Structure of *Pvu*II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment

Weimin Gong, Margaret O'Gara, Robert M. Blumenthal[1] and Xiaodong Cheng*

W.M.Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA and [1]Department of Microbiology and Immunology, Medical College of Ohio, Toledo, OH 43699-0008, USA

ABSTRACT

We have determined the structure of *Pvu*II methyltransferase (M.*Pvu*II) complexed with *S*-adenosyl-L-methionine (AdoMet) by multiwavelength anomalous diffraction, using a crystal of the selenomethionine-substituted protein. M.*Pvu*II catalyzes transfer of the methyl group from AdoMet to the exocyclic amino (N4) nitrogen of the central cytosine in its recognition sequence 5′-CAGCTG-3′. The protein is dominated by an open α/β-sheet structure with a prominent V-shaped cleft: AdoMet and catalytic amino acids are located at the bottom of this cleft. The size and the basic nature of the cleft are consistent with duplex DNA binding. The target (methylatable) cytosine, if flipped out of the double helical DNA as seen for DNA methyltransferases that generate 5-methylcytosine, would fit into the concave active site next to the AdoMet. This M.*Pvu*II α/β-sheet structure is very similar to those of M.*Hha*I (a cytosine C5 methyltransferase) and M.*Taq*I (an adenine N6 methyltransferase), consistent with a model predicting that DNA methyltransferases share a common structural fold while having the major functional regions permuted into three distinct linear orders. The main feature of the common fold is a seven-stranded β-sheet (6↓ 7↑ 5↓ 4↓ 1↓ 2↓ 3↓) formed by five parallel β-strands and an antiparallel β-hairpin. The β-sheet is flanked by six parallel α-helices, three on each side. The AdoMet binding site is located at the C-terminal ends of strands β1 and β2 and the active site is at the C-terminal ends of strands β4 and β5 and the N-terminal end of strand β7. The AdoMet–protein interactions are almost identical among M.*Pvu*II, M.*Hha*I and M.*Taq*I, as well as in an RNA methyltransferase and at least one small molecule methyltransferase. The structural similarity among the active sites of M.*Pvu*II, M.*Taq*I and M.*Hha*I reveals that catalytic amino acids essential for cytosine N4 and adenine N6 methylation coincide spatially with those for cytosine C5 methylation, suggesting a mechanism for amino methylation.

INTRODUCTION

DNA methyltransferases (Mtases) transfer a methyl group from *S*-adenosyl-L-methionine (AdoMet) to a given position of a particular DNA base within a specific DNA sequence. The resulting methylation can protect the DNA from a cognate restriction endonuclease or can have epigenetic effects on gene expression. The DNA Mtases belong to two families: one methylates C5, a ring carbon of cytosine, yielding 5-methylcytosine (5mC), while the second family methylates the exocyclic amino group (NH_2) of cytosine or adenine yielding *N*4-methylcytosine (N4mC) or *N*6-methyladenine (N6mA) respectively. Two of the 5mC Mtases have been structurally characterized as covalent reaction intermediate complexes with their DNA substrates (1,2); one of these, M.*Hha*I has been characterized in complexes with structural analogs of DNA in three different methylation states, unmethylated, hemimethylated and fully methylated (3,4).

The primary sequences of the 5mC Mtases share a set of conserved motifs (I–X) in a constant linear order (5–9). The majority of these motifs are responsible for three basic functions of the 5mC Mtases: AdoMet binding, sequence-specific DNA binding and catalysis of methyl transfer. In contrast, the amino-Mtases (which generate N6mA or N4mC) belong to three groups characterized by distinct linear orders for the conserved motifs (10). The three groups are named α (including Mtases such as Dam), β (including Mtases such as M.*Pvu*II) and γ (including Mtases such as M.*Taq*I). To date only one DNA amino-Mtase has been structurally characterized, the group γ N6mA Mtase M.*Taq*I (11).

While the M.*Taq*I structure has been determined only in the absence of DNA, it is sufficient to allow general structural comparison with the 5mC Mtases. Both M.*Hha*I and M.*Taq*I are bilobal structures: one lobe contains a catalytic domain with both the active site for methyl transfer and the AdoMet binding site and the other lobe contains a target (DNA) recognition domain (TRD). The catalytic domains of the two proteins exhibit very

---

*To whom correspondence should be addressed at present address: Department of Biochemistry, Emory University School of Medicine, 1510 Clifton Road, Atlanta, GA 30322, USA. Tel: +1 404 727 8491; Fax: +1 404 727 3746; Email: xcheng@emory.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

similar three-dimensional folding (12). This folding pattern is also present in M.*Hae*III, another 5mC Mtase, in catechol *O*-Mtase, a single domain small molecule AdoMet-dependent Mtase, in VP39, an mRNA cap-specific RNA 2′-*O*-Mtase and in glycine *N*-Mtase (2,13–15). The folding similarity includes the positions of conserved amino acid side chains involved in either AdoMet binding or catalysis; only the binding of AdoMet reported for glycine *N*-Mtase differs from the consensus pattern (15). Guided by this common catalytic domain structure, sequence alignment of amino-Mtases suggests that for all amino-Mtases to fit the consensus M.*Hha*I/M.*Taq*I catalytic domain structure, despite having different motif orders, different sets of topological connections would be required for the three DNA amino-Mtase groups (10).

Determining the structure of *Pvu*II methyltransferase (M.*Pvu*II), a group β N4mC Mtase, would thus address two important questions about DNA Mtases. First, do the N4mC Mtases in fact match the consensus catalytic domain structure seen between M.*Taq*I and M.*Hha*I (12)? Second, are the major structural elements of amino-Mtases connected in three different orders, as suggested by their primary sequences (10)? M.*Taq*I itself did not provide a strong test for this model because the group γ and 5mC Mtases have essentially the same motif order; they differ only in the position of motif X (10). No Mtase from group α or β has been structurally characterized before this report.

M.*Pvu*II, part of the restriction–modification system from the Gram-negative bacterium *Proteus vulgaris* (16), modifies the internal cytosine of the recognition sequence 5′-CAGCTG-3′ (17) to generate N4mC (18). *Pvu*II endonuclease, which cleaves duplex DNA at the center of the same recognition sequence to generate blunt-ended products, was structurally characterized earlier (19,20). With this report, the *Pvu*II restriction–modification system becomes the first system for which the structures of the cognate endonuclease and methyltransferase have both been determined.

MATERIALS AND METHODS

Overexpression and crystallization

Overexpression and purification of and selenomethionine (SeMet) incorporation into M.*Pvu*II have been described previously (21). To crystallize the M.*Pvu*II–AdoMet binary complex, 0.2 mM AdoMet was added to the pre-purified protein (~5 μM) and the mixture was further purified by cation exchange chromatography (21). M.*Pvu*II and selenomethionyl M.*Pvu*II, complexed with AdoMet, both crystallized in the monoclinic space group P2₁ with unit cell dimensions of $a = 48.8$ Å, $b = 112.4$ Å, $c = 59.3$ Å and $\beta = 109.2°$ (21). There are two molecules per crystallographic asymmetric unit cell, termed molecules A and B. X-Ray diffraction data were collected using a MarResearch imaging plate detector on beamline X12-C at the National Synchrotron Light Source, Brookhaven National Laboratory, and processed using the HKL software package (22). Multiwavelength anomalous diffraction (MAD; 23) data to 3.3 Å resolution (Table 1) were collected on a single frozen SeMet crystal at three different wavelengths, corresponding to the inflection point λ1 (minimum Δ*f*′) and the peak λ2 (maximum Δ*f*″) of the Se-containing crystal absorption spectrum and a third wavelength (λ3) remote from the peak position (21). A higher resolution data set (up to 2.8 Å) used

for final model refinement was collected from a native crystal (λ4 = 1.072 Å, 180° rotation, 1.5° increment, 90 s exposure).

SeMet MAD phasing

There are a total of 18 possible Se sites per asymmetric unit (nine per molecule). To locate the Se positions, we calculated the anomalous and isomorphous difference Patterson maps at the Harker section ($v = 1/2$) among data sets collected at wavelengths λ1, λ2 and λ3. A number of peaks were observed, which corresponded to possible Se sites and the cross vectors between them (21). Five Se sites were first manually determined from the Patterson maps. These five sites were used to calculate initial estimates of phases, to compute the difference and Bijvoet difference Fourier synthesis and to search for additional Se sites. Finally, a total of 12 Se sites were determined and confirmed by the two-fold non-crystallographic symmetry (NCS) operator, revealed by a self-rotation function (21).

Table 1. Statistics of experimental SeMet MAD data with rejection criteria $I/\sigma(I) \geq 2$

	λ1	λ2	λ3
Wavelength (Å)	0.98233	0.98211	0.92
Energy (eV)	12 621	12 624	13 476
Resolution range (Å)	∞–3.30		
Completeness (%)	94.5	92.2	94.5
$R_{linear} = \Sigma\|I - <I>\|/\Sigma I$	0.048	0.051	0.044
$<I/\sigma>$	17.0	16.1	18.6
Observed reflections	29 236	27 108	27 410
Unique reflections	8 904	8 790	8 913
Anomalous pairs	7 576	8 212	7 475
Highest resolution shell (Å)	3.36–3.30		
Completeness (%)	90.0	87.5	90.5
$R_{linear} = \Sigma\|I - <I>\|/\Sigma I$	0.097	0.101	0.087
$<I/\sigma>$	9.6	8.9	11.1
Unique reflections	425	391	431

These 12 Se positions were used for MAD phasing by treating the data from each wavelength as a multiple isomorphous replacement experiment with the inclusion of anomalous scattering (MIRAS): native with native anomalous scattering (λ3), derivative isomorphous (λ1) and derivative isomorphous with anomalous scattering (λ2; Table 2). The MAD–MIRAS phases were improved using 40% solvent content by four, four and eight cycles of solvent leveling (24) following each of three envelope determinations (25). The solvent-leveled map was used to refine the NCS operator and to construct the averaging mask. The phases were further improved using 16 rounds of Furey's averaging protocols (25). The electron density was averaged within the mask, the density for each molecule was replaced with the average and the 'averaged' density map was inverted to obtain new phases. The resulting phases were combined with the original solvent-leveled

MAD-MIRAS phases. The process was cycled until convergence was obtained (16 cycles).

Refinement

The starting Cα backbone for molecule A was traced using the skeleton in program O (26) with reference to three maps at 3.3 Å resolution: MAD-MIRAS, solvent flattened and density averaged. The Se positions also provide markers for selenomethionine in the polypeptide chain tracing. The atomic coordinates for molecule B were generated by the two-fold NCS operator. After the initial model building, the atomic model was subjected to refinement against 2.8 Å resolution data from a native crystal (Table 3). Initially, a strict NCS was invoked, assuming that two NCS-related molecules are strictly identical. The two models were refined by simulated annealing and least squares minimizations using the X-PLOR program suite (27). Seven rounds of refinement and model rebuilding brought the crystallographic R factor to 0.22. The model was further refined by a restraint NCS, with two NCS-related atoms restrained in their average positions. An additional five rounds of refinement, refitting and placing ordered water molecules brought the R factor to 0.19 and R_{free} to 0.28 (Table 3).

