Electronic Supplementary Material (ESI) for Chemical Science.

Supporting Online Information

Overcoming Conservation in TALE-DNA Interactions: A Minimal Repeat Scaffold Enables Selective Recognition of an Oxidized 5-Methylcytosine

Sara Maurer, Benjamin Buchmuller, Christiane Ehrt, Julia Jasper, Oliver Koch and Daniel Summerer*

Faculty of Chemistry and Chemical Biology
TU Dortmund University
CP-02-143
Otto-Hahn-Str. 4a
44227 Dortmund
*Daniel.Summerer@tu-dortmund.de

Table of Contents

Experimental Section

Construction of Mutant TALE Module Plasmids

All TALE module plasmids with the respective mutations and deletions were cloned on TALE module plasmid pHD5 using site-directed cassette mutagenesis. pHD5 and the hybridized inserts with mutations (see List of Oligonucleotides) were digested using restriction enzymes *NcoI* and *XhoI* (New England Biolabs). Inserts were then ligated into digested pHD5 plasmid using T4 DNA ligase (New England Biolabs) and used for TALE assembly.

Golden Gate Assembly for TALE Libraries

TALE genes were assembled according to a previously published protocol[1] with library TALE module plasmids created via site-directed mutagenesis from TALE module plasmid pHD5. 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 five libraries with N-terminal GFP tag and C-terminal His6 tag.

Expression and Purification of TALE Proteins

For expression and purification of TALEs with GFP tag, overnight cultures of *E. coli* BL21(DE3) Gold transformed with a TALE expression plasmid grown in LB media supplemented with 50 µg/mL carbenicillin at 37 °C were diluted 50-fold into LB medium supplemented with the same antibiotic and were incubated at 37 °C and 200 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 1,400 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).

Primer Extension Assay

Reactions were performed as described previously[2]. Briefly, templates HEY2b and modified versions HEY2b_C6->5mC, 5hmC, 5fC, and 5caC (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) using a phosphorimager.

DNase I Competition Assay

Reactions were performed as described previously[3]. 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 $CaCl_2$) 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.

Evaluation of Natural TALE Repeat Variants

Amino acid sequences of known and/or predicted TAL effector (TALE) repeats to date were retrieved from Pfam version 31.0 (status February, 2018)[4] with annotations based on their match to the hidden Markov model (HMM) PF03377. The dataset included 12,712 alignments from NCBI and 6,271 alignments from UniProt covering TALE repeats of *Xanthomonadaceae*, *Burkholderiaceae*, *Legionellaceae*, and *Vibrionaceae*. This approach is similar to TALE prediction with software tools developed for *Xanthomonas* species.[5] The alignments were unified resulting in about 1,300 unique TALE repeat sequences (about 900 unique TALE sequences independent from the RVD) and analysed for truncation of the helix α domain (canonical residues 1–11) or deletions within or around the RVD (canonical residues 11–15) using custom computer code (Figure SI 4). Of 131 putatively truncated TALE repeats, we identified three predicted protein products of *Ralistonia solanacearum* and *Burkholderia rhizoxinica* and three protein products of *Xanthomonas spp.*, in which the initial 4 to 8 amino acids of the helix α domain were excluded by a potential start codon; whether or not these repeats are functional remains unclear. False positive truncations were due to atypical residues in the helix α domain or additional residues in between two TALE repeats causing misalignment of the HMM, which were checked manually. Likewise, we identified about 110 unique TALE repeats, in which one or more residues were missing within positions 11–15, in about 100 of these cases, a single residue of the RVD was deleted. The other cases comprised the truncated repeats mentioned before, and a single TALE-like protein of *Burkholderia rhizoxinica*, BAT,[6] in which both RVD residues were missing in 3 of 27 repeats. Although instances with multiple deletions were detected by the HMM, there was no natural TALE repeat variant lacking more than two residues within or around the RVD. Despite their evolutionary distance and dissimilarity in sequence composition, functional TALE repeats in *Xanthomonas spp*., *Ralistonia spp.*, and *Burkholderia spp.* seem to be conserved in length of their helix α domain.[7]

Homology Modeling

The modeling workflow was performed using MOE 2015.10[8] and Chimera.[9] As starting point, the crystal structure pdb 4gjr (chains A, I and J) was used for the model of R**** (CDK2NA sequence context) with DNA. The base pairs of the DNA strand were mutated to the corresponding base pairs of the used DNA sequence with Chimera. The respective cytosine was modified to 5-carboxycytosine in MOE, resulting in a DNA model for caC. The template structure was prepared with the “structure preparation” functionality and aligned to the target sequence. Subsequently, a homology model was created using the homology modeling module of MOE. Here, the option “include selected atoms as environment for

induced fit option" was enabled to transfer the modified DNA to the homology model and to consider it during model generation. Additionally, models of the DNA-unbound protein were generated for the wt TALE protein and for the S**** (CDKN2A sequence context) and R**** (CDKN2A and BRCA1 sequence context) variants with pdb 3v6p as template. In all cases, 25 main chain models with 3 side chain samples each were created and refined using the "Fine" setting with an RMS gradient of 0.1 kcal/(mol·Å²). Apart from that, standard settings were used together with the default force field Amber10:EHT.[10]

Molecular Dynamics (MD) Simulation

The homology models of the wt TALE construct, the S**** variant in the CDKN2A sequence context, and the R**** in the sequence contexts CDKN2A and BRCA1 were used as initial structures for the MD simulation analyses. MD simulations were performed using the AMBER16 simulation package.[11] The ff14SB force field[12] was applied for protein parameterization. The proteins were placed in a truncated octahedral box with TIP3P water molecules (extension of at least 12 Å of any protein atom) and the system was neutralized using Na+ ions. In all subsequently described steps, the SHAKE algorithm was applied to hydrogen atoms, and the cut-off for non-bonded interactions was set to 8 Å. The energy minimization and equilibration steps were performed using SANDER. The modeled structures were subjected to an energy minimization in two steps. At first, 50 steps of the Steepest Descent minimization algorithm were applied, followed by a maximum of 200 steps of Conjugate Gradient reaching an RMS gradient of 0.0001 kcal/(mol·Å²) with a harmonic positional restraint of 25 kcal/(mol·Å²) on all protein atoms. Subsequently, this procedure was repeated reducing the restraints to 5 kcal/(mol·Å²). The temperature was adjusted to 300 K using the Berendsen temperature coupling over 50 ps of NVT simulation retaining the protein restraint. Subsequently, 150 ps MD simulation of an NPT ensemble were performed at a constant isotropic pressure of 1 atm and a protein restraint of 5 kcal/(mol·Å²). The restraints were incrementally reduced over 50 ps of a NVT ensemble simulation and the last 10 ps of simulation were performed in the absence of any restraints. In the production phase, 250 ns simulations of an NVT ensemble were performed on GPU using PMEMD.[13] In this run, the temperature was set to 300 K using the Berendsen's weak coupling algorithm and the Particle Mesh Ewald method was applied for the treatment of long-range interactions. Atomic coordinates, energies, and temperatures were sampled every 20 ps.

