Recognition of Flanking DNA Sequences by EcoRV Endonuclease Involves Alternative Patterns of Water-mediated Contacts*

Nancy C. Horton and John J. Perona‡

From the Department of Chemistry and Interdepartmental Program in Biochemistry and Molecular Biology, University of California, Santa Barbara, California 93106-9510

The 2.1-Å cocrystal structure of EcoRV endonuclease bound to 5′-CGGGATATCCCC, in a crystal lattice isomorphous with the cocrystallized undeckamer 5′-AAAGATATCTT previously determined, shows novel base recognition in the major groove of the DNA flanking the GATATC target site. Lys104 of the enzyme interacts through water molecules with the exocyclic N-4 amino groups of flanking cytosines. Steric exclusion of water molecule-binding sites by the 5-methyl group of thymine drives the adoption of alternative water-mediated contacts with AT versus GC flanks. This structure provides a rare example of structural adaptability in the recognition of different DNA sequences by a protein and suggests preferred strategies for the expansion of target site specificity by EcoRV.

Restriction endonucleases function in all prokaryotes as components of defensive restriction-modification systems and are superb models for the study of protein-DNA interactions owing to their exceptionally high sequence specificities. The type II restriction-modification systems are the best studied from a structural, functional, and mechanistic perspective and are composed of a homodimeric endonuclease and monomeric methylase (1, 2). The crystal structures of the following six restriction enzymes have been determined: EcoRI (3), EcoRV (4–6), BamHI (7, 8), PvuII (9), Cfr10I (10), and FokI (11). With the exception of Cfr10I and EcoRI, the enzymes have been solved in both the absence and presence of DNA. The structures reveal extensive complementarity at the protein-DNA interfaces, which appears to explain the high specificities of up to 10^16-fold in cleavage rate constants for the specific sites (Ref. 12 and references therein). An additional, less appreciated, contribution to specificity may also arise from DNA-induced conformational changes in the enzymes.

EcoRV has recently emerged as the best studied of the restriction endonucleases from both structural and mechanistic standpoints. This enzyme cleaves the sequence 5′-GATATC-3′ at the center TA step in a blunt-ended fashion, generating 5′-phosphate groups (13). The cocrystal structure of the EcoRVDNA complex reveals a tight network of hydrogen bonding and electrostatic and van der Waals contacts at the protein-DNA interface over the entire hexameric DNA site (4). Specificity at the outer two base pairs of each half-site is determined by hydrogen bonding with discriminating base functional groups in the major groove. At the center step the DNA is bent sharply by 50° into the major groove, so that protein cannot penetrate to contact the hydrogen bonding moieties. Indirect readout is thus implicated in specificity at this position. This may originate in part from differences in the energetic cost of partially unstacking the center TA step relative to CG or GC steps (14).

Whereas the specificity of EcoRV in vivo is limited to the hexamer target site GATATC, the crystal structures show that the enzyme also contacts 2–3 base pairs of DNA to either side of this site (4). These contacts with flanking DNA in five cocrystal structures with 5′-AAAGATATCTT at 2.0–2.1-Å resolution (5, 6, 14), and with the decamer 5′-GGGGATATCCCC at 3.0 Å, are primarily with the sugar-phosphate backbone (4, 5). Several studies have shown that the enzyme is sensitive to perturbation of these contacts. Replacement of phosphate groups with phosphorothioates showed that δp and δs substitutions directly 3′ to GATATC reduce V_max/K_m toward a decamer substrate by 4- and 50-fold, respectively (15). Furthermore, mutation of four amino acids contacting flanking phosphate groups reduces k_cat/K_m toward a 20-mer substrate containing GATATC (16). The most striking effect occurs with the mutant R226A within the C-terminal subdomain, which lowers activity by nearly 10^5-fold relative to the wild-type enzyme.

