Biochem. J. (1987) 242, 645–653 (Printed in Great Britain)

Probing the sequence-specific interaction of the cyclic AMP receptor protein with DNA by site-directed mutagenesis

Manda E. GENT,* Angela M. GRONENBORN,†‡ R. Wayne DAVIES*§ and G. Marius CLORE†

*Department of Biochemistry and Applied Molecular Biology, University of Manchester Institute of Science and Technology, Manchester M60 1QD, U.K., and †Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, Federal Republic of Germany

Mutants in the DNA-binding helix of the cyclic AMP receptor protein (CRP), as well as mutants in a synthetic DNA-binding site derived from the sequence in the *lac* regulatory region, have been constructed by oligonucleotide-directed mutagenesis, and used to study the effect of selected amino acid substitutions on CRP-mediated transcriptional activity and on sequence-specific DNA binding. It has been shown that mutation of Arg-180 to Lys or Leu abolishes both CRP-mediated expression of β-galactosidase *in vivo* and CRP binding of DNA as measured by immunoprecipitation. In contrast, the mutation of Arg-185 to Leu or Lys and the mutation of Lys-188 to Leu does not appear to influence these two parameters significantly. On the DNA side, both substitutions studied, namely the exchange of the G·C base pair in position 2 of the consensus $T_1G_2T_3G_4A_5$ motif into an A·T base pair and the exchange of the A·T base pair in position 5 for a G·C base pair, abolish specific binding. Implications of these findings with respect to the present models for specific CRP-DNA recognition are discussed.

INTRODUCTION

The cyclic AMP receptor protein (CRP) is involved in the regulation of a large number of genes in *Escherichia coli* by binding in the presence of cyclic AMP to specific target sites near the promoter of each gene that it regulates. This interaction results either in stimulation of transcription, as in the case of the *lac* operon, or in inhibition of transcription, as in the case of its own structural gene (Adhya & Garges, 1982, Ullman & Danchin, 1983; de Crombrugghe *et al.*, 1984).

CRP is a dimer of $M_r \sim 47000$ composed of two identical subunits (Anderson *et al.*, 1971; Riggs *et al.*, 1971) and its amino acid sequence has been deduced from the nucleotide sequence of the cloned gene (Aiba *et al.*, 1982; Cossart & Gicquel-Sanzey, 1982). The protein consists of two structural domains (McKay & Steitz, 1981; McKay *et al.* 1982): the larger *N*-terminal domain contains the cyclic AMP-binding site (Krakow & Pastan, 1973; Eilen *et al.*, 1978; McKay *et al.*, 1982) whereas the smaller *C*-terminal domain is involved in specific DNA binding (Ebright *et al.*, 1984*a*). Several models for the specific interaction of CRP with DNA have been put forward (Steitz *et al.*, 1983; Weber & Steitz, 1984; Ebright *et al.*, 1984*b*). As no structure of a CRP–DNA complex is available at the present time, the precise details of this interaction are still unknown. Based on conservation of amino acid sequence and structural homologies between several DNA-binding proteins (McKay & Steitz, 1981; Takeda *et al.*, 1983; Pabo & Sauer, 1984), it has been proposed that in sequence-specific DNA binding a helix–turn–helix motif interacts with the DNA in such a way that amino acids located in one of the helices contact base pairs in the major groove of the DNA and thereby determine specificity (Sauer *et al.*, 1982; Anderson *et al.*, 1982; Steitz *et al.*, 1982). This is helix F for CRP, spanning the region from Arg-180 to Lys-188.

The two models proposed for the specific interaction of CRP with its DNA-binding site in the *lac* operon (Weber & Steitz, 1984; Ebright *et al.*, 1984*b*; Ebright, 1986) involve the following features: Arg-180, the first residue of the recognition helix, contacts the G·C base pair in the major groove at position 2 of the TGTGA motif of the binding site in a bidentate hydrogen-bonding fashion (Seeman *et al.*, 1976); Glu-181, the second residue of the recognition helix, contacts the G·C base pair at position 4, forming a hydrogen bond between one carboxylate oxygen atom and the amino group of the cytosine and, in addition, one of the models proposes a further hydrogen bond with the amino group on the adjacent adenine of the A·T base pair at position 3 (Weber & Steitz, 1984). Arg-185 is supposed to form a specific hydrogen bond with either the adeninine N-7 of the A·T base pair in position 5 of the pentanucleotide recognition sequence (Weber & Steitz, 1984) or with the O-4 of the thymine (Ebright *et al.*, 1984*b*). One of the models (Weber & Steitz, 1984) proposes a further bidentate hydrogen bonding scheme between Lys-188 and the guanine of the G·C base pair in position 4, while the other model (Ebright, 1986) proposes a non-specific contact with the phosphate of T_1pG_2. In addition, a variety of non-specific contacts with other phosphates of the backbone have been proposed (Weber & Steitz, 1984; Ebright, 1986).

Using site-directed mutagenesis it is possible to test the present models for such specific contacts between particular amino acids on the protein and defined bases in the DNA site by generating mutant proteins in which the proposed contact amino acid is replaced by one that is no longer capable of making an interaction of the type proposed. In addition, the sequence of the DNA site may

Abbreviation used: CRP, cyclic AMP receptor protein.

‡ To whom correspondence and reprint requests should be addressed.

§ Present address: Allelix Inc., 6850 Goreway Drive, Mississauga, Ontario L4V 1B1, Canada.

also be altered in order to identify contacted bases. Using a genetic approach such a contact point was identified in the CRP-*lac* regulatory system between Glu-181 and the G·C base pair in position 4 of the $T_1G_2T_3G_4A_5$ consensus sequence (Ebright *et al.*, 1984*a*).

In this study we investigate the effects of a variety of CRP mutants located in the DNA-binding helix F on specific DNA binding and transcription activity as well as the binding of wild-type CRP to synthetic DNA-binding sites whose sequences are derived from the *lac* promoter region.

MATERIALS AND METHODS

Bacterial strains and plasmids

The *E. coli* strain used for production of the wild type and mutant CRP proteins was pp47 (*crp*$^-$), kindly provided by H. Aiba. The mutant *crp* gene was inserted on a 980 bp *Eco*RI/*Hin*dIII fragment into the inducible expression vector PINIII A1, carrying the *lac UV5* and *lpp* gene promoters and the *lac I* gene (Masui *et al.*, 1983).

Molecular cloning

Preparation of plasmid DNA, restriction endonuclease cleavage, isolation and labelling of fragments, ligation, and transformation were carried out as described by Maniatis *et al.* (1982).