Analysis of the resulting trajectories were performed using CPPTRAJ of the AMBER16 simulations package and VMD[14] for 12,500 frames as extracted from the production run. The clustering was performed using a hierarchical agglomerative approach with average linkage and an RMSD threshold of 3 Å in CPPTRAJ. Finally, a principal component analysis

was performed using the ProDy[15] interface of the Normal Mode Wizard (NMWiz) plugin in VMD. Based on the protein's Cα atoms, normal modes for the MD trajectories were derived.

List of Oligonucleotides

All Oligos were ordered from Sigma-Aldrich and Metabion. 5-(methyl) cytosine nucleotides are marked as **C**. 5-(hydroxymethyl) cytosine nucleotides are marked as **X**. 5-(formyl) cytosine nucleotides are marked as **Y**. 5-(carboxyl) cytosine nucleotides are marked as **Z**.

Oligonucleotides for Site-Directed Cassette Mutagenesis

	Modification	**Sequence**
o1680	VAIASN***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAAT AAGCAAGCGCTCGAAACGGTGCAGGAG
o1681	VAIASQ***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCCAG AAGCAAGCGCTCGAAACGGTGCAGGAG
o1682	VAIASH***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCCAT AAGCAAGCGCTCGAAACGGTGCAGGAG
o1683	VAIASD***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCGAT AAGCAAGCGCTCGAAACGGTGCAGGAG
o1684	VAIASE***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCGAA AAGCAAGCGCTCGAAACGGTGCAGGAG
o1685	VAIASS***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1686	VAIAST***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCACC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1687	VAIASY***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCTAT AAGCAAGCGCTCGAAACGGTGCAGGAG
o1688	VAIASK***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAAA AAGCAAGCGCTCGAAACGGTGCAGGAG
o1689	VAIASR***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCCGT AAGCAAGCGCTCGAAACGGTGCAGGAG
o1690	VAIASW***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCTGG AAGCAAGCGCTCGAAACGGTGCAGGAG
o1702	VAIANN**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAAT GGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1703	VAIANQ**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCAG GGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1704	VAIANH**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCAT GGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1705	VAIAND**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACGAT GGCAAGCAAGCGCTCGAAACGGTGCAGGAG

o1706	VAIANE**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACGAA GGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1707	VAIANS**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAGC GGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1708	VAIANT**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACACC GGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1709	VAIANY**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACTAT GGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1710	VAIANK**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAAA GGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1711	VAIANR**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCGT GGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1712	VAIANW**GG	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACTGG GGCAAGCAAGCGCTCGAAACGGTGCAGGAG
o1713	VAIANN***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAAT AAGCAAGCGCTCGAAACGGTGCAGGAG
o1714	VAIANN***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCAG AAGCAAGCGCTCGAAACGGTGCAGGAG
o1715	VAIANQ***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCAT AAGCAAGCGCTCGAAACGGTGCAGGAG
o1716	VAIANH***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACGAT AAGCAAGCGCTCGAAACGGTGCAGGAG
o1717	VAIAND***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACGAA AAGCAAGCGCTCGAAACGGTGCAGGAG
o1718	VAIANE***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1719	VAIANS***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACACC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1720	VAIANT***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACTATA AGCAAGCGCTCGAAACGGTGCAGGAG
o1721	VAIANK***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACAAA AAGCAAGCGCTCGAAACGGTGCAGGAG
o1722	VAIANR***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACCGT AAGCAAGCGCTCGAAACGGTGCAGGAG
o1723	VAIANW***	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAACTGG AAGCAAGCGCTCGAAACGGTGCAGGAG
o1724	VAIAN***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAATGGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1725	VAIAQ***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCAGGGC AAGCAAGCGCTCGAAACGGTGCAGGAG

o1726	VAIAH***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCATGGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1727	VAIAD***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCGATGGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1728	VAIAE***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCGAAGGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1729	VAIAS***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCGGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1730	VAIAT***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCACCGGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1731	VAIAY***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCTATGGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1732	VAIAK***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAAAGGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1733	VAIAR***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCGTGGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1734	VAIAW***G	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCTGGGGC AAGCAAGCGCTCGAAACGGTGCAGGAG
o1735	VAIAN****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAATAAG CAAGCGCTCGAAACGGTGCAGGAG
o1736	VAIAQ****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCAGAAG CAAGCGCTCGAAACGGTGCAGGAG
o1737	VAIAH****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCATAAG CAAGCGCTCGAAACGGTGCAGGAG
o1738	VAIAD****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCGATAAG CAAGCGCTCGAAACGGTGCAGGAG
o1739	VAIAE****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCGAAAAG CAAGCGCTCGAAACGGTGCAGGAG
o1740	VAIAS****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAGCAAG CAAGCGCTCGAAACGGTGCAGGAG
o1741	VAIAT****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCACCAAG CAAGCGCTCGAAACGGTGCAGGAG
o1742	VAIAY****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCTATAAG CAAGCGCTCGAAACGGTGCAGGAG
o1743	VAIAK****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCAAAAAG CAAGCGCTCGAAACGGTGCAGGAG
o1744	VAIAR****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCCGTAAG CAAGCGCTCGAAACGGTGCAGGAG
o1745	VAIAW****	TTTTCCATGGCCTGACCCCGGACCAAGTGGTGGCTATCGCCTGGAAG CAAGCGCTCGAAACGGTGCAGGAG

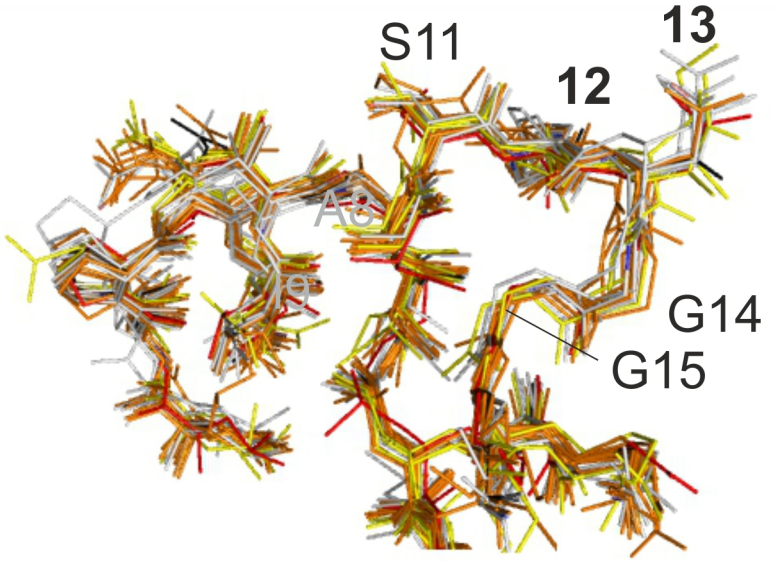
o1746	reverse primer	TTTTCTCGAGGGTCTCCTGCACCGTTTCGAGCGCTTG