Table 2. Treatment of SeMet MAD data as MIRAS at 3.3 Å resolution

Native wavelength		λ3	
Derivative wavelength	λ1	λ2	λ3
isomorphous/anomalous	iso	iso/ano	ano
Phasing power[a]	2.350	1.830/1.410	1.080
R-Kraut[b]	0.028	0.028/0.037	0.029
R-Cullis[c]	0.516	0.601/–	–
Figure of merit	0.389	0.365/0.308	0.265
Overall figure of merit	0.619		

[a]Phasing power = r.m.s.($<F_H>/E$), where F_H is the calculated 'heavy atom' structure factor amplitude and E is the residual lack of closure.
[b]R-Kraut = $\Sigma|F_{PH} - F_{PH}^{cal}|/\Sigma|F_{PH}|$ for centric data and $\Sigma(|F_{PH+} - F_{PH+}^{cal}| + |F_{PH-} - F_{PH-}^{cal}|)/\Sigma|F_{PH+} + F_{PH-}|$ for acentric data.
[c]R-Cullis = $\Sigma||F_{PH} \pm F_P| - F_H|/\Sigma|F_{PH} \pm F_P|$ for centric data, where F_{PH} and F_P are the observed structure factor amplitudes for the 'derivative' and 'native' data sets and F_H is the calculated 'heavy atom' structure factor amplitude.

RESULTS

Structure determination

M.*Pvu*II is produced in two forms, resulting from translation initiators 13 codons apart (17). The shorter form of M.*Pvu*II, starting from the internal translation initiator at Met14, was overexpressed in *Escherichia coli* and purified both in native and selenomethionine-substituted forms (21). The M.*Pvu*II polypeptide chain is 323 amino acids long (numbered 14–336). Diffraction data (Table 1) were collected at three X-ray wavelengths from a crystal of the selenomethionyl M.*Pvu*II–AdoMet complex, so that MAD could be used to extract the phases (23). Following the suggestion of Ramakrishnan (28,29), multiwavelength data were treated as if they were from a conventional MIRAS experiment.

A total of 12 (out of 18 possible) Se sites per asymmetric unit were determined from Patterson maps and were used for MAD phasing with a figure of merit of 0.62 at 3.3 Å resolution. The MAD-MIRAS map, coupled with two-fold non-crystallographic symmetry averaging, was accurate enough to permit an initial interpretation (21) and the model was finally refined to 2.8 Å resolution with a crystallographic R factor of 0.19 and an R_{free} value of 0.28.

Table 3. Structural refinement of M.*Pvu*II at 2.8 Å resolution

Wavelength (Å)	1.072	
Resolution range (Å)	∞–2.8	2.85–2.8 (highest resolution shell)
Completeness (%)	99.7	97.5
$R_{linear} = \Sigma|I - <I>|/\Sigma I$	0.052	0.226
$<I/\sigma>$	11.5	5.5
Observed reflections	54 787	
Unique reflections	14 886	730
R_{factor}[a]$= \Sigma|F_o - F_c|/\Sigma|F_o|$	0.193	
R_{free}[a]	0.283	
Non-hydrogen protein atoms	4455	
r.m.s. deviation from ideality		
Bond lengths (Å)	0.02	
Bond Angles (°)	2.9	
Dihedrals (°)	24.9	
Improper (°)	2.2	

[a]R_{factor} and R_{free} are calculated for ~92 and 8% of the data respectively.

Overview of the M.*Pvu*II structure

The polypeptide chain folds into a structure with a V-shaped cleft, big enough to accommodate duplex DNA (Fig. 1). The V-shaped cleft is formed by three loops on one side and a three-helix bundle on the other side. The methyl donor AdoMet binds at the bottom of the cleft, which consists of a twisted 10 stranded β-sheet around which six α-helices are arranged on both sides.

Figure 2 shows the topology diagrams of M.*Pvu*II, M.*Hha*I and M.*Taq*I. For clarity and convenience, we retain the nomenclature of Schluckebier *et al.* (12) for the secondary structure assignment and of Posfai *et al.* (5) for the conserved motifs. Loops or turns are designated by their flanking secondary structures; two of them are termed the glycine-containing G loop (loop 1-A) and the proline-containing P loop (loop 4-D) (10). The catalytic domains of the three structures are all of the α/β type with a central β-sheet sandwiched between two layers of α-helices: helices αC, αD and αE located on one side and helices αZ, αA and αB on the opposite side of the sheet (Fig. 2a). The β-sheets in the three structures all contain five central adjacent parallel β-strands with strand order 5, 4, 1, 2, 3 and one antiparallel hairpin (β6 and β7) next to strand β5. The order of parallel strands is reversed once between β4 and β1. The majority of the active amino acids from conserved motifs (circled in Fig. 2) are located at the carboxyl ends or in loop regions outside the carboxyl ends of these parallel β-strands. In all three structures the AdoMet binding site is located at the carboxyl ends of strands β1 and β2 and the amino end of helix αC;
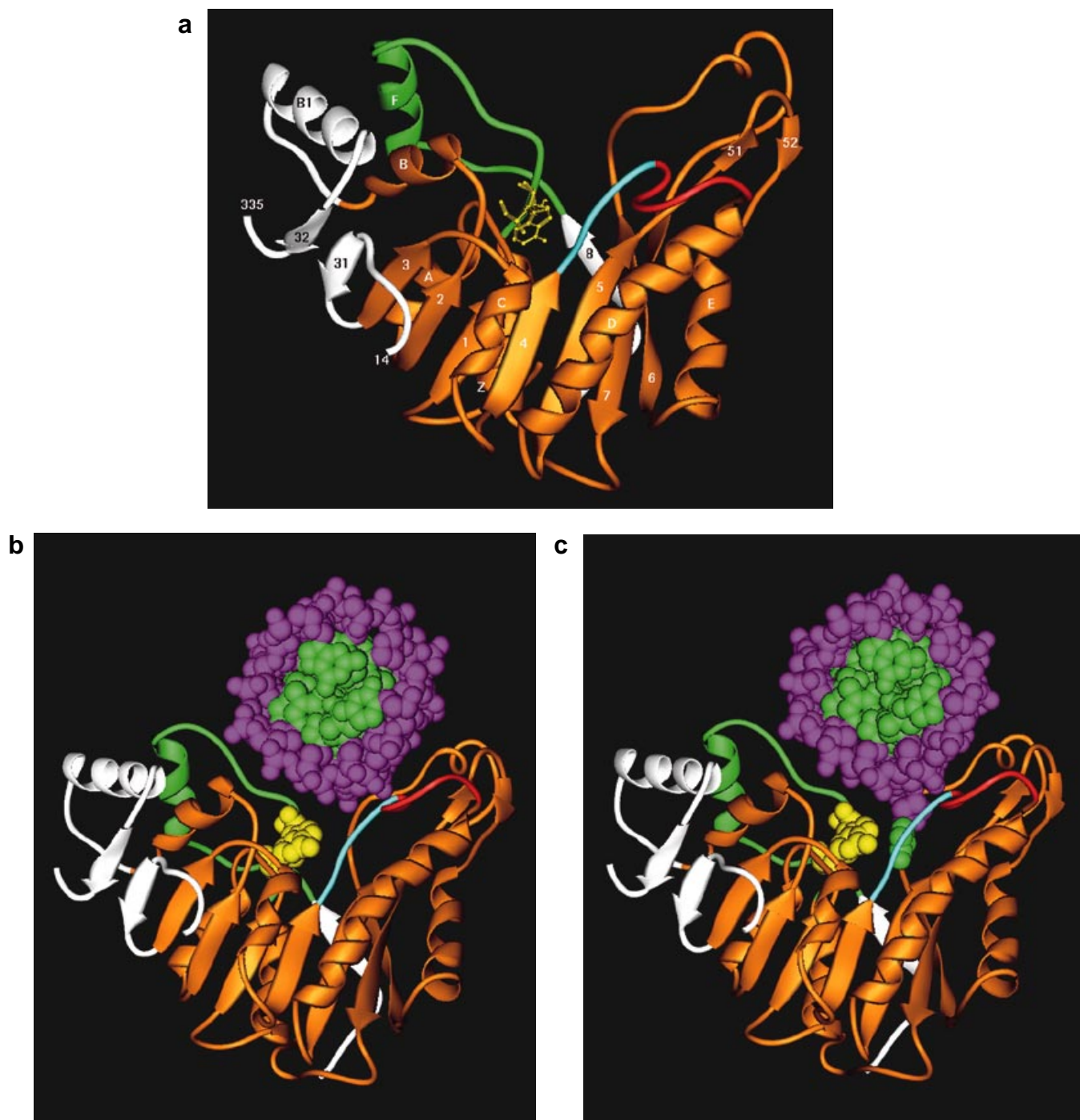
Figure 1. Ribbon (70) diagram of M.*Pvu*II. (**a**) The protein folds into a structure with a V-shaped cleft. AdoMet (in ball-and-stick representation) is bound at the bottom of the cleft. The regions of M.*Pvu*II that are structurally most similar to M.*Hha*I and M.*Taq*I are shown in brown and the less similar regions are shown in white and green. The green region is part of the putative DNA target recognition domain (TRD). The catalytic P loop is in pale blue and red. The pale blue part contains conserved amino acids Ser53–Pro–Pro–Phe56 and the red part is flexible (high thermal factors), consistent with potential conformational change upon DNA binding. (**b**) M.*Pvu*II docked to cognate DNA, taken from the R.*Pvu*II–DNA structure (19). The DNA phosphate backbone and sugar rings are in purple, the DNA bases are in green and AdoMet is in yellow. (**c**) M.*Pvu*II docked to DNA with a flipped cytosine (see Fig. 5).

