

# **Structural modelling of a type I DNA methyltransferase**

**David T.F. Dryden, Shane S. Sturrock and Markus Winter**

Institute of Cell and Molecular Biology, University of Edinburgh, Darwin Building, The King's Buildings, Mayfield  
Road, Edinburgh EH9 3JR, UK

Correspondence should be addressed to D.T.F.D.

tel: 0131 650 5376

fax: 0131 668 3870

2, 632-635

Sir -- We have identified, by amino acid sequence alignment and secondary structure prediction, a domain in the modification subunits of type I DNA restriction- modification ( R-M ) systems that is identical to that found in the simpler methyltransferases ( mtases) of type II R-M systems and which is responsible for catalysing DNA methylation. Using the known structures of type II mtases, we have constructed a tertiary structure model of this domain, and a model of the entire mtase part of a type I system bound to DNA. The models correlate very well with all of the available data on type I systems and strongly suggest an evolutionary link between all mtases in type I and type II R-M systems.

The function of a bacterial DNA R-M system is to maintain the methylation state of the chromosome by methylating the hemimethylated DNA produced during replication and restrict viral propagation by cleaving unmodified viral DNA<sup>1,2</sup>. In a type I R-M system, the restriction endonuclease, modification mtase and sequence recognition functions require different subunits, R, M and S, in one large multifunctional enzyme<sup>1,2</sup>. The subunit stoichiometry is believed to be R<sub>2</sub>M<sub>2</sub>S<sub>1</sub>, but the M<sub>2</sub>S<sub>1</sub> form can function alone as a mtase of molecular mass 170,000. The type I systems have been divided into families of related enzymes. The S subunits contain two DNA target recognition domains (TRDs) which each recognise one part of the bipartite targets of type I systems. The target for *EcoKI*, one of the best characterised type I systems, is AAC ( N<sub>6</sub> ) GTGC. The TRDs are joined by two short sequences which are conserved in all S subunits. The M subunits methylate the adenine nucleotide, one on each strand, at each of the underlined positions in the sequence above, using S-adenosyl methionine ( SAM ) as cofactor. The M subunit of *EcoKI* contains a N-terminal m\* region and a C-terminal tail region which gives *EcoKI* its preference for methylating hemimethylated targets rather than unmodified ones<sup>3,4</sup>, fig. 1a. The R subunit cleaves DNA at an apparently random site remote from the target sequence after extensive DNA translocation driven by ATP hydrolysis<sup>1,2</sup>.

The type II R-M systems comprise separate endonuclease and mtase enzymes<sup>1,2,5,6</sup>. The tertiary structure of *HhaI* C5 cytosine mtase and *TaqI* N6 adenine mtase<sup>7,8</sup> shows that this class of mtases, with molecular masses between 20,000 and 40,000, generally contain only one TRD and a highly conserved catalytic domain which binds the SAM<sup>9,10</sup>. The conservation of sequence motifs and structural domains strongly suggests evolution from a common ancestor<sup>10,11</sup>. In contrast to the type I systems, the type II mtases generally

recognise short, symmetric DNA targets and methylate unmodified DNA as efficiently as hemimethylated DNA.

