Reprinted from *J. Mol. Biol.* (1990) **213**, 159–166

# Crystallization of *Escherichia coli* Catabolite Gene Activator Protein with its DNA Binding Site

## The Use of Modular DNA

**Steve C. Schultz, George C. Shields and Thomas A. Steitz**

J. Mol. Biol. (1990) 213, 159–166

# Crystallization of *Escherichia coli* Catabolite Gene Activator Protein with its DNA Binding Site

## The Use of Modular DNA

**Steve C. Schultz, George C. Shields† and Thomas A. Steitz**

*Department of Molecular Biophysics and Biochemistry and Howard Hughes Medical Institute Yale University, New Haven, CT 06511, U.S.A.*

*(Received 16 August 1989; accepted 28 December 1989)*

To obtain crystals of the *Escherichia coli* catabolite gene activator protein (CAP) complexed with its DNA-binding site, we have searched for crystallization conditions with 26 different DNA segments ≥28 base-pairs in length that explore a variety of nucleotide sequences, lengths, and extended 5′ or 3′ termini. In addition to utilizing uninterrupted asymmetric *lac* site sequences, we devised a novel approach of synthesizing half-sites that allowed us to efficiently generate symmetric DNA segments with a wide variety of extended termini and lengths in the large size range (≥28 bp) required by this protein. We report three crystal forms that are suitable for X-ray analysis, one of which (crystal form III) gives measurable diffraction amplitudes to 3 Å resolution. Additives such as calcium, *n*-octyl-β-D-glucopyranoside and spermine produce modest improvements in the quality of diffraction from crystal form III. Adequate stabilization of crystal form III is unexpectedly complex, requiring a greater than tenfold reduction in the salt concentration followed by addition of 2-methyl-2,4-pentanediol and then an increase in the concentration of polyethylene glycol.

### 1. Introduction

The catabolite gene activator protein (CAP‡), also known as the cAMP receptor protein (CRP), when complexed with its allosteric effector, cyclic AMP (cAMP), binds to specific DNA sequences within several *Escherichia coli* operons and either activates (Zubay *et al.*, 1970; Epstein *et al.*, 1975; de Crombrugghe & Pastam, 1978) or represses (Mouva *et al.*, 1981; Aiba, 1983) transcription from the adjacent promoter. The crystal structure of CAP with cAMP bound was solved at 2·9 Å (1 Å = 0·1 nm) resolution (McKay & Steitz, 1981; McKay *et al.*, 1982) and its co-ordinates refined at 2·5 Å resolution to an *R*-factor of 20·6% (Weber & Steitz, 1987). Comparison of the structure of CAP with that of *cro* repressor demonstrated a helix-turn-helix motif (Steitz *et al.*, 1982), a structure that is now known to be utilized by many DNA-binding proteins for sequence-specific interactions with DNA (Ohlendorf & Matthews, 1983; Pabo & Sauer, 1984). A model for the CAP–DNA complex was constructed in which *B*-DNA was oriented on the helix-turn-helix of CAP by electrostatic complementarity (Steitz *et al.*, 1982; Weber & Steitz, 1984). This model proposed that R180, E181, R185 and K188 could project from the recognition helix of CAP into the major groove of *B*-DNA and interact with functional groups on the edges of base-pairs in the GTGA sequence of the CAP binding site (Weber & Steitz, 1984). Studies of mutant CAP proteins show that indeed R180 and E181 are necessary for proper DNA binding, but R185 and K188 are not (Ebright *et al.*, 1984, 1987; Gent *et al.*, 1987). Studies of altered binding site sequences have demonstrated that the sequence GNGA is important for sequence specific binding (Jansen *et al.*, 1987).

The model for the CAP–DNA complex also proposed a modest bend in the DNA to produce interactions with 22 of the 24 bp of DNA that ethylation interference experiments suggested was the minimum size of the CAP binding site (Majors,

† Present address: Department of Chemistry, Lake Forest College, Lake Forest, IL, U.S.A.

‡ Abbreviations used: CAP, catabolite gene activator protein; CRP, cAMP receptor protein; cAMP, cyclic AMP; bp, base-pair(s); DMT, 5′-dimethoxytrityl; DTT, dithiothreitol; PEG, polyethylene glycol; MPD, 2-methyl-2,4-pentanediol.



0022–2836/90/090159–08 $03.00/0

© 1990 Academic Press Limited

1977). The anomalous mobility of CAP–DNA complexes in non-denaturing polyacrylamide gels has been interpreted to indicate that CAP does indeed bend DNA (Wu & Crothers, 1984). Subsequent experiments establish that a CAP binding site of at least 28 bp is required for full affinity (Liu-Johnson *et al.*, 1986). Calculations of the electrostatic charge potential on CAP show that large regions of positive potential exist on the sides of the CAP protein. Models of CAP–DNA complexes that place the negative phosphate backbone of DNA into these regions of positive potential suggest that CAP may bend its DNA binding site by as much as 150° (Warwicker *et al.*, 1987). Such a bend in the DNA appears to be necessary to achieve interactions between CAP and a binding site of 28 bp.

