Functional characterization and molecular evolution of *E. coli* DNA methyltransferase

by

Sanjay Chahar

A Thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy in Biochemistry

Approved Thesis Committee

Prof. Dr. Albert Jeltsch

Prof. Dr Georgi Muskhelishvili

Prof. Dr. Sebastian Springer

Prof. Dr. Alfred Pingoud

Date of Defense: December 14, 2009

School of Engineering and Science

List of Publications

1. Liebert, K., Horton, J. R., **Chahar, S**., Orwick, M., Cheng, X., & Jeltsch, A. (2007) Two alternative conformations of the AdoHcy bound to Escherichia Coli DNA adenine methyltransferase and the implication of conformational changes in regulating the catalytic cycle. *J. Biol. Chem.* **282** *, 22848-55*.

2. **Chahar, S**., Elsawy, H., Ragozin, S., and Jeltsch, A. (2009) Changing the DNA recognition specificity of *E.coli* DNA methyltransferase (EcoDam) enzyme by directed evolution. *J. Mol. Biol.* doi:10.1016/j.jmb.2009.09.027.

3. Elsawy, H., Podobinschi, S., **Chahar, S**., and Jeltsch, A. (2009) Transition from EcoDam to T4Dam DNA recognition mechanism without loss of activity and specificity. *ChemBioChem*. **10**, 2488-93.

Manuscripts in preparation

4. **Chahar, S**. *et al.,* Role of *E. coli* DNA methyltransferase (EcoDam) as a global regulator of gene expression in *E. coli*.

5. **Chahar, S**. *et al.,* The *E. coli* methyltransferase (EcoDam) enzyme binds to polyphosphates in cells and meditates potential connections between metabolic signaling and epigenetic modifications.

Acknowledgements

I would like to thank my Prof. Dr. Albert Jeltsch who has made his support in a number of ways, his excellent guidance, valuable discussions and encouragement which helped me in accomplishing this PhD dissertation. I would like to express my sincere gratitude to him for providing me the opportunity to work in his lab and introducing me to the exciting field of protein engineering and molecular epigenetics.

I am indebted to all of my present and former colleagues and friends to support me and providing a enjoyable working environment. It is a pleasure for me to show my deepest gratitude to Prof. Dr. Georgi Muskhelishvili and his group members for their constant support and encouragement.

I am thankful to Prof. Dr. Sebastian Springer and Prof. Dr. Alfred Pingoud for being co-referees of my PhD thesis. I am grateful to Prof. Dr. Xaiodong Cheng for his cooperation to our lab and his valuable discussions.

Finally, it is honor for me to thank Jacobs University for providing me the opportunity to accomplish my PhD dissertation in its outstanding international environment.

Abstract

DNA from most prokaryotes and eukaryotes contains methylated bases, i.e. 4-methylcytosine (N4mC), 5-methylcytosine (5mC) and 6-methyladenine (N6mA). These modifications are introduced after DNA replication by DNA methyltransferases (MTases), which catalyze methyl group transfer from the donor S-adenosyl-L-methionine (AdoMet), producing S-adenosyl-L-homocysteine (AdoHcy) and methylated DNA. DNA-adenine methylation at specific GATC sites plays a pivotal role in bacterial gene expression, DNA replication, mismatch repair, and bacterial virulence among gram-negative bacteria.

DNA MTases are attractive model system to study how proteins recognize specific sequences of DNA and how their specificity changes during molecular evolution. DNA methyltransferases (MTases) allow powerful combinations of *in vivo* and *in vitro* mutagenesis and DNA-based screening for enzymatic properties, which makes them one of the most powerful model system for directed *in vitro* evolution of proteins. In the present PhD work, through combination of various approaches of *in vitro* evolution, we have successfully changed the DNA recognition specificity of *E. coli* DNA methyltransferase (EcoDam) enzyme, that recognizes GATC palindromic sequence and methylates adenine at N^6 position, and have generated novel variants of EcoDam with new recognition specificity.

Understanding the pathways of natural evolution is a major scientific challenge. The EcoDam/T4Dam pair represents a very interesting model case to study molecular evolution. The EcoDam and T4Dam DNA-(adenine N6)-methyltransferases both methylate the adenine residue in GATC sites. These enzymes are highly related in amino acid sequence, but they deviate in their contact to the first base pair of the target sequence. EcoDam contacts Gua1 with K9 (which corresponds to T4Dam A6), while T4Dam contacts Gua1 with R130 (which corresponds to EcoDam Y138). In a rational protein design study of EcoDam, we have “transplanted” the T4Dam DNA recognition into EcoDam and show that the EcoDam K9A/Y138R double mutant is highly active and specific. We also studied the intermediates of this transition and report that the evolutionary transition from EcoDam to T4Dam might be driven by a selection pressure for increased catalytic activity.

Besides their prominent biological roles, Dam MTases are an ideal model system for a structure-based drug design studies. We have studied the inhibition mechanism of inhibitor-58 on EcoDam and characterized its biochemical properties.

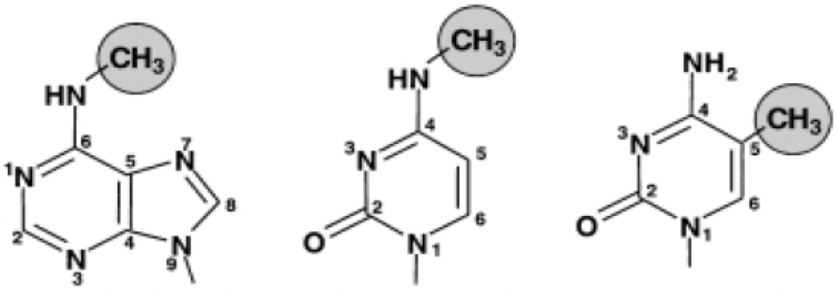
GATC sites are the natural substrate for EcoDam to methylate adenine at N6 position. The methylation status of GATC sites (methylated, unmethylated, and hemimethylated) can affect specific binding of DNA-interacting proteins. The presence of this tetranucleotide in the promoter or the regulatory sequences can affect gene expression by regulating binding of RNA polymerase or transcriptional regulators. We have studied the global transcriptional changes in the dam negative *E. coli* strain SCS110 in the presence and absence of EcoDam and report that EcoDam can serve as regulator of gene transcription. Bacterial cells encounter varied environmental changes and make appropriate adjustments to ensure their survival. We have studied the polyphosphates accumulation generated during stress response in *E. coli* and report that EcoDam binds to polyphosphates in cells and meditates potential connections between metabolic signaling and epigenetic modifications.

Table of contents

1 Introduction

1.1 DNA methylation

DNA from most prokaryotes and eukaryotes contains methylated bases (**Figure 1.0**), i.e. 4-methylcytosine (N4mC), 5-methylcytosine (5mC) and 6-methyladenine (N6mA). These methylated bases are natural components of the DNA which add another level of information to the cellular genome. The methylation modifications are introduced after DNA replication by DNA methyltransferases (MTases), which catalyze methyl group transfer from the donor S-adenosyl-L-methionine (AdoMet), producing S-adenosyl-L-homocysteine (AdoHcy) and methylated DNA (review: Cheng, 1995; Cheng and Blumenthal, 1999; Cheng and Roberts, 2001; Jeltsch, 2002). In *Escherichia coli* about 2% of adenines and 1% of cytosines are methylated (review Barras and Marinus, 1989).



***N*⁶-methyladenine** ***N*⁴-methylcytosine** **5-methylcytosine**

Figure 1.0: Structures of methylated bases occurring in DNA (Jeltsch, 2002)

AdoMet-dependent methyltransferases (MTases) have been characterized structurally by X-ray crystallography and NMR. These include DNA MTases, RNA MTases, protein MTases and small molecule MTases (Cheng, 2001). Generally, the DNA MTases recognize a specific sequence and utilize a so-called "base flipping" mechanism (Klimasauskas *et al.*, 1994) to rotate the target base within that sequence out of the DNA helix and insert it into the enzyme's active site pocket. Base flipping takes place in a biphasic manner, first the target base is rotated out of the DNA in a very fast reaction and

later the target base is tightly contacted by the enzyme and positioned in the active site pocket (Liebert *et al.*, 2004). The methylation does not interfere with the Watson/Crick pairing properties of adenine and cytosine, however it positions the methyl group in the major groove of DNA where it can be easily detected by DNA interacting proteins (Jeltsch, 2002). DNA methylation and other epigenetic mechanisms often play central regulatory roles in the control of the cellular physiology (Wion and Casadesus, 2006) as described in more detail in chapter 4 of the thesis.

1.2 Types and distribution of DNA methyltransferases

Based on the chemistry of methylation, DNA methyltransferases (MTases) can be subdivided into two groups: the N-type which forms a C-N bond (adenine-N^6 and cytosine-N^4 MTases) and the C-type which forms a C-C bond (cytosine-C5 Mtases). In general, both types of MTases are two-domain proteins comprising one large and one small domain with the DNA binding cleft being located at the domain interface. The large domain contains a set of conserved amino acid motifs, which differ between the N- and the C-MTases (Wilson, 1992; Malone, 1995). The N-MTases can be further subdivided into three classes (α, β, and γ) that differ from each other with respect to the position of insertion of the small domain into the framework of the large domain as well as by a circular permutation of the amino acid sequence of the large domain (**Figure 1.1**) (Faumann, 1999; Jeltsch, 1999). These differences result in a different arrangement of the most conserved amino acid sequence motifs that is characteristic for each class. The C-MTases are most similar to the γ class of N-MTases. Later, three new classes (ζ, δ and ε) have been reported for N-MTases. However, M.BssHI is the only DNA MTase for which the ζ architecture has been confirmed (Bujnicki, 2002).

Cytosine-C5 DNA modification is the most prevalent DNA modification in eukaryotic genomes. However, DNA methylation is not a ubiquitous feature of the eukaryotic genomes because some organisms, including *Saccharomyces cervisiae*, *Schizosaccharomyces pombe*, or *Caenorhabditis elegans*, lack detectable DNA

methylation (Colot and Rossignol, 1999). So far, C5-methylation has been found in invertebrates, fungi, protozoa, and all of the higher plants and vertebrates studied.

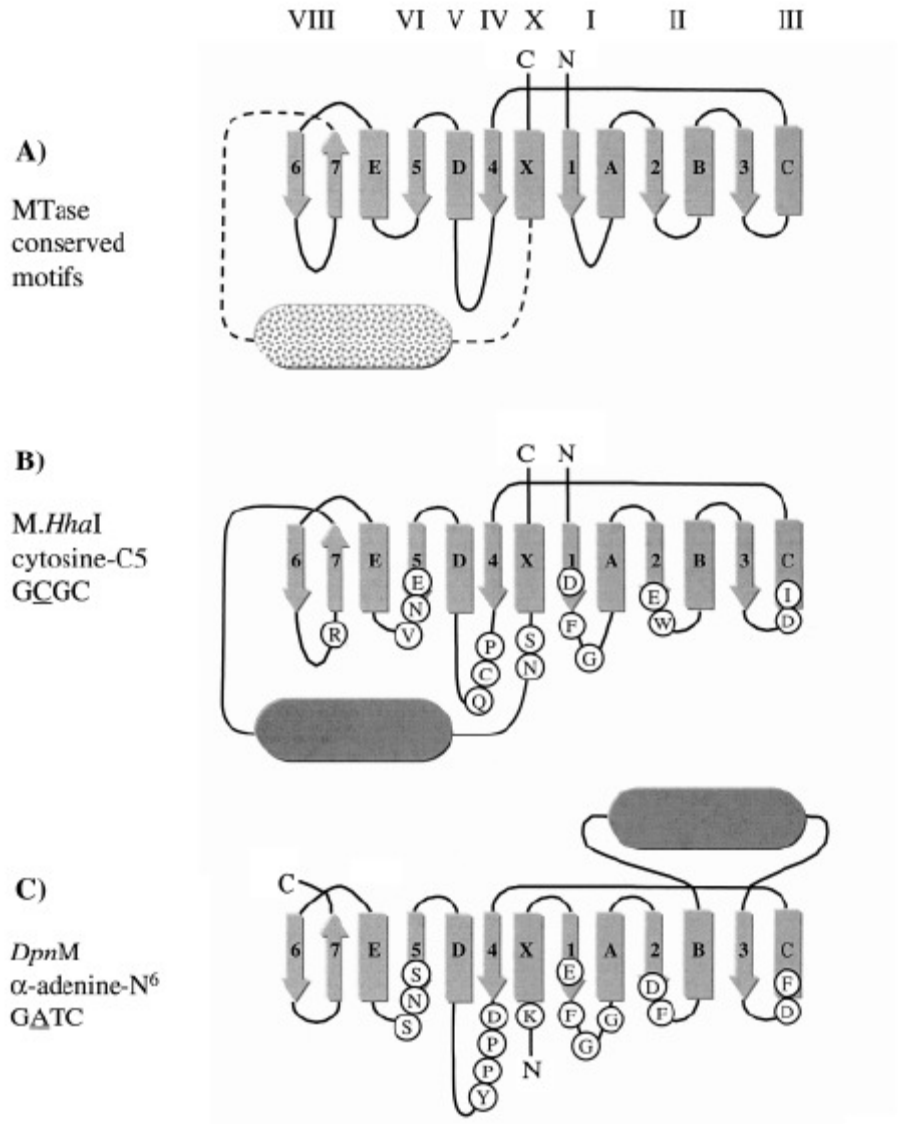


Figure 1.1: Comparison of the topologies and location of conserved motifs in DNA MTases of different families. A) Locations of the conserved motifs displayed on the framework of the structure of the large domain of cytosine-C5 MTases. B) Topology and locations of important amino acid residues in cytosine-C5 MTases. C) Topology and locations of important amino acid residues in α-N MTases. (Jeltsch, 2002)

The N-modified bases are present in prokaryotes (N^4-methylcytosine and N^6-methyladenine) and some lower eukaryotes (only N^6-methyladenine) (Jeltsch and Gumport, 2004). It has been reported that N^6-methyladenine is present in DNA from

several unicellular eukaryotes, including members of the genera *Chlamydomonas*, *Chlorella*, *Oxytricha*, *Paramecium*, and *Tetrahymena* (Ratel *et al*., 2006).

1.3 Prokaryotic DNA methyltransferases

Most prokaryotic DNA MTases are parts of restriction/modification (RM) systems, which are widely distributed in the bacterial and archeal kingdoms (Pingoud *et al*., 2005; Jeltsch, 2002). More than 1000 MTases of different specificity have been found in bacteria (Roberts *et al.,* 2000). Bacterial DNA methyltransferases are generally small in size (300-500 amino acids) (Jeltsch, 2002; Cheng, 1995). They usually methylate the DNA specifically within or next to short recognition sequences, which are often palindromic. These systems comprise two enzymes, a restriction endonuclease that specifically recognizes and cleaves DNA within short, and often plaindromic sequences. The cellular DNA is protected from cleavage by a corresponding MTase, because it modifies the DNA within the same sequence and prevents endonuclease action. The methylation pattern basically imprints a bar code on the DNA that allows the bacteria to distinguish between foreign and its own DNA. This mechanism is employed to protect the bacteria against phage infection, as well as the uptake of DNA from the environment. These systems thereby constitute a barrier to horizontal gene transfer that might have an important role in bacterial speciation (Jeltsch, 2003).

A second group of prokaryotic MTases is not accompanied by a restriction enzyme called "solitary DNA MTases" or "orphan DNA MTases" as exemplified by the *E. coli* DNA adenine methyltransferase (EcoDam) and *Caulobacter crescentus* CcrM (cell cycle-regulated DNA MTase). EcoDam recognizes GATC palindromic sequence and methylates adenine at N^6-position. It belongs to the α-class of exocyclic methyltransferases. On the other hand, CcrM is classified in the β group and is widely present in the α-subdivision of proteobacteria. It binds to and methylates adenosine in the sequence 5'-GANTC-3' where 'N' is any nucleotide (Collier *et al*., 2007; Wright *et al*., 1997). CcrM regulates cell cycle events in *Caulobacter crescentus* (Marczynski *et al*., 2002). EcoDam and CcrM are of independent evolutionary origin. It is speculated that

each might have evolved from an ancestral restriction-modification system that has lost its restriction component, leaving an 'orphan' methylase devoted solely to epigenetic genome modification (Marinus *et al*., 2009).

E. coli DNA methyltransferase (EcoDam) is a single polypeptide of 32 kDa (Herman and Modrich, 1982) that is encoded by the dam gene located at 74 min on the genetic map (Marinus, 1973; Bachmann, 1990). The natural substrate for EcoDam is hemimethylated DNA, the configuration of DNA immediately behind the replication fork. EcoDam binds to the template and slides processively along the DNA, methylating about 55 GATC sites per binding event (Urig *et al*., 2002). EcoDam competes with two other proteins, MutH and SeqA, for hemimethylated GATC substrate sites. These two proteins act before Dam to participate in the removal of replication errors (MutH) and to form the compacted and properly supercoiled chromosome structure for the nucleoid (SeqA). Increasing the cellular level of EcoDam causes a decrease in the amount of hemimethylated DNA, and prevents these two proteins from carrying out their functions, leading to an increased mutation rate and a change in supercoiling of the chromosome, respectively (Herman and Modrich, 1981; Marinus *et al*., 1984; Lobner-Olesen, *et al*., 2003).

There are about 130 molecules of Dam per *E. coli* cell, and this level is optimal to allow a period of time between synthesis of the extending nucleotide chains and methylation of the GATC sequences within them (Boye *et al*., 1992). The actual time between synthesis and methylation can be rapid for plasmid molecules (2-4 s) (Stancheva *et al*., 1999) or about 1 min for chromosomal DNA in slow-growing cells with a doubling time of about 100 min (Campbell and Kleckner, 1988). Increase or decrease in the number of Dam molecules can profoundly alter the physiological properties of the cell. The cellular level of EcoDam is regulated mainly at the transcriptional level. Five distinct promoters have been identified contributing expression of dam gene transcript. The major dam promoter (P2) is located 3kb upstream of the gene and is regulated by the growth rate (Lobner-Olesen, *et al*., 1992).

Although EcoDam is a highly processive enzyme, it may become less processive at GATC sites flanked by specific DNA sequences (Peterson and Reich, 2006). Reduced

processivity may allow competition between EcoDam and specific DNA-binding proteins, thus permitting the formation of nonmethylated GATCs (Ringquist and Smith, 1992; Wang and Church, 1992; Hale *et al*., 1994; Tavazoie and Church, 1998). The number of unmethylated sites in the chromosome varies depending on the growth phase and the growth rate. The unmethylated dam sites appear to be mostly (Ringquist and Smith, 1992) or completely (Palmer and Marinus, 1994) modified in strains overproducing Dam, suggesting that the enzyme competes with other DNA-binding proteins at these specific sites. In addition, some GATC sites in DNA structures (non-B-form DNA such as H-DNA) (Parniewski *et al*., 1990) are relatively resistant to methylation at the normal cellular level of enzyme. Palindromic structures containing GATCs are also relatively resistant to Dam methylation (Allers and Leach, 1995).

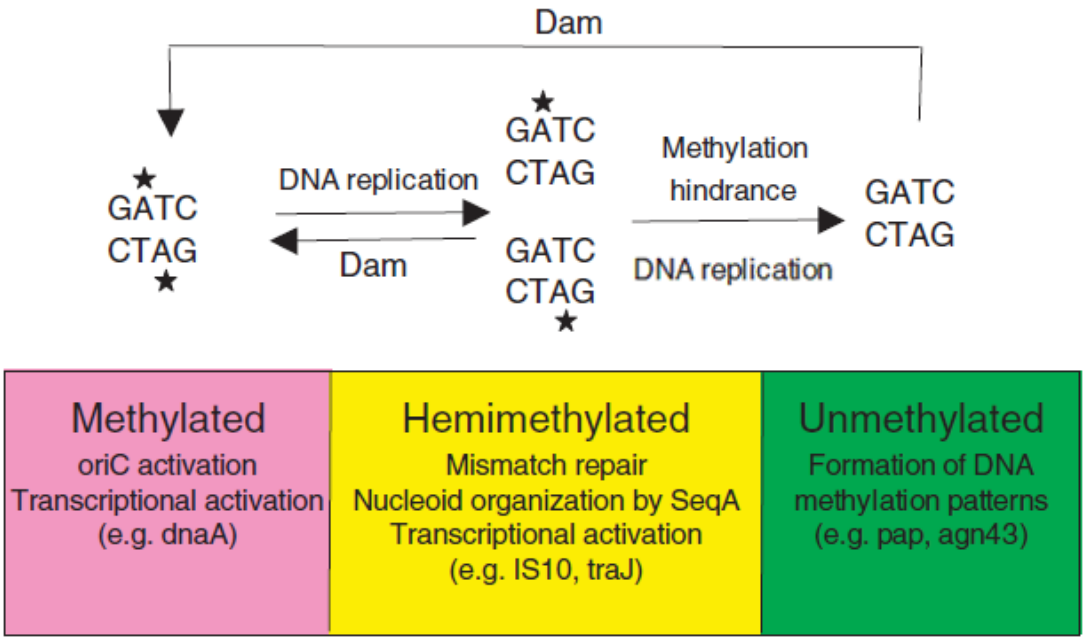


Figure 1.2: State of GATC site methylation in gamma proteobacteria. GATC sites can be fully methylated, hemimethylated or unmethylated depending on the growth phase and growth rate of bacteria (Marinus *et al*., 2009).

