Miniature CRISPR-Cas12 endonucleases – Programmed DNA targeting in a smaller package

Giang T. Nguyen[a,b], Yukti Dhingra[a,b], and Dipali G. Sashital[a*]

[a] Roy J. Carver Department of Biochemistry, Biophysics, and Molecular Biology, Iowa State University, Ames, IA 50010

[b] Equal contribution

[*]Correspondence to sashital@iastate.edu

Abstract

CRISPR associated (Cas) endonucleases specifically target and cleave RNA or DNA based on complementarity to a guide RNA. Cas endonucleases – including Cas9, Cas12a, and Cas13 – have been adopted for a wide array of biotechnological tools, including gene editing, transcriptional modulation, and diagnostics. These tools are facilitated by ready reprogramming of guide RNA sequences and the varied nucleic acid binding and cleavage activities observed across diverse Cas endonucleases. However, the large size of most Cas endonucleases (950–1,400 amino acids) can restrict applications. The recent discovery of miniature Cas endonucleases (400-800 amino acids) provides the potential to overcome this limitation. Here we review recent advances in understanding the structural mechanisms of two miniature Cas endonucleases, Cas12f and Cas12j.

Short title: Miniature CRISPR-Cas12 endonucleases

Keywords: CRISPR-Cas; Cas12; gene editing; endonucleases; biotechnology

Introduction

The use of Cas9, Cas12a and Cas13 for biotechnology has prompted significant interest in their structure and function over the past decade [1,2]. These RNA-guided endonucleases are derived from CRISPR-Cas systems, a diverse collection of defense mechanisms in bacteria and archaea [3]. Defense against viruses and plasmids relies on Cas effector proteins, which universally use guide RNAs to target a complementary nucleic acid from an invading genome [4–11]. However, Cas effectors are highly evolutionarily diverged, leading to a variety of nucleic acid targeting and cleavage mechanisms [12]. Cas effectors may be multi-protein complexes (class1) or single protein endonucleases (class 2). Even within these classes, Cas effectors diverge greatly in composition and activity, leading to the classification of types and sub-types based on the presence of signature Cas effectors. Among these, the type V systems are the most diverse, with 12 sub-types (type V-A through V-K) and substantial variation between the signature Cas12 proteins. Cas12a (formerly Cpf1), the first discovered type V endonuclease, has been extensively studied biochemically and structurally, and widely adopted for genome editing and diagnostic tools [10,13,22–24,14–21].

The function of Cas12a can be broken down into multiple stages including guide RNA-protein (RNP) formation, target discovery and initiation of DNA unwinding, complementary base pairing between guide RNA and target DNA to form an R-loop, and finally target DNA cleavage (Figure 1A) [1]. These wide-ranging functions underly the large, multi-domain architecture of Cas12a, which contains several different domains to undertake these varied tasks (Figure 1B). Metagenomic analysis of microbial communities have recently uncovered smaller type V Cas endonucleases, Cas12f (400–700 amino acids, also known as Cas14) from uncultured bacteria and archaea and Cas12j (700-800 amino acids, also known as CasΦ) from the genomes of huge phages [25–27]. These small Cas12 proteins are about half the size of Cas12a and lack detectable sequence identity, except in the conserved RuvC nuclease domain (Figure 1C,D). Although Cas12f and Cas12j are devoid of certain functional domains compared to Cas12a, these proteins retain many of the functionalities of their larger counterparts [26,28,29]. Recent structural and functional studies of these two proteins have revealed how they can accomplish these tasks despite their diminutive size. Here, we use Cas12a as a point of comparison for these two smaller Cas12 proteins to highlight their structural mechanisms.

Overall architecture of Cas12 RNPs

Similar to other Cas endonucleases, including Cas9 and Cas13, Cas12a assumes a typical bilobed architecture, consisting of recognition (REC) and nuclease (NUC) lobes (Figure 1B) [13–15,17,18,22]. The Cas12a RNP contains a CRISPR RNA (crRNA), which acts as a guide for DNA target binding. Cas12a contains an RNase active site for processing the primary crRNA transcript [30]. Both the RNase-containing WED domain and the DNA-cleaving RuvC domain are contained within the Cas12a NUC lobe, while the REC lobe interacts extensively with the guide region of the crRNA [13–15,30].

The Cas12j RNP is a true miniature of Cas12a, containing a similar bilobed structure consisting of smaller REC and NUC lobes and a single crRNA, but lacking several comparable Cas12a domains (Figure 1D) [31,32]. Despite the smaller size of the Cas12j protein, the crRNA guide is similar in size to that of Cas12a (~43 nt). Both Cas12a and Cas12j recognize stem-loops within the crRNA using WED domains. Similar to Cas12a, the guide region of the crRNA extends along the length of the REC domain in Cas12j, presenting this region for base-pairing with the target DNA.

In contrast to Cas12j, the Cas12f RNP structure diverges substantially from Cas12a. Cas12f functions as a dimer and requires an additional trans-activating crRNA (tracrRNA) [26,33,34]. Similar to Cas9 and other Cas12 endonucleases that require a tracrRNA [35], the crRNA and tracrRNA can be linked together to form a single-guide RNA (sgRNA) [8] that can be used to activate Cas12f [26]. Within the dimer, two Cas12f subunits use the sgRNA as a scaffold for RNP assembly [33,34]. Assembly of the monomers in an asymmetric fashion enables formation of functional REC and NUC lobes with the sgRNA positioned in a central channel between the two lobes (Figure 1C). The Cas12f sgRNA is substantially larger than the crRNAs of Cas12a and Cas12j, comprising a 37 nt-crRNA linked to a 140 nt-tracrRNA. The tracrRNA portion of the sgRNA consists of 5 stem-loops mainly recognized by the WED and RuvC domains of both Cas12f monomers, which play a central role in dimerization. In addition, both Cas12f monomers contribute individual REC domains to form an extended recognition channel, along which the crRNA guide region extends, similar to other Cas endonucleases (Figure 1C).

