Studies of

Transcription-Activation-Like-Effectors

Interacting with the Epigenetic Mark

N^6-Methyl Adenine

Dissertation

submitted for the degree of
Doctor of Natural Sciences
(Dr. rer. nat.)

presented by
Sarah Flade, M.Sc.

Technische Universität Dortmund
Faculty of Chemistry and Chemical Biology

Dortmund, 2017

Eidesstattliche Erklärung

Hiermit versichere Ich an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst habe. Alle in dieser Arbeit in Anspruch genommenen Quellen und Hilfsmittel wurden kenntlich vermerkt.

Dortmund, den 13. Juli 2017

Gutachter 1: Prof. Daniel Summerer, TU Dortmund.

Gutachter 2: Prof. Daniel Rauh, TU Dortmund.

Disputation: Dortmund, den 09. November 2017.

Fakultät für Chemie und Chemische Biologie.

Table of Contents

Acknowledgement

1. Introduction

1.1. Epigenetics

The human body is buildup of over 200 diverse kinds of cell types, yet all share the genomic DNA sequence as an almost identical genetic blueprint.[1] This genetic information is heritable and passed on to the next generation after cell division. Even though all cells arise from the same stem cell, one genotype can result in different phenotypes. This further implies that genes determining the phenotype are permanently turned on while others are switched off. This pattern of active and inactive genes is stably kept and transferred to the new cell. When this phenomenon was discovered, the question was raised whether additional information is passed on to the daughter cell beyond the canonical DNA-sequence.

Conrad Waddington proposed in the early 1940s the term of "epigenetics".[2] He coined this term as a variety of developmental processes above the level of the genome.[3] Mechanisms by which the genotype of a cell emerges into several special phenotypes. Waddington knew these developmental processes were separate from the genome and he was able to bridge the gap between genotype and phenotype.[4] If an organism is fully developed its overall genome remains stable, apart from random mutations. The functional determination, which part of the genome is transcribed and therefore which genes are expressed are based on several mechanisms. For instance, biochemical modifications at single DNA bases or histone modifications determine the accessibility of the DNA.

More than 60 years ago, it was discovered that cytosine (C) deoxynucleotides could carry methyl groups (5-methyl cytosine, 5mC) in eukaryotic genomic DNA. Since then, 5mC has been intensively studied for its role as carrier of epigenetic information. In 1975 two key publications on 5mC as well as X-chromosome inactivation due to methylation events, demonstrated that the modification of C could serve as an epigenetic mark.[5,6] In their work Holliday, Pugh and Riggs suggested that the DNA sequence could be methylated *de novo* and that this mark could be passed on during cell division while the additional information is stored in the form of methyl groups on the DNA itself (see figure 1). They speculated that DNA binding proteins could sense and interpret the presence of methyl groups in the DNA.

Moreover, the methylation of the DNA could directly silence a gene which would bring the phenotype of a cell into being.[7]
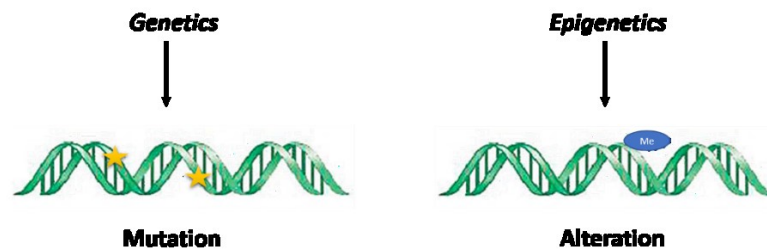
Figure 1: Differentiation between genetics and epigenetics. A mutation is a change in the setup of the DNA and might result from imperfect DNA replication. Epigenetics on the other hand are reversible and based on alterations of gene expression or the accessibility of the gene.

1.1 Methylation of Cytosine

Today methylation of the 5th position of Cytosine is one of the best studied epigenetic modifications. 5mC is the most common modification in vertebrates. In mammalian cells this epigenetic mark predominantly appears in so called CpG dinucleotides.[7] Regions with high CpG density at transcription starting sites (TSS) are shown to have an influence on gene control, e.g. methylation on TSS can block initiation of the DNA polymerase while the methylation on the gene body itself does not impair transcription.[7] Other studies of targeted *de novo* methylation and demethylation revealed that genes can be switched from their expressed to their silenced state and vice versa.[8] Another function of 5mC is to maintain the stability of the genome. Methylation of cytosine in repeat regions like centromeres are considered important for chromosomal stability due to suppression of transposons.[9] Due to this manifold involvement in cellular processes, 5mC is an important factor to keep cellular processes functional. In normal cells, CpG islands are usually not methylated. In contrast, tumor cells exhibit global hypo-methylation of DNA. This loss of epigenetic modifications leads to methylation deficiency in repetitive DNA sequences, causing silent CpGs to become active. Moreover, coding regions as well as introns exhibit reduced methylation which finally results in chromosomal and therefore genomic instability. This further causes the reactivation of transposable elements and favors meiotic recombination.[10,11] Along with the overall

hypo-methylation of tumor DNA, hyper-methylated tumor suppressor genes were identified. [10] In contrast to normal cells, tumor suppressor genes become hyper-methylated which results in the silencing of gene expression. In this major event of cancer development many genes can be affected which play a significant role in cell cycle control, apoptosis as well as DNA repair. A known example for hyper-methylation of a tumor suppressor gene is the breast cancer susceptibility gene 1 (BRCA1) and its consequential inactivation in the development of breast cancer. [11] Taken together, 5mC plays a major role in regulation of gene expression and maintaining the genome stability. [12]

DNA methylation is part of a complex epigenetic network. Next to the modification of DNA, other epigenetic events are involved. For example, histone proteins do not only function in DNA packaging but their molecular structure can influence gene expression. Here, the epigenetic information is stored on the histone in form acetylation of lysine (K), methylation of K and arginine (R), as well as phosphorylation of serine (S). These modifications are post-translational introduced to the histone tails and can be in a single, double or even triple fashion, e.g. histone 3 can be single methylated at lysine residue 3 (H3K3me) which leads to activation of transcription. [10] These modifications on the histone proteins influence gene transcription and DNA repair. However, this "histone code" is very complex and it is difficult to predict the outcome of a modification event, since the functional consequences do not rely on a single modification but on the interplay of several modifications. While acetylation is in general associated with activation, the outcome of methylation events depends on the targeted amino acid and the residue position within the histone tail. E.g. methylation at H3K9 and H3K20 in combination with H4K20 leads to the repression of gene transcription. [13–16] In cancer cells, the histones exhibit a different pattern of methylation and acetylation compared to normal cells. Hence, the common interplay of histone modifications might be disrupted in cancer cells.

Epigenetic signals are reversible. The mechanism to introduce and remove, as well as maintain epigenetic marks involves participation of DNA-methyltransferases (DMT) and Demethylases. In DNA, DMTs are essential for setting up methylation patterns and catalyze the methylation of C to 5mC using S-adenosyl methionine (SAM) as methyl donor. [17] For a long time it was believed that passive dilution of 5mC based on replication of the cell and absence

of methylation maintenance is the only pathway for demethylation of 5mC.[18] However, with the identification of oxidized 5mC species (5-hydroxymethyl cytosine, hmC; 5-formyl cytosine, fC and 5-carboxy ctyosine, caC) an active demethylation mechanism in mammalians was proposed. In this active demethylation pathway 5mC is oxidized to hmC by the ten-eleven translocation dioxygenase 1 (TET1).[19–21] Further oxidation of hmC by TET leads to fC and caC, which are used as a substrate for the thymine-DNA-glycosylase (TDG).[19,20,22,23] In this step, TDG creates abasic sites which can be restored to C via base excision repair (see figure 2).[8,24,25] While hmC is the most stable and abundant epigenetic modification of the cytosine derivates known so far, the overall genomic content is variable depending on the cell type. The abundance of fC and caC shows a far lower frequency than 5mC and hmC. Despite their roles in the active demethylation pathway, the involvement of the oxidized variants in other biochemical processes is yet unclear [24,26–31].
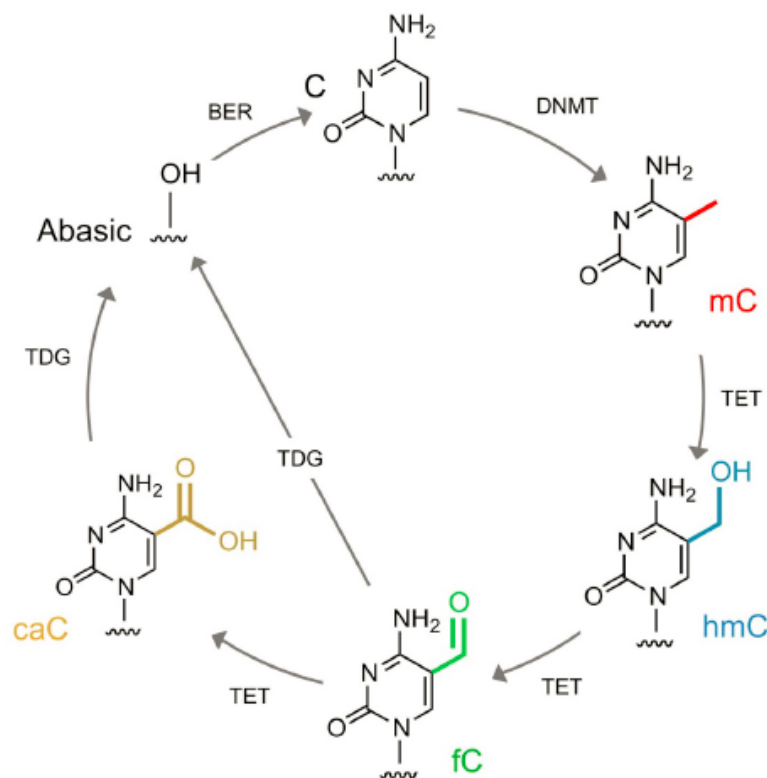
Figure 2: Methylation of cytosine by DMT and active demethylation of 5mC via the TET enzyme to higher oxidized modifications. fC and caC are catalyzed to their abasic from by TDG and are further restored to cytosine by base excision repair (BER) [32]

1.2 Adenine Methylation in Prokaryotic Genomes

Based on its widespread distribution the most common and best studied DNA modification in eukaryotes is 5mC. It exerts a dominant function by establishing the epigenetic setting and by playing a putative role in DNA-based physiological disorders like cancer. However, 5mC is mostly absent in prokaryotic genomes. By contrast, N^6-methyl adenine (6mA) is the most dominant epigenetic mark in bacteria. 6mA entails many functions in the prokaryotic cell and methylation of A in *E. coli* occurs via the DNA adenine methyltransferase (Dam). Embedded in the sequence context of GATC, Dam transfers a methyl group to the adenine amino group using S-adenosyl methionine (SAM) as a methyl donor. The leftover S-adenosyl-homocysteine (SAH) is regenerated to SAM in an ATP-dependent reaction (see figure 3).[33] The Dam enzyme is the first methyltransferase (MTase) identified in γ-proteobacteria. It is classified as a SAM-dependent α-MTase. Based on a 10 amino acids domain, Dam shares significant homology with DpnIIA and MboIA methyltransferases. The Dam MTase is widespread in enteric bacteria and can be found e.g. in *Salmonella spp.*, *Yersinia spp*. and *Vibrio cholerae*.[33] Commonly, methylation occurs after DNA replication. The replication fork passes the methylation site trailed by Dam which processes hemi- and unmethylated GATC with similar efficiency.[34] Since the Dam enzyme can be a *de novo* or a *maintenance* MTase, 6mA acts as a transient or stable epigenetic mark depending on the event.[34,35]
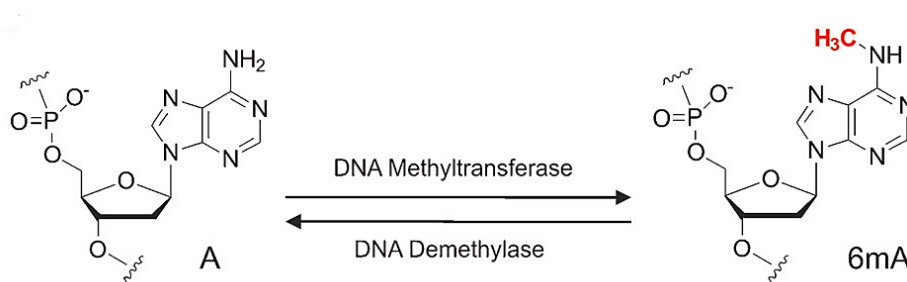
Figure 3: The DNA nucleobase adenosine (A) is modified to its epigenetic version of N^6-methyl-adenosine (6mA) by the transfer of a methyl group. This transfer is performed by methyltransferases and can be reversible. Image from [36]

1.2.1 Adenine Methylation during DNA Replication

6mA is involved in sister-strand synthesis in *E. coli*. Only in the presence of methylated adenine at the origin of replication (oriC) the initiator protein DnaA binds and separates the two DNA strands.[37] During DNA replication, GATC sites of oriC are not immediately fully methylated but remain hemi-methylated for a major part of the cell cycle. Consequently, the newly synthesized DNA does not carry 6mA as a signal and *de novo* methylation is hindered by the protein named SeqA, which negatively regulates initiation of DNA replication. This hemi-methylated state of the oriC prevents the start of a new replication cycle.[35,38] Additionally, the discrimination of non-methylated daughter strands enables *E. coli* to employ the methylated parental strand as a template for mismatched DNA bases guided by recognition of the MutHLS protein complex.[39]

1.2.2 Adenine Methylation for Genome Defense

6mA functions as signal for genome defense embedded in the restriction-modification-system (RM). The RM-system is believed to be evolved protecting bacteria against bacteriophages and can be found in many prokaryotic and archaeal life forms. It distinguishes between host DNA and foreign invader DNA.[40] The RM-system encodes pairs of intracellular enzyme activities: An endonuclease and its opposing partner DNA methyltransferase.[41] The endonuclease cleaves invading viral DNA at short recognition sites which are not modified by methylation. Meanwhile, the host DNA is protected by 6mA. The unmodified invading DNA is rarely methylated fast enough to be protected from restriction and thus the RM-system provides a simple form of genome protection along with self and non-self-discrimination (see figure 4).[33]
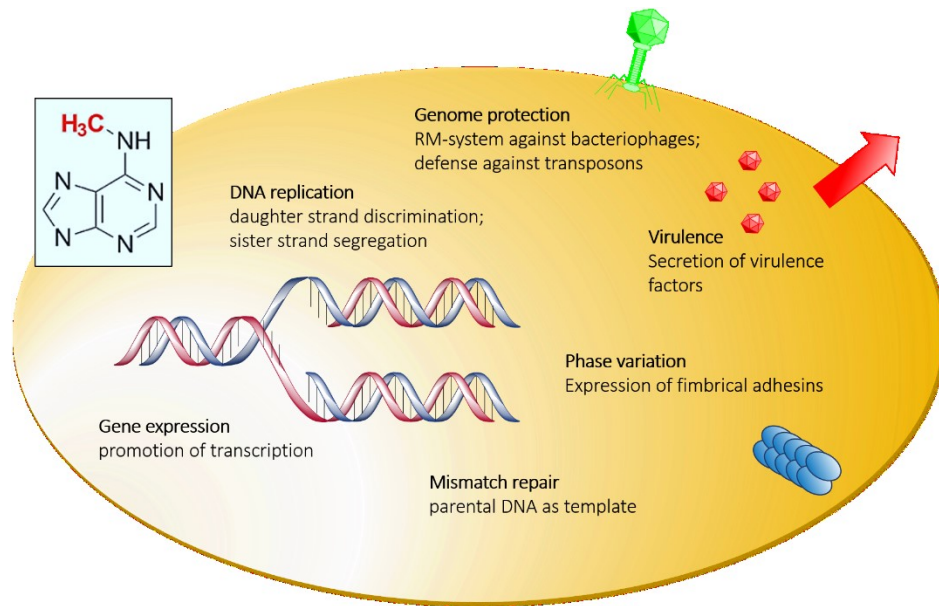
Figure 4: Functions of adenine methylation in prokaryotic genomes. 6mA plays a role in the protection of the genome against viruses. Also, it is involved in DNA replication, sister strand synthesis and the repair of mismatched DNA. The establishment of pathogenic virulence and adhesins expression via phase variation are also 6mA mediated. Image is adapted from [33].

1.2.3 Adenine Methylation in the Regulation of Virulence Factors

Another task of 6mA is the regulation of virulence factors. Uropathogenic *E. coli* harbor a pyelonephritis-associated pili (pap)-operon, which encodes fimbrical adhesins that allow the pathogen to adhere to the urinary tract epithelium. Normally, the pap operon is switched off, based on distinctive methylation pattern in the pap regulatory region.[42] This regulatory region contains two GATC sites (proximal and distal). In the off state, the Leucin responsive protein (Lrp) is bound at three downstream sites blocking RNA polymerase transcription. This binding of Lrp additionally reduces the affinity of Lrp-upstream binding, creating a feedback loop to keep the operon state switched off.[43] While the Lrp protein binds to the downstream site, methylation of the proximal GATC is prevented. The distal GATC is free and becomes methylated. This off state is kept, until the PapI protein is present.[42] Above a critical threshold, the switching factor PapI stimulates the Lrp translocation from downstream to upstream binding site. Distal GATC methylation is now prevented, while proximal GATC is methylated promoting the on state of the pap operon. A positive feedback loop sustains the on state and allows expression of fimbrical adhesins. Thus, the pathogen is able to invade the urinary bladder and cause urinary tract infection (see figure 5).[33,42,44,45]
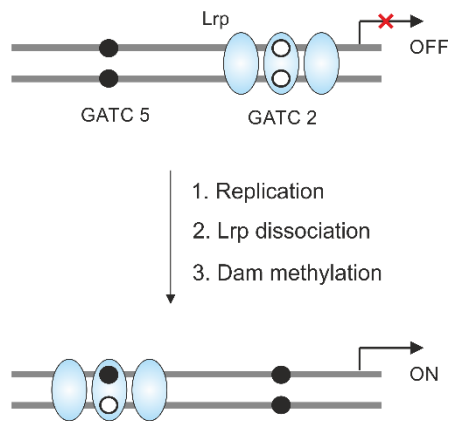
Figure 5: Schematic display of the Pap operon. This regulatory domain comprises two GATC motifs which can exhibit a differential methylation status. This pattern allows the expression of fimbrical adhesins to promote the invasion and adhesion to the urinary mucosa and finally results in tract infection.

1.3 Adenine Methylation in Eukaryotic Genomes

5mC and 6mA were discovered almost simultaneously a few decades ago. However, at that time research of 6mA was mainly focused on prokaryotic genomes. Due to the low abundance of 6mA in higher organisms and constraints in technological detection its potential epigenetic function has remained elusive until recently. With the use of highly sensitive modern mass spectrometry it became possible to detect trace amounts present in the cells, usually in the range of parts per million. This low occurrence, compared to 5mC, raised further doubts about the functional significance of 6mA in multicellular organisms and vertebrate tissue.[46–48] Nevertheless, a few early reports existed which accounted the presence of 6mA in eukaryotic DNA. Unicellular eukaryotes like ciliates and green algae were reported to contain both 6mA and 5mC in their genomes.[49] However, the biological relevance of 6mA regarding involvement in epigenetic processes in higher organisms had remained vague.

In May 2015 studies have shown the abundance of 6mA in three phylogenetically distinct eukaryotes *Chlamydomonas reinhardtii*[50], *Drosophila melanogaster*[51] and *Caenorhabditis elegans*[52]. First indications of potential epigenetic functions of 6mA in eukaryotic genomes were described along with new detection methods like UHPLC-MRM-MS/MS (ultra-high-performance liquid chromatography-triple quadrupole mass spectrometry), 6mA targeted MeDIP-sequencing and other restriction assisted sequencing of DNA samples containing 6mA.[53] Parallel studies on claw frog (*Xenopus laevis*)[54] and zebrafish (*Danio rerio*)[55] attracted further attention to eukaryotic 6mA and spurred speculations how widespread this modification could be in vertebrates and mammalian genomes.[36,56,57] Soon the occurrence of methylated adenine in mouse embryonic stem cells (mESC) was published[58] and has been questioned[59]. Interestingly, the abundance of 6mA seemed to vary: Most tested organisms and tissues exhibited minute levels of 6mA (e.g. 0.00009 % 6mA/A in *Xenopus*), except for the green algae, which displayed the presence of 6mA in 84 % of all genes. In case of the fruit fly the levels of 6mA are dependent on the developmental stage of the organism. In the embryogenesis of *Drosophila*, the abundance of 6mA shows a 70-fold increase in initial stages compared to the adult stage (from 0.001 % to 0.07 % A/6mA). Comparable results were obtained in zebrafish and mice. The content of methylated adenine is doubled in the zebrafish zygote and increased further to a maximum of 0.1 % A/6mA in the 32/64 cell stadium. With the progressing development of the zebrafish, the abundance of 6mA dropped indicating a

role in the functional regulation of embryogenesis. While methylated adenine is evenly distributed in *C. elegans*, it is highly enriched in transcription starting sites in *C. reinhardtii* and in transposable elements of *D. melanogaster* implying not only a single role for eukaryotic 6mA but also a species-specific function.

Moreover, the studies revealed the existence of several 6mA-introducing and removing enzymes. In *D. melanogaster*, *C. elegans* and mESC 6mA-demethylases (and methyltransferase DAMT1 in the case of *C. elegans*) could be identified. In the fruit fly, the demethylase DMAD (diogxgenase *Drosphila* DNA 6mA demethylase) showed similar structural domains as the mammalian TET-enzyme. It was shown that this demethylase is specific for 6mA and does not demethylate 5mC, hmC or 6mA in RNA. Alterations in the expression of the demethylase lead to increased lethality of the *Drosophila* embryo. Comparable to these findings, deficiency of the identified mouse Alkbh1 demethylase caused reduction in litter size and resulted finally in embryonic lethality indicating a regulation mechanism for 6mA.

