Identification of a contact between arginine-180 of the catabolite gene activator protein (CAP) and base pair 5 of the DNA site in the CAP–DNA complex

(protein–DNA interaction/helix–turn–helix motif/cyclic AMP receptor protein/Escherichia coli transcription)

XIAOPING ZHANG and RICHARD H. EBRIGHT*
Department of Chemistry and Waksman Institute, Rutgers University, New Brunswick, NJ 08855

Communicated by Jon Beckwith, March 30, 1990 (received for review January 26, 1990)


The Escherichia coli catabolite gene activator protein (CAP; also referred to as the cAMP receptor protein, CRP) is a sequence-specific DNA-binding protein involved in transcription regulation; CAP functions by binding to specific DNA sites located at or near promoters (1, 2). The three-dimensional structure of CAP has been determined to 2.5-Å resolution by x-ray diffraction analysis (3). The protein is a dimer of two identical subunits, each of which is 209 amino acids long and contains a helix–turn–helix DNA-binding motif (see refs. 4 and 5). A detailed model has been proposed for the structure of the CAP–DNA complex (refs. 6–10; Fig. 1). In the model, the CAP–DNA complex is twofold symmetric: one subunit of the CAP dimer interacts with one half of the DNA site; the other subunit of the CAP dimer interacts in a twofold symmetry-related fashion with the other half of the DNA site. One contact between an amino acid of the helix–turn–helix motif of CAP and a base pair of the DNA half-site has been identified experimentally (7, 11, 12); i.e., amino acid 2 of the recognition α-helix of CAP, Glu-181, has been shown to contact base pair 7 of the DNA half-site.

The model for the structure of the CAP–DNA complex predicts that amino acid 1 of the recognition α-helix of CAP, Arg-180, makes a specificity-determining contact with base pair 5 of the DNA half-site (refs. 6 and 7; Fig. 1B). The model predicts that the guanidinium side chain of Arg-180 makes hydrogen bonds with the guanine N7 atom and the guanine O6 atom of the canonical base pair, G-C, at base pair 5 of the DNA half-site. Note that only G-C (not A-T, C-G, or T-A) can make the two putative hydrogen bonds.

We have tested experimentally the role of Arg-180 in specificity at base pair 5 of the DNA half-site. Our approach was to replace Arg-180 by an amino acid unable to contact base pair 5, and, then, to ask whether this replacement affects specificity at base pair 5. This general approach has been designated the ‘‘loss-of-contact approach’’ and has been used successfully in investigation of amino acid–base pair contacts by CAP, lac repressor, λ repressor, λ Cro, AraC protein, the σ70 subunit of E. coli RNA polymerase, and the σ11 subunit of Bacillus subtilis RNA polymerase (12–20). To eliminate the ability of amino acid 180 to contact base pair 5, we have replaced Arg-180 by glycine, which has no side chain, and by alanine, which has a short, one-carbon side chain.

We have used site-directed mutagenesis to construct two substituted CAP variants: [Gly180]CAP and [Ala180]CAP. We have investigated the profiles of specificity of wild-type CAP, [Gly180]CAP, and [Ala180]CAP with respect to base pairs 5 and 7 of the DNA half-site. We have found that replacement of Arg-180 of CAP by glycine or alanine eliminates the ability of CAP to distinguish between canonical and noncanonical base pairs at base pair 5 of the DNA half-site. The effect is position-specific; replacement of Arg-180 by glycine or alanine does not eliminate the ability of CAP to distinguish between canonical and noncanonical base pairs at base pair 7 of the DNA site. These results indicate that Arg-180 of CAP controls specificity at base pair 5. We conclude, in agreement with the model in Fig. 1 (6–10), that Arg-180 of CAP makes a specificity-determining contact with base pair 5 in the CAP–DNA complex.

MATERIALS AND METHODS

Bacterial Strains. The E. coli K-12 strains constructed in this work (Table 1) are derivatives of strain XA102 (Δlac–proABX111 argEam metB ara rpoB nal Su–2; ref. 21). The crp and strA markers are from strain CA8445 (HfrH ∆crp-45 ∆cyo-854 strA thi; ref. 22). The fnr-1 and zci::Tn10 markers are from strain ECL323 (fnr-1 zci::Tn10 ∆lacU169 araD139 motA strA thi; ref. 23). The fnr-1 marker was included to prevent Fnr-dependent transcription of lacP1 (–68T–55A) (see ref. 24).