Oligonucleotides for Primer Extension Assay and DNase I Competition Assay

		Sequence
o465	HEY2b_C6-->mC	TGGATTCCCACTCTTCAGCCCCAGCGTTACAGCATCTTCAG TGGCTTCTTCCACCGTGAGCTCTTCCGTTTCCACATCC
o466	radiolabeled primer for PEX	GGATGTGGAAACGGAAGA
o476	HEY2b_C6-->C	TGGATTCCCACTCTTCAGCCCCAGCGTTACAGCATCTTCAG TGGCTTCTTCCACCGTGAGCTCTTCCGTTTCCACATCC
o520	HEY2b_C6-->hmC	TGGATTCCCACTCTTCAGCCCCAGCGTTACAGCATCTTCAG TGGCTTCTTCCACCGTGAGCTCTTCXGTTTCCACATCC
o1422	HEY2b_C6-->fC	TGGATTCCCACTCTTCAGCCCCAGCGTTACAGCATCTTCAG TGGCTTCTTCCACCGTGAGCTCTTCYGTTTCCACATCC
o1423	HEY2b_C6-->caC	TGGATTCCCACTCTTCAGCCCCAGCGTTACAGCATCTTCAG TGGCTTCTTCCACCGTGAGCTCTTCZGTTTCCACATCC
o1892	HEY2b_5´Cy5, 3´Cy3 labeled reverse primer for FRET	GGATGTGGAAACGGAAGA
o1516	BRCA1(18)_C6→C	CTTCCTCTTCCGTCTCTTTCCTTTTACGTCATCCGGGGGCA GACT
o1518	BRCA1(18)_C6→mC	CTTCCTCTTCCGTCTCTTTCCTTTTACGTCATCCGGGGGCA GACT
o1521	BRCA1(18)_C6→hmC	CTTCCTCTTCCGTCTCTTTCCTTTTAXGTCATCCGGGGGCA GACT
o1524	BRCA1(18)_C6→fC	CTTCCTCTTCCGTCTCTTTCCTTTTAYGTCATCCGGGGGCA GACT
o1527	BRCA1(18)_C6→caC	CTTCCTCTTCCGTCTCTTTCCTTTTAZGTCATCCGGGGGCA GACT
o2504	BRCA1(18)_5´Cy5, 3´Cy3 labeled reverse primer for FRET	CCCCCGGATGACGTAAAA
o1591	CDKN2A(18)_C6→C	GGCCAGCCAGTCAGCCGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o1592	CDKN2A(18)_C6→mC	GGCCAGCCAGTCAGCCGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o1593	CDKN2A(18)_C6→hmC	GGCCAGCCAGTCAGCXGAAGGCTCCATGCTGCTCCCCGC CGCCGGC

o1594	CDKN2A(18)_C6→fC	GGCCAGCCAGTCAGCYGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o1595	CDKN2A(18)_C6→caC	GGCCAGCCAGTCAGCZGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o2503	CDKN2A(18)_5´Cy5, 3´Cy3 labeled reverse primer for FRET	CATGGAGCCTTCGGCTGA
o3098	CDKN2A(18)_C5→G, C6→C	GGCCAGCCAGTCAG**G**CGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o3099	CDKN2A(18)_C5→G, C6→mC	GGCCAGCCAGTCAG**G**CGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o3100	CDKN2A(18)_C5→G, C6→hmC	GGCCAGCCAGTCAG**G**XGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o3101	CDKN2A(18)_C5→G, C6→fC	GGCCAGCCAGTCAG**G**YGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o3102	CDKN2A(18)_C5→G, C6→caC	GGCCAGCCAGTCAG**G**ZGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o3103	CDKN2A(18)_5´Cy5, 3´Cy3 labeled reverse primer for FRET	CATGGAGCCTTCG**C**CTGA
o3104	CDKN2A(18)_C5→T, C6→C	GGCCAGCCAGTCAG**T**CGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o3105	CDKN2A(18)_C5→T, C6→mC	GGCCAGCCAGTCAG**T**CGAAGGCTCCATGCTGCTCCCCGC CGCCGGC
o3106	CDKN2A(18)_C5→T, C6→hmC	GGCCAGCCAGTCAG**T**XGAAGGCTCCATGCTGCTCCCCGCC GCCGGC
o3107	CDKN2A(18)_C5→T, C6→fC	GGCCAGCCAGTCAG**T**YGAAGGCTCCATGCTGCTCCCCGCC GCCGGC
o3108	CDKN2A(18)_C5→T, C6→caC	GGCCAGCCAGTCAG**T**ZGAAGGCTCCATGCTGCTCCCCGCC GCCGGC
o3109	CDKN2A(18)_5´Cy5, 3´Cy3 labeled reverse primer for FRET	CATGGAGCCTTCG**A**CTGA

Figure SI 1: Alignments of 24 TALE Repeats Shows Conserved Main and Side Chain Conformations



Alignments of 24 TALE repeats covering all standard RVDs from two crystal structures (pdb 3UGM and 3V6T, conducted with Pymol 1.3).[16] Repeats with RVD HD, NN, NI and NG are coloured grey, red, orange and yellow, respectively.

