Structural Insights into the Specific Recognition of 5-methylcytosine and 5-hydroxymethylcytosine by TAL Effectors

Lulu Liu [1,3], Yuan Zhang [2], Menghao Liu [1,3], Wensheng Wei [2,3], Chengqi Yi [2,3,4] and Jinying Peng [2]

*1 - **Academy for Advanced Interdisciplinary Studies,** Peking University, Beijing 100871, China*
*2 - **State Key Laboratory of Protein and Plant Gene Research,** School of Life Sciences, Peking University, Beijing 100871, China*
*3 - **Peking-Tsinghua Center for Life Sciences,** Peking University, Beijing 100871, China*
*4 - **Department of Chemical Biology and Synthetic and Functional Biomolecules Center,** College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China*

***Correspondence to Jinying Peng:** Fax: +8610 62752895. jypengpku@pku.edu.cn.*

https://doi.org/10.1016/j.jmb.2019.11.023
Edited by James Berger

Abstract

Transcription activator-like effectors (TALEs) recognize DNA through repeat-variable diresidues (RVDs), and TALE-DNA interactions are sensitive to DNA modifications. Our previous study deciphered the recognition of 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) by TALEs. Here, we report seven crystal structures of TALE-DNA complexes. The 5mC-specific RVD HA recognizes 5mC through van der Waals interactions and exhibits highly similar loop conformation to natural RVDs. The degenerate RVD RG contacts 5mC and 5hmC via van der Waals interactions as well; however, its loop conformation differs significantly. The loop conformations of universal RVD R* and 5hmC-specific RVD Q* are similar to that of RG, while the interactions of R* with C/5mC/5hmC and Q* with 5hmC are mediated by waters. Together, our findings illustrate the molecular basis for the specific recognition of 5mC and 5hmC by multiple noncanonical TALEs and provide insights into the plasticity of the TALE RVD loops.

Introduction

Transcription activator-like effectors (TALEs) are virulence factors secreted by pathogenic bacteria *Xanthomonas*, which act within various plant species by binding to promoter sequences and activating the expression of individual plant genes to support bacterial infection [1–3]. The specificity of TAL effectors is determined by a modular DNA-binding domain composed of a variable number of tandem repeats, with each repeat recognizing one specific DNA base pair. Remarkably, each repeat is comprised of 33–35 (typically 34) highly conserved amino acids, with the exception of two hypervariable residues called repeat-variable diresidues (RVDs) at positions 12 and 13 [4,5]. Experimental and computational approaches have partially deciphered the code of DNA recognition by RVDs. The four most frequently used RVDs, NI, NG, HD, and NN, were found to preferentially bind to A, T, C, and G/A, respectively [1,4]. The complete RVD-DNA recognition code has also been deciphered via screening of all possible RVD combinations [6,7]. The modular architecture of the TALE repeats provides multiple programmable tools for genome manipulation by TALEs fused with functional domains, such as transcription activators, repressors, or nucleotide endonucleases, to create transcriptional modulators and gene editing tools (TALEN) [8–11]. Although CRISPR-Cas9 is currently the most widely used genetic manipulation tool, TALE-based technologies have their unique applications. For instance, CRISPR-Cas9 poses a challenge for mitochondrial DNA (mtDNA) use, as it is difficult to import the guide RNA component into mitochondria [12]. However, mitochondria-targeted TALEN (mito-TALENs) are successfully used to selectively eliminate mitochondrial pathogenic mutations, making it an effective therapy for human mitochondrial diseases caused by mutations in mtDNA, such as Leber's hereditary optic neuropathy, ataxia, neurogenic muscle weakness, and retinitis pigmentosa [13–15].

Crystal structures of TALE-DNA complexes demonstrate that TALEs form a right-handed superhelical assembly and wrap around the DNA major groove [16,17]. All of the repeats in the TALE-DNA complexes exhibit nearly identical conformations. The first RVD residue (His or Asn) does not directly interact with the nucleobase, but its side chain engages in a hydrogen bond (H bond) with the carbonyl oxygen of the conserved Ala8 to stabilize the proper loop conformation [16,17]. The second residue makes a direct base-specific contact with the DNA sense strand, suggesting that the TALE-DNA interaction is sensitive to DNA chemical modifications.

Besides the four canonical nucleotides A, T, G and C, methylated cytosine represents the “fifth base” of mammalian genomes, which constitutes ~1% of all DNA bases and primarily occurs symmetrically in the context of CpG dinucleotides [18,19]. In mammalian genomes, approximately 70–80% of CpGs are methylated [20]. As an important epigenetic marker, 5mC regulates diverse biological processes, including X chromosome inactivation, gene expression and silencing, maintenance of genome stability, and genomic imprinting [21,22]. 5mC can be sequentially oxidized to 5-hydroxymethylcytosine (5hmC) by ten-eleven translocation (TET) family proteins [23,24]. The levels of 5hmC are highly variable in different tissues (~1%–10% of 5mC) [25,26]; neuronal tissues contain the highest levels of 5hmC, some somatic tissues such as the kidney and heart exhibit moderate levels of 5hmC, and DNA from the spleen and endocrine glands possesses the lowest amounts of 5hmC [25]. In mammalian genomes, 5hmC is particularly enriched in promoters and gene bodies of actively transcribed genes [27,28]. Given its tissue-specific and genome-wide distribution, several studies have found that 5hmC is a stable epigenetic modification, and dysregulation of 5hmC is frequently observed in cancers [29–32].

Unlike CRISPR-cas9 that is base-pairing dependent and hence modification insensitive, TALE-DNA interactions are sensitive to DNA modifications and provide unique opportunity for novel applications. For instance, this sensitivity can be used to detect modified bases. The RVDs NG and N* (the asterisk represents the deletion of the second amino acid of the RVD), which can tolerate 5mC, are used to overcome the genome manipulation sensitivity in vivo [33–36] and to detect 5mC in synthesized oligonucleotide sequences with a high resolution in vitro [37,38]. Compared to the widely used 5mC–antibody-based methylated DNA immunoprecipitation (MeDIP), TALE-based analysis of 5mC exhibits higher resolution and sensitivity and strand-specificity [39]. Engineered TALEs that combine NG, N*, and HD are used as DNA binding receptors to directly distinguish C, 5mC, and 5hmC in defined DNA sequences [40]. Furthermore, studies revealed that some size-reduced RVD loops (G*, S*, and T*) bind to C, 5mC, and 5hmC with similar affinities, indicating further applications of TALE-based tools [39,41]. In a previous study, we deciphered the recognition of 5mC and 5hmC by TALEs [42], identified the novel 5mC-specific RVD HA (the binding affinity to 5mC is about twofold of that to C and 5hmC) and 5hmC-specific RVDs Q* (the binding affinity to 5hmC is about twofold of that to C and 5mC) (Fig. S2a), the degenerate RVD RG that recognizes both 5mC and 5hmC, and the universal RVD R* that recognizes unmodified C, 5mC, and 5hmC. Utilizing these novel RVDs, we performed methylation-dependent gene activation, genome editing, and locus-specific 5hmC detection. Here, we report a total of seven structures of TALE-DNA complexes and elucidate the molecular basis of recognition of 5mC and 5hmC by noncanonical TALE RVDs (HA, RG, R*, and Q*). Our study also gives insights into the plasticity of the TALE RVD loops.

Results

The overall structures of the TALE-DNA complexes

We solved the crystal structures of noncanonical TALEs in complex with modified DNA. TALE proteins containing residues corresponding to positions 231–720 of the 11.5 repeats TAL effector dHax3 [43] were crystallized in complex with a 17-base pair (bp), chemically synthesized target DNA oligonucleotide, which was modified from dHax3-mCG, as reported previously [35]. A -CGCG- sequence was included, and the first C opposite the RVD of repeat 6 was also synthesized as 5mC and 5hmC (Fig. 1a and b, S1a and b).

