Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 5178–5182, May 1994
Biochemistry

Chemical synthesis of a fully active transcriptional repressor protein

(affinity purification/gel retardation assays/*in vitro* transcription)

GLORIA DEL SOLAR*†, FERNANDO ALBERICIO‡, RAMÓN ERITJA§, AND MANUEL ESPINOSA*

*Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Cientificas, Velázquez, 144, E-28006 Madrid, Spain; ‡Life Sciences Research Group, Millipore Corporation, Bedford, MA 01730; and §Centro de Investigación y Desarrollo, Consejo Superior de Investigaciones Cientificas, Jordi Girona, 18, E-08034 Barcelona, Spain

Communicated by Donald R. Helinski, January 28, 1994

ABSTRACT **Plasmid pLS1-encoded 45-amino acid transcriptional repressor CopG (formerly RepA) has been chemically synthesized. A one-step purification of the synthetic protein has been developed, which yields high levels of pure protein with low or no contamination of truncated products. We have compared some properties of the chemical CopG protein with those of the biologically purified CopG. The two proteins were indistinguishable in (*i*) their ability to generate specific protein–DNA complexes, (*ii*) their capacity to protect a restriction site included within the CopG DNA target, and (*iii*) in their *in vitro* capacity to specifically repress synthesis of *copG* mRNA.**

Modulation of gene expression from a given promoter can be mediated by DNA-binding proteins. Such transcriptional control is achieved through various mechanisms, including long-distance DNA–protein and protein–protein interactions (refs. 1 and 2; reviewed in ref. 3), protein-mediated curvatures in regions upstream of promoters (4, 5), and steric hindrance of the binding of RNA polymerase to the promoter. The generation of local DNA deformations in the promoter region (6, 7) or contacts between the DNA-binding protein and RNA polymerase can also modulate promoter activity (8). Information on DNA–protein complexes can be obtained from gel retardation assays (9) or through a variety of "footprinting" techniques (10). However, the determination of the conformation of the complexes usually is the ultimate goal (11). Such information can be comprehensive if it is accompanied by mutational analysis, which is facilitated when genetic systems designed to select the desired mutations are available (12, 13). However, laborious procedures of overproduction and purification of the mutant proteins are often still required.

One way to overcome these difficulties is to attempt the chemical synthesis of all or a part of the desired protein. However, these procedures often do not yield active products, especially when a certain peptide size is surpassed. We have developed a procedure to chemically synthesize and purify a well-characterized transcriptional repressor and we have tested the activity of the final product. To this end, we have combined our knowledge of the properties of the CopG repressor (formerly RepA; ref. 14) with the advantages of 9-fluorenylmethoxycarbonyl (Fmoc) chemistry. This chemical procedure has been improved, and a one-step purification method has been developed (15). We report the complete chemical synthesis of the plasmid pLS1-encoded transcriptional repressor CopG by a relatively simple method which yields milligrams of highly pure and fully active protein. Our findings open avenues for the analysis of protein structure and function and for the generation of custom-made protein mutants.

MATERIALS AND METHODS

DNA Isolation and Manipulations. The plasmid used was pLS1Δ24, an *in vitro* clockwise deletion derivative (including coordinates 4240–401) of plasmid pLS1 (16). Plasmid DNA was prepared from *Streptococcus pneumoniae* and purified by two consecutive CsCl/ethidium bromide gradients, essentially as described (14). Digestions of DNA with restriction enzymes (New England Biolabs) were performed as specified by the supplier. Restriction fragments were purified from 0.8% agarose gels with the Geneclean kit (Bio-101).

Purification of CopG from a Biological Source. CopG (formerly RepA; ref. 14) protein was overproduced with the *Escherichia coli* BL21(DE3) host/pET5 vector expression system (17). The *copG* gene was overexpressed after induction with isopropyl β-D-thiogalactopyranoside, and its product was purified by chromatography on agarose and heparin-agarose columns, as described (14). A further step of purification and concentration was achieved by a second heparin-agarose column, using a 0.1–1 M NaCl gradient. Concentration of the purified CopG protein used here was 20 μg/ml, as calculated by determination of CopG amino acid composition.