Guide RNA processing by Cas12a and Cas12j

The lack of tracrRNA requirement for Cas12a and Cas12j underscores another similarity between the two proteins. While Cas12f likely recruits a host RNase to cleave the pre-crRNA:tracrRNA duplex [26] as observed in other tracrRNA-requiring systems [35], Cas12j, like Cas12a, processes its own precursor-crRNA (pre-crRNA) into mature crRNA [29]. Cas12a harbors an RNase active site within the WED domain, which binds specifically to the hairpin of the crRNA repeat [17,30]. The Cas12a RNase is specific for ribose, and thus cleaves RNA only

[30]. In contrast, Cas12j does not harbor a distinct RNase active site, and instead uses the RuvC active site to process pre-crRNA [29]. Although Cas12g, a related type V endonuclease, has been shown to cleave RNA using its RuvC domain [36], Cas12j is unique in its dual-purposing of a single active site for both crRNA maturation and target DNA cleavage. This multi-tasking domain accounts for some of the reduction in the overall size of Cas12j.

Target searching and DNA unwinding

Cas endonucleases must accomplish the challenging task of rapidly searching DNA to find a target that is complementary to the crRNA. DNA-binding Cas effectors universally simplify this searching process by first searching for protospacer adjacent motif (PAM) sequences that are located adjacent to target regions [8,10,37,38]. PAM searching reduces the number of locations at which DNA unwinding must be initiated to determine whether a crRNA complement is present. In addition, PAM recognition is associated with DNA destabilization, enabling initiation of DNA unwinding and annealing of the crRNA-target DNA heteroduplex (Figure 1A).

Like Cas12a [10], the best-characterized Cas12f and Cas12j orthologs recognize 5’-T-rich PAMs [28,29]. In all three Cas12 proteins, PAM recognition occurs through hydrogen bonding and van der Waals interactions between the protein and both strands of the PAM (Figure 2A, B, C). Three conserved lysines in the loop-lysine helix-loop (LKL) region of the PAM interaction (PI) domain of Cas12a are in proximity of the PAM sequence (Figure 2A) [15–18]. However, only one of these lysine residues forms a hydrogen bond with the PAM, suggesting PAM readout occurs mainly via shape recognition. The other two lysine residues within the LKL region facilitate DNA unwinding. Two conserved prolines in the LKL region help to properly orient the lysine residues, along with conserved methionine and acidic residues, for insertion into the PAM region, promoting unwinding of the PAM-proximal base pairs. Rotation of the backbone between the PAM and the first nucleotide of the target is thought to be stabilized by interactions with a lysine in the WED domain, referred to as a “phosphate lock” (Figure 2A).

PAM readout in both Cas12j and Cas12f is facilitated by specific hydrogen bonding and van der Waals interactions with residues located in the WED and RECI domains (Figure 2B, C). Similar to Cas12a, Cas12j harbors a PI domain containing two conserved lysine residues, one of which reads out a base in the PAM and the other of which facilitates unwinding of the DNA target [31,32] (Figure 2C). In contrast, Cas12f does not contain a comparable PI domain containing an LKL-like region (Figure 1B, Figure 2B) [33,34]. Instead, an arginine in a RECI domain helix facilitates DNA unwinding by inserting between the unwound base pairs (Figure 2B) [34]. Similar to Cas12a, the WED domain in both Cas12j and Cas12f recognize backbone phosphate groups between the PAM and the first nucleotide of the target, suggesting that they share similar phosphate locking mechanisms for heteroduplex formation with other Cas endonucleases (Figure 2A, B, C right panels) [15,17,18,31–34].

Conformational changes in target DNA-bound state enable DNA cleavage

Following PAM-recognition and DNA unwinding, the guide RNA can fully hybridize with the target strand of the DNA, forming an R-loop (Figure 1A). R-loop formation induces significant conformational rearrangements within each Cas12 protein. In Cas12a, the distance between the REC and NUC lobe widens, allowing accommodation of the heteroduplex in a positively-charged channel formed between the two lobes (Figure 3A) [13,17]. A similar positively charged central channel formed by REC and NUC lobes is formed in the Cas12f dimer (Figure 3B) [33,34]. In contrast, Cas12j adopts a T-shaped architecture instead of the typical bilobed architecture upon R-loop formation, in which the RECI and RECII domains move about 50 Å away from each other to accommodate and then wrap around the crRNA-DNA (Figure 3C) [31,32].

Conformational changes upon target binding also serve to activate the RuvC nuclease domains in each Cas12 protein. Notably, all three Cas12 proteins contain a “lid” region of the RuvC domain, which blocks access to the active site prior to target binding (Figure 3A, B, C) [19,31,32,34]. Upon conformational rearrangement of the REC and NUC lobes, the Cas12a lid loop adopts an α-helical structure, which unblocks the catalytic site in the RuvC domain and allows Cas12a to adopt an active conformation (Figure 3A) [19,22]. Cas12j and Cas12f conformational changes similarly cause alterations in the lid region that likely lead to nuclease activation (Figure 3B, C) [31,32,34], indicating that the smaller Cas12 proteins are regulated in a similar manner to Cas12a. Although the Cas12f dimer consists of two RuvC domains, only one domain is active (Figure 3B left) [33,34]. Upon DNA binding, the active RuvC domain undergoes a conformational change that allows access to the DNA substrate (Figure 3B right)

[34]. The lid domain of the second RuvC remains “closed” (Figure 3B left), and the domain is further inactivated by insertion of two purine bases of the tracrRNA at the catalytic site.

Upon activation, Cas12 proteins use their single active RuvC site to first cleave the displaced NTS, followed by TS cleavage [22,28,29,31]. Cleavage of the NTS and the formation of the complete R-loop is thought to allow the TS to enter the RuvC catalytic site [23,24], although the exact mechanism of TS cleavage by Cas12a has not been structurally characterized. Cas12a is proposed to share a similar TS cleavage mechanism with Cas12b, in which the TS bends to bind in the RuvC active site [19,39]. Interestingly, substrate DNA is found in a bent configuration in the RuvC catalytic pocket of Cas12f, which is similarly observed in Cas12b (Figure 3B) [34,39]. After NTS cleavage in Cas12j, bending of the unwound TS at the PAM distal end positions the TS in a parallel orientation to the NTS providing a geometry favorable for cleavage of the TS by RuvC active site [31]. These structural observations suggest that Cas12a, Cas12f and Cas12j may use similar mechanisms to cleave the TS.

