Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 10957–10961, November 1994
Biochemistry

Three-dimensional structure of the adenine-specific DNA methyltransferase M·*Taq* I in complex with the cofactor *S*-adenosylmethionine

J. LABAHN*, J. GRANZIN*, G. SCHLUCKEBIER*, D. P. ROBINSON†, W. E. JACK†, I. SCHILDKRAUT†, AND W. SAENGER*‡

*Institut für Kristallographie, Freie Universität, Takustrasse 6, D-14195 Berlin, Germany; and †New England Biolabs, 32 Tozer Road, Beverly, MA 01915

Communicated by Hamilton O. Smith, July 8, 1994 (received for review March 3, 1994)

ABSTRACT **The *Thermus aquaticus* DNA methyltransferase M·*Taq* I (EC 2.1.1.72) methylates N^6 of adenine in the specific double-helical DNA sequence TCGA by transfer of —CH_3 from the cofactor *S*-adenosyl-L-methionine. The x-ray crystal structure at 2.4-Å resolution of this enzyme in complex with *S*-adenosylmethionine shows α/β folding of the polypeptide into two domains of about equal size. They are arranged in the form of a C with a wide cleft suitable to accommodate the DNA substrate. The N-terminal domain is dominated by a nine-stranded β-sheet; it contains the two conserved segments typical for N-methyltransferases which form a pocket for cofactor binding. The C-terminal domain is formed by four small β-sheets and α-helices. The three-dimensional folding of M·*Taq* I is similar to that of the cytosine-specific *Hha* I methyltransferase, where the large β-sheet in the N-terminal domain contains all conserved segments and the enzymatically functional parts, and the smaller C-terminal domain is less structured.**

DNA-methyltransferases (MTases) are a family of enzymes that occur in nearly all living organisms. They catalyze the transfer of —CH_3 from the cofactor *S*-adenosyl-L-methionine (AdoMet) to cytosine C5 (C-MTases) or cytosine N^4 or adenine N^6 (N-MTases) in di- to octanucleotide target sequences of double-stranded DNA (1). In bacteria, all three types of MTases are found and implicated in the protection of DNA from their own restriction endonucleases and in mismatch repair (2). In eukaryotes only C-MTases have been observed so far; they are involved in cell differentiation, genome imprinting, mutagenesis, and regulation of gene expression (3).

The C-MTases are a homogeneous class of molecules with three-dimensional structures probably similar to the structure described recently for the M·*Hha* I enzyme from *Haemophilus haemolyticus* (4). This is because their amino acid sequences show sequential arrangement of 10 conserved segments (I to X) from the N to the C terminus (5); segments I (DXFXGXG, with X = any amino acid) and IV (FPCQ) are implicated in binding of AdoMet, and the cysteine in IV is involved in the transfer of —CH_3. In contrast, the N-MTases show only two of the conserved segments (6). They correspond to segments I and IV in the C-MTases, namely I (DXFXGXG), which can degenerate so much that only one glycine is retained, and II (DPPY), where aspartate can be replaced by asparagine or serine, and tyrosine by phenylalanine. Because these two segments can occur in reversed order—i.e., one or the other N-terminal (7)—the N-MTases are a more heterogeneous class of molecules. When the amino acid sequences of only those N-MTases that recognize TNNA (N = any nucleotide) are compared, an additional segment III is found (8). It spans 38 amino acids, has no equivalent in C-MTases, and occurs sequentially—i.e., I, II, III. The mechanism of methyl transfer is different in C- and N-MTases. In the former the conserved cysteine SH in segment IV attacks C6 of cytosine to form a covalent intermediate with resonance-stabilized carbanionic C5 that is able to take up CH_3^+ with subsequent release of H^+ and breaking of the S—C6 bond (9, 10). In N-MTases, the reaction mechanism is not clear. We know that the configuration of CH_3^+ is inverted after transfer (11) and that it probably attacks adenine N^6 directly. The methylation mechanism as known in organic chemistry—i.e., electrophilic attack at adenine N1 followed by Dimroth rearrangement to yield N^6-methyladenine—is not observed (12).

