# Overexpression, purification, crystallization and preliminary X-ray diffraction analysis of the pMV158-encoded plasmid transcriptional repressor protein CopG

F. Xavier Gomis-Rüth[a,*], María Solà[a], Rosa Pérez-Luque[a], Paloma Acebo[b], M. Teresa Alda[b], Ana González[c], Manuel Espinosa[b], Gloria del Solar[b], Miquel Coll[a]

[a] *Centre d'Investigació i Desenvolupament, C.S.I.C., Jordi Girona, 18-26, 08034 Barcelona, Spain*
[b] *Centro de Investigaciones Biológicas, C.S.I.C., Velázquez, 144, 28006 Madrid, Spain*
[c] *European Molecular Biology Organization, c/o DESY, Notkestraße 85, 22603 Hamburg, Germany*

Received 16 February 1998

**Abstract Plasmid pMV158 encodes a 45 amino acid transcriptional repressor, CopG, which is involved in copy number control. A new procedure for overproduction and purification of the protein has been developed. The CopG protein thus obtained retained its ability to specifically bind to DNA and to repress its own promoter. Purified CopG protein has been crystallized using the sitting-drop vapor diffusion method. The crystals, belonging to orthorhombic space group $C222_1$ (cell constants $a = 67.2$ Å, $b = 102.5$ Å, $c = 40.2$ Å), were obtained from a solution containing methylpentanediol, benzamidine and sodium chloride, buffered to pH 6.7. Complete diffraction data up to 1.6 Å resolution have been collected. Considerations about the Matthews parameter account for the most likely presence of three molecules in the asymmetric unit (2.27 Å$^3$/Da).**



*Key words:* Plasmid; Transcriptional repressor protein; CopG; Crystallization; X-ray analysis

## 1. Introduction

Bacterial plasmids are extrachromosomal DNA elements which replicate in an autonomous and controlled way within their hosts. Thus, plasmids maintain a fixed number of copies, although copy number of a given plasmid may vary within different hosts and under different growth conditions [1–3]. Copy number is controlled by plasmid-encoded elements that regulate the key stage of replication, the initiation event. Initiation is usually mediated by a plasmid-encoded protein (generically termed Rep) which, in many cases, is rate-limiting. Control systems consist in plasmid-encoded negative regulatory circuits, which operate by correcting deviations from the average copy number in individual cells [1]. The control systems do so by either increasing or decreasing the rate of initiation of replication frequency.

In some plasmids, replication is controlled by the joint action of a constitutively expressed antisense RNA, and of a transcriptional repressor protein [4–6]. Such is the case of the streptococcal replicon pMV158, which is the prototype of a family of multicopy plasmids replicating by the rolling circle mechanism [6]. In pMV158, the genes encoding the repressor CopG and the initiator of replication RepB proteins are co-transcribed from the single promoter $P_{cr}$ (see Fig. 1A; [7,8]). The *copG* gene product is a 5.1 kDa protein (previously addressed as RepA; [8]) which binds to a DNA region that includes the promoter $P_{cr}$, thus hindering the binding of the host RNA polymerase. DNase I and hydroxyl radical footprinting experiments performed with purified CopG showed that its DNA target spans 48 nucleotides (pMV158 coordinates 581–628), the protein binding to five successive helix turns by the same face of the DNA helix. The center of the CopG-DNA binding region contains a 13 bp sequence showing a two-fold rotational symmetry [7]. Mutations and/or deletions inactivating *copG* gene lead to plasmids with a five-fold increased copy number [4,6]. Since the genetic organization of plasmids of the pMV158 family is similar, the same regulatory circuit is expected to exist in all of them [4,6,9]. A region exhibiting an intrinsic DNA curvature was found at the vicinity of the CopG target region and, upon addition of purified CopG protein, a strong induced bend was found [10]. The angle of the CopG-induced bend has been calculated to be at least of 100 degrees (G. del Solar, unpublished). DNA bends or DNA loops mediated by proteins binding at the vicinity of promoters seem to be a common mechanism to repress or to activate gene expression [11–13].

