*Proc. Natl. Acad. Sci. USA*
Vol. 90, pp. 7548–7552, August 1993
Biochemistry

# Facilitated distortion of the DNA site enhances *Eco*RI endonuclease–DNA recognition

(protein–DNA interaction/energetics/base analogue/specificity/hydrogen bond)

DAVID R. LESSER†§, MICHAEL R. KURPIEWSKI†, TIMOTHY WATERS¶‖, BERNARD A. CONNOLLY**, AND LINDA JEN-JACOBSON†‡‡

†Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260; ¶Department of Biochemistry, The University of Southampton, Southampton SO9 3TU, United Kingdom; and **Department of Biochemistry and Genetics, The University, Newcastle upon Tyne, NE2 4HH, United Kingdom

*Communicated by Peter H. von Hippel, April 20, 1993*

ABSTRACT **We have measured the binding of *Eco*RI endonuclease to a complete set of purine-base analogue sites, each of which deletes one functional group that forms a hydrogen bond with the endonuclease in the canonical GAATTC complex. For five of six functional group deletions, the observed penalty in binding free energy is +1.3 to +1.7 kcal/mol. For two of these cases (replacement of adenine N7 with carbon) a single protein–base hydrogen bond is removed without deleting an interstrand Watson–Crick hydrogen bond or causing structural "adaptation" in the complex. This observation establishes that the incremental energetic contribution of one protein–base hydrogen bond is about −1.5 kcal/mol. By contrast, deletion of the N6-amino group of the inner adenine in the site improves binding by −1.0 kcal/mol because the penalty for deleting a protein–base hydrogen bond is outweighed by facilitation of the required DNA distortion ("kinking") in the complex. This result provides direct evidence that the energetic cost of distorting a DNA site can make an unfavorable contribution to protein–DNA binding.**

The highly selective recognition of the DNA sequence GAATTC by *Eco*RI endonuclease involves a number of interdependent contributions to specificity. The structure of endonuclease–DNA cocrystalline complexes (1, 2) shows that the endonuclease makes hydrogen bonds or nonpolar contacts with nearly all major-groove functional groups of the bases. The protein also makes symmetrical contacts to DNA phosphates on both strands at pNpGAApTTC. These contacts are indispensable to recognition of the canonical site (3, 4), and a subset (pNGAApTTC) contributes to discrimination against sites with one incorrect base pair (*Eco*RI* sites) by "adapting" in a stereotypical pattern (3), suggesting that formation of the canonical set of protein–phosphate contacts is sequence dependent.

Taking into account the favorable energetic contributions of the protein–base and protein–phosphate interactions and the large favorable contribution (5) of the "hydrophobic effect," we proposed (3) that the energetic cost of distorting the DNA in the complex makes a significant unfavorable contribution to the interaction with the correct DNA site and an even more unfavorable contribution to interaction with incorrect natural DNA sites.

In the canonical complex, the DNA adopts a pronounced torsional kink (2) in the middle of the recognition site that unwinds the DNA and widens the major groove to permit insertion of the recognition elements of the endonuclease (1, 2). This kinked geometry differs from that of the free DNA (6) in an exaggerated negative roll angle between the central base pairs (−57°; compared with free DNA −6°; J. M. Rosenberg, personal communication) such that these base pairs are completely unstacked; this alone may cost +8 kcal/mol (7). Other important features are an exaggerated positive roll angle between the second and third base pairs (+33°, free DNA +1°), an unusual stacking of the adenine bases (A3 on A2, Fig. 1, *Inset*), and the formation of a three-center (bifurcated) hydrogen bond (2) between the N6-amino group of A2 and the O4 atoms of thymines 4′ and 5′.

Direct evidence that DNA distortion can make an unfavorable contribution to the energetics of protein–DNA interaction has hitherto been lacking. To probe the functional role of the unusual local geometry at the central DNA kink, we have deleted the N6-amino groups of each of the adenine bases (A2 and A3, Fig. 1) by substitution with the synthetic base-analogue purine (P). (In purine, the 6-$NH_2$ group of adenine is replaced by a hydrogen atom.) Both amino groups donate hydrogen bonds (2) to the side-chain carbonyl of Asn-141, but deletion of the N6-amino group has markedly different effects at the two adenine positions.

Deletion of single hydrogen-bonding groups (on G1, A2, or the imidazole N7 of A3) penalizes binding free energy by +1.3 to +1.7 kcal/mol. In marked contrast, the substitution A3 → P *improves* protein–DNA binding by −1.0 kcal/mol. We infer that the favorable effect of A3 → P arises because the penalty for deleting a protein–base hydrogen bond is outweighed by facilitation of the required DNA distortion in the complex.