We obtained the structures of the TALE-DNA complex, including specific RVDs HA opposite 5mC and Q* opposite 5hmC (designated HA-5mC and Q*-5hmC), degenerate RVD RG opposite 5mC and 5hmC (designated RG-5mC and RG-5hmC), and universal RVD R* opposite C, 5mC, and 5hmC (designated R*-C, R*-5mC, and R*-5hmC). Of these structures, there are two complexes in each asymmetric unit of HA-5mC, R*-C, and R*-5mC, and four complexes in each asymmetric unit of RG-5mC, RG-5hmC, R*-5hmC, and Q*-5hmC, and all the individual TALE-DNA complexes are nearly identical (Figs. S2b–d). As previously reported, the overall structures of all monomer complexes in these crystals described earlier are nearly identical [16] and are arranged in a consecutive, right-handed, superhelical assembly. The superhelical TALE structures wrap around the major groove of the DNA double helix, which is in a relatively unperturbed B-form (Fig. 1c). Each 34–amino acid repeat comprises two helices connected by an RVD loop, which extends into the DNA major groove and forms direct contact with the corresponding DNA base in the sense strand. All repeats exhibit an almost identical conformation, with root mean square deviations (RMSDs) of 0.19–0.29 Å over all 34 C_α atoms, except for those containing RVD RG, R*, and Q*, with an RMSD of approximately 1.15 Å among HA and RG, R*, or Q* over the three C_α atoms of residues 12–14. The loop conformations of RG, R*, and Q* differ significantly from those of HA and other previously reported canonical RVDs whose first residue is either His or Asn. Their specific distinctions and the resulting effects on specific recognition by TALEs are discussed further.

RVD HA specifically contacts 5mC through van der Waals interactions

The TALE protein with RVD HA in repeat 6 was crystallized with DNA duplex containing 5mC

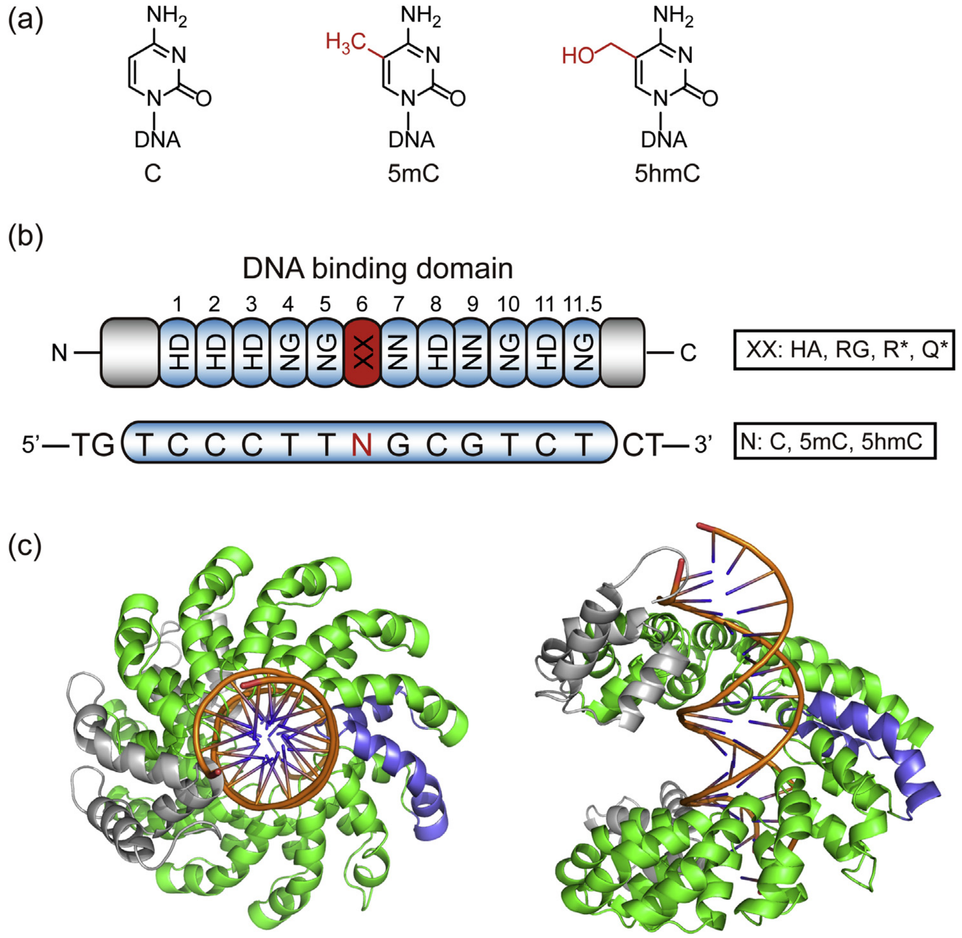

Fig. 1. The overall structures of the TALE-DNA complexes. (a) Chemical structures of C, 5mC, and 5hmC. (b) The sense strand of the DNA oligonucleotides and the corresponding TAL effectors (TALEs) used for crystallization. There are 11.5 repeats included in the TALEs, and the RVD of repeat 6 was mutated to HA, RG, R*, or Q*. TALEs were crystallized in complex with a 17-base pair, chemically synthesized target DNA oligonucleotides. The target DNA contained a -CGCG- sequence in the sense strand, and the first C opposite repeat 6 was synthesized as C, 5mC, or 5hmC. The structures are named by the RVD in repeat 6 of the TALE and the corresponding modified or unmodified C in the DNA sense strand. For example, the crystal structure of RVD HA in complex with 5mC is designated HA-5mC. (c) Overall structure of TALEs bound to DNA. The 11.5 repeats form a right-handed, superhelical structure that wraps around the DNA major groove. The 11.5 repeats are shown in green except that the repeat 6 is shown in slate, the flanking N- and C-terminals are shown in gray.

opposite HA (designated as HA-5mC). The final crystal structure was refined to a resolution of 2.48 Å, with two uniform complexes included (Fig. S2b, Table S1). The electron density is well defined from repeat 1 through repeat 11 (Fig. 2a).

All of the repeats in the HA-5mC structure form highly similar two-helix bundles, with RMSDs of 0.19–0.29 Å overall 34 C_α atoms (Fig. 2b). The structures demonstrate that the first residue of natural RVDs, either His or Asn, does not directly interact with DNA, while its side chain conformation is invariant and makes a direct hydrogen bond to the carbonyl oxygen atom of the conserved Ala8, thereby constraining the RVD loop. However, for HA opposite 5mC, while the HA loop backbone conformation remains unchanged compared with canonical RVDs, the side chain of His12 extends slightly deeper into DNA major groove, forming an alternative H bond between His12 and Ser11 of the next repeat (Fig. 2c). Previously reported NG-5mC also exhibits H-bond interactions with Ser11 [35,41]. These TALE-DNA complexes demonstrate that Ser11 is an excellent alternative to Ala8, as both interact with His12 or Asn12 to maintain a normal RVD loop conformation.

The recognition of 5mC by HA is similar to that of NG, as previously reported [35]. In HA-5mC, His12 stabilizes the RVD loop, and the short side chain of Ala13 not only provides sufficient space to accommodate the 5-methyl group of 5mC but also allows for optimal nonpolar van der Waals interactions between the side chain methyl group of Ala13 and the 5-methyl group of 5mC. The distance between these two groups is 3.35 Å (Fig. 2c and d). However, the distance may be too large for an unmodified cytosine (C) to make van der Waals interaction with the Ala13 side chain due to the lack of 5-methyl group, while the 5-hydroxymethyl group of 5hmC is relatively bulky and likely to introduce steric clash with the side chain of Ala13. Thus, our result can explain the specificity of HA for 5mC rather than C or 5hmC.

