coli* K type I R-M systems *in vivo* (Eskridge *et al.*, 1967) but is *in vitro* (Eskin *et al.*, 1973; Bandyopadhyay *et al.*, 1985). The ocr protein product of gene 0.3, a 26 kDa dimer (Mark & Studier, 1981), is responsible for this protection (Studier, 1975). *Ocr* is the first gene on T7 DNA and is the first expressed after infection (Studier, 1972, 1973). Ocr is thought to act as a polyanion mimic of DNA and competes with DNA for the DNA-binding site of the type I R-M enzymes (Dunn *et al.*, 1981; Bandyopadhyay *et al.*, 1985). The inhibition may be similar to the interaction of uracil glycosylase inhibitor with the DNA repair enzyme uracil glycosylase (Mol *et al.*, 1995; Savva & Pearl, 1995; Putnam *et al.*, 1999). Ocr is very acidic: 34 of its amino acids are either aspartic or glutamic acids and only six are arginine or lysine (Dunn *et al.*, 1981) (Fig. 1). Structure predictions (Dryden *et al.*, 1995; Sturrock & Dryden, 1997) have presented a working model of the architecture of the binding site for ocr on a type I R-M enzyme. A crystal structure will provide a clear insight into how ocr mimics DNA and the nature of the interaction of ocr with type I R-M enzymes.

2. Results and discussion

2.1. Protein production and characterization

The wild-type ocr protein was produced by transformation of *E. coli* strain BL21(DE3)-pLysS with the plasmid pAR2993. Transformed cells were plated out on agar with 20 µg ml^{-1} ampicillin and 25 µg ml^{-1} chloramphenicol to produce single colonies. A single colony was used to innoculate flasks containing 500 ml LB media supplemented with ampicillin and chloramphenicol. When the optical density at 650 nm reached 0.6, IPTG was added to a final concentration of 0.4 m*M*. After incubation for 2 h at 310 K, the cells were harvested by centrifugation. The cells were sonicated and cell debris removed by centrifugation for 3 h at 30 000*g*. The supernatant was loaded onto a 200 × 16 mm DEAE Sepharose fast flow ion-exchange column equilibrated with 20 m*M* Tris–HCl, 7 m*M* β-mercaptoethanol, 300 m*M* NH_4Cl at pH 8 at 48 ml h^{-1}. A 500 ml gradient from 0.3 to 1 *M* NH_4Cl was run overnight at 24 ml h^{-1}. Frac-

Table 1
Data sets.

λ (Å)	0.87	1.2	0.9777	0.9793
Max. resolution	1.8	1.85	1.8	1.8
Completeness (final shell)	99.3 (98.5)	97 (85)	99.1 (96.7)	99.1 (97.1)
R_{merge} (final shell)	4.9 (9.6)	4.8 (12)	5.7 (13.5)	4.7 (15.9)
Total reflections	71916	65144	81164	72034
Unique	8927	8922	9012	9033
Anomalous diffraction ratio (%)	4.98	5.20	6.51	5.19
Se (f', f'')	−1.0, 3.0	−1.4, 0.73	−6.9, 0.5	−6.9, 0.5
Cs (f', f'')	0.0, 3.0	0.0, 5.3	0.2, 3.7	0.2, 3.7

† Anomalous diffraction ratio = $\langle||F+| - |F-||\rangle/\langle 1/2(|F+| + |F-|)\rangle$.

Table 2
Fractional coordinates of the Se and Cs atoms used to generate the Patterson peaks shown in Fig. 3.

	x	*y*	*z*
Se1	0.08063	0.27275	0.24952
Se2	0.20312	0.16840	0.51414
Se3	0.08258	−0.00231	0.04150
Se4	0.10748	−0.31247	0.00709
Cs1	0.15214	−0.41763	0.18218
Cs1	0.08887	0.12099	−0.11881
Cs1	0.33431	0.20409	0.09952

tions containing ocr were put through a HiLoad 16/60 Superdex 200 preparative grade gel-filtration column in 20 m*M* Tris–HCl, 7 m*M* β-mercaptoethanol, 300 m*M* NH_4Cl pH 8 buffer. An additional step was tried using a 150 × 16 mm phenyl Sepharose affinity column equilibrated in 20 m*M* Tris–HCl, 7 m*M* β-mercaptoethanol, 50 m*M* NH_4Cl, 2.4 *M* $(NH_4)_2SO_4$ buffer at pH 8. Protein was applied to the column and flushed with this buffer for 30 min at 60 ml h^{-1} and protein was then eluted at 48 ml h^{-1} with a 500 ml gradient from 2.4 to 0 *M* $(NH_4)_2SO_4$. This achieved a higher level of purity but was not found to be necessary for the production of good crystals. Chromatography columns were from Amersham Pharmacia. The purified ocr protein was stored in 20 m*M* Tris–HCl, 7 m*M* β-mercaptoethanol, 300 m*M* NH_4Cl pH 8 buffer containing 50% glycerol at 253 K.

