Epigenome editing ; application of genome editing tools

2016.06.09

M1 Yamaji Kyohei

Contents

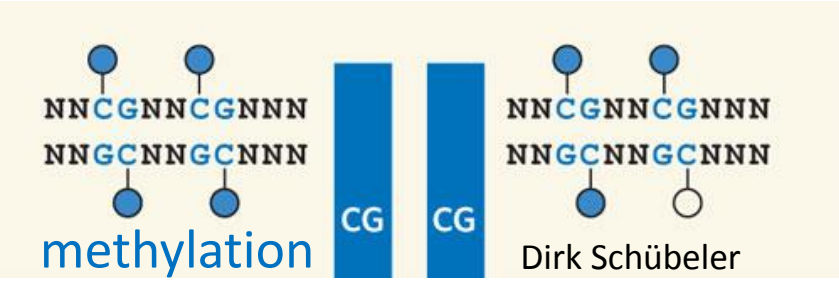
Epigenetics ; chemical modifications

DNA/Histone modifications can control its gene expression patterns.

DNA methylation pattern is different between stem cells and differentiated cells

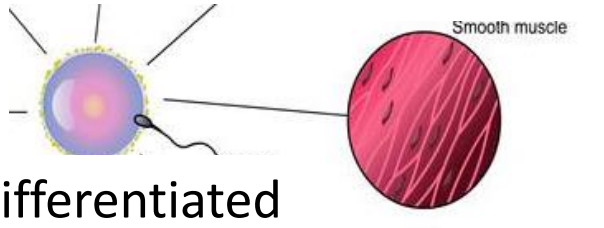
stem cells

Differentiated cells



Dirk Schübeler
Nature **462**, 296-297

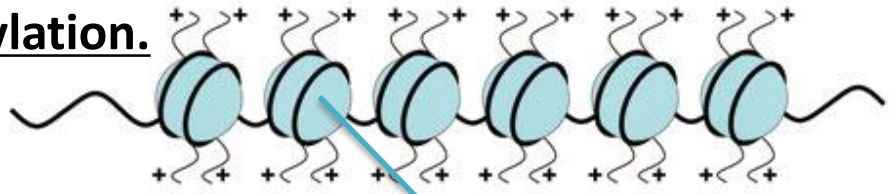
differentiated



undifferentiated

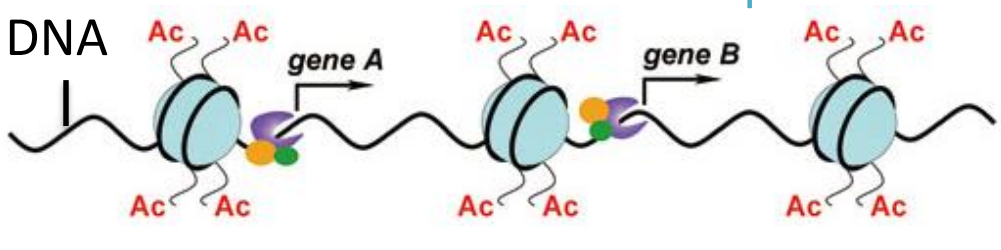
Expression of some genes is suppressed in differentiated cells with reduced global Histone acetylation.

Differentiated cells



Expression of gene A and B

stem cells



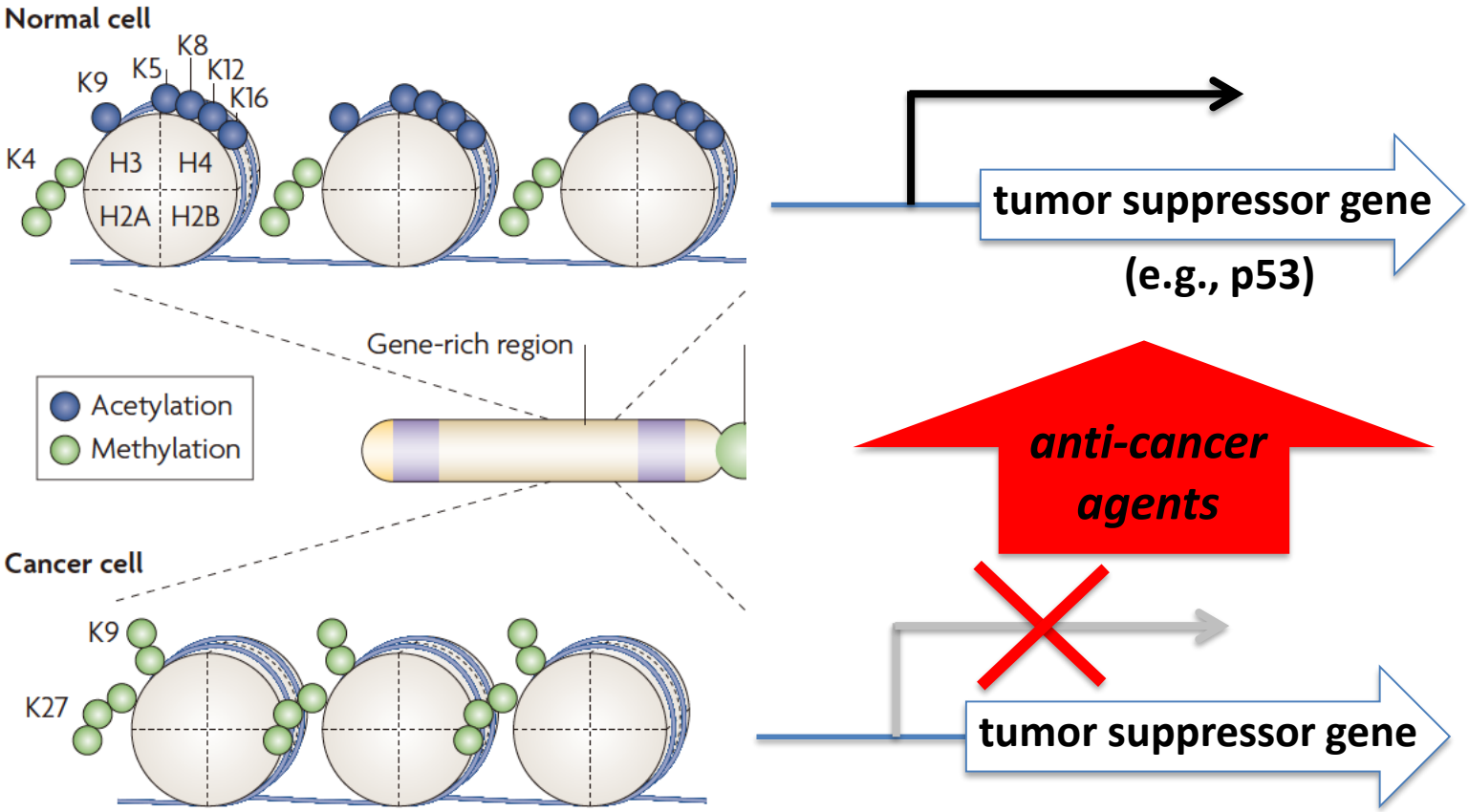
Expression of gene A and B

Loring, J F, *et al. Cell Research* (2014) **24**:143–160

Catalysis medicine

In some cancer cells, tumor suppressor genes are suppressed with decreased Histone acetylation.

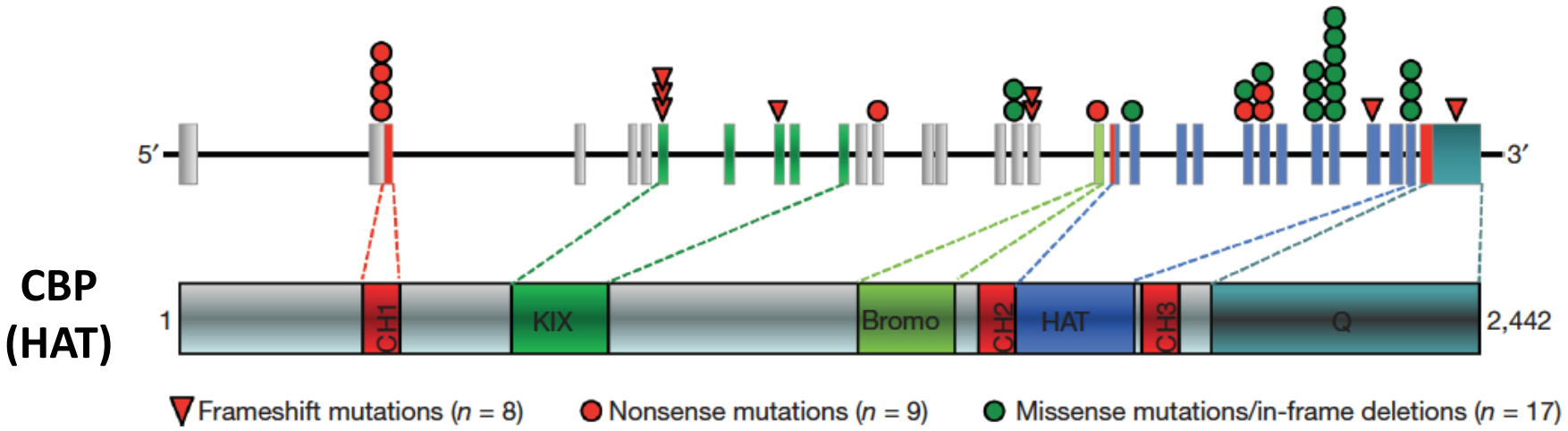
Inducing Histone acetylation can be a hopeful anti-cancer strategy.



Manel Esteller (2007)

Cataysis medicine

Mutations of HAT are frequently found in B-cell lymphoma.



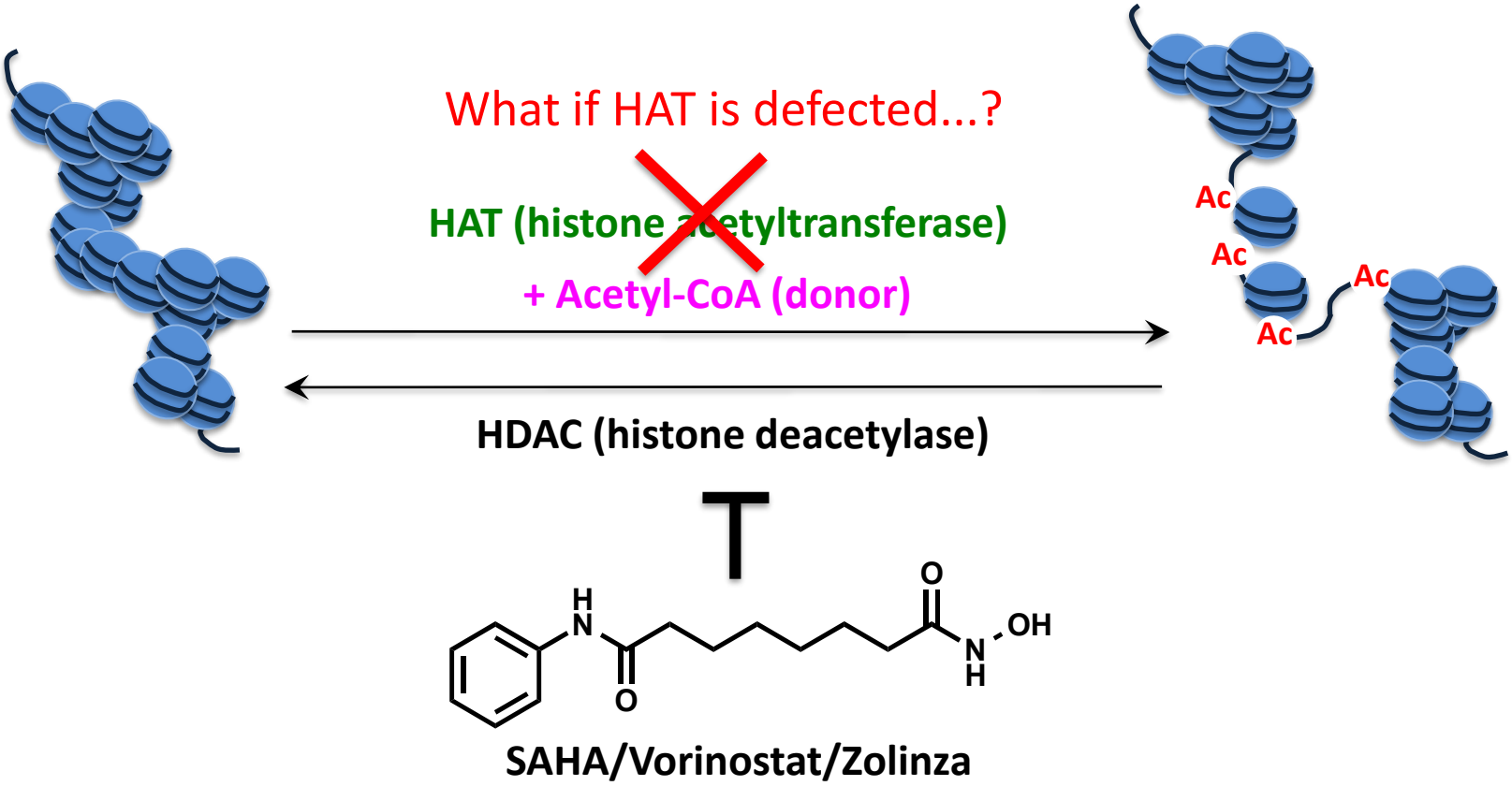
Inactivating mutations of acetyltransferase genes in B-cell lymphoma

Pasqualucci, L. *et al. Nature* 2011, *471*, 189.

HAT-independent Histone acetylation can be the solution?

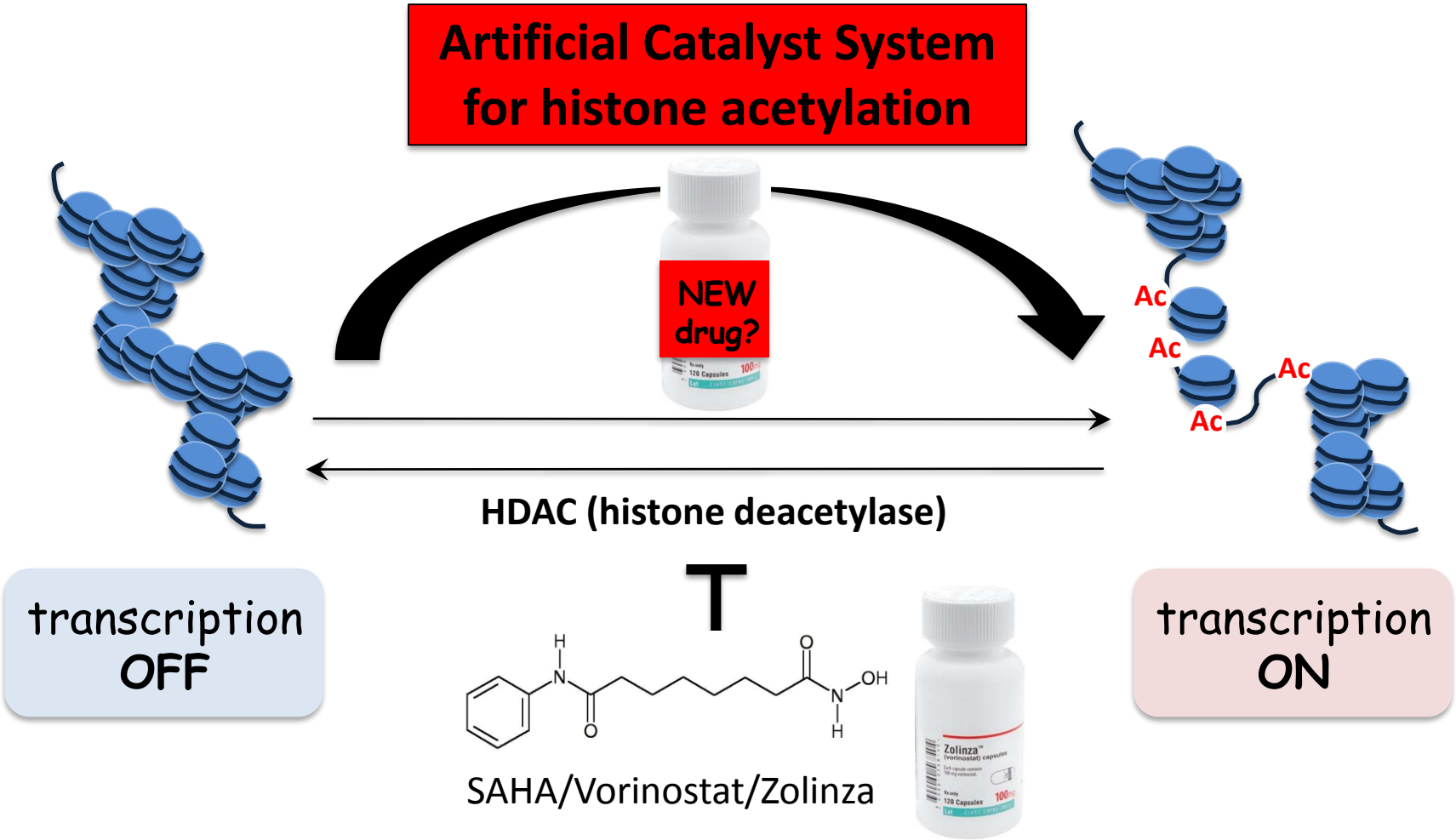
Cataysis medicine

Histone acetylation level is regulated on the balance of two catalysts called HAT and HDAC.