To determine the precise interactions of CAP with its DNA binding site, and to explore the nature and extent of CAP-induced distortion of the DNA, we have surveyed a large range of conditions and DNA sequences for crystallizing a CAP–DNA complex. Jordan *et al.* (1985) have demonstrated the importance of testing a variety of DNA lengths and extended 5′ and 3′ termini in crystallizing protein–DNA complexes. We have followed two strategies in designing the DNA sequences used in these experiments: one utilizes blunt-end sequences 28, 29 and 30 bp in length, and a second, novel approach employs half-site sequences that allow an efficient scan of symmetric DNA fragments with a wide variety of lengths and extended 5′ or 3′ termini. We report three crystal forms from three different DNA sequences that produce crystals suitable for X-ray analysis. One crystal form was obtained from a full-length 29 bp fragment and two crystal forms were obtained from half-site sequences, which should produce an effective length of 30 bp plus 5′-dG or 5′-dGC base overhangs. All three of these crystal forms diffract to better than 4·0 Å; crystals of CAP and the 30 base sequence with a 5′-dG overhang diffract to 3·0 Å when properly stabilized.

```
                    Full sites

28     AATTAATGTGAGTTAGCTCACTCATTAG
       TTAATTACACTCAATCGAGTGAGTAATC

29     ACTTTAATGTGAGTTAGCTCACTCATAAC
       TGAAATTACACTCAATCGAGTGAGTATTG

30     CAATTAATGTGAGTTAGCTCACTCATTAGG
       GTTAATTACACTCAATCGAGTGAGTAATCC

                    Half sites

16       GAAAAGTGTGACATAT
12       CTTTTCACACTG

17      CGAAAAGTGTGACATAT
13      GCTTTTCACACTG

18     GCGAAAAGTGTGACATAT
14     CGCTTTTCACACTG

19    CGCGAAAAGTGTGACATAT
15    GCGCTTTTCACACTG

20   GCGCGAAAAGTGTGACATAT
16   CGCGCTTTTCACACTG
```

**Figure 1.** DNA sequences used in crystallization experiments with CAP. The full-site sequences correspond to the CAP operator site in the *lac* operon except that the 29 sequence contains 3 base sequence changes on the 5′ end and 1 on the 3′ end. The half-site sequences correspond to the half of the CAP site in the *gal* operon that is proximal to the polymerase binding site (de Crombrugghe *et al.*, 1984).

## 2. Experimental

### (a) *Oligodeoxyribonucleotides*

Oligodeoxyribonucleotides were synthesized on an Applied Biosystems (Foster City, CA) model 380B DNA synthesizer on 10 $\mu$m columns using the manufacturer's chemicals and protocols. The 5′-dimethoxytrityl (DMT) protecting group was not removed after the last cycle. Deprotection of the bases was done according to the manufacturer's protocols. The oligodeoxyribonucleotides were purified by high-pressure liquid chromatography on a 2·54 cm Vydac $C_4$ column using a gradient of acetonitrite in 10 mM-triethylammonium acetate (pH 6·5) with the DMT protecting group remaining, and again after the DMT group was removed by treatment with 80% (v/v) acetic acid for 15 min. Sequences more than 20 nucleotides long were purified first by preparative polyacrylamide gel electrophoresis (with the DMT group removed) and then by high-pressure liquid chromatography. Extinction coefficients for the oligodeoxyribonucleotides were calculated using the formula given by Fasman (1975). Equimolar quantities of the 2 strands were mixed to a final concentration of 1 to 2 mM in 5 mM-sodium cacodylate (pH 7·4), 0·2 M-NaCl, 0·5 mM-EDTA. The strands were annealed by placing this solution at 90 °C and cooling to 4 °C in 30 to 60 min.

The sequences of the full-site oligodeoxyribonucleotides (see Fig. 1) are derived from the CAP binding site in the *lac* operon; the 29-mer differs from the *lac* operon by 3 bp changes on the 5′ end and 1 bp change on the 3′-end. The sequences of the half-site oligodeoxyribonucleotides (see Fig. 1) are derived from the half of the CAP binding site in the *gal* operon closest to the RNA polymerase site; this sequence precisely matches the consensus CAP binding site (de Crombrugghe *et al.*, 1984). These oligodeoxyribonucleotides were hybridized to their various complementary strands to produce duplex DNAs with single-stranded overhanging nucleotides of varying lengths. The labels given to these sequences are listed in Fig. 2.

