UNIVERSITY OF CRETE
Department of Biology

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Protein Biotechnology

STRUCTURAL STUDIES OF THE M.*Bse*CI METHYLTRANSFERASE FROM *Geobacillus stearothermophilus*

by

DIMITRIOS A. MITSIKAS

Advisor:
Dr. Michael Kokkinidis
Prof. Emeritus of Structural Biology

Examining Committee:
Dr. Irene Athanassakis, Prof. of Immunology
University of Crete, Dept. of Biology

Dr. Ioannis Pavlidis, Assist. Prof. of Biochemistry
University of Crete, Dept. of Chemistry

October 2021
Heraklion, Greece

Abstract

DNA modification methyltransferase BseCI (M.*Bse*CI) from *Geobacillus stearothermophilus*, is an adenine-specific methyltransferase, associated with a restriction endonuclease. It catalyses the methylation of the N6 atom of 3′ adenine in the sequence 5′-ATCGAT-3′ providing, by extent, a defence mechanism against bacterial infection from bacteriophages. Although cytosine-specific methyltransferases have been intensively studied, our knowledge concerning adenine-specific methyltransferases still remains very limited. The work presented here reports the determination of the crystal structure of M.*Bse*CI, both in its apo form (2.5 Å) and in complex with the cofactor product *S*-adenosyl-L-homocysteine and its cognate DNA sequence in three different methylation states, corresponding to the natural substrate (2.4 Å), the end product of the methylation reaction (2.6 Å), and the unmethylated state (2.3 Å). Like other methyltransferases, the enzyme adopts a bilobed shape and contains the characteristic methylase fold, despite the lack of sequence homology to other enzymes of that family. Surprisingly, extensive structural similarities exist even among DNA target recognition domains of previously determined N6-adenine methyltransferases, suggesting that this group of enzymes might be more uniform than had been expected. The findings reported here, other than merely adding another piece to the puzzle of methyltransferases, provide further insight into the structural basis of adenine-specific methylation and give the opportunity for structural comparisons with various previously determined methyltransferases.

[ΠΕΡΙΛΗΨΗ: Η DNA μεθυλτρανσφεράση BseCI (M.*Bse*CI) απο το βακτήριο *Geobacillus stearothermophilus*, είναι μια μεθυλτρανσφεράση αδενίνης, η οποία σχετίζεται με μια περιοριστική ενδονουκλεάση. Καταλύει τη μεθυλίωση του ατόμου N6 της 3′ αδενίνης στην αλληλουχία 5′-ATCGAT-3′ παρέχοντας, κατ' επέκταση, έναν αμυντικό μηχανισμό έναντι μόλυνσης του βακτηρίου από βακτηριοφάγους. Αν και οι μεθυλτρανσφεράσες κυτοσίνης έχουν μελετηθεί εντατικά, οι γνώσεις μας σχετικά με τις μεθυλτρανσφεράσες αδενίνης παραμένουν περιορισμένες. Στην εργασία αυτή περιγράφεται ο προσδιορισμός της τρισδιάστατης δομής της M.*Bse*CI, τόσο στη μορφή του αποενζύμου (2.5 Å), όσο και στη μορφή συμπλόκου με το συμπαράγοντα *S*-αδενοσυλ-L-ομοκυστεΐνη και το DNA-στόχο σε τρεις διαφορετικές καταστάσεις μεθυλίωσης, που αντιστοιχούν στο φυσικό υπόστρωμα (2.4 Å), στο τελικό προϊόν της αντίδρασης της μεθυλίωσης (2.6 Å), και στη μη μεθυλιωμνη κατάσταση (2.3 Å). Όπως και οι υπόλοιπες μεθυλτρανσφεράσες, το ένζυμο αποτελείται από δύο διακριτές επικράτειες και περιέχει το χαρακτηριστικό μοτίβο μεθυλάσης, παρά την έλλειψη ομολογίας στην αλληλουχία της με άλλα ένζυμα της ίδιας οικογένειας. Παραδόξως, παρατηρούνται εκτεταμένες δομικές ομοιότητες και στην περιοχή αναγνώρισης του DNA σε σχέση με άλλες λυμένες N6-μεθυλτρανσφεράσες, υποδηλώνοντας πως η συγκεκριμένη ομάδα ενζύμων ίσως να είναι πιο ομοιογενής απ' ότι αναμενόταν. Τα αποτελέσματα της δουλειάς αυτής, πέρα από την προσθήκη μιας ακόμα μεθυλτρανσφεράσης στην ήδη υπάρχουσα βιβλιογραφία, παρέχουν περαιτέρω πληροφορίες σχετικά με τη δομική βάση της N6-μεθυλίωσης και δίνουν την ευκαιρία για συγκριτική μελέτη με άλλες ήδη πάρχουσες μεθυλτρανσφεράσες.]

Acknowledgements

I am first of all indebted to my advisor, Professor Michael Kokkinidis, for the time and work he devoted to me, for sharing his wisdom and for making me part of his lovely team.

I am also grateful to Dr. Nicholas M. Glykos for all the crystallographic data he provided and —of course— for putting me on the right track back in the days when I was a young and innocent undergrad.

Many thanks to Kallia Kouyianou for letting me access her work, for providing me with her crystals, for the excellent collaboration and for all those times she tried to soothe me whenever I reached a tense impasse.

I should also thank Dr. Ioannis Pavlidis and Dr. Irene Athanassakis for their valuable comments on the first draft of this thesis.

Dina Kotsifaki for taking me by the hand to show me in practice the intricacies of wet lab, and for helping me with my first steps in the lab.

Dr. Efstratios Mylonas for the useful answers to every question I had, and for collaborating with me in other projects of the lab.

Dr. Maria Amprazi for being both a colleague and a friend, and for letting me work by her side.

Alexis Molfetas for his help and advice in all matters crystallographic and for annoying me with his 1.5 Å crystals.

Dr. Katerina Kefala for helping with the data collection and for all the funny chattering in the office, Mary Providaki for her help with gel-shift assays, Kyriaki Liouta for being a good friend and for trusting me to help her with her work, and of course every other member of the group for the nice, long chats we had while procrastinating in the lab.

My friends here in Heraklion, who made these two years enjoyable, and particularly Angeliki and Stavroula for the lovely trips we took in Crete.

My friends Christina, Nikos, Mària, Dimitra, Ioanna, Kyriakos, for being close and supportive to me, each one in their own unique way, no matter the distance separating us.

My family for their continuous moral and material support.

Table of Contents

Chapter 1

Introduction

Ever since the discovery of the DNA structure more than 60 years ago, the molecular mechanisms that preserve the genetic information encoded by DNA and guarantee its faithful inheritance across generations have been placed at the forefront of scientific interest. The numerous studies that emerged, soon led to the realisation that DNA molecules, though chemically rather inert, are typically involved in interactions with other macromolecules throughout their entire life cycle. Replication, transcription, recombination, packing, transposition, modification, all require specific interactions with proteins, clearly manifesting the importance of understanding protein-nucleic acid interactions for essentially interpreting some of the most significant biological processes.

DNA modifying enzymes play key roles in virtually all the aforementioned processes; from transcription factors exerting their regulatory effect in gene expression, to helicases and topoisomerases that alter the topology of the DNA, they indicate an elegant, albeit complex interaction pattern prevalent in all forms of life. One of the most interesting subfields within the DNA modifying enzymes family is that of DNA methylation. This base-specific modification contributes to many biological pathways including gene regulation, genomic imprinting and —speaking of prokaryotes— protection of host genetic material against foreign genome aggression.

DNA methyltransferases (MTases) comprise a family of enzymes responsible for transferring a methyl group from the cofactor *S*-adenosyl-L-methionine (AdoMet) to a pyrimidine ring carbon (C-MTases) forming C5-methylcytosine (5mC), or to the exocyclic amino groups of cytosine and adenine (N-MTases), yielding either N4-methylcytosine (N4mC) or N6-methyladenine (N6mA) [Fig. 1.1]. The methyl groups in these positions protrude towards the major groove of B-DNA duplexes in such way that they do not affect the Watson-Crick base pairing, whilst making them an easy target for recognition by proteins. By these means, methylation increases structural diversity of DNA by adding extra information that is not encoded in its sequence.

Although the identification of modified nucleobases dates back to the early 1920s —with the discovery of C5-methylcytosine (Johnson & Coghill, 1925)— DNA methylation was not reported until 1948 on calf-thymus DNA using paper chromatography (Hotchkiss, 1948). However, the underlying mechanisms of DNA methylation were first described a few years later as part of the restriction-modification (RM) systems in prokaryotes (Arber & Dussoix, 1962). A series of papers published by Gold (Gold *et al.*, 1963; Gold *et al.*, 1964; Gold & Hurwitz, 1964*a*,*b*; Hurwitz *et al.*, 1964*a*,*b*) further supported the restric-

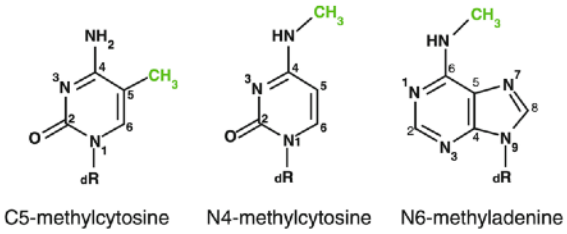

Fig. 1.1 Structures of the methylated bases that occur in DNA. Adapted from *DNA Methyltransferases - Role and Function*, by A. Jeltsch & R. Jurkowska, 2016, p. 2, Springer.

tion-modification hypothesis, providing as well substantial evidence for the occurrence and importance of AdoMet-dependent DNA and RNA methylation. The discovery of the first MTase in *Escherichia coli*, *Eco*Dam, at that time (Gold & Hurwitz, 1964*a*) —a solitary MTase not belonging in the RM system— further shifted the interest towards recombinant expression of cloned enzymes. These findings, together with the first purification (Smith & Wilcox, 1970) and application (Dana & Nathans, 1971) of a restriction endonuclease, resulted in a jointly awarded 1978 Nobel Prize in Physiology or Medicine for Werner Arber, Hamilton Smith and Daniel Nathans, sparking an era of many breakthrough discoveries in the field of DNA modifying enzymes. To date, a huge number of structural, biochemical and evolutionary studies have shed light on many important insights into MTases.

All DNA MTases follow a ternary complex mechanism where the catalytically competent complex includes the enzyme, the DNA substrate and the AdoMet cofactor, with the binding of the two latter components either being sequentially ordered or random. The discovery of the bacterial 5mC-MTase M.*Hha*I (Cheng *et al.*, 1993; Klimašauskas *et al.*, 1994) along with various studies that followed, led to a conceptual breakthrough regarding the catalytic mechanism of all DNA MTases: the to-be-methylated base is completely rotated out of the DNA helix and inserted into a catalytic pocket; a mechanism so-called as "base flipping". Comparisons of various DNA MTases also led to the discovery of a specific motif characteristic for C-MTases, which is organised upon a two-domain Rossmann fold architecture (Klimašauskas *et al.*, 1989; Lauster *et al.*, 1989; Cheng, 1995). This motif was also extensively found in other prokaryotic (Lauster *et al.*, 1987; Guschlbauer, 1988) or mammalian MTases (Bestor *et al.*, 1988), indicating a conserved signature motif for all AdoMet-dependent MTases directly involved in the catalytic process (Cheng, 1995; Martin & McMillan, 2002; Jeltsch, 2002). This fold is known today as the AdoMet-dependent MTase fold. The high similarities among all groups of MTases suggest that these enzymes are monophyletic. Presumably, the two subdomains originated from duplication of an ancestral AdoMet-binding Rossmann fold domain, one of which diverged to generate the binding sites for different methylation substrates (Malone *et al.*, 1995).

This thesis focuses on the structural studies of the *Bse*CI modification N6mA-MTase from *Geobacillus stearothermophilus*. In addition to the apoenzyme, which had already been crystallised by past members of the group, this work also reports the cocrystallisation of M.*Bse*CI with the AdoHcy cofactor and its cognate DNA duplex in three different methylation states, corresponding to the natural substrate, the end product of the methylation reaction and the unmethylated state. Contrary to C-MTases, which have been extensively studied, our knowledge concerning N-MTases is still limited. Determination and analysis of all four structures will offer the opportunity to compare the structural characteristics of various MTases, and to clarify the structural basis of adenine-specific methylation and the conformational changes that the enzyme-DNA complex undergoes during enzymatic turnover, and will further aid in understanding DNA binding and sequence discrimination by a catalytic protein.

1.1 Structural details of prokaryotic DNA methyltransferases

Three major roles of DNA methylation have been identified in prokaryotes: i) distinction of self and non-self DNA, ii) direction of post-replicative mismatch repair, and iii) control of DNA replication and cell cycle. The first is associated with RM systems and serves as a defence against infection of bacteria by virulent phages, and account for the overwhelming majority of DNA MTases in prokaryotes. In addition to a MTase, RM systems also contain a restriction endonuclease (ENase) that specifically cleaves the invading non-methylated phage DNA, while host DNA is protected by a specific methylation pattern (Jeltsch, 2002).

Table 1.1 Classification of restriction-modification (RM) systems.

Type	Composition	Cofactors	DNA restriction pattern
I	Multiple subunits	Mg^{2+} & ATP	Random cleavage far from asymmetrical recognition sites
II	Separate enzymes for restriction and modification	Mg^{2+}	Cleavage within symmetrical recognition sequences
IIs	Separate enzymes for restriction and modification	Mg^{2+}	Cleavage at fixed distance from symmetrical recognition sequences
III	Separate enzymes for restriction and modification	Mg^{2+} & ATP	Cleavage at fixed distance from symmetrical recognition sequences
IV	Several subunits	Mg^{2+} & GTP	Cleavage of DNA containing methylated nucleotides

RM systems are classified into four types according to their structural features, cofactor requirements and pattern of DNA cleavage, as summarised in Table 1.1 (Wilson & Murray, 1991; Jeltsch, 2002, Loenen *et al.*, 2014). In types I and III the DNA MTase is part of a multisubunit protein complex involved in both restriction and modification, contrary to type II systems, in which the MTase and ENase activities are exerted by two distinct enzymes, encoded by different genes. Type IV systems, on the other hand, differ from classic RM systems as they lack a MTase, restricting only modified DNA.

Regardless of their classification, all bacterial DNA MTases are two-domain proteins, with the DNA binding cleft located at their interface. Sequence analyses revealed that all C-MTases contain a set of up to ten conserved amino acid motifs (I-X) and a variable region near the C-terminus that forms part of the target recognition domain (TRD), which is responsible for sequence-specific DNA recognition. N-MTases contain only nine of these conserved sequence motifs (I-VIII and X) and are divided into three subclasses (α, β and γ) depending on variations in the relative positions of these conserved motifs. Each group is therefore dominated by a specific MTase family; groups α and γ consist almost exclusively of N6mA-MTases, while group β contains mainly N4mC-MTases and a few N6mA-MTases (Malone *et al.*, 1995; Scavetta *et al.*, 2000).

The N-terminal domain incorporates both the AdoMet binding site and the catalytic centre of the enzyme, and shares among all DNA MTases a common structural core made up of a six-stranded parallel

β-sheet, with a seventh strand inserted in an antiparallel manner between the fifth and sixth strand (6↑ 7↓ 5↑ 4↑ 1↑ 2↑ 3↑). This seven-stranded β-sheet is flanked by α-helices and the entire architecture acts as a topological switch in the middle, distinguishing two separate subdomains with Rossmann folds (Jeltsch, 2002; Martin & McMillan, 2002). As shown in Fig. 1.2, the AdoMet binding site is located in the right subdomain (β1-β3), while the binding cleft for the target base in the left one (β4-β7), and both sites form hydrophobic pockets located in equivalent positions within the subdomains (Malone *et al.*, 1995; Jeltsch, 2002). The N-terminal domain also harbours all the ten characteristic amino acid motifs, which are in-

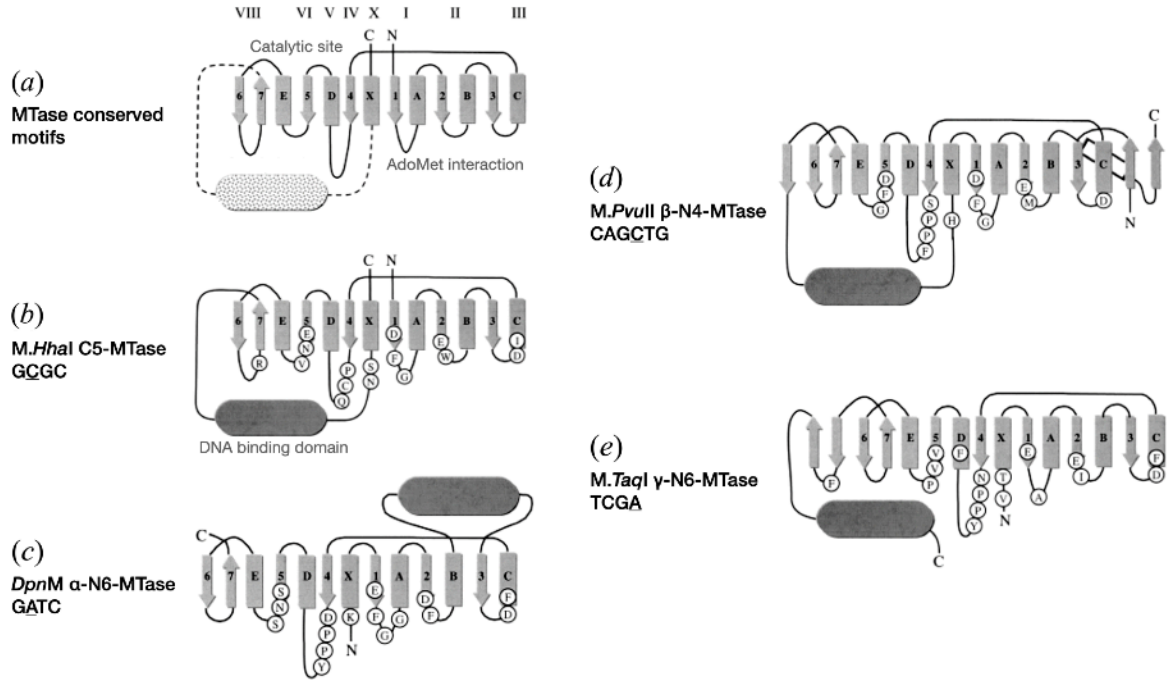

Fig. 1.2 Topology diagrams and location of conserved residues in prokaryotic DNA MTases of different families. (*a*) Schematic representation of the general structure of the DNA MTase fold. (*b*) General structure of a 5mC-MTase. (*c*)-(*e*) General structures of three N-MTases that belong in each of the three α, β and γ subgroups. Adapted from A. Jeltsch, *Chem. Eur. J. Chem. Biol.* 2002, *3*, p. 277.

volved in similar interactions across MTases, despite variations in their orientation. Besides motif I (DxFxGxG), which is universally conserved across all MTase families, motifs IV (GFPCQ) and VI (ENV) are most conserved in C-MTases, whereas motif IV ([DNS]PP[YF]) in N-MTases (Kumar *et al.*, 1994; Cheng & Roberts, 2001; Jeltsch, 2002). On the contrary, the C-terminal domain varies in amino acid sequence, size and structure among different DNA MTases. This structural heterogeneity goes in parallel with variations in function, as this domain forms many —but not all— of the sequence-specific contacts between the enzyme and the DNA, which mediate the recognition of the target site that is characteristic for each MTase (Jeltsch, 2002).

The most intriguing structural and mechanistic feature of DNA MTases is the base flipping mechanism, in which the target base is completely rotated out of the DNA helix and binds into a hydrophobic pocket. This pocket is formed by residues located in the IV, VI and VIII motifs of the N-terminal domain, at the ends of β4 and β5 and the beginning of β7, respectively (Jeltsch, 2002). Base flipping has been identified in other DNA-interacting enzymes as well, including many DNA repair enzymes (Cheng & Roberts, 2001). Perhaps, such enzymes require intimate contacts with the aromatic ring of the target nucleobases or,

in other cases, an open space for the catalytic machinery to act on the DNA backbone (Roberts, 1995; Lloyd & Cheng, 1997; Roberts & Cheng, 1998). The base flipping mechanism is also very likely to be supported by contacts of the enzyme to the phosphate groups flanking the flipped base (Cheng & Blumenthal, 1996). In any case, the structural distortions of DNA incidental to this unusual mechanism are naturally obvious, and vary from retention of an almost B-DNA-like structure (Klimašauskas *et al.*, 1994), to extensive rearrangements prompting the formation of large clefts (Reinisch *et al.*, 1995) or the compression of the DNA backbone (Goedecke *et al.*, 2001). Fig. 1.3 offers a plain view of what is just mentioned.