Antibodies and Western blotting

The polyclonal anti-CRP antibodies, the electrophoretic transfer from SDS/polyacrylamide protein gel to nitrocellulose and the method of immunodetection were as described previously (Gronenborn & Clore, 1986).

Site-specific mutagenesis

CRP mutants were constructed by oligonucleotide-directed mutagenesis (Zoller & Smith, 1984). The template was single-stranded M13mp8 containing the *crp* gene on a 951 bp *Bam*H1 fragment. As mutagenic primers we used 15–20 bp oligonucleotides containing either single mismatches [e.g. Arg-180 (CGT)→Leu (CTT), Arg-185 (CGC)→Leu (CTC), Ser-128 (TCA)→Ala (GCA), Gln-193 (CAG)→Leu (CTG)] or double mismatches [e.g. Thr 182 (ACC)→Val (GTC), Lys-188 (AAG)→Leu (CTG)], and 23–26 bp oligonucleotides containing triple mismatches [e.g. Arg-180 (CGT)→Lys (AAG)], approximately in the middle of each oligonucleotide. The double mutant Val-183→Leu+Arg-185→Lys was created with a 23-mer oligonucleotide carrying a single mismatch in the valine codon (GTG→CTG) and a triple mismatch in the arginine codon (CGC→AAG). The deletion mutant of the F-helix was created during the mutagenesis of Arg-180 since looping out of the intervening sequence is possible due to the fact that the sequence AAACC around the position coding for amino acids 181/182 is identical with a sequence further downstream at a position coding for amino acids 201/202. Phage harbouring the mutations were initially identified by dot-blot hybridization using the ^{32}P-labelled mutagenic oligonucleotide as a probe. The frequency of mutation was between 1% and 25%. After plaque purification the single strand of a mutant clone was sequenced by the dideoxy method (Sanger *et al.*, 1977). This verified the occurrence of the desired mutations as the sole ones within the entire coding sequence. The mutated *crp* gene was subsequently cloned into the plasmid PINIII A1 as an *Eco*RI/*Hin*dIII fragment and the recombinant plasmid was used for transformation of pp47. The sequence of the mutant gene in the plasmid was again checked by sequencing (Chen & Seeburg, 1985).

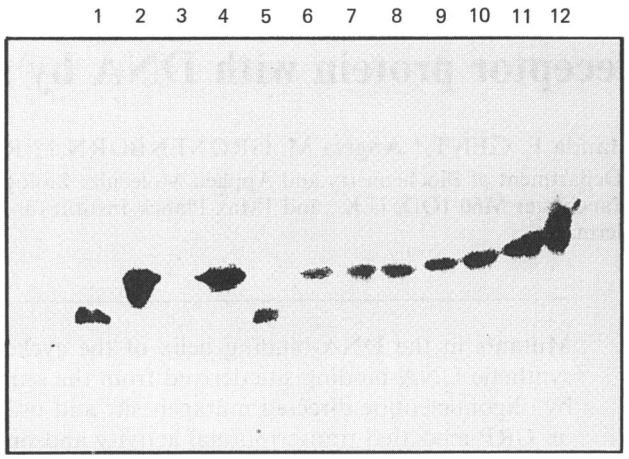

Fig. 1. Western blot of cell extracts of *E. coli* pp47 containing the various mutant *crp* genes

Lane 1, PIN III ΔF helix; lane 2, purified wild-type CRP; lane 3, PIN AIII; lane 4, PIN AIII wild type CRP; lane 5, PIN AIII ΔF helix; lane 6, PIN AIII Ser-128→Ala; lane 7, PIN AIII Arg-180→Leu; lane 8, PIN AIII Arg-180→Lys; lane 9, PIN AIII Arg-185→Leu; lane 10, PIN AIII Val-183→Leu+Arg-185→Lys; lane 11, Pin AIII Gln-193→Leu; lane 12, PIN AIII Lys-188→Leu. The Figure shows an autoradiogram of a blot of a 18% SDS/polyacrylamide gel of cell extracts after incubation with rabbit anti-CRP serum (1:20 dilution) and detection with iodinated goat anti-rabbit antibodies (Clore & Gronenborn, 1986).

CRP purification

CRP was purified from an overproducing *E. coli* strain harbouring the heat-inducible plasmid pPLcCRP1 which carries the *crp* gene under transcriptional control of the λ promoter P_L (Gronenborn & Clore, 1986). It was > 98% pure as judged by SDS/polyacrylamide-gel electrophoresis. Its concentration was determined spectrophotometrically using $\epsilon_{278} = 4.1 \times 10^4$ M^{-1}·cm^{-1} for the dimer (Takahashi *et al.*, 1982).

β-galactosidase assays

β-Galactosidase activity in cell extracts harvested from the exponential growth phase was measured according to the procedure described by Miller (1972).

DNA binding and immunoprecipitation

Binding of CRP to DNA was performed in 20 mM-Tris/HCl (pH 7.5)/100 mM-NaCl/5 mM-cyclic AMP/2 mM-dithiothreitol/1 mM-EDTA/0.01% bovine serum albumin/0.05% Nonidet P40/25% (v/v) glycerol in a total volume of 600 μl. After incubation at 0 °C for 45 min, 50 μl of serum containing anti-CRP antibodies was added and the mixture was incubated for further 2 h at 0 °C. The cyclic AMP·CRP–DNA–antibody complex was immunoprecipitated with 100 μl of washed formalin-fixed *Staphylococcus aureus* (10 mg) for 1 h at 0 °C. To reduce background binding the precipitate was washed