In the homogeneous class of C-MTases, M·*Hha* I serves as a structural prototype. We describe here the three-dimensional structure of an N-MTase, M·*Taq* I from *Thermus aquaticus*, in complex with the cofactor AdoMet.§ This enzyme (EC 2.1.1.72) methylates adenine N^6 in the target sequence TCGA, consists of 421 amino acids (47,856 Da) (13, 14), and contains the three conserved sequences I [EPACAHG (amino acid residues 45–51)], II [NPPY (105–108)], and III [amino acid residues 141–178].

METHODS

M·*Taq* I was produced by using the expression vector pPR594 provided by P. Riggs (New England Biolabs) (15). This plasmid was cut with *Hin*dIII and *Nco* I and treated with calf intestinal phosphatase to remove 5′ PO_4 groups. Two fragments were ligated to this vector: (*i*) a 1.2-kb *Bst*XI/*Hin*dIII fragment from pSW149M-4A (13) that contained all but the N terminus of the M·*Taq* I coding region and (*ii*) an oligonucleotide duplex adapter with *Bst*XI and *Nco* I cohesive ends that completed the M·*Taq* I gene and positioned the initiation codon downstream of the *lacZ* ribosome-binding site in pPR594. The two strands of the duplex were 5′-C ATG GGT CTC CCT CCT CTC CTC TCT CTC CCT TCT AAC GCT GCT CCT AGG AGC-3′ and 5′-CT AGG AGC AGC GTT AGA AGG GAG AGA GAG GAG AGG AGG GAG ACC-3′. The plasmid with the expected restriction pattern was named pAGL15 and used in the expression studies described here. Expression was in *Escherichia coli* strain ER1821 (E. Raleigh, New England Biolabs): F⁻ e14⁻ (*mcrA*⁻) *endA1 supE44 thi-1* Δ(*mcrC–mrr*)114::IS*10*.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MTase, DNA-methyltransferase; C-MTase, cytosine-C5 DNA-methyltransferase; N-MTase, N DNA-methyltransferase; AdoMet, *S*-adenosyl-L-methionine; MIRAS, multiple isomorphous replacement including anomalous scattering.

‡To whom reprint requests should be addressed.

§The atomic coordinates have been deposited in the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, NY 11973 (reference 1ADM).

Table 1. Crystallographic and MIRAS data

Parameter	Native	K_2PtCl_4	$[PtCl_2(C_2H_4)]_2$	$C_6H_3(OH)$-(COOH)HgCl
Wavelength, Å	1.5418	1.008	1.5418	1.00
Resolution, Å	2.4	2.7	3.2	4.0
Observations	220,712	154,082	58,965	31,494
Unique reflections	37,305	25,769	14,351	8,453
Completeness, %	94.9	93.7	93.5	97.6
R_{sym}, %	9.1	7.9	6.0	4.1
Mean fractional isomorphous difference	—	0.18	0.17	0.16
R_{VECREF}	—	0.46	0.49	0.53
No. of sites/no. of new sites	—	4/4	8/4	2/2
Figure of merit (all derivatives)	0.688			

$R_{Sym} = \Sigma_i\Sigma_j|\langle I_i\rangle - I_{ij}|/\Sigma_i\Sigma_j I_{ij}$, where I_{ij} are the individual measurements contributing to the mean reflection intensity $\langle I_i\rangle$; mean fractional isomorphous difference $= \Sigma|F_{PH} - F_P|/\Sigma F_P$, where F_P and F_{PH} are the protein and derivative structure amplitudes, respectively; $R_{VECREF} = [\Sigma w(\mathbf{u})(P_o(\mathbf{u}) - P_c(\mathbf{u}))/\Sigma w(\mathbf{u})P_o^2(\mathbf{u})]^{1/2}$, $\mathbf{u} \in P$ (Patterson vector space), where P_o and P_c are the observed and calculated Patterson function, $w(\mathbf{u})$ is a function of $\mathbf{u}$, the resolution d_{min}, and the point group symmetry (17).

