(Cambridge Univ. Press, Cambridge, 1987), p. 209.
39. R. D. Blandford, C. S. Kochanek, I. Kovner, R. Narayan, *Science* **245**, 824 (1989); E. L. Turner, *Ann. N.Y. Acad. Sci.* **571**, 319 (1989).
40. N. T. Roseveare, *Mercury's Perihelion from Le Verrier to Einstein* (Clarendon Press, Oxford, 1982).
41. I. I. Shapiro, C. C. Counselman III, R. W. King, *Phys. Rev. Lett.* **36**, 555 (1976).
42. R. H. Dicke and H. M. Goldenberg, *Astrophys. J. Suppl. Ser.* **27**, 131 (1974); R. H. Dicke, *Science* **184**, 419 (1974).
43. C. M. Will, *Phys. Rep.* **113**, 345 (1984).
44. T. M. Brown *et al.*, *Astrophys. J.* **343**, 526 (1989).
45. For review and references see (*8, 43*), and C. M. Will, in *300 Years of Gravitation*, S. W. Hawking and W. Israel, Eds. (Cambridge Univ. Press, Cambridge, 1987), p. 80.
46. R. A. Hulse and J. H. Taylor, *Astrophys J.* **195**, L51 (1975).
47. J. H. Taylor, L. A. Fowler, P. M. McCulloch, *Nature* **277**, 437 (1979).
48. A. Einstein, *Sitzungsber. Preuss. Akad. Wiss.* **1916**, 688 (1916).
49. For a review of the quadrupole approximation for gravitational radiation reaction, see T. Damour, in *300 Years of Gravitation*, S. W. Hawking and W. Israel, Eds. (Cambridge, Univ. Press, Cambridge, 1987), p. 128.
50. J. H. Taylor and J. M. Weisberg, *Astrophys. J.* **345**, 434 (1989).
51. T. Damour and J. H. Taylor, *ibid.*, in press.
52. C. M. Will and D. M. Eardley, *Astrophys. J.* **212**, L91 (1977); C. M. Will and H. W. Zaglauer, *ibid.* **346**, 366 (1989).
53. C. W. F. Everitt *et al.*, in *Near Zero: New Frontiers of Physics*, J. D. Fairbank, B. S. Deaver, Jr., C. W. F. Everitt, P. F. Michelson, Eds. (Freeman, New York, 1988), p. 587.
54. I. Ciufolini, *Int. J. Mod. Phys. A* **4**, 3083 (1989).
55. B. Mashhoon, H. J. Paik, C. M. Will, *Phys. Rev. D* **39**, 2825 (1989).
56. P. L. Bender, N. Ashby, M. A. Vincent, J. M. Wahr, *Adv. Space Res.* **9** (no. 9), 113 (1989).
57. R. D. Reasenberg, R. W. Babcock, J. F. Chandler, I. I. Shapiro, in *Proceedings of the International Symposium on Experimental Gravitational Physics*, P. Michelson, H. En-ke, G. Pizzella, Eds. (World Scientific, Singapore, 1988), p. 3.
58. P. W. Worden, in *Near Zero: New Frontiers of Physics*, J. D. Fairbank, B. S. Deaver, Jr., C. W. F. Everitt, P. F. Michelson, Eds. (Freeman, San Francisco, 1988), p. 766.
59. R. F. C. Vessot, *Adv. Space Res.* **9** (no. 9), 21 (1989).
60. K. S. Thorne, in *300 Years of Gravitation*, S. W. Hawking and W. Israel, Eds. (Cambridge Univ. Press, Cambridge, 1987), p. 330.
61. Supported in part by NSF grant PHY 89-22140.

Research Articles

The Energetic Basis of Specificity in the Eco RI Endonuclease–DNA Interaction

DAVID R. LESSER,* MICHAEL R. KURPIEWSKI, LINDA JEN-JACOBSON†

High sequence selectivity in DNA-protein interactions was analyzed by measuring discrimination by Eco RI endonuclease between the recognition site GAATTC and systematically altered DNA sites. Base analogue substitutions that preserve the sequence-dependent conformational motif of the GAATTC site permit deletion of single sites of protein-base contact at a cost of +1 to +2 kcal/mol. However, the introduction of any one incorrect natural base pair costs +6 to +13 kcal/mol in transition state interaction energy, the resultant of the following interdependent factors: deletion of one or two hydrogen bonds between the protein and a purine base; unfavorable steric apposition between a group on the protein and an incorrectly placed functional group on a base; disruption of a pyrimidine contact with the protein; loss of some crucial interactions between protein and DNA phosphates; and an increased energetic cost of attaining the required DNA conformation in the transition state complex. Eco RI endonuclease thus achieves stringent discrimination by both "direct readout" (protein-base contacts) and "indirect readout" (protein-phosphate contacts and DNA conformation) of the DNA sequence.

PROTEINS THAT INTERACT WITH PARTICULAR TARGET SEquences in DNA may show sequence selectivities ranging from stringent to fairly permissive, depending on the requirements imposed by their functions. Extreme selectivity is exemplified by restriction endonucleases, which must efficiently cleave small (4 to 6 base pairs) recognition sites on foreign DNA, but must avoid potentially lethal cleavage of the cellular genome at sites that differ by as little as one base pair. By contrast, gene-regulatory proteins bind at larger sites (12 to 30 bp) and some bind a series of related sites in a graduated fashion (*1, 2*).

Sequence specificity is determined in part by protein contacts to the DNA bases (direct readout). Structural studies of the Eco RI endonuclease–DNA complex show that both strands of the GAATTC site are recognized by hydrogen bonds with each purine base (*3, 4*) and contacts to the pyrimidines (*4*). However, the tightly complementary surfaces in DNA-protein complexes (*3–6*) also include extensive contacts to the DNA backbone.

It has been suggested that the sequence-dependence of DNA conformation (*7*) might provide an indirect readout (*6, 7*) by affecting the attainment of optimal complementarity both for protein-base hydrogen bonding (*8*) and for the precisely positioned (*3, 5, 6*) interactions between protein and DNA phosphates. However, there has been no evidence to indicate which phosphate interactions are indispensable to recognition, which make only nonspecific contributions to binding free energy, and which (if any) are altered when a protein interacts with a closely related but incorrect site. For Eco RI endonuclease, it has been suggested (*9*) that indirect readout may also contribute to specificity because the energy required to attain the "kinked" DNA conformation in the complex (*3*) is least unfavorable for the GAATTC site.

To determine the roles of direct protein-DNA contacts and of sequence-dependent DNA conformation, we analyzed the energetics of stringent discrimination, using a rigorous measure of the Eco RI–DNA interaction in transition state complexes. We manipulated

The authors are in the Department of Biological Sciences, University of Pittsburgh, Pittsburgh PA 15260.

*Present address: Navy Drug Screening Laboratory, Oakland, CA 94627.
†To whom correspondence should be addressed.

the DNA side of the interface to explore those factors most likely to be generally applicable to other DNA binding proteins, and to identify the contacted functional groups on the bases, independent of any subsequent revisions of the amino acid residue assignments at the endonuclease-DNA interface (see "Note added in proof"). Base analogue substitutions that preserve the conformational motif of the GAATTC site permit the deletion of single sites of protein-base hydrogen bonding or van der Waals interactions to thymine methyl groups, without altering protein-phosphate interactions, leading to small (about +1.5 kcal/mol) and position-independent energy differentials. In contrast, any incorrect natural base pair produces position-dependent energy differentials (+6 to +13 kcal/mol) which are too large to be attributed to changes in protein-base interactions alone, but also include large contributions from changes in protein-phosphate interactions and differences in the energetic cost of distorting the DNA. Thus, stringent discrimination against natural DNA sites derives from interdependent changes in both direct and indirect readouts.

Experimental strategy for quantifying sequence discrimination. To evaluate the energetics of discrimination systematically, we examined all nine possible sites that differ from the canonical GAATTC by a single base pair. We embedded the canonical site and all sites with either an incorrect base pair or a base analogue substitution in essentially the same duplex 17-nucleotide (nt) oligomer (Fig. 1) (*10*) to maintain constant flanking sequences. In order for cleavage in each half-site to give rise to a distinguishable product, we placed the cleavage sites off center. Every site with one base pair substitution (for example the site $\frac{\text{AAATTC}}{\text{TTTAAG}}$) contains both a normal $\frac{\text{GAA}}{\text{CTT}}$ half-site and an altered half-site ($\frac{\text{AAA}}{\text{TTT}}$ in this example). We show for each site (Table 1) the predicted changes in hydrogen bonding and the potential for an unfavorable steric apposition of hydrogen bond donors or acceptors (*8, 11*).

The reaction of free endonuclease and DNA leading to the free product cleaved in both DNA strands includes enzyme interactions with the DNA substrate, reaction intermediates, and products. In order that structure-function correlations not be obscured by contributions from these various complexes, we considered only the interactions in the activated (transition state) enzyme–DNA ([E–DNA]‡) complex for the first irreversible bond-breaking step. The relative probability of reaching this transition state (from free enzyme and DNA) is a measure of discrimination between any two DNA sites. This probability, for each DNA site, is related to the standard free energy change ($\Delta G^\circ_I\ddagger$) in forming an activated [E–DNA]‡ complex from the free enzyme (E) and DNA. The transition state interaction energy $\Delta G^\circ_I\ddagger$ accounts for both the assembly of the enzyme-DNA (E-DNA) complex from the free molecules and the activation of the E-DNA complex to the transition state (*12*)

$$\text{E} + \text{DNA} \underset{\Delta G^\circ_{ED}}{\overset{K_A}{\rightleftharpoons}} \text{E} - \text{DNA} \underset{\Delta G^\circ_{\ddagger}}{\rightleftharpoons} [\text{E} - \text{DNA}]\ddagger$$

where K_A is the observed equilibrium association constant and ΔG°_{ED} the standard free energy for formation of the E-DNA complex, and $\Delta G^\circ\ddagger$ is the standard free energy of activation for the first bond-breaking step. Thus, $\Delta G^\circ_I\ddagger = \Delta G^\circ_{ED} + \Delta G^\circ\ddagger$, [where $\Delta G^\circ_{ED} = -RT \ln K_A$ and $\Delta G^\circ\ddagger = RT \ln (kT/h) - RT \ln k_c$]; and where k_c is the cleavage rate constant for the first bond-breaking step, k is the Boltzmann constant, h is the Planck constant, R is the ideal gas constant and T the absolute temperature.

The overall probability of the first bond-breaking step is given by

$$k_c \times K_A = (kT/h) \exp(-\Delta G^\circ_I\ddagger/RT)$$

K_A is directly measured (in the absence of cleavage) by omitting Mg^{2+} (*13–17*); k_c is measured (Fig. 2) by first forming E-DNA complexes in the presence of excess enzyme, and then adding Mg^{2+} to initiate the reaction.

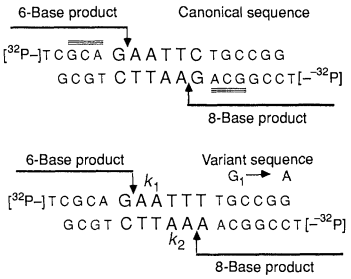

Fig. 1. Design of duplex 17-nt substrates. The recognition hexanucleotide is shown in large letters and cleavage sites are indicated by arrows. Double lines show the conserved three-base flanking region to each side. Methods of synthesis and purification by HPLC have been described (*19*). An equimolar mixture of the complementary strands was labeled with ^{32}P at the 5′-ends (*17*) and purified over NENsorb 20 to remove ATP. Complementary single strands were annealed; duplex products were purified by electrophoresis on nondenaturing 12.5 percent polyacrylamide gels, repurified on NENsorb 20, reannealed, and verified to be entirely in the duplex form by nondenaturing gel electrophoresis.

