DNA Sequence Determinants for Binding of the Escherichia coli Catabolite Gene Activator Protein*

Angelo Gunasekera, Yon W. Ebright, and Richard H. Ebright†

From the Department of Chemistry and Waksman Institute, Rutgers University, New Brunswick, New Jersey 08855

(Received for publication, November 26, 1991)

The consensus DNA site for binding of the Escherichia coli catabolite gene activator protein (CAP) is 22 base pairs in length and is 2-fold symmetric: 5'-AAATGTGATCAGATCACATTT-3'. Positions 4 to 8 of each half of the consensus DNA half-site are the most strongly conserved. In this report, we analyze the effects of substitution of DNA base pairs at positions 4 to 8, the effects of substitution of thymine by uracil and by 5-methylcytosine at positions 6, 7, and 8, and the effect of dam methylation of the 5'-GATC-3' sequence at positions 7 to 10. All DNA sites having substitutions of DNA base pairs at positions 4 to 8 exhibit lower affinities for CAP than does the consensus DNA site, consistent with the proposal that the consensus DNA site is the ideal DNA site for CAP. Specificity for T:A at position 4 appears to be determined solely by the thymine 5-methyl group. Specificity for T:A at position 6 and specificity for A:T at position 8 appear to be determined in part, but not solely, by the thymine 5-methyl group. dam methylation has little effect on CAP-DNA complex formation. The thermodynamically defined consensus DNA site spans 28 base pairs. All, or nearly all, DNA determinants required for maximal affinity for CAP and for maximal thermodynamically defined CAP-DNA ion pair formation are contained within a 28-base pair DNA fragment that has the 22-base pair consensus DNA site at its center. The quantitative data in this report provide base-line thermodynamic data required for detailed investigations of amino acid-base pair and DNA fragment that has the 22-base pair consensus DNA half-site, and the effects of substitution of DNA base pairs at positions 5, 7, and 8 of the consensus DNA half-site (12-18; Fig. 1). Two contacts between amino acids of the helix-turn-helix motif of CAP and DNA base pairs of the DNA half-site have been demonstrated experimentally (12-16): Arg-180 of CAP has been shown to contact the DNA base pair at position 5 of the DNA half-site, and Glu-181 of CAP has been shown to contact the DNA base pair at position 7 of the DNA half-site.

Important basic information regarding the chemistry and thermodynamics of protein-DNA complex formation can be obtained by measuring the effects of substitution of DNA base pairs, one by one, of the DNA site. Takeda and co-workers (19, 20) have termed this experimental approach "systematic base pair substitution analysis." Several studies of the DNA sequence determinants for binding of CAP have been reported previously. Ebright and co-workers (15, 17, 18) have examined the effects of substitution of DNA base pairs at position 7 of the E. coli lac DNA site for CAP and at positions 6 and 7 of the consensus DNA site for CAP. Jansen et al. (21) have examined, semiquanti-}

*This work was supported by National Institutes of Health Grant GM41376 (to R. H. E.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

†To whom correspondence should be sent: Waksman Institute, Rutgers University, New Brunswick, NJ 08855. Tel: 908-932-5179; Fax: 908-932-5735.

1The abbreviations used are: CAP, catabolite gene activator protein; 5-MeC, 5-methylcytosine; HPLC, high performance liquid chromatography; MOFS, 4-morpholinooethanesulfonic acid.

2In this report, positions within the DNA site for CAP are numbered as in Refs. 2, 3, 5-8, and 13-18. A different numbering convention is used in Ref. 12.
length. The results provide base-line thermodynamic information required for physical, biochemical, and genetic studies of amino acid-base pair and amino acid-phosphate contacts in this protein-DNA complex.

MATERIALS AND METHODS

CAP—CAP was the kind gift of Dr. J. Krakow (Hunter College of CUNY) and had been purified as described in Ref. 23. The fraction of CAP molecules active in sequence-specific DNA binding (0.64) was determined by titration of DNA fragment ICAP (6) under stoichiometric binding conditions; all data are reported in terms of molar concentrations of active CAP dimers.

Synthetic DNA Sites—40-base pair incompletely symmetric double-stranded DNA fragments containing the consensus DNA site for CAP, or substituted derivatives of the consensus DNA site for CAP (sequences in Figs. 2A, 3A, and 4), were prepared as described in Ref. 6. Uracil, 5-methylcytosine, and N6-methyladenine were introduced using, respectively, deoxyuridine-β-cyanoethlyphosphoramidite (ABN), 5-methyldeoxycytidine-β-cyanoethylphosphoramidite (ABN), and N6-methyldeoxyadenosine-β-cyanoethylphosphoramidite (Pharmacia LKB Biotechnology Inc.).

