4. M. Ptashne, *A Genetic Switch: Phage λ and Higher Organisms* (Cell Press and Blackwell Scientific Publications, Cambridge, MA, ed. 2, 1992).
5. D. L. Friedman, *Curr. Opin. Genet. Dev.* **2**, 727 (1992).
6. H. R. Horvitz and I. Herskowitz, *Cell* **68**, 237 (1992); S. K. Kim and D. Kaiser, *Annu. Rev. Microbiol.* **46**, 117 (1992).
7. Y. Brun, G. Marczynski, L. Shapiro, *Annu. Rev. Biochem.* **63**, 419 (1994).
8. M. J. McFall-Ngai and E. G. Ruby, *Science* **254**, 1491 (1991); S. Long and B. J. Staskawicz, *Cell* **73**, 921 (1993).
9. F. Jacob and J. Monod, *Cold Spring Harbor Symp. Quant. Biol.* **26**, 389 (1961).
10. R. Thomas, *J. Theor. Biol.* **153**, 1 (1991); M. Kaufman, J. Urbain, R. Thomas, *ibid.* **114**, 527 (1985); L. Glass and S. A. Kauffman, *ibid.* **39**, 103 (1973).
11. D. Bray, R. B. Bourret, M. I. Simon, *Mol. Biol. Cell* **4**, 469 (1993); D. C. Hauri and J. Ross, *Biophys. J.* **68**, 708 (1995).
12. A. Arkin and J. Ross, *Biophys. J.* **67**, 560 (1994).
13. The in vivo transcription rate in *E. coli* has been directly observed at 42 nucleotides (nt) per second [S. L. Gotta, O. L. Miller Jr., S. L. French, *J. Bacteriol.* **173**, 6647 (1991)]. The overall in vivo rate is 12 to 19 nt/s [A. Kornberg and T. A. Baker, *DNA Replication* (W. H. Freeman, New York, 1992), p. 246]. Higher rates, in the range of 50 to 60 nt/s, are also observed, particularly in vitro. See also U. Vogel and K. F. Jensen, *J. Bacteriol.* **176**, 2807 (1994).
14. Steady-state concentrations of all cellular proteins are determined by the dynamic balance between production and losses. Assumptions for the equations in Fig. 2A include (i) transcription initiation is the rate-limiting step; (ii) the repressing protein-promoter interaction has first-order kinetics (higher order kinetics enter through the repression term, $Y(t)$, in Eq. 5); (iii) transcription and translation are coupled so that the number of proteins produced per transcript is a constant (A_{GENE}) independent of transcription rate but differing from gene to gene (*16*) [O. Yarchuk, N. Jacques, J. Guillerez, M. Dreyfus, *J. Mol. Biol.* **226**, 581 (1992)]; and (iv) for low (<10?) MOI, protein production is proportional to MOI (RNAP, ribosome, or substrate availability is limiting at high MOI) [K. Luk and K. Mark, *J. Virol.* **64**, 183 (1983); G. P. Zambetti and R. C. Shuster, *Mol. Gen. Genet.* **193**, 322 (1984)]. In vivo experiments at an average MOI result in a Poisson distribution of MOIs within the experimental cell population, with the exception of induced monolysogens for which the initial MOI is 1. We neglect (i) additional concentration reduction attributable to dilution in growing cells and (ii) additional production from newly replicated viral DNA.
15. A. Levine, A. Bailone, R. Devoret, *J. Mol. Biol.* **131**, 655 (1979).
16. M. A. Shea and G. K. Akers, *ibid.* **181**, 211 (1985).
17. The circuit in Fig. 4 includes the functions essential to the path choice decision. Functions such as the OOP antisense RNA whose role in λ regulation appears to be nonessential for the path choice are not included (*1*) [L. Krinke and D. L. Wulff, *Genes Dev.* **1**, 1005 (1987); *ibid.* **4**, 2223 (1990)].
18. U. Schmeissner, D. Court, H. Shimatake, M. Rosenberg, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 3191 (1980).
19. A. Das, *J. Bacteriol.* **174**, 6711 (1992).
20. S. W. Mason, J. Li, J. Greenblatt, *J. Biol. Chem.* **267**, 19418 (1992); J. Greenblatt *et al.*, *Nature* **364**, 401 (1993).
21. J. W. Roberts, *Cell* **52**, 5 (1988).
22. B. C. Hoopes and W. R. McClure, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3134 (1985); Y. S. Ho and M. Rosenberg, *J. Biol. Chem.* **260**, 11838 (1985); R. W. Simons and N. Kleckner, *Annul Rev. Genet.* **22**, 567 (1988).
23. M. A. Hoyt, D. M. Knight, A. Das, H. I. Miller, H. Echols, *Cell* **31**, 565 (1982).
24. C. Herman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10861 (1993).
25. J. A. Noble *et al.*, *ibid.*, p. 10866.
26. P. Kourilsky, *Mol. Gen. Genet.* **122**, 183 (1973).
27. D. I. Wulff and M. Rosenberg, in (*3*), pp. 53–74.
28. J. W. Roberts and R. Devoret, in (*3*), pp. 123–144.
29. L. M. De Vargas and A. Landy, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 588 (1991).
30. A. Landy, *Annu. Rev. Biochem.* **58**, 913 (1989); *Curr. Opin. Genet. Dev.* **3**, 699 (1993).
31. J. M. Leong *et al.*, *J. Mol. Biol.* **189**, 603 (1986).
32. L. H. Hartwell and T. A. Weinart, *Science* **246**, 629 (1989); L. H. Hartwell, *Genetics* **129**, 975 (1991).
33. A. Campbell and D. Botstein, in (*3*), pp. 365–380.
34. S. Casjens, G. Hatfull, R. Hendrix, *Semin. Virol.* **3**, 383 (1992).
35. Measurements of the signal protein half-lives are available. The rate of signal protein transcript production can be estimated for the critical operons, but with diminishing certainty for downstream genes on P_R and P_L operons as a result of polarity effects. The kinetics of the P_R, P_L, P_{RM}, and P_{RE} promoters have been extensively studied. The average proteins/transcript parameter is not available.
36. For a kinetic model of the switch, see (*16*).
37. Descriptive of Cro repression of P_R, P_L, and P_{RM}. Descriptive of CI repression of P_R and P_L. CI interaction with P_{RM} is more complex (*4, 16*) [B. J. Meyer, R. Maurer, M. Ptashne, *J. Mol. Biol.* **139**, 163 (1980); B. J. Meyer and M. Ptashne, *ibid.*, p. 195].
38. We thank I. Hershowitz, D. Botstein, D. Kaiser, S. Kustu, and A. Campbell for valuable suggestions.

RESEARCH ARTICLE

Structure of Bam HI Endonuclease Bound to DNA: Partial Folding and Unfolding on DNA Binding

Matthew Newman,* Teresa Strzelecka,† Lydia F. Dorner, Ira Schildkraut, Aneel K. Aggarwal‡

The crystal structure of restriction endonuclease Bam HI complexed to DNA has been determined at 2.2 angstrom resolution. The DNA binds in the cleft and retains a B-DNA type of conformation. The enzyme, however, undergoes a series of conformational changes, including rotation of subunits and folding of disordered regions. The most striking conformational change is the unraveling of carboxyl-terminal α helices to form partially disordered "arms." The arm from one subunit fits into the minor groove while the arm from the symmetry related subunit follows the DNA sugar-phosphate backbone. Recognition of DNA base pairs occurs primarily in the major groove, with a few interactions occurring in the minor groove. Tightly bound water molecules play an equally important role as side chain and main chain atoms in the recognition of base pairs. The complex also provides new insights into the mechanism by which the enzyme catalyzes the hydrolysis of DNA phosphodiester groups.