HDAC inhibitor is actually used as an anti-cancer drug.

HAT-independent histone acetylation by artificial catalyst system

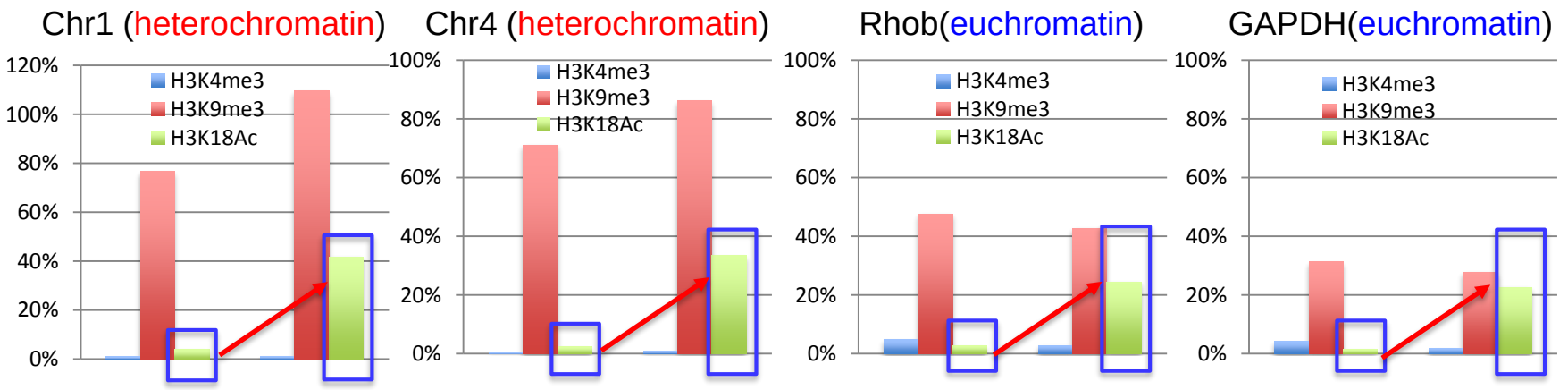


Current catalysis medicine

SynCAc acetylate genome-wide histones nonspecifically

The result of *In nucleo* ChIP (No compound vs SynCAc) by Dr.Ishiguro

•HeLa nuclei were acetylated with 50μM 8DMAP, 10μM 3NMD-8R @25℃ for 3hrs in SynCAc.
(8DMAP : batch-C by Dr. Amamoto, 3MMD-8R : KY2457 by Dr.Yamatsugu)



Both hetero/euchromatin are targets of SynCAc.
(heterochromatin acetylation might be stronger)

Without targeting, genome-wide histone acetylation is observed

Gene targeting and catalysis medicine

The usefulness of Gene targeting of artificial catalyst

- **medical application**
- **applications as biological tools**

Applications of artificial catalyst as biological tools

- **roles of various acylations on histone**
- **roles of modifications on various residues of histone**

→internal controls are easily obtained in targeting of catalyst

For further applications…

- **genome-wide screening (acetylation)**

And etc…

Gene targeting can expand the possibility of artificial catalyst.

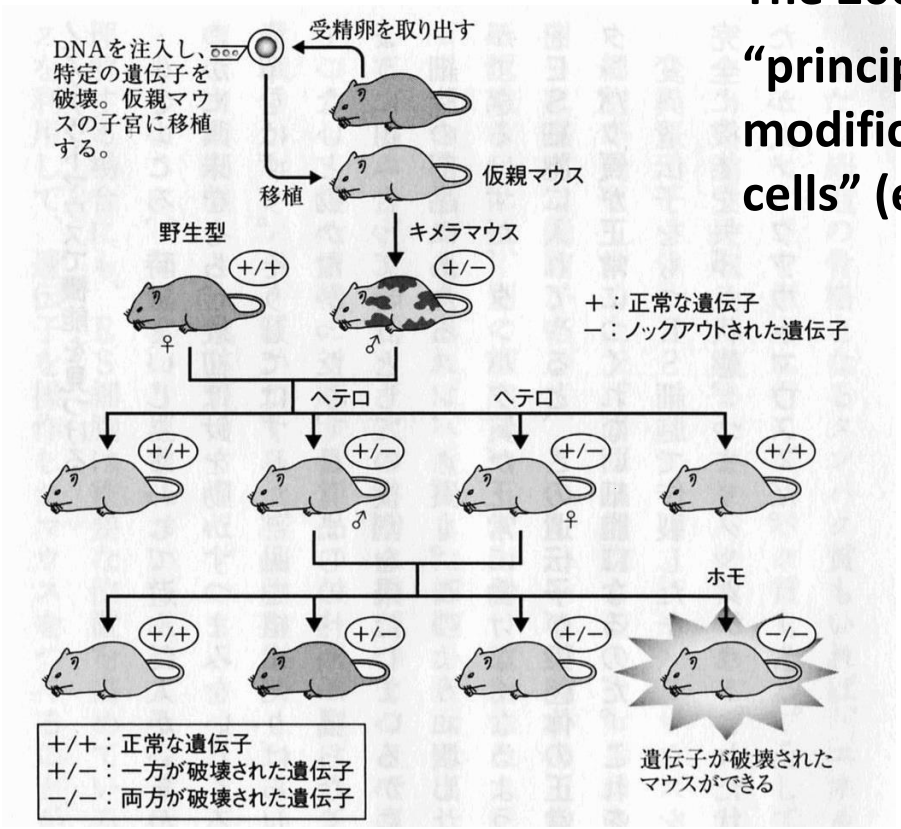
Contents

The utility of Genome editing

Genome editing

Genetic engineering in which DNA is inserted, deleted, or replaced in the genome of an organism using engineered nucleases, or "molecular scissors." (wikipedia)

General procedures for producing knockout/knockin mice



The 2007 Nobel Prize in Physiology or Medicine

"principles for introducing specific gene modifications by the use of embryonic stem cells" (established : 1989)

"the beginning of new era in genetics"

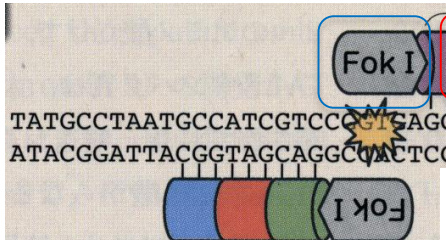
However…

Takes so much time and trouble

Genome editing tools have made this drastically easier

Genome editing tools

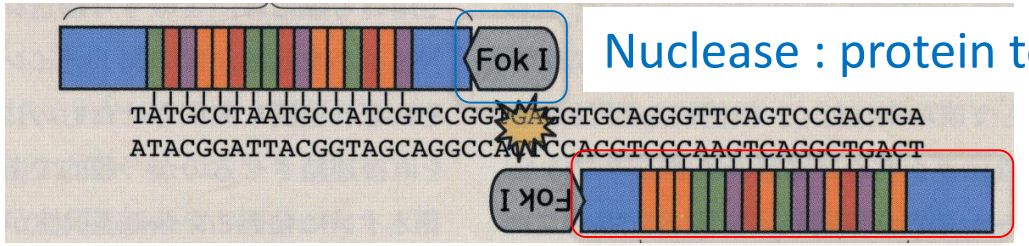
1st generation : ZFN (zinc finger nuclease)



ZF modules : protein to bind DNA

Nuclease : protein to cleave DNA

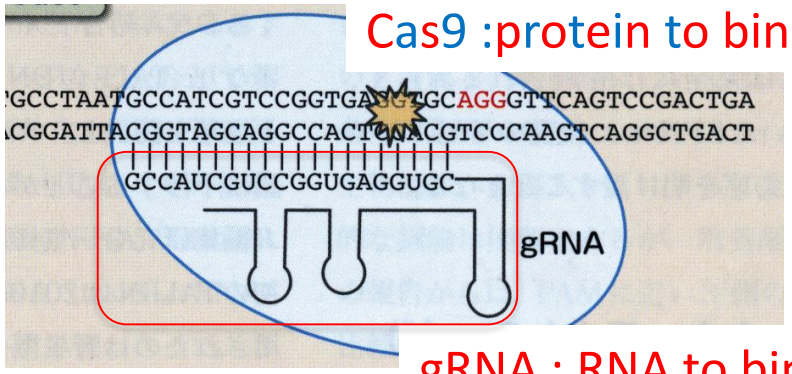
2nd generation : TALEN (transcription-activator-like effector nuclease)



Nuclease : protein to cleave DNA

TALE : protein to bind DNA

3rd generation : CRISPR/Cas9



Cas9 :protein to bind and cleave DNA

gRNA : RNA to bind DNA

Genome editing tools…

recognize specific sequences

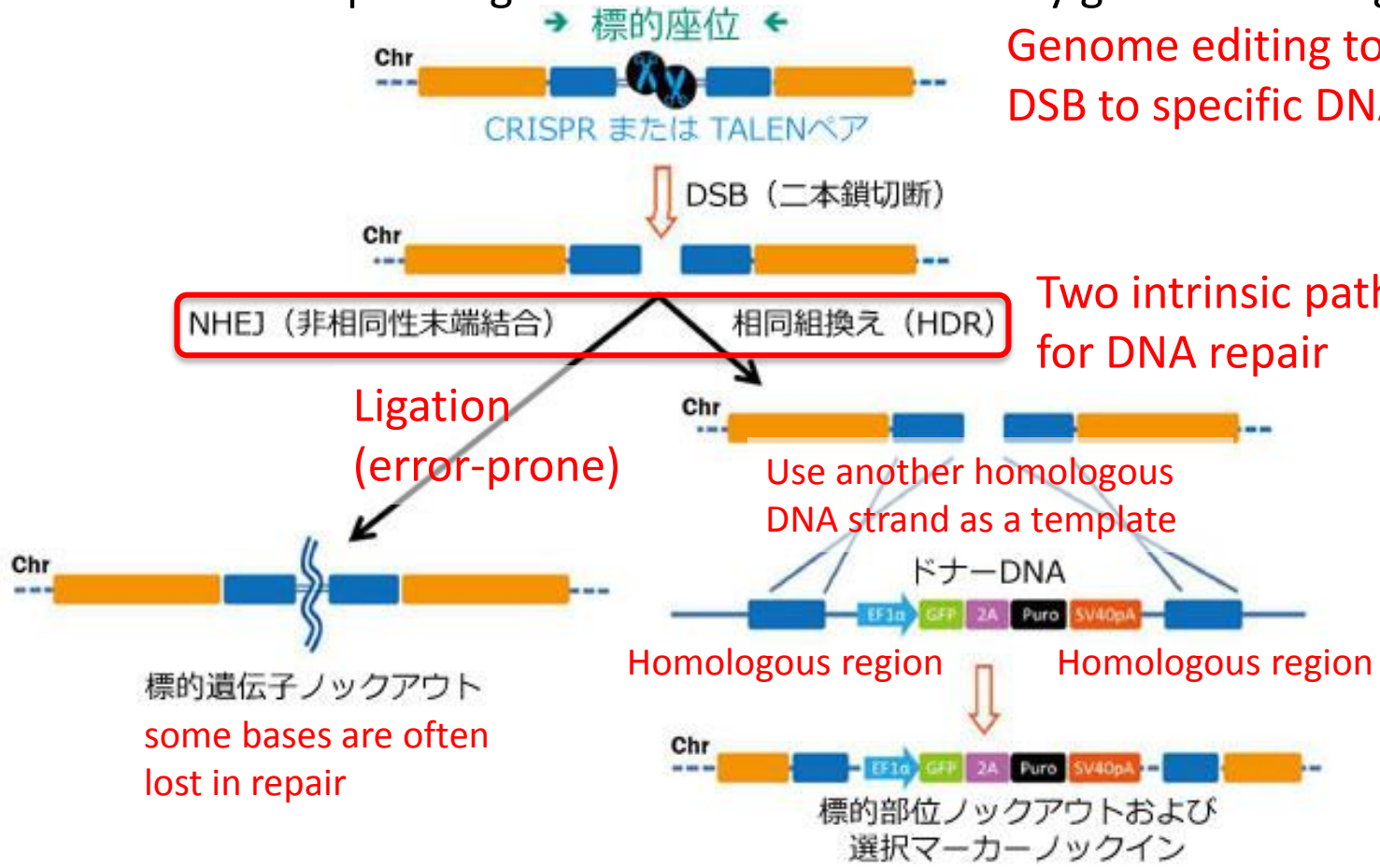
cleave DNA strand

実験医学別冊　今すぐ始めるゲノム編集　TALEN＆CRISPR.Cas9の必須知識と実験プロトコール　山本卓編

Principle of genome editing

genome editing tools promote DNA repair dependent mutations to specific DNA regions

Scheme for the specific gene knockout or knockin by genome editing tools



Genome editing tools introduce DSB to specific DNA region

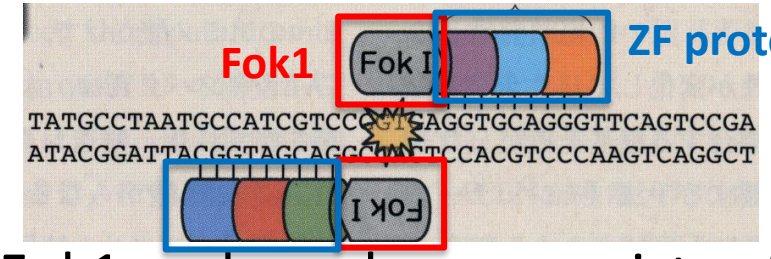
Two intrinsic pathways for DNA repair

https://www.cosmobio.co.jp/support/technology/a/crispr-talen.asp

Contents

Fok1 endonuclease

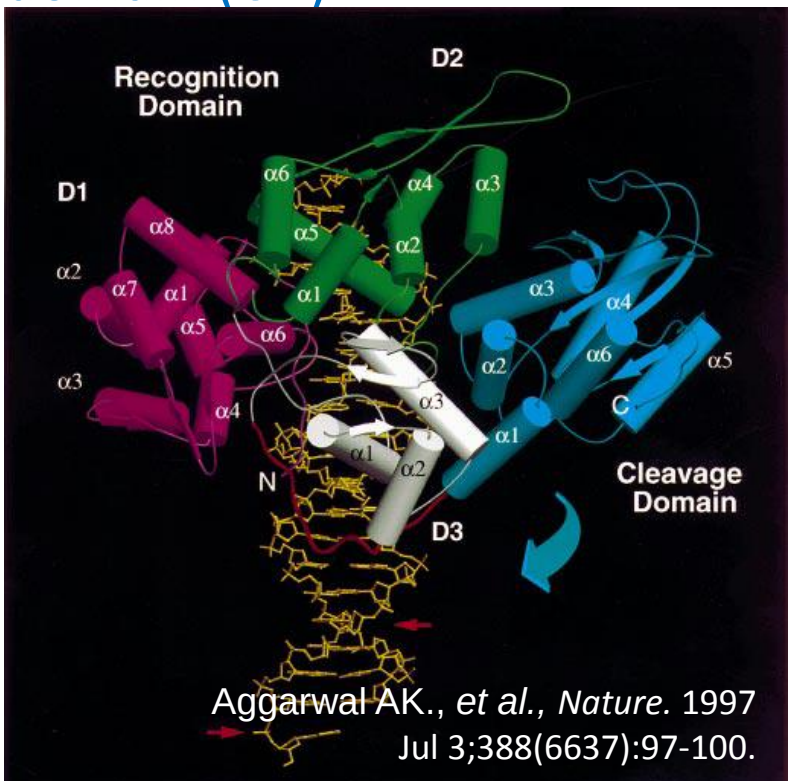
ZFN consists of Fok1 endonuclease cleavage domain and ZF protein.