After generating a double-stranded break, Cas12a releases the PAM-distal DNA, but remains bound tightly to the PAM-proximal DNA. Therefore, Cas12a is maintained in an activated state that is able to generate trans cleavage of non-target ssDNA and dsDNA [20–22,40,41]. Cas12j and Cas12f similarly remain activated following target cleavage [26,29]. These activities have been exploited for diagnostic tool development (Figure 4) [20,26,42], although the relevance of this non-specific cleavage in vivo remains unclear [43].

Cleavage efficiencies and specificities of compact Cas12 endonucleases

Although Cas12f and Cas12j retain the activities of larger Cas endonucleases, their compact size comes at the expense of efficiency. Small Cas12 enzymes cleave target DNA much more slowly than Cas12a, resulting in slow cleavage kinetics and poor genome editing outcomes in human cells [28,29,44]. These slower kinetics may be a result of reduction of stabilizing contacts between the smaller proteins and the RNA-DNA heteroduplex. Indeed, protein engineering of Cas12f to introduce additional favorable contacts for contacting the nucleic acids greatly enhanced the efficiency of Cas12f-mediated gene activation and editing in human cells [44]. Additional structural features within Cas12f and Cas12j may also attenuate cleavage rate. Helix α7 of the Cas12j RECI domain blocks access of single-stranded DNA to the active site, regulating cleavage activity [31]. Intriguingly, truncation of helix α7 causes a substantially higher cleavage rate compared to the wild type Cas12j.

Slower cleavage kinetics also likely impact the specificity of smaller Cas12 enzymes. As putative immune effector proteins, Cas endonucleases are expected to have some level of non-specificity toward their target DNA. Indeed, Cas12a can cleave target sequences with up to four mismatches [40]. While the specificities of Cas12f and Cas12j remain to be extensively studied, early reports indicate that they may have low tolerance for mismatches in the PAM-proximal region of the target [31,45]. Importantly, this low mismatch tolerance is alleviated upon improvement of Cas12j cleavage efficiency through truncation of α7 [31], suggesting that the structural features of small Cas12 proteins that reduce their cleavage efficiencies also increase their specificity. Thus, efforts to engineer more efficient versions of Cas12j and Cas12f may have a side effect of also increasing their potential for off-target cleavage.

Conclusion

CRISPR-Cas9 and Cas12a are versatile tools for a variety of applications (Figure 4). However, the large size of Cas9 and Cas12a is not suitable for packing into viral vectors, including commonly used adeno-associated virus vectors, which are limited to cargo sizes of <5 kb for efficient viral packaging [46]. Compact Cas proteins, such as Cas12f and Cas12j, should enable the creation of precise genome-editing tools that are easier to deliver to somatic tissues (Figure 4). The similarities in cleavage mechanisms between Cas12a and miniature Cas effectors suggest that these proteins will be useful tools, while their slower cleavage kinetics suggest that they require further optimization. Indeed, recent studies have greatly improved the initial poor efficiencies of Cas12f gene editing through both protein and guide RNA engineering [44,45,47,48]. These recent advancements underscore the importance of continued exploration of Cas12 diversity to expand the CRISPR toolbox.

Acknowledgements

We thank members of the Sashital lab for helpful discussion. D.G.S. acknowledges funding from the National Institutes of Health (GM140876 and GM115874).

Declaration of Interests

The authors declare no conflict of interest.

References

1. Murugan K, Babu K, Sundaresan R, Rajan R, Sashital DG: **The Revolution Continues: Newly Discovered Systems Expand the CRISPR-Cas Toolkit**. *Mol Cell* 2017, **68**:15–25.

2. Pickar-Oliver A, Gersbach CA: **The next generation of CRISPR–Cas technologies and applications**. *Nat Rev Mol Cell Biol* 2019, **20**:490–507.

3. Marraffini LA: **CRISPR-Cas immunity in prokaryotes**. *Nature* 2015, **526**:55–61.

4. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P: **CRISPR provides acquired resistance against viruses in prokaryotes**. *Science* 2007, **315**:1709–1712.

5. Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin E V., Van Der Oost J: **Small CRISPR RNAs guide antiviral defense in prokaryotes**. *Science* 2008, **321**:960–964.

6. Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP: **RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex**. *Cell* 2009, **139**:945–956.

7. Garneau JE, Dupuis MÈ, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S: **The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA**. *Nature* 2010, **468**:67–71.

8. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity**. *Science* 2012, **337**:816–821.

9. Gasiunas G, Barrangou R, Horvath P, Siksnys V: **Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria**. *Proc Natl Acad Sci U S A* 2012, **109**.

10. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler

P, Volz SE, Joung J, Van Der Oost J, Regev A, et al.: **Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System**. *Cell* 2015, **163**:759–771.

11. Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, Shmakov S, Makarova KS, Semenova E, Minakhin L, et al.: **C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector**. *Science* 2016, **353**.

12. Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, Charpentier E, Cheng D, Haft DH, Horvath P, et al.: **Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants**. *Nat Rev Microbiol* 2020, **18**:67–83.

13. Dong D, Ren K, Qiu X, Zheng J, Guo M, Guan X, Liu H, Li N, Zhang B, Yang D, et al.: **The crystal structure of Cpf1 in complex with CRISPR RNA**. *Nature* 2016, **532**:522–526.

14. Gao P, Yang H, Rajashankar KR, Huang Z, Patel DJ: **Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition**. *Cell Res* 2016, **26**:901–913.

15. Yamano T, Nishimasu H, Zetsche B, Hirano H, Slaymaker IM, Li Y, Fedorova I, Nakane T, Makarova KS, Koonin E V., et al.: **Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA**. *Cell* 2016, **165**:949–962.