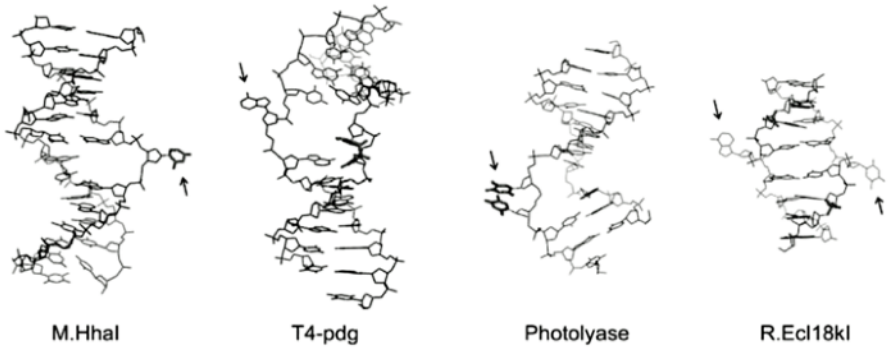

Fig. 1.3 Examples of base flipping observed in crystal structures of protein-DNA complexes. Left to the right: target base flipping (M.*Hha*I; PDB entry 1mht), opposite base flipping (T4 ENase V; PDB entry 1vas), damaged dinucleotide flipping (DNA photolyase; PDB entry 1tez), dual base flipping in a base pair (Ecl18kl ENase; PDB entry 2fqz). Arrows point at flipped bases. Protein residues are omitted for clarity. Adapted from *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution*, by H. Grosjean, 2009, p. 38, Landes Bioscience.

As for the cofactor binding site, it is remarkably conserved in all MTases, with residues from motifs I-III and X involved in contacts with almost every putative hydrogen bond donor or acceptor of the AdoMet cofactor, and in several hydrophobic interactions [Fig. 1.4] (Tran *et al.*, 1998; Jeltsch, 2002). In γ-N-MTases, motif V is also implicated in the binding of AdoMet, as a highly conserved Phe residue functionally replaces a Phe that is missing from motif I (Jeltsch, 2002). The by-product of methylation, *S*-adenosyl-L-homocysteine (AdoHcy), which does not carry a positively charged sulfur atom, binds often differently to the pocket, in a way that the homocysteine moiety interacts with the active site residues located in the IV motif (Schluckebier *et al.*, 1997). It should be noted here that AdoMet can bind to MTases in two different orientations, an active, where the transferable methyl group points towards the target base, and a non-reactive, found in structures crystallised in absence of cognate duplex oligonucleotides. The latter orientation revealed a favourable alignment for cation π-interactions with specific residues, and although the functional relevance of this observation is still unknown, it is speculated that plays a physiological role in storing AdoMet for future use, while preventing its spontaneous hydrolysis in absence of substrate DNA (O'Gara *et al.*, 1999). Another interesting finding is that some MTases include more than one AdoMet binding sites. Several instances have shown that binding of a second AdoMet can allosterically influence the active site of the MTase, but it is yet unknown whether the second cofactor can be directly channeled into the active site (Bergerat & Guschlbauer, 1990; Bergerat *et al.*, 1991; Adams & Blumenthal, 1997; Malygin *et al.*, 2001).

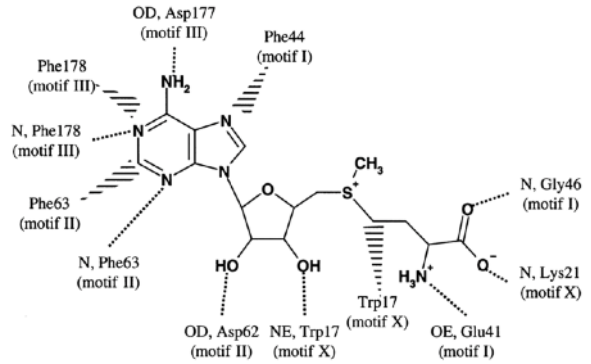

Fig. 1.4 Schematic representation of the AdoMet cofactor and its contacts in *Dpn*M MTase's binding site. Hydrogen bonds are indicated by dotted lines and hydrophobic contacts by dashed wedges. Adapted from A. Jeltsch, *Chem. Eur. J. Chem. Biol.* 2002, *3*, p. 279.

1.2 DNA methylation in prokaryotes

Despite AdoMet being a very effective donor for methyl groups, study of the catalytic mechanism of 5mC-MTases revealed that methylation in this position involves some peculiarities, as cytosine is an electron-poor aromatic system and the C5 atom is not intrinsically reactive. Particularly, the catalysis does not account for a direct spontaneous attack on the activated methylsulfonium group of AdoMet, but initially a nucleophilic attack of a Cys residue, located in the PCQ motif (motif IV), on the C6 position of the cytosine ring, forming a covalent bond between the enzyme and the substrate base. Thereby, this concomitant increase in the negative charge density at the C5 position allows the attack on the cofactor's methyl group. The structure of M.*Hha*I also revealed that the nucleophilic attack of the Cys residue might be facilitated by a transient protonation of the cytosine ring at the endocyclic N3 atom by an enzyme-derived acid (typically a conserved Glu residing in the VI motif). Additionally, an Arg residue from the RxR motif (motif VIII) might possibly be involved in the stabilisation of both the Glu and the Cys base. The addition of the methyl group to the nucleobase is followed by a deprotonation of the C5 atom, catalysed by a so far unknown proton acceptor, which breaks the covalent bond between the enzyme and the base through an elimination reaction (Cheng & Roberts, 2001; Jeltsch, 2002).

Methylation of the exocyclic amino groups of cytosine and adenine is not a trivial process too, since neither of these bases displays nucleophilicity at the exocyclic amino groups, as their free electron pairs are conjugated to the aromatic systems. Detectable nucleophilicity is in fact only observed at the N1 and N3 positions of adenine and the N3 position of cytosine, which are not targets for methylation. DNA N-MTases are characterised by the conserved [DNS]PP[YF] motif (motif IV) located in the active site at a position analogous to the PCQ motif in C-MTases. According to the structure of M.*Taq*I (Goedecke *et al.*, 2001), the side chain of the [DNS] residue and the main chain carbonyl group of the first Pro residue serve as hydrogen bond acceptors for the protons of the exocyclic N6 amino group. The tetrahedral arrangement of these groups induces a shift in hybridisation of the nitrogen atom from sp^2 to sp^3, therefore localising the free electron pair at the N6 position. Moreover, the structure of N4mC-MTase M.*Pvu*II revealed that the [DNS] residue can also function as a proton acceptor, where the Ser residue interacts with a Glu

residue by a charge relay system allowing the former to act as a base (Gong *et al.*, 1997). Probably the [DNS] residue in only transiently protonated and immediately transfers the proton to other residues and finally to water. It is likely that this proton transfer occurs late in the reaction when the N–CH_3 has been almost formed. Thus, a cationic transition state is expected, which is further stabilised by cation π-interactions with surrounding aromatic residues (Schluckebier *et al.*, 1998). This theory is further supported by the presence of at least one aromatic residue in the catalytic site of N-MTases interacting with the flipped base, such as the [YF] residue in motif IV, whereas 5mC-MTases do not usually incorporate aromatic residues near the active site (Holz *et al.*, 1999; Jeltsch *et al.*, 1999; Wong & Reich, 2000).

To have a better understanding of the biochemistry of bacterial DNA MTases, both issues of substrate binding kinetics and DNA recognition should also be addressed. In general, AdoMet and DNA can bind to the enzyme either randomly or sequentially ordered. In the latter case the exact order differs among various MTases. Studies on T4 Dam MTase, for example, suggest that the enzyme exists in two distinct conformations: one binds the substrates in a random order, the other binds first the AdoMet cofactor. In M.*Hha*I, M.*Msp*I and *Eco*Dam, binding of DNA precedes that of AdoMet, whereas in M.*Eco*RI, M.*Eco*RV, M.*Eca*I and M.*Rsr*I this sequence is reversed (Jeltsch, 2002). Recent evidences, though, imply a random substrate binding order for M.*Hha*I, in a similar manner to T4 Dam (Vilkaitis *et al.*, 2001).

DNA recognition and binding in MTases usually proceed via a multistep reaction, just like in many other DNA-binding enzymes: an initial non-specific binding to the DNA is followed by simple diffusion scanning for the target site (Surby & Reich, 1996; Gowher & Jeltsch, 2000). The sequence specificity is achieved by a multitude of direct and water mediated contacts between the enzyme and the DNA substrate. Biochemical data suggest a similar contact pattern for 5mC-MTases (Kiss *et al.*, 2001); M.*Hha*I and M.*Hae*III, for instance, contact the major groove with residues located in the enzymes' small domain (Jeltsch, 2002). However, N6mA-MTase M.*Taq*I directly contacts both the major and the minor groove, with residues located in the N- and C-terminal domain, respectively. This was also observed in M.*Eco*RV, leading to the conclusion that the N-terminal domain of N-MTases is substantially involved in sequence recognition (Beck *et al.*, 2001). Interestingly, however, specificity of MTases is not as high as that of restriction endonucleases (which are a paradigm of highly-specific DNA-interacting enzymes), because they usually do not saturate all possible contacts in the major groove (Jeltsch, 2002). This may explain why some DNA MTases methylate sites that slightly differ from their canonical ones (Cerritelli *et al.*, 1989; Pingoud & Jeltsch, 2001; Evdokimov *et al.*, 2002). Moreover, the observation that the energetic contribution of contacts is almost always favourable even with mismatched substrates, demonstrates that no strong, specific interactions are formed between the enzyme and the target base (Jeltsch, 2002).

Last but not least, the processivity of DNA methylation is an issue that has not been sufficiently investigated. It has been observed for some MTases, which are components of bacterial RM systems (M.*Hha*I, M.*Hpa*II, M.*Eco*RI, M.*Eco*RV), that they are not able to catalyse several rounds of methylation on one target molecule in a processive way (Surby & Reich, 1996; Gowher & Jeltsch, 2000; Jeltsch, 2002). Yet, some others, such as M.*Ccr*M and M.*Sss*I, were shown to methylate DNA in a processive manner (Renbaum & Razin, 1992; Berdis *et al.*, 1998), and similar results were more recently obtained for

the *Eco*Dam MTase (Urig *et al.*, 2002). But these are solitary MTases that are not part of RM systems, and it is still unclear how this difference is attributed to their functional role.

1.3 The BseCI DNA methyltransferase

M.*Bse*CI DNA MTase from *Geobacillus stearothermophilus* is a type II N6mA-MTase. It exists as a monomer with an apparent molecular size of 66.7 kDa. The optimum temperature for enzyme activity has been determined at 50-55 °C and the optimum pH at approximately 7.4. The enzyme is inhibited by NaCl and KCl at concentrations greater than 50 m*M*, does not require Mg^{2+} for activity and —like most prokaryotic DNA MTases— it is not active on single-stranded DNA. M.*Bse*CI catalyses the transfer of a methyl group from AdoMet to the N6 atom of the 3′ adenine in the sequence 5′-ATCGAT-3′ similarly to its isoschizomer M.*Cla*I (Rina *et al.*, 1992; Rina & Bouriotis, 1993).

Comparison of the M.*Bse*CI amino acid sequence (579 aa) with sequences of other MTases has shown that M.*Bse*CI contains the two motifs that are common among all N6mA-MTases, GxG and NPPY, located at positions 47-59 and 128-137, respectively. Additionally, M.*Bse*CI and twelve other N6mA-MTases (M.*Ban*III, M.*Taq*I, M.*Pst*I, M.*Hpa*I, M.*Bsu*BI, M.*Hinc*II, M.*Cvi*BIII, M.*Eco*57I, M.*Tth*HB8, M.*Pae*R7I, M.*Acc*I, M.*Bst*VI) indicated a relatively high degree of homology in a third region, around positions 174-188; all the aforementioned enzymes share the 5′-TNNA-3′ element as a subset of their recognition sequence (Rina *et al.*, 1994). This moderately conserved region is presumably involved in the recognition of the methylation site by 5′-NTNNAN-3′-specific N6mA-MTases (Ito *et al.*, 1990). A fourth less conserved amino acid motif, DFLxx, following region II, has been previously observed in five N6mA-MTases (M.*Taq*I, M.*Bst*VI, M.*Vci*BIII, M.*Pae*R71, M.*Tth*HB8I) and found to be present in M.*Bse*CI as well (Rina *et al.*, 1994). All the above enzymes recognise the sequence 5′-NTCGAN-3′ as a target for adenine methylation and the above motif has not been found in other N6mA-MTases with different recognition sites (González & Vásquez, 1993). Having said that, although M.*Bse*CI exhibits the conserved motifs of N6mA-MTases, overall it lacks a detailed homology to MTases of this group, with the exception of the isomethyleric M.*Ban*III, to which it shows the greatest similarity (53% identity).

Chapter 2

Protein Preparation, Crystallisation and Preliminary Analysis of M.BseCI Crystals

2.1 Protein expression and purification

M.*Bse*CI (UniProt accession code: P43423) was purified based on the protocols of Athanasiadis *et al.* (1997) & Kapetaniou *et al*. (2007). The following paragraphs describe only the preparation and crystallisations of the M.*Bse*CI ternary complexes with the AdoHcy cofactor and a specific 10 bp DNA duplex in two different methylation states, the hemimethylated (hereafter referred to as M.*Bse*CI_{HM}) and its fully methylated counterpart (hereafter referred to as M.*Bse*CI_{FM}). Data and partially solved models for the apoenzyme and the ternary complex with AdoHcy and the unmethylated counterpart (hereafter referred to as M.*Bse*CI_{UM}) were kindly provided by Dr. Nicholas M. Glykos.

The *bseCIM* gene encoding M.*Bse*CI (GenBank accession code: X79509) was cloned as previously described by Rina & Bouriotis, 1993, inserted into a pET-26b(+) vector (Novagen) containing a C-terminal 6×His-tag and transformed into *E. coli* BL21 (DE3) cells. A sufficient amount of soluble protein for structural studies was obtained after expression using the following conditions. Cells were grown in 2 l LB medium containing 100 mg ml^{-1} ampicillin and 50 mg ml^{-1} kanamycin at 37 °C until OD_{600} reached ~0.8. The culture was induced with 1m*M* isopropyl *β*-D-thiogalactopyranoside (IPTG) for 4 h at 37 °C and harvested by centrifugation at 4000*g* for 30 min at 4 °C. The precipitated cells were resuspended in 50 ml lysis buffer containing 50 m*M* Na_2HPO_4/NaH_2PO_4 pH 8.0, 300 m*M* NaCl, 5m*M* imidazole, 1m*M* PMSF, 20 mg ml^{-1} leupeptin and 150 mg ml^{-1} benzamidine, and homogenised by sonication. The precipitate was removed by centrifugation at 18 500*g* for 1 h at 4 °C. Purification was performed via the His-tag by affinity chromatography at 4 °C on a 10 ml Ni-NTA chelating column (Qiagen) pre-equilibrated in lysis buffer, and initially washed stepwise with 10 m*M* and 20 m*M* imidazole. With a subsequent increase in imidazole concentration until 300 m*M*, M.*Bse*CI started to elute at 30 m*M* imidazole. Eluates containing more than 80% of homogenous M.*Bse*CI, as judged by 10% SDS-PAGE [Fig. 2.1(*a*)], were pooled, dialysed extensively against 20 m*M* Tris-HCl pH 8.0, 200 m*M* NaCl and concentrated using Amicon Ultra-15 filters. The concentrated solution was further purified using a

Sephacryl S-200 gel-filtration column (XK-26/60, Cytiva) pre-equilibrated with one column volume of dialysis buffer. The flow rate was 0.5 ml min^{-1} and the 2 ml eluates collected were monitored at 280 nm. A typical elution profile of the Sephacryl S-200 column is shown in Fig. 2.1(*b*)-(*c*). M.*Bse*CI eluted at a volume that is consistent with the presence of a monomer. The protein-containing eluates were finally pooled and concentrated to approximately 10 mg ml^{-1}.

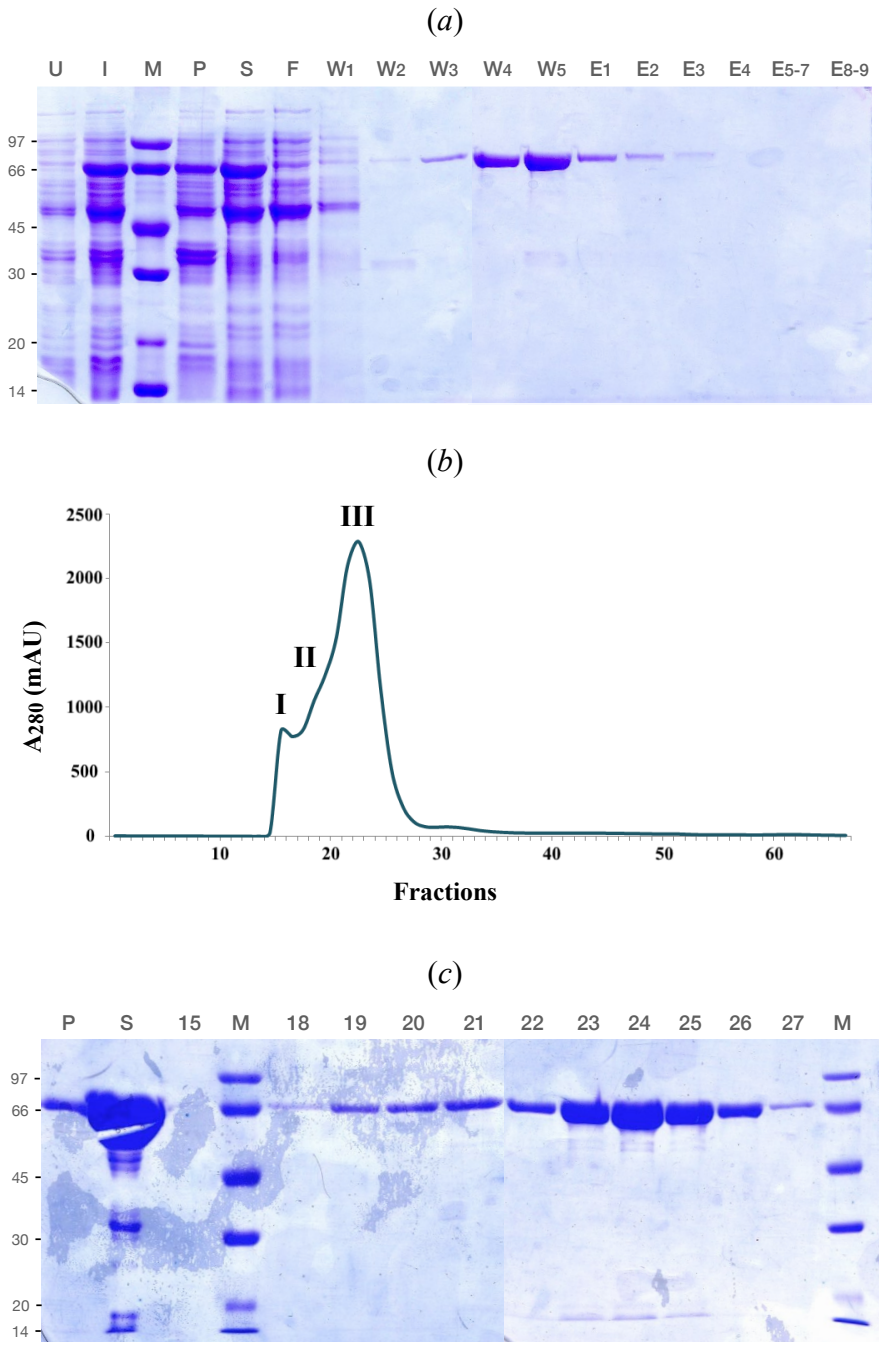

Fig. 2.1 (*a*) 10% SDS-PAGE of fractions collected from Ni-NTA purification. **U**: uninduced culture; **I**: induced culture; **M**: LMW SDS marker (GE Healthcare); **P**: pellet; **S**: supernatant; **F**: flow-through; **W1**: 10 m*M* imidazole wash; **W2**: 20 m*M* imidazole wash; **W3**: 30 m*M* imidazole wash; **W4**: 50 m*M* imidazole wash; **W5**: 100 m*M* imidazole wash; **E1-2**: 150 m*M* imidazole eluates; **E3-4**: 200 m*M* imidazole eluates; **E5-9**: 300 m*M* imidazole eluates. (*b*) Elution profile of the Sephacryl S-200 column. Peak I corresponds to eluate 15, peak II corresponds to eluates 18-20 and peak III corresponds to eluates 21-27. (*c*) 10% SDS-PAGE of fractions collected from size-exclusion chromatography. **P**: pellet; **S**: protein sample prior to gel filtration; **M**: LMW SDS marker (GE Healthcare); **15**: void; **18-27**: protein-containing eluates.

2.2 DNA preparation and gel-shift assay

M.*Bse*CI was subsequently complexed with both a hemimethylated and a fully methylated DNA duplex carrying the specific recognition site (in bold typeface). The lyophilised single-stranded 10 bp-long oligodeoxynucleotides used for the M.*Bse*CI_{HM} ternary complex (5′-GC**ATCGA^{m}T**CG-3′ and its complementary unmethylated strand, 5′-CG**ATCGAT**GC-3′) and the M.*Bse*CI_{FM} ternary complex (5′-CG**ATCGA^{m}T**GC-3′ and its complementary strand, 5′-GC**ATCGA^{m}T**CG-3′) were synthesised by VBC Biotech. The complementary oligonucleotides were annealed in 300 m*M* NaCl at equal concentrations by heating to 100 °C for 5 min, followed by slow cooling until reaching room temperature to produce DNA duplexes. M.*Bse*CI-DNA mixtures (in 20 m*M* Tris buffer pH 8.0 with 50 m*M* NaCl, 10 m*M* $CaCl_2$ and 3 m*M* AdoHcy) were prepared to give final concentrations of 75 μ*M* protein and 100 μ*M* duplex DNA. The ternary complexes were formed after incubation at 50 °C for 1 h, followed by an overnight incubation at 18 °C, and this protein mixture was used for crystallisation trials. The reason behind the universal use of AdoHcy instead of AdoMet is that the latter one is chemically very unstable, both in solution and in dry state, suffering from fast oxidisation and spontaneous degradation (especially in pH values above 5), therefore raising the risk to create a mixed population of ternary complexes containing either form of the cofactor.