Taken together these recent findings imply functional epigenetic roles of 6mA. In RNA, the profound effects of adenine methylation are extensively studied. In mammalian non-coding RNA and mRNA, 6mA can be catalyzed by demethylases generating intermediates like N^6-hydroxymethyl adenine (6hmA) and N^6-fomyl adenine (6fA). Since these recently discovered 6mA variants are stable modifications in RNA, they might possess regulatory functions.[60] Furthermore, RNA carrying an adenine methyl group can be destabilized regarding Watson-Crick base-pairing to induce a switching mechanism that alters the RNA structure. Thus, binding of proteins to RNA can be influenced by the presence as well as the positioning of 6mA at the protein binding site. The same feature could be assigned to adenine N^6-methyl groups present in DNA. It could affect gene transcription and inhibit or facilitate the binding of transcription factors or other DNA binding proteins.

Since *C. elegans* and *D. melanogaster* have little or no 5mC in their genome, it is possible that the altering presence of this modification compared to other higher organisms, is the result of evolutionary events.[56,57] Here, 6mA could play a predominant role in epigenetic processes and the identification of 6mA demethylases and methyltransferases implies that adenine methylation could be present in many more organisms. With the ongoing development and

continuous improvement of sensitive detection techniques, the function and precise distribution of eukaryotic 6mA could be explored throughout all domains of life (see figure 6).
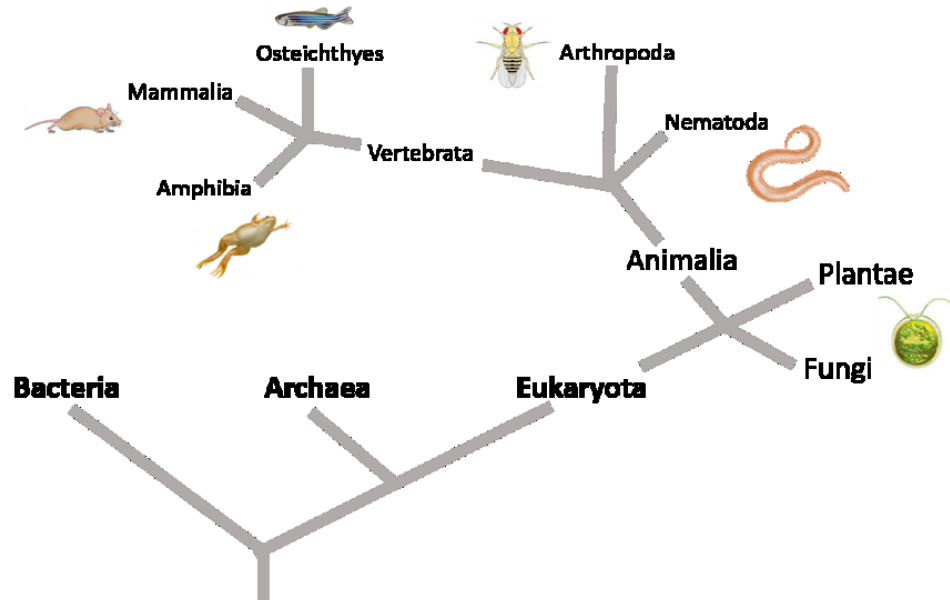
Figure 6: Overview of 6mA which was predominantly present in bacteria and archaea until recent findings. Three studies in phylogenetically distinct organisms reported the presence of 6mA in eukaryotic genomes. Further publications illuminated the abundance of 6mA in higher organisms like amphibia and mammalians. Image was adapted from [36].

1.4 Transcription-Activator-Like Effector Proteins

Transcription-activator-like effectors (TALEs) are DNA binding proteins that are injected into plant cells by numerous phytopathogenic bacteria. After transfer to the host cell via the type III secretion system the TALE protein enters the nucleus and binds to effector specific promotors.[61] Subsequently this will impair the host defense mechanisms or trigger a response beneficial for the bacterium. TAL effectors are key virulence factors for plant pathogens and can reprogram plant cells into mimicking eukaryotic transcription factors.[62] TALEs have been found predominantly in species of *Xanthomonas*, but related homologs exist in *Ralstonia solanacearum* and *Burkholderia rhizoxinica* as well.[63,64]
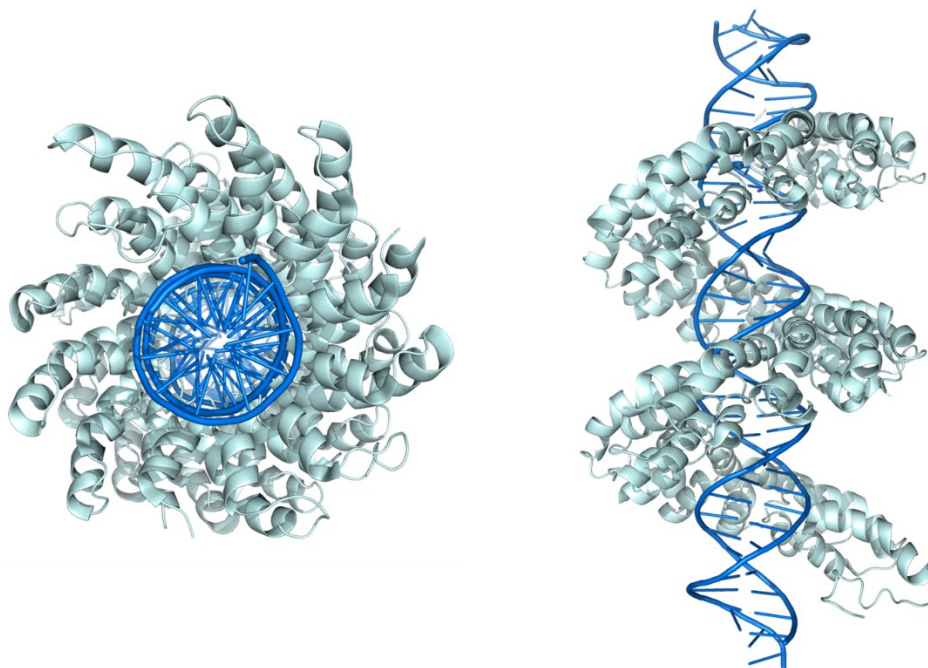
Figure 7: Crystal structure of the TALE protein PthXo1 binding to its target DNA from top view (left) and side view (right). The TALE (light blue) wraps itself along the major groove of the target DNA (blue) via concatenated repeats facilitating recognition and binding of the protein. Adapted from [65] and PDB entry 3UGM.

The central DNA binding domain of TALEs consist of a unique structure which shows no similarity to known DNA binding proteins. In 2009, Boch and Bonas presented the setup of this central domain, along with the study of Moscou and Bogdanove which confirmed their findings by computational analysis.[66,67] Soon after the first reports on TALEs its crystal structure (see figure 7) was elucidated.[65] The high modularity enabled the development of

toolkits for customized TALE assembly. These are often based on Golden Gate cloning [68,69,65,71], but also methods combining solid phase synthesis and standard cloning like fast ligation based automatable solid-phase high throughput (FLASH) exist. A detailed protocol on how the TALEs proteins are cloned in a Golden Gate (GG) approach is presented in section 6.2.

1.4.1 TALE Scaffold and Binding Mode

TALE proteins are composed of three major parts: The N-terminal region (NTR), the central DNA binding domain and the C-terminal region (CTR). In Nature, the N-terminus of the protein contains the type III secretion signal which provides access to the host plant. The C-terminal region bears the nucleus localization sequence (NLS) as well as an activator domain. While the N-terminus binds the 5'-end of the target DNA, the C-terminal region faces the 3'-end. In between the N-and C-terminus the central repeat domain (CRD) is located. The CRD consists of tandem, concatenated repeats which each comprise 33 to 35 amino acids. These repeats are nearly identical and highly conserved except for the amino acid position 12 and 13. These positions are called repeat variable di-residues (RVD) and facilitate the TAL effector binding to the DNA sequence. TALEs recognize their target DNA via a simple one-to-one code (see figure 8): In nature, several RVDs are observed, however the most naturally occurring RVDs are HD, NG, NN and NI which respectively specify the DNA nucleobases C, T, G/A and A. Other less frequent RVDs like NS and N* exhibit lose specificity (the asterisk incorporates a missing amino acid at position 13 forming a 33 amino acid repeat).[72,73] RVDs NH and NK are less common and show increased specificity for guanine (G). The central repeat domain is terminated with a truncated repeat containing the first 20 amino acids and is commonly referred to as half or last repeat (LR).[66]

Each TALE repeat consists of two α-helices with the RVD residing in an interhelical loop domain. The residue at position 12 in each RVD is facing away from the DNA interacting with the proteins' backbone carbonyl at position 8 thus stabilizing the interhelical loop. The residue at position 13 faces the DNA and looms deep into the major grove facilitating sequence-specific contact with a single DNA nucleobase (see figure 8). It has been observed that most contacts between the RVD and DNA bases are resulting from direct hydrogen bonds (e.g. HD in contact with cytosine between the aspartate (D) carboxyl-group and 4-amino group of C),

highly steric packing (e.g. alpha carbon of glycine (G) of NG to extra cyclic methyl carbon of thymine base) or reduced specificity via steric exclusion of alternative DNA nucleobases (NI RVD).[74] Moreover, studies revealed that HD and NN repeats procure a significant contribution to overall TALE binding affinity, while NI and N* repeats appear to result in reduced activity and lower specificity.[73] Depending on the target sequence context selectivity effects have been reported.[75,76]
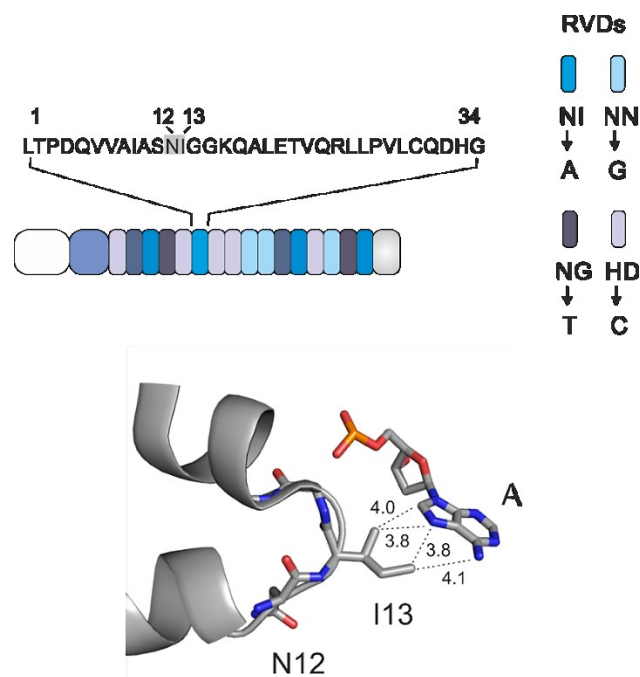
Figure 8: Framework of TAL effector proteins. Each TALEs consists of multiple concatenated, conserved repeats which comprise 34 amino acids (left). Position 12 and 13 of each repeat are variable residues (RVD) and form a loop which is embedded into two α-helices (below). The RVDs of the TALE protein facilitate the contact and binding to the target DNA and follow a simple cipher in which the RVDS NI, NN, NG and HD specifically recognize A, G/A, T and C nucleobases (right). Adapted from [77].

Besides the CRD, the N-terminus of TALE proteins also plays a key role when it comes to DNA binding. Crystallographic resolution of the N-terminus illuminated that this region is crucial for full activity of TALE proteins.[78] Although the underlying mechanism is not yet well understood it was shown that the NTR comprises 4 continuous "cryptic" repeats analogous to the canonical repeats of the CRD. These additional N-terminal repeats follow the TALE taxonomy and are named repeat 0, -1, -2 and -3. The cryptic repeats provide the main part of binding energy. They exhibit positively charged patches evoked by arginine and lysine residues.[78,79] These conserved residues contribute to the overall basic charge of the TAL

protein implying promoted nucleic acid binding by the NTR, however, this binding occurs to be non-specific and sequence independent of the target DNA .[78,80,81]

Despite deciphering the TALE code and insights of the binding affinity of TALE proteins to DNA, little is known how the proteins search and recognize their target. Studies of other DNA binding proteins observed either a mechanical salt-dependent "hopping mechanism" in which the protein is involved in cycles of binding and unbinding to the DNA or a "sliding mechanism" whereas the protein slides along the DNA helix. TALE proteins have been described to be a hybrid of both.[82] A model suggested that TALE proteins take a conformation wrapping itself loosely along the DNA helix during nonspecific search. Upon encountering a target site, TALEs compress and "verify" the local sequence thus resulting presumably in stable binding.

1.4.2 DNA Binding Proteins and TALE Applications

In the past decade, several approaches have emerged which offer targeting and manipulation of any gene of interest from a wide range of different cell types and organisms. The discovery of programmable proteins like TALES, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR-CAS) and Zinc Fingers (ZF) equipped scientists with an unmatched ability to alter cells and organisms in their genomic setup. In the case of TALEs and ZFs the versatility of is based on the modularity of the DNA binding domain enabling recognition of any DNA sequence. Zinc Fingers consist of 30 amino acids and contain two conserved pairs of cysteines and histidines which are stabilized by one zinc ion. Each finger is built-upon an antiparallel β-sheet recognizing 3 to 4 DNA nucleotides and inserting an α-helix structure into the major groove of the DNA. ZFs were the first programmable DNA binding proteins harnessing genome editing approaches. However, design and selection of each ZF turned out to be laborious and time consuming since neighboring ZFs influenced specificity and enhanced off target effects. The fortuitous discovery of TALEs offered an easier assembly approach. The TALE modules appeared to be mostly independent of neighboring influences and were publicly available, which allowed inexpensive and quick generation of custom TALEs. Only 3 years after the TALE recognition mode was decoded, the CRISPR-CAS platform was developed. While TALEs and Zinc Fingers exhibit highly modular DNA binding domains, the CRISPR-CAS system relies on RNA-DNA recognition. The advantage of the CRISPR-CAS system is based on the easy RNA design and a single constant Cas9 nuclease. For DNA cleavage, the Cas9 is guided by an RNA (gRNA) which will hybridize to its target DNA sequence to facilitate binding of the CRISPR-CAS system. CRISPR-CAS has been shown to be effective in most established model organisms and cell cultures. However, it showed off target effects just like TALEs and Zinc Fingers. In all three protein models, most off-target and specificity problems have been overcome by simple adjustments e.g. truncation of the gRNA. The most exciting aspects of these DNA targeting proteins is how efficiently they have been used and how broad their effective application can be reported in many model organisms from crop plants and livestock to humans. For example, TALEs can be exploited for gene editing. Hereby the protein is fused to a nuclease domain (e.g. TALE with FokI domain) and introduces double strand breaks *in vivo* which are repaired by non-homologous end joining (NHEJ) or homologous recombination (HR). Thus, deletions, insertions or mutations can be established in the genome. Moreover, DNA binding proteins can be harnessed for tracking or imaging studies when connected to a

fluorescent domain like GFP (green fluorescent protein). Likewise, the fusion of the DNA binding protein to a transcriptional regulator facilitates expression or inhibition of the target gene. Besides alterations to the genome, the epigenome can also be manipulated. A study showed that site specific removal of cytosine methylation is possible by cloning TET catalytic domains to TALE proteins.[83] Similarly, TALES linked to histone demethylase protein LSD1 altered histone methylation and the introduction of methylation by fusing TALE proteins with methyltransferases has been reported to be successful.[68–71]

1.4.3 TALEs and Epigenetics

TALE proteins can also be used for targeted transcription regulation and altering the epigenetic setup. A critical property of TALEs is their central DNA binding domain. For any genome targeting approach high sequence selectivity is required. Consequently, to avoid any off target binding effects of TALEs, maximal RVD selectivity is needed and studies on mutational analysis have offered improved RVDs.[87,88] Interestingly, another critical characteristic of TALEs is their sensitivity towards DNA nucleobases which carry epigenetic modifications. For instance, the RVD HD is shown to be sensitive to 5mC. If 5mC is present in the target sequence, TALE binding is abolished.[89–91] 5mC has been identified to be present in CpG islands embedded in many promotor regions. This critical regulatory element is generally chosen for knock out studies of genes with therapeutic and biotechnological interest. This negative impact on TALE binding has led to the development of customized RVDs which exhibit similar binding affinity to C and 5mC.[89] It was suggested that the TALE repeat N* can efficiently accommodate methylated cytosine and thus overcome the methylation sensitivity of TALEs. However, this discriminatory binding affinity can also be exploited to develop TALE repeats binding to the full range of epigenetic modified nucleobases like 5hmC, 5fC and 5caC.[12,92–94] Moreover, binding affinities of TALES to other modified DNA like 6mA or N^4-methyl cytosine (N4mC) bases are yet unknown. In the light of recent studies of 6mA being present in eukaryotic genomes in organisms from all domains of life, it is important to understand and investigate the specific binding capability of TALEs to this epigenetic nucleobase.

2. Aim of Work

The specific detection of epigenetic marks might open the door for the diagnosis and subsequent treatment of numerous diseases which are based on a disorder of the DNA methylation pattern as known for the case of 5mC. In the past decade, lots of effort went into the development of protein-DNA recognition systems. The present work is located at the interface of DNA-protein interaction studies and the influence of epigenetic signals with the specific task to illuminate the binding behavior of TALEs to DNA harboring 6mA modifications. The first step is the interrogation of the RVD NI interacting with DNA representing naturally occurring methylation sites in uropathogenic *E. coli* (Pap operon). For further insights, the canonical NI will be mutated to RVDs facilitating a less polar surface and the resulting interaction with 6mA will be examined. Moreover, polar RVDs will be created which could enable H-bonding to the N^6-amino group or N^7-O-atom of adenine and thus exhibit differential TALE binding. DNA. A second approach is the expansion of the N^6-methyl group with bulkier substituents in the form of SAM analogs, which can be transferred to the GATC motif by the Dam methyltransferase. Here, we interrogate the ability of TALEs to accept large groups facing the major groove of the DNA paving the way for smart TALE design and labeling strategies with bulky fluorophore labels.

3. Results and Discussion

3.1 Selection of TALE Target-DNA

For the interrogation of TALE binding affinity to epigenetically modified DNA a sequence context was chosen which is regulated via the reversible methylation of adenine. Two naturally occurring 17-nucleotide long sequences, namely Pap2 and Pap5 were chosen. Both hold a regulatory GATC site within the *E. coli* pap operon (see figure 9). Methylation of the N^6-position of adenine within the proximal (Pap2) and distal (Pap5) GATC site occurs via the Dam methyltransferase. The methylation determines binding of the leucine-responsive regulatory protein and leads to the expression of pyelonephritis-associated pili (Pap). Ultimately this results in the adhesion of uropathogenic *E. coli* (UPEC) to the mucosa cells as a basic mechanism for many urinary tract infections.[42,44,45,95]

Pap2: 5´-TAAAAG**A**TCGTTTAAAT
       5´-TAAAAG**A**TCGTTTAAA
       5´-TAAAAG**A**TCGTTTAA

Pap5: 5´-TAGACG**A**TCTTTTATGC
       5´-TAGACG**A**TCTTTTATG
       5´-TAGACG**A**TCTTTTA

      nt: 1   5   10   15
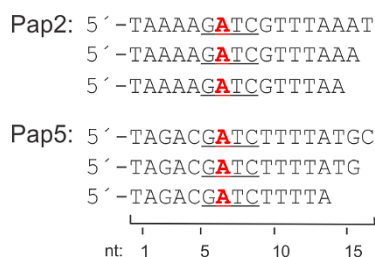
Figure 9: Overview of selected target DNA sequence for TALE targeting. The GATC sites are underlined with A or 6mA of interest in red. The sequences are 17 nucleotides (nt) long and the nucleobase of interest is located at position 7 of the oligos. Additional truncated oligos were designed for a later investigation sequence length dependency of TALE proteins. Adapted from [77].

3.2 TALE Assembly and Expression

The TALE proteins used in this work were generated based on the *Xanthomonas axonopodis* scaffold. A protocol for the respective golden gate assembly was established by the group of Daniel Voytas.[96,97] Hereby, the constructed TALE proteins contain a truncated N-terminus (+136 amino acids from the first canonical repeat) which harbors a thioredoxin (TRX) or green fluorescent protein (GFP)-tag. In contrast to natural TALEs, the proteins assembled in the lab lack the translocation signal, the nuclear localization signal (NLS) and the activation domain

in the C-terminus. The TALE repeat arrays were assembled according to the sequences of Pap2 and Pap5. After the first and second round of the Golden Gate reaction the *E. coli* BL21 cells were transformed with the generated plasmids for selection. The protein was isolated and purified via C-terminal His6 affinity tag by Ni-NTA chromatography. The purity and correct protein size was verified by SDS-PAGE analysis (see figure 10). For detailed description of TALE assembly and expression see methods 6.2.2 and 6.2.3.
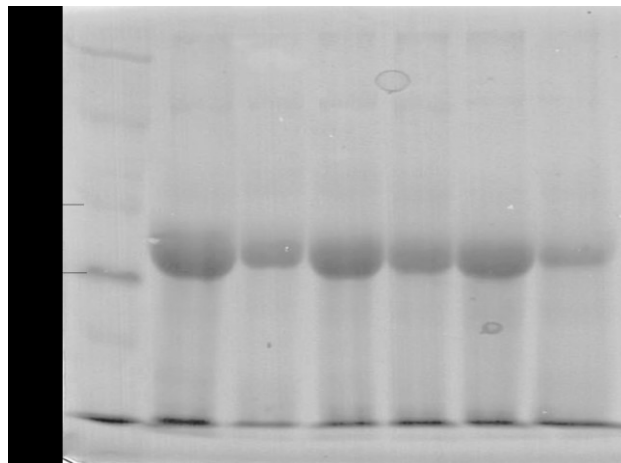
Figure 10: Exemplary image of 8 % SDS resolving gel containing expressed and purified TALE proteins with thioredoxin tag. From left to right: TALE_Pap2_NI, TALE_Pap2_NL and TALE_Pap2_NV (first and second elution). Constructed TALEs contain 17 repeats (including T_0) and carry a mass of approx. 86.32 kDa. As a marker, the protein broad range maker from NEB was used (2-212 kDa, P7702S). Image is adapted from [77].