The product $k_c \times K_A$, analogous to the parameter k_{cat}/K_m derived from steady-state kinetics (where k_{cat} and K_m are the catalytic and Michaelis-Menten constants), avoids two difficulties for Eco RI endonuclease: (i) k_{cat} reflects a rate-limiting step after bond cleavage (*13, 14*); (ii) K_m is not equal to the enzyme-substrate binding constant (*18*) and may be dominated by kinetic constants for different reaction steps when comparing different DNA substrates.

Discrimination between substrates 1 and 2 (the relative probabilities of reaching the transition state) can be cast in terms of a discrimination energy $\Delta\Delta G^\circ_I\ddagger$, the difference in the energetic costs of attaining the transition state

$$\Delta\Delta G^\circ_I\ddagger = -RT \ln [(k_{c_1} \times K_{A_1})/(k_{c_2} \times K_{A_2})]$$

Although the individual constants K_A and k_c may be affected by nonproductive binding (*12*) or the presence of nonspecific complexes as intermediates in formation of the cleavage-competent complex (*14*), these effects cancel out in $k_c \times K_A$ and $\Delta\Delta G^\circ_I\ddagger$.

Discrimination in both binding and cleavage. Measured values of K_A at physiological pH (Table 2) show that GAATTC binds better than any site containing one or more incorrect base pairs ($\Delta\Delta G^\circ_{ED}$ is +4 to +6 kcal/mol), but $\Delta\Delta G^\circ_{ED}$ for these sites (Table 2) has no simple correspondence to the predicted changes in hydrogen bonds to the bases (Table 1). Even the inverted site

Table 1. Predicted changes in protein interactions with the purine bases at sites with one incorrect base pair (incorrect purine in boldface). A crystallographic model (*3*) proposed symmetrical interactions with the purine bases on both strands of the canonical GAATTC site by hydrogen bond donors (Don) and acceptors (Acc) on the protein. Current information (*4*) confirms that protein side chains are in position to make these interactions, but amino acid residue assignments may be revised.

Site	H-bonds Missing	H-bonds Conserved*	Unfavorable Appositions
GAATTC	None	All	None
	First-base substitutions		
TAATTC	Don 1→N^7-G	Don 2→O^4-T	None
AAATTC	Don 2→O^6-G	Don 1→N^7-A	Don 2↔N^6-A
CAATTC	Don 2→O^6-G and Don 1→N^7-G	None	Don 2↔N^4-C
	Second-base substitutions		
GCATTC	Don 3→N^7-A	Acc 1←N^4-C	None
GGATTC	Acc 1←N^6-A	Don 3→N^7-G	Acc 1↔O^6-G
GTATTC	Don 3→N^7-A and Acc 1←N^6-A	None	Acc 1↔O^4-T
	Third-base substitutions		
GACTTC	Don 4→N^7-A	Acc 2←N^4-C	None
GAGTTC	Acc 2←N^6-A	Don 4→N^7-G	Acc 2↔O^6-G
GATTTC	Don 4→N^7-A and Acc 2←N^6-A	None	Acc 2↔O^4-T

*G → T and A → C substitutions retain hydrogen bond acceptor (carbonyl) or donor (amino) groups, respectively, in virtually the same positions (*55*) that they occupied in the canonical base. Thus, only the interaction with purine-N^7 is necessarily deleted. G → A and A → G substitutions conserve the interaction to N^7.

(**CTTAAG**, incorrect bases in boldface) that represents nonspecific binding has about the same K_A as sites predicted to form 11 of the 12 protein-base hydrogen bonds.

Interpretation of $\Delta\Delta G°_{ED}$ must take account of three factors. (i) Incorrect base pairs have nonadditive effects on binding free energies because they perturb not only protein-base interactions, but also protein-phosphate interactions and the DNA conformation; (ii) Binding to sites containing too many incorrect base pairs (here two or more, for example, GGATCC) is in a distinct nonspecific mode; (iii) The observed binding constants K_A include contributions from binding to nonspecific sites on each oligonucleotide. This may be described by $K_A = K_S + N_{NS}K_{NS}$ where K_S is the intrinsic equilibrium association constant for the hexanucleotide recognition site, K_{NS} is the association constant for nonspecific sites in the same oligonucleotide, and N_{NS} is the number of such nonspecific sites (16 to 20 in these 17-nt oligomers) (Fig. 1). The nonspecific contribution is significant for sites with one incorrect base pair because binding is nearly isoenergetic to that at nonspecific sites, but negligible for GAATTC (where $K_S >> N_{NS}K_{NS}$). These factors should be considered in interpreting the differential binding of any protein to various nucleic acid sites.

Although these measurements of K_A were made in the absence of Mg^{2+} to avoid DNA cleavage, several lines of evidence show that K_A is not affected by the presence of Mg^{2+}. (i) Identical values of K_A were obtained in the presence and absence of Mg^{2+} for the slowly cleaved GAGTTC and GTATTC sites (Table 2). (ii) We have used NH_2-terminal deletion derivatives of the endonuclease that retain full binding specificity but cannot cleave DNA (*19*) to show that Mg^{2+} does not affect the measured K_A for GAATTC or sites with one incorrect base pair (*20*).

To permit direct measurement of first-order cleavage rate contants for the bond-breaking steps, high enzyme and substrate concentrations were used to drive all sites into enzyme-substrate complexes. For each DNA site, k_1 represents cleavage in the constant GAA half-site (Fig. 2) and k_2 represents the cleavage in the altered half-site; k_4 and k_3 are the corresponding constants for cleavage of the nicked intermediates. The following features are apparent from data (Table 3) obtained at physiological pH (*21*).

(i) The canonical GAATTC site is cleaved at equal rates (Fig. 2C) in each strand; that is, $k_1 = k_2$. Cleavage of the nicked intermediates (k_3 or k_4) is only slightly slower so that double-strand cleavage occurs without significant accumulation of nicked intermediates, as it does in pBR322 DNA (*13, 14*). (ii) There is discrimination (up to 10^6-fold) between DNA sites in the first-order rate constants for strand scission (Table 3). (iii) For all single-substituted sites (except GACTTC), cleavage in the GAA half-site (k_1) is faster than in the variant half-site (k_2) and nicked intermediates accumulate because k_3 and k_4 are reduced.

We chose k_1 as the appropriate cleavage rate constant k_c to calculate $\Delta G°_I‡$ because it reflects cleavage in a locally constant GAA half-site and thus avoids any effect of a base substitution in the immediate vicinity of the scissile bond. The significant variation of the rate constant k_1 in response to base substitutions in the other half-site is a consequence of the architecture of the complex, in

Table 2. Changes in free energies of formation of E-DNA complexes ($\Delta\Delta G°_{ED}$) and transition state complexes ($\Delta\Delta G°_I‡$) for variant DNA sites. Each site (incorrect base pair in boldface) was embedded in the same 17-nt context (Fig. 1). The vertical arrow denotes cleavage in the GAA half-site represented by the rate constant k_1. Calculated free energy differences are relative to the GAATTC site. Equilibrium association constants K_A (mean ± SD; n = 3 to 8) were obtained by rate-competition (*56*) and are expressed as moles of duplex 17-nt oligomer, and endonuclease dimers as the active species (*13, 17*). Unlabeled competing 17-nt duplex DNA (5 nM to 1.6 μM, as appropriate for each site) was equilibrated with Eco RI (50 pM) for 30 minutes in binding buffer (10 mM bis-tris-propane, 5 μM dithiothreitol, 1 mM EDTA, bovine serum albumin at 50 mg/ml, 95 mM NaCl, pH 7.5, 25 °C). Duplex [5′-^{32}P]-TCGCGAATTCGCG (equimolar to the endonuclease) was added, and samples were removed at various times and filtered through nitrocellulose membranes. The K_A was calculated (*56*). For each test oligonucleotide, K_A was not affected by increasing the enzyme concentration to 200 pM or increasing the unlabeled competing oligonucleotide concentration fivefold. Values for K_A were independently confirmed by direct binding measurements (*17*) for the canonical 17-nt oligomer and by equilibrium competition (*19, 56*) for other sites. Data for k_1 are from Table 3.

Site	K_A* (M^{-1})	$\Delta\Delta G°_{ED}$§ (kcal/mol)	k_1 (sec^{-1})	Relative $(k_1 \times K_A)$‖ $(M\text{-}sec)^{-1}$	$\Delta\Delta G°_I‡$ (kcal/mol)
GAATT C CTTAA↑G	$3.3(\pm1.6)\times10^{11}$	0	$3.4(\pm0.7)\times10^{-1}$	1	0
		First-base substitutions			
TAATT C ATTAA↑G	$3.1(\pm1.6)\times10^{8}$	4.1±0.5	$5.0(\pm0.3)\times10^{-3}$	1.5×10^{-5}	6.6±0.5
AAATT C TTTAA↑G	$1.8(\pm1.0)\times10^{8}$	4.4±0.5	$8.4(\pm1.7)\times10^{-4}$	1.4×10^{-6}	8.0±0.5
CAATT C GTTAA↑G	$5.3(\pm2.2)\times10^{7}$	5.2±0.4	$2.0(\pm0.7)\times10^{-4}$	1.0×10^{-7}	9.5±0.5
		Second-base substitutions			
GCATT C CGTAA↑G	$4.4(\pm1.7)\times10^{7}$	5.3±0.4	$6.7(\pm0.5)\times10^{-5}$	2.7×10^{-8}	10.3±0.4
GGATT C CCTAA↑G	$2.6(\pm1.7)\times10^{7}$	5.6±0.5	$6.2(\pm1.5)\times10^{-6}$	1.5×10^{-9}	12.0±0.6
GTATT C CATAA↑G	$1.6(\pm0.8)\times10^{8}$	4.5±0.5	~2 ×10^{-7}	2.9×10^{-10}	~13
		Third-base substitutions			
GACTT C CTGAA↑G	$5.9(\pm0.6)\times10^{7}$	5.1±0.3	$4.5(\pm0.6)\times10^{-4}$	2.5×10^{-7}	9.0±0.3
GAGTT C CTCAA↑G	$1.1(\pm0.9)\times10^{8}$	4.7±0.6	$8.9(\pm0.3)\times10^{-5}$	8.9×10^{-8}	9.6±0.3
GATTT C CTAAA↑G	$3.4(\pm0.7)\times10^{7}$	5.4±0.4	$2.0(\pm0.5)\times10^{-6}$	6.2×10^{-10}	12.5±0.4
		Double substitutions¶			
GGATCC	$9.7(\pm4.8)\times10^{7}$	4.8±0.5	No cleavage		
GAGCTC	$4.3(\pm2.0)\times10^{7}$	5.3±0.4	No cleavage		
AAATTA	$2.0(\pm1.0)\times10^{7}$	5.7±0.5	No cleavage		
		Inverted site			
CTTAAG	$4.7(\pm2.6)\times10^{7}$	5.2±0.5	No cleavage		

*K_A was also measured for the slowly cleaved GAGTTC and GATTTC sites by gel-retardation (*57*) with and without 3 mM Mg^{2+}; NaCl concentration in the buffer was decreased to 89 mM when Mg^{2+} was present. Measured values ±Mg^{2+} were essentially identical to those obtained by rate-competition in the absence of Mg^{2+}. §Calculated from the ratio of K_A for each variant site to K_A for the GAATTC site. ‖This ratio measures discrimination, relative to GAATTC. ¶We also found no cleavage of sites containing two substitutions in one half-site (TACTTC, CACTTC, GCCTTC, GGGTTC, GCGTTC) or one substitution in each half-site (GACGTC, GACTTT). The only double-substituted sites cleaved (product ≤2 percent, after 24 hours) were AACTTC and GGCTTC.