A selenomethionine-substituted form of ocr was produced by transforming the methionine-deficient *E. coli* strain B834. This strain (kindly provided by Dr W. Hunter, Dundee University, Scotland) was transformed with pAR2993 as with BL21(DE3) and grown in minimal media supplemented with 20 μg ml^{-1} of all amino acids except cysteine (to prevent reutilization of sulfur to make normal methionine) and with 20 μg ml^{-1} seleno-L-methionine along with 20 μg ml^{-1} ampicillin.

Electrospray mass spectra were measured using a Themoquest LCQ mass spectrometer (Thermoquest Corporation, Paradise, Hemel Hempstead, England). Native ocr gave M_r = 13 677 ± 1 Da, compared with a mass of M_r = 13 959 ± 1 Da for the selenium form. The difference in mass of 47 Da × 6 indicates 100% incorporation of six Se atoms.

2.2. Crystallization and X-ray data analysis

Crystallization was carried out using the hanging-drop vapour-diffusion method. The drop consisted of equal volumes (normally 2 μl) of protein solution and well solution. Crystallization trials were performed with Hampton Research sets I and II and crystalline precipitate was observed with polyethylene glycol (PEG) 8000, Tris–HCl buffer and sodium acetate. Refinement of these conditions led to well formed crystals in 32% PEG, 300 m*M* NH_4Cl, 0.1 *M* Tris–HCl pH 8.0 and 0.2 *M* sodium acetate at 291 K for both native and selenomethionine forms of the protein (type I). This crystal form was metastable and after about 10 d the crystals gradually dissolved. In some of the drops, poorly formed plate-like crystals reappeared after days or weeks (type II).

Caesium acetate was also used in crystal trials and found to produce large well formed crystals of type I. Trials with rubidium acetate and potassium acetate yielded progressively poorer crystals ranging between the caesium and sodium acetate crystal quality. No crystals were obtained when ammonium acetate was used as an additive. Crystals of type I, irrespective of size or external morphology, diffracted poorly. An initial data set for the wild type was collected to 2.5 Å in space group *C*2, with unit-cell parameters *a* = 88.18, *b* = 105.96, *c* = 73.34 Å, β = 120.319°.

The observation that caesium seemed to help crystal growth was further investigated. Using CsCl, a large plate-like crystal (Fig. 2) grew within two weeks under the following conditions: well solution 32% PEG 8000, 0.1 *M* Tris–HCl pH 8.2, 0.2 *M* CsCl, 2.8% glycerol; drop 2 μl well solution plus 2 μl of protein solution at 16 mg ml^{-1}. This selenomethionine form of the ocr crystals could be flash-frozen in liquid nitrogen without use of additional cryoprotectant. These crystals also belong to space group *C*2 and have unit-cell parameters *a* = 78.78, *b* = 37.66, *c* = 35.49 Å, β = 98.41°. Anomalous data sets were collected for the selenomethionine crystals on station 9.5 at the Daresbury synchrotron and station BM30 at the ESRF, Grenoble (Table 1).