Figure SI 2: Map of Exemplary TALE Plasmid

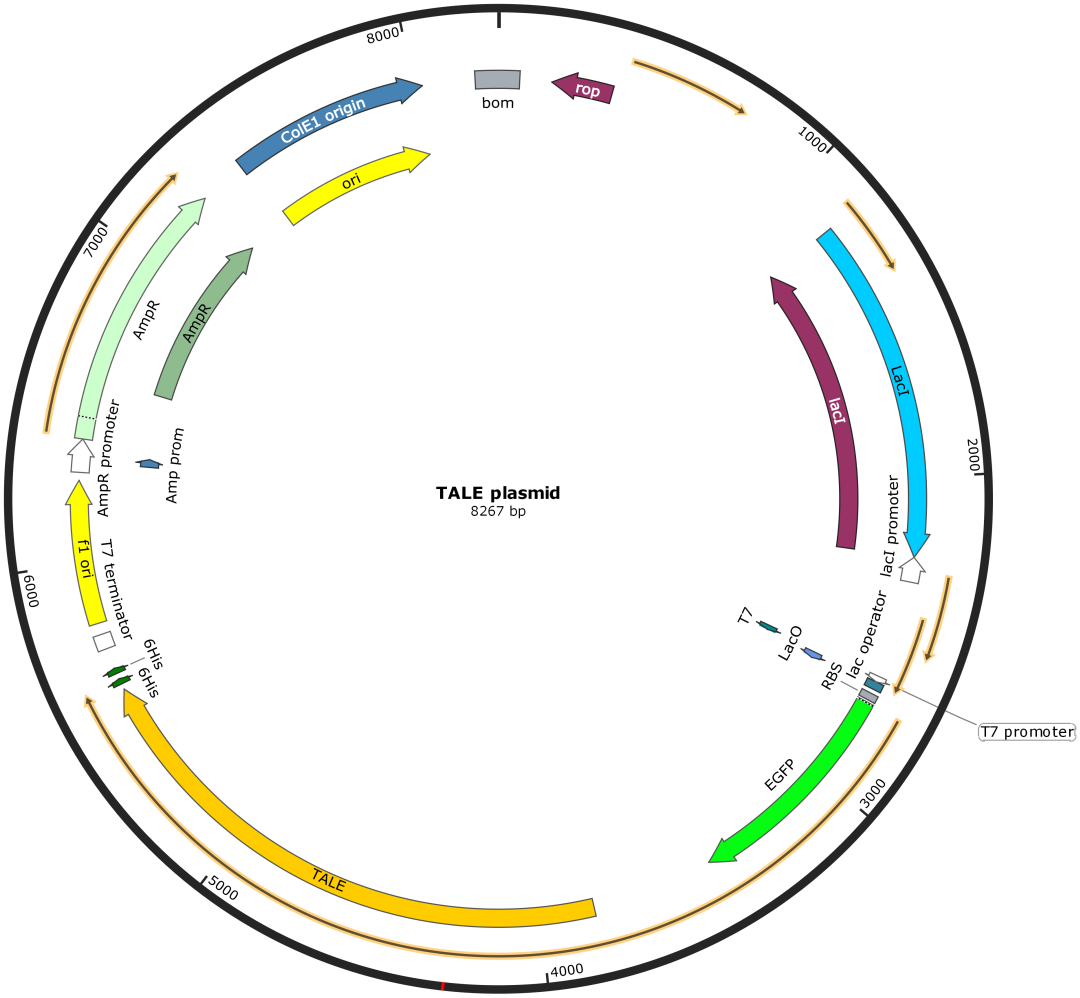


Figure SI 3: Protein Sequences (HEY2b, BRCA1(18) and CDKN2A(18) context) of Exemplary TALE Protein with N-terminal GFP Domain

GFP_TALE_HEY2b

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVT
TLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKTLGYSQQQQE
KIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGV
GKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQR
LLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQA
LETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**N**
GGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPV
LCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETV
QRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIASHDGGK
QALESIVAQLSRPDPALAALTNDHLLEHHHHHH*

GFP_TALE_BRCA1(18)

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVT
TLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKTLGYSQQQQE
KIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGV
GKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQR
LLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQA
LETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**H**
DGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPV
LCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETV
QRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGK
QALESIVAQLSRPDPALAALTNDHLLEHHHHHH

GFP_TALE_CDKN2A(18)_RVD4_HD

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVT
TLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKTLGYSQQQQE
KIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGV
GKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQR
LLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQA
LETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**N**
IGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPV
LCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETV
QRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGK
QALESIVAQLSRPDPALAALTNDHLLEHHHHHH

GFP_TALE_CDKN2A(18)_RVD4_NN

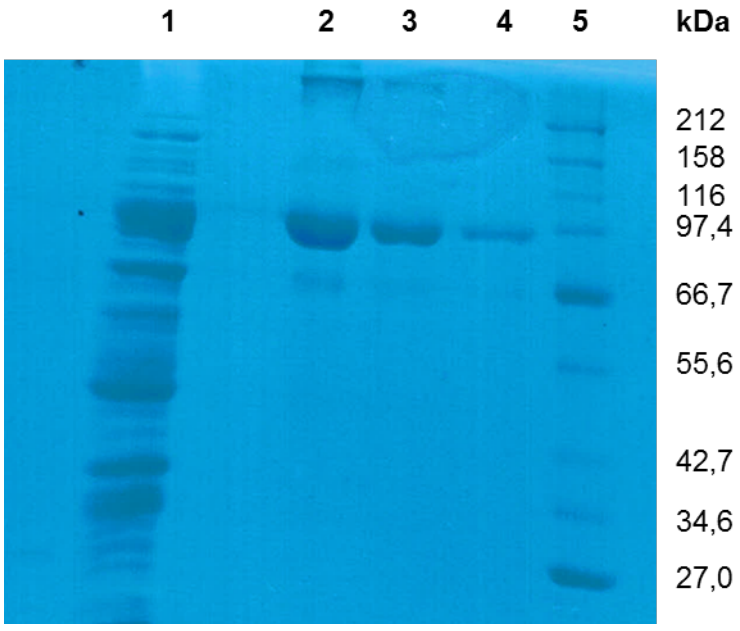
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVT
TLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKTLGYSQQQQE
KIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGV
GKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQR
LLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQA
LETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**N**
IGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPV
LCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETV
QRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGK
QALESIVAQLSRPDPALAALTNDHLLEHHHHHH

GFP_TALE_CDKN2A(18)_RVD4_NG

MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVT
TLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIE
LKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKANFKIRHNIEDGSVQLADHYQQNT
PIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKTLGYSQQQQE
KIKPKVRSTVAQHHEALVGHGFTHAHIVALSQHPAALGTVAVKYQDMIAALPEATHEAIVGV
GKQWSGARALEALLTVAGELRGPPLQLDTGQLLKIAKRGGVTAVEAVHAWRNALTGAPLNLT
PDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLC
QDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQR
LLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQA
LETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**N**
IGGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQV
VAIAS**NN**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHG
LTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPV
LCQDHGLTPDQVVAIAS**HD**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NI**GGKQALETV
QRLLPVLCQDHGLTPDQVVAIAS**NG**GGKQALETVQRLLPVLCQDHGLTPDQVVAIAS**NN**GGK
QALESIVAQLSRPDPALAALTNDHLLEHHHHHH

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

Figure SI 4: SDS PAGE Analysis of Exemplary Ni-NTA-Purified TALE Protein Expression of HEY2b TALE wt



Lane 1: Ni-NTA wash fraction

Lane 2: purified first elution of HEY2b wt TALE

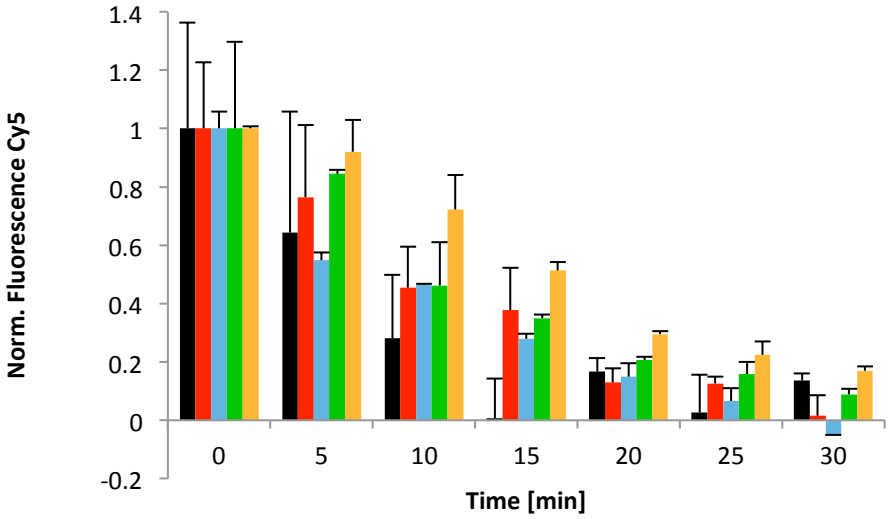
Lane 3: purified second elution HEY2b wt TALE