and the active site at the carboxyl ends of strands β4 and β5 and the amino end of the strand β7 (see below). The N- and C-termini of the folded polypeptide are within the AdoMet binding region in all three structures: located in the region between helix αZ and strand β1 (M.*Hha*I in Fig. 2b), prior to helix αZ (M.*Taq*I in Fig. 2c) and between helix αB and strand β3 (M.*Pvu*II in Fig. 2d).
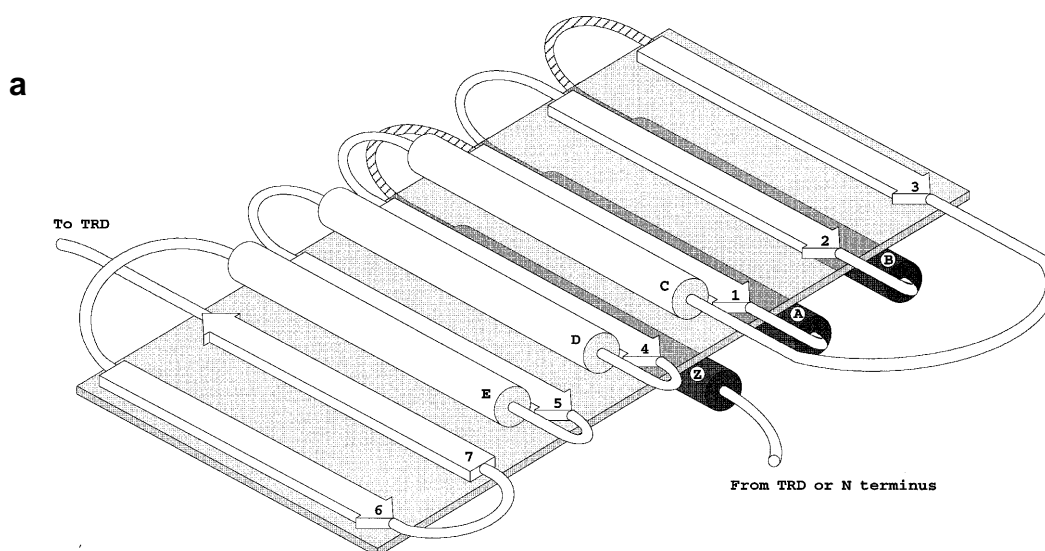
a

Figure 2. Topology diagrams. (**a**) The consensus methylase fold, derived from DNA Mtases (M.*Hha*I, M.*Hae*III, M.*Taq*I and M.*Pvu*II) and one small molecule AdoMet-dependent Mtase (catechol *O*-Mtase). The main feature of the fold is a region of five parallel β-strands (5, 4, 1, 2, 3) followed by an antiparallel β-hairpin (strands 6 and 7), surrounded by six helices, three (αC, αD, αE and αZ, αA, αB) on each side of the β-sheet. The dashed loops can be broken to become the N- and C-termini. Opposite page: (**b**) M.*Hha*I, (**c**) M.*Taq*I and (**d**) M.*Pvu*II diagrams indicate their similarity in the catalytic domains. α-Helices are shown as rectangles (lettered) and β-strands as broad arrows (numbered). Common elements of secondary structure among the three enzymes are shown in similar positions. Conserved or functionally important amino acids from motifs I–X are circled. The β-strands (ten in M.*Pvu*II, seven in M.*Hha*I and nine in M.*Taq*I) form a β-sheet. In M.*Pvu*II, a region (dashed line) between strands β7 and β8 is not modeled in the current structure. (**e**) Predicted topological folding for group β amino-Mtases from an earlier study (10).

M.*Pvu*II fits the consensus fold for AdoMet-dependent Mtases
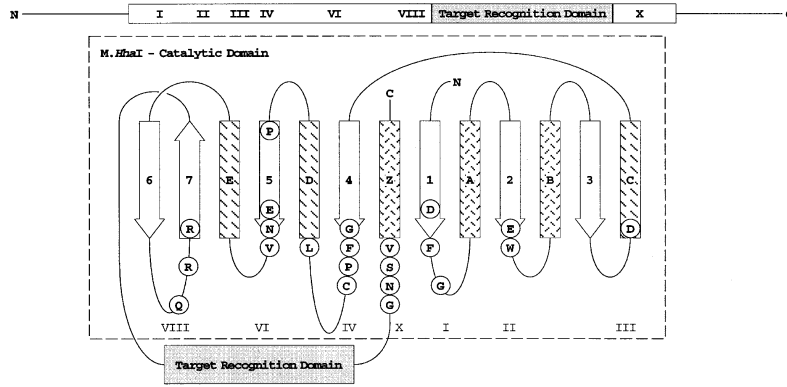
The TRD, which is associated with sequence-specific DNA recognition, lies in the smaller domain of the three bilobal Mtases discussed above. In the current structure of M.*Pvu*II, this domain comprises only one helix (αF) and its associated loops (Fig. 2d). It is interesting how the TRD is connected to the catalytic domain in the three Mtases. In 5mC Mtases such as M.*Hha*I, helix αZ (motif X, part of the catalytic domain) is folded from the C-terminus, following the TRD. Thus, there are two connections between the catalytic domain and TRD (Fig. 2b). In group γ N6mA Mtase M.*Taq*I, helix αZ originates from the N-terminus and the TRD is linked to the catalytic domain through β9 only (Fig. 2c). Thus, in both M.*Taq*I and M.*Hha*I the functional regions are in the order (amino→carboxyl) AdoMet binding region, active site region and TRD, the major difference between them being that helix αZ is moved from the N-terminus in M.*Taq*I to the C-terminus in M.*Hha*I.

As predicted (10), the most pronounced difference in topology between M.*Pvu*II and both M.*Hha*I and M.*Taq*I is the connection between the AdoMet binding and active site regions: the two regions are connected via the putative TRD (helix αF) in the order (amino→carboxyl) active site region, TRD and AdoMet binding region (Fig. 2d). The active site and AdoMet binding regions of M.*Pvu*II fit the consensus structure of M.*Hha*I/M.*Taq*I/M.*Hae*III/ catechol *O*-Mtase, regardless of the motif order in the primary sequence. We call this common catalytic domain structure the AdoMet-dependent methylase fold (Fig. 2a). This fold has also been observed in the RNA Mtase VP39, though helix αE is replaced by a β-strand (14).
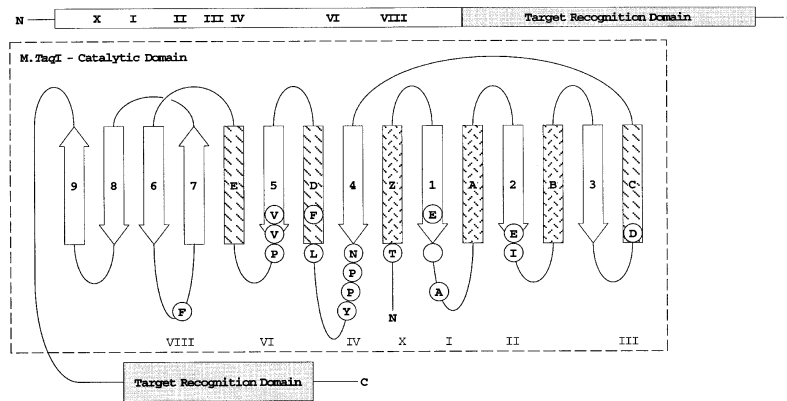
We had predicted the folding of group β amino-Mtases, including M.*Pvu*II (Fig. 2e), based on structure-guided sequence analysis (10). Overall, the prediction is quite accurate, though there are some significant differences between the prediction and the current model. Unexpectedly, part of the AdoMet binding region (β3–αC or motif III) is located upstream of the active site region, near the N-terminus of the polypeptide. This arrangement preserves the crossover between strands β3 and β4, but splits the coding for the AdoMet binding region into two distant parts of the gene. The β3–αC secondary structure (motif III) was predicted to originate from the C-terminus as a contiguous part of the AdoMet binding region. This prediction, which would result in no crossover connection between strands β3 and β4, was made in part because of the very short distance between the N-terminus of another group β Mtase (M.*Bam*HII) and its strand β4 (Fig. 3). However, this crossover has been observed in all currently available DNA Mtase structures (5mC Mtases M.*Hha*I and M.*Hae*III, group β Mtase M.*Pvu*II and group γ Mtase M.*Taq*I), as well as in catechol *O*-Mtase, glycine *N*-Mtase and the RNA Mtase VP39, and is predicted to occur in group α amino-Mtase structures, with the crossover connection in a separate domain comprising the TRD (10).

Such a crossover is necessary to generate a so-called topological switch point (30), at which the strand order is reversed and loops connected to the carboxyl ends of the two adjacent strands (β1 and β4 in Fig. 2a) go in opposite directions. The positions of concave active sites can be predicted from such switch points in different types of α/β twisted open sheet structures, including arabinose
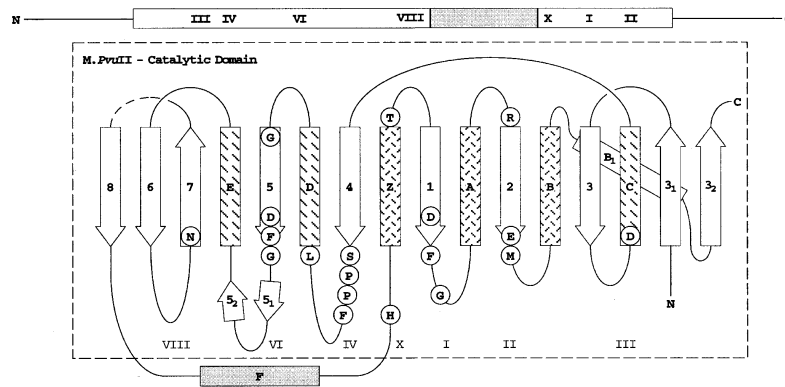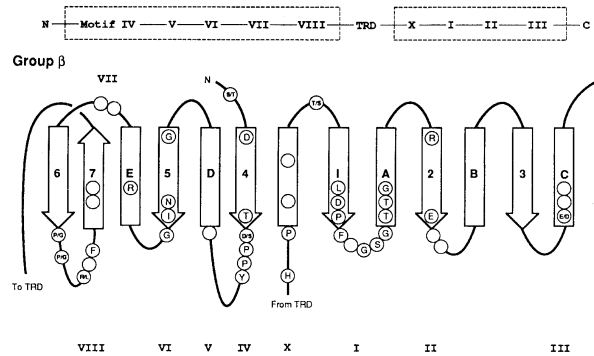
b