These studies show that the distal enzyme-DNA contacts outside the target site significantly increase the catalytic rate, perhaps by helping to orchestrate the mutual conformational changes in the enzyme and DNA which occur en route to the transition state. The biological role of EcoRV (and other type II enzymes) requires that all flanking sequences permit target site cleavage with high catalytic efficiency. However, this does not preclude the existence of some sequence preference. Indeed, 10- and 500-fold variations in binding constants with different flanks have been found in vitro studies of EcoRV and EcoRI, respectively (17, 18). Additionally, within the context of cleavage of the “star” site GTTAGTC, EcoRV prefers 5′-G and 3′-C on either side of the target (19, 20). These flanking sequence selectivities are modest compared with the 10^10-fold and greater specificities versus base substitutions internal to the GATATC site, but they nonetheless suggest that target site expansion may be feasible. This is a prospect of considerable practical importance for the development of new restriction enzymes specific for 8–10 base recognition sites.

Attempts to alter the substrate specificity of the EcoRI and EcoRV restriction endonucleases have not been successful thus far (21–24), indicating a need for further basic studies. Whereas EcoRV is the best studied enzyme of the class, its structure bound to DNA with GC flanks is determined at only 3.0 Å resolution (4). This is insufficient for describing detailed

* This work was supported by National Institutes of Health Grant GM53763, American Chemical Society-Petroleum Research Fund Grant 30427-G4 (to J. J. P.), and American Cancer Society Predoctoral Fellowship PP-88-015-01-GMC (to N. C. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed. Tel.: 805-893-7389; Fax: 805-893-4120.

‡ To whom correspondence should be addressed. Tel.: 805-893-7389; Fax: 805-893-4120.
protein-DNA interactions, particularly those involving solvent molecules. Moreover, large differences in DNA and protein conformation, most probably arising from altered crystal packing contacts, have been noted in the structures of the enzyme bound to specific DNA sites possessing AT versus GC flanks (4–6, 14). Therefore, to characterize the structural aspects of the flanking interactions better, we have determined a new cocrystral structure of EcoRV bound to 5′-CGGGATATCCC at 2.1 Å resolution. We find that Lys104 makes novel water-mediated interactions with hydrogen bonding functional groups of the flanking GC bases. By contrast, no interactions of Lys104 with bases of the flanking DNA are present in the five 2.0–2.1 Å resolution complexes with 5′-AAAGATATCCTT (5, 6, 14). These data provide insight into new modes of flanking sequence recognition by EcoRV and suggest preferred alternatives for engineering sequence specificity for larger DNA sites into the enzyme.

**MATERIALS AND METHODS**

**Purification and Cocrystallization of EcoRV and DNA**—Wild-type EcoRV was prepared as described and stored as an ammonium sulfate pellet at 4°C (6). The self-complementary oligonucleotide 5′-CGGGATATCCC was synthesized for cocrystallization by standard methods and purified on a Rainin PureDNA high pressure liquid chromatography column developed in a gradient of triethylammonium acetate/acetonic acid. Detritiation was performed on the column (25). The DNA was lyophilized and stored at −20°C until ready for use, when it was brought to a concentration of 10 mg/ml in 50 mM Tris (pH 7.5), 1 mM EDTA. Cocryocrystals of EcoRV complexed with 5′-CGGGATATCCC were grown by vapor diffusion from solutions containing 0.17 mM EcoRV and 0.34 mM DNA, in the presence of 15% PEG 4K, 100 mM imidazole (pH 6.5), and 150 mM NaCl (final conditions). The crystals are in the well described P1 lattice (5, 6) with cell dimensions isomorphous to the previous complexes (Table I). Nearly all of the molecular packing interactions in this crystal are identical with those of the cocryocrystal with 5′-AAAGATATCCTT (dAT). Only base stacking of the unpaired 5′ nucleotide (C versus A) on the dimerization domain of an adjoining molecule is very slightly altered because of the differing sizes of the pyrimidine and purine rings. The crystals have an enzyme dimer and duplex DNA in the asymmetric unit, so that this packing interaction is not made on the opposite end of the molecule. Conserved features observed on each of the two flanks occur in the context of different nearby lattice contacts and are consequently likely to reflect true aspects of the flanking interactions in solution. Thus, comparison of the cocryocrystal structures of EcoRV complexed with dAT versus dGC provides an excellent opportunity to elucidate detailed differences in the interactions of EcoRV with AT versus GC flanks.