Formation of ternary complexes was monitored and evaluated by gel electrophoresis in TAE buffer (200 m*M* Tris base, 100 m*M* acetic acid, 5 m*M* EDTA) on a non-denaturing 10% polyacrylamide gel. The gel was incubated in EtBr in order to detect the DNA bands under UV light, and subsequently stained with Coomassie Blue to visualise protein bands. As can be seen in Fig. 2.2, binding is successful but a few peculiarities can be noticed. First, a faint band appears in lane 5 where the apoenzyme was loaded; this is probably due to the fact that M.*Bse*CI, as a DNA binding enzyme, might carry —even after purification— cellular DNA to some extent. Second, comparison of band intensity in lanes 2 and 4 indicates a stronger binding affinity of the M.*Bse*CI•AdoHcy complex for the fully methylated DNA duplex. On one hand hand, this is inconsistent with the general affinity of MTases for hemimethylated substrates, but on the other, the ternary complex with the fully methylated duplex is the biologically relevant end product of the methylation reaction. In addition, a faint free-DNA band corresponding to the excess of unbound free DNA should have been also expected in both lanes 2 and 4, given the stoichiometry of the enzyme-DNA reactants (1:1.3). Whatever the case is, two things can be deduced: i) binding is successful and ii) the binary complex shows almost a 2-fold increase in biding affinity for the fully methylated DNA duplex compared to its hemimethylated counterpart, probably because it is the biologically relevant reaction product.

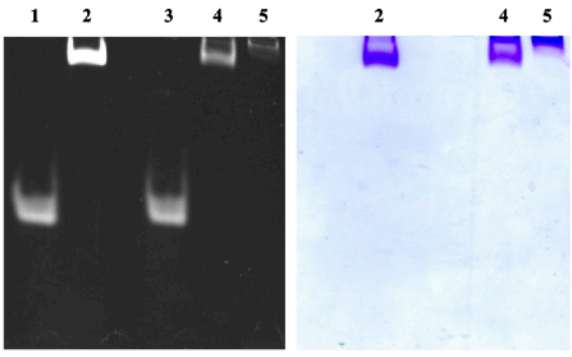

Fig. 2.2 Left panel: Electrophoretic mobility shift assay. Lane 1: free fully methylated DNA; Lane 2: M.*Bse*CI_{FM}; Lane 3: free hemimethylated DNA; Lane 4: M.*Bse*CI_{HM}; Lane 5: M.*Bse*CI apoenzyme. M.*Bse*CI was mixed with DNA duplexes and AdoHcy in a stoichiometry of 1:1.3:3. Right panel: Same gel stained with Coomassie Blue.

2.3 Crystallisations and data collection

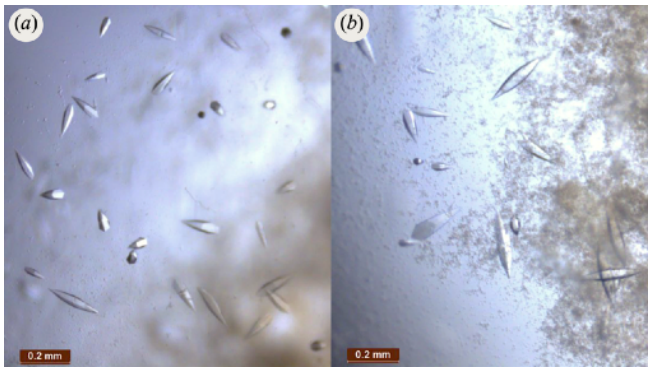

Fig. 2.3 Crystals of (*a*) M.*Bse*CI$_{HM}$ and (*b*) M.*Bse*CI$_{FM}$.

Crystallisation conditions were initially screened by the sitting drop vapour-diffusion method at 20 °C using the OryxNano crystallisation robot (Douglas Instruments), Linbro 96-well optimisation plates and various commercially available crystallisation kits (JBScreen JCSG++ HTS, Structure Screen 1,2 and MIDAS-HT96). The crystallisation drops were a mixture of 400 nl protein solution and 100 nl reservoir solution. Crystals for both ternary complexes were first obtained from MIDAS-HT96 (Molecular Dimensions) under 28%(*v*/*v*) glycerol ethoxylate, 0.1 *M* Tris–HCl pH 9.0 and the crystallisation conditions were then manually refined using the conventional hanging-drop vapour-diffusion method in 24-well Linbro cell-culture plates. The drops consisted of 2 µl protein solution and 1 µl reservoir solution and were equilibrated against 1000 µl of reservoir solution at 20 °C. Crystals of reasonable size were obtained with a reservoir solution composed of 26-30%(*v*/*v*) glycerol ethoxylate and 0.1 *M* Tris–HCl pH 9.5. However, in the case of M.*Bse*CI$_{HM}$, rapid crystal formation within 24 h and thereby bad crystal packing, gave rise to crystals of insufficient quality that usually got dissolved within 2 d. Therefore, a controlling nucleation approach using paraffin oil (300 µl) was employed. Better quality small prismatic crystals at the length of approximately 0.1-0.2 mm [Fig. 2.3] were finally appeared within 7 d.

X-ray diffraction data were collected on a PILATUS 6M detector with an active area of 424×435 mm^2, at a wavelength of 0.98 Å using beamline BL13–XALOC at the ALBA Synchrotron (Juanhuix *et al.*, 2014). Prior to data collection, single crystals were mounted on a nylon loop and then, without being cryoprotected, flash-cooled to 100 °C by storing in liquid nitrogen. Using an oscillation range of 0.25°, 1440 images were collected to a maximum resolution of 2.4 Å for the M.*Bse*CI$_{HM}$ complex. Similarly, for the M.*Bse*CI$_{FM}$ complex, using an oscillation range of 0.5°, 720 images were collected to a maximum resolution of 2.6 Å. Fig. 2.4 shows typical diffraction images recorded from the M.*Bse*CI$_{HM}$ and M.*Bse*CI$_{FM}$ crystals. Data were processed and scaled using *iMOSFLM* (Battye *et al.*, 2011) and *AIMLESS* (Evans & Murshudov, 2013) from the *CCP*4 program suite (Winn *et al.*, 2011).

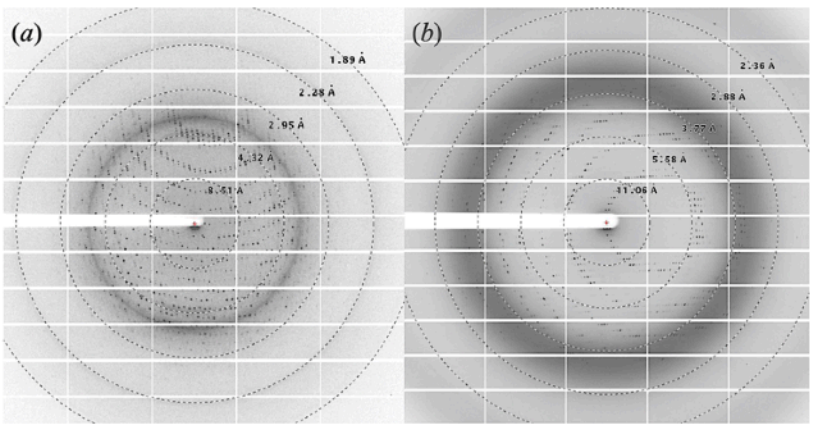

Fig. 2.4 X-ray diffraction images of (*a*) M.*Bse*CI$_{HM}$ and (*b*) M.*Bse*CI$_{FM}$ crystals. The resolution shells are shown by concentric circles.

2.4 Preliminary analysis and structure determination

Analysis of the diffraction data indicated that both crystals belong to the hexagonal space group $P6_1$ with unit cell parameters $a=b=87.7$, $c=156.1$ Å, $\gamma=120°$ and $a=b=88.0$, $c=155.5$ Å, $\gamma=120°$ for the M.*Bse*CI$_{HM}$ and M.*Bse*CI$_{FM}$ crystals, respectively. Assuming one molecule per asymmetric unit, the Matthews coefficient V_M is 2.66 Å^3 Da^1 (corresponding to a solvent content of 53.8%) for the M.*Bse*CI$_{HM}$ crystal, while for the M.*Bse*CI$_{FM}$ crystal, V_M is 2.67 Å^3 Da^1 (corresponding to 53.9% solvent content). Data-quality assessment for both crystals indicated low *Z*-scores and high completeness in ice-ring sensitive areas, implying that no ice-rings were detected. The distribution of $|L|$ values indicated a twin fraction of 0.0, suggesting no twining problems. A summary of the crystallographic parameters and data-collection statistics is given below in Table 2.1. As for the M.*Bse*CI$_{UM}$ ternary complex, crystals were previously obtained by Kapetaniou *et al*. M.*Bse*CI$_{UM}$ crystallises in $P6_1$ space group with unit cell parameters $a=b=87.0$, $c=156.4$ Å, $\gamma=120°$. The asymmetric unit is comprised of one molecule and 53.2% solvent content.

Table 2.1 Data collection and processing. Values for the outer shell are given in parentheses.

	M.*Bse*CI$_{HM}$	**M.*Bse*CI$_{FM}$**
Diffraction source	BL13–XALOC	BL13–XALOC
Wavelength (Å)	0.9792	0.9795
Temperature (K)	100	100
Detector	PILATUS 6M–Dectris	PILATUS 6M–Dectris
Crystal-detector distance (mm)	389.84	505.10
Rotation range per image (°)	0.25	0.5
Total rotation range (°)	105	85
Exposure time per image (s)	0.25	1.0
Space group	$P6_1$	$P6_1$
a, *b*, *c* (Å)	87.7, 87.7, 156.1	88.0, 88.0, 155.5
α, β, γ (°)	90, 90, 120	90, 90, 120
Mosaicity (°)	0.65	0.58
Resolution range (Å)	52.03-2.40 (2.49-2.20)	44.01-2.60 (2.72-2.60)
Total No. of reflections	130119 (14231)	82317 (8630)
No. of unique reflections	24491 (2662)	19732 (2101)
Completeness (%)	92.4 (94.9)	94.1 (82.4)
Redundancy	5.3 (5.3)	4.2 (4.1)
$\langle I/\sigma(I)\rangle$	7.8 (1.4)	5.1 (0.7)
R_{merge}†	0.101 (0.910)	0.104 (2.254)
Overall *B* factor from Wilson plot (Å^2)	48.83	60.36

†$R_{merge}=\Sigma_h\Sigma_j|\langle I_h\rangle - I_{h,j}| / \Sigma_h\Sigma_j I_{h,j}$ where *h* represents the unique reflections and *j* their symmetry-equivalent contributors.

The important data set was a platinum derivative of the M.*Bse*CI-DNA crystals. The derivatised crystals diffracted to only 3.25 Å and were (not unexpectedly) non-isomorphous with the native crystals. The platinum anomalous signal, however, was sufficiently strong to allow a unambiguous determination of the position of the anomalous scatterers from the maximum entropy anomalous difference Patterson maps (Glykos & Kokkinidis, 2000), and to give (through SAD phasing performed with the program *CRANK* from *CCP*4) a phase set with a mean figure of merit of 0.53 after density modification. The corresponding protein density map showed clear protein-like features, with a prominent β-sheet decorated with α-helices (corresponding to what turned out to be the N-terminal domain). Very little of the C-terminal domain and no evidence for the presence of DNA could be discerned from that map. A rough backbone trace was build at 3.25 Å using the program *BUCCANEER* (Cowtan, 2012) from *CCP*4. This model was used as a search model in a *PHASER* (McCoy *et al.*, 2007) run performed against a medium 2.2 Å resolution native dataset. The solution from *PHASER* was clear and unique with a translation function *Z*-score of 13.4. This initial rough model was extended and refined through successive cycles of *BUCCANEER* and density improvement with *ARP/wARP* (Langer *et al.*, 2008), leading to a map with clear and pronounced electron density for the DNA oligonucleotide. Further model building and refinement to a resolution of 2.3 Å (as useful data hardly reach 2.2 Å, mainly due to strong anisotropicity along the polar axis of the $P6_1$ group) led to the final model that converged to an R/R_{free} of 19.6/22.2%. Both M.*Bse*CI_{HM} and M.*Bse*CI_{FM} ternary complexes were solved by molecular replacement using as a search model the aforementioned M.*Bse*CI_{UM} complex. Solutions from *PHASER* were definite with a translation function *Z*-score of 26.2 for M.*Bse*CI_{HM} and 42.4 for M.*Bse*CI_{FM}. Final models converged to an R/R_{free} of 17.4/20.6% and 18.0/21.2%, respectively. As for the previously obtained crystals of the apoenzyme (Athanasiadis *et al.*, 1997), these diffract to 2.5 Å and belong to the monoclinic space group $P2_1$ with unit cell dimensions a=53.7, b=85.7, c=151.8 Å and β=95.1°. The asymmetric unit contains two molecules and has a solvent content of 54%. Interestingly, although the active form of DNA MTases is usually a monomer, dimeric assemblies in asymmetric units occur rather frequently in these enzymes. A self-rotation function has been calculated using the program *POLARRFN* (written by Kabsch, W. and distributed with the *CCP*4 suite) in order to identify local symmetries in the asymmetric unit. The self-rotation function [Fig. 2.5], calculated using data between 10 and 3 Å with a radius of integration of 23 Å and a rotation space sampling step of 5°, indicated the presence of a non-crystallographic 2-fold axis. Phases were solved by multiple isomorphous replacement including anomalous scattering data. Prior to interpretation, the MIRAS electron density was modified by solvent flattering and non-crystallographic symmetry averaging and finally was improved by

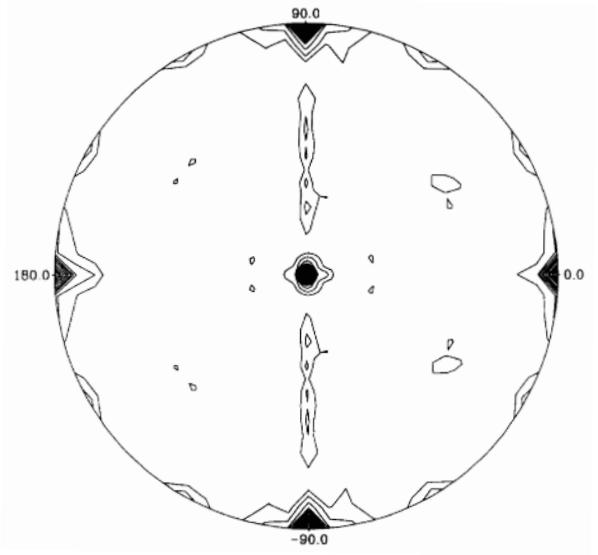

Fig. 2.5 Stereographic projection of the self-rotation function for κ=180° for native M.*Bse*CI crystals. Adapted from A. Athanasiadis *et al.*, *Acta Crystallogr. D Biol. Crystallogr.* 1997, *53*, p. 479.

phase refinement through iterative skeletonization. Further building and refinement led to the final model with an R/R_{free} of 16.8/20.6%.

Crystallographic refinement was performed with the program *COOT* (Emsley & Cowtan, 2004) and the *PHENIX* software package (Liebschner *et al.*, 2019). Initial cycles of coordinate refinement and addition of the most well-ordered water molecules were followed by several rounds of TLS refinement interspersed with further addition of water molecules, fitting of discretely disordered residues and occasional rounds of rebuilding. The refinement process was finally completed with successive rounds of ADP refinement. To keep this section tidied up, the important steps of refinement process for each model can be found in the Appendix (Tables A1-A4).

2.5 Structural evaluation

Other than the values of the R and R_{free} factors, the quality and reliability of a model is also assessed by the deviation of its stereochemistry from the ideal values. Table 2.2 below shows the results from the evaluation analysis of each model as provided by *MolProbity* (Williams *et al.*, 2018). Overall, all four models have been solved with an excellent geometry. There is a consistent indication of two side-chain outliers (Phe^{79} and Tyr^{136}) for the ternary complexes and one side-chain outlier (Glu^{509}) for the apoenzyme. These are density-based deviations and are contributed to the residues' distinct interaction pattern. An r.m.s.d. bond angle outlier reported in M.*Bse*CI$_{UM}$ model (Asp^{285}) is due to steric clashes with Lys^{480} of a X-Y, X, Z+1/6 symmetry-related molecule. Fig. 2.6 also shows the Ramachandran plots for each model. More than 95% of amino acid residues fall into energetically favoured regions and no outliers are reported. Finally, the electron density in the final $2mF_o–DF_c$ maps is well-defined [Fig. 2.7]. At 1.0σ contour level, the polypeptide backbone is continuous throughout most of the models. The majority of residues, including those interacting with the DNA and the cofactor, have well-defined and interpretable electron density. Omit-maps were also used to check some loop regions. Weak or no density only accounts —not unexpectedly— for some solvent-exposed residues and parts of the N- and C-terminus. Lastly, the DNA duplexes have clear discernible density, allowing even for the modelling of the methyl groups. It should be noted at this point that in M.*Bse*CI$_{HM}$ the orientation of the bound DNA duplex is predetermined by the methylated adenine. However, the fully methylated and unmethylated counterparts are entirely symmetrical and thus can be crystallised bound in two different orientations, since the base pairs on either side of the recognition sequence are not implicated in specific interactions. And it actually does so, as in M.*Bse*CI$_{FM}$ and M.*Bse*CI$_{UM}$ the flipped adenine is that of the strand which is already methylated in the hemimethylated duplex (probably this orientation leads to better crystal packing between the end base pairs of the DNA duplexes of adjacent crystal units). Except for this subtle difference, all three complexes are virtually identical, and M.*Bse*CI$_{HM}$ has been arbitrarily used for the creation of the following figures. In case of structurally or biologically important distinctions, each complex will be referred explicitly.

Table 2.2 Overall crystallographic data and refinement results.

	M.*Bse*CI	**M.*Bse*CI$_{HM}$**	**M.*Bse*CI$_{FM}$**	**M.*Bse*CI$_{UM}$**
Space group	$P2_1$	$P6_1$	$P6_1$	$P6_1$
Asymmetric unit (No. of molecules)	2	1	1	1
Resolution range (Å)	24.71 - 2.50	42.93 - 2.40	44.00 - 2.60	28.49 - 2.30
Overall reflections	46244	24387	19649	29350
Completeness (%)	97.11	91.74	93.70	98.72
R†	0.168	0.174	0.180	0.196
R_{free}‡	0.206	0.206	0.212	0.222
Nonhydrogen protein atoms	8604	4530	4504	4414
DNA / AdoHcy / CH_3 atoms	–	399 / 26 / 1	399 / 26 / 2	399 / 26 / –
Water molecules	215	187	47	241
R.m.s. deviations				
Bonds (Å)	0.003	0.003	0.003	0.003
Angles (°)	0.695	0.552	0.614	0.606
Dihedrals (°)	10.70	15.25	15.11	15.27
Impropers (°)	0.055	0.044	0.049	0.049
All-atom contacts				
Clashcore	4.37	4.32	4.24	4.01
Protein geometry				
Poor rotamers (%)	0.21	0.40	0.40	0.40
Favored rotamers (%)	98.28	96.99	96.57	97.31
Ramachandran outliers (%)	0	0	0	0
Ramachandran favoured (%)	98.10	95.11	95.43	96.63
Rama distribution Z-score	-0.54 ± 0.25	-1.96 ± 0.31	-1.82 ± 0.33	-1.40 ± 0.32
C_β deviations (%)	0	0	0	0
Bad bonds (%)	0	0	0	0
Bad angles (%)	0	0	0	0.02
Peptide omegas				
Cis prolines (%)	0	0	0	0
Nucleic acid geometry				
Bad bonds (%)	–	0	0	0
Bad angles (%)	–	0	0	0

†$R=\Sigma_h||F_o| - |F_c|| / \Sigma_h|F_o|$. ‡$R_{free}$ is calculated using a randomly chosen subset (5%) of reflections not used in the refinement.