c

d

e

```
                20        30        40        50        60        70        80        90       100       110       120
                |         |         |         |         |         |         |         |         |         |         |
                     3₁      3     C             4                      D            5      5₁  5₂       E

PvuII  (N4)  14 mlnfgkkpayttsngsMYigDsLElLesFpeeSIsLVMTSPPFaLqrkkeygnleqheYVdWFlsFakvvnKkLKpdGSFVVDFGgaYMkgvparsiynfrvlirmide
BamHI                        LYngDcLElFkqVpdenVDtIFaDPPFnL              YLdWYykWIdEciRVLKpgGSLFIYNIPkWn
BamHII                 --------mkenIgdcTIDLtVTSPPYdd                    YsfnFeetaqElyRVtKegGvVVWvVGdkth
Cfr9I                      LFegDaLsvLrrLpsgSVrcIVTSPPYwg                FLnrLvtIFsEakRVLtddGTLWVNIGdgYt
CfrBI                      LLqgNcaEtLkkLpdeSVNLVFTSPPYyn                YLsLLrsVIkEchRVLsegrfFVINVspvLI
MthZI                      IYfkNsaD-MneLkdkSINLVVTSPPYpM                MheeLekVWhEvdRVtapgGvVIINIGdatr
SmaI                       IIvgDarEaVqgLdseiFDcVVTSPPYwg                YIkdLvdLFrDvrRtLKddGTLWLNIGdsYt
XcyI                       LFegDaLtvLrrLpsgSVrcVVTSPPYwg                FLhrLvaIFaEvkRVLtddGTLWLNIGdgYt
DpnII 2 (N6)               LVhsDtFkfLskMkpeSMDMIFaDPPYfL                kheFnrkWIrlakeVLKpnGTVWIsgslhnI
EcaI                       FIwsDnslaLnrL/5/kakLIYlDPPYat                YLeFMrrrLilmreILdddGTIYVhIGhqML
HinfI                      ILkgDcIEkLktIpneSIDLIFaDPPYfM                YdsFcelWLkEckRILKstGSIWVigsfqnI
HpaI                       LIcsNaIkaLknLeenSIDLIITDPPYnL                YLeFsheWLeEcyRVLKphGTIYIfMGmkYI
KpnI                       IYigDnLtyLqgL/5/TIDFcYiDPPYnt                WMsFLlprLfhahKMLKdtGiIaIsIddyeF
MboI C                     IYqgNcIDfMshFqdnSIDMIFaDPPYfL                IYnFnheWIaQarqLLKdnGTIWIsgthhnI
MboII                      IhqmNcFDfLdqVenkSVQLaViDPPYnL                FLpFtyrWIdkvldkLdkdGSLYIfntpfnc
RsrI                       YdvcDcLDtLakLpddSVQLIIcDPPYnI                YIgWakrWLaEaeRVLsptGSIaIfgGlqYq
HhaII                      knkmNgIDlFklIpdkaVkIaFfDDgYrg                tdeIIqqFInEfeRVLpnGyLFLwVdkfhL
                           |←Motif III→|  |←Motif IV→|             |←——Motif V——→|  |←Motif VI→|
```

```
               130       140       150       160       170       180       190       200       210       220       230
                |         |         |         |         |         |         |         |         |         |         |
               ————————6—————  ——————7———————                                                    ——————————————8——→

PvuII  (N4)  123 vgfflaedfywfnpskLpspiewvnkrkirvkdavNtvwwfsktewpksditkvlapysdrmkkliedpdkfytpktrpsghdigksfskdnggsippnllqisnsesn
                                      |←VIII→|                            ↑   ↑          ↑              ↑
```

```
               240       250       260       270       280       290       300       310       320       330
                |         |         |         |         |         |         |         |         |         |
               —   F   ——   ——   Z   ——  ——1——  ——   A   ——  ——2——  ——   B   ——       B₁——  —3₂—→

PvuII  (N4)  232 gqylancklmgikaHpArFPakLpefFIrmlTepdDLVVDiFgGCSnTTGlVAeresRkWIsFEMkpEYvaasafRFldnniseekitdiynrilngesldlnsii
BamHI                          kfneLsvkLLdrIItmSTneGDVVLDPFgCSGTTfaVsemlgRkWIGFELgnce--IikeRL                          ↑
BamHII                         HpAiFPekLaedhIlswSneGDIVFDPFmcSGTTAkMAalnnRkYIGtEIskEYcdIaneRL
Cfr9I                          HfAtFPteLIrpcIlaSTkpGDYVLDPFfCSGTvGvVcqqedRqYVGIELnpEYvdIavnRL
CfrBI                          HpAtFPygLaerVIkyySfknDVIILDPFaCSGTTAkaAidlgRrFVmcEIskQYidLiiegt
MthZI                          raAaYPfeLayrLInmySimGDWVLDPFlCTGTTmiaAacagRnsIGYELdhNFkdLiesRI
SmaI                           HfAvFPraMarlcVlagSrpGGkVLDPFfCSGTTGvVcqeldRecVGIELneEYasLakeRI
XcyI                           HfTtFPpeLIrpcIhaSTepGDYVLDPFfCSGTvGlVcqdenRqYVGIELnpEYvtLaadRL
DpnII 2(N6)                    HpTqKPeyLLerIIlaSTkeGDYILDPFvCSGTTGvVAkrlgRrFIGIDaekEYlkIarkRL
EcaI                           ypTeKnfnMMklIVgaSSnpGDLVIDPFcGSGSTlhaAsllgRkWIGIDeslfaaktvmkRF
HinfI                          HsTqKPesLLykVIlsSSkpnDVVLDPFfCTGTTGaVAkalgRnYIGIEreqkYidVaekRL
HpaI                           HpTqKPeaLYerMIlaSSneGDIVLDPFvCSGTlnfVckhlnRsgIGIDInkEYieMakeRL
KpnI                           fdTpKAlnYImsIIncmakpdaLILDPFaCSGTTAhaA/9/sRktIlMEsnhpItkthiayk
MboI C                         HpTqKPlgLLsrIIlsSTqkdDLILDPFsCSGTTGiaGvlldRnYIGIEgelEFleLskrRY
MboII                          HiTpKPrdLIerIIraSSnpnDLVLDcFmcSGTTAiVAkklgRnFIGcDMnaEYvnqanfvL
RsrI                           HpTqKPaaVIerLVralShpGstVLDffaCSGvTArVAiqegRnsIctDaapvFkeYyqkqL
HhaII                          HthsKPieMqkqLIlaTTqeGDLILDPasCGgySvfecckqtnRnFIGcDLifgddeneqd--
                               |←——Motif X——→|  |←——Motif I——→|  |←Motif II→|
```
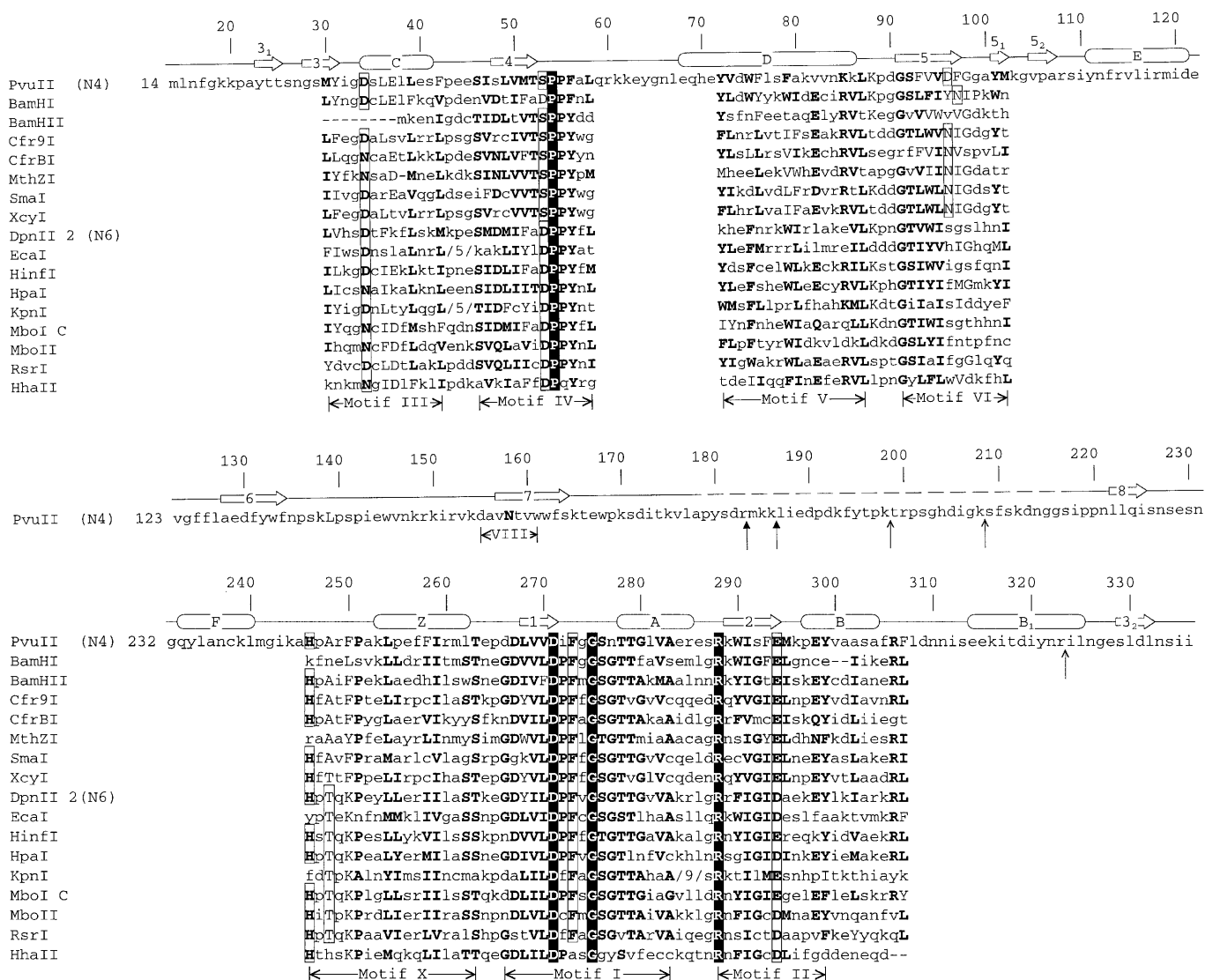
Figure 3. Sequence alignment of group β amino-Mtases including eight N4mC and nine N6mA Mtases (there is some uncertainty on assignments, particularly for M.*Hin*fI). Conserved amino acids are grouped as (E, D, Q, N), (V, L, I, M), (F, Y, W), (G, P, A), (K, R) and (S, T), using standard one letter abbreviations. Invariant amino acids are shown as white letters against a black background; conserved positions are indicated by bold letters within a box. Lesser degrees of conservation are shown, in decreasing order, by bold and upper case letters, while non-conserved positions are shown as lower case letters. A dash (–) indicates a deletion relative to other sequences and a slash (/) followed by a number indicates an insertion and its size. Motifs I–X are labeled using the nomenclature of Posfai *et al.* (5). The secondary structures of M.*Pvu*II are indicated by cylinders (α-helices) and arrows (β-strands) drawn directly above the amino acids forming them. The dashed line indicates a flexible region (amino acids 179–216) which is not modeled in the current structure. This region includes four out of five preferred trypsin cleavage sites indicated by arrows (34).

binding protein (31), carboxypeptidase (32) and tyrosyl-tRNA synthetase (33).

