Norwegian University
of Life Sciences

Master's Thesis 2019   30 ECTS
Faculty of Biosciences (BIOVIT – IPV)

Review of CRISPR-assosiated proteins, their functions and patent situations

Nikita Fonarev
Chemistry and Biotechnology, Molecular Biology

Acknowledgements

This master thesis was carried out at the Faculty of Biosciences at the Norwegian University of Life Sciences, with Professor Trine Hvoslef-Eide and PhD Candidate Anders Keim Wulff-Vester as supervisors.

First of all, I would like to thank Trine for the opportunity to write this MSc thesis at Faculty of Biosciences on such a short notice, and for being both enthusiastic and inspiring during the whole process.

Secondly, I would like to thank Anders for the tremendous amount of guidance, good advices and involvement in writing process of the thesis.

Last but not least, big thanks to Hanna Sofie Schou Grytli, Mikkel Schäffer Botnen, Kaveh Nikouee, Oskar Nordberg, Kalle Gundersen, Adrian Sveberg, Marius Magnusson and Rebekka Bedringås for all the support along the way.

Ås, May 2019

Nikita Fonarev

Sammendrag

CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats) er en ny revolusjonerende metode innen genredigering. CRISPR-Cas, opprinnelig oppdaget som et forsvarsmekanisme i bakterier og archaea ble fort et foretrukket verktøy for en rekke bruksområder innen genteknologi, mye takket være enkel design og bruk av programmerbare nukleaseenzymer, både *in vitro* og *in vivo*.

Siden CRISPR nuklease Cas9 er det mest brukte genredigeringsverktøyet innen CRISPR-Cas systemer, er andre CRISPR-Cas proteiner fortsatt i stor grad uutforsket. CRISPR-Cas9 «hype» har inntil nylig satt andre Cas-proteiner i skyggen av komplekset. Det, og andre aspekter som patentsituasjonen rundt Cas9 har brakt forskernes oppmerksomhet til å studere og analysere andre Cas-proteiner, på jakt etter forbedringer og analoger til CRISPR-Cas9.

Det har lenge vært en del forvirring rundt CRISPR-Cas-proteiner over lang tid, i stor grad grunnet mangel av en felles klassifiseringssystem og nomenklatur for CRISPR-Cas systemer. Ikke koordinert forskning har ført til en økning av oppdagete CRISPR-Cas proteiner, men mange av disse var homologe proteiner ført inn under ulike navn. Det har vært flere forsøk på å oppnå et klassifikasjonssystem for Cas-proteiner for å opprettholde dette raskt voksende feltet i genredigeringsverktøy.

Denne masteroppgaven har som formål å sette sammen en enkel oversikt over funksjoner og patentsituasjon rundt kjente CRISPR-Cas proteiner, samt utføre en analyse av Cas systemer for å identifisere mulige alternativer til det mye brukte CRISPR-Cas9 komplekset som kan brukes til genredigering.

Abstract

CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats) systems are a new revolutionary gene editing tool. CRISPR-Cas was originally discovered as a defense mechanism in bacteria and archaea. CRISPR has quickly become a preferred tool for genome editing applications over the course of last few years thanks to the ease of design and use of programmable nuclease enzymes, both *in vivo* and *in vitro*.

Even though CRISPR nuclease Cas9 is the most used gene editing tool in CRISPR-Cas systems, other CRISPR-Cas proteins remain largely unexplored. The CRISPR-Cas9 “hype” has until recently left other Cas proteins in the shadow of the complex. That, and other aspects such as patent situation around Cas9 has brought researchers attention to studying and analyzing of other Cas-proteins in search for improvements and analogs of CRISPR-Cas9.

There has been some confusion around CRISPR-Cas proteins for some time, due to absence of classification and nomenclature system for CRIPR-Cas systems. Non-coordinated researches resulted in a quick growth of discovered CRISPR-Cas proteins, but a number of them were homologues denoted under more than one name. There have been several attempts on achieving a classification system for the Cas-proteins in order to maintain this quickly growing field in genome editing tools.

This thesis aims to make a simple overview of functions and patent situation around known CRISPR-Cas proteins, as well as analyzing alternatives to the widely used CRISPR-Cas9 complex for genome editing purposes.

Table of contents

Introduction

Genome editing is a process of permanent modification at a specific genomic site in a cell. Genome editing experiments can be designed to perform genetic modifications, such as gene insertion, or gene deactivation. Gene insertion leads to adaption of a new gene or a set of genes in the target cell genome, which will result in acquiring new functions for the target cell, for example resistance to a certain disease. Gene deactivation can result in gene knockout and is particularly useful in the battle against genetic disorders.

Before the discovery of nucleases as a mean of performing genetic modifications, researchers mainly relied on random spontaneous mutations, demonstrated in the mid-twentieth century by Mendel, Morgan, Avery *et.al.* (Muller, 1927). Using Muller's techniques, alternations in target genome were performed by enhancing mutations with chemical and radiation treatments. Later on, another methods like transposon insertion were successfully performed on some organisms. Much like methods proposed by Muller et.al. those were both unpredictable, and often resulted in off-target activity – changes in the random or unwanted sites of the genome, other that desired region, or genes (Carroll, 2017).

The first breakthrough in genome engineering came in 1970-1980s (Scherer & Davis, 1979), when researchers reported successful targeted genome editing in yeast cells (Rothstein, 1983) and mice (Thomas et al., 1986). The process required use of homologous recombination, delivering remarkably precise targeting, but at the price of low efficiency. Additionally, gene targeting was limited by the absence of cultivable stem cells other than mice, which made adaption for use in other species practically unavailable (Mansour et al., 1988).

The situation changed in 1996, when Kim *et.al.* published their work on the first ZFNs (Zinc-Finger Nucleases) (Kim et al., 1996) and fusing of zinc-fingers together with *FokI* nuclease. It was based on the work of Miller *et.al.* (Miller et al., 1985), who previously reported the discovery of zinc-fingers in 1985. This new technology was tested both *in vitro* (Smith et al., 2000) – on microorganisms, cells and biological molecules outside of their usual biological surroundings, and *in vivo* (Bibikova et al., 2001) – on living organisms, and/or cells. These discoveries made it possible to start a new era of modern genome editing, and perform genome alternations in both model organisms (Bibikova et al., 2002), animal (Mani et al., 2005), human (Kandavelou et al., 2009; Urnov et al., 2005) and plant cells (Townsend et al., 2009).

Around the same time, another important discovery was made – a DNA binding molecule discovered in plant virulence factors – a so-called TALE motif (Transcription Activator-Like Effector) gave rise to TALENs – Transcription Activator-Like Effector Nucleases (Moscou & Bogdanove, 2009; Boch et al., 2009). TALENs were designed to perform in almost the same way as ZFNs, both complexes use *FokI* nuclease, but the DNA-binding mechanism is different (see Table 1). TALENs were proven to perform at the same rate efficiency as ZFNs, but appeared to have lower cytotoxicity and hence lower off-target activity in cells (Ramalingam et al., 2014).

TALENs had some quite useful advantages compared to ZNFs – they were easier to generate and had better target-specificity. At the same time TALENs proved to be more difficult to deliver into mammalian cells (Holkers et al., 2012), and plants (Chen & Gao, 2013). In addition, high initial pricing of ~$5,000 per target made TALENs practically unavailable for small laboratories.

Before TALENs could establish themselves as a viable alternative to ZFNs another genome-editing tool was discovered – an adaptive immunity mechanism in bacteria and archaea – CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats. Briefly, bacteria and archaea can use a set of proteins as a defense mechanism for degradation of complimentary sequences present within previously encountered invading viral and plasmid DNA or RNA. CRISPR-Cas systems use parts of viral DNA to compose short CRISPR RNA fragments (crRNA). Those guide RNAs can then detect and degrade viral nucleic acids with help of certain Cas'es (Nishimasu et al., 2014).

This mechanism has been studied and adopted for use in genome engineering. The most used CRISPR-associated protein – Cas9 was studied and proved to be an endonuclease. Together with crRNA Cas9 forms a complex functionning as an RNA-guided endonuclease with RNA-directed target sequence recognition and protein mediated DNA cleavage. (Gasiunas et al., 2012). CRISPR-Cas9 quickly became a point of interest. Several research group have successfully engineered and performed genome editing experiments with CRISPR-Cas9 in many organisms, mammalian cells and plants (Hatoum-Aslan et al., 2013; Feng et al., 2013; Cong et al., 2013; Cho et al., 2013; Woo et al., 2015).

CRISPR-Cas systems have been reportedly performing at comparable or higher target efficiency as TALEs and zinc-fingers (Chandrasegaran & Carroll, 2016). There have been reports of successful simultaneous introductions of multiple guide RNAs into cells for multiplex gene editing, a process easily achieved with CRISPR compared to TALENs and ZFNs (Cong et al., 2013). Cas9-mutant nucleases have been used to perform single strand break, or knocking out a single nucleotide, giving arise to nickases. Nickases have been used to produce single strand breaks with overhangs for precise homology directed repair, resulting in precise gene integration and insertion (Shen et al., 2014).

The benefits of CRISPR-Cas9 systems have brought researchers attention to other Cas proteins. Many studies have been performed in order to find possible alternatives to Cas9 and get a better understanding of the CRISPR-Cas locus. Several Cas proteins have shown either DNase activity, RNase activity, or both. Bioinformatic analyses of CRISPR locus of several organisms containing CRISPR genes have shown approximately 65 Cas orthologues divided into two classes, six types, and 30 subtypes, based on CRISPR-Cas classification system, proposed by Makarova *et. al.* (Makarova & Koonin, 2015; Makarova et al., 2017; Makarova et al., 2015).

Up to this date there is still no complete overview of CRISPR-Cas systems, Cas proteins and their functions. This thesis aims to gather such information and clear up the situation around CRISPR-Cas proteins for better understanding of those programmable nucleases.

Materials & Methods

A variety of research papers have been collected and analyzed in order to get an overview over CRISPR-associated proteins. A total of 56 (140 including orthologues) proteins have been studied; Their functions and applications are shortly described. All scientific papers used during the research are listed in the “References” chapter.

Results

Programmable nucleases as tools for efficient and precise genome editing

Short presentation of other programmable nucleases is required in order to achieve better understanding of CRISPR in context of genome editing tools. The discovery of programmable nucleases able to perform a DNA and/or RNA cleavage at the desired target-site in genome has become a breakthrough in genome engineering. There are several means to perform genome editing - ZFNs (figure 1a), TALENS (figure 1b) and CRISPR (figure 1c).

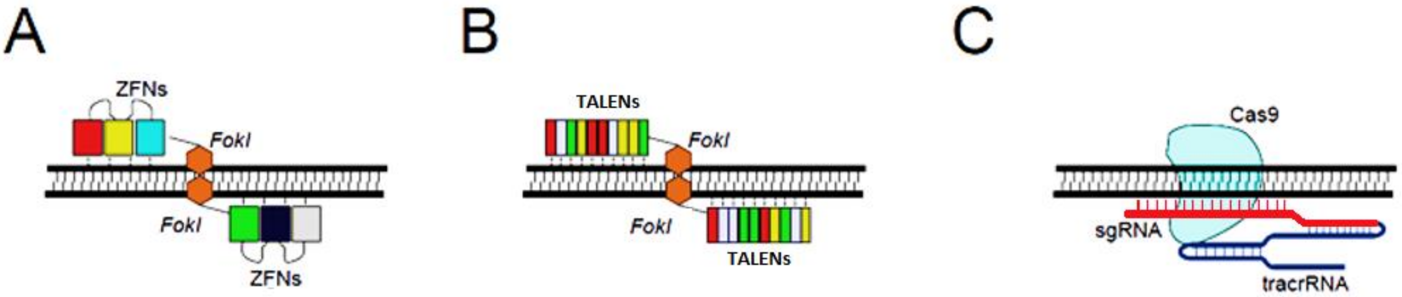

Figure 1. *Ptglab.com. (2019). CRISPR-Cas9, TALENs and ZFNs - the battle in gene editing. Available at: https://www.ptglab.com/news/blog/crispr-cas9-talens-and-zfns-the-battle-in-gene-editing/ (Accessed 9 Mar. 2019).*
A: ZFNs – two ZFNs, constructed to recognize and bind to specific sites at opposite DNA strands; a FokI restriction enzyme dimer cleaves DNA at the target site. B: TALENs – two TALENs, designed to recognize and bind to target site at opposite DNA strands; A FokI dimer cleaves target DNA. C: CRISPR-Cas9 system, the target site is recognized complementary, a bond is formed between the genomic DNA and crRNA (sgRNA + tracrRNA), Cas9 nuclease performs DNA cleavage.

