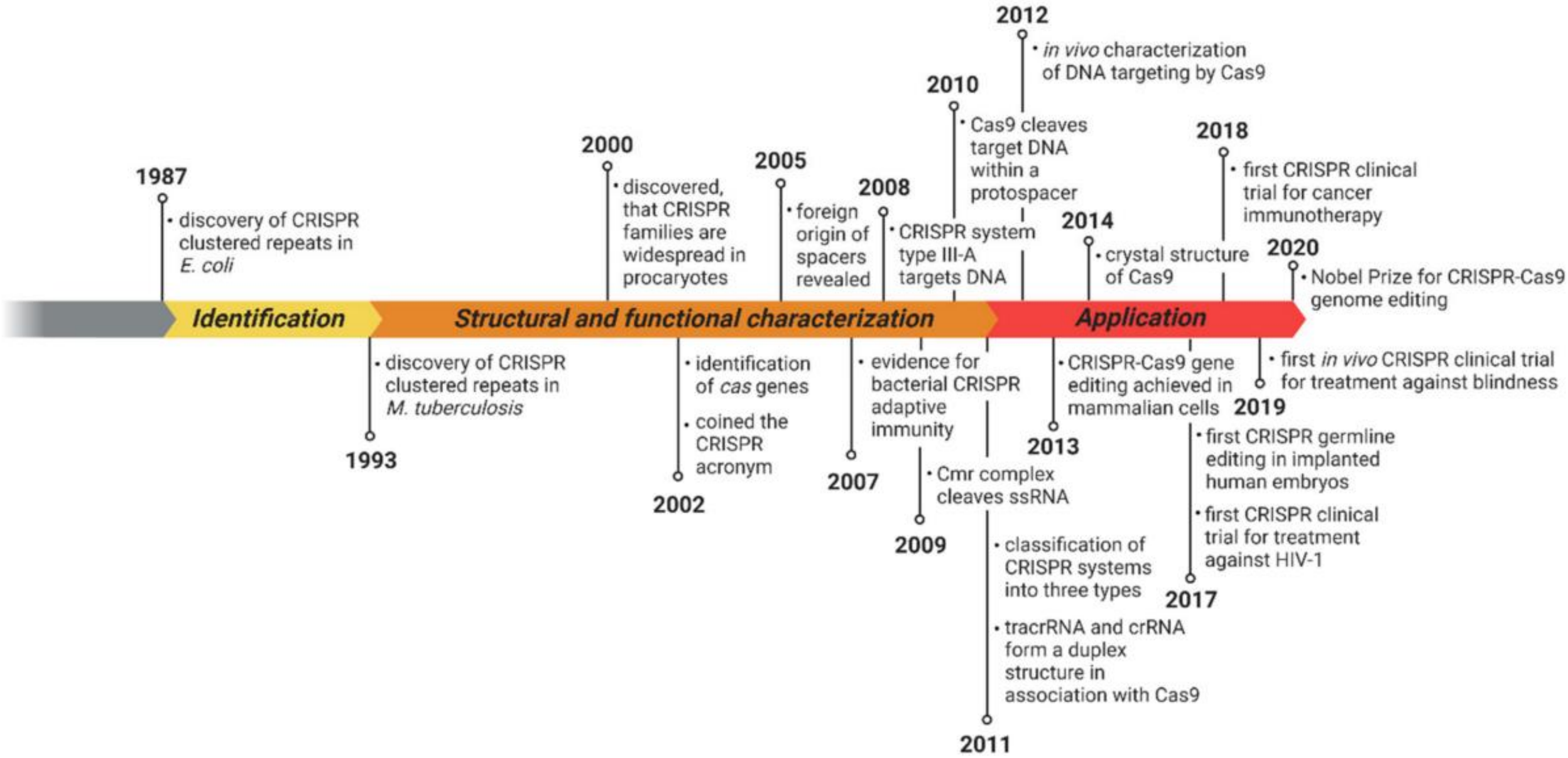

1987
• discovery of CRISPR clustered repeats in *E. coli*
Identification
1993
• discovery of CRISPR clustered repeats in *M. tuberculosis*
2000
• discovered, that CRISPR families are widespread in procaryotes
2002
• identification of *cas* genes
• coined the CRISPR acronym
2005
• foreign origin of spacers revealed
2007
• evidence for bacterial CRISPR adaptive immunity
2008
• CRISPR system type III-A targets DNA
2009
• Cmr complex cleaves ssRNA
2010
• Cas9 cleaves target DNA within a protospacer
2011
• classification of CRISPR systems into three types
• tracrRNA and crRNA form a duplex structure in association with Cas9
2012
• *in vivo* characterization of DNA targeting by Cas9
2013
• CRISPR-Cas9 gene editing achieved in mammalian cells
2014
• crystal structure of Cas9
2017
• first CRISPR germline editing in implanted human embryos
• first CRISPR clinical trial for treatment against HIV-1
2018
• first CRISPR clinical trial for cancer immunotherapy
2019
• first *in vivo* CRISPR clinical trial for treatment against blindness
2020
• Nobel Prize for CRISPR-Cas9 genome editing
Structural and functional characterization
Application

Table 1. Selected Cas proteins and their functions. Types (Roman numerals) are colored black, and subtypes (Roman numerals with a letter) are colored gray [14,21].

Protein	Association in Type or Subtype	Function
Cas1	I, II, IV, IV (assumed) III-A, III-B	DNA nuclease
Cas2	I, II, V III-A, III-B, VI (some)	RNA nuclease
Cas3	I	DNA nuclease and helicase
Cas4	II, V I (most)	DNA nuclease
Cas5	IV I-C, III (some)	pre-crRNA processing
Cas6	I (most), III-A, III-B	pre-crRNA processing
Cas7	I, III, IV	RNA recognition, crRNA binding
Cas8	I (most)	large subunit of Cascade complex
Cas9	II	DNA nuclease
Cas10	I (some), III (most)	large subunit of Csm or Cmr complex
Cas11	III I (some), IV (some)	small subunit of effector complexes
Cas12	V	crRNA processing, DNA nuclease
Cas13	VI	crRNA processing, RNA nuclease

Table 2. Cas protein association with CRISPR functional steps in type I–VI systems. An asterisk represents that protein being potentially fused to a large subunit in some subtypes. Underlined proteins are present in multiple copies. Proteins colored in gray are dispensable or missing in some subtypes or variants. Abbreviations: LS, large subunit; RT, reverse transcriptase; SS, small subunit; ?, unknown [14].