### (b) *Protein*

In order to construct an expression vector that overproduces CAP at higher levels than plasmid pHA (Aiba *et al.*, 1982), the *Bam*HI–*Eco*RI fragment from this plasmid was cloned into pUC-19 and the resulting construct was transformed into *E. coli* DH1 (Hanahan, 1983). Cells containing this plasmid were grown to saturation (16 to 24 h) in 2× YT medium (Messing, 1983) containing 500 mg of ampicillin/l. CAP protein was purified by a

```
28-H    16      G------ATAT
        12      C------

29-1    16      G------ATAT
        13     GC------

30-1    16      G------ATAT
        14    CGC------

29-2    17     CG------ATAT
        12      C------

30-H    17     CG------ATAT
        13     GC------

31-1    17     CG------ATAT
        14    CGC------

32-1    17     CG------ATAT
        15   GCGC------

30-2    18    GCG------ATAT
        12      C------

31-2    18    GCG------ATAT
        13     GC------

32-H    18    GCG------ATAT
        14    CGC------

33-1    18    GCG------ATAT
        15   GCGC------

34-1    18    GCG------ATAT
        16  CGCGC------

32-2    19   CGCG------ATAT
        13     GC------

33-2    19   CGCG------ATAT
        14    CGC------

34-H    19   CGCG------ATAT
        15   GCGC------

35-1    19   CGCG------ATAT
        16  CGCGC------

34-2    20  GCGCG------ATAT
        14    CGC------

35-2    20  GCGCG------ATAT
        15   GCGC------

36-H    20  GCGCG------ATAT
        16  CGCGC------

A31-2   18A   ACG------ATAT
        13     GC------

C31-2   18C   CCG------ATAT
        13     GC------

T31-2   18T   TCG------ATAT
        13     GC------

LAC31-2 18L GCAATTAATGTGAGTTAA
        13L  GTTAATTACACTC
```

**Figure 2.** DNA segments obtained by combining appropriate strands of the half-site segments presented in Fig. 1. Variations in sequence for the 31-2 segment are listed.

procedure similar to that described by Ghosaini *et al.* (1988). Approximately 100 g of cells were lysed in 500 ml of 50 mM-potassium phosphate (pH 7·5), 0·2 M-NaCl, 0·5 mM-EDTA, 2 mM-DTT, 0·02% (w/v) sodium azide, by treatment with lysozyme and sonication. This mixture was centrifuged and the supernatant was brought to 30% saturated ammonium sulfate at 0°C. The supernatant of this mixture was collected and brought to 70% saturated ammonium sulfate. The pellet from this precipitation was collected and dissolved in 200 ml of 50 mM-potassium phosphate (pH 7·5), 0·5 mM-EDTA, 2 mM-DTT, 0·02% (w/v) sodium azide and dialyzed against the same buffer. The protein was loaded onto a 100-ml cAMP-agarose (Pharmacia) column and the column was washed with 1·5 l of 0·5 M-potassium phosphate, 0·1 mM-EDTA, 2 mM-DTT, 0·02% sodium azide. CAP was eluted with 0·5 mM-cAMP in 0·5 M-potassium phosphate, 0·1 mM-EDTA, 2 mM-DTT, 0·02% sodium azide. The protein was concentrated to 10 to 15 mg/ml in a Centriprep 30 (Amicon) spin concentrator and samples of 350 $\mu$l of this solution were passed through a Superose 12 HR 10/30 Pharmacia fast protein liquid chromatography column in 5 mM-Tris (pH 7·5), 0·2 M-NaCl, 0·1 mM-EDTA, 2 mM-DTT, 0·02% sodium azide. The protein was concentrated to 4 to 6 mg/ml if necessary in a Centriprep 10 Amicon spin concentrator. The yield was highly variable; 10 to 100 mg were obtained from 100 g of cells.

### (c) *Crystallization*

#### (i) *Chemicals*

Polyethylene glycol (PEG), 2-(*N*-morpholino)ethanesulfonic acid (Mes), *n*-octyl-$\beta$-D-glucopyranoside, spermine, spermidine, and adenosine 3′:5′-cyclic monophosphate (cAMP) were purchased from Sigma. Ammonium sulfate and Tris were purchased from Schwarz/Mann Biotech. 2-Methyl-2,4-pentandiol (MPD), gold label, was purchased from Aldrich. All of these chemicals were used without additional purification.

#### (ii) *Crystallizations*

Crystals were grown by hanging drop vapor diffusion unless otherwise stated. A typical crystallization contained 4 to 6 mg CAP/ml, a 1·5 molar excess of DNA, 50 mM-buffer, 0·2 M-NaCl, 0·1 M-$MgCl_2$, 2 mM-cAMP. In the hanging drop vapor diffusion experiments, the initial concentrations of salts, precipitants and buffers in the hanging drop were half those in the well, such that a 2-fold increase in concentration was expected for all components in the drop during the course of the experiment. The concentrations reported are the final conditions, i.e. those of the well solution. The amounts and types of precipitating agents used in the various crystallization experiments are given in Results. Crystallizations by dialysis were done in 20-$\mu$l microdialysis buttons with

Spectra-por 7 (Spectrum) dialysis membrane, molecular mass cutoff = 1000 Da.