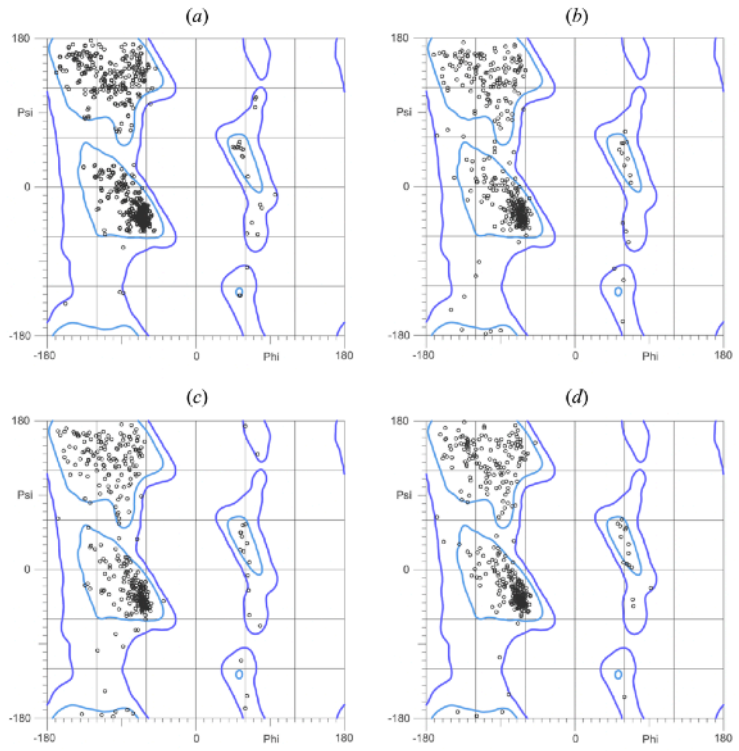

Fig. 2.6 Ramachandran plots generated by *MolProbity* for the final refined models of (*a*) M.*Bse*CI, (*b*) M.*Bse*CI$_{HM}$, (*c*) M.*Bse*CI$_{FM}$ and (*d*) M.*Bse*CI$_{UM}$.

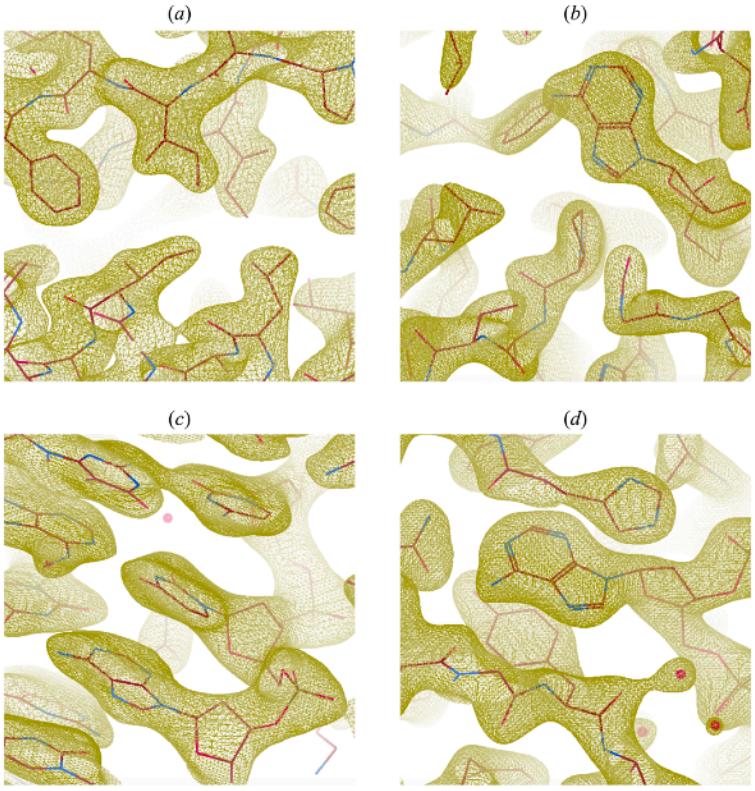

Fig. 2.7 Maximum entropy estimate of the final $2mF_o–DF_c$ electron density maps contoured at 1.0σ for (*a*) M.*Bse*CI, (*b*) M.*Bse*CI$_{HM}$, (*c*) M.*Bse*CI$_{FM}$ and (*d*) M.*Bse*CI$_{UM}$ models as produced by the program *GraphEnt* (Glykos & Kokkinidis, 2000), illustrating the quality of phase determination. The final models in the snapshots are shown superimposed using a liquorice representation.

Chapter 3

Structural Analysis of M.BseCI

3.1 Overview of the structure

The structure of M.*Bse*CI is illustrated in Fig. 3.1(*a*). The protein, with dimensions of approximately 90×60×40 Å^3, is clearly divided into two domains of comparable size; an N-terminal domain and a C-terminal domain. The two domains are connected by a single loop and arranged in the form of a 'C'-shape with a cleft sufficiently wide to accommodate double-stranded DNA.

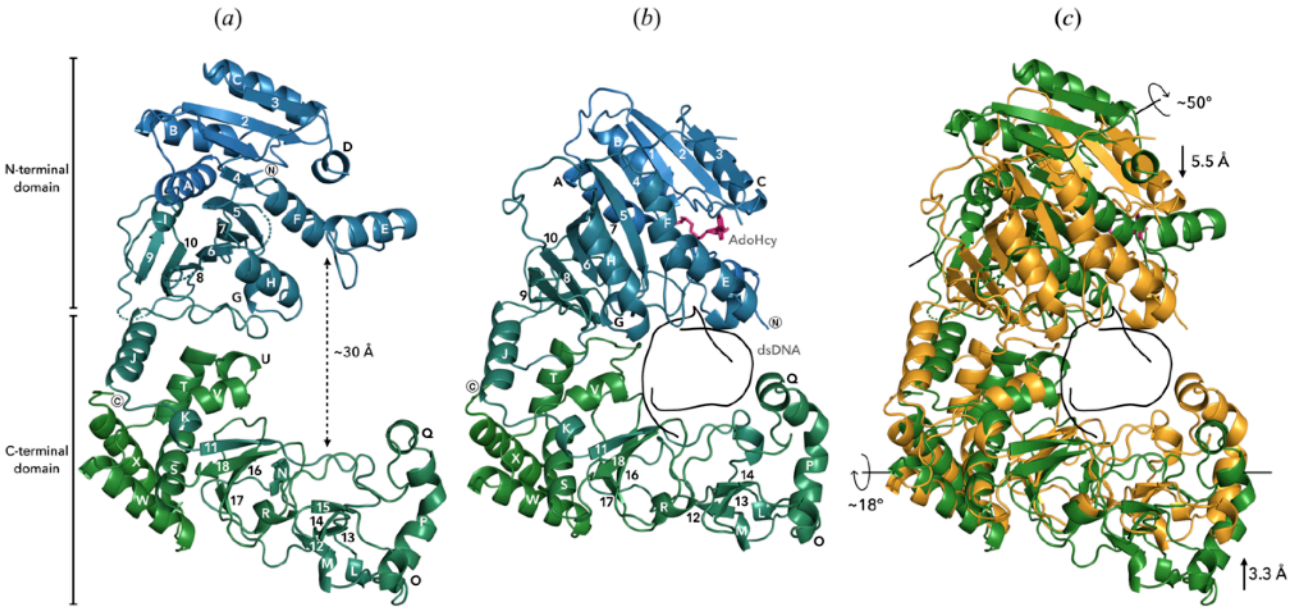

Fig. 3.1 (*a*) Three-dimensional structure of M.*Bse*CI apo form in cartoon representation. View showing the folding of the polypeptide chain in two domains, N-terminal (upper) and C-terminal (lower), which form a cleft wide enough (~30 Å) to accommodate double-stranded DNA. The chain is coloured from the N- (blue) to C-terminus (green). The individual α-helices and β-strands are labelled with letters and numbers, respectively. (*b*) Cartoon representation of the ternary complex (here M.*Bse*CI$_{HM}$) in the same view as in (*a*). The double-stranded DNA (dsDNA) is shown in black ribbon and the AdoHcy cofactor in magenta. (*c*) Superposition of structures shown in panels (*a*) and (*b*). The apoenzyme is shown in green and the M.*Bse*CI$_{HM}$ polypeptide chain in orange colour. Upon DNA binding, the N- and C-terminal domains undergo a rotation of about 50° and 18°, respectively, around the indicated axes. These rotational movements are combined with a shift of both domains towards the DNA, as indicated by the arrows at the right side of the panel. The mean displacement for the N- and C-terminal domains correspond to 5.5 and 3.3 Å, respectively.

The N-terminal domain (amino acid residues 1-301) is a mixed α/β structure consisting of a twisted ten-stranded β-sheet, decorated on both sides by α-helices oriented more or less parallel to the β-strands. Helices αA, αB, αC, αI run across one face of the sheet, whereas helices αD, αE, αF, αG and αH run across the other. This configuration is recognisable as a MTase fold, which is conserved and present in several other MTases (Malone *et al.*, 1995). The consensus fold begins with αA and continues with alternate parallel β-strands and α-helices, with the exception of the antiparallel strand β7 that immediately follows β6. The N-terminal domain of M.*Bse*CI appears to be constructed by the insertion of two α-helices (αE, αG) and a three-stranded antiparallel β-sheet (β8 through β10) flanked by an α-helix (αI). The rest of the protein forms the C-terminal domain (amino acid residues 302-579) and has not such an extensive secondary structure. The polypeptide chain diverts away from the N-terminal domain forming two successive helices (αJ and αK) followed by a β-strand (β11) that forms the edge of a small four-stranded antiparallel β-sheet. It then coils away in the domain following an almost circular path through which a short two-stranded β-sheet (β12 and β13), a small β-hairpin (β14 and β15) and a few short helical components (αL, αM, αN and αO) are formed, before it returns to the rest of the four-stranded β-sheet (β16 to β18) via helices αP, αQ and αR. The last part of the C-terminal domain is composed of an α-helical cluster (αS through αX).

The two domains are connected covalently by the loop between β10 and αJ, which is positioned at the domains' interface. Their orientation is partially determined by the protrusion of some structural elements towards the cleft and close to this loop, such as helices αG, αH, αU and αV. The cohesive forces, other than van der Waals contacts, include hydrogen bonds (Thr^{190}-Tyr^{513}, Asn^{300}-Tyr^{523}, Tyr^{303}-Glu^{523}) and a salt bridge between Glu^{195} and Lys^{302}. Another salt bridge, formed between Lys^{294} (located in the last position of β10) and Glu^{310} (located in the first position of αJ), possibly plays a significant role in the retention of the bilobed shape.

The cleft between the two domains has a diameter of approximately 30 Å, large enough to accommodate double-stranded DNA, and its surface is predominantly occupied by positively charged side chains capable of interacting with the DNA backbone (see also Fig. 3.3(*a*) in the following section). Upon DNA binding and formation of the ternary complex [Fig. 3.1(*b*)], the protein undergoes a small hinge movement closing its grip around the DNA. Compared to the structure of the apoenzyme, no extensive backbone conformational changes are observed [see Fig. A1 of the Appendix and corresponding legend], but both domains change their relative orientation, while the loop between β10 and αJ moves out of the cleft. As shown in Fig. 3.1(*c*), both domains rotate independently, and whilst the rotation is less prominent for the C-terminal domain, the N-terminal domain rotates significantly about 50° around an axis that goes across its β-sheet. These movements are also coupled with a slight displacement of both domains towards the DNA molecule. The reposition of the domains allows for interactions with the DNA, and brings the cofactor and the catalytically essential residues close to the target nucleobase without excessively distorting the protein's structural framework.

Overall structural characteristics of M.*Bse*CI and comparison of its sequence with that of several MTases, confirms that the protein belongs to the γ-group of type II N6mA-MTases. The sequences aligned in Fig. A2 of the Appendix all belong to γ-N6mA-MTases and recognise the 5′-NTNNAN-3′ element as a target sequence for adenine methylation, whereas seven of them share the 5′-TCGA-3′ element as a subset

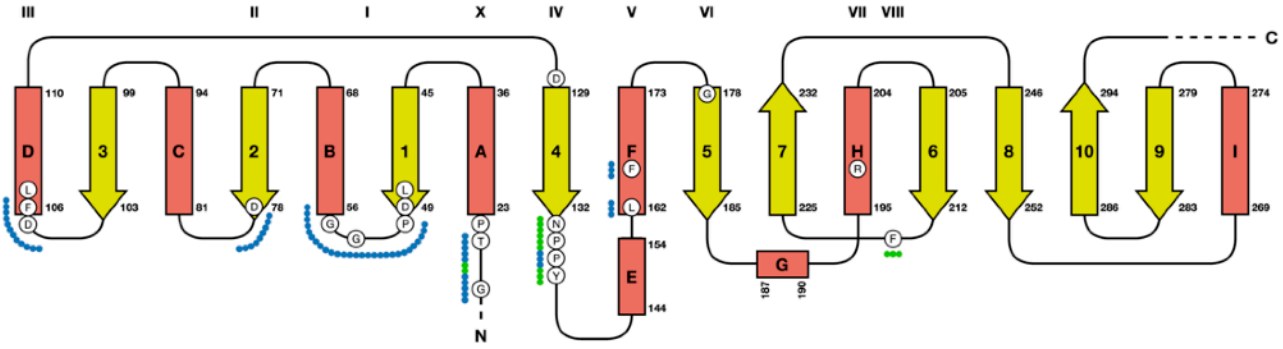

Fig. 3.2 Topology diagram of the M.*Bse*CI N-terminal domain based on the crystal structure of the apoenzyme. Rectangles and arrows indicate α-helices (lettered) and β-strands (numbered), respectively. Highly conserved amino acid residues are circled and the corresponding motifs are shown on top of the diagram. Blue dotted lines indicate regions interacting with the cofactor, and green dotted lines indicate regions interacting with the target nucleobase.

of their recognition site. The nine conserved amino acid motifs (I-VIII and X) residing in the N-terminal domain vary in the degree of conservation, from strongly conserved, to motifs that are more tolerant to primary sequence variation. The Pro-X-X-Gly-X-Gly sequence (or a close relative) of motif I, for example, is strongly conserved in a wide variety of AdoMet-dependent MTases, including as well N4mC-MTases, 5mC-MTases, RNA MTases and protein MTases, and is recognisable from the sequence alone without structural guidance (Cheng & Roberts, 2001). The tight loop formed by this glycine-rich region is important for positioning the AdoMet adenosyl moiety in the correct conformation to ensure close contact with the main chain of the protein. The diprolyl Asn-Pro-Pro-Tyr sequence (motif IV) that mainly forms the active site, is also universally conserved among γ-group MTases, and has mostly Asp and Ser variants in its first position regarding the other two groups (Malone *et al.*, 1995). On the other hand, motif VII bears little resemblance among MTases and motif VIII cannot be unambiguously identified. As shown schematically in the topology diagram of Fig. 3.2, the most conserved motifs form the core structure of the protein and constitute most of the structural elements that surround the active site. In M.*Bse*CI these include four α-helices (αA, αB, αC, αD), four β-strands (β1, β2, β3, β4) and associated loop regions, named *l*-A (preceding αA), *l*-A1 (between αA and β1), *l*-1B (between β1 and αA), *l*-2C (between β2 and αC), *l*-3D (between β3 and αD) and *l*-4E (between β4 and αE). As happens with all MTases, the majority of highly conserved or invariant residues (depicted in circles in the topology diagram of Fig. 3.2) are predominantly clustered around the active site and occur in loops adjacent to ordered secondary structure elements. These residues are implicated in the binding of the AdoMet/AdoHcy, the target base, and the methyl group transfer catalysis, and appear to be crucial for the function of the enzyme. Other conserved residues appear to maintain the framework of the fold. Table 3.1 outlines the interaction pattern of the most conserved residues. The alignment also shows the variability in sequence lengths between motifs. This is understandable in structural terms, since these motifs form a rigid structural framework that can allow flexibility in the regions separating them. The C-terminal domain that follows is directly involved in sequence-specific DNA recognition and thus shows the greatest heterogeneity in terms of size and sequence composition. In fact, no sequence conservation could be observed among the C-terminal domains of the γ-MTases aligned here, despite recognising a common target (data not shown). All the above justify the lack of detailed homology of M.*Bse*CI to other MTases, even to those belonging in closely related groups. Other than the C-terminal domain (or its equivalent target recognition domain (TRD) in other

MTase families), which shows significant variability among AdoMet-dependent MTases, the high degree of structural similarity among various MTase folds is neither reflected by a corresponding degree of sequence conservation. Similar observations have been made with other sets of enzymes, such as amino-hydrolases (Holm & Sander, 1997). This tendency for a structural fold to persist longer than sequence similarity can allow recognition of function among proteins no longer sharing considerable sequence identity, or even among unannotated proteins.

Table 3.1 Interactions of conserved residues.†

Conserved residue	Secondary structure	Motif	Conserved Interactions (Hydrogen bonds, hydrophobic with AdoHcy, and salt bridges)
G17*	*l*-A	X	–
T21			T21:N–AdoHcy:OXT, T21:$O^{\gamma 1}$–AdoHcy:O, T21:$O^{\gamma 1}$–N133:$N^{\delta 2}$
P22*			–
L48	*β*1	I	–
D49			D49:$O^{\delta 1}$–A51:N
P50*	*l*-1B		–
A51			A51:O–AdoHcy:N, A51:N–D49:$O^{\delta 1}$
G53*			G53:N–D78:$O^{\delta 2}$
G55*			–
D78	*β*2	II	D78:$O^{\delta 1}$–AdoHcy:O2′, D78:$O^{\delta 2}$–AdoHcy:O3′, D78:$O^{\delta 2}$–G53:N
F79	*l*-2C		Hydrophobic–AdoHcy
D105	*l*-3D	III	D105:$O^{\delta 1}$–AdoHcy:N6
F106	*α*D		F106:N–AdoHcy:N1
L107			L107:O–K153:N^{ζ}
D128	*l*-D4	IV	D128:$O^{\delta 1}$–K175:$N^{\zeta 2}$, D128:$O^{\delta 2}$–R46:N
N133	*l*-4E		N133:$N^{\delta 2}$–Adenine:N1, N133:$O^{\delta 1}$–Y136:OH, N133:$N^{\delta 2}$–T21:$O^{\gamma 1}$
P134*			P134:O–Adenine:N6
P135*			Hydrophobic–AdoHcy
Y136			Hydrophobic–Adenine, Y136:N–Adenine:N7, Y136:OH–N133:$O^{\delta 1}$
L162	*α*F	V	Hydrophobic–AdoHcy
Y163			–
F166			Hydrophobic–AdoHcy
G178*	*β*5	VI	–
R198	*α*H	VII	R198:$N^{\eta 1}$–Y188:O, R198:$N^{\eta 1}$–T191:O, R198:$N^{\eta 2}$–Y301:O
F219	*l*-67	VIII	Hydrophobic–Adenine

†Main chain hydrogen bonds that stabilise secondary structure elements are omitted. *These residues are likely to be structural determinants as well.

3.2 Cofactor binding site

The cofactor binds to the N-terminal domain near the surface of the protein [Fig. 3.3(*a*)] in a pocket that is mainly formed by the conserved regions I, II, III and X. The shape of the pocket is such that AdoHcy is accommodated in an extended conformation, with the sulfur atom, to which the methyl group is attached in AdoMet, located 4.2 Å away from N6 of the target adenine [Fig. 3.3(*b*)]. The bound cofactor has no direct contacts with the DNA. The structure around the binding site is better illustrated in Fig. 3.3(*c*). The binding pocket for the purine ring of AdoHcy is formed by the hydrophobic side chains of Phe^{79}, Phe^{166}, Pro^{135}, Leu^{162}, and negatively charged side chain Asp^{105}, which interacts directly via O$^{\delta1}$ with adenine N6, 2.63 Å. The aromatic ring of Phe^{166} is juxtaposed to the AdoHcy purine ring in an edge-to-face π-stacking configuration, while that of Phe^{79} lies in parallel in a face-to-face π-stacking interaction. The backbone NH of Phe^{106} also forms a hydrogen bond with adenine N1, 3.05 Å. The highly conserved Gly^{53} (and Gly^{55}) in motif I (part of a classical Rossmann fold *β*1-*α*B-*β*2*)* are required for the sharp bend and hydrogen bonding between ribose O2′H and O3′H, and side chain O$^{\delta1}$ and O$^{\delta2}$ of Asp^{78} (2.72 Å and 2.93 Å, respectively) at the end of strand *β*2. The methionine carboxylate OXT and O1 form hydrogen bond contacts to Thr^{21} NH and O$^{\gamma1}$, 2.93 Å and 2.51 Å, respectively, and there is also a short contact between the methionine NH and Ala^{51} backbone O, 2.59 Å. Several other residues contribute to the formation of the cofactor binding pocket (e.g. Gly^{17}, Ala^{18}, Phe^{20}, Asp^{49}, Cys^{52}, Glu^{56}, Leu^{57}, Lys^{104}, Leu^{107}).