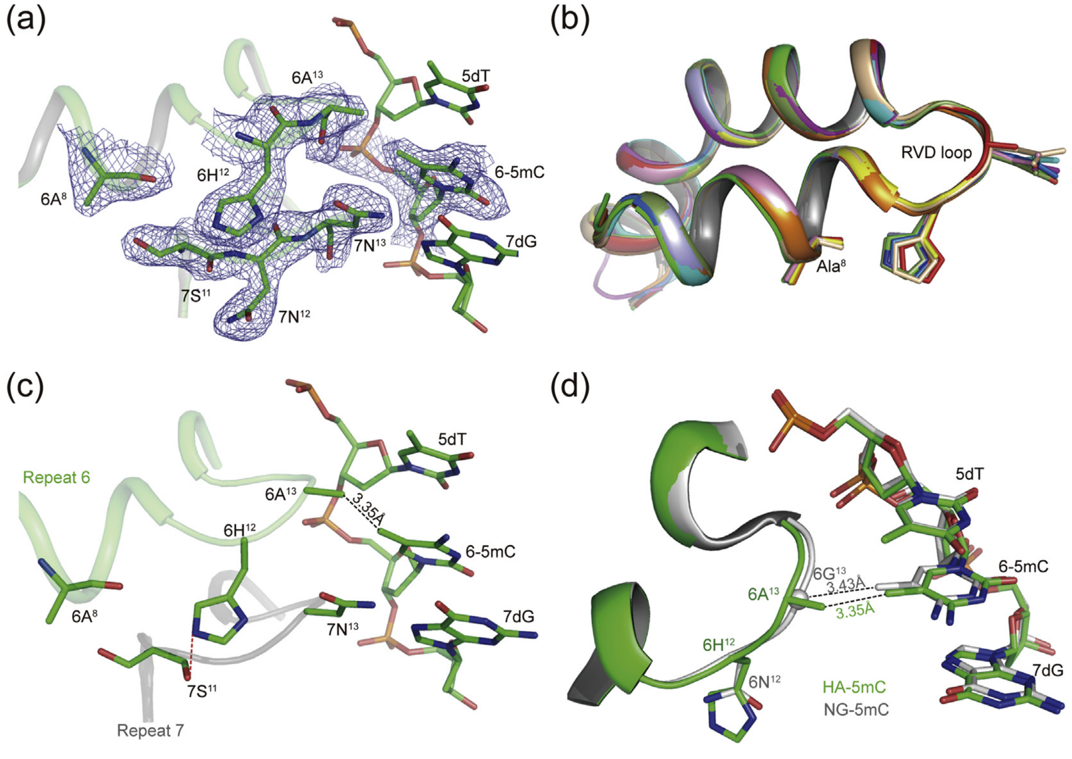

Fig. 2. RVD HA specifically contacts 5mC through a van der Waals interaction. (a) 2Fo-Fc electron density map (slate blue mesh), contoured at 1.0σ, of the key site of the HA-5mC complex, where RVD HA and 5mC reside. (b) Structural superimposition of all 11.5 repeats in the HA-5mC complex structure. All of the repeats can be superimposed with RMSDs of 0.19–0.29 Å overall 34 C_α atoms. Repeat 6, which contains RVD HA, is colored in red. 6His12 ($6H^{12}$, His12 of repeat 6) is slightly farther from the conserved 6Ala8 when compared with His12 or Asn12 in other repeats. (c) Structural basis for the recognition of 5mC by RVD HA. The side chain of His12 forms an H bond (red dashed line) with the conserved Ser11. The side chain methyl group of Ala13 directly contacts the 5-methyl group of 5mC through a van der Waals interaction (black dashed line). The distance between them is 3.35 Å. (d) A comparison of 5mC recognized by HA (green stick model, green cartoon) with previously reported NG-5mC (gray stick model, gray cartoon). The methyl groups of Ala13 or the C_α atoms of Gly13 contact the 5-methyl group of 5mC through a van der Waals interaction.

In our previous study, we found that the affinity of HA for 5mC is stronger than NG [42]. Thus, we speculate that the van der Waals interaction between the two methyl groups in HA-5mC is stronger than that between C_α and the methyl group in NG-5mC, probably because of the shorter distance between the two methyl groups in HA-5mC. In addition to NG, N* was also used to detect 5mC, and its affinity for 5mC is also weaker than HA [37,38]. Based on the crystal structure, we speculate that the RVD loop of N* is shorter due to the lack of an amino acid, and thus the distance to the methyl group of 5mC is too large to form an optimal van der Waals interaction. This is consistent with observations from cellular assays that N* can also tolerate 5hmC [40]. Our previous study also revealed that the affinity and specificity of NA for 5mC are similar to HA, and it is reasonable considering that both Asn12 and His12 are involved in stabilizing the invariant conformation of the RVD loop and the same Ala13 interacting with 5mC directly through van der Waals interactions [42].

Degenerate RVD RG forms a distinct loop conformation

The first residue of canonical RVDs, either His or Asn, plays a highly conserved role in maintaining invariant loop conformations. Here, we cocrystallized TALEs containing the noncanonical degenerate RVD RG, in which the first residue was mutated to Arg, with DNA duplex including 5mC or 5hmC opposite RG (designated as RG-5mC and RG-5hmC, respectively) for the first time. The structures were refined to resolutions of 3.10 Å and 3.09 Å, respectively (Table S1). Unlike HA-5mC, their unit cells comprise four complexes (designated A, B, C, and D), and the qualities of the electron density for complexes A and B are significantly greater than those of C and D (Figs. S2c, S3 and S4). For complexes A and B, the electron density from repeat 1 through 11, particularly the key sites where RG and the modified C reside, is clearly observed (Fig. 3a and b).

The RG-containing repeats (repeat 6) and RVD loop of RG-5mC and RG-5hmC exhibit almost identical features and can be superimposed with an RMSD of 0.153 Å over all of the 34 C_α atoms. However, the loop conformation of RG is significantly different from that of the HA in HA-5mC, with an RMSD of approximately 1.15 Å between HA-5mC and RG-5mC or RG-5hmC over the three C_α atoms (residues 12–14) that constitute the RVD loop (Fig. 3c). The RG loop deviates from the DNA major groove and shifts toward the 3' end of the DNA, and the corresponding DNA base pair shifts toward the RVD (Fig. 3d and e).

Previous studies illustrated that the first amino acid of RVD, either His12 or Asn12, does not contact with DNA directly, but the side chain of His12 or Asn12 forms a direct H bond with the carbonyl oxygen atom of the conserved Ala8 to stabilize the conformation of the RVD loop [16,17]. However, in the case of RG, the side chain amino group of Arg12 donates an H bond to the phosphate group of the DNA duplex. Furthermore, the main chain amino group of Arg12, rather than its side chain, forms an H bond with carbonyl oxygen atom of Ala8. Therefore, through these two H-bond interactions, Arg12 supports the formation of a more stable loop conformation, which is distinct from those formed by His12 or Asn12. In addition, the main chain carbonyl oxygen of the RVD second residue Gly13 flips to the DNA major groove. Meanwhile, the corresponding 5mC or 5hmC horizontally shifts toward the DNA major groove (Fig. 3d–g). As a result, the main chain carbonyl oxygen of Gly13 forms van der Waals interactions with the 5-methyl group of 5mC and the 5-hydroxyl group of 5hmC, with distances of 4.10 Å and 3.65 Å, respectively (Fig. 3f and g). Thereby, we speculate that unmodified C cannot form van der Waals interaction with the carbonyl oxygen of Gly13 due to the lack of 5-methyl or 5-hydroxymethyl group, and thus RG is able to discriminate 5mC and 5hmC from C.

Universal RVD R* recognizes C, 5mC, and 5hmC via water-mediated interactions

We further obtained structures of the TALE-DNA complex, with the noncanonical universal RVD R* (the asterisk represents the deletion of the second residue of the RVD) opposite unmodified C, 5mC, and 5hmC (designated as R*-C, R*-5mC, and R*-5hmC, respectively). The structures were refined to 2.20 Å, 2.49 Å, and 3.03 Å resolutions, respectively (Table S2). For R*-C and R*-5mC, we observed two complexes (designated A and B) in each asymmetric unit (Fig. S2b). The electron densities from repeat 1 through 11, particularly the key sites where R* and C or 5mC reside, are clearly observed (Fig. 4a and b). In contrast, the R*-5hmC unit cell contains four complexes (designated A, B, C, and D), similar to the structures of RG-5mC/5hmC, the qualities of the electron density of complexes A and B are greater than those of C and D (Figs. S2d and S5), and the electron densities of their key sites can be clearly observed (Fig. 4c).

When superimposed, repeat 6 of R*-C, R*-5mC, and R*-5hmC exhibit highly similar conformations, with RMSDs of 0.14–0.31 Å over all of the 34 C_α atoms (Fig. 4d). Because TALE RVDs are followed immediately by two conserved glycine residues (Gly14 and Gly15), R* is equivalent to RG except Gly13 is missing. When we compared R*-5mC with RG-5mC and HA-5mC, we found that the R* loop conformation is similar to that of RG rather than that of HA; however, the deletion of Gly13 results in a

