Engineering Novel TALE Repeats for the Selective and Direct Detection of Epigenetic Cytosine Modifications

Dissertation

Submitted for the degree of

Doctor of Natural Sciences

(Dr. rer. nat.)

Presented by

Sara Maurer, M.Sc.

at the

Faculty of Chemistry and Chemical Biology

Dortmund 2018

This work was prepared from October 2014 to March 2018 in the group of Prof. Dr. Daniel Summer at the Universität Konstanz and the Technische Universität Dortmund. Funding was provided by the Konstanz Research School Chemical Biology and the Technische Universität Dortmund.

List of Publications

[1] S. Maurer, B. Buchmuller, C. Ehrt, J. Jasper, O. Koch and D. Summerer, "Overcoming Conservation in TALE-DNA Interactions: A Minimal Repeat Scaffold Enables Selective Recognition of an Oxidized 5-Methylcytosine." *Chem. Sci.,* **2018**, DOI: 10.1039/C8SC01958D.

[2] S. Maurer, M. Giess, O. Koch and D. Summerer, "Interrogating Key Positions of Size-Reduced TALE Repeats Reveals a Programmable Sensor of 5-Carboxylcytosine." *ACS Chem. Biol.*, **2016**, 11, 3294-3299.

[3] P. Rathi, S. Maurer and D. Summerer, "Selective recognition of N4-methylcytosine by engineered transcription activator-like effectors." *Philos. Trans. R. Soc. Lond. B Biol. Sci*., **2018**, 373, 1748.

[4] P. Rathi, S. Maurer, G. Kubik and D. Summerer, "Isolation of Human Genomic DNA Sequences with Expanded Nucleobase Selectivity." *J. Am. Chem. Soc.*, **2016**, 138 (31), 9910-9918.

[5] D. Summerer, M. Giess, G. Kubik, S. Maurer*, Patent application,* EP 3214183 A1, **2016**, filed by the University of Konstanz.

Table of Contents

Acknowledgments

At first, I want to thank Prof. Dr. Daniel Summerer for giving me the opportunity to join his group and to work on a very interesting thesis project. I very much appreciated working in his group and want to thank him for his constant support, good advice, and willingness to answer all my questions.

I would also like to thank PD Dr. Leif Dehmelt for kindly agreeing to be the second evaluator of this thesis.

My special thanks to all the present and former members of the Summerer group: Dr. Moritz Johannes Schmidt, Dr. Grzegorz Kubik, Dr. Sarah Flade, Dr. Preeti Rathi, Anna Witte, Anne Jung, Álvaro Muñoz-López, Benjamin Buchmuller, Jan Wolffgramm, Mario Gieß, and Dr. Shubhendu Palei – thank you all for the great working atmosphere and the good times both inside and outside the lab. I would also like to say thank you to all the students which have been working in the Summerer group, with special thanks to my students Brinja Kosel and Britta Schulte.

Furthermore, I want to thank Katharina Kuhr, Simone Eppmann, Martina Reibner, Petra Alhorn, and Ulrich Schoppe for their help with general lab and ordering management.

I want to thank Dr. Oliver Koch from the TU Dortmund for the good collaboration. Thank you also to Julia Jasper for the homology models of the TALEs, and Christiane Ehrt for the molecular dynamics simulations.

Furthermore, I want to thank the Hartig group, the Marx group (both Universität Konstanz), the Rauh group, and the Dehmelt group (both TU Dortmund) for sharing their laboratory equipment.

I would like to thank the Konstanz Research School Chemical Biology and the TU Dortmund for funding.

Special thanks also to Dr. Ingrid Richter and Jan Wolffgramm for editing and proofreading my thesis, as well as for their helpful advice and support.

Finally, I sincerely thank all my friends and family, and in particular my parents, for their constant support throughout my studies and my PhD.

Abstract

Genetic information within a cell is stored in the nucleotide sequence of deoxyribonucleic acid (DNA). In a multicellular organism, all somatic cells contain the same genetic material i.e. have the same genotype. Nevertheless, they can differ greatly in their morphologies and functions, which is commonly attributed to a differentiated pattern of gene expression. Responsible for the formation of a cell´s specific phenotype are epigenetic modifications, which act as an additional layer of information that modulates chromatin structure and genome function within a cell. Being superimposed on the nucleotide sequence of DNA, epigenetic modifications consist of chemical modifications to DNA nucleobases and histone proteins which structure the chromatin. One of the first discovered epigenetic DNA modifications in mammals is 5-methylcytosine (5mC)[3], which has been linked to transcriptional repression and hence has been shown to influence a multitude of cellular and developmental processes such as genome stability, X-chromosome inactivation or imprinting[4]. Furthermore, aberrant methylation patterns have been linked to the development of numerous diseases including mental illnesses and cancer[4-5]. With the discovery of the oxidized variants of 5mC namely 5-hydroxymethylcytosine (5hmC)[6], 5-formylcytosine (5fC)[7], and 5-carboxylcytosine (5caC)[7a, 8] as additional cytosine modifications, the question of their epigenetic relevance has emerged and has since been investigated. In order to get a better understanding of their individual functions, programmable detection methods with high selectivity for the individual cytosine variants, preferentially with applicability in living organisms, are needed.

In the present study, modified transcription activator-like effector (TALE) proteins were used as DNA-binding scaffolds to investigate the individual cytosine modifications in a sequence-specific, direct, and selective manner. TALE proteins consist of an array of structurally highly conserved repeat units of which each makes specific contact to one of the DNA nucleotides. This one-to-one contact between the TALE repeat and the nucleobase is enabled by the hypervariable di-residue (repeat variable di-residue, RVD) positioned at amino acid residues 12 and 13 within each TALE repeat. Recognition thereby follows the so-called TALE code according to which RVD amino acids HD (His12, Asp13), NG (Asn12, Gly13), NN (Asn12, Asn13) and NI (Asn12, Ile13) specifically recognize nucleobases C, T, G and A, respectively[9]. The programmable sequence-specificity in combination with their sensitivity towards the epigenetic nucleobase 5mC and 5hmC[10] provided the idea for engineering size-reduced TALE repeats with novel selectivity for the chemically and structurally distinct oxidized 5mC derivatives. In a first project, saturation mutagenesis of RVD position 12 in combination with deletions of position 13 or deletion of position 13 and 14 provided two TALE libraries with size-reduced repeats that were screened for their binding ability towards C and the four modified cytosine variants. This resulted in a total of 200 new

repeat-DNA interactions and the discovery of repeat P* (Pro12, * = deletion at pos. 13) as the first programmable sensor of 5caC[2]. On the basis of these results, the study was extended to a more thorough investigation of a truncated TALE repeat scaffold with maximal flexibility for new repeat-nucleobase interaction modes. In a second project, six different mutant repeat libraries containing one to four deletions in the RVD loop region (repeat positions 12 to 15) combined with substitutions of repeat positions 11 or 12 were generated and libraries were screened for their selective binding to C, 5mC, 5hmC, 5fC and 5caC. The development of a Förster resonance energy transfer (FRET)-based DNase I competition assay allowed screening of TALE mutant libraries in a 384-well plate format and yielded a total number of 330 new repeat-nucleobase interactions. Intriguingly, this study revealed repeat R**** (Arg11, **** = deletions at positions 12 to 15) as the first size-reduced TALE repeat, which selectively and directly binds an oxidized 5mC variant. The structure and selectivity of repeat R**** was studied in more detail using molecular dynamic (MD) simulations and homology models. Our results showed a completely new binding mode towards 5caC, where Arg11 has a dual function in stabilizing the repeat structure and selectively recognizing 5caC. The results obtained from these studies denote a considerable step towards a better understanding of how repeat-nucleobase interactions can be modulated or manipulated and reveal the surprising engineering potential and adaptability of TALE repeats.

Zusammenfassung

Die genetische Information einer Zelle ist in der Nukleotidsequenz der Desoxyribonukleinsäure (DNS, engl.: DNA) gespeichert. In einem mehrzelligen Organismus besitzen alle somatischen Zellen das gleiche genetische Material, d. h. sie haben den gleichen Genotyp. Nichtsdestotrotz unterscheiden sie sich grundlegend in ihrer Morphologie und ihren Funktionen, was auf ein differenziertes Genexpressionsmuster innerhalb jeder Zelle zurückzuführen ist. Verantwortlich für die Ausbildung des Phänotyps einer individuellen Zelle sind epigenetische Modifikationen, die in Überlagerung mit der DNA-Nukleotidsequenz eine zusätzliche Informationsschicht darstellen und die Chromatinstruktur sowie die Funktionen des Genoms maßgeblich beeinflussen. Epigenetische Modifikationen bestehen aus zwei Arten chemischer Modifikationen, denen der DNA und denen von Histonproteine. Eine der am besten untersuchten epigenetischen DNA-Modifikationen in Säugetieren ist 5-Methylcytosin (5mC)[3], welches durch Methylierung von Cytosin an der Position C5 innerhalb des Desoxyribosemoleküls entsteht. Diese DNA-Modifikation steht in Verbindung mit transkriptioneller Repression und beeinflusst dadurch eine große Anzahl zellulärer und entwicklungsphysiologischer Prozesse wie z.B. die Stabilität des Genoms, die X-Chromosom-Inaktivierung oder *Imprinting*[4]. Darüber hinaus konnten anormale Methylierungsmuster mit der Entstehung verschiedener Krankheiten, darunter psychische Krankheiten und Krebs, in Verbindung gebracht werden[4-5]. Mit der Entdeckung der oxidierten Formen von 5-Methylcytosin, darunter 5-Hydroxymethylcytosin (5hmC)[6], 5-Formylcytosin (5fC)[7] und 5-Carboxylcytosin (5caC)[7a, 8], als weitere Cytosin-Modifikationen, entstand die Frage nach deren epigenetischer Bedeutung. Um die biologischen Funktionen der einzelnen Cytosin-Modifikationen besser zu erforschen und zu verstehen, werden programmierbare, selektive Detektionsmethoden benötigt, die sich zur Anwendung im lebenden Organismus eignen.

In der hier vorliegenden Arbeit wurden modifizierte Transkriptionsaktivator-artige Effektor (TALE) - Proteine auf ihre Verwendung zur sequenz-spezifischen, direkten und selektiven Detektion der einzelnen modifizierten Cytosinbasen in DNA untersucht. TALE Proteine bestehen aus einer Abfolge von strukturell hochkonservierten Wiederholungseinheiten, die jeweils eine einzelne DNA-Nukleobase erkennen. Die spezifischen Eins-zu-eins-Erkennung zwischen der Wiederholungseinheit des TALEs und der DNA-Nukleobase geschieht durch zwei hypervariable Aminosäure an Position 12 und 13, dem sogenannten *repeat variable di-residue* (RVD), innerhalb einer TALE Wiederholungseinheit. Die Erkennung folgt dabei dem sogenannten *TALE code*, wonach RVD Aminosäuren HD (His12, Asp13), NG (Asn12, Gly13), NN (Asn12, Asn13) und NI (Asn12, Ile13) jeweils spezifisch die DNA-Nukleobasen C, T, G und A erkennen[9].

Die programmierbare Sequenzspezifizität von TALE-Proteinen in Kombination mit deren bereits bekannten Sensitivität gegenüber den epigenetischen Cytosin-Modifikationen 5mC und 5hmC[10], lieferte die Idee zur Entwicklung artifizieller, verkleinerter TALE-Wiederholungseinheiten mit neuen Selektivitäten für die chemisch und strukturell anspruchsvolleren oxidierten 5-Methylcytosin-Varianten. In einem ersten Teilprojekt wurden über Kombination von Sättigungsmutagenese der Wiederholungseinheit-Position 12 und einer Deletion der nebenliegenden Position 13 sowie der Positionen 13 und 14 zwei TALE-Wiederholungseinheit-Bibliotheken erstellt (Bibliotheken X* und X** mit X = eine der 20 kanonischen Aminosäuren an Position 12 und * = Deletion der Position 13, sowie ** = Deletion von Positionen 13 und 14). Die verkleinerten Wiederholungseinheiten wurden im Folgenden auf ihre Bindungsselektivität gegenüber den fünf verschiedenen Cytosin-Nukleobasen getestet. Dies resultierte in der Untersuchung von 200 unbekannten TALE-Nukleobase-Interaktionen und führte zur Entdeckung von TALE-Wiederholungseinheit P* (Pro12, * = Deletion von Position 13) als ersten programmierbaren Sensor für 5caC[2].

Basierend auf diesen Ergebnissen wurde die Untersuchung auf noch stärker verkleinerte TALE-Wiederholungseinheit-Strukturen ausgeweitet, was eine maximale Flexibilität zur Entstehung von neuen TALE-Nukleobasen-Interaktionen ermöglichen sollte. Daher wurden in einem zweiten Teilprojekt sechs weitere TALE-Wiederholungseinheit-Bibliotheken hergestellt, die eine bis vier Deletionen (Positionen 12 bis 15 innerhalb der TALE-Wiederholungseinheit) und Substitutionen von Wiederholungseinheit-Positionen 11 und 12 aufwiesen. Die resultierenden TALE-Wiederholungseinheiten wurden ebenfalls auf ihre Bindungsselektivität gegenüber C, 5mC, 5hmC, 5fC und 5caC getestet. Ein neu entwickelter Förster-Resonanzenergietransfer (FRET)-basierter DNase I-Kompetitionsassay ermöglichte eine Hochdurchsatzuntersuchung der TALE-Wiederholungseinheiten in 384 parallelen Ansätzen (*384-well plate format*). Dadurch konnten 330 unbekannten TALE-Nukleobasen-Interaktionen untersucht werden. Dies führte zur Entdeckung von Wiederholungseinheit R**** (Arg11, **** = Deletion von Positionen 12 bis 15) als erste, verkleinerte TALE-Wiederholungseinheit mit selektiver, programmierbarer und direkter Erkennung einer oxidierten 5mC-Variante. Genauere Untersuchungen der Struktur der Wiederholungseinheit sowie der Selektivität von R**** über Moleküldynamik (MD)-Simulationen und die Erstellung von Homologiemodellen, zeigten einen komplett neuen Bindungsmechanismus bei der direkten Erkennung von 5caC. Die Ergebnisse aus der vorliegenden Arbeit stellen somit einen wesentlichen Beitrag zum besseren Verständnis der Manipulation von TALE-Nukleobasen-Interaktionen dar und konnten darüber hinaus die erstaunliche Adaptierbarkeit von TALE-Wiederholungseinheiten aufzeigen.

Abbreviations

A	Adenine
Å	Angstrom
aa	Amino acid
AdoHyc	S-adenosyl-I-homocysteine
AdoMet	S-adenosyl-I-methionine
APS	Ammonium persulfate
ATRX	Alpha-thalassemia mental retardation X-linked
BAH	Bromo-adjacent homology domain
BER	Base excision repair
bp	Base pair
BRCA1	Breast cancer 1
BS-seq	Bisulfite sequencing
C	Cytosine
CAB-seq	Chemical modification-assisted bisulfite sequencing
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CGI	CpG island
cm	Centimeter(s)
CMS	5-Methylenesulfonate
CpG	Cytosine-guanine diphosphate
CRD	Central repeat domain
CTR	C-terminal region
Cy3/Cy5	Cyanine 3 / Cyanine 5
Da	Dalton
DDD	Dickerson-Drew Dodecamer
DIP-CAB-seq.	DNA immunoprecipitation-coupled CAB-seq
DNA	Deoxyribonucleic acid
Dnmt	DNA methyltransferase
ds	Double-stranded

DSBH	Double-stranded β-helix
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
fCAB-seq	5fC chemical modification-assisted bisulfite sequencing
Fe	Iron
FRET	Fluorescence resonance energy transfer
gDNA	Genomic DNA
GFP	Green fluorescent protein
GLIB	Glucosylation, periodate oxidation, biotinylation
h	Hour
HD	Histidine aspartate (repeat variable di-residue)
HDAC	Histone deacetlyase
Hey2	*hes-related family bHLH transcription factor with YRPW motif 2*
His6	Hexahistidine tag
hMeDIP-seq	hydroxymethyl-DNA immunoprecipitation-sequencing
HOTAIR	Hox transcription antisense intergenic RNA
HR	Homology repair
IP	Imaging plate
k	Kilo
KF	Klenow fragment
L	Liter(s)
LAD	Lamina-associated domain
lincRNA	Long intergenic noncoding RNA
lncRNA	Long noncoding RNA
MBD	Methyl-CpG-binding domain
MeCP	Methyl-CpG-binding protein

MeDIP	Methylated DNA immunoprecipitation
mESC	Mouse embryonic stem cells
mg	Milligram(s)
mM	Millimolar
MGMT	O6-methylguanine-DNAmethyltransferase
min	Minute(s)
miRNA	Micro RNA
MLH1	MutL homolog 1
mRNA	Messenger RNA
NCP	Nucleosome core particle
ng	Nanogram(s)
NG	Asparagine glycine (repeat variable di-residue)
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NI	Asparagine isoleucine (repeat variable di-residue)
Ni-NTA	Nickel-nitrilotriacetic acid
nm	Nanometer(s)
nM	Nanomolar
NN	Asparagine asparagine (repeat variable di-residue)
NTR	N-terminal region
OxBS-seq	Oxidative bisulfite sequencing
PAGE	Polyacrylamide gel electrophoresis
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PDB	Protein data bank
PEX	Primer Extension
PGC	Primordial germ cells
PHD	Plant homodomain
PRC	Pre-replication complex
PWWP	Conserved proline-tryptophan-tryptophan-proline domain

RedBS-seq	Reductive bisulfite sequencing
RNA	Ribonucleic acid
rpm	Revolution(s) per minute
rRNa	Ribosomal RNA
RT	Room temperature
RVD	Repeat variable di-residue
S	Second(s)
SAM	S-adenosyl-L-methionine
SDS	Sodium dodecyl sulfate
SMRT-seq	Single molecule real time sequencing
T	Thymine
TAB-seq	TET-assisted bisulfite sequencing
TAD	Topologically associated domain
TALE	Transcription activator-like effector
TALEN	Transcription activator-like effector nuclease
TBE	Tris-borate-EDTA
TDG	Thymine DNA glycosylase
TET DNA dioxygenase	Ten-eleven translocation DNA dioxygenase
TF	Transcription factor
TRD	Transcriptional repressor domain
tRNA	Transfer RNA
TSS	Transcription start site
U	Uracil
UHRF1	Ubiquitin-like plant homeodomain and RING finger domain 1
µm	Micrometer(s)
µM	Micromolar
V	Volts
WT	Wild type
w/o	without
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

ZFP	Zinc finger protein
ZMW	Zero-mode waveguide
ZNF	Zinc finger nuclease
° C	Degree Celsius
5mC	5-Methylcytosine
5hmC	5-Hydroxymethylcytosine
5fC	5-Formylcytosine
5caC	5-Carboxylcytosine
*****	Deletion

List of Tables

List of Figures

1 Introduction

1.1 Deoxyribonucleic acid (DNA) and the genome

1.1.1 The DNA

Discovered first in 1869 by Friedrich Miescher, deoxyribonucleic acid (DNA) was long doubted to be the molecule that is responsible for storing a cell´s genetic information. At this time, DNA was thought to be a long polymer built from four chemically very similar building blocks. This chemically simple makeup did not seem to be compatible with the sophisticated process of storing the genetic information. Only in the 1950s when X-ray diffraction analysis revealed the double-stranded helical structure of DNA, today known as Watson-Crick DNA structure[11], the DNA´s potential for encoding genetic information and replication became obvious[12].

DNA is composed of four basic different building blocks called nucleotides (Figure 1-1). Each of these DNA nucleotides consists of the five-carbon sugar deoxyribose attached to a phosphate and one of the four nucleobases adenine (A), guanine (G), cytosine (C), and thymine (T). The nucleotides are covalently linked through a phosphodiester bond, which is formed between the 5´-phosphate group and the 3´-OH group of two adjacent nucleotides. This alternating sugar-phosphate-sugar-phosphate array forms the backbone of a DNA strand. Dictated by the way the nucleotides are linked, each DNA strand displays a chemical orientation with a 5´ and a 3´ end (referring to the carbon positions in the deoxyribose).

In the natural form of DNA, the double helix of two complementary DNA strands with antiparallel orientation are held together via hydrogen bonds between the nucleobases of the two DNA strands. To accomplish this, the nucleobases are located on the inside of the double helix whereas the deoxyribose-phosphate backbone faces towards the outside. Hydrogen bonds are formed between the purine nucleobases (A and G) and the pyrimidine nucleobases (T and C) where, according to the Watson-Crick base-pairing, A always pairs with T, and G with C (Figure 1-1). Three hydrogen bonds connect a C•G base pair, and two hydrogen bonds form between a A•T base pair. The so formed base pairs can nicely stack on each other since the aromatic nucleobases are planar. This base-stacking is further stabilized by hydrophobic interactions and van der Waals forces, leading to an energetically favorable packing of the nucleobases inside the DNA double helix with equal distances between the deoxyribose units in the DNA backbone. Through intertwining of the two DNA strands, the DNA double helix is formed[13].

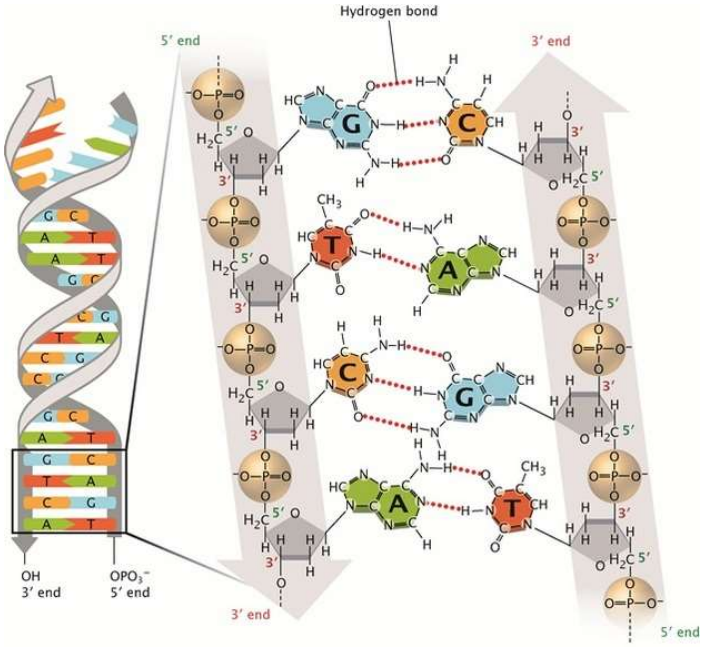


Figure 1-1 Watson-Crick base pairing in DNA. Watson-Crick base pairing in DNA occurs via hydrogen bonds between the nucleobases. Two hydrogen bonds are formed between and adenine (A) and a thymine (T) base pair (A•T), and three hydrogen bonds are formed between a cytosine (C) and guanine (G) base pair (C•G). This facilitates base-stacking and leads to an energetically favorable packing of the nucleobases inside the DNA double helix. Modified from[14].

1.1.2 Conformations of the DNA double helix

The most common conformation of the DNA double helix is the DNA B-form (Figure 1-2 a). Initially proposed by Watson and Crick, B-DNA is a right-handed helix with a helix diameter of 20 Angstrom (Å), a helix pitch (length of a complete helical turn of DNA) of 34 Å, and an axial rise (distance between adjacent planar bases) of 3.4 Å. Hence, in solution approximately 10.5 base pairs make up one complete helical turn. A characteristic feature of the B-DNA is the presence of two distinct grooves called the major and the minor groove. These two distinct grooves arise because the glycosidic bonds of one base pair are not positioned exactly opposite each other and therefore creating a wide (12 Å) and deep (8.5 Å) major groove, and a narrower (6 Å) and shallower (7.5 Å) minor groove. While the Watson-Crick hydrogen bonds of the DNA helix are not available to the surrounding solvent, many other different functional groups of the purine and pyrimidine nucleobases are accessible via the major and the minor groove which provide a distinct surface for different interaction partners. Due to its large size, the major groove is particularly important for protein-DNA interactions, with the thymine methyl group protruding into the groove and several hydrogen bond donors and acceptors being available[13, 15].

A second right-handed helix form of DNA is the A-DNA (Figure 1-2 b). This conformation of DNA is mostly found in dehydrated states of DNA, when humidity is less than 75 % (e.g. in DNA-crystals), but is also present in double-stranded RNA (dsRNA) and in RNA/DNA hybrids. Like B-DNA, A-DNA is made up from two antiparallel strands of nucleotides that are connected via Watson-Crick hydrogen bonding between the complementary nucleobases. A-DNA is wider and shorter than B-helix and has a narrow and very deep major groove and a quite broad and shallow minor groove. Compared to the B-form, A-DNA has a helix diameter of 23 Å, a helix pitch of 28 Å and an axial rise of 2.55 Å. Furthermore, the C-3´ of the deoxyribose lies outside the plane that is formed by the other four carbons of the sugar (C-3´ endo conformation) whereas in B-DNA the C-2´ is outside the plane (C-2´ endo conformation). This difference in the so-called sugar puckering leads to a 19 ° tilting of the base pairs in A-DNA, which is responsible for many of the structural differences between the A- and B-form of DNA[13, 15].

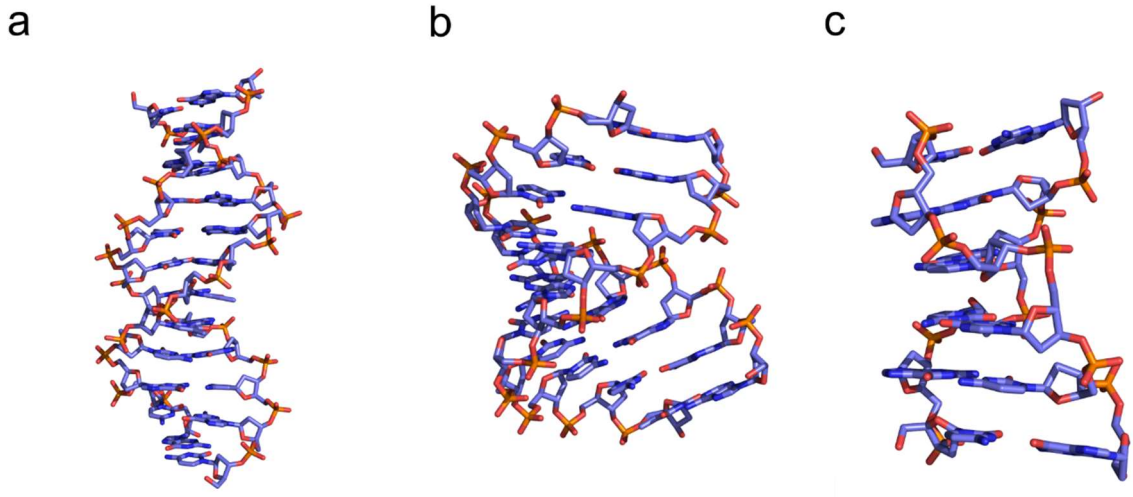


Figure 1-2 Helical structures of DNA. (a) B-DNA forms a right-handed helix with a large major groove. It is the most common form of DNA (PDB entry 1BNA)[16]. (b) The A-form of DNA is a right-handed, structurally condensed helix (PDB entry 440D)[17]. (c) In comparison to B- and A-DNA, Z-DNA forms a left-handed helix (PDB entry 4FS5)[18].

A third biological active form of DNA is Z-DNA (Figure 1-2 c). This form of DNA differs profoundly from the previously described A- and B-forms in that it is a left-handed helix. Z-DNA can form under certain circumstances in DNA sequences of alternating pyrimidine-purine nucleotides in high salt conditions, or in supercoiled DNA. For a long time, Z-DNA was believed to have no biological relevance, but later studies have shown that certain proteins interact with Z-DNA very strongly and suggested a biological role in defense of viral diseases[19].

1.1.3 The genome and DNA organization within the cell

The genetic information of each cell is stored in the nucleotide sequence of the two DNA strands using the four-letter alphabet of the nucleobases A, G, C and T. Within this sequence, segments that contain the information for a protein are called coding regions or genes. The other parts of DNA, which were erroneously considered to be junk-DNA in earlier times, are called noncoding regions. These for example contain tandem repeats, telomeres, transposable elements, or regions that code for functional RNAs such as transfer RNA (tRNA) or ribosomal RNA (rRNA). Additionally, they contain regions of important regulatory function, i.e. promoter regions that allow regulation of gene expression. The entirety of the genetic information stored in all cells of an organism is called the genome.

In eukaryotic cells the DNA is stored in the nucleus. Due to the large size of most eukaryotic genomes, the DNA cannot simply be stored as a long polymer within the nucleus but has to be packed into a more compact and condensed form. For example, each human cell contains about two meter of DNA that must fit into the nucleus with a diameter of less than 10 µm. At the same time, specific regions of the genome need to be accessible in a coordinated and dynamic fashion so that the cell can react to the ever-changing environment via transcriptional activation or repression of certain genomic regions. To achieve this, eukaryotic DNA is associated with specialized proteins that manage folding and coiling of DNA, leading to its multilevel organization into chromatin and chromosomes (Figure 1-3).

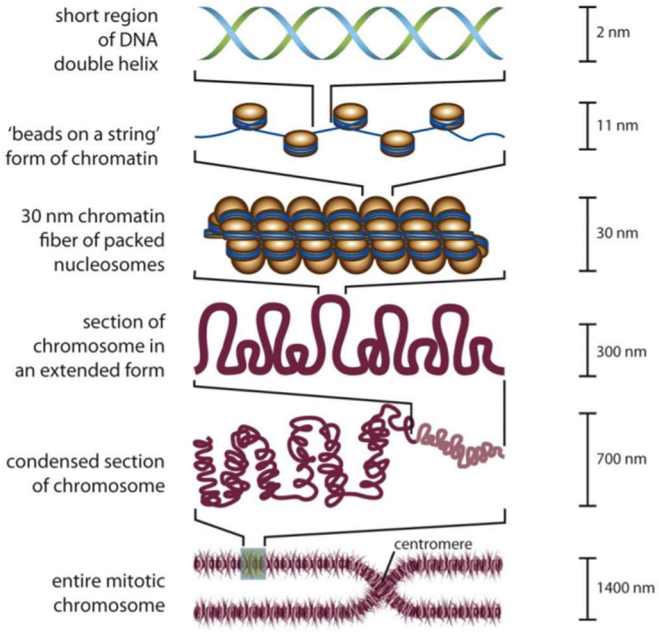


Figure 1-3 DNA packaging into chromosomes. Different levels of chromatin organization within the cell nucleus. Modified from[20].

The first and basic layer of organization within the chromosome is the formation of nucleosomes. Thereby, DNA associates with the proteins of the nucleosome resulting in a chromatin structure in which the DNA´s length is reduced to about one third. In each nucleosome, a DNA segment of approximately 165 bp is wrapped around a disc-shaped, octameric histone protein complex, the so-called nucleosome core particle (NCP). Each NCP consists of two copies of each of the core histone proteins H2A, H2B, H3 and H4. A common structural motif of these four histone proteins is the histone-fold motif, which is formed from three α-helices connected via two loops. The histone-fold motif is needed for interaction between the histone proteins to form heterodimers of H2A/H2B and H3/H4. Two H3/H4 heterodimers are forming a H3/H4 tetramer which associates with two H2A/H2B dimers, and together they form the NCP. Protruding from the NCP are the N-terminal ends of the histones, the so-called histone tails, which are subject to numerous covalent modifications that have an influence on the chromatin structure[12]. In this first layer of organization, the nucleosomes are spaced via linker DNA which can vary from roughly 10 to 80 bp in length and leads to the characteristic “beads-on-a-string” resemblance of the 10 nm chromatin fiber. In the second organization level, a fifth histone protein H1 interacts with the nucleosomes and further condenses the chromatin to the 30 nm fiber. How exactly the chromatin is

packed in the 30 nm fiber is not completely understood yet, but most likely follows the Zigzag model[21]. Furthermore, the chromatin is arranged spatially within the nucleus. The model of fractal globule formation in the local chromatin environment was provided by Dekker *et al.* in 2009[22] and suggests the partitioning of chromatin into certain chromosomal territories or domains such as for example the topologically associating domains (TADs) or the lamina associating domains (LADs). This has provided insights into short- and long-range interactions between distant genomic regions within a chromosome or even different chromosomes and their role in transcriptional regulation or chromosomal translocation[23].

1.1.4 DNA replication

During DNA replication, the genetic material of a cell is duplicated. Since the DNA copies are passed on to new daughter cells in the subsequent cell division, it is very important that DNA replication is complete and accurate to maintain the genomic integrity of the daughter cell. DNA replication occurs in a semiconservative manner[24] meaning that each of the two parental DNA strands serves as a template for an entire new strand.

The DNA replication can be divided into three parts: (1) initiation, (2) elongation, and (3) termination. Prior to the initiation, the pre-replication complex (PRC) is assembled at the origin of replication which is the starting point of replication in the genome. The two DNA strands are unwound and separated by helicases, which leads to the establishment of the replication fork. DNA replication is performed by DNA polymerases. These enzymes have a high fidelity in copying DNA with an error rate of only about 1 mistake in 10^9 nucleotides copied. This high fidelity is crucial for the faithful transmission of copied genetic information to the daughter cells and results from the DNA polymerase´s intrinsic 3´ – 5´-exonuclease activity. DNA polymerases can only synthesize a new DNA strand in the 5´ – 3´ direction, but every time a nucleotide is incorporated into the new DNA strand it is proof-read and eventually cleaved out by the 3´ – 5´ exonuclease activity. The unidirectional synthesis ability of DNA polymerase requires specific structures when using both DNA strands as templates. In the elongation phase, the so-called leading strand can be readily duplicated by the DNA polymerase in the synthesis direction 5´ to 3´. Thereby, new nucleotides are attached to the free 3´-OH group of the previously synthesized nucleotide via a phosphodiester bond. However, this is not possible for the second DNA strand, the antiparallel lagging strand. Instead, DNA primases synthesize short RNA primers, so-called Okazaki fragments, on the lagging strand which are then elongated by the DNA polymerase. After elongation, the RNA primers are

removed, and the DNA fragments are ligated. In eukaryotes, DNA replication occurs simultaneously at multiple origins of replication. Since the chromosome is a linear molecule, replication forks meet at some points and terminate replication. Due to the linearity, DNA polymerase cannot reach the very end of the chromosome. Therefore, each time a cell replicates, parts at the end of the chromosomes, the telomeres, are shortened. The telomeres prevent the loss of those chromosome parts that contain genes. Due to the finite length of the telomeric regions, DNA replication in somatic cells can only occur a certain number of times before DNA loss prevents further DNA replication[12, 25].

1.1.5 From DNA to protein

Every cell needs to access and express the genetic information stored on its DNA to be able to fulfill its specific functions. The genome of each eukaryotic cell is basically identical; however, different cells have distinct functions which are defined by the expression of only a specific subset of their entire genetic information. Deciding which part of the genome must be used in a specific type of cell is a very complex task that involves various control processes, also including epigenetic mechanisms. The general objective of gene expression is to transfer the information contained in a specific sequence of DNA into a complementary sequence of RNA. Like DNA, RNA also consists of four nucleotides that are covalently linked to form a long polymer. Instead of deoxyribose, in RNA the sugar moiety is ribose and the nucleobase T is replaced by uracil (U). Despite these differences, the genetic information is retained in the nucleotide sequence of RNA referring to this process as transcription. Depending on the information contained in the DNA sequence, several types of RNA molecules can be transcribed. The RNAs are basically divided into (protein)-coding and noncoding RNAs. RNA coding for a specific protein is called messenger RNA (mRNA) and contains the information of the protein´s respective gene sequence on the DNA. mRNA is used to transmit this information from the DNA in the nucleus to the cytoplasm where it is used as a template to synthesize the protein it encodes for. Hence, as its name implies, mRNA serves as a messenger molecule which is degraded after protein synthesis. The group of noncoding RNA contains several different types of RNA including transfer RNA (tRNA) and ribosomal RNA (rRNA). In contrast to protein-coding mRNA, these two RNA types do not code for a protein but are functional molecules themselves and perform actively in the process of translation during protein biosynthesis. During translation, the mRNA sequence is translated into a protein sequence at the ribosome (Figure 1-4). rRNA is part of the ribosome, a cytosolic RNA/protein complex that catalyzes protein biosynthesis, whereas tRNA is responsible for providing the amino acid building blocks used for making a protein to the

ribosome. In addition to tRNA and rRNA, the group of noncoding RNAs also includes RNA molecules that control gene expression; e.g. long noncoding RNAs (lncRNAs) or long intergenic noncoding RNAs (lincRNA) such as Hox transcript antisense intergenic RNA (HOTAIR)[26].

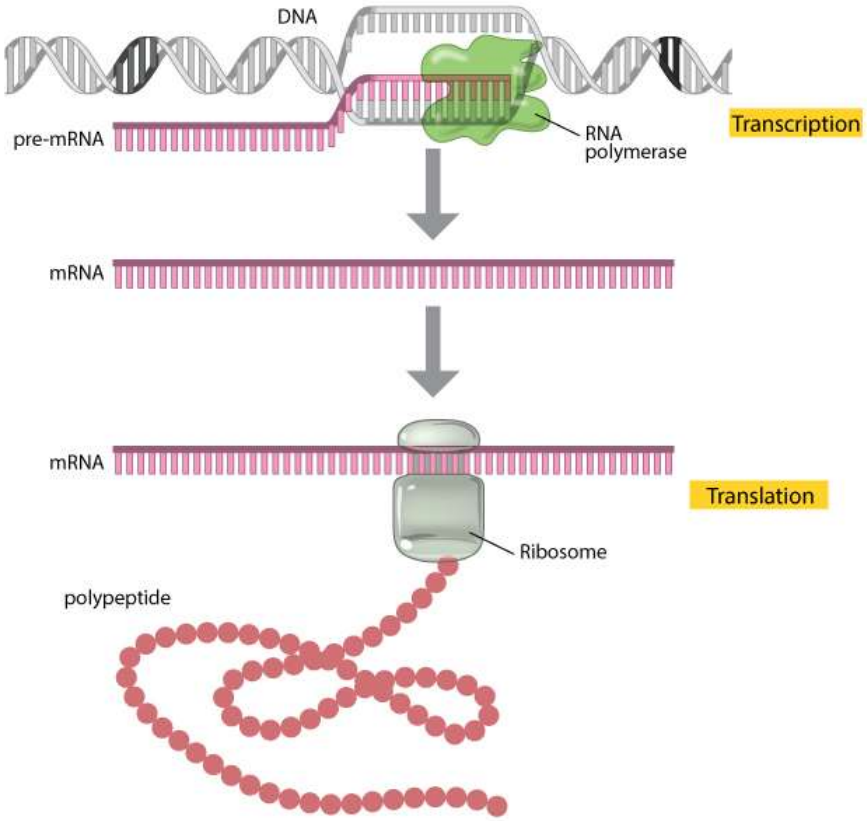


Figure 1-4 Schematic overview of transcription and translation. Protein biosynthesis is divided into two parts: During transcription, the RNA polymerase makes a pre-mRNA transcript from a DNA template. After processing of the pre-mRNA transcript, the resulting mRNA is translated into the protein amino acid sequence at the ribosome. Modified from[27].

Transcription in eukaryotes is a complex process involving a range of proteins and factors. It can be divided into three general steps: (1) initiation, (2) elongation, and (3) termination. RNA polymerases are the enzymes that are responsible for making RNA from a DNA segment. In eukaryotes, there are three different RNA polymerases named RNA polymerase I, II and III, which are each responsible for making distinct types of RNA. RNA polymerases I and III transcribe genes that encode for tRNA, rRNA, and various small RNAs, whereas RNA polymerase II is responsible for transcribing genes that encode for proteins and thus produces mRNA. Before transcription starts, the pre-initiation complex consisting of a set of transcription factors and the RNA polymerase is assembled at the gene promoter. The promoter is

a specific DNA sequence that lies upstream (in 5´ direction) of the transcription start site (TSS) of the respective gene. In many RNA polymerase II-transcribed genes, the promoter contains a T and A rich sequence called TATA-box. This sequence is recognized by one of the transcription factors and leads to the gradual assembly of the pre-initiation complex. After RNA polymerase is released from the initiation complex, it moves along the DNA and synthesizes the mRNA molecule. This elongation phase is tightly coupled with RNA processing steps such as 5´-capping and 3´-polyadenylation of the nascent RNA molecule in eukaryotes. These specific modifications of the RNA molecule provide a signal for complete transcription, transcription termination and nuclear export of the newly synthesized RNA molecule. After transcription is completed and the readily processed mRNA molecule is exported from the nucleus, it is recruited to the ribosomes in the cytoplasm where a second process – translation of the mRNA transcript – takes place. As the name “translation” implies, the information stored in the mRNA nucleotide sequence is read and translated into a protein´s amino acid sequence. This translation of information is dictated by the genetic code. Thereby, three consecutive RNA nucleotides (a codon) codes for one of the 20 canonical amino acids. Since there are 64 (4^3) different codon possibilities but only 20 amino acids, each amino acid is encoded by several codons meaning that the genetic code is redundant. At the ribosome, the mRNA codons are read, and the respective protein is assembled. During this process, tRNA molecules transport the required amino acid building blocks to the ribosome. Faithful transfer of information encoded in the codon sequence of the mRNA is indispensable at this step. Therefore, each tRNA molecule displays a three nucleotide-long anticodon sequence that is (a) specific for the amino acid that is carried by the respective tRNA molecule, and (b) is complementary to the mRNA codon that encodes for the specific amino acid. Translation always starts with a start codon (AUG) coding for methionine and ends with one of the three stop codons UAA, UGA, or UAG that do not code for an amino acid. The stop codon signals the ribosome to terminate translation and to release the newly synthesized protein into the cytoplasm. Once released from the ribosome, the protein must fold into its correct three-dimensional structure to be functional.

1.2 Epigenetics

Although the majority of all cells in a multicellular organism have the same genotype, they differ greatly in their patterns of gene expression and exhibit cell type specific phenotypes. This phenomenon could not be explained by the genetic mechanisms known at the time of the early 20th century and hinted at an additional layer of information stored on the DNA and inherited during cell division – without changing the underlying DNA nucleotide sequence. A first definition of the term "epigenetics" was made by Conrad Waddington in 1942, describing it as "a branch of biology which studies the causal interactions between genes and their products which bring the phenotype into being"[28]. Since then, the term epigenetics itself has undergone an evolution resulting in our present understanding thereof as "the study of mitotically and/or meiotically heritable changes in gene functions that cannot be explained by changes in the DNA sequence"[29]. The Greek prefix *epi* meaning "upon, on top" of refers to the nature of epigenetic modifications being superimposed on the DNA sequence/chromatin. Epigenetic modifications consist of (1) modifications of DNA (e.g. methylation), and (2) modifications of histone proteins (methylation, acetylation, ubiquitinylation). Together they form the epigenome – a tightly interacting network of heritable epigenetic modifications within the cell, shaping its chromatin and modulating genome function. Shaping the chromatin structure has fundamental implications on the accessibility of the DNA i.e. for transcription factors or other DNA binding proteins. Epigenetic modifications provide an important mechanism to control gene expression by restructuring the chromatin into transcriptionally inactive heterochromatin or transcriptionally active euchromatin. Tightly regulated gene expression plays important roles for development and differentiation of cells, since it ultimately determines a cell´s specific phenotype. Giving an idea of how important and fundamental epigenetic mechanisms are, it seems obvious that aberrant epigenetic modifications are a severe risk to cellular integrity and are potential triggers for disease[20, 30].

Today, epigenetic research mainly focusses on further unravelling the function and significance of epigenetic modification crosstalk, as well as studying the function of individual epigenetic marks, the mechanisms that introduce and maintain them, and their global regulation within the cell.

1.2.1 Epigenetic DNA modifications

1.2.1.1 Cytosine modifications in mammalian genomes

Methylation of the pyrimidine fifth position of cytosine (5-methylcytosine, 5mC) is a highly conserved epigenetic modification occurring in plants, animals and fungi[5b]. Being the first epigenetic mark identified[3], 5mC was found to play fundamental roles in genome stability, tissue-specific gene expression, cell differentiation, and developmental processes such as X-chromosome inactivation and imprinting[4-5, 31] in mammals. The wide distribution of 5mC as well as its importance in cellular processes have led to the common designation of 5mC as the "fifth base". Methylation of the mammalian genome almost exclusively occurs in the context of palindromic cytosine-phosphate-guanine (CpG)[5b], where both strands of DNA are methylated in an asymmetric manner. Methylated CpGs mainly occur in transposons, intergenic regions, and gene bodies, but less than 10 % are found in CpG islands (CGIs) – specific CG dense regions of the genome which are associated with promoters of housekeeping and developmental regulator genes[32]. Around 60 % of all human promoters are associated with CGIs; even if they have been suggested to be largely unmethylated, a subset undergoes tissue-specific methylation during development[5a, 33] and becomes methylated in differentiated tissues[34].

In somatic cells of the adult organism CpG methylation levels are usually stably maintained at a relatively high level of 60-80 % and patterns are faithfully inherited to daughter cells[5a, 35]. However, DNA methylation was found to be highly dynamic and undergoing drastic changes during specific developmental stages such as the embryonic preimplantation development and developing primordial germ cells (PGCs)[36]. In the latter, initial global demethylation is important for setting up pluripotency and erasing the paternal epigenetic methylation pattern[37]. The rapid decrease in methylation levels of the zygotic paternal genome could not be explained by replication-dependent dilution of 5mC[36b], but rather hinted at an active demethylation mechanism[38]. However, this mechanism was not revealed until 2009, when the oxidized 5mC derivative 5-hydroxymethylcytosine (5hmC) was shown to accumulate in certain tissues. Subsequently, the human ten-eleven translocation (TET) 1 DNA dioxygenase was identified to generate 5hmC from 5mC[6]. In the following, it was also shown that the other two TET family members, TET2 and TET3, are able to oxidize 5mC to 5hmC[39]. This is only the first TET-mediated oxidation step in an active demethylation cycle (Figure 1-5) where 5hmC is further oxidized to first 5-formylcytosine (5fC)[7] and then 5-carboxylcytosine (5caC)[7a, 8]. Restoration of unmodified cytosine (demethylation) is completed by thymine DNA glycosylase (TDG) that cleaves 5fC or 5caC from the DNA leading to an abasic site. Lastly, the abasic site is restored to cytosine through base excision repair (BER)[8, 40].

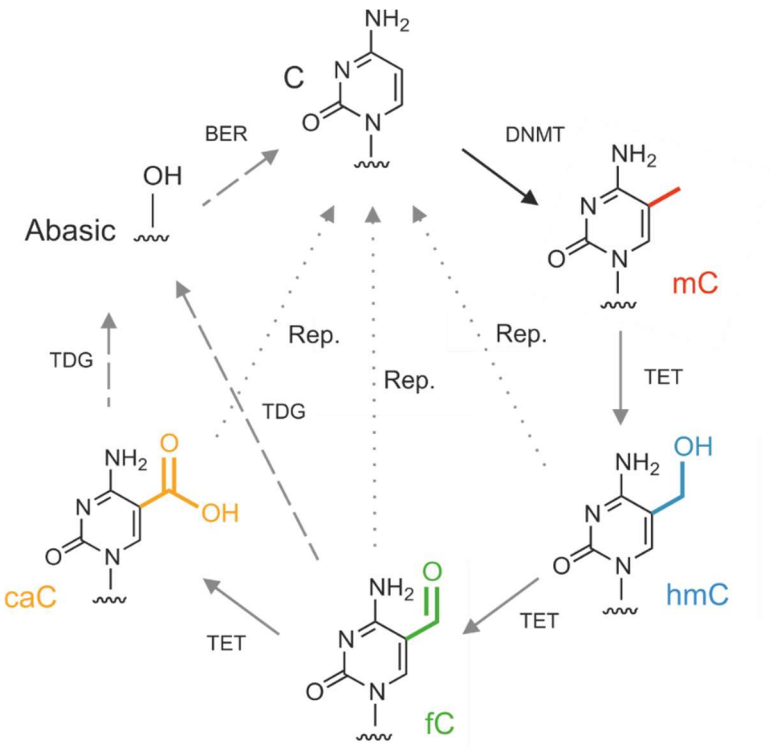


Figure 1-5 Cycle of active demethylation. Formation of 5-methylcytosine (5mC) occurs via methylation of cytosine at the carbon 5-position and is catalyzed by DNA methyltransferases (Dnmts). In the following, stepwise oxidation of 5mC to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and then 5-carboxylcytosine (5caC) is catalyzed by ten-eleven translocation (TET) DNA dioxygenases. Both, 5fC and 5caC are substrates for thymine DNA glycosylase (TDG) and can be excised from the DNA which leads to the formation of an abasic site. In a last step, cytosine is restored from the abasic site by base-excision repair (BER). Modified from[41].