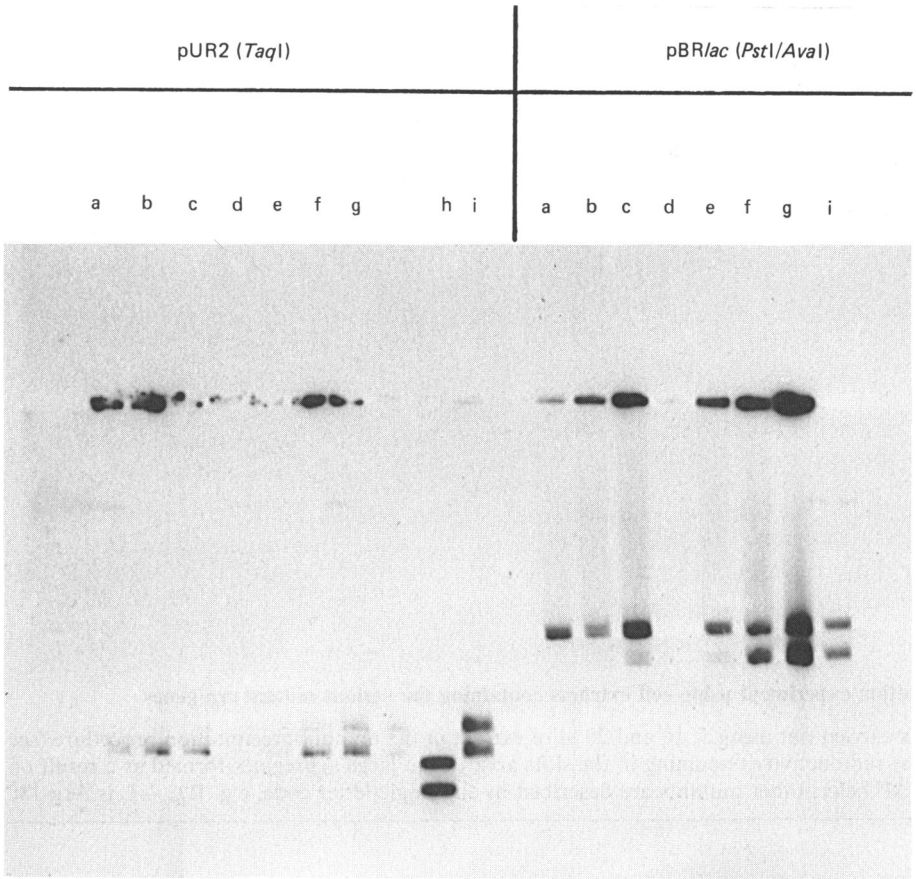

Fig. 2. Immunoprecipitation experiment with wild-type CRP using *Taq*I-digested with PUR2 DNA (left-hand panel) and *Pst*I/*Ava*I-digested pBR332 *lac* DNA (right-hand panel)

Lanes a, b and c, 5, 10 and 20 μl, respectively, of crude extract from *E. coli* ΔH1Δ*trp* harboring the plasmid pPLcCRP1 grown at 42 °C induced state); lane d, 20 μl of crude extract from the same strain grown at 30 °C (repressed state); lanes e, f and g, 1, 3 and 5 μl of purified CRP (5 μM stock solution); lane h, heat-denatured single-stranded PUR2/*Taq*I DNA; lane i, input DNA. Lane d serves as a control as no plasmid-encoded CRP is produced in the repressed low-temperature (30 °C) state; when the temperature of the culture is shifted from 30 °C to 42 °C in mid-exponential phase (lanes a, b and c), however, CRP production is switched on up to levels approx. 100 times that of the strain in the non-induced state (Gronenborn & Clore, 1986).

with 0.5 ml of wash buffer containing 20 mM-Tris/HCl, pH 8.0, 150 mM-NaCl, 2 mM-dithiothreitol, 1 mM-EDTA, 0.01% bovine serum albumin, 0.05% Nonidet P40 and 10 mg of sheared calf thymus DNA/ml. The DNA was eluted from the immune complex in 200 μl of 2% SDS in TE buffer (10 mM-Tris/HCl/1 mM-EDTA, pH 7.0), the supernatant was phenol-extracted twice, chloroform-extracted once and finally ethanol-precipitated. The immunoprecipitated DNA fragments were separated on 1% agarose gels and the dried gels were autoradiographed at −70 °C with intensifying screens for 24–48 h.

Polyacrylamide-gel electrophoresis of DNA–protein complexes

This was essentially carried out as described previously (Garner & Revzin, 1981; Kolb *et al.*, 1983*a*). Samples containing different amounts of CRP were incubated for 20 min at room temperature. Electrophoresis was performed on 5% or 8% polyacrylamide gels for 2–3 h in 30 mM-Tris/90 mM-borate/2.5 mM-EDTA/0.2 mM-cyclic AMP. The gels were autoradiographed at −70 °C with intensifying screens for 12–24 h.

RESULTS

DNA binding by mutant proteins

The following CRP mutants were constructed and used in our study: Arg-180→Lys; Thr-182→Val; Arg-185→Leu; Val-183→Leu+Arg-185→Lys; Lys-188→Leu; Ser-128→Ala; and ΔF. The latter contains the Arg-180→Leu substitution and a deletion of helix F spanning the region from residue 182 to residue 201 (Sims *et al.*, 1985). In order to assess the DNA-binding properties of the mutant proteins we used an immunoprecipitation method which can be carried out on crude cell extracts (McKay, 1981). As a prerequisite for this method to be valid, it was ascertained that polyclonal antibodies raised against purified wild type CRP (Gronenborn & Clore, 1986) also recognized the different mutant proteins. This was done by a Western blot experiment in which an SDS/polyacrylamide protein gel of *E. coli* cell extracts was blotted onto nitrocellulose and probed with anti-CRP antibodies. This clearly showed that all constructed mutant proteins are indeed recognized by our antiserum (Fig. 1). Similar results were obtained using a dot-blot immunoassay of the same protein extracts (results not shown).

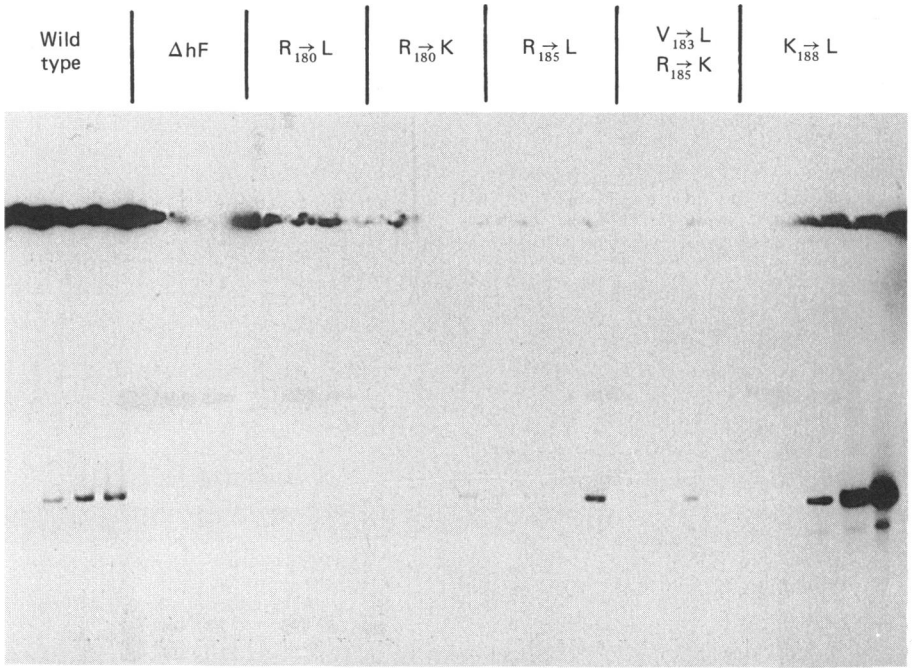

Fig. 3. Immunoprecipitation experiment using cell extracts containing the various mutant *crp* genes

Each experiment was carried out using 5, 10 and 20 μl of extract in the immunoprecipitation procedure (see the Materials and methods section). The radioactivity remaining in the slots arises from large aggregates formed as a result of non-specific DNA binding. Key: ΔhF, ΔF helix; other mutants are described by the single-letter code, e.g. $R_{180} \rightarrow L$ is Arg-180 → Leu.