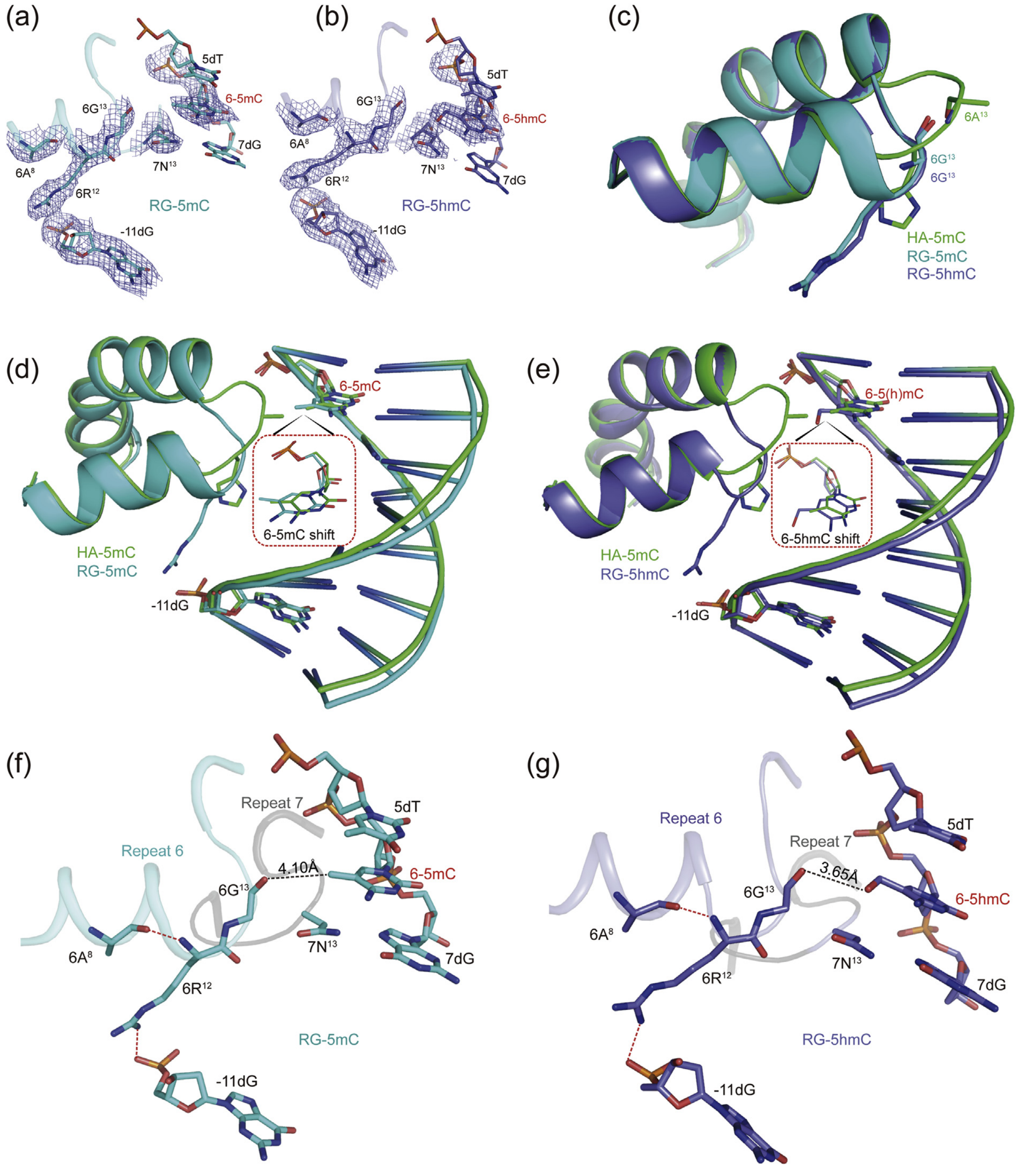

Fig. 3. Degenerate RVD RG forms a distinct loop conformation. (a) and (b) 2Fo-Fc electron density maps (slate blue mesh), contoured at 1.0σ, of the key sites of the RG-5mC (cyan stick model, cyan cartoon) and RG-5hmC (slate blue stick model, slate blue cartoon) structures, where RVD RG and 5mC or 5hmC reside. (c) Structural superimposition of repeat 6 with RVD RG or HA. Repeat 6 of RG-5mC (cyan) and RG-5hmC (slate blue) exhibit similar structures and can be superimposed with an RMSD of 0.153 Å over all of the 34 C_α atoms, while their loop conformations differed significantly when compared to HA (green). (d) and (e) Comparison of the repeat 6 and their corresponding DNA bases for RG-5mC, RG-5hmC, and HA-5mC. The RG loops are shifted away from the DNA major groove and toward the 3′ end of the DNA, and the corresponding DNA bases migrate to the RVD. (f) and (g) Structural basis for the recognition of 5mC and 5hmC by RVD RG. The Arg12 residue forms two H bonds (red dashed lines) with Ala8 and the DNA duplex, and the carbonyl oxygen of Gly13 forms a van der Waals interaction (black dashed line) with the 5-methyl group of 5mC or the 5-hydroxyl group of 5hmC.

Please cite this article as: L. Liu, Y. Zhang, M. Liu, et al., Structural Insights into the Specific Recognition of 5-methylcytosine and 5-hydroxymethylcytosine by TAL Effec..., Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.11.023

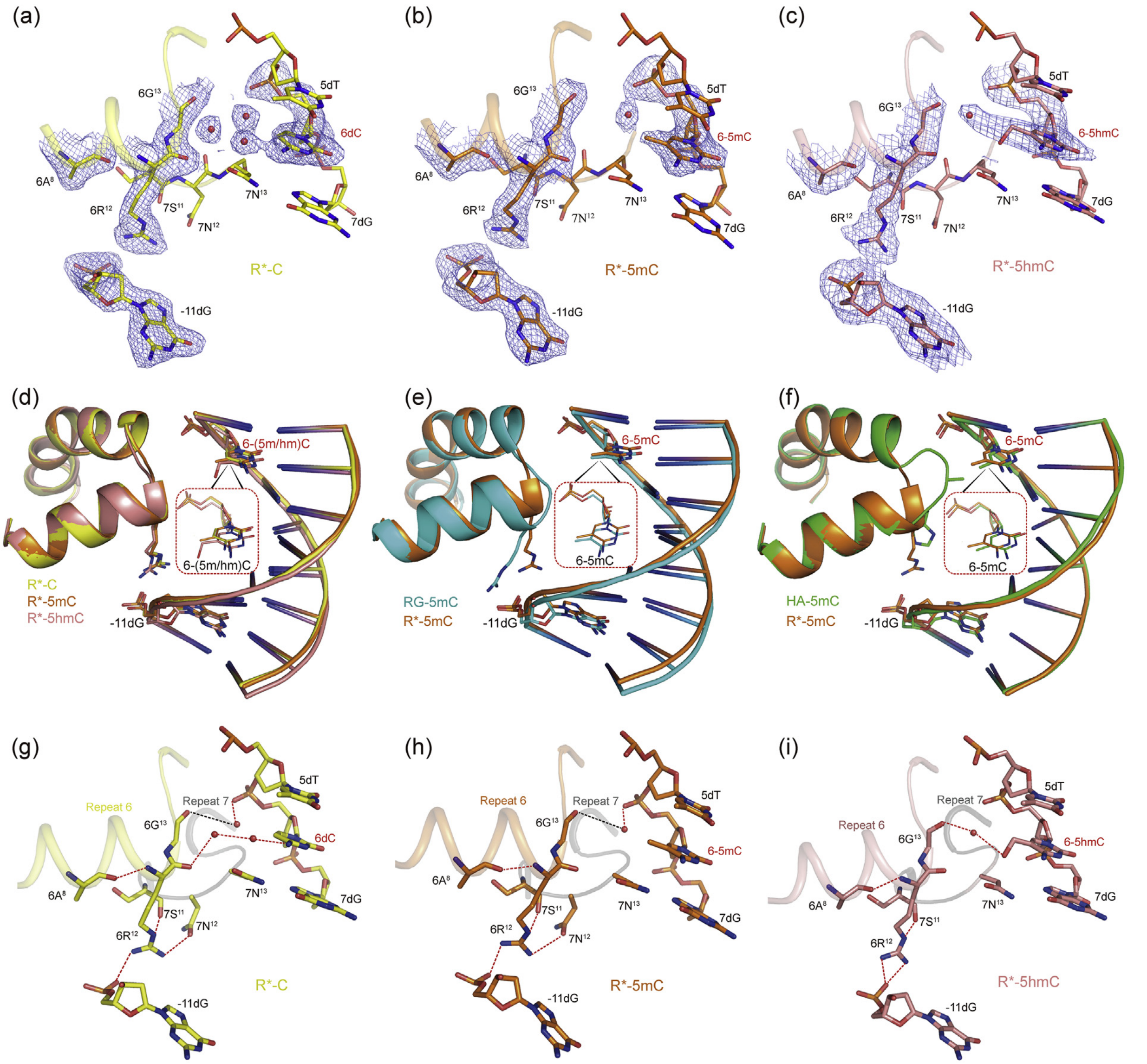

Fig. 4. Universal RVD R* accommodates C, 5mC, and 5hmC via water-mediated interactions. (a)–(c) 2Fo-Fc electron density maps (slate blue mesh), contoured at 1.0σ, of the key sites of R*-C (yellow), R*-5mC (orange), and R*-5hmC (salmon), water molecules shown as red spheres. (d) Structural superimposition of the repeat 6 and the corresponding DNA bases of R*-C, R*-5mC, and R*-5hmC uncover highly similar conformations with RMSDs of 0.14–0.31 Å. (e) and (f) Comparison of repeat 6 and the corresponding DNA bases of R*-5mC with that of RG-5mC and HA-5mC. The loop conformation and the corresponding DNA base location of R* are more similar to those of RG than of HA; however, deletion of Gly13 results in a truncated RVD loop that extends less deeply into the DNA major groove. (g)–(i) Structural basis for the recognition of C, 5mC, and 5hmC by RVD R*. The residue Arg12 of R* maintains a rather stable loop conformation through a total of four H-bond interactions with DNA and related amino acids, and Gly13 (originally Gly14) accommodates C, 5mC, or 5hmC through water-mediated interactions.

truncated RVD loop that extends less deeply into the DNA major groove (Fig. 4e and f).