1.4 Role of DNA methylation in prokaryotes

1.4.1 Initiation of chromosome replication

Persistent hemimethylated sites have been detected at the origin of chromosome replication, *oriC*, and the region surrounding it (Campbell and Kleckner, 1990). This region includes the *dnaA* gene, which is located 43kb from *oriC*. DnaA initiates chromosome replication by binding to *oriC*, facilitating duplex opening to load DnaB helicase and DNA polymerase III holoenzyme.

Persistence of the hemimethylated state is due to the high density of GATC sequences in *oriC* (11 in 245bp) and the promoter region of *dnaA* (8 in 219bp), providing multiple binding sites for the SeqA protein. The SeqA induced hemimethylated state in this region of chromosome lasts about one third of the cell cycle (sequestration), preventing reinitiation from *oriC* from occurring more than once per cell cycle. *oriC* and *dnaA* promoter region must be fully methylated for efficient initiation of DNA replication. After the initiation event, other mechanisms ensure that DnaA is not in proper conformation for initiation. Among these mechanisms is a reduction in the transcription of the *dnaA* gene. Thus EcoDam coordinates the DNA replication to cell cycle and ensures that initiation occurs only once per cell cycle (Braun *et al*., 1985; Yamaki *et al*., 1988). In *Vibrio cholerae*, both Dam methylation and SeqA are essential (Julio *et al*., 2001; Saint-Dic *et al*., 2008), and SeqA overproduction causes DNA replication arrest (Saint-Dic *et al*., 2008).

1.4.2 Dam directed mismatch repair

During the process of DNA replication, errors may arise in the newly synthesized DNA strand and have to be removed from the newly synthesized strand and not from the parental strand. In *E. coli* and *S. enterica*, this discrimination is achieved by virtue of the hemimethylated state of DNA behind the replication fork: the newly synthesized DNA is not methylated, but the parental strand is methylated (Pukkila *et al*., 1983). The base

mismatch formed by a replication error (*e.g.* G-T) is recognized and bound by the MutS protein, which recruits the MutL protein (a molecular matchmaker) to form a ternary complex with MutH. The latent endonuclease activity of MutH is unmasked by the complex, and the enzyme cleaves the unmethylated strand 5' to the G at a nearby GATC site. MutH is then displaced from the complex by the UvrD helicase. UvrD unwinds DNA and the exposed single strand is degraded by exonucleases until the mismatch is removed. The resultant gap is filled in by the DNA polymerase III holoenzyme, and the nick is sealed by DNA ligase. Finally, the hemimethylated GATC is symmetrically methylated by Dam (Iyer *et al*., 2006). MutH is active on hemimethylated, but not on fully methylated DNA, thus mismatch repair action is confined to the hemimethylated region behind the replication fork. Both, the lack of Dam methylation and the overproduction of Dam lead to an increase in the spontaneous mutation frequency (Marinus and Morris, 1974; Herman and Modrich, 1981). Overproduction of Dam leads to premature methylation of newly synthesized DNA, thereby preventing MutH action if a mismatch is present. In turn, lack of Dam results in the loss of strand discrimination, leading to the use of parental strand as a template for mismatch repair with a 50% probability. Single- and double-stranded breaks have been detected in the chromosome of *dam* mutants as a consequence of mismatch repair (Marinus and Morris, 1974; Wang and Smith, 1986). Homologous recombination is required to repair the double-stranded breaks, explaining the fact why mutations inactivating homologous recombination are synthetically lethal in *dam* mutant background (Marinus, 2000).

1.4.3 Regulation of gene expression

The state of GATC sites (methylated, unmethylated, and hemimethylated) can affect specific binding of DNA-interacting proteins. The presence of this tetranucleotide in the promoter or the regulatory sequences can affect gene expression by regulating binding of RNA polymerase or transcriptional regulators. There is evidence that specific protein binding maintains about 36 unmethylated GATCs in the *E. coli* chromosome (Casadesus and Low, 2006). Nine such GATCs are in the cyclic AMP-binding protein (CAP)-binding sites preceding the *mtlA, cdd, flhD, gcd, ycdZ, yffE, ppiA,* and *proP* operons (Wang and

Church, 1992), suggesting that gene expression might be modulated by Dam methylation through differential CAP binding.

Studies on the pap operon have provided the most detailed evidence that unmethylated GATCs are involved in transcriptional control (Casadesus and Low, 2006). Pyelonephritis-associated pilus (Pap) expression is regulated by a phase variation mechanism in which individual cells either express pili (phase on) or not (Phase off). *Pap* Pilus gene expression is in phase-on state when GATC1028 is unmethylated and GATC1130 is methylated. Conversely, in the phase-off state, the methylation state at these two sites is reversed. The mechanism of phase variation involves competition between Dam and the transcriptional activators Lrp and PapI. Lrp is required for methylation protection of GATC1130, and both Lrp and PapI are required for methylation protection of GATC1028 (Casadesus and Low, 2006).
Several *E. coli* promoters have GATC sites in their -10 or -35 region. These include promoter regions for the *sulA*, *trpS*, *trpR*, *tyrR*, and *glnS* genes, and expression of these genes is increased in *dam* mutants (Plumbridge, 1987; Barras and Marinus, 1989; Marinus, 1996).

1.4.4 Bacterial virulence

A role for EcoDam as a virulence factor has been observed for a growing list of bacterial pathogens (Mahan *et al*., 2000; Mahan and Low 2001; Heusipp *et al*., 2007). Although, Dam methylation is not essential for viability of *E.coli* (Bale, 1979), *dam* is an essential gene in *Vibrio cholerae*, *Yersinia pseudotuberculosis,* at least under tested growth conditions (Julio, 2001). However, in both *Yersinia pseudotuberculosis* and *Vibrio cholerae*, virulence attenuation is observed if Dam methylase is overproduced (Chen *et al.*, 2003; Julio *et al*., 2002). In *Salmonella enterica,* and *Haemophilus influenzae,* lack of Dam methylation causes attenuation of virulence in model animals (Taylor *et al*., 2005 and Watson *et al*., 2004). However, involvement of Dam methylation in bacterial virulence is not universal; for instance, Dam^- mutants of *Shigella flexneri* are not attenuated (Honma *et al*., 2004).

1.5 Crystal structure of *E. coli* DNA methyltransferase (EcoDam)

The crystal structure of *E. coli* DNA methyltransferase (EcoDam) has been solved at 1.89 Å resolution in a ternary complex containing EcoDam, AdoHcy, and a 12-mer oligodeoxynucleotide duplex containing a single, centrally located GATC target site (Horton *et al.,* 2006). EcoDam contains two domains: a seven-stranded catalytic domain (residues 1-56 and 145-270) harboring the binding site for AdoHcy and a DNA binding domain (residues 57-144) consisting of a five-helix bundle and a β-hairpin loop (residues 118-139) that is conserved in the family of GATC-related MTase orthologs (Yang *et al*., 2003) (**Figure 1.3**). Three regions were found to be disordered: residues 188-197 immediately after the active-site D181-P-P-Y184 motif (after strand β4), residues 247-259 between strands β6 and β7, and residues 271-278 at the C-terminus.

The biochemical and structural data demonstrate that in the absence of AdoMet in EcoDam the flipped target adenine does not enter the active site pocket but binds to an alternative site on the surface of the enzyme, which might resemble an intermediate in the base-flipping pathway. The orphan thymine can adopt an intrahelical or an extrahelical position, which illustrates the structural flexibility of the DNA in complex with the MTase (Horton *et al.,* 2006).

1.6 DNA recognition by *E. coli* DNA methyltransferase (EcoDam) enzyme

EcoDam recognizes GATC palindromic sequence and methylates adenine at N-6 position. The methylation target, the adenine of the second base-pair in GATC (Ade2), flips out from the DNA helix. The specific interactions with the remaining bases of the site occur in the DNA major groove. The EcoDam molecule binds as a monomer in solution and spans ten base-pairs, four base-pairs on the 5' side and five base-pairs on the 3' side of the flipped-out target adenine.

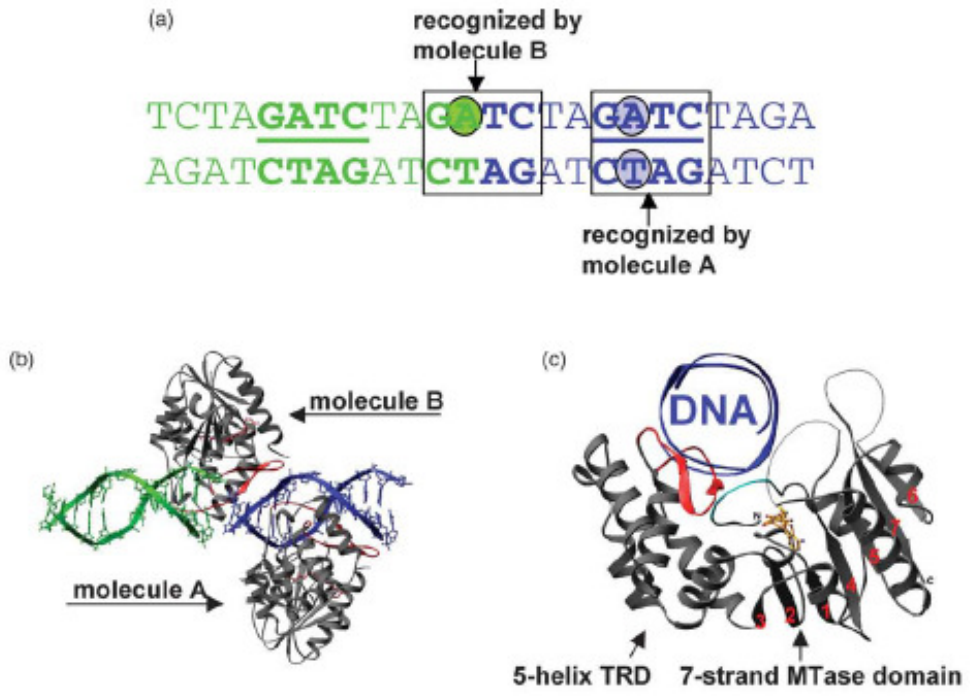


Figure 1.3: Structure of the EcoDam-AdoHcy-12mer DNA complex. (a) Two DNA duplexes (green and blue) are stacked head-to-end, with one GATC site in the middle of each duplex and one in the joint of two duplexes. The nucleotides in extrahelical positions are in shaded circles. (b) Molecule A binds to the GATC site in the middle of each DNA duplex, while EcoDam molecule B binds to the joint of two DNA duplexes. (c) EcoDam contains two domains: a seven-stranded catalytic domain that harbors the binding for AdoHcy (in stick model) and a DNA-binding domain consisting of a five-helix bundle and a β-hairpin (red) that is conserved in the family of GATC-related MTase orthologs (Horton *et al.,* 2006).

The amino acid residues from the β-hairpin make the majority of base specific interactions, but K9 from the N-terminal loop also forms a base contact. The recognition of the first base-pair is an interesting deviation between T4Dam and EcoDam. In the T4Dam structure (Horton *et al.,* 2005), the first guanine of the GATC site is contacted by R130 with bifurcated hydrogen bonds to the N7 and O6 atom of Gua1. R130 is located at the end of the β-hairpin, but it is not conserved among the Dam-related MTases. The EcoDam structure shows that Gua1 interacts, *via* the N7 and O6 atoms respectively, with two side chains, K9 and Y138 (at the position corresponding to EcoDam K9, T4Dam has an alanine (Ala).

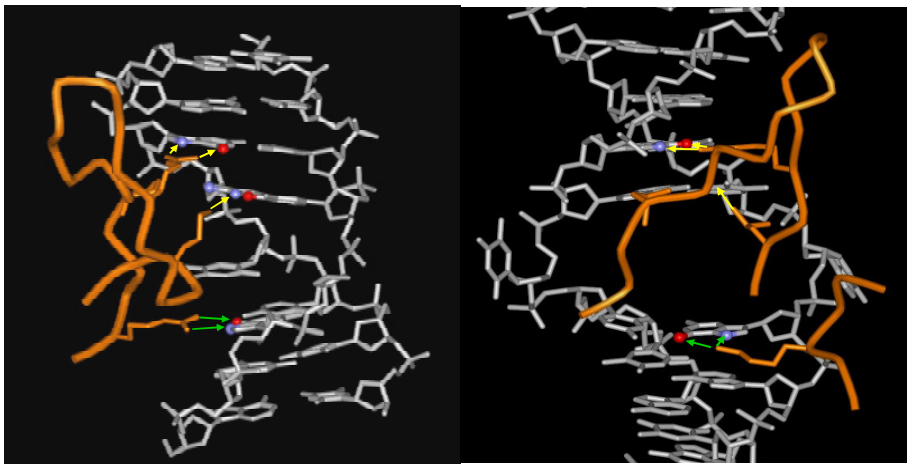


Figure 1.4: Comparison of the DNA recognition by T4Dam (left) and EcoDam (right). The figure shows the conserved ß-hairpin, which interact via R116 and R124 with C4 and via and M114/P126 and L122/P134 with T3 in T4Dam and EcoDam, respectively (yellow arrows). The recognition of G1 is mediated by Arg130 (from the β-hairpin) in case of T4Dam, but by K9 in EcoDam (green arrows) (Horton *et al.,* 2006).

The third base-pair of GATC makes van der Waals contacts with two hydrophobic side-chains of L122 and P134. P134 has been shown to be central for discrimination of the TA base pair (Horton *et al.,* 2005). The guanine in the fourth base-pair of GATC interacts *via* its O6 and N7 atoms with the guanidino group of R124 in a bifurcated hydrogen bonding pattern. Mutagenesis studies with EcoDam have shown that exchange of any residue involved in specific interactions with the GATC site leads to major changes in DNA recognition specificity of EcoDam (Horton *et al.,* 2005, Horton *et al.,* 2006).

1.7 DNA methyltransferases (MTases) and protein engineering

DNA MTases are particularly prone to in vitro evolution approaches as these enzymes modify DNA at specific sites. Thus, the blueprint of enzyme activity and specificity can be detected on the DNA coding for the protein, giving a unique coupling of genotype (DNA sequence) and phenotype (enzymatic properties) on the individual DNA molecule. This facilitates directed evolution experiments enormously, because screening can be performed with a library of DNA molecules coding for different MTases variants and, consequently, carrying different methylation patterns.

1.7.1 Protein engineering

Chemical reactions in living organisms are catalyzed by enzymes, the vast majority of which are proteins. Enzyme's efficiency and precision are unmatched by most conventional industrial catalysts: they allow reactions to occur billions of times faster than would be possible without them. However, naturally occurring enzymes often lack features necessary for commercial applications: enzymes have evolved to accelerate only biological reactions, under the narrow set of conditions that are compatible with life. However, enzymes could be engineered and evolved in a targeted manner for their successful broad range applications.

Evolution is a process in which the selection for specific traits is accomplished by applying environmental pressure. In nature, genetic diversity is obtained by spontaneous mutations that occur during DNA replication or by recombination events. In a pool of genetically diverse organisms the one most suited for survival (best suited for carrying out all relevant reactions at the specific habitat) is selected for, i.e. has a larger probability for passing its genetic material to the next generation. In recent years, it has become clear that the evolutionary process can be mimicked at the lab bench.

Protein engineering describes the process of altering the structure of an existing protein to improve its properties (Lutz *et al*., 2009). Protein engineering involves three steps: choosing the right method of engineering (engineering strategies, such as rational design or randomization), making those changes (mutagenesis) and evaluating the protein variants for improved properties (screening or selection). There are two general strategies for protein engineering: rational design (Bornscheuer and Pohl, 2001; Bornscheuer, 2002) and directed evolution (Farinas *et al*., 2000).

1.7.2 Rational protein design

Initial efforts at enzyme engineering took a so-called rational-design approach. In rational design, precise changes in amino acid sequence are preconceived based on a detailed

knowledge of protein structure, function and mechanism, and are then introduced using site directed mutagenesis (Chen, 1999). This technology holds strong promise for optimizing the desired properties for commercial applications. It also greatly enhances our basic understanding of enzyme binding and catalytic mechanisms, thus increasing the success of future enzyme engineering efforts and laying the foundation for functional prediction of new protein sequences in databases (Chen, 2001). Success was not only measured in terms of increased specificity, but also of stability, the ability to fold, and catalytic activity. However, many of the early attempts were disappointing, as these interdependent properties are hard to predict, further enhanced by incomplete understanding of the underlying mechanisms required to improve the desired enzyme properties. The power of rational design has been demonstrated by the generation of a faster superoxide dismutase, already one of the fastest known enzymes in nature (Getzoff, *et al*., 1992) and complete inversion of coenzyme specificities for both isocitrate and isopropylmalate dehydrogenases (Chen *et al*., 1996; Chen *et al*., 1999).

1.7.3 Directed enzyme evolution

Directed evolution mimics Darwinian evolution in the test tube, and involves the generation and selection of a molecular library with sufficient diversity for the altered function to be represented (Arnold, 1998; Zhao, 1997). Directed evolution differs from natural evolution in two key aspects: (i) natural evolution occurs under multiple and variable selection pressures, whereas directed evolution is accomplished under controlled selection pressure for predetermined functions; (ii) in directed evolution, 'non-natural' functions, of practical use, can be obtained through the design of appropriate selection schemes, whereas natural evolution favors functions advantageous to the survival of the organism (Schmidt-Dannert, 2001; Williams *et al*., 2004).

Directed evolution does not require information about how enzyme structure relates to its function (Stemmer, 1994; Kuchner *et al*., 1997). This technique employs a random process like error-prone PCR which is used to create a library of mutagenized genes. Genetic selection or high-throughput screening subsequently identifies the mutants that

possess improved properties. The sorted genes might be subjected to further cycles of mutagenesis and screening to enhance the original beneficial mutations (**Figure 1.5**). Recently, directed evolution has been significantly improved using *in vitro* recombination or DNA shuffling (Stemmer, 1994). These methods rapidly combine beneficial mutations that arise from random mutagenesis and significantly expand sequence diversity derived from small pool of homologous genes.

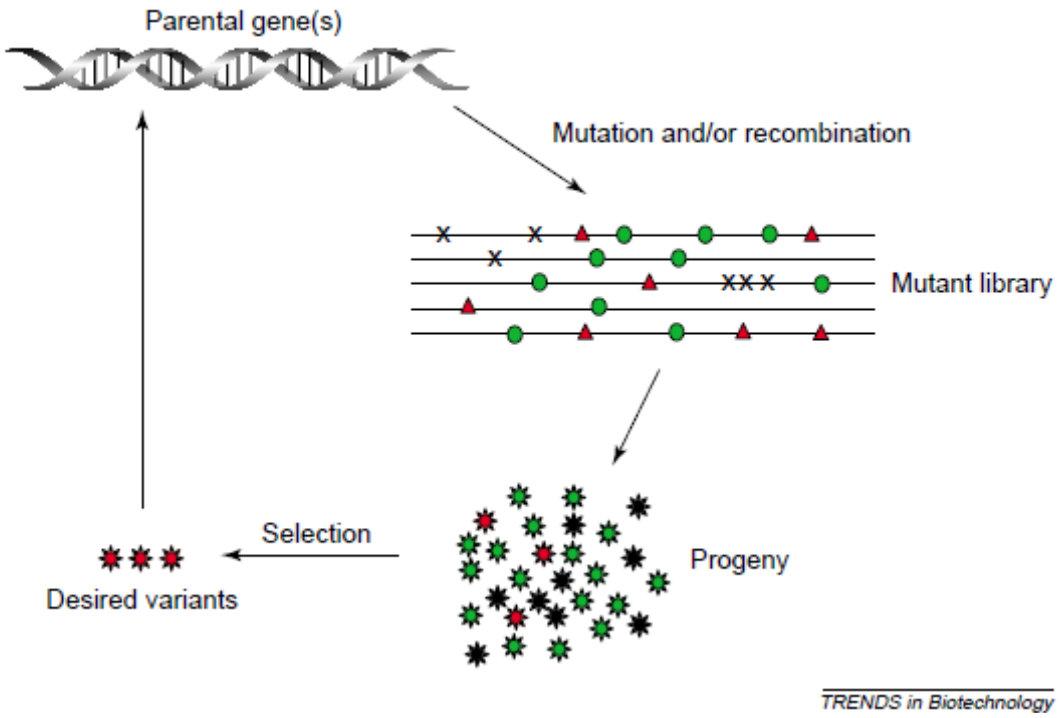


Figure 1.5: A typical scheme for directed evolution. Individual variants in a mutant library (right) differ from the parental genes by containing mutations that are functionally neutral (green circle), deleterious (X) or beneficial (red triangle). Progeny expressing these mutant genes are screened for the function of interest. The gene variants isolated from the improved progeny (red) can be used as parents for the subsequent cycle of evolution (Chatterjee, 2006).

