INTRODUCTION

The original research plan envisaged new approaches to oligonucleotide-mediated targeted DNA methylation. The key elements of the research plan were the following:

1. Use of triplex-forming-oligonucleotides (TFO) as targeting moduls for targeted DNA methylation
2. Enzymatic (SNAP-tag mediated) coupling of CG-specific DNA(cytosine-5)-methyltransferases (C5-MTase) to the TFO
3. Use of mutant versions and complementing fragments of the CG-specific C5-MTase M.SssI

During the work the original research plan was modified. The changes were introduced to solve unexpected problems or to explore new research directions:

a) The M.SssI-SNAP-TFO conjugates had low solubility and could not be purified in sufficient quantity. Fortunately, we could find another CG-specific MTase (M.MpeI), whose TFO conjugates proved more soluble.
b) Another obstacle was the instability and poor solubility of MTase fragments. Our efforts to circumvent these problems led to the discovery that circularly permuted variants of M.SssI and M.MpeI have MTase activity.
c) We developed a simple *E. coli* genetic assay for detecting sequence-specific DNA-protein interactions.

The work related to the TFO-guided DNA methylation was carried out in collaboration with Dr. Elmar Weinhold (RWTH Aachen University, Aachen, Germany), with whom we originally applied for the ERA Chemistry grant.

MAIN RESULTS

TFO-SNAP-M.SssI fusions

The SNAP-tag is a modified version of a guanine-alkyltransferase, which can transfer a variety of molecules onto itself if the molecule carries a benzylguanine moiety. We constructed plasmids that express, in an IPTG-inducible fashion, SNAP-M.SssI-His6 fusion proteins in *E. coli*. To explore

ways to increase targeting specificity, different variants of the SNAP-M.SssI-His6 hybrid protein were tested. The proteins varied in the type of the M.SssI part (wild-type, T313H, Q147L and C141S mutants). The T313H, Q147L substitutions were known to confer decreased DNA binding affinity, whereas the mutant with the C141S replacement had very low activity but wild-type DNA binding affinity (Ślaska-Kiss et al, 2012). The chimeric proteins retained the activities of the fusion partners (i.e. MTase and SNAP-tag activity). The plasmids expressing the SNAP-M.SssI fusion proteins from the T7 phage promoter were stable in the *E. coli* host ScarabXpress T7lac. The use of this strain turned out to be very important, because other *E. coli* host strains producing T7 RNA polymerase were not viable when transformed with plasmids expressing M.SssI from the T7 promoter. Production of soluble chimeric proteins required growing of the culture at room temperature. The overproduced SNAP-M.SssI-His6 (wild-type) enzyme was fairly soluble. Two other variants (SNAP-M.SssI(C1541S)-His6 and SNAP-M.SssI(Q147L)-His6) were less soluble, we had to add 1% TritonX -100 and higher concentration (0.3 M) of NaCl to the extraction buffer to keep the proteins in solution. We developed purification procedures involving a single affinity chromatography step on Ni-agarose column. The yield was 2 mg ~90% pure protein from 400 ml culture. The SNAP-M.SssI(T3131H)-His6 variant was produced in much lower amounts than the other variants, and purification of the protein was not successful.

Our collaborating colleagues in Aachen synthesized the benzylguanine-PEG_{35}-TFO, and optimized conditions of the coupling reaction with SNAP-M.SssI-His6. Unfortunately, the TFO-SNAP-M.SssI-His6 conjugates had low solubility and could not be purified in sufficient amounts for further work.

TFO-SNAP-M.MpeI fusions

Fortunately, the deadlock caused by the physicochemical properties of M.SssI could be solved by using another CG-specific C5-MTase: M.MpeI. Like M.SssI, M.MpeI is a prokaryotic C5-MTase sharing the specificity of mammalian DNA methyltransferases: CG/CG → ^{m5}CG/^{m5}CG. Our attention was drawn to M.MpeI by a paper describing the X-ray crystal structure of the M.MpeI-DNA complex (Wojciechowski et al., 2013). The fact that, in contrast to M.SssI, M.MpeI could be crystallized, suggested to us that M.MpeI might be more soluble than M.SssI, and thus could be

easier to work with. We constructed plasmids expressing SNAP-M.MpeI-His6 or SNAP-M.MpeI(Q142L)-His6 and purified the hybrid proteins. The substitution Q142L in M.MpeI corresponds to Q147L in M.SssI and presumably leads to similar phenotype.