Strains XAE400, XAE451, XAE452, XAE453, XAE471, XAE472, and XAE473 were constructed as follows. Strain XAE14 (XA102 ∆crp-45 strA fnr-1 zci::Tn10 pHA5) was lysogenized with λ434plac5 and with derivatives of λ434plac5 having substitutions at base pairs –68 and –55, or base pairs –66 and –57, of lacP1 (Table 2); lysogens were identified by blue colony color on 5-bromo-4-chloro-3-indolyl β-D-galactoside indicator plates. The resulting lysogens were

Abbreviation: CAP, catabolite gene activator protein.

*To whom reprint requests should be addressed: Waksman Institute, Rutgers University, New Brunswick, NJ 08855.

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.
grown in liquid culture in the absence of ampicillin selection; isolates that had lost plasmid pHAS were identified by red colony color on maltose/tetrazolium indicator plates. The isolates utilized were established to be single-copy lysogens, based on segregation patterns and on levels of CAP-independent β-galactosidase expression.

**Bacteriophage.** A list of bacteriophage used or constructed in this work is presented in Table 2. M13mp2-lacP1(-68C; -55G) and M13mp2-lacP1(-68T; -55A) were constructed from M13mp2 by site-directed mutagenesis (see method in ref. 26). Derivatives of λh343plac5 having substitutions at base pairs −68 and −55 of lac were constructed by homologous recombination between λh343plac5-P11L62 and the appropriate derivatives of M13mp2 (see method in ref. 27). Recombinants were identified by light blue plaque color on 5-bromo-4-chloro-3-indolyl β-d-galactoside indicator plates.
and were verified by nucleotide sequence determination of the lac promoter region.

**Plasmids Encoding CAP and CAP Derivatives.** The plasmid encoding wild-type CAP (pHA5) has been described (28). Plasmids encoding [Gly180]CAP (pTK180G) and [Ala180]CAP (pTK180A) were constructed by use of site-directed mutagenesis to introduce substitutions into M13mp18-CRP (see method in ref. 26), followed by subcloning of the 3.5-kilobase BamHI-BamHI fragment into the BamHI site of plasmid pBR322 (29). For each plasmid, the nucleotide sequence of the complete crp structural gene and promoter was verified. Plasmids were constructed and maintained in strain CA8445 (HfrH Δcrp-45 Δacy-a-854 strA thi; ref. 22).

**In Vivo DNA Binding.** Experimentation. Seven E. coli K-12 “tester strains” were used: XAE400, XAE451, XAE452, XAE453, XAE471, XAE472, and XAE473 (Table 1). Each tester strain has two important components. (i) Δcrp-45, a deletion of the gene that encodes wild-type CAP (22). Δcrp-45 is &gt;5 kilobases long; it deletes all homology to the 3.5-kilobase crp insert of the plasmids used in this study (30). (ii) lacZ, the gene that encodes β-galactosidase. In tester strain XAE400, lacZ is placed under the control of the wild-type lacP1 DNA site for CAP. In the remaining six tester strains, lacZ is placed under the control of derivatives of the lacP1 DNA site for CAP that have G:C → A:T, G:C → C:G, and G:C → T:A substitutions at base pair 5 of each DNA half-site, or G:C → A:T, G:C → C:G, and G:C → T:A substitutions at base pair 7 of each DNA half-site (sequences in Fig. 2). The lacZ gene is present on a ϕ343lac5 prophase stably integrated into the bacterial chromosome at attL.

To analyze the profile of specificity of CAP variant X, the plasmid encoding X was introduced into each of the seven tester strains. Note that each of the resulting plasmid-bearing strains was identical except for two base pairs in lacP1.

To analyze the profile of specificity of CAP variant X, the plasmid encoding X was introduced into each of the strains (see Materials and Methods). For example, the set of strains constructed to analyze the specificity of wild-type CAP carried pHA5; these strains are designated XAE400/CRP, XAE451/CRP, etc.

Therefore, in each of the resulting plasmid-bearing strains the intracellular concentration of CAP was equal. For each plasmid-bearing strain, the differential rate of β-galactosidase synthesis was determined by the method of Miller (31). Cultures were grown in LB medium (31) containing ampicillin at 100 μg/ml. Data were corrected for background—i.e., for the differential rate of β-galactosidase synthesis in the absence of CAP or CAP variant.