16. Yamano T, Zetsche B, Ishitani R, Zhang F, Nishimasu H, Nureki O: **Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1**. *Mol Cell* 2017, **67**:633-645.e3.

17. Swarts DC, van der Oost J, Jinek M: **Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a**. *Mol Cell* 2017, **66**:221-233.e4.

18. Stella S, Alcón P, Montoya G: **Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage**. *Nature* 2017, **546**:559–563.

19. Stella S, Mesa P, Thomsen J, Paul B, Alcón P, Jensen SB, Saligram B, Moses ME, Hatzakis NS, Montoya G: **Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity**. *Cell* 2018, **175**:1856-1871.e21.

20. Chen JS, Ma E, Harrington LB, Da Costa M, Tian X, Palefsky JM, Doudna JA: **CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity**. *Science* 2018, **360**:436–439.

21. Li S-Y, Cheng Q-X, Liu J-K, Nie X-Q, Zhao G-P, Wang J: **CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA**. *Cell Res* 2018, **28**:491–493.

22. Swarts DC, Jinek M: **Mechanistic Insights into the cis- and trans-Acting DNase Activities of Cas12a.** *Mol Cell* 2019, **73**:589-600.e4.

23. Zhang L, Sun R, Yang M, Peng S, Cheng Y, Chen C: **Conformational Dynamics and Cleavage Sites of Cas12a Are Modulated by Complementarity between crRNA and DNA**. *iScience* 2019, **19**:492–503.

24. Cofsky JC, Karandur D, Huang CJ, Witte IP, Kuriyan J, Doudna JA: **CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks**. *Elife* 2020, **9**:e55143.

25. Shmakov S, Smargon A, Scott D, Cox D, Pyzocha N, Yan W, Abudayyeh OO, Gootenberg JS, Makarova KS, Wolf YI, et al.: **Diversity and evolution of class 2 CRISPR-Cas systems**. *Nat Rev Microbiol* 2017, **15**:169–182.

26. Harrington LB, Burstein D, Chen JS, Paez-Espino D, Ma E, Witte IP, Cofsky JC, Kyrpides NC, Banfield JF, Doudna JA: **Programmed DNA destruction by miniature CRISPR-Cas14 enzymes**. *Science* 2018, **362**:839–842.

27. Al-Shayeb B, Sachdeva R, Chen LX, Ward F, Munk P, Devoto A, Castelle CJ, Olm MR, Bouma-Gregson K, Amano Y, et al.: **Clades of huge phages from across Earth's ecosystems**. *Nature* 2020, **578**:425–431.

28. Karvelis T, Bigelyte G, Young JK, Hou Z, Zedaveinyte R, Budre K, Paulraj S, Djukanovic V, Gasior S, Silanskas A, et al.: **PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage**. *Nucleic Acids Res* 2020, **48**:5016–5023.

• This work showed that 10 Cas12f orthologs recognize and cleave dsDNA in a PAM dependent manner. The authors use a biochemical approach to demonstrate that Cas12f orthologs can protect against invading dsDNA, similar to the much larger class 2 CRISPR effectors.

29. Pausch P, Al-Shayeb B, Bisom-Rapp E, Tsuchida CA, Li Z, Cress BF, Knott GJ, Jacobsen SE, Banfield JF, Doudna JA: **CRISPR-CasΦ from huge phages is a hypercompact genome editor**. *Science* 2020, **369**:333–337.
• This study reported that Cas12j (CasΦ) uses a single active site for crRNA processing and DNA cleavage. Similar to Cas12a, Cas12j does not require tracrRNA. Additionally, Cas12j is capable of low levels of genome editing in eukaryotic cells.

30. Fonfara I, Richter H, BratoviÄ M, Le Rhun A, Charpentier E: **The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA**. *Nature* 2016, **532**:517–521.

31. Pausch P, Soczek KM, Herbst DA, Tsuchida CA, Al-Shayeb B, Banfield JF, Nogales E, Doudna JA: **DNA interference states of the hypercompact CRISPR–CasΦ effector**. *Nat Struct Mol Biol* 2021, **28**:652–661.
•• The authors solved cryo-EM structures of Cas12j (CasΦ) in pre- and post-DNA-bound states to elucidate the DNA binding and cleavage mechanisms. The study also created mutant Cas12j variants with cleavage rates ~20-fold faster compared to wild-type.

32. Carabias A, Fuglsang A, Temperini P, Pape T, Sofos N, Stella S, Erlendsson S, Montoya G: **Structure of the mini-RNA-guided endonuclease CRISPR-Cas12j3**. *Nat Commun* 2021, **12**:4476.
•• The authors present a cryoEM structure of Cas12j3/R-loop complex after DNA cleavage. The structure provides molecular details for PAM recognition, R-loop formation and DNA cleavage.

33. Takeda SN, Nakagawa R, Okazaki S, Hirano H, Kobayashi K, Kusakizako T, Nishizawa T, Yamashita K, Nishimasu H, Nureki O: **Structure of the miniature type V-F CRISPR-Cas effector enzyme**. *Mol Cell* 2021, **81**:558-570.e3.
•• This study reported the first cryo-EM structure of Cas12f-sgRNA-target DNA complex revealing an asymmetric homodimer assembly with a single guide RNA. Cas12f1 shares the mechanistic conservation among the type V Cas12 enzymes.

34. Xiao R, Li Z, Wang S, Han R, Chang L: **Structural basis for substrate recognition and**

cleavage by the dimerization-dependent CRISPR-Cas12f nuclease. *Nucleic Acids Res* 2021, **49**:4120–4128.

•• In this paper, the cryo-EM structures of Cas12f-sgRNA-target DNA and Cas12f-sgRNA complexes were determined in order to understand the mechanisms underlying substrate recognition and cleavage. The structures reveal an asymmetric Cas12f homodimer with a conformational change in one of the RuvC domains activating cleavage.

35. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E: **CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III**. *Nature* 2011, **471**:602–607.