Initial sequence-based secondary structure analyses of CopG indicated the existence of a putative bihelical unit separated by a short turn. Comparison with transcriptional repressors indicated that CopG could have a putative helix-turn-helix motif, both helices spanning from segments Ser-12 to Met-24 (α-helix A), and from Lys-28 to Glu-37 (α-helix B), out of the 45 residues of CopG (see Fig. 1C). In addition, a sequence-based homology search against the proteins deposited in the Protein Data Bank indicated that CopG shares significant putative structural homology with the phage P22 Arc and Mnt repressors, *met* myoglobin, a central 37-residue diphtheria toxin catalytic domain segment, segment 61 to 104 of adenylate kinase, and the first two thirds of the phage λ-Cro repressor. However, analysis of the tertiary structures of these proteins around the theoretical homology region yielded no significant conclusions. In fact, the regular secondary structure elements deviate in location and spatial distribution, reaching from a single long helix in the case of diphtheria toxin to the helix-turn-helix of the Cro repressor, considered a homology model for CopG [8], or the strand-turn-helix-turn-helix structure of the Arc repressor. To investigate the detailed three-dimensional structure of CopG, we have undertaken its improved overexpression, developed a new method for large scale purification and achieved its crystallization, hereby presenting the preliminary crystallographic results.

*Corresponding authors: F. Xavier Gomis-Rüth and Miquel Coll;
Fax: +34 (3) 204 59 04.
E-mail: xgrcri@cid.csic.es; mcccri@cid.csic.es


*PII* S0014-5793(98)00219-1

This is the first report on the crystallization of a plasmid-encoded repressor protein.

## 2. Materials and methods

### 2.1. Biological material

*Escherichia coli* BL21(DE3) ($r_B^-$ $m_B^-$, *gal*, *ompT*, *int*::$P_{lacUV5}$-T7 gene 1 *imm21 nin5*) was employed as the host for overexpression of *copG*. This strain has a single copy of the phage T7 RNA polymerase gene under the control of the IPTG-inducible *lac*UV5 promoter, integrated into the chromosome [14]. Plasmids used were the pMV158 derivative pJS3, which carries a *cat* gene [15], and the commercially available plasmid pALTER-1 (Promega). This latter plasmid carries genes for resistance to tetracycline, and contains the T7 RNA polymerase-directed ϕ10 promoter. Plasmid pALTER-1 was doubly digested with *Pst*I and *Cla*I (inactivating the *tet* gene), and the large fragment was ligated to the 2296 base pair long *Hgi*AI (compatible with *Pst*I) *Cla*I fragment of plasmid pJS3. This latter fragment contains promoterless *copG* and *repB* genes and an intact *cat* gene. *E. coli* BL21(DE3) transformants were selected for resistance to chloramphenicol (30 μg ml$^{-1}$). The resulting expression plasmid, termed pALT7:crcat (Fig. 1B), carries genes *copG*, *repB* and *cat* under the control of the T7 ϕ10 promoter. The nucleotide sequence of the entire *copG* and *repB* genes in the expression plasmid was determined.

### 2.2. Large scale purification of CopG

Cultures of *E. coli* BL21(DE3) harboring pALT7:crcat were grown in M9 medium [16] to middle exponential phase ($3\times10^8$ colony forming units per ml), treated with 1 mM IPTG for 7 min, and then with rifampicin (200 μg ml$^{-1}$) to selectively express genes under the ϕ10 promoter [8]. Stability of CopG in the cultures was measured and resulted to be very high. Thus, the induced cultures were incubated for 12–18 h, which resulted in an increased yield of CopG protein. Cells were collected by centrifugation and washed twice with buffer A (500 mM NaCl, 20 mM Tris/HCl pH 7.5, 1 mM EDTA, 1 mM DTT, 1% ethylene glycol). All subsequent steps were carried out at 4°C. Pellets from 4 l cultures were pooled and concentrated 100-fold. The cell paste was passed twice through a chilled cell-pressure French Press (140 mPa). Cell debris was removed by high speed centrifugation (30 000 rpm in a Beckman Ti45 rotor), and the supernatant was treated with ammonium sulphate to 50% saturation. After 30 min on ice, samples were ultracentrifuged as above, and the CopG containing supernatant was collected. This was exhaustively dialyzed against buffer A containing 1.5 M NaCl. Samples were applied to a 1885 ml ($5\times96$ cm) acrylamide (BioGel P10, Bio-Rad) column equilibrated with the same buffer, and the flow rate was set at 30 ml h$^{-1}$. Fractions (11 ml) were analyzed for the presence of CopG by SDS-tricine-polyacrylamide gel electrophoresis. As standards, the commercially available Mark12 protein mixture (Novex) was used. Peak fractions were pooled and dialyzed against buffer B (20 mM Tris/HCl pH 8.0, 1 mM EDTA, 5 mM DTT, 5% ethylene glycol) containing 100 mM KCl. The sample was applied to a 98 ml ($2.5\times20$ cm) heparin-agarose (Affi-Gel, Bio-Rad) column, and the flow rate was set at 70 ml h$^{-1}$. CopG eluted at 250 mM KCl in a 100–400 mM KCl gradient. Fractions containing CopG were pooled and passed through a tangential flow device (Mini-ultrasette, Filtron) with a cut-off of 30 kDa. At this stage, CopG was considered to be pure enough (more than 99.9%) for crystallization. The filtrate was concentrated by passage through cartridges (Filtron, cut-off 1 kDa) until the CopG concentration was about 15 mg ml$^{-1}$. The yield obtained was of 1.5 mg of CopG per 4 l starting culture. CopG concentration was determined by quantitative amino acid analysis.