Purified M·*Taq* I was crystallized by using sitting drop vapor diffusion. M·*Taq* I at 0.1 mM in a buffer containing 100 mM NaCl, 10 mM 2-mercaptoethanol, 20 mM Tris·HCl (pH 7.5) was equilibrated against 12% (wt/vol) 6-kDa polyethylene glycol in the same buffer. Attempts were made to cocrystallize M·*Taq* I with DNA containing the sequence TCGA flanked on both sides by two to six additional nucleotides. As shown by difference Fourier syntheses, however, no DNA was incorporated in the crystals, but the addition of oligonucleotides had a significantly positive effect on the final size of the crystals. Consequently we added 0.35 mM GC-TCGAGC to all crystallization experiments. AdoMet was located in the electron density without having been added to the crystallization buffer, indicating that the complex between M·*Taq* I and AdoMet is stable enough to survive purification.

The first characterization of the space group indicated an orthorhombic primitive cell with the systematic extinction and Laue symmetry of space group $P2_12_12_1$, but a Patterson synthesis calculated with the native data showed a prominent vector at $u = 0.31$, $v = 0.45$, $w = 0.5$ that was 23% of the height of the origin peak. This vector indicated the occurrence of a translation symmetry and explains the observed extinctions along (001). The space group assignment was consequently changed to $P2_12_12$. The unit cell dimensions are $a = 130.66$ Å, $b = 141.83$ Å, $c = 53.15$ Å, and there are two molecules in the asymmetric unit. All x-ray diffraction data were collected on MAR Research Image plate scanners using different x-ray sources [synchrotron at HasyLab/European Molecular Biology Laboratory outstation at the Deutsches Electronen Synchrotron (DESY) Hamburg and the Science and Energy Research Council (SERC) Daresbury Laboratory, and a standard source: Enraf Nonius FR571 rotating-anode generator]. The measurements were evaluated, scaled, and merged by the program MOSFLM (16). Two of three derivative datasets, K_2PtCl_4 and $C_6H_3(OH)(COOH)HgCl$, were measured near the absorption edge (Hamburg, X31: wavelength = 1.008 Å, Daresbury: wavelength = 1.0 Å) to optimize the anomalous signal. The native data set and one additional derivative data set were collected on a rotating-anode generator.

The crystal structure of the complex formed between M·*Taq* I and AdoMet has been determined at 2.7-Å resolution by multiple isomorphous replacement including anomalous scattering data (MIRAS) (Table 1). Prior to interpretation, the MIRAS electron density was modified by solvent flattening and noncrystallographic symmetry averaging and finally was improved by phase refinement with iterative skeletonization [PRISM (18)]. The model was built on an Evans & Sutherland (Salt Lake City) workstation using the software package O (19) and refined by using the programs XPLOR (20) and TNT (21) (Table 2). A total of 74 water molecules were included and their parameters were refined. Omit-maps were used to check the AdoMet binding site and some loop regions. The two molecules in the asymmetric unit were treated independently.

Three regions (amino acid residues 1–21, 113–129, and 411–421), forming intermolecular contacts, have less-well-defined electron density and could consequently be interpreted only at half rms below the mean rms electron density, and there was no electron density for residues 19–20 and 115–117.

RESULTS AND DISCUSSION

There are two molecules in the asymmetric unit, and they are related by a noncrystallographic symmetry and show slight conformational differences, especially in loop regions and in the methionine part of AdoMet. The rms difference between the corresponding main-chain atoms is 1.1 Å (1.7 Å for all atoms); when the flexible regions (residues 1–20, 113–120, and 406–421) are omitted from the analysis, the rms difference of the main-chain atoms decreases to 0.6 Å (1.2 Å for all atoms).

For simplicity, only molecule I will be referred to in all further discussions.

The three-dimensional structure of M·*Taq* I is of the α/β-type. It is clearly divided into two domains with comparable size (Fig. 1). The domains are connected by a single short loop and arranged in the shape of a C with a cleft sufficiently wide to accommodate double-stranded DNA. The overall dimensions of the molecule are 80 × 55 × 40 Å.

The N-terminal domain (amino acid residues 1–243) is spherical (Fig. 2); its central part is formed by a nine-stranded β-sheet which is decorated on both sides by α-helices. The C-terminal domain (amino acid residues 244–421) is elongated in shape; at its N terminus, αF, αJ, and αK form a three-helix bundle.