Rational design and directed evolution both have their distinct advantages and yet the technologies are complementary. Rational design has been used to introduce key residues or structural elements that are not usually attainable via a random process. Random mutagenesis could then be employed to generate subtle changes that would fine- tune

protein packing and function. Conversely, random mutagenesis can also provide useful information by screening improved variants and finding out substitutions (hot-spots) where a more precise 'rational' strategy would be implemented. Although, either rational design or directed evolution can be very effective, a combination of both strategies will probably represent the most successful route for improving the properties and function of an enzyme (Altamirano *et al*., 2000; Kuchner *et al*., 1997; Chen *et al*., 2001).

1.7.4 Applications and achievements of protein engineering

As mentioned above, enzymes from nature rarely have the combined properties necessary for industrial chemical production such as high activity and selectivity on non-natural substrates and toleration of high concentrations of organic media over a wide range of conditions (like decreasing substrate and increasing product concentrations) that will be present during the course of a manufacturing process. A continuously expanding applications of enzymes for the chemical, pharmaceutical and food industries create a growing demand for enzymes that exhibit higher operational stability, higher specificity and enantio selectivity, as well as for those that have new activities on natural and unnatural substrates. With the advances in protein engineering technologies, a variety of enzyme properties can be altered simultaneously, if the appropriate method of mutagenesis, selection and screening are employed (Luetz *et al*., 2008). The success of industrial applications of enzymes has not been limited to hydrolytic enzymes (such as lipases, esterases, acylases), but extending the scope to enzymes that perform a wide range of transformations, including asymmetric reduction, oxidation, and carbon-carbon bond formation (Schoemaker *et al*., 2003; Schmid *et al*., 2001).

2 Research objectives of the thesis

DNA methylation represents an important epigenetic modification in the *E. coli* regulating several important biological processes. The present PhD work in the thesis was focused on the functional characterization and molecular evolution of *E. coli* DNA methyltransferase (EcoDam) enzyme.

2.1 Directed evolution of *E. coli* DNA methyltransferase (EcoDam) enzyme

DNA methyltransferases (MTases) allow powerful combinations of *in vivo* and *in vitro* mutagenesis and DNA-based screening for enzymatic properties, which makes this approach one of the most powerful applications of directed *in vitro* evolution of proteins. In order to better understand the mechanism of DNA recognition by proteins and enzymes, which underlies gene regulation in all organisms, we aimed to change the DNA recognition specificity of DNA MTases in different approaches. We aimed to set up a system of mutual co-evolution of DNA MTase and target site that broaden our understanding of the process of molecular evolution in great detail. In the present study, we have successfully changed the DNA recognition specificity of EcoDam from GATC to GATT by combining different approaches of mutagenesis with restriction protection at GATT site for selection and screening. Selected mutants showed up to 1500 fold change in specificity *in vitro* and methylated the new GATT target site with rates comparable to the rate of GATC methylation by the wildtype enzyme.

2.2 Kinetic investigation of the mechanism of action of inhibitor-58

A role of the Dam enzyme as a virulence factor has been observed for a growing list of bacterial pathogens (Low *et al.,* 2001). However, DNA-adenine methylation is not present in higher eukaryotes including humans. These observations raise the possibility that dam inhibitors may be useful as anti-microbial agents. In the present study, we investigated the kinetic mechanism of action of one new inhibitor (inhibitor-58) from a

subset of compounds that address the various binding pockets of the enzyme. The results from kinetic analysis clearly showed that inhibitor 58 possess competitive mode of inhibition for AdoMet, while uncompetitive for DNA.

2.3 Role of EcoDam in gene regulation

GATC sites are the natural substrate for EcoDam to methylate adenine at N6 position. The methylation status of GATC sites (methylated, unmethylated, and hemimethylated) can affect specific binding of DNA-interacting proteins. In the present study, we investigated the gene expression profile of dam$^-$ *E. coli* strain (JM110) in the presence and absence of EcoDam. Herein we report that EcoDam expression can affect the expression of certain genes and can serve as a transcriptional regulator for gene expression. The present study also revealed that EcoDam can stimulate expression by binding to TANAC sites which has not been observed before.

2.4 Role of EcoDam in bridging metabolic signals and epigenetic modifications

Bacterial cells encounter varied environmental changes and make appropriate adjustments to ensure their survival. Herein we report that the *E. coli* methyltransferase (EcoDam) enzyme binds to polyphosphates in cells and meditates potential connections between metabolic signaling and epigenetic modifications.

2.5 Supporting co-author work

Sanjay Chahar (S. C.) was involved in the study of two alternative conformations of *S*-Adenosyl-L-homocysteine bound to *Escherichia coli* DNA adenine methyltransferase (EcoDam) and the implication of conformational changes in regulating the catalytic cycle. Results of the study are described in the appendix chapter 2, and the author contribution to the study is mentioned in the appendix chapter 4.

S. C. was also involved in the study of transition from EcoDam to T4Dam DNA recognition mechanism without loss of activity and specificity. Results of the study are described in the appendix chapter 3, and the author contribution to the study is mentioned in the appendix chapter 4.

3 Results and discussion

3.1 Directed evolution of *E. coli* DNA methyltransferase (EcoDam) enzyme

In this section the directed evolution approaches employed to change the DNA recognition specificity of *E. coli* DNA methyltransferase (EcoDam) enzyme will be discussed and the entire details of the whole approach is described in the publication of these results (Chahars, S., *et al.,* 2009), which is attached as Appendix chapter 1 to this thesis.

In our attempt to change the DNA recognition specificity of EcoDam, we first started our experiments by site saturation mutagenesis at the two residues (R124 and P134) that are directly involved in the recognition of the TC part of the recognition sequence (GATC). After one round of selection, the R124S/P124S double mutant showed a new target recognition specificity (GATT) with low activity. P134S substitution still kept the protein to maintain its wildtype property of recognition of TC base pair at the third position of recognition sequence (GATC), while R124S substitution helped the protein to recognize altered base pair at fourth position (TA) and leaded the protein to achieve a new target specificity (GATT). As suggested by Arnold and co-workers (Bloom *et al*., 2006), a step wise accumulation of single amino acid substitutions favors the protein to achieve the desired goal. Therefore, the R124S/P134S mutant was used as starting template to introduce further mutations. After full gene randomization and selection, 4 clones with improved methylation at GATT site were identified which contained 18 different mutations in addition to the original R124S/P134S mutation. This result demonstrates that complex protein properties can be improved by gradual accumulation of beneficial mutations. Protein variants containing up to 9 additional mutations (II-2) were found and some variants were found to contain similar additional mutations (II-3 and II-5, II-4 and II-6). To evaluate the combinatorial effects of these beneficial mutations, 4 selected mutants (II-1, II-2, II-3 and II-4) were *in vitro* recombined using DNA shuffling method

(Crameri *et al*., 1998). Interestingly, two clones (III-1 and III-2), with improved methylation at GATT site, containing shuffled mutations from previous round were identified, mimicking the natural evolutionary process by introducing these natural mechanisms into directed evolution.

We introduce a new method for directed evolution approach where enzyme libraries are combined with substrate libraries and selection for new variants is achieved by a unique coupling of genotype (DNA sequence) to phenotype (enzyme property) on individual DNA molecule. We have used BspEI restriction protection assay to select mutants with new target specificity. Identification of false positive clones (mutations within BspEI cleavage site) underscores strong selection system.

Evolutionary analysis of enzyme families suggests that drastic changes in enzyme function might require considerable changes in polypeptide backbones (Matsumura, 2000). Such changes will probably not occur during the current in vitro evolution process, in which enzymes are mainly improved by point mutation with a significant bias for transitions over transversions, thus limiting access to a broader spectrum of substitutions. In contrast, natural mutations often result from recombination and DNA repair events that generate deletions, insertions, duplications or fusions. Such mutations alter the spacing between amino acid residues and polypeptide chain segments and can result in larger changes in specificity and new catalytic activities. Certainly new methods of library generation will be needed in the field of protein design that could combine natural evolutionary pathways and current strategies of protein evolution, to make this approach even more powerful in future.

One additional limitation of directed evolution approach is the prerequisite for a sensitive and efficient method for screening a large number of potential mutants. Mutant library populations represent a pool from which the best mutants have to be retrieved on the basis of detected enzymatic activity. Isolated mutants might subsequently serve as an improved starting point for additional rounds of mutagenesis to accumulate beneficial mutations for the best results. Identification and isolation of best mutant(s) is a crucial

step in this process. In cases when the envisaged enzymatic activity or property is essential for viability and growth (e.g. overcoming increasing concentration of antibiotics or providing an essential nutrient) (Stemmer, 1994; Naki *et al*., 1998) selection might be applicable (Fastrez, 1997). Otherwise, screening based on the development of a detectable signal connected to the targeted catalytic activity needs to be done. This process may be directed either towards the identification and isolation of very few highly active mutants from an essentially inactive background population, or towards the quantitative measurement of the activity exhibited by each cell or colony for the identification of most active mutants out of a moderately active background population.

3.2 Kinetic investigation of the mechanism of action of inhibitor-58 on EcoDam

A role of the Dam enzyme, as a virulence factor has been observed for a growing list of bacterial pathogens (Low *et al.,* 2001). However, DNA-adenine methylation is not present in higher eukaryotes including humans. These observations raise the possibility that Dam inhibitors may be useful as anti-microbial agents. Inhibition of Dam by small molecule inhibitors may provide the basis for the development of a new class of antibiotics. Besides their prominent biological roles, Dam MTases are an ideal model system for structure-based drug design approaches. The strength of the Dam system is based on (i) the unique amount of structural information on non-specific, semi-specific, specific DNA complexes, (ii) a very good understanding of the dynamic processes taking place during DNA recognition and coenzyme binding, and (iii) the large number of specific biochemical assays already established that allow to direct study all the details of DNA and cofactor interactions with the enzyme.

3.2.1 Experiments and Results

Through virtual screening and molecular docking approaches, our collaborators (Prof. Dr. Cheng, Dr. Alex MacKerell) have selected 20 compounds that address the various binding pockets of enzyme, for biochemical investigations. Inhibition was measured in our lab and the mechanism of action of inhibition of inhibitor-58 (I-58) on EcoDam which appeared to be the most potent compound has been studied.

Steady state analysis, using hemimethylated 20mer DNA substrate and labeled AdoMet, has been performed to study the inhibition mechanism of inhibitor-58 (I-58). To this end, the concentration of AdoMet and inhibitor was varied between 0.74 and 5.5 µM (AdoMet) and 5-20 µM (I-58), respectively. DNA was held constant at 0.5 µM, and the enzyme was used at 50 nM concentration. Initial slopes were derived by linear regression of methylation kinetics comprising eight time points each (t=10 min). The data were fitted to various models describing the inhibition of an enzyme, fitting was only successful with a competitive model (Figure 1).

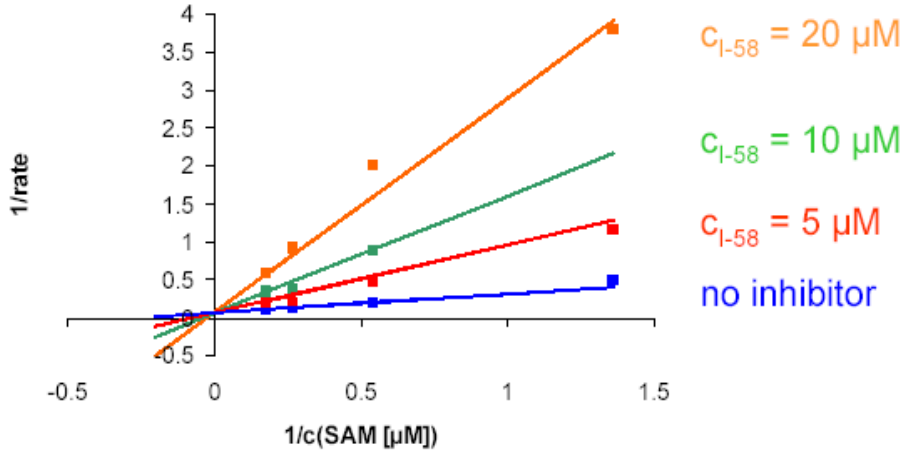


Figure 1: Double-reciprocal plot analysis of EcoDam in presence of inhibitor-58, with varying SAM (AdoMet) concentration. Orange, green, red and blue color lines indicate 20 µM, 10 µM and 5 µM inhibitor and no inhibitor conditions, respectively. All data points are based on complete 8 time point kinetics.

In a second series, DNA (0.2-2 µM) and inhibitor concentrations (5-20 µM) were varied. Enzyme was used at 50 nM, AdoMet at 1.8 µM concentration. Fitting was only possible to an un-competitive model (Figure 2).

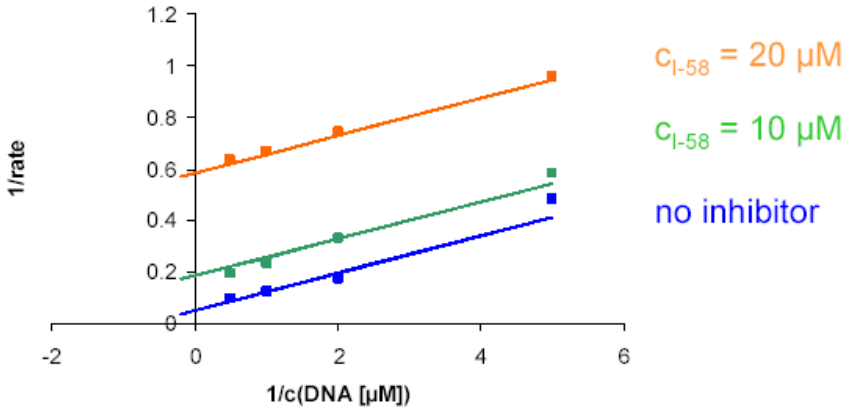


Figure 2: Double-reciprocal plot analysis of EcoDam in presence of inhibitor 58, with varying DNA concentration. Orange, green and blue color lines indicate 20 µM, 10 µM inhibitor and no inhibitor conditions, respectively. All data points are based on complete 8 time point kinetics.

The results obtained after kinetic analysis are Km(AdoMet)=3.9 (±0.2) µM, Ki(I-58)=2 (±0.1) µM. The binding affinity of the inhibitor to the enzyme is approximately twice that of AdoMet. The Km(DNA) was 1.1(±0.5) µM. The Kcat for the enzyme is 11 (±3) min^{-1}. The results showed that EcoDam first binds to DNA, and then to AdoMet. To confirm this kinetic mechanism, the inhibition effect of AdoHcy (S-adenosyl-homocysteine) on Dam enzyme has also been studied. AdoHcy is a structural analogue of AdoMet and should compete with AdoMet for its binding. Kinetic analysis clearly showed that SAH is competitive for AdoMet (Figure 3) and uncompetitive for DNA (Figure 4).

These data show that AdoHcy (SAH) and inhibitor-58 have similar mechanism of inhibition. AdoHcy is known to bind AdoMet binding pocket of the enzyme, therefore, it is likely that inhibitor-58 (I-58) as well binds to the SAM (AdoMet) binding site on the enzyme.

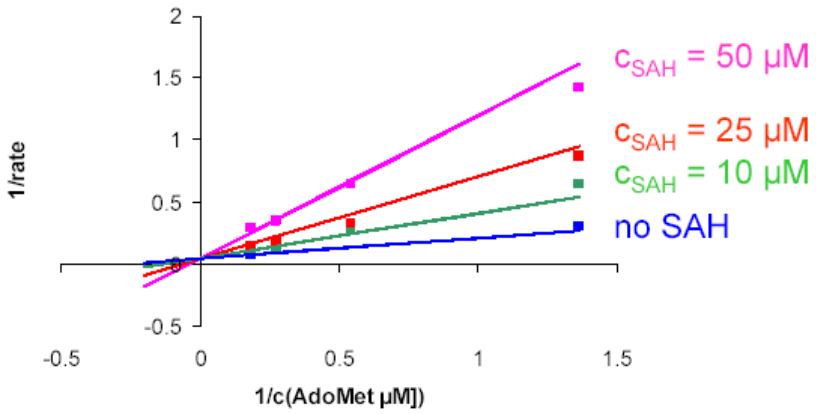


Figure 3: Double-reciprocal plot analysis of EcoDam in presence of inhibitor SAH, with varying SAM (AdoMet) concentration. Pink, red, green and blue color lines indicate 50 µM, 25 µM , 10 µM inhibitor (SAH) and no inhibitor conditions, respectively. All data points are based on complete 8 time point kinetics.

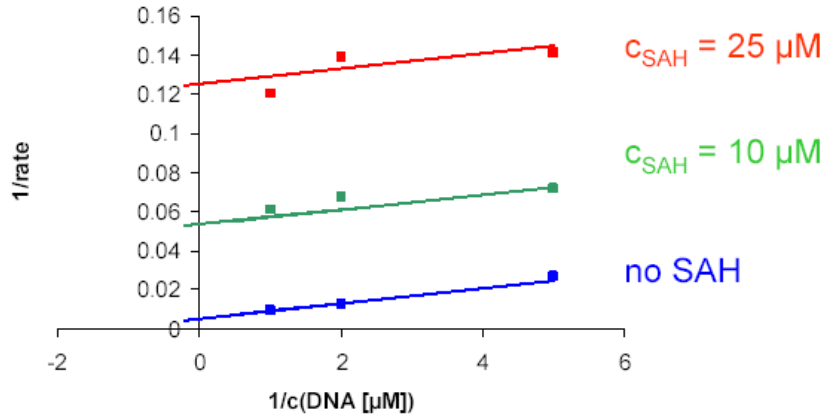


Figure 4: Double-reciprocal plot analysis of EcoDam in presence of inhibitor SAH, with varying DNA concentration. Red, green and blue color lines indicate 25µM, 10µM inhibitor (SAH) and no inhibitor conditions, respectively. All data points are based on complete 8 time point kinetics.

3.3 Role of EcoDam in gene transcription regulation in *E. coli*

Proteins specifically interacting with DNA often contact the DNA in the major groove (Garvie *et al.,* 2001). The methylation of an AT base pair at the adenine-N6 atom places the methyl group directly into the major groove of the DNA, where it might influence the interaction of DNA binding proteins with the DNA (Wion *et al.,* 2006). Therefore, adenine-N6 methylation has the capacity to influence the interaction of transcription factors with the DNA and, thereby, affecting gene expression. In addition, it has been reported that GATC sites are not randomly distributed, and often overlap with sequences recognized by global regulators (Henaut *et al.,* 1996; Wang and Church, 1992; Tavazoie and Church, 1998; Oshima *et al.,* 2002). This supports the notion that EcoDam might play a role in global gene regulation. A possible role of Dam methylation in the regulation of gene expression has been studied by several groups using EcoDam deletion *E. coli* strains with variable outcome results: Oshima *et al.,* (2002) reported that the expression of a large number of genes was affected by the dam deficiency. Genes involved in aerobic respiration, stress and SOS responses, amino acid metabolism and nucleotide metabolism were expressed at higher levels in the dam negative cells. Transcription of genes participating in anaerobic respiration, flagella biosynthesis, chemotaxis and motility was decreased in the dam mutant strain. The promoters of most of the Dam-controlled genes were also found to contain GATC sequences (Oshima *et al.,* 2002). Later, Løbner-Olesen *et al.,* (2003) reported that cells lacking Dam methyltransferase only showed a modest change in transcription of 17 genes. Up-regulation was observed for 8 genes mainly belonging to the SOS regulon (Løbner-Olesen, 2003). Robbins-Manke *et al.* (2005) found that >200 genes are expressed at a higher level in a dam negative strain, which was associated with a derepression of LexA-regulated SOS genes as well as the up-regulation of other non-SOS genes involved in DNA repair (Robbins-Manke, 2005). In addition, there is solid evidence that EcoDam is involved in phase variation and pathogenicity of γ-proteobacteria (reviews: Hernday, 2002; Hernday, 2004; Heusipp, 2007). For example, the expression of pyelonephritis-associated pili (Pap) in uropathogenic *E. coli* is epigenetically controlled by the

methylation state of the two GATC sites in the Pap regulon (Hernday, 2003), which in part determines the phase variation of pili formation.

However, the lists of genes affected in these different experiments were barely overlapping indicating a large variance in response. One reason for this large variance may lie in the usage of EcoDam negative *E. coli* strains, which might have acquired compensatory additional genetic and/or epigenetic changes during their cultivation in the absence of EcoDam. Another problem could be secondary effects caused by the absence of EcoDam. Because of its involvement in DNA repair and cell cycle control, many genes from these pathways are expected to react on the loss of EcoDam. In order to overcome the problems associated with established strains and secondary effects of the loss of EcoDam, we decided to study gene expression in EcoDam negative SCS110 cells short time after re-introducing EcoDam.