The global genomic content of 5hmC, 5fC, and 5caC was found to be cell type dependent, and unlike 5mC levels, highly variable in different tissues[7, 42]. The most abundant and stable of the oxidized 5mC forms is 5hmC, with levels of approximately 5 % of 5mC. The other two oxidized variants, 5fC and 5caC, have comparably lower levels. Levels of 5fC range from 0.06 % to 0.6 % of 5mC, while the level of 5caC is approximately 0.01 % of 5mC[7-8, 43]. In adult mice, 5hmC was shown to be present at high levels in the central nervous system, and was also found in kidney, heart, spleen and thymus tissue, as well as in mouse embryonic stem cells (mESCs)[6b, 7a, 42b-d]. So far, highest steady-state levels for 5hmC were detected in mouse and human brain[44]. Interactions of 5hmC with specific proteins lead to the suggestion that 5hmC, just like 5mC, might function as a distinct epigenetic mark itself justifying its designation as the 6th nucleobase[45]. Similar to 5hmC, 5fC also has been detected in many cell types including all major organs, mESCs and human neurons[7, 44, 46]. Several analyses of genomic 5fC profiles

showed its presence at regulatory elements such as enhancers and promoters[47], and interactions with specific 5fC reader proteins were revealed[45a]. These findings combined together with the report of 5fC being a predominantly stable base[42a] are raising the question whether 5fC itself has independent epigenetic functions or if it is rather a semi-permanent base with a limited life-time[48]. Global levels of 5caC are comparably low, but were detected in a few cell types as well as mESCs[7a, 8, 43]. Recently, elevated levels of 5caC were reported for human breast cancer, glioma, medulloblastoma and ependymoma tissues[49]. Despite its rather low abundance, several 5caC reader proteins were identified[45a, 50] and the ability of both 5fC and 5caC to interfere with cellular processes such as transcriptional regulation was demonstrated[51].

While active demethylation and normal patterns of TET-oxidation products occur in pre-implantation development, PGCs, pluripotency and differentiation[52], aberrant levels of cytosine modifications have been linked to various diseases. For example, global reduction of 5hmC levels have been observed in various cancers, probably due to dysfunctional TET proteins, increased production of TET-inhibiting metabolite 2-hydroxygluturate, or from silencing of TET proteins due to hypermethylation of their encoding genes[53]. Additionally, aberrant methylation of normally unmethylated CpG islands in somatic cells is associated with pathogenic developments. Hypermethylation of so-called tumor suppressor genes is one example where unusual methylation leads to a loss of function in a gene, which ultimately promotes cancer development. Also, methylation of DNA repair genes such as breast cancer 1 (BRCA1), MGMT O6-methylguanine-DNA methyltransferase (MGMT) and mutL homolog 1 (MLH1), which leads to their transcriptional silencing, is associated with cancer. Even if various connections between aberrant patterns of epigenetic cytosine modifications and abnormal cell behavior have been made, there is still further investigation required to fully understand the epigenetic network, its mechanisms, and the biological relevance of each of the individual cytosine modifications, especially of the less abundant 5fC and 5caC oxidation variants.

1.2.1.2 DNA methyltransferases (Dnmts)– Writers of DNA methylation

In mammals, three different DNA methyltransferases namely Dnmt1, Dnmt3a, and Dnmt3b are responsible for the introduction and the maintenance of cytosine DNA methylation. Methylation occurs by transfer of the methyl group from the universal methyl donor *S*-adenosyl-L-methionine (AdoMet) to the C5 position of cytosine, catalyzed by the Dnmt enzyme (Figure 1-6).

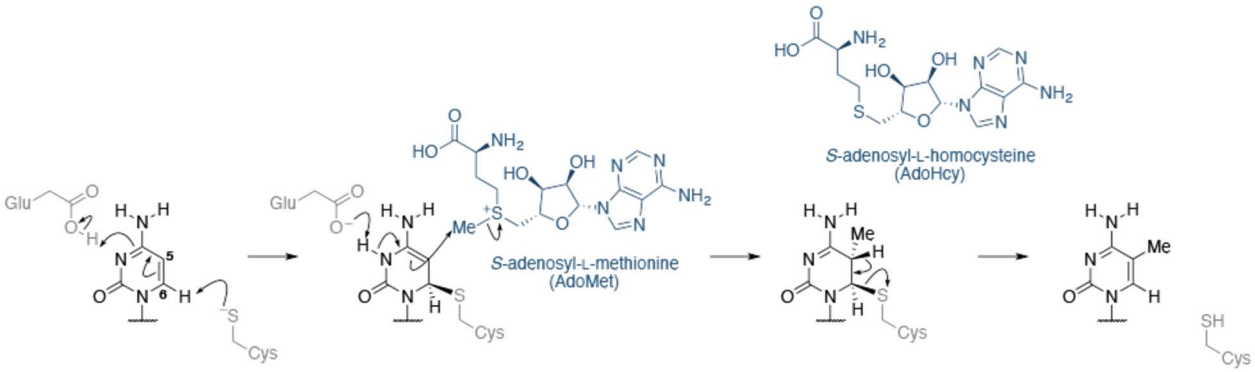


Figure 1-6 Mechanism of cytosine methylation. DNA methyltransferases contain a conserved cysteine residue, which in its deprotonated form (thiolate anion) serves as a nucleophile for attacking the C6 atom of cytosine. In the following step, a nucleophilic attack on the methyl group of the cofactor *S*-adenosyl-L-methionine (AdoMet) takes place, converting AdoMet into S-adenosyl-L-homocysteine (AdoHcy) and transferring the methyl group onto the cytosine C5 position. In a last step, the enzyme is released via a β-elimination. Modified from[54].

Dnmt1, together with its obligate partner, the ubiquitin-like plant homeodomain and RING finger domain 1 (UHRF1)[55], was found to preferentially act on hemimethylated substrates. It is therefore classified as a maintenance methyltransferase with the primary responsibility to restore the methylation pattern of a cell after chromosome replication and repair. In contrast, Dnmt3a and Dnmt3b were found to introduce cytosine methylation on nonmethylated DNA, and are therefore called *de novo* methyltransferases[56]. *De novo* methyltransferases are associated with the enzymatically inactive regulatory factor Dnmt3-like protein (Dnmt3L)[57] and are especially important during early embryonal development when specific methylation patterns must be established. Inactivation of Dnmt3a and Dnmt3b in embryonic stem cells have confirmed their role in mammalian development[58]. A fourth Dnmt family member, Dnmt2, was found to have no discernable effect on DNA maintenance or *de novo* methylation, but instead was found to methylate tRNA[59].

Mammalian Dnmts are composed of a highly conserved catalytic C-terminal domain and a variable N-terminal regulatory domain (Figure 1-7). The catalytic domain harbors the active center of the enzyme,

which contains the amino acid motif AdoMet-dependent MTase fold[60]. This motif is responsible for binding of Dnmt to its cofactor *S*-adenosyl-L-methionine (AdoMet) and targeting of the cytosine substrate. The C-terminal signature motifs I, IV, VI, IX and X are most conserved in all cytosine methyltransferases; motifs I and X are involved in cofactor binding, whereas motifs IV and VI have catalytic functions. The N-terminal regulatory domain differs between the Dnmts. In the case of the *de novo* methyltransferases, Dnmt3a and Dnmt3b, the N-terminal part contains a ubiquitous eukaryotic protein module containing the the strongly conserved amino acid motif Pro-Trp-Trp-Pro (PWWP). The PWWP motif may be involved in unspecific DNA binding and was shown to mediate binding to methylated lysine on histones[61]. Also located in the N-terminal regulatory domain is the alpha-thalassemia mental retardation X-linked (ATRX)-related cysteine-rich region. This domain contains a C2-C2 zinc finger and an atypical plant homodomain (PHD) and is responsible for protein-protein interactions. For example, Dnmt3L interacts via this domain with histone modifications and leads to the recruitment or activation of the *de novo* methyltransferase Dnmt3a[62]. Characteristic for the Dnmt1 N-terminal regions are the proliferating cell nuclear antigen (PCNA)-interacting domain and the replication foci-targeting (RFT) domain. Since Dnmt1 is responsible for the maintenance of the methylation status in a cell, the DNA methyltransferase is localized at the replicating foci during the S-phase of the cell cycle where it interacts with the DNA replication machinery.

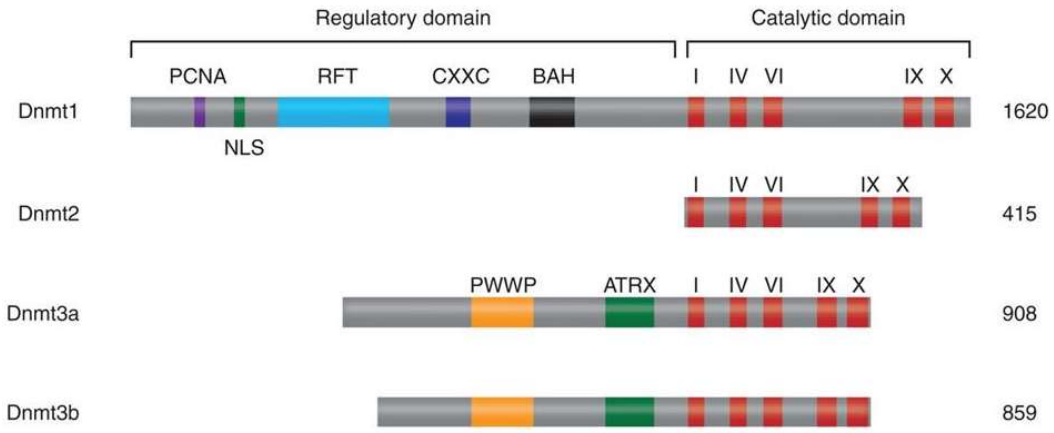


Figure 1-7 Schematic depiction of the mammalian DNA methyltransferase (Dnmt) protein family. The catalytic domains are conserved among the Dnmt family members Dnmt1, Dnmt2 and Dnmt3. The signature motifs I, IV, VI, IX and X are most conserved in all cytosine methyltransferases, whereas they differ largely in their N-terminal regulatory domain. Domain abbreviations: PCNA-interacting domain (PCNA), nuclear localization signal (NLS), replication foci-targeting domain (RFT), bromo-adjacent homology domain (BAH), domain containing "proline-tryptophan-tryptophan-proline" motif (PWWP), cysteine-rich domain implicated in binding DNA that contains CpG dinucleotides (CxxC). Modified from[63].

1.2.1.3 Methyl-CpG-binding proteins (MBDs) – Readers of DNA methylation

DNA methylation within the CpG context is generally associated with transcriptional silencing in mammalian genomes. This can be accomplished in two ways; either DNA methylation abolishes binding of transcription factors and thereby prevents transcription, or methyl-CpG-binding proteins are recruited to the methyl-CpG sites and promote transcriptional silencing processes. These "readers" of DNA methylation are proteins that specifically bind to methylated CpG dinucleotides and mediate the epigenetic crosstalk between DNA methylation, histone modifications and chromatin organization. Among these proteins, the methyl-CpG-binding domain (MBD) protein family is the most prominent one including methyl-CpG-binding protein 2 (MeCP2), MBD1, MBD2, MBD3, and MBD4 (Figure 1-8). Other protein families are the Kaiso family and the SET-Ring finger-associated (SRA) domain family[64]. Common to all MBD protein family members is the MBD domain which binds single symmetrically-methylated CpGs[65]. However, MBD3 as well as the most recently discovered MBD5 and MBD6 were shown to behave differently. MBD3 is binding to unmethylated DNA and to 5hmC-containing DNA, while MBD5 and MBD6 do not bind to methylated DNA[45d, 66].

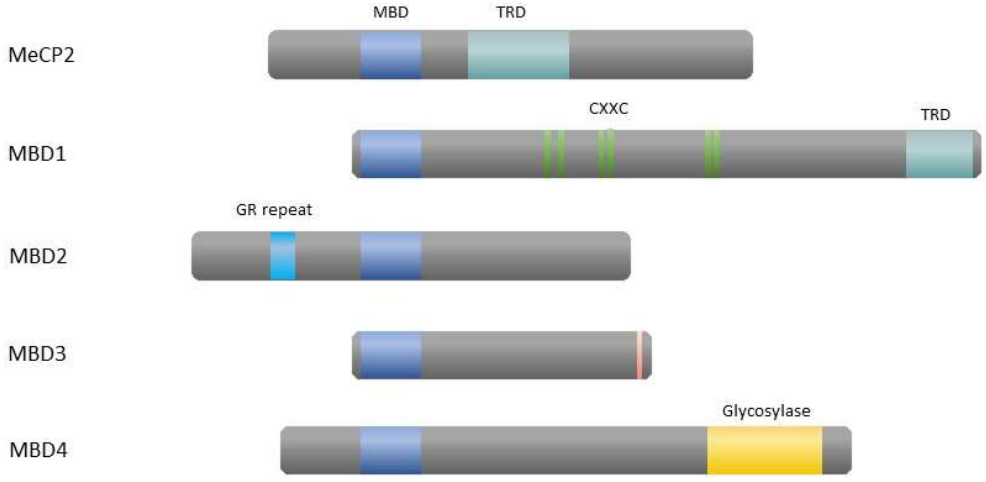


Figure 1-8 Schematic depiction of the methyl-CpG-binding domain (MBD) protein family. The MBD protein family consists of the five proteins MeCP2, MBD1, MBD2, MBD3, and MBD4. Domain abbreviations: Methyl-CpG-binding domain (MBD), transcriptional repression domain (TRD), cysteine-rich domain implicated in binding DNA that contains CpG dinucleotides (CxxC). Modified from[30a].

In addition to the MBD domain, different catalytic or interaction domains, including transcription repression domains (TRDs), unmethylated-CpG-binding zinc finger (CxxC) domains or glycosylases, are required for the individual functions of each MBD family member[67]. This allows MBD protein family members to interact with other components of the epigenetic-based transcriptional repression machinery, most importantly histone deacetylases or chromatin remodelers. An example for the interplay of DNA methylation and histone deacetylation is the interaction of MeCP2 with the Sin3A histone deacetylase (HDAC) complex, which is believed to be responsible for transcriptional silencing[68].

1.2.1.4 Ten-eleven translocation (TET) DNA dioxygenases – Erasers of DNA methylation

Active DNA demethylation in mammals occurs via the TET/TDG pathway shown in Figure 1-5. The main enzymes acting in this pathway are TET enzymes which are iron (II)/α-ketoglutarate (Fe (II)/α-KG) -dependent DNA dioxygenases. Common to all the TET family members (TET1, TET2 and TET3) is a C-terminal core catalytic domain, which consists of a double-stranded β-helix (DSBH) domain and a cysteine-rich domain (Figure 1-9)[69]. The DSBH is important for colocalization of Fe (II), α-KG and DNA for oxidation of the cytosine modification and is further stabilized in the DNA-bound state by the cysteine-rich domain wrapping around the DSBH domain. TET family members TET1 and TET3 were found to exist in various splicing isoforms, with some of them being truncated at the N-terminus and lacking the DNA-binding CXXC zinc finger domain that is present in full-length TET1 and TET3 (TET1e and TET3FL, respectively). Isoforms of TET1 and TET3 were associated with different developmental stages or cell types[70]. TET enzymes preferentially act on 5mC in the context of CpGs. Thereby, oxidation of 5mC to 5hmC is favored over conversion of 5hmC to 5fC, or 5fC to 5caC, which is probably due to differential binding and catalytic activity on the single differently modified cytosines[7a] or their variable ability for hydrogen abstraction[71].

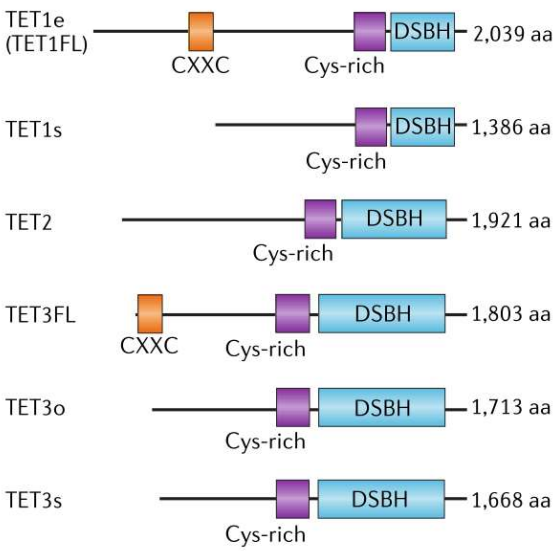


Figure 1-9 Domain structure of ten-eleven translocation (TET) DNA dioxygenases. The catalytic cysteine-rich and double-stranded β-helix (DSBH) domains are located at the C-terminus. Full-length TET1 (TET1FL) and TET3 (TET3FL) contain an N-terminal CxxC domain. Modified from[52].

1.2.1.5 Structural and chemical properties of cytosine modifications in DNA

By displaying their unique chemistries via the DNA major groove of the double-stranded DNA, 5mC and its oxidized variants provide a specific major groove surface for interaction partners. In mammalian genomes, a majority of the DNA binding proteins interact via the DNA minor or major groove. Consequently, they are exposed to a diverse pattern of modified cytosines, which can favor or prevent specific interactions. This modulation of DNA-protein interactions can effectively regulate biochemical processes such as gene transcription or restructuring of the chromatin. Furthermore, the influences of the individual modified cytosines on the duplex stability, local structure, or flexibility of DNA are of great interest regarding regulating processes that depend on the specific local state of the DNA. Subsequently, the impact of 5mC and its oxidized variants has been extensively studied in terms of their chemical and structural properties. For many of these studies, the well-studied prototypic B-DNA oligomer with the sequence d[CGCGAATTCGCG]$_2$, the Dickerson-Drew Dodecamer (DDD)[16], served as a DNA template.

Dickerson-Drew Dodecamers with single 5mC, 5hmC, 5fC or 5caC incorporated into the DNA sequence showed that regular Watson-Crick base pairing is favored[72] over wobble base pairing[72c, 73]. Even if they do not interfere with regular base pairing, individual oxidized 5mCs are still able to influence the stability

or the local structure of the DNA duplex. It was shown that in a DDD, where a single oxidized 5mC was incorporated, 5caC stabilizes the DDD whereas the presence of 5mC, 5hmC or 5fC had only minor to no effects on DNA duplex stability. Increased DDD stability was evidenced by an increase in the melting temperature of the dodecamer[72c, 74]. This is likely to be attributed to electronic dipole-dipole interactions associated with 5caC[40a, 73] rather than resulting from improved base-stacking geometry. Also, 5caC – unlike 5hmC and 5fC – does not promote base-pair opening within the neighboring DNA sequences which further supports its stabilizing effect[72c]. The influence of enriched modified cytosines within a DNA sequence on duplex stability has been studied for 5fC. While DNA duplex stability was reported to not be influenced by the increased number of formylated cytosines, the DNA duplex conformation appeared to be altered[75]. Crystal structures showed an unknown DNA conformation with helical underwinding, termed F-DNA, which was hypothesized to be involved in the recruitment of 5fC-reader proteins. This finding was challenged by the crystal structure reported by Hardwick *et al.* where crystals were obtained from the same sequence but using different crystallization conditions[76]. In this case, the crystal structure did not show a conformational change in the DNA, which was further supported by NMR data. Lastly, looking at the DNA flexibility and how this is influenced by the different cytosine variants, a study by Ngo *et al.* reported a remarkable increase in DNA flexibility upon introduction of a single 5fC copy into a 90 bp long DNA duplex[77]. This enhanced flexibility was also seen for DNA duplexes containing 5hmC, but the effect was less evident. In contrast, no measurable effect was detected for 5caC. Consistent with previous results[78], 5mC decreases DNA duplex flexibility which might be attributed to the bulky methyl group that restricts conformational fluctuation[79]. DNA flexibility was previously linked to enhanced nucleosomal mechanic stability[80], allowing the DNA to bind to the histone octamer in a more stable manner and leading to a limited exposure of the DNA to the transcription machinery. This again correlates with the fact that most 5fCs are found near poised enhancers and at the transcription start sites of low-expression genes[47a].

In summary, the overall effects of modified cytosines on DNA stability and conformation are rather moderate, however DNA flexibility appeared to be influenced by the presence of distinct cytosine variants. In contrast, the chemical properties of 5mC and its oxidized variants differ greatly from each other and therefore provide targets for specific DNA-protein interactions. A basis for the differential recognition of the modified cytosines are their large C5-substituents that have an increased steric demand and vary in their conformational flexibility, as well as their individual ability to donate or accept hydrogen bonds. The formation of an intra-nucleobase hydrogen bond between the exocyclic N4-amino group and the formyl or carbonyl group oxygen was observed for 5fC and 5caC, respectively[72c, 81] (Figure

1-10). This leads to a fixation of the C=O bond in the plane of the nucleobase. In contrast, the C-O bond in 5hmC was found to be conformationally more flexible than the formyl or carboxyl groups, with an observed partial occupancy of two conformations within the DNA duplex[72c].

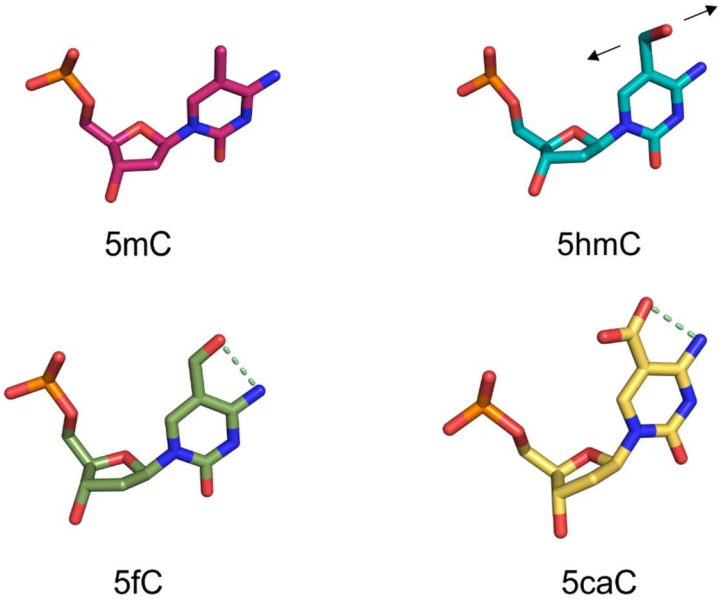


Figure 1-10 Crystal structures of the cytosine derivatives in DNA duplexes. Conformational flexibility observed for the C-O bond of 5-hydroxymethylcytosine (5hmC) is indicated by black arrows. Intramolecular hydrogen bonds observed in 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) are indicated through dashed lines (green). PDB entries: 4GJP for 5mC[10c], 4I9V for 5hmC[72c], 4QC7 for 5fC[72c], 4PWM for 5caC[72c].

While these effects influence and guide nucleobase-specific interactions with DNA-binding proteins, the chemical properties of the C5-substituent also have an effect on the ability of a nucleobase to be cleaved out by TDG in the active demethylation process. The electron-withdrawing C5-sustituent of 5fC increases the N1 acidity leading to a weaker N1-C1´ bond and thereby enhances the leaving group potential of the nucleobase[40a, 73]. Together with its ability to form strong polar interactions at the TDG enzyme recognition site[82], this favors TDG-mediated 5fC excision over excision of 5caC[40a]. Even if 5caC similarly interacts with TDG[83], its leaving-group potential is hampered by the C5-carboxyl group, which is negatively charged in physiological pH and therefore does not promote an N1 acidity like in 5fC[73].

1.3 Detecting cytosine modifications in the genome

The DNA major groove code established by the modified cytosines 5mC, 5hmC, 5fC, and 5caC constitutes an important informational layer, which is superimposed on the DNA sequence. In order to elucidate the functions of the individual modified cytosines, several methods for their specific detection at single-base resolution have been developed over the past years. Since cytosine modifications do not alter Watson-Crick base-pairing of canonical nucleobases, they cannot be detected by conventional methods that rely on sequencing analysis of the Watson-Crick sequence. Alternative detection methods exploit the information of the major groove, which is unique to each of the modified cytosines. This has led to the development of chemistry-based sequencing methods including chemical conversion reactions, affinity enrichment with specific tags or antibodies, nanopore sequencing and single-molecule real-time sequencing (SMRT) as well as the design of protein scaffolds with each of these methods exploiting the differential chemical or structural properties of the individual cytosine modifications.

1.3.1 Chemistry-based detection methods, SMRT and nanopores

To this day, the "gold standard" technique for single-base resolution detection of modified cytosines is bisulfite-sequencing (BS-seq). In this method, bisulfite is used to catalyze the hydrolytic deamination of cytosine to uracil, while methylated cytosines are not converted[84] (Figure 1-11). Since uracil has the same Watson-Crick base-pairing properties as thymine, cytosine and 5mC can be distinguished in a following sequencing reaction. Consequently, a genomic map at single-base resolution can be obtained to distinguish C and 5mC. Conventional BS-seq does not allow to discriminate between 5mC and 5hmC since 5hmC, like 5mC, is resistant to deamination[85]. A number of chemical and enzymatic methods have been developed that allow to distinguish between 5mC and 5hmC. In oxidative bisulfite sequencing (oxBS-seq)[86] (Figure 1-11) 5hmC is selectively oxidized to 5fC using $KRuO_4$ before subjecting the DNA to bisulfite treatment. Along with C and 5caC, 5fC is deaminated to uracil during the bisulfite reaction. By subtracting an oxBS-seq dataset from a BS-seq dataset obtained from the same DNA sample, the presence of 5hmC at single-base resolution can be revealed. A second method to resolve 5mC and 5hmC is TET-assisted bisulfite sequencing (TAB-seq)[87]. In a first step 5hmC is glucosylated using T4 phage β-glucosyltransferase protecting it from TET1-mediated oxidation in the following step. Consequently, 5mC and 5fC are iteratively oxidized to 5caC that is deaminated during bisulfite treatment, and only the protected 5hmC is not deaminated. This way, TAB-seq allows a direct readout of 5hmC at single-base resolution. In combination with a BS-seq data set, 5mC distribution can be determined as well. Also for

the detection of 5fC and 5caC several BS-seq-based detection methods such as reductive BS-seq (redBS-seq)[88], 5fC chemical modification-assisted BS-seq (fCAB-seq)[47a] or DNA immunoprecipitation-coupled CAB-seq (DIP-CAB-seq)[89] have been established. While these methods were successfully applied in numerous studies to detect the individual modified cytosines at single-base resolution, the main drawbacks are the harsh reaction conditions during the bisulfite treatment as well as the bisulfite-induced depyrimidination of DNA[90].

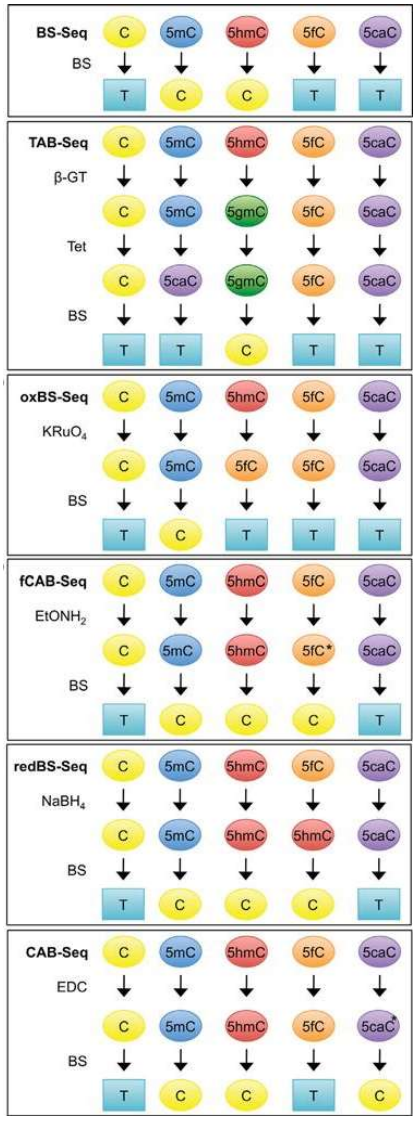


Figure 1-11 Overview of single-base resolution methods for the detection of cytosine modifications. Bisulfite-sequencing-based methods for the detection of 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) in DNA. Modified from[91].

Alternative methods to detect cytosine modifications at single-base resolution without the need for chemical transformation have been on the rise. For example, single-molecule real-time sequencing (SMRT-seq)[92] is a method in which fluorescently labeled nucleotides are incorporated opposite the template DNA strand by a processive DNA polymerase that is attached to a zero-mode waveguide (ZMW)[93]. This allows the observation of a single fluorescent nucleotide being incorporated at a time by detection of the fluorescent label, which is cleaved off upon incorporation. The kinetics for nucleotide incorporation depend on the nature of the nucleobase in the DNA strand and if a modification is present, DNA polymerase typically pauses. As a result, differentially modified nucleobases can be detected by their individual kinetic signature. Combining this technique with chemical labeling of the modified nucleobases to increase their bulkiness and to facilitate their discrimination by the polymerase has allowed simultaneous mapping of 5hmC, 5fC and 5caC at almost single-base resolution in a fungal model organism[94]. Another emerging single-molecule sequencing method is nanopore sequencing. One distinct single-stranded DNA template is passed through a nanopore and nucleotide-dependent changes in an electric current within the nanopore are detected. This promising technique has since been successfully used for the discrimination of C, 5mC, 5hmC, 5fC and 5caC[95]. While large genomes (e.g. the human genome) remain challenging for SMRT-seq, nanopore sequencing allowed the detection of 5mC in the human genome with 82 % accuracy[96].

1.3.2 Protein scaffolds for the detection of epigenetic cytosine modifications

1.3.2.1 Scaffolds without sequence-specificity

Protein scaffolds that recognize a specific modification in DNA, sometimes in a constrained sequence context i.e. the CpG dinucleotide, have been widely used for the profiling and mapping of different cytosine modifications in DNA. By combining an affinity-based enrichment for a specific modified nucleobase with a high-throughput sequencing method, e.g. next-generation sequencing (NGS), genome-wide distribution maps for the modified nucleobases have been obtained that provide valuable biological insights into their specific localization and distribution patterns. One commonly used protein scaffold for enrichment is antibodies. For example, methods such as methylated DNA immunoprecipitation-sequencing (MeDIP-seq)[97] and methyl-binding protein-sequencing (MBD-seq) were developed and widely used for methylome analyses. Using a 5hmC-specific antibody, an analogue method known as hydroxymethyl-DNA immunoprecipitation-sequencing (hMeDIP-seq) was developed for the detection of 5hmC[98] and allowed the genome-wide mapping of this modified nucleobase[99]. The

antibody-enrichment strategy was further extended to the detection of 5fC and 5caC, and in combination with 5mC and 5hmC-specific antibodies allowed the genome-wide analysis of all oxidized 5mC variants[47b]. An improvement for selective antibody binding was made through the use of antibodies that recognized chemically converted nucleobases. In the case of 5hmC which was converted to cytosine 5-methylenesulfonate (CMS), anti-CMS antibodies showed enhanced selectivity as well as less density bias and reduced background noise[100]. Following this strategy, several other methods were developed that showed improved selectivity and resilience, for example hMe-Seal[101] and GLIB (glucosylation, periodate oxidation, biotinylation)[100].

Despite the valuable information on genome-wide distribution of 5mC and its oxidized variants that was obtained in numerous studies using antibodies, this method also faces some serious drawbacks consisting of (1) poor resolution of distribution maps due to the length of the DNA fragment and the size of the antibody, (2) biased amplification of enriched signals that are frequent but very weak, and (3) enrichment methods do not allow absolute quantifications.

1.3.2.2 Programmable and sequence specific DNA binding proteins

Compared to protein scaffolds with no or only constrained sequence selectivity, the advantages of protein scaffolds with programmable sequence selectivity lie in the increased resolution and the direct enrichment of the desired target DNA sequence. One of the first protein scaffolds that were used for programmable binding of DNA were zinc finger proteins (ZFPs) (Figure 1-12). These naturally occurring protein domains are among the most frequent types of DNA-binding motifs found in eukaryotes and were initially discovered in *Xenopus laevis* transcription factor IIIA (TFIIIA)[102]. An individual zinc finger domain consists of approximately 30 amino acids in a conserved ββα configuration with two invariant pairs of cysteine and histidine (Cys_2-His_2). These coordinate a zinc (II) ion tetrahedrally and thereby shape the characteristic “finger” structure.

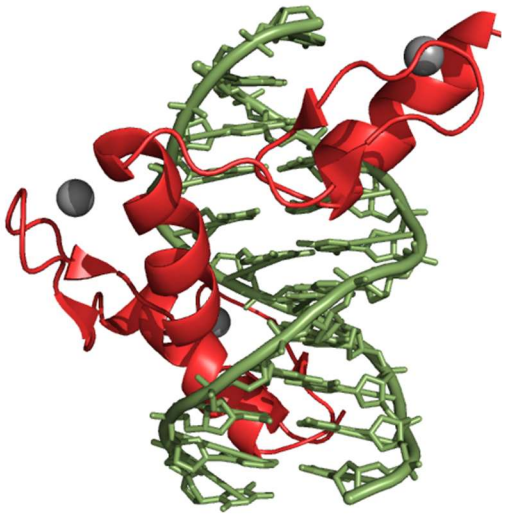


Figure 1-12 Crystal structure of a Zif628 zinc finger–DNA complex. Zif628 contains three Cys_2-His_2 zinc fingers (red) that are coordinated by zinc ions (gray spheres), and wrap around the DNA (green). (PDB entry 1AAY)[103]

Recognition of DNA occurs via amino acids on the surface of the α-helix, which make specific contact to three consecutive base pairs of DNA via hydrogen bonding in the DNA major groove. Each zinc finger domain within a multiple-finger ZFP functions as an individual module with its DNA binding specificity determined by a number of variable amino acids within the finger. Adjacent zinc finger domains were shown to be joined by flexible linkers[104]. This discovery was a key to the application of zinc finger domains as proteins for programmable DNA recognition. It allowed the development of unnatural ZFPs with more than three zinc finger domains which increased the length of their target DNA sequence to 18 base pairs[105]. Based on this proof-of-principle, several methods for the construction of ZFPs with tailored DNA-binding specificity have been developed i.e. using large combinatorial libraries or rational design[106]. With the newly obtained ability to target specific DNA sequences within a genome, zinc finger DNA-domains developed into a DNA-binding domain for effector domains such as transcription factors, integrases or nucleases resulting in chimeric proteins that were used for targeted genome engineering. Providing a tool for targeted modulation and modification of genomic loci, chimeric ZFPs such as ZFP transcription factors (ZFP-TFs) and zinc finger nucleases (ZFNs) were applied in various model organisms as well as the human genome to investigate their therapeutic potential[107]. With respect to the detection of epigenetic modifications, in combination with MBDs, ZFPs were also applied for the detection of 5mC, and showed sensitivity towards 5hmC and 5caC in DNA[108].

1.4 Transcription activator-like effector (TALE) proteins

1.4.1 Origin of TALE proteins

Transcription activator-like effectors are naturally occurring DNA-binding proteins predominantly found in the Gram-negative plant pathogen bacterial genus *Xanthomonas*. The TAL effector- or AvrBs3/PthA family represents the largest effector family found in *Xanthomonas* spp. They function as transcriptional activators of plant genes involved in infection and disease resistance[109]. Key to the pathogenicity of most of the *Xanthomonas* pathovars is a type III secretion system (T3SS) that is required for translocation of TALEs into the host cell. Spanning both bacterial membranes, the T3SS is extended by a hollow conduit, the Hrp pilus, that traverses the plant cell wall and allows to translocate the TALE protein into the plant cell cytoplasm[110]. After translocation to the nucleus, TALEs activate transcription of genes[111] involved in bacterial growth, lesion development, and cell hypertrophy[109a].

1.4.2 TALE protein structure

TALE proteins can be divided into three structural parts: (1) the N-terminal region (NTR) containing the type III secretion signal, (2) the central repeat domain (CRD) mediating binding of DNA, and (3) the C-terminal region (CTR) which contains a nuclear localization signal (NLS) and an activation domain (AD) (Figure 1-13). The CRD consists of an array of 13 to 28 repeat units of which each is made up from 33 to 35 highly conserved amino acid residues, terminated by a truncated "half-repeat"[112]. Within each repeat unit, the highly conserved amino acid sequence only differs in two hypervariable residues at positions 12 and 13, which are therefore called repeat variable di-residue (RVD) and are responsible for selective nucleobase recognition[113]. The one-to-one recognition thereby follows the TALE code where RVDs consisting of the amino acids HD (His12, Asp13), NG (Asn12, Gly13), NI (Asn12, Ile13), and NN (Asn12, Asn13) (NH; Asn12, His13) selectively recognize C, T, A, and G, respectively[9].

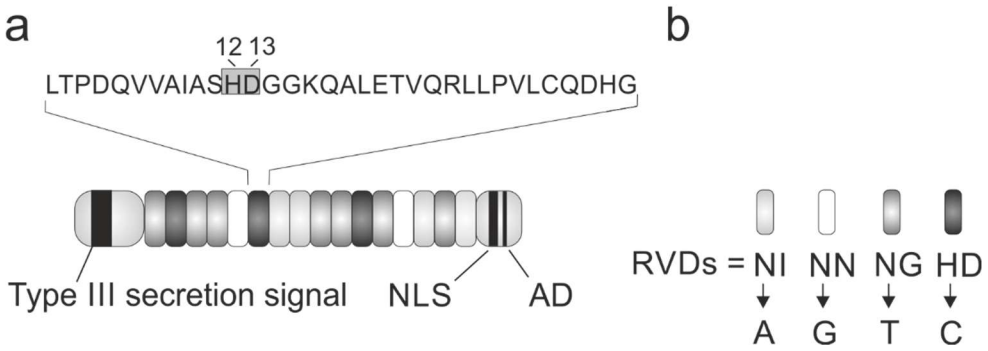


Figure 1-13 Schematic structure of a transcription activator-like effector (TALE) protein and the canonical TALE code. (a) Structure of a TALE with N-terminal domain containing the type III secretion signal, the central repeat domain (CRD), and the C-terminal domain containing the nuclear localization signal (NLS) and the activator domain (AD). (b) Canonical RVDs NI (Asn12, Ile13), NN (Asn12, Asn13), NG (Asn12, Gly13), and HD (His12, Asp13) recognizing adenine (A), guanine (G), thymine (T) and cytosine (C), respectively. Modified from[2]. Copyright 2016 American Chemical Society.

Crystal structures of natural TALE PthXo1 and the artificial TALE dHax3 revealed that in the DNA-bound state each TALE repeat unit forms a highly similar structure consisting of two α-helices designated as helix a and helix b, and a connecting loop that contains the RVD. Making specific RVD-nucleobase contacts, the TALE forms a right-handed superhelix and wraps around the DNA major groove (Figure 1-14)[113].

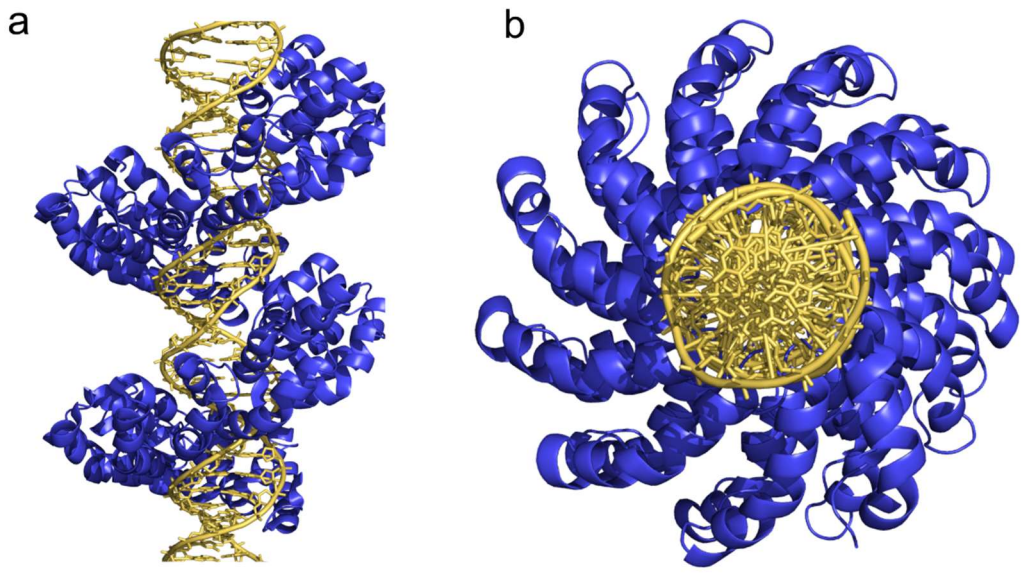


Figure 1-14 Crystal structures of transcription activator-like effector (TALE) proteins. Crystal structure of a transcription activator-like effector (TALE) protein bound to its DNA target in a side view (a), and top view (b). The TALE is shown in blue, DNA in yellow. (PDB entry 3UGM)[113a].

Besides the canonical repeat units found in the CRD, the N-terminus harbors additional cryptic repeats that are structurally very similar to those found in the CRD and contribute to nonspecific binding of the DNA. Initially thought to contain only two cryptic repeats (repeats -1, and 0)[113] the NTR was later found to harbor a total of four cryptic repeats (termed repeats -3, -2, -1, and 0)[114]. Similar to the canonical repeats, they consist of two α-helices and a connecting loop. Repeat 0 tightly packs against the first canonical repeat of the CRD and continues the superhelical structure of the CRD, whereas repeat -1 makes van der Waals contact with the methyl group of the 5´thymine base through the indole ring of a conserved tryptophan residue (W232). Since most natural TALE target sites are preceded by a thymine (referred to as T_0), the 5´ T rule was thought to be mandatory for TALEs to be functional[9b]. Several studies have since suggested that T_0 is required when designing optimal TALE DNA-binding domains[114-115]. However, alternative NTRs i.e. an NTR with the double amino acid substitution W232R/Q231S in repeat -1 accepted G_0[116] and further suggested that the necessity for T_0 is influenced by the repeat number, RVD-composition, and the presence of RVD HD at position 1 (HD1)[117]. Lastly, numerous positively charged amino acids in the NTR contribute to nonspecific binding of TALE to the DNA suggesting that the NTR serves as an indispensable “nucleation site” for efficient binding of TALEs to DNA

both *in vitro* and *in vivo*[114]. The binding specificity is promoted by the CRD via specific RVD-nucleobase interactions, which are explained in more detail in chapter 1.4.3. However, nonspecific binding is also found in the context of the CRD, which displays positively charged patches along its inner surface. In this context, residues Lys16 and Gln17 within each TALE repeat are important for nonspecific contact to the negatively charged DNA phosphate backbone[113b, 114].

1.4.3 Repeat variable di-residue (RVD) interactions

With more than 20 RVD sequences observed in naturally occurring TAL effectors, RVDs HD, NG, NI, NN, NS, HG, and N* (* meaning a deletion of position 13), are accounting for nearly 90 % of all RVDs[112]. Among these, the most specific RVDs that were since applied for the design of numerous artificial TALE proteins are HD, NG, NI and NN that specifically target C, T, A, and G, respectively[9]. Within this context, sequence-specific contacts to the DNA are exclusively made by the second RVD amino acid residue (position 13). In contrast, the first amino acid residue of the RVD (position 12) makes contact to the carbonyl oxygen of amino acid residue 8 within each repeat leading to a stabilization of the RVD-loop structure[113]. The specific RVD-nucleobase interactions are based on either hydrogen bonds or nonpolar van der Waals interactions between the side chain or the backbone of the respective amino acid at position 13 and the nucleobase. For example, the aspartate 13 (Asp13) of the RVD HD makes specific contact to cytosine through a hydrogen bond between its carboxylate oxygen and the cytosine N4 amino group (Figure 1-15). Hydrogen bond-mediated recognition occurs also in the case of RVD NN between the asparagine residue (Asn13) and the N7 of the corresponding guanine base. This interaction is also observed when adenosine replaces guanosine; an observation that might explain the more relaxed selectivity of RVD NN compared to RVD HD. In the case of the RVDs NG and NI, nucleobase contact is made through van der Waals interactions. Consequently, RVD NG binds thymine via van der Waals interactions between the backbone Cα of glycine 13 (Gly13) and the methyl group of the thymine base, while in RVD NI the aliphatic side chain of isoleucine 13 (Ile13) makes van der Waals contact to the C8 and N7 of the adenine purine ring[113].

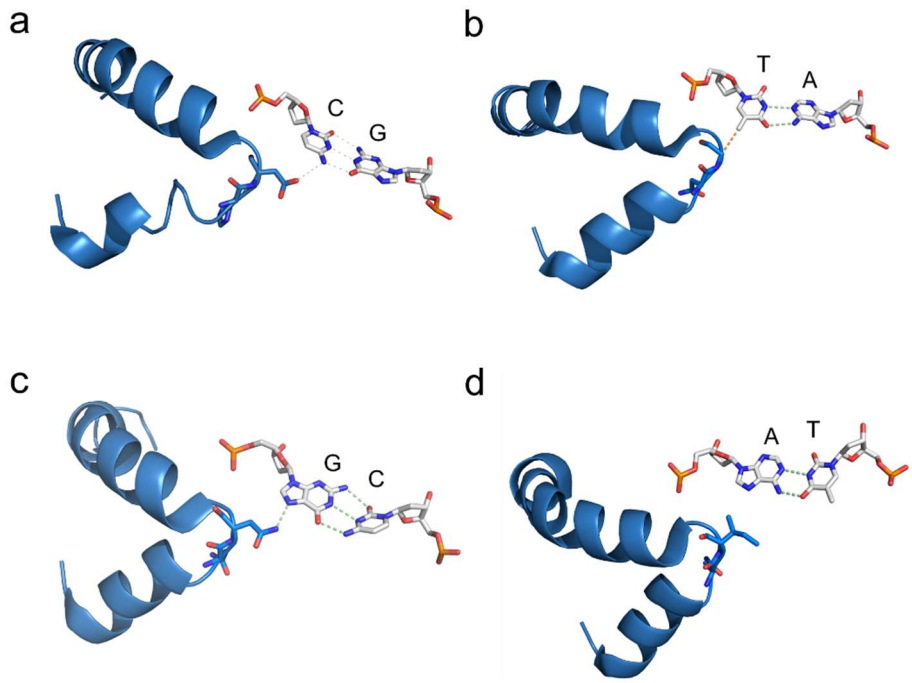


Figure 1-15 Repeat variable di-residue (RVD)-nucleobase interactions for the canonical RVDs HD, NG, NN, and NI. (a) RVD HD forms a hydrogen bond between the aspartate 13 and the cytosine N4 amino group. (b) RVD NG binds thymine through van der Waals interactions between the backbone Cα of glycine and the methyl group of thymine. (c) In RVD NN, the asparagine residue makes a hydrogen bond to the N7 of guanine. (d) RVD NI interacts with adenine via van der Waals interactions between the aliphatic side chain of isoleucine and C8/N7 of adenine. (PDB entry 3UGM)[113a]

Despite the fact that these RVDs have naturally evolved to be highly specific by having the highest nucleobase affinity, much work has been put into the design of alternative RVDs to further improve the performance of TALE-based genome editors. Several studies engaged in creating large RVD libraries via saturation mutagenesis of repeat positions 12 and 13 and investigated the selectivity profile and performance of these RVDs. This helped to get a better understanding of the general rules that govern RVD specificity and led to the discovery of several interesting alternative RVDs[118]. The specific interactions between RVD and the respective nucleobase dictated by their chemical and steric properties are a prerequisite for selective binding of a specific DNA target sequence. However, individual RVD contacts with a nucleobase are influenced by the sequence context, neighboring effects, position, and overall TALE length. The impact of these influences on TALE affinity and specificity will be discussed in the following chapter.

1.4.4 Influences on RVD specificity and TALE protein affinity

Both TALE affinity and specificity are not only determined by individual RVD/nucleobase interactions but are also influenced by a number of global effects. Specificity of an RVD cannot exclusively be explained by the canonical recognition code, but is modulated by additional parameters such as global repeat context and direct neighbor repeat effects as well as by the position of the given repeat within the CRD. Variances in specificity due to context effects were most prominent in repeat NN, which already has an intrinsic dual specificity for G and A. In some cases, the context effects led to the preferred binding of mismatched sequences rather than the predicted sequence. However, repeat NN is more specific for G if its direct N- or C-terminal neighbor repeat is NI[119]. Furthermore, RVD specificity is influenced by the position of the repeat within the repeat array. Repeats located at the N-terminal end of the CRD, which is binding the 5´ end of the DNA target sequence, were found to be less tolerant to mismatches than the same repeats located at the C-terminus. This has direct implications for overall TALE affinity, with TALE-DNA recognition underlying a certain polarity effect in which N-terminal repeats are more important for promoting TALE affinity than C-terminal repeats[115b, 119-120]. One of these studies further revealed that TALE affinity is significantly influenced by the repeat composition due to individual repeats/RVDs having varying binding strength. For example, RVD NG was argued to be the strongest RVD, followed by RVDs NN and HD with slightly less binding strength, and a comparably weak RVD NI[115b]. Thus, the positioning of several strong RVDs at the N-terminal end of the repeat array might be beneficial for TALE affinity. It is important to not exceed a certain number of repeats within a TALE, since the total length of the CRD again has an influence on TALE specificity. With increasing numbers of repeats providing extra binding energy this might eventually lead to a loss in TALE specificity[119, 121].