We have compared the sequences of all the M subunits of type I systems with the sequences of the  $\gamma$  class of type II N6 adenine mtases<sup>5</sup>, fig 1b. This identified the same six sequence motifs in the M subunits as found in the type II  $\gamma$  class mtases<sup>12</sup>. Secondary structure predictions over the 200 amino acid region containing the six conserved blocks show  $\alpha$  helix- $\beta$  strand- $\alpha$  helix repeats over the whole region, fig 1c. The structural features predicted by the PredictProtein PHD program<sup>13</sup> line up well with those actually found in the crystal structure of the *TaqI* mtase catalytic domain<sup>8</sup>. The insertions and deletions in the type I sequences, shown in fig. 1c., occur on surface loops in the *TaqI* structure. This alignment of sequence and secondary structure suggests that the M subunits of type I systems contain a catalytic domain virtually identical to that in *TaqI*. Using our sequence alignments and the *TaqI* tertiary structure, we have modelled the tertiary structure of the *EcoKI* M subunit catalytic domain, fig 2a. The model domain comprises two independent subassemblies joined by a loop between  $\beta$  strands 3 and 4. The first subassembly has the form of the Rossmann mononucleotide binding fold<sup>14</sup>. The second subassembly is similar except for antiparallel  $\beta$  strands, 6 and 7, near the end of the domain. The model domain indicates plausible locations for the sites of mutagenesis<sup>15</sup> and proteolysis<sup>4</sup> which have been examined in *EcoKI*, table 1. One can see that G177 and N266 are located close to the cofactor and that F269 is poised at the edge of the active site where it can interact with the target base if it is flipped out of the DNA helix. Proteolysis yields stable polypeptides from the cleavage sites to the C terminus of the M subunit<sup>4</sup>. Proteases can only reach the sites buried on the  $\beta$  sheet when the N-terminal half of the subunit has been digested and the SAM binding site lost. The exposed loop containing the R279 - V282 site is protected by SAM binding from digestion, suggesting a conformational change possibly analogous to that observed for the similar loop in *HhaI* mtase<sup>7</sup>.

Following the recent suggestions that a type I mtase has approximate two-fold rotational symmetry<sup>4, 16, 17</sup>, we have constructed a model of the entire mtase part of a type I system bound to DNA using two copies of the type II *HhaI* mtase structure, figure 2b. The location of the start of the modelled domain, which contains two m\* mutation sites at S144 and R153<sup>3</sup>, suggests that the N-terminal m\* region is folded up against the back of the

catalytic domain. The tail of the M subunit is sensitive to the methylation state of the DNA target<sup>4</sup>. To allow one catalytic domain, bound at one half of the DNA target, to be sensitive to the methylation state of the other half of the target, it appears necessary to place the tail of each M subunit against the DNA and TRD forming the second half of the target site, fig. 2c. In the absence of further structural information, the m\* region and the tail region are shown as a single block for each M subunit in fig. 2c.

In addition to being consistent with biochemical results, including recent DNA footprinting<sup>18</sup> and protein-DNA crosslinking<sup>19</sup> experiments, the model is consistent with measurements of the hydrodynamic shape of the protein<sup>17, 20</sup>. The effect of DNA binding to the *EcoR124I* type I mtase on small angle X-ray scattering is particularly striking and has been interpreted as a clamping of the DNA between the M subunits<sup>17</sup> exactly as shown in fig. 2b, c. The proposed contacts between the M subunits and both parts of the DNA target sequence could, by altering the subunit orientation in response to target methylation, give *EcoKI* its specificity for methylating half-methylated DNA and cleaving unmodified DNA by the whole enzyme, as originally predicted by Burckhardt *et al*<sup>21</sup>.

The apparent success of the modelling procedure also strongly suggests an evolutionary link between all mtases in type I and type II R-M systems. Almost all of the known R-M systems are of the type II class which contain a simple modification mtase, but recent results suggest that the more complex type I R-M systems may also be widespread in nature<sup>2, 22</sup>. The link between type I and II mtases may be extended to suggest that all of the DNA mtases are constructed from one or more copies of the catalytic domain and one or more TRDs. These domains can be on the same polypeptide chain or on separate subunits. In some DNA mtases, namely the C5 cytosine<sup>6</sup>, N6 adenine  $\beta$  class<sup>5</sup> and type III N6 adenine mtases<sup>1, 2</sup>, the catalytic domain is interrupted by the TRD. This could be accommodated structurally due to the nature of the two substructures within the catalytic domain and could arise by gene duplication and rearrangement. The presence of additional domains and subunits in type I and type III R-M systems compared to the type II R-M systems allows the acquisition of a more sophisticated range of enzymatic responses on binding to a DNA target.

## **Acknowledgements**

This work would not have been possible without the generous provision of X ray coordinates and information prior to publication by G. Schluckebier and Professor W. Saenger (Berlin) and Dr. X. Cheng (Cold Spring Harbor). We are also grateful to Professor N.E. Murray and Dr. A. Coulson (Edinburgh) for their support. M. Winter is supported by a Darwin Trust Studentship, S.S. Sturrock by a BBSRC CASE award in conjunction with MasPar Computer Corporation, and D.T.F. Dryden by a Royal Society University Research Fellowship.