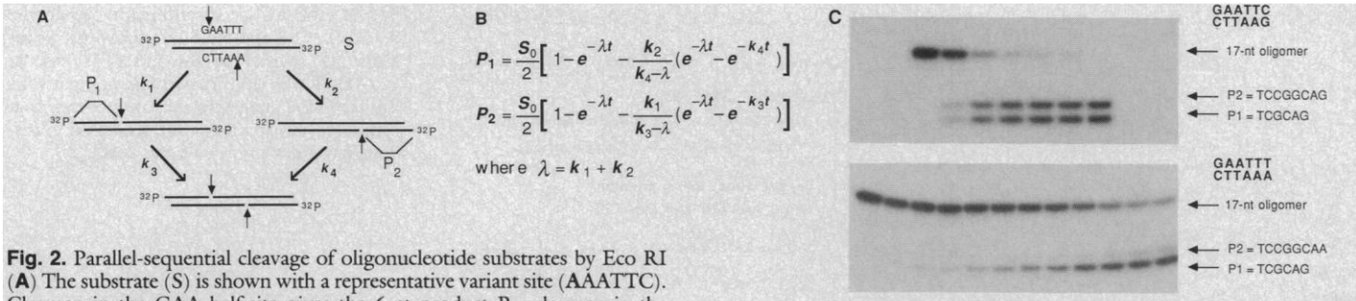

Fig. 2. Parallel-sequential cleavage of oligonucleotide substrates by Eco RI (**A**) The substrate (S) is shown with a representative variant site (AAATTC). Cleavage in the GAA half-site gives the 6-nt product P_1; cleavage in the variant (AAA) half-site gives the 8-nt product P_2. The model is valid when (Table 3) intact substrate and nicked intermediates are completely enzyme-bound and product release is achieved by quenching. (**B**) The equations show the time dependence of appearance of P_1 and P_2 as a function of the four first-order kinetic constants $k_1 - k_4$. The factor of 1/2 in each equation reflects the fact that both substrate strands are labeled, and each product can represent no more than one-half of the initial substrate radioactivity. Kinetic constants were computed as described in Table 3. (**C**) Autoradiograms show representative time courses for cleavage of (upper gel) GAATTC (times 0, 1.3, 3.3, 6.4, 8.6, 11.1 seconds, and 15 minutes, lanes left to right) and (lower gel) AAATTC (times 0, 1, 2, 3, 5, 10, 15, 20, 40, 80, 165, 300 minutes, lanes left to right).

which each half-site receives recognition contacts from both protein subunits and is cleaved in a catalytic cleft formed from elements of both subunits (*3*). Thus, the interaction energy over the entire transition state complex must be considered; k_1 reports on global (not only local) features of the transition state complex (*22*).

Stringent discrimination and the components of $\Delta\Delta G^\circ_I\ddagger$. Any single base pair substitution produces a large (+6.6 to +13 kcal/mol) unfavorable $\Delta\Delta G^\circ_I\ddagger$ (Table 2 and Fig. 3A). The changes in protein-base interactions (Table 1) account for the hierarchy (*23*) of substitutions at each base position, but $\Delta\Delta G^\circ_I\ddagger$ for each site (relative to the canonical GAATTC) includes additional factors.

The interaction energy $\Delta G^\circ_I\ddagger$ for any site is the aggregate of protein-base (ΔG_{base}) and protein-phosphate (ΔG_{phos}) interaction terms and a "reorganization" term ΔG_{reorg} (*24, 25*) that includes any other entropic or conformational factors. The discrimination energy $\Delta\Delta G^\circ_I\ddagger$ between GAATTC and any site containing an incorrect base pair (Fig. 3, A and B, case 1) may include not only changes in the protein-base contacts ($\Delta\Delta G_{base}$, Fig. 3B), but also an energy differential associated with altered protein contacts with DNA phosphates ($\Delta\Delta G_{phos}$) and increased costs of driving variant DNA sites into the kinked conformation ($\Delta\Delta G_{reorg}$). If all three terms change, then the separate contributions of each cannot be extracted from the comparison of GAATTC to any site containing an incorrect natural base pair. That $\Delta\Delta G_{base}$, $\Delta\Delta G_{phos}$, and $\Delta\Delta G_{reorg}$ contribute to discrimination is shown by the following. (i) Sites containing base analogues that delete protein-base hydrogen bonding sites or thymine methyl groups show that the contributions (1 to 2 kcal/mol) of single hydrogen bonds or interactions with methyl groups are too small to account for the large energetic discrimination against natural base pair substitutions. (ii) Consistent estimates of 1 to 2 kcal/mol for the energy differentials associated with changes in individual protein-base hydrogen bonds are obtained by confining comparisons to pairs of sites containing natural base pair substitutions at the same position (Fig. 3, A and B, case 2). (iii) The pattern of protein-phosphate contacts in complexes containing an incorrect natural base pair always shows major characteristic differences from the pattern in the canonical GAATTC complex. Isosteric sites with base analogue substitutions that preserve the sequence-dependent conformation of the canonical GAATTC do not show these differences.

Table 3. Cleavage rate constants at physiological pH. First-order strand scission constants were obtained under single-turnover conditions [concentrations of both enzyme (1.6 μM) and DNA (0.7 μM) were >20 times above the equilibrium dissociation constant (K_D) for the worst of the sites (Table 2)]. Enzyme-DNA (17-nt oligomer) complexes were equilibrated for 30 minutes in binding buffer (Table 2), pH 7.5, 25°C. Nondenaturing gels showed that all DNA sites were completely enzyme-bound; increased enzyme concentrations (5.6 μM) did not affect the reaction rates. Reactions were initiated by adding $MgCl_2$ to 5 mM, samples were removed at various times and quenched by addition of urea to 8M and EDTA to 50 mM. Products were separated by electrophoresis on 16 percent polyacrylamide, 8M urea gels and quantitated by microdensitometry and scintillation counting. Best-fitting values of the four rate constants were computed simultaneously by nonlinear least-squares fits to data on products P_1 and P_2 (Fig. 2) and full-length strands from both remaining substrate and nicked intermediate ($BD = S_0 - P_1 - P_2$). Partially denaturing gels that separated uncut duplex from intact single strands (derived from nicked duplexes) gave the same values, with equations that take account of remaining substrate and intermediates. Some sites were checked by annealing one labeled strand to its unlabeled complement such that only P_1 or P_2 was labeled; the rate constants obtained were the same. Values are means ± SD of three to five independent experiments for each oligonucleotide. All hexanucleotide recognition sites were oriented as in Fig. 1, but observed cleavage rate asymmetries do not depend on orientation, because an oligonucleotide with a TAATTC site in the opposite orientation yielded the same values for k_1 through k_4. Rate constants do not depend on whether Mg^{2+} is added to E-DNA complexes or DNA is added to E-Mg^{2+} complexes, as has been observed for the GAATTC site (*50*).

Site	k_1 (sec^{-1})	k_2 (sec^{-1})	k_3 (sec^{-1})	k_4 (sec^{-1})
GAATTC	$3.4(\pm0.7)\times10^{-1}$	$3.2(\pm0.8)\times10^{-1}$	$2.5(\pm0.9)\times10^{-1}$	$2.2(\pm0.6)\times10^{-1}$
TAATTC	$5.0(\pm0.3)\times10^{-3}$	$2.2(\pm1.5)\times10^{-4}$	$1.4(\pm1.6)\times10^{-5}$	$1.6(\pm1.1)\times10^{-3}$
AAATTC	$8.4(\pm1.7)\times10^{-4}$	$1.5(\pm0.3)\times10^{-4}$	$2.8(\pm0.6)\times10^{-5}$	$3.8(\pm1.7)\times10^{-5}$
CAATTC	$2.0(\pm0.7)\times10^{-4}$	$1.1(\pm0.2)\times10^{-5}$	*	*
GCATTC	$6.7(\pm0.5)\times10^{-5}$	$6.3(\pm1.9)\times10^{-6}$	*	*
GGATTC	$6.2(\pm1.5)\times10^{-6}$	Nil§	Nil	NMF‖
GTATTC	~2 $\times10^{-7}$	Nil	Nil	NMF
GACTTC	$4.5(\pm0.6)\times10^{-4}$	$7.2(\pm0.7)\times10^{-4}$	$1.0(\pm1.1)\times10^{-5}$	$1.1(\pm0.9)\times10^{-4}$
GAGTTC	$8.9(\pm0.3)\times10^{-5}$	$9.3(\pm5.9)\times10^{-6}$	*	*
GATTTC	$2.0(\pm0.5)\times10^{-6}$	Nil	Nil	NMF

*The amount of 8-nt product (P_2) observed in these cases was too small (≤4% of initial substrate radioactivity after 8 hours) to permit accurate calculation of kinetic constants k_3 and k_4. One or more individual experiments yielded calculated values of zero. §Nil, no product observed. ‖NMF, no meaningful figure. For the GGATTC, GTATTC, and GATTTC sites, no cleavage in the variant half-site was observed; thus the substrate for the reaction described by k_4 never appears.

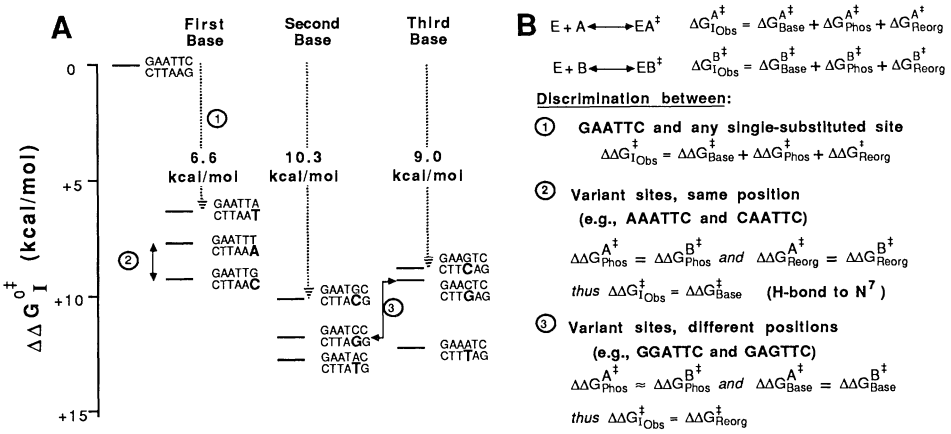

Fig. 3. Analysis of discrimination energies. (**A**) $\Delta\Delta G^{\circ}_{I}\ddagger$ for all nine single-substituted sites (Table 2), relative to the GAATTC site at zero. (**B**) Three kinds of pairwise comparisons between sites, analyzing the components that contribute to observed differences in transition-state interaction energies $\Delta\Delta G^{\circ}_{I}\ddagger$.

Small energetic penalties for hydrogen bond deletion in isosteric sites. In the first base position, substitution of 7-deazaguanine (^{7C}G in the site ^{7C}GAATTC) deletes the hydrogen bond to N^7 of guanine, at a cost ($\Delta\Delta G^{\circ}_{ED}$) of +1.1 kcal/mol (Table 4). A 2-aminopurine substitution (^{2A}P in the site ^{2A}PAATTC) (*26*) deletes the hydrogen bond to O^6 of guanine, with $\Delta\Delta G^{\circ}_{ED}$ = +1.7 kcal/mol. The $\Delta\Delta G^{\circ}_{I}\ddagger$ for these isosteric sites (about +2.5 kcal/mol) is also small.