To conclude, studies based on natural and artificial TALE proteins highlight the existence of multiple interplaying factors that influence RVD specificity and TALE affinity. Although there is no exact rule to achieve optimal TALE specificity and affinity regarding the TALE protein structure, these studies further increased our understanding of the TALE/DNA interaction. Furthermore, they provide helpful guidelines for improving TALE design, which should be considered when designing TALE proteins for genome engineering applications (chapter 1.4.6) or the detection of epigenetic cytosine modifications (chapter 1.5).

1.4.5 Binding mode of TALE proteins.

In a cellular context, all DNA-binding proteins face the challenge of finding and selectively binding their distinct target sequence while being exposed to millions of nonspecific target sites. The search for the target sequence needs to be efficient. Once it is encountered, it requires rapid dynamics and stable binding to the target sequence. This task has been long described with the model of facilitated diffusion, in which nonspecific binding of a DNA-binding protein is followed by “sliding” and/or “hopping” along the DNA[122]. In the sliding period, the protein constantly makes contact to the DNA by following its helical structure in a rotating movement around the DNA[123], while the protein quickly dissociates and re-associates with the DNA during the hopping period. TALE proteins are structurally very different from other known DNA-binding proteins. They form a superhelix wrapped around the DNA molecule[113] (chapter 1.4.2). In an attempt to unravel the TALE-binding mechanism, recent studies (based on single-molecule techniques) proposed a two-state search mechanism, which differs from the traditional models for nonspecific search along the DNA[124]. In the search state (the first state) nonspecific contact with the DNA is initiated by the TALE NTR which has been previously reported to act as nucleation site for TALE binding[114, 125]. The NTR thereby facilitates the one-dimensional diffusion along the DNA which is a prerequisite for rapid target search. During the nonspecific search, the superhelical CRD is in a loose conformational state that allows to track the sequence of the DNA helix mediated by electrostatic interactions between specific residues of the CRD and the DNA phosphate backbone. TALEs were suggested to use a rotationally decoupled mechanism during the nonspecific search in which they do not rotate along the helical path of the DNA, but rather follow the direct trajectory allowing for faster translocation along the DNA. The state of nonspecific search is interrupted by steps of longer duration binding events where the CRD is checking local DNA sequences. Unfavorably steric and electrostatic interactions thereby prevent the TALE from associating with incorrect target sites. However, if the target sequence is encountered, the TALE transitions into the recognition state (second state). In this dynamic transition process, the CRD undergoes a conformational change into a more compressed superhelical structure. The conformational change is mediated by the requisite contacts between the CRD and the DNA backbone, leading to the correct alignment of the RVDs and other residues within the CRD and eventually resulting in stable binding[124].

1.4.6 Application of TALE proteins

Due to their simple and programmable DNA recognition mode, TALE proteins have developed into a popular and widely used DNA-binding scaffold. Allowing for a very flexible design, the single-nucleobase recognition mode of TALE repeats provides a clear advantage over other more confined DNA-binding scaffolds such as for example ZFPs, which recognize nucleobase triplets. Individual TALE repeats can be easily assembled into proteins of various lengths without the need for reengineering linkage regions. The development of several methods for TALE assembly such as Golden Gate cloning[115a], high-throughput solid-phase assembly[126], and ligation-independent cloning methods[127] has facilitated the construction of TALEs and allows for their rational design to target virtually any desired DNA sequence. Exploiting the programmability, TALE proteins have been used for the creation of sequence-specific genome engineering tools known as transcription activator-like effector nucleases (TALENs) by fusing TALEs with the dimeric nuclease *Fok*I. Using TALENs, sequence-specific double strand breaks can be created within the double-stranded DNA. Double strand breaks activate DNA double strand repair mechanisms such as non-homologous end joining (NHEJ) or homologous recombination (HR). While NHEJ is error-prone and often results in deletions or insertions, homologous recombination requires a homologous or near-homologous DNA template that is used to guide the repair of the damaged DNA site. TALENs allow to modify genomes at targeted DNA sites and since have been used in a variety of model organisms, i.e. for the investigation of gene functions, the generation of disease models and even for treating genetic disorders in humans[128]. Further applications of TALEs fused to alternative effector domains such as transcriptional activators[129]/repressors[130], or demethylases[131] have been reported, as well as their use as fluorescent probes to study chromatin dynamics[108c, 132].

An interesting characteristic of TALE proteins is their sensitivity towards epigenetic modifications in DNA[10c]. This can be detrimental when using TALEs on genomic DNA samples or directly in living organisms. However, it also represents a powerful tool for *in vivo* studies of epigenetic modification dynamics or modification-dependent control over specific genomic loci if combined with TALE repeats that are sensitive towards epigenetic nucleobases. Pursuing this strategy, several TALE repeats have been developed, which display altered selectivities for epigenetic nucleobases. Recent innovations and achievements in this field are discussed in more detail in the following chapter (chapter 1.5).

1.5 RVD design for selective detection of epigenetic cytosine modifications

Since the discovery that TALE proteins can recognize methylated DNA, a large number of TALE repeats have been investigated for their potential to bind epigenetic cytosine modifications. One of the first examples was RVD HD, which was shown to be sensitive towards a C5-methyl group installed on its target nucleobase cytosine. Whereas methylation of cytosine abolished binding of RVD HD, 5mC is readily bound by RVD NG[10c]. This was attributed to the similarities between its canonical target base thymine and 5mC, as both contain a 5´methyl group that can be recognized by RVD NG through a hydrophobic interaction with the Cα methylene moiety of RVD residue 13 (Gly13)[113]. In the case of RVD HD, the presence of a methyl group at this position presumably induces a steric clash with Asp13 and does not allow the formation of a hydrogen bond between the aspartate and the cytosine 4-amino group (Figure 1-15). This sensitivity was even observed for single 5mC positions within a target DNA sequence and showed only little dependence on sequence context and position within the DNA target sequence[10b, 133]. This was extended to the oxidized 5mC variant 5hmC, with results revealing RVD NG to bind 5mC, but not C or 5hmC, and RVD N* binding C and 5mC, but not 5mC[10d, 134]. This was extended further by applying these RVDs in affinity enrichment experiments where 5mC and 5mC could be detected at single nucleotide positions within a genomic target sequence. An advantage of TALEs over other conventional approaches such as antibody strategies is their strand-selectivity and their higher sensitivity[10a, 135].

Inspired by the TALEs sensitivity for 5mC and 5hmC, RVDs with selectivities for each of the modified cytosine variants should be designable if the unique chemical properties and the size of their 5´ substituents are considered. Since the individual cytosine variants differ quite drastically from each other in their chemical properties, steric demand and flexibility (chapter 1.2.1.5) this could be the key to creating RVDs with individual selectivities (Figure 1-16).

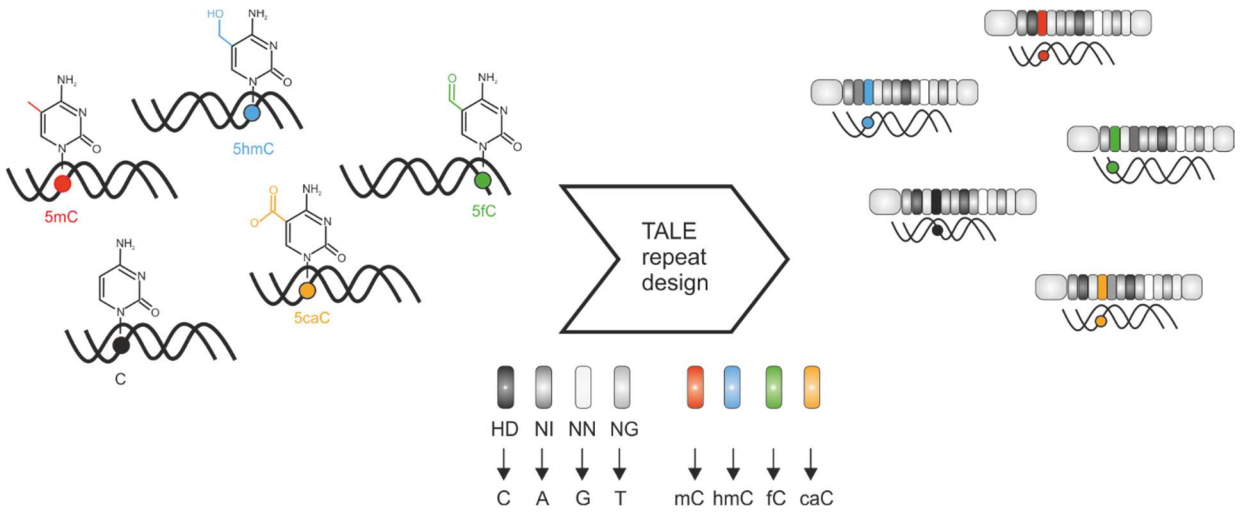


Figure 1-16 Extending the transcription activator-like effector (TALE) repeat repertoire. Design of artificial mutant TALE repeats for the selective and direct detection of the cytosine modifications 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) in DNA.

This inspired several strategies involving directed mutation or randomization of RVD positions 12 and 13 and/or their neighboring positions as well as deletion of key residues within the RVD region and led to several interesting new repeats with specific selectivities and applications thereof. In one of these studies for example, randomization of positions 11 to 14 within the repeat led to the discovery of an alternative artificial repeat that binds 5mC with higher affinity than C[136]. Another approach combined saturation mutagenesis of RVD positions 12 with either saturation mutagenesis or deletion of position 13, respectively. This afforded a large library of repeat mutants that were screened for their activation potential in a reporter gene assay. As a result, two repeats with enhanced specificity for 5mC, namely RVDs HA (His12, Ala13) and NA (Asn12, Ala13), were discovered as well as the 5mC-binding RVD FS (Phe12, Ser13)[137].

Increasing the rather narrow space between the RVD loop and the DNA backbone by reducing the size of the repeat should be a strategy to allow for completely new interaction possibilities between TALE repeat and nucleobase. In the course of the present work, this idea was thoroughly investigated for its potential to yield selective TALE repeats for each of the individual cytosine variants 5mC, 5mC, 5fC and 5caC. The results thereof are presented in chapter 3.

2 Aim of this work

The aim of this work was the design and engineering of novel TALE repeats with individual selectivities for the oxidized 5mC derivatives 5hmC, 5fC and 5caC. The discovery of the oxidized 5-methylcytosines as intermediates in the cycle of active demethylation, has posed the question of their individual biological functions and their relevance as independent epigenetic marks next to 5mC. In order to address these questions, methods for the direct and selective detection of the individual oxidized 5-methylcytosines are needed.

In this study, TALE proteins were employed as DNA-binding scaffolds due to their inherent advantage of high programmability and their reported sensitivity towards epigenetic cytosine modifications[10]. Naturally occurring TALE repeats are structurally very conserved, and the space between the RVD loop and the DNA backbone is limited[113]. Due to the increased steric demand and unique chemical properties of their 5-substituents, the oxidized 5mC derivatives might require alternative repeat structures that enable completely new binding modes. Reducing the size of the TALE repeat and creating an enlarged space between repeat loop and nucleobase should provide the basis for new repeat-nucleobase interactions. Hence, rational design and engineering of size-reduced and truncated TALE repeat scaffolds was used in this study to extend the TALE code towards the recognition of oxidized 5mC derivatives.

High-throughput screening of large libraries of size-reduced TALE repeats with different patterns of amino acid deletions or substitutions of key residues within the repeat loop aimed at a better understanding of how TALE repeat-nucleobase interactions are modulated. Following this strategy, thorough investigation of the engineering potential and adaptability of TALE repeats should yield new TALE repeats with individual selectivities for the oxidized 5mC variants 5hmC, 5fC and 5caC. With this at hand, selective and sequence-specific detection of these nucleobases in DNA would be greatly facilitated. Importantly, this TALE-based detection method largely benefits from the flexibility and ease of application provided by the inherent programmability of TALE proteins. Being able to localize and selectively detect the oxidized 5mC variants in a programmable manner in an *in vivo* environment represents a crucial step towards the understanding of the biological roles and epigenetic relevance of 5hmC, 5fC and 5caC.

3 Results and Discussion

3.1 Introduction

Crystal structures of TALE proteins bound to DNA revealed a very narrow space between the RVD loop and the 5-positions of cytosines, with 3.4 – 3.7 Å between the Cα and the amide C=O bond of amino acid residue 13 and the 5-methyl group of thymine or 5mC (Figure 3-1). Since the oxidized 5mC variants exhibit larger 5-subsituents with increased steric demand and unique chemical properties, the space limitations imposed by the natural RVD loop most likely abolish binding of these cytosine variants. This was for example seen in the case of RVD HD, where the 5-methyl group of 5mC abolished RVD-nucleobase recognition due to steric clashing. Furthermore, the RVD loop is restricted in its flexibility due the formation of an intra-repeat contact of loop residue 12 with loop residue 8 (alanine, Ala8). This leads to stabilization and preorganization of the RVD loop structure.

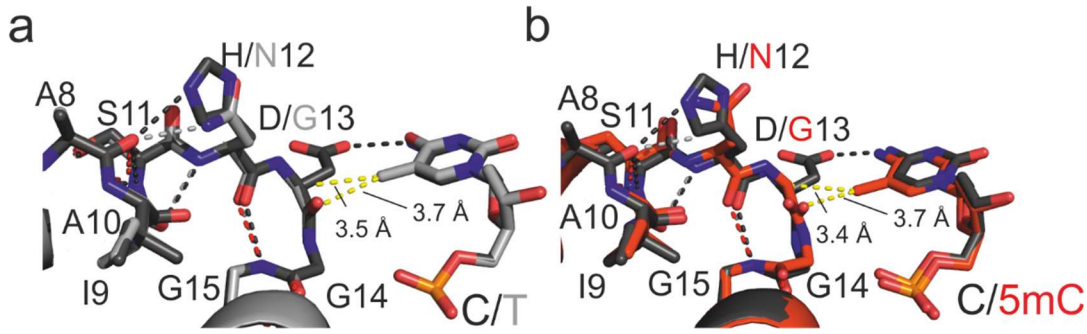


Figure 3-1 Superimposed crystal structures of repeat variable di-residue (RVD) HD and NG with cytosine and thymine. (a) Crystal structure overlay of RVD HD (black) binding to C (black) and RVD NG (gray) binding to T (gray) (both PBD entry 3V6T)[113b] (b) Superimposed crystal structures of RVD HD (black) binding to C (black) and RVD NG (red) binding to 5mC (red) (PDB 4GJP)[10c]. Modified from[2]. Copyright 2016 American Chemical Society.

Generally, RVD loop and repeat organization are highly conserved in naturally occurring TALE proteins. TALE repeats from *Xanthomonas* spp. showed an overall strong conservation in both sequence[138] and structure[113], which also included the conformation of their amino acid side chains (Figure.6-1). Likewise, truncations of TALE repeats are very rarely observed in nature. Of 1,396 naturally occurring unique TALE repeat sequences identified from amino acid sequences with annotated or predicted TALE repeat domains deposited in NCBI and UniProt, only five had deletions within their loop region. In the case of *Xanthomonas* spp. TALE repeats, the loop region of none of the TALE repeats had more than one deletion[138].

One of these naturally occurring truncated repeats with a deletion in the RVD loop is N* (Asn12, * = deletion at position 13). Due to the missing amino acid, the RVD loop of N* does not extend as deep into the DNA major groove as it is the case for other TALE repeats[113a]. This allowed N* to accommodate the sterically more demanding nucleobase 5mC with the methyl group protruding into the DNA major groove[134]. However, N* is not able to recognize the oxidized 5mC derivative 5hmC which bears a flexible hydroxyl moiety at the same position[10d].

These overall findings encouraged the investigation of the TALE repeat engineering potential in order to create scaffolds for new TALE-nucleobase interactions and selectivities for the oxidized 5mC derivatives with their larger side chains. The critical aspect seems to be the RVD loop, which due to its size does not leave the space for recognition of nucleobases with larger 5-substituents. Consequently, reducing the size of the repeat loop could allow the selective recognition of the oxidized 5mC derivatives 5hmC, 5fC, and 5caC.

3.2 Engineering of size-reduced TALE repeats: Libraries X* and X**

To investigate the binding potential of size-reduced TALE repeats, saturation mutagenesis of repeat position 12 was combined with one or two deletions of adjacent repeat positions 13 and 14 (Figure 3-2 b). This resulted in TALE repeats that are shortened by one or two amino acids and potentially results in more space for the accommodation of large cytosine 5-substituents. Even if the amino acid at repeat position 12 usually does not participate in specific nucleobase interaction, the altered overall loop geometry could influence behavior of individual amino acid residues within the loop and thus enable new interactions. The respective mutant TALE repeats were constructed via NNK codon mutagenesis and Quickchange reactions on TALE repeat plasmids (see chapter 5.2.12) that were subsequently assembled into vector plasmid pGFP-ENTRY via Golden Gate assembly[115a]. Hence, TALEs with N-terminal GFP domain, a shortened *AvrBs3*-type TALE N-terminus with +136 aa starting from the canonical repeat 1, and a C-terminal His6 tag (Figure 3-2 a) were created. The GFP-TALE fusion proteins could be expressed at high levels and were purified making use of the His6-tag.

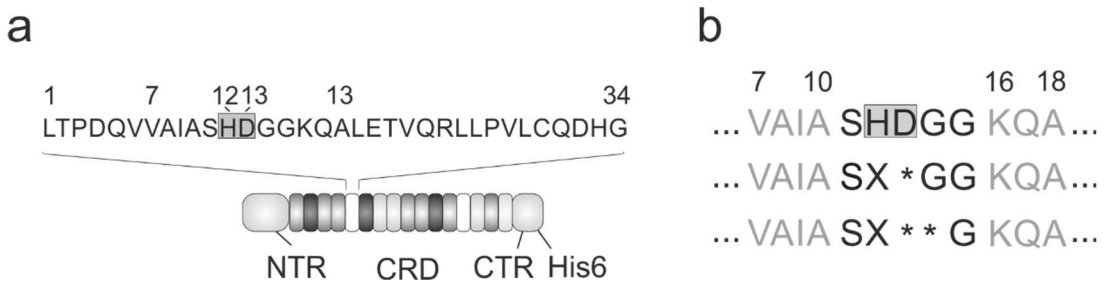


Figure 3-2 Transcription activator-like effector (TALE) protein scaffold and repeat library design for libraries X* and X.** (a) Mutant TALE proteins contained a N-terminal GFP-domain, a shortened *AvrBs3*-type TALE N-terminus (+136 amino acids starting from the canonical repeat 1), the respective mutant TALE repeat at repeat position five within the central repeat domain (CRD), and a C-terminal His6 tag for protein purification. (b) Library design for mutant TALE libraries X* and X** (X = randomization of repeat position 12 to each of the 20 canonical amino acids, * = deletion of repeat position 13, ** = deletion of repeat positions 13 and 14).Modified from[1-2]. Copyright 2018 Royal Society of Chemistry and 2016 American Chemical Society.

Two libraries with 20 mutant TALE proteins each were created. Library X* (X = randomization of repeat residue 12 to each of the 20 canonical amino acids, * = deletion of repeat residue 13) contained TALE repeats with a single mutation of repeat residue 12 and a deletion of repeat residue 13. Library X** contains repeats with a randomization of repeat position 12 and a double deletion of residues 13 and 14. All TALEs were assembled to target a 18-nucleotide sequence of *Danio rerio* Hey2_b gene, with each TALE protein containing one mutant repeat opposite C at position six (C6) in the DNA target sequence. For an *in vitro* screening, synthetic oligonucleotides containing C or one of the cytosine variants at position C6 in the Hey2_b target were employed in a DNA polymerase accessibility assay[10b] to study the affinity of the mutant TALE repeats towards the different cytosines (Figure 3-3 a and b). In this assay, the TALE competes with *E. coli* DNA polymerase (Klenow fragment 5´ – 3´ exo$^-$, KF exo$^-$) for the accessibility of the synthetic DNA-primer complex containing the TALE target sequence. TALE binding inversely correlates with primer extension by DNA polymerase and can be analyzed via polyacrylamide gel electrophoresis (PAGE). TALE binding prevents DNA polymerase from accessing the DNA, which results in a small nonextended primer-template complex. If TALE binding if abolished, DNA polymerase produces the large extension product (Figure 3-3 c).

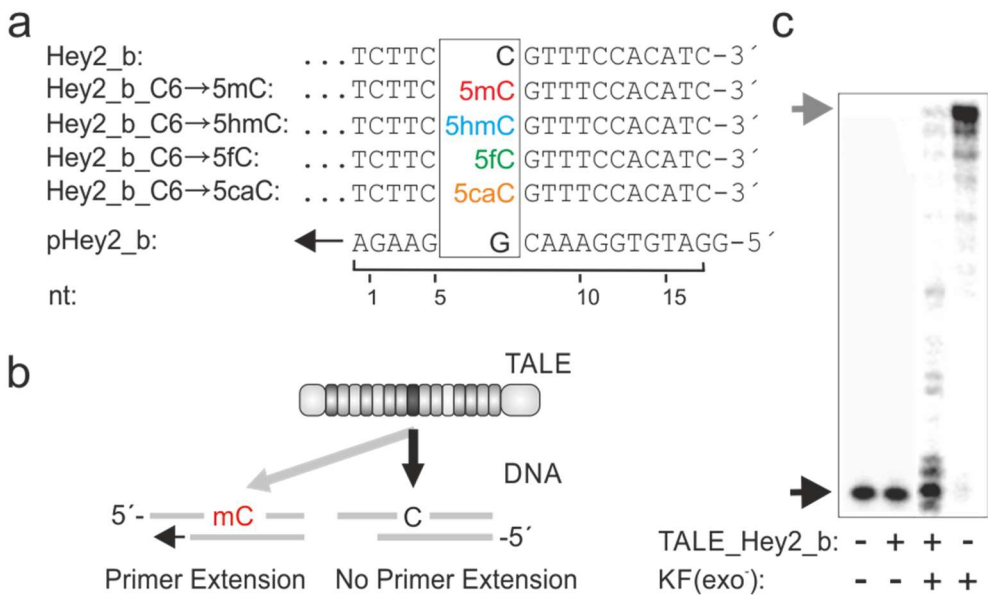


Figure 3-3 Polymerase accessibility assay. (a) Oligonucleotides with the Hey2_b target sequence used in the DNA polymerase accessibility assay. C at position six of the target sequence was substituted with 5mC, 5hmC, 5fC and 5caC. (b) Principle of DNA polymerase accessibility assay shown for C and 5mC. The target sites of TALE and DNA polymerase overlap in the DNA-primer complex. The character of the modified cytosine at position C6 determines TALE affinity. Both scenarios, TALE binding and non-binding to DNA is shown with a black or gray arrow, respectively. (c) DNA polymerase accessibility assay as shown in (b) using 8.325 nM primer-template complex, 416 nM TALE and 25 U of KF exo$^-$. Primer extension product and nonextended primer are marked with a gray and black arrow, respectively. Modified from [2]. Copyright 2016 American Chemical Society.

Using the here described DNA polymerase accessibility assay, a total number of 400 TALE repeat-nucleobase interactions were studied and binding profiles for each mutant TALE repeat towards the five different cytosine nucleobases were obtained.

3.2.1 Binding profiles TALE repeat mutants from library X**

TALEs repeat mutants with two deletions of the adjacent repeat positions 13 and 14 showed a very prominent loss in overall affinity (Figure 3-4). This effect was strongest for library members N**(Asn12, deletions of residues 13 and 14) and H** (His12, deletions of residue 13 and 14). Also, no selectivity towards one of the cytosines was seen for either of the two mutant TALE repeats. For the other double-deletion mutant TALE repeats, primer extension levels ranged from roughly 40 % to 100 % (with 100 % primer extension meaning no TALE binding) for the individual cytosine variants. Among all mutant TALE repeats from this library, 5mC was generally bound strongest of all cytosine variants. Affinity for the

oxidized derivative 5fC was second highest for most of the repeat mutants (K**, A**, C**, G**, F**, Y**, and P**). In contrast, C, 5hmC, and 5caC were generally only weakly or not bound at all. Interestingly, the expectedly smallest of all the repeat mutants, G**, showed the highest overall affinity with the observed binding preference for 5mC and 5fC (Figure 3-4).

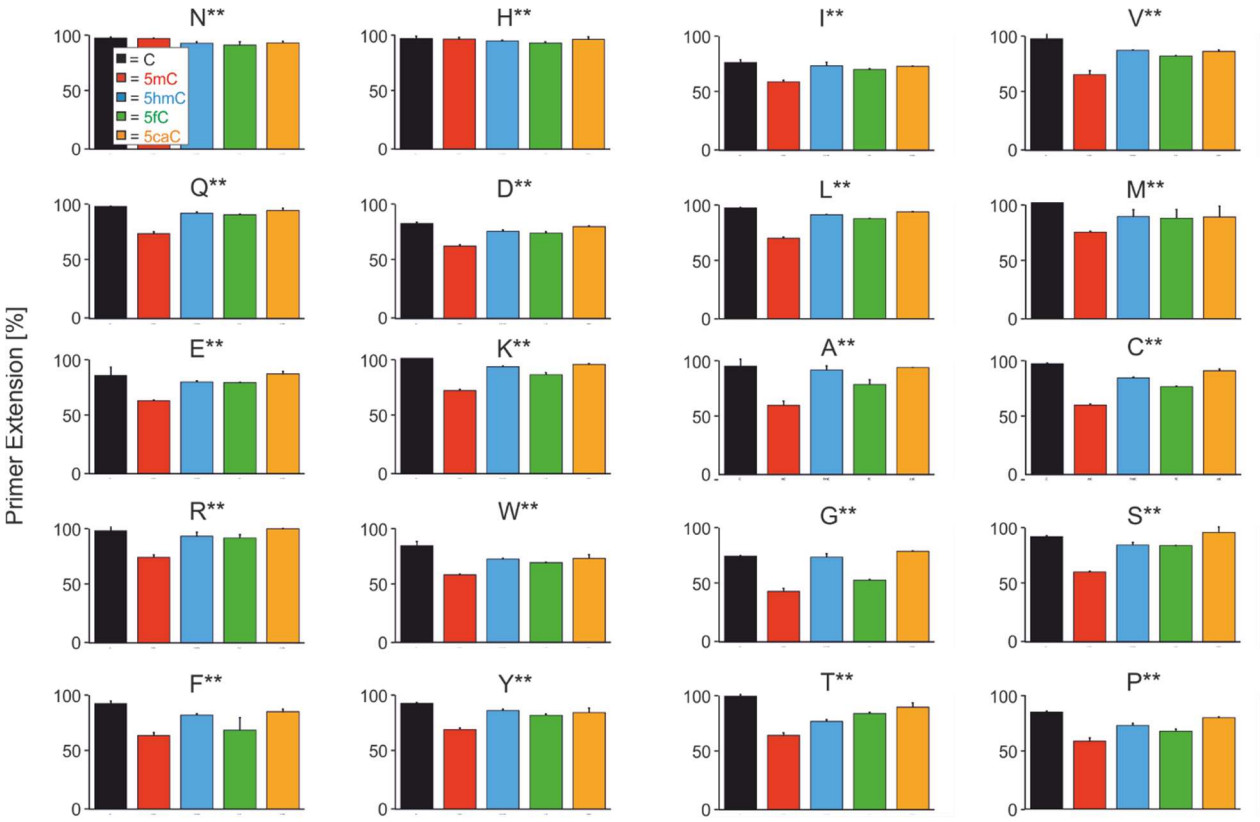


Figure 3-4 Binding profiles for the double-deletion transcription activator-like effector (TALE) repeat mutants of library X.** X = randomization of repeat residue 12, * = deletions of position 13 and 14**.** Results were obtained from a DNA polymerase accessibility assay. The assay was performed as described in chapter 5.2.13 with reactions carried out in duplicates using 83.3 nM TALE protein. The graph shows levels of primer extension in the presence of the respective TALE repeat mutant and the respective cytosine variant as a measurement of TALE affinity for the individual cytosines. Primer extension product formation in a reaction performed w/o TALE was set as 100 %. Modified from[2]. Copyright 2016 American Chemical Society.

Intriguingly, the deletion of two amino acid residues in the RVD loop did not result in the anticipated selectivities for the larger oxidized cytosine variants, but rather resulted in an overall reduced affinity. Also, substitution of the amino acid residue 12 did not remarkably influence repeat specificity. A general binding preference for 5mC was seen for amino acids with polar side chains (e.g. S** and T**), charged side chains (R**, K**) as well as hydrophobic side chains (A**, L**) at repeat position 12. Therefore, it

seems likely that the variation at this first RVD position has a moderate effect on RVD specificity, whereas the double deletion of position 13 and 14 has a dominating effect on repeat affinity. In natural occurring TALE repeats, amino acid residue at position 14 is a conserved Glycine residue, that does not interact with the nucleobase, but rather faces towards the 5´phosphate. However, deletion of this residue in combination with deletion of the adjacent amino acid at position 13 apparently leads to an unfavorable repeat structure that reduces affinity of the TALE mutants to DNA.

3.2.2 Binding profiles TALE repeat mutants from library X*

In contrast to the TALE repeat mutants from the double-deletion library X**, repeats with randomization of repeat position 12 and a single deletion of amino acid residue at position 13 (library X*) showed a very diverse range of binding profiles, with an overall higher affinity towards the different cytosine nucleobases (Figure 3-5).

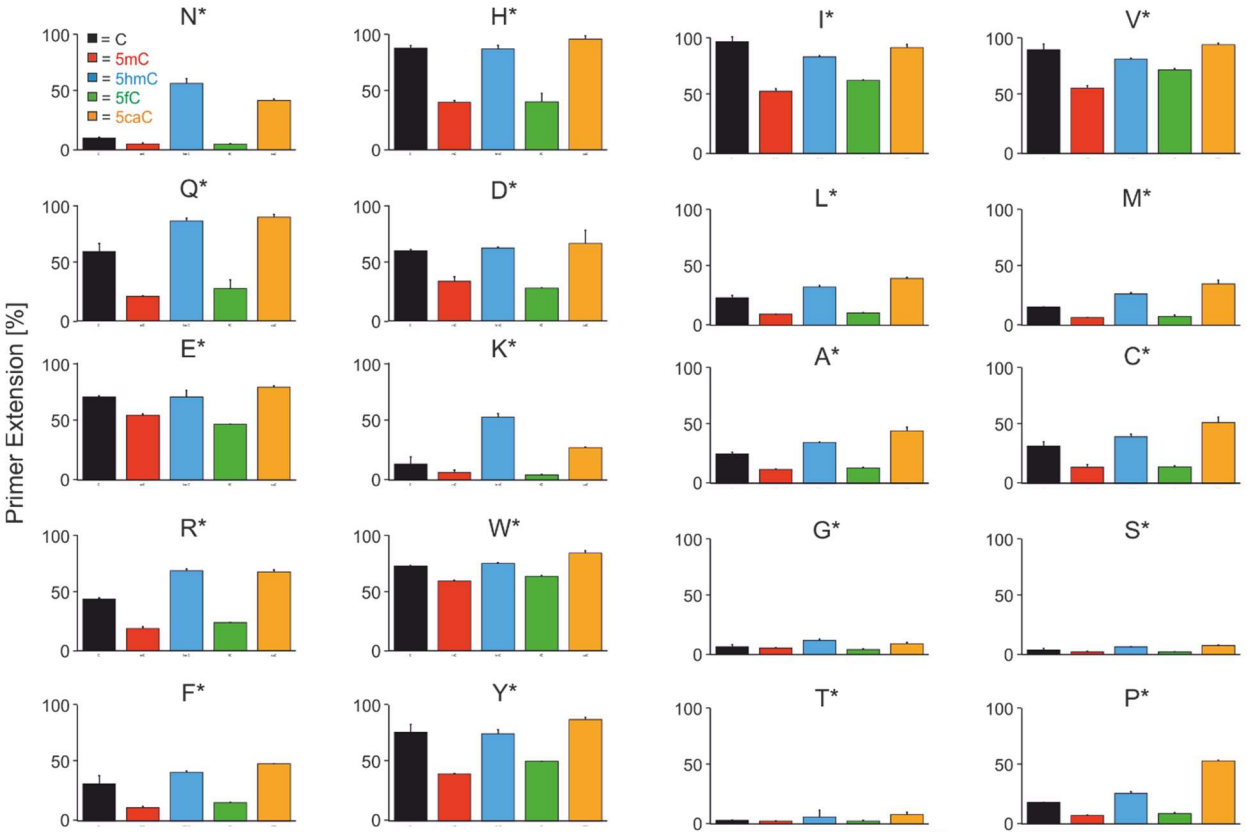


Figure 3-5 Binding profiles for the transcription activator-like effector (TALE) repeat mutants from library X*. X = saturation mutagenesis of position 12, * = deletion of amino acid position 13. Results were obtained from DNA polymerase accessibility performed as described in 5.2.13 with reactions carried out in duplicates using 83.3 nM TALE protein. The graph shows levels of primer extension in the presence of the respective TALE repeat mutant and the respective cytosine variant as a measurement of TALE affinity for the individual cytosines. Primer extension product formation in a reaction performed w/o TALE was set as 100 %. Modified from[2]. Copyright 2016 American Chemical Society.

The repeat mutants of this library are related to the previously studied, naturally occurring TALE repeat N* which was shown to bind C, 5mC and 5fC, but not 5hmC and 5caC. Similar to this binding profile and the one seen for G**, TALE repeat mutants of the single deletion library X* generally showed a binding preference for 5mC, with 5fC bound second most strong. This hints at a similarity in repeat-nucleobase

interaction for these two cytosine variants. Considering their size, 5fC can be better compared to 5hmC than to 5mC. However, crystal structures of DNA containing the modified cytosines revealed that the size of the 5-substituent is not the only determinant for their steric demand[72]. Contrasting to 5hmC, which was observed to exist in two conformations due to flexibility of its hydroxyl group, the formyl group of 5fC is fixed in a plane with the nucleobase through a hydrogen bond with the N4-amino group of cytosine (Figure 1-10). The same is true for 5caC, whose carboxyl-group is also fixed in one plane with the nucleobase. Therefore, similarities between 5mC and 5fC in repeat-nucleobase interaction most likely results from similarity in providing small, nonpolar methyl or methylene groups in the direction of the RVD loop, whereas 5hmC and 5caC have polar characters and display an oxygen towards the backbone.

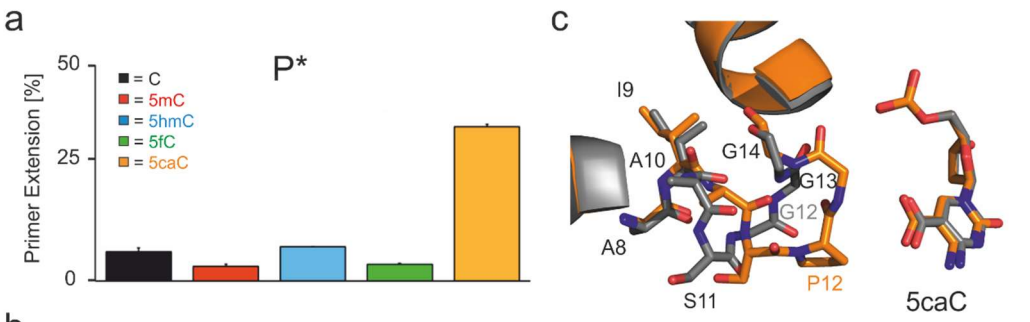
Besides the general binding preference for 5mC and 5fC observed in the single-deletion mutant library, several groups of mutant TALE repeats with distinct binding behaviors were identified. Substitution of asparagine with the alternative polar amino acid glutamine resulting in repeat mutant Q* showed same binding preferences, but also a loss of affinity for all five cytosine nucleobases (Figure 3-5). Among the amino acids with charged side chains, the anionic amino acids aspartic acid (D*) and glutamic acid (E*) showed reduced affinity compared to the cationic amino acids arginine (R*) and lysine (K*). Especially interesting in this group is repeat K*, which differed in its binding profile in favoring 5caC over 5hmC. The group of amino acids with aromatic side chains showed a comparably low binding affinity towards the individual cytosine variants, with the exception of phenylalanine (F*). Here, the size of the side chain might play a role, especially in the case of the very low affinity observed for tryptophan (W*). Homology modeling studies of this repeat in the presence of 5mC-containing DNA revealed that tryptophan at position 12 is forced between the RVD loop and the nucleobase which results in an unlikely downward movement of the backbone. In the absence of DNA, tryptophan occupies the space of the nucleobase (Figure 6-2). Therefore, the steric clash between the large side chain of tryptophan leading to an unusual repeat conformation when interacting with DNA likely explains the low affinity of the mutant repeat W*. Another interesting mutant repeat group is the one of amino acids with nonpolar side chains, including valine (V*), isoleucine (I*), alanine (A*), leucine (L*), methionine (M*), and cysteine (C*). Whereas the latter four repeat mutants show high affinity to all five cytosine nucleobases and the commonly observed preference for 5mC and 5fC, V* and I* show comparatively low affinity. Structurally, this could be explained by the amino acid *β*-branch, which is common to valine and isoleucine, but not to the other four amino acid residues (A*, L*, M* and C*). Indeed, homology models revealed that repeats A*, L*, M*, and C* are placed in a similar orientation when interacting with DNA that contains 5mC (Figure 6-3).

In contrast, in the case of V* interacting with 5mC-containing DNA, a different orientation of the repeat amino acid is observed in which the valine side chain is positioned very closely to the 5-methyl group of 5mC, which could cause steric clash and therefore prevent binding (Figure 6-3).

Furthermore, a very interesting binding behavior was observed for the rather inhomogeneous group of glycine (G*), the smallest repeat investigated in this study, and the two polar repeats containing serine (S*) and threonine (T*), which both bear a hydroxyl group. Common to these three repeat mutants was a very high affinity to all five cytosine nucleobases, leading to their designation as universal binders for cytosine and its modified derivatives. Homology models of these three repeat mutants suggested that their repeat helix is shortened, and the RVD-containing loop is reduced in size (Figure 6-4). In the case of S* and T* the helix is additionally stabilized by a hydrogen bond to the carbonyl group of repeat residue 8 (Ala8). This unique repeat structure even allowed the accommodation of the large, charged cytosine nucleobases 5hmC and 5caC. In the case of G*, the small repeat size as well as the missing side chains allowed the binding of all five cytosines with high affinity.

3.2.3 Mutant TALE repeat P* as a selective sensor of 5-carboxylcytosine

A very particular binding profile was obtained for TALE repeat mutant P*. The repeat containing a proline at postion 12 of the repeat bound C, 5mC, 5hmC and 5fC with variable affinity, but showed an especially low affinity for 5caC (Figure 3-5 and Figure 3-6).



K_i [nM]	C	mC	hmC	fC	caC	T	G	A
TALE repeat P*	172.8 ± 21.7	60.2 ± 7.4	202.9 ± 24.6	62 ± 5.7	339.1 ± 10.9	276.4 ± 89.5	> 1000	>1000

Figure 3-6 Selectivity of transcription activator-like effector (TALE) repeat mutant P* to 5-carboxylcytosine (5caC). (a) Binding profile of TALE mutant repeat P* obtained from DNA polymerase accessibility assay. Duplicate reactions were performed using 8.325 nM primer-template complex, 416 nM TALE and 25 U KF exo$^-$. (b) K_i constants for P* binding to the five cytosine nucleobases and the other three canonical nucleobases (T, G, A) were obtained from DNA polymerase accessibility assay as in (a) but with different TALE concentration. (c) Homology model of TALE mutant repeat P* interacting with 5caC, superimposed on TALE mutant repeat G* interacting with 5caC. Modified from[2]. Copyright 2016 American Chemical Society.

This selectivity was even more pronounced at higher TALE concentrations (Figure 3-6a) and was shown to be consistent in the context of all eight nucleobases (including T, G and A). This led to the designation of P* as a selective sensor of 5caC, which had not been reported previously. K_i constants for TALE mutant repeat P* were obtained by DNA polymerase accessibility assay (Figure 3-6 and Figure 6-5). The K_i for P* interacting with DNA containing 5caC was measured as 339.1 nM. Comparably lower K_i values were obtained for the other cytosine variants (Figure 3-6 b). To get a better understanding of the unusual binding selectivity of P*, homology models of this mutant repeat interacting with DNA containing 5caC were analyzed. These models suggested that the pyrrolidine ring of proline is facing the 5-carboxyl group, preventing a backbone conformation as seen for the universal binder G*, S* and T* (Figure 3-6 c). The conformational restrictions imposed by the pyrrolidine ring thus likely prevent accommodation of the 5-carboxyl group of 5caC, but still allow binding of cytosine and the other cytosine variants 5mC, 5hmC, and 5caC.

3.3 Truncated TALE repeat scaffolds for the selective recognition of oxidized 5-methylcytosine variants

The previous library designs with a combination of randomization of repeat position 12 and one or two deletions of adjacent repeat positions 13 and 14 (libraries X* and X**, respectively) afforded valuable insights into the engineering potential of TALE repeats and resulted in a number of mutant TALE repeats with interesting binding profiles. Concluding from these studies, the introduction of two deletions at position 13 and 14 leads to an overall reduced affinity of the mutant TALE repeats. However, to enable completely new interaction modes between a TALE repeat and the oxidized 5mC variants, more drastic structural changes of the TALE repeat are required. Besides the intra-repeat interactions between the side chains of amino acid residue 12 with repeat residue Ala8, TALE repeats also engage in a number of inter-repeat interactions with their neighbor repeats that are critical for TALE structure and the ability to bind DNA (Figure 3-7).

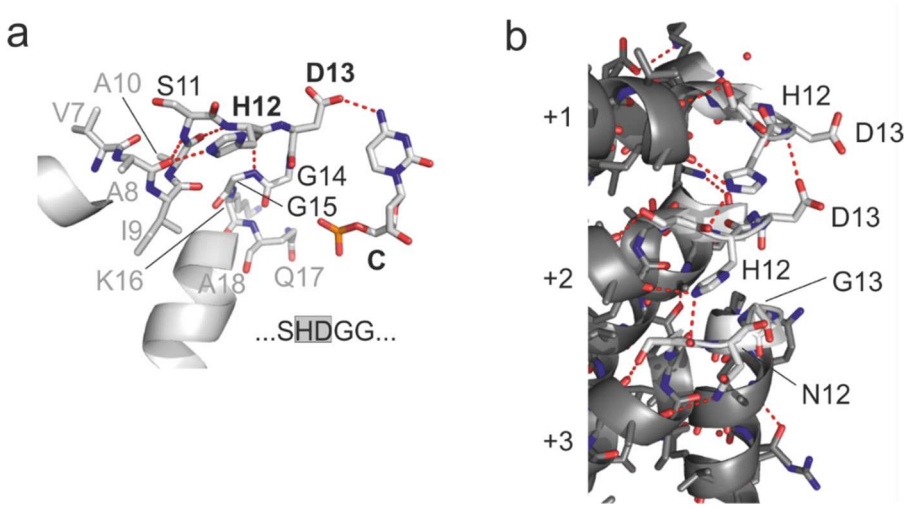


Figure 3-7 Intra- and inter-repeat interactions of transcription activator-like effector (TALE) repeats. (a) Crystal structure of TALE repeat HD (His12, Asp13) interacting with cytosine[113]. Hydrogen bonds involved in both repeat variable di-residue (RVD)-nucleobase and intra-repeat interactions are shown as red dashed lines. (b) Crystal structure of DNA-unbound TALE (PDB entry 3V6P)[113b] showing three consecutive exemplary TALE repeats and their inter-repeat interactions. Hydrogen bonds are shown as red dashed lines. Modified from[1]. Copyright Royal Society of Chemistry 2018.

In order to reconcile the idea of a truncated TALE repeat scaffold, which allows maximal freedom for creating new nucleobase interactions and at the same time maintains essential inter-repeat interactions, careful engineering approaches are required.

3.3.1 Library design and screening of truncated TALE repeats

To investigate further truncated TALE repeat scaffolds for their selective recognition of the five cytosine nucleobases, six mutant repeat library designs with one to four deletions of repeat positions 12 to 15 in combination with randomization of repeat positions 11 or 12 were created (Figure 3-8 b).

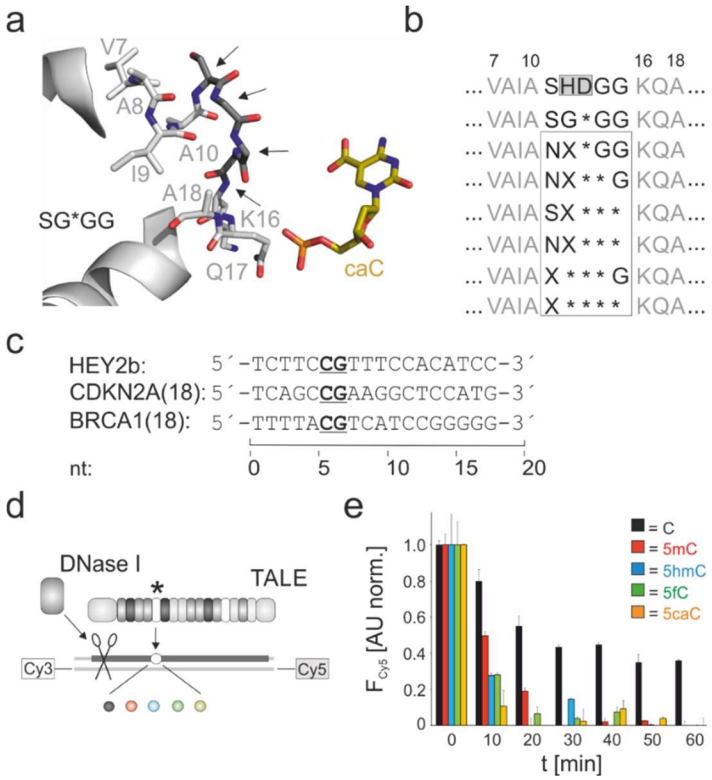


Figure 3-8 Library design and DNase I screening assay. (a) Model of mutant transcription activator-like effector (TALE) repeat G* interacting with 5-carboxylcytosine (5caC). Black arrows indicate positions chosen for randomization or deletion. (b) Design of mutant TALE libraries. Positions 11 or 12 (X) are randomized to amino acids tryptophan (W), tyrosine (Y), histidine (H), aspartic acid (D), glutamic acid (E), glutamine (Q), asparagine (N), serine (S), threonine (T), arginine (R), or lysine (K), and deletion of positions 12 to 15 (* = deletion). (c) Target sequences Hey2_b, CDKN2A(18) and BRCA1(18) used for the DNase I assay. CpG that is targeted by the mutant TALE repeats in bold and underlined. (d) Schematic representation of DNase I screening assay using Cy3/Cy5-double-labeled DNA-oligonucleotides. (e) Representative Cy5 fluorescence time course of DNase I experiment using 0.5 µM wt TALE containing RVD HD, 0.1 µM DNA and 1 U DNase I. Cy5 fluorescence was background-corrected by subtracting Cy5 fluorescence from a sample without TALE and normalized to a control without DNase I and then to Cy5 fluorescence of C at timepoint t = 0. Modified from[1]. Copyright Royal Society of Chemistry 2018.

