

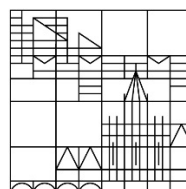
Deciphering Epigenetic Cytosine Modifications with Transcription Activator Like Effector Proteins

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

**Presented by
Grzegorz Kubik M. Sc. RWTH**

at the

Universität
Konstanz



**Faculty of Sciences
Chemistry Department**

Date of the oral examination: 22.01.2016

First referee: Prof. Dr. Daniel Summerer

Second referee: Prof. Dr. Jörg Hartig

Third referee: Prof. Dr. Valentin Wittmann

INDEX

Acknowledgements.....	VI
Abstract	VII
Zusammenfassung	VIII
Abbreviation Index.....	IX
1. Introduction.....	- 1 -
1.1 The genetic Code	- 2 -
1.2 Epigenetics.....	- 4 -
1.2.1 Detection of epigenetic cytosine modifications	- 6 -
1.2.1.1 Reactivity based approaches.....	- 7 -
1.2.1.2 Binding based approaches	- 9 -
1.2.1.3 Processive Readers	- 10 -
1.3 Programmable Sequence Selective DNA binding Proteins	- 12 -
1.3.1 Zinc Finger Proteins.....	- 12 -
1.3.2 Transcription Activator Like Effector Proteins.....	- 13 -
1.3.2.1 Structure.....	- 13 -
1.3.2.2 TALE-DNA Binding Mode	- 14 -
1.3.2.3 TALE Manufacturing.....	- 15 -
1.3.2.4 TALE Applications	- 16 -
2. Aim of the work	- 18 -
3. Results and Discussion.....	- 20 -
3.1 Detection of 5-methylcytosine with TALEs	- 21 -
3.1.1 TALE scaffold and expression	- 21 -
3.1.2 Revalidation of methylation sensitivity of TALEs	- 22 -
3.1.3 5mC Dependent <i>in vitro</i> Inhibition of DNA Synthesis by TALEs	- 23 -
3.1.4 <i>In vitro</i> 5mC detection in total DNA by TALEs	- 27 -
3.1.5 5mC detection at single nucleotide resolution with TALEs	- 31 -
3.2 Detection of 5-hydroxymethylcytosine with TALEs.....	- 34 -
3.3 Molecular Recognition of fC and caC.....	- 39 -
3.4 Sensitivity of TALE RVDs to N4-methylcytosine	- 40 -
4. Summary and Outlook	- 42 -
5. Materials and Methods.....	- 44 -

5.1 Materials.....	- 45 -
5.1.1 Lab equipment.....	- 45 -
5.1.2 Consumables.....	- 47 -
5.1.3 Chemicals, Reagents, Formulations	- 48 -
5.1.4 Enzymes.....	- 50 -
5.1.5 Kits	- 50 -
5.1.6 Buffers and Media.....	- 51 -
5.1.7 Oligonucleotides	- 53 -
5.1.8 Plasmids.....	- 56 -
5.1.9 Bacterial Strains.....	- 57 -
5.1.10 TALE Proteins.....	- 57 -
5.2 Methods.....	- 58 -
5.2.1 General Methods	- 58 -
5.2.1.1 Liquid culture.....	- 58 -
5.2.1.2 Plate culture	- 58 -
5.2.1.3 Preparation of electro competent bacteria	- 58 -
5.2.1.4. Preparation of chemical competent bacteria.....	- 59 -
5.2.1.4.1 Standard procedure	- 59 -
5.2.1.4.2 Procedure for BL21(DE3) and BL21(DE3)Gold strain	- 59 -
5.2.1.5 Transformation	- 60 -
5.2.1.5.1 By Heat shock.....	- 60 -
5.2.1.5.2 By electroporation	- 60 -
5.2.1.6 PCR Purification	- 60 -
5.2.1.7 Agarose-Gel Electrophoresis.....	- 60 -
5.2.1.8 SDS-Gel Electrophoresis.....	- 61 -
5.2.1.9 Plasmid Isolation	- 61 -
5.2.1.10 DNA Sequencing.....	- 61 -
5.2.2 Construction of TALE expression plasmids.....	- 62 -
5.2.2.1 With Non-RVD variability	- 62 -
5.2.2.2 Without Non-RVD variability by TALE assembly.....	- 62 -
5.2.3 Creation of monomer plasmids with non-canonical RVDs.....	- 63 -

5.2.4 Protein expression and Purification.....	- 63 -
5.2.5 Electromobility Shift Assays	- 64 -
5.2.6 Primer Extension Reactions.....	- 64 -
5.2.7 qPCR Experiments on synthetic oligonucleotides	- 65 -
5.2.8 Extraction of total DNA.....	- 66 -
5.2.9 Genomic 5mC-Detection Procedure using Bisulfite Sequencing	- 66 -
5.2.10 Streptavidin-Agarose bead preparation.....	- 67 -
5.2.11 Genomic 5mC-Detection Procedure using TAL-Effectors.....	- 67 -
5.2.12 Reduction of fC containing DNA.....	- 68 -
6. Supporting Figures	- 69 -
References	- 99 -
Table of Figures.....	- 106 -
Reprint Permissions.....	- 107 -

ACKNOWLEDGEMENTS

With handing in my PhD thesis I would like to acknowledge a few people for their support. Without those peoples help, I guess, I would never have been able to make it that far.

At first I would like to thank Prof. Dr. Daniel Summerer for giving me the opportunity to work on this very interesting and fruitful project and for believing in my ability to change the scientific field from chemistry to molecular biology. Furthermore, I acknowledge all the inspiring discussions we had and his support and open door and open mind during my stay in his group. Then I would like to say “thank you” to my colleague Moritz Schmidt for introducing me to many standard molecular biological techniques, the scientific discussions and his friendship. I also want to acknowledge my thesis committee members Prof. Hartig and Prof. Welte for their advices in the report meetings, and the groups of Prof. Hartig, Prof. Marx, Prof. Scheffner, Prof. Deuerling and Dr. Böttcher for access to their equipment. Furthermore I thank Prof. Wittmann for taking responsibility as third referee and Prof. Marx for chairing the oral exam. I acknowledge the Konstanz Research School Chemical Biology for providing such a beneficial research and training atmosphere and the DFG priority program SPP1623 for financial support and scientific exchange.

I also thank my friends Melanie Rettig, Christoph Malbertz, Vera Hoffmann and Preeti Rathi for proof reading, they raised many good points and helped to significantly improve my thesis. In addition I acknowledge my friends Christoph Malbertz, Dominik Limper, Lisa Locher, Daniel Rösner, Daniela John, Vera Hoffmann, Preeti Rathi, Stefan Haubenreisser and Jan Schulz-Wachler for being good friends, all the scientific and non-scientific discussions and the good times and all the fun we had. A special “thank you” is dedicated to Miriam Stock, for spending more than eight pleasant and warm years at my side, and to my Family, especially my parents Daria and Jan Kubik and my brother Kajetan, for their continuous support and inspiration, for raising me to the person I am, for all the good advices and for providing me the opportunity to unfold my interests and gather a variety of experiences.

In addition I want to acknowledge Prof. Dr. Meike Niggemann for all the support she gave me all the way through my studies and later on as well. In her group I made my first steps into research and learned to think in a scientific way, to work independently and straightforward, to take advantage of my curiosity and to stay open minded.

ABSTRACT

Epigenetic modification at the 5-position of cytosine is an important regulatory element of mammalian gene expression with essential roles in genome stability, development and disease.^[1-3] The range of cytosine modifications, long narrowed to only 5-methylcytosine (5mC), has been recently extended by the discovery of 5-hydroxymethyl-, 5-formyl- and 5-carboxylcytosine (hmC, fC and caC). These are key intermediates of the active demethylation pathway, but may also represent new epigenetic marks with individual biological roles.^[4-6] This increase in chemical complexity has created a pressing need for analytical approaches that enable unraveling of their functions, and shaped new challenges for already established methods with respect to their sensitivity and selectivity. The crucial step of any such approach is the method used for the differentiation of the cytosine 5-modifications from one another.^[7] In this study a new strategy is reported that bears the potential to simplify detection of those epigenetic marks and could be used for diagnostics. This approach employs Transcription Activator Like Effector (TALE) proteins as molecular sensors. TALEs have recently emerged as a new platform for the design of programmable DNA binding domains offering a simple one to one recognition code for any user defined DNA sequence.^[8-12] This study reports direct differentiation between single genomic cytosine (C) and 5mC *in vitro*^[13] based on the 5mC sensitivity of Repeat Variable Diresidue (RVD) **HD**.^[14,15] Furthermore, 5mC levels at single positions and little dependence on the sequence context or the position in the target sequence can be quantitatively detected.^[16] In addition, differentiation between C, 5mC and hmC has been achieved by additional employment of RVDs **NG** and **N***, that means, an expanded programmability of nucleobase recognition.^[17] Moreover, combination of the, within this study developed, assay with selective reduction of fC to hmC has also enabled the detection of fC, and the ability of RVD **HD** to discriminate between C, 5mC and N-4-methylcytosine (N4mC) has been discovered. This provides a strong basis for further reengineering of TALEs towards discrimination ability of all known DNA modifications.

ZUSAMMENFASSUNG

Epigenetische Modifizierungen von Cytosin an der 5-Position sind ein wichtiges regulatorisches Element in der Genexpression in Säugetieren und spielen eine zentrale Rolle bei der Genomstabilität, in der Entwicklungsbiologie und bei der Ausbildung von Krankheiten.^[1-3] Die Palette an Cytosin-Modifizierungen war lange nur auf 5-Methylcytosin (5mC) beschränkt, wurde jedoch kürzlich durch die Entdeckung von 5-Hydroxymethyl-, 5-Formyl- und 5-Carboxylcytosin (hmC, fC und caC) erweitert. Diese stellen wichtige Zwischenprodukte in der aktiven DNS-Demethylierung dar, können aber auch neue epigenetische Marker mit individuellen biologischen Funktionen sein.^[4-6] Diese Erweiterung der chemischen Komplexität von Cytosin-Modifizierungen führte zu der Notwendigkeit neue analytische Ansätze, zur Erforschung ihrer Funktionen, zu entwickeln, aber auch bereits etablierte Verfahren im Hinblick auf ihre Empfindlichkeit und Selektivität weiterzuentwickeln. Der entscheidende Schritt eines jeden solchen Verfahrens ist die Strategie, die für die Unterscheidung der Cytosin-5-Modifikationen voneinander verwendet wird.^[7] Diese Studie stellt eine neue Strategie vor, die das Potenzial hat die Detektion dieser epigenetischen Marker zu vereinfachen und in der Diagnostik verschiedenster Krankheiten eingesetzt werden kann. Dieser Ansatz basiert auf Transcription Activator Like Effector (TALE) Proteinen als molekulare Sensoren. TALEs wurden kürzlich als neue Plattform für programmierbare DNS-Bindungsdomänen etabliert und bieten einen einfachen 1:1 DNS-Erkennungsmodus, der es erlaubt auf jede beliebige DNS-Sequenz zu zielen. ^[8-12] Diese Studie zeigt, dass TALEs zwischen einzelnen genomischen Cytosinen (C) und 5mCs *in vitro* unterscheiden können.^[13] Dies basiert auf der 5mC Sensitivität des **HD** RVD (Repeat Variable Di-residue).^[14,15] Weiterhin können Modifizierungsniveaus an einzelnen Positionen ohne signifikante Abhängigkeit vom Sequenzkontext oder der Position in der Zielsequenz quantitativ ermittelt werden.^[16] Zusätzlich wurde die Unterscheidung zwischen C, 5mC und hmC durch den zusätzlichen Einsatz der RVDs **NG** und **N*** ermöglicht, was eine erweiterte Programmierbarkeit der Nukleobasen-Erkennung bedeutet.^[17] Weiterhin konnte der Nachweis von fC durch Kombination der, im Rahmen dieser Studie entwickelten, Methode mit selektiver Reduktion von fC zu hmC erreicht werden. Außerdem wurde gezeigt, dass RVD **HD** nicht nur zwischen C und 5mC unterscheiden kann, sondern zusätzlich eine individuelle Selektivität für N4mC aufweist, welche sich deutlich von der für C und 5mC unterscheidet. Somit stellt diese Studie eine solide Grundlage für die Weiterentwicklung von TALE Proteinen dar, um in Zukunft alle bekannten DNS-Modifizierungen erkennen und unterscheiden zu können.

ABBREVIATION INDEX

Abbreviation	Meaning
*	Amino acid deletion
5mC	5-methyl cytosine
A	Adenine
A	Alanine
aaRS	Aminoacyl tRNA synthetase
AD	Activation Domain
Amp	Ampicillin
ATP	Adenosintriphosphate
AU	Arbitrary Units
BER	Base Excision Repair
BS	Sodium bisulfite
BS-Seq	Bisulfite Sequencing
C	Cytosine
C	Cysteine
CAB-Seq	Chemical modification-Assisted Bisulfite sequencing
caC	5-carboxy cytosine
CpG	cytosine - guanine dinucleotide
C _T	Threshold cycles
D	Aspartic acid
DNA	Deoxyribonucleic Acid
DNMT	DNA Methyl Transferase
DNS	Desoxyribonukleinsäure
dNTP	Deoxy ribonucleotide triphosphate
Ds	Double stranded
E	Glutamic acid
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide
EMSA	Electro Mobility Shift Assay
fC	5-formyl cytosine
fCAB-Seq	5fC-assisted bisulfite sequencing
FD	Functional Domain
Fig.	Figure
FLASH	Fast Ligation-based Automatable Solid-phase High-throughput
G	Glycine
G	Guanine
gDNA	Genomic DNA
GFP	Green Fluorescent Protein
GG	Golden Gate

GG1	Golden Gate reaction 1
H	Histidine
His-tag	6His affinity tag
hmC	5-hydroxymethyl cytosine
Hrp-T3S	Hypersensitive response and pathogenicity Type III Secretion
I	Isoleucine
IC	Inhibitory Constant
ICA	Iterative Capped Assembly
IPD	Interpuls Duration
JBP1	J-binding protein 1
K	Lysine
KF(exo-)	Large fragment of DNA polymerase I (Klenow (exo-))
KRAB	Krüppel-associated box
KRuO ₄	Potassium perruthenate
L	Leucine
LR	Last Repeat
M	Methionine
M	Molar
MBD	Methyl-CpG-Binding Domain
MeCP2	Methyl CpG binding Protein 2
MeDIP	Methylated DNA Immunoprecipitation assays
mM	Millimolar
mRNA	Messenger ribonucleic acid
mU	Milliunit
N	Asparagine
N4mC	N4-methyl cytosine
NaBH ₄	Sodium borohydride
NGS	Next Generation Sequencing
NLS	Nuclear Localization Signal
nM	Nanomolar
nt	Nucleotide
oxBS-Seq	Oxidative Bisulfite Sequencing
P	Proline
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PD	Passive Dilution
PEX	Primer Extension
Pi	Phosphate
PPi	Diphosphate
Q	Glutamine
qPCR	Quantitative PCR

R	Arginine
redBS-Seq	Reducing Bisulfite Sequencing
RNA	Ribonucleic Acid
RVD	Repeat Variable Diresidue
S	Serine
SAH	S-adenoslyhomocysteine
SAM	S-adenosyl methionine
SDS	Sodium dodecylsulfonate
S-Fig.	Supporting figure
SMRT	Single-Molecule Real Time
spec	Spectinomycine
SRDX	EAR-repression domain
T	Thymine
T	Threonine
T3S	Type III Secretion signal
T4 BGT	T4 Phage β -Glucosyltransferase
TAB-Seq	TET-Assisted Bisulfite sequencing
TALE	Transcription Activator Like Effector
TDG	Thymine-DNA-Glycosylase
TET	Ten-Eleven Translocation dioxygenase
Tet	Tetracycline
TET1	Ten-Eleven Translocation dioxygenase 1
THF	Tetrahydrofolate
tRNA	Transfer Ribonucleic Acid
TRX	Thioredoxin
U	Uracil
V	Valine
W	Tryptophan
ZFP	Zinc Finger Protein
α -KG	α -ketoglutarate

1. INTRODUCTION

Parts of the following subchapters:

1.2 Epigenetics

1.2.1 Detection of epigenetic cytosine modification

1.2.1.1 Reactivity based approaches

1.2.1.2 Binding based approaches

1.2.1.3 Processive Readers

have been adapted with permission
from my publication:

ACS Chem. Biol., **2015**, 10 (7), pp 1580–1589.

Copyright © 2015 American Chemical Society

1.1 The genetic Code

The genetic information of an organism is stored on the deoxyribonucleic acid (DNA), coded by the sequence of the four canonical nucleobases cytosine (C), thymine (T), guanine (G) and adenine (A). (See Fig. 1A) These nucleobases are linked to deoxyribose which is phosphorylated at the 5' position, forming the monomeric units of DNA, called nucleotides. In cell, DNA is present as a dimer of a sense strand and its reverse antiparallel cognate. In this dimer, C always pairs with G, and T with A, via hydrogen bonds, so called Watson-Crick base pairing, leading to a right handed double helical structure of the DNA.^[18] (See Fig. 1B) Within this helix, neighboring base pairs interact via stacking interactions of the π -orbitals of the nucleobases, forming the inside part of the DNA. The backbone of the DNA is formed by the sugar and phosphate moieties of each nucleotide, leading to an overall negatively charged surface of the dimer with two grooves, different in size, enabling access to the nucleobases.^[19]

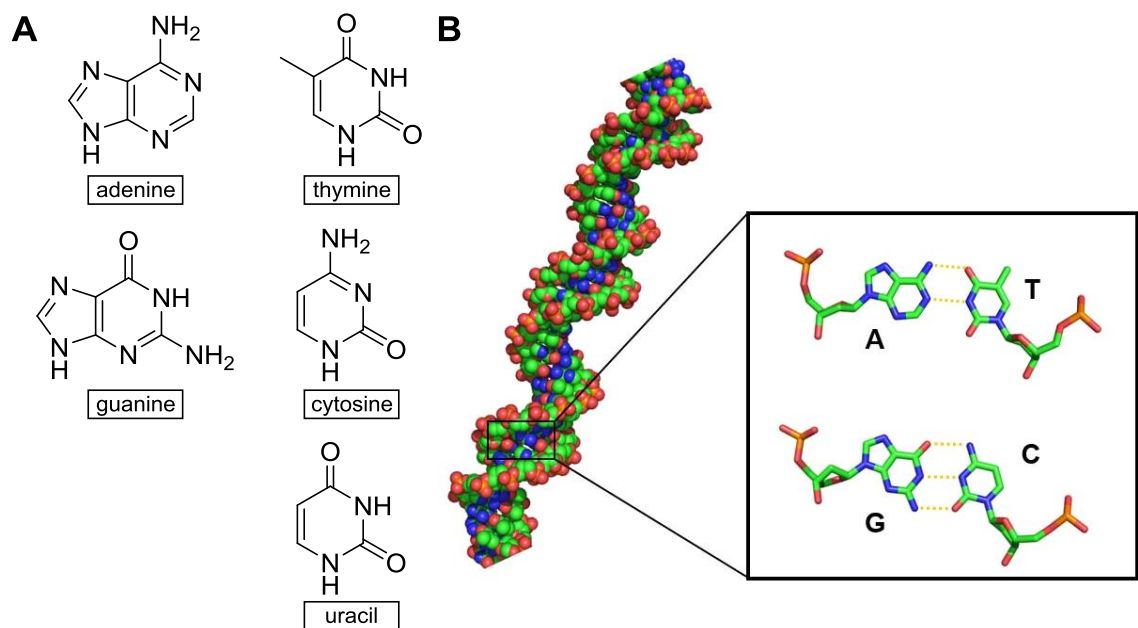


Figure 1: DNA structure and composition. **A:** Canonical nucleobases adenine (A), guanine (G), thymine (T), cytosine (C) and uracil (U) **B:** Crystal structure of DNA duplex forming a double helix. In the box, A forming Watson-Crick base pair with T, via hydrogen bonds and C with G, respectively (PDB entry: 3UGM).^[20]

The, compared to covalent bonds, weak manner of hydrogen bonds is crucial for switching DNA from its resting state to the active state by dissociation of the two strands, thus enabling biochemical processes, such as replication of the DNA prior to proliferation or transcription, to take place.^[21]

To process the information that is stored on the DNA, and bring the phenotype into being, it is first transcribed to a messenger ribonucleic acid (mRNA). RNAs are also composed of nucleotides, with the main differences of bearing ribose as carbohydrate moiety and uracil (U) replacing T. The mRNA is then transported to the ribosomes in the cytoplasm

where it is translated into a defined sequence consisting of an alphabet of in total 20 canonical amino acids. Thereby, a set of three nucleotides, the so called codon, encodes for one specific amino acid or a stop signal. In most organisms three stop codons are present and the remaining ones are coding for amino acids with different frequencies. For example only one codon out of 64 possibilities is coding for methionine (**M**) but each, arginine (**R**), leucine (**L**) and serine (**S**) are encoded by six codons. The amino acids are delivered to the aminoacyl recognition site (A site) of the ribosome by so called transfer ribonucleic acids (tRNA), which are specifically loaded with the respective amino acid by aminoacyl tRNA synthetases (aaRS), and are incorporated into the growing peptide chain.

To overcome the limitations of this 20 letters amino acid alphabet and for example to enable the addition of new chemistries to proteins, various techniques have been developed to (site specifically) introduce unnatural amino acids. One of those is the use of orthogonal aaRS/tRNA pairs like the tyrosyl system that derived from archaea *methanocaldococcus jannaschii*. This aaRS/tRNA pair was engineered to encode selectively a large variety of structurally to tyrosine related unnatural amino acids by suppressing the so called amber codon (CUA), which is a stop signal in most other organisms.^[22]

1.2 Epigenetics

Since nearly all cells within an organism have the same genome and, therefore, are carrying the same information, the question arose how it is regulated that different types of cells can develop out of one stem cell, creating a highly complex organism, and how this information is stored and provided to the next generation of cells. In 1939 WADDINGTON introduced the term “epigenetics” as “the causal interactions between genes and their products, which bring the phenotype into being”.^[23] Later on, it has been defined as reversible, heritable regulation mechanism in gene expression without any alterations in the DNA sequence.^[24,25] Today, it is evident that those additional information, that allow cells to differentiate and to respond to influences by the environment, are stored on the DNA itself in the form of epigenetic modifications of the canonical nucleobases.^[2,7] Methylation of cytosine at the 5 position is by now the best studied epigenetic mechanism. In mammals it predominantly occurs in CpG dinucleotides. It is estimated that 70 – 80 % of them are methylated throughout the genome. Non-methylated CpGs are mainly found in dense areas, so called CpG islands, close to gene promoter regions. Hence, 5-methylcytosine (5mC) is a major regulator of gene expression and plays a crucial role in developmental processes, cell differentiation and in a variety of diseases as for example different cancer types and mental diseases. Therefore, its specific distribution patterns can be used as biomarkers for diagnostics.^[1,3,7,24–26,27] The vital part of methylation as dynamic and reversible regulatory mechanism is its introduction and removal. DNA-methyl transferases (DNMT), thereby, catalyze the methylation of C using S-adenosylmethionine (SAM) as methyl donor. SAM is consumed to S-adenosylhomocysteine (SAH) which is subsequently hydrolyzed to homocysteine and adenosine.^[28] DNMTs are responsible for the *de novo* methylation of cytosine, as well as for maintenance methylation after replication, making 5mC a relatively stable and inheritable nucleobase.^[1,7] After a long time, passive dilution (PD) being the only known 5mC demethylation pathway by replication and absence of maintenance methylation, amongst other proposed mechanisms, one active demethylation pathway has been proven to take place in mammals.

The uncovering of this pathway has been spurred by the discovery of oxidized 5mC variants (5-hydroxymethylcytosine, hmC; 5-formylcytosine, fC; 5-carboxylcytosine, caC) in mammalian DNA as key intermediates of this mechanism. First 5mC is oxidized by ten-eleven translocation dioxygenase 1 (TET1) to hmC.^[29] In this reaction the enzyme transfers a dioxygen molecule on α -ketoglutarate (α -KG) and 5mC via highly reactive Fe(IV)/Fe(III) oxo species. α -ketoglutarate is converted to succinate via oxidative decarboxylation within this reaction.^[30] Iterative oxidation of hmC by TET dioxygenases leads to fC and caC which can be excised by thymine-DNA-glycosylase (TDG) forming abasic sites that can be restored to C by base excision repair (BER).^[4–6,31,32] (See Fig. 2)

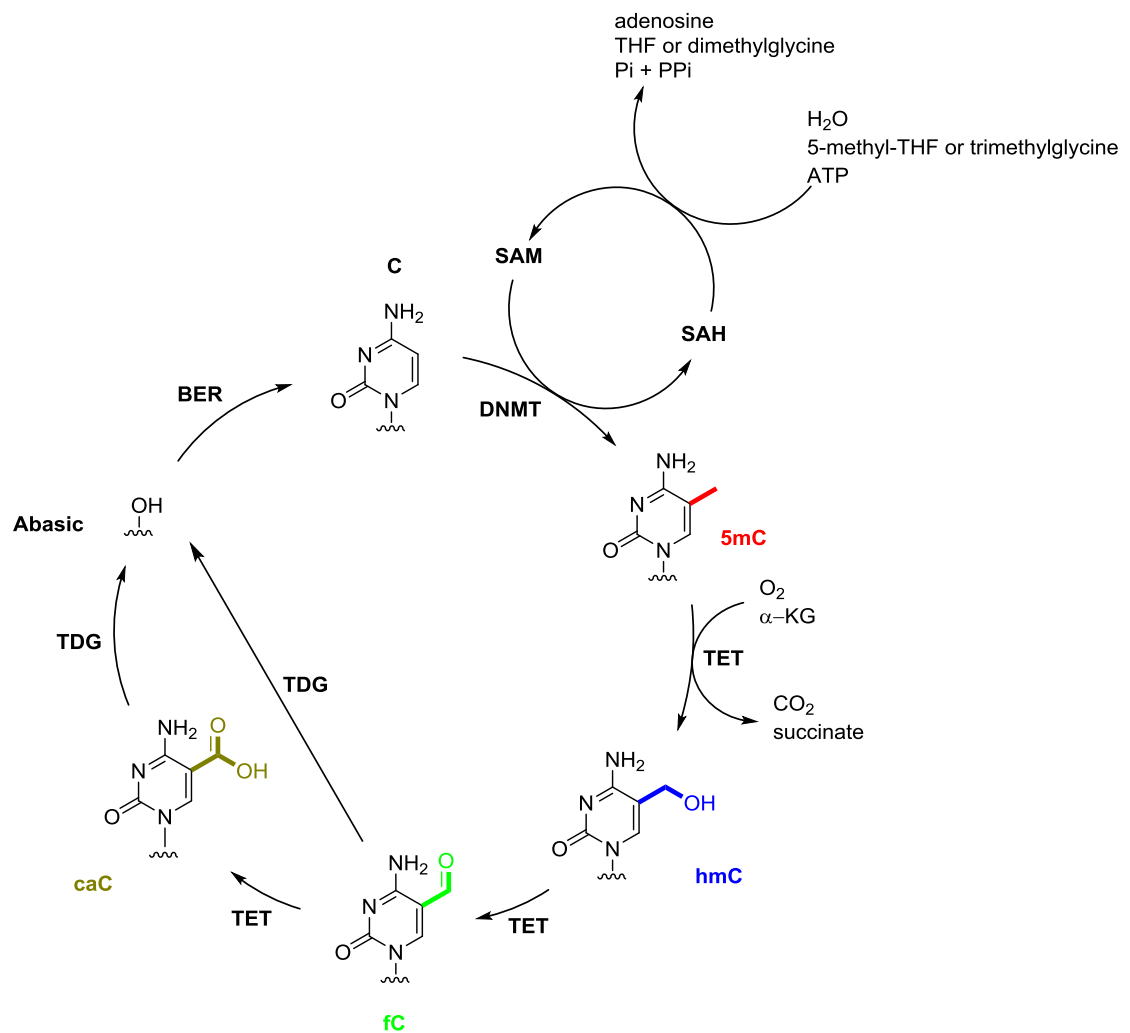


Figure 2: Active methylation and demethylation cycle of cytosine with cycle for regeneration of methyl donor. DNMT: DNA-methyl transferase; TET: ten-eleven translocation dioxygenase; TDG: Thymine-DNA-glycosylase; BER: Base Excision Repair; SAM: S-adenosylmethionine; SAH: S-adenosylhomocysteine; THF: tetrahydrofolate; α-KG: α-ketoglutarate; ATP: adenosinotriphosphate; Pi: phosphate; PPi: diphosphate; C: cytosine; 5mC: 5-methylcytosine; hmC: 5-hydroxymethylcytosine; fC: 5-formylcytosine; caC: 5-carboxylcytosine.^[7]

The genomic content of the three oxidized 5mC derivatives has been shown to be variable in different cell types, whereas hmC is the most stable and most abundant one. In several mammalian cell types it occurs in the lower double-digit percent range relative to 5mC. The occurrence of fC and caC is at a lower frequency, for fC often being one to three orders of magnitude lower than hmC, and caC occurring approximately one additional order of magnitude lower than fC.^[5-7,31,33] Despite their role in active demethylation it is elusive if oxidized C derivatives play some other roles in biochemical processes. For example, it has been shown that brain cells exhibit very high contents of hmC, whereas in several types of cancer cells the contents are low. This difference in occurrence frequency may be explained by differences in the kinetics of maintenance methylation and subsequent oxidation, depending on the cell proliferation rate. However,

profiling studies regarding the spatial distribution of oxidized C derivatives in embryonic stem cells have shown specific distribution patterns at different promoter types, transcription start sites and gene bodies.^[7,34–39]

On the one hand modifications of C at 5 position show no pronounced effect on Watson-Crick base pairing or base pair geometry, except for the case of dense occurrence of fC, but may have an impact on DNA duplex stability. caC causes an increased melting temperature, whereas the effects exhibited by the other modifications are subtle.^[40,41,42] On the other hand, the fact that each of the five C variants provides unique properties, like their ability to accept or donate hydrogen bonds and their steric demand, which are displayed to the major groove of the DNA, provides the potential for selective recognition of those modifications by former and/or so far uncharacterized reader proteins and, therefore, to serve as potential sites for protein recruitment or repulsion. Even fC and hmC, which are similar in size, exhibit differences in their conformational flexibility. Whereas the C=O of the aldehyde moiety of fC is fixed in the same plane as the heterocycle by an intramolecular hydrogen bond to the N4 amino function, hmC is found in two different conformations, providing noticeable structural differences bearing the potential for selective recognition.^[43] Indeed recent fishing studies already provided first candidate proteins that are capable of recognizing oxidized 5mC variants and may be involved in biochemical processes.^[7,41,42,44,45]

However, also post-translational modifications on histones, e.g. ubiquitination, methylation and acetylation, have been shown to have a regulatory role and can be linked to disease. These findings have led to a more wide definition of the term epigenetics, meaning any molecular signature on chromosomes to register, signal or perpetuate altered activity states, leading to stably heritable phenotype without changes in the DNA sequence.^[26,46]

1.2.1 Detection of epigenetic cytosine modifications

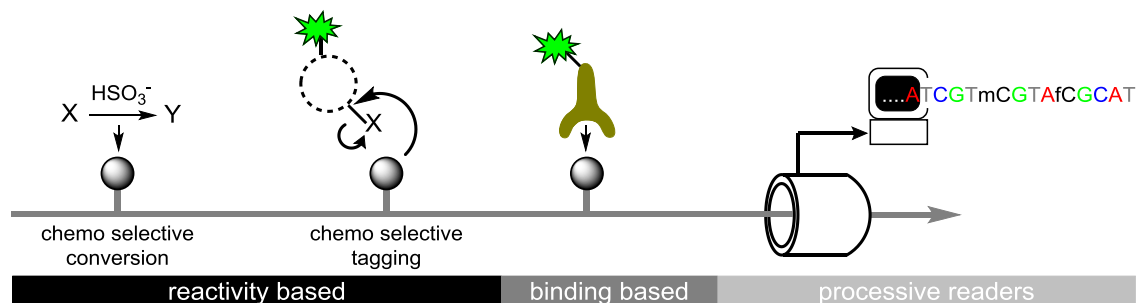


Figure 3: Overview of current approaches for the detection of epigenetic cytosine modifications.

The finding of 5mC being a valuable mark for different diseases has spurred the development of various techniques to detect this modification in genomic DNA samples. The later discovery of oxidized C derivatives has given an impulse to further develop

these methods towards full selectivity to all C variants, but it has also caused inventions of completely new approaches. In general this approaches can be divided into three categories. First, reactivity based methods, second, binding based approaches by direct molecular recognition and third, processive readers. (See Fig. 3) In addition, combinations of this methods have been used to achieve full selectivity as well. In the following sections the most important of those methods will be shortly introduced.

1.2.1.1 Reactivity based approaches

In general two different ideas have been pursued, one being the chemo selective conversion of nucleobases in one another, doing so exploiting their difference in reactivity and second the chemo selective tagging of modified C based on the difference in reactivity of the modifications towards electrophiles and nucleophiles. The chemo selective conversion approach Bisulfite Sequencing (BS-Seq) is by far the most impactful approach for the detection of 5mC. This method takes advantage of difference in deamination ability of C and 5mC, when heated with sodium bisulfite (BS). Within the reaction, C is converted to U but 5mC not to T.^[47] This way an epigenetic information is converted into a canonical nucleobase mutation and can be easily detected by common sequence analysis techniques, as polymerase chain reaction (PCR), Sanger sequencing or next-generation sequencing (NGS), with C being read as T and 5mC as C.

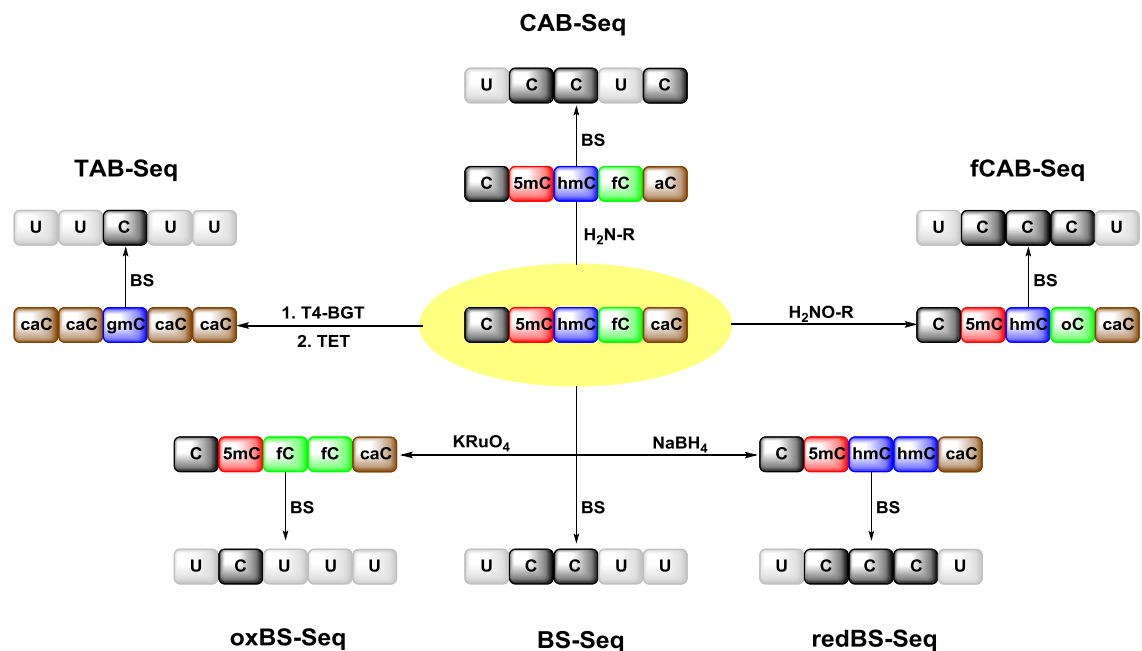


Figure 4: Strategies for achieving full selectivity with BS-Seq regarding cytosine modifications. **First step:** chemical modification of epigenetic nucleobase of interest. Reagents and catalysts stated at the arrows. **Second step:** bisulfite conversion of DNA sample leading to unaltered Watson-Crick base pairing or to altered base pairing (indicated as C in black box or U in grey box, respectively). C: cytosine; 5mC: 5-methylcytosine, hmC: 5-hydroxymethylcytosine; fC: 5-formylcytosine; caC: 5-carboxylcytosine; aC: 5-amidecytosine; oC: 5-oximecytosine; gmC: glucosylated 5-hydroxymethylcytosine; U: uracil.