The two domains are connected covalently by the loop between β9 in the N-terminal domain and αF in the C-terminal domain. The orientation between the two domains is determined by the insertion of the loop between αJ and αK

Table 2. Structure refinement

Resolution, Å	10.0–2.4
R factor, %	21.2
Observations [$F \geq 1\sigma(F)$]	30,150
rms Δbond lengths, Å	0.014
rms Δbond angles, °	2.42

R factor $= \Sigma|F_o - F_c|/\Sigma F_o$.

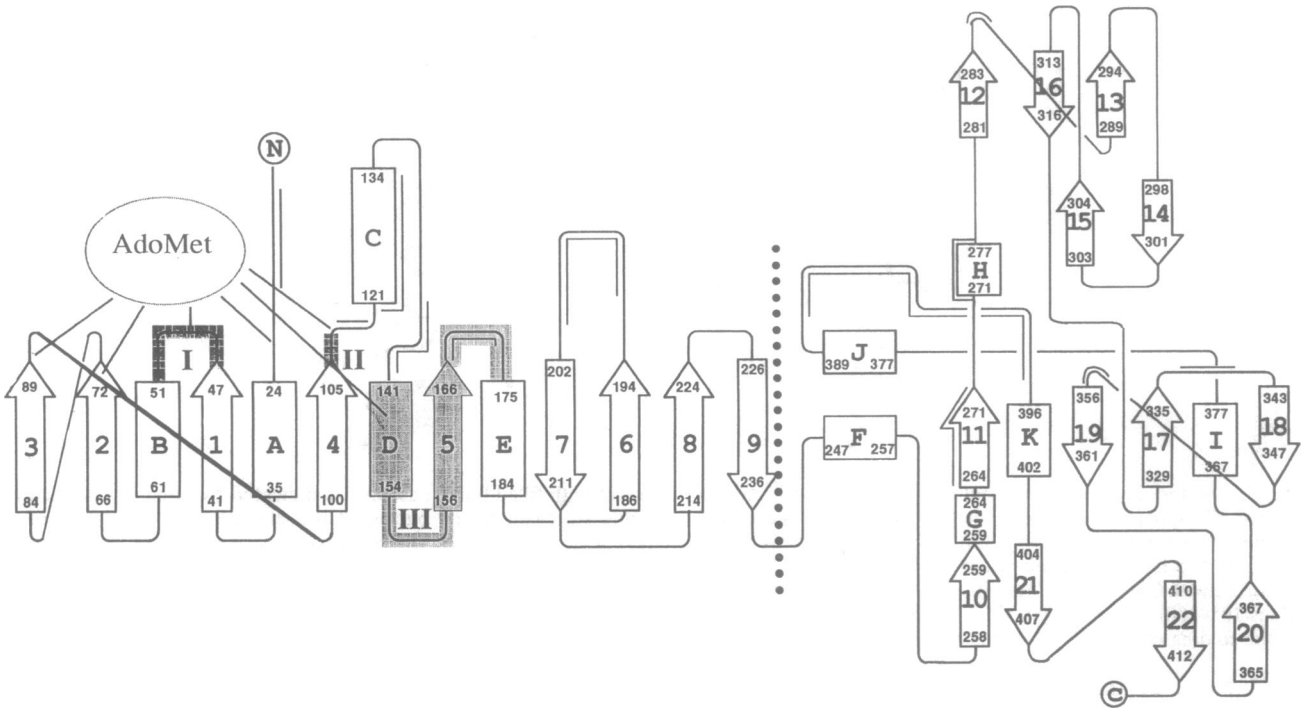

FIG. 1. Organization of the secondary structure elements of M·*Taq* I. The polypeptide is folded into two domains separated by the loop between β-strand 9 (β9) and α-helix F (αF) (dotted line). Small figures denote N-terminal and C-terminal amino acids in the β-strands (arrows numbered 1–22) and in the α-helices (rectangles labeled A–K). The AdoMet binding site is marked, and double lines in several loops, at αC, αH, and β11 indicate those segments of M·*Taq* I that could possibly interact with the DNA double helix. In adenine MTases there are two conserved segments, I (EPACAHG, amino acid residues 45–51, at the C terminus of β1) and II (NPPY, amino acid residues 105–108, at the C terminus of β4); in enzymes that are specific for TNNA there is another segment, III (amino acid residues 141–178 in αD-β5-αE); these three segments are outlined by shading and marked I, II, and III.