3.3.1 Experiments and Results

We compared the gene expression profile of dam$^-$ *E. coli* cells (SCS110) in presence of induced EcoDam versus control cells having only empty plasmid (pUC8). Cells were treated identically (growth phase, antibiotics, shaking etc), induced by addition of IPTG (1mM) and grown until mid log phase (OD_{600} 0.5). Then total RNA was extracted, converted into fluorescently labelled cDNA and hybridized to E. coli cDNA arrays, to study changes in gene expression. Two sets of experiments were conducted to allow for quality assessment.

In our data sets, we identified 192 genes in total that showed statistically significant expression differences, 86 of them showed >1.5 fold up- or downregulation in the presence of EcoDam. We then extracted information on promoters and operon structure from RegulonDB, which was available for about 50% of all genes. This also provided the sigma factors used for transcription of these genes. Then, the number of GATC and TANAC sites present in the whole data set and in the subset of up- or down-regulated genes was calculated and statistics determined. GATC sites were investigated because

they represent the target for enzymatic methylation, TANAC because previous experiments suggest a preferential binding of EcoDam to such sites (X. Cheng, personal communication).

		all genes	genes with promoter	genes with promoter information								
				s70			GATC			TANAC		
				obs	exp	P	obs	exp	p	obs	exp	p
EcoDam / control	all	192	89	67	69.5	0.3	21	16.1	0.112	56	48.4	0.065
	Dam up	99	50	42	39.0	0.202	8	9.0	0.44	37	27.2	0.004
	Dam down	93	39	25	30.4	0.033	13	7.0	0.016	19	21.2	0.289

Table 1: ***Summary of the analysis*** The first two columns give the number of differentially regulated genes, and the numbers of genes for which promoter information was available. For these genes, we then have counted how many of them are driven by sigma 70. This number is given under 'obs' (for observed). On the basis of the whole dataset the number of genes expected to be driven by these promoters was calculated and given as 'exp' (for expected) and p-values for over- or under-representation calculated ('p' only the relevant p-value is listed here). In the next triples of columns, the numbers of GATC and TANAC sites observed, expected and the relevant p-values are given. P-values are shaded by significance (yellow for $p<0.1$ and orange for $p<0.01$).

Sigma 70 genes are under-represented in EcoDam repressed genes and (statistically not significantly) enriched in EcoDam activated genes. This result suggests that high EcoDam activity is favoring expression of log-phase metabolic genes from sigma 70. In the down-regulated fraction, genes with GATC sites were highly overrepresented, suggesting that the methylation of GATC sites has a repressing effect on the expression of many genes. Among the upregulated genes, we found an enrichment of TANAC sequences suggesting that binding of EcoDam to such sites may stimulate gene expression, perhaps by competing with the binding of a repressor.

Table 2: Examples of differentially expressed genes in micro array experiment. All of the listed genes showed down regulation of expression in the presence of EcoDam.

Gene	Blattner number	Number of GATC sites	Relative expression change (wt/control)	p-value	Gene product
frdA	b4152	4	0.39	0.03	fumarate reductase, flavoprotein subunit
srlA	b2703	1	0.27	0.03	D-glucitol-specific enzyme II of phosphotransferase system
gcvT	b2904	1	0.32	0.01	Aminomethyl transferase,
narG	b1227	1	0.40	0.01	Nitrate reductase alpha-subunit

Table 3: Examples of differentially expressed genes in micro array experiment. All of the listed genes showed up regulation of expression in the presence of EcoDam.

Gene	Blattner number	Number of TANAC sites	Relative expression change (wt/control)	p-value	Gene product and function
atpA	b3734	2	1.4	0.04	Membrane-bound ATP synthase, F1 sector, alpha-subunit
atpI	b3739	3	1.45	0.04	Membrane-bound ATP synthase subunit, F1-F0-type proton-ATPase
rpsL	b3342	1	1.53	0.03	30S ribosomal subunit protein S12
rfaQ	b3632	1	1.48	0.003	Heptose region of LPS core

3.3.2 Real-Time PCR results

To confirm our DNA microarray results, quantitative expression of selected genes was analysed by Real-Time RT-PCR analysis. Real time PCR was performed for 8 individual genes that showed significant effect of GATC and TANAC sequences on gene expression during data analysis. When we analyze RT PCR data for EcoDam vs. Control, we observed 2 to 4 fold down-regulation of gene expression in 4 candidate genes (srlA, gcvT, frdA, narG) where GATC site(s) are present in promoter region. We also observed 1.5 to 2 fold up-regulation of gene expression for 4 candidate genes (atpA, atpI, rpsL, rfaQ) where TANAC site(s) are present in promoter region.

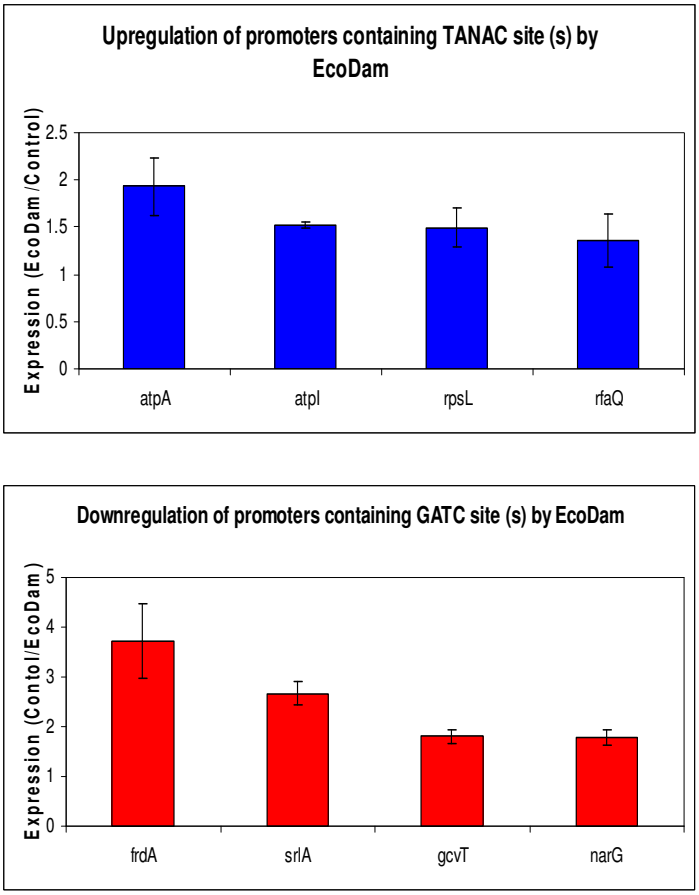


Figure 5: Real time RT PCR of 4 genes up-regulated in the presence of EcoDam (blue bars, all containing at least one TANAC site in their promoter region) and 4 genes down-regulated in the presence of EcoDam (red bars, all containing at least one GATC site in their promoter region).

3.3.3 Validation of our data by CFP reporter assay

Data obtained from DNA microarray analysis and RT-PCR analysis clearly demonstrated that EcoDam affects the expression of genes having GATC and TANAC sequences within their promoter region. To validate the effect of EcoDam on the expression by these promoters, we cloned each promoter region upstream of CFP reporter gene and studied how the expression is affected in presence of EcoDam. JM110 cells (Dam^-) were transformed with Promoter-CFP construct in presence and in absence of EcoDam and the CFP fluorescence of the E. coli cells measured in a Tecan-Safire fluorescence reader. Cell density was normalized by OD_{600} for each measurement. We observed the same effect of Dam methylation on gene expression as we observed in DNA micro array and Real-time PCR experiments.

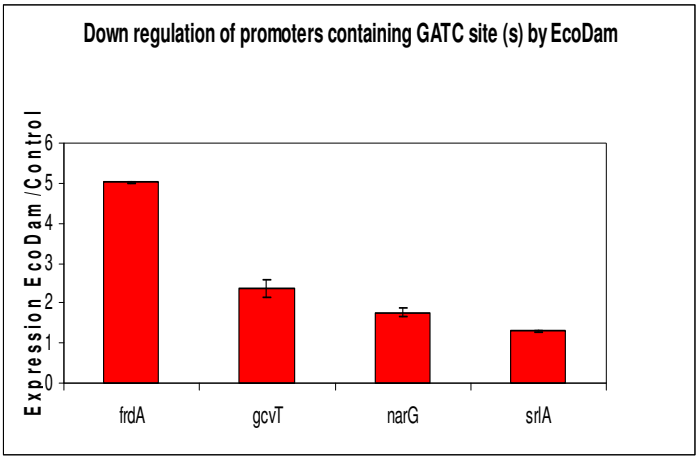


Figure 6: Down regulation of CFP expression by five promoters containing GATC site(s).

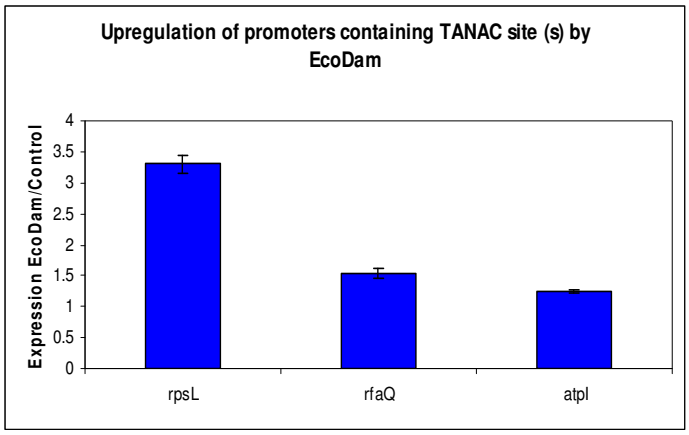


Figure 7: Up regulation of CFP expression by three promoters containing TANAC site(s).

To study the influence of the GATC or TANAC sequence on this effect in detail, we mutated these sites in some of the promoters and measured CFP fluorescence in presence of EcoDam. Mutational analysis was not possible with all studied promoters because of the presence of multiple GATC or TANAC sites present in the promoters. Interestingly, the GATC suppressive and TANAC simulative effects on expression in presence of EcoDam was abolished after mutating these sites. These results suggest a direct role of the GATC and TANAC sites on the transcriptional changes.

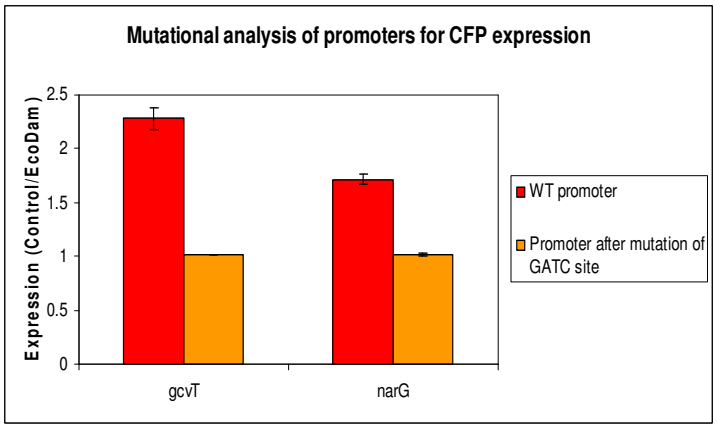


Figure 8: Comparative CFP expression of wild type promoters versus promoters after mutation of the GATC site. Wild type promoters showed up to 2.2 fold-down expression (gcvT), while GATC mutated promoters did not show fold-down expression in presence of EcoDam.

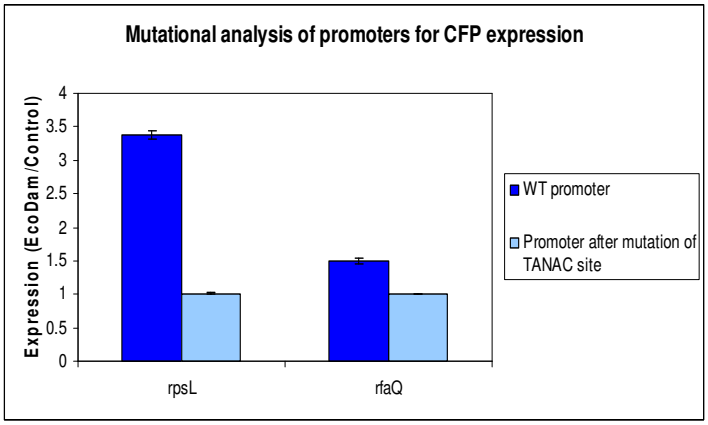


Figure 9: Comparative CFP expression of wild type promoters versus promoters after mutation of the TANAC site. Wild type promoters showed up to 3.4 fold-up expression (rpsL), while TANAC mutated promoters did not show fold-up expression in presence of EcoDam.

3.3.4 Discussion

Microorganisms such as *E. coli* live in environments subjected to rapid changes and their survival depends on their ability to regulate the expression of genes coding for the enzymes and transport protein required for growth in the altered environment. Dam deficient *E. coli* strains exhibit pleiotropic changes, including an increased mutational rate, uncoordinated DNA replication initiation, transcriptional alterations, hypersensitivity to DNA damage and a dependence on recombinational repair (Løbner-Olesen, *et al.,* 1994; Marinus, *et al.,* 1984; Palmer, *et al.,* 1994). Herein, we studied the global transcriptional changes in the dam negative *E. coli* strain SCS110 in the presence and in the absence of EcoDam. Our micro array, RT-PCR and CFP fluorescence data showed that, a subset of genes (srlA, gcvT, frdA, narG) having GATC sites within their promoter region (-400 bp to +1) showed down-regulation (>1.5) of gene expression in the presence of EcoDam. We also observed that, a subset of genes (atpA, atpI, rpsL, rfaQ) having TANAC site within their upstream region (-400 bp to +1) showed up-regulation (>1.5) of gene expression in the presence of EcoDam. It has been previously reported by others (Robbins-Manke *et al.,* 2005; Løbner-Olesen, *et al.,* 2003), that change in expression profile in dam deficient strains is not always correlated with the presence of GATC sites within their upstream promoter region, however they reported changes in

expression of certain genes in Dam methylation-dependent manner. Our finding also is in agreement with their finding that altered gene expression has not always been correlated with the presence of GATC sites within their upstream promoter region. Our data support the notion that EcoDam might be involved in regulation of certain genes either directly by blocking the recognition site of transcriptional regulators (through GATC site methylation) or indirectly. Here, we propose a mechanism of methylation independent indirect regulation of gene expression by EcoDam, and suggest that EcoDam binding to TANAC site might influence binding of other transcriptional regulators thereby affecting gene expression.

3.4 Role of EcoDam in bridging metabolic signals to epigenetic modifications

Bacterial cells switch in a stochastic manner between different physiological states to ensure their survival under various environmental conditions. Different systems have positive and negative effects on gene expression and metabolism to ensure the better fitness of the organism in the varied environmental conditions. One of such responses is the "stringent response" which is induced upon amino acid starvation and triggers a physiological reprogramming of the bacterial cell. Other responses include phase variation or phenotypic switch under fluctuating selective pressures, and stress responses induced upon approaching the stationary phase. These responses can contribute to regulation of many aspects of microbial cell biology that are sensitive to changing environment and nutrient availability like growth, adaptation, secondary metabolism, survival, persistence, cell division, motility, biofilms, development, competence and virulence (Brown and Kornberg, 2004; Stumpf and Foster, 2005; Mclnerney *et al.*, 2006).

Signaling of the stringent and other stress associated responses is achieved by the accumulation of effector molecules generated in different metabolic pathways (Potrykus *et al.*, 2008). These effector molecules include guanosine pentaphosphate (pppGpp), guanosine tetraphosphate (ppGpp) and long chain polyphosphates, which have long been implicated in the regulation of bacterial adjustments to stress conditions and can

profoundly affect cellular physiology in response to nutrient deprivation (Potrykus *et al.*, 2008; Keasling *et al.*, 1993; Manganelli, R., 2007; Kornberg, *et al.*, 1999). In *E. coli* two enzymes, RelA and SpoT, control the levels of (p)ppGpp: RelA (ppGpp synthetase I) synthesizes ppGpp in response to uncharged tRNA (the consequence of amino acid starvation), SpoT possesses weak synthetase activity and is the sole hydrolase for ppGpp degradation (Ferullo *et al.*, 2008; Braeken *et al.*, 2006). In *E. coli*, polyphosphates (polyP) are synthesized by the polyphosphate kinase (PPK), an inner membrane associated enzyme that catalyses the reversible conversion of the terminal phosphate of ATP to polyP. Another membrane associated enzyme, the exopolyphosphatase PPX, is involved in polyP degradation and contributes to the maintenance of the polyP dynamic balance in the cell (Kuroda *et al.*, 1997).

3.4.1 Experiments and Results

In order to study the biological functions of EcoDam and its relevance to different physiological states of the cell, EcoDam-GFP fusion protein was expressed inside HMS174 (DE3) *E. coli* strain. Unexpectedly, GFP-EcoDam was found to be localized in discrete spots at either a single or both cell poles in the majority of cells. We also observed in some cells that the whole cell was fluorescent, while some cells were fluorescing with additional distinct spots distributed through the cell in the form of patches. Localisation of EcoDam in spots was not related to the expression level and also observed at very low expression (pers. communication, M. Schlickenrieder, PhD thesis, Giessen, 2006).

In 2008, Zhao *et al*. reported a spot-localisation of a group II intron encoded reverse transcriptase (Zhao *et al*., 2008). They further showed that this effect is mainly due to polyphosphate binding of the reverse transcriptase by demonstrating that *E. coli* strains having altered polyphosphate response do not show spotty localisation of the reverse transcriptase. Interestingly, the reverse transcriptase spots coincide in their localization with the spots corresponding to the EcoDam. This led us to hypothesize that polyphosphate binding could be the reason of the spotty localisation of EcoDam as well.

This connection is reasonable, since we have already observed strong binding of EcoDam to phosphocellulose (Urig *et al.*, 2002). This observation lead us to assume that EcoDam might be involved in bridging epigenetic signals (DNA methylation) to metabolic signals (stress response generated PolyP), which may represent a new direct connection of metabolic state, epigenetic signalling and gene regulation in bacteria.

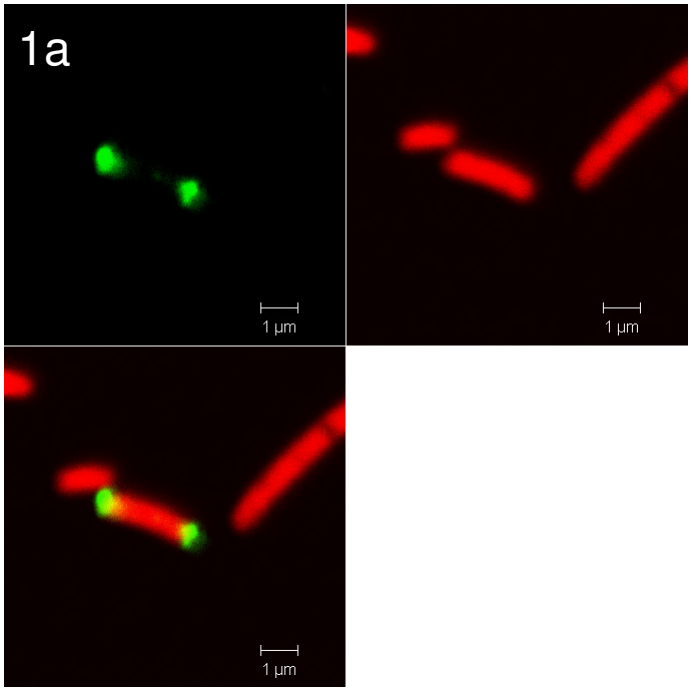
To investigate if the spotty localisation of EcoDam is related to polyphosphate, the EcoDam-GFP fusion protein was expressed inside different *E. coli* strains that have altered polyphosphate response and all are derivative of HMS174 (DE3). These include *gppA, uhbT, wcaK, ynbC, zntR* (which accumulate polyP), and *ppK* mutant which has reduced polyP level (Zhao *et al.*, 2008).

In the first instance, EcoDam-GFP fusion protein was expressed in wildtype HMS174 (DE3) where EcoDam-GFP showed a spotty distribution (Figure 1A and 1B) similar as seen before. In majority of the wildtype HMS174 (DE3) population (60%), EcoDam-GFP fusion protein was polar localized which was present either at one or both ends (Figure 1A and 1B). A subpopulation of cells (25%) showed patches of EcoDam spots distributed throughout the cell (Figure 1c). 5% of the cells showed homogenous distribution of EcoDam. Remaining 10% of the population showed no detectable GFP fluorescence signals. GFP alone always showed homogenous distribution inside cell. Costaining with propidium iodide (PI) showed that the spots do not overlap with the DNA (Figure 1A). Similar spotty localization of EcoDam-GFP fusion protein was observed in Tuner pLysS *E. coli* strain as well, independent of expression level and was observed at very low expression level (2mM IPTG).

