Protein—DNA recognition complexes: Conservation of structure and binding energy in the transition state

ARTICLE *in* BIOPOLYMERS · JANUARY 1997

Impact Factor: 2.39 · DOI: 10.1002/(SICI)1097-0282(1997)44:2<153::AID-BIP4>3.0.CO;2-U · Source: PubMed

1 AUTHOR:

Linda Jen-Jacobson
University of Pittsburgh
53 PUBLICATIONS **2,013** CITATIONS

Linda Jen-Jacobson
Department of
Biological Sciences
University of Pittsburgh
Pittsburgh PA 15260

Protein–DNA Recognition Complexes: Conservation of Structure and Binding Energy in the Transition State

Abstract: *This paper considers how enzymes that catalyze reactions at specific DNA sites have been engineered to overcome the problem of competitive inhibition by excess nonspecific binding sites on DNA. The formation of a specific protein–DNA recognition complex is discussed from both structural and thermodynamic perspectives, and contrasted with formation of nonspecific complexes. Evidence (from Eco*RI *and Bam*HI *endonucleases) is presented that a wide variety of perturbations of the DNA substrate alter binding free energy but do not affect the free energy of activation for the chemical step; that is, many energetic factors contribute equally to the recognition complex and the transition-state complex. This implies that the specific recognition complex bears a close resemblance to the transition-state complex, such that very tight binding to the recognition site on the DNA substrate does not inhibit catalysis, but instead provides energy that is efficiently utilized along the path to the transition state. It is suggested that this view can be usefully extended to ''noncatalytic'' site-specific DNA-binding proteins like transcriptional activators and general transcription factors.* © *1997 John Wiley & Sons, Inc. Biopoly* ***44:*** *153–180, 1997*

Keywords: *protein–DNA recognition; binding energy; transition state; sequence-specific binding; nonspecific binding; adaptation; enzyme catalysis; thermodynamics*

INTRODUCTION

As it has become clear that gene expression is modulated in space and time to govern biological organization and function, it has become equally clear that we cannot understand this modulation without a fairly detailed knowledge of the specificities and functions of a large class of proteins that interact with particular DNA sites. Some such proteins are extremely specific for a particular DNA sequence (e.g., restriction nucleases, repressors), others are more permissively selective for a class of related DNA sequences (e.g., prokaryotic RNA polymerases binding to promoters with relatively loose ''consensus'' sequences). Once site selection has occurred, the function of a DNA-recognition protein is sometimes evidently catalytic (e.g., restriction endonucleases) or unambiguously noncatalytic (e.g., repressors), but some ''noncatalytic'' proteins may play a role in a catalytic event. For example, many transcription factors do not themselves catalyze the making or breaking of covalent bonds, but instead serve as organizing centers for the assembly of multiprotein complexes that include a catalytic protein (RNA polymerase). It seems somewhat arbitrary to declare such transcription factors to be ''noncatalytic,'' since their function may be necessary (albeit not sufficient) for catalysis, and we

Correspondence to: Linda Jen-Jacobson
Contract grant sponsor: NIH; contract grant number: GM-29207

CCC 0006-3525/97/020153-28

would likely not make a parallel judgment against a noncatalytic subunit that associates with a catalytic subunit on a more permanent basis. Indeed, if the DNA-recognition and assembly-organizing functions require a molecular precision comparable to that of a transition state, it may be heuristically useful to think about such a protein as we would think about an enzyme.

Taking this view of sequence-specific DNA-binding proteins, it becomes apparent that they have been engineered by evolution to deal with some problems not commonly encountered by other kinds of enzymes. The most challenging of these is that the substrate—a specific DNA site—is immersed in a huge molar excess of other DNA sequences that bear a strong chemical resemblance to the substrate. The problem differs fundamentally from the problem of specificity for small-molecule substrates, because a DNA binding protein cannot exclude these other sites from its ''active site'' or ''binding pocket'' on the basis of size or shape. These other DNA sites may be either very poor substrates or competitive inhibitors, but in either case they reduce the concentration of free protein and make accurate selection of the ''correct'' recognition site more difficult, as was pointed out by von Hippel et al.[1] and Lin and Riggs[2] over 20 years ago. In fact, there are only a few other examples of enzymes whose function requires them to carry out accurate substrate recognition and catalysis in the presence of such molecular distraction, and no other examples in which the concentration ratio of ''distractors'' to true substrates is so high. This paper will consider how this challenge is met by site-specific DNA-recognition proteins, drawing illustrative examples principally from work on restriction endonucleases in the author's laboratory, but emphasizing the generality of the protein-engineering principles that have emerged.

BINDING TO SPECIFIC AND NONSPECIFIC DNA SITES

All known site-specific and site-selective DNA-binding proteins also have a finite affinity for nonspecific DNA.[3–5] The term ''nonspecific binding'' may be defined as binding that is not localized to any particular site, but is equiprobable at any point along the DNA. The macroscopic or apparent binding constant alone is an insufficient criterion of ''nonspecific'' binding,[6,7] because it may be difficult to distinguish nonspecific from localized ''adaptive'' binding (see below) on this basis. It is preferable to have an experimental demonstration that the protein makes no localized ''footprint'' on the test DNA, using chemical interference methods.[7–9]

The sequence requirements of a ''nonspecific'' binding site may differ from protein to protein. For restriction endonucleases and *lac* repressor, a replacement of a *single* base pair (bp) in the recognition site (to produce a ''noncognate site'') may drastically reduce binding, yet does not necessarily reduce the binding constant to the level characteristic of totally nonspecific DNA binding. The ''noncognate binding mode'' differs from both the specific and the nonspecific modes.[6,7,9,11–13] On the other hand, for both *Eco*RI and *Bam*HI restriction endonucleases, sites that differ from their 6 bp cognate sites by *two* or more base pairs can be considered as nonspecific, since they are not cleaved, exhibit only delocalized footprints, and their binding constants are as low as those for any completely nonspecific DNA sequence.[6,13]

It is of interest to ask how well various proteins discriminate in binding between specific and nonspecific DNA sites. However, it is not simple to extract useful comparisons from the literature, in part because protein–DNA binding constants are very sensitive to solution conditions, including the nature and concentrations of cations and anions,[14–16] pH,[15,17,18] temperature,[19–26] and osmotic strength.[27–29] Furthermore, specific binding and nonspecific binding usually have different dependences on these variables.[7,15,27–29] A second problem in comparisons is that for nonspecific binding, the macroscopic or apparent binding constant depends on the length of the test DNA, because every base pair represents the beginning of a potential nonspecific binding site, so at a given molar concentration, a longer DNA molecule provides a higher concentration of nonspecific binding sites.[31] For nonspecific sites we therefore use the ''intrinsic'' binding constant, that is, the per-site binding constant. A third problem is that binding to the specific recognition site may be far more sensitive to surrounding DNA sequence than has hitherto been recognized. Our work on *Eco*RI endonuclease (Table I; cf. Figure 9) indicates that the three flanking base pairs on either side of the recognition site may alter the specific binding constant by as much as 500-fold. It is unclear whether or how to account for flanking-sequence effects in comparing different proteins to each other.

Table I compares site-specific and nonspecific DNA binding for a number of proteins. Allowing for the difficulties noted above, the following general conclusions emerge:

Table I Binding of Site-Specific Proteins to Specific Sites and Nonspecific DNA

	Specific Site	K_S (M^{-1})	K_{NS}[a] (M^{-1})	Ratio (K_S/K_{NS})	Conditions	Reference
Repressors						
λ Cro	O_R3	5.0×10^{11}	6.8×10^5	7.4×10^5	0.1*M* KCl; pH 7.4; 0°C	61
	O_R1	1.2×10^{11}		1.8×10^5		
	O_R2	8.3×10^9		1.2×10^4		
λ cI repressor	O_R1	1.3×10^9	1.2×10^4	1.1×10^5	0.2*M* KCl, 2 m*M* Ca^{2+}; pH 7.3; 22°C	60
	O_R1	8.3×10^9	2.5×10^3	3.3×10^6	0.15*M* KCl, 2.5 m*M* Mg^{2+}, 1 m*M* Ca^{2+}; pH 7.0; 20°C	30
	O_R2	1.5×10^9		6.0×10^5		
	O_R3	1.6×10^8		6.4×10^4		
lac Repressor	O^{sym}[b]	1.2×10^{11}	4.2×10^3	2.9×10^7	0.2*M* KCl; pH 7.3; 24°C	12
trp Repressor	Wild-type operator	4.5×10^8	2.4×10^4	1.9×10^4	0.25*M* KGlu, 0.15*M* NaCl, 20 m*M* Mg^{2+}; pH 7.5; 25°C; 2 m*M* tryptophan	22
MetJ repressor	Single metbox	6.2×10^5	2×10^3	3.1×10^2	0.05*M* KCl, 0.04*M* NaCl, 0.01*M* NaI, 5 m*M* Mg^{2+}; pH 7.0; 25°C	23
P22 Arc repressor (dimer to first half-site)	Half operator	3.2×10^9	3.8×10^5	8.4×10^3	0.1*M* KCl, 3 m*M* Mg^{2+}; pH 7.5; 20°C	62
(2nd dimer to 2nd half-site)	Full operator	6.7×10^{12}		1.8×10^7		
Restriction endonucleases						
*Eco*RV	GATATC	2.6×10^9	1.0×10^5	2.6×10^4	0.1*M* KCl; pH 7.0; 22°C	7
*Bam*HI	GGATCC	3.8×10^9	9.1×10^4	4.1×10^4	0.12*M* NaF; pH 7.3; 22°C	65
*Eco*RI	GAATTC	1.7×10^{11}	3.6×10^3[c]	4.7×10^7	0.18*M* NaCl; pH 7.3; 22°C	Seq. 1 (Figure 9)
		1.4×10^{10}		3.9×10^6		Seq. 2
		2.4×10^9		6.7×10^5		Seq. 4
		3.1×10^8		8.6×10^4		Seq. 6
Transcription factors						
Catabolite gene activator protein (CAP)	*lac* Promoter	8.4×10^{10}	7.3×10^5	1.2×10^5	No salt; pH 8.0; 20°C; 5 μM cAMP	142
		3.1×10^9	2×10^5	1.6×10^4	0.1*M* NaCl, 10 m*M* Mg^{2+}; pH 8.0; 37°C; 2 m*M* cAMP	143
	gal Promoter	2.6×10^8	2×10^5	1.3×10^3		
TATA binding protein	Adenovirus MLP	5×10^8	2×10^3	2.5×10^5	0.06*M* KCl; pH 8.0; 24°C	144
Integration host factor (IHF)	λ attP H′	6.1×10^8	3.5×10^5	1.7×10^3	0.06*M* KCl; pH 8.3; 25°C	32

[a] Intrinsic association constant for nonspecific sites. Where intrinsic constants were calculated by original authors, those values are shown; if not, values were calculated as $K_{NS} = (K_{obs})/[2(L - n)]$, where L is length of DNA fragment used and n is the crystallographically defined site size for the *specific* complex.

[b] O^{sym} consists of an inverted repeat of 10 bp of the left half of the wild-type operator.

[c] K_{NS} for the nonspecific CTTAAG site was the same when embedded in the four sequence contexts shown for the GAATTC site, within experimental error.

1. Reported equilibrium association constants for the specific DNA recognition sites (K_S) vary considerably, ranging from about $10^8 M^{-1}$ to $10^{11} M^{-1}$. It is impossible to tell how much of this range reflects true functional differences among the proteins and how much reflects different degrees of optimization of solution conditions in the various systems.
2. Intrinsic equilibrium association constants for nonspecific DNA (K_{NS}^{int}) are generally in the range 10^3–$10^5 M^{-1}$. Much of this apparent variation may reflect the steep dependence on salt concentration.
3. The ratio of intrinsic binding constants (K_S/K_{NS}^{int}) may cover a range from about 10^3-fold to about 10^7-fold. This ratio is often called the ''binding specificity ratio.'' [Note that for *Eco*RI endonuclease, specific binding is context sensitive (Table I), but the nonspecific binding constant is found to be invariant in these sequence contexts, so the binding specificity ratio can be modulated 500-fold by flanking sequence effects (Table I). This phenomenon has not been carefully studied for the other proteins.]

The significance of these relative affinities may be appreciated by considering the number of binding sites of each kind available under some representative conditions. At one limit, consider a hypothetical protein that recognizes a 6-bp site in a bacterial genome of 4×10^6 bp. There are about 1000 cognate sites, but about 4×10^6 nonspecific sites—a 4000-fold molar excess. To determine the relative amounts of protein bound to specific and nonspecific sites, we write a partition function of the form

$$\frac{[P \cdot DNA]_S}{[P \cdot DNA]_{NS}} = \frac{K_S[DNA]_S}{K_{NS}[DNA]_{NS}}$$

where subscripts S and NS respectively denote the specific and nonspecific sites or complexes. If K_S/K_{NS} is near the lower limit of 10^3, then about four times more protein is bound to nonspecific than to specific DNA sites. If K_S/K_{NS} is about 10^6, then specific binding should predominate by about a factor of 250.

At the other limit, consider a protein that recognizes a unique site in a mammalian genome of 3×10^9 bp, such that potential nonspecific binding sites are in excess to the specific site by a factor of 3×10^9. For $K_S/K_{NS} \approx 10^3$, only 1 in 3 million protein molecules will be bound at the correct recognition site. Even at the high end of the specificity range $K_S/K_{NS} \approx 10^6$, only about 1 in 3000 proteins will bind to the unique recognition site. (This may in part explain why site recognition in organisms with complex genomes often involves complexes of multiple proteins—their effective specificities will multiply!)

Is competitive nonspecific binding as dominant in vivo as these crude calculations suggest? Nash and co-workers[32] have used in vivo footprinting to analyze the intracellular occupancy of sites targeted by the site-specific *E. coli* integration host factor (IHF). They conclude that although the canonical target site has full occupancy and variant sites have varying degrees of occupancy, the protein is bound predominantly to nonspecific sites, and there is very little free protein in the cell.

The in vivo evidence makes it plain that the problem of competing nonspecific sites is indeed a serious one. We can divide the problem into two questions:

—How does the protein locate and bind the correct site?

—How does the protein achieve catalysis at the correct site?

CONCEPTS OF SPECIFICITY DETERMINATION

Early thought was dominated by ''direct readout'' models,[33,34] which proposed that proteins could distinguish unambiguously between DNA sites by a set of direct hydrogen-bonding contacts to the bases in the DNA major groove. As crystal structures of DNA–protein complexes were solved, some obvious common features emerged. In cognate complexes, proteins form extensive well-matched complementary surfaces with the edges of the DNA bases and with the sugar–phosphate backbone. There are direct hydrogen bonds and van der Waals contacts to bases,[35–42] primarily in the major groove and sometimes also in the minor groove.[39] Some proteins (e.g., *Eco*RI endonuclease) contact every base in the recognition site,[40] but in larger sites only some of the base pairs or one member of a pair are recognized.[36] Constrained protein–base contacts are made by the polypeptide backbone or short polar side chains. More flexible contacts may be made by long side chains, which may be held in precise position by interaction with nearby residues[36] (''buttressing'') or because the side-chain contacts two

groups on the same base (''bidentate'' recognition)[33,38] or adjacent bases (''bridging'').[36,40] In many cases, hydrogen bonding is mediated by specifically bound H_2O molecules.[43–47]

Protein–phosphate contacts were once considered to make only nonspecific (Coulombic) contributions to binding free energy. However, all cognate DNA–protein complexes show some key phosphates tightly constrained by hydrogen bonding from polypeptide backbone NH groups and/or short polar side chains. In many cases, a single amino acid sidechain contacts both a base and a phosphate and there are frequently ''networks'' of hydrogen bonds that connect the residues making phosphate contacts to each other or to residues contacting bases, suggesting that particular phosphate contacts have a role in specific recognition. In the *Eco*RI–DNA interface, contacts to six specific DNA phosphates appear to act as ''clamps'' to position base-recognition elements and stabilize the distorted DNA conformation.[6,42,48,49] In many specific complexes, phosphate contacts immediately 5′- to the recognition site and at the center of the site (or half-site) appear to fulfill this ''clamping'' role.[36,50–53]

It is now also generally recognized that the sequence-dependent conformational properties of DNA (see below) play an integral role in specificity for most sequence-specific proteins.[6,35–37,52–58] The sequence-dependent conformation of free DNA and (when bound DNA is distorted) its distortability will necessarily influence precise positioning of backbone and base functional groups for formation of protein–DNA complementarity. The term ''indirect readout''[43] is used to indicate that base sequence determines the work required to achieve the precise DNA conformation in the complex. The concepts of ''direct'' and ''indirect'' readouts of information are not mutually exclusive, but intimately intertwined determinants of specificity; most sequence-specific proteins make use of both, although the *relative* contributions may vary.