## MATERIALS AND METHODS

*Eco*RI endonuclease was prepared as described (8). All experiments used duplex 17-nt oligomers (Fig. 1, *Inset*) with the same flanking sequence used in our earlier studies (3, 9). Oligonucleotides containing base analogues were synthesized (10, 11), end-labeled, and purified (3). Melting temperatures ($t_m$) of oligonucleotides with one or two purines substituted for adenines were ≥65°C at 0.2 M KCl (D. Szwajkajzer, L.J.-J., and K. Breslauer, unpublished data). Equilibrium binding constants $K_A$ were measured at pH 7.5, 135 mM NaCl, 25°C (9). The dissociation rates of endonuclease–DNA complexes were measured as described (9), except that *Eco*RI endonuclease was at 10 nM and $^{32}$P-labeled oligonucleotide was at 12 nM. Single-turnover cleavage reactions were done and analyzed as described (9), except that $MgCl_2$ was present at 12 mM. Ethylation-interference footprinting was done as described (3, 4, 12).

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.

§Present address: U.S. Navy Drug Screening Laboratory, Oakland, CA 94627.
‖Present address: Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, United Kingdom.
‡‡To whom reprint requests should be addressed.

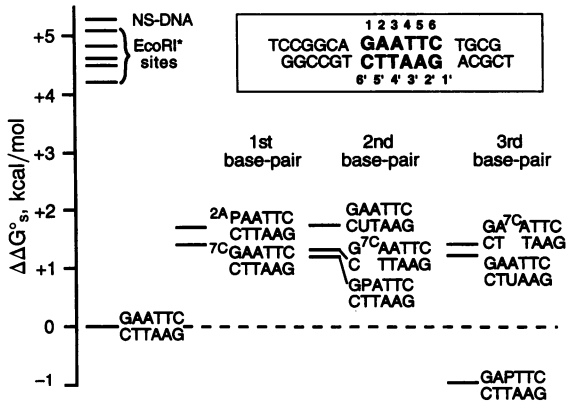


FIG. 1. Observed $\Delta\Delta G^\circ_s$ relative to the GAATTC site (- - -) for sites with one base analogue, imbedded in the 17-nt context (*Inset*). Values for *Eco*RI* sites, pyrimidine analogues, and "nonspecific" DNA (NS-DNA, the inverted site CTTAAG) are from ref. 3. P = purine; $^{7c}$A = 7-deazaadenine; $^{2A}$P = 2-aminopurine; $^{7c}$G = 7-deazaguanine.

## RESULTS AND DISCUSSION

**The Energetic Contribution of a Protein–Base Hydrogen Bond.** The rationale and detailed methods for measuring enzyme–substrate binding free energies $\Delta G^\circ_s$ from equilibrium association constants $K_A$ have been described (3). Changes are expressed as $\Delta\Delta G^\circ_s$ relative to the binding free energy for the canonical GAATTC site. $Mg^{2+}$ is omitted to measure $K_A$, but $Mg^{2+}$ neither induces major structural rearrangements in the complex (4) nor affects $K_A$ (3, 4).

We first consider the energetic contribution of each major-groove functional group on a purine base to endonuclease–DNA binding, measured with oligonucleotides containing a base-analogue substitution in only one strand of the recognition site. Each analogue removes one functional group (Table 1) that forms a hydrogen bond with the endonuclease in the canonical complex (2). We find a small penalty of +1.3 to +1.7 kcal/mol in five of six cases (Table 1, Fig. 1). (The exceptional case of A3 → P is considered below.) This value does not depend upon the hydrogen-bonding partner of each group (2), which may be an uncharged side chain (e.g., A2-N7 with Asn-141), a charged side chain (e.g., A3-N7 with Arg-145), or mediated by a water molecule (G1-N7 and G1-O6). The contribution of any one functional group cannot be determined from symmetrical double-strand substitutions (13–15) without unjustified assumptions.

Table 1. Binding parameters for single-analogue sites

| Site | Group deleted | $K_A$, $M^{-1}$ | $\Delta\Delta G^\circ_s$, kcal/mol |
|---|---|---|---|
| GAATTC | | 1.1 (±0.2) × $10^{10}$ | 0 |
| $G_1$ analogues | | | |
| $^{7c}$GAATTC | N7 imidazole | 1.0 (±0.2) × $10^9$ | +1.4 ± 0.2 |
| $^{2A}$PAATTC | O6 keto | 6.2 (±0.7) × $10^8$ | +1.7 ± 0.2 |
| $A_2$ analogues | | | |
| GPATTC | N6 amino | 1.3 (±0.3) × $10^9$ | +1.3 ± 0.2 |
| G$^{7c}$AATTC | N7 imidazole | 1.2 (±0.4) × $10^9$ | +1.3 ± 0.2 |
| $A_3$ analogues | | | |
| GAPTTC | N6 amino | 6.0 (±0.8) × $10^{10}$ | −1.0 ± 0.1 |
| GA$^{7c}$ATTC | N7 imidazole | 1.0 (±0.2) × $10^9$ | +1.4 ± 0.2 |

Each site, imbedded in a 17-nt oligomer (Fig. 1) with an unmodified complementary strand, was tested in both orientations, e.g., P in positions 3 or 3'. Values of $K_A$ for both orientations were in close agreement; means ± SD of ≥6 determinations on each orientation are shown. The difference in standard binding free energy between the unmodified site and each analogue site is $\Delta\Delta G^\circ_s = -RT \ln (K_A/K_A^{unmod})$.