ZFN needs to be designed to put the cleavage site in between

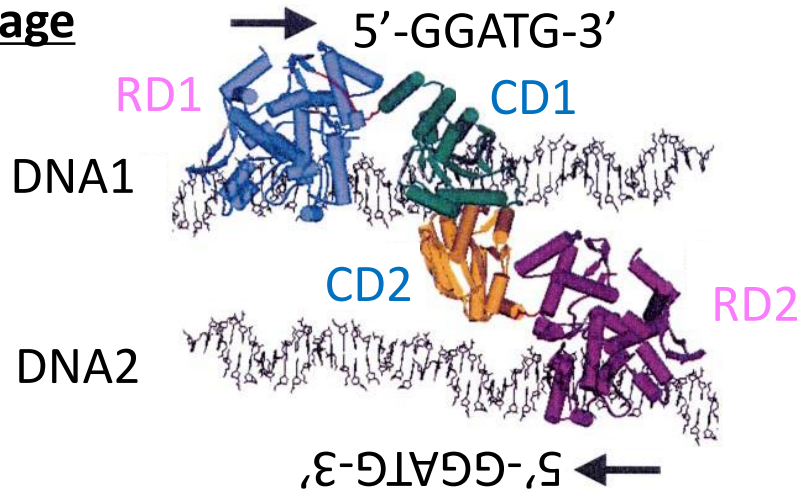
Fok1 endonuclease consists of Recognition domain(RD) and Cleavage domain (CD)

Recognition domain binds 5'-GGATG-3'
(natural Fok1 is sequence specific)



Aggarwal AK., *et al., Nature.* 1997 Jul 3;388(6637):97-100.

Cleavage domain dimerization is necessary for the cleavage

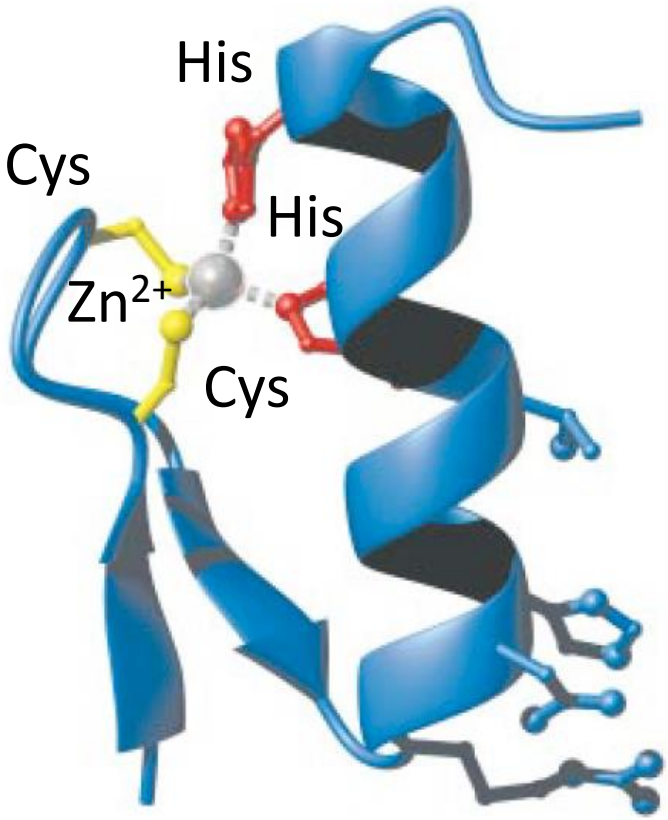


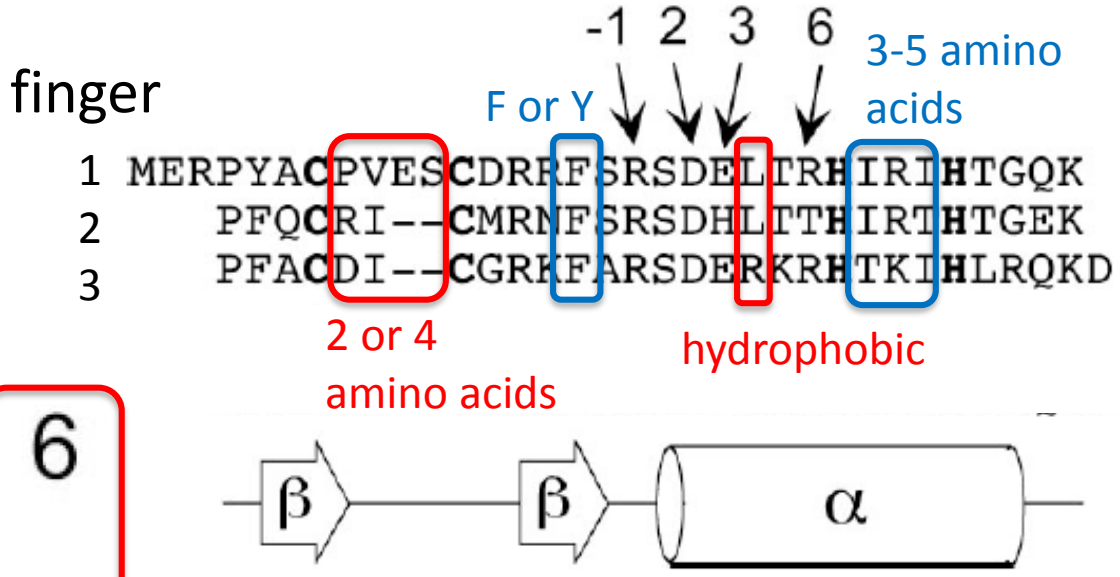
Aggarwal AK., *et al., J. Mol. Biol.* (2001) 309, 69±78

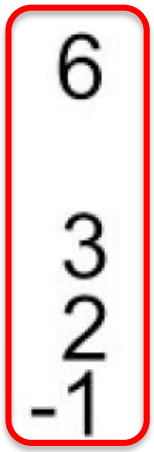
structure of zinc finger protein

ZF motif contains Zn, which interacts with His2Cys2 and stabilizes ββα

Structure of finger 2 from Zif286, a kind of ZF motifs







Potential DNA binding residues

(numbered from the start of α helix)

Residues closest to DNA binds DNA

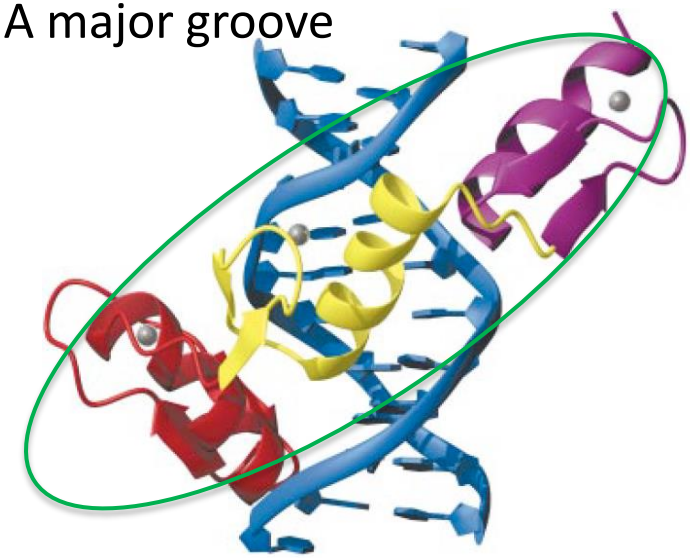
(distance depends on the structure of ZF)

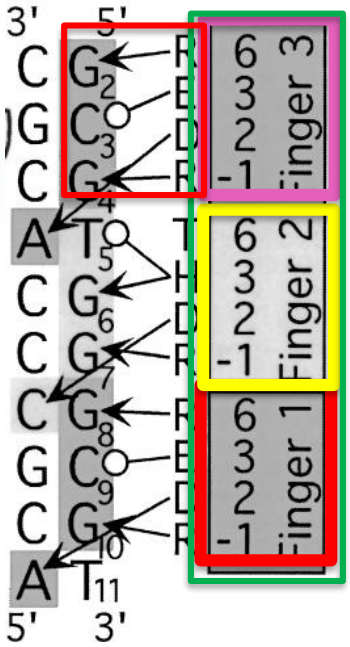
Pabo *CO., et al., Annu. Rev. Biophys. Biomol. Struct*. 2001. 70:313–40

DNA recognition by zinc finger protein

How ZF motifs recognize DNA

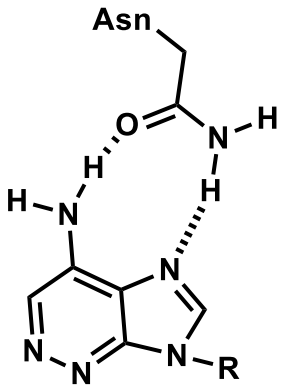
ZF motifs interact with DNA major groove



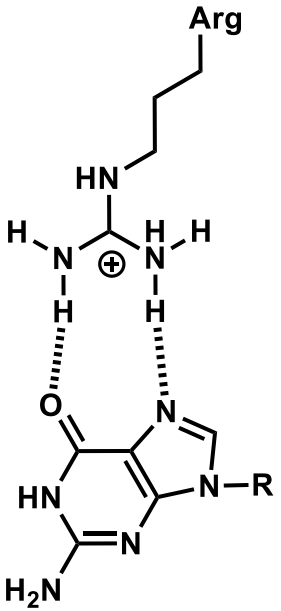


One residue recognizes one nucleobase

Asparagine-adenine



Arginine-guanine



For sufficient sequence specificity, 3-6 motifs are usually combined.

Pabo *CO., et al., Annu. Rev. Biophys. Biomol. Struct*. 1999. 3:183–212
Pabo *CO., et al., Annu. Rev. Biophys. Biomol. Struct*. 2001. 70:313–40

Stabilizing residues of zinc finger-DNA complex

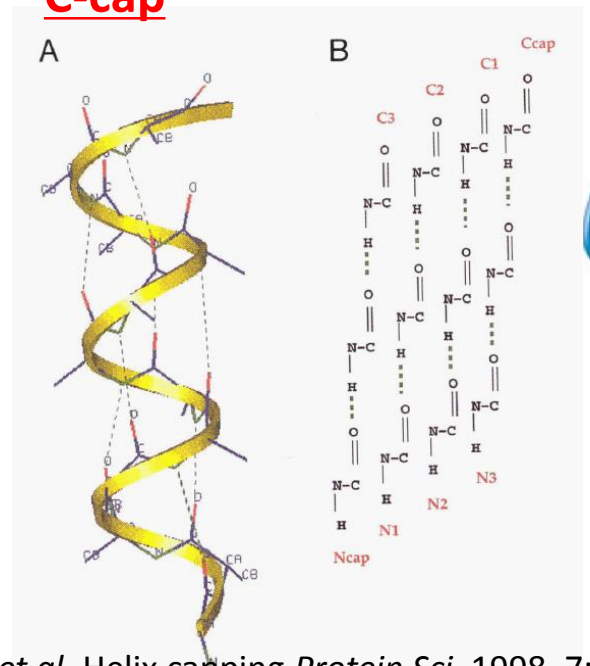
Other interactions between ZF motifs and DNA

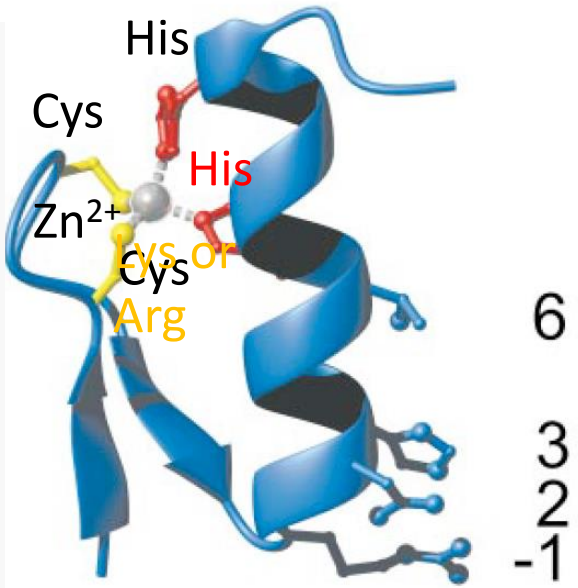
a. His (Nδ)-phosphate of the DNA backbone (5'- primary strand)
b. Lys or Arg-phosphate of the DNA backbone (the same phosphate as a. or its 3'-strand)
c. Lys in the linker (TGEKP)-phosphate of the DNA backbone (5'- primary strand)

The linker (TGEKP) contributes to the stability of ZF-DNA complex

T : make hydrogen bond with NH in the next Helix (C-cap)
G : complete the C-cap
E : stabilize finger-finger interaction (contact R of HXRXH in the next finger)
K : contact DNA backbone
P : π–π stacking with -1 residue in the next finger

C-cap





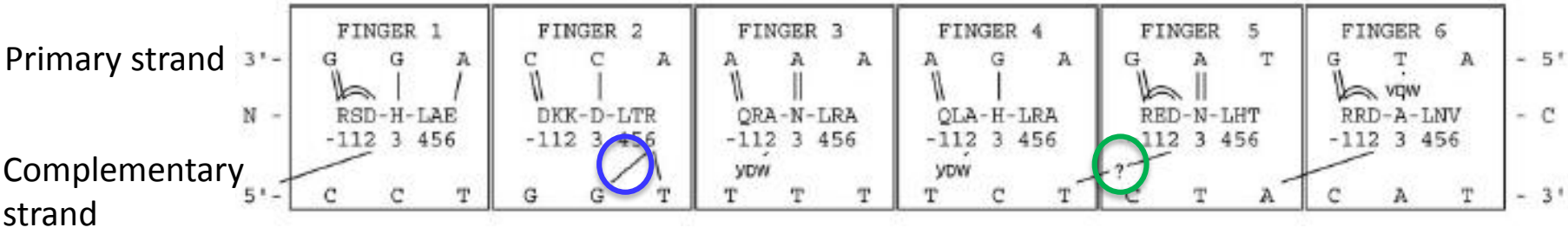
Rose GD., *et al.*,Helix capping.*Protein Sci.* 1998, 7:21-38

Drawback of zinc finger protein

drawback of ZFN

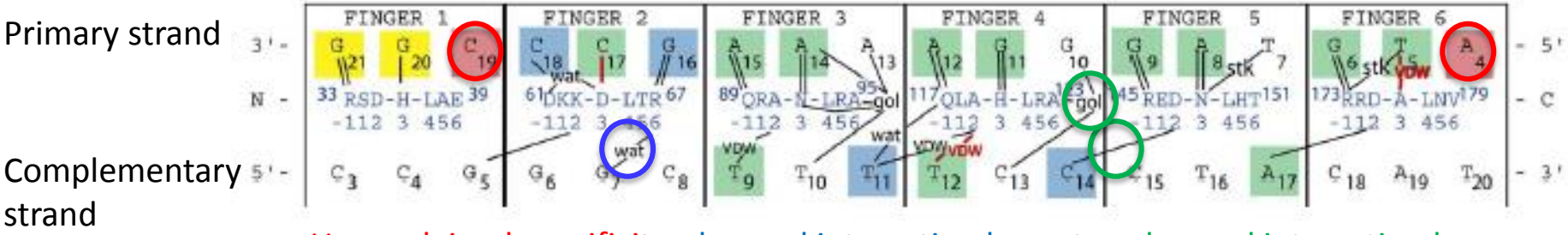
Assembled motifs often fail to recognize predicted sequence