A single affinity chromatography step on Ni-agarose yielded ~1 mg of ~90% pure protein of the WT as well as of the Q142L variant from 100 ml IPTG-induced culture. The purified SNAP-M.MpeI-His6 proteins were shown to have SNAP-tag as well as MTase activity. Coupling of the purified proteins to TFO worked well, and the TFO-SNAP-M.MpeI-His6 conjugates were soluble and could be purified free of other components of the coupling reaction.

Specificity of targeted DNA methylation by the TFO-SNAP-M.MpeI conjugate was tested *in vitro* using a plasmid (Litcon54), which had four Psp1406I recognitions sites (5'-AACGTT/5'-AACGTT). Psp1406I was selected to assay targeted DNA methylation because CG-specific methylation (AA^{m5}CGTT/AA^{m5}CGTT) was known to block Psp1406I cleavage. The plasmid was engineered to contain the TFO binding site 5'-TTTTCTCTCTTTTTTTTTTTTTT in the vicinity of one of the Psp1406I sites. Because C5-MTases methylate palindromic substrate sites in two independent binding events involving different binding orientations, it seemed likely that the MTase bound to the DNA *via* the targeting domain can, for steric reasons, methylate only one strand of closely located substrate sites (^{m5}CG/CG). Therefore, it was important to test whether hemimethylation, *i.e*. methylation of only one of the strands, is sufficient to block Psp1406I cleavage. We showed that CG-specifically hemimethylated Psp1406I sites (AA^{m5}CGTT/AACGTT) are resistant to Psp1406I digestion, thus the assay system was suitable for detecting methylation of the CG site located in close distance to the TFO binding site.

Using this assay system we could show that the TFO-SNAP-M.MpeI conjugate preferentially methylated the addressed CG site, i.e. the CG within the Psp1406I recognition site located adjacently to the TFO binding site. Interestingly, the level of methylation specificity (the amount of partially digested fragments arising from protection of the addressed site *vs.* the amount of partially digested fragments resulting from off-target methylation) depended on the plasmid conformation: the specificity was the highest with linear plasmid, lower with relaxed circular plasmid, and the lowest with supercoiled plasmid. Because due to conformational constraints the average distance in space between individual CG sites on the same plasmid

molecule decreases in the order: linear > open circular > supercoiled, it could be concluded that non-addressed CG sites were methylated by MTase molecules bound at the addressed CG site but reaching distant, non-addressed CG sites. The unacceptable level of off-target methylation has been a general problem in different approaches of targeted DNA methylation. The underlying mechanism has not been investigated, non-addressed methylation was thought to be mainly caused by free enzyme molecules, which are not bound to the binding site of the targeting domain. Our results suggest that off-target methylation can also occur by correctly bound MTase molecules. This mechanism is likely to be valid for all types of targeting modules not just for TFO. A manuscript reporting these results (Giesbertz, A., Szentes, S., Kiss, A. and Weinhold, E., title pending) is in preparation. Testing of the TFO-SNAP-M.MpeI conjugates in mammalian cell cultures in the lab of Dr. Marianne Rots (Groningen) is in progress.

Complementing M.SssI and M.MpeI fragments for the TFO approach

We had previously shown that certain truncated, inactive fragments of M.SssI can functionally complement each other, i.e. they can form active enzyme when synthesized together in the same *E. coli* cell. During the course of the present project we demonstrated that also M.MpeI has the capacity of fragment complementation: the fragment combinations Mpe[1-241] + Mpe[246-395] and Mpe[1-309] + Mpe[246-395] produced detectable MTase activity *in vivo*. The numbers in square brackets indicate the first and the last amino acids of the fragment.