3.3 TALE-Binding to N^6-Methyl Adenine

3.3.1 Interrogation of the natural RVD NI binding to A and 6mA

NI is the most frequently found natural occurring RVD in TALEs for binding of the nucleobase A.[66,67,87,98] However, it is unknown how the NI TALE repeat interacts with an N^6-methylated adenine. The steric changes introduced by the N^6-methyl group might weaken or abolish the target affinity. It is well known that isoleucine at position 13 in the RVD NI (I13) generates a hydrophobic surface at the Hoogsteen face of A. Crystal structures report atomic distances between 3.8 and 4.1 angstrom between A and I13 (see figure 11).[65,99]
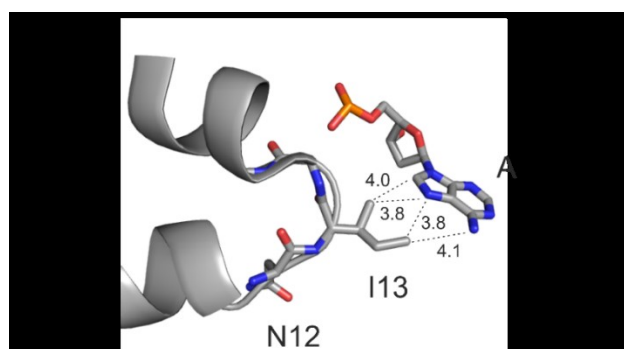
Figure 11: Crystal Structure of the RVD NI to the nucleobase adenine (A). Dotted lines indicated the distances from position 12 and 13 in angstrom. Adapted from [77].

Furthermore, a hydrophobic interaction between the N^6-amino hydrogen of A and the O^4 of a paired thymine nucleobase might move the N^6-methyl group towards the side chain of I13. Therefore, an interaction of the TALE RVD NI with 6mA seems possible. To test this hypothesis, expressed TALEs (Pap2_NI) were employed in electromobility shift assays (EMSA) using a GFP fluorophore tag as a read out. TALEs targeting the Pap2 sequence were incubated together with unlabeled DNA oligonucleotide duplexes comprising either A or 6mA at position 7 opposite the TALEs RVD NI. The result displayed the formation of defined TALE-DNA complexes with increased electromobility in comparison to sole TALE protein. Interestingly the results do not show differences in TALE binding to DNA containing either A or 6mA (see figure 12.

a
```
MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQN
PGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGERQHMDSPDLGTVD
LRTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNAL
TGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQR
LLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALE
TVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGK
QALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASH
DGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVA
IASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPD
QVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLC
QDHGLTPDQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLLEHHHHHH*
```

b

DNA/TALE: 2.5   1.25   0.25

TALE_Pap2_(NI): + + + + + + +
Pap2:   + - + - + - -
Pap2_6mA: - + - + - + -

c

= A   = 6mA

TALE_Pap2_(NI): +   +
Pap2:   +   -
Pap2_6mA: -   +

Figure 12: Interaction of TALE RVD NI with A and 6mA-containing DNA oligos carried out in EMSA assay. **A**: Amino acid sequence of expressed TALE_Pap2_NI targeting the Pap2 DNA sequence with the RVD NI in bold black. **B**: EMSA gel with TALE Pap2_NI in complex with various concentrations of DNA duplexes, free TALE protein in the right well shows lower electromobility than TALE-DNA complexes. **C**: Quantification of EMSA assay with a TALE/DNA ratio of 6.5:1 (n=3). Image was adapted from [77].

It could be possible, that no discrimination of 6mA was observed due to a low nucleobase selectivity for the EMSA assay. In order to investigate the aforementioned results in further detail, a primer extension assay was carried out. The primer extension assay provides higher selectivity and sensitivity and therefore might unveil subtle differences in TALE binding to A or 6mA which are likely not to be seen in an EMSA assay (see figure 13 and methods 6.2.5).

Figure 13: Principle of TALE controlled primer extension assay. The TALE is shown as carton with the RVD of interest (here NI) in red which is positioned opposite the nucleobase A or 6mA (red, imbedded in GATC site). Image was adapted from [77]

If the TALE protein is binding to its target sequence, the DNA polymerase is blocked and elongation of the primer template is inhibited. The 5'-32P labelled primer is hybridized to a template oligo containing the Pap2 target sequence and further incubated with

TALE_Pap2_NI protein. Eventually, dNTPs and Klenow Fragment were added to the mixture and further incubated. The reaction was resolved and monitored by denaturing polyacrylamide gel electrophoresis (PAGE) with 9 M urea. The extension product of the PAGE gel was quantified and employed as a measure of TALE binding (see figure 14).

Figure 14: Primer extension assay with TALE_Pap2_NI binding to DNA oligo complexes. **B**: Image of urea PAGE gel with reactions containing 25 mU Klenow Fragment (exo-), 41.5 µM dNTP, 8.325 primer-template complex and various concentrations of TALE_Pap2_NI (541 nM, 333 nM, 166.5 nM and 8.32 nM) resulting in the displayed TALE/DNA ratios. **B**: Quantification of urea PAGE gel as a measure of TALE binding to A and 6mA (n=3). The image was adapted from [77].

Despite greater sensitivity the primer extension assay also showed no differences in TALE_Pap2 binding. The same result was observed over varying TALE concentrations. Similarly, targeting the Pap5 sequence context yielded the same results. In order to exclude masking of potential selectivity caused by excessive binding energy, the experiments were repeated with truncated TALE_Pap2_NI variants. To achieve this, the truncated TALE versions were targeting only 16 or 15 nt Pap2 DNA duplexes. Again, no differences could be detected in the primer extension assay (see figure 15). A possible explanation is the probable accommodation and tolerance of 6mA by the RVD NI. Both experiments suggest additionally, that the N^6-methyl group is barely involved in hydrophobic interactions with TALE.

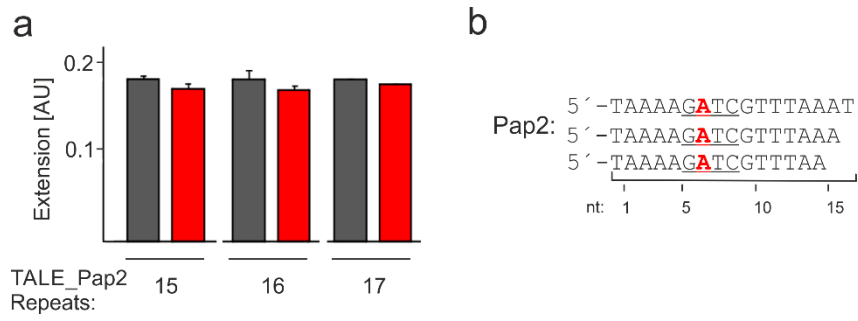**Figure 15**: Primer extension assay with truncated TALE_Pap2_NI binding to DNA oligo complexes with 17, 16 and 15 nt. **A**: Quantification of the extension product as a measure of TALE binding in the primer extension assay (n=3). **B**: Pap2 sequence context with truncated versions used in this experiment. The nucleobase A or 6mA (red) is embedded in the GATC sequence. Adapted from [77].

3.3.2 Interrogation of Non-Polar RVDs Binding to A and 6mA

Previous structural studies reported that adenine can be harbored by RVDs which carry a mutation in the amino acids at position 13. In the RVD NI the hydrophobic side chain directly faces the A nucleobase without favorable contact and may not exhibit a major contribution to the overall TALE affinity to DNA. Thus, NI was reported to be a weak RVD [73,99]. However, it also indicates that isoleucine (see results 3.3.1) and another wide range of amino acids at position 13 could tolerate adenine and possibly N^6-methyl adenine when there is no steric clash. Mutants with hydrophobic amino acids at position 13 seem to accommodate the adenine nucleobase but this remains still elusive.[99] For further modulation of the hydrophobic surface of the RVD NI, the position 13 was replaced by a shorter valine (V) and a more flexible leucine (L). The generated RVDs NV and NL can provide deeper insights into the interaction of TALE proteins with 6mA. For testing the mutated RVDs, TALEs were generated that targets again the Pap2 sequence context. The RVDs were positioned opposite A or 6mA and primer extension experiments were carried out (TALE_Pap2_NL and TALE_Pap2_NV). In both cases, the TALE proteins exhibited similar binding to the DNA target compared to the wildtype RVD NI (see figure 16).
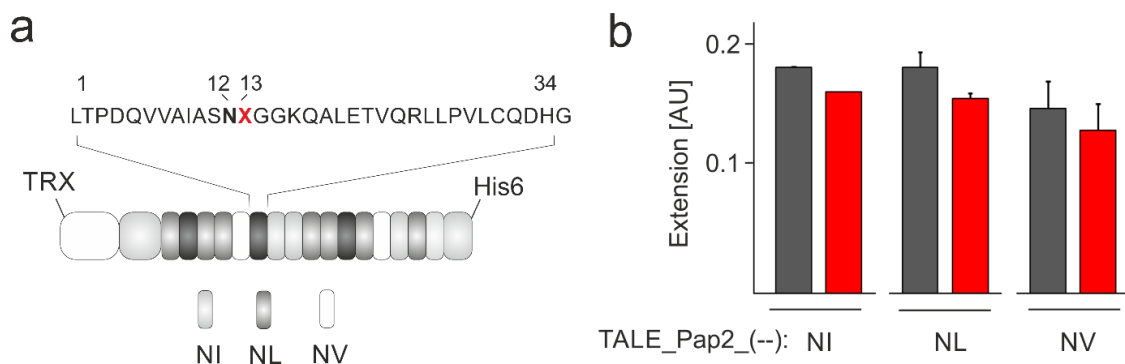
Figure 16: Modulation of the hydrophobic surface of RVD NI. **A**: Principle of the TALE scaffold carrying the RVD NL, NV or NI at position 7, which is placed opposite A or 6mA. The mutations represent hydrophobic shorter (V) or more flexible (L) sidechains facing the nucleobase A or 6mA. **B**: Quantification of the primer extension assay as a measure of TALE binding carrying the RVDs NL and NV compared to wildtype NI (n=3). Adapted from [77].

This observation was also consistent for the Pap5 context (see figure 17). These findings imply that non-polar hydrophobic RVDs NL and NV accommodate A and 6mA without specific coordination and affinity. Furthermore, the N^6-methyl group is tolerated and does not lead to a steric clash. This pronounced flexibility prevents hydrophobic interactions and is consistent with reported crystal structures [99].
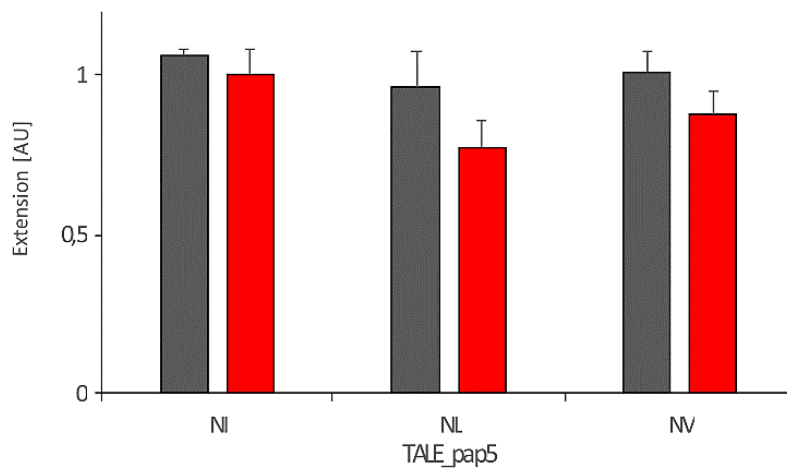
Figure 17: Quantification of the primer extension assay as a measure of TALE binding carrying the RVDs NL and NV compared to wildtype NI. Modulation of the hydrophobic surface of RVD NI. The mutations represent hydrophobic shorter (V) or more flexible (L) sidechains facing the nucleobase A or 6mA. Here the Pap5 sequence context was tested with TALE_Pap5 comprising either NI, NL or NV at repeat position 7. Adapted from [77]

To verify these data, models of TALE_Pap2_NI, TALE_Pap2_NV and TALE_Pap2_NL were conducted based on available crystal structures. All models were generated with DNA containing either A or 6mA at position 7 of the Pap2 operon. The model for the wildtype RVD NI displays the rotation of the terminal methyl group in the side chain of I13 to accommodate 6mA. Similarly, a conformational change can be observed for L13. Both amino acid side chains prevent unfavorable interaction with the N^6-methyl group via slight conformational changes. For the shorter side chain of V13 compared to I13 and L13, a larger distance to the nucleobase is revealed. Here, no rotation is observed for the accommodation of 6mA. All three models suggest that the N^6-methyl group can be harbored in the RVD variants. The larger sidechains of NI and NL adapt and slightly rotate to conformations avoiding steric clash. The amino acids do not undergo hydrophobic interactions that lead to a measurable change in selectivity (see figure 18).
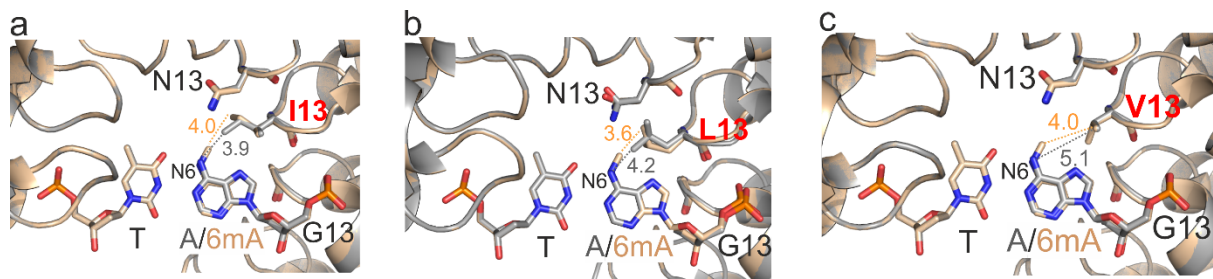
Figure 18: Superimposed models of TALE_Pap2 with mutated RVD containing I13, L13 or V13 (red). The TALE is shown as cartoon and binds to DNA carrying A or 6mA opposite the relevant RVD. Model containing A is displayed in gray, 6mA in light brown. **A**: The RVD wit Ile at position 13 interaction with A and 6mA. **B**: RVD NL interacting with A and 6mA. **C**: Model of RVD NV opposite of A or 6mA. Models were created in collaboration with Julia Jaspers and Oliver Koch, TU Dortmund.

3.3.3 Interrogation of Polar RVDs Binding to A and 6mA

Previous studies have described that hydrophobic RVD mutants do not clash with adenine. This finding is consistent with the results of the present work. Models of RVD variants containing non-polar amino acids like proline, methionine and cysteine can accommodate the nucleobase adenine and therefore do not give rise to any steric constraints. Similarly, it has been reported that polar amino acids which do not form an H-bond to the opposite adenine adapt a conformation to avoid a steric clash.[99] The naturally occurring RVD NN with polar asparagine (N13) is known to bind both, adenine and guanine. Crystal structures have unveiled that the side chain amide group of N13 forms an H-bond with the N^7-atom of adenine.[99] Yet, the binding affinity of NN with its polar surface to 6mA is unknown. Therefore, primer extensions with the RVD NN embedded in TALE_Pap2 were performed. The results exhibit similar binding to A and 6mA. The N^6-methyl group shows little influence of the overall TALE binding affinity (see figure 19). Modelling of TALE_Pap2 with RVD NN indicate, that the N13 slightly rotates accommodating the methyl group of 6mA which results in a loss of the H-bond to the N^7-amino group.

Figure 19: Interaction of TALE_Pap2 containing the polar RVD NN wit A and 6mA. **A**: Quantification of primer extension. The TALE/DNA ratio was 60:1 (n=3). **B**: Superimposed modelling of TALE_Pap2_NN interacting with A (gray) or 6mA (light brown). Red dotted line indicates H-bonding. Models were created by Julia Jaspers and Oliver Koch, TU Dortmund.

The next experiment was to examine the polar RVD NE. This RVD contains glutamic acid (E) at position 13 of the repeat. Reported *in vivo* activity assays with TALE nucleases utilizing NE

showed no binding to the nucleobase A.[87,100] However, the models indicate that the side chain of E13 can reach further towards the major groove of the DNA and therefore direct H-bonding with the N^6-methyl group can be observed in the crystal structure.[99] These observations suggest that the binding of the RVD NE to 6mA could experience more influence by the N^6-methyl group. Surprisingly, the performed primer extensions did not indicate differences in binding of TALE_Pap2_NE to A or 6mA. Modelling of this interaction supports the hypothesis that the side chain of E13 reorients itself to accommodate the nucleobase. This results in a loss of the hydrogen bond to the adenine N^6-methyl group. Furthermore, NE forms a hydrogen bond to its own backbone amide. This conformational change causes the precedent RVD to rearrange in order to avoid an unfavorable steric clash (see figure 20)
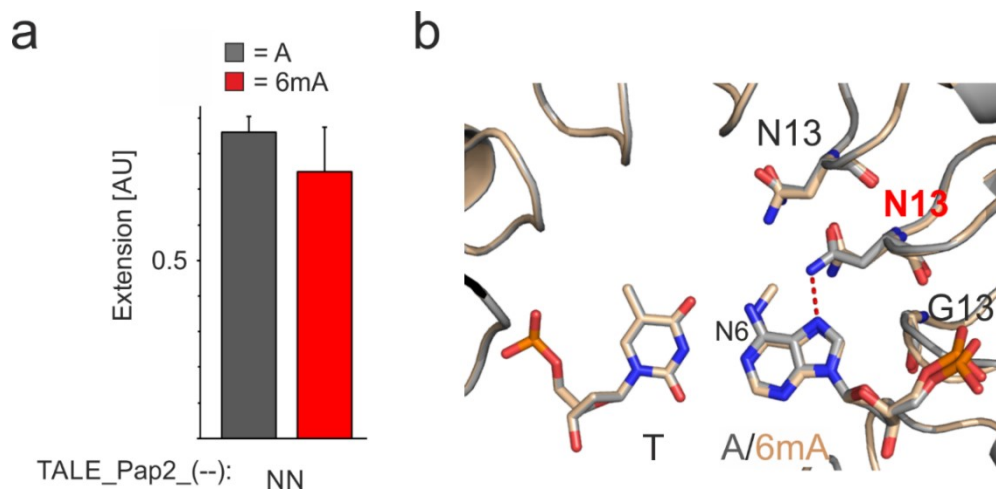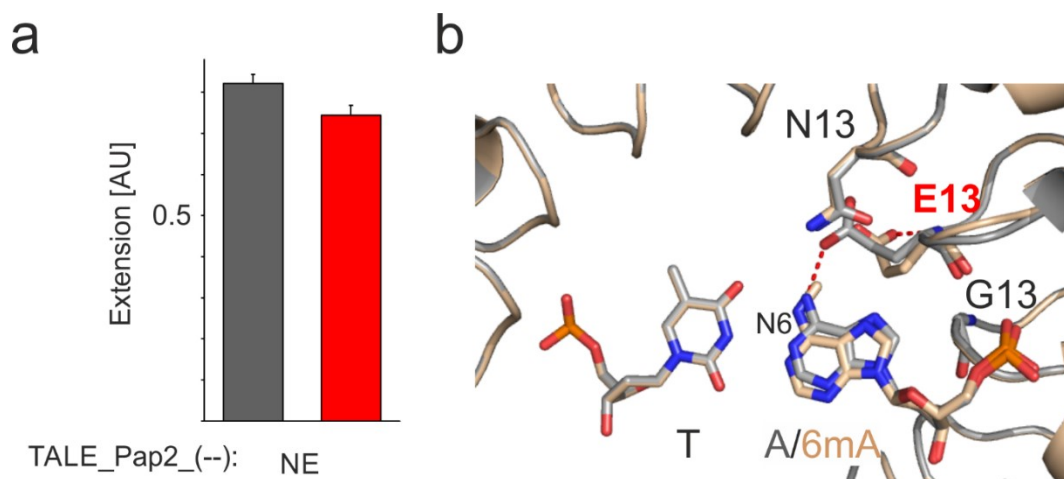

Figure 20: Interaction of TALE_Pap2 containing the polar RVD NE wit A and 6mA. **A**: Quantification of primer extension. The TALE/DNA ratio was 60:1. **B**: Superimposed modelling of TALE_Pap2_NN interacting with A (gray) or 6mA (light brown). Models were created by Julia Jaspers and Oliver Koch, TU Dortmund.

Taken together, the findings of non-polar RVDs (NI, NL and NV) as well as polar RVDs (NN, NE) do not exhibit differential binding to A or 6mA with the respective TALE protein. This indicates that the TALE repeat framework is not selective towards A and 6mA. The methylation insensitivity allows robust binding of the DNA by the TALE protein and enables user-defined binding to DNA loci which potentially contain epigenetically modified adenine. This makes it a useful tool for *in vivo* applications, where actual epigenetic modifications could occur depending on the target sequence. Consequently, testing TALEs with varied RVDs placed opposite other DNA nucleobases like thymine, cytidine and guanine did not result in

differential binding. These data indicate that TALE repeats have a high intrinsic motility to undergo conformational changes while accommodating A and 6mA while remaining a robust DNA binder.