THERMODYNAMICS OF SPECIFIC AND NONSPECIFIC BINDING

There is no single site-specific DNA-binding protein for which we have a satisfyingly complete thermodynamic description of either the specific or the nonspecific interaction with DNA (see Ref. 59 for a comprehensive review). Some extensively studied systems are the *lac*,[11,12,14,15] *trp*,[21,22] λ cI,[18,24,58,60] λ Cro,[20,61] Arc,[62,63] and Mnt[64] repressors, and the *Eco*RI[6,9,16,54] and *Bam*HI endonucleases.[65] Although there are certainly significant differences between these systems, some informative general principles have emerged. One of these is that the overall favorable standard binding free energy (ΔG°_{bind}) is the net of large favorable and unfavorable contributions.

The case of the *Eco*RI endonuclease (Figure 1) illustrates this point. The ΔG°_{bind} for forming the ''specific'' recognition complex (in the absence of the Mg^{2+} cofactor) at a GAATTC site is about −15 kcal/mol at near-physiological salt concentration (0.18*M*). The following decomposition of ΔG°_{bind} should be understood as approximate, with uncertainties as high as ±50% in estimating some classes of contributions.

In this specific complex, the protein makes both hydrogen-bond contacts and nonpolar contacts with the bases at functional groups that lie in the DNA major groove.[40–42] It is not straightforward to extract the energetic contribution of each of these contacts from experiments, because the effect of even the most subtle modification of the protein or DNA may represent the net of opposing contributions, including the bond removed, perturbation of local or global protein structure, changes in solvation, and DNA conformational effects.[9] Nevertheless, judicious choice of base–analogue experiments indicates that each protein–base hydrogen bond or nonpolar contact contributes about −1.4 kcal/mol,[6,54] for a total contribution of about −30 kcal/mol.

The crystal structures of *Eco*RI–cognate DNA complexes with substrate and product[40–42] show that the protein makes ionic and/or hydrogen-bonding contacts with 14 DNA phosphates (Figure 2a); for six of these there is contact with both phosphoryl oxygens. Six of the 14 phosphates are contacted only by positively charged protein side chains, 4 of the 14 are contacted only by uncharged groups (e.g., a peptide NH), and 4 of the 14 are contacted by both charged and uncharged groups. In the case of the phosphate at pGAATTC (Figure 2b), one phosphoryl oxygen is contacted by two charged side chains and the other by one charged and one uncharged side chain. Thus, the protein makes 24 direct interactions in all with DNA phosphoryl oxygens.

It is far from certain how to account for the total contribution of these protein–phosphate contacts to ΔG°_{bind}. Is it appropriate to count phosphates, to count phosphoryl oxygens, or to count protein functional groups that interact with phosphoryl oxygens? How does one account for the ∼ 14 water molecules that are tightly bound to the phosphates and contribute to the complementarity at the protein–DNA in-

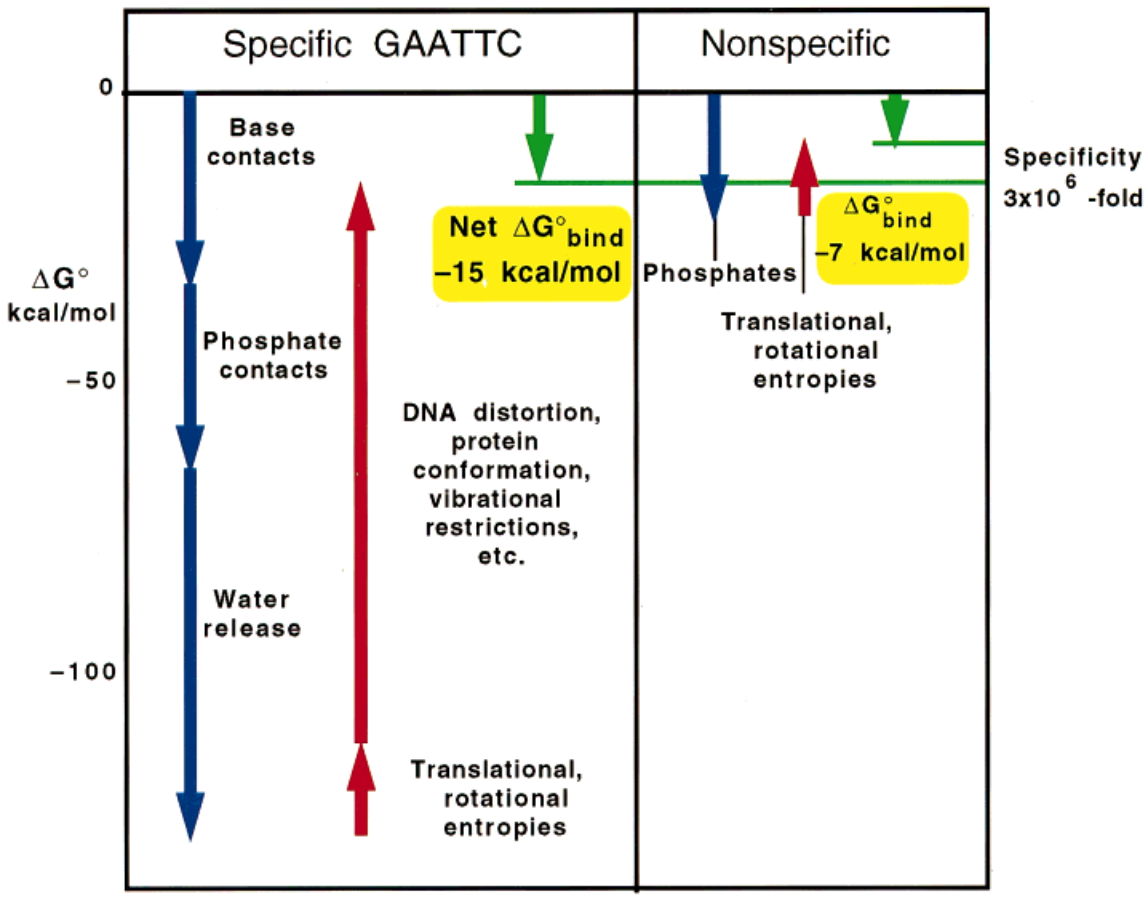

FIGURE 1 Energetic components of specific and nonspecific binding of *Eco*RI endonuclease. Favorable contributions are represented by blue arrows, unfavorable by red and observed net ΔG°_{bind} by green (values for sequence no. 1, Figure 9, at pH 7.3, 0.18*M* salt, 22°C). All components are estimated and may vary from true values by ±50% (see text). For specific binding, the large unfavorable contribution from ''DNA distortion, protein conformation and vibrational restrictions'' is estimated as the difference between the sum of other components and observed ΔG°_{bind}. The division into separate components should *not* be taken to mean that these contributions to complex formation are independent of each other; they are not.[6]

terface or for long range electrostatic effects (i.e., global ion redistribution upon binding)?[66]

Concepts based solely upon stoichiometric cation displacement from DNA phosphate by charged side chains[14] would appear to be inadequate to deal with the complexity of the interactions in the protein–DNA recognition interface. For example, the salt dependence of *Eco*RI binding to a cognate site (M. Kurpiewski, D. J. Chi, D. Cao, and L. Jen-Jacobson, unpublished results) indicates that a net of approximately 12 cations are displaced, but it is not obvious how this relates to the protein–phosphate interactions visualized in the crystal structure of the complex.[40–42] One approach is to calculate the contribution of the polyelectrolyte effect[15,67,68] from the salt-dependence data, and then add a contribution for the interactions of uncharged groups with phosphoryl oxygens (Figure 2a). An alternative is to calculate separately the contributions from interactions of charged[67–69] and uncharged groups with phosphoryl oxygens. It turns out that the free-energy contribution of a salt-bridge with a charged side chain is not very different from the contribution of a hydrogen bond from an uncharged group to phosphate as determined from studies with mutant proteins.[69] The two approaches yield similar total contributions of about −30 kcal/mol for the protein–phosphate interactions.

Of the 14 phosphates that interact with protein, six (three on each DNA strand at pNpGAApTTC)

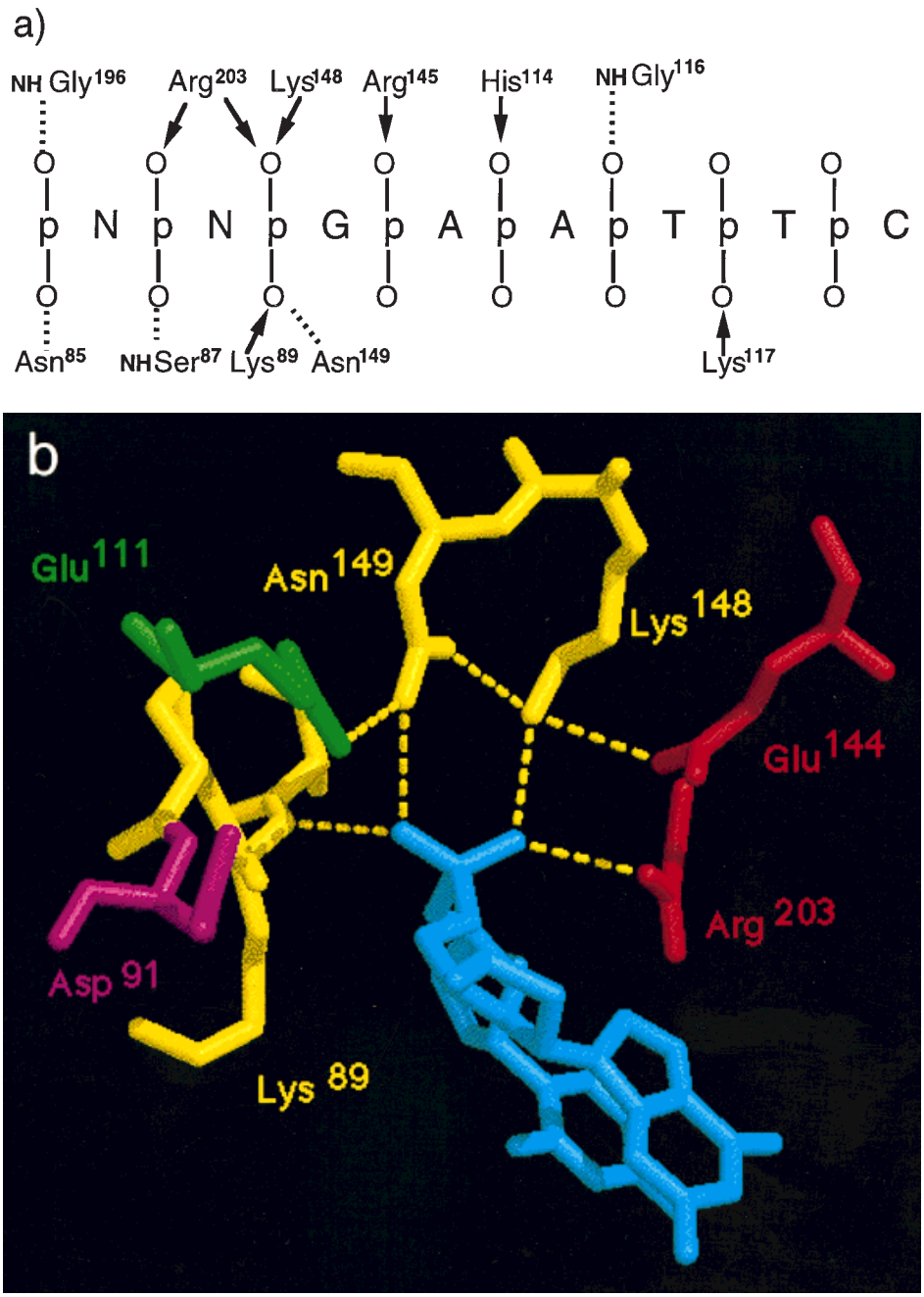

FIGURE 2 Protein–phosphate interactions in the *Eco*RI endonuclease–DNA complex. (a) Schematic of direct interactions with phosphoryl oxygens in two crystal structures.[40,117] Compare these crystallographically defined contacts with ethylation-interference patterns, which identify the phosphates important for binding (Figure 8a). This schematic omits several water-mediated interactions, since phosphates receiving *only* water-mediated contacts do not show interference and are thus not crucial to binding. The flanking bases (N) differ in the two co-crystals,[40,117] but are not directly contacted by protein. (b) Networked hydrogen bonding interactions with the ''clamp'' phosphate at pGAATTC. The DNA (CpG) is in blue. The *pro-Sp* phosphoryl oxygen points to right and *pro-Rp* oxygen to left. Active-site residues Glu111 (green) and Asp91 (magenta) are at left. Arg203 makes water-mediated hydrogen-bonds (not shown) to *G* for recognition. Glu144 is hydrogen-bonded to Arg145 (see Figures 4 and 5). Reprinted by permission from Kurpiewski et al.[49]

have been described[6] as crucial ''clamps'' that act to position the protein recognition elements that contact the DNA bases. These particular phosphates are recognized by the protein with extremely high geometric precision.[48,49] It has been noted that these ''clamp'' phosphates are positioned so that the protein can exert a torsional strain[70] to ''kink'' the DNA, which is in turn necessary for insertion of the base-recognition elements into the DNA major groove.[40] Thus, sequence specificity is determined primarily by the cooperative action of the protein–phosphate and protein–base contacts.[6]

The other major factor contributing favorably to ΔG°_{bind} is the release of bound water from nonpolar surfaces (hydrophobic effect), but there is substantial uncertainty about its magnitude. Formation of the specific *Eco*RI complex, like that of many other protein–DNA complexes,[12,19–26,71] is characterized by a large negative change in heat capacity ΔC°_p (Refs. 19 and 71, and J. T. Ames, L. E. Engler, and L. Jen-Jacobson, unpublished results). Record and co-workers[19] estimated a contribution of −120 kcal/mol from the hydrophobic effect based upon the assumption that the measured negative ΔC°_p derived entirely from the release of ordered water from nonpolar surfaces, and later[71] −135 kcal/mol, taking account of the positive contribution to ΔC°_p of water release from polar surfaces. The uncertainties in these determinations were $> \pm 50\%$. We have computed from the crystal structure[40–42] that approximately 2500 Å^2 of *nonpolar* surface area is buried upon specific complex formation. If we calculate the hydrophobic free energy contribution from burial of this area using the oft-cited average value of −25 cal/Å^2 (Ref. 72) this implies a contribution of −60 kcal/mol, which we tentatively adopt in Figure 1. An attempt to reconcile the magnitudes of the ''microscopic'' and ''macroscopic'' hydrophobic effects[73] has produced the suggestion that −47 cal/Å^2 is a more appropriate value; this would imply a hydrophobic contribution of about −120 kcal/mol.

