Transition from Nonspecific to Specific DNA Interactions along the Substrate-Recognition Pathway of Dam Methyltransferase

John R. Horton1,4, Kirsten Liebert2,4, Stanley Hattman3, Albert Jeltsch2,*, and Xiaodong Cheng1,*

1Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, Georgia 30322
2School of Engineering and Science, International University Bremen, Campus Ring 1, 28759 Bremen, Germany
3Department of Biology, University of Rochester, Rochester, New York 14627

Summary

DNA methyltransferases methylate target bases within specific nucleotide sequences. Three structures are described for bacteriophage T4 DNA-adenine methyltransferase (T4Dam) in ternary complexes with partially and fully specific DNA and a methyl-donor analog. We also report the effects of substitutions in the related \textit{Escherichia coli} DNA methyltransferase (EcoDam), altering residues corresponding to those involved in specific interaction with the canonical GATC target sequence in T4Dam. We have identified two types of protein-DNA interactions: discriminatory contacts, which stabilize the transition state and accelerate methylation of the cognate site, and anti-discriminatory contacts, which do not significantly affect methylation of the cognate site but disfavor activity at noncognate sites. These structures illustrate the transition in enzyme-DNA interaction from nonspecific to specific interaction, suggesting that there is a temporal order for formation of specific contacts.

Introduction

Many DNA binding proteins are noncatalytic and exert their effects by binding at appropriate locations on the double helix. A solution structure of the \textit{E. coli} lac repressor DNA binding domain bound to nonspecific DNA revealed that the same set of protein residues can switch from an electrostatic interaction with the DNA backbone in a nonspecific complex to a specific binding mode with DNA base pairs in the cognate operator sequence (Kalodimos et al., 2004). Nonspecific DNA-protein interactions may accelerate target-site encounters in vivo by permitting the protein to sample many such sites in each binding event within a sliding range of ~100 base pairs before dissociating from the DNA (Halford and Marko, 2004). Proteins such as DNA methyltransferases (MTases) recognize specific nucleotide sequences with the added constraint of having to bring catalytic side chains and the target nucleotide together. In a process termed “base flipping,” DNA MTases rotate the target nucleoside ~180° along the
flanking phosphodiester bonds such that the flipped nucleoside projects into the catalytic pocket (Klimasauskas et al., 1994).

Dam DNA-adenine MTases, which methylate the exocyclic amino nitrogen (N6) of Ade in GATC (Hattman et al., 1978; Lacks and Greenberg, 1977), are widespread among enteric bacteria and some of their bacteriophages (Hattman and Malygin, 2004). Dam methylation at GATC sites plays pivotal roles in bacterial and phage gene expression (Hattman, 1999; Hernday et al., 2003; Julio et al., 2002; Løbner-Olesen et al., 2003; Oshima et al., 2002), DNA replication (Messer and Noyer-Weidner, 1988), mismatch repair (Modrich, 1989; Yang, 2000), phase variation (Hernday et al., 2003), and bacterial virulence among Gram-negative bacteria (Garcia-Del Portillo et al., 1999; Heithoff et al., 1999). Dam methylation is essential for the virulence of a growing list of bacterial pathogens, including *Salmonella typhimurium* (Garcia-Del Portillo et al., 1999; Heithoff et al., 1999; Heithoff et al., 2001), *Neisseria meningitidis* (Bucci et al., 1999), *E. coli* (Krabbe et al., 2000), *Yersinia pseudotuberculosis* (Julio et al., 2001), *Vibrio cholerae* (Julio et al., 2001), *Pasteurella multocida* (Chen et al., 2003), and *Haemophilus influenzae* (Watson et al., 2004). *Salmonella* mutants that lack Dam activity are attenuated for virulence in mice and confer protection against murine typhoid fever (Dueger et al., 2001; Dueger et al., 2003), and inactivation of Dam attenuates *H. influenzae* virulence (Watson et al., 2004). These observations raise the possibility that Dam inhibitors might have broad antimicrobial action.

Bacteriophage T4Dam contains two domains (Yang et al., 2003): a catalytic domain that binds the methyl donor S-adenosyl-L-methionine (AdoMet) and a DNA binding domain that contains a β hairpin loop (residues 110–130 in T4Dam), which is conserved in the family of GATC-related MTase orthologs (Lauster et al., 1987). T4Dam, like EcoDam (Urig et al., 2002), modifies DNA processively, i.e., repetitively continuing catalytic function by exchanging methylation product S-adenosyl-L-homocysteine (AdoHcy) for AdoMet without dissociating from the DNA (Zinoviev et al., 2003). In a ternary complex with AdoHcy and a 12 bp synthetic DNA duplex, T4Dam interacts only with the sugar-phosphate backbone despite the presence of a central GATC sequence (Yang et al., 2003). Here we report three crystal structures of T4Dam bound to longer DNA substrates. These structures afford snapshots that suggest a possible mechanism for a pathway of stepwise recognition of the specific target sequence. Residues were identified that mediate nonspecific or sequence-specific contacts. Two Arg residues (R130 and R116) can switch roles from a purely electrostatic interaction with DNA phosphates in the nonspecific complex to a highly specific binding mode with base pairs in the specific or semispecific complex. On the other hand, the phosphate-interacting residues R95 and N118 in the specific complex are not involved in any DNA interaction in the nonspecific complex. We also investigated whether the information derived from T4Dam structure has relevance to understanding the family of bacterial Dam MTase orthologs. To this end we studied the effect of site-directed mutations in EcoDam altered at residues that are conserved within the Dam family and are known to be involved in specific interaction with the GATC target site in T4Dam. Three EcoDam variants showed changes in target-sequence specificity. R124A had an overall reduction in catalytic activity but methylated two noncanonical sequences (GATT and GATG) faster than the canonical GATC. In contrast, variants P134A and P134G retained full enzymatic activity on GATC but gained the ability to methylate the noncanonical sequences GACC and GAAC. In addition, while the L122A variant had slightly reduced activity, it showed a dramatic increase in specificity due to a loss in ability to methylate noncanonical sites. These strong effects correlate with the central role of these residues in the enzyme-DNA interface.
Results