In the libraries with one to three deletions of positions 13 to 15, position 11 was kept constant as either serine (S) or asparagine (N), resulting in libraries NX*GG, NX**G, SX***, and NX***. Position 12 was randomized to amino acids that are able to undergo polar interactions, namely tryptophan (W), tyrosine (Y), histidine (H), aspartic acid (D), glutamic acid (E), glutamine (Q), asparagine (N), serine (S), threonine (T), arginine (R) or lysine (K). In a fifth library (X***G) containing deletions of positions 12 to 14, position 11 was randomized to one of the 11 different amino acids. Furthermore, library X**** combined randomization of position 11 with four deletions (positions 12 to 15). The individual mutant TALE repeat plasmids were each created by site-directed cassette mutagenesis (chapter 5.2.12) using synthetic oligonucleotides, which contained the respective modifications. The mutant TALE repeat plasmids were each assembled into pGFP_ENTRY plasmids via Golden Gate assembly[115a] resulting in individual mutant TALEs targeting a 18-nucleotide long sequence of the *D. rerio* Hey2_b gene. TALE proteins with an N-terminal GFP domain, a shortened *AvrBs3*-type TALE N-terminus, and a C-terminal His6 tag were obtained (as in chapter 3.2). Within the 18 repeat long TALEs, individual mutant TALE repeats were located to target C6 of the DNA target sequence. For the newly developed FRET-based DNase I competition assay, the 18-nucleotide target sequence was provided as a synthetic oligonucleotide template with C6 modified to C, 5mC, 5hmC, 5fC, or 5caC. The individual target sequences were hybridized to a Cy3/Cy5-double-labeled DNA primer. In this assay, TALE binding prevents cleavage of the target DNA by DNase I hence maintaining spatial proximity between the Cy3/Cy5 FRET pair. If TALE protein binding is abolished (e.g. due to a cytosine modification), DNase I cleavage of the DNA target will lead to spatial separation of Cy3 and Cy5 and therefore result in a decrease in FRET acceptor (Cy5) fluorescence, which was then used as a read-out. Suitability of the assay for differentiating of individual mutant repeat-nucleobase interactions was approved using wt TALE containing RVD HD (SHDGG) opposite the modified cytosines. This afforded differential kinetics for the individual cytosines with the slowest kinetic observed for C, and a high signal/noise ratio (Figure 3-8 e). The FRET assay was therefore decided to be suitable for the screening of the mutant TALEs to study their binding behavior towards the different cytosines.

3.3.2 Screening of truncated TALE repeat libraries

Transferring the FRET assay to a 384-well plate format, a total of 330 unknown repeat-nucleobase interactions were investigated. This screening afforded a number of interesting binding profiles for individual mutant TALE repeats (Figure 3-9 and Figures 6-18 – 6-23). Among the six different libraries, NX*GG and NX**G showed highest overall affinity towards the DNA target. In both libraries, a pronounced selectivity for 5mC was observed. In the case of library NX**G, this preference was especially prominent for mutant repeats containing aromatic amino acids tryptophan (NW**G), tyrosine (NY**G) and histidine (NH**G) at the randomized position 12. A similar preference was observed for repeats containing amino acids with charged side chains such as arginine (NR**G) and lysine (NK**G), or amino acids with polar, uncharged side chains such as serine (NS**G) and asparagine (NN**G).

Binding towards the other cytosine nucleobases was less strong, and highest selectivities were seen with mutant repeats NS**G, NH**G and especially NY**G, whereas repeats NW**G, NR**G and NK**G showed least selective binding behavior. Interestingly, none of the mutant TALE repeats showed binding to 5caC. This observation coincided with results from screenings of libraries X* and X** (chapters 3.2.1 and 3.2.2) and is most likely attributed to the charged state and large steric demand of 5caC[2].

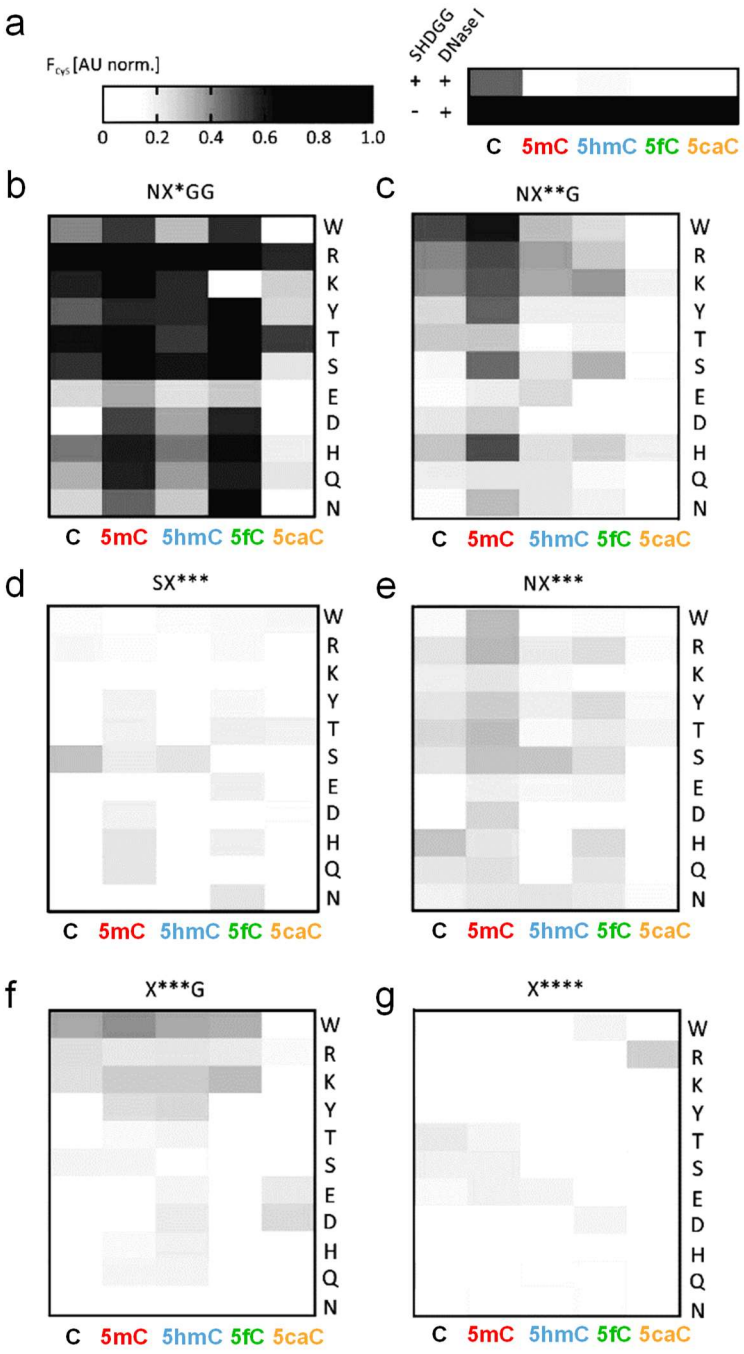


Figure 3-9 DNase I assay screening results for truncated mutant transcription activator-like effector (TALE) repeats. (a) Data of control reactions with wt TALE containing RVD HD (SHDGG) with and w/o DNase I. Conditions as in Figure 3-8 e. (b)-(g) Data obtained from DNase I assay screening for each of the six libraries (library indicated above each heatmap). Randomized amino acid at position X is indicated for each lane on the right and cytosine nucleobase for each column below. Reactions were conducted in duplicates using 5 µM TALE, 0.1 DNA and 1 U DNase I. Heatmaps show normalized Cy5 fluorescence (normalization as in Figure 3-8 e at time point t=25 min. Modified from[1]. Copyright Royal Society of Chemistry 2018.

In the case of library NX*GG, which was the library with the strongest overall binding affinity, most mutant repeats bound 5fC similarly strong as 5mC, resulting in binding profiles with a dual nucleobase preference for 5mC and 5fC. Binding of 5fC was especially pronounced in the cases of repeats containing arginine (NR*GG) and threonine (NT*GG) at the randomized position 12, however, overall selectivity in these two repeats was rather weak. Interestingly, strong binding of 5mC in combination with weak to no binding of C was observed for mutant repeats NQ*GG and ND*GG. These could provide interesting candidates for *in vivo* applications studying the presence of C and 5mC at specific genomic loci.

The next set of libraries, sharing the common characteristic of three deletions (positions 13 to 15), showed rather weak affinity and indistinct selectivity profiles. Additionally, overall binding was reduced if position 11 was a serine (library SX***) than compared to asparagine (library NX***). Repeat mutant SS*** showed higher affinity towards C, which was however the only stronger interaction observed among the TALE repeats of this library. Compared to this, TALE mutants of library NX*** showed a slightly stronger overall affinity, with most repeats again preferentially binding 5mC, followed by 5fC and C. Here, repeats NW***, NR***, NT***, NS***, and ND*** showed the highest affinity, with the most pronounced selectivity for 5mC seen with repeats NW*** and ND***. Furthermore, repeat NH*** exhibited a binding profile where C was bound strongest. Other than that, repeat binding profiles in this library are rather indistinctive with all cytosines being bound rather weakly.

In contrast to the two libraries SX*** and NX***, the third library with deletion of three residues (X***G) distinguishes from the previous two in that the randomization is shifted to position 11 and positions 12 to 14 are deleted. This resulted in a general reduction of affinity accompanied by a loss of specificity. Within this library, highest affinity was seen for repeat mutants W***G and K***G, however, both exhibited a rather diffuse binding profile with no apparent selectivity. Binding performance of the remaining mutant repeats was generally very weak. Introduction of an additional deletion at position 15 resulted in library X**** which provided the smallest repeats studied in the context of this work. Repeat structures containing four deletions basically had no affinity for target DNAs, irrespective of amino acid at the randomized position 11, with the single exception of repeat R****.

3.3.3 Repeat R**** provides a new interaction mode with the oxidized cytosine variant 5-carboxylcytosine

The intriguing binding profile of repeat R**** with a single, positive selectivity for 5caC was especially surprising since this nucleobase was generally bound least by all other mutant repeats. The interaction of R**** with 5caC was the strongest binding observed for this nucleobase among the 66 TALE repeats tested. A 30 min time-course DNase I experiment (Figure 3-10 a) reveals the unusual binding profile of this repeat. For a more quantitative analysis, K_i constants for the mutant repeat R**** were determined using the previously described DNA polymerase accessibility assay. This resulted in a K_i constant of 420.5 nM for repeat R**** binding to target DNA containing 5caC. In contrast, the other cytosines were bound only weakly even at highest applicable TALE protein concentrations (protein solubility limit) (Figure 3-10 c).

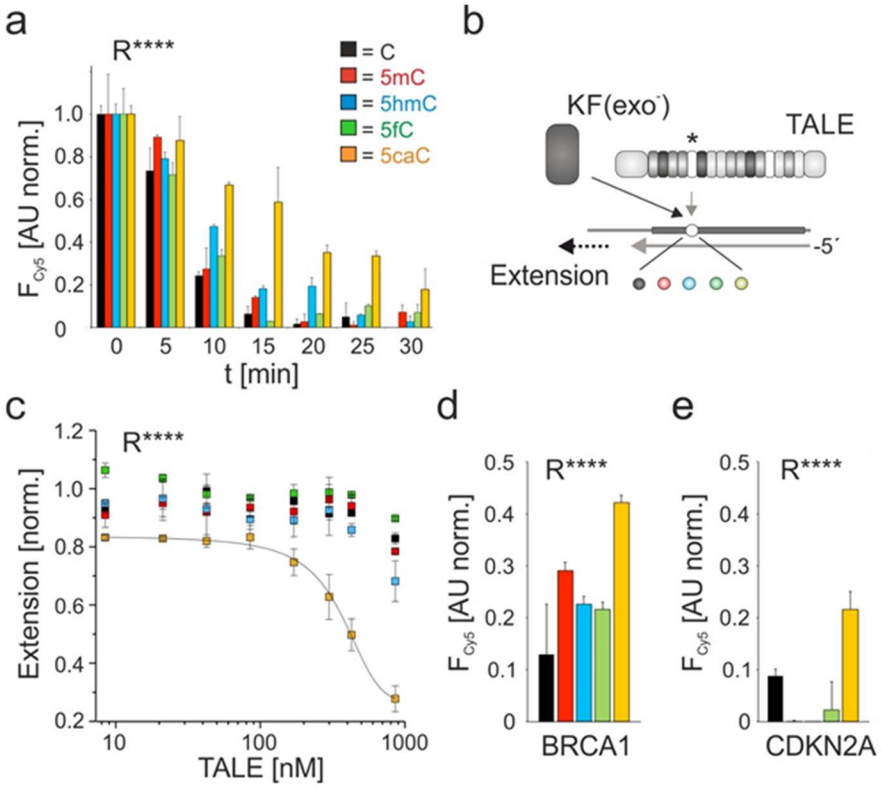


Figure 3-10 Repeat R** as selective binder of 5-carboxylcytosine.** (a) 30 min time-course DNase I experiment as in Figure 3-8 e using 5 µM TALE R****, 0.1 µM DNA and 1 U DNase I. (b) Schematic representation of DNA polymerase accessibility assay as in Figure 3-3 b and c. (c) K_i-determination for mutant TALE R**** using DNA polymerase accessibility assay with 8.3 nM primer/template complex and 25 U KF exo$^-$. Black line represents dose response fit. (d) DNase I experiment for TALE R**** in the context of BRCA1(18) performed as in Figure 3-10 a with 7.5 µM TALE R**** and 0.1 µM DNA. Graph shows Cy5 fluorescence at time point t=30 min. (e) as Figure 3-10 d, but with TALE R**** targeting DNA sequence CDKN2A(18) and 3.5 µM TALE. Modified from[1]. Copyright Royal Society of Chemistry 2018.

In order to exclude a possible context dependency for the 5caC-selectivity of R****, the respective mutant TALE repeat was assebled into two additional TALEs each targeting an 18-nucleotide long DNA sequence of human gene sequences cyclin-dependent kinase inhibitor 2A (CDKN2A) and breast cancer 1 (BRCA1), respectively. In this TALE setup, the mutant TALE repeat was again positioned opposite a CpG within the target sequence, with C modified to each of the five cytosine variants (Figure 3-8 c). Applying these TALEs in the DNasI assay confirmed the 5caC-selectivity of repeat R**** and could thus demonstrate the sequence independency of this selectivity.

A very intriguing question was, how the truncated repeat R**** retains its ability to bind DNA despite the drastic structural changes in its repeat setup. As mentioned earlier, within the DNA-unbound TALE protein state repeats engage in a number of inter-repeat interactions that lead to a preorganisation of the TALE protein. This interaction network made up from van der Waals and polar interactions involves both the helical regions as well as the RVD loop of a repeat[113]. Results from the DNase I screening assay for library X**** revealed a markedly reduced overall affinity compared to other mutant TALE repeat libraries with fewer deletions. This might point towards a dissability of these mutant repeats to commit in essential (inter-repeat) interactions hence eliminating DNA binding ability. With R**** being the clear and single exception within the X**** library, this hints at a unique interaction mode, which distinguishes R**** from the other library members. In order to get a better understanding of the interaction mechanisms of repeat R****, molecular dynamics (MD) simulations with 250 ns trajectories for DNA-undbound TALE[139] CDKN2A(18) containing wt TALE repeat SHDGG, S****, or R****, respectively, were performed (Figure 3-11 and Figure 6-6). Whereas TALE containing wt repeat SHDGG remained quite stable and only showed minimal movements, substitution with repeat S**** led to an extensive kinking at the repeat deletion site (Figure 3-11 a and b), which might result from removed inter-loop interactions. Interestingly, this kinking effect was not observed when repeat R**** was inserted into the TALE protein. Instead, the final equilibrated structure of TALE protein containing repeat R**** greatly resembled the one with wt repeat SHDGG (Figure 3-11 c and d, Figures 6-7 and 6-8). An analysis of side chain dynamics revealed a new inter-loop interaction consisting of a salt bridge between Arg11 of R**** with Asp13 of the preceding HD repeat (Figure 3-8 c and Figure 6-9 and 6-11). The nature of this interaction implied a sequence dependency since a neighbour repeat HD requires a C nucleobase at the respective DNA target position. Therefore, side chain dynamics for repeat R**** in a TALE targeting BRCA1 where the preceding repeat is NI (DNA nucleobase A in the target sequence, Figure 3-8 c) were additionally performed. In this case, no polar interaction between the Arg 11 of repeat R****

and Ile13 of the preceeding repeat NI can be made. R**** selectivity towards 5caC was maintained in the BRCA1 context with MD simulations revealing alternative stabilization interactions including polar interactions between Arg11 of R**** and Asn12, which is not nucleobase specific (Figure 6-10 and 6-12). This alternative interactions could serve as an explanation for the sequence independency of R**** selectivity. Selectivity of TALE repeat mutant R**** was also retained in sequences where RVDs NN or NG (G or T in target sequence) preceeded mutant TALE repeat R**** (Figure 6-17). Both RVDs bear the same identified hydrogen bond donors and acceptors as RVD NI. This further confirms the context independency of the 5caC-selectivity of R**** and indicates its applicability for programmable 5caC detection.

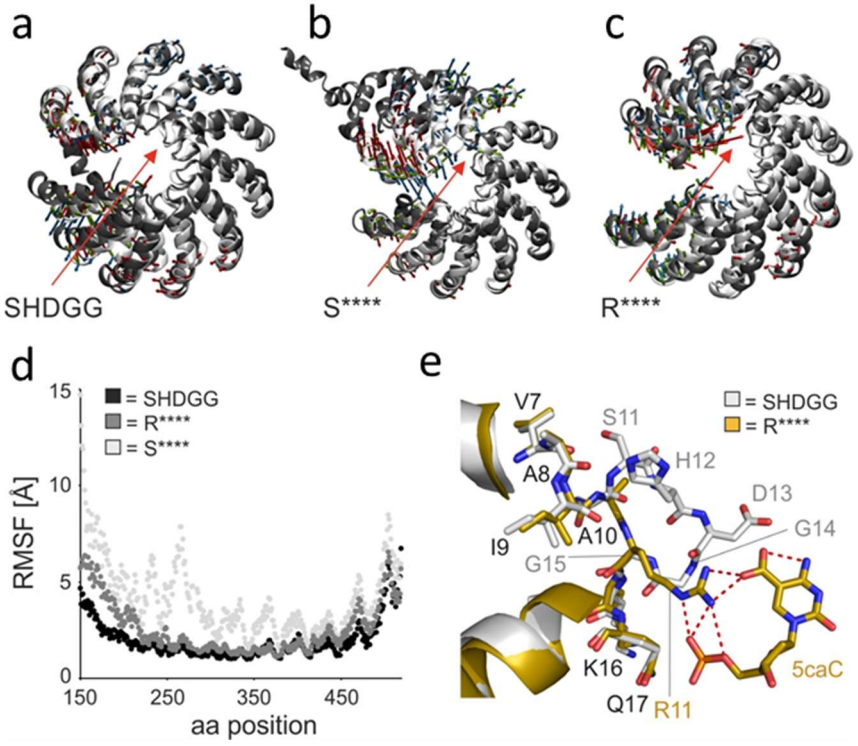


Figure 3-11 Molecular dynamics simulations of wild type and mutant transcription activator-like effector (TALE) repeats. Visualization of conformational changes of transcription activator-like effector (TALE) proteins containing a single SHDGG (a), S**** (b), or R**** (c) repeat (red arrows). White cartoons show the initial structure for the MD production run after equilibration, gray cartoons show representative structures of the cluster that represents the final simulation models at the end of the production run. The small arrows indicate the first three principal components (green, PC1; red, PC2; blue, PC3) resulting from a principal component analysis of the MD trajectory based on the backbone atoms´ movements. (d) Root mean square fluctuations (RMSF) for the protein residues sampled over the 250 ns production run. (e) Crystal structure of TALE-DNA complex containing wt repeat SHDGG (white) opposite C (not shown) superimposed to homology model of repeat R**** opposite 5caC (yellow). Hydrogen bonds are shown as red dashed lines. Modified from[1]. Copyright Royal Society of Chemistry 2018.

Finally, the interaction mechanism by which R**** selectively recognizes 5caC remained to be elucidated. Therefore, homology modeling studies of a TALE targeting CDKN2A(18) bearing repeat mutant R**** in complex with target DNA were performed. Alignments of this model with crystal structures of TALE bearing wt repeat SHDGG[113a] revealed a completely altered repeat organization in the case of R**** (Figure 3-11 e). Due to the deletion of the conserved loop residues 11 to 14 in mutant repeat R****, a large cavity is formed that leads to a vastly increased distance between Arg11 Cα and the 5caC carboxyl C-atom (8.6 Å). This is a clear difference to the otherwise very narrow space observed in conventional repeats, where for example in repeat NG the distance between Cα of Gly13 and the 5-methyl group of T is only 3.4 Å[113]. In contrast to the markedly altered loop structure seen in mutant repeat R****, the structural conformation of the adjacent helix residues remains mainly unchanged. While Cα of Arg11 exactly occupies the Cα-position of Gly15, neighbor residues Ala10 as well as side chains of residues Gln16 and Lys17 are conformationally preserved. Side chain conformation of the latter two amino acid residues is especially important for DNA phosphate-backbone interaction of TALE to DNA[113]. Together with the retained conformation of the repeat helical structures, this might constitute an important factor contributing to integrity and functionality of the mutant repeat R****. With the Cα of Arg11 taking up the exact position of the Gly15 Cα, the arginine side chain seemed to orient similar to the backbone of Gly14 and G15 of natural repeats reaching in the direction of the 5caC nucleobase. Recognition of 5caC thereby ultimately occurs via electrostatic interactions between the arginine guanidinium group and both the 5caC-carboxyl group and the backbone phosphate, thus providing a feasible explanation for the selective recognition of this nucleobase.

In summary, mutant repeat R**** presents the first programmable receptor which selectively and directly recognizes an oxidized 5mC variant in DNA. The selective interaction of the mutant repeat with 5caC was enabled by complete reshaping of the TALE repeat structure, resulting in a truncated TALE repeat. Albeit the fundamental structural differences compared to conventional TALE repeats, mutant repeat R**** retains conformation of essential residues in the loop-adjacent positions, while at the same time engaging in a completely new nucleobase binding mode. In this respect, residue Arg11 adopts a dual role by both forming alternative inter-loop interactions to stabilize overall TALE structure, and selectively recognizing 5caC. This result is especially encouraging, since it demonstrates the general adaptability of TALE repeats and their potential to be reengineered in order to selectively and directly target nucleobases beyond the canonical nucleobases.

4 Summary and Outlook

In the course of this work, a total number of 106 different size-reduced mutant TALE repeats were created, assembled into full-length TALE proteins, and tested *in vitro* for their selectivity towards C, 5mC and the oxidative 5mC variants 5hmC, 5fC, and 5caC in DNA. Studying a total of 530 new repeat-nucleobase interactions, the first programmable receptor for the selective and direct recognition of an oxidized 5mC derivative in DNA was developed.

In a first project, two mutant TALE repeat libraries with saturation mutagenesis of repeat position 12 and a deletion of adjacent positions 13 (library X*), or positions 13 and 14 (library X**) were created. Binding performance of each of the size-reduced mutant TALE repeats towards DNA target sequences containing C, 5mC, 5hmC, 5fC, or 5caC was assessed employing a DNA polymerase accessibility assay. Concluding from the individual binding profiles, deletions of repeat positions 13 and 14 seemed to have a dominating effect which drastically reduced overall TALE affinity. This was mostly independent from amino acid residue at the randomized position 12. However, in the case of only a single deletion (position 13, library X*), different amino acid residues at position 12 led to differential binding of the individual cytosine nucleobases. Library X* contained mutant TALE repeats with altered, mostly dual selectivity (5mC and 5fC), but also universal binders with strong affinity towards all five cytosines (G*, S*, T*). Finally, mutant repeat P* was discovered as the first programmable sensor of 5caC. Homology models of selected mutant TALE repeats interacting with the different cytosine nucleobases allowed to get insights into the individual recognition modes.

In a second project, engineering of TALE repeats was extended to investigate a truncated TALE repeat scaffold, which would allow maximum flexibility for the establishing of new binding modes towards modified cytosines, especially the more bulky oxidized 5mCs. In this context, six different library designs with one to four deletions of repeat positions 12 to 15 and mutation of positions 11 or 12 to amino acid residues that are able to undergo polar interactions were investigated. Using a newly developed FRET-based DNase I competition assay, binding of mutant TALE repeats towards C, 5mC, 5hmC, 5fC, and 5caC was investigated in a high-throughput manner. This resulted in 330 new repeat-nucleobase interactions. Concluding from the binding profiles of individual mutant TALE repeats, affinity generally decreased with increasing numbers of deletions introduced in the repeat loop region. The smallest mutant TALE repeats (library X****) studied in the course of this work showed greatly reduced overall affinity. More detailed investigations of repeat S**** by MD simulations revealed that low affinity could be due to disturbed inter-repeat interactions resulting from altered repeat architecture due to deletions.

However, at the same time this massive restructuring of the TALE repeat paves the way for completely new repeat-nucleobase interaction modes. In this respect, mutant repeat R**** exhibited an intriguing selectivity for a single oxidized 5mC. This repeat indeed presents the first programmable receptor which selectively and directly recognizes an oxidized 5mC derivative in DNA. More thorough investigation of repeat R**** revealed a sequence independency of this effect. Furthermore, MD simulations and homology modeling studies of mutant repeat R**** were employed to elucidate the unique selectivity of this modified repeat. Results indicated a unique, dual role of residue Arg11 which both engages in the formation of a new inter-loop interaction to stabilize overall TALE structure, and selectively recognizes 5caC through an electrostatic interaction mediated by its side chain.

Studying the TALE repeat structure-function relationship, this work considerably helped to get a better understanding of how repeat-nucleobase interactions are created and how they can be modulated or manipulated. Furthermore, this revealed a surprisingly high engineering potential and adaptability of TALE repeats. In the future, mutant TALE repeats with new selectivities can be used for detection of the oxidized 5mC variants in genomic DNA contexts. The main advantage of TALE proteins is their programmability, which offers a very flexible DNA-binding scaffold that can be equipped with selective sensors for specific nucleobases such as repeat R****. Compared to conventional detection methods such as bisulfite-based methods, SMRT or nanopore-based methods, TALE proteins have the great advantage of being applicable in a cellular environment to directly assess DNA modification status in the living organism. In comparison to antibody-based methods, TALE proteins are advantageous in that they allow detection at single-nucleotide resolution and offer strand-specificity. Engineered TALE repeats with a selective and direct recognition potential furthermore obliviate the need for modifying or tagging individual modified nucleobases, which is required in other detection methods. Engineered TALE proteins hence provide a promising, alternative DNA modification detection method that benefits from great flexibility and adaptability. In the case of repeat R**** with its unique selectivity for 5caC, further experiments need to prove its applicability in an *in vivo* environment. Furthermore, TALE repeats for the selective and direct recognition of the other oxidized 5mC variants, 5hmC and 5fC, are required to complete the set of selective TALE repeats. Therefore, additional library designs need to be developed and more mutant TALE repeats need to be investigated for their discrimination potential and their selectivity. In the end, this will result in a complete tool box of engineered TALE repeats with individual selectivities for 5mC, 5hmC, 5fC, and 5caC that can be applied to study the biological function and epigenetic relevance of these DNA modifications.

5 Materials and Methods

5.1 Materials

5.1.1 Oligonucleotides

All oligonucleotides were ordered from Sigma Aldrich and Metabion. 5-methylcytosine nucleotides are marked as C, 5-hydroxymethylcytosine nucleotides are marked as X, 5-formylcytosine nucleotides are marked as Y, and 5-carboxylcytosine nucleotides are marked as Z. All oligonucleotides were ordered as desalted or HPLC-purified 100 µM solutions and diluted to a 10 µM working stock.

5.1.1.1 Sequencing primers

Table 5-1 Sequencing primers

Primer	Description	Sequence
o247	TALE sequencing primer	GGGTTATGCTAGTTATTGCTCAG
o315	TALE sequencing primer	CGTAGAGGATCGAGATC
o814	TALE module plasmid seq. primer	AATGCGCAAACCAACCC
o890	TALE sequencing primer	AGATATGATTGCGGCCCTG
o1408	TALE sequencing primer	AGAAGCGCGAT5CACATG

5.1.1.2 Primers for site-directed mutagenesis of TALE module plasmids

Table 5-2 Primers for site-directed mutagenesis.

Primer	Description	Sequence
o1191	NNK*_rev	GGCGATAGC5CAC5CACTTGGTCCGGAGTCAGGCCATG
o1192	NNK*_fw	GACCAAGTGGTGGCTATCGCCAGCNNKGGCGGCAAGCAAGCGCTCGAAA
o1618	QC primer p735(NNK*→A*)_fw	GGTGGCTATCGCCAGCGCAGGCGGCAAGCAAGCGC
o1619	QC primer p735(NNK*→A*)_rev	CGAGCGCTTGCTTGCCGCCTGCGCTGGCGATAGCC

Primer	Description	Sequence
o1620	QC primer p735(NNK*→D*)_fw	GTGGCTATCGCCAGCGATGGCGGCAAGCAAG
o1621	QC primer p735(NNK*→D*)_rev	CGCTTGCTTGCCGCCATCGCTGGCGATAGCC
o1622	QC primer p735(NNK*→G*)_fw	GTGGCTATCGCCAGCGGTGGCGGCAAGCAAGC
o1623	QC primer p735(NNK*→G*)_rev	GCGCTTGCTTGCCGC5CACCGCTGGCGATAGCC
o1624	QC primer p735(NNK*→H*)_fw	GGCTATCGCCAGCCATGGCGGCAAGCAAG
o1625	QC primer p735(NNK*→H*)_rev	CGCTTGCTTGCCGCCATGGCTGGCGATAG
o1626	QC primer p735(NNK*→I*)_fw	GGCTATCGCCAGCATTGGCGGCAAGCAAGCG
o1627	QC primer p735(NNK*→I*)_rev	CTTGCTTGCCGCCAATGCTGGCGATAGC5CAC
o1628	QC primer p735(NNK*→L*)_fw	GTGGCTATCGCCAGCCTGGGCGGCAAGCAAGC
o1629	QC primer p735(NNK*→L*)_rev	CTTGCTTGCCGCCCAGGCTGGCGATAGC5CACC
o1630	QC primer p735(NNK*→M*)_fw	GGCTATCGCCAGCATGGGCGGCAAGCAAG
o1631	QC primer p735(NNK*→M*)_rev	CGCTTGCTTGCCGCCCATGCTGGCGATAG

Primer	Description	Sequence
o1632	QC primer p735(NNK*→T*)_fw	GGCTATCGCCAG5CACCGGCGGCAAGCAAG
o1633	QC primer p735(NNK*→T*)_rev	CGCTTGCTTGCCGCCGGTGCTGGCGATAG
o1634	QC primer p735(NNK*→W*)_fw	GGTGGCTATCGCCAGCTGGGGCGGCAAGCAAG
o1635	QC primer p735(NNK*→W*)_rev	GCGCTTGCTTGCCGCCCCAGCTGGCGATAGCC
o1636	QC primer p735(NNK*→A**)_fw	GGTGGCTATCGCCAGCGCAGGCAAGCAAGCGC
o1637	QC primer p735(NNK*→A**)_rev	GAGCGCTTGCTTGCCTGCGCTGGCGATAGC5CAC
o1638	QC primer p735(NNK*→R**)_fw	GGTGGCTATCGCCAGCCGTGGCAAGCAAGCGC
o1639	QC primer p735(NNK*→R**)_rev	GAGCGCTTGCTTGC5CACGGCTGGCGATAGC5CAC
o1640	QC primer p735(NNK*→D**)_fw	GGCTATCGCCAGCGACGGCAAGCAAGC
o1641	QC primer p735(NNK*→D**)_rev	GCGCTTGCTTGCCGTCGCTGGCGATAG
o1642	QC primer p735(NNK*→C**)_fw	GGCTATCGCCAGCTGCGGCAAGCAAGCG
o1643	QC primer p735(NNK*→C**)_rev	GCGCTTGCTTGCCGCAGCTGGCGATAGC

Primer	Description	Sequence
o1644	QC primer p735(NNK*→Q**)_fw	GTGGCTATCGCCAGCCAGGGCAAGCAAGCGC
o1645	QC primer p735(NNK*→Q**)_rev	GAGCGCTTGCTTGCCCTGGCTGGCGATAGCC
o1646	QC primer p735(NNK*→E**)_fw	GTGGCTATCGCCAGCGAAGGCAAGCAAGCGC
o1647	QC primer p735(NNK*→E**)_rev	GAGCGCTTGCTTGCCTTCGCTGGCGATAGCC
o1648	QC primer p735(NNK*→H**)_fw	GTGGCTATCGCCAGCCATGGCAAGCAAGCGC
o1649	QC primer p735(NNK*→H**)_rev	GAGCGCTTGCTTGCCATGGCTGGCGATAGCC
o1650	QC primer p735(NNK*→I**)_fw	GTGGCTATCGCCAGCATTGGCAAGCAAGCGC
o1651	QC primer p735(NNK*→I**)_rev	GAGCGCTTGCTTGCCAATGCTGGCGATAGCC
o1652	QC primer p735(NNK*→L**)_fw	GGTGGCTATCGCCAGCCTGGGCAAGCAAGCGC
o1653	QC primer p735(NNK*→L**)_rev	GAGCGCTTGCTTGCCCAGGCTGGCGATAGC5CAC
o1654	QC primer p735(NNK*→K**)_fw	GGCTATCGCCAGCAAAGGCAAGCAAGCG
o1655	QC primer p735(NNK*→K**)_rev	GCGCTTGCTTGCCTTTGCTGGCGATAGC

Primer	Description	Sequence
o1656	QC primer p735(NNK*→M**)_fw	GTGGCTATCGCCAGCATGGGCAAGCAAGCGC
o1657	QC primer p735(NNK*→M**)_rev	GAGCGCTTGCTTGCCCATGCTGGCGATAGCC
o1780	QC primer p735(NNK*→N**)_fw	GGTGGCTATCGCCAGCAATGGCAAGCAAGCGCTC
o1781	QC primer p735(NNK*→N**)_rev	GCGCTTGCTTGCCATTGCTGGCGATAGC5CAC5CAC
o1680	Site-dir. mutagenesis pHD5(VAIASN***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAATAA GCAAGCGCTCGAAACGGTGCAGGAG
o1681	Site-dir. mutagenesis pHD5(VAIASQ***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCCAGAA GCAAGCGCTCGAAACGGTGCAGGAG
o1682	Site-dir. mutagenesis pHD5(VAIASH***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCCATAA GCAAGCGCTCGAAACGGTGCAGGAG
o1683	Site-dir. mutagenesis pHD5(VAIASD***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCGATAA GCAAGCGCTCGAAACGGTGCAGGAG
o1684	Site-dir. mutagenesis pHD5(VAIASE***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCGAAAA GCAAGCGCTCGAAACGGTGCAGGAG
o1685	Site-dir. mutagenesis pHD5(VAIASS***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAGCAA GCAAGCGCTCGAAACGGTGCAGGAG
o1686	Site-dir. mutagenesis pHD5(VAIAST***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAG5CACCA AGCAAGCGCTCGAAACGGTGCAGGAG
o1687	Site-dir. mutagenesis pHD5(VAIASY***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCTATAA GCAAGCGCTCGAAACGGTGCAGGAG

Primer	Description	Sequence
o1688	Site-dir. mutagenesis pHD5(VAIASK***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAAAAA GCAAGCGCTCGAAACGGTGCAGGAG
o1689	Site-dir. mutagenesis pHD5(VAIASR***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCCGTAA GCAAGCGCTCGAAACGGTGCAGGAG
o1690	Site-dir. mutagenesis pHD5(VAIASK***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCTGGAA GCAAGCGCTCGAAACGGTGCAGGAG
o1691	Site-dir. mutagenesis pHD5(VAIANN*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAATGG CGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1692	Site-dir. mutagenesis pHD5(VAIANQ*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCAGG GCGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1693	Site-dir. mutagenesis pHD5(VAIANH*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCATGG CGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1694	Site-dir. mutagenesis pHD5(VAIAND*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACGATGG CGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1695	Site-dir. mutagenesis pHD5(VAIANE*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACGAAG GCGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1696	Site-dir. mutagenesis pHD5(VAIANS*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAGCG GCGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1697	Site-dir. mutagenesis pHD5(VAIANT*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAA5CACCG GCGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1698	Site-dir. mutagenesis pHD5(VAIANA*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACTATGG CGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1699	Site-dir. mutagenesis pHD5(VAIANK*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAAAGG CGGCAAGCAAGCGCTCGAAACGGTGCAGGAG

Primer	Description	Sequence
o1700	Site-dir. mutagenesis pHD5(VAIANR*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCGTGG CGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1701	Site-dir. mutagenesis pHD5(VAIANW*GG)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACTGGG GCGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1702	Site-dir. mutagenesis pHD5(VAIANN**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAATGG CAAGCAAGCGCTCGAAACGGTGCAGGAG
o1703	Site-dir. mutagenesis pHD5(VAIANQ**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCAGG GCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1704	Site-dir. mutagenesis pHD5(VAIANH**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCATGG CAAGCAAGCGCTCGAAACGGTGCAGGAG
o1705	Site-dir. mutagenesis pHD5(VAIAND**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACGATGG CAAGCAAGCGCTCGAAACGGTGCAGGAG
o1706	Site-dir. mutagenesis pHD5(VAIANE**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACGAAG GCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1707	Site-dir. mutagenesis pHD5(VAIANS**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAGCG GCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1708	Site-dir. mutagenesis pHD5(VAIANT**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAA5CACCG GCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1709	Site-dir. mutagenesis pHD5(VAIANY**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACTATGG CAAGCAAGCGCTCGAAACGGTGCAGGAG
o1710	Site-dir. mutagenesis pHD5(VAIANK**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAAAGG CAAGCAAGCGCTCGAAACGGTGCAGGAG
o1711	Site-dir. mutagenesis pHD5(VAIANR**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCGTGG CAAGCAAGCGCTCGAAACGGTGCAGGAG

Primer	Description	Sequence
o1712	Site-dir. mutagenesis pHD5(VAIANW**G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACTGGG GCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1713	Site-dir. mutagenesis pHD5(VAIANN***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAATAA GCAAGCGCTCGAAACGGTGCAGGAG
o1714	Site-dir. mutagenesis pHD5(VAIANQ***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCAGAA GCAAGCGCTCGAAACGGTGCAGGAG
o1715	Site-dir. mutagenesis pHD5(VAIANH***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCATAA GCAAGCGCTCGAAACGGTGCAGGAG
o1716	Site-dir. mutagenesis pHD5(VAIAND***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACGATAA GCAAGCGCTCGAAACGGTGCAGGAG
o1717	Site-dir. mutagenesis pHD5(VAIANE***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACGAAAA GCAAGCGCTCGAAACGGTGCAGGAG
o1718	Site-dir. mutagenesis pHD5(VAIANS***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAGCAA GCAAGCGCTCGAAACGGTGCAGGAG
o1719	Site-dir. mutagenesis pHD5(VAIANT***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAA5CACCA AGCAAGCGCTCGAAACGGTGCAGGAG
o1720	Site-dir. mutagenesis pHD5(VAIANY***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACTATAA GCAAGCGCTCGAAACGGTGCAGGAG
o1721	Site-dir. mutagenesis pHD5(VAIANK***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAAAAA GCAAGCGCTCGAAACGGTGCAGGAG
o1722	Site-dir. mutagenesis pHD5(VAIANR***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCGTAA GCAAGCGCTCGAAACGGTGCAGGAG
o1723	Site-dir. mutagenesis pHD5(VAIANW***)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACTGGAA GCAAGCGCTCGAAACGGTGCAGGAG

Primer	Description	Sequence
o1724	Site-dir. mutagenesis pHD5(VAIAN***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAATGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1725	Site-dir. mutagenesis pHD5(VAIAQ***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCAGGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1726	Site-dir. mutagenesis pHD5(VAIAH***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCATGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1727	Site-dir. mutagenesis pHD5(VAIAD***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCGATGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1728	Site-dir. mutagenesis pHD5(VAIAE***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCGAAGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1729	Site-dir. mutagenesis pHD5(VAIAS***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1730	Site-dir. mutagenesis pHD5(VAIAT***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGC5CACCGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1731	Site-dir. mutagenesis pHD5(VAIAY***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCTATGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1732	Site-dir. mutagenesis pHD5(VAIAK***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAAAGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1733	Site-dir. mutagenesis pHD5(VAIAR***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCGTGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1734	Site-dir. mutagenesis pHD5(VAIAW***G)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCTGGGGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1735	Site-dir. mutagenesis pHD5(VAIAN****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAATAAGCAAGCGCTCGAAACGGTGCAGGAG

Primer	Description	Sequence
o1736	Site-dir. mutagenesis pHD5(VAIAQ****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCAGAAGCA AGCGCTCGAAACGGTGCAGGAG
o1737	Site-dir. mutagenesis pHD5(VAIAH****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCATAAGCA AGCGCTCGAAACGGTGCAGGAG
o1738	Site-dir. mutagenesis pHD5(VAIAD****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCGATAAGCA AGCGCTCGAAACGGTGCAGGAG
o1739	Site-dir. mutagenesis pHD5(VAIAE****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCGAAAAGCA AGCGCTCGAAACGGTGCAGGAG
o1740	Site-dir. mutagenesis pHD5(VAIAS****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAAGCA AGCGCTCGAAACGGTGCAGGAG
o1741	Site-dir. mutagenesis pHD5(VAIAT****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGC5CACCAAGC AAGCGCTCGAAACGGTGCAGGAG
o1742	Site-dir. mutagenesis pHD5(VAIAY****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCTATAAGCA AGCGCTCGAAACGGTGCAGGAG
o1743	Site-dir. mutagenesis pHD5(VAIAK****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAAAAAGCA AGCGCTCGAAACGGTGCAGGAG
o1744	Site-dir. mutagenesis pHD5(VAIAR****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCGTAAGCA AGCGCTCGAAACGGTGCAGGAG
o1745	Site-dir. mutagenesis pHD5(VAIAW****)	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCTGGAAGCA AGCGCTCGAAACGGTGCAGGAG
o1746	Site-dir. mutagenesis pHD5_reverse primer	TTTTCTCGAGGGTCTCCTG5CACCGTTTCGAGCGCTTG

5.1.1.3 Primers for DNA polymerase accessibility assay

Table 5-3 Primers for DNA polymerase accessibility assay

Primer	Description	Sequence
o465	HEY2b_C6→5mC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCTTCT TC5CACCGTGAGCTCTTCCGTTTC5CACATCC
o476	HEY2b	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCTTCT TC5CACCGTGAGCTCTTCCGTTTC5CACATCC
o520	HEY2b_C6→5mC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCTTCT TC5CACCGTGAGCTCTTCXGTTTC5CACATCC
01422	HEY2b_C6→5caC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCTTCT TC5CACCGTGAGCTCTTCZGTTTC5CACATCC
o1423	HEY2b_C6→5fC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCTTCT TC5CACCGTGAGCTCTTCYGTTTC5CACATCC
o466	radiolabeled HEY2b_rev	GGATGTGGAAACGGAAGA

5.1.1.4 Primers for EMSA

Table 5-4 Primers for EMSA

Primer	Description	Sequence
o465	HEY2b_C6→5mC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCTTCTT C5CACCGTGAGCTCTTCCGTTTC5CACATCC
o476	HEY2b	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCTTCTT C5CACCGTGAGCTCTTCCGTTTC5CACATCC
o520	HEY2b_C6→5mC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCTTCTT C5CACCGTGAGCTCTTCXGTTTC5CACATCC
o1422	HEY2b_C6→5caC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCTTCTT C5CACCGTGAGCTCTTCZGTTTC5CACATCC
o1423	HEY2b_C6→5fC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCTTCTT C5CACCGTGAGCTCTTCYGTTTC5CACATCC
o1152	HEY2b_rev	GGATGTGGAAACGGAAGAGCT5CACGGTGGAAGAAGC5CACTGAAGAT GCTGTAACGCTGGGGCTGAAGAGTGGGAATCCA

5.1.1.5 Primers for DNase I competition assay

Table 5-5 Primers for DNase I competition assay

Primer	Description	Sequence
o465	HEY2b_C6→5mC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCT TCTTC5CACCGTGAGCTCTTCCGTTTC5CACATCC
o476	HEY2b_C6→C	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCT TCTTC5CACCGTGAGCTCTTCCGTTTC5CACATCC
o520	HEY2b_C6→5mC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCT TCTTC5CACCGTGAGCTCTTCXGTTTC5CACATCC
o1422	HEY2b_C6→5caC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCT TCTTC5CACCGTGAGCTCTTCZGTTTC5CACATCC
o1423	HEY2b_C6→5fC	TGGATTCC5CACTCTTCAGCCCCAGCGTTACAGCATCTTCAGTGGCT TCTTC5CACCGTGAGCTCTTCYGTTTC5CACATCC
o1892	HEY2b_rev_5´ Cy5_3´ _Cy3	5´Cy5-GGATGTGGAAACGGAAGA-3´Cy3

Primer	Description	Sequence
o1591	CDKN2A(18)_C6→C	GGCCAGCCAGTCAGCCGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o1592	CDKN2A(18)_C6→5mC	GGCCAGCCAGTCAGCCGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o1593	CDKN2A(18)_C6→5hmC	GGCCAGCCAGTCAGCXGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o1594	CDKN2A(18)_C6→5fC	GGCCAGCCAGTCAGCYGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o1595	CDKN2A(18)_C6→5caC	GGCCAGCCAGTCAGCZGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o2503	CDKN2A(18)_rev_5´Cy5_3´ Cy3	Cy5-CATGGAGCCTTCGGCTGA-Cy3
o1516	BRCA1(18)_C6→C	CTTCCTCTTCCGTCTCTTTCCTTTTACGTCATCCGGGGGCAGACT
o1518	BRCA1(18)_C6→5mC	CTTCCTCTTCCGTCTCTTTCCTTTTACGTCATCCGGGGGCAGACT
o1521	BRCA1(18)_C6→5mC	CTTCCTCTTCCGTCTCTTTCCTTTTAXGTCATCCGGGGGCAGACT
o1524	BRCA1(18)_C6→5fC	CTTCCTCTTCCGTCTCTTTCCTTTTAYGTCATCCGGGGGCAGACT
o1527	BRCA1(18)_C6→5caC	CTTCCTCTTCCGTCTCTTTCCTTTTAZGTCATCCGGGGGCAGACT
o2504	BRCA1(18)_rev_5´Cy5_ 3´Cy3	5´Cy5-CCCCCGGATGACGTAAAA-3´Cy3
o3098	CDKN2A(18)_C5→G_C6→C	GGCCAGCCAGTCAG**G**CGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o3099	CDKN2A(18)_C5→G_C6→ 5mC	GGCCAGCCAGTCAG**G**CGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o3100	CDKN2A(18)_C5→G_C6→ 5hmC	GGCCAGCCAGTCAG**G**XGAAGGCTCCATGCTGCTCCCCGCCGCCGGC

Primer	Description	Sequence
o3101	CDKN2A(18)_C5→G_C6→ 5fC	GGCCAGCCAGTCAG**G**YGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o3102	CDKN2A(18)_C5→G_C6→ 5caC	GGCCAGCCAGTCAG**G**ZGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o3103	CDKN2A(18)_C5→G_rev 5´Cy5_3´Cy3	5´Cy5 CATGGAGCCTTCG**C**CTGA 3´Cy3
o3104	CDKN2A(18)_C5→T_C6→C	GGCCAGCCAGTCAG**T**CGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o3105	CDKN2A(18)_C5→T_C6→ 5mC	GGCCAGCCAGTCAG**T**CGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o3106	CDKN2A(18)_C5→T_C6→ 5hmC	GGCCAGCCAGTCAG**T**XGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o3107	CDKN2A(18)_C5→T_C6→ 5fC	GGCCAGCCAGTCAG**T**YGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o3108	CDKN2A(18)_C5→T_C6→ 5caC	GGCCAGCCAGTCAG**T**ZGAAGGCTCCATGCTGCTCCCCGCCGCCGGC
o3109	CDKN2A(18)_C5→T_rev 5´Cy5_3´Cy3	5´Cy5 CATGGAGCCTTCG**A**CTGA 3´Cy3

5.1.2 Plasmids

Table 5-6 List of plasmids

Plasmid	Features	Supplier	Resistance
pHD5_NNK* (pSaM735)	pHD5_12(NNK)_13*	this work	Tetracycline
pHD5_NNK** (pAnI510)	pHD5_12(NNK)_13*_14*	AG Summerer	Tetracycline
pHD5_P* (pSaM828)	pHD5_12P_13*	this work	Tetracycline
pHD5_Y* (pSaM829)	pHD5_12Y_13*	this work	Tetracycline

Plasmid	Features	Supplier	Resistance
pHD5_Q* (pSaM830)	pHD5_12Q_13*	this work	Tetracycline
pHD5_K* (p1284)	pHD5_12K_13*	this work	Tetracycline
pHD5_VAIANH*GG (p1578)	pHD5_11N_12H_13*	this work	Tetracycline
pHD5_VAIANY*GG (p1579)	pHD5_11N_12Y_13*	this work	Tetracycline
pHD5_VAIASN*** (p1581)	pHD5_12N_13*_14*_15*	this work	Tetracycline
pHD5_VAIASQ*** (p1582)	pHD5_12Q_13*_14*_15*	this work	Tetracycline
pHD5_VAIASH*** (p1583)	pHD5_12H_13*_14*_15*	this work	Tetracycline
pHD5_VAIASD*** (p1584)	pHD5_12D_13*_14*_15*	this work	Tetracycline
pHD5_VAIASE*** (p1585)	pHD5_12E_13*_14*_15*	this work	Tetracycline
pHD5_VAIASS*** (p1586)	pHD5_12S_13*_14*_15*	this work	Tetracycline
pHD5_VAIAST*** (p1587)	pHD5_12T_13*_14*_15*	this work	Tetracycline
pHD5_VAIASY*** (p1588)	pHD5_12Y_13*_14*_15*	this work	Tetracycline
pHD5_VAIASK*** (p1589)	pHD5_12K_13*_14*_15*	this work	Tetracycline
pHD5_VAIASR*** (p1590)	pHD5_12R_13*_14*_15*	this work	Tetracycline
pHD5_VAIASW*** (p1591)	pHD5_12W_13*_14*_15*	this work	Tetracycline