Protein-DNA selectivity is a central event in many biological processes. Type II restriction endonucleases are attractive systems for studying selectivity because of their high specificity and great variety. Almost 2400 type II restriction enzymes representing 188 different specificities have now been discovered (*1*). They generally recognize DNA sequences that vary between four to eight base pairs (bp), and require only Mg^{2+} to catalyze the hydrolysis of DNA. Their sequence specificity is remarkable; for example, the activity can be a million times lower as the result of a single base pair change within the recognition sequence (*1*). This specificity is crucial for the prevention of accidental cleavage at the many nonspecific sites in a DNA sequence.

Bam HI endonuclease (from *Bacillus amyloliquefaciens* H) binds as a dimer to the symmetrical sequence 5′-GGATCC-3′. The enzyme cleaves DNA after the 5′-G on each strand to produce 4-bp (5′) staggered ends. The structure of Bam HI has been determined in the absence of DNA and consists of a central β sheet with α helices on both sides (*2, 3*). The structure shows striking resemblance to the endonuclease Eco RI model (*4, 5*) despite the lack of sequence similarity between the two enzymes. We have now determined the structure of Bam HI bound to a 12-bp DNA fragment containing its recognition sequence.

As we anticipated from the structure of the free enzyme, the DNA binds in the large cleft of the Bam HI dimer. The DNA retains a regular B-DNA–like conformation, and there are no major bends or kinks. The enzyme undergoes a series of conformational changes on DNA binding. The most striking of these is at the carboxyl-terminal end of the protein. This region is an ordered α helix (α7) in the free protein, but it

M. Newman, T. Strzelecka, and A. K. Aggarwal are in the Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA. L. F. Dorner and I. Schildkraut are at New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA.

*Present address: Department of Biochemistry and Molecular Biology, University of Leeds, Leeds LS2 9JT, UK.
†Present address: Laboratory of Chemical Physics National Institutes of Health, Bethesda, MD 20892, USA.
‡To whom correspondence should be addressed.

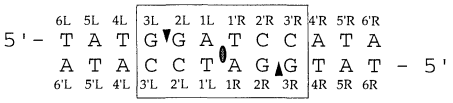

Fig. 1. Sequence of the palindromic 12-bp DNA fragment cocrystallized with Bam HI. The 6-bp recognition sequence is shown in a box. The cleavage sites are indicated by arrowheads. The sequence is divided into left (L) and right (R) half-sites and the nucleotides numbered 1 to 6 starting from the center.

assumes an extended, partially disordered conformation on DNA binding. We refer to the extended conformation as an arm. The arm from one subunit fits into the minor groove of the DNA, whereas the arm from the symmetry related subunit lies parallel to the sugar-phosphate backbone. This introduces an unavoidable asymmetry to the complex because the minor groove is too narrow to accommodate both arms. Although disordered regions are known to become ordered on DNA binding, as occurs in the NH_2-terminal arm of lambda (λ) repressor (*6*) or the recognition helices of bZip and helix-loop-helix proteins (*7*), we now show that an ordered α helix unfolds and becomes partially disordered on DNA binding.

The specific base pair contacts occur primarily in the major groove of the DNA, with a few interactions occurring in the minor groove. The major groove contacts are formed mainly by atoms at the NH_2-terminal end of a parallel four-helix bundle. Every hydrogen bond donor and acceptor group in the major groove is involved in direct or water-mediated bonds with the protein. Although Bam HI shares structural similarity with Eco RI, none of the interactions could have been anticipated from the Eco RI–DNA complex (*4*, *5*).

Structure determination. The crystallization of Bam HI–DNA complex has been described (*8*). Reproducible cocrystals could only be obtained with a 12-bp DNA fragment containing 5′-TAT ends (Fig. 1). The crystallographic asymmetric unit contains one Bam HI–DNA complex in which the two halves of the complex are related by a noncrystallographic twofold symmetry axis. The structure of the complex was solved by combining molecular replacement (MR) phases with heavy atom phases derived from an iodinated derivative (Table 1). No symmetry averaging was used in the structure determination and therefore the crystal structure presents two independent half-complexes. This enabled us to cross-validate most of the observations drawn from the structure.

The iodinated derivative was prepared by substituting iodouracils for four thymine residues (4L, 4R, 5′L, and 5′R) on the 12-bp DNA fragment. The isomorphous difference Patterson map showed clear positions for all four iodine atoms. When additional iodinated derivatives were being prepared, we solved the structure of Bam HI by itself (*2*, *3*). This enabled us to construct a search model consisting of Bam HI dimer for MR. The rotation function was solved (3.5 to 12 Å) with program MERLOT (*9*). Two peaks corresponding to the noncrystallographic twofold symmetry of the complex were observed in the rotation function map. One of these peaks was refined by the Patterson correlation refinement option of program X-PLOR (*10*). Translation function was also performed with X-PLOR (4.0 to 10 Å), which produced a top peak of 16.4 σ. The solution corresponding to the top peak was improved by rigid body refinement of the Bam HI dimer followed by individual subunits, to yield an *R* factor of 46.7 percent. The solution conformed closely to the packing model derived earlier (*8*). The validity of the MR method was also established by calculating a difference Fourier map from MR phases and iodine isomorphous differences (3.5 to 8 Å). The highest peaks in the map coincided with the known positions of iodines.

An electron density map calculated with phases from the rigid-body refined model of Bam HI dimer was discontinuous. The map was improved by refining Bam HI dimer with the use of simulated annealing (X-PLOR) to 2.5 Å resolution and combining model phases with the iodine phases. The resulting map was sufficiently continuous to permit the building of DNA (which had been omitted) and extensive rebuilding of the protein. However, no density could be seen for the COOH-terminal α helices of the Bam HI dimer. Using the Photon Factory data to 2.2 Å resolution, we performed several rounds of simulated annealing and positional refinement on the DNA and the rebuilt Bam HI dimer but residues 194 to 213 were omitted from COOH-terminal end of each monomer. The resulting map showed densities for portions of the missing COOH-terminal polypeptides, but at locations that were radically different from that in the free enzyme (Fig. 2). Moreover, the densities were asymmetric between the two subunits. For the left subunit, residues 194 to 208 were built leading back toward the core of the protein, and for the right subunit, residues 194 to 198 were built leading into the minor groove (Fig. 2). (Although there is some indication that the COOH-terminal end of right subunit can adopt a similar conformation to that of the left subunit for a small percentage of the time, the primary conformation is in the minor groove.)

As soon as more of the structure was correctly placed and identified (3208 protein atoms and 465 DNA atoms), positional

Table 1. X-ray data collection and phasing statistics for Bam HI–DNA complex. The cocrystals belong to space group $P2_12_12_1$, with unit cell parameters a = 108.8 Å, b = 81.9 Å, c = 68.8 Å, and one complex in the crystallographic asymmetric unit (*8*). Data from the native 1 and the iodouracil crystals were collected at room temperature, with a Xuong-Hamlin area detector (*33*) operating on a Rigaku RU200 rotating anode x-ray generator. Data from the native 2 crystal (8°C) were collected at beamline 6A2 of the Photon Factory (KEK, Japan), with the use of Weissenberg geometry (*34*) and an oscillation range of 4.5° (cassette radius of 286.5 mm). Fuji imaging plates were used to record the diffraction patterns and these were digitized by means of a Fujix BA100 scanner. The data were processed with WEIS (*35*). Scaling, merging, and refinement of heavy atom parameters were done with CCP4 (*36*).