T4Dam Complexed with a 13-mer Oligonucleotide: Nonspecific Interactions and the 1/4-Site Recognition

We crystallized a ternary complex containing T4Dam, AdoHcy, and a 13-mer oligonucleotide containing a single GATC target site. The DNA duplex contains a 5′-overhanging Ade in one strand and a 5′-overhanging Thy in the other strand such that the Ade and Thy would form a base pair at the joint of two molecules if the DNA duplexes were stacked head to tail. Unexpectedly, the helical axes of the two DNA molecules in the crystal were shifted relative to one another by ~12 Å (Figure 1A). In the crystallographic asymmetric unit, two Dam monomers (molecules A and B) bind each DNA duplex (Figure 1A). Dam molecule A primarily binds to a single DNA duplex spanning eight base pairs. It contacts two phosphates 5′ to each GATC in both strands via hydrogen bonds or electrostatic interactions. These are mediated to one strand by the side chains of Q12 and S13 in the N-terminal loop and by R130 and N133 with the complementary DNA strand (Figure 1B).

Dam molecule B binds the joint between the two DNA duplexes. In addition to phosphate interactions, which span 12 base pairs between the duplexes, R116 of molecule B makes bifurcated hydrogen bonding interactions with Gua in the G:C base pair at position 3 (Figures 1C and 1D). Unexpectedly, the next potential G:C base pair (position 2) is unpaired, along with the 5′-overhanging terminal Ade (position 1). The 5′-overhanging Thy of the next DNA molecule approaches, becomes extrahelical, and stacks with the Cyt of the G:C base pair (at position 3), while the phenyl ring of F111 stacks on the other side (Figure 1C). The carbonyl O4 atom of the Thy makes a van der Waals contact with M114, while its methyl group is in contact with P126 (Figure 1D). The amino acid residues involved in basespecific interactions (F111, M114, R116, and P126) and with the two phosphates 5′ to the Gua (R91 and N118) are invariant in the Dam family of GATC MTases (see Figure 1 of Yang et al., 2003). It seems as if T4Dam traps the sequence at the joint in a conformation that mimics part (~1/4) of the recognition sequence.

T4 Dam Complexed with a 15-mer Oligonucleotide

We designed a self-complementary 15-mer oligonucleotide with the end sequence of the duplex representing part of the GATC target sequence (see Figure 2A). In addition, we reduced the ratio of protein to DNA to approximately half of that previously used because we had observed two Dam molecules bound per DNA duplex. The new ternary complex of T4Dam, DNA, and AdoHcy crystallized by forming three layers of a large enzyme-substrate network (Figure S1). The enzyme-substrate complexes along one layer contain four DNA duplexes (blue, green, magenta, and orange) and five T4Dam molecules (Figure 2A).

We observed three basic features of the ternary-complex layers. First, all of the joints between neighboring DNA duplexes are occupied by a Dam molecule (Figure 2A). Three of these Dam molecules have very similar protein-DNA interactions (the 3/4-site recognition), so we designated all three of them as molecule C (to distinguish from the A and B molecules shown in Figure 1). These occupy the joints between DNA duplexes, which are shown in Figure 2A as blue and green, green and magenta, and orange and blue (of the next set of four), respectively. Second, only one Dam (molecule E in Figure 2A) is bound to the specific GATC site in the middle of the orange DNA, exhibiting full-site recognition in what appears to be a catalytically competent conformation. Third, molecule D occupies the joint between magenta and orange DNA and interacts with a noncanonical site (see below).

Cell. Author manuscript; available in PMC 2009 March 17.
The 3/4-Site Recognition

The adjacent green and blue DNA duplexes are stacked head to tail, aided by F111 stacking with 5′ Thy bases from the two duplexes (Figure 2B). There are additional base-specific interactions between the enzyme and the DNA in the joint of the two duplexes. As shown in Figure 2C, R116, P126, and M114 interact with one half-site (green duplex of Figure 2B), while S112 and R130 interact with the second half-site (blue in Figure 2B). The side chain guanidino groups of R116 and R130 form the same bifurcated hydrogen bonds with the N7 and O6 atoms of the two 5′ Gua bases, respectively. The two overlapping Thy bases differ in their interaction with T4Dam. M114 and P126 make van der Waals contacts with the methyl group and O4 atom of the green Thy base while leaving it available to make Watson-Crick hydrogen bonds with a complementary Ade base. In contrast, S112 makes two hydrogen bonds with the Watson-Crick edge of the blue Thy base, whose methyl group interacts with the Ca atom of conserved G128.