3.4 Interrogation of TALEs to Bulky N^6-Alkynyl Substituents

The tolerance of TALEs harboring methyl groups at the N^6-position of adenine raised the question if the protein can accommodate even larger substituents at this position. The recognition and binding mode of TALE proteins is directed towards the DNA major groove. This feature makes TALE proteins a useful capture tool for analytical approaches. For example, TALEs sensitive to methylation of cytosine can be used as selective fishing tools for DNA containing epigenetic bases like 5mC and 5hmC, thus delivering information about the presence and status of modified cytosine embedded in DNA samples based on TALE binding or abolished binding.[12,32,92,93,101,102] All these applications would profit of fluorescently labeled DNA for detection. However, the optimal site to attach DNA labels is the 5-position of cytosine and thymine directed towards the DNA major groove. This site is recognized by methylation sensitive TALEs and could abolish binding. Thus, the N^6-position could be an ideal site to introduce DNA labels which are tolerated during TALE binding. This blind spot could enable effective binding of TALE proteins to methylated and fluorophore-labeled DNA. Therefore, it was intriguing to investigate the introduction of larger substituents at this position.

For the introduction of a label at the adenine position, S-adenosyl methionine (SAM) analogs were used. In nature, SAM is a common cosubstrate for enzymes which are involved in the transfer of methyl groups. In this process SAM is hydrolyzed to S-adenosyl homocysteine (SAH) and adenine. In bacteria, the Dam methyltransferase utilizes SAM to transfer a CH_3-group to the N^6-position of adenine. For the approach of interrogating larger substituents, the Dam enzyme represents a practical tool to introduce labels in the form of SAM analogs at adenine. The SAM analogs contain alkynyl-substituents which replace the electrophilic methyl group at the sulfonium center and therefore enlarge the moiety of the replaced methyl group. The SAM analogs were synthesized and then tested for acceptance by the Dam enzyme. In this assay, SAM-free Dam enzyme is needed to ensure a high yield of transferred groups and to avoid preference of the common SAM cofactor. The obtained results indicate that the synthesized SAM analogs are accepted by the Dam enzyme and that the substituents were successfully transferred to the A position within the GATC motif. The synthesis of the SAM variants, as well as the testing of the Dam tolerance was conducted in collaboration with the group of Elmar Weinhold, RWTH Aachen (see figure 21).
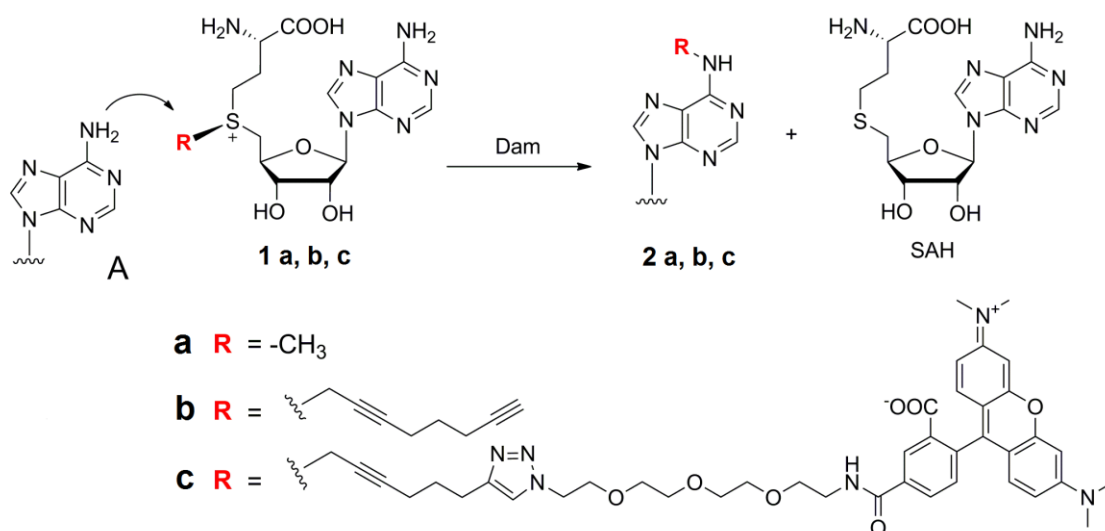
Figure 21: Modification of adenine at the N^6-position embedded in the GATC motif of the pap operon (Pap2). The *E. coli* Dam enzyme transfers a methyl group from the SAM substrate to adenine (2a). In the reaction homocysteine (SAH) is produced. Other SAM analogs (1b and 1c) are accepted by the Dam transferase and can be transferred to the N^6-position of adenine (2b and 2c). The SAM analogs as well as the expression of SAM-free Dam enzyme were provided by the group of Elmar Weinhold, RWTH Aachen.

In our approach, we interrogated if labeled DNA would interfere with TALE binding. The SAM analog 1b was of special interest. This alkynyl substituent exhibits a certain flexibility and does not hold a branched structure. Another advantage is, that 1b can be used for posterior labeling. Via the Cu(I)-catalyzed [3+2] azide-alkyne cycloaddition further labels and fluorophores can be introduced. To test our hypothesis, DNA duplexes with the sequence context of the pap operon 2 were modified by the Dam enzyme with SAM and other synthesized SAM analogs (see also methods 6.2.4). After the transfer of the group of interest (1a, b or c) to the N^6-position of adenine, the DNA was purified and subjected to mass spectrometry. The analysis by electrospray ionization time-of-flight (ESI-TOF) indicated quantitative labeling of both DNA strands with the substituent (see also appendix 7-11). Yet, the primer extension assay performed with labeled DNA (2b) and TALE_Pap2_NI protein did not exhibit differential binding of TALE protein (see figure 22). The TALE does not discriminate between DNA containing adenine or the 2b-labeled nucleobase. This result suggests that the larger alkynyl substituent is tolerated and well accommodated by TALEs even when the label is present in both DNA strands (see appendix 11). Models of this scenario indicate that the substituent can arrange and thread facing the open half of the groove towards the surface of the TALE-DNA complex. Furthermore, the model exhibits the possibility that even larger

substituents could be accommodated by TALEs when connected via the terminal alkyne moiety of 1b.
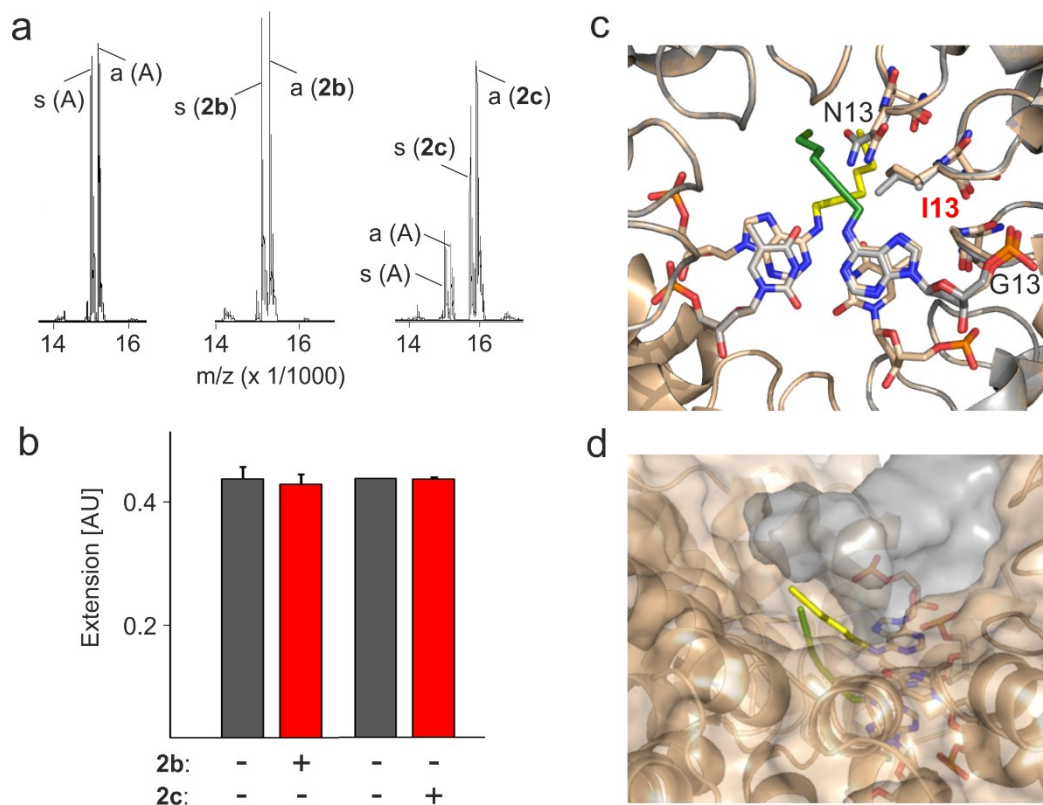
Figure 22: Binding of TALE_Pap2_NI to larger substituents. **A**: Analysis of DNA oligonucleotides by ESI-TOF. The first spectrum shows the Pap2 DNA without labeling followed by labeling reaction resulting in 2a and 2b (from left to right), s = sense strand, a = antisense strand. Masses: s calc: 14977.8, s found: 14977.9; a calc: 15162.8, a found: 15163.9; **2b:** s calc: 15082.8, s found: 15083; a calc: 15267.8, a found: 15268; **2c**: s calc: 15713.1, s found: 15714; a calc: 15898.1, a found: 15898.7. **B**: Quantification and analysis of TALE_Pap2_NI binding to DNA with or without substituents **2b** and **2c**. The TALE-DNA ratio was 60:1 (n=3). **C**: Superimposed models of TALE_Pap2_NI interacting with DNA either containing two adenines or **2a** in each strand. The substituent **2a** is positioned opposite the RVD NI in the sense strand (green sticks) as well as in the antisense strand embedded in the palindromic GATC motif (yellow sticks). **D**: Same superimposed model as in C but with differently orientated surface presentation. See also appendix 8 and 9.

Based on the information of the modeling that the TALEs could accept even larger groups, the labeling reaction was repeated with the bulkier cofactor 1c. This cofactor is derived from the molecule 1b and extended by a tetramethylrhodamin (TAMRA) fluorophore via the Cu(I)-catalyzed [3+2] azide-alkyne cycloaddition (see figure 21/1c). Analysis by ESI-TOF reported a successful labeling reaction but with reduced yield compared to 1b (see appendix 11). Nonetheless, primer extension again did not show any distinction in TALE_Pap2_NI binding

to either DNA containing A or 2c nucleobase. This finding implies that the bulky TAMRA group of 2c is exposed to the surface of the complex and that the N^6 position is a blind spot for larger substituents. The N^6-adenine position holds the possibility for even larger fluorophore labels to be tolerated by TALE proteins (see figure 23).
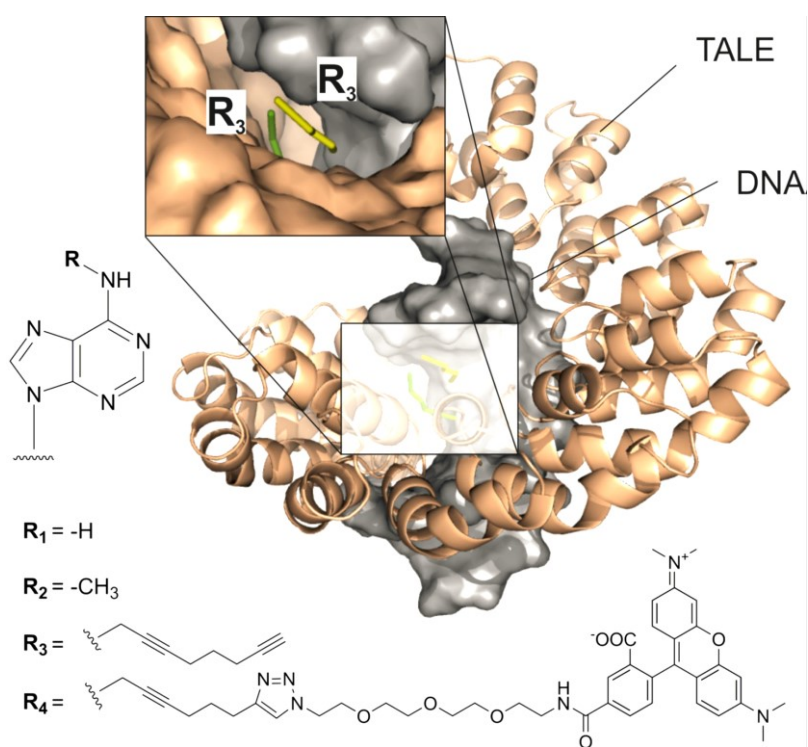
Figure 23: Model of TALE-Pap2_NI binding to DNA comprising substituents of the N^6-adenine position. The TALE protein is targeting DNA towards the major groove. Substitution of the N^6-methyl group by bulkier substituents still allows accommodation by the RVD NI. Even larger groups as a dialkynyl substituent and expanded fluorophore dye like TAMRA can be accepted by TALEs at this site. The bulky substituents are exposed to the surface of the TALE-DNA complex. Image from [77].

4. Summary and Outlook

TALE proteins are DNA binding proteins and useful tools for genome targeting and engineering approaches. TALEs represent convenient, programmable tools to target DNA for *in vivo* and *in vitro* applications. These proteins exhibit a high flexibility which is based on their framework of modular repeats. TALEs consist of multiple concatenated repeats which are highly conserved but harbor a variable part (RVD). Each RVD enables the selective recognition of one DNA nucleobase and stabilizes the interaction between DNA target and TALE protein. Each TALE can be assembled in the lab which makes the proteins versatile, effective targeting tools. However, knowledge is required about their binding behavior to DNA, especially if the target sequence comprises epigenetically modified nucleobases. Earlier studies reported methylation sensitive TALEs of 5-methyl cytosine and many approaches were developed to overcome this sensibility or to utilize it as effective sensory tools.[12,32,89,93] However, apart from epigenetic 5-methyl cytosine and its higher oxidized derivatives, the binding ability of TALE proteins to DNA containing the epigenetic mark of N^6-methyl adenine (6mA) was unknown. To shed light on this, we examined the binding of TALEs to the 6mA. This epigenetic mark is predominantly found in prokaryotic genomes but was recently reported to be present in eukaryotic organisms. In this work, we find that the naturally occurring RVD NI, which recognizes the DNA nucleobase A, is not sensitive towards 6mA. Model-assisted studies with RVD variants generating more hydrophobic surfaces and the disposition for hydrogen bonding to the N^6-amino group or N^7-atom of adenine showed that 6mA is well accommodated by the RVD mutants. Based on the modelling results, it was suggested that even larger groups at the N^6-position could be tolerated. Therefore, we used bulky N^6-alkynyl-substituents which can be transferred to the GATC motif by the Dam enzyme. These substituents can be utilized and extended for even larger groups like rhodamine dyes by simple click chemistry reaction. And indeed, TALEs can also accommodate bulky N^6-substitutions. Taken together, we report the first TALE binding studies to the epigenetic base of 6mA and its bulky substituents. Unlike cytosine and 5mC, 6mA is well tolerated by natural and modified RVDs and therefore, 6mA likely does not interfere in TALE based targeting approaches. In conclusion, these findings report the N^6-position of adenine as a blind spot for TALE proteins which allows the introduction of labels without interferences. These results allow valuable insights for

prospective approaches using TALEs and expand their scope of application as capture probes with labeled DNA.

As complementary approach, the identification of further RVD variants, which could recognize 6mA, could deliver further insights into the TALE DNA interaction. Libraries containing RVDs with mutations of the amino acid alphabet at position 13 might identify 6mA methylation sensitive TALE RVDs.[102] Also, structure-function studies on the repeat loop could deliver further information about TALEs interacting with N^6-substituents and other potential DNA fluorophores.[92] Consequently, these finding could be expanded to other DNA contexts, which hold for example the epigenetic mark N^4-methyl cytosine, which like 6mA, is predominantly found in prokaryotic genomes. Furthermore, crystal structures of TALES interacting with epigenetically modified DNA could support recent findings. With these approaches, further insights to TALEs interacting with epigenetic bases could be illuminated and used as useful knowledge for TALE construction, smart design of fluorescently labeled DNA as well as targeting and engineering approaches.

5. Zusammenfassung und Ausblick

TALE Proteine sind DNA bindende Proteine und sind ein nützliches Werkzeug für biologische Anwendungen, wie zum Beispiel die gezielte Einführung von Modifikationen im Genom einer Zelle. TALEs stellen ein günstiges, programmierbares Instrument dar, mit welchem DNA-Sequenzen in Experimenten genau angesteuert werden können. Diese Anwendung ist für *in vivo* und *in vitro* Versuchen möglich. TALE Proteine besitzen eine sehr hohe Flexibilität, welche auf einem Gerüst von sogenannten Repeat-Modulen basiert. Diese Proteine bestehen aus mehreren miteinander verketteten Repeats, welche hoch konserviert sind, jedoch einen variablen Bereich besitzen (RVD). Jeder dieser variablen Bereiche ermöglicht die selektive Erkennung von DNA Nukleinbasen und stabilisiert zusätzlich die Interaktion zwischen DNA und TALE Protein. Der modulare Aufbau der TALEs erlaubt die einfache Assemblierung dieser Proteine im Labor womit zahlreiche und vielfältige Anwendungen durchgeführt werden können. Jedoch setzt dies Wissen über das Bindeverhalten der TALE Proteine zur DNA voraus, insbesondere wenn die Zielsequenz epigenetische Modifizierungen beinhalten könnte. Frühere Studien berichten über 5-Methyl-Cytosin-sensitive TALEs und es wurden zahlreiche Ansätze entwickelt diese Sensitivität zu umgehen oder sie als sensorisches Werkzeug zu nutzen.[12,32,89,93] Jedoch ist bisher nicht bekannt, abgesehen von 5-Methyl-Cytosin (5mC) und seinen oxidierten Derivaten, wie TALE Proteine and DNA bindet, die die epigenetische Base N^6-Methyl-Adenosin (6mA) in seiner Sequenz trägt. Um dies zu beleuchten, untersuchten wir die Bindung von TALEs mit 6mA. Dieses epigenetische Signal kommt hauptsächlich in prokaryotischen Genomen vor, jedoch wurde es kürzlich auch in eukaryotischen Organismen nachgewiesen.[51,54–56,103]. In dieser Arbeit wird berichtet, dass das natürlich vorkommende RVD NI, welches die Nukleinbase Adenosin erkennt, nicht sensitiv gegenüber 6mA ist. Model-basierte Studien mit verschiedenen RVD Varianten, welche eine hydrophobere Oberfläche oder H-Brückenbildung zur N^6-Gruppe oder dem N^7-Atom ermöglichen könnten, zeigten, dass 6mA von den RVDs durch kleinere konformative Änderungen innerhalb des TALE Proteins akzeptiert werden kann. Es erschien möglich, dass sogar größere Gruppen an der N^6-Position durch die RVDs beherbergt werden könnte. Daher wurden in weiteren Experimenten größere N^6-Alkynyl-Substituenten an die entsprechende Position angebracht. Ein Vorteil dieser Substituenten ist, dass diese von der Dam Methyltransferase akzeptiert werden und auf das GATC Sequenzmotiv übertragen werden können. Diese Substituenten können durch Click-

Chemie mit zusätzlichen Gruppen, wie beispielsweise Rhodamin Farbstoffen, erweitert werden. Tatsächlich binden TALE Proteine auch an DNA, welche diese größeren N^6-Gruppen trägt. Im Gegensatz zu 5mC, kann 6mA von natürlichen und modifizierten RVDs beherbergt werden und zeigt vermutlich keinen Einfluss auf die TALE Bindung zur DNA. Diese Ergebnisse zeigen, dass die N^6-Position als ein blinder Punkt für TALE Proteine sein könnte, welcher die Einführung von Labeln ohne Einflussnahme auf die Protein-DNA Bindung erlaubt. Diese Ergebnisse erlauben wertvolle Einblicke für weitere TALE Anwendungen und erweitern den Einsatzumfang von TALE Proteinen mit markierter DNA.

Komplementäre Ansätze zur Identifizierung von 6mA-sensitiven RVD Varianten könnten die Einblicke in die TALE-DNA Interaktion vertiefen. Die Generierung von RVD-Bibliotheken, welche Mutationen des gesamten Aminosäuren-Alphabetes and Position 13 tragen, könnte 6mA-bindende RVD Varianten zu Tage befördern.[102] Des Weiteren könnten weitere Struktur-Funktions-Studien über die Repeat-Schleife weitere Informationen über TALEs, die mit größeren N^6-Gruppen und potentiellen Fluorophoren interagieren, liefern.[92] Diese Ergebnisse könnten auf andere DNA-Sequenzkontexte erweitert werden, welche beispielsweise die epigenetische Modifikation N^4-Methyl-Cytosin beinhalten, die wie 6mA überwiegend in prokaryotischen Genomen vorkommt. Außerdem könnte die Erzeugung von Kristallstrukturen von TALEs in Interaktion mit epigenetischen Basen bisherige Ergebnisse unterstützen und möglicherweise weitere Einblicke liefern. Diese Ergebnisse und Ansätze können verwendet werden für die Generierung von TALE Proteinen, sowie das smarte Design von Fluorophor-markierter DNA.

6. Materials and Methods

6.1 Materials

6.1.1 Oligonucleotides

DNA oligonucleotides were purchased from Sigma-Aldrich (0.025 µmol scale, purified by desalting) or from Metabion for methylated oligos (A). Oligonucleotides were stocked as 100 µM solutions in water. Working concentration of primers were adjusted to 10 µM.