Programmable nucleases cleave DNA or RNA in order to knock out genes, perform gene correction or transgene addition of a new set of genes (figure 2).

Gene disruption by NHEJ (Non-homologous end joining, figure 2A) is a process of deactivating a gene, or a set of genes by cleaving the gene, without a homologous template available for DSB (Double-Strand Break) repair. Instead of repairing the gene by using a copy of gene from either a donor or other copy of the same gene present in the genome, this process simply repairs DSB by ligating DNA. This process may result in a so-called frameshift mutation. Frameshift mutation is loss of nucleotides that leads to loss of protein function encoded by the gene, often because of premature stop codon (Robertson et al., 2009).

Gene correction by HDR (Homology Directed Repair, figure 2B) is another way of altering the target genome with the help from DNA repair mechanisms of the cell. HDR repair is only possible if a homologous copy of the gene is present in the genome. A WT (wild type) copy of the gene can be delivered to the target-site and used as a template for DSB repair. It is useful if a mutated or defected gene is no longer functional. The WT gene will be used as a template for the repair of the target sequence and the gene can restore its functions (Robertson et al., 2009).

HDR can be used to adopt a new gene or a set of genes for a so-called transgene addition (Figure 2C). The pathway is similar to gene correction by HDR, but instead of WT gene a new gene previously not present in the cell, and often adapted from another organism will be delivered as a template for HDR repair pathway (Robertson et al., 2009).

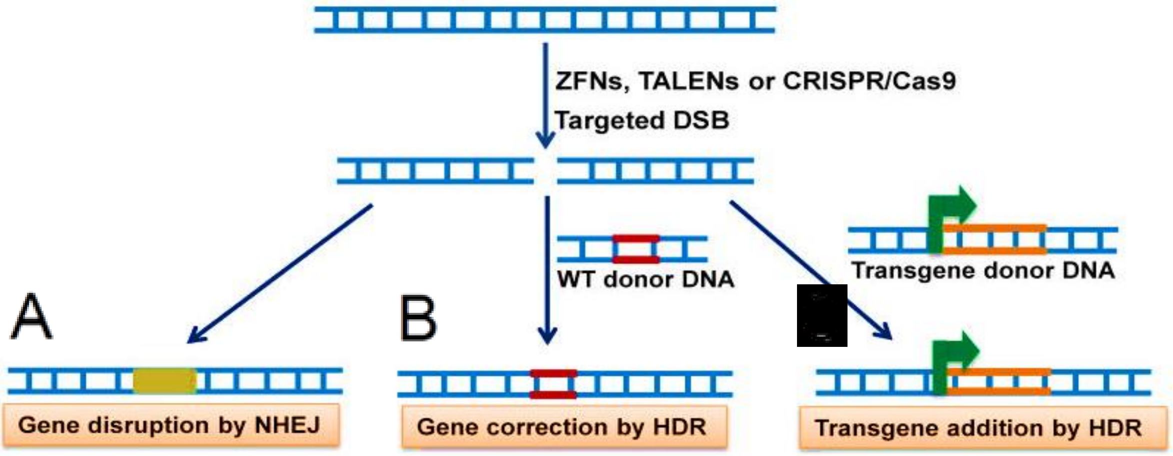

Figure 2. *Chandrasegaran, S. & Carroll, D. (2016). Origins of programmable nucleases for genome engineering. Journal of molecular biology, 428 (5): 963-989. Genome engineering by ZFNs, TALENs or CRISPR-Cas9. Graphic representation of how programmable nucleases are used to perform either gene knock out by NHEJ(A), gene correction by HDR(B) or addition of new genes by HDR (C).*

Programmable nucleases offer a wide spectrum of opportunities with areas of use such as genomic modifications in model organisms, disease vectors and organisms, crop plants, human cells, livestock and primates(Ma et al., 2013; Ramalingam et al., 2014; Aryan et al., 2013; Genovese et al., 2014; Ghorbal et al., 2014; Haun et al., 2014; Carlson et al., 2012; Niu et al., 2014; Tan et al., 2013). Despite the differences in functionality, design and applications (table 1), programmable nucleases have one thing in common – means of performing effective and successful genome editing (Segal & Meckler, 2013).

Table 1. *Based on Ptglab.com. (2019). CRISPR-Cas9, TALENs and ZFNs - the battle in gene editing. Available at: https://www.ptglab.com/news/blog/crispr-cas9-talens-and-zfns-the-battle-in-gene-editing/ (Accessed 9 Mar. 2019).*

Feature	ZFNs	TALENs	CRISPR-Cas9
Recognized DNA target length	9–18 base pairs	30–40 base pairs	18-22 base pairs + PAM sequence
Means of target sequence recognition	DNA–protein interactions	DNA–protein interactions	DNA–RNA interactions by Watson-Crick base pairing
Means of target cleavage and repair	Double-strand break performed by a *FokI* restriction enzyme dimer	Double-strand break performed by a *FokI* restriction enzyme dimer	Both single- and double-strand breaks performed by Cas9 nuclease
Preparation	Challenging. ZFNs libraries are available, but the final complex must be tested for target specificity.	Easier than ZFNs. TALE motifs with target specificities are well defined. Several TALEs per nucleotide are available.	Easy. Guide RNA must be programmed to be complimentary to the target sequence.
Commercial pricing	Very expensive ($4,000 to $7,000 per target)	Expensive ($3,360-$5,000 per target)	Cheap ($500 per target)
Targeting efficiency	Variable*	Moderate	High‡
Off-target effects	Variable*	Low	Moderate‡
Multiple targets	Difficult	Difficult	Easy
Viral delivery	Easy	Moderate	Moderate
Advantages and disadvantages	Neighboring ZFNs can affect each other's specificity. *FokI* performs double strand break when in dimer form. A total of two ZFNs must be designed – one for 5'-3' strand and one for 3'-5' strand upstream and downstream the target sequence. One ZFN binds to three nucleotides of the target sequence.	Good specificity and little off-target activity. *FokI* performs double strand break when in dimer form. A total of two TALENs must be designed – one for 5'-3' strand and one for 3'-5' strand upstream and downstream the target sequence. One TALE is required per nucleotide of the target sequence.	PAM downstream of target DNA/RNA sequence is required to perform complex binding. Compared to protein-DNA interactions - easy to use and prepare due to DNA-RNA interactions. Complex tolerates mismatches between guideRNA and target site, some mismatches and off-target activity can occur.

**Depending on design of ZFN. Can vary from high to low.*

‡Depending on design of the guide RNA and target site.

ZFNs

The first endonucleases used for genome editing were ZFNs. ZFNs are composed of the endonuclease called *FokI,* and zinc-fingers proteins, which are a family of naturally occurring transcription factors.

ZFNs are DNA binding molecules that can be arranged in a linear polar fashion, and work by recognizing trinucleotide sequences of different lengths and provide a desired on-target specificity, both *in vitro* and *in vivo.* Each zinc-finger has a common backbone, but a variety of free amino acids makes them specific for certain nucleotides. Alternations of the free amino acids on the α-helix leads to a nucleotide-specific bond between the amino acids and the complimentary nucleotides in the target genome sequence.

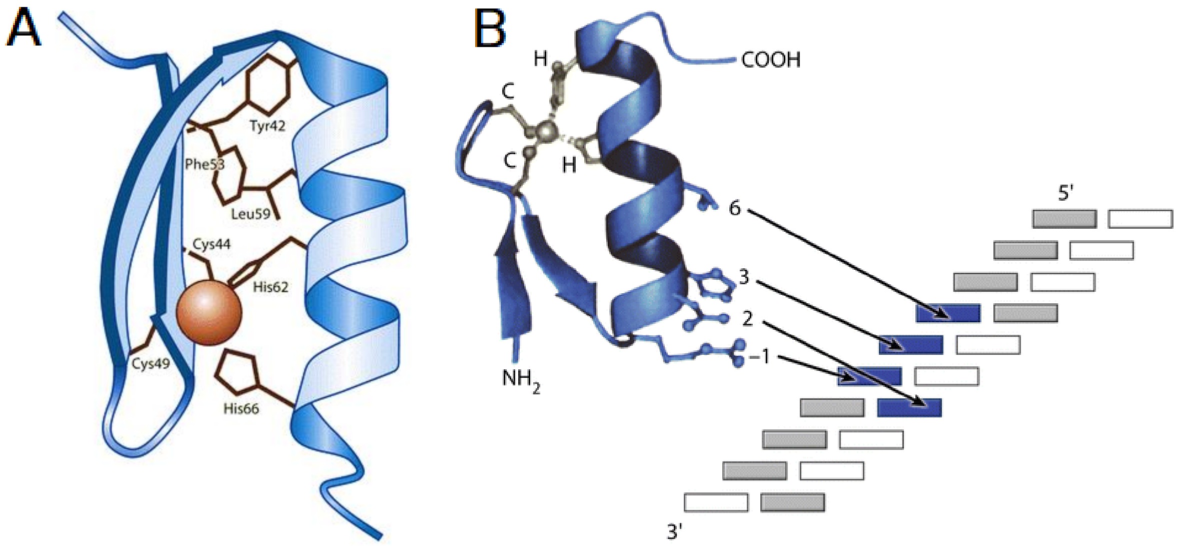

Figure 3. *Klug, A. (2010). The discovery of zinc-fingers and their applications in gene regulation and genome manipulation. Annual Review of Biochemistry, 79: 213-231.*

A: Graphic representation of a zinc-finger protein – double β-sheet and a single α-helix, stabilized by Cys2-His2 site and a Zn-molecule (shown in brown).
B: DNA binding mechanism of ZFNs. A total of four free amino acids of the α-helix are forming a bond to target site. Amino acids in position one, three and four of α-helix binding to the 3'-5' strand of the target DNA, and amino acid number two stabilizing the bond by attaching ZFN to a single nucleotide on the complimentary 5'-3' strand.

A zinc-finger consists of two main components, as shown in figure 3A:

The first component - an α-helix, uses hydrogen bonds interactions from the amino acids and forms a triple bond to three nucleotides (a triplet) on one strand of the DNA (figure 3B) (Pavletich, 1991). Furthermore, discovered by Klug (Klug, 2010), there is a fourth interaction from the second position in the α-helix to the complementary DNA strand. The other main component in addition to the α-helix is a highly conserved cys$_2$-his$_2$ site that is fundamental in the protein folding of the zinc-fingers by coordinating a Zn-molecule. The cys$_2$-his$_2$ site and the three amino acids Tyr42, Phe53, and Leu59 are forming a hydrophobic structural core of the complex. The numbers are referring to the position of the amino acids in the protein sequence, counted from N-terminal to C-terminal, noted as -NH_2 and -COOH respectively (figure 3B).

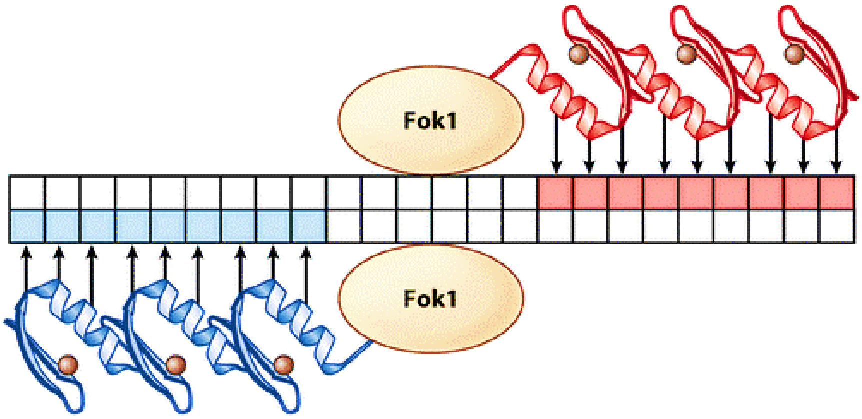

Figure 4. *Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. Annual Review of Biochemistry, 79: 213-231.*

Mechanism of DNA binding and FokI dimer formation by zinc-finger nucleases.

The main disadvantage of ZFNs is zinc-finger motif specificity influence of neighbor ZFNs making it difficult and time-consuming to design a ZFNs experiment without negative interactions within the complex. This phenomenon is referred to as cytotoxicity and makes specificity prediction challenging. Solid preparations prior to experiment must be made to achieve satisfactory results and limit off-target activity of the complex.