### 2.3. DNase I footprinting and in vitro transcription assays

A 260 bp *Dde*I-*Sty*I DNA fragment from plasmid pMV158 (coordinates 499–758, containing the target of CopG) was labelled at its 5′-*Dde*I end. DNA (0.67 nM) was incubated with various amounts of CopG protein and further treated with DNase I as described [8]. For in vitro transcription, the DNA fragment (5 nM) used as template was the 842 bp *Ban*I-*Pst*I fragment (coordinates 214–1056). This fragment contains promoters $P_{cr}$ and $P_{ctII}$, this latter directing synthesis of the countertranscript RNA II. In vitro synthesis of the run-off *cop-rep* mRNA and of the entire RNA II, in the absence or in the presence of protein CopG (50 and 100 nM), was performed as described [7].

### 2.4. Crystallization and X-ray diffraction analysis

Sitting-drop vapor diffusion method crystallization trials were performed at 20°C and 4°C using aqueous protein of 13–16 mg ml$^{-1}$ concentration in 10 mM Tris/HCl pH 7.5, further 275 mM in NaCl and 0.5 mM in $NaN_3$. Crystal screens were purchased from Hampton Research. Drops were prepared mixing 2 μl of protein solution and 2 μl of precipitating agent solution and allowed to equilibrate against a 0.25 ml reservoir containing the same solution. Crystals were mounted in thin-wall glass capillaries directly from the drop or after addition of harvesting buffer for further analysis. Alternatively, they were directly taken out of the original drop with a cryo loop and flash-frozen in a liquid nitrogen cryo stream (Oxford Cryosystems) at 100 K. Diffraction data at room temperature or cryo conditions were collected on a 300 mm MAR Research image plate detector attached to a Rigaku RU200 rotating anode generator providing monochromatized CuKα X-ray radiation. The data were processed with MOSFLM v. 5.41 [17] and programs of the CCP4 suite [18]. Further data were collected at the tuneable EMBL X31 beamline of the Deutsches Elektronensynchrotron in Hamburg (Germany).

## 3. Results and discussion

The new plasmid vector employed for the overproduction of CopG (Fig. 1B), in conjunction with the procedure developed here for the scale-up purification of CopG protein, have proved to be excellent. The degree of purification achieved has been very high (Fig. 2A), and CopG has been shown to be soluble under all conditions tested. This, and the remarkable high stability of CopG within the cell cultures, has allowed us to purify large quantities of this rather small protein. To test

Table 1
Data collection and processing statistics

| | Native (3 crystals) | Et-Hg-Cl derivative |
|---|---|---|
| Cell constants (*a*, *b*, *c*, in Å) | 67.19, 102.49, 40.21 | 67.54, 102.33, 40.20 |
| Resolution range (in Å) | 25.82–1.57 | 30.43–1.80 |
| X-ray source | rotating anode | synchrotron |
| Radiation wavelength (λ in Å) | 1.5418 | 1.002 |
| No. of measured reflections | 91 693 | 38 505 |
| No. of unique reflections | 18 358 | 11 102 |
| Completeness (in %)/$R_{merge}$/intensity [$I/\sigma(I)$] in | | |
| resolution range (19.61–1.60 Å) | 98.1/0.080/2.5 | |
| resolution range (1.69–1.60 Å) | 96.4/0.126/3.0 | |
| resolution range (19.59–1.80 Å) | | 98.9/0.037/15.7 |
| resolution range (1.85–1.80 Å) | | 87.0/0.177/4.0 |
| Average multiplicity | 5.0 | 3.5 |
| No. of measured data with anomalous contribution ($R_{merge}$) | | 11 297 (0.043) |

$R_{merge} = (\sum_h \sum_j |I(h)_j - \langle I(h)\rangle|)/(\sum_h \sum_j I(h)_j)$; $I(h)_j$ is the observed intensity of the *j*th measurement of reflection *h*, and $\langle I(h)\rangle$ the mean intensity of reflection *h*.