Then, the EcoDam-GFP was expressed in a HMS174 (DE3) *gppA* mutant strain which has a deletion in the guanosine-5'-triphosphate-3'-diphosphate pyrophosphatase, an enzyme which dephosphorylates pppGpp to produce ppGpp, and processively hydrolyzes intracellular polyP liberating orthophosphate (Keasling *et al.*, 1993). This strain (*gppA*) shows an increased polyphosphate content and the polyphosphate is no longer localized at the polar spots (Zhao *et al.*, 2008). In agreement with our model, we observed a

homogenous distribution of EcoDam in the *gppA* strain (Figure 2). The *gppA* strain, which has accumulated polyphosphate moiety, showed homogenous distribution of EcoDam-GFP fusion protein and was present in the majority of cells (80%). 10% of the population showed patches of spots which were found to be polar or some time distributed throughout the cell. Remaining 10% of the population showed no detectable GFP fluorescence signals.

WT HMS174 (DE3)



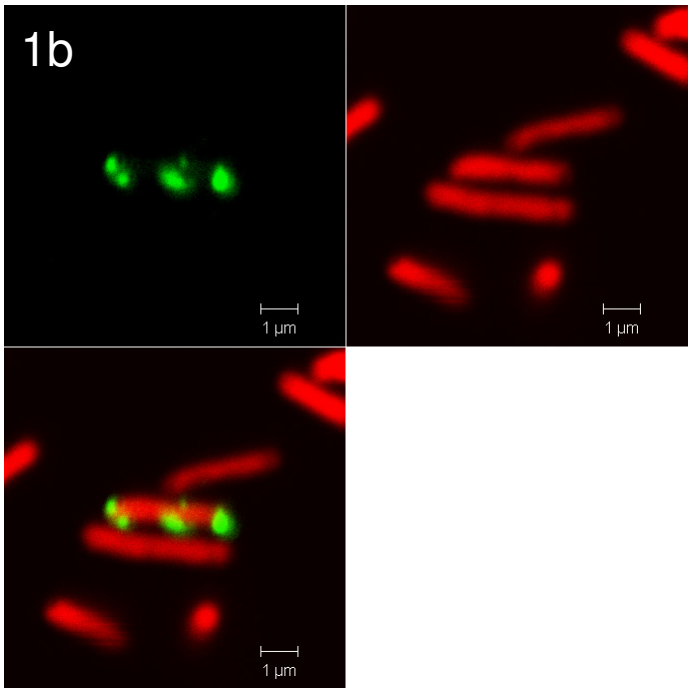


Figure 1A (1a and1b): EcoDam-GFP fusion protein showing spotty localization inside wildtype HMS174 (DE3). Spots were mostly polar (60%), present either at one or both ends.

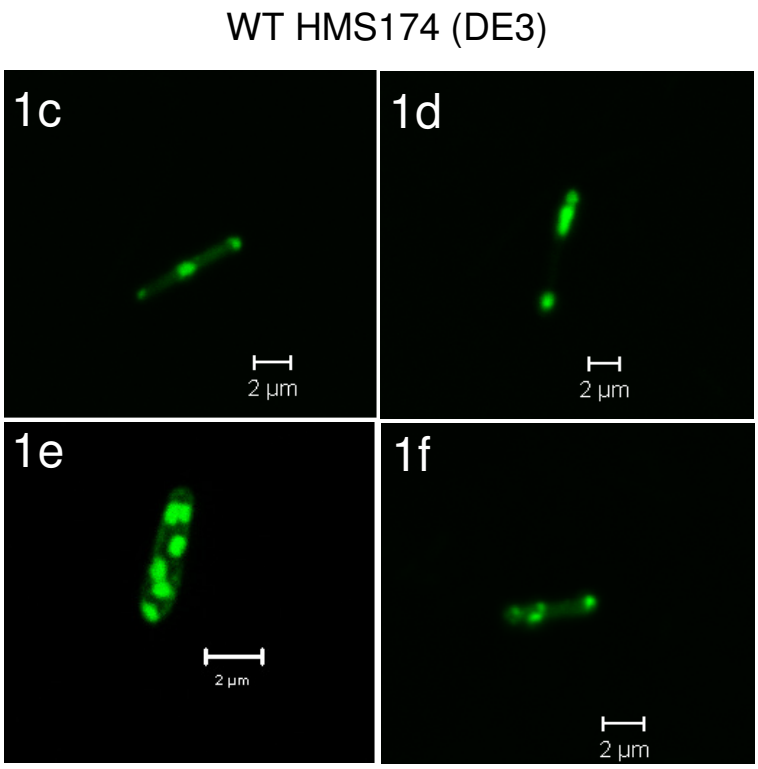


Figure 1B (1c, 1d, 1e and 1f): EcoDam-GFP fusion protein showing spotty localization inside wildtype HMS174 (DE3). Spots were mostly polar (60%), present either at one or both ends. A subpopulation of cells (25%) showed patches of EcoDam spots distributed throughout the cell (1c and 1f).

gppA mutant

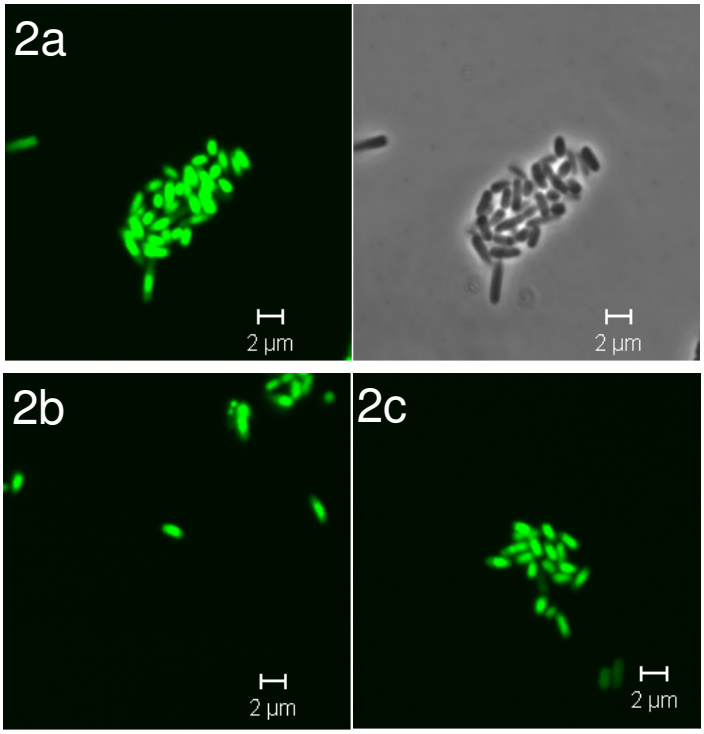


Figure 2: EcoDam-GFP fusion protein showing homogenous distribution inside *gppA* mutant.

The homogenous distribution of EcoDam inside *gppA* mutant encouraged us to investigate other mutants (*uhbT, wcaK, ynbC, zntR*) which have accumulated polyphosphates. Interestingly, like *gppA*, all of the other mutants (*uhbT, wcaK, ynbC, zntR*) showed diffused localization of EcoDam in the majority of cells (80%) (Figure 3a, 3b, 3c, and 3d).

To confirm the conclusion that diffused EcoDam localization is due to accumulation of polyphosphates inside the mutants, we also examined EcoDam-GFP localisation in a *ppk* disruptant strain which has lower level of polyP. As expected, we observed EcoDam-GFP fusion protein was spotty. These spots followed a similar pattern like in wildtype HMS174 (DE3) (Figure 4a, 4b and 4c).

The most well studied stress response is RelA-dependent (ppGpp synthetase I) stringent response, which leads to the accumulation of ppGpp. It has been reported that in addition

to the well studied transcriptional changes caused by switching from normal sigma factor (σ^{70}) to alternative sigma factors (σ^{S}, σ^{32}, σ^{54}), the ppGpp accumulation promotes the arrest of replication, but the mechanism of such cell cycle arrest remains unclear (Chiaramello *et al.*, 1990; Rokeach *et al.*, 1986; Schreiber *et al.*, 1995; Potrykus *et al.*, 2008; Magnusson *et al.*, 2005; Braeken *et al.*, 2006). It is known that (p)ppGpp is also able to inhibit exopolyphosphatase (PPX) activity, causing a 100- to 1000-fold increase of polyP cellular level (Kuroda *et al.*, 1997). The (p)ppGpp-mediated polyP accumulation is responsible (through unknown mechanism) for the induction of two genes essential for the full development of the stringent response: recA and rpoS (Shiba *et al.*, 2000).

Recently in 2008, Ferullo *et al.* reported that replication arrest, which is a characteristic feature of stringent response, is RelA dependent and EcoDam and SeqA proteins are essential to enforce full stringent response. To understand the involvement of EcoDam during stringent response, we performed an experiment where stringent response is triggered by addition of serine hydroxamate to the medium (which is a derivative of serine and acts as a competitive inhibitor of seryl-tRNA synthetase) (Tosa *et al.*, 1971).

When the wildtype HMS174 (DE3) cells were grown in the presence of serine hydroxamate for 1.5 hour, the EcoDam-GFP fusion protein was no longer present in the form of spots, but a homogenous distribution of EcoDam-GFP fluorescence throughout the cell was observed (Figure 5a, 5b and 5c). Interestingly, the homogenous distribution effect was abolished when the cells were recovered by growing them in normal LB media in the absence of serine hydroxamate for 2hours. The recovered cells again showed spotty EcoDam-GFP fluorescence (Figure 6a, 6b, 6c, and 6d). These results indicate that induction of stringent response by addition of serine hydroxamate, leads to a redistribution of EcoDam inside the cell.

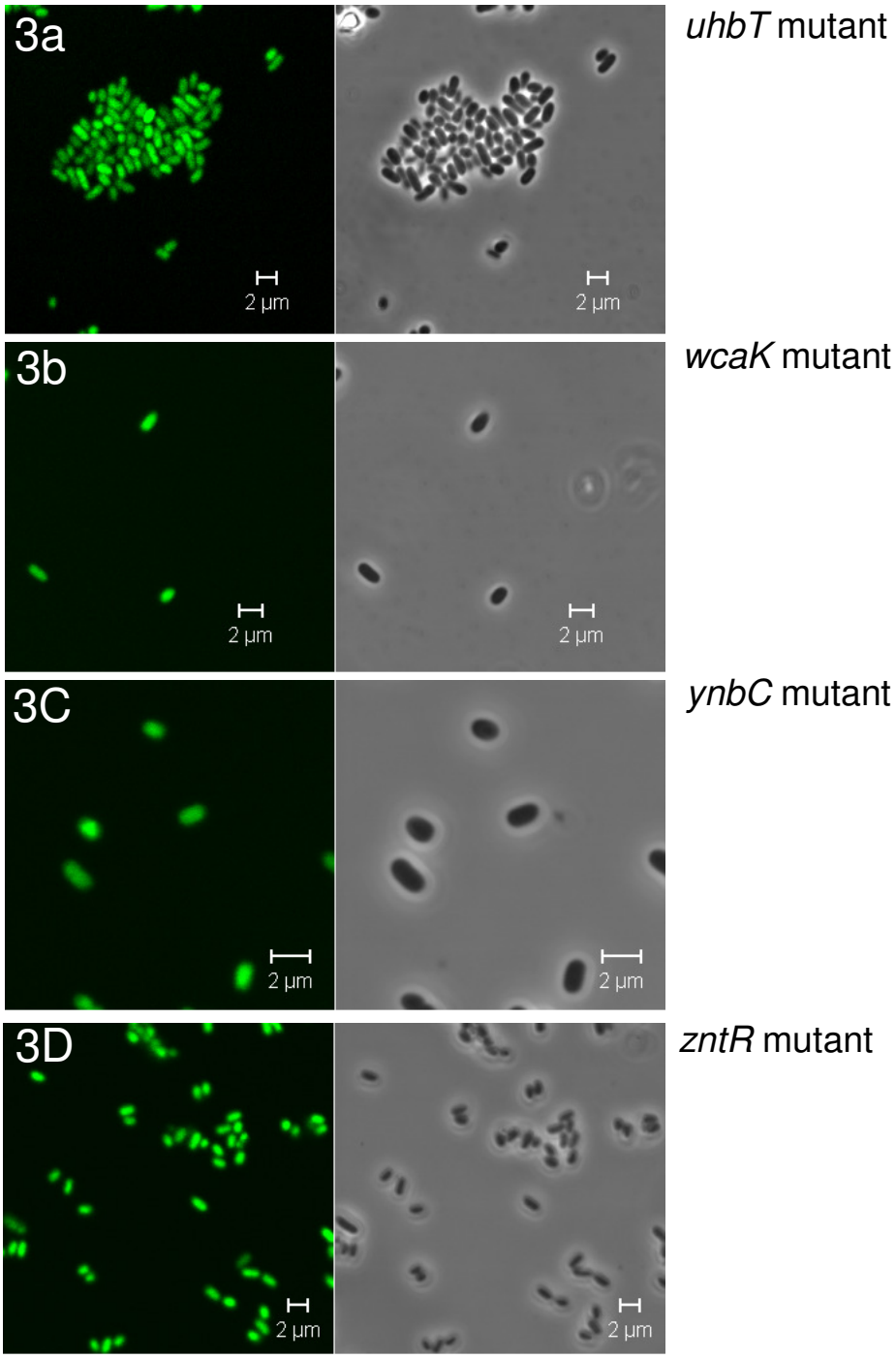


Figure 3: EcoDam-GFP fusion protein showing homogenous distribution inside mutants (*uhbT, wcaK, ynbC, zntR*) having increased level of polyphosphates.

ppk mutant

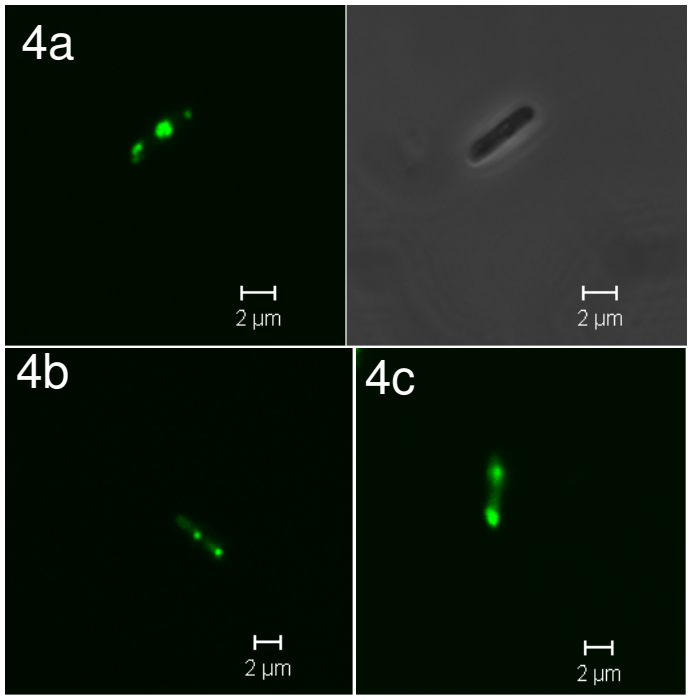


Figure 4: EcoDam-GFP fusion protein showing spotty localization inside *ppk* disruptant which has lower level of polyphosphates.

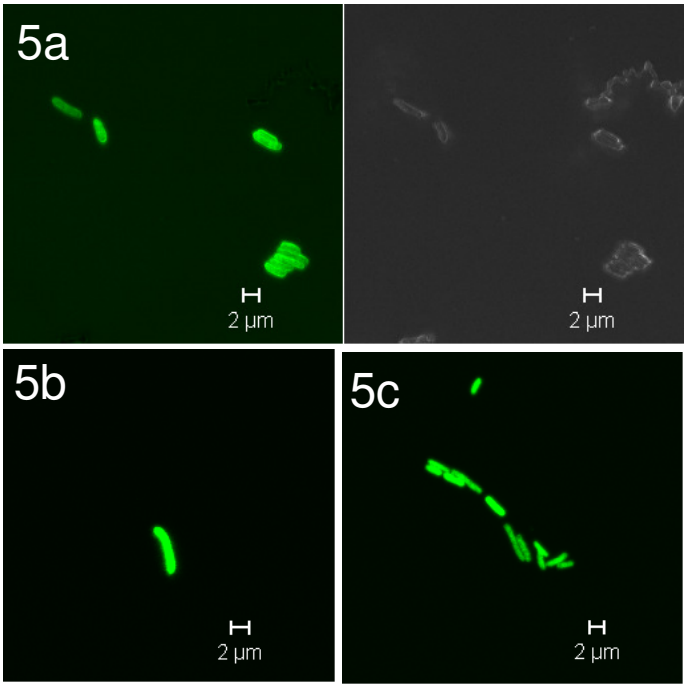


Figure 5: EcoDam-GFP fusion protein showing diffused localization upon treatment of cells with serine hydroxamate.

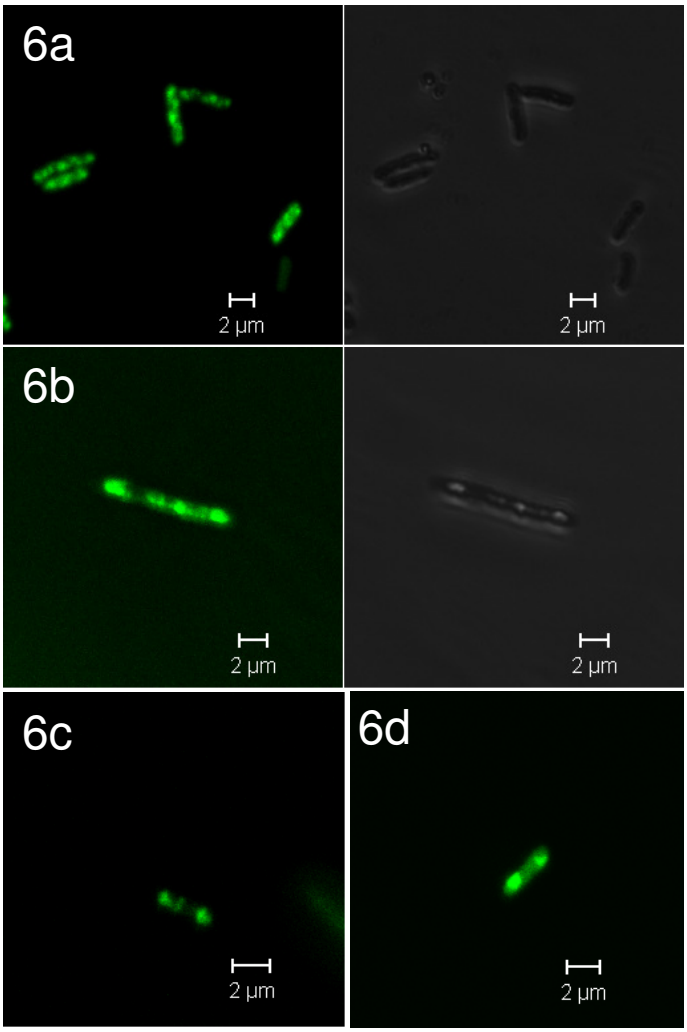


Figure 6: HMS174 (DE3) cells regain spotty localization of EcoDam-GFP fusion protein when the cells are recovered from serine hydroxamate.

3.4.2 Discussion

It has been reported that bacterial cells can undergo various physiological reprogramming causing a dramatic alteration in gene expression including the reduction of rRNA synthesis and increased transcription of amino acid biosynthesis genes during stress response generated by nutrition deprivation (Potrykus *et al.*, 2008; Barker *et al.*, 2001). In bacteria as well as other organisms, control of the cell cycle is an important component of stress responses. Ferullo *et al.* (2008) successfully characterized the DNA replication arrest triggered by ppGpp accumulation during stress response and reported that DNA methylation and SeqA binding to non-origin loci are necessary to enforce full stringent response in *E. coli,* thereby affecting both initiation of replication and chromosome segregation. Herein, we studied the involvement of EcoDam during stress response. We characterized *E. coli* after ppGpp accumulation (stress response regulator) and 6 different other mutants (*gppA, uhbT, wcaK, ynbC, zntR* and *ppk*) with altered polyP level and report that the polyP and ppGpp accumulation sequesters EcoDam as was observed by diffused localization of EcoDam inside polyP and ppGpp accumulated cells. It has been shown that ppGpp and polyP molecules plays very important roles during stress response regulation in the bacteria (Potrykus *et al.*, 2008; Shiba *et al.*, 2000; Riccardo *et al.*, 2007; Kornberg *et al.*, 1999). It has been reported that ppGpp molecules do not show any direct association with EcoDam (Ferullo *et al.*, 2008). Albeit the fact that we do not have any biochemical data yet to show the direct association of EcoDam with polyP, we presume that EcoDam sequestration through its association with accumulated polyP provides another mean for cell cycle control during stress response. We hypothesize that during the stress response, EcoDam sequestration to non-origin loci mediated by polyP, promotes the bacteria for cell cycle arrest by delaying re-methylation at *OriC* region. This is in agreement with the previous finding that arrest with two or more replicated chromosomes during stress response, provides an opportunity for recombinational repair if one of the chromosomes becomes damaged and two potential templates for gene expression (Ferullo *et al.*, 2008). In addition, we speculate that sequestration of EcoDam might help the bacteria to generate altered gene expression profile during stress response. It has been shown that polyP can bind to the RNA polymerase and changes its specificity

to stress response associated gene promoters (Shuichi *et al.*, 1997). Here we propose a model that EcoDam binds to the signaling molecules (polyP) accumulated during stress response and is sequestered to different sites of action to non-origin loci, which might impart to both: the cell cycle arrest (a defense mechanism), and a modulation of the expression of stress associated genes by changes in Dam methylation. Herein we report how metabolic changes associated with varied environmental conditions are connected with epigenetic modifications, a direct connection for bridging metabolic signaling to epigenetic modifications.