Table 1: List of oligonucleotides used for primer extension assay.

Lab Code	Name/Locus	Sequence (5'-3')
oSaB_899	Lrp_Pap5	TTC TGC GAA AAG AAA GTC CGT AAA AAT TCA TTT AGA CGA TCT TTT ATG C
oSaB_873	Lrp_Pap5	TTC TGC GAA AAG AAA GTC CGT AAA AAT TCA TTT AGA CGA TCT TTT ATG C
oSaB_770	Lrp_Pap5	GCA TAA AAG ATC GTC TA
oSaB_900	Lrp_Pap2	GTT TTG TTC TAG TTT AAT TTT GTT TTG TGG GTT AAA AGA TCG TTT AAA T
oSaB_874	Lrp_Pap2	GTT TTG TTC TAG TTT AAT TTT GTT TTG TGG GTT AAA AGA TCG TTT AAA T
oSaB_875	Lrp_Pap2	ATT TAA ACG ATC TTT TA

Table 2: List of oligonucleotides used for electromobility shift assay.

Lab Code	Name/Locus	Sequence (5'-3')
oSaB_900	Lrp_Pap2	GTT TTG TTC TAG TTT AAT TTT GTT TTG TGG GTT AAA AGA TCG TTT AAA T
oSaB_874	Lrp_Pap2	GTT TTG TTC TAG TTT AAT TTT GTT TTG TGG GTT AAA AGA TCG TTT AAA T
oSaB_1815	Lrp_Pap2	ATT TAA ACG ATC TTT TAA CCC ACA AAA CAA AAT TAA ACT AGA ACA AAA C

Table 3: List of oligonucleotides used for site directed mutagenesis.

Lab Code	Name/Locus	Sequence (5'-3')
oSaB_808	pNI6_I13L_fw	TCG CCA GCA ACC TGG GCG GCA AGC AAG CGC T
oSaB_809	pNI6_I13L_rv	GCT TGC TTG CCG CCC AGG TTG CTG GCG ATA GC
oSaF_1110	pNI6_I13V_fw	CTA TCG CCA GCA ACG TTG GCG GCA AGC AAG CGC TCG
oSaF_1111	pNI6_I13V_rv	GCG CTT GCT TGC CGC CAA CGT TGC TGG CGA TAG CCA C
oSaF_1315	pNI6_I13E_fw	GTG GTG GCT ATC GCC AGC AAC GAA GGC GGC AAG CAA
oSaF_1316	pNI6_I13E_rv	CGC TTG CTT GCC GCC TTC GTT GCT GGC GAT AGC CAC

Table 4: List of oligonucleotides used for sequencing

Lab Code	Name/Locus	Sequence (5'-3')
oSaB_247	Seq. Primer	GGG TTA TGC TAG TTA TTG CTC AG
oDaS_315	Seq. Primer	CGT AGA GGA TCG AGA TC
oSaB_734	Seq. Primer	TTG ATG CCT GGC AGT TCC CT
oSaB_735	Seq. Primer	CGA ACC GAA CAG GCT TAT GT
oSaB_814	Seq. Primer	AAT GCG CAA ACC AAC CC
oSaB_890	Seq. Primer	AGA TAT GAT TGC GGC CCT G

6.1.2 Plasmids

Table 5: List of plasmids via TALE assembly and site directed mutagenesis.

Lab Code	Name/Locus		
pSaB327	pNI6_I13V	Tet	QC pNI6 Module
pSaB328	pNI6_I13L	Tet	QC pNI6 Module
pSaB339	pET-TRX-FINAL_FINAL_ENTRY	Carb	Backbone Vector
pSaB340	pap_lrp5_16 in pSaB339	Carb	TALE plasmid
pVaS342	pap_lrp2_16 in pSaB339	Carb	TALE plasmid
pVaS343	pap_lrp2_15 in pSaB339	Carb	TALE plasmid
pVaS344	pap_lrp2_14 in pSaB339	Carb	TALE plasmid
pVaS345	pap_lrp5_15 in pSaB339	Carb	TALE plasmid
pVaS360	pap_lrp2_16.NV6 in pSaB339	Carb	TALE plasmid
pVaS361	pap_lrp2_16.NL6 in pSaB339	Carb	TALE plasmid
pSaF469	pap_lrp5_16 NL in pSaB339	Carb	TALE plasmid
pSaF470	pap_lrp5_16 NV in pSaB339	Carb	TALE plasmid
pAnI521	pET_GFP_FINAL_ENTRY	Carb	Backbone Vector
pSaF522	pHD6_D13E	Tet	QC pHD6 Module
pSaF523	pNI6_I13E	Tet	QC pNI6 Module
pMoS715	pET_araBAD-GFP_FINAL_ENTRY	Carb	Backbone Vector
pSaF1105	pap_lrp2_16_NN6 in pSaB339_TRX	Carb	TALE plasmid
pSaF1106	pap_lrp2_16_NE6 in pSaB339_TRX	Carb	TALE plasmid
pSaF1107	pap_lrp2_16_NI in pAnI521_GFP	Carb	TALE plasmid
pSaF1108	pap_lrp2_16_NL in pAnI521_GFP	Carb	TALE plasmid
pSaF1109	pap_lrp2_16_NV in pAnI521_GFP	Carb	TALE plasmid
pSaF1110	pap_lrp2_16_NN in pAnI521_GFP	Carb	TALE plasmid
pSaF1111	pap_lrp2_16_NE in pAnI521_GFP	Carb	TALE plasmid
pSaF1112	pap_lrp5_16_NI in pAnI521_GFP	Carb	TALE plasmid
pSaF1113	pap_lrp5_16_NL in pAnI521_GFP	Carb	TALE plasmid
pSaF1114	pap_lrp5_16_NV in pAnI521_GFP	Carb	TALE plasmid
pSaF1115	pap_lrp5_16_NN in pAnI521_GFP	Carb	TALE plasmid
pSaF1116	pap_lrp5_16_NE in pAnI521_GFP	Carb	TALE plasmid
pSaF1117	pap_lrp2_16_NI in pMoS715_GFP	Carb	TALE plasmid
pSaF1118	pap_lrp2_16_NV in pMoS715_GFP	Carb	TALE plasmid
pSaF1119	pap_lrp2_16_NN in pMoS715_GFP	Carb	TALE plasmid
pSaF1120	pap_lrp2_16_NE in pMoS715_GFP	Carb	TALE plasmid
pSaF1121	pap_lrp5_16_NI in pMoS715_GFP	Carb	TALE plasmid
pSaF1122	pap_lrp5_16_NL in pMoS715_GFP	Carb	TALE plasmid
pSaF1123	pap_lrp5_16_NV in pMoS715_GFP	Carb	TALE plasmid
pSaF1124	pacyc_trcPromotor_Lrp2_16_GATC_mplum	Cam	Target plasmid

6.1.3 TALE Proteins

Table 6: Overview of expressed TALE proteins.

Lab Code	Name/Locus	Plasmid
TALE_026	Lrp_Pap5_16_NI_TRX	pSaB340
TALE_027	Lrp_Pap2_16_NI_TRX	pVaS342
TALE_029	Lrp_Pap2_15_NI_TRX	pVaS343
TALE_030	Lrp_Pap2_14_NI_TRX	pVaS344
TALE_031	Lrp_Pap5_15_NI_TRX	pVaS345
TALE_032	Lrp_Pap5_14_NI_TRX	pVaS346
TALE_033	Lrp_Pap2_16_NV_TRX	pVaS360
TALE_034	Lrp_Pap2_16_NL_TRX	pVaS361
TALE_035	Lrp_Pap5_16_NV_TRX	pSaF470
TALE_036	Lrp_Pap2_16_NL_TRX	pSaF469
TALE_112	Lrp_Pap2_16_NN_TRX	pSaF1105
TALE_113	Lrp_Pap2_16_NE_TRX	pSaF1106
TALE_316	pap_lrp2_16_NI_GFP	pSaF1107
TALE_317	pap_lrp2_16_NL_GFP	pSaF1108
TALE_318	pap_lrp2_16_NV_GFP	pSaF1109
TALE_319	pap_lrp2_16_NN_GFP	pSaF1110
TALE_320	pap_lrp2_16_NE_GFP	pSaF1111
TALE_321	pap_lrp5_16_NI_GFP	pSaF1112
TALE_322	pap_lrp5_16_NL_GFP	pSaF1113
TALE_323	pap_lrp5_16_NV_GFP	pSaF1114
TALE_324	pap_lrp5_16_NN_GFP	pSaF1115
TALE_325	pap_lrp5_16_NE_GFP	pSaF1116
TALE_326	pap_lrp2_16_NI_araBAD_GFP	pSaF1117
TALE_327	pap_lrp2_16_NV_araBAD_GFP	pSaF1118
TALE_328	pap_lrp2_16_NN_araBAD_GFP	pSaF1119
TALE_329	pap_lrp2_16_NE_araBAD_GFP	pSaF1120
TALE_330	pap_lrp5_16_NI_araBAD_GFP	pSaF1121
TALE_331	pap_lrp5_16_NL_araBAD_GFP	pSaF1122
TALE_332	pap_lrp5_16_NV_araBAD_GFP	pSaF1123

6.1.4 Bacterial Strains

Table 7: Overview of bacterial strains used for cloning and expression of TALE proteins.

Lab Code	*E. coli* Strain	Genotype	Supplier
BL21 DE3	sMoS308	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5	Stratagene
BL21 DE3 Gold	sMoS309	F- ompT gal dcm lon hsdS (rB– mB–) λ(DE3 [laci lacUV5-T7 gene 1 ind1 sam7 nin5)	Agilent
GH371	sDaS58	F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu)7697 galU galK rpsL (StrR) endA1 nupG fhuA::IS2 upp-	iGEM
K12 ER 2925(Dam-)	sSaF374	*ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2 galT22 mcrA dcm-6 hisG4 rfbD1 R(zgb210::Tn10)TetS endA1 rpsL136 dam13::Tn9 xylA-5 mtl-1 thi-1 mcrB1 hsdR2*	New Englang Biolabs
K12 ER 2738(Dam+)	sSaF375	F´ *proA+B+ lacI^q Δ(lacZ)M15 zzf::Tn10(Tet^R)/ fhuA2 glnV Δ(lac-proAB) thi-1 Δ(hsdS-mcrB)5*	AG Goettrup, Konstanz

6.1.5 Media

Table. 8: Overview of media used for optimal growth of *E. coli.*

Type of Media	Chemical	Concentration
LB-medium (Lysogeny Broth) (pH = 7.0)	Tryptone	10 g/L
	Yeast extract	5 g/L
	NaCl	5 g /L
LB-agar (pH = 7.0)	Tryptone	10 g/L
	Yeast extract	5 g/L
	NaCl	5 g/L
	Agar-Agar	15 g/L
SOC-medium (pH = 7.0)	Tryptone	2 % (w/v)
	Yeast extract	0.5 % (w/v)
	NaCl	10 mM
	KCl	2.5 mM
	$MgCl_2$	10 mM
	$MgSO_4$	10 mM
	Glucose	20 mM

6.1.6 Buffers

6.1.6.1 Buffers for Protein Purification

Table 9: Buffers for native purification of TALE proteins with GFP-tag.

Native purification	Chemicals	Concentration
Qiagen lysis buffer (pH = 8.0)	NaH_2PO_4	50 mM
	NaCl	300 mM
Qiagen wash buffer 1 20 mM Imidazole (pH = 8.0)	NaH_2PO_4	50 mM
	NaCl	300 mM
	Imidazole	20 mM
Qiagen wash buffer 2 50 mM Imidazole (pH = 8.0)	NaH_2PO_4	50 mM
	NaCl	300 mM
	Imidazole	50 mM
Qiagen elution buffer 500 mM Imidazole (pH = 8.0)	NaH_2PO_4	50 mM
	NaCl	300 mM
	Imidazole	500 mM
PBS (4x) (pH = 8.0)	NaCl	548 mM
	KCl	43 mM
	Na_2HPO_4	69 mM
	KH_2PO_4	24 mM

Table 10: Buffers for denaturing workup of TALE proteins with TRX-tag.

Native purification	Chemicals	Concentration
Taq lysis buffer (pH = 9.0)	Tris*HCl	100 mM
	NaCl	300 mM
	MgCl2	2.5 mM
	TritonX-100	0.1 % (v/v)
Buffer-Z wash buffer 1 (pH = 8.0)	Urea	8 M
	NaCl	100 mM
	Hepes	20 mM
	Imidazole	20 mM
Buffer-Z wash buffer 2 (pH = 8.0)	Urea	8 M
	NaCl	100 mM
	Hepes	20 mM
	Imidazole	50 mM
Buffer-Z elution buffer (pH = 8.0)	Urea	8 M
	NaCl	100 mM
	Hepes	20 mM
	Imidazole	500 mM
PBS (4 x) (pH = 8.0)	NaCl	548 mM
	KCl	43 mM
	Na_2HPO_4	69 mM
	KH_2PO_4	24 mM

6.1.6.2 Buffer for Dialysis of TALE Proteins

Table 11: Overview of used dialysis buffers.

Dialysis buffer	Chemicals	Concentration
5 x TALE storage buffer (pH = 7.5)	Tris*HCl	100 mM
	NaCl	1 mM
	Glycerol	50 % (v/v)

6.1.6.3 Reaction Buffers

Table 12: Reaction buffers for Primer Extension.

Primer Extension	Chemicals	Concentration
5 x TALE storage buffer (pH = 7.5)	Tris*HCl	100 mM
	NaCl	1 mM
	Glycerol	50 % (v/v)
5 x TALE binding buffer	Tris*HCl (pH = 8)	100 mM
	NaCl	250 mM
	$MgCl_2$	25 mM
	BSA	0.5 mg/ml
	Glycerol	50 % (v/v)
KF storage buffer	Tris*HCL (pH = 7.5)	25 mM
	EDTA	0.1 mM
	DTT	1 mM
	Glycerol	50 % (v/v)

Table 13: Reaction buffers for EMSA.

EMSA	Chemicals	Concentration
5 x BGrK1 (pH = 7.5)	Tris*HCl (pH = 8)	100 mM
	NaCl	250 mM
	$MgCl_2$	25 mM

Table 14: Reaction buffers for alkylation by Dam methyltransferase.

Alkylation	Chemicals	Concentration
Dam reaction buffer (10 x) (pH = 8)	Tris*HCl (pH = 8) EDTA DTT	1500 mM 100 mM 20 mM
Dam storage buffer (pH = 7.5)	Tris*HCl KCl EDTA DTT Glycerol	50 mM 50 mM 10 mM 1 mM 52.5 % (v/v)

6.1.6.4 Buffers for Gel-Electrophoresis

Table 15: Buffers and gel composition for denaturing PAGE.

Denaturing PAGE	Chemicals	Concentration
TBE running buffer (10 x)	Tris*HCl Boric Acid EDTA	890 mM 890 mM 20 mM
TBE running buffer/Urea	Tris*HCl Urea Boric Acid EDTA	890 mM 9 M 890 mM 20 mM
Denaturing PAGE gel (10 %)	Sequencing gel concentrate 10 x TB / 9 M urea 9 M urea APS (10 % w/v) TEMED	40 ml 10 ml 50 ml 800 µl 40 µl
Denaturing loading buffer	Formamide EDTA Bromophenol blue	80 % (v/v) 2 mM 1 g

Table 16: Buffers and gel composition for EMSA.

EMSA	Chemicals	Concentration
10 x TAE	Tris*HCl Boric Acid EDTA	890 mM 890 mM 20 mM
EMSA native gel	Gel 40 concentrate 10 x TAE buffer MilliQ-H_2O APS (10 % w/v) TEMED	2 ml 0.5 ml 7.4 ml 100 µl 10 µl

Table 17: Buffers for Agarose Gel-Electrophoresis.

Agarose Electrophoresis	Chemicals	Concentration
TBE running buffer (0.5 x)	Tris*HCl	45 mM
	Boric Acid	45 mM
	EDTA	1 mM
NEB gel loading dye (6 x)	FicollR-400	15 % (v/v)
	EDTA	66 mM
	Tris*HCl	19.8 mM
	SDS	0.102 % (w/v)
	Bromophenol blue	0.09 % (w/v)
Staining solution	Ethidium bromide	0.1 % (w/v)

Table 18: Buffers and gel composition for SDS-PAGE.

SDS-PAGE	Chemicals	Concentration
SDS loading buffer (4 x) pH = 6.8	Tris*HCl	200 mM
	Glycerol	40 % (v/v)
	2-Mercaptoethanol	1 % (v/v)
	SDS	8 % (w/v)
	Bromophenol blue	0.008 % (w/v)
SDS running buffer (10 x)	Tris Base	30.3 g/L
	Glycerol	144 g/L
	SDS	10 g/L
Resolving gel (8 %)	MilliQ-H_2O	3808 µl
	Gel 40 (40 % Acrylamid-, Bisacrylamid 37,5:1)	1092 µl
	Tris (1.5 M, pH = 8.8)	420 µl
	SDS (10 % w/v)	66.7 µl
	APS (10 % w/v)	66.7 µl
	TEMED	6.67 µl
Stacking gel (5 %)	MilliQ-H_2O	1000 µl
	Gel 40 (40 % Acrylamid-, Bisacrylamid 37,5:1)	208 µl
	Tris (0.5 M, pH = 6.8)	420 µl
	SDS (10 % w/v)	16.7 µl
	APS (10 % w/v)	16.7 µl
	TEMED	1.67 µl
Coomassie staining	MilliQ-H_2O	40 % (v/v)
	Acetic Acid	10 % (v/v)
	Methanol	50 % (v/v
	Coomassie Brilliant Blue G250	0.1 % (w/v)
Coomassie destaining	MilliQ-H_2O	70 % (v/v)
	Acetic Acid	10 % (v/v)
	Methanol	20 % (v/v)

6.1.6.5 Buffers for Chemically Competent Cells.

Table 19: Buffers for preparation of chemically competent *E. coli* cells.

Buffer	Chemicals	Concentration
$MgCl_2$-buffer	$MgCl_2$	100 mM
$CaCl_2$-buffers	$CaCl_2$	50 mM
Storage-buffer	$CaCal_2$	50 mM
	Glycerol	15 % (v/v)

6.1.7 Antibiotic Stocks, Enzymes and Kits.

Table 20: Overview of antibiotic stocks. All antibiotics were dissolved in their respective solvent and were subsequently sterile filtered with a 0.2 μM syringe filter.

Antibiotic Stock	Solvent	Concentration
Chloramphenicol (Cam)	Ethanol	34 mg/L
Carbenicillin (Carb)	Ethanol/MilliQ-H2O-H_2O (1:1 v/v)	50 mg/L
Spectinomycin (Spec)	MilliQ-H_2O	100 mg/L
Tetracycline (Tet)	Ethanol	12.5 mg/L

Table 21. Overview of enzymes used for various reactions.

Enzymes	Commercial Supplier
BsaI	New England Biolabs
BsmBI	New England Biolabs
Dam mtase (S-adenosylmethionine free)	AG Weinhold, Aachen
DpnI	New England Biolabs
DpnII	New England Biolabs
KF(exo-)	New England Biolabs
Lysozyme	Fluka
Plasmid safe nuclease	Biozym
Phusion polymerase	New England Biolabs
Taq polymerase	NEB/self-expressed
T4-DNA ligase	New England Biolabs
T4 polynucleotidekinase	Fermentas
T5 exonuclease	New England Biolabs

Table 22: Kits.

Kits	CommercialSupplier
BCA Protein Assay	Thermo Scientific
BCA Protein Assay (DTT compatible)	Thermo Scientific
GeneJET Gel Extraction Kit	Thermo Scientific
GeneJET Plasmid Miniprep Kit	Thermo Scientific

6.1.8 Reagents and Chemicals

Table 23: Used reagents and chemicals in this work.

Reagents and Chemicals	Commercial Supplier
Acetic acid	Carl Roth
Agar-Agar (Cobe I)	Carl Roth
Agarose	Biozym
Ammonium persulfate (APS)	Carl Roth
Boric Acid	Carl Roth
Bovine Serum Albumin (BSA)	Cell Signaling
Bromphenol blue	Carl Roth
Calcium Chloride	Fisher Scientific
Carbenicillin disodium salt	Carl Roth
Chloramphenicol	Carl Roth
Coomassie Brilliant Blue G250	Carl Roth
Disodiumhydrogen phosphate	Sigma Aldrich
dNTPs	New England Biolabs
Dithiothreitol (DTT)	Carl Roth
Ethanol p.a.	Sigma Aldrich
Ethanol 96%	Fisher Scientific
Ethidium bromide	Sigma Aldrich
Ethylenediaminetetraacetate (EDTA)	Carl Roth
Formamide	Acros
[gamma-^{32}P]-ATP	Hartmann Analytics
Glucose	Carl Roth
Glycerol	Carl Roth
Hepes	Carl Roth
Hydrochloric Acid (37%)	VWR
Imidazole	Abcr
Isopropanol (99.9%)	Fisher Scientifc
Isopropyl ß-D-1thiogalactopyranoside (IPTG)	Carl Roth
Kanamycinsulfate	Carl Roth
L-Arabinose	Carl Roh
LB-agar	Carl Roth
LB-medium	Carl Roth
LE-agarose	Carl Roth
Magnesium chloride hexaydrate	Acros
Magnesium sulfate heptahydate	Merck
Methanol	Sigma Aldrich
Nickel Nitrilotriacetic acid (Ni-NTA resin)	Thermo Scientific
Phenylmethanesulfonyl fluoride (PMSF)	Carl Roth
Potassium chloride	Carl Roth
Potassium dihydrogenphosphate monohydrate (KH2PO4)	Sigma Aldrich
Protein Marker Broad Range (2-212 kDa)	New England Biolabs

Rotiphorese Gel 40	Carl Roth
Rotiphorese sequencing gel concentrate	Carl Roth
S-adenosylmethonine	New England Biolabs
Sodium chloride (NaCl)	Carl Roth
Sodium dihydrogen phosphate monohydrate	Merck
Sodium dodecyl sulfate (SDS)	Carl Roth
Sodium hydroxide (NaOH)	Fisher Scientific
Spectinomycin	Alfa Aeser
Tetracyclin hydrochlorid	Carl Roth
Tetramethylethyldiamine (TEMED)	Carl Roth
Tris(hydroxymethyl)-aminomethan (TRIS)	Sigma Aldrich
Tryptone	Carl Roth
Triton X-100	Fluka
Unstained protein marker 2-212 kDa	New England Biolabs
Urea	Carl Roth
Xylene cyanole	Carl Roth
Yeast extract	Carl Roth
2-Log DNA ladder	New England Biolabs
2-Mercaptoethanol	Merck

6.1.9 Lab Equipment

Table 24: Used lab equipment for this work.