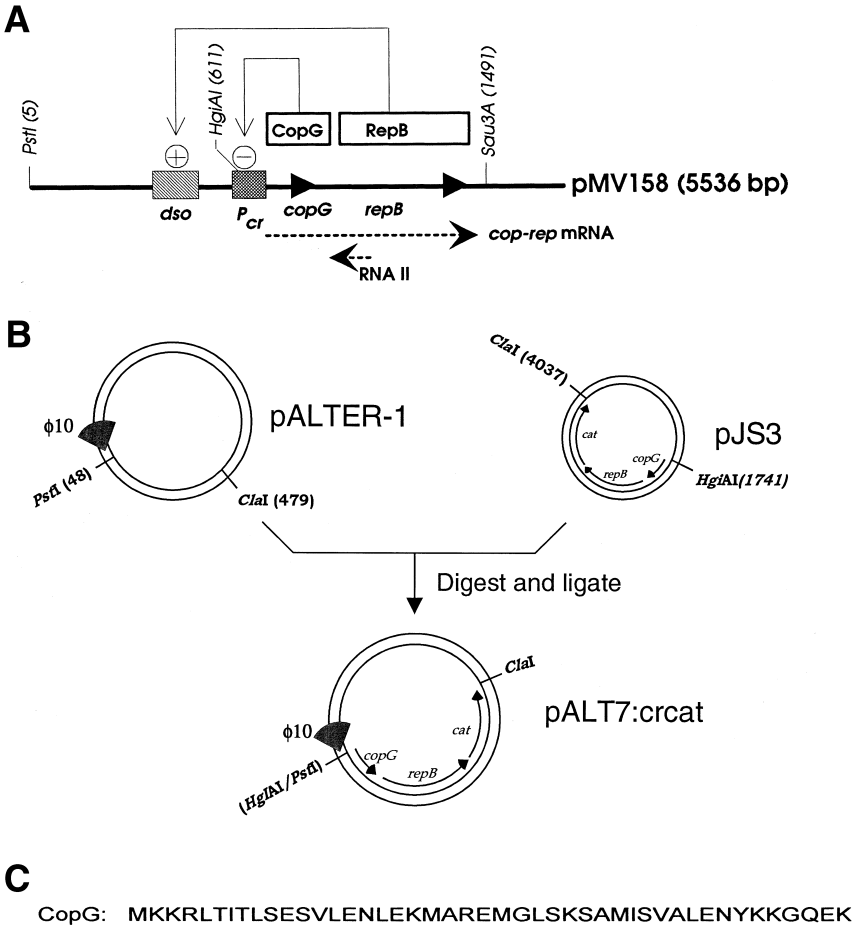


Fig. 1. The control region of plasmid pMV158 and construction of the expression plasmid. Coordinates of pMV158 are from [23,24]. A: Schematic map of the leading strand replication and control region of pMV158. The initiator of replication RepB protein is the positive effector, acting at the plasmid double strand origin (*dso*; stippled rectangle). The negative elements of the regulatory circuit are the antisense RNA II (inhibiting translation of *repB*) and CopG (repressing synthesis of the *copG-repB* mRNA). Pertinent restriction sites and coordinates of cleavage are indicated. Symbols: broken lines, RNAs; empty boxes, proteins; filled boxes, *dso* and promoter $P_{cr}$. B: Scheme of the construction of the expression vector pALT7:crcat employed for overexpression of *copG*. Coordinates of plasmid pJS3 are from [15]. Maps are not drawn to scale. C: Amino acid sequence of CopG as determined from sequencing [8] and from chemical synthesis [25] of the protein.