Plasmid	Features	Supplier	Resistance
pHD5_VAIANN*GG (p1592)	pHD5_11N_12N_13*	this work	Tetracycline
pHD5_VAIANQ*GG (p1593)	pHD5_11N_12Q_13*	this work	Tetracycline
pHD5_VAIAND*GG (p1594)	pHD5_11N_12D_13*	this work	Tetracycline
pHD5_VAIANE*GG (p1595)	pHD5_11N_12E_13*	this work	Tetracycline
pHD5_VAIANS*GG (p1596)	pHD5_11N_12S_13*	this work	Tetracycline
pHD5_VAIANT*GG (p1597)	pHD5_11N_12T_13*	this work	Tetracycline
pHD5_VAIANK*GG (p1598)	pHD5_11N_12K_13*	this work	Tetracycline
pHD5_VAIANR*GG (p1599)	pHD5_11N_12R_13*	this work	Tetracycline
pHD5_VAIANW*GG (p1600)	pHD5_11N_12W_13*	this work	Tetracycline
pHD5_VAIANN**G (p1601)	pHD5_11N_12N_13*_14*	this work	Tetracycline
pHD5_VAIANQ**G (p1602)	pHD5_11N_12Q_13*_14*	this work	Tetracycline
pHD5_VAIANH**G (p1603)	pHD5_11N_12H_13*_14*	this work	Tetracycline
pHD5_VAIAND**G (p1604)	pHD5_11N_12D_13*_14*	this work	Tetracycline
pHD5_VAIANE**G (p1605)	pHD5_11N_12E_13*_14*	this work	Tetracycline
pHD5_VAIANS**G (p1606)	pHD5_11N_12S_13*_14*	this work	Tetracycline

Plasmid	Features	Supplier	Resistance
pHD5_VAIANT**G (p1607)	pHD5_11N_12T_13*_14*	this work	Tetracycline
pHD5_VAIANY**G (p1608)	pHD5_11N_12Y_13*_14*	this work	Tetracycline
pHD5_VAIANK**G (p1609)	pHD5_11N_12K_13*_14*	this work	Tetracycline
pHD5_VAIANR**G (p1610)	pHD5_11N_12R_13*_14*	this work	Tetracycline
pHD5_VAIANW**G (p1611)	pHD5_11N_12W_13*_14*	this work	Tetracycline
pHD5_VAIANN*** (p1612)	pHD5_11N_12W_13*_14*_15*	this work	Tetracycline
pHD5_VAIANQ*** (p1613)	pHD5_11N_12Q_13*_14*_15*	this work	Tetracycline
pHD5_VAIANH*** (p1614)	pHD5_11N_12H_13*_14*_15*	this work	Tetracycline
pHD5_VAIAND*** (p1615)	pHD5_11N_12D_13*_14*_15*	this work	Tetracycline
pHD5_VAIANE*** (p1616)	pHD5_11N_12E_13*_14*_15*	this work	Tetracycline
pHD5_VAIANS*** (p1617)	pHD5_11N_12S_13*_14*_15*	this work	Tetracycline
pHD5_VAIANT*** (p1618)	pHD5_11N_12T_13*_14*_15*	this work	Tetracycline
pHD5_VAIANY*** (p1619)	pHD5_11N_12Y_13*_14*_15*	this work	Tetracycline
pHD5_VAIANK*** (p1620)	pHD5_11N_12K_13*_14*_15*	this work	Tetracycline
pHD5_VAIANR*** (p1621)	pHD5_11N_12R_13*_14*_15*	this work	Tetracycline

Plasmid	Features	Supplier	Resistance
pHD5_VAIANW*** (p1622)	pHD5_11N_12W_13*_14*_15*	this work	Tetracycline
pHD5_VAIAN***G (p1623)	pHD5_11N_12*_13*_14*	this work	Tetracycline
pHD5_VAIAQ***G (p1624)	pHD5_11Q_12*_13*_14*	this work	Tetracycline
pHD5_VAIAH***G (p1625)	pHD5_11H_12*_13*_14*	this work	Tetracycline
pHD5_VAIAD***G (p1626)	pHD5_11D_12*_13*_14*	this work	Tetracycline
pHD5_VAIAE***G (p1627)	pHD5_11E_12*_13*_14*	this work	Tetracycline
pHD5_VAIAT***G (p1628)	pHD5_11T_12*_13*_14*	this work	Tetracycline
pHD5_VAIAY***G (p1629)	pHD5_11Y_12*_13*_14*	this work	Tetracycline
pHD5_VAIAK***G (p1630)	pHD5_11K_12*_13*_14*	this work	Tetracycline
pHD5_VAIAR***G (p1631)	pHD5_11R_12*_13*_14*	this work	Tetracycline
pHD5_VAIAW***G (p1632)	pHD5_11W_12*_13*_14*	this work	Tetracycline
pHD5_VAIAN**** (p1633)	pHD5_11N_12*_13*_14*_15*	this work	Tetracycline
pHD5_VAIAQ**** (p1634)	pHD5_11Q_12*_13*_14*_15*	this work	Tetracycline
pHD5_VAIAD**** (p1635)	pHD5_11D_12*_13*_14*_15*	this work	Tetracycline
pHD5_VAIAE**** (p1636)	pHD5_11E_12*_13*_14*_15*	this work	Tetracycline

Plasmid	Features	Supplier	Resistance
pHD5_VAIAS**** (p1637)	pHD5_11S_12*_13*_14*_15*	this work	Tetracycline
pHD5_VAIAT**** (p1638)	pHD5_11T_12*_13*_14*_15*	this work	Tetracycline
pHD5_VAIAY**** (p1639)	pHD5_11Y_12*_13*_14*_15*	this work	Tetracycline
pHD5_VAIAK**** (p1640)	pHD5_11K_12*_13*_14*_15*	this work	Tetracycline
pHD5_VAIAR**** (p1641)	pHD5_11R_12*_13*_14*_15*	this work	Tetracycline
pET_MBP_TALE-070_6His (pSaM532)	MBP-TALE-070 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_MBP_TALE-071_6His (pSaM533)	MBP-TALE-071 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-100_6His (pSaM586)	GFP-TALE-100 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-101_6His (pSaM587)	GFP-TALE-101 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-102_6His (pSaM588)	GFP-TALE-102 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-103_6His (pSaM589)	GFP-TALE-103 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-137_6His (pSaM651)	GFP-TALE-137 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-169_6His (pSaM736)	GFP-TALE-169 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-170_6His (pSaM737)	GFP-TALE-170 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-171_6His (pSaM738)	GFP-TALE-171 fusion protein with C-terminal His-tag	this work	Carbenicillin

Plasmid	Features	Supplier	Resistance
pET_GFP_TALE-172_6His (pSaM739)	GFP-TALE-172 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-173_6His (pSaM740)	GFP-TALE-173 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-174_6His (pSaM741)	GFP-TALE-174 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-175_6His (pSaM742)	GFP-TALE-175 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-176_6His (pSaM743)	GFP-TALE-176 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-177_6His (pSaM744)	GFP-TALE-177 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-178_6His (pSaM745)	GFP-TALE-178 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-179_6His (pSaM746)	GFP-TALE-179 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-180_6His (pSaM747)	GFP-TALE-180 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-181_6His (pSaM748)	GFP-TALE-181 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-182_6His (pSaM749)	GFP-TALE-182 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-183_6His (pSaM774)	GFP-TALE-183 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-184_6His (pSaM775)	GFP-TALE-184 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-185_6His (pSaM776)	GFP-TALE-185 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-186_6His (pSaM777)	GFP-TALE-186 fusion protein with C-terminal His-tag	this work	Carbenicillin

Plasmid	Features	Supplier	Resistance
pET_GFP_TALE-187_6His (pSaM778)	GFP-TALE-187 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-188_6His (pSaM779)	GFP-TALE-188 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-189_6His (pSaM780)	GFP-TALE-189 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-203_6His (pSaM831)	GFP-TALE-203 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-204_6His (pSaM832)	GFP-TALE-204 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-205_6His (pSaM833)	GFP-TALE-205 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-206_6His (pSaM834)	GFP-TALE-206 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-207_6His (pSaM835)	GFP-TALE-207 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-208_6His (pSaM836)	GFP-TALE-208 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-209_6His (pSaM837)	GFP-TALE-209 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-210_6His (pSaM838)	GFP-TALE-210 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-211_6His (pSaM839)	GFP-TALE-211fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-212_6His (pSaM840)	GFP-TALE-212 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-213_6His (pSaM841)	GFP-TALE-213 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-214_6His (pSaM842)	GFP-TALE-214 fusion protein with C-terminal His-tag	this work	Carbenicillin

Plasmid	Features	Supplier	Resistance
pET_GFP_TALE-215_6His (pSaM843)	GFP-TALE-215 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-216_6His (pSaM844)	GFP-TALE-216 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-217_6His (pSaM845)	GFP-TALE-217 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-218_6His (pSaM846)	GFP-TALE-218 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-219_6His (pSaM847)	GFP-TALE-219 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-220_6His (pSaM848)	GFP-TALE-220 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-221_6His (pSaM849)	GFP-TALE-221 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-222_6His (pSaM850)	GFP-TALE-222 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-223_6His (pSaM851)	GFP-TALE-223 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-224_6His (pSaM852)	GFP-TALE-224 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-225_6His (pSaM853)	GFP-TALE-225 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-226_6His (pSaM854)	GFP-TALE-226 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-227_6His (pSaM855)	GFP-TALE-227 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-228_6His (pSaM856)	GFP-TALE-228 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-229_6His (pSaM857)	GFP-TALE-229 fusion protein with C-terminal His-tag	this work	Carbenicillin

Plasmid	Features	Supplier	Resistance
pET_GFP_TALE-230_6His (pSaM858)	GFP-TALE-230 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-231_6His (pSaM859)	GFP-TALE-231 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-232_6His (pSaM860)	GFP-TALE-232 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-233_6His (pSaM861)	GFP-TALE-233 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-271_6His (pSaM962)	GFP-TALE-271 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-272_6His (pSaM963)	GFP-TALE-272 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-273_6His (pSaM964)	GFP-TALE-273 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-274_6His (pSaM965)	GFP-TALE-274 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-275_6His (pSaM966)	GFP-TALE-275 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-276_6His (pSaM967)	GFP-TALE-276 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-277_6His (pSaM968)	GFP-TALE-277 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-278_6His (pSaM969)	GFP-TALE-278 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-279_6His (pSaM970)	GFP-TALE-279 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-280_6His (pSaM971)	GFP-TALE-280 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-281_6His (pSaM972)	GFP-TALE-281 fusion protein with C-terminal His-tag	this work	Carbenicillin

Plasmid	Features	Supplier	Resistance
pET_GFP_TALE-282_6His (pSaM973)	GFP-TALE-282 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-283_6His (pSaM974)	GFP-TALE-283 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-284_6His (pSaM975)	GFP-TALE-284 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-285_6His (pSaM976)	GFP-TALE-285 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-286_6His (pSaM977)	GFP-TALE-286 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-287_6His (pSaM978)	GFP-TALE-287 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-288_6His (pSaM979)	GFP-TALE-288 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-311_6His (pSaM1096)	GFP-TALE-311 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-312_6His (pSaM1097)	GFP-TALE-312 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-313_6His (pSaM1098)	GFP-TALE-313 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-314_6His (pSaM1099)	GFP-TALE-314 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-385_6His (pSaM1192)	GFP-TALE-385 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-475_6His (pSaM1196)	GFP-TALE-475 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_6His_GFP-TALE-476 (pSaM1197)	GFP-TALE-476 fusion protein with N-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-492_6His (pSaM1216)	GFP-TALE-492 fusion protein with C-terminal His-tag	this work	Carbenicillin

Plasmid	Features	Supplier	Resistance
pET_GFP_TALE-493_6His (pSaM1217)	GFP-TALE-493 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-494_6His (pSaM1218)	GFP-TALE-494 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-495_6His (pSaM1236)	GFP-TALE-495 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-496_6His (pSaM1237)	GFP-TALE-496 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-497_6His (pSaM1238)	GFP-TALE-497 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-498_6His (pSaM1239)	GFP-TALE-498 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-499_6His (pSaM1240)	GFP-TALE-499 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-500_6His (pSaM1241)	GFP-TALE-500 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-501_6His (pSaM1242)	GFP-TALE-501 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-502_6His (pSaM1243)	GFP-TALE-502 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-503_6His (pSaM1244)	GFP-TALE-503 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-504_6His (pSaM1245)	GFP-TALE-504 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-505_6His (pSaM1246)	GFP-TALE-505 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-506_6His (pSaM1247)	GFP-TALE-506 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-507_6His (pSaM1248)	GFP-TALE-507 fusion protein with C-terminal His-tag	this work	Carbenicillin

Plasmid	Features	Supplier	Resistance
pET_GFP_TALE-508_6His (pSaM1249)	GFP-TALE-508 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-509_6His (pSaM1250)	GFP-TALE-509 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-510_6His (pSaM1251)	GFP-TALE-510 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-511_6His (pSaM1252)	GFP-TALE-511 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-512_6His (pSaM1253)	GFP-TALE-512 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-513_6His (pSaM1254)	GFP-TALE-513 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-514_6His (pSaM1255)	GFP-TALE-514 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-515_6His (pSaM1256)	GFP-TALE-515 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-516_6His (pSaM1257)	GFP-TALE-516 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-517_6His (pSaM1258)	GFP-TALE-517 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-518_6His (pSaM1259)	GFP-TALE-518 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-519_6His (pSaM1260)	GFP-TALE-519 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-520_6His (pSaM1261)	GFP-TALE-520 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-521_6His (pSaM1262)	GFP-TALE-521 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-522_6His (pSaM1263)	GFP-TALE-522 fusion protein with C-terminal His-tag	this work	Carbenicillin

Plasmid	Features	Supplier	Resistance
pET_GFP_TALE-523_6His (pSaM1264)	GFP-TALE-523 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-524_6His (pSaM1265)	GFP-TALE-524 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-525_6His (pSaM1266)	GFP-TALE-525 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-526_6His (pSaM1267)	GFP-TALE-526 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-527_6His (pSaM1268)	GFP-TALE-527 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET-MBP_GG2-entry plasmid (pAnI509)	GG2 entry plasmid with N-terminal MBP and C-terminal 6His	AG Summerer	Carbenicillin
pET_GFP_GG2-entry plasmid (pAnI521)	GG2 entry plasmid with N-terminal GFP and C-terminal 6His	this work	Carbenicillin
pET_6His_GFP_GG2-entry plasmid (pPrR773)	GG2 entry plasmid with N-terminal GFP and 6His	AG Summerer	Carbenicillin
pET_GFP_TALE-598_6His (pSaM1676)	GFP-TALE-598 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-599_6His (pSaM1677)	GFP-TALE-599 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-600_6His (pSaM1678)	GFP-TALE-600 fusion protein with C-terminal His-tag	this work	Carbenicillin
pET_GFP_TALE-601_6His (pSaM1679)	GFP-TALE-601 fusion protein with C-terminal His-tag	this work	Carbenicillin

5.1.3 TALEs

Table 5-7 List of TALEs

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-070	p532	TALE-MBP-1297_LRHD	HD NG NG HD HD NN NG NG NG HD HD NI HD NI NG HD HD	this work
TALE-071	p533	TALE-MBP-1258	HD NN NG HD HD NI NI NG NI NN HD NG NG HD NG	this work
TALE-100	p586	TALE-GFP-1297_LRNG	HD NG NG HD HD NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-101	p587	TALE-GFP-1297_LRHD	HD NG NG HD HD NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-102	p588	TALE-GFP-1257	NG NG NI HD NG NN HD NG NN HD NG HD HD HD NN HD NG	this work
TALE-103	p589	TALE-GFP-1258	HD NN NG HD HD NI NI NG NI NN HD NG NG HD NG	this work
TALE-137	p651	TALE-GFP-1297_LRNG_HD5→N*	HD NG NG HD **VAIASN*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-169	p736	TALE-GFP-1297-LRNG-HD5(12NNK_13*)	HD NG NG HD **(NNK*)** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-170	p737	TALE-GFP-1297-LRNG-HD5(12NNK_13*_14*)	HD NG NG HD **(NNK**)** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-171	p738	TALE-GFP-1297-LRNG-HD5(12A_13*)	HD NG NG HD **VAIASA*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-172	p739	TALE-GFP-1297-LRNG-HD5(12C_13*)	HD NG NG HD **VAIASC*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-173	p740	TALE-GFP-1297-LRNG-HD5(12D_13*)	HD NG NG HD **VAIASD*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-174	p741	TALE-GFP-1297-LRNG-HD5(12E_13*)	HD NG NG HD **VAIASE*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-175	p742	TALE-GFP-1297-LRNG-HD5(12F_13*)	HD NG NG HD **VAIASF*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-176	p743	TALE-GFP-1297-LRNG-HD5(12G_13*)	HD NG NG HD **VAIASG*** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-177	p744	TALE-GFP-1297-LRNG-HD5(12H_13*)	HD NG NG HD **VAIASH*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-178	p745	TALE-GFP-1297-LRNG-HD5(12I_13*)	HD NG NG HD **VAIASI*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-179	p746	TALE-GFP-1297-LRNG-HD5(12K_13*)	HD NG NG HD **VAIASK*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-180	p747	TALE-GFP-1297-LRNG-HD5(12L_13*)	HD NG NG HD **VAIASL*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-181	p748	TALE-GFP-1297-LRNG-HD5(12M_13*)	HD NG NG HD **VAIASM*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-182	p749	TALE-GFP-1297-LRNG-HD5(12P_13*)	HD NG NG HD **VAIASP*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-183	p774	TALE-GFP-1297-LRNG-HD5(12Q_13*)	HD NG NG HD **VAIASQ*** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-184	p775	TALE-GFP-1297-LRNG-HD5(12R_13*)	HD NG NG HD **VAIASR*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-185	p776	TALE-GFP-1297-LRNG-HD5(12S_13*)	HD NG NG HD **VAIASS*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-186	p777	TALE-GFP-1297-LRNG-HD5(12T_13*)	HD NG NG HD **VAIAST*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-187	p778	TALE-GFP-1297-LRNG-HD5(12V_13*)	HD NG NG HD **VAIASV*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-188	p779	TALE-GFP-1297-LRNG-HD5(12W_13*)	HD NG NG HD **VAIASW*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-189	p780	TALE-GFP-1297-LRNG-HD5(12Y_13*)	HD NG NG HD **VAIASY*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-203	p831	TALE-GFP-1297-LRNG-HD5(12A_13*_14*)	HD NG NG HD **VAIASA**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-204	p832	TALE-GFP-1297-LRNG-HD5(12C_13*_14*)	HD NG NG HD **VAIASC**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-205	p833	TALE-GFP-1297-LRNG-HD5(12D_13*_14*)	HD NG NG HD **VAIASD**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-206	p834	TALE-GFP-1297-LRNG-HD5(12E_13*_14*)	HD NG NG HD **VAIASE**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-207	p835	TALE-GFP-1297-LRNG-HD5(12F_13*_14*)	HD NG NG HD **VAIASF**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-208	p836	TALE-GFP-1297-LRNG-HD5(12G_13*_14*)	HD NG NG HD **VAIASG**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-209	p837	TALE-GFP-1297-LRNG-HD5(12H_13*_14*)	HD NG NG HD **VAIASH**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-210	p838	TALE-GFP-1297-LRNG-HD5(12I_13*_14*)	HD NG NG HD **VAIASI**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-211	p839	TALE-GFP-1297-LRNG-HD5(12K_13*_14*)	HD NG NG HD **VAIASK**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-212	p840	TALE-GFP-1297-LRNG-HD5(12L_13*_14*)	HD NG NG HD **VAIASL**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-213	p841	TALE-GFP-1297-LRNG-HD5(12M_13*_14*)	HD NG NG HD **VAIASM**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-214	p842	TALE-GFP-1297-LRNG-HD5(12N_13*_14*)	HD NG NG HD **VAIASN**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-215	p843	TALE-GFP-1297-LRNG-HD5(12P_13*_14*)	HD NG NG HD **VAIASP**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-216	p844	TALE-GFP-1297-LRNG-HD5(12Q_13*_14*)	HD NG NG HD **VAIASQ**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-217	p845	TALE-GFP-1297-LRNG-HD5(12R_13*_14*)	HD NG NG HD **VAIASR**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-218	p846	TALE-GFP-1297-LRNG-HD5(12S_13*_14*)	HD NG NG HD **VAIASS**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-219	p847	TALE-GFP-1297-LRNG-HD5(12T_13*_14*)	HD NG NG HD **VAIAST**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-220	p848	TALE-GFP-1297-LRNG-HD5(12V_13*_14*)	HD NG NG HD **VAIASV**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-221	p849	TALE-GFP-1297-LRNG-HD5(12W_13*_14*)	HD NG NG HD **VAIASW**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-222	p850	TALE-GFP-1297-LRNG-HD5(12Y_13*_14*)	HD NG NG HD **VAIASY**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-223	p851	TALE-GFP-1297-LRNG-HD5(VAIANN***)	HD NG NG HD **VAIANN*** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-224	p852	TALE-GFP-1297-LRNG-HD5(VAIANQ***)	HD NG NG HD **VAIANQ*** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-225	p853	TALE-GFP-1297-LRNG-HD5(VAIANH***)	HD NG NG HD **VAIANH*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-226	p854	TALE-GFP-1297-LRNG-HD5(VAIAND***)	HD NG NG HD **VAIAND*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-227	p855	TALE-GFP-1297-LRNG-HD5(VAIANE***)	HD NG NG HD **VAIANE*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-228	p856	TALE-GFP-1297-LRNG-HD5(VAIANS***)	HD NG NG HD **VAIANS*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-229	p857	TALE-GFP-1297-LRNG-HD5(VAIANT***)	HD NG NG HD **VAIANT*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-230	p858	TALE-GFP-1297-LRNG-HD5(VAIANY***)	HD NG NG HD **VAIANY*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-231	p859	TALE-GFP-1297-LRNG-HD5(VAIANK***)	HD NG NG HD **VAIANK*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-232	p860	TALE-GFP-1297-LRNG-HD5(VAIANR***)	HD NG NG HD **VAIANR*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-233	p861	TALE-GFP-1297-LRNG-HD5(VAIANW***)	HD NG NG HD **VAIANW*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-271	p962	TALE-GFP-1297-LRNG-HD5(VAIASH***)	HD NG NG HD **VAIASH*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-272	p963	TALE-GFP-1297-LRNG-HD5(VAIASD***)	HD NG NG HD **VAIASD*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-273	p964	TALE-GFP-1297-LRNG-HD5(VAIASY***)	HD NG NG HD **VAIASY*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-274	p965	TALE-GFP-1297-LRNG-HD5(VAIASK***)	HD NG NG HD **VAIASK*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-275	p966	TALE-GFP-1297-LRNG-HD5(VAIASR***)	HD NG NG HD **VAIASR*** ** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-276	p967	TALE-GFP-1297-LRNG-HD5(VAIASW***)	HD NG NG HD **VAIASW***** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-277	p968	TALE-GFP-1297-LRNG-HD5(VAIANQ**G)	HD NG NG HD **VAIANQ**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-278	p969	TALE-GFP-1297-LRNG-HD5(VAIANS**G)	HD NG NG HD **VAIANS**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-279	p970	TALE-GFP-1297-LRNG-HD5(VAIANY**G)	HD NG NG HD **VAIANY**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-280	p971	TALE-GFP-1297-LRNG-HD5(VAIAH***G)	HD NG NG HD **VAIAH***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-281	p972	TALE-GFP-1297-LRNG-HD5(VAIAT***G)	HD NG NG HD **VAIAT***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-282	p973	TALE-GFP-1297-LRNG-HD5(VAIAN****)	HD NG NG HD **VAIAN**** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-283	p974	TALE-GFP-1297-LRNG-HD5(VAIAQ****)	HD NG NG HD **VAIAQ****** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-284	p975	TALE-GFP-1297-LRNG-HD5(VAIAD****)	HD NG NG HD **VAIAD****** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-285	p976	TALE-GFP-1297-LRNG-HD5(VAIAE****)	HD NG NG HD **VAIAE****** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-286	p977	TALE-GFP-1297-LRNG-HD5(VAIAT****)	HD NG NG HD **VAIAT****** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-287	p978	TALE-GFP-1297-LRNG-HD5(VAIAY****)	HD NG NG HD **VAIAY****** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-288	p979	TALE-GFP-1297-LRNG-HD5(VAIAW****)	HD NG NG HD **VAIAW****** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-311	p1096	TALE-GFP-1297-LRNG-HD5(VAIAH****)	HD NG NG HD **VAIAH****** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-312	p1097	TALE-GFP-1297-LRNG-HD5(VAIAS****)	HD NG NG HD **VAIAS****** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-313	p1098	TALE-GFP-1297-LRNG-HD5(VAIAK****)	HD NG NG HD **VAIAK****** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-314	p1099	TALE-GFP-1297-LRNG-HD5(VAIAR****)	HD NG NG HD **VAIAR****** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-385	p1192	TALE-GFP-1297-LRNG-NG5	HD NG NG HD **NG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-475	p1196	TALE-GFP-CDKN2A-HD5(VAIAR****)	HD NI NN HD **VAIAR****** NN NI NI NN NN HD NG HD HD NI NG NN HD NG NN HD NG HD HD HD	this work
TALE-476	p1197	TALE-GFP-1297-LRNG	HD NG NG HD HD NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-492	p1216	TALE-GFP-CDKN2A(18)-HD5(VAIAR****)	HD NI NN HD **VAIAR****** NN NI NI NN NN HD NG HD HD NI NG NN	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-493	p1217	TALE-GFP-BRCA1(18)-HD5(VAIAR****)	NG NG NG NI **VAIAR****** NN NG HD NI NG HD HD NN NN NN NN NN	this work
TALE-494	p1218	TALE-GFP-BRCA1(18)	NG NG NG NI HD NN NG HD NI NG HD HD NN NN NN NN NN	this work
TALE-495	p1236	TALE-GFP-1297-LRNG-HD5(VAIASN***)	HD NG NG HD **VAIASN***** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-496	p1237	TALE-GFP-1297-LRNG-HD5(VAIASQ***)	HD NG NG HD **VAIASQ***** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-497	p1238	TALE-GFP-1297-LRNG-HD5(VAIASE***)	HD NG NG HD **VAIASE***** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-498	p1239	TALE-GFP-1297-LRNG-HD5(VAIASS***)	HD NG NG HD **VAIASS***** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-499	p1240	TALE-GFP-1297-LRNG-HD5(VAIAST***)	HD NG NG HD **VAIAST***** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-500	p1241	TALE-GFP-1297-LRNG-HD5(VAIANN*GG)	HD NG NG HD **VAIANN*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-501	p1242	TALE-GFP-1297-LRNG-HD5(VAIANQ*GG)	HD NG NG HD **VAIANQ*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-502	p1243	TALE-GFP-1297-LRNG-HD5(VAIANH*GG)	HD NG NG HD **VAIANH*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-503	p1244	TALE-GFP-1297-LRNG-HD5(VAIAND*GG)	HD NG NG HD **VAIAND*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-504	p1245	TALE-GFP-1297-LRNG-HD5(VAIANE*GG)	HD NG NG HD **VAIANE*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-505	p1246	TALE-GFP-1297-LRNG-HD5(VAIANS*GG)	HD NG NG HD **VAIANS*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-506	p1247	TALE-GFP-1297-LRNG-HD5(VAIANT*GG)	HD NG NG HD **VAIANT*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-507	p1248	TALE-GFP-1297-LRNG-HD5(VAIANY*GG)	HD NG NG HD **VAIANY*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-508	p1249	TALE-GFP-1297-LRNG-HD5(VAIANK*GG)	HD NG NG HD **VAIANK*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-509	p1250	TALE-GFP-1297-LRNG-HD5(VAIANR*GG)	HD NG NG HD **VAIANR*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-510	p1251	TALE-GFP-1297-LRNG-HD5(VAIANW*GG)	HD NG NG HD **VAIANW*GG** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-511	p1252	TALE-GFP-1297-LRNG-HD5(VAIANN**G)	HD NG NG HD **VAIANN**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-512	p1253	TALE-GFP-1297-LRNG-HD5(VAIANH**G)	HD NG NG HD **VAIANH**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-513	p1254	TALE-GFP-1297-LRNG-HD5(VAIAND**G)	HD NG NG HD **VAIAND**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work

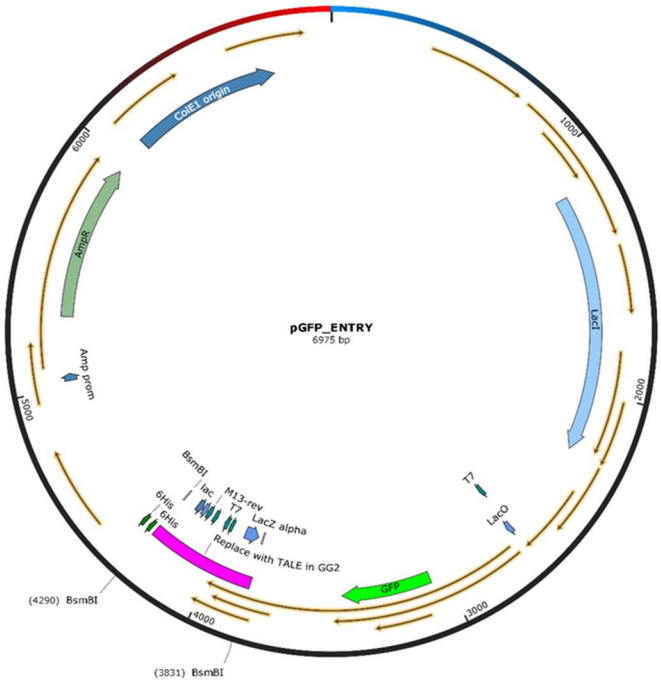
TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-514	p1255	TALE-GFP-1297-LRNG-HD5(VAIANE**G)	HD NG NG HD **VAIANE**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-515	p1256	TALE-GFP-1297-LRNG-HD5(VAIANT**G)	HD NG NG HD **VAIANT**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-516	p1257	TALE-GFP-1297-LRNG-HD5(VAIANK**G)	HD NG NG HD **VAIANK**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-517	p1258	TALE-GFP-1297-LRNG-HD5(VAIANR**G)	HD NG NG HD **VAIANR**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-518	p1259	TALE-GFP-1297-LRNG-HD5(VAIANW**G)	HD NG NG HD **VAIANW**G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-519	p1260	TALE-GFP-1297-LRNG-HD5(VAIAN***G)	HD NG NG HD **VAIAN***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-520	p1261	TALE-GFP-1297-LRNG-HD5(VAIAQ***G)	HD NG NG HD **VAIAQ***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work

TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-521	p1262	TALE-GFP-1297-LRNG-HD5(VAIAD***G)	HD NG NG HD **VAIAD***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-522	p1263	TALE-GFP-1297-LRNG-HD5(VAIAE***G)	HD NG NG HD **VAIAE***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-523	p1264	TALE-GFP-1297-LRNG-HD5(VAIAS***G)	HD NG NG HD **VAIAS***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-524	p1265	TALE-GFP-1297-LRNG-HD5(VAIAY***G)	HD NG NG HD **VAIAY***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-525	p1266	TALE-GFP-1297-LRNG-HD5(VAIAK***G)	HD NG NG HD **VAIAK***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-526	p1267	TALE-GFP-1297-LRNG-HD5(VAIAR***G)	HD NG NG HD **VAIAR***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-527	p1268	TALE-GFP-1297-LRNG-HD5(VAIAW***G)	HD NG NG HD **VAIAW***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work

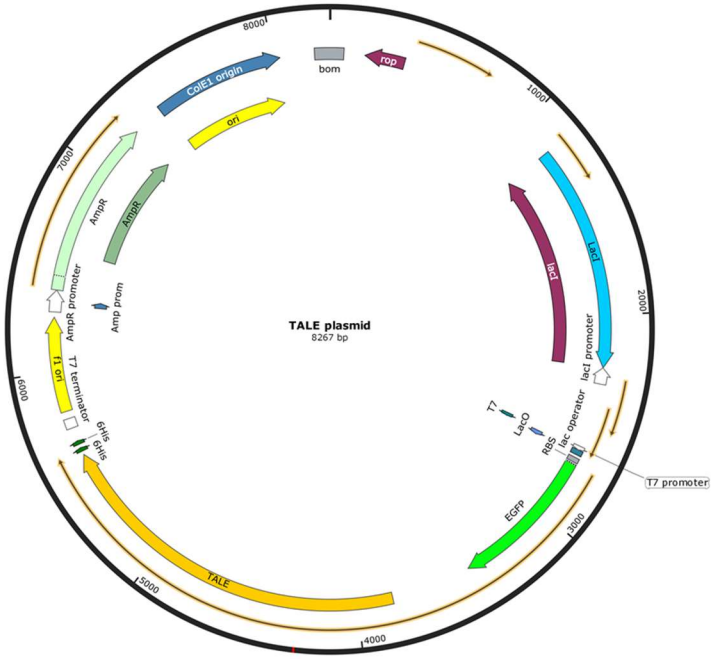
TALE	Plasmid	Features	RVD Sequence	Supplier
TALE-598	p1676	TALE-GFP-1297-LRNG-HD4(NN)_HD5(VAIAR****)	HD NG NG **NN** **VAIAR***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-599	p1677	TALE-GFP-1297-LRNG-HD4(NG)_HD5(VAIAR****)	HD NG NG **NG** **VAIAR***G** NN NG NG NG HD HD NI HD NI NG HD NG	this work
TALE-600	p1678	TALE-GFP-CDKN2A(18)-HD4(NN)_HD5(VAIAR****)	HD NI NN **NN** **VAIAR****** NN NI NI NN NN HD NG HD HD NI NG NN	this work
TALE-601	p1679	TALE-GFP-CDKN2A(18)-HD4(NG)_HD5(VAIAR****)	HD NI NN **NG** **VAIAR****** NN NI NI NN NN HD NG HD HD NI NG NN	this work

5.1.4 Plasmid maps

5.1.4.1 Map of TALE entry vector pGFP-ENTRY



5.1.4.2 Exemplary GFP-TALE library plasmid



5.1.5 TALE protein sequences

5.1.5.1 GFP_TALE_Hey2b

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHM
KQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADK
QKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMD
ELYKTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGK
QWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIAS**HD**GGKQALE
TVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGK
QALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLL
PVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLT
PDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**N**
GGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGKQALESI
VAQLSRPDPALAALTNDHLLEHHHHHH*

5.1.5.2 GFP_TALE_BRCA1(18)

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHM
KQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADK
QKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMD
ELYKTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGK
QWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIAS**NG**GGKQALE
TVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGK
QALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLL
PVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLT
PDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**N**
NGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALESI
VAQLSRPDPALAALTNDHLLEHHHHHH

5.1.5.3 GFP_TALE_CDKN2A(18)

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHM
KQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADK
QKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMD
ELYKTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGK
QWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIAS**HD**GGKQALE
TVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGK
QALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLL
PVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLT
PDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**N**
IGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALESI
VAQLSRPDPALAALTNDHLLEHHHHHH

5.1.5.4 GFP_TALE_CDKN2A(18)_NN4

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHM
KQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADK
QKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMD
ELYKTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGK
QWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIAS**HD**GGKQALE
TVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGK
QALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLL
PVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLT
PDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**N**
IGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALESI
VAQLSRPDPALAALTNDHLLEHHHHHH

5.1.5.5 GFP_TALE_CDKN2A(18)_NG4

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHM
KQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADK
QKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMD
ELYKTLGYSQQQQEKIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGVGK
QWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLTPDQVVAIAS**HD**GGKQALE
TVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGK
QALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLL
PVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLT
PDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**N**
IGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALESI
VAQLSRPDPALAALTNDHLLEHHHHHH

RVDs of canonical repeats are shown bold, variable repeat of this study is red.

5.1.6 List of Chemicals

Table 5-8 List of Chemicals

Chemical	Commercial Supplier	CAS
Acetic Acid	Carl Roth	64-19-7
Agarose	Biozym	9012-36-6
Ammonium peroxodisulfate	Carl Roth	7727-54-0
Boric acid	Carl Roth	10043-35-3
Bromophenol blue	Sigma Aldrich	115-39-9
Calcium chloride	Fisher Scientific	10043-52-4

Chemical	Commercial Supplier	CAS
Carbenicillin disodium salt	Carl Roth	4800-94-6
Chloramphenicol	Carl Roth	56-75-7
Coomassie Brilliant Blue G250	Carl Roth	6104-58-1
Ethanol p.a.	Sigma Aldrich	64-17-5
Ethanol 96 %	Fisher Scientific	64-17-5
Ethidium bromide	Sigma Aldrich	1239-45-8
Ethylenediaminetetraacetate (EDTA)	Carl Roth	60-00-4
Formamide	Acros	75-12-7
Glycine	Carl Roth	56-40-6
Glycerole (Rotipuran)	Carl Roth	56-81-5
Hepes	Carl Roth	7365-45-9
Hydrochloric acid (37 %)	VWR	7647-01-0
Imidazole	abcr	288-32-4
2-Propanol	Fisher Scientific	67-63-0
Kanamycinsulfate	Carl Roth	25389-94-0
LB agar (Lennox)	Carl Roth	-
LB-medium (Lennox)	Carl Roth	-
Magnesium chloride hexadydrate	Acros	7791-18-6
Magnesium sulfate heptahydrate	Merck	10034-99-8
Methanol	Sigma Aldrich	67-56-1
N,N, N', N'-Tetramethyleneethylenediamine	Carl Roth	110-18-9
Phenylmethylsulfonyl fluoride	Carl Roth	329-98-6
Potassium chloride	Carl Roth	7447-40-7
Rotiphorese Gel 40 (37.5:1)	Carl Roth	-
Rotiphorese sequencing gel concentrate	Carl Roth	-
Sodium chloride	Carl Roth	7647-14-5
Sodium dihydrogen phosphate monohydrate	Merck	10049-21-5
Sodium dodecyl sulfate	Carl Roth	151-21-3
Sodium hydroxide	Fisher Scientific	1310-73-2

Chemical	Commercial Supplier	CAS
Sodium phosphate dibasic dihydrate	Sigma Aldrich	10028-24-7
Tetracycline hydrochloride	Carl Roth	64-75-5
Tris(hydroxymethyl)aminomethane	Sigma Aldrich	77-86-1
Triton-X	Fluka	9002-93-1
Urea	Carl Roth	57-13-6
Xylencyanol	Carl Roth	2650-17-1
Yeast extract	Carl Roth	8013-01-2
2-Mercaptoethanol	Merck	60-24-2

5.1.7 Buffers

Table 5-9 List of buffers

Buffers for chemically com*E. coli* cells	Chemical	Concentration
$MgCl_2$ buffer	$MgCl_2$	100 mM
$CaCl_2$ buffer	$CaCl_2$	50 mM
Storage buffer	$CaCl_2$ Glycerol	50 mM 15 % (v/v)

SDS	Chemical	Concentration
SDS loading buffer (4 x) pH 6.8	Tris-HCl Glycerol 2-Mercaptoethanol SDS Bromphenol blue	200 mM 40 % (v/v) 1 % (v/v) 8 % (w/v) 0.008 % (w/v)
SDS running buffer (10 x)	Tris Base Glycine SDS	30.3 g/L 144 g/L 10 g/L

SDS	Chemical	Concentration
Resolving gel (8 %)	MilliQ water	3808 µL
	Rotiphorese Gel 40	1092 µl
	Tris (1.5 M, pH 8.8)	420 µL
	SDS (10 % w/v)	66.7 µL
	APS (10 % w/v)	66.7 µL
	TEMED	6.67 µL
Stacking gel (5 %)	MilliQ water	1000 µL
	Rotiphorese Gel 40	208 µL
	Tris (0.5 M, pH 6.8)	420 µL
	SDS (10 % w/v)	16.7 µL
	APS (10 % w/v)	16.7 µL
	TEMED	1.67 µL
Comassie staining	MilliQ water	40 % (v/v)
	Acetic Acid	10 % (v/v)
	Methanol	50 % (v/v)
	Comassie Brilliant Blue G250	0.1 % (w/v)

PAGE Gel Electrophoresis	Chemical	Concentration
TBE (10x)	Tris	890 mM
	Boric Acid	890 mM
	EDTA	20 mM
Denaturing PAGE gel (10 %)	Sequencing gel concentrate	40 mL
1xTBE	10 x TBE/9 M Urea	10 mL
	9 M Urea	50 mL
	APS	800 µL
	TEMED	40 µL

Agarose Gel Electrophoresis	**Chemical**	**Concentration**
Ethidiumbromid staining solution	Ethidiumbromide	0.1 % (w/v)
Gel loading dye (NEB) 6x pH 8	Ficoll®-400 EDTA Tris-HCl SDS Bromophenol blue	15 % (v/v) 66 mM 19.8 mM 0.102 % (w/v) 0.09 % (w/v)

Protein Purification (native)	**Chemical**	**Concentration**
Taq lysis buffer	Tris-HCl NaCl $MgCl_2$ Triton X-100	10 mM 300 mM 2.5 mM 0.1 % (v/v)
Qiagen lysis buffer pH 8	NaH_2PO4 NaCl	50 mM 300 mM
Qiagen wash buffer I pH 8	NaH_2PO4 NaCl Imidazole	50 mM 300 mM 20 mM
Qiagen wash buffer II pH 8	NaH_2PO4 NaCl Imidazole	50 mM 300 mM 50 mM
Qiagen elution buffer pH 8	NaH_2PO4 NaCl Imidazole	50 mM 300 mM 500 mM
PBS (4 x) pH 8	NaCl KCl $Na_2HPO_4 \times 2H_2O$ KH_2PO_4	32 g/L 3.2 g/L 12.24 g/L 3.2 g/L

DNA polymerase accessibility assay	**Chemical**	**Concentration**
Hybridization buffer 5 x BGrK1, pH 8	Tris NaCl $MgCl_2$ BSA Glycerol	100 mM 250 mM 25 mM 0.5 mg/mL 25 % (v/v)
5 x TALE storage buffer pH 7.5	NaCl Tris-HCl Glycerol	1 mM 100 mM 50 % (v/v)
KF exo $^-$ storage buffer pH 7.4	Tris-HCl DTT EDTA Glycerol	25 mM 1 mM 0.1 mM 50 % (v/v)
Sample buffer (gel loading)	Formamide EDTA Bromophenol blue	80 % (v/v) 2 mM 1 g/L
Running buffer (PAGE) 10 x TBE with Urea pH 8	Tris Boric acid EDTA Urea	890 mM 890 mM 20 mM 9 M

FRET	**Chemical**	**Concentration**
Hybridization buffer 5 x BGrK1, pH 8	Tris NaCl $MgCl_2$ BSA Glycerol	100 mM 250 mM 25 mM 0.5 mg/mL 25 % (v/v)
5 x TALE storage buffer pH 7.5	NaCl Tris-HCl Glycerol	1 mM 100 mM 50 % (v/v)

FRET	Chemical	Concentration
DNase I storage buffer pH 7.5	Tris-HCl $CaCl_2$ Glycerol	50 mM 10 mM 50 % (v/v)

EMSA	Chemical	Concentration
Hybridization buffer 5 x BGrK1 pH 8	Tris-HCl NaCl $MgCl_2$ Glycerol	100 mM 250 mM 25 mM 25 % (v/v)
5 x TALE storage buffer pH 7.5	NaCl Tris-HCl Glycerol	1 mM 100 mM 50 % (v/v)
Running buffer gel 10 x TAE buffer	Tris EDTA pH adjusted with acetic acid	400 mM 25 mM

5.1.8 Antibiotics

Antibiotics were dissolved in the respective solvent and sterile filtered (0.22 µM syringe filter).

Table 5-10 List of antibiotics

Antibiotic Stocks (1000x)	Solvent	Concentration
Carbenicillin	Ethanol/MilliQ (1:1 v/v)	100 mg/L
Kanamycin	MilliQ	50 mg/L
Spectinomycin	MilliQ	100 mg/L
Tetracyclin	MilliQ	12.5 mg/L

5.1.9 Biomolecular Reagents

Table 5-11 List of biomolecular reagents

Reagents	Commercial Supplier
[ɣ-^{32}P]-ATP	Hartmann Analytics
2-log DNA Ladder	New England Biolabs
dNTPs	New England Biolabs
Page Ruler Plus Prestained Protein Ladder	Thermo Fisher Scientific
HisPur Ni-NTA Resin	Thermo Fisher Scientific

Table 5-12 List of enzymes

Enzymes	Commercial Supplier
BsaI	New England Biolabs
BsmBI	New England Biolabs
DNase I	Thermo Fisher Scientific
DpnI	New England Biolabs
KF (exo-)	New England Biolabs
Lysozyme	Sigma Aldrich
NcoI	New England Biolabs
Plasmid Safe	Biozym
Taq Polymerase	New England Biolabs
T4 DNA Ligase	New England Biolabs
T4 Polynucleotide Kinase	Fermentas
XhoI	New England Biolabs

5.1.10 Commercial kits

Table 5-13 List of kits

Kits	Commercial Supplier
BCA Protein Assay	Thermo Fisher Scientific
GeneJET Gel extraction	Thermo Fisher Scientific
GeneJET PCR purification	Thermo Fisher Scientific
NucleoSpin Plasmid Easy Pure	Macherey Nagel

5.1.11 Disposables

Table 5-14 List of disposables

Disposables	Commercial Supplier
96 well plates transparent	Carl Roth
96 deep well plates	Sarstedt
96 well plates black	Greiner
96 fritted deep well plates	Thermo Fisher Scientific
Microplate 384 well, PS, F-Bottom, small volume, black	Greiner
Amicon Ultra MWCO 3 kDa	Millipore
Cannulas Sterican 0.60 x 80 mm	B.Braun
Cannulas Sterican 0.80 x 80 mm	B.Braun
Cannulas Sterican 0.90 x 80 mm	B.Braun
Cuvettes	Sarstedt
Electroporation cuvettes 1 mm	Carl Roth
Falcon tubes 15 ml	Sarstedt
Falcon tubes 50 ml	Sarstedt
Glas beads	VWR
Gloves Nitrile	VWR
Multiply Pro 0.2 mL reaction tube	Sarstedt
Nunc Sealing Tape, breathable, sterile	Thermo Fisher Scientific
Pipette tips (10 µL, 200 µL, 1000 µL)	Sarstedt

Disposables	Commercial Supplier
Reaction tubes 1.5 mL	Sarstedt
Reaction tubes 2.0 mL	Sarstedt
Scalpel	Braun
Sephadex g-25	GE Healthcare
Serological pipette 10 mL	Sarstedt
Serological pipette 25 mL	Sarstedt
Slide-A-Lyzer MINI Dialysis Unit MWCO 10 kDa	Thermo Fisher Scientific
Slide-A-Lyzer Dialysis Cassette MWCO 10 kDa	Thermo Fisher Scientific
Syringe (Omnifix-F, Omnifix 10 mL, 25 mL)	B.Braun
Syringe filter 0.2 µM	Sarstedt
UV cuvettes	Sarstedt
ZelluTrans dialysis membrane T1 MWCO 3.5 kDa	Carl Roth

5.1.12 Lab equipment

Table 5-15 Laboratory equipment

Equipment	Commercial Supplier
Accu Block Digital Dry Bath	Scientific Industries
Balance AX224	Sartorius
Balance PLJ	Kern
Biophotometer Plus	Eppendorf
Camera Powershot G10	Canon
Centrifuge 5810 R	Eppendorf
Centrifuge Sorvall Lynx 6000	Thermo Fisher Scientific
Electrophoresis Power Supply EV233	Consort
Electrophoretic gel system MIDICELL PRIMO EC330	Thermo Scientific
Gel documentation BDAdigital	Biometra
Gel chamber SDS	Biorad
Gel Dryer MGD 4534	VWR

Equipment	Commercial Supplier
Incubator INCU-line	VWR
Incubator I26	New Brunswick Scientific
Infinite M1000 plate reader	Tecan
Magnetic stirrer MR Hei-Mix S	Heidolph
Magnetic stirrer MR Hei-standard	Heidolph
Microwave MW82N	Samsung
Multichannel pipette Xplorer 10 µL (12 channels)	Eppendorf
Multichannel pipette Xplorer 100 µL (12 channels)	Eppendorf
Multichannel pipette Xplorer 300 µL (12 channels)	Eppendorf
Multichannel pipette Picus 1200 1200 µL (12 channels)	Sartorius
Nanodrop 2000	Thermo Fisher Scientific
PCR Cycler SimpliAmp	Thermo Fisher Scientific
PCR Cycler My Cycler	BioRad
pH Meter FiveEasy	Mettler Toledo
Phosphor screen cassettes BAS-Cassette 2025	Fuji
Pipetteboy Easypet	Eppendorf
Plate incubator	Memmert
Power Pac Basic	Biorad
Research Plus Pipettes (2.5, 10, 100, 1000 µL)	Eppendorf
Phosphor screen cassettes BAS-Cassette 2025	Fuji
Scanner CanoScan 9000F	Canon
Tabletop Centrifuge Sprout	Heathrow Scientific
Tabletop Centrifuge 5417R	Eppendorf
Tabletop Centrifuge 5424	Eppendorf
Thermocell Cooling and Heating Block	BIOER
Thermomixer comfort	Eppendorf
Thermomixer compact	Eppendorf

Equipment	Commercial Supplier
Titramax 1000	Heidolph
Typhoon FLA 9500	GE Healthcare
UV-Star UV-table (312 nm)	Biometra
Vortex Genie20	Scientific Industries
Ultra-low temperature freezer U725	New Brunswick Scientific
UV table UV star 312 nm	Biometra
Waterbath JB Aqua 12Plus	Grant

5.1.13 Software

Table 5-16 Software

Software	Supplier
BioDoc Gel documentation	Analytik Jena
ChemDraw 3D 16.0	Cambridgesoft
CorelDraw Graphic Suite X6	Softronic
GraphPad Prism 7	GraphPad Software
ImageJ	National Institutes of Health
ImageQuant TL	GE Healthcare
MS Office 2016	Microsoft
OriginPro 2017G	OriginLab Corporation
Pymol	Schrödinger LLC
Quantitiy One	BioRad

5.2 Methods

5.2.1 Liquid Bacterial Culture

Liquid cultures were handled in glass Erlenmeyer flasks or plastic Falcon tubes. Sterile LB-medium was supplemented with the required antibiotic(s) at a concentration of 100 µg/mL (Carbenicillin), 50 µg/mL (Kanamycin), 100 µg/mL (Spectinomycin) and 12.5 µg/mL (Tetracyclin) and incubated at 37 °C and 180 rpm shaking.