In the third position, disrupting the hydrogen bond from N^6 of adenine [by the isosteric (*27*) substitution of N^6-methyladenine (^{m}A) in the site GA^{m}ATTC] produces $\Delta\Delta G^{\circ}_{ED}$ of +1.2 kcal/mol (Table 4). Similar values are observed for base-analogue sites that delete hydrogen bonding groups (adenine-N^6 or adenine-N^7) in the second or third positions (*26*). $\Delta\Delta G^{\circ}_{I}\ddagger$ for GA^{m}ATTC (+2.6 kcal/mol) is also small.

By contrast, the nonisosteric substitution A → ^{7C}G in the third-base position (GA^{7C}GTTC) produces a large $\Delta\Delta G^{\circ}_{I}\ddagger$ (+11 kcal/mol, Table 4), comparable to the +9.6 kcal/mol observed for the GAGTTC site, but with an additional penalty of +1.4 kcal/mol for deletion of the hydrogen bond to N^7 of guanine. As controls, ^{7C}G outside the recognition site (^{7C}GCAGAATTC) has no effect on the interaction energy ($\Delta\Delta G^{\circ}_{I}\ddagger$ = 0), and ^{7C}G on the opposite strand of variant CAATTC site ($_{7C}^{CAATTC}$GTTAAG, Table 4) gives the same $\Delta\Delta G^{\circ}_{I}\ddagger$ as that normally observed for CAATTC (+9 kcal/mol) (Table 2).

The earlier crystallographic model (*3*) proposed no interactions with the pyrimidines, but the revised model (*4*) reveals pyrimidine contacts consistent with biochemical evidence (*28*). Deletion of any one thymine methyl group (*26*) (in isosteric GAAUTC or GAATUC sites) exacts only a small penalty in $\Delta\Delta G^{\circ}_{ED}$ (+1.2 kcal/mol for GAAUTC; +1.7 kcal/mol for GAATUC). There is little or no further penalty in $\Delta\Delta G^{\circ}_{I}\ddagger$. Similarly, replacing C in one half-site with A to form the isosteric (*29*) G·A mismatch site increases $\Delta\Delta G^{\circ}_{ED}$ by only about +1 kcal/mol. Symmetrical uracil substitution in both half-sites produces strictly additive effects on $\Delta\Delta G^{\circ}_{ED}$ (*26*).

For isosteric sites, preservation of the DNA conformation has the effect that $\Delta\Delta G_{reorg} \approx 0$, and the phosphate contacts (below) show no major changes from those in the GAATTC complex, so that $\Delta\Delta G^{\circ}_{phos} \approx 0$ (when assessed in the E-DNA complex). Furthermore, K_A closely approximates K_S (because $K_S >> N_{NS}K_{NS}$). With the obscuring factors eliminated for these sites, $\Delta\Delta G^{\circ}_{ED} = \Delta\Delta G_{base}$. [The small additional difference ($\Delta\Delta G^{\circ}\ddagger$ about +1 kcal/mol) between $\Delta\Delta G^{\circ}_{ED}$ and $\Delta\Delta G^{\circ}_{I}\ddagger$ for some isosteric sites reflects an unfavorable protein-phosphate interaction realized only in the transition state (see below).] Thus, $\Delta\Delta G^{\circ}_{ED}$ for the isosteric sites shows that deletion of a single protein-base hydrogen bond or a van der Waals contact to a thymine methyl group can provide only a small contribution (+1.1 to +1.7 kcal/mol) to site discrimination.

Discrimination increments within each series of incorrect DNA sites. The incremental energies of each protein-purine base interaction can also be extracted by comparing complexes (Fig. 3, A and B, case 2) with different incorrect base pairs (at the same base position) to each other rather than to the GAATTC complex. Both members of any such pair lack the normal pyrimidine base on the opposing strand, and thus suffer equal penalties from loss of any pyrimidine interaction. The $\Delta\Delta G_{base}$ values obtained, for each set of three substitutions in three different positions, are consistent with each other, with the proposed hydrogen-bonding model (Table 1), and with the values from isosteric base-analogue sites, indicating that the $\Delta\Delta G_{phos}$ and $\Delta\Delta G_{reorg}$ terms are the same for any subsitution at a given base position. Thermodynamic analysis (*30, 31*) confirms that the $\Delta\Delta G_{phos}$ terms are equivalent for the nine sites containing single base pair subsitutions, and high-resolution footprinting (below) shows that these sites share a common pattern of changes in the protein-phosphate contacts.

At the first base position, the canonical G is recognized (Fig. 4) by hydrogen bonds to O^6 and N^7. The hierarchy of first-position substitutions T > A > C (Table 2) results from successive steps of

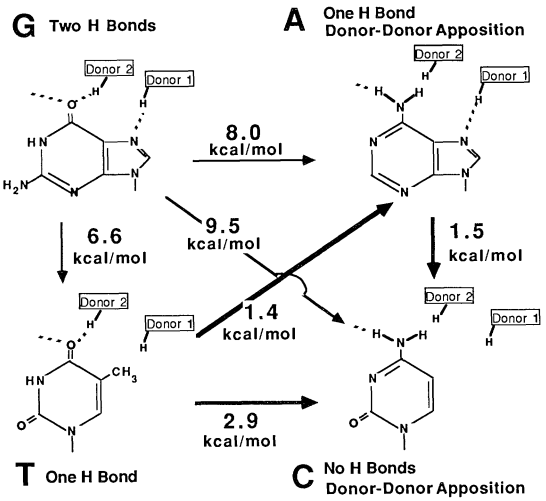

Fig. 4. Transition-state interaction free energy differences for sites containing substitutions at the first base-pair position. Only the first base of the hexanucleotide site is shown, interacting with hydrogen bond donor groups (rectangles) on the protein (Table 1). Light arrows show $\Delta\Delta G^{\circ}_{I}\ddagger$ between GAATTC and (clockwise) AAATTC, CAATTC, and TAATTC sites. Heavy arrows show differences in $\Delta G^{\circ}_{I}\ddagger$ between pairs of variant sites. Watson-Crick hydrogen bonds to the opposite DNA strand are shown extending to the left.

loss of one hydrogen bond, loss of one hydrogen bond with introduction of an unfavorable steric apposition, and loss of two hydrogen bonds with an unfavorable apposition (Table 1).

The change from **A**AATTC to **C**AATTC (Fig. 4) represents the deletion of the imidazole ring, the cost of which (+1.5 kcal/mol) represents the incremental binding energy (*32*) of the hydrogen bond to N^7 of adenine. Comparison of **T**AATTC to **A**AATTC (Fig. 4) shows that each complex has lost and retained one hydrogen bond. The difference is that N^6 of adenine is presented in unfavorable steric apposition (*8, 11*) to a hydrogen bond donor on the protein, and this apparently costs +1.4 kcal/mol. Thus, there is an energetic penalty associated with the introduction of an incorrect functional group on a base, distinguishable from the deletion of a protein-base hydrogen bond. Comparison of **T**AATTC to **C**AATTC shows that the latter both deletes the hydrogen bond to O^4 of thymine and presents N^4 of cytosine unfavorably apposed to a hydrogen bond donor on the protein. The total penalty is +2.9 kcal/mol, implying that removal of the hydrogen bond to O^4 of thymine costs about +1.5 kcal/mol.

These orderly energetic increments imply that disruption of one hydrogen bond to guanine (G → A or G → T substitutions) does not disrupt the other hydrogen bond to the conserved functional group on the first base (Fig. 4). This conclusion is consistent with the increments observed for isosteric sites containing ^{7C}G or ^{2A}P substitutions (Table 4). A parallel analysis of substitutions for the canonical A in the second position shows that the substitution hierarchy C > G > T represents a comparable stepwise change in the protein-base interactions and leads to quantitatively consistent conclusions: Removal of the hydrogen bond to N^7 of adenine costs about +1 kcal/mol, the unfavorable apposition of a hydrogen bond acceptor on the protein to O^6 of guanine contributes +1.7 kcal/mol, and the total penalty for the C → T substitution is +2.7 kcal/mol, implying that removal of the hydrogen bond from N^4 of cytosine costs about +1 kcal/mol. Our data for third-base substitutions (Table 2) lead to the same conclusions.

The existence of these orderly hierarchies (in all three base positions) implies that a base substitution may disrupt one or two hydrogen bonds while conserving all other hydrogen bonds to the bases; there is no catastrophic collapse of the protein-base hydrogen bonding interface. This conclusion is supported by two independent lines of evidence. (i) The salt-insensitive component of binding free energy (*31*), attributable largely to the protein-base interactions (*17, 19*), is reduced by only 1 to 2 kcal/mol in complexes with one incorrect base pair (*30*). (ii) The "missing contact" footprinting method shows that the **T**AATTC site retains contacts at AA in the variant half-site and at GAA in the normal half-site (*33*).

Thus, when we confine comparisons to pairs of sites with substitutions at the same base position (Fig. 3, A and B; case 2), the data for all nine single-substituted sites yield a consistent estimate of 1 to 2 kcal/mol for the incremental energetic contribution of each protein-base interaction. Formation of each protein-base interaction is an exchange process in which hydrogen bonds to water are surrendered by particular functional groups. The incremental energies of the interactions reflect the differential contributions of these exchange processes (*8, 25*). The quantitatively consistent effects of hydrogen bond deletions or unfavorable appositions (or both) at the purine bases imply that the protein places hydrogen bond donors or acceptors in approximately the same positions in both the canonical complex and complexes with an incorrect base pair.

The combined data from base-analogue and natural base pair substitutions confirm direct readout of sequence information by hydrogen bonding to the purine bases and contacts to the complementary pyrimidines. They also demonstrate that changes in protein-base interactions produce energetic effects too small to account for stringent discrimination against natural base pair substitutions.

Phosphate contacts in recognition—the canonical complex. The protein appears to be in proximity to the DNA backbone over a span of about 10 bp (*3*), but only six symmetry-related phosphates (three per strand) have a crucial role in recognition. The phosphate contacts at **pA**p**G**A**A**p**T**TC facilitate base recognition in that they anchor and orient protein recognition elements within the major groove of the DNA, and also act as clamps to stabilize the kinked DNA conformation in the complex (*3*). Kinking widens the major groove for insertion of the protein recognition elements (*3*).

The importance of these phosphate contacts is confirmed by the concordance of data from ethylation-interference footprinting (Figs. 5A and 6A) (*34, 35*) and photofootprinting (*35*). The DNA backbone at the phosphate contacts C**p**A**p**GAATTC is buried in clefts in the protein (*3*). Photofootprinting (*35*) shows that these contacts severely constrain the rotational mobility of the two upstream base pairs on each strand. The contacts at GAA**p**TTC (site of the central kink) are made by tight packing of the protein against the DNA (*3, 4, 36*) and the functional significance of these contacts is confirmed by a stereoselective effect of phosphorothioate substitution on both binding and cleavage (*37*).

Ethylation of any one of these three phosphates (six per duplex) strongly inhibits binding (Figs. 5A and 6A); the positions and strengths of the interferences are completely symmetrical on the two DNA strands. Strong interference [a large effect of ethylation of a

Table 4. Changes in free energies of formation of E–DNA complexes ($\Delta\Delta G^\circ_{ED}$) and transition state complexes ($\Delta\Delta G^\circ_{\ddagger}$) for base-analogue sites. Free energy differences and $k_1 \times K_A$ are given relative to the GAATTC site (Table 2). The methods were the same as those in Table 2. Oligonucleotides (flanking sequences as in Fig. 1) containing ^{7C}G were synthesized on an Applied Biosystems DNA synthesizer, with the use of fully protected 7-deaza-2′-deoxyguanosine β-cyanoethyl phosphoramidite (*58*). The protected N^6-methyl-2′-deoxyadenosine β-cyanoethyl-phosphoramidite was from Pharmacia. Oligonucleotides containing 2-aminopurine (^{2A}P) were prepared as described (*26*). K_A values for the three isosteric sites were identical by rate-competition and direct-binding methods. Reversing the orientation of isosteric sites in the 17-nt oligomer had no effect on the observed values of K_A and k_1.