		Adaptation	Expression	Interference	
		Spacer Integration	**pre-crRNA Processing**	**Effector Complex**	**Target Cleavage**
class 1	type I	Cas1, Cas2, Cas4	Cas6	Cas7, Cas5, SS *, Cas8/LS	Cas3″, Cas3′
	type III	Cas1, Cas2, RT	Cas6	Cas7, Cas5, SS, Cas10/LS	Cas10/LS
	type IV	Cas1, Cas2	Cas6	Cas7, Cas5, SS, Csf1/LS	?
class 2	type II	Cas1, Cas2, Cas4	RNase III	Cas9	Cas9
	type V	Cas1, Cas2, Cas4	Cas12	Cas12	Cas12
	type VI	Cas1, Cas2	Cas13	Cas13	Cas13

Table 3. Classification of class 1 CRISPR–Cas systems [14].

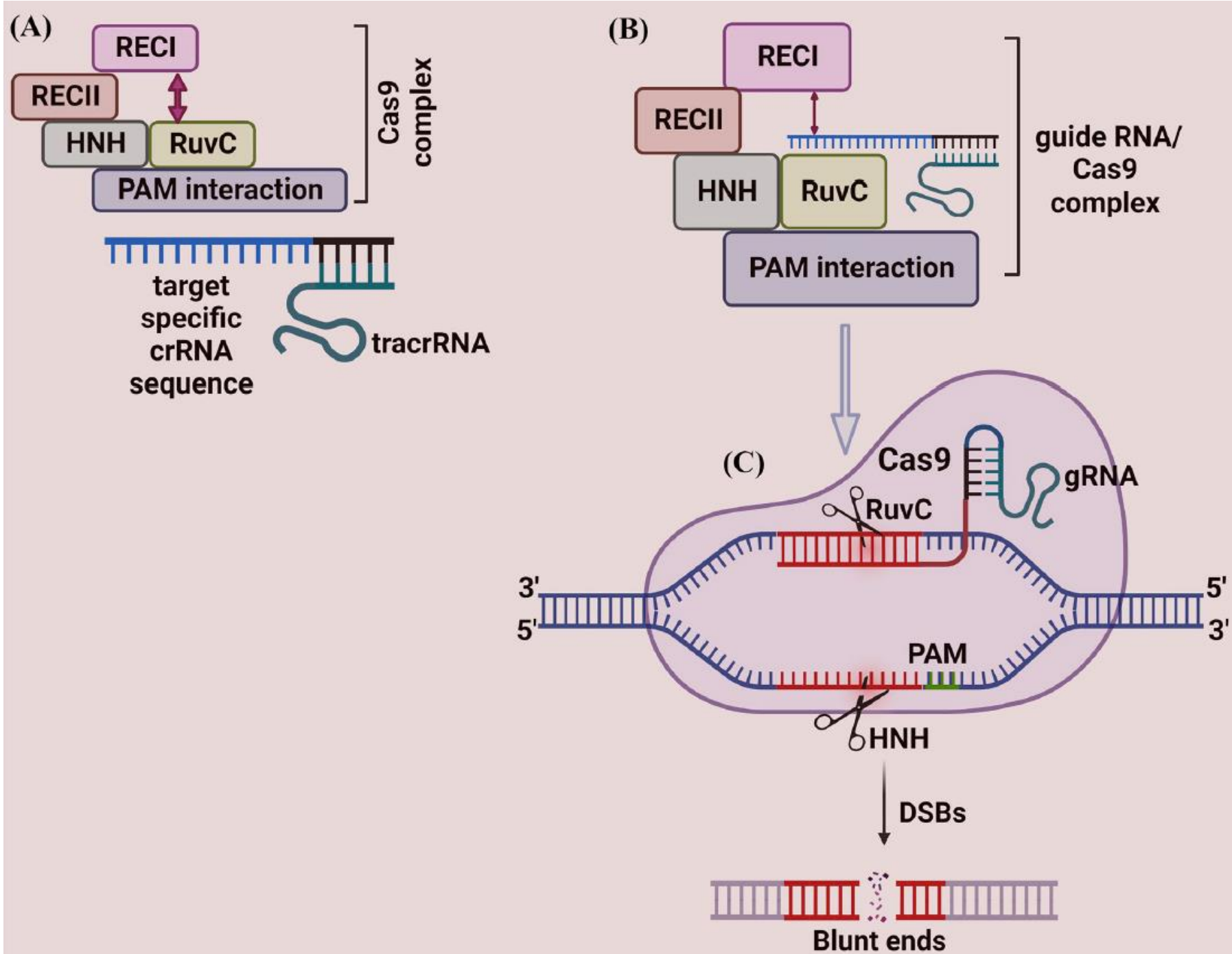
Class	Type	Subtype	Variant	Native Target	Origin
1	III	III-A		DNA + RNA	*Staphylococcus epidermidis* (SERP2463–SERP2455)
		III-B		DNA + RNA	*Pyrococcus furiosus* (PF1131–PF1124)
		III-C		DNA + RNA	*Methanothermobacter thermautotrophicus* (MTH328–MTH323)
		III-D		RNA?	*Synechocystis sp. 6803* (sll7067–sll7063)
		III-E		RNA?	*Candidatus Scalidua brodae* (SCABRO_02601, SCABRO_02597, SCABRO_02593, SCABRO_02595)
		III-F		DNA?	*Thermotoga lettingae TMO* (Tlet_0097–Tlet_0100)
	IV	IV-A			*Thioalkalivibrio sp. K90mix* (TK90_2699–TK90_2703)
		IV-B			*Rhodococcus jostii RHA1* (RHA1_ro10069–RHA_ro10072)
		IV-C		DNA?	*Thermoflexia bacterium* (D6793_05715–D6793_05700)

Table 4. Classification of class 2 CRISPR–Cas systems. An asterisk represents a variant that was formerly classified as variant I-U3 [14].