The fluorescence spectrum was measured to determine the absorption edge of selenium and data sets were measured at four wavelengths; the inflection point of the edge (0.9777 Å, λ_2), the peak of the white-line absorption (0.9793 Å, λ_3) and at two remote wavelengths (0.87 Å, λ_1; 1.2 Å, λ_4). The merging *R* factor between Friedel pairs within each data set does indicate a slightly larger anomalous signal at the absorption edge (6.9% compared with the remote *R* factors of 5.7) and this is supported by the size of the anomalous diffraction ratios for the four data sets (Table 1). Anomalous Patterson difference maps for each of the four data sets were calculated (Fig. 3). Major differences between these maps is explained by the unexpected discovery that the crystals incorporated significant amounts of caesium, which has a large anomalous contribution over this wavelength range (Fig. 4).

```
....,....1....,....2....,....3....,....4....,....5....,....6
AMSNMTYNNVFDHAYEMLKENIRYDDIRDTDDLHDAIHMTADNAVPHYYADIFSVMASEG
        HHHHHHHHHHHHH          HHHHHHHHHHHHH    HHHHHHHHHHHHH

....,....7....,....8....,....9....,....10...,....11...,....12
IDLEFEDSGLMPNTKDVIRILQARIYEQLTIDLWEDAEDLLNEYLEEVEEYEEDEE
  EEEE         HHHHHHHHHHHHHHHH HHHHHHHHHHHHHHHHHHHH
```

Figure 1
Sequence of ocr. The secondary-structure prediction for ocr is from the program *PHD* (Rost & Sander, 1993).

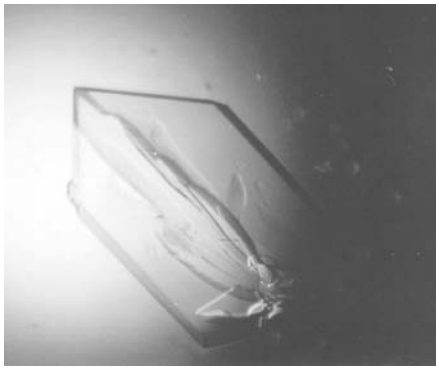

Figure 2
A crystal of the selenium form of ocr with dimensions 0.4 × 0.4 × 0.1 mm.

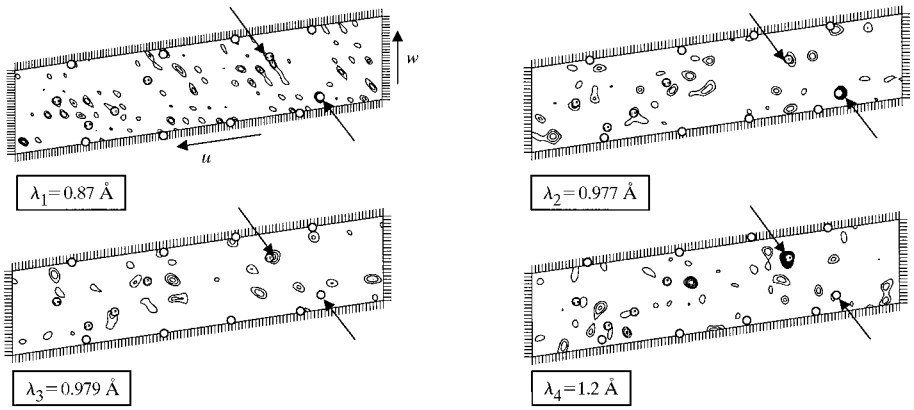

Figure 3
Harker sections calculated at $y = 0$. The Fourier coefficients for calculating the Patterson map for each of the four data sets are $(|F_i+| - |F_i-|)^2$. The section runs from $u = 0$ to 1 in 108ths and from $w = 0$ to 0.5 in 48ths. The stippled circles correspond to the Harker peaks resulting from the Cs—Cs vectors from three independent caesium ions. The pale filled circles correspond to the Harker peaks of the Se—Se vectors from four of the Se atoms. Patterson maps were calculated and plotted using the program *CNS* (Brunger *et al.*, 1998). Arrows indicate positions of the strongest selenium and caesium peaks.

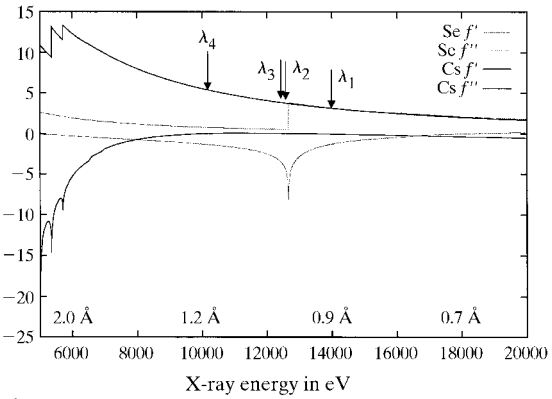

Figure 4
Plot of the f' and f'' values for Cs (dark lines) and Se (gray lines) from 6000 eV (2 Å) to 18 000 eV (0.7 Å) and plotted using tools from the web page http://www.bmsc.washington.edu/scatter.