The goal of this subproject was to try to exploit the phenomenon of MTase fragment complementation (Ślaska-Kiss *e*t al., 2012; Chaikind and Ostermeier, 2014) for TFO-mediated targeted DNA methylation. Because the TFO-approach requires large amount of purified protein, a prerequisite of any further work was to find complementation-proficient M.SssI or M.MpeI fragments, which are soluble and can be purified. First we tried two N-terminal M.SssI fragments (Sss[1-239] and Sss[1-304]), which were known to have high complementation efficiency with appropriate C-terminal partners (Ślaska-Kiss *et* al., 2012). Plasmids expressing the fusion proteins His6-SNAP-Sss[1-239] or His6-SNAP-Sss[1-304] from the arabinose-inducible P_{BAD} promoter were created. His6-SNAP-Sss[1-239] was produced in large amount but was insoluble, whereas

His6-SNAP-Sss[1-304] was highly soluble and could be purified by affinity chromatography on Ni-agarose (Table 1).

In the envisaged strategy only the N-terminal fragment would carry the TFO-SNAP-tag, the complementing C-terminal fragment would not be coupled to TFO. Unfortunately, the well-complementing C-terminal fragments such as Sss[241-386] and Mpe[246-395] had low solubility, and could not be purified. We tried to increase solubility of the Sss[241-386] and Mpe[246-395] fragments by cloning the corresponding gene segments in plasmid vectors designed to add solubility-enhancing tags to the C-terminus of the protein of interest. Three vectors were tried (all from Addgene): pET-Trx-His6-LIC, pET-mOCR-His6-LIC and pET-MBP-His6-LIC. Phenotypes of the fusion proteins are shown in Table 1.

protein	N or C[1]	properties
His6-SNAP-Sss[1-239]	N	large amount but insoluble
His6-SNAP-Sss[1-304]	N	large amount, soluble, could be purified
Sss[241-386]-Trx-His6	C	large amount, but insoluble
Sss[241-386]-mOCR-His6	C	undetectable on SDS gel
Sss[241-386]-MBP-His6	C	substantial amount, soluble, could be purified
Mpe[1-241] (unfused)	N	undetectable on SDS gel
His6-SNAP-Mpe[1-309]	N	detectable on SDS gel, soluble
Mpe[246-395]His6:	C	medium amount, insoluble
Mpe[246-395]Trx-His6:	C	medium amount, insoluble
Mpe[246-395]MBP-His6:	C	large amount, insoluble

Table 1. Production and solubility of M.SssI and M.MpeI fragments fused to different tags. [1]N and C designate N- and C-terminal fragments, respectively. Trx stands for thioredoxin and MBP for maltose binding protein.

When we combined *in vitro* the N- and C-terminal M.SssI fragments, which could be purified (His6-SNAP-Sss[1-304] + Sss[241-386]-MBP-His6), we could detect weak MTase activity.

Circularly permuted variants of M.SssI and M.MpeI

This subproject was initiated as an effort to solve the solubility problem we had with C-terminal fragments (see above). As most C5-MTases, M.MpeI and M.SssI consist of a single polypeptide chain, which folds into a large and a small domain comprising approximately the N-terminal two-third and the C-terminal one-third of the amino acid sequence, respectively (Wojciechowski et al., 2013; Koudan et al, 2004). We hypothesized that poor solubility of the C-terminal fragments could be the result of displacement of the hydrophobic C-terminal α-helix. In the native enzyme this α-helix folds back into the large domain, whereas in a separately expressed C-terminal fragment, where it lacks its natural environment, is probably exposed to the solvent. We thought that this problem could be circumvented by using fragments obtained from circularly permuted (CP) MTase variants, in which the C-terminal α-helix is covalently linked to the N-terminus of the enzyme. The proximity of the N- and C-termini in the structural models of M.MpeI and M.SssI (Wojciechowski et al., 2013; Koudan et al, 2004) suggested that covalently linking the native termini would not grossly distort the structure of the enzymes, and construction of enzymatically active circular permutants could be possible.

We constructed plasmids expressing CP variants of M.MpeI and M.SssI. Eleven of the fourteen cpM.MpeI variants and six of the seven cpM.SssI variants had detectable MTase activity. Based on CP topology, the artificially created CP variants showing MTase activity can be classified in ten types. The numbers in the names of the permutants indicate the new N-terminal amino acid (Figure 1).