Similar to RG, in the case of R*-C, R*-5mC and R*-5hmC, the main chain amino group of Arg12 forms an H bond with the carbonyl oxygen atom of Ala8 and the side chain amino group of Arg12 contacts the same phosphate group of the DNA duplex. In addition, the Arg12 of R* forms two additional H bonds with Ser11 (in repeat 7), Asn12 (in repeat 7), or DNA duplex (Fig. 4g–i). In summary, Arg12 of R* maintains a rather stable loop conformation through a total of four H-bond interactions with DNA duplex and related amino acids.

The truncated RVD loop of R* extends less deeply into the DNA major groove, thus Gly13 (originally Gly14) is located at a considerable distance (>5.5 Å)

Please cite this article as: L. Liu, Y. Zhang, M. Liu, et al., Structural Insights into the Specific Recognition of 5-methylcytosine and 5-hydroxymethylcytosine by TAL Effec..., Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.11.023

from the corresponding C, 5mC, and 5hmC bases and does not directly interact with them (Fig. 4g–i). In the R*-C structure, the carbonyl oxygen of Arg12 points to the DNA major groove, and there are two water molecules between Arg12 and C that form H bonds with the carbonyl oxygen of Arg12 and the amino group of C, respectively. Furthermore, an H bond also exists between the two water molecules, resulting in a water-mediated indirect interaction between Arg12 and C. There is also a water molecule between the carbonyl oxygen of Gly13 and the phosphate group of C, and it forms van der Waals and H-bond interactions with them, respectively, resulting in a water-mediated interaction between Gly13 and the phosphate skeleton of C (Fig. 4g). The case of R*-5mC is similar to R*-C; however, there is no water-mediated H bond between the carbonyl oxygen of Arg12 and the amino group of 5mC (Fig. 4h). We speculate that the 5-methyl group of 5mC would possibly introduce steric clash with the two water molecules presented in R*-C. Consistent with our hypothesis, the two water molecules are also absent from R*-5hmC possibly due to the presence of the 5-hydroxymethyl group. Instead, the main chain carbonyl oxygen of Gly13 interacts with the 5-hydroxyl group of 5hmC through water-mediated H bonds (Fig. 4i). In summary, Arg12 of R* maintains a rather stable loop conformation through a total of four H-bond interactions. Moreover, the truncated loop of R* is located at a considerable distance from the corresponding bases, allowing for flexible water-mediated interactions with C, 5mC, and 5hmC.

RVD Q* recognizes 5hmC through water-mediated H bonds

Previous screening identified 5hmC-selective RVDs, although their affinities are relatively weak [42]. We crystallized a TALE containing 5hmC-specific RVD Q* in complex with 5hmC-containing DNA (Q*-5hmC). The structure includes four

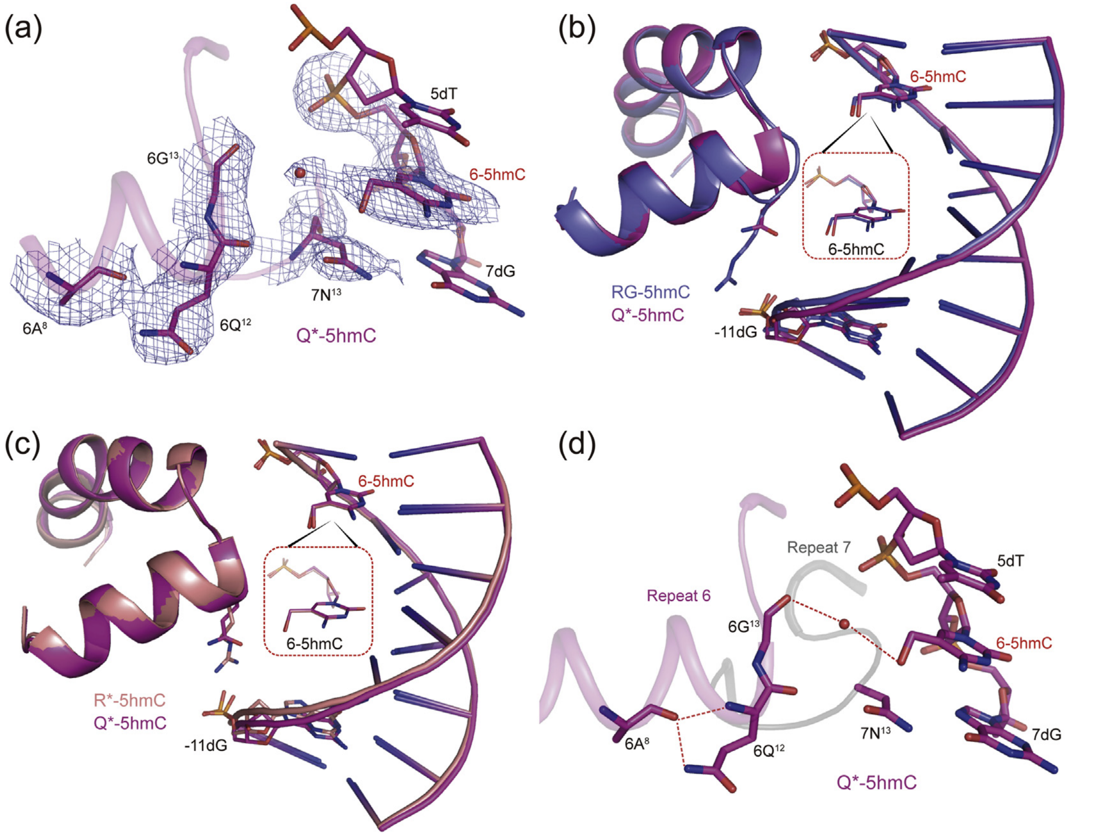

Fig. 5. The RVD Q* recognizes 5hmC through water-mediated H bonds. (a) 2Fo-Fc electron density maps (slate blue mesh), contoured at 1.0σ, of the key sites of Q*-C (light magenta), water molecules shown as red spheres. (b) and (c) Structural superimposition of the repeats 6 and the corresponding DNA bases of Q*-5hmC, RG-5hmC, and R*-5hmC. The loop conformation of Q* is similar to that of RG, except that the truncated loop of Q* is farther from the DNA major groove. The Q* loop in Q*-5hmC and the R* loop in R*-5hmC, as well as the two 5hmC, are precisely superimposed. (d) Structural basis for the recognition of 5hmC by RVD Q*. The amino groups of Gln12 form two H bonds with the carbonyl oxygen of Ala8 and Gly13 interacts with the hydroxyl group of 5hmC via water-mediated H bonds.

Please cite this article as: L. Liu, Y. Zhang, M. Liu, et al., Structural Insights into the Specific Recognition of 5-methylcytosine and 5-hydroxymethylcytosine by TAL Effec..., Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.11.023

complexes (designated A, B, C, and D) and is finally refined to 2.99 Å resolution (Fig. S2d, Table S2). Complexes A and B exhibit higher quality of electron density (Fig. S6). The electron density of the key sites where Q* and 5hmC reside is well defined (Fig. 5a).

When comparing Q*-5hmC with RG-5hmC, we found that the loop conformation of Q* is similar to that of RG, except that the truncated loop of Q* is further deviated from the DNA major groove due to deletion of the 13th amino acid (Fig. 5b). Meanwhile, the Q* loop in Q*-5hmC and the R* loop in R*-5hmC, as well as the two 5hmC bases, are precisely superimposed (Fig. 5c).