3.5 Two alternative conformations of *S*-Adenosyl-L-homocysteine bound to *Escherichia coli* DNA adenine methyltransferase and the implication of conformational changes in regulating the catalytic cycle

In this study, the crystal structure of *Escherichia coli* DNA adenine methyltransferase (EcoDam) in a binary complex with cofactor product AdoHcy (S-adenosyl-homocysteine) has been studied. The results of the study reported that bound AdoHcy showed two alternative conformations, extended or folded. The extended conformation represents the catalytically competent conformation and the folded conformation prevents catalysis. The largest difference between the binary and tertiary structures was in the conformation of N-terminal hexapeptide (^{9}KWAGGK^{14}). Details of the study can be found in the publication of the results (Liebert *et al*., 2007) which is attached in the appendix as chapter 2.

3.6 Transition from EcoDam to T4Dam DNA recognition mechanism without loss of activity and specificity

The EcoDam and T4Dam DNA-(adenine N6)-methyltransferases both methylate the adenine residue in GATC sites. In this study, T4Dam DNA recognition module was transplanted into EcoDam DNA recognition module. The activity and specificity of the transplanted EcoDam variant and the other intermediates of this transition were

investigated. The results of the study showed that the transplanted EcoDam variant (K9A/Y138R) is highly active and specific, and evolutionary transition from EcoDam to T4Dam might be driven by a selection pressure for increased catalytic activity. Details of the study can be found in the publication of the results (Elsawy *et al*., 2009) which is attached in the appendix as chapter 3.

4 Materials and methods

4.1 Kinetic investigation of the mechanism of action of inhibitor-58 on EcoDam

4.1.1 Protein expression and purification

EcoDam gene was cloned inside pET28a (Clonetech) expression vector with T7 inducible promoter. Wild type EcoDam was expressed inside *E.coli* HMS174 (DE3) (*F-recA 1 hsdR($r_{K12}^{-}m_{K12}^{+}$) Rif ^{r}*) (Novagen) host strain as His_6-fusion protein and purified using one-column chromatography (Qiagen Ni-NTA) as described previously (Urig et al., 2002).

4.1.2 20-mer oligonucleotide substrate used for kinetics

The sequences of the 20-mer oligonucleotide substrates (MWG, Ebersbach, München) used for kinetics are as follows:

dam20: 5-gcgacagtgatcggcctgtc-3

mad20m: 5- Bt-gacaggcc gxtc actgtcgc-3

x: N6-Methyladenin Bt = Biotin

DNA Methylation was analyzed in 50mM HEPES (pH7.5) 50mM Nacl, 1mM EDTA, 0.5 mM DTT, 0.2g/l BSA containing 0.76µM [methyl-3H]-AdoMet (NEN) at 37°C as described (Roth and Jeltsch 2000).

4.2 Role of EcoDam in gene transcription regulation in *E. coli*

4.2.1 EcoDam gene cloning inside pUC8 vector

The EcoDam gene was amplified from pET28a-EcoDam construct and cloned in pUC8 vector. EcoRI and HindIII enzyme restriction sites were used for ligation inside pUC8 vector. Following primers were used for EcoDam amplification: forward primer (EcoRI site is introduced) 5-agatataccgaattcgatgggcagcagc-3, reverse primer (HindIII site) 5-gtgcggccgcaagcttttattttttcgcggg-3. Expression of EcoDam was performed in dam$^-$ *E. coli* SCS110 cells (Stratagene, Germany) and RNA samples were prepared for micro array analysis. Control cells were having only empty pUC8 vector.

4.2.2 Micro array experiment

To compare the gene expression profile of dam$^-$ *E.coli* cells (SCS110) (rpsL (Strr) thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44D (lac-proAB) [F' traD36 proAB lacI^{q}ZΔM15]) (Stratagene, Germany) in presence of induced EcoDam versus control cells having only empty plasmid (pUC8), overnight cultures were diluted to 1:100 in fresh LB media containing ampicillin (75µg/ml) and grown at 37°C until OD_{600} reached 0.2. Cultures were induced by addition of IPTG (1mM) and were further grown until OD_{600} reached 0.5. Total RNA was extracted from each sample using the Qiagen RNeasy midi kit. For each DNA microarray set, 50µg total RNA was primed with 3µg of random hexamers (Invitrogen) and labelled by reverse transcription in presence of Cy3- and Cy5-conjugated dCTP (Amersham Bioscience). Dye incorporation efficiencies of at least 10 pmol/ng for each labelling reaction were determined with a NanoDrop™device and the software provided by the manufacturer. For each experiment, an equal amount of 1µg of Cy3-labeled cDNA sample was combined with a Cy5-labeled cDNA sample and hybridized to an *E. coli* K12 V2 OciChip™ microarray following protocols recommended by the supplier (www.ocimumbio.com). After 16 hours of hybridization, the microarrays were washed and scanned at 10 µm resolution with a PerkinElmer ScanExpress scanner. Scanned array images were analysed using the TM4 software

package (Saeed *et al.,* 2003). Spots were quantified and the quality of each spot was verified using the TIGR Spotfinder software provided (Saeed *et al.,* 2003).
The quality control (QC) score was calculated depending on signal-to-noise ratio for every channel in spot and shape quality for each spot and calculating P-values for each channel as result of a t-test comparing the spot pixel set and surrounding background pixel set using the TIGR Spotfinder software (Saeed *et al.,* 2003). Data was normalized by locally weighted linear regression (Cleveland *et al.,* 1988) and a one-class t-test (Pan 2002) was applied to replicated experiments to obtain genes with significant p values ($p<0.05$) using the TIGR MIDAS software.

4.2.3 Real-Time PCR experiment

Real-Time PCR experiments were performed as described previously (Blot et al. 2006). In brief, QuantiTect SYBR Green one-step Real-Time RT-PCR reactions (Qiagen GmbH, Hilden, Germany) were performed in triplicate, following the manual of the manufacturer and using an Mx3000PTM Real-Time cycler (Stratagenes, La Jolla, CA, USA).

4.2.4 Promoter-CFP vector construct

Each promoter was amplified (-400 upstream of +1) and cloned into pBAD24rplL-CFP vector back bone (Kindly provided by Dr Michael Burger, group Prof. Muskhelishvili, Jacobs University Bremen). For this the pBAD24rplL-CFP vector was digested with EcoRV and NheI enzymes (NEB, Germany) and the 4369 bp fragment was gel purified using NucleoSpin extract II column (Macherey-Nagel) . Each PCR amplified promoter was digested with NheI (NEB, Germany) enzyme, phosphorylated by poly nucleotide kinase (PNK) (NEB, Germany) treatment and purified using NucleoSpin column (Macherey-Nagel). NheI digested and phosphorylated PCR products were ligated to 4369 bp digested fragment from pBAD24rplL-CFP vector.

4.2.5 CFP fluorescence measurement

Each promoter-CFP vector construct was transformed in JM110 (Stratagene, Germany) cells. A single colony was inoculated in 5ml LB media along with ampicillin (75µg/ml)

and grown overnight at 37°C in shaker at 200rpm. The overnight culture was diluted to 1:100 in 10ml fresh LB media containing ampicillin (75µg/ml). Cells were grown until OD_{600} reached 0.5. 1.0 OD_{600} culture was resuspended in 200µl PBS. CFP fluorescence of the resuspended culture was measured in a 96 well plate in Tecan-Safire fluorescence reader following manual provided by the supplier (Tecan group ltd., Switzerland).

4.2.6 Primers set used for Real time-PCR analysis

RT frdA forward: 5-atgcgtagccataccgttg-3

RT frdA reverse: 5-cccgctactgtatcgtgaaag-3

RT srlA forward: 5-tggtgcagagtggtttatcg-3

RT srlA reverse: 5-aggctaatcaacagcggaag-3

RT gcvT forward: 5-tacgaacaacacacgctttg-3

RT gcvT reverse: 5-gcatgatgttcgtcgatttg-3

RT narG forward: 5-gatctgccaaaccatgaacc-3

RT narG reverse: 5-gcatcatcgggtatttcagg

RT atpI forward: 5-gtctgtgtcgctcgtgagtc-3

RT atpI reverse: 5-gctgaacagcaatccacttg-3

RT atpA forward: 5-tgaattccaccgaaatcagc-3

RT atpA reverse: 5-aatgcggataacaccgtcac-3

RT rpsL forward: 5-acgtggcgtatgtactcgtg-3

RT rpsL reverse: 5-gtcagacgaacacggcatac-3

RT rfaQ forward: 5-tgcgatatcatggggatatg-3

RT rfaQ reverse: 5-caaaatagggatggtgtcctg-3

4.2.7 (- 400 bp upstream) DNA sequences of promoters containing GATC site

frdA promoter

tcaaacgggaccaaatgaatatcggttaccgtcgcctggctcatacaaggcgtctccacctccagcactccacgatcggcaaag
aaacgacggatctccgccataatcgccgcgcgttttaataagttaggaatggatgcgctcggctgccaggatgccgtttcgctcat
agttaaatctccagtttttgacaagggcacgaagtctactcgcaacgcgacggcgagacaaattttacgcaggaatcaaacagc
ggtgggcagtgactaaaaaaagcacgatctgatggtttagtaattaaattaatcatcttcagtgataatttagccctcttgcgcacta
aaaaaatcgatctcgtcaaatttcagacttatccatcagactatactgttgtacctata (transcriptional start site).

srlA promoter

ctgctgggcattcgggttattagcgaaatcgccgcccatctgcatcacattgtgctgcggctcaagcaggaagccgccagacgg
cgttccggtggtcgtatcagtctcagtaggttttggcttgctgctacaggcggcaagcaacacaaaaaggggaagcaatgttaca
taacgacgcttgaacatgaggggtccatttaacagattcaaccaggggcaagtatggtaaagcatcacgccccgcacaaggaa
gcggtagtcactgcccgatacggactttacataactcaactcattcccctcgctatccttttattcaaactttcaaattaaaatatttatc
tttcattttgcgatcaaaataacacttttaaatctttcaatctgattagattaggttg (transcriptional start site).

gcvT promoter

gccgcaacttatccgccaggcaatgggattaaacgatttgcctgaatggctgcgttaaaaatttctcctctgttgtttatttgataccc
atcacactttcatctcccggtttttcgccgggagattttcctcatttgaaataaactaatttcacctccgttttcgcattatattttctaatg
ccattatttttgatttagtgtttttgacatttttttagctcttaatattgtcttattcaaattgactttctcatcacatcatctttgtatagaaact
ggtgtatttttggtttttattctgtcgcgatttttgcatttttaaccataagctaatgtgatgatcaattttaccttatggttaacagtctgt
ttcggtggtaagttcaggcaaaa (transcriptional start site).

narG promoter

cacttactttagtaagctcctgggattcattcacttgccgccttcctgtaaaccgaattatatagagtaaaatatttgattatccctttgcg
cggcatgatgtcgcgctttttttatgcgtcatttagttacaacatactaatgttatatggtttatttcgccggatttcattaagagccatta
atatgttacccatggggaatactccttaatacccatctgcataaaaatcttaatagtttaaataactacaggtataaaacgtcttaattta
cagtctgttatgtggtggctgttaattatcctaaaggggtatcttaggaatttactttatttttcatccccatcactcttgatcgttatcaatt
cccacgctgtttcagagcgttaccttgcccttа (transcriptional start site).

4.2.8 (- 400 bp upstream) DNA sequences of promoters containing TANAC site

atpI promoter

tggtaagcagaaaataagtcattagtgaaaatatcagtctgctaaaaatcggcgctaagaaccatcattggctgttaaaacattatta
aaaatgtcaatgggtggtttttgttgtgtaaatgtcatttattaaaacagtatctgttttagactgaaatatcataaacttgcaaaggcat
catttgccaagtaaataaatatgctgtgcgcgaacatgcgcaatatgtgatctgaagcacgctttatcaccagtgtttacgcgttattt
acagtttttcatgatcgaacagggttagcagaaaagtcgcaattgtatgcactggaaaaatatttaaacatttattcacctttggcta
cttattgtttgaaatcacgggggcgcaccgtataatttgaccg (transcriptional start site).

atpA promoter (Transcription unit: **atp**BEFHAGDC)

atpBp2
atttgaccgctttttgatgcttgactctaagccttaaagaaagtttatacgacacgcggcatacctcgaagggagcaggagtgaaa
aacgtgatgtctgtgtcgctcgtgagtcgaaacgttgctcggaagcttctgctcgttcagttactggtggtgatagcaagtggattg
ctgttcagcctcaaagacccctttctggggcgtctctgcaataagcgggggcctggcagtctttctgcctaacgttttgtttatgatatt
tgcctggcgtcaccaggcgcatacaccagcgaaaggccgggtggcctggacattcgcatttggcgaagctttcaaagttctggc
gatgttggtgttactggtggtggcgttggcggttttaaaggcggtattcttgc (transcriptional start site).

atpBp1
ttttcatgatcgaacagggttagcagaaaagtcgcaattgtatgcactggaaaaatatttaaacatttattcacctttggctacttattg
tttgaaatcacgggggcgcaccgtataatttgaccgctttttgatgcttgactctaagccttaaagaaagtttatacgacacgcggc
atacctcgaagggagcaggagtgaaaaacgtgatgtctgtgtcgctcgtgagtcgaaacgttgctcggaagcttctgctcgttca
gttactggtggtgatagcaagtggattgctgttcagcctcaaagacccctttctggggcgtctctgcaataagcgggggcctggca
gtctttctgcctaacgttttgtttatgatattt (transcriptional start site).

rpsL promoter

tgccttacgccgcgaactcgccaactacgatgttattttgaggttttgaggcgtttatgctgcacacattacatcgctcaccctggct
gacggattttgctgcgctgctgcgtctgctcagtgaaggagacgaactgctattattgcaagatggcgtaactgccgcagttgac
ggtaaccgctaccttgaaagtctgcgtaatgcccccattaaggtctatgccctgaacgaagaccttattgcccgcggtttgactggt
caaatttcgaacgacatcattctcattgactatactgatttcgtcagacttacggttaagcaccccagccagatggcctggtgatgg
cgggatcgttgtatatttcttgacacctttttcggcatcgccctaaaattcggcg (transcriptional start site).

rfaQ promoter

ttttcgataggccggagccttacgtccgcgcacccagagccgtatccagatcagcggctgaataaggtagagaagggcggtgt
aaagcaattcgagcatagtaaatagctgacttatggatgtgctggggattctatgtatttagctgtggctttaccattacttttcccgttt
ttgacttaaatagcttcagtttggtctgatctgccgctacatcttcatttttttttgtatttttatgcgattcattgaaactcggccccattttca
aatctacataggccgtactgacattatcgaaatgctatttttttatctatttgatttttatgattaaagtatattttgtgtataaaaatcattcg
ggtcggattgctgcgaaagaaatgatacactagcacgt (transcriptional start site).

4.2.9 Primers set used to clone respective promoter in pBAD24rplL-CFP vector

atpI forward primer: 5-tggtaagcagaaaataagtcattagtgaaaatatcagtctgc-3

atpI reverse primer: 5-ctcgtcctgagctagccggtcaaattatacggtgcg-3

rpsL forward primer: 5-tgccttacgccgcgaactcgccaactacgatg-3

rpsL reverse primer: 5-ctcgtcctgagctagccgccgaattttagggcgatgc-3

rfaQ forward primer: 5-ttttcgataggccggagccttacgtccgcg-3

rfaQ reverse primer: 5-ctcgtcctgagctagcacgtgctagtgtatcatttctttcgc-3

frdA forward primer: 5-tcaaacgggaccaaatgaatatcggttaccgtcg-3

frdA reverse primer: 5-ctcgtcctgagctagctataggtacaacagtatagtctg-3

srlA forward primer: 5-ctgctgggcattcgggttattagcgaaatcgc-3

srlA reverse primer: 5-ctcgtcctgagctagccaacctaatctaatcagattg-3

gcvT forward primer: 5-gccgcaacttatccgccaggcaatgggattaaacg-3

gcvT reverse primer: 5-ctcgtcctgagctagcttttgcctgaacttaccaccg-3

narG forward primer: 5-cacttactttagtaagctcctgggattcattcacttgc-3

narG reverse primer: 5-ctcgtcctgagctagctaagggcaaggtaacgctctg-3

4.3 Primers set used for GATC site mutation within respective promoters

gcvT forward primer2: 5-gatacccatctcgagttcatctcccg-3

gcvT reverse primer2: 5-ggtaaaattgaccgtcacattagc-3

narG forward primer2: 5-cacttactttagtaagctcctgggattcattcacttgc-3

narG reverse primer2: 5-ggaattgataacgaccaagagtgatg-3

4.3.1 Primers set used for TANAC site mutation within respective promoters

rfaQ forward primer2: 5-gtggctttaccattacttttc-3

rfaQ reverse primer2: 5-acgtgctcgagtatcatttc-3 (XhoI site is created ctcgag)

rpsL forward primer: 5-tgccttacgccgcgaactcgccaactacgatg-3

rpsL reverse primer2: 5-ctgacgaaatcaatctcgagaatgagaatgatg-3 (XhoI site is created ctcgag)

4.4 Role of EcoDam in bridging metabolic signals to epigenetic modifications

4.4.1 Bacterial strains

Mutant *E.coli* strains (*uhbT, wcaK, ynbC, zntR* and *ppk*) were purchased from laboratory of Prof Lovett ST, Massachusetts, United states. Wildtype HMS174 (DE3) (*F- recA 1 hsdR*($r_{K12}^{-}m_{K12}^{+}$) *Rif*^{r}) strain was purchased from Novagen biotech.

4.4.2 EcoDam-GFP fusion construct

EcoDam-GFP fusion construct was generated by ligating a 2.2 kb fragment from pACYC184 vector backbone (chloramphenicol ^{R}) to a 2 kb fragment from pET28a-EcoDam GFP construct. (T7 promoter). GFP open reading frame is fused to the C-terminal of EcoDam open reading frame in the pET28a-EcoDam-GFP construct. .

Following primers were used to amplify respective fragments:

Pacyc184 forward pr (KpnI restriction site is introduced)
5-aggcgtttaagggtaccaataactgc-3
Pacyc184 reverse pr (SphI restriction site is introduced)
5-cgctagcagcatgccatagtgactgg-3
pET28a EcoDam-GFP forward primer (KpnI restriction site is introduced)
5-tcttccccatcggtaccgtcggcgatataggc-3
pET28a EcoDam-GFP reverse pr (SphI restriction site is introduced)
5-cagggcgcgtcgcatgcgccaatccggatatag-3

4.4.3 Expression and Induction

EcoDam-GFP fusion construct was transformed inside $CaCl_2$ chemical competent cells by heat shock method. Individual colony was grown over night at 37°C in LB medium with appropriate antibiotic (chloramphenicol). Overnight primary culture was diluted to 1:100 and grown in fresh 20ml LB media with chloramphenicol and cells were grown until the OD_{600} of 0.4 was reached. Cells were induced by the addition of 0.5 mM IPTG and grown for another 2 hours.

4.4.4 Sample preparation and Microscopic analysis

850 µl of induced culture was mixed with 100 µl of formaldehyde (40%, Sigma, Germany) and was incubated for 1 hour at the room temperature along with 50 µl of 1M phosphate buffer. The cells were centrifuged at 3000rpm for 3 min. Cell pellet was gently resuspended in the appropriate volume of PBS.

The microscopic analysis was done using a confocal laser scanning microscopic (LSM510-Meta, Zeiss Germany), with the images taken using the 100x objective with oil immersion. Image analysis was performed using the LSM510 software, Release 3.0 (Carl Zeiss Jena GmbH). GFP fluorescence was detected by excitation at 490nm and recording the emission signal with a broad-pass filter ranging from 475 to 525nm.