The immunoprecipitation assay is based on the ability to separate DNA to which at least one protein molecule is bound from free DNA by precipitation with a polyclonal antibody raised against the DNA-binding protein (McKay, 1981; Clore *et al.*, 1982). Since radioactively labelled DNA is used in the binding experiment, fragments which are bound can be easily detected by separation on agarose gels with subsequent autoradiography. Note that as the experiments are carried out with a large excess of antibody any small variations in the affinity of the wild type and mutant proteins for the antibody will have no effect on the results. Fig. 2 shows the result of an immunoprecipitation experiment using two different plasmids which both contain a specific binding site for CRP. In the left-hand panel the experiment was carried out using pUR2 DNA (Rüther, 1980) which after digestion with *Taq*I yields two fragments, 1443 bp and 1242 bp long, the smaller of which carries the CRP-binding site of the *lac* region. Increasing amounts either of cell extract of *E. coli* cells harbouring the *crp* gene on a heat-inducible plasmid (lanes a–c) (Gronenborn & Clore, 1986) or of purified CRP (lanes e–g) result in the predominant retention of the smaller fragment. The right-hand panel shows the results of a similar experiment using DNA fragments derived from a pBR322 derivative carrying a 203 bp fragment of the *lac* region (Schaeffer *et al.*, 1982). In this case the plasmid DNA was digested with the restriction enzymes *Pst*I and *Ava*I, yielding a 2382 bp fragment containing the CRP-binding site and a smaller 2184 bp fragment without a specific site. Here too we find selective retention of the fragment containing the binding site, this time the larger one, for small concentrations of CRP. If the protein concentrations, however, are high, non-specific binding also occurs, manifesting itself by the appearance of the smaller fragment as well as increasing amounts of radioactivity in the wells.

Fig. 3 shows the equivalent experiment carried out using mutant extracts. For most mutants the retention of the fragment carrying the CRP binding site is substantially less than that for the wild-type protein. The Lys-188 → Leu mutant, however, exhibits a substantially higher degree of binding to the specific-site-containing fragment. All mutants were also investigated for β-galactosidase expression, the results of which are listed in Table 1. The amount of β-galactosidase produced by the wild-type clone was set to 1 and the results of the mutant β-galactosidase assays were normalized to this value, and are the average of at least three independent experiments. Three classes of mutants emerge: (i) strains that produce virtually no β-galactosidase, comprising the vector without any gene insert, the ΔF-helix mutant, both Arg-180 mutants and the Val-183/Arg-185 double mutant; (ii) strains that show slightly reduced or comparable levels of β-galactosidase expression with respect to the wild-type clone, namely the Ser-128 → Ala mutant, the Arg-185 → Leu mutant, the Glu-193 → Leu mutant and the Thr-182 → Val mutant; and (iii) one example with elevated levels of β-galatosidase, namely the Lys-188 → Leu mutant. It is interesting to note that the latter mutant is the one which appears to show an increased level of binding to the specific DNA site. Screening of the above mutants for sugar utilization on colour indicator plates shows the same trend as that observed in the β-galactosidase assays and the results are summarized in Table 2. In this respect it might be worth pointing out that there is a consistent and noticeable difference between the Arg-180 → Leu and Arg-180 → Lys mutant. Whereas the Leu substitution results in white

Table 1. Relative levels of β-galactosidase expression

The relative error in these measurements is $< \pm 15\%$.

Protein	Relative level
CRP wild type	1
PIN	0.01
ΔF-helix	0.01
Ser-128→Ala	0.5
Arg-180→Leu	0.02
Arg-180→Lys	0.05
Thr-182→Val	0.7
Arg-185→Leu	0.87
Val-183→Leu+Arg-185→Lys	0.04
Lys-188→Leu	1.5
Gln-193→Leu	0.9

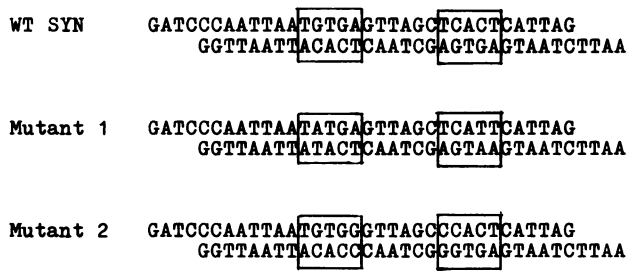

Fig. 4. Oligonucleotides used in the construction of the synthetic CRP-binding sites

Each 34-mer was synthesized (six in total), the two corresponding ones annealed and the double-stranded fragments were subsequently cloned into *Eco*RI/*Bam*H1-digested pBR322.

colonies on EMB plates, the Lys mutant shows slightly pink colonies, thus demonstrating some intermediate level of sugar utilization, albeit a rather low one.

DNA binding of wild-type CRP to mutant DNA sites

DNA binding by purified wild-type CRP was studied using the gel retardation assay (Garner & Revzin, 1981; Fried & Crothers, 1981; Kolb *et al.*, 1983*b*). In addition to a 203 bp fragment covering the wild-type *lac* promoter region, we used fragments containing synthetic CRP-binding sites. These were constructed by cloning the synthetic oligonucleotides shown in Fig. 4 into pBR322. Digestion of the resulting plasmids with *Eco*RI and *Sal*I yielded 310 b fragments containing the CRP-binding site at the *Eco*RI end.