Lab Equipment	Commercial Supplier
Autoclave	Systec
Balance ABT 220-4M	Kern
Balance PlJ 3500-2NM	Kern
Balance SE622	VWR
Biophotometer Plus	Eppendorf
Centrifuge 5810R	Eppendorf
Centrifuge Universal 320R	Hettich
Centrifuge Sorvall Evolution RC	Thermo Scientific
Concentrator plus speed-vac	Eppendorf
Contamat FHT 111M	Thermo Scientific
Electrophoresis power supply EV233	Consort
Electrophoresis power supply EV243	Consort
Electrophoresis power supply EV245	Consort
Electrophoresis power supply	BioRad
Fridge ProLine (4°C)	Liebherr
Freezer ProLine (-20°C)	Liebherr
SDS-PAGE Gel electrophoresis system	BioRad
Gel electrophoresis system Midicell Primo EC330	Thermo Scientific
Gel electrophoresis system 45-2010-i	Peqlab
Gel documentation BDA digital	Biometra
Gel Dryer MGD 4534	VWR
Homogenizer Emulsi-Flex-C5	Avestin
Heating block digital dry bath	Labnet
Ice machine AF20	Scotsman
Incubator	Memmert
Incubator INCU-line	VWR
Infinite M200 plate reader	Tecan
Incubator 1000	Heidolph
Magnetic stirrer MR Hei-standard	Heidolph
Magnetic stirrer MR3002	Heidolph
Magnetic stirrer RCT Classic	IKA
Microwave MW82N	Samsung
Millipore filter system	ELGA
Multichannel pipette 10 µl (8 channels) Research Plus	Eppendorf
Multichannel pipette Xplorer 100 µl (12 channels)	Eppendorf
PCR Cycler (Mycycler)	BioRad
PCR Cycler (Primus 25 advanced)	Peqlab
pH Meter	Mettler Toledo
Phosphor screen cassettes BAS 2025	Fuji
Phosphor imager Molecular Imager Chemi-Doc	BioRad

Pipette boy Easypet	Eppendorf
Pipettes Research Plus	Eppendorf
Phosphor imager Pharos-FX Plus	BioRad
Phosphorimager Typhoon FLA 9500	GE Healtcare
Research Plus Pipettes (2.5, 10, 100, 1000 µl)	Eppendorf
Sequencing Gel System	BioRad
Sequencing Gel System	Carl Roth
Shaking Incubator Ecotron	Infors HAT
Shaking Incubator I26	New Brunswick Scientific
Shaker Unimax 1010	Heidolph
Sonicator Digital	Branson
Tabletop Centrifuge 5417R	Eppendorf
Tabletop Centrifuge 5424	Eppendorf
Tabletop Centrifuge Minispin	Carl Roth
Thermomixer comfort	Eppendorf
Thermomixer compact	Eppendorf
Thermomixer 96 well plate	BioER
Thermomixer 96 well plate MB-102	BioER
Ultra-low temperature freezer (-80°C)	New Brunswick Scientific
UV-Star UV-table (312 nm)	Biometra
Vacuum pump	Vaccubrand
Vortex Genie 2	Scientific Industries
Water bath JB Aqua 12 Plus	Grant

6.1.10 Consumables

Table 25: Single-use products consumed in this work.

Consumables	Commercial Supplier
96-deep well plates	Sarstedt
96-well plates transparent	Carl Roth
96-well plates, flat bottom	Greiner
Amicon Ultra MWCO 3 kDa	Millipore
Cannulas Sterican 0.60 x 80 mm	Braun
Cannulas Sterican 0.80 x 120 mm	Braun
Cannulas Sterican 0.90 x 40 mm	Braun
Colum (PP 5ml)	Qiagen
Cuvettes semi-mikro	VWR
Falcon tube 15 ml	Sarstedt
Falcon tube 30 ml	Sarstedt
Glas beads	Carl Roth
Gloves Nitrile	VWR
PCR tubes 0.2 ml	Sarstedt
Petri dishes	Sarstedt
Pipette tips	Sarstedt
Reaction tubes 1.5 ml	Sarstedt
Reaction tubes 2 ml	Sarstedt
Scalpel	Mediwar
Self adhesive cover film	Ratiolab
Sephadex columns G25	Life Technologies
Serological pipette 10 ml / 25 ml	Sarstedt
Slide-A-Lyzer Mini Dialysis Unit MWCO: 10 kDa	Thermo Scientific
Syringe	Henke Sass Wolf
Syringe needle	Braun
Syringe filter 0.2 µM	Sarstedt
UV-transparent disposable cuvettes	Sarstedt
Whatman paper 3 mm	Whatman
ZelluTrans dialysis membrane visking Cellulose	Carl Roth

6.1.11 Software

Table 26: Software used for documentation, analysis, quantification and presentation of results.

Software	Supplier
BioDoc Analyse Geldokumentation	Analytic Jena
ChemDraw Professional 15	Cambridgesoft
CorrelDraw Graphic Suite X6	Softronic
FinchTV	Geospiza
ImageJ 1.50i	Nat. Institute of Health
MS Office	Microsoft
Serial Cloner	Serial Basics
Pymol	DeLanoScientific LLC
Quantity One	BioRad

6.2 Methods

6.2.1 General Biomolecular Methods

6.2.1.1 Preparation of Chemically Competent Cells

20 ml of LB-media was inoculated with a single *E. coli* colony picked from an agar plate. The culture was incubated at 37°C overnight shaking at 180 rpm. The next day, 800 ml of LB-medium was inoculated with 8 ml of the overnight culture. The culture was incubated at 37°C until the $OD_{600\ nm}$ of 0.4 was reached. The culture was cooled on ice and equally distributed to 16 flacon tubes (50 ml). Cells were harvested by centrifugation at 4000 rpm at 4°C for 10 min. The supernatant was discarded and the pellet were resuspended in 10 ml $MgCl_2$ buffer. The cell suspension was centrifuged again and the remaining pellets were resuspended in 10 ml $CaCl_2$ buffer. The bacterial suspension was incubated for 30 min on ice and again centrifuged at 4°C, 4000 rpm for 10 min. The supernatant was discarded and the pellets were pooled and resuspended in an overall volume of 4 ml storage buffer. The cell suspension was aliquoted and snap frozen with liquid nitrogen. The cells were stored at -80°C.

6.2.1.2 Transformation of Chemically Competent Cells

About 10 to 100 ng of plasmid solution was mixed with 50 µl of chemically competent E. coli cells (50 µl aliquot). The suspension was mixed carefully by stirring with a pipette tip and incubated on ice for 30 min. Then, the cells were heat shocked at 42°C for 30 seconds, incubated again on ice for 2 min. The cells were rescued with 1 ml pre-warmed SOC by incubation at 37°C for 1 h (1000 rpm shaking). After incubation, the cells were plated on LB-agar plated supplied with the respective antibiotic and incubated at 37°C overnight.

6.2.1.3 Isolation of Plasmids

For isolation of plasmids, 5 ml of LB-media was inoculated with a single colony picked from an agar plated. The culture was supplied with the respective antibiotic and incubated at 37°C overnight. The next day, the culture was pelleted by centrifugation at 4000 rpm for 10 min and the supernatant was discarded. The pellet was resuspended in 250 µl of resuspension buffer and transferred to a 1.5 Eppendorf tube following the manufacturers protocol (GenJET

plasmid Minprep Kit, Qiagen). Then, 250 µl lysis solution were added and the rubes were inverted several times followed by addition of neutralization buffer and repeated inversion. After centrifugation for 5 min at 1400 rpm the supernatant was isolated from the cell debris and transferred to the supplied columns. After centrifugation for 1 min, 500 ml of washing buffer was added and the flow through was discarded. The columns were centrifuged for 1 min to remove remaining washing buffer. Finally, the columns were placed in a new 1.5 ml Eppendorf tube and air dried to remove residual ethanol. 50 µl of MilliQ-H_2O were added to the column and the plasmids were eluted by centrifugation at 14000 rpm for 2 min.

6.2.1.4 Sanger Sequencing of DNA

5 µl of 30-75 ng of DNA, 2.5 µl of respective primers (2.5 µM) and 2.5 µl of MilliQ-H_2O were mixed and send for sequencing in a 1.5 ml Eppendorf tube to GATC biotech company. Sequence traces were analyzed using Finch TV software and aligned to theoretical sequences using Serial Cloner software.

6.2.1.5 Agarose Gel-Electrophoresis

Agarose powder was resuspended in 0.5 x TBE buffer and heated in a microwave until it was completely dissolved. The solution was cooled to approximately 40°C and then poured into a gel casting tray. A comb was placed on top and the polymerized gel was transferred to a gel chamber filled with 0.5 x TBE buffer. Prepared DNA samples were mixed with DNA loading dye and carefully loaded onto the gel. The gel was run for 1 h at 120 V or adjusted conditions for higher resolution. The gel was incubated in an aqueous ethidium bromide solution for 5 min and destained in MilliQ-H_2O. The gel was put on a Biometra UV-table (254 nm) and the image was collected by a gel documentation system. If necessary, DNA fragments could be extracted from the agarose gel by excising the fragment from the gel with a scalpel followed by use of GenJET gel extraction kit (Qiagen).

6.2.2 Cloning and Modification of Plasmids

6.2.2.1 Site Directed Mutagenesis

For introduction of site-directed mutagenesis of TALE module plasmids (repeat position 13) primers were designed following the guidelines from Agilent for QuikChange site directed mutagenesis. Forward and reverse primers exhibit melting temperatures of ≥ 78°C, a GC content above 40 % and the ends should possibly contain a G or C. The lengths of the primers varied to match the required melting temperatures, but were commonly between 30 to 45 bp. The desired mutational site should be placed in the middle part of the sequence with 15 to 20 bp flanking it.

The reaction was performed in a total volume of 25 µl. The reaction mix contained 10 to 50 ng template plasmid, 200 µM dNTPs, 200 nM forward and reverse primers, as well as Pfu Ultra DNA polymerase and its respective reaction buffer. The reaction was carried out with an initial denaturation step for 30 s at 95°C, followed by 16 cycles of 30 s at 95°C, annealing for 60 s at 55°C and elongation at 60 s per kb at 68°C. The final elongation was conducted at 68°C for 7 min followed by storage of the reaction at 4°C. Subsequently, the reaction was digested with *DpnI* (0.5 µl, 10 U) at 37°C for 1 h. Without further purification, 4 µl were transformed into 50 µl of chemically competent *E. coli* GH371 cells. Clones were isolated and checked for correct mutagenesis by Sanger sequencing.

6.2.2.2 TALE Assembly

TALE plasmids were assembled according to a two-step Golden Gate protocol published by Cermark et al in 2011[97]. The module repeat plasmids were assembled according to the desired RVD sequence using the module plasmids numbered in that order. The first intermediate array called pFUS_A can include up to 10 repeat modules. The next repeat modules are cloned into a second array pFUS_B (see figure 24). In the first Golden Gate reaction, the single module plasmids as well as the backbone intermediate arrays are subjected to digestion by *BsaI* enzyme at 37°C for 5 min followed by ligation with T4 DNA ligase at 16°C for 10 min. The reaction is incubated for 10 to 15 cycles, then heated to 50°C for 5 min and terminated at 80°C for 5 min. 0.5 µl of ATP (25 mM stock) and 0.5 µl of plasmid safe nuclease are added to the 25 µl reaction and incubated at 37°C for 1 h. Eventually, the reaction was heat inactivated

by incubation at 70°C for 20 min. Approximately 4 µl of this reaction mix was transformed into *E. coli* GH371 cells without any further purification steps. Positive clones could be pre-selected by a blue-white screening if XGal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) was added to the plates. Up to 4 white clones were picked for further analysis by colony PCR. Therefore, 1 µl of the respective bacterial suspension was diluted to a 25 µl reaction volume supplemented with 250 µM dNTPs, 100 nM of forward and reverse primes oSaB374/oSaB375, 40 mU Taq polymerase and thermo pol buffer. The reaction mix was incubated at 94°C for 2 min, followed by 35 cycles of 94°C, 55°C 30 s, 68°C 2 min and a final extension step at 68°C for 2 min. The results of the colony PCR were monitored on a 1 % agarose gel for correct size. If the clones were identified with the correct size, 5 ml of overnight cultures supplied with spectinomycin were inoculated with the respective bacterial suspension. 1 µl of each bacterial suspension was spotted to a safe plate.
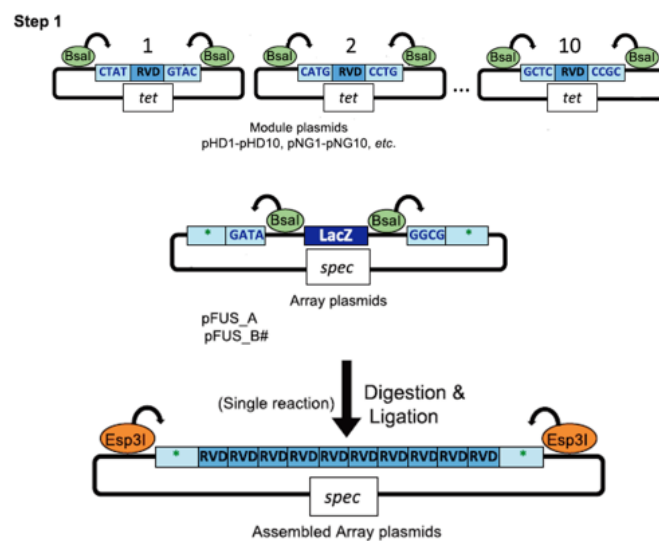
Figure 24: First step of Golden Gate assembly of TALE protein. The module plasmids are arranged according to the sequence of interest and assembled to an array plasmid via a restriction and ligation reaction. Adapted from [97]

The next day, plasmids were isolated and subjected to Golden Gate 2 reaction. In this reaction, the intermediate arrays pFUS_A and pFUS_B, along with the last repeat and the backbone vector pET_TRX_ENTRY or pET_GFP_ENTRY are joined (see figure 25). A 10 µl digestion with *Bsmb*I restriction enzyme and ligation reaction mixture is prepared as in the

first step (10 cycles). Plasmid safe treatment was omitted in this step. About 4 µl of the reaction was transformed into expression strain *E. coli* BL21 (DE3). The transformed cells were plated on LB-agar plates supplemented with carbenicillin. The next day 4 white colonies were picked and monitored by colony PCR (primers oSaB890/oDaS247). If the size was correct, 5 ml of LB-media cultures were inoculated with the respective bacterial suspension and incubated at 37°C overnight. The next day the plasmids were isolated and sent for sanger sequencing to check for correct sequence using sequencing primers oDaS315, oSaB890 and oDaS247.

Figure 25: Second step of Golden Gate assembly of TALE protein. Each array plasmid can hold up to 10 repeat modules. Consequently, the array repeats and the last repeat are joined in a second restriction and ligation reaction and connected to the backbone vector, which holds the N-terminal and C-terminal domains of the protein. Adapted from [97].
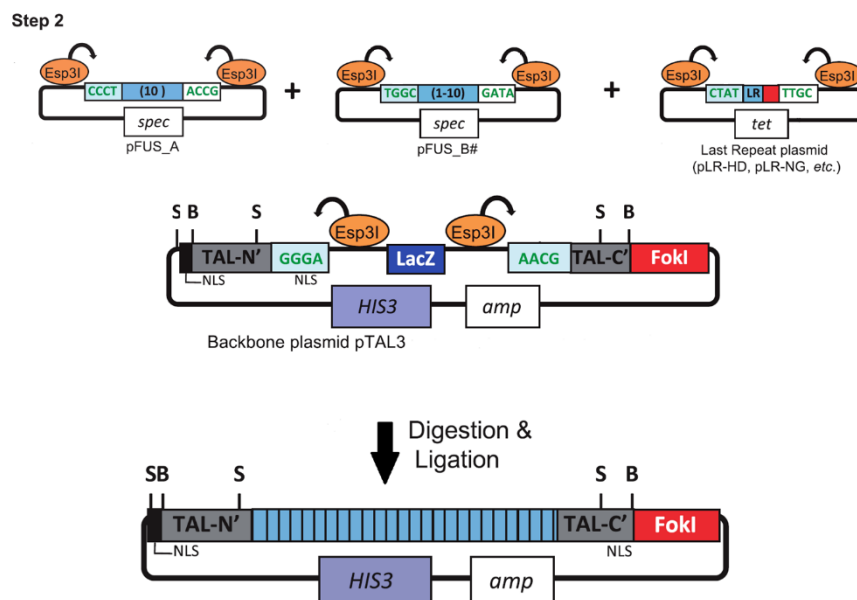
6.2.3 Expression and Purification

6.2.3.1 Expression and Purification of TALE Proteins with TRX-Tag

Single clones of *E. coli* BL21 (DE3) Gold transformed with TALE expression plasmids (pSaB339_pET_TRX_final_entry) picked and 5 ml LB-media supplemented with 50 µg/ml carbenicillin was inoculated. The culture was incubated at 37°C overnight. The next day, the overnight cultures were diluted 40-fold into LB-medium supplemented with carbenicillin. The culture was grown at 37°C and 250 rpm shaking until an $OD_{600\ nm}$ of 0.4 was reached and induced with a final concentration of 0.2 mM IPTG. After 5 h of incubation, expression culture was harvested by centrifugation (15 min, 4000 rpm, 4°C). The supernatant was discarded and the remaining pellet was lysed in Taq lysis-buffer supplied with 1 mM PMSF and 50 µg/ml lysozyme by shaking at room temperature at 1400 rpm for 30 min. The suspension was pelleted by centrifugation (20 min, 9000 rpm, 4°C) and the supernatant was discarded. Under denaturing conditions, the TALE protein was extracted from the pellet and bound to Ni-NTA resin according to the manufactures protocol. The resin was washed two times with 4 x PBS-buffer four times with washing buffer 1 containing 20 mM imidazole and once with washing buffer 2 with 50 mM imidazole. Finally, the protein was eluted from the column with elution buffer containing 500 mM imidazole. The respective fractions were pooled and monitored by SDS-PAGE.

6.2.3.2 Expression and Purification of TALE Proteins with GFP-Tag

Single clones of *E. coli* BL21 (DE3) Gold transformed with TALE expression plasmids based on pET_GFP_final_entry (pAnI521) and expressed as described above. After 5 h of expression the cultures were harvested by centrifugation (15 min, 4000 rpm, 4°C). The remaining pellets were lysed in Qiagen lysis buffer containing 1 mM PMSF and 50 µg/ml lysozyme and additionally disrupted using French press. Then, the solution was centrifuged (15 min, 9.000 rpm, 4°C), the supernatant was collected and extracted with Ni-NTA). For purification, the same procedure was applied as described for TRX-TALEs above.

6.2.3.3 SDS-PAGE

TALE proteins were monitored on 8 % SDS-PAGE gels. Proteins were analyzed on 8 % SDS-gels by PAGE. The resolving gel was prepared from 3808 µl MilliQ-water, 1092 µl Acrylamide

concentrate (40 %), 420 µl Tris (pH = 8.8), 66.7 µl SDS solution (10 %), 66.7 µl APS solution (10 %) and 6.67 µl TEMED. After 30 min of polymerization, the stacking gel prepared from 1000 µl MilliQ-water, 208 µl Acrylamide concentrate, 420 µl Tris (pH = 6.8), 16.7 µl SDS solution, 16.7 µl APS solution and 1.67 µl TEMED was poured onto the resolving gel. Protein samples were mixed with 10 µl of SDS-PAGE loading buffer and denatured at 95°C for 5 min. Subsequently, 15 µl of the samples were loaded onto the gel. The gel was run at 120 V for 1.5 h. Finally, after staining the gel with brilliant blue solution for about 30 to 60 min, the gel was destained and the image of the gel was documented.

6.2.3.4 Dialysis and Storage of TALE Proteins

The TALE proteins were placed into dialysis tubes (molecular weight cut off 10 kDa) and dialyzed against 1 x TALE storage buffer at 4°C. The buffer was changed several times until the concentration of imidazole from the elution buffer was diluted to a picomolar range. The purity of the TALE proteins was controlled by SDS-PAGE. Quantification of the protein concentration was performed with BCA assay (Thermo Fisher, compatible with DTT reducing agent) followed by snap freezing of 20 µl aliquots and long-term storage at -80°C.