The 28-base pair incompletely symmetric double-stranded DNA fragment containing the consensus DNA site for CAP (DNA fragment ICAP28; sequence in Fig. 5) was prepared as follows. 28-residue oligodeoxyribonucleotides corresponding to the top and bottom strands were synthesized using solid-phase β-cyanoethylphosphoramidite chemistry on an Applied Biosystems 380A automated synthesizer. Products were deprotected and were purified using a two-step protocol, consisting of DEAE-5PW (TSK) anion-exchange HPLC followed by PRP-100 (Hamilton) reversed-phase HPLC. The self-complementary oligodeoxyribonucleotide (5 pmol) was purified from unincorporated [γ-32P]ATP and T4 polynucleotide kinase (24). The labeled top strand oligodeoxyribonucleotide was purified from unincorporated [γ-32P]ATP and T4 polynucleotide kinase (24). The labeled self-complementary oligodeoxyribonucleotide was purified from unincorporated [γ-32P]ATP and T4 polynucleotide kinase by chromatography on a Nensorb-20 mini-Kel-F column (Du Pont). The labeled self-complementary oligodeoxyribonucleotide (2.5 nmol) was annealed to yield double-stranded DNA in 10 μl of 25 mM Tris-HCl (pH 8.0) and 500 mM NaCl, by heating 10 min at 90 °C followed by gradual cooling (30–36 h) to 22 °C. Specific activities obtained ranged from 0.4 to 0.8 Bq/fmol.

Non-denaturing polyacrylamide gel electrophoresis followed by autoradiography indicated ≥95% duplex formation with the incompletely symmetric DNA fragments and 60–80% duplex formation with the completely symmetric DNA fragments. Under the reaction conditions utilized for the nitrocellulose filter binding experiments (see next section), the presence of 5–50% non-duplex DNA did not affect the measured equilibrium binding constants; therefore, annealed DNA fragments were used without further purification.

DNA concentrations were determined from absorbance at 260 nm; nucleotide extinction coefficients utilized were 2,200 M−1 cm−1 for single-stranded DNA and 6,500 M−1 cm−1 for double-stranded DNA (25).

Nitrocellulose Filter Binding Experiments—Filter binding experiments were performed as described in Refs. 6, 17, and 18. Experiments were performed using Schleicher & Schuell BA83 filter membranes (0.22 μm × 55 mm) and a Hoeffer filter manifold. Standard experiments were performed in 500 μl of buffer containing 10 mM MOPS/NaOH (pH 7.3), 200 mM NaCl, 0.1 mM dithiothreitol, 50 μg/ml bovine serum albumin, and 0.2 mM CAMP. For studies of salt dependence, the NaCl concentration of the assay buffer was varied to yield the desired monovalent cation concentration. 32P-Labeled DNA (1–30 pm) and CAP (5–50,000 pm; in >5-fold excess over the DNA concentration) were equilibrated at room temperature for 60 min and were then filtered in ≤10-15 s using suction through filters presoaked in assay buffer minus bovine serum albumin. Filters were dried, and Cerenkov radiation was determined in a Beckman LS5000TD scintillation counter. All data were corrected for background (i.e. radioactivity retained or the filter in the absence of CAP).

Representative data are presented in Fig. 6. Values of Kobs were extracted by nonlinear regression (Marquardt-Levenberg algorithm as implemented in SigmaPlot 4.0 (Jandel Scientific) using the equation

\[ B = \frac{B_{\text{max}}[\text{CAP}]_T}{(1/K_{\text{obs}}) + [\text{CAP}]_T} \]

where [CAP]T denotes the free CAP concentration, B denotes radioactivity retained on the filter, and B0 denotes radioactivity retained on the filter at saturation. [CAP]T and B were inputs to the nonlinear regression; Kobs and B0 were unconstrained outputs. \( \Delta G_{\text{obs}} \), the binding free energy for CAP-DNA complex formation, was calculated using the equation

\[ \Delta G_{\text{obs}} = -RT\ln(K_{\text{obs}}) \]

The difference between the binding free energy for CAP-DNA complex formation in the case of DNA site i versus the binding free energy for CAP-DNA complex formation in the case of the consensus DNA site, was calculated using the equation

\[ \Delta G_{\text{obs}} = \Delta G_{\text{obs}}(\text{DNA site i}) - \Delta G_{\text{obs}}(\text{DNA site consensus}) \]

High precision data were obtained where Kobs = \( 1 \times 10^4 \) M−1 to Kobs = \( 2 \times 10^2 \) M−1. Where Kobs < 1 \( \times 10^2 \) M−1, the value Kobs = \( 1 \times 10^4 \) M−1 was used to calculate a minimum estimate for \( \Delta G_{\text{obs}} \).