Disordered regions

As noted above, there are two molecules per crystallographic asymmetric unit cell, termed molecules A and B. The current model of molecule A contains residues a16–a178, a217–a335 and one AdoMet, while molecule B contains residues b16–b56, b69–b178, b215–b335 and one AdoMet. The r.m.s. deviation between 269 common Cα atoms of the final refined two molecules is 0.6 Å. In both molecules ~40 amino acids (Pro179–Gly216), located immediately after strand β7 and before

strand β8, were not modeled in the current structure because of poor electron density. This poor density suggests that these amino acids are very flexible. Consistent with this flexibility, four out of five preferred trypsin cleavage sites are within this 40 amino acid region: the primary cleavages occur on the carboxyl sides of Arg183 and Lys186 and are followed by slower cleavages carboxyl of Lys198, Lys208 and Arg323 (34). In fact, SDS–PAGE analysis of dissolved crystals indicates that some M.*Pvu*II crystals contained limited amounts of protein that had been cleaved in this region. It is noteworthy in this regard that some 5mC Mtases are naturally made as two separate polypeptides that associate in the cell to form active enzyme (35,36).

In molecule B part of the catalytic P loop (amino acids 57–68) was also not modeled due to poor electron density. However the corresponding P loop in molecule A was modeled, though Leu58–Asn66 (red in Fig. 1) possessed the highest crystallographic thermal factors in the current refined structure. This flexibility may be due to the absence of the DNA in the crystal and suggests a potential conformational change upon DNA binding. Similarly, the catalytic P loop in M.*Hha*I, which contains the key catalytic amino acids Pro80–Cys81, undergoes a massive conformational change upon binding DNA, moving ~25 Å toward the corresponding DNA binding cleft of the protein (1).

AdoMet binding

The binding site for AdoMet is adjacent to the carboxyl ends of strands β1, β2, the amino end of helix αC and the loop prior to helix αZ, regions that contain conserved motifs I, II, III and X respectively (Fig. 4). The interactions between AdoMet and M.*Pvu*II are almost identical to those between AdoMet and M.*Hha*I (1), M.*Taq*I (11), catechol *O*-Mtase (13) and VP39 (14). Amino acid side chains interacting with AdoMet are found in spatially equivalent positions, except that Phe273 of M.*Pvu*II and Phe18 of M.*Hha*I are in the G loop, while the corresponding Phe146 of M.*Taq*I is in helix αD (Fig. 2).

In motif I of group β Mtases (Asp–X–Phe–X–Gly), the amino acids Asp and Gly are invariant. In M.*Pvu*II these correspond to Asp271 and Gly275 (Fig. 3). The side chain carboxylate of Asp271 (β1) makes two hydrogen bonds to the main chain amide group of Phe273 (G loop) and the side chain hydroxyl of Thr279 (αA) and these bonds stereochemically constrain the β1–loop–αA structure. A negatively charged amino acid corresponding to Asp271 has been found in the same position of motif I in all DNA Mtases sequenced so far, including Asp16 of M.*Hha*I and Glu45 of M.*Taq*I (9,10). The main chain amide group of Gly275 (G loop) hydrogen bonds to the side chain carboxylate of Glu294 (β2), which is another conserved negatively charged amino acid (motif II) that interacts with the ribose hydroxyls of AdoMet. Comparable backbone–side chain interactions occur in M.*Hha*I (Gly20–Glu40) and M.*Taq*I (Ala49–Glu71).

In M.*Pvu*II the AdoMet binding α/β cluster (αZ→β1→αA→β2→αB) is further stabilized by the interactions of Arg288 (an invariant arginine among group β Mtases located prior to strand β2; Fig. 3) with the side chain of Thr263 (loop Z-1) and backbone carboxyls of both Thr263 (loop Z-1) and Glu286 (loop A-2). Only three structurally characterized AdoMet binding proteins interact with AdoMet in substantially different ways from the nucleic acid Mtases and catechol *O*-Mtase. One of these proteins is the *E.coli* MetJ repressor (37), for which AdoMet is a co-repressor and not a substrate; another is the reactivation domain of *E.coli* methionine synthase (38), which uses AdoMet in a flavodoxin-coupled reductive methylation of cobalamin. The third is glycine *N*-Mtase which does have a region structurally very similar to the consensus AdoMet-binding regions, though that is not where AdoMet was bound in the reported structure (15).

AdoMet binding and target base binding sites are structurally similar to one another

The M.*Pvu*II protein has approximate two-fold pseudo symmetry around the center of the cleft, due in part to the structural similarity of the AdoMet binding site to the active site. These sites are each dominated by comparable α/β clusters, αZ→β1→αA→ β2→αB and αC→β4→αD→β5→αE; the former includes motifs I, II and X and forms the bulk of the AdoMet binding region and the latter includes motifs IV–VI and forms the bulk of the active site region. The two α/β clusters can be superimposed by rotating strands β1 and β2 onto strands β4 and β5 (Fig. 4b). This yields an r.m.s. deviation of 0.7 Å for the Cα atoms of these β-strands. Similar superimposability has also been observed for the α/β clusters of the 5mC Mtases M.*Hha*I and M.*Hae*III and the N6mA Mtase M.*Taq*I (10). This observation has led to the suggestion that the original Mtases arose after gene duplication converted an AdoMet binding protein into a protein that bound two molecules of AdoMet (see also 39–42) and that the two halves then diverged (10). Regardless of the evolutionary model, the M.*Pvu*II structure suggests that this internal structural repeat is a feature common to most AdoMet-dependent Mtases. Only the reactivation domain of *E.coli* methionine synthase does not fit this pattern (38).

DISCUSSION

Predicted DNA binding and base flipping

It is very likely that the V-shaped cleft of the protein is where DNA binds. In the absence of large scale protein conformational changes, the cleft is large enough to accommodate double-stranded DNA without steric hindrance (Fig. 1b). Positively charged groups, capable of interacting with the DNA phosphate backbone, are prominent on the surface of the cleft from the P loop (Arg60–Lys–Lys62), loop 5-E (Lys103 and Arg108) and loops 6–7 (Lys138, Lys148–Arg–Lys150, Arg152 and Lys154). We have docked a 13mer B-DNA duplex containing the *Pvu*II recognition sequence, taken from the R.*Pvu*II–DNA structure (19), against the basic face of the cleft (Fig. 1b). The fit of the DNA in the cleft is extremely convincing, with the protein occupying a distance of ~37 Å along the axis of the double helix, which suggests that M.*Pvu*II intimately contacts a 10 nt stretch including the 6 nt recognition sequence.

The M.*Hha*I–DNA structure provided the first example of base flipping (1). Several other types of enzymes are now also known or believed to use this approach (43,44), including the DNA repair enzymes T4 endonuclease V and human uracil-DNA glycosylase (45,46). The M.*Pvu*II structure is consistent with a base flipping mechanism. Base flipping is a process by which an enzyme can rotate a DNA nucleotide out of the double helix, breaking only the base pairing hydrogen bonds and trapping it in a protein binding pocket. In our docking model the DNA is positioned such that the target cytosine is in the helix and the NH_2 group to be methylated is far from the active CH_3 group of AdoMet (Fig. 1b). Thus it is likely that M.*Pvu*II (an amino-Mtase) flips the cytosine out of the DNA helix to access the target amino group (Fig. 1c), in a manner similar to that employed by 5mC Mtases, M.*Hha*I and M.*Hae*III (1–4). The structure of M.*Taq*I and spectroscopic data for M.*Eco*RI suggest that these two amino-Mtases flip the target adenine out of DNA (47,48).

Although it is possible to predict where DNA binds, we cannot identify any known DNA binding motifs in the current structure that might be responsible for DNA sequence specificity. Furthermore, there is no obvious similarity between M.*Pvu*II and the structures of R.*Pvu*II or MyoD in complex with DNA (19,49); both are homodimeric proteins recognizing the same DNA sequence, CAGCTG. R.*Pvu*II uses a β-ribbon motif to interact with nucleotides in the DNA major groove, while the myogenic
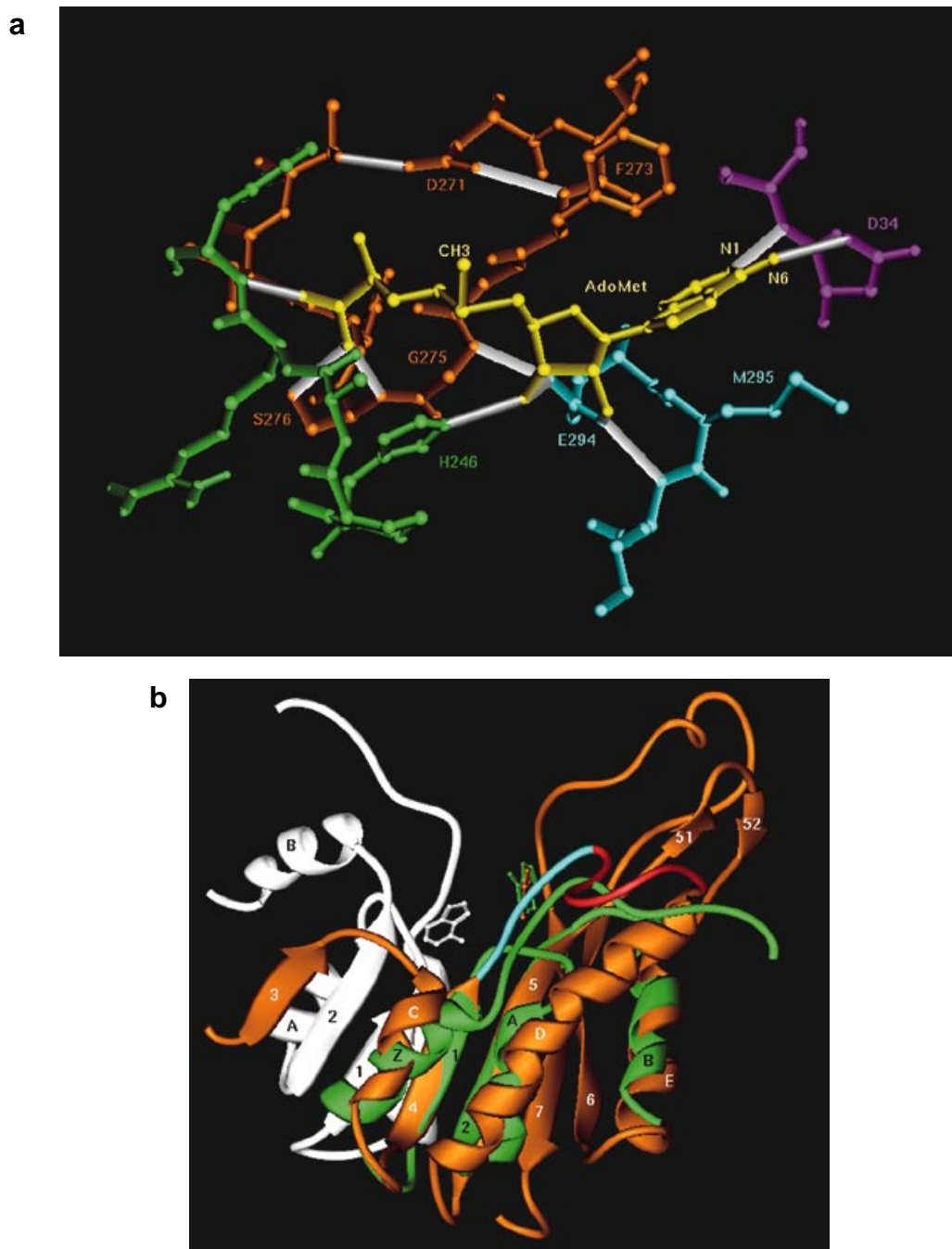
Figure 4. AdoMet binding site. (**a**) AdoMet is involved in contacts with four regions, formed by four motifs, the G loop (motif I in brown), strand β (motif II in pale blue), helix αC (motif III in purple) and loop F-Z (motif X in green). (**b**) Superimposition of the two α /β clusters. The first cluster, αZ→β1→αA→β2→αB, in white, is rotated (now in green) with respect to the second cluster, αC→β4→αD→β5→αE, in brown, to achieve the most overlap possible. Also shown are the positions, relative to the respective α/β clusters, of the AdoMet adenosyl moiety (white and green) and the target cytosine ring, in brown (inferred from M.*Hha*I–DNA structure, see Fig. 5).