5.2.2 Plate culture

A sterile LB-Agar solution was heated to liquidize, cooled down to approximately 40 °C and then supplemented with the required antibiotic(s) at a concentration of 100 µg/mL (Carbenicillin), 50 µg/mL (Kanamycin), 100 µg/mL (Spectinomycin) and 12.5 µg/mL (Tetracylin). LB-Agar was poured into plastic Petri dishes (Ø 92 mm). *E. coli* were plated close to the flame using glass beads.

5.2.3 Agarose Gel Electorphoresis

Agarose powder was resuspended in 0.5 x TBE buffer and heated to boiling at 600 V in a microwave until the solid was completely dissoveld. After cooling the agarose mixture to approximately 40 °C, a gel of the required size was casted in a casting chamber equipped with a comb. The polymerized gel was transferred into a gel chamber filled with 0.5 x TBE running buffer. Samples were mixed with DNA loading buffer and loaded onto the agarose gel. Gels were run at 80 V for 1h (for a 1 % w/v gel) and stained in an aqueous ethidiumbromide bath for approximately 10 min. After destaining in water, the gels were analyzed on a Biometra UV table at 254 nm equipped with a camera. To extract DNA fragments, the respective DNA band was cut out from the gel using a scalpel and was then extracted using the GeneJET Gel Extraction Kit.

5.2.4 Preparation of chemically competent *E. coli*

A single *E. coli* colony was used to inoculate 20 mL of LB-medium. The culture was incubated at 37 °C and 180 rpm shaking overnight (16 h). 8 mL of this overnight culture were used to inoculate 800 mL LB-medium and the culture was incubated at 37 °C and 180 rpm shaking until an OD_{600} of 0.4 was reached. The culture was then cooled on ice for 30 min and then distributed onto 16 50 mL Falcon tubes. After centrifugation at 4 °C (3500 x g, 10 min) the supernatant was discarded, and the pellets were each

resuspended in ice-cold 100 mM $MgCl_2$ buffer (10 mL). The cell suspension was centrifuged at 4 °C (3500 x g, 10 min), the supernatant was again discarded, and the pellets were each resuspended in ice-cold 50 mM $5caCl_2$ buffer (10 mL) and incubated on ice for 30 min. After centrifugation at 4 °C (3500 x g, 10 min), the supernatant was again discarded, and the pellets were resuspended in a total volume of 4 mL storage buffer (50 mM $5caCl_2$ + 15 % (v/v) glycerol). The cells were divided into 25 µL aliquots, shock-frozen in liquid nitrogen and then stored at -80 °C.

5.2.5 Transformation of DNA plasmids using chemically competent cells

For the transformation of chemically competent *E. coli* cells, 10-100 ng of plasmid solution was mixed with a 25 µL cell aliquot. Tubes were flicked gently three times and were incubated on ice for 30 min. The cells were heat shocked at 42 °C for 30 seconds and then rescued with 500 µL prewarmed SOC (37 °C) followed by incubation at 37 °C (800 rpm) for 1h in a Thermomixer. The cells were plated on LB-Agar plates with the respective antibiotic(s) and incubated at 37 °C over night.

5.2.6 Preparation of electrocompetent *E. coli*

A single *E. coli* colony was picked from an agar plate and used to inoculate 20 ml LB-medium in an Erlenmeyer flask. The culture was incubated at 37 °C overnight (16 h) at 180 rpm shaking. 1200 mL of LB-medium were inoculated with 12 mL of the overnight culture and incubated at 37 °C and 180 rpm shaking until an OD_{600} of 0.4 was reached. The culture was cooled on ice for 30 min and was then distributed on six 250 mL centrifugation bottles and was centrifuged at 4 °C (3500 x g, 10 min). The supernatant was discarded, and the pellets were immediately resuspended in 200 mL ice-cold MilliQ water each and then again centrifuged at 4 °C (3500 x g, 10 min). This washing step was repeated once, and the pellets were then resuspended in 15 mL ice-cold MilliQ water with 10 % (v/v) glycerol and centrifuged at 4 °C (3500 x g, 10 min). The pellets were then resuspended in a total volume of 1.5 mL ice-cold MilliQ water with 10 % (v/v) glycerol, immediately shock-frozen in liquid nitrogen as 50 µL aliquots and then stored at -80 °C.

5.2.7 Transformation of DNA plasmids using electrocompetent cells

For the transformation of electrocompetent cells, 10-100 ng of plasmid solution were mixed with a 50 µL cell aliquot by gently flicking the tube three times. The cells were transferred into an ice-cooled electroporation cuvette (1 mm, BioRad) and transformed at 1800 V (6 ms time constant). After addition of 1 mL prewarmed SOC (37 °C), the cells were incubated at 37 °C (800 rpm) for 1 h in a Thermomixer and then plated on LB-agar plates supplemented with the respective antibiotic(s). Plates were incubated at 37 °C overnight.

5.2.8 DNA Sequencing

For Sanger sequencing, 5 µL of a 80-100 ng/µL DNA sample were mixed with 2.5 µL of the respective sequencing primer (2.5 µM). The volume was adjusted to 10 µL with MilliQ water and the sample was sent to GATC. Sequencing traces were analyzed using the software SnapGene.

5.2.9 TALE assembly and library construction

TALE genes were assembled according to a previously published protocol[115a] with library TALE module plasmids created via site-directed mutagenesis (chapter 5.2.12) from TALE module plasmid pHD5. For the construction of Golden Gate 1 constructs, library TALE module plasmids were pooled and used in a single Golden Gate 1 reaction for each library. Resulting Golden Gate 1 constructs were assembled into entry plasmid pAnI521 in Golden Gate 2 reactions, resulting in plasmids coding for the respective TALE proteins of the eight libraries with N-terminal GFP tag and C-terminal His6 tag.

The respective TALE module plasmids were assembled into the Golden Gate 1 array vectors (pFUS_A and pFUS_B) by a protocol of repetitive digestion with restriction enzyme BsaI (5 U) at 37 °C and ligation with T4 DNA ligase (200 U) at 16 °C (10 cycles). After a final digestion at 50 °C for 5 min and enzyme heat inactivation at 80 °C for 5 min, ATP (0.5 mM) and PlasmidSafe were added to each Golden Gate 1 reaction (incubation at 37 °C for 1h followed by heat inactivation at 70 °C for 20 min) to digest all non-circular DNA. A 25 µL cell aliquot of chemically competent GH371 cells was transformed with 2.5 µL of the Golden Gate 1 reaction mixture according and was plated on agar plates supplemented with 100 µg/mL Spectinomycin and X-Gal (50 µg/mL) for blue-white screening and incubated at 37 °C overnight. The next day, 4 positive clones (white) were picked and further analyzed for their correctly assembled TALE inserts in a colony PCR using primers o734 and o735. For this, 1 µL of a bacterial suspension was

mixed with dNTPs (0.25 mM), primers o 734 and o735 (each 0.1 mM), Thermo Pol buffer (1 x) and Taq polymerase (1 U) in a reaction volume of 25 µL and subjected to the following PCR protocol:

Segment	Temperature [°C]	Time
Initial denaturation	94	2 min
Denaturation **Annealing** **Elongation**	94 55 68	30 s 30 s 1.75 min (35 x)
Final Elongation	68	2 min
Storage	4	∞

The results were monitored on a 1 % agarose gel. Overnight cultures of colonies with the correctly assembled TALE insert were prepared, plasmids were isolated and then used for Golden Gate 2 reactions.

For the construction of pFUS array vectors with library TALE modules, the respective mutant TALE repeats of each library were pooled and used in a Golden Gate1 reaction. Instead of picking 4 white colonies the next day, blue colonies were cut out from the plates and the white colonies were scratched from the plates and used as a Golden Gate 1 pool for the subsequent Golden Gate 2 reactions. Library coverage was monitored to be at least 3-fold.

Golden Gate 2 reactions were performed according to Golden Gate 1 reactions described above. Isolated Golden Gate 1 array plasmids were assembled into the final entry plasmid pET_GFP_ENTRY (p521) using restriction enzyme BsmBI and T4 DNA ligase in five cycles of repetitive digestion and ligation. For the Golden Gate 2 reactions, no PlasmidSafe digest was performed. 2.5 µL of the reaction volume was transformed into a 25 µL aliquot of chemically competent GH371 cells, plated on agar plates supplemented with carbenicillin (100 µg/mL) and X-Gal (50 µg/mL), and incubated at 37 °C overnight. The next day, 4 positive clones (white) were analyzed for the correctly assembled TALE in a colony PCR as described above using primers o890 and o247. Correct insert length was checked on a 1 % agarose gel. Overnight cultures of clones with correctly assembled TALEs were prepared, the plasmids were isolated and subjected to Sanger Sequencing to check for sequence integrity of the TALE plasmids.

5.2.10 Expression of TALE-GFP fusion proteins

For expression and purification of TALEs with GFP tag, overnight cultures of *E. coli* BL21(DE3) Gold transformed with a TALE expression plasmid were grown in LB media supplemented with 100 µg/mL carbenicillin at 37 °C overnight (16 h). The culture was diluted 50-fold into LB medium supplemented with the same antibiotic and was incubated at 37 °C and 180 rpm shaking until an OD_{600} of ~0.4 was reached. IPTG was added to a concentration of 0.2 mM and the culture was harvested by centrifugation after 5 h of further incubation under the same conditions. The pellet was lysed in Lysis buffer (10 mM Tris-HCl, 300 mM NaCl, 2.5 mM $MgCl_2$, 0.1 % Triton X-100, pH = 9) containing 1 mM PMSF and 50 µg/ml Lysozyme (Sigma Aldrich) by shaking at room temperature at 1400 rpm for 30 min. The suspension was pelleted by centrifugation; the supernatant was collected and extracted with Ni-NTA (Thermo Scientific). Ni-NTA was washed two times with 4 x PBS-Buffer (0.55 M NaCl, 43 mM KCl, 69 mM $Na_2HPO_4 \cdot 2H_2O$, 24 mM KH_2PO_4, pH = 8), four times with wash buffer (50 mM $NaH_2PO_4 \cdot H2O$ and 300 mM NaCl pH = 8) containing 20 mM imidazole and once with wash buffer containing 50 mM imidazole. The protein was eluted three times with wash buffer containing 500 mM imidazole. Pooled elution fractions were added to a dialysis tube (Carl Roth) and dialyzed against TALE Storage Buffer (20 mM Tris-HCl pH = 7.5, 200 mM NaCl, 10 % Glycerol, 1 mM DTT). Purity of the TALE protein was analyzed on SDS PAGE stained with Gelcode blue (Thermo Scientific) and quantified by a BCA assay (Pierce). The proteins were snap-frozen and stored in aliquots at -80 °C in TALE storage buffer including 0.1 mg/ml bovine serum albumin (BSA, New England Biolabs).

5.2.11 SDS Gel

Purified TALE proteins were analyzed using SDS gel electrophoresis. For an 8 % resolving gel, 3808 µL MilliQ water, 1092 µL Rotiphorese Gel 40 (Carl Roth), 420 µL Tris (1.5 M, pH 8.8), 66.7 µL SDS (10 % w/v), 66.7 µL APS (10 % w/v) and 6.67 µL TEMED were mixed and poured into a gel caster (BioRad). After polymerization, a 5 % stacking gel from 1000 µL MilliQ water, 208 µL Rotiphorese Gel 40 (Carl Roth), 420 µL Tris (0.5 M, pH 6.8), 16.7 µL SDS (10 % w/v), 16.7 µL APS (10 % w/v) and 1.67 µL TEMED was prepared and poured on top of the resolving gel. After polymerization protein samples were mixed with SDS loading buffer and denatured at 95 °C for 5 min. 10 µL of each sample were loaded onto the gel and gels were run at 120 V for 1.5 h. Gels were stained in brilliant blue staining solution for 60 min. After destaining in H_2O, gels were documented using scanner (CanoScan 9000F, Canon).

5.2.12 Site-directed mutagenesis

5.2.12.1 Quickchange PCR

To introduce site-directed mutations and deletions into the TALE module plasmids, primers were designed according to the guidelines of the QuikChange II Site-directed Mutagenesis Kit (Agilent). Briefly, forward and reverse primers were designed to exhibit a melting temperature (T_m) of ≥ 78 ° C calculated on the the basis of

$$Tm = 81.5 + 0.41(\%GC) - \left(\frac{675}{N}\right) - \%mismatch$$

with N = length of the primer and %GC = GC content of the primer. The primers were designed to start with a G or C and had a minimal GC content of 40 %. Primer length was varied to match the required T_m but generally ranging from 30 to 45 bp. In addition, primers were designed to have a non-self-complementary overlap on one site. PCR reactions were performed in a 50 mL reaction according to:

Component	Volume [µL]	Concentration
Template plasmid	1	75 ng
Primer forward (10 µM]	2.5	400 nM
Primer reverse (10 µM)	2.5	400 nM
dNTPs (10 mM)	1	200 µM
Pfu Buffer (10 x)	5	1 x
MilliQ water	37	
Pfu DNA polymerase	1	1.25 U

The following PCR program that was used to perform Quickchange PCR:

Segment	Temperature [°C]	Time
Initial denaturation	95	4 min
Denaturation **Annealing** **Elongation**	95 55 68	30 s 1 min 9 min (25 x)
Final Elongation	68	10 min
Storage	4	∞

PCR reactions were digested with DpnI (1 µL, 20 U) overnight at 37 °C without further purification followed by heat inactivation at 80 °C for 20 min. 2 µL of the reaction mixture were transformed into a 25 µL cell aliquot of chemically competent cells (see chapter 5.2.5).

5.2.12.2 Cassette mutagenesis

For the introduction of mutations and higher numbers of deletions into TALE module plasmids, site-directed cassette mutagenesis using the restriction enzymes NcoI and XhoI was used. Forward primers were designed carrying the respective mutations and deletions and a NcoI restriction site at their 5´end. A universal reverse primer with a XhoI restriction site at the 5´end was designed to overlap with the forward primer outside the area of mutations and deletions. First, the respective forward primer was hybridized to the reverse primer according to the following protocol in a 25 µL reaction:

Component	Volume [µL]	Concentration
Primer forward (µM)	1	4 µM
Primer reverse (µM)	1	4 µM
NEB2 (10 x)	2.5	1 x
MilliQ water	20.5	

Reactions were incubated at 95 °C for 5 minutes and then cooled down to RT for 30 min. In each reaction, 1 µL dNTPs (200 µM), 2.5 µL NEB2, 21 µL MilliQ water and 0.5 µL KF exo- (25 U) to perform primer extension in a 50 µL reaction at 37 °C for 1h followed by heat inactivation at 75 °C (30 min). Reactions were purified using the GeneJET PCR purification kit (Thermo Scientific) and DNA inserts were eluted in 50 µL of MilliQ water. The digest of the inserts was performed in a 60 µL reaction volume:

Component	Volume [µL]	Concentration
DNA insert (elution from Purif.)	50	
NcoI	0.5	10 U
XhoI	0.5	10 U
Cutsmart (10 x)	6	1 x
MilliQ water	3	

Reactions were incubated at 37 °C for 1 h and again purified using the GeneJET PCR purification kit. In parallel, the template plasmid (TALE module plasmid) was digested with the same restriction enzymes in a 150 µL reaction:

Component	Volume [µL]	Concentration
Template plasmid	100	10 µg
NcoI	2.5	50 U
XhoI	2.5	50 U
Cutsmart (10 x)	15	1 x
MilliQ water	30	

Reactions were incubated at 37 °C for 2 h and purified using the GeneJET PCR purification kit.

Inserts and digested template plasmids were ligated using a molar ratio of insert/plasmid of 10:1 in a 10 µL reaction:

Component	Volume [µL]	Concentration
Template plasmid **Insert**	2 2	 Molar ratio insert/plasmid 10:1
T4 DNA ligase buffer (10 x)	1	1 x
T4 DNA ligase	0.5	200 U
MilliQ water	4.5	

Reactions were incubated at 16 °C overnight (16 h) followed by heat inactivation at 65 °C for 10 min. 4 µL of the reaction mixture was transformed into a 25 µL cell aliquot of chemically competent cells (chapter 5.2.5).

5.2.13 DNA polymerase accessibility assay

Templates HEY2b and modified versions HEY2b_C6_55mC, 55mC, 55fC, and 55caC (o476, o465, o520, o1423 or o1422) and respective primer pHEY2b (o466, 5´-^{32}P-end labeled) were annealed at 100 nM and 33.3 nM in 6 µL Hybridization buffer (40 mM Tris-HCl (pH = 8.0), 100 mM NaCl, 10 mM $MgCl_2$, 0.2 mg/ml BSA, 10 % glycerol) by incubation at 95 °C for 5 min and then at room temperature for 30 min. TALE proteins were added in 6 µL TALE storage buffer (200 mM NaCl, 20 mM Tris-HCl (pH = 7.5), 1 mM DTT, 10 % glycerol) at varying concentrations and mixtures were incubated at room temperature for 30 min. 12 µL of a mixture of 25 mU Klenow fragment of *E. coli* DNA polymerase I (3´-5´-exo-(KF(exo-)), New England Biolabs) and 200 µM of each dNTP in binding buffer (30 mM Tris-HCl (pH = 8.0), 150 mM NaCl, 5 mM $MgCl_2$, 0.1 mg/ml BSA, 0.5 mM DTT, 7.5 % glycerol) were added and the mixtures were incubated at room temperature for 15 min. 12 µL PAGE loading buffer (80 % formamide, 2 mM EDTA) were added and the mixtures were incubated at 95 °C for 5 min and cooled on ice. Mixtures were analyzed by denaturing PAGE gel (9 M urea). Gels were dried on a gel drier for 3 h and then exposed to a Storage Phosphor Screen (Fuji) overnight followed by imaging using a Typhoon FLA9500.

5.2.14 EMSA

Oligonucleotide oGrK476, oGrKo465, oGrK520, oGrK1422 and oGrK1423 were hybridized each with oGrK1152 or oSaM2170 (Cy5 labeled) in 1 x BGrK2 buffer (20 mM Tris, 50 mM NaCl, 5 mM $MgCl_2$ and 5 % v/v glycerol, pH = 8 adjusted using HCl) by incubating at 95 °C for 5 min and at RT for 30 min. Then, 0.5 pmol of the respective DNA duplex was incubated with 0.1 pmol of the respective TALE protein in 1x BGrK2 buffer with a final volume of 10 µL. The reaction was kept at room temperature for 30 min and then on ice for 30 min. The EMSA gel (0.5 x TAE buffer, 8 % Rotiphorese Gel 40 (Carl Roth), 0.1 % APS and 0.01 % TEMED) was pre-run at 4 °C for 30 min in a Mini Protean vertical electrophoresis cell (BioRad), loaded with 5 µL sample and the run for 90 min at 4 °C and 70 V. GFP and Cy5 fluorescence of the gels was recorded with a Typhoon FLA-9500 laser scanner (GE Healthcare) and analyzed using the software Image J. TALE-DNA complex formation was calculated by building the ratios of the shifted band (TALE-DNA complex) and the complete area of the same lane.

5.2.15 DNase I competition assay

For the DNase I competition assay either HEY2b templates o476, o465, o1422 and o1423, BRCA1(18) templates o1516, o1518, o1521, o1524 and o1527 or CDKN2A(18) templates o1591, o1592, o1593, o1594 and o1595 were hybridized with their respective reverse complement o1892 (5´-Cy5 and 3´-Cy3 labeled, HEY2b), o2504 (5´-Cy5 and 3´-Cy3 labeled, BRCA1(18)) or o2503 (5´-Cy5 and 3´-Cy3 labeled, CDKN2A(18)) at a concentration of 200 nM in 3 µL hybridization buffer (40 mM Tris-HCl (pH = 8.0), 100 mM NaCl, 10 mM $MgCl_2$, 0.2 mg/ml BSA, 10 % glycerol) by incubation at 95 °C for 5 min and then at room temperature for 30 min. TALEs were added in 3 µL TALE storage buffer (200 mM NaCl, 20 mM Tris-HCl (pH = 7.5), 1 mM DTT, 10 % glycerol) at indicated concentrations and mixtures were incubated at room temperature for 30 min. 6 µL of a mixture of 1 U Dnase I (Thermo Scientific) in Dnase I buffer (20 mM Tris-HCl (pH = 7.5), 5 mM $MgCl_2$ and 0.2 mM 5caCl_2) were added and the plate containing the mix was placed immediately in a TECAN M1000 plate reader, pre-heated at 37 °C. Fluorescence by Cy5 emission was measured at 665 nm every 5 min. Cy5 fluorescence of a control in absence of TALE was subtracted from the data and the ratio of Cy5 fluorescence of the TALE samples to that of a control without Dnase I were plotted as relative Cy5 fluorescence.

6 Appendix

6.1 TALE repeat conformation analysis

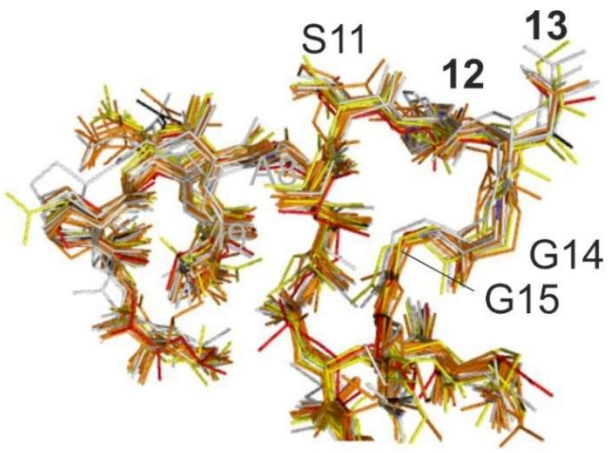


Figure 6-1 TALE repeat conformation. Alignments of 24 repeats from two different TALE crystal structures that cover all standard RVDs (PDB entries 3UGM[113a] and 3V6T[113b]). RVDs HD, NG, NN and NI are colored gray, yellow, red and orange, respectively. Modified from[1]. Copyright Royal Society of Chemistry 2018.

6.2 Homology modeling data of mutant TALE repeats from library X*

6.2.1 Homology models of mutant TALE repeat W*

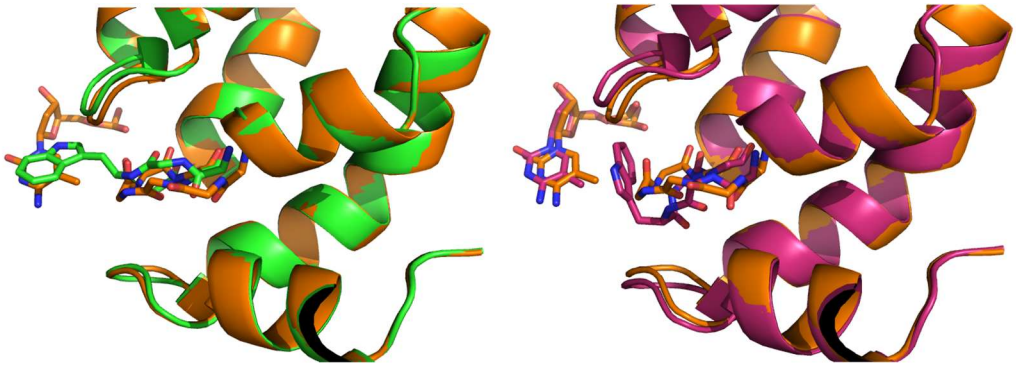


Figure 6-2 Homology models of transcription activator-like effector (TALE) mutant repeat W*. Left side: repeat NG interacting with 5mC (orange), original x-ray structure PDB entry 4GJR[10c] and repeat W* without DNA (green): In the absence of DNA, Trp12 occupies the nucleobase space. Right side: repeat NG (orange) and repeat W* (pink) interacting with DNA containing 5mC. In the presence of DNA, the side chain of Trp12 is forced between the nucleobase and the RVD loop. Modified from[2]. Copyright 2016 American Chemical Society.

6.2.2 Homology models of mutant TALE repeats L*, M*, A*, and V*.

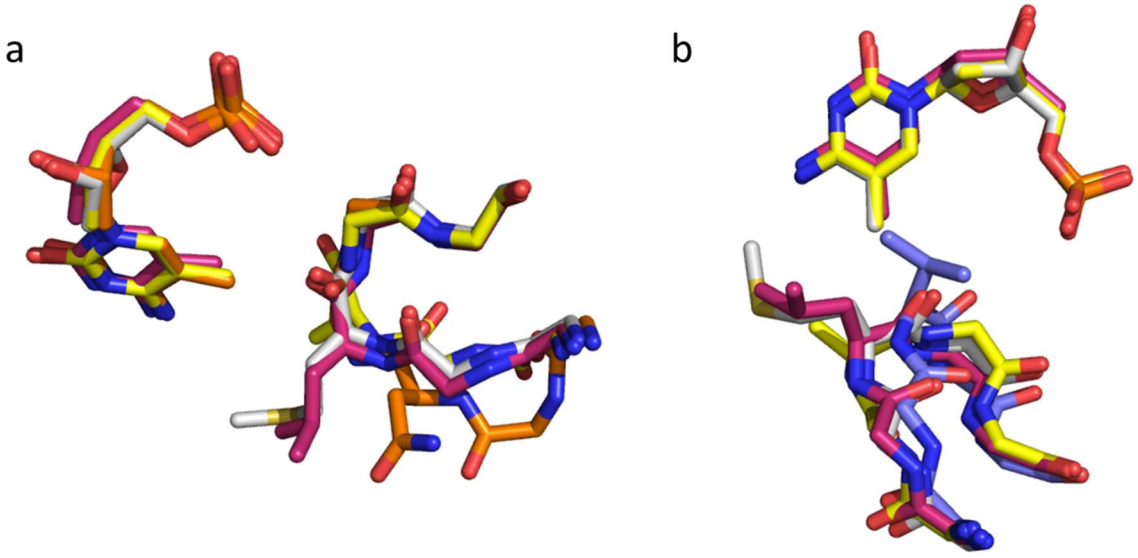


Figure 6-3 Homology models of repeats NG, L*, M*, A*, and V*. (a) Repeats NG (orange, original x-ray structure PDB entry 4GJR[10c]), L* (pink), M* (gray) and A* (yellow) interacting with 5mC. (b) Repeats V* (blue), L* (pink), M* (gray) and A* (yellow) interacting with 5mC. The V* *β*-branch side chain has a different orientation positioning in it in close proximity to the methyl group of 5mC. This could prevent binding of the nucleobase. Modified from[2]. Copyright 2016 American Chemical Society.

6.2.3 Homology models of mutant TALE repeats S*, T*, and G*.

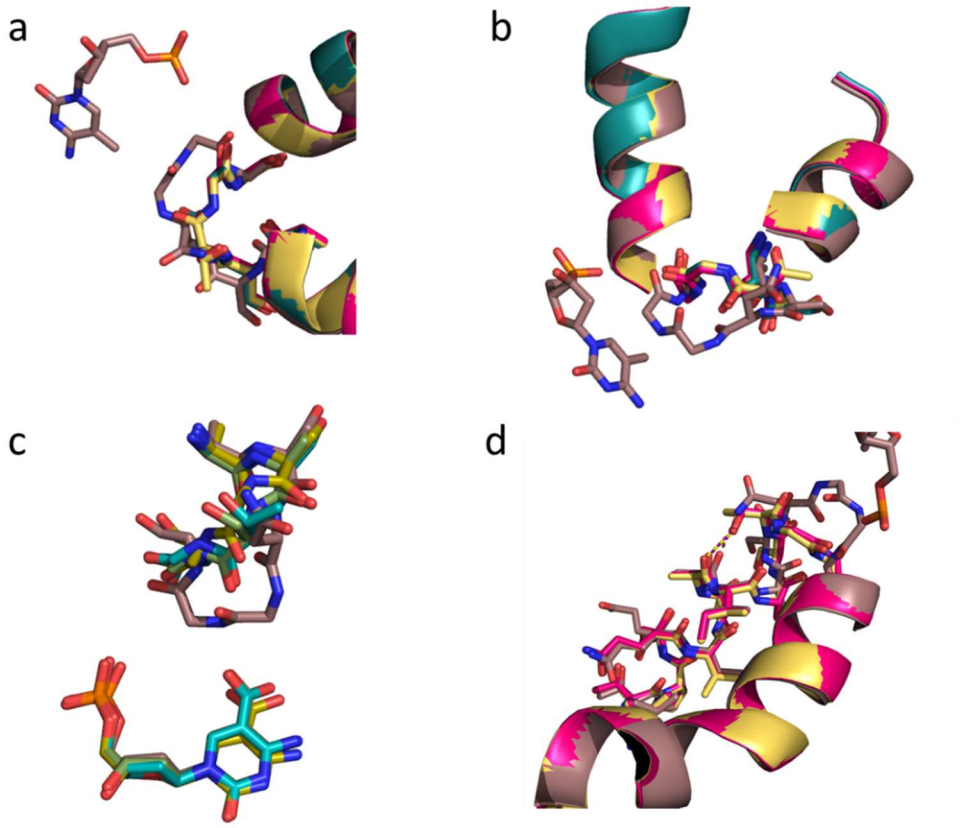


Figure 6-4 Homology models of repeats NG, S*, T* and G*. (a) and (b) show two different orientations of repeat NG interacting with 5mC (brown, original x-ray structure PDB entry 4GJR), and repeats G* (cyan), S* (pink) and T* (yellow) without DNA. (c) repeat NG interacting with 5mC (brown, original x-ray structure PDB entry 4GJR[10c]), and repeats G* (yellow), S* (green) and T* (cyan) interacting with 5caC. (d) repeat NG interacting with 5mC (brown, original x-ray structure PDB 4GJR), repeat S* interacting with 5caC (yellow), and repeat T* interacting with 5caC (pink). Modified from[2]. Copyright 2016 American Chemical Society.

6.3 K_i-determination for TALE mutant repeat P*.

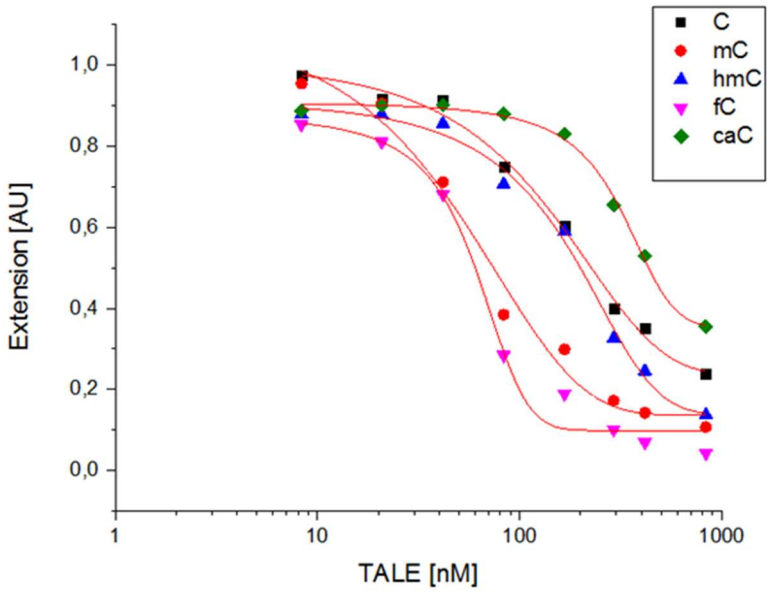


Figure 6-5 K_i-determination of mutant repeat P*. K_i-determination for TALE_Hey_2b containing mutant repeat P* interacting with C, 5mC, 5hmC, 5fC and 5caC. Fits are shown as red curves. Modified from[2]. Copyright 2016 American Chemical Society.

6.4 Molecular dynamics data for mutant TALE repeat R****.

6.4.1 Initial structures after equilibration for TALEs CDKN2A(18) bearing mutant TALE repeats S****, R****, or wt repeat SHDGG.

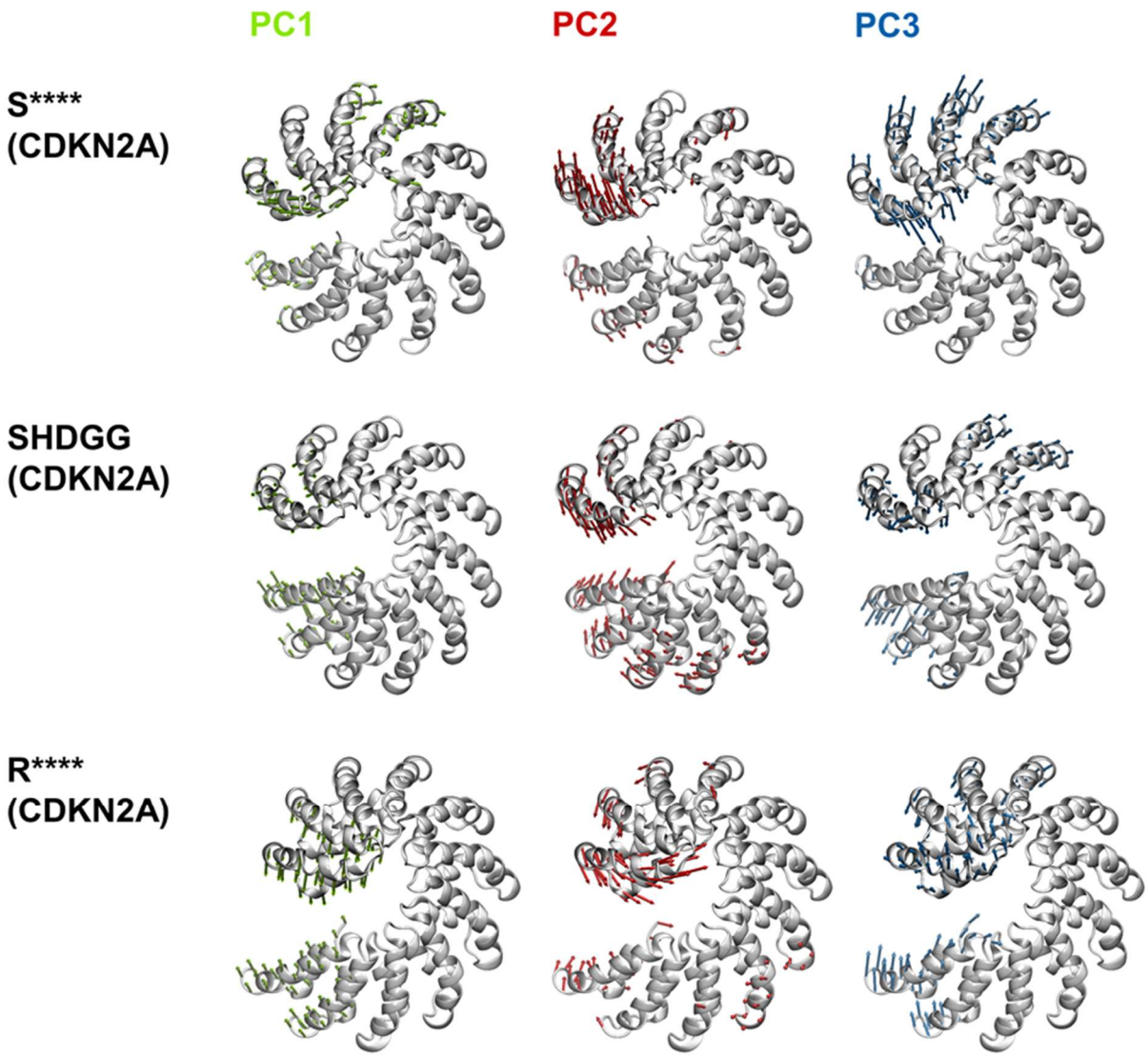


Figure 6-6 PCA analysis for TALEs CDKN2A(18) with wildtype repeat SHDGG, S**, or R**** at position of repeat five.** Directions of the first three PCA-derived modes used for the generation of Figure 3-11 a-c. The initial structures after equilibration are shown. Modified from [1]. Copyright Royal Society of Chemistry 2018.

6.4.2 RMSF plots for TALEs bearing mutant TALE repeats S****, R****, or wt repeat SHDGG.

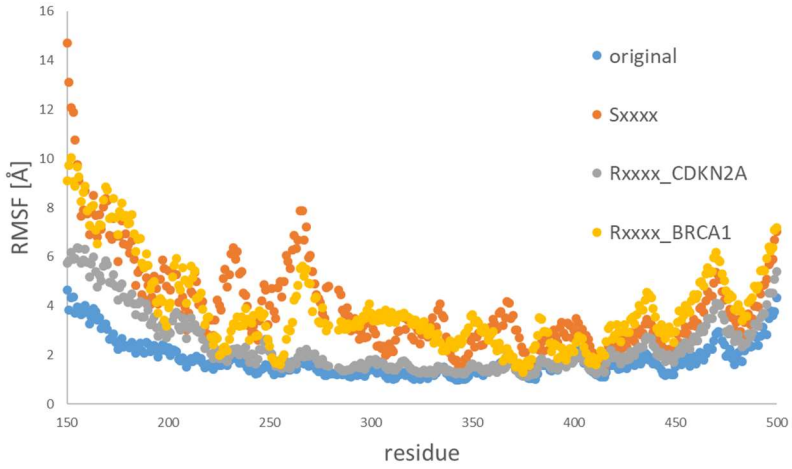


Figure 6-7 RMSF plot for TALEs CDKN2A(18) and BRCA(18) containing mutant repeats (alignment of all repeats). RMSF plot for TALEs bearing repeats SHDGG (wt, targeting C, blue), S**** (orange), or R**** (gray) in the context of TALE targeting CDKN2A, and for repeat R**** (yellow) in the context of TALE targeting BRCA1. Simulations in the absence of DNA. Modified from[1]. Copyright Royal Society of Chemistry 2018.

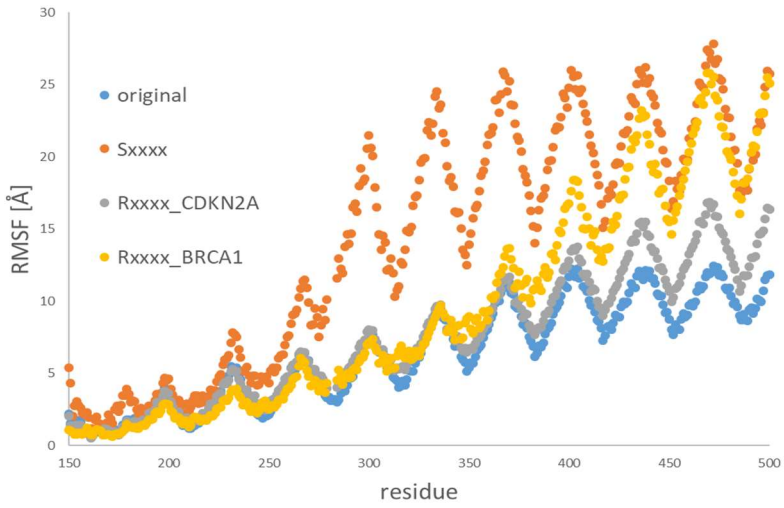


Figure 6-8 RMSF plot for TALEs CDKN2A(18) and BRCA1(18) containing mutant TALE repeats (alignment of the first N-terminal repeat). RMSF plot for TALEs bearing repeats SHDGG (wt, targeting C, blue), S**** (orange), or R**** (gray) in the context of TALE targeting CDKN2A, and for repeat R**** (yellow) in the context of TALE targeting BRCA1. Simulations in the absence of DNA. Modified from[1]. Copyright Royal Society of Chemistry 2018.

6.4.3 Analysis of hydrogen bonding for mutant TALE repeat R**** in the context CDKN2A(18).

donor	acceptor	occupancy
LYS134-Side	GLU138-Side	17,90%
ARG133-Side	ILE97-Main	0,03%
GLU138-Main	LYS134-Main	31,33%
LEU137-Main	ARG133-Main	20,16%
ARG133-Side	SER99-Side	3,94%
ARG133-Side	ALA132-Main	0,21%
ALA132-Main	VAL129-Main	0,01%
ALA132-Main	VAL128-Main	33,54%
LYS134-Side	THR109-Side	0,90%
ARG133-Main	ILE131-Main	8,42%
LYS134-Side	GLN105-Side	3,83%
ARG133-Side	ALA96-Main	0,06%
ARG133-Side	SER99-Main	1,73%
ARG133-Side	ASN165-Side	1,08%
ARG133-Side	**ASP101-Side**	**61,07%**
ARG133-Side	SER163-Main	0,32%
ARG133-Side	ILE131-Main	0,74%
ARG133-Side	ASN164-Main	0,02%
LYS134-Side	GLN135-Side	2,32%
LYS134-Side	ASP101-Side	0,18%
LYS134-Side	SER99-Main	0,03%
LYS134-Side	HIE100-Main	0,73%
ARG133-Side	GLN135-Side	0,12%
ALA136-Main	ARG133-Main	0,02%
ARG133-Side	HIE100-Main	0,02%

Figure 6-9 Hydrogen bonding analysis for mutant TALE repeat R** in the CDKN2A(18) context.** Analysis of hydrogen bonding observed for R**** in TALE designed for CDKN2A(18) context during a 250 ns MD trajectory. Hydrogen bonds via repeat loop to downstream (towards C-terminus) and upstream (towards N-terminus) repeats are highlighted in red and green, respectively. Modified from[1]. Copyright Royal Society of Chemistry 2018.

6.4.4 Analysis of hydrogen bonding for mutant TALE repeat R**** in the context BRCA1(18).

donor	acceptor	occupancy
LEU137-Main	ARG133-Main	12,98%
LYS134-Side	GLU138-Side	13,20%
GLU138-Main	LYS134-Main	17,23%
ALA132-Main	VAL128-Main	14,32%
ARG133-Main	ILE131-Main	3,95%
ARG133-Side	ASN165-Main	0,22%
ARG133-Side	ASN165-Side	1,19%
ARG133-Side	ILE131-Main	0,83%
ARG133-Side	**ASN100-Side**	**8,18%**
LYS134-Side	THR109-Side	1,42%
ARG133-Side	ASN164-Main	0,08%
ARG133-Side	ALA162-Main	0,22%
ARG133-Side	SER163-Main	0,42%
LYS134-Side	GLN105-Side	2,30%
LYS134-Main	**ASN100-Side**	**10,52%**
ASN100-Side	ALA132-Main	1,44%
LYS134-Side	GLN135-Side	1,16%
ARG133-Side	SER99-Side	2,46%
ARG133-Side	ILE101-Main	0,13%
ARG133-Side	ALA130-Main	0,72%
LYS134-Side	**ILE101-Main**	**7,36%**
ARG133-Side	GLN135-Side	0,76%
ALA132-Main	VAL129-Main	2,10%
ARG133-Side	ALA132-Main	0,58%
LYS134-Side	GLN105-Main	0,01%
ARG133-Side	ALA96-Main	3,38%
ARG133-Side	SER99-Main	0,73%
ALA136-Main	ARG133-Main	0,82%
LYS134-Side	GLY102-Main	0,08%
ARG133-Side	VAL129-Main	0,04%
ARG133-Side	SER163-Side	0,02%

Figure 6-10 Hydrogen bonding analysis for mutant TALE repeat R** in the BRCA1(18) context.** Analysis of hydrogen bonding observed for R**** in TALE designed for BRCA1(18) context during a 250 ns MD trajectory. Hydrogen bonds via repeat loop to downstream (towards C-terminus) and upstream (towards N-terminus) repeats are highlighted in red and green, respectively. Modified from[1]. Copyright Royal Society of Chemistry 2018.

6.4.5 Stabilizing hydrogen bonding of mutant TALE repeat R**** in the sequence context CDKN2A(18).

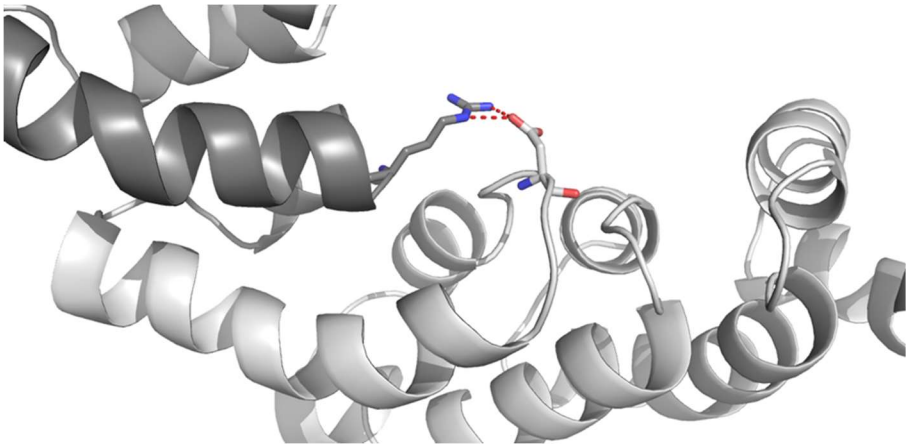


Figure 6-11 Hydrogen bonding observed for mutant repeat R** in the CDKN2A(18) context.** Stabilizing hydrogen bond between Arg11 of repeat R**** and Asp13 of preceding repeat HD seen in MD simulations of DNA-unbound TALE designed for target sequence CDKN2A. Snapshot at 249 ns. Repeat R**** in gray, other repeats in white. Modified from[1]. Copyright Royal Society of Chemistry 2018.

6.4.6 Stabilizing hydrogen bonding of mutant TALE repeat R**** in sequence context BRCA1(18).

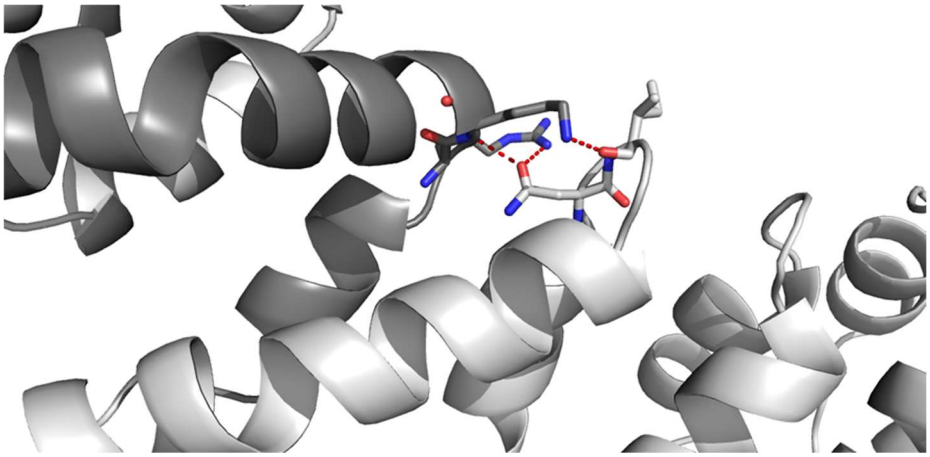


Figure 6-12 Hydrogen bonding observed for mutant repeat R** in BRCA1(18) context.** Stabilizing hydrogen bond between repeat R**** and preceding repeat NI seen in MD simulations of DNA-unbound TALE designed for target sequence BRCA1(18). Snapshot at 238 ns. Repeat R**** in gray, other repeats in white. Modified from[1]. Copyright Royal Society of Chemistry 2018.