RESULTS

Effects of Substitution of DNA Base Pairs

We have performed equilibrium DNA binding experiments to investigate the effects of substitution of DNA base pairs at positions 4 to 8 of the consensus DNA half-site. We have used as ligands 40-base pair DNA fragments having the consensus DNA site (DNA fragment ICAP; Ref. 6) and 15 40-base pair DNA fragments having derivatives of the consensus DNA site with substitutions of DNA base pairs at positions 4 to 8 of each DNA half-site (sequences in Fig. 2A). The results obtained are presented in Table I and Fig. 2B.

Position 4—The consensus base pair at position 4 of the DNA half-site is T:A (3–5). CAP exhibits moderate specific-
DNA Sequence Determinants for Binding of CAP

TABLE I

<table>
<thead>
<tr>
<th>DNA site</th>
<th>$K_{ds}$</th>
<th>$\Delta G_{ds}$</th>
<th>$\Delta G_{ds}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>kcal/mol</td>
<td>kcal/mol</td>
<td>kcal/mol</td>
</tr>
<tr>
<td>ICAP-4A-SYM</td>
<td>7.0 $\pm$ 0.3 $\times 10^{9}$</td>
<td>14.64 $\pm$ 0.02</td>
<td></td>
</tr>
<tr>
<td>ICAP-4C-SYM</td>
<td>9.5 $\pm$ 0.7 $\times 10^{9}$</td>
<td>12.12 $\pm$ 0.04</td>
<td></td>
</tr>
<tr>
<td>ICAP-4G-SYM</td>
<td>1.5 $\pm$ 0.1 $\times 10^{10}$</td>
<td>13.73 $\pm$ 0.04</td>
<td></td>
</tr>
<tr>
<td>ICAP-5A-SYM</td>
<td>1.5 $\pm$ 0.2 $\times 10^{9}$</td>
<td>12.38 $\pm$ 0.07</td>
<td></td>
</tr>
<tr>
<td>ICAP-5C-SYM</td>
<td>3.0 $\pm$ 0.4 $\times 10^{7}$</td>
<td>10.09 $\pm$ 0.07</td>
<td></td>
</tr>
<tr>
<td>ICAP-5T-SYM</td>
<td>&lt;1.0 $\times 10^{7}$</td>
<td>&lt;9.45 $\times 10^{7}$</td>
<td></td>
</tr>
<tr>
<td>ICAP-6A-SYM</td>
<td>4.0 $\pm$ 1.0 $\times 10^{7}$</td>
<td>11.61 $\pm$ 0.13</td>
<td></td>
</tr>
<tr>
<td>ICAP-6C-SYM</td>
<td>5.3 $\pm$ 1.0 $\times 10^{7}$</td>
<td>11.78 $\pm$ 0.10</td>
<td></td>
</tr>
<tr>
<td>ICAP-6G-SYM</td>
<td>3.0 $\pm$ 0.7 $\times 10^{7}$</td>
<td>11.88 $\pm$ 0.06</td>
<td></td>
</tr>
<tr>
<td>ICAP-7A-SYM</td>
<td>1.0 $\times 10^{7}$</td>
<td>&lt;9.45 $\times 10^{7}$</td>
<td></td>
</tr>
<tr>
<td>ICAP-7C-SYM</td>
<td>2.0 $\times 10^{7}$</td>
<td>12.30 $\pm$ 0.12</td>
<td></td>
</tr>
<tr>
<td>ICAP-7T-SYM</td>
<td>9.3 $\times 10^{7}$</td>
<td>11.20 $\pm$ 0.11</td>
<td></td>
</tr>
<tr>
<td>ICAP-8A-SYM</td>
<td>2.0 $\times 10^{7}$</td>
<td>14.50 $\times 10^{7}$</td>
<td></td>
</tr>
<tr>
<td>ICAP-8C-SYM</td>
<td>3.3 $\pm$ 0.7 $\times 10^{7}$</td>
<td>11.50 $\pm$ 0.11</td>
<td></td>
</tr>
</tbody>
</table>

FIG. 2. A, DNA sites used to assess the effects of substitution of DNA base pairs. DNA sites are present on 40-base pair DNA fragments; sequences flanking the DNA sites are identical to the sequences flanking the E. coli lac DNA site for CAP (6). DNA fragment ICAP has the consensus DNA site for CAP (3-6). B, binding free energy changes ($\Delta G_{ds}$) upon substitution of DNA base pairs. The sequence in large characters corresponds to positions 4 to 8 of the consensus DNA half-site. Solid bars show the binding free energy changes upon the indicated substitutions. Binding free energy changes are for DNA sites substituted in both DNA half-sites.