(iii) *Space group determination and data collection*

The space group, cell dimensions and limits of diffraction were determined from precession photographs using a Charles Supper camera on a Rigaku RU-200PL rotating anode X-ray generator. Diffraction data were collected on a Xuong-Hamlin multiwire area detector system using a Rigaku RU-300 rotating anode generator as the X-ray source. The data were collected and reduced using software from the University of California at San Diego.

## 3. Results

### (a) *Crystallizations from polyethylene glycol*

All of the sequences listed in Figure 2 were used in hanging drop crystallization experiments with PEG as the precipitant. The conditions tested were 10% and 15% (w/v) PEG 3350, at pH values 5·0, 5·5, 6·0, 6·5, 7·0, 7·5 and 8·0, with either 0·2 M-NaCl or 0·2 M-NaCl, 0·1 M-$MgCl_2$. Crystalline material was obtained with most of the sequences (see Table 1) at pH 5 to 7 unless otherwise noted. Four of the sequences gave crystals that were clearly superior in shape and size to crystals from the other sequences. These four sequences were used in further crystallization experiments in which the pH, PEG and salts were varied to produce larger crystals. The space group, cell dimensions and limits of diffraction were determined for three of the crystal forms (see Table 2); the 28-H sequence diffracted too poorly (>15 Å) to determine the space group.

To test for the presence of DNA in crystal forms I, II and III, crystals of each were washed, dissolved and then analyzed on a $C_4$ Vydac HPLC column using a gradient of acetonitrile in 10 mM-triethyl ammonium acetate (pH 6·5). All of these crystal forms contained DNA. Accurate determination of the quantities of CAP protein in these crystals was not accomplished and, therefore, the precise ratio of CAP to DNA was not determined. However, the crystals did contain more than half of the DNA introduced into the hanging drop.

Crystals of the CAP-31-2 complex (which contains 30 bp duplex DNA and a 5′-dG overhang on each end) diffracted better and were more reproducible than those grown with the other sequences. We studied a wide variety of growth conditions for improving diffraction from these crystals; the optimal conditions for growing CAP-31-2 crystals are currently: 5 to 10% (w/v) PEG 3350, 0·2 M-NaCl, 0·1 M-$CaCl_2$, 0·05 M-Mes (pH 5 to 6), 2 mM-cAMP, 2 mM-spermine, 0·3% (w/v) *n*-octyl-β-D-glucopyranoside, 0·02% $NaN_3$, 2 mM-DTT in hanging drop vapor diffusion. The crystals grew in two to five days. Identical crystals of CAP-31-2 were obtained from PEG 3350, 8000 and 10,000; no crystals were obtained with PEG 400.

### (b) *Crystallizations from other conditions*

All attempts to crystallize the CAP–DNA complex with ammonium sulfate as the precipitant have been unsuccessful. The sequences tried were: 28-H, 31-2, 32-2, 28, 29 and 30 at pH values of 5·0, 5·5, 6·0, 6·5, 7·0, 7·5 and 8·0.

Crystallization was attempted with MPD for sequences 29, 31-2 and 32-2. Crystals were obtained only with the 32-2 sequence. These crystals diffracted to only >10 Å. Crystals of CAP-31-2 grown with 5% MPD added to a normal PEG crystallization also diffracted poorly (>6 Å).

Crystals of CAP–cAMP complexed with the 29, 31-2 and 32-2 sequences could also be grown by dialyzing from 0·5 M-NaCl/0·1 M-$CaCl_2$ to 0·02 M-NaCl/0·01 M-$CaCl_2$ at pH 5·0 to 7·0. Diffraction from these crystals was consistently more mosaic and to lower resolution (4·5 Å for CAP-31-2) than crystals grown from PEG in hanging drop vapor diffusion. Crystals of CAP-31-2 grown by dialysis were frequently long, thin and multiple, and therefore were often not useful for X-ray diffraction.

Crystals of CAP-32-2 grew best by dialysis of 6 mg CAP/ml and a 1·5 molar excess of 32-2 DNA in 5% PEG 3350, 0·2 M-NaCl, 0·1 M-$CaCl_2$, 0·05 M-Mes (pH 6), 0·02% $NaN_3$, 2 mM-DTT, against the same solution with 2 mM-cAMP added. The crystals grew as hexagonal needles in one to three days.