4.4.5 Serine hydroxamate experiment

Wildtype HMS174 (DE3) cells were transformed with EcoDam-GFP construct. Single colony was picked and grown overnight at 37°C in the presence of chloramphenicol. Overnight culture was diluted to 1:100 and grown in fresh 20ml LB media along with appropriate antibiotic (chloramphenicol) until OD_{600} of 0.2 was obtained. Culture was induced by addition of 0.5mM IPTG and grown until OD_{600} of 0.5 was obtained. Induced culture was splitted into two parts (10ml each). 10ml induced culture was grown with serine hydroxamate (1mg/ml) for 1.5 hour. 850 µl of the culture was taken for microscopic analysis and rest of the serine hydroxamate treated culture was centrifuged at 3000 rpm for 10 min. Cells were resuspended in equal volume of fresh LB media and recovered by growing them at 37°C for 2 hours along with 0.5mM IPTG in the presence of chloramphenicol. After 2 hours of recovery sample was taken for microscopic analysis.

5 References

1 Adams, R.L., McKay, E.L., Craig, L.M. and Burdon, R.H. (1979) Methylation of mosquito DNA. *Biochim Biophys Acta.* **563**, 72-81.

2 Allers, T. and Leach, D.R. (1995) DNA palindromes adopt a methylation-resistant conformation that is consistent with DNA cruciform or hairpin formation in vivo. *J. Mol. Biol.* **252**, 70-85.

3 Altamirano, M.M., Blackburn, J.M., Aguayo, C. and Fersht, A.R. (2000) Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature.* **403**, 617–622.

4 Appel, D., Lutz-Wahl, S., Fischer, P., Schwaneberg, U. and Schmid, R.D. (2001) A P450 BM-3 mutant hydroxylates alkanes, cycloalkanes, arenes and heteroarenes. J Biotechnol. **88**, 167-71.

5 Arnold, F.H. (1998) Design by directed evolution. *Acc Chem Res.* **31**, 125–131.

6 Ashapkin, V.V., Kutueva, L.I. and Vanyushin, B.F. (2002) The gene for domains rearranged methyltransferase (DRM2) in *Arabidopsis thaliana* plants is methylated at both cytosine and adenine residues. *FEBS Lett.* **532,** 367-72.

7 Bachmann, B.J. (1990) Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol Mol Biol Rev.* **54**, 130-197.

8 Bale, A., d'Alarcao, M. and Marinus, M.G. (1979) Characterization of DNA adenine methylation mutants of *Escherichia coli* K12. *Mutat Res.* **59**, 157-65.

9 Barker, M.M., Gaal, T. and Gourse, R.L. (2001) Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. *J. Mol. Biol.* **305**, 689-702.

10 Barras, F. and Marinus, M.G. (1989) The great GATC: DNA methylation in *E. coli. Trends Genet*, **5**, 139-143.

11 Bloom, J.D., Labthavikul, S.T., Otey, C.R. and Arnold, F.H. (2006) Protein stability promotes evolvability. *Proc. Natl. Acad. Sci*. USA. **103**, 5869-74.

12 Bornscheuer, U. T. and Pohl, M. (2001) Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol.* **5**, 137-143.

13 Bornscheuer, U. T. (2002) Methods to increase enantioselectivity of lipases and esterases. *Curr. Opin. Biotechnol.* **13,** 543-547.

14 Bosma, T., Damborsky, J., Stucki, G. and Janssen, D.B. (2002) Biodegradation of 1,2,3-Trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene. *Appl Environ Microbiol.* **68**, 3582–3587.

15 Boye, E., Marinus, M.G. and Løbner-Olesen, A. (1992) Quantitation of Dam methyltransferase in *Escherichia coli*. *J Bacteriol.* **174**, 1682-5.

16 Braun, R.E., O'Day, K. and Wright, A. (1985) Autoregulation of the DNA replication gene dnaA in *E. coli* K-12. *Cell.* **40**, 159-69.

17 Braeken, K., Moris, M., Daniels, R., Vanderleyden, J. and Michiels, J. (2006) New horizons for (p)ppGpp in bacterial and plant physiology. *Trends Microbiol.* **14**, 45-54.

18 Brown, M.R. and Kornberg, A. (2004) Inorganic polyphosphate in the origin and survival of species. *Proc. Natl. Acad. Sci*. USA. **101**, 16085-7.

19 Buchholz, F. and Stewart, A.F. (2001) Alteration of Cre recombinase site specificity by substrate-linked protein evolution. *Nat. Biotechnol.* **19**, 1047-52.

20 Bujnicki, J.M. (2002) Sequence permutations in the molecular evolution of DNA methyltransferases. *BMC Evol. Biol.* 2:3.

21 Campbell, J.L. and Kleckner, N. (1988) The rate of Dam-mediated DNA adenine methylation in *Escherichia coli*. *Gene.* **74**, 189-90.

22 Campbell, J.L. and Kleckner, N. (1990) *E. coli* oriC and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell.* **62**, 967-79.

23 Canada, K.A., Iwashita, S., Shim, H. and Wood, T.K. (2002) Directed evolution of toluene *ortho*-monooxygenase for enhanced 1-naphthol synthesis and chlorinated ethene degradation. *J Bacteriol.* **184**, 344-349.

24 Casadesus, J. and Low, D. (2006) Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev.* **70**, 830-56.

25 Chatterjee, R. and Yuan, L. (2006) Directed evolution of metabolic pathways. *Trends Biotechnol.* **24**, 28-38.

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

27 Cheng, X. Blumenthal, R.M. (1999) *S-adenosylmethionine-dependent Methyltransferases: Structures and functions*. World Scientific Publishing, Singapore.

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

29 Chen, L., Paulsen, D. B., Scruggs, D. W., Banes, M. M., Reeks, B. Y. and Lawrence, M. L. (2003) Alteration of DNA adenine methylase (Dam) activity in *Pasteurella multocida* causes increased spontaneous mutation frequency and attenuation in mice. *Microbiology.* **149**, 2283-2290.

30 Chen, R. (1999) A general strategy for enzyme engineering. *Trends Biotechnol.* **17**, 344-345.

31 Chen, R. (2001) Enzyme engineering: rational redesign versus directed evolution. *TRENDS in Biotechnol*. **19**, 13-17.

32 Chen, R. (2001) Enzyme Technologies for Pharmaceutical and Biotechnological Application. (Kirst, H. A., Yeh, w. k. and Zmijewski, M. J. eds) Marcel Dekker Inc.

33 Chen, R., Greer, A. and Dean, A.M. (1996) Redesigning secondary structure to invert coenzyme specificity in isopropylmalate dehydrogenase. *Proc. Natl. Acad. Sci*. USA. **93**, 12171-12176.

34 Chiaramello, A.E., Zyskind, J.W. (1990) Coupling of DNA replication to growth rate in *Escherichia coli*: a possible role for guanosine tetraphosphate. *J Bacteriol.* **172**, 2013-9.

35 Cleveland, W., Devlin, S., Grosse, E. (1988) Regression by local fitting. *Journal of Econometrics.* **37**, 87--114.

36 Collier, J., McAdams, H. H., and Shapiro, L. (2007) A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc. Natl. Acad. Sci*. USA. **104**, 17111-17116.

37 Colot, V. and Rossignol, J.L. (1999) Eukaryotic DNA methylation as an evolutionary device. *Bioessays*. **21**, 402-411.

38 Dai, M. and Copley, S.D. (2004) Genome shuffling improves degradation of the anthropogenic pesticide pentachlorophenol by *Sphingobium chlorophenolicum* ATCC 39723. *Appl Environ Microbiol.* **70**, 2391–2397.

39 Farinas, E., Schwaneberg, U., Anton, G. and Arnold, F. H. (2001) Directed evolution of a cytochrome P450 monooxygenase for alkane oxidation. *Adv. Synth. Catal.* **343**, 601-606.

40 Fastrez, J. (1997) *In vivo* versus *in vitro* screening or selection for catalytic activity in enzymes and abzymes. *Mol. Biotechnol.* **7**, 37-55.

41 Faumann, E.B., Blumenthal, R.M. and Cheng, X. (1999) Structure and evolution of AdoMet-dependent methyltransferases. In Cheng, X. and Blumenthal, R.M. (eds), *S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions.* World Scientific Publishing, Singapore, pp. 1–38.

42 Ferullo, D.J. and Lovett, S.T. (2008) The stringent response and cell cycle arrest in *Escherichia coli*. *PLoS Genet.* **4**, e1000300.

43 Fong, S., Machajewski, T.D., Mak, C.C. and Wong, C.H. (2000) Directed evolution of D-2-keto-3-deoxy-6-phosphogluconate aldolase to new variants for the efficient synthesis of D- and L-sugars. *Chem. Biol.* **7**, 873-883.

44 Garvie, C.W. and Wolberger, C. (2001) Recognition of specific DNA sequences. *Mol Cell.* **8**, 937-46.

45 Georgiou, G. and DeWitt, N. (1999) Enzyme beauty. *Nature Biotechnol.* **17**, 1161–1162.

46 Gershenson, A., Schauerte, J.A., Giver, L. and Arnold, F.H. (2000) Tryptophan phosphorescence study of enzyme flexibility and unfolding in laboratory-evolved thermostable esterases. *Biochemistry.* **39**, 4658–4665.

47 Getzoff, E.D., Cabelli, D.E., Fisher, C.L., Parge, H.E., Viezzoli, M.S., Banci, L., and Hallewell, R.A. (1992) Faster superoxide dismutase mutants designed by enhancing electrostatic guidance. *Nature.* **358**, 347-351.

48 Hale, W.B., van der Woude, M.W. and Low, D.A. (1994) Analysis of nonmethylated GATC sites in the *Escherichia coli* chromosome and identification of sites that are differentially methylated in response to environmental stimuli. *J Bacteriol.* **176**, 3438-41.

49 Henaut, A., Rouxel, T., Gleizes, A., Moszer, I., and Danchin, A. (1996) Uneven distribution of GATC sequence motifs in the *E. coli* chromosome, its plasmids and its phages. *J. Mol. Biol.* **257**, 574-585.

50 Herman, G.E. and Modrich, P. (1982) *Escherichia coli* dam methylase. Physical and catalytic properties of the homogeneous enzyme. *J. Biol. Chem.* **257**, 2605-12.

51 Herman, G.E. and Modrich, P. (1981) *Escherichia coli* K-12 clones that overproduce dam methylase are hypermutable. *J Bacteriol.* **145**, 644-6.

52 Hernday, A., Krabbe, M., Braaten, B. and Low, D. (2002) Self-perpetuating epigenetic pili switches in bacteria. *Proc. Natl. Acad. Sci*. USA. **99**, 16470-6.

53 Hernday, A., Braaten, B. and Low, D. (2004) The intricate workings of a bacterial epigenetic switch. *Adv Exp Med Biol.* **547**, 83-89.

54 Hernday, A.D., Braaten, B.A., Broitman-Maduro, G., Engelberts, P. and Low, D.A. (2004) Regulation of the pap epigenetic switch by CpxAR: phosphorylated CpxR inhibits transition to the phase ON state by competition with Lrp. *Mol Cell.* **16**, 537-47.

55 Hernday, A.D., Braaten, B.A. and Low, D.A. (2003) The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. *Mol Cell.* **12**, 947-57.

56 Heusipp, G., Fälker, S. and Schmidt, M.A. (2007) DNA adenine methylation and bacterial pathogenesis. *Int J Med Microbiol.* **297**, 1-7.

57 Honma, Y., Fernandez, R. E. and Maurelli, A. T. (2004) A DNA adenine methylase mutant of *Shigella flexneri* shows no significant attenuation of virulence. *Microbiology.* **150**, 1073-1078.

58 Horton, J.R., Liebert, K., Hattman, S., Jeltsch, A. and Cheng, X. (2005) Transition from nonspecific to specific DNA interactions along the DNA recognition pathway of dam methyltransferase. *Cell.* **121**, 349-61.

59 Horton, J.R., Liebert, K., Bekes, M., Jeltsch, A. and Cheng, X. (2006) Structure and substrate recognition of the *Escherichia coli* DNA Adenine methyltransferase. *J. Mol. Biol.* **358**, 559-570.

60 Iyer, R.R., Pluciennik, A., Burdett, V. and Modrich, P.L. (2006) DNA Mismatch Repair: Functions and Mechanisms. *Chem. Rev.* **106**, 302-323.

61 Jeltsch, A. (1999) Circular permutations in the molecular evolution of DNA methyltransferases. *J. Mol. Evol*. **49**, 161-164.

62 Jeltsch, A. (2002) Beyond Watson and Crick : DNA Methylation and Molecular Enzymology of DNA Methyltransferases. *ChemBioChem*, **3**, 274-293.

63 Jeltsch, A. (2003) Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? Gene. **317**, 13-16.

64 Jeltsch, A. and Gumport, R.I. (2004) DNA modification: Bacterial methyltransferases. In: Encyclopedia of Biological Chemistry (Lennarz, W.J. and Lane, D.M., eds.) Academic press, San Diego, CA, ISBN 0124437109.

65 Julio, S.M., Heithoff, D.M., Provenzano, D., Klose, K.E., Sinsheimer, R.L., Low D.A. and Mahan, M.J. (2001) DNA adenine methylase is essential for viability and plays a role in the pathogenesis of *Yersinia pseudotuberculosis* and *Vibrio cholerae*. *Infect Immun.* **69**, 7610-5.

66 Julio, S.M., Heithoff, D.M., Sinsheimer, R.L., Low, D.A. and Mahan, M.J. (2002) DNA adenine methylase overproduction in *Yersinia pseudotuberculosis* alters YopE expression and secretion and host immune responses to infection. *Infect. Immun.* **70**, 1006-1009.

67 Kazlauskas, R.J. and Bornscheuer, U.T. (2009) Finding better protein engineering strategies. *Nat Chem Biol.* **5**, 526-529.

68 Keasling, J.D., Bertsch, L. and Kornberg, A. (1993) Guanosine pentaphosphate phosphohydrolase of *Escherichia coli* is a long-chain exopolyphosphatase. *Proc. Natl. Acad. Sci*. USA. **90**, 7029-33.

69 Klimasauskas, S., Kumar, S., Roberts, R.J. and Cheng, X. (1994) HhaI methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357-369.

70 Kornberg, A., Rao, N.N. and Ault-Riche, D. (1999) Inorganic polyphosphate: a molecule of many functions. *Annu Rev Biochem.* **68**, 89-125.

71 Kuchner, O. and Arnold, H.F. (1997) Directed evolution of enzyme catalysts. *Trends Biotechnol.* **15**, 523-30.

72 Kuroda, A., Murphy, H., Cashel, M. and Kornberg, A. (1997) Guanosine tetra- and pentaphosphate promote accumulation of inorganic polyphosphate in *Escherichia coli*. *J. Biol. Chem.* **272**, 21240-3.

73 Kuroda, A. and Kornberg, A. (1997) Polyphosphate kinase as a nucleoside diphosphate kinase in *Escherichia coli* and *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci*. USA. **94**, 439-42.

74 Kusano, S. and Ishihama, A. (1997) Functional interaction of *Escherichia coli* RNA polymerase with inorganic polyphosphate. *Genes Cells.* **2**, 433-41.

75 Liebert, K., Hermann, A., Schlickenrieder, M. and Jeltsch, A. (2004) Stopped-flow and mutational analysis of base flipping by the dam DNA-(adenine-N6)-methyltransferase. *J Mol Biol*., **341**, 443-54.

76 Li, Q.S., Schwaneberg, U., Fischer, P. and Schmid, R.D. (2000) Directed evolution of the fatty-acid hydroxylase P450 BM-3 into an indole-hydroxylating catalyst. *Chemistry.* **6**, 1531-6.

77 Løbner-Olesen, A., Boye, E. and Marinus, M.G. (1992) Expression of the *Escherichia coli* dam gene. *Mol Microbiol.* **6**, 1841-51.

78 Løbner-Olesen, A., Hansen, F.G., Rasmussen, K.V., Martin, B., Kuempel, P.L. (1994) The initiation cascade for chromosome replication in wild-type and Dam methyltransferase deficient *Escherichia coli* cells. *EMBO J.* **13**, 1856-62.

79 Løbner-Olesen, A., Marinus, M.G. and Hansen, F.G. (2003) Role of SeqA and Dam in *Escherichia coli* gene expression: a global/microarray analysis. *Proc. Natl. Acad. Sci.* USA. **100**, 4672-7.

80 Low, D. A., Weyand, N. J. and Mahan, M. J. (2001) Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infect Immun.* **69**, 7197–7204.

81 Luetz, S., Giver, L. and Lalonde, J. (2008) Engineered enzymes for chemical production. *Biotechnol Bioeng.* **101**, 647-53.

82 Lutz, S. and Bornscheuer, U.T. (eds) (2009) Proetin engineering handbook (Wiley-VCH, Weinheim, 2009).

83 Magnusson, L.U., Farewell, A. and Nyström, T. (2005) ppGpp: a global regulator in *Escherichia coli*. *Trends Microbiol.* **13**, 236-42.

84 Mahan, M.J., Heithoff, D.M., Sinsheimer, R.L. and Low, D.A. (2000) Assessment of bacterial pathogenesis by analysis of gene expression in the host. *Ann. Rev. Genet.* **34**, 139-164.

85 Mahan, M. J. and Low, D. A. (2001) DNA methylation regulates bacterial gene expression and virulence. *ASM News.* **67**, 356-361.

86 Malone, T., Blumenthal R.M. and Cheng, X. (1995) Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes. *J Mol Biol*., **253**, 618-32.

87 Manganelli, R. (2007) Polyphosphate and stress response in *Mycobacteria*. *Mol Microbiol.* **65**, 258-60.

88 Marczynski, G.T. and Shapiro, L. (2002) Control of chromosome replication in *Caulobacter crescentus*. *Annu Rev Microbiol*. **56**, 625-56.

89 Marinus, M.G. (1973) Location of DNA methylation genes on the *Escherichia coli* K-12 genetic map. *Mol Gen Genetic.* **127**, 47-55.

90 Marinus, M.G. and Morris, R.N. (1974) Biological function for 6-methyladenine residues in the DNA of *Escherichia coli* K12. *J. Mol. Biol.* **85**, 309-22.

91 Marinus, M.G., Poteete, A. and Arraj, J.A. (1984) Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli* K-12. *Gene.* **28**, 123-5.

92 Marinus, M.G. (2000) Recombination is essential for viability of an *Escherichia coli* dam (DNA adenine methyltransferase) mutant. *J Bacteriol.* **182**, 463-8.

93 Marinus, M.G. (2006) Methylation of DNA in: Neidhardt, F.C., Curtiss, R., Ingraham J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M. and Umbarger, H.E. (Eds.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, ASM Press, Washington, DC, 1996, pp. 782–791.

94 Marinus, M.G. and Casadesus, J. (2009) Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev.* **33**, 488-503.

95 Matsumura I. (2000) *IBC's Fifth Annual World Congress on Enzyme Technologies*, Las Vegas.

96 Matsumura, I. and Ellington, A.D. (2001) In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J. Mol. Biol.* **305**, 331-339.

97 McInerney, P., Mizutani, T. and Shiba, T. (2006) Inorganic polyphosphate interacts with ribosomes and promotes translation fidelity *in vitro* and *in vivo*. *Mol Microbiol.* 60, 438-47.

98 McLoughlin, S.Y., Jackson, C., Liu, J.W. and Ollis, D. (2005) Increased expression of a bacterial phosphotriesterase in *Escherichia coli* through directed evolution. *Protein Expr Purif.* **41**, 433-440.

99 Miyazaki, K., Wintrode, P.L., Grayling, R.A., Rubingh, D.N. and Arnold, F.H. (2000) Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J. Mol. Biol.* **297**, 1015-26.

100 Morawski, B., Quan, S. and Arnold, F.H. (2001) Functional expression and stabilization of horseradish peroxidase by directed evolution in *Saccharomyces cerevisiae. Biotechnol Bioeng.* **76**, 99-107.

101 Morley, K.L. and Kazlauskas, R.J. (2005) Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.* **23**, 231-237.

102 Naki, D., Paech, C., Ganshaw, G. and Schellenberger V. (1998) Selection of subtilisin-hyperproducing *Bacillus* in a highly structured environment. *Appl. Environ. Microbiol.* **49**, 290-294.

103 Oshima, T., Wada, C., Kawagoe, Y., Ara, T., Maeda, M., Masuda, Y., Hiraga, S. and Mori, H. (2002) Genome-wide analysis of deoxyadenosine methyltransferase-mediated control of gene expression in *Escherichia coli*. *Mol Microbiol.* **45**, 673-95.

104 Palmer, B.R. and Marinus, M.G. (1994) The dam and dcm strains of *Escherichia coli*-a review. *Gene.* **143**, 1-12.

105 Parniewski, P., Kwinkowski, M., Wilk, A. and Klysik, J. (1990) Dam methyltransferase sites located within the loop region of the oligopurine-oligopyrimidine sequences capable of forming H-DNA are undermethylated in vivo. *Nucleic Acids Res.* **18**, 605-11.

106 Peterson, S.N. and Reich, N.O. (2006) GATC flanking sequences regulate Dam activity: evidence for how Dam specificity may influence pap expression. *J. Mol. Biol.* **355**, 459-72.

