Triple helix-targeted DNA methylation with DNA methyltransferase-oligodeoxynucleotide conjugates

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades einer Doktorin der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Anna Giesbertz (M.Sc.)

aus

Voerendaal, Niederlande

Berichter*:* *Prof. Dr. Elmar Weinhold*

Prof. Dr. Antal Kiss

Tag der mündlichen Prüfung*:* *09.11.2018*

Table of Contents

Abbreviations

A	adenine
AdoHcy	*S*-adenosyl-L-homocysteine
AdoMet	*S*-adenosyl-L-methionine
Amp	ampicillin
BARAC	biarylazacyclooctynone
BG	benzylguanine
Bp	base pair(s)
C	cytosine
Cas	CRISPR associated
CRISPR	clustered regularly interspaced short palindromic repeats
CuAAC	copper-catalyzed alkyne azide cycloaddition
Cv	column volume(s)
DBCO	dibenzoazacyclooctyne
dCas9	dead Cas9
DEAE	diethylaminoethyl
DIBO	dibenzocyclooctyne
DIFO	difluorinated cyclooctyne
DIMAC	dimethoxyazacyclooctyne
DNA	deoxyribonucleic acid
Ds	double-stranded
E.coli	*Escherichia coli*
EpCAM	epithelial cell adhesion molecule
eq.	equivalent(s)
ESI	electrospray ionization
FPLC	fast protein liquid chromatography
G	guanine
H	hour(s)
hAGT	human *O*6-alkylguanine DNA alkyltransferase

Hp	*N*-methylhydroxypyrrole
HPLC	high performance liquid chromatography
Im	*N*-methylimidazole
k_{app}	apparent catalytic rate constant
LB	lysogeny broth
LNA	locked nucleic acid
Min	minute(s)
MS	mass spectrometry
MTase	methyltransferase
NTA	nitriloacetic acid
OD	optical density
ODN	oligodeoxynucleotide
OVD	over digest
PCQ	prolin-cysteine-glutamine
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
pI	isoelectrical point
POI	protein of interest
Py	*N*-methylpyrrole
REase	restriction endonuclease
RNA	ribonucleic acid
Rp	reversed-phase
Rpm	rotations per minute
Rt	room temperature
S	second(s)
SDS-PAGE	sodium dodecylsulfate polyacrylamide gel electrophoresis
sgRNA	single guide RNA
SPAAC	strain-promoted alkyne azide cycloaddition
T	thymine
TALE	transcription activator-like effector

TAMRA	5(6)-carboxytetramethylrhodamine
TB	terrific broth
TBE	TRIS/borate/EDTA
TE	TRIS/EDTA
TFO	triple helix-forming oligodeoxynucleotide
TFS	triple helix-forming site
t_R	retention time
U	unit(s)
UV	ultraviolet light
Vis	visible light

Three and one letter codes of amino acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1 Introduction

1.1 DNA

Nature's elegant solution for storing the genetic information of an organism and passing this information on to next generations lies in the structure of DNA. DNA, short for deoxyribonucleic acid, is a large biopolymer that typically consists of two complementary nucleotide strands composed of a hydrophilic phosphate sugar backbone and one of four different nucleobases adenine (A), cytosine (C), thymine (T) and guanine (G), connected to each sugar by a β-glycosidic bond. The nucleobases of one strand form hydrogen bonds with the nucleobases of the other strand according to Watson-Crick base pairing (Fig. 1.1 A). This specific base pairing leads to the formation of a right-handed double helix with the two strands running in opposite directions (Fig. 1.1 B). A further stabilization comes from π stacking interaction of the nucleobases[1].

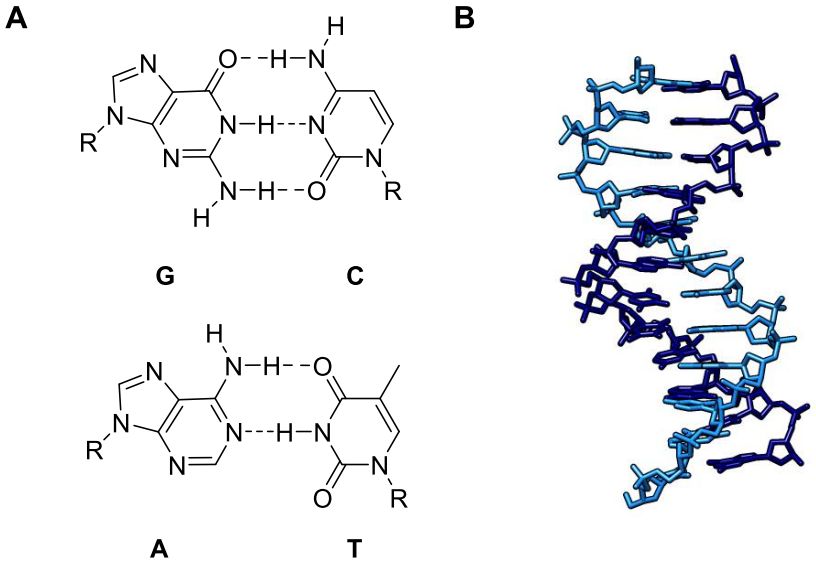
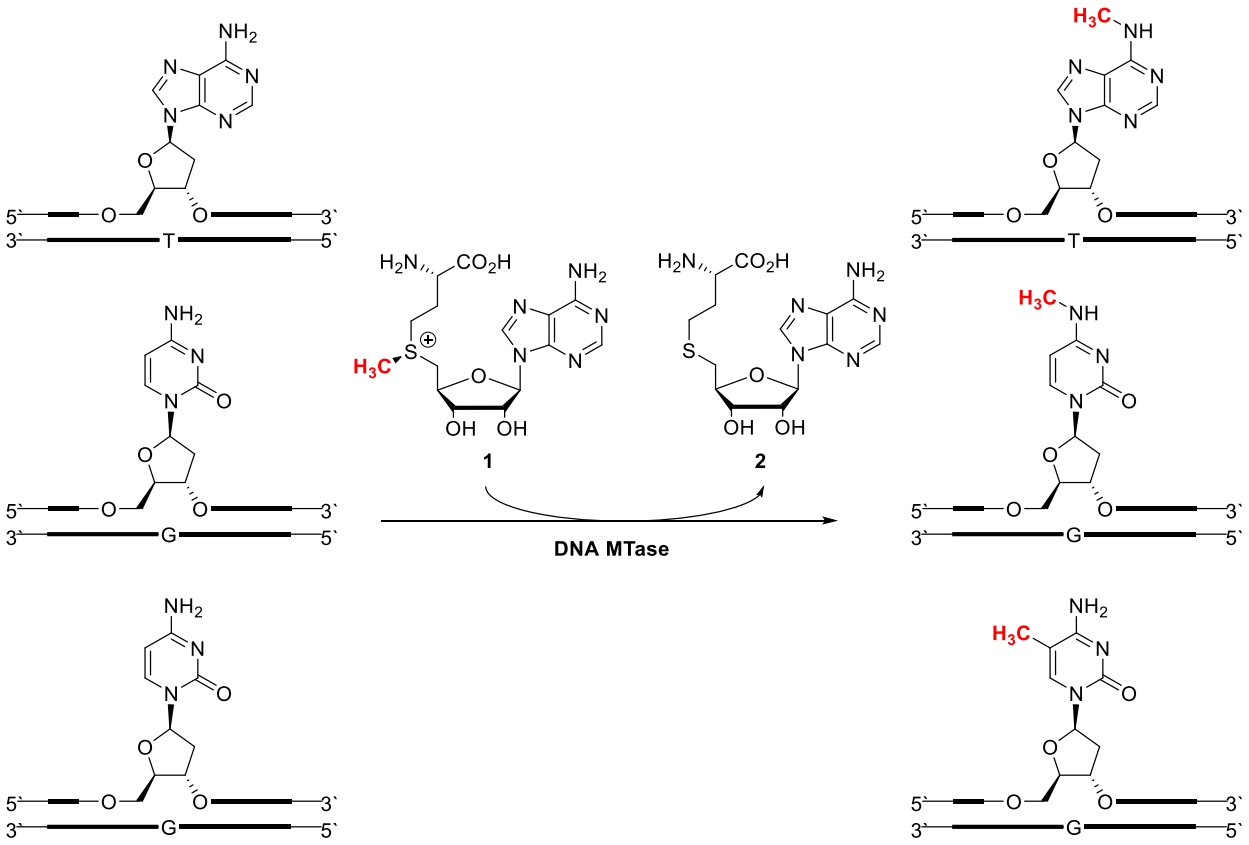


Fig. 1.1: DNA. **A:** Base pairing according to Watson and Crick. **B:** Helical structure of B DNA.

The nucleotide sequence dictates the structural basis of a cell. During DNA replication, the original strands separate and serve as templates for the formation of new strands to which the genetic information is exactly copied[2].

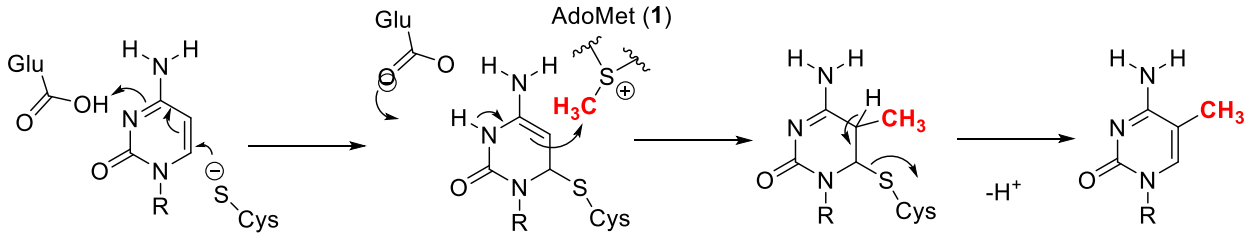
1.2 DNA methylation

Despite of the identical information cells inherit through DNA replication, multicellular organisms have many cell types with different functions. Responsible for this cellular differentiation are epigenetic changes, heritable modifications of DNA and histone proteins without alteration of the nucleotide sequence[3]. An epigenetic mechanism that is recognized as a key process involved in the expression of genes[4], and thus the characterization of a cell, is DNA methylation. DNA methylation is the specific and covalent transfer of a methyl group on DNA and is catalyzed by DNA methyltransferases (MTases). DNA MTases recognize a certain nucleotide sequence and transfer an activated methyl group, provided by the cofactor *S*-adenosyl-L-methionine (AdoMet, **1**), to either adenine or cytosine nucleobases. *S*-adenosyl-L-homocysteine (AdoHcy, **2**) is formed as a byproduct (Scheme 1.1).



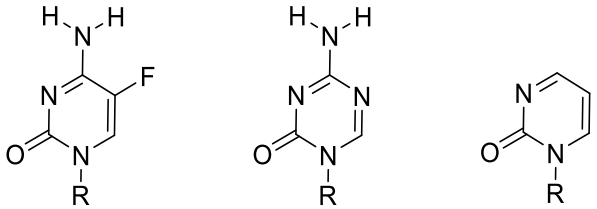
Scheme 1.1: DNA MTase-catalyzed DNA methylation, leading to the formation of *N*6-methyladenine, *N*4-methylcytosine and *C*5-methylcytosine, respectively.

Based on the different methylated nucleobases found[5, 6, 7] the DNA MTases are divided into two classes: N-DNA MTases, subdivided into DNA adenine-N6 and cytosine-N4 MTases, catalyze the methylation of the exocyclic nitrogens of adenine and cytosine, respectively[8]. C5-DNA MTases catalyze the methylation at the C5 position of cytosine[9] and it is the most abundant type of methylation naturally found in eukaryotes. C5-DNA MTases share a common primary structure with ten conserved sequence motifs numbered I to X[8]. The target recognition domain, that lies within a long variable region between motif VIII and IX, accounts for the recognition of a specific sequence on the DNA. The variety in this region is less between DNA MTases with similar sequence specificity[10, 11]. The conserved motifs are assigned to common functions of the DNA MTases: Motif I shows similarities with sequences in other AdoMet-dependent MTases[12, 13] and is involved in cofactor binding. Motif IV, VI and VIII are related to catalysis. Motif IV, the PCQ motif, is the most invariable and contains the catalytic site[11]. A cysteine within this motif is a key player in the catalytic mechanism of C5-DNA MTases that was postulated by Wu and Santi[14, 15]. The thiol group of the cysteine residue attacks the C6 position of the target cytosine to form a covalent intermediate. A glutamate stabilizes the intermediate by neutralizing the generated negative charge[16]. The increased nucleophilic character of the C5 atom activates it to a nucleophilic attack on AdoMet (**1**). After methyl transfer, the proton at C5 is eliminated and the aromatic ring is restored (Scheme 1.2).



Scheme 1.2: Catalytic mechanism of C5-DNA MTases based on the crystal structure of M.HhaI[17].

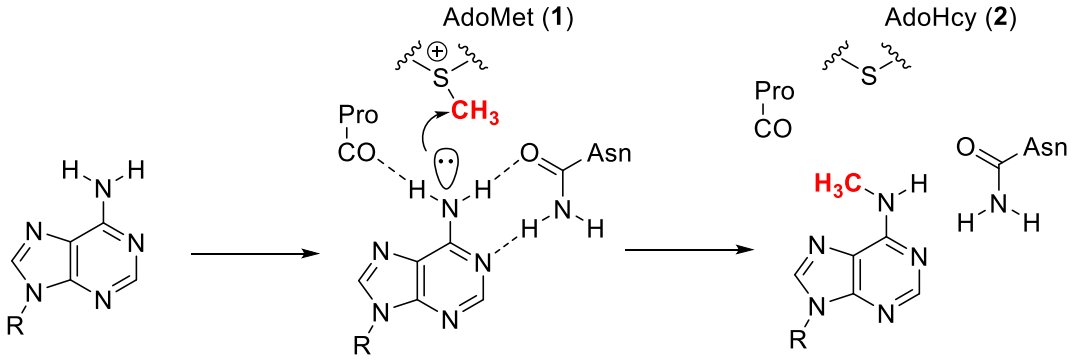
The covalent intermediate can be trapped by substituting cytosine with analogs (Fig 1.2) such as 5-fluorocytosine (**3**)[18], 5-azacytosine (**4**)[19] or zebularine (**5**)[20] and herewith preventing completion of the catalytic cycle. In this way, C5-DNA MTases can be inhibited.



3 **4** **5**

Fig. 1.2: Cytosine analogs resulting in a trapped covalent intermediate of the C5-MTase catalytic cycle.

The N-DNA MTases have nine conserved motifs that correspond to the motif I to VIII and X within the C5-DNA MTases. N-DNA MTases are divided into three groups on the order of regions responsible for cofactor binding and catalysis[21]. The catalytic mechanism of N-DNA MTases was revealed by the structure of the N6 adenine MTase M.EcoRI, showing direct methyl group transfer to the N6 atom[22] The negative polarization of the N6 atom of adenine *via* hydrogen bond formation with an asparagine residue and a proline residue was additionally postulated to cause the activation of the N6 atom for a nucleophilic attack (Scheme 1.3)[23].



Scheme 1.3: Suggested catalytic mechanism of N-DNA MTases, based on the structure of M.TaqI co-crystallized with DNA[23].

As the nucleobases are buried deep within the helical structure, there is not sufficient space for the methylation reaction to take place[24]. DNA MTases use a so-called base-flipping mechanism and have an active role in opening the DNA to access the target nucleobase[25]. The target nucleobase is elegantly rotated 180° out of the helix, without breaking any covalent bonds or heavily disturbing the DNA helix otherwise[24]. Observations made with structures of M.HhaI and duplex oligonucleotides with an abasic site at the target base showed that it is in fact the backbone that is rotated[26, 27]. The extrahelical conformation is stabilized through interactions with the DNA MTase[28]. The nucleobase flips along and is stabilized in the catalytic pocket by several residues of the enzyme. Once the methyl transfer has succeeded, the nucleobase is flipped back. Only DNA, DNA MTase and AdoMet (**1**) are present during the base-flipping process[29]. Co-crystallization of the DNA MTase M.HhaI with DNA and AdoHcy (**2**) (Fig. 1.3) led to this discovery[17] and the mechanism was later not only confirmed for other C5-DNA MTases[30, 31] but also for N-DNA MTases[23].

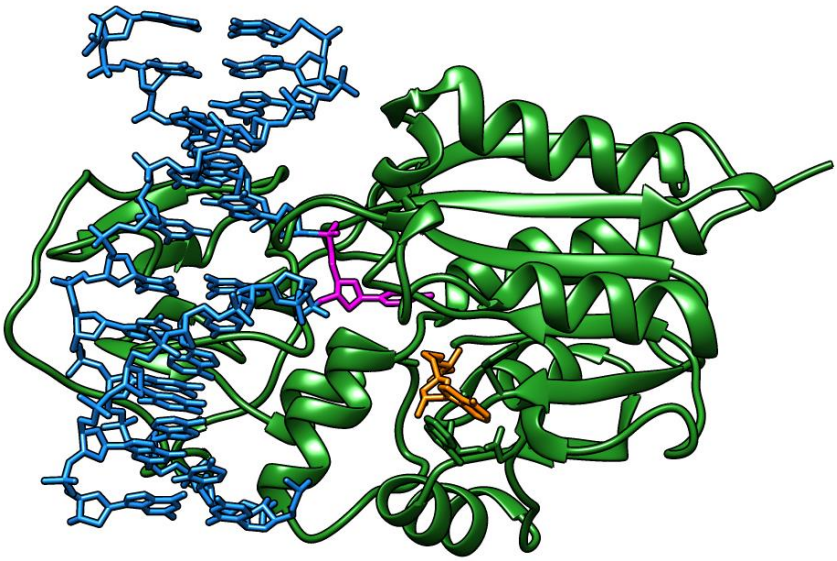


Fig. 1.3: Structure of M.HhaI (green) in complex with DNA (blue) and the target base (magenta) flipped out of the helix. The cofactor is indicated in orange[32]. PDB ID 5MHT.

1.2.1 DNA methylation in prokaryotes

In prokaryotes, both N- and C-DNA methylation take place. One function of DNA methylation in prokaryotes is to discriminate between own DNA and foreign DNA, which is carried out by means of modification-restriction systems[33]. These systems work with a combination of two enzymes. A DNA MTase sequence-specifically methylates own DNA, protecting it against restriction. A methylation sensitive restriction endonuclease (REase) cleaves DNA specifically at the recognition site of the DNA MTase. In this way the bacterium is protected against foreign DNA, as the latter is usually not methylated and thus will be cleaved by the REase.

DNA methylation in prokaryotes also serves to regulate DNA replication and control post replicative mismatch repair. An example where the latter role is well understood is the *E. coli* deoxyadenosine methylase system, where adenine in GATC sequences is methylated[34]. Just after DNA replication, when the palindromic DNA sequence is in a hemimethylated state, the deoxyadenosine MTase methylates the adenine to a state of full methylation. In the short timeframe between hemimethylated and full-methylated DNA, post replicative mismatches can be repaired, as it is possible to distinguish the methylated parental from the unmethylated daughter strand[16].

1.2.2 DNA methylation in eukaryotes

DNA methylation in eukaryotes is generally limited to cytosine-C5 methylation. In mammalians it is nearly restricted to the methylation within CpG sequences and occurs only at certain CpG sites, resulting in methylation patterns. The patterns differ between different cell types within an organism but are conserved in corresponding cell types between organisms of the same kind[35]. Owing to the mutagenic nature of 5-methylcytosine, CpG sites are generally scarcely represented in the mammalian genome. Yet certain regions possess a high CpG density, termed CpG islands, and they are often located in the promoter region, near the 5`-end of the gene[36]. CpG islands are usually in an unmethylated state, whereas the regions with low CpG density are methylated. DNA methylation patterns carry the epigenetic information that defines the character of a cell. The establishment and maintenance of DNA methylation patterns is accomplished by the cooperation of the mammalian DNA MTases DNMT1, DNMT3a/b and DNMT3L, together with histone modifications, passive and active demethylation by ten eleven translocation dioxygenases[37] and other epigenetic marks[38].

Mainly accountable for setting the DNA methylation patterns are DNMT3a and DNMT3b, traditionally defined as *de novo* DNA MTases, as they introduce new methylation marks onto DNA during early stages of mammalian development[39]. DNMT3a and DNMT3b were discovered with the help of expressed sequence tag databases[40]. Although about 85% of the sequences within the catalytic domains of DNMT3a and DNMT3b share similarities, DNMT3a works in a distributive and DNMT3b in a processive manner[41]. DNMT3a also shows high catalytic activity at non-CpG sites[42] and is associated with non-CpG methylation within embryonic stem cells, germ cells and differentiated neuronal cells[43]. DNMT3L is, in contrast to the other DNMTs, catalytically inactive. It has a regulatory function and can either stimulate methylation by DNMT3a/b *in vitro* and *in vivo*[44] or compete with DNMT3a/b to reduce methylation[45].

Preservation of the methylation patterns is the main function of DNMT1, the first mammalian DNA MTase that was cloned[46]. The genome contains a large quantity of DNMT1. It is present mainly in dividing cells and is maximally expressed during DNA replication. DNMT1 assures that the original methylation marks on the parental strands are inherited by the daughter strands after replication, by methylation of the unmethylated strand in hemi-methylated DNA[47]. This function is reflected in its preference to bind to hemimethylated DNA[48]. The specificity towards hemimethylated DNA is nevertheless insufficient to maintain methylation patterns genome wide[38] and DNMT3a/b are required to cooperate with DNMT1, especially in CpG rich regions, for example in embryonic stem cells[49]. DNMT1 on the other hand is known to also exhibit *de novo* methylation[50], showing that the functions of the DNMTs are not that sharply confined. DNMTs can be guided to the desired genomic regions by the interaction with certain proteins or histone modifications[38].

Many diseases are associated with abnormal DNA methylation[51]. This could be a result of incorrect establishment of DNA methylation as in the case with the *immunodeficiency, centromeric instability and facial anomalies* syndrome that is caused by mutations in the DNMT3b gene[39]. Other diseases, such as the autoimmune disease *systemic lupus erythematosus*, are caused by defective maintenance of the methylation patterns[52, 53]. Aberrant DNA methylation is often accompanied with the contraction or expansion of repeats and there are many diseases linked to repeat instability[54]. Finally, as DNA methylation is involved in genomic imprinting, the loss of imprinting also plays a role in several disorders, including cancer[51]. Irregular DNA methylation is involved in almost every form of human cancer[55]. Global hypomethylation of DNA results in genomic instability[51], whereas hypermethylation of CpG islands within promoter regions inactivates tumor-suppressor genes[56].

Of all types of cancer, carcinomas that develop from epithelial cells are the most common and usually show overexpression of the epithelial cell adhesion molecule (EpCAM). EpCAM enables cell-cell adhesion and has been identified as a signal transducer in the development of cancer[57].
The EpCAM gene could actively and permanently be downregulated by CpG methylation in the promoter region but with accompanying toxicity due to the non-targeted manner of the methylation. These observations indicated the potential of DNA methylation in silencing genes and the technical challenges associated with the need to target methylation to specific regions[58].

1.2.3 Bacterial CpG-specific DNA MTases

Most bacterial C5-DNA MTases catalyze methylation of cytosine within a recognition sequence that is longer than the CpG sequence. Only a few known bacterial C5-DNA MTases share their recognition sequence with that of mammalian DNA MTases, making them interesting tools for studying DNA methylation in eukaryotes.

1.2.3.1 M.SssI

M.SssI originates from the *Spiroplasma* sp. strain MQ1. It was cloned, expressed in *E. coli*, and characterized[59]. M.SssI catalyzes the methylation of cytosine within the CpG sequence without differentiating between hemimethylated and unmethylated DNA. In the absence of magnesium ions, M.SssI works in a processive manner[60]. This changes to a distributive mode of methylation when magnesium ions are present, as they decrease the affinity of M.SssI towards DNA. M.SssI also shows type I-like topoisomerase activity in the presence of magnesium ions[61].

The crystal structure of M.SssI has not yet been determined. Therefore, a model of the ternary complex of M.SssI, DNA and AdoHcy (**2**), based on the structures of M.HhaI and M.HaeIII, was created (Fig. 1.4), showing which amino acid residues could be important for binding to the DNA and to the cofactor[62].

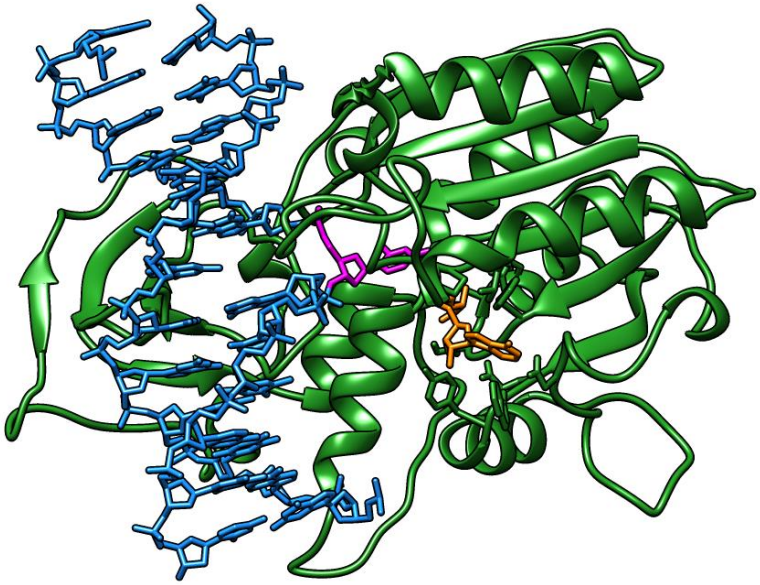


Fig. 1.4: Model of M.SssI (green) in complex with DNA (blue) and AdoHcy (**2**, orange)[62]. The flipped-out base is colored magenta.

The *in vivo* production of inactive N-terminal and inactive C terminal fragments of M.SssI in one *E. coli* cell can lead to the assembly of active enzyme, but only with a few percent activity of the native (full-length) enzyme[63].

1.2.3.2 M.MpeI

M.MpeI was only recently discovered in the CpG poor *Mycoplasma penetrans*[30]. M.MpeI catalyzes the *in vivo* and *in vitro* methylation of cytosine within the CpG sequence. The cocrystal structure with DNA confirms that M.MpeI uses the base-flipping mechanism to access the target cytosine (Fig. 1.5). A structural comparison between M.MpeI, DNMT1 and M.HhaI shows a few conserved residues, among which is the catalytic cysteine. In contrast to DNMT1, the residue that temporarily replaces the flipped-out base and interacts with the orphan guanine is an intercalating glutamine residue and not a guanine of the 5`-side of the substrate strand[30].

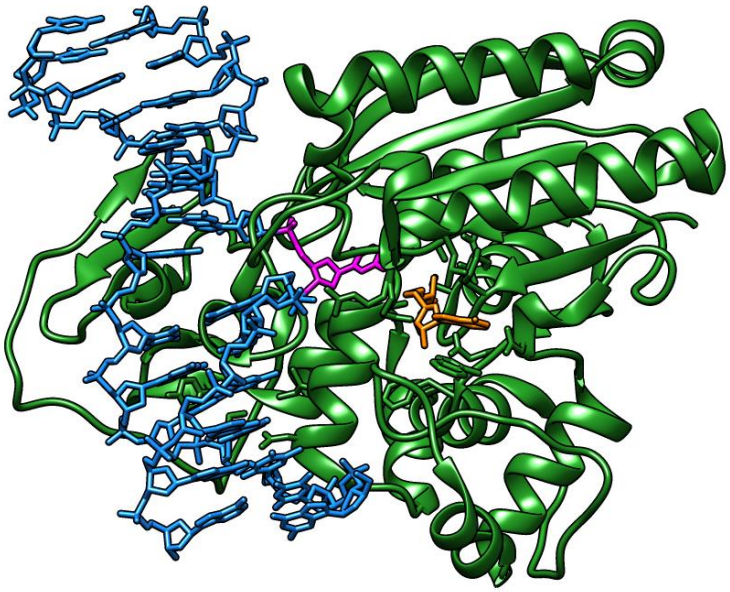


Fig. 1.5: Structure of M.MpeI (green) in complex with target DNA (blue). The target cytosine (magenta) is flipped out of the helix. The cofactor is indicated in orange[30]. PDB ID: 4DKJ.

1.3 Targeted DNA methylation

Methylation of specific CpG sites would be interesting to study and control gene expression[64]. The concept of targeted DNA methylation is based on increasing the DNA MTase specificity by covalently coupling the DNA MTase to a targeting device. The targeting device binds highly specifically to a DNA sequence near an addressed site and anchors the DNA MTase. The DNA MTase now preferentially methylates at the targeted site (Fig. 1.6)[65]. Several systems can serve as a targeting device and thereby indirectly increase the specificity of the DNA MTase.

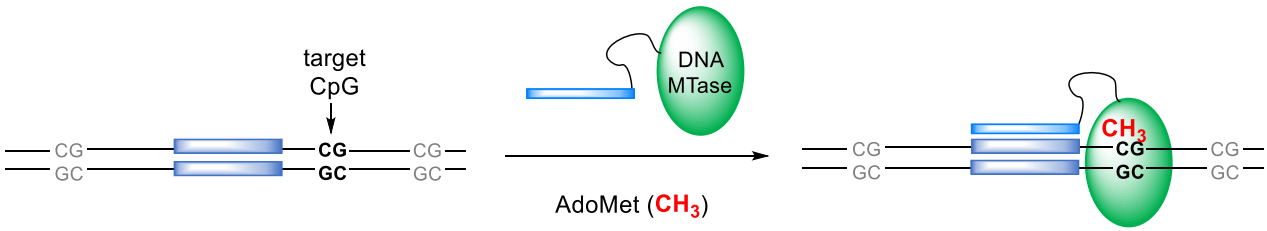


Fig. 1.6: General concept of targeted DNA methylation. The targeted DNA sequence is depicted in dark blue, the targeting device in light blue, the DNA MTase in green and the methyl group in red.

1.3.1 Synthetic polyamides

Based on the structures of the naturally occurring minor groove binders distamycin A and netropsin, small polyamides comprising of *N*-methylpyrroles (Py), *N*-methylimidazoles (Im) and *N*-methylhydroxypyrroles (Hp), were designed to bind DNA sequences with high specificity and affinity[66]. Pairs of Im/Py, Py/Im, Hp/Py and Py/Hp each recognize a Watson-Crick base pair (bp)[66]. The covalent linkage of two antiparallelly oriented polyamide strands with a composition that corresponds to the desired target DNA sequence leads to folding into a hairpin in the minor groove (Fig. 1.7)[67, 68].

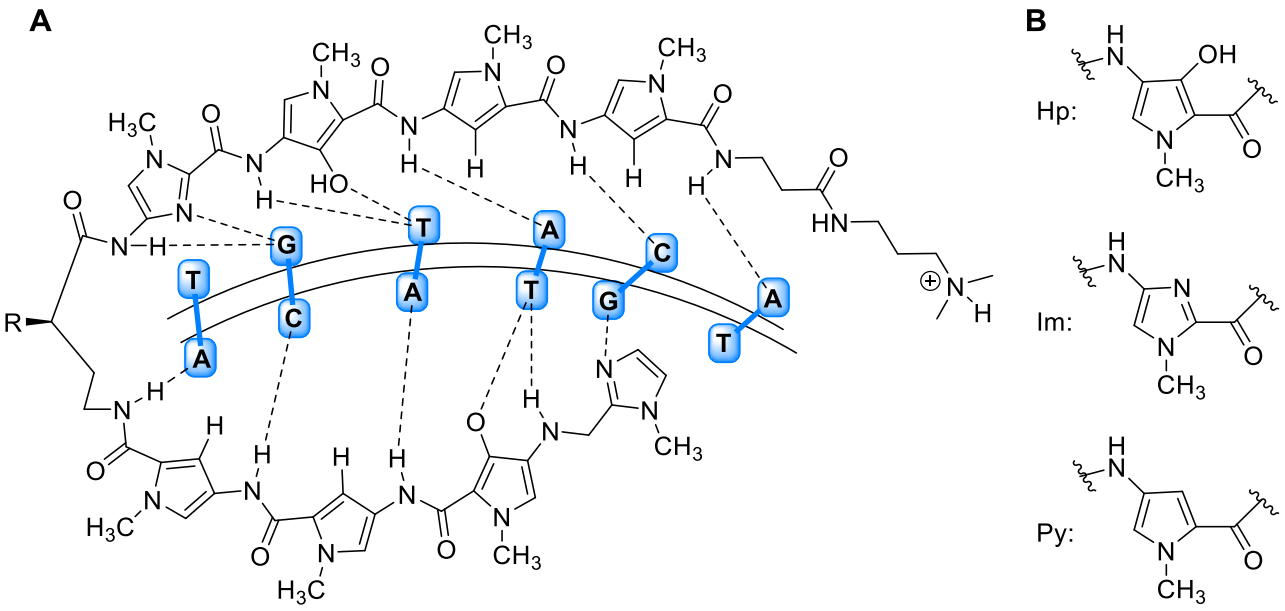


Fig. 1.7: Polyamide minor groove binders. **A:** Eight-membered polyamide binding to a DNA target sequence of 6 bp (blue)[68]. **B:** Structures of *N*-methylhydroxypyrrole (Hp), *N*-methylimidazole (Im) and *N*-methylpyrrole (Py).

Polyamides adjust perfectly to the curve of the DNA minor groove[69]. They are easily taken up by the cell and any desired DNA sequence can be targeted with suitably composed polyamides.

A disadvantage is that the polyamides target short sequences which occur multiple times within the human genome. To increase the target sequence length, hairpin modules were combined into so-called tandem-hairpin polyamides that could target longer DNA sequences[70, 71].

The repertoire of synthetic polyamides was further expanded by constructing other motifs and finding new heterocyclic ring systems with improved affinity and specificity compared to the original three[68]. Polyamides have successfully been used for downregulating and activating genes[68].

1.3.2 Zinc finger proteins

A zinc finger protein was the first device used for targeted DNA methylation[65]. A zinc finger is a structural motif within proteins that is stabilized by chelation of a zinc ion. The most common zinc finger motif is the Cys_2His_2 motif. It consists of about 30 amino acids and folds into a ββα structure. The zinc ion is coordinated by the conserved cysteine and histidine residues. Each zinc finger motif recognizes and binds three bp within a DNA sequence, by inserting the α-helix in the major groove of the DNA (Fig. 1.8). Linkage of several zinc finger motifs increases the sequence specificity and a unique sequence within the

DNA can be targeted[72]. The sequence specificity can be further customized by exchanging certain amino acids within a zinc finger motif[73].

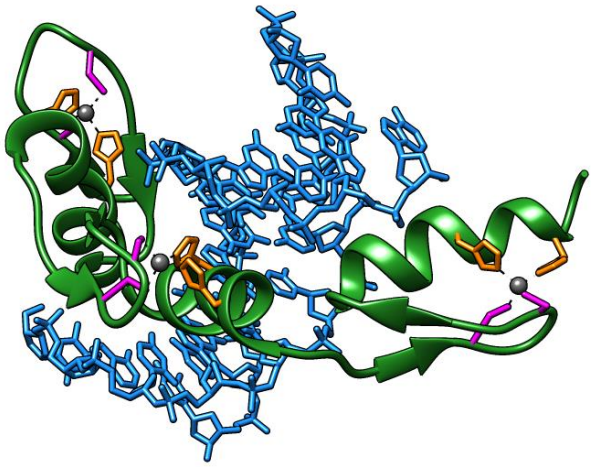


Fig. 1.8: Structure of DNA (blue) in complex with Zif268 (green), a zinc finger protein with the Cys_2Hys_2 motif. Three zinc fingers bind to the DNA by inserting their α-helix into the major groove. Each zinc finger coordinates a Zn^{2+} ion (gray) by the conserved cysteine (magenta) and histidine (orange) residues[74]. PDB ID 1AAY.