Another downside of ZFNs is that the endonuclease used in ZFNs – *FokI* – must form a dimer in order to perform cleavage. A complex of two *FokI,* one on each DNA strand, are necessary to perform a successful DNA cleavage. A total of two different ZFNs – one on 5’-3’ strand and one on 3’-5’ strand, that have to recognize different, but closely located nucleotide sequences must be designed for a single cleavage (figure 4). The advantages of that are limitations linked to off-target activity (Kim, 1996) (see table 1).

TALENs

Just like ZFNs, TALENS perform DNA cleavage by forming a *FokI* dimer. TALENs are formed from a series of TALEs - highly conserved repeats, where a single TALE recognizes one specific nucleotide. The construction of engineered TALE repeat domain requires use of multiple and nearly identical sequences. TALENs can be designed to perform with high specificity. Unlike ZFNs TALENs are not affected by a presence of neighbor TALENs, which makes them easier to construct (see table 1). The process of designing a TALENs-*FokI* complex is rather rapid using a DNA-code of the target-binding site and composing a complimentary DNA-binding TALE domain that repeats domains to individual bases in target-binding site in the genome. TALENs deliver high-success rate and can be adopted for use in essentially any DNA sequence of interest.

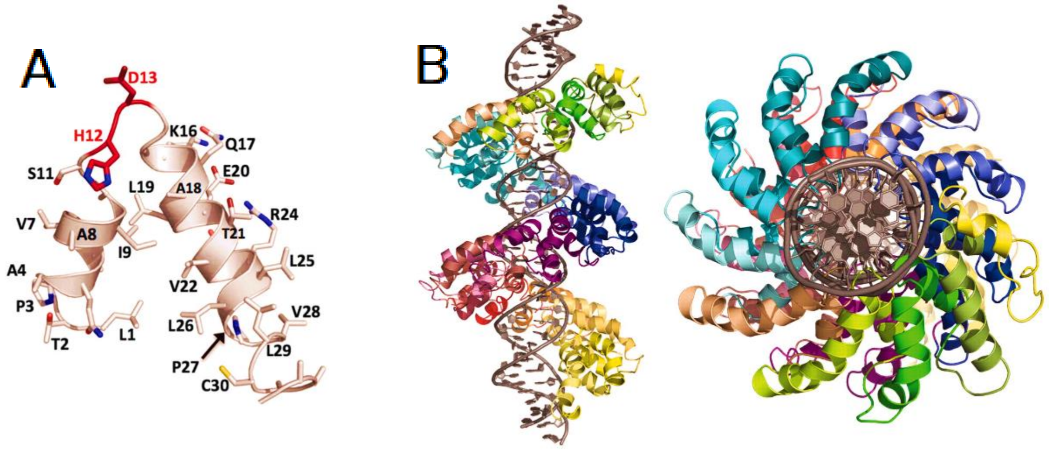

Figure 5. *Mak, A. N.-S., Bradley, P., Cernadas, R. A., Bogdanove, A. J. & Stoddard, B. L. (2012). The crystal structure of TAL effector PthXo1 bound to its DNA target. Science, 335 (6069): 716-719.*

A. A single TALE protein crystal structure.
B. TALENs bound to major groove of DNA sequence.

TALENs complex consist of a TALE repeat domain - individual TALE repeats arranged in an array to bind specifically to a single base each. The bond is formed by two hypervariable residues at 12th and 13th position in the TALE protein (marked red in figure 5A), located between two α-helixes (Boch et al., 2009; Moscou & Bogdanove, 2009). The protein is V-shaped and forms a superhelix around the DNA, positioning 12th and 13th residue of the TALE in the major groove of the DNA, where the residue 13 makes a base-specific contact with the DNA (figure 5B). Nearly all engineered TALE repeat arrays available today use four different domains to make the base-specific bond – NN for recognition of guanine, NI for adenine, HD for cysteine, and NG for thymine. It has been reported that another residue – NK – makes even better base recognition than NN (which can also recognize adenine) and forms a bond with guanine, but NK repeats show less activity than NN (Joung & Sander, 2013).

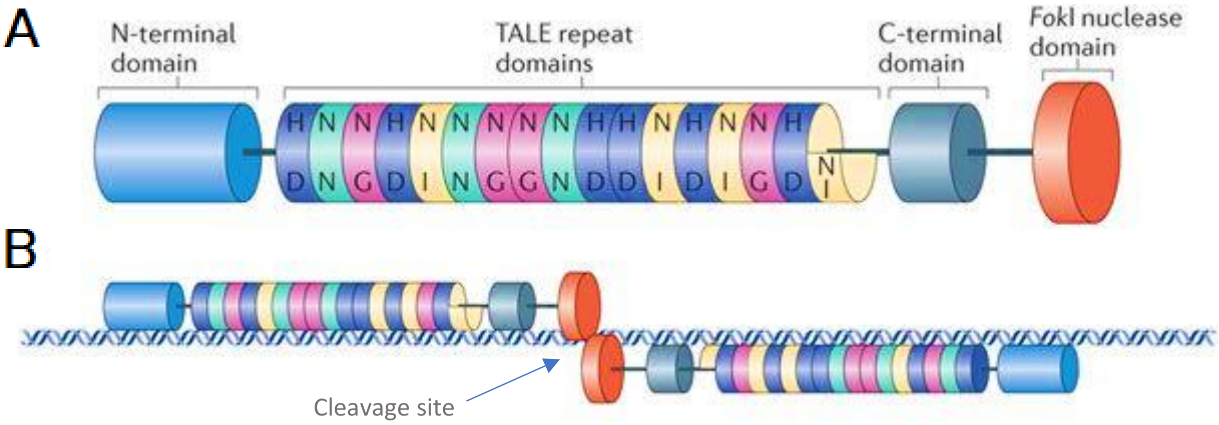

Figure 6. *Joung, J. K. & Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. Nature reviews Molecular cell biology, 14 (1): 49.*

A. Graphical representation of TALENs, with TALE repeat domains that bind specifically to single nucleotides.
B. Mechanism of DNA binding by TALENs, with formation of FokI dimer.

Figure 6A shows TALENs domain somposition, including N-, and C-terminals, TALE repeat domain and *FokI* nuclease domain. Figure 6B shows binding pattern of TALENs, similar to that of ZFNs. Two TALENs are complimentary bound to both DNA strands upstream and downstream of cleavage site. *FokI* dimer is formed at the cleavage site to perform DSB (Streubel et al., 2012).

CRISPR-Cas9

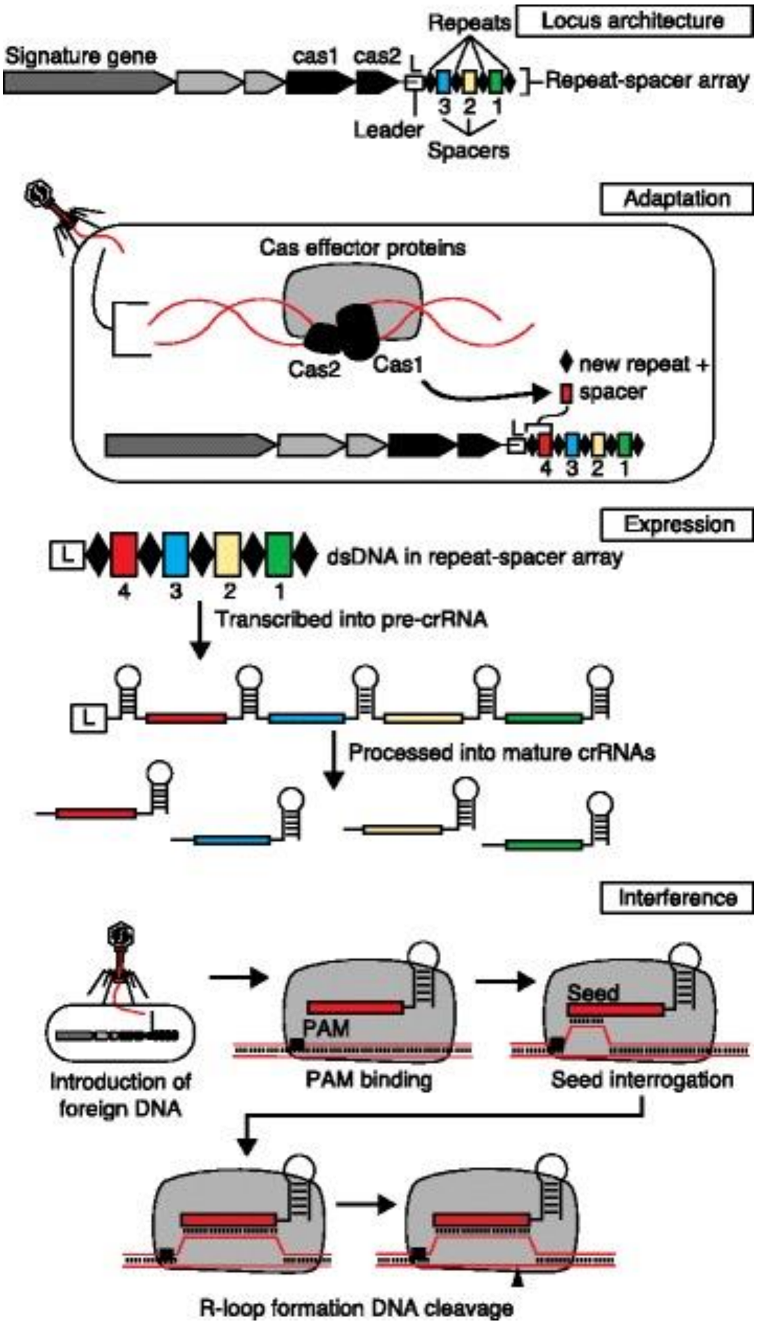

Figure 7. *Barrangou, R. (2015). Diversity of CRISPR-Cas immune systems and molecular machines. Genome Biology, 16 (1): 247.*

CRISPR-Cas systems. CRISPR-loci architecture and the three steps of CRISPR-Cas immunity response – adaption, expression and interference.

CRISPR-Cas systems were first discovered as a part of adaptive immunity biological process in bacteria and archaea. During the last years a complex called CRISPR-Cas9 has been successfully used as a genome editing tool. The whole process of CRISPR editing is dependent on a series of smaller processes. Protospacer Adjacent Motif, or shortly PAM - a short nucleotide sequence, usually three or five nucleotides, have to be recognized by the complex in order to start the process. CRISPR-systems use PAM sequences to differentiate between invading and own DNA or RNA. This target recognition is achieved and performed by a seed sequence, which residues at the 5' end of the crRNA spacer (Barrangou, 2015).

CRISPR-Cas9 complex consist of two main components - a guide RNA and a CRISPR-associated protein Cas9 - an endonuclease which can perform double strand breaks. The guide RNA is a user-composed, specially designed sequence of approximately 20 nucleotides responsible for recognizing target sequence and complimentary binding of the target gene. The guide RNA can be designed and modified accordingly to user's desire to target a specific area of the genome that is to be altered by the complex.

Each Cas9 protein has a specific PAM sequence (see Cas9, table 2), for example 5'-NGG-3', required for target-site recognition (Jinek et al., 2012).

In nature, the process of CRISPR-Cas immunity response is based on three steps: Adaptation, expression and interference (figure 7).

During adaptation step, after foreign DNA is detected, Cas effector proteins will cleave the invasive DNA. Small parts of this DNA are then adapted as spacers - part of repeat-spacer array.

During the expression stage, CRISPR-array is transcribed into pre-crRNA and is further processed into mature crRNA. Mature crRNA is composed of both partial CRISPR spacer sequences and partial CRISPR repeats, together those will form a mature CRISPR guide RNA (Hsu & Zhang, 2014).

In the last stage - interference - crRNA will guide CRISPR-Cas towards PAM sequence for complimentary binding to the foreign DNA. Once the PAM sequence is detected, the complex can bind to the foreign DNA by forming a bond between seed sequence and the target. If the level of correspondence between guide crRNA and the foreign DNA is high, the bond between crRNA and the foreign DNA will extend over the seed sequence and further to the spacer region. The result is formation of an R-loop, and eventually cleavage of the target DNA approximately three bases upstream of the PAM (Barrangou, 2015).

The advantages of CRISPR-Cas9 over ZFNs and TALENs lies in RNA-DNA interactions, providing amongst other much easier design for any genomic targets, easy off-target prediction and multiplexing – the possibility of modifying several genomic sites simultaneously.
One of the main disadvantages of the CRISPR-Cas9 *in vivo* is relatively high tolerance of mismatches. CRISPR-Cas9 tolerates up to 25% of mismatches (one to six base pairs) between guide RNA and target-sequence, potentially leading to an increased level of off-target activity and cytotoxicity (Hsu et al., 2013).