With upcoming of oxidized C derivatives this methodology has been studied according its selectivity towards this new epigenetic bases. It turned out that under this conditions hmC forms a cytosine-5-methylsulfonate that exhibits unaltered base pairing and is therefore read as C.^[48] Whereas fC and caC are read as T, due to deformylation and decarboxylation, respectively, followed by deamination to U.^[6,36] To overcome this lack in selectivity different tactics have been developed. One approach is to selectively interconvert hmC and fC into each other by reduction with sodium borohydride (NaBH₄) in case of reducing bisulfite sequencing (redBS-Seq)^[35] or oxidation with potassium perruthenate (K₂RuO₄) in case of oxidative bisulfite sequencing (oxBS-Seq).^[36,49] Comparing BS-Seq, redBS-Seq and oxBS-Seq results allows to distinguish between C, 5mC, hmC and fC. Alternatively tagging supported approaches can be used. In TET-assisted bisulfite sequencing (TAB-Seq) hmC is protected by enzymatic glucosylation using T4 β -glucosyltransferase (T4 BGT).^[50] In a second step 5mC and fC are oxidized to caC by TET1 while the protected hmC stays unaffected by the enzyme. In subsequent BS-Seq only hmC is read as a C.^[51] Similar approaches are 5fC-assisted bisulfite sequencing (fCAB-Seq) where fC is protected from deamination by converting it with a substituted hydroxylamine into an oxime^[37] and chemical modification-assisted bisulfite sequencing (CAB-Seq) where amidation of the carboxy function with EDC and a primary amine results in a slower deamination enabling the detection of caC.^[52] (See Fig. 4) The data of the latter two techniques can be compared with standard BS-Seq data to differentiate between C, fC and caC. BS-Seq-based approaches bear pivotal advantages, as single nucleotide resolution, strand specificity and the quantitative level analysis of the target modification. But intrinsic drawbacks are harsh reaction conditions that lead to sample destruction of up to 99% and the comparably complicated and costly downstream analysis due to the reduction in sequence complexity.^[53]

The above described tagging approaches that have been used to provide more selectivity to BS-Seq can also be exploited for mapping studies. For this a moiety (e.g. biotin) is attached to the tag that can be selectively pulled out of the sample and the attached DNA is subsequently analyzed by next generation sequencing. This provides information about the distribution of the modification of interest in the genome. This method benefits from its technical ease and, therefore, low costs and the high sensitivity of the downstream analysis. However, the major drawback is its low resolution that is limited by the fragment size of the DNA, being normally in the range of several hundreds of base pairs.^[37,38,54,55,56]

1.2.1.2 Binding based approaches

These approaches take advantage of the sensitivity of DNA binding molecules towards epigenetic modifications. Till now only DNA binders with or without fixed sequence selectivity have been reported to be used for this kind of assay. Antibodies are established for a long time for the typing and profiling of 5mC and can be employed for genome-wide affinity enrichment of 5mC-containing DNA and subsequent analysis by e.g. quantitative polymerase chain reaction (qPCR) or next generation sequencing in methylated DNA immunoprecipitation assays (MeDIP). (See Fig. 5) The benefits of MeDIP assays are their simplicity and cost-effectivity, but they suffer from the inherent drawbacks of any non-sequence-selective affinity enrichment approach. One of those drawbacks is poor resolution, which is defined by the fragment size of the genomic DNA library, others are non-quantitative results, lacking of strand selectivity and density bias in 5mC rich sequences. Consequently it only provides information about the potentially biased relative abundance of 5mC in large genomic windows. Nevertheless, its benefits make MeDIP a widely used assay, and therefore antibodies for the full set of cytosine modifications have been developed and employed for genomic profiling.^[39,57]

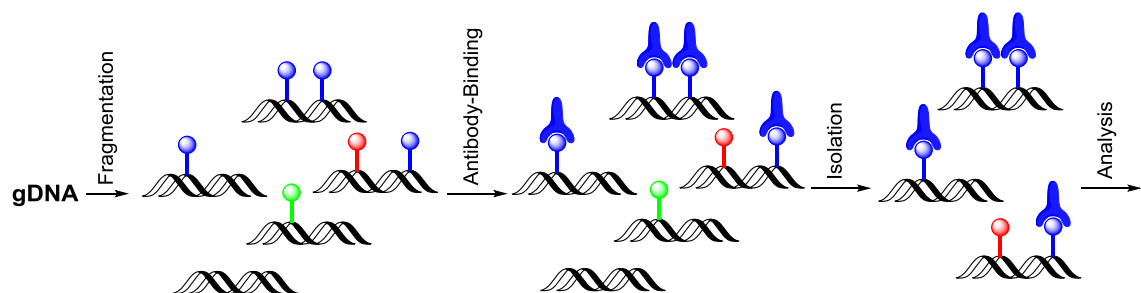


Figure 5: Principle of antibody based analysis of epigenetic nucleobase detection. First DNA sample is fragmented and next treated with antibodies, selective for modification of interest, followed by subsequent isolation of bound DNA fragments and downstream analysis.^[7]

Alternatively the methyl-CpG-binding domain (MBD) of MeCP2 and others can be used for affinity enrichment of fragments containing 5mC. This domain provides a fixed sequence selectivity towards CpG dinucleotides.^[58] It has been shown that some MBD bind selectively 5mC and others bind both, 5mC and hmC, but by now it is unknown if MBDs can be further evolved to provide full selectivity for all five cytosine nucleobases.^[45,59] Furthermore J-binding protein 1 (JBP1) from Trypanosomes has been employed for hmC analysis. In nature it binds glucosylated 5-hydroxymethyluracil but has been shown to also accommodate glucosylated hmC. This enables affinity enrichment of hmC containing DNA after glucosylation with T4 BGT.^[56] One of the major benefits of DNA binding molecules is that they are not limited to *in vitro* analysis but can also be used for *in vivo* visualization of the respective cytosine modification.^[60]

Another approach is the use of modification sensitive restriction enzymes combined with downstream analysis, those endonucleases allow for detection of epigenetic modifications in a more or less fixed sequence context. Combination with glycosylation has also been performed to gain more selectivity.^[61]

1.2.1.3 Processive Readers

Processive readers bear the big advantage that, compared to approaches based on chemical conversion and affinity enrichment, they do not depend on further downstream analysis. They directly convert the DNA Sequence and the modification status into a specific output signal. One kind of processive readers is based on nanopores, either formed by proteins or introduced into solid materials. In this set up a single pore is connecting two chambers that are separated by an insulating barrier and filled with electrolyte. Therefore, an applied electric current flows only through the pore. Charged macromolecules such as DNA can also migrate through the pore and cause a modulation of the ionic current.^[7,62] Thereby, the different structures of the four canonical nucleobases and the epigenetic nucleobases can be distinguished by the difference in the ionic current modulation they cause.^[63] Especially protein nanopores have been successfully employed for this approach, because they can be expressed with structural homogeneity and can be easily engineered with high precision on the basis of crystal structures.^[7,64] Furthermore nanopores have been improved by combining them with a phi29 DNA polymerase at the entrance to control the kinetic of DNA translocation through the pore.^[65] Till now it has been proved that nanopores provide enough selectivity to directly distinguish between the four canonical nucleobases and 5mC and hmC and also the detection of fC and caC has been reported,^[66] but this approach suffers from its high technical demand and relatively low accuracy of 92 – 98 % which could be improved by multiple reads.^[7,67]

Another approach is based on the use of DNA polymerases that offer very high read lengths. In the so called single-molecule real time (SMRT) DNA sequencing single engineered phi29 polymerases are immobilized in a so called zero-mode waveguide in zeptoliter-sized nanostructures.^[68] This makes the detection of fluorescence signals possible with very low background. During the processive synthesis of a single DNA molecule the fluorescence signals, which are generated during nucleotide binding until the labeled pentaphosphate is released, can be continuously monitored. Therefore, four phospholinked deoxynucleotides, which are individually labeled with fluorophores, are used. Dependent on the bypassed, templating nucleobase and the epigenetic modification, unique kinetic signatures of the fluorescence signals can be observed during DNA synthesis. In particular differences in the duration between two fluorescence pulses (interpulse duration, IPD) can be exploited to detect epigenetic modification.^[69] It

has been shown that 5mC and hmC lead to moderate IPD increases compared to C. Thereby, the difference between 5mC and hmC is very small and also the relatively small difference to C makes this analysis challenging. On the other hand fC and especially caC cause a much stronger IPD increase.^[70] To gain more selectivity and more bigger effects it has been shown, that the IPD can be more prominent increased by prior tagging of the modified cytosine nucleobases using selective chemistries, indicating a size-dependence of this effect.^[5,7,36,54,71] Full selectivity towards all epigenetic nucleobases without chemical tagging or chemical conversion may be some day achieved by further engineering of the used polymerase.^[72] However, this method till now suffers from its high technical demand and its restriction to *in vitro* application. Hence, the development of a new strategy that combines benefits and avoids drawbacks of the reported methods is highly favorable and may simplify diagnostics and enable highly specific applications *in vivo*.

1.3 Programmable Sequence Selective DNA binding Proteins

1.3.1 Zinc Finger Proteins

For a long time zinc fingers have been the only proteins that can bind DNA in a sequence selective manner, providing the possibility for being designed to target user defined sequences. Indeed Zinc finger proteins (ZFP) are probably the most frequent DNA-binding domains in human transcription factors. Each finger recognizes 3-4 DNA nucleotides and consists of an anti-parallel β -sheet that contains two conserved cysteines (C) and an α -helix that contains two conserved histidines (H). Those four conserved amino acids of each finger recruit a zinc ion that causes the peptide to fold into its tertiary structure. (See Fig. 6) The α -helix of each finger can bind to the DNA specifically via the major groove.^[73–75,76]

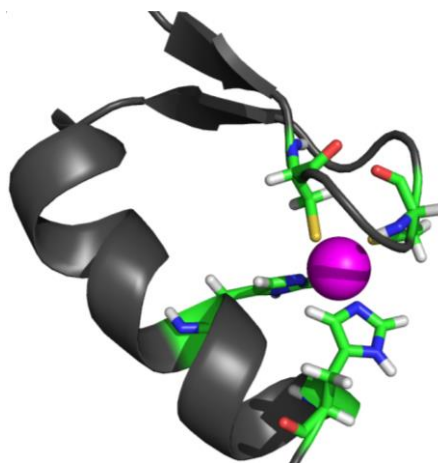


Figure 6: Crystal structure of the 10th C2H2 zinc finger of human Zinc finger protein 406 (PDB entry: 2ELM)^[77] α -helix and β -sheet shown as grey cartoon, cysteines and histidines responsible for recruitment of zinc ion (magenta sphere) shown as colored sticks.

Virtually any desired DNA sequence could be targeted by the combination of multiple fingers, each of those engineered to bind a specific triplet DNA codon, but by 2014 not for every codon a specific finger was available.^[78] Furthermore, context effects, in especially through interactions of neighboring fingers, make the design of ZFPs challenging. In addition ZFPs fused to nuclease domains which have been validated *in vitro* often fail in *in vivo* genome editing. One of the reasons for this is the lack of enough specificity in case that many similar DNA sequences are present in the genome.^[73–75,78] The approach of programmable and sequence selective DNA binding by modular proteins has been revolutionized by the discovery of Transcription Activator Like Effector (TALE) proteins that can overcome the drawbacks of ZFPs.^[78] This scaffold for DNA binding domains provides the basis for this work and will be introduced in detail in the following.

1.3.2 Transcription Activator Like Effector Proteins

Transcription activator like effector proteins have been discovered in the plant pathogenic *Xanthomonas* bacteria. Those bacteria infect their hosts' cells with this type of effector proteins, where they are localized to the nucleolus and upregulate expression of beneficial genes for bacterial virulence or downregulate defense systems.^[9] (See Fig. 7A)

1.3.2.1 Structure

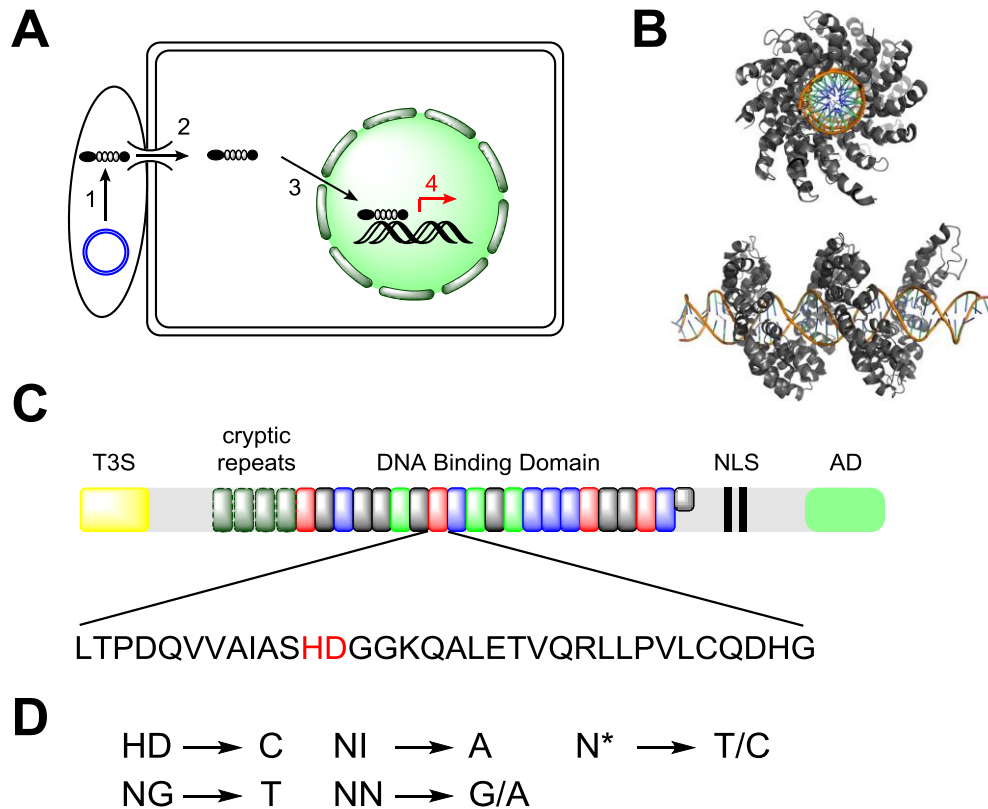


Figure 7: **A:** Mechanism of pathogeny of *Xanthomonas* bacteria 1: Expression TALE-proteins. 2: Translocation of TALEs into host cell. 3: Localization of TALEs into nucleolus. 4: Binding of TALEs to DNA target and upregulation of gene expression. **B:** Crystal structure of TALE (grey) bound to DNA. (PDB entry: 3UGM)^[20] **C:** Schematic structure of TALEs with N-terminal secretion signal (T3S), C-terminal activation domain (AD) and nuclear localization signal (NLS) and DNA binding domain including N-terminal cryptic repeats and exemplary amino acid sequence of one monomeric unit with Repeat Variable Diresidue (RVD) depicted in red letters. **D:** RVD selectivities of the four most common RVDs and the N* RVD (* means amino acid deletion).

In general TALEs consist of three major parts, the N-terminal region, the C-terminal region and the central DNA Binding domain. Plant cells are often infected with a mix of effector proteins via the hypersensitive response and pathogenicity type III secretion (Hrp-T3S) system, but also other secretion systems have been found.^[9] The signals necessary for the secretion and translocation into the plant cell are normally localized in the N-terminus of the proteins. The C-terminus bears the nuclear localization signal

(NLS), which makes the host cell to transport TALEs into the nucleolus where they can bind to their target site and cause effects on gene expression via the C-terminal activation domain (AD).^[9,20] (See Fig. 7C)

The DNA binding domain consists of concatenated highly conserved repeat units which are formed by predominantly 33-35 amino acids bundled in two α -helices and loop regions. The first, short α -helix spans amino acids 3 to 11 and is connected via a loop to the longer and bent α -helix, spanning amino acids 14 to 33. The amino acids in the loop at position 12 and 13 (so called Repeat-Variable Diresidue; RVD) are hyper variable and are responsible for nucleotide specific DNA binding via the DNA major groove, making TALEs strand specific binders winding along the DNA.^[9,20,79] (See Fig. 7B) The length of the DNA binding domain varies, being in nature normally found between 1.5 and 33.5 repeat units, the last one normally being a truncated one and therefore referred to as half repeat. However a minimum number of 6.5 repeats has been shown to be necessary for induction of targeted gene expression.^[8,9] The described so called canonical repeats are preceded by cryptic repeats showing structural similarity to the canonical repeats. This extended N-terminal DNA binding region is involved in specific and nonspecific DNA binding and is considered as the “nucleation site” for DNA binding.^[20,80]

1.3.2.2 TALE-DNA Binding Mode

The N-terminus of TALE proteins binds an initial T (referred to as position 0) of the target DNA sequence via a non-polar Van der Waals interaction of a tryptophan (**W**) located in the cryptic -1 repeat and the methyl group of T. In order to overcome this restriction to target sequences starting with a T, TALEs have been designed and evolved to being able to bind any of the four nucleobases at position 0.^[81] Having a closer look on the binding mode of the canonical repeat units, the amino acid at position 12 of each repeat contacts to the first short α -helix within the repeat unit, stabilizing the loop geometry. Whereas the amino acid at position 13 is exclusively responsible for sequence-specific DNA contacts. Other amino acids within a repeat unit are involved in sequence-unspecific interactions with the DNA backbone. The most abundant RVDs in nature are **NI**, **NN (NH)**, **NG** and **HD**, thereby recognizing A, G, T and C respectively,^[7,9,20,79] whereas the degree of specificity varies from RVD to RVD.^[11,12,20,82] But also target sequence context dependent effects on selectivity have been reported,^[83,84] as well as variability of amino acids 4 and 32 within a repeat unit has been reported to influence efficiency of TALE nucleases (TALEN) *in vivo*, suggesting an impact on DNA binding affinity.^[85]

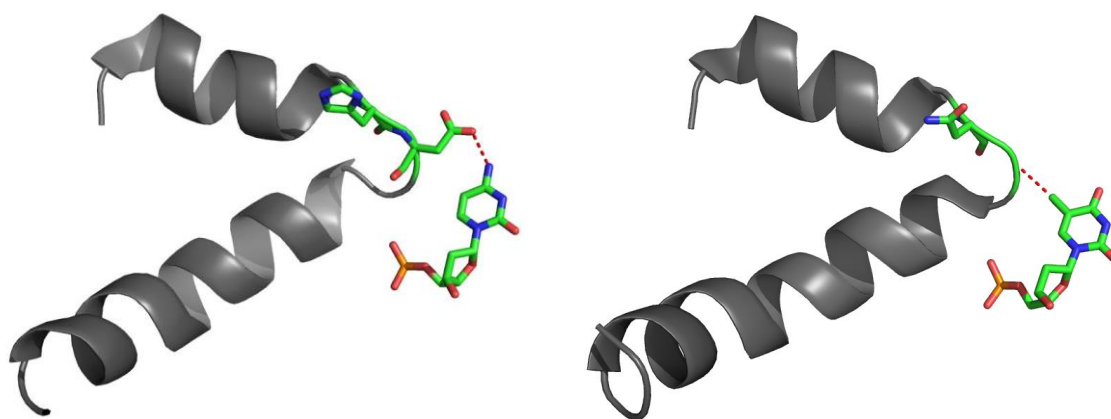


Figure 8: Crystal structure of TALE monomeric units bound to their target nucleobase. Left: RVD **HD** bound to cytosine. Hydrogen bond indicated as red dotted line. Right: RVD **NG** bound to thymine. Van der Waals interaction between 5-methyl group and glycine methylene group indicated as red dotted line. (PDB entry: 3UGM)^[20]

Studies have also revealed that the N-terminal repeat units contribute more to affinity than C-terminal repeats.^[86] Repeat units with RVD **HD** bind cytosine via a hydrogen bond of the aspartates carboxylate and the cytosine's N4 and Van der Waals interaction of aspartates methylene group and the C5 of cytosine.^[20,79] (See Fig. 8) Studies have shown that RVD **HD** is sensitive towards methylation at 5 position of cytosine. This methylation disturbs the hydrogen bond between **HD** and C by its steric demand, and therefore lowers the affinity.^[14,15,87] *In vitro* assays using fluorescence polarization measurements have shown a maximum 2-3 fold reduced signal for TALEs binding to single methylated DNA than to non-methylated. This sensitivity has been overcome by the use of size reduced RVD **N***, in which the amino acid at position 13 is deleted, and has been shown to bind C and 5mC with similar affinity.^[15,88]

Due to the structural similarity to thymine RVD **NG** is supposed to also recognize 5mC by Van der Waals interaction between the glycine methylene moiety and the 5-methylgroup. Indeed studies have shown, that the affinity to 5mC is similar to the affinity to T.^[87]

1.3.2.3 TALE Manufacturing

Due to their modularity, expression plasmids for TALEs can be easily created. For this, a large variety of methods has been developed based on either standard cloning^[89] or Golden Gate cloning,^[90–92] but also methods combining a solid phase and standard cloning like Fast Ligation-based Automatable Solid-phase High-throughput (FLASH) Assembly^[93] and Iterative Capped Assembly (ICA)^[94] have been developed.

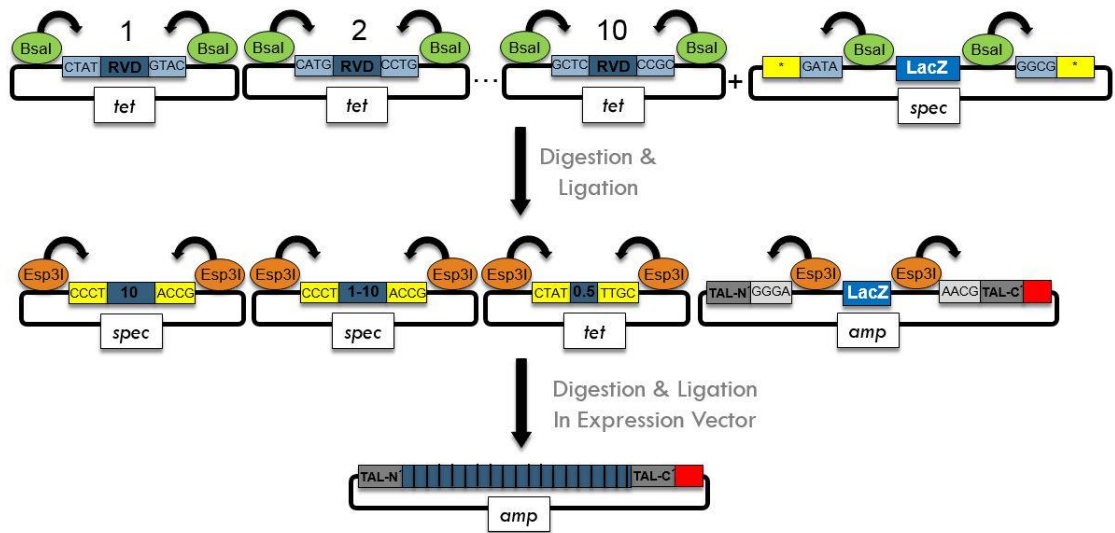


Figure 9: Principle of Golden Gate assembly of TALE expression plasmids: *Bsal* and *Esp3I*: Type IIS restriction enzymes; *tet*: tetracycline resistance; *spec*: spectinomycine resistance; *amp*: ampicillin resistance; *TAL-N*: TALE N-terminus; *TAL-C*: TALE C-terminus; *LacZ*: Reporter gene for blue/white screening; red box: optional catalytic domain.^[90]

One method that will be described in detail is based on Golden Gate (GG) cloning and was developed in the VOYTAS laboratory.^[90] In this method up to ten monomer units are assembled into an array plasmid in one pot. By restriction of the starting plasmids with type IIS restriction enzymes fragments with unique single stranded overhangs are created that assemble directly in the correct order into the array backbone and are ligated by ligase enzymes. In a second step several fragments from the array plasmids from GG1 as well as a fragment coding for the last repeat are cloned into the expression vector using the same procedure but different restriction enzyme. (See Fig. 9) In addition to the N- and C-terminal regions of TALE proteins the expression vector also encodes for additional protein domains fused to the TALE DNA binding domain. TALEs can be then directly expressed in target cells, e.g. for gene editing, or be recombinantly expressed, e.g. in *Escherichia coli* (*E. coli*).^[17,90]

1.3.2.4 TALE Applications

In addition to the natural upregulation of genes, TALEs can be exploited for many other purposes. The main application for TALEs is by now the use for gene editing. For this, TALE DNA binding domains are fused to nuclease domains, providing so called TALE nucleases (TALENs), to introduce *in vivo* double strand breaks that can be repaired by non-homologous end joining or homologous recombination, introducing deletions, insertions and/or mutations.^[83,89,92,95] But also fusions to recombinases have been developed for this use.^[96]

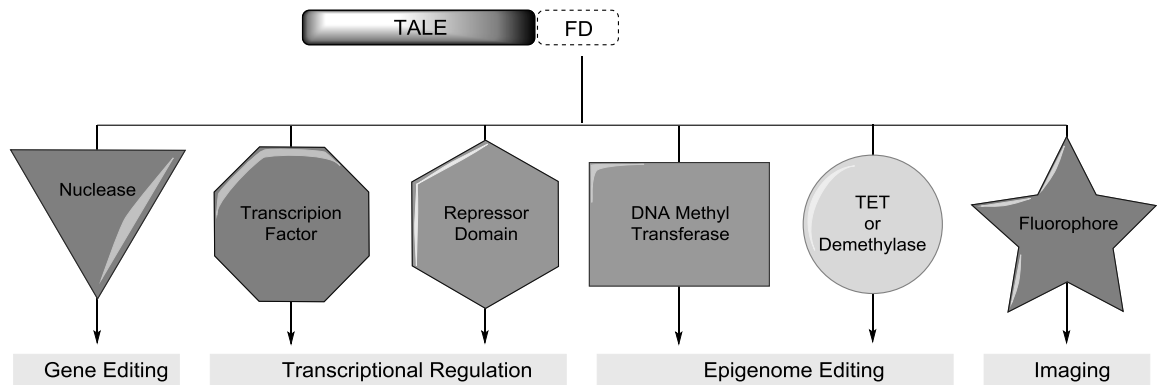


Figure 10: General overview of exploited chimeric Proteins consisting of TALE DNA Binding Domain and Functional Domain (FD).

In addition modifications on the epigenome of living cells have been performed with TALE fusion proteins.^[97] Using TET catalytic domains a site specific removal of cytosine methylation has been achieved on the one hand^[98] and on the other hand site specific methylation has been reported using fusions to DNMTs.^[99] Moreover, the alteration of histone methylation has been achieved e.g. by the use of chimeric TALE LSD1 histone demethylase proteins.^[100]

Furthermore, studies have shown that TALE DNA binding domains can also be used for suppression of reporter genes, but also fusions to repressor domains such as the Krüppel-associated box (KRAB)^[101] domain and the EAR-repression domain (SRDX)^[102] have been used in mammalian cells.

Recently fusions to fluorescent protein domains, such as Green Fluorescent Protein (GFP), have been reported for *in vivo* tracking of satellite DNA.^[103] This makes TALEs a platform for very versatile use for both *in vitro* and *in vivo* applications.

2. AIM OF THE WORK

As described above fusions of TALEs to other protein domains can be exploited for a large variety of *in vivo* applications, such as gene editing, erasing or introducing of epigenetic modifications, histone modification, gene suppression and *in vivo* visualization of genetic loci. Together with the reported sensitivity of the TALE DNA binding domain they bear the high potential for being used for such *in vivo* purposes dependent on the epigenetic status of the target locus.

The aim of this work is to take the first step towards those epigenetic modification dependent *in vivo* applications by investigation of the modification sensitivity in more detail. Since 5mC is by now the best known and investigated epigenetic mark that has been correlated to a variety of disease and biological functions, this work will focus on the sensitivity of TALE DNA binding domains towards cytosine modifications *in vitro*.

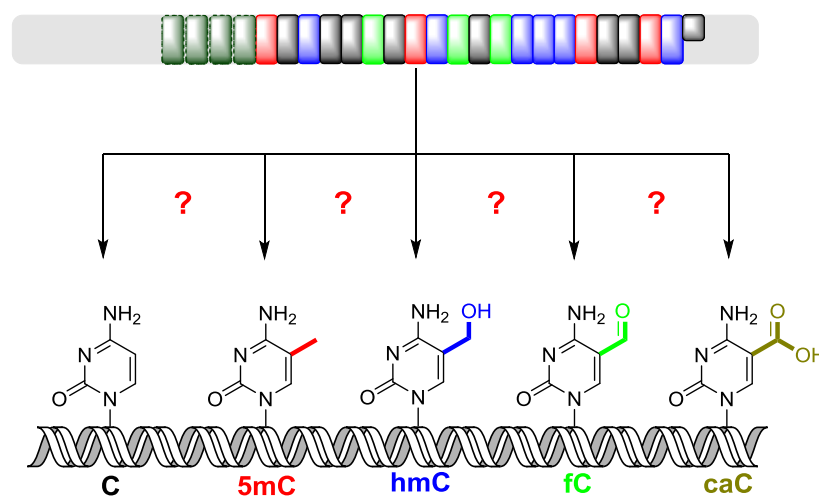


Figure 11: Investigation of discriminatory ability of TALE repeat units.

First, the reported effect of cytosine methylation at 5 position will be revalidated and investigated in more detail according to its maximum discriminatory ability, sequence context effects and effects on TALE DNA binding with respect to the position in the TALE DNA binding domain that faces the modified cytosine. Furthermore, an assay will be developed that allows the combination of TALE DNA binding with downstream analysis techniques with high throughput potential to detect cytosine modifications in real biological samples, and overcomes many of the drawbacks of up to date methods for epigenetic analysis. This assay would also provide the basis for TALEs being exploited for diagnostics of epigenetic related disease. Next, the discriminatory ability of TALEs for the differentiation between 5mC and the less abundant oxidized 5-methylcytosine nucleobases hmC, fC and caC will be investigated. Since fC and caC occur at very low levels, compared to hmC, the main focus of this work will be on the discrimination of C, 5mC and hmC, but also some attempts towards the other two cytosine variants will be made. In addition, the effect of methylation of the amino function at 4 position of cytosine will be investigated.

3. RESULTS AND DISCUSSION

The results and discussions of this section have been partly published in

Angew. Chem. Int. Ed. **2014**, *53*, 6002–6006.

© 2014 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

and

ChemBioChem **2015**, *16*, 228–231.

© 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

and

J. Am. Chem. Soc., **2015**, *137* (1), pp 2–5.

Copyright © 2015 American Chemical Society

All adaptations and reprints have been made with permission.

3.1 Detection of 5-methylcytosine with TALEs

3.1.1 TALE scaffold and expression

The TALE proteins used in this work are defined by a shortened N-terminal and C-terminal region compared to the AvrBs3 scaffold found in *Xanthomonas* bacteria, missing the translocation signal, as well as the nuclear localization signal and the transcriptional activation domain. Furthermore, they bear an N-terminal thioredoxin (TRX) tag, and a C-terminal 6His affinity tag (His-tag) for purification. (See Fig. 12A)

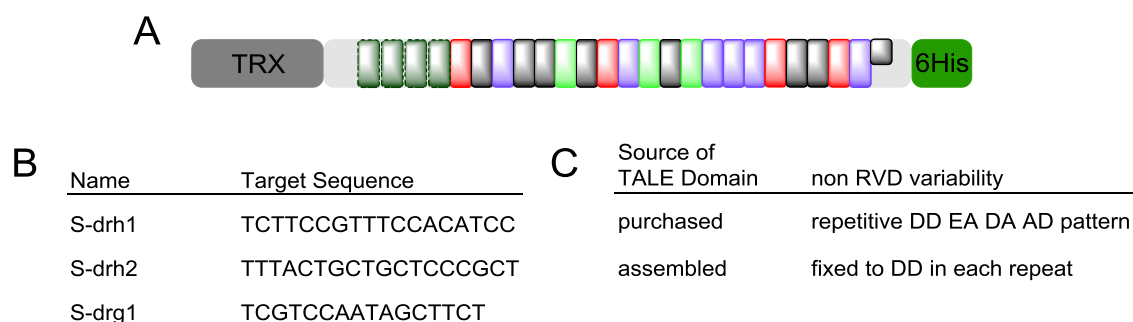


Figure 12: Illustration of TALE scaffold. **A:** General architecture of used TALEs with N-terminal thioredoxin (TRX) and C-terminal 6 histidine (6His) tag. **B:** Target sequences of the first three investigated TALEs. **C:** Amino acid pattern in the repeat units at positions 4 (first letter) and 32 (second letter) dependent on the source of the TALE DNA-binding domain.

For investigation of 5mC discriminatory ability of TALEs, three sequences in the genes *hey2* and *gria3a*, respectively, of zebrafish (*danio rerio*) were chosen (See figure 12B), and have already been described in literature.^[89] TALE expression plasmids were created by amplifying each of the DNA binding domains and a part of the TALE N- and C-termini by polymerase chain reaction (PCR) from a commercially available mammalian shuttle plasmid. The His-tag was introduced by the reverse primer. Furthermore, two restriction sites have been introduced that were used to clone the fragment into the bacterial expression vector, providing the N-terminal TRX-tag, as well as a T7 promoter, an inducible lacO repressor sequence and ampicillin resistance. The proteins have been expressed in *E. coli* and purified via Ni-NTA immobilization. TALE T-drh101 that is targeting Sequence S-drh1 bears a **NG** last repeat unit, which means that there is a misbinding at the last position of the target sequence.

In contrast to all later on with the VOYAS assembly tool box created TALEs, the TALEs T-drh101, T-drh201 and T-drg101 provide non RVD variability at the amino acid positions 4 and 32 within the repeat units. (See figure 12C)

3.1.2 Revalidation of methylation sensitivity of TALEs

To revalidate the cytosine methylation sensitivity of RVDs **HD** and **NG**, TALE-drh101 was incubated with four different synthetic double stranded (ds) ³²P-labelled DNA fragments bearing variations in the TALE target sequence S-drh1 and the binding was monitored by electro mobility shift assay (EMSA). Binding to its unmodified target (Et-drh101) was observed for TALE-drh101 at a 10:1 Protein to DNA ratio, but no binding has been detected when six Cs in this sequence have been replaced by 5mC (Et-drh102), proving the reported sensitivity of **HD** to cytosine methylation. To investigate the performance of RVD **NG** in discrimination between C and 5mC, six Ts of the target sequence have been exchanged by either cytosine (Et-drh103) or 5-methylcytosine (Et-drh104). For both, no binding of TALE-drh101 was observed in the range of Protein to DNA ratios at which binding to Et-drh101 was observed, suggesting a lower affinity of RVD **NG** to 5mC than to T. (See figure 13) This lower affinity is not in agreement with literature data, but in this case six instead of three positions have been changed to 5mC and also sequence context dependent effects cannot be ruled out.

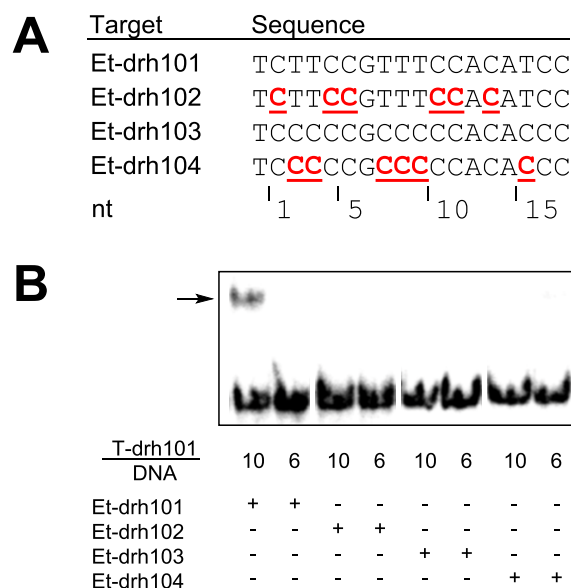


Figure 13: EMSA to validate 5mC sensitivity of RVDs **HD** and **NG**. **A:** Target sequences in oligo nucleotides used for EMSA. 5mC depicted in red in bold and underlined. **B:** EMSA with different TALE to DNA ratios. TALE-DNA complex marked with an arrow.^[13]

After proving the ability of RVD **HD** to distinguish between C and 5mC, the maximum discriminatory potential of TALEs had to be explored. Since EMSAs are not a suitable analysis method with respect to real genomic samples, further investigations have been carried out using an assay set up that can be coupled with powerful downstream analysis techniques.