Thus, the favorable free energy from direct protein–base and protein–phosphate contacts contributes only about half of the total favorable ΔG°_{bind}. The sum of the favorable factors (base and phosphate contacts and hydrophobic effect) is on the order of −120 kcal/mol, implying that there must be large unfavorable factors (approx. +105 kcal/mol) in opposition.

Complex formation restricts rotational and translational freedom of the protein and DNA; this makes an unfavorable entropic contribution to ΔG°_{bind} on the order of +15 kcal/mol at 22°C.[71,74] The remaining balance of +90 kcal/mol derives from a variety of conformational and configurational factors, whose identities we can surmise but whose quantitative contributions to ΔG°_{bind} we cannot deconvolute at present. These include induced structure in the protein (e.g., residues 121–138 are disordered in the *Eco*RI apoenzyme but ordered in the complex[75]), restrictions on the internal motions of protein and DNA,[76–78] restrictions on the motion of tightly bound water molecules trapped at the protein–DNA interface,[22,79] and the energy required to distort or ''strain''[70] the DNA into the unusual ''kinked'' conformation in the complex.

It is also important to bear in mind that whereas type II restriction endonucleases bind as preformed dimers to a palindromic site comprised of two equivalent half-sites, some DNA-binding proteins assemble cooperatively into higher order oligomers only upon binding. For example, strong cooperativity between bacteriophage P22 Arc repressor dimers as they bind to operator half-sites powerfully enhances the overall binding constant and is necessary for repression in vivo.[62,63] The λ cI and *E. coli* methionine (MetJ) repressors bind cooperatively not to operator half-sites, but to adjacent nonidentical operators.[23,80] For bZIP proteins interacting with cyclic AMP-responsive element target sites, both dimerization and DNA binding are promoted by the trans-activator protein Tax, which interacts directly with the basic segments of the bZIP elements.[81] In such cases of coupled binding and oligomerization, the protein–protein interaction energy is considered as a distinct contribution to the overall binding energy of the specific complex.

The situation for nonspecific binding is simpler (Figure 1). It is useful to think of the nonspecific complex as a loose association in which the protein makes favorable Coulombic[67,68] and/or hydrogen-bond interactions with DNA phosphates, but where no intimate protein–DNA interface is formed. We presume that few or no protein–base contacts are made and no DNA distortion occurs, generalizing from observations in the only cases in which crystal structures of both specific and nonspecific complexes are presently available: *Eco*RV endonuclease[82] and glucocorticoid receptor DNA-binding domain.[83]

For many proteins, the total number of protein–phosphate interactions in the nonspecific complex appears (from the salt-dependence of binding) to be larger than in the specific complex.[4,7,15,84] Binding of *Eco*RI and *Bam*HI endonucleases to nonspecific DNA is accompanied by a near-zero change in heat capacity ΔC°_p (Ref. 65 and J. T. Ames et al., unpublished results), like that of other site-selective pro-

teins[12,20,22] and like binding of the nonspecific single-strand binding protein (SSB) to oligodeoxypyrimidines.[85] This contrasts with the strongly negative ΔC_p° of about −1.5 kcal/mol for formation of the specific *Eco*RI (Ref. 19 and J. T. Ames et al., unpublished results) and *Bam*HI complexes,[65] implying that nonspecific binding involves no significant favorable contribution from the hydrophobic effect, and no significant unfavorable contributions from restricted vibrations of protein, DNA, or trapped water. Nonspecific binding for *Bam*HI and *Eco*RI is enthalpy driven at all temperatures, like the binding of the nonspecific SSB protein[85] and nonspecific binding of lac repressor,[12] whereas it is reportedly entropy driven for λ Cro protein.[20] The entropy change in nonspecific *Eco*RI and *Bam*HI binding is near zero or slightly unfavorable, implying that the positive ΔS° from the polyelectrolyte effect at best approximately balances the unfavorable loss of translational and rotational entropies. We conclude that formation of the nonspecific complex is driven by enthalpically favorable protein–phosphate interactions.

Taken together, these observations further reinforce a model of the nonspecific complex as a loose association with ordered (but not tightly bound) water still present on nonpolar surfaces. This view is important to understanding the role of the nonspecific complex in site location.

ROLE OF NONSPECIFIC BINDING IN SITE LOCATION

It has long been appreciated that a protein cannot efficiently locate a rare recognition site on DNA by free three-dimensional diffusion in solution.[86–88] Many, perhaps all, site-specific DNA binding proteins therefore make use of linear diffusion or ''sliding''[86–94] along a DNA molecule, which results in association with the target site at rates faster than the diffusion-controlled ''limit.''[86,87,91,93] Although it was at first proposed[95] that sliding might not be relevant to site location in vivo because of inhibition by physiological salt or Mg^{2+} concentrations and/or the binding of other proteins to the DNA, it has now been shown convincingly that there is a strong correlation between the ability of mutant proteins to undergo linear diffusion in vitro and their biological function in vivo.[96,97]

Even at a recognition site contained in a small oligonucleotide, it appears highly likely that initial binding does not lead immediately to the specific complex. Evidence for an intermediate that precedes specific binding comes from studies of the dependence of the measured association rate constant k_a on salt concentration. If the apparent k_a is a true elementary second-order rate constant for a single-step association, it should have only a weak dependence on salt concentration, derived from ''screening'' of DNA phosphates by associated cations.[15,91] On the other hand, if the formation of the specific complex (E′S′) proceeds through an intermediate ES, that is,

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} E'S' \qquad (1)$$

then the measurement of the *initial* rate of association (such that the process characterized by k_{-2} is negligible) yields an apparent second-order association constant k_a^{app} given by (if $k_2 \ll k_{-1}$)

$$k_a^{app} = \frac{k_1 k_2}{k_{-1}} = K_1 k_2 \qquad (2)$$

In such circumstances, k_a^{app} appears to be strongly dependent on salt concentration (decreases with increasing [salt]) because the rate constant k_{-1} for the dissociation step ES → E + S is strongly salt dependent.[15,91] An observed strong salt dependence of k_a^{app}($d \log k_a^{app}/d \log[\text{salt}] \gg 1$) indicates the presence of intermediates in the specific binding of *E. coli* RNA polymerase[15,98] and *Eco*RI endonuclease (Figure 3). For *Eco*RI, the final specific complex E′S′ is that visualized in the crystal structure[40–42] and the first intermediate or initial collision complex is the same ''nonspecific complex'' we described above. Other proteins (e.g., RNA polymerase[15,98]) may have a series of intermediate forms bound to DNA.

Note that the terms used to describe the nonspecific complex (''not localized,'' ''loose association'') are completely consistent with the idea that this is also the same complex involved in ''sliding'' along the DNA. To say that the intrinsic binding constants for the various base sequences encountered are approximately equal is equivalent to saying that the DNA presents an isopotential surface for sliding. Only the recognition site (or very closely related sites) represents a potential ''well'' at which stronger binding occurs. It follows that any site-selective protein that is capable of linear diffusion or sliding along the DNA will attain its specific protein–DNA complex by passing through one or more intermediate complexes.

Consider a protein sliding along a DNA mole-

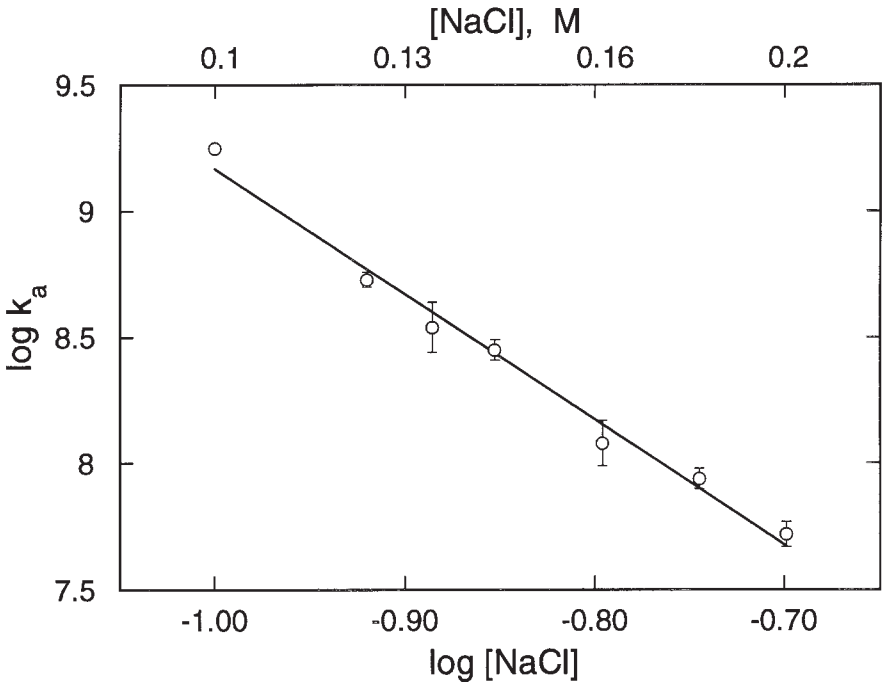

FIGURE 3 Evidence for an intermediate in the association of *Eco*RI endonuclease with TCGCGAATTCGCG. The logarithm of the apparent second-order association rate constant (k_a) is plotted against log[NaCl]. The measured slope is -5.0, whereas a slope of ≈ -1.5 is expected if there is no intermediate in binding. (With no intermediate, the expected slope[15,91] of a plot of $d \log k_a/d$ log[salt] is $-0.12z$, compared with $-0.88z$ for the dependence $d \log K_A/d$ log[salt], where z is the number of cations stoichiometrically released upon binding). Association was measured by the membrane-filter method[145] at pH 7.4, 22°C. Initial linear rates of association were determined at 5–7 s intervals for a total of about 60 s, and values of k_a calculated according to Eq. (7).[145] Values shown are means $\pm$ *SD* of at least 4 determinations.

cule, rapidly (and randomly) scanning to and fro until it has a chance encounter with its recognition site. As it binds preferentially to its recognition site, it will ''pause'' upon reaching that site, because the higher affinity for the specific site is produced largely by a slower rate of dissociation from that site. It has been observed experimentally[99] that *Eco*RI endonuclease makes detectable pauses even at sites that differ from the recognition site by one incorrect base-pair (*Eco*RI* sites), despite the fact that the equilibrium affinity for such sites is barely distinguishable from that for nonspecific DNA.[6] This is consistent with the observation[6,13] that the enzyme makes a localized ethylation-interference footprint at *Eco*RI* sites, but none on nonspecific DNA.

Although we lack complete information for many systems, it seems certain that conversion of the nonspecific to the specific complex will require not only the formation of protein–base contacts and the release of bound water from nonpolar surfaces, but also conformational changes in the protein and/or the DNA to form the intimate recognition interface. Such changes are implied by the way in which the protein ''enfolds'' the DNA in many complexes[40,100–102] by the frequent distortions in bound DNA[39,40,57,82,102] by comparison of nonspecific to specific complexes[7,12,13,15,82–84] and in some cases by solution studies such as fluorescence stopped-flow[103,104] or quench-flow DNA-footprinting kinetics.[26]

We presently know relatively little about what ''triggers'' this recognition process. It is puzzling because the speed of sliding is apparently incompatible with a mechanism by which the protein continually reaches into one or both DNA grooves to ''sample'' the functional groups on the bases. Current models propose that the sliding protein detects sequence-dependent differences in local DNA structure (e.g., twist and roll)[105]—differences in equilibrium conformation and/or the range of anisotropic fluctuations.[106,107] The important conclusion of this

section is that site-specific or site-selective DNA binding proteins also bind to nonspecific DNA not by some unhappy accident of their structure, but because it is to their advantage to do so. It is a devil's bargain, however, because to function at its recognition site, the protein must find a way to overcome the competitive inhibition by nonspecific DNA.

ASSEMBLY OF THE RECOGNITION INTERFACE AND THE CATALYTIC SITE

One of the keys to an enzyme achieving extremely high DNA sequence specificity is the physical coupling of recognition to catalysis, such that efficient catalysis occurs only when the protein–DNA interface is correctly assembled at the correct recognition site. Among the enzymes with the highest site specificity are the type II restriction endonucleases, which are unique in that a protein homodimer recognizes the two identical half-sites of a palindromic sequence and catalyzes two reactions (i.e., cleavage in both DNA strands) using two symmetrical ''active sites'' via two apparently identical transition states.[6] It is usually the case that both reactions occur in a single binding event.[89] Strikingly, each recognition interface and each active site is formed from elements of both subunits, such that the rate of catalysis in both half-sites is responsive to correct recognition in both half-sites, even when (for endonucleases that make staggered cuts) the reactive phosphodiester bonds may be separated by 17–19 Å. This forms a striking contrast with dimeric transcription factors, in which each monomer is usually (for an exception, see NF-κB[100,101]) responsible for recognizing a distinct half-site. We illustrate with the case of *Eco*RI endonuclease, although the same general principles apply for the *Bam*HI[45] and *Eco*RV[82] endonucleases, which have little primary sequence homology to *Eco*RI, but whose catalytic sites are strikingly similar.[108,109]

The *Eco*RI recognition interface and catalytic site consist of an extremely intricate network that connects, principally via hydrogen bonds, the recognized functional groups on the DNA bases, the protein side chains that recognize those groups, the phosphate where strand scission occurs, the protein side chains directly involved in catalyzing hydrolysis of the phosphodiester bond, and the crucial ''clamp'' phosphates. This network is so constructed that the precise location of recognition elements is ''coupled'' to the precise location of catalytic elements, inextricably coupling catalysis to correct sequence recognition.

Figures 4 and 5 show the relevant portions of the catalytic site in the crystal structure of the specific E′S′ complex in the absence of Mg^{2+}.[40,42] The side chains of Asp91 and Glu111 lie in position to chelate the Mg^{2+} ion required for catalysis. The Mg^{2+} also coordinates to one phosphoryl oxygen of the scissile phosphate, polarizing it for nucleophilic attack and probably also stabilizing the pentacovalent transition state.[42] The catalytic role of Glu111 is supported by genetic evidence[110] and there are acidic sidechains precisely analogous to Glu111 and Asp91 in the active sites of *Eco*RV and *Bam*HI endonucleases.[108,109] The position of Asp91 is maintained by the interaction of residues 86–89 with the upstream phosphates (see Figure 2a). Glu111 lies in a β element, anchored (Figure 4) by the interaction of His114 and Gly116-NH with phosphates and a nonpolar recognition interaction between the side chain of Gln115 and a thymine methyl group. The position of Gly116-NH relative to the ''clamp'' phosphate at GAApTTC is especially critical: this phosphate is in an unusual conformation because of the central ''kink'' that exists only in the complexed DNA, and this interaction is exquisitely sensitive to geometric perturbation.[48] Note also that the catalytic residue Glu111 is ''wired'' (via hydrogen bonds with Asn149) to the ''clamp'' phosphate at pGAATTC (Figure 2b), which also helps anchor Asp91.

The scissile phosphate is further polarized by coordination to Lys113 and to the guanidino group of Arg145, which also donates a hydrogen bond to the imidazole-N7 of the inner adenine (Figure 4). Arg145 is the only protein side chain that is directly involved in both base recognition and catalysis, providing communication between the processes. It is especially noteworthy that the Arg145 side chain shown in Figures 4 and 5 does not derive from the same subunit that provides the other elements for recognition in this half-site. So, for example (Figure 5a), the N6-amino group of the inner adenine and both N7 and N6 of the outer adenine are recognized via hydrogen bonds with the side chain of Asn141, coming from a different subunit than Arg145. This interdigitation of the subunits provides communication between the half-sites.