**Fig. 1. (a)** The M subunit (529 amino acids) of *EcoKI* showing the N-terminal region containing the m\* mutations (153 amino acids long)<sup>3</sup>, the region of homology with the catalytic domain of type II mtases (200 amino acids) and the tail region (~150 amino acids)<sup>4</sup>.

**(b)** Sequence alignment showing six conserved blocks in the archetypal M subunits of the three type I families, the M subunit of *Mycoplasma pulmonis*<sup>22</sup>, the first non-enteric bacterial type I R-M system and *TaqI* type II mtase, the archetypal member of the  $\gamma$  class of N6 adenine mtases<sup>5</sup>. Six additional M subunits and 14 type II  $\gamma$  class mtases were also included in the alignment but are not shown since they are almost identical to the archetypal M subunits<sup>1,2</sup> or their alignment is similar to that previously published<sup>12</sup>. The numbers in brackets refer to the number of amino acids not shown and which are not in the conserved blocks.

**(c)** Secondary structure predictions<sup>13</sup>, for the region in **(b)** above for the type I families and the type II  $\gamma$  class of mtases, compared with the secondary structure observed in the catalytic domain of the *TaqI* mtase crystal structure<sup>8</sup>. Arrows and rectangles indicate  $\beta$  strands and  $\alpha$  helices respectively. The strands and helices in the *TaqI* structure are numbered as in reference 10.

**Fig. 2. (a)** A tertiary structure model, constructed using FRODO<sup>23</sup> and SYBYL ( Tripos ) software, of the catalytic domain from amino acids 141 to 380 in the M subunit of *EcoKI* based on the sequence alignment in Fig. 1 and the structure of the *TaqI* mtase<sup>8</sup>. The  $\alpha$  helices,  $\beta$  strands and SAM are coloured magenta, yellow and blue respectively. The helices and strands are numbered as given in fig. 1c. The sites of mutagenesis and proteolysis are coloured red and green respectively and numbered as in table 1.

**Fig. 2. (b)** A front view of a partial model of a type I mtase bound to DNA constructed using two copies of the structure of the type II *HhaI* mtase bound to DNA<sup>7</sup>. The catalytic domain of *HhaI* has been replaced with our modelled domain with helices, strands and SAM coloured as in fig. 2a. The target bases are flipped out of the DNA helix, shown in red, to project into the catalytic domain and are 10 base pairs apart as found in the *EcoKI* recognition sequence. The grey spacefilled structures are the two TRDs proposed to form the S subunit. The actual structure of these domains in type I mtases is unknown so the diagram shows the TRD of *HhaI*. They are located in two successive major grooves in agreement with DNA footprinting and protein-DNA crosslinking experiments<sup>18, 19</sup>.

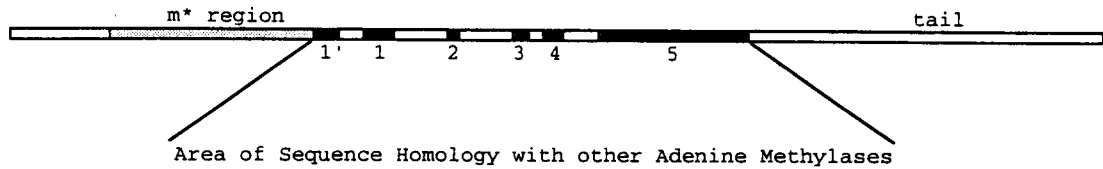
**Fig. 2. (c)** A schematic rear view of the complete mtase structure bound to DNA. The M subunits are shown in a bold outline with the postulated position of the N-terminal m\* and C-terminal tail regions and the catalytic domain indicated. The S subunit is shown shaded with the TRD joined by the two short spacer regions which are well conserved in all S subunits. The DNA helix, shown as a cylinder, lies on the S subunit and between the M subunits in agreement with earlier models<sup>17, 21</sup>.

