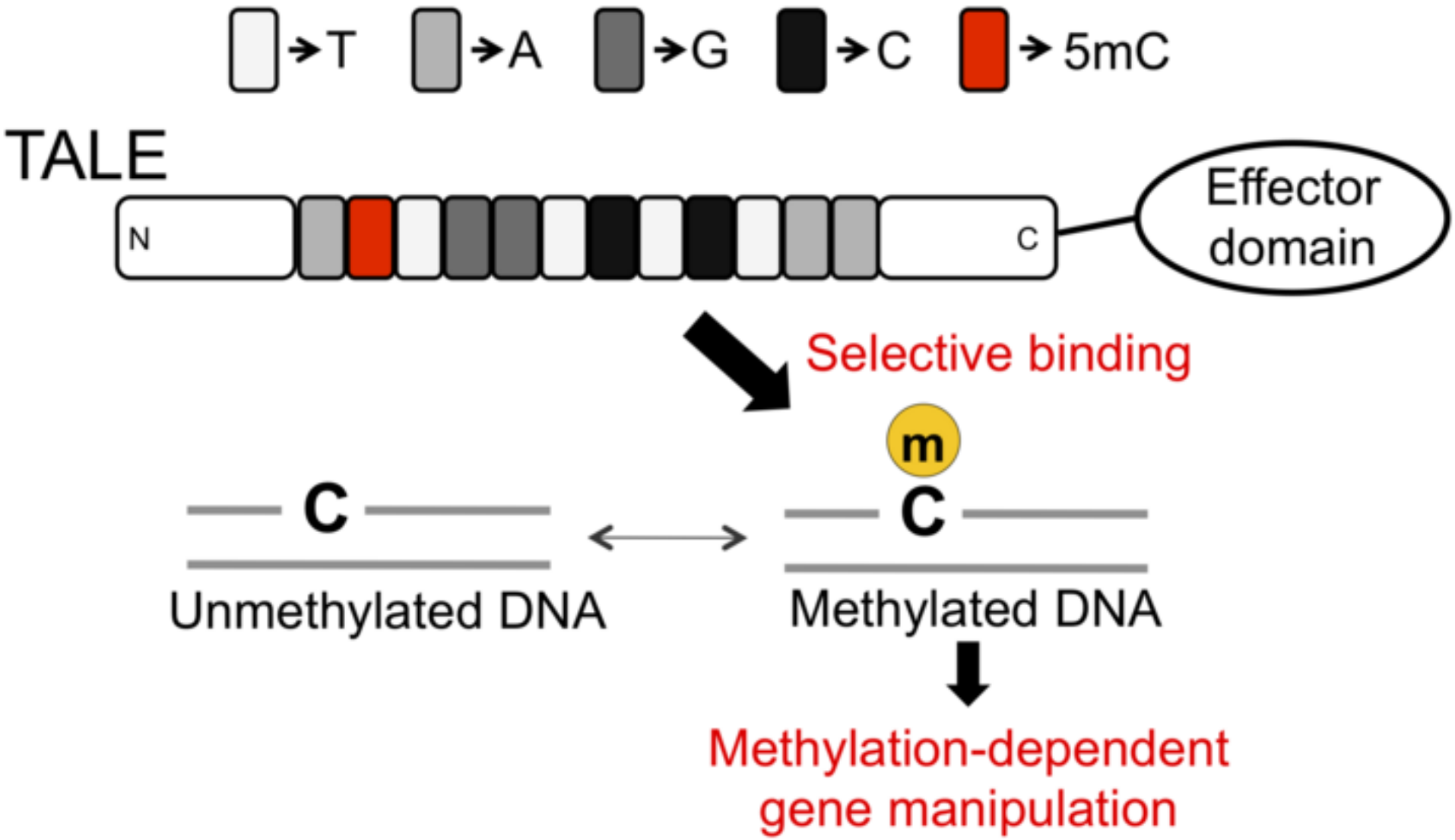
*Graphical Abstract
Click here to download high resolution image



Modified nucleobase-specific gene regulation using engineered transcription activator-like effectors

Shogo Tsuji[§], Miki Imanishi*
Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan

* Corresponding author at: Institute for Chemical Research, Kyoto University, Uji, Kyoto 611-0011, Japan.
E-mail address: imiki@scl.kyoto-u.ac.jp

[§] Present address: National Center for Geriatrics and Gerontology, Obu, Aichi 474-8511, Japan

Abbreviations:

TALE, transcription activator-like effector; RVD, repeat variable diresidues; 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxylcytosine; CRISPR, clustered regularly interspaced short palindromic repeats; gRNA, guide RNA

Epigenetic modification, as typified by cytosine methylation, is a key aspect of gene regulation that affects many biological processes. However, the biological roles of individual methylated cytosines are poorly understood. Sequence-specific DNA recognition tools can be used to investigate the roles of individual instances of DNA methylation. Transcription activator-like effectors (TALEs), which are DNA-binding proteins, are promising candidate tools with designable sequence specificity and sensitivity to DNA methylation. In this review, we describe the bases of DNA recognition of TALEs, including methylated cytosine recognition, and the applications of TALEs for the study of methylated DNA. In addition, we discuss TALE-based epigenome editing and oxidized methylated cytosine recognition.

Keywords: DNA methylation, DNA binding, epigenetics, gene manipulation, gene targeting

1. Introduction

In mammals, DNA methylation occurs mainly at the fifth position of cytosine bases in CpG dinucleotides. The resulting 5-methylcytosine (5mC) is an important epigenetic mark that regulates gene expression, chromatin remodeling, and genome stability [1]. The typical biological change induced by 5mC is transcriptional repression due to the restricted DNA binding of transcription factors. Due to the importance of 5mC, many 5mC detection methods have been developed, including bisulfite sequencing [2] and detection with anti-5mC antibodies or 5mC-binding proteins [3-5]. These methods have revealed that DNA methylation status changes dynamically during biological events and diseases [6]. However, the biological functions of locus-specific 5mC are poorly understood. One strategy to explore the biological roles of locus-specific 5mC is the use of molecular tools that recognize 5mC in a sequence-specific manner in living cells. In this context, we and others have focused on sequence-specific DNA-binding molecules, especially transcription activator-like effectors (TALEs).