There is considerable doubt about the identity of the base that abstracts a proton from water in the hydrolytic mechanisms of *Eco*RI and other restriction endonucleases with strongly analogous active sites. Among the proposed candidates in *Eco*RI are a second water molecule,[42] or the neighboring phosphate at GApATTC[111]; however, a phosphate with $pK_a < 2$ is almost certainly too weak a base and there is no structural indication that the pK_a would

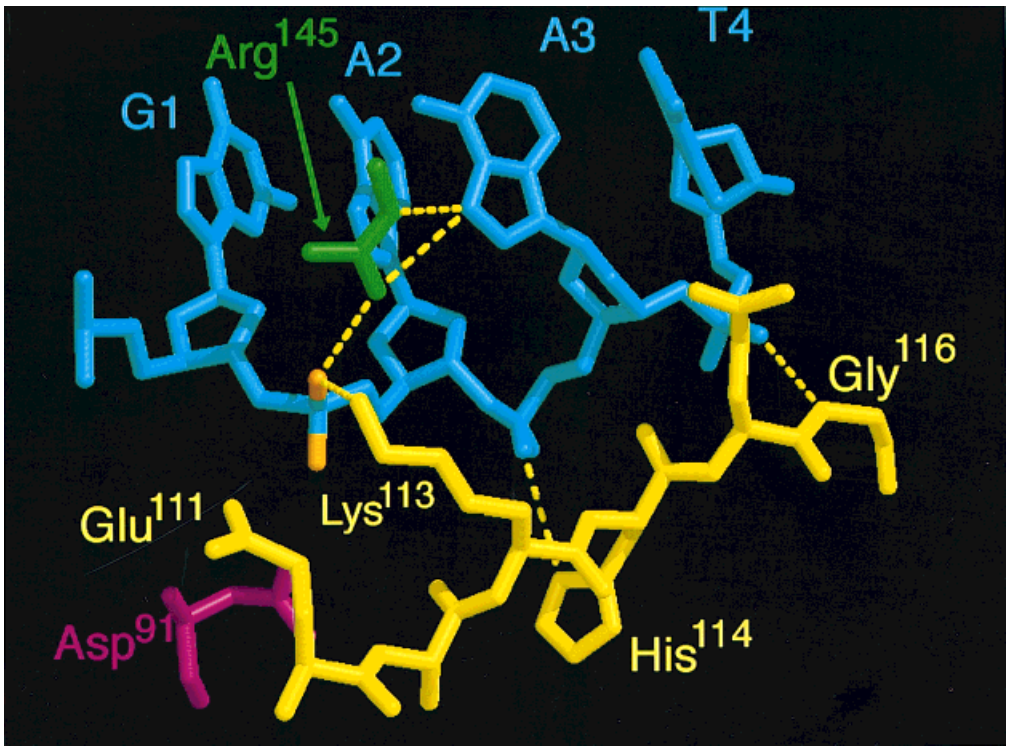

FIGURE 4 Region of the catalytic site in the *Eco*RI endonuclease–DNA complex. Only one DNA strand is shown in blue. The phosphoryl oxygens at the scissile phosphate are in orange. Note that Arg145 (green) participates in recognition by hydrogen bonding to A3-N7, and also polarizes the scissile phosphate by hydrogen bonding. The sidechains of Glu111 and Asp91 (magenta) chelate a Mg^{2+} ion in the transition state.[42] All protein residues shown are from the same subunit. Models were drawn with MIDAS Plus software using the atomic coordinates of the *Eco*RI endonuclease-TCGCGAATTCGCG complex[40–42]; Brookhaven Protein Data Bank accession number 1rie.

be anomalously high.[112] A two-metal mechanism has been proposed for *Eco*RV[112,113] and may be possible[109] for *Bam*HI, which has Glu in the position analogous to Lys113 of *Eco*RI.

The α-helix that provides the recognition residues Asn141 at its N-terminus and Arg145 at its midpoint is anchored at its C-terminus by the interaction of Lys148 and Asn149 with the ''clamp'' phosphate at pGAATTC (Figure 2b). This phosphate serves as an organizing center for both recognition and catalytic elements through networked interactions of base recognition residues (Arg203 and Arg145 via hydrogen bonding to Glu144) and catalytic residues Glu111 and Asp91 (Figure 2). The relative positions of these symmetrical phosphates in the two half-sites are altered by rotation about the helix axis as a result of DNA unwinding caused by the central ''kink.''[49] This provides additional coupling between the recognition and catalytic processes and the DNA distortion in the complex.

This structure shows that the ''catalytic site'' is essentially completely assembled when binding occurs in the absence of Mg^{2+}. When Mn^{2+} (or Mg^{2+}) is diffused into the protein–DNA cocrystals, DNA cleavage occurs, but the crystals do not shatter and the resulting enzyme-product–Mn^{2+} complex is nearly isomorphous (protein and DNA backbones superimposable with rms error of 0.5 Å) with the starting protein–DNA complex.[42] This strongly implies that there are no *major* rearrangements upon addition of Mg^{2+} or on the way to the transition state, but there are certainly *minor* rearrangements required in the immediate vicinity of the scissile phosphodiester bond.

THE SPECIFIC E′S′ COMPLEX CLOSELY RESEMBLES THE TRANSITION STATE

The transition state cannot be ''seen'' in a crystal structure, of course, but the methods of solution

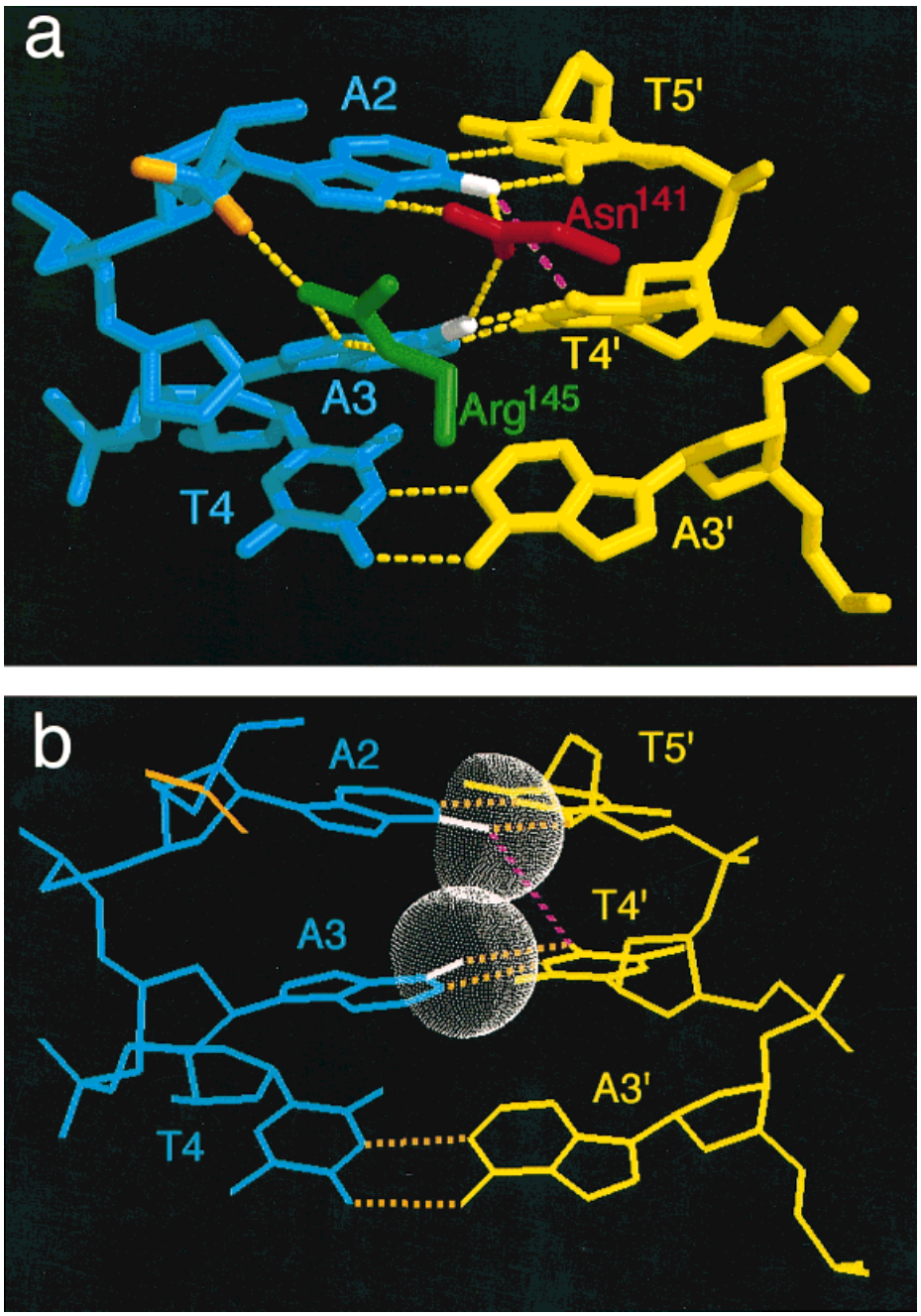

FIGURE 5 Networked interactions at the AAT base pairs of the *Eco*RI endonuclease–DNA complex. The two DNA strands are shown in blue and yellow. The phosphoryl oxygens at the scissile phosphate are in orange. Note the three-center hydrogen bond (magenta) from A2-N6 to T4′-O4. (a) Recognition of the adenine bases by Asn141 and Arg145 from opposite subunits. Note again the hydrogen bond from Arg145 to the scissile phosphate (cf. Figure 4) and recognition of both adenines in one half-site by Asn141. (b) Wireframe view of the same region, showing the van der Waals spheres of A2-N6 and A3-N6 (white) in steric clash. The protein sidechains and their hydrogen bonds have been removed for clarity.

biochemistry allow us to analyze whether specific factors that affect the energetics of the E′S′ complexes for *Eco*RI and *Bam*HI endonucleases also affect the energetics of the transition states. To do this, we have used the structural perturbation method,[9] in which the DNA substrates are modified

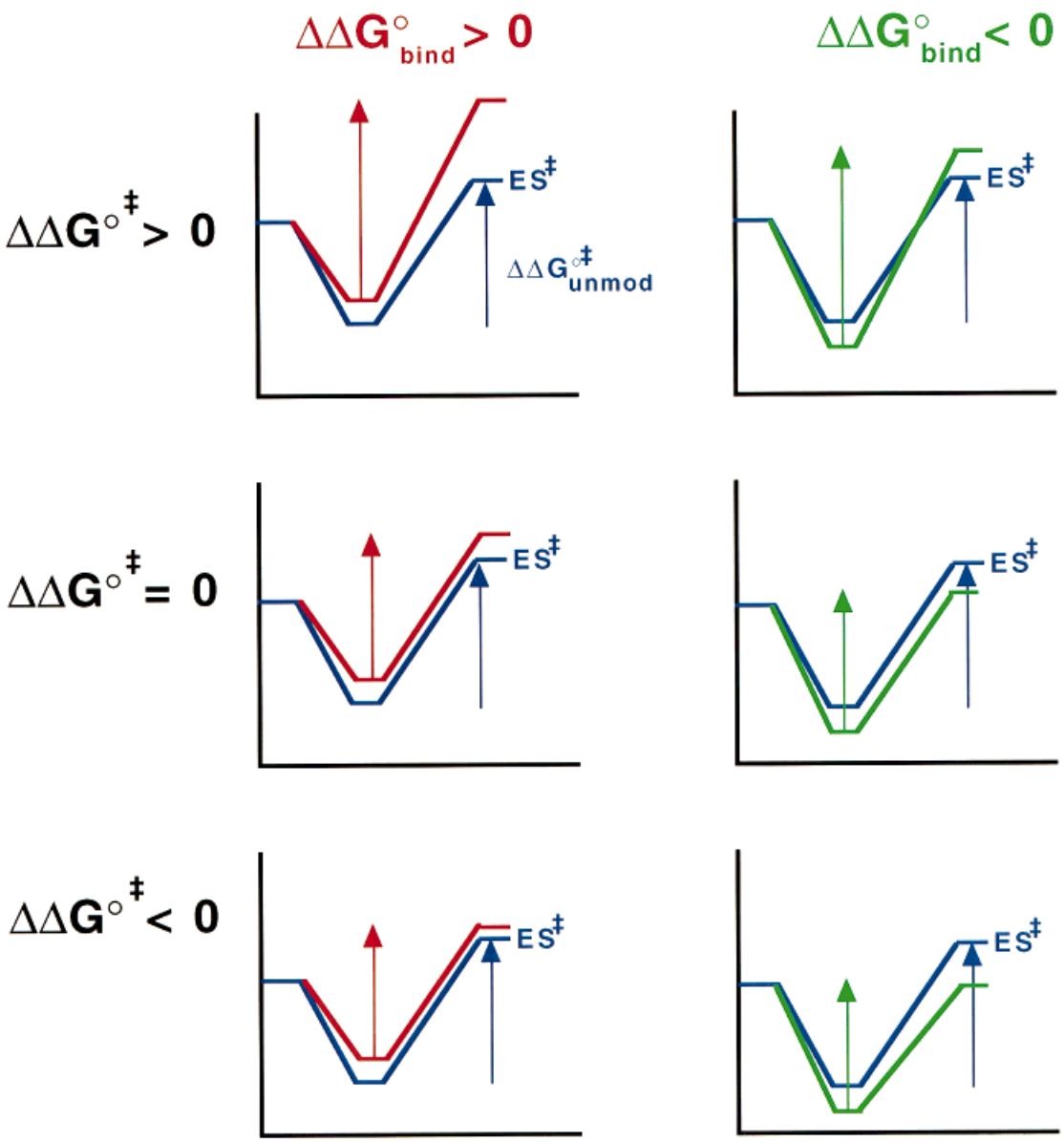

FIGURE 6 Reaction coordinate diagrams showing modifications that affect ΔG°_{bind} and/or $\Delta G^{\circ\ddagger}$. Each progress curve shows only one enzyme–substrate complex (e.g., E′S′), at its free-energy minimum; any binding intermediates (e.g., ES preceding E′S′) are omitted for clarity. In each panel, the reaction profile of the unmodified substrate is in blue and standard free energy of activation $\Delta G^{\circ\ddagger}$ for the unmodified substrate is indicated by the blue arrow. Reaction profiles are in red for modifications that inhibit binding ($\Delta\Delta G^{\circ}_{bind} > 0$) and in green for modifications that improve binding ($\Delta\Delta G^{\circ}_{bind} < 0$). Activation free energies for modified substrates are indicated by corresponding red and green arrows.

with base analogues or phosphate analogues so as to alter the binding free energy ΔG°_{bind} by an amount $\Delta\Delta G^{\circ}_{bind}$. Depending on the modified group, the binding free energy may be less or more favorable than for the unmodified substrate (Figure 6). In each case, the free energy of activation $\Delta G^{\circ\ddagger}$ may increase, remain the same, or decrease. We have in fact observed numerous occurrences of all six cases in studies of *Eco*RI and *Bam*HI endonucleases.