Position 6—The consensus base pair at position 6 of the DNA half-site is T:A (3-5). CAP prefers T:A to nonconsensus base pairs by 1.4–1.5 kcal/mol/DNA half-site. The rank order of preference is T:A > A:T = G:C > C:G.

Position 7—The consensus base pair at position 7 of the DNA half-site is G:C (3-5). CAP exhibits strong specificity for G:C at position 7. CAP prefers G:C to nonconsensus base pairs by 2.4 to 2.6 kcal/mol/DNA half-site. The rank order of preference is G:C > T:A > A:T = G:C.

Position 8—The consensus base pair at position 8 of the DNA half-site is A:T (3-5). CAP exhibits moderate to strong specificity for A:T at position 8. CAP prefers A:T to nonconsensus base pairs by 1.2–1.7 kcal/mol/DNA half-site. The rank order of preference is A:T > G:C = C:G > T:A.

**Effects of Substitution of Thymine Residues by Uracil and by 5-Methylcytosine**

The thymine 5-methyl group projects into the DNA major groove (27, 28). The thymine 5-methyl group can be a critical determinant for specificity for A:T or T:A in protein-DNA interaction. It has been shown experimentally that the thymine 5-methyl group is the sole determinant of specificity for A:T at one position in the DNA half-site for the lac repressor (29) and that the thymine 5-methyl group is the sole determinant of specificity for T:A at one position in the DNA half-site for the lac repressor (30).

We have performed equilibrium DNA binding experiments to investigate the effects of substitution of thymine by uracil and by 5-methylcytosine. We have used as ligands a 40-base pair DNA fragment having the consensus DNA site (DNA fragment ICAP; Ref. 6), three 40-base pair DNA fragments having derivatives of the consensus DNA site with substitutions of thymine by uracil at positions 4, 6, or 8 of each DNA half-site (sequences in Fig. 3A), and three 40-base pair DNA fragments having derivatives of the consensus DNA site...
DNA Sequence Determinants for Binding of CAP

DNA half-site is T:A (3–5). CAP prefers T:A to U:A at position 4 of the DNA half-site by 0.7 kcal/mol/DNA half-site. The magnitude of the preference for T:A versus U:A at position 4 of the DNA half-site is comparable to the magnitude of the preference for T:A versus A:T, C:G, or G:C (0.5–1.3 kcal/mol/DNA half-site; Table I and Fig. 2B). These results indicate that the 5-methyl group is a determinant of specificity for T:A at position 4 of the DNA half-site.

CAP does not prefer T:A to 5-MeC:G at position 4 of the DNA half-site. The absence of a preference for T:A versus 5-MeC:G at position 4 of the DNA half-site is in marked contrast to the preference for T:A versus C:G (0.5 kcal/mol/DNA half-site; Table I and Fig. 2B). These results further indicate that the 5-methyl group is a determinant of specificity for T:A at position 4 of the DNA half-site.

It is striking that, within experimental error, the apparent incremental binding free energy contribution of the 5-methyl group is equal in the comparison of T:A versus U:T (0.7 kcal/mol/DNA half-site) and in the comparison of 5-MeC:G versus C:G (0.6 kcal/mol/DNA half-site). Taken together, the results indicate that the 5-methyl group is a determinant of specificity for T:A at position 6 of the DNA half-site.

Position 6—The consensus base pair at position 6 of the DNA half-site is T:A (3–5). CAP prefers T:A to U:A at position 6 of the DNA half-site by 1.0 kcal/mol/DNA half-site. The magnitude of the preference for T:A versus U:A at position 6 of the DNA half-site is comparable to the magnitude of the preference for T:A versus A:T, C:G, or G:C (1.4 to 1.5 kcal/mol/DNA half-site; Table I and Fig. 2B). These results indicate that the 5-methyl group is a determinant of specificity for T:A at position 6 of the DNA half-site.

However, CAP also prefers T:A to 5-MeC:G at position 6 of the DNA half-site, by 0.8 kcal/mol/DNA half-site. This result indicates that the 5-methyl group is not the sole determinant of specificity for T:A at position 6 of the DNA half-site. The apparent incremental binding free energy contribution of the 5-methyl group is significantly different in the comparison of T:A versus U:T (1.0 kcal/mol/DNA half-site) and in the comparison of 5-MeC:G versus C:G (0.6 kcal/mol/DNA half-site). Taken together, the results indicate that the role of the thymine 5-methyl group in specificity at position 6 of the DNA half-site is complex.