3.1.3 5mC Dependent *in vitro* Inhibition of DNA Synthesis by TALEs

The aim was to develop a biochemical assay that produces a signal or not dependent on the binding of TALEs to DNA. This signal could then be detected by downstream analysis techniques. Since DNA can be easily analyzed by a variety of long established methods like e.g. quantitative polymerase chain reaction or next generation sequencing, it was investigated if TALE binding can be coupled to *in vitro* DNA synthesis.

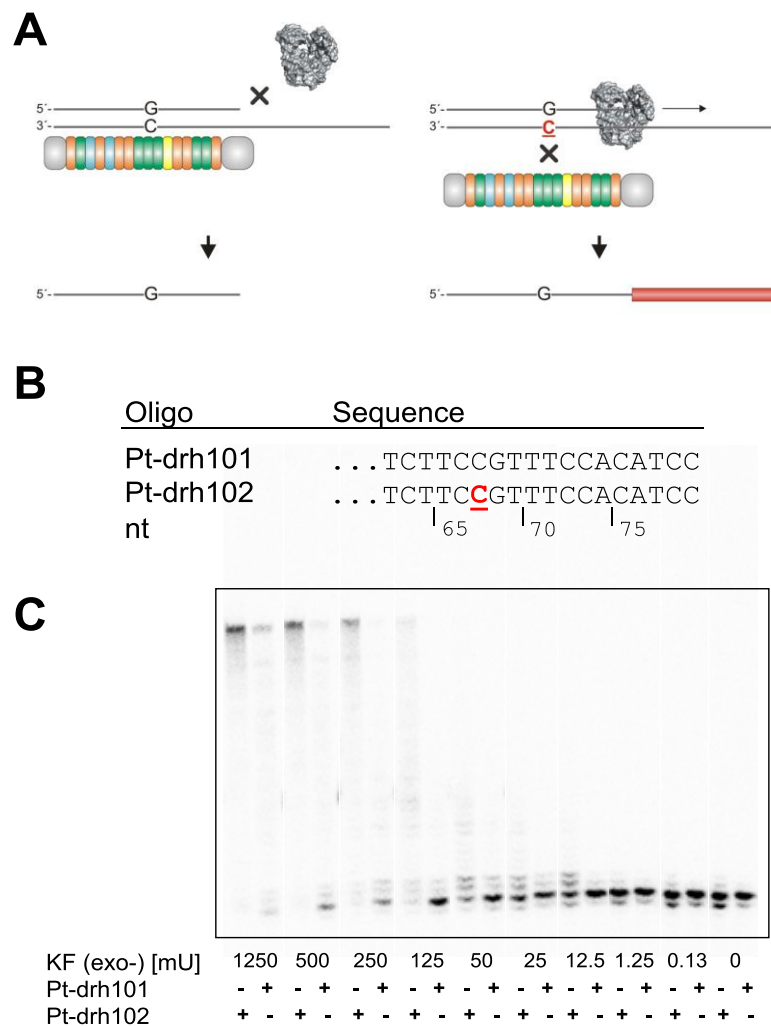


Figure 14: Primer extension assay for 5mC detection. **A:** General principle of Polymerase (grey protein surface) inhibition by TALEs (colored cartoon) in case of non-methylated cytosine opposite of RVD HD but not in case of 5mC. **B:** Target sequences in Primer extension templates, 5mC depicted in bold red and underlined. **C:** Screening for optimal KF(exo-) amount in primer extension reaction for T-drh101.^[13]

The general idea was to inhibit a DNA polymerase dependent on the methylation status of the DNA. The newly synthesized DNA strand could then serve as signal in downstream analysis. (See Fig. 14A) For this, a 18 nt long reverse complement 5'-³²P-labelled primer (p-drh101) was hybridized to synthetic 79 nt DNA fragments of zebrafish *hey2* gene bearing the target sequence S-drh1 at the 3' end, with or without

single cytosine methylation and was incubated with a 149 fold excess of T-drh101. Next the large fragment of DNA polymerase I (Klenow (exo-); KF(exo-)) at varying amounts and dNTPs were added and Primer extension was performed for 15 minutes at ambient temperature. Afterwards, the reaction was stopped and the samples were denatured and analyzed by denaturing polyacrylamide gel electrophoresis (PAGE). No extension product was observed at 125 mU KF(exo-) for non-methylated DNA but significant primer extension was observed for methylated DNA. (See Fig. 14C) This proved that TALEs are able to inhibit a DNA polymerase, when bound to a primer-template duplex and remain their discriminatory ability. In Addition no inhibition of KF(exo-) was observed when T-drh101 was absent (See S-Fig. S1), showing that primer extension is not influenced by cytosine methylation itself.

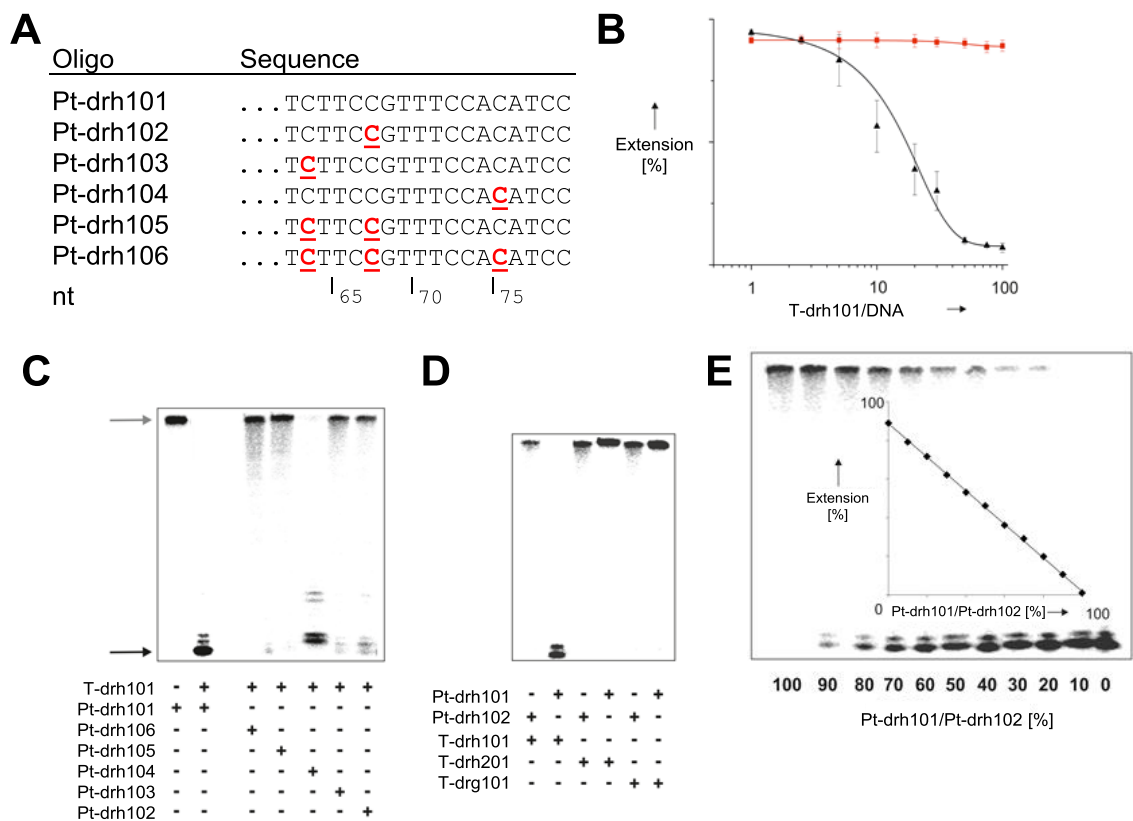


Figure 15: In detail investigation of 5mC detection with T-drh101. **A:** Target sequences in Primer extension templates, 5mC depicted in bold red and underlined. **B:** Dependence of KF(exo-) inhibition on TALE concentration. Experiments with varying T-drh101/target DNA ratios. Extension product/total lane ratios were plotted against T-drh101 concentration. Pt-drh101: black triangles, Pt-drh102: red squares. **C:** Dependence of KF(exo-) inhibition on the number and position of 5mC. PAGE analysis of primer extension reactions containing 125 mU KF(exo-), 8.3 nM primer-template with 5mC at different positions in presence or absence of 833 nM T-drh101. Extended/not extended primers are marked with a grey and black arrow. **D:** Sequence selectivity of KF(exo-) inhibition. Experiments were conducted with three different TALEs targeting sequences present (T-drh101) or absent (T-drh201, T-drg101) from Pt-drh101, comparing methylated and non-methylated DNA. **E:** Analysis of 5mC level at a single position. Experiments were performed with mixes of Pt-drh101 and Pt-drh102. Insert shows linear regression of extension product plotted against Pt-drh101 to Pt-drh102 ratio.^[13]

Next, this inhibitory effect was investigated in more detail. For this, T-drh101 concentration was varied in the primer extension (PEX) and IC_{50} value for non-methylated DNA was determined ($IC_{50} = 83 \pm 8 \text{ nM @ } 125 \text{ mU KF(exo-)}$) by dose response curve fit. Since no significant effect of TALE-drh101 on DNA synthesis in case of methylated DNA has been observed no IC_{50} value has been determined. (See Fig. 15B) Next the effect of multiple methylations and of the position of methylation within the target sequence was investigated. For this, templates with different methylation patterns have been hybridized with primer p-drh101 and PEX assay has been performed as described above with T-drh101 to DNA ratio of 100:1. In templates Pt-drh103 and Pt-drh104 cytosines at position 1 and 13, respectively, have been exchanged by 5mC. Templates Pt-drh105 and Pt-drh106 were bearing two or three 5mC at positions 1, 5 and 1, 5, 13, respectively. As expected inhibition of KF(exo-) was not observed in case of multiple methylations, so TALE binding was inhibited. Interestingly the position of the methylation in the target sequence has been shown to have an impact on TALE binding. When it is opposite the C-terminal repeat units the binding is much less disturbed than in case of methylation opposite repeat units in the central or N-terminal part of the DNA binding domain, indicating a polarity of TALE-binding, in which the N-terminus contributes more than the C-terminus, as it was already reported by MECKLER *et al.*^[86] (See Fig. 15C) But also a reflection of the large distance to the KF(exo-)-binding site cannot be ruled out.^[13]

To investigate the specificity of TALEs to their desired target sequence, cross selectivity has been tested by conducting PEX assay as above on templates Pt-drh101 and Pt-drh102 using TALEs T-drh101, T-drh201 and T-drg101. Only T-drh101 inhibited primer extension in case of non-methylated target sequence (See Fig. 15D), indicating high specificity of TALEs to their target in agreement with literature data.^[83]

An ideal method for the detection of 5mC should not only provide information about the modification status, but also about the modification level. This is necessary due to the fact that biological samples in general consist of a mixture of cells, e.g. blood samples. For validation of this method to being able to also provide level information PEX as above has been conducted on mixtures of Pt-drh101 and Pt-drh102 with varying ratios, simulating methylation levels from 0 % to 100 %. As shown in figure 15E differences in methylation level of just 10 % can be sufficiently detected and correlate in a strictly linear manner.

Discrimination ability for 5mC detection has also been tested for TALEs T-drh201 and T-drg101 in sequences S-drh2 and S-drg1, respectively, to investigate the application scope of this method. For this, primer extensions as above have been performed on the cognate templates without 5mC or with single 5mC at seven and four scattered positions, respectively. Only one single methylation opposite a **HD** RVD very close to the

C-terminus of T-drh201 did not cause significant inhibition of TALE binding. Even though some sequence context dependent effects can be observed regarding the perturbation ability of methylation on TALE-binding, 100 % of the investigated methylations opposite of N-terminal or central repeat units have been shown to significantly disrupt polymerase inhibition. (See Fig. 16B/C) This indicates a broad applicability of TALEs for 5mC detection. It should be mentioned that PEX conditions have not been optimized for the single investigated methylation positions, providing some potential for further improvement in this sequence contexts.

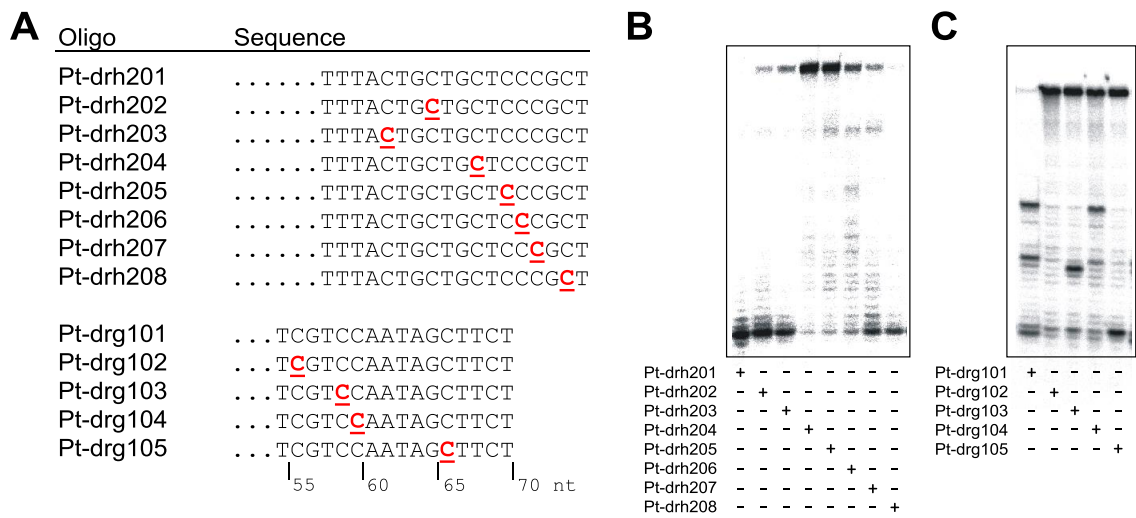


Figure 16: Investigation of 5mC detection ability of TALEs T-drh201 and T-drg101. **A:** Target sequences in Primer extension templates, 5mC depicted in bold red and underlined. **B/C:** Dependence of KF(exo-) inhibition on the position of 5mC. PAGE analysis of primer extension reactions containing 500 mU (**B**) or 25 mU (**C**) KF(exo-), 8.3 nM primer-template with 5mC at different positions in presence of 416 nM T-drh201 (**B**) or 833 nM T-drg101 (**C**).^[13]

Furthermore, IC₅₀ values for T-drh201 and T-drg101 and the cognate non-methylated DNA templates have been determined, being 24 ± 7 nM @ 500 mU KF(exo-) and 67 ± 4 nM @ 25 mU KF(exo-), respectively. This difference in affinity of the three TALEs could be correlated to the length of the bound target sequence. T-drh201 binds an 18 nt sequence and provides the highest inhibitory effect on KF(exo-), T-drh101, designed to bind also an 18 nt sequence, but due to the above mentioned mismatch at the last repeat (LR) is the effective target sequence length of 17 nt possibly reflected by the medium affinity. T-drg101, the shortest TALE, provides also the lowest inhibitory ability. But also an influence by the sequence context, in especially by the composition of the Target sequence, cannot be ruled out and should be considered due to the different modes RVDs bind their cognate nucleobase.

Finally, to test if the chosen signaling strategy would provide high sensitivity, reading out primer extension products by qPCR was investigated. Primer extension reactions were repeated with templates Pt-drh101 and Pt-drh102 as above using a 5'-biotin labeled

the reaction mixtures up to a viscosity limit of a 3×10^4 -fold mass excess over primer during the hybridization step. Above 3×10^2 , no full-length primer extension product was obtained, suggesting masking of KF(exo-) activity. The experiment was repeated with a 3×10^4 -fold excess of salmon sperm DNA and increasing amounts of KF(exo-). From an amount of 5 units per reaction, both primer extension and high 5mC-discrimination were fully restored. A similar result was achieved by increasing the primer extension time. (See Fig. 17)

This high robustness towards the presence of off target DNA offered promising perspectives for a direct application to complex biological samples. Next, total DNA from zebrafish (*danio rerio*) was extracted and the TALE target sequences and their surrounding have been determined to be completely non-methylated by bisulfite sequencing (See S-Fig. S7) and identity of the *hey2* locus has been verified by amplification of a 541 bp fragment and Sanger-sequencing. (See S-Fig. S9) The sequence matched Genbank entry NM_131622.2 with a single point mutation at position 1168, outside of the relevant region of later experiments. A part of this gDNA was then enzymatically methylated with CpG methyltransferase and methylation levels at positions of interest were confirmed to be 100 %. (See S-Fig. S4)^[13] Next, the denaturing washing conditions of streptavidin bound DNA were optimized because the initial approach with sodium hydroxide led to a dramatic loss of DNA which is unsuitable for the application with real genomic DNA (gDNA) due to much lower concentration and therefor lower amounts.

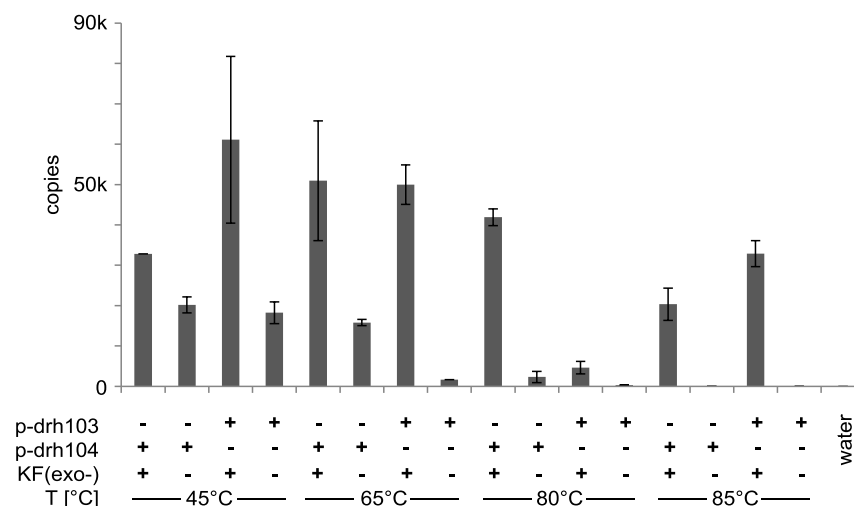


Figure 18: Optimization of bead washing conditions with heat denaturation of DNA hybrids.

It was tested to remove the template strand by incubation at increased temperature in a non-denaturing buffer. Furthermore two different primer lengths have been tested to avoid effects of the tetra ethylene glycol linked biotin moiety on the TALE binding but

Finally non-methylated and methylated genomic DNA were independently hybridized to, via tetra ethylene glycol linker 5' biotinylated, primer p-drh302. Then, primer extension was conducted as above and stopped by addition of ethylene diamine tetra acetate (EDTA) solution. Biotinylated DNA was immobilized on streptavidin beads and non-biotinylated DNA was removed using the above described optimized washing conditions and dried beads were analyzed by qPCR using primers that bind upstream of the S-drh3 target sequence in the desired extension product.

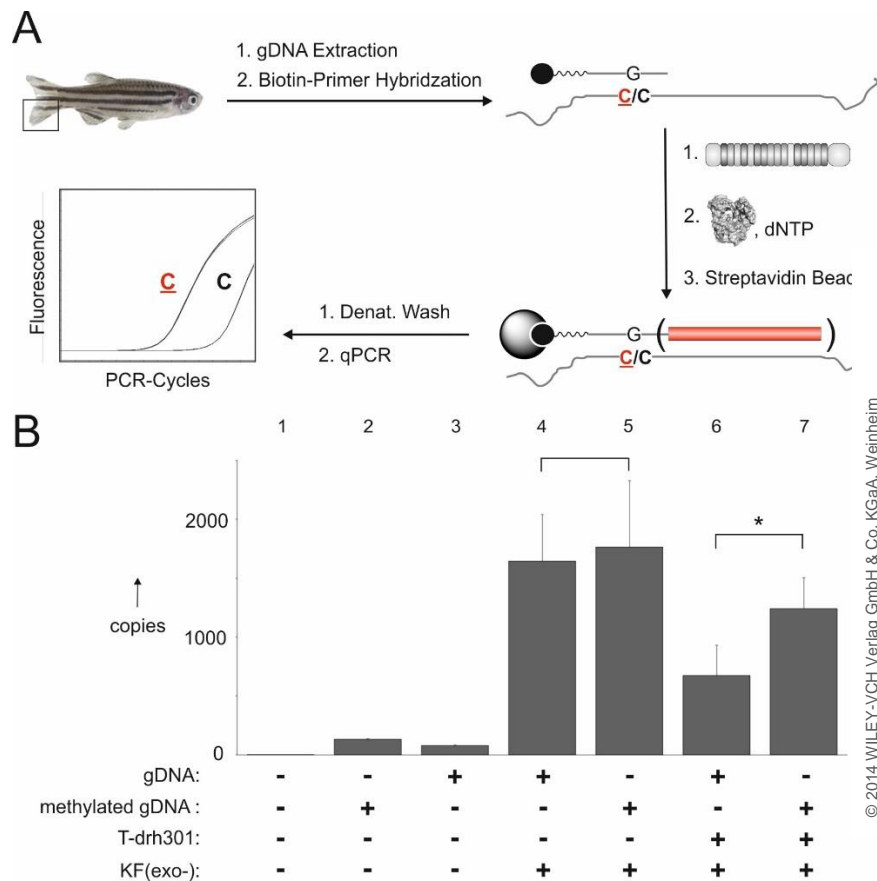


Figure 20: Detection of single 5mC in eukaryotic total DNA. **A:** Workflow for 5mC detection using TALEs. Biotin is shown as black dot, streptavidin-bead as grey sphere, other elements as before. **B:** 5mC-dependent, TALE-controlled primer extension. Reactions contained 100 ng non-methylated or methylated total DNA and 0.8 nM p-drh302 in presence or absence of 25 mU KF(exo-) and/or 1160 nM TALE T-drh301. Washed streptavidin beads were used in qPCR for quantification of extension product. * $p < 0.05$ from Student *t* test.^[13]

Again amplification was only observed when KF(exo-) was present, meaning selective removal of the non-biotinylated template DNA and indicating effective primer extension under this conditions. Furthermore, in absence of T-drh301 no significant difference in primer extension was observed for the two different DNAs, indicating similar quality. Even though some impact on primer extension of T-drh301 was observed in case of methylated DNA, in case of non-methylated DNA a significant, approximately two fold decrease in primer extension product formation was observed, when compared to

methyated DNA. (See Fig. 20) This shows that T-drh301 is able to detect its target sequence in a large and complex eukaryotic genome and to distinguish between methylated and non-methylated cytosines. Furthermore it proves the method to being able to be coupled with a selective, sensitive and scalable assay read-out.

3.1.5 5mC detection at single nucleotide resolution with TALEs

The above described method enables the investigation of the methylation status of a user defined DNA sequence. A single methylation is sufficient to disturb the binding of a given TALE protein to its target sequence. As it has been shown above, many positions, except those at the very 3' end, are able to be detected, but they cannot be distinguished. This causes the resolution to be the size of the bound target sequence and therefore being in the range of 16 to 20 nt. Even though the provided resolution is one order of magnitude better than provided by antibody based approaches, by now this method cannot keep up with the single nucleotide resolution offered by bisulfite sequencing.

To overcome this lack in resolution all cytosines that are not of interest should be faced with universal repeat units binding C and 5mC with same affinity. The reported **N*** RVD is known to bind both C and 5mC with similar affinity.^[15] So, to investigate if this RVD could be used for single nucleotide resolution detection of 5mC, TALE T-drh102 was assembled bearing RVD **N*** in repeats 1 and 10 instead of **HD**. Template oligonucleotides with different methylation patterns, regarding positions 1, 5 and 10 (See Fig. 21B) were hybridized to 5'-³²P-labelled p-drh101 and primer extensions were conducted as above and analyzed by denaturing PAGE.

First, primer extension was performed on template Pt-drh101 to validate that T-drh102 is able to bind to the target region and inhibit primer extension. An IC_{50} value of 58 ± 8 nM @ 25 mU KF(exo-) was determined, being much lower than for T-drh101, possibly reflecting the reported impact of amino acids at position 4 and 32 within the repeat units. According to the reported lowered efficiency of TALENs without variability at positions 4 and 32, this data indicate lower affinity of non-variably assembled TALEs to their target. To verify this lower affinity being caused by the lacking non RVD variability, TALE T-drh103 was assembled, providing the same architecture and RVD sequence as T-drh101 just lacking the non RVD variability. The IC_{50} value of T-drh103 was determined to 50 ± 7 nM @ 25 mU KF(exo-) being in the same range as the IC_{50} value of T-drh102 and considering the errors being identical. This proves the impact of the non RVD variability on the affinity of TALEs to their target.

Next, PEX on Pt-drh102 was conducted and it was shown that the discriminatory ability at position 5 remained. (See Fig. 21C bars 1 and 2) Further on, the remaining methylation patterns were investigated in PEX conducted as above. Always when methylation was present at position 5 of the target sequence TALE binding was perturbed and primer was

extended independent on the methylation status at positions 1 and 10. Was methylation absent at position 5, T-drh102 was always able to bind the target sequence and inhibit primer extension by KF(exo-), again independent on the methylation status at positions 1 and 10. Concluding that the methylation status at positions facing a universal RVD has no impact on TALE binding to DNA. (See Fig. 21C)

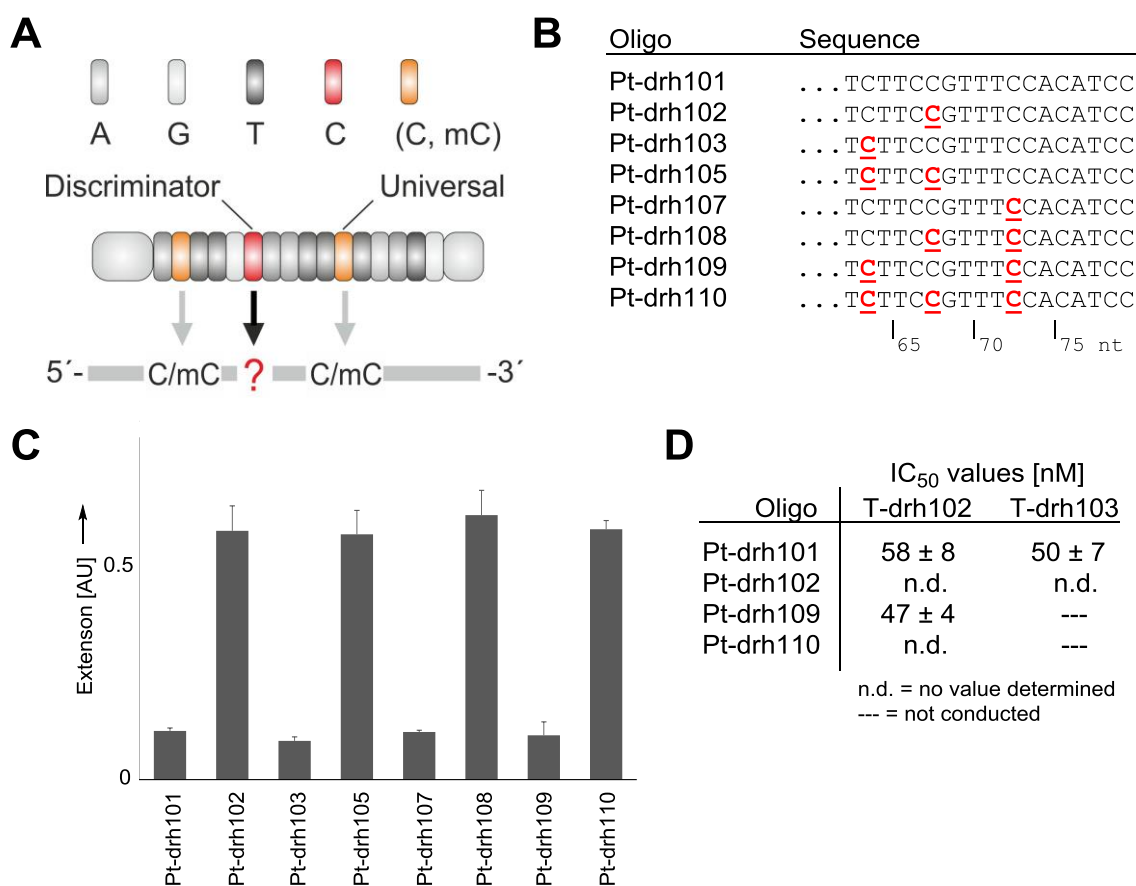


Figure 21: Single nucleotide resolution with TALEs **A:** TALE design strategy for single nucleotide resolution of mC detection. A single discriminatory TALE repeat (red) is used to recognize the C-position of interest (red question mark), the modification status of other C positions in the target sequence is ignored by using universal TALE repeats (orange) for their recognition. **B:** Target sequences in Primer extension templates, 5mC depicted in bold red and underlined. **C:** Analysis of TALE T-drh102 bearing RVD HD at position 5 and RVDs N* at positions 1 and 10 for its ability to detect mC with single nucleotide resolution. Primer extension reactions containing 8.325 nM primer-template complex in presence of 416 nM TALE and 25 mU KF(exo-) were analyzed by denaturing PAGE. Full length extension product was quantified and shown in column diagram. AU = Arbitrary Unit. **D:** IC₅₀ values determined in primer extension assay with 25 mU KF(exo-).^[16]

To validate this more accurately, measurements of IC₅₀ values for at position 5 methylated and non-methylated templates in case of full methylation and full non-methylation at position 1 and 10 were conducted. In case of methylation at position 5 no values have been determined due to the too low impact of T-drh102 on primer extension. For templates lacking methylation at this position, very similar values have been

determined with slightly stronger binding of T-drh102 to DNA methylated opposite of the non-discriminatory **N*** RVD. (See Fig. 21D)

In addition it was investigated if the methylation level can be analyzed independent of the methylation status opposite of universal RVDs. For this, Mixtures of Pt-drh101 and Pt-drh102, Pt-drh109 and Pt-drh110 and Pt-drh101 and Pt-drh109, respectively, were applied to analysis by T-drh102 inhibited primer extension. As before increasing of the methylation level at position of interest leads to a linear increase in primer extension product formation. The methylation status at positions 1 and 10 opposite the universal RVDs did not show an impact on primer extension. Also no impact was observed when methylation was kept at 0 % at position 5 opposite the discriminatory **HD** RVD and varied at positions 1 and 10. (See Fig. 22).

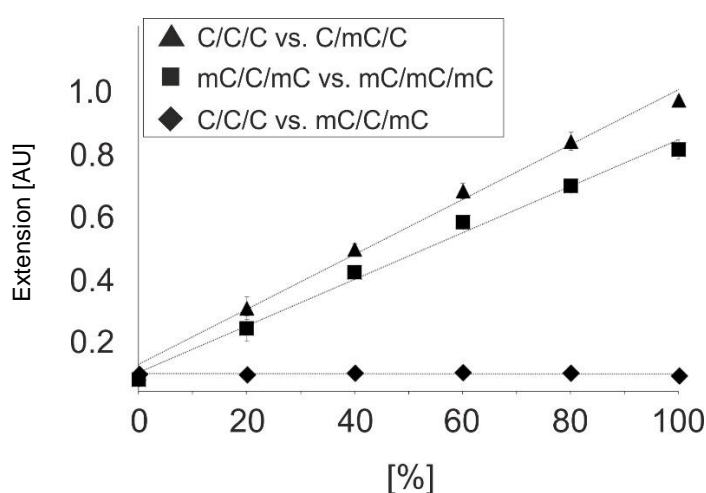


Figure 22: Analysis of the 5mC modification level at position 5 by TALE T-drh102 using template mixtures as follows: triangles: Pt-drh101 and Pt-drh102, squares: Pt-drh109 and Pt-dth110, diamonds: Pt-drh101 and Pt-drh109. Lines were obtained from linear regression.^[16]

In conclusion it has been shown that with intelligent design, this TALE based approach can be used to analyze the cytosine methylation status and level at a user defined single nucleotide position, employing a universal RVD. Since **N*** has not been tested in the context of the other cytosine modifications, there may remain a need for the development of a truly universal repeat unit as **N*** may not bind to hmC, fC and caC.

3.2 Detection of 5-hydroxymethylcytosine with TALEs

Since an ideal method for detection of epigenetic marks should not only be able to detect 5mC but also to distinguish between the oxidized derivatives of 5-methylcytosine, next, TALEs were employed for the programmable detection of 5-hydroxymethylcytosine, because this modification is much more abundant than fC and caC.

First, the ability of RVD **HD** to discriminate between C, 5mC and hmC was tested. Primer extensions on templates containing either single C, 5mC or hmC at position 5 in the target sequence with T-drh103 at varying concentrations were performed as above and analyzed by denaturing PAGE and phosphor imaging. No difference was observed for methylated and hydroxymethylated DNA. In both cases, even at high TALE to DNA ratios, no significant inhibition of primer extension by KF(exo-) was detectable. This shows that the developed method lacks selectivity between 5mC and hmC, meaning that both modifications would cause the same signal.