Most of the residues interacting with the cofactor appear to be well conserved in γ-N6mA-MTases. Phe^{106}, Phe^{166} and Pro^{135} are utterly conserved. Thr^{21}, Asp^{78}, Asp^{105} and Leu^{162} have only Ser, Glu, Ser and Tyr variants, respectively, whereas Ala^{51} has mainly Ser and occasionally Gly variants. Phe^{79} is poorly conserved and substituted in all cases by a nonpolar aliphatic side chain (Ile, Leu, Val), with M.*Ban*III be-

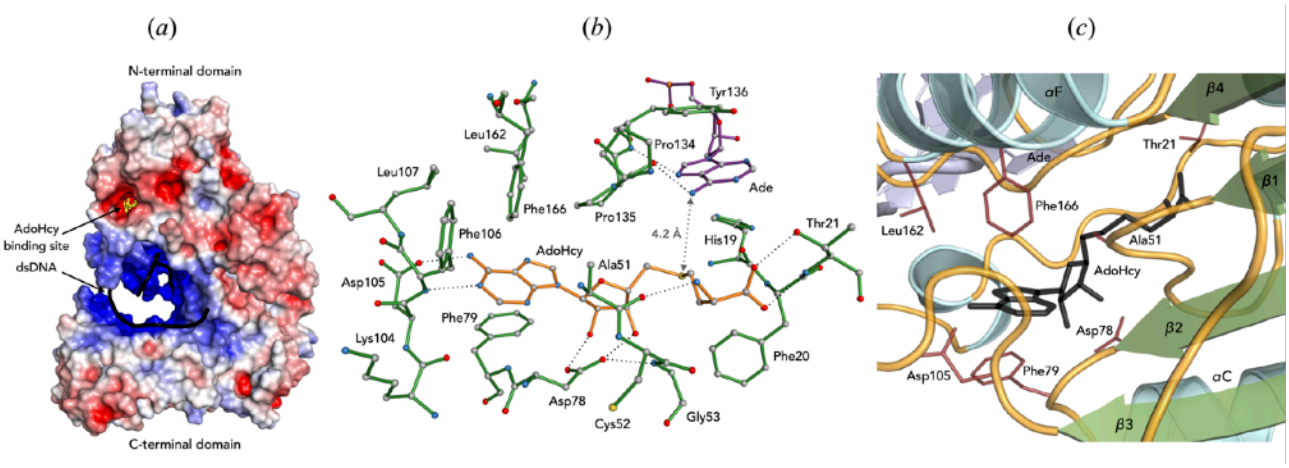

Fig. 3.3 (*a*) Molecular surface representation of M.*Bse*CI (here M.*Bse*CI$_{HM}$) colour-coded according to the electrostatic potential from -5 kT/e (red) to +5 kT/e (blue). AdoHcy is shown in yellow liquorice and DNA backbone in black ribbon representation. (*b*) Schematic representation showing the binding of AdoHcy and part of the flipped adenine's (Ade) interactions. Atoms are shown is standard colours: C, grey; N, blue; O, red; S, yellow; P, orange. Bonds are colour-coded: AdoHcy, orange; target adenine, purple; protein, green. Dashed lines indicate hydrogen bonds. Not all residues forming the binding site are shown. (*c*) View of the structure around the cofactor binding pocket in cartoon representation. AdoHcy is shown in black liquorice representation, and the side chains of a few conserved residues that form the binding pocket are also shown (in red colour) to provide a better spatial understanding.

ing the only exception having a Glu in its equivalent position. This conservation is expanded, up to a certain degree, in other MTase families as well. Table 3.2 includes representatives from various MTase families that share the common fold. The majority of bonds is formed by residues in motifs I, II and III, and to a lesser extent by residues in motif X. Motif IV is occasionally implicated in direct interactions with the cofactor. The O2´H and O3´H groups always form hydrogen bonds with an acidic Asp/Glu side chain, and the interaction between an Asp side chain and N6 of the adenine ring is also entirely conserved in type II MTases. All the residues involved in contacts presented in Table 3.2 are critical for binding and positioning the cofactor, as well as residues of the GxG motif (motif I), which is a structural determinant of the cofactor binding site. Surprisingly, mutational studies on Asp^{78} in an other N6mA-MTase, *Eco*RV, which is analogous to Asp^{105} of M.*Bse*CI, showed no evidence for reduced affinity towards AdoMet or reduced catalytic activity (Roth *et al.*, 1998). This result is very different from what one would have expected, given the conservation of this residue across type II MTase families. Aromatic residues involved in hydrophobic contacts (not shown in Table 3.2) are also conserved among the various families for both creating a hydrophobic environment and stabilising the cofactor's adenosyl moiety.

Table 3.2 Interactions between AdoMet/AdoHcy and amino acid residues in M.*Bse*CI and other MTases.*

		Atom in other MTases						
Cofactor atom	Atom in M.*Bse*CI	M.*Dpn*M	M.*Pvu*II	M.*Taq*I	M.*Hha*I	VP39	CheR	COMT
OXT	T21:N (X)	K21:N (X) A48:N (I)	S276:N (I) S276:O^{γ} (I)	T23:N (X) T23:$O^{\gamma 1}$ (X)	G23:N (I) S305:O^{γ} (X)	Q39:N G72:N	–	S72:O^{γ}
O	T21:$O^{\gamma 1}$ (X)	G46:N (I) E41:$O^{\varepsilon 2}$ (I)	–	E22:$O^{\varepsilon 1}$ (X)	L21:N (I) S305:O^{γ} (X)	H74:$N^{\delta 1}$	T94:$O^{\gamma 1}$ R98:$N^{\eta 1}$	V42:N
N	A51:O (I)	L49:N (I) D194:$O^{\delta 2}$ (IV)	–	E45:$O^{\varepsilon 1}$ (I) N105:$O^{\delta 1}$ (IV)	–	G68:O Y66:OH D138:$O^{\delta 1}$	A123:O R230:O E129:$O^{\varepsilon 1}$	G66:O S72:O^{γ} D141:$O^{\delta 2}$
O2´	D78:$O^{\delta 1}$ (II)	D62:$O^{\delta 1}$ (II)	E294:$O^{\varepsilon 1}$ (II)	E71:$O^{\varepsilon 1}$ (II)	E40:$O^{\varepsilon 1}$ (II) W44:$N^{\varepsilon 1}$ (II)	D95:$O^{\delta 1}$	D154:$O^{\delta 1}$	E90:$O^{\varepsilon 1}$
O3´	D78:$O^{\delta 2}$ (II)	W17:$N^{\varepsilon 1}$ (X) D62:$O^{\delta 2}$ (II)	H246:$N^{\delta 1}$ (X) E294:$O^{\varepsilon 2}$ (II)	E71:$O^{\varepsilon 2}$ (II)	E40:$O^{\varepsilon 2}$ (II)	D95:$O^{\delta 2}$	D154:$O^{\delta 2}$	E90:$O^{\varepsilon 2}$
N1	F106:N (III)	F178:N (III)	D34:N (III)	F90:N (III)	I61:N (III)	–	–	S119:O^{γ}
N3	–	F63:N (II)	–	I72:N (II)	W41:N (II)	–	–	–
N6	D105:$O^{\delta 1}$ (III)	D177:$O^{\delta 2}$ (III)	D34:$O^{\delta 1}$ (III)	D89:$O^{\delta 2}$ (III)	D60:$O^{\delta 1}$ (III)	V116:O	A38: O N212:$O^{\varepsilon 1}$	S119:O^{γ} Q120:$O^{\varepsilon 1}$
N7	–	–	–	–	–	V116:N	–	–

*M.*Dpn*M is an α-N6mA-MTase (Tran *et al.*, 1998), M.*Pvu*II is a β-N4mC-MTase (Gong *et al.*, 1997), M.*Taq*I is a γ-N6mA-MTase (Labahn *et al.*, 1994), M.*Hha*I is a 5mC-MTase (Klimašauskas *et al.*, 1994), VP39 is a RNA-MTase (Hodel *et al.*, 1996), CheR is a protein MTase (Djordjevic & Stock, 1997), COMT is a small molecule MTase (Vidgren *et al.*, 1994). Motifs are given in parentheses.

Having said that, the existing literature shows that in some AdoMet-binding proteins the cofactor's pocket might bear no substantial resemblance to the common α/β cluster, therefore binding the cofactor in a widely different manner. Such examples are the dimeric *E.coli* MetJ repressor, which uses AdoMet allosterically as a co-repressor and not as a substrate (Somers & Phillips, 1992), and the reactivation domain of *E.coli* methionine synthase, which uses AdoMet in a flavodoxin-coupled reductive methylation of cobal-

amin (Dixon *et al.*, 1996). Another example, the glycine N-MTase (GNMT), shares a region structurally very similar to the consensus AdoMet-binding fold, but binds AdoMet in a different position (Fu *et al.*, 1996). Other than that, the structural similarity of AdoMet-binding sites stretches across MTases, bearing resemblance, as well, to the Rossmann fold of the NAD-/FAD-binding motif, which is also comprised of a *β-α-β* structure that contains a similar glycine-rich turn and a conserved Asp (or Glu) that forms a hydrogen bond to the 2´OH of the adenosine ribose (Wierenga *et al.*, 1986).

3.3 DNA binding and stabilisation of the extrahelical adenine

The DNA duplex is bound in such way that the minor groove faces the N-terminal domain and the major groove the C-terminal domain of the enzyme [Fig. 3.4(*a*)]. Its overall conformation shows extensive distortions compared to the B-DNA form. The most prominent one is —not unexpectedly— the complete rotation of the target adenine (A7; for numbering see Fig. 3.4) together with its deoxyribose moiety and the neighbouring phosphodiester groups out of the DNA helix. The deoxyribose ring of the target adenine has a C2´-*endo* conformation and the target base adopts an *anti* orientation about the glycosidic bond. The exclusion of the target adenine from the helix is accompanied by substantial distortions of the DNA. This can be seen by comparing the oligonucleotide structure with that of a prototype B-DNA with the same sequence. First, as shown in Fig. 3.4(*b*), half of the DNA backbone (base pairs C1:G20 through C5:G16) seems to retain an almost B-DNA-like conformation, but the rest suffers from extensive rearrangements. Part of the non-target, methylated strand that hosts bases G11 through A13 seems underwound by approximately 50° in relation to the B-DNA form over this segment. Second, the space left by the rearranged A7 is partially occupied by the unpaired T14, via a small compression in the DNA backbone, that places it towards the centre of the double helix. This thymine is likely to be stabilised via a parallel-displaced π-stacking interaction with G6 of the target-strand [Fig. 3.4(*c*)-(*d*)]. Similar findings have been reported for M.*Taq*I (Goedecke *et al.*, 2001), but not for other DNA MTases, in which amino acid residues are found in the vacated space resulting from extruded target nucleotides. Third, all but two base pairs (A3:T18 and C5:G16) deviate from coplanarity, exhibiting features of either buckle, propeller twist or inclination [Fig. 3.4 (*c*)]. The slightly heavier distortions observed for the base pairs at the ends of the duplex (C1:G20 and C10:G11) are probably due to crystal packing forces. These packing interactions have no effect on the observed protein-DNA interactions, since both base pairs extend out of the enzyme.

Although the extrahelical positioning of the target A7 is expected, the squeezing of the backbone chain around the area, between the rearranged nucleotides A7 and G9, has only been observed in M.*Taq*I and in orphan N6mA-MTase from *C. defficile* CamA (Zhou *et al.*, 2021). The three phosphate groups P_0, P_1 and P_2 (where P_0 is the phosphate group of A7) yield a triangle in space with inter-phosphate distances of 5.5, 5.8 and 5.9 Å [Fig. 3.4(*e*)-(*f*)]. Similarly to CamA, but unlike M.*Taq*I, the rearrangement of the phosphate groups is balanced by a positively charged Lys side chain (Lys^{14} in M.*Bse*CI), which is posi-

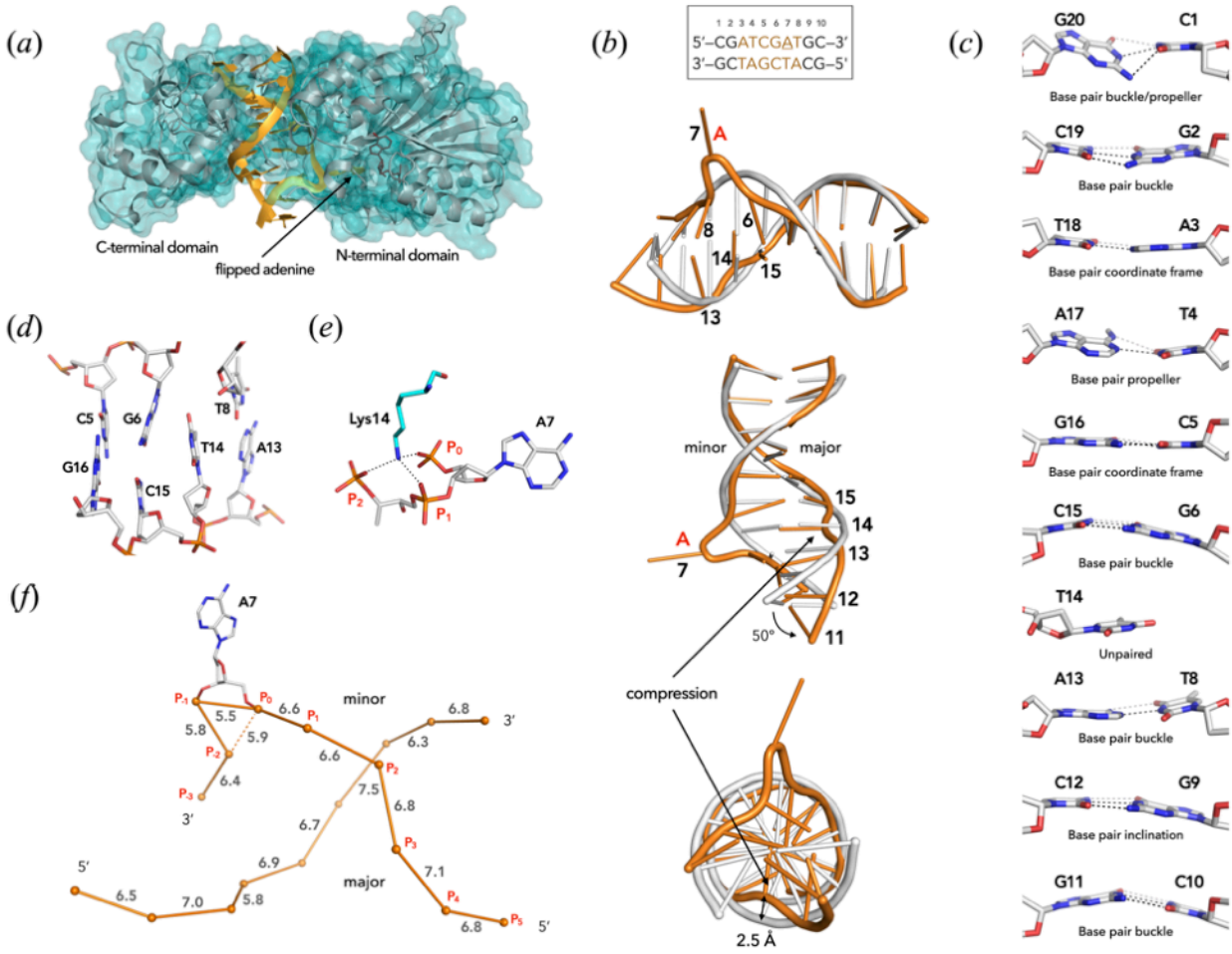

Fig. 3.4 (*a*) Cartoon representation of the M.*Bse*CI ternary complex showing the orientation of the bound DNA. (*b*) B-form DNA (white) superimposed with the M.*Bse*CI-bound DNA molecule (orange), showing the conformational distortions of the latter. The box on the top shows the bound DNA sequence with its target sequence in orange colour and the target adenine underlined. (*c*) Coplanar distortions among the ten base pairs of the M.*Bse*CI-bound DNA. A17, which can be seen here methylated, is symmetric to the A7 base in the target sequence. (*d*) Stabilisation of the unpaired T14. (*e*) Lys^{14} sits in the centre of the triangle formed by phosphate groups P_0, P_1 and P_2 (where P_0 is the phosphate group of the target A7 base) stabilising this conformational rearrangement. (*f*) Inter-phosphate distances of M.*Bse*CI-bound DNA (normally inter-phosphate distances in B-DNA are roughly 7 Å). Atoms in liquorice representations are shown is standard colours: C, white; N, blue; O, red; P, orange, with the exception of panel (*e*) where C atoms the amino acid residue are coloured in cyan. M.*Bse*CI_{HM} was used for the analysis. Structural rearrangements/ distortions in M.*Bse*CI_{FM} and M.*Bse*CI_{UM} were found to be almost identical.

tioned opposite the P_0, P_1, P_2 plane facing the centre of it [Fig. 3.4(*e*)]. This local distortion, together with the interstrand-stacked partner thymine (T14), play a role in stabilising the extrahelical conformation.

The flipped A7 adenine is inserted into a hydrophobic active site cleft formed primarily by three aromatic residues [Fig. 3.5]. The adenine is stacked between Tyr^{136} and Phe^{219}, and sealed off with His^{19} at the bottom of the cleft. The polar groups of the A7 ring (N1, N6 and N7) are involved in hydrogen bonds with the side chain N$^{\delta 2}$ of Asn^{133} (interacting with N1), the main chain carbonyl O of Pro^{134} (interacting with N6) and the main chain amide N of Tyr^{136} (interacting with N7). Asn^{133}, Pro^{134} and Tyr^{136} belong to the catalytically active site sequence NPPY of motif IV and Phe^{219} is part of motif VIII. Tyr^{136} in the apoenzyme is dislodged from the active site, facing the opposite direction, towards the cofactor binding cavity.

The interaction pattern established by these four residues is conserved among N-MTases (Malone *et al.*, 1995; Schluckebier *et al.*, 1995; Goedecke *et al.*, 2001). Since adenine-specific MTases incorporate

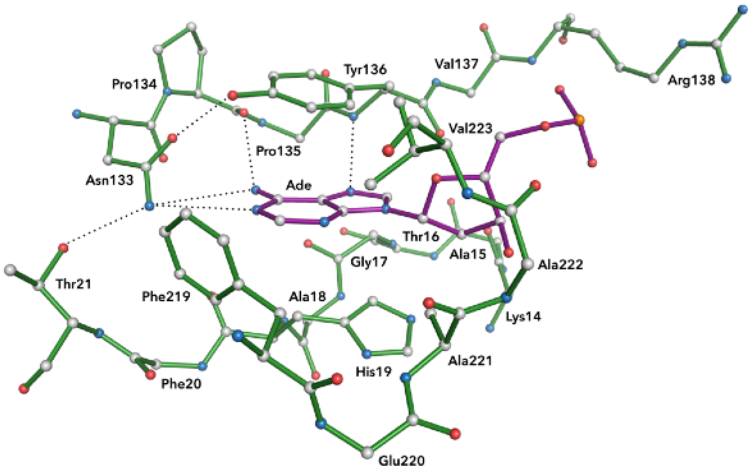

Fig. 3.5 Schematic representation showing the binding of the flipped adenine (Ade) in M.*Bse*CI$_{HM}$. Side chains not interacting with Ade, except for Arg^{138}, are omitted. Atoms are shown is standard colours: C, grey; N, blue; O, red; P, orange. Bonds are colour-coded: target adenine, purple; protein, green. Dotted lines indicate hydrogen bonds.

Asn or Asp in the first position of the diprolyl sequence, one could argue that the presence of either residue defines the specificity for adenine in the active-site cleft. However, in reality, substrate specificity between N6mA-MTases and N4mC-MTases might occasionally overlap. The binding cleft is also created by the side chains of Ala^{221}, Val^{223} and Thr^{21}. It seems that interactions mediated by hydrophobic residues are much more important for base flipping than hydrogen bonds to the flipped base. It has been shown that in case of removal of even a single hydrophobic residue by mutagenesis, base flipping is almost abolished, whereas mutations of residues that contact the target base via hydrogen bonds, only affect the catalytic activity (Jeltsch *et al.*, 1999). Arg^{138}, which directly follows the catalytic NPPY sequence of motif IV, is part of a loop that is involved in sequence-specific contacts to the minor groove. Studies on M.*Hha*I (Cheng *et al.*, 1993; Klimašauskas *et al.*, 1994) revealed that the loop following the catalytic motif of 5mC-MTases is also implicated in binding the minor groove of DNA, locking simultaneously the target cytosine in the extrahelical conformation. The hypothesis that the loop following the catalytic sequence might sterically block the flipping back of extrahelical nucleotides was also later postulated by Goedecke *et al.* after studies on M.*Taq*I. Given the universally conserved structure of the catalytic domain in DNA-MTases, it is possible that the loops succeeding the catalytic motifs have a similar function in sequence-specific recognition and stabilisation of extrahelical nucleotides.