Position 8—The consensus base pair at position 8 of the DNA half-site is A:T (3–5). CAP prefers A:T to U:A at position 8 of the DNA half-site by 1.0 kcal/mol/DNA half-site. The magnitude of the preference for A:T versus U:A at position 8 of the DNA half-site is comparable to the magnitude of the preference for A:T versus C:G, C:C, or T:A (1.2 to 1.7 kcal/mol/DNA half-site; Table I and Fig. 2B). These results indicate that the 5-methyl group is a determinant of specificity for A:T at position 8 of the DNA half-site.

However, CAP also prefers A:T to G:5-MeC at position 8 of the DNA half-site, by 1.4 kcal/mol/DNA half-site. This result indicates that the 5-methyl group is not the sole determinant of specificity for A:T at position 8 of the DNA half-site. The apparent incremental binding free energy contribution of the 5-methyl group is significantly different in the comparison of A:T versus A:U (1.0 kcal/mol/DNA half-site) and in the comparison of G:5-MeC versus G:C (0.3 kcal/mol/DNA half-site). Taken together, the results indicate that the role of the 5-methyl group in specificity at position 8 of the DNA half-site is complex.
Effect of dam Methylation

In E. coli, 5'-GATC-3' sequences are methylated on both DNA strands at the N6 atom of adenine, by action of the E. coli dam methylase (22). dam methylation can have large effects on protein-DNA complex formation (31, 32). The consensus DNA site for CAP contains one 5'-GATC-3' sequence in each DNA half-site (positions 7 to 10 of the DNA half-site). We have performed equilibrium DNA binding experiments to investigate the effect of dam methylation of the consensus DNA site. We have used as ligands a 40-base pair DNA fragment having a derivative of the consensus DNA site with N6-methyladenine at positions 8 and 9 of each DNA half-site (Fig. 4). The results obtained are presented in Table III. The results indicate that CAP exhibits nearly equal affinities for the consensus DNA site and for the derivative of the consensus DNA site with N6-methyladenine at positions 8 and 9 of each DNA half-site.

Effects of DNA Fragment Length

Crothers and co-workers (33), using primer extension mapping and gel mobility shift experiments, have established that the E. coli lac DNA site for CAP spans 28-34 base pairs. CAP-induced DNA bending brings the outer segments of the DNA site into contact with the flanks of the CAP dimer, enabling the CAP dimer (≈55 Å in diameter) to contact an extended DNA site (≥95 Å in length if not bent; 12, 33, 34).

In this report, we have analyzed the effects of DNA fragment length on CAP-DNA complex formation with the consensus DNA site and have determined the minimum effective DNA fragment length for CAP-DNA complex formation with the consensus DNA site. We have used as ligands a 40-base pair incompletely symmetric DNA fragment having the consensus DNA site (DNA fragment ICAP; Ref. 6) and a 28-base pair incompletely symmetric DNA fragment having the consensus DNA site (DNA fragment ICAP28; sequence in Fig. 5). In the incompletely symmetric DNA fragments, the sequences flanking the consensus DNA site were identical to the sequences flanking the E. coli lac DNA site for CAP. In addition, we have used as ligands 30-, 28-, 26-, 24-, and 22-base pair completely symmetric DNA fragments having the consensus DNA site (sequences in Fig. 5). In the completely symmetric DNA fragments, the sequences flanking the consensus DNA site were identical to the upstream sequence flanking the E. coli lac DNA site for CAP.

Crothers and co-workers (33), using primer extension mapping and gel mobility shift experiments, have established that the E. coli lac DNA site for CAP spans 28-34 base pairs. CAP-induced DNA bending brings the outer segments of the DNA site into contact with the flanks of the CAP dimer, enabling the CAP dimer (≈55 Å in diameter) to contact an extended DNA site (≥95 Å in length if not bent; 12, 33, 34).

In this report, we have analyzed the effects of DNA fragment length on CAP-DNA complex formation with the consensus DNA site and have determined the minimum effective DNA fragment length for CAP-DNA complex formation with the consensus DNA site. We have used as ligands a 40-base pair incompletely symmetric DNA fragment having the consensus DNA site (DNA fragment ICAP; Ref. 6) and a 28-base pair incompletely symmetric DNA fragment having the consensus DNA site (DNA fragment ICAP28; sequence in Fig. 5). In the incompletely symmetric DNA fragments, the sequences flanking the consensus DNA site were identical to the sequences flanking the E. coli lac DNA site for CAP. In addition, we have used as ligands 30-, 28-, 26-, 24-, and 22-base pair completely symmetric DNA fragments having the consensus DNA site (sequences in Fig. 5). In the completely symmetric DNA fragments, the sequences flanking the consensus DNA site were identical to the upstream sequence flanking the E. coli lac DNA site for CAP.