Chemical Synthesis of CopG. Chain assembly was carried out on a Millipore 9050 Plus automated peptide continuous-flow synthesizer, starting with a Fmoc-Lys(Boc)-PAC-PEG-PS resin (Millipore; ref. 18), where Boc is *tert*-butyloxycarbonyl, PAC is a peptide acid linker, PEG is a polyethylene glycol spacer, and PS is the styrene/1% divinylbenzene copolymer constituting the resin beads. The required Fmoc-amino acids (5 equivalents) were incorporated by 1- to 2-hr couplings (as the synthesis progressed, coupling times were extended) mediated by benzotriazolyl *N*-oxytris(pyrrolidinio)phosphonium hexafluorophosphate/1-hydroxybenzotriazole/*N*,*N*-diisopropylethylamine (5, 5, and 10 equivalents, respectively) in *N*,*N*-dimethylformamide. Deprotection of the Fmoc group was done with 1,8-diazabicyclo[5.4.0]undec-7-ene/piperidine-*N*,*N*-dimethylformamide, 2:2:96 (vol/vol), for 7–12 min. After each coupling, an acetylation step with 0.3 M acetic anhydride and 0.3 M *N*,*N*-diisopropylethylamine in *N*,*N*-dimethylformamide for 10 min was carried out to block the unreacted amino groups and therefore to facilitate the final purification. After removal of the Fmoc group from the last amino acid, Biolinker *S*-Trt [*S*-(trityl)thioglycolylaminoethylsulfonylethyl *p*-nitrophenyl carbonate; 2 equivalents; Nikkon Millipore, Tokyo; ref. 15] was incorporated into the amino-terminal function in the presence of 1-hydroxybenzotriazole (2 equiv-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Fmoc, 9-fluorenylmethoxycarbonyl; B-CopG, biologically produced CopG; Ch-CopG, chemically synthesized CopG.

†To whom reprint requests should be addressed.

alents) in *N,N*-dimethylformamide for 2 hr at 25°C. Cleavage of the peptide from the resin was done by reagent R (trifluoroacetic acid/thioanisole/ethanedithiol/anisole, 90:5:3:2) for 2 hr at 25°C (19). The resulting sulfhydryl-containing peptide was dissolved in 6 M guanidine hydrochloride/0.4 M Tris·HCl, pH 7.5, and mixed with an iodoacetamide resin (≈1.1 equivalents; Nikkon Millipore, Tokyo), for 2 hr at 25°C. In this step only the desired peptide possessing the sulfhydryl group was immobilized by covalent bonding to the resin. Then, 1 equivalent of 2-mercaptoethanol was added for 30 min at 25°C to block the remaining iodoacetamide groups. The solution was filtered off, and the resin was washed with acetic acid/water, 1:1, and with methanol. The target peptide was released by treatment of the resin with aqueous 5% NH_4OH for 2 hr at 25°C and the solution was filtered off. The resin was then washed twice with aqueous 5% NH_4OH. The combined washings were lyophilized. The overall yield (synthesis and purification) was 5%.

Gel Retardation Assays. Plasmid pLS1Δ24 DNA was doubly digested with *Fnu*4HI and *Sty* I to give fragments of 1327, 1287, 753, 252, 123, 93, and 3 bp. The target of CopG is included within the 252-bp fragment, and is located ≈100 bp from both ends. DNA (20 nM) was incubated with various concentrations of CopG proteins for 10 min at 20°C in 20 mM Tris·HCl, pH 8.0/1 mM EDTA/5 mM dithiothreitol. Reaction mixtures were loaded on nondenaturing 5% polyacrylamide gels, electrophoresed at 10 V/cm, stained, and photographed. Several pictures of the same gel, with different exposure times, were taken, and the negatives were used for densitometric scannings.

CopG Protection of *Apa*LI Restriction Site. Plasmid pLS1Δ24 DNA was linearized with *Eco*RI (coordinate 3170), phenol-treated, and precipitated. DNA (6.5 nM) was incubated (final volume, 15 μl) with various amounts of CopG proteins for 10 min at 20°C in *Apa*LI buffer (10 mM Tris·HCl, pH 8.0/50 mM KCl/7 mM $MgCl_2$/7 mM 2-mercaptoethanol). Then 6 units of *Apa*LI (which cuts at pLS1 coordinate 607) was added, and the incubation was continued at 37°C for 25 min. Reaction mixtures were electrophoresed in 1% agarose gels, stained, and photographed as above.