In addition to the interactions between M.*Bse*CI and the flipped adenine, the enzyme makes contacts to the whole DNA molecule. Phosphate contacts —which are important for properly positioning the bound DNA— span almost the entire length of the backbone, but base-specific interactions are encountered only within the six base pairs of the recognition site. Of these, six are water-mediated, and the rest of them are direct contacts which principally determine sequence specificity. The interactions between M.*Bse*CI and DNA are illustrated in Figure 3.6. Both domains are almost equally involved in DNA binding. Of the residues that interact directly with the bases (besides those interacting with A7), Arg^{138} and Gln^{140} reside in the loop after the NPPY motif and Arg^{159} in the loop between αE and αF. Lys^{338} is located in the loop following β11, Trp^{437} is in the loop following αQ, and both Tyr^{513} and Arg^{518} lie in the loop between αV and αT. The side chains of Arg^{159} and Trp^{437} form hydrogen bonds with the exocyclic O atoms of T18 in the A3:T18 base pair. The Gln^{140} side chain contacts the exocyclic N3 of A17 in the T4:A17 base pair, while Arg^{138} and Lys^{338} stabilise the 5C:16G base pair via hydrogen bonds to the O2 of 5G and O6 of 16G, respectively. The side chain of Arg^{138} also contacts the exocyclic N3 atom of 6G in the 6G:15C base pair, while Arg^{518} contacts both the O6 and N7 atoms of the same base. Finally, the T8:A13 base pair is stabilised by interactions of the main chain amide N and carbonyl O atoms of Tyr^{513} with the exocyclic O4

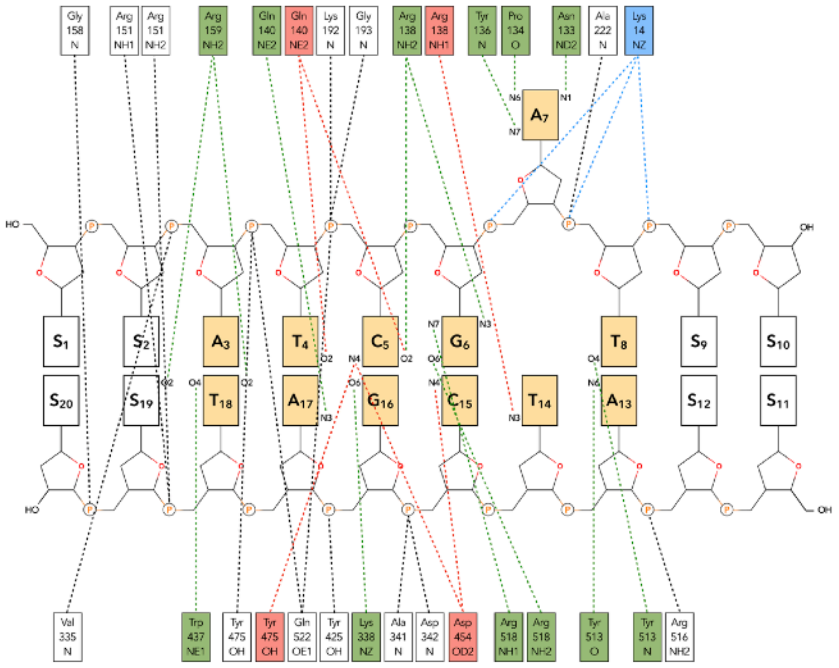

Fig. 3.6 Illustration of hydrogen bonds and salt bridges between M.*Bse*CI and the DNA substrate. The bases of the recognition sequence are shown with yellow background. Direct specific contacts between amino acid residues of the N-terminal domain (upper row) and the C-terminal domain (lower row) are shown in green colour. Water-mediated specific contacts are shown in red colour. Amino acid residues forming non-specific contacts with the phosphate groups are shown with white background. Lys[14] which is involved in backbone squeezing in the target strand is shown in blue colour.

of T8 and N6 of A13, respectively. The interaction pattern from the side of the minor groove is similar to that of M.*Taq*I (and to some extent to that of CamA), in terms of the structural elements involved in binding, with —mostly loop— regions between motifs IV-V and VI-VII, the region preceding motif X and the region around motif VIII, approaching the DNA; but other than that, no extensive similarities for the residues involved in specific binding could be observed. Nevertheless, comparison of the C-terminal domains of M.*Bse*CI and M.*Taq*I exhibits some interesting findings. Although one would expect substantial structural variations, given the domains' insignificant sequence conservation (18.0% identity) and their distinct role in specific DNA recognition, much of the fold is structurally conserved, including the four-stranded antiparallel β-sheet and many of the helical elements that constitute the core of the domain. Additionally, a number of residues responsible for sequence-specific binding, exhibit structural (found in similar positions) and chemical (have similar physicochemical properties and interact with the same functional groups of the DNA molecule) similarities. Val[335], Lys[338], Asp[342], Arg[516] and Arg[518] in M.*Bse*CI correspond to the analogous residues Phe[268], Arg[271], Ser[272], Lys[337] and His[394] respectively, in the structure of M.*Taq*I (see Fig. A3 of the Appendix), and interact with the same functional groups in the bound DNA. Such similarities are present in the N-terminal domains to a much lesser extent, where only two residues, Arg[138] and Gln[140] in M.*Bse*CI, are analogous to Lys[139] and Tyr[117] respectively, in M.*Taq*I. What is really interesting, is that CamA exhibits a striking structural similarity to M.*Bse*CI (Fig. A4 of the Appendix), although exhibiting slightly larger sequence identity (23.5%) and recognising a completely different target sequence (5′-CAAAA<u>A</u>-3′; underlining indicates the target A). Similarities, though, in residues involved in DNA binding are not evinced, except very few implicated in phosphate group contacts, like Asp[324]/Asp[353], Gln[522]/Gln[532], Tyr[475]/Tyr[476], in M.*Bse*CI and CamA, respectively. These findings have clear implications in evolu-

tionary processes, and come in contrast to the hypothesis that an ancestral catalytic domain may have been fused with various target recognition domains. Such concerns have also been raised by Reinisch *et al.* (1995), after observing similarities between the target recognition domains of M.*Hae*III and M.*Hha*I. As it seems, the different groups of DNA MTases might be, paradoxically, more uniform than had been expected.

Interestingly, no hydrogen bond contacts appear between M.*Bse*CI and the methyl group attached to A17 in ether M.*Bse*CI_{HM} or M.*Bse*CI_{FM}. And the overall interaction pattern seems to be maintained among the three complexes, providing no putative explanation with respect to the greater affinity of the enzyme for the hemimethylated over the unmethylated form of its substrate DNA. All protein-nucleic acid contacts, either direct or water-mediated, can be seen schematically in Fig. A5 of the Appendix.

3.4 Catalytic mechanism

As mentioned in the previous section, side chains of the conserved amino acid residues belonging to the catalytic motif IV, and Phe^{219} of motif VIII, are the structural determinants of the catalytic site. The distance between the sulfur atom of AdoHcy and N6 is 4.2 Å in M.*Bse*CI_{HM} (4.5 Å in M.*Bse*CI_{FM} and 4.1 Å in M.*Bse*CI_{UM}), which is slightly shorter compared to the corresponding distance of 4.7 Å observed for M.*Taq*I (Goedecke *et al.*, 2001) and 5 Å observed for 5mC-MTase M.*Hha*I (Klimašauskas *et al.*, 1994), and slightly longer compared to the 3.8 Å distance in CamA (Zhou *et al.*, 2021). This structural arrangement is consistent with the hypothesis that methylation in M.*Bse*CI occurs via a direct methyl group transfer mechanism.

The catalytic mechanism of N6-adenine methylation has been thoroughly studied by Goedecke *et al.* on M.*Taq*I and is remarkably conserved in M.*Bse*CI, as well. In brief, the enzyme catalyses the exchange of a proton at N6 of adenine with a methyl group, via an S_N2 reaction without the formation of a covalent intermediate, in the exact same manner as M.*Taq*I. Asn^{133} and Pro^{134} are analogous to Asn^{105} and Pro^{106} in M.*Taq*I, serving as hydrogen bond acceptors for the protons of the exocyclic N6 group [Fig. 3.5] and contributing to its activation for nucleophilic attack. Tyr^{136} is the equivalent of Tyr^{108} in M.*Taq*I, which stabilises the positively charged 6-methylammonium intermediate and the cationic transition state by cation-π interactions. The hydrogen bonding pattern [Fig. 3.5] is of critical importance for the maintenance of catalytic activity. Disruption of the bond between Asn^{133} and N1 of the adenine ring probably leads to catalytic inactivation, as mutational studies on analogous residues of M.*Eco*RV, *Eco*KI, *Bcg*I and EcoDam have shown (Roth *et al.*, 1998). Alterations in all four residues of EcoDam's DPPY motif, also abolished MTase activity (Guyot *et al.*, 1993).

Similarly again to M.*Taq*I, the formation of the planar 6-methylamino group in the final step of the reaction probably leads to steric overlaps with the active site residues, triggering by extent the flipping of the methylated adenine back into the DNA helix. Closer inspection of M.*Bse*CI_{FM}, which corresponds to the end product of the methylation reaction, shows that the methylated adenine is slightly drifted away

from the side chains of Phe^{219} and Asn^{133} by an average of 0.3-0.4 Å, compared to its position in M.*Bse*CI$_{HM}$, and that the hydrogen bond between the side chain N$^{\delta 2}$ of Asn^{133} and N1 of the adenine ring is disrupted. The changes in the interaction pattern, and particularly the weakening of the π-stacking interaction with Phe^{219}, cause a loss in thermodynamic stability, which should be gained by interactions in the DNA helix.

3.5 Structural basis for recognition of hemimethylated DNA

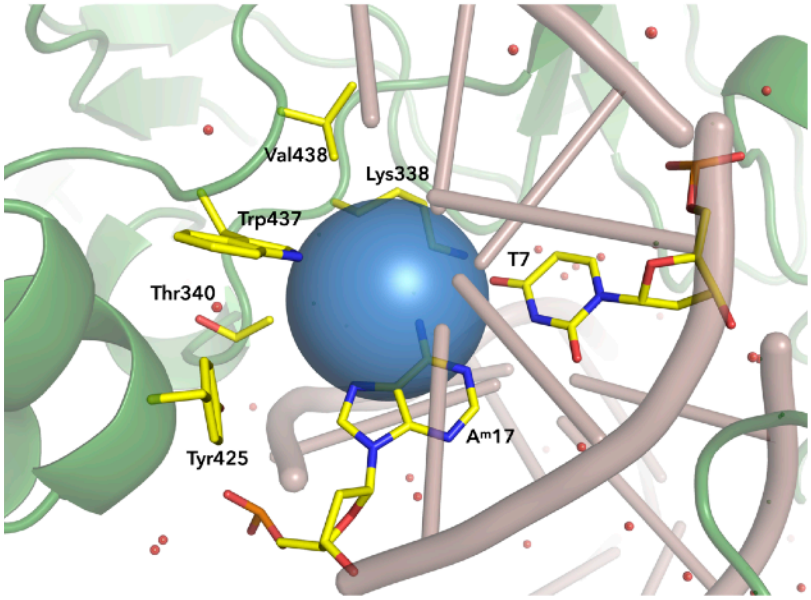

Fig. 3.7 Schematic representation showing the structural environment around the methyl group of the non-target adenine. M.*Bse*CI is shown in green and DNA in light pink colour. Side chains facing towards the methyl group and A^{m}17:T4 base pair are shown in liquorice representation with atoms coloured as follows: C, yellow; N, blue; O, red; P, orange. The methyl group is represented as a blue sphere with a 2 Å radius, corresponding to its actual van der Waals radius. Water molecules are shown as red spheres.

Despite the amount of structural information that has been acquired in recent decades regarding bacterial MTases, the mechanism by which they discriminate hemimethylated target sequences remains still unclear. This discrimination mechanism has been elucidated for a well-studied eukaryotic DNA methylase, DNMT1, which is responsible for maintaining the CpG methylation pattern; the preferential affinity for hemimethylated DNA sequences is offered by UHRF1, which facilitates the loading of DNMT1 onto appropriate sites of newly replicated DNA (Avvakumov *et al.*, 2008). But hypotheses postulated so far for bacterial MTases are very limited, and vary from residues interacting directly with the pre-existing methyl group in the non-target strand (Reinisch *et al.*, 1995), to spontaneous turn arounds and rebindings of the enzyme until finding the correct position (Zhou *et al.*, 2021). The question that arises here is whether the crystal structures of M.*Bse*CI ternary complexes could offer any tangible information on how the enzyme finds the correct orientation upon DNA binding.

As mentioned above, the methyl group in the non-target strand is not involved in any hydrogen bond contacts with neighbouring residues. However, the surrounding area is formed by a considerable number of hydrophobic and neutral amino acid residues, despite its close proximity to the negatively charged DNA backbone and its active role in direct specific DNA binding. The side chains of Tyr^{425}, Trp^{437}, Val^{438}, all lie around the methyl group forming the core of a small hydrophobic pocket around the T4:A^{m}17 base pair and the base pairs astride [Fig. 3.7]. The side chains of Trp^{437} and Lys^{338} (the only

charged residue in the area) are placed in such way that their hydrogen bond donors are facing away from the A^m17 nucleobase, participating in specific contacts with the bases on either side of the T4:A^m17 pair. The pocket is also formed by a moderately hydrophilic Thr^{340} side chain with the $O^{\gamma 1}$ group facing away from the methyl group. Such unusual structural arrangement around the non-target, methylated adenine might therefore imply a possible scanning mechanism for M.*Bse*CI: the correct positioning of the enzyme onto the hemimethylated target DNA sequence is likely to be driven by hydrophobic forces with the pre-existing methyl group of the non-target adenine. Yet, this assumption could be wrong, as structural information cannot provide the full picture of what normally happens inside the cellular environment, and the possibility of the existence of another reading mechanism or even the presence of another protein that mediates the correct positioning of the enzyme needs to be further studied.

Chapter 4

Conclusions

The structure of a second γ-N6mA-MTase, M.*Bse*CI, both in its apo form and in complex with its cognate DNA duplex in three different methylation states, allows us to better understand the structural basis of adenine-specific methylation and begin comparing mechanisms of gaining access to substrate nucleotides, specific binding, and catalysis. It is obvious that features of the adenine methylation reaction are broadly conserved within this class of enzymes, as happens with other MTase classes. Both M.*Bse*CI and M.*Taq*I —as expected—flip the target adenine out of the DNA helix and into active site pockets located in their N-terminal domains, which have similar structural organisation and topology. Contrary to the sequence and structural conservation of the N-terminal domains, there are substantial differences in the C-terminal domains, which make the majority of the base-specific contacts to the recognition sequence. Yet, a significant portion in both enzymes' C-terminal domain exhibits an obvious structural homology, and similarities also extend to the amino acid residues involved in sequence-specific binding and their contacts pattern. And although one would expect that this fold is likely to be necessary for the recognition of the 5′-TCGA-3′ element, this structural conservation expands to even a greater extent to the orphan CamA N6mA-MTase, raising questions about the evolutionary origin of the sequence-specific binding apparatus in all N6mA-MTases.

The DNA distortion in all three aforementioned adenine MTases also exhibits similarities, with its backbone around the target flipped adenine differing dramatically from the normal structure of the B-DNA form. The obvious backbone squeezing around the flipped base, the absence of any shifting in the base pairing, as well as the absence of residues occupying the vacated space opposite the unpaired base, might be characteristics found only in γ-N6mA-MTases. However, the temporal sequence of events and the mechanistic insights that promote the flipping of the base have yet to be determined (Zhou *et al.*, 2021); it is still unknown whether the squeezing precedes and triggers the extrusion of the base, or follows it and prevents the base from flipping back into the DNA helix. And there is also the problem of returning the methylated adenine into the helix. Notably the base flipping mechanism seems to be quite versatile. So-far insistently studied M.*Hha*I, M.*Hae*III and M.*Taq*I, all use three different means of stabilising DNA structure with a flipped base. There is still a number of questions that needs to be answered regarding base flipping, including how it is initiated and to what extent it is related to the recognition of the target sequence. Previous findings in M.*Hha*I bound to a DNA substrate with an abasic site at the position of the target cytosine, show that the enzyme still moves the sugar-phosphate backbone to the "flipped-out" posi-

tion (O'Gara *et al.*, 1998). Similar findings have been also observed for the flipped-out abasic nucleotide in different glycosylase-DNA complexes (Cheng & Roberts, 2001). It seems likely then that the base *per se* does not promote the structural changes of the DNA, but it is rather the backbone which is targeted for rotation by the enzyme, and the base is simply carried along with it (Cheng & Roberts, 2001; Roberts & Cheng, 1998). On top of that, base flipping in M.*Hha*I probably works with any base at the target position, meaning that the methylation reaction is sensitive to the base itself, but the base flipping step is not (Klimašauskas & Roberts, 1995; Yang *et al.*, 1995). Aside from the base flipping mechanism, questions about the recognition of the target site are also raised; it is not certain, for instance, whether the contacts between the protein and the DNA that are seen in these structures are the ones responsible for sequence-specific recognition during the initial binding, and perhaps this initial recognition is cooperatively related to the flipping mechanism.

The most intriguing question that arises, though, about DNA MTases, is how the enzyme discriminates the hemimethyated sequences over their unmethylated counterparts. Inspection of the area around the methylated non-target adenine in the M.*Bse*CI ternary complex did reveal a small, relatively hydrophobic platform in the C-terminal domain, implying that hydrophobic forces with the pre-methylated adenine might guide the correct positioning of the enzyme in the presence of hemimethylated sequences. Further inspection of the M.*Taq*I ternary complex, however, did not reveal any similarities in the structural arrangement of the enzyme around the methylated base. Instead, the methyl group is juxtaposed against a Ser residue, similar to 5mC-M.*Hae*III for which Reinisch *et al.* (1995) had hypothesised a connection with a possible discrimination mechanism, indicating that such mechanisms might be unique for each bacterial DNA MTase. The discrimination mechanism still remains an elusive process across all MTase families and solely crystallographic data will hardly provide a definite answer to this problem.

As it is obvious, the determination of the structure of M.*Bse*CI is not just a plain addition of another MTase to the existing literature. DNA MTases constitute a large family of enzymes with an apparent structural homology and a similar catalytic mechanism. Interpreting the functional and mechanistic features of adenine-specific MTases, gives the opportunity to compare and better understand characteristics of various other MTases, and aggrandise the interest towards further research regarding the enzymology and biology of DNA MTases. More structures of related MTases with various recognition sites are necessary to understand DNA recognition by these enzymes, providing a unique tool to further study the molecular evolution of DNA recognition and elucidate the functional relevance of the contacts provided by DNA binding enzymes. Structures of MTases complexed with both non-specific and specific DNA in unflipped state could also shed light on the conformational changes of the enzyme-DNA complex during enzymatic turnover, and aid in delving deeper into the physicochemical mechanism of base flipping. Accumulation of knowledge in DNA methylation can even further shift the interest towards important applications in drug development. Since DNA methylation has been found to be occasionally linked with bacterial pathogenicity, and adenine methylation in particular is not present in mammalian nuclear DNA, various drug-design initiatives could potentially emerge by exploiting such enzymes in favour of blocking methylation and thus slowing down bacterial growth.

In summary, while the four structures presented here offer a broad view on the structural characteristics of M.*Bse*CI and its obvious similarities to previously solved MTases, there is a dire need for bet-

ter understanding binding, catalytic and kinetic mechanisms of both existing and newly identified MTases. The work reported here will hopefully provide the stimulus for further studying and deeper understanding of DNA modifying enzymes, up to the extent of making feasible the exploitation of their unusual features for modern therapeutic targets.

✢

References

Avvakumov, G.; Walker, J.; Xue, S.; Li, Y.; Duan, S.; Bronner, C.; Arrowsmith, C.; Dhe-Paganon, S. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **2008**, *455*, 822-825.

Adams, G. M.; Blumenthal, R. M. The *Pvu*II DNA (cytosine-*N4*)-methyltransferase comprises two trypsin-defined domains, each of which binds a molecule of *S*-adenosyl-L-methionine. *Biochemistry* **1997**, *36*, 8284-8292.

Arber, W.; Dussoix, D. Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage λ. *J. Mol. Biol.* **1962**, *5*, 18-36.

Athanasiadis, A.; Papanikolau, Y.; Rina, M.; Papadovasilaki, M.; Dauter, Z.; Petratos, K.; Bouriotis, V.; Kokkinidis, M. Purification, crystallization and preliminary X-ray analysis of the M.*Bse*CI DNA methyltransferase from *Bacillus stearothermophilus*. *Acta Crystallogr. D Biol. Crystallogr.* **1997**, *53*, 477-479.

Battye, T. G.; Kontogiannis, L.; Johnson, O.; Powell, H. R.; Leslie, A. G. *iMOSFLM*: a new graphical interface for diffraction-image processing with *MOSFLM*. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 271-281.

Beck, C.; Cranz, S.; Solmaz, M.; Roth, M.; Jeltsch, A. How does a DNA interacting enzyme change its specificity during molecular evolution? A site-directed mutagenesis study at the DNA binding site of the DNA-(adenine-N6)-methyltransferase *Eco*RV. *Biochemistry* **2001**, *40*, 10956-10965.

Berdis, A. J.; Lee, I.; Coward, J. K.; Stephens, C.; Wright, R.; Shapiro, L.; Benkovic, S. J. A cell cycle-regulated adenine DNA methyltransferase from *Caulobacter crescentus* processively methylates GANTC sites on hemimethylated DNA. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 2874-2879.

Bergerat, A.; Guschlbauer, W. The double role of methyl donor and allosteric effector of S-adenosyl-methionine for Dam methylase of *E. coli*. *Nucleic Acids Res.* **1990**, *18*, 4369-7435.

Bergerat, A.; Guschlbauer, W.; Fazakerley, G. V. Allosteric and catalytic binding of *S*-adenosylmethionine to *Escherichia coli* DNA adenine methyltransferase monitored by ^{3}H NMR. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 6394-6397.

Bestor, T.; Laudano, A.; Mattaliano, R.; Ingram, V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.* **1988**, *203*, 971-983.

Cerritelli, S.; Springhorn, S. S.; Lacks, S. A. DpnA, a methylase for single-strand DNA in the *Dpn* II restriction system, and its biological function. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 9223-9227.

Cheng, X. Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 293-318.

Cheng, X.; Blumenthal, R. M. Finding a basis for flipping bases. *Structure* **1996**, *4*, 639-645.