**RESULTS**

EcoRV was cocrystallized with the undecamer duplex 5′-CGGGATATCCC possessing a 5′-C overhang (dGC) in the absence of divalent metals. The crystals are in the well described P1 lattice (5, 6) with cell dimensions isomorphous to the previous complexes (Table I).Nearly all of the molecular packing interactions in this crystal are identical with those of the cocryocrystal with 5′-AAAGATATCCTT (dAT). Only base stacking of the unpaired 5′ nucleotide (C versus A) on the dimerization domain of an adjoining molecule is very slightly altered because of the differing sizes of the pyrimidine and purine rings. The crystals have an enzyme dimer and duplex DNA in the asymmetric unit, so that this packing interaction is not made on the opposite end of the molecule. Conserved features observed on each of the two flanks occur in the context of different nearby lattice contacts and are consequently likely to reflect true aspects of the flanking interactions in solution. Thus, comparison of the cocryocrystal structures of EcoRV complexed with dAT versus dGC provides an excellent opportunity to elucidate detailed differences in the interactions of EcoRV with AT versus GC flanks.

The overall structure of this EcoRV-DNA complex is very similar to that of EcoRV cocrystallized with dAT (Fig. 1 (5)), with no significant differences in structure detectable within the GATATC target site. The quaternary structure is also identical with that of the dAT cocryocrystal lacking divalent metal ions. The relative orientations of the DNA binding/catalytic domains in the two structures differ by only a 0.4–0.6° rotation and a 0.22-Å difference in the center of mass separation of the two subunits. These differences are within the level of coordinate error, estimated from Luzzati plots at roughly 0.2–0.25 Å for each structure (data not shown). Small but perhaps significant intersubunit rotations of 1.0–1.5° are present in comparisons of either cocryocrystal lacking divalent metals, with isomorphous crystals of metal-containing ternary complexes (data not shown (5, 6)).

Nearly all of the previously observed major and minor groove enzyme-DNA contacts, both with base functional groups and sugar-phosphate moieties, are also present in this complex. The only significant exception is the conformation of the Arg211 side chain in both subunits. In the dAT structure the guanidinium group makes a direct electrostatic interaction with a DNA phosphate at GpATATC in subunit I and a water-mediated interaction with a DNA phosphate at C CGGGATATCCC, Wat, water.

**Table I**

Crystallographic data collection and refinement statistics

| DNA sequence | Resolution | Space group | a (Å) | b (Å) | c (Å) | α (°) | β (°) | γ (°) | Data coverage (%) | R<sub>merge</sub> | R<sub>free</sub> (%) | R<sub>merge</sub> (%) | R<sub>free</sub> (%) | R<sub>free</sub> (%) | R<sub>free</sub> (%) | r.m.s. bonds | r.m.s. ang | No. waters |
|--------------|------------|-------------|-------|-------|-------|-------|-------|-------|------------------|----------------|----------------|------------------|----------------|----------------|----------------|-------------|-------------|-------------|
| CGGGATATCCC  | 2.1        | P1          | 49.4  | 50.2  | 64.1  | 96°   | 109°  | 108°  | 85               | 0.053          | 27.4           | 0.196            | 0.294          | 0.012          | 1.86°          | 236         |

<sup>a</sup> Includes all data in the intensity range I/σ(I) > 3.0.

<sup>b</sup> Overall B factor is determined from a Wilson plot of the structure factor data using a low resolution cut-off of 4.7 Å.

<sup>c</sup> Refinement was carried out using a low resolution cut-off of 6.0 Å.
The location of the polypeptide segment Arg 221-Arg226 and the position of Q loops (R and R') and Q loops (Q and Q') bind in the major and minor grooves of the DNA, respectively. The location of the polypeptide segment Arg221, Arg226 and the position of Lys119-Asn120 are shown in each subunit. These amino acids interact with the flanking DNA.