Item	Native 1	Native 2	Iodouracil
Source	Xuong-Hamlin	Photon Factory	Xuong-Hamlin
Wavelength (Å)	1.542	0.970	1.542
Resolution (Å)	2.5	2.2	3.5
Measured reflections	43,264	103,881	30,505
Unique reflections	19,359	30,078	6,529
Data coverage (%)	88.4	94.2	79.7
Signal [%>3σ(I)]	64.6	73.9	87.9
R_{merge}* (%)	6.8	11.8	8.5
Isomorphous phasing statistics:			
R_{iso}† (%)			16.2
R_{cullis}‡ (%)			67.0
Phasing power§			1.2/1.0
Figure of merit‖			0.31/0.62
Refinement statistics (10 to 2.2 Å):			
Reflections [$F > 2\sigma(F)$]		28,405	
R factor (%)		18.9	
rms bond lengths (Å)¶		0.018	
rms bond angles (°)¶		2.8	

*Merging *R* factor ($R_{merge} = \Sigma|I_{obs} - \langle I\rangle|/\Sigma\langle I\rangle$) calculated for all data. †Mean isomorphous difference ($R_{iso} = \Sigma|F_p - F_{ph}|/\Sigma F_p$, where F_p and F_{ph} are the structure factor amplitudes of the native and derivative data, respectively). ‡R_{cullis} is the mean isomorphous lack of closure error divided by the mean isomorphous difference calculated over all the centric data to 3.5 Å resolution. §Phasing power is defined as $\langle|F_h|\rangle$/rms(ε) (where $\langle|F_h|\rangle$ is the mean calculated amplitude for the heavy atom model and rms(ε) is the weighted root-mean-square (rms) lack of closure error for the isomorphous differences). Values given are for the acentric and centric terms, respectively. ‖Figure of merit for the acentric and centric terms, respectively. ¶The rms deviation from ideal values.

and temperature factor refinements were performed to 2.2 Å resolution (X-PLOR). Water molecules (215 total) were subsequently identified and included in the refinement. The geometry of DNA was improved when we used the program TNT (*11*) for the final cycles of positional refinement. X-PLOR was then used to obtain a final set of temperature factors.

Overall structure. The Bam HI dimer binds to DNA from the major groove side. This is similar to Eco RI (*5*) but opposite to Eco RV (*12*) and Pvu II (*13*), which bind from the minor groove side. The DNA fits into the large cleft of the Bam HI dimer and appears surrounded by the enzyme (Fig. 3A). Bam HI binds to DNA in a "crossover" manner. That is, the left (L) Bam HI subunit cleaves and makes most of its phosphate contacts to the left DNA half-site, but makes almost all of its base-pair contacts to the right half-site of DNA. The opposite is true for the right (R) subunit. This crossover mode of binding differs from that of dimeric repressors, such as λ and 434, where contacts from each subunit are more or less confined to a single DNA half-site (*14*). Thus, the two half-sites of Bam HI recognition sequence are interdependent in that mutations at the R half-site can affect cleavage at the L half-site, and vice versa.

The Bam HI subunit structure consists of a large six-stranded mixed β sheet (β^3, β^4, β^5, β^6, β^7, and β^1), which is sandwiched on both sides by α helices (α^1, α^2, α^3, α^{3A}, α^4, α^5, and α^6). Strands β^3, β^4, and β^5 are antiparallel and form a β meander; strands β^5, β^6, and β^7 are parallel and resemble a Rossmann or mononucleotide binding fold, with α^4 and α^6 acting as the cross over helices (Fig. 3). The dimer interface is formed primarily by α^4 and α^6, which pair with the corresponding helices from the symmetry related subunit to form a parallel four-helix bundle (Fig. 3C). The NH_2-termini of these helices are directed toward the major groove of DNA. Their dipole moments should make an important electrostatic contribution to stabilization of the enzyme-DNA complex. Most of the amino acids that recognize the Bam HI DNA sequence are located near the NH_2-terminus of the four-helix bundle. The catalytic residues Asp^{94}, Glu^{111}, and Glu^{113} are clustered at one end of the β meander, close to the scissile phosphate group.

The DNA retains a B-DNA–like conformation over the central 10 bp (*15*). There are no major bends or kinks of the type seen in Eco RI and Eco RV (specific) DNA complexes (*5*, *12*). Thus, like Pvu II DNA, which also retains a B-DNA–like conformation (*13*), Bam HI DNA shows that severe DNA distortion is not necessary for specific complex formation. DNA distortion had been suggested as a way of positioning the scissile phosphate group in the active site of a specific complex (*16*). For Bam HI and Pvu II, it may be differences in protein conformations or "docking" modes on specific compared to nonspecific sites that position or exclude the scissile phosphate group from the active site.

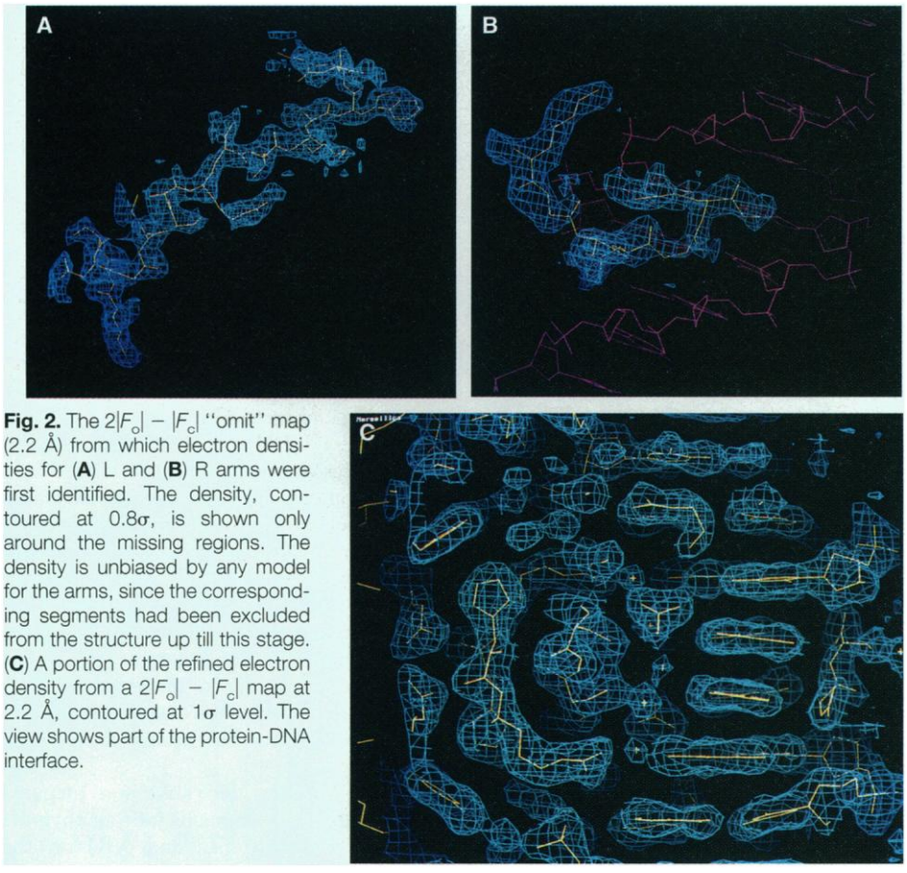

Fig. 2. The $2|F_o| - |F_c|$ "omit" map (2.2 Å) from which electron densities for (**A**) L and (**B**) R arms were first identified. The density, contoured at 0.8σ, is shown only around the missing regions. The density is unbiased by any model for the arms, since the corresponding segments had been excluded from the structure up till this stage. (**C**) A portion of the refined electron density from a $2|F_o| - |F_c|$ map at 2.2 Å, contoured at 1σ level. The view shows part of the protein-DNA interface.

Protein conformational changes. In contrast to the DNA, the Bam HI enzyme undergoes a series of conformational changes on complex formation (Fig. 4). Some of these such as rigid body motion of subunits, ordering of disordered loops, and local side-chain and main chain rearrangements have been observed in other protein-DNA complexes (*14*). However, the observed unwinding of the COOH-terminal α helices to assume extended, partially disordered conformations on DNA binding is unprecedented. These changes are described below.

1) The first conformational change is a quaternary change (Fig. 4). The Bam HI subunits rotate toward the cleft and clamp onto the DNA. The magnitude of this rotation is approximately 19° about an axis roughly parallel to the DNA helical axis (~20° to it) and passing through the middle of the subunit-subunit interface. The rotation leads to a narrowing of the DNA binding cleft. The distance between residues Gly^{194R} and Gly^{194L} across the DNA binding cleft changes from 31 Å in the free enzyme to 21 Å in the enzyme-DNA complex. Many of the contacts between the enzymes and the sugar-phosphate backbone of DNA could not be formed in the absence of this large rigid body motion. The λ Cro dimer also undergoes a large quarternary change on DNA binding (*17*). As in the case of Bam HI, the Cro monomers rotate (from 35° to 50°) about an axis roughly along the DNA axis.