in the C-terminal domain into a wide cavity in the N-terminal domain that is formed by loops between β5-αE and β6-β7. The cohesive forces between the surfaces of these loops are by van der Waals contacts and seven hydrogen bonds involving Tyr[191], Ser[202], Asp[219], and Phe[245] and Leu[387], Arg[389], and Asp[390] in the N- and C-terminal domains, respec-

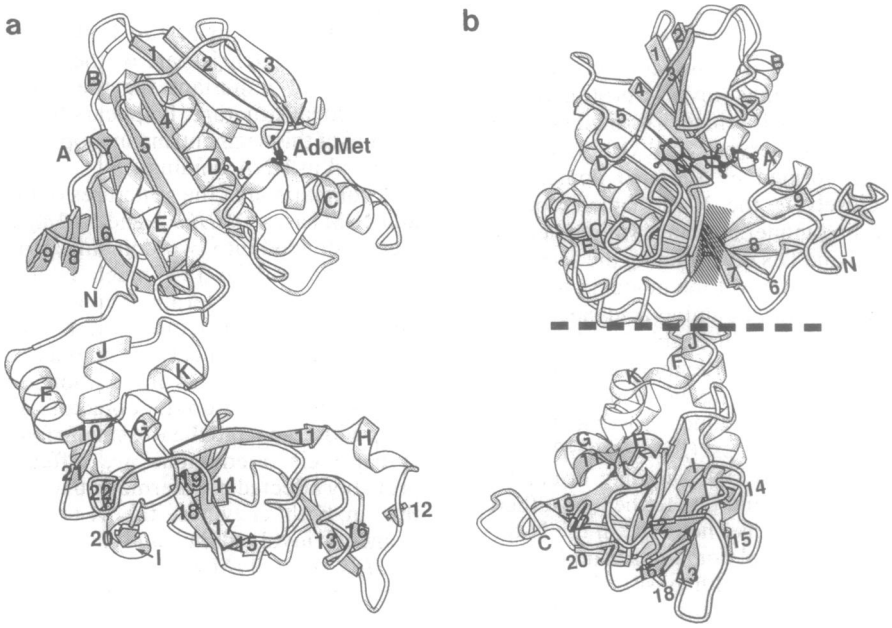

FIG. 2. Three-dimensional structure of M·*Taq* I in ribbon presentation. (*a*) View showing the folding of the polypeptide into two domains, N-terminal (upper) and C-terminal (lower), which form a cleft (to the right) wide enough to accommodate double-helical DNA. The N terminus is marked "N," the C terminus is above β17, and the connection between the domains is only through the loop between β9 and αF. The individual β-strands and α-helices are labeled according to Fig. 1; AdoMet is indicated. (*b*) Molecule rotated 90° around the vertical relative to *a*. The two domains are held in position relative to each other by noncovalent interactions between loop αJ-αK in the C-terminal domain and loops β5-αE and β6-β7 in the C-terminal domain. DNA would be bound with the helix axis about horizontal (marked by thick broken line); for methyl transfer the adenine in the sequence TCGA is proposed to unstack and to loop out into the pocket indicated by a shaded bar, between the loop segment N-terminal to αA and the loop β4-αC. Drawn with MOLSCRIPT (22).