These changes tell us about the relative effects of the perturbation on the E′S′ and transition-state ($ES^{\ddagger}$) complexes. For example, if we perturb a par-

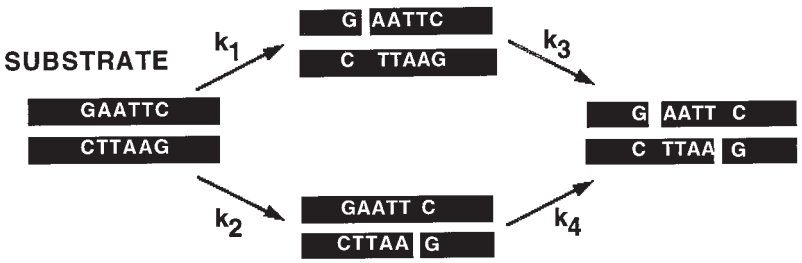

FIGURE 7 Parallel-sequential pathway for double-strand cleavage of an *Eco*RI site GAATTC. Rate constants k_1, k_2 and k_3, k_4 can be distinguished experimentally using eccentric site location in oligonucleotides.[6] When DNA is modified in one half-site, k_1 and k_4 always represent cleavages in the unmodified half-site, whereas k_2 and k_3 are for the modified half-site.

ticular interaction in E′S′ and find that $\Delta\Delta G^{\circ\ddagger} \neq 0$, the perturbation has differential effects on the free-energy levels of E′S′ and ES‡. This is often difficult to interpret in a structural sense. Some of these cases with biological importance are considered in later sections.

On the other hand, if $\Delta\Delta G^{\circ\ddagger} = 0$, this indicates that the interaction makes the same free-energy contribution to E′S′ and the transition state. If we can interpret $\Delta\Delta G^{\circ}_{\text{bind}}$ based upon a crystal structure of the E′S′ complex, we can infer that the same interpretation applies to the ES‡ complex.

An unusual feature of type II restriction endonucleases is that both equilibrium and kinetic aspects of binding to DNA can be studied in the absence of catalysis simply by omitting the required Mg^{2+} cofactor.[6,16,103] It is therefore relatively simple to partition the reaction into ''binding'' [see Eq. (1) above] and ''catalytic'' (see Figure 7) steps, which is generally much more difficult for other enzymes. It is important to note, however, that this definition of the catalytic steps includes a term for the binding of Mg^{2+} to the enzyme–DNA complex, so that the apparent rate constant $k^{\text{app}}_{\text{cleave}}$ is given by

$$k^{\text{app}}_{\text{cleave}} = k_{\max} \frac{[Mg^{2+}]}{K_D^{Mg} + [Mg^{2+}]} \qquad (3)$$

where K_D^{Mg} is the apparent equilibrium dissociation constant for Mg^{2+} from the protein–DNA complex. The experimental value of K_D^{Mg} is about 3 ± 1 m*M*, regardless of whether Mg^{2+} is added to the enzyme–DNA complex or enzyme is added to the DNA–Mg^{2+} complex.[114] This value does not change significantly for substrates that are poorly bound and cleaved.[114]

As noted above, restriction endonucleases catalyze two distinct cleavage reactions in the two DNA strands (Figure 7). In the unmodified palindromic substrate these cleavages occur at precisely equal rates, but when a modification is made in one half-site the rates may (and often do) become unequal. We adopt the convention (Figure 7) in which for an asymmetrically modified DNA site, k_1 represents the cleavage rate constant in the unmodified half-site and k_2 the cleavage rate constant in the modified half-site. Because of the interdigitation of subunits and the communication between the half-sites, the rate constant k_1 reports on global rather than strictly local features of the complex,[6] whereas k_2 is more subject to strictly local perturbation by modifications near the reaction center. We therefore concentrate on k_1 in considering general features of the transition state.

Table II shows a variety of modified DNA substrates for *Eco*RI endonuclease where there is an effect on $\Delta G^{\circ}_{\text{bind}}$, but for which there is little or no change in k_1, and therefore $\Delta\Delta G^{\circ\ddagger} \approx 0$. In many of these cases, there is also no change in k_2. These cases include substitutions that disrupt one or more protein–base contacts, perturb one or more protein–phosphate contacts, and/or affect the energy required to distort the DNA. In all of these cases, unchanged ethylation-interference footprints (compare Figure 8a with 8b) and unaltered salt dependence of binding indicate that there are no ''adaptive'' changes in the modified complexes (see below).

The simplest instances are those in which a single protein–DNA interaction is deleted by removing a functional group on a base (deleting adenine-N7 by substituting 7-deaza-adenine [^{7C}A] or thymine-CH_3 by substituting deoxyuridine). In these situations $\Delta\Delta G^{\circ}_{\text{bind}} > 0$ and $\Delta\Delta G^{\circ\ddagger} \approx 0$ (Table II). From these we can infer equal free-energy contributions in E′S′ and ES‡ from the hydrogen-bond contacts Asn141 → A2-N7 and Arg145 → A3-N7 (Figure 5a) and, similarly, from the nonpolar contacts among

Table II Modifications with Equal Effects on E′S′ and ES‡

Site[a]	Contact Disrupted or Affected Property	$\Delta\Delta G^{\circ}_{bind}$ [b] (kcal/mol)	k_1 (s^{-1})[c] Unmodified Half-Site	$\Delta\Delta G_1^{\circ\ddagger}$ [d] (kcal/mol)	$\Delta\Delta G_I^{\circ\ddagger}$ [e] (kcal/mol)	k_2 (s^{-1}) Modified Half-Site	$\Delta\Delta G_2^{\circ\ddagger}$ [d] (kcal/mol)
1 2 3 4 5 6 G A A T T C C T T A A G 6′5′4′3′2′1′		0	0.8 ± 0.1	0	0	0.8 ± 0.1	0
	Direct protein–base contacts only						
G[7c]**A**ATTC C TTAAG	H-bond Asn141	+1.2 ± 0.2	0.8 ± 0.1	+0.1 ± 0.1	+1.3 ± 0.2	0.7 ± 0.1	+0.1 ± 0.1
GA[7c]**A**TTC CT TAAG	H-bond Arg145	+1.4 ± 0.2	0.7 ± 0.1	+0.1 ± 0.2	+1.5 ± 0.3	0.1 ± 0.1	+1.4 ± 0.1
GAA**U**TC CTTAAG	Nonpolar Ala142/Gln115	+1.2 ± 0.2	0.9 ± 0.1	−0.1 ± 0.2	+1.1 ± 0.3	0.7 ± 0.1	+0.1 ± 0.1
	Direct contacts and conformational or other factors						
GA**P**TTC CTTAAG	H-bond Asn141 Clash A2-A3 Watson–Crick H-bond	−1.0 ± 0.1	1.1 ± 0.2	−0.2 ± 0.2	−1.2 ± 0.2	0.7 ± 0.1	+0.1 ± 0.1
GA**P**TTC CTT**P**AG	As above, both half-sites	−1.7 ± 0.1	1.0 ± 0.2	−0.1 ± 0.2	−1.8 ± 0.2	0.9 ± 0.1	−0.1 ± 0.1
G**P**ATTC CTTAAG	H-bond Asn141 Clash A2-A3 Watson–Crick H-bond H-bond A2 → T4′	+1.3 ± 0.2	0.9 ± 0.1	−0.1 ± 0.1	+1.2 ± 0.2	0.3 ± 0.1	+0.6 ± 0.1
G**P**ATTC CTTA**P**G	As above, both half-sites	+2.3 ± 0.1	0.7 ± 0.1	+0.1 ± 0.2	+2.4 ± 0.2	0.9 ± 0.2	−0.1 ± 0.2
G**PP**TTC CTTAAG	Two H bonds Asn141 Clash A2-A3 2 Watson–Crick H bonds H bond A2 → T4′	+0.7 ± 0.1	0.7 ± 0.1	+0.1 ± 0.1	+0.8 ± 0.1	0.2 ± 0.1	+0.8 ± 0.1

GA**P**TTC CT**U**AAG	H bond Asn141 Clash A2-A3 Watson–Crick H bond Nonpolar Ala142/Gln115	−0.2 ± 0.1	1.1 ± 0.2	−0.2 ± 0.2	−0.4 ± 0.2	0.5 ± 0.1	+0.2 ± 0.1
G**P**ATTC C**U**TAAG	H bond Asn141 Clash A2-A3 Watson–Crick H bond H bond A2 → T4′ Nonpolar Gly140	+3.0 ± 0.2	0.6 ± 0.2	+0.2 ± 0.2	+3.2 ± 0.3	0.5 ± 0.1	+0.3 ± 0.1
G**PP**TTC CT**U**AAG	See GPPTTC above, nonpolar Ala42/Gln115	+1.8 ± 0.2	0.7 ± 0.1	+0.1 ± 0.2	+2.0 ± 0.2	0.3 ± 0.1	+0.6 ± 0.3
GA **P**TTC CT^{4H}**T**AAG	H bond Asn141, H bond Ala142 carbonyl Clash A2-A3 2 Watson–Crick H bonds H bond A2 → T4′	+2.1 ± 0.2	0.7 ± 0.1	+0.1 ± 0.2	+2.2 ± 0.2	±0.1 ± 0.1	+1.2 ± 0.1
			Protein–phosphate contacts				
Rp-**p**$_s$GAATTC CTTAAG	P-S$^-$ toward Lys148 and Arg203	+0.7 ± 0.1	0.8 ± 0.1	0 ± 0.1	+0.7 ± 0.2	0.6 ± 0.1	+0.2 ± 0.2
Rp-**p**$_s$GAATTC CTTAAG**p**$_s$-Rp	As above, both half-sites	+1.4 ± 0.1	1.0 ± 0.1	−0.1 ± 0.1	+1.3 ± 0.2	0.8 ± 0.1	0 ± 0.1

[a] Each site was embedded in the 17-nt oligomer TCCGGCAGAATTCTGCG; sites with single substitutions were tested in both orientations. Note that this construct differs from that used for Figure 1, such that the oligonucleotide with the unmodified $\overline{\text{GAATTC}}$ site shows $\Delta\Delta G^\circ_{bind} = +2.2$ kcal/mol relative to the reference sequence of Figure 1 (see Sequence 1, Figure 9). P = purine; ^{7c}A = 7-deazaadenine; ^{4H}T = 5-methyl-2-pyrimidinone.

[b] Observed difference in standard binding free energy between the unmodified and modified site: $\Delta\Delta G^\circ_{bind} = -RT \ln(K_A^{mod}/K_A^{unmod})$. Equilibrium binding measurements were at pH 7.3, 22°C, 150 m*M* NaCl. Reference K_A for unmodified site was $3.3 \times 10^9 M^{-1}$; means ± SD of ≥ 3 determinations for each site are shown.

[c] Means ± SD of ≥ 3 determinations; first-order strand scission constants were obtained under single-turnover conditions using 12 m*M* Mg^{2+} to initiate the reaction. For symmetrical sites, k_1 represents cleavage in the bottom strand.

[d] Observed difference in standard free energy of activation $\Delta G^{\circ\ddagger}$, relative to the unmodified site = $\Delta\Delta G^{\circ\ddagger} = -RT \ln(k_{cleave}^{mod}/k_{cleave}^{unmod})$.

[e] Differences in transition state interaction free energy (unmodified half-site); calculated as $\Delta\Delta G_1^{\circ\ddagger} = -RT \ln[(k_1 \times K_A)_{mod}]/[k_1 \times K_A)_{unmod}]$.

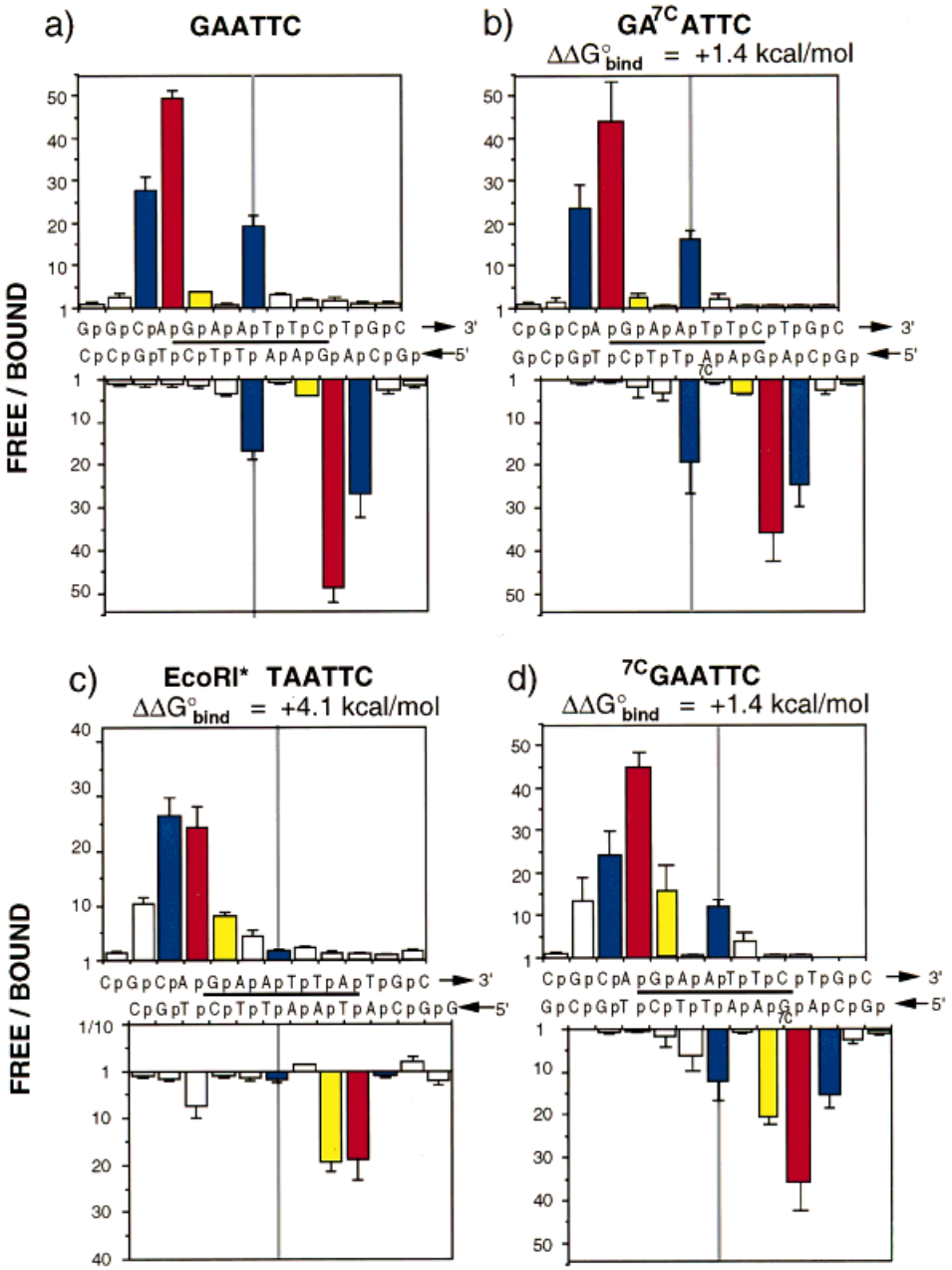

FIGURE 8 Ethylation-interference footprints made by *Eco*RI endonuclease on 17-nt substrates[6] containing canonical GAATTC or variant sites. Each bar shows the degree to which an ethyl group at each position interferes with binding. The interference axis (ordinate) is positive in both directions. A value > 1 means that ethylation inhibits binding; a value < 1 means that ethylation stimulates binding. The ''primary clamp'' phosphate is denoted by red bars, the ''supplementary clamp'' phosphates by blue bars, and the scissile phosphate by yellow bars. The dashed vertical line shows the center of symmetry in the canonical complex (panel a). Panel b shows *no* adaptation. Panel c shows *profound* adaptive changes involving loss of ''supplementary clamp'' contacts in both half-sites. Panel d shows *mild* adaptation, primarily expressed as increased interference at the scissile bonds GpAATTC.

T4-CH_3, Ala142, and Gln115. Thus, these contacts that the crystal structure shows in E′S′ must also exist in the transition-state complex $ES^{‡}$.