These observations indicate that T4Dam preferentially binds at the joint of two duplexes, which mimics damaged DNA or altered recognition sites. This is surprising but consistent with biochemical data that suggest that binding specificity for DNA MTases is determined by the nucleotides flanking the target nucleotide and that DNA MTases bind more tightly to substrates containing mismatches at the target base (Cheng and Roberts, 2001; Klimasauskas and Roberts, 1995). In other words, DNA MTases do not depend on the flipped target base for recognition-site interaction. Presence of only one-half of the recognition site on one strand was sufficient for stable complex formation with T4Dam provided that the 5′ G:C base pairs were present at both ends of the palindrome (Hattman and Malygin, 2004). This is exactly what we observed for the joint formed by the 3/4-site complex.

Full-Site Recognition Involves a Protein Side Chain Intercalation

Only one T4Dam (molecule E) occupies a GATC site (orange DNA) (Figure 3A). The β hairpin makes nearly the same specific interactions with DNA bases in the major groove as observed on the 3/4 site. F111 and S122 both insert their side chains into the DNA from the major-groove side (Figure 3A). Although the target Ade is flipped out of the duplex, its electron density was not very well ordered in the active site (see below for details of active-site interactions). The side chain of S112 occupies the space left by the flipped Ade, forming two hydrogen bonds with the “orphaned” Thy, similar to that observed in the 3/4-site complex. This S112 interaction restores hydrogen bonding to the polar edge of the orphaned Thy and replaces its stacking to the flanking base pairs (Figure 3A). The Thy-S112 interaction is similar to other protein-side-chain-orphaned base interactions such as those for Gua-Q237 of DNA-cytosine MTase HhaI (Klimasauskas et al., 1994), Thy-Y162 of human 3-methyladenine DNA glycosylase (Lau et al., 1998), and Cyt-N149 of human 8-oxoguanine DNA glycosylase (Bruner et al., 2000).

The phenyl ring of F111 intercalates into the DNA helix and stacks between the adjacent A:T base pair and the Thy:S112 “base-amino acid” pair, resulting in a local doubling in helical rise (Figure 3A). The intercalation of amino acids between DNA base pairs from the major-groove side has been described for several protein-DNA complexes. In the M.HaeIII-DNA complex, Ile221 lies between the stacked bases and opens a gap in the DNA so that the orphaned Gua pairs with an adjacent Cyt (Reinisch et al., 1995). In the very short patch-repair endonuclease-DNA complex, three aromatic residues intercalate into the DNA next to the TG mismatch (Tsutakawa et al., 1999). In the HincII restriction endonuclease-DNA complex, a Gln side chain intercalates between two base pairs on either side of the recognition site (Horton et al., 2002). In addition, intercalation by the repair enzyme formamido-pyrimidine-DNA glycosylase, in which the F111-M75 residue pair is stacked between the A:T base pair and the base-amino acid pair Cyt:R109, has been observed from the minor-groove side of DNA (Serre et al., 2002).
Interaction with a Noncanonical Site

F111 intercalation by molecule E into the central AT stacking effectively causes a one-base-pair lengthening of the DNA molecule depicted in orange (Figures 4A and 4B). The expansion is propagated toward one end of the DNA molecule, resulting in two disordered nucleotides of the neighboring duplex (magenta). The 5′-overhanging Thy of the magenta DNA is pushed out and apparently becomes disordered, resulting in the Cyt of the next base pair stacking with F111 of Dam molecule D. The side chain of S112 approaches the Cyt base with the side chain hydroxyl oxygen and the exocyclic amino nitrogen N4 of the Cyt at a van der Waals distance, partly because of repulsion force between the N4 amino nitrogen (NH₂) and the main chain amide nitrogen (NH) (Figure 4C). The interaction between S112 and Cyt is sufficient to displace the complementary Gua and make it disordered. The side chain of R130 skips the next A:T base pair and interacts with the Gua of the adjacent downstream G:C base pair (Figure 4C). Since the presence of a Gua downstream of a GATC (or modified TATC) site does not support catalysis (data not shown), we assume this complex exemplifies the interaction of T4Dam with an isolated TC dinucleotide site in the DNA, which does not lead to DNA methylation.

Stabilization of the Flipped Adenine in the Presence of Sinefungin

Thus far we had prepared ternary complexes using the methylation reaction product AdoHcy. The protein-AdoHcy interactions for each protamer are nearly identical to those described in the T4Dam-AdoHcy binary complex (Yang et al., 2003). In the full-site recognition complex between Dam molecule E and the orange DNA (Figure 3A), the target Ade is flipped out but not fully ordered in the active site. We reasoned that product AdoHcy might signal the enzyme to release from the target site in order to exchange for AdoMet prior to the next methyl transfer. Thus, stable binding of the flipped Ade in the active-site pocket probably requires Ado-Met, as has been suggested for EcoDam (Liebert et al., 2004). Therefore, we used the AdoMet analog sinefungin (adenosyl ornithine) to prepare a new ternary complex because it also carries a formal positive charge on the ε amino group (Figure 3B).