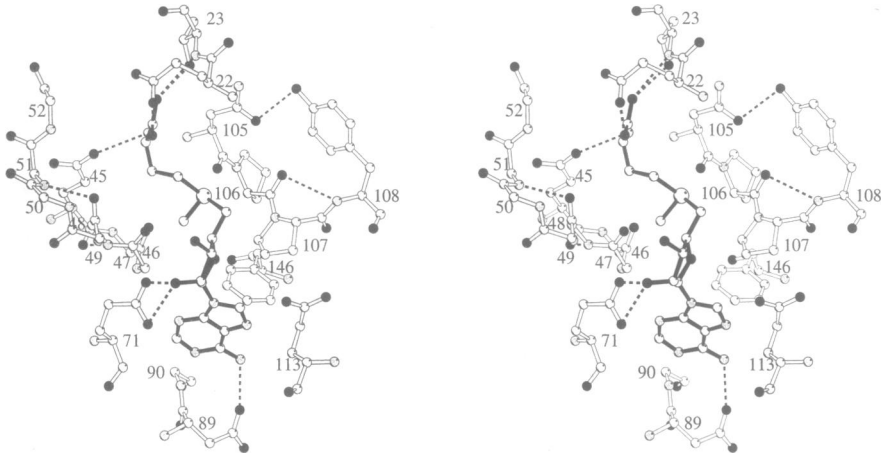

FIG. 3. Stereoview of the AdoMet binding to M·*Taq* I. AdoMet is drawn with filled bonds; atoms are indicated by open (C), shaded (N), and filled (O) spheres and a large shaded sphere (S); and hydrogen bonds are dashed. The pucker of the ribose is C3′-*endo* C4′-*exo*, the conformation of torsion angle O4′—C4′—C5′—S is +*gauche*, and adenine is in the *anti* position. The conserved segment I, EPACAHG, residues 45–51, is to the left (main-chain atoms only except for Glu^{45}) and at the N-terminal end of αB, which is part of a classical Rossmann fold (β1-αB-β2), with Glu^{71} at the C terminus of β2 in the correct position to hydrogen bond to O2′H of the nucleoside. Conserved segment II (NPPY, residues 105–108) is to the right. The side chains of Phe^{146} in conserved segment III and of Glu^{113} are in van der Waals contact with adenine. Note the folding of I and II, which are both stabilized by two internal hydrogen bonds each. Drawn with MOLSCRIPT (22).

tively.¶ None of these amino acids is conserved in N-MTases.

The cofactor AdoMet is inserted into a cavity in the N-terminal domain formed by the two conserved regions I and II (Fig. 3). I is at the C terminus of β1 and is part of a classical Rossmann fold (β1-αB-β2) (23). This fold requires the conserved Gly^{51} for the sharp bend and hydrogen bonding between ribose O2′H and the carboxylate of Glu^{71} (2.5 Å and 2.6 Å, respectively) at the C terminus of β2. The shape of the cavity is such that AdoMet is accommodated in an extended conformation (see legend of Fig. 3). Adenine is held in position by van der Waals contacts to the side chains of Glu^{113} and Phe^{146} and by hydrogen bonding between adenine N^6 and Asp^{89} $O^{\delta 2}$, 2.96 Å. The methionine carboxylate O^1 is in hydrogen bonding contact to Thr^{23} NH and O^{γ}, 3.12 Å and 3.16 Å, respectively, and there is short contact between O2 and Glu^{22} $O^{\epsilon 1}$, 2.5 Å, which must involve protonation of one of the two carboxylates. A salt bridge is formed between methionine $—NH_3^+$ and the conserved Glu^{45} $O^{\epsilon 1}$, 3.16 Å.

The width of the cleft formed by the two domains of M·*Taq* I corresponds closely to the 21-Å diameter of the B-DNA double helix. Using model building, we studied the complex between M·*Taq* I and DNA containing the sequence TCGA, but we did not try to interpret the model in terms of individual protein–DNA contacts, as amino acid side chains may rotate and the protein contacts probably involve the N terminus and some or all of the eight flexible loop segments indicated in Fig. 1. Upon DNA binding the conformation of the loops may be affected and even induce movement of the domains, because three of the loops (β5-αE, β6-β7, and αJ-αK) that might be involved in DNA binding orient the domains relative to each other. In addition to the loops, there are secondary structure elements like helix αC in the N-terminal domain and, in the C-terminal domain, strand β11 and helix αH. The loops are studded with positively charged amino acids Arg, Lys, and His, which are in positions to interact with DNA phosphate groups (Fig. 4).

A direct transfer of the methyl group from the M·*Taq* I–AdoMet complex onto the TCGA adenine residue in undistorted M·*Taq* I and B-DNA appears impossible, as the distance of approximately 15 Å between methyl donor and acceptor is too far, and there are no obvious side chains that would be capable of assisting in this reaction. Instead, it is more probable that the DNA is conformationally distorted so that the N^6 of adenine is brought so close to the cofactor that it can accept the CH_3^+ group directly. As reported for the cognate M·*Hha* I–DNA complex (24), this could be achieved by unstacking and "looping out" of the target adenosine in TCGA such that adenine is located in the gap (indicated by shading in Fig. 2) formed between the loop at the N terminus of αA and the loop β4-αC.