DNA-binding molecules are widely used for sequence-specific gene manipulation in combination with various functional domains, such as transcriptional regulators, endonucleases, and epigenetic modifiers. For gene manipulation at desired genomic loci, DNA-binding molecules must be designed to bind specific sequences. C2H2-type zinc finger proteins were the

first proteins reported with programmable sequence specificity (Fig. 1A) [7-9]. A single zinc finger motif recognizes 3 specific bases; its specificity is determined by the amino acid sequence of its α-helix [8]. By aligning specific zinc finger motifs in tandem, artificial proteins with desirable sequence specificity have been generated [10]. However, the DNA recognition of zinc finger motifs is also affected by adjacent zinc finger motifs. Satisfactory DNA binding is not always achieved after the modular assembly of zinc finger motifs [11, 12], which necessitates the laborious redesigning of zinc finger proteins.

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) system is an impressive new tool for sequence-specific gene manipulation (Fig. 1B) [13-15]. The CRISPR/Cas9 system recognizes a specific DNA sequence based on the Watson–Crick base-pairing between the 20 bases at the 5′-terminal region of the guide RNA (gRNA) and 1 strand of the target DNA. The Cas9 nuclease is introduced to the target region as a Cas9/gRNA complex, resulting in a double-strand DNA break at the target site [16]. The catalytically deactivated Cas9 that fuses with the functional domains enables various manipulations other than DNA cleavage [17, 18]. Using the CRISPR/Cas9 system, gene-specific manipulation can be achieved by simply designing a gRNA that is complementary to the target DNA. Due to its simple programmability, the CRISPR/Cas9 system has been rapidly adopted and widely used as a standard tool for genome manipulation.

However, the simple DNA recognition mode of the CRISPR/Cas9 system has disadvantages for epigenetic research because Watson–Crick hybridization is insensitive to cytosine modifications. Thus, the gRNA cannot discriminate between the various epigenetic states of a target sequence. Furthermore, changing the DNA recognition mode of gRNA to render it sensitive to cytosine modification appears to be difficult. On the other hand, in most DNA–protein interactions, DNA recognition is mediated by the side chains of the amino acid residues of the DNA-binding proteins. The DNA-binding sequence specificity of proteins should be manipulable via modification of these interacting amino acid residues. In fact, an engineered zinc finger protein has shown sensitivity to methylated cytosine bases [19]. However, zinc finger proteins have disadvantages in terms of programmability. In contrast, TALEs are promising candidates for epigenetic research and therapy (Fig. 1C). TALEs have easily programmable sequence selectivity and sensitivity to cytosine modification. In this review, we focus on TALE specificity for 5mC and other modified nucleobases. In addition, we describe the applications of TALEs in epigenetic research.

2. DNA-binding properties of TALE proteins

2.1. Binding mode of TALEs

TALEs consist of 3 domains: an N-terminal domain that harbors the type III secretion and translocation signal, a central repeat domain that comprises a series of tandem repeats of typically 34 highly conserved amino acids, and a C-terminal domain that contains nuclear localization signals (NLSs) and an activation domain (AD) (Fig. 2A, B) [20]. The DNA-binding specificities of TALEs are determined mainly by the central repeat domain. Each repeat recognizes 1 target base without significant effects from the neighboring repeats [21-23]. The base preferences of these repeats are determined by 2 amino acids, at positions 12 and 13, called repeat variable diresidues (RVD) (Fig. 2C). Owing to the simple recognition codes of RVD for DNA bases, TALEs can be readily designed to bind to desired DNA sequences. This programmable sequence specificity has made TALEs an attractive tool for targeted gene manipulation including transcriptional regulation, genome editing, live-cell imaging of specific chromatin loci, chromatin immunoprecipitation, and epigenetic modification (Fig. 3).

2.2. DNA recognition by the central repeat domain

More than 20 naturally occurring RVDs have been identified. Among them, NG (asparagine and glycine), HD (histidine and aspartic acid), NI (asparagine and isoleucine), and NN (asparagine and asparagine) are commonly used for the specific recognition of the nucleobases thymine, cytosine, adenine, and guanine/adenine, respectively [21-23]. As the RVD NN recognizes guanine and adenine, the RVD NH (asparagine and histidine), which is more guanine-specific but less active, is also used for recognition of guanine (Fig. 2A, inset) [24, 25]. The crystal structure of a TALE–DNA complex revealed that each repeat consists of 2 α-helixes with a loop between the helixes (Fig. 2C) [26, 27]. The loop region contains an RVD and interacts with the corresponding DNA base on the sense strand of the target DNA, thereby determining the specificity of the RVD. For base recognition, only the amino acid at position 13 directly makes contact with the corresponding base, whereas the amino acid at position 12 interacts with the amino acid at position 8 of the same repeat and stabilizes the proper loop conformation [26, 27]. Comprehensive analyses covering all 400 possible RVD combinations revealed a few novel, functional RVDs [28, 29]. These studies demonstrated the potential of engineered TALE repeats with desired base specificity.

2.3. DNA recognition by the N-terminal domain

Although the central repeat domain predominantly defines TALE DNA-binding specificity, the N-terminal domain also contributes to TALE DNA binding. In nature, almost all TALE-binding sites are preceded by a conserved T nucleotide, called T_0 [21, 22]. The importance of T_0 for efficient TALE DNA binding has been experimentally demonstrated [22, 30, 31]; the presence of T_0 restricts the sites that can be targeted by TALE-based gene manipulation. The crystal structures of TALEs have shown that their N-terminal domains contain 4 degenerate repeats, termed repeats −3, −2, −1, and 0 [26, 27, 32, 33]. Each of the 4 repeats contains 2 α-helixes with an adjoining loop, similar to the canonical repeats. Structural data suggested that the side chain of W232 in the adjoining loop of repeat −1 makes contact with T_0 by van der Waals interactions [26]. The R266 residue in the helix of repeat −1 may also interact with T_0 [33]. In addition, the RVD in repeat −1 reportedly affects T_0 specificity [34]. Although multiple amino acid residues may participate in T_0 recognition, several studies have shown that point mutations at W232 greatly modify T_0 specificity [35-37]. Furthermore, the molecular evolution of the loop region of repeat −1 has created TALE variants that tolerate all 4 nucleotides at position zero [35, 36]. These variants could simplify the design of the binding sites of TALE proteins. Furthermore, these studies demonstrated the feasibility of using directed evolution methods to alter the DNA recognition modes of TALE proteins.