**TABLE III**

<table>
<thead>
<tr>
<th>DNA fragment</th>
<th>( K_{\text{on}} )</th>
<th>( -\Delta G_{\text{on}} )</th>
<th>( \Delta G_{\text{on}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICAP</td>
<td>7.0 ± 0.3 x 10^6 kcal/mol</td>
<td>14.6 ± 0.02 kcal/mol</td>
<td>[-0.88 m⁺log[M⁺]]</td>
</tr>
<tr>
<td>ICAP-DAM</td>
<td>2.0 ± 0.2 x 10^7 kcal/mol</td>
<td>15.25 ± 0.06 kcal/mol</td>
<td>-0.6</td>
</tr>
</tbody>
</table>

**DISCUSSION**

**Specificity at Positions 4 to 8 of the DNA Half-site**

**Position 4**—Our results show that CAP exhibits moderate specificity for the consensus base pair T:A at position 4, and that this specificity is determined solely by the thymine 5-
methyl group. In the structure of the CAP-DNA complex, no amino acid of CAP forms H-bonds with the base pair at position 4 of the DNA half-site (12). We propose that specificity for T:A at position 4 results from either: (i) energetically favorable van der Waals interactions between CAP and the thymine 5-methyl group of T:A at position 4; (ii) effects of the thymine 5-methyl group of T:A at position 4 on the geometry of the DNA phosphates at position 4; (iii) effects of the thymine 5-methyl group of T:A at position 4 on the solvation of the DNA phosphates at position 4; or (iv) effects of the thymine 5-methyl group of T:A at position 4 on DNA twist or DNA bending. These mechanisms are not mutually exclusive. With respect to mechanism i, it is noteworthy that in the structure of the CAP-DNA complex the Cg of N1, and N2 atoms of the side chain of Arg-180 of CAP appear to be in van der Waals contact with the thymine 5-methyl group of T:A at position 4 (12). With respect to mechanisms ii and iii, it is noteworthy that in the structure of the CAP-DNA complex the side chain of Thr-168 of CAP and the peptide backbone NH groups of Arg-169 and Gln-170 of CAP form H-bonds with the top strand DNA phosphate 5' to position 4 and that the side chain of Arg-169 of CAP forms an ion pair with the top strand DNA phosphate 3' to position 4 (12; Fig. 1B). High resolution X-ray crystallographic investigations of B-DNA duplexes indicate that at T:A nucleotide pairs there is a site-bound water molecule immobilized between the thymine 5-methyl group and the top strand DNA phosphate 5' to the nucleotide pair; in contrast, at U:T, A:T, C:G, or G:C nucleotide pairs there is no equivalent site-bound water molecule (39, 40; see also Ref. 41). This suggests one detailed, although hypothetical, version of mechanism iii, as follows. Upon binding of CAP to a DNA site having T:A at position 4, formation of H-bonds between CAP and the top strand DNA phosphate 5' to position 4 displaces a site-bound water molecule from position 4; displacement of the site-bound water molecule results in an increase in entropy and a corresponding increment of binding free energy. Upon binding of CAP to DNA sites having U:T, A:T, C:G, or G:C at position 4 this increment of binding free energy is absent.

Position 4 is the first nucleotide pair within the six-nucleotide pair DNA segment in contact with the helix-turn-helix motif of CAP (positions 4 to 9 of the DNA half-site; Ref. 12). We point out that, like CAP, λ repressor and 434 repressor exhibit moderate specificity for T:A at the first nucleotide pair within the six-nucleotide pair DNA segment in contact with the helix-turn-helix motif (20, 42, 43) and that, like in the structure of the CAP-DNA complex, in the structure of the λ repressor-DNA complex and in the structure of the 434 repressor-DNA complex, no amino acid forms H-bonds with the base pair at the first nucleotide pair within the six-nucleotide pair DNA segment in contact with the helix-turn-helix motif (42, 44). It appears likely that CAP, λ repressor, and 434 repressor use similar or identical mech-