Fig. 5(*a*) shows the results of this binding assay for the 203 bp *lac* wild-type fragment and for the 310 bp fragment containing the synthetic wild-type sequence. The appearance of the band corresponding to the specific 1:1 DNA–CRP complex occurs at equivalent concentrations of CRP for both fragments (lane d). Thus the synthetic binding site is recognized by CRP with a similar affinity as the wild-type site. Note also the abnormal electrophoretic mobility for the 203 bp fragment which carries the CRP-binding site approximately in the middle. A comparison of CRP binding to the 203 bp wild-type *lac* fragment and a mutant 310 bp fragment carrying a G→A change at position 2 of the TGTGA motif and a G→A change in the symmetrically related AGTGT sequence is shown in Fig. 5(*b*). At CRP concentrations where the formation of the specific 1:1 complex is already complete (lanes a–c) for the 203 bp fragment, no complex formation is observed with the mutant fragment. At higher concentrations, however, retention of the mutant fragment in the wells occurs, most likely due to the formation of non-specific complexes carrying a large number of protein molecules. Fig. 5(*c*) shows the comparison for all three 310 b fragments. As can be seen in lanes f–j, binding to the synthetic wild-type sequence occurs at CRP concentrations above 3×10^{-9} M, whereas binding to both mutant sites (lanes a–e) occurs only above 1.5×10^{-7} M. The latter do not show any evidence for the formation of a specific 1:1 complex. Similar experiments carried out with 34 bp *Bam*H1/*Eco*RI fragments just comprising the synthetic sequences again demonstrate specific complex formation for the synthetic wild-type sequence; no specific 1:1 complex, however, could be detected for either mutant site (results not shown).

DISCUSSION

The mutants we have constructed comprise several amino acid exchanges which are central to the various

Table 2. Colour of CRP mutants on EMB indicator plates

Mutant	Colour when grown with: Lactose	Galactose	Maltose
PIN	White	White	White
CRP wild type	Dark purple	Dark purple/green shine	Dark purple/green shine
ΔF-helix	White	White	White
Ser-128→Ala	Dark purple	Dark purple/green shine	Dark purple/green shine
Arg-180→Leu	White	White	White
Arg-180→Lys	Slightly pink	White	Slighty pink
Thr-182→Val	Dark purple	Dark purple/green shine	Dark purple/green shine
Arg-185→Leu	Purple*	Purple	Purple
Val-183→Leu / Arg-185→Lys	White	White	White
Lys-188→Leu	Dark Purple/green shine	Dark purple/green shine	Dark purple/green shine
Gln-193→Leu	Dark purple	Dark purple/green shine	Dark purple/green shine

* Slightly lighter than wild type.