36. Yan WX, Hunnewell P, Alfonse LE, Carte JM, Keston-Smith E, Sothiselvam S, Garrity AJ, Chong S, Makarova KS, Koonin E V, et al.: **Functionally diverse type V CRISPR-Cas systems**. *Science* 2019, **363**:88–91.

37. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C: **Short motif sequences determine the targets of the prokaryotic CRISPR defence system**. *Microbiology* 2009, **155**:733–740.

38. Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, Van Der Oost J, Brouns SJJ, Severinov K: **Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence**. *Proc Natl Acad Sci U S A* 2011, **108**:10098–10103.

39. Yang H, Gao P, Rajashankar KR, Patel DJ: **PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease**. *Cell* 2016, **167**:1814-1828.e12.

40. Murugan K, Seetharam AS, Severin AJ, Sashital DG: **CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects.** *J Bol Chem* 2020, **295**:5538–5553.

41. Fuchs RT, Curcuru J, Mabuchi M, Yourik P, Robb GB: **Cas12a trans-cleavage can be modulated in vitro and is active on ssDNA, dsDNA, and RNA**. *bioRxiv* 2019, doi:10.1101/600890.

42. Kaminski MM, Abudayyeh OO, Gootenberg JS, Zhang F, Collins JJ: **CRISPR-based**

diagnostics. *Nat Biomed Eng* 2021, **5**:643–656.

43. Marino ND, Pinilla-Redondo R, Bondy-Denomy J: **CRISPR-Cas12a targeting of ssDNA plays no detectable role in immunity**. *bioRxiv* 2022, doi:10.1101/2022.03.10.483831.

44. Xu X, Chemparathy A, Zeng L, Kempton HR, Shang S, Nakamura M, Qi LS: **Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing**. *Mol Cell* 2021, **81**:4333-4345.e4.
• The authors develop a CasMINI system to engineer compact and efficient Cas effectors that are capable of gene activation, base editing, and gene editing. They engineer Cas12f with potential use for gene regulation and gene editing.

45. Kim DY, Lee JM, Moon S Bin, Chin HJ, Park S, Lim Y, Kim D, Koo T, Ko JH, Kim YS: **Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus**. *Nat Biotechnol* 2022, **40**:94–102.
• The authors modified natural guide RNA of Un1Cas12f1 to improve the cleavage activity in eukaryotic cells transforming the CRISPR–Cas12f system into a highly efficient and specific genome-editing tool via adenovirus delivery.

46. Wu Z, Yang H, Colosi P: **Effect of genome size on AAV vector packaging.** *Mol Ther* 2010, **18**:80–86.

47. Wu Z, Zhang Y, Yu H, Pan D, Wang Y, Wang Y, Li F, Liu C, Nan H, Chen W, et al.: **Programmed genome editing by a miniature CRISPR-Cas12f nuclease**. *Nat Chem Biol* 2021, **17**:1132–1138.
• In this study, the authors use biochemistry to characterize properties of AsCas12f1 and demonstrate its ability to be used for genome editing tool in both bacteria and human cells with various delivery methods.

48. Wang Y, Sang S, Zhang X, Tao H, Guan Q, Liu C: **Efficient Genome Editing by a Miniature CRISPR-AsCas12f1 Nuclease in Bacillus anthracis**. *Front Bioeng Biotechnol* 2022, **9**.