Class	Type	Subtype	Variant	Native Target	Origin
2	V	V-F	V-F1	DNA	*Uncultured archaeon* (NDOCEIEL_00008–NDOCEIEL_00011)
			V-F1*	DNA	*Bacillus thuringiensis HD-771* (BTG_31928)
			V-F2	DNA	*Uncultured archaeon* (ICDLJNLD_00049–ICDLJNLD_00052)
			V-F3		*Candidatus Micrarchaeota archaeon* (COU37_03050–COU37_03065)
			V-U1		*Gordonia otitidis* (GOOTI_RS19525)
			V-U2		*Cyanothece sp. PCC 8801* (PCC8801_4127)
			V-U4		*Rothia dentrocariosa M567* (HMPREF0734_01291)
		V-G		RNA	*Hot springs metagenome* FLYL01000025.1 (182949–185252)
		V-H			*Hypersaline lake sediment metagenome (JGI)* (Ga0180438_100006283)
		V-I		DNA	*Freshwater metagenome (JGI)* (Ga0208225_100001036)
		V-K			*Cyanothece sp. PCC 8801* (PCC8801_2993–PCC8801_2997)
	VI	VI-A		RNA	*Leptotrichia shahii* (B031_RS0110445)
		VI-B	VI-B1	RNA	*Prevotella buccae* (HMPREF6485_RS00335–HMPREF6483_RS00340)
			VI-B2	RNA	*Bergeyella zoohelcum* (HMPREF9699_02005–HMPREF9699_02006)
		VI-C		RNA?	*Fusobacterium perfoetens* (T364_RS0105110)
		VI-D		RNA	*Ruminococcus bicirculans* (RBI_RS12820)

Cas9 Protein

Streptococcus pyogenes



Protein name	Host organism	sgRNA sequence size	PAM sequence	Target	Cut site
High–Fidelity Cas9	**HypaCas9**				
eSpCas9	**FokI–Fused dCas9**				
xCas9 and SpRY/SpG: Altered PAM Recognition					
Cas9	*S. pyogenes*	20	5′-NGG-3′	dsDNA	5′ of PAM
Cas9	*S. pyogenes*	–	5′-NAC, NTG, NTT, and NCG-3′	DNA	5′ of PAM
Cas9	*F. novicida*	20	5′-NGG-3′	DNA	5′ of PAM
Cas9	*S. aureus*	21	5′-NNGRRT-3′	DNA	5′ of PAM
Cas9	*Neisseria meningitidis*	24	5′-NNNNGATT-3′	DNA	5′ of PAM
Cas9	*S. thermophilus*	20	5′-NNAGAAW 5′	DNA	5′ of PAM
Cas9	*S. thermophilus*	20	5′-NGGNG-3	DNA	5′ of PAM
Cas9	*Campylobacter jejuna*	22	NNNNACAC and NNNRYAC	DNA	5′ of PAM
C2c1	*Alicyclobacillus acidoterrestris*	20	T-rich PAM	DNA	5′ of PAM
Cpf1	*Prevotella and Francisella*	20	TTTV	DNA	5′ of PAM
Cpf1	*Acidaminococcus sp.*	24	5′-TTTN-3′	DNA	3′ of PAM
Cas12a	*Acidaminococcus sp.*	–	Thymine-rich PAM	DNA	5′ of PAM
Cas13	*Lb*	28	Non-G nucleotide at the 3′ protospacer flanking site (PFS)	ssRNA	–
Cas14	*Uncultivated archaea*	–	–	ssDNA	–

Pros and Cons of Cas9 Protein

The advantage of the CRISPR/Cas9 system is the design simplicity with more efficiency than existing ZFN and TALEN systems [8]. Multiplexed genome editing is another significant advantage of Cas9, which can be achieved by designing multiple sequence-specific gRNAs simultaneously [8].

gRNA, which guides the Cas9 to cleave dsDNA or ssDNA showed higher on- and off-target mutations in the targeted organisms [9, 55]. Even the finest accessible CRISPR/Cas system that uses HDR also induces undesirable mutations. However, these off-target effects can be reduced by utilizing the modified Cas9 version called null Cas9 (nCas9), which generates a nick in only one strand than the DSBs [55]. However, 100% off-targets cannot be reduced using nCas9, which requires upgraded Cas9 versions in the future.

Cas12 Protein

Acidaminococcus species (*As*Cas12a)

Lachnospiraceae bacterium (LaCas12a)