Lane 4: purified third elution of HEY2b wt TALE

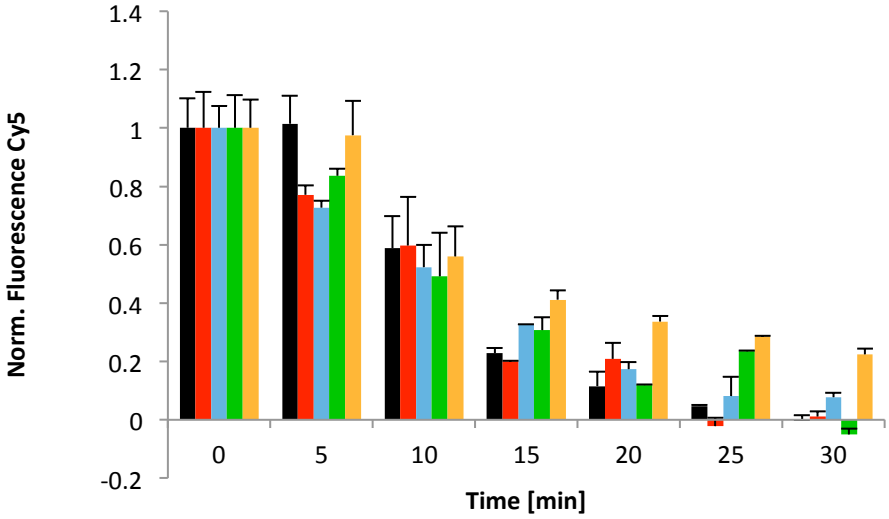
Lane 5: Protein Marker

Figure SI 5: 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.**

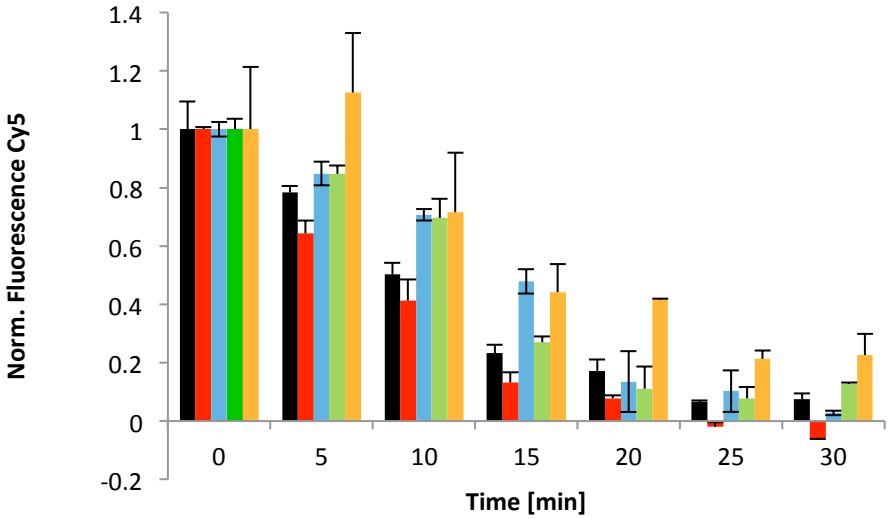
CDKN2A(18)_HD



CDKN2A(18)_NN

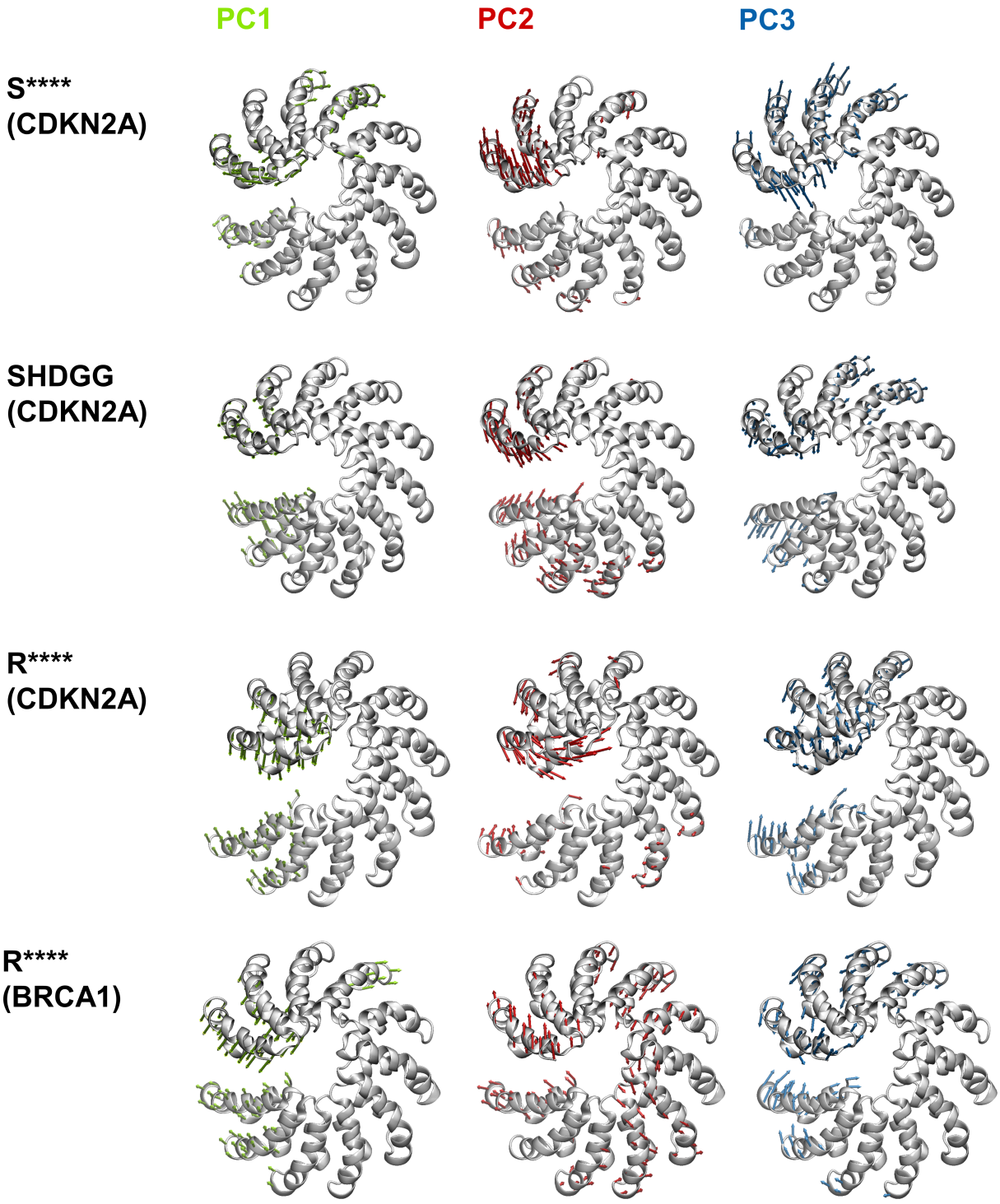


CDKN2A(18)_NG



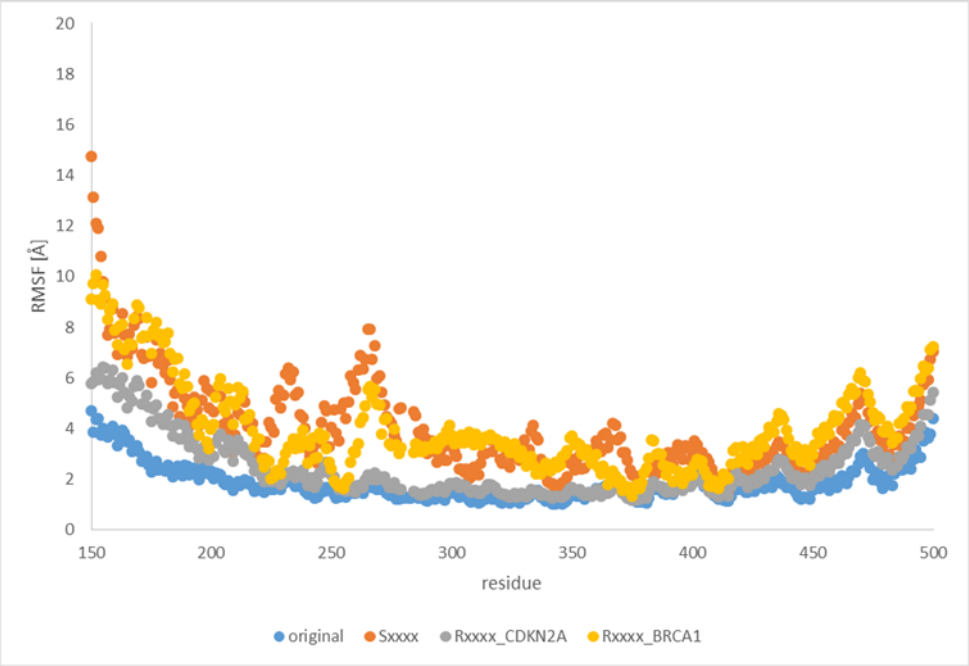
Experiments were performed in duplicates using 10 µM Tale, 0.1 µM DNA, and 1 unit DNaseI.

Figure SI 6: The Directions of the First Three PCA-Derived Modes Used for the Generation of Fig. 4 f-h of the Manuscript together with the Analyis of TALE with Repeat R** binding to BRCA1(18) sequence.**



The Initial Structures after System Equilibration are Shown.