Since the conserved segment III (amino acid residues 141–178) was found only in TNNA-recognizing adenine MTases, it was thought to be implicated in binding of the target DNA (8). In M·*Taq* I, this segment is contained in αD-β5-αE, which is part of the interior of the N-terminal domain and important for the stabilization of its three-dimensional structure. It appears that only amino acids at the termini of this segment and in loops αD-β5 and β5-αE may be involved in DNA binding.

A comparison of the structures of the two prototypes of C- and N-MTases, M·*Hha* I and M·*Taq* I, shows that both are clearly divided into two domains. However, the N-terminal domain in M·*Hha* I is much larger than the C-terminal domain, and the C terminus folds back onto the N-terminal domain. In both structures, the N-terminal domain is dominated by a large β-sheet, but numbers and orientations of the strands are different, as are location, orientation, and number of α-helices. The AdoMet cofactors are in comparable positions with respect to the Rossmann folds (β1-αA-β2 in M·*Hha* I) but their orientations are different and, in M·*Hha* I, probably affected by intermolecular contacts (4). It is striking that in both enzymes, the N-terminal domains carry all elements involved in catalytic activity and contain most (M·*Hha* I) or all (M·*Taq* I) of the conserved segments. This suggests that their three-dimensional structures are more conserved than those of the C-terminal domains, which are more variable in their amino acid sequence, consistent with the construction of hybrid enzymes, where the C-terminal domains are exchanged without impeding enzymatic activity (25, 26).

¶Tyr^{191} O^{η}···$O^{\delta 1}$ Asp^{390}, 3.27 Å; Tyr^{191} O^{η}···$O^{\delta 2}$ Asp^{390}, 2.74 Å; Ser^{202} O^{γ}···$O^{\delta 2}$ Asp^{390}, 2.64 Å; Ser^{202} O^{γ}···O Asp^{390}, 3.52 Å; Asp^{219} $O^{\delta 1}$, $O^{\delta 2}$···Arg^{389} $N^{\eta 1}$, $N^{\eta 2}$, 2.95 Å, 2.98 Å; Phe^{245} NH···O Leu^{387}, 2.98 Å.

The technical assistance of Andreas Milde and Cynthia Reeves is gratefully acknowledged. This work was supported by the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 344 and

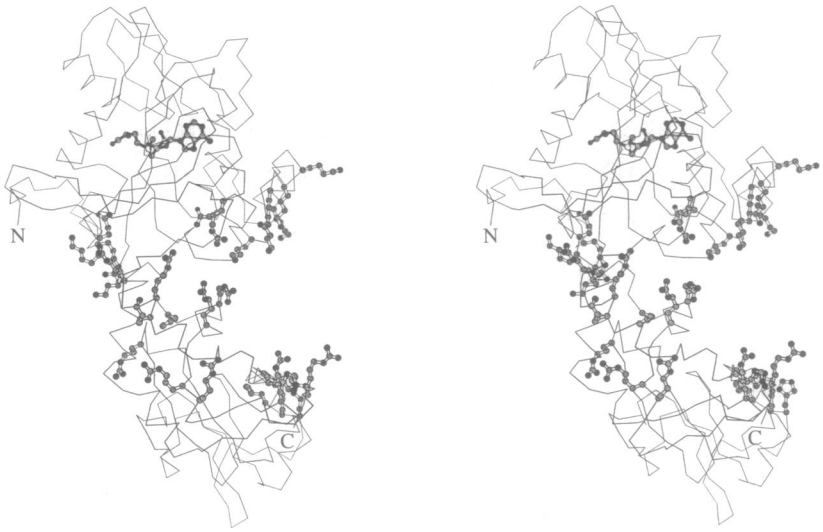

FIG. 4. Stereoview of a C^{α} model of M·*Taq* I. The cofactor AdoMet and the positively charged amino acid residues His, Lys, and Arg in the proposed DNA-binding cleft are drawn in ball and stick representation. Drawn with MOLSCRIPT (22).