Predicted sequence : GGA-CCA-AAA-AGA-GAT-GTA



actual sequence : GGC-CCG-AAA-AGG-GAT-GTA

Weak specificity



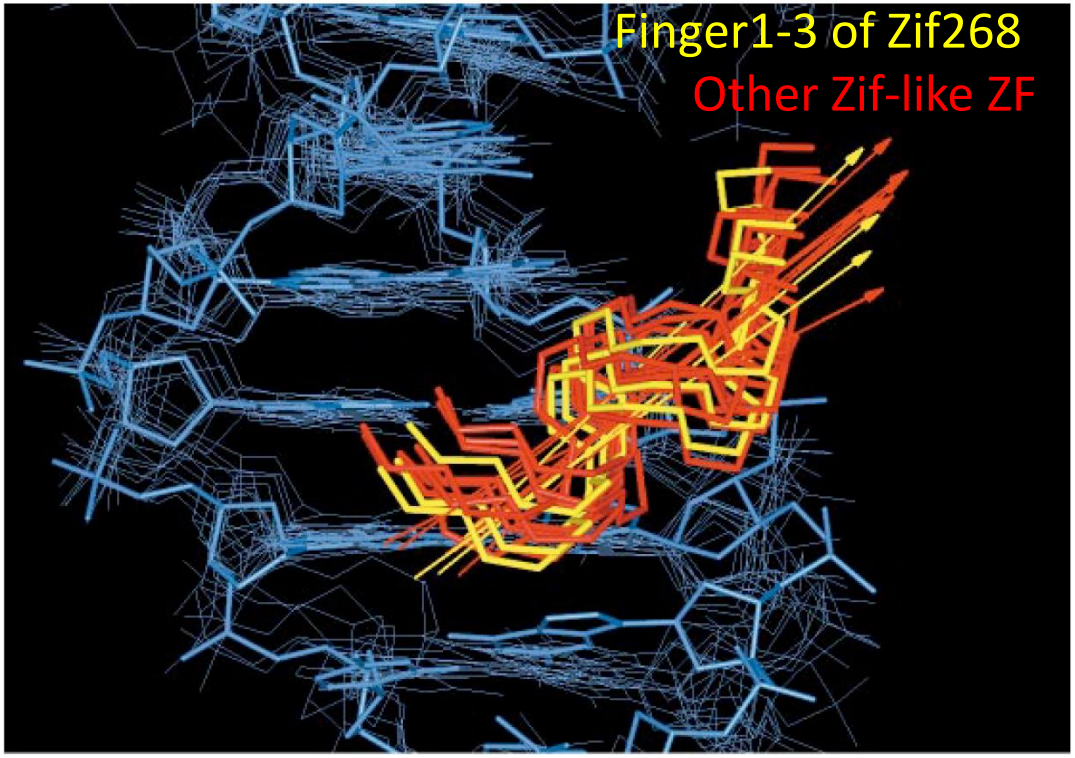
Unexpalained specificity, changed interaction by water, changed interaction by glycerol(gol, necessary for crystallization, substitution of water?)

Horton, NC. *et al., J. Mol. Biol.* (2006) 363, 405–421

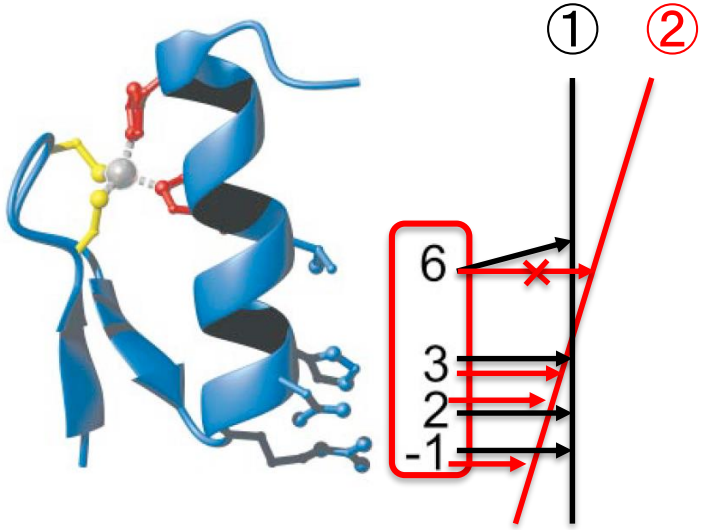
Structural insight behind the DNA recognition by ZF

reason for drawback of ZFN

The broad range of variation in the docking arrangement of ZF-DNA complex



Pabo *CO., et al., Annu. Rev. Biophys. Biomol. Struct*. 2001. 70:313–40



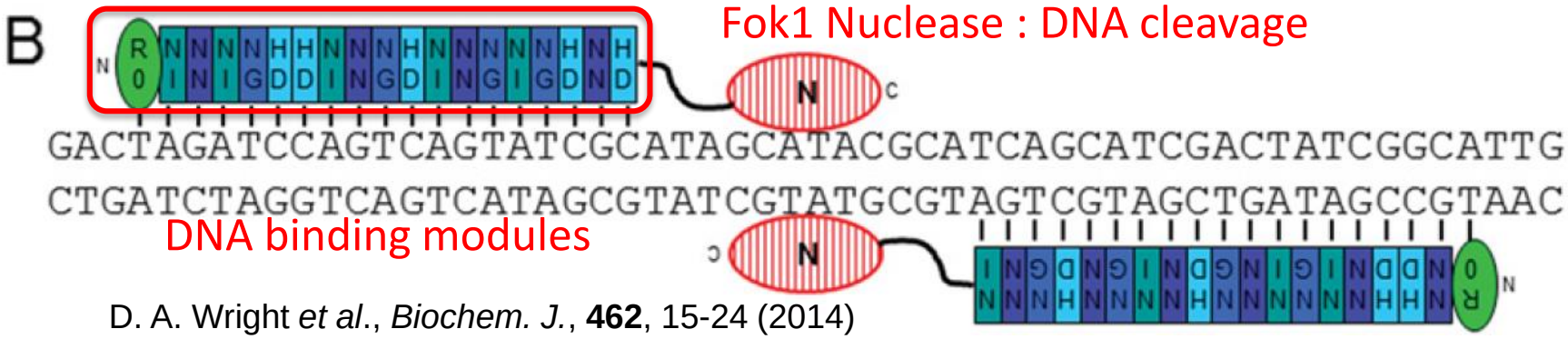
The change in docking arrangement sometimes leads to the loss of base recognition by some residues

Modular assembly may change the docking arrangement

Contents

Overlook on TALEN

TALEN consists of Fok1 nuclease domain and TALE (DNA binding protein)



D. A. Wright *et al*., *Biochem. J.*, **462**, 15-24 (2014)

TALEN needs to be designed to put the cleavage site in between like ZFN

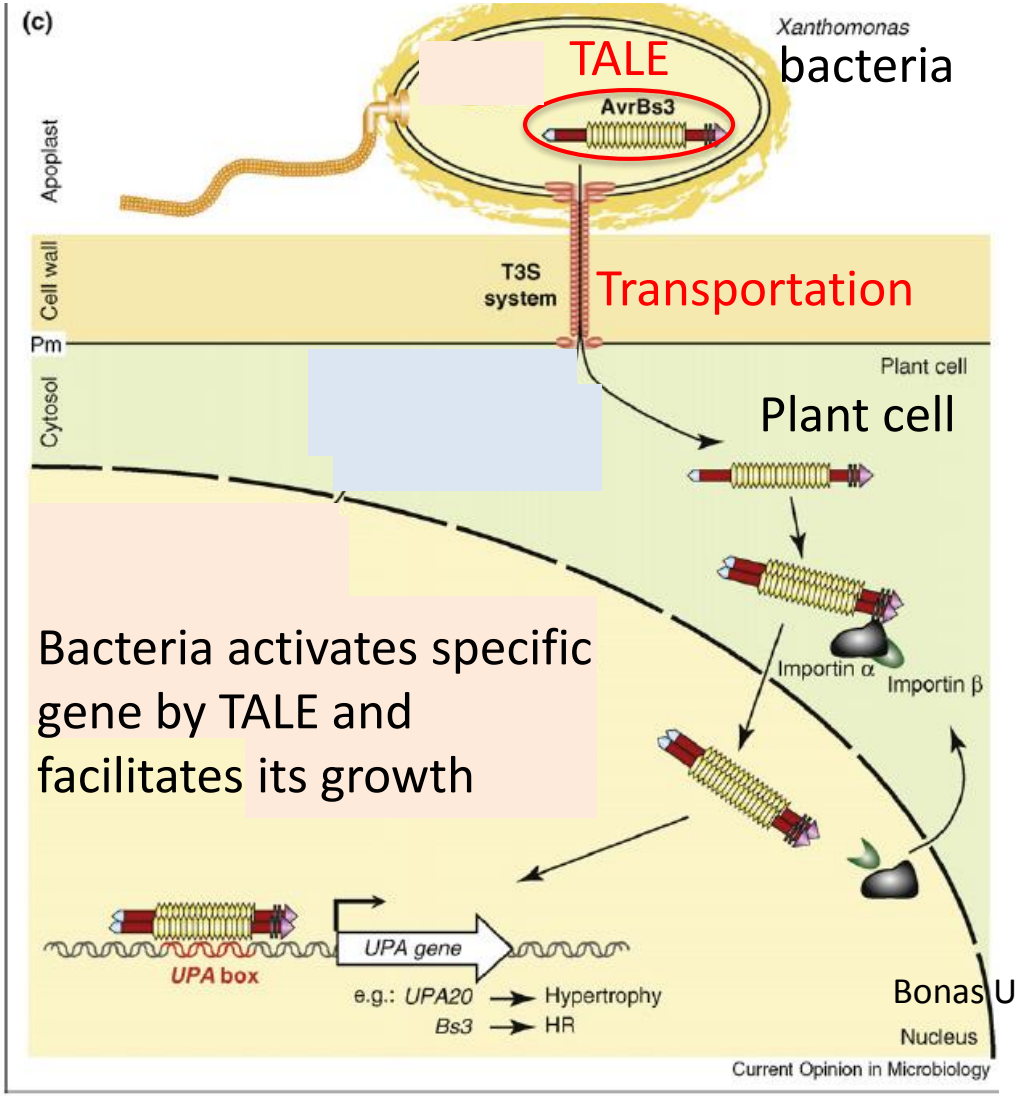
TALEN has overcome ZFN in terms of its easiness in designing and specificity

Predicted sequence is (almost) always recognized. (drawback of ZFN is solved)
Only restraint is that the target sequence has to start from thymine
High specificity (many more bases can be recognized than the other two tools)

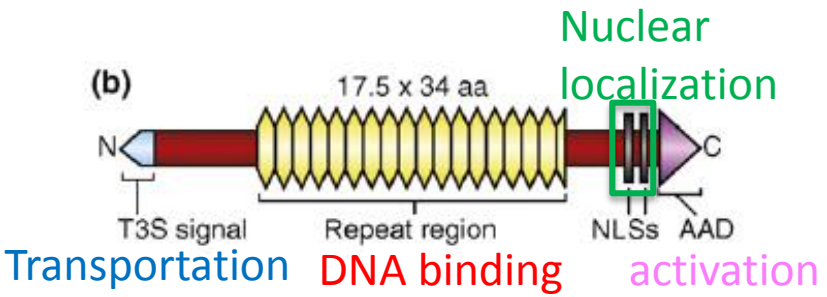
Easy to design, usable to various sequences, and high accuracy of targeting

TALE in natural world

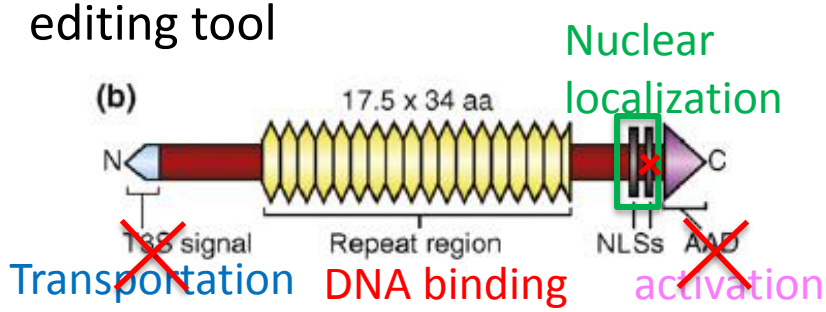
TALE (Transcription-activator like effector) in plant infectious bacteria



Schematic view of domains in TALE



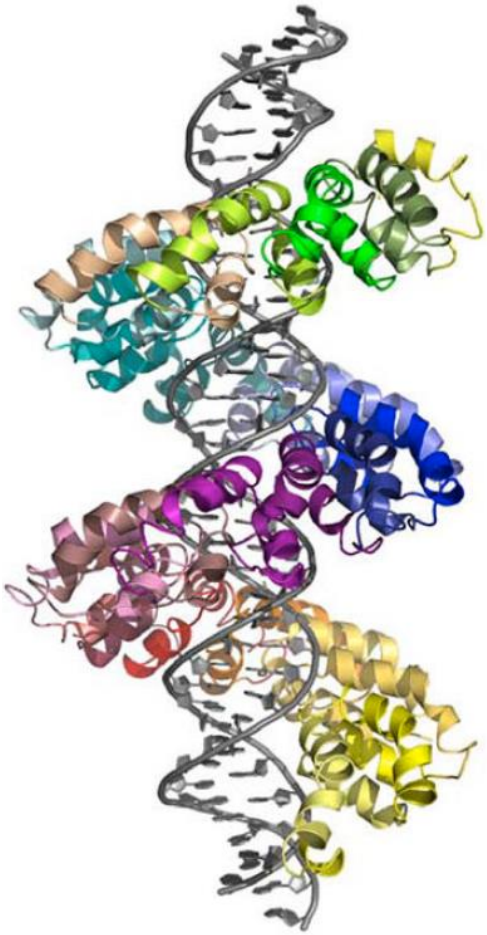
TALE in application as a genome editing tool



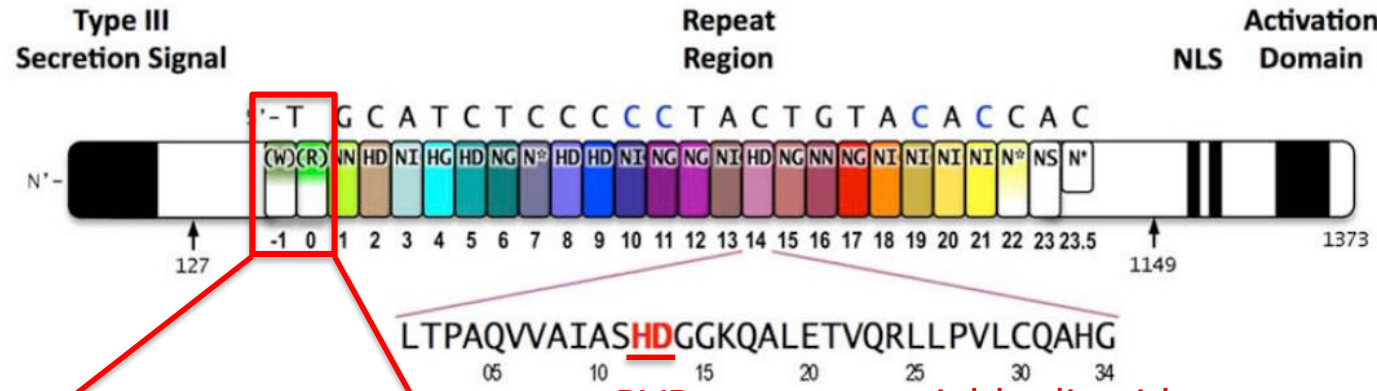
Bonas U. *et al., Curr Opin Microbiol*. 2009 Feb;12(1):37-43.