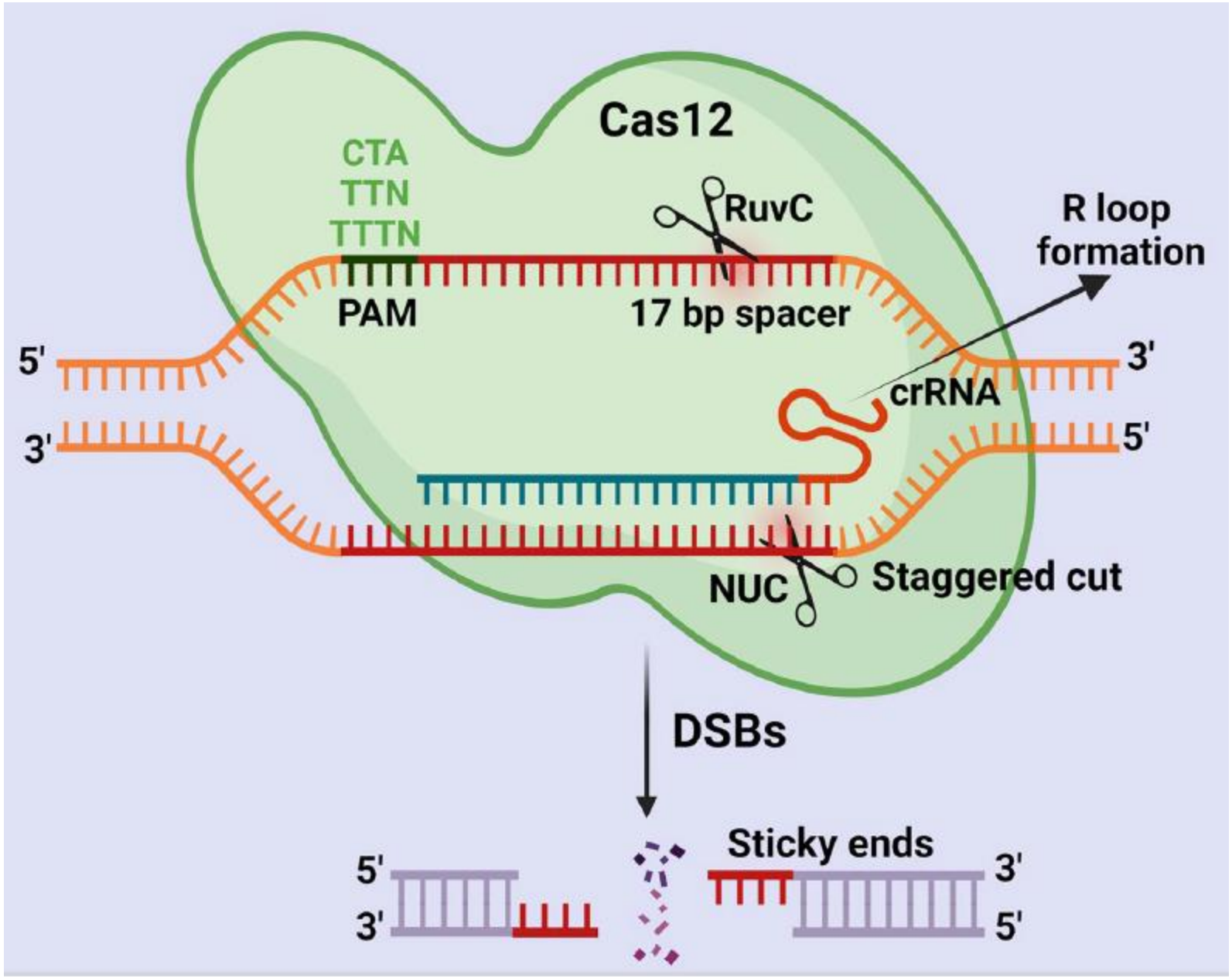


Fig. 4 Schematic illustration of CRISPR/Cas12 mechanism. The Cas12 protein requires only the crRNAs to generate DSBs. Cas12 protein cleaves the target region beside a PAM sequence (CTA, TTN, TTTN) with the help of the RuvC and nuclease lobe (NUC) domains. Once Cas12 starts encountering, it initiates R-loop, which forms base-pair hybridization between the crRNA and the target DNA strand. During this step, Cas12 matches the < 17 bp of the target sequence and leads to an R-loop formation. Once R-loop is formed, the Cas12 protein uses its active RuvC domain and generates a staggering cut in the non-target strand with the help of the PAM sequence

CRISPR/Cas9 vs CRISPR/Cpf1

CRISPR:	**CRISPR/Cas9**	**CRISPR/Cpf1**
Size	Big in size (~4.1 kb)	Small in size (~3.8 kb)
Simplicity	Complex	Simpler
Guide RNA	Cas9 requires crRNA and tracrRNA (~100 nt)	Cpf1 requires only crRNA (~42 nt)
Domains	Cas9 contains two nuclease domains: a HNH nuclease domain and a RuvC nuclease domain	Cpf1 contains only one nuclease domain: a RuvC nuclease domain
PAM	5'-NGG-3' (where "N" is any nucleobase). G-rich PAM	5'-TTTV-3' (AsCpf1; where "Y" is a pyrimidine (C/T) and "V" is A/G/C nucleobase). T-rich PAM
Blunt/Sticky	Cuts both strands in a DNA molecule at the same position, leaving behind blunt ends (blunt end cuts)	Cuts with a sticky-end DNA double-stranded break (staggered end cuts) of 4 or 5 nucleotides overhang, enhancing the efficiency of genetic insertions and specificity during DNA repair
Cutting site	Cuts target DNA closer to PAM site (proximal to recognition site). Generates indels that destroy the PAM site and prevents further rounds of cutting.	Cuts target DNA further away (18-23 nt) from PAM site (distal from recognition site). No disruption of PAM site and allows multiple rounds of cutting.

