The EMBO Journal Vol.17 No.24 pp.7404–7415, 1998

The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator

F.Xavier Gomis-Rüth, Maria Solà, Paloma Acebo[1], Antonio Párraga, Alicia Guasch, Ramón Eritja[2], Ana González[3], Manuel Espinosa[1], Gloria del Solar[1] and Miquel Coll[4]

Institut de Biologia Molecular de Barcelona, CSIC, Jordi Girona, 18–26, 08034 Barcelona, [1]Centro de Investigaciones Biológicas, CSIC, Velázquez, 144, 28006 Madrid, Spain, [2]EMBL, Meyerhofstraße 1, 69117 Heidelberg and [3]EMBL, c/o DESY, Notkestraße 85, 22603 Hamburg, Germany

[4]Corresponding author
e-mail: mcccri@cid.csic.es

The structure of the 45 amino acid transcriptional repressor, CopG, has been solved unliganded and bound to its target operator DNA. The protein, encoded by the promiscuous streptococcal plasmid pMV158, is involved in the control of plasmid copy number. The structure of this protein repressor, which is the shortest reported to date and the first isolated from a plasmid, has a homodimeric ribbon–helix–helix arrangement. It is the prototype for a family of homologous plasmid repressors. CopG cooperatively associates, completely protecting several turns on one face of the double helix in both directions from a 13-bp pseudosymmetric primary DNA recognition element. In the complex structure, one protein tetramer binds at one face of a 19-bp oligonucleotide, containing the pseudosymmetric element, with two β-ribbons inserted into the major groove. The DNA is bent 60° by compression of both major and minor grooves. The protein dimer displays topological similarity to Arc and MetJ repressors. Nevertheless, the functional tetramer has a unique structure with the two vicinal recognition ribbon elements at a short distance, thus inducing strong DNA bend. Further structural resemblance is found with helix–turn–helix regions of unrelated DNA-binding proteins. In contrast to these, however, the bihelical region of CopG has a role in oligomerization instead of DNA recognition. This observation unveils an evolutionary link between ribbon–helix–helix and helix–turn–helix proteins.

Keywords: CopG/plasmid/protein–DNA complex/transcriptional repressor/X-ray crystal structure

Introduction

Bacterial plasmid replication is highly regulated, and stable maintenance in the host depends on constant plasmid copy number (Nordström, 1983; Novick, 1987; Nordström and Austin, 1989). In multicopy plasmids replicating by the rolling circle mechanism, this event is controlled by the availability of the *rep* gene-encoded initiator of replication protein (Rep), which introduces a site- and strand-specific nick at the origin of replication, *dso*. In the promiscuous streptococcal 5536-bp pMV158 plasmid, the synthesis of the 24.5-kDa RepB initiator protein is under the control of the products of two genes, *rnaII* and *copG*. The former encodes a 50-nucleotide antisense RNA, complementary to a region of the *copG–repB* mRNA that includes the putative ribosome binding site, thus acting through translational attenuation. The *copG* gene codes for the transcriptional repressor CopG, previously referred to as RepA (del Solar and Espinosa, 1992; del Solar *et al*., 1995). This is the smallest naturally occurring transcriptional repressor biochemically described to date, and has been purified as a homodimer of identical 45-residue subunits. CopG represses its own synthesis and that of the RepB protein by binding to the single *copG–repB* promoter (P_{cr}) region, which contains a 13-bp imperfect inverted repeat (IR) sequence. It is the prototype for the Cop-family of homologous bacterial plasmid repressors (see Figure 1; del Solar *et al*., 1989).

An intrinsic DNA curvature has been observed around the CopG DNA target region, which includes several A-tracts, spaced 10-bp upstream of the IR. Upon addition of CopG, this intrinsic curvature is greatly enhanced (Pérez-Martín *et al*., 1989). Hydroxyl radical and DNase I footprint experiments at low CopG:DNA ratios show cooperative protection covering at least four successive DNA turns, two on each side of the pseudo twofold axis of the IR (del Solar *et al*., 1990). The protected sites are located on the same face of the double helix and include the promoter –35 site, the overlapping IR and the –10 region. Binding of CopG thus could hinder host RNA polymerase binding (del Solar *et al.*, 1989, 1990; Pérez-Martín *et al.*, 1989). At high protein:DNA ratios, the protected region can extend up to 100 bp (our unpublished results). These observations indicate formation of a multiprotein complex nucleating from the 13-bp IR and extending over several helical turns of DNA. Formation of these multiprotein complexes is apparently favoured by DNA binding, since the free protein in solution remains dimeric up to 780 μM concentrations, as revealed by analytical ultracentrifugation analysis (data not shown).

In order to examine the detailed three-dimensional structure of CopG, we have crystallized it and solved its structure when non-liganded and when in complex with a 19-bp dsDNA fragment encompassing the 13-bp IR. Further implications for other members of the Cop family of plasmid repressors, similarities with ribbon–helix–helix motif (RHH) repressors of the MetJ/Arc superfamily, and resemblance with other unrelated helix–turn–helix motif (HTH) DNA-binding proteins of different organisms are discussed.

Results and discussion

Protomer structure

A sequence-based predicted bihelical region in CopG that complies with the requirements of a HTH-DNA binding

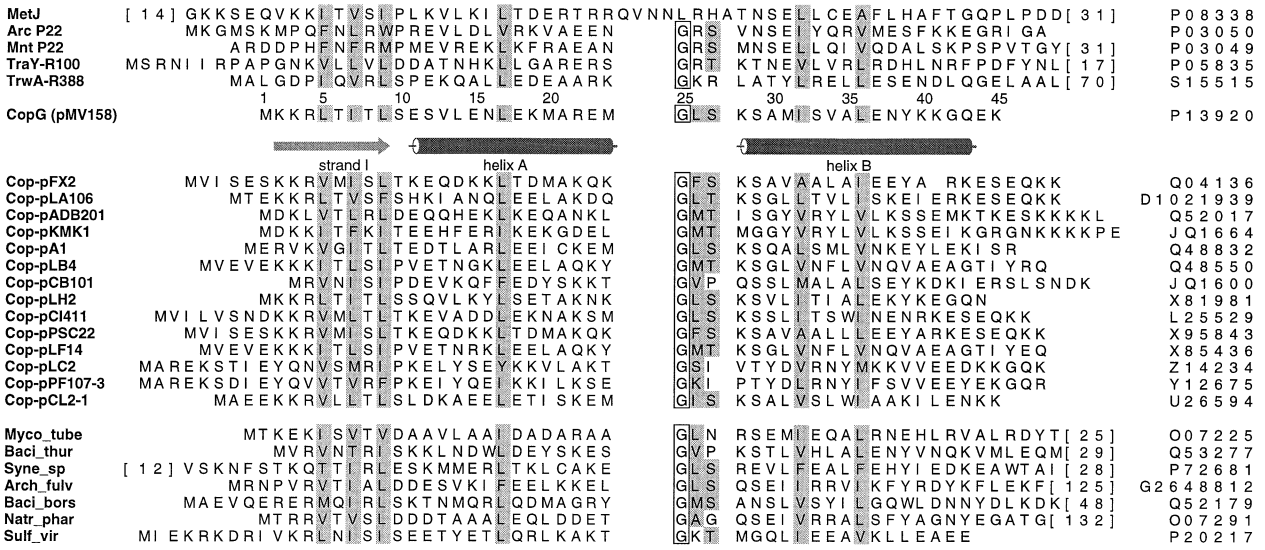

Fig. 1. Five members of the MetJ/Arc superfamily of proteins and 14 further integrants of the Cop family, comprising homologous plasmid gene products, are structurally aligned with the CopG sequence. The glycine residues of the turn connecting helix A and B of the RHH-motif is framed, conserved (hydrophobic) positions are highlighted by grey background, as are the highly conserved Thr/Ser residues in the mentioned turn. The numbering and regular secondary-structure regions correspond to the CopG structure. Seven additional aligning bacterial (from *Mycobacterium tuberculosis, Bacillus thuringiensis, Synechocystis sp., Archaeoglobus fulgidus, Bacillus borstelensis* and *Natronobacterium pharaonis*) and viral (from *Sulfolobus*) hypothetical gene products or ORFs are also displayed. The Swiss-Prot/PIR sequence data bank access code for each sequence is provided for reference. The numbers in parentheses indicate additional residues that are not shown in the alignment.

motif (see below; Ohlendorf *et al*., 1982; del Solar *et al.*, 1989), together with the size and dimeric structure of the protein, led to the hypothesis that CopG and related plasmid-encoded repressors belonged to the family of Cro proteins of bacteriophages λ and 434. Circular dichroism measurements of CopG indicated a consensus average content of >50% α-helix, and 10–35% β-strands and turns, compatible with this prediction. This hypothesis has now been proven to be only partially correct.