***In Vitro* Transcription Assays.** Transcripts were synthesized *in vitro* from the 842-bp *Ban* I–*Pst* I pLS1 fragment (coordinates 214–1056), which contains two promoters: P_{cr} and P_{ctII}. The P_{cr} promoter contains the CopG target, whereas the P_{ctII} promoter (used as internal control) is not repressed by CopG. Reaction mixtures (50 μl) contained 40 mM Tris·HCl (pH 8.0), 10 mM $MgCl_2$, 150 mM KCl, 0.1 mM EDTA, 5% (vol/vol) glycerol, 200 μM GTP, 200 μM ATP, 50 μM CTP, 50 μM [α-^{32}P]UTP (4 Ci/mmol; 1 Ci = 37 GBq) and 5 nM DNA fragment. Transcription was initiated by the addition of 0.15 unit of *Escherichia coli* RNA polymerase, and samples were incubated for 10 min at 37°C. When CopG proteins were used, mixtures were incubated with the proteins for 10 min at 20°C before the addition of RNA polymerase. Samples were treated and electrophoresed as described (20).

RESULTS AND DISCUSSION

The protein synthesized, CopG, is the smallest natural transcriptional repressor so far described (20). It is encoded by

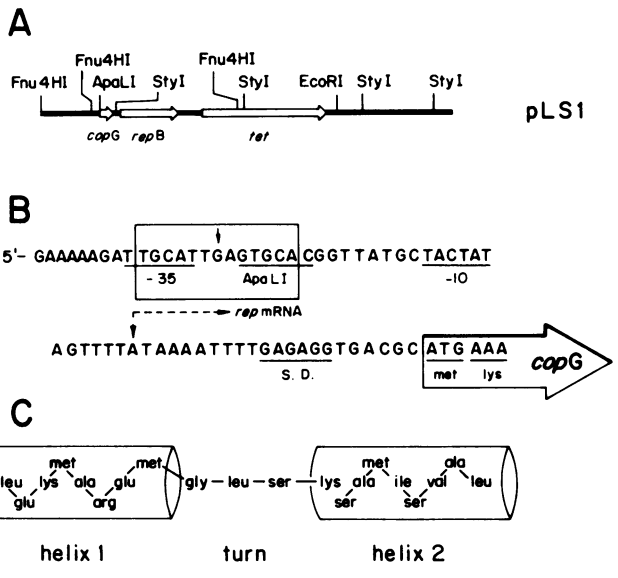

FIG. 1. Relevant features of plasmid pLS1. (*A*) Schematic map with relevant restriction sites. (*B*) The P_{cr} promoter, within which a 13-bp symmetric element (symmetric bases are underlined; arrow points to the center of symmetry) and a single *Apa*LI restriction site are included. The transcription initiation site (arrowhead), the putative ribosome-binding (Shine–Dalgarno, S.D.) site, and the first two codons of CopG are indicated. (*C*) Schematic representation of the putative helix–turn–helix motif of CopG (residues 17–36).