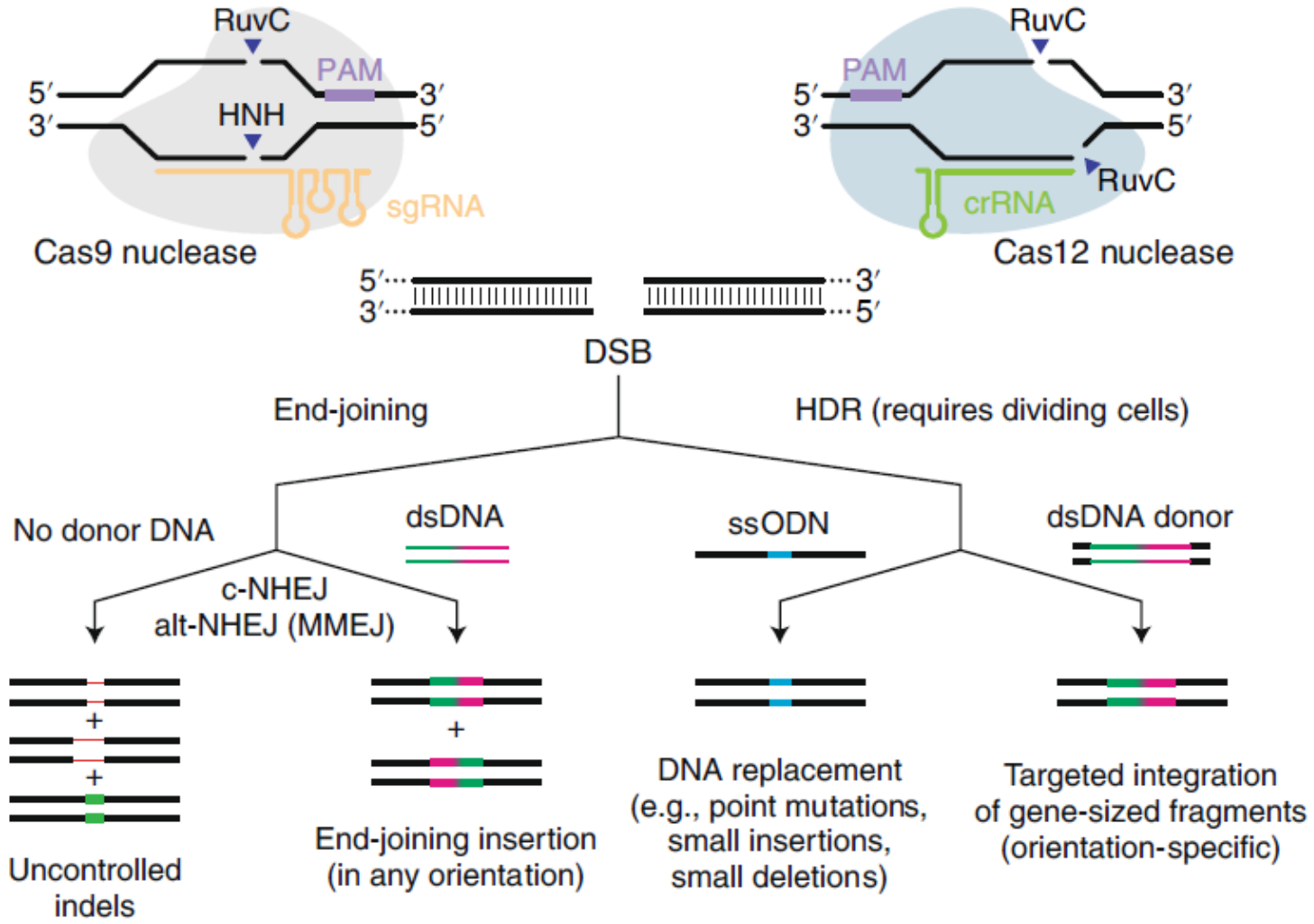


Fig. 2 | Genome editing with Cas nucleases. Cas9 nucleases are guided by guide RNAs to generate predominantly blunt-end DSBs using two distinct nuclease domains (RuvC and HNH). The DSBs occur within the protospacer, typically preceding three base pairs upstream of the PAM. Cas12 nucleases are guided by crRNAs to cleave both strands of DNA with a single RuvC-like nuclease domain. Cas12 cuts DNA in a staggered orientation within PAM-distal regions of the protospacer. Genome editing with CRISPR–Cas nucleases results from two major arms of DNA repair. End-joining mechanisms, such as classical nonhomologous end-joining (c-NHEJ) and microhomology-mediated end-joining (MMEJ, or alt-NHEJ), result in uncontrolled, but predictable, indels for gene disruption. In the presence of a donor DNA template, homology directed repair (HDR) is a competing (typically less efficient) pathway that occurs mostly in dividing cells and is used to install targeted mutations or to knock in larger DNA sequences. Insertion sizes with single-stranded oligonucleotide donors (ssODNs) are typically limited by synthesis capabilities[118].

Pros and Cons of Cas12 Protein

Like Cas9, Cas12 protein was also considered as a sole member of the CRISPR family for genome editing. But in most circumstances, Cas12 was deemed superior to Cas9 protein because Cas12 protein generates staggered DSBs and promotes HDR repair mechanism instead of both NHEJ & HDR [67]. Cas12 system also overcomes disadvantages associated with diagnostic strategies. For example, diagnosing SARS-COV-2 utilizing quantitative qRT-PCR demands a longer time to get the results [66]. But the CRISPR/Cas12-based DETECTR detects the SARS-CoV-2 within 1 h [66]. These results proved the advantage of the CRISPR/Cas12 system, which can also be utilized in detecting newly emerging viruses in the future.

The rapid advancement of the CRISPR-Cas12 system for genome editing has proved revolutionary for the life sciences. Despite this technology's wide application areas, the CRISPR/Cas12 system faces several significant flaws. For example, the CRISPR/Cas12 system is dependent on host cell DNA repair machinery with or without the presence of a template [68]. Although this system has been successfully used to acquire accurate DNA insertion into the desired genomic loci, its effectiveness varies depending on the cell type. DNA repair through HDR is also associated with active cell division, making these tools ineffective in cell division (for example, neurons) [69]. However, significant continuing research aims to tailor the Cas12 system further to ensure accurate DNA insertion into the targeted genome. Apart from this drawback, this system has a wide range of applicability, and ongoing endeavors are striving to generate enhanced CRISPR/Cas12 for robust genome engineering.