Three units of 45-residue CopG, named A, B and C, are present in the unliganded protein crystal asymmetric unit. The structures of the three molecules are very similar, as denoted by r.m.s. deviation values of 0.34 Å (for 40 common C_α atoms, deviating <1 Å, of molecules A and B), 0.33 Å (for 38 common C_α atoms of molecules A and C), and 0.42 Å (for 41 common C_α atoms of molecules B and C), although some side-chain conformations differ (see below). The polypeptide chains begin with a segment in extended conformation (strand I, Figures 1 and 2A), reaching Leu9. At Glu11, the first helix A starts with its axis rotated ~70° away from the direction of strand I. This helix finishes with Met24, then enters a glycine-mediated turn leading to helix B, which begins with Lys28 and extends to the C-terminal end. The axis of helix B is rotated ~90° with respect to helix A, owing to the presence of Gly25. The bihelical topology is the same as that found in proteins with an HTH-motif structure involved in DNA major-groove recognition (Matthews *et al*., 1982; Pabo and Sauer, 1984; Harrison and Aggarwal, 1990), although in CopG this element has a different function (see below). The HTH-like motif is held by an intramolecular interior hydrophobic cluster made up by the side-chains of Leu17, Met20 and Met24 (from helix A), and Leu26, Met31, Ile32 and Val34 (from the posterior turn and helix B). Furthermore, up to five salt bridges involving residues from helices A and B, and located on the exterior of the protomer structure may stabilize the arrangement of these two helices (Glu18 $O_{\varepsilon1}$–Arg22 N_ε, Glu18 $O_{\varepsilon2}$–Lys28 N_ζ, Glu18 $O_{\varepsilon1}$–Arg22 $N_{\eta1}$, Glu23 $O_{\varepsilon1}$–Lys19 N_ζ, Glu37 $O_{\varepsilon2}$–Lys40 N_ζ). Interestingly, these residues do not form bonds in molecule A, although they do so in both molecules B and C (see Figure 3A). Accordingly, the establishment of these ionic interactions does not appear to be indispensable for structure maintenance. The only significant additional intramolecular stabilization may occur due to van der Waals interactions between the side-chains of Ile7 and Leu9 (from strand I), and Leu17 (helix A) and Ile32 (helix B), respectively, at distances of ~3.9 Å.

Dimer and oligomer structure

CopG has been described as a dimer. In our unliganded protein structure, two CopG units (molecules A and B) associate in a symmetrical manner via a local twofold axis (Figure 2A) with intimate contact, as signified by a common surface of 1615 Å^2 and a number of intermolecular contacts below 4 Å (20 hydrogen bonds, 56 van der Waals contacts and one salt bridge). This association leads to the formation of a two-stranded twisted antiparallel sheet or β-ribbon, based on 10 inter main-chain hydrogen bonds, comprising the N-terminal segment from Met1 to Glu11 of each molecule. This renders, together with the previously described bihelical element, an RHH-structure for each protein chain. This β-ribbon is presented to the exterior from a hydrophobic and dimeric protein core, mainly formed by internal residues coming from both strands and the four helices. This hydrophobic cluster, enclosing the intramolecular one described previously, reaches from the base of the sheet to the opposite surface of the dimer, and is made up by the side-chains of Leu5, Ile7 and Leu9 (from each strand I), Val13, Leu14, Leu17, Met20, Ala21 and Met24 (helices A), and Leu26, Met31, Ile32, Val34, Ala35, Leu36 and Tyr39 (coming from the turn connecting helices A and B, and the latter helix from each polypeptide chain). Beyond Gly42, the protein main-

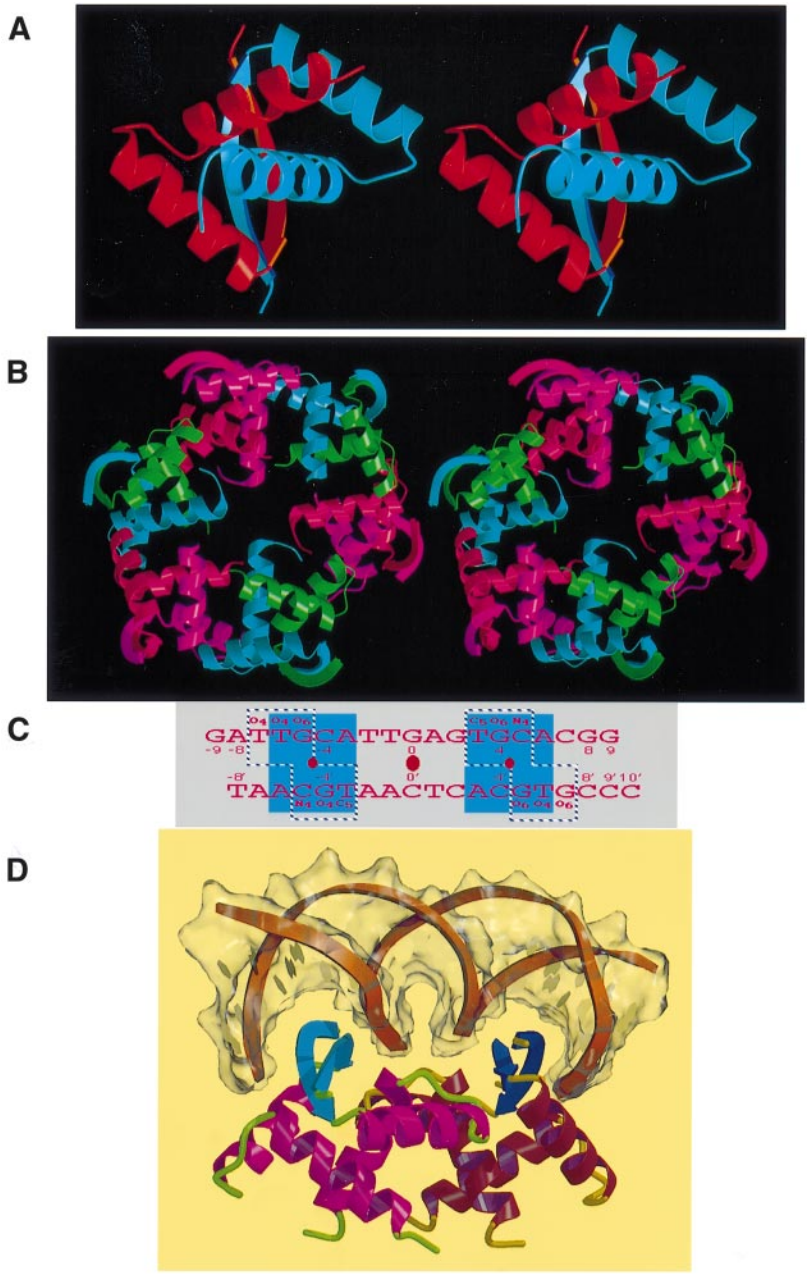

Fig. 2. (**A**) Stereo figure showing two CopG monomers forming a dimer, which contains a β-ribbon (red and blue arrows), the recognition element of the DNA major groove. (**B**) Stereo cartoon of the CopG dodecamer superhelical oligomeric structure in the unliganded crystal. (**C**) Sequence of the 19-bp dsDNA used for co-crystallization experiments. The twofold axis of the operator is denoted by a large ellipsoid at the centre of the sequence. Two symmetrical sequences are indicated by blue boxes left and right of the centre. They have internal twofold symmetry, as denoted by the small ellipsoids. The bases recognized by CopG are indicated by dashed boxes. Base atoms participating in DNA–protein interactions are indicated. (**D**) The structure of two CopG dimers complexed with the 19-bp dsDNA containing the IR, superimposed with its semi-transparent Connolly surface.

chains project away from the dimer body, with no further intra- or inter-contacts, explaining the flexibility of the C-terminal segment of CopG, as characterized by above-average temperature factors in this region and partial disorder (see Materials and methods). Helix A of each unit displays charged side-chains pointing to the exterior of the dimer, making no crystal contacts. The helices B cross each other around the position of Ala35 at ~90° and establish a series of dimer-maintaining contacts, in particular a double hydrogen bond made up by Asn38 of each molecule, and a further one established between Asn16A $N_{\delta 2}$ and Glu37B $O_{\varepsilon 1}$. Therefore, a mutation of Ala35 (the distance Ala35A C_β to Ala35B C_β is 4.3 Å) to a residue with a bulky side-chain, or the disruption of the latter hydrogen bonds, would weaken or even prevent the dimeric association.