**Table 1.** Sites of proteolysis and mutation in the putative catalytic domain of *EcoKI*.

<u>Proteolytic cleavage sites in catalytic domain<sup>4</sup></u>	<u>Comment</u>
R168     ↓ E169	Trypsin cleavage at partially exposed loop, fragment from amino acid 169 to ~440. On $\beta$ strand 1.
V263     ↓ A264	Elastase or trypsin cleavage at buried site on $\beta$ strand.
A264     ↓ T265	Fragments from cleavage site to end of M subunit. On $\beta$
R305     ↓ A306	strands 4 and 5.
R279     ↓ T280	Trypsin or chymotrypsin cleavage at exposed loop.
F281     ↓ V282	Fragments from cleavage site to end of M subunit. SAM binding prevents cleavage and stabilises whole subunit. On loop between $\beta$ strand 4 and $\alpha$ helix D.
<u>Site-directed mutations<sup>15</sup></u>	
G177D	Abolishes SAM binding, insoluble when expressed at 37°C, soluble at 25°C. Mtase structure and DNA binding normal. In motif 1 between $\beta$ strand 1 and $\alpha$ helix A.
N266D, F269G, F269C	SAM and DNA binding normal but inactive. In motif 4 at end of $\beta$ strand 4.
F269Y, F269W	SAM and DNA binding normal. Activity 25% and $\leq 5\%$ of wild-type mtase respectively.

1. Wilson, G. G. & Murray, N. E. *Annu. Rev. Genet.* **25**, 585-627 (1991).
2. Barcus, V. A. & Murray, N. E. In *Population genetics of bacteria.* (Baumberg, S., Young, J. P. W., Saunders, S. R. & Wellington, E. M. H. eds), pp. 31-58, Society for General Microbiology Symposium **52**, Cambridge University Press (1995).
3. Kelleher, T. E., Daniel, A. S. & Murray, N. E. *J. Mol. Biol.* **221**, 431-440 (1991).
4. Cooper, L. P. & Dryden, D. T. F. *J. Mol. Biol.* **236**, 1011-1021 (1994).
5. Wilson, G. G. *Methods Enzymol.* **216**, 259-279 (1992).
6. Kumar, S. *et al.* *Nucleic Acids Res.* **22**, 1-10 (1994).
7. Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. *Cell* **76**, 357-369 (1994).
8. Labahn, J. *et al.* *Proc. Natl. Acad. Sci. USA* **91**, 10957-10961 (1994).
9. Cheng, X. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 293-318 (1995).
10. Schluckebier, G., O'Gara, M., Saenger, W. & Cheng, X. *J. Mol. Biol.* **247**, 16-20 (1995).
11. Lauster, R. *J. Mol. Biol.* **206**, 313-321 (1989).
12. Noyer-Weidner, M., Walter, J., Terschüren, P.-A., Chai, S. & Trautner, T. A. *Nucleic Acids Res.* **22**, 4066-4072 (1994).
13. Rost, B. & Sander, C. *J. Mol. Biol.* **232**, 584-599 (1993).
14. Rossman, M. G., Liljas, A., Bränden, C.-I. & Banaszak, L. J. In *The Enzymes, 3rd ed. vol XI.* (Boyer, P. D. ed), pp. 61-102, Academic Press (1975).
15. Willcock, D. F., Dryden, D. T. F. & Murray, N. E. *EMBO J.* **13**, 3902-3908 (1994).
16. Meister, J. *et al.* *EMBO J.* **12**, 4585-4591 (1993)
17. Taylor, I. A., Davis, K. G., Watts, D. & Kneale, G. G. *EMBO J.* **13**, 5772-5778 (1994).
18. Powell, L. M. & Murray, N. E. *Nucleic Acids Res.* **23**, 967-974 (1995).
19. Chen, A., Powell, L. M., Dryden, D. T. F., Murray, N. E. & Brown, T. *Nucleic Acids Res.* **23**, 1177-1183 (1995).
20. Powell, L. M., Dryden, D. T. F., Willcock, D. F., Pain, R. H. & Murray, N. E. *J. Mol Biol.* **234**, 60-71 (1993).
21. Burckhardt, J., Weisemann, J., Hamilton, D. L. & Yuan, R. *J. Mol. Biol.* **153**, 425-440 (1981).
22. Dybvig, K. & Yu, H. *Mol. Microbiol.* **12**, 547-560 (1994).
23. Jones, T. A. *J. Appl. Crystallogr.* **11**, 268-272 (1978).