6.2.4 Enzymatic Alkylation of DNA Duplexes by Dam Methyltransferase

Oligonucleotide pairs oSaB900/oSaF1815 and oSaB874/oSaF1815 were hybridized in 15.5 µl Dam reaction buffer by incubation mix at 95°C for 5 min. The hybridization was then allowed to room temperature for 30 min. SAM analogues were added (80 µM final concentration) and the pH was adjusted to 8.0 if necessary. After addition of 0.5 µl *E. coli* Dam enzyme (51.5 µM stock, 1.65 µg/ml) the reaction was incubated at 37 °C for 4 h in a final reaction volume of 312.5 µl. The reaction mix was then purified with a PCR cleanup kit (Qiagen) and the amount of DNA was quantified by UV absorption measurement. The reaction was checked for correct alkylation by sending samples to Metabion international AG to perform ESI-TOF mass spectrometry.

6.2.5 Primer Extension

Oligonucleotides o875 and o770 were ^{32}P-radiolabeled at the 5'-end using [gamma]-^{32}P-ATP and T4 polynucleotide kinase. The labeled oligos were purified by G-25 gel filtration. The template oligonucleotides and radiolabeled primers (oSaB900/oSaB874, oSaB899/oSaB873) were diluted to 6 µl of hybridization buffer and incubated at 95°C for 5 min. The hybridization mix (was then allowed to cool to room temperature for 30 min. 6 µl of varying concentrations of TALE protein diluted in 1 x TALE storage buffer were added to the hybridization mix and incubated for another 30 min at room temperature. 12 µl of Klenow Fragment *E. coli* DNA polymerase (30 mU, 3'-5' exo-) were added to the reaction resulting in a final concentration of 8.325 nM and 25 nM for primer and template, respectively. The reaction was incubated for 15 min, then stopped with addition of 12 µl of PAGE loading buffer and incubation at 95°C for 5 min. The samples were loaded onto a pre-run denaturing PAGE gel (9 M urea) and it was run in 1 x TBE buffer at 120 W for 1 h. Then, the gel was transferred to Whatman paper and covered with cling film. The gels were dried for a minimum of 1 h and a phosphor plate was exposed to it overnight. The next the image of the gel was collected by a phosphor imager and analyzed by ImageJ/Quantity One software.

6.2.6 Electromobility Shift Assay

Oligonucleotide pairs oSaB900/oSaF1815 and oSaB874/oSaF1815 were incubated at 95°C for 5 min in 1 x BGrK1 buffer. After hybridization of the oligos, the samples were cooled to room temperature for 30 min. 1.5 pmol of TALE protein with N-terminal GFP domain was incubated with varying concentrations of DNA (2.5 pmol – 0.5 pmol) at room temperature for 30 min and additional 30 min on ice. The reaction was carried out in 1 x BGrK1 buffer resulting in a final volume of 10 µL. 5 µl of the reaction mix were loaded onto a pre-run EMSA gel and run with 70 V for 90 min at 4 °C. The GFP-fluorescence of the TALE-proteins was recorded with a Typhoon FLA-9500 laser scanner and TALE-DNA complexes were analyzed with ImageJ software.

7. References

1. He, C. & Cole, P., Introduction to Epigenetics. *Chem Rev*. **115**, 2223–2224 (2015).

2. Waddington, C. H. Preliminary notes on the development of the wings in normal and mutant strains of Drosophila. *Proc. Natl. Acad. Sci.* **25**, 299-307 (1939).

3. Waddington, C. The epigenotype. *Int. J. Epidemiol.* **41**, 10-13 (2012).

4. Jamniczky, H. A. *et al.* Rediscovering Waddington in the post-genomic age. *BioEssays* **32**, 553–558 (2010).

5. Holliday, R. & Pugh, J. E. DNA Modification Mechanisms and Gene Activity during Development. *Source Sci. New Ser.* **187**, 226–232 (1975).

6. Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Genome Res.* **14**, 9–25 (1975).

7. Jones, P. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484-492 (2012).

8. Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. *Science* **80,** 612-616 (2010).

9. Takai, D. & Jones, P. A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 3740–5 (2002).

10. Esteller, M. Epigenetics in Cancer. *N. Engl. J. Med.* **358**, 1148–1159 (2008).

11. Rodríguez-Paredes, M. & Esteller, M. Cancer epigenetics reaches mainstream oncology. *Nat. Med.*, 330-339, (2011).

12. Kubik, G. & Summerer, D. TALEored Epigenetics: A DNA-Binding Scaffold for Programmable Epigenome Editing and Analysis. *ChemBioChem* **17**, 975-980 (2016).

13. Talbert, P. B. & Henikoff, S. Histone variants--ancient wrap artists of the epigenome. *Nat Rev Mol Cell Biol* **11**, 264–275 (2010).

14. Mendenhall, E., Williamson, K., Reyon, D. & Zou, J. Locus-specific editing of histone modifications at endogenous enhancers. *biotechnology* **31**, 1133-1136 (2013).

15. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat. Biotechnol.* **28**, 1057–1068 (2010).

16. Fischle, W. & Schwarzer, D. Probing Chromatin-modifying Enzymes with Chemical Tools. *ACS Chem. Biol.* **11**, 689–705 (2016).

17. Law, J. & Jacobsen, S. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204-220 (2010).

18. Han, D. *et al.* A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. *Mol. Cell* **63**, 711–719 (2016).

19. Kohli, R. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502***, 472-479 (2013).

20. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333***, 1300-1303*. (2011).

21. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).

22. Hu, L. *et al.* Structural insight into substrate preference for TET-mediated oxidation. *Nature* **527**, 118–22 (2015).

23. Liu, M. Y. *et al.* Mutations along a TET2 active site scaffold stall oxidation at 5-hydroxymethylcytosine. *Nat. Chem. Biol.* **13**, 181-187(2016).

24. Lu, X., Zhao, B. & He, C. TET family proteins: oxidation activity, interacting molecules, and functions in diseases. *Chem. Rev.* **115**, 2225-2239 (2015).

25. Shen, L., Song, C., He, C. & Zhang, Y. Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu. Rev.* **83**, 585-614 (2014).

26. Kriukienė, E., Liutkevičiūtė, Z. & Klimašauskas, S. 5-Hydroxymethylcytosine–the elusive epigenetic mark in mammalian DNA. *Chem. Soc. Rev.* **41**, 6916-6930 (2012).

27. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* **6**, 1049–1055 (2014).

28. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333***, 1300-1303 (2011).

29. Song, C. X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).

30. Raiber, E.-A. E. *et al.* 5-Formylcytosine alters the structure of the DNA double helix. *Nat. Struct. Mol. Biol.* **22**, 44–9 (2015).

31. Pfaffeneder, T. *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* **10**, 574–81 (2014).

32. Kubik, G. & Summerer, D. Deciphering Epigenetic Cytosine Modifications by Direct Molecular Recognition. *ACS Chem. Biol.* **10**, 1580–1589 (2015).

33. Wion, D. & Casadesús, J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* **4**, 183–192 (2006).

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

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

36. Summerer, D. N6-Methyladenine: A Potential Epigenetic Mark in Eukaryotic Genomes. *Angew. Chemie Int. Ed.* **54**, 10714-10716 (2015).

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

38. Casadesús, J. & Low, D. a. Epigenetic gene regulation in the bacterial world. *Microbiol. Mol. Biol. Rev.* **70**, 830–856 (2006).

39. Hsieh, P. Molecular mechanisms of DNA mismatch repair. *Mutat. Res. Repair* **486**, 71–87 (2001).

40. Bertani, G. & Weigle, J. Host controlled variation in bacterial viruses. *J. Bacteriol.* **62**, 113-121 (1953).

41. Boyer, H. W. DNA Restriction and Modification Mechanisms in Bacteria. *Annu. Rev. Microbiol.* **25**, 153–176 (1971).

42. Woude, M. van der, Braaten, B. & Low, D. Epigenetic phase variation of the pap operon in Escherichia coli. *Trends Microbiol.* **4**, 5-9 (1996).

43. Hernday, A., Braaten, B. & Low, D. The intricate workings of a bacterial epigenetic switch. *Adv. Syst. Biol.* **547**, 83-89 (2004).

44. Hernday, A., Krabbe, M., Braaten, B. & Low, D. Self-perpetuating epigenetic pili switches in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16470–16476 (2002).

45. Hernday, A., Braaten, B. & Low, D. The mechanism by which DNA adenine methylase and PapI activate the pap epigenetic switch. *Mol. Cell* **12**, 947-957 (2003).

46. Vanyushin, B., Tkacheva, S. & Belozersky, A. Rare bases in animal DNA. *Nature* **225, 948-949** (1970).

47. Vanyushin, B. Adenine methylation in eukaryotic DNA. *Mol. Biol.* **39**, 473-481 (2005).

48. Pfeifer, G. P. Epigenetics: An elusive DNA base in mammals. *Nature* **532**, 319-320 (2016).

49. Hattman, S., Kenny, C., Berger, L. & Pratt, K. Comparative study of DNA methylation in three unicellular eukaryotes. *J. Bact.* **135**, 1156–1157 (1978).

50. Fu, Y. *et al.* N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. *Cell* **161**, 879–892 (2015).

51. Zhang, G. *et al.* N6-methyladenine DNA modification in Drosophila. *Cell* **161**, 893–906 (2015).

52. Greer, E., Blanco, M., Gu, L., Sendinc, E. & Liu, J. DNA methylation on N 6-adenine in C. elegans. *Cell* **161**, 868-878 (2015).

53. Luo, G.-Z. *et al.* Characterization of eukaryotic DNA N6-methyladenine by a highly sensitive restriction enzyme-assisted sequencing. *Nat Commun* **7**, 1–6 (2016).

54. Koziol, M. M. J. *et al.* Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat. Struct. Mol. Biol.* **23**, 24–30 (2016).

55. Liu, J. *et al.* Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nat. Commun.* **7**, 13052 (2016).

56.   Luo, G.-Z., Blanco, M. A., Greer, E. L., He, C. & Shi, Y. DNA N(6)-methyladenine: a new epigenetic mark in eukaryotes? *Nat. Rev. Mol. Cell Biol.* **16**, 705–10 (2015).

57.   Heyn, H. & Esteller, M. An adenine code for DNA: A second life for N6-methyladenine. *Cell* **161**, 710–713 (2015).

58.   Wu, T. T. P. *et al.* DNA methylation on N6-adenine in mammalian embryonic stem cells. *Nature* **532,** 1–18 (2016).

59.   Schiffers, S. *et al.* Quantitative LC–MS Provides No Evidence for m6dA or m4dC in the Genome of Mouse Embryonic Stem Cells and Tissues. *Angew. Chemie* **5**, 1–5 (2017).

60.   Fu, Y. *et al.* FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat. Commun.* **4,** 1798 (2013).

61.   Römer, R. *et al*. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science* **318**, 645-648 (2007).

62.   Boch, J. & Bonas, U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.* **48**, 419-436 (2010).

63.   Lange, O. *et al.* Breaking the DNA-binding code of Ralstonia solanacearumTAL effectors provides new possibilities to generate plant resistance genes against bacterial wilt disease. *New Phytol.* **199**, 773-786 (2013).

64.   Lahaye, T. Programmable DNA-binding proteins from Burkholderia provide a fresh perspective on the TALE-like repeat domain. *Nucleic Acids Res.* **42,** 7436–7449 (2014).

65.   Mak, A. A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. A. J. & Stoddard, B. L. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science (80-.).* **335,** 716–9 (2012).

66.   Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326,** 1509–12 (2009).

67.   Moscou, M. & Bogdanove, A. A simple cipher governs DNA recognition by TAL effectors. *Science* **326***,* 1501-1501 (2009).

68.   Pennisi, E. The tale of the TALEs. *Science* **338,** 1408–11 (2012).

69.   Bogdanove, a. J. A. & Voytas, D. F. TAL Effectors: Customizable Proteins for DNA Targeting. *Science (80-.).* **333,** 1843–1846 (2011).

70.   Weber, E., Gruetzner, R., Werner, S., Engler, C. & Marillonnet, S. Assembly of Designer TAL Effectors by Golden Gate Cloning. *PLoS One* **6,** e19722 (2011).

71.   Sanjana, N. E. *et al.* A transcription activator-like effector toolbox for genome engineering. *Nat Protoc* **7,** 171–192 (2012).

72.   Deng, D. *et al.* Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science (80-.).* **335,** 720–723 (2012).

73.   Streubel, J., Blücher, C., Landgraf, A. & Boch, J. TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.* **30**, 593-595 (2012).

74. Doyle, E., Stoddard, B., Voytas, D. & Bogdanove, A. TAL effectors: highly adaptable phytobacterial virulence factors and readily engineered DNA-targeting proteins. *Trends Cell Biol.* **23**, 390-398 (2013).

75. Juillerat, A., Dubois, G., Valton, J., Thomas, S. & Stella, S. Comprehensive analysis of the specificity of transcription activator-like effector nucleases. *Nucleic Acids Res.***42**, 5390-5402 (2014).

76. Rogers, J., Barrera, L., Reyon, D. & Sander, J. Context influences on TALE-DNA binding revealed by quantitative profiling. *Nature* **6**, 7440, (2015).

77. Flade, S. *et al.* The N6-Position of Adenine Is a Blind Spot for TAL-Effectors That Enables Effective Binding of Methylated and Fluorophore-Labeled DNA. *ACS Chem. Biol.* (2017)

78. Gao, H., Wu, X., Chai, J. & Han, Z. Crystal structure of a TALE protein reveals an extended N-terminal DNA binding region. *Cell Res.***22**, 1716-1720 (2012).

79. Mak, A., Bradley, P. & Bogdanove, A. TAL effectors: function, structure, engineering and applications. *Curr. Opin.* **23**, 93-99 (2013).

80. Lamb, B., Mercer, A. & Barbas, C. Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Res.* **41**, 9779-9785 (2013).

81. Schreiber, T. & Bonas, U. Repeat 1 of TAL effectors affects target specificity for the base at position zero. *Nucleic Acids Res.* **42,** 7160–7169 (2014).

82. Cuculis, L., Abil, Z., Zhao, H. & Schroeder, C. M. TALE proteins search DNA using a rotationally decoupled mechanism. *Nat. Chem. Biol.* **12**, 831–837 (2016).

83. Maeder, M. L. *et al.* Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* **31,** 1137–1142 (2013).

84. Li, K. *et al.* Manipulation of prostate cancer metastasis by locus-specific modification of the CRMP4 promoter region using chimeric TALE DNA methyltransferase and demethylase. *Oncotarget* **6**, 10030- 10044 (2015).

85. Thakore, P. I. P., Black, J. B. J., Hilton, I. I. B. & Gersbach, C. A. C. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods* **13,** 127–137 (2016).

86. Kungulovski, G. & Jeltsch, A. Epigenome Editing: State of the Art, Concepts, and Perspectives. *Trends Genet.* **32,** 101–113 (2016).

87. Yang, J. *et al.* Complete decoding of TAL effectors for DNA recognition. *Cell Res.* **24,** 628–31 (2014).

88. Miller, J. C. *et al.* Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat. Methods* **12,** 465–471 (2015).

89. Valton, J., Dupuy, A., Daboussi, F., Thomas, S. & Maréchal, A. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J. Biol.* **287**, 38427-38432(2012).

90. Kubik, G., Schmidt, M., Penner, J. & Summerer, D. Programmable and Highly Resolved In Vitro Detection of 5-Methylcytosine by TALEs. *Angew. Chemie Int.* **53**, 6002-6006 (2014).

91. Bultmann, S., Morbitzer, R., Schmidt, C. & Thanisch, K. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res.* **40**, 5368-5377 (2012).

92. Maurer, S., Giess, M., Koch, O. & Summerer, D. Interrogating Key Positions of Size-Reduced TALE Repeats Reveals a Programmable Sensor of 5-Carboxylcytosine. *ACS Chem. Biol.* **11**, 3294-3299 (2016).

93. Kubik, G. & Summerer, D. Achieving single-nucleotide resolution of 5-methylcytosine detection with TALEs. *ChemBioChem* **16**, 228–231 (2014).

94. Kubik, G., Batke, S. & Summerer, D. Programmable Sensors of 5 - Hydroxymethylcytosine. *J. Am. Chem. Soc.* **137**, 2–5 (2015).

95. Stephenson, S. A.-M. & Brown, P. D. Epigenetic Influence of Dam Methylation on Gene Expression and Attachment in Uropathogenic Escherichia coli. *Front. public Heal.* **4,** 131 (2016).

96. Miller, J. *et al.* A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **29**, 143-148 (2011).

97. Cermak, T. *et al.* Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.* **39,** 7879 (2011).

98. Cong, L. Le, Zhou, R. H., Kuo, Y. C., Cunniff, M. & Zhang, F. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. *Nat. Commun.* **3,** 968 (2012).

99. Deng, D., Yan, C., Wu, J., Pan, X. & Yan, N. Revisiting the TALE repeat. *Protein Cell* **5,** 297–306 (2014).

100. Juillerat, A., Pessereau, C., Dubois, G. & Guyot, V. Optimized tuning of TALEN specificity using non-conventional RVDs. *Sci. Rep.* **5**, 8150 (2015).

101. Rathi, P., Maurer, S., Kubik, G. & Summerer, D. Isolation of Human Genomic DNA Sequences with Expanded Nucleobase Selectivity. *J. Am. Chem. Soc.* **138,** 9910–9918 (2016).

102. Maurer, S., Koch, O. & Summerer, D. Interrogating Key Positions of Size-Reduced TALE Repeats Reveals a Programmable Sensor of 5-Carboxylcytosine. *ACS chemical biology* **11**, 3294-3299 (2016).

103. Katz, D., Edwards, T., Reinke, V. & Kelly, W. A C. elegans LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* **137**, 308-320 (2009).

8. Appendix

```
ATGAGCGATAAAATTATTCACCTGACTGACGACAGTTTTGACACGGATGTACTCAAAGCGGACGGGGCGATCCTCGTCGATTT
CTGGGCAGAGTGGTGCGGTCCGTGCAAAATGATCGCCCCGATTCTGGATGAAATCGCTGACGAATATCAGGGCAAACTGACCG
TTGCAAAACTGAACATCGATCAAAACCCTGGCACTGCGCCGAAATATGGCATCCGTGGTATCCCGACTCTGCTGCTGTTCAAA
AACGGTGAAGTGGCGGCAACCAAAGTGGGTGCACTGTCTAAAGGTCAGTTGAAAGAGTTCCTCGACGCTAACCTGGCCGGTTC
TGGTTCTGGCAACGCCAGCACATGGACAGCCCAGATCTGGGTACCGTGGACTTGAGGACACTCGGTTATTCGCAACAGCAAC
AGGAGAAAATCAAGCCTAAGGTCAGGAGCACCGTCGCGCAACACCACGAGGCGCTTGTGGGGCATGGCTTCACTCATGCGCAT
ATTGTCGCGCTTTCACAGCACCCTGCGGCGCTTGGGACGGTGGCTGTCAAATACCAAGATATGATTGCGGCCCTGCCCGAAGC
CACGCACGAGGCAATTGTAGGGGTCGGTAAACAGTGGTCGGGAGCGCGAGCACTTGAGGCGCTGCTGACTGTGGCGGGTGAGC
TTAGGGGGCCTCCGCTCCAGCTCGACACCGGGCAGCTGCTGAAGATCGCGAAGAGAGGGGGAGTAACAGCGGTAGAGGCAGTG
CACGCCTGGCGCAATGCGCTCACCGGGGCCCCCCTGAACCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGG
CAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTA
TCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACT
CCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTG
CCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACAATGGCGGCAAGCAA
GCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAG
C**AACATT**GGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACC
AAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGAC
CATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCACGATGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTT
GCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCG
AAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACGGT
GGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCCGGACCAAGTGGT
GGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCC
TGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTG
CTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATCGCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGT
GCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACTCCGGACCAAGTGGTGGCTATC
GCCAGCAACATTGGCGGCAAGCAAGCGCTCGAAACGGTGCAGCGGCTGTTGCCGGTGCTGTGCCAGGACCATGGCCTGACCCG
GACCAAGTGGTGGCTATCGCCAGCAACGGTGGCGGCAAGCAAGCGCTCGAAAGCATTGTGGCCCAGCTGAGCCGGCCTGATCC
GGCGTTGGCCGCGTTGACCAACGACCATCTGCTGGAACATCACCATCACCATCACTGA
```

Appendix 1: Open reading frame of sequenced expression plasmid encoding the TALE_Pap2_NI. The light and dark blue marked codons will code for the repeat RVD NI.

```
MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRG
IPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGERQHMDSPDLGTVDLRTLGYSQQQQEKIKPKVRSTV
AQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGP
PLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPD
QVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
NIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQA
LETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRL
LPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQD
HGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASN
IGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALESIVAQLSRPDPALAALTNDHLLEHHHHHH*
```

Appendix 2: Amino acid sequence of TALE_Pap2 with RVD NI at position 7 and thioredoxin tag.

```
MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKYGIRG
IPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGERQHMDSPDLGTVDLRTLGYSQQQEKIKPKVRSTV
AQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGKQWSGARALEALLTVAGELRGP
PLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPD
QVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS
HDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQA
LETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRL
LPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQD
HGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQ
VVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASN
NGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALESIVAQLSRPDPALAALTNDHLLEHHHHHH*
```