Zinc finger proteins have been used to assay methylation in *E. coli*[75, 76] and control methylation in mammalian cell lines[77, 78]. They have not only been genetically fused to DNA MTases but also to REases, to control DNA cleavage[79]. Off-target methylation was minimized with a construction of a DNA MTase-zinc finger fusion protein with reduced DNA binding affinity of the DNA MTase component and dominating properties of the zinc fingers[80]. Another strategy to reduce background methylation is based on the fact that inactive N- and C-terminal fragments of certain bacterial C5-DNA MTases can assemble to active enzyme if they are expressed in the same *E. coli* cell[63, 81] and was demonstrated with the complementation of inactive M.HhaI fragments that were fused to zinc fingers, leading to active enzyme with a high specificity towards the target site[82].

1.3.3 Transcription activator-like effector proteins

Transcription activator-like effector (TALE) proteins originate from plant pathogens and are secreted into plants, where they bind the nucleus and imitate eukaryotic transcription factors to alter gene expression and herewith promote the bacterial infection[83]. TALE proteins have three conserved regions. The N-terminal region contains the system required for secretion. The C-terminal region contains signals for nuclear localization and activation[83]. The central repeat domain is responsible for specific DNA binding. It consists of typically 34 amino acids long repetitive sequences. A set of highly variable amino acids at the 12th and 13th position within each repeat, referred to as the repeat-variable-diresidue, determines DNA binding. The RVDs directly correlate to one of each of the four nucleobases[84, 85]. Upon recognition of the DNA target sequence, the loosely wrapped TALE protein binds more tightly to the DNA (Fig. 1.9)[86].

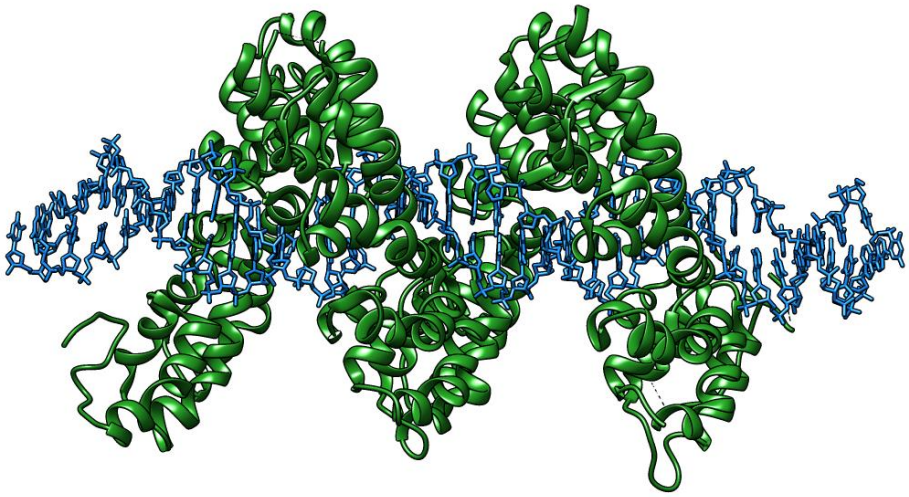


Fig. 1.9: Crystal structure of the central repeat domain of PthXo1 TALE (green) wound around its target DNA (blue)[87]. PDB ID: 3UGM.

With the knowledge of which RVD recognizes which nucleobase, custom designed TALEs were constructed and fused with all kinds of proteins *e.g.* for specific DNA cleavage[88], specific CpG demethylation[89], chromatin modification[90] and targeted DNA methylation[91].

A disadvantage is the negative impact of 5-methylcytosine and other epigenetic modifications on the binding of TALE proteins[92].

1.3.4 CRISPR/dCas9

CRISPR stands for clustered regularly interspaced short palindromic repeats and is a type of DNA sequences found in many bacteria. Segments of repetitive DNA sequences are alternated with built-in fragments of foreign DNA, deriving from viral attacks. The CRISPR system works closely together with CRISPR associated (Cas) proteins to ward off viruses[93].

In a minimal set up, like the CRISPR/Cas9 system, the Cas9 endonuclease is guided to exogenous DNA with the help of two RNA molecules. This set up has been engineered into a system where the functions of the two RNA molecules are fused into one, single guide RNA (sgRNA)[94]. By replacing the Cas9 with a catalytically inactive endonuclease (dCas9) and co-expressing it with the sgRNA, dCas9 could be directed to a target DNA sequence and sterically block transcription[95]. Usage of dCas9 instead of Cas9, makes the system function merely as a targeting device based on an RNA-guided DNA-binding complex. By fusion of dCas9 with various effectors, the system can not only be used for sterically blocking transcription but also for the active up- and downregulation of transcription[96, 97, 98], histone modification[99, 100] and targeted DNA methylation[101, 102, 103] and demethylation (Fig. 1.10)[104].

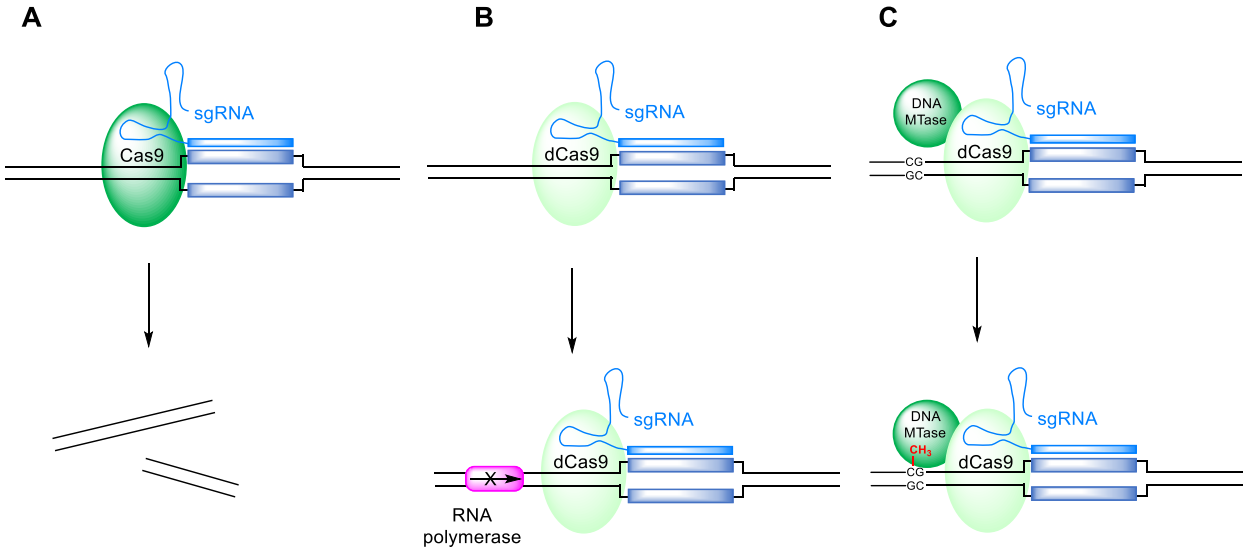


Fig. 1.10: RNA-guided targeting by CRISPR/Cas9. The sgRNA (light blue) binds to the target DNA sequence (blue) and guides either **A:** Active Cas9 (green) to cleave the DNA, **B:** Inactive Cas9 (dCas9, light green) to sterically block RNA polymerase (magenta) or **C:** dCas9 genetically fused to an effector (*e.g.* a DNA MTase, green) to actively modify the DNA at the target site. Fig. adapted from [95].

This method of targeting requires only two components. A readily constructed fusion protein can be addressed to a different target sequence, simply by redesigning the sgRNA, which is relatively effortless compared to engineering zinc finger proteins or TALE proteins for each desired target sequence. A well designed sgRNA can achieve highly specific and efficient targeting[105]. The robustness of the system also allows for targeting multiple genes, by expressing more than one sgRNA from a single plasmid[106, 107].

1.3.5 Triple helix-forming oligodeoxynucleotides

The ability of nucleic acids to form three-stranded structures in the presence of magnesium ions was found only four years after the postulation of the helical double-stranded DNA structure by Watson and Crick[108]. Later it was discovered that short oligodeoxynucleotides, called triple helix-forming oligodeoxynucleotides (TFO), can bind in the major groove of DNA to form triple-helical structures. TFOs typically have a length of 10 to 30 nucleotides and bind specifically to oligopurine-oligopyrimidine sequences, forming Hoogsteen or reverse Hoogsteen hydrogen bonds to the purine strand[109]. Although the need for oligopurine-oligopyrimidine stretches limits the use of TFOs, these stretches are overrepresented in the human genome, especially in the promoter regions[110]. Triple helices are stabilized by polyamines such as spermine[111, 112] or divalent cations like magnesium, zinc or calcium ions[113, 114], which reduce the electrostatic repulsion between the three negatively charged strands. The TFOs can be categorized into two classes, based on the relative orientation of the third strand to the oligopurine strand (Fig. 1.11). In the pyrimidine motif, the third strand is parallelly oriented to the oligopurine strand. Protonated C binds to a G-C bp strand and T binds to an A-T bp with Hoogsteen hydrogen bonds[115]. In the purine motif, the third strand has an antiparallel orientation to the oligopurine strand and G, A and T bind with reverse Hoogsteen hydrogen bonds to G-C, A-T and A-T bp, respectively[116]. Generally, TFOs with a pyrimidine motif are preferred for the triplex formation with AT-rich sequences, whereas for GC-rich sequences the purine motif is the preferred choice[117].

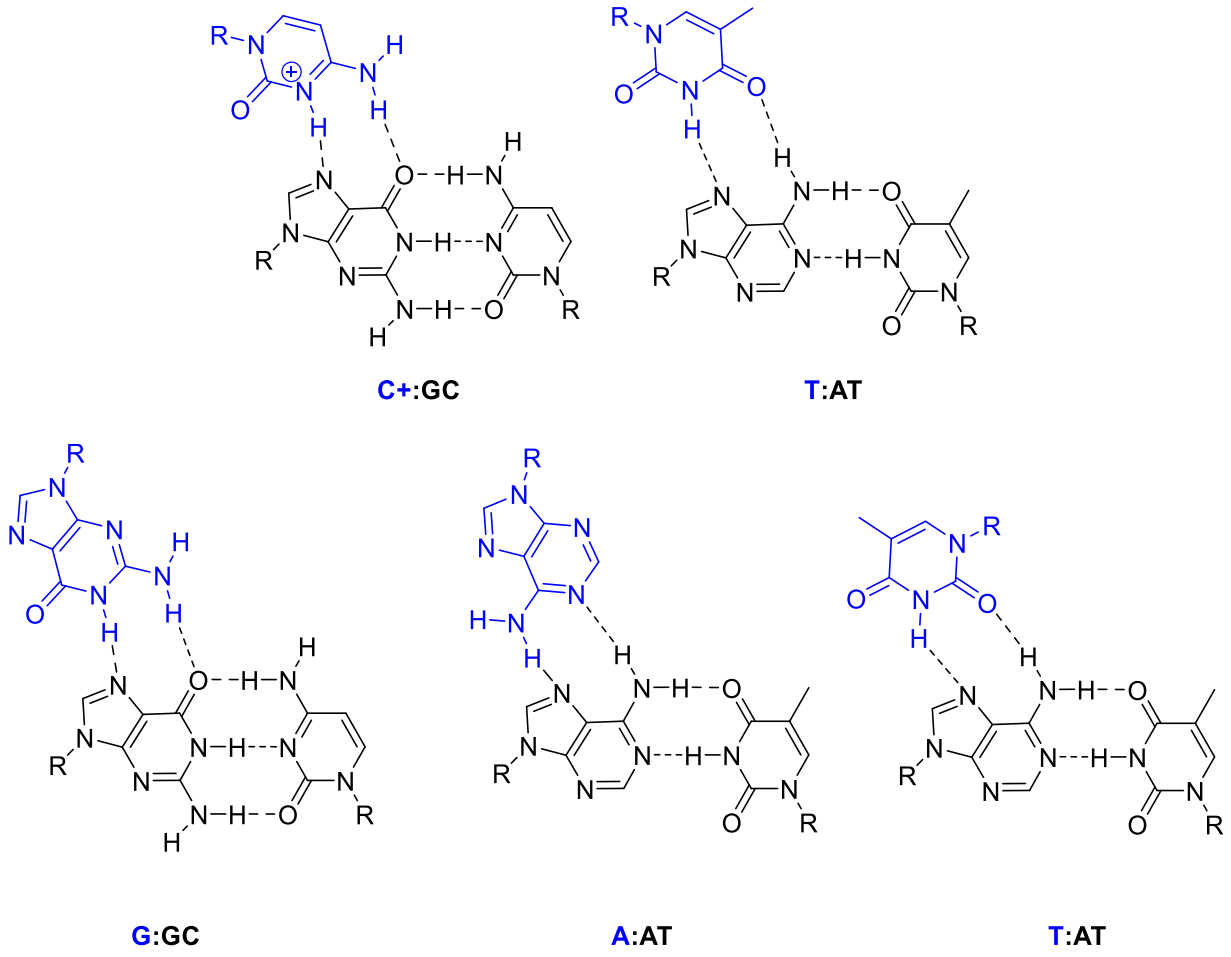


Fig. 1.11: Triple helix binding motifs. The third strand nucleobases are colored blue.

The conditions within living cells are not optimal for triple helices. First, a neutral pH is problematic for TFOs with pyrimidine motif that bind to G-C duplexes, as C needs to be protonated. Also, the required magnesium ion concentration to stabilize triple helices is much higher than the physiological concentration[113] and the relatively high concentration of potassium ions within the cell stabilizes the association of G-rich TFOs into other structures like G-quadruplexes[118]. Finally, the oligodeoxynucleotides are prone to be degraded by nucleases[119]. Modification of the nucleobases, the sugar or the phosphate backbone can tackle these limitations (Fig. 1.12).

A well-known modification is the substitution of C with methylcytosine (**6**). The higher pKa value of 5-methylcytosine allows the pyrimidine motif TFOs to be used at a neutral pH[120, 121]. Another way to address the pH restriction is the use of pseudoisocytosine (**7**), which has the same hydrogen bonding pattern as protonated C[122].

To lower the need for magnesium ions, T can be replaced with 5-propynyluracil (**8**). This stabilizes the triple helix by increased base stacking interactions[123]. Replacement of the backbone with an uncharged backbone (*e.g.* morpholinos[124] or peptide nucleic acid, PNA[125]) not only diminishes the requirement of Mg^{2+}, but also makes the TFOs more resistant to nuclease degradation. Using 2'-methoxylated nucleosides[126] or locked nucleic acid (LNA)[127] stabilizes the C3`-endo conformation of the sugar and thereby minimizes the distortion of the duplex when the triple helix is formed. This also increases the resistance towards nucleases.

The introduction of modified bases like 6-thioguanine (**9**) could prevent the formation of G quartets and thereby reduce potassium ion mediated inhibition of triple helix formation[128]. Inhibition by potassium can also be reduced by replacing the phosphate backbone with a positively charged backbone. Other backbone modifications to reduce the susceptibility of TFOs towards nucleases include phosphorothioates[129] and phosphoroamidates linkages[130].

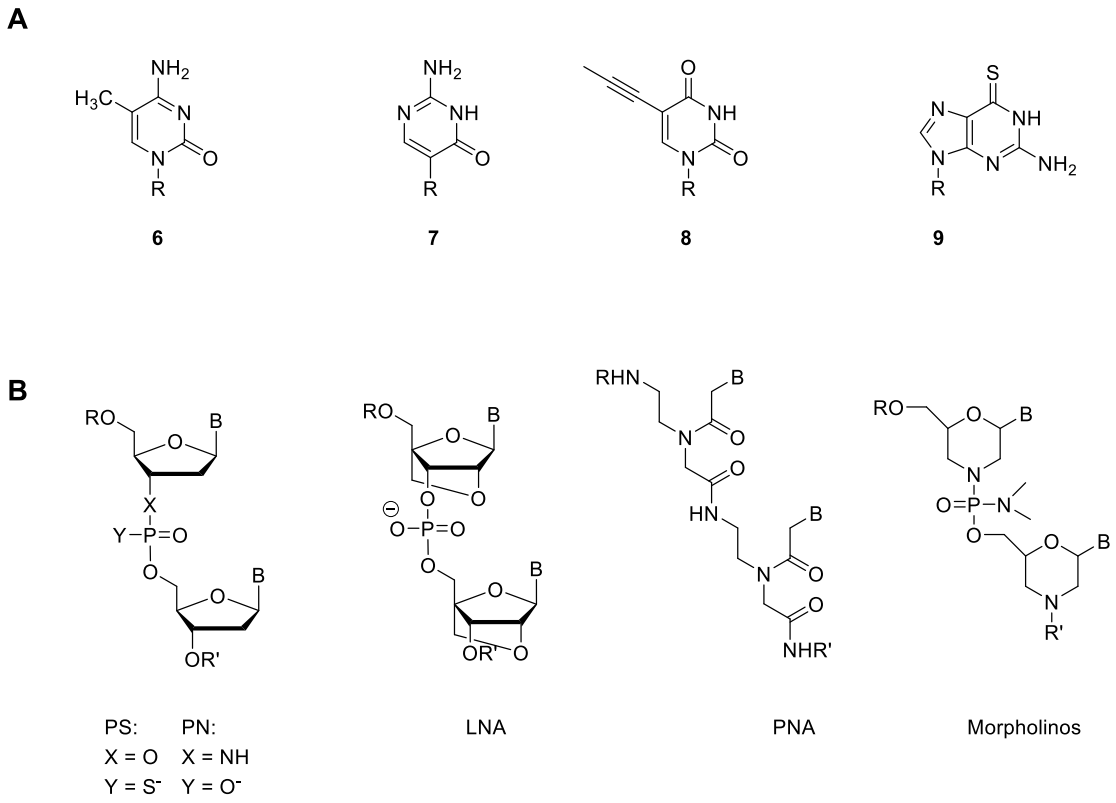
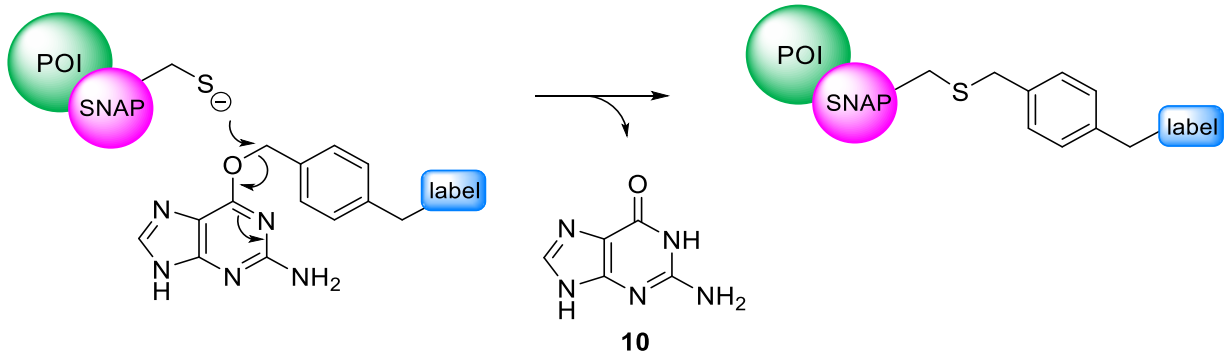


Fig. 1.12: Modifications to increase triplex stability. **A:** Modified nucleobases: methylcytosine, pseudoisocytosine, 5-propynyluracil and 6-thioguanine. **B:** Modified backbones: phosphorothioates, phosphoroamidates, LNA, PNA and morpholinos connecting the nucleobases.

TFOs have an anionic character and a relative high molecular weight, which makes cellular uptake challenging[131]. This issue can be addressed by conjugating the TFOs with cationic lipids[132], cellpermeable agents[133], DNA condensing agents[134], dendrimers[135] or hydrophobic compounds like cholesterol[136]. TFOs have been successfully used for the modulation of gene activity, both *in vitro* and *in vivo*[109, 131].

1.4 SNAP-tag technology

The SNAP-tag is a protein-tag, based on the human repair enzyme *O*6-alkylguanine DNA alkyltransferase (hAGT), whose natural function is to repair *O*6-alkylated guanine residues in DNA by covalently transferring the alkyl group to its own active site. As the reaction is irreversible, the enzyme itself is hereby inactivated. The substrate specificity of hAGT is relatively low and *O*6-benzylguanine (BG) is also accepted as a substrate, even when the benzyl ring carries a substituent in the para position[137]. By directed evolution of hAGT, the reaction rate towards BG derivatives was improved and the reaction could be used to efficiently label fusion proteins with small synthetic molecules (Scheme 1.4). The SNAP-tag can be fused to either the C- or the N-terminus of the protein of interest (POI)[138].



Scheme 1.4: Selective covalent coupling of a SNAP-tag fusion protein with a BG derivative under cleavage of guanine (**10**).

Limitations that traditional protein tags cope with, are overcome with the SNAP-tag technology. The method does not have to rely on the intrinsic properties of a protein that can be genetically encoded *e.g.* autofluorescence, but it can utilize synthetic probes with better (fluorescent) properties. Also, the SNAP-tag method is not tailor-made for one specific application but can be used in different locations in the cell and with different expression hosts. The reaction is highly specific, as the BG-substrates do not react with other proteins. The method lends itself for the labeling of fusion proteins *in vitro* and *in vivo*, where in the latter case, the labeling efficiency depends on the cell permeability of the BG-substrate[139]. Since the reaction rate is independent of the nature of the label, the variety of synthetic probes is almost endless and the properties of the POI can be extended as desired[139]. The large spectral shift of available dyes makes it possible to image up to four BG-fluorophores in one cell[140]. The stability of the fluorescence-labeled fusion proteins allows tracking for over periods of hours[141]. By substitution at the para position of the

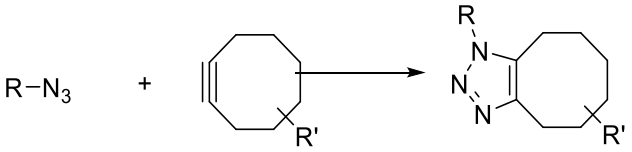
benzyl ring with an oligonucleotide, the SNAP-tag technology can even be used to mediate protein-DNA coupling[142].

Another protein-tag that works in the same way is the CLIP-tag, that reacts specifically with *O*2-benzylcytosine substrates and thus the CLIP-tag and the SNAP-tag can be used simultaneously[143].

1.5 Strain-promoted click chemistry

The classic example for click chemistry is the copper-catalyzed alkyne azide cycloaddition (CuAAC)[144]. Since one of its main players, the azide, is a small molecule, metabolically stable and unreactive towards biological functionality, it is a popular chemical reporter[145]. Yet the cytotoxicity of copper and its ability to block or reduce protein activity by binding their active sites[146, 147], makes the CuAAC not suitable for studying living systems[148] and it created a need for alternative reactions.

One of those alternatives is found by Bertozzi *et al.*[149] and is based on an observation made by Wittig and Krebs, when they reacted a cyclooctyne with phenyl azide which led to the rapid formation of a stable triazole[150]. The huge bond angle deformation of the acetylene group in the cyclic system is the driving force in this so-called strain-promoted alkyne azide cycloadditon (SPAAC), as it makes up for 18 kcal/mol of strain. This energy is released in the transition state upon reaction with an azide[151], making the use of auxiliary reagents such as copper obsolete (Scheme 1.5). Importantly, the selectivity of the reaction remains unchanged.



Scheme 1.5: Strain-promoted [3+2]c cycloaddition of azides and cyclooctynes[149].

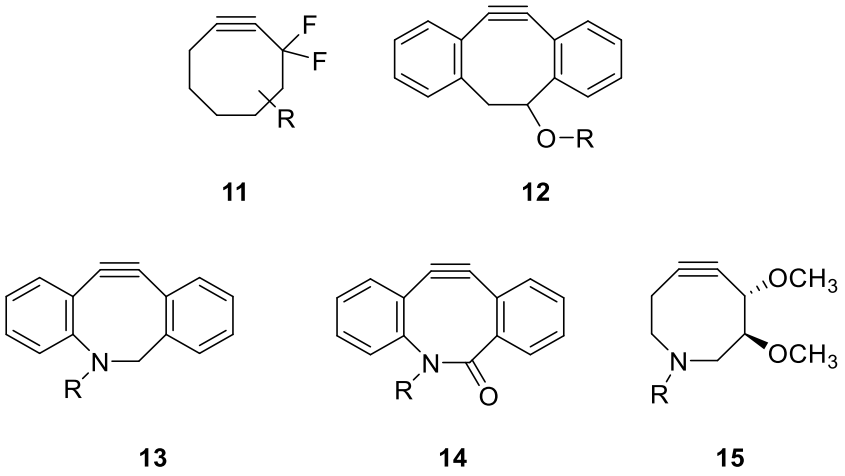
Bertozzi *et al.* further improved the reaction by attaching fluorine substituents to the cyclooctyne ring[146]. The combination of ring strain and electron withdrawing groups led to kinetics that were comparable with those of the equivalent copper-catalyzed reactions. For a second generation of difluorinated cyclooctynes (DIFO, **11**), the synthesis route was simplified, and higher yields were obtained[152].

An increased reactivity can also be achieved through augmented ring strain *via* a dibenzo structure, as Boons *et al.* showed[153]. The reactivity of these dibenzocyclooctynes (DIBO, **12**) was similar to that of the DIFO compounds. Ortho hydrogen atoms of the aromatic rings contribute to the stability of the compound, by protecting the alkyne against nucleophilic attack.

An sp^2 like center (such as nitrogen) in the cyclooctyne ring further elevates the reaction rate, as was shown with DBCO **13**[154] and BARAC **14**[155]. With DIMAC **15**, an increased water solubility was shown

and herewith the sensitivity towards azide detection[156]. The incorporation of a sulfur atom into the structure does enhance the reagent stability but reduces reaction kinetics[157].

Next to SPAAC other copper-free click reactions have been developed, such as the strain-promoted alkyne nitrone cycloadditon[158] and reactions with strained alkenes[159, 160].



11 **12** **13** **14** **15**

Fig. 1.13: Different cyclooctyne systems for strain-promoted cycloadditions.

2 Objectives

As DNA methylation is an essential process within the epigenetic system of an organism, it would be very practical to have a tool to direct methylation to specific sites on DNA and study locus-specific effects. Triple helix-targeted DNA methylation has already been proven to be a successful strategy for this purpose. In previous work, maleimide chemistry was used to couple a triple helix- forming oligodeoxynucleotide (TFO) to an M.SssI variant lacking the catalytic cysteine, to avoid maleimide inactivation of the protein. The conjugate was shown to specifically methylate the addressed CpG *in vitro*[161], but the use of maleimide chemistry required a DNA MTase with a highly reduced DNA MTase activity.

The aim of this work is the construction of DNA MTase-TFO conjugates by SNAP-tag mediated protein-DNA coupling and the evaluation of their DNA methylation specificity. The use of SNAP-tag technology should allow for coupling of TFO with a broader spectrum of DNA MTases including mammalian DNA MTases. Latter would be advantageous for future purposes in therapeutic applications. The TFO, that was previously designed to target a select sequence in the EpCAM promoter[162], is used here as well.

The first part of the thesis focuses on the construction of the DNA MTase-TFO conjugates and its components. DNA MTase-SNAP fusion proteins will be expressed in *E.coli* and their DNA MTase activity evaluated on λ DNA. The TFO will be modified with a BG-function *via* a combination of NHS chemistry and azide-alkyne cycloaddition click chemistry. Since it has been reported that copper mediates the hydrolysis of DNA in the presence of ascorbic acid[163], the strain-promoted azide-alkyne cycloaddition (SPAAC) is preferred, as it obviates the use of copper. The BG-modified TFO will then be covalently coupled with the DNA MTase fusion proteins in a SNAP-tag reaction.

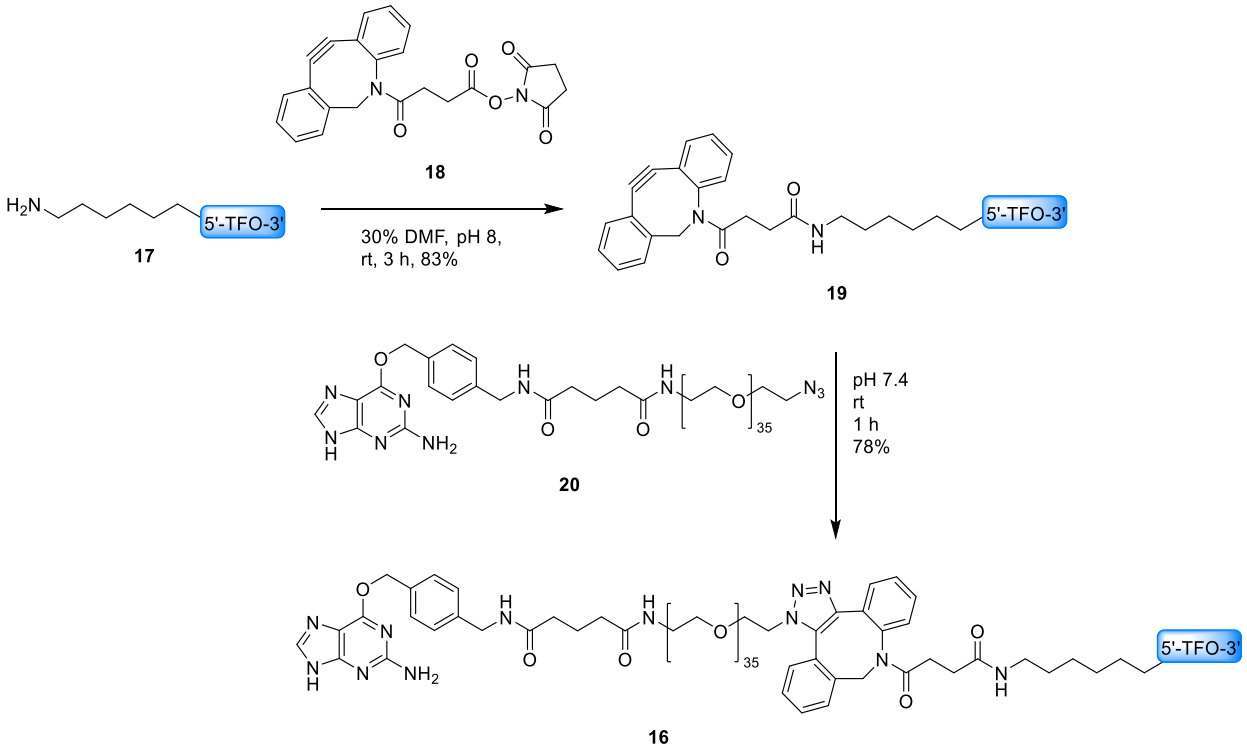
In the second part, the DNA MTase-TFO conjugates will be characterized by their UV absorbance and their DNA MTase activity on λ DNA.

The final part is centered around the *in vitro* DNA methylation specificity of the constructed DNA MTase-TFO conjugates. This will be tested in a modification-restriction assay with plasmids that either contain or lack the triple helix-forming site (TFS). Bisulfite sequencing should further evaluate the *in vitro* DNA methylation specificity.

3 Results and discussion

3.1 Synthesis of BG-PEG$_{35}$-TFO

The synthesis of BG-PEG$_{35}$-TFO (**16**) consists of two steps (Scheme 3.1). In the first reaction, a DBCO group is attached to amino-TFO (**17**). The NHS-activated DBCO compound **18** reacts with the amino function of **17** under release of NHS. In the second reaction, the cyclic acetylene group in the DBCO moiety of **19** is reacted with the azide group of BG-PEG$_{35}$-N$_3$ (**20**) in a strain promoted alkyne azide cycloaddition (SPAAC) under formation of the stable triazole **16**.



Scheme 3.1: Two-step synthesis of BG-PEG$_{35}$-TFO (**16**).

The coupling of amino-TFO (**17**) and DBCO-NHS (**18**) is carried out in an aqueous environment (phosphate buffer pH 8.0 with 30% DMF) at rt. Analysis of the reaction after 3 h by reversed-phase HPLC (rp HPLC) (Fig. 3.1 A) shows that the amino-TFO (**17**) is almost completely converted.

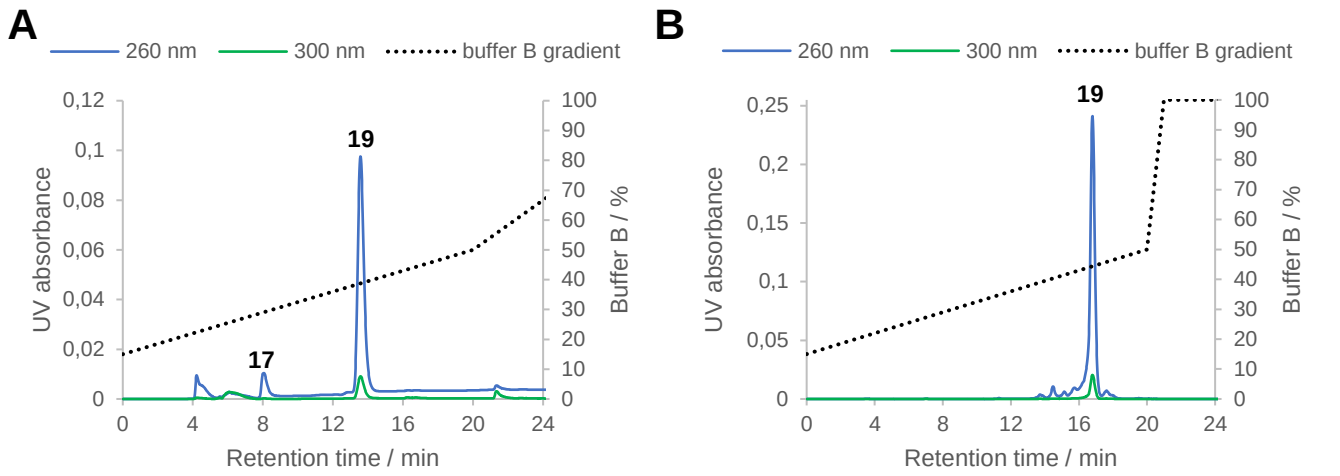


Fig. 3.1: rp HPLC analysis of **A** the reaction of amino-TFO (**17**) and DBCO-NHS (**18**) to DBCO-TFO (**19**) after 3 h and **B** DBCO-TFO (**19**) after purification.

An increased 300/260 nm UV ratio in the product is expected because of the incorporation of the DBCO moiety and indicates that the signal at t_R = 13.6 min corresponds to the product **19**. Isolation of **19** by rp HPLC results in a yield of 83%.

The cycloadditon of DBCO-TFO (**19**) and BG-PEG$_{35}$-N$_3$ (**20**) is also carried out at rt. The HPLC chromatogram (Fig. 3.2 A) shows that DBCO-TFO (**19**) is almost completely consumed after 1 h and a new signal is formed.

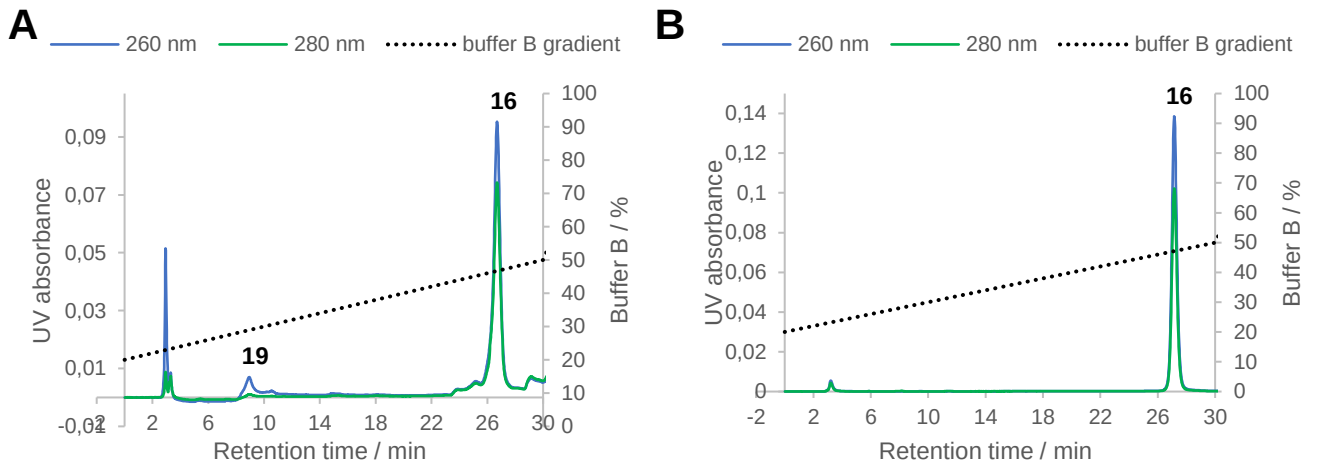


Fig. 3.2: rp HPLC analysis of the reaction of **A** DBCO-TFO (**19**) with BG-PEG$_{35}$-N$_3$ (**20**) to BG-PEG$_{35}$-TFO (**16**) after 1 h and **B** BG-PEG$_{35}$-TFO (**16**) after purification.