6.5 Exemplary SDS-gel from purification of TALEs containing mutant repeats.

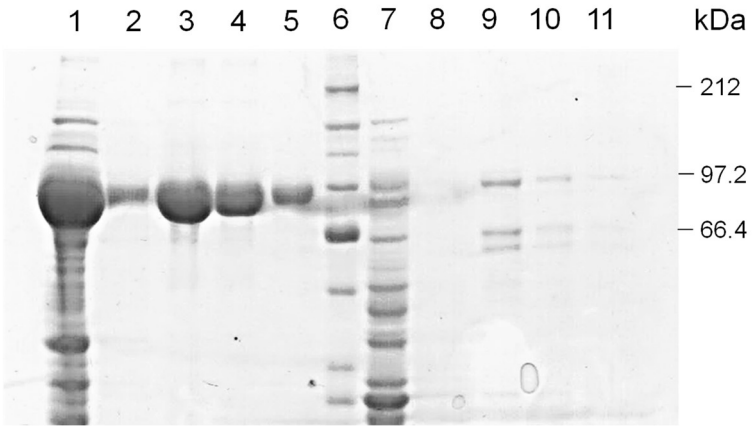


Figure 6-13 Exemplary SDS-gel from TALE purification. Lanes 1 to 5: Induction with 200 mM IPTG. Lane 1: NiNTA wash, Lane 2: wash (PBS), Lanes 3-5: Elution 1 to 3. Lane 6: Protein marker. Lanes 7 to 11: without induction. Lane 7: NiNTA wash, Lane 8: wash (PBS), Lanes 9 to 11: Elution 1 to 3.

6.6 K_i-determination for mutant TALE repeat R****.

6.6.1 K_i-determination for mutant TALE repeat R**** (C and 5mC).

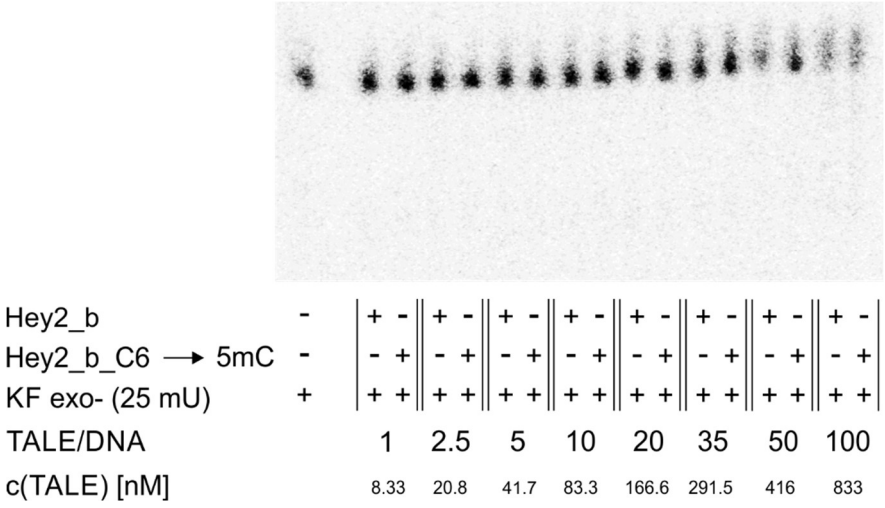


Figure 6-14 PAGE results for K_i-determination of mutant repeat R** (C and 5mC).** PAGE results of a DNA polymerase accessibility assay for a TALE targeting an 18 nt sequence of zebrafish gene Hey2_b with mutant repeat R**** opposite C or 5mC at target position six.

6.6.2 K_i-determination for mutant TALE repeat R**** (5hmC and 5fC).

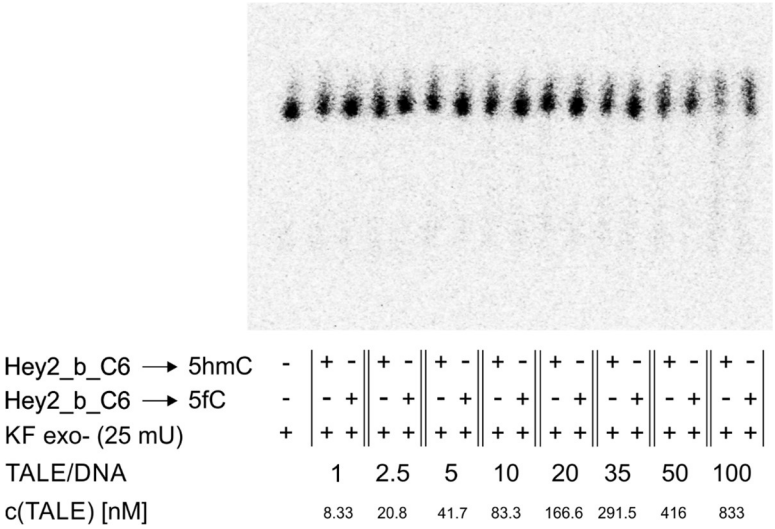


Figure 6-15 PAGE results for K_i-determination of mutant repeat R** (5hmC and 5fC).** PAGE results of a DNA polymerase accessibility assay for a TALE targeting an 18 nt sequence of zebrafish gene Hey2_b with mutant repeat R**** opposite 5hmC or 5fC at target position six.

6.6.3 K_i-determination for mutant TALE repeat R**** (5caC).

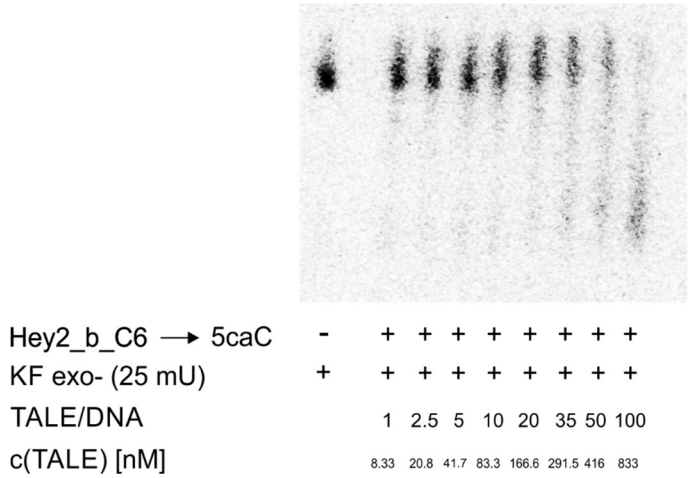
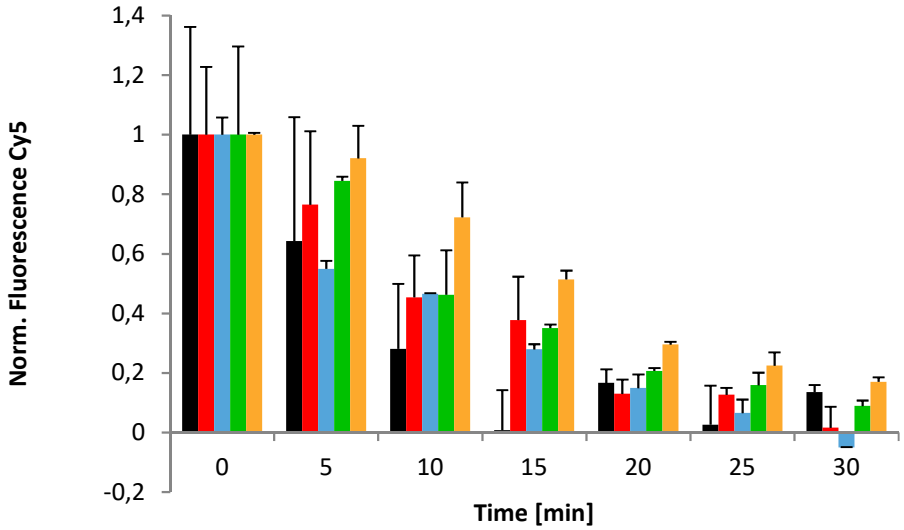


Figure 6-16 PAGE results for K_i-determination of mutant repeat R** (5caC).** PAGE results of a DNA polymerase accessibility assay for a TALE targeting an 18 nt sequence of zebrafish gene Hey2_b with mutant repeat R**** opposite 5caC at target position six.

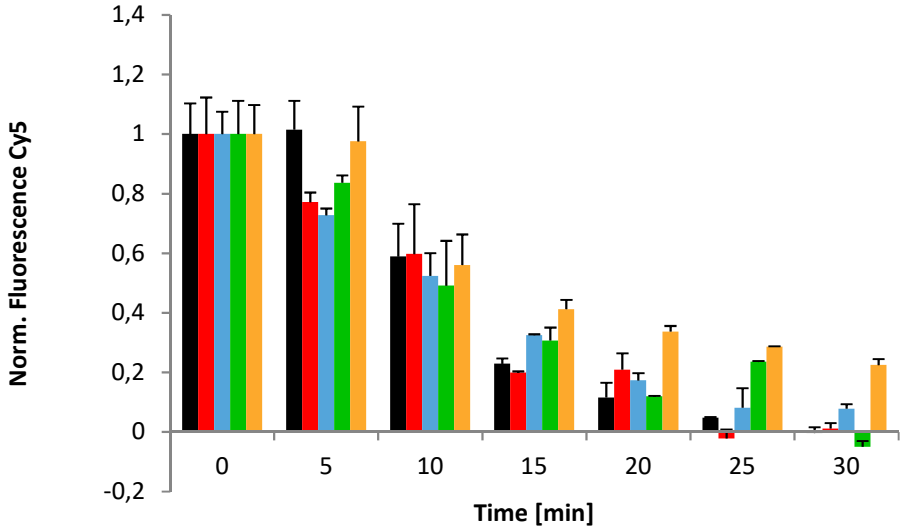
6.7 DNase I competition assay results for libraries NX*GG, NX**G, SX***, NX***, X***G, and X****.

6.7.1 Context variation for mutant TALE repeat R****.

CDKN2A(18)_HD



CDKN2A(18)_NN



CDKN2A(18)_NG

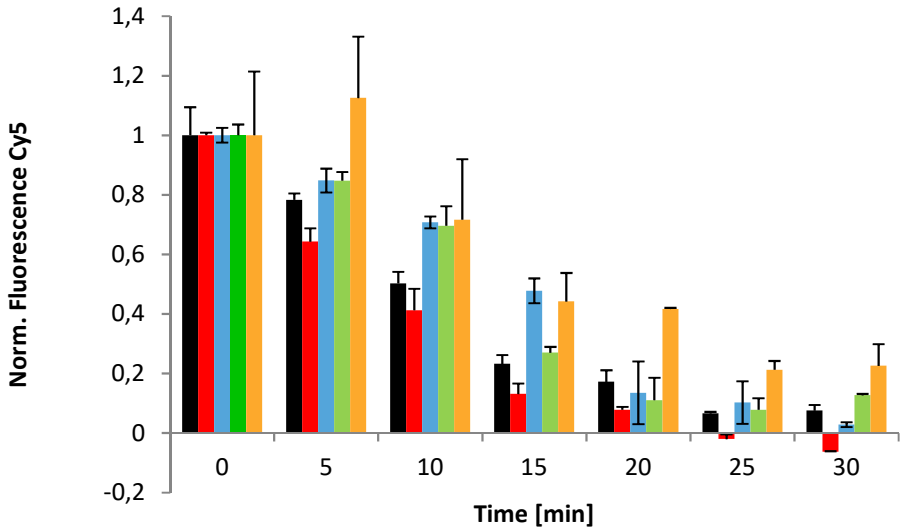
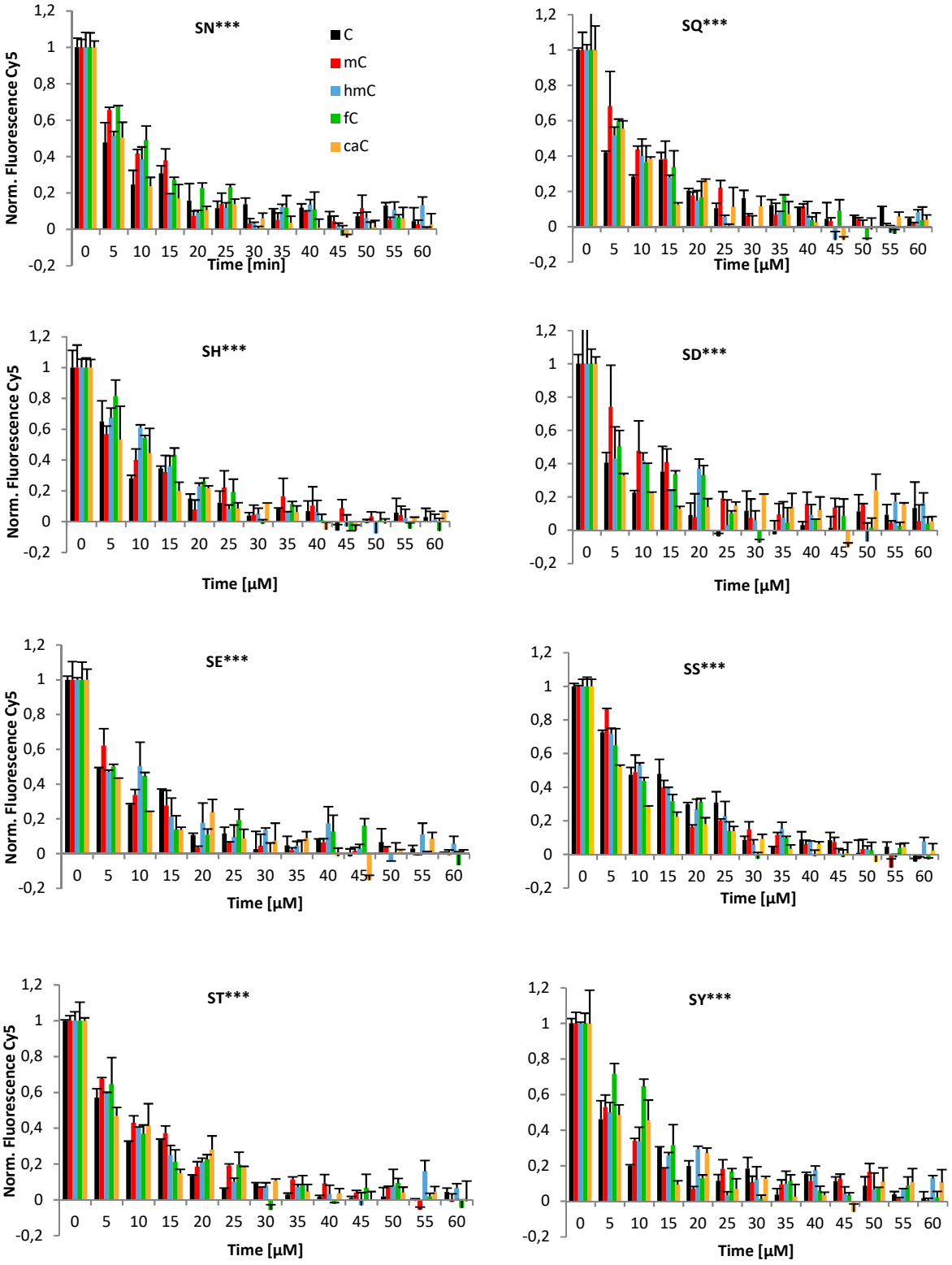


Figure 6-17 Context independence of mutant repeat R** for 5caC.** DNase I time course experiment with TALEs CDKN2A(18) containing RVDs HD, NN, or NG at position four and mutant RVD R**** at position five.

6.7.2 DNase I competition assay screening results for TALE mutant repeats from libraries NX*GG, NX**G, SX***, NX***, X***G, and X****.



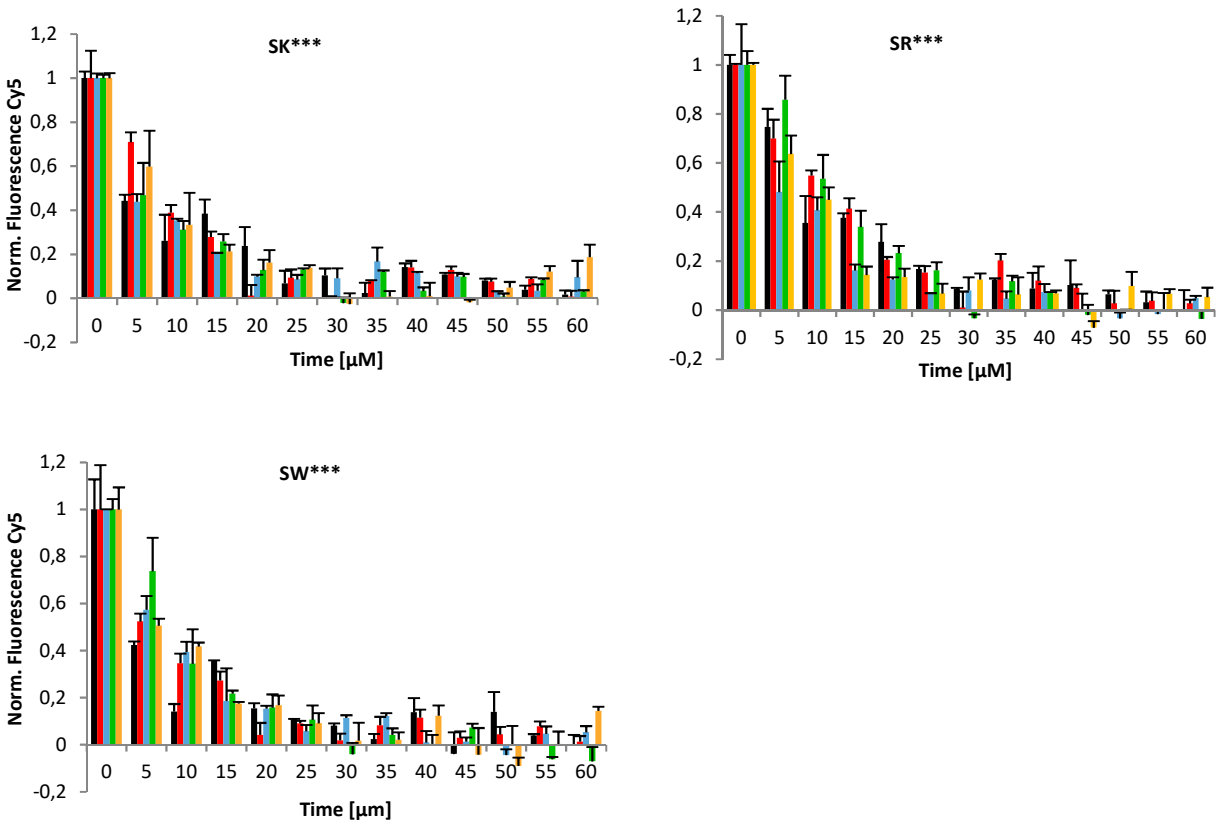
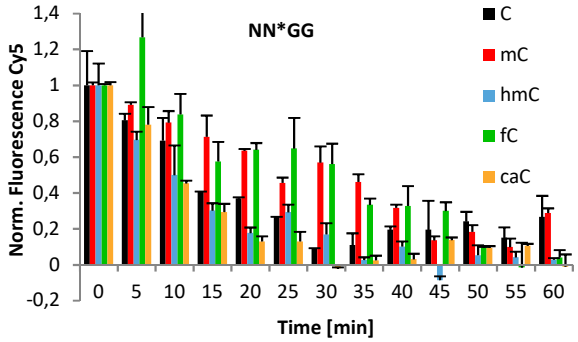
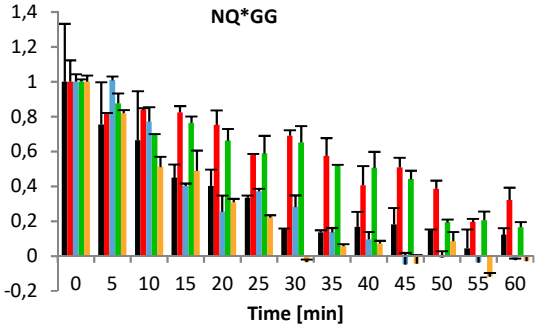
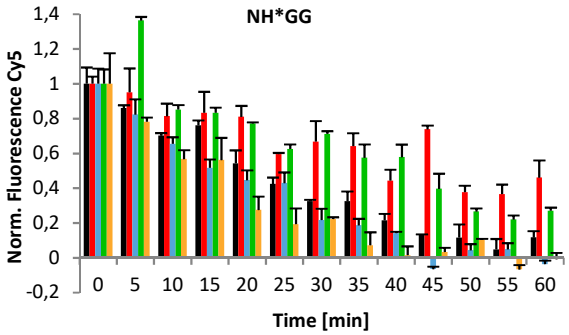
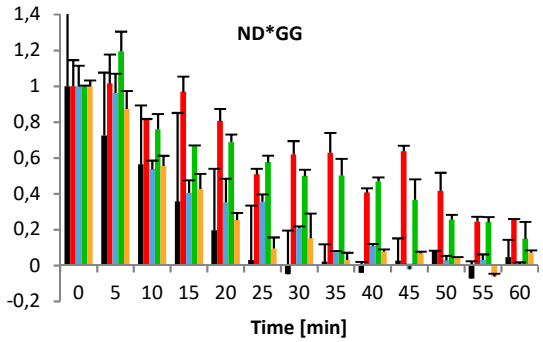


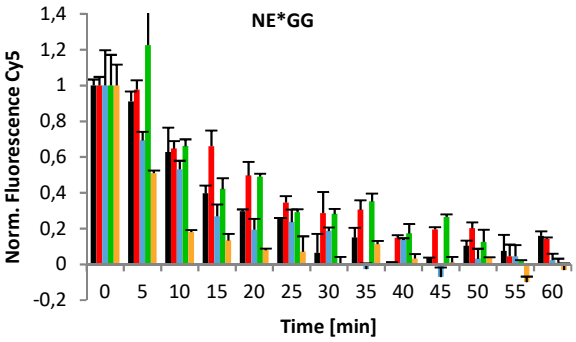
Figure 6-18 Screening results for library SX*.** DNase I time course experiment for library SX*** using 5µM TALE, 0.1 µM DNA, and 1 unit DNase I.

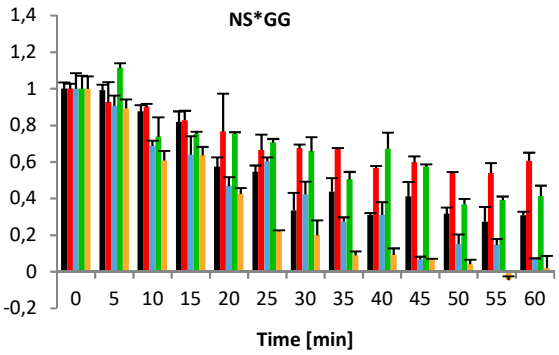

NN*GG
C
mC
hmC
fC
caC
Norm. Fluorescence Cy5
Time [min]

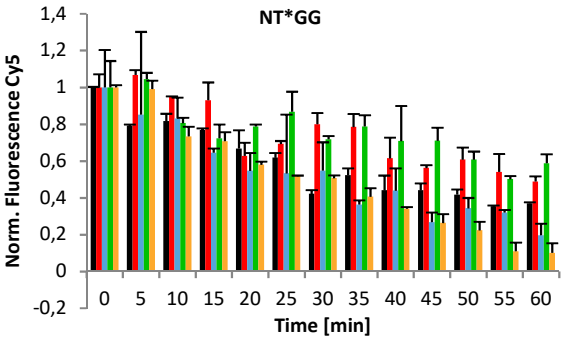

NQ*GG
Time [min]

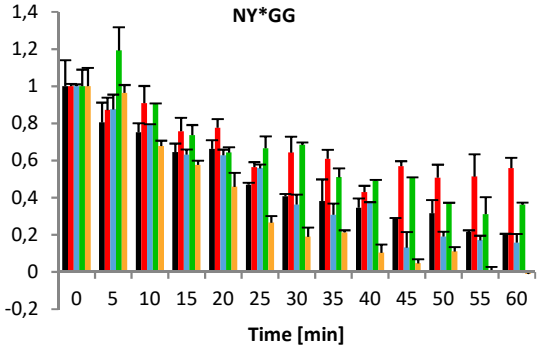

NH*GG
Norm. Fluorescence Cy5
Time [min]


ND*GG
Time [min]


NE*GG
Norm. Fluorescence Cy5
Time [min]


NS*GG
Time [min]


NT*GG
Norm. Fluorescence Cy5
Time [min]


NY*GG
Time [min]

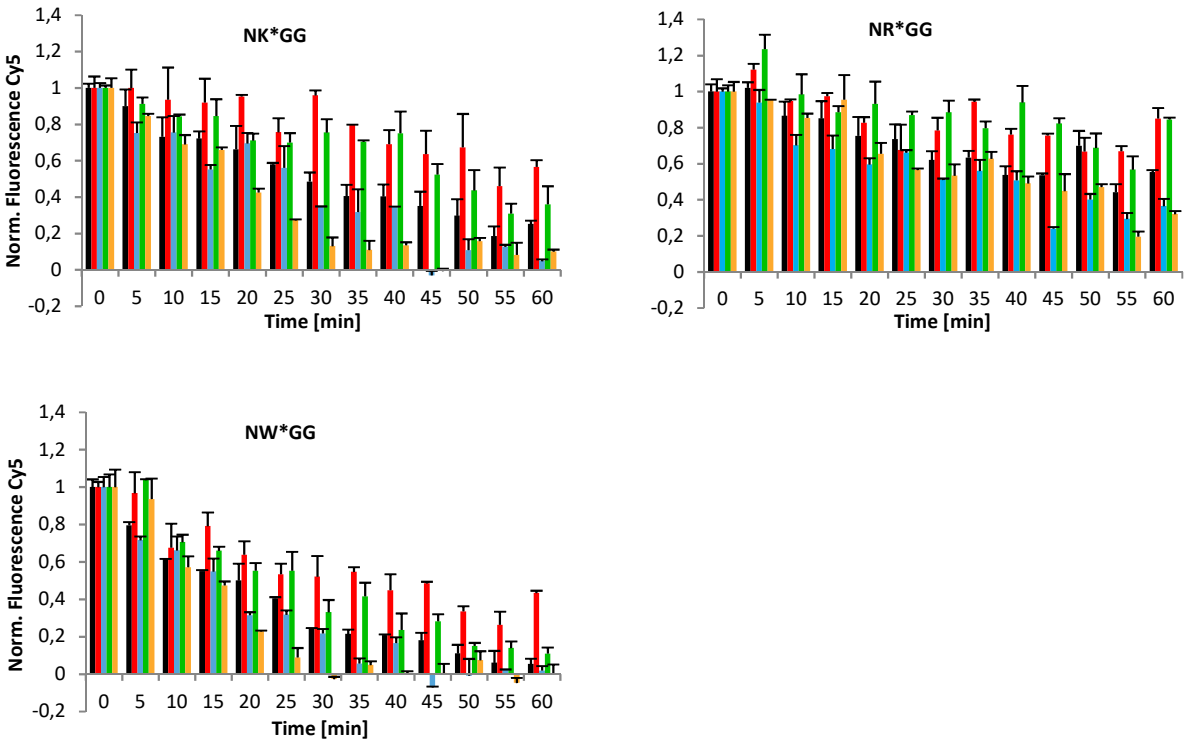
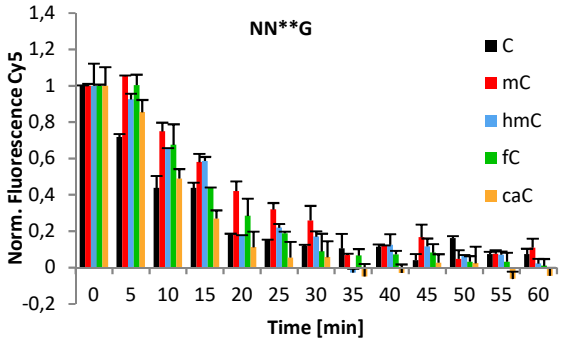
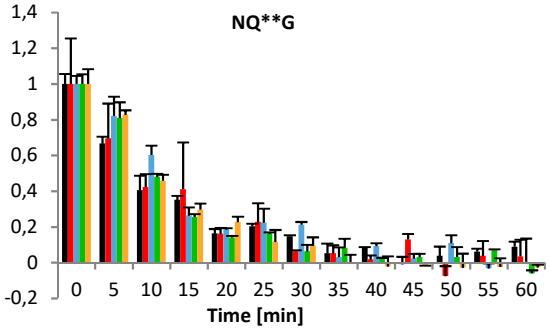
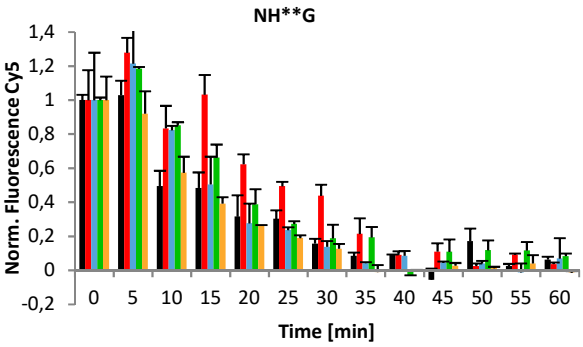
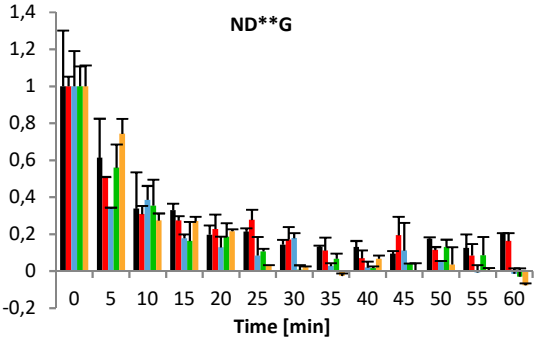


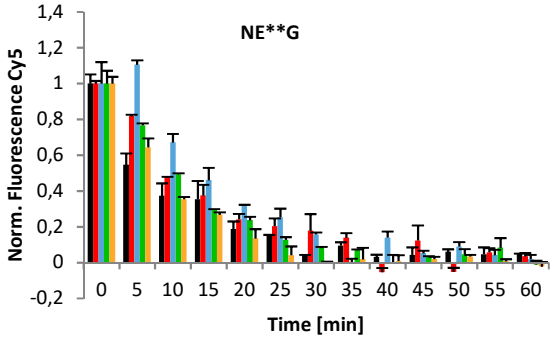
Figure 6-19 Screening results for library NX*GG. DNase I time course experiment for library NX*GG using 0.5 µM TALE, 0.1 µM DNA, and 1 unit DNase I.

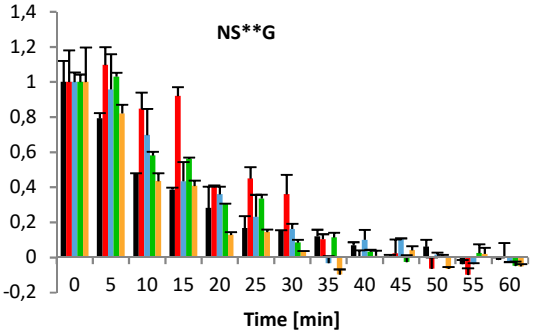

NN**G
C
mC
hmC
fC
caC
Norm. Fluorescence Cy5
Time [min]

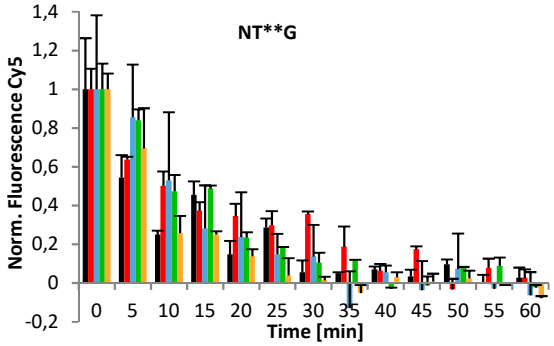

NQ**G
Time [min]

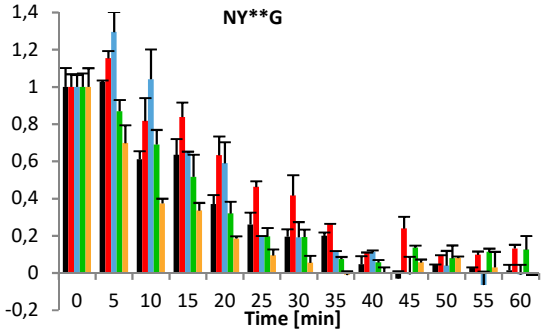

NH**G
Norm. Fluorescence Cy5
Time [min]


ND**G
Time [min]


NE**G
Norm. Fluorescence Cy5
Time [min]


NS**G
Time [min]


NT**G
Norm. Fluorescence Cy5
Time [min]


NY**G
Time [min]

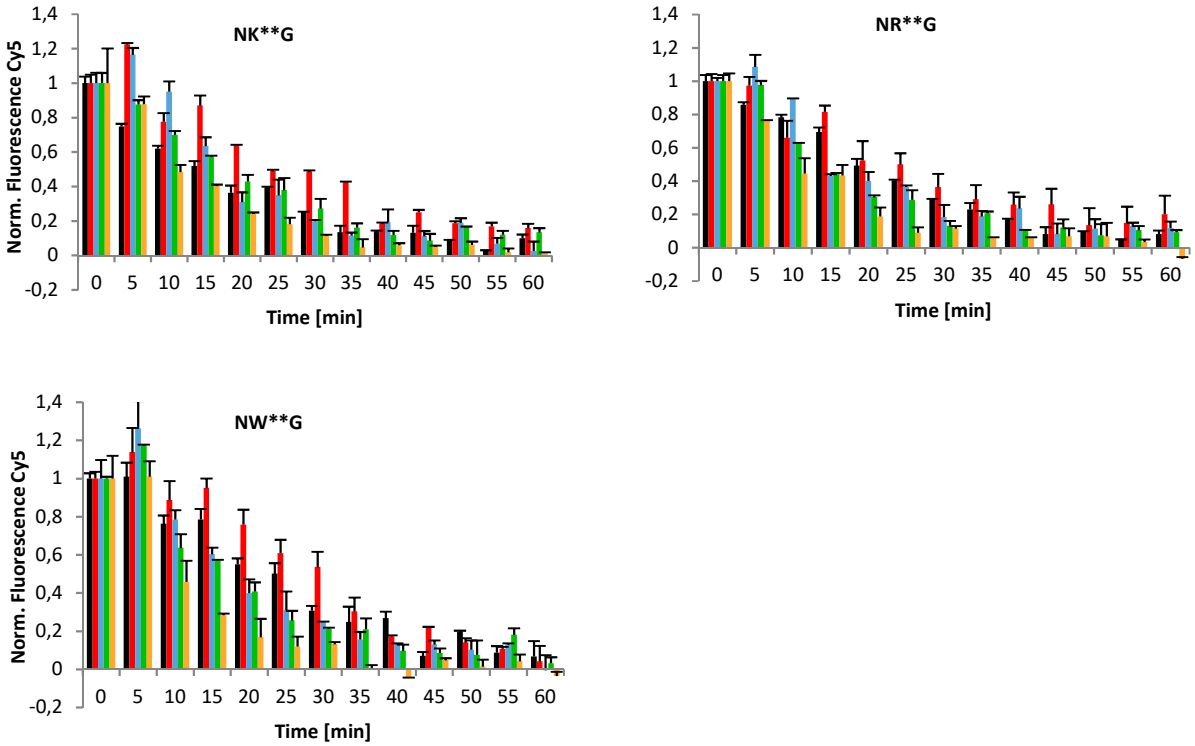
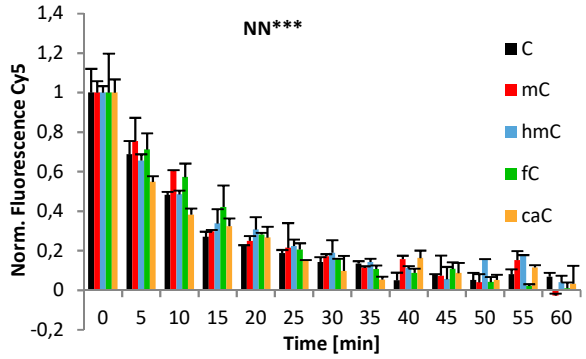
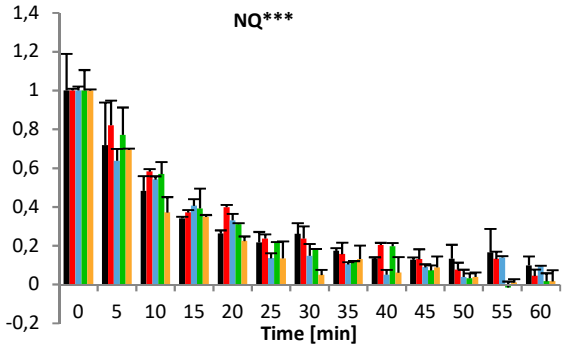
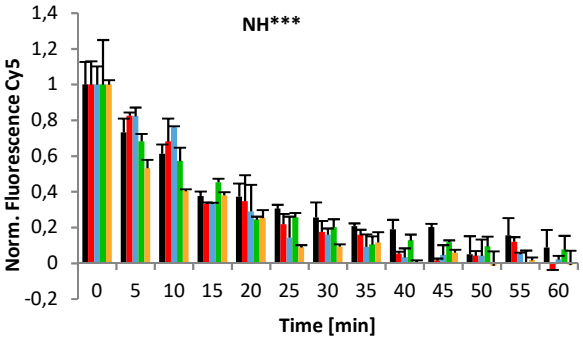
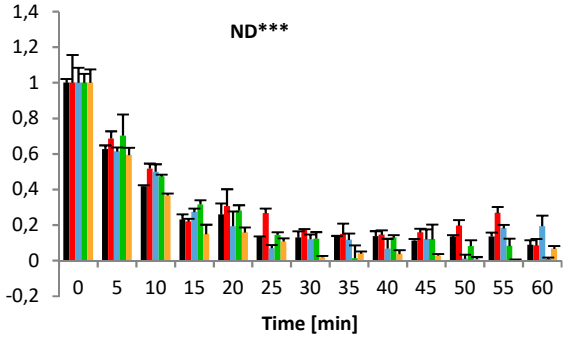


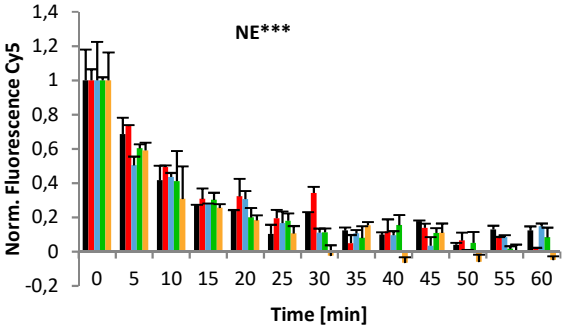
Figure 6-20 Screening results for library NXG.** DNase I time course experiment for library NX**G using 5 µM TALE, 0.1 µM DNA, and 1 unit DNase I.

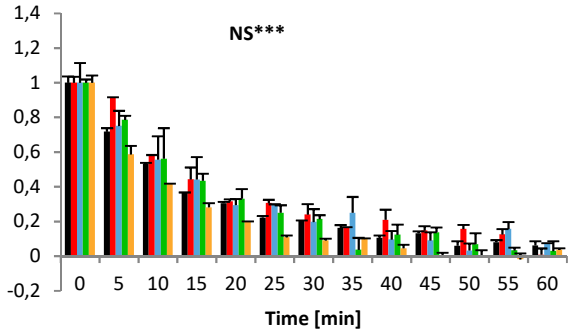

NN***
C
mC
hmC
fC
caC
Norm. Fluorescence Cy5
Time [min]

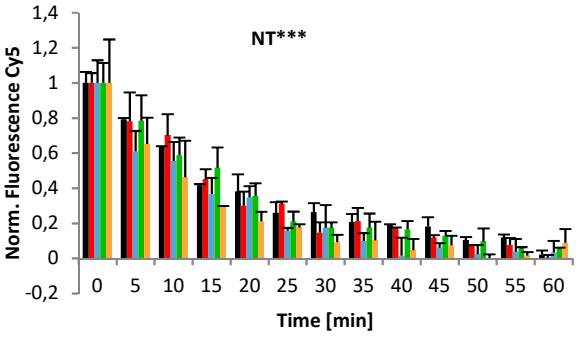

NQ***
Time [min]

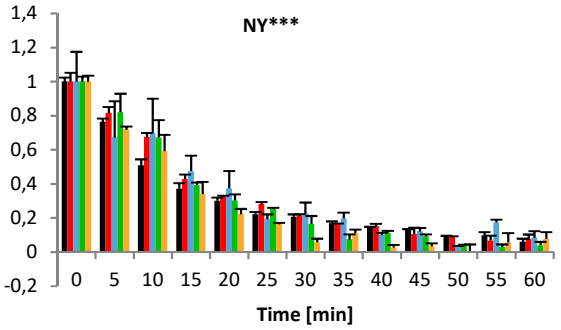

NH***
Norm. Fluorescence Cy5
Time [min]


ND***
Time [min]


NE***
Norm. Fluorescence Cy5
Time [min]


NS***
Time [min]


NT***
Norm. Fluorescence Cy5
Time [min]


NY***
Time [min]

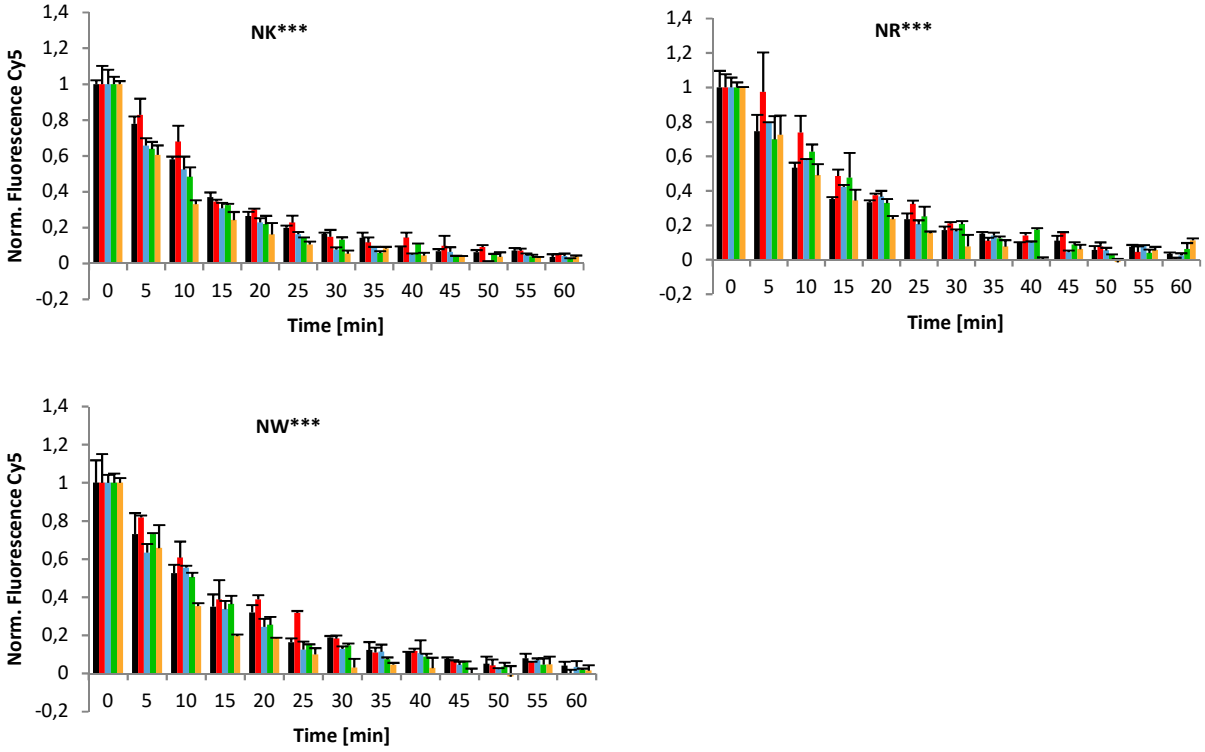
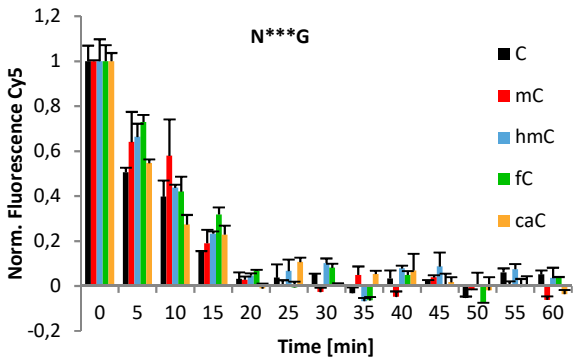
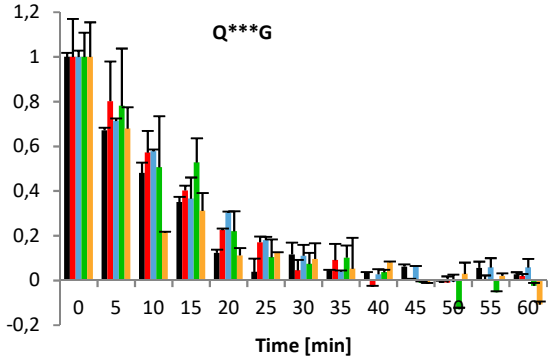
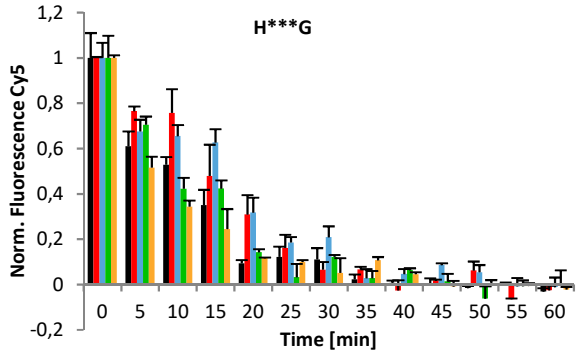
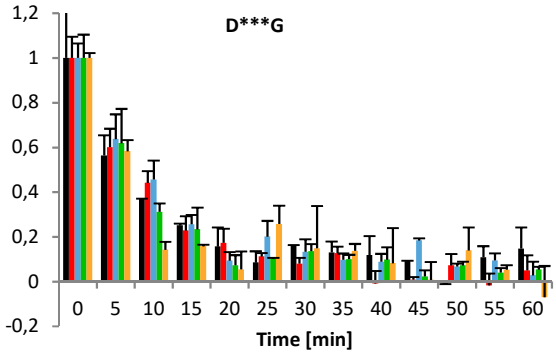


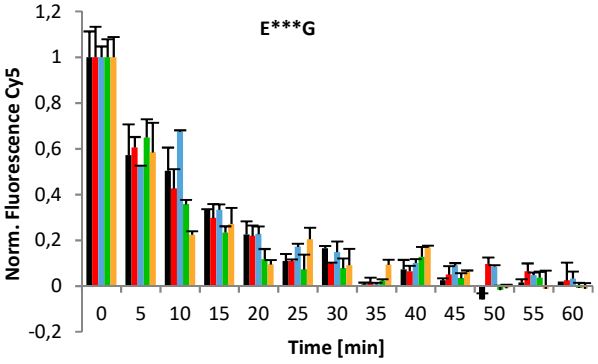
Figure 6-21 Screening results for library NX*.** DNase I time course experiment for library NX*** using 5 µM TALE, 0.1 µM DNA, and 1 unit DNase I.