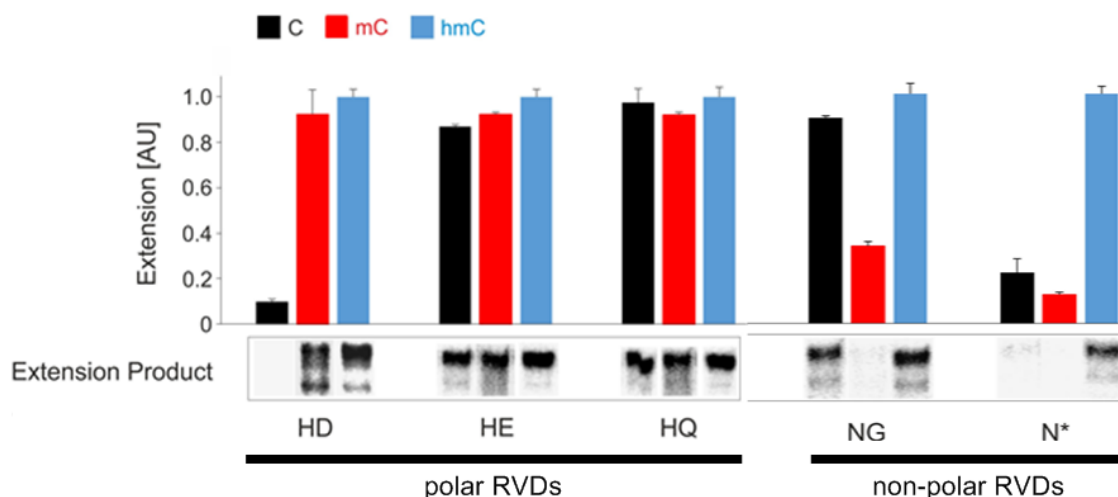
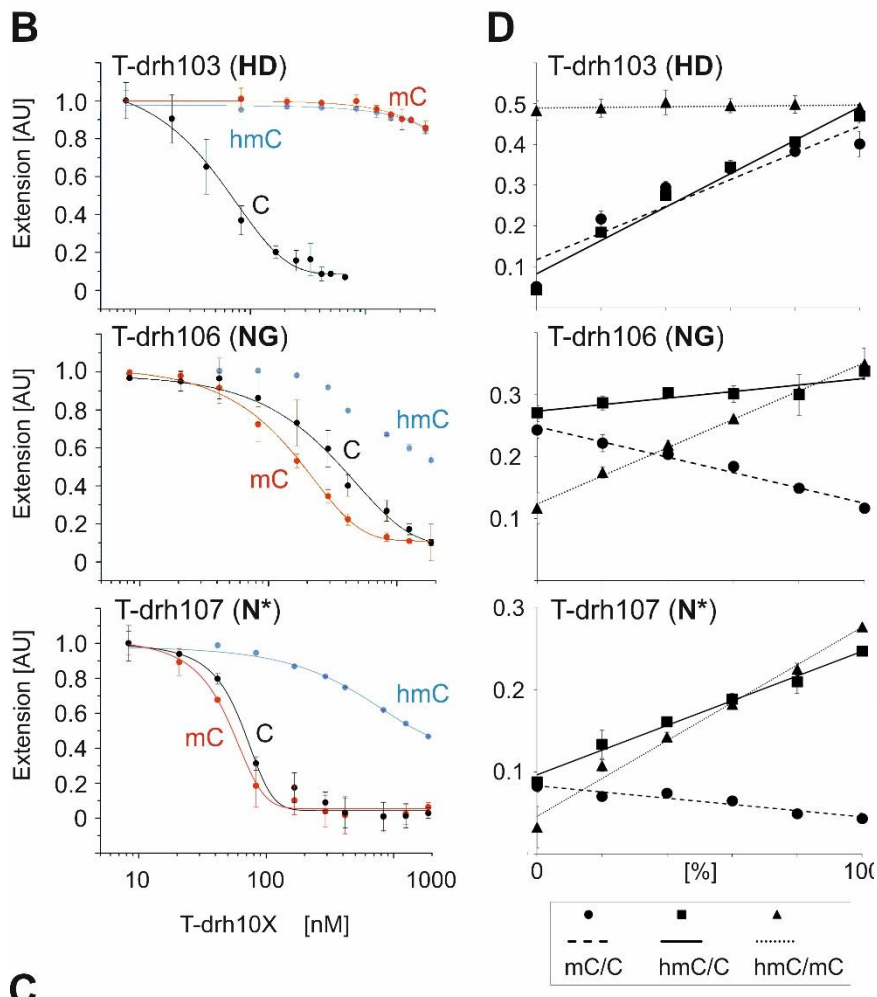


Figure 23: Sensitivity of RVDs to cytosine variants. TALEs bearing RVD **HD** and polar RVDs **HE** and **HQ** and nonpolar RVDs **NG** and **N*** for the differentiation between C (black), mC (red), and hmC (blue) in the template, analyzed in primer extension assays. Variable RVD was present in TALE repeat 5. Error bars are from duplicate experiments. Data are normalized for hmC to facilitate comparison of selectivities.^[17]

To overcome this lack of selectivity, TALEs T-drh104 and T-drh105 were assembled bearing polar RVDs **HE** and **HQ**, respectively, in repeat unit 5. In case of **HE** this means an increase in the conformational flexibility of the sidechain due to the additional methylene moiety, potentially facilitating hydrogen bonding. However, in primer extension analysis this RVD change resulted in a significantly reduced binding to C without promoting the binding to 5mC or hmC, thus being unable to differentiate between any of the three epigenetic nucleobases. In case of an installed amide group instead of the carboxylate and with the extended sidechain length (RVD **HQ**) a similar result was obtained.

Oligo	Sequence
Pt-drh101	...TCTTCCGTTTCCACATCC
Pt-drh102	...TCTTCCGTTTCCACATCC
Pt-drh111	...TCTTCCGTTTCCACATCC

| 65 | 70 | 75 nt



	K_i [nM]	C	mC	hmC	Selectivity ²
T-drh103 (HD)		50 ± 7	n. d.	n. d.	n. d.
T-drh106 (NG)		357 ± 33	163 ± 5	~2000 ¹	2.2
T-drh107 (N*)		60 ± 4	50 ± 7	~1500 ¹	25

¹ Estimated

² Selectivities, i.e. K_i ratios for cognate vs. two noncognate nucleobases.

Copyright © 2015 American Chemical Society

Figure 24: Quantitative analysis of affinity and selectivity of TALEs with nonpolar TALE repeats. **A:** templates bearing a single C, mC (red), or hmC (blue) at position 5 **B:** KF(exo⁻) inhibition profiles by T-drh103, T-drh106, and T-drh107 with primer template complex. Data were normalized for hmC and fitted using a dose response function (fits are shown as line). Error bars are from duplicate experiments. **C:** Inhibition constants K_i of T-drh103, T-drh106, and T-drh107 with primer template complexes of Fig. 24B. **D:** Inhibition experiments with T-drh103, T-drh106, and T-drh107 with primer template complexes of Fig. 24B employed in mixtures as indicated by arrows (left and right end of x-axis corresponds to 100 % of respective nucleobase). Y-axis as in Fig. 24B. Error bars are from duplicate experiments.^[17]

These data suggest that TALE repeats bearing RVDs **HE** or **HQ** may be prevented to form selective hydrogen bonding to C, 5mC or hmC due to unfavorable steric interactions of the enlarged side chains. (See Fig. 23)

A second approach to solve this lack in selectivity was to employ size reduced RVDs. TALEs bearing RVD **NG** (T-drh106), **N*** (T-drh107) and **S*** (T-drh108), respectively in repeat unit 5 instead of **HD** were assembled. **NG** has been reported to bind T and 5mC with similar affinity, being able to accommodate the 5-methylgroup via Van Der Waals interaction due to the lower steric demand, because of the lacking sidechain of glycine. Furthermore it has been reported that **NG** provides no significant affinity to C due to the large distance of cytosine's C5 and glycine's methylene. Considering the increased steric demand and polarity of hmC, **NG** is expected to not bind to hmC and thus enabling the detection of hmC by comparison with data collected using **HD**. In detail, binding of **HD** correlates to C at the investigated position, not binding correlates to 5mC or hmC, whereas binding of **NG** correlates to 5mC (or T in case of point mutation) and not binding correlates to C or hmC. Indicating the presence of hmC when both RVDs do not bind. Primer extensions were conducted as above and indeed inhibition of the extension product formation was observed when **NG** was facing 5mC and reduced binding was observed when facing C or hmC, making **NG** a selective sensor for 5mC. Primer extension reactions for determination of the IC_{50} values for T-drh106 revealed that binding to hmC ($IC_{50} = \sim 2 \mu M @ 25 \text{ mU KF(exo-)}$) is stronger inhibited than to C ($IC_{50} = 357 \pm 33 \text{ nM @ 25 mU KF(exo-)}$), indicating unfavorable steric effects of the hydroxyl group. RVD **NG** provides an approximately 2 fold selectivity between 5mC ($IC_{50} = 163 \pm 5 \text{ nM @ 25 mU KF(exo-)}$) and C and approximately 12 fold selectivity between 5mC and hmC. Next, discrimination ability of RVD **N*** was tested. Primer extension reactions with varying amounts of T-drh107 revealed that the above observed similar affinity to C ($IC_{50} = 60 \pm 4 \text{ nM @ 25 mU KF(exo-)}$) and 5mC ($IC_{50} = 50 \pm 7 \text{ nM @ 25 mU KF(exo-)}$) of **N*** can also be observed when placed in repeat unit 5. Strongly reduced inhibition of primer extension, and thus lower affinity, was observed when hmC ($IC_{50} = \sim 1.5 \mu M @ 25 \text{ mU KF(exo-)}$) was opposite **N***. Even though **N*** is, due to the deletion of amino acid 13, smaller than **NG**, accommodation of hmC is highly unfavored, but the estimated inhibitory constant is approximately 25 % lower than for **NG**, indicating higher affinity of **N*** to hmC than **NG**. This results show a 25 and 30 fold selectivity of **N*** for C and 5mC, respectively, than to hmC, providing a selective sensor for hmC. (See Fig. 24B/C) Using T-drh108 with the **S*** RVD, provided very similar results as observed for **N***. Additionally unpolar RVD **NV** was tested and as expected, due to the steric demand of the unpolar sidechain, no binding to any of the three cytosine variants was observed. (See S-Fig. S14)

Furthermore, different modification level profiles for this target sequence have been recorded using TALEs T-drh103, T-drh106 and T-drh107, showing strictly linear behavior of primer extension product formation when modification levels change. The effects were all in agreement with the prior determined selectivities. Showing that already low levels of the modification of interest can be detected reliably. (See Fig. 24D)

In addition, selectivity of RVDs **HD**, **NG** and **N*** to the canonical nucleobases T, G and A was investigated, showing that under this conditions G and A are not bound by any of the three RVDs, whereas T is bound by **NG** and **N*** but not by **HD**. This shows that the method may be biased by point mutations, but should be considered as very robust, since the probability for such a mutation in a large eukaryotic genome at the locus of interest is very low. However Sanger sequencing of the region of interest would provide evidence about the sample identity and reveal the presence of mutations.

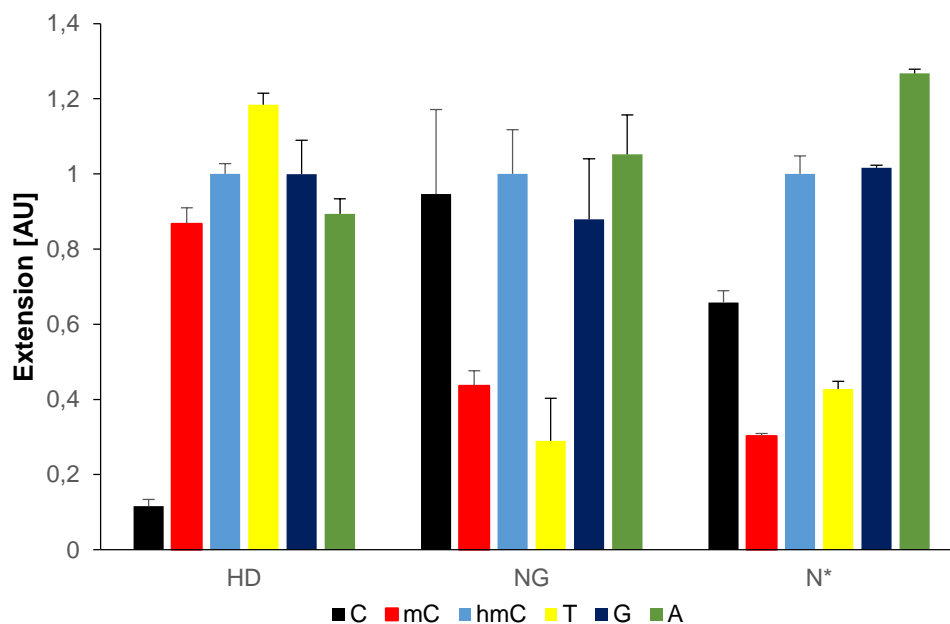


Figure 25: Selectivity profile of RVDs **HD**, **NG** and **N*** for the four canonical nucleobases and epigenetic nucleobases 5mC and hmC. Primer extension reactions containing 8.3 nM primer-template complex in presence of 416 nM TALE and 25 mU KF(exo-) were analyzed by denaturing PAGE. Full length extension product was quantified and shown in column diagram. AU = arbitrary units.^[17]

For validation, the observed discriminatory effects have been tested in the second sequence context S-drh2, to prove that no specific sequence context effects have been observed. For this reason, TALEs T-drh202, T-drh203 and T-drh204 were assembled with RVDs **HD**, **NG** and **N***, respectively, in repeat 10. The discriminatory effects that have been observed in this sequence context were in agreement with above observations, (See S-Fig. S22) even though **N*** shows significant affinity to hmC in this context, but remaining a clearly higher affinity to C and 5mC. This higher affinity to hmC

may either simply reflect the position more close to the C-terminus of the TALE protein or be caused by some sequence context effects, as it has been reported that hmC is found in two different conformations in DNA.^[41] So one explanation for sequence context effects might be that the surrounding of the hmC causes either the one or the other conformation to be more favored. This two conformations may have a different steric demand to the major groove, one providing a face more similar to 5mC and being bound better and the other being bound worse by RVD **N***. However, this data show that this approach can be easily transferred to other sequences, showing a high robustness since there was only little optimization conducted for the second context. Also, in this sequence RVD **S*** showed similar results to that collected with **N***.

In summary for each of the three most abundant cytosine variants C, 5mC and hmC, one selective sensing RVD has been found. C can be directly detected using RVD **HD**, 5mC can be directly distinguished from C and hmC when using **NG** and hmC can be detected using **N*** or **S*** RVD. To distinguish all three modifications from each other two, separate analysis, e. g. using **N*** and **HD** as sensing RVD need to be compared.

Since **N*** has been shown not to bind to hmC with high affinity, for reliable single nucleotide resolution a new universal RVD, in best case binding to all known cytosine variants, will have to be developed, since any hmC in the target sequence will have an impact on TALE binding and may cause false results, but the principle for gaining single nucleotide resolution stays unaffected.

3.3 Molecular Recognition of fC and caC

Next, the sensitivity of TALEs to the two remaining oxidized cytosine variants fC and caC was investigated. For this TALEs T-drh103, T-drh106, T-drh107 and T-drh108 bearing RVDs **HD**, **NG**, **N*** and **S***, respectively, in repeat 5, were incubated with target DNA bearing fC or caC opposite the investigated RVD. Then, primer extension reactions were performed as above and analyzed by denaturing PAGE. The same was performed for Sequence context S-drh2 using TALEs T-drh204 and T-drh205 bearing RVDs **N*** and **S***, respectively. (See S-Fig. S23 and S24)

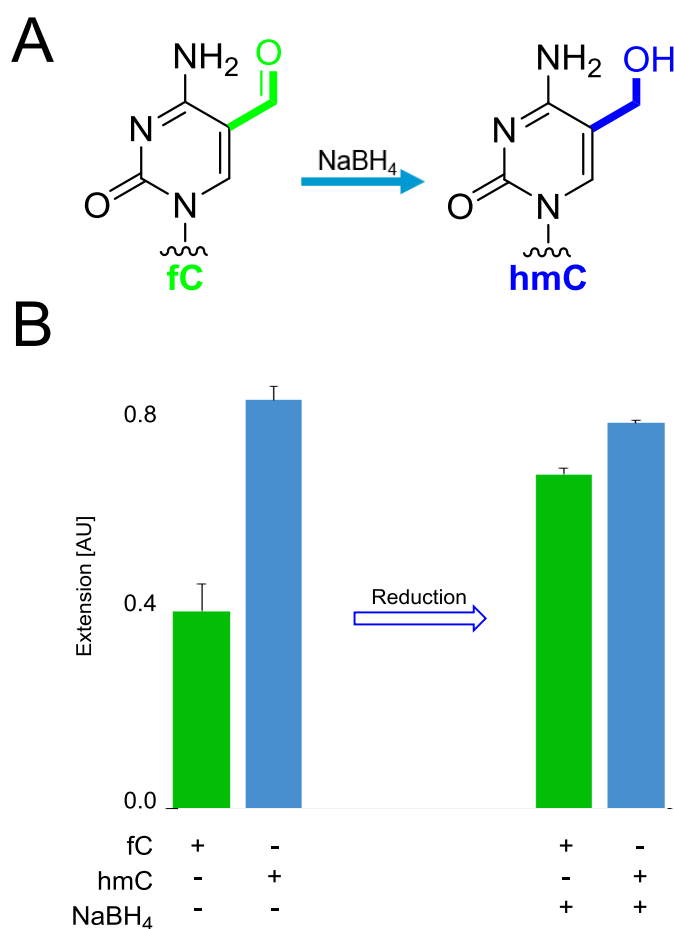


Figure 26: **A:** Selective reduction of fC to hmC by NaBH₄. **B:** Selectivity of TALE repeat N* for hmC and fC. DNA-binding before and after reduction of fC to hmC by NaBH₄ was analyzed in primer extension assay containing 8.3 nM primer-template complex in presence of 416 nM TALE T-drh107 and 25 mU KF(exo-) were analyzed by denaturing PAGE. Full length extension product was quantified and shown in column diagram. AU = arbitrary units.

The results revealed that fC is bound by RVDs **N*** and **S*** efficiently, leading to inhibition of primer extension. This results are in agreement with reported structural data where fC forms intramolecular hydrogen bonds with the amine function at 4 position of cytosine, with the aldehyde oxygen being fixed to point into the direction of the amino moiety, providing a similar face to the major groove as 5mC. On the other hand caC is not bound

by any of the RVDs, interestingly in S-drh2 where some binding to hmC was observed a clear decreased affinity to caC has been observed, when compared to hmC. This may be caused by an overall higher steric demand of caC or by the difference in polarity. This data mean, that hmC and caC and fC, 5mC and C cannot be reliably distinguished by RVDs **N*** and **S***. Due to the different sensitivity of **N*** and **S*** to fC and caC, this approach was coupled to selective chemical reduction of fC to hmC using sodium borohydride. When comparing the primer extension assay results of untreated and reduced DNA fC can be distinguished from 5mC. As shown in figure 26 after treatment of fC containing DNA with sodium borohydride an increased primer extension product formation, similar to hmC containing DNA, can be observed. This shows a decreased binding affinity of TALE protein, bearing RVD **N*** opposite the investigated cytosine. This indicates successful and selective reduction but evidence should be gained from mass spectrometry analysis. Coupling of the method to reported selective oxidation of hmC to fC with potassium perruthenate is not suitable due to the denaturing reaction conditions, making the application to genomic DNA inconvenient.

In conclusion, employing a simple chemical conversion of fC to hmC makes it possible to differentiate between C, 5mC, hmC/caC and fC using TALE proteins. Differentiation ability between hmC and caC has still to be achieved, but due to the much lesser occurrence of caC, compared to hmC, the applicability of a purely binding based assay may be challenging.

3.4 Sensitivity of TALE RVDs to N4-methylcytosine

An epigenetic variant of cytosine that is found in bacteria is N4-methylcytosine (N4mC). Its role is still elusive but being presumably part of the host protection system, blocking endogenous restriction enzymes and thus allowing them only to digest invading DNA. However in mammals this modification is either absent or extremely rare.^[106]

To have a more complete view on the sensitivity of TALEs to cytosine modifications RVDs **HD**, **NG** and **N*** were tested opposite of N4mC in primer extension assay as above and results were compared to those obtained with C, 5mC and hmC containing DNA.

First results have shown that **N*** and **NG** binds N4mC with similar affinity as 5mC and **HD** was indicated to bind N4mC stronger than 5mC. (See S-Fig. S23) For a closer look IC₅₀ value of T-drh103 was determined for a synthetic DNA template with an N4mC moiety at position 5 (Pt-drh114).

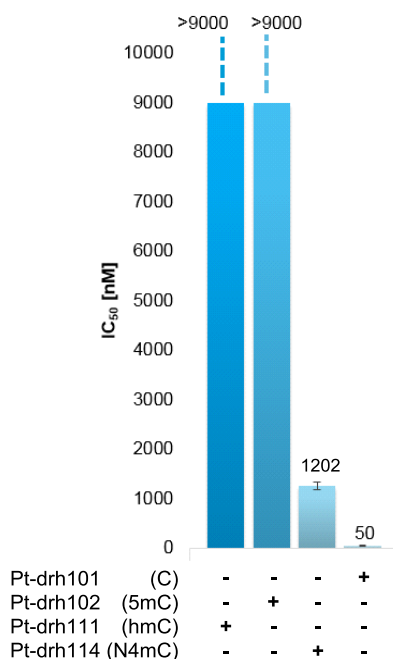


Figure 27: Comparison of IC_{50} values of TALE T-drh103 for hmC, 5mC, N4mC and C containing DNA.

The determined value of $IC_{50} = 1202 \pm 40$ nM @ 25 μ M KF(exo-) means dramatically reduced binding of **HD** to N4mC than to C by a factor of approximately 24 but significantly increased affinity compared to 5mC of at least 7 fold. (See Fig. 27) The weaker binding, compared to C, can be explained by the perturbation of the hydrogen bond between the aspartates carboxylate and the cytosine's amino group, since the N4-methyl group is expected to point into the major groove so that the remaining hydrogen of the methylated amine can be employed for Watson-Crick base pairing. The higher affinity compared to 5mC may be due to the difference in steric demand of methylation at 4 or 5 position of Cytosine and the possibility of Van der Waals interaction of the aspartates methylene moiety with C5 of Cytosine. However even without employment of new RVDs N4mC can be distinguished from 5mC and C by RVD **HD** due to the strong difference in binding affinity.

4. SUMMARY AND OUTLOOK

In summary, TALE proteins have been demonstrated to be able to detect 5mC and oxidized variants in a complex, eukaryotic genome. This methodology offers direct, programmable and conversion-free detection of cytosine modifications at 5 position with high resolution. The, within this study, observed discrimination is significantly stronger than previously reported for *in vitro* assays and provides robustness to a broad scope of target sequences and modification positions within the complex.^[13]

Furthermore a new strategy for TALE design with increased resolution of 5mC detection by TALEs with single nucleotide resolution was described. This has been achieved by combining a single discriminatory TALE repeat for investigation of the modification status and level of a C position of interest with universal TALE repeats not considering cytosine modifications at other positions. This concept has been proven for the discrimination of C and 5mC with high sensitivity and robustness regarding multiple positions in the target to be ignored, using **HD** as sensing RVD and **N*** as universal RVD.^[16]

In addition sensing ability of TALEs was expanded to selectively recognize C, 5mC, and hmC, respectively, in user-defined, canonical DNA sequences by the use of RVDs **HD**, **NG** and **N***, respectively. This defines a toolbox of TALE repeats, each providing high selectivity for its cognate nucleobase in the presence of the other two non-cognate nucleobases.^[17] Furthermore it has been shown that by application of selective reduction, also fC can be distinguished from 5mC and C with **N*** as sensing RVD, and N4mC can be distinguished from 5mC and C with sensing RVD **HD** by the difference in affinity.

The former reported sensitivity of TALEs to Cytosine methylation has thus been investigated in more detail and turned into a powerful assay for the detection of a variety of epigenetic Cytosine modifications that either already are or may raise as bio markers for different diseases such as different cancer types, providing a tool for future application in diagnostics.^[7]

Based on this first studies further engineering of TALEs promises to gain full selectivity to all cytosine modifications by the development of repeat units with unique selectivities to one special cognate epigenetic nucleobase. In addition a universal repeat may be found enabling unbiased single nucleotide resolution in epigenetic modification analysis with TALEs.

Thus TALEs, as fully programmable and genetically encoded sensors for epigenetic modifications, applicable in both *in vitro* and *in vivo*, offer a platform for diverse epigenetic technologies by combination with functional domains such as fluorescent proteins, transcriptional activators and repressors, nucleases, DNMTs and TET proteins.

5. MATERIALS AND METHODS

The methods of this section have been partly published in supplementary information of the following publications

Angew. Chem. Int. Ed. **2014**, *53*, 6002–6006.

© 2014 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

and

ChemBioChem **2015**, *16*, 228–231.

© 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

and

J. Am. Chem. Soc., **2015**, *137* (1), pp 2–5.

Copyright © 2015 American Chemical Society

All adaptations and reprints have been made with permission.

5.1 Materials

5.1.1 Lab equipment

Type	Model	Manufacturer
Agarose Gel System	SGU 020T-02	C.B.S. Scientific
Balance	PCJ 3500_2NM	Kern
Balance	SE622	VWR
Balance	ABT 220-4M	Kern
Centrifuge	5417R	Eppendorf
Centrifuge	5424	Eppendorf
Centrifuge	5810R	Eppendorf
Centrifuge	Universal 320R	Hettich
Centrifuge	iFuge M08VT	Neuation
Concentrator	Plus 5305	Eppendorf
Electroporator	Eporator	Eppendorf
Gel dryer	MGD-4534	VWR
Gel visualization System	UV star with Powershot G10	Biometra
Heating block	AccuBlock	LabNet
Imaging Plate	BAS-IP2040	Fuji
Incubator	Incu-Line	VWR
Magnetic Stirrer	Stir CB161	Stuart
Magnetic Stirrer	MR Hei-Standard	Heidolph
Magnetic Stirrer	RCT classic	IKA
Orbital shaker	STD 1000 shaker	VWR
PCR Cycler	PeqLab	PeqLab
PCR Cycler	MyCycler	BioRAD
pH meter	Fire Easy FE20	Mettler Toledo
Molecular Imager	FX Pro	BioRAD
Photometer	BioPhotometer Plus	Eppendorf
Pipette	0.1-2.5 µl Research	Eppendorf
Pipette	0.1-2.5 µl Research Plus	Eppendorf
Pipette	0.5-10 µl Research Plus	Eppendorf
Pipette	2-20 µl Research Plus	Eppendorf
Pipette	10-100 µl Research Plus	Eppendorf
Pipette	20-200 µl Research	Eppendorf
Pipette	100-1000 µl Research Plus	Eppendorf
Pipette	0.5-10 µl Research Plus 8 Chanel	Eppendorf
Pipette	10-100 µl Research 8 Chanel	Eppendorf
Pipette	Rotafiller 3000	HS
Pipette	Pipetboy	Integra
Pipette	Easypet	Eppendorf

Pipette	Multipette plus	Eppendorf
Plate Reader	Infinite M200	TECAN
Power supply	EV233	Consort
Power supply	EV243	Consort
Power supply	EV245	Consort
qPCR Cycler	iCycler	BioRAD
SDS Gel System	Mini Protean Tetra Cell	BioRAD
Sequencing Gel System	Sequi-Gen	BioRAD
Shaking incubator	Ecotron	Infors HT
Software	Origin 8.6 Pro	OriginLab
Software	BioDoc Analyze	Analytik Jena AG
Software	iCycler	BioRAD
Software	MS Office	Microsoft
Software	ChemBioDraw	Cambridgesoft
Software	SerialCloner	SerialBasics
Software	Finch TV	Geospiza
Software	Quantity One	BioRAD
Thermomixer	Comfort	Eppendorf
Thermomixer	Compact	Eppendorf
Thermomixer	PocketBloc	BioEr
UV-Lamp	VL-6.LC	VILBER
Vortexer	Vortex-Genie 2	Scientific Industries
Water sterilizer	Synthesis A10	Millipore

5.1.2 Consumables

Type	Model	Supplier
Pipette	86.1254.001	Sarstedt
Pipette	86.1685.001	Sarstedt
Pipette	Pasteur 230 mm	WU Mainz
Syringe	20 ml (24 ml) NORM-JECT Luer	Henke Sass Wolf
Syringe needle	Sterican 0.90x40 mm	Braun
Syringe needle	Sterican 0.60x80 mm	Braun
Syringe needle	Sterican 0.80x120 mm	Braun
Micro centrifuge tube	72.695.500	Sarstedt
Micro centrifuge tube	72.706	Sarstedt
Micro centrifuge tube	72.737.002	Sarstedt
Cuvette	67.758	Sarstedt
Cuvette	67.742	Sarstedt
Cuvette	Electroporation cuvettes 1 mm	Roth
Dialysis tube	Visking Cellulose	Roth
Dialysis device	Slide-A-Lyzer Mini Dialysis unit	Thermo Scientific
Micro titter plate	96 PCR Plate non-skirted	VWR
Column	PP column 1 ml	Qiagen
Column	PP column 5 ml	Qiagen
Column	Centrifuge Columns, 0.8 mL	Pierce
Concentrator	Amicon Ultra 0.5 mL – 3K	Millipore
Concentrator	Amicon Ultra 0.5 mL – 50K	Millipore
Concentrator	Amicon Ultra 0.5 mL – 100K	Millipore
Pipette Tips	70.1130	Sarstedt
Pipette Tips	70.760.002	Sarstedt
Pipette Tips	70.762	Sarstedt
Cover films	Self-adhesive, soft PVC	Ratiolab
Column	Illustra G-25 Micro Spin	GE Healthcare
Tube	62.547.254	Sarstedt
Tube	62.554.512	Sarstedt
Gas	C 206 GLS	Roth
Petri dishes	82.1473.001	Sarstedt
Sterile filter	83.1826.001	Sarstedt

5.1.3 Chemicals, Reagents, Formulations

Name	Grade	Abbreviation	Supplier
2-Log DNA Ladder			NEB
(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)	≥99.5 %	HEPES	Roth
[gamma-P32]ATP			Hartmann Analytic
2-(N-morpholino)ethanesulfonic acid-monohydrate	≥98 %	MES	Fisher
Ammonium persulfate	≥98 %	APS	Roth
Ammonium sulfate	≥99.5 %		Merck
Boric acid	≥99.8 %		Roth
Bovine Serum Albumin			Cell Signaling
Brilliant blue G250			Roth
Bromophenol blue			Roth
Calcium chloride			Fisher
Carbenicilin	≥98 %		Roth
Chloramphenicol	≥98.5 %		Roth
ColorPlus Prestained Protein Ladder			NEB
Disodiumhydrogen phosphate	≥98.5 %		Sigma Aldrich
Dithiothreitol	≥99 %	DTT	Roth
Ethanol	≥99.8 %		Sigma Aldrich
Ethidium bromide			Sigma Aldrich
Ethylenediaminetetraacetic acid	≥99 %	EDTA	Roth
Form amide	≥99.5 %		Acros
Glycerol			Roth
Hydrochloric acid			VWR
Imidazol	99 %		ABCR
Isopropanol	≥99.9 %		Fisher
Isopropyl β-D-1-thiogalactopyranoside	≥99 %	IPTG	Roth
LB-Agar			Roth
LB-Medium			Roth
LE Agarose			Roth
Magnesium chloride-hexahydrate	≥99 %		Acros
Magnesium sulfate-heptahydrate	≥99.5 %		Merck
Manganese chloride-tetrahydrate	≥99 %		Riedel-de Haën

Nickel Nitrilotriacetic acid		Ni-NTA	Thermo Scientific
Phenylmethanesulfonylfluoride	≥99 %	PMSF	Roth
Potassium chloride	≥99.5 %	KCl	Roth
Potassium dihydrogenphosphate-monohydrate	≥99.5 %		Sigma Aldrich
Protein Marker, Broad Range (2-212 kDa)			NEB
Rotiphorese Gel40			Roth
S-adenosylmethionine			NEB
Salmon Sperm DNA			Ambion
Sequencing gel concentrate			Roth
Sodium chloride	≥99.5 %	NaCl	Roth
Sodium dodecyl sulfate	≥99 %	SDS	Roth
Sodium Hydroxide		NaOH	Fisher
Spectinomycin		Spec	Alfa Aeser
Streptavidin-Agarose			Sigma Aldrich
Tetracycline		Tet	Roth
Tetramethylethylenediamine	≥99 %	TMEDA	Roth
Tris(hydroxymethyl)aminomethane	≥99.9 %	Tris	Sigma Aldrich
Triton X-100			Fluka
Tryptone			Roth
Urea	≥99.5 %		Roth
Xylene cyanol			Roth
Yeast Extract			Roth

5.1.4 Enzymes

Application	Name	Supplier
Restriction	Bsal	NEB
Restriction	BsmBI	NEB/Thermo
Restriction	XhoI	NEB
Restriction	NcoI	NEB
Restriction	Plasmid Safe	Biozym
Restriction	DpnI	NEB
Restriction	HaeIII	NEB
Primer extension	KF(exo-)	NEB
PCR	Phusion	NEB
PCR	Taq	NEB/Self-made
Lysis	Lysozyme	Fluka
Ligation	T4-DNA ligase	NEB
Labeling	T4 Polynucleotidekinase	Fermentas
Methylation	M.SssI	NEB

5.1.5 Kits

Type	Name	Supplier
PCR Purification	GeneJet PCR purification	Thermo Fisher Scientific
Gel extraction	GeneJet Gel extraction	Thermo Fisher Scientific
Plasmid isolation	GeneJet Plasmid miniprep	Thermo Fisher Scientific
Bisulfite conversion	Epitect	Qiagen
Total DNA Extraction	DNeasy Blood & Tissue	Qiagen
DNA Purification	QiaAmp	Qiagen
qPCR master mix	GoTaq	Promega
BCA	Pierce BCA Protein Assay Kit - Reducing Agent Compatible	Thermo Fisher Scientific

5.1.6 Buffers and Media

- **SOB⁺⁺:**
 - 20 g/l tryptone
 - 5 g/l yeast extract
 - 0.5 g/l NaCl
 - 0.186 g/l KCl
 - Dissolve in water
 - Heat sterilize the medium and add
 - 10 mmol/l MgCl₂
 - 10 mmol/l MgSO₄

- **TB Buffer**
 - 10 mM HEPES
 - 15 mM CaCl₂
 - 250 mM KCl
 - adjust pH to 6.7 and autoclave
 - Add 55 mM MnCl₂

- **SOC:**
 - 0.58 g/l NaCl
 - 2.03 g/l MgCl₂ · 6H₂O
 - 2.46 g/l MgSO₄ · 7H₂O
 - 5 g/l Yeast Extract
 - 20 g/l Trypton
 - Dissolve in MilliQ-water and adjust pH to 7-7.5
 - Add 20 ml of 1M sterile glucose solution after autoclaving

- **TBE buffer**
 - 89 mM Tris
 - 89 mM Boric acid
 - 2 mM EDTA (diluted from 0.5 M solution, pH = 8)

- **ThermoPol® Reaction Buffer (NEB formula)**
 - 20 mM Tris-HCl
 - 10 mM (NH₄)₂SO₄
 - 10 mM KCl
 - 2 mM MgSO₄·7H₂O
 - 0.1% Triton® X-100
 - Adjust pH to 8.8

- **Taq Lysis Buffer**
 - 10 mM Tris-HCl
 - 300 mM NaCl
 - 2,5 mM MgCl₂
 - 0.1 % Triton X-100
 - adjust pH to 9

- **Buffer-Z**
 - 8 M Urea
 - 0.1 M NaCl
 - 20 mM HEPES
 - Adjust pH to 8

- **4x PBS**
 - 548 mM NaCl
 - 43 mM KCl
 - 69 mM Na₂HPO₄·2H₂O
 - 3.2 g KH₂PO₄
 - adjust pH to 8

- **TALE storage Buffer:**
 - 200 mM NaCl
 - 20 mM Tris
 - 50% Glycerol

- **Hybridization Buffer (2xBGrK1):**
 - 40 mM Tris-HCl (pH = 8)
 - 100 mM NaCl
 - 10 mM MgCl₂
 - 0,2 mg/ml BSA
 - 25% glycerol
 - Vortex

- **Binding buffer**
 - 30 mM Tris-HCl (pH = 8.0)
 - 150 mM NaCl
 - 5 mM MgCl₂
 - 0.1 mg/ml BSA
 - 0.5 mM DTT
 - 7.5 % glycerol

- **CutSmart Buffer**
 - 50 mM KOAc
 - 20 mM Tris-acetate
 - 10 mM Mg(OAc)₂
 - 100 µg/mL BSA
 - pH = 7.9

- **NEB2 Buffer**
 - 50 mM NaCl
 - 10 mM Tris-HCl
 - 10 mM MgCl₂
 - 1 mM DTT
 - pH = 7.9