H-Met-Lys-Lys-Arg-Leu-Thr-Ile-Thr-Leu-Ser-Glu-Ser-

-Val-Leu-Glu-Asn-Leu-Glu-Lys-Met-Ala-Arg-Glu-Met-

-Gly-Leu-Ser-Lys-Ser-Ala-Met-Ile-Ser-Val-Ala-Leu-

-Glu-Asn-Tyr-Lys-Lys-Gly-Gln-Glu-Lys-H

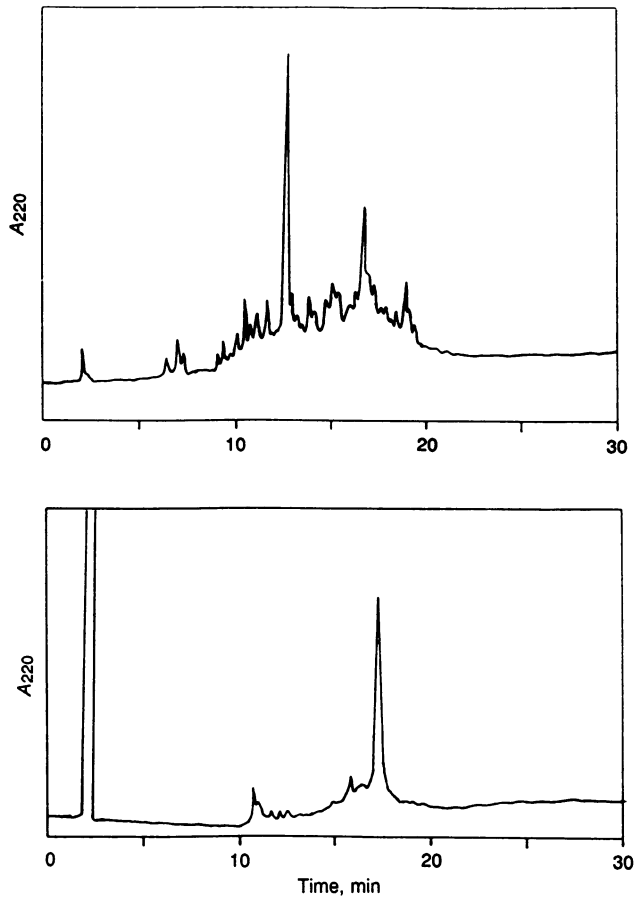

FIG. 2. Composition and synthesis of CopG. (*Top*) Amino acid composition of CopG, as determined by automatic sequencing of the biologically synthesized protein (14). (*Middle* and *Bottom*) HPLC of crude preparation and Bio-linker purification step, respectively. HPLC conditions: C_{18} resin; elution with a 20-min linear gradient from 1:9 to 6:4 of 0.1% trifluoroacetic acid/acetonitrile vs. 0.1% trifluoroacetic acid/water; flow rate, 1 ml/min.

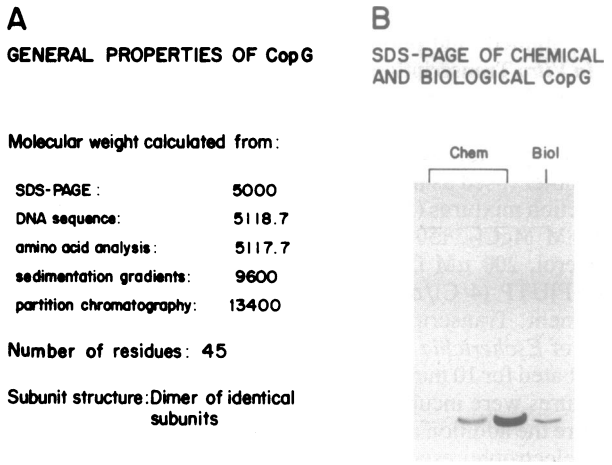

FIG. 3. Features of CopG. (*A*) General properties of B-CopG, as derived from a number of physicochemical analyses. (*B*) Comparison of electrophoretic mobility and purity of chemically synthesized CopG (Ch-CopG) and B-CopG in a denaturing SDS/17% polyacrylamide gel stained with Coomassie brilliant blue.

the promiscuous plasmid pLS1 (Fig. 1; ref. 16), but similar repressors have been described for several other bacterial plasmids (14, 20). CopG is a fully sequenced 45-amino acid polypeptide (14), which controls the synthesis of the plasmid initiator of replication, the RepB protein, and thus controls plasmid copy number (14, 21). The *copG* and *repB* genes are cotranscribed from the P_{cr} promoter, which is included in the target of CopG (20). The CopG target also includes a 13-bp symmetric element which overlaps the −35 region of the P_{cr} promoter (Fig. 1). Due to its small size, the predicted structure of CopG is simple: one β-sheet at the N terminus, the helix–turn–helix motif typical of many DNA-binding proteins (22, 23), and a positively charged random-coil C terminus (14). The DNA region protected by purified CopG protein spans 45 bp, which may be due either to the strong DNA bend introduced by CopG around the P_{cr} promoter (24) or to a very high cooperative binding of CopG to its target DNA, perhaps mediated by protein–protein interactions (20). These kinds of interactions, leading to dimer or tetramer formation, seem to be a key feature for an efficient regulatory role of the *lac* repressor (LacR), and they should be considered for any oligomeric DNA-binding protein (25).

Chemical synthesis of the entire CopG protein is problematic because of the length of the sequence, the presence of four methionine residues (Fig. 2), and the strong tendency of the protein to bind to the purification supports, yielding low or no recovery. In fact, several solid-phase syntheses based upon the methodology developed by Merrifield (26) were attempted with various chemistries without any significant recovery of the protein. This contrasts with the results obtained for the LacR-(1–56) headpiece (27), perhaps due to the positively charged C end of CopG, or to the greater methionine content of CopG (four residues) as compared with the truncated LacR synthetic peptide (two residues). Finally, a successful synthesis was achieved with a Fmoc/*t*Bu mild-orthogonal scheme (28). After the cleavage of the peptide from the resin, a chemical purification was carried out by using an affinity-type purification procedure based upon the specific reaction between the sulfhydryl group contained in the peptide and an iodoacetamide group linked to a solid support (15). The use of this purification method is critical because it yields relatively homogeneous material in a single step with excellent recoveries (Fig. 2).