107 Pingoud, A., Fuxreiter, M., and Pingoud, V. (2005) Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life. Sci*. **62**, 685-707.

108 Plumbridge, J. (1987) The role of dam methylation in controlling gene expression. *Biochimie.* 69, 439-43.

109 Potrykus, K. and Cashel, M. (2008) (p)ppGpp: still magical? *Annu Rev Microbiol.* **62**, 35-51.

110 Pukkila, P.J., Peterson, J., Herman, G., Modrich, P. and Meselson, M. (1983) Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *Escherichia coli*. *Genetics.* **104**, 571-82.

111 Ratel, D., Ravanat, J.L., Berger, F. and Wion, D. (2006) N6-methyladenine: the other methylated base of DNA. *Bioessays.* **28**, 309-315.

112 Ringquist, S. and Smith, C.L. (1992) The *Escherichia coli* chromosome contains specific, unmethylated dam and dcm sites. *Proc. Natl. Acad. Sci.* USA. 89, 4539-43.

113 Robbins-Manke, J.L., Zdraveski, Z.Z., Marinus, M. and Essigmann, J.M. (2005) Analysis of global gene expression and double-strand-break formation in DNA adenine methyltransferase- and mismatch repair-deficient *Escherichia coli*. *J Bacteriol.* **187**, 7027-37.

114 Roberts, R.J., and Macelis, D. (2000) REBASE-restriction enzymes and methylases. *Nucleic Acids Res.* **28**, 306-307.

115 Rokeach, L.A. and Zyskind, J.W. (1986) RNA terminating within the *E. coli* origin of replication: stringent regulation and control by DnaA protein. *Cell.* **46**, 763-71.

116 Roodveldt, C. and Tawfik, D.S. (2005) Directed evolution of phosphotriesterase from *Pseudomonas diminuta* for heterologous expression in *Escherichia coli* results in stabilization of the metal-free state. *PEDS.* **18**, 51-58.

117 Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V. and Quackenbush, J. (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques.* **34**, 374-8.

118 Saint-Dic, D., Kehrl, J., Frushour, B. and Kahng, L.S. (2008) Excess SeqA leads to replication arrest and a cell division defect in *Vibrio cholerae*. *J Bacteriol.* **190**, 5870-8.

119 Schmidt-Dannert, C. (2001) Directed evolution of single proteins, metabolic pathways and viruses. *Biochem.* **40**, 13125-13136.

120 Schmid, A., Dordick, J.S., Hauer, B., Kiener, A., Wubbolts, M. and Witholt, B. (2001) Industrial biocatalysis today and tomorrow. *Nature.* **409**, 258-268.

121 Schoemaker, H.E., Mink, D. and Wubbolts, M.G. (2003) Dispelling the myths--biocatalysis in industrial synthesis. *Science.* **299**, 1694-7.

122 Schreiber, G., Ron, E.Z. and Glaser, G. (1995) ppGpp-mediated regulation of DNA replication and cell division in *Escherichia coli*.

123 Shiba, T., Tsutsumi, K., Ishige, K. and Noguchi, T. (2000) Inorganic polyphosphate and polyphosphate kinase: their novel biological functions and applications. *Biochemistry (Mosc).* **65**, 315-23.

124 Song, J.K. and Rhee, J.S. (2001) Enhancement of stability and activity of phospholipase A1 in organic solvents by directed evolution. *Biochim Biophys Acta.* **1547**, 370-378.

125 Stancheva, I., Koller, T. and Sogo, J.M. (1999) Asymmetry of Dam remethylation on the leading and lagging arms of plasmid replicative intermediates. *EMBO J.* **18**, 6542-51.

126 Stefan, A., Radeghieri, A., Gonzalez Vara y Rodriguez, A. and Hochkoeppler, A. (2001) Directed evolution of beta-galactosidase from *Escherichia coli* by mutator strains defective in the 3'-->5' exonuclease activity of DNA polymerase III. *FEBS Lett.* **493**, 139-43.

127 Stemmer, W.P. (1994) DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci*. USA. **91**, 10747-51.

128 Stemmer, W.P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature.* **370**, 389-91.

129 Stumpf, J.D. and Foster, P.L. (2005) Polyphosphate kinase regulates error-prone replication by DNA polymerase IV in *Escherichia coli*. *Mol Microbiol.* **57**, 751-61.

130 Suzuki, T., Yasugi, M., Arisaka, F., Yamagishi, A. and Oshima, T. (2001) Adapt**at**ion of a thermophilic enzyme**,** 3-isopropylmalate dehydrogenase**,** to low temperatures. *Protein Engineering.* **14**, 85-91.

131 Tavazoie, S. and Church, G.M. (1998) Quantitative whole-genome analysis of DNA-protein interactions by in vivo methylase protection in *E. coli*. *Nat Biotechnol.* **16**, 566-71.

132 Taylor, V. L., Titball, R. W. and Oyston, P. C. F. (2005) Oral immunization with a dam mutant of *Yersinia pseudotuberculosis* protects against plague. *Microbiology.* **151**, 1919-1926.

133 Tosa, T. and Pizer, L.I. (1971) Effect of serine hydroxamate on the growth of *Escherichia coli*. *J Bacteriol.* **106**, 966-71.

134 Umeno, D. and Arnold, H.F. (2004) Evolution of a pathway to novel long-chain carotenoids. *J Bacteriol.* **186**, 1531-1536.

135 Urig, S., Gowher, H., Hermann, A., Beck, C., Fatemi, M., Humeny, A. and Jeltsch, A. (2002) The *Escherichia coli* dam DNA methyltransferase modifies DNA in a highly processive reaction. *J. Mol. Biol.* **319**, 1085-96.

136 Van Den Berg, S., Lofdah, P.A., Hard, T. and Berglund, H. (2006) Improved solubility of TEV protease by directed evolution. *J Biotechnol.* **121**, 291-298.

137 van Kampen, M.D., Rosenstein, R., Götz, F. and Egmond, M.R. (2001) Cloning, purification and characterisation of *Staphylococcus warneri* lipase 2. *Biochim Biophys Acta.* **1544***,* 229-41.

138 Wang, T.C. and Smith, K.C. (1986) Inviability of dam recA and dam recB cells of *Escherichia coli* is correlated with their inability to repair DNA double-strand breaks produced by mismatch repair. *J Bacteriol.* **165**, 1023-5.

139 Wang, M.X. and Church, G.M. (1992) A whole genome approach to in vivo DNA-protein interactions in *E. coli*. *Nature.* **360**, 606-10.

140 Watson, M. E., Jarisch, J. and Smith, A. L. (2004) Inactivation of deoxyadenosine methyltransferase (dam) attenuates *Haemophilus influenzae* virulence. *Mol. Microbiol.* **55**, 651-654.

141 Williams, G. J., Nelson, A. S. and Berry, A. (2004) Directed evolution of enzymes for biocatalysis and the life sciences. *Cellul. Mol. Sci.* **61**, 3034-3046.

142 Wilson, G.G. (1992) Amino acid sequence arrangements of DNA methyltransferases. *Methods Enzymol*. **216,** 259-279.

143 Wion, D. and Casadesus, J. (2006) N6-methyladenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol*. **4,** 183-192.

144 Wright, R., Stephens, C., and Shapiro, L. (1997) The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its essential functions are conserved in *Rhizobium meliloti* and *Caulobacter crescentus*. *J Bacteriol.* **179**, 5869-5877.

145 Yamaki, H., Ohtsubo, E., Nagai, K. and Maeda, Y. (1988) The oriC unwinding by dam methylation in *Escherichia coli*. *Nucleic Acids Res.* **16**, 5067-73.

146 Yang, Z., Horton, J.R., Zhou, L., Zhang, X.J., Dong, A., Zhang, X., Schlagman, S.L., Kossykh, V., Hattman, S. and Cheng, X. (2003) Structure of the bacteriophage T4 DNA adenine methyltransferase. *Nature Struct. Biol.* **10**, 849-855.

147 Zhang, J.H., Dawes, G. and Stemmer, W.P. (1997) Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. *Proc. Natl. Acad. Sci*. USA. **94**, 4504-4509.

148 Zhao, H. and Arnold, F. H. (1997) Functional and non-functional mutations distinguished by random recombination of homologous genes. *Proc. Natl. Acad. Sci*. USA. **94**, 7997-8000.

149 Zhao, H. and Arnold, F. H. (1997) Optimization of DNA Shuffling for High Fidelity Recombination. *Nucleic Acids Res.* **25**, 1307-1308.

150 Zhao, H., Giver, L., Shao, Z., Affholter, J. A. and Arnold, F. H. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat. Biotechnol.* **16**, 258-261.

151 Zhao, H. and Arnold, F.H. (1999) Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng.* **12**, 47–52.

152 Zhao, J., Niu, W., Yao, J., Mohr, S., Marcotte, E.M. and Lambowitz, A.M. (2008) Group II intron protein localization and insertion sites are affected by polyphosphate. *PLoS Biol.* **6**, e150.

6 Appendix

Chapter 1: Changing the DNA recognition specificity of E.coli DNA methyltransferase (EcoDam) enzyme by directed evolution.

The article has been removed due to copyright issues.

Chapter 2: Two alternative conformations of *S*-Adenosyl-L-homocysteine bound to *Escherichia coli* DNA adenine methyltransferase and the implication of conformational changes in regulating the catalytic cycle

The article has been removed due to copyright issues.

Chapter 3: Transition from EcoDam to T4Dam DNA recognition mechanism without loss of activity and specificity

The article has been removed due to copyright issues.

Chapter 4:

Author contribution description

1 Changing the DNA recognition specificity of *E.coli* DNA methyltransferase (EcoDam) enzyme by directed evolution.

Sanjay Chahar (S. C.) has contributed to design of the study and has performed following experiments: generation of mutant libraries, screening, purification and kinetic characterization of wildtype EcoDam and its mutants identified in different generations, as well as participated in data interpretation and writing of manuscript. In addition, S. C. was involved in supervision of students working in the project.

2 Two alternative conformations of *S*-Adenosyl-L-homocysteine bound to *Escherichia coli* DNA adenine methyltransferase and the implication of conformational changes in regulating the catalytic cycle.

S. C. has performed following experiments: purification of wildtype EcoDam and its K14A mutant protein, steady-state methylation experiments using varying concentration of AdoMet and measurement of Km and Kcat values. S. C. has also studied the effect of N- and C-terminal his-tag fusion in the identification of recognition of first base pair recognition sequence (GATC) by EcoDam.

3 Transition from EcoDam to T4Dam DNA recognition mechanism without loss of activity and specificity.

S. C. has contributed to the design of the study, the planning of the experiments and discussion of results.

4 Kinetic investigation of the mechanism of action of inhibitor-58 on EcoDam

S. C. has performed following experiments: purification of EcoDam protein, and kinetic experiments to investigate the mechanism of inhibition of inhibitor-58 on EcoDam.

5 Role of EcoDam in gene transcription regulation

S. C. has contributed to the design of study and performed following experiments: micro array experiments, cloning of all constructs, expression analysis by Real-time PCR and CFP fluorescence measurements, mutational analysis of promoters, as well as participated in data interpretation and writing of manuscript.

6 Role of EcoDam in bridging metabolic signals to epigenetic modifications

S. C. has contributed to the design of study and performed following experiments: generation of EcoDam-GFP fusion construct, microscopic analysis of EcoDam-GFP fusion protein inside wildtype and all mutant strains, serine hydroxamate experiment, as well as participated in data interpretation and writing of manuscript.

Chapter 5: Supplementary table

Supplementary table 1: List of expressed genes in micro array experiment (wt/control).

S.No.	Gene blattner No.	Gene name	Expression ration (wt/control)	p-value
1	b3242	aaeX	0,843558	0,016416
2	b1336	abgT	0,854453	0,031662
3	b2316	accD	0,914002	0,047749
4	b4069	acs	1,730283	0,012662
5	b3134	agaW	0,787917	0,029786
6	b0606	ahpF	1,258587	0,031145
7	b0516	allC	1,551233	0,041031
8	b4087	alsA	0,576159	0,046522
9	b2957	ansB	0,503588	0,006513
10	b4055	aphA	1,356754	0,027272
11	b3210	arcB	0,728133	0,024431
12	b1721	arpB_2	2,248844	0,029492
13	b2221	atoD	1,230082	0,006091
14	b3737	atpE	1,454872	0,049644
15	b1228	b1228	0,429582	0,033807
16	b1995	b1995	1,546087	0,016327
17	b2078	baeS	0,81407	0,043576
18	b3537	bcsF	1,080719	0,013554
19	b3337	bfd	2,574596	0,022878
20	b2132	bglX	1,408835	0,037139
21	b0774	bioA	2,789568	0,029726
22	b1987	cbl	0,577334	0,006783
23	b3607	cysE	1,325559	0,026335
24	b3934	cytR	0,667599	0,00321
25	b0621	dcuC	0,246841	0,046031
26	b3639	dfp	1,276494	0,003558
27	b2133	dld	1,389276	0,026749
28	b3540	dppF	0,687369	0,017697
29	b2140	dusC	0,61674	0,01466
30	b2595	ecfD	0,92964	0,048365
31	b2685	emrA	1,375907	0,047336
32	b0596	entA	1,766215	0,016469
33	b2370	evgS	0,742358	0,028985
34	b4079	fdhF	0,52402	0,014648
35	b0518	fdrA	1,367567	0,035495
36	b4290	fecB	2,690966	0,015204
37	b0592	fepB	3,588796	0,039551
38	b0229	fhiA	0,523347	0,010807
39	b1948	fliP	1,135354	0,002009
40	b4152	frdC	0,39395	0,038985
41	b4151	frdD	0,410707	0,007655
42	b3844	fre	0,865652	0,037719
43	b3946	fsaB	1,433954	0,023599
44	b0084	ftsI	1,262582	0,029392
45	b3127	garP	0,471893	0,011644
46	b2904	gcvH	0,328295	0,008394
47	b0809	glnQ	0,628536	0,012906

48	b2241	glpA	0,23043	0,027573
49	b2243	glpC	0,36145	0,043037
50	b3426	glpD	0,297834	0,035704
51	b3925	glpX	0,338012	0,022715
52	b0654	gltJ	0,514013	0,034857
53	b0962	helD	0,748872	0,047327
54	b4175	hflC	1,549622	0,038208
55	b1507	hipA	0,831674	0,034448
56	b1508	hipB	0,843038	0,031475
57	b2022	hisB	0,69319	0,036226
58	b2514	hisS	1,666786	0,001791
59	b0161	htrA	1,242264	0,035229
60	b0139	htrE	1,574154	0,004778
61	b2723	hycC	0,59614	0,045981
62	b3686	ibpB	1,920045	0,020342
63	b4018	iclR	1,275782	0,048126
64	b3769	ilvM	1,894493	0,034015
65	b2528	iscA	1,118812	0,029237
66	b0186	ldcC	0,8403	0,007995
67	b4033	malF	0,687131	0,039839
68	b1620	malI	1,379482	0,014862
69	b1621	malX	1,259806	0,0052
70	b1530	marR	1,472555	0,03887
71	b4345	mcrC	2,007409	0,012925
72	b1047	mdoC	2,004737	0,038945
73	b1663	mdtK	0,716545	0,038618
74	b0732	mngB	0,725513	0,029891
75	b2392	mntH	1,292689	0,006699
76	b4233	mpl	1,500625	0,008356
77	b3068	mug	0,703093	0,021657
78	b0085	murE	0,701855	0,04613
79	b2831	mutH	1,320133	0,003008
80	b2205	napG	1,325032	0,003671
81	b1227	narI	0,405378	0,015328
82	b1226	narJ	0,484072	0,04301
83	b4237	nrdG	1,635251	0,00797
84	b1633	nth	0,743579	0,025478
85	b2281	nuoI	1,210966	0,011133
86	b2743	pcm	1,297186	0,040778
87	b3951	pflD	1,34399	0,003651
88	b4099	phnI	2,116893	0,020372
89	b1090	plsX	1,143035	0,020488
90	b1702	ppsA	0,752322	0,045558
91	b0467	priC	0,529628	0,001126
92	b2585	pssA	1,2612	0,013631
93	b1849	purT	0,630163	0,022553
94	b3629	rfaS	1,484301	0,003129
95	b2608	rimM	0,798086	0,014146
96	b0953	rmf	0,764277	0,024666
97	b3984	rplA	1,228871	0,006831
98	b3983	rplK	1,278209	0,02401
99	b2606	rplS	1,534087	0,036366
100	b3636	rpmG	1,611678	0,014972

101	b3303	rpsE	0,838116	0,03118
102	b2144	sanA	1,482234	0,000564
103	b2796	sdaC	2,361938	0,039744
104	b0098	secA	2,171689	0,044519
105	b3591	selA	0,559115	0,01052
106	b2703	srlE	0,279832	0,036208
107	b2744	surE	1,113275	0,038249
108	b1886	tar	0,800983	0,036099
109	b3119	tdcR	1,144212	0,035472
110	b0436	tig	1,814706	0,004429
111	b1252	tonB	1,717583	0,014363
112	b4191	ulaR	1,07602	0,042242
113	b4324	uxuR	1,306078	0,003692
114	b4258	valS	1,525378	0,013051
115	b3630	waaP	1,567946	0,025484
116	b1749	xthA	0,646871	0,007714
117	b0251	yafY	0,691003	0,025119
118	b0293	yagZ	1,375238	0,041655
119	b0354	yaiL	1,24527	0,041407
120	b0489	ybbK	1,077863	0,022444
121	b0490	ybbL	0,836951	0,030361
122	b0513	ybbY	0,509326	0,000578
123	b0735	ybgE	1,361133	0,006914
124	b0813	ybiF	0,881815	0,000826
125	b0868	ybjS	1,0367	0,028894
126	b0916	ycaQ	2,037275	0,03844
127	b0941	ycbT	1,523048	0,032655
128	b0946	ycbW	1,40354	0,035512
129	b1013	ycdC	0,803763	0,048735
130	b1005	ycdF	0,74934	0,043422
131	b1016	ycdN	1,100882	0,038934
132	b1036	ycdZ	1,761959	0,043402
133	b1055	yceA	1,064019	0,008305
134	b1111	ycfQ	0,83804	0,026986
135	b1321	ycjX	0,794922	0,002518
136	b1426	ydcH	0,829354	0,003413
137	b1477	yddM	0,54278	0,038603
138	b1634	ydgR	1,682702	0,028968
139	b1643	ydhI	0,500096	0,013297
140	b1706	ydiU	0,445006	0,006479
141	b1847	yebF	0,532297	0,00571
142	b1867	yecD	1,131784	0,047175
143	b1931	yedK	0,719031	0,039035
144	b2005	yeeV	0,515573	0,026758
145	b2016	yeeZ	0,886575	0,039873
146	b2085	yegR	0,663203	0,019236
147	b2100	yegV	0,608452	0,045739
148	b2120	yehM	0,420215	0,002257
149	b2290	yfbQ	1,226043	0,048365
150	b2327	yfcA	1,117113	0,017781
151	b2333	yfcP	0,415606	0,042677
152	b2430	yfeW	1,689084	0,006282
153	b2593	yfiH	1,227494	0,030974

154	b2625	yfjI	0,805474	0,020177
155	b2630	yfjN	0,736801	0,017019
156	b2644	yfjY	1,955499	0,012629
157	b2645	yfjZ	0,772825	0,027375
158	b2853	ygeI	0,456383	0,045276
159	b2920	ygfH	3,553213	0,008407
160	b3157	yhbT	0,645925	0,014965
161	b3525	yhjH	0,797041	0,029546
162	b3539	yhjV	1,780148	0,044107
163	b3705	yidC	1,702772	0,006365
164	b3755	yieP	0,749269	0,03819
165	b4065	yjcE	2,98077	0,046942
166	b4110	yjcZ	1,532222	0,00455
167	b4130	yjdL	0,4279	0,004614
168	b4185	yjfM	0,939171	0,028142
169	b4189	yjfO	0,813555	0,014656
170	b4277	yjgZ	1,891974	0,001948
171	b4306	yjhP	0,577404	0,040673
172	b4332	yjiJ	1,278964	0,004201
173	b4354	yjiY	6,683928	0,036029
174	b4363	yjjB	0,826144	0,026688
175	b4379	yjjW	0,794598	0,039609
176	b0392	ykiA	2,141203	0,020686
177	b1527	yneK	0,804765	0,032211
178	b1583	ynfB	0,750075	0,007488
179	b1585	ynfC	0,789946	0,00739
180	b1757	ynjE	0,558182	0,039528
181	b1760	ynjH	1,02542	0,027318
182	b2376	ypdI	1,129645	0,037399
183	b2845	yqeG	1,37948	0,032647
184	b3047	yqiH	0,65556	0,010365
185	b3050	yqiJ	1,189024	0,049226
186	b3098	yqjD	0,646807	0,042422
187	b4210	ytfF	1,229751	0,029433