The observed $\Delta\Delta G^\circ_s$ for any base substitution may include (3) changes in protein–base ($\Delta\Delta G_{base}$) and protein–phosphate interactions ($\Delta\Delta G_{phos}$), as well as a "reorganization" term ($\Delta\Delta G_{reorg}$) that includes any other changes in entropic (e.g., conformational) contributions to the interaction. The much larger energetic penalties for *Eco*RI* sites reflect unfavorable changes in all three terms (3). For most base-analogue sites, there is little change in the ethylation-interference footprint or in the salt dependence of binding (see below, Fig. 2*b*), suggesting that $\Delta\Delta G_{phos}$ is negligible for these sites.

It is usually impossible to allocate an observed $\Delta\Delta G^\circ_s$ among these components in any rigorous way, but the simplest cases are the sites that delete adenine-N7 (G$^{7c}$AATTC and GA$^{7c}$ATTC, Table 1), because these do not perturb interstrand Watson–Crick hydrogen bonding, do not alter protein–phosphate interactions, and have almost no effect (our unpublished data) on the cleavage-rate constant $k_1$ (see below). Each adenine N7 forms only a single hydrogen bond with a single protein side chain (2). We infer that the net $\Delta\Delta G^\circ_s$ in these cases most nearly represents the loss of a protein–base hydrogen bond, establishing that a single protein–base hydrogen bond contributes only about −1.5 kcal/mol to the overall binding free energy. This increment represents an exchange energy, in which each of the groups exchanges hydrogen bonds to water for hydrogen bonds to each other. Although other functional-group deletions (Table 1) may produce about the same net effect on $\Delta\Delta G^\circ_s$, there are other compensating changes (see below) in addition to deletion of a protein–base hydrogen bond.

**The Anomalous Favorable Effect of Deleting A3-N6.** Given the consistent energetic penalties for all other deletions of a hydrogen-bonding group on a purine base (Table 1, Fig. 1), one would have expected the same small penalty for deletion of the N6-amino group of A3, which forms a hydrogen bond (2) with Asn-141. It is, therefore, striking that the A3 → P substitution (GAPTTC) *improves* binding by −1.0 kcal/mol (Table 1, Fig. 1). The opposite effects of A2 → P and A3 → P substitutions are most dramatically demonstrated by the rates of dissociation of the complexes (Fig. 2*a*). Compared with the half-life of the canonical complex, that of the GPATTC complex is ≈4-fold shorter and that of the GAPTTC complex is ≈5-fold longer.

**Protein–Phosphate Contacts.** Protein–phosphate contacts in the GPATTC and GAPTTC complexes are nearly indistinguishable from those in the canonical complex by two criteria. (*i*) Ethylation-interference footprints (3, 4, 9, 12) for both purine-substituted sites are qualitatively and quantitatively similar to those on the canonical GAATTC site (data not shown). The only change was that the GAPTTC site showed no interference with binding at the scissile GpA in the unmodified strand, whereas the GAATTC site showed weak, but reproducible, interference at this point (3). Interference at GpA is generally absent for sites that are cleaved more rapidly than normal (9) and stronger for slowly cleaved sites (3). (*ii*) We determined the dependence of the dissociation rate constants $k_d$ upon monovalent cation concentration. The slope of a plot of log $k_d$ against log $[M^+]$ is related to the net thermodynamic equivalent number of counterions that recondense on DNA phosphates during complex dissociation (16, 17). The invariant slopes (6.5 ± 0.5) of these plots (Fig. 2*b*) imply that protein–phosphate contacts are little affected by purine substitution. By contrast, for complexes containing a single incorrect "natural" base pair, footprinting shows a loss of two important anchoring phosphate contacts in the substituted half-site, and a net of two fewer cations are released upon complex formation (3).

**Deletion of Multiple Protein–Base Interactions.** In most cases, sites with two-base analogues showed binding free energy changes ($\Delta\Delta G^\circ_s$), as predicted by the sums of $\Delta\Delta G^\circ_s$ values for the single-analogue sites, within experimental