Cas13 Protein

Leptotrichia shahii (*Lsh*Cas13a),

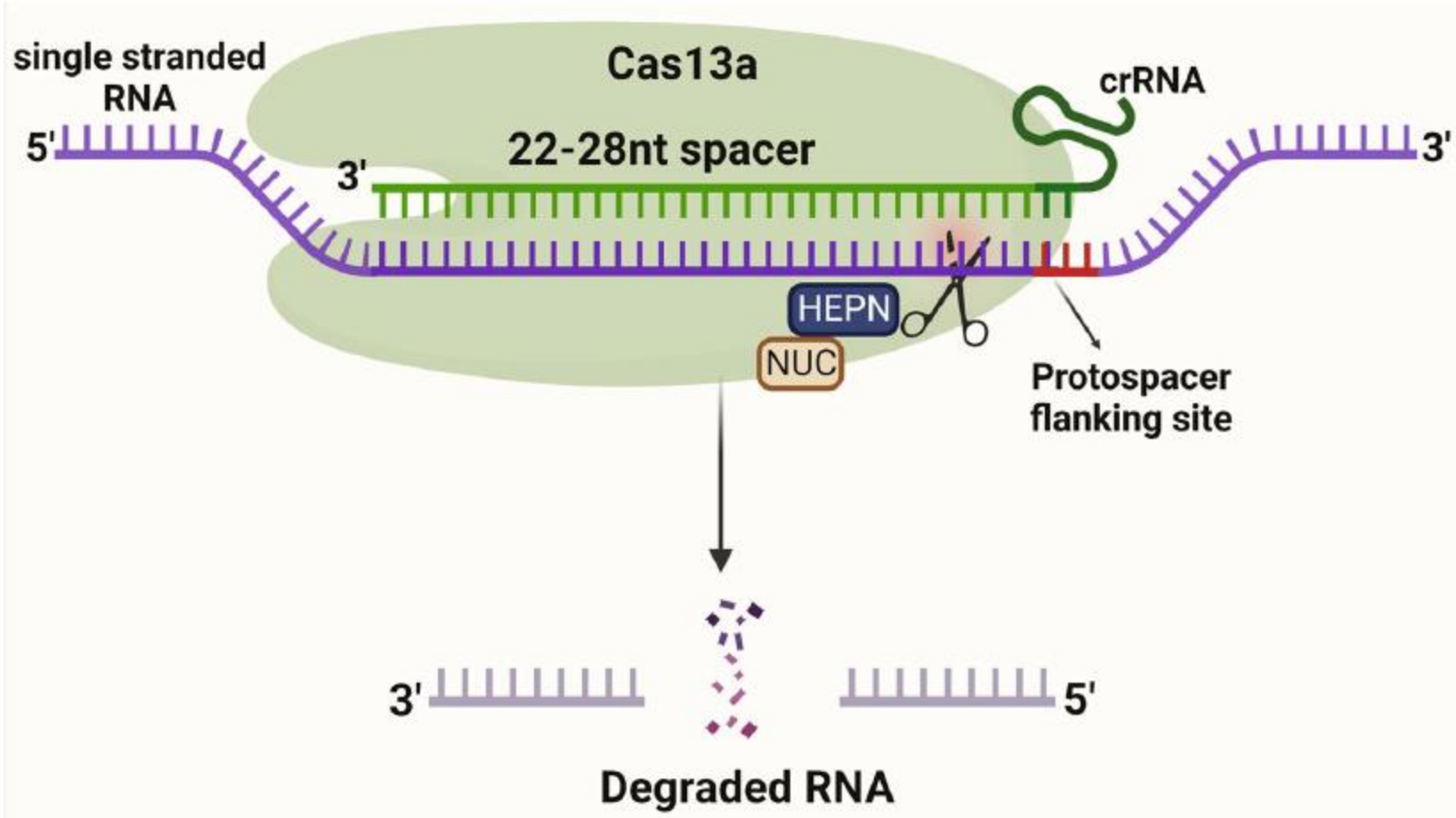
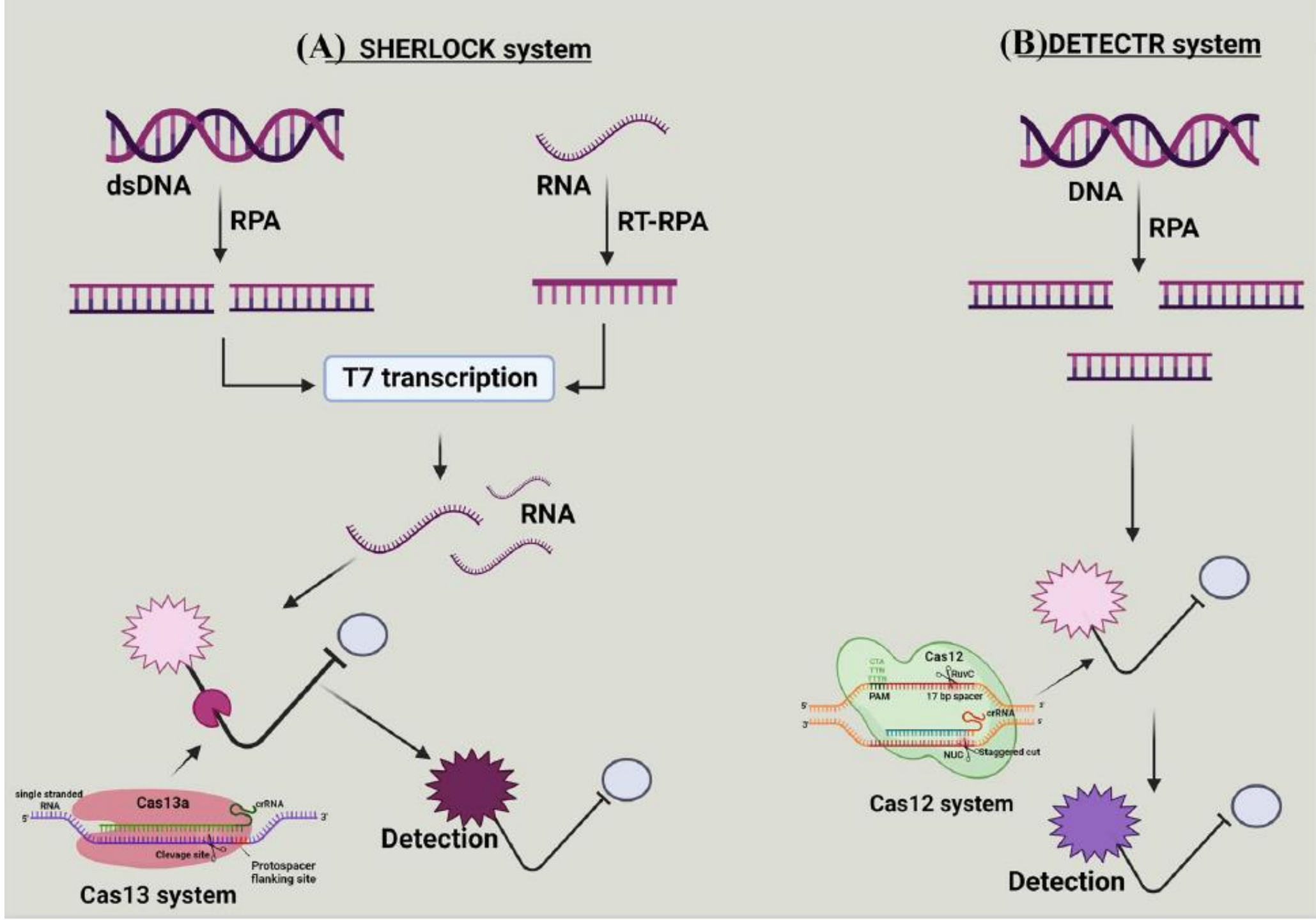


Fig. 6 Schematic illustration of CRISPR/Cas13a mechanism. Cas13a protein is activated through a single crRNA. Cas13a protein comprises crRNA, NUC lobes, and two nucleotide-binding (HEPN) RNase domains for targeting RNA. The Cas13a cleaves ssRNA, upon recognizing the target sequence (22–28 nt) complementary to the crRNA spacer. The target sequence is flanked by a protospacer-flanking site (PFS) at the 3′-end and crRNA binds together and cleaves the target region of ssRNA without the tracrRNA


(A) SHERLOCK system
dsDNA
RPA
RNA
RT-RPA
T7 transcription
RNA
single stranded RNA
Cas13a
crRNA
Cleavage site
Protospacer flanking site
Cas13 system
Detection
(B)DETECTR system
DNA
RPA
Cas12
RuvC
PAM
17 bp spacer
crRNA
NUC
Staggered cut
Cas12 system
Detection

Cas13 Protein

Prevotella sp. (*Psp*Cas13b)