2) The second conformational change is the local folding of residues 79 to 91 which connect strands β^3 and β^4. These residues were disordered in free Bam HI but become ordered on DNA binding (Fig. 4). They now incorporate a single turn of helix (named α^{3A}) between residues 81 to 84, and an extended region between residues 89 to 92. The region lies close to the phosphate backbone of nucleotides Thy^4 and Gua^3, which is only possible because residues 90 and 91 are both glycines. The region also forms part of the dimer interface in the complex, making contacts to the twofold related helix, α^6. The total change in solvent accessible area due to dimer formation is 2050 $Å^2$, an increase of approximately 500 $Å^2$ from the free enzyme. An extra salt bridge is also formed across the dimer interface between Lys^{87L} and Glu^{167R}.

The folding of this region appears to be

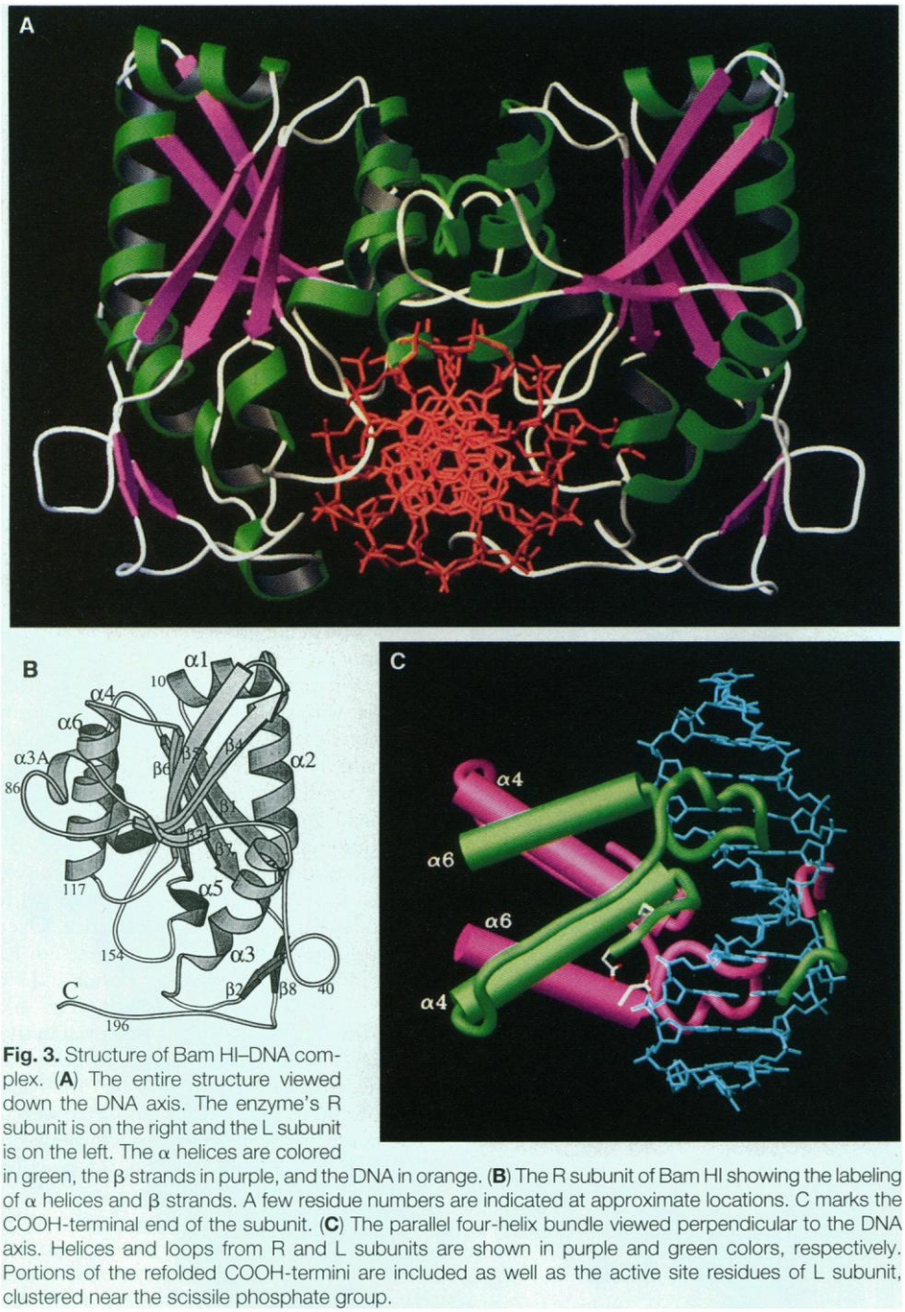

Fig. 3. Structure of Bam HI–DNA complex. (**A**) The entire structure viewed down the DNA axis. The enzyme's R subunit is on the right and the L subunit is on the left. The α helices are colored in green, the β strands in purple, and the DNA in orange. (**B**) The R subunit of Bam HI showing the labeling of α helices and β strands. A few residue numbers are indicated at approximate locations. C marks the COOH-terminal end of the subunit. (**C**) The parallel four-helix bundle viewed perpendicular to the DNA axis. Helices and loops from R and L subunits are shown in purple and green colors, respectively. Portions of the refolded COOH-termini are included as well as the active site residues of L subunit, clustered near the scissile phosphate group.

driven by a set of contacts established between the main chain atoms of residues 90 and 91 and the phosphate backbone of nucleotides Thy[4] and Gua[3]. A similar folding is seen in other protein-DNA complexes; for example, the NH_2-terminal arm of λ repressor (*6*), the basic helices of bZip and helix-loop-helix proteins (*7*), the second zinc module of glucorticoid receptor (*18*), and the DNA recognition loop of Eco RV (*12*) are all examples of regions that become ordered on DNA binding. Direct interactions between the regions and the DNA provide the necessary free energy for the disorder to order transition. Disordered regions allow better molding of proteins to their DNA sites, but their resulting overall binding energies would be less than that with performed structures (*19*). However, for Bam HI and glucorticoid receptor the new dimer contacts, established by the DNA induced refolding, should enhance the overall binding energies.

3) The third type of conformational change is a local rearrangement of the loop (residues 152 to 157) preceding helix α^6. Many of the side chain and main chain atoms that interact specifically with DNA are located on this loop. The side chain of Arg^{155} swings about 90° from its conformation in the free enzyme to form specific base contacts with Gua[3]. In order to participate in this interaction, the main chain of Arg^{155} occupies a region on the Ramachandran plot that is not normally occupied by nonglycine residues; its unfavorable torsion angles are stabilized by a hydrogen bond between Ile^{117}N and Arg^{155}O (2.8 Å). In both subunits, there is also a local distortion or kinking of helix α^3 on DNA binding. This enables residues at the NH_2-terminal end of the helix to make direct contacts with the DNA backbone.

4) The most striking conformational change is the unfolding of α helices at the COOH-terminal of the protein (residues 194 to 213). This region contained an ordered α helix (α^7) in the free protein, but when bound to DNA it becomes extended and partially disordered. We refer to the extended conformation as an "arm." The arm from the R subunit binds to the minor groove of the DNA, whereas the arm from the twofold related L subunit folds back toward the core of the protein (Fig. 4). The minor groove is too narrow to accommodate arms from both subunits, if they folded symmetrically into the groove.