Data reduction. Ratios of equilibrium dissociation constants, K_{D,1}/K_{D,2}, were calculated from the differential rates of β-galactosidase synthesis as described in ref. 10 (see also refs. 13 and 14). Briefly,

\[
\theta = Z/Z_M.
\]

where \(\theta\) denotes the fractional occupancy of lacP1 by CAP; \(Z\) denotes the measured differential rate of β-galactosidase synthesis; and \(Z_M\) denotes the maximum differential rate of β-galactosidase synthesis (equal to 5000 for the tester strains used, determined separately; see ref. 10). And

\[
\frac{K_{D,1}}{K_{D,2}} = \frac{\theta(1-\theta)}{\theta(1-\theta)}
\]

where \(K_{D,1}\) denotes the equilibrium dissociation constant for interaction of CAP variant X with the DNA site in strain 1, and \(K_{D,2}\) denotes the equilibrium dissociation constant for interaction of the identical CAP variant X with the DNA site in strain 2.

This method makes the assumption that, for a given fractional occupancy of lacP1, CAP variant X stimulates β-galactosidase synthesis to the same extent as does wild-type CAP. In most instances, this assumption is valid. In previous work, we have measured ratios of equilibrium dissociation constants exhibited by wild-type CAP and by more than 20 substituted CAP variants—both in vivo, using the method above, and in vitro, using the nitrocellulose filter binding assay (ref. 10; R.H.E., X.Z., A. Gunasekera, M. Smith, and T. Kunkel, unpublished results). We have found good agreement between the in vivo results and the in vitro results.

The method above yields high-precision data for \(\theta = 0.002\) to \(\theta = 0.95\) (precision of estimate of \(\theta\) typically within 10%). When \(\theta < 0.002\), the value \(\theta = 0.002\) was used to calculate a minimum estimate for the ratio \(K_{D,1}/K_{D,2+}\).

### RESULTS

We have measured the profiles of specificity of wild-type CAP, [Gly180]CAP, and [Ala180]CAP with respect to G:C, A:T, G:C, and T:A at base pairs 5 and 7 of the DNA half-site. Fig. 2 shows the DNA sites used as ligands in the analysis. They consist of the wild-type lacP1 DNA site for CAP, three derivatives of the lacP1 DNA site for CAP that have G:C → A:T, G:C → C:G, and G:C → T:A substitutions at base pair 5

Table 1. E. coli K-12 strains constructed in this study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>XAE400</td>
<td>XA102 Δcrp-45 strA fur-1 zci::Tn10 ϕ343lac5</td>
</tr>
<tr>
<td>XAE451</td>
<td>XA102 Δcrp-45 strA fur-1 zci::Tn10 ϕ343lac5P1(-68A;--55T)</td>
</tr>
<tr>
<td>XAE452</td>
<td>XA102 Δcrp-45 strA fur-1 zci::Tn10 ϕ343lac5P1(-68C;--55G)</td>
</tr>
<tr>
<td>XAE453</td>
<td>XA102 Δcrp-45 strA fur-1 zci::Tn10 ϕ343lac5P1(-68T;--55A)</td>
</tr>
<tr>
<td>XAE471</td>
<td>XA102 Δcrp-45 strA fur-1 zci::Tn10 ϕ343lac5P1(-66A;--57T)</td>
</tr>
<tr>
<td>XAE472</td>
<td>XA102 Δcrp-45 strA fur-1 zci::Tn10 ϕ343lac5P1(-66C;--57G)</td>
</tr>
<tr>
<td>XAE473</td>
<td>XA102 Δcrp-45 strA fur-1 zci::Tn10 ϕ343lac5P1(-66T;--57A)</td>
</tr>
</tbody>
</table>

### Table 2. Bacteriophage used in this study

<table>
<thead>
<tr>
<th>Bacteriophage</th>
<th>Source or ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ϕ343lac5</td>
<td>13</td>
</tr>
<tr>
<td>ϕ343lac5P1(-68A;--55T)</td>
<td>This work</td>
</tr>
<tr>
<td>ϕ343lac5P1(-68C;--55G)</td>
<td>This work</td>
</tr>
<tr>
<td>ϕ343lac5P1(-68T;--55A)</td>
<td>This work</td>
</tr>
<tr>
<td>ϕ343lac5P1(-66A;--57T)</td>
<td>10</td>
</tr>
<tr>
<td>ϕ343lac5P1(-66C;--57G)</td>
<td>10</td>
</tr>
<tr>
<td>ϕ343lac5P1(-66T;--57A)</td>
<td>10</td>
</tr>
<tr>
<td>M13mp2 CRP</td>
<td>25</td>
</tr>
<tr>
<td>M13mp2-lacP1(-68A;--55T)</td>
<td>T. Kunkel</td>
</tr>
<tr>
<td>M13mp2-lacP1(-68C;--55G)</td>
<td>This work</td>
</tr>
<tr>
<td>M13mp2-lacP1(-68T;--55A)</td>
<td>This work</td>
</tr>
<tr>
<td>M13mp18-CRP</td>
<td>10</td>
</tr>
<tr>
<td>M13mp18-180G</td>
<td>This work</td>
</tr>
<tr>
<td>M13mp18-180A</td>
<td>This work</td>
</tr>
</tbody>
</table>
of each DNA half-site, and three derivatives of the lacP1 DNA site for CAP that have G-C → A-T, G-C → C-G, and G-C → T-A substitutions at base pair 7 of each DNA half-site. The data were obtained using the quantitative in vivo DNA binding assay described in ref. 10 (see also refs. 13 and 14). In Tables 3 and 4, the data are expressed as the ratio \(K_D/K_{DP+}\), i.e., as the equilibrium dissociation constant for the interaction of CAP variant \(X\) with a substituted lacP1 DNA site, divided by the equilibrium dissociation constant for the interaction of the identical CAP variant \(X\) with the wild-type lacP1 DNA site.