Site	K_A (M^{-1})	$\Delta\Delta G^\circ_{ED}$ (kcal/mol)	k_1 (sec^{-1})	Relative $(k_1 \times K_A)$ $(M\text{-}sec)^{-1}$	$\Delta\Delta G^\circ_{\ddagger}$ (kcal/mol)
		Isosteric substitutions			
^{7C}GAATTC / CTTAA↑G	$4.9(\pm1.2)\times10^{10}$	1.1±0.3	$3.8(\pm0.2)\times10^{-2}$	1.7×10^{-2}	2.4±0.4
^{2A}PAATTC / CTTAA↑G	$1.9(\pm0.2)\times10^{10}$	1.7±0.3	$9.4(\pm1.4)\times10^{-2}$	1.5×10^{-2}	2.5±0.3
GA^{m}ATTC / CT TAA↑G	$4.5(\pm0.1)\times10^{10}$	1.2±0.3	$3.0(\pm0.3)\times10^{-2}$	1.2×10^{-2}	2.6±0.3
		Non-isosteric substitutions			
GA^{7C}GTT C / CT CAA↑G	$3.3(\pm1.2)\times10^{8}$	4.1±0.4	$2.9(\pm1.6)\times10^{-6}$	8.5×10^{-9}	11.0±0.5
CAATTC / ^{7C}GTTAA↑G	$1.4(\pm0.6)\times10^{8}$	4.6±0.4	$2.0(\pm0.4)\times10^{-4}$	2.5×10^{-7}	9.0±0.4

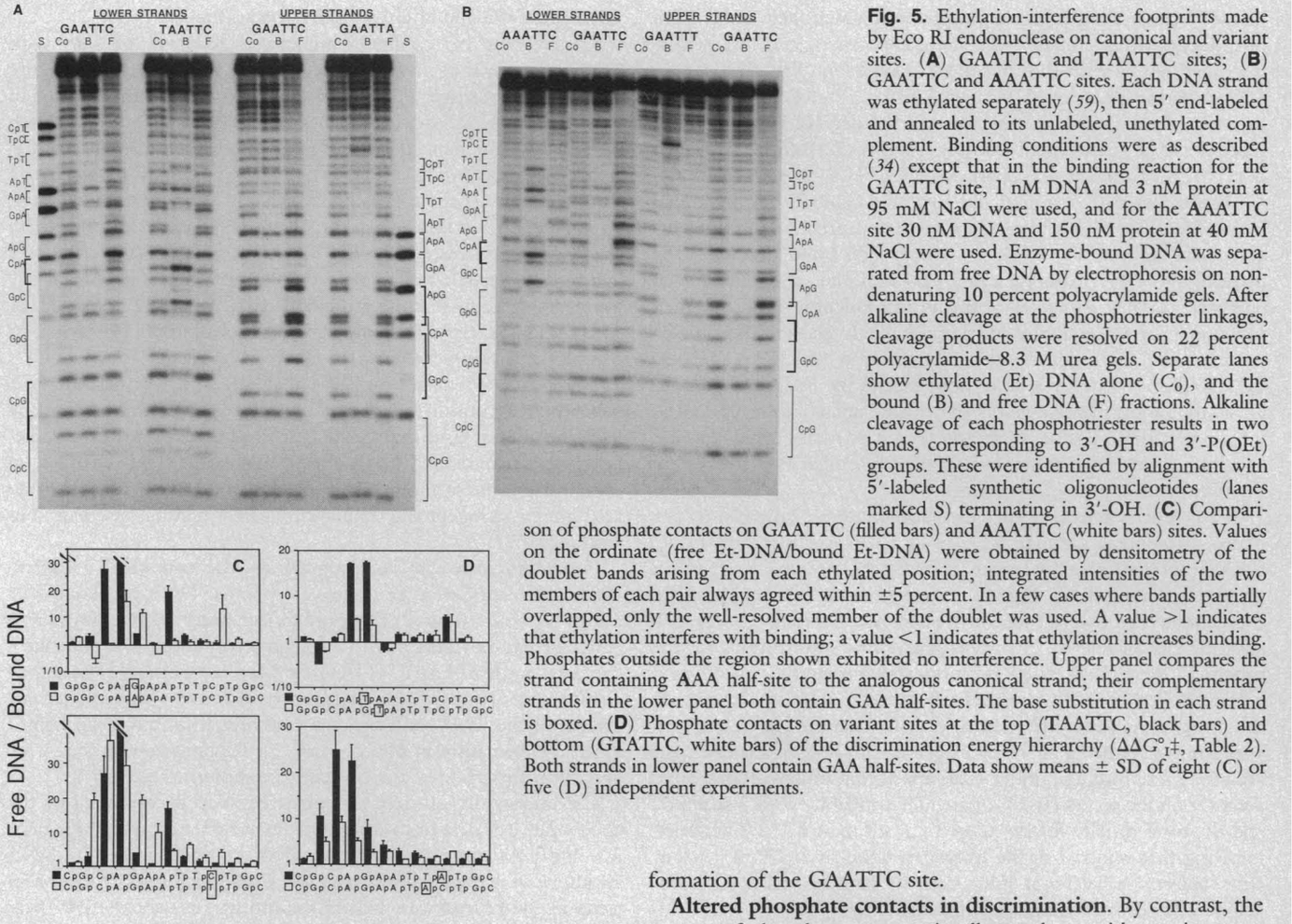

Fig. 5. Ethylation-interference footprints made by Eco RI endonuclease on canonical and variant sites. (**A**) GAATTC and TAATTC sites; (**B**) GAATTC and **A**AATTC sites. Each DNA strand was ethylated separately (*59*), then 5′ end-labeled and annealed to its unlabeled, unethylated complement. Binding conditions were as described (*34*) except that in the binding reaction for the GAATTC site, 1 nM DNA and 3 nM protein at 95 mM NaCl were used, and for the **A**AATTC site 30 nM DNA and 150 nM protein at 40 mM NaCl were used. Enzyme-bound DNA was separated from free DNA by electrophoresis on nondenaturing 10 percent polyacrylamide gels. After alkaline cleavage at the phosphotriester linkages, cleavage products were resolved on 22 percent polyacrylamide–8.3 M urea gels. Separate lanes show ethylated (Et) DNA alone (C_0), and the bound (B) and free DNA (F) fractions. Alkaline cleavage of each phosphotriester results in two bands, corresponding to 3′-OH and 3′-P(OEt) groups. These were identified by alignment with 5′-labeled synthetic oligonucleotides (lanes marked S) terminating in 3′-OH. (**C**) Comparison of phosphate contacts on GAATTC (filled bars) and **A**AATTC (white bars) sites. Values on the ordinate (free Et-DNA/bound Et-DNA) were obtained by densitometry of the doublet bands arising from each ethylated position; integrated intensities of the two members of each pair always agreed within ±5 percent. In a few cases where bands partially overlapped, only the well-resolved member of the doublet was used. A value >1 indicates that ethylation interferes with binding; a value <1 indicates that ethylation increases binding. Phosphates outside the region shown exhibited no interference. Upper panel compares the strand containing **A**AA half-site to the analogous canonical strand; their complementary strands in the lower panel both contain GAA half-sites. The base substitution in each strand is boxed. (**D**) Phosphate contacts on variant sites at the top (**T**AATTC, black bars) and bottom (**GT**ATTC, white bars) of the discrimination energy hierarchy ($\Delta\Delta G°_I‡$, Table 2). Both strands in lower panel contain GAA half-sites. Data show means ± SD of eight (C) or five (D) independent experiments.

particular phosphate on binding energy (*38*)] must reflect the cooperative loss of other contacts, because any single (independent) protein-phosphate interaction makes only a small (1.3 kcal/mol) incremental contribution to binding energy (*19, 39*) and thus produces only weak interference.

Weak interference (reproducible three- to fourfold inhibition of binding) (Figs. 5C and 6A) is exemplified by the phosphates at **p**CAG**p**AAT**p**TC. This group includes the scissile G**p**A bond (*40*). The complete absence of interference (*38*) at one phosphate within the recognition site (GA**p**ATTC) supports the view that interference reflects functionally significant contacts rather than mere steric hindrance.

For the isosteric ^{7C}**G**AATTC (Fig. 6D), ^{2A}**P**AATTC and GA-^{m}**A**TTC sites, the pattern of strong interferences at the three clamp phosphates on each strand is the same as that of the GAATTC site (Fig. 6A) and thermodynamic analysis (*30*) of counter-ion displacement from DNA phosphates during endonuclease binding confirms that the number of cations released (8 ± 1) is the same as for the GAATTC site (*16, 17, 19*). The only difference is the enhanced interference (Fig. 6D, diagonal stripes) at the scissile G**p**A bond. This local adaptation may destabilize the transition state and thus change the activation energy ($\Delta\Delta G°‡$) by +1.4 kcal/mol (Table 4). [Complete loss of interference at this phosphate is correlated with stabilization of the transition-state interaction ($\Delta\Delta G°_I‡ = -1.7$ kcal/mol) in a half-site containing a sterospecific phosphorothioate [Sp-GAA**p(s)**TTC, where **p(s)** denotes the phosphorothioate (*37*)]. Thus all three clamp contacts on each strand can be made at base analogue sites that preserve the sequence-dependent backbone conformation of the GAATTC site.

Altered phosphate contacts in discrimination. By contrast, the pattern of phosphate contacts in all complexes with one incorrect base pair (for example, **A**AATTC, Figs. 5B and 6B) differs significantly from that in the canonical complex. To a first approximation, the altered pattern is the same for all complexes with one incorrect base pair (Figs. 5 and 6B). The distinguishing feature of this pattern is asymmetric loss of two "supplementary" clamp contacts (for example, **p**AAAA**p**TTC, recognition site underlined) in the strand containing the altered half-site. The virtual disappearance of interference at these points (Fig. 5C, upper; Fig. 6B, lower) implies that bound protein can accommodate an ethyl group at either phosphate. Loss of these contacts makes a crucial energetic contribution to discrimination between canonical and variant sites. Thermodynamic methods (*31*) confirm that the variant complexes have altered phosphate contacts. For every site with one incorrect base pair (*30*), a net of 2 ± 0.7 fewer cations are released than in formation of the GAATTC complex.

In all the variant complexes, the clamps immediately upstream of the recognition sequence (**p**AAATTC) are still the points of strongest interference in both DNA strands (Fig. 5 and Fig. 6, B and C). These "primary" clamps are indispensable to binding for canonical, variant, and isosteric sites, and are thus an essential part of recognition of bases in the hexanucleotide site.

Increased interference occurs at some points in the variant complexes. On the complementary strand of the **A**AATTC site, the phosphate upstream at **p**CAGAATTT (Fig. 5C, lower) shows strong interference, suggesting increased interaction with this phosphate. The phosphate at the scissile bond shows increased interference in both half-sites of the variant complexes (Fig. 5, C and D; Fig. 6B, diagonal stripes). This may reflect a local structural change

that destabilizes the transition state complex, thus contributing to discrimination. The phosphate at AAATTCpT (Fig. 5C, upper) shows strong interference, where there was none in the canonical or isosteric complexes. This phosphate (Fig. 6B, horizontal stripes) lies directly across the DNA helix from the abnormally strong cluster of contacts in the GAA half-site of this incorrect complex. The DNA in the incorrect complex may be asymmetrically kinked as a result of loss of the supplementary clamp contacts in the variant half-site (**p**AAAA**p**TTC) and enhanced contacts in the GAA half-site (**p**C-**p**A**p**GAATTT). A new contact with the phosphate at AAATTCpT may be made as part of this adjustment, or an ethyl group here may interfere sterically with DNA distortion in the "over-kinked" GAA half-site.