Figure SI 7: RMSF Plot (alignment of all repeats) for TALEs with Repeat SHDGG (original C-binding), R** and S**** Binding to DNA with CDKN2A Sequence and for TALE with Repeat S**** Binding to DNA with BRCA1 Sequence Context (Simulations in the Absence of DNA)**



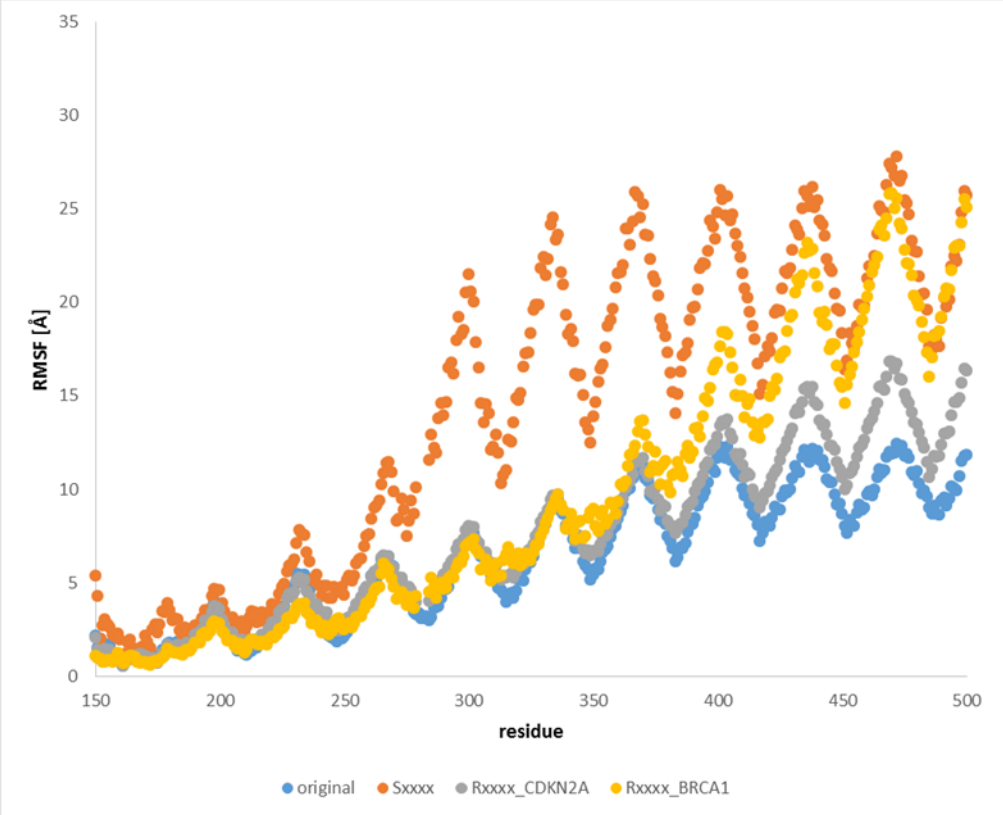
original = wt TALE with repeat SHDGG

Sxxxx = TALE with repeat S****

Rxxxx_CDKN2A = TALE with repeat R**** designed for sequence CDKN2A

Rxxxx_BRCA1 = TALE with repeat R**** designed for sequence BRCA1

Figure SI 8: RMSF Plot (alignment of the first N-terminal repeat) for TALEs with Repeat SHDGG (original C-binding), R** and S**** Binding to DNA with CDKN2A Sequence and for TALE with Repeat S**** Binding to DNA with BRCA1 Sequence Context (Simulations in the Absence of DNA)**