This is the first study describing artificially designed, circularly permuted variants of C5-MTases. Although the MTase activity shown by most CP variants was less than 5% of the WT level, our results provide new experimental evidence for the remarkable structural plasticity of C5-MTases. We expected that permuting the enzymes between the two domains would cause the least perturbation to the native structure, thus Type H permutants (cp245/280M.MpeI and

cp243M.SssI) would have the highest activity. Surprisingly, Type B permutants (cp62M.MpeI and cp58M.SssI, whose N-terminus is between the weakly conserved motifs II and III, Figure 1), were by far the most active (~50% of the WT).

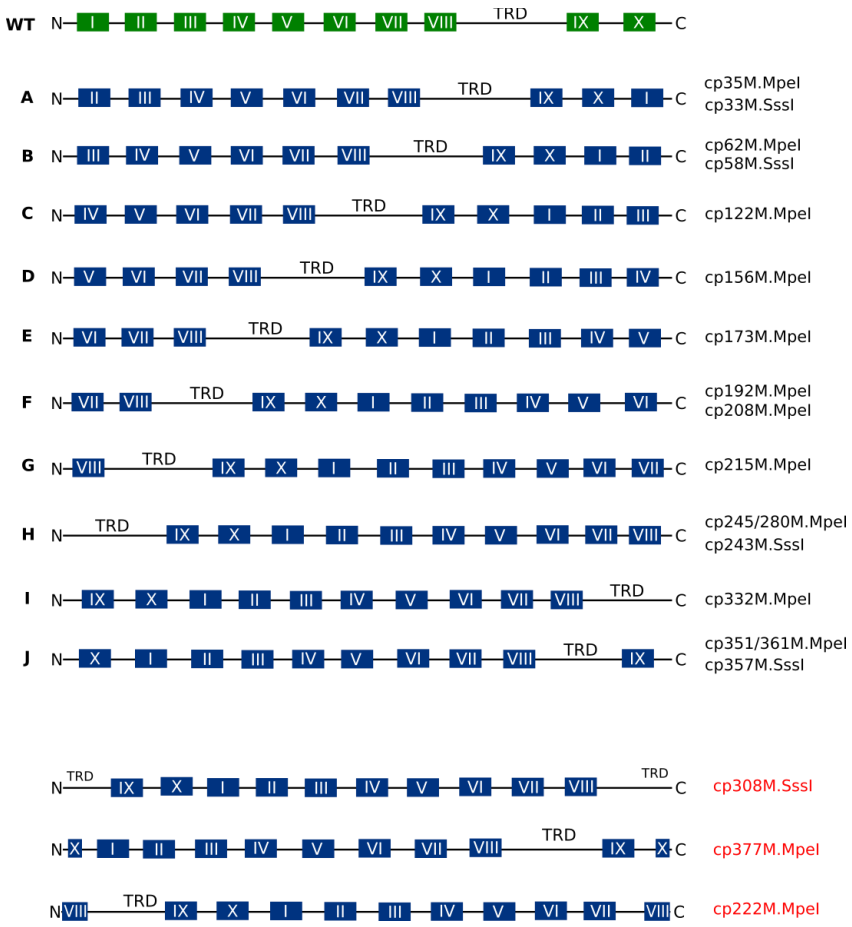

Figure 1: Schematic representation of the circularly permuted MTase variants constructed in this work. Conserved motifs characteristic for C5-MTases are marked with roman numerals. TRD, Target Recognizing Domain. The variants lacking MTase activity are in the bottom of the figure and are shown with names in red.

Type H permutation, in which the new amino-end is between motif VIII and the TRD, has already been observed in five natural C5-MTases: M.BssHII, M.Alw26I, M2.Eco31IC, M.Esp3I and M2.BsaI. To see if there are natural circularly permuted C5-MTases having permutation topology

different from Type H, we performed a computer search of C5-MTase sequences deposited in the REBASE database (http://rebase.neb.com/rebase/). The search found 28 C5-MTase sequences satisfying the criterion of motif X preceding motif I. These included the five enzymes already known to have circularly permuted sequence (see above) and 23 new enzymes. The latter group contains four enzymes (M2.BcoDI, M.BsmAI, M2.EcoMI and M.BsmBI), for which there is biochemical evidence showing that the reaction product is C5-methylcytosine. The other nineteen enzymes are putative C5-MTases identified only by the presence of the characteristic sequence motifs. Amino acid sequence alignments showed that all natural circularly permuted C5-MTases found previously or in our search have the Type H arrangement, indicating that permutation leaving the two domains intact is favored by Nature.