A third CopG molecule (termed C) is present in the asymmetric unit that, via a crystallographic dyad, forms a further dimer of the type molecule C–molecule C. Therefore, dimers form either with a local (molecule A–molecule B; see above) or crystallographic axis. The dimers AB and CC interact with a contact surface of 528 $Å^2$ and 56 contacts below 4 Å, including 14 van der Waals contacts and six hydrogen bonds, mainly through

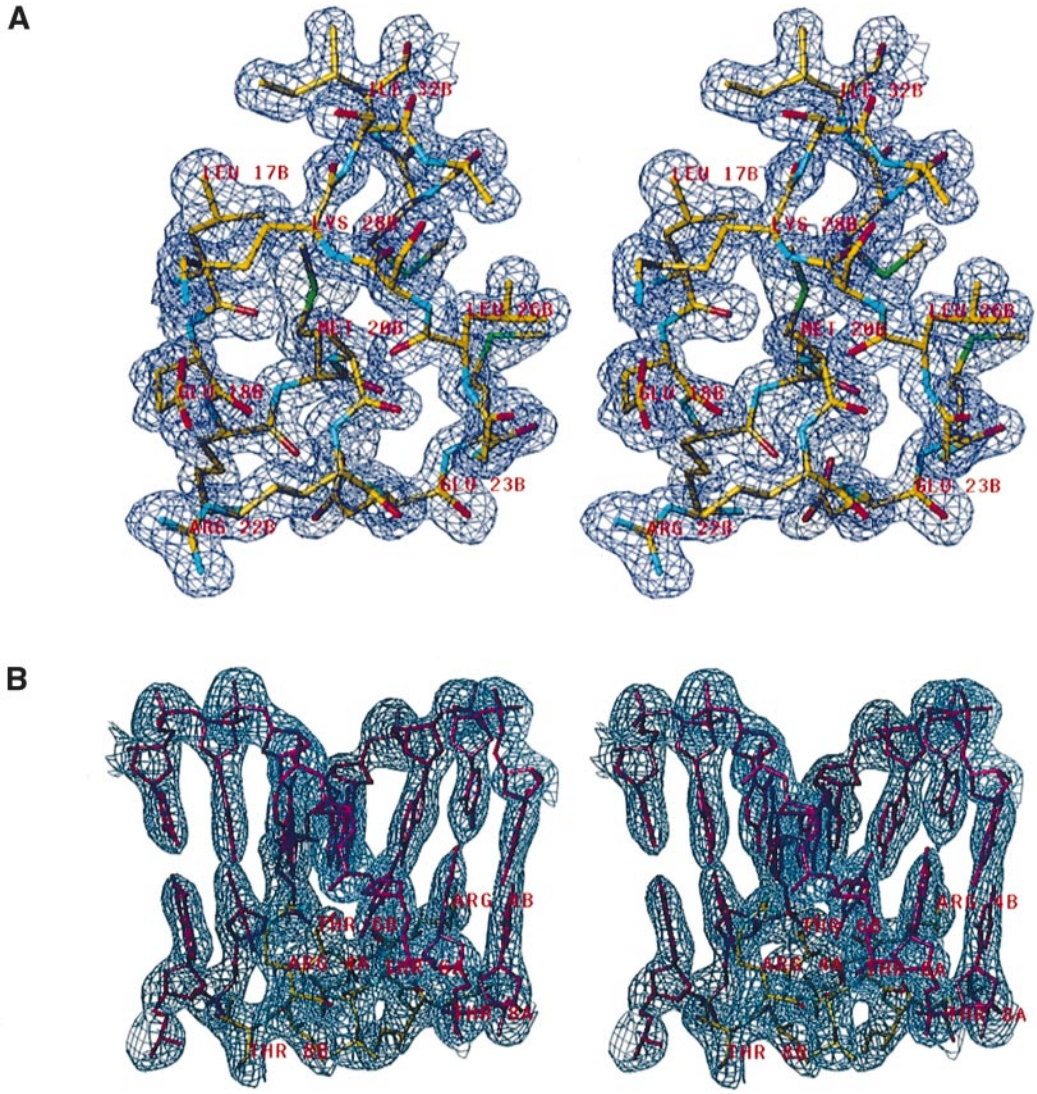

Fig. 3. (**A**) Stereo plot displaying detail of the unliganded CopG structure, superimposed over the final ($2F_{obs}$–F_{calc})-electron density map, contoured at 1.25 σ around the turn connecting helices A and B of the HTH-motif of molecule B. Selected residues are labelled. (**B**) Stereo image of the detail of the final CopG-DNA model, superimposed with the calculated ($2F_{obs}$–F_{calc})-electron density map, contoured at 1 σ, and highlighting the interaction between the protein dimer β-ribbon (atom colour code sticks) and the centre of the DNA major groove (magenta sticks). Selected residues are labelled.

the turn and helix B of molecule B with equivalent parts of a molecule C, with both helices running almost parallel. In addition, the AB dimer contacts a crystallographically related vicinal AB dimer with a common surface of 526 Å^2 and 49 contacts below 4 Å. The contacts observed at this interface are similar, but not identical to those observed for the AB–CC dimer–dimer interface. Mainly hydrophobic interactions between the side-chains of Met24, Val13, Leu26, Val34 and Ala30, and two hydrogen bonds are established. These interactions generate in the crystal a right-handed helical superstructure encompassing 12 CopG molecules per turn, with the DNA-binding regions pointing to outside (Figure 2B).

CopG in complex with DNA

The structure of CopG in complex with a 19-bp dsDNA containing the 13-bp pseudosymmetric element (see Figures 2C and D, and 3B) reveals a tetramer composed of two dimers, related by a crystallographic dyad, interacting in the same way as two AB dimers in the unliganded structure (see above). No significant differences are observed in the main-chains when compared with the free protein structure, as denoted by an r.m.s. deviation of 0.56 Å for all C_α atoms of the tetramer. Each of the two polypeptide chains of the dimer is defined from residue 1 to 43. CopG interacts with DNA bases via the N-terminal β-ribbon, and with backbone phosphate groups via residues at the N-terminus of helix B (see Figures 2D and 4A,B), inducing a bend of 60° in the nucleic acid moiety. The bend is produced by marked compression of both minor and major grooves facing the protein. The minor groove located at the center of the operator becomes extremely narrow (1.9 Å). The major grooves are also tighter (8.6 Å) than regular B-DNA (11.7 Å) in the area of interaction with the β-ribbon (Figure 4C). The DNA backbone follows a smooth path in general, except for the zone close to the centre of the operator, where the minor groove is compressed and the base pairs are rather inclined. In