The new crystal contains two T4Dam molecules (not shown), one bound in the joint of two DNA duplexes, similar to the Dam C molecules in Figure 2B, and the other bound to the specific GATC site in the middle of one duplex, similar to the Dam E molecule in Figure 3A. The flipped Ade is surrounded (via hydrogen bonds, π stacking, and hydrophobic interactions) by amino acids belonging to the conserved catalytic D171-P-P-Y174 motif (Malone et al., 1995), Y181, K11, and sinefungin (Figure 3C). The Ade N6-amino group that becomes methylated forms a pair of hydrogen bonds; one is to the side chain of D171, and the other is to the backbone carbonyl oxygen between the two proline residues P172 and P173. The target amino nitrogen is at a distance of less than 3 Å away from the sinefungin ε amino group, which is out of the plane of the constrained Ade base. This structural arrangement suggests that the target nitrogen lone pair is deconjugated and positioned for an inline direct methyl-group transfer as suggested for the TaqI DNA-adenine MTase (Goedecke et al., 2001). The ε amino group of sinefungin forms a hydrogen bond with the hydroxyl of Y181, which in turn interacts with the main chain carbonyl of T8. The opposite face of the flipped Ade is in a face-to-face π stacking with the aromatic ring of Y174.

Biochemical Analysis of EcoDam Variants

EcoDam has considerable sequence similarity (25% identity) to T4Dam (Hattman et al., 1985) but has significantly higher sequence conservation with Dam enzymes from pathogenic bacteria. For example, the E. coli and S. typhimurium Dam proteins are 92% identical (differing at only 22 of 278 residues) and have no gaps in their alignment. Because of the biological importance of the Dam family, we investigated whether the T4Dam structures contribute to understanding the function of these orthologs. To this end we studied the effects of substituting...
Ala for residues in EcoDam (Figure 5A) that correspond to those involved in T4Dam-specific interaction with its target GATC site. This includes Y119 (F111 in T4Dam), N120 (S112 in T4Dam), L122 (M114 in T4Dam), R124 (R116 in T4Dam), and P134 (P126 in T4Dam). All of these residues are highly conserved among Dam orthologs. In addition, we mutated residues R137, Y138, and K139 since these could assume the function of T4Dam Arg130 (Figure 5A).

The R124A and Y119A variants were the most strongly affected by the Ala substitution; their catalytic activity was reduced more than 100-fold (Figure 5B, and see T4Dam R116 and F111 in Figure 2B). N120A, N120S, and L122A were affected only slightly. DNA binding by the R124A variant was reduced 10-fold (accounting for only one-tenth of the drop in catalytic activity), while binding of Y119A, P134A, P134G, and K139A was reduced 2- to 3-fold (Figure 5C). The other variants (N120A, N120S, L122A, R137A, Y138A, and K139A) did not display any appreciable difference in DNA binding compared to the wild-type.

To further investigate the process of DNA recognition, the rate of DNA methylation by the wild-type and variant enzymes was determined using duplexes containing a single hemimethylated target (N6-methyl-Ade in the bottom strand, third base pair in Figure 5A). This ensured that only one strand of the DNA was subject to methylation (i.e., the Ade of the top strand, second base pair in Figure 5A). The duplexes contained the canonical GATC site or a variant with a single base substitution at either the first, third, or fourth base pair of the target sequence (see Figure 5A); these variant sites are designated here as “near-cognate” sites (a total of nine). In this fashion, a specificity profile of the wild-type enzyme and its variants was obtained (Figure 6). Wild-type EcoDam is a very specific enzyme because near-cognate sites were modified 100- to 1000-fold more slowly than the cognate site (Figure 6A). The first position of the GATC sequence is recognized less accurately than the third and fourth base, in agreement with earlier findings (Liebert et al., 2004). It is interesting that the contact by R130 of T4Dam to this base is not well conserved among other members of the Dam family (e.g., substituted by Y in EcoDam; see Figure 5A) and that the R130-Gua1 contact is not yet formed in the 1/4-site complex (compare Figures 1C and 2B). EcoDam variants altered at residues that might be involved in the recognition of the first base pair (R137A, Y138A, and K139A) did not exhibit any strong changes in methylation activity or specificity (Figure S2).

In contrast to the first position, the third and fourth bases of GATC are recognized more accurately. At both positions, transitions (Thy3 to Cyt or Cyt4 to Thy) are less deleterious than transversions, indicating that conservative exchanges are more tolerable. The contact between T4Dam R116 and Gua4 (Figure 2C) is conserved among Dam MTases (e.g., Figure 5A). We determined the specificity profile of the corresponding Eco-Dam R124A variant (Figure 6B). In agreement with the T4Dam structure, GATG and GATT sites were methylated by R124A faster than the canonical GATC site. In contrast, wild-type EcoDam methylation of these two near-cognate sites was three orders of magnitude slower than methylation of GATC. Thus, while R124A has a 100-fold-reduced rate of DNA methylation at GATC sites relative to wild-type EcoDam, it methylated GATG and GATT sites 2- to 3-fold faster than GATC and 30- to 40-fold faster than the wt enzyme modified GATG or GATT. Therefore, R124A has lost the discriminatory requirement for a C:G base pair at the fourth position of GATC. In order to analyze this information more quantitatively, we have defined a specificity factor by integrating the relative methylation activities at all near-cognate sites (Experimental Procedures). A comparison of specificity factors for the recognition of position 4 (S4) reveals that the R124A variant has an 8000-fold-changed relative preference for methylation of near-cognate sites modified at the fourth position (Figure 6F). No other variant showed such a strong change in S4. Furthermore, the R124A variant retained (or even increased) its specificity for the first and third positions in GATC, so this is a base pair-specific change (Figure 6B).