Asymmetry of phosphate contacts and cleavage rates. The phosphate contacts on the two DNA strands of the variant complexes show striking asymmetry (Fig. 6, compare B to A). There is a strong cluster of contacts in the GAA half-site and loss of the two supplementary clamp contacts in the AAA half-site. A "negative interference" at **p**CAAAATTC (Fig. 5C, upper) suggests that binding is enhanced when the protein cannot contact this phosphate, consistent with the adaptations to increase contact in the GAA half-site.

In eight of nine complexes with one incorrect base pair, the loss of phosphate contacts in the variant half-site can be correlated with greatly reduced cleavage rates relative to the GAA half-site ($k_1 > k_2$). However, in the GACTTC site, an anomalously high k_2 for cleavage in the GAC half-site ($k_2 > k_1$) (Table 3) is correlated with retention of the contacts at **p**AGAC**p**TTC (Fig. 6C, lower). The contacts in the GAA half-site show the same changes as in the other eight variant sites (Fig. 6, B and C, upper). These correlations imply that alterations of phosphate contacts in the E-DNA complexes also affect the transition-state complexes. Furthermore, ethylation-interference footprints on slowly cleaved sites (GTATTC and GA**G**TTC) are the same in the presence and absence of Mg^{2+} (*41*), so we need not invoke any major rearrangements of the E-DNA complex upon binding Mg^{2+}.

Invisible energy in recognition and discrimination. The observed binding free energy ($\Delta G°_{ED}$) at GAATTC is −16 kcal/mol (calculated from K_A, Table 2), representing the net of all favorable and unfavorable factors in complex formation. Identifiable protein-base and protein-phosphate interactions make favorable contributions; by estimating their sum in the canonical complex, we can estimate (by difference) the magnitude and sign of ΔG_{reorg} (Fig. 3B), which represents all other factors taken together. Each protein-base hydrogen bond contributes about −1.4 kcal/mol (Fig. 4, Table 4), each interaction to a thymine methyl group −1.2 to −1.7 kcal/mol, and each protein-phosphate interaction (including counter-ion release) about −1.3 kcal/mol (*19, 39*), indicating a sum ($\Delta G_{base} + \Delta G_{phos}$) of the order of −36 kcal/mol. Thus ΔG_{reorg} makes a net unfavorable contribution to binding energy.

The energy represented by ΔG_{reorg} is itself an aggregate. There are favorable contributions from water release on binding (*42*) and the formation of electrostatic networks (*3*) that interconnect the protein-base and protein-phosphate interactions. These do not completely compensate for the unfavorable losses of translational, rotational and vibrational entropies upon binding, including the energy required to stabilize the kinked DNA conformation in the canonical complex (*3*).

Cleavage of any DNA site requires that the enzyme and DNA achieve a precise transition state configuration in the immediate vicinity of the scissile phosphodiester bond, but for a macromolecular substrate it is not known whether the global transition state properties (including all aspects of enzyme and substrate conformations) must be rigorously conserved. For sites with one incorrect base-pair, there are two alternative views of the global properties of the transition state complexes. (i) All transition state complexes must attain precisely the same protein-phosphate contacts and overall DNA conformation as in the canonical GAATTC complex, but the energetic cost ($\Delta G°_{I}‡$) of attaining these is much greater. (ii) Transition state complexes for incorrect sites may differ from the canonical transition state complex in both global DNA conformation and the absence of some critical stabilizing protein-phosphate contacts, such that $\Delta G°_{I}‡$ is increased. These two constructs have the same operational consequences in that the term ($\Delta\Delta G_{phos} + \Delta\Delta G_{reorg}$) contributes to the large differentials in $\Delta\Delta G°_{I}‡$ (+6.6 to +10.3 kcal/mol, Fig. 3A, case 1). After accounting for these factors and their position-dependence, the small steps in $\Delta\Delta G°_{I}‡$ at each base position (1 to 2 kcal/mol, Fig. 3A, case 2) directly reflect the (stabilizing or destabilizing) realization of protein-base interaction differences in the transition state.

Direct and indirect readout of sequence information. The energy required for any DNA site to reach the transition state thus depends upon both direct readout (protein-base contacts) and indirect readout (protein-phosphate contacts and DNA conformation). These are completely interdependent for natural DNA sites, so that separation of ΔG_{base}, ΔG_{phos}, and ΔG_{reorg} should be regarded as only a conceptual convenience. The role of these interdependent factors in discrimination is best appreciated by considering the complexes formed at four classes of DNA sites.

1) The free GAATTC site in solution has been shown by two-dimensional NMR (*43*) to embody a predisposition to the kinked conformation; thus, a minimum (relative to all other sequences) of additional stabilization is required to attain the canonical transition state complex. Both the primary and supplementary clamps (**p**A**p**GAA**p**TTC) are formed, so that functional groups are placed in position to form the full complement of protein-base interactions.

2) Isosteric sites containing a base analogue preserve the conformation of the GAATTC motif, so that the supplementary clamps are formed (Fig. 6D) and the energy required to drive the DNA into the required conformation is little different from that for GAATTC. Thus, only direct readout is perturbed, the change in binding energy ($\Delta\Delta G°_{ED}$) reflects only $\Delta\Delta G_{base}$, and we can infer by difference the large contribution of indirect readout ($\Delta\Delta G_{phos} + \Delta\Delta G_{reorg}$ terms) to discrimination against nonisosteric sites. The aggregate energetic penalty ($\Delta\Delta G°_{I}‡$) for isosteric sites is position-independent, in contrast to the position-dependent effects of natural base pair changes (Fig. 3A).

3) Sites with one incorrect base pair have an altered DNA conformation, so that the supplementary clamps cannot form in the variant half-site. Loss of these contacts in turn increases the energetic cost of driving the DNA into the required transition state conformation. In these "adaptive" complexes, indirect readout (the sum $\Delta\Delta G_{phos} + \Delta\Delta G_{reorg}$) is perturbed in addition to direct readout ($\Delta\Delta G_{base}$). The "adaptive" complexes are asymmetric in phosphate contacts, in cleavage rates and probably in DNA kinking.

When an incorrect base pair is introduced at different positions within the site, perturbation of the same kind of base interaction (Fig. 3A, case 3) is associated with different aggregate energetic penalties ($\Delta\Delta G°_{I}‡$). This is presumably because base substitutions at different positions have different effects on DNA conformation (and thus on $\Delta\Delta G_{reorg}$).

The interdependent changes in protein-base, protein-phosphate, and DNA-conformation terms make it impossible to account on a contact-by-contact basis for energetic differences when comparing "adaptive" complexes to either the canonical or the nonspecific complex (below). Similar nonadditive effects have been observed for mutant *lac* operators (*44*). In the consensus half-sites of the λ and *cro*

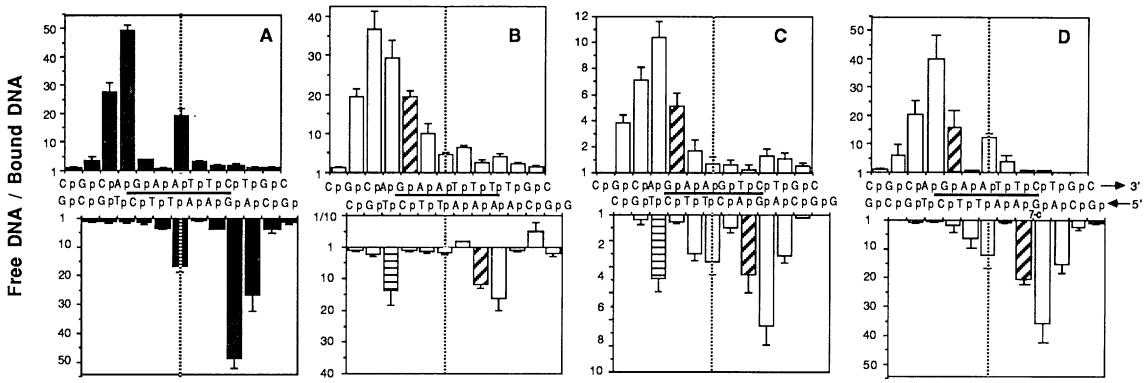

Fig. 6. Symmetries of the footprints on the sites GAATTC (**A**), AAATTC (**B**), GACTTC (**C**), and ^{7C}GAATTC (**D**). Lower panels contain variant half-sites. Data show means ± SD of five to eight independent experiments. The central hexanucleotide is underlined. The interference axis (ratio of free to bound ethylated DNA) is positive in both directions. Dotted lines indicate the center of symmetry in the canonicalcomplex. Horizontally striped bars (Band C) show the novel interference to the 3′-side of the hexanucleotide site on the variant strand. Diagonally striped bars (B to D) emphasize increased interference at the scissile bonds.

operators (*2*), single base pair changes produce large energetic penalties, providing high discrimination between the set of related operator sites and nonspecific DNA. We believe this reflects perturbation of both direct and indirect readouts. However, single base pair changes at five positions (three of which involve only aliphatic contacts to thymine methyl groups) in the nonconsensus half-sites of the bacteriophage operators produce smaller effects on binding energy (0.2 to 1 kcal/mol), with little or no energetic distinction between different incorrect bases (*2*). This permits graded discrimination between related operator sites.

4) At sites that differ from GAATTC by two or more base pairs, the energy required to attain the kinked conformation is too large to be overcome, so that only nonspecific complexes are formed. In these complexes, binding is completely delocalized along the DNA (*45*), no cleavage occurs, and the probable absence of DNA distortion removes an unfavorable contribution to the apparent binding energy. There is essentially no nonelectrostatic component of binding free energy (*30, 31*), suggesting no hydrogen-bonding to the bases, but there is about one more electrostatic interaction with DNA phosphate than in the canonical complex (*30*). Because the nonspecific complexes are nearly isoenergetic (Table 2) to the "adaptive" complexes, binding at nonspecific sites affects both the measured binding (K_A) and cleavage (k_1) of oligonucleotides containing sites with one incorrect base pair. Increased partitioning into the nonspecific mode is reflected in weakened intensities of the interference patterns (Fig. 5D); this partitioning factor varies about fivefold from the top (TAATTC) to the bottom (GTATTC) of the energetic hierarchy.

In vivo sequence discrimination. The GAATTC sites (about 1000) in the *E. coli* genome are protected by methylation (*46*) in both strands [$k_1 \times K_A$ is reduced by a factor of 3×10^8 (*27*)], but there are also about 18,000 unmethylated sites (*14, 47*) that differ by only a single base pair [Eco RI* sites (*23*)]. Our data provide insight as to how Eco RI endonuclease avoids lethal cleavage of incorrect DNA sites.

It is inappropriate to infer in vivo specificity from k_{cat}/K_M values measured in vitro. In an *E. coli* cell, reaction proceeds at high concentrations of endonuclease (*48*) and nonspecific genomic DNA sites. At Eco RI* sites, the low cleavage rate constants (Table 3) relative to the dissociation rate constant are such that reaction at these sites is equilibrium-controlled (Michaelis-Menten kinetics apply). By contrast, the small numbers of unmethylated GAATTC sites in vivo (one to ten on an incoming plasmid or bacteriophage genome) make it meaningless to apply either steady-state (Briggs-Haldane) or equilibrium (Michaelis-Menten) assumptions used in vitro under conditions of multiple turnover. Furthermore, the relevant biological point is functional disruption of a DNA region by cleavage, rather than product accumulation by multiple turnover.