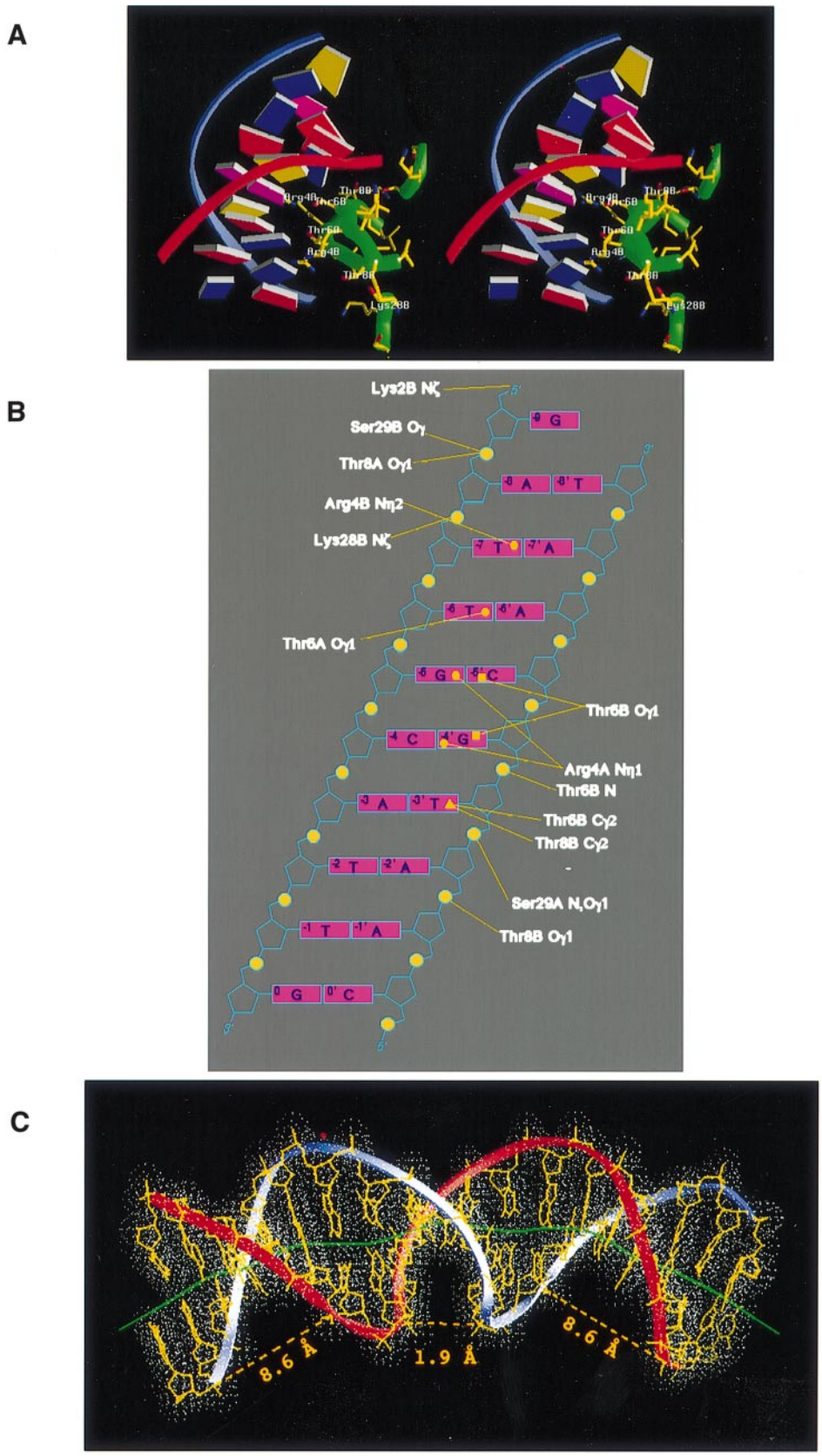

Fig. 4. (**A**) Schematic detail of the interactions between one CopG dimer and one half of the target operator in stereo. The protein segments are depicted in green. Colour codes for the DNA bases are: magenta, cytosines; yellow, guanines; blue, thymines; red, adenines. (**B**) Scheme of the interactions between CopG protein atoms [(A) and (B) characterize the two protein chains of a dimer] and the left half of the pseudosymmetric operator. Squares display interactions with $N_4/N_6/N_7$ atoms of cytosine or adenine bases, full spheres with O_4/O_6 atoms of thymine or guanine, and triangles with C_5 atoms of thymines. (**C**) The 60° bent structure of the 19-bp operator DNA when complexed with the CopG tetramer, superimposed over its van der Waals dot surface. The DNA local helix axis is depicted in green. Dashed lines unite the phosphorous atoms between which the grooves' openings have been measured. The associated distances (after subtraction of the van der Waals radii) are also shown.

addition to the interdimer crystallographic contacts that define the functional tetramer, other crystal symmetry contacts are made by the DNA part provided with G and C overhangs (see Figure 2C), rendering a right-handed continuous dsDNA with an extended left-handed superhelical twist due to interactions with the protein. The arbitrarily chosen length of the dsDNA used for crystallization places the centre of the IR at different faces of the DNA helix, thus preventing any additional protein–protein (i.e. intertetramer) contact similar to those seen in the unliganded crystal structure. However, such interactions may occur in the case of longer oligonucleotides (i.e. 22 bp) or when dealing with plasmid DNA containing the entire CopG target.

Base recognition contacts

Figure 2C shows the pseudosymmetric operator sequence included in the 19-bp oligonucleotide encompassing the 13-bp IR in our structure. The blue boxes indicate the true symmetric elements at each half of the operator. These two elements have internal symmetry themselves and overlap the base recognition sites of each dimer of CopG (dashed boxes), although they are not identical. Although each CopG dimer has local twofold symmetry, the two β-strands contact different bases, not matching the twofold symmetry of the DNA sequence. As inferred from Figures 2C, and 4A and B, the following contacts determine the recognition on the left side: Thy-7 O_4 (nomenclature according to Figure 2C) is H-bonded to Arg4B $N_{\eta 2}$, Thy-6 O_4 to Thr6A $O_{\gamma 1}$, Gua-5 O_6 to Arg4A $N_{\eta 1}$, Cyt-5 N_4 to Thr6B $O_{\gamma 1}$, Gua-4 O_6 to Arg4A $N_{\eta 1}$, and the C_5 methyl group of Thy-3′ is sandwiched between the $C_{\gamma 2}$ methyl groups of Thr6B and Thr8B. The asymmetrical recognition contacts of the two β-strands of a CopG dimer is exemplified by the different interactions of the threonine residues at position 6 of each strand. While Thr6A $O_{\gamma 1}$ interacts via a hydrogen bond with acceptor Thy-6 O_4, Thr6B $O_{\gamma 1}$ does so with a donor, namely Cyt-5′ N_4. This is possible because these two protein side-chains establish an H-bond with each other, Thr6B $O_{\gamma 1}$ being the donor. Accordingly, this group can further act as an acceptor for the cytosine amino group, while Thr6A $O_{\gamma 1}$ remains a donor. An asymmetric recognition by the β-ribbon was also observed in the Arc/operator complex (Raumann *et al*., 1994). Although the recognition within each half of the CopG operator is asymmetrical, there is symmetry in the interaction between the left and right halves. Indeed, there is a crystallographic dyad between the two dimers of the functional tetramer and through the centre of the operator (base pair Gua0–Cyt0′). The DNA sequence is not palindromic, and so it has dual occupancy with two equally populated orientations around the dyad (see Materials and methods). However, the chemical groups recognized at both halves are identical. As shown by the dashed recognition boxes of Figure 2C, the diverging base in both halves is that at position 7: a thymine on the left side (Thy-7) and a guanine on the right side (Gua-7′). Nonetheless, the two chemical groups involved in interactions are acceptors, namely O_4 and O_6, both interacting with the guanidinium group Arg4B $N_{\eta 2}$.

Backbone interactions

Further contacts occur between the protein part and the DNA backbone (see Figure 4B). These are not directly dependent on the DNA sequence, and comprise H-bonds between the side-chains of Thr8, Lys28, and Ser29 of each CopG chain and DNA phosphate groups. In addition, the phosphates at positions 3 and –3′ establish contacts with the protein main-chain amide nitrogens of Lys28 and Ser29 of molecule A. These residues are located at the first turn of CopG helix B. Thus, the DNA phosphate group N-caps the helix and is located on the axis of the helix dipole. This kind of interaction has also been observed in Arc and MetJ (Somers and Phillips, 1992; Raumann *et al*., 1994) and in other, unrelated DNA-binding molecules such as CAP (Schultz *et al*., 1991). In CopG molecule B, helix B also points its N-terminus towards a DNA phosphate group, but in this case there are no direct H-bonds. In the unliganded structure, chloride ions occupy the location of the phosphate groups N-capping helix B, confirming the electropositive potential of these points.

Implications for the members of the Cop family of plasmid repressors

As mentioned above, CopG is the prototype for a whole family of Cop repressors of plasmid origin. A sequence alignment of pMV158-CopG with 14 members of this family (Figure 1) displays features that are compatible with the structure described here. Varying lengths of the members are accounted for by longer N- and C-terminal segments. All sequences display the glycine-mediated turn connecting two α-helices, and similar residues at key positions, i.e. those required for the maintenance of the hydrophobic core structure of the dimer. Thus, all members of the family are likely to share the RHH structure, the functional unit being a dimer of dimers. Interestingly, several members display up to eight additional residues at the N-terminus, reminiscent of the Arc structure (see below), where five residues, disordered in the unliganded structure, contribute to stabilization of the protein-DNA complex (Raumann *et al*., 1994). The CopG-protein was found with helix B ending at Gly42-Lys45 (depending on the molecule). This region could be easily elongated for up to eight residues (case of Cop-pADB201 or Cop-pKMK1) without disturbing the oligomer structure.

Furthermore, some unrelated hypothetical gene products or ORFs of varying length of bacterial and viral origin also display significant sequence similarity, putatively sharing the domain architecture with the Cop family and Arc/Met superfamily (see Figure 1).

Comparison with other members of the MetJ/Arc superfamily of repressors

The structures of three components of the MetJ/Arc superfamily have been published non-liganded and/or in complexes with their target DNA, namely both 53-residue Arc and 82-residue Mnt repressors of *Salmonella typhimurium* bacteriophage P22, and *Escherichia coli* 104-residue methionine repressor protein MetJ (Rafferty *et al*., 1989; Breg *et al*., 1990; Somers and Phillips, 1992; Burgering *et al*., 1994; Raumann *et al*., 1994). This superfamily further comprises the TraY and TrwA proteins of conjugative plasmids, as shown by sequence homology (Breg *et al*., 1990; Inamoto and Ohtsubo, 1990; Nelson and Matson, 1996; Moncalián *et al*., 1997). The common motif of this superfamily is a dimerizing 40–45 residue