3. Recognition of cytosine methylation by TALEs

3.1. 5mC recognition by natural RVDs

Bultmann *et al*. reported that TALE-based transcriptional activation did not occur in the presence of 5mC in the target genomic DNA sequence [38]. In the TALE repeat that contains the RVD HD as a C-binder, the aspartate residue of the RVD interacts with the cytosine N4 atom via hydrogen bond [26, 27]. Structural and biochemical analyses suggested that an additional methyl moiety in 5mC causes steric clash between the aspartate residue and the methyl moiety, resulting in a binding penalty [39, 40], which explains the silencing of the activity of TALEs containing the RVD HD at highly methylated target sequences.

On the other hand, RVDs that recognize 5mC have been reported. For examples, the RVD N* (where * represents a deletion) tolerates 5mC [39]. This is presumably because the deletion of the amino acid at position 13 prevents steric clash with the additional methyl moiety of 5mC. The commonly used RVD NG, which is specific for thymine, has also been reported to bind 5mC because of the structural similarity between thymine and 5mC [40]. These findings have

enabled the application of TALE-based technology for the investigation of methylated DNA sequences [39, 41].

3.2. Discrimination of methylation states by natural and engineered RVDs

Using RVDs that recognize 5mC, TALE-based gene manipulation can be applied to any sequences with cytosine methylations. In addition, using RVDs that discriminate between cytosine and 5mC, cytosine methylation can be assessed at the desired sequences. For example, the RVD HD has a stronger preference for unmodified cytosine than for 5mC [42]. Furthermore, new TALE scaffolds with enhanced abilities to discriminate between C and 5mC, based on engineered DNA backbone interactions, have been recently reported [43]. Based on the strong unmodified cytosine preference of the RVD HD, Kubik *et al*. reported sequence-specific 5mC detection, of even a single cytosine methylation, in oligo DNA and extracted genomic zebrafish DNA [42, 44].

Although the strong preference of the RVD HD for unmodified cytosine is useful for the detection of cytosine methylation, an RVD with a strong 5mC preference is needed to enable the use of TALE-based technology to selectively manipulate methylated sites (Fig. 4). The RVDs NG and N* can tolerate 5mC, but their methylation discrimination abilities are not high [42, 45]. Therefore, to obtain new RVDs that bind to 5mC but not cytosine, bacterial 1-hybrid screening of a TALE repeat library was conducted [45]. In the study, a TALE repeat library was constructed by randomizing 4 amino acid residues, which covered RVDs and their neighboring residues (positions 11–14). The identified repeat having “ASAA” at positions 11-14 showed a strong preference for 5mC. The component amino acids, alanine and serine, are small, which probably minimizes steric hindrance between the RVD loop and the methyl moiety of 5mC, as for the RVD N*.

3.3. TALE application for methylated DNA sequences

The methylation discriminating ability of TALEs enables the use of TALE-based technology in a cytosine methylation-dependent manner. Specific human genomic DNA sequences have been isolated and 5mC has been quantitatively analyzed at single nucleotide positions using modified base-specific TALEs as programmable affinity probes [46]. A split protein-fused TALE system, which reports TALE–DNA binding as an easily detectable light signal with a low signal-to-background ratio[47], is another method to detect the methylation states of specific genomic DNA loci. In the system, a pair of TALEs fused with to the N- or C-terminal portions of a fluorescent or luminescent protein localize to adjacent sites by binding to the target DNA,

resulting in the reconstitution of the reporter protein activity. DNA methylation at endogenous genomic sites was detected using the methyl CpG binding domain of MBD1 [48] as the partner module of a sequence-specific TALE in a split fluorescence complementation system [49]. The methylation states of specific genomic DNA loci have also been detected by split luminescence analysis using TALEs with the ability to discriminate DNA methylation [50]. The biggest advantage of these systems is the ability to detect locus-specific DNA methylation in living cells because the split reporter-fused TALEs are genetically encodable. This live-cell methylation detection system is of great clinical and research value because of its sequence selectivity and non-invasiveness, which are improvements over previous 5mC detection methods. In addition to simply detecting cytosine methylation, various TALE-based gene manipulations, such as transcriptional activation and genome editing, have been achieved in a cytosine methylation-dependent manner [45, 51]. These live-cell applications provide new ways to explore the biological functions of modified cytosines.

4. TALEs-based control of DNA methylation states at specific genomic sites

Another approach to study the roles of individual instances of cytosine methylation is to manipulate the methylation states of specific genomic sites. To this end, TALEs are fused with methylation catalysts [52]. For example, TALEs have been fused to the catalytic domains of DNA methyltransferases, resulting in the successful methylation of target promoter regions [53, 54]. The consequent silencing of the target genes demonstrated that the methylation was biologically functional. On the contrary, TALEs fused with the catalytic domain of ten-eleven translocation methylcytosine dioxygenase, which catalyzes the first step of cytosine demethylation, reportedly removed DNA methylation in the target region [54-56]. These gene-specific demethylation tools have potential for the clinical reactivation of genes, such as the tumor suppressor genes that are inactivated in certain cancer cells. Non-TALE DNA-binding molecules, including zinc finger proteins [57, 58] and gRNA/Cas9 [59-62] complexes, have also been used in DNA methylation and demethylation. For site-specific modulation of DNA methylation states, the CRISPR/Cas9 system may have a great advantage in design feasibility, although TALEs and zinc finger proteins have superiority abilities to distinguish between 5mC and C.

5. Oxidized 5mC recognition by TALEs

Biological processes often involve active DNA demethylation, an enzymatic process that removes or modifies the methyl moiety from 5mC (Fig. 5). During the demethylation reaction, 5mC is first oxidized to 5-hydroxymethylcytosine (5hmC), then to 5-formylcytosine (5fC), 5-carboxylcytosine (5caC), and unmodified cytosine. A small number of reader proteins for these oxidized 5mC bases have been identified [63]. The existence of specific reader proteins suggests that these oxidized 5mC bases are not just intermediates during the demethylation process, but specific epigenetic marks with individual biological roles. Therefore, TALE recognition of oxidized 5mC is a matter of considerable interest. The abilities of the natural RVDs HD, NG, and N* to bind to cytosine, 5mC, and 5hmC were evaluated; none of the RVDs bound to 5hmC [64]. To obtain new RVDs with a preference for 5hmC and other oxidized 5mC moieties, TALE repeat mutants have been screened. Maurer *et al*. screened a size-reduced TALE repeat library, in which the TALEs had deletions of the position 13 residue to reduce the steric hindrance between the RVD loop and the C5 atom of modified cytosine, for binding to cytosine, 5mC, 5hmC, 5fC, and 5caC [65]. The repeats G*, S*, and T* showed similar affinities for all 5 cytosine nucleobases [65]. Zhang *et al.* reported a comprehensive investigation of the recognition ability of all 400 possible RVDs for 5hmC and 5mC and identified several RVDs that are specific for 5hmC [51]. Using these artificial RVDs, TALEs can be designed for various sequences that contain oxidized 5mC. However, RVDs with a strong preference for only 1 cytosine derivative have not been identified. The next goal in this field is to identify RVDs that bind a specific modified cytosine with high selectivity, which would enable modification-specific gene manipulation and greatly aid the exploration of the biological roles of cytosine modification.