**Specificity at Base Pair 5 of the DNA Half-Site.** Wild-type CAP exhibits robust specificity for G-C at base pair 5 of the DNA half-site (Table 3). Wild-type CAP interacts 200-fold, 2000-fold, and 30-fold more strongly with the wild-type lacP1 DNA site for CAP than it interacts with the derivatives of the lacP1 DNA site that have, respectively, A-T, C-G, and T-A at base pair 5 of each DNA half-site. These values correspond to specificity free energies of \(-3.2\) kcal/mol, \(-4.7\) kcal/mol, and \(-2.1\) kcal/mol.

In contrast, [Gly\(^{180}\)]CAP and [Ala\(^{180}\)]CAP exhibit essentially no specificity between G-C and A-T, C-G, or T-A at base pair 5 of the DNA half-site (Table 3). [Gly\(^{180}\)]CAP and [Ala\(^{180}\)]CAP interact essentially equally with each of the four DNA sites. (In fact, [Gly\(^{180}\)]CAP and [Ala\(^{180}\)]CAP exhibit a border-line-significant preference for the derivative of the lacP1 DNA site that has the noncanonical base pair T-A at base pair 5 of the DNA half-site.) These results demonstrate a significant change in the energetics of specificity at base pair 5 of the DNA half-site upon substitution of Arg-180.

**Specificity at Base Pair 7 of the DNA Half-Site.** As a control, we have asked whether the observed change in the energetics of specificity upon substitution of Arg-180 of CAP is position-specific. To do this, we have measured the profiles of specificity of wild-type CAP, [Gly\(^{180}\)]CAP, and [Ala\(^{180}\)]CAP with respect to G-C, A-T, C-G, and T-A at a second position within the DNA half-site, i.e., at base pair 7 of the DNA half-site.

Wild-type CAP exhibits robust specificity for G-C at base pair 7 of the DNA half-site (refs. 10, 12; Table 4). Wild-type CAP interacts >5000-fold, \(>5000\)-fold, and 700-fold more strongly with the wild-type lacP1 DNA site for CAP than it interacts with the derivatives of the lacP1 DNA site that have, respectively, A-T, C-G, and T-A at base pair 7 of each DNA half-site.

Within the limits of the data, [Gly\(^{180}\)]CAP and [Ala\(^{180}\)]CAP retain wild-type specificity (i.e., retain specificity for G-C) at base pair 7 of the DNA half-site (Table 4). [Gly\(^{180}\)]CAP and [Ala\(^{180}\)]CAP interact at least 1 order of magnitude more strongly with the wild-type lacP1 DNA site for CAP than with the derivatives of the lacP1 DNA site that have A-T, C-G, and T-A at base pair 7 of each DNA half-site. It has not been possible to quantify with more precision the specificities of [Gly\(^{180}\)]CAP and [Ala\(^{180}\)]CAP with respect to base pair 7 of the DNA half-site (due to the low DNA-binding affinities of [Gly\(^{180}\)]CAP and [Ala\(^{180}\)]CAP; R.H.E., A. Gunasekera, and P. S. Pendergrast, unpublished results). Nevertheless, the qualitative pattern is clear.

**DISCUSSION**

Replacement of Arg-180 of CAP by glycine or alanine eliminated specificity at base pair 5 of the DNA half-site but did not eliminate specificity at base pair 7 of the DNA half-site. These results indicate that Arg-180 of CAP controls specificity at base pair 5 of the DNA half-site. We conclude, in agreement with the model in Fig. 1 (6–10), that Arg-180 of CAP makes a specificity-determining contact with base pair 5 of the DNA half-site in the CAP–DNA complex.