Structure of TALEN

Mechanisms for DNA binding

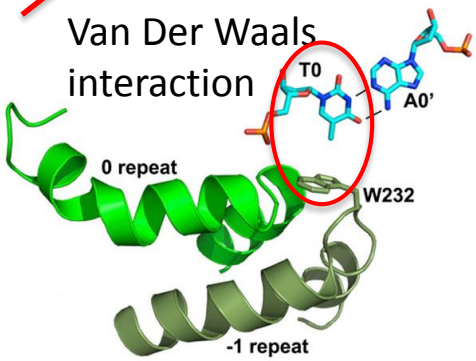


Schematic view of TALE sequence



RVD : repeat-variable diresidue

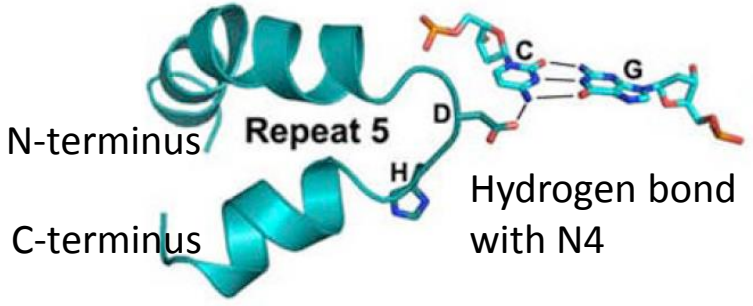
Van Der Waals interaction



The repeat (0, -1) recognizes Thymine

'HD' recognizes Cytosine

Van Der Waals interaction



Hydrogen bond with N4

Interactions between TALE and DNA major groove

A. N. Mak *et al.*, *Science*, **335**, 716-719 (2012)

DNA sequence recognition by TALE

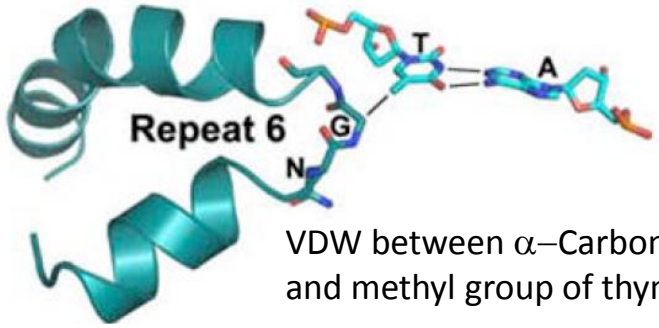
RVDs and their specificity

Interact with base

Stabilize TALE

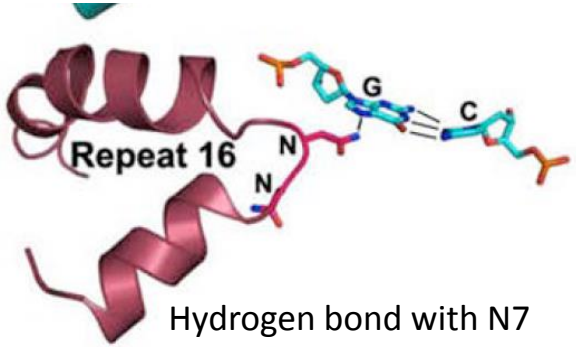
Bogdanove AJ., *et al., Science.* 2009 Dec 11;326(5959):1501.

Interaction between 'NG' and Thymine



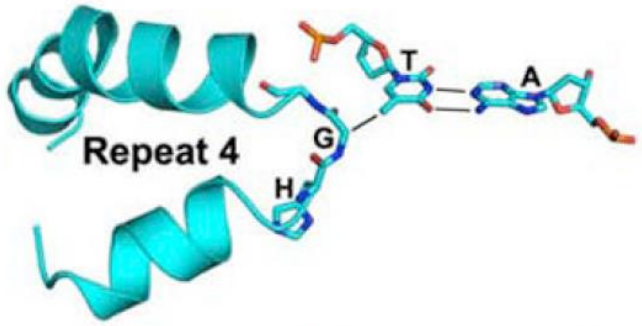
VDW between α–Carbon of Gly and methyl group of thymine

Interaction between 'NN' and Guanine



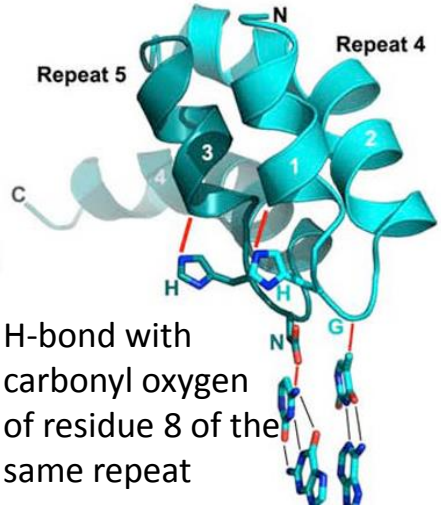
Hydrogen bond with N7 nitrogen of Guanine or Adenine (both purine ring)

Interaction between 'NN' and Thymine



VDW between α–Carbon of Gly and methyl group of thymine

N or H (the 1st of RVD) stabilize the structure



H-bond with carbonyl oxygen of residue 8 of the same repeat

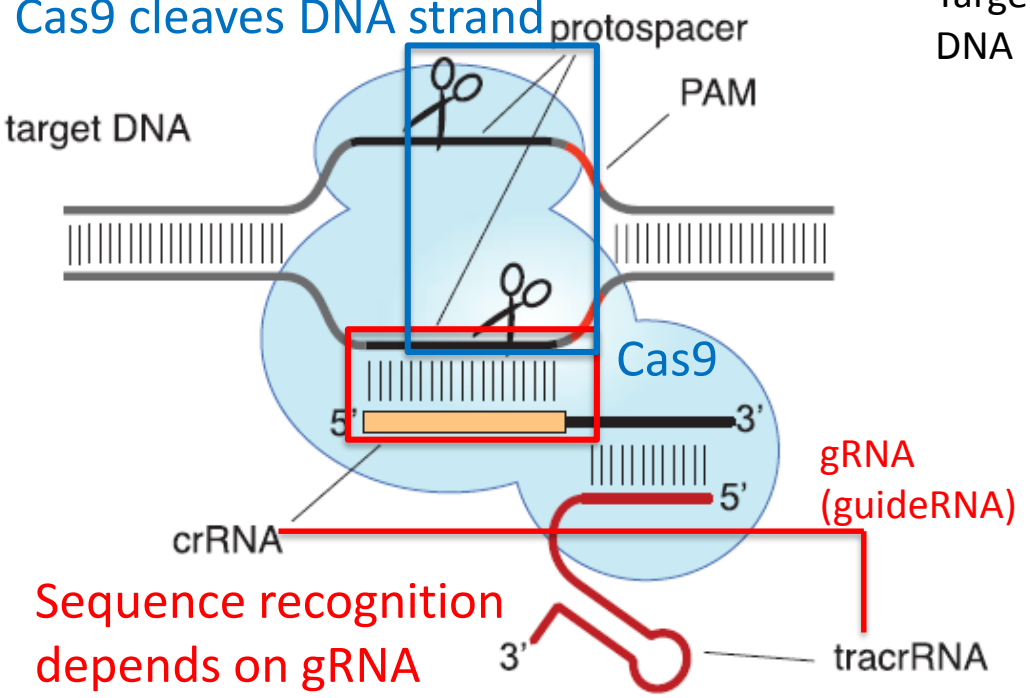
A. N. Mak *et al.*, *Science*, **335**, 716-719 (2012)

Contents

Overlook on CRISPR/Cas9

CRISPR/Cas9 consists of Cas9 protein and gRNA

Cas9 cleaves DNA strand



Jinek, M., *et al.* (2012) *Science,* 337, 816-21

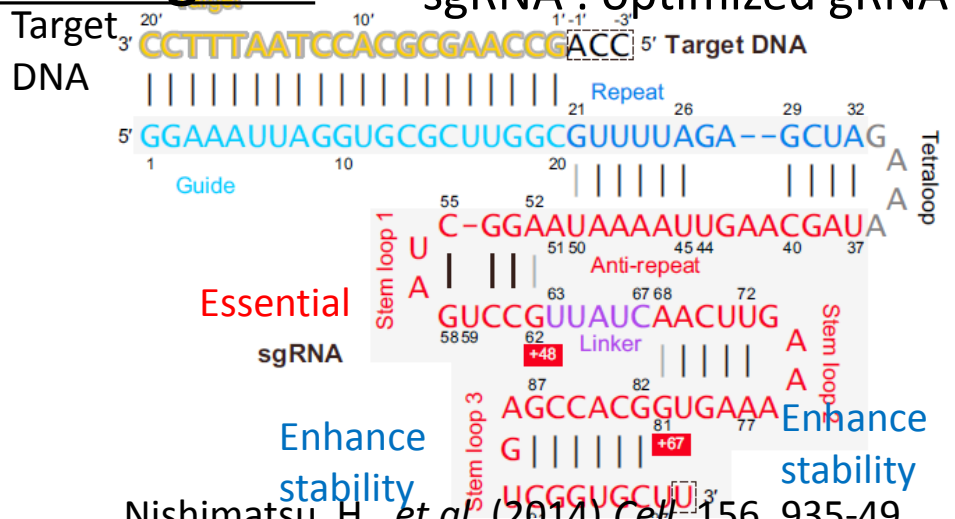
DNA recognition depends on RNA, easily modified

restraint of PAM sequence(only 3~6 base)

Multiplex editing

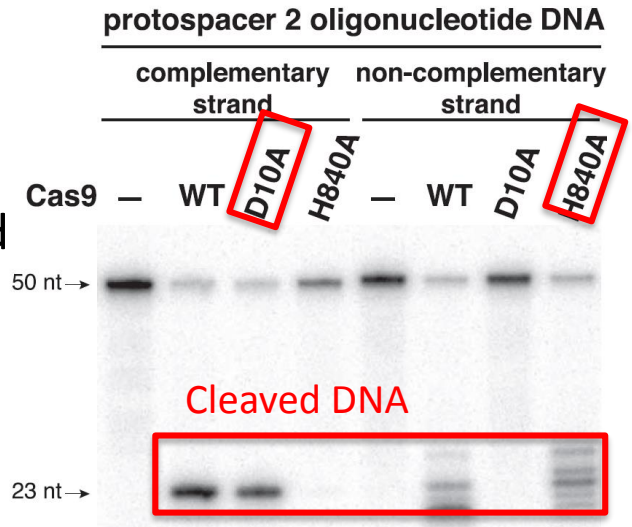
Little cost and effort for design

sgRNA : optimized gRNA



Nishimatsu, H., *et al.* (2014) *Cell,* 156, 935-49

Residues responsible for DNA cleavage

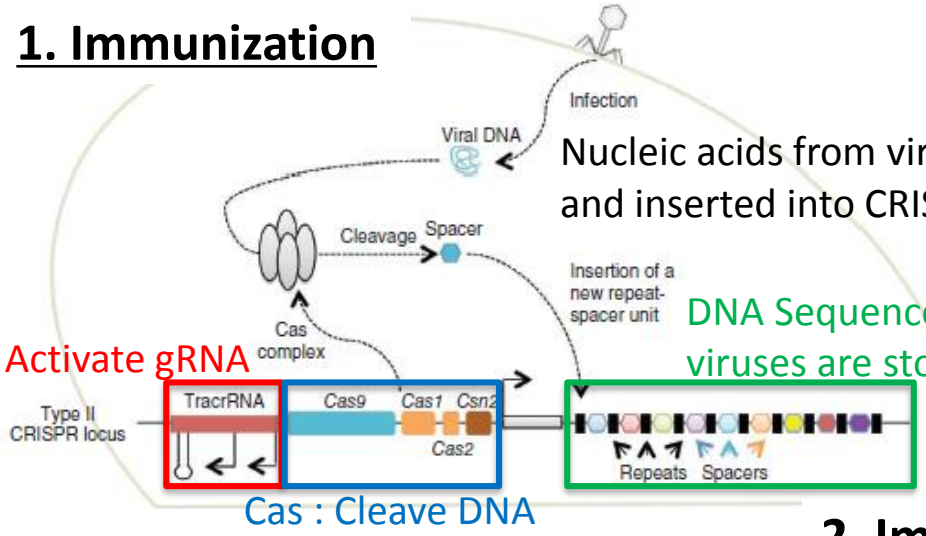


Jinek, M., *et al.* (2012) *Science,* 337, 816-21

CRISPR/Cas9 in natural world

CRISPR/Cas is originally an immune system against nucleic acids from viruses

1. Immunization



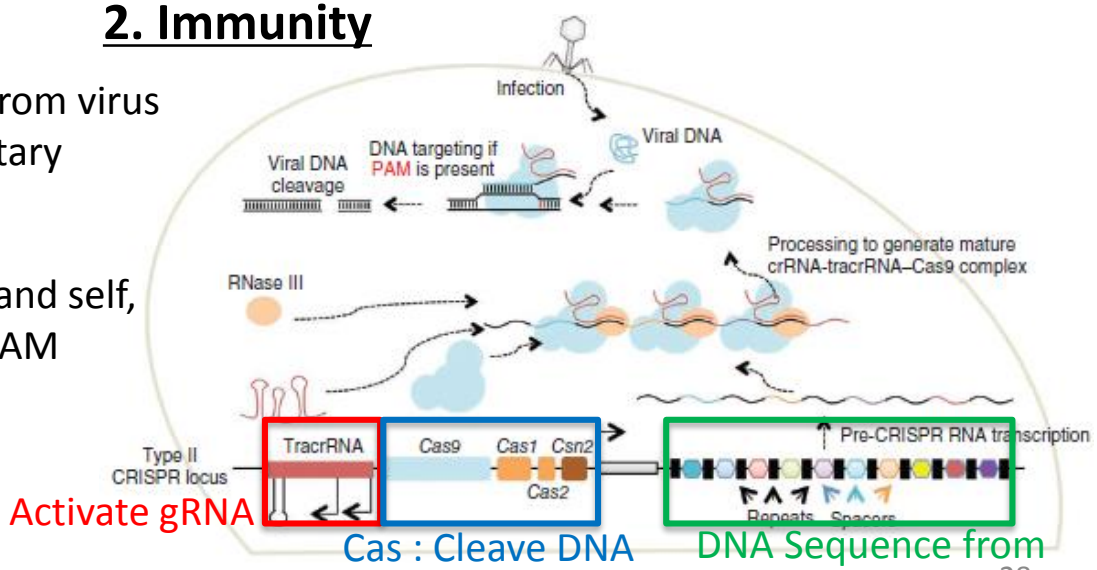
Nucleic acids from viruses are cleaved and inserted into CRISPR locus

DNA Sequence from viruses are stored

2. Immunity

Cas9 cleaves invading DNA from virus recognizing the complementary sequence to the stored one.

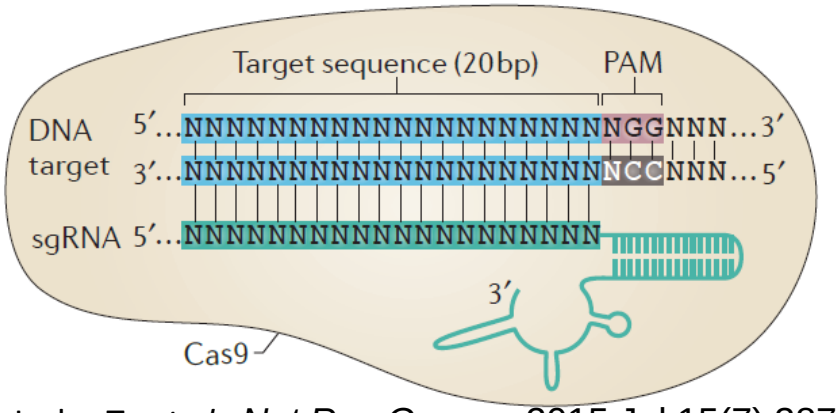
To distinguish unself (virus) and self, Cas9 recognizes preceding PAM sequence.