The R and L arms can only be traced to residues 198 and 208, respectively. The ensuing residues are presumed to be disordered in the complex. The Gly^{194} residue anchors the beginning of each arm to DNA by a hydrogen bond between its NH group and phosphate of $Thy^{1'}$. The R arm continues a path toward the back side of DNA and into the minor groove. This positions it directly opposite from the four-helix bundle, making a series of contacts in the minor groove (described below). The L arm follows a path that is roughly parallel to the sugar-phosphate backbone of the DNA (Fig. 5). Met^{198L} contacts the DNA backbone directly, whereas Lys^{207L} reaches over to form a salt bridge with the phosphate group of a translationally related DNA fragment. The latter contact is likely to be preserved with a longer DNA fragment. The other basic residues on the L arm face the core of the protein so that Arg^{201} forms a salt bridge to Glu^{69} and Lys^{204} is 3.8 Å from the carbonyl group of Asn^{73}. The conformation of the L arm is also stabilized by the packing of Trp^{206} against Tyr^{75} and Tyr^{96} (Fig. 5).

The question arises as to what drives the refolding of the COOH-terminal helices (α^7) on DNA binding? The R arm clearly gains a new set of interactions in the minor groove of DNA. The L arm gains a possible salt bridge between Lys^{207} and a phosphate group, and a new set of interactions are established within the protein. These interactions may tip the free energy balance from helical to extended and partially disordered conformations on DNA binding. α-Helices are only marginally stable by themselves (*20*), their configuration in proteins is stabilized by tertiary interactions. In

free Bam HI, α^7 was the most mobile ($\langle B \rangle$ = 33 Å^2) and solvent accessible helix of the enzyme (*3*), suggesting a low-energy threshold between folded and unfolded states. The kinking of helix α^3 and the movement of Gly^{194} to form the anchoring hydrogen bond could also contribute to the destabilization of the COOH-terminal helix on DNA binding. Crystal packing effects cannot explain the refolding of the COOH-terminal regions because there are no crystal lattice contacts to the R arm and only a couple of contacts to the L arm (one of which mimics contact to a longer piece of DNA). In the free enzyme, the COOH-terminal region makes a few crystal contacts, but these are far exceeded by interactions established within the protein (*21*).

Refolding of secondary structure occurs most dramatically in the structure of influenza haemagglutinin, when induced by a pH change (*22*). The structure of Bam HI–DNA complex provides an example of an α helix refolding on DNA binding. Spolar and Record (*23*) have analyzed the large negative heat capacity changes accompanying site specific DNA binding in terms of quarternary rearrangement of protein subunits and the ordering of disordered regions. Bam HI shows that unraveling of secondary structural elements can also be used to create part of the protein-DNA interface. Moreover, the entropic cost of forming a complex, in some cases, could be partially offset by regions becoming more disordered on DNA binding.

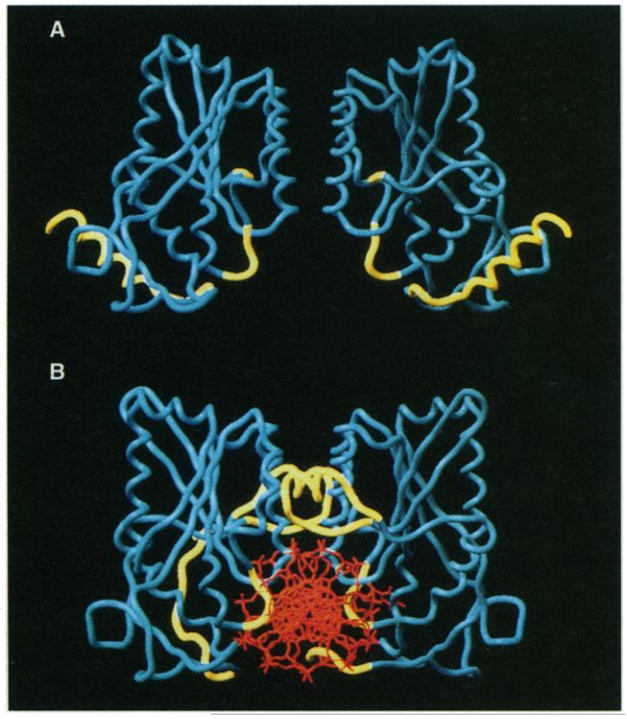

Fig. 4. Structures of (**A**) free and (**B**) DNA bound forms of Bam HI. Regions that undergo conformational change are shown in yellow color. Upon DNA binding, the residues between strands β^3 and β^4 become ordered but the COOH-terminal helices unwind and adopt extended, partially disordered conformations (R and L arms).

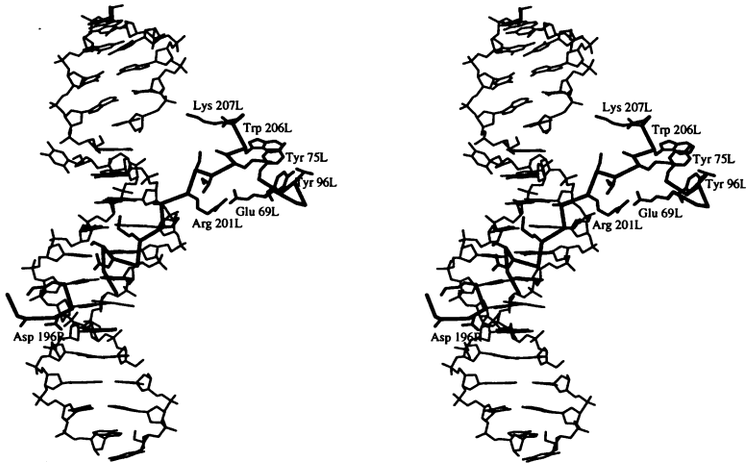

Fig. 5. A stereo view of the refolded R and L arms. The R arm enters the DNA minor groove while the L arm follows the sugar-phosphate backbone. The DNA is shown with a few nucleotides from the screw-related DNA which stacks to form a pseudocontinuous helix.

Protein-DNA recognition. The Bam HI complex displays an impressive range of protein-DNA interactions. For example, side chain atoms, main chain atoms, tightly bound water molecules, all contribute toward recognition of the Bam HI sequence. Interactions occur in both the major and the minor grooves of DNA. In the major groove, every hydrogen bond donor and acceptor group takes part in direct or water-mediated hydrogen bonds with the protein. This complementarity at the protein-DNA interface ensures that only the Bam HI recognition sequence can make all of the necessary interactions. Any other DNA sequence would lead to a loss of hydrogen bonds or steric clashes (or both).

1) The major groove contacts are formed primarily by regions at the NH_2-terminal ends of helices α^4 and α^6. These helices dimerize with the corresponding helices from the twofold related subunit to form a parallel four-helix bundle. The α^4 helices from subunits L and R form a V shape that penetrates the major groove at the center of the 6-bp recognition sequence (Fig. 3C). Helices α^6 themselves do not enter the major groove, but the loops preceding them interact with the ends of the recognition sequence. Over the 6-bp recognition sequence, there are 12 direct protein-DNA hydrogen bonds and six water-mediated hydrogen bonds (Fig. 6). Thus, the hydrogen bonding potential in the major groove is totally satisfied. Most of the specific contacts to the R-DNA half-site come from the L subunit, and vice versa. The contacts are almost identical between the two DNA half-sites, and we describe them for the R half-site only.

The outer G·C base pair is contacted by the 152–157 loop that precedes helix α^6 (Fig. 6). Two hydrogen bonds are donated by Arg^{155} to Gua^3 [$N^{\eta 2}$. . .N7 (2.9 Å), N^{ϵ}. . .O6 (2.6 Å)]. Interaction between arginine and guanine is probably the most commonly observed (and predictive) interaction in protein-DNA complexes (*14*). However, the hydrogen-bonding pattern we observe differs from usual pattern in that N^{ϵ} and $N^{\eta 2}$ rather than $N^{\eta 1}$ and $N^{\eta 2}$ of Arg^{155} form the bidentate hydrogen bonds to guanine. The unusual orientation of Arg^{155} is stabilized by a salt bridge to Glu^{161} [$N^{\eta 1}$. . .$O^{\epsilon 2}$ (2.9 Å), $N^{\eta 2}$. . .$O^{\epsilon 1}$ (2.7 Å)]. The outer cytosine base, $Cyt^{3'}$, makes a single hydrogen bond with Asp^{154} [N4. . .$O^{\delta 1}$ (2.9 Å)]. Hydrogen bonds between aspartate and cytosine residues have also been observed for the fourth and fifth zinc fingers of the human glioblastoma

GL1-DNA complex (*24*).