Figures

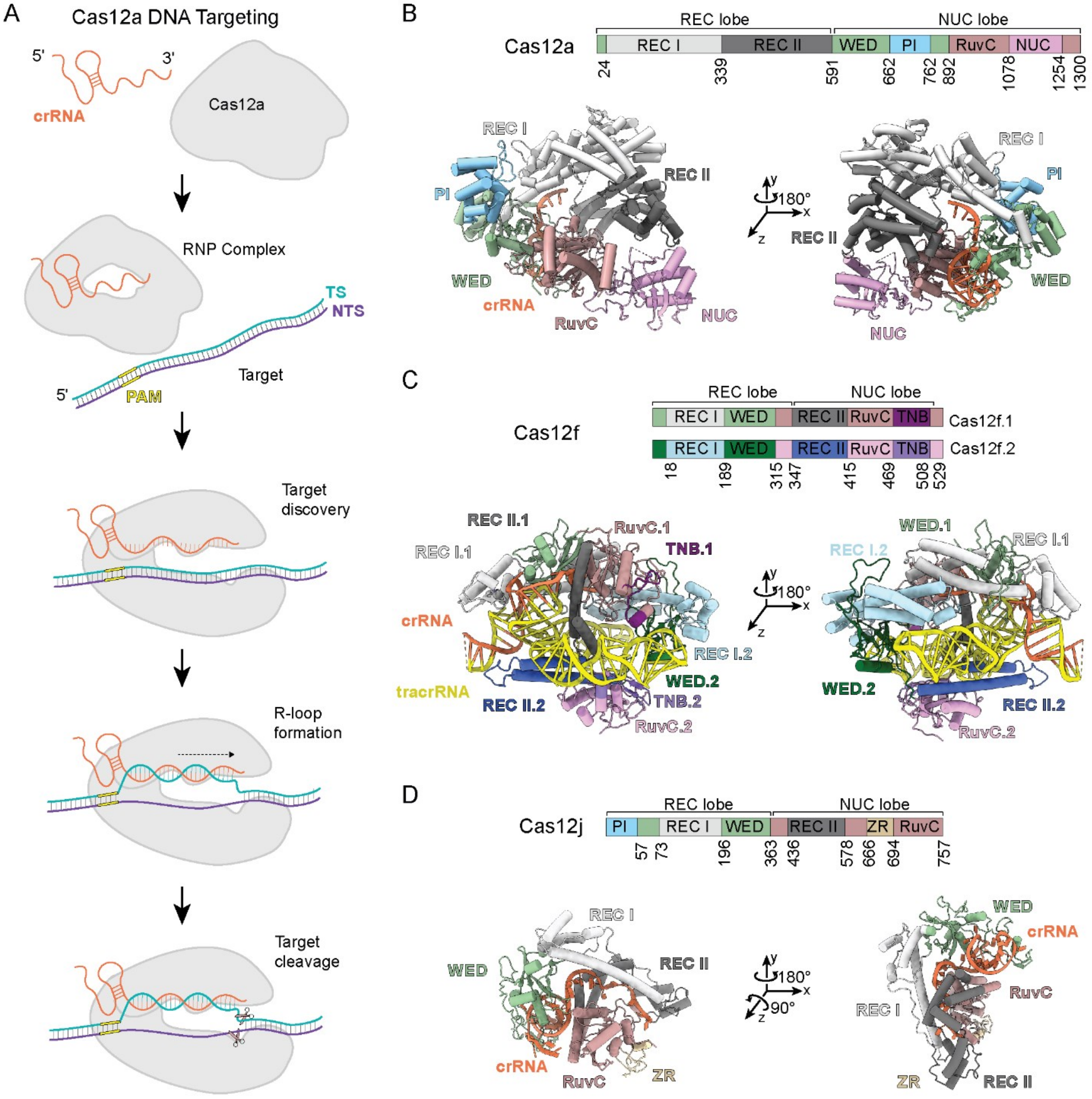

Figure 1. Mechanism and architecture of Cas12 proteins. (A) Outline of major stages of Cas12a mediated target cleavage. Cas12a forms a ribonucleoprotein complex with the crRNA. Cas12a-RNA searches for PAM sequences in the foreign DNA. Upon PAM recognition, the target and non-target strands are unwound and an R-loop is formed simultaneously. The complex binds to the target on the basis of complementarity to the crRNA. The RuvC active site of Cas12a then cleaves the non-target strand, followed by the target strand, generating a staggered double stranded

break. (B, C, D) Domain organization and architecture of Cas12a from the bacterium *Francisella novicida*, Cas12f1 from an uncultured archaeon and Cas12j from a huge phage bound to their respective crRNAs. Domains comprise of Recognition (REC), PAM interaction (PI), Wedge (WED), Nuclease (Nuc), ZR (zinc finger), TNB (target nucleic acid binding). PDB IDs: 5ng6 (Cas12a-RNA); 7l48 (Cas12f-RNA); 7m5o (Cas12j-RNA).

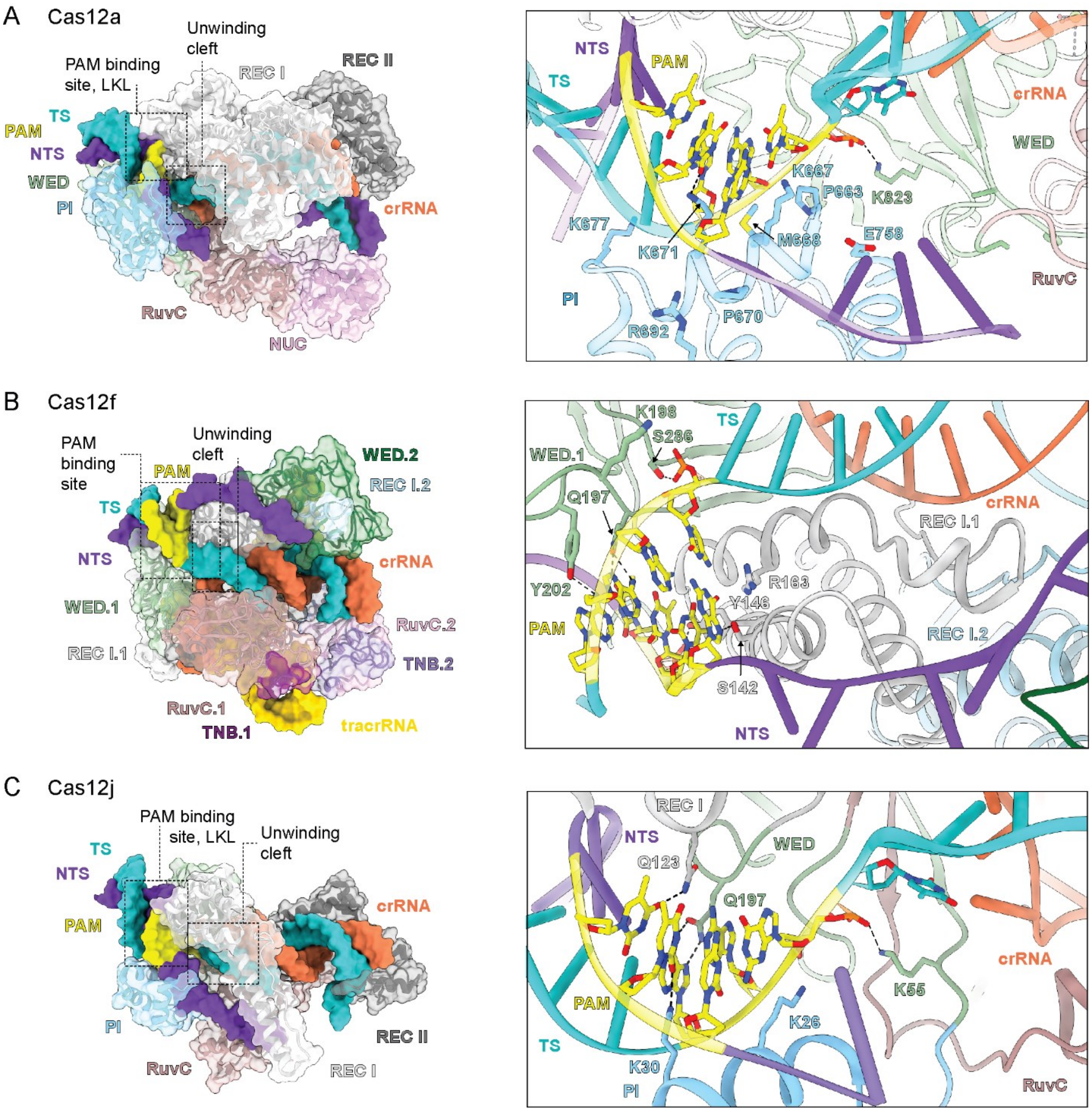

Figure 2. PAM recognition and initiation of DNA unwinding by Cas12 proteins. (A) Surface view of Cas12a bound to a DNA target and close up view of Cas12a recognition of a 5’-TTTN-3’ PAM. PDB ID: 5fnv (Cas12a-RNA-DNA). PAM recognition is mediated by PI and WED domains. The LKL motif of PI domain specifies the PAM by a combination of direct nucleotide recognition and shape readout. A lysine residue from the WED domain forms a hydrogen bond with the phosphate group of the last nucleotide of the PAM, also referred to as a phosphate lock, that helps stabilize the formation of the RNA-DNA hybrid at the unwinding initiation point. (B)