In contrast to Arg12 of R*, which forms a total of four H bonds with DNA and TALE, the main chain and side chain of Gln12 form two H bonds with the carbonyl oxygen of Ala8. Therefore, the loop conformation of Q* is less stable than that of R* (Fig. 5d). Similar to R*-5hmC, the carbonyl oxygen of Gly13 in Q*-5hmC interacts with the 5-hydroxyl group of 5hmC through water-mediated H bonds (Fig. 5d). Our previous cellular screening assay indicated that the binding affinity of Q* for 5hmC is outcompeted by R* [42], indicating that the first residue of RVD also contributes to binding affinity by forming a rather stable RVD loop conformation.

Discussion

Previous structural studies of TALE-DNA complexes focused mainly on canonical RVDs and revealed the first residue, either His12 or Asn12, whose side chains form direct hydrogen bonds with the carbonyl oxygen atom of the conserved Ala8, thereby constraining highly similar RVD loop conformations [16,17,35]. In the case of HA, the first residue of which is consistent with canonical RVDs, our crystal structures demonstrate that the loop conformation of HA is similar to that of canonical RVD loops as predicted. While for degenerate RVD RG, its first residue is mutated to noncanonical Arg and the mutated residue Arg has a larger side chain than canonical His or Asn. The structures show that the loop conformation of RG differs significantly, which deviates from the DNA major groove and shifts toward the 3’ end of the DNA. For universal RVD R* and 5hmC-specific RVD Q*, their first residues are still noncanonical; in addition, the second residues are missing, creating the truncated loops. The loop conformations of R* and Q* are similar to that of RG, except that the truncated loops of R* and Q* are further deviated from the DNA major grooves (Fig. 6a and S7). Meanwhile, the DNA bases corresponding to RVDs RG, R*, and Q* migrate to the RVDs (Fig. 6b). It is reasonable considering that when His12 or Asn12 is substituted with Arg12 or Gln12, the interactions with Ala8 are formed via the main chain amino group instead of the side chain. Hence, the loop conformations of RG, R*, and Q* are similar and deviate from the DNA major groove and meanwhile move toward the DNA 3′ end. In summary, mutations of His12 or Asn12 to Arg12 or Gln12 result in significant changes to RVD loop conformations, which greatly enhances our understanding of the plasticity of the TALE RVD loops.

Structures of natural TALEs showed that only the second residue of RVD directly interacts with the DNA base, while the first residue merely serves to stabilize the normal RVD loop conformation [16,17,35]. However, subsequent studies that deciphered canonical and noncanonical TALE RVDs for DNA recognition indicated that the first residue also

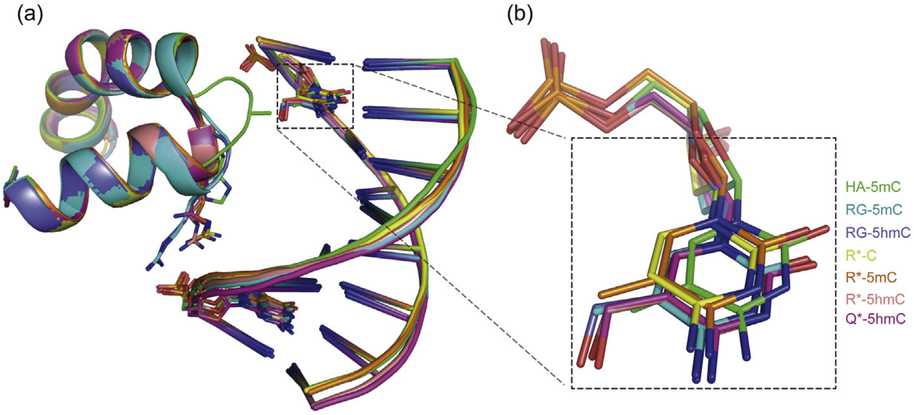

Fig. 6. Structural superimposition of the repeat 6 and the corresponding DNA bases of all the seven TALE-DNA complex structures. (a) The loop conformation of HA is similar to that of normal RVD loops, while that of degenerate RVD RG differs significantly, which deviates from the DNA major groove and shifts towards the 3′ end of the DNA. The loop conformations of universal RVD R* and 5hmC-specific RVD Q* are similar to that of RG, except that the truncated loops of R* and Q* are further deviated from the DNA major grooves due to deletion of the 13th amino acids. (b) The DNA bases corresponding to RVDs RG, R*, and Q* migrate to the RVDs.

modulates binding strength and specificity [6,7]. Our structures here may interpret the molecular basis of the observation. For degenerate RVD RG, besides the H bond formed by the main chain, the side chain of Arg12 forms an H bond with the phosphate group of the DNA (Fig. 3f and g). Compared with RG, the side chain of Arg12 in R* makes two additional H bonds with adjacent amino acids (Fig. 4g−i), while in Q*, the side chain of Gln12 forms only one H bond with Ala8 (Fig. 5d). Therefore, R* possesses the most stable loop conformation with Arg12 forming a total of four H bonds with DNA and TALE. R* could recognize C/5mC, although the interactions between Gly13 and C/5mC are quite weak. Thus, we hypothesize that the first residue of the RVD is also involved in recognition by maintaining a rather stable loop conformation to compensate for the weak direct contact with DNA bases. Consistent with this hypothesis, although RVDs Q* and R* sharing the nearly identical loop conformation and the same interaction manner with 5hmC base (Fig. 5c), the binding affinity of Q* for 5hmC is weaker than that of R* [42], probably because the loop conformation of Q* is less stable due to the less side-chain-mediated interactions. Furthermore, unlike R*, the labile loop of Q* may be unable to compensate for the potential weak direct interaction with C or 5mC. Therefore, Q* is selective for 5hmC owing to their relatively stronger direct interactions (Fig. 5d). Hence, our structures containing noncanonical TALE RVDs illustrate that the first residue of the RVDs contributes to the binding affinity and specificity as well.

It has been demonstrated that the side chains of His12 or Asn12 form a direct H bond with the conserved Ala8, thereby constraining the RVD loop [16,17]. However, in HA-5mC, the side chain of His12 forms an H bond to carbonyl oxygen of the conserved Ser11 (in the next repeat), with a loop conformation highly similar to that of normal RVDs. Furthermore, a previously reported NG-5mC structure also displays an H-bond interaction between the side chains of Asn12 and Ser11 [35]. These structures demonstrate that Ser11 is an excellent alternative to Ala8 and interacts with His12 or Asn12 to maintain a normal RVD loop conformation. This mechanism is of great significance for *Xanthomonas* bacteria. When Ala8 is mutated or deleted, Ser11 can act as a substitute to interact with the first residue of RVDs, maintaining the normal RVD loop conformations, thereby recognize the corresponding DNA base, and ultimately achieve infection of the host plant.

Unlike CRISPR-Cas9 that is insensitive to DNA modifications, TALE-DNA interaction is modification sensitive and thus provides the possibility for modification-dependent applications. For instance, novel 5mC-selective RVDs can be used to detect 5mC with higher resolution and sensitivity than 5mC antibody [39]. Meanwhile, methylation-dependent gene activation and genome editing can be achieved using these novel RVDs [42]. Furthermore, mitochondria-targeted TALEN (mito-TALENs) is proved to be an effective therapy for human mitochondrial diseases [13,14]; the modification-sensitive characteristic of TALEs enables its potential applications in the treatment of mitochondrial diseases associated with DNA modifications. Our crystals of TALE-DNA complexes elucidated the structural basis of specific recognition of 5mC and 5hmC by noncanonical TALE RVDs and provided insights into the plasticity of RVD loops. This enhances our understanding of TALE-DNA interactions and may promote the applications of TALEs in genetic manipulation and future precision therapy.

Our TALE-DNA complex structures indicate that some RVDs would probably recognize two other cytosine modifications 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). For instance, the truncated loops of R* and Q* are located at considerable distances from the corresponding C, 5mC, and 5hmC bases; therefore, R* and Q* would probably also tolerate and recognize 5fC and 5caC through somehow water-mediated interactions. TALE RVDs recognizing 5fC and 5caC remain to be screened by biochemical experiments, and further crystal structures are needed to illustrate the molecular bases.

Our previous publication reveals that RVDs HA, RG, R*, and Q* can also recognize thymine, which is most similar to 5mC [42]. This has little effect on the applications of TALE-based gene editing tools, because these tools are sequence-specific, and cytosine and modified cytosines can be differentiated from thymine by sequencing. Just in case, caution should be taken against DNA sequences with identical flanking sequence but containing either modified cytosines or T, although this possibility is very small.