Mali, P., *et al., Nature Methods*, 2013, 10, 957-62

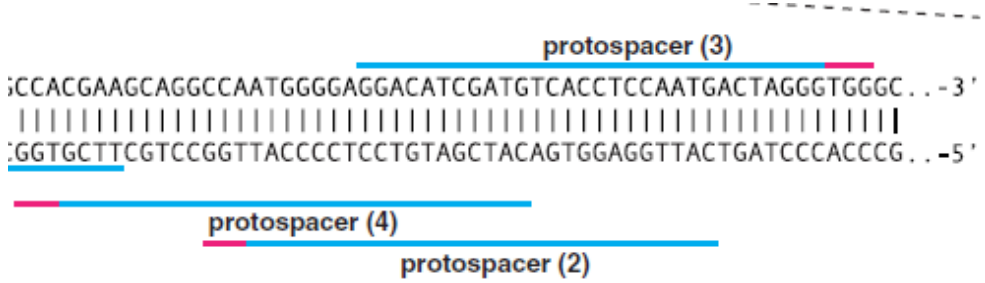
Design and structure of CRISPR/Cas9

Restraint of PAM sequence is not a problem for practical use



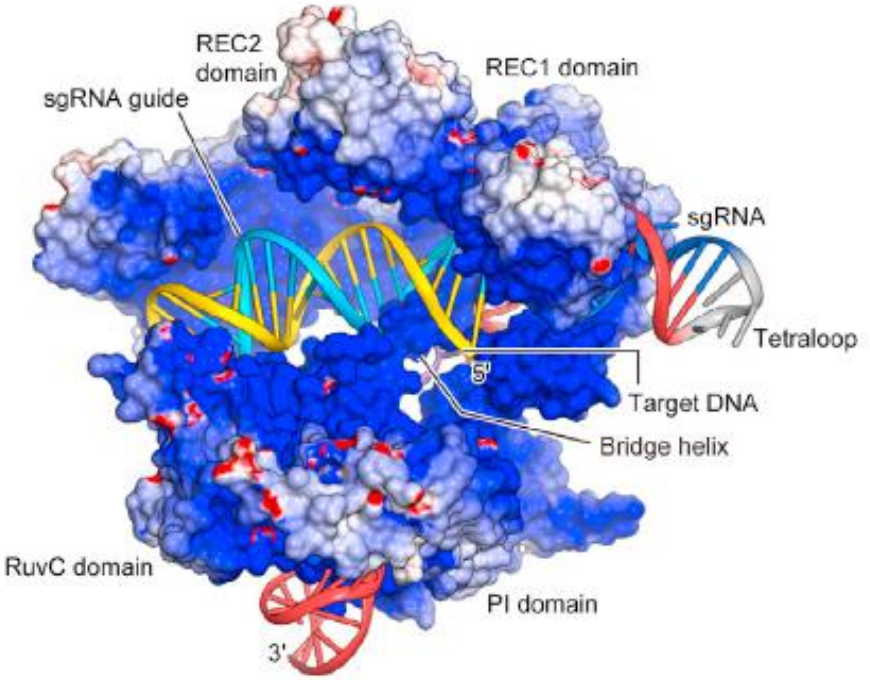
Jacks, T., *et al., Nat Rev Cancer.* 2015 Jul;15(7):387-95.

gRNA must include PAM (NGG) but often several can be designed for 1 gene



Cong, L., *et al.* (2013) *science,* 339, 819-22

sgRNA; DNA; Cas9 ternary complex



Nishimatsu, H., *et al.* (2014) *Cell,* 156, 935-49

positively charged groove of CRISPR/Cas9 interacts with negatively charged sgRNA; DNA complex

See appendix for detailed interactions

Specificity of CRISPR/Cas9

Drawback? of CRISPR/Cas9

TALEN is often better than CRISPR/Cas9 in Specificity

Number of off-target sites of CRISPR/Cas9

sgRNA name	Targeted chromatin state	sgRNA guiding sequence	PAM	No. of binding sites
sgRNA 12	Active promoter	GGGGACGCGCTGGCTTCCCG	GGG	13
sgRNA 11	Active promoter	GGACCGGCTCCCTGGCGGTC	GGG	15
sgRNA 2	DNase1(+)intergenic	GAACACAAAGCATAGACTGC	GGG	16
sgRNA 8	DNase1(-)intergenic	GCCACTTCTAAGCCCTTGAT	GGG	42
sgRNA 3	DNase1(+)intergenic	GGCCCAGACTGAGCACGTGA	TGG	43
sgRNA 1	DNase1(+)intergenic	GGGAAAGACCCAGCATCCGT	GGG	49
sgRNA 5	Active promoter	GCGGTACGCCGCTTCAGTGA	GGG	51
sgRNA 9	DNase1(-)intergenic	GAAACTGGTCCCGTTTACAG	GGG	120
sgRNA 10	DNase1(-)intergenic	GATGAGATAATGATGAGTCA	GGG	201
sgRNA 7	Inactive promoter	GCCTAGGCAGTGGGGGTGCA	GGG	284
sgRNA 4	Active promoter	GGCACTGCGGCTGGAGGTGG	GGG	483
sgRNA 6	Inactive promoter	GGCCCTGCAATGTCAAGGGA	GGG	1,281

(No. of binding sites: 10, 100, 1,000)

Kuscu, C., *et al.* (2014) *nature,* 32, 677-83

Depending on sgRNA, 10-1,000 off-target sites exist.

CRISPR/Cas9 and TALEN are not so different in terms of mismatch tolerance per recognition bases

CRISPR/Cas9 : 1-3 mismatches / 20 bases

TALEN : 1-2 mismatches / 18 repeats

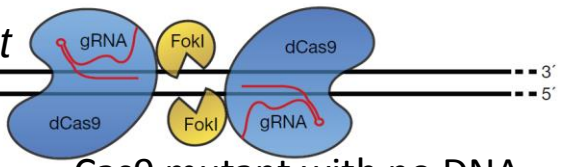
Mali, P., *et al.* (2013) *nature biotechnology,* 31, 833-38

Strategies to improve specificity of CRISPR/Cas9

1. Truncate sgRNA target sequence by 2 or 3 nucleotides (bases which weakly recognize target sequence)
2. two Cas9s for one target gene.
3. Cas9 mutants

2-1. Cas9-Fok1

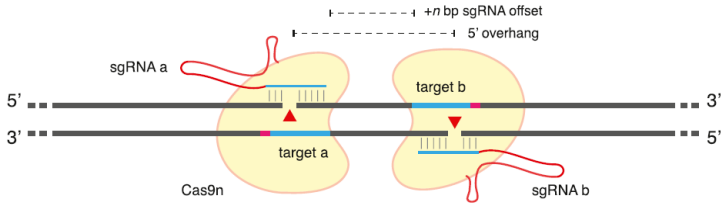
Fok1 : cleave DNA



Joung JK., *et al. Nat Biotechnol.* 2014 Jun;32(6):569-76

Cas9 mutant with no DNA cleavage ability

2-2. Cas9 nickase (cleave only one of the DNA double strand)



Ran, F. A. *et al.*, *Cell* 154, 1380–1389 (2013).

CRISPR/Cas9 mutant with high specificiy

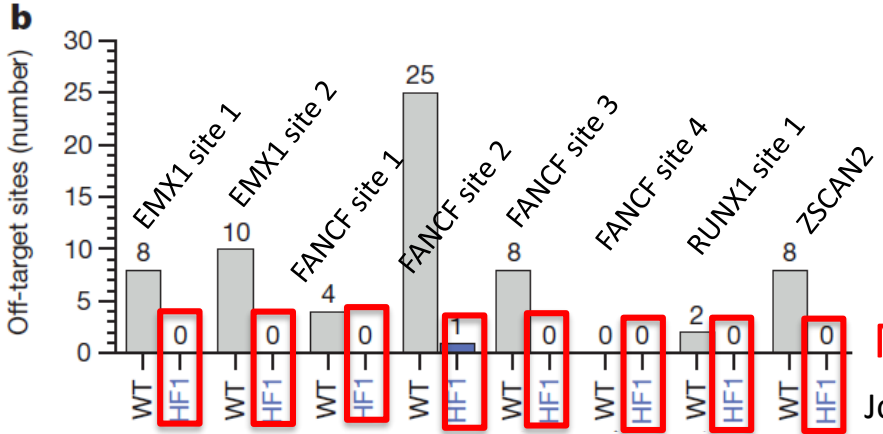
Construction of highly specific Cas9 mutant

Problems about former attempts to increase specificity

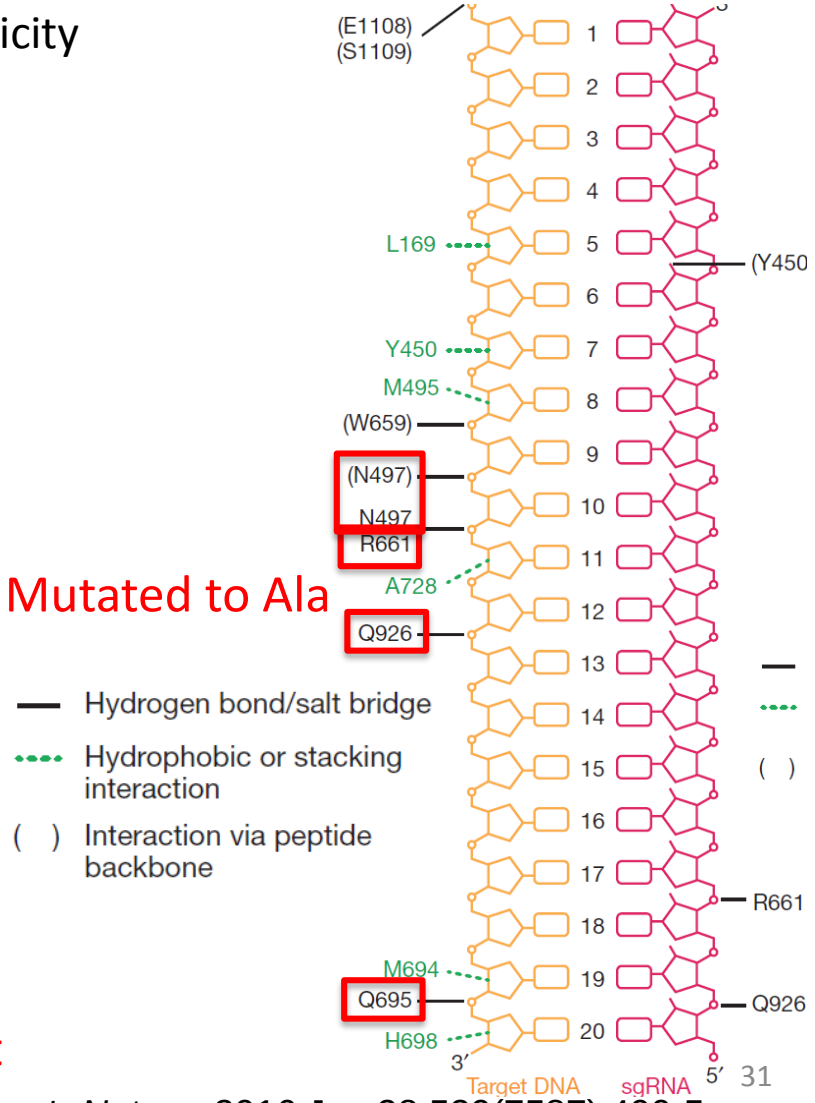
- unproven efficacies on genome-wide scale
- potential to create more new off-target
- need for fusion of functional domains to Cas9
- reduction in targeting range
- more challenges for delivery

Necessity for the development of a robust and easily used strategy

Mutations to reduce Cas9-DNA interactions lead to the increase in specificity



Mutant



Mutated to Ala

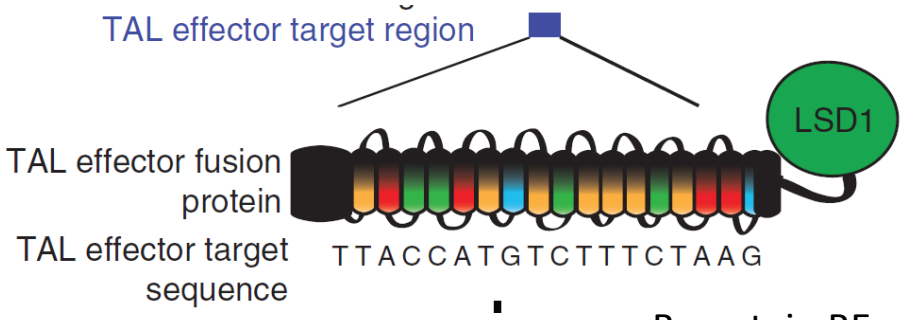
Joung JK., *et al., Nature.* 2016 Jan 28;529(7587):490-5

Contents

Epigenome editing tools

two ways to deliver effectors to specific DNA locus

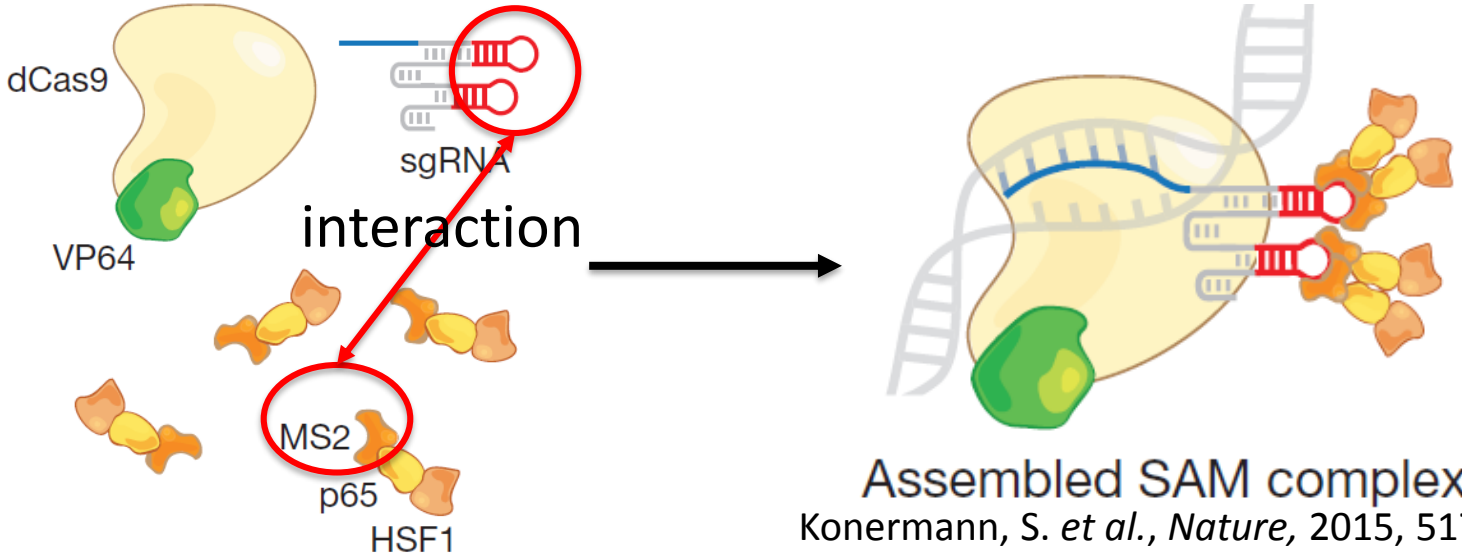
1. Directly fuse effector domains to DNA binding protein



LSD1 : histone demethylase

Bernstein BE., *et al., Nat Biotechnol.*, 2013 Dec;31(12):1133-6.