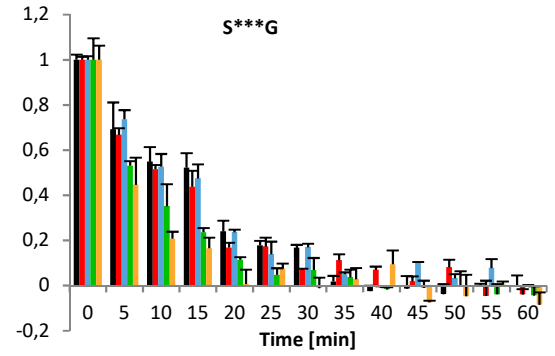

N***G
C
mC
hmC
fC
caC
Norm. Fluorescence Cy5
Time [min]

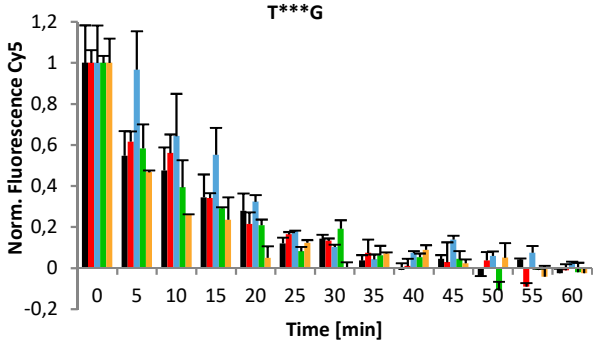

Q***G
Time [min]

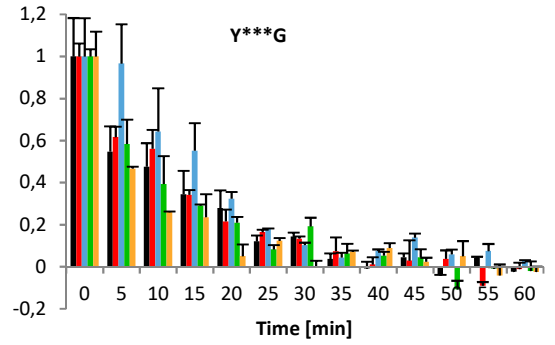

H***G
Norm. Fluorescence Cy5
Time [min]


D***G
Time [min]


E***G
Norm. Fluorescence Cy5
Time [min]


S***G
Time [min]


T***G
Norm. Fluorescence Cy5
Time [min]


Y***G
Time [min]

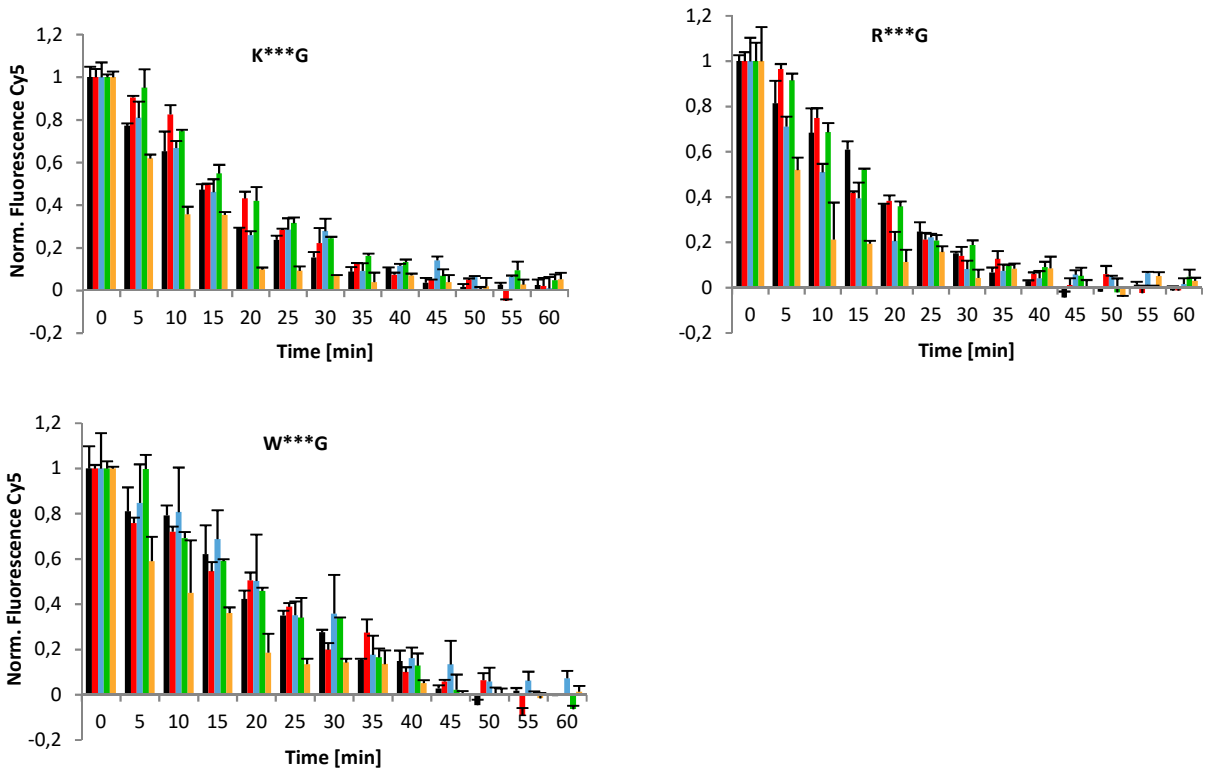
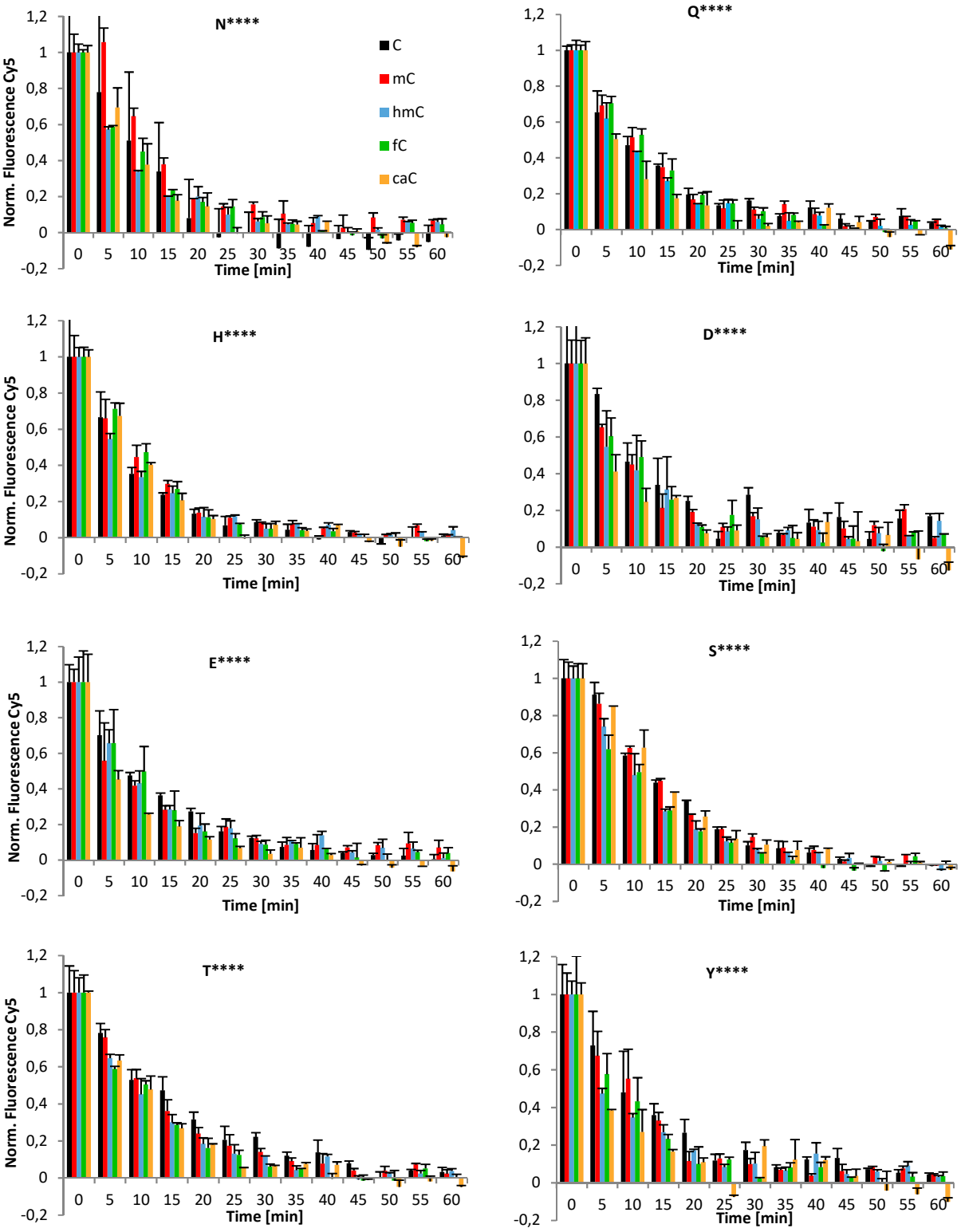


Figure 6-22 Screening results for library X*G.** DNase I time course experiment for library X***G using 5 µM TALE, 0.1 µM DNA, and 1 unit DNase I.


N****
Q****
H****
D****
E****
S****
T****
Y****
C
mC
hmC
fC
caC
Norm. Fluorescence Cy5
Time [min]
1,2
1
0,8
0,6
0,4
0,2
0
-0,2
0 5 10 15 20 25 30 35 40 45 50 55 60

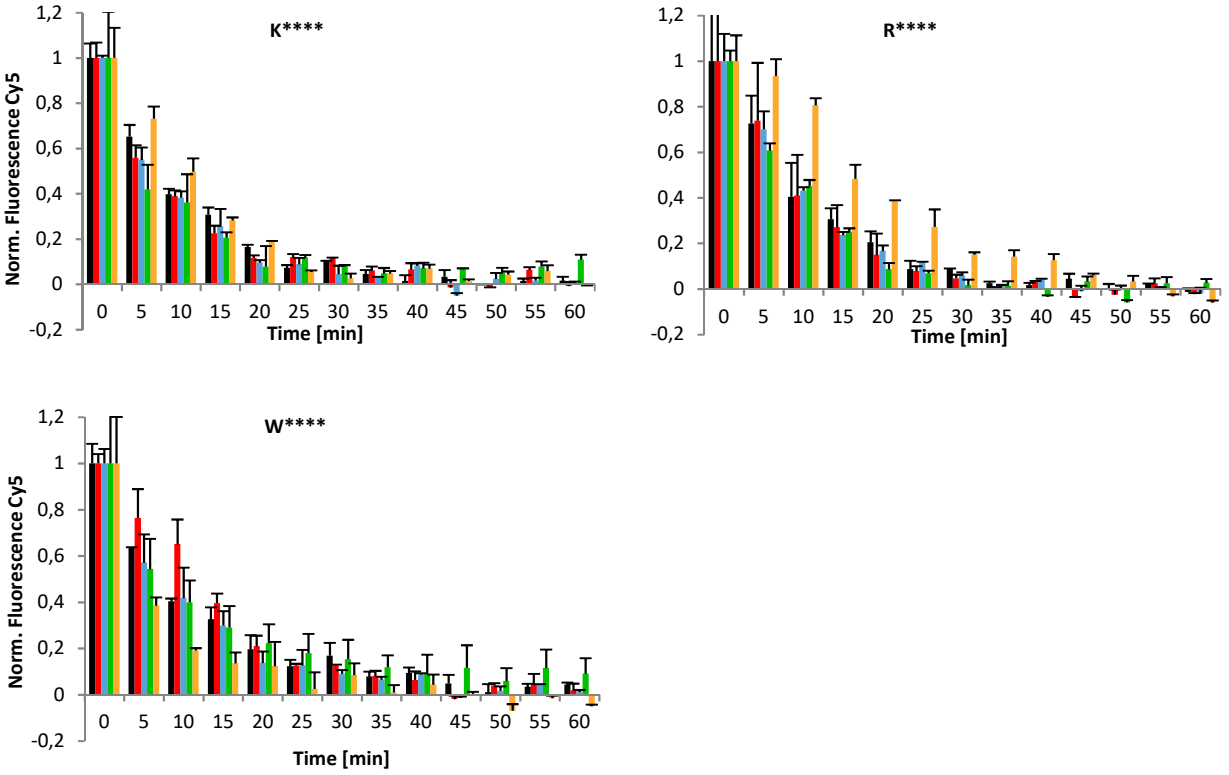


Figure 6-23 Screening results for library X**.** DNase I time course experiment for library X**** using 5 µM TALE, 0.1 µM DNA, and 1 unit DNase I.

7 References

[1] S. Maurer, B. Buchmuller, C. Ehrt, J. Jasper, O. Koch, D. Summerer, *Chemical Science*, **2018**, DOI: 10.1039/C8SC01958D.

[2] S. Maurer, M. Giess, O. Koch, D. Summerer, *ACS Chemical Biology* **2016**, *11*, 3294-3299.

[3] a) R. Holliday, J. Pugh, *Science* **1975**, *187*, 226-232; b) A. D. Riggs, *Cytogenetic and Genome Research* **1975**, *14*, 9-25.

[4] Z. D. Smith, A. Meissner, *Nature Reviews Genetics* **2013**, *14*, 204.

[5] a) A. Bird, *Genes & Development* **2002**, *16*, 6-21; b) J. A. Law, S. E. Jacobsen, *Nature reviews. Genetics* **2010**, *11*, 204-220.

[6] a) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930-935; b) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930.

[7] a) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300-1303; b) T. Pfaffeneder, B. Hackner, M. Truß, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angewandte Chemie International Edition* **2011**, *50*, 7008-7012.

[8] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*.

[9] a) M. J. Moscou, A. J. Bogdanove, *Science* **2009**, *326*, 1501; b) J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, U. Bonas, *Science* **2009**, *326*, 1509-1512.

[10] a) P. Rathi, S. Maurer, G. Kubik, D. Summerer, *Journal of the American Chemical Society* **2016**, *138*, 9910-9918; b) G. Kubik, M. J. Schmidt, J. E. Penner, D. Summerer, *Angewandte Chemie International Edition* **2014**, *53*, 6002-6006; c) D. Deng, P. Yin, C. Yan, X. Pan, X. Gong, S. Qi, T. Xie, M. Mahfouz, J.-K. Zhu, N. Yan, Y. Shi, *Cell Res* **2012**, *22*, 1502-1504; d) G. Kubik, S. Batke, D. Summerer, *Journal of the American Chemical Society* **2015**, *137*, 2-5.

[11] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737.

[12] A. J. B. Alberts, J. Lewis, et al., *Molecular Biology of the Cell*, 6th edition ed., Garland Publishing, New York, **2014**.

[13] J. L. T. J. M. Berg, G. J. Gatto, L. Stryer, *Biochemistry*, 8th edition ed., Springer Palgrave Macmillan, **2015**.

[14] L. Pray, *Vol. 1(1):100*, Nature Education, **2008**.

[15] R. R. Sinden, *DNA Structure and Function*, Academic Press, **1994**.

[16] H. R. Drew, R. M. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura, R. E. Dickerson, *Proceedings of the National Academy of Sciences* **1981**, *78*, 2179-2183.

[17] Y. G. Gao, H. Robinson, A. H. Wang, *European journal of biochemistry* **1999**, *261*, 413-420.

[18] T. Chatake, T. Sunami, *Journal of inorganic biochemistry* **2013**, *124*, 15-25.

[19] A. Rich, S. Zhang, *Nature Reviews Genetics* **2003**, *4*, 566.

[20] L. Armstrong, *Epigenetics*, Garland Science, **2014**.

[21] J. Bednar, R. A. Horowitz, S. A. Grigoryev, L. M. Carruthers, J. C. Hansen, A. J. Koster, C. L. Woodcock, *Proceedings of the National Academy of Sciences* **1998**, *95*, 14173-14178.

[22] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker, *Science* **2009**, *326*, 289-293.

[23] a) M. R. Hübner, M. A. Eckersley-Maslin, D. L. Spector, *Current opinion in genetics & development* **2013**, *23*, 89-95; b) Jesse R. Dixon, David U. Gorkin, B. Ren, *Molecular Cell*, *62*, 668-680.

[24] M. Meselson, F. W. Stahl, *Proceedings of the National Academy of Sciences of the United States of America* **1958**, *44*, 671-682.

[25] S. P. Bell, A. Dutta, *Annual Review of Biochemistry* **2002**, *71*, 333-374.

[26] J. L. Rinn, M. Kertesz, J. K. Wang, S. L. Squazzo, X. Xu, S. A. Brugmann, L. H. Goodnough, J. A. Helms, P. J. Farnham, E. Segal, H. Y. Chang, *Cell* **2007**, *129*, 1311-1323.

[27] S. Clancy, W. Brown*, Vol. 1(1):101*, Nature Education, **2008**.

[28] C. H. Waddington, *Endeavour* **1942**, *1*, 18-20.

[29] a) A. D. Riggs, R. A. Martienssen, R. V. E. A., *Cold Spring Harbor Laboratory Press* **1996**, *32*, 1-4; b) A. D. Riggs, T. N. Porter, *Cold Spring Harbor Laboratory Press* **1996**, *32*, 29-45.

[30] a) C. D. Allis, M.-L. Caparros, T. Jenuwein, D. Reinberg, *Epigenetics*, 2nd ed., Cold Spring Harbor Laboratory Press, **2015**; b) B. E. Bernstein, A. Meissner, E. S. Lander, *Cell* **2007**, *128*, 669-681.

[31] a) P. A. Jones, *Nat Rev Genet* **2012**, *13*, 484-492; bR. Jaenisch, A. Bird, *Nature Genetics* **2003**, *33*, 245.

[32] a) F. Larsen, G. Gundersen, R. Lopez, H. Prydz, *Genomics* **1992**, *13*, 1095-1107; b) J. Zhu, F. He, S. Hu, J. Yu, *Trends in Genetics*, *24*, 481-484.

[33] L. Z. Strichman-Almashanu, R. S. Lee, P. O. Onyango, E. Perlman, F. Flam, M. B. Frieman, A. P. Feinberg, *Genome Research* **2002**, *12*, 543-554.

[34] M. J. Fazzari, J. M. Greally, *Nature Reviews Genetics* **2004**, *5*, 446.

[35] M. G. Goll, T. H. Bestor, *Annual Review of Biochemistry* **2005**, *74*, 481-514.

[36] a) P. Hajkova, S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter, M. A. Surani, *Mechanisms of Development* **2002**, *117*, 15-23; b) W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, *403*, 501; c) J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, J. Walter, *Current Biology*, *10*, 475-478; d) H. Sasaki, Y. Matsui, *Nature Reviews Genetics* **2008**, *9*, 129.

[37] S. Feng, S. E Jacobsen, W. Reik, *Epigenetic Reprogramming in Plant and Animal Development, Vol. 330*, **2010**.

[38] L. Shen, C.-X. Song, C. He, Y. Zhang, *Annual Review of Biochemistry* **2014**, *83*, 585-614.

[39] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Z. h. a. n. g. Y, *Nature* **2010**, *466*.

[40] a) A. Maiti, A. C. Drohat, *J Biol Chem* **2011**, *286*; b) A. R. Weber, C. Krawczyk, A. B. Robertson, A. Kuśnierczyk, C. B. Vågbø, D. Schuermann, A. Klungland, P. Schär, *Nature Communications* **2016**, *7*, 10806.

[41] M. L. Álvaro, S. Daniel, *The Chemical Record* **2018**, *18*, 105-116.

[42] a) M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *Nat Chem Biol* **2015**, *11*, 555-557; b) M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *Nat Chem* **2014**, *6*, 1049-1055; c) M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmiller, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angewandte Chemie International Edition* **2010**, *49*, 5375-5377; d) D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLOS ONE* **2010**, *5*, e15367.

[43] 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.

[44] M. Wagner, J. Steinbacher, T. F. J. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell, *Angewandte Chemie International Edition* **2015**, *54*, 12511-12514.

[45] a) Cornelia G. Spruijt, F. Gnerlich, Arne H. Smits, T. Pfaffeneder, Pascal W. T. C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, Arie B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146-1159; b) M. Mellén, P. Ayata, S. Dewell, S. Kriaucionis, N. Heintz, *Cell* **2012**, *151*, 1417-1430; c) C. Frauer, T. Hoffmann, S. Bultmann, V. Casa, M. C. Cardoso, I. Antes, H. Leonhardt, *PLOS ONE* **2011**, *6*, e21306; d) O. Yildirim, R. Li, J.-H. Hung, Poshen B. Chen, X. Dong, L.-S. Ee, Z. Weng, Oliver J. Rando, Thomas G. Fazzio, *Cell* **2011**, *147*, 1498-1510.

[46] D. Condliffe, A. Wong, C. Troakes, P. Proitsi, Y. Patel, L. Chouliaras, C. Fernandes, J. Cooper, S. Lovestone, L. Schalkwyk, J. Mill, K. Lunnon, *Neurobiology of aging* **2014**, *35*, 1850-1854.

[47] a) C.-X. Song, Keith E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G.-L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678-691; b) L. Shen, H. Wu, D. Diep, S. Yamaguchi, Ana C. D'Alessio, H.-L. Fung, K. Zhang, Y. Zhang, *Cell* **2013**, *153*, 692-706; c) X. Lu, D. Han, B. S. Zhao, C.-X. Song, L.-S. Zhang, L. C. Dore, C. He, *Cell Res* **2015**, *25*, 386-389; d) F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello, S. Oliviero, *Cell Reports* **2015**, *10*, 674-683; e) M. Iurlaro, G. R. McInroy, H. E. Burgess, W. Dean, E.-A. Raiber, M. Bachman, D. Beraldi, S. Balasubramanian, W. Reik, *Genome Biology* **2016**, *17*, 141; f) X. Wu, A. Inoue, T. Suzuki, Y. Zhang, *Genes & Development* **2017**, *31*, 511-523; g) C. Zhu, Y. Gao, H. Guo, B. Xia, J. Song, X. Wu, H. Zeng, K. Kee, F. Tang, C. Yi, *Cell Stem Cell* **2017**, *20*, 720-731.e725.

[48] M. Su, A. Kirchner, S. Stazzoni, M. Müller, M. Wagner, A. Schröder, T. Carell, *Angewandte Chemie International Edition* **2016**, *55*, 11797-11800.

[49] a) M. Eleftheriou, A. J. Pascual, L. M. Wheldon, C. Perry, A. Abakir, A. Arora, A. D. Johnson, D. T. Auer, I. O. Ellis, S. Madhusudan, A. Ruzov, *Clinical Epigenetics* **2015**, *7*, 88; b) A. Ramsawhook, L. Lewis, B. Coyle, A. Ruzov, *Clinical Epigenetics* **2017**, *9*, 18.

[50] a) J. Xiong, Z. Zhang, J. Chen, H. Huang, Y. Xu, X. Ding, Y. Zheng, R. Nishinakamura, G.-L. Xu, H. Wang, S. Chen, S. Gao, B. Zhu, *Molecular Cell* **2016**, *64*, 913-925; b) H. Hashimoto, Y. O. Olanrewaju, Y. Zheng, G. G. Wilson, X. Zhang, X. Cheng, *Genes & Development* **2014**, *28*, 2304-2313; c) D. Wang, H. Hashimoto, X. Zhang, B. G. Barwick, S. Lonial, L. H. Boise, P. M. Vertino, X. Cheng, *Nucleic Acids Research* **2017**, *45*, 2396-2407.

[51] L. Wang, Y. Zhou, L. Xu, R. Xiao, X. Lu, L. Chen, J. Chong, H. Li, C. He, X.-D. Fu, D. Wang, *Nature* **2015**, *523*, 621-625.

[52] X. Wu, Y. Zhang, *Nature Reviews Genetics* **2017**, *18*, 517.

[53] a) E.-A. Raiber, D. Beraldi, S. Martínez Cuesta, G. R. McInroy, Z. Kingsbury, J. Becq, T. James, M. Lopes, K. Allinson, S. Field, S. Humphray, T. Santarius, C. Watts, D. Bentley, S. Balasubramanian, *npj Genomic Medicine* **2017**, *2*, 6; b) G. Pfeifer, W. Xiong, M. Hahn, S.-G. Jin, *Cell and Tissue Research* **2014**, *356*, 631-641; c) H. Mou, Z. Kennedy, D. G. Anderson, H. Yin, W. Xue, *Genome Medicine* **2015**, *7*, 53; d) Y.-H. Kim, D. Pierscianek, M. Mittelbronn, A. Vital, L. Mariani, M. Hasselblatt, H. Ohgaki, *Journal of Clinical Pathology* **2011**, *64*, 850-852.

[54] https://www.atdbio.com/content/56/Epigenetics. 09/08/2018.

[55] a) M. Bostick, J. K. Kim, Est, xe, P.-O. ve, A. Clark, S. Pradhan, S. E. Jacobsen, *Science* **2007**, *317*, 1760-1764; b) J. Sharif, M. Muto, S.-i. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano, H. Koseki, *Nature* **2007**, *450*, 908.

[56] M. Okano, S. Xie, E. Li, *Nature Genetics* **1998**, *19*, 219.

[57] M. S. Kareta, Z. M. Botello, J. J. Ennis, C. Chou, F. Chédin, *Journal of Biological Chemistry* **2006**, *281*, 25893-25902.

[58] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247-257.

[59] M. G. Goll, F. Kirpekar, K. A. Maggert, J. A. Yoder, C.-L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, T. H. Bestor, *Science* **2006**, *311*, 395-398.

[60] R. Z. Jurkowska, T. P. Jurkowski, A. Jeltsch, *ChemBioChem* **2011**, *12*, 206-222.

[61] a) A. Dhayalan, A. Rajavelu, P. Rathert, R. Tamas, R. Z. Jurkowska, S. Ragozin, A. Jeltsch, *Journal of Biological Chemistry* **2010**, *285*, 26114-26120; b) H. Wu, H. Zeng, R. Lam, W. Tempel, M. F. Amaya, C. Xu, L. Dombrovski, W. Qiu, Y. Wang, J. Min, *PLOS ONE* **2011**, *6*, e18919.

[62] a) D. Jia, R. Z. Jurkowska, X. Zhang, A. Jeltsch, X. Cheng, *Nature* **2007**, *449*, 248; b) S. K. T. Ooi, C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S.-P. Lin, C. D. Allis, X. Cheng, T. H. Bestor, *Nature* **2007**, *448*, 714.

[63] E. Li, Y. Zhang, *Cold Spring Harbor Perspectives in Biology* **2014**, *6*.

[64] a) R. R. Meehan, J. D. Lewis, S. McKay, E. L. Kleiner, A. P. Bird, *Cell* **1989**, *58*, 499-507; b) A. P. Bird, A. P. Wolffe, *Cell* **1999**, *99*, 451-454; c) B. Hendrich, A. Bird, *Molecular and Cellular Biology* **1998**, *18*, 6538-6547; d) G. J. P. Filion, S. Zhenilo, S. Salozhin, D. Yamada, E. Prokhortchouk, P.-A. Defossez, *Molecular and Cellular Biology* **2006**, *26*, 169-181; e) M. Unoki, T. Nishidate, Y. Nakamura, *Oncogene* **2004**, *23*, 7601.

[65] X. Nan, R. R. Meehan, A. Bird, *Nucleic Acids Research* **1993**, *21*, 4886-4892.

[66] a) M. Saito, F. Ishikawa, *Journal of Biological Chemistry* **2002**, *277*, 35434-35439; b) S. Laget, M. Joulie, F. Le Masson, N. Sasai, E. Christians, S. Pradhan, R. J. Roberts, P.-A. Defossez, *PLOS ONE* **2010**, *5*, e11982.

[67] Q. Du, P.-L. Luu, C. Stirzaker, S. J. Clark, *Epigenomics* **2015**, *7*, 1051-1073.

[68] a) X. Nan, H.-H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, A. Bird, *Nature* **1998**, *393*, 386; b) P. L. Jones, G. C. Jan Veenstra, P. A. Wade, D. Vermaak, S. U. Kass, N. Landsberger, J. Strouboulis, A. P. Wolffe, *Nature Genetics* **1998**, *19*, 187.

[69] W. A. Pastor, L. Aravind, A. Rao, *Nature Reviews Molecular Cell Biology* **2013**, *14*, 341.

[70] a) H. Zhang, X. Zhang, E. Clark, M. Mulcahey, S. Huang, Y. G. Shi, *Cell Research* **2010**, *20*, 1390; b) N. Liu, M. Wang, W. Deng, C. S. Schmidt, W. Qin, H. Leonhardt, F. Spada, *PLOS ONE* **2013**, *8*, e62755; c) S.-G. Jin, Z.-M. Zhang, Thomas L. Dunwell, Matthew R. Harter, X. Wu, J. Johnson, Z. Li, J. Liu, Piroska E. Szabó, Q. Lu, G.-l. Xu, J. Song, Gerd P. Pfeifer, *Cell Reports* **2016**, *14*, 493-505.

[71] L. Hu, J. Lu, J. Cheng, Q. Rao, Z. Li, H. Hou, Z. Lou, L. Zhang, W. Li, W. Gong, M. Liu, C. Sun, X. Yin, J. Li, X. Tan, P. Wang, Y. Wang, D. Fang, Q. Cui, P. Yang, C. He, H. Jiang, C. Luo, Y. Xu, *Nature* **2015**, *527*, 118.

[72] a) L. Lercher, M. A. McDonough, A. H. El-Sagheer, A. Thalhammer, S. Kriaucionis, T. Brown, C. J. Schofield, *Chemical Communications* **2014**, *50*, 1794-1796; b) D. Renciuk, O. Blacque, M. Vorlickova, B. Spingler, *Nucleic Acids Research* **2013**, *41*, 9891-9900; c) 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.

[73] A. Maiti, A. Z. Michelson, C. J. Armwood, J. K. Lee, A. C. Drohat, *Journal of the American Chemical Society* **2013**, *135*, 15813-15822.

[74] M. Sumino, A. Ohkubo, H. Taguchi, K. Seio, M. Sekine, *Bioorganic & Medicinal Chemistry Letters* **2008**, *18*, 274-277.

[75] E.-A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi, S. Balasubramanian, *Nat Struct Mol Biol* **2015**, *22*, 44-49.

[76] J. S. Hardwick, D. Ptchelkine, A. H. El-Sagheer, I. Tear, D. Singleton, S. E. V. Phillips, A. N. Lane, T. Brown, *Nature Structural &Amp; Molecular Biology* **2017**, *24*, 544.

[77] T. T. M. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, T. Ha, *Nature Communications* **2016**, *7*, 10813.

[78] a) K. B. Geahigan, G. A. Meints, M. E. Hatcher, J. Orban, G. P. Drobny, *Biochemistry* **2000**, *39*, 4939-4946; b) U. Mirsaidov, W. Timp, X. Zou, V. Dimitrov, K. Schulten, A. P. Feinberg, G. Timp, *Biophysical Journal* **2009**, *96*, L32-L34; c) S. Derreumaux, M. Chaoui, G. Tevanian, S. Fermandjian, *Nucleic Acids Research* **2001**, *29*, 2314-2326.

[79] A. Pérez, Chiara L. Castellazzi, F. Battistini, K. Collinet, O. Flores, O. Deniz, Maria L. Ruiz, D. Torrents, R. Eritja, M. Soler-López, M. Orozco, *Biophysical Journal* **2012**, *102*, 2140-2148.

[80] Thuy T. M. Ngo, Q. Zhang, R. Zhou, Jaya G. Yodh, T. Ha, *Cell* **2015**, *160*, 1135-1144.

[81] a) M. Martin, L. Ulrike, S. Dimitrios, P. Toni, G. F. A., D. C. A., K. S. C., K. Konstantin, C. Thomas, *Chemistry – A European Journal* **2011**, *17*, 13782-13788; b) A. Burdzy, K. T. Noyes, V. Valinluck, L. C. Sowers, *Nucleic Acids Research* **2002**, *30*, 4068-4074.

[82] L. S. Pidugu, J. W. Flowers, C. T. Coey, E. Pozharski, M. M. Greenberg, A. C. Drohat, *Biochemistry* **2016**, *55*, 6205-6208.

[83] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G.-L. Xu, C. Luo, H. Jiang, C. He, *Nat Chem Biol* **2012**, *8*, 328-330.

[84] 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.

[85] Y. Huang, W. A. Pastor, Y. Shen, M. Tahiliani, D. R. Liu, A. Rao, *PLOS ONE* **2010**, *5*, e8888.

[86] M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934-937.

[87] M. Yu, Gary C. Hon, Keith E. Szulwach, C.-X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J.-H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368-1380.

[88] M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nat Chem* **2014**, *6*, 435-440.

[89] X. Lu, D. Han, B. S. Zhao, C.-X. Song, L.-S. Zhang, L. C. Doré, C. He, *Cell Research* **2015**, *25*, 386.

[90] K. Tanaka, A. Okamoto, *Bioorganic & Medicinal Chemistry Letters* **2007**, *17*, 1912-1915.

[91] A. I. Bernstein, P. Jin, in *Epigenetic Technological Applications* (Ed.: Y. G. Zheng), Academic Press, Boston, **2015**, pp. 39-53.

[92] a) B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, S. W. Turner, *Nat Meth* **2010**, *7*, 461-465; b) Z. Feng, G. Fang, J. Korlach, T. Clark, K. Luong, X. Zhang, W. Wong, E. Schadt, *PLOS Computational Biology* **2013**, *9*, e1002935.

[93] M. J. Levene, J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, W. W. Webb, *Science* **2003**, *299*, 682-686.

[94] 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, M. Tahiliani, Y. Wang, T. A. Clark, J. Korlach, P. J. Pukkila, L. Aravind, A. Rao, *Proceedings of the National Academy of Sciences* **2014**, *111*, E5149-E5158.

[95] a) M. Wanunu, D. Cohen-Karni, R. R. Johnson, L. Fields, J. Benner, N. Peterman, Y. Zheng, M. L. Klein, M. Drndic, *Journal of the American Chemical Society* **2011**, *133*, 486-492; b) 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; c) 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; d) Z. L. Wescoe, J. Schreiber, M. Akeson, *Journal of the American Chemical Society* **2014**, *136*, 16582-16587.

[96] J. T. Simpson, R. E. Workman, P. C. Zuzarte, M. David, L. J. Dursi, W. Timp, *Nature Methods* **2017**, *14*, 407.

[97] M. Weber, J. J. Davies, D. Wittig, E. J. Oakeley, M. Haase, W. L. Lam, D. Schübeler, *Nature Genetics* **2005**, *37*, 853.

[98] a) H. Wu, A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, *Genes & Development* **2011**, *25*, 679-684; b) H. Stroud, S. Feng, S. Morey Kinney, S. Pradhan, S. E. Jacobsen, *Genome Biology* **2011**, *12*, R54; c) K. E. Szulwach, X. Li, Y. Li, C.-X. Song, J. W. Han, S. Kim, S. Namburi, K. Hermetz, J. J. Kim, M. K. Rudd, Y.-S. Yoon, B. Ren, C. He, P. Jin, *PLOS Genetics* **2011**, *7*, e1002154; d) S.-G. Jin, X. Wu, A. X. Li, G. P. Pfeifer, *Nucleic Acids Research* **2011**, *39*, 5015-5024; e) G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*; f) Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang, Yujiang G. Shi, *Molecular Cell* **2011**, *42*, 451-464.

[99] a) H. Wu, Y. Zhang, *Genes & Development* **2011**, *25*, 2436-2452; b) K. Williams, J. Christensen, K. Helin, *EMBO reports* **2012**, *13*, 28-35; c) M. R. Branco, G. Ficz, W. Reik, *Nature Reviews Genetics* **2011**, *13*, 7.

[100] W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*.

[101] C.-X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C.-H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat Biotech* **2011**, *29*, 68-72.

[102] J. Miller, A. D. McLachlan, A. Klug, *The EMBO Journal* **1985**, *4*, 1609-1614.

[103] M. Elrod-Erickson, M. A. Rould, L. Nekludova, C. O. Pabo, *Structure (London, England : 1993)* **1996**, *4*, 1171-1180.

[104] a) Y. Nakaseko, D. Neuhaus, A. Klug, D. Rhodes, *Journal of Molecular Biology* **1992**, *228*, 619-636; b) D. Neuhaus, Y. Nakaseko, J. W. R. Schwabe, A. Klug, *Journal of Molecular Biology* **1992**, *228*, 637-651.

[105] Q. Liu, D. J. Segal, J. B. Ghiara, C. F. Barbas, *Proceedings of the National Academy of Sciences* **1997**, *94*, 5525-5530.

[106] a) Y. Choo, A. Klug, *Proceedings of the National Academy of Sciences* **1994**, *91*, 11163-11167; b) A. C. Jamieson, S.-H. Kim, J. A. Wells, *Biochemistry* **1994**, *33*, 5689-5695; c) H. Wu, W. P. Yang, C. F. Barbas, *Proceedings of the National Academy of Sciences* **1995**, *92*, 344-348; d) E. Rebar, C. Pabo, *Science* **1994**, *263*, 671-673; e) D. J. Segal, B. Dreier, R. R. Beerli, C. F. Barbas, *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 2758-2763; f) R. R. Beerli, C. F. Barbas Iii, *Nature Biotechnology* **2002**, *20*, 135.

[107] a) A. C. Jamieson, J. C. Miller, C. O. Pabo, *Nature Reviews Drug Discovery* **2003**, *2*, 361; b) D. Carroll, *Genetics* **2011**, *188*, 773-782; c) R. R. Beerli, D. J. Segal, B. Dreier, C. F. Barbas, *Proceedings of the National Academy of Sciences* **1998**, *95*, 14628-14633; d) R. R. Beerli, B. Dreier, C. F. Barbas, *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 1495-1500.

[108] a) J. R. Porter, S. H. Lockwood, D. J. Segal, I. Ghosh, in *Engineered Zinc Finger Proteins: Methods and Protocols* (Eds.: J. P. Mackay, D. J. Segal), Humana Press, Totowa, NJ, **2010**, pp. 365-382; b) C. I. Stains, J. L. Furman, D. J. Segal, I. Ghosh, *Journal of the American Chemical Society* **2006**, *128*, 9761-9765; c) C. Lungu, S. Pinter, J. Broche, P. Rathert, A. Jeltsch, *Nature Communications* **2017**, *8*, 649; dY. Liu, Y. O. Olanrewaju, X. Zhang, X. Cheng, *Biochemistry* **2013**, *52*, 9310-9317.

[109] a) S. Kay, U. Bonas, *Current Opinion in Microbiology* **2009**, *12*, 37-43; b) S. Schornack, A. Meyer, P. Römer, T. Jordan, T. Lahaye, *Journal of Plant Physiology* **2006**, *163*, 256-272; c) F. F. White, B. Yang, *Plant Physiology* **2009**, *150*, 1677-1686.

[110] E. Weber, T. Ojanen-Reuhs, E. Huguet, G. Hause, M. Romantschuk, T. K. Korhonen, U. Bonas, R. Koebnik, *Journal of Bacteriology* **2005**, *187*, 2458-2468.

[111] a) U. Bonas, R. E. Stall, B. Staskawicz, *Molecular and General Genetics MGG* **1989**, *218*, 127-136; b) P. Römer, S. Hahn, T. Jordan, T. Strauß, U. Bonas, T. Lahaye, *Science* **2007**, *318*, 645-648; cS. Kay, S. Hahn, E. Marois, G. Hause, U. Bonas, *Science* **2007**, *318*, 648-651.

[112] J. Boch, U. Bonas, *Annual Review of Phytopathology* **2010**, *48*, 419-436.

[113] a) A. N.-S. Mak, P. Bradley, R. A. Cernadas, A. J. Bogdanove, B. L. Stoddard, *Science* **2012**, *335*, 716-719; b) D. Deng, C. Yan, X. Pan, M. Mahfouz, J. Wang, J.-K. Zhu, Y. Shi, N. Yan, *Science* **2012**, *335*, 720-723.

[114] H. Gao, X. Wu, J. Chai, Z. Han, *Cell Res* **2012**, *22*, 1716-1720.

[115] a) 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; b) J. F. Meckler, M. S. Bhakta, M.-S. Kim, R. Ovadia, C. H. Habrian, A. Zykovich, A. Yu, S. H. Lockwood, R. Morbitzer, J. Elsäesser, T. Lahaye, D. J. Segal, E. P. Baldwin, *Nucleic Acids Research* **2013**, *41*, 4118-4128; c) L. Cong, R. Zhou, Y.-c. Kuo, M. Cunniff, F. Zhang, *Nat Commun* **2012**, *3*, 968; d) 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, G. P. Dulay, K. L. Hua, I. Ankoudinova, G. J. Cost, F. D. Urnov, H. S. Zhang, M. C. Holmes, L. Zhang, P. D. Gregory, E. J. Rebar, *Nat Biotech* **2011**, *29*, 143-148.

[116] B. M. Lamb, A. C. Mercer, I. I. I. C. F. Barbas, *Nucleic Acids Research* **2013**, *41*, 9779-9785.

[117] a) T. Schreiber, U. Bonas, *Nucleic Acids Research* **2014**, *42*, 7160-7169; b) B. P. Hubbard, A. H. Badran, J. A. Zuris, J. P. Guilinger, K. M. Davis, L. Chen, S. Q. Tsai, J. D. Sander, J. K. Joung, D. R. Liu, *Nat Meth* **2015**, *advance online publication*.

[118] a) J. Yang, Y. Zhang, P. Yuan, Y. Zhou, C. Cai, Q. Ren, D. Wen, C. Chu, H. Qi, W. Wei, *Cell Research* **2014**, *24*, 628; b) J. C. Miller, L. Zhang, D. F. Xia, J. J. Campo, I. V. Ankoudinova, D. Y. Guschin, J. E. Babiarz, X. Meng, S. J. Hinkley, S. C. Lam, D. E. Paschon, A. I. Vincent, G. P. Dulay, K. A. Barlow, D. A. Shivak, E. Leung, J. D. Kim, R. Amora, F. D. Urnov, P. D. Gregory, E. J. Rebar, *Nat Meth* **2015**, *12*, 465-471; c) A. Juillerat, C. Pessereau, G. Dubois, V. Guyot, A. Marechal, J. Valton, F. Daboussi, L. Poirot, A. Duclert, P. Duchateau, *Sci. Rep.* **2015**, *5*.

[119] J. M. Rogers, L. A. Barrera, D. Reyon, J. D. Sander, M. Kellis, J. Keith Joung, M. L. Bulyk, *Nat Commun* **2015**, *6*.

[120] A. Garg, J. J. Lohmueller, P. A. Silver, T. Z. Armel, *Nucleic Acids Research* **2012**, *40*, 7584-7595.

[121] a) J. P. Guilinger, V. Pattanayak, D. Reyon, S. Q. Tsai, J. D. Sander, J. K. Joung, D. R. Liu, *Nature methods* **2014**, *11*, 429-435; b) F. C. Rinaldi, L. A. Doyle, B. L. Stoddard, A. J. Bogdanove, *Nucleic Acids Research* **2017**, *45*, 6960-6970.

[122] O. G. Berg, R. B. Winter, P. H. Von Hippel, *Biochemistry* **1981**, *20*, 6929-6948.

[123] P. C. Blainey, G. Luo, S. C. Kou, W. F. Mangel, G. L. Verdine, B. Bagchi, X. S. Xie, *Nature Structural &Amp; Molecular Biology* **2009**, *16*, 1224.

[124] a) L. Cuculis, Z. Abil, H. Zhao, C. M. Schroeder, *Nat Commun* **2015**, *6*; b) L. Cuculis, Z. Abil, H. Zhao, C. M. Schroeder, *Nat Chem Biol* **2016**, *advance online publication*.

[125] A. N.-S. Mak, P. Bradley, A. J. Bogdanove, B. L. Stoddard, *Current Opinion in Structural Biology* **2013**, *23*, 93-99.

[126] D. Reyon, S. Q. Tsai, C. Khayter, J. A. Foden, J. D. Sander, J. K. Joung, *Nature Biotechnology* **2012**, *30*, 460.

[127] J. L. Schmid-Burgk, T. Schmidt, V. Kaiser, K. Höning, V. Hornung, *Nature Biotechnology* **2012**, *31*, 76.

[128] N. Sun, H. Zhao, *Biotechnology and Bioengineering* **2013**, *110*, 1811-1821.

[129] a) R. Geißler, H. Scholze, S. Hahn, J. Streubel, U. Bonas, S.-E. Behrens, J. Boch, *PLOS ONE* **2011**, *6*, e19509; b) F. Zhang, L. Cong, S. Lodato, S. Kosuri, G. M. Church, P. Arlotta, *Nature Biotechnology* **2011**, *29*, 149.

[130] M. M. Mahfouz, L. Li, M. Piatek, X. Fang, H. Mansour, D. K. Bangarusamy, J.-K. Zhu, *Plant Molecular Biology* **2012**, *78*, 311-321.

[131] M. L. Maeder, J. F. Angstman, M. E. Richardson, S. J. Linder, V. M. Cascio, S. Q. Tsai, Q. H. Ho, J. D. Sander, D. Reyon, B. E. Bernstein, J. F. Costello, M. F. Wilkinson, J. K. Joung, *Nat Biotech* **2013**, *31*, 1137-1142.

[132] a) H. Ma, P. Reyes-Gutierrez, T. Pederson, *Proceedings of the National Academy of Sciences* **2013**, *110*, 21048-21053; b) Y. Miyanari, C. Ziegler-Birling, M.-E. Torres-Padilla, *Nature Structural &Amp; Molecular Biology* **2013**, *20*, 1321.

[133] K. Grzegorz, S. Daniel, *ChemBioChem* **2015**, *16*, 228-231.

[134] J. Valton, A. Dupuy, F. Daboussi, S. Thomas, A. Maréchal, R. Macmaster, K. Melliand, A. Juillerat, P. Duchateau, *Journal of Biological Chemistry* **2012**, *287*, 38427-38432.

[135] P. Rathi, S. Maurer, D. Summerer, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2017**.

[136] S. Tsuji, S. Futaki, M. Imanishi, *Chemical Communications* **2016**, *52*, 14238-14241.

[137] Y. Zhang, L. Liu, S. Guo, J. Song, C. Zhu, Z. Yue, W. Wei, C. Yi, *Nature Communications* **2017**, *8*, 901.

[138] J. Grau, M. Reschke, A. Erkes, J. Streubel, R. D. Morgan, G. G. Wilson, R. Koebnik, J. Boch, *Scientific Reports* **2016**, *6*, 21077.

[139] a) H. Wan, J.-p. Hu, K.-s. Li, X.-h. Tian, S. Chang, *PLOS ONE* **2013**, *8*, e76045; b) N. Tochio, K. Umehara, J.-i. Uewaki, H. Flechsig, M. Kondo, T. Dewa, T. Sakuma, T. Yamamoto, T. Saitoh, Y. Togashi, S.-i. Tate, *Scientific Reports* **2016**, *6*, 37887; c) H. Flechsig, *PLOS ONE* **2014**, *9*, e109919; d) B. I. M. Wicky, M. Stenta, M. Dal Peraro, *PLOS ONE* **2013**, *8*, e80261.

Eidesstattliche Versicherung (Affidavit)

_____ _____

Name, Vorname
(Surname, first name)

Matrikel-Nr.
(Enrolment number)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden, § 63 Abs. 5 Hochschulgesetz NRW.

Die Abgabe einer falschen Versicherung an Eides statt ist strafbar.

Wer vorsätzlich eine falsche Versicherung an Eides statt abgibt, kann mit einer Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft werden, § 156 StGB. Die fahrlässige Abgabe einer falschen Versicherung an Eides statt kann mit einer Freiheitsstrafe bis zu einem Jahr oder Geldstrafe bestraft werden, § 161 StGB.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offence can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offences of this type is the chancellor of the TU Dortmund University. In the case of multiple or other serious attempts at deception, the candidate can also be unenrolled, Section 63, paragraph 5 of the Universities Act of North Rhine-Westphalia.

The submission of a false affidavit is punishable.

Any person who intentionally submits a false affidavit can be punished with a prison sentence of up to three years or a fine, Section 156 of the Criminal Code. The negligent submission of a false affidavit can be punished with a prison sentence of up to one year or a fine, Section 161 of the Criminal Code.

I have taken note of the above official notification.

_____ _____

Ort, Datum
(Place, date)

Unterschrift
(Signature)

Titel der Dissertation:
(Title of the thesis):

_____

_____

_____

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Titel selbstständig und ohne unzulässige fremde Hilfe angefertigt habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht.
Die Arbeit hat in gegenwärtiger oder in einer anderen Fassung weder der TU Dortmund noch einer anderen Hochschule im Zusammenhang mit einer staatlichen oder akademischen Prüfung vorgelegen.

I hereby swear that I have completed the present dissertation independently and without inadmissible external support. I have not used any sources or tools other than those indicated and have identified literal and analogous quotations.

The thesis in its current version or another version has not been presented to the TU Dortmund University or another university in connection with a state or academic examination.*

*Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the PhD thesis is the official and legally binding version.

_____ _____

Ort, Datum
(Place, date)

Unterschrift
(Signature)