Materials and methods

DNA synthesis and purification

DNA oligonucleotides containing 5mC and 5hmC were synthesized on an ABI Expedite 8909 nucleic acid synthesizer. The modified nucleotides were specifically incorporated at the desired positions using commercially available phosphoramidites (Glen Research). The DNAs were then deprotected using standard methods recommended by Glen Research manual. Purification was performed using Glen-Pak DNA purification cartridges (Glen Research) according to the manufacturer's instructions. Next, we performed urea-PAGE (polyacrylamide gel electrophoresis) to further improve DNA purity. The purified DNAs were validated by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) (Figs. S8a and b). Normal DNA oligonucleotides were purchased from Invitrogen. The ssDNA was annealed with 1.2-fold molar amount of the antisense strand by heating at 95 °C for 5 min and slow cooling to 4 °C over a period of 8 h, to a final concentration of 1 mM. The annealing buffer contained 10 mM Tris pH 7.5 and 100 mM NaCl.

Annealed modified dsDNA sequences:
5′- TGTCCCTTCGCGTCTCT- 3'.
3′- ACAGGGAAGCGCAGAGA- 5'.
The modified cytosine is underlined.

Protein expression and purification

Overexpression and purification of TALE proteins were performed following previously published protocols [16]. All TALEs with amino acids 231–720 of dHax3 [43] were subcloned into the pET21b vector (Novagen). Mutations of these proteins to include the desired RVDs were introduced using the Easy Mutagenesis System (Transgen Biotech). Plasmids encoding the engineered TALE proteins were transformed into *E. coli* BL21 (DE3) competent cells. Cells were grown at 37 °C and 220 rpm, and protein expression was induced with 0.5 mM isopropyl β-D-thiogalactoside (IPTG) when the OD600 reached 0.8. Following induction, cells were grown at 22 °C and 220 rpm for an additional 16 h. The cells were harvested by centrifugation at 4 °C for 30 min at 3000 g and homogenized in buffer containing 25 mM Tris-HCl pH 8.0 and 150 mM NaCl. We then performed sonication to lyse the cells, the sonicator probe was set to a frequency of 20 kHz, and the sample was subjected to a total of 40 min of sonication (sonication for 5 s with a 5 s interval between each sonication) until the sample was clear. After sonication, the cell lysis was centrifuged at 4 °C for about 40 min at 12,000 rpm. Thereafter, the supernatant was applied to a Ni^{2+}-nitrilotriacetate affinity resin (HISTrap, GE Healthcare) (Buffer A: 10 mM Tris-HCl pH 8.0, 150 mM NaCl and Buffer B: 10 mM Tris-HCl pH 8.0, 150 mM NaCl and 500 mM imidazole), a heparin column (GE Healthcare) (Buffer A: 10 mM Tris-HCl pH 8.0, 100 mM NaCl and Buffer B: 10 mM Tris-HCl pH 8.0, 1 M NaCl), and finally, a HiLoad 16/600 Superdex 200 pg (GE Healthcare) (Buffer GF: 25 mM MES pH 6.0, 50 mM NaCl, 5 mM $MgCl_2$, 10 mM DTT). The peak fractions from gel filtration were collected and concentrated to 5 mg/ml, flash frozen in liquid nitrogen, and stored at −80 °C. The molecular weight and purity of TALE proteins were determined by SDS-PAGE (Fig. S8c).

Crystallization

Before crystallization, the 5 mg/ml TALE proteins and 1 mM annealed dsDNA oligonucleotides were mixed in a 1:1.5 M ratio and incubated at 4 °C for at least 30 min. The TALE-DNA complex crystals were grown at 18 °C by sitting-drop vapor diffusion in a mother solution containing 8–10% PEG3350 (w/v), 10% ethanol, and 0.1 M MES pH 6.7 (TALE-DNA and mother solution were mixed with 1: 1 vol ratio). The crystals appeared within 1–2 days and grew to full size over approximately a week (Fig. S8d). As the initial diffractions of the crystals were not sufficient to accurately assign side chains, we optimized by dehydration. Crystal dehydration was performed by a serial transfer of the protein crystal into dehydrating solutions (50 μl), which are composed of the original mother solution supplemented with increasing concentrations of the precipitant PEG400 (HA-5mC, R*-C, and R*-5mC) or glycerine (RG-5mC, RG-5hmC, R*-5hmC, and Q*-5hmC), beginning with 5% (v/v) and increasing to 30%, in 5% increments. The crystals were incubated for 5 min at 18 °C in each condition. After dehydration, the crystals were harvested using fiber loops and stored in liquid nitrogen.

Data collection and structure determination

The TALE-DNA complex data sets were collected at the SSRF (Shanghai Synchrotron Radiation Facility, Shanghai) beamlines BL17U, BL18U1, and BL19U1 with Mar CCD [44]. All collected data sets were integrated and scaled with the HKL2000 and HKL3000 packages [45,46]. Further processing was carried out with programs from the CCP4 suite [47]. The initial models of the TALE-DNA complexes were determined by molecular replacement with the reported TALE-DNA complex structure (PDB accession code: 4GJP) as the original searching model using the program PHASER [48]. The structure was refined with WinCoot by building the remaining models into the electron density map [49] followed by refinement using Refmac5 using CCP4 [50]. All structure figures were prepared with PyMOL [51] using complex A of each structure except that RG-5mC used complex B. Data collection and structural refinement statistics are summarized in Tables S1 and S2.

Accession numbers

Atomic coordinates and structure factors described in this work have been deposited in the Protein Data Bank with accession codes **6JVZ** (HA-5mC), **6JW3** and **6JW4** (RG-5mC and RG-5hmC), **6JW0**, **6JW1**, and **6JW2** (R*-C, R*-5mC, and R*-5hmC), and **6JW5** (Q*-5hmC).

Acknowledgments

The authors thank Nieng Yan and Dong Deng (Tsinghua University, Beijing, China) for providing

Please cite this article as: L. Liu, Y. Zhang, M. Liu, et al., Structural Insights into the Specific Recognition of 5-methylcytosine and 5-hydroxymethylcytosine by TAL Effec..., Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.11.023

plasmid encoding the engineered dHax3, enabling our mutations of TALE proteins containing desired RVDs. We also gratefully acknowledge the technical suggestions of Dong Deng, Hu Zeng, and Chenxu Zhu with crystallization. We thank National Center for Protein Sciences at Peking University in Beijing, China for facility support. Use of Shanghai Synchrotron Radiation Facility (BL17U, BL18U1 and BL19U1) for X-ray data collection is also gratefully acknowledged. This work was supported by the Fok Ying Tung Education Foundation (no. 161018 to C.Y.) and the National Natural Science Foundation of China (nos. 91953201 and 21825701 to C.Y.).

Author contributions

L.L., J.P., and C.Y. designed the experiments; L.L. prepared the crystals, collected and processed X-ray data; Y.Z. assisted in preparing the plasmids encoding TALE proteins; M.L. and W.W. gave some key suggestions; L.L. wrote the manuscript; J.P. edited the manuscript.

Conflicts of interest

The authors declare no competing financial interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.11.023.

Received 11 July 2019;
Received in revised form 12 November 2019;
Accepted 27 November 2019
Available online xxxx

Keywords:
Crystal structures;
TAL effectors;
5-Methylcytosine;
5-Hydroxymethylcytosine;
Recognition

Abbreviations used:

5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; TALEs, Transcription activator-like effectors; RVDs, repeat-variable diresidues; HA, histidine and alanine; RG, arginine and glycine; R*, arginine and the deletion of the second residue of the RVD; Q*, glutamine and the deletion of the second residue of the RVD; NG, asparagine and glycine; N*, asparagine and the deletion of the second residue of the RVD; HD, histidine and aspartic acid; mtDNA, mitochondrial DNA; mito-TALENs, mitochondria-targeted TALEN.

References

[1] S. Kay, S. Hahn, E. Marois, G. Hause, U. Bonas, A bacterial effector acts as a plant transcription factor and induces a cell size regulator, Science 318 (2007) 648–651.

[2] P. Romer, S. Hahn, T. Jordan, T. Strauss, U. Bonas, T. Lahaye, Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene, Science 318 (2007) 645–648.

[3] J. Boch, U. Bonas, Xanthomonas AvrBs3 family-type III effectors: discovery and function, Annu. Rev. Phytopathol. 48 (2010) 419–436.