original = wt TALE with repeat SHDGG

Sxxxx = TALE with repeat S****

Rxxxx_CDKN2A = TALE with repeat R**** designed for sequence CDKN2A

Rxxxx_BRCA1 = TALE with repeat R**** designed for sequence BRCA1

Figure SI 9: Analysis of Hydrogen Bonding by TALE wt (Original) Repeat SHDGG in DNA-Unbound TALE Designed for Sequence CDKN2A during 250 ns MD Trajectory. Inter-Repeat Hydrogen Bonding via RVD loops to Downstream (Direction C-Terminus) and Upstream (Direction N-Terminus) repeats are highlighted in red and green, respectively)

original C (SHDGG)

donor	acceptor	occupancy
ASN169-Side	ASP135-Side	0,13%
SER133-Side	VAL129-Main	81,78%
LEU141-Main	GLY137-Main	6,28%
LYS138-Side	GLU142-Side	9,42%
GLY170-Main	**ASP135-Main**	**8,22%**
HID134-Main	ILE131-Main	0,87%
ALA140-Main	GLY136-Main	8,07%
ASP135-Main	**ASN169-Side**	**45,10%**
ALA132-Main	VAL128-Main	22,46%
HID134-Side	ALA130-Main	21,31%
HIE100-Side	ALA132-Main	0,54%
GLU142-Main	LYS138-Main	3,81%
SER133-Side	ALA96-Main	0,17%
SER133-Main	VAL129-Main	7,54%
HID134-Side	ILE131-Main	0,92%
HIE100-Side	HID134-Main	0,07%
LYS138-Side	THR109-Side	1,14%
LYS138-Side	GLN139-Side	3,08%
ASN169-Side	ASP135-Main	0,29%
LYS138-Side	GLN105-Side	1,00%
LYS138-Side	GLN105-Main	0,02%
HIE100-Side	ASP135-Side	1,30%
LYS138-Side	GLU108-Side	0,03%
HID134-Main	ALA132-Main	0,19%
HIE100-Side	SER133-Main	0,36%
GLN139-Main	GLY136-Main	1,66%
GLY102-Main	ASP135-Side	0,01%
GLY137-Main	HID134-Main	0,22%
SER133-Main	ALA130-Main	0,01%
LYS138-Side	**ASP101-Main**	**17,47%**
LYS138-Side	ASP101-Side	0,15%
SER133-Side	GLN93-Side	0,02%
SER133-Side	ALA130-Main	0,26%
LYS138-Side	GLY102-Main	0,02%
GLN105-Side	LYS138-Main	0,01%
ASP101-Main	ASP135-Side	0,02%
THR109-Side	LYS138-Main	0,04%

Figure SI 10: Analysis of Hydrogen Bonding by TALE Repeat S** in DNA-Unbound TALE Designed for Sequence CDKN2A during 250 ns MD Trajectory. Inter-Repeat Hydrogen Bonding via RVD loops to Downstream (Direction C-Terminus) and Upstream (Direction N-Terminus) repeats are highlighted in red and green, respectively)**

S****_CDKN2A

donor	acceptor	occupancy
LYS134-Side	ASP101-Main	1,38%
SER133-Main	ILE131-Main	0,33%
ALA132-Main	VAL128-Main	0,09%
SER133-Side	ILE131-Main	0,11%
LYS134-Side	HIE100-Main	0,14%
ALA132-Main	VAL129-Main	1,78%
HIE100-Side	ALA132-Main	0,20%
LYS134-Side	GLN135-Side	0,24%
ALA136-Main	SER133-Main	1,00%
LYS134-Side	GLY102-Main	0,02%
LEU137-Main	SER133-Main	22,78%
GLU138-Main	LYS134-Main	17,77%
SER133-Side	GLN135-Side	3,34%
LYS134-Side	GLN105-Side	2,34%
LYS134-Side	ASP101-Side	0,89%
ALA136-Main	SER133-Side	0,90%
GLN105-Side	LYS134-Main	0,02%
SER133-Side	HIE100-Side	0,17%
LYS134-Side	GLU138-Side	10,42%
LYS134-Side	THR109-Side	1,46%
LYS134-Side	GLN105-Main	0,03%
SER133-Side	HIE100-Main	0,02%
SER133-Side	ASP101-Side	3,10%
LYS134-Main	ASP101-Side	4,22%
GLN135-Side	SER133-Side	0,12%
SER133-Side	ALA130-Main	0,02%
GLN105-Side	ALA132-Main	0,02%
LYS134-Main	GLN105-Side	0,02%
SER133-Main	ALA130-Main	0,02%
LYS168-Side	LYS134-Main	0,03%
GLN135-Main	SER133-Side	0,07%
LYS134-Side	ASP120-Side	0,14%
ALA132-Main	ALA162-Main	1,56%
SER133-Side	ASN165-Main	0,06%
ALA132-Main	SER163-Main	0,05%

Figure SI 11: Analysis of Hydrogen Bonding by TALE Repeat R** in DNA-Unbound TALE Designed for Sequence CDKN2A during 250 ns MD Trajectory. Inter-Repeat Hydrogen Bonding via RVD loops to Downstream (Direction C-Terminus) and Upstream (Direction N-Terminus) repeats are highlighted in red and green, respectively)**

R****_CDKN2A

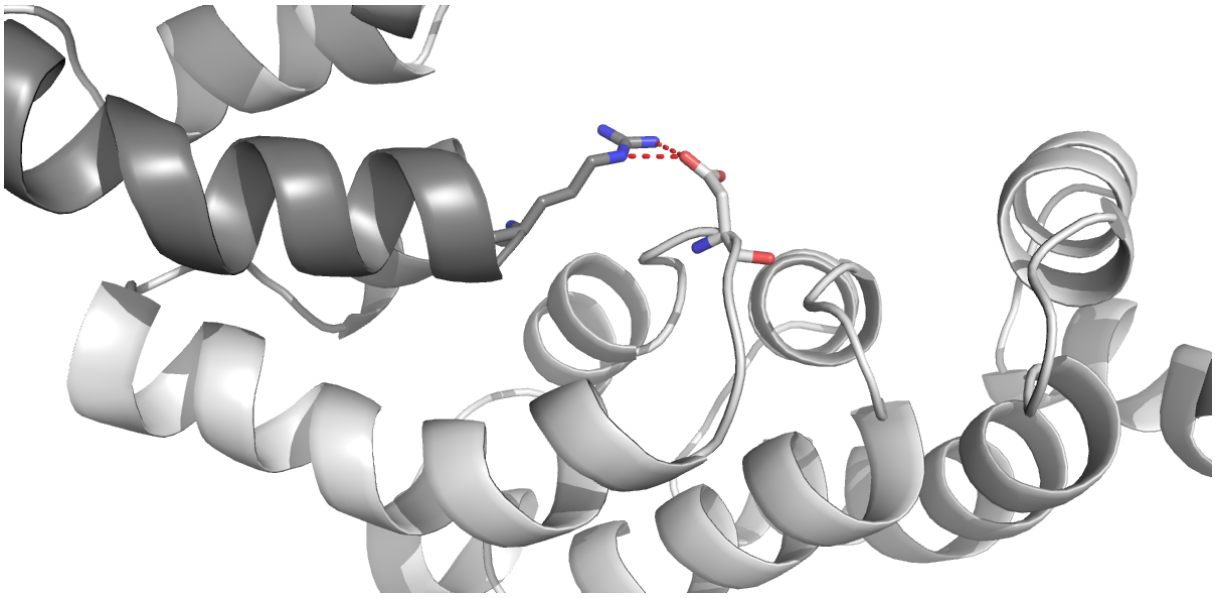
donor	acceptor	occupancy
LYS134-Side	GLU138-Side	17,90%
ARG133-Side	ILE97-Main	0,03%
GLU138-Main	LYS134-Main	31,33%
LEU137-Main	ARG133-Main	20,16%
ARG133-Side	SER99-Side	3,94%
ARG133-Side	ALA132-Main	0,21%
ALA132-Main	VAL129-Main	0,01%
ALA132-Main	VAL128-Main	33,54%
LYS134-Side	THR109-Side	0,90%
ARG133-Main	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 SI 12: Analysis of Hydrogen Bonding by TALE Repeat R** in DNA-Unbound TALE Designed for Sequence BRCA1 during 250 ns MD Trajectory. Inter-Repeat Hydrogen Bonding via RVD loops to Downstream (Direction C-Terminus) and Upstream (Direction N-Terminus) repeats are highlighted in red and green, respectively)**