The newly introduced exocyclic amino groups of the guanosine flanking bases are outside of the region contacted by the minor groove binding Q loops (Fig. 1) and surface loops at residues Lys119, Asn120 and are consequently without effect on the structure. However, differences between the dAT and dGC cocrystal structures appear in the major groove interactions of the flanking base pairs with the enzyme. On both sides of the GATATC target the exocyclic N-4 amino groups of flanking cytosines interact through water molecules to Lys104 on an enzyme surface loop (Fig. 3, A and B). Water molecules bridging Lys104 to the flanking bases are not present in any of the dAT structures previously determined (5, 6) and could not be visualized in the cocrystal with the decamer 5'-GGGATATCCC (4) owing to the lower resolution (3 Å) of this structure. In subunit I Lys104 bridges through a network of water molecules binding in the major groove at the two flanking CG pairs (Fig. 3A and Table II). The water molecules hydrogen-bond with protein at three positions as follows: the Lys104 side chain amine, the backbone at amino acids Ala181-Gly182, and the backbone at Phe105. Water molecules 3 and 5 also bridge through water molecules to the DNA phosphates at Cyt1 and Cyt9 (Table II). In this subunit Lys104 interacts through one water molecule to the N-4 of Cyt11 and through two water molecules to the O-6 of its base-pairing partner at Gua2, whereas its connection to the inner flanking pair Gua2-Cyt10 is through a chain of three water molecules: Wat1,3,4. Recognition of this inner pair occurs via water molecules more closely connected to the Ala181-Gly182 main chain. The Gly182 amide also serves as a hydrogen bond acceptor from the N-4 of Cyt9, providing part of the discrimination for the outer GC base pairs of GATATC (Fig. 3, A and B). Similar water-mediated interactions of Ala181 and Gly182 with the inner Ade3-Thy10 pair are observed in the dAT structures (4–6, 14). However, in these structures no direct or water-mediated recognition of base functional groups on the outer flanking Ade2-Thy11 pair occurs.

In subunit II the Lys104-amine group interacts directly with two water molecules, one of which (Wat12) bridges directly to the N-4 of the inner flanking cytosine (Fig. 3B). Wat12 also interacts with another water (Wat13) which donates a hydrogen bond to the main chain amide of Gly182. While preserving the common feature of water-mediated flanking sequence recognition through Lys104, the interactions made in subunits I and II thus clearly differ. This appears attributable to the proximity of a crystal packing contact made by the 5-overhanging cytosine (Cyt1) in subunit I. The base of Cyt1 packs onto the peptide main chain in the dimerization domain of an adjoining molecule, which provides stabilization in a
TABLE II
Interactions of water molecules mediating flanking sequence recognition

| Subunit I | Subunit II |
|-----------|-----------|
| Atom      | Distance  | Atom      | Distance  |
| Wat¹      |          | Wat¹      |          |
| Lys¹⁰⁴,Ne | 2.8      | Lys¹⁰⁴,Ne | 2.5      |
| Cyt¹¹:N⁴ | 3.2      | Wat⁴      | 2.6      |
| Wat⁴      | 2.6      | Wat²      | 2.6      |
| Wat²      | 2.6      | Wat²      | 2.6      |
| Wat⁵      | 2.9      | Wat¹      | 2.6      |
| Cyt¹¹:N⁴ | 2.7      | Wat²      | 2.7      |
| Wat¹      | 2.6      | Wat²      | 2.5      |
| Wat²      | 2.6      | Wat¹      | 2.5      |
| Wat²      | 2.6      | Wat¹      | 2.5      |
| Wat³      | 2.6      | Wat¹      | 2.5      |
| Wat¹      | 2.6      | Wat²      | 2.5      |
| Wat²      | 3.2      | Wat³      | 3.2      |
| Cyt¹¹:N⁴ | 3.2      | Wat³      | 3.2      |
| Wat³      | 3.2      | Wat³      | 3.2      |
| Wat³      | 3.2      | Wat³      | 3.2      |

a This water bridges through one other water molecule to the side chain of His¹⁰⁵.
b This water bridges through other water molecules to the main chain amides of Lys¹⁰⁴ and Phe¹⁰⁵ and to the DNA phosphate at C-9.
c This water bridges through one other water molecule to the DNA phosphate at C-9.

in a longer duplex DNA. In subunit II, which lacks this lattice intermolecular hydrogen bonding, the 5'-overhanging cytosine is disordered. Moreover, interactions between Lys¹⁰⁴ and the outer flanking base pairs are altered by only 1 Å (34), and this shift should be readily accommodated by small rearrangements of the waters.