The attachment of long and bulky groups to amino-TFO (**17**) leads to a decrease of the polarity with each step. This is indirectly observed through the retention time t_R of the TFO component, that increases with each step. The product BG-PEG$_{35}$-TFO (**16**) is isolated by rp HPLC in 78% yield.

3.2 M.SssI-TFO

3.2.1 Construction and characterization of SNAP-M.SssI

SNAP-M.SssI is expressed in SG1709 *E.coli* cells. After purification by nickel-NTA chromatography, the product is further purified by cation exchange chromatography (Fig. 3.3).

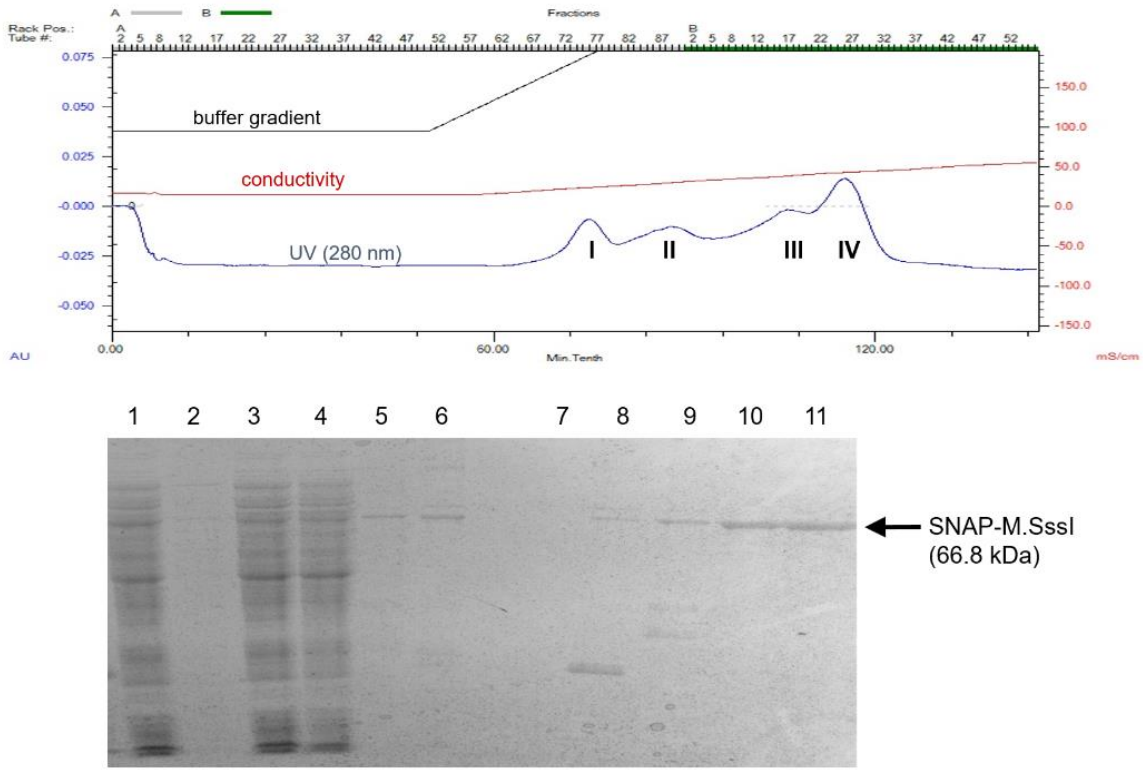


Fig. 3.3: Purification of SNAP-M.SssI by nickel-NTA chromatography (lane 1 to 6) followed by cation exchange chromatography (Poros HS 50 cation exchange column; lane 7 to 11 and chromatogram above). **Lane 1:** lysate; **lane 2:** flow through, first fraction; **lane 3:** flow through, last fraction; **lane 4:** first wash fraction (10 mM imidazole); **lane 5:** last wash fraction (10 mM imidazole); **lane 6:** eluate (200 mM imidazole); **lane 7:** FPLC flow through; **lane 8:** peak I; **lane 9:** peak II; **lane 10:** peak III; **lane 11:** peak IV. Same amounts of protein (not volume) were analyzed in lane 8 to 11.

The fractions containing SNAP-M.SssI were pooled and concentrated by ultrafiltration. The DNA MTase activity of SNAP-M.SssI with AdoMet (**1**) was tested on λ DNA and compared with wild type M.SssI (Fig. 3.4).

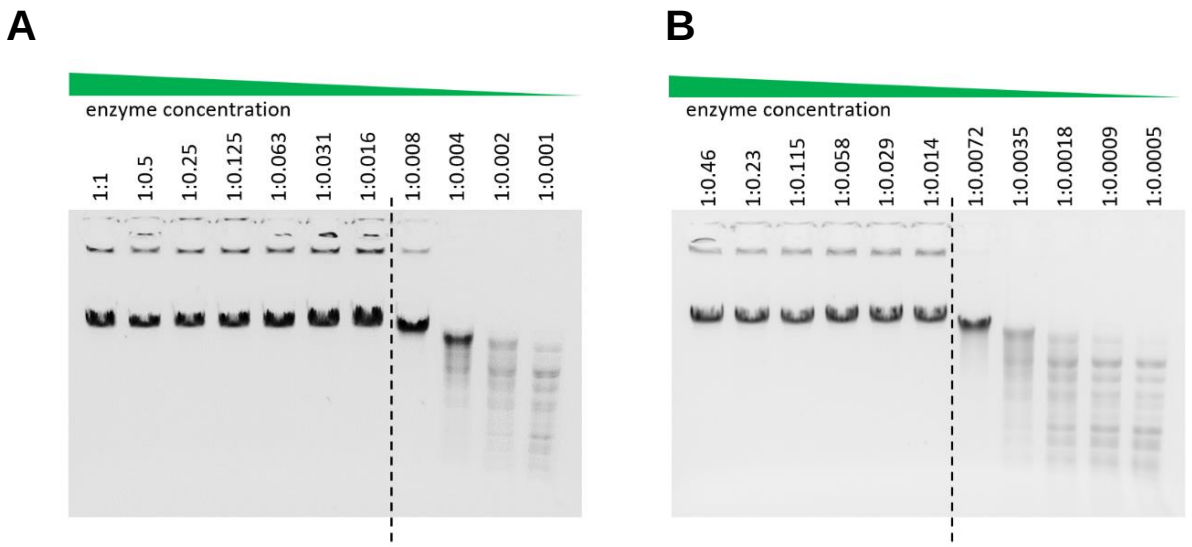


Fig. 3.4: Modification-restriction assay with AdoMet (**1**) of **A** wild type M.SssI starting with 1 eq. of DNA MTase with respect to the number of CpG sites on ds λ DNA and **B** SNAP-M.SssI starting with 0.46 eq. of DNA MTase with respect to the number of CpG sites on ds λ DNA. The enzyme concentration is reduced by half in each lane. Molar ratios are indicated above the gels. After incubation at 37°C for 1 h and inactivation at 65°C for 20 min, R.BstUI is added and incubation continued at 60°C for 1 h.

The assay with wild type M.SssI (Fig. 3.4 A) starts with 1 eq. with respect to the number of CpG sites on ds λ DNA. Up to the seventh lane, the λ DNA is not fragmented which means that 0.016 eq. of wild type M.SssI is sufficient to protect the λ DNA against restriction and wild type M.SssI has thereby an apparent catalytic rate constant k_{app} of 64 h^{-1}. In the assay with SNAP-M.SssI (Fig. 3.4 B), that starts with 0.46 eq., six lanes are fully protected against restriction by R.BstUI. The sixth lane contains 0.014 eq. of SNAP-M.SssI with respect to the number of CpG sites on ds λ DNA. The apparent catalytic rate constant k_{app} of SNAP-M.SssI is 70 h^{-1} and in the same range as k_{app} of wild type M.SssI. From this it can be concluded that the SNAP-tag has no apparent influence on the DNA MTase activity of M.SssI.

To test the SNAP-tag activity, SNAP-M.SssI was incubated with BG-TAMRA (**21a**). The reaction was carried out at 37°C for 30 min in the dark, to protect the fluorophore from light and prevent photobleaching. Fig. 3.5 B shows the analysis by SDS-PAGE.

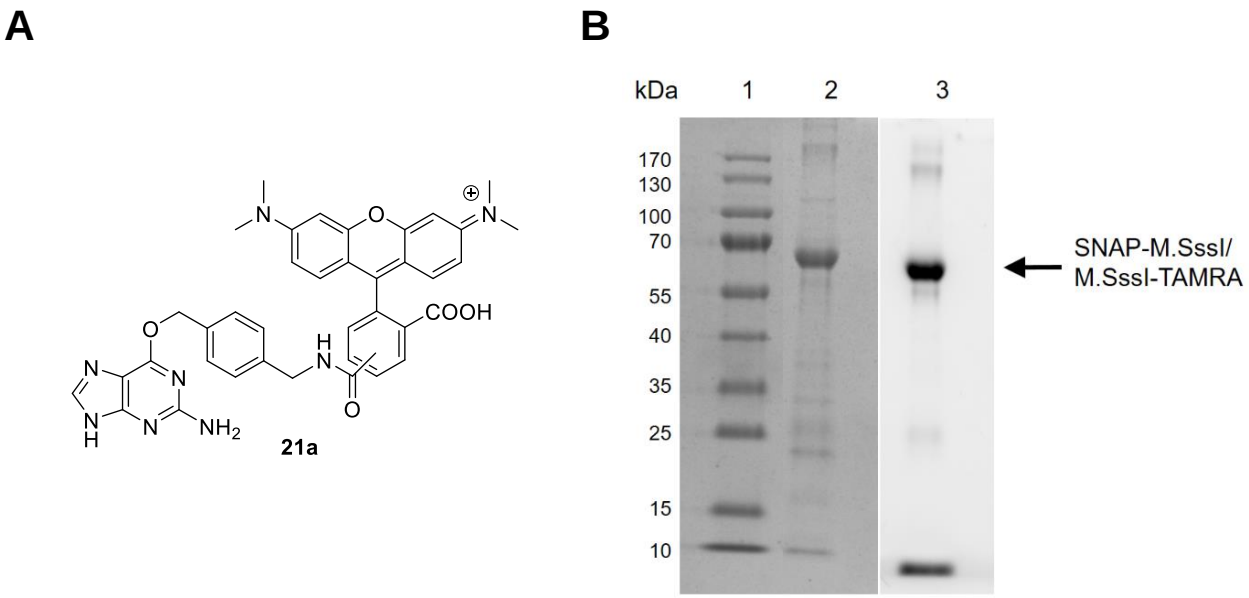
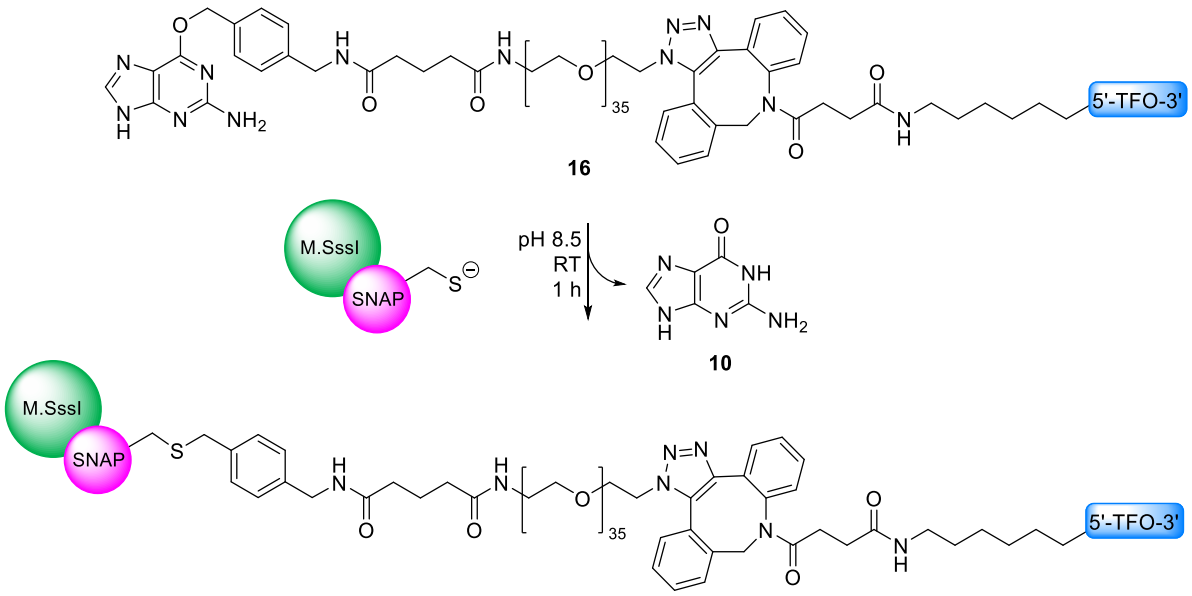


Fig. 3.5: SNAP-tag activity test of SNAP-M.SssI with BG-TAMRA (**21a**). **A:** structure of BG-TAMRA (**21a**). Only one isomer is used for the reaction. **B:** SDS-PAGE analysis. **Lane 1:** marker; **lane 2:** reaction after 30 min, Coomassie stained; **lane 3:** reaction after 30 min, fluorescence detection with a UV-transilluminator (312 nm).

Analysis of the Coomassie stained gel (lane 2) shows one band with the expected molecular mass for SNAP-M.SssI. BG-TAMRA (**21a**) is too small to induce a shift on the SDS-PAGE gel when it is covalently bound to SNAP-M.SssI. However, the rhodamine fluorophore in **21a** can be made visible with a UV illuminator. The band at the position of SNAP-M.SssI that is observed when analyzing the gel under UV light (lane 3) establishes the SNAP-tag activity of SNAP-M.SssI.

3.2.2 Conjugation of SNAP-M.SssI with BG-PEG$_{35}$-TFO



Scheme 3.2: Conjugation of SNAP-M.SssI with BG-PEG$_{35}$-TFO (**16**). Only one isomer is shown.

M.SssI-TFO is constructed by incubating SNAP-M.SssI with an excess of BG-PEG$_{35}$-TFO (**16**) (Scheme 3.2). DTT (1 mM) is added to the buffer to prevent loss of SNAP-tag activity. Analysis of the SNAP-tag reaction at different reaction times (Fig. 3.6) shows that the reaction is very fast. Although the SNAP-tag reaction itself is relatively insensitive to the pH, stability studies of SNAP-M.SssI and M.SssI-TFO show the requirement of a basic environment (pH 8.5) and a high salt concentration of at least 100 mM NaCl. The conversion does not significantly increase after 1 h.

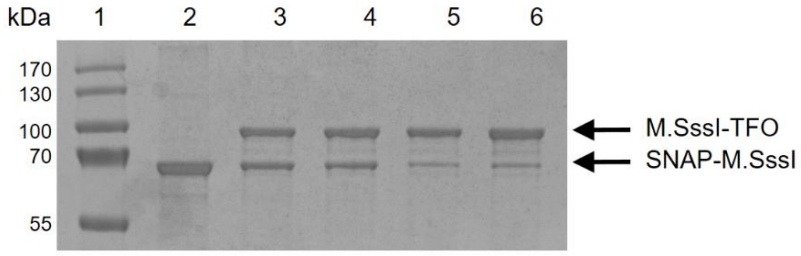
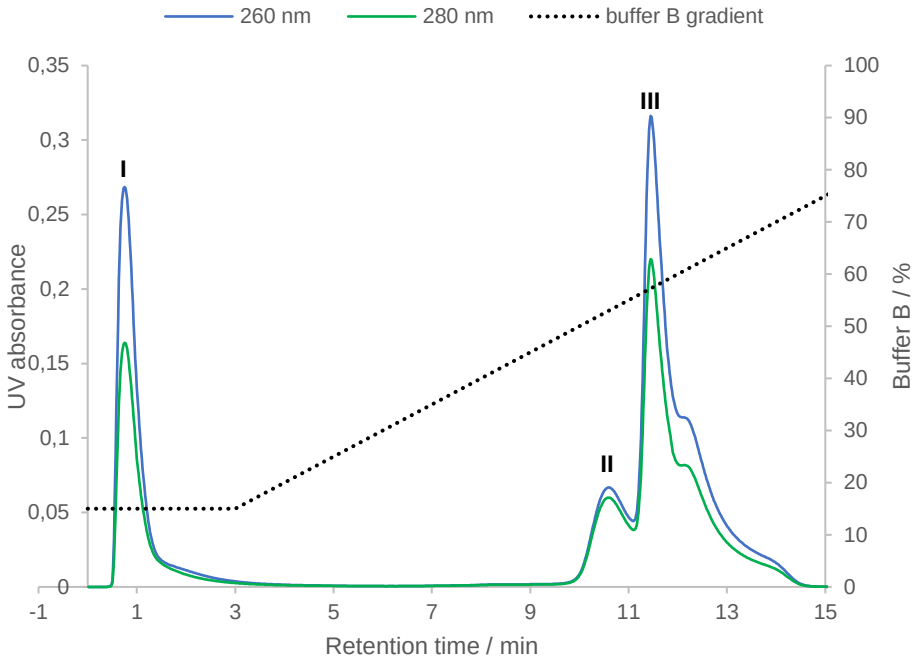


Fig. 3.6: SNAP-tag reaction of SNAP-M.SssI with BG-PEG$_{35}$-TFO (**16**). **Lane 1:** marker; **lane 2:** SNAP-M.SssI; **lane 3:** 0 min; **lane 4:** 10 min; **lane 5:** 30 min; **lane 6:** 60 min.

The product was purified by anion exchange chromatography. Fig. 3.7 shows the HPLC chromatogram and the corresponding SDS-PAGE analysis.

A



B

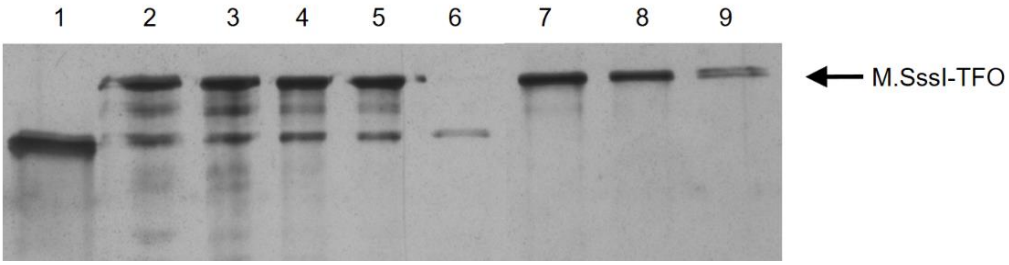


Fig. 3.7: Purification of M.SssI-TFO by anion exchange chromatography (Poros HQ 10 anion exchange column). **A:** HPLC chromatogram. **B:** SDS-PAGE analysis. **Lane 1:** SNAP-M.SssI; **lane 2:** reaction mixture after 1 h; **lane 3:** reaction mixture after 1 h and centrifugation; **lane 4:** reaction mixture after dilution in HPLC buffer; **lane 5:** reaction mixture after dilution in HPLC buffer and centrifugation; **lane 6:** peak I (SNAP-M.SssI); **lane 7-8:** peak II (M.SssI-TFO); **lane 9:** peak III (BG-PEG$_{35}$-TFO (**16**)).

The positively charged SNAP-M.SssI does not bind to the anion exchange column and elutes with the front. M.SssI-TFO and BG-PEG$_{35}$-TFO (**16**) are eluted at higher NaCl concentrations. The excess of BG-PEG$_{35}$-TFO (**16**) cannot be separated completely from the conjugation product. A low yield of less than 5% was observed for M.SssI-TFO. The low yield cannot result from precipitation of the product during the preparation steps, indicated by similar amounts of protein in lanes 2 to 5 on the SDS gel (Fig. 3.7 B). SDS-PAGE analysis of the collected fractions shows the presence of conjugation product in the BG-PEG$_{35}$-TFO-fraction (lane 9). The contamination of the conjugate with BG-PEG$_{35}$-TFO (**16**) is not visible on the SDS gel, but it nevertheless presents a problem for further use of the M.SssI-TFO conjugate.

The purification of M.SssI-TFO was attempted with different combinations of anion and cation exchange columns and different orders of purification steps. The problems of poor recovery and incomplete separation (either of SNAP-M.SssI and M.SssI-TFO or of M.SssI-TFO and BG-PEG$_{35}$-TFO (**16**)) remained and the result described above could not be improved.

3.3 M.MpeI-TFO

The bacterial M.MpeI also methylates cytosine within the CpG recognition sequence. Both enzymes are similar in size and have similar amino acid sequences[164]. In contrast to M.SssI, M.MpeI could be cocrystallized with DNA, which might indicate a higher stability of M.MpeI. This, along with the similarities to M.SssI, were considerations to use a SNAP-tag fusion protein of M.MpeI for the SNAP-tag mediated conjugation with BG-PEG_{35}-TFO.

3.3.1 Construction and characterization of SNAP-M.MpeI

SNAP-M.MpeI was expressed in SG1709 *E.coli* cells and purified. Fig. 3.8 shows the SDS-PAGE analysis of the purified protein after nickel-NTA chromatography.

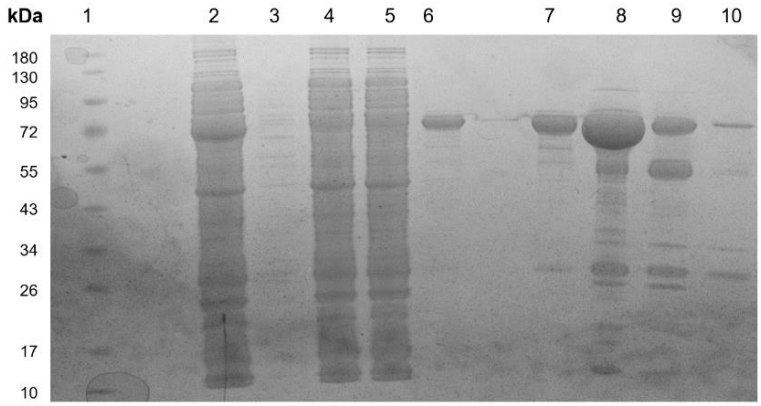


Fig. 3.8: Purification of SNAP-M.MpeI by nickel-NTA chromatography. **Lane 1:** marker; **lane 2:** lysate; **lane 3:** flow through, first fraction; **lane 4:** flow through, last fraction; **lane 5:** first wash fraction (10 mM imidazole); **lane 6:** last wash fraction (10 mM imidazole); **lane 7:** elution with 50 mM imidazole; **lane 8:** elution with 100 mM imidazole; **lane 9:** elution with 150 mM imidazole; **lane 10:** elution with 250 mM imidazole.

The fraction that is eluted with 100 mM imidazole contains most of the desired protein and is concentrated by dialysis. The DNA MTase activity of SNAP-M.MpeI with AdoMet (**1**) is tested on λ DNA and compared with wild type M.MpeI (Fig. 3.9).

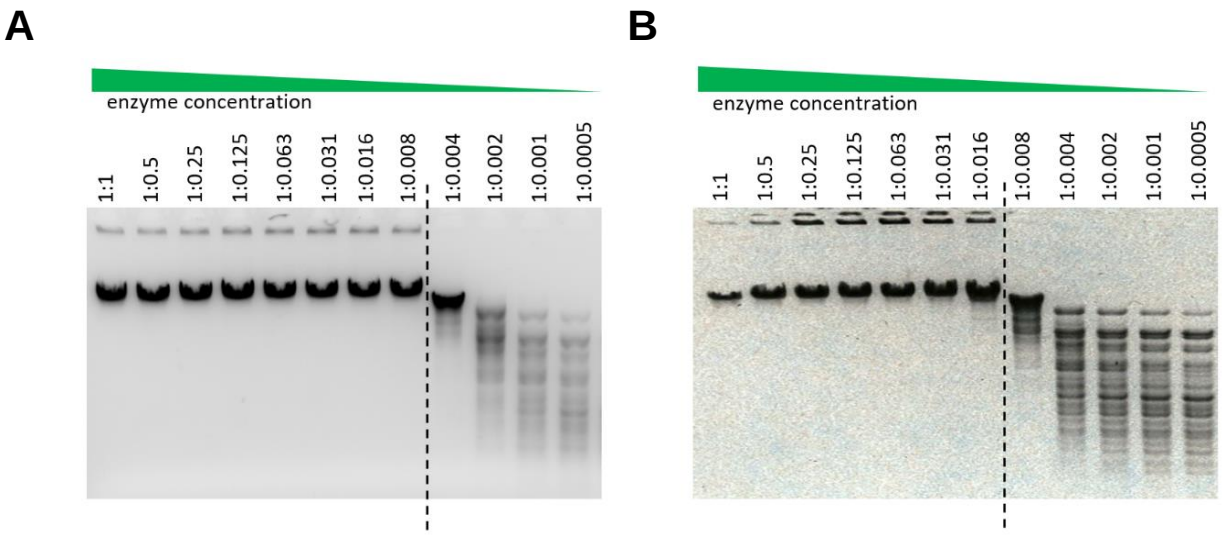
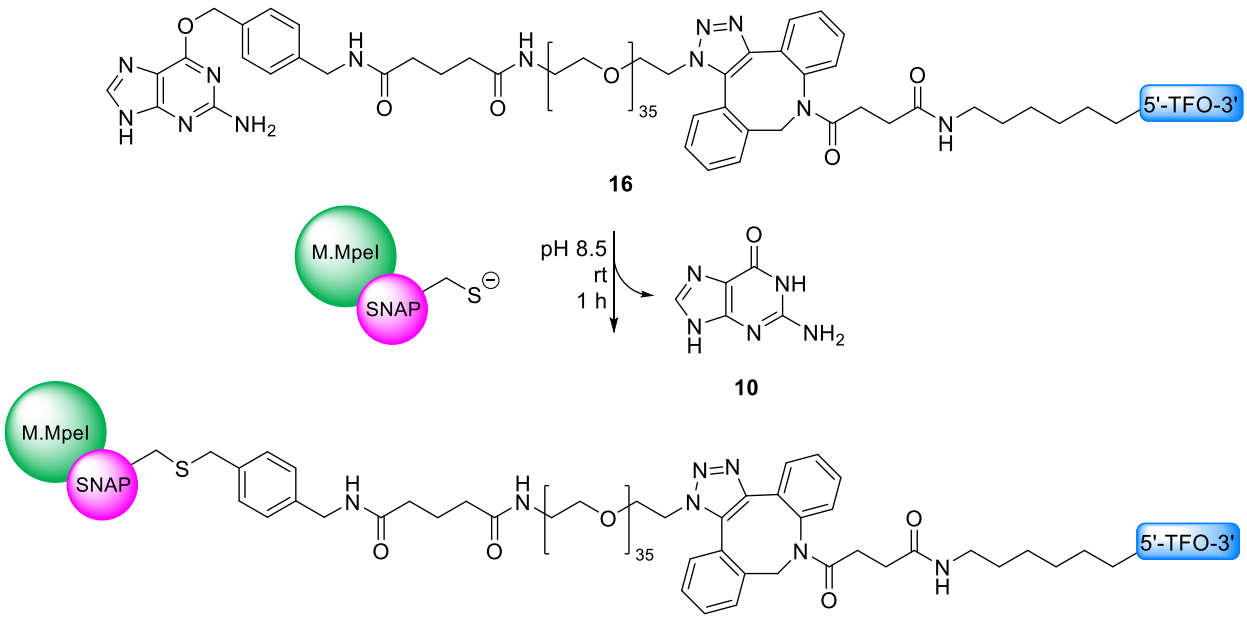


Fig. 3.9: Modification-restriction assay with AdoMet (**1**) of **A** SNAP-M.MpeI and **B** wild type M.MpeI. The first lane contains 1 eq. of DNA MTase with respect to the number of CpG sites on ds λ DNA. The enzyme concentration is reduced by half in each lane. Molar ratios are indicated above the gels. After incubation at 37°C for 1 h and inactivation at 65°C for 20 min, R.BstUI is added and incubation continued at 60°C for 1 h.

0.0078 eq. of SNAP-M.MpeI is sufficient to fully protect λ DNA (k_{app} = 128 h^{-1}) , whereas for full protection by wild type M.MpeI, 0.016 eq. of DNA MTase is needed. The addition of the SNAP-tag to the protein does not seem to significantly alter the DNA MTase activity, as expected from the results obtained with SNAP-M.SssI (Chapter 3.2.1).

3.3.2 Conjugation of SNAP-M.MpeI with BG-PEG$_{35}$-TFO



Scheme 3.3: Conjugation of SNAP-M.MpeI with BG-PEG$_{35}$-TFO (**16**). Only one isomer is shown.

For the conjugation, SNAP-M.MpeI and BG-PEG$_{35}$-TFO (**16**) are incubated at pH 8.5 and rt (Scheme 3.3). The addition of DTT (1 mM) prevents loss of SNAP-tag activity. To facilitate the purification by anion exchange chromatography, only 0.6 eq. of BG-substrate **16** are used. The conversion does not increase after 1 h. To make sure the conjugate does not precipitate before injection onto the anion exchange column, a sample of the reaction mixture is taken between each sample preparation step and analyzed by SDS-PAGE (Fig. 3.10 C).

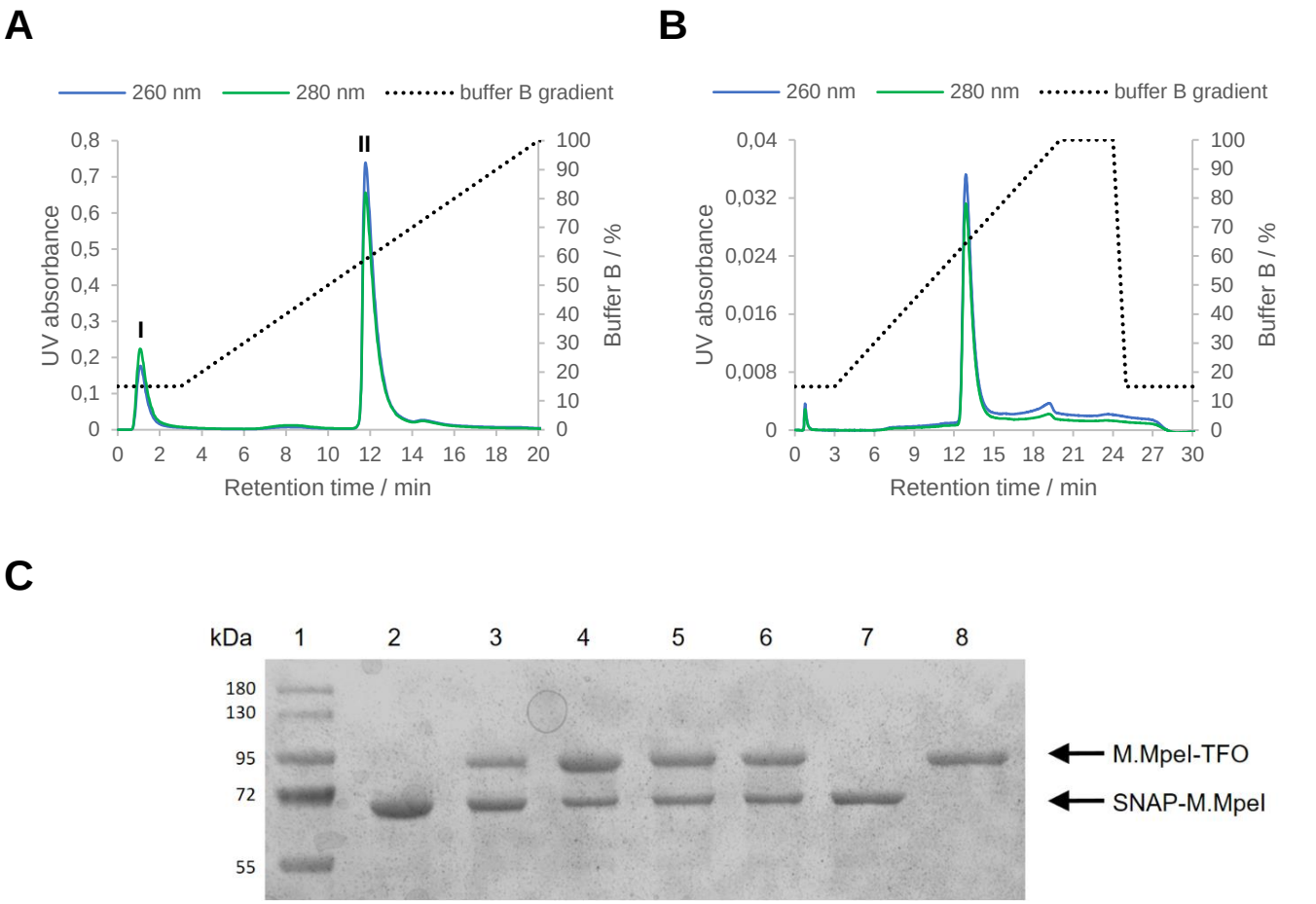


Fig. 3.10: Purification of M.MpeI-TFO by anion exchange chromatography (Poros HQ 10 anion exchange column). HPLC chromatogram of **A** the reaction mixture after 1 h and **B** the product (fraction II). **C:** SDS-PAGE analysis. **Lane 1:** marker; **lane 2:** SNAP-M.MpeI; **lane 3:** reaction mixture after 1 h; **lane 4:** reaction mixture after 1 h and centrifugation; **lane 5:** reaction mixture after dilution in HPLC buffer; **lane 6:** reaction mixture after dilution in HPLC buffer and centrifugation; **lane 7:** fraction I; **lane 8:** fraction II.

Fig. 3.10 shows the anion exchange chromatogram of the reaction mixture after 1 h, the corresponding SDS gel and the analysis of fraction II. Since all BG-PEG$_{35}$-TFO (**16**) has reacted, the conjugate can easily be isolated. SNAP-M.MpeI is positively charged, so the unreacted SNAP-M.MpeI does not bind to the anion exchange column and elutes directly with the front. The negatively charged M.MpeI-TFO is eluted from the column at a higher NaCl concentration.

3.3.3 Characterization of M.MpeI-TFO

M.MpeI-TFO was characterized by its UV-absorbance and its DNA MTase activity on λ DNA. Fig. 3.11 shows the UV absorption spectra of BG-PEG$_{35}$-TFO (**16**), SNAP-M.MpeI and M.MpeI-TFO.