The relative probability of binding and first-strand cleavage at an Eco RI* site is best represented by $k_1 \times K_A$. Even for the best Eco RI* site TAATTC (*49, 50*), this probability is 50,000 times lower (Table 2) than at GAATTC. Summing over all Eco RI* sites, using the estimated in vivo concentration of endonuclease (*48*), and accounting for competition by nonspecific DNA, we calculate that the *E. coli* genome would sustain about 20 single-strand nicks in a 30-minute generation (*51*). However, single-strand nicks at Eco RI* sites do not progress to double-strand cuts. The rate constant for second-strand cleavage (k_3) is about 3×10^{-5} sec^{-1} for the best of the sites, compared to a dissociation rate constant (k_d) that is at least 20,000 times faster (*52*). The enzyme therefore dissociates from Eco RI* sites before making a second-strand cut (*49, 50*), and the nick may then be repaired by DNA ligase (*53*).

At sites that differ from GAATTC by two or more base pairs, the enzyme forms only nonspecific complexes that are cleavage-incompetent (Table 2). Such sites act as effective competitive inhibitors of cleavage at Eco RI* sites in vivo because they are in large excess and bind endonuclease nearly as well as Eco RI* sites (Table 2).

General principles of stringent discrimination. In the canonical GAATTC complex, formation of the protein-base and protein-phosphate contacts is inextricably associated with the correct sequence-dependent DNA conformation. In this sense the direct and indirect readouts of sequence information are inseparable.

The incremental energetic contribution of each protein-base hydrogen bond is 1 to 2 kcal/mol. There is no need to postulate hydrogen bonds of unusual strengths. The large $\Delta\Delta G°_I‡$ required for stringent discrimination against sites with one incorrect base-pair represents not only the loss of one or two protein-base hydrogen bonds or interactions with thymine methyl groups, but also includes a number of other factors. For example, binding at some incorrect sites includes destabilizing interactions that have no energetic counterpart in the canonical complex, as anticipated by von Hippel and Berg (*8*). A misplaced functional group exacts an additional penalty (1 to 2 kcal/mol) due to steric repulsion. In addition, for "adaptive" complexes some of the canonical set of protein-phosphate interactions do not form and there are increased energetic costs associated with achieving the required conformations of DNA and protein. Different base pair substitutions appear to have nonadditive energetic effects (*24, 44*), because the free energy loss attending each substitution derives in large part from phosphate-interaction or conformational terms that should not be "counted" more than once.

These same principles should apply to stringent discrimination by nonenzymatic repressors between a set of related operator sites and

nonoperator DNA. Permissive discrimination among the set of recognized sites (multiple operators) may result when variable base positions have little effect on indirect readout or make protein contacts that are not cooperative with other elements of the DNA-protein interface. Even variable bases not directly contacted by protein can modulate discrimination (*54*) by affecting DNA conformation.

Note added in proof. After this paper was submitted for publication, the structural model (*3*) of the protein in the Eco RI endonuclease-DNA complex was revised. Our experimental approach to the energetics of protein-DNA interaction is based on systematically varying the DNA site rather than the protein, and thus our data and interpretations are unaffected by any revisions in the protein structure. We use no designations of specific amino acid residues in this paper. Our citations of the earlier structural model (*3*) refer to those features (notably the "kinked" DNA in the complex) that have remained the same according to the current information (*4*). Our view of specificity goes beyond the protein-base contacts emphasized in earlier models (*3, 23*) because we accord crucial roles in sequence discrimination to sequence-dependent differences in DNA conformation and protein-phosphate contacts.

REFERENCES AND NOTES

1. T. Maniatis *et al.*, *Cell* **5**, 109 (1975); A. Johnson, B. J. Meyer, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1783 (1978); A. D. Johnson *et al.*, *Nature* **294**, 217 (1981); Y. Takeda, D. H. Ohlendorf, W. F. Anderson, B. W. Matthews, *Science* **221**, 1020 (1983); M. Ptashne, *A Genetic Switch* (Cell Press, Cambridge, MA, 1986); N. Benson, P. Sugiono, P. Youderian, *Genetics* **188**, 121 (1988); N. Benson and P. Youderian, *ibid.* **121**.
2. Y. Takeda, A. Sarai, V. Rivera, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 439 (1989); A. Sarai and Y. Takeda, *ibid.*, p. 6513.
3. J. A. McClarin *et al.*, *Science* **234**, 1526 (1986); C. A. Frederick *et al.*, *Nature* **309**, 327 (1984).
4. Y. Kim, J. C. Grable, R. Love, P. J. Greene, J. M. Rosenberg, *Science* **249**, 1307 (1990).
5. J. E. Anderson, M. Ptashne, S. C. Harrison, *Nature* **326**, 846 (1987); S. R. Jordan and C. O. Pabo, *Science* **242**, 893 (1988); A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, S. C. Harrison, *ibid.*, p. 899; C. Wolberger, Y.-C. Dong, M. Ptashne, S. C. Harrison, *Nature* **335**, 789 (1988).
6. A. Otwinowski *et al.*, *Nature* **336**, 321 (1988).
7. H. R. Drew *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2179 (1981); R. E. Dickerson and H. R. Drew, *J. Mol. Biol.* **149**, 761 (1981); R. E. Dickerson, *ibid.* **166**, 419 (1983); D. J. Patel, L. Shapiro, D. Hare, *Annu. Rev. Biophys. Chem.* **16**, 423 (1987).
8. P. H. von Hippel and O. G. Berg, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1608 (1986).
9. M. C. Needels *et al.*, *ibid.* **86**, 3579 (1989); J. Heitman and P. Model, *Proteins: Struct. Funct. Genet.* **7**, 185 (1990). The logic of these inferences depends on the identification (*3*) of Arg^{200} as the side chain interacting with guanine and the assumption that no additional side chain interacts with guanine in either wild-type or mutant proteins.
10. All binding and cleavage-rate measurements and footprinting analyses were made at 25°C on the same series of duplex 17-nt oligomers; the measured T_m of the 17-nt oligomer containing a **T**AATTC site is 66°C. Others have used octamers or decamers that are not entirely in duplex form, may suffer from thermal "fraying" [D. J. Patel *et al.*, *Biochemistry* **21**, 428 (1982)] of the ends of the duplexes, and lack some of the phosphate contacts outside the recognition site.
11. A. R. Fersht *et al.*, *Nature* **314**, 235 (1985).
12. R. Wolfenden, *ibid.* **223**, 704 (1969); G. E. Lienhard, *Science* **180**, 149 (1973); R. L. Schowen, in *Transition States of Biochemical Processes*, R. D. Gandour and R. L. Schowen, Eds. (Plenum Press, New York, 1978), p. 77; M. I. Page, in *The Chemistry of Enzyme Action*, M. I. Page, Ed. (Elsevier, Amsterdam, 1984), chap. 1; A. Fersht, in *Enzyme Structure and Mechanism* (Freeman, New York, ed. 2, 1985), chap. 2, 3, 12.
13. P. Modrich and D. Zabel, *J. Biol. Chem.* **251**, 5866 (1976).
14. B. J. Terry, W. E. Jack, P. Modrich, in *Gene Amplification and Analysis*, vol. 5, *Restriction Endonucleases and Methylases*, J. G. Chirikjian, Ed. (Elsevier/North-Holland, Amsterdam, 1987), p. 103.
15. S. E. Halford and N. P. Johnson, *Biochem. J.* **191**, 593 (1980).
16. B. J. Terry, W. E. Jack, R. A. Rubin, P. Modrich, *J. Biol. Chem.* **258**, 9820 (1983).
17. L. Jen-Jacobson *et al.*, *ibid.*, p. 14638.
18. $K_M = k_c/k_a + k_d/k_a$, where k_a is the association rate constant and k_d is the dissociation rate constant; for Eco RI $k_c >> k_d$, and therefore K_M differs substantially from the equilibrium dissociation constant for the E–DNA complex; see Fersht in (*12*, p. 101). K_M determined in vitro with catalytic enzyme concentrations is of dubious relevance to the in vivo situation of high enzyme and nonspecific DNA concentrations, and small numbers of substrate sites.
19. L. Jen-Jacobson, D. Lesser, M. Kurpiewski, *Cell* **45**, 619 (1986).
20. D. R. Lesser, thesis, University of Pittsburgh, 1988.
21. Observed rate constants are virtually independent of salt concentration in the range 0.04 to 0.22 M, in accord with observations of others (*14*).
22. The side chains involved in the catalytic mechanism have not been identified, so that the structure of the transition state in the local vicinity of the scissile phosphodiester bond is unknown.
23. The quantitative hierarchical relation between substitutions at each base position resembles the semiquantitative hierarchy proposed by J. M. Rosenberg and P. Greene [*DNA* **1**, 117 (1982)] from fragmentary data on frequencies of double-strand scissions at Eco RI* sites [B. Polisky *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3310 (1975); R. C. Gardner, A. J. Hovarth, J. Messing, R. J. Shepherd, *DNA* **1**, 109 (1982)]. However, the earlier proposals did not include the effects of unfavorable appositions (Table 1).
24. W. P. Jencks, *Adv. Enzymol.* **43**, 219 (1975); *Proc. Natl. Acad. Sci. U.S.A.* **78**, 4046 (1981).
25. A. Fersht, *Trends Biochem Sci.* **12**, 301 (1987); *Biochemistry* **27**, 1577 (1988).
26. Oligonucleotides containing 2-aminopurine (^{2A}P) substitutions were synthesized by T. Waters and B. A. Connolly (University of Southampton, U.K.) We have made detailed thermodynamic, kinetic, and footprinting analyses of complexes containing this and other purine base analogues (7-deazaadenine; purine) (D. Lesser, M. Kurpiewski, T. Waters, B. A. Connolly, L. Jen-Jacobson, in preparation) and pyrimidine analogues (uracil, 5-bromocytosine) (D. Lesser, M. Kurpiewski, L. Jen-Jacobson, in preparation) in all possible positions.
27. Methylating the inner adenine on both strands does not perturb the DNA conformation relative to the GAATTC site [C. A. Frederick *et al.*, *J. Biol. Chem.* **263**, 17872 (1988)]. However, $\Delta\Delta G^\circ_I\ddagger$ for a site with both of the inner adenines methylated increases to +11.5 kcal/mole (D. Lesser, M. Kurpiewski, C. Yee, L. Jen-Jacobson, unpublished results).
28. C. A. Brennan, M. D. VanCleve, R. I. Gumport, *J. Biol. Chem.* **16**, 7270 (1986); A. Fliess *et al.*, *Nucleic Acids Res.* **14**, 3463 (1986); L. W. McLaughlin, F. Benseler, E. Graser, N. Piel, S. Scholtissek, *Biochemistry* **26**, 7238 (1987).
29. W. N. Hunter, T. Brown, O. Kennard, *J. Biomolec. Struct. Dynam.* **4**, 173 (1986).
30. Analysis of the salt dependence of binding (*31*) shows that the net number of cations released is 8 ± 0.5 for GAATTC (*16, 17, 19*), 8 ± 1 for the isosteric sites, and 6 ± 0.5 for all sites with a single incorrect base pair. The binding energy (at 0.2 M salt) is about 50 percent salt-sensitive (electrostatic) and 50 percent salt-insensitive (nonelectrostatic) for both GAATTC sites and single-substituted sites. By contrast, binding to nonspecific DNA is >90 percent electrostatic (9 ± 1 cations released) by these criteria (D. Lesser and L. Jen-Jacobson, in preparation).
31. M. T. Record, Jr., T. M. Lohman, P. L. deHaseth, *J. Mol. Biol.* **107**, 145 (1976); M. T. Record, Jr., A. F. Anderson, T. M. Lohman, *Q. Rev. Biophys.* **11**, 103 (1978).
32. Formation of the **A**AATTC complex is an exchange reaction where the hydrogen bond between water and donor 1 is replaced by the hydrogen bond to N^7 of adenine. If donor 1 on the protein retains its hydrogen bond to water in the **C**AATTC complex, the apparent $\Delta\Delta G^\circ_I\ddagger$ may be identified with a true incremental binding energy.
33. Method described in A. Brunelle and R. F. Schleif, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6673 (1987); M. Kurpiewski and L. Jen-Jacobson, in preparation.
34. A.-L. Lu, W. E. Jack, P. Modrich, *J. Biol. Chem.* **256**, 13200 (1981).
35. M. M. Becker, D. Lesser, M. Kurpiewski, A. Baranger, L. Jen-Jacobson, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6247 (1988).
36. S. D. Yanofsky *et al.*, *Proteins* **2**, 273 (1987).
37. W. J. Stec, in *Reviews on Heteroatom Chemistry*, S. Oae, Ed. (MYU, Tokyo, 1988), vol. 1, p. 367; D. Lesser, M. Kurpiewski, M. Koziolkiewicz, W. J. Stec, L. Jen-Jacobson, in preparation.
38. The degree of interference depends on how ethylation at a single phosphate perturbs the equilibrium dissociation constant K_D and on protein and DNA concentrations. Interferences at **pAp**GAA**p**TTC increase with decreasing concentrations of DNA (*35*), reflecting a large change in K_D for ethylated DNA ($K_D \sim 10$ nM, compared to 1 pM for unethylated DNA). Interferences at **pC**AG**p**AAT**p**T**p**C are weak and nearly independent of DNA concentration; the lowest ethylated DNA concentration used (50 pM) is sufficient to give substantial binding, indicating a smaller change (50- to 100-fold) in K_D. No interference at GA**p**ATTC is observed at any DNA concentration.
39. T. M. Lohman, P. L. de Haseth, M. T. Record, *Biochemistry* **19**, 3522 (1980).
40. Ethylation of the scissile phosphate at G**p**AATTC has only a weak effect on binding, but strongly inhibits cleavage even at enzyme and DNA concentrations high enough to overcome any effect on enzyme-DNA binding. Ethylation at GA**p**ATTC has no effect at all on binding, but strongly inhibits cleavage. Ethylation of the "primary clamp" phosphates **p**GAATTC inhibits both binding and cleavage; (M. Kurpiewski and L. Jen-Jacobson, unpublished observations).
41. M. Kurpiewski and L. Jen-Jacobson, in preparation.
42. A large negative ΔC°_p (standard heat capacity change) has been measured for the Eco RI endonuclease-cognate DNA interaction and is largely attributable to the hydrophobic effect" (water release from nonpolar surfaces upon binding); the contribution of this favorable "driving force" to binding free energy can be as much as −100 kcal/mol [J.-H. Ha, R. S. Spolar, M. T. Record, Jr., *J. Mol. Biol.* **209**, 801 (1989)].
43. Interproton distances calculated from two-dimensional (2D)–NMR [W. Nerdal, D. R. Hare, B. R. Reid, *Biochemistry* **28**, 10008 (1989)] indicate that CGCGAATTCGCG in aqueous solution has "type I" and "type II" kinks strongly resembling those in the crystal structure of the endonuclease-DNA complex (*3*). Discontinuities in backbone conformation at these points were also observed by ^{31}P-NMR [J. Ott and F. Eckstein, *Biochemistry* **24**, 2530 (1985)]. Although unusually low helical twist angles occur at the kinks in the crystal structure of the same dodecamer (*7*), the free DNA shows more pronounced kinks in solution than it does in the crystalline state.
44. M. C. Mossing and M. T. Record, Jr., *J. Mol. Biol.* **186**, 295 (1985); M. T.