*Cell. Author manuscript; available in PMC 2009 March 17.*
We found a similar base pair-specific loss of specificity associated with T4Dam residues P126 and M114, which recognize the T:A base pair at the third position of GATC (see Figure 5A). Naturally occurring variant phage enzymes (T2Dam<sup>h</sup> and T4Dam<sup>h</sup>), which efficiently modify GACC sites in addition to the canonical GATC site (Brooks and Hattman, 1978), contain a P126S substitution (Miner et al., 1989). In addition, P126G, -A, or -C substitutions behaved in a Dam<sup>h</sup>-like fashion (Miner et al., 1989). In this regard, it is perhaps not surprising that EcoDam P134A and P134G variants had normal catalytic activity at GATC sites (Figure 5B). However, P134A exhibited a significant increase in methylation rates of GA<sub>AC</sub> and GA<sub>CC</sub> substrates, with GACC being modified at almost the same rate as canonical GATC (Figure 6C). This change in preference corresponds to a more than 100-fold loss in sequence discrimination at the third base when compared to wild-type EcoDam. Further shortening of the side chain of P134 to glycine eliminated discrimination between GAAC and GACC, which were methylated at a rate about 10-fold lower than GATC (Figure 6D). The change in P134A and P134G recognition of the third base pair is illustrated in Figure 6G, where the specificity factor for recognition of the third base (S<sub>3</sub>) is compared for all variants. This ratio is shifted 1200- to 1500-fold in comparison to wild-type EcoDam. However, these changes do not result in simple loss of specificity at the third position: GA<sub>TC</sub> sites are still preferred about 10-fold relative to GA(A/C)C, while GAGC sites are methylated at least 1000-fold more slowly.

The activity of the L122A variant of EcoDam (M114 in T4Dam; Figures 2B and 2C) is not appreciably reduced (Figure 5B). Intriguingly, however, no methylation activity was detectable for any of the near-cognate sites (Figure 6E). This indicates that the L122A variant has a significantly improved specificity (Figure 6H). This can be rationalized by assuming that the side chain of L122 is required to stabilize the whole protein-DNA interface. Whereas the L122A change alone does not severely reduce catalytic activity on the normal GATC substrate, a combination of L122A with the change of any of the base pairs in the recognition site may disturb synergistically the protein-DNA interface, and this could explain the complete loss of activity.

In addition to residues making base-specific contacts, we studied the aromatic residue that intercalates into the DNA (Y119 in EcoDam, F111 in T4Dam; Figure 3A) and the adjacent hydrophilic residue that contacts the orphaned Thy (N120 in EcoDam, S112 in T4Dam; Figure 3A). As shown in Figure 5B, Y119A was the second-most-affected variant. This suggests that intercalation of the aromatic ring into the DNA is an important step in enzyme catalysis, possibly involved in initiating or stabilizing base flipping. In contrast, removal of the side chain of N120 (N120A) had only a minor effect on methylation rate, although the structure of the specific T4Dam-DNA complex suggests an important role for this amino acid in base flipping. This finding is consistent with the fact that base flipping is fast and not rate limiting in catalysis for EcoDam and other DNA MTases (Allan et al., 1999; Beck and Jeltsch, 2002; Liebert et al., 2004).

**Discussion**

DNA recognition by proteins is essential for specific expression of genes in any living organism. Although the principle of proteins recognizing DNA sequences by contacts in the major groove has been known for decades (Seeman et al., 1976), there is no general code allowing one to deduce amino acid motifs from their target DNA sequences. Notable exceptions are the C2H2-type zinc fingers, where the DNA recognition process is sufficiently understood to define a DNA recognition code of this family of proteins (Pabo et al., 2001). Consequently, the rational design not only of DNA-interacting enzymes but also of even noncatalytic proteins is still in its infancy.
Here we describe six unique T4Dam-DNA interactions along the substrate-recognition pathway (Figure 7). Surprisingly, both protein and DNA components undergo very little overall conformational change upon binding. The protein structures of the nonspecific, semispecific, and specific complexes can be superimposed within 0.4–0.8 Å of root-mean-square deviation with that of the binary T4Dam-AdoHcy complex (PDB code 1Q0S). The DNA component is primarily in the B form, except for the one-base-pair expansion caused by F111 intercalation and the flipping of the target nucleoside out of the DNA helix and into the enzyme’s catalytic pocket in the specific complex. However, three prominent orientations of T4Dam relative to the DNA helical axis were observed. The β hairpin loop, whose axis is defined in parallel to the β strands forming the hairpin, sits almost perpendicular (~80°) to the DNA axis in the nonspecific complexes (Figure 7A), where R130 forms one phosphate contact. Nonspecific interactions also occur in the DNA minor groove (Figure 7B), where the protein tilts to ~45° relative to the DNA and the second Arg of the hairpin loop (R116) forms one of the phosphate interactions. Direct interactions with bases occur only in the DNA major groove (Figures 7C-7F), where the angle between the axes of the β hairpin and DNA is ~30° in the 1/2-site complex and ~25° in the 3/4-site and full-site complexes. These results suggest that T4Dam moves along the DNA and rotates up and down as a rigid body relative to the DNA.