To address why the apparently nonspecific water-mediated interactions between Lys¹⁰⁴ and the outer flanking base pairs are not present in dAT, we superimposed the dAT and dGC structures based on polypeptide backbone atoms within the DNA-binding domains (root mean square deviation = 0.24 Å for the superposition of 244 amino acids (6)). This shows that in both subunits one of the water molecules bound to Lys¹⁰⁴ is blocked from binding to flanking AT pairs by one of the thymine C-5 methyl groups (Fig. 5, A and B). In subunit II the steric hindrance by Thy¹⁰⁴, occluding binding of Wat¹⁰⁴, is sufficient to account for why no water-mediated contacts with Lys¹⁰⁴ are observed in dAT. This is because Wat¹⁰⁴ itself interacts directly with Cyt¹¹⁴ (Fig. 5B). However, in subunit I it is more difficult to explain why the steric exclusion of Wat¹⁰⁴ should also disrupt the Wat¹⁰⁴ interactions (Fig. 5A). Whereas all the waters and groups of the flanking DNA and nearby protein are clearly visible in OMIT electron density reorient their two hydrogen bond donor and acceptor groups to provide equivalent interactions with flanking TA pairs. Moreover, it is also very unlikely that discrimination could be achieved against flanking GC or AT base pairs in which the purine and pyrimidine rings are exchanged. This is because the positions of the major groove hydrogen-bonding sites in these pairs are altered by only 1 Å (34), and this shift should be readily accommodated by small rearrangements of the waters.

manner apparently similar to the continuation of base-stacking in a longer duplex DNA. In subunit II (see Fig. 1 and Ref. 6). The DNA flank interacting with subunit I is in red and that interacting with subunit II is in gray. The shift in position of C-11 of the outer flanking pair in subunit II is evident. Interactions of water molecules 1–7 of subunit I and 1–3' of subunit II are delineated in Table II, and their hydrogen-bonding contacts are shown in Fig. 3, A and B. Protein atoms of Lys¹⁰⁴ and Ala¹⁵¹-Gly¹⁰⁴ are shown in green for subunit I and in yellow for subunit II.
subunit II, as noted above). By contrast, considerably lower B-factors, suggestive of reduced mobilities, are associated with the target site DNA and many of its interacting protein segments. Thus, removal of any single interaction in the flanks might be sufficient to disrupt a number of adjoining contacts as well.

**DISCUSSION**

*Influence of Crystal Packing Interactions on the Structure of the GATATC Target Site—*Comparison of the high resolution, isomorphous cocrystal structures of EcoRV complexed to dAT and dGC shows that no significant differences in DNA conformation are detectable within the GATATC target site. This finding differs from that of Winkler and colleagues (4, 5) who reported significant differences in DNA conformation between cocrystals with AT versus GC flanks. These differences appear particularly significant in the sugar-phosphate backbone at the scissile center-base step, where the 3.0-Å cocrystal structure with GC flanks shows that the DNA at this position possesses a nearly A-like conformation (4). However, this structure was determined in an orthorhombic crystal lattice, with distinctly altered intermolecular packing contacts involving both the enzyme and the ends of the DNA. Therefore, it was not possible to deduce whether the different conformations observed inside the GATATC site were a consequence of the changed lattice contacts or arose as a propagated effect from the different flanks. Because this dGC cocrystal structure is in a lattice isomorphous to the dAT structure, it is now possible to state definitively that flanking DNA sequence has no effect on DNA structure inside the target site.

*Strategies for the Recognition of Alternative DNA Sequences by Proteins—*The dGC structure reported here illustrates three new structural features associated with flanking sequence interactions by EcoRV. First, it reveals a role for a new segment of the enzyme at Lys^{104}-Phe^{105} in the recognition process. Second, it shows that the water-mediated recognition of base functional groups in the flanks is not limited to the inner flanking pair, as suggested by the dAT structures, but includes as well the adjoining base pair farther from the target site. Third, it provides a rare detailed example of how a protein is able to use alternative intermolecular contacts to recognize different DNA sequences. In this case, an important driving force for the new interactions, and the consequent recruitment of Lys^{104} in recognition, arises from steric occlusion of water-binding sites by thymine C-5 methyl groups.