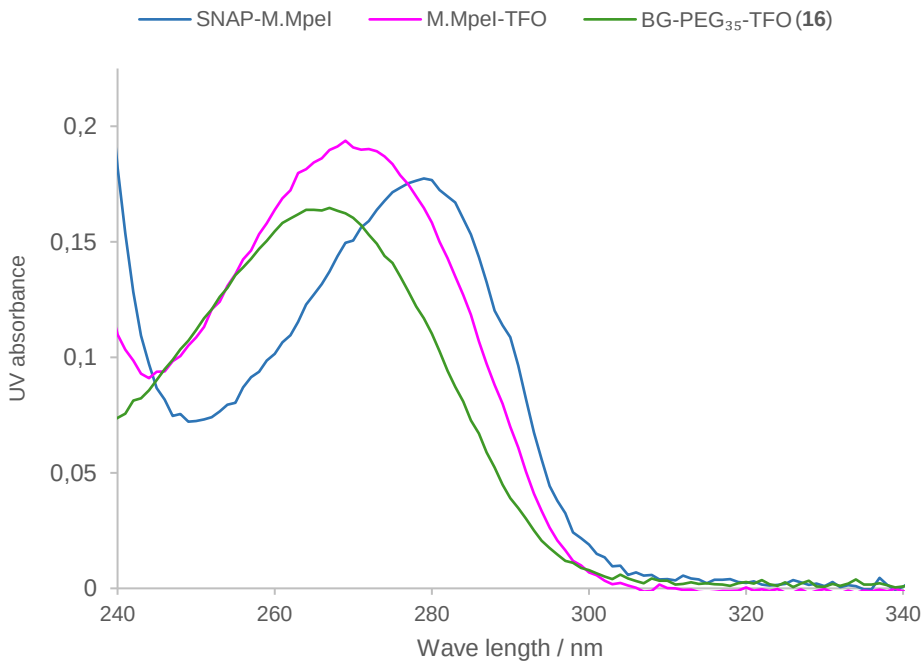


Fig. 3.11: UV absorption spectra of BG-PEG$_{35}$-TFO (**16**, green), SNAP-M.MpeI (blue) and M.MpeI-TFO (magenta).

The absorption spectrum of SNAP-M.MpeI shows a maximum around 280 nm, which is a result of the aromatic amino acids of the protein. DNA typically absorbs at a maximum around 260 nm. The absorption maximum of BG-PEG$_{35}$-TFO (**16**) is slightly higher because of the embodiment of the DBCO-moiety. The M.MpeI-TFO conjugate has an absorption maximum that is in between that of the protein and the DNA component.

The theoretical 260/280 nm UV ratio of M.MpeI-TFO can be calculated with equation (3.1), using the UV ratios of BG-PEG$_{35}$-TFO (**16**) and SNAP-M.MpeI, that can be determined from their absorption spectra shown in Fig. 3.11, and the known extinction coefficients of BG-PEG$_{35}$-TFO (**16**) (ε_{260} = 192300 $M^{-1}\cdot cm^{-1}$) and SNAP-M.MpeI (ε_{280} = 71655 $M^{-1}\cdot cm^{-1}$).

$$\frac{\varepsilon_{260}(M.MpeI-TFO)}{\varepsilon_{280}(M.MpeI-TFO)} = \frac{\varepsilon_{260}(BG-PEG_{35}-TFO)+\varepsilon_{260}(SNAP-M.MpeI)}{\varepsilon_{280}(BG-PEG_{35}-TFO)+\varepsilon_{280}(SNAP-M.MpeI)} \quad (3.1)$$

260/280 nm UV ratios of 1.39, 0.57 and 1.03 are observed for BG-PEG$_{35}$-TFO (**16**), SNAP-M.MpeI and M.MpeI-TFO, respectively. Latter is in line with the calculated 260/280 nm UV ratio of 1.11, which also agrees with the 260/280 nm UV ratio found in the analysis by anion exchange chromatography (Fig. 3.10 A and B).

The DNA MTase activity of M.MpeI-TFO with AdoMet (**1**) was tested on λ DNA in a modification-restriction assay (Fig. 3.12) and compared to that of SNAP-M.MpeI (Fig. 3.9 A).

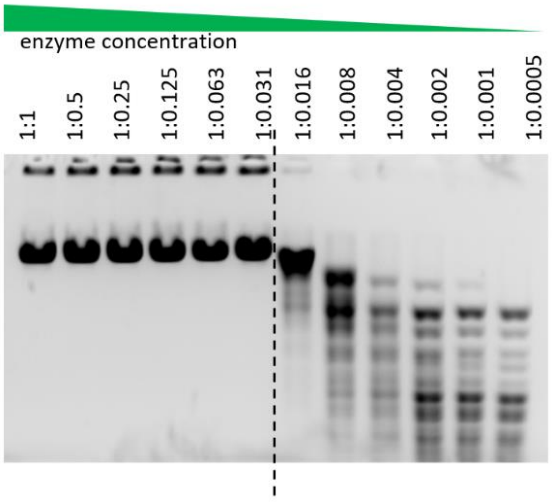


Fig. 3.12: Modification-restriction assay with AdoMet (**1**) of M.MpeI-TFO. The first lane contains 1 eq. of DNA MTase with respect to the number of CpG sites on ds λ DNA. The enzyme concentration is reduced by half in each lane. Molar ratios are indicated above the gel. After incubation at 37°C for 1 h and inactivation at 65°C for 20 min, R.BstUI is added and incubation continued at 60°C for 1 h.

To fully protect λ DNA, 0.03125 eq. of M.MpeI-TFO is needed, which means M.MpeI-TFO has an apparent catalytic rate constant k_{app} of 32 h^{-1}. Herewith, the activity is reduced by a factor of 4 compared to SNAP-M.MpeI (k_{app} = 128 h^{-1}, Fig. 3.9 A). Apparently there is a repulsive effect between the DNA and the negatively charged TFO that could lead to a decrease of DNA MTase activity of the M.MpeI-TFO conjugate.

The DNA MTase activity with AdoMet (**1**) was also tested on λ DNA under the conditions of the DNA methylation specificity assay and with various NaCl concentrations. The agarose gels in Fig. 3.13 show that at a concentration of 200 mM NaCl the activity is slightly affected. At a concentration of 500 mM NaCl, the M.MpeI-TFO cannot methylate the DNA efficiently anymore. Furthermore, this experiment shows that the DNA MTase activity of M.MpeI-TFO is not altered in the actual standard conditions for the DNA methylation specificity assay (Fig. 3.13 A) compared to the conditions for the activity test (Fig. 3.12).

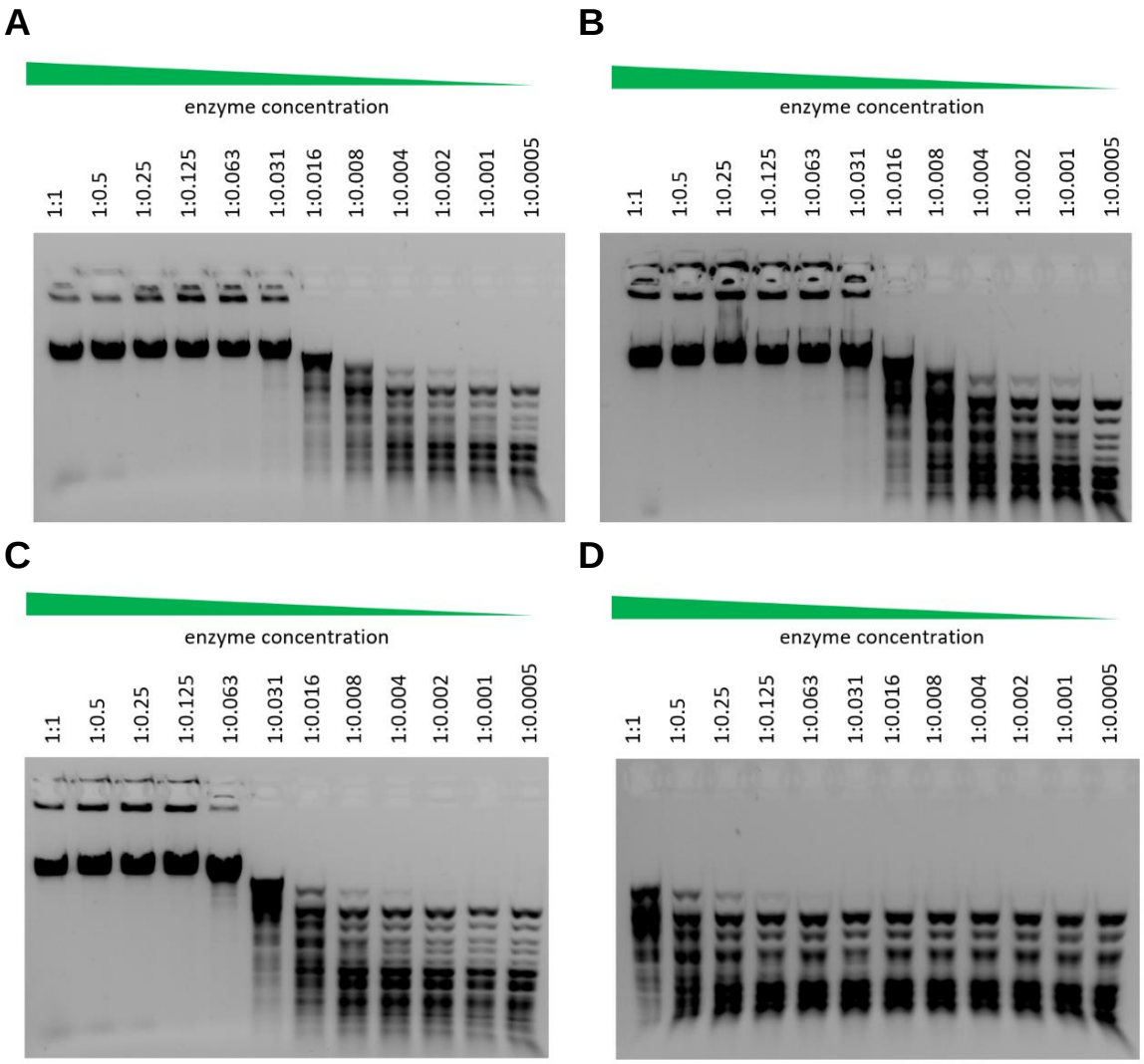


Fig. 3.13: Modification-restriction assay on λ DNA of M.MpeI-TFO with AdoMet (**1**) in a buffer with **A** 50 mM **B** 100 mM **C** 200 mM and **D** 500 mM NaCl, in addition to $MgCl_2$ (10 mM) and TRIS-HCl (10 mM), pH 7.5. Each first lane contains 1 eq. of DNA MTase with respect to the number of CpG sites on ds λ DNA. The enzyme concentration is reduced by half in each lane. After incubation at 30°C for 1 h and inactivation at 65°C for 20 min, R.BstUI is added and incubation continued at 60°C for 1 h.

3.4 M.MpeI(Q142L)-TFO

A variant of M.MpeI, M.MpeI(Q142L), has also been fused to the SNAP-tag protein. It is the M.MpeI equivalent of the M.SssI(Q147L) variant, in which a glutamine in the catalytic domain is replaced with a leucine residue. M.SssI(Q147L) binds to the DNA with lower affinity and has a reduced DNA MTase activity[165]. M.MpeI(Q142L) is expected to have the same properties. SNAP-M.MpeI(Q142L) was also reacted with BG-PEG_{35}-TFO (**16**) to construct a low affinity conjugate.

3.4.1 Construction and characterization of SNAP-M.MpeI(Q142L)

SNAP-M.MpeI(Q142L) was expressed in SG1709 *E.coli* cells and purified. A comparison of the purification of the mutant to that of SNAP-M.MpeI (Fig. 3.14 and Fig. 3.8, respectively) shows that, in the case of SNAP-M.MpeI(Q142L), the desired protein is more concentrated in the 100 mM imidazole fraction and this fraction contains less impurities. This ultimately leads to a higher yield (20.6 mg SNAP-M.MpeI(Q142L) from 400 mL culture vs. 19.7 mg SNAP-M.MpeI from 600 mL culture).

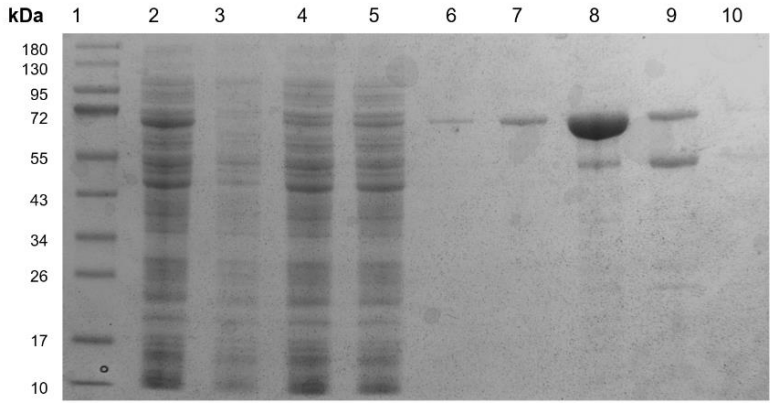


Fig. 3.14: Purification of SNAP-M.MpeI(Q142L) by nickel-NTA chromatography. **Lane 1:** marker; **lane 2:** lysate; **lane 3:** flow through, first fraction; **lane 4:** flow through, last fraction; **lane 5:** first wash fraction (10 mM imidazole); **lane 6:** last wash fraction (10 mM imidazole); **lane 7:** elution with 50 mM imidazole; **lane 8:** elution with 100 mM imidazole; **lane 9:** elution with 150 mM imidazole; **lane 10:** elution with 250 mM imidazole.

The DNA MTase activity of SNAP-M.MpeI(Q142L) was tested in a modification-restriction assay (Fig. 3.15). As expected, the DNA MTase activity is highly reduced compared to that of SNAP-M.MpeI (k_{app} = 1 h^{-1} vs. 128 h^{-1}) and SNAP-M.MpeI(Q142L) is not able to protect λ DNA significantly.

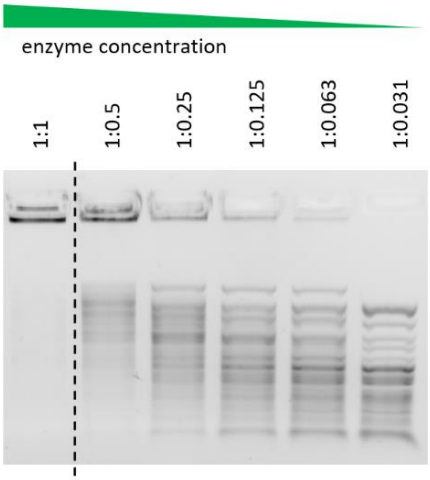
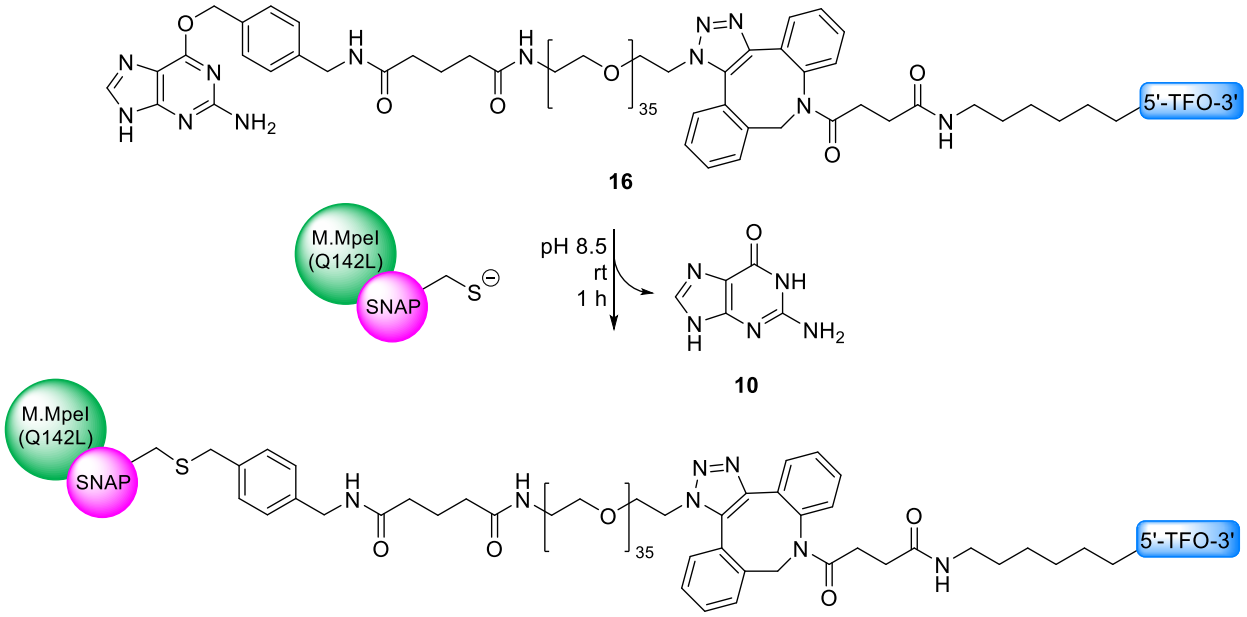


Fig. 3.15: Modification-restriction assay with AdoMet (**1**) of SNAP-M.MpeI(Q142L). The first lane contains 1 eq. of DNA MTase with respect to the number of CpG sites on ds λ DNA. The enzyme concentration is reduced by half in each lane. After incubation at 37°C for 1 h and inactivation at 65°C for 20 min, R.BstUI is added and incubation continued at 60°C for 1 h.

3.4.2 Conjugation of SNAP-M.MpeI(Q142L) with BG-PEG$_{35}$-TFO

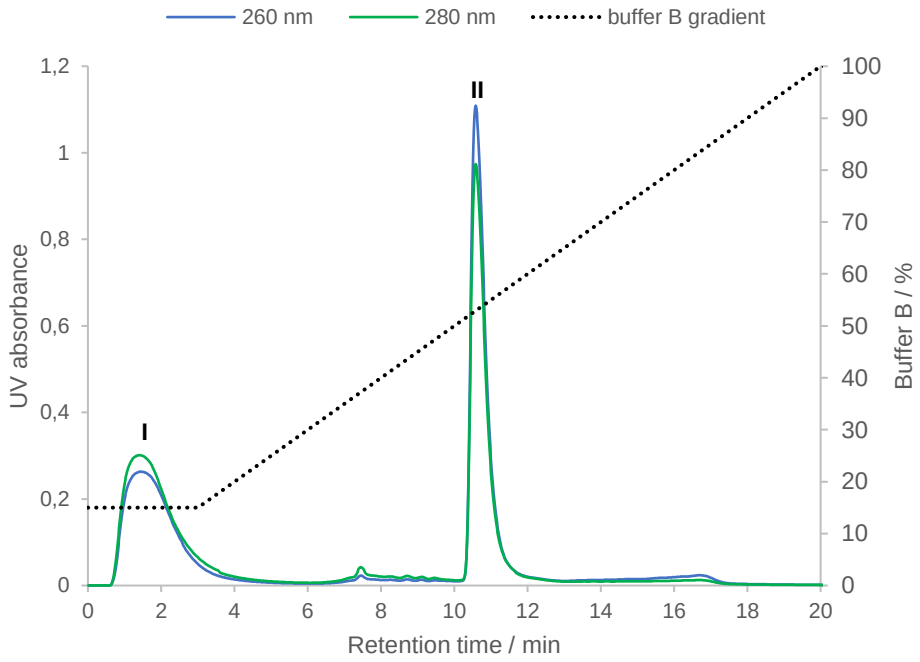


Scheme 3.4: Conjugation of SNAP-M.MpeI(Q142L) with BG-PEG$_{35}$-TFO (**16**). Only one isomer is shown.

SNAP-M.MpeI(Q142L) was incubated with BG-PEG$_{35}$-TFO (**16**) at pH 8.5 and rt (Scheme 3.4). Loss of SNAP-tag activity was prevented by the addition of DTT (1 mM). After 1 h the reaction was terminated and the purification is analogous to that of M.MpeI-TFO. SDS-PAGE analysis of the reaction mixture

(Fig. 3.16 B) shows that the product does not precipitate before injection onto the anion exchange column. As before, usage of substoichiometric amounts of BG-PEG$_{35}$-TFO (**16**) circumvents the problem of having to separate unreacted BG-substrate **16** from the product (Fig. 3.16 A).

A



B

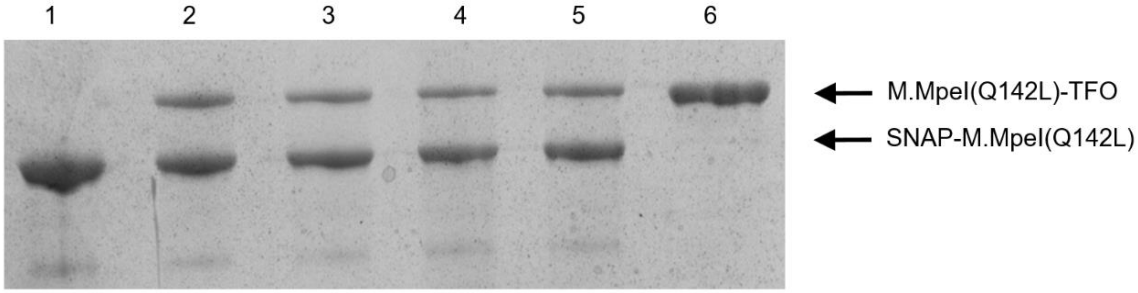


Fig. 3.16: Purification of M.MpeI(Q142L)-TFO by anion exchange chromatography (Poros HQ 10 anion exchange column). **A:** HPLC chromatogram. **B:** SDS-PAGE analysis. **Lane 1:** SNAP-M.MpeI(Q142L); **lane 2:** reaction mixture after 1 h; **lane 3:** reaction mixture after 1 h and centrifugation; **lane 4:** reaction mixture after dilution in HPLC buffer; **lane 5:** reaction mixture after dilution in HPLC buffer and centrifugation; **lane 6:** fraction II.

3.4.3 Characterization of M.MpeI(Q142L)-TFO

The DNA MTase activity of M.MpeI(Q142L)-TFO was tested in a modification-restriction assay. Fig. 3.17 shows that the M.MpeI(Q142L)-TFO has a reduced activity with λ DNA compared to SNAP-M.MpeI(Q142L) (Fig. 3.15). Stoichiometric amounts of M.MpeI(Q142L)-TFO with respect to the number of CpG sites only partially protect λ DNA against fragmentation. As before with M.MpeI-TFO, the reduction of activity for M.MpeI(Q142L)-TFO can be explained by a further reduction of binding affinity due to the presence of the negatively charged TFO.

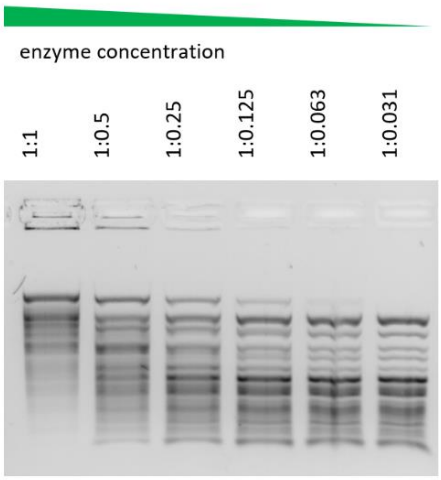


Fig. 3.17: Modification-restriction assay with AdoMet (**1**) of M.MpeI(Q142L)-TFO. The first lane contains 1 eq. of DNA MTase with respect to the number of CpG sites on λ DNA. The enzyme concentration is reduced by half in each lane. After incubation at 37°C for 1 h and inactivation at 65°C for 20 min, R.BstUI was added and incubation continued at 60°C for 1 h.

3.5 Triple helix formation

To see if the designed TFO can in fact form a triple helix with the target sequence inserted into the plasmid DNA, triplex formation was analyzed with PAGE (Fig. 3.18). For this purpose, duplex DNA was formed by hybridization of the 54+ and the 54- ODN by incubation at 95°C for 3 min and then slowly cooling to rt within 1 h. Next, amino-TFO (**17**) was added and the mixture was incubated at rt overnight. Triplex formation was investigated with different ratios of duplex DNA to amino-TFO (**17**) (1:1, 1:2 and 1:5).

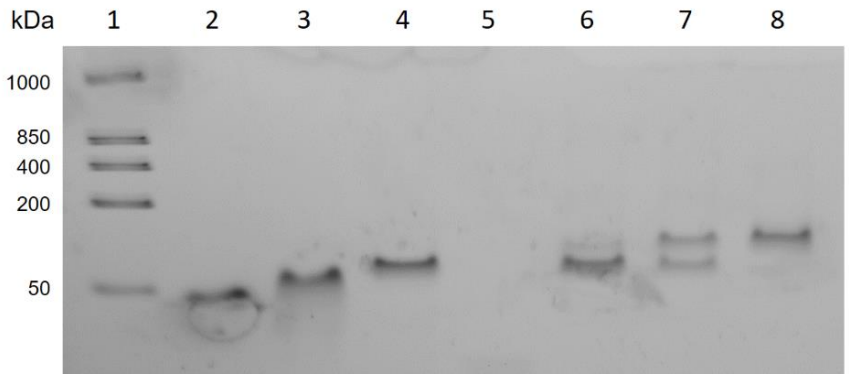


Fig. 3.18: PAGE analysis of triplex helix formation of amino-TFO (**17**) with 54+/54- duplex DNA. **Lane 1:** marker; **lane 2:** 54+ ODN; **lane 3:** 54- ODN; **lane 4:** dsDNA; **lane 5:** Amino-TFO (**17**); **lane 6:** dsDNA (200 nM) + **17** (200 nM); **lane 7:** dsDNA (200 nM) + **17** (400 nM); **lane 8:** dsDNA (200 nM) + **17** (1000 nM).

The 24 nucleotides long amino-TFO (**17**) is not visible on the gel (Fig. 3.18, lane 5). The hybridized duplex DNA (lane 4) migrates slower through the gel than the single-stranded ODN 54+ and 54- (lane 2 and 3). The addition of the third strand to the duplex DNA to form a triplex leads to an even slower fmigration in the gel. In case of a 1 to 1 ratio of duplex DNA to amino-TFO (**17**), where both components have a concentration of 200 nM, there is hardly any triple helix formed (lane 6) and no electromobility shift is observed. About 50% of duplex DNA has formed a triple helix with the amino-TFO (**17**) when the concentration of amino-TFO (**17**) is 400 nM (lane 7). When using 1000 nM amino-TFO (**17**), quantitative triple helix formation is observed (lane 8).

It must be considered that this experiment is merely to establish the ability of amino-TFO (**17**) to form a triple helix with the target DNA sequence. Although the same buffer (10 mM TRIS-HCl, 50 mM NaCl, 10 mM $MgCl_2$, pH 7.5) is used, the duplex DNA is different from the plasmid DNA used in the DNA methylation specificity assay but is required to visualize triple helix formation by PAGE.

3.6 Specificity of targeted DNA methylation by M.MpeI-TFO conjugates

The methylation specificity by the DNA MTase-TFO conjugates was tested *in vitro* in a modification-restriction assay on a plasmid that holds the triple helix-forming site (TFS). This plasmid, called Litcon54, was originally designed to test the methylation specificity of the M.SssI(C141S)-TFO conjugate that was constructed by maleimide crosslinking[161, 162]. The same TFO sequence was used for the M.SssI(C141S)-TFO conjugate, making the Litcon54 plasmid compatible with the conjugates in this work as well.
Litcon54 has four recognition sites for the CpG methylation-sensitive REase R.Psp1406I (AACGTT), of which one is closely located to the TFS and holds the targeted CpG. If there is no protection by methylation, R.Psp1406I cuts the plasmid four times (Fig. 3.19, left), resulting into four DNA fragments. However, if the DNA MTase-TFO forms a triple helix with the TFS and can methylate site A, digestion with R.Psp1406I results in only three DNA fragments (Fig. 3.19, right).

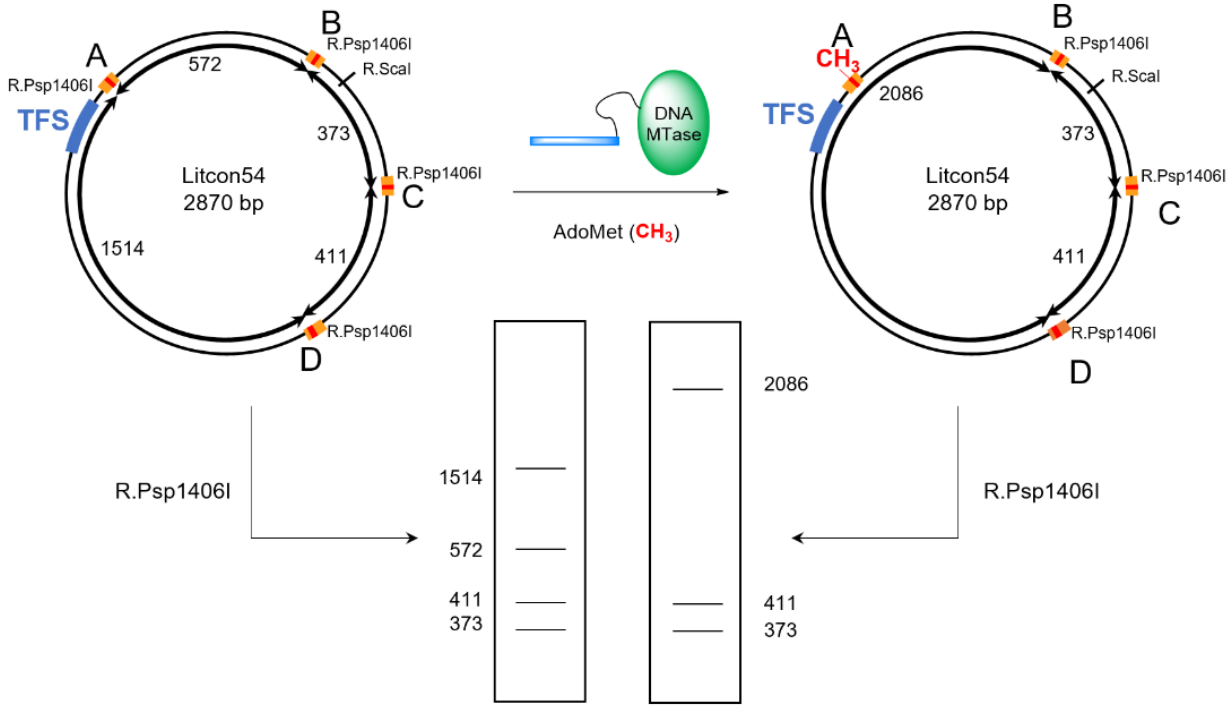
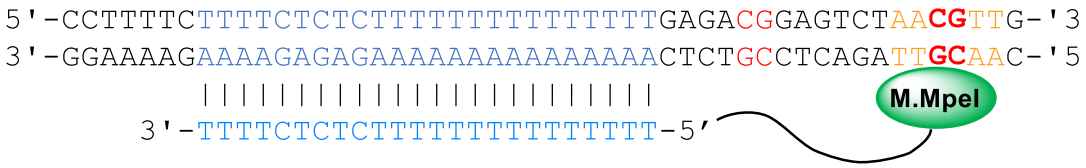


Fig. 3.19: Assay to test targeted DNA methylation specificity. Expected fragmentation pattern **left:** before and **right:** after targeted methylation by the M.MpeI-TFO conjugate, when restricted with R.Psp1406I.

As a negative control, the experiment was carried out with Litcon47 plasmid. This plasmid has the same structure as Litcon54, except that it does not hold the TFS (Fig. 3.20). The TFO can not bind to the DNA and thus can not direct the DNA MTase to the addressed site. As a result, no specific methylation pattern is expected.

Litcon54:



Litcon47:

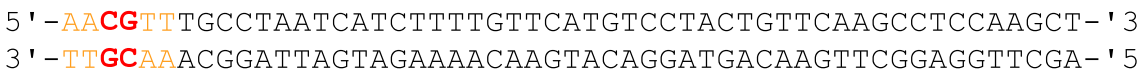


Fig. 3.20: Inserts of Litcon54 and Litcon47. Triple helix formation and thus specific methylation of the target site (red and bold) is only possible with Litcon54, as it contains the TFS (dark blue) for the TFO (light blue).

3.6.1 *In vitro* methylation of Litcon54 and Litcon47 by M.MpeI-TFO

DNA methylation specificity of the M.MpeI-TFO conjugate is tested not only with supercoiled (∞), but also with open-circular (O) and linear (–) Litcon54. The open-circular and linear forms are generated by incubating supercoiled plasmid DNA with Nb.BsmI (cleaving only one strand) and R.ScaI (cleaving both strands), respectively (Fig. 3.21 C). After the modification-restriction assay, supercoiled and open-circular forms are incubated with ScaI as well, so all possible fragments would have the same size in all three methylation assays (Fig. 3.21 A). The additional cleavage of the 373 bp fragment in two smaller fragments (252 bp and 121 bp) by R.ScaI leads to a fragmentation pattern of five bands when the plasmid is not protected and to a pattern of four bands when the plasmid is protected at the targeted site A (Fig. 3.21 B).

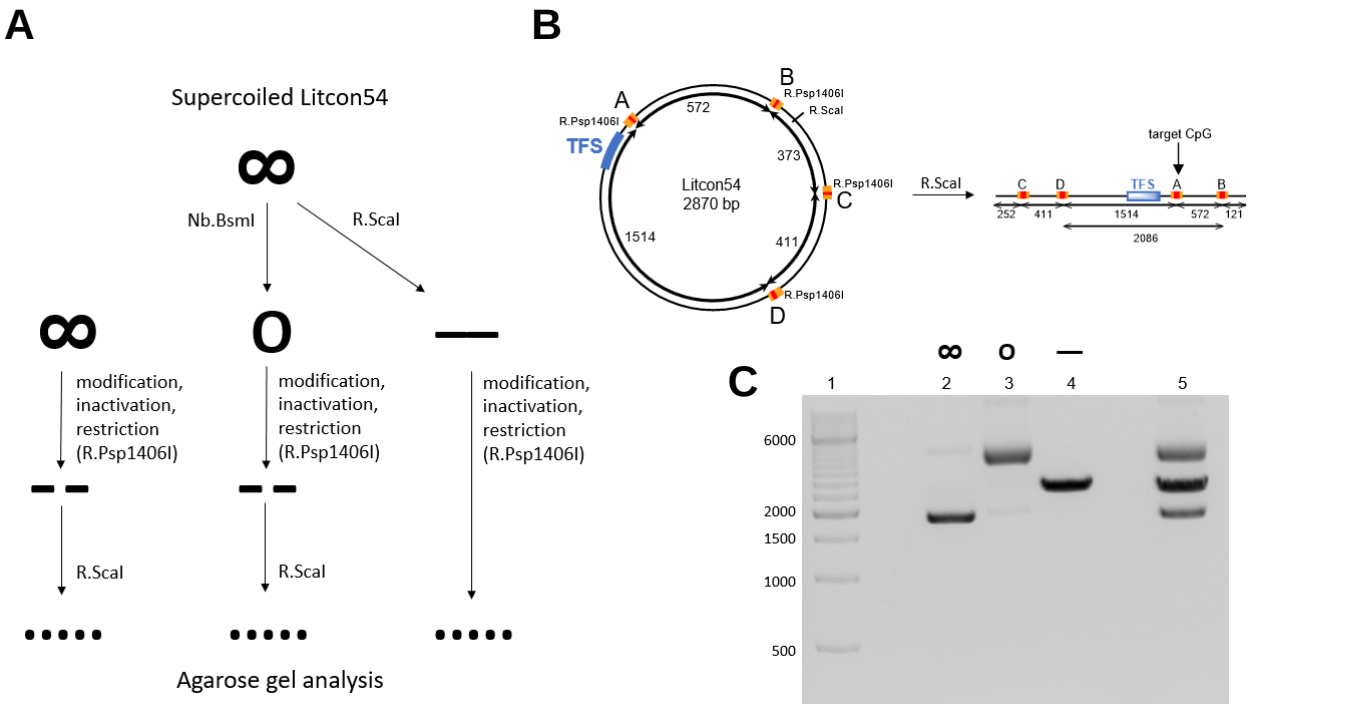


Fig. 3.21 A: Experimental setup to assay DNA methylation specificity of M.MpeI-TFO on supercoiled (∞), open-circular (O) and linear (–) Litcon54. **B:** Linearization of Litcon54 with R.ScaI. **C:** Agarose gel analysis of **lane 2:** supercoiled (∞), **lane 3:** open-circular (O) and **lane 4:** linear (–) Litcon54. **Lane 1:** marker; **lane 5:** Loading a small amount of each DNA form in one single lane.