### (c) *Effect of salts and additives on crystal quality*

Crystals of CAP-31-2 were grown from PEG using the optimal conditions described above except with

**Table 1**
*Results of PEG crystallization experiments of CAP with half-site and full-site DNA sequences*

| DNA | NaCl | $MgCl_2$ |
|---|---|---|
| 28-H | Crystal form I | Crystal form I |
| 29-1 | Needles/rods | Plates |
| 29-2 | Plates/rods | Rods |
| 30-H | Plates/rods | Rods |
| 30-1 | Fibers/slabs | A.P.† |
| 30-2 | A.P. | A.P. |
| 31-1 | Plates | Rods |
| 31-2 | Needles | Crystal form III |
| 32-H | Rods/needles | Fibers |
| 32-1 | A.P. | A.P. |
| 32-2 | A.P. | Crystal form IV |
| 33-1 | Plates | Plates |
| 33-2 | Microcrystals | A.P. |
| 34-H | A.P. | Microcrystals (pH > 7) |
| 34-1 | A.P. | A.P. |
| 34-2 | Microcrystals | Rods |
| 35-1 | A.P. | A.P. |
| 35-2 | A.P. | A.P. |
| 36-H | Rods (pH > 7) | A.P. |
| A31-2 | N.D.‡ | Rods |
| C31-2 | N.D. | Needles/blocks |
| T31-2 | N.D. | Plates |
| LAC31-2 | N.D. | Plates |
| 28 | A.P. | Slabs |
| 29 | Plates | Crystal form II |
| 30 | A.P. | A.P. |

† A.P., amorphous precipitate.
‡ N.D., not determined.

**Table 2**
*X-ray analysis of CAP–DNA crystals suitable for diffraction*

| Crystal form: | I | II | III | IV |
|---|---|---|---|---|
| DNA | 28–H | 29 | 31–2 | 32–2 |
| Diffraction | >15 Å | 4·0 Å | 3·0 Å | 3·5 Å |
| Space group | N.D.† | $C2$ | $C222_1$ | $P6_{1,5}22$ |
| Cell dimensions | N.D. | $a = 155$ Å<br>$b = 78$ Å<br>$c = 123$ Å<br>$\beta = 96°$<br>$V_m = 3{\cdot}0$ Å$^3$/Da assuming 2 dimers/asymmetric unit | $a = 138{\cdot}0$ Å<br>$b = 152{\cdot}6$ Å<br>$c = 76{\cdot}0$ Å<br>$V_m = 3{\cdot}0$ Å$^3$/Da assuming 1 dimer/asymmetric unit | $a = b = 112{\cdot}7$ Å<br>$c = 136$ Å<br>$V_m = 3{\cdot}7$ Å$^3$/Da assuming 1 monomer/asymmetric unit |

† N.D., not determined.

$MgCl_2$, $CaCl_2$, $BaCl_2$ or $MgSO_4$ as the added salt. The crystals from $CaCl_2$ diffracted slightly better (3·5 Å) than those from $MgCl_2$ (4 Å) and crystallized from approximately 2% lower concentrations of PEG. Crystals from $BaCl_2$ or $MgSO_4$ diffracted slightly worse (4·5 Å) than those from $MgCl_2$ (4 Å).

Table 3 lists various chemicals added to PEG crystallizations of CAP with the 31-2 sequence. The limit of diffraction was noticeably improved by spermine and *n*-octyl-*β*-D-glucopyranoside. These additives also reduced the amount of diffuse scatter in the diffraction pattern.

### (d) *Stabilization conditions*

Crystals of CAP-31-2 mounted directly from the hanging drop exhibited variable cell dimensions and variable resolution limits of diffraction; we found that transferring these crystals from the drops in which they were grown into an appropriate stabilizing solution was essential. Hanging drops containing the crystals in 5 to 8% PEG, 0·2 M-NaCl, 0·1 M-$CaCl_2$, 0·05 M-Mes (pH 5 to 6), 2 mM-cAMP, 2 mM-spermine, 0·3% *n*-octyl-*β*-D-glucopyranoside, 2 mM-DTT, were first equilibrated by vapor diffusion against 3% PEG, 10 mM-$MgCl_2$, 20 mM-Mes (pH 5·75), 2 mM-cAMP, 2 mM-DTT, for 18 to 24 hours and the crystals were then transferred into this solution. Harvesting with solutions containing higher salt or PEG concentrations, or at temperatures below 18°C, resulted in cracked crystals. The crystals were then transferred to the following sequence of solutions: 5% PEG, 5% MPD; 5% PEG, 10% MPD; 10% PEG, 10% MPD; 15% PEG, 10% MPD; 20%PEG, 10% MPD; all of these solutions contained 10 mM-$MgCl_2$, 20 mM-Mes (pH 5·75), 2 mM-cAMP, 0·02% $NaN_3$, 2 mM-DTT. Omitting MPD from these transfers resulted in cracked crystals.

Crystals stabilized in these conditions diffracted slightly better (3·0 Å) than crystals mounted directly from hanging drops (3·5 Å). The cell dimensions of these stabilized crystals were also much more reproducible at $a = 138{\cdot}0(\pm 0{\cdot}5)$ Å, $b = 152{\cdot}6(\pm 0{\cdot}5)$ Å, $c = 76{\cdot}0(\pm 0{\cdot}5)$ Å than crystals mounted from drops, $a = 140$ to 141 Å, $b = 154{\cdot}4$ Å, $c = 75$ to 80 Å.