6. Conclusion

Sequence-specific epigenome targeting has important applications in basic research and therapy. TALEs, which have easily programmable sequence selectivity and sensitivity to cytosine modification, are among the most promising molecular tools for epigenome targeting. By combining TALEs that can bind to modified cytosine with various functional domains, sequence-specific gene manipulation can be applied to genomic regions with epigenetic modifications. Furthermore, the discovery of TALE RVDs specific for individual cytosine modifications would enable the analysis of the dynamically changing epigenetic states of specific genomic loci in living cells. TALEs are expected to be powerful tools for epigenetic research and therapy.

Acknowledgments

This work was supported by JSPS KAKENHI (16H03281) and partially by JST/CREST (JPMJCR14W3) (M.I.).

References

[1] J.A. Law, S.E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals, Nat. Rev. Genet. 11 (2010) 204-220.

[2] M. Frommer, L.E. McDonald, D.S. Millar, C.M. Collis, F. Watt, G.W. Grigg, P.L. Molloy, C.L. Paul, A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 1827-1831.

[3] F. Santos, B. Hendrich, W. Reik, W. Dean, Dynamic reprogramming of DNA methylation in the early mouse embryo, Dev. Biol. 241 (2002) 172-182.

[4] C. Desjobert, M. El Mai, T. Gerard-Hirne, D. Guianvarc'h, A. Carrier, C. Pottier, P.B. Arimondo, J. Riond, Combined analysis of DNA methylation and cell cycle in cancer cells, Epigenetics 10 (2015) 82-91.

[5] S. Celik-Uzuner, Y. Li, L. Peters, C. O'Neill, Measurement of global DNA methylation levels by flow cytometry in mouse fibroblasts, In Vitro Cell. Dev. Biol. Anim. 53 (2017) 1-6.

[6] Z.D. Smith, A. Meissner, DNA methylation: roles in mammalian development, Nat. Rev. Genet. 14 (2013) 204-220.

[7] C.O. Pabo, E. Peisach, R.A. Grant, Design and selection of novel Cys2His2 zinc finger proteins, Annu. Rev. Biochem. 70 (2001) 313-340.

[8] N.P. Pavletich, C.O. Pabo, Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å, Science 252 (1991) 809-817.

[9] E.J. Rebar, C.O. Pabo, Zinc finger phage: affinity selection of fingers with new DNA-binding specificities, Science 263 (1994) 671-673.

[10] D.J. Segal, B. Dreier, R.R. Beerli, C.F. Barbas, 3rd, Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 2758-2763.

[11] C.L. Ramirez, J.E. Foley, D.A. Wright, F. Muller-Lerch, S.H. Rahman, T.I. Cornu, R.J. Winfrey, J.D. Sander, F. Fu, J.A. Townsend, T. Cathomen, D.F. Voytas, J.K. Joung, Unexpected failure rates for modular assembly of engineered zinc fingers, Nat. Methods 5 (2008) 374-375.

[12] M. Imanishi, A. Nakamura, T. Morisaki, S. Futaki, Positive and negative cooperativity of modularly assembled zinc fingers, Biochem. Biophys. Res. Commun. 387 (2009) 440-443.

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

[14] L. Cong, F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems, Science 339 (2013) 819-823.

[15] P. Mali, L. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, J.E. Norville, G.M. Church, RNA-guided human genome engineering via Cas9, Science 339 (2013) 823-826.

[16] F. Jiang, J.A. Doudna, CRISPR-Cas9 Structures and Mechanisms, Annu. Rev. Biophys. 46 (2017) 505-529.

[17] S. Konermann, M.D. Brigham, A.E. Trevino, J. Joung, O.O. Abudayyeh, C. Barcena, P.D. Hsu, N. Habib, J.S. Gootenberg, H. Nishimasu, O. Nureki, F. Zhang, Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Nature 517 (2015) 583-588.

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

[19] Y. Choo, Recognition of DNA methylation by zinc fingers, Nat. Struct. Biol. 5 (1998) 264-265.

[20] K. Gu, B. Yang, D. Tian, L. Wu, D. Wang, C. Sreekala, F. Yang, Z. Chu, G.L. Wang, F.F. White, Z. Yin, R gene expression induced by a type-III effector triggers disease resistance in rice, Nature 435 (2005) 1122-1125.

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

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

[23] H. Scholze, J. Boch, TAL effector-DNA specificity, Virulence 1 (2010) 428-432.

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

[25] J. Streubel, C. Blucher, A. Landgraf, J. Boch, TAL effector RVD specificities and

efficiencies, Nat. Biotechnol. 30 (2012) 593-595.

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

[27] D. Deng, C. Yan, X. Pan, M. Mahfouz, J. Wang, J.K. Zhu, Y. Shi, N. Yan, Structural basis for sequence-specific recognition of DNA by TAL effectors, Science 335 (2012) 720-723.

[28] J. Yang, Y. Zhang, P. Yuan, Y. Zhou, C. Cai, Q. Ren, D. Wen, C. Chu, H. Qi, W. Wei, Complete decoding of TAL effectors for DNA recognition, Cell Res. 24 (2014) 628-631.

[29] 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, Improved specificity of TALE-based genome editing using an expanded RVD repertoire, Nat. Methods 12 (2015) 465-471.

[30] P. Romer, T. Strauss, S. Hahn, H. Scholze, R. Morbitzer, J. Grau, U. Bonas, T. Lahaye, Recognition of AvrBs3-like proteins is mediated by specific binding to promoters of matching pepper Bs3 alleles, Plant Physiol. 150 (2009) 1697-1712.