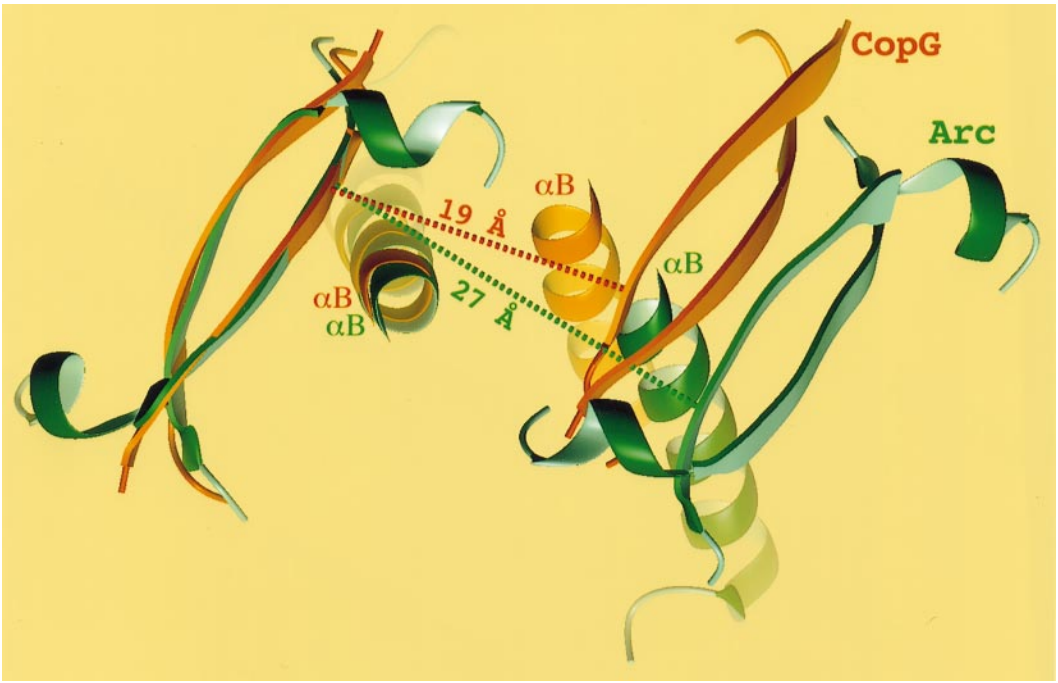

Fig. 5. Structural comparison of the consecutive recognition β-ribbons of Arc (green) and CopG (brown) tetramers in their complexes with DNA after optimal superimposition of one protein dimer (left). Helices B from both dimers that strongly interact in CopG, but not in Arc, are also depicted. Arc displays an additional N-terminal 3_{10}-helix that interacts with the DNA major groove.

RHH structure forming a two stranded β-sheet that recognizes the DNA major groove, with a glycine-mediated turn connecting both helices of each chain (Somers and Phillips, 1992; Raumann *et al.*, 1994) (see Figure 1). The exception is MetJ, which does not possess this turn (Figure 1), but unites helices A and B by a longer connecting loop. Mnt and MetJ display further structural elements mainly associated with co-repressor binding or oligomerization. The N-terminal regions mediate much of the DNA sequence specificity, as noted by the substitution of the first six Mnt residues with residues 1–9 of Arc, producing a mutant that recognises the Arc operator (Knight and Sauer, 1989). The six N-terminal Arc residues, disordered in the absence of DNA, form tandem reverse turns in the complex that are stabilized by interactions with DNA, with no major differences between the unliganded and the liganded structures (Raumann *et al.*, 1994).

Despite being eight residues shorter, CopG dimer most resembles the Arc structure with no insertion or deletion within the aligned segment (see Figure 1), as indicated by an r.m.s. deviation of 1.28 Å for 78 topologically equivalent C_α-atoms (from a dimer). This similarity is higher than with Mnt, (r.m.s. deviation 1.80 Å for 78 equivalent C_α-atoms). With MetJ the r.m.s. deviation is only 1.02 Å for 78 equivalent C_α-atoms, but the overall structure differs significantly from CopG due to a number of insertions (see Figure 1).

Nevertheless, the high structural similarity in the RHH-motifs of Arc, Mnt, MetJ and CopG repressors, there are major differences between these proteins in the regions associated with cooperative interdimer contact. MetJ dimers bind to DNA tandem repeats or Met boxes, and the entire length of helix A is involved in cooperative interdimer interactions (Phillips *et al*., 1989; Wild *et al*., 1996). The dimers wrap around the DNA helix. Between two and five such Met boxes have been found in several operators. In contrast, no association states higher than a tetramer have been reported for Mnt or Arc when bound to DNA. The functional unit of these repressors is a tetramer that recognizes a symmetric operator. Arc dimers interact using the loop between helices A and B to form the tetramer that binds to one face of the double helix. CopG also establishes tetramers that bind to one face of the DNA, but, in contrast to Arc, CopG dimers strongly interpenetrate as they form the DNA-bound active tetramer. Multiple contacts occur between residues constituting the loops connecting the three regular secondary structure elements of each chain and helix A, and the latter helix itself of the other dimer, and between almost the entire helices B from both dimers (Figure 2D). The tight interaction is formed by a small hydrophobic core encompassing 10 apolar residues from both dimers forms, together with two hydrogen bonds. The looser interdimer contacts in both Arc and MetJ have the inability to form tetramers in the unliganded crystal structures as an effect, as CopG does (Rafferty *et al.*, 1989; Schildbach *et al*., 1995). Furthermore, the surface occluded by the two dimers upon tetramerization is 292 $Å^2$ in Arc, whereas it is 526 $Å^2$ in CopG. The latter results in a different arrangement of the two consecutive major groove-recognising β-ribbons. A much shorter distance separates these elements in CopG (19 Å) than in Arc (27 Å) (see Figure 5). The direct effect of this shorter spacing is a much larger DNA bending in the CopG/operator complex (60°; calculated as described in Materials and methods) than in the Arc/operator complex (22°). Although in the latter a compression of the minor groove is observed at the center of the operator, in contrast to CopG, it opens the major groove at both left and right recognition sites.

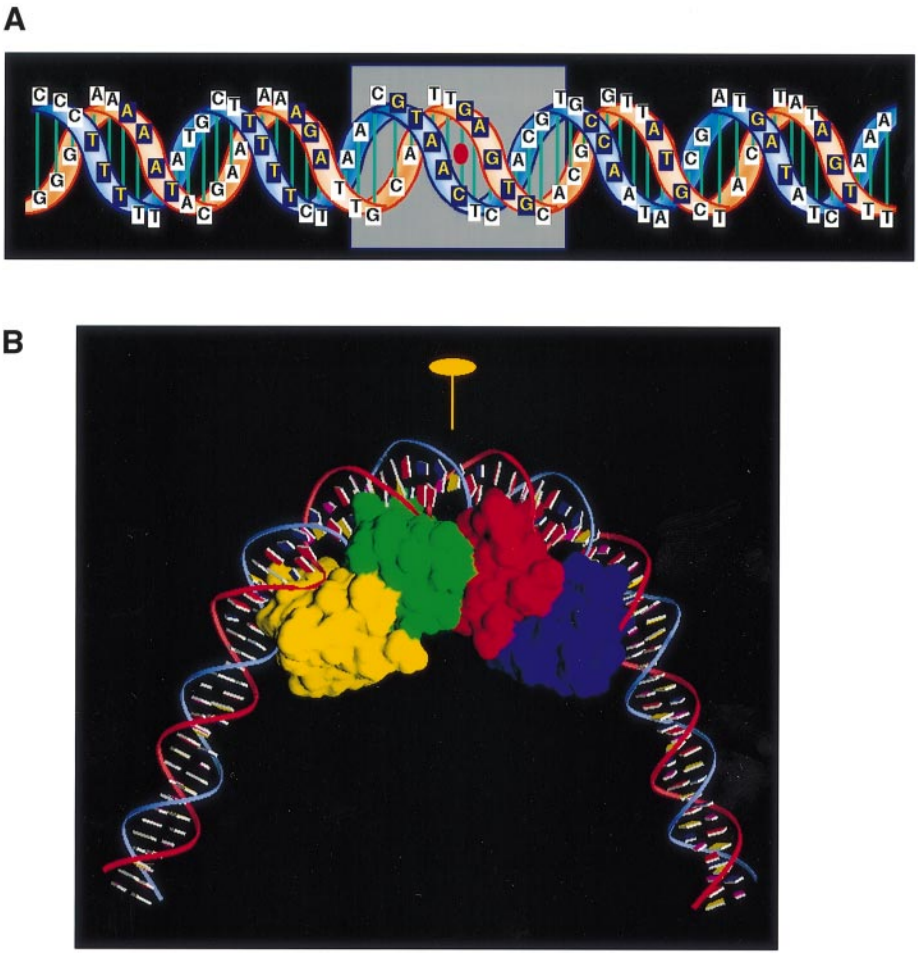

Fig. 6. (**A**) Schematic diagram of the protected DNA bases at the *copG–repB* promoter (yellow letters on dark blue background), as determined from hydroxyl radical footprinting. The 13-bp symmetric element is boxed (grey shading), with the (pseudo) dyad indicated by a red ellipse. The protected bases due to cooperative CopG-binding, covering four DNA turns, are all at one face of the helix (adapted from del Solar *et al.*, 1990).(**B**) Four CopG-dimers (characterized by coloured, solid Connolly surfaces) covering four double-helix turns bend the DNA by ~120°. The view is perpendicular to the operator axis.

CopG lacks the N-terminal 3_{10}-helix (or tandem reverse turns, see Figure 5) which in Arc and MetJ, together with the DNA recognising β-ribbon, binds the DNA major groove and prevents its compression. These comparisons suggest that proteins of this superfamily recognize different DNA sequences by changing both the chemical identity of β-sheet residues and the way the tetramer is formed.