## 4. Discussion

Crystallization experiments for several protein–DNA complexes have defined DNA length and extended 5′ or 3′ termini as critical variables in growing crystals suitable for X-ray analysis (Jordan *et al.*, 1985; McClarin *et al.*, 1986). Attempts in this laboratory to crystallize the CAP–DNA complex (at low ionic strength) with DNA sequences 16 and 22 bp long resulted only in crystals of CAP·cAMP without DNA. Liu-Johnson *et al.* (1986) subsequently found that CAP requires DNA sequences longer than other operator proteins for maximum affinity. By using DNA sequences of 28 bp or longer, we have obtained crystals of the CAP–DNA complex.

**Table 3**
*Effect of chemical additives on diffraction from CAP-31-2 crystals grown from PEG in 0·2 M-NaCl/0·1 M-$MgCl_2$*

| Additive | Limit of diffraction (Å) | Comments |
|---|---|---|
| No additives | 4·0 | |
| 5% MPD | 6·0 | Increased diffuse scatter |
| 5% Glycerol | 6·0 | Increased diffuse scatter |
| 0·5% *n*-Octyl-*β*-D-glucopyranoside | 3·5 | Requires 2% higher PEG concentrations for crystallization |
| 2 mM-Spermine | 3·5 | Requires 2% lower PEG concentrations for crystallization |
| 2 mM-Spermidine | 3·5 | Requires 2% lower PEG concentrations for crystallization |
| 5% Methanol | 4·0 | No apparent differences from crystallizations without methanol |

**Table 4**
*Average intensities from data collection on CAP-31-2 crystals*

| Shell lower limit (Å) | Average merged intensity | Average $Y/\mathrm{sig}(Y)$† | Number of reflections (Refs) with $Y/\mathrm{sig}(Y)$ in given range | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0 | <2 | <5 | <10 | <20 | <40 | <60 | >60 | Refs | Obs. |
| 5·13 | 552·52 | 36·405 | 0 | 84 | 317 | 497 | 838 | 1778 | 1123 | 1089 | 5726 | 24,327 |
| 4·07 | 271·07 | 16·672 | 8 | 339 | 835 | 1092 | 1513 | 1447 | 350 | 54 | 5638 | 19,870 |
| 3·56 | 133·80 | 7·277 | 20 | 864 | 1810 | 1497 | 988 | 300 | 17 | 0 | 5496 | 16,582 |
| 3·23 | 65·33 | 3·499 | 67 | 1728 | 2482 | 894 | 176 | 2 | 0 | 0 | 5349 | 14,534 |
| 3·00 | 40·85 | 2·164 | 81 | 2734 | 2259 | 208 | 13 | 0 | 0 | 0 | 5295 | 13,235 |
| Total: | 217·90 | 13·538 | 176 | 5749 | 7703 | 4188 | 3528 | 3527 | 1490 | 1143 | 27,504 | 88,548 |

† $Y/\mathrm{sig}(Y)$ is the average merged intensity/sigma (average merged intensity).

### (a) *Effect of DNA parameters on crystal growth*

We have pursued two approaches to survey a large variety of DNA sequences, DNA lengths and overhanging bases in our attempts to obtain crystals of the CAP–DNA complex suitable for X-ray analysis. The first approach utilizes blunt-ended fragments 28, 29 and 30 bp long with asymmetric nucleotide sequences. These fragments maintain the natural sequence of the CAP binding site in the *lac* operon in an uninterrupted DNA segment. However, large-scale synthesis and purification of oligonucleotides of this length is difficult, which limits the number of such sequences that can be reasonably tested. The second approach employs half-site sequences with a 4 bp self-complementary overhang on the 3′ end that is able to self hybridize on the protein and provide a completely symmetric binding sequence. We synthesized ten oligonucleotides $\leq 20$ bases in length that could be mixed to generate 19 different double-stranded segments with symmetric zero, one or two base overhangs on the 5′ or 3′ end (see Figs 1 and 2). Note that this method produces only symmetric overhangs, whereas Jordan *et al.* (1985) also tested asymmetric overhangs.

The use of half-site sequences as described here is an important approach for systematically screening a wide variety of DNA lengths and overhanging nucleotides in the search for crystals of protein–DNA complexes suitable for X-ray diffraction. This method is particularly useful in the study of complexes with DNA segments larger than about 20 bp, since reliable purification of longer sequences becomes more difficult on the milligram scale required for crystallization experiments. This method also allows the use of dyad symmetric DNA without the problems of hairpin formation that accompany sequences that are internally self-complementary. In principle, utilization of modular DNAs as described here could be expanded to the assembly of any length oligomer to explore crystallization of proteins with very large DNA sequences, or for investigating large complexes that involve multiple binding sites and proteins such as those proposed for the *mal* operon (Raibaud *et al.*, 1989) or the Cro protein (Maniatis *et al.*, 1975).