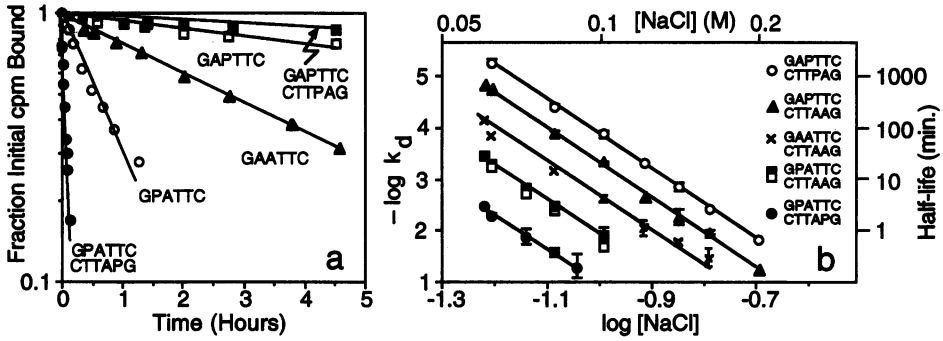


FIG. 2. (*a*) Kinetics of dissociation of endonuclease–DNA complexes at 60 mM NaCl. The long lifetimes here are a consequence of the low salt concentration (*b*). (*b*) Salt dependence of dissociation rate constants. Values are means ± SDs of three to five independent measurements; some error bars are too small to be seen at this scale. Data for the GAPTTC site are means ± SDs of both orientations in the 17-nt oligomer (Fig. 1). Both orientations of the GPATTC site are shown.

error. The A2 → P substitution (Table 2) elicits an approximately additive penalty in combination with a second analogue in the same half-site or the other half-site. We also find (our unpublished work) additive penalties with A2 → $^{7c}$A and A3 → $^{7c}$A substitutions in all combinations. Ethylation-interference footprints for these double-analogue sites are nearly identical to those for single-analogue sites and the unmodified GAATTC site (data not shown). The moderate penalties and absence of "adaptations" for these sites are in contrast to *Eco*RI* sites, where one incorrect base pair produces altered phosphate contacts and a large energetic penalty (3, 18); two incorrect base pairs have markedly nonadditive effects (3, 18).

Remarkably, the favorable effect of the A3 → P substitution can partly or completely compensate for the energetic cost of deleting other protein–base interactions (Table 2 and Fig. 3). The observed effects on $\Delta\Delta G^\circ_S$ are approximately additive (within experimental error), regardless of whether a second analogue is introduced in the same half-site or in the other half-site. (There are small deviations from the "expected" values for the $\frac{\text{GPPTTC}}{\text{CTTAAG}}$ and $\frac{\text{GAPTTC}}{\text{CTUAAG}}$ sites. These deviations are real, as indicated by dissociation-rate measurements similar to those in Fig. 2*a*, but we are reluctant to draw any inferences at this time.)

Table 2. Energetic additivity in double base-analogue sites

| Site | $\Delta\Delta G^\circ_S$, kcal/mol | |
|---|---|---|
| | Predicted | Observed |
| **$A_3$ → P** | | |
| G$^{7c}$APTTC<br>C TTAAG | +0.3 ± 0.3 | +0.4 ± 0.2 |
| GAPTTC<br>CTTPAG | −2.0 ± 0.2 | −1.7 ± 0.1 |
| GAPTTC<br>CUTAAG | +0.8 ± 0.2 | +1.0 ± 0.2 |
| GPPTTC<br>CTTAAG | +0.3 ± 0.2 | +0.7 ± 0.1 |
| GAPTTC<br>CTUAAG | +0.2 ± 0.2 | −0.2 ± 0.1 |
| **$A_2$ → P** | | |
| GPATTC<br>CUTAAG | +3.1 ± 0.3 | +3.0 ± 0.2 |
| GPATTC<br>CTTAPG | +2.6 ± 0.2 | +2.3 ± 0.1 |
| GP$^{7c}$ATTC<br>CT TAAG | +2.7 ± 0.2 | +2.7 ± 0.3 |
| GPATTC<br>CTUAAG | +2.5 ± 0.3 | +2.1 ± 0.2 |

Binding constants $K_A$ were determined, and $\Delta\Delta G^\circ_S$ was calculated, as described for Table 1. Predicted value is the sum of observed $\Delta\Delta G^\circ_S$ values for each of the constituent single-analogue sites (Table 1). Other sites and observed $\Delta\Delta G^\circ_S$ used in calculation were $\frac{\text{GAATTC}}{\text{CTUAAG}}$ (+1.2 ± 0.2) and $\frac{\text{GAATTC}}{\text{CUTAAG}}$ (+1.8 ± 0.2).

Some comparisons are particularly instructive: The GPPTTC site deletes two protein–base hydrogen bonds, yet binds −0.6 kcal/mol better than the GPATTC site that deletes only one. Note also that the $\frac{\text{GAPTTC}}{\text{CTUAAG}}$ site deletes two interactions from 1 bp, yet still binds slightly better ($\Delta\Delta G^\circ_S$ = −0.2 kcal/mol) than the unmodified GAATTC site.