Leibniz-Programm, by the Bundesministerium für Forschung und Technologie, and by the European Community Project SC1-CT90-0472(TSTS).

1. Anderson, J. E. (1993) *Curr. Opin. Struct. Biol.* **3,** 24–30.
2. Noyer-Weidner, M. & Trautner, T. A. (1993) in *DNA Methylation: Molecular Biology and Biological Significance*, eds. Jost, J. P. & Saluz, H. P. (Birkhäuser, Basel), pp. 39–108.
3. Tilghman, S. M. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 8761–8762.
4. Cheng, X., Kumar, S., Posfai, J., Pflugrath, J. W. & Roberts, R. J. (1993) *Cell* **74,** 299–307.
5. Posfai, J., Bhagwat, A. S., Posfai, G. & Roberts, R. J. (1989) *Nucleic Acids Res.* **17,** 2421–2435.
6. Klimasauskas, S., Timinskas, A., Mekevicius, S., Butkiene, D., Butkus, V. & Janulaitis, A. (1989) *Nucleic Acids Res.* **17,** 9823–9832.
7. Wilson, G. G. (1993) *Methods Enzymol.* **216,** 259–279.
8. Janulaitis, A., Vaisvila, R., Timinskas, A., Klimasauskas, S. & Butkus, V. (1992) *Nucleic Acids Res.* **20,** 6051–6056.
9. Wu, J. C. & Santi, D. V. (1987) *J. Biol. Chem.* **262,** 4778–4786.
10. Chen, L., MacMillan, A. M., Chang, W., Ezaz-Nikpay, K., Lane, W. S. & Verdine, G. L. (1991) *Biochemistry* **30,** 11018–11025.
11. Ho, D. K., Wu, J. C., Santi, D. V. & Floss, H. G. (1991) *Arch. Biochem. Biophys.* **284,** 264–269.
12. Pogolotti, A. L., Jr., Ono, A., Subramaniam, R. & Santi, D. V. (1988) *J. Biol. Chem.* **263,** 7461–7464.
13. Slatko, B. E., Benner, J. S., Jager-Quinton, T., Moran, L. S., Simcox, T. G., Van Cott, E. M. & Wilson, G. G. (1987) *Nucleic Acids Res.* **15,** 9781–9793.
14. Barany, F., Slatko, B., Danzitz, M., Cowburn, D., Schildkraut, I. & Wilson, G. G. (1992) *Gene* **112,** 91–95.
15. Jack, W. E., Greenough, L., Dorner, L. F., Xu, S.-Y., Strzelecka, T., Aggarwal, A. K. & Schildkraut, I. (1991) *Nucleic Acids Res.* **19,** 1825–1829.
16. Science and Engineering Research Council (1979) *CCP4*, Collaborative Computing Project No. 4 (Daresbury Lab., Sci. Eng. Res. Council, Warrington, U.K.).
17. Tickle, I. (1991) in *Isomorphous Replacement and Anomalous Scattering*, eds. Wolf, W., Evans, P. R. & Leslie, A. G. W. (Daresbury Lab., Sci. Eng. Res. Council, Warrington, U.K.), pp. 87–95.
18. Baker, D., Bystroff, C., Fletterick, R. J. & Agard, D. A. (1993) *Acta Crystallogr.* **D49,** 429–439.
19. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjelgaard, M. (1991) *Acta Crystallogr.* **A47,** 110–119.
20. Brünger, A. T., Kuriyan, J. & Karplus, M. (1987) *Science* **235,** 458–460.
21. Tronrud, D. E., Ten Eyck, L. F., Matthews, B. W. (1987) *Acta Crystallogr.* **A43,** 489–501.
22. Kraulis, J. P. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
23. Wierenga, R. K., Terpstra, P. & Hol, W. G. J. (1986) *J. Mol. Biol.* **187,** 101–107.
24. Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. (1994) *Cell* **76,** 357–369.
25. Mi, S. & Roberts, R. J. (1993) *Nucleic Acids Res.* **20,** 4811–4816.
26. Klimasauskas, S., Nelson, J. L. & Roberts, R. J. (1991) *Nucleic Acids Res.* **19,** 6183–6190.