[31] P. Romer, S. Recht, T. Strauss, J. Elsaesser, S. Schornack, J. Boch, S. Wang, T. Lahaye, Promoter elements of rice susceptibility genes are bound and activated by specific TAL effectors from the bacterial blight pathogen, Xanthomonas oryzae pv. oryzae, New Phytol. 187 (2010) 1048-1057.

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

[33] S. Stella, R. Molina, I. Yefimenko, J. Prieto, G. Silva, C. Bertonati, A. Juillerat, P. Duchateau, G. Montoya, Structure of the AvrBs3-DNA complex provides new insights into the initial thymine-recognition mechanism, Acta Crystallogr. D Biol. Crystallogr. 69 (2013) 1707-1716.

[34] T. Schreiber, U. Bonas, Repeat 1 of TAL effectors affects target specificity for the base at position zero, Nucleic Acids Res. 42 (2014) 7160-7169.

[35] B.M. Lamb, A.C. Mercer, C.F. Barbas, 3rd, Directed evolution of the TALE N-terminal domain for recognition of all 5' bases, Nucleic Acids Res. 41 (2013) 9779-9785.

[36] S. Tsuji, S. Futaki, M. Imanishi, Creating a TALE protein with unbiased 5'-T binding, Biochem. Biophys. Res. Commun. 441 (2013) 262-265.

[37] E.L. Doyle, A.W. Hummel, Z.L. Demorest, C.G. Starker, D.F. Voytas, P. Bradley, A.J. Bogdanove, TAL effector specificity for base 0 of the DNA target is altered in a complex, effector- and assay-dependent manner by substitutions for the tryptophan in cryptic repeat -1, PLoS One 8 (2013) e82120.

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

[39] J. Valton, A. Dupuy, F. Daboussi, S. Thomas, A. Marechal, R. Macmaster, K. Melliand, A. Juillerat, P. Duchateau, Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation, J. Biol. Chem. 287 (2012) 38427-38432.

[40] D. Deng, P. Yin, C. Yan, X. Pan, X. Gong, S. Qi, T. Xie, M. Mahfouz, J.K. Zhu, N. Yan, Y. Shi, Recognition of methylated DNA by TAL effectors, Cell Res. 22 (2012) 1502-1504.

[41] A. Dupuy, J. Valton, S. Leduc, J. Armier, R. Galetto, A. Gouble, C. Lebuhotel, A. Stary, F. Paques, P. Duchateau, A. Sarasin, F. Daboussi, Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALEN, PLoS One 8 (2013) e78678.

[42] G. Kubik, M.J. Schmidt, J.E. Penner, D. Summerer, Programmable and highly resolved in vitro detection of 5-methylcytosine by TALEs, Angew. Chem. Int. Ed. Engl. 53 (2014) 6002-6006.

[43] P. Rathi, A. Witte, D. Summerer, Engineering DNA Backbone Interactions Results in TALE Scaffolds with Enhanced 5-Methylcytosine Selectivity, Sci. Rep. 7 (2017) 15067.

[44] G. Kubik, D. Summerer, Achieving single-nucleotide resolution of 5-methylcytosine detection with TALEs, Chembiochem 16 (2015) 228-231.

[45] S. Tsuji, S. Futaki, M. Imanishi, Sequence-specific recognition of methylated DNA by an engineered transcription activator-like effector protein, Chem. Commun. 52 (2016) 14238-14241.

[46] P. Rathi, S. Maurer, G. Kubik, D. Summerer, Isolation of Human Genomic DNA

Sequences with Expanded Nucleobase Selectivity, J. Am. Chem. Soc. 138 (2016) 9910-9918.

[47] H. Hu, H. Zhang, S. Wang, M. Ding, H. An, Y. Hou, X. Yang, W. Wei, Y. Sun, C. Tang, Live visualization of genomic loci with BiFC-TALE, Sci. Rep. 7 (2017) 40192.

[48] B. Hendrich, A. Bird, Identification and characterization of a family of mammalian methyl-CpG binding proteins, Mol. Cell. Biol. 18 (1998) 6538-6547.

[49] C. Lungu, S. Pinter, J. Broche, P. Rathert, A. Jeltsch, Modular fluorescence complementation sensors for live cell detection of epigenetic signals at endogenous genomic sites, Nat. Commun. 8 (2017) 649.

[50] S. Tsuji, K. Shinoda, S. Futaki, M. Imanishi, Sequence-specific 5mC detection in live cells based on the TALE-split luciferase complementation system, Analyst 143 (2018) 3793-3797.

[51] Y. Zhang, L. Liu, S. Guo, J. Song, C. Zhu, Z. Yue, W. Wei, C. Yi, Deciphering TAL effectors for 5-methylcytosine and 5-hydroxymethylcytosine recognition, Nat. Commun. 8 (2017) 901.

[52] G. Kungulovski, A. Jeltsch, Epigenome Editing: State of the Art, Concepts, and Perspectives, Trends Genet. 32 (2016) 101-113.

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

[54] C.L. Lo, S.R. Choudhury, J. Irudayaraj, F.C. Zhou, Epigenetic Editing of Ascl1 Gene in Neural Stem Cells by Optogenetics, Sci. Rep. 7 (2017) 42047.

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

[56] K. Li, J. Pang, H. Cheng, W.P. Liu, J.M. Di, H.J. Xiao, Y. Luo, H. Zhang, W.T. Huang, M.K. Chen, L.Y. Li, C.K. Shao, Y.H. Feng, X. Gao, Manipulation of prostate cancer metastasis by locus-specific modification of the CRMP4 promoter region using chimeric TALE DNA methyltransferase and demethylase, Oncotarget 6 (2015) 10030-10044.