Symmetrical purine substitution at positions 3 and 3′ increases the favorable effect to an approximately additive extent, implying that the favorable influence of A3 → P substitution can be realized twice in one complex. The effect on $k_d$ is also greater than for either single substitution (Fig. 2 *a* and *b*). Steady-state kinetic studies (13) using 10-bp oligonucleotides (lacking some phosphate contacts and only marginally in duplex form) failed to detect improved binding of the $\frac{\text{GAPTTC}}{\text{CTTPAG}}$ site [$K_M$ is a poor measure of DNA-binding affinity for this enzyme (3)]. The penalty for A2 substitution was nevertheless apparent in that A2,2′ → P reduced $k_{cat}/K_M$ more strongly (≈16-fold) than did A3,3′ → P (≈2-fold), by affecting $K_M^{DNA}$ but not affecting $k_{cat}$.

**Effects on Cleavage Rate Constants.** Among all the base-analogue sites (Table 1), only A3 → P substitution increases the cleavage-rate constant ($k_1$) in the unmodified GAA half-site (≈1.7-fold, Table 3, Fig. 4). Using $k_1 \times K_A$ as a metric to avoid strictly local substitution effects that influence $k_2$, we calculate that the transition-state interaction free energy (3) is more favorable (relative to that in the canonical complex) for the GAPTTC site ($\Delta\Delta G^{\circ\ddagger}_1 = -1.3$ kcal/mol). For other analogue sites, $k_1$ is, in some cases, unaffected (e.g., G$^{7c}$AATTC, GA$^{7c}$ATTC) and, in other cases, is inhibited (3) as much as 10-fold ($^{7c}$GAATTC). In most cases, cleavage in the modified half-site ($k_2$) is preferentially inhibited.

The GAATTC site is cleaved in both DNA strands at identical rates (3), but for the GAPTTC site $k_1/k_2 \approx 2$. This rate asymmetry is independent of orientation in the duplex: A3 → P stimulates cleavage of the G1′–A2′ bond, whereas A3′ → P stimulates cleavage of the G1–A2 bond (Fig. 4).

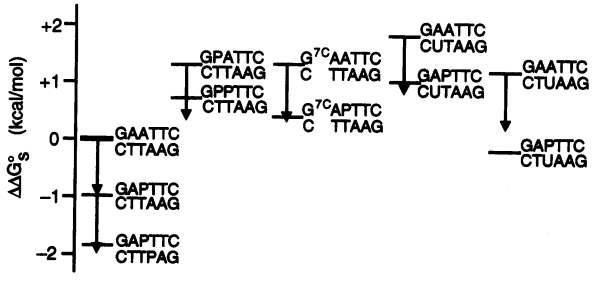


FIG. 3. Superimposition of the favorable effect of purine–3-substitution. Each arrow indicates the effect ($\Delta\Delta G^\circ_S$ = −1 kcal/mol) of single A3 → P substitution. Data are from Tables 1 and 2.

Table 3. Cleavage parameters for purine-substituted sites

| Site | $k_1$, $sec^{-1}$ | $\Delta\Delta G_1^{\circ\ddagger}$, kcal/mol | $k_2$, $sec^{-1}$ | $\Delta\Delta G_2^{\circ\ddagger}$, kcal/mol |
|---|---|---|---|---|
| GAATTC / CTTAAG | 1.1 ± 0.2 | 0 | 1.1 ± 0.2 | 0 |
| GPATTC / CTTAAG | 1.3 ± 0.2 | −0.1 ± 0.1 | 0.4 ± 0.1 | +0.6 ± 0.1 |
| GAPTTC / CTTAAG | 1.8 ± 0.2 | −0.3 ± 0.1 | 0.9 ± 0.1 | +0.1 ± 0.1 |
| GPATTC / CTTAPG | 1.0 ± 0.1 | +0.1 ± 0.1 | 1.2 ± 0.2 | −0.1 ± 0.1 |
| GAPTTC / CTTPAG | 1.4 ± 0.2 | −0.1 ± 0.1 | 1.3 ± 0.1 | −0.1 ± 0.1 |

Sites were imbedded in 17-nt oligomers (Fig. 1). For each single-strand substitution, the hexanucleotide was tested in both orientations—e.g., P in positions 3 or 3′. $k_1$ represents cleavage in the unmodified DNA strand, and $k_2$ represents cleavage in the modified DNA strand (3). Values of $k_1$ and $k_2$ for both orientations were in close agreement; thus the means ± SDs of the two orientations (six measurements) are presented. For symmetrical sites, $k_1$ represents the formation of 6-nt product, and $k_2$ represents the formation of 8-nt product (Fig. 3). These data were at 12 mM $Mg^{2+}$; values reported earlier (3, 9) were at 5 mM $Mg^{2+}$. The difference in standard free energies of activation is $\Delta\Delta G^{\circ\ddagger} = -RT \ln (k/k^{unmod})$, where subscripts 1 and 2 denote the cleavages characterized by $k_1$ and $k_2$.