whether this new procedure of purification yielded an active protein, we performed two tests: (i) DNA binding of CopG to its target DNA, and (ii) CopG-mediated specific inhibition of transcription from $P_{cr}$ promoter. Binding of purified CopG was tested by DNase I footprinting experiments on the coding strand. The results showed the presence of five protected regions, which encompass a 13 bp symmetric element, and the −35 and −10 regions of promoter $P_{cr}$ (Fig. 2B). Thus, at a molar ratio of 6 CopG dimers per DNA molecule, more than 90% of the DNA was bound by CopG, indicating that the protein purified by this new procedure binds efficiently to its DNA target [8]. In vitro transcription assays were performed with a template pMV158-DNA fragment which carries promoter $P_{cr}$ (the one regulated by CopG), and the constitutive $P_{ctII}$ promoter (directing synthesis of the antisense RNA II). The DNA (5 nM) was incubated without CopG or with two concentrations of the protein (50 and 100 nM) prior to the addition of the RNA polymerase. The results (Fig. 2C) showed that CopG specifically inhibited synthesis of the *cop-rep* mRNA from promoter $P_{cr}$, whereas transcription from $P_{ctII}$ was unaffected by the presence of the protein. These results indicate that CopG retained its functionality as a transcriptional repressor.

Initially, crystal screens I and II based on the incomplete factorial approach and on sparse matrix sampling [19,20], systematic ammonium sulfate, sodium/potassium phosphate, sodium chloride and sodium citrate screenings considering different pH values and salt concentrations were assayed. Needles were obtained with 70% 2-methyl-2,4-pentanediol (MPD)/0.1 M HEPES pH 7.5. After several optimization steps including additives and detergents, best crystals suitable for X-ray analysis were obtained from drops mixed from 4 μl protein solution and 2 μl of precipitating agent solution, consisting of 72% MPD/0.1 M HEPES pH 6.7, and 0.5 μl of 20% (w/v) benzamidine/HCl solution (Fig. 3A). These crystals did not resist harvesting (with 90–100% MPD) nor direct mounting into a capillary for X-ray analysis. Considering the high percentage of MPD in the crystallization conditions, flash-freezing was experimented. These crystals diffracted properly (Fig. 3B) beyond 1.6 Å resolution and belong to orthorhombic space group $C222_1$ with cell constants $a = 67.2$ Å, $b = 102.5$ Å, $c = 40.2$ Å. Complete diffraction data have been collected from three crystals using cryo-cooling. The crystals result in being very isomorphous to each other and the data mergeable. A summary of the data collection and processing statistics is provided in Table 1.