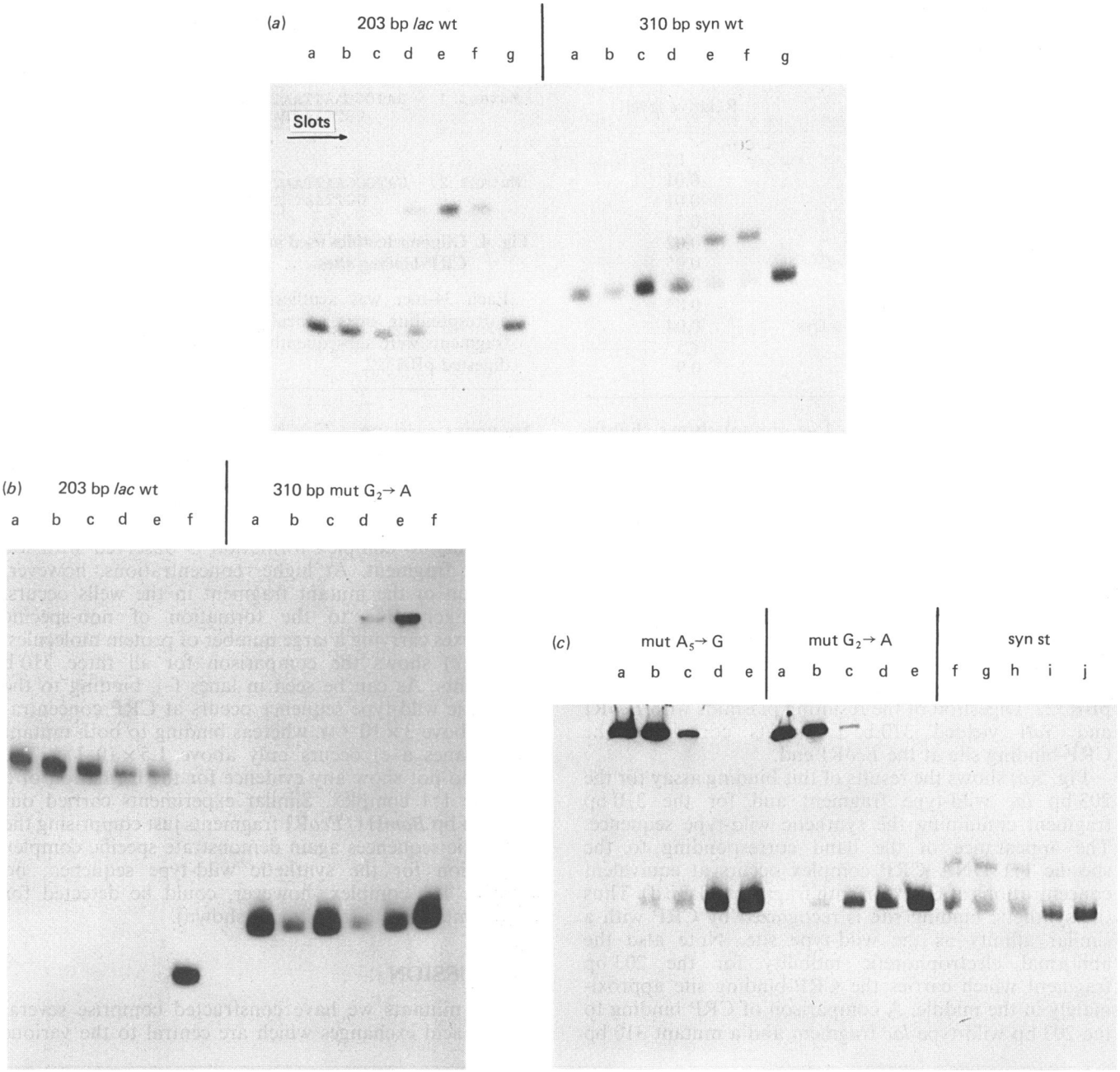

Fig. 5. Gel retardation assays using purified CRP and various DNA binding sites

(*a*) Comparison of CRP binding to the 203 bp *lac* wild-type (wt) fragment and to the synthetic (syn) 310 bp *lac* fragment. Lanes a, b, c, d, e, f, and g: 6.2×10^{-10} M-, 1.6×10^{-9} M-, 3.2×10^{-9} M-, 6.4×10^{-9} M-, 1.56×10^{-8} M-, 3.2×10^{-8} M- and 0 M-CRP, respectively. (*b*) Comparison of CRP binding to the 203 bp *lac* wild-type fragment and to the synthetic *lac* site containing the TATGA mutation. Lanes a, b, c, d, e and f: 1.5×10^{-8} M-, 3×10^{-8} M-, 6×10^{-8} M-, 1.5×10^{-7} M-, 3×10^{-7} M- and 0 M-CRP, respectively. (*c*) Comparison of CRP binding to all three synthetic CRP binding sites. Lanes a, b, c, d, e, f, g, h, i, and j: 6×10^{-7} M-, 3×10^{-7} M-, 1.5×10^{-7} M-, 6×10^{-8} M-, 3×10^{-8} M-, 1.5×10^{-8} M-, 6×10^{-9} M-, 3×10^{-9} M-, 1.5×10^{-9} M- and 6×10^{-10} M-CRP, respectively.

models that have been proposed for the specific interaction of CRP with DNA (Weber & Steitz, 1984; Ebright *et al.*, 1984*b*; Ebright, 1986) and thus enable us to test them experimentally.

As is evident from the assays *in vivo* (β-galatosidase expression as measured by the enzyme assay and by colour indicator plates) as well as the DNA-binding experiments using immunoprecipitation, mutants involving Arg-180, which either abolishes one or two of the possible hydrogen bonds, lead to loss of specific recognition. Thus we conclude that Arg-180 is indeed a crucial amino acid in determining specificity, either because it is involved directly in hydrogen bonding with the DNA site or because it plays a crucial role in determining the correct conformation of the turn between helices E and F. The role of Glu-181 for specific

recognition has already been examined (Ebright *et al.*, 1984*a*) and further studies using mutants at this position indicate a strong specificity between the Glu and the G·C base pair at position 4 (Ebright *et al.*, 1985). Our results for the substitution of Arg-185 by Leu, which abolishes any possibility for specific hydrogen bonding by the amino acid side chain, indicate, in contrast with the models, that the functional group of this amino acid cannot be important for determining specificity. Since, however, we have no data with respect to the non-specific DNA-binding properties of this mutant protein, we cannot exclude non-specific interactions with the phosphate backbone in this case.

The most surprising result with respect to the models concerns the mutant at position 188. Substitution of Lys-188 by Leu increases the affinity for DNA fragments containing the specific recognition site of the protein and enhances transcriptional activity. Thus it is unlikely that the functional group of Lys-188 makes a positive contribution to the DNA-binding affinity, be it at the non-specific or specific level. On the contrary, it seems that replacing the rather bulky lysine side chain by the smaller leucine one allows the protein to achieve a better fit on the DNA target site.

The double mutant Val-183→Leu+Arg-185→Lys abolishes specific binding. In view of the results obtained with the Arg-185→Leu mutant, we attribute this loss in specific binding to the Val→Leu substitution, although this can only strictly be proven if the Arg-185→Lys single mutant also shows no substantial loss in specific binding. It is noteworthy that this valine is one of the best conserved amino acids in the recognition helix of the helix–turn–helix motif in DNA-binding proteins (Anderson *et al.*, 1982; Gicquel-Sanzey & Cossart, 1982; Pabo & Sauer, 1984). Most known sequences show either a valine or isoleucine in this position. Inspection of the crystal structures (McKay *et al.*, 1982; Anderson *et al.*, 1981; Pabo *et al.*, 1982) reveals that the side chains of these amino acids point towards the other helix of the helix–turn–helix motif, and are part of the hydrophobic interface between the two helices. The amino acid on helix E of CRP which is closest to Val-183 is Arg-169. All other proteins show a considerably less bulky side chain at the equivalent position, mostly Gln, but also Thr, Ile or Leu. Thus increasing the size of the side chain in helix F from Val to Leu in CRP may possibly interfere with the proper alignment of the two helices with respect to each other, thereby influencing the correct positioning of the DNA-binding face of helix F on the DNA.

With respect to the mutations in the specific DNA site we have to consider the following findings. Wild-type CRP shows the strongest binding towards the site in the *lac* operon which contains two symmetrically related XGTGA sequences, with the binding reduced in a stepwise fashion for the *mal* sequence, which contains one intact XGTGA sequence with the symmetrically related one only containing the TG dinucleotide at positions 3 and 4, and the *gal* sequence, which also contains one intact XGTGA pentamer but only a single T at the symmetrically related position 3. For the *lac* L8 and L29 mutations, which contain a substitution of the G·C base pair for an A·T base pair at position 4 of the pentanucleotide, specific binding is also greatly reduced (Kolb *et al.*, 1983*b*). Thus it seems clear that at least one intact XGTGA motif is necessary for specific recognition by CRP. This is consistent with the finding that binding of CRP to the *lac* operator site, which also contains one TGTGA pentanucleotide, can be detected by footprinting (Schmitz, 1981).

Our mutations contain symmetrically related base substitutions, namely a G·C to A·T change at position 2 and an A·T to G·C change at position 5. Thus in both cases no intact TGTGA sequence is present. As a consequence, specific DNA binding of CRP is abolished, pointing to the importance of these mutated positions. The G·C base pair at position 2 had been implicated in specific binding in the existing models, via hydrogen bond formation between Arg-180 and the N-7 and O-4 of the guanine ring. Thus changing the guanine to an adenine prevents this bidentate hydrogen bonding scheme with the loss of at least one specific hydrogen bond. The A·T base pair at position 5 has also been proposed as a candidate for specific hydrogen bonding involving the guanidinium group of Arg-185. Our protein mutants, however, show that Arg-185 can be substituted by a leucine without substantial loss in specific binding. On the other hand, changing the A·T base pair at position 5 to a G·C base pair again abolishes specific binding. This points to the importance of this A·T base pair for specific recognition, especially as it is highly conserved in a number of specific binding sites for repressor (Ebright, 1986), but does not necessarily imply the loss of a specific interaction with an amino acid side chain via hydrogen bonding. It is quite clear that sequence-specific variations of DNA structure may play an additional role in specific recognition by DNA-binding proteins. Indeed, in the crystal structure of the only available high resolution protein–DNA complex to date, namely the *Eco*RI-DNA complex (Frederick *et al.*, 1984), the DNA structure exhibits distinct distortions from classical regular B-DNA. Further, there is evidence for structural changes induced by specific binding of CRP to its target site which manifest themselves as 'bending' or 'kinking' of the DNA (Wu & Crothers, 1984; Gronenborn *et al.*, 1984; Kotlarz *et al.*, 1986); the molecular details, however, of the altered DNA structure within the complex are still unknown. The other alternative is a hydrophobic interaction involving the methyl group of the thymine of the A·T base pair with an amino acid side chain. A possible candidate for such an interaction is Ser-179.