Contacts to the middle G·C base pair arise from the 152–157 loop, the NH_2-terminal end of helix α^4, and a water molecule (Fig. 6). Asn^{116} donates a single hydrogen bond to Gua^2 through its terminal amino group [$N^{\delta 2}$. . .O6 (3.1 Å)]. The amino group is properly positioned by a hydrogen bond between $O^{\delta 1}$ of Asn^{116} and the main chain nitrogen of Ser^{118}. The N7 atom of Gua^2 is hydrogen-bonded to a tightly bound water molecule [N7. . .O (3.0 Å)], which in turn is linked to Arg^{122R} [$N^{\eta 2}$. . .O (3.0 Å)] and Asn^{116} [$O^{\delta 1}$. . .O (2.6 Å)]. The hydrogen-bonding potential of $Cyt^{2'}$ is satisfied by a bond to the main chain carbonyl of Asp^{154} [N4. . .O (2.7 Å)]. Hydrogen bonds between main chain carbonyl groups and amino groups of cytosine bases have also been observed in Eco RV–DNA and Gal4-DNA complexes (*12*, *25*). The close approach of the main chain atoms to $Cyt^{2'}$ helps to explain why methylation of this residue (at atom N4) by Bam HI methylase would sterically inhibit the endonuclease from binding to its specific site.

In the Bam HI–DNA complex, only the thymine base of the inner A·T base pair is involved in a hydrogen bond with the protein (Fig. 6). The Asn^{116}, which donated a hydrogen bond to Gua^2, also donates a hydrogen bond to $Thy^{1'}$ [$N^{\delta 2}$. . .O4 (2.9 Å)]. Thus, the side chain of Asn^{116} bridges between Gua^2 and $Thy^{1'}$. The nonpolar portion of Asn^{116} side chain also forms van der Waals contacts with the methyl group of $Thy^{1'}$, a situation similar to the Gln^{29}-$Thy^{3'}$ interaction in the 434 repressor–operator complex (*26*). There are no direct hydrogen bonds between Ade^1 and the en-

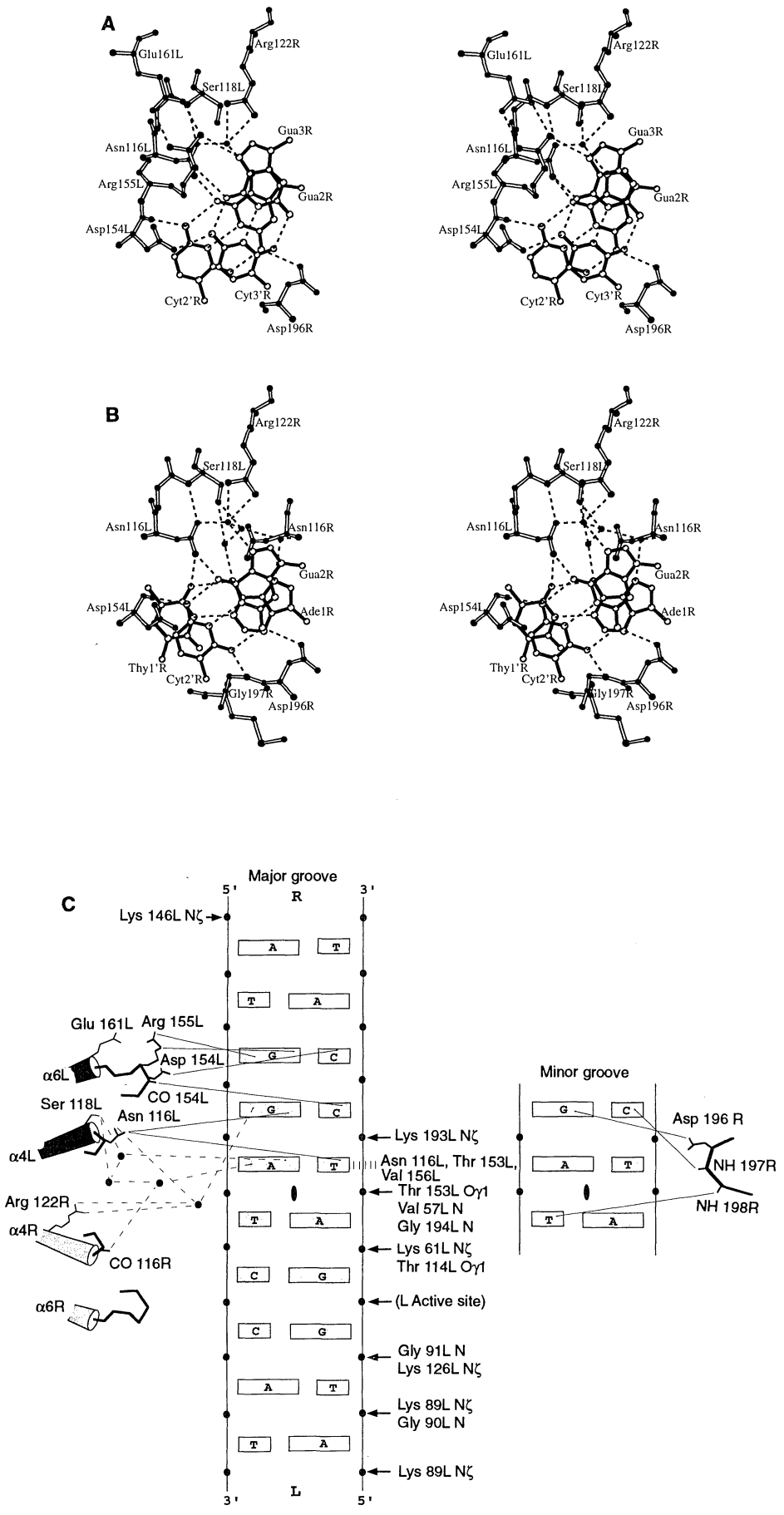

Fig. 6. Protein-DNA interactions. (**A**) A stereo view of the interactions between the enzyme and the outer (GC) and middle (GC) base pairs of the recognition sequence. (**B**) A stereo view of the interactions between the enzyme and the middle (GC) and inner (AT) base pairs of the recognition sequence. Dashed lines indicate hydrogen bonds. Conserved water molecules are shown as black spheres. Interactions are shown for both the major and minor groove edges of the base-pairs. (**C**) A sketch summarizing contacts between Bam HI and DNA: (i) Contacts to base pairs in the major groove of R half-site. The 6-bp recognition sequence is shown shaded. Direct hydrogen bonds between the enzyme and the bases are shown by solid lines, water mediated hydrogen bonds by dashed lines, and van der Waal contacts by a row of vertical lines. Base pair contacts to the L half-site are exactly the same. (ii) Hydrogen bonds between base pairs and R arm in the minor groove. (iii) Contacts between the L subunit and the sugar-phosphate backbone of DNA. The R subunit makes identical, twofold symmetric contacts. Note that the L subunit makes most of its base pair contacts to the R half-site, but most of its backbone contacts to the L half-site.

zyme; they are instead mediated by an intricate network involving three highly localized water molecules (Fig. 6). Similar water-mediated recognition of base pairs has been observed in a number of protein-DNA complexes, most notably the *trp* repressor-operator complex (*27*).