### Table IV

**Effects of DNA fragment length**

<table>
<thead>
<tr>
<th>DNA fragment</th>
<th>$K_{obs}$</th>
<th>$-\Delta G_{obs}$</th>
<th>$\Delta G_{o}$</th>
<th>Number of ion pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>m$^{-1}$ kcal/mol kcal/mol m$^{-1}$ m$^{-1}$, integer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICAP</td>
<td>7.0 ± 0.3 \times 10^{10}</td>
<td>14.64 ± 0.02</td>
<td>0.3</td>
<td>8</td>
</tr>
<tr>
<td>SICAP28</td>
<td>2.2 ± 0.2 \times 10^{10}</td>
<td>13.96 ± 0.05</td>
<td>0.5</td>
<td>7.8 ± 0.4</td>
</tr>
<tr>
<td>SICAP26</td>
<td>3.8 ± 0.8 \times 10^{9}</td>
<td>12.93 ± 0.04</td>
<td>1.0</td>
<td>6.8 ± 0.4</td>
</tr>
<tr>
<td>SICAP24</td>
<td>2.1 ± 0.4 \times 10^{9}</td>
<td>11.23 ± 0.10</td>
<td>2.7</td>
<td>ND</td>
</tr>
<tr>
<td>SICAP22</td>
<td>3.0 ± 0.4 \times 10^{8}</td>
<td>10.09 ± 0.07</td>
<td>3.9</td>
<td>ND</td>
</tr>
</tbody>
</table>

$^a$ Values of $K_{obs}$ and $-\Delta G_{obs}$ were calculated as described under “Materials and Methods” and are reported ± S.D.

$^b$ $m$ denotes the number of protein-DNA ion pairs formed upon protein-DNA complex formation. Values of $m$ were determined from the dependence of $\log K_{obs}$ on $-\log[Na^+]$ at constant pH and temperature (pH = 7.3; temperature = 23 °C) as described in Refs. 6 and 44-46 and are reported ± 1 S.E.
DNA Sequence Determinants for Binding of CAP

**Position 5**—Our results show that CAP exhibits strong specificity for the consensus base pair G:C at position 5. In the structure of the CAP-DNA complex, the side chain of Arg-180 of CAP forms two H-bonds with G:C at position 5 of the DNA half-site, i.e., one H-bond with the guanine N7 atom and one H-bond with the guanine O6 atom (12, 16). The specificity for G:C at position 5 is consistent with the formation of these two H-bonds. The rank order of preference among DNA base pairs at position 5 also is consistent with the formation of these two H-bonds (cf. Refs. 27 and 28). Model building suggests that the side chain of Arg-185 could form one H-bond with G:C at position 8 (with the guanine O6 atom, with unfavorable effects on solvation of the guanine N7 atom) but zero H-bonds with G:C or T:A at position 8.

In the structure of the CAP-DNA complex, the Orn1 atom of the side chain of Glu-181 of CAP appears to be in van der Waals contact with the thymine 5-methyl group of A:T at position 8 (12). The effect of the thymine 5-methyl group in specificity at position 8 may be related to this contact.

**Position 6**—Our results show that CAP exhibits moderate specificity for the consensus base pair T:A at position 6 and that this specificity is determined in part, but not solely, by the thymine 5-methyl group. In the structure of the CAP-DNA complex, no amino acid of CAP forms H-bonds with the base pair at position 6 of the DNA half-site (12, 18). We propose that specificity for T:A at position 6 results from either: (i) energetically favorable van der Waals interactions between CAP and the thymine 5-methyl group of T:A at position 6; (ii) sequence-dependent effects on the geometry of the DNA phosphates at position 6; (iii) sequence-dependent effects on the solvation of the DNA phosphates at position 6; or (iv) sequence-dependent effects on DNA twist or DNA bending. These mechanisms are not mutually exclusive. With respect to mechanisms ii and iii, it is noteworthy that in the structure of the CAP-DNA complex the side chain of Lys-188 of CAP forms an ion pair with the top strand DNA phosphate 5′ to position 6 (12, Fig. 1B). With respect to mechanism iv, it is noteworthy that in the structure of the CAP-DNA complex there is a large (ΔG = +40 kcal/mol) kink between positions 6 and 7 (12) and that the sequence 5′-TG-3′ has been shown to favor DNA flexibility and DNA bending (45-47).

**Position 7**—Our results show that CAP exhibits strong specificity for the consensus base pair G:C at position 7. In the structure of the CAP-DNA complex, the side chain of Glu-181 of CAP forms one H-bond with G:C at position 7 of the DNA half-site, i.e., one H-bond with the cytosine N3 atom (12-15). The strong specificity for G:C at position 7 is consistent with the formation of this H-bond. The rank order of preference among DNA base pairs at position 7 also is consistent with the formation of this H-bond (cf. Refs. 27 and 28). Model building suggests that the side chain of Glu-181 could form one H-bond with T:A at position 7 (with the adenine N7 atom, with unfavorable effects on solvation of the adenine N7 atom), but zero H-bonds with A:T or C:G at position 7.