2. Use of RNA binding protein

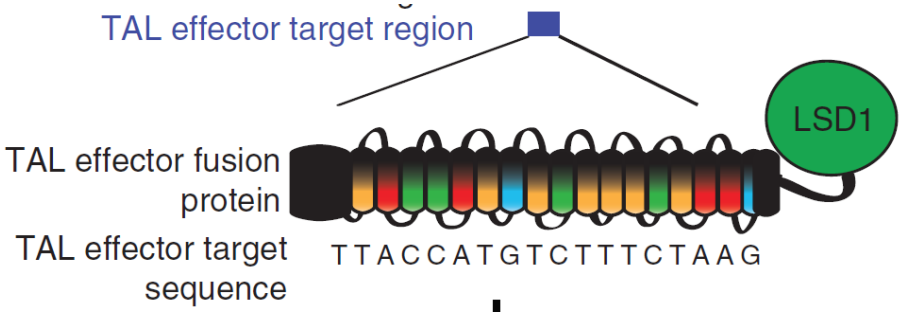


Konermann, S. *et al.*, *Nature,* 2015, 517, 583-8

Use of histone demethylase fused TALE

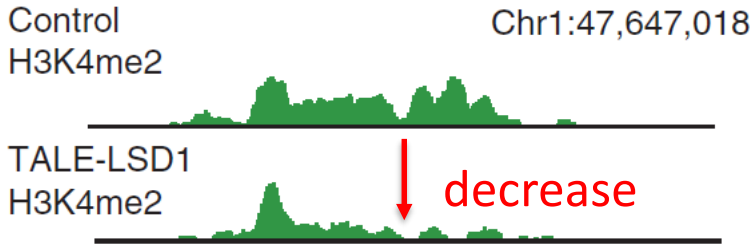
two ways to deliver effectors to specific DNA locus

1. Directly fuse effector domains to DNA binding protein

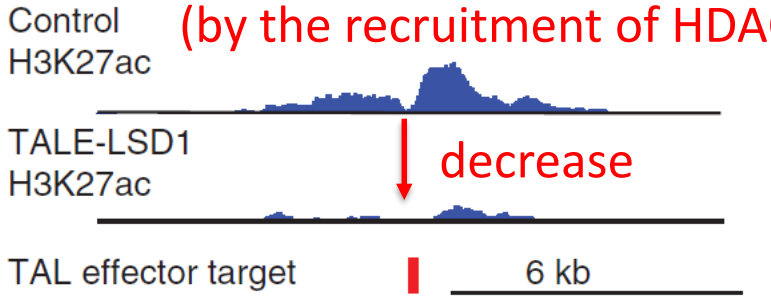


LSD1 : histone demethylase

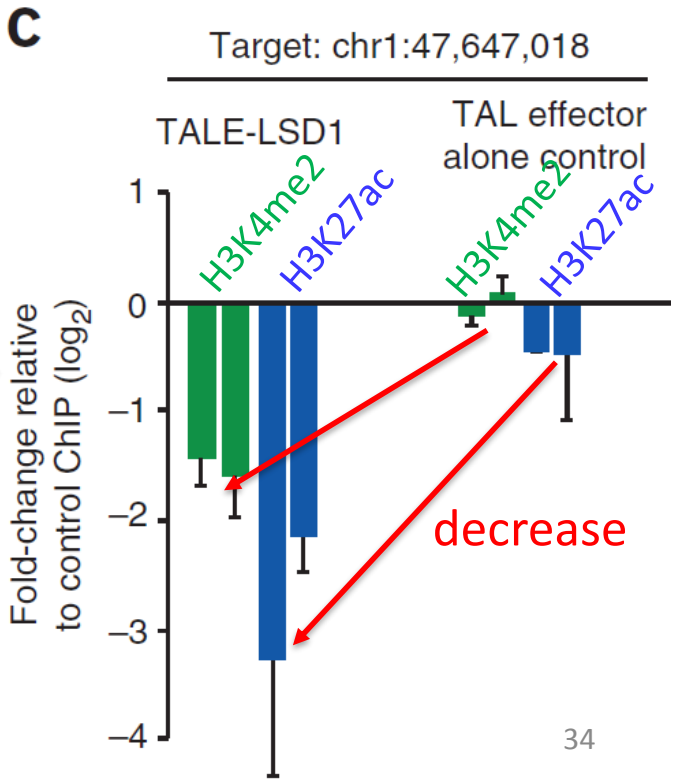
LSD1 catalyzed the removal of H3K4 methylation around targeted region



Also, removal of H3K27ac was observed (by the recruitment of HDAC)



Quantitative data

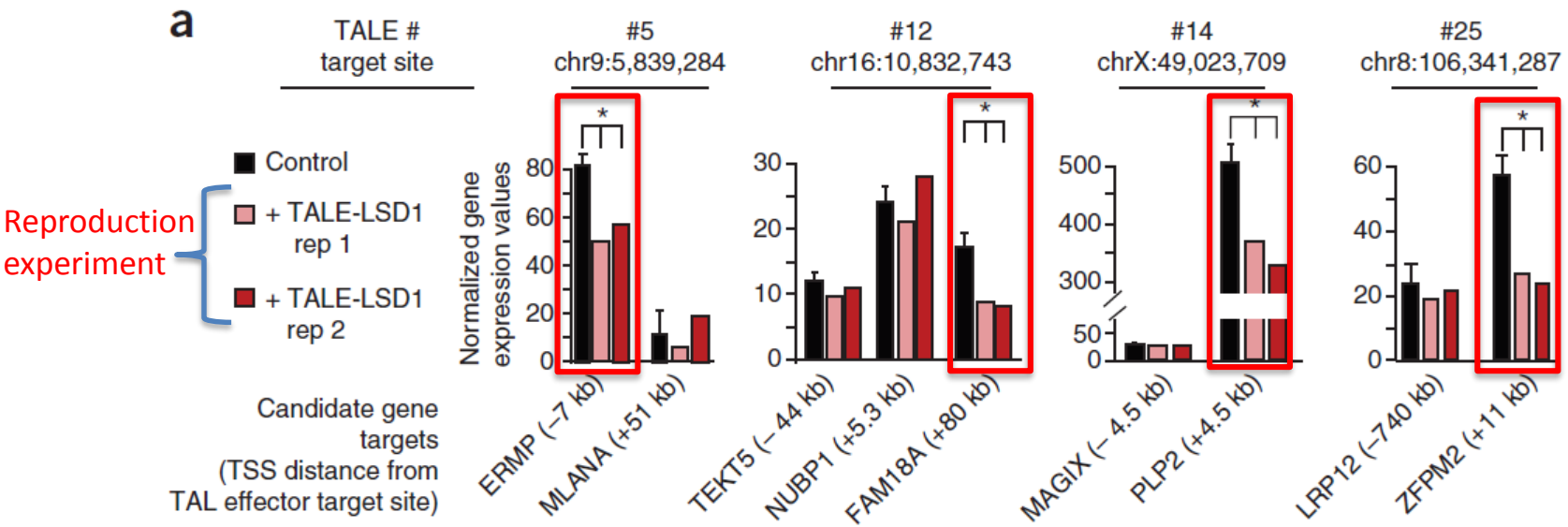


Epigenetics study using TALE-LSD1①

Prior study indicate that sequence elements enriched for H3K27ac and H3K4me2 exhibit enhancer activity

TALEN-LSD1 enables post-translational modifications to be artificially introduced and their effects to be evaluated.

Four of nine enhancers tested, nearby genes were inactivated



Why did the other show little change?

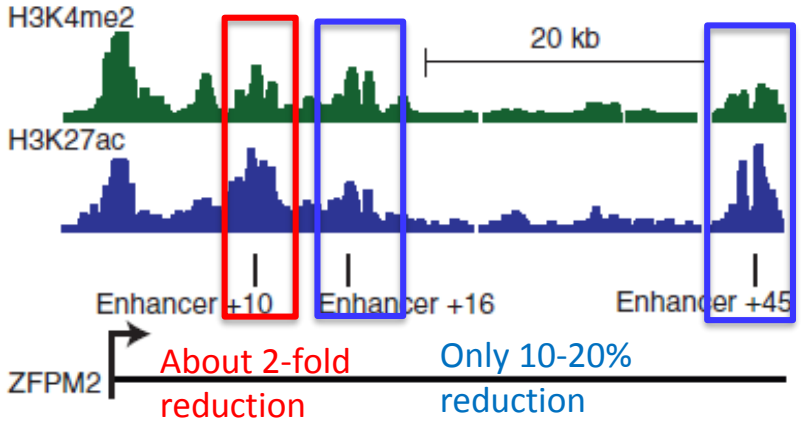
Weak transcriptional effect under detection threshold?

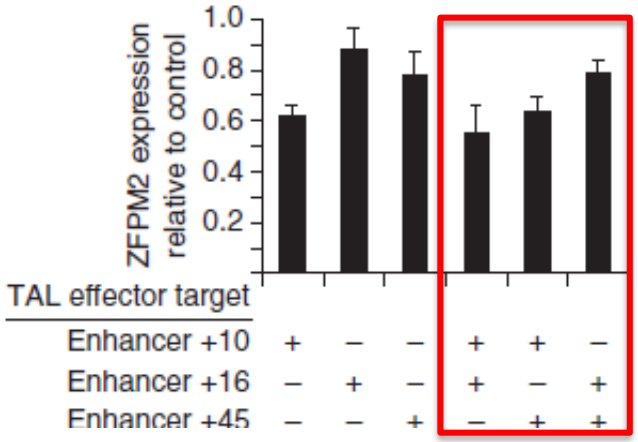
No function as an enhancer in the cell tested?

Epigenetics study using TALE-LSD1②

In the same enhancer, the region where TALEN-LSD1 is targeted may change the degree of effect

Three enhancers for ZFPM2 were tested, yielding different inactivation of gene expression





Effect of targeting paired enhancers is less than the sum of their individual effects

↓

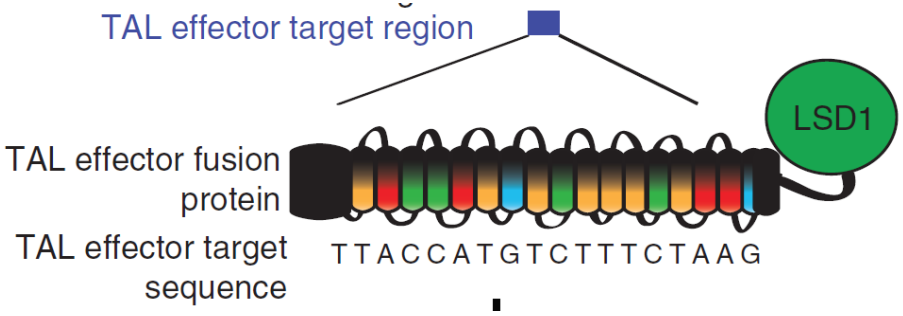
Redundant effect of the multiple enhancers

TALE-LSD1 fusions are valid tools to study complex regulatory interactions among multiple enhancers and genes in a locus

Specific gene activation tool using CRISPR/Cas9

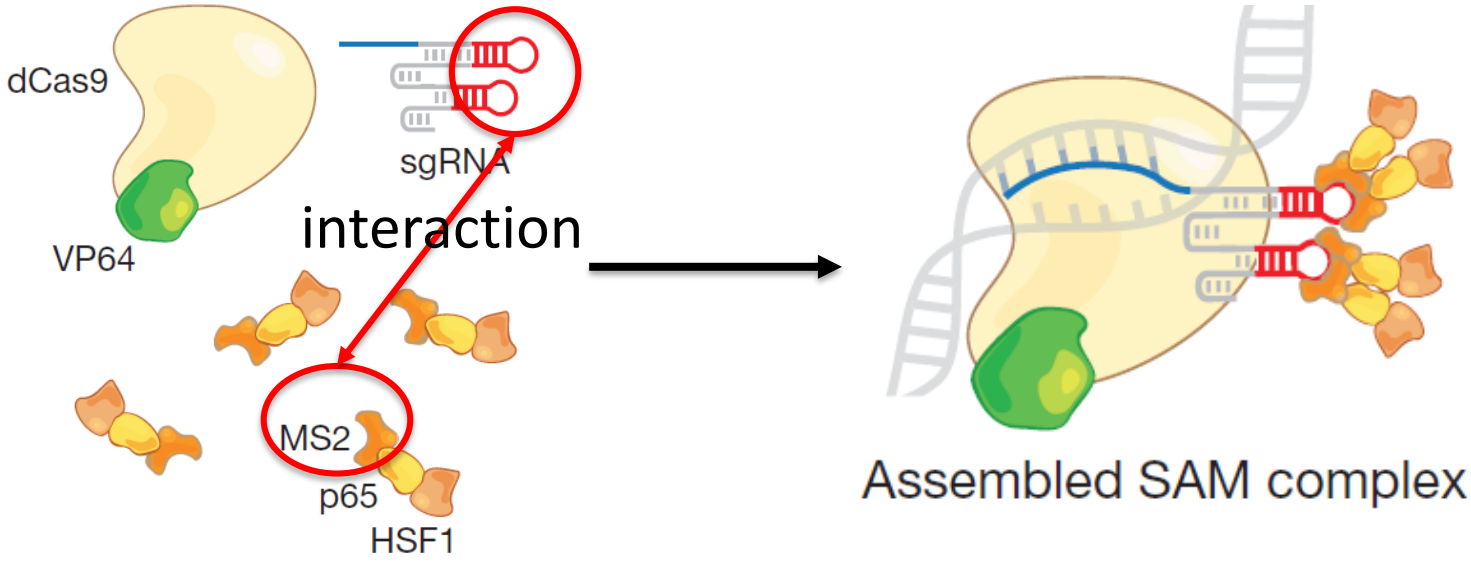
two ways to deliver effectors to specific DNA locus

1. Directly fuse effector domains to DNA binding protein



LSD1 : histone demethylase

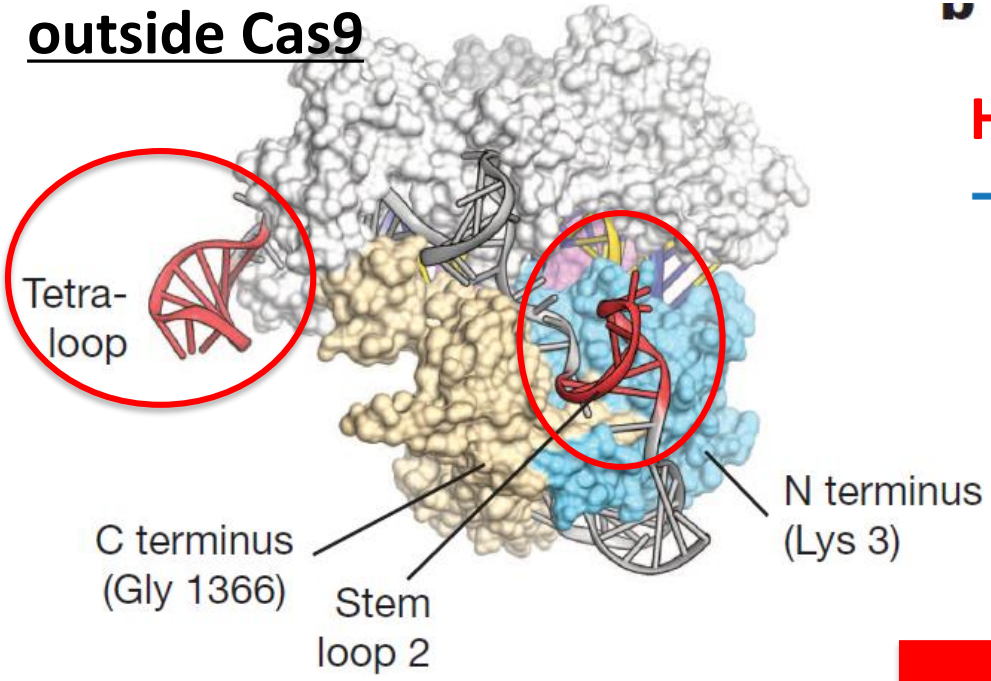
2. Use of RNA binding protein



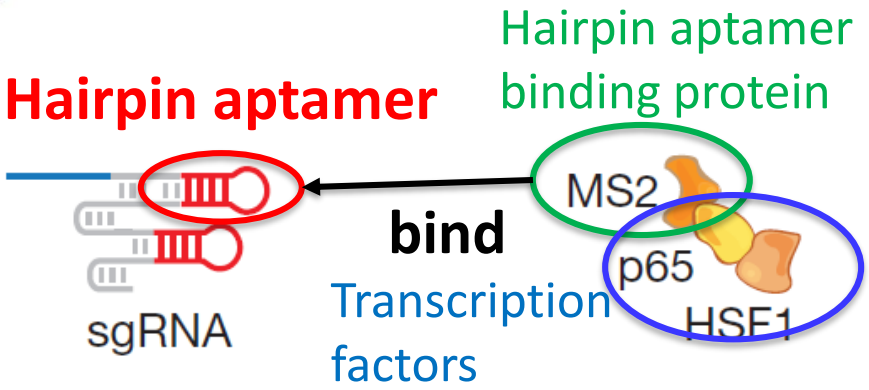
Structure based design of SAM system

Structure based design for construction of RNA aptamer

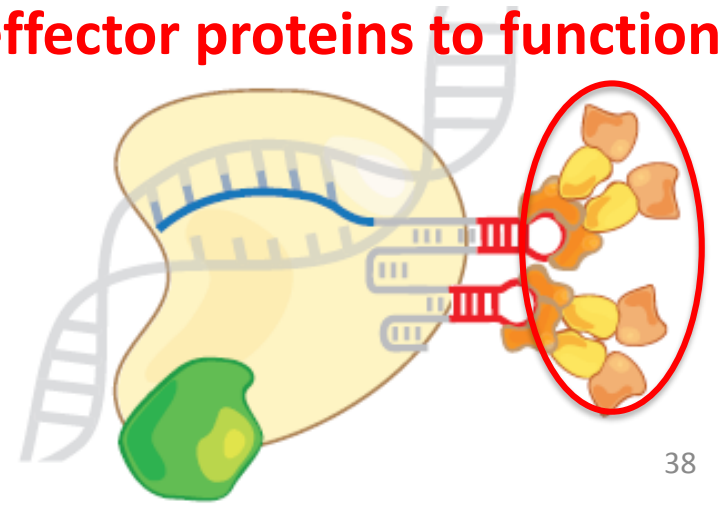
Tetraloop and stem loop2 is exposed outside Cas9