a. M subunit of EcoKI



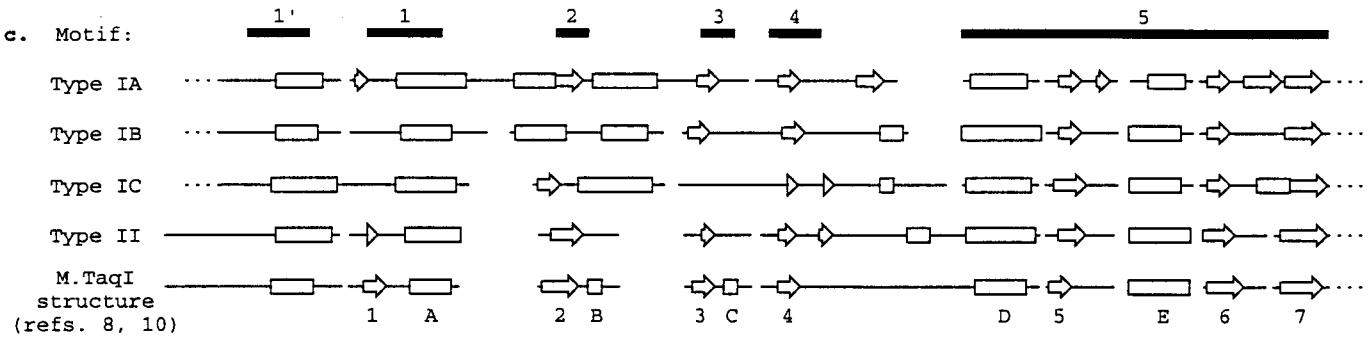
b. Motif

	1'	1	2	3	4
M. EcoKI	(146) GQYFTPRPLIKTI	(11) VQDPAAGTAGFLIEA	(26) FIGLELV	(25) IRLGNTL	(08) KAHIVATNPPF
M. EcoAI	(161) GEFYTPRAVTRFM	(11) IMPDPCGTGGFLACA	(21) IHGVEKK	(21) IRHDNTL	(09) QLDVIVTNPPF
M. EcoR124I	(196) GEFYTPQHVSCLI	(13) IYDPAAGSGSLLLQA	(12) FFGQEIN	(22) IKLGNTL	(09) PFDAIVSNPPY
M. Mycoplasma	(195) GEFYTPSKVSELL	(13) AYDPACGSGSLLIKL	(09) IYGQEVK	(22) LRSGDTL	(10) SFDCIVANPPF
M. TaqI	(18) GRVETPPEVVDPM	(11) VLEPACADGPFLLRAF	(09) FVGVEID	(11) GILADFL	(06) AFDLILGNPPY

Motif

	5
M. EcoKI	(17) NKQLCFMQHIIETL HPGGRAAVVVPDNLV F EGGKGTDIRRDLMDKCHLHTILRLPTGIFYA QGVKTNVLFFTK (165)
M. EcoAI	(19) ETADLFLQLIVEVL AKNGRAAVVLPDGLTF GEGVTKIKIKLLTEECNLHTIVRLPNGVFNPHYTGIKTNLLFFTK (120)
M. EcoR124I	(28) KADFAFVLHALNYL SAKGRAAVVCFPGIFY RGGAEQKIRQYLVDDNNYVETVISLAPNLFY TTIAVNIVLFSK (114)
M. Mycoplasma	(26) YADFAFLQHMLFHVNKDNGIIASVFSLGILSRKSPKAEDIRKYIIDKNYIDTIIIFLPPNLFYN TSIESCIIVARK (116)
M. TaqI	(32) NLYGAFLEKAVRLL KPGGVLVFVVPATWLV LEDFALLREFLARE KTSVYYLGEVFPQ KKVSAVVIRFQK (211)

c. Motif:



(refs. 8, 10)