Other mechanisms may also be operative in determining the positions of the waters in different flanking sequence contexts. For example, in the dAT structures, a water molecule bridges the Ala^{181} main chain carbonyl group to the N-6 of Ade^{3} of the inner flanking pair (5, 6). Comparative analysis of the dAT and dGC structures shows that these two groups are 6.0 Å apart in dAT and 7.0 Å apart in dGC, a difference outside estimated errors given that the coordinates of each individual structure are determined within 0.2–0.25 Å. The greater separation of hydrogen-bonding groups accounts for the absence of this water in the dGC structure (there is no steric overlap with other water molecule-binding sites). The altered relative positions of protein and DNA groups may originate in intrinsic sequence-dependent variations in B-DNA structure (33), which can play a key role in favoring particular conformations adopted by DNA when bound to proteins. High resolution analyses of DNase I bound to different DNA sequences have provided a detailed description of the operation of this mechanism (37, 38).

Another example of the recognition of alternative sequences by a protein is that of the estrogen receptor, which is able to adapt to different nonconsensus target sites via the rearrangement of a lysine side chain at the protein-DNA interface (35).
this case, the lysine adopts a new orientation in response to an unfavorable juxtaposition with the N-6 of an adenine residue substituted for a guanine in the consensus site. In its new orientation the lysine also makes new intramolecular contacts within the protein. In contrast, conformations of Lys104 in the dAT versus dGC structures of EcoRV differ less dramatically (Fig. 5, A and B). In subunit II of dGC the most important difference is in the dihedral angle $\gamma$–$\delta$–$\epsilon$–$\zeta$, so that $\zeta$ points toward the DNA (Fig. 5B). In subunit I there are instead significant differences in many of the Lys$^{104}$ side chain dihedrals in dGC (Fig. 5A), the effect of which is to improve the linearity of the Ne$^2$–H–O hydrogen-bonding contact relative to that which would obtain in dAT.

The structure of EcoRV endonuclease bound to dGC thus reveals additional strategies for the recognition of alternative DNA sequences, in addition to the rearrangement of amino acids at the interface as seen in the estrogen receptor-DNA complexes. These exploit the potential for water-mediated recognition, showing how alternative arrangements of water molecules can form in the areas of protein-binding sites where the
DNA sequence to be recognized is not unique. These new patterns can be driven by steric occlusion or by the intrinsic sequence-dependent variations in DNA structure. It is expected that further examples of alternative water-mediated contacts, as a structural underpinning for broad specificity, will be found in other protein-DNA complexes. High resolution studies of altered complexes in the engrafted homeodomain (39) and glucocorticoid receptor (40) systems, in which the mutated protein has an altered binding specificity, also provide insight into this question.

Implications for Specificity—Flanking interactions are important in wild-type EcoRV as they contribute to the overall binding energy, as well as to $k_{cat}/K_m$ (16). The effects of differing flanking sequences on the association constant $K_a$ has been measured to be $\sim$10-fold for EcoRV (17) and 500-fold for EcoRI (18). Whereas no systematic study is available on the effect of flanking sequences on catalytic rate by EcoRV, the cleavage rates for the EcoRI substrates were found to be unaffected by the sequence of the flanking regions (18). Therefore, at least in the case of EcoRI, the energetic effect of flanking sequences on the transition state is equivalent to that on the ground state (18), and effects that stabilize the ground state are predicted to stabilize the transition state by the same amount, thus increasing $k_{cat}/K_m$. Such a linear free energy relationship has been observed using base analogues in the cognate sequence of EcoRV.2 This also indicates that the structures of the ground and transition states are very similar (18), validating structural studies on the ground state.