→appropriate for scaffold for effector proteins



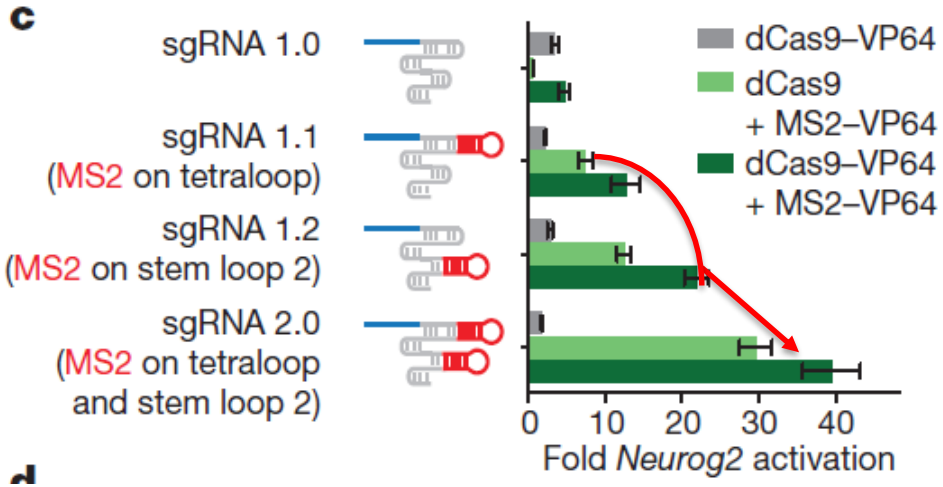
Good environment for effector proteins to function



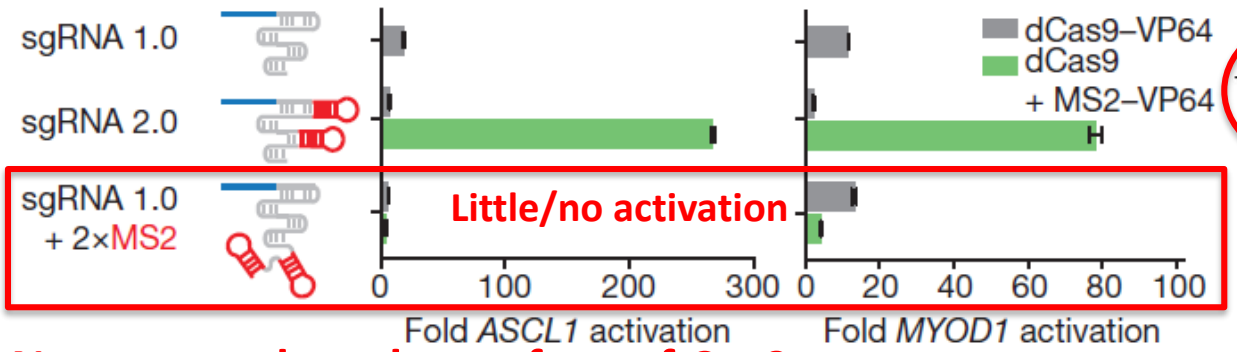
Activation of specific gene achieved with SAM

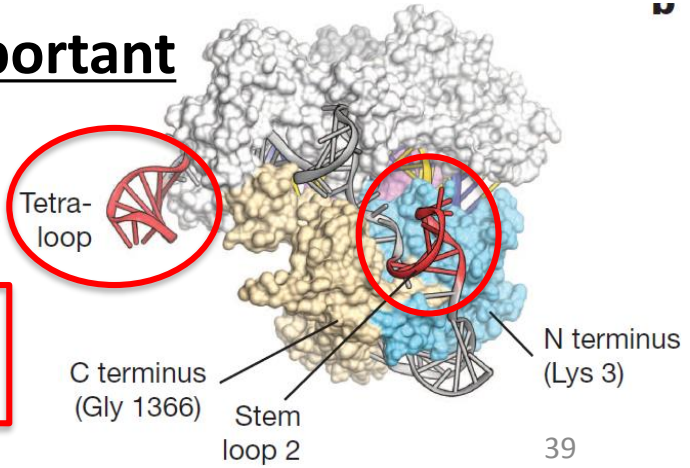
Confirmation of the validity of the design

Efficient gene activation was achieved dependent on the number of hairpin aptamers



Structural position of hairpin aptamers is important





Not exposed on the surface of Cas9

Contents

Summary

- genome editing tools facilitated the study of gene functions

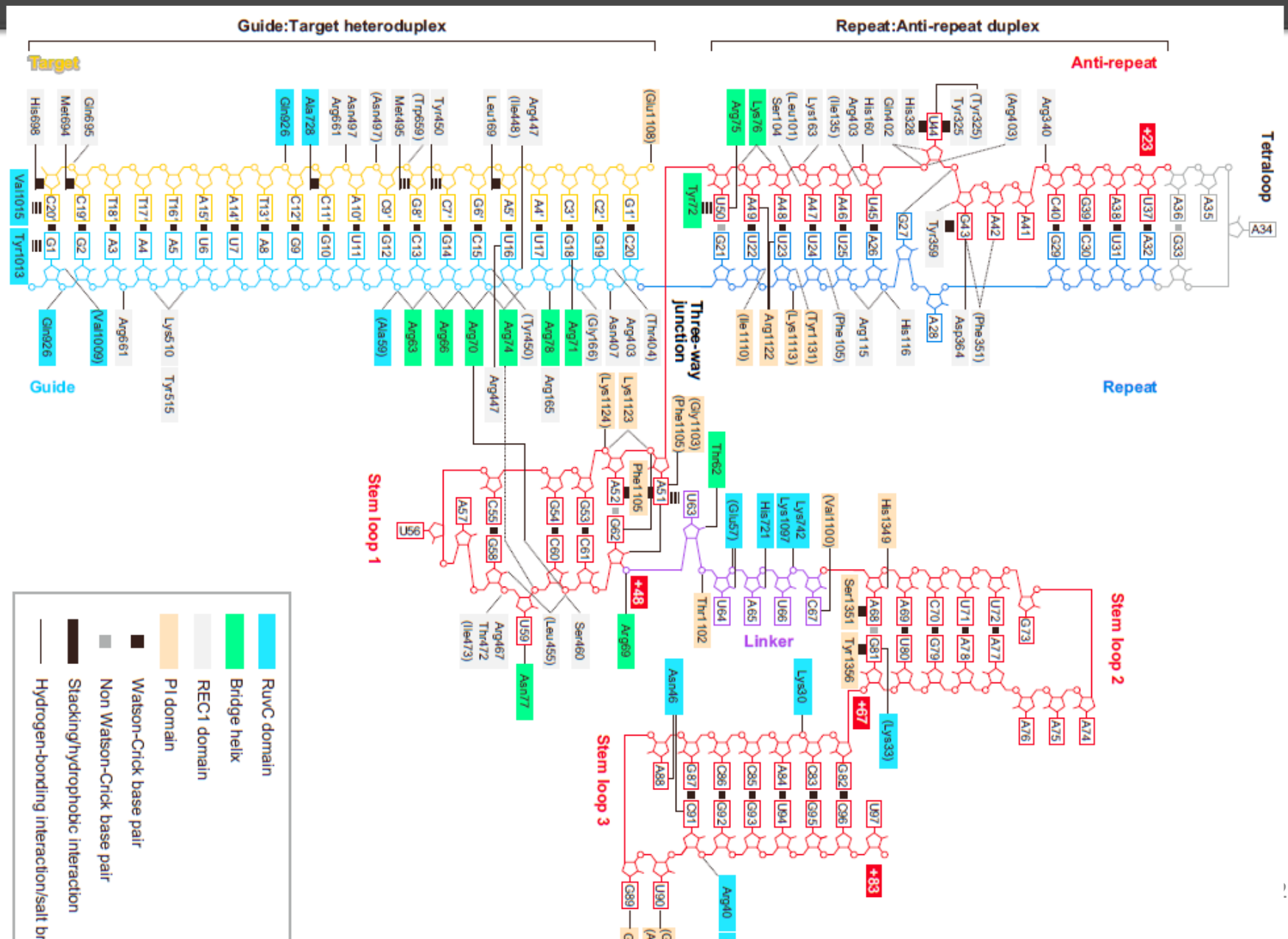
- ZFN was the first to be invented but has not been widely used due to its difficulty in designing

- TALEN overcame the drawback of ZFN due to its simplicity in recognizing DNA sequence.

- CRISPR/Cas9, using gRNA for recognition of DNA sequence, rendered genome editing much easier and is now intensively studied

- genome editing tools are applicable to edit epigenome, enabling more direct study of epigenetics.

Appendix; detailed interactions of Cas9-gRNA-DNA

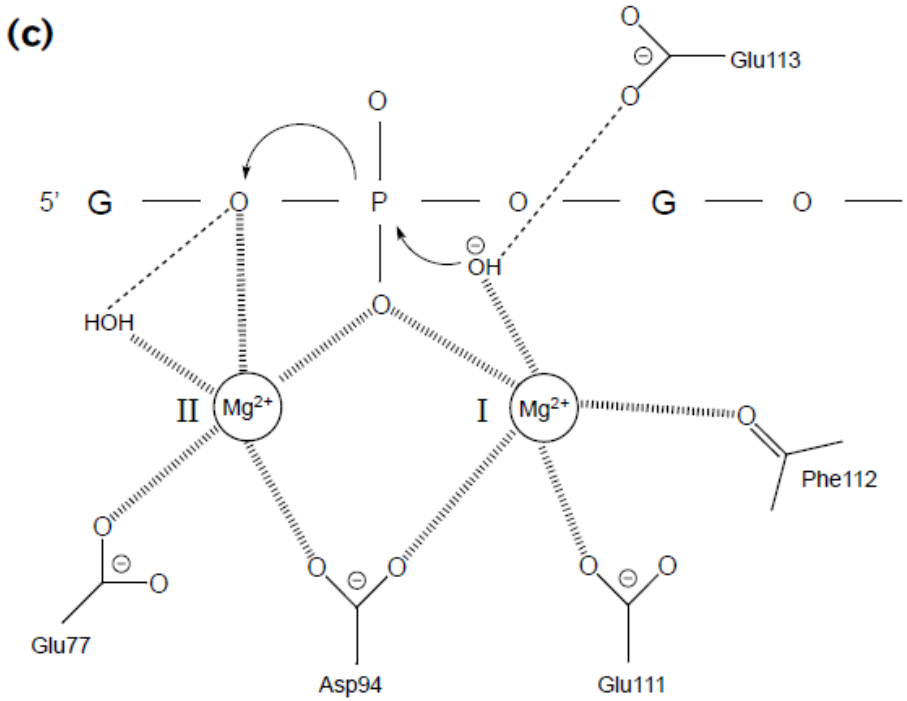


ZFN (Zinc-finger nuclease)

ZFN consists of Fok1 endonuclease cleavage domain and ZF protein.

Fok1 endonuclease consists of Recognition domain and Cleavage domain

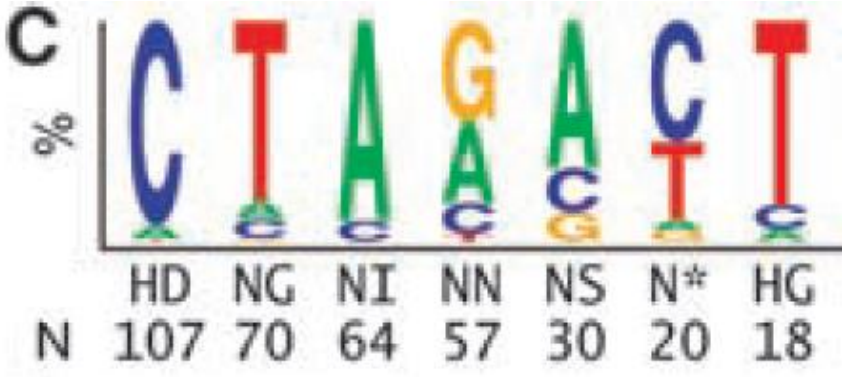
Recognition domain binds 5’-GGATG-3’



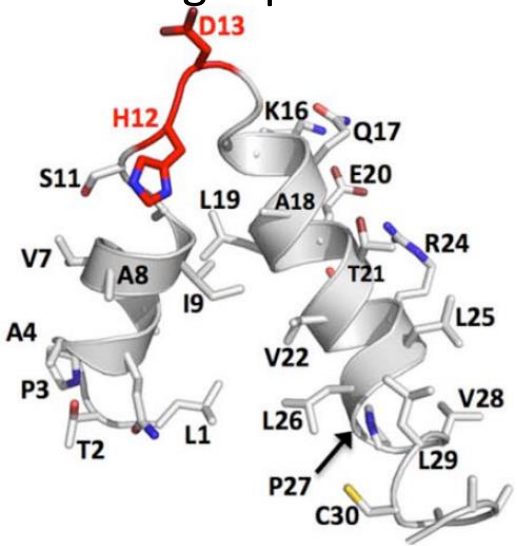
Cole P A, *et al. Chem. Rev.*, **2015**, *115* (6), pp 2419–2452

TALEN (transcription-activator-like effector nuclease)

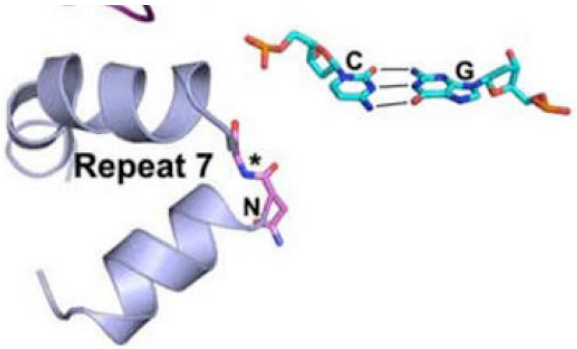
RVDs and their specificity



Structure of the DNA binding repeat



'N*' extends less deeply into the major groove



CRISPR/Cas9

Application to genome-scale gene activation screen