Crystals were observed with most of the sequences used in these experiments (see Table 1). Four of the crystal forms were clearly superior in size and morphology to those obtained from the other sequences. Crystals of CAP-28H (form I) diffracted poorly (>15 Å) and were not pursued. Crystals of CAP with the 29 bp *lac* site (form II) produced very mosaic diffraction patterns and were difficult to reproduce; conditions for improving diffraction from these crystals are being investigated. Crystals of CAP-32-2 (form IV) are reproducible but tend to grow as needles with a thickness of only 0·1 mm. Crystals of CAP-31-2 (form III) are very reproducible, grow quite large (up to 1·0 mm × 0·5 mm × 0·5 mm), and diffract isotropically to 3·5 Å with very little mosaic spread. The latter crystal form is the one we have pursued.

To study the effect of protruding bases on crystal growth, we changed the 5′-dG nucleotide of the 31-2 sequence to dA, dC and dT. Different crystal morphologies occurred for each of these sequences; none was suitable for X-ray diffraction. To study the effect of base sequence on crystallization, we changed the nucleotide sequence from that of *gal* to that of the left half of the CAP binding site in the *lac* operon, while maintaining the length and 5′-dG overhang of the 31-2 sequence (see Fig. 2, LAC31-2). Once again the crystals were morphologically distinct from the CAP-31-2 crystals and were unsuitable for X-ray analysis. These results emphasize the importance of overhanging bases and DNA sequence on crystal growth, presumably due to their effect on packing in the crystal. The effect of the DNA sequence (as with the LAC31-2 segment) on packing could result from sequence specific differences in the proposed bend of the DNA in the complex.

### (b) *Effect of crystallization conditions on crystal quality*

A wide range of conditions for crystallizing the CAP-31-2 complex were explored. Divalent metal ions ($Mg^{2+}$, $Ca^{2+}$, or $Ba^{2+}$) and pH values between 5 and 6 were required for formation of these crystals; precipitation could be from PEG or by dialysis

to low salt. Crystals grown from PEG with $CaCl_2$ consistently diffracted better than those grown from other conditions. Spermine and *n*-octyl-β-D-glucopyranoside improved the limit of diffraction from crystals grown with $MgCl_2$, but not those grown with $CaCl_2$. However, these additives did reduce the amount of diffuse scatter and mosaic character in diffraction from crystals grown in PEG with $CaCl_2$.

### (c) *Stabilization of CAP-31-2 crystals*

Determining adequate conditions for harvesting and stabilizing the CAP-31-2 crystals proved to be rather difficult. Initial attempts to harvest the crystals into conditions similar to those in which they grew, resulted in poorer diffraction and extensive cracking of the crystals. A survey of precipitant concentrations, salts, pH and temperature determined that low concentrations of precipitant (0 to 3% PEG), low salt (10 mM-$MgCl_2$) and temperatures above 18°C were critical for maintaining the quality of the crystals. After harvesting of the crystals, transfers to higher concentrations of PEG resulted in improved diffraction, as did transfers to solutions containing MPD; however, such treatment generally resulted in cracked crystals. By transferring the crystals stepwise first to 10% MPD, 5% PEG and then stepwise to 10% MPD, 20% PEG at room temperature in 10 mM-$MgCl_2$, cracking was avoided, the crystals diffracted to higher angle (3·0 Å) and the mosaic character of the diffraction pattern was reduced. The beneficial effect of MPD on these crystals is surprising, since including MPD in the crystallization results in poorly diffracting crystals. This procedure also results in more reproducible cell dimensions with variations less than 0·5 Å; the *c* cell dimension varied as much as 6% in crystals mounted directly from the hanging drop.

### (d) *Conclusions*

The experiments described here strongly reinforce earlier conclusions (Jordan *et al.*, 1985; Anderson *et al.*, 1984) concerning the importance of DNA length and extended 5′ and 3′ termini on crystallization of protein–DNA complexes. Our search has produced no means of predicting which lengths and overhangs will provide crystals suitable for X-ray diffraction. Rather, our results emphasize the dramatic effect that subtle changes in DNA length, sequence and overhanging bases have on crystal formation. The use of half-site sequences as described here proved very effective in systematically screening a wide variety of DNA lengths and overhanging bases and thereby obtaining crystals suitable for X-ray diffraction.

The proposed bend of up to 150° presents a unique problem in crystal packing if end-to-end stacking is to occur as observed in all other crystals of operator–DNA complexes. Consequently, crystallization of the CAP–DNA complex may be especially sensitive to the length, sequence and termini of the DNA fragment. The three crystal forms described here may help provide interesting insights into each of these variables. Comparisons of the CAP-31-2 and CAP-32-2 structures will show how different DNA termini can pack in these crystals. The CAP-29 complex may provide information regarding the sequence dependent nature of the bend, both by comparisons with the other crystal forms (which contain the *gal* sequence) and by the asymmetry of the 29 sequence.