2) Minor groove contacts are formed by the refolded COOH-terminal arm of *R* subunit (Fig. 6). As noted above, there is insufficient room within the minor groove to accommodate arms from both subunits if they were to fold symmetrically in the groove. The possibility of steric clashes between symmetrically folded NH_2-terminal arms was similarly suggested as the cause of λ repressor's asymmetric binding to its DNA site (*28*). The R arm of Bam HI makes specific interactions with both DNA half-sites. There are hydrogen bonds between Asp^{196R} and Gua^{2R} [$O^{\delta 1}$. . .N2 (3.0 Å)], Gly^{197R} and $Cyt^{2'R}$ [N. . .O2 (3.0 Å)], Met^{198R} and $Thy^{1'L}$ [N. . .O2 (2.8 Å)], and van der Waals contacts between Gly^{197R} and Ade^{1R} [C^{α}. . .C2 (3.4 Å)].

3) The interactions between the enzyme and the sugar-phosphate backbone of DNA are extensive. Each subunit makes 11 direct hydrogen bonds to DNA phosphate groups (Fig. 6C), with approximately eight more mediated via water molecules. The residues involved come from several segments of the enzyme: the NH_2-terminal end of helix α^3 (57 to 61), the ordered region just before strand β^4 (89 to 91), the NH_2-terminal end of helix α^4 (111 to 126), the NH_2-terminal end of helix α^6 (146 to 165), and residues just before the refolded COOH-terminus (193 to 194). Of the direct phosphate contacts, seven involve side chain groups (five of which are lysines) and four are from chain NH groups (involving three glycine residues). The L subunit contacts seven consecutive phosphates from 5L to 2'R, whereas the R subunit contacts the noncrystallographic symmetry related 5R to 2'L. Thus, contacts to the central four nucleotides of the recognition sequence overlap between the two strands.

Active sites and catalytic mechanism. Type II restriction enzymes require only Mg^{2+} as a cofactor to catalyze the hydrolysis of DNA phosphodiesters, leaving free 5' phosphate and 3' hydroxyl groups. The reaction is considered to proceed by an in-line displacement of the 3' leaving group, in which an activated water molecule acts as the attacking nucleophile (*29*). The active sites are found to be structurally similar for the four restriction enzymes whose structures are known. Residues Asp^{94}, Glu^{111} and Glu^{113} in Bam HI can be spatially aligned with residues Asp^{91}, Glu^{111} and Lys^{113} in Eco RI (*5*), residues Asp^{74}, Asp^{90} and Lys^{92} in Eco RV (*12*), and residues Asp^{58}, Glu^{68} and Lys^{70} in Pvu II (*13*). Several mechanisms have been proposed for the way these enzymes might activate a water molecule for a nucleophilic attack (*12, 30*). For Bam HI, based on the structure of the free enzyme, we suggested a general base mechanism in which the acidic residues Asp^{94} and Glu^{111} coordinate a Mg^{2+} at the active site, while Glu^{113} acts as a general base to deprotonate the attacking water molecule (*2*). Based on the structure of the complex, however, a two metal mechanism may in fact be more likely.

The Bam HI–DNA complex shows that the geometries of the L and R active sites are similar. In particular, three water molecules labeled as A, B, and C are positioned at almost identical positions in the two active sites (Fig. 7). Water molecules A and B are separated by about 4 Å, and occupy positions that bear resemblance to the two metal sites in the exonuclease domain of DNA polymerase I (*31*). Water molecule A is located centrally at the active site, within hydrogen-bonding distances from the carboxylate groups of Glu^{111} and Glu^{113}, the nonesterified oxygens of the scissile phosphate group, and the main chain carbonyl group of Phe^{112}. Water molecule B is located toward the 3' end of the scissile phosphate, a short hydrogen-bonding distance from Glu^{111} and a long hydrogen-bonding distance from the leaving O3' atom (Fig. 7). Also, Asp^{94} could hydrogen-bond to water molecule B, if it adopted the configuration observed in the free enzyme (*2*). In the presence of Mg^{2+}, water molecules A and B could easily be substituted by cations, functioning in the manner postulated for DNA polymerase I (*31*), with the cation at site A helping to activate the attacking water molecule with the cation at site B helping to stabilize the negative charge on the 3' oxyanion leaving group, and both cations helping to stabilize the pentacovalent transition site. Water molecule C may be a candidate for the attacking water molecule. It is hydrogen-bonded to the phosphate group of the neighboring 3' nucleotide (Ade^{1}), and is the most linearly positioned water molecule with respect to the scissile P-O3' bond (angle of about 150°). At present, its distances from the phosphorus atom of the scissile phosphate, the carboxylate group of Glu^{113}, and water molecule A are >4 Å, but these could change on Mg^{2+} binding.

Comparison to Eco RI. The structural similarity between Bam HI and Eco RI was noted previously (*2*). The resemblance was unexpected because of the lack of sequence similarity between the two enzymes. The cores of both enzymes consist of a central β sheet with α helices on both sides. Two α helices from each subunit come together to form a parallel four-helix bundle. The NH_2-terminal end of the four-helix bundle is directed toward the major groove of DNA in both enzymes. Despite this similarity, none of the protein-DNA contacts observed in the Bam HI–DNA complex could have been anticipated from the structure of Eco RI–DNA complex (*5*). Although the angles between the helices of the bundle are similar in both enzymes, the rms difference for Cα atoms is high (3.0 Å). Moreover, the inner and outer arms that precede the he-

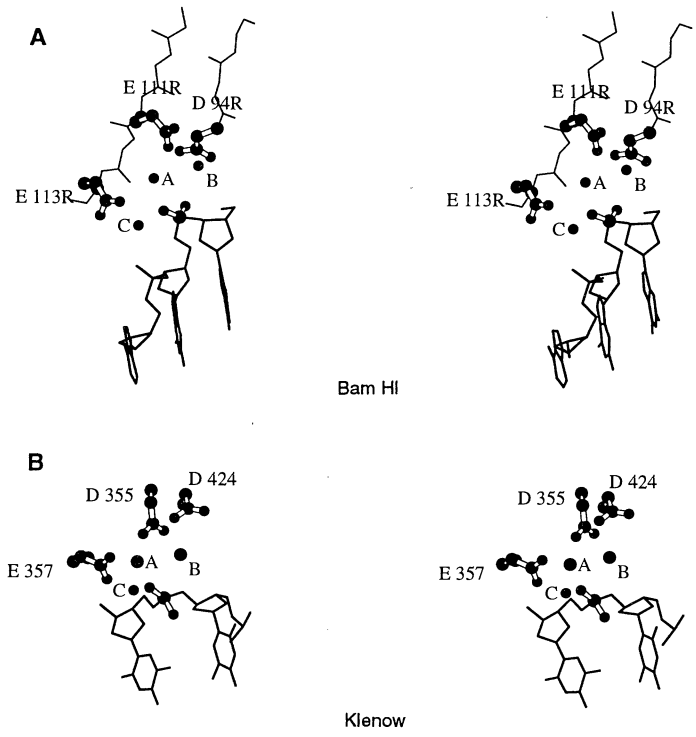

Fig. 7. Stereo views of the (**A**) Bam HI and (**B**) DNA polymerase I (Klenow) active sites. The views have been selected to emphasize the similarities between the active sites. The Bam HI active site is from the right DNA half-site, and A, B, and C are labels for three water molecules. In the polymerase active site, A and B are labels for two divalent cations and C is a label for a water molecule. The active site residues and the scissile phosphate group are shown in ball and stick representation.

lices of the bundle in Eco RI are missing in Bam HI. The beginning portion of the inner arm described as the extended chain motif, which makes many of the protein-DNA contacts (especially to pyrimidines) in Eco RI (*5*), is also absent in Bam HI. Thus, the four-helix bundle of Bam HI and Eco RI cannot be viewed as a conserved DNA recognition motif, in the way that the helix-turn-helix (HTH) motif or the zinc-finger is viewed in transcription factors (*14*). The large insertions and deletions lead to different geometries for base pair contacts in Bam HI and Eco RI. The presence of a four-helix bundle in the two enzymes may be due less to similarity of their DNA sites (GGATCC compared to GAATTC), but more to their similar cleavage properties; both enzymes cleave with a 4-bp stagger to produce 5′ overhanging ends. If such is the case, the four-helix bundle may be regarded more as a conserved dimerization motif for positioning the active sites at a separation of about 17 to 19 Å along the DNA axis, rather than for conserved DNA contacts (*2, 3*). Interestingly, Eco RV and Pvu II are structurally similar but recognize their DNA sites by different elements, namely, antiparallel β strands in Pvu II and a loop region in Eco RV (*12, 13, 32*). Both enzymes cleave DNA to produce blunt ends with no stagger of base pairs, and their similarity may again be due to a need to position the active sites at a separation of only a couple of angstroms along the DNA axis.