CRISPR guide RNAs

In order to achieve target specificity *in vivo* and *in vitro* CRISPR systems use series of small RNA molecules, called guide RNAs. *In vivo*, guide RNAs are acquired directly from invading viral nucleic acids. *In vitro,* however, guide RNAs can be programmed to target specific sites in the genome based on sequence complementarity. There are three different types of guide RNAs:

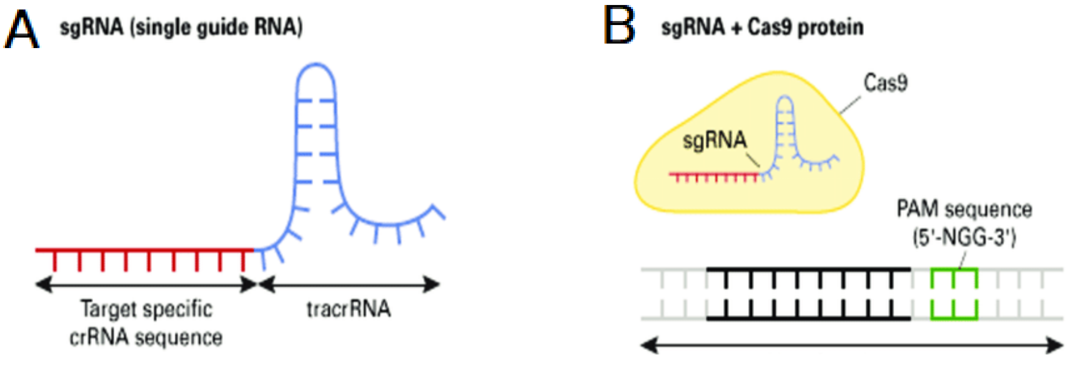

Figure 8. *El-Bassyouni, H. & Ahmed Mohammed, M. (2018). Genome Editing: A Review of Literature*.

sgRNA

SgRNA, or single guide RNA is a RNA sequence that contains both crRNA and tracrRNA (figure 8A). SgRNA is a crucial part in CRISPR-Cas9 gene targeting process (figure 8B), required for both sequence targeting, and nuclease recruiting for permanent genome alterations.

crRNA and pre-crRNA

CrRNA, or CRISPR RNA is a short guide sequence (~20 nucleotides) used for complimentary binding to the target-site in the genome. In nature, spacers acquired during adaption stage are used to compose crRNA, allowing the complex to bind specifically to the invading target sequence.

Process of crRNA maturation starts under expression stage (figure 7), when CRISPR repeat-spacer arrays are transcribed into precursor crRNAs – pre-crRNA. Further processing of pre-crRNA results in mature crRNA sequence. Mature crRNA is a set of repeat fragments acquired directly from viral invading DNAs. Nearly all CRISPR-Cas systems, apart from subtypes II, V-B and V-E, use Cas6 to cleave pre-crRNA to generate mature crRNA. For gene editing purposes, crRNA is user-designed prior to the experiment. User-designed crRNA is designed to bind complimentary to target-site in target genome. (Karvelis et al., 2013)

tracrRNA

In Class II CRISPR-Cas subtypes II, V-B and V-E, process of crRNA maturation is slightly different and is dependent on an additional RNA molecule called tracrRNA. TracrRNA is *trans*-encoded small RNA sequence complimentary to repeat regions of crRNA required for Cas9-nuclease recruiting. (Deltcheva et al., 2011) The term *trans* points to tracrRNA origin, being processed from spacer-repeat region of CRISPR, composed from spacers acquired from viral DNA. TracrRNA is vital for the process of cleavage by CRISPR-Cas9 complex. Deletion of tracrRNA from CRISPR-Cas9 complex results in deactivation of the whole process. (Karvelis et al., 2013) Main role of tracrRNAs in CRISPR systems is maturation of crRNA by directing pre-crRNA processing. CrRNA maturation involves two steps:

→ First processing event:

TracrRNA binds to repeat sequences of pre-crRNA resulting in formation of a double-stranded RNA duplex. In contrast to other CRISPR systems, duplex is recognized by a ribonuclease – RNase III, instead of Cas6. Recruiting of RNase III results in a site-specific cleavage of pre-crRNA. The products of cleavage are space-repeat-space units. (Deltcheva et al., 2011)

→ Second processing event:

Under the second process event, space-repeat-space units undergo further processing. During this stage, removing 5'-end of the space-repeat-space unit result in 39-42 nucleotides mature crRNA sequence. (Deltcheva et al., 2011)

Major types of CRISPR-Cas systems and their subtypes

In 2011 Makarova *et.al.* (Makarova et. Al, 2011b) suggested new classification and nomenclature for CRISPR-Cas systems. Before that, classification of CRISPR-associated proteins was primarily based on Cas1 phylogeny since it was assumed to be the only Cas preserved amongst all CRISPR systems. Since that time, a large amount of research has shown that CRISPR-Cas systems can be divided and classified based on CRISPR-Cas locus architecture (figure 9) (Koonin & Makarova, 2019).

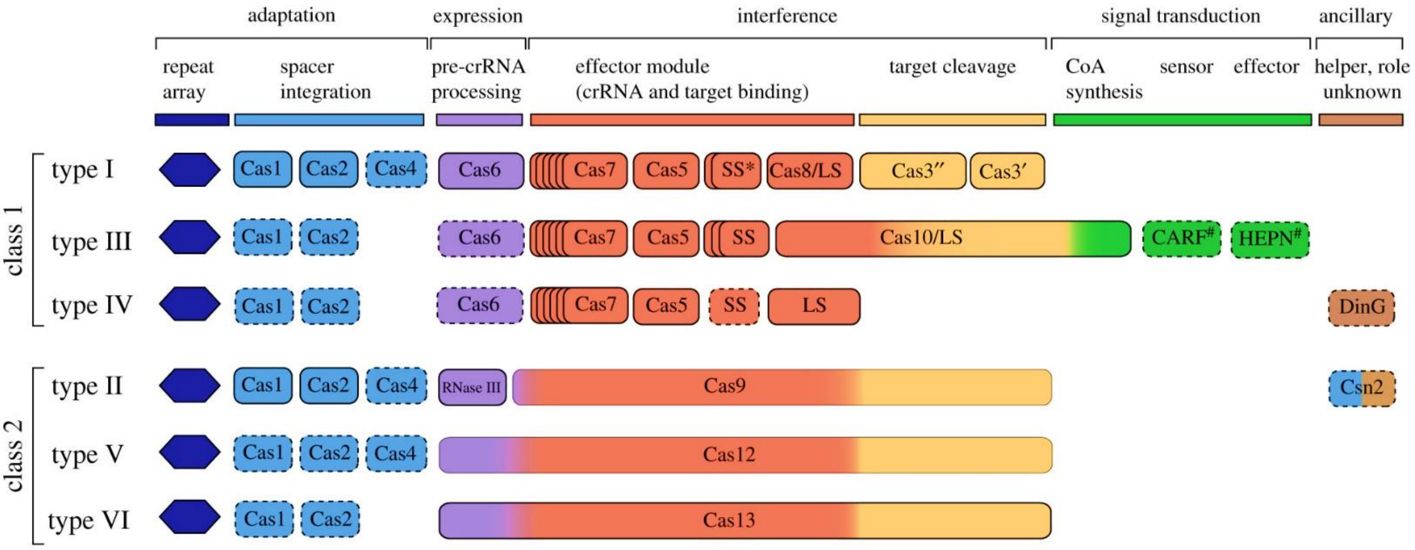

Figure 9. *Koonin, E. V. & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. Philosophical Transactions of the Royal Society B, 374 (1772): 20180087.*

CRISPR locus architecture of CRISPR systems.

Figure 9 shows typical CRISPR locus architecture for all known CRISPR systems, with their respective types and Cas proteins involved in different steps of the complex. As shown on the figure, there are two distinct classes of CRISPR-Cas'es – class I (figure 10A) and class II (figure 10B). Class I and II are further divided into several types with distinctive subtypes. The reason why CRISPR systems are divided in such lies in class I Cas systems use of multiple single proteins for expression, interference and adaption steps, while class II systems use single multidomain proteins for the same purposes.

Class I and II CRISPR systems differ in the mechanism of action and target-binding motifs, as for example PAM motif recognition. In class I CRISPR-Cas systems PAM motif is located upstream of seed sequence, while in class II it is located either up, - or downstream for seed sequence (figure 10 A and B) (Leenay et al., 2016).

Detailed locus composition of CRISPR Class I and Class II is presented in Appendix III and Appendix IV respectively.

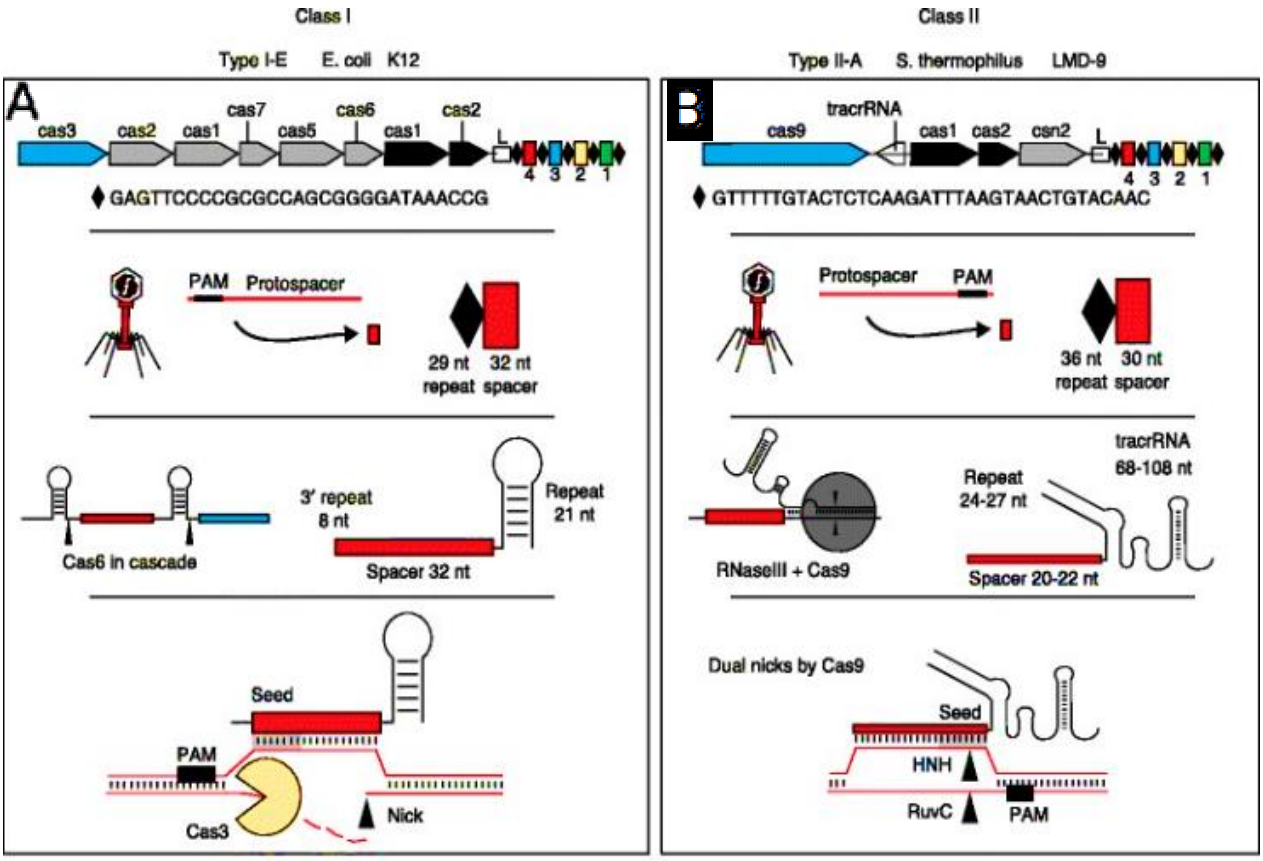

Figure 10. *Barrangou, R. (2015). Diversity of CRISPR-Cas immune systems and molecular machines. Genome Biology, 16 (1): 247.*
Class I and class II CRISPR-Cas systems.
A. Class I loci architecture, protospacer composition, R-loop formation and DNA cleavage mechanism, with Cas3 as an example.
B. Class II loci architecture, protospacer composition, R-loop formation and DNA cleavage mechanism, with Cas9 as an example.