Considering all the available data, we propose that the most important amino acids for the specific interaction of CRP with DNA are those lying in the top left region of the helical wheel as illustrated in Fig. 6, with amino acids in position 1 and 2 being responsible for correct positioning of CRP on the DNA. The Gly in position 5 is clearly not involved in any specific contacts but is reasonably close to the phosphate backbone. Consequently, we predict that substitution by a bulky amino acid in this position may lead to a decrease in the DNA-binding affinity, whereas substitution by an amino acid which is capable of forming a hydrogen bond to the phosphate oxygen, such as serine, might lead to an increase in the non-specific binding affinity. As far as we are aware there is no indication that any amino acid beyond position Arg-185 is involved in specific contacts with the DNA site; indeed our results show that neither the functional group of Arg-185 nor that of Lys-188 appears to be important in specific recognition. Thus we think it most likely that only the *N*-terminal part of helix F makes specific contacts with the exposed edges of the

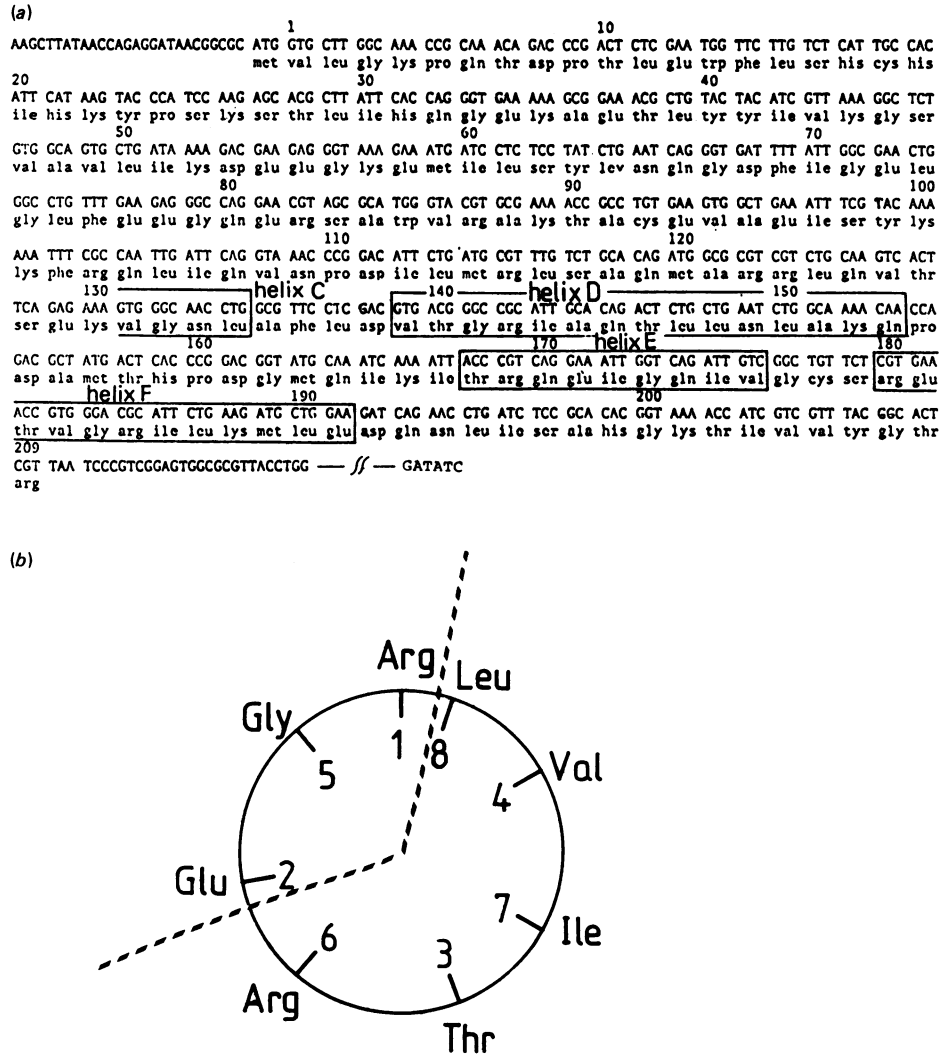

Fig. 6. (*a*) Nucleotide sequence of the *crp* structural gene (Aiba *et al.*, 1982; Cossart & Gicquel-Sanzey, 1982) indicating some of the structural elements of the three-dimensional structure, and (*b*) helical wheel for the first eight amino acids of helix F

bases in the major groove. Indeed, this finding may point to a rather simple explanation for the involvement of protein helices in DNA binding. It has been shown that polypeptide helices exhibit a dipole with the negative pole at the *C*-terminus (Hol *et al.*, 1978). Therefore it is quite possible that the positive charge at the *N*-terminal end of an α-helical protein segment will be attracted to the negatively charged phosphate backbone, thereby positioning this helix with respect to the DNA. Specific amino acids at this *N*-terminal end of the helix will then be involved in the recognition of particular base pairs, with an Arg at position 1 interacting with a G·C base pair and a Glu, Gln or Asn interacting with an A·T base pair. The particular orientation of the helix with respect to the major groove will subsequently be determined by the best fit of the individual side chains to the particular DNA structure. Whether a specific set of rules, as has been proposed by Ebright (1986), will emerge for all cases where protein helices are involved in recognition of a DNA sequence will have to await more experimental data on a variety of DNA-binding proteins, and no doubt the introduction of mutations into these helices will help to answer this question.

We thank H. Aiba for the gift of pp47, S. Busby for the gift of pBR322 containing the 203 bp *lac* insert, S. Minter for the synthesis of the oligonucleotides used in the mutagenesis, H. Erfle for the synthesis of oligonucleotides used to construct the mutant DNA-binding sites, C. Jansen for help with the immunoprecipitation and S. Gärtner for skilful technical assistance. We also thank R. Ebright for communicating results prior to publication. M.G. thanks P. Sims for advice and encouragement. This work was supported by the Max-Planck Gesellschaft and Grant Gr 658/3-1 from the Deutsche Forschungsgemeinschaft to A.M.G. and G.M.C., and by a grant from the SERC (U.K.) to R.W.D. M.G. acknowledges support from a SERC pre-doctoral fellowship, from a short term EMBO travelling fellowship during her stay in Martinsried, and from Allelix Inc.