The agarose gel analysis in Fig. 3.22 A shows the fragmentation patterns of the three different Litcon54 forms after modification with M.MpeI-TFO and subsequent restriction with R.Psp1406I and, in the case of open-circular and supercoiled plasmid forms, with additional R.ScaI digestion. Lane 5 shows the methylation pattern of linear Litcon54 that is fragmented with R.Psp1406I. Binding of the conjugate does not influence fragmentation at the target CpG.

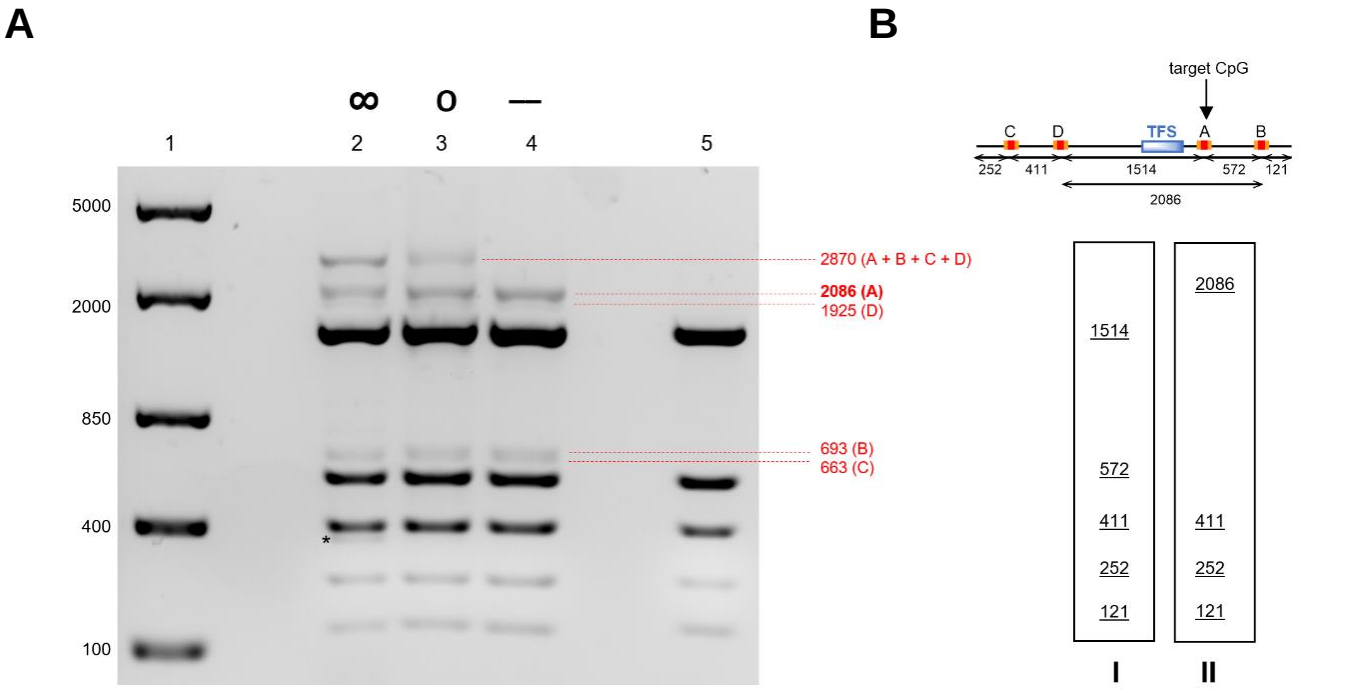


Fig. 3.22: DNA methylation specificity assay of M.MpeI-TFO with Litcon54. **A:** Agarose gel analysis. **Lane 1:** marker; methylation of **lane 2:** supercoiled (∞), **lane 3:** open-circular (O) and **lane 4:** linear (–) Litcon54 (1 h at 30°C, followed by 20 min inactivation at 65°C) and consecutive restriction with R.Psp1406I (1 h at 37°C) and R.ScaI (lane 2 and 3, 1 h at 37°C); **lane 5:** unmethylated Litcon54 restricted with R.Psp1406I and R.ScaI. The bands that are generated by monoprotection of the respective CpG sites or protection of all 4 challenged sites are indicated with red dashed lines and the corresponding fragment size. The band marked with an asterisk in lane 2 is the 373 bp fragment due to incomplete linearization with R.ScaI. **B:** Expected fragmentation patterns of unprotected (I) and site A-specifically protected (II) linear Litcon54 after restriction with R.Psp1406I.

In general, the methylation rate is quite low, and a small fraction of the plasmid gets protected. Specific DNA methylation is indicated by the 2086 bp fragment, that combines the 1514 bp and the 572 bp fragment due to protection of the targeted CpG (site A). The 2086 bp fragment is observed in all three forms and is the strongest protected band in the sample with linear Litcon54 (lane 4). Although fragments that come from single protection of the off-target CpG sites (i.e. the 663 bp, the 693 bp and the 1925 bp fragment) are visible in all three forms, the completely protected 2870 bp fragment is only visible in the

samples with supercoiled and open-circular plasmid (lane 2 and lane 3, respectively). The methylation specificity appears to be depending on the topological state of the Litcon54 and is the highest with linear Litcon54.

The methylation of linear Litcon54 is investigated in more detail by analyzing the reaction after different incubation times. To see whether preincubation has an effect on binding of the TFO to the TFS and on the methylation specificity, M.MpeI-TFO is incubated with linear Litcon54 at 30°C for 1 h before adding AdoMet (**1**). The agarose gel analysis is shown in Fig. 3.23.

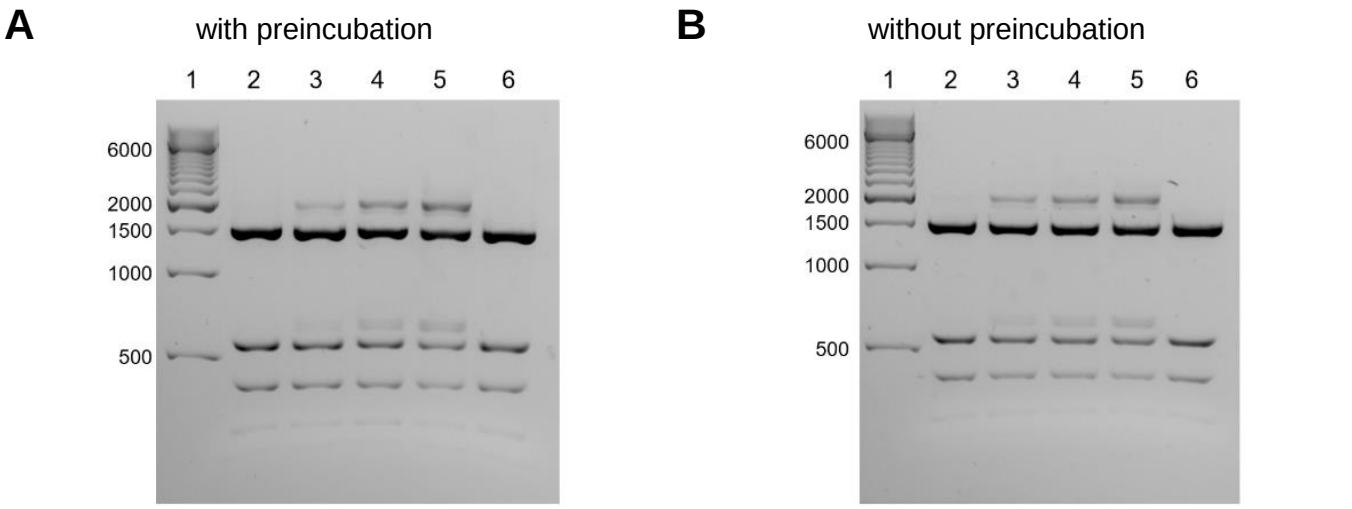


Fig. 3.23: DNA mespecificity assay of M.MpeI-TFO with linear Litcon54 after different incubation times with **A** preincubation and **B** no preincubation of M.MpeI-TFO and linear Litcon54 (1 h at 30°C). **Lane 1:** Marker; methylation of linear Litcon54 at 30°C after **lane 2:** 0 min, **lane 3:** 10 min, **lane 4:** 30 min, **lane 5:** 60 min, followed by 20 min inactivation at 65°C and consecutive restriction with R.Psp1406I (1 h at 37°C); **lane 6:** unmethylated linear Litcon54 restricted with R.Psp1406I.

Preincubation of M.MpeI-TFO with linear Litcon54 does not have a significant effect on the specificity. In both cases, the fragment resulting from specific methylation can be observed after 10 min. Single-protected fragments resulting from protection at non-addressed sites also appeared after 10 min of methylation with M.MpeI-TFO. Both specific methylation and non-specific methylation increase within the time span of 1 h. As before, no multiple-protected fragments can be observed with linear Litcon54.

To further evaluate the specificity of the M.MpeI-TFO conjugate, the methylation specificity of M.MpeI-TFO is assayed with Litcon47, a plasmid that lacks the site for triple helix formation (Fig. 3.20). The experiment is carried out on both supercoiled and linear Litcon47, the extreme scenarios with respect to the specificity as concluded from Fig. 3.22. The result is compared with the methylation specificity of M.MpeI-TFO with Litcon54 (Fig. 3.24).

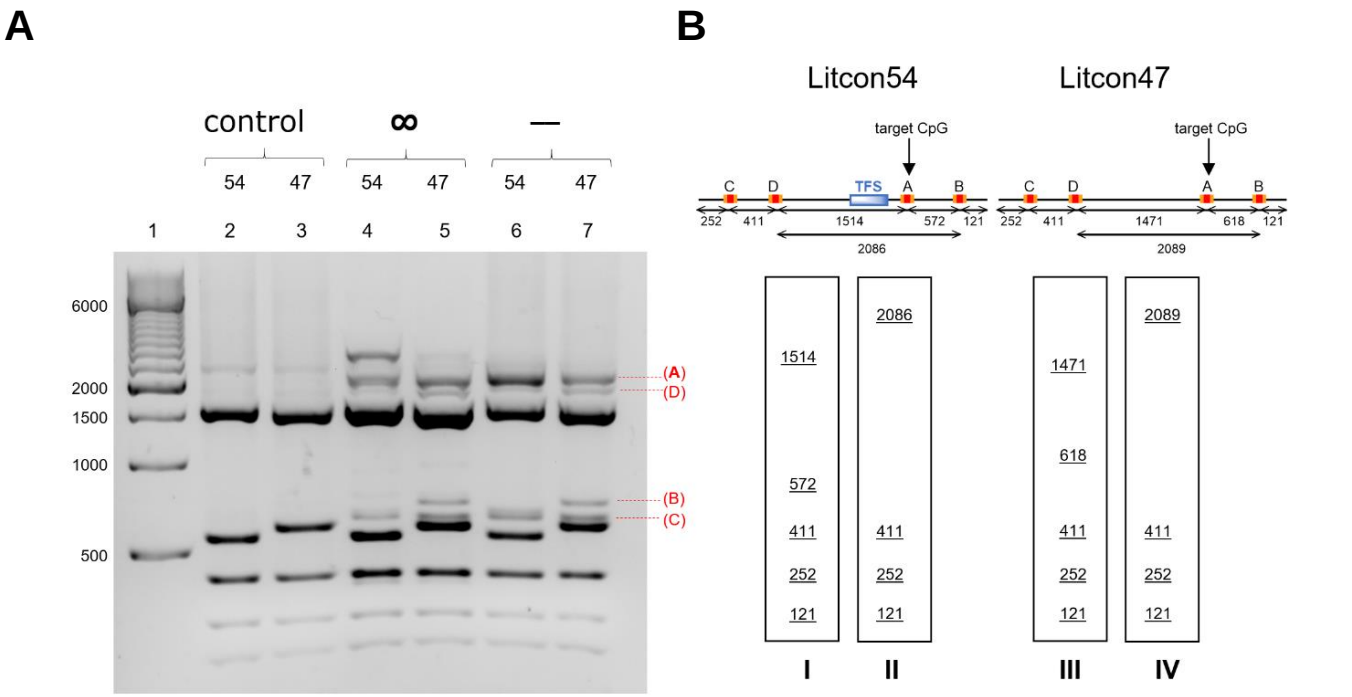


Fig. 3.24: DNA methylation specificity assay of M.MpeI-TFO with Litcon54 and Litcon47. **A:** Agarose gel analysis. **Lane 1:** marker; **lane 2:** restriction control of linear Litcon54; **lane 3:** restriction control of linear Litcon47; **lane 4:** supercoiled Litcon54; **lane 5:** supercoiled Litcon47; **lane 6:** linear Litcon54; **lane 7:** linear Litcon47 (red letters mark the bands that are generated by monoprotection of the respective CpG sites). **B:** Expected fragmentation patterns of unprotected (I) and site A-specifically protected (II) linear Litcon54 and unprotected (III) and site A-specifically protected (IV) linear Litcon47 after restriction with R.Psp1406I and, in the case of supercoiled and open-circular forms, R.ScaI..

In linear Litcon54 (lane 6), the fragment that is generated by methylation of the targeted CpG site has the highest intensity of all the generated fragments. Although in linear Litcon47 (lane 7) the corresponding band is less pronounced, the intensity is higher compared to that of the band that is generated by protection of site D. The fragments are similar in size and the intensities on the gel should also be similar. This is remarkable, as it means that there is a preference towards site A compared to site D. Thus, a comparison of the fragments generated by protection at site A and D is not very meaningful. Comparing fragment (A) with the fragments (B) and (C) and taking into consideration that the band intensity also depends on the

fragment size, shows that there is a preference towards site A with linear Litcon54 (lane 6), whereas with linear Litcon47 (lane 7) there is not. The fragment in supercoiled Litcon54 (lane 4) with the lowest mobility appears to have multiple protection at all challenged sites, as observed in Fig. 3.22 (lane 2). This full-length band is not observed with supercoiled Litcon47. The samples for supercoiled and linear Litcon47 (lane 5 and lane 7, respectively) look similar.

The fragments observed in the experiments have different origins. In the case of Litcon54, a part of the M.MpeI-TFO is not bound to the TFS and can methylate non-specifically, resulting in single-protected fragments observed in all three forms. In the case of Litcon47, where there is no TFS for M.MpeI-TFO to form a triple helix, the higher amount of unbound M.MpeI-TFO results in more non-specific mono-methylation. Immobilized M.MpeI-TFO methylates at the target site (A), but also leads to intramolecular non-specific methylation, resulting in multiple-protected fragments. Intramolecular non-specific methylation appears to be depending on immobilized M.MpeI-TFO, as there is very little multiple-methylation observed in the case of supercoiled Litcon47 where there is no immobilized M.MpeI-TFO. Multiple-protected Litcon54 could be achieved *via* a mechanism, in which the bound M.MpeI-TFO reaches over to the CpG sites B, C and D. This would be less likely to occur in linear Litcon54 due to the higher flexibility of the DNA ends compared to the supercoiled and open-circular forms. Another possibility could be that the immobilized M.MpeI-TFO starts by methylating site A and, without dissociating from the DNA, goes around in a circle, consecutively protecting site B, C and D against fragmentation by R.Psp1406I. Such a processive mechanism would not lead to protection of all four sites in linear Litcon54, because the challenged CpG sites are in front of and behind the TFS (Fig. 3.21 B). Processive methylation was observed with M.SssI in the absence of Mg^{2+}[61]. Because of the similarity between M.MpeI and M.SssI and the fact that there is Mg^{2+} present in the assay, a processive mechanism is less likely to be the case.

In conclusion, preferential methylation of the target site is present in all three conformational forms of Litcon54 DNA. Non-specific methylation, that is also observed in all forms, is either intermolecular methylation that results from free M.MpeI-TFO or intramolecular methylation, which is a result from bound M.MpeI-TFO. The specificity of targeted DNA methylation is the highest in linear Litcon54, as here the non-specific intramolecular methylation is suppressed.

3.6.2 *In vitro* methylation of Litcon54 by M.MpeI-TFO under various NaCl concentrations

The modification-restriction experiments on λ DNA showed that the DNA MTase activity of M.MpeI-TFO decreases when the NaCl concentration is increased (Chapter 3.3.3, Fig. 3.13). This is expected because the binding affinity of the DNA MTase should be reduced in a more salt-rich environment. A decreased DNA MTase binding affinity could be advantageous for specific DNA methylation at the target site, as DNA methylation will be more dominated by the TFO binding and off-target methylation should be reduced. The *in vitro* methylation of linear Litcon54 by M.MpeI-TFO is tested under various NaCl concentrations, to see if an increased NaCl concentration has a positive effect on the specificty of M.MpeI-TFO towards the target site. Linear Litcon54 is incubated with M.MpeI-TFO in buffers that contain the standard 50 mM NaCl, but also 100 mM and 200 mM NaCl, in addition to TRIS-HCl (10 mM) and $MgCl_2$ (10 mM). Fig. 3.25 shows the agarose gel analysis.

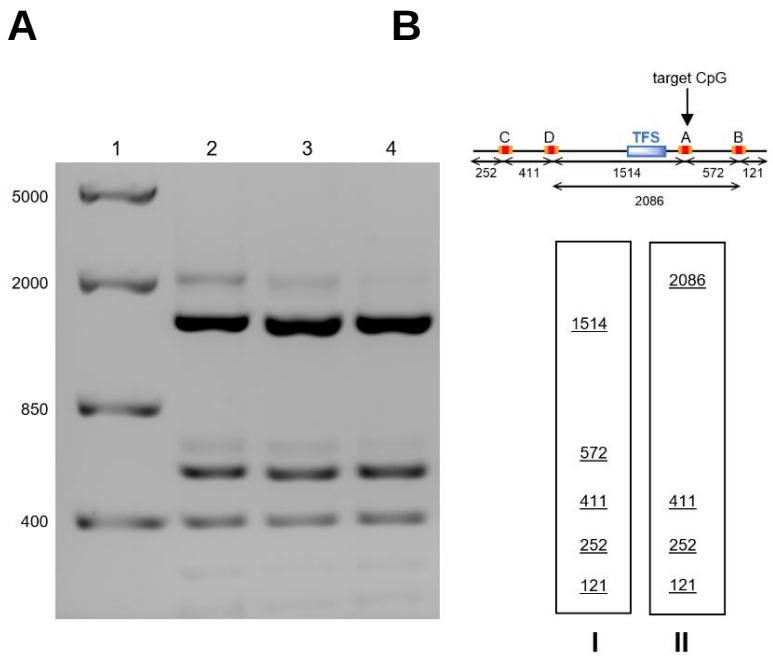


Fig. 3.25: DNA methylation specificity assay of M.MpeI-TFO with linear Litcon54 under various NaCl concentrations. **A:** Agarose gel analysis. **Lane 1:** Marker; **lane 2:** 50 mM NaCl; **lane 3:** 100 mM NaCl; **lane 4:** 200 mM NaCl. **B:** Expected fragmentation pattern of unprotected (I) and site A-specifically protected (II) linear Litcon54 after restriction with R.Psp1406I.

The overall methylation rate is even lower than observed before (Chapter 3.6.1) when increasing the NaCl concentration in the buffer. The DNA methylation does not become more specific with an increasing NaCl concentration and bands that result from off-target methylation are still faintly visible. Although a hundred-fold concentration of sodium ions is needed to accomplish the same triple helix stability as with magnesium ions, sodium ions on themselves can principally stabilize triple helices. If sodium ions are added to a solution that already contains magnesium ions, they gradually displace the magnesium ions and thereby decrease the stability of the triple helix[166, 167]. This may explain why DNA methylation specificity towards site A is not increased.

3.6.3 *In vitro* methylation of linear Litcon54 by M.MpeI(Q142L)-TFO

To test the *in vitro* methylation specificity of M.MpeI(Q142L)-TFO, an excess of the conjugate was incubated with linear Litcon54 at 30°C. The methylation reaction was stopped by heat inactivation after 20 h and restriction with R.Psp1406I led to the fragmentation patterns shown in Fig. 3.26 A.

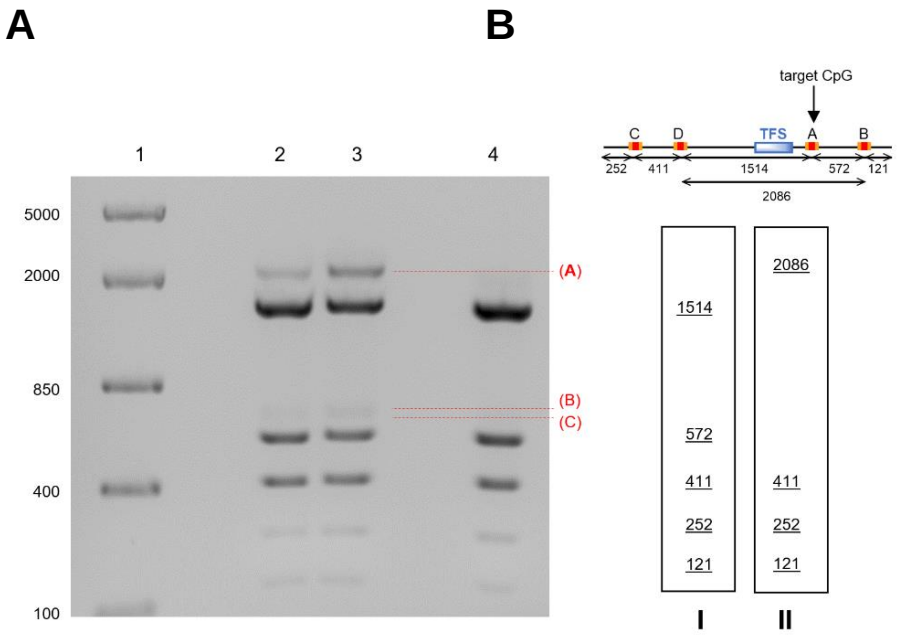


Fig. 3.26: DNA methylation specificity assay of M.MpeI(Q142L)-TFO with linear Litcon54. **A:** Agarose gel analysis. **Lane 1:** Marker; **lane 2:** 10 eq.; **lane 3:** 20 eq.; **lane 4:** Restriction control. **B:** Expected fragmentation pattern of unprotected (I) and site A-specifically protected (II) linear Litcon54 after restriction with R.Psp1406I.

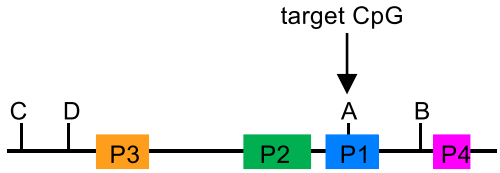
The amount of methylation is still very low after 20 h of incubation. The higher intensity of the specific 2086 bp fragment compared to the fragments that result from other single protected sites, indicates that there is a preference for the targeted CpG site. Usage of 20 eq. instead of 10 eq. leads to a stronger band at 2086 bp. Most interestingly, non-specific bands arising from mono methylation of site B and C are hardly visible, indicating that specificity of the M.MpeI(Q142L)-TFO towards site A is higher than that of M.MpeI-TFO (Fig. 3.22, lane 4). The concept of increasing specificity by reducing the binding affinity of the DNA MTase has been used before[80] and was more recently verified for M.SssI(Q147L), that was used for targeted DNA methylation with a CRISPR/dCas9 system[103]. Since this mutation in M.SssI corresponds to M.MpeI(Q142L), the same positive effect could be expected here as well.

3.6.4 Pyrosequencing of *in vitro* methylated Litcon54

The modification-restriction analysis described above considers only four of a total of 177 CpG sites on Litcon54. To get a broader perspective of the methylation specificity, various regions of Litcon54 samples (linear and supercoiled), that were methylated with M.MpeI-TFO, were sequenced and analyzed to see if there is also CpG methylation in regions other than the target region. Pyrosequencing is a sequence-by-synthesis method based on the detection of light upon the release of pyrophosphate when the complementary nucleotide is being added by the DNA polymerase[168]. Treatment of the samples with bisulfite in an acidic environment prior to sequencing helps to discriminate between cytosine and 5-methylcytosine as cytosine will deaminate to uracil whereas 5-methylcytosine remains unchanged[169, 170]. Combination of bisulfite conversion with pyrosequencing offers a quantitative methylation analysis of individual cytosine residues.

Four regions P1 (A155-T258), P2 (A310-G504), P3 (C1556-T1600) and P4 (G2453-G2501) were sequenced (Fig. 3.27 A). Of the four CpG sites analyzed before, only the target CpG (A) lies within a sequenced region. Untreated linear Litcon54, meaning that it was not incubated with either conjugate or AdoMet (**1**), was used as a control. Unfortunately, sequencing of the region that contains the target CpG (P1) failed. Fig. 3.27 B shows the methylation percentages found for the analyzed CpGs in region P2, P3 and P4.

A



B

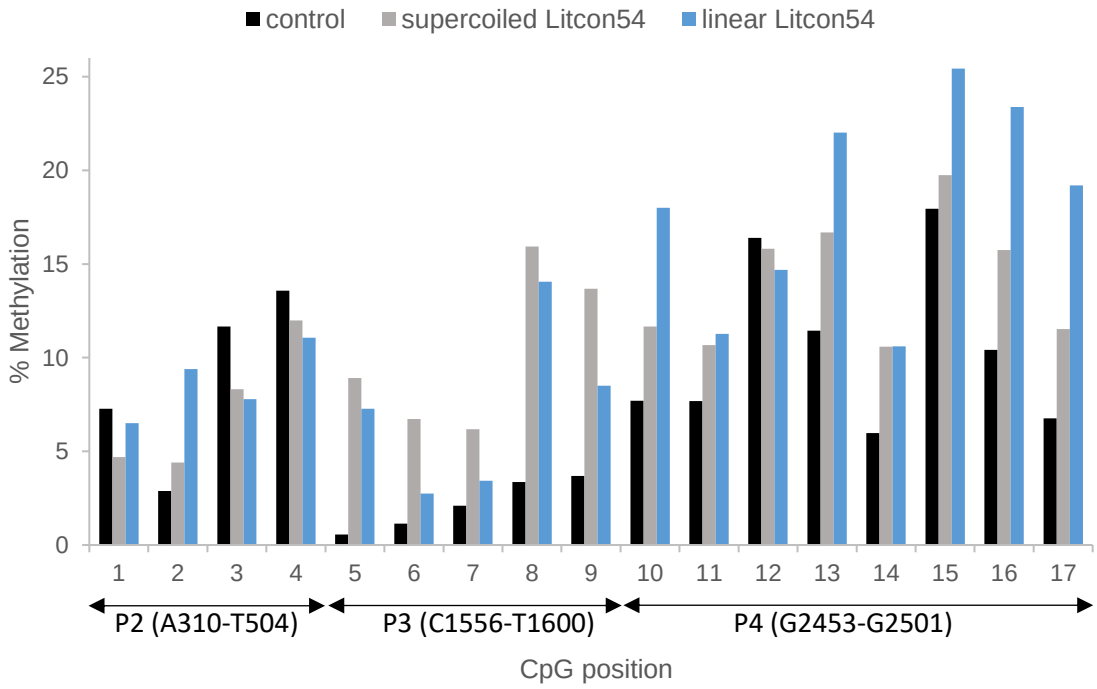


Fig. 3.27: Pyrosequencing of bisulfite-treated Litcon54. **A:** Location of regions P1 – P4 on Litcon54. **B:** methylation percentages found for supercoiled Litcon54 (gray) and linear Litcon54 (blue) and untreated linear Litcon54 (black). 17 CpGs were analyzed. Sequencing was planned and performed by A. Mposhi (Department of Hepatology and Gastroenterology/Department of Pathology and Medical Biology, University Medical Centre Groningen, Groningen, The Netherlands).

Since sequencing of the target region (P1) failed, no conclusions can be drawn with regards to the methylation levels of the non-specific regions compared to the target region. The significant methylation of the control sample within region P2 and P4, that could be caused by incomplete bisulfite conversion, makes the results of the linear and supercoiled samples within region P2 and P4 unreliable. When only considering region P3, the methylation percentage of linear Litcon54 is slightly lower compared to supercoiled Litcon54. This is in accordance with the observation made before, that a higher specificity towards the target site can be achieved with linear Litcon54.

3.6.5 Comparison to M.SssI(C141S)-TFO and other targeting systems

For a long time, M.SssI was the only known bacterial DNA MTase with a CpG recognition sequence. The recently discovered M.MpeI proved to be a good alternative in approaches to make tools for studying CpG methylation, as it also methylates within the CpG sequence[30]. In contrast to M.SssI, M.MpeI is much easier to handle and protein precipitation is not an issue.

The maleimide coupling chemistry that was previously used, requires a DNA MTase without the catalytic cysteine residue. Because of this, M.SssI(C141S)-TFO has only a few percent DNA MTase activity left. If this mutation is combined with a mutation to reduce DNA binding affinity, for example by replacing Gln147 with leucine[165], it leads to an inactive DNA MTase. On the contrary, SNAP-tag technology provides an improved coupling strategy that can be used to conjugate a targeting device to basically any desirable DNA MTase. This allows for the coupling with DNA MTases where the DNA MTase activity is not compromised by the removal of catalytic residues. With this coupling method it is possible to use the M.MpeI mutant corresponding to M.SssI(Q147L), M.MpeI(Q142L), to create a system with an active low DNA affinity MTase, as was demonstrated here.

SNAP-tag mediated coupling can also overcome the major disadvantage that chemical based targeting devices, like TFOs, have, because the coupling with the protein of interest, including the purification, used to be a chemical reaction. SNAP-tag mediated coupling can also be performed *in vivo* which in principle allows to produce the conjugates in the cell. This could make TFO targeting devices as interesting as protein targeting devices, like zinc fingers or TALEs. Moreover, designing TFOs takes much less effort than for example engineering zinc fingers to target a unique sequence in the DNA. Another interesting targeting device is the recently discovered CRISPR/dCas9 system that has some advantages over TFOs. Besides the simplicity with which the specificity can be designed, the CRISPR/dCas9 also binds very tightly to the DNA, leading to more efficient methylation of the target site (Sabrina Otto, unpublished). In this approach, the M.MpeI-TFO is used as an M.MpeI-ODN to hybridize with an extended 3` end of the sgRNA and hereby guided to the targeted site.

4 Summary and outlook

4.1 Summary

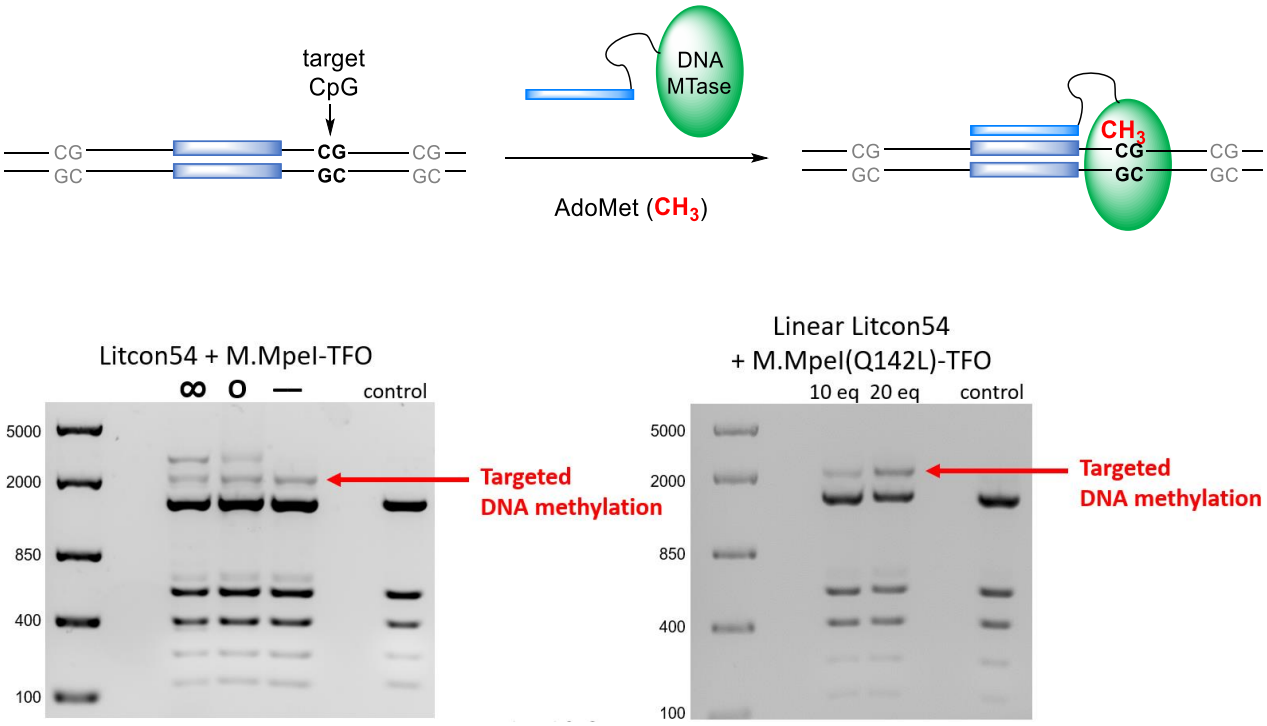
In this work, three bacterial DNA methyltransferases (MTases), that methylate cytosine within the CpG recognition sequence, were coupled to a triple helix-forming oligodeoxynucleotide (TFO) to create constructs for targeted DNA methylation. The constructs were designed to target and methylate a single CpG site located in the promoter region of the epithelial cell adhesion molecule (EpCAM), of which the overexpression has been found in many carcinomas and is correlated to the methylation status within the promoter region[58]. The coupling strategy was based on the SNAP-tag technology, with which the benzylic carbon of a para-substituted benzylguanine (BG) is connected to a variant of *O*6-alkylguanine DNA alkyltransferase (hAGT) that is fused with a DNA MTase. The BG modification of the TFO occurred in two steps and BG-PEG_{35}-TFO (**16**) was obtained with an overall yield of 65%. SNAP-tag fusion proteins of M.SssI, M.MpeI and the low DNA affinity mutant M.MpeI(Q142L) were then coupled to the BG-modified TFO **16**.

The purification of M.SssI-TFO by anion exchange chromatography proved to be challenging and the product that was obtained with a yield of less than 5% was contaminated with BG-PEG_{35}-TFO (**16**), which could not be removed from the product. M.SssI-TFO was therefore not further investigated. The conjugates M.MpeI-TFO and M.MpeI(Q142L)-TFO could be purified by anion exchange chromatography with a yield of 55% and 16%, respectively.

The DNA methylation specificity of M.MpeI-TFO was tested *in vitro* with a modification-restriction assay using Litcon54 DNA, a plasmid that was designed for the TFO to bind near a CpG site, and Litcon47, a similar plasmid lacking the triple helix-forming site (TFS). The methylation status of the challenged CpG sites was verified by incubation with the CpG methylation-sensitive restriction endonuclease R.Psp1406I. The experiment was carried out on supercoiled (∞), open-circular (O), and linear (–) plasmid. Specificity towards site A was observed with all three forms of Litcon54. Non-specific methylation was also observed and derived from methylation by both unbound and bound M.MpeI-TFO. The latter could be suppressed in linear Litcon54, which led to a higher specificity towards the target site compared to the supercoiled and open-circular forms. Preincubation of M.MpeI-TFO with linear Litcon54 did not affect the specificity. Higher amounts of NaCl in the reaction buffer, that would decrease the affinity of M.MpeI-TFO with the

DNA, possibly also reduced the stability of the triple helix and therefore did not increase the specificity towards the targeted site.

The highest specificity towards the targeted site was achieved with 20 eq. of M.MpeI(Q142L)-TFO and linear Litcon54. Not only was non-specific methylation by bound conjugate suppressed due to the use of linear plasmid, but also the reduced DNA binding affinity of the DNA MTase led to minimal non-specific methylation by free M.MpeI(Q142L)-TFO.