The results do not exclude the possibility that Arg-180 of CAP controls specificity at base pair 5 of the DNA half-site indirectly—i.e., without a direct contact. For example, another protein functional group could make the contact with

---

**Table 3.** In vivo data: Recognition of DNA sites symmetrically altered at positions 5 and 18

<table>
<thead>
<tr>
<th>DNA site</th>
<th>Wild-type CAP</th>
<th>[Gly(^{180})]CAP</th>
<th>[Ala(^{180})]CAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Z)</td>
<td>(K_D/K_{DP+})</td>
<td>(Z)</td>
</tr>
<tr>
<td>5A;18T</td>
<td>260</td>
<td>200</td>
<td>230</td>
</tr>
<tr>
<td>5C;18G</td>
<td>25</td>
<td>2000</td>
<td>320</td>
</tr>
<tr>
<td>5T;18A</td>
<td>1100</td>
<td>30</td>
<td>980</td>
</tr>
</tbody>
</table>

P\(^{+}\) denotes the wild-type lacP1 DNA site; substituted derivatives of the lacP1 DNA site are identified by the position and sequence of the substitutions (sequences in Fig. 2). Tester strains XAE400, XAE451, XAE452, and XAE453 were used. Values of \(Z\) are presented for reference only and have been corrected for CAP-independent \(\beta\)-galactosidase expression.
base pair 5, Arg-180 being responsible solely for stabilizing the proper orientation of this other protein functional group. However, we consider this more complex possibility unlikely. In addition, the results do not exclude the possibility that Arg-180 of CAP contacts base pair 4 or base pair 6 of the DNA half-site in addition to base pair 5.

The results in this report regarding Arg-180 are complementary to the results in previous reports regarding Glu-181, the amino acid that contacts base pair 7 of the DNA half-site (7, 11, 12). Replacement of Glu-181 by glycine, alanine, valine, or leucine eliminates specificity at base pair 7 but does not eliminate specificity at base pair 5 (ref. 12; R.H.E., X.Z., A. Gunasekera, M. Smith, and T. Kunkel, unpublished results).

The identification of two amino acid–base pair contacts in the CAP–DNA complex—the contact by Arg-180 and the contact by Glu-181—suffices to define the orientation of the helix–turn–helix motif of CAP with respect to DNA in the CAP–DNA complex. The results indicate that amino acid 1 of the recognition α-helix of CAP (Arg-180) is farther from the twofold axis of the DNA site in the CAP–DNA complex than is amino acid 2 of the recognition α-helix of CAP (Glu-181). This orientation is as shown in the model in Fig. 1 (6–10). This orientation is inconsistent with the alternative model suggested by Müller-Hill and coworkers (34).

It will be of interest to construct the 17 remaining possible amino acid substitutions at amino acid 180 of CAP, and to assess the effects on DNA-sequence recognition at base pair 5 of the DNA half-site. We anticipate that this will provide insights regarding the chemistry of affinity and specificity.

We thank Dr. Thomas A. Kunkel and Ms. Mary Smith for training and assistance in site-directed mutagenesis, and Dr. Yon Ebright for oligonucleotide synthesis. This work was supported by National Institutes of Health Grant GM41376 to R.H.E.


Table 4. *In vivo* data: Recognition of DNA sites symmetrically altered at positions 7 and 16

<table>
<thead>
<tr>
<th>DNA site</th>
<th><em>Z</em></th>
<th><em>Kd/Kd,p</em></th>
<th><em>Z</em></th>
<th><em>Kd/Kd,p</em></th>
<th><em>Z</em></th>
<th><em>Kd/Kd,p</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>7A:16T</td>
<td>0</td>
<td>&gt;5000</td>
<td>0</td>
<td>&gt;8</td>
<td>1</td>
<td>&gt;20</td>
</tr>
<tr>
<td>7C:16G</td>
<td>0</td>
<td>&gt;5000</td>
<td>0</td>
<td>&gt;8</td>
<td>0</td>
<td>&gt;20</td>
</tr>
<tr>
<td>7T:16A</td>
<td>70</td>
<td>700</td>
<td>0</td>
<td>&gt;8</td>
<td>0</td>
<td>&gt;20</td>
</tr>
</tbody>
</table>

*P+* denotes the wild-type lacI gene; substituted derivatives of the lacI gene are identified by the position and sequence of the substitutions (sequences in Fig. 2). Tester strains XAE400, XAE471, XAE472, and XAE473 were used. Values of *Z* are presented for reference only and have been corrected for CAP-independent β-galactosidase expression.