Complementation between fragments of circularly permuted M.MpeI and M.SssI

This work was started on the hypothesis that poor solubility of C-terminal fragments of M.MpeI and M.SssI was due to the exposure of the C-terminal α-helix to the solvent, and that the native fold could be restored by circularly permuting the amino acid sequence so that the originally C-terminal α-helix becomes covalently linked to the large domain. For obtaining fragments with the desired sequence arrangement, Type J permutants (cp351MmMpeI, cp361M.MpeI and cp357M.SssI) starting with conserved motif X (Figure 1) appeared to be optimal, because splitting them between motif VIII and the TRD would create two polypeptides, which approximately correspond to the two domains. Complementation capacity was tested by cloning the corresponding gene segments in compatible expression plasmid vectors, introducing both plasmids into the same cell, and testing the methylation status of the plasmid preparations isolated from induced cultures. Contrary to expectations, neither the Mpe[361-244] + Mpe[245-360] nor the Sss[357-242] + Sss[243-356] fragment combinations showed MTase activity.

After the failure with Type J, we tried Type B permutants for complementing fragments. Of all permutants, Type B enzymes (cp62M.MpeI and cp58M.SssI, Figure 1), showed the highest MTase activity. To test the complementation potential of fragments prepared from cp62M.MpeI, two plasmid pairs were constructed. The plasmids pB-Mpe[192-61] and pOB-Mpe[62-191] expressed fragments split at H192 in conserved motif VI, whereas the plasmids pB-Mpe[280-61]

and pOB-Mpe[62-279] expressed fragments split at E280, in the region between motif VIII and the TRD. The fragment combination Mpe[192-61] plus Mpe[62-191] did not show complementation. In contrast, the fragments Mpe[280-61] plus Mpe[62-279] resulted in partial methylation of the plasmid preparation indicating that the Mpe[280-61] and Mpe[62-279] peptides derived from cp62M.MpeI formed active MTase.

The complementation capacity of the Class B permutant version of M.SssI (cp58M.SssI) was tested in similar experiments. The permutation site of cp58M.SssI exactly corresponds to that of cp62M.MpeI (Figure 1). Two bisection sites were tested, both were designed to fall between motif VIII and the TRD. The split site at V243 determined the fragments Sss[58-242] and Sss[243-57], whereas the one at N276 determined Sss[58-275] and Sss[276-57]. When tested for complementation in *E. coli*, the fragment pair Sss[58-242] + Sss[243-57] was found to be inactive, whereas the Sss[58-275] + Sss[276-57] fragment pair showed weak MTase activity.

The experiments described above analyzed complementation ability of fragments, which were precise cleavage products of the parental cpMTase without gap or extraneous amino acids. Interestingly, wild-type M.SssI and M.MpeI showed efficient complementation even if the fragments had long overlapping segments (Ślaska-Kiss *et* al., 2012 and our unpublished observations with M.MpeI). To test whether circularly permuted variants have the same capacity, the plasmids pB-Mpe[192-61] and pOB-Mpe[62-279], which express fragments with a 87 bp overlap were assayed for complementation. The plasmid DNA isolated from arabinose-induced double-transformants showed similar level of protection against Hin6I digestion as the plasmid preparation purified from cells expressing the precisely cleaved Mpe[280-61] + Mpe[62-279] fragments. Most importantly, the separately produced Mpe[192-61] and Mpe[62-279] fragments were soluble.

A poster describing results obtained with circularly permuted C5-MTases was presented at the Hungarian Molecular Life Sciences meeting (Eger, 2017), and a manuscript (Albert, P., Varga, B, Zsibrita, N. and Kiss, A., Circularly permuted variants of two CG specific prokaryotic DNA methyltransferases) is almost ready for submission.