- **B&W**
 - 1 M NaCl
 - 5 mM Tris
 - 0.5 mM EDTA
 - Adjust pH to 7.5

- **PCR reconstitution buffer**
 - 5 mM Tris
 - 50 mM NaCl
 - pH = 7.5

5.1.7 Oligonucleotides

Name	Lab Code	Sequence	Supplier
Et-drh101	oGrK_415	AGA CGG CAT ACG ATC TTC CGT TTC CAC	Sigma
		ATC CAC CAC ATC CCA ACA GAG CAG CGG	Aldrich
		GAG CAG CAG TAA AGA TCG GAA GAG C	
Et-drh102	oGrK_416	AGA CGG CAT ACG AT <u>C</u> TT <u>C</u> <u>CGT</u> TT <u>C</u> <u>CAC</u>	Sigma
		AT <u>C</u> CAC CAC ATC CCA ACA GAG CAG CGG	Aldrich
		GAG CAG CAG TAA AGA TCG GAA GAG C	
Et-drh103	oGrK_418	AGA CGG CAT ACG ATC CCC CGC CCC CAC	Sigma
		ACC CAC CAC ATC CCA ACA GAG CAG CGG	Aldrich
		GAG CAG CAG TAA AGA TCG GAA GAG C	
Et-drh104	oGrK_419	AGA CGG CAT ACG ATC <u>CCC</u> <u>CGC</u> <u>CC</u> C CAC	Sigma
		<u>AC</u> C CAC CAC ATC CCA <u>ACA</u> GAG <u>CAG</u> CGG	Aldrich
		GAG CAG CAG TAA AGA TCG GAA GAG C	
Er-drh101	oGrK_417	GCT CTT CCG ATC TTT ACT GCT GCT CCC	Sigma
		GCT GCT CTG TTG GGA TGT GGT GGA TGT	Aldrich
		GGA AAC GGA AGA TCG TAT GCC GTC T	
Er-drh103	oGrK_420	GCT CTT CCG ATC TTT ACT GCT GCT CCC	Sigma
		GCT GCT CTG TTG GGA TGT GGT GGG TGT	Aldrich
		GGG GGC GGG GGA TCG TAT GCC GTC T	
Pt-drh101	oGrK_476	TGG ATT CCC ACT CTT CAG CCC CAG CGT	Sigma
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	Aldrich
		CGT GAG CTC TTC CGT TTC CAC ATC C	
Pt-drh102	oGrK_465	TGG ATT CCC ACT CTT CAG CCC CAG CGT	Sigma
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	Aldrich
		CGT GAG CTC TTC <u>CGT</u> TTC CAC ATC C	
Pt-drh103	oGrK_515	TGG ATT CCC ACT CTT CAG CCC CAG CGT	Sigma
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	Aldrich
		CGT GAG CT <u>C</u> TTC CGT TTC CAC ATC C	
Pt-drh104	oGrK_516	TGG ATT CCC ACT CTT CAG CCC CAG CGT	Sigma
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	Aldrich
		CGT GAG CTC TTC CGT TTC <u>CAC</u> ATC C	
Pt-drh105	oGrK_517	TGG ATT CCC ACT CTT CAG CCC CAG CGT	Sigma
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	Aldrich
		CGT GAG CT <u>C</u> TTC <u>CGT</u> TTC CAC ATC C	
Pt-drh106	oGrK_518	TGG ATT CCC ACT CTT CAG CCC CAG CGT	Sigma
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	Aldrich
		CGT GAG CT <u>C</u> TTC <u>CGT</u> TTC <u>CAC</u> ATC C	
Pt-drh107	oGrK_712	TGG ATT CCC ACT CTT CAG CCC CAG CGT	Sigma
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	Aldrich
		CGT GAG CTC TTC CGT TT <u>C</u> CAC ATC C	
Pt-drh108	oGrK_1055	TGG ATT CCC ACT CTT CAG CCC CAG CGT	Sigma
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	Aldrich
		CGT GAG CTC TTC <u>CGT</u> TT <u>C</u> CAC ATC C	
Pt-drh109	oGrK_1057	TGG ATT CCC ACT CTT CAG CCC CAG CGT	Sigma
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	Aldrich
		CGT GAG CT <u>C</u> TTC CGT TT <u>C</u> CAC ATC C	
Pt-drh110	oGrK_1056	TGG ATT CCC ACT CTT CAG CCC CAG CGT	Sigma
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	Aldrich
		CGT GAG CT <u>C</u> TTC <u>CGT</u> TT <u>C</u> CAC ATC C	
Pt-drh111	oGrK_520	TGG ATT CCC ACT CTT CAG CCC CAG CGT	metabion
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	
		CGT GAG CTC TTC <u>CGT</u> TTC CAC ATC C	
Pt-drh112	oGrK1423	TGG ATT CCC ACT CTT CAG CCC CAG CGT	metabion
		TAC AGC ATC TTC AGT GGC TTC TTC CAC	
		CGT GAG CTC TTC <u>CGT</u> TTC CAC ATC C	

Pt-drh113	oGrK1422	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC <u>CGT</u> TTC CAC ATC C	metabion
Pt-drh114	oGrK_1236	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC <u>CGT</u> TTC CAC ATC C	IBA
Pt-drh115	oGrK_1170	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC TGT TTC CAC ATC C	Sigma Aldrich
Pt-drh116	oGrK_1171	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC GGT TTC CAC ATC C	Sigma Aldrich
Pt-drh117	oGrK_1172	TGG ATT CCC ACT CTT CAG CCC CAG CGT TAC AGC ATC TTC AGT GGC TTC TTC CAC CGT GAG CTC TTC AGT TTC CAC ATC C	Sigma Aldrich
p-drh101	oGrK_466	GGA TGT GGA AAC GGA AGA	Sigma Aldrich
p-drh102	oGrK_467	Biotin-GGA TGT GGA AAC GGA AGA	Sigma Aldrich
p-drh103	oDaS_892	Biotin-TEG-GGA TGT GGA AAC GGA AGA	Sigma Aldrich
p-drh104	oDaS_887	Biotin-TEG-TTG GGA TGT GGT GGA TGT GGA AAC GGA AGA	Sigma Aldrich
p-drh105	oGrK_1181	GGA TGT GGA AAC AGA AGA	Sigma Aldrich
p-drh106	oGrK_1182	GGA TGT GGA AAC CGA AGA	Sigma Aldrich
p-drh107	oGrK_1183	GGA TGT GGA AAC TGA AGA	Sigma Aldrich
Pt-drh201	oGrK_702	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA CTG CTG CTC CCG CT	Sigma Aldrich
Pt-drh202	oGrK_703	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA CTG <u>CTG</u> CTC CCG CT	Sigma Aldrich
Pt-drh203	oGrK_790	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA <u>CTG</u> CTG CTC CCG CT	Sigma Aldrich
Pt-drh204	oGrK_791	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA CTG CTG <u>CTC</u> CCG CT	Sigma Aldrich
Pt-drh205	oGrK_792	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA CTG CTG <u>CTC</u> CCG CT	Sigma Aldrich
Pt-drh206	oGrK_793	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA CTG CTG CTC <u>CCG</u> CT	Sigma Aldrich
Pt-drh207	oGrK_794	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA CTG CTG CTC <u>CCG</u> CT	Sigma Aldrich
Pt-drh208	oGrK_795	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA CTG CTG CTC CCG <u>CT</u>	Sigma Aldrich
Pt-drh209	oGrK1393	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA CTG CTG <u>CTC</u> CCG CT	metabion
Pt-drh210	oGrK1481	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA CTG CTG <u>CTC</u> CCG CT	metabion

Pt-drh211	oGrK1480	ACA TTT AAA TCC AAC ATT TAA AAC GCT CCC ACT TCA GTT CCC CAC GGT CGG TAT GGT TTA CTG CTG <u>CTC</u> CCG CT	metabion
p-drh201	oGrK_787	GCA GCG GGA GCA GCA GTA AA	Sigma Aldrich
Pt-drg101	oGrK_704	GAA CAT CAC TGA GAA ACC TTT CCA TCT CAA TTA CAA CGT GGA CAA TCT GGA GTC GTC CAA TAG CTT CTC	Sigma Aldrich
Pt-drg102	oGrK_705	GAA CAT CAC TGA GAA ACC TTT CCA TCT CAA TTA CAA CGT GGA CAA TCT GGA <u>GTC</u> GTC CAA TAG CTT CTC	Sigma Aldrich
Pt-drg103	oGrK_706	GAA CAT CAC TGA GAA ACC TTT CCA TCT CAA TTA CAA CGT GGA CAA TCT GGA GTC <u>GTC</u> CAA TAG CTT CTC	Sigma Aldrich
Pt-drg104	oGrK_707	GAA CAT CAC TGA GAA ACC TTT CCA TCT CAA TTA CAA CGT GGA CAA TCT GGA GTC GTC <u>CAA</u> TAG CTT CTC	Sigma Aldrich
Pt-drg105	oGrK_708	GAA CAT CAC TGA GAA ACC TTT CCA TCT CAA TTA CAA CGT GGA CAA TCT GGA GTC GTC CAA TAG <u>CTT</u> CTC	Sigma Aldrich
p-drg101	oGrK_788	TGA CTG AGA <u>AGC</u> TAT TGG ACG A	Sigma Aldrich
Pt-drh301	oGrK_1030	AAC ATT TAA ATC CAA CAT TTA AAA CGC TCC CAC TTC AGT TCC CCA CGG TCG GTA TGG TTT ACT GCT GCT CCC GCT GCT C	Sigma Aldrich
Pt-drh302	oGrK_1031	AAC ATT TAA ATC CAA CAT TTA AAA CGC TCC CAC TTC AGT TCC CCA CGG TCG GTA TGG TTT ACT GCT GCT <u>CCC</u> GCT GCT C	Sigma Aldrich
p-drh301	oGrK_1029	GAG CAG CGG GAG CAG CAG TA	Sigma Aldrich
p-drh302	oGrK_1004	Biotin-TEG-CAA CAG AGC AGC GGG AGC AGC AGT A	Sigma Aldrich
dHax_hom_Notl_His6_rev_pET-24a	oDaS_231	TTT TGC GGC CGC TCA GTG ATG GTG ATG GTG ATG TTC CAG CAG ATG GTC ATT CG	metabion
dHax_hom_KpnI_fwd_pET32_before_T7	oDaS_247	GGG TTA TGC TAG TTA TTG CTC AG	metabion
HEY2_Bis_a_r	oDaS_308	TTT TTG GTA CCG TGG ACT TGA GGA CAC TCG GTT ATT CGC AAC AGC	metabion
HEY2_Bis_b_f	oDaS_315	CGT AGA GGA TCG AGA TC	metabion
HEY2_Bis_b_r	oDaS_751	CCC AAA ATA TTA ACT TAA ATT TAA C	Sigma Aldrich
HEY2_Bis_b_r	oDaS_754	GTT TGA GGT TTT TTA AAG AGG T	Sigma Aldrich
HEY2_Bis_d_r	oDaS_756	TAG TAG TTG TTG TAG TTG TAG TTT AAA TTT	Sigma Aldrich
hey2_PCR_fwd	oDaS_757	CTA CTT TCC TTT ATA TAC AAA ACA TAA AAA	Sigma Aldrich
	oDaS_760	TGC CAC CGC TGT CCA CT	Sigma Aldrich
	oDaS_762	TCC ACT TCT CTG CTT TCG	Sigma Aldrich
	oDaS_763	AGA AGC CAC TGA AGA TGC	Sigma Aldrich
qp_drh3_fwd	oDaS_903	GGA ACT GAA GTG GGA GCG T	Sigma Aldrich
qp_drh3_rev	oDaS_904	GAT GAT ATC AAA GCA TAC ATC TCA	Sigma Aldrich
qp-drh1_fwd	oGrK_621	CCC ACT CTT CAG CCC CA	Sigma Aldrich

pGrK471	pET_TRX_TALE-054(1257)_6His	Amp
pGrK484	pS*5	Tet
pGrK485	pNV5	Tet
pGrK501	pET_TRX_TALE-056_6His	Amp
pGrK506	pET_TRX_TALE-055_6His	Amp
pGrK507	pET_TRX_TALE-058_6His	Amp
pGrK582	pET_TRX_TALE-094_6His	Amp
pGrK585	pET_TRX_TALE-095_6His	Amp
pGrK624	pS*10	Tet
pGrK631	pET_TRX_TALE-116_6His	Amp

5.1.9 Bacterial Strains

E. coli Strain	Lab Code	Endogenous Resistance	Supplier
BL21 DE3	sMoS308		Marx Group
BL21 (DE3) GOLD	sMoS309	Tet	Hartig Group
DH5 alpha or DH10B	sSaB193- sSaB277		Addgene TALE Tool Box
GH371	sDaS58		iGEM
Top10	sDaS46		Lemke Group

5.1.10 TALE Proteins

TALE	Lab Code	Plasmid	RVD Sequence	Assembly
T-drh101	TALE_002	pDaS170	HD NG NG HD HD NN NG NG NG HD HD NI HD NI NG HD NG	No
T-drh102	TALE_016	pVaS348	N* NG NG HD HD NN NG NG NG HD N* NI HD NI NG HD NG	Yes
T-drh103	TALE_041	plrT446	HD NG NG HD HD NN NG NG NG HD HD NI HD NI NG HD NG	Yes
T-drh104	TALE_018	pVaS350	HD NG NG HD HE NN NG NG NG HD HD NI HD NI NG HD NG	Yes
T-drh105	TALE_021	pDaS362	HD NG NG HD HQ NN NG NG NG HD HD NI HD NI NG HD NG	Yes
T-drh106	TALE_038	pGrK442	HD NG NG HD NG NN NG NG NG HD HD NI HD NI NG HD NG	Yes
T-drh107	TALE_037	pGrK441	HD NG NG HD N* NN NG NG NG HD HD NI HD NI NG HD NG	Yes
T-drh108	TALE_055	pGrK506	HD NG NG HD S* NN NG NG NG HD HD NI HD NI NG HD NG	Yes
T-drh109	TALE_056	pGrK501	HD NG NG HD NV NN NG NG NG HD HD NI HD NI NG HD NG	Yes
T-drh201	TALE_002	---	NG NG NI HD NG NN HD NG NN HD NG HD HD HD NN HD NG	No
T-drh202	TALE_054	pGrK471	NG NG NI HD NG NN HD NG NN HD NG HD HD HD NN HD NG	Yes
T-drh203	TALE_095	pGrK585	NG NG NI HD NG NN HD NG NN NG NG HD HD HD NN HD NG	Yes
T-drh204	TALE_094	pGrK582	NG NG NI HD NG NN HD NG NN N* NG HD HD HD NN HD NG	Yes
T-drh205	TALE_116	pGrK631	NG NG NI HD NG NN HD NG NN S* NG HD HD HD NN HD NG	Yes
T-drh301	TALE_028	pGrK430	NI HD NG NN HD NG NN HD NG HD HD HD NN HD NG NN HD NG HD	Yes
T-drg101	TALE_003	pDaS237	HD NN NG HD HD NI NI NG NI NN HD NG NG HD NG	No

5.2 Methods

5.2.1 General Methods

5.2.1.1 Liquid culture

Liquid cultures were handled in glass Erlenmeyer flasks or conical plastic tubes. Sterile LB medium was supplemented with 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol, 12.5 µg/mL tetracycline and 100 µg/mL spectinomycin, respectively. Medium was inoculated with a single bacterial clone and incubated at 37 °C at 200 rpm shaking for overnight.

5.2.1.2 Plate culture

Sterile LB-Agar was molten supplemented with the respective antibiotic (50 µg/mL carbenicillin, 34 µg/mL chloramphenicol, 12.5 µg/mL tetracycline and 100 µg/mL spectinomycin, respectively) and poured into a plastic petri dish. When solid, bacterial culture was plated close to a flame using glass beads.

5.2.1.3 Preparation of electro competent bacteria

800 ml of LB medium were inoculated with 8 ml of bacterial overnight culture and incubated at 200 rpm shaking and 37 °C till an $OD_{600} = 0.4$ was reached. The culture was cooled on ice. In the further procedure the cultures were kept on ice as much as possible. Culture was centrifuged at 4 °C (Eppendorf 5804R) for 10 min at 4000 rpm. Supernatant was discarded and the pellet was resuspended in 320 ml of cold sterile MilliQ-Water with a Pipette on Ice. Culture was centrifuged at 4 °C (Eppendorf 5804R) for 15 min at 4000 rpm and the supernatant was discarded and the pellet was resuspended in 160 ml of cold sterile MilliQ-Water supplemented with 10 % glycerol. Then culture was centrifuged at 4 °C (Eppendorf 5804R) for 15 min at 4000 rpm and the supernatant was discarded and the pellet was resuspended in 2.4 ml cold sterile MilliQ-Water supplemented with 10 % glycerol. A volume of 25 µL was transferred to 0.2 ml PCR tubes on ice and immediately frozen in liquid Nitrogen. Storage was performed at -80 °C.

5.2.1.4. Preparation of chemical competent bacteria

5.2.1.4.1 Standard procedure

800 ml of LB medium were inoculated with 8 ml of bacterial overnight culture and incubated at 200 rpm shaking and 37 °C till an $OD_{600} = 0.5$ was reached. The culture was cooled on ice. In the further procedure the cultures were kept on ice as much as possible. Culture was centrifuged at 4 °C (Eppendorf 5804R) for 10 min at 4000 rpm. Supernatant was discarded and the pellet was resuspended in 80 ml of cold 100 mM $MgCl_2$ with a Pipette on Ice. Culture was centrifuged at 4 °C (Eppendorf 5804R) for 15 min at 4000 rpm and the supernatant was discarded and the pellet was resuspended in 80 ml of cold 50 mM $CaCl_2$ and incubated on Ice. Then culture was centrifuged at 4 °C (Eppendorf 5804R) for 15 min at 4000 rpm and the supernatant was discarded and the pellet was resuspended in 4 ml 50 mM $CaCl_2$ supplemented with 15 % glycerol. A volume of 50 μ L was transferred to 0.2 ml PCR tubes on ice and immediately frozen in liquid Nitrogen. Storage was performed at -80 °C.

5.2.1.4.2 Procedure for BL21(DE3) and BL21(DE3)Gold strain

800 ml of SOB⁺⁺ medium were inoculated with 5 ml of bacterial overnight culture and incubated at 200 rpm shaking and 25-30 °C till an $OD_{600} = 0.5$ was reached. The culture was chilled on ice for at least 10 min. In the further procedure the cultures were kept on ice as much as possible. Culture was centrifuged at 4 °C (Eppendorf 5804R) for 10min at 4000 rpm. Supernatant was discarded and the pellet was resuspended in 200 ml of cold TB-Buffer with a Pipette on Ice and incubated on ice for 10 min. After centrifugation at 4 °C (Eppendorf 5804R) for 10 min at 4000 rpm the supernatant was discarded and the pellet was resuspended in 30 ml of cold TB-Buffer and 2.24 ml cold DMSO were added and incubated for 10 min on ice. A volume of 200 μ L was transferred to 0.2 ml PCR tubes on ice and immediately frozen in liquid Nitrogen. Storage was performed at -80 °C.

5.2.1.5 Transformation

5.2.1.5.1 By Heat shock

Up to 5 µL Plasmid solution or crude reaction mixture are placed in a 1.5 mL micro centrifuge tube and cooled on ice. 25-200 µl bacterial cells suspension (thawed on ice) are added and mixed gently by pipetting and incubated on ice for 30 min. Next heat shock at 42 °C is performed for 30-45 sec and the tube is subsequently placed on ice for another 2 min. 1 ml pre-warmed SOC medium (37 °C) is then added and suspension is incubated at 37 °C and shaking at 600 rpm for 1h. Finally a part of the suspension is plated on LB-Agar plates with corresponding antibiotic using glass beads.

5.2.1.5.2 By electroporation

Up to 1 µL Plasmid solution is placed in a 1.5 mL micro centrifuge tube and cooled on ice 25 µl bacterial cells suspension (thawed on ice) are added and mixed gently by pipetting and transferred to a precooled 1 mm Electroporation Cuvette and pulsed with 1800 V for approximately 5 ms. 1 ml pre-warmed SOC medium (37 °C) is added immediately and suspension is incubated at 37 °C and shaking at 600 rpm for 1h. Finally a part of the suspension is plated on LB-Agar plates with corresponding antibiotic using glass beads.

5.2.1.6 PCR Purification

For PCR purification, the GeneJET PCR Purification Kit was used according to the manufacturer's protocol. Specifically, the crude reaction mix, binding buffer and isopropanol were mixed in a 1:1:1 ratio and loaded on a PCR purification column. After centrifugation for 1 min at 14000 rpm, 700 µl wash buffer were loaded on the column followed by centrifugation at 14000 rpm for 1 min. After an additional centrifugation at 14000 rpm for 1 min, the DNA was eluted by loading nuclease-free water on the column followed by incubation at ambient temperature for 1 min and centrifugation for 1 min at 14000 rpm.^[13]

5.2.1.7 Agarose-Gel Electrophoresis

Agarose is suspended in 0.5xTBE Buffer at 0.75 to 3 % and is molten in the microwave. When cold enough it is poured in to the gel casting chamber and when solidified placed into the running chamber filled with 0.5xTBE buffer. Sample is mixed with 1/5 volume of 6x loading dye (50% glycerol supplemented with bromophenol blue and xylene cyanol) and loaded on the agarose gel. Running was performed at 120 V, and 30 mA for 1 to 2.5 h. Afterwards the gel was stained with ethidium bromide and visualized by UV light.

5.2.1.8 SDS-Gel Electrophoresis

Proteins were analyzed on 8% SDS-gels by PAGE. The resolving gel was prepared from 3808 μ l MilliQ-water, 1092 μ l Acrylamide concentrate (40 %), 420 μ l Tris (pH = 8.8), 66.7 μ l SDS solution (10 %), 66.7 μ l APS solution (10 %) and 6.67 μ l TMEDA. It was directly poured into between two glass plates and covered with isopropanol, after 30 min the isopropanol was removed and the resolving gel was covered with stacking gel prepared from 1000 μ l MilliQ-water, 208 μ l Acrylamide concentrate (40 %), 420 μ l Tris (pH = 6.8), 16.7 μ l SDS solution (10 %), 16.7 μ l APS solution (10 %) and 1.67 μ l TMEDA.

Samples were mixed with a 1/3 volume of SDS-PAGE Loading buffer, incubated at 95 °C for 5 min and subsequently loaded on the gel and running was performed at 120 V, and 30 mA for 1 to 1.5 hours. Afterwards the gel was stained with brilliant blue.

5.2.1.9 Plasmid Isolation

For plasmid isolation, the GeneJET Plasmid Miniprep Kit was used according to the manufacturer's protocol. Specifically, 5 ml overnight culture were pelleted and the supernatant was removed. The pellet was resuspended in 250 μ l resuspension buffer and transferred to a 1.5 ml micro centrifuge tube. Next 250 μ l Lysis solution were added and tubes were inverted 6 times. Latest after 5 min 350 μ l Neutralization buffer were added and the tubes were inverted 6 times. Centrifugation was performed for 5 min at 14000 rpm and the supernatant was subsequently loaded on the supplied columns and vacuum was applied from below and washing was performed twice by addition of 750 μ l Wash buffer. Next the columns were placed in a collecting tube and centrifuged at 14000 rpm for 1 min. Finally the columns were placed in a 1.5 ml micro centrifugation tube and 60 μ l MilliQ-water were applied and incubated for 1 min. The plasmids were then eluted by centrifugation at 14000 rpm for 1 min.

5.2.1.10 DNA Sequencing

For sequencing 10 μ l of a 30-75 ng DNA solution supplemented with 2.5 μ M primer were put into a 1.5 ml micro centrifuge tube and sent to GATC. Sequence traces were analyzed with Finch TV software.

5.2.2 Construction of TALE expression plasmids

5.2.2.1 With Non-RVD variability

Fragments containing the N- and C-terminal sequences of a *Xanthomonas axonopodis* pathovar *citri* TALE^[107] with variable numbers of internal repeats were amplified from plasmids TALEN1297, TALEN1257 and TALEN1258^[89] (Addgene plasmids 32279, 32280, 32281) with primers oDaS308 and oDaS231 and cloned into the KpnI and EagI sites of plasmid pET32a-TRX (Addgene plasmid 11516). This resulted in plasmids pET_TRX_TAL1257_short, pET_TRX_TAL1258_short and pET_TRX_TAL1297_short coding for the respective TALE proteins T-drh101, T-drh201 and T-drg101, respectively with N-terminal thioredoxin (TRX) tag and S-tag, and a C-terminal His6 tag.

5.2.2.2 Without Non-RVD variability by TALE assembly

TALEs were assembled according to a previously published protocol^[90] with slight adjustments using pTRX_ENTRY in golden gate 2 reactions. The respective repeat units from the module vectors were cloned into the respective pFUS array vector by repetitive digestion at 37 °C for 5 min with 5 U BsaI restriction enzyme and ligation at 16 °C for 10 min with 200 U T4 DNA ligase in a 10 µl total volume of T4 DNA ligase buffer for 10 cycles with final digestion step for 5 min at 50 °C and enzyme heat inactivation at 80 °C for 5 min. Next 0.5 µl of a 10 mM ATP solution and 0.5 µl of Plasmid-Safe nuclease were added and the reaction mixture was incubated at 37 °C for 1 h to digest all non-circular DNA. After heat inactivation of the nuclease for 20 min at 70 °C GH371 bacteria were transformed with 2.5 µl of the reaction mixture. Next day for each construct 4 single clones were picked and each suspended in 20 µl MilliQ-water and a bit of this suspension was saved by spotting on a LB-Agar plate supplemented with spectinomycin. Each of the bacterial suspensions was used for colony PCR to identify correctly assembled variants. For this 1 µl of the respective bacterial suspension was diluted into a 25 µl total volume of Thermopol buffer supplemented with dNTPs at 250 µM each, 100 nM of primer oSaB_734 and oSAB_735 and 40 mU Taq polymerase. The reactions mixture was applied to following conditions: 2 min 94 °C, then 35 cycles of 30 sec 94 °C, 30 sec 55 °C, 1:45 min 68 °C and final extension step at 68 °C for 2 min and infinite hold at 4 °C at the end. Results were monitored on agarose gel and overnight cultures were prepared for each array plasmid from one correct colony of the safe plate. Next day plasmids were isolated and array plasmids and last repeat sequence were cloned into the expression vector pTRX_ENTRY using the same technique as above but using BsmBI restriction enzyme and only one digestion and restriction cycle and final digestion step at 37 °C for 15 min. No plasmid safe digest was performed. Bacterial strain BL21(DE3) or BL21(DE3)Gold were transformed with 2.5 µl of the

reaction mixture and plated on LB-Agar plates supplemented with carbenicilin. Next day colonies were picked and saved as above and colony PCR was performed as above using primers oSaB_890 and oDaS_247 and extension time of 3 min. Overnight cultures from clones with correct insert size have been prepared and next day plasmids were isolated and sequence identity has been verified by sanger sequencing using primers oDaS_315, oSaB_890, oJoP_410 and oDaS_247.^[17]

5.2.3 Creation of monomer plasmids with non-canonical RVDs

Module plasmids pS*5, pNV5, pS*10 were created by a standard cloning approach. Oligonucleotides oGrK_1272 / oGrK_1273, oGrK_1274 / oGrK_1275 or oGrK_1472 / oGrK_1473 were hybridized and the overhangs were filled up using Klenow fragment of *E. coli* DNA polymerase I (3'-5'-exo-). After PCR purification, the DNA fragments were digested using NcoI and XhoI, purified and cloned into the backbone of pHD5 using T4 DNA Ligase.^[17]

5.2.4 Protein expression and Purification

E. coli BL21(DE3)Gold or BL21(DE3) was transformed with the respective TALE expression plasmid and a single clone was picked and was grown overnight in LB media supplemented with 50 µg/ml carbenicilin (and 12.5 µg/ml tetracycline) at 37 °C and 200 rpm shaking. This culture was diluted 50-fold into LB medium supplemented with 50 µg/ml carbenicilin and the culture was incubated at 37 °C and 200 rpm shaking until an OD₆₀₀ of ~0.4 was reached. Protein expression was induced with 0.2 mM IPTG and the culture was harvested after 5 h of incubation by centrifugation (10 min, 3320 x g, room temperature). The pellet was lysed with Taq-lysis-buffer containing 1 mM PMSF and 50 µg/ml lysozyme and shaking at room temperature at 1400 rpm for 30 min. The suspension was pelleted by centrifugation (20 min, 7690 x g, room temperature) and the supernatant was discarded. Protein was extracted by strong shaking at room temperature for 2 h from pellet under denaturing conditions using Buffer Z and bound to Ni-NTA according to the manufacturer's protocol. Ni-NTA was washed 2 times with 4 x PBS-Buffer containing 8 M urea, 4 times with Buffer Z containing 20 mM imidazole, 1 time Buffer Z containing 50 mM imidazole and eluted 3 times with Buffer Z containing 500 mM imidazole. Pooled elution fractions were added to a dialysis tube and dialyzed against TALE Storage Buffer supplemented with 1 mM DTT. Purity and quantity of TALE proteins was analyzed by SDS PAGE and staining with Brilliant blue and a BCA assay and proteins were stored in aliquots at -80 °C in TALE storage buffer.^[17]

5.2.5 Electromobility Shift Assays

Oligonucleotides Er-drh101 and Er-drh103 were 5'-³²P-end labeled using γ -³²P-ATP and T4 polynucleotide kinase and purified using a G-25 gel filtration column. Oligonucleotides were hybridized pairwise (Et-drh101/Er-drh101; Et-drh102/Er-drh101; Et-drh103/Er-drh103; Et-drh104/Er-drh103) at a concentration of 12.5 nM, respectively, by incubating at 95 °C for 5 min and subsequently at room temperature for 30 min in Hybridization buffer supplemented with 100 ng/ μ L salmon sperm DNA. For TALE-binding, to 6 μ L of the hybridized DNA duplexes, 6 μ L of TALE storage buffer supplemented with 1 mM DTT with varying concentrations of TALE proteins were added and mixtures were incubated at room temperature for 30 min. Mixtures were loaded onto an analytical, non-denaturing PAGE-gel (0.5 mm, 50 cm length) that was run in 0.25 x TB buffer at 12 W and room temperature. Gel was dried and data recorded on a phosphorimager.^[13]

5.2.6 Primer Extension Reactions

Templates as shown in figures and primer p-drh101, p-drh105, p-drh106, p-drh107, p-drh201 or p-drg101 (5'-³²P-end labeled as above) were diluted from 5x stocks (500 nM and 166.5 nM, respectively) into 6 μ L Hybridization buffer supplemented with or without 100 ng/ μ L salmon sperm DNA. to final concentrations of 100 nM and 33.3 nM and hybridized by incubating at 95 °C for 5 min and subsequently at room temperature for 30 min. TALE proteins were added in 6 μ L TALE storage buffer supplemented with 1 mM DTT at varying concentrations as shown in figures and mixtures were incubated at room temperature for 30 min. 12 μ L of a mixture of 25-500 mU Klenow fragment of E. coli DNA polymerase I (3'-5'-exo-) and 200 μ M of each dNTP in binding buffer supplemented with or without 50 ng/ μ L salmon sperm DNA and 0.5 mM DTT were added, resulting in final concentrations of 8.325 nM and 25 nM for primer and template, respectively. These mixtures were incubated at room temperature for additional 15 min. 12 μ L of PAGE loading buffer (80 % formamide, 2 mM EDTA) were added, the mixtures were incubated at 95 °C for 5 min and cooled to 4 °C. Mixtures were loaded onto a pre-run, analytical denaturing PAGE gel (9 M urea, 0.5 mm, 50 cm length) and gel was run in 1 x TBE at 120 W and 40 – 50 °C. Gel was dried and data recorded on a phosphorimager. For IC₅₀ determination, fitting was performed in Origin 8.6.0G (Originlab) using the Dose Response function without initial parameters.

To investigate the influence of additional, complex off-target DNA on TALE binding (See Fig. 17), the hybridization of templates and primer p-drh101 were conducted as above using hybridization buffer containing 0.1 – 10 μ g/ μ L salmon sperm DNA in a 6 μ L total volume. This corresponds to a mass ratio of 2.89×10^2 - 2.89×10^4 , calculated with the molecular weight of double stranded target sequence S-drh1 (10379.9 g/mol). 10 μ g/ μ L was thereby a saturating amount in terms of solubility. 6 μ L of a 1.665 μ M T-drh101

solution in TALE storage buffer were added, resulting in a TALE/target DNA ratio of 50:1. Mixtures were incubated at room temperature for 30 min. 12 μ L of binding buffer containing 0.125 to 10.0 U Klenow Fragment of *E. coli* DNA polymerase I (3'-5'-exo-) and 200 μ M of each dNTP were added and the resulting mixtures were incubated at room temperature for additional 15 to 720 min. 12 μ L of PAGE loading buffer were added, the mixtures were incubated at 95 $^{\circ}$ C for 5 min and cooled to 4 $^{\circ}$ C. Mixtures were loaded onto a pre-run, analytical denaturing PAGE gel (9 M urea, 0.5 mm, 50 cm length) and gel was run in 1 x TBE at 120 W and 40 – 50 $^{\circ}$ C. Gel was dried and data recorded on a phosphorimager.^[13]

5.2.7 qPCR Experiments on synthetic oligonucleotides

For qPCR experiments, primer extension reactions were conducted as above with primer p-drh102 replacing 5'-32P-radiolabelled primer p-drh101. Reactions were quenched with 12 μ L of 20 mM EDTA. Mixtures were diluted with water to 500 μ L and 10 μ L of streptavidin-agarose beads (pre-incubated 4 h in 20 mM NaOH) were added. Suspensions were shaken for 30 min and agarose beads were washed 4 times with 20 mM NaOH, 2 times with water and 2 times with PBS buffer (each 100 μ L). To not introduce bias into copy numbers of primer extension products, streptavidin agarose was used in a 400-fold excess of biotin binding sites over molecules of added primer. For PCR reactions, agarose beads were diluted in PBS to contain 1/10⁵ of above primer extension reactions in 10 μ L of bead suspension. PCR reactions (50 μ L) were conducted with using 25 μ L GoTaq qPCR Master Mix containing BRYT Green dye, 200 nM of each primer qp-drh1_fwd and primer qp-drh1_rev and 10 μ L of above diluted bead suspension in a BioRad i-Cycler with filters FAM-490 using the following program: 2 min 95 $^{\circ}$ C, then 60 cycles of 0:15 min 95 $^{\circ}$ C, 1:00 min 60 $^{\circ}$ C. Reproduction experiment to obtain standard deviations of C_T values included the complete experimental procedure starting from primer extension reactions. For quantification of copy numbers present in agarose bead PCR reactions, a calibration curve was generated using the same PCR conditions and dilutions of Pt-drh101. Copy numbers were obtained by linear regression. PCR products were additionally analyzed by agarose gel electrophoresis and staining with ethidium bromide.

5.2.8 Extraction of total DNA

The tale fin of a male individual of Zebrafish (*Danio rerio*, Konstanz wild type) was used for extraction of total DNA. To the tissue, 20 µl proteinase K and 180 µl Buffer ATL were added and it was dissolved at 56 °C and 1400 rpm shaking. Next, 200 µl buffer AL and 200 µl ethanol were added and the mixture was loaded on a DNeasy Mini spin column and centrifuged for 1 min at 8000 rpm. The column was washed with 500 µl of buffer AW1 and 500 µl of buffer AW2 and DNA was eluted using 2 x 200 µl nuclease-free water. DNA was left unfragmented and was quantified by UV measurement (260 nm). To analyze the sequence of the *hey2* locus, a 541 bp part was amplified with primers oDaS760 and oGrK796 (S-Fig. S8). The sequence matched Genbank entry NM_131622.2 with a single point mutation at position 1168, outside of the relevant region of later experiments (S-Fig. S9).^[13]

5.2.9 Genomic 5mC-Detection Procedure using Bisulfite Sequencing

95.4 µg of total DNA were enzymatically fragmented by digestion with 250 U HaeIII in CutSmart Buffer for 2 h at 37 °C in a 500 µl total volume and purified following the PCR Purification protocol. 25 µg of the fragmented total DNA were enzymatically methylated using 20 U M.SssI CpG methyltransferase in 200 µl NEB2 Buffer supplemented with 640 µM S-adenosylmethionine by incubation at 37 °C for 14 h (after 4 h of reaction time, additional 4 µl S-adenosylmethionine (32 mM) were added) and purified following the PCR Purification protocol. Fractions of both original and enzymatically methylated total DNA were bisulfite converted according to the manufacturer's instructions. For this purpose, 85 µl bisulfite mix and 35 µl DNA protect buffer were added to 50 ng DNA in 20 µl nuclease-free water and the mixture was incubated according to the following protocol: 95 °C, 5 min; 60 °C, 25 min; 95 °C, 5 min; 60 °C, 85 min; 95 °C, 5 min; 60 °C, 175 min; 20 °C, overnight. The reaction mix was supplemented with 560 µl freshly prepared buffer BL (containing 10 µg/ml carrier RNA) and loaded on an EpiTect spin column. After centrifugation for 1 min at 14000 rpm the column was washed with 500 µl BW buffer and then incubated for 15 min at ambient temperature with 500 µl BD buffer and subsequently centrifuged for 1 min at 14000 rpm. The column was washed twice with 500 µl BW buffer subsequently centrifuged at 14000 rpm for 1 min and dried at 56 °C for 5 min. After elution of the converted DNA with 20 µl nuclease-free water the same region as above was amplified via a nested, repeated PCR approach using 5 µl of converted DNA solution as template and primer pair oDaS754 / 751 for first PCR in a 100 µl total volume and 10 µl of first PCR mix as template and primer pair oDaS756 / 757 for second PCR in a 100 µl total volume (S-Fig. S6). The PCR reactions were performed using Taq-Polymerase in 1x NEB ThermoPol Buffer and following cycling

parameters: 95 °C, 1 min then 45 cycles 95 °C, 20 sec; 55 °C, 30 sec; 68 °C, 60 sec, followed by a final extension step at 68 °C for 5 min. After the second PCR the mixture was purified according to the PCR purification protocol. The forward strand of these PCR products was sequenced by Sanger sequencing using primer oDaS757.