Interestingly, the phosphate-interacting residues R95 and N118, which hydrogen bond with the first and second phosphates 5’ to the Gua4 in the specific complex (or in any complex involving R116-Gua4 interaction; Figures 7C-7F), are not involved in any DNA interaction in the nonspecific complexes (Figures 7A and 7B). Two different pairs of residues (Q12 and S13, and R130 and N133) interact with the two phosphates 5’ to each Gua of the GATC palindrome in the nonspecific complex (see Figure 1B). In contrast, two Arg residues (R130 and R116) can switch roles from a purely electrostatic interaction with the DNA phosphate in the nonspecific complexes (Figures 7A and 7B) to a highly specific binding mode with base pairs of the specific or semi-specific complexes (Figures 7C-7F). A similar switch in interaction with DNA was observed for the residue R22 of E. coli lac repressor (Kalodimos et al., 2004). This switch effectively reorients T4Dam, thereby positioning the enzyme’s active-site pocket to accommodate the flipped target base. After catalysis, the enzyme moves away from the target site and rotates back into the perpendicular orientation, exposing the active site to solvent and allowing AdoHcy to exchange for AdoMet. This mechanism would ensure that base flipping and methyl transfer specifically occur in a complex with cognate GATC sites and that AdoHcy/AdoMet exchange is possible after each turnover without dissociation from the DNA.

Our data suggest a temporal order for the formation of specific contacts during the one-dimensional sliding of T4Dam along the DNA. The contact of R116 to the fourth base pair of the GATC site is observed in the 1/4- and 3/4-site recognition complexes. Next, the contacts of P126 and M114 to the third base pair are formed. All of these residues are strictly conserved within the Dam MTase family. The contact of R130 to Gua1 that is specific to T4Dam is formed later. This result agrees with a similar conclusion drawn from rapid kinetics experiments with M.EcoRV variants (Beck and Jeltsch, 2002). In this enzyme, substitutions of amino acids conserved in the enzyme family (such as N136A) interfered with specific complex formation at an early state, while substitutions of amino acids characteristic of EcoRV (such as R145A) interfered with complex formation at later stages. This finding might illustrate a general pathway for changes of DNA specificity of proteins and enzymes during molecular evolution. The recent study on human DNA-repair protein O6-alkylguanine-DNA alkyltransferase (AGT) suggested that the recruitment of multiple AGT molecules to the same region of DNA might aid the search for DNA damage through a process of directional bias (Daniels et al., 2004). However, such directional bias was only observed for the repair of single-stranded DNA by AGT but not for double-stranded DNA, and the system cannot be directly compared to the Dam MTases because T4Dam and EcoDam move along double-stranded DNA (Figure 2A), whereas AGT forms polymers.
We analyzed the biochemical effects of altering the contacts described above in double-mutant cycles (Fersht et al., 1992). This involved shortening the respective amino acid side chains and using DNA substrates with near-cognate sites. We found that it was possible to predictably design MTase variants that no longer recognize one specific base pair within their recognition site. The EcoDam R124A variant displayed a change in specificity because it had a significantly higher catalytic activity toward a near-cognate site. In addition, the EcoDam P134A variant (the analog of the T4Dam<sup>b</sup> MTase) methylated a near-cognate site at almost the same rate as wild-type EcoDam modified the canonical site, indicating a broadened specificity (Figure 6).

In this work we have identified two types of protein DNA contacts, discriminatory and antidiscriminatory. A discriminatory contact is one that stabilizes the transition state of enzymatic catalysis and specifically accelerates the reaction with the cognate site. The contact between R116 of T4Dam (R124 of EcoDam) and the Gua<sub>4</sub> is an example of a discriminatory contact. Disruption of the contact by removal of the amino acid side chain led to a strongly reduced activity of the enzyme variant. An antidiscriminatory contact, e.g., the contact between P126 of T4Dam (P134 of EcoDam) and the third base pair of the recognition site, is one that does not significantly accelerate the reaction with the cognate site but disfavors activity at near-cognate sites because steric clashes may occur if the wrong DNA sequence is bound. This would strongly interfere with methylation of most noncanonical DNA sequences and lead to an efficient counterselection against methylation of nontarget sites. This is illustrated by the high activity and broadened specificity of EcoDam variants P134A and P134G.

**Experimental Procedures**

**Crystallography**

T4Dam was expressed and purified as described previously (Kossykh et al., 1995; Yang et al., 2003). AdoHcy or sinefungin was added at a 3:1 molar excess before concentration. The protein was concentrated to ~60 mg/ml and later incubated with annealed oligonucleotide at various protein/DNA ratios for at least 2 hr on ice before setup of crystallization trials.

Three data sets were collected (Table S1). Data set 1 (PDB code 1YF3) was collected from a crystal containing the 13-mer oligonucleotide (with a protein/DNA ratio of ~2:1) and AdoHcy, which was grown with a reservoir solution of 20% PEG MME 5000, 100 mM citrate-phosphate (pH 6.4), and 30 mM ammonium sulfate. Data set 2 (PDB code 1YFJ) was collected from an orthorhombic-form crystal containing the 15-mer oligonucleotide (with a protein/DNA ratio of ~1:1) and AdoHcy, which was grown under 16% PEG 6000, 100 mM MES (pH 6.0), 200 mM ammonium acetate, 10 mM CaCl<sub>2</sub>, and 10% ethylene glycol. Crystals containing the 15-mer oligonucleotide grew in several forms within the same drop. Varying the protein/DNA ratio and concentration of PEG 6000 could control crystal growth so that generally only one form appeared. Data set 3 (PDB code 1YFL) was collected from a crystal grown with a protein/DNA ratio of ~2:1 using the blunt-end 16-mer DNA and sinefungin under 7% PEG 6000, 100 mM MES (pH 6.2), 200 mM ammonium acetate, 10 mM CaCl<sub>2</sub>, and 10% glycerol.