A good solution to the Patterson maps has been obtained using the program *SOLVE* and most of the Harker peaks can be explained by four Se atoms and three Cs atoms. The strongest selenium peak (u = 0.17, 0, 0.08) is in the map using data calculated at the absorption edge λ_2 = 0.98 Å (Fig. 3). This Patterson peak is also strong in the map calculated from data measured at λ_1 = 0.87 Å. The selenium signal is very much reduced at the longer wavelengths above the selenium absorption edge. The anomalous Patterson peaks calculated with data measured at λ_3 = 0.981 Å and λ_4 = 1.2 Å are dominated by the caesium signal. This is clearly seen by the presence of the strong peak u = 0.31, v = 0, w = 0.38 (Fig. 3). The changes in intensities of the anomalous Patterson peaks can be explained by inspection of Fig. 4, which shows that the anomalous contribution of caesium varies between 6 electrons and 4 electrons over the four wavelengths, while the selenium contribution is only significant at the shorter wavelengths. The position of the heavy atoms (Table 2) determined by *SOLVE* (Terwilliger & Berendzen, 1999) has provided an electron-density map that is of sufficient quality to solve the structure. Preliminary refinement has confirmed the assignment of the selenium and caesium positions.

We thank the Leverhulme Trust for providing funds for this research work (Grant F/158/BC) and the staffs of ESRF and Daresbury SRS for help with data collection. DD thanks the Royal Society for a University Research Fellowship and CA was supported by a Darwin Trust studentship. Plasmid pAR1993 was a kind gift from William Studier and Alan Rosenberg (Brookhaven National Laboratories).

References

Bandyopadhyay, P. K., Studier, F. W., Hamilton, D. L. & Yuan, R. (1985). *J. Mol. Biol.* **182**, 567–578.
Berge, T., Ellis, D. J., Dryden, D. T. F., Edwardson, J. M. & Henderson, R. M. (2000). *Biophys. J.* **79**, 479–484.
Bickle, T. A. & Kruger, D. H. (1993). *Microbiol. Rev.* **57**, 434–450.
Brunger, A. T., Adams, P. A., 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.
Dryden, D. T. F., Sturrock, S. S. & Winter, M. (1995). *Nature Struct. Biol.* **2**, 632–635.
Dunn, J. J., Elzinga, M., Mark, K.-K. & Studier, F. W. (1981). *J. Biol. Chem.* **256**:2579–2585.
Ellis, D. J., Dryden, D. T., Berge, T., Edwardson, J. M. & Henderson, R. M. (1999). *Nature Struct. Biol.* **6**, 15–17.
Eskin, B., Lautenberger, J. A. & Linn, S. (1973). *J. Virol.* **11**, 1020–1023.
Eskridge, R., Weinfield, H. & Paigen, K. (1967). *J. Bacteriol.* **93**, 835–844.
Kruger, D. H. & Bickle, T. A. (1983). *Microbiol. Rev.* **47**, 345–360.
Kruger, D. H. & Schroeder, C. (1981). *Microbiol. Rev.* **45**, 9–51.
Mark, K.-K. & Studier, F. W. (1981). *J. Biol. Chem.* **256**, 2573–2578.
Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W. & Tainer, J. A. (1995). *Cell*, **82**, 701–708.
Murray, N. E. (2000). *Microbiol. Mol. Biol. Rev.* **64**, 412–434.
Putnam, C. D., Shroyer, M. J., Lundquist, A. J., Mol, C. D., Arvai, A. S., Mosbaugh, D. W. & Tainer, J. A. (1999). *J. Mol. Biol.* **282**, 331–346.
Rost, B. & Sander, C. (1993). *J. Mol. Biol.* **232**, 584–599.
Savva, R. & Pearl, L. H. (1995). *Nature Struct. Biol.* **2**, 752–757.
Studier, F. W. (1972). *Science*, **176**, 367–376.
Studier, F. W. (1973). *J. Mol. Biol.* **79**, 237–248.
Studier, F. W. (1975). *J. Mol. Biol.* **94**, 283–295.
Studier, F. W. & Bandyopadhyay, P. K. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 4677–4681.
Sturrock, S. S. & Dryden, D. T. F. (1997). *Nucleic Acids Res.* **25**, 3408–3414.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.