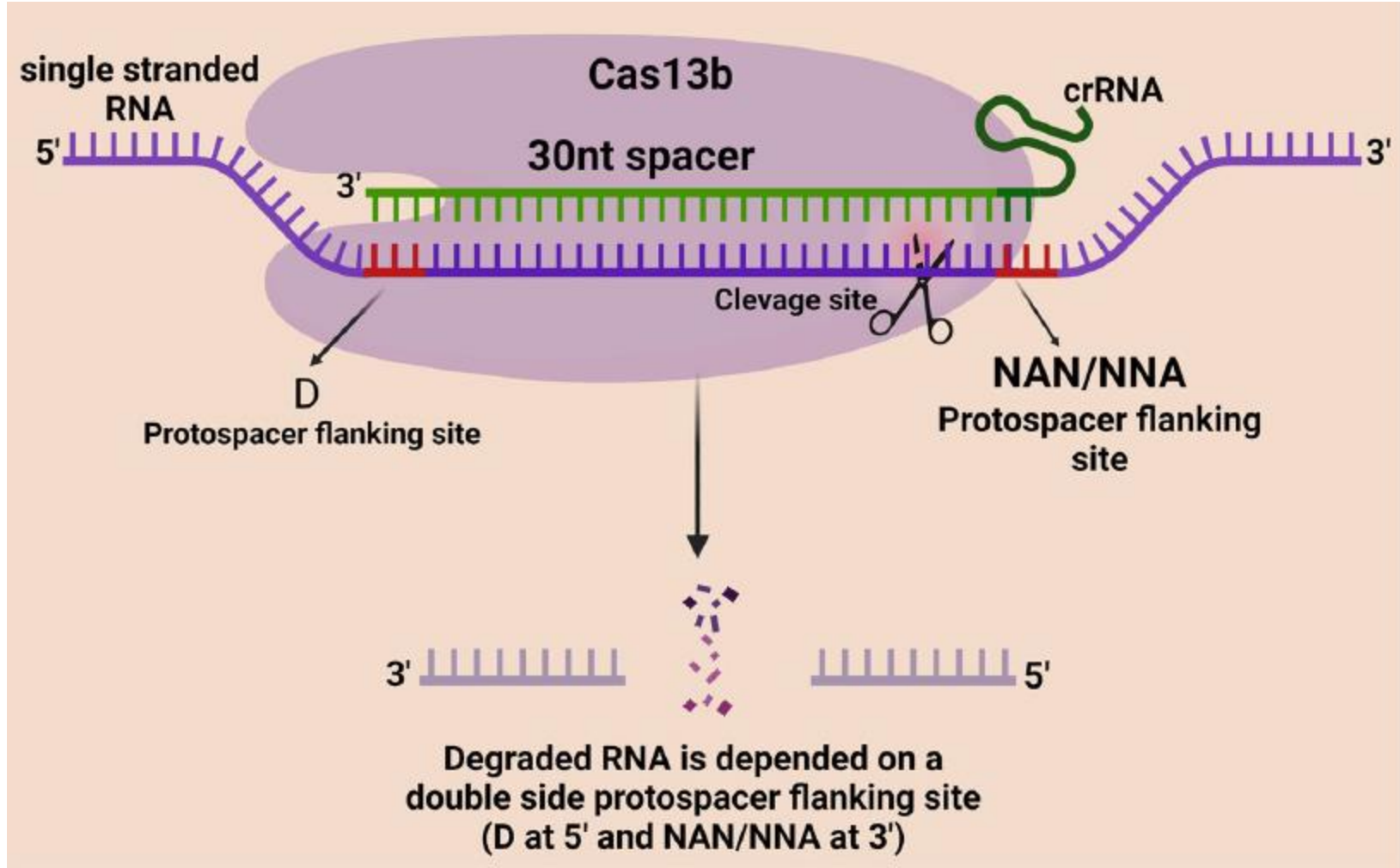
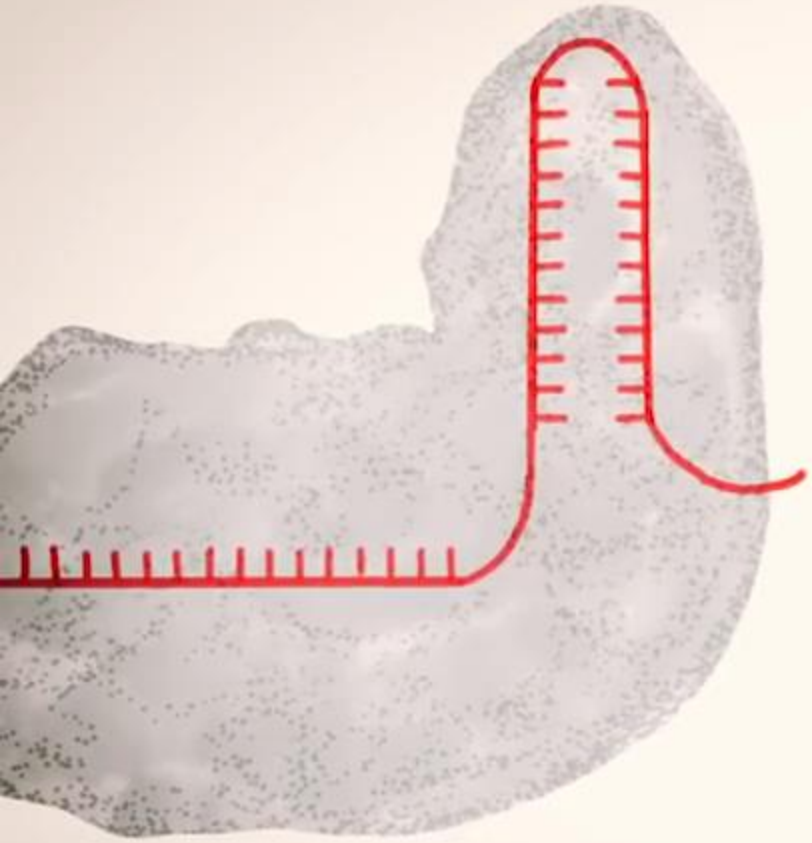


Fig. 7 Schematic illustration of CRISPR/Cas13b mechanism. Cas13b protein is associated with the mature crRNA. This ÇRISPR/Cas13b complex searches for the target ssRNA and induces precise conformational changes at the ssRNA target with the help of the Protospacer flanking site (PFS), which flanks RNA targeting at the 5′ end and PAM sequence (NAN/NNA) at the 3′ end, resulting in nonspecific RNA cleavage