Analysis of some properties of the biologically produced CopG (B-CopG) protein (Fig. 3*A*) showed that it purifies as a dimer of identical subunits, which agrees with the twofold

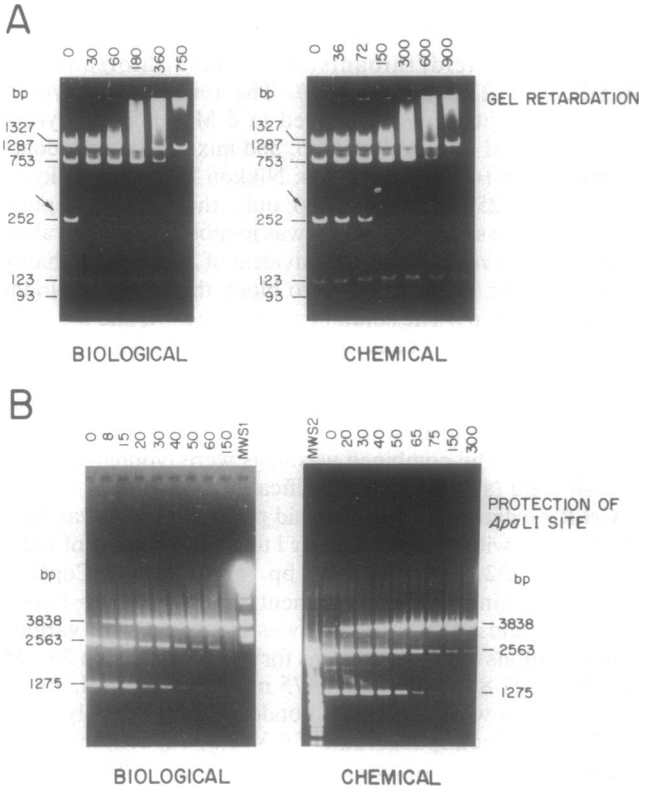

FIG. 4. Interaction of Ch-CopG and B-CopG with the DNA target. (*A*) CopG–DNA complex formation analyzed by gel retardation assays. DNA (20 nM) doubly digested with *Fnu*4HI and *Sty* I (size of the fragments is indicated on the left) was incubated with chemical or biological CopG protein (at the indicated nanomolar concentrations). DNA–protein complexes formation was analyzed in nondenaturing 5% polyacrylamide gels. The band containing the CopG target is indicated (arrow). (*B*) Specific protection of the *Apa*LI site of pLS1 by chemical and biological CopG proteins. DNA (6.5 nM) linearized with *Eco*RI was incubated with the indicated nanomolar concentrations of chemical or biological CopG. The DNA was then subjected to restriction with *Apa*LI, followed by electrophoresis in 1% agarose gels. The sizes of the linear and doubly digested DNA are indicated. MW1 and MW2, "molecular weight" standards.

symmetry of the CopG target (Fig. 1). The entire amino acid sequence of B-CopG has been determined (14), which allowed us a precise determination of its molecular weight (Fig. 3*A*). A reliable estimation on the molar concentration of the purified B-CopG protein resulted from the analysis of its amino acid composition. The molar extinction coefficient of CopG could not be measured, because the single aromatic residue of the protein was not enough to obtain absorbance information by simple methods. Several amounts of Ch-CopG from the main peak of the chromatogram after Biolinker purification (Fig. 2) were run in a 17% polyacrylamide denaturing gel (Fig. 3*B*) and compared with a known amount of B-CopG (determined as indicated above). Densitometric scanning of the gel allowed us to calculate that the total amount of pure Ch-CopG obtained was 200 μg. Scaling up the procedure gave milligram amounts of purified Ch-CopG. No significant contamination of the chemical protein with truncated peptides was found (Fig. 3*B*), indicating that the synthesis procedure can yield crystallographic-quality purified protein.