**Biochemical Experiments**

The sequence of the 20-mer oligonucleotide substrate was 5′-GCG

ACAGTGATCCGCTGTCC-3′/5′-GACAGCCCGTACTGTGCC-3′,

where M is N6-methyl-Ade. In addition, nine substrates were used that contain near-cognate sites, differing in one base pair from GATC at the first, third, or fourth position. Oligonucleotides were purchased from MWG (Ebersbach, München) in purified form. Site-directed mutagenesis was performed as described (Liebert et al., 2004). Wild-type EcoDam and its variants were expressed in *E. coli* HMS174 cells as His<sub>6</sub>-fusion proteins and purified using one-column chromatography (Qiagen Ni-NTA) essentially as described previously (Urig et al., 2002).
DNA methylation was analyzed in 50 mM HEPES (pH 7.5), 50 mM NaCl, 1 mM EDTA, 0.5 mM DTT, 0.2 g/l BSA containing 0.76 μM [methyl-3H]-AdoMet (NEN) at 37°C as described (Roth and Jeltsch, 2000) using 0.5 μM oligonucleotide substrate and 0.6 μM enzyme (single-turnover conditions). Initial turnover rates were derived by linear regression of the initial part of the reaction progress curves. Catalytic rates larger than 1 min⁻¹ were derived by a fit of the reaction progress curves to a single exponential. All variants were purified at least twice, and at least three independent kinetic experiments were performed with each preparation. Standard errors of the rate constants derived never exceeded ±40%. In addition, multiple-turnover (steady-state) methylation rates were determined using 25–100 nM enzyme. These analyses revealed very similar relative changes in catalytic activities as determined under single-turnover conditions (data not shown).

DNA binding experiments were analyzed by surface plasmon resonance using a BiaCore X instrument in buffer HBS-EP (10 mM HEPES [pH 7.4], 150 mM NaCl, 3 mM EDTA, 0.005% Surfactant P20) as recommended by the supplier. Experiments were performed at a flow rate of 10 μl/min at ambient temperature using streptavidin-coated chips (Sensor Chip SA, BiaCore). Chips were loaded with 400–800 resonance units of the biotinylated 20-mer substrate oligonucleotide. Protein concentrations were varied from 10 nM to 500 nM.

To compare specificities of EcoDam and its variants (Figure 6H), a specificity factor was defined as the ratio between the rates of methylation of the canonical site and the rates of methylation of all near-cognate sites, viz.

\[
S = \frac{K_{GATC}}{(K_{AATC} + K_{TATC} + K_{GATC} + K_{GAGC} + K_{GACC} + K_{GATG} + K_{GATA} + K_{GATT})},
\]

where \(k_{wxyz}\) specifies the rate of methylation of the \(wxyz\) substrate.

Similarly, specificity factors for the recognition of each position of the target sequence (Figures 6F and 6G) were defined as the ratio between the rates of methylation of all near-cognate sites modified at other positions and the rates of methylation of substrates modified at this position, viz.

\[
S^3 = \frac{K_{AATC} + K_{TATC} + K_{GATC} + K_{GAGC} + K_{GACC} + K_{GATG} + K_{GATA} + K_{GATT}}{(K_{GAGC} + K_{GACC} + K_{GATG} + K_{GATA} + K_{GATT})}
\]

\[
S^4 = \frac{K_{AATC} + K_{TATC} + K_{GATC} + K_{GAGC} + K_{GACC} + K_{GATG} + K_{GATA} + K_{GATT}}{(K_{GATG} + K_{GATA} + K_{GATT})}
\]

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgements**

We thank Dr. Zhe Yang for discussion and Dr. Robert M. Blumenthal for comments on the manuscript. These studies were supported in part by US Public Health Services grant GM29227 to S.H., US Public Health Services grant GM49245 and the Georgia Research Alliance to X.C., and grants of the German BMBF (BioFuture programme) and DFG (JE 252/2) to A.J.
References


Chen L, Paulsen DB, Scruggs DW, Banes MM, Reeks BY, Lawrence ML. Alteration of DNA adenine methylase (Dam) activity in Pasteurella multocida causes increased spontaneous mutation frequency and attenuation in mice. Microbiol 2003;149:2283–2290.


Figure 1. Structure of T4Dam-AdoHcy-13-mer DNA

(A) The two DNA molecules, shown at right with the helical axes projecting out of the page, are shifted relative to one another perpendicularly to the DNA axis (PDB code 1YF3).

(B) Schematic summary of the protein-DNA contacts in the nonspecific complex (molecule A, gray) and the 1/4-site recognition complex (molecule B, red).

(C) F111 of the hairpin loop of the joint binding Dam (molecule B) stacks with the 5′ Thy (blue).

(D) Specific interactions are observed for R116-Gua, P126-Thy, and M114-Thy.
Figure 2. Structure of T4Dam-AdoHcy-15-mer DNA

(A) All joints between two DNA duplexes are occupied by Dam molecules, labeled as C or D, while only one specific GATC site is bound by molecule E (PDB code 1YFJ).

(B) F111 in the hairpin loop of Dam molecule C stacks with two 5' Thy (blue and green).