Strategies for Target Site Expansion of EcoRV—Even though the interaction of EcoRV with flanking DNA bases is nonspecific, the existence of these contacts with potentially discriminating groups in the major groove suggests that expanding the target site selectivity via protein engineering approaches may be feasible. In the wild-type complexes, lack of specificity for flanking DNA is most evidently due to the absence of direct, discriminatory protein-DNA hydrogen bonds. However, a further rationale is that the water-mediated contacts with the bases, and the interactions with the sugar-phosphate backbone, are less well ordered than those internal to GATATC. This is reflected in the relatively high crystallographic B-factors (ranging from 25 to 45 Å$^2$) of the two flanking DNA base pairs, the enzyme surface loops in which $\text{Lys}^{104}$ and amino acids 221–226 reside, and the intervening waters. Therefore, an important starting point toward introducing expanded specificity is to decrease the mobility of these groups. It seems likely that one way to accomplish this objective is to build in direct protein interactions with the flanks.

The dGC crystal structure draws attention to the possibility that direct interactions with the flanking DNA bases might be introduced via mutation of $\text{Lys}^{104}$. Introduction of these direct interactions has potential both to improve the stability of the flanking regions and to eliminate the ambiguities sometimes associated with obtaining specificity through water-mediated hydrogen bonding. Modeling from the dGC structure shows that substitution with Arg$^{\text{a}}$ permits, without steric clashes, direct contact in the major groove with the O-4 of a thymine base directly 3' to the target site (Fig. 6B). The closest approach of the Arg guanidinium to the O-6 of a guanine base modeled at the position of the 3'-cytosine is 3.7 Å, so that direct contact can be envisioned with the requirement for only small structural rearrangements (Fig. 6A).

The design of new distal contacts to provide discrimination for flanking bases at the catalytic step may be aided by the fact that Gly$^{182}$ makes interactions both internal and external to GATATC (Fig. 3). New flanking interactions, if made by a segment of protein already directly interacting with the hexamer target, might more readily participate in the cooperative conformational changes that occur en route to the transition state. In fact, modeling shows that several substitutions at the adjacent Ala$^{181}$ can position hydrogen-bonding amino acids in the major groove, at the flanking base pair directly adjacent to the target site (37). For example, the A181E substitution allows direct contact with the N-4 of a 5'-cytosine (Fig. 6A). Alternatively, Lys$^{181}$ can be modeled to donate hydrogen bonds to both the N-7 and O-6 atoms of a 5'-guanine base. In the context of an adjacent TA base pair (Fig. 6B), the interaction of Lys$^{181}$ with the N-7 of the 5'-adenine could provide some additional stability, with specificity for the 3'-thymine arising from Arg$^{104}$. Thus either of the A181E or A181K mutations, perhaps together with the introduction of K104R, should provide a reasonable starting point for the rational extension of site specificity in EcoRV.2 The designs might also be aided by consideration of the properties of enzyme mutants that interact with the target site and that introduce deficiencies at the catalytic step. Introduction of new specific flanking interactions, in the context of a mutant background, might provide additional binding energy useful in reconstituting the required conformational changes upon DNA binding. Unlike the present circumstance with wild-type EcoRV, however, the sequence-specific nature of these contacts would create the potential for specific rather than nonspecific rate enhancement.

Acknowledgments—We thank Norbert Reich for helpful suggestions and Mark Rould for providing the local scaling program MAXSCALE. Fig. 3, A and B, was produced with the program MAXIMAGE written by M. Rould.