Class I CRISPR-Cas systems utilize multi-protein complexes. Class I is divided into three types: I, III and IV, and 12 subtypes. Class I CRISPR systems represent about 90% of the CRISPR-Cas locus discovered in bacteria and archaea. (Makarova et al., 2017)

Class II CRISPR-Cas systems use single-protein effectors (Sternberg & Doudna, 2015). Class II is divided into three types – II, V and VI, and further into ten subtypes. Class II CRISPR systems represent the last ten percent of CRISPR locus in bacteria and is absent in archaea. Direct repeats of class II systems can be both palindromic (inverted-reverse sequence, reads the same back and forward) and non-palindromic. (Komor et al., 2017)

Following CRISPR-Cas types description is based on research by Makarova, Koonin and Haft (Haft et.al, 2005; Makarova et.al 2011,2015a,2015b,2018; Koonin et.al., 2017; Koonin & Makarova, 2019).

Type I CRISPR-Cas systems

All type I CRISPR-Cas systems contain a signature gene – *Cas3.* Type I systems are encoded by a single operon containing Cas1 and Cas2, subunit proteins of Cascade effector complex, including large subunit, small subunit (often fused to a large subunit), *Cas5*, *Cas6*, *Cas7* and *Cas8* genes.

The CRISPR-Cas type I systems are divided into eight subtypes, all target DNA sequences:

I-A

Signature genes for subtype I-A are Cas8 alternative *Cas8a1* (large subunit), and *Cas11*. Cas3 is often split into two domains – helicase Cas3' and HD nuclease Cas3''. Csa5 is often present as a small unit.

I-B

Cas8b serves as a signature gene for the subgroup. Unlike I-A subtype, Cas3 is not split into two domains.

I-C

Cas8c is a signature gene for the subgroup I-C. The subgroup usually lacks *Cas6* gene, and Cas5 replaces its catalytic functions.

I-D

The HD domain (nuclease domain) is associated with the large subunit instead of Cas3.

I-E

Lacks *Cas4* gene.

I-F

Lacks *Cas4* gene, and Cas2 is fused to Cas3, there is no separate gene for small subunit (missing or fused to large subunit).

I-F variant

Same as I-F, but additionally lacks *Cas8* gene.

I-U

CRISPR-Cas proteins that show similarity to type I systems architecture, but biological functions are yet unknown.

Type II CRISPR-Cas systems

The signature gene for this CRISPR-Cas system is *Cas9. Cas9* encodes a multidomain protein that combines all the functions of effector complexes and the target DNA cleavage. The protein is essential for the maturation of the crRNA.

Every CRISPR-Cas type II locus contains *Cas1* and *Cas2* in addition to *Cas9* genes and requires tracrRNA for proper functioning. Type II CRISPR-Cas system has been developed into a powerful genome-engineering tool during the past years.

Type II CRISPR-Cas system are divided into four subtypes, all target DNA sequences:

II-A

Lacks *Cas4* gene. Has an additional protein – Csn2 (signature protein for the subtype*).* Csn2's function is spacer acquisition and integration.

II-B

Subtype II-B systems do not possess the C*sn2* gene, but has a fourth distinct gene from Cas4 family, that is also associated with subtypes I-A to I-D.

II-C

Is the newest subtype in the type II CRISPR-Cas systems. II-C subtype possesses only three genes – *Cas1*, *Cas2* and *Cas9*, more common in sequenced bacterial genomes.

II-C variant

Same as II-C, but has alternative types of Cas1 and Cas2 proteins and a *Cas4* gene which is absent in subtype II-C.

Type III CRISPR-Cas systems:

The signature gene for the type III CRISPR-Cas systems is *Cas10*. Most of type III CRISPR-Cas systems do not encode their own distinct Cas1 and Cas2 proteins, but use crRNAs produced by CRISPR arrays associated with type I or II CRISPR-Cas systems.

Type III CRISPR-Cas is divided into five subtypes:

III-A

Subtype III-A often possess *Cas1*, *Cas2*, *Csm6* and *Cas6* genes. Has only two Cas7 copies, in comparison to III-B and III-C, where both have three, and III-D that has four copies. Targets DNA and RNA.

III-B

Subtype III-B lacks *Cas1*, *Cas2* and *Csm6* genes and is dependent on other CRISPR-Cas systems that are present in the same genome. Targets DNA and RNA.

III-C

Resembles III-B, but has different locus architecture. Lacks *Cas6* gene. Targets DNA and RNA.

III-D

Has four copies of *Cas7* gene and an additional unidentified gene, lacks *Cas5* gene. Presumably targets RNA.

Type IV CRISPR-Cas systems

The *Csf1* gene can be considered a signature gene for the type IV CRISPR-Cas systems, that lacks *Cas1* and *Cas2* genes. Type IV systems possess an effector complex that consist of highly reduced large subunit Csf1, two genes for RAMP proteins of the Cas5 (*Csf3*) and Cas7 (*Csf2*) groups, and in some cases a gene for small subunit.

Type IV CRISPR-Cas systems consist of two distinct subtypes:

IV-A

Contains a helicase Csf4, and Cas6 analogue.

IV-B

Contains a gene for a small alpha helical protein, presumably a small subunit, lacks *csf4* and *Cas6* genes.

Type V CRISPR-CAS systems

Signature gene for type V systems is *Cpf1* (*Cas12*). This Cas-protein is a large protein that contains nuclease domain RuvC, homologous to Cas9, but lacks the second nuclease domain present in all Cas9 systems – NHN. Type V CRISPR-Cas systems target DNA, and are composed of seven subtypes:

V-A

Consists of Cpf1 multidomain protein (Cas12a), Cas4 nuclease, and *Cas1* and *Cas2* genes.

V-B

This subtype has another variant of Cpf1, often referred to as Cas12b, Cas4 is fused to Cas1. Unlike V-A, V-B subtype uses tracrRNA.

V-B variant

Same as V-B, but different locus architecture.

V-C

Contains another Cpf1 analog – Cas12c. Lacks Cas4 and Cas2 and has slightly different locus architecture.

V-D

Same as V-C, but has different locus architecture and Cas12d variant.

V-E

Same as V-A, but uses tracrRNA and yet another Cas12 analogue – Cas12e.

V-U

Tentative. This subtype is for CRISPR-Cas proteins that show similarity to type V systems architecture, but biological functions are yet unknown. There are total five subtype V-U variants that differ in Cpf1 composition.

Type VI CRISPR-CAS systems

Type VI CRISPR-Cas systems have a common signature gene – *Cas13*. Type VI CRISPR-Cas systems target RNA. Majority of VI types lack *Cas1* and *Cas2* genes.

VI-A

Contains Cas13a, Cas1 and Cas2.

VI-B1

Lacks *Cas1* and *Cas2* genes, has an additional *Csx28* gene and alternative Cas13b.

VI-B2

Same as VI-B2, but lacks Csx28 and has an additional Csx27 protein instead.

VI-C

Lacks *Cas1* and *Cas2* genes, composed of Cas13c only.

VI-D

Same as VI-A, but has an additional "WYL" gene of unknown function.

An overview of CRISPR-associated proteins

CRISPR-Cas proteins are a number of proteins typically found in the CRISPR-locus in a variety of microorganisms, such as bacteria and archaea. The CRISPR locus composition tends to differ in those organisms. Those differences provide possibility to group CRISPR systems in microorganisms based on the composition of CRISPR locus – see “Major types of CRISPR-Cas system and their subtypes”.

CRISPR-Cas proteins provide different functions in the CRISPR-Cas systems and are involved in antiviral defense against viral nucleic acids. Since the discovery of CRISPR-Cas systems, there have been many attempts to categorize, study and adapt CRISPR-Cas proteins for use in gene engineering and biotechnology studies. An understanding of CRISPR-Cas systems composition and functions could provide huge advantage for scientific applications.

CRISPR-Cas9 and its variants are probably the most used Cas-complexes in modern biotechnology, but the study of Cas’es gives indications that other Cas-proteins can be substitutes or even better alternatives to Cas9. Following is a short overview of CRISPR-Cas proteins discovered during this study, with a short description of their functions in the CRISPR-Cas systems when available.

In this study, CRISPR-associated proteins are divided into two groups: Essential proteins – Cas’es, and additional proteins – Cxx’es. “Cxx” is not an official name, and is only used in this thesis to describe the group of additional Cas proteins as a whole. The abbreviation “xx” indicates two letters of the short name of the protein, for example “sy” in Csy.

See Appendix I and II for a short list of all CRISPR-associated protein and their functions.

Essential CRISPR-Cas proteins

CRISPR-associated proteins have a variety of different functions in the complex, such as:

Nuclease – a restriction enzyme that can perform cleavage of phosphodiester bonds between nucleotide chains, such as DNA or RNA. Cleavage results in smaller nucleotide units.

DNAse – a nuclease specific to DNA chain cleavage, also called deoxyribonuclease.

RNAse – a nuclease specific to RNA chain cleavage, also called ribonuclease.

Endonuclease – a nuclease that performs non-specific cleavage of nucleotide sequence chain.

Exonuclease – a nuclease that can only perform cleavage at the end of nucleotide sequence chain, and one nucleotide at a time.

Exoribonuclease - a ribonuclease that can only perform cleavage at the end of ribonucleotide sequence chain, and one at a time.

Helicase – an enzyme able to separate duplex nucleic acids.

Integrase – an enzyme able to integrate nucleic acids into DNA or RNA sequences.

Endodeoxyribonuclease – a restriction enzyme that possess both deoxyribonuclease and endonuclease catalytic functions.

ATPase – an enzyme catalyzing ATP degradation to ADP + free phosphate ion, releasing energy that enzyme can use to catalyze chemical reactions.

RAMP - Repeat Associated Mysterious Protein, a family of proteins containing RRM (RNA recognition motif).

Casposase – CRISPR-Cas transposase.

Transposase – an enzyme able to bind transposons (short DNA sequence) and move them to another site in the genome.

Polymerase – an enzyme that catalyzes DNA or RNA polymer synthesis.

Reverse transcriptase – an enzyme that catalyzes complimentary DNA from RNA template.

Cyclase – an enzyme that can catalyzes cyclic compounds – chemical compounds formed as a ring.

Slicer protein – an enzyme able to degrade nucleic chains.

Cas1

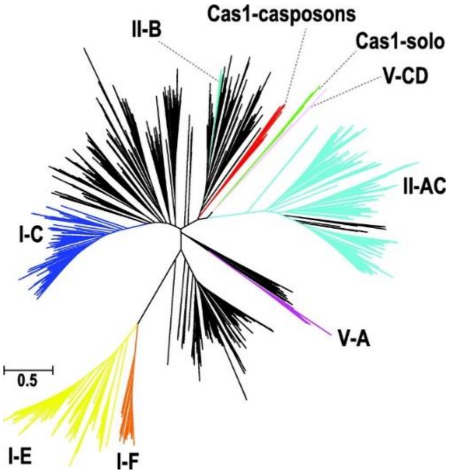

Figure 11. *Makarova, K. S., Wolf, Y. I. & Koonin, E. V. (2018). Classification and nomenclature of CRISPR-Cas systems: where from here? The CRISPR journal, 1 (5): 325-336.*

CRISPR-Cas phylogenetic tree based on Cas1 similarity.

Based on research of Cas1 activity, Cas1 proteins might be mobile elements - so called casposons. Purified Cas1 casposase can integrate specific sequences into random target sites, both short oligonucleotides with inverted repeat sequences, and mini-casposons (Hickman & Dyda, 2015). Cas1 proteins are asymmetrical homodimers with each monomer having an N-terminal β-sheet domain and C-terminal α-helical domain (James Nunez et.al.) In CRISPR systems, Cas1 protein is a metal-depended DNA nuclease, that possess endonuclease activity, and is needed for the process of viral DNA disintegration. The removal of the gene from the genome in *E. coli* resulted in increased sensitivity to DNA damage and chromosomal segregation (Makarova & Koonin, 2015).

Additionally, Cas1 has been linked to physical and genetic interactions with key components of DNA repair systems, implicating its involvement in DNA repair mechanisms (Kim et al., 2013).

Cas1 is the most preserved Cas protein in the CRISPR genome, and for quite a long time was used for classification of CRISPR systems (figure 11) before they were being classified by CRISPR locus architecture. (Makarova et al., 2017) Crystal structure of Cas1 is shown in figure 12 as a part of Cas1-Cas2 complex.

Cas2

Cas2 proteins are symmetrical homodimers with a core ferredoxin fold. Active site mutants of Cas2 can acquire spacers, indicating a non-enzymatic role of Cas2 during CRISPR-Cas immunity (James Nunez et al). Different homologues of Cas2 have shown RNase activity, specific to U-rich regions, and double stranded DNase activityl. Most important role of Cas2 proteins in CRISPR-systems is as a subunit of Cas1-Cas2 complex (Makarova & Koonin, 2015). Crystal structure of Cas2 is shown in figure 12 as a part of Cas1-Cas2 complex.