transcription factor MyoD is a basic helix–loop–helix protein. The lack of obvious similarity may reflect the disparate roles of these three CAGCTG-recognizing proteins. DNA Mtases carry out base flipping (within specific nucleotide sequences) so they can access the atom to be methylated on the target nucleotide. Such a mechanism is not required for other sequence-specific proteins, such as transcription factors (for which specific binding

is the main role) and restriction endonucleases (which only act on the readily accessible DNA phosphate backbone).

As mentioned before, only two 5mC Mtases, M.*Hha*I and M.*Hae*III, have been structurally characterized in complex with their DNA substrates. The protein–base contacts in the recognized sequence are expected to differ between M.*Hha*I and M.*Hae*III due to their different specificity and, indeed, the folding of the

corresponding TRDs is different (2). However, both TRDs contain a shared feature: two recognition loops (1,2,44). In the M.*Pvu*II structure, two loops (prior to and after helix αF) on the other side of the V-shaped cleft could easily fit into the concave face of the major or minor groove of B-form DNA. These two loops, which may correspond to the two 5mC recognition loops, are held in place through scaffolding made up of three helices, αF, αB and αB$_1$. A similar pair of recognition loops has also been proposed for M.*Taq*I (47). The reason for such conservation may be that sequence recognition is a part of the base flipping mechanism and loops, instead of the more rigid structures of α-helix or β-strand, are used for discriminating DNA sequences flexibly and effectively.

Predicted catalytic mechanism for DNA amino methylation

What we call the catalytic P loop of the amino-Mtases was found in early sequence comparisons and called an 'Asp–Pro–Pro–Tyr motif' based on its sequence (50,51). A later comparison suggested it might correspond to Pro–Cys (motif IV) in 5mC Mtases, even though the reaction mechanisms of the two families of Mtases appear to be quite distinct (52). The structural comparison of M.*Hha*I and M.*Taq*I has confirmed that the Pro–Cys and Asn–Pro–Pro–Tyr motifs of these two enzymes are spatially equivalent (12) and thus, by analogy, are referred to as motif IV (10). Motif IV has the consensus sequence Ser–Pro–Pro–Tyr for N4mC Mtases, Asp–Pro–Pro–Tyr for groups α and β N6mA Mtases and Asn–Pro–Pro–Tyr for group γ N6mA Mtases (10,53,54). However, as we discuss below, Ser→Asp→Asn must not present an essential functional difference. We note that these consensus sequences are not absolute and there is still a problem in distinguishing N4mC from N6mA Mtases just on the basis of amino acid sequence (see Fig. 3).

The flipped cytosine, taken from the M.*Hha*I–DNA structure, can be docked surprisingly well into the M.*Pvu*II active site, located at the bottom of the V-shaped cleft. By superimposition of the common α/β-sheet structures, the active site amino acids in M.*Hha*I from the catalytic P loop and strands β5 and β7 overlap the corresponding amino acids in M.*Pvu*II: Gly78–Phe–Pro–Cys81 onto Ser53–Pro–Pro–Phe56 (P loop), Glu119 onto Asp96 (β5) and Arg165 onto Asn158 (β7) (Fig. 5a). In M.*Hha*I these amino acids interact with the target cytosine: Arg165 interacts with O2, Glu119 with N3 and N4, the main chain carbonyl of Phe79 with N4 and Cys81 covalently bonds to C6. Though M.*Pvu*II also interacts with cytosine, we do not observe the identical amino acids in the same structural elements in M.*Pvu*II. However, as noted above, different amino acids are spatially equivalent in the two enzymes.

One can easily model the interactions between the polar edge of the flipped cytosine and M.*Pvu*II (shown in brown in Fig. 5a). The target atom, cytosine N4, would have two possible hydrogen bond partners: the hydroxyl group of Ser53 and the main chain carbonyl of Pro54 (the first two amino acids of the highly conserved motif IV). Also from this conserved motif, the phenyl ring of Phe56 could make van der Waals contacts with the cytosine ring; Phe56 occupies a position similar to Cys81 of M.*Hha*I. Asn158 (β7), which does not appear to be conserved among the N4mC Mtases, might hydrogen bond to cytosine O2.

Asp96 (Asn in most of the other N4mC Mtases) may hydrogen bond with and activate the Ser53 hydroxyl group (Asp96:O$_{δ2}$...Ser53:O$_γ$ = 2.7 Å), thereby facilitating proton

transfer from the cytosine amino group through the Ser and eventually to the Asp (Fig. 6a). If this occurs, the protonated Asp96 might then hydrogen bond to the N3 of the cytosine. Ser53 and Asp96 thus appear to belong to a charge relay system analogous to that seen in the serine proteinases (55).

Most importantly, the distance of the AdoMet methyl group to the cytosine N4 is ~4 Å in our docking model, sufficiently close to permit methyl group transfer. For comparison, in the structures of M.*Hha*I–DNA complexes the substrate cytosine C5–AdoMet methyl distance is ~2.9 Å (56); the product 5mC methyl–AdoHcy sulfur is also ~2.9 Å (4). Thus, our model suggests that methylation of the exocyclic amino group results from a direct attack of the activated cytosine N4 on the AdoMet methyl group, in analogy with the previously proposed mechanism for DNA adenine methylation (12,57,58).

In the group β N4mC Mtases, Ser53 of M.*Pvu*II is conserved except in M.*Bam*HI, which has Asp at this position (Fig. 3); a conserved Asp is present in the same place in the group β N6mA Mtases, as well as in the group α N6mA Mtases (10). Modeling suggests that Asp in this position of the P loop could interact with cytosine N4 and N3 (M.*Bam*HI) or adenine N6 and N1 (Fig. 6b). The Asp carboxyl group could hydrogen bond with cytosine N4 (NH$_2$) or adenine N6 (NH$_2$), thereby increasing the nucleophilicity of the nitrogen and serving as a trap for the amino-leaving proton, when the methyl group transfers to the nitrogen from AdoMet. In that case the protonated carboxyl group could hydrogen bond with cytosine N3 or adenine N1. If this is correct, the conserved Asp in M.*Bam*HI and the N6mA Mtases may be functionally comparable to Asp96 in M.*Pvu*II. Ser53 in M.*Pvu*II may compensate for the fact that Asp96 is too far from the cytosine N4 for direct interaction (Figs 5a and 6a), but this does not explain why most N4mC Mtases do not simply have Asp in place of Ser, as is seen in M.*Bam*HI.

When the structures of M.*Pvu*II and M.*Taq*I, a group γ N6mA Mtase, are superimposed at their common α/β-sheet structures, Asn105 of Asn–Pro–Pro–Tyr (P loop) in M.*Taq*I is present in place of Ser53 of Ser–Pro–Pro–Phe in M.*Pvu*II; and two hydrophobic amino acids, Phe196 (loops 6–7) and Val163 (β5), of M.*Taq*I replace the positions of two polar/charged groups (Asn158 and Asp96) of M.*Pvu*II. These hydrophobic amino acids, particularly Phe196, are likely to make van der Waals contacts with the target nucleotide (12). The carboxamide of M.*Taq*I Asn105 could interact with both adenine N6 and N1 (Fig. 6c), similar to the role Asn229 of thymidylate synthase plays in hydrogen bonding to dUMP (see figure 3 of 59). However, Asn229 of thymidylate synthase plays a contributory but non-essential role in catalysis (60). In contrast to Asp or Ser, it is unlikely that Asn can accept a proton. Therefore, the only role obvious at the present time played by the Asn of the group γ N6mA Mtases is in positioning the substrate adenine, while the methylation would result from a direct attack of the AdoMet methyl group on the adenine N6 with a general base (possibly a highly ordered water molecule) assisting the proton transfer that occurs at N6.