[57] F. Li, M. Papworth, M. Minczuk, C. Rohde, Y. Zhang, S. Ragozin, A. Jeltsch, Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences

and repress expression of target genes, Nucleic Acids Res. 35 (2007) 100-112.
[58] W. Nomura, C.F. Barbas, 3rd, In vivo site-specific DNA methylation with a designed sequence-enabled DNA methylase, J. Am. Chem. Soc. 129 (2007) 8676-8677.
[59] S.R. Choudhury, Y. Cui, K. Lubecka, B. Stefanska, J. Irudayaraj, CRISPR-dCas9 mediated TET1 targeting for selective DNA demethylation at BRCA1 promoter, Oncotarget 7 (2016) 46545-46556.
[60] X. Xu, Y. Tao, X. Gao, L. Zhang, X. Li, W. Zou, K. Ruan, F. Wang, G.L. Xu, R. Hu, A CRISPR-based approach for targeted DNA demethylation, Cell Discov. 2 (2016) 16009.
[61] S. Morita, H. Noguchi, T. Horii, K. Nakabayashi, M. Kimura, K. Okamura, A. Sakai, H. Nakashima, K. Hata, K. Nakashima, I. Hatada, Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions, Nat. Biotechnol. 34 (2016) 1060-1065.
[62] X.S. Liu, H. Wu, X. Ji, Y. Stelzer, X. Wu, S. Czauderna, J. Shu, D. Dadon, R.A. Young, R. Jaenisch, Editing DNA Methylation in the Mammalian Genome, Cell 167 (2016) 233-247 e217.
[63] J. Song, G.P. Pfeifer, Are there specific readers of oxidized 5-methylcytosine bases?, Bioessays38 (2016) 1038-1047.
[64] G. Kubik, S. Batke, D. Summerer, Programmable sensors of 5-hydroxymethylcytosine, J. Am. Chem. Soc. 137 (2015) 2-5.
[65] S. Maurer, M. Giess, O. Koch, D. Summerer, Interrogating Key Positions of Size-Reduced TALE Repeats Reveals a Programmable Sensor of 5-Carboxylcytosine, ACS Chem. Biol. 11 (2016) 3294-3299.

Figure legends:

Fig. 1 Schematic representation of the DNA binding of designable DNA-binding molecules. (A) Engineered zinc finger array, in which each module recognizes 3 bases. (B) In the CRISPR/Cas9 system, the gRNA/Cas9 complex recognizes target DNA containing the protospacer adjacent motif (PAM) sequence (red). The 5′-region (blue) of the gRNA undergoes Watson–Crick base-pairing with the target DNA. (C) A TALE contains multiple repeats (gray) that recognize a single base pair.

Fig. 2 DNA binding of a TALE. (A) Schematic representation of a TALE, consisting of an N-terminal domain, a central repeat domain, and a C-terminal domain. The N-terminal domain contains 4 degenerate repeats (−3, −2, −1, and 0). The C-terminal domain contains NLSs and an AD. The amino acid sequence of a repeat is shown. The RVD and their recognized DNA bases are shown in the right box. (B) Structure of the DNA-binding domain of a TALE, PthXo1, bound to its target DNA sequence [PDB: 3UGM]. (C) Structure of a repeat containing the RVD HD [PDB: 3UGM]. The RVDs are shown in red.

Fig. 3 Applications of TALEs. (A) TALEs fused with transcriptional activators increase the transcription of target genes. (B) TALEs fused with transcriptional repressors decrease the transcription of target genes. (C) The catalytic domain of the restriction endonuclease FokI can be fused with TALEs to construct TALE nucleases (TALENs), which work as pairs. When a pair binds to adjacent target sites, the FokI domains dimerize and induce a double-strand DNA break at the target site. (D) TALEs fused with a fluorescent protein are used to visualize repetitive genomic sequences in living cells. (E) TALEs fused with a tag sequence are used to perform chromatin immunoprecipitation assays. (F) TALEs fused with epigenetic modifiers alter the epigenetic states of target sites.

Fig. 4 Schematic of methylation-dependent gene manipulation by TALEs. TALEs with methylation-discriminating RVDs bind to target sites in a methylation state-dependent manner. The selective binding of TALEs can achieve methylation-dependent gene manipulation.

Fig. 5 Cytosine methylation and active demethylation. DNMT: DNA methyltransferase. TET: Ten-eleven translocation dioxygenase. TDG: thymine DNA glycosylase. BER: base excision repair.