Cas1-Cas2 complex

The initial stage of CRISPR-Cas immunity involves the integration of foreign DNA spacer segments into the host genomic CRISPR locus. Two CRISPR-associated proteins are required for the acquisition step of adaptation, in which fragments of foreign DNA are incorporated into the host CRISPR locus – Cas1 and Cas2. Cas1 and Cas2 are the only proteins conserved among almost all CRISPR-Cas systems (Nuñez et al., 2015).

Purified CRISPR Cas1-Cas2 complex can integrate protospacers into CRISPR locus, indicating that the two proteins together form a DNA integrase (Hickman & Dyda, 2015).

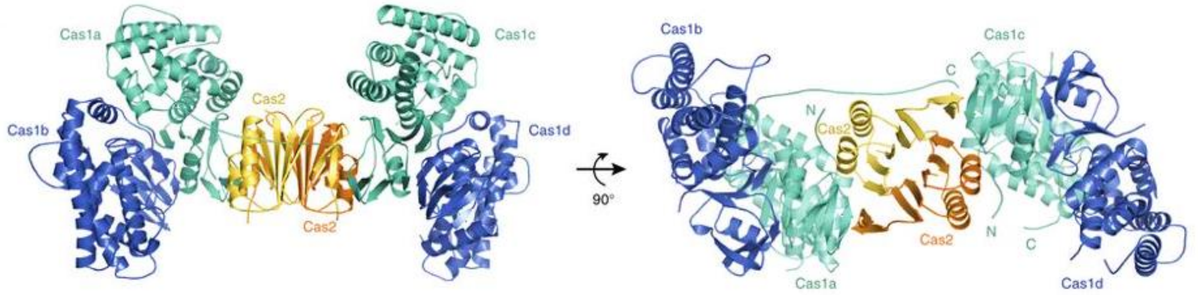

Figure 12. *Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W. & Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. Nature structural molecular biology, 21 (6): 528.*

Crystal structure of Cas1- Cas2 complex - Cas2 dimer (yellow and orange) and two Cas1 dimers blue and teal.

Cas1-Cas2 complex is an asymmetrical complex consisting of two Cas1 dimers (Cas1a-b and Cas1c-d) and a Cas2 dimer. Cas1a and Cas1c make contact with the Cas2 dimer, but no contacts between Cas1b or Cas1d and Cas2 were observed. The interface between Cas1 and Cas2 consists of hydrogen-bonding, electrostatic and hydrophobic interactions (Nuñez et al., 2014).

Cas3

Cas3 proteins have two domains - Cas3’ helicase and Cas3’’ HD nuclease. Cas 3 proteins are nuclease helicases with single strand DNA-stimulated ATPase activity coupled to unwinding of DNA-DNA and RNA-DNA duplexes. Cas3’ is involved in delivery of nuclease activity to CASCADE complex (see crRPNs). Cas3’’ HD domain has ATP-independent nuclease activity that targets ssDNA. Cas3 is essential for crRNA-guided DNA interference of CRISPR systems (Sinkunas et al., 2011).

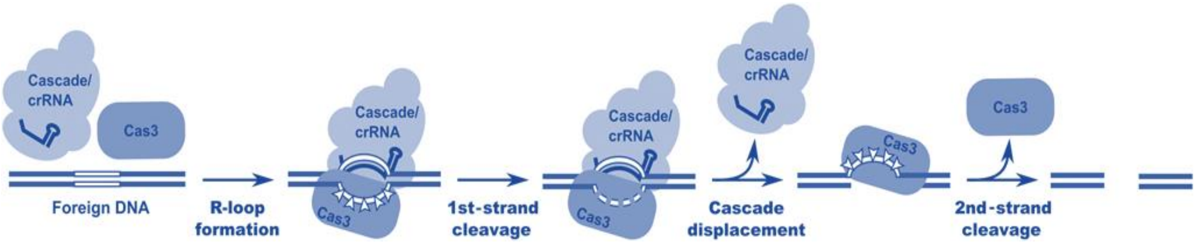

Figure 13. *Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. & Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. The EMBO journal, 30 (7): 1335-1342.*

Representation of Cas3 recruitment by CASCADE surveillance complex for directed DNA cleavage.

Cas3 proteins are involved in cleavage of the invading DNA (figure 13). In CRISPR-Cas systems, Cas3 is a motor protein responsible for nuclease activity of CASCADE-crRNA complex (Makarova, 2015).

Cas4

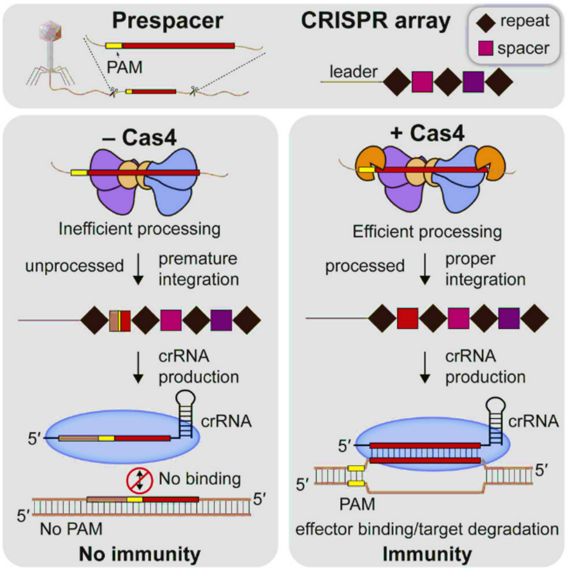

Figure 14. *Lee, H., Zhou, Y., Taylor, D. W. & Sashital, D. G. (2018). Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. Molecular cell, 70 (1): 48-59. e5.*

Cas4 role in CRISPR systems, where the protein is taking part in incorporating spacers into CRISPR array for viral immunity.

Cas4 is a nuclease with three-cysteine C-terminal cluster; it possesses 5'-3' ssDNA exonuclease activity and is a reverse transcriptase (Makarova et al., 2017). Cas4 plays a role in acquiring of new viral DNA sequences and incorporating those into the host genome for further crRNA production (figure 14). Cas4 has a RecB domain(a nuclease); Some Cas4 variants have shown exonuclease activity *in vitro* and are characterized as 5'-3' single strand DNA exonucleases (Lee et al., 2018).

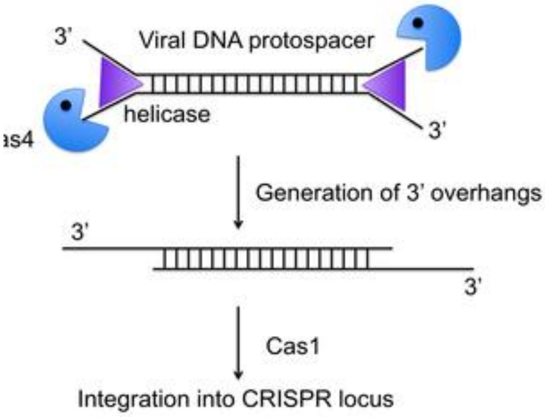

Figure 15. Zhang, J., Kasciukovic, T. & White, M. F. (2012). The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster. PLoS One, 7(10): e47232.

Process of spacer generation by Cas4 for integration into CRISPR locus.

In their study Zhang *et. al.* (Zhang et al., 2012) have shown that Cas4 protein families perform as 5'-3' DNA exonucleases *in vivo* too. Based on evidence that Cas4 can form complexes with Cas1 and Cas2 the group suggested that the activity of Cas4 is dependent on its partner proteins, one possible role of Cas4 is generating recombinogenic 3'-5'-ssDNA overhangs in duplex DNA protospacers selected for incorporation into the genome (figure 15) (Zhang et al., 2012).

Cas4 has ancestral connection to Csa1, which is a Cas protein specific to archaea. It has been suggested to rename Csa1 to Cas4'. Cas4 and Csa1 has shown connection to Cas1 and Cas2 in some organisms, leading to an assumption that Cas4 and Csa1 are participating in spacer acquisition pathway (Plagens et al., 2012).

Cas5

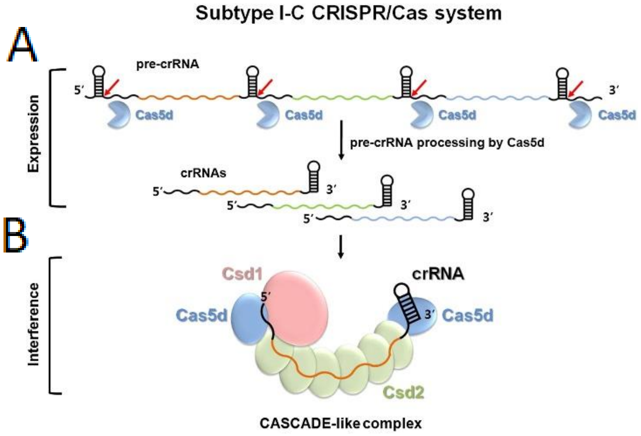

Figure 16. *Nam, K. H., Kurinov, I. & Ke, A. (2011). Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity. Journal of Biological Chemistry, 286 (35): 30759-30768.*

Cas5 role in CRISPR-Cas expression and interference stages of crRNA maturation.

Cas5 is involved in interactions with large subunit of the CASCADE surveillance complex, Cas7 and binding the 5'-handle of crRNA (figure 16A). In subtype I-C Cas5 replaces Cas6 functions, and performs as an endoribonuclease (Barrangou et. al, 2007).

Cas5 plays an important role in as pre-crRNA processor in crRNA maturation. Protein cleaves pre-crRNA into smaller crRNAs during expression stage. Additionally, together with Cas7* and Cas8*, Cas5 forms CASCADE-like interference complex, suggesting further crRNA-mediated DNA silencing by the complex (figure 16B). Cas5 CASCADE-like complex shows higher specificity for the repeat region of crRNA than CASCADE complex itself. In an experiment performed by Mohanraju *et.al.* (Mohanraju et al., 2016) alterations of loop sequence of crRNA repeat region had little effect on Cas5 CASCADE-like complex, but the same changes disrupted formation of CASCADE complex. Mohanraju *et.al.* (Mohanraju et al., 2016) suggested that increase in specificity is mediated by presence of either Cas7 or Cas8 in the complex. Based on Cass *et.al.* research, it is suggested that Cas8 is responsible for the specificity increase (Cass et al., 2015).

*In newest CRISPR-Cas nomenclature, Csd1 and Csd2 (shown in figure 16) are renamed to Cas7 and Cas8 respectively.

Cas6

In a study of Cas6 performed by Carte *et.al.* (Carte et al., 2008) has been identified as an endoribonuclease, belonging to RAMP protein family (nucleases containing G-rich regions). Cas6 functions have been tested both *in vivo* and *in vitro*. Cas6 is taking a part in crRNA maturation by cleaving precursor CRISPR RNAs within the repeat sequences. The protein is able to catalyze site-specific cleavage within each repeat, and release individual invader targeting units. The process starts with Cas6 binding to a 5'- handle of pre-crRNA, and further cleaving in the 3'- handle of CRISPR repeat RNA. Cas6 cleavage products undergo further processing in order to generate smaller mature psiRNAs (RNA polymerase III-based plasmid that produces short RNAs (figure 17) (InvivoGen)) (Wang et al., 2011).

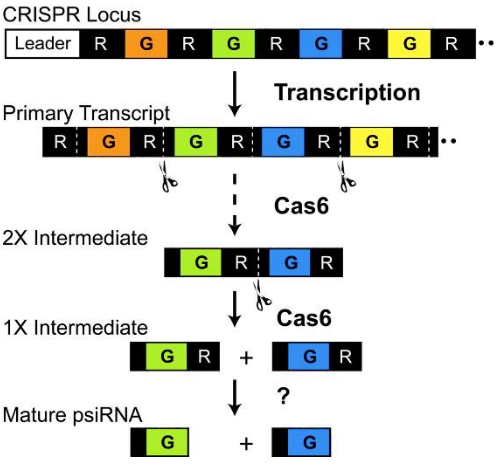

Figure 17. *Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M. & Terns, M. P. (2010). Binding and cleavage of CRISPR RNA by Cas6. Rna, 16 (11): 2181-2188.*
CrRNA maturation by Cas6.

Cas7

Cas7 is another subunit of CASCADE complex, participating in the process of binding crRNA. The protein is often present in several copies in the CASCADE complexes, and is responsible for the formation of the helical groove. Cas7 has a stabilizing role in CASCADE – three points of contact stabilize the complex: one between Cas7 and the guide region of crRNA, and two more involved in conserved protein-protein contacts between Cas7 subunits (figure 18B).