REFERENCES

Adhya, S. & Garges, S. (1982) Cell **29**, 287–289
Aiba, H., Fujimoto, S. & Ozaki, N. (1982) Nucleic Acids Res. **10**, 1345–1362
Anderson, W. F., Ohlendorf, D. H., Takeda, Y. & Matthews, B. W. (1981) Nature (London) **290**, 754–758

Anderson, W. F. Takeda, Y., Ohlendorf, D. H. & Matthews, B. W. (1982) J. Mol. Biol. **159**, 745–755
Anderson, W. B., Schneider, A. B., Emmer, M., Perlman, R. L. & Pastan, I. (1971) J. Biol. Chem. **246**, 5929–5937
Chen, E. Y. & Seeburg, P. H. (1985) DNA **4**, 165–170
Clore, G. M., Gronenborn, A. M. & Davies, R. W. (1982) J. Mol. Biol. **155**, 447–466
Cossart, P. & Gicquel-Sanzey, B. (1982) Nucleic Acids Res. **10**, 1363–1378
de Crombrugghe, B. Busby, S. & Buc, H. (1984) in Biological Regulation and Development (Goldberger, R. F. & Yamamoto, K. R., eds.), vol. 3B, pp. 129–167, Plenum, New York
Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwith, J. (1984*a*) Nature (London) **311**, 232–235
Ebright, R. H., Cossart, P., Gicquel-Sanzey, B. & Beckwirth, J. (1984*b*) Proc. Natl. Acad. Sci. U.S.A. **81**, 7274–7278
Ebright, R. H., Kolb, A., Buc, H., Cossart, P., Gicquel-Sanzcy, B., Krakow, J., Kunkel, T. & Beckwith, J. (1985) in Fourth Conversation in Biomolecular Stereodynamics (Sarma, R. H., ed.), pp. 210–211, Albany, NY
Ebright, R. H. (1986) in Protein Structure, Folding, and Design (Oxender, D., ed.) A. R. Liss, New York, in the press
Eilen, E., Pampeno, C. & Krakow, J. S. (1978) Biochemistry **17**, 2469–2473
Frederick, C. A., Grable, J., Melia, M., Samudzi, C., Jen-Jacobsen, L., Wang, B. C., Greene, P., Boyer, H. W. & Rosenberg, J. M. (1984) Nature (London) **309**, 327–330
Fried, H. & Crothers, D. (1981) Nucleic Acids Res. **9**, 6505–6525
Garner, M. & Revzin, A. (1981) Nucleic Acids Res. **9**, 3047–3060
Gioquel-Sanzey, B. & Cossart, P. (1982) EMBO J. **5**, 591–595
Gronenborn, A. M. & Clore, G. M. (1986) Biochem. J. **236**, 643–649
Gronenborn, A. M., Nermut, M. V., Eason, P. & Clore, G. M. (1984) J. Mol. Biol. **179**, 751–757
Hol, W. G. J., van Duijnen, P. T. & Berendsen, H. J. C. (1978) Nature (London) **273**, 443–446
Kolb, A., Busby, S. Herbert, M., Kotlarz, D. & Buc, H. (1983*a*) EMBO J. **2**, 217–222
Kolb, A., Spassky, A., Chapon, G., Blazy, B. & Buc, H. (1983*b*) Nucleic Acids Res. **11**, 7833–7852
Kotlarz, D. Fritsch, A. & Buc, H. (1986) EMBO J. **5**, 647–652
Krakow, J. S. & Pastan, I. (1973) Proc. Natl. Acad. Sci. U.S.A. **70**, 2529–2533
Maniatis, T. Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
Masui, Y., Coleman, J. & Inouye, M. (1983) in Experimental Manipulation of Gene Expression (Inouye, M., ed.) Academic Press, New York
McKay, R. D. G. (1981) J. Mol. Biol. **145**, 471–488
McKay, D. B. & Steitz, T. A. (1981) Nature (London) **290**, 744–749
McKay, D. B., Weber, I. T. & Steitz, T. A. (1982) J. Biol. Chem. **257**, 9518–9524
Miller, J. (1972) Experiments in Molecular Genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y. & Matthews, B. W. (1982) Nature (London) **298**, 718–723
Pabo, C. O. & Lewis, M. (1982) Nature (London) **298**, 443–447
Pabo, C. D. & Sauer, R. T. (1984) Annu. Rev. Biochem. **53**, 293–321
Pabo, C. O., Kroratin, W., Jeffery, W. & Sauer, R. T. (1982) Nature (London) **298**, 443–447
Riggs, A. D., Reiness, G. & Zubay, G. (1971) Proc. Natl. Acad. Sci. U.S.A. **68**, 1222–1225
Rüther, U. (1980) Mol. Gen. Genet. **178**, 475–477
Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
Sauer, R. Yocum, R. Doolittle, R., Lewis, M. & Pabo, C. (1982) Nature (London) **298**, 447–451
Schaeffer, F., Kolb, A. & Buc, H. (1982) EMBO J. **1**, 99–105
Schevitz, R. W., Otwonowski, Z., Joachimiak, A., Lawson, C. L. & Sigler, P. (1985) Nature (London) **317**, 782–786
Schmitz, A. (1981) Nucleic Acids Res. **9**, 277–292
Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) Proc. Natl. Acad. Sci. U.S.A. **73**, 804–808
Sims, P. F. G., Minter, S. J., Stancombe, R. Gent, M. E., Andrews, J., Waring, R. B., Towner, P. & Davies, R. W. (1985) Biochimie **67**, 841–847
Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F. & Matthews, B. W. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 3097–3100
Steitz, T. A., Weber, I. T. & Matthew J. B. (1983) Cold Spring Harbor Symp. Quant. Biol. **47**, 419–426
Takahashi, M., Gronenborn, A. M., Clore, G. M., Blazy, B. & Baudras, A. (1982) FEBS Lett. **139**, 37–40
Takeda, Y., Ohlendorf, D., Anderson, W. & Mathews, B. (1983) Science **221**, 1020–1026
Ullman, A. & Danchin, A. (1983) Adv. Cyclic Nucleotide Res. **15**, 1–53
Weber, I. T. & Steitz, T. A. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 3973–3977
Wu, H. M. & Crothers, D. M. (1984) Nature (London) **308**, 509–513
Zoller, M. J. & Smith, M. (1984) DNA **3**, 479–488

Received 4 August 1986/8 October 1986; accepted 13 November 1986