Substitutions of purine (P) for adenine delete contacts Asn141 → A-N6 (see Figure 5a) but also have effects on the energy required to distort the DNA.[54] Substitution of purine for A3 in either half-site (or both) improves binding free energy ($\Delta\Delta G^{\circ}_{bind} < 0$) because the unfavorable effect of deleting the hydrogen bond Asn141 → A3-N6 is outweighed by relief of the steric clash between the adjacent adenines (Figure 5b) and removal of an interstrand Watson–Crick hydrogen bond (Figures 5a,b) that constrains this base pair from adopting its unusual tip angle in the complex. If purine is substituted for A2 (in one or both half-sites), we observe $\Delta\Delta G^{\circ}_{bind} > 0$ as the net of removal of the contact Asn141 → A2-N6 (unfavorable), relief of the steric clash (favorable), and removal of a three-center hydrogen bond between A2-N6 and T4′-O4 on the opposite strand (Figure 5) that apparently stabilizes the distorted DNA conformation. These energetic allocations have been confirmed by analysis of complete thermodynamic pseudocycles[9,115] in which a series of base analogue sites are used to assort the energetic components in various combinations, yielding consistent values of $\Delta\Delta G^{\circ}_{bind}$. For example, the three-center hydrogen bond is also disrupted by substituting 5-methyl-2-pyrimidinone (^{4H}T) for T-4′ or the Watson–Crick hydrogen bond is disrupted by substituting ^{4H}T for T-5′.

For all these purine substitutions, we observe $\Delta\Delta G^{\circ‡} \approx 0$ for cleavage in one or both half-sites. This strongly implies not only that the hydrogen bonds from Asn^{141} to the N6-amino groups of both A2 and A3 are maintained in the transition state, but also that the various ''conformational'' factors have the same influence in the transition state that they do in the E′S′ complex. Note especially that the steric clash between adjacent adenines (Figure 5b) is produced by the unusual tip angle of base pair 3, a peculiar feature of the ''kinked'' DNA conformation, and therefore the overall DNA conformation in the transition state must closely resemble that in the E′S′ complex.

Various base–analogue substitutions for the central AATT bases can be combined in one substrate with progressively larger and approximately additive effects on $\Delta\Delta G^{\circ}_{bind}$.[9,13,54] This additivity indicates that structural adaptation has not occurred for these derivatives. In some cases of double and triple substitutions (Table II), we still observe $\Delta\Delta G^{\circ‡} \approx 0$, despite the fact that many different factors in complex stability are affected. For example, the P · ^{4H}T substitution in the third base pair disrupts two direct protein–base contacts, deletes two interstrand Watson–Crick hydrogen bonds and the three-center hydrogen bond A2 → T4′, and relieves the A2-A3 clash. The fact that $\Delta\Delta G^{\circ‡} \approx 0$ shows that *all* the factors perturbed in such multiple substitutions make equal contributions in E′S′ and $ES^{‡}$. By contrast, introduction of even one guanine analogue (7-deazaguanine [^{7C}G] or 2-aminopurine) in the first base position leads to *adaptive* changes (Figure 8d) and inhibition of cleavage ($\Delta\Delta G^{\circ‡} > 0$). Furthermore, the effects of multiple substitutions on binding and cleavage are much greater than expected from additivity.[9,13,54]

The inference that the overall DNA conformation of the E′S′ complex persists in $ES^{‡}$ is further supported by observations[49] on substitutions of chiral phosphorothioates (Ps) at the ''clamp'' phosphate pGAATTC, around which many recognition and catalytic residues are organized (see above). An Rp-Ps derivative points a P—S^{-} at the side chains of Lys148 and Arg203 (see Figure 2b) and P═O at Lys89 and Asn149. Because of the very tight packing of protein side chains around this phosphate even the slightly increased (by 0.6 Å) bond length of the P—S^{-} produces an inhibitory effect on binding (Table II), which is exaggerated when the substitution occurs in both strands, but there is no alteration in cleavage rate ($\Delta\Delta G^{\circ‡} = 0$), implying that the energetic effects in E′S′ and $ES^{‡}$ are equal. The same is true for the Rp-Ps substitution in both DNA strands (Table II). As noted previously, the precise relative positions of these phosphates in the two half-sites are determined by the DNA distortion seen in the crystal structure of the E′S′ complex, so it follows that the same DNA distortion must be maintained in $ES^{‡}$.

In most of the cases in Table II there is little or no change in the cleavage rate constant k_2 for the modified half-site. The most informative exception is the ^{7C}A substitution for A3, where k_2 decreases about 10-fold. From the crystal structure of the E′S′ complex, we infer that this substitution prevents the interaction of Arg145 with A3-N7 (Figures 4 and 5a). We believe that in the absence of this interaction, the flexible Arg145 side chain is no longer held in position to interact with (and polarize) the scissile phosphate GpAATTC, so the rate of hydrolysis is strongly decreased. This is an excellent example of how a DNA modification may affect a strictly local feature of the transition state, reported by k_2, without affecting the more global features of the transition state as reported by k_1. That a struc-

tural explanation for this transition state effect can be deduced from the E′S′ structure supports the hypothesis that the interactions of Arg145 are maintained from E′S′ to ES‡.

There are also several examples for the complex of *Bam*HI endonuclease with its DNA recognition site GGATCC where base–analogue substitutions produce equal effects on the transition-state and the E′S′ complex:

(a) Replacement of T with U disrupts the T-CH_3 interaction[45] with Val156 ($\Delta\Delta G^\circ_{bind}$ = +1.9 ± 0.2 kcal/mol), but does not affect $\Delta\Delta G^{\circ\ddagger}$.
(b) Replacement of G2 by ^{7C}G disrupts water-mediated hydrogen bonds[45] from Asn116 and Arg122 from the other subunit; there is a significant effect on binding ($\Delta\Delta G^\circ_{bind}$ = +2.6 ± 0.2 kcal/mol), but $\Delta\Delta G^{\circ\ddagger} = 0$.
(c) Replacement of A by ^{7C}A disrupts a water-mediated hydrogen bond from Asn116-peptide-NH.[45] There is a relatively large binding penalty ($\Delta\Delta G^\circ_{bind}$ = +2.7 ± 0.3 kcal/mol), and $\Delta\Delta G^{\circ\ddagger} \approx 0$.

In contrast to all the preceding cases that have involved the modification of direct protein–base or protein–phosphate contacts, we have also found a class of substrate modifications that do not alter direct protein–DNA contacts, but apparently affect ΔG°_{bind} by changing the energy required to distort the DNA. These involve changes in the *noncontacted* base sequences flanking the recognition site. We have studied this most systematically for the *Eco*RI endonuclease–DNA interaction (M. R. Kurpiewski et al., unpublished results), but the same kinds of phenomena are observed for *Eco*RV[7] and *Bam*HI.[65]

Figure 9 shows data for six representative 22-bp oligonucleotides (of 40 sequences analyzed), each containing the GAATTC recognition site but differing in flanking sequence. Relative to the reference sequence, altering the flanking three base pairs on either side produces a range of $\Delta\Delta G^\circ_{bind}$ up to +3.7 kcal/mol, equivalent to a 500-fold inhibition of binding. For all sequences except the flanking A-tract (known to confer unusual rigidity),[106,116] there is no change in cleavage rate constant. That is, the free energy level of ES‡ is affected by the flanking sequence by the same increment as that of E′S′.

A number of lines of evidence indicate that flanking sequence variation influences the protein–DNA interaction by affecting DNA conformation or distortability rather than by altering particular protein–base or protein–phosphate contacts:

(a) There are no protein contacts to the base pairs outside the GAATTC site in the crystal structures of the protein complexes with TCGCGAATTCGCG[40–42] or TCGTGGAATTCCACG.[117] The two sequences differ in ΔG°_{bind} by 2.6 kcal/mol (the former context plots at 0,0, identical to the reference sequence shown in Figure 9; the latter context to four base pairs is identical to a sequence in Figure 9), but the crystal structures show identical protein–base and protein–phosphate contacts in the complexes.
(b) A number of conservative base–analogue substitutions in the two base pairs flanking the site have no effect at all on ΔG°_{bind}.
(c) The ethylation-interference footprints are indistinguishable for the sequences shown in Figure 9, indicating that there is no major change in protein–phosphate contacts (M. R. Kurpiewski et al., unpublished results).
(d) This inference is supported by the observation that all the sequences have identical salt dependencies of equilibrium binding; that is, $d \log K_{obs}/d \log[NaCl]$ is invariant. This indicates equal stoichiometries of cation release upon binding.
(e) For four of the sequences in Figure 9 (sequences 1, 2, 4, 6) measurements of K_A as a function of solution osmolarity indicate that the stoichiometry of water release upon complex formation is the same within experimental error (D. Cao, D. J. Chi, M. R. Kurpiewski, and L. Jen-Jacobson, unpublished results).
(f) Flanking sequences that promote strong specific binding at the GAATTC site do not improve binding at *Eco*RI* sites (where ''adaptive'' complexes form without the same DNA distortion[118]) and do not alter the intrinsic binding constant for nonspecific sites (M. R. Kurpiewski et al., unpublished results).

The only explanation consistent with all these data is that flanking sequence influences the energy required to achieve the DNA distortion in the complex. It is hardly surprising that base sequence should influence the local flexibility of DNA, although it is somewhat unexpected that the changes in ΔG°_{bind} (Figure 9) are so large. The fact that flanking sequence has equal effects on the E′S′ and ES‡ complexes further confirms the inference from our base–analogue and phosphate–analogue data (Table II) that the two complexes have very similar overall DNA conformations.

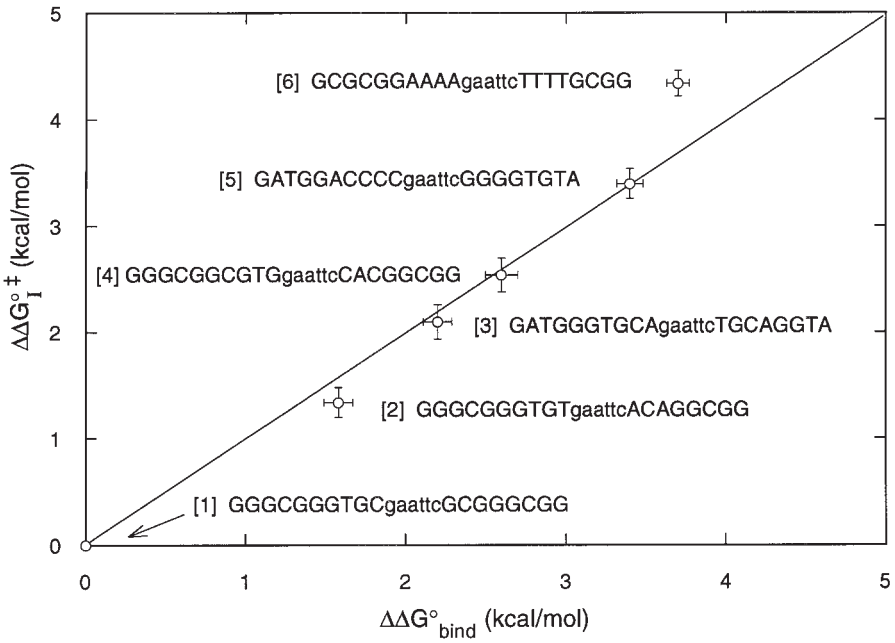

FIGURE 9 Effect of flanking sequence context on binding and cleavage by *Eco*RI endonuclease. All values are given relative to the reference sequence at 0,0. The equilibrium binding affinity to the reference sequence and the sequence in the crystal structure[40] is the same. The transition-state interaction free energy difference (ordinate) is given by $\Delta\Delta G^{\circ\ddagger}_{I} = \Delta\Delta G^{\circ}_{bind} + \Delta\Delta G^{\circ\ddagger}$. The line of unit slope therefore represents $\Delta\Delta G^{\circ\ddagger}_{I} = \Delta\Delta G^{\circ}_{bind}$, that is, $\Delta\Delta G^{\circ\ddagger} = 0$. Only the sequence with flanking AAAA has $\Delta\Delta G^{\circ\ddagger}$ significantly different from 0.

BINDING ENERGY IN THE ENZYME–SUBSTRATE AND ENZYME–TRANSITION STATE COMPLEXES

We can now return with more insight to the dilemma confronting a site-specific DNA-binding protein. In order to overcome the competitive effect of the excess nonspecific binding sites and localize to a recognition site long enough to carry out a function, it must bind very tightly to the recognition site. (This is true for both catalytic and ''noncatalytic'' DNA-binding proteins.) The oft-quoted ''fundamentalist'' position[119] on enzyme catalysis, however, holds that enzymes achieve catalytic power only by stabilizing the transition state and very tight binding to the substrate has classically been regarded as ''anticatalytic.'' That is, tighter substrate binding would simply drive the E′S′ complex into a deeper energy well (Figure 6), creating a larger barrier $\Delta G^{\circ\ddagger}$ between E′S′ and ES‡, thereby decreasing the reaction rate. In this view, increased specificity (tighter binding to the ''correct'' substrate) and increased catalytic efficiency appear to be mutually contradictory objectives in constructing any enzyme.

A resolution of this conceptual difficulty was presented over 20 years ago by Albery and Knowles[122] in the postulate of ''uniform binding,'' which referred to enzyme–substrate interactions that contribute precisely equally to stabilizing E′S′ and the transition state complex ES‡. As a result of uniform binding, the addition of a stabilizing interaction to the E′S′ complex not only increases binding specificity but also leaves the activation free energy $\Delta G^{\circ\ddagger}$ unchanged, so the problem of the ''deeper energy well'' no longer pertains. In the terms used in this paper, adding such a stabilizing interaction produces $\Delta\Delta G^{\circ}_{bind} < 0$ and $\Delta\Delta G^{\circ\ddagger} = 0$. It is crucial to note that this also increases the overall catalytic efficiency, because the reaction rate is determined by the total free energy barrier in going from free enzyme and DNA to the ES‡ complex. This is known as the transition-state interaction free energy change, and is given by

$$\Delta G^{\circ\ddagger}_{I} = \Delta G^{\circ}_{bind} + \Delta G^{\circ\ddagger} \qquad (4)$$

and for any modification of the enzyme or the substrate, the corresponding changes are

$$\Delta\Delta G_{I}^{\circ\ddagger} = \Delta\Delta G_{bind}^{\circ} + \Delta\Delta G^{\circ\ddagger} \quad (5)$$

It follows that so long as the E′S′ and ES‡ complexes are equally affected so that $\Delta\Delta G^{\circ\ddagger} = 0$, any modification that improves binding will also increase the overall rate of reaction.

This situation is easier to understand from a structural perspective for macromolecular substrates than for small-molecule substrates. The interface between an enzyme and a macromolecular substrate (e.g., a DNA recognition site) is much larger than the binding site for a small-molecule substrate, so it is possible to build more and more enzyme–substrate interactions in aid of recognition specificity without these being too close to the actual reaction center. This separation in space also permits many specific enzyme–substrate interactions to be made and maintained throughout the reaction without undue constraint on the changes that must take place in the immediate vicinity of the reaction site to reach the transition state. Thus, in speaking of enzyme–substrate complexes with very extended interfaces, it is useful to distinguish between the ''local'' and ''global'' features of the transition-state interactions. The idea is similar to the ''split-site'' hypothesis of Menger[121] and its refinement by Murphy.[122]

The data presented above indicate that for enzymes that work at specific DNA sites, the ''global'' transition-state interactions contribute to ''uniform binding.'' The ''local'' transition-state interactions are likely to stabilize the transition state preferentially in the immediate vicinity of the reaction center. As we accumulate more and more examples of specific interactions that contribute equally to E′S′ and ES‡, and evidence that structural features like the DNA distortion are the same in both complexes, we gain increasing confidence that the transition-state complex must have a structure extremely close to the structure of the E′S′ complex as it exists, for example, in a crystal. For this reason we have termed the E′S′ complex a ''pretransition state'' complex,[49,54] to indicate that it already embodies many structural features of the transition-state complex. Most importantly, the close resemblance between E′S′ and ES‡ should be a general phenomenon for proteins that interact with specific DNA sites.