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

[5] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, et al., Breaking the code of DNA binding specificity of TAL-type III effectors, Science 326 (2009) 1509–1512.

[6] J.C. Miller, L. Zhang, D.F. Xia, J.J. Campo, I.V. Ankoudinova, D.Y. Guschin, et al., Improved specificity of TALE-based genome editing using an expanded RVD repertoire, Nat. Methods 12 (2015) 465–471.

[7] J. Yang, Y. Zhang, P. Yuan, Y. Zhou, C. Cai, Q. Ren, et al., Complete decoding of TAL effectors for DNA recognition, Cell Res. 24 (2014) 628–631.

[8] A.J. Bogdanove, D.F. Voytas, TAL effectors: customizable proteins for DNA targeting, Science 333 (2011) 1843–1846.

[9] R. Morbitzer, P. Romer, J. Boch, T. Lahaye, Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors, P Natl Acad Sci USA 107 (2010) 21617–21622.

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

[11] J.C. Miller, S.Y. Tan, G.J. Qiao, K.A. Barlow, J.B. Wang, D.F. Xia, et al., A TALE nuclease architecture for efficient genome editing, Nat. Biotechnol. 29 (2011), 143-U9.

[12] M. Hashimoto, S.R. Bacman, S. Peralta, M.J. Falk, A. Chomyn, D.C. Chan, et al., MitoTALEN: a general approach to reduce mutant mtDNA loads and restore oxidative phosphorylation function in mitochondrial diseases, Mol. Ther. 23 (2015) 1592–1599.

[13] S.R. Bacman, S.L. Williams, M. Pinto, S. Peralta, C.T. Moraes, Specific elimination of mutant mitochondrial genomes in patient-derived cells by mitoTALENs, Nat. Med. 19 (2013) 1111–1113.

[14] P. Reddy, A. Ocampo, K. Suzuki, J. Luo, S.R. Bacman, S.L. Williams, et al., Selective elimination of mitochondrial mutations in the germline by genome editing, Cell 161 (2015) 459–469.

[15] S.R. Bacman, J.H.K. Kauppila, C.V. Pereira, N. Nissanka, M. Miranda, M. Pinto, et al., MitoTALEN reduces mutant mtDNA load and restores tRNA(Ala) levels in a mouse model of heteroplasmic mtDNA mutation (vol 24, pg 1696, 2018), Nat. Med. 24 (2018) 1940.

[16] D. Deng, C.Y. Yan, X.J. Pan, M. Mahfouz, J.W. Wang, J.K. Zhu, et al., Structural basis for sequence-specific

recognition of DNA by TAL effectors, Science 335 (2012) 720–723.

[17] A.N.S. 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.

[18] M.G. Goll, T.H. Bestor, Eukaryotic cytosine methyltransferases, Annu. Rev. Biochem. 74 (2005) 481–514.

[19] M. Ehrlich, R.Y.H. Wang, 5-Methylcytosine in eukaryotic DNA, Science 212 (1981) 1350–1357.

[20] S.K.T. Ooi, A.H. O'Donnell, T.H. Bestor, Mammalian cytosine methylation at a glance, J. Cell Sci. 122 (2009) 2787–2791.

[21] A. Bird, DNA methylation patterns and epigenetic memory, Genes Dev. 16 (2002) 6–21.

[22] M.M. Suzuki, A. Bird, DNA methylation landscapes: provocative insights from epigenomics, Nat. Rev. Genet. 9 (2008) 465–476.

[23] M. Tahiliani, K.P. Koh, Y.H. Shen, W.A. Pastor, H. Bandukwala, Y. Brudno, et al., Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1, Science 324 (2009) 930–935.

[24] S. Ito, A.C. D'Alessio, O.V. Taranova, K. Hong, L.C. Sowers, Y. Zhang, Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification, Nature 466 (2010) 1129–1133.

[25] D. Globisch, M. Munzel, M. Muller, S. Michalakis, M. Wagner, S. Koch, et al., Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates, PLoS One 5 (2010).

[26] S. Kriaucionis, N. Heintz, The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain, Science 324 (2009) 929–930.

[27] M. Mellen, P. Ayata, S. Dewell, S. Kriaucionis, N. Heintz, MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous System, Cell 151 (2012) 1417–1430.

[28] M. Yu, G.C. Hon, K.E. Szulwach, C.X. Song, L. Zhang, A. Kim, et al., Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome, Cell 149 (2012) 1368–1380.

[29] M. Bachman, S. Uribe-Lewis, X.P. Yang, M. Williams, A. Murrell, S. Balasubramanian, 5-Hydroxymethylcytosine is a predominantly stable DNA modification, Nat. Chem. 6 (2014) 1049–1055.

[30] M.C. Haffner, A. Chaux, A.K. Meeker, D.M. Esopi, J. Gerber, L.G. Pellakuru, et al., Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers, Oncotarget 2 (2011) 627–637.

[31] S.G. Jin, Y. Jiang, R.X. Qiu, T.A. Rauch, Y.S. Wang, G. Schackert, et al., 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations, Cancer Res. 71 (2011) 7360–7365.

[32] C.G. Lian, Y.F. Xu, C. Ceol, F.Z. Wu, A. Larson, K. Dresser, et al., Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma, Cell 150 (2012) 1135–1146.

[33] J. Valton, A. Dupuy, F. Daboussi, S. Thomas, A. Marechal, R. Macmaster, et al., Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation, J. Biol. Chem. 287 (2012) 38427–38432.

[34] A. Dupuy, J. Valton, S. Leduc, J. Armier, R. Galetto, A. Gouble, et al., Targeted gene therapy of xeroderma pigmentosum cells using meganuclease and TALEN, PLoS One 8 (2013), e78678.

[35] D. Deng, P. Yin, C. Yan, X. Pan, X. Gong, S. Qi, et al., Recognition of methylated DNA by TAL effectors, Cell Res. 22 (2012) 1502–1504.

[36] J.B. Hu, Y. Lei, W.K. Wong, S.Q. Liu, K.C. Lee, X.J. He, et al., Direct activation of human and mouse Oct4 genes using engineered TALE and Cas9 transcription factors, Nucleic Acids Res. 42 (2014) 4375–4390.

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

[38] 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. 53 (2014) 6002–6006.

[39] 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.

[40] G. Kubik, S. Batke, D. Summerer, Programmable sensors of 5-hydroxymethylcytosine, J. Am. Chem. Soc. 137 (2015) 2–5.

[41] 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.

[42] Y. Zhang, L.L. Liu, S.J. Guo, J.H. Song, C.X. Zhu, Z.W. Yue, et al., Deciphering TAL effectors for 5-methylcytosine and 5-hydroxymethylcytosine recognition, Nat. Commun. 8 (2017).

[43] M.M. Mahfouz, L.X. Li, M. Shamimuzzaman, A. Wibowo, X.Y. Fang, J.K. Zhu, De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks, P Natl Acad Sci USA 108 (2011) 2623–2628.

[44] Q.S. Wang, K.H. Zhang, Y. Cui, Z.J. Wang, Q.Y. Pan, K. Liu, et al., Upgrade of macromolecular crystallography beamline BL17U1 at SSRF, Nucl. Sci. Tech. 29 (2018).

[45] Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode, Methods Enzymol. 276 (1997) 307–326.

[46] W. Minor, M. Cymborowski, Z. Otwinowski, M. Chruszcz, HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes, Acta Crystallogr D Biol Crystallogr 62 (2006) 859–866.

[47] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, et al., Overview of the CCP4 suite and current developments, Acta Crystallogr. D 67 (2011) 235–242.

[48] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R.J. Read, Phaser crystallographic software, J. Appl. Crystallogr. 40 (2007) 658–674.

[49] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. D 60 (2004) 2126–2132.

[50] A.A. Vagin, R.A. Steiner, A.A. Lebedev, L. Potterton, S. McNicholas, F. Long, et al., REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use, Acta Crystallogr D Biol Crystallogr 60 (2004) 2184–2195.

[51] N. Alexander, N. Woetzel, J. Meiler, Bcl::Cluster : a method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System, IEEE Int Conf Comput Adv Bio Med Sci 2011 (2011) 13–18.

Please cite this article as: L. Liu, Y. Zhang, M. Liu, et al., Structural Insights into the Specific Recognition of 5-methylcytosine and 5-hydroxymethylcytosine by TAL Effec..., Journal of Molecular Biology, https://doi.org/10.1016/j.jmb.2019.11.023