REFERENCES
1. Roberts, R. J. & Halford, S. E. (1993) Nucleases (Linn, S. M., Lloyd, R. S., and Roberts, R. J., eds) 2nd Ed., pp. 35–88, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
2. Pingoud, A. & Jeltsch, A. (1997) Eur. J. Biochem. 246, 1–22
3. Rosenberg, J. (1991)Curr. Opin. Struct. Biol. 1, 104–113
4. Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R., Martin, P. D., Petrats, K. & Wilson, K. S. (1993) EMBO J. 12, 1731–1735
5. Kostrewa, D. & Winkler, F. K. (1995) Biochemistry 34, 683–696
6. Perona, J. J. & Martin, A. M. (1997) J. Mol. Biol. 273, 207–225
7. Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1994) Structure 2, 439–452
8. Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1995) Science 269, 656–663
9. Cheng, X., Balandrian, K., Schildkraut, I. & Anderson, J. E. (1994) EMBO J. 13, 3927–3935
10. Bozic, D., Grauzlis, S., Siksyns, V. & Huber, R. (1996) J. Mol. Biol. 255, 176–186
11. Wah, D. A., Hirsch, J. A., Dorner, L. F., Schildkraut, I. & Aggarwal, A. (1997) Nature 388, 97–100
12. Jen-Jacobson, L. (1995) Methods Enzymol. 259, 305–344
13. Schildkraut, I., Banner, C. D. B., Rhodes, C. S. & Parekh, S. (1984) Gene (Amst.) 1, 327–329
14. Horton, N. C. & Perona, J. J. (1998) J. Mol. Biol. 277, 779–787
15. Thorogood, H., Grasby, J. A. & Connolly, B. A. (1996) J. Mol. Biol. 259, 885–8862
16. Wenz, C., Jeltsch, A. & Pingoud, A. (1996) J. Biol. Chem. 271, 5565–5573
17. Engel, L. K., Wysz, K. K. & Jen-Jacobson, L. (1997) J. Mol. Biol. 269, 82–101
18. Jen-Jacobson, L. (1997) Biopolymers 44, 153–180
19. Taylor, J. D. & Halford, S. E. (1989) Biochemistry 28, 6198–6207
20. Taylor, J. D. & Halford, S. E. (1992) Biochemistry 31, 90–97
21. Heinman, J. & Model, P. (1996) EMBO J. 9, 3369–3378
22. Flores, H., Osuna, J., Heinman, J. & Soberon, X. (1995) Gene (Amst.) 157, 295–301
23. Wenz, C., Selent, U., Wende, W., Jeltsch, A., Wolfe, H. & Pingoud, A. (1994) Biochim. Biophys. Acta 1219, 73–80
24. Lania, T., Selent, U., Wenz, C., Wende, W., Schulz, A., Adiraj, M., Katt, S. B. & Pingoud, A. (1996) Protein Eng. 9, 1005–1010
25. Aggarwal, A. K. (1990) Methods: A Companion to Methods in Enzymology, 1, 25–90
26. Otwinowski, Z. (1993) in Proceedings of the CCP 4 Study Weekend: Data Collection and Processing, January, 29–30, 1993, Sawrey, L., Isaacs, N.,

2 A. M. Martin, M. D. Sam, N. O. Reich, and J. J. Perona, manuscript in preparation.
3 Mutations at Ala$^{181}$ with the aim of target site expansion have already been constructed (Dr. A. Pingoud, personal communication).
EcoRV Flanking Sequence Recognition

27. Brunger, A. T., Kuriyan, J. & Karplus, M. (1987) *Science* **235**, 458–460
28. Sack, J. S. (1988) *J. Mol. Graphics* **6**, 224–225
29. Finzel, B. C. (1995) *Acta Crystallogr. D* **51**, 450–457
30. Parkinson, G., Vojtechovsky, J., Clowney, L., Brunger, A. T., & Berman, H. M. (1996) *Acta Crystallogr. D* **52**, 57–64
31. Dayringer, H., Tramontano, A., Sprang, S. & Fletterick, R. J. (1986) *J. Mol. Graphics* **4**, 82
32. Perry, K. M., Fauman, E. B., Finer-Moore, J. S., Montfort, W. R., Maley, G. F., Maley, F. & Stroud, R. M. (1990) *Proteins* **8**, 315–333
33. Steitz, T. A. (1990) *Q. Rev. Biophys.* **23**, 205–280
34. Seeman, N. C., Rosenberg, J. M. & Rich, A. (1976) *Proc. Natl. Acad. Sci. U. S. A.* **73**, 804–808
35. Schwabe, J. W. R., Chapman, L. & Rhodes, D. (1995) *Structure* **3**, 201–213
36. Deleted in proof
37. Suck, D., Lahm, A. & Oefner, C. (1988) *Nature* **332**, 464–468
38. Weston, S. A., Lahm, A. & Suck, D. (1992) *J. Mol. Biol.* **226**, 1237–1256
39. Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T. & Pabo, C. O. (1997) *Structure* **5**, 1047–1054
40. Gewirth, D. T. & Sigler, P. B. (1995) *Nat. Struct. Biol.* **2**, 386–394