The diversity of protein-DNA contacts in restriction enzymes may be a reflection of the selective pressure that they face. A bacterial chromosome contains many sites that differ by only a single base pair from the specific site. These nonspecific sites would be unmethylated, and their accidental cleavage by the host restriction enzyme could prove fatal for the cell. Restriction enzymes must therefore evolve a protein-DNA interface that is highly complementary toward their cognate DNA site. The interface needs to be easily disrupted by noncognate sequence that differ by even a single base pair and the effect transmitted to the active sites. The search for "tight" complementarity may be one of the reasons for the surprising diversity of protein sequences amongst restriction enzymes (*1*).

REFERENCES AND NOTES

1. R. J. Roberts and S. E. Halford, in *Nucleases*, S. M. Linn, R. S. Lloyd, R. J. Roberts, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ed. 2, 1993), pp 35–88.
2. M. Newman, T. Strzelecka, L. F. Dorner, I. Schildkraut, A. K. Aggarwal, *Nature* **368**, 660 (1994).
3. ———, *Structure* **2**, 439 (1994).
4. Y. Kim, J. C. Grable, R. Love, P. J. Greene, J. M. Rosenberg, *Science* **249**, 1307 (1990).
5. J. M. Rosenberg, *Curr. Opin. Struct. Biol.* **1**, 104 (1991).
6. S. R. Jordan and C. O. Pabo, *Science* **242**, 893 (1988).
7. T. Ellenberger, *Curr. Opin. Struct. Biol.* **4**, 12 (1994).
8. T. Strzelecka, M. Newman, L. F. Dorner, I. Schlidkraut, A. K. Aggarwal, *J. Mol. Biol.* **239**, 430 (1994).
9. P. M. D. Fitzgerald, *J. Appl. Crystallogr.* **20**, 383 (1988).
10. A. T. Brunger, *X-PLOR v3.0 Manual* (Yale University, New Haven, CT, 1992).
11. D. E. Tronrud, L. F. Ten Eyck, B. W. Matthews, *Acta Crystallogr.* **A43**, 489 (1987).
12. F. K. Winkler *et al.*, *EMBO J.* **12**, 1781 (1993).
13. X. Cheng, K. Balendiran, I. Schildkraut, J. E. Anderson, *ibid.* **13**, 3927 (1994).
14. S. C. Harrison and A. K. Aggarwal, *Annu. Rev. Biochem.* **59**, 993 (1990); C. O. Pabo and R. T. Sauer, *Annu. Rev. Biochem.* **61**, 1053 (1992).
15. The small deviations from regular B-DNA were determined with program CURVES [R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.* **6**, 655 (1989)]. The average helical parameters are 32.3° for twist and 3.47 Å for rise. The role angles are generally small (< |8°|), but the propeller twist increases to an average of −19.2° at the central AT step. The minor groove width varies from 10.7 to 12.8 Å (compared to 11.6 Å for B-DNA), and the major groove width varies from 19.3 to 22.2 Å (compared to 18.4 Å for B-DNA). In order to achieve end-to-end stacking within the cocrystal, the terminal nucleotides of the Bam HI 12-nt fragment are distorted in an unexpected manner. Thy^{6R} swings completely out of the DNA helix; its conformation stabilized by stacking interactions with tyrosine residues on the surfaces of screw-related Bam HI subunits. $Ade^{6'L}$, which would normally form base pairs with Thy^{6L}, is disordered and cannot be seen in the electron density map. The resulting gap in the helix is filled by base Thy^{6R} (on the translationally related DNA fragment), which disengages from its Watson-Crick mate $Ade^{6'R}$. These distortions convey the importance of stacking interactions in the formation of a DNA helix. Thy^{6R} would rather lose two hydrogen bonds with $Ade^{6'R}$ in order to maintain stacking interactions within the helix. These distortions could also explain why reproducible cocrystals were only obtained with a 12-nt oligomer containing 5′-TAT ends, and not 5′-GCG ends (*8*).
16. F. K. Winkler, *Curr. Opin. Struct Biol.* **2**, 93 (1992).
17. R. G. Brennan, S. L. Roderick, Y. Takeda, B. W. Matthews, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8165 (1990).
18. B. F. Luisi *et al.*, *Nature* **352**, 497 (1991).
19. A. D. Frankel and P. S. Kim, *Cell* **65**, 717 (1991).
20. T. E. Creighton, *Proteins: Structures and Molecular Properties* (Freeman, New York, ed. 2, 1993).
21. Residues 194 to 213 at the COOH-terminus lose ~253 $Å^2$ of solvent-accessible surface area as a result of crystal contacts, but more than 730 $Å^2$ are lost because of contacts within the protein. This argues that stabilization of the helix at the COOH-terminus of the free enzyme is due primarily to tertiary interactions within the protein and is not due to crystal contacts. There are no published solution data that we are aware of on the comparison between free and bound conformations of Bam HI. Fluorescence measurements (using suitably tagged fluorescent probes) may be particularly effective in monitoring the conformational transitions we describe in this article.
22. P. A. Bullough, F. M. Hughson, J. J. Skehel, D. C. Wiley, *Nature* **371**, 37 (1994).
23. R. S. Spolar and M. T. Record Jr., *Science* **263**, 777 (1994).
24. N. P. Pavletich and C. O. Pabo, *ibid.* **261**, 1701 (1993).
25. R. Marmorstein, M. Carey, M. Ptashne, S. C. Harrison, *Nature* **356**, 408 (1992).
26. J. E. Anderson, M. Ptashne, S. C. Harrison, *ibid.* **326**, 846 (1987); A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne, S. C. Harrison, *Science* **242**, 899 (1988).
27. Z. Otwinowski *et al.*, *Nature* **335**, 321 (1988).
28. A. Sarai and Y. Takeda, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6513 (1989).
29. B. A. Connolly, F. Eckstein, A. Pingoud, *J. Biol. Chem.* **259**, 10760 (1984); J. A. Grasby, B. A. Connolly, *Biochemistry* **31**, 7855 (1992).
30. A. Jeltsch, A. Jurgen, H. Wolfes, G. Maass, A. Pingoud, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8499 (1993).
31. L. S. Beese and T. A. Steitz, *EMBO J.* **10**, 25 (1991).
32. A. Athanasiadis *et al.*, *Nature Struct. Biol.* **1**, 469 (1994).
33. N.-H. Xuong, S. T. Freer, R. Hamlin, C. Nielsen, W. Vernon, *Acta Crystallogr.* **A34**, 289 (1978).
34. N. Sakabe, *Nucl. Instrum. Method Phys. Res.* **A303**, 448 (1991).
35. T. Higashi, *J. Appl. Crystallogr.* **22**, 9 (1989).
36. CCP4, The SERC collaborative computing project No. 4 (Daresbury Laboratory, Warrington WA4 4AD, UK, 1979).

37. We thank N. Sakabe and A. Nakagawa for facilitating data collection at Photon Factory; R. Knott for DNA synthesis; T. Steitz and J. Jaeger for coordinates; X. Jiang for discussion; and W. Hendrickson for comments on the manuscript. Supported by NIH grant GM-44006 (A.K.A.); and a Fogarty International Fellowship (M.N.). Coordinates have been deposited in the Brookhaven Protein Database. They can also be obtained by email (Aggarwal@cuhhca.hhmi.Columbia.edu).

6 January 1995; accepted 14 June 1995