Cleavage-rate asymmetry is a sensitive indicator of subtle "adaptation" in the complex (9) and here indicates that the endonuclease subtly "adapts" to the GAPTTC site to preferentially facilitate transition-state complementarity for cleavage in the unmodified half-site. When symmetry is restored (GAPTTC / CTTPAG, Table 3) the rate constants assume near-normal values, presumably because the complex cannot adapt to improve transition-state interactions in both half-sites simultaneously (9). ($k_1$ and $k_2$ are designated arbitrarily for symmetrical sites, where they do not differ significantly from each other.)

The complex that registers in our binding measurements and is seen in the crystal structure (1, 2) has already undergone major conformational changes in both DNA and protein (1, 2, 8) and is, thus, not a conventional "enzyme–substrate complex" but is a "pretransition-state" complex, poised to enter the transition state. $Mg^{2+}$ can be diffused into endonuclease–DNA crystals, and DNA cleavage occurs without

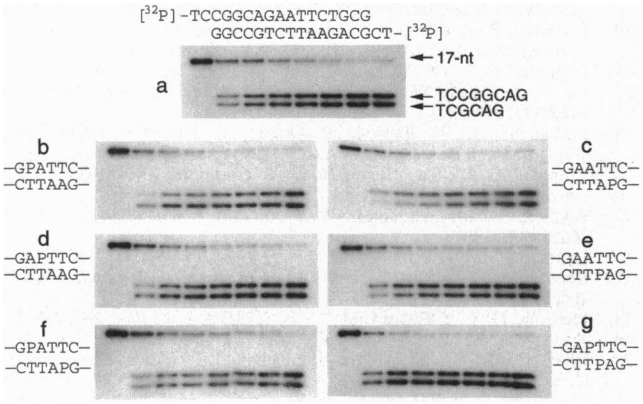


FIG. 4. Single-turnover cleavage reactions. Reaction times in seconds (left to right for each panel) were as follows: (*a*) 0, 0.9, 1.9, 2.6, 3, 4, 5, 6; (*b*) 0, 0.8, 1.5, 2.4, 3, 4, 5, 8; (*c*) 0, 1, 1.7, 2.1, 3, 4, 5, 8; (*d*) 0, 0.9, 1.1, 2, 2.7, 3.2, 4, 5; (*e*) 0, 0.9, 1.5, 2, 2.5, 3.1, 4, 5; (*f*) 0, 0.9, 1.8, 2.2, 3, 4, 5, 8; (*g*) 0, 0.9, 1.8, 2.2, 3, 3.5, 4, 5. Only a fraction of the complete time course (13 points) used for calculation (Table 3) is shown.

shattering the crystals (2), implying no *major* rearrangements from the pretransition state to transition state to enzyme–product complexes. Our data also indicate that the DNA conformation is very similar in the pretransition-state and transition-state complexes. For some base-analogue sites (e.g., $G^{7c}$AATTC) binding is inhibited, but the first-order cleavage rate constant is near-normal, so substitution has almost equal energetic effects on the pretransition-state and transition-state complexes (i.e., $\Delta\Delta G_S^\circ \approx \Delta\Delta G_1^{\circ\ddagger}$). This is also true for the GAPTTC / CTTPAG site, where a favorable effect on the energy of DNA distortion (see below) produces a large improvement in binding ($\Delta\Delta G_S^\circ = -1.7$ kcal/mol), but where $k_1$ and $k_2$ are normal ($\Delta\Delta G^{\circ\ddagger} \approx 0$, Table 3). This implies that the DNA conformational factors affected by A3,3′ → P substitution are the same in the pretransition-state and transition-state complexes.

**Purine Substitutions and the Energetics of DNA Distortion.** We previously postulated (3) that the energetic cost of distorting the free DNA into the required "kinked" conformation contributes significantly to the overall energetics of forming the canonical complex. The observation that A3 → P substitution enhances endonuclease binding, even while deleting a protein–base hydrogen bond, provides direct evidence that this contribution can be experimentally manipulated.

The favorable effect of A3 → P is obviously not attributable to $\Delta\Delta G_{base}$ alone. Because both the footprints and the salt dependence of binding (Fig. 2*b*) are little perturbed, $\Delta\Delta G_{phos}$ is probably negligible for the GAPTTC site. The observed $\Delta\Delta G_S^\circ$ (−1.0 kcal/mol, Table 1) is likely the net of an unfavorable contribution for deletion of the hydrogen bond between A3-N6 and Asn-141 (estimated at +1.5 kcal/mol) and a favorable effect on $\Delta\Delta G_{reorg}$, which we estimate to be ≈−2.5 kcal/mol.

We infer that A3 → P substitution reduces the unfavorable contribution to $\Delta\Delta G_{reorg}$ from distorting the DNA into the unique conformation found in the endonuclease–DNA complex. The central DNA "kink" has several geometric features that position the two adenines in each half-site, so that Asn-141 can make a "bridging" hydrogen bond between N6-amino groups on adjacent adenines (2). These features include not only the unusual roll angles of the A·T bp (Fig. 5) but also a particular stacking arrangement in which the vector connecting the C1′ atoms of A2 and A3 is nearly parallel to the helix axis (twist angle, ≈2°), whereas the twist angle between the C1′ atoms of the complementary bases (T5′ and T4′) is that of undistorted B-DNA (≈36°). This brings adjacent adenines (e.g., A2 and A3) into near-perfect overlap, as viewed down the helix axis.