Surface view of Cas12f bound to a DNA target and close up view of Cas12f recognition of a 5’-TTTN-3’ PAM. PDB ID: 7c7l (Cas12f-RNA-DNA). Note that PAM recognition by Cas12f is mediated by WED and RECI domains of one subunit, 12f.1. Cas12f uses a similar phosphate lock to Cas12a. (C) Surface view of Cas12j bound to a DNA target. Close up view of Cas12j recognition of a 5’-TTN-3’ PAM. PDB ID: 7lys (Cas12j-RNA-DNA). PAM recognition is mediated by PI and WED domains, similar to Cas12a. In all close-ups, notable residues are labeled with single-letter amino acid code. Domains and guide RNA are colored as in Figure 1, target strand of DNA is in teal, non-target strand is in purple.

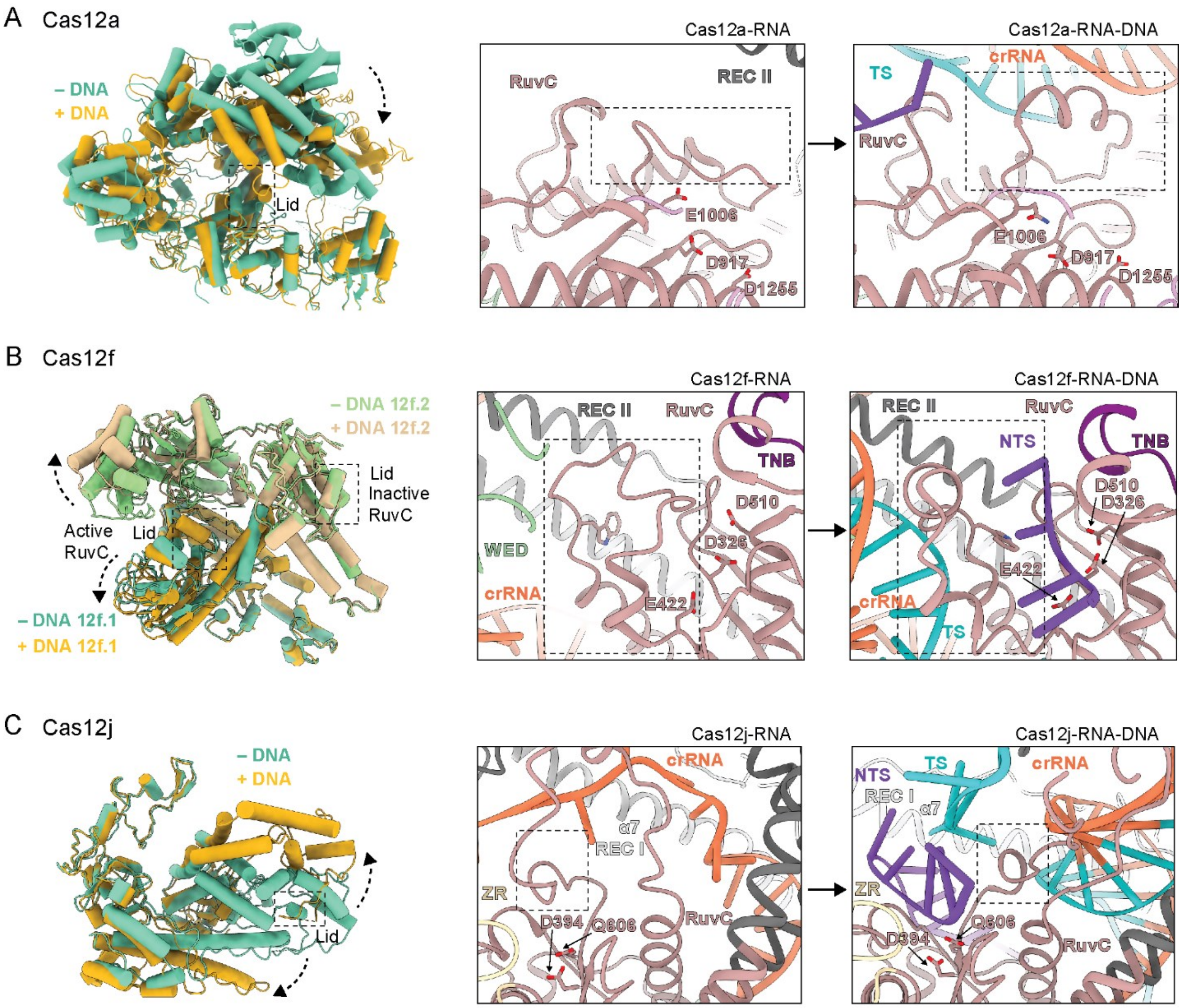

Figure 3. Conformational changes in DNA bound state activate Cas12 cleavage activity. (Left panels A, B, C) Overlays of RNA bound and RNA-DNA bound Cas12a, Cas12f and Cas12j with regions undergoing a conformation changes outlined in boxes. The nucleic acids were omitted from the structures for clarity. Arrows for Cas12a (A) and Cas12j (C) indicate larger movements in the protein structure, mainly occurring in the REC and RuvC domains. Note that only one of the subunits in Cas12f (12f.1 in B) undergoes a conformational change. There is no movement in RuvC for 12f.2. (Right panels A, B, C) Close up view of lid motifs in RNA bound and RNA-DNA bound states, colored as in Figures 1 and 2. Active site residues have been shown and labeled with single amino acid letter codes. Lid motif in RuvC domains of all three proteins transition to a helical state upon DNA binding. Cas12a lid motif is residues 1008-1021, Cas12f lid motif is residues 424-442 and Cas12j lid motif is residues 610-616. PDB IDs: 5nfv (Cas12a-RNA), 6gtg