This strategy for increasing specificity by lowering the free energy of both E′S′ and ES‡ has a potential problem in that its continued application will eventually make product-release rate-limiting. Hackney[74] has pointed out that this difficulty can be circumvented by the use of ''induced fit'' to introduce an energetically *unfavorable* factor that raises the free energy levels of E′S′, ES‡, and EP equally. It is well known, however, that induced fit of the enzyme cannot increase specificity.[123] DNA-binding proteins have therefore made use of an analogous solution to the problem by distorting the DNA substrate rather than the enzyme. In this scenario, differences in distortability between different DNA sequences can contribute to specificity. It has been suggested, for example, that the energy required to achieve the ''kinked'' DNA conformation in the *Eco*RI complex is least unfavorable for the GAATTC site, and more unfavorable for any other site.[6,54] This use of the net unfavorable energy of DNA distortion may explain why so many site-specific DNA binding proteins distort their DNA sites.

A particularly cogent discussion of catalysis pertinent to the pretransition state strategy has been presented by Benkovic and co-workers.[124] A key concept is that interactions in the E′S′ complex that contribute to the ''work required to go between the ground and transition states'' will enhance catalysis, whereas those that exert ''force'' (i.e., the ability to do work) that is not directed along the reaction coordinate toward the transition state will inhibit catalysis. The pretransition state complex, embodying many of the structural features of ES‡, is explicitly conceived as a step along the reaction coordinate, such that processes completed in forming E′S′ (e.g., DNA distortion, release of water from nonpolar surfaces, conformational changes in the enzyme) do not have to be repeated when progressing to ES‡. This strategy for using binding energy to provide specificity and drive catalysis was explicitly discussed by Jencks two decades ago.[125]

For *Eco*RI and *Bam*HI endonucleases, each E′S′ complex must be very nearly perfectly optimized by evolution as a step on the path to the transition state. For this reason, it is rare to observe any modification to the substrate that increases a catalytic rate constant (although examples are known[48,54]) and such cases show relatively minor rate enhancements. By contrast, many substrate modifications decrease catalytic rate constants by as much as six orders of magnitude, for reasons outlined below.

''ADAPTIVE'' COMPLEXES AND SITE DISCRIMINATION

A corollary of the forgoing discussion is that any protein–DNA complex that deviates structurally

from the pretransition-state configuration will lie somewhere off the direct path to the transition state, and that catalysis will consequently be slowed. This is important because enzyme–substrate binding complexes can change (''adapt'') in various ways to minimize binding free energy, but transition states require a much higher level of precision, dictated by the local requirements of the reaction chemistry.

Consider an enzyme binding to an incorrect DNA site that differs not much (say, by one base pair) from the correct recognition site. Some of the correct, energetically favorable, protein–base interactions may be impossible; there may be an additional energetic penalty[6] for the incorrect apposition of hydrogen-bond donors (acceptors) on the protein to donors (acceptors) on the DNA; the hydrophobic driving force may be decreased if a less intimate interface is formed; and the cost of distorting the DNA may be increased.[6] The protein may nevertheless attempt to compensate partially for these factors by forming new favorable interactions with the DNA, so as to minimize $\Delta G^\circ_{\text{bind}}$. In doing this, the protein need not be constrained to minor variations on the structural theme of the ''correct'' complex, but can ''adapt'' to any degree into completely new structures so long as this minimizes $\Delta G^\circ_{\text{bind}}$. As a result, $\Delta\Delta G^\circ_{\text{bind}}$ may be small even though profound structural change has occurred.

The crucial point, though, is that such structures have an altered relationship of protein and DNA and *are not on the direct path to the transition state,* and consequently will have reduced rates of catalysis. Whereas the energetic costs paid on the way to the canonical pretransition-state E′S′ need not be repaid in going to $ES^\ddagger$, they do need to be paid (at least in part) in going from an ''adapted'' ES complex to $ES^\ddagger$. Thus, the pretransition-state concept provides a rationale for the exercise of specificity in the catalytic step as well as in protein–DNA binding.

Experimental evidence for such ''adaptive'' changes has been obtained for the restriction endonucleases *Eco*RI,[6,9,114] *Eco*RV,[7] and *Bam*HI.[65] In each case, binding to sites with one incorrect base pair triggers dramatic stereotypical changes in phosphate contacts[6] (Figure 8c), in the salt dependence of binding (Refs. 6 and 7, and M. R. Kurpiewski et al., unpublished results), in the pH dependence of binding,[7] and in the observed ΔC°_p of binding.[65] (Not all these properties have been examined for all three enzymes.) In the case of *Eco*RI, crystallographic evidence[118] confirms major structural changes (in both protein and DNA) in the complex with the *Eco*RI* site GACTTC. Similar adaptations are observed when *Eco*RI binds to a site containing one or two N6-methyl-adenines.[114] In all these situations, the profound structural change produces profound change in the elementary cleavage-rate constants, for *Eco*RI ranging from 70-fold to $> 10^6$-fold for various *Eco*RI* sites[6] and up to 600,000-fold for a double-methylated site.[114] When the DNA modifications are confined to one half-site, the adaptation is usually structurally asymmetric[6,13,114] (Figure 8c,d) and the catalytic rate constants in the normal half-site (k_1) and the abnormal half-site (k_2) change unequally, in some cases markedly so.[6,9,13,114] This is important to discrimination in vivo because the rate constant for second-strand cleavage at *Eco*RI* sites is so low that the enzyme dissociates before second-strand cleavage can occur, leaving a repairable single-strand nick rather than a potentially lethal double-strand break.[6]

Even relatively mild structural perturbations can change catalytic rate constants, as illustrated by the ^{7C}GAATTC site (Figure 8d). Adaptation is signaled largely by increased ethylation interference at the upstream phosphate pCAGAATTC (white bar) in the unmodified half-site and at the scissile bond GpAATTC (yellow bars) in both half sites; the crucial ''clamp'' contacts at pApGAApTTC are essentially unperturbed. The overall $\Delta\Delta G^\circ_{\text{bind}}$ is small (+1.4 kcal/mol), but cleavage rate constants decrease 10-fold in the unmodified half-site (k_1) and 30-fold in the modified half-site (k_2). These subtle asymmetric structural adaptations apparently cause enough difference between the ''adapted'' E′S′ complex and $ES^\ddagger$ to affect $\Delta\Delta G^{\circ\ddagger}$ significantly.

Structural adaptation may well turn out to be a universal feature of the binding of site-specific proteins to variant DNA sequences. Adaptive changes in complexes of *lac* repressor with mutant (O^c) operator sites were first proposed[11] to account for nonadditive changes in $\Delta G^\circ_{\text{bind}}$ for various single base-pair changes. More recently, Frank et al.[12] have observed that complexes with variant *lac* operators also show differences in the salt dependence of binding and in ΔC°_p upon complex formation. Adaptive binding of Cro repressor to mutant operators was inferred[20] from altered $\Delta S^\circ_{\text{bind}}/\Delta C^\circ_p$ ratios. Both thermodynamic changes[25] and a crystal structure of a noncognate complex[126] show that the glucocorticoid receptor DNA-binding domain binds adaptively at noncognate sites. Structural adaptation has also been invoked to explain discrimination between alternative target sequences by transcription factor

NF-κB[127] and members of the transcription factor family bZIP (e.g., GCN4).[128]

PRODUCTIVE AND NONPRODUCTIVE BINDING OF ''NONCATALYTIC'' PROTEINS

Might the concept of the pretransition-state complex be useful in thinking about DNA-binding proteins that are usually considered ''noncatalytic,'' such as general transcription factors? Although information on these systems is too incomplete to provide a conclusive answer, this way of considering interacting protein–DNA systems may suggest some interesting experimental questions.

As an example, consider the assembly of the preinitiation complex of eukaryotic RNA polymerase II. The ultimate catalytic event is the initiation of transcription—that is, the formation of the first internucleotide bond. The primary DNA-binding activity for this complex resides[129] in the TATA-box binding protein (TBP) of transcription factor IID (TFIID). There are now crystal structures of the TBP–DNA complex[39,130,131] and of ternary complexes in which TFIIB[132] or TFIIA[133,134] are also bound. These structures reveal a dramatic DNA distortion (bending and minor-groove opening) upon binding of TBP. This distortion is retained almost without change in the two ternary complexes. It appears that the DNA distortion is necessary for TFIIA and TFIIB (binding on opposite sides of TBP) to make their interactions with DNA and with TBP.

The analogies with the pretransition-state complex are intriguing. Payment is made early in the process for the major unfavorable energy of DNA distortion, utilizing the binding energy provided by the binary interaction of TBP with a ''good'' promoter. Both the precision in the DNA distortion and concomitant conformational change in TBP[39] seem to be required to progress to the next step(s), which although not themselves catalytic, are essential for the ultimate assembly of the catalytic complex. If binding does not lead to the correct overall conformation, including the requisite DNA distortion, the complex presumably will not lie on the path to the ultimate transition state. So far as we know, however, there are no biochemical data to determine if TBP forms stable but ''adaptive'' (and therefore transcriptionally inactive or nonproductive) complexes at some promoters or promoter variants.

The interaction of transcriptional activators with TBP, TBP-associated factors, or other general transcription factors may provide not only additional binding energy but also additional DNA-site specificity.[135–137] For example, an intrinsically ''poor'' promoter may support only weak transcription because the favorable free energy available from the direct interaction of TBP with that promoter is insufficient to supply the sequence-dependent free energy required to distort the DNA. With an activator protein bound at an appropriate location, however, the additional binding energy available from interaction of activator with one or more proteins of the preinitiation complex may make the initiation of transcription possible. The binding energy available from interaction between TBP (or TFIIB) and various mutant forms of the GAL4 activator correlates reasonably well with the ability of the mutant proteins to activate transcription in vivo.[138]

The additional free energy made available by protein–protein interactions may be used in various ways. For example, the bacterial catabolite gene activator protein CAP, which contains two distinct activating regions,[139] appears to stimulate the formation of the initial or ''closed'' complex between RNA polymerase and the *lac* promoter[140] but at the *gal* promoter it also stimulates conversion to the catalytically active ''open'' complex.[141] In either case, the net effect is to provide energy along the path to the transition state.

CONCLUSIONS

After binding to their specific recognition sites, site-specific DNA-binding proteins may carry out either obviously catalytic functions or assembly functions that lead to catalytic events. Such proteins must bind very tightly to their recognition sites in order to overcome competitive inhibition by the large excess of nonspecific DNA sites. Tight binding to the substrate might be anticatalytic if it retarded progression to the transition state, but some DNA binding proteins have avoided this potential difficulty by utilizing the stabilizing interactions of the protein–DNA recognition complex to stabilize the transition state. That is, the complementary interface and strong binding interactions, formed in the specific recognition complex and extending far beyond the immediate vicinity of the reaction center, are retained in the transition-state complex, so that the recognition complex and the transition-state complex share a strong structural resemblance. As a

result, favorable free energy components of binding, including those from direct protein–base and protein–phosphate contacts and from the hydrophobic effect, are efficiently utilized to promote formation of the transition-state complex. The free-energy costs of entropy losses, of conformational changes in the protein, and of distorting the DNA recognition site are paid once in the binding-recognition step, and need not be paid again in the catalytic step.

This overall strategy also enhances accurate selection of correct DNA sites for catalysis, because binding at variant or incorrect DNA sites often (perhaps always) requires ''adaptive'' structural changes in the complex. These altered structures no longer bear a close resemblance to the transition state and thus do not lie on the direct path to the transition state. Consequently, when a completely correct protein–DNA recognition interface is not formed, rates of catalysis are dramatically slower.

This work was supported by a grant from NIH (GM-29207). I am indebted to my students and co-workers who made this work possible, in particular Michael Kurpiewski, Lisa Engler, Jennifer Ames, David Chi, Dengfeng Cao, Jianwen Zhang, and David Lesser. We are grateful to Bernard Connolly, Wojciech Stec, and the members of their laboratories, who provided purified oligodeoxynucleotides containing base and phosphate backbone analogs for these studies.

REFERENCES

1. von Hippel, P. H. & McGhee, J. D. (1972) *Ann. Rev. Biochem.* **41,** 231–300.
2. Lin, S.-Y. & Riggs, A. D. (1975) *Cell* **4,** 107–111.
3. von Hippel, P. H. & Berg, O. G. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 1608–1612.
4. Record, M. T., Jr. & Spolar, R. S. (1990) in *Nonspecific DNA-Protein Interactions*, Revzin, A., Ed., CRC Press, Boca Raton, FL, pp. 33–69.
5. von Hippel, P. H. (1994) *Science* **263,** 769–770.
6. Lesser, D. R., Kurpiewski, M. R. & Jen-Jacobson, L. (1990) *Science* **250,** 776–786.
7. Engler, L. E., Welch, K. K. & Jen-Jacobson, L. (1997) *J. Mol. Biol.* **269,** 82–101.
8. Siebenlist, U. & Gilbert, W. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 122–126.
9. Jen-Jacobson, L. (1995) *Methods Enzymol.* **259,** 305–344.
10. Thielking, V., Alves, J., Fliess, A., Maass, G. & Pingoud, A. (1990) *Biochemistry* **29,** 4682–4691.
11. Mossing, M. C. & Record, M. T., Jr. (1985) *J. Mol. Biol.* **186,** 295–305.
12. Frank, D. E., Saecker, R. M., Bond, J. P., Capp, M. W., Todikov, O. V., Melcher, S. E., Levandoski, M. M. & Record, M. T., Jr. (1997) *J. Mol. Biol.* **267,** 1186–1206.
13. Jen-Jacobson, L., Lesser, D. R. & Kurpiewski, M. R. (1991) in *Nucleic Acids and Molecular Biology*, Vol. 5, Eckstein, F. & Lilley, D. M. J., Eds., Springer-Verlag, Heidelberg, pp. 141–170.
14. Record, M. T., Jr., Lohman, T. M. & deHaseth, P. L. (1976) *J. Mol. Biol.* **107,** 145–158.
15. Record, M. T., Jr., Ha, J.-H. & Fisher, M. A. (1991) *Methods Enzymol.* **208,** 291–343.
16. Terry, B. J., Jack, W. E., Rubin, R. A. & Modrich, P. (1983) *J. Biol. Chem.* **258,** 9820–9825.
17. Jen-Jacobson, L., Kurpiewski, M., Lesser, D., Grable, J., Boyer, H. W., Rosenberg, J. M. & Greene, P. J. (1983) *J. Biol. Chem.* **258,** 14638–14646.
18. Senear, D. F. & Ackers, G. K. (1990) *Biochemistry* **29,** 6568–6577.
19. Ha, J.-H., Spolar, R. S. & Record, M. T., Jr. (1989) *J. Mol. Biol.* **209,** 801–816.
20. Takeda, Y., Ross, P. D. & Mudd, C. P. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 8180–8184.
21. Jin, L., Yang, J. & Carey, J. (1993) *Biochemistry* **32,** 7302–7309.
22. Ladbury, J. E., Wright, J. G., Sturtevant, J. M. & Sigler, P. B. (1994) *J. Mol. Biol.* **218,** 669–681.
23. Hyre, D. E. & Spicer, L. D. (1995) *Biochemistry* **34,** 3212–3221.
24. Merabet, E. & Ackers, G. K. (1995) *Biochemistry* **34,** 8554–8563.
25. Lundbäck, T. & Härd, T. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 4754–4759.
26. Petri, V., Hsieh, M. & Brenowitz, M. (1995) *Biochemistry* **34,** 9977–9984.
27. Parsegian, V. A., Rand, R. P. & Rau, D. C. (1995) *Methods Enzymol.* **259,** 43–94.
28. Garner, M. M. & Rau, D. C. (1995) *EMBO J.* **14,** 1257–1263.
29. Sidorova, N. Y. & Rau, D. C. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 12272–12277.
30. Senear, D. F. & Batey, R. (1991) *Biochemistry* **30,** 6677–6688.
31. von Hippel, P. H. (1979) in *Biological Regulation and Development*, Vol. 1, Goldberger, R. F., Ed., Plenum Press, N.Y., pp. 279–347.
32. Yang, S.-W. & Nash, H. A. (1995) *EMBO J.* **14,** 6292–6300.
33. Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. USA* **73,** 804–808.
34. Rosenberg, J. M. & Greene, P. (1982) *DNA* **1,** 117–124.
35. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988) *Science* **242,** 899–907.
36. Harrison, S. C. & Aggarwal, A. K. (1990) *Ann. Rev. Biochem.* **59,** 933–969.