Figure1

Click here to download high resolution image

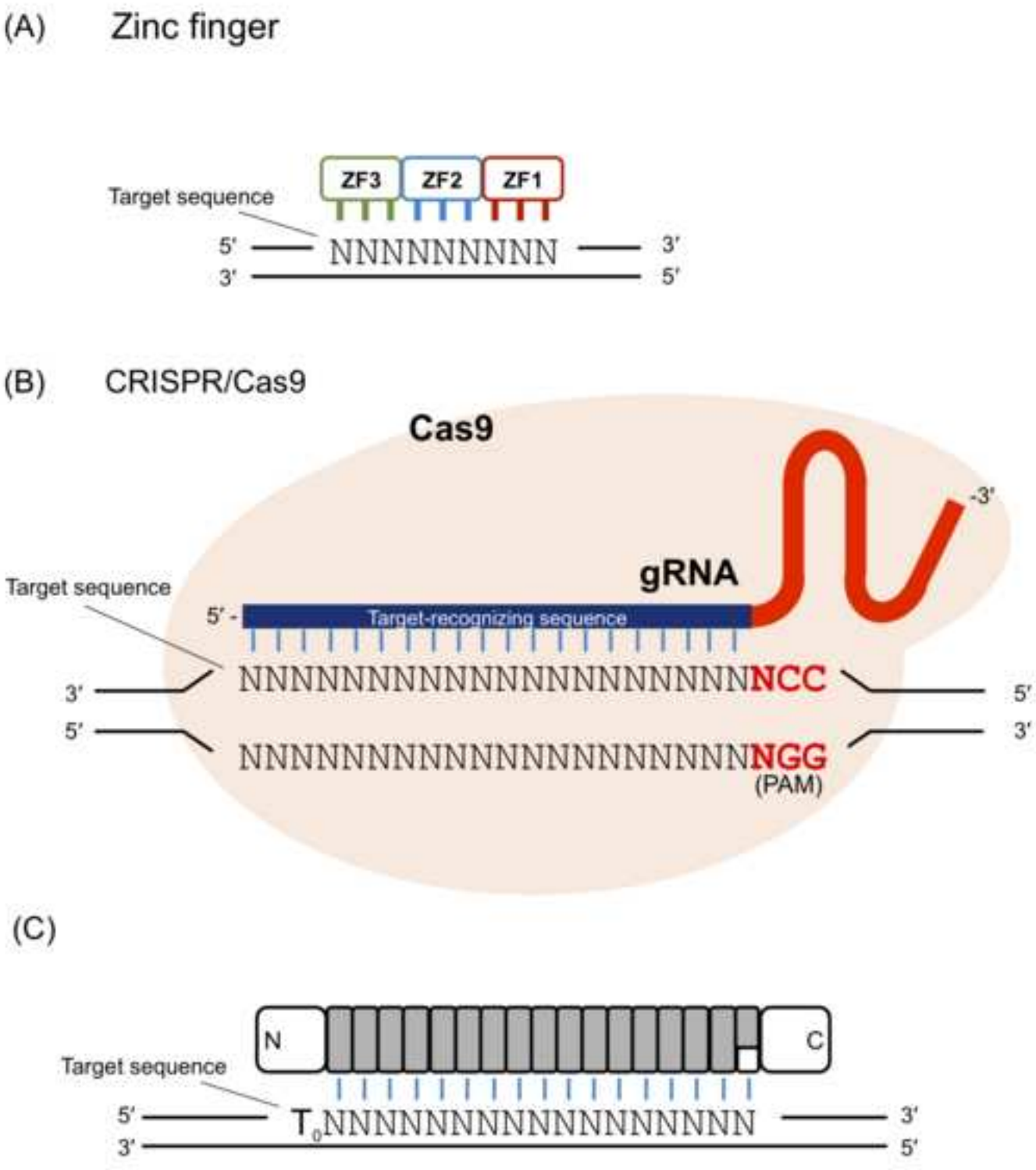


Figure2

Click here to download high resolution image

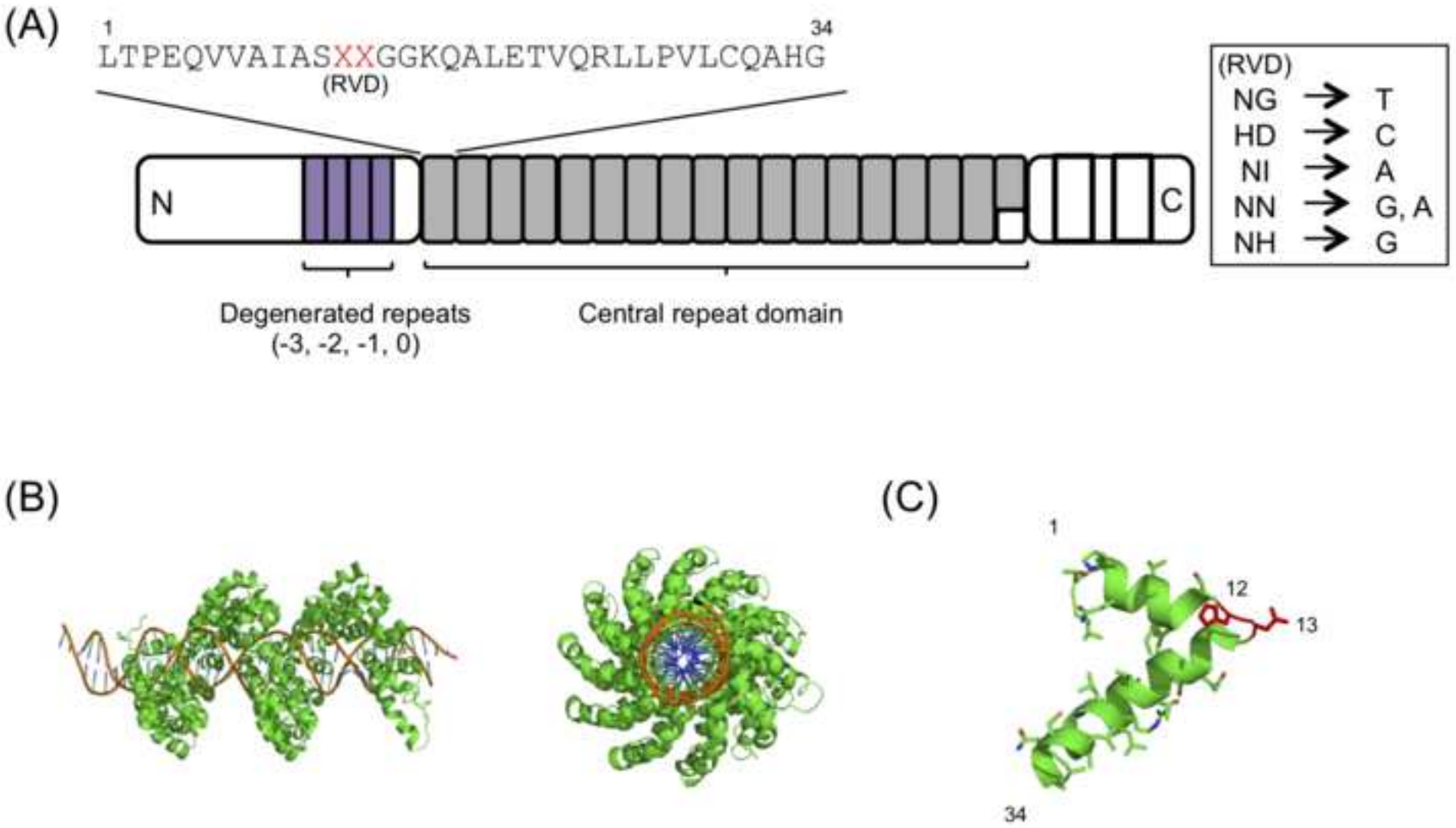


Figure3