Currently, we are pursuing a solution for the structure of the CAP-31-2 complex. We are able reproducibly to collect high-quality data to 3·0 Å resolution (Table 4) and are attempting to determine phases using heavy-atom isomorphous replacement and molecular replacement techniques.

We thank Jonathan Friedman and Adrian Goldman for their helpful advice. We thank Aino Ruusala and Grace Sun for synthesis of the oligonucleotides, and Paul Raccuia and Ji Qin for their excellent technical assistance. This research was supported by an American Cancer Society postdoctoral fellowship grant PF-2823 to S.C.S. and a National Institutes of Health grant GM-22778 to T.A.S.

## References

Aiba, H., Fujimoto, S. & Ozaki, N. (1982). *Nucl. Acids Res.* **10**, 1345–1361.

Aiba, M. (1983). *Cell*, **32**, 141–149.

Anderson, J. E., Ptashne, M. & Harrison, S. C. (1984). *Nature (London)*, **309**, 327–331.

de Crombrugghe, B. & Pastam, I. (1978). In *The Operon* (Miller, J. H. & Reznikoff, W. S., eds), pp. 303–324, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

de Crombrugghe, B., Busby, S. & Buc, H. (1984). *Science*, **224**, 831–838.

Ebright, R. H., Cossart, P., Gieguel-Sanzey, B. & Beckwith, J. (1984). *Nature (London)*, **311**, 232–235.

Ebright, R. H., Kolb, A., Buc, H., Kunkel, T. A., Krakow, J. S. & Beckwith, J. (1987). *Proc. Nat. Acad. Sci., U.S.A.* **84**, 6083–6087.

Epstein, W., Rothman-Denes, L. B. & Hesse, J. (1975). *Proc. Nat. Acad. Sci., U.S.A.* **72**, 2300–2304.

Fasman, G. D. (1975). Editor of *Handbook of Biochemistry and Molecular Biology*, 3rd edit., vol. I, p. 589, CRC Press, Cleveland, OH.

Gent, M. E., Gronenborn, A. M., Davies, R. W. & Clore, G. M. (1987). *Biochem. J.* **242**, 645–653.

Ghosaini, L. R., Brown, A. M. & Sturtevant, J. M. (1988). *Biochemistry*, **27**, 5257–5261.

Hanahan, D. (1983). *J. Mol. Biol.* **166**, 557–580.

Jansen, C., Gronenborn, A. M. & Clore, G. M. (1987). *Biochem. J.* **246**, 227–232.

Jordan, S. R., Whitcombe, T. V., Berg, J. M. & Pabo, C. O. (1985). *Science*, **230**, 1383–1385.

Liu-Johnson, H.-N., Gartenberg, M. R. & Crothers, D. M. (1986). *Cell*, **47**, 995–1005.

Majors, J. E. (1977). Ph.D. thesis, Harvard University, Cambridge, MA.

Maniatis, T., Ptashne, M., Backman, K., Kleid, D., Flashman, S., Jeffrey, A. & Maurer, R. (1975). *Cell*, **5**, 109–113.

McClarin, J. A., Fredrick, C. A., Wang, B. C., Green, P., Boyer, H. W., Grable, J. & Rosenberg, J. M. (1986). *Science*, **234**, 1526–1541.

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.

Messing, J. (1983). *Methods Enzymol.* **101**, 28.

Mouva, R. N., Green, P., Nakamura, K. & Inouye, M. (1981). *FEBS Letters*, **128**, 186–190.

Ohlendorf, D. H. & Matthews, B. W. (1983). *Annu. Rev. Biophys. Bioeng.* **12**, 259–284.

Pabo, C. O. & Sauer, R. T. (1984). *Annu. Rev. Biochem.* **53**, 293–321.

Raibaud, O., Vidal-Ingigliardi, D. & Richet, E. (1989). *J. Mol. Biol.* **205**, 471–485.

Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F. & Matthews, B. W. (1982). *Proc. Nat. Acad. Sci., U.S.A.* **79**, 3097–3100.

Warwicker, J., Engelman, B. P. & Steitz, T. A. (1987). *Proteins: Struct. Funct. Genet.* **2**, 283–289.

Weber, I. T. & Steitz, T. A. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 3973–3977.

Weber, I. T. & Steitz, T. A. (1987). *J. Mol. Biol.* **198**, 311–326.

Wu, H.-M. & Crothers, D. M. (1984). *Nature (London)*, **308**, 509–513.

Zubay, G., Schwartz, D. & Beckwith, D. (1970). *Proc. Nat. Acad. Sci., U.S.A.* **66**, 104–110.

*Edited by R. Huber*