To find out whether the Ch-CopG protein retained its activity in the formation of specific DNA–protein complexes, we compared the behavior of the two purified CopG proteins by two *in vitro* assays. First, gel retardation showed that both proteins efficiently and specifically bound to a 252-bp pLS1

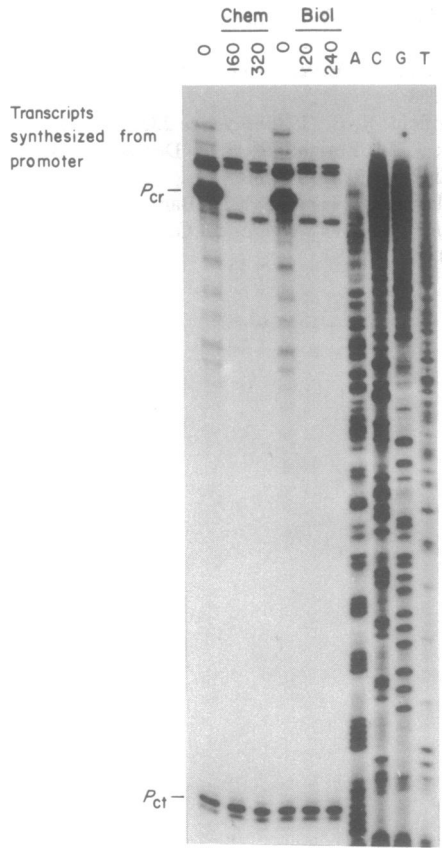

FIG. 5. Specific inhibition of RNA synthesis from the P_{cr} promoter by chemical and biological CopG proteins. Concentration of DNA in reaction mixtures was 5 nM; concentrations of CopG (nM) are indicated above each lane. The four rightmost lanes show a marker DNA sequence. Run-off transcripts above P_{cr} are probably due to artifacts generated by the protruding ends of the DNA fragments.

DNA fragment containing the target of CopG (Fig. 4*A*). Three or four orders of CopG-DNA complexes were found, and their generation depended upon the DNA/protein ratio employed. However, 2 times more Ch-CopG than B-CopG was needed to obtain total specific complex formation. At high protein concentration, smearing of the larger bands was observed. This could be due to nonspecific binding rather than overloading artifacts, since the smallest fragments were unaffected (Fig. 4*A*).

The second assay was aimed at showing CopG protection of the single pLS1 *Apa*LI site, which is partially included within the 13-bp symmetric element (Fig. 1). Full protection of the restriction site was obtained with both proteins (Fig. 4*B*), and again twice as much Ch-CopG as B-CopG was needed to achieve the same level of protection. The above results indicate that Ch-CopG has about 50% activity as compared with the B-CopG protein. Densitometric scanning of the gels allowed us to obtain a rough estimation of the K_d of both proteins, which was in the range of at least 10^{-9} M (assuming the active fraction of the proteins to be 1). A DNA fragment containing the symmetric element and the *Apa*LI restriction site of pLS1, surrounded by foreign DNA, showed low affinity for CopG (data not shown), indicating that adjacent pLS1 sequences are needed for good CopG–DNA target recognition.

Functionality of Ch-CopG was assayed by comparing the ability of Ch- and B-CopG proteins to act as specific transcriptional repressors *in vitro*. For this purpose, we used a pLS1 DNA fragment containing two promoters: P_{cr} (containing the CopG target; Fig. 1) and P_{ctII} (from which a plasmid-encoded 50-nt-long antisense RNA is transcribed; ref. 21). Analysis of transcripts synthesized in the presence or in the absence of the repressors showed that run-off transcripts from P_{cr} were specifically inhibited by both types of CopG protein, whereas the production of full transcripts from promoter P_{ctII} was unaffected by the presence of either protein (Fig. 5). We conclude that only promoter P_{cr} is fully and specifically repressed by both CopG proteins.

The results presented here show that the chemical synthesis of an active protein is a powerful tool for the analysis of DNA–protein interactions. The behaviors of Ch-CopG and B-CopG were indistinguishable, both in the generation of DNA–protein complexes and in their ability to repress transcription from the *copG* promoter. The chemical procedure used does not seem to affect the dimerization of CopG and allows proper folding of this chemically prepared protein. Moreover, the proposed chemical structure of CopG, obtained by protein isolation, purification, and sequencing (14), has been confirmed here by the classical scheme of unambiguous synthesis that yields a fully active protein. Since CopG is the smallest transcriptional repressor known, this should be a good model system to perform "mutational" studies by direct synthesis coupled with structural determinations by nuclear magnetic resonance spectrometry. Synthesis of specifically designed Ch-CopG mutants may lead to a rapid characterization of critical residues involved in DNA recognition. This in turn would allow the design of a specific DNA site-directed mutagenesis program to correlate *in vitro* activities of Ch-CopG mutants with their *in vivo* functional roles.