Record, Jr., and M. C. Mossing, in *RNA Polymerase and Regulation of Transcription*, W. Reznikoff *et al.*, Eds. (Elsevier, New York, 1987), pp. 62–83.
45. Complexes with ethylated nonspecific DNA can be isolated by nondenaturing gel electrophoresis, but show no discrete (localized) footprints.
46. A. Dugaiczyk, J. Hegepeth, H. W. Boyer, H. M. Goodman, *Biochemistry* **13**, 503 (1974).
47. P. H. von Hippel, in *Biological Regulation and Development*, R. F. Goldberger, Ed. (Plenum, New York, 1979), vol. 1, p. 279; T. D. Schneider, G. D. Stormo, L. Gold, A. Ehrenfeucht, *J. Mol. Biol.* **188**, 415 (1986).
48. In an *E. coli* cell, the concentration of Eco RI endonuclease dimers is 0.8 μM (*13*) and the concentration of nonspecific DNA sites is ~0.01 M (*46*). At 0.2 M salt, these nonspecific competitors increase the apparent K_D of GAATTC sites to ~2.5 μM.
49. In accord with this analysis, it has been observed that plasmid pCM21 DNA, which contains several Eco RI sites, is nicked at the GAATTA site [J. O. Bishop, *J. Mol. Biol.* **128**, 545 (1979)], which has the best k_1 of any variant site (Table 3). The GAATTA site is also nicked in a pBR322 derivative lacking a GAATTC site (*50*).
50. K. King, S. J. Benkovic, P. Modrich, *J. Biol. Chem.* **264**, 11807 (1989).
51. We have "reconstructed" this situation in vitro, using near-physiological concentrations of endonuclease (1.6 μM), duplex [5′-^{32}P]TCGCAGAATTATGCCGG (3.2 μM), and a mixture containing all other Eco RI* sites (29 μM), 0.2 M NaCl, pH 7.5, 37°C. Nonspecific DNA was included at 0.2 mM. We observed only nicks in the GAA half-site, at a velocity of 2 μM/sec, corresponding in vivo (for an *E. coli* cell of volume 10^{-15} liter) to about seven nicks per 30-minute generation.
52. The measured dissociation rate constant (k_d) for the endonuclease-TAATTC complex at 0.08 M NaCl is ~0.7 sec^{-1}. At physiological salt (~0.2 M), k_d should be increased about 30-fold (*19*), whereas cleavage-rate constants are salt-independent (*14, 20*), so k_d/k_3 should be about 6×10^5.
53. P. Modrich, Y. Anraku, I. R. Lehman, *J. Biol. Chem.* **248**, 7495 (1973); A. Sugino *et al.*, *ibid.* **252**, 3987 (1977); J. Heitman, N. D. Zinder, P. Model, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2281 (1989); J. D. Taylor and S. E. Halford, *Biochemistry* **28**, 6198 (1989).
54. G. B. Koudelka, S. C. Harrison, M. Ptashne, *Nature* **326**, 886 (1987); G. B. Koudelka, P. Harbury, S. C. Harrison, M. Ptashne, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 4633 (1988).
55. N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804 (1976).
56. S.-Y. Lin and A. D. Riggs, *J. Mol. Biol.* **72**, 671 (1972).
57. M. M. Garner and A. Revzin, *Nucleic Acids Res.* **9**, 3047 (1981); M. Fried and D. M. Crothers, *ibid.*, p. 6505.
58. The protected 7-deaza-2′-deoxyguanosine β-cyanoethyl phosphoramidite (^{7C}G) was synthesized by P. A. Gottlieb. For synthetic method and early results on Eco RI cleavage of palindromic 8-bp oligomers containing ^{7C}G, see F. Seela, H. Driller, *Nucleic Acids Res.* **14**, 2319 (1986).
59. U. Siebenlist and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 122 (1980).
60. We thank C. Yee for technical assistance; P. A. Gottlieb for the ^{7C}G phosphoramidite; J. S. Franzen for assistance in DNA melting studies; B. R. Reid, M. T. Record, Jr., and J. M. Rosenberg for communicating results prior to publication; and P. Modrich, C. O. Pabo, M. T. Record, Jr., J. M. Rosenberg, and R. S. Spolar for comments on the manuscript. Supported by NIH research grant GM-29207.

3 November 1989; accepted 10 May 1990

Cell Cycle Control of DNA Replication by a Homologue from Human Cells of the p34^{cdc2} Protein Kinase

GENNARO D'URSO, ROBERT L. MARRACCINO, DANIEL R. MARSHAK, JAMES M. ROBERTS

The regulation of DNA replication during the eukaryotic cell cycle was studied in a system where cell free replication of simian virus 40 (SV40) DNA was used as a model for chromosome replication. A factor, RF-S, was partially purified from human S phase cells based on its ability to activate DNA replication in extracts from G_1 cells. RF-S contained a human homologue of the *Schizosaccharomyces pombe* p34^{cdc2} kinase, and this kinase was necessary for RF-S activity. The limiting step in activation of the p34 kinase at the G_1 to S transition may be its association with a cyclin since addition of cyclin A to a G_1 extract was sufficient to start DNA replication. These observations suggest that the role of p34^{cdc2} in controlling the start of DNA synthesis has been conserved in evolution.

THE TIMING OF EVENTS DURING THE SOMATIC CELL CYCLE can be explained by the hypothesis that any one event does not begin until certain previous ones have been completed (*1*). Consequently, mutations that prevent particular cell cycle functions can cause cell cycle arrest (*2*). The arrest points often coincide with the time in the cell cycle when the affected gene products are required. Mutations in the *Saccharomyces cerevisiae CDC28* gene (*3*), or the *Schizosaccharomyces pombe cdc2* gene (*4*) both arrest the cell cycle at two points—just before the onset of DNA synthesis and just before entry into mitosis. In *S. cerevisiae*, the control point at the G_1-S boundary is called START and represents the major point at which cell growth and division are coordinated (*5*). In *S. pombe* the major coordination point for growth and division is at G_2-M, the second point in the cell cycle regulated by *cdc2*, although under certain conditions growth and division can also be coordinated at G_1-S (*6*). The *cdc2* and *CDC28* genes are homologous (*7*) and functionally interchangeable (*8*). They encode a 34-kD serine-threonine protein kinase, designated p34 (*9*).

The mitotic function of p34 has been evolutionarily conserved. Maturation promoting factor (MPF), initially isolated from vertebrate eggs, can induce meiotic maturation of G_2-arrested oocytes without new protein synthesis (*10*). MPF was later isolated from somatic cells of various species and shown to be active during mitosis, but inactive or latent during interphase (*11*). One component of human MPF is a 34-kD serine-threonine kinase that is 63 percent identical (amino acid sequence) to the *cdc2* protein kinase (p34^{cdc2}) of *S. pombe* (*12*). Furthermore, the human *cdc2* gene can rescue the mitotic function of *cdc2* in *S. pombe cdc2* mutants (*13*).

It is not known whether the G_1-S function of p34 occurs in higher eukaryotes. In mammalian cells, as in *S. cerevisiae*, the start of DNA synthesis reflects a commitment to complete the remainder of the cell division cycle (*14*). Hence, the focus of many regulatory signals that promote or inhibit cell duplication ultimately is on the transition from G_1 into S. Cells that cease proliferation usually exit during G_1 and enter a resting state called G_0 (*15*). The kinetics of

G. D'Urso is in the Department of Basic Sciences, Fred Hutchinson Cancer Center, 1124 Columbia Street, Seattle, WA 98104, and the Department of Pathology, University of Washington, Seattle, WA 98195. R. L. Marraccino is in the Department of Basic Sciences, Fred Hutchinson Cancer Center, 1124 Columbia Street, Seattle, WA 98104. D. R. Marshak is at the Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724. J. M. Roberts is in the Department of Basic Sciences, Fred Hutchinson Cancer Center, 1124 Columbia Street, Seattle, WA 98104, and the Department of Biochemistry, University of Washington, Seattle, WA 98195.