CopG displays structural homology with otherwise unrelated DNA-binding proteins

Topology analysis against proteins of known three-dimensional structure, as perfomed with DALI (Holm and Sander, 1995), reveals not only the structural similarity with the other three members of the MetJ/Arc superfamily, but also with the DNA-recognizing HTH-motif of some unrelated proteins of diverse origin. Between 28 and 31 residues are similar (of 45 residues of a protomer) in CopG and in *E.coli* response regulator NarL (Baikalov *et al*., 1996; Protein Data Bank accession No. 1rnl), MATa1 of the MATa1/MATa2 homeodomain heterodimer from *Saccharomyces cerevisiae* (Li *et al*., 1995; PDB access code 1yrn), fruitfly paired (Gln-50) class cooperative homeodomain (Wilson *et al*., 1995; PDB access code 1fjl) and human Oct-1 POU domain (Klemm *et al*., 1994; PDB access code 1oct). Interestingly, no other proteins with two consecutive helices (e.g. globins), but only DNA-binding proteins have been found in the similarity searches.

The HTH-motif is formed by a pair of α-helices linked by a tight turn in otherwise distinctly different structures. It is mainly defined by a central 20-residue segment corresponding to particular residues in the first DNA-binding protein structures solved containing this motif, phage λ and 434 repressor proteins (Pabo and Lewis, 1982; Aggarwal *et al*., 1988), Cro proteins (Anderson *et al*., 1981; Wolberger *et al*., 1988), and the *E.coli* cAMP-binding protein CAP (McKay and Steitz, 1981). Both the first and second helices, the latter being the DNA recognition helix, may be extended at their N- and C-termini. The properties of the HTH-element have been extensively described and reviewed (Pabo and Sauer, 1984; Harrison and Aggarwal, 1990).

In the CopG sequence and structure, all requirements for a fully functional HTH-motif are fulfilled (Figure 1). The 20-residue motif extends from Leu17 (position 1) to Leu36 (position 20). Hydrophobic relative positions 4, 8, 10, 15, 16, 18 are Met20, Met24, Leu26, Met31, Ile32, and Val34, respectively, and restricted side-chain positions 5 and 9 are Ala21 and Gly25, respectively; the latter with

main-chain conformational angles located in the region defined for left-handed α-helices (Φ = 65–66°, Ψ = 28–34°). Hydrogen bond donor/acceptor positions 11–13 and 17 are occupied by Ser27, Lys28, Ser29 and Ser33. Finally, the solvent-exposed, hydrophilic positions 1 to 3, 6, 7 and 11 are Leu17, Glu18, Lys19, Arg22, Glu23 and Ser27, respectively. Thus, the only position deviating somewhat from the prescriptions is the relative position 1, Leu17 in CopG, which is hydrophobic in all sequences discussed for members of the CopG/MetJ/Arc-repressor superfamily (see Figure 1). This residue establishes a van der Waals interaction with the side-chain of residue 9 (CopG numbering, Figure 1) of the N-terminal ribbon (see above), a conserved hydrophobic position. In addition, the C-terminal half of helix B is less hydrophobic than in most of the recognition helices, consistent with the role of protein–protein interactions of this helix in CopG.

Therefore, CopG displays a fully defined HTH-motif structure that is involved not in DNA-binding, but in the maintenance of the intrinsic dimeric functional structure and cooperativity.

Conclusions

The structures of CopG described here, free and in complex with its target DNA, show a dimeric RHH conformation, with the β-ribbon binding to the major groove of the DNA, so that its side-chains contact the bases. The two α-helices stabilize the dimer, together with the β-sheet forming a hydrophobic core. Helix B and the turn connecting helices A and B form the dimer–dimer interface in the tetramer. The active DNA-binding oligomerization state is a tetramer (dimer of dimers) contacting almost two turns of the double helix at one face, with the interaction surface between dimers identical to that present in the free protein crystal.

Electrophoretic mobility shift assays showed four discrete DNA-repressor complexes, indicating that at least four CopG dimers bind to the whole target region, which is consistent with preliminary stoichiometry determinations (data not shown). Based on these band-shift assays and data from DNase I and hydroxyl radical footprinting (see Figure 6A), indicating four dimers protecting four consecutive helical turns at one face of the DNA, formation of a multiprotein complex nucleating from the CopG 13-bp IR and extending along the DNA in both directions is suggested. Accordingly, and based on the complex structure, a model is proposed, in which the target plasmid DNA is bent by ~120°, consistent with electron microscopy and gel retardation experiments (Pérez-Martín *et al.*, 1989; our unpublished data). In this model, AB–AB types of interdimer interactions are assumed to generate the four protein dimers interacting with the target DNA (see Figure 6B), the primary binding locus for the CopG repressor being the 13-bp pseudosymmetric operator. Once the first two dimers are bound, two more dimers would bind cooperatively in both 3′- and 5′-directions, even though the sequences flanking the IR are neither homologous to the operator nor symmetrical. The dimer–dimer affinity and the intrinsically flexible promoter sequence could compensate for the lower binding affinity of these secondary sites. In this extended manner, the repressor would cover not only the P_{cr} promoter –35 RNA polymerase recognition sequence, but also the –10 box and, possibly, UP sites upstream of the operator. The result would be an enhanced repression of the P_{cr} promoter, which is in agreement with the strict control of the plasmid replication.

In conclusion, the presence of a bihelical region (HTH-motif) embedded in the RHH motif that does not recognize DNA, but is involved in oligomerization, in CopG and in other proteins containing a β-ribbon recognition substructure, suggests that they are evolutionarily related to those, in which the HTH-motif participates in DNA-binding. The former proteins usually associate in tetramers in order to recognize a (pseudo)symmetric sequence and are highly economic as they are shorter, and thus require less DNA to be encoded. In contrast, most of the latter act as dimers and usually display additional substructures playing a role in interprotein contacts.

Materials and methods

Protein purification and crystallization

CopG was produced by heterologous overexpression, purified and crystallized as described (Gomis-Rüth *et al.*, 1998). Unliganded protein crystals belong to orthorhombic space group $C222_1$ with cell constants a = 67.2 Å, b = 102.5 Å and c = 40.2 Å, and diffract beyond 1.6 Å resolution using conventional CuK_α radiation. Three molecules are present in the asymmetric unit.

The complex between CopG and a 19-bp dsDNA encompassing the 13-bp IR sequence (Figure 2C) was crystallized from drops equilibrated by the hanging drop vapour diffusion method against 60–78% 2-methyl-2,4-pentanediol, buffered with 0.1 M HEPES to pH 6.7–7.5, or with 0.1 M sodium acetate to pH 5.0. These crystals, diffracting beyond 2.6 Å resolution, belong to tetragonal spacegroup $P4_322$ with cell constants a = b = 40.3 Å, c = 221.6 Å, and one CopG-dimer–half-DNA-operator complex per asymmetric unit.

Structure solution and refinement

The free CopG structure was solved by the multiple isomorphous replacement method using anomalous diffraction (MIRAS). Data from native unliganded crystals and two heavy-ion derivatives obtained by co-crystallization, CH_3CH_2HgCl and $K_3UO_2F_5$, were measured at 100 K on the in-house 300-mm-MAR Research image plate detector attached to a Rigaku RU200 rotating anode generator, providing graphite-monochromatized CuK_α X-radiation. The data were processed with MOSFLM v. 5.41 (Leslie, 1991) and programs of the CCP4 suite (CCP4, 1994). Further data at the CH_3CH_2HgCl-mercury absorption edge (f''_{max} = 1.002 Å) were collected at the tuneable EMBL X31 beamline of the Deutsches Elektronensynchrotron (DESY) in Hamburg (Germany) using cryo-cooling. Initial phases were computed with MLPHARE (CCP4, 1994), yielding an overall figure of merit for the resolution range between 20 and 2.5 Å of 0.58. Table I provides a summary of the data collection, processing and phasing statistics. Subsequent density modification with DM (Cowtan and Main, 1996) increased the figure of merit to 0.88. An initial Fourier-map was inspected on a Silicon Graphics Workstation using Turbo-Frodo (Roussel and Cambilleau, 1989). At the first glance, the N-terminal strand–helix motif of two molecules became visible and was constructed as a polyalanine model. Successive cycles comprising least-squares reciprocal space refinement with X-PLOR version 3.851 (Brünger, 1991) and, in subsequent cycles, with SHELX-97 (Sheldrick, 1990), using the same set of flagged reflections for free *R*-factor monitoring and parameters (Engh and Huber, 1991), phase combination and manual model building, gradually permitted completion of the model. At the initial stages, non-crystallographic symmetry (NCS) restraints were applied for refinement and used for density averaging with DM and programs of the RAVE package (Jones, 1992). At the end, isotropic temperature factors were refined with SHELX-97. The final model comprises 130 of 135 residues of the chemical sequence corresponding to three monomers as present in the asymmetric unit (numbered A1-A43, B1–B45, and C1-C42, respectively), 194 solvent molecules (labelled W004 to W197) and three chloride anions (labelled Cl 1W to Cl 3W), tentatively assigned at the basis of strong positive difference density, reasonable temperature factors upon refinement (15.2, 15.5, and 26.3 $Å^2$, respectively), the absence of electron density for ion clusters containing oxygen atoms as phosphate