A simple E. coli genetic assay for detecting sequence-specific DNA-protein interactions

During our work with different approaches to targeted DNA methylation we have often felt the need of being able to test whether a protein can bind to a specific DNA sequence or not. The published methods are complicated, either require purification of the protein of interest, or involve the construction of fusion proteins.

We wished to develop a technique, in which binding of the protein of interest to a specific DNA sequence leads to a measurable effect in *E. coli*. The envisaged assay exploits the well-characterized regulatory mechanism of the *E. coli lac* operon (Figure 2). The assay utilizes two plasmids. One of the plasmids carries the gene of the Lac repressor (*lacI*) with the recognition sequence of the tested protein inserted into the *lacI* promoter. The target protein is expressed, in an inducible fashion, from a compatible plasmid. The two plasmids are co-transformed into a LacI deficient *E. coli* host. It is assumed that if the target protein binds to the *lacI* promoter, it will interfere with transcription of the *lacI* gene, leading to induction of the *lac* operon and production of β-galactosidase, whose activity can be easily measured.

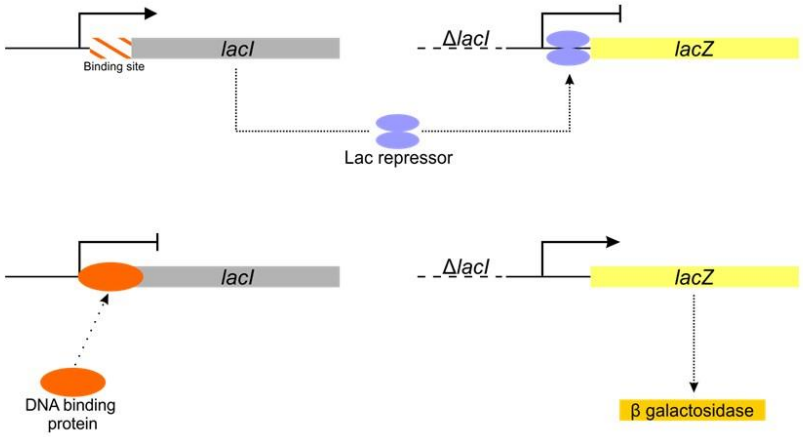

Figure 2. Concept of the *in vivo* DNA binding assay.

The assay was tested with zinc finger proteins and CRISPR-dCas9 recognizing specific sequences, but is expected to be generally applicable to all sequence-specific DNA binding proteins for which the target site is known. A poster (Sarolta Szentes, Eszter Zsigmond, Pál Salamon and Antal Kiss: A simple genetic system to study sequence-pecific DNA-binding proteins in *E. coli*) was presented at the Hungarian Molecular Life Sciences Meeting 2017, in Eger, and a manuscript with the same title and authors is in preparation.

REFERENCES

Ślaska-Kiss, K., Tímár, E. and Kiss, A. (2012) Complementation between inactive fragments of SssI DNA methyltransferase. BMC Molecular Biology, 13: 17.

Chaikind, B. and Ostermeier, M. (2014) Directed Evolution of Improved Zinc Finger Methyltransferases. *PLoS ONE*, 9: e96931.

Darii, M.V., Cherepanova, N.A., Subach, O.M., Kirsanova, O.V., Rasko, T., Ślaska-Kiss, K., Kiss, A., Deville-Bonne, D., Reboud-Ravaux, M. and Gromova, E.S. (2009) Mutational analysis of the CG recognizing DNA methyltransferase SssI: Insight into enzyme-DNA interactions. *Biochimica et Biophysica Acta*, 1794: 1654-1662.

Koudan, E.V., Bujnicki, J.M., Gromova, E.S. (2004) Homology modeling of the CG-specific DNA methyltransferase SssI and its complexes with DNA and AdoHcy. *J Biomol Struct Dyn,* 22: 339-345.

Wojciechowski, M., Czapinska, H. and Bochtler, M. (2013) CpG underrepresentation and the bacterial CpG-specific DNA methyltransferase M.MpeI. *Proc. Natl. Acad. Sci. USA*, 110: 105-110.