Cheng, X.; Kumar, S.; Posfai, J.; Pflugrath, J. W.; Roberts, R. J. Crystal structure of the *Hha*I DNA methyltransferase complexed with *S*-adenosyl-L-methionine. *Cell* **1993**, *74*, 299-307.

Cheng, X.; Roberts, R. J. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.* **2001**, *29*, 3784-3795.

Cowtan, K. Completion of autobuilt protein models using a database of protein fragments. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68*, 328-335.

Danna, K.; Nathans, D. Specific cleavage of simian virus 40 DNA by restriction endonuclease of *Hemophilus influenzae*. *Proc. Natl. Acad. Sci. USA* **1971**, *68*, 2913-2917.

Dixon, M.; Huang, S.; Matthews, R.; Ludwig, M. The structure of the C-terminal domain of methionine synthase: presenting S-adenosylmethionine for reductive methylation of B_{12}. *Structure* **1996**, *4*, 1263-1275.

Djordjevic, S.; Stock, A. Crystal structure of the chemotaxis receptor methyltransferase Cher suggests a conserved structural motif for binding *S*-adenosylmethionine. *Structure* **1997**, *5*, 545-558.

Emsley, P.; Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2126-2132.

Evans, P. R.; Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69*, 1204-1214.

Evdokimov, A. A.; Zinoviev, V. V.; Malygin, E. G.; Schlagman, S. L.; Hattman, S. Bacteriophage T4 Dam DNA-[N^6-adenine]methyltransferase. Kinetic evidence for a catalytically essential conformational change in the ternary complex. *J. Biol. Chem.* **2002**, *277*, 279-286.

Fu, Z.; Hu, Y.; Konishi, K.; Takata, Y.; Ogawa, H.; Gomi, T.; Fujioka, M.; Takusagawa, F. Crystal structure of glycine *N*-methyltransferase from rat liver. *Biochemistry* **1996**, *35*, 11985-11993.

Glykos, N. M.; Kokkinidis, M. *GraphEnt*: a maximum-entropy program with graphics capabilities. *J. Appl. Crystallogr.* **2000**, *33*, 982-985.

Goedecke, K.; Pignot, M.; Goody, R. S.; Scheidig, A. J.; Weinhold, E. Structure of the N6-adenine DNA methyltransferase M.*Taq*I in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **2001**, *8*, 121-125.

Gold, M.; Hausmann, R.; Maitra, U.; Hurwitz, J. The enzymatic methylation of RNA and DNA. VIII. Effects of bacteriophage infection on the activity of the methylating enzymes. *Proc. Natl. Acad. Sci. USA* **1964**, *52*, 292-297.

Gold, M.; Hurwitz, J.; The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. V. Purification and properties of the deoxyribonucleic acid-methylating activity of *Escherichia coli*. *J. Biol. Chem.* **1964a**, *239*, 3858-3865.

Gold, M.; Hurwitz, J. The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. VI. Further studies on the properties of the deoxyribonucleic acid methylation reaction. *J. Biol. Chem.* **1964b**, *239*, 3866-3874.

Gold, M.; Hurwitz, J.; Anders, M. The enzymatic methylation of RNA and DNA, II. On the species specificity of the methylation enzymes. *Proc. Natl. Acad. Sci. USA* **1963**, *50*, 164-169.

Gong, W.; O'Gara, M.; Blumenthal, R. M.; Cheng, X. Structure of *Pvu*II DNA-(cytosine N4) methyltransferase, an example of domain permutation and protein fold assignment. *Nucleic Acids Res.* **1997**, *25*, 2702-2715.

González, E.; Vásquez, C. Characterization of the *bstVIRM* genes encoding the *Bacillus stearothermophilus* V restriction-modification system. *Gene* **1993**, *131*, 103-106.

Gowher, H; Jeltsch, A. Molecular enzymology of the *Eco*RV DNA-(adenine-N^6)-methyltransferase: kinetics of DNA binding and bending, kinetic mechanism and linear diffusion of the enzyme on DNA. *J. Mol. Biol.* **2000**, *303*, 93-110.

Grosjean, H. *DNA and RNA modification enzymes: Structure, Mechanism, Function and Evolution*; Landes Bioscience: Austin, Texas, **2009**.

Guschlbauer, W. The DNA and *S*-adenosylmethionine-binding regions of *Eco*Dam and related methyltransferases. *Gene* **1988**, *74*, 211-214.

Guyot, J.; Grassi, J.; Hahn, U.; Guschlbauer, W. The role of the preserved sequences of Dam methylase. *Nucleic Acids Res.* **1993**, *21*, 3183-3190.

Hodel, A.; Gershon, P.; Shi, X.; Quiocho, F. The 1.85 Å structure of vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends. *Cell* **1996**, *85*, 247-256.

Holm, L.; Sander, C. An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins* **1997**, *28*, 72-82.

Holz, B.; Dank, N.; Eickhoff, J. E.; Lipps, G.; Krauss, G.; Weinhold, E. Identification of the binding site for the extrahelical target base in N^6-adenine DNA methyltransferases by photo-cross-linking with duplex oligodeoxyribonucleotides containing 5-iodouracil at the target position. *J. Biol. Chem.* **1999**, *274*, 15066-15072.

Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* **1948**, *175*, 315-332.

Hurwitz, J.; Gold, M.; Anders, M. The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. III. Purification of soluble ribonucleic acid-methylating enzymes. *J. Biol. Chem.* **1964a**, *239*, 3462-3473.

Hurwitz, J.; Gold, M.; Anders, M. The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. IV. The properties of the soluble ribonucleic acid-methylating enzymes. *J. Biol. Chem.* **1964b**, *239*, 3474-3482.

Ito, H.; Sadaoka, A.; Kotani, H.; Hiraoka, N.; Nakamura, T. Cloning, nucleotide sequence, and expression of the H*inc*II restriction-modification system. *Nucleic Acids Res.* **1990**, *18*, 3903-3911.

Jeltsch, A. Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chem. Eur. J. Chem. Biol.* **2002**, *3*, 275-293.

Jeltsch, A.; Jurkowska, R. *DNA Methyltransferases - Role and Function*; Springer International Publishing AG Switzerland, **2016**.

Jeltsch, A.; Roth, M.; Friedrich, T. Mutational analysis of target base flipping by the *Eco*RV adenine-N^6 DNA methyltransferase. *J. Mol. Biol.* **1999**, *285*, 1121-1130.

Johnson, T. B.; Coghill, R. D. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the Tubercle bacillus. *J. Am. Chem. Soc.* **1925**, *47*, 2838-2844.

Juanhuix, J.; Gil-Ortiz, F.; Cuní, G.; Colldelram, C.; Nicolás, J.; Lidón, J.; Boter, E.; Ruget, C.; Ferrer, S.; Benach, J. Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *J. Synchrotron. Radiat.* **2014**, *21*, 679-689.

Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants **1993**, *J. Appl. Cryst. 26*, 795-800.

Kapetaniou, E. G.; Kotsifaki, D.; Providaki, M.; Rina, M.; Bouriotis, V.; Kokkinidis, M. Purification, crystallization and preliminary X-ray analysis of the *Bse*CI DNA methyltransferase from *Bacillus stearothermophilus* in complex with its cognate DNA. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2007**, *63*, 12-14.

Kiss, A.; Pósfai, G.; Zsurka, G.; Raskó, T.; Venetianer, P. Role of DNA minor groove interactions in substrate recognition by the M.*Sin*I and M.*Eco*RII DNA (cytosine-5) methyltransferases. *Nucleic Acids Res*. **2001**, *29*, 3188-3194.

Klimašauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **1994**, *76*, 357-369.

Klimašauskas, S.; Roberts, R. M.*Hha*I binds tightly to substrates containing mismatches at the target base. *Nucleic Acids Res.* **1995**, *23*, 1388-1395.

Klimašauskas, S.; Timinskas, A.; Menkevicius, S.; Butkiene, D.; Butkus, V.; Janulaitis, A. Sequence motifs characteristic of DNA[cytosine-N4]methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. *Nucleic Acids Res.* **1989**, *17*, 9823-9832.

Kumar, S.; Cheng, X.; Klimašauskas, S.; Mi, S.; Posfai, J.; Roberts, R. J.; Wilson, G. G. The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* **1994**, *22*, 1-10.

Kumar, S.; Johnson, W. S.; Tomasz, M. Orientation isomers of the mitomycin C interstrand cross-link in non-self-complementary DNA. Differential effect of the two isomers on restriction endonuclease cleavage at a nearby site. *Biochemistry* **1993**, *32*, 1364-1372.

Labahn, J.; Granzin, J.; Schluckebier, G.; Robinson, D.; Jack, W.; Schildkraut, I.; Saenger, W. Three-dimensional structure of the adenine-specific DNA methyltransferase M.*Taq*I in complex with the cofactor *S*-adenosylmethionine. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 10957-10961.

Langer, G.; Cohen, S. X.; Lamzin, V. S.; Perrakis, A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **2008**, *3*, 1171-1179.

Lauster, R.; Kriebardis, A.; Guschlbauer, W. The GATATC-modification enzyme *Eco*RV is closely related to the GATC-recognizing methyltransferases *Dpn*II and *dam* from *E. coli* and phage T4. *FEBS Lett*. **1987**, *220*, 167-176.

Lauster, R.; Trautner, T. A.; Noyer-Weidner, M. Cytosine-specific type II DNA methyltransferases. A conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.* **1989**, *206*, 305-312.

Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L. W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *Phenix*. *Acta Crystallogr. D Struct. Biol.* **2019**, *75*, 861-877.

Lloyd, R. S.; Cheng, X. Mechanistic link between DNA methyltransferases and DNA repair enzymes by base flipping. *Biopolymers* **1997**, *44*, 139-151.

Loenen, W. A.; Dryden, D. T.; Raleigh, E. A.; Wilson, G. G.; Murray, N. E. Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Res.* **2014**, *42*, 3-19.

Malone, T.; Blumenthal, R. M.; Cheng. X. Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J. Mol. Biol.* **1995**, *253*, 618-632.

Malygin, E. G.; Evdokimov, A. A.; Zinoviev, V. V.; Ovechkina, L. G.; Lindstrom, W. M.; Reich, N. O.; Schlagman, S. L.; Hattman, S. A dual role for substrate *S*-adenosyl-L-methionine in the methylation reaction with bacteriophage T4 Dam DNA-[*N*6-adenine]-methyltransferase. *Nucleic Acids Res.* **2001**, *29*, 2361-2369.

Martin, J. L.; McMillan, F. M. SAM (dependent) I AM: the S-adenosylmethionine-dependent methyltransferase fold. *Curr. Opin. Struct. Biol.* **2002**, *12*, 783-793.

McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *Phaser* crystallographic software. *J. Appl. Crystallogr.* **2007**, *40*, 658-674.

O'Gara, M.; Horton, J.; Roberts, R.; Cheng, X. Structures of *Hha*I methyltransferase complexed with substrates containing mismatches at the target base. *Nat. Struct. Biol.* **1998**, *5*, 872-877.

O'Gara, M.; Zhang, X.; Roberts, R. J.; Cheng, X. Structure of a binary complex of *Hha*I methyltransferase with *S*-adenosyl-L-methionine formed in the presence of a short non-specific DNA oligonucleotide. *J. Mol. Biol.* **1999**, *287*, 201-209.

Pingoud, A.; Jeltsch, A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **2001**, *29*, 3705-3727.

Reinisch, K. M.; Chen, L.; Verdine, G. L.; Lipscomb, W. N. The crystal structure of HaeIII methyltransferase convalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* **1995**, *82*, 143-153.

Renbaum, P.; Razin, A. Mode of action of the *Spiroplasma* CpG methylase M.SssI. *FEBS Lett.* **1992**. *313*, 243-247.

Rina, M.; Bouriotis, V. Cloning, purification and characterization of the *Bse*CI DNA methyltransferase from *Bacillus stearothermophilus*. *Gene* **1993**, *133*, 91-94.

Rina, M.; Dialektakis, D.; Clark, D.; Pagomenou, M.; Bouriotis, V. Isolation and identification of restriction endonuclease BseCI. *Nucleic Acids Res.* **1992**, *20*, 1807.

Rina, M.; Markaki, M.; Bouriotis. V. Sequence of the cloned *bseCIM* gene: M.*Bse*CI reveals high homology to M.*Ban*III. *Gene* **1994**, *150*, 71-73.

Roberts, R. J. On base flipping. *Cell* **1995**, *82*, 9-12.

Roberts, R. J.; Cheng, X. Base flipping. *Annu. Rev. Biochem.* **1998**, *67*, 181-198.

Roth, M.; Helm-Kruse, S.; Friedrich, T.; Jeltsch, A. Functional roles of conserved amino acid residues in DNA methyltransferases investigated by site-directed mutagenesis of the *Eco*RV adenine-N^6-methyltransferase. *J.Biol. Chem.* **1998**, *273*, 17333-17342.

Scavetta, R. Structure of *Rsr*I methyltransferase, a member of the *N*6-adenine β class of DNA methyltransferases. *Nucleic Acids Res.* **2000**, *28*, 3950-3961.

Schluckebier, G.; Kozak, M.; Bleimling, N.; Weinhold, E.; Saenger, W. Differential binding of *S*-adenosylmethionine *S*-adenosylhomocysteine and Sinefungin to the adenine-specific DNA methyltransferase M.*Taq*I. *J. Mol. Biol.* **1997**, *265*, 56-67.

Schluckebier, G.; O'Gara, M.; Saenger, W.; Cheng, X. Universal catalytic domain structure of AdoMet-dependent methyltransferases. *J. Mol. Biol.* **1995**, *247*, 16-20.

Smith, H. O.; Wilcox, K. W. A Restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties. *J. Mol. Biol.* **1970**, *51*, 379-391.

Somers, W.; Phillips, S. Crystal structure of the *met* represser-operator complex at 2.8 Å resolution reveals DNA recognition by *β*-strands. *Nature* **1992**, *359*, 387-393.

Surby, M. A.; Reich, N. O. Contribution of facilitated diffusion and processive catalysis to enzyme efficiency: implications for the *Eco*RI restriction-modification system. *Biochemistry* **1996**, *35*, 2201-2208.

Tran, P. H.; Korszun, Z. R.; Cerritelli, S.; Springhorn, S. S.; Lacks, S. A. Crystal structure of the DpnM DNA adenine methyltransferase from the *Dpn*II restriction system of *Streptococcus pneumoniae* bound to *S*-adenosylmethionine. *Structure* **1998**, *6*, 1563-1575.

Urig, S.; Gowher, H.; Hermann, A.; Beck, C.; Fatemi, M.; Humeny, A.; Jeltsch, A. The *Escherichia coli* dam DNA methyltransferase modifies DNA in a highly processive reaction. *J. Mol. Biol.* **2002**, *319*, 1085-1096.

Wierenga, R.; Terpstra, P.; Hol, W. Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. *J. Mol. Biol.* **1986**, *187*, 101-107.

Vidgren, J.; Svensson, L.; Liljas, A. Crystal structure of catechol *O*-methyltransferase. *Nature* **1994**, *368*, 354-358.

Vilkaitis, G.; Merkiene, E.; Serva, S.; Weinhold, E.; Klimašauskas, S. The mechanism of DNA cytosine-5 methylation. Kinetic and mutational dissection of *Hha*I methyltransferase. *J. Biol. Chem.* **2001**, *276*, 20924-20934.

Williams, C.; Headd, J.; Moriarty, N.; Prisant, M.; Videau, L.; Deis, L.; Verma, V.; Keedy, D.; Hintze, B.; Chen, V.; Jain, S.; Lewis, S.; Arendall, W.; Snoeyink, J.; Adams, P.; Lovell, S.; Richardson, J.; Richardson, D. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2017**, *27*, 293-315.

Wilson, G. G.; Murray, N. E. Restriction and modification systems. *Annu. Rev. Genet.* **1991**, *25*, 585-627.

Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. Overview of the *CCP*4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 235-242.

Wong, D. L.; Reich, N. O. Identification of tyrosine 204 as the photo-cross-linking site in the DNA-*Eco*RI DNA methyltransferase complex by electrospray ionization mass spectrometry. *Biochemistry* **2000**, *319*, 15410-15417.

Yang, A.; Shen, J.; Zingg, J.; Mi, S.; Jones, P. *Hha*I and *Hpa*II DNA methyltransferases bind DNA mismatches, methylate uracil and block DNA repair. *Nucleic Acids Res.* **1995**, *23*, 1380-1387.

Zhou, J.; Horton, J.; Blumenthal, R.; Zhang, X.; Cheng, X. *Clostridioides difficile* specific DNA adenine methyltransferase CamA squeezes and flips adenine out of DNA helix. *Nat. Commun.* **2021**, *12*, 1-12.

Appendix

Table A1. Summary of the important steps during refinement of M.*Bse*CI model.

Building, rebuilding & model correction	**Refinement strategy**	***R* (%)**	R_{free} **(%)**
Model rebuilding (*RESOLVE*).	Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	17.0	22.1
Rotamers fitting, addition/deletion of H_2O molecules & loops remodelling.	TLS refinement with SA (T_i=4000 K, cool rate: 50 K). Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	16.9	21.0
Rotamers fitting, addition/deletion of H_2O molecules and removal of part of the N-terminal.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	16.3	20.7
Rotamers fitting, addition/deletion of H_2O molecules and removal of various disordered side chains.	TLS refinement with SA (T_i=4000 K, cool rate: 50 K). Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	16.3	20.4
Rotamers & Ramachandran outliers fitting. Addition/deletion of H_2O molecules. Addition of several side chains.	Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	16.6	20.6
Rotamers & Ramachandran outliers fitting. Addition/deletion of H_2O molecules. Removal of disordered loops.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	17.2	20.3
B-factor outliers. Addition/deletion of H_2O molecules.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained)	16.8	20.6
Residues modeled: Chain A: 23-110, 123-157, 160-220, 223-235, 242-257, 260-578 Chain B: 20-111, 122-220, 223-238, 242-257, 259-576			

Table A2. Summary of the important steps during refinement of M.*Bse*CI_{HM} model.

Building, rebuilding & model correction	**Refinement strategy**	***R* (%)**	**R_{free} (%)**
Automated solution from molecular replacement	Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	20.9	24.6
Rotamers fitting, addition of D573, I574, E575, AdoHcy and H_2O molecules.	Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	19.0	23.1
Rotamers fitting & DNA correction. Addition/deletion of H_2O molecules.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	18.6	22.2
Model rebuilding (*RESOLVE*).	Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained)	18.0	21.3
Rotamers fitting, addition/deletion of H_2O molecules. Addition of CH_3 group & loops remodelling.	TLS refinement with SA (T_i=4000 K, cool rate: 50 K). Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	16.6	20.6
Rotamers & Ramachandran outliers fitting. Addition of T9. Addition/deletion of H_2O molecules & loops remodelling.	TLS refinement with SA (T_i=4000 K, cool rate: 50 K). Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	18.2	21.9
Model rebuilding (*RESOLVE*).	Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	17.3	20.6
Rotamers fitting. Addition/deletion of H_2O molecules & loops remodelling	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	16.1	20.5
Model rebuilding (*RESOLVE*).	Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	16.3	20.5
Rotamers & Ramachandran outliers fitting. Addition/deletion of H_2O molecules & loops remodelling. Removal of K239 side chain.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	17.3	20.5
B-factor outliers. Addition/deletion of H_2O molecules.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	17.4	20.6
Residues modeled: 9-109, 122-576			

Table A3. Summary of the important steps during refinement of M.*Bse*CI$_{FM}$ model.

Building, rebuilding & model correction	Refinement strategy	R (%)	R_{free} (%)
Automated solution from molecular replacement	TLS refinement with SA (T_i=4000 K, cool rate: 50 K). Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	20.7	25.0
Rotamers fitting & removal of disordered loops. Addition/deletion of H_2O molecules.	TLS refinement with SA (T_i=4000 K, cool rate: 50 K). Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	18.8	22.2
Rotamers fitting & removal of disordered loops. Addition/deletion of H_2O molecules.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained)	18.9	21.9
Rotamers & Ramachandran outliers fitting. Removal of disordered residues. Addition/deletion of H_2O molecules.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained)	18.5	21.4
B-factor outliers. Addition/deletion of H_2O molecules.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained)	18.0	21.2
Residues modeled: 9-109, 122-238, 242-576			

Table A4. Summary of the important steps during refinement of M.*Bse*CI$_{UM}$ model.