The GAPTTC site deletes both the N6-amino group of A3 and a Watson–Crick hydrogen bond, so the following factors may contribute to the favorable effect:

(*i*) Relief of steric clash: The particular stacking arrangement of A2–A3 and the unusual roll angle of bp 2 and 3 bring the N6 atoms of A2 and A3 closer than the sum of their van der Waals radii (Fig. 5). This unfavorable clash is relieved in the GAPTTC site. The simplest expectation is that deletion of A2-N6 should also relieve crowding, but the GPATTC site did not show a comparable enhancement of binding (Table 1). However, A2 → P substitution (unlike A3 → P) prohibits formation of the bifurcated hydrogen bond T5′═O←(A2-N6)→O═T4′. We think it likely that the observed $\Delta\Delta G_S^\circ$ for A2 → P (+1.3 kcal/mol, Table 1) includes both a favorable effect from relief of the amino-group clash and an approximately compensating unfavorable contribution from loss of the bifurcated hydrogen bond, in addition to the unfavorable deletion of the hydrogen bond between protein and A2-N6.

(*ii*) Reduced constraint: In the canonical GAATTC complex (Fig. 5) the Watson–Crick hydrogen bond between

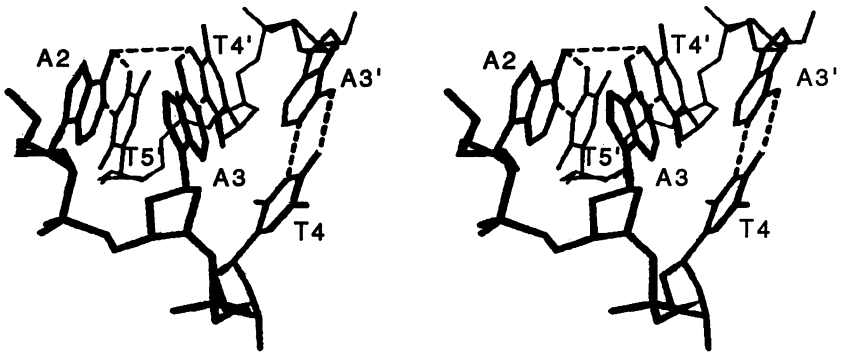


FIG. 5. Stereoscopic view of the structure of sequence AAT (left to right) in the endonuclease–DNA complex. - - -, Watson–Crick hydrogen bonds and the bifurcated hydrogen bond (top center) between A2 and T4'. This graphic was provided by J. M. Rosenberg, from a refinement of the crystal structure (1).

A3-N6 and T4'-O4 constrains A3 from adopting its unusual position, but in the GAPTTC site deletion of this hydrogen bond reduces constraint. We have found (M.R.K., B.A.C., and L.J.-J., unpublished work) an analogous, but smaller, improvement in binding ($\Delta\Delta G^\circ_S$ = −0.3 kcal/mol) when T4' is replaced by 5-methyl-2-pyrimidinone, presumably because deletion of two Watson–Crick hydrogen bonds more than compensates for the loss of a protein contact (2) to T4'-O4. This may also permit minor adjustment of the position of A3 to relieve steric clash. The T4'-O4 atom is not in steric conflict with another group, so the net favorable effect of deleting T4'-O4 is less than for deletion of A3-N6.

Both the steric and constraint factors involve bp 3 and its interactions with bp 2. The sequence AAT may have intrinsically low "deformability" that tends to exclude it from regions of DNA distortion (19), but in this case formation of the endonuclease–DNA interface produces favorable factors (direct contacts, counterion release, and water release from nonpolar surfaces) that compensate for the energetic cost of deformation. Any modification that enhances distortability improves the observed binding free energy.

The extreme negative roll angle at the central kink precludes interaction of the major-groove functional groups of A3·T4' with those of A3'·T4, so it is reasonable that the favorable effect of A3 → P can be superimposed (Table 2) on the effect of A3' → P substitution in the other half-site.

A structure of the $\overset{\text{GPATTC}}{\underset{\text{CTTAPG}}{\cdots\cdots}}$ site derived by NMR-restrained molecular dynamics (20) indicates that A2,2' → P has little effect on the free DNA conformation. The A2,2' → P and A3,3' → P substitutions have identical effects in reducing the curvature of the GAATTC site (21), yet we find they affect the endonuclease–DNA interaction in opposite directions. Thus, these measures of free DNA conformation or curvature may not be sensitive indicators of the potential for (or energetic cost of) adopting particular distortions (kinking, unwinding) required in a protein–DNA complex.