4.2 Outlook

This work demonstrates the potential of the SNAP-tag method as a coupling strategy for protein-DNA conjugates designed for targeted DNA methylation, but the constructs created here with this method only showed a partial preference towards the target site. It is obviously possible to try to optimize the experimental conditions in order to increase specificity. With this it is always important to keep in mind that *in vitro* conditions are different from *in vivo* conditions. To move to a direction where the specificity is sufficient for therapeutic applications, it is essential to optimize the elements of the constructs. It was shown here that the DNA MTase with a reduced DNA binding affinity led to a higher specificity for the targeted site. This result suggests that choosing such a DNA MTase might be preferred. An alternative would be to choose a mammalian DNA MTase, as its intrinsic activity is usually lower, and it will have less immunity issues when delivered in the organism. Successful gene silencing *via* targeted DNA methylation with the mammalian DNA MTase DNMT3a fused to an engineered zinc finger protein has been demonstrated[78, 171]. Alongside optimizing the DNA MTase, it is a question whether to keep using a TFO as targeting device or implement for example the CRISPR/dCas9 system in a tool for targeted DNA methylation.

Targeted DNA methylation is only a part of the relatively young field of epigenetic editing, in which epigenetics and gene targeting are combined. Other epigenetic processes like histone modifications and DNA demethylation can be targeted to a specific gene as well, and the direct instruction of epigenetic marks for gene expression has already been shown in several studies[64]. Epigenetic editing can help elucidating DNA methylation and its relation to other epigenetic processes, as the whole epigenetic network is still far from completely understood. With sufficient specificity of the targeting devices, epigenetic editing could become a powerful tool of sustainably reprogramming any gene and curing gene-related diseases.

5 Experimental

5.1 Materials

5.1.1 Chemicals

2-Mercaptoethanol	Serva
2-(*N*-morpholino)ethanesulfonic acid (MES)	Gerbu
3-(*N*-morpholino)propanesulfonic acid (MOPS)	Gerbu
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Gerbu
5(6)-Carboxytetramethylrhodamine-*N*-hydroxysuccinimidyl ester (TAMRA-NHS)	Thermo Fisher Scientific
Acetic acid (96%)	Merck
Acetone	Sigma Aldrich
Acetonitrile, HPLC grade	Fisher Chemical
Acrylamid/bisacrylamid (37.5:1)	Serva
Agar	AppliChem
Agarose	Sigma Aldrich
Ammonium persulfate (APS)	Serva
Ampicillin, sodium salt	Gerbu
BG-NH_2	Dr. G. Hanz, RWTH Aachen University
BG-PEG_{35}-N_3	Dr. G. Hanz, RWTH Aachen University

Bis-TRIS	Gerbu
Boric acid	Gerbu
Bromophenol blue, sodium salt	Serva
Calcium chloride	Fluka
Coomassie Brilliant Blue R250	Serva
Coomassie Plus Protein Assay Reagent	Serva
Dibenzoazacyclooctyne-*N*-hydroxysuccinimidyl ester (DBCO-NHS)	Sigma-Aldrich
Dimethylsulfoxide (DMSO)	Acros Organics
Disodiumhydrogen phosphate	Gerbu
Dithiothreitol (DTT)	Gerbu
Ethanol	Merck
Ethylenediaminetetraacetic acid (EDTA)	Gerbu
Formic acid	J. T. Baker
GelRed (10000 x)	Biotium
Glycerol	Gerbu
Glycerol (99%)	Sigma
Guanidinium hydrochloride	Gerbu
His Select Nickel Affinity Gel	Sigma
Hydrochloric acid (37%)	Merck
Imidazole, hydrochloride	Fluka
Isopropanol	Fisher Chemical

Isopropyl-β-D-thiogalactopyranoside (IPTG)	Gerbu
Magnesium acetate	Fluka
Magnesium chloride, pentahydrate	J. T. Baker
Methanol, HPLC grade	Fisher Scientific
Methanol, LC/MS grade	Fisher Chemical
N,*N*-Dimethylformamide (DMF)	Sigma Aldrich
Potassium acetate	Fluka
Potassium chloride	Merck
Potassiumdihydrogen phosphate	Merck
Proteosilver Silver stain kit	Sigma Aldrich
S-Adenosyl-L-methionine, butanedisulfonate (AdoMet)	Knoll Bioresearch SA
Sodium acetate	J. T. Baker
Sodium ascorbate	Fluka
Sodium chloride	Grüssing
Sodium hydroxide	Sigma Aldrich
Sodiumdodecylsulfat (SDS)	Gerbu
Tetramethylethylenediamine (TEMED)	Acros Organics
Triethylamine (TEA)	Fisher Scientific
Triethylammonium acetate (TEAAc)	Sigma Aldrich
Trifluoroacetic acid (TFA)	Acros Organics
TRIS-acetate	Fluka

TRIS-HCl	AppliChem
TRIS-X	Gerbu
Triton X-100	Fluka
Tryptone	BD Biosciences
Yeast extract	BD Biosciences
Zinc chloride	Grüssing

5.1.2 DNA and plasmids

The plasmids pSNAP(T7)-M.MpeI, pSNAP(T7)-M.MpeI(Q142L) and pSNAP(T7)-M.SssI encoding the respective SNAP-MTase fusion proteins were constructed by Dr. S. Szentes (Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Science, Szeged, Hungary).

The plasmids Litcon54 and Litcon47 were constructed within our lab by Dr. M. Maluszynska-Hoffman[162] and S. Otto, M.Sc. respectively. To this purpose, the commercially available plasmid Litmus28i (New England Biolabs) was digested with R.SacI and R.KpnI and the obtained vector fragment was incubated with the desired insert, that was produced by hybridization of the 54+/- and the 47+/- deoxynucleotides, respectively, and will have SacI- and KpnI-compatible single-stranded ends.

All oligodeoxynucleotides (ODN) were purchased from Integrated DNA Technologies. Their abbreviations and sequences are listed below.

Abbreviation	Sequence
Amino-TFO (**17**)	5'-H_2N-C_6-TTTTTTTTTTTTTTTCTCTCTTTT-3'
54_TFS+	5'-CCTTTTCTTTTCTCTCTTTTTTTTTTTTTTTGAGACGGAGTC TAACGTTGGTAC-3'
54_TFS-	5'-CAACGTTAGACTCCGTCTCAAAAAAAAAAAAAAAGAGAGA AAAGAAAAGGAGCT-3'
47_TFS+	5'-AACGTTTGCCTAATCATCTTTTGTTCATGTCCTACTGTTCAAG CCTCCAAGCTGTAC-'3
47_TFS-	5'-AGCTTGGAGGCTTGAACAGTAGGACATGAACAAAAGATGATT AGGCAAACGTTAGCT-'3

λ DNA was purchased from Thermo Scientific.

5.1.3 Markers

O'Range ruler	Thermo Scientific
FastRuler MR	Thermo Scientific
PAGE ruler prestained	Thermo Scientific
PAGE ruler unstained	Thermo Scientific

5.1.4 Enzymes and proteins

BSA	Thermo Fisher Scientific
Nb.BsmI	New England Biolabs
Proteinase K	Qiagen
R.BstUI	New England Biolabs
R.Psp1406I	Thermo Fisher Scientific
R.ScaI	Thermo Fisher Scientific

5.1.5 *Escherichia coli* strains

DH5α	provided internally
ScarabXpress T7lac (SG1709)	Scarab Genomics

5.1.6 Growth media and antibiotics

LB Miller	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl
LB Lennox	10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl
TB	12 g/Ltryptone, 24 g/L yeast extract, 4% glycerol, 13.3 mM KH_2PO_4, 9.2 mM K_2HPO_4
Medium for plates	10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 15 g/L agar
Ampicillin (Amp)	100 μg/mL

5.1.7 Consumables

Centrifugal units	Amicon ultra-4 10K (Millipore)
	JumboSep 10K (Pall)
Cryogenic tube (2 mL)	Nalgene
Dialysis tubing	Spectra/Por 4 (Spectrum Laboratories)
Disinfectant	Bacillol AF (Bode)
Electroporation cuvette	Bio-Rad
Filter paper MN 827 B	Macherey-Nagel
Microreactors	Eppendorf
PCR purification kit	Qiagen
PCR tubes	Thermofisher

Petri dishes	Greiner Bio-One
Pipetting tips	Gilson
Plastic cuvettes (2 mL)	Plastibrand
Polyethersulfone filters for organic solvents (0.2 μm pores)	Sartorius Stedim Biotech
Precision wipes	Kimberly Clark
Qiaprep mini spin kit	Qiagen
Cellulose acetate filters for aqueous solvents (0.2 μm pores)	Sartorius Stedim Biotech
Syringe filters (0.2 μm)	Whatman

5.1.8 Instruments

Agarose gel electrophoresis system	MiniSub Cell GT (Bio-Rad)
	connected to a PowerPac 300 (Bio-Rad)
Block thermostat	HBT 130 (HLC)
Centrifugal evaporator	Savant SPD111V SpeedVac Concentrator (Thermo Scientific)
	with refrigerated vapor trap RVT 100 (Savant)
Centrifuges	Microcentrifuge (V < 2 mL), model 5410 (Eppendorf)
	Avanti J-20 High-Performance Centrifuge (V > 2mL)
	with rotors JS-4.3 and JA-17 (Beckman Coulter)
Electroporator	MicroPulser electroporator (Bio-Rad)
	BioLogic DuoFlow FPLC (Bio-Rad),

FPLC system	with BioFrac fraction collector (Bio-Rad), peristaltic pump Abimed Minipuls 3 (Gilson) BioLogic DuoFlow software version 5.10 (Bio-Rad)
Gel dryer	Model 583 gel dryer (Bio-Rad)
Gel imager	Gel iX (Intas)
High vacuum pump	RZ 6 (Vacuubrand)
HPLC system	Breeze binary pump system 1525 and 2487 Dual λ absorbance detector (Waters) controlled with Empower 2 software (Waters)
Incubators for bacterial growth	Innova 4330 shaker (New Brunswick Scientific) GFL-3031 shaker (GFL) Heraeus B6 function line (Kendro Laboratory Products)
pH-meter	766 Calimatic (Knick)
Pipettes	Pipetman (Gilson)
SDS-PAGE system	Mini PROTEAN 3 Cell (Bio-Rad) connected to a PowerPac 300 (Bio-Rad)
Shaker	KS 250 basic (IKS Labortechnik)
Sonicator	Sonifier II W-250 (Branson)
Sterilizer	Varioklav steam sterilizer (H+P Labortechnik)
Thermocycler	Mastercycler personal (Eppendorf)
UV/Vis spectrometer	Cary 3E (Varian) with quartz cuvettes (Hellma)
Vortexer	Vortex-Genie 2 (Scientific Industries Inc.)

Water purification system	Milli-Q Academic with Elix 3 water system (Millipore)

5.2 Methods

5.2.1 Instrumental methods

5.2.1.1 DNA sequencing

DNA sequencing was performed at the Fraunhofer Institute for Molecular Biology and Applied Ecology (IME, Aachen). Plasmid DNA (100 ng) was mixed with sequencing primer (2 μL, 10 μM) and brought to a total volume of 30 μL with 10 mM TRIS-HCl, pH 8.5.

5.2.1.2 Bisulfite sequencing

Bisulfite sequencing of modified Litcon DNA was done by A. Mposhi (Department of Hepatology and Gastroenterology/Department of Pathology and Medical Biology, University Medical Centre Groningen, Groningen, The Netherlands). Samples were methylated as described in chapter 5.6.3. Non-linear Litcon DNA was linearized afterwards with R.ScaI. All samples were purified using the PCR purification Kit (Qiagen) according to the instruction of the manufacturer. The DNA was eluted from the column with H_2O (50 μL).

5.2.1.3 High performance liquid chromatography (HPLC)

Analysis of modified TFOs and other chemical substances was carried out with reversed-phase HPLC (rp HPLC) using an AquaPerfect C-18 column (250 x 4.6 mm, 5 μm, 120 Å) with an AquaPerfect C-18 precolumn (20 x 4.6 mm, 5 μm, 120 Å) from MZ Analysentechnik at a flow rate of 1 mL/min. Purification was done either with the same column or with a semi preparative Prontosil C-18 column (250 x 8.0 mm, 5 μm, 120 Å) with a C-18 precolumn (33 x 8.0 mm, 5 μm, 120 Å) from Bischoff at a flow rate of 3 mL/min. Analysis and isolation of the bioconjugation products was carried out by anion exchange

chromatography using a Poros HQ 10 anion exchange column (100 x 4.6 mm, 10 µm) from Applied Biosystems at a flow rate of 2 mL/min. Elution buffers were filtered before use. Before each run, the HPLC system was equilibrated with starting buffer for at least 10 min. Samples were diluted with buffers corresponding to starting conditions and centrifuged before injection. The following methods were used:

Method 1:

Buffer A: 10 mM TEAAc, pH 7.0

Buffer B: 10 mM TEAAc, 70% CH_3CN, pH 7.0

t / min	*B* / %
0	15
20	50
32	100
42	100
44	15

Detection at 260 nm and 300 nm

Method 2:

Buffer A: 10 mM TEAAc, pH 7.0

Buffer B: 10 mM TEAAc, 70% CH_3CN, pH 7.0

t / min	*B* / %
0	15
20	50
21	100
31	100
32	15

Detection at 260 nm and 300 nm

Method 3:

Buffer A: 10 mM TEAAc, pH 7.0

Buffer B: 10 mM TEAAc, 70% CH_3CN, pH 7.0

t / min	*B* / %
0	20
30	50
35	100
40	100
41	20

Detection at 260 nm and 280 nm

Method 4:

Buffer A: 0.01% TFA, pH 2.75

Buffer B: 0.01% TFA, 70% CH_3CN, pH 2.75

t / min	*B* / %
0	20
20	50
30	80
40	100
46	100
47	20

Detection at 284 nm and 555 nm

Method 5:

Buffer A: 0.01% TFA, pH 2.75

Buffer B: 0.01% TFA, 70% CH_3CN, pH 2.75

t / min	*B* / %
0	20
35	60
37	100
40	100
41	20

Detection at 284 nm and 555 nm

Method 6:

Buffer A: 20 mM TRIS-HCl, 10% glycerol, pH 8.5

Buffer B: 20 mM TRIS-HCl, 1 M NaCl, 10% glycerol, pH 8.5

t / min	*B* / %
0	15
3	15
20	100
24	100
25	15

Detection at 260 nm and 280 nm

5.2.1.4 Fast protein liquid chromatography (FPLC)

For purification of SNAP-M.SssI, FPLC was used with a Poros HS 50 cation exchange column (20 mL) from Applied Biosystems. Buffers were filtered, and the column was equilibrated with 5 cv of starting buffer before use. Samples were loaded onto the column with a peristaltic pump. A salt gradient (200-1000 mM NaCl) was used to elute SNAP-M.SssI from the column. UV detection was at 280 nm and purification was performed at 4°C.

5.2.1.5 UV/Vis spectroscopy

UV/Vis spectroscopy was used to determine concentrations of chemicals, DNA and proteins. Measurements were carried out in quartz cuvettes (Hellma) with a 1 cm light path. Concentrations were determined according to the Beer-Lambert law using known extinction coefficients. For dsDNA, the absorption at 260 nm was measured and the concentration was determined with the empirical formula 1 OD_{260} = 50 µg/mL. Protein concentrations were determined according to the Bradford method[172]. In a plastic cuvette, 100 µL sample and 900 µL Coomassie Plus Protein Assay Reagent were mixed together and incubated for 10 min at rt. The absorption at 595 nm was measured and the concentration was determined using a BSA calibration curve.

5.2.1.6 Mass spectrometry

Mass spectrometry was carried out with an LTQ Orbitrap XL electrospray ionization (ESI) mass spectrometer from Thermo Scientific. Samples were dissolved in methanol acidified with formic acid and measured in positive ion mode.

5.2.2 Molecular biological methods

5.2.2.1 Preparation of chemically competent *E.coli* cells

LB Lennox medium (200 mL) was inoculated with an overnight culture (2 mL) of SG1709 *E.coli* cells and incubated at 37°C and 130 rpm. When an optical density OD_{600} of 0.7–1.0 was reached, the cell cultures were poured into prechilled centrifuge beakers and cooled on ice for 10 min. The cultures were centrifuged (2790 x g, 4°C, 20 min) and the supernatant was decanted. The pellets were resuspended in $CaCl_2$ (40 mL, 100 mM) and centrifuged (2790 x g, 4°C, 10 min). The supernatant was decanted and the pellets were resuspended in $CaCl_2$ (40 mL, 100 mM) again. After cooling the suspensions on ice for 30 min, the cells were centrifuged (2790 x g, 4°C, 10 min) and the supernatant was decanted. The pellets were resuspended in $CaCl_2$ (4 mL, 100 mM) with glycerol (15%), aliquots (100 µL) were frozen in liquid nitrogen and finally stored at -80°C.

5.2.2.2 Transformation of chemically competent *E.coli* cells

Frozen suspensions of chemically competent *E.coli* cells (100 µL) were thawed on ice, mixed with DNA (50 ng) and incubated on ice for 30 min. The cell suspension was then incubated at 42°C for 90 s and after cooling on ice for 60 s, LB Miller medium (900 µL) was added and incubation continued at 37°C for 45 min. Finally, cells (100 µL) were plated on LB/Amp/0.2% glucose plates and incubated at 37°C overnight.

5.2.2.3 Plasmid DNA isolation

LB Miller medium with Amp (100 µg/mL) was inoculated with a colony of cells holding the desired plasmid and incubated overnight at 37°C and 225 rpm. The next day, the overnight culture was centrifuged (13000 x g, 4°C, 10 min) and the DNA was purified using the QIAprep Spin Miniprep Kit (Qiagen) according to the manufacturer's instruction. DNA was eluted from the column with H_2O (50 µL).

5.2.2.4 Preparation of linear and open-circular Litcon DNA

For the preparation of linear Litcon DNA, supercoiled Litcon DNA (10 µg) was incubated with R.ScaI (10 OVD) in 75 µL of 10 mM Bis-TRIS propane, 10 mM $MgCl_2$, 100 mM KCl, 100 µg/mL BSA, pH 6.5) at 37°C for 1 h and then at 65°C for 20 min.
To obtain open-circular Litcon DNA, supercoiled Litcon DNA (10 µg) was incubated with Nb.BsmI (10 OVD) in 75 µL of 50 mM TRIS-HCl, 100 mM NaCl, 10 mM $MgCl_2$, 100 µg/mL BSA, pH 7.9) at 65°C for 1 h and then at 80°C for 20 min.
Afterwards, both DNA forms were purified using the PCR purification kit (Qiagen) according to the instruction of the manufacturer. The DNA was eluted from the column with H_2O (50 µL).

5.2.2.5 Agarose gel electrophoresis

Agarose (210 mg or 300 mg) was dissolved in 0.5 x TBE buffer (30 mL; 44.5 mM TRIS-X, 44.5 mM boric acid, 1 mM EDTA, pH 8.0) by heating in a microwave oven. Subsequently, GelRed (3 µL) was added and the solution was poured into the agarose gel electrophoresis chamber. A plastic comb was placed into the chamber to create wells. After the gel was solidified, the comb was removed, and the gel was covered with 0.5 x TBE buffer. Samples were supplemented with Loading Dye and loaded into the wells. After electrophoresis, DNA bands were visualized with a UV transilluminator (312 nm).

5.2.2.6 Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE)

Analysis of proteins was done by SDS-PAGE on NuPAGE gels. For the running gel, (12%) Bis-TRIS (357 mM, pH 6.5–6.8), acrylamide/bisacrylamide (37.5:1, 12%), APS (0.1%, w/v) and TEMED (0.04%, v/v) were mixed together and poured between two glass plates, that were fixed in a casting stand, and covered with isopropanol. After polymerization of the running gel, the isopropanol was removed and for the stacking gel (5%), Bis-TRIS (357 mM, pH 6.5–6.8), acrylamide/bisacrylamide (37.5:1, 5%), APS (0.1%, w/v) and TEMED (0.04%, v/v) were mixed together. This mixture was immediately applied on top of the running gel. A comb was placed in the still liquid phase to generate wells. After polymerization, the gels were either stored in wet paper towels at 4°C or directly used.
Protein samples were mixed with SDS buffer (final concentration 50 mM TRIS-HCl, 2% SDS, 10% glycerol, 320 mM β-mercaptoethanol, 0.05% Coomassie Brilliant Blue R250, pH 6.8), incubated at 95°C for 10 min and shortly centrifuged. The comb was removed, the gel was placed in the electrophoresis apparatus and both top and bottom chambers were filled with NuPAGE running buffer (50 mM MOPS, 50 mM TRIS-X, 5 mM EDTA, 0.1% SDS (w/v)). The samples were pipetted into the wells and electrophoresis was carried out at 150 V for 40 min to 1 h. Afterwards, the gel was either incubated in Coomassie staining solution (10% acetic acid, 40% ethanol, 0.05% Coomassie Brilliant Blue R250) overnight and destained (10% acetic acid, 40% ethanol) the next day, or incubated in fixation solution (10% acetic acid, 30% ethanol) overnight and treated with silver staining (ProteoSilver silver stain kit, Sigma Aldrich) according to the manual of the manufacturer.

5.2.2.7 Polyacrylamide gel electrophoresis (PAGE)

PAGE was used to analyze triple helix formation of amino-TFO (**17**) with 54+/54- duplex DNA. Gels (12%), consisting of TRIS-HOAc (10 mM), $Mg(OAc)_2$ (10 mM), pH 7.4, acrylamide/bisacrylamide (37.5:1, 12%), APS (0.1%, w/v) and TEMED (0.04%, v/v) were made as described in chapter 5.2.2.6. Electrophoresis was carried out in an SDS-free buffer (10 mM TRIS-HOAc, 10 mM $Mg(OAc)_2$, pH 7.4) at 4°C and 90 V for 2.5 to 3 h. Afterwards, the gels were stained with GelRed for 30 min and analyzed under a UV transilluminator (312 nm).

5.3 Protein expression

5.3.1 Expression and purification of SNAP-M.SssI

LB Miller medium with Amp (100 µg/mL) was inoculated with a colony of *E.coli* harboring pSNAP(T7)-M.SssI and incubated overnight at 37°C and 225 rpm. The next day, TB/Amp medium (3 x 200 mL) with $ZnCl_2$ (100 µM) was inoculated with 2 mL of the overnight culture and incubated at 37°C and 200 rpm. When an optical density OD_{600} of 0.6–0.7 was reached, protein expression was induced with IPTG (0.4 mM). After further incubation at 23°C and 200 rpm for 20 h, the cells were harvested by centrifugation (2799 x g, 4°C, 30 min).

The cell pellets were resuspended in 5 mL buffer (50 mM TRIS-HCl, 200 mM NaCl, 5% glycerol, 1 mM DTT, pH 8.0) and the cells were disrupted by sonication (duty cycle: 30%, output control: 6-8, duration: 5 min, in 30 s steps and in-between a 1 min break). After centrifugation (39,800 x g, 4°C, 1 h), the supernatant was diluted twofold with equilibration buffer (50 mM TRIS-HCl, 500 mM NaCl, 1 mM imidazole, pH 7.5) and applied to a His-Select Nickel Affinity Column (2.5 mL), that was equilibrated with 4 cv of the same buffer. The column was then washed with 10 cv of wash buffer (50 mM TRIS-HCl, 500 mM NaCl, 10 mM imidazole, pH 7.5). The protein was eluted with 8 cv of elution buffer (50 mM TRIS-HCl, 500 mM NaCl, 200 mM imidazole, pH 7.5). The eluate was diluted with 30 mL of FPLC buffer (6.7 mM MES, 6.7 mM HEPES, 6.7 mM NaOAc, 1 mM EDTA, 10% glycerol, pH 8.5) and loaded on the FPLC for further purification by cation exchange chromatography (Poros HS 50, 20 mL). Proteins were eluted with an NaCl gradient from 200 mM to 1000 mM in FPLC buffer. Fractions containing SNAP-M.SssI were pooled and concentrated by ultrafiltration (JumboSep 10 K, Amicon ultra-4 10 K). The solution was mixed with glycerol to an end concentration of 50% (v/v) and stored at -20°C. The yield of SNAP-M.SssI was 1.89 mg.

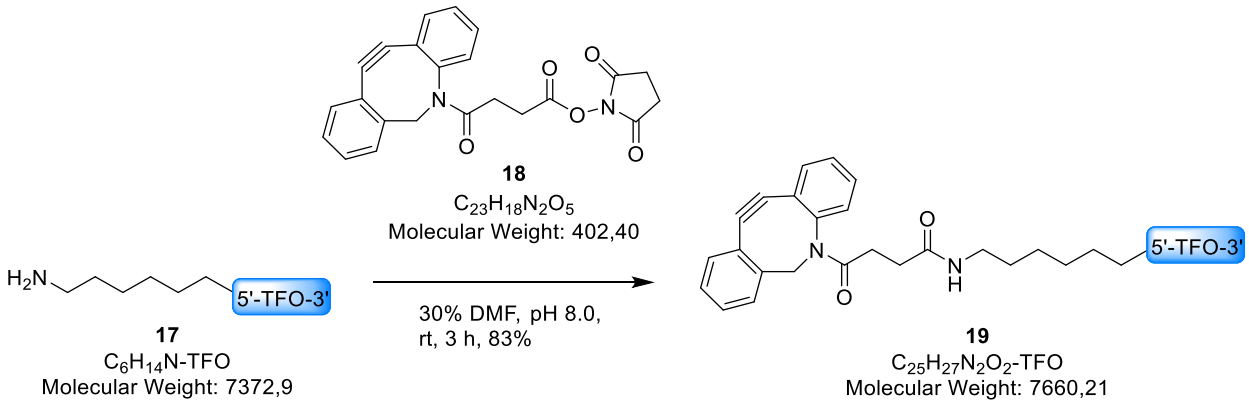
5.3.2 Expression and purification of SNAP-M.MpeI and SNAP-M.MpeI(Q142L)

LB Miller medium with Amp (100 µg/mL) was inoculated with a colony of *E.coli* harboring pSNAP(T7)-M.MpeI or pSNAP(T7)-M.MpeI(Q142L) and incubated overnight at 37°C and 225 rpm. The next day, TB/Amp medium (3 x 200 mL for pSNAP(T7)-M.MpeI, 2 x 200 mL for pSNAP(T7)-M.MpeI(Q142L)) was inoculated with 2 mL of the overnight culture and incubated at 37°C and 200 rpm. When an optical density OD_{600} of 0.6–0.7 was reached, protein expression was induced with IPTG (0.4 mM). After further incubation at 23°C and 200 rpm for 20 h, the cells were harvested by centrifugation (2799 x g, 4°C, 30 min).

The cell pellets were resuspended in 5 mL buffer (50 mM TRIS-HCl, 300 mM NaCl, 5% glycerol, 1 mM DTT, pH 8.0) and the cells were disrupted by sonication (duty cycle: 30%, output control: 6-8, duration: 5 min, in 30 s steps and in-between a 1 min break). Triton X-100 (final concentration 1%) was added and the disrupted cells were left on ice for 10 min. After centrifugation (39,800 x g, 4°C, 1 h), the supernatant was diluted twofold with equilibration buffer (50 mM TRIS-HCl, 500 mM NaCl, 1 mM imidazole, pH 7.5) and applied to a His-Select Nickel Affinity Column (2.5 mL), that was equilibrated with 4 cv of the same buffer. The column was then washed with 10 cv of wash buffer (50 mM TRIS-HCl, 500 mM NaCl, 10 mM imidazole, pH 7.5). Proteins were eluted using increasing imidazole concentrations in 2.5 mL elution buffer (50 mM TRIS-HCl, 500 mM NaCl, pH 7.5). The first fraction was obtained with 50 mM imidazole, the second fraction with 100 mM imidazole, the third fraction with 150 mM and the fourth and fifth fraction with 250 mM imidazole. Fractions containing the desired protein were pooled and concentrated by dialysis against buffer (50 mM TRIS-HCl, 100 mM NaCl, 10 mM β-mercaptoethanol, 50% glycerol, pH 7.5). The protein solution was stored at -20°C. 19.7 mg of SNAP-M.MpeI and 20.6 mg of SNAP-M.MpeI(Q142L) were obtained according to this protocol.

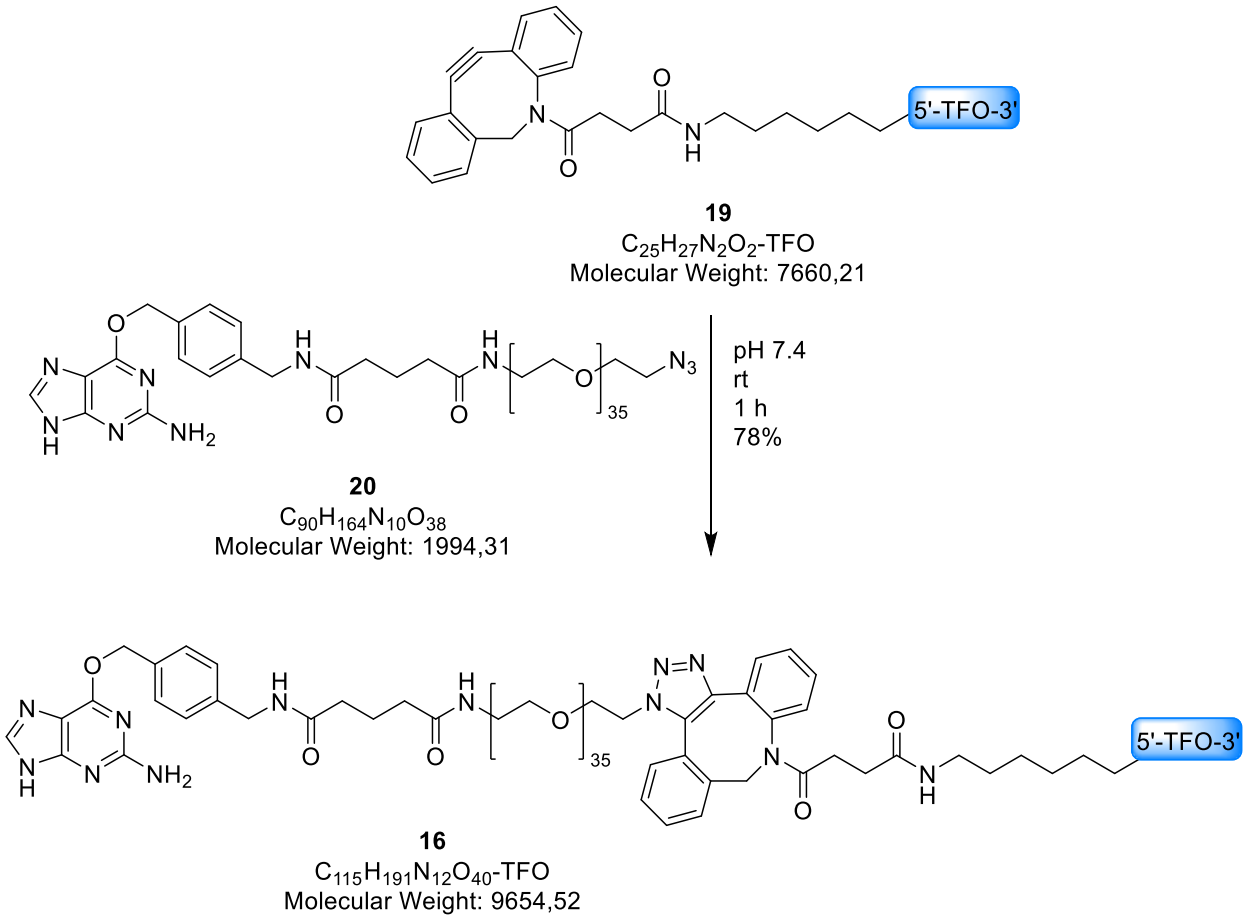
5.4 Syntheses

5.4.1 DBCO-TFO



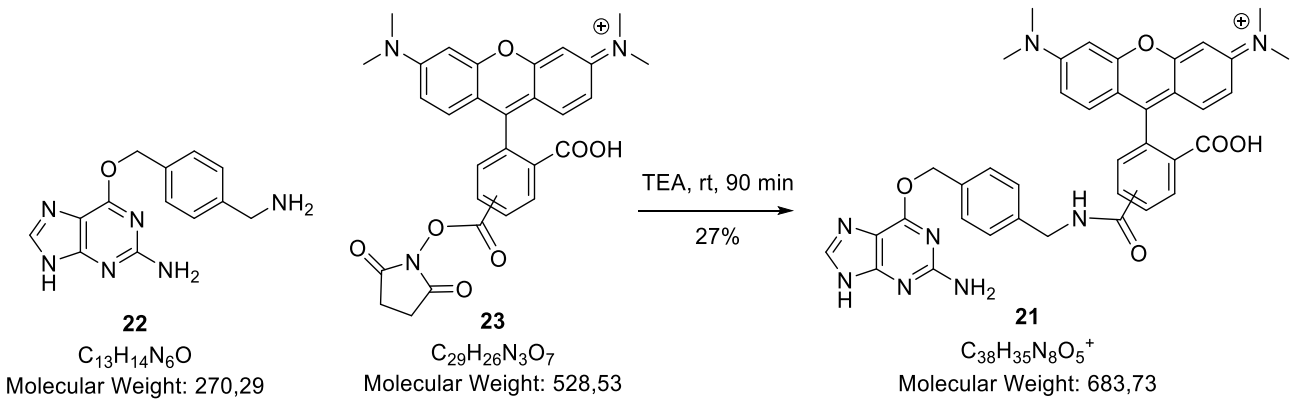
A solution (600 µL) of amino-TFO (**17**, 150 nmol, 250 µM) and DBCO-NHS (**18**, in DMF, 750 nmol, 1.25 mM) in reaction buffer (140 mM NaCl, 2.5 mM KCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, 30% DMF, pH 8.0) was incubated at rt for 3 h. The reaction was analyzed with rp HPLC (method 1). After complete conversion, the product **19** (t_R = 13.6 min) was isolated by rp HPLC (method 2) with the same analytical column. The solvent was removed by lyophilisation and the dry product **19** was dissolved in TE buffer pH 8.0. The yield of DBCO-TFO (**19**) was 125 nmol (83%), based on UV absorption (ε_{260} = 192300 $M^{-1}cm^{-1}$).

5.4.2 BG-PEG$_{35}$-TFO



A solution (400 µL) of DBCO-TFO (**19**, 100 nmol, 250 µM) and BG-PEG$_{35}$-N$_3$ (**20**, 396 nmol, 1 mM) in reaction buffer (140 mM NaCl, 2.5 mM KCl, 8 mM Na_2HPO_4, 1.5 mM KH_2PO_4, pH 7.4) was incubated at rt for 1 h. The product **16** was isolated by rp HPLC (method 3, t_R = 26.7 min) with the same analytical column. The solvent was removed by lyophilisation and the dry product **16** was dissolved in TE buffer pH 8.0. The yield of BG-PEG$_{35}$-TFO (**16**) was 78 nmol (78%), based on UV absorption (ε_{260} = 192300 $M^{-1}cm^{-1}$).