(C) Specific interactions are mediated by R116, P126, M114, S112, G128, and R130. Malygin et al. (1999) suggested that T4Dam makes hydrogen bonds with either the Gua N7- or O6-keto groups (or both) in forming the complex.
Figure 3. Intercalation of the T4Dam F111

(A) Interactions between molecule E and a canonical GATC site. A dashed light-blue circle labels the flipped-out Ade. The region of intercalation of T4Dam into the DNA is labeled by a dashed dark-blue circle and shown enlarged in the right panel. F111 of molecule E intercalates between the AT base pair and the Thy:S112 “base-amino acid” pair.

(B) Chemical structures of AdoMet, AdoHcy, and sinefungin.

(C) Active-site conformation in the presence of sinefungin (PDB code 1YFL). An invariant N-terminal residue K11 interacts with the side chains of D171 and Y174 as well as the backbone carbonyl oxygen of G9; the same D171-K11-Y174 interactions were observed in the binary structure of T4Dam-AdoHcy (Yang et al., 2003). The D171-K11-Y174 interaction is likely to be critical for normal function since a K11S substitution virtually abolishes enzyme activity (V.G. Kossykh, S.L. Schlagman, and S.H., unpublished data). The amino group of K11 is also close to the ring N1 atom of the target Ade. The mutant of the corresponding Lys in M.EcoRV (K16R) showed an altered specificity toward the target base (Roth and Jeltsch, 2001).
Figure 4. Interactions with a Noncanonical Site

(A) F111 intercalation by molecule E into the central AT stacking of the DNA molecule depicted in orange effectively causes a one-base-pair lengthening. The expansion results in two disordered nucleotides (shaded) of the neighboring duplex (magenta).

(B) Interactions between molecule D and a noncanonical site. The 5′-overhanging Thy of the magenta DNA is pushed out and apparently becomes disordered, resulting in the Cyt of the next base pair stacking with F111 of Dam molecule D.

(C) Detailed interactions of R130 and the external G:C base pair and S112-Cyt are shown.
Figure 5. Biochemical Analysis of EcoDam Variants
(A) Schematic summary of protein-DNA base contacts in the specific complex and sequence alignment of the β hairpin loop of T4Dam (G110-T131, recognition sequence GATC), EcoDam (G118-K139, recognition sequence GATC), and EcoRV (C122-P143, recognition sequence GATATC). The flipped target base is labeled as a shaded X. Point mutations made in the EcoDam are indicated (note the differences in numbering of residues). It should be noted that the normal in vivo substrate for T4Dam is phage DNA containing glucosylated 5-hydroxymethyl-Cyt (hmCyt) in place of Cyt. Phage hmCyt-containing DNAs (with or without the presence of glucosylation) are not methylated by EcoDam (Hattman, 1970). As seen in the structures presented here, neither of the Cyt bases in the palindromic GATC site (Cyt1 or Cyt4) makes contact with T4Dam. In the specific complex, the shortest distance between the protein and these bases is 6.3 Å from Cyt4 to V178 and 7.7 Å from Cyt1 to K129. In this regard, EcoDam has insertions in both places, viz. six additional residues adjacent to V178 and two additional residues adjacent to K129 (see Figure 1 of Yang et al., 2003). These additional residues in EcoDam might sterically clash with the hydroxymethyl group on either hmCyt base (or both) and prevent the enzyme from methylating the DNA.

(B and C) Single-turnover DNA methylation rates (B) and DNA binding affinities (C) of wild-type EcoDam and its variants. EcoDam variants were cloned, expressed in E. coli, and purified to homogeneity.
Figure 6. Specificity Profiles of EcoDam

(A–E) Single-turnover methylation rates of wild-type and variants are given for the cognate GATC (light-blue bars) as well as all nine near-cognate substrates. On the horizontal axis the three positions of the GATC site that are mutated are given (G = GATC, T = GATC, C = GATC). On the right axis the new base introduced at each position is specified (for an example, see Figure S2). The methylation rates of the respective pairs of enzyme and substrate are given on the vertical axis (note the logarithmic scaling). (A) Wild-type, (B) R124A, (C) P134A, (D) P134G, and (E) L122A.

(F–H) Specificity factors (defined in Experimental Procedures) of EcoDam variants for recognition of the fourth (S⁴) (F) and third positions (S³) (G) of the GATC sequence and overall specificity factors (H). All values are given as relative changes with respect to the wild-type. The specificity factor of wild-type EcoDam was 540; the value was increased at least 30-fold in the case of the L122A variant. Because no activity at near-cognate sites could be detected with the L122A variant, the specificity factor given here is a lower limit, indicated by the arrow. The specificities of the R124A, P134A, and P134G variants were dramatically reduced. The
specificity factors of all other variants did not show large deviations when compared with the wild-type enzyme.
Figure 7. Snapshots of T4Dam-DNA Interactions Illustrated by Orientation of the Protein Hairpin Loop Relative to the DNA Axis
(A) Nonspecific complexes with R130 involved in phosphate contact.
(B) Nonspecific complex with R116 involved in phosphate contact.
(C) The 1/4-site complex with R116 involved in base-specific contact and N118 and R130 in phosphate contacts.
(D) The 3/4-site complex with R116 and R130 involved in base-specific contacts and N118 in phosphate contact.
(E) Interaction with a noncanonical site.
(F) A full-site complex.