Cas7 crystal structure revealed "right hand" formation consisting of fingers, palm and thumb domains (figure 18A).

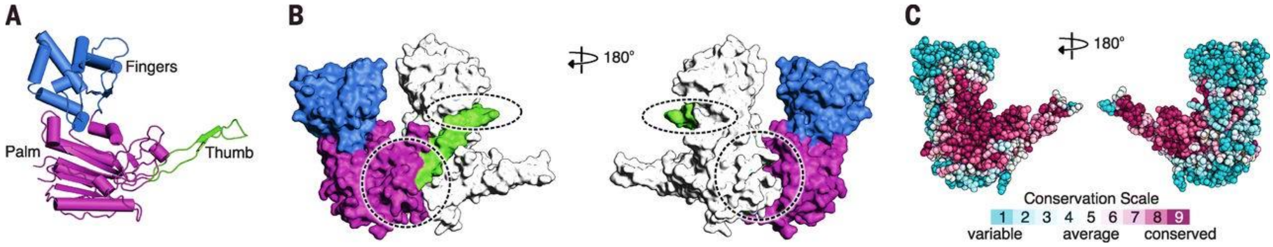

Figure 18. *Mulepati, S., Héroux, A. & Bailey, S. (2014). Crystal structure of a CRISPR RNA guided surveillance complex bound to a ssDNA target. Science, 345 (6203): 1479-1484.*
Crystal structure of Cas7.

Thumb and finger domains are involved in structural stabilization. The thumb extends to neighboring Cas7 proteins ensuring further stabilization of Cas7 subunits in the complex by interactions with the fingers. The palm is a conserved area (figure 18C) that contains the modified RNA-recognition motif (RRM) and is responsible for crRNA-Cas7 interactions (Mulepati et al., 2014).

Cas8

Large subunit of Cascade complexes, involved in PAM recognition. Cas8 interacts physically with Cas5-Cas7-crRNA complex, and it has been showed that Cas8 responds to PAM sequence when binding to nucleic acids. There are two residues of Cas8, called Cas8' and Cas8''. Cas8'' has yet not shown any significant homology to any protein in database searches.

Functions of the Cas8' proteins have been tested for Cas8' alone, and in complex with Cas5-Cas7. Isolated Cas8' has proven to be able to form complexes with R-loop substrates, and predicted PAM sequence. Cas8' in assays together with Cas5– Cas7 proteins have shown signs of a distinct binding complex, leading to a suggestion that Cas8' probably can adjust how Cas5–Cas7 can precisely assemble on the substrate, thereby controlling its aggregation. Cas8' has shown signs of single strand RNA nuclease activity *in vitro*.

Deactivating Cas8 *in vivo* caused interference in crRNA binding by CASCADE complex, deactivating Cas8 *in vitro* causes alterations in catalytic activity (Cass et al., 2015).

Cas9

The most used CRISPR-Cas protein – Cas9 is an RNA-guided DNA cleaving endonuclease that generates DSB in target sequences through base pairing to the CRISPR guide RNA. In CRISPR Cas9 system, tracrRNA forms a double stranded stem, which allows recruitment of Cas9 in order to perform DSB. Since guide RNA is easy programmable, achieving target specificity with Cas9 complexes is an easy task.

Cas9 has two nuclease domains, called HNH and RuvC-like domain. HNH domain can perform DNA cleavage on the complimentary strand, while RuvC-like domain targets non-complimentary strand and cleave it in site-specific manner (Sternberg & Doudna, 2015).

Cas9 is a multidomain protein. Apart from nuclease domains and crRNA-guided DNA interference and silencing it is involved in crRNA maturation. In presence of Cas9, ribonuclease RNase III is recruited to perform tracrRNA maturation, which results in mature crRNA (Jinek et al., 2012).

Since Cas9 has become a vital part of widely used genome editing tool complex, search for Cas9 variants has resulted in discovery of a variety of Cas9 homologues. Most notable difference between those Cas9 proteins is PAM sequence recognition. Usually, Cas9 nuclease is targeting 3'- NGG - 5' sequence, but a large number of Cas9 with alternative PAM has been discovered (table 2) (Komor et al., 2017).

Table 2. *Komor, A. C., Badran, A. H. & Liu, D. R. (2017). CRISPR-based technologies for the manipulation of eukaryotic genomes. Cell, 168 (1-2): 20-36. Short overview of alternative Cas9 proteins and their properties.*

Name	Construct size (nucleotides)	PAM sequence	Cleavage pattern, complimentary/non-complimentary strand break position
SpCas9	1368	5'- NGG -3'	18/17
FnCas9	1629	5'- NGG -3'	18/17
St1Cas9	1121	5'- NNAGAAW -3'	18/17
St3Cas9	1409	5'- NGGNG -3'	18/17
NmCas9	1082	5'- NNNNGATT -3'	22/21
SaCas9	1053	5'- NNGRRT – 3'	19/18
VQR SpCas9	1368	5'- NGA -3'	18/17
EQR SpCas9	1368	5'- NGAG -3'	18/17
VRER SpCas9	1368	5'- NGCG -3'	18/17
RHA FnCas9	1629	5'- YG -3'	18/17
KKH SaCas9	1053	5'- NNNRRT -3'	19/18

SpCas9 remains the most used analogue of Cas9 protein due to its well-known characterization, balance between PAM complexity and construct size. SpCas9s PAM sequence is well represented in human genome and occurs every 8-12 bp, making genome targeting quite easy, but at the same time increasing the chances of off-target activity of the complex (Hsu et al., 2013).

Cas10

Cas10 encodes a multidomain protein containing a palm domain, similar to that in cyclases and polymerases of the PolB family. Cas10 is the large subunit of effector complexes of type III systems, and is composed of two domains – CRISPR Palm polymerase and HD nuclease.

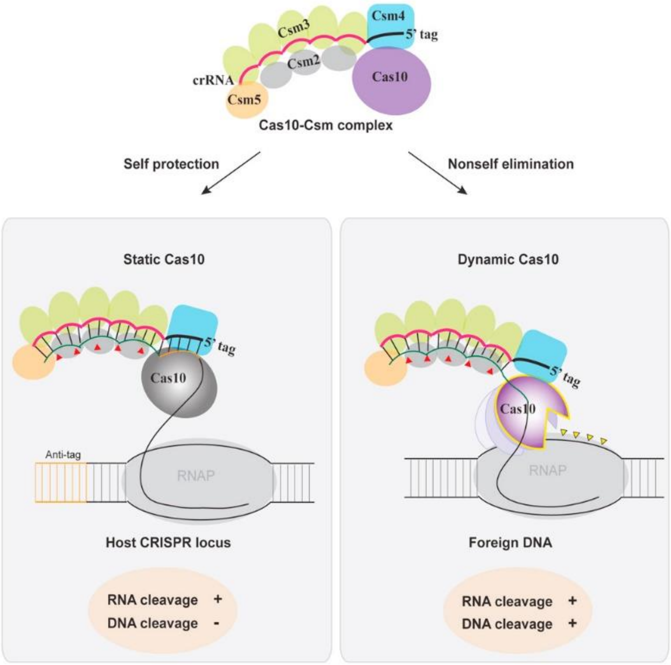

Figure 19. *Wang, L., Mo, C. Y., Wasserman, M. R., Rostøl, J. T., Marraffini, L. A. & Liu, S. (2019). Dynamics of Cas10 Govern Discrimination between Self and Non-self in Type III CRISPR-Cas Immunity. Molecular cell, 73 (2): 278-290. e4.*

Two different states of Cas10

HD nuclease domain of Cas10 is involved in crRNA biogenesis or targeting stage of CRISPR immunity. Palm polymerase domain functions and roles in CRISPR systems are still unknown.

In a study by Hatoum-Aslan *et.al*. (Hatoum-Aslan et al., 2014) deactivation of Palm polymerase domain in Cas10 resulted in CRISPR immunity systems failure, leading to suggestion that Palm domain might play either a structural role in Cas10 folding and stability or catalytic role in crRNA biogenesis or viral DNA targeting. Further testing showed that Cas10 plays a functional role in crRNA biogenesis, possibly DNA recognition and/or cleavage by sliding along the DNA and scanning for targets (Hatoum-Aslan et al., 2014).

In their newly published study Wang *et.al.* (Wang et al., 2019) confirmed assumption about Cas10 function in discrimination between self,- and invading DNA. Cas10 is a DNase/RNase responsible for DNA degradation. Cas10 is found in static state in CRISPR locus, but displaying conformational changes in presence of viral DNA and implying DNase functions, resulting in distinct behaviors (figure 19) (Wang et al., 2019).

Cas11

Cas11 is another protein in the CRISPR-Cas family, also known as SS – small subunit of CASCADE surveillance complex, often fused to large subunit (Shah et al., 2019; Majumdar et al., 2015).

The protein has shown endodeoxyribonuclease activity and can bind DNA and metal ions (manganese and/or magnesium) (UniProtKB, 2019). It is possible that Cas11 is participating in maintaining CRISPR repeat elements. In CASCADE surveillance complex, Cas11 has protein-protein interactions with Cas7 indicating its part in stabilization of the complex (Majumdar & Terns, 2019).

There has been an attempt to determine Cas11 functions based on crystal structure of the protein; however, the results showed that the interactions of Cas11 with the CASCADE were below detection limit of the biophysical techniques used by the research group, and its functions and role in CRISPR-systems remain unknown (Reeks et al., 2013).

Cpf1

Cas12, better known as Cpf1 is a subtype V-A, class II CRISPR-Cas nuclease, which has been used as a programmable genome editing tool. Cpf1 is a single-RNA-guided enzyme; It recognizes thymidine-rich PAM motifs (table 3) and can perform both DNA and RNA breaks (Zetsche et al., 2015; Strohkendl et al., 2018).

Cpf1 is an alternative to type CRISPR-Cas9 systems that performs at even better rate than Cas9, due to lower cytotoxicity and tolerance for mismatches, which greatly reduces off-target activities (table 4). In several editing experiments (Kim et al., 2016; Kim et al., 2017; Kleinstiver et al., 2016) genome editing with Cpf1 showed little to none mismatches during protein activity. Unlike Cas9, Cpf1 can process its own precursor crRNA, and does not require additional proteins like RNase III. Additionally, Cpf1 is smaller, shows RNase activity, some Cpf1 homologues(subtype V-B and V-E) do not require tracrRNA (Zaidi et al., 2017). Another remarkable difference compared to CRISPR-Cas9 is PAM site recognition. While CRISPR-Cas9 PAM is located downstream of Cas9 DSB site, Cpf1 PAM is located upstream of its cleavage site (figure 20) (Rusk, 2019).

Table 3. *Komor, A. C., Badran, A. H. & Liu, D. R. (2017). CRISPR-based technologies for the manipulation of eukaryotic genomes. Cell, 168 (1-2): 20-36. Short overview of alternative Cpf1 proteins and their properties.*

Name	Construct size (nucleotides)	PAM sequence	Cleavage pattern, complimentary/non-complimentary strand break position
AsCpf1	1307	5’- TTTN -3’	24/19
LbCpf1	1228	5’- TTTN -3’	24/19

Table 4. *Short comparison of Cas9 and Cpf1 properties. Based on research by (Sternberg & Doudna, 2015; Kleinstiver et al., 2016; Strohkendl et al., 2018)*

Feature	Cas9	Cpf1
Guide RNA	sgRNA (tracrRNA+crRNA)	crRNA
Guide RNA processing	RNase III	Cpf1
tracrRNA	Present	Not needed*
Recognized DNA target length	18-22nt + PAM (3-8nt)	24nt + PAM (4nt)
Guide RNA length	~100nt	~42nt
Nuclease domain	RuvC-like + NHN	RuvC
Cleavage pattern	Blunt end	Staggered 5’-overhang
PAM sequence	Variable (see table 2)	5’- TTTN -3’
PAM location	2-3 bp downstream of DSB	18-23 bp upstream of DSB
PAM site preservation after DSB	Destroyed	Preserved
Multiplex genome editing	Yes	Yes
Targeting efficiency	High	Slightly lower
Off-target effects	Moderate	Low
Mismatch tolerance	1-6bp	Variable