A second AdoMet molecule

Some Mtases, including M.*Pvu*II, appear to bind two molecules of AdoMet (34,61,62), one of which affects the selectivity of the protein towards substrate and non-specific DNA sequences (for M.*Eco*Dam; 61,62). Extra electron density (from $2F_o − F_c$, $F_o − F_c$
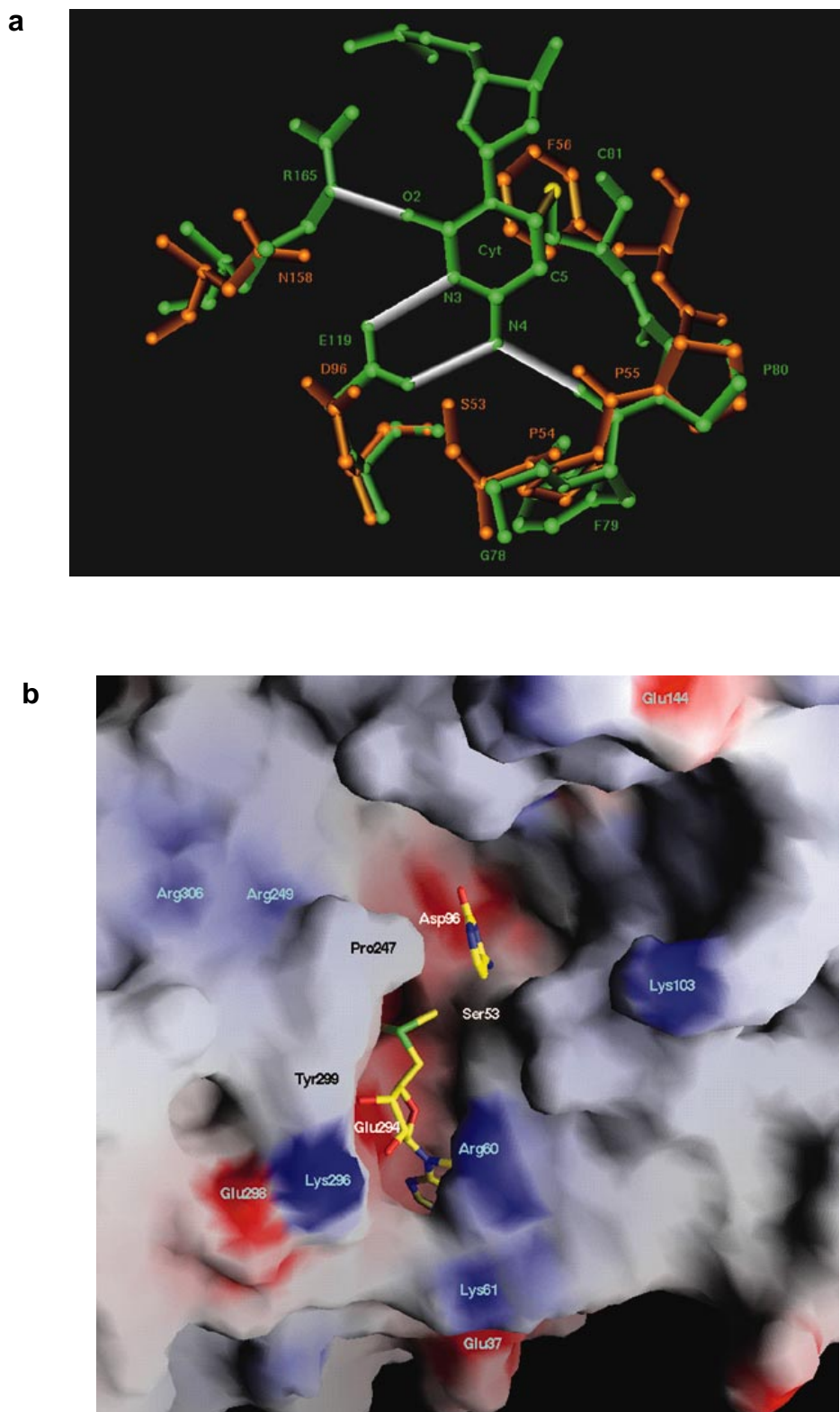
Figure 5. Active site. (**a**) Superimposition of the active sites in M.*Pvu*II (brown) and M.*Hha*I (green). Amino acids shown are from the P loop (motif IV) and strands β5 (motif VI) and β7 (motif VIII) (see Fig. 2). In the complex between M.*Hha*I and a transition state analog substrate, Cys81 is linked by a covalent bond (yellow) to C6 of the target cytosine (1). The cytosine is recognized by a number of hydrogen bonds (in white). (**b**) Close-up GRASP (71) representation displayed at the level of the solvent accessible surface. Color coded purple for positive (~20 K_BT), red for negative (~–20 K_BT) and white for neutral, where K_B is the Boltzmann's constant and T is the temperature. The AdoMet and the modeled target cytosine ring are in stick representation, with yellow for carbon, purple for nitrogen, red for oxygen, green for sulfur. The second AdoMet binding site is formed by the first AdoMet molecule at the bottom, Tyr299 and Pro247 on the left side and the main chains of Pro55 and Phe56 (between Ser53 and Arg60) on the right side.
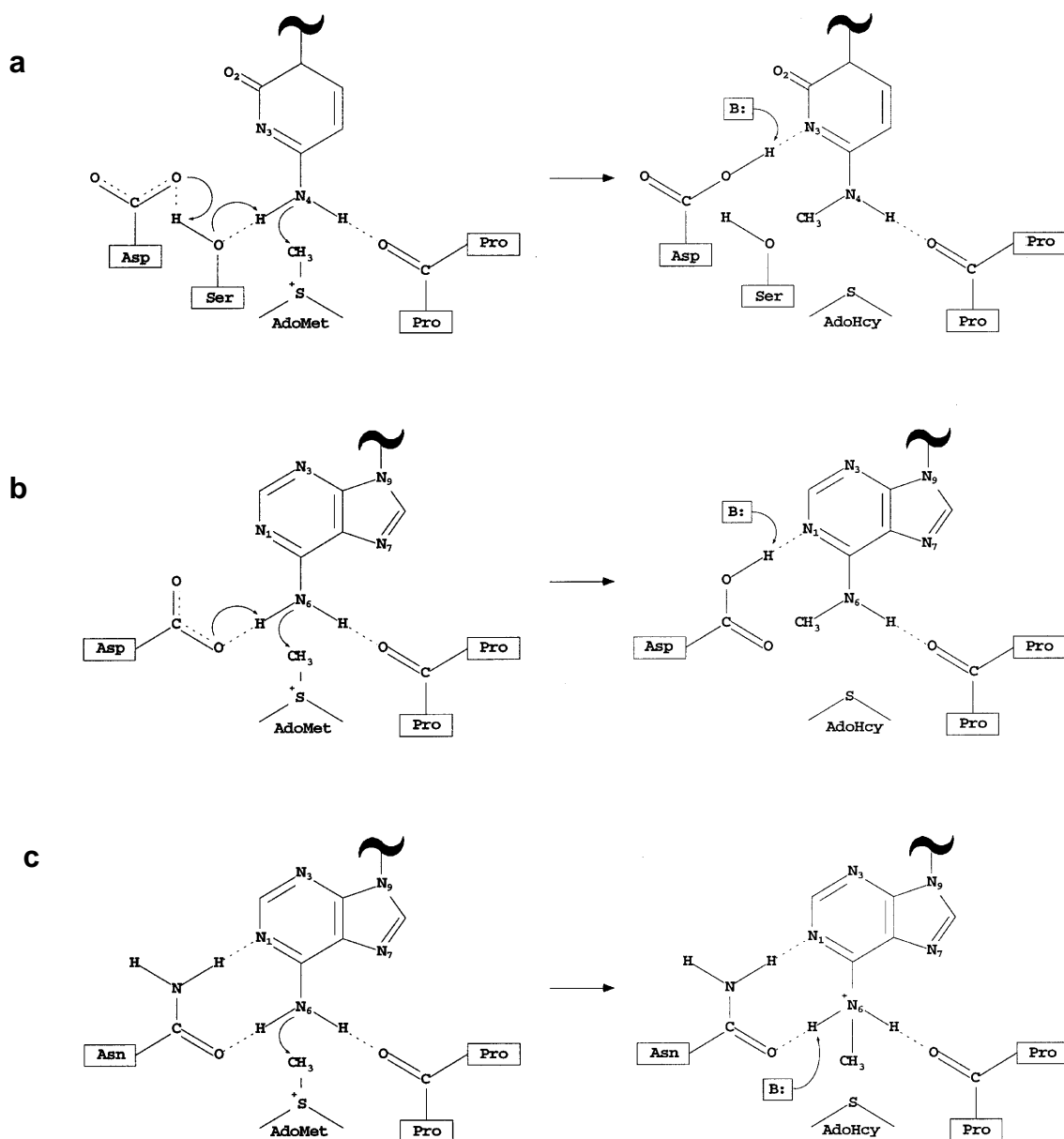
Figure 6. Proposed reaction mechanisms for amino-Mtases with the P loop containing (**a**) Ser (as in M.*Pvu*II), (**b**) Asp and (**c**) Asn. A general base (B:), which could be a water molecule, might be needed to eliminate the proton.

and initial MAD-MIRAS maps) was found near the first AdoMet in molecule A. This may be a second AdoMet, as the density can be fitted well to an AdoMet adenosyl moiety with the methionine moiety extending into the solvent. This second AdoMet binding site is formed by the first AdoMet molecule at the bottom, Tyr299 (αB) and His246–Pro247 (loop F-Z) on one side and the main chains of Pro55 and Phe56 (P loop) on the other (Fig. 5b). The adenine sits above the ribose ring of the first AdoMet. Most interestingly, the second AdoMet ribose oxygens interact with the side chain of Glu37 of crystallographic symmetry-related molecule B. This interaction, analogous to the first AdoMet–Glu294 (β2, motif II), may stabilize the second AdoMet in molecule A, due to the different crystal packing environment.

However, despite the structural similarity of the AdoMet binding and active sites (Fig. 4b), this second AdoMet molecule does not occupy the active site (Fig. 5b). Instead, the second AdoMet occupies a space equivalent to the solvent channel in the M.*Hha*I–DNA structure, where a network of well-ordered water molecules, including that proposed as the general base for eliminating the C5 proton, mediates contacts between the target cytosine, AdoHcy and M.*Hha*I (see figure 3 of 4).

Evolutionary relationships among the DNA Mtases

As noted above, the structure of M.*Pvu*II confirms two predicted features of DNA Mtase structure. First, all DNA Mtases

structurally characterized to date have AdoMet adenine binding pockets that are superimposable onto their methylatable base binding pockets (10; see Fig. 4b). Second, all DNA Mtases structurally characterized to date share a common α/β architecture for their catalytic domains, making different topological connections to accommodate the permuted linear orders of functional regions in their genes (see Fig. 2).

These two features have implications for models of the evolutionary relationships among DNA Mtases. The internal symmetry provided by the two binding pockets, each formed by a comparable set of α helices and β strands, is suggestive of evolution by gene duplication (10). Subsequent gene fusion could have converted the resulting small molecule Mtase to a DNA Mtase by adding a TRD; some DNA Mtases are still produced in two separate pieces that associate to form active enzyme and one piece is essentially the TRD while the other is the catalytic domain (35,36).

The second feature, common structure despite permuted gene orders, raises a question. Do the four groups of DNA Mtases (α, β, γ and 5mC) represent divergence from a common ancestor or convergence from separate Mtase lineages? Matthews *et al*. (63) have proposed a set of six criteria for distinguishing divergence from convergence: the DNA sequences of the genes should be similar, the amino acid sequences of the proteins should be similar, the three-dimensional structures should be similar, the enzyme–substrate interactions should be similar, the catalytic mechanisms should be similar and '...those segments of the polypeptide chain that are critical for catalysis are in the same sequence in the respective proteins (i.e. insertions and deletions are allowed, but not transpositions)'. There is as yet no structure for a Mtase of the α group, but Mtases from the other three groups (where the information is known) satisfy all except the last criterion (Fig. 2): the DNA Mtase groups have the major functional regions in three permuted gene orders (10). We can only note that several proteins have been found to remain structurally and functionally intact following circular permutation of their genes (64–68) and that genetic mechanisms for gene permutation have been proposed (69). Whether convergence or divergence describes the relationship between the DNA Mtases, it is clear that the N4mC Mtases such as M.*Pvu*II do not represent a separate subfamily of enzymes.