Appendix 3: Amino acid sequence of TALE_Pap5 with RVD NI at position 7 and thioredoxin tag.

```
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYP
DHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV
YIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVT
AAGITLGMDELYKTLGYSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQ
VVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASN
IGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQAL
ETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLL
PVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDH
GLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQV
VAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNI
GGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALE
SIVAQLSRPDPALAALTNDHLLEHHHHHH*
```
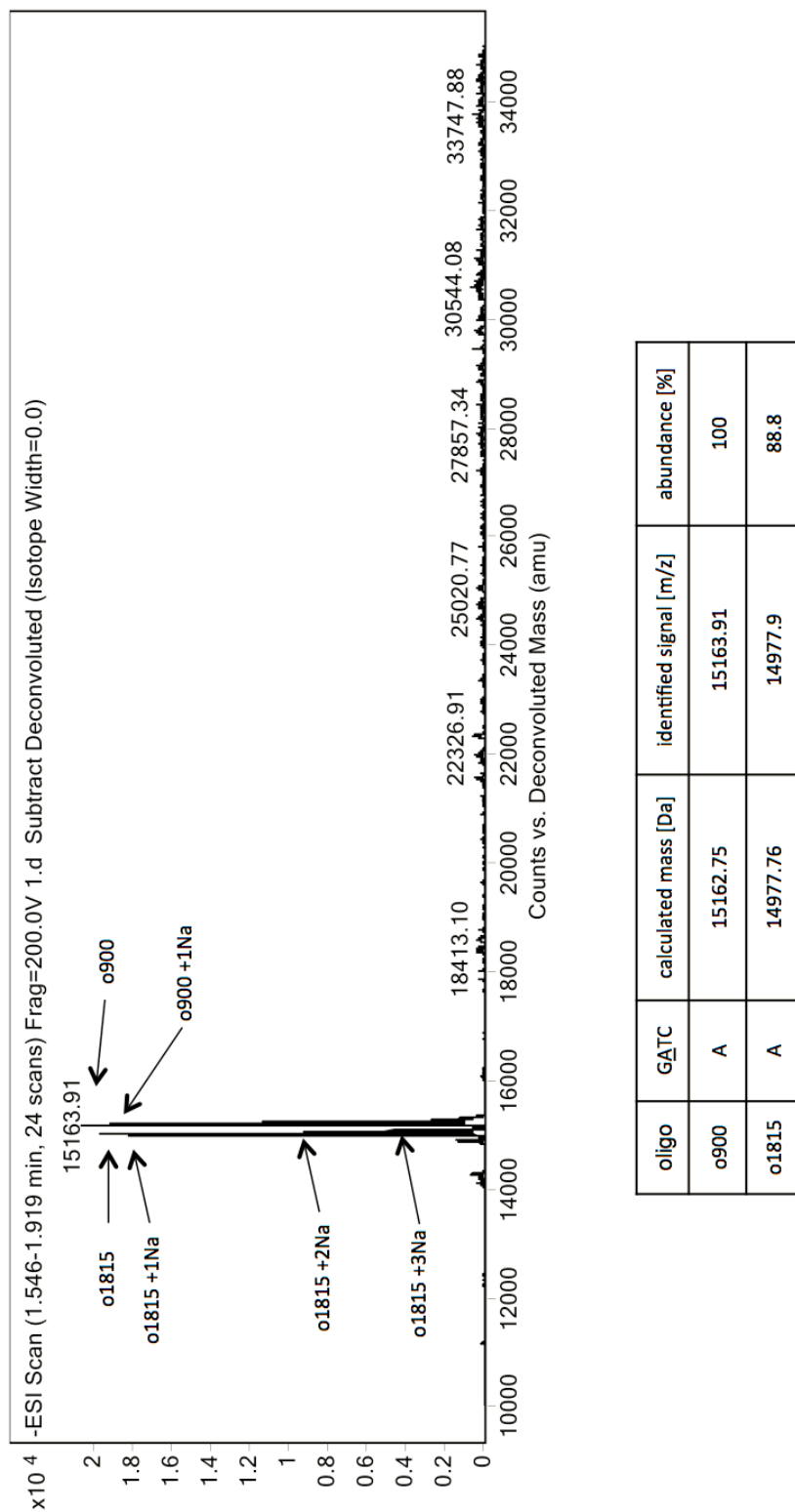
Appendix 4: Amino acid sequence of TALE_Pap2 with RVD NI at position 7 and green fluorescent protein tag.

```
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYP
DHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNV
YIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVT
AAGITLGMDELYKTLGYSQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALP
EATHEAIVGVGKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQ
VVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASN
IGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQAL
ETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLPDQVVAIASNGGGKQALETVQRLLP
VLCQDHGLTPDQVVAIASHDGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHG
LTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGGGKQALETVQRLLPVLCQDHGLTPDQVV
AIASNGGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNIGGKQALETVQRLLPVLCQDHGLTPDQVVAIASNGG
GKQALETVQRLLPVLCQDHGLTPDQVVAIASNNGGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALES
IVAQLSRPDPALAALTNDHLLEHHHHHH*
```

Appendix 5: Amino acid sequence of TALE_Pap5 with RVD NI at position 7 and green fluorescent protein tag.
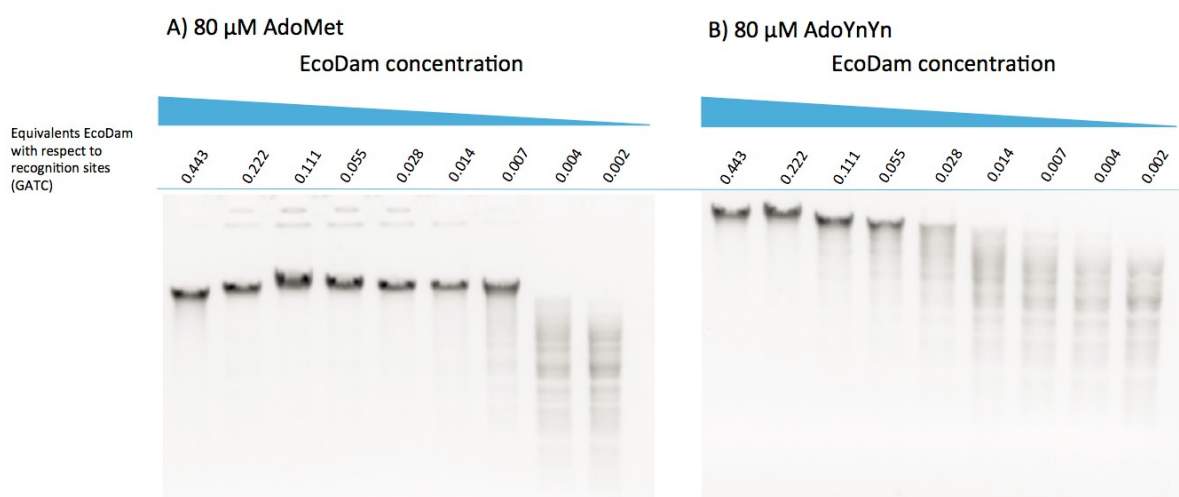
Appendix 6: SAM (AdoMet) and synthesized analogs (AdoYnYn, 1b) are accepted by the Dam enzyme (EcoDam). λ-DNA is labeled by the Dam enzyme with SAM and cofactor 1b. In the reaction mix the concentration of the labeling enzyme was varied in equivalents of enzyme in respect to GATC recognition sites. The Dam transferase is self-expressed and not purchased by a custom provider to ensure a functional, SAM-free enzyme. After the labeling reaction, the mix analyzed by restriction digest. The assay was performed by the Elmar Weinhold group, RWTH Aachen.

oligo	GA<u>T</u>C	calculated mass [Da]	identified signal [m/z]	abundance [%]
o900	A	15162.75	15163.91	100
o1815	A	14977.76	14977.9	88.8

Appendix 7: ESI-TOF MS analysis of unmodified oligonucleotide duplex of the Pap2 sequence context (oligos o900 and o1815).
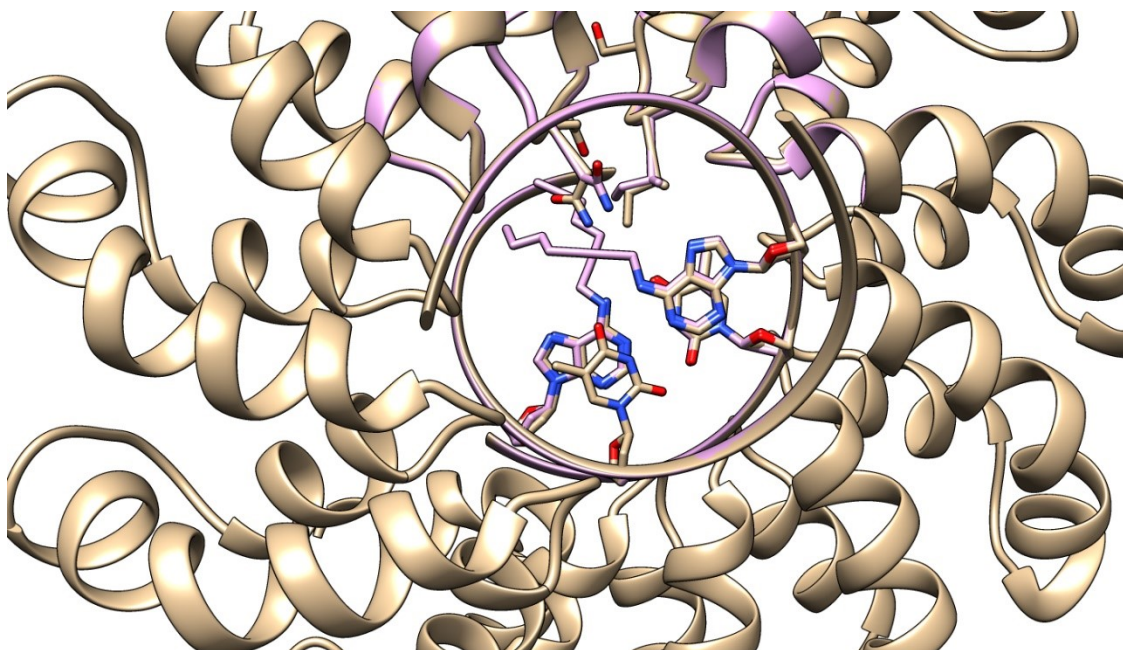
Appendix 8: Model of TALE (light brown) with the RVD NI accommodating substituent 1b.(violet) Model was generated by Julia Jasper and Oliver Koch, TU Dortmund.
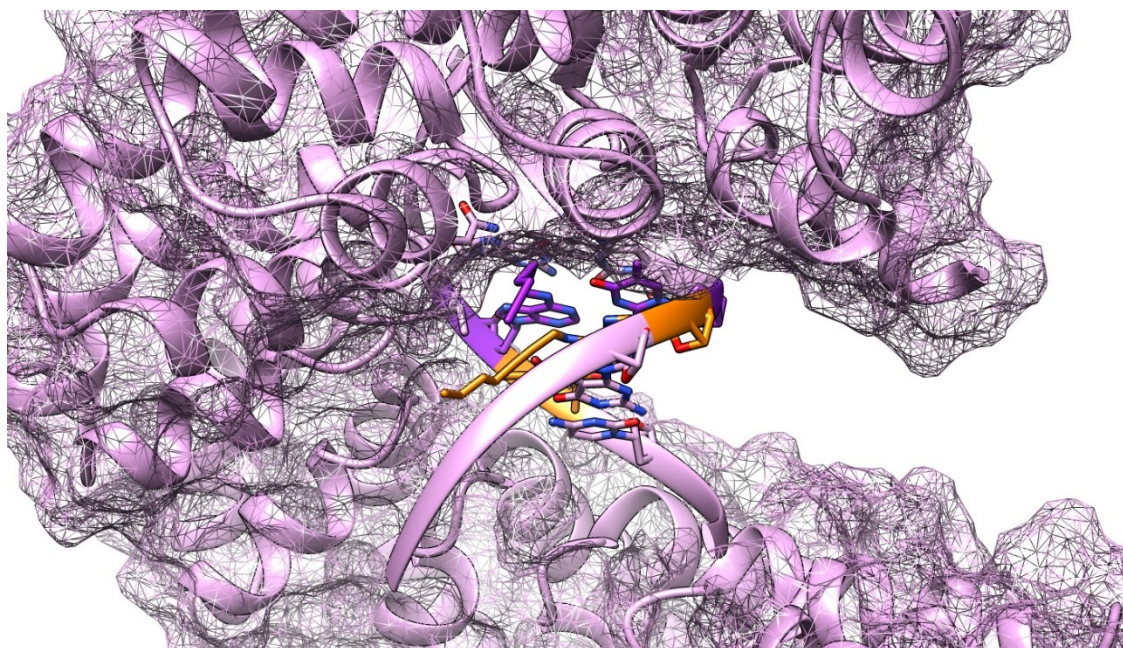
Appendix 9: Model of TALE (violet) with the RVD NI accommodating substituent 1b.(orange) Model was generated by Julia Jasper and Oliver Koch, TU Dortmund.
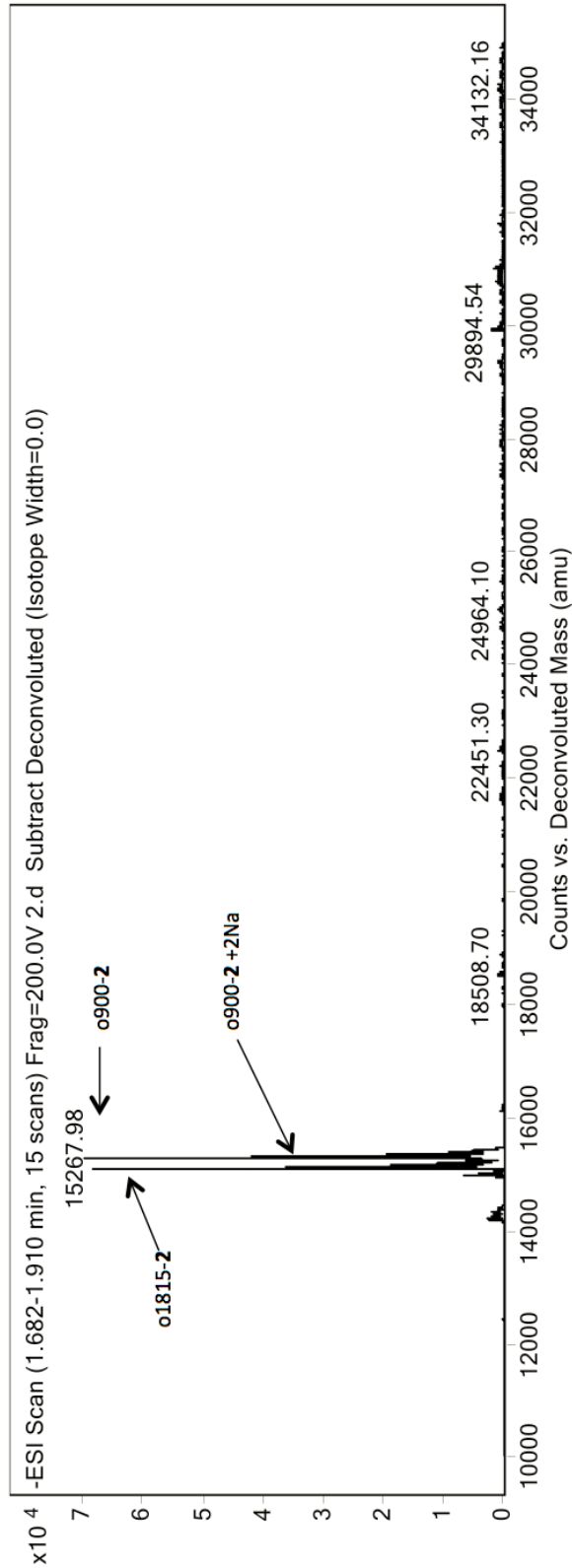
Appendix 10: ESI-TOF MS analysis of N6-modified oligonucleotide duplex of the Pap2 sequence context (oligos o900 and o1815) by *E. coli* Dam methyltransferase with cofactor 1b.

oligo	GATC	mass oligo [Da]	mass added by **2** [Da]	calculated mass [Da]	identified signal [m/z]	abundance [%]
o900	A	15162.75	105.07	15267.82	15267.98	100 %
o1815	A	14977.76	105.07	15082.82	15082.96	97.7 %

The spectrum label: -ESI Scan (1.666-2.949 min, 80 scans) Frag=200.0V 4.d Subtract Deconvoluted (Isotope Width=0.0)

Axis: Counts vs. Deconvoluted Mass (amu)

Peaks labeled: o900-3 (15898.68), o1815-3, o900, o1815

oligo	GATC	mass oligo [Da]	mass added by 3 [Da]	calculated mass [Da]	identified signal [m/z]	abundance [%]
o900	A	15162.75	735.35	15898.1	15898.68	100 %
o1815	A	14977.76	735.35	15713.11	15713.96	81.9 %

Appendix 11: ESI-TOF MS analysis of N^6-modified oligonucleotide duplex of the Pap2 sequence context (oligos o900 and o1815) by *E. coli* Dam methyltransferase with cofactor 1c.

9. Abbreviations

A	adenine
6mA	N^6-methyl adenine
6hmA	N6-hydroxymethyl adenine (in RNA)
6fA	N6-formyl adenine (in RNA)
AD	activation domain
Amp	ampicillin
ATP	adenosine triphosphate
AU	arbitrary units
BER	Base Excision Repair
bp	base pair
°C	degree Celsius
C	cytosine
5mC	5-methyl cytosine
5hmC	5-hydroxymethyl cytosine
5caC	5-carboxymethyl cytosine
5fC	5-formyl cytosine
N4mC	N^4-methyl cytosine
CpG	cytosine preceding guanine dinucleotide
CRISPR	clustered regularly interspaced short palindromic repeats
CRD	central repeat domain
CTR	C-terminal region
D	aspartic acid
Dam	deoxyadenosine methyltransferase
DNA	deoxyribonucleic acid
gDNA	genomic DNA
DMT	DNA methyl transferase
dNTP	deoxy ribonucleotide triphosphate
ds	double stranded
E	glutamic acid
EMSA	electromobility shift assay
ESI-TOF	Electrospray ionization time of flight

FLASH	Fast Ligation-based Automatable Solid-phase High-throughput
G	guanine
G	glycine
GFP	green fluorescent protein
GG	golden gate cloning
H	histidine
HD	RVD HD
His6	histidine tag
I	isoleucine
K	lysine
L	leucine
Lrp	leucine responsive protein
LR	last repeat
meDIP	methylated DNA immunoprecipitation
mESC	mouse embryonic stem cells
min	minutes
ml	milliliter
mM	millimolar
mU	milliunits
MTase	methyltransferase
N	asparagine
NE	RVD NE
NG	RVD NG
ng	nanogram
NHEJ	non-homologous end joining
NI	RVD NI
NL	RVD NL
NLS	nuclear localization sequence
nM	nanomole
nm	nanometer
NN	RVD NN
nt	nucleotide
NV	RVD NV

OD	optical density
oriC	origin of replication
Pap	pyelonephritis associated pili
PEX	primer extension
R	arginine
RM	restriction-modification
RNA	ribonucleic acid
gRNA	guiding RNA
rpm	revolutions per minute
RVD	repeat variable di-residue
S	serine
SAM	S-adenosyl methionine
SAH	S-adenosyl homocysteine
T	Thymine
TALEs	transcription-activation-like effector
TAMRA	tetramethylrhodamine
TET1	ten-eleven-translocase 1
TDG	thymine DNA glycosylase
TRX	thioredoxin
TSS	transcription starting site
UHPLC-MRM-MS/MS	ultra-high-performance liquid chromatography triple quadrupole mass spectrometry
UPEC	uropathogenic *E. coli*
V	valine
V	Volt
W	Watt
ZF	Zinc finger
µg	microgram
µl	microliter
µM	micromol

Note: chemical abbreviations can be found in the section 6.1.8.

10. Curriculum Vitae

Personal Details

Name:	Sarah Flade, M.sc.
Date of birth:	30.06.1988
Place of birth:	Neuenbürg, Germany
Nationality:	German

Education

Since 10/2013

PhD student (Dr. rer. nat.)

at the University of Konstanz and TU Dortmund

in the field of microbiological epigenetics and molecular

biology of TALE proteins. Supervision: Prof. Daniel Summerer.

10/2013 – 03/2016

Research Scholarship

of the Konstanz Research School Chemical Biology and fellow

of the Konstanz Graduate Program.

10/2011 - 08/2013

Master of Science Chemical Biology

at the Karlsruhe Institute of Technology

Master thesis: *"Structure-Activity-Relationship of the*

antimicrobial and synergistic peptides PGLa and Magainin 2".

Supervision: Prof. Anne Ulrich.

10/2008 – 07/2011

Bachelor of Science Biology

at the Karlsruhe Institute of Technology

Bachelor thesis: *"Effects of ethanol on early development in*

zebrafish". Supervision: Prof. Uwe Strähle, Universität

Heidelberg.

Publications

1. **Sarah Flade**, Julia Jasper, Mario Gieß, Maytas Juhasz, Andreas Dankers, GrzegorZ Kubik, Oliver Koch, Elmar Weinhold und Daniel Summerer, "The N6-Position of Adenine is a Blind Spot for TAL-Effectors that Enables Effective Binding of Methylated and Fluorophore-Labeled DNA", *ACS Chemical Biology*, **2017**.
   DOI: 10.1021/acschembio.7b00324.

Poster Presentations

1. Sarah Flade, Julia Jasper, Maytas Juhasz, Andreas Dankers, Grzegors Kubik, Oliver Koch, Elmar Weinhold und Daniel Summerer, *The N6-Position of Adenine is a Blind Spot for TAL-Effectors that Enables Effective Binding of Methylated and Fluorophore-Labeled DNA*, poster presented at the annual "Day of Chemistry" of the TU Dortmund, 2017, Dortmund.

2. Sarah Flade and Daniel Summerer, *Recognition and Control of Adenine-N^6-Methylation by TALEs*, poster presented at the 7th annual retreat of the Konstanz Research School Chemical Biology, 2015, Gültstein.

3. Sarah Flade and Daniel Summerer, *Recognition of N6-Methyladenine in DNA by Transcription-activation-like effectors*, Poster presented at the 6th annual retreat of the Konstanz Research School Chemical Biology, 2014, Bad Herrenalb.