Table I. Data collection, processing and phasing statistics

	CopG (unliganded) Native	CopG/ CH_3CH_2-Hg-Cl	CopG/ CH_3CH_2-Hg-Cl	CopG/ $K_3UO_2F_5$	CopG/ 19-bp dsDNA
No. of crystals used	3	1	1	1	1
Space group	$C222_1$	$C222_1$	$C222_1$	$C222_1$	$P4_322$
Cell constants a, b, c (Å)	67.19, 102.49, 40.21	67.54, 102.33, 40.20	67.43, 102.43, 40.14	67.00, 102.88, 40.31	40.30, 40.30, 221.56
Resolution range (Å)	25.82–1.57	30.43–1.80	32.62–2.38	32.79–2.38	40.16–2.56
X-ray source	rot. anode	synchrotron	rot. anode	rot. anode	synchrotron
Radiation wavelength (λ in Å)	1.5418	1.002	1.5418	1.5418	0.9058
No. of measurements	91 693	38 505	17 722	18 639	78 954
No. of unique reflections	18 358	11 102	5720	5809	6252
Whole range					
Completeness (%)	98.1	98.9	98.4	98.6	96.3
R_{merge} (%)[a]	8.0	3.7	2.1	2.3	6.2
Intensity [I/σ(I)]	10.3	15.7	25.9	23.6	8.1
resolution range (Å)	19.61–1.60	19.59–1.80	19.61–2.38	19.69–2.38	19.86–2.56
Last shell					
Completeness (%)	96.4	87.0	98.9	95.4	86.8
R_{merge} (%)[a]	12.6	17.7	4.8	5.8	45.5
Intensity [I/σ(I)]	3.0	4.0	14.2	11.8	1.7
resolution range (Å)	1.69–1.60	1.85–1.80	2.51–2.38	2.50–2.38	2.70–2.56
Average data multiplicity	5.0	3.5	3.1	3.2	5.9
R_{Deriv}[b]	–	19.3	11.5	8.7	–
Number of heavy-ion sites	–	2	2	1	–
Anomalous refinement	–	yes	yes	yes	–
Phas.power (20–2.5 Å, acentr.)[c]	–	1.66	1.70	0.56	–
R_{Cullis} (20–2.5 Å, acentr.)[d]	–	0.68	0.67	0.95	–

[a] $R_{merge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - <I(hkl)>| / \Sigma_{hkl}\Sigma_i I_i(hkl)$
[b] $R_{Deriv} = \Sigma_{hkl} |F_P - F_{PH}| / \Sigma_{hkl} F_P$
[c] Phas. power = r.m.s. ($|F_H|$/ E), where E = the residual lack of closure.
[d] $R_{Cullis} = \Sigma_{hkl} ||F_{PH} \pm F_P| - F_{H(calc)}| / \Sigma_{hkl} |F_{PH} \pm F_H|$

Table II. Final refinement model statistics

	free CopG	CopG–DNA complex
Resolution range used for refinement (Å)	8.0–1.6	20.0–2.56
No. of reflections used	16 782	6024
Crystallographic R_{factor} (free R_{factor})[a]	0.202 (0.277)	0.247 (0.316)
No. of active non-hydrogen protein/DNA atoms (inactive)	1011 (6)	1439 (0)
R.m.s. deviation from target values for		
bonds (Å)	0.009	0.008
angular distance (Å)	0.025	–
angles (°)	2.203	1.453
R.m.s. deviation for bonded Bs for active atoms (Å^2)	3.2	2.3
Chloride anions	3	–
Solvent molecules	194	13

[a] $R_{factor} = \Sigma_{hkl} ||F_{obs}| - k |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$; free R_{factor}, same for a test set of 7% of reflections not used during refinement.

or sulfate, and the composition of the crystallization solution. All residues are placed in stereochemically favourable regions, as visualized by a Ramachandran plot, with just C-terminal residue C41 in a somewhat generously allowed region. The main-chains are defined by clear electron density. Just the side-chains of the three C-terminal residues of molecule B (B43-B45), the last residue of molecule A (A43), the last-but-one amino acid of molecule C (C41), the first two residues of molecule C (C1-C2), all of them part of flexible surface-located regions, and the guanidyl group of Arg22A are in somewhat poor density. The side-chains of all three N-terminal methionine residues from atom S_δ onwards have been set to occupancy zero. Table II provides a summary of the final model refinement statistics.

CopG/DNA complex data were also collected at 100 K at the EMBL X11 beamline of DESY, processed with MOSFLM v. 5.51 and programs of the CCP4 suite (see Table I). The structure was solved by molecular replacement calculations using AMoRe (Navaza, 1994), employing the refined free CopG dimer as a searching model. After determination of the correct position of the protein part, this was fixed during two-body searches further using a 3 bp-model. Different solutions matched the oligonucleotide with some base pair overlaps. The model was improved and completed in successive cycles of manual model building and crystallographic refinement, including bulk solvent correction and constrained temperature factor refinement employing X-PLOR, and applying NOE-like restraints for the DNA base pairs. Since a crystallographic twofold axis runs through the centre of the dsDNA moiety and the sequence is not perfectly palindromic (see Figure 2C), the DNA is present in dual occupancy, as observed in other protein/DNA complexes (Ferré-D'Amaré *et al.*, 1993; Becker *et al.*, 1998). This was accounted in the construction of the dsDNA in the two orientations and half occupancy for all atoms (sequence according to Figure 2C). The Fourier maps therefore represent the average electron density of the two DNA orientations. These maps are, however, of excellent quality (see Figure 3B), with all DNA bases and backbone atoms defined. The protein residues involved in contacts with DNA are also clearly defined with no indication of disorder nor deviation from the crystallographic symmetry. This is owing to the fact that the bases contacted by the protein conform to the dyad symmetry, with the exception of those at position –7 and –7′ (see Figure 2C); but even here the contacted groups of the bases are identical (see Results and discussion). The final model encompasses the two CopG protein chains (defined from residues 1 to 43, and labelled A1-A43 and B1-B43), two dsDNA chains with half occupancy and 13 solvent molecules, all defined by proper electron

density. Table II provides a summary of the final model refinement statistics.

Figures and solid surfaces have been calculated with MOLSCRIPT (Kraulis, 1991), MOLMOL (Koradi *et al*., 1996), VOLUMES/BOBSCRIPT (Esnouf, 1997), GRASP (Nicholls *et al*., 1993) and Turbo-Frodo. Accessibility surface calculations have been computed with X-PLOR using a probe radius of 1.6 Å. DNA curvature and axis were determined with CURVES (Lavery and Sklenar, 1988). Model quality control checks were performed with WHAT IF (Vriend, 1990). The final coordinates have been deposited with the Protein Data Bank in Brookhaven (accession Nos 2cpg and 1b01).

Acknowledgements

We thank Rosa Pérez-Luque for help in crystallization work, and Maite Alda in the overproduction and purification of CopG. This work was supported by grants PB95-0224 from the Ministerio de Educación y Ciencia, Spain, BIO97-0347 from CICYT, Spain, and by the Generalitat de Catalunya (Centre de Referència en Biotecnologia and grant 1997SGR-275). Data collection at the EMBL X31 beamline at the DORIS storage ring, DESY, Hamburg, was supported by the European Union Large Installation Project CHGE-CT93-0040.

References

Aggarwal,A.K., Rodgers,D.W., Drottar,M., Ptashne,M. and Harrison,S.C. (1988) Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science*, **242**, 899–907.

Anderson,W.F., Takeda,Y., Ohlendorf,D.H. and Matthews,B.W. (1981) Structure of the *cro* repressor from bacteriophage λ and its interaction with DNA. *Nature*, **290**, 754–758.

Baikalov,I., Schröder,I., Kaczor-Grzeskowiak,M., Grzeskowiak,K., Gunsalus,R.P. and Dickerson,R.E. (1996) Structure of the *Escherichia coli* response regulator NarL. *Biochemistry*, **35**, 11053–11061.

Becker,S., Groner,B. and Müller,C.W. (1998) Three-dimensional structure of the Stat3β homodimer bound to DNA. *Nature*, **394**, 145–151.

Breg,J.N., van Opheusden,J.H.J., Burgering,M.J.M., Boelens,R. and Kaptein,R. (1990) Structure of Arc repressor in solution: evidence for a family of β-sheet DNA-binding proteins. *Nature*, **346**, 586–589.