(Cas12a-RNA-DNA); 7l48 (Cas12f-RNA), 7l49 (Cas12f-RNA-DNA); 7m5o (Cas12j-RNA), 7odf (Cas12j-RNA-DNA)

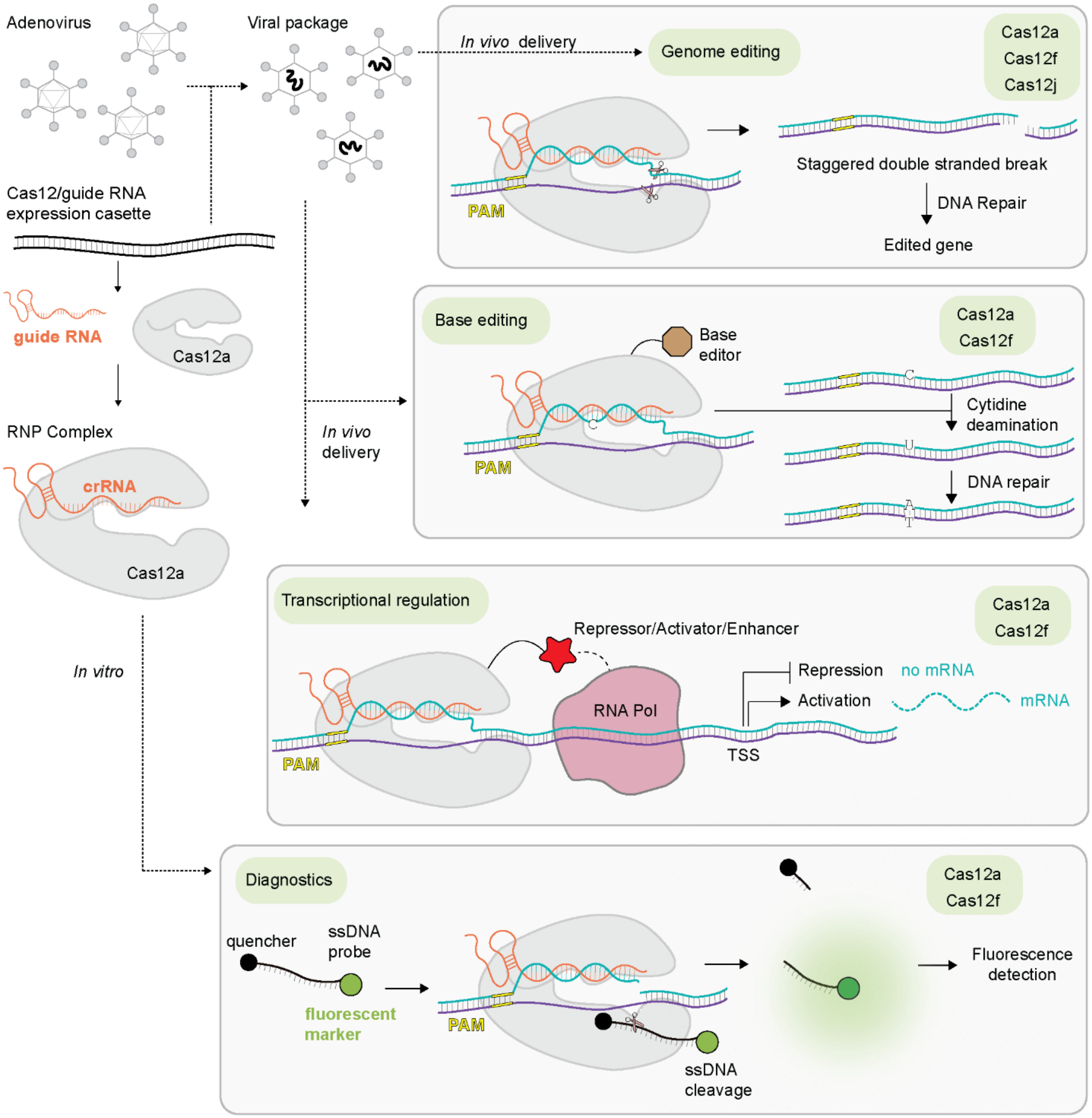

Figure 4. Applications of Cas12 proteins. Adenoviruses are commonly used for packaging and *in vivo* delivery of a guide RNA and Cas12 encoding DNA cassette. In vivo delivery of Cas12 and its guide RNA can be used for genome editing or transcriptional regulation. The capability of Cas12a, Cas12f and Cas12j has been indicated as green boxes associated with the techniques described. Cas12j has recently been shown to be capable of genome editing and its capability for use in other applications remains to be tested. For genome editing, the guide RNA is

complementary to the target and once the RNP bind to target sequence, cleavage occurs generating a dsDNA break which is then repaired to result into corrected or desired sequence. To correct a mismatched base, a base editor like deaminase can be attached to Cas12. The base editor can then modify the base needed in the target sequence followed by DNA repair. For transcriptional regulation, a gene for the modulator is added to the cassette with the guide RNA and Cas12. Binding of the RNP to a target upstream of the RNA polymerase promoter will place the modulator close to the RNA polymerase for activation or repression to occur. Cas12 bound to RNA or the RNP (ribonucleoprotein) complex is directly used for *in vitro* applications. For diagnostic assays like detection of nucleic acids, target recognition/cleavage is followed by cleavage of the ssDNA probe provided with the RNP complex. A fluorophore and quencher are attached to the two ends of the ssDNA respectively. ssDNA cleavage will release the fluorophore from its quencher and fluorescence can be detected as an indication of presence of the target in the given sample.