**Position 8**—Our results show that CAP exhibits moderate to strong specificity for the consensus base pair A:T at position 8 and that this specificity is determined in part, but not solely, by the thymine 5-methyl group. In the structure of the CAP-DNA complex, the side chain of Arg-185 of CAP forms one H-bond with A:T at position 8 of the DNA half-site, i.e., one H-bond with the thymine O6 atom (12). The moderate to strong specificity for A:T at position 8 is consistent with the formation of this H-bond. The rank order of preference among DNA base pairs at position 8 also is consistent with the formation of this H-bond (cf. Refs. 27 and 28). Model building suggests that the side chain of Arg-185 could form one H-bond with C:G at position 8 (with the guanine O6 atom, with unfavorable effects on solvation of the guanine N7 atom) but zero H-bonds with G:C or T:A at position 8.

**Effect of dam Methylation**

Although the consensus DNA site for CAP contains one dam methylation site in each DNA half-site (positions 7 to 10 of the DNA half-site), our results show that dam methylation of the consensus DNA site has little effect on CAP-DNA complex formation. In the structure of the CAP-DNA complex (12), no residues of CAP are close to the adenine N6 atom (the target of dam methylation) at positions 8 and 9 of the DNA half-site. The absence of a large effect on binding of CAP of dam methylation of the consensus DNA site is consistent with the structure.

**Effects of DNA Fragment Length**

**CAP-DNA Complex Formation**—Our results indicate that the minimum effective DNA fragment length for binding of CAP to the consensus DNA site is 28 base pairs. All, or nearly all, DNA determinants required for maximal affinity for CAP are present in a 28-base pair DNA fragment having the 22-base pair consensus DNA site at its center. In contrast, not all DNA determinants required for maximal affinity for CAP are present in a 26-base pair DNA fragment having the 22-base pair consensus DNA site at its center. These results are in agreement with the results obtained by Crothers and co-workers (33) using the Escherichia coli DNA site for CAP.

In the structure of the CAP-DNA complex, all CAP-DNA contacts are to DNA determinants present in a 28-base pair DNA segment (12). Lys-26 of CAP contacts the bottom strand phosphate 5′ to position −5 of the DNA half-site (present in a 28-base pair DNA fragment but not present in shorter DNA fragments); Lys-166 of CAP contacts the bottom strand phosphate 5′ to position −2 of the DNA half-site (present in a 26-base pair DNA fragment but not present in shorter DNA fragments); and His-199 of CAP contacts the bottom strand phosphate 5′ to position −1 of the DNA half-site (present in a 24-base pair DNA fragment but not present in shorter DNA fragments). It appears likely that the difference in affinity between the 28- and 26-base pair DNA fragments having the consensus DNA site (0.5 kcal/mol) reflects the absence of the contact by Lys-26 of CAP in the case of the 28-base pair DNA fragment, that the difference in affinity between the 28- and 26-base pair DNA fragments having the consensus DNA site (1.7 kcal/mol) reflects the absence of the contact by Lys-166 of CAP in the case of the 24-base pair DNA fragment, and that the difference in affinity between the 24- and 22-base pair DNA fragments having the consensus DNA site (1.2 kcal/mol) reflects the absence of the contact by His-199 of CAP in the case of the 22-base pair DNA fragment.
the consensus DNA site (6). Our present results indicate that CAP likewise makes eight thermodynamically defined ion pairs with 30- and 28-base pair DNA fragments having the consensus DNA site. We conclude that the DNA phosphates that participate in the eight thermodynamically defined ion pairs in the case of the 40-base pair DNA fragment also are present in the 30- and 28-base pair DNA fragments but that at least one of these DNA phosphates is not present in the 26-base pair DNA fragment. However, we caution that end effects may complicate interpretation of salt dependence experiments with short DNA fragments (48, 49).

In the structure of the CAP-DNA complex as reported in Ref. 12, there are nine CAP-DNA ion pairs. The CAP-DNA ion pairs involve Lys-26 of one CAP subunit and Lys-166, Lys-188, Arg-169, and His-199 of both CAP subunits. Lys-26 of one CAP subunit forms an ion pair with the bottom strand phosphate 5′ to position −3 of one DNA half-site (present in a 28-base pair DNA fragment, but not present in shorter DNA fragments). It appears likely that the difference in the number of thermodynamically defined ion pairs between the 28- and 26-base pair DNA fragments having the consensus DNA site (eight versus seven) reflects the absence of the ion pair by Lys-26 of CAP in the case of the 26-base pair DNA fragment.

Acknowledgments—We thank Dr. J. Krakow for purified CAP.
We also thank Dr. Y. Takeda for important discussions.

REFERENCES