5.4.3 BG-TAMRA



22
$C_{13}H_{14}N_6O$
Molecular Weight: 270,29

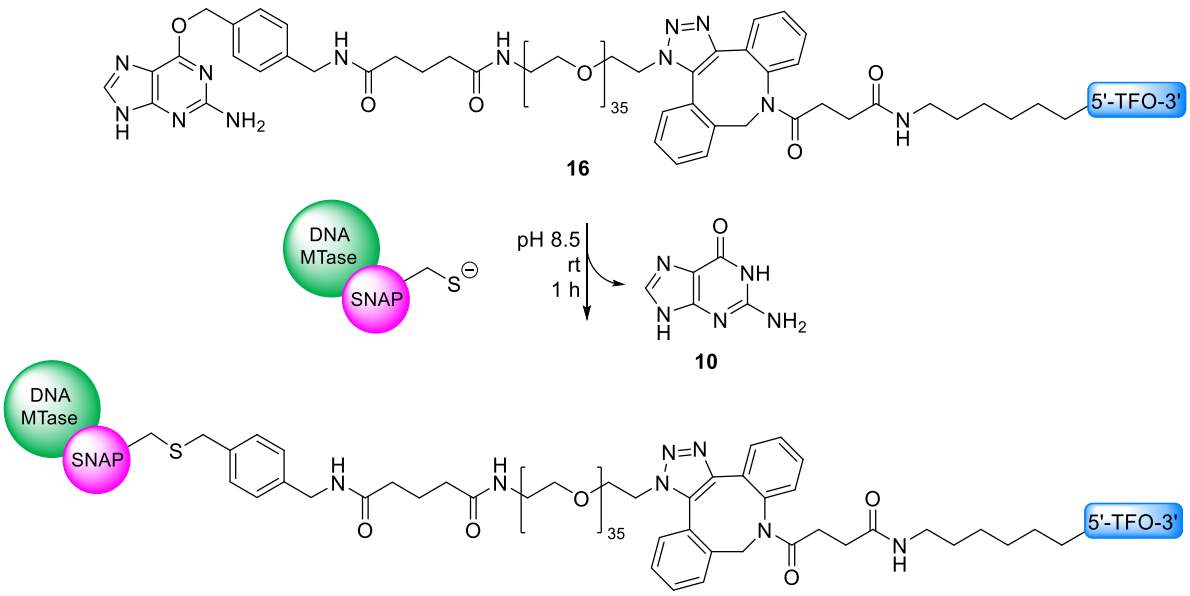
23
$C_{29}H_{26}N_3O_7$
Molecular Weight: 528,53

21
$C_{38}H_{35}N_8O_5^+$
Molecular Weight: 683,73

A solution (100 µL) of BG-NH_2 (**22**, 2 µmol, 20 mM), TAMRA-NHS (**23**, in DMSO, 2 µmol, 20 mM) and TEA (3.6 µmol, 36 mM) in DMF was incubated at rt for 90 min. The reaction was monitored by rp HPLC (method 4). Both isomers **21b** (t_R = 26 min) and **21a** (t_R = 27.5 min) were isolated by semi-preparative rp HPLC (method 5). After removal of the solvent by lyophilisation the dry products **21a** and **21b** were dissolved in DMSO. The total yield of BG-TAMRA (**21**) was 540 nmol (27%), based on UV absorption (ε_{553} = 80000 $M^{-1}\cdot cm^{-1}$).

MS (ESI, pos.) *m/z* (relative intensity): 683.27 (100) $[M]^+$, 532.22 (55) [M-guanine]$^+$.

5.5 Protein-TFO conjugate synthesis and purification



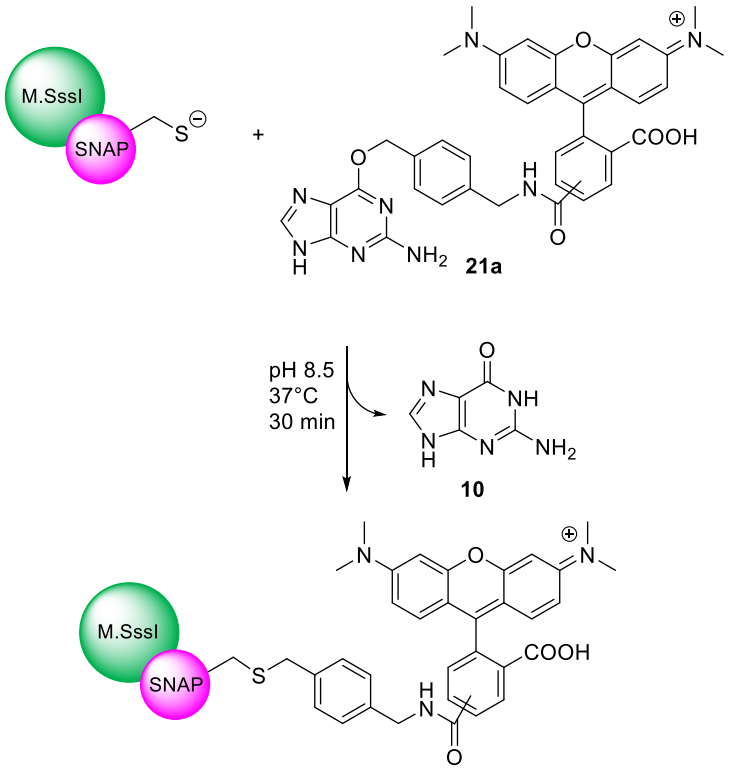
A solution of SNAP-MTase and BG-PEG$_{35}$-TFO (**16**) in buffer (50 mM TRIS-HCl, 100 mM NaCl, 1 mM DTT, pH 8.5) was incubated at rt for 1 h. The solution was centrifuged (16,000 x g, 4°C, 10 min) and the supernatant was diluted twofold with buffer (10 mM TRIS-HCl, 200 mM NaCl, 10% glycerol, pH 8.5) and centrifuged (16000 x g, 4°C, 10 min) again. The conjugate was purified by anion exchange chromatography using an HPLC system (method 6). Fractions containing the conjugate were concentrated by ultrafiltration (Amicon ultra-4 10 K). The concentration was determined with the Bradford assay.

This general protocol was used for the conjugation of SNAP-M.SssI, SNAP-M.MpeI and SNAP-M.MpeI(Q142L) with BG-PEG$_{35}$-TFO (**16**). The specific details for each conjugation and the respective yields are listed in the table below.

BG-PEG$_{35}$-TFO +	**SNAP-M.SssI**	**SNAP-M.MpeI**	**SNAP-M.MpeI(Q142L)**
SNAP-MTase	943 µg, 10 µM	4 mg, 20 µM	2.2 mg, 11 µM
BG-PEG$_{35}$-TFO (**16**)	15 µM, 1.5 eq.	12 µM, 0.6 eq.	6.5 µM, 0.6 eq.
V / µL	1411	3000	3000
t_R / min	10.6	11.8	10.7
Y / %	4.1	55	16

5.6 Activity assays

5.6.1 SNAP-tag activity test of SNAP-M.SssI with BG-TAMRA



A solution (50 µL) of SNAP-M.SssI (5 µM) and BG-TAMRA (**21a**)[1] (10 µM, 2 eq.) in buffer (50 mM TRIS-HCl, 100 mM NaCl, 1 mM DTT, pH 8.5) was incubated at 37°C for 30 min. To protect the fluorophore from light, the reaction tube was covered with aluminum foil. A sample (2 µg) was run on an SDS gel. The gel was first analyzed under a UV transilluminator (312 nm) before staining it with Coomassie.

[1] Either the 5- or the 6-isomer of BG-TAMRA (**21**) was used.

5.6.2 Modification-restriction assay with λ DNA

20 µL of a solution (40 µL) of λ DNA (50 µg/mL), AdoMet (**1**, 160 µM) and the DNA MTase of interest (i.e. SNAP-MTase fusion protein or conjugation product, 196 nmol, 4.9 µM) in reaction buffer (50 mM TRIS-HCl, 50 mM NaCl, 10 mM EDTA, 5 mM DTT, pH 8.5), were added to a solution (20 µL) of λ DNA (50 µg/mL) and AdoMet (**1**, 160 µM) in reaction buffer and mixed. 20 µL of the latter solution, were added to another 20 µL of λ DNA (50 µg/mL) and AdoMet (**1**, 160 µM) in reaction buffer and mixed again. The process was repeated to create a dilution series, where the DNA MTase concentration was reduced by 50% from each dilution to the next dilution while the other components were kept constant. The samples were incubated at 37°C for 1 h and then at 65°C for 20 min to terminate the methylation reaction. R.BstUI (10 U) in buffer (30 µL, 50 mM KOAc, 20 mM TRIS-HOAc, 10 mM $Mg(OAc)_2$, 100 µg/mL BSA, pH 7.9) was added to each sample and incubation was continued at 60°C for 1 h. Proteinase K (1 µL, 600 mAU/mL) was added and the samples were incubated at 53°C for 1 h. The samples were supplemented with loading buffer (10 µL, 6x) and aliquots (10 µL) were loaded on an agarose gel (1%). After electrophoresis at 90 V, the DNA bands were visualized with a UV transilluminator (312 nm).

To investigate the DNA MTase activity of M.MpeI-TFO with λ DNA under conditions used to test targeted DNA methylation specificity, different NaCl concentrations were used for the modification reactions with reaction buffers containing 50, 100, 200 or 500 mM NaCl, in addition to $MgCl_2$ (10 mM) and TRIS-HCl (10 mM), pH 7.5. The methylation reaction was carried out at 30°C for 1 h and then at 65°C for 20 min to stop the methylation.

5.6.3 Modification-restriction assay with Litcon54 DNA or Litcon47 DNA and M.MpeI-TFO

A solution (10 µL) of Litcon54 DNA or Litcon47 DNA (360 ng, 20 nM), M.MpeI-TFO (20 nM) and AdoMet (**1**, 80 µM) in buffer (10 mM TRIS-HCl, 50 mM NaCl, 10 mM $MgCl_2$, pH 7.5) was incubated at 30°C for 1 h and then at 65°C for 20 min to inactivate the DNA MTase. A solution (5 µL) of R.Psp1406I (10 U) in buffer (33 mM TRIS-HOAc, 20 mM $Mg(OAc)_2$, 66 mM KOAc, 100 µg/mL BSA, pH 7.9) was added and the mixture was incubated at 37°C for 1 h. Supercoiled and open-circular forms of Litcon-DNA were further fragmented by adding a solution (5 µL) of R.ScaI (10 U) in buffer (10 mM bis-TRIS propane, 10 mM $MgCl_2$, 100 mM KCl, 100 µg/mL BSA, pH 6.5) and incubation at 37°C for 1 h and then at 80°C for 20 min. The samples were supplemented with loading buffer and loaded on an agarose gel (0.7%). After electrophoresis at 80 V, the DNA bands were visualized with a UV transilluminator (312 nm).

5.6.4 Modification-restriction assay with linear Litcon54 DNA and M.MpeI(Q142L)-TFO

Linear Litcon54 DNA (360 ng, 20 nM) was incubated with M.MpeI(Q142L)-TFO (200 nM or 400 nM) and AdoMet (**1**, 80 µM) in buffer (10 µL, 10 mM TRIS-HCl, 50 mM NaCl, 10 mM $MgCl_2$, pH 7.5) and incubated at 30°C for 20 h and then at 65°C for 20 min to inactivate the DNA MTase. R.Psp1406I (10 U) in buffer (5 µL, 33 mM TRIS-HOAc, 20 mM $Mg(OAc)_2$, 66 mM KOAc, 100 µg/mL BSA, pH 7.9) was added and the mixture was incubated at 37°C for 1 h. The samples were supplemented with loading buffer, loaded on an agarose gel (0.7%) and after electrophoresis at 80 V, the DNA bands were visualized with a UV transilluminator (312 nm).

6 References

[1] J. D. Watson, F. H. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid, *Nature* **1953**, *171*, 737-738.

[2] J. M. Berg, J. L. Tymoczko, L. Stryer, *Biochemistry*, W.H. Freeman, Basingstoke, **2012**.

[3] C. Dupont, D. R. Armant, C. A. Brenner, Epigenetics: definition, mechanisms and clinical perspective, *Semin Reprod Med* **2009**, *27*, 351-357.

[4] T. B. Miranda, P. A. Jones, DNA methylation: the nuts and bolts of repression, *J Cell Physiol* **2007**, *213*, 384-390.

[5] D. B. Dunn, J. D. Smith, Occurrence of a new base in the deoxyribonucleic acid of a strain of Bacterium coli, *Nature* **1955**, *175*, 336-337.

[6] M. Ehrlich, M. A. Gama-Sosa, L. H. Carreira, L. G. Ljungdahl, K. C. Kuo, C. W. Gehrke, DNA methylation in thermophilic bacteria: N4-methylcytosine, 5-methylcytosine, and N6-methyladenine, *Nucleic Acids Research* **1985**, *13*, 1399-1412.

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

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

[9] M. Ehrlich, R. Y. Wang, 5-Methylcytosine in eukaryotic DNA, *Science* **1981**, *212*, 1350-1357.

[10] S. Klimasauskas, J. L. Nelson, R. J. Roberts, The sequence specificity domain of cytosine-C5 methylases, *Nucleic Acids Res* **1991**, *19*, 6183-6190.

[11] J. Posfai, A. S. Bhagwat, G. Posfai, R. J. Roberts, Predictive motifs derived from cytosine methyltransferases, *Nucleic Acids Res* **1989**, *17*, 2421-2435.

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

[13] D. Ingrosso, A. V. Fowler, J. Bleibaum, S. Clarke, Sequence of the D-aspartyl/L-isoaspartyl protein methyltransferase from human erythrocytes. Common sequence motifs for protein, DNA, RNA, and small molecule S-adenosylmethionine-dependent methyltransferases, *J Biol Chem* **1989**, *264*, 20131-20139.

[14] J. C. Wu, D. V. Santi, Kinetic and catalytic mechanism of HhaI methyltransferase, *J Biol Chem* **1987**, *262*, 4778-4786.

[15] L. Chen, A. M. MacMillan, W. Chang, K. Ezaz-Nikpay, W. S. Lane, G. L. Verdine, Direct identification of the active-site nucleophile in a DNA (cytosine-5)-methyltransferase, *Biochemistry* **1991**, *30*, 11018-11025.

[16] A. Jeltsch, Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases, *Chembiochem* **2002**, *3*, 274-293.

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

[18] D. G. Osterman, G. D. DePillis, J. C. Wu, A. Matsuda, D. V. Santi, 5-Fluorocytosine in DNA is a mechanism-based inhibitor of HhaI methylase, *Biochemistry* **1988**, *27*, 5204-5210.

[19] S. Gabbara, A. S. Bhagwat, The mechanism of inhibition of DNA (cytosine-5-)-methyltransferases by 5-azacytosine is likely to involve methyl transfer to the inhibitor, *Biochemical Journal* **1995**, *307*, 87-92.

[20] L. Zhou, X. Cheng, B. A. Connolly, M. J. Dickman, P. J. Hurd, D. P. Hornby, Zebularine: A Novel DNA Methylation Inhibitor that Forms a Covalent Complex with DNA Methyltransferases, *Journal of Molecular Biology* **2002**, *321*, 591-599.
[21] T. Malone, R. M. Blumenthal, X. Cheng, Structure-guided analysis reveals nine sequence motifs conserved among DNA amino-methyltransferases, and suggests a catalytic mechanism for these enzymes, *J Mol Biol* **1995**, *253*, 618-632.
[22] A. L. Pogolotti, Jr., A. Ono, R. Subramaniam, D. V. Santi, On the mechanism of DNA-adenine methylase, *J Biol Chem* **1988**, *263*, 7461-7464.
[23] K. Goedecke, M. Pignot, R. S. Goody, A. J. Scheidig, E. Weinhold, Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog, *Nat Struct Biol* **2001**, *8*, 121-125.
[24] S. Kumar, X. Cheng, S. Klimasauskas, S. Mi, J. Posfai, R. J. Roberts, G. G. Wilson, The DNA (cytosine-5) methyltransferases, *Nucleic Acids Res* **1994**, *22*, 1-10.
[25] S. Klimasauskas, T. Szyperski, S. Serva, K. Wuthrich, Dynamic modes of the flipped-out cytosine during HhaI methyltransferase-DNA interactions in solution, *EMBO J* **1998**, *17*, 317-324.
[26] R. J. Roberts, X. Cheng, Base flipping, *Annu Rev Biochem* **1998**, *67*, 181-198.
[27] M. O'Gara, J. R. Horton, R. J. Roberts, X. Cheng, Structures of HhaI methyltransferase complexed with substrates containing mismatches at the target base, *Nat Struct Biol* **1998**, *5*, 872-877.
[28] J. R. Horton, G. Ratner, N. K. Banavali, N. Huang, Y. Choi, M. A. Maier, V. E. Marquez, A. D. MacKerell, Jr., X. Cheng, Caught in the act: visualization of an intermediate in the DNA base-flipping pathway induced by HhaI methyltransferase, *Nucleic Acids Res* **2004**, *32*, 3877-3886.
[29] R. J. Roberts, On base flipping, *Cell* **1995**, *82*, 9-12.
[30] M. Wojciechowski, H. Czapinska, M. Bochtler, CpG underrepresentation and the bacterial CpG-specific DNA methyltransferase M.MpeI, *Proc Natl Acad Sci U S A* **2013**, *110*, 105-110.
[31] K. M. Reinisch, L. Chen, G. L. Verdine, W. N. Lipscomb, The crystal structure of HaeIII methyltransferase convalently complexed to DNA: an extrahelical cytosine and rearranged base pairing, *Cell* **1995**, *82*, 143-153.
[32] M. O'Gara, R. J. Roberts, X. Cheng, A structural basis for the preferential binding of hemimethylated DNA by HhaI DNA methyltransferase, *J Mol Biol* **1996**, *263*, 597-606.
[33] G. G. Wilson, N. E. Murray, Restriction and Modification Systems, *Annual Review of Genetics* **1991**, *25*, 585-627.
[34] W. Messer, M. Noyer-Weidner, Timing and targeting: the biological functions of Dam methylation in E. coli, *Cell* **1988**, *54*, 735-737.
[35] A. Hermann, H. Gowher, A. Jeltsch, Biochemistry and biology of mammalian DNA methyltransferases, *Cell Mol Life Sci* **2004**, *61*, 2571-2587.
[36] A. Bird, DNA methylation patterns and epigenetic memory, *Genes Dev* **2002**, *16*, 6-21.
[37] R. M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation, *Nature* **2013**, *502*, 472-479.
[38] A. Jeltsch, R. Z. Jurkowska, New concepts in DNA methylation, *Trends Biochem Sci* **2014**, *39*, 310-318.
[39] M. Okano, D. W. Bell, D. A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development, *Cell* **1999**, *99*, 247-257.
[40] M. Okano, S. Xie, E. Li, Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases, *Nat Genet* **1998**, *19*, 219-220.
[41] H. Gowher, A. Jeltsch, Molecular enzymology of the catalytic domains of the Dnmt3a and Dnmt3b DNA methyltransferases, *J Biol Chem* **2002**, *277*, 20409-20414.

[42] A. Aoki, I. Suetake, J. Miyagawa, T. Fujio, T. Chijiwa, H. Sasaki, S. Tajima, Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases, *Nucleic Acids Res* **2001**, *29*, 3506-3512.

[43] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, Human DNA methylomes at base resolution show widespread epigenomic differences, *Nature* **2009**, *462*, 315-322.

[44] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, Structure and function of mammalian DNA methyltransferases, *Chembiochem* **2011**, *12*, 206-222.

[45] F. Neri, A. Krepelova, D. Incarnato, M. Maldotti, C. Parlato, F. Galvagni, F. Matarese, H. G. Stunnenberg, S. Oliviero, Dnmt3L antagonizes DNA methylation at bivalent promoters and favors DNA methylation at gene bodies in ESCs, *Cell* **2013**, *155*, 121-134.

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

[47] R. Holliday, J. E. Pugh, DNA modification mechanisms and gene activity during development, *Science* **1975**, *187*, 226-232.

[48] M. Fatemi, A. Hermann, S. Pradhan, A. Jeltsch, The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA, *J Mol Biol* **2001**, *309*, 1189-1199.

[49] G. Liang, M. F. Chan, Y. Tomigahara, Y. C. Tsai, F. A. Gonzales, E. Li, P. W. Laird, P. A. Jones, Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements, *Mol Cell Biol* **2002**, *22*, 480-491.

[50] M. Fatemi, A. Hermann, H. Gowher, A. Jeltsch, Dnmt3a and Dnmt1 functionally cooperate during de novo methylation of DNA, *Eur J Biochem* **2002**, *269*, 4981-4984.

[51] K. D. Robertson, DNA methylation and human disease, *Nat Rev Genet* **2005**, *6*, 597-610.

[52] B. Richardson, DNA methylation and autoimmune disease, *Clin Immunol* **2003**, *109*, 72-79.

[53] B. Richardson, L. Scheinbart, J. Strahler, L. Gross, S. Hanash, M. Johnson, Evidence for impaired T cell DNA methylation in systemic lupus erythematosus and rheumatoid arthritis, *Arthritis Rheum* **1990**, *33*, 1665-1673.

[54] A. López Castel, J. D. Cleary, C. E. Pearson, Repeat instability as the basis for human diseases and as a potential target for therapy, *Nature Reviews Molecular Cell Biology* **2010**, *11*, 165.

[55] M. Esteller Epigenetics in Cancer, *New England Journal of Medicine* **2008**, *358*, 1148-1159.

[56] P. A. Jones, P. W. Laird, Cancer epigenetics comes of age, *Nat Genet* **1999**, *21*, 163-167.

[57] D. Maetzel, S. Denzel, B. Mack, M. Canis, P. Went, M. Benk, C. Kieu, P. Papior, P. A. Baeuerle, M. Munz, O. Gires, Nuclear signalling by tumour-associated antigen EpCAM, *Nat Cell Biol* **2009**, *11*, 162-171.

[58] B. T. F. van der Gun, R. Wasserkort, A. Monami, A. Jeltsch, T. Rasko, K. Slaska-Kiss, R. Cortese, M. G. Rots, L. de Leij, M. H. J. Ruiters, A. Kiss, E. Weinhold, P. M. J. McLaughlin, Persistent downregulation of the pancarcinoma-associated epithelial cell adhesion molecule via active intranuclear methylation, *Int J Cancer* **2008**, *123*, 484-489.

[59] P. Renbaum, D. Abrahamove, A. Fainsod, G. G. Wilson, S. Rottem, A. Razin, Cloning, characterization, and expression in Escherichia coli of the gene coding for the CpG DNA methylase from Spiroplasma sp. strain MQ1(M.SssI), *Nucleic Acids Res* **1990**, *18*, 1145-1152.

[60] P. Renbaum, A. Razin, Mode of action of the Spiroplasma CpG methylase M.SssI, *FEBS Lett* **1992**, *313*, 243-247.
[61] K. Matsuo, J. Silke, K. Gramatikoff, W. Schaffner, The CpG-specific methylase SssI has topoisomerase activity in the presence of Mg2+, *Nucleic Acids Res* **1994**, *22*, 5354-5359.
[62] E. V. Koudan, J. M. Bujnicki, E. S. Gromova, Homology modeling of the CG-specific DNA methyltransferase SssI and its complexes with DNA and AdoHcy, *J Biomol Struct Dyn* **2004**, *22*, 339-345.
[63] K. Slaska-Kiss, E. Timar, A. Kiss, Complementation between inactive fragments of SssI DNA methyltransferase, *BMC Mol Biol* **2012**, *13*, 17.
[64] M. L. de Groote, P. J. Verschure, M. G. Rots, Epigenetic Editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes, *Nucleic Acids Res* **2012**, *40*, 10596-10613.
[65] G. L. Xu, T. H. Bestor, Cytosine methylation targetted to pre-determined sequences, *Nat Genet* **1997**, *17*, 376-378.
[66] P. B. Dervan, Molecular recognition of DNA by small molecules, *Bioorg Med Chem* **2001**, *9*, 2215-2235.
[67] M. Mrksich, M. E. Parks, P. B. Dervan, Hairpin Peptide Motif. A New Class of Oligopeptides for Sequence-Specific Recognition in the Minor Groove of Double-Helical DNA, *Journal of the American Chemical Society* **1994**, *116*, 7983-7988.
[68] P. B. Dervan, R. M. Doss, M. A. Marques, Programmable DNA binding oligomers for control of transcription, *Curr Med Chem Anticancer Agents* **2005**, *5*, 373-387.
[69] D. S. Goodsell, M. L. Kopka, D. Cascio, R. E. Dickerson, Crystal structure of CATGGCCATG and its implications for A-tract bending models, *Proc Natl Acad Sci U S A* **1993**, *90*, 2930-2934.
[70] I. Kers, P. B. Dervan, Search for the optimal linker in tandem hairpin polyamides, *Bioorg Med Chem* **2002**, *10*, 3339-3349.
[71] Y. Kawamoto, A. Sasaki, K. Hashiya, S. Ide, T. Bando, K. Maeshima, H. Sugiyama, Tandem trimer pyrrole-imidazole polyamide probes targeting 18 base pairs in human telomere sequences, *Chem Sci* **2015**, *6*, 2307-2312.
[72] Q. Liu, D. J. Segal, J. B. Ghiara, C. F. Barbas, 3rd, Design of polydactyl zinc-finger proteins for unique addressing within complex genomes, *Proc Natl Acad Sci U S A* **1997**, *94*, 5525-5530.
[73] J. R. Desjarlais, J. M. Berg, Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins, *Proc Natl Acad Sci U S A* **1993**, *90*, 2256-2260.
[74] M. Elrod-Erickson, M. A. Rould, L. Nekludova, C. O. Pabo, Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions, *Structure* **1996**, *4*, 1171-1180.
[75] C. D. Carvin, R. D. Parr, M. P. Kladde, Site-selective in vivo targeting of cytosine-5 DNA methylation by zinc-finger proteins, *Nucleic Acids Res* **2003**, *31*, 6493-6501.
[76] A. R. McNamara, P. J. Hurd, A. E. Smith, K. G. Ford, Characterisation of site-biased DNA methyltransferases: specificity, affinity and subsite relationships, *Nucleic Acids Res* **2002**, *30*, 3818-3830.
[77] F. Li, M. Papworth, M. Minczuk, C. Rohde, Y. Zhang, S. Ragozin, A. Jeltsch, Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes, *Nucleic Acids Res* **2007**, *35*, 100-112.
[78] A. G. Rivenbark, S. Stolzenburg, A. S. Beltran, X. Yuan, M. G. Rots, B. D. Strahl, P. Blancafort, Epigenetic reprogramming of cancer cells via targeted DNA methylation, *Epigenetics* **2012**, *7*, 350-360.

[79] S. Durai, M. Mani, K. Kandavelou, J. Wu, M. H. Porteus, S. Chandrasegaran, Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells, *Nucleic Acids Res* **2005**, *33*, 5978-5990.

[80] A. E. Smith, K. G. Ford, Specific targeting of cytosine methylation to DNA sequences in vivo, *Nucleic Acids Res* **2007**, *35*, 740-754.

[81] W. Choe, S. Chandrasegaran, M. Ostermeier, Protein fragment complementation in M.HhaI DNA methyltransferase, *Biochem Biophys Res Commun* **2005**, *334*, 1233-1240.

[82] B. Chaikind, M. Ostermeier, Directed evolution of improved zinc finger methyltransferases, *PLoS One* **2014**, *9*, e96931.

[83] J. Boch, U. Bonas, Xanthomonas AvrBs3 family-type III effectors: discovery and function, *Annu Rev Phytopathol* **2010**, *48*, 419-436.

[84] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, U. Bonas, Breaking the code of DNA binding specificity of TAL-type III effectors, *Science* **2009**, *326*, 1509-1512.

[85] M. J. Moscou, A. J. Bogdanove, A simple cipher governs DNA recognition by TAL effectors, *Science* **2009**, *326*, 1501.

[86] L. Cuculis, Z. Abil, H. Zhao, C. M. Schroeder, Direct observation of TALE protein dynamics reveals a two-state search mechanism, *Nature Communications* **2015**, *6*, 7277.

[87] A. N. Mak, P. Bradley, R. A. Cernadas, A. J. Bogdanove, B. L. Stoddard, The crystal structure of TAL effector PthXo1 bound to its DNA target, *Science* **2012**, *335*, 716-719.

[88] M. Christian, T. Cermak, E. L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A. J. Bogdanove, D. F. Voytas, Targeting DNA double-strand breaks with TAL effector nucleases, *Genetics* **2010**, *186*, 757-761.

[89] M. L. Maeder, J. F. Angstman, M. E. Richardson, S. J. Linder, V. M. Cascio, S. Q. Tsai, Q. H. Ho, J. D. Sander, D. Reyon, B. E. Bernstein, J. F. Costello, M. F. Wilkinson, J. K. Joung, Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins, *Nat Biotechnol* **2013**, *31*, 1137-1142.

[90] E. M. Mendenhall, K. E. Williamson, D. Reyon, J. Y. Zou, O. Ram, J. K. Joung, B. E. Bernstein, Locus-specific editing of histone modifications at endogenous enhancers, *Nat Biotechnol* **2013**, *31*, 1133-1136.

[91] D. L. Bernstein, J. E. Le Lay, E. G. Ruano, K. H. Kaestner, TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts, *J Clin Invest* **2015**, *125*, 1998-2006.

[92] S. Bultmann, R. Morbitzer, C. S. Schmidt, K. Thanisch, F. Spada, J. Elsaesser, T. Lahaye, H. Leonhardt, Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers, *Nucleic Acids Res* **2012**, *40*, 5368-5377.

[93] B. Wiedenheft, S. H. Sternberg, J. A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea, *Nature* **2012**, *482*, 331-338.

[94] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, *Science* **2012**, *337*, 816-821.

[95] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, W. A. Lim, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression, *Cell* **2013**, *152*, 1173-1183.

[96] L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, L. S. Qi, CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes, *Cell* **2013**, *154*, 442-451.

[97] D. Bikard, W. Jiang, P. Samai, A. Hochschild, F. Zhang, L. A. Marraffini, Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system, *Nucleic Acids Res* **2013**, *41*, 7429-7437.

[98] S. Konermann, M. D. Brigham, A. E. Trevino, J. Joung, O. O. Abudayyeh, C. Barcena, P. D. Hsu, N. Habib, J. S. Gootenberg, H. Nishimasu, O. Nureki, F. Zhang, Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, *Nature* **2014**, *517*, 583.

[99] I. B. Hilton, A. M. D'Ippolito, C. M. Vockley, P. I. Thakore, G. E. Crawford, T. E. Reddy, C. A. Gersbach, Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers, *Nat Biotechnol* **2015**, *33*, 510-517.

[100] P. I. Thakore, A. M. D'Ippolito, L. Song, A. Safi, N. K. Shivakumar, A. M. Kabadi, T. E. Reddy, G. E. Crawford, C. A. Gersbach, Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements, *Nat Methods* **2015**, *12*, 1143-1149.

[101] A. Vojta, P. Dobrinic, V. Tadic, L. Bockor, P. Korac, B. Julg, M. Klasic, V. Zoldos, Repurposing the CRISPR-Cas9 system for targeted DNA methylation, *Nucleic Acids Res* **2016**, *44*, 5615-5628.

[102] T. Xiong, G. E. Meister, R. E. Workman, N. C. Kato, M. J. Spellberg, F. Turker, W. Timp, M. Ostermeier, C. D. Novina, Targeted DNA methylation in human cells using engineered dCas9-methyltransferases, *Scientific Reports* **2017**, *7*, 6732.

[103] Y. Lei, X. Zhang, J. Su, M. Jeong, M. C. Gundry, Y.-H. Huang, Y. Zhou, W. Li, M. A. Goodell, Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein, *Nature Communications* **2017**, *8*, 16026.

[104] X. Xu, Y. Tao, X. Gao, L. Zhang, X. Li, W. Zou, K. Ruan, F. Wang, G. L. Xu, R. Hu, A CRISPR-based approach for targeted DNA demethylation, *Cell Discov* **2016**, *2*, 16009.

[105] M. M. Harrison, B. V. Jenkins, K. M. O'Connor-Giles, J. Wildonger, A CRISPR view of development, *Genes Dev* **2014**, *28*, 1859-1872.

[106] A. M. Kabadi, D. G. Ousterout, I. B. Hilton, C. A. Gersbach, Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector, *Nucleic Acids Res* **2014**, *42*, e147.

[107] X. Ma, Q. Zhang, Q. Zhu, W. Liu, Y. Chen, R. Qiu, B. Wang, Z. Yang, H. Li, Y. Lin, Y. Xie, R. Shen, S. Chen, Z. Wang, Y. Chen, J. Guo, L. Chen, X. Zhao, Z. Dong, Y. G. Liu, A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants, *Mol Plant* **2015**, *8*, 1274-1284.

[108] G. Felsenfeld, D. R. Davies, A. Rich, FORMATION OF A THREE-STRANDED POLYNUCLEOTIDE MOLECULE, *Journal of the American Chemical Society* **1957**, *79*, 2023-2024.

[109] M. Duca, P. Vekhoff, K. Oussedik, L. Halby, P. B. Arimondo, The triple helix: 50 years later, the outcome, *Nucleic Acids Res* **2008**, *36*, 5123-5138.

[110] J. R. Goñi, X. de la Cruz, M. Orozco, Triplex-forming oligonucleotide target sequences in the human genome, *Nucleic Acids Research* **2004**, *32*, 354-360.

[111] K. J. Hampel, P. Crosson, J. S. Lee, Polyamines favor DNA triplex formation at neutral pH, *Biochemistry* **1991**, *30*, 4455-4459.

[112] T. Thomas, T. J. Thomas, Selectivity of polyamines in triplex DNA stabilization, *Biochemistry* **1993**, *32*, 14068-14074.

[113] M. Rougee, B. Faucon, J. L. Mergny, F. Barcelo, C. Giovannangeli, T. Garestier, C. Helene, Kinetics and thermodynamics of triple-helix formation: effects of ionic strength and mismatches, *Biochemistry* **1992**, *31*, 9269-9278.

[114] S. W. Blume, J. Lebowitz, W. Zacharias, V. Guarcello, C. A. Mayfield, S. W. Ebbinghaus, P. Bates, D. E. Jones, Jr., J. Trent, N. Vigneswaran, D. M. Miller, The integral divalent cation within

the intermolecular purine*purine. pyrimidine structure: a variable determinant of the potential for and characteristics of the triple helical association, *Nucleic Acids Res* **1999**, *27*, 695-702.

[115] H. E. Moser, P. B. Dervan, Sequence-specific cleavage of double helical DNA by triple helix formation, *Science* **1987**, *238*, 645-650.

[116] P. A. Beal, P. B. Dervan, Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation, *Science* **1991**, *251*, 1360-1363.

[117] P. Vekhoff, A. Ceccaldi, D. Polverari, J. Pylouster, C. Pisano, P. B. Arimondo, Triplex formation on DNA targets: how to choose the oligonucleotide, *Biochemistry* **2008**, *47*, 12277-12289.

[118] W. M. Olivas, L. J. Maher, 3rd, Competitive triplex/quadruplex equilibria involving guanine-rich oligonucleotides, *Biochemistry* **1995**, *34*, 278-284.

[119] P. S. Eder, R. J. DeVine, J. M. Dagle, J. A. Walder, Substrate specificity and kinetics of degradation of antisense oligonucleotides by a 3' exonuclease in plasma, *Antisense Res Dev* **1991**, *1*, 141-151.

[120] J. S. Lee, M. L. Woodsworth, L. J. P. Latimer, A. R. Morgan, Poly(pyrimidine) poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH, *Nucleic Acids Research* **1984**, *12*, 6603-6614.

[121] L. E. Xodo, G. Manzini, F. Quadrifoglio, G. A. van der Marel, J. H. van Boom, Effect of 5-methylcytosine on the stability of triple-stranded DNA--a thermodynamic study, *Nucleic Acids Res* **1991**, *19*, 5625-5631.

[122] A. Ono, P. O. P. Ts'o, L. S. Kan, Triplex formation of oligonucleotides containing 2'-O-methylpseudoisocytidine in substitution for 2'-deoxycytidine, *Journal of the American Chemical Society* **1991**, *113*, 4032-4033.

[123] L. Lacroix, J. Lacoste, J. F. Reddoch, J. L. Mergny, D. D. Levy, M. M. Seidman, M. D. Matteucci, P. M. Glazer, Triplex formation by oligonucleotides containing 5-(1-propynyl)-2'-deoxyuridine: decreased magnesium dependence and improved intracellular gene targeting, *Biochemistry* **1999**, *38*, 1893-1901.

[124] J. Basye, J. O. Trent, D. Gao, S. W. Ebbinghaus, Triplex formation by morpholino oligodeoxyribonucleotides in the HER-2/neu promoter requires the pyrimidine motif, *Nucleic Acids Res* **2001**, *29*, 4873-4880.

[125] P. E. Nielsen, M. Egholm, R. H. Berg, O. Buchardt, Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide, *Science* **1991**, *254*, 1497-1500.

[126] M. Shimizu, A. Konishi, Y. Shimada, H. Inoue, E. Ohtsuka, Oligo(2'-O-methyl)ribonucleotides. Effective probes for duplex DNA, *FEBS Lett* **1992**, *302*, 155-158.

[127] M. Petersen, J. Wengel, LNA: a versatile tool for therapeutics and genomics, *Trends Biotechnol* **2003**, *21*, 74-81.

[128] W. M. Olivas, L. J. Maher, 3rd, Overcoming potassium-mediated triplex inhibition, *Nucleic Acids Res* **1995**, *23*, 1936-1941.