Click here to download high resolution image

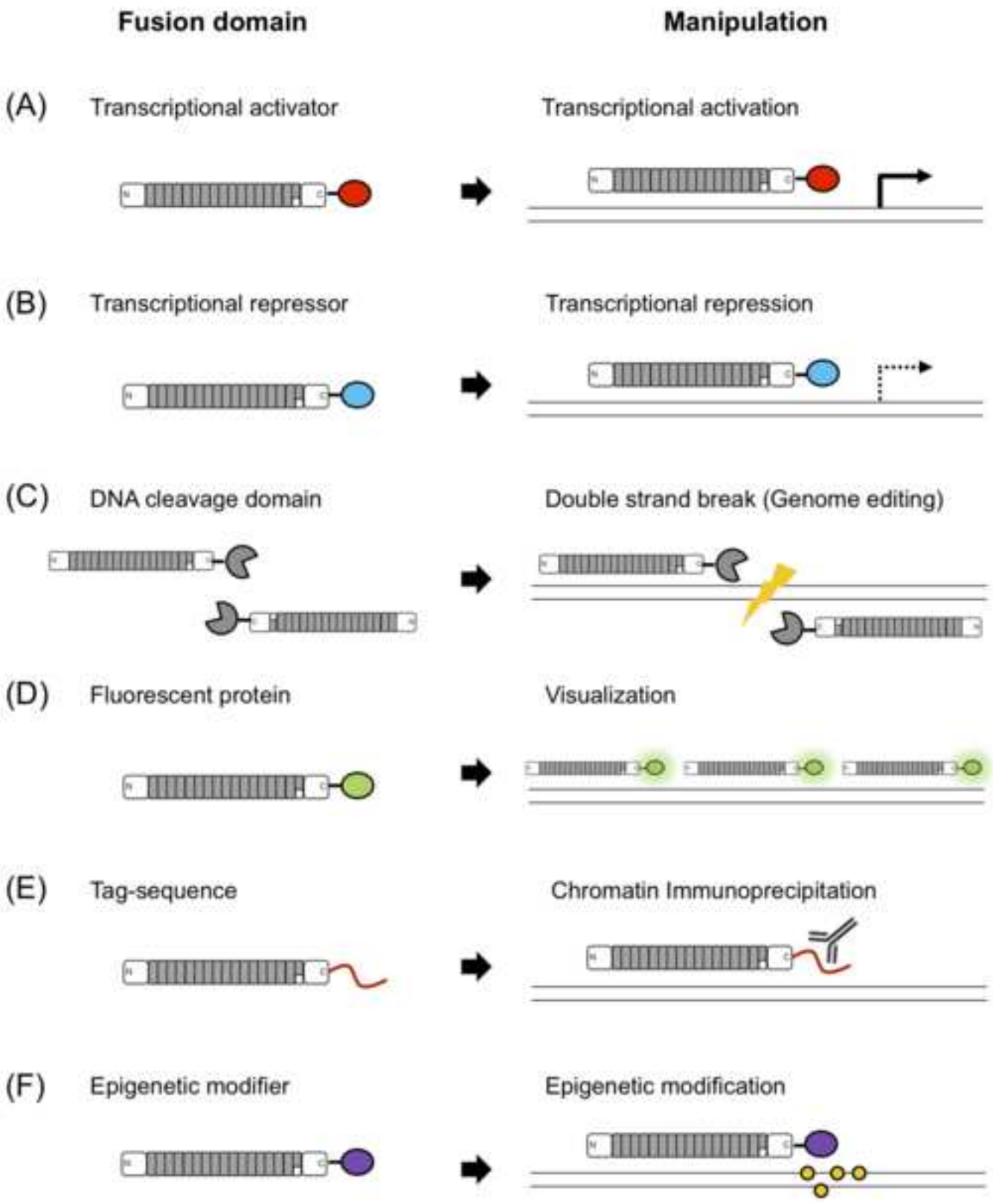


Figure4

Click here to download high resolution image

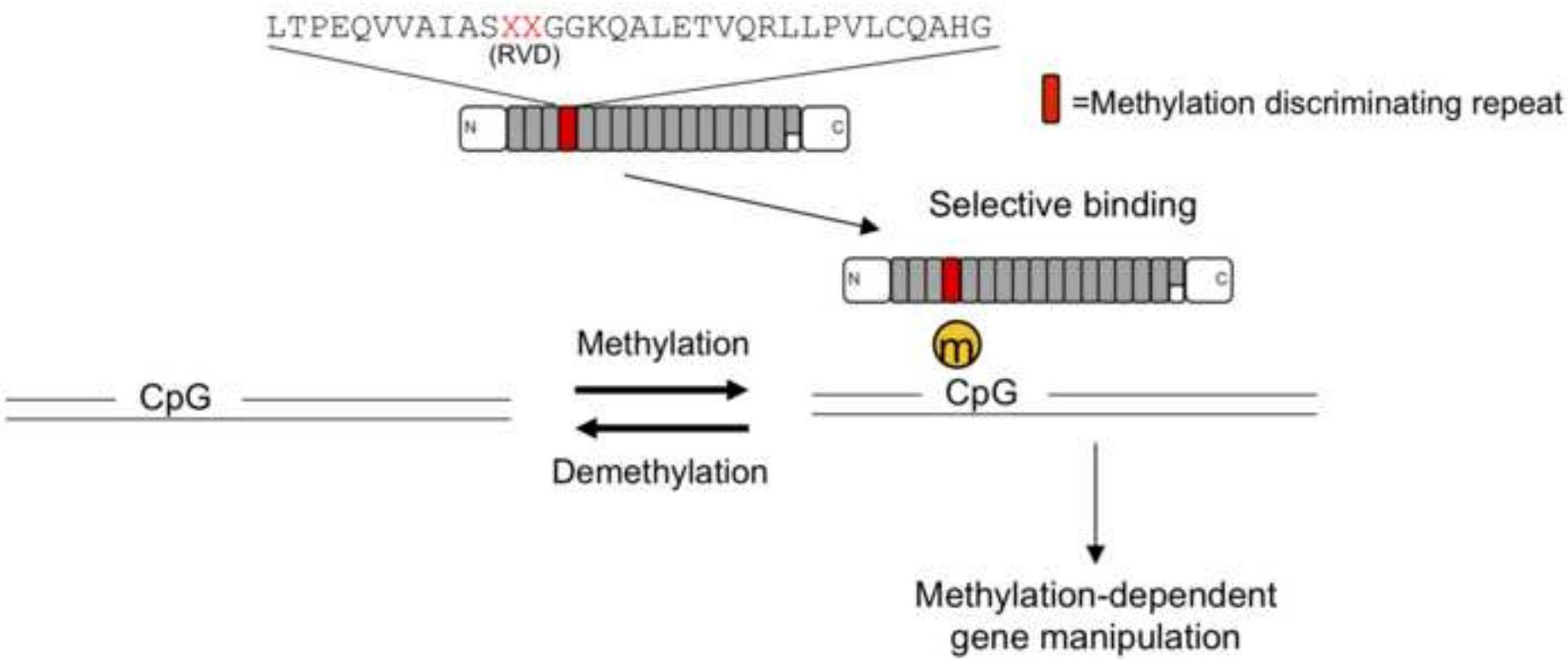


Figure5

Click here to download high resolution image