These sequencings confirmed the absence of CpG methylation at the target sequence S-drh3 in original total DNA and quantitative methylation at the S-drh3 target sequence in enzymatically methylated DNA (S-Fig. S7).^[13]

5.2.10 Streptavidin-Agarose bead preparation

1 ml Streptavidin-Agarose beads (Suspension in 0.01 M sodium phosphate, 0.15 M NaCl, 0.02 % sodium azide, pH = 7.2, Sigma Aldrich) were centrifuged (14000 rpm, 2 min) and 500 µl of the supernatant were removed. 500 µl 1 x B&W Buffer were added, vortexed, centrifuged (14000 rpm, 2 min) and 500 µl of the supernatant were removed. This washing step was repeated two times and beads were incubated after addition of 1 mL 1 x B&W Buffer supplemented with 100 ng/µl salmon sperm DNA at room temperature and 900 rpm shaking for 30 min, subsequently centrifuged (14000 rpm, 2 min) and 1 mL of the supernatant was removed. After washing with 500 µl 1 x B&W Buffer the beads were resuspended by addition of 1.5 ml 1 x B&W Buffer yielding 2 ml of 0.5 x bead suspension.^[13]

5.2.11 Genomic 5mC-Detection Procedure using TALE-Effectors

100 ng fragmented total zebrafish DNA and Primer p-drh302 (20 fmol) were hybridized in 6 µL hybridization buffer by incubating at 95 °C for 10 min and subsequently allowing to cool down to room temperature with an initial temperature decrease of 1 °C/min. TALE proteins were added in 6 µL TALE storage buffer supplemented with 1 mM DTT at varying concentrations and mixtures were incubated at 25 °C for 60 min. 12 µL of a mixture of 25 mU Klenow Fragment of E. coli DNA polymerase I (3'-5'-exo-) and 83 µM of each dNTP in binding buffer supplemented with 0.5 mM DTT were added. These mixtures were incubated at room temperature for additional 15 min. 36 µl of 20 mM EDTA, 450 µl 1 x B&W buffer supplemented with 100 ng/µl salmon sperm DNA and 10 µl of 0.5 x streptavidin-agarose beads were added and mixtures were incubated at 25 °C and 900 rpm shaking for 90 min. The mixtures were transferred to centrifugal columns and buffer was removed by centrifugation (5000 rpm, 10 sec). Beads were washed one time with 500 µl 1 x B&W buffer. Beads were transferred to a micro tube with 500 µl 1 x B&W buffer and incubated at 85 °C and 900 rpm shaking for 15 min. Subsequently, beads were transferred to a preheated centrifugal column and the buffer was removed by centrifugation (5000 rpm, 10 sec). Subsequently, beads were washed with 500 µl PCR-reconstitution buffer and resuspended in 40 µl water.

PCR reactions (25 μ L) were conducted using 12.5 μ L GoTaq qPCR Master Mix containing BRYT Green dye, 5 μ L of primer mixture containing each primer qp-drh3_fwd and primer qp-drh3_rev at 1 μ M and 7.5 μ L of above bead suspension in a BioRad i-Cycler with filters FAM-490 using the following program: 2 min 95 °C, then 80 cycles of 0:15 min 95 °C, 1:00 min 60 °C. qPCR was validated in terms of product homogeneity and selectivity by test reactions on zebrafish total DNA and salmon sperm blocking DNA and analysis by recording of melting curves and agarose gel electrophoresis. Results were reproduced in multiple independent experiments including the complete experimental procedure starting from primer extension reactions and employing two independently expressed and purified lots of T-drh301. For quantification of copy numbers present in agarose bead PCR reactions, a calibration curve was generated using the identical PCR conditions and dilutions of original and methylated fragmented total Zebrafish DNA (S-Fig. S28, no significant impact of methylation on qPCR efficiency was observed). Copy numbers were obtained by linear regression. qPCR data were subjected to a Student t-test for analysis of statistical significance. For qPCRs shown in Figure 20 in presence of KF(exo-), four replicate experiments conducted under identical conditions but fully independent (including the complete experimental procedure starting with primer extension reactions) were used. P-values were $p = 0.02$ for columns 6/7 and $p = 0.0062$ for columns 4/6. Differences of comparisons between other experimental sets were not significant ($p = 0.74$ for columns 4/5, and $p = 0.14$ for columns 5/7).^[13]

5.2.12 Reduction of fC containing DNA

4 μ l of 100 μ M Oligonucleotide were diluted by the addition of 26 μ l MilliQ-water and 10 μ l of a freshly prepared 1 M sodium borohydride solution were added. The mixture was vortexed and centrifuged briefly and kept at ambient temperature in the dark for 1 h at 1000 rpm shaking. Pressure was released every 15 min by opening and closing the cap, vortexing and brief centrifugation. Afterwards 20 μ l of a 750 mM sodium acetate solution at pH = 5 were added and reaction mixture was kept at ambient temperature for at least 10 min and subsequently purified via illustra G-25 microspin column.^[35]

6. SUPPORTING FIGURES

The Supporting Figure of this section have been partly published in supplementary information of the following publications

Angew. Chem. Int. Ed. **2014**, *53*, 6002–6006.

© 2014 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

and

ChemBioChem **2015**, *16*, 228–231.

© 2015 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

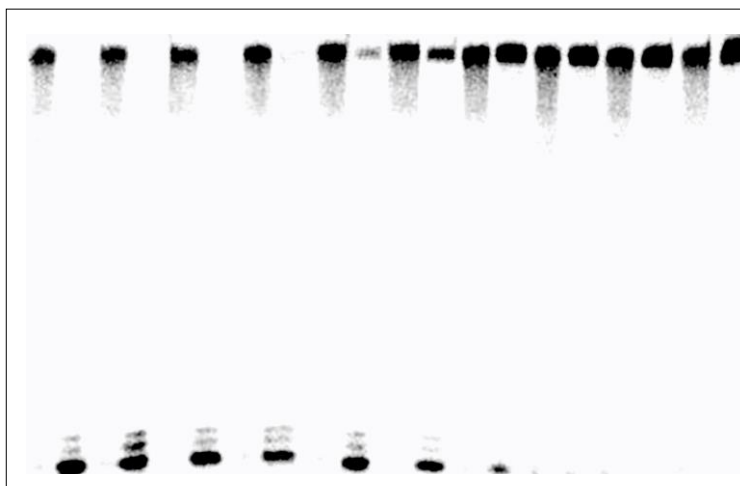
and

J. Am. Chem. Soc., **2015**, *137* (1), pp 2–5.

Copyright © 2015 American Chemical Society

All adaptations and reprints have been made with permission.

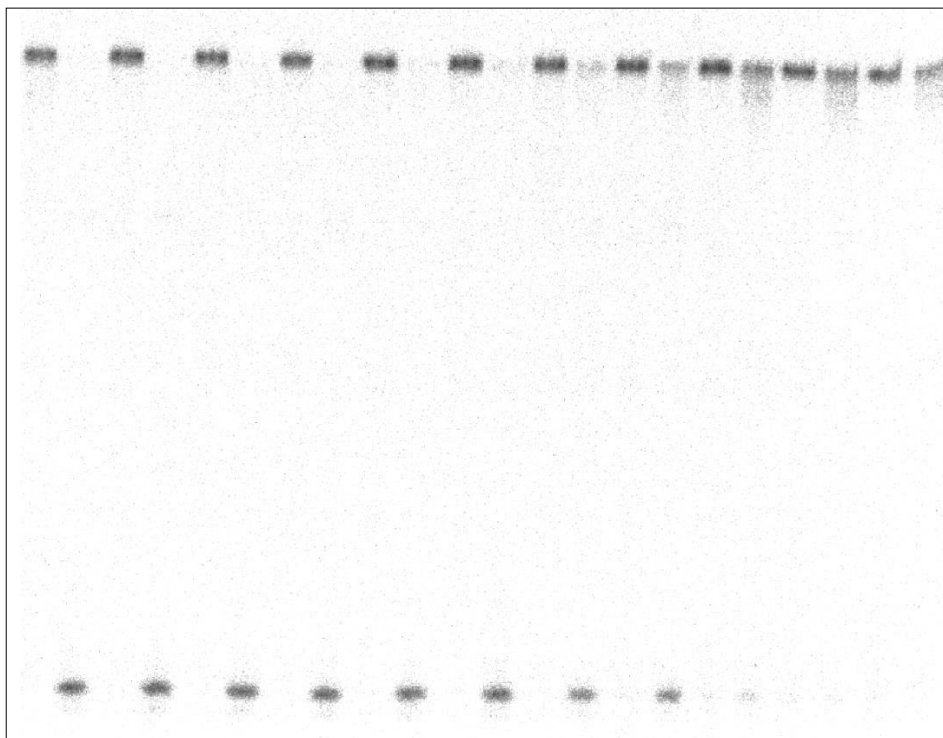
S-Figure S1: Primer Extension Gel with T-drh101 for IC₅₀ determination



TALE / DNA:	100	75	50	30	20	10	5	2.5	1	0				
c(TALE) [nM]:	833	624	416	250	167	83	42	21	8.3	0				
Pt-drh101	-	+	-	+	-	+	-	+	-	+	-	+	-	+
Pt-drh102	+	-	+	-	+	-	+	-	+	-	+	-	+	-

Amount of Klenow Fragment: 0.125 U each
 KI = 83.0 ± 7.6 nM

S-Figure S2: Primer Extension Gel with T-drh201 for IC₅₀ determination

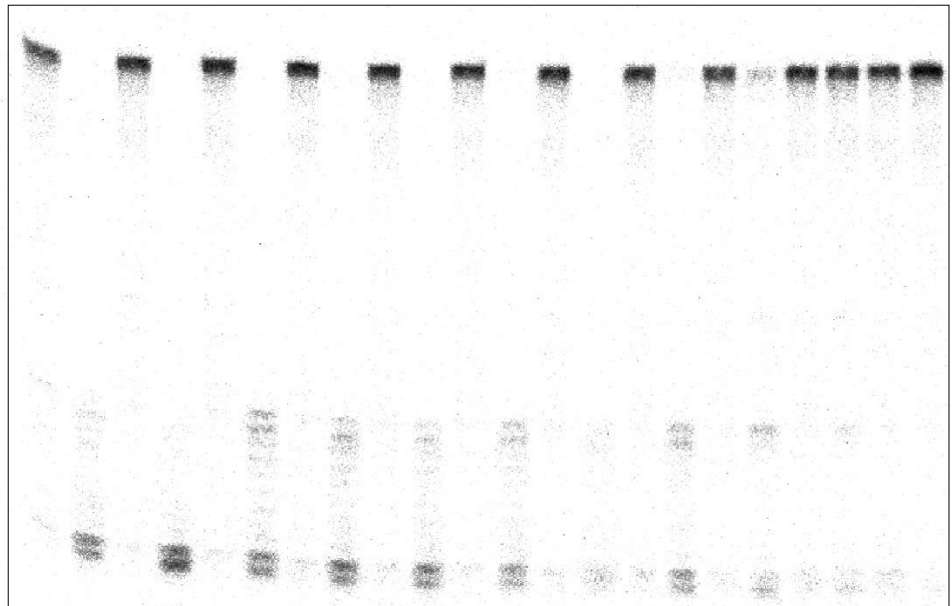


Pt-drh204	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Pt-drh201	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
TALE/DNA	100	50	25	20	15	10	5	2,5	1	0,5	0													
c(TALE) [nM]	833	416	208	167	125	83.3	41.6	20.8	8.33	4.16	0													

Amount of Klenow Fragment: 500 mU each

KI = 24.1 ± 6.6 nM

S-Figure S3: Primer Extension Gel with T-drg101 for IC₅₀ determination

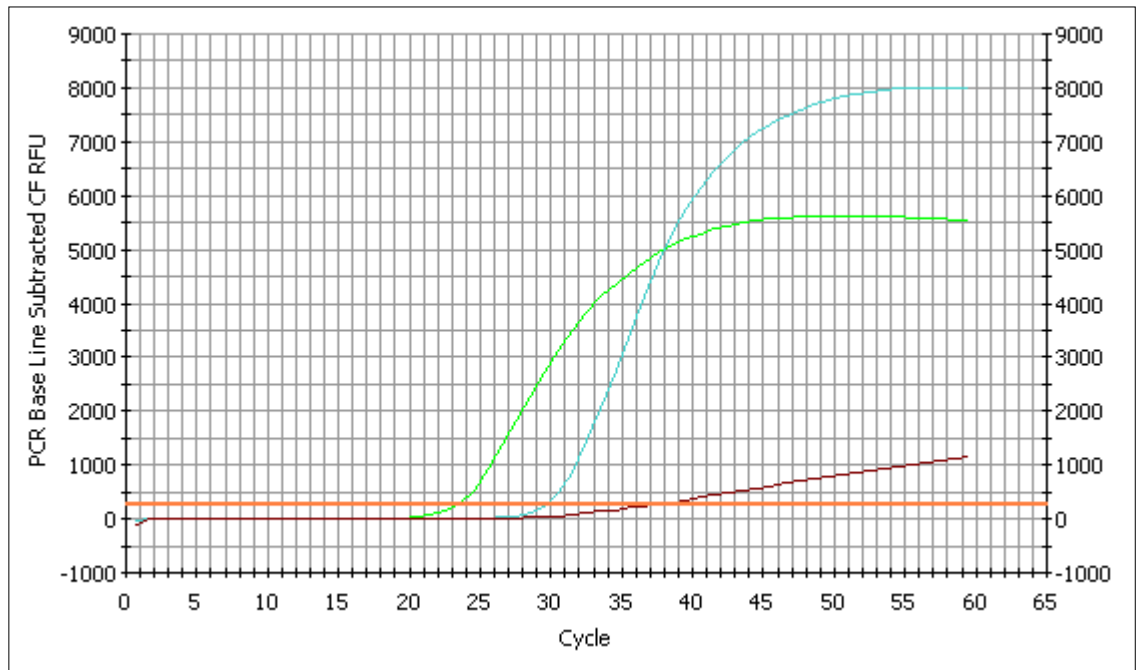


Pt-drg102	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Pt-drg101	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
TALE/DNA	200		150		100		75		50		40		30		20		10		5		0	
c(TALE) [nM]	1665		1249		833		624		416		333		250		167		83.3		41.6		0	

Amount of Klenow Fragment: 25 mU each

KI = 67.4 ± 4.0 nM

S-Figure S4: Example reaction curves of qPCR analysis of mC-dependent primer extensions

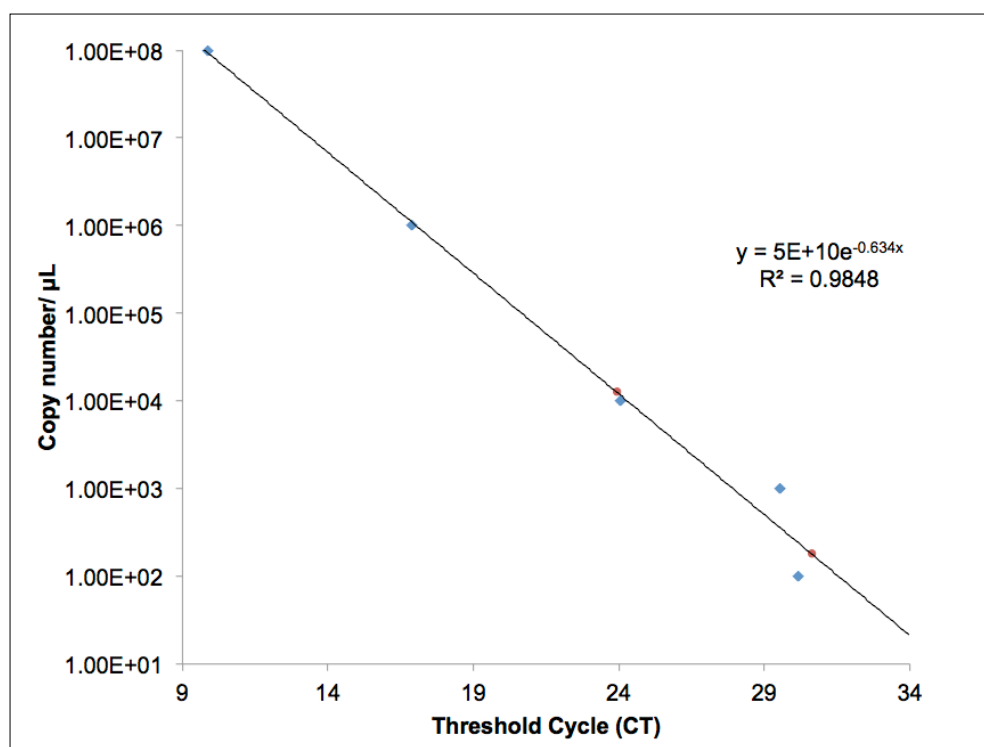


Green curve: Pt-drh102 (oGrK_465)

Blue curve: Pt-drh101 (oGrK_476)

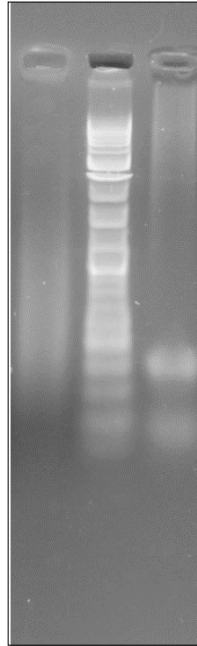
Brown curve: no template

S-Figure S5: Reaction curves of qPCR analysis for concentration calibration on Pt-drh101



C_T values of calibration PCR reactions using Pt-drh101 (oGrK_476) are shown as blue squares and of PCR reaction using agarose bead-immobilized primer extension products as red circles, respectively. An average C_T value for non-methylated template of 23.9 ± 0.9 and for methylated template of 30.7 ± 1.3 was obtained, correlating to $1.31 \times 10^4 \pm 4.9 \times 10^2$ and $1.76 \times 10^2 \pm 7.5$ molecules, i.e. a 71-fold higher formation of extension product for non-methylated DNA.

S-Figure S6: Agarose Gel of nested, repeated PCRs on bisulfite converted total DNA *D. rerio* for amplification of *hey2* locus

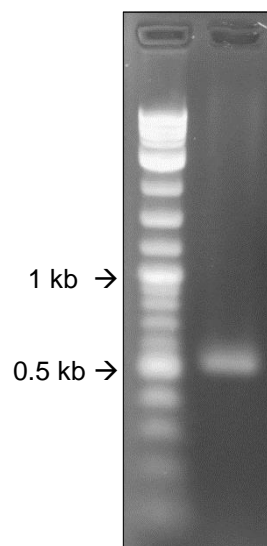


Lane 1: First PCR on bisulfite converted DNA

Lane 2: Ladder

Lane 3: PCR on 5 % of first PCR Mix of bisulfite converted DNA

S-Figure S8: Agarose Gel of PCR on *D. rerio* hey2 locus



Lane 1: DNA Ladder

Lane 2: PCR Product

S-Figure S9: Alignment of sequenced hey2 locus of *D. rerio* individual used in this study with Genbank entry NM_131622.2.

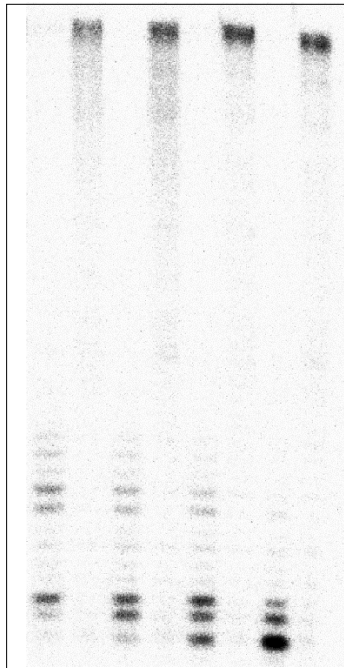
Seq_1	453	-----CCACTTCTCTGCTTTTCGTTATCAGCGACCGTTCACGCAGCAGCTGCTGCAGC	402
Seq_2	901	ACCGCTGTCCACTTCTCTGCTTTTCGTTATCAGCGACCGTTCACGCAGCAGCTGCTGCAGC	960
Seq_1	401	TGCAGCTCAAACCTTCCCTCTATCATTTCCCGCTGGATTCCCACTCTTCAGCCCCAGCGT	342
Seq_2	961	TGCAGCTCAAACCTTCCCTCTATCATTTCCCGCTGGATTCCCACTCTTCAGCCCCAGCGT	1020
Seq_1	341	TACAGCATCTTCAGTGGCTTCTTCCACCGTGAGCTCTTCCGTTTCCACATCCACCACATC	282
Seq_2	1021	TACAGCATCTTCAGTGGCTTCTTCCACCGTGAGCTCTTCCGTTTCCACATCCACCACATC	1080
Seq_1	281	CCAACAGAGCAGCGGGAGCAGCAGTAAACCATAACCGACCGTGGGGAAGTGAAGTGGGAGC	222
Seq_2	1081	CCAACAGAGCAGCGGGAGCAGCAGTAAACCATAACCGACCGTGGGGAAGTGAAGTGGGAGC	1140
Seq_1	221	GTTTTAAATGTTGGATTTAAATGTTGGTCGTCTTCCATGCTTTGTACATAAAGGAAAGCA	162
Seq_2	1141	GTTTTAAATGTTGGATTTAAATGTTGGACGTCTTCCATGCTTTGTACATAAAGGAAAGCA	1200
Seq_1	161	GCGGCTATTGTGCCTGCTTCGGTCAGCAGCATGGGCTTTTGTCTTCCCTCTACACTTGTGC	102
Seq_2	1201	GCGGCTATTGTGCCTGCTTCGGTCAGCAGCATGGGCTTTTGTCTTCCCTCTACACTTGTGC	1260
Seq_1	101	ACATATGCAGCGTCAAACCTAAGCCAACATTCTGGGAAGAAAAGAAAGAGTTTTTACACG	42
Seq_2	1261	ACATATGCAGCGTCAAACCTAAGCCAACATTCTGGGAAGAAAAGAAAGAGTTTTTACACG	1320
Seq_1	41	TCGCACTGTGTTGGAAACCGTAAAGGAAGTTTGTCTTCTGTT-----	1
Seq_2	1321	TCGCACTGTGTTGGAAACCGTAAAGGAAGTTTGTCTTCTGTTTAAACAGTGCCTGCATAAA	1380

Sequence 1: *D. rerio* individual (S-drh3 target sequence highlighted in blue)

Sequence 2: Genbank entry NM_131622.2

A single A->T point mutation was identified (highlighted in red)

S-Figure S10: Primer Extension Gel with T-drh102 With Different Methylation Patterns



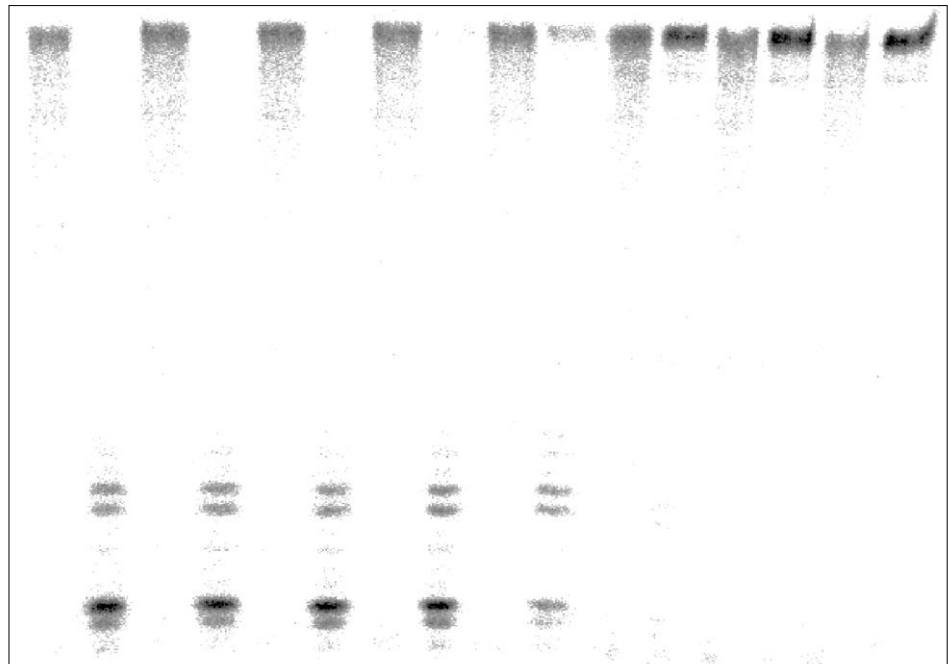
Methylation at Position

2	-	-	+	+	-	-	+	+
6	-	+	-	+	-	+	-	+
11	-	-	-	-	+	+	+	+

Concentration of TALE: 416 nM each

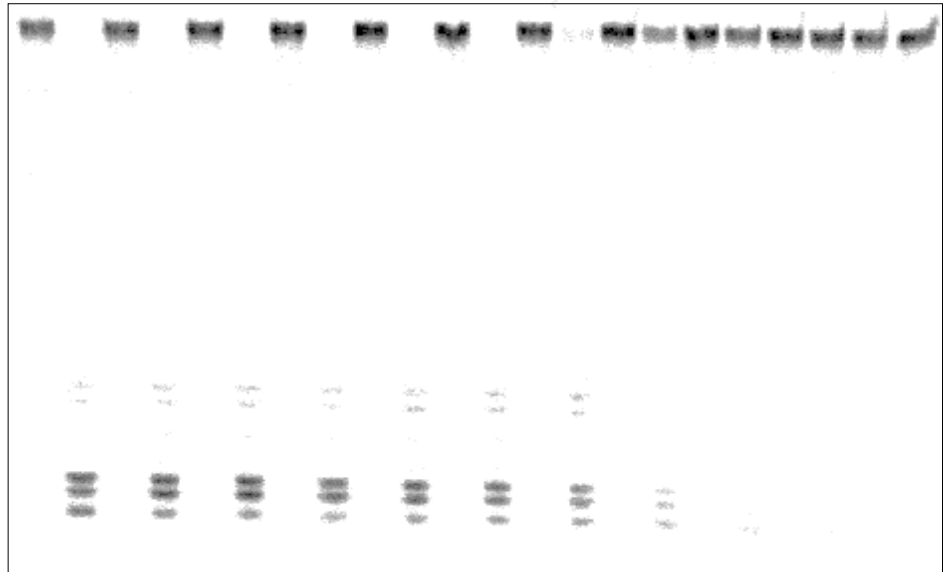
Amount of Klenow Fragment: 25 mU each

S-Figure S11: Primer Extension Gel with T-drh102 for IC₅₀ determination



Pt-drh102	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Pt-drh101	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
TALE/DNA	100		50		30		20		10		5		2,5		1	
c(TALE) [nM]	833		416		250		167		83.3		41.6		20.8		8.3	
Amount of Klenow Fragment: 25 mU each																

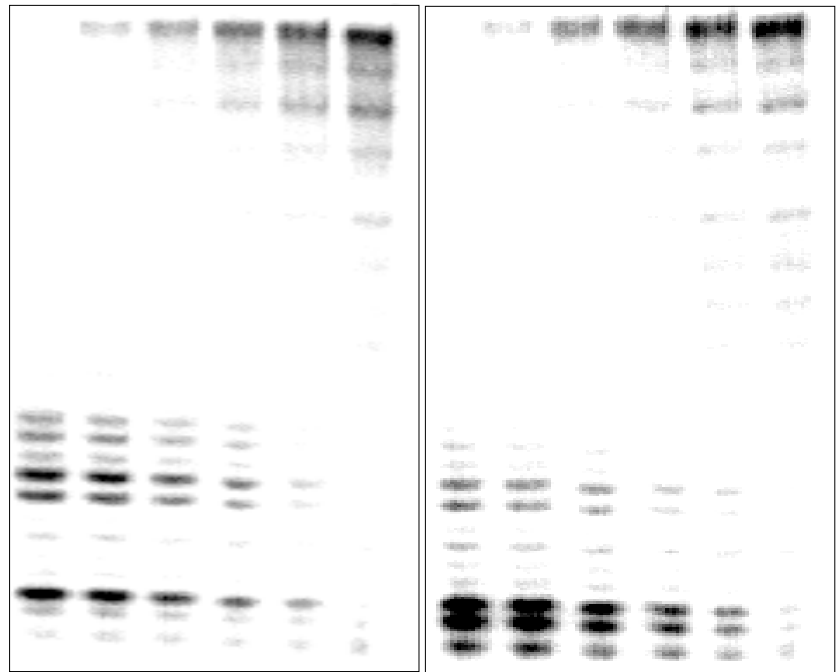
S-Figure S12: Primer Extension Gel with T-drh102 for IC₅₀ determination



Pt-drh110	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Pt-drh109	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
TALE/DNA	233		150		75		50		30		20		10		5		2,5		1		0			
c(TALE) [nM]	1940		1249		624		416		250		167		83.3		41.6		20.8		8.3		0			

Amount of Klenow Fragment: 25 mU each

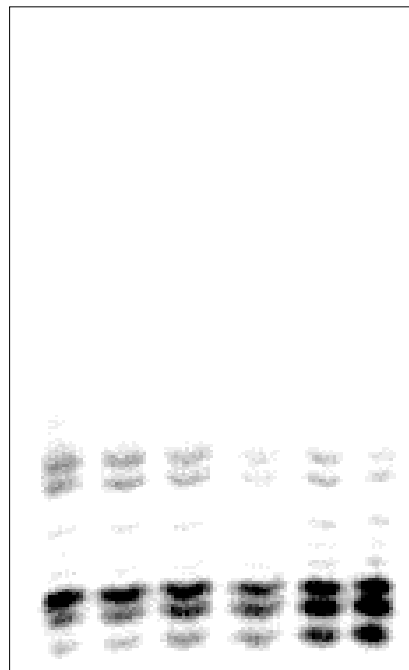
S-Figure S13: Modification Level measurements for T-drh102



Pt-drh101 [%]
 Pt-drh102 [%]
 Pt-drh109 [%]
 Pt-drh110 [%]

100 80 60 40 20 0
 0 20 40 60 80 100

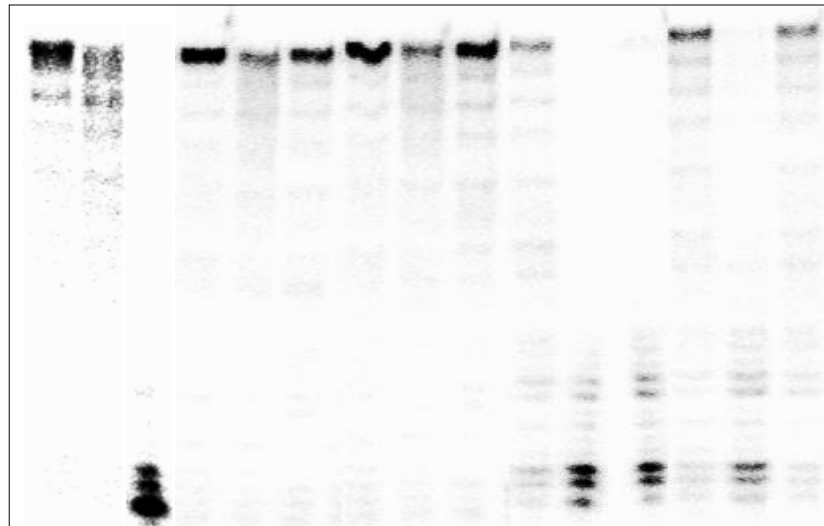
100 80 60 40 20 0
 0 20 40 60



Pt-drh101 [%]
 Pt-drh109 [%]

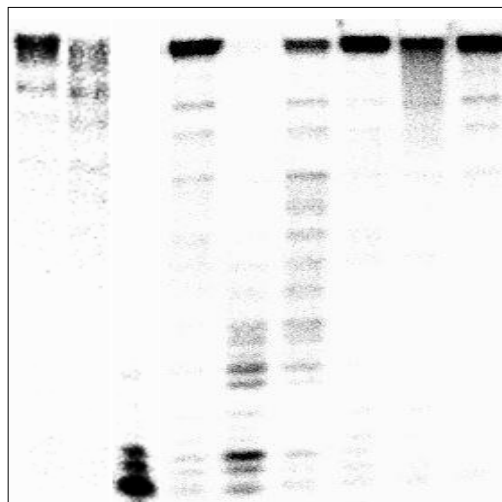
100 80 60 40 20 0
 0 20 40 60 80 100

S-Figure S14: Primer Extension Gel for RVD comparison



Pt-drh111(hmC)	+	-	-	+	-	-	+	-	-	+	-	-	+	-	-
Pt-drh102(5mC)	-	+	-	-	+	-	-	+	-	-	+	-	-	+	-
Pt-drh101(C)	-	-	+	-	-	+	-	-	+	-	-	+	-	-	+
RVD	HD			HE			HQ			N*			NG		
T-drh(xxx)	103			104			105			106			107		

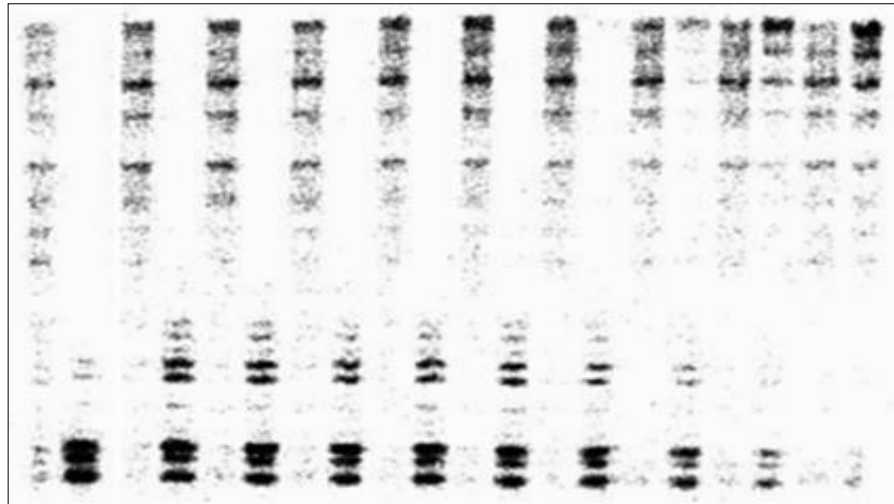
Amount of respective TALE: 416 nM each (TALE T-drh10X = 833 nM)
 Amount of Klenow Fragment: 25 mU each



Pt-drh111(hmC)	+	-	-	+	-	-	+	-	-
Pt-drh102(5mC)	-	+	-	-	+	-	-	+	-
Pt-drh101(C)	-	-	+	-	-	+	-	-	+
RVD	HD			S*			NV		
T-drh (xxx)	103			108			109		

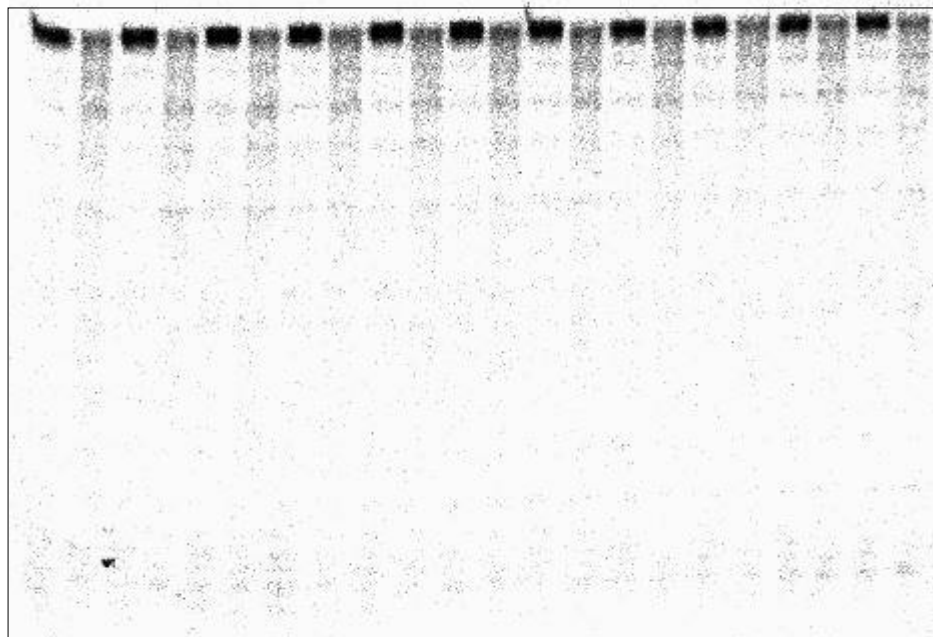
Amount of respective TALE: 833 nM each
 Amount of Klenow Fragment: 25 mU each

