

## Extraordinary Structures of Orphan Methyltransferases with Their Substrate DNA

**John Horton, Jujun Zhou, Clayton Woodcock, Xing Zhang, Xiaodong Cheng**

*Department of Epigenetics and Molecular Carcinogenesis, The University of Texas -M.D. Anderson Cancer Center, Houston TX, USA*

*jrhorton@mdanderson.org*

Recognition of DNA by proteins, both sequence and structure specific, is important in the functioning of the cell, such as in the processes of replication, transcription, and DNA repair. Twenty-five years after base flipping, a phenomenon whereby a base in normal B-DNA is swung completely out of the helix into an extrahelical position, was first observed in *HhaI* methyltransferase, we are still learning from and surprised by structures of protein-DNA complexes. The novel structures of the bacterium *Caulobacter crescentus* cell cycle-regulated DNA adenine methyltransferase (CcrM), as well as the newly discovered CamA enzyme (named for *Clostridioides difficile* adenine methyltransferase A) in complexes with double-stranded DNA containing their recognition sequence, will be discussed. Each of these enzymes affect their DNA substrate in a unique number of ways that are critical for their level of discrimination of their recognition DNA sequence.

CcrM in *C.crescentus* is responsible for maintenance methylation immediately after replication and methylates the adenine of hemimethylated GANTC. CcrM contains an N-terminal methyltransferase domain and a C-terminal nonspecific DNA-binding domain. CcrM is a dimer, with each monomer contacting primarily one DNA strand: the methyltransferase domain of one molecule binds the target strand, recognizes the target sequence, and catalyzes methyl transfer, while the C-terminal domain of the second molecule binds the non-target strand. The DNA contacts at the five base pair recognition site results in dramatic DNA distortions including bending, unwinding and base flipping. The two DNA strands are pulled apart, creating a bubble comprising four recognized base pairs. The five bases of the target strand are recognized meticulously by stacking contacts, van der Waals interactions and specific Watson-Crick polar hydrogen bonds to ensure high enzymatic specificity.

In the developed world, *C. difficile* is one of the leading causes of hospital-acquired infections. CamA-mediated methylation of the last adenine in CAAAAA is required for normal sporulation and biofilm production by this bacterium, a key step in disease transmission. Thus, selective inhibition of CamA has great therapeutic potential. CamA contains an N-terminal methyltransferase domain as well as a C-terminal DNA recognition domain. Major and minor groove DNA contacts in the recognition site involve base-specific hydrogen bonds, van der Waals contacts and the Watson-Crick pairing of a rearranged A:T base pair. These interactions provide sufficient sequence discrimination to ensure high specificity. In addition, this DNA methyltransferase has unusual features that may aid in discovery of a new selective antibiotic to combat *C. difficile* infection.

Knowledge acquired from these structures may also relate to other projects in our laboratory relating to mammalian epigenetics.

**Keywords: DNA methyltransferase, N6 adenine methylation**