NOTE ADDED IN PROOF

Since acceptance of this paper, the structure of an AdoMet-dependent protein methyltransferase has been published (72). The *Salmonella typhimurium* CheR protein matches the consensus Mtase structure very well, including the binding of AdoHcy in the expected AdoMet pocket.

ACKNOWLEDGEMENTS

REFERENCES

1 Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng.X. (1994) *Cell*, **76**, 357–369.
2 Reinisch,K.M., Chen,L., Verdine,G.L. and Lipscomb,W.N. (1995) *Cell*, **82**, 143–153.
3 O'Gara,M., Klimasauskas,S., Roberts,R.J. and Cheng,X. (1996) *J. Mol. Biol.*, **261**, 634–645.
4 O'Gara,M., Roberts,R.J. and Cheng,X. (1996) *J. Mol. Biol.*, **263**, 597–606.
5 Posfai,J., Bhagwat,A.S., Posfai,G. and Roberts,R.J. (1989) *Nucleic Acids Res.*, **17**, 2421–2435.
6 Lauster,R., Trautner,T.A. and Noyer-Weidner,M. (1989) *J. Mol. Biol.*, **206**, 305–312.
7 Som,S., Bhagwat,A.S. and Friedman,S. (1987) *Nucleic Acids Res.*, **15**, 313–332.
8 Cheng,X., Kumar,S., Posfai,J., Pflugrath,J.W. and Roberts,R.J. (1993) *Cell*, **74**, 299–307.
9 Kumar,S., Cheng,X., Klimasauskas,S., Mi,S., Posfai,J., Roberts,R.J. and Wilson,G.G. (1994) *Nucleic Acids Res.*, **22**, 1–10.
10 Malone,T., Blumenthal,R.M. and Cheng,X. (1995) *J. Mol. Biol.*, **253**, 618–632.
11 Labahn,J., Granzin,J., Schluckebier,G., Robinson,D.P., Jack,W.E., Schildkraut,I. and Saenger,W. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 10957–10961.
12 Schluckebier,G., O'Gara,M., Saenger,W. and Cheng,X. (1995) *J. Mol. Biol.*, **247**, 16–20.
13 Vidgren,J., Svensson,L.A. and Liljas,A. (1994) *Nature*, **368**, 354–358.
14 Hodel,A.E., Gershon,P.D., Shi,X. and Quiocho,F.A. (1996) *Cell*, **85**, 246–257.
15 Fu,Z., Hu,Y., Konishi,K., Takata,Y., Ogawa,H., Gomi,T., Fujioka,M. and Takusagawa,F. (1996) *Biochemistry*, **35**, 11985–11993.
16 Gingeras,T.R., Greenough,L., Schildkraut,I. and Roberts,R.J. (1981) *Nucleic Acids Res.*, **9**, 4525–4536.
17 Blumenthal,R.M., Gregory,S.A. and Cooperider,J.S. (1985) *J. Bacteriol.*, **164**, 501–509.
18 Butkus,V., Klimasauskas,S., Petrauskiene,L., Maneliene,Z., Lebionka,A. and Janulaitis,A.A. (1987) *Biochim. Biophys. Acta*, **909**, 201–207.
19 Cheng,X., Balendiran,K., Schildkraut,I. and Anderson,J.E. (1994) *EMBO J.*, **13**, 3927–3935.
20 Athanasiadis,A., Vlassi,M., Kotsifaki,D., Tucker,P.A., Wilson,K.S. and Kokkinidis,M. (1994) *Nature Struct. Biol.*, **1**, 469–475.
21 O'Gara,M., Adams,G.M., Gong,W., Kobayashi,R., Blumenthal,R.M. and Cheng,X. (1997) *Eur. J. Biochem.*, in press.
22 Otwinowski,Z. (1993). In Sawyer,L., Issacs,N. and Bailey,S (eds), *Data Collection and Processing*. SERC Daresbury Laboratory, Warrington, UK, pp. 56–62.
23 Hendrickson,W.A. (1991) *Science*, **254**, 51–58.
24 Wang,B.C. (1985) *Methods Enzymol.*, **115**, 90–112.
25 Furey,W. and Swaminathan,S. (1996) *Methods Enzymol.*, **277**, in press
26 Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgard,M. (1991) *Acta Crystallogr.*, **A47**, 110–119.
27 Brünger,A.T. (1992) *X-PLOR. A System for X-ray Crystallography and NMR*, version 3.1. Yale University, New Haven, CT.
28 Ramakrishnan,V., Finch,J.T., Graziano,V., Lee,P.L. and Sweet,R.M. (1993) *Nature*, **362**, 219–223.
29 Ramakrishnan,V. and Biou,V. (1996) *Methods Enzymol.*, **276A**, 538–557.
30 Bränden,C.-I. (1980) *Q. Rev. Biophys.*, **13**, 317–338.
31 Gilliland,G.L. and Quiocho,F.A. (1981) *J. Mol. Biol.*, **146**, 341–362.
32 Rees,D.C., Lewis,M. and Lipscomb,W.N. (1983) *J. Mol. Biol.*, **168**, 367–387.
33 Brick,P., Bhat,T.N. and Blow,D.M. (1988) *J. Mol. Biol.*, **208**, 83–98.
34 Adams,G.M. and Blumenthal,R.M. (1997) *Biochemistry*, in press.
35 Karreman,C. and de Waard,A. (1990) *J. Bacteriol.*, **172**, 266–272.
36 Lee,K.-F., Kam,K.-M. and Shaw,P.-C. (1995) *Nucleic Acids Res.*, **23**, 103–108.
37 Somers,W.S. and Phillips,S.E.V. (1992) *Nature*, **359**, 387–393.

38 Dixon,M.M., Huang,S., Matthews,R.G. and Ludwig,M. (1996) *Structure*, **4**, 1263–1275.
39 Lauster,R. (1988) *Gene*, **74**, 243.
40 Lauster,R. (1989) *J. Mol. Biol.*, **206**, 313–321.
41 Tao,T., Walter,J., Brennan,K.J., Cotterman,M.M. and Blumenthal,R.M. (1989) *Nucleic Acids Res.*, **17**, 4161–4175.
42 Guyot,J.-B. and Caudron,B. (1994) *C.R. Acad Sci. Paris III Sci. Vie.*, **317**, 20–24.
43 Roberts,R.J. (1995) *Cell*, **82**, 9–12.
44 Cheng,X. and Blumenthal,R.M. (1996) *Structure*, **4**, 639–645.
45 Vassylyev,D.G., Kashiwagi,T., Mikami,Y., Ariyoshi,M., Iwai,S., Ohtsuka,E. and Morikawa,K. (1995) *Cell*, **83**, 773–782.
46 Slupphaug,G., Mol,C.D., Kavli,B., Arvai,A.S., Krokan,H.E. and Tainer,J.A. (1996) *Nature*, **384**, 87–92.
47 Schluckebier,G., Labahn,J., Granzin,J., Schildkraut,I. and Saenger,W. (1995) *Gene*, **157**, 131–134.
48 Allan,B.W. and Reich,N.O. (1996) *Biochemistry*, **35**, 14757–14762.
49 Ma,P.C., Rould,M.A., Weintraub,H. and Pabo,C.O. (1994) *Cell*, **77**, 451–459.
50 Hattman,S., Wilkinson,J., Swinton,D., Schlagman,S., Macdonald,P.M. and Mosig,G. (1985) *J. Bacteriol.*, **164**, 932–937.
51 Chandrasegaran,S. and Smith,H.O. (1988) In Sarma,R.H. and Sarma,M.H. (eds), *Structure and Expression: from Proteins to Ribosomes*. Adenine Press, New York, NY, Vol. 1, pp. 149–156.
52 Klimasauskas,S., Timinskas,A., Menkevicius,S., Butkiene,D., Butkus,V. and Janulaitis,A. (1989) *Nucleic Acids Res.*, **17**, 9823–9832.
53 Wilson,G.G. (1992) *Methods Enzymol.*, **216**, 259–279.
54 Wilson,G.G. and Murray,N.E. (1991) *Annu. Rev. Genet.*, **25**, 585–627.
55 Warshel,A., Naray-Szabo,G., Sussman,F. and Hwang,J.K. (1989) *Biochemistry*, **28**, 3629–3637.
56 Kumar,S., Horton,J.H., Jones,G.D., Walker,R.T., Roberts,R.J. and Cheng,X. (1997) *Nucleic Acids Res.*, **25**, 2773–2783.
57 Pogolotti,A.L., Ono,A., Subramaniam,R. and Santi,D.V. (1988) *J. Biol. Chem.*, **263**, 7461–7464.
58 Ho,D.K., Wu,J.C., Santi,D.V. and Floss,H.G. (1991) *Arch. Biochem. Biophys.*, **284**, 264–269.
59 Gerlt,J.A. (1994) *Curr. Opin. Struct. Biol.*, **4**, 593–600.
60 Liu,L. and Santi,D.V. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 8604–8608.
61 Bergerat,A. and Guschlbauer,W. (1990) *Nucleic Acids Res.*, **18**, 4369–4375.
62 Bergerat,A., Guschlbauer,W. and Fazakerley,G.V. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 6394–6397.
63 Matthews,B.W., Remington,S.J., Grhtter,M.G. and Anderson,W.F. (1981) *J. Mol. Biol.*, **147**, 545–558.
64 Luger,K., Hommel,U., Herold,M., Hofsteenge,J. and Kirschner,K. (1989) *Science*, **243**, 206–210.
65 Buchwalder,A., Szadkowski,H. and Kirschner,K. (1992) *Biochemistry*, **31**, 1621–1630.
66 Hahn,M., Piotukh,K., Borriss,R. and Heinemann,U. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 10417–10421.
67 Ritco-Vonsovici,M., Minard,P., Desmadril,M. and Yon,J.M. (1995) *Biochemistry*, **34**, 16543–16551.
68 Graf,R. and Schachman,H.K. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 11591–11596.
69 Gilbert,W. (1985) *Science*, **228**, 823–824.
70 Carson,M. (1991) *J. Appl. Crystallogr.*, **24**, 958–961.
71 Nicholls,A., Sharp,K.A. and Honig,B. (1991) *Proteins Struct. Funct. Genet.*, **11**, 281–296.
72 Djordjevic,S. and Stock,A.M. (1997) *Structure*, **5**, 545–558.