[129] J. Lacoste, J. C. Francois, C. Helene, Triple helix formation with purine-rich phosphorothioate-containing oligonucleotides covalently linked to an acridine derivative, *Nucleic Acids Res* **1997**, *25*, 1991-1998.

[130] H. Torigoe, Y. Hari, M. Sekiguchi, S. Obika, T. Imanishi, 2'-O,4'-C-methylene bridged nucleic acid modification promotes pyrimidine motif triplex DNA formation at physiological pH: thermodynamic and kinetic studies, *J Biol Chem* **2001**, *276*, 2354-2360.

[131] T. G. Uil, H. J. Haisma, M. G. Rots, Therapeutic modulation of endogenous gene function by agents with designed DNA-sequence specificities, *Nucleic Acids Res* **2003**, *31*, 6064-6078.

[132] H. Porumb, H. Gousset, R. Letellier, V. Salle, D. Briane, J. Vassy, M. Amor-Gueret, L. Israël, E. Taillandier, Temporary ex Vivo Inhibition of the Expression of the Human Oncogene HER2 (NEU) by a Triple Helix-forming Oligonucleotide, *Cancer Research* **1996**, *56*, 515-522.

[133] V. M. Macaulay, P. J. Bates, M. J. McLean, M. G. Rowlands, T. C. Jenkins, A. Ashworth, S. Neidle, Inhibition of aromatase expression by a psoralen-linked triplex-forming oligonucleotide targeted to a coding sequence, *FEBS Letters* **1995**, *372*, 222-228.

[134] O. Boussif, F. Lezoualc'h, M. A. Zanta, M. D. Mergny, D. Scherman, B. Demeneix, J. P. Behr, A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine, *Proceedings of the National Academy of Sciences* **1995**, *92*, 7297-7301.

[135] L. M. Santhakumaran, T. Thomas, T. J. Thomas, Enhanced cellular uptake of a triplex-forming oligonucleotide by nanoparticle formation in the presence of polypropylenimine dendrimers, *Nucleic Acids Res* **2004**, *32*, 2102-2112.

[136] K. Cheng, Z. Ye, R. V. Guntaka, R. I. Mahato, Enhanced Hepatic Uptake and Bioactivity of Type α1(I) Collagen Gene Promoter-Specific Triplex-Forming Oligonucleotides after Conjugation with Cholesterol, *Journal of Pharmacology and Experimental Therapeutics* **2006**, *317*, 797-805.

[137] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, A general method for the covalent labeling of fusion proteins with small molecules in vivo, *Nat Biotechnol* **2003**, *21*, 86-89.

[138] A. Juillerat, T. Gronemeyer, A. Keppler, S. Gendreizig, H. Pick, H. Vogel, K. Johnsson, Directed evolution of O6-alkylguanine-DNA alkyltransferase for efficient labeling of fusion proteins with small molecules in vivo, *Chem Biol* **2003**, *10*, 313-317.

[139] A. Keppler, M. Kindermann, S. Gendreizig, H. Pick, H. Vogel, K. Johnsson, Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro, *Methods* **2004**, *32*, 437-444.

[140] A. Keppler, C. Arrivoli, L. Sironi, J. Ellenberg, Fluorophores for live cell imaging of AGT fusion proteins across the visible spectrum, *Biotechniques* **2006**, *41*, 167-170, 172, 174-165.

[141] A. Keppler, H. Pick, C. Arrivoli, H. Vogel, K. Johnsson, Labeling of fusion proteins with synthetic fluorophores in live cells, *Proc Natl Acad Sci U S A* **2004**, *101*, 9955-9959.

[142] G. J. Gu, M. Friedman, C. Jost, K. Johnsson, M. Kamali-Moghaddam, A. Pluckthun, U. Landegren, O. Soderberg, Protein tag-mediated conjugation of oligonucleotides to recombinant affinity binders for proximity ligation, *N Biotechnol* **2013**, *30*, 144-152.

[143] A. Gautier, A. Juillerat, C. Heinis, I. R. Correa, Jr., M. Kindermann, F. Beaufils, K. Johnsson, An engineered protein tag for multiprotein labeling in living cells, *Chem Biol* **2008**, *15*, 128-136.

[144] H. C. Kolb, M. G. Finn, K. B. Sharpless, Click Chemistry: Diverse Chemical Function from a Few Good Reactions, *Angew Chem Int Ed Engl* **2001**, *40*, 2004-2021.

[145] J. A. Prescher, C. R. Bertozzi, Chemistry in living systems, *Nat Chem Biol* **2005**, *1*, 13-21.

[146] J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli, C. R. Bertozzi, Copper-free click chemistry for dynamic in vivo imaging, *Proc Natl Acad Sci U S A* **2007**, *104*, 16793-16797.

[147] E. Lallana, E. Fernandez-Megia, R. Riguera, Surpassing the use of copper in the click functionalization of polymeric nanostructures: a strain-promoted approach, *J Am Chem Soc* **2009**, *131*, 5748-5750.

[148] E. M. Sletten, C. R. Bertozzi, Bioorthogonal chemistry: fishing for selectivity in a sea of functionality, *Angew Chem Int Ed Engl* **2009**, *48*, 6974-6998.

[149] N. J. Agard, J. A. Prescher, C. R. Bertozzi, A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems, *J Am Chem Soc* **2004**, *126*, 15046-15047.
[150] G. Wittig, A. Krebs, Zur Existenz niedergliedriger Cycloalkine, I, *Chemische Berichte* **1961**, *94*, 3260-3275.
[151] R. B. Turner, A. D. Jarrett, P. Goebel, B. J. Mallon, Heats of hydrogenation. IX. Cyclic acetylenes and some miscellaneous olefins, *Journal of the American Chemical Society* **1973**, *95*, 790-792.
[152] J. A. Codelli, J. M. Baskin, N. J. Agard, C. R. Bertozzi, Second-generation difluorinated cyclooctynes for copper-free click chemistry, *J Am Chem Soc* **2008**, *130*, 11486-11493.
[153] X. Ning, J. Guo, M. A. Wolfert, G. J. Boons, Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions, *Angew Chem Int Ed Engl* **2008**, *47*, 2253-2255.
[154] M. F. Debets, S. S. van Berkel, S. Schoffelen, F. P. Rutjes, J. C. van Hest, F. L. van Delft, Aza-dibenzocyclooctynes for fast and efficient enzyme PEGylation via copper-free (3+2) cycloaddition, *Chem Commun (Camb)* **2010**, *46*, 97-99.
[155] J. C. Jewett, E. M. Sletten, C. R. Bertozzi, Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones, *J Am Chem Soc* **2010**, *132*, 3688-3690.
[156] E. M. Sletten, C. R. Bertozzi, A hydrophilic azacyclooctyne for Cu-free click chemistry, *Org Lett* **2008**, *10*, 3097-3099.
[157] G. de Almeida, E. M. Sletten, H. Nakamura, K. K. Palaniappan, C. R. Bertozzi, Thiacycloalkynes for copper-free click chemistry, *Angew Chem Int Ed Engl* **2012**, *51*, 2443-2447.
[158] X. Ning, R. P. Temming, J. Dommerholt, J. Guo, D. B. Ania, M. F. Debets, M. A. Wolfert, G. J. Boons, F. L. van Delft, Protein modification by strain-promoted alkyne-nitrone cycloaddition, *Angew Chem Int Ed Engl* **2010**, *49*, 3065-3068.
[159] S. S. van Berkel, A. T. Dirks, M. F. Debets, F. L. van Delft, J. J. Cornelissen, R. J. Nolte, F. P. Rutjes, Metal-free triazole formation as a tool for bioconjugation, *Chembiochem* **2007**, *8*, 1504-1508.
[160] M. L. Blackman, M. Royzen, J. M. Fox, Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity, *J Am Chem Soc* **2008**, *130*, 13518-13519.
[161] B. T. van der Gun, M. Maluszynska-Hoffman, A. Kiss, A. J. Arendzen, M. H. Ruiters, P. M. McLaughlin, E. Weinhold, M. G. Rots, Targeted DNA methylation by a DNA methyltransferase coupled to a triple helix forming oligonucleotide to down-regulate the epithelial cell adhesion molecule, *Bioconjug Chem* **2010**, *21*, 1239-1245.
[162] M. Maluszynska-Hoffman, Super-specific DNA methylation by a DNA methyltransferase coupled with a triple helix-forming oligonucleotide, **2008**.
[163] S. H. Chiou, DNA- and protein-scission activities of ascorbate in the presence of copper ion and a copper-peptide complex, *J Biochem* **1983**, *94*, 1259-1267.
[164] P. Albert, B. Varga, N. Zsibrita, A. Kiss, Circularly permuted variants of two CG-specific prokaryotic DNA methyltransferases, *PLoS One* **2018**, *13*, e0197232.
[165] M. V. Darii, N. A. Cherepanova, O. M. Subach, O. V. Kirsanova, T. Raskó, K. Ślaska-Kiss, A. Kiss, D. Deville-Bonne, M. Reboud-Ravaux, E. S. Gromova, Mutational analysis of the CG recognizing DNA methyltransferase SssI: Insight into enzyme–DNA interactions, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2009**, *1794*, 1654-1662.
[166] N. Sugimoto, P. Wu, H. Hara, Y. Kawamoto, pH and Cation Effects on the Properties of Parallel Pyrimidine Motif DNA Triplexes, *Biochemistry* **2001**, *40*, 9396-9405.

[167] F. Romina, S. Bruna, M. Giorgio, Q. Franco, X. L. E., Effect of cations on purine·purine·pyrimidine triple helix formation in mixed-valence salt solutions, *European Journal of Biochemistry* **1999**, *260*, 801-809.
[168] C. T. Harrington, E. I. Lin, M. T. Olson, J. R. Eshleman, Fundamentals of Pyrosequencing, *Archives of Pathology & Laboratory Medicine* **2013**, *137*, 1296-1303.
[169] R. Y. Wang, C. W. Gehrke, M. Ehrlich, Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues, *Nucleic Acids Res* **1980**, *8*, 4777-4790.
[170] M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands, *Proc Natl Acad Sci U S A* **1992**, *89*, 1827-1831.
[171] S. Nunna, R. Reinhardt, S. Ragozin, A. Jeltsch, Targeted methylation of the epithelial cell adhesion molecule (EpCAM) promoter to silence its expression in ovarian cancer cells, *PLoS One* **2014**, *9*, e87703.
[172] M. M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal Biochem* **1976**, *72*, 248-254.

Appendix

Analysis of BG-TAMRA (21)

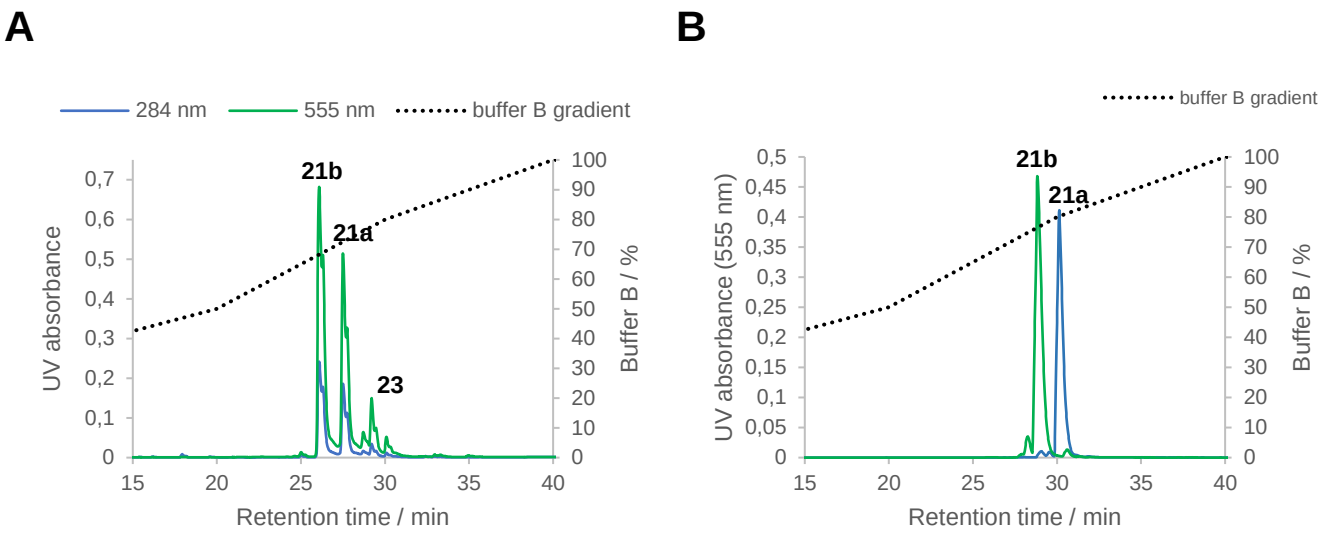


Fig. 1: rp HPLC analysis of A the reaction of BG-NH_2 (**22**) and TAMRA-NHS (**23**) to form BG-TAMRA (**21**) after 90 min (method 4) and B the purified BG-TAMRA isomers **21a** (blue) and **21b** (green) (method 4).

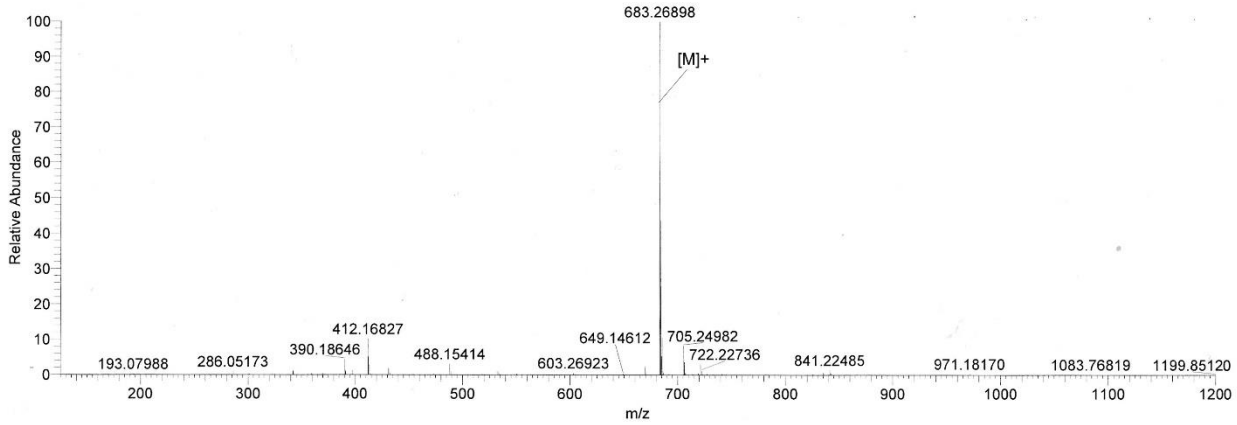


Fig. 2: MS of **21a** (ESI, positive ion mode).

Amino acid sequences of SNAP-M.SssI, SNAP-M.MpeI and SNAP-M.MpeI(Q142L)

The sequences of SNAP-M.SssI, SNAP-M.MpeI and SNAP-M.MpeI(Q142L) are listed below. The interchanged leucine and glutamine residues within SNAP-M.MpeI and SNAP-M.MpeI(Q142L) are highlighted. The physical parameters were calculated with the ProtParam tool (www.expasy.com).

SNAP-M.SssI

Molecular weight: 66845.60 g/mol

Theoretical pI: 8.56

Extinction coefficient (280 nm): 72560 $M^{-1} \cdot cm^{-1}$

MASTMDIKLT GEFAMDKDCE MKRTTLDSPL GKLELSGCEQ GLHEIKLLGK GTSAADAVEV PAPAAVLGGP
EPLMQATAWL NAYFHQPEAI EEFPVPALHH PVFQQESFTR QVLWKLLKVV KFGEVISYQQ LAALAGNPAA
TAAVKTALSG NPVPILIPCH RVVSSSGAVG GYEGGLAVKE WLLAHEGHRL GKPGLGPAGG SPGLECMSKV
ENKTKKLRVF EAFAGIGAQR KALEKVRKDE YEIVGLAEWY VPAIVMYQAI HNNFHTKLEY KSVSREEMID
YLENKTLSWN SKNPVSNGYW KRKKDDELKI IYNAIKLSEK EGNIFDIRDL YKRTLKNIDL LTYSFPCQDL
SQQGIQKGMK RGSGTRSGLL WEIERALDST EKNDLPKYLL MENVGALLHK KNEEELNQWK QKLESLGYQN
SIEVLNAADF GSSQARRRVF MISTLNEFVE LPKGDKKPKS IKKVLNKIVS EKDILNNLLK YNLTEFKKTK
SNINKASLIG YSKFNSEGYV YDPEFTGPTL TASGANSRIK IKDGSNIRKM NSDETFLYMG FDSQDGKRVN
EIEFLTENQK IFVAGNSISV EVLEAIIDKI GGSHHHHHH

SNAP-M.MpeI

Molecular weight: 68244.35 g/mol

Theoretical pI: 8.99

Extinction coefficient (280 nm): 71655 $M^{-1} \cdot cm^{-1}$

```
MASTMDIKLT GEFAMDKDCE MKRTTLDSPL GKLELSGCEQ GLHEIKLLGK GTSAADAVEV PAPAAVLGGP
EPLMQATAWL NAYFHQPEAI EEFPVPALHH PVFQQESFTR QVLWKLLKVV KFGEVISYQQ LAALAGNPAA
TAAVKTALSG NPVPILIPCH RVVSSSGAVG GYEGGLAVKE WLLAHEGHRL GKPGLGPAGG SDSNKDKIKV
IKVFEAFAGI GSQFKALKNI ARSKNWEIQH SGMVEWFVDA IVSYVAIHSK NFNPKIERLD RDILSISNDS
KMPISEYGIK KINNTIKASY LNYAKKHFNN LFDIKKVNKD NFPKNIDIFT YSFPCQDLSV QGLQKGIDKE
LNTRSGLLWE IERILEEIKN SFSKEEMPKY LLMENVKNLL SHKNKKNYNT WLKQLEKFGY KSKTYLLNSK
NFDNCQNRER VFCLSIRDDY LEKTGFKFKE LEKVKNPPKK IKDILVDSSN YKYLNLNKYE TTTFRETKSN
IISRPLKNYT TFNSENYVYN INGIGPTLTA SGANSRIKIE TQQGVRYLTP LECFKYMQFD VNDFKKVQST
NLISENKMIY IAGNSIPVKI LEAIFNTLEF VNNEELEHHH HHH
```

SNAP-M.MpeI(Q142L)

Molecular weight: 68229.38 g/mol

Theoretical pI: 8.99

Extinction coefficient (280 nm): 71655 $M^{-1} \cdot cm^{-1}$

```
MASTMDIKLT GEFAMDKDCE MKRTTLDSPL GKLELSGCEQ GLHEIKLLGK GTSAADAVEV PAPAAVLGGP
EPLMQATAWL NAYFHQPEAI EEFPVPALHH PVFQQESFTR QVLWKLLKVV KFGEVISYQQ LAALAGNPAA
TAAVKTALSG NPVPILIPCH RVVSSSGAVG GYEGGLAVKE WLLAHEGHRL GKPGLGPAGG SDSNKDKIKV
IKVFEAFAGI GSQFKALKNI ARSKNWEIQH SGMVEWFVDA IVSYVAIHSK NFNPKIERLD RDILSISNDS
KMPISEYGIK KINNTIKASY LNYAKKHFNN LFDIKKVNKD NFPKNIDIFT YSFPCQDLSV LGLQKGIDKE
LNTRSGLLWE IERILEEIKN SFSKEEMPKY LLMENVKNLL SHKNKKNYNT WLKQLEKFGY KSKTYLLNSK
NFDNCQNRER VFCLSIRDDY LEKTGFKFKE LEKVKNPPKK IKDILVDSSN YKYLNLNKYE TTTFRETKSN
IISRPLKNYT TFNSENYVYN INGIGPTLTA SGANSRIKIE TQQGVRYLTP LECFKYMQFD VNDFKKVQST
NLISENKMIY IAGNSIPVKI LEAIFNTLEF VNNEELEHHH HHH
```

Litcon54 DNA plasmid sequence

All 177 CpG sites are marked in red. The inserted sequence is underlined. The TFS is marked in blue. The four challenged CpG sites within the R.Psp1406I recognition sequence (AACGTT) are indicated in bold red. The target CpG is bold and circled.

GTTAACTACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGT
ATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTC
GCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGA
TCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA**CG**TT
CTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGC
CGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGA
ATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAA
CCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGAC
GAGCGTGACACCACGATGCCTGTAGCAATGGCAACAA**CG**TTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCG
GCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTG
CTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTA
GTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTACCCCGGTTGATAATCAGAAAAGCCCCAAAAACAG
GAAGATTGTATAAGCAAATATTTAAATTGTAA**CG**TTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCAT
TTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGTTCCAGTT
TGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACG
TGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTA
GAGCTTGACGGGGAAAGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAG
CGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTAAAAGGATCTAGGTGAAGATCCT
TTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGAT
CTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG
GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCC
GTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA
GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGT
TCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCT
TCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAA
ACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGG
AGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTAATGTGAGTT
AGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTT
CACACAGGAAACAGCTATGACCATGATTACGCCAAGCTACGTAATACGACTCACTATAGGGCAGATCTTCGAATGCATCGCGC
GCACCGTACGTCTCGAGGAATTCCTGCAGGATATCTGGATCCACGAAGCTTCCCATGGTGACGTCACCGGTTCTAGATACCTA
GGTGAGCTCCTTTTTCTTTTTCTCTCTTTTTTTTTTTTTTTGAGACGGAGTCTAA**CG**TTGGTACCCTCTAGTCAAGGCCTATAGT
GAGTCGTATTACGGACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGC
ACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCG
AATGGCGCTTCGCTTGGTAATAAAGCCCGCTTCGGCGGGCTTTTTTTT

Litcon47 DNA plasmid sequence

All 176 CpG sites are marked in red. The inserted sequence is underlined. The four challenged CpG sites within the R.Psp1406I recognition sequence (AACGTT) are indicated in bold red.

GTTAACTACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGT
ATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAACATTTCCGTGTC
GCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGA
TCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAA**CGTT**
CTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGC
CGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGA
ATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAA
CCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGAC
GAGCGTGACACCACGATGCCTGTAGCAATGGCAACAA**CGTT**GCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCG
GCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTG
CTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTA
GTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCA
TTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTACCCGGTTGATAATCAGAAAAGCCCCAAAAACAG
GAAGATTGTATAAGCAAATATTTAAATTGTAA**CGTT**AATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCAT
TTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGCCCGAGATAGGGTTGAGTGTTGTTCCAGTT
TGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACG
TGAACCATCACCCAAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCCGATTTA
GAGCTTGACGGGGAAAGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAG
CGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCGCGTAAAAGGATCTAGGTGAAGATCCT
TTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGAT
CTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG
GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCC
GTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCA
GTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGT
TCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCT
TCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAA
ACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGG
AGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTAATGTGAGTT
AGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTT
CACACAGGAAACAGCTATGACCATGATTACGCCAAGCTACGTAATACGACTCACTATAGGGCAGATCTTCGAATGCATCGCGC
GCACCGTACGTCTCGAGGAATTCCTGCAGGATATCTGGATCCACGAAGCTTCCCATGGTGACGTCACCGGTTCTAGATACCTA
GGTGAGCTAA**CGTT**TGCCTAATCATCTTTTGTTCATGTCCTACTGTTCAAGCCTCCAAGCTGTACCCTCTAGTCAAGGCCTAT
AGTGAGTCGTATTACGGACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGC
AGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATG
GCGAATGGCGCTTCGCTTGGTAATAAAGCCCGCTTCGGCGGGCTTTTTTTT

Pyrosequencing of Litcon54

Region	CpGs	Primer sequences
P1	8	Fw: TAGTTGGAGAAAGGGGGATGT
		Rv: GGGACACCGCTGATCGTTTAATAAACTCCTTTTCTTTTCTCTCTT
		Seq: TGTAAGGAGATTAAGTTGGG
P2	16	Fw: AGAGAGAAAAGAAAAGGAGTTTATT
		Rv: GGGACACCGCTGATCGTTTAACACCCCAAACTTTACACTTTAT
		Seq: AGAAAAGGAGTTTATTTAGG
P3	5	Fw: GATTTTAAAAAATTTGATTTGGGTGATGG
		Rv: GGGACACCGCTGATCGTTTAATTCCAATTTAAAACAAAAATCCACTATT
		Seq: ATGGTTTAAGTAGTGGGTTAT
P4	8	Fw: GTTTTTTTGTGATTGGTGAGTATTTAA
		Rv: GGGACACCGCTGATCGTTTACCAATAATAAACACTTTTAAAATTCTACT
		Seq: AAGTTATTTTGAGAATAGTGTA

Universal biotin primer sequence: GGGACACCGCTGATCGTTTA

Bisulfite converted sequence of the reverse strand

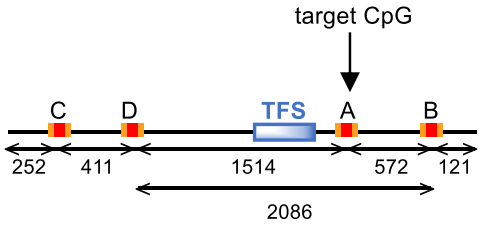
The four challenged CpG sites within the R.Psp1406I recognition sequence (AACGTT) are in bold red. The regions of interest P1, P2, P3 and P4 are highlighted and the CpG sites within the regions are marked in red. The analyzed CpGs are additionally highlighted in gray.

AAAAAAAAGTTCGTCGAAGCGGGTTTTATTATTAAGCGAAGCGTTATTCGTTATTTAGGTTGCGTAATTGTTGGGA
AGGGCGATCGGTGCGGGTTTTTTCGTTATTACGTTAGTTGGCGAAAGGGGGATGTGTTGTAAGGCGATTAAGTTGG
GTAACGTTAGGGTTTTTTTAGTTACGACGTTGTAAAACGACGGTTAGTTCGTAATACGATTTATTATAGGTTTTGA
TTAGAGGGTATTAACGTTAGATTTCGTTTTAAAAAAAAAAAAAAAGAGAGAAAAGAAAAGGAGTTTATTTAGGTAT
TTAGAATCGGTGACGTTATTATGGGAAGTTTCGTGGATTTAGATATTTTGTAGGAATTTTTCGAGACGTACGGTGC
GCGCGATGTATTCGAAGATTTGTTTTATAGTGAGTCGTATTACGTAGTTTGGCGTAATTATGGTTATAGTTGTTTT
TTGTGTGAAATTGTTATTCGTTTATAATTTTATATAATATACGAGTCGGAAGTATAAAGTGTAAAGTTTGGGGTGT
TTAATGAGTGAGTTAATTTATATTATATGTGAGTAAAAGGTTAGTAAAAGGTTAGGAATCGTAAAAAGGTCGCGTT
GTTGGCGTTTTTTTATAGGTTTCGTTTTTTTGACGAGTATTATAAAAATCGACGTTTAAGTTAGAGGTGGCGAAAT
TCGATAGGATTATAAAGATATTAGGCGTTTTTTTTTGGAAGTTTTTTCGTGCGTTTTTTTGTTTCGATTTTGTCGT
TTATCGGATATTTGTTCGTTTTTTTTTTTTCGGGAAGCGTGGCGTTTTTTTATAGTTTACGTTGTAGGTATTTTAG
TTCGGTGTAGGTCGTTCGTTTTAAGTTGGGTTGTGTGTACGAATTTTTCGTTTAGTTCGATCGTTGCGTTTTATTC
GGTAATTATCGTTTTGAGTTTAATTCGGTAAGATACGATTTATCGTTATTGGTAGTAGTTATTGGTAATAGGATTA
GTAGAGCGAGGTATGTAGGCGGTGTTATAGAGTTTTTGAAGTGGTGGTTTAATTACGGTTATATTAGAAGAATAGT
ATTTGGTATTTGCGTTTTGTTGAAGTTAGTTATTTTCGGAAAAAGAGTTGGTAGTTTTTGATTCGGTAAATAAATT
ATCGTTGGTAGCGGTGGTTTTTTTGTTTGTAAGTAGTAGATTACGCGTAGAAAAAAAGGATTTTAAGAAGATTTTT
TGATTTTTTTTACGGGGTTTGACGTTTAGTGGAACGAAAATTTACGTTAAGGGATTTTGGTTATGAGATTATTAAA
AAGGATTTTTATTTAGATTTTTTTACGCGTTTTGTAGCGGCGTATTAAGCGCGGCGGGTGTGGTGGTTACGCGTAG
CGTGATCGTTATATTTGTTAGCGTTTTAGCGTTCGTTTTTTTCGTTTTTTTTTTTTTTTTTCGTTACGTTCGTT
TTTTTCGTTAAGTTTTAAATCGGGGGTTTTTTTTAGGGTTTCGATTTAGTGTTTTACGGTATTTCGATTTTAAAAA
ATTTGATTTGGGTGATGGTTTACGTAGTGGGTTATCGTTTTGATAGACGGTTTTTCGTTTTTTGACGTTGGAGTTT
ACGTTTTTTTAATAGTGGATTTTTGTTTTAAATTGGAATAATATTTAATTTTATTTCGGGTTATTTTTTTGATTTAT
AAGGGATTTTGTCGATTTCGGTTTATTGGTTAAAAAATGAGTTGATTTAATAAAAATTTAACGCGAATTTTAATAA
AATATTAACGTTTATAATTTAAATATTTGTTTATATAATTTTTTTGTTTTTGGGGTTTTTTTGATTATTAATCGGG
GTAAATTAATTTAAAGTATATATGAGTAAATTTGGTTTGATAGTTATTAATGTTTAATTAGTGAGGTATTTATTTT
AGCGATTTGTTTATTTCGTTTATTTATAGTTGTTTGATTTTTCGTCGTGTAGATAATTACGATACGGGAGGGTTTA
TTATTTGGTTTTAGTGTTGTAATGATATCGCGAGATTTACGTTTATCGGTTTTAGATTTATTAGTAATAAATTAGT
TAGTCGGAAGGGTCGAGCGTAGAAGTGGTTTTGTAATTTTATTCGTTTTTATTTAGTTTATTAATTGTTGTCGGGA
AGTTAGAGTAAGTAGTTCGTTAGTTAATAGTTTGCGTAACGTTGTTGTTATTGTTATAGGTATCGTGGTGTTACGT
TCGTCGTTTGGTATGGTTTTATTTAGTTTCGGTTTTTAACGATTAAGGCGAGTTATATGATTTTTTATGTTGTGTA
AAAAAGCGGTTAGTTTTTTCGGTTTTTCGATCGTTGTTAGAAGTAAGTTGGTCGTAGTGTTATTATTTATGGTTAT
GGTAGTATTGTATAATTTTTTTATTGTTATGTTATTCGTAAGATGTTTTTTTGTGATTGGTGAGTATTTAATTAAG
TTATTTTGAGAATAGTGTATGCGGCGATCGAGTTGTTTTTGTTCGGCGTTAATACGGGATAATATCGCGTTATATA
GTAGAATTTTAAAAGTGTTTATTATTGGAGAACGTTTTCGGGGCGAAAATTTTTAAGGATTTTATCGTTGTTGAG
ATTTAGTTCGATGTAATTTATTCGTGTATTTAATTGATTTTTAGTATTTTTTATTTTTATTAGCGTTTTTGGGTGA
GTAAAAATAGGAAGGTAAAATGTCGTAAAAAAGGGAATAAGGGCGATACGGAAATGTTGAATATTTATATTTTTTT
TTTTTTAATATTATTGAAGTATTTATTAGGGTTATTGTTTTATGAGCGGATATATATTTGAATGTATTTAGAAAAA
TAAATAAATAGGGGTTTCGCGTATATTTTTTCGAAAAGTGTTATTTGACGTAGTTAAT

Possible fragments of linear Litcon54

Fragments that are observed on the agarose gels are marked with an asterisk. The site A-specific band is additionally bold.

	Protection of site	Fragment size (bp)
No protection:	-	121*
	-	252*
	-	411*
	-	572*
	-	1514*
Single protection:	A	**2086***
	B	693*
	C	663*
	D	1925*
Double protection:	A + B	2207
	A + C	**2086*** + 663*
	A + D	2497
	B + C	693* + 663*
	B + D	693* + 1925*
	C + D	2177
Triple protection:	A + B + C	2207 + 663*
	A + B + D	2618
	A + C + D	2749
	B + C + D	693* + 2177
Quadruple protection:	A + B + C + D	2870*



Acknowledgements

I would like to express my greatest gratitude to all the people who helped realizing this work.

In the first place I want to thank Elmar, for giving me the opportunity to work in his lab on this exciting theme, that suited my interests very well. Thank you for your guidance and your support.
Thank you, Antal, not only for agreeing to be my second examiner, but also for all the interesting discussions and the useful suggestions for new experiments. Thanks for your hospitality during my stay in Szeged. Also, thank you, ánd Saci, for providing me with the plasmids for the SNAP-tag fusions proteins, which were essential for this work. I especially want to thank Saci for her friendship that developed from a few nice collaborations, both in Aachen and in Szeged. Thank you for showing me around in Szeged and introducing me into the world of Rétes and fitness croissants.
I also thank everyone else that was involved in the ERA chemistry project for the nice and motivating meetings, that always felt like a little reunion. A huge thank you to Archie, for planning and performing the pyrosequencing experiments, and for patiently answering my many questions.
Off course, I also want to thank all my colleagues and former colleagues, for creating a pleasant working atmosphere in the lab and for all the fun things we did together outside working hours, like cookie baking, canoeing, going to the Rhein in Flammen and singing (with Gisi!). Thanks to Britta, for introducing me into the group! Matyas, thank you for being my desk neighbor for all those years and for always offering a helping hand in the lab. Felix, thank you for your help, whether it was related to analyzing gels or fixing computers. Thanks for making delicious Carbonara, I am looking forward to the recipe.. Thank you, Andreas, for your useful advice on the microbiological experiments. I thank Kerstin for helping me with protein expressions and her expertise on FPLC. I reluctantly also “thank” Claudia for all the, sometimes frustrating, time I had to spend with her in the lab. I did learn a lot about HPLC though…
Finally, I want to thank my family. Pieter, my information source when it comes to traveling (by train), German grammatic or glucose and insulin. As my older brother you were my big example. Thank you for setting the bar so high. Thank you, papa en mama. You are more than my personal taxi and the lady who gets new cheese. You raised me, you allowed me to explore my creativity. And you always supported me; through thick and thin.

Curriculum vitae

Personal data

Name: Giesbertz, Anna Louise

Date of birth: 6th October 1988

Place of birth: Voerendaal, the Netherlands

Nationality: Dutch

Education

04/2013 – 07/2018:	PhD studies under supervision of Prof. Dr. Elmar Weinhold Thesis title: "Triple helix-targeted DNA methylation with DNA MTase-oligodeoxynucleotide conjugates"
10/2010 – 02/2013:	Master of Science in Chemistry, RWTH Aachen University Aachen, Germany; Master thesis title: "Synthesis of a double activated AdoMet analogue for one step fluorescence labeling of DNA"
10/2006 – 07/2010:	Bachelor of Science in Chemistry, RWTH Aachen University Aachen, Germany
2000 – 2006:	Secondary school Sintermeertencollege Heerlen, the Netherlands

Publications

- Hanz, G. M., Jung, B., Giesbertz, A., Juhasz, M., Weinhold, E. Sequence-specific Labeling of Nucleic Acids and Proteins with Methyltransferases and Cofactor Analogues. *J. Vis. Exp.* **2014**, (93), e52014, DOI: 10.3791/52014.
- Grunwald, A., Dahan, M., Giesbertz, A., Nilsson, A., Nyberg, LK., Weinhold, E., Ambjörnsson, T., Westerlund, F., Ebenstein, Y. Bacteriophage strain typing by rapid single molecule analysis. Nucleic Acids Res. **2015,** 43(18), e117, DOI: 10.1093/nar/gkv563.