Building, rebuilding & model correction	Refinement strategy	R (%)	R_{free} (%)
Rotamers fitting & DNA correction. Addition/deletion of H_2O molecules.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	21.5	25.8
Rotamers fitting & removal of disordered loops. Addition/deletion of H_2O molecules.	TLS refinement with SA (T_i=4000 K, cool rate: 50 K). Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	21.4	25.4
Model rebuilding (*RESOLVE*).	Individual sites & isotropic B-factors refinement. X-ray/stereochemistry weight optimisation (geometry was restrained).	20.6	25.2
Rotamers & Ramachandran outliers fitting. Addition/deletion of H_2O molecules.	TLS refinement with SA (T_i=4000 K, cool rate: 50 K). Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	19.2	21.7
Rotamers & Ramachandran outliers fitting. Removal of disordered residues. Addition/deletion of H_2O molecules.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	19.6	22.1
B-factor outliers. Addition/deletion of H_2O molecules.	TLS refinement. Individual sites & isotropic B-factors refinement. X-ray/stereochemistry — ADP weight optimisation (geometry was restrained).	19.6	22.2
Residues modeled: 10-109, 124-236, 242-379, 382-550, 553-576			

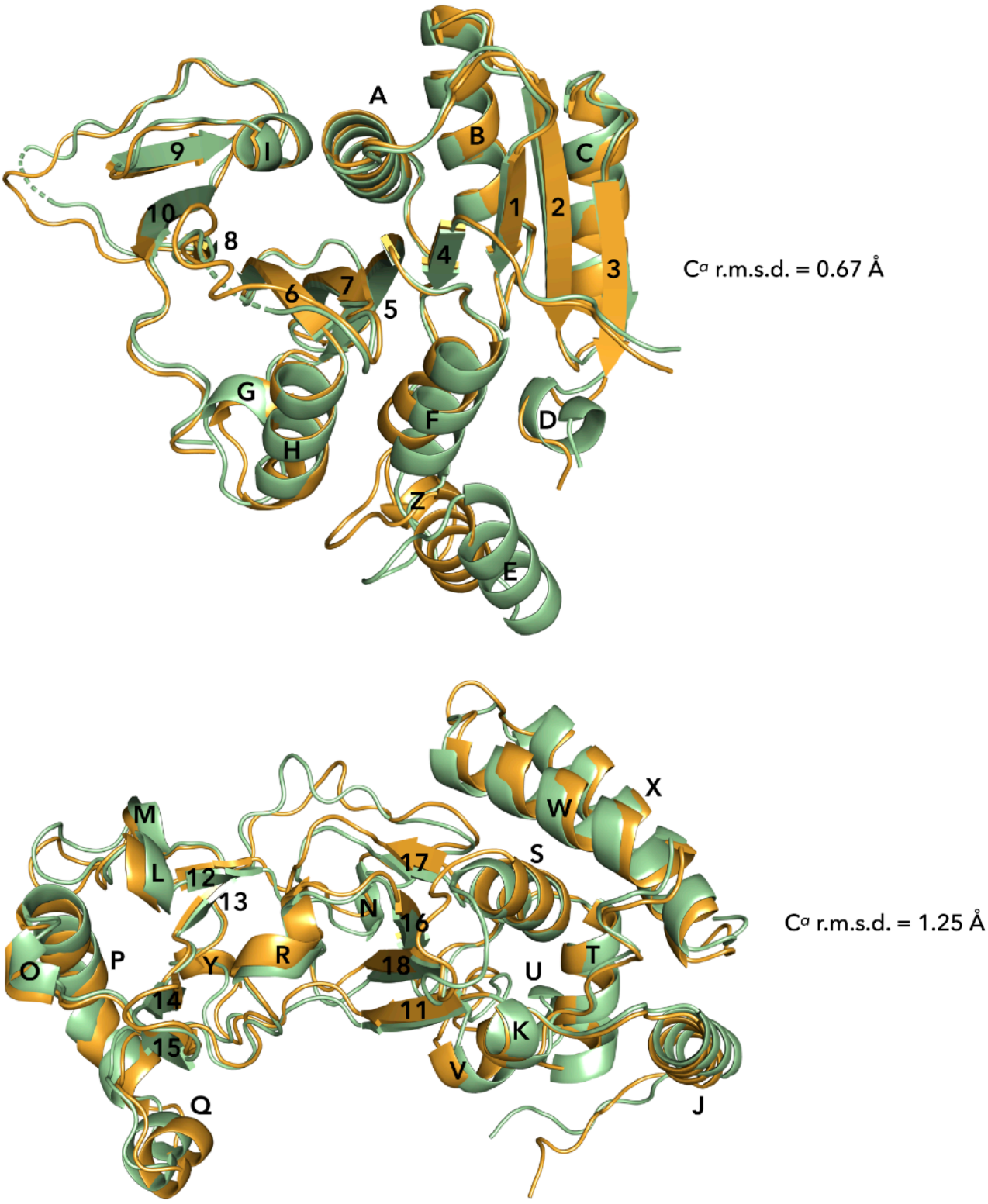

Fig. A1 Superposition of the M.*Bse*CI apoenzyme (green colour) and ternary complex (orange colour). DNA and AdoHcy are omitted. The two panels compare the structures of the N-terminal (upper) and C-terminal (lower) domains. The corresponding C^{α} r.m.s. deviations are 0.67 Å and 1.25 Å, respectively. Reformations of the polypeptide backbone upon DNA binding are minor, and pertain to components Z, Y, U, N and D, which interconvert between helical structures and coils. A small helical component of the N-terminal domain is also visible in the crystal structures of the complexes (not shown here). Such changes are not necessarily the result of stabilisation from DNA binding, but might be also attributed to insignificant, arbitrary, small changes in the torsion angles of the backbone, thus making the molecular graphics program interpret differently the secondary structure elements.

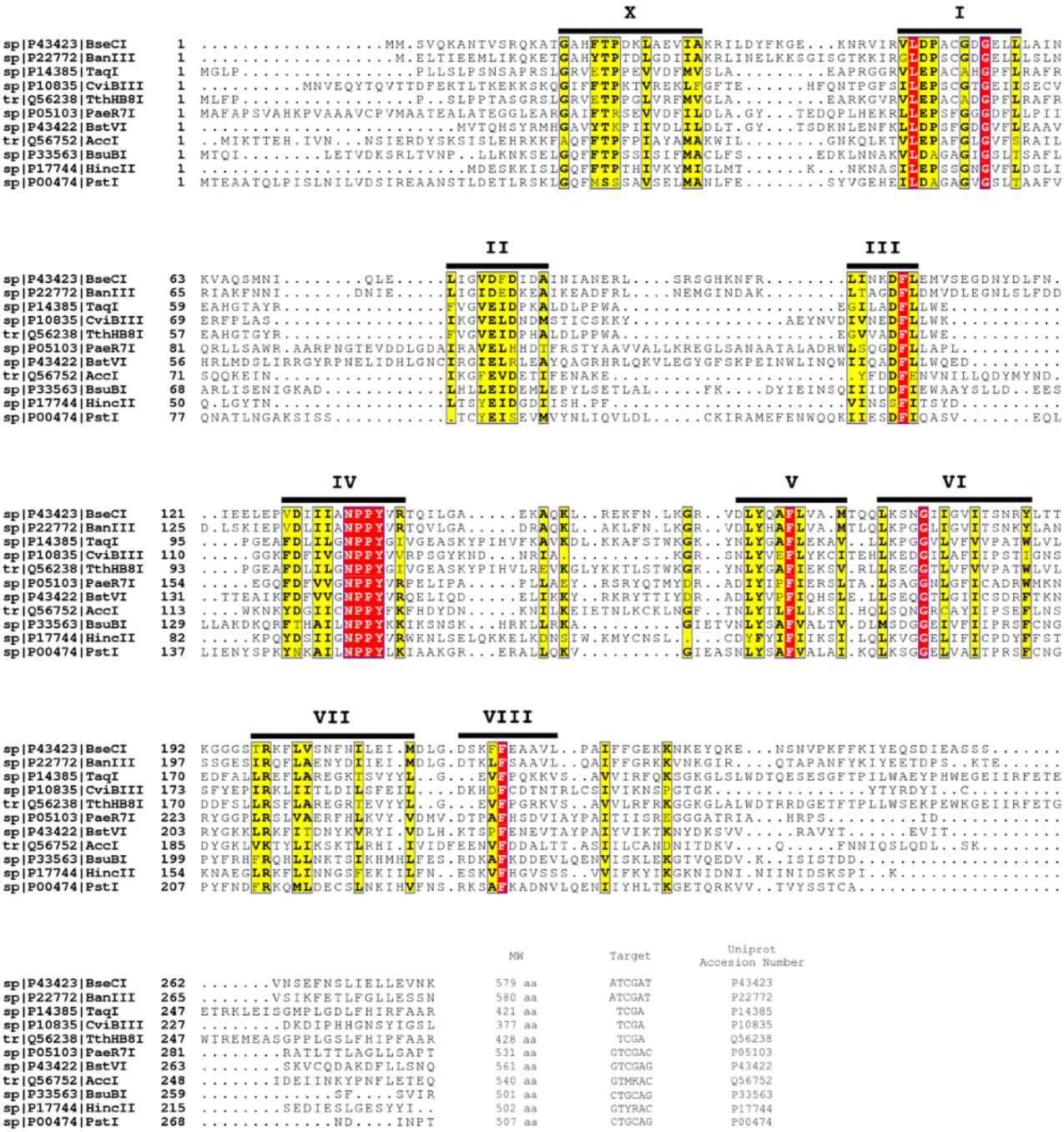

Fig. A2 Sequence alignement of eleven γ-N6mA-MTases. Shown here is only the alignment of their conserved N-terminal domains. Invariant amino acids are shown as white letters against a red background. Highly to moderately conserved positions are indicated against a yellow background. The bars above the sequences represent the limits of the nine conserved motifs.

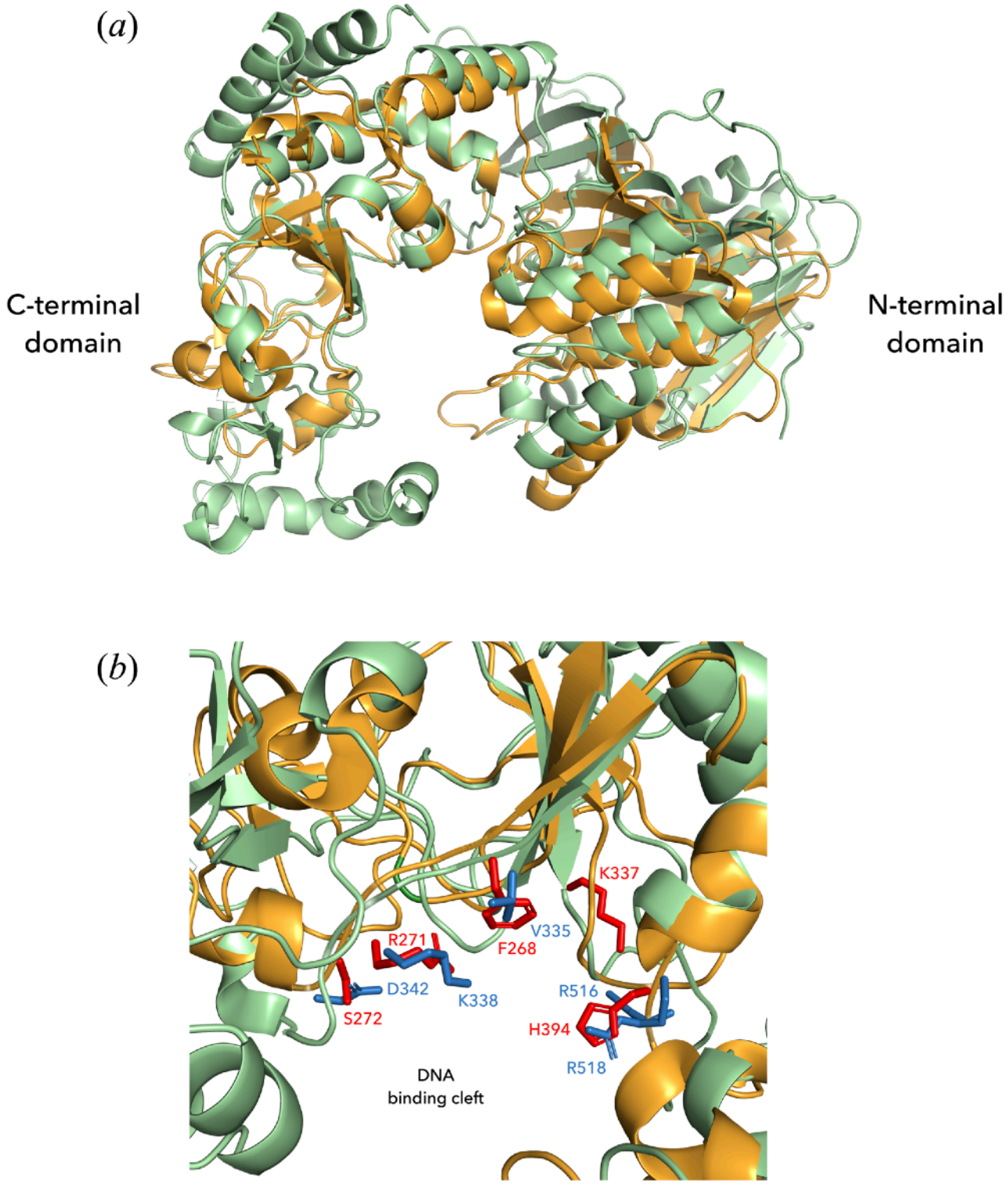

Fig. A3 M.*Bse*CI (green colour) and M.*Taq*I (orange colour) superimposed on their C-terminal domains. Panel (*a*) shows the similarities between the polypeptide chains of the C-terminal domains. The corresponding C^{α} r.m.s. deviation is 1.74 Å. Panel (*b*) illustrates the similarities between particular residues involved in specific DNA binding. Side chains belonging in amino acid residues of M.*Bse*CI are coloured blue, while those of M.*Taq*I are coloured red.

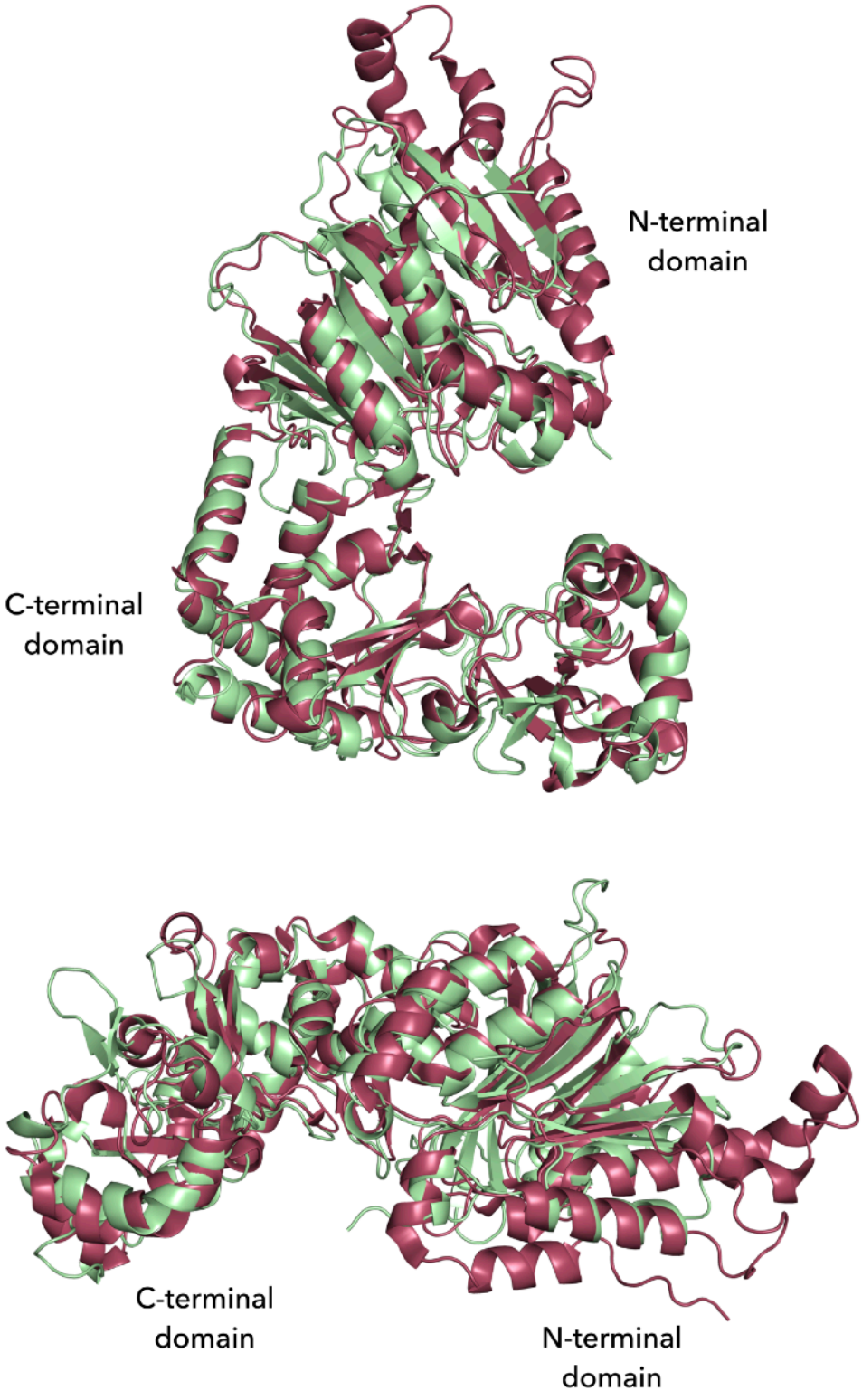

Fig. A4 Superposition of M.*Bse*CI (green colour) and CamA (red colour) adenine-specific MTases. Top view (upper) and side view (lower). The corresponding C^{α} r.m.s. deviation is 1.71 Å. DNA and cofactors are omitted.

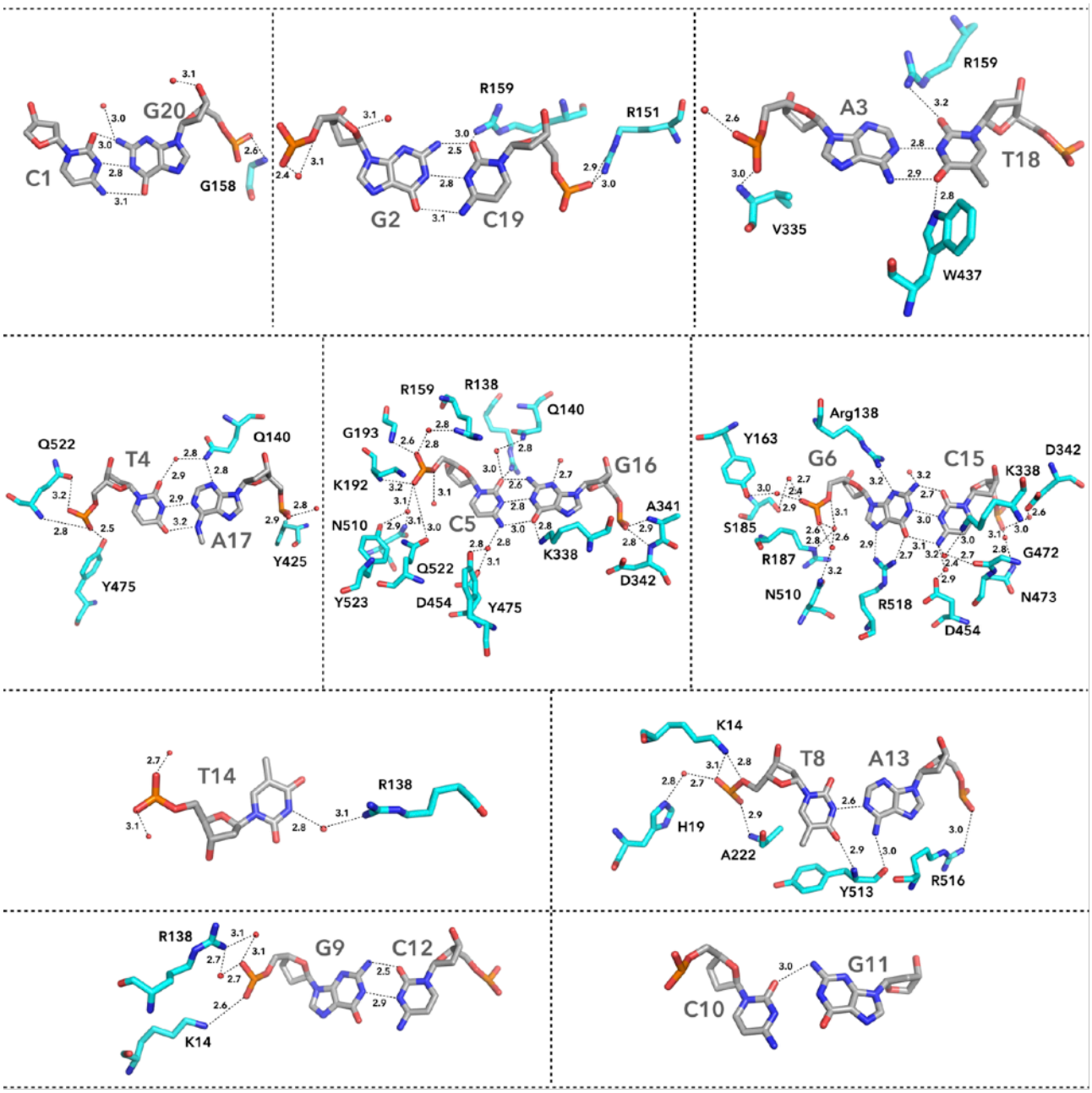

Fig. A5 Schematic illustration of the M.*Bse*CI-mediated DNA interactions. Atoms in liquorice representations are coloured as follows: C, grey in DNA and cyan in protein; N, blue; O, red; P, orange. Water molecules are shown as red spheres. Dotted lines represent hydrogen bond contacts, with their corresponding distances labeled accordingly (in Å).