R****_BRCA1

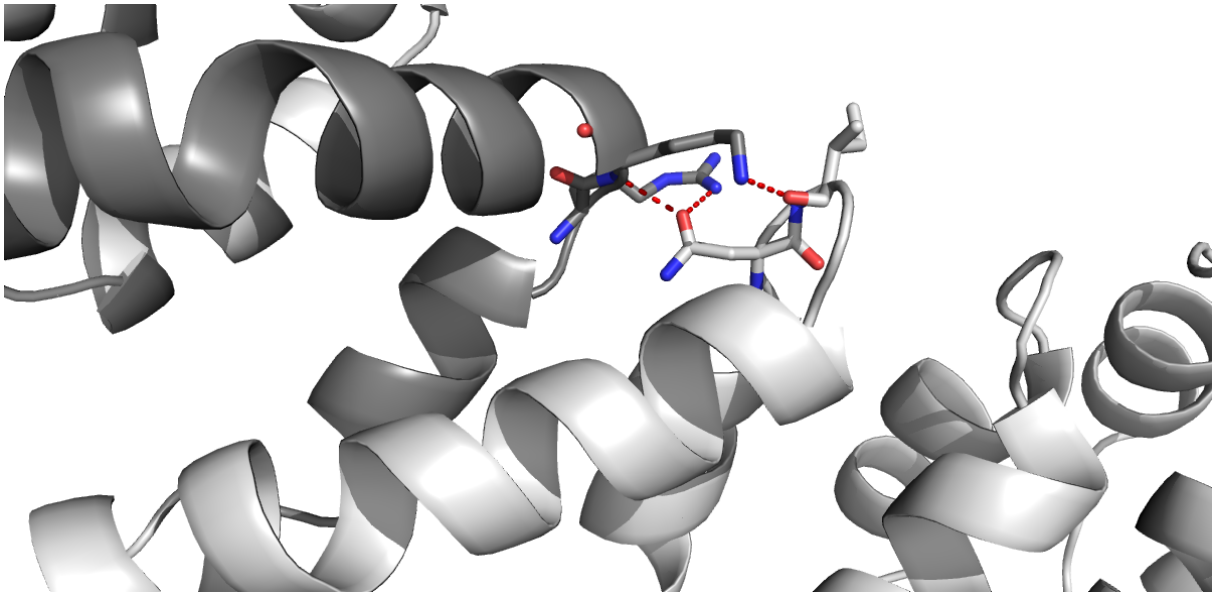
donor	acceptor	occupancy
LEU137-Main	ARG133-Main	12,98%
LYS134-Side	GLU138-Side	13,20%
GLU138-Main	LYS134-Main	17,23%
ALA132-Main	VAL128-Main	14,32%
ARG133-Main	ILE131-Main	3,95%
ARG133-Side	ASN165-Main	0,22%
ARG133-Side	ASN165-Side	1,19%
ARG133-Side	ILE131-Main	0,83%
ARG133-Side	**ASN100-Side**	**8,18%**
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 SI 13: Stabilizing Hydrogen Bond between R** Repeat and Preceding HD Repeat in DNA-Unbound Form Observed during MD Simulations (snapshot after 249 ns)**



R**** repeat in grey, other repeats in white.

Figure SI 14: Stabilizing Hydrogen Bond Between R** Repeat and Preceding NI Repeat in DNA-Unbound Form Observed during MD Simulations (Snapshot after 238 ns)**



R**** repeat in grey, other repeats in white.

References

[1] 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.

[2] G. Kubik, M. J. Schmidt, J. E. Penner, D. Summerer, *Angewandte Chemie International Edition* **2014**, *53*, 6002-6006.

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

[4] R. D. Finn, P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry, A. L. Mitchell, S. C. Potter, M. Punta, M. Qureshi, A. Sangrador-Vegas, G. A. Salazar, J. Tate, A. Bateman, *Nucleic Acids Res* **2016**, *44*, D279-285.

[5] J. Grau, M. Reschke, A. Erkes, J. Streubel, R. D. Morgan, G. G. Wilson, R. Koebnik, J. Boch, *Sci Rep* **2016**, *6*, 21077.

[6] G. Lackner, N. Moebius, L. Partida-Martinez, C. Hertweck, *J Bacteriol* **2011**, *193*, 783-784.

[7] aO. de Lange, C. Wolf, P. Thiel, J. Kruger, C. Kleusch, O. Kohlbacher, T. Lahaye, *Nucleic Acids Res* **2015**, *43*, 10065-10080; bO. de Lange, C. Wolf, J. Dietze, J. Elsaesser, R. Morbitzer, T. Lahaye, *Nucleic Acids Res* **2014**, *42*, 7436-7449.

[8] , Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, **2018**.

[9] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J Comput Chem* **2004**, *25*, 1605–1612.

[10] aP. R. Gerber, K. Müller, *Journal of Computer-Aided Molecular Design* **1995**, *9*, 251–268; bD. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, *J Comput Chem* **2005**, *26*, 1668–1688.

[11] D. A. Case, R. M. Betz, C. D. S, I. T. E. I. I. Cheatham, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, Homeyer, S. Izadi, P. Panowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, P. A. Kollman, University of California, San Francisco, **2016**.

[12] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J Chem Theory Comput* **2015**, *11*, 3696–3713.

[13] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, R. C. Walker, *J Chem Theory Comput* **2013**, *9*, 3878–3888.

[14] W. Humphrey, A. Dalke, K. Schulten, *Journal of Molecular Graphics* **1996**, *14*, 33–38.

[15] A. Bakan, L. M. Meireles, I. Bahar, *Bioinformatics (Oxford, England)* **2011**, *27*, 1575–1577.

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