S-Figure S15: Primer Extension Gels with T-drh103 for IC₅₀ determination



Pt-drh102	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-		
Pt-drh101	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+		
TALE/DNA	80		60		50		40		30		20		10		5		2,5		1	
c(TALE) [nM]	666		500		416		333		250		167		83.3		41.6		20.8		8.33	

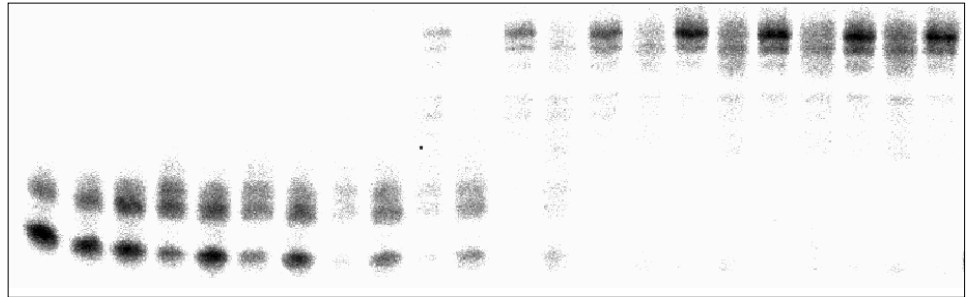
Amount of Klenow Fragment: 25 mU each



Pt-drh111	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Pt-drh102	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+
TALE/DNA	400		300		250		200		150		100		50		25		10		1		0	
c(TALE) [nM]	3330		2498		2081		1665		1249		833		416		208		83.3		8.33		0	

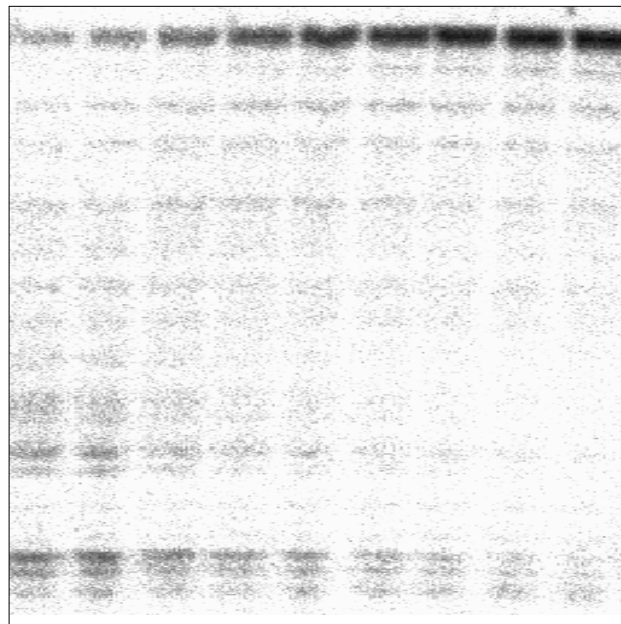
Amount of Klenow Fragment: 25 mU each

S-Figure S16: Primer Extension Gel with T-drh106 for IC₅₀ determination



Pt-drh102	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-		
Pt-drh101	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+		
TALE/DNA	222		150		100		50		35		20		10		5		2,5		1		0	
c(TALE) [nM]	1844		1249		833		416		291		167		83.3		41.6		20.8		8.33		0	

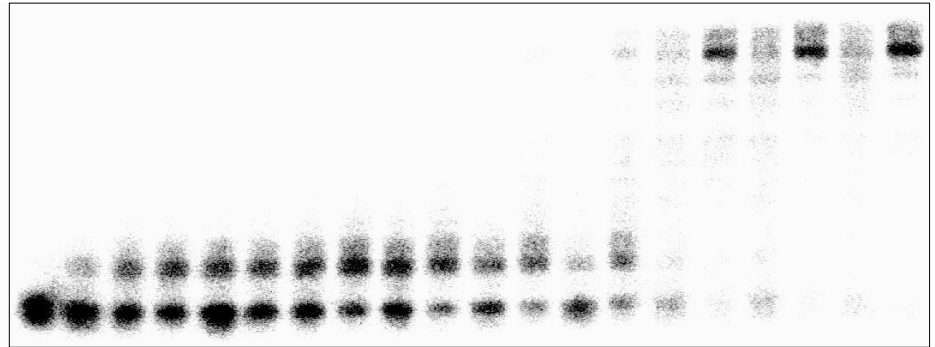
Amount of Klenow Fragment: 25 mU each



Pt-drh111	+	+	+	+	+	+	+	+	+									
TALE/DNA	222		150		100		50		35		20		10		5		1	

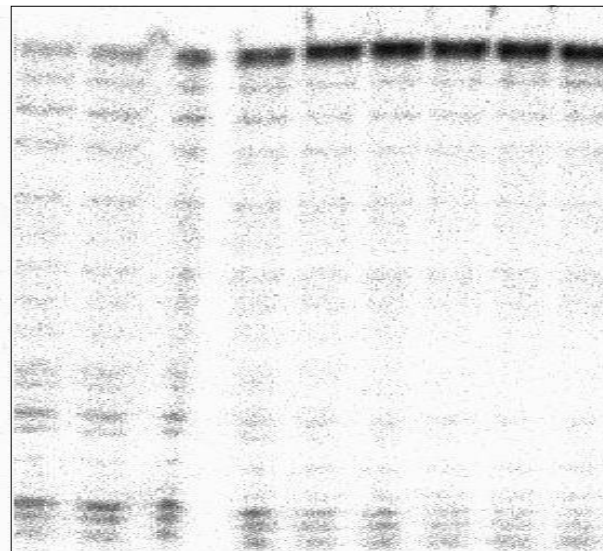
Amount of Klenow Fragment: 25 mU each

S-Figure S17: Primer Extension Gel with T-drh107 for IC₅₀ determination



Pt-drh102	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Pt-drh101	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+		
TALE/DNA	225		150		100		50		35		20		10		5		2,5		1			
c(TALE) [nM]	1873		1249		833		416		291		167		83.3		41.6		20.8		8.33			

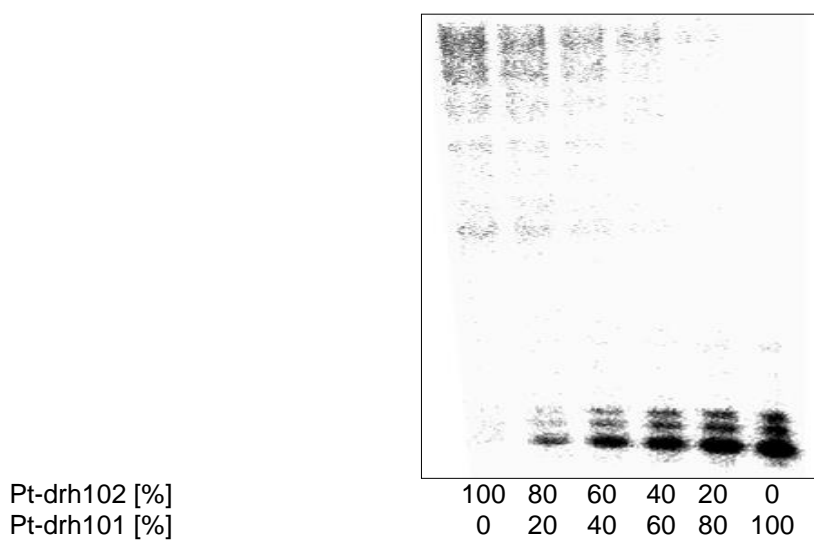
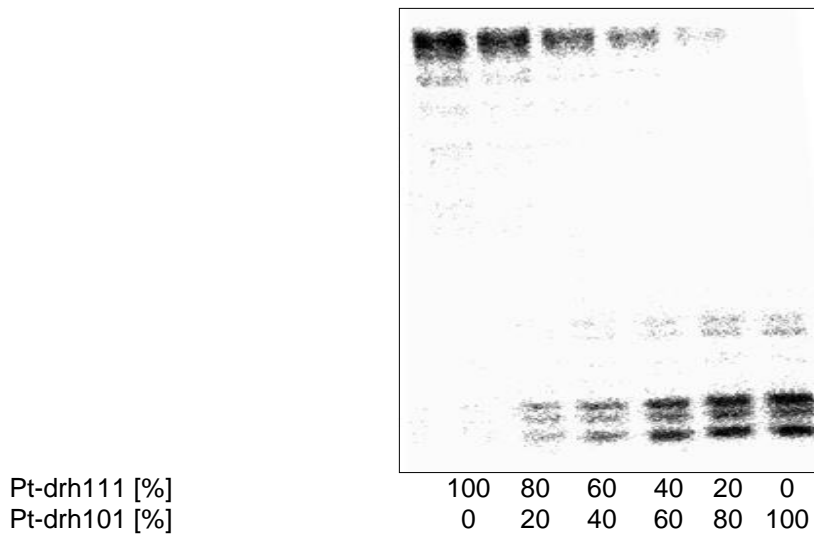
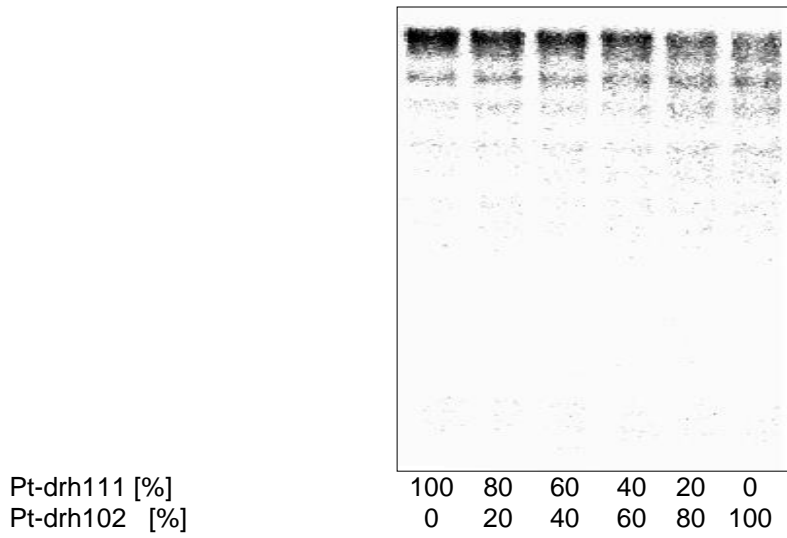
Amount of Klenow Fragment: 25 mU each



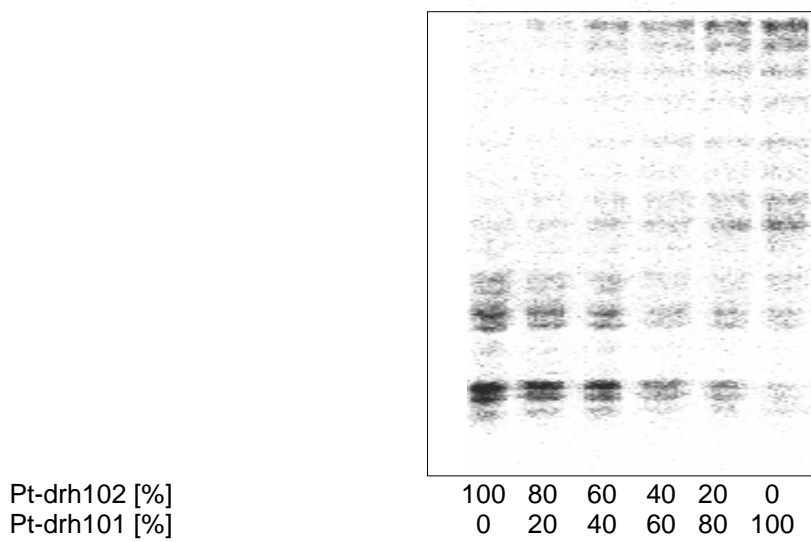
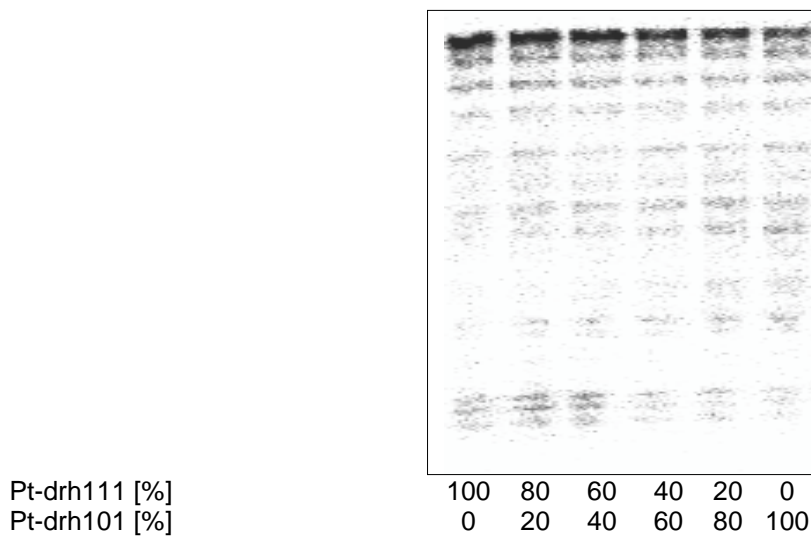
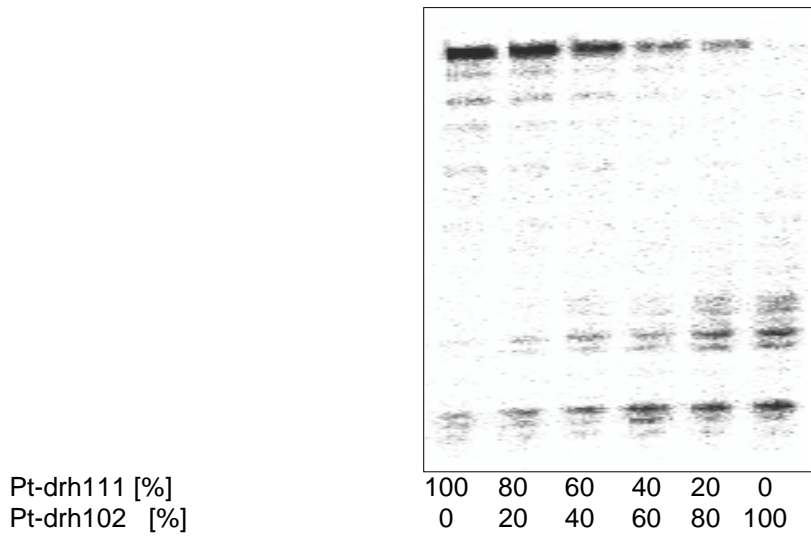
Pt-drh111	+	+	+	+	+	+	+	+	+
TALE/DNA	225	150	100	50	35	20	10	5	1

Amount of Klenow Fragment: 25 mU each

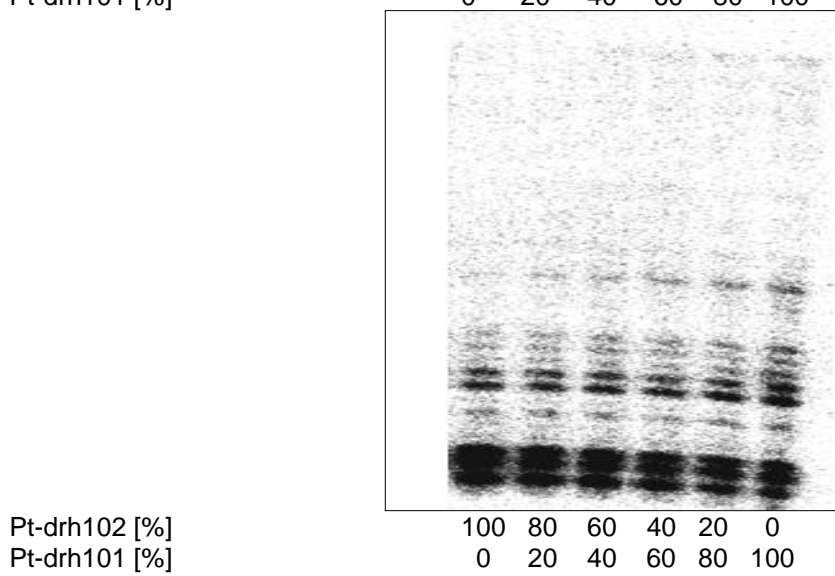
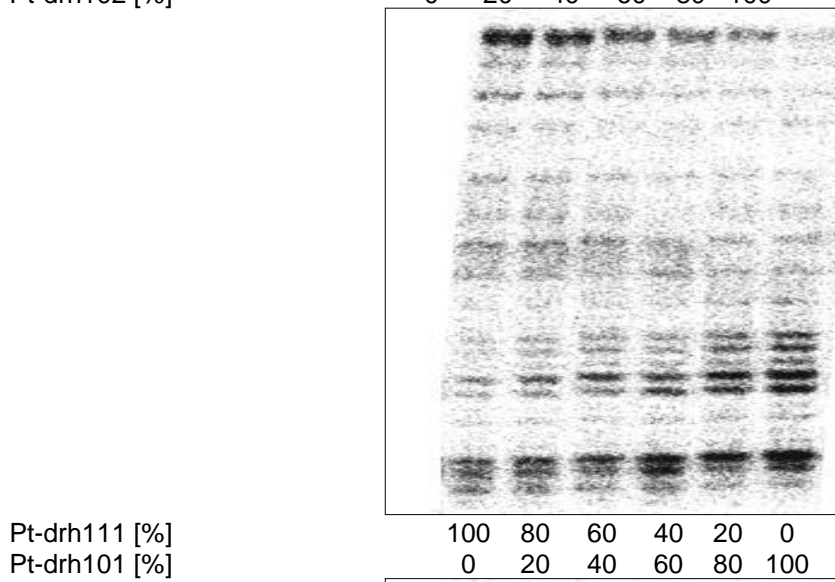
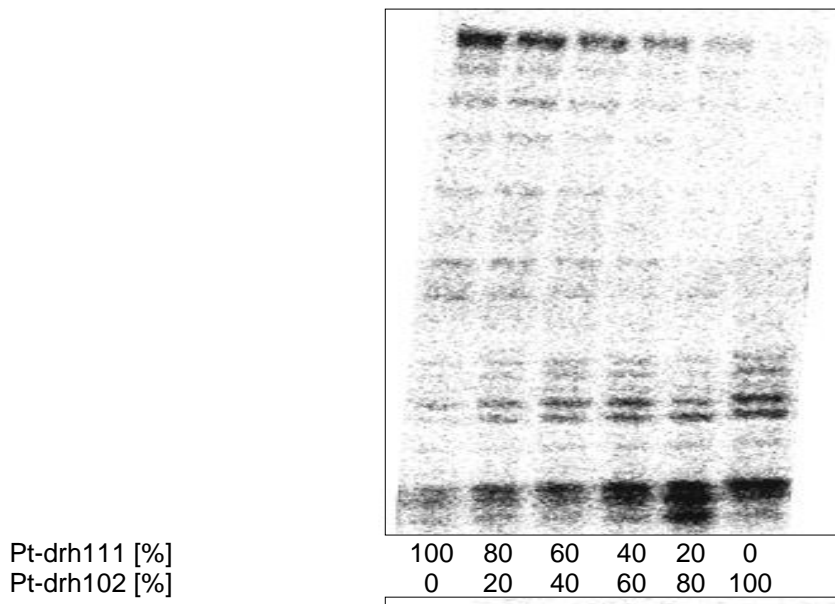
S-Figure S18: Modification Level measurement for T-drh103



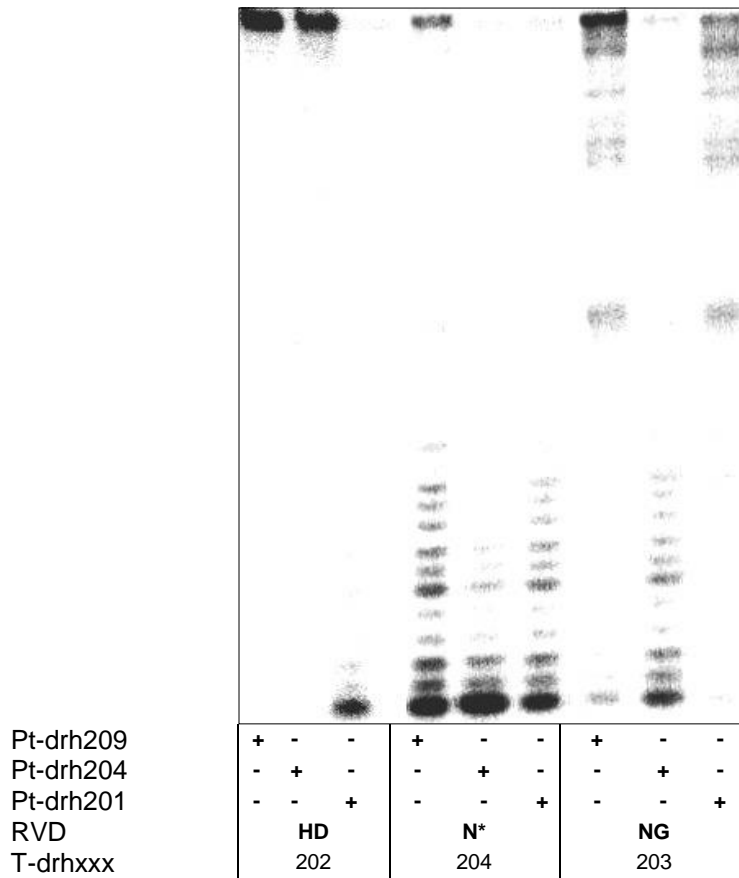
S-Figure S19: Modification Level measurement for T-drh106



S-Figure S20: Modification Level measurement for T-drh107

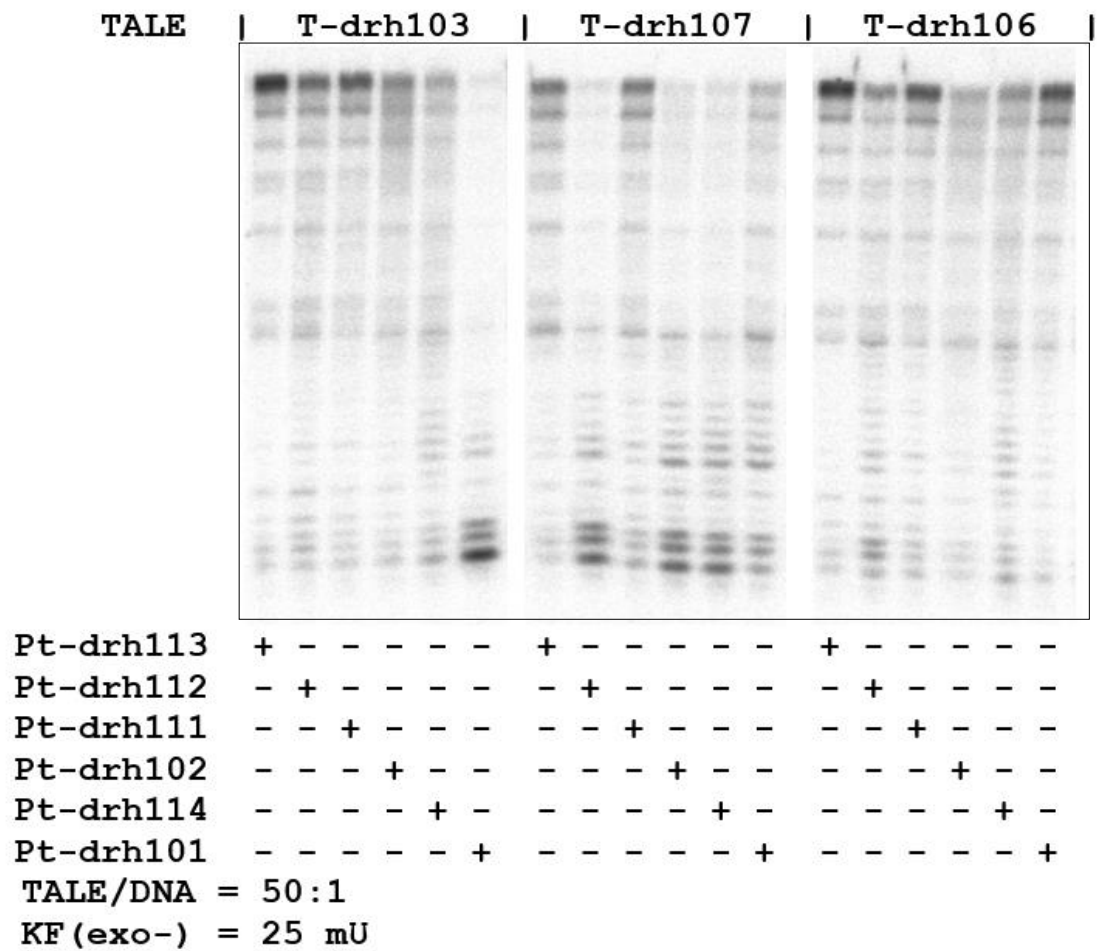


S-Figure S22: Primer Extension Gel for RVD comparison in Sequence context S-drh2

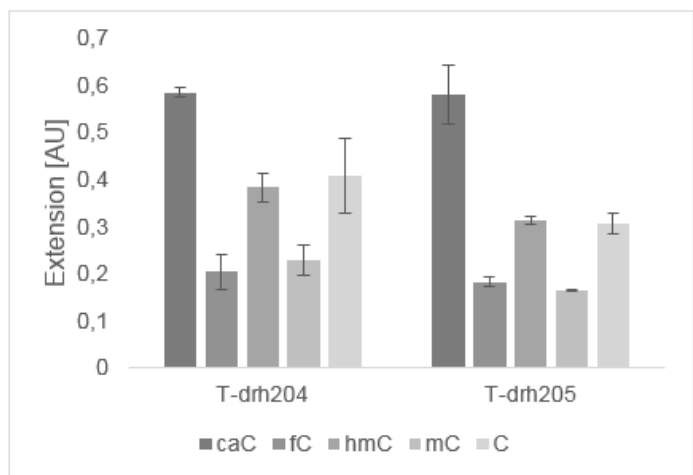
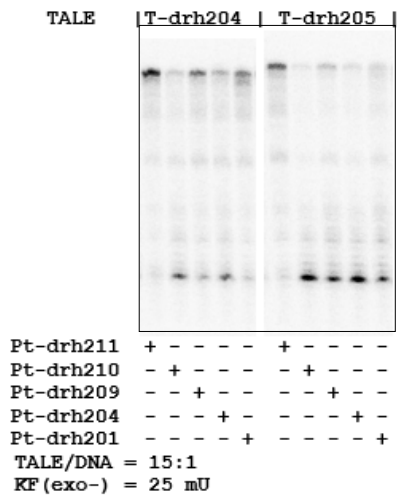
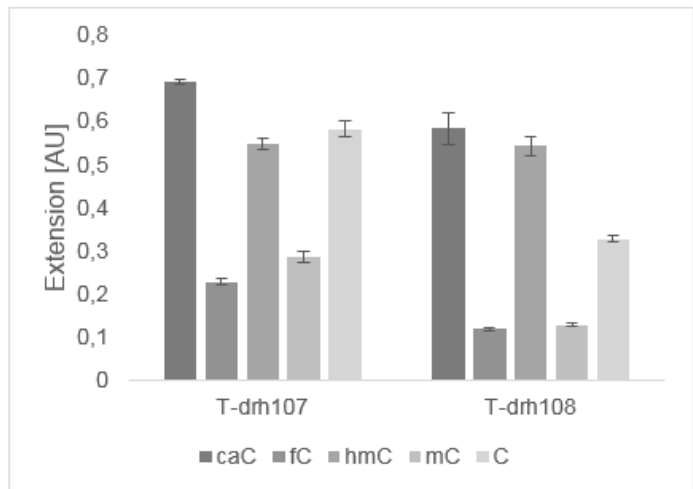
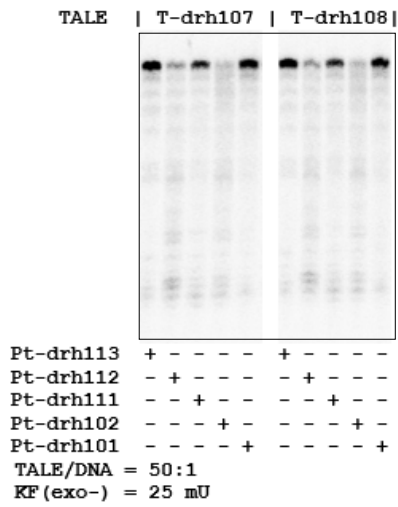


Amount of respective TALE: 146 nM each (T-drh202 = 208 nM)
 Amount of Klenow Fragment: 25 mU each

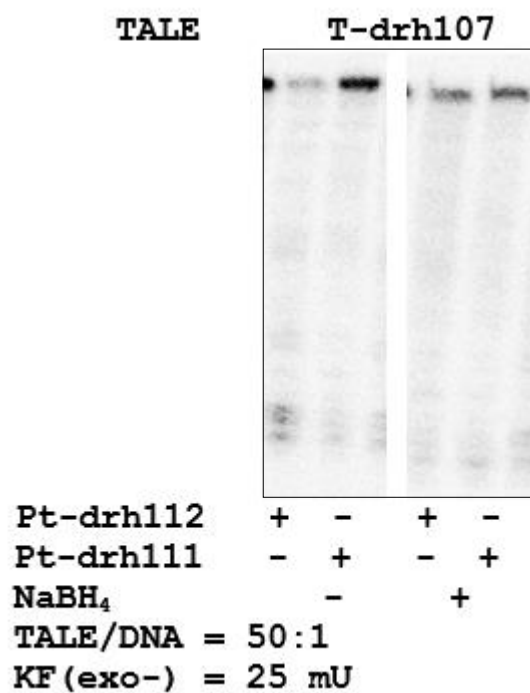
S-Figure S23: Primer Extension Gel for RVD selectivity comparison to all C modifications in Sequence context S-drh1



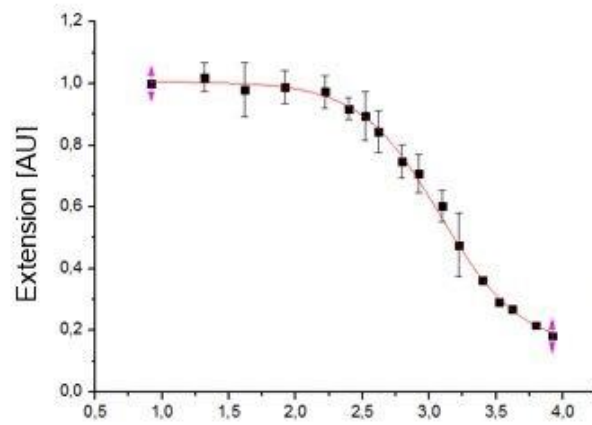
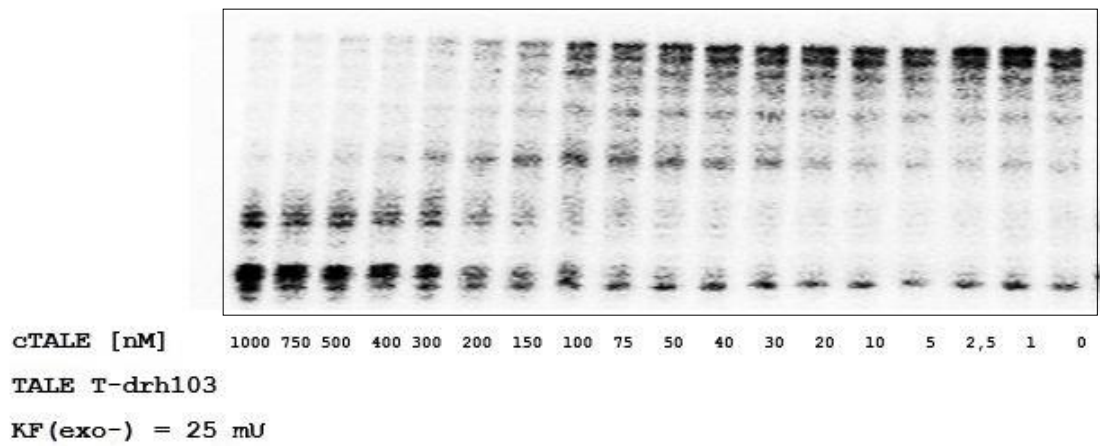
S-Figure S24: Primer Extension Gel for N* and S* selectivity comparison in Sequence context S-drh1 and S-drh2



S-Figure S25: Primer Extension Gel for reduction experiment

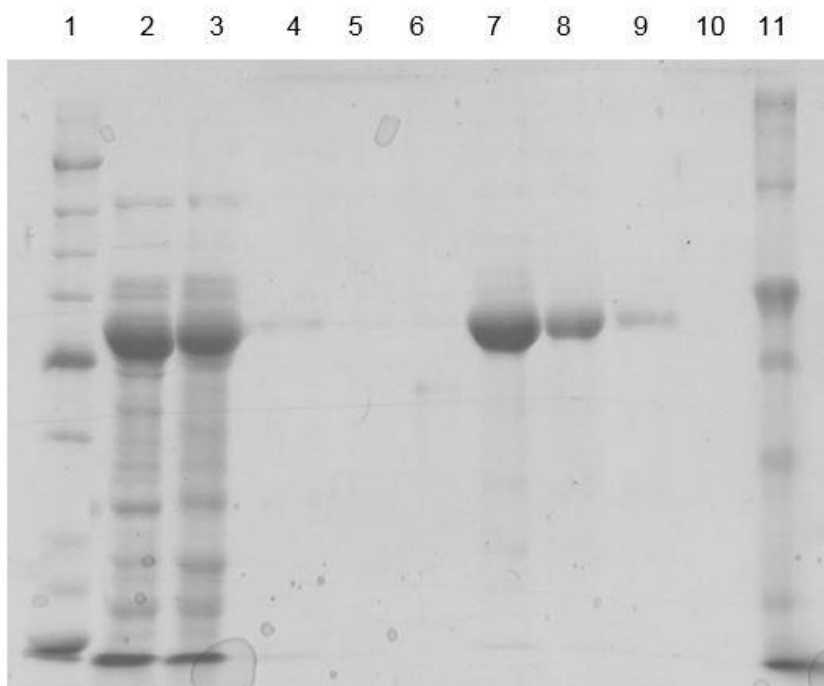


S-Figure S26: Primer Extension Gel with T-drh103 for IC₅₀ determination to N4mC



Modell	DoseResp	
Gleichung	$y = A1 + (A2-A1)/(1 + 10^{((LOGx0-x)^p)})$	
Chi-Quadr Reduziert	0,22855	
Kor. R-Qu	0,99855	
	Wert	Standardf
A1	0,1431	0,00712
A2	1,00579	0,01037
LOGx0	3,07984	0,0143
p	-1,44568	0,06834
span	0,88269	0,01489
EC20	3135,398	114,70803
EC50	1201,817	39,577
EC80	460,6640	32,37716

S-Figure S27: SDS PAGE Analysis of exemplary TALE expression and Ni-NTA-Purification for T-drh104



Lane 1: Ladder

Lane 2: Supernatant after Lysis

Lane 3: Flow through after Ni-NTA Binding

Lane 4: Wash fraction with 4xPBS containing 8 M Urea

Lane 5: Wash fraction with Buffer Z containing 20 mM Imidazole

Lane 6: Wash fraction with Buffer Z containing 50 mM Imidazole

Lane 7: Elution 1 with Buffer Z containing 500 mM Imidazole

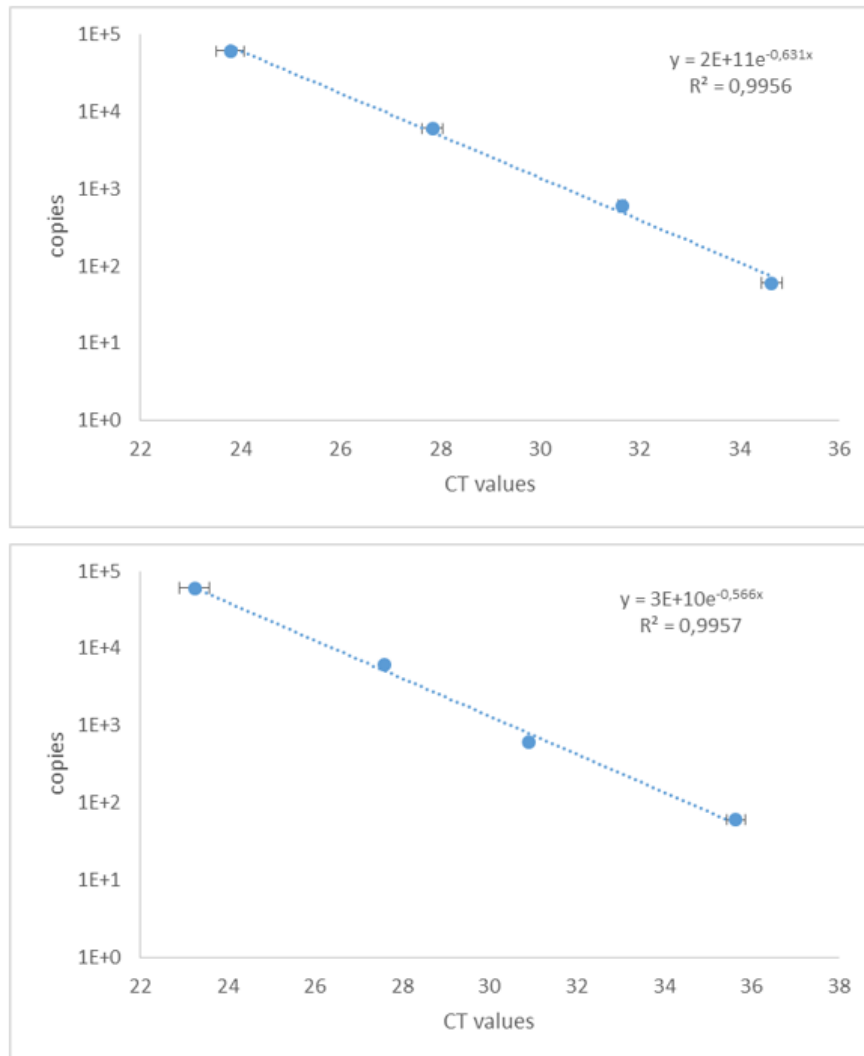
Lane 8: Elution 2 with Buffer Z containing 500 mM Imidazole

Lane 9: Elution 3 with Buffer Z containing 500 mM Imidazole

Lane 10: Empty

Lane 11: Ladder

S-Figure S28: qPCR concentration calibration with fragmented Zebrafish total DNA



C_T values of calibration PCR reactions for original DNA are shown in the upper graph and for methylated DNA in the lower graph.