Research was financed by Comisión Interministerial de Ciencia y Tecnologia, Grant BIO91-0691 to (M.E.) and Grant SAL90-0828 (to R.E.). The Special Actions Program of the Consejo Superior de Investigaciones Cientificas stimulated and partially supported this research.

1. Ptashne, M. (1986) *Nature (London)* **322,** 697–701.
2. Schleif, R. (1988) *Science* **241,** 1182–1187.
3. Adhya, S. (1989) *Annu. Rev. Genet.* **23,** 227–250.
4. Nash, H. A. (1990) *Trends Biochem. Sci.* **15,** 222–227.
5. Rojo, F., Zaballos, A. & Salas, M. (1990) *J. Mol. Biol.* **211,** 713–725.
6. Travers, A. A. (1991) *Curr. Opin. Mol. Biol.* **1,** 114–122.
7. Lilley, D. M. J. (1991) *Nature (London)* **354,** 359–360.
8. Nuez, B., Rojo, F. & Salas, M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 11401–11405.
9. Lane, D., Prentki, P. & Chandler, M. (1992) *Microbiol. Rev.* **56,** 509–528.
10. Tullius, T. D., Dombroski, B. A., Churchill, M. E. A. & Kam, L. (1987) *Methods Enzymol.* **155,** 537–558.
11. Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991) *Science* **253,** 1001–1007.
12. Reidhaar-Olson, J. F. & Sauer, R. T. (1988) *Science* **241,** 53–57.
13. Kleina, L. G. & Miller, J. H. (1990) *J. Mol. Biol.* **212,** 295–318.
14. del Solar, G., de la Campa, A. G., Pérez-Martín, J., Choli, T. & Espinosa, M. (1989) *Nucleic Acids Res.* **17,** 2405–2420.
15. Funakoshi, S., Fukuda, H. & Fujii, N. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 6981–6985.
16. Lacks, S. A., López, P., Greenberg, B. & Espinosa, M. (1986) *J. Mol. Biol.* **192,** 753–765.
17. Rosenberg, A. H., Lade, B. N., Chui, D., Lin, S.-W., Dunn, J. J. & Studier, F. W. (1987) *Gene* **56,** 125–135.
18. Barany, G., Albericio, F., Biancalana, S., Bontems, S. L., Chang, J. L., Eritja, R., Ferrer, M., Fields, C. G., Fields, G. B., Lyttle, M. H., Solé, N. A., Tian, Z. T., van Abel, R. J., Wright, P. B., Zalipsky, S. & Hudson, D. (1992) in *Peptides: Proceedings of the Twelfth American Peptide Symposium*, eds. Smith, J. A. & Rivier, J. (ESCOM Science, Leiden, The Netherlands), pp. 603–604.

19. Albericio, F., Kneib-Cordonier, N., Biancalana, S., Gera, L., Masada, R. I., Hudson, D. & Barany, G. (1990) *J. Org. Chem.* **55**, 3730–3743.
20. del Solar, G., Pérez-Martín, J. & Espinosa, M. (1990) *J. Biol. Chem.* **265**, 12569–12575.
21. del Solar, G. & Espinosa, M. (1992) *Mol. Microbiol.* **6**, 83–94.
22. Ohlendorf, D. H., Anderson, W. F. & Matthews, B. W. (1983) *J. Mol. Evol.* **19**, 109–114.
23. Brennan, R. G. & Matthews, B. W. (1989) *J. Biol. Chem.* **264**, 1903–1906.
24. Pérez-Martín, J., del Solar, G., Lurz, R., de la Campa, A. G., Dobrinski, B. & Espinosa, M. (1989) *J. Biol. Chem.* **264**, 21334–21339.
25. Chakerian, A. E. & Matthews, K. S. (1992) *Mol. Microbiol.* **6**, 963–968.
26. Merrifield, R. B. (1986) *Science* **232**, 341–347.
27. Shin, J. A., Ebright, R. H. & Dervan, P. B. (1991) *Nucleic Acids Res.* **19**, 5233–5236.
28. Fields, G. B., Tian, Z. T. & Barany, G. (1991) in *Synthetic Peptides: A User's Guide*, ed. Grant, G. (Freeman, Salt Lake City), pp. 77–183.