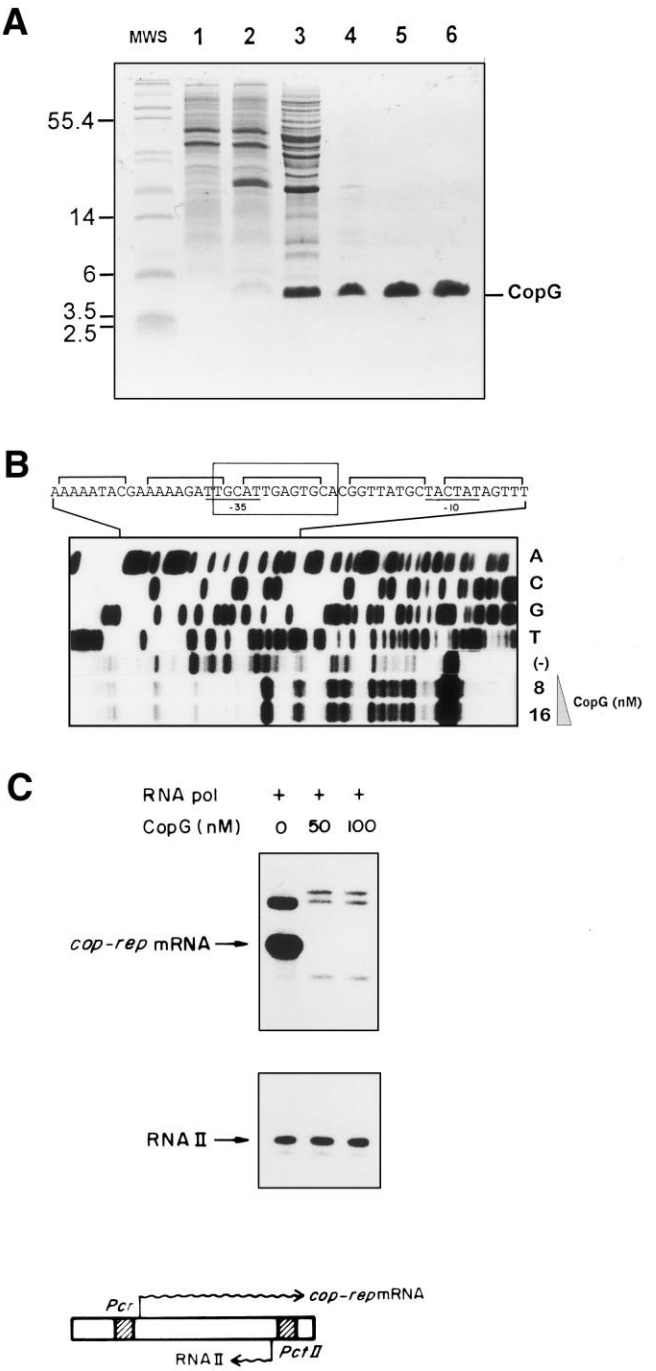


Fig. 2. CopG purification and functionality. A: Steps of purification as shown by electrophoresis on a 16% SDS-tricine-polyacrylamide gel. The proportion of the total volume of the samples loaded was: 0.017% of uninduced (lane 1) or induced (lane 2) cultures; 0.08% of ammonium sulphate supernatant (lane 3); 0.04% of peak fractions from acrylamide column (lane 4); 0.08% of peak fractions from heparine agarose column (lane 5); and 0.08% of the purified CopG preparation after ultrafiltration through cartridges with cut-off of 30 kDa and 1 kDa (lane 6). This sample was used for the preparation of crystals. MWS, molecular weight standards (relevant sizes, in kDa, indicated on the left). B: DNase I protection of the CopG target by purified protein. Two amounts of protein (8 and 16 nM, in monomer concentration) were used. The DNA sequence of the same fragment (A, C, G, T) is also shown. The nucleotide sequence of the labelled coding strand of pMV158, and the CopG-protected sites are indicated above, as well as the −35 and −10 regions of promoter $P_{cr}$ (underlined) and the 13 bp symmetric element (boxed). C: Transcripts synthesized from promoters $P_{cr}$ and $P_{ctII}$ in the absence (0) or presence (50 and 100 nM, in monomer concentration) of purified CopG protein. Bands corresponding to the run-off *cop-rep* mRNA and to the full length antisense RNA II are indicated. The band above the run-off transcript could be due to artefacts observed in DNA fragments with protruding ends [8]. Relative positions of the promoters in the DNA template are schematized below.

Considerations about the packing [21] renders a dimer ($V_m = 3.40$ Å$^3$/Da; solvent content 64%), a trimer ($V_m = 2.27$ Å$^3$/Da; solvent content 45%) or a tetramer ($V_m = 1.70$ Å$^3$/Da; solvent content 27%) as possible, being three molecules in the asymmetric unit the most probable.

As no phase information is available and no positive results

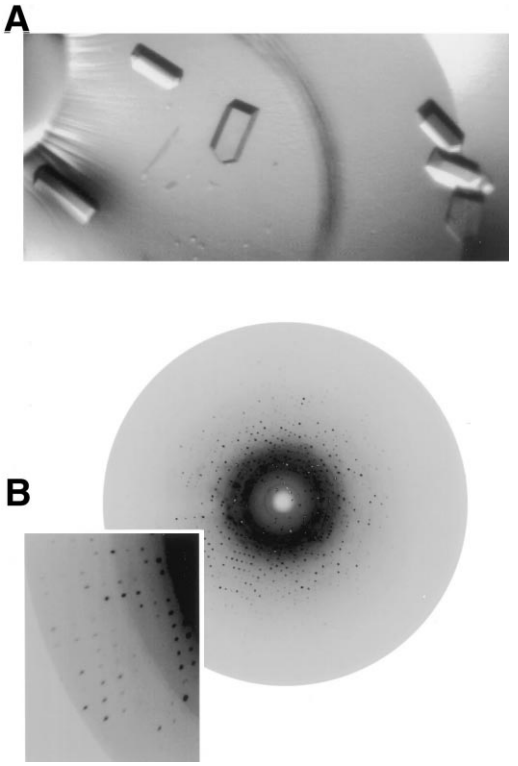


Fig. 3. A: Orthorhombic CopG crystals of maximal dimensions 0.5×0.2×0.2 mm. B: X-ray diffraction pattern recorded on a MAR Research image plate detector. The rotation range was 1 degree and the crystal to detector distance 90 mm. The resolution at the border of the plate (amplified area) is 1.57 Å.