37. Steitz, T. A. (1990) *Quart. Rev. Biophys.* **23,** 205–208.
38. Pabo, C. O., Aggarwal, A. K., Jordan, S. R., Beamer, L. J., Obeysekare, U. R. & Harrison, S. C. (1990) *Science* **247,** 1210–1213.
39. Kim, J. L. & Burley, S. K. (1994) *Nat. Struct. Biol.* **1,** 638–657.
40. Rosenberg, J. M. (1991) *Curr. Opin. Struct. Biol.* **1,** 104–113.
41. Kim, Y., Choi, J., Grable, J. C., Greene, P., Hager, P. & Rosenberg, J. M. (1994) in *Structural Biology, The State of the Art. Proc. of the Eighth Conversation SUNY,* Sarma, R. H. & Sarma, M. K., Eds., Adenine Press, New York.
42. Choi, J., Kim, Y., Wilkosz, P. A., Chandrasekhar, K., Grable, J. C., Greene, P., Hager, P. & Rosenberg, J. M. (in press) *Proteins Struct. Funct. Genet.*
43. Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmostein, R. Q., Luisi, B. F. & Sigler, P. (1988) *Nature* **335,** 321–329.
44. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Afflolter, M., Otting, G. & Wütrich, K. (1994) *Cell* **78,** 211–223.
45. Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1995) *Science* **269,** 656–663.
46. Wolberger, C. (1996) *Curr. Opin. Struct. Biol.* **6,** 62–68.
47. Schwabe, J. W. R. (1997) *Curr. Opin. Struct. Biol.* **7,** 126–134.
48. Lesser, D. R., Grajkowski, A., Kurpiewski, M. R., Koziolkiewicz, M., Stec, W. J. & Jen-Jacobson, L. (1992) *J. Biol. Chem.* **267,** 24810–24818.
49. Kurpiewski, M. R., Koziolkiewicz, M., Wilk, A., Stec, W. J. & Jen-Jacobson, L. (1996) *Biochemistry* **35,** 8846–8854.
50. Ebright, R. H. (1986) in *Protein Structure Design and Folding,* Oxender, D. L., Ed., Alan R. Liss, New York, pp. 207–219.
51. Kissinger, C. R., Beishan, L., Marin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990) *Cell* **63,** 579–590.
52. Schultz, S. C., Shields, G. C. & Steitz, T. A. (1991) *Science* **253,** 1001–1007.
53. Parkinson, G., Wilson, C., Gunasekera, A., Ebright, Y. W., Ebright, R. H. & Berman, H. M. (1996) *J. Mol. Biol.* **260,** 395–408.
54. Lesser, D. R., Kurpiewski, M. R., Waters, T., Connolly, B. A. & Jen-Jacobson, L. (1993) *Proc. Natl. Acad. Sci.* **90,** 7548–7552.
55. Koudelka, G. B. (1993) in *Nucleic Acids and Molecular Biology,* Vol. 7, Eckstein, F. & Lilley, D. M. J., Eds., Springer-Verlag, Heidelberg, pp. 16–27.
56. Harrington, R. E. & Winicov, I. (1994) *Progr. Nucleic Acid Res.* **47,** 195–269.
57. Suzuki, M. & Yagi, N. (1995) *Nucleic Acids Res.* **23,** 2083–2091.
58. Strahs, D. & Brenowitz, M. (1994) *J. Mol. Biol.* **244,** 494–510.
59. Härd, T. & Lundbäck, T. (1996) *Biophys. Chem.* **62,** 121–139.
60. Sauer, R. T., Jordan, S. R. & Pabo, C. O. (1990) *Adv. Protein Chem.* **40,** 1–61.
61. Takeda, Y., Sarai, A. & Rivera, V. M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 439–443.
62. Brown, B. M. & Sauer, R. T. (1993) *Biochemistry* **32,** 1354–1363.
63. Smith, T. L. & Sauer, R. T. (1995) *J. Mol. Biol.* **249,** 729–742.
64. Waldburger, D. D. & Sauer, R. T. (1995) *Biochemistry* **34,** 13109–13116.
65. Engler, L. E., Kucera, R., Schildkraut, I. & Jen-Jacobson, L., manuscript in preparation.
66. Misra, V. K., Hecht, J. L., Sharpt, K. A., Friedman, R. A. & Honig, B. (1994) *J. Mol. Biol.* **238,** 264–280.
67. Zhang, W., Bond, J. P., Anderson, C. F., Lohman, T. M. & Record, M. T., Jr. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 2511–2516.
68. Padmanabhan, S., Zhang, W., Capp, M. W., Anderson, C. F. & Record, M. T., Jr. (1997) *Biochemistry* **36,** 5193–5206.
69. Johnson, K. A. & Benkovic, S. J. (1990) in *The Enzymes,* Vol. XIX, Sigman, D. S. & Boyer, P. D., Eds., Academic Press, New York, pp. 159–211.
70. Kumar, S., Duan, Y., Kollman, P. A. & Rosenberg, J. M. (1994) *J. Biomol. Struct. Dynam.* **12,** 487–525.
71. Spolar, R. S. & Record, M. T., Jr. (1994) *Science* **263,** 777–784.
72. Richards, F. M. (1977) *Ann. Rev. Biophys. Bioeng.* **6,** 151–176.
73. Sharp, K. A., Nicholls, A., Fine, R. & Honig, B. (1991) *Science* **252,** 106–109.
74. Hackney, D. (1990) in *The Enzymes,* Vol. XIX, Sigman, D. S. & Boyer, P. D., Eds., Academic Press, New York, pp. 1–36.
75. Chandrasekhar, K., Wilkosz, P., Samudzi, C. & Rosenberg, J. M., manuscript in preparation.
76. Sturtevant, J. M. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 2236–2240.
77. Duan, Y., Wilkosz, P. & Rosenberg, J. (1996) *J. Mol. Biol.* **264,** 546–555.
78. Finucane, M. D. & Jardetzky, O. (1995) *J. Mol. Biol.* **253,** 576–589.
79. Morton, C. J. & Ladbury, J. E. (1996) *Protein Sci.* **5,** 2115–2118.
80. Brenowitz, M., Senear, D. F. & Ackers, G. F. (1989) *Nucleic Acids Res.* **17,** 3747–3755.
81. Baranger, A. M., Palmer, C. R., Hamm, M. K., Giebler, H. A., Brauweiler, A., Nyborg, J. K. & Schepartz, A. (1995) *Nature* **376,** 606–608.

82. Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K. & Wilson, K. S. (1993) *EMBO J.* **12,** 1781–1795.
83. Luisi, B. F., Xu, W. X., Otwinowski, Z., Freedman, L. P., Yamamoto, K. R. & Sigler, P. B. (1991) *Nature* **352,** 497–505.
84. Takahashi, M., Blazy, B., Baudras, A. & Hillen, W. (1983) *J. Mol. Biol.* **167,** 895–899.
85. Ferrari, M. E. & Lohman, T. M. (1994) *Biochemistry* **33,** 12896–12910.
86. Riggs, A. D., Bourgeois, S. & Cohn, M. (1970) *J. Mol. Biol.* **53,** 401–417.
87. Winter, R. B., Berg, O. G. & von Hippel, P. H. (1981) *Biochemistry* **20,** 6961–6977.
88. von Hippel, P. H. & Berg, O. G. (1989) *J. Biol. Chem.* **264,** 675–678.
89. Terry, B. J., Jack, W. E. & Modrich, P. (1987) in *Gene Amplification and Analysis, Vol. V, Restriction Endonucleases and Methylases,* Chirikjian, J. G., Ed., Elsevier/North-Holland, Amsterdam, pp. 103–117.
90. Jack, W. E., Terry, B. J. & Modrich, P. (1992) *Proc. Natl. Acad. Sci. USA* **79,** 4010–4014.
91. Lohman, T. M. (1985) *CRC Crit. Rev. Biochem.* **19,** 191–245.
92. Nardone, G., George, J. & Chirikjian, J. G. (1986) *J. Biol. Chem.* **261,** 12128–12133.
93. Barkley, M. D. (1981) *Biochemistry* **20,** 3833–3842.
94. Surby, M. A. & Reich, N. O. (1996) *Biochemistry* **35,** 2201–2208.
95. Ehbrecht, H.-J., Pingoud, A., Urbanke, C., Maass, G. & Gualerzi, C. (1985) *J. Biol. Chem.* **260,** 6160–6166.
96. Dowd, D. R. & Lloyd, R. S. (1990) *J. Biol. Chem.* **265,** 3424–3431.
97. Jeltsch, A., Wenz, C., Stahl, F. & Pingoud, A. (1996) *EMBO J.* **15,** 5104–5111.
98. Leirmo, S. & Record, M. T., Jr. (1990) in *Nucleic Acids and Molecular Biology,* Vol. 5, Eckstein, F. & Lilley, D. M. J., Eds., Springer-Verlag, Heidelberg, pp. 123–151.
99. Jeltsch, A., Alves, J., Wolfes, H., Maass, G. & Pingoud, A. (1994) *Biochemistry* **33,** 10215–10219.
100. Ghosh, G., Duyne, G. V., Ghosh, S. & Sigler, P. B. (1995) *Nature* **373,** 303–310.
101. Müller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. (1995) *Nature* **373,** 311–317.
102. Rice, P. A. (1997) *Curr. Opin. Struct. Biol.* **7,** 86–93.
103. Halford, S. E. & Johnson, N. P. (1983) *Biochem. J.* **211,** 405–415.
104. Alves, J., Urbanke, C., Fliess, A., Maass, G. & Pingoud, A. (1989) *Biochemistry* **28,** 7879–7888.
105. Gorin, A. A., Zhurkin, V. B. & Olson, W. K. (1995) *J. Mol. Biol.* **247,** 34–48.
106. Zhurkin, V. B., Ulyanov, N. B., Gorin, A. A. & Jernigan, R. L. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 7046–7050.
107. Grove, A., Galeone, A., Mayol, L. & Geiduschek, E. P. (1996) *J. Mol. Biol.* **260,** 120–125.
108. Winkler, F. K. (1992) *Curr. Opin. Struct. Biol.* **2,** 93–99.
109. Aggarwal, A. K. (1995) *Curr. Opin. Struct. Biol.* **5,** 11–19.
110. King, K., Benkovic, S. J. & Modrich, P. (1989) *J. Biol. Chem.* **264,** 11807–11815.
111. Jeltsch, A., Alves, J., Wolfes, H., Maas, G. & Pingoud, A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8499–8503.
112. Kostrewa, D. & Winkler, F. K. (1995) *Biochemistry* **34,** 683–696.
113. Vipond, I. B., Baldwin, G. S. & Halford, S. E. (1995) *Biochemistry* **34,** 687–704.
114. Jen-Jacobson, L., Engler, L. E., Lesser, D. R., Kurpiewski, M. R., Yee, C. & McVerry, B. (1996) *EMBO J.* **15,** 2870–2882.
115. Zhang, J., Kurpiewski, M. R. & Jen-Jacobson, L., manuscript in preparation.
116. Leroy, J. L., Charretier, E., Kochoyan, M. & Guéron, M. (1988) *Biochemistry* **29,** 4227–4234.
117. Grable, J. C. (1990) Ph.D. thesis, University of Pittsburgh, Pittsburgh.
118. Wilkosz, P. (1993) Ph.D. thesis, University of Pittsburgh, Pittsburgh.
119. Schowen, R. L. (1978) *Transition States of Biochemical Processes,* Plenum Press, New York, chap. 2.
120. Albery, W. J. & Knowles, J. R. (1976) *Biochemistry* **15,** 5631–5640.
121. Menger, F. M. (1992) *Biochemistry* **34,** 5368–5373.
122. Murphy, D. J. (1992) *Biochemistry* **34,** 4507–4510.
123. Fersht, A. (1985) *Enzyme Structure and Mechanism,* 2nd ed., W. H. Freeman & Co., New York.
124. Cannon, W. R., Singleton, S. F. & Benkovic, S. J. (1996) *Nature Struct. Biol.* **3,** 821–833.
125. Jencks, W. P. (1975) *Adv. Enzymol.* **43,** 219–410.
126. Gewirth, D. T. & Sigler, P. B. (1995) *Nature Struct. Biol.* **2,** 386–394.
127. Müller, C. W., Rey, F. A. & Harrison, S. C. (1996) *Nature Struct. Biol.* **3,** 224–227.
128. Metallo, S. J. & Schepartz, A. (1994) *Chem. Biol.* **1,** 143–151.
129. Burley, S. K. & Roeder, R. G. (1996) *Ann. Rev. Biochem.* **65,** 769–799.
130. Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993) *Nature* **365,** 320–327.
131. Kim, Y.-C., Geiger, H. H., Hahn, S. & Sigler, P. B. (1993) *Nature* **365,** 512–520.
132. Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G. & Burley, S. K. (1995) *Nature* **377,** 119–127.

133. Geiger, J. H., Hahn, S., Lee, S. & Sigler, P. B. (1996) *Science* **272,** 830–837.
134. Tan, S., Hunziker, Y., Sargent, D. R. & Richmond, T. J. (1996) *Nature* **381,** 127–134.
135. Frankel, A. D. & Kim, P. S. (1991) *Cell* **65,** 717–719.
136. Goodrich, J. A., Cutler, G. & Tjian, R. (1996) *Cell* **84,** 825–830.
137. Ptashne, M. & Gann, A. (1997) *Nature* **386,** 569–577.
138. Wu, Y., Reece, R. J. & Ptashne, M. (1996) *EMBO J.* **15,** 3951–3963.
139. Niu, W., Kim, Y., Tau, G., Heyduk, T. & Ebright, R. H. (1996) *Cell* **87,** 1123–1134.
140. Malan, T. P., Buc, H. & McClure, W. R. (1982) *J. Mol. Biol.* **157,** 493–525.
141. Bell, A., Gaston, K., Williams, R., Chapman, K., Kolb, A., Buc, H., Minchin, S., Williams, J. & Busby, S. (1990) *Nucleic Acids Res.* **18,** 7243–7250.
142. Fried, M. G. & Crothers, D. M. (1984) *J. Mol. Biol.* **172,** 241–262.
143. Kolb, A., Spassky, A., Chapon, C., Blazy, B. & Buc, H. (1983) *Nucleic Acids Res.* **11,** 7833–7852.
144. Hahn, S., Buratowski, S., Sharp, P. A. & Guarente, L. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5718–5722.
145. Riggs, A. D., Bourgeois, S. & Cohn, M. (1970) *J. Mol. Biol.* **53,** 401–417.