S-Figure S29: Amino acid sequence of used TALE proteins

Scaffold Sequence:

MSD KII HLT DDS FDT DVL KAD GAI LVD FWA EWC
GPC KMI API LDE IAD EYQ GKL TVA KLN IDQ NPG
TAP KYG IRG IPT LLL FKN GEV AAT KVG ALS KGQ
LKE FLD ANL AGS GSG ERQ HMD SPD LGT VDL RTL
GYS QQQ QEK IKP KVR STV AQH HEA LVG HGF THA
HIV ALS QHP AAL GTV AVK YQD MIA ALP EAT HEA
IVG VGK QWS GAR ALE ALL TVA GEL RGP PLQ LDT
GQL LKI AKR GGV TAV EAV HAW RNA LTG APL N
--- **DNA Binding Domain** ---
SIV AQL SRP DPA LAA LTN DHL LEH HHH HH*

Red: TRX

Blue: TALE N-terminus

Orange: TALE C-terminus

Green: His-tag

Repeat Unit Sequence for assembled TALEs:

LTPDQVVAIAS**XX**GGKQALETVQRLLPVLCQDHG

Red: RVD

Underlined: Positions that are variable in TALEs T-drh101, T-drh201 and T-drg101

REFERENCES

- [1] J. A. Law, S. E. Jacobsen, *Nat Rev Genet* **2010**, *11*, 204–220.
- [2] P. A. Jones, *Nat Rev Genet* **2012**, *13*, 484–492.
- [3] M. Rodríguez-Paredes, M. Esteller, *Nat Med* **2011**, 330–339.
- [4] A. Maiti, A. C. Drohat, *Journal of Biological Chemistry* **2011**, *286*, 35334–35338.
- [5] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
- [6] Y.-F. He, B.-Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li et al., *Science* **2011**, *333*, 1303–1307.
- [7] G. Kubik, D. Summerer, *ACS Chem. Biol.* **2015**, 150501095946007.
- [8] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, U. Bonas, *Science* **2009**, *326*, 1509–1512.
- [9] J. Boch, U. Bonas, *Annu. Rev. Phytopathol.* **2010**, *48*, 419–436.
- [10] A. J. Bogdanove, D. F. Voytas, *Science (New York, N.Y.)* **2011**, *333*, 1843–1846.
- [11] M. J. Moscou, A. J. Bogdanove, *Science* **2009**, *326*, 1501.
- [12] J. Yang, Y. Zhang, P. Yuan, Y. Zhou, C. Cai, Q. Ren, D. Wen, C. Chu, H. Qi, W. Wei, *Cell Res* **2014**, *24*, 628–631.
- [13] G. Kubik, M. J. Schmidt, J. E. Penner, D. Summerer, *Angew. Chem. Int. Ed.* **2014**, *53*, 6002–6006.
- [14] S. Bultmann, R. Morbitzer, C. S. Schmidt, K. Thanisch, F. Spada, J. Elsaesser, T. Lahaye, H. Leonhardt, *Nucleic Acids Research* **2012**, *40*, 5368–5377.
- [15] J. Valton, A. Dupuy, F. Daboussi, S. Thomas, A. Marechal, R. Macmaster, K. Melliand, A. Juillerat, P. Duchateau, *Journal of Biological Chemistry* **2012**, *287*, 38427–38432.
- [16] G. Kubik, D. Summerer, *ChemBioChem* **2015**, *16*, 228–231.
- [17] G. Kubik, S. Batke, D. Summerer, *J. Am. Chem. Soc.* **2015**, *137*, 2–5.
- [18] J. D. WATSON, CRICK, F. H. C., *Nature* **1953**, *171*, 737–738.
- [19] W. Janning, E. Knust, *Genetik. Allgemeine Genetik, Molekulare Genetik, Entwicklungsgenetik*, Thieme, Stuttgart, **2008**.
- [20] A. N.-S. Mak, P. Bradley, R. A. Cernadas, A. J. Bogdanove, B. L. Stoddard, *Science* **2012**, *335*, 716–719.
- [21] J. M. Berg, J. L. Tymoczko, L. Stryer, N. D. Clarke, *Biochemistry*, Freeman and Company, New York, **op. 2002**.
- [22] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444.

- [23] C. H. Waddington, *Proc Natl Acad Sci U S A.* **1939**, *25*, 299–307.
- [24] M. Esteller, *N Engl J Med* **2008**, *358*, 1148–1159.
- [25] E. Olkhov-Mitsel, B. Bapat, *Cancer Med* **2012**, *1*, 237–260.
- [26] R. Bonasio, S. Tu, D. Reinberg, *Science* **2010**, *330*, 612–616.
- [27] a) B. M. Nugent, C. L. Wright, A. C. Shetty, G. E. Hodes, K. M. Lenz, A. Mahurkar, S. J. Russo, S. E. Devine, M. M. McCarthy, *Nat Neurosci* **2015**, *18*, 690–697; b) S. Sharma, T. K. Kelly, P. A. Jones, *Carcinogenesis* **2010**, *31*, 27–36; c) S. I. Sherwani, H. A. Khan, *Gene* **2015**.
- [28] T. Shlomi, J. D. Rabinowitz, *Nat Chem Biol* **2013**, *9*, 293–294.
- [29] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930; b) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind et al., *Science* **2009**, *324*, 930–935.
- [30] a) C. Loenarz, C. J. Schofield, *Trends in Biochemical Sciences* **2011**, *36*, 7–18; b) L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang, Y. Xu, *Cell* **2013**, *155*, 1545–1555.
- [31] T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeyer, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012.
- [32] a) X. Lu, B. S. Zhao, C. He, *Chem. Rev.* **2015**, *115*, 2225–2239; b) L. Shen, C.-X. Song, C. He, Y. Zhang, *Annu. Rev. Biochem.* **2014**, *83*, 585–614.
- [33] S. Liu, J. Wang, Y. Su, C. Guerrero, Y. Zeng, D. Mitra, P. J. Brooks, D. E. Fisher, H. Song, Y. Wang, *Nucleic Acids Research* **2013**, *41*, 6421–6429.
- [34] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *Nature Chem* **2014**, *6*, 1049–1055.
- [35] M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nature Chem* **2014**, *6*, 435–440.
- [36] M. J. Booth, M. R. Branco, G. Ficzy, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934–937.
- [37] C.-X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu et al., *Cell* **2013**, *153*, 678–691.
- [38] E.-A. Raiber, D. Beraldi, G. Ficzy, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik, S. Balasubramanian, *Genome Biol* **2012**, *13*, R69.
- [39] L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H.-L. Fung, K. Zhang, Y. Zhang, *Cell* **2013**, *153*, 692–706.
- [40] a) L. Lercher, M. A. McDonough, A. H. El-Sagheer, A. Thalhammer, S. Kriaucionis, T. Brown, C. J. Schofield, *Chem. Commun.* **2014**, *50*, 1794–1796; b) D. Renciuik, O. Blacque, M. Vorlickova, B. Spingler, *Nucleic Acids Research* **2013**, *41*, 9891–9900; c) E.-A. Raiber, P.

- Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi, S. Balasubramanian, *Nat Struct Mol Biol* **2014**, 22, 44–49; d) A. Chentsova, E. Kapourani, A. Giannis, *Beilstein J. Org. Chem.* **2014**, 10, 7–11.
- [41] M. W. Szulik, P. S. Pallan, B. Nocek, M. Voehler, S. Banerjee, S. Brooks, A. Joachimiak, M. Egli, B. F. Eichman, M. P. Stone, *Biochemistry* **2015**, 54, 1294–1305.
- [42] M. Sumino, A. Ohkubo, H. Taguchi, K. Seio, M. Sekine, *Bioorganic & Medicinal Chemistry Letters* **2008**, 18, 274–277.
- [43] L. Wang, Y. Zhou, L. Xu, R. Xiao, X. Lu, L. Chen, J. Chong, H. Li, C. He, X.-D. Fu et al., *Nature* **2015**.
- [44] a) M. Iurlaro, G. Ficuz, D. Oxley, E.-A. Raiber, M. Bachman, M. J. Booth, S. Andrews, S. Balasubramanian, W. Reik, *Genome biology* **2013**, 14, R119; b) C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan et al., *Cell* **2013**, 152, 1146–1159; c) La Francois, Christopher J., Y. H. Jang, T. Cagin, W. A. Goddard, L. C. Sowers, *Chem. Res. Toxicol.* **2000**, 13, 462–470.
- [45] M. Mellén, P. Ayata, S. Dewell, S. Kriaucionis, N. Heintz, *Cell* **2012**, 151, 1417–1430.
- [46] a) S. L. Berger, T. Kouzarides, R. Shiekhattar, A. Shilatifard, *Genes & Development* **2009**, 23, 781–783; b) A. Bird, *Nature* **2007**, 447, 396–398; c) E. V. JABLONKA, M. J. LAMB, *Annals of the New York Academy of Sciences* **2002**, 981, 82–96; d) T. P. Jurkowski, M. Ravichandran, P. Stepper, *Clin Epigenet* **2015**, 7, 1141.
- [47] M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, *Proceedings of the National Academy of Sciences* **1992**, 89, 1827–1831.
- [48] a) Y. Huang, W. A. Pastor, Y. Shen, M. Tahiliani, D. R. Liu, A. Rao, *PLoS ONE* **2010**, 5, e8888; b) R. P. Darst, C. E. Pardo, L. Ai, K. D. Brown, M. P. Klade in *Current Protocols in Molecular Biology* (Eds.: F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. Seidman, J. A. Smith, K. Struhl), John Wiley & Sons, Inc, Hoboken, NJ, USA, **2001**; c) M. Shiraishi, H. Hayatsu, *DNA Res.* **2004**, 11, 409–415.
- [49] P. Schüler, A. K. Miller, *Angew. Chem. Int. Ed.* **2012**, 51, 10704–10707.
- [50] J. JOSSE, A. KORNBERG, *The Journal of biological chemistry* **1962**, 237, 1968–1976.
- [51] M. Yu, G. C. Hon, K. E. Szulwach, C.-X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park et al., *Cell* **2012**, 149, 1368–1380.
- [52] X. Lu, C.-X. Song, K. Szulwach, Z. Wang, P. Weidenbacher, P. Jin, C. He, *J. Am. Chem. Soc.* **2013**, 135, 9315–9317.
- [53] a) N. Plongthongkum, D. H. Diep, K. Zhang, *Nature reviews. Genetics* **2014**, 15, 647–661; b) M. J. Booth, E.-A. Raiber, S. Balasubramanian, *Chemical reviews* **2015**, 115, 2240–2254; c) S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner et al., *J. Am. Chem. Soc.* **2013**, 135, 14593–14599.

- [54] C.-X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C.-H. Chen, W. Zhang, X. Jian et al., *Nature biotechnology* **2011**, 29, 68–72.
- [55] a) W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov et al., *Nature* **2011**, 473, 394–397; b) L. Zhang, K. E. Szulwach, G. C. Hon, C.-X. Song, B. Park, M. Yu, X. Lu, Q. Dai, X. Wang, C. R. Street et al., *Nat Comms* **2013**, 4, 1517.
- [56] A. B. Robertson, J. A. Dahl, C. B. Vagbo, P. Tripathi, H. E. Krokan, A. Klungland, *Nucleic Acids Research* **2011**, 39, e55.
- [57] G. Ficiz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, 473, 398–402.
- [58] N. Li, M. Ye, Y. Li, Z. Yan, L. M. Butcher, J. Sun, X. Han, Q. Chen, X. zhang, J. Wang, *Methods* **2010**, 52, 203–212.
- [59] V. Valinluck, *Nucleic Acids Research* **2004**, 32, 4100–4108.
- [60] a) E. J. Oakeley, *Pharmacology & therapeutics* **1999**, 84, 389–400; b) K. Yamagata, *Methods* **2010**, 52, 259–266.
- [61] a) Z. Sun, J. Terragni, T. Jolyon, J. G. Borgaro, Y. Liu, L. Yu, S. Guan, H. Wang, D. Sun, X. Cheng et al., *Cell Reports* **2013**, 3, 567–576; b) H. Wang, S. Guan, A. Quimby, D. Cohen-Karni, S. Pradhan, G. Wilson, R. J. Roberts, Z. Zhu, Y. Zheng, *Nucleic Acids Research* **2011**, 39, 9294–9305; c) B. Khulan, R. F. Thompson, K. Ye, M. J. Fazzari, M. Suzuki, E. Stasiak, M. E. Figueroa, J. L. Glass, Q. Chen, C. Montagna et al., *Genome research* **2006**, 16, 1046–1055; d) A. L. Brunner, D. S. Johnson, S. W. Kim, A. Valouev, T. E. Reddy, N. F. Neff, E. Anton, C. Medina, L. Nguyen, E. Chiao et al., *Genome research* **2009**, 19, 1044–1056; e) D. Cohen-Karni, D. Xu, L. Apone, A. Fomenkov, Z. Sun, P. J. Davis, Kinney, Shannon R Morey, M. Yamada-Mabuchi, S.-y. Xu, T. Davis et al., *Proceedings of the National Academy of Sciences of the United States of America* **2011**, 108, 11040–11045.
- [62] a) J. J. Kasianowicz, E. Brandin, D. Branton, D. W. Deamer, *Proceedings of the National Academy of Sciences of the United States of America* **1996**, 93, 13770–13773; b) L. Song, M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, J. E. Gouaux, *Science (New York, N.Y.)* **1996**, 274, 1859–1866.
- [63] a) M. Akeson, D. Branton, J. J. Kasianowicz, E. Brandin, D. W. Deamer, *Biophysical journal* **1999**, 77, 3227–3233; b) Wallace, Emma V. B., D. Stoddart, A. J. Heron, E. Mikhailova, G. Maglia, T. J. Donohoe, H. Bayley, *Chem. Commun.* **2009**, 46, 8195; c) D. Stoddart, A. J. Heron, E. Mikhailova, G. Maglia, H. Bayley, *Proceedings of the National Academy of Sciences* **2009**, 106, 7702–7707; d) W.-W. Li, L. Gong, H. Bayley, *Angew. Chem. Int. Ed.* **2013**, 52, 4350–4355; e) E. A. Manrao, I. M. Derrington, M. Pavlenok, M. Niederweis, J. H. Gundlach, J. Mathe, *PLoS ONE* **2011**, 6, e25723; f) A. H. Laszlo, I. M. Derrington, H. Brinkerhoff, K. W. Langford, I. C. Nova, J. M. Samson, J. J. Bartlett, M. Pavlenok, J. H. Gundlach, *Proceedings of the National Academy of Sciences* **2013**, 110, 18904–18909.

- [64] a) M. Faller, *Science* **2004**, *303*, 1189–1192; b) I. M. Derrington, T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis, J. H. Gundlach, *Proceedings of the National Academy of Sciences* **2010**, *107*, 16060–16065.
- [65] a) K. R. Lieberman, G. M. Cherf, M. J. Doody, F. Olasagasti, Y. Kolodji, M. Akeson, *J. Am. Chem. Soc.* **2010**, *132*, 17961–17972; b) E. A. Manrao, I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, J. H. Gundlach, *Nat Biotechnol* **2012**, *30*, 349–353.
- [66] Z. L. Wescoe, J. Schreiber, M. Akeson, *J. Am. Chem. Soc.* **2014**, *136*, 16582–16587.
- [67] J. Schreiber, Z. L. Wescoe, R. Abu-Shumays, J. T. Vivian, B. Baatar, K. Karplus, M. Akeson, *Proceedings of the National Academy of Sciences* **2013**, *110*, 18910–18915.
- [68] a) Chaisson, Mark J P, J. Huddleston, M. Y. Dennis, P. H. Sudmant, M. Malig, F. Hormozdiari, F. Antonacci, U. Surti, R. Sandstrom, M. Boitano et al., *Nature* **2015**, *517*, 608–611; b) J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman et al., *Science* **2009**, *323*, 133–138.
- [69] a) B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, S. W. Turner, *Nature methods* **2010**, *7*, 461–465; b) A. J. Berman, S. Kamtekar, J. L. Goodman, J. M. Lázaro, M. de Vega, L. Blanco, M. Salas, T. A. Steitz, *The EMBO journal* **2007**, *26*, 3494–3505; c) D. Summerer, *Chembiochem : a European journal of chemical biology* **2010**, *11*, 2499–2501.
- [70] T. A. Clark, X. Lu, K. Luong, Q. Dai, M. Boitano, S. W. Turner, C. He, J. Korlach, *BMC Biol* **2013**, *11*, 4.
- [71] a) C.-X. Song, T. A. Clark, X.-Y. Lu, A. Kislyuk, Q. Dai, S. W. Turner, C. He, J. Korlach, *Nature methods* **2012**, *9*, 75–77; b) L. Chavez, Y. Huang, K. Luong, S. Agarwal, L. M. Iyer, W. A. Pastor, V. K. Hench, S. A. Frazier-Bowers, E. Korol, S. Liu et al., *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, E5149-58.
- [72] J. Aschenbrenner, M. Drum, H. Topal, M. Wieland, A. Marx, *Angewandte Chemie (International ed. in English)* **2014**, *53*, 8154–8158.
- [73] J. F. Petolino, *In Vitro Cell.Dev.Biol.-Plant* **2015**, *51*, 1–8.
- [74] F. D. Urnov, E. J. Rebar, M. C. Holmes, H. S. Zhang, P. D. Gregory, *Nat Rev Genet* **2010**, *11*, 636–646.
- [75] H. R. Jabalameli, H. Zahednasab, A. Karimi-Moghaddam, M. R. Jabalameli, *Gene* **2015**, *558*, 1–5.
- [76] a) M. Moore, A. Klug, Y. Choo, *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 1437–1441; b) D. Carroll, J. J. Morton, K. J. Beumer, D. J. Segal, *Nat Protoc* **2006**, *1*, 1329–1341; c) M. A. Hossain, J. J. Barrow, Y. Shen, M. I. Haq, J. Bungert, *J. Cell. Biochem.* **2015**, n/a.

- [77] N. Tochio, M. Yoneyama, S. Koshiba, S. Watanabe, T. Harada, T. Umehara, A. Tanaka, T. Kigawa, S. Yokoyama, *Solution structure of the 10th C2H2 zinc finger of human Zinc finger protein 406*, **2008**.
- [78] A. F. Gilles, M. Averof, *EvoDevo* **2014**, 5, 43.
- [79] D. Deng, C. Yan, X. Pan, M. Mahfouz, J. Wang, J.-K. Zhu, Y. Shi, N. Yan, *Science* **2012**, 335, 720–723.
- [80] H. Gao, X. Wu, J. Chai, Z. Han, *Cell Res* **2012**, 22, 1716–1720.
- [81] a) E. L. Doyle, A. W. Hummel, Z. L. Demorest, C. G. Starker, D. F. Voytas, P. Bradley, A. J. Bogdanove, *PLoS ONE* **2013**, 8, e82120; b) B. M. Lamb, A. C. Mercer, C. F. Barbas, *Nucleic Acids Research* **2013**, 41, 9779–9785; c) E. L. Doyle, N. J. Booher, D. S. Standage, D. F. Voytas, V. P. Brendel, J. K. VanDyk, A. J. Bogdanove, *Nucleic Acids Research* **2012**, 40, W117-W122.
- [82] M. Bochtler, *Biological Chemistry* **2012**, 393.
- [83] A. Juillerat, G. Dubois, J. Valton, S. Thomas, S. Stella, A. Marechal, S. Langevin, N. Benomari, C. Bertoni, G. H. Silva et al., *Nucleic Acids Research* **2014**, 42, 5390–5402.
- [84] T. Schreiber, U. Bonas, *Nucleic Acids Research* **2014**, 42, 7160–7169.
- [85] T. Sakuma, H. Ochiai, T. Kaneko, T. Mashimo, D. Tokumasu, Y. Sakane, K.-i. Suzuki, T. Miyamoto, N. Sakamoto, S. Matsuura et al., *Sci. Rep.* **2013**, 3.
- [86] J. F. Meckler, M. S. Bhakta, M.-S. Kim, R. Ovadia, C. H. Habrian, A. Zykovich, A. Yu, S. H. Lockwood, R. Morbitzer, J. Elsaesser et al., *Nucleic Acids Research* **2013**, 41, 4118–4128.
- [87] D. Deng, P. Yin, C. Yan, X. Pan, X. Gong, S. Qi, T. Xie, M. Mahfouz, J.-K. Zhu, N. Yan et al., *Cell Res* **2012**, 22, 1502–1504.
- [88] R. Moore, A. Chandrasah, L. Bleris, *ACS synthetic biology* **2014**, 3, 708–716.
- [89] J. D. Sander, L. Cade, C. Khayter, D. Reyon, R. T. Peterson, J. K. Joung, J.-R. J. Yeh, *Nat Biotechnol* **2011**, 29, 697–698.
- [90] T. Cermak, E. L. Doyle, M. Christian, L. Wang, Y. Zhang, C. Schmidt, J. A. Baller, N. V. Somia, A. J. Bogdanove, D. F. Voytas, *Nucleic Acids Research* **2011**, 39, e82.
- [91] E. Weber, R. Gruetzner, S. Werner, C. Engler, S. Marillonnet, M. Bendahmane, *PLoS ONE* **2011**, 6, e19722.
- [92] N. E. Sanjana, Le Cong, Y. Zhou, M. M. Cunniff, G. Feng, F. Zhang, *Nat Protoc* **2012**, 7, 171–192.
- [93] D. Reyon, S. Q. Tsai, C. Khayter, J. A. Foden, J. D. Sander, J. K. Joung, *Nat Biotechnol* **2012**, 30, 460–465.
- [94] A. W. Briggs, X. Rios, R. Chari, L. Yang, F. Zhang, P. Mali, G. M. Church, *Nucleic Acids Research* **2012**, 40, e117.

- [95] a) E. Pennisi, *Science (New York, N.Y.)* **2012**, 338, 1408–1411; b) M. Christian, T. Cermak, E. L. Doyle, C. Schmidt, F. Zhang, A. Hummel, A. J. Bogdanove, D. F. Voytas, *Genetics* **2010**, 186, 757–761; c) S. Boissel, J. Jarjour, A. Astrakhan, A. Adey, A. Gouble, P. Duchateau, J. Shendure, B. L. Stoddard, M. T. Certo, D. Baker et al., *Nucleic Acids Research* **2013**; d) M. Beurdeley, F. Bietz, J. Li, S. Thomas, T. Stoddard, A. Juillerat, F. Zhang, D. F. Voytas, P. Duchateau, G. H. Silva, *Nat Comms* **2013**, 4, 1762.
- [96] A. C. Mercer, T. Gaj, R. P. Fuller, C. F. Barbas, *Nucleic Acids Research* **2012**, 40, 11163–11172.
- [97] S. Konermann, M. D. Brigham, A. Trevino, P. D. Hsu, M. Heidenreich, Le Cong, R. J. Platt, D. A. Scott, G. M. Church, F. Zhang, *Nature* **2013**.
- [98] M. L. Maeder, J. F. Angstman, M. E. Richardson, S. J. Linder, V. M. Cascio, S. Q. Tsai, Q. H. Ho, J. D. Sander, D. Reyon, B. E. Bernstein et al., *Nat Biotechnol* **2013**, 31, 1137–1142.
- [99] a) D. L. Bernstein, Le Lay, John E., E. G. Ruano, K. H. Kaestner, *J. Clin. Invest.* **2015**, 125, 1998–2006; b) K. Li, J. Pang, H. Cheng, W.-P. Liu, J.-M. Di, H.-J. Xiao, Y. Luo, H. Zhang, W.-T. Huang, M.-K. Chen et al., *Oncotarget* **2015**, 6, 10030–10044.
- [100] a) E. M. Mendenhall, K. E. Williamson, D. Reyon, J. Y. Zou, O. Ram, J. K. Joung, B. E. Bernstein, *Nature biotechnology* **2013**, 31, 1133–1136; b) *Science-Business eXchange* **2013**, 6; c) F. Falahi, A. Sgro, P. Blancafort, *Front. Oncol.* **2015**, 5.
- [101] a) Z. Zhang, E. Wu, Z. Qian, W.-S. Wu, *Sci. Rep.* **2014**, 4, 7338; b) Le Cong, R. Zhou, Y.-c. Kuo, M. Cunniff, F. Zhang, *Nat Comms* **2012**, 3, 968.
- [102] M. M. Mahfouz, L. Li, M. Piatek, X. Fang, H. Mansour, D. K. Bangarusamy, J.-K. Zhu, *Plant Mol Biol* **2012**, 78, 311–321.
- [103] a) Y. Miyanari, C. Ziegler-Birling, M.-E. Torres-Padilla, *Nat Struct Mol Biol* **2013**, 20, 1321–1324; b) K. Thanisch, K. Schneider, R. Morbitzer, I. Solovei, T. Lahaye, S. Bultmann, H. Leonhardt, *Nucleic Acids Research* **2013**.
- [104] S. Wang, Y. Long, J. Wang, Y. Ge, P. Guo, Y. Liu, T. Tian, X. Zhou, *J. Am. Chem. Soc.* **2014**, 136, 56–59.
- [105] R. Wilson, *Nucleic Acid Therapeutics (Formerly Oligonucleotides)* **2011**, 21, 437–440.
- [106] a) M. Ehrlich, G. G. Wilson, K. C. Kuo, C. W. Gehrke, *Journal of bacteriology* **1987**, 169, 939–943; b) D. Chung, J. Farkas, J. R. Huddleston, E. Olivar, J. Westpheling, *PLoS ONE* **2012**, 7, e43844; c) J. Beaulaurier, X.-S. Zhang, S. Zhu, R. Sebra, C. Rosenbluh, G. Deikus, N. Shen, D. Munera, M. K. Waldor, A. Chess et al., *Nature communications* **2015**, 6, 7438; d) M. Yu, L. Ji, D. A. Neumann, D.-H. Chung, J. Groom, J. Westpheling, C. He, R. J. Schmitz, *Nucleic Acids Research* **2015**.
- [107] J. C. Miller, S. Tan, G. Qiao, K. A. Barlow, J. Wang, D. F. Xia, X. Meng, D. E. Paschon, E. Leung, S. J. Hinkley et al., *Nat Biotechnol* **2010**, 29, 143–148.

TABLE OF FIGURES

Figure 1: DNA structure and composition.....	- 2 -
Figure 2: Active methylation and demethylation cycle of cytosine.	- 5 -
Adapted from <i>ACS Chem. Biol.</i> , 2015 , 10 (7), pp 1580–1589.	
Figure 3: Current approaches for the detection of epigenetic cytosine modifications.-	6 -
Figure 4: Strategies for full selectivity of BS-Seq regarding cytosine modifications. ...	7 -
Figure 5: Principle of antibody based analysis of epigenetic nucleobase detection. ...	9 -
Adapted from <i>ACS Chem. Biol.</i> , 2015 , 10 (7), pp 1580–1589.	
Figure 6: Crystal structure of human Zinc finger protein 406.....	- 12 -
Figure 7:TALE Overview.....	- 13 -
Figure 8: Crystal structure of TALE monomeric units bound to their target	- 15 -
Figure 9: Principle of Golden Gate assembly.....	- 16 -
Figure 10: General overview of exploited chimeric TALE Proteins.	- 17 -
Figure 11: Investigation of discriminatory ability of TALE repeat units.....	- 19 -
Figure 12: Illustration of TALE scaffold.	- 21 -
Figure 13: EMSA to validate 5mC sensitivity of RVDs HD and NG.	- 22 -
Adapted from <i>Angew. Chem. Int. Ed.</i> 2014 , 53, 6002–6006.	
Figure 14: Primer extension assay for 5mC detection.....	- 23 -
Adapted from <i>Angew. Chem. Int. Ed.</i> 2014 , 53, 6002–6006.	
Figure 15: In detail investigation of 5mC detection ability with T-drh101.	- 24 -
Adapted from <i>Angew. Chem. Int. Ed.</i> 2014 , 53, 6002–6006.	
Figure 16: Investigation of 5mC detection ability of TALEs T-drh201 and T-drg101.-	26 -
Figure 17: 5mC detection in presence of large amount off target DNA.....	- 27 -
Figure 18: Optimization of bead washing conditions	- 28 -
Figure 19: Determination of IC ₅₀ value for T-drh301.....	- 29 -
Figure 20: Detection of single 5mC in eukaryotic total DNA.	- 30 -
Reprinted from <i>Angew. Chem. Int. Ed.</i> 2014 , 53, 6002–6006. © 2014 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim	
Figure 21: Single nucleotide resolution with TALEs	- 32 -
Adapted from <i>ChemBioChem</i> 2015 , 16, 228–231.	
Figure 22: Analysis of the 5mC modification level at single position.....	- 33 -
Adapted from <i>ChemBioChem</i> 2015 , 16, 228–231.	
Figure 23: Sensitivity of RVDs to cytosine variants.	- 34 -
Adapted from <i>J. Am. Chem. Soc.</i> , 2015 , 137 (1), pp 2–5.	
Figure 24: Quantitative analysis of affinity and selectivity of TALEs	- 35 -
Reprinted from <i>J. Am. Chem. Soc.</i> , 2015 , 137 (1), pp 2–5. Copyright © 2015 American Chemical Society	
Figure 25: Selectivity profile of RVDs HD, NG and N*.....	- 37 -
Figure 26: A: Selective reduction of fC to hmC by NaBH ₄	- 39 -
Figure 27: Comparison of IC ₅₀ values of TALE T-drh103	- 41 -

REPRINT PERMISSIONS

**JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS**

Aug 11, 2015

This Agreement between Grzegorz Kubik ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3685870218214
License date	Aug 11, 2015
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Angewandte Chemie International Edition
Licensed Content Title	Programmable and Highly Resolved In Vitro Detection of 5-Methylcytosine by TALEs
Licensed Content Author	Grzegorz Kubik, Moritz J. Schmidt, Johanna E. Penner, Daniel Summerer
Licensed Content Date	May 6, 2014
Pages	5
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	Yes, including English rights
Number of languages	2
Languages	German English Polish
Title of your thesis / dissertation	Deciphering epigenetic cytosine modifications with TALEs
Expected completion date	Dec 2015
Expected size (number of pages)	120
Requestor Location	Grzegorz Kubik Kahlenbergstr. 7 Stolberg (Rhld), Germany 52224 Attn: Grzegorz Kubik
Billing Type	Invoice
Billing Address	Grzegorz Kubik Kahlenbergstr. 7 Stolberg (Rhld), Germany 52224 Attn: Grzegorz Kubik
Total	0.00 EUR
Terms and Conditions	

TERMS AND CONDITIONS

**JOHN WILEY AND SONS LICENSE
TERMS AND CONDITIONS**

Aug 11, 2015

This Agreement between Grzegorz Kubik ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3685870063504
License date	Aug 11, 2015
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	ChemBioChem
Licensed Content Title	Achieving Single-Nucleotide Resolution of 5-Methylcytosine Detection with TALEs
Licensed Content Author	Grzegorz Kubik, Daniel Summerer
Licensed Content Date	Dec 17, 2014
Pages	4
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	Yes, including English rights
Number of languages	2
Languages	German English Polish
Title of your thesis / dissertation	Deciphering epigenetic cytosine modifications with TALEs
Expected completion date	Dec 2015
Expected size (number of pages)	120
Requestor Location	Grzegorz Kubik Kahlenbergstr. 7 Stolberg (Rhld), Germany 52224 Attn: Grzegorz Kubik
Billing Type	Invoice
Billing Address	Grzegorz Kubik Kahlenbergstr. 7 Stolberg (Rhld), Germany 52224 Attn: Grzegorz Kubik
Total	0.00 EUR
Terms and Conditions	

TERMS AND CONDITIONS



RightsLink®

[Home](#)[Account Info](#)[Help](#)

Title: Programmable Sensors of 5-Hydroxymethylcytosine

Author: Grzegorz Kubik, Sabrina Batke, Daniel Summerer

Publication: Journal of the American Chemical Society

Publisher: American Chemical Society

Date: Jan 1, 2015

Copyright © 2015, American Chemical Society

Logged in as:
Grzegorz Kubik

[LOGOUT](#)

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

[BACK](#)[CLOSE WINDOW](#)

Copyright © 2015 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customercare@copyright.com



RightsLink®

[Home](#)[Account Info](#)[Help](#)

ACS Publications
Most Trusted. Most Cited. Most Read.

Title: Deciphering Epigenetic Cytosine Modifications by Direct Molecular Recognition

Author: Grzegorz Kubik, Daniel Summerer

Publication: ACS Chemical Biology

Publisher: American Chemical Society

Date: Jul 1, 2015

Copyright © 2015, American Chemical Society

Logged in as:
Grzegorz Kubik
Account #:
3000948537

[LOGOUT](#)

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

[BACK](#)[CLOSE WINDOW](#)

Copyright © 2015 [Copyright Clearance Center, Inc.](#) All Rights Reserved. [Privacy statement](#). [Terms and Conditions](#).
Comments? We would like to hear from you. E-mail us at customer care@copyright.com