were obtained from molecular replacement calculations performed with AMoRe [22], the search for heavy-ion derivatives was started. Because no suitable harvesting buffer was found, putative heavy-ion containing derivatives were prepared by adding small amounts of product directly to the drops and allowing to equilibrate for several days or weeks. No positive results were obtained by this method. An alternative approach consisting of co-crystallizing CopG in the presence of small amounts of heavy-ion compounds proved to be successful. A crystal obtained in the presence of Et-Hg-Cl was measured (see Table 1) at the Hg $f_{max}'$ wavelength of 1.002 Å at the DESY (Hamburg), rendering a clear resolution-consistent peak in the anomalous Patterson map consistent with position (0.106, 0.196, 0.470; height 4.53 σ). The collection of further derivative data for proper phasing is currently under way.

*Acknowledgements:* This work was supported by Grants PB95-0224 from the Ministerio de Educación y Ciencia, Spain (to M.C.), and BIO97-0347 from CICYT (to M.E.). The support of the Consejo Superior de Investigaciones Científicas, through an Action Program on Protein Structure and Function and the Generalitat de Catalunya (Centre de Referència en Biotecnologia) is acknowledged. Data collection at EMBL-Hamburg was supported by the European Union Large Installation Project CHGE-CT93-0040.

## References

[1] Nordström, K. (1983) Plasmid 9, 1–7.
[2] Novick, R.P. (1989) Annu. Rev. Microbiol. 43, 537–565.
[3] Espinosa, M., del Solar, G., Rojo, F. and Alonso, J.C. (1995) FEMS Microbiol. Lett. 130, 111–120.
[4] del Solar, G. and Espinosa, M. (1992) Mol. Microbiol. 6, 83–94.
[5] Brantl, S. (1994) Mol. Microbiol. 14, 473–483.
[6] del Solar, G., Acebo, P. and Espinosa, M. (1995) Mol. Microbiol. 18, 913–924.
[7] del Solar, G., de la Campa, A.G., Pérez-Martín, J., Choli, T. and Espinosa, M. (1989) Nucleic Acids Res. 17, 2405–2420.
[8] del Solar, G., Pérez-Martín, J. and Espinosa, M. (1990) J. Biol. Chem. 265, 12569–12575.
[9] Kwak, J.-H. and Weisblum, B. (1994) J. Bacteriol. 176, 5044–5051.
[10] Pérez-Martín, J., del Solar, G., Lurz, R., de la Campa, A.G., Dobrinski, B. and Espinosa, M. (1989) J. Biol. Chem. 264, 21334–21339.
[11] Lee, D.H. and Schleif, R.F. (1989) Proc. Natl. Acad. Sci. USA 86, 476–480.
[12] Ebright, R.H. (1993) Mol. Microbiol. 8, 797–802.
[13] Busby, S. and Ebright, R.H. (1997) Mol. Microbiol. 23, 853–859.
[14] Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Methods Enzymol. 185, 60–89.
[15] Ballester, S., Alonso, J.C., López, P. and Espinosa, M. (1990) Gene 86, 71–79.
[16] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[17] Leslie, A.G.W. (1991) in: D. Moras, A.D. Podjanry and J.C. Thierry (Eds.), Crystallographic Computing V, Oxford University Press, Oxford, pp. 27–38.
[18] CCP4 (1994) Acta Crystallogr. Sect. D, 50, 760–763.
[19] Carter, C.W.J. and Carter, C.W. (1979) J. Biol. Chem. 254, 12219–12223.
[20] Jancarik, J. and Kim, S.-H. (1991) J. Appl. Crystallogr. 24, 409–411.
[21] Matthews, B.W. (1968) J. Mol. Biol. 33, 491–497.
[22] Navaza, J. (1994) Acta Crystallogr. Sect. A 50, 157–163.
[23] Lacks, S.A., López, P., Geenberg, B. and Espinosa, M. (1986) J. Mol. Biol. 192, 753–765.
[24] Priebe, S.D. and Lacks, S.A. (1989) J. Bacteriol. 171, 4778–4784.
[25] del Solar, G., Albericio, F., Eritja, R. and Espinosa, M. (1994) Proc. Natl. Acad. Sci. USA 91, 5178–5182.