Position-specific influences on DNA conformation may also affect the interactions with other proteins. Substitutions of the central base pairs of the phage 434 operator influence repressor binding (22), even though they do not directly contact the protein (23). The T5' → U substitution has been reported (24) to improve $k_{cat}/K_M^{DNA}$ for *Eco*RI methylase.

**The Functional Role of DNA Distortion.** Whenever the conformation of protein-bound DNA differs from the lowest-energy conformation of the free DNA, the energy required to distort the DNA will contribute unfavorably to the protein–DNA interaction. The overall magnitude of this contribution probably cannot be determined experimentally because of the difficulty of deconvoluting it from the other energetic factors affecting complex formation.

For any DNA-binding protein, DNA distortion may be required to bring bases and/or particular phosphates into optimal geometry to interact with protein. For sequence-specific enzymes like restriction endonucleases, the distorted DNA conformation may lie on the pathway toward the transition state, so the energetic cost for distorting the DNA would have been borne in any event to attain the transition state and need not be paid again.

The observed effect of deleting even one functional group shows how local interactions of precisely placed elements contribute to the sequence dependence of forming a DNA–protein interface. The energetic cost of distorting the DNA to achieve protein–DNA complementarity will likely be greater for incorrect natural sites other than GAATTC, and this will contribute significantly to sequence discrimination (3). Our observations also open the way to designing oligonucleotide sites that improve upon the interactions of normal recognition sites with sequence-specific proteins.

We thank M. D. Barkley, P. J. Hagerman, and J. M. Rosenberg for critical reading of the manuscript. This work was supported, in part, by Grant GM-29207 from the National Institutes of Health. B.A.C. is a Research Fellow of the Lister Institute of Preventive Medicine.

1. Kim, Y., Grable, J. C., Love, R., Greene, P. J. & Rosenberg, J. M. (1990) *Science* **249**, 1307–1309.
2. Rosenberg, J. M. (1991) *Curr. Opin. Struct. Biol.* **1**, 114–122.
3. Lesser, D. R., Kurpiewski, M. R. & Jen-Jacobson, L. (1990) *Science* **250**, 776–786.
4. Becker, M. M., Lesser, D., Kurpiewski, M., Baranger, A. & Jen-Jacobson, L. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6247–6251.
5. Ha, J.-H., Spolar, R. S. & Record, M. T., Jr. (1989) *J. Mol. Biol.* **209**, 801–816.
6. Yanagi, K., Prive, G. G. & Dickerson, R. E. (1991) *J. Mol. Biol.* **217**, 201–214.
7. Delcourt, S. G. & Blake, R. D. (1991) *J. Biol. Chem.* **266**, 15160–15169.
8. Jen-Jacobson, L., Lesser, D. R. & Kurpiewski, M. (1986) *Cell* **45**, 619–629.
9. Lesser, D. R., Grajkowski, A., Kurpiewski, M. R., Koziolkiewicz, M., Stec, W. J. & Jen-Jacobson, L. (1992) *J. Biol. Chem.* **267**, 24810–24818.
10. Connolly, B. A. (1991) in *Oligonucleotides and Analogues: A Practical Approach*, ed. Eckstein, F. (IRL, New York), pp. 155–183.
11. Connolly, B. A. (1992) *Methods Enzymol.* **211**, 36–53.
12. Lu, A.-L., Jack, W. E. & Modrich, P. (1981) *J. Biol. Chem.* **256**, 13200–13206.
13. McLaughlin, L. W., Benseler, F., Graeser, E., Piel, N. & Scholtissek, S. (1987) *Biochemistry* **26**, 7238–7245.
14. Brennan, C. A., Van Cleve, M. D. & Gumport, R. I. (1986) *J. Biol. Chem.* **261**, 7270–7278.
15. Aiken, C. R., McLaughlin, L. W. & Gumport, R. I. (1991) *J. Biol. Chem.* **266**, 19070–19078.
16. Lohman, T. M. (1985) *CRC Crit. Rev. Biochem.* **19**, 191–245.
17. Record, M. T., Jr., Ha, J.-H. & Fisher, M. A. (1991) *Methods Enzymol.* **208**, 291–343.
18. Mossing, M. C. & Record, M. T., Jr. (1985) *J. Mol. Biol.* **186**, 295–305.
19. Travers, A. A. (1991) *Curr. Opin. Struct. Biol.* **1**, 114–122.
20. Clore, G. M., Oschkinat, H., McLaughlin, L. W., Benseler, F., Happ, C. S., Happ, E. & Gronenborn, A. M. (1988) *Biochemistry* **27**, 4185–4197.
21. Diekmann, S. & McLaughlin, L. W. (1988) *J. Mol. Biol.* **202**, 823–834.
22. Koudelka, G. B., Harbury, P., Harrison, S. C. & Ptashne, M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4633–4637.
23. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988) *Science* **242**, 899–907.
24. Reich, N. O. & Danzitz, M. J., Jr. (1992) *Biochemistry* **31**, 1937–1945.