Brünger,A.T. (1991) Crystallographic phasing and refinement of macromolecules. *Curr. Opin. Struct. Biol.*, **1**, 1016–1022.

Burgering,M.J.M., Boelens,R., Gilbert,D.E., Breg,J.N., Knight,K.L., Sauer,R.T. and Kaptein,R. (1994) Solution structure of dimeric Mnt repressor (1–76). *Biochemistry*, **33**, 15036–15045.

CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.

Cowtan,K.D. and Main,P. (1996) Phase combination and cross validation in iterated density-modification calculations. *Acta Crystallogr.*, **D52**, 43–48.

del Solar,G. and Espinosa,M. (1992) The copy number of plasmid pLS1 is regulated by two *trans*-acting plasmid products: the antisense RNA II and the repressor protein, RepA. *Mol. Microbiol.*, **6**, 83–94.

del Solar,G., de la Campa,A.G., Pérez-Martín,J., Choli,T. and Espinosa,M. (1989) Purification and characterization of RepA, a protein involved in the copy number control of plasmid pLS1. *Nucleic Acids Res.*, **17**, 2405–2420.

del Solar,G.H., Pérez-Martín,J. and Espinosa,M. (1990) Plasmid pLS1-encoded RepA protein regulates transcription from repAB promoter by binding to a DNA sequence containing a 13-base pair symmetric element. *J. Biol. Chem.*, **265**, 12569–12575.

del Solar,G., Acebo,P. and Espinosa,M. (1995) Replication control of plasmid pLS1: efficient regulation of plasmid copy number is exerted by the combined action of two plasmid components, CopG and RNA II. *Mol. Microbiol.*, **18**, 913–924.

Engh,R.A. and Huber,R. (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr.*, **A47**, 392–400.

Esnouf,R.M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.*, **15**, 133–138.

Ferré-D'Amaré,A.R., Predergast,G.C., Ziff,E.B. and Burley,S.K. (1993) Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature*, **383**, 38–45.

Gomis-Rüth,F.X., Solà,M., Pérez-Luque,R., Acebo,P., Alda,M.T., González,A., Espinosa,M., del Solar,G. and Coll,M. (1998) Overexpression, purification, crystallization and preliminary X-ray diffraction analysis of the pMV158-encoded plasmid transcriptional repressor protein CopG. *FEBS Lett.*, **425**, 161–165.

Harrison,S.C. and Aggarwal,A.K. (1990) DNA recognition by proteins with the helix–turn–helix motif. *Annu. Rev. Biochem.*, **59**, 933–969.

Holm,L. and Sander,C. (1995) Dali : a network tool for protein structure comparison. *Trends Biochem. Sci*., **20**, 478–480.

Inamoto,S. and Ohtsubo,E. (1990) Specific binding of TraY protein to *oriT* and the promoter region for the *traY* gene of plasmid R100. *J. Biol. Chem.*, **265**, 6461–6466.

Jones,T.A. (1992) A yaap, asap, @#*? A set of averaging programs. In Dodson,E.J., Gover,S. and Wolf,W. (eds), *Molecular replacement*. SERC Daresbury Laboratory, Warrington, UK, pp. 91–105.

Klemm,J.D., Rould,M.A., Aurora,R., Herr,W. and Pabo,C.O. (1994) Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules. *Cell*, **77**, 21–32.

Knight,K.N. and Sauer,R.T. (1989) DNA binding specificity of the Arc and Mnt repressors is determined by a short region of N-terminal residues. *Proc. Natl Acad. Sci. USA*, **86**, 797–801.

Koradi,R., Billeter,M. and Wüthrich,K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.*, **14**, 51–55.

Kraulis,P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Lavery,R. and Sklenar,H. (1988) The definition of generalised helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.*, **6**, 63–91.

Leslie,A.G.W. (1991) Macromolecular data processing. In Moras,D., Podjanry,A.D. and Thierry,J.C. (eds), *Crystallographic computing V*. Oxford University Press, Oxford, UK, pp. 27–38.

Li,T., Stark,M.R., Johnson,A.D. and Wolberger,C. (1995) Crystal structure of the MAT**a**1/MATα2 homeodomain heterodimer bound to DNA. *Science*, **270**, 262–269.

Matthews,B.W., Ohlendorf,D.H., Anderson,D.F. and Takeda,Y. (1982) The DNA loop model for ara repression: AraC protein occupies the proposed loop sites *in vivo* and repression-negative mutations lie in these same sites. *Proc. Natl Acad. Sci. USA*, **83**, 3654–3658.

McKay,D.B. and Steitz,T.A. (1981) Structure of catabolyte gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature*, **290**, 744–749.

Moncalián,G., Grandoso,G., Llosa,M. and de la Cruz,F. (1997) OriT-processing and regulatory roles of TrwA protein in plasmid R388 conjugation. *J. Mol. Biol.*, **270**, 188–200.

Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, **A50**, 157–163.

Nelson,W.C. and Matson,S.W. (1996) The F plasmid TraY gene product binds DNA as a monomer or a dimer—structural and functional implications. *Mol. Microbiol.*, **20**, 1179–1187.

Nicholls,A., Bharadwaj,R. and Honig,B. (1993) GRASP: graphical representation and analysis of surface properties. XXXVII. Annual Meeting of the Biophysical Society, Washington DC, USA, February 14–18, 1993. *Biophys. J.*, **64**, A166–A166.

Nordström,K. (1983) Control of plasmid regulation. *Plasmid*, **9**, 1–7.

Nordström,K. and Austin,S.J. (1989) Mechanisms that contribute to the stable segregation of plasmids. *Annu. Rev. Genet.*, **23**, 37–69.

Novick,R.P. (1987) Plasmid incompatibility. *Microbiol. Rev.*, **51**, 381–395.

Ohlendorf,D.H., Anderson,W.F., Fisher,R.G., Takeda,Y. and Matthews,B.W. (1982) The molecular basis of DNA-protein recognition inferred from the structure of *Cro* repressor. *Nature*, **298**, 718–723.

Pabo,C.O. and Lewis,M. (1982) The operator-binding domain of lambda repressor: structure and DNA recognition. *Nature*, **298**, 443–447.

Pabo,C.O. and Sauer,R.T. (1984) Protein–DNA recognition. *Annu. Rev. Biochem.*, **53**, 293–321.

Pérez-Martín,J., del Solar,G.H., Lurz,R., de la Campa,A.G., Dobrinski,B. and Espinosa,M. (1989) Induced bending of plasmid pLS1 DNA by the plasmid-encoded protein RepA. *J. Biol. Chem.*, **264**, 21334–21339.

Phillips,S.E.V. *et al*. (1989) Cooperative tandem binding of *met* repressor of *Escherichia coli*. *Nature*, **341**, 711–715.

Rafferty,J.B., Somers,W.S., Saint-Girons,I. and Phillips,S.E.V. (1989) Three-dimensional crystal structures of *Escherichia coli met* repressor with and without corepressor. *Nature*, **341**, 705–710.

Raumann,B.E., Rould,M.A., Pabo,C.O. and Sauer,R.T. (1994) DNA recognition by β-sheets in the Arc repressor–operator crystal structure. *Nature*, **367**, 754–757.

Roussel,A. and Cambilleau,C. (1989) Turbo-Frodo. In *Silicon Graphics Geometry Partners Directory.* Silicon Graphics, Mountain View, CA, pp. 77–79.

Schildbach,J.F., Milla,M.E., Jeffrey,P.D., Raumann,B.E. and Sauer,R.T. (1995) Crystal structure, folding and operator binding of the hyperstable Arc repressor mutant PL8. *Biochemistry*, **34**, 1405–1412.

Schultz,S.C., Shields,G.C. and Steitz,T.A. (1991) Crystal structure of a CAP–DNA complex: the DNA is bent by 90°. *Science*, **253**, 1001–1007.

Sheldrick,G.M. (1990) Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallogr.*, **A46**, 467–473.

Somers,W.S. and Phillips,S.E.V. (1992) Crystal structure of the *met* repressor–operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature*, **359**, 387–393.

Vriend,G. (1990) WHAT IF: a molecular modelling and drug design program. *J. Mol. Graph.*, **8**, 52–56.

Wild,C.M., McNally,T., Phillips,S.E.V. and Stockley,P.G. (1996) Effects of systematic variation of the minimal *Echerichia coli met* consensus operator site—*In vivo* and *in vitro met* repressor binding. *Mol. Microbiol.*, **21**, 1125–1135.

Wilson,D.S., Guenther,B., Desplan,C. and Kuriyan,J. (1995) High resolution crystal structure of a paired (Pax) class cooperative homeodomain dimer on DNA. *Cell*, **82**, 709–719.

Wolberger,C., Dong,Y.C., Ptashne,M. and Harrison,S.C. (1988) Structure of a phage 434 Cro/DNA complex. *Nature*, **335**, 789–795.

Received July 31, 1998; revised and accepted October 23, 1998