Gene editing

Disabling genes

Fixing mutations

Pros and Cons of Cas13 Protein

Like Cas9 and Cas12, CRISPR/Cas13 is also a robust, precise, and versatile RNA-targeting system, which opens novel research horizons in diverse fields. Compared with earlier RNA manipulation systems, CRISPR/Cas13 offers numerous advantages. For example, its modular construction, which consists of a single protein effector module and an RNA guide module, allows for significant scalability by enabling the production of whole libraries of various guide RNAs in addition to easy and quick design [83]. The recently discovered Cas13 mutant versions (dCas13, Cas13x), which function as programmable RNA-binding proteins, efficiently target different effectors to specific RNAs to induce specific mutations [83]. Due to the inherent crRNA biogenesis, multiple RNAs can be targeted using Cas13 precisely. Compared with RNAi, CRISPR/Cas13 mediated genome modifications are not limited to targeting cytoplasmic transcripts. In addition, Cas13 enables faster downregulation of gene expression by directly knocking out the cytoplasmic mRNA transcripts [83]. Recently, CRISPR/Cas13-based SHERLOCK and SHERLOCKv2 also played a vital role in developing novel molecular diagnostic tools to detect viruses, including SARS-CoV-2. Apart from these major advantages, CRISPR/Cas13 faced off-target mutations [55], which is the major drawback of this system. However, future research would help overcome this obstacle and aid

Cas14 Protein

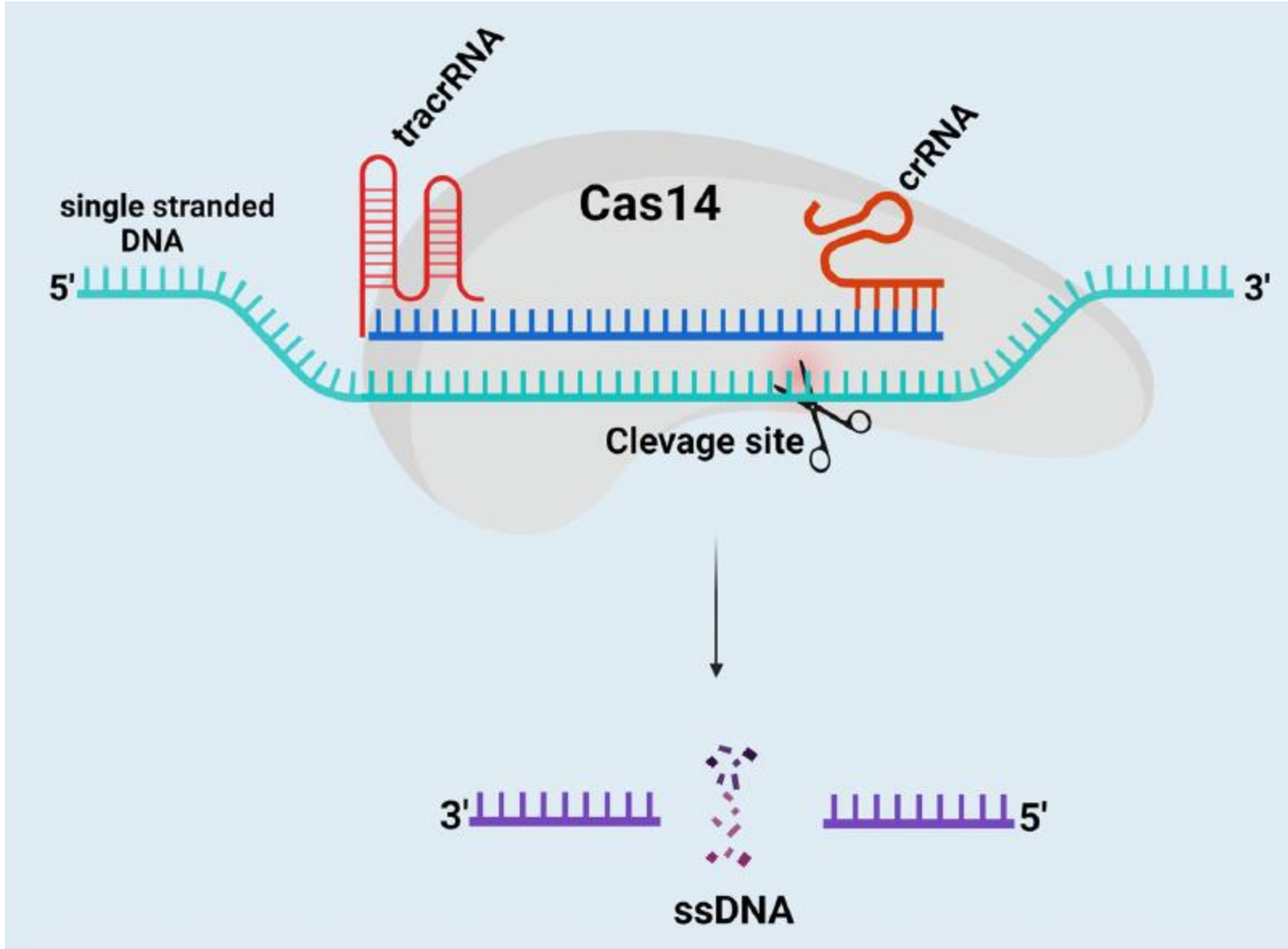


Fig. 8 Schematic illustration of CRISPR/Cas14 mechanism. The Cas14 protein comprises both tracrRNA and crRNA to target ssDNA. Cas14 protein recognizes the ssDNA with the help of tracrRNA and crRNA, mediates seed sequence interaction with the target ssDNA, and cleaves the ssDNA, not dsDNA or ssRNA. The cleavage efficiency of the Cas14 protein is more specific than Cas9, Cas12, and Cas 13 proteins without the presence of the PAM region

Pros and Cons of Cas14 Protein

The Cas14 protein, which edits dsDNA, ssDNA, and RNA, holds various advantages over traditional Cas9 protein. For example, the Cas14 protein that has around 500 amino acid is very small, so this protein could be easily delivered to target any tissue than Cas9 protein [86]. Additionally, the Cas14 protein's selectivity improves the fidelity of single nucleotide polymorphism (SNP) [83]. Finally, the Cas14 protein has less limiting PAM requirements (uses only T-rich sequences) [83], allowing it to edit more targeted genomic sequences than the Cas9 protein. However, only a few studies have been demonstrated, which requires extensive studies to assist Cas14 in several fields of research in the future.