**Subtypes V-B and V-E use tracrRNA*

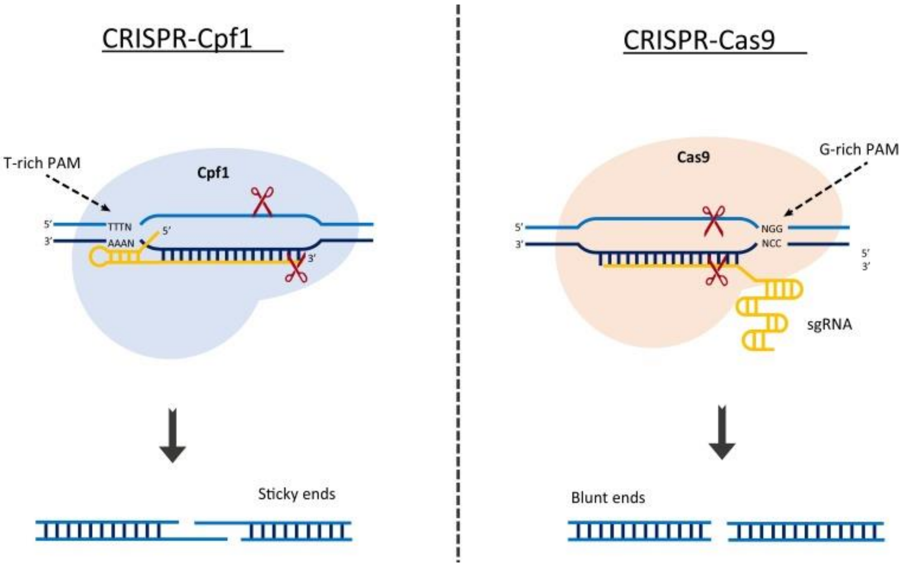

Figure 20. *Zaidi, S. S.-e.-A., Mahfouz, M. M. & Mansoor, S. (2017). CRISPR-Cpf1: a new tool for plant genome editing. Trends in plant science, 22 (7): 550-553. Process of DSB by Cpf1 compared to Cas9.*

CRISPR-Cpf1 is a smaller complex compared to CRISPR-Cas9, meaning easier delivery of the complex to the cell. Regarding mismatches, unlike Cas9 Cpf1 does not tolerate double mismatches between guide RNA and target-site. Only exception is the 3’-end of Cpf1 crRNA, where double mismatches are tolerated between positions 19-24, and single mismatches are tolerated at positions one, eight and nine. In a study by Kleinstiver *et.al.,* deletion of four to six base pairs at the 3’-end of Cpf1 crRNA had no effect on Cpf1 targeting ability (Kleinstiver et al., 2016). One of the main advantages of Cpf1 compared to Cas9 is low off-target activity. In an experiment by Kim *et.al*. (Kim et al., 2016) Cpf1 showed six off-target sites for LbCpf1 and 12 for AsCpf1. In contrast, Cas9 had off-target activity on over 90 sites. In the same experiment Kim *et.al.* were able to demonstrate that preassembled, recombinant Cpf1 had no off-target activity at all (Kim et al., 2016).

Cas13

Cas13 is a subtype VI CRISPR-Cas ribonuclease. Four distinct types of Cas13 protein are discovered (figure 21A) – Cas13a (subtype VI-A), Cas13b (subtype VI-B), Cas13c (subtype VI-C), Cas13d (subtype VI-D). All four variants possess ability to perform crRNA guided targeting of RNA provided by HEPN domains. Cas13 is capable of generating mature crRNAs (short and long repeats, spacers) by cleaving own CRISPR-array (Smargon et al., 2017).

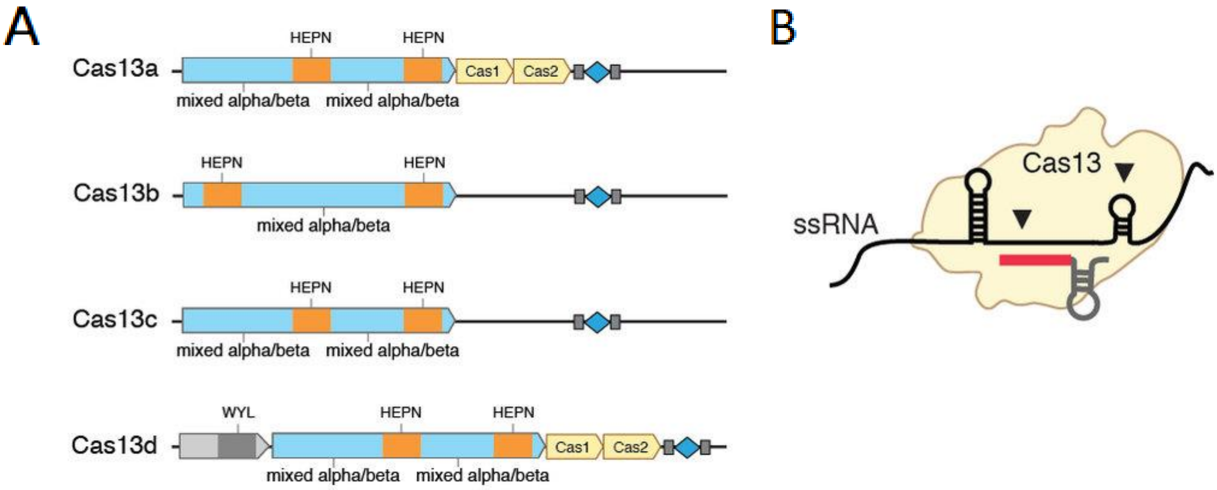

Figure 21. *Abudayyeh, O., Gootenberg, J. (2019). Cas13 — Zhang Lab. Zhang Lab. Available at: https://zlab.bio/cas13 (Accessed 13 May 2019).*
A: Locus architecture of four known Cas13 proteins.
B: Process of single stranded RNA cleavage by RNA guided Cas13.

Cas13 is a powerful platform for RNA manipulation, showing resemblance to Cas9 and Cpf1 proteins. CRISPR-Cas13 shows programmable RNase activity, both specific and non-specific, allowing *in vivo* targeting applications in mammalian and plant cells. Cas13 has been used for *in vivo* applications such as RNA knockdown, RNA editing and nucleic acid detection.

Cas13 mediated RNA knockdown cleaves targeted transcripts by relying on dual HEPN domains of Cas13 subtypes (figure 21B, HEPN nuclease domains marked as triangles), and can be used for gene expression alterations by degrading mRNA (messenger RNA). Alteration efficiency varies depending on Cas13 systems, and shows up to 90-95% knockdown efficiency.

RNA editing by Cas13 allows temporal alternation of genetic transcripts by the REPAIR (RNA Editing for Programmable A to I Replacement) system. REPAIR system works by fusing adenosine deaminases to Cas13 complexes for further RNA alternation by the complex.

In vitro nucleic acid detection by Cas13 allows specific single-nucleotide distinction in target sequence. This process can be used to amplify signals from molecules, even at extremely low concentrations (Abudayyeh and Gootenberg, 2019).

Secondary CRISPR-Cas proteins

Secondary Cas proteins are a group of CRISPR-associated proteins. These proteins are often homologues of essential CRISPR-Cas proteins or are involved in CRISPR-Cas protein regulations. Large number of secondary Cas proteins has been found since discovery of CRISPR systems (see Appendix II), but their functions are often unknown. Majority of those proteins are involved in CRISPR surveillance complexes (Makarova et al., 2011b).

Csb1, Csb2, Csb3

Csb proteins belong to I-U subtype of CRISPR-Cas systems (Makarova et al., 2017). Function unknown.

Csc1

Csc1 protein has a G-rich region and is a part of Cas5 group. The protein is a part of RAMP protein superfamily. Even though that has not been officially confirmed for this particular protein, RAMP superfamily proteins containing a G-loop are usually associated with RNA binding and catalysis of crRNA processing RNases. (Makarova et al., 2011a; Makarova et al., 2011b)

Csc2

Csc2 belongs to the RAMP superfamily (Makarova et al., 2011a). Function unknown.

Csn2

Csn2 protein as a double-stranded non-specific DNA-binding protein regulated by the presence of Ca2+. Csn2 is arranged in a tetrameric ring structure, composed of an α/β domain and an α-helical domain (Nam et al., 2011).

Csx1

Csx1 is a type III-B CRISPR protein often located in close proximity to Cmr complexes, but is not a part of the complex. Csx1 shows properties of a metal-independent, temperature-dependent ssRNA nuclease that cleaves selectively after adenosine repeats (Sheppard et al., 2016).

Csx3

Csx3 is a type III- B CRISPR-associated RNase (Yan et al., 2015).

Csx10

Csx10 protein, a fusion of Cas5 and Cas7, has two RAMP-like RRM domains that have a G-rich loop each. Some Csx10 variants are components of CASCADE system, fused to small subunit and Cas7 group RAMP proteins (see Cas5 and Cas7 for functional description) (Makarova et al., 2011b).

Csx15/20

Csx15/20 is a type III CRISPR peptidase protein linked to crRNA maturation. Wrongly catalogued as type I-U in CDD database (Shah et al., 2019).

Csx19, Csx24

Csx19 and Csx24 are likely type III CRISPR-associated proteins (Shah et al., 2019). Function unknown.

Csx14, Csx16, Csx17, Csx18

Csx14, Csx16, Csx17 and Csx18 proteins belong to subtype III-U (Makarova et al., 2017). Functions unknown.

Csx26

Csx26 is a type III CRISPR-associated putative small subunit protein (Makarova et al., 2017). Function unknown.

Csx27

Accessory protein. Represses Cas13 mediated RNA interference (Smargon et al., 2017).

Csx28

Accessory protein. Enhances Cas13 mediated RNA interference (Smargon et al., 2017).

CsaX

CsaX is a subtype III-U CRISPR-associated protein (Makarova et al., 2017). Function unknown.

crRNPs CRISPR surveillance complexes

CrRNPs –CRISPR ribonucleoproteins are multiprotein complexes composed of Cas protein subunits. Ribonucleoprotein complexes often show either RNase activity and/or DNase activity and have similar functions as multidomain proteins of CRISPR-Cas class II.

In a newly published work, Dolan *et.al.* have performed CASCADE-Cas3 induced genome editing in human embryonic stem cells. Dolan *et.al* show that genome editing by type I CRISPR systems is crRNA guided and programmable. This could be an important milestone in genome engineering using programmable nucleases. Introducing class I CRISPR systems as yet another group of programmable nucleases disproves common belief that class I CRISPR systems are not suitable for genome engineering (Dolan et al., 2019).

CASCADE surveillance complex

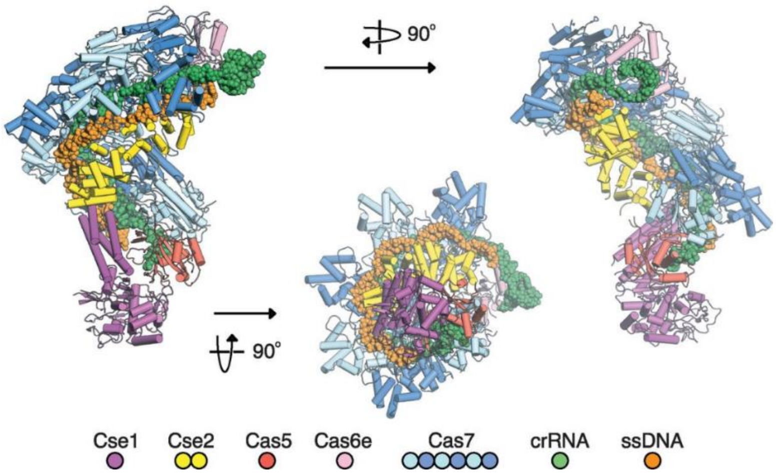

Figure 22. *Mulepati, S., Héroux, A. & Bailey, S. (2014). Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. Science, 345 (6203): 1479-1484.*
Structural composition of CASCADE surveillance complex, composed of 11 Cas subunits.

CASCADE - Clustered Regularly Interspaced Short Palindromic Repeat – associated Complex for Antiviral Defense is a large protein complex that is responsible for response to a viral nucleic acid infection. CASCADE complex is a type I-E surveillance complex, composed of 11 subunits of five Cas-proteins: Cse1, Cse2, Cas7, Cas5 and Cas6e, and a 61-nucleotide crRNA (figure 22). The body of the complex is formed by six Cas7 subunits, wrapped around the crRNA, and Cse2 dimer. Cas6 and the 3'-end of crRNA are connected to a Cas7 protein at the ”head” of the complex, Cas5 is positioned at the 5'-end. Cse1 is positioned at the N-terminal, while Cse2 is positioned at the C-terminal of the protein.

CASCADE mechanism is similar to CRISPR-Cas9. CASCADE recognizes a DNA-sequence as foreign if the sequence has one or more regions, complimentary to crRNA of the complex, and a protospacer adjacent motif (PAM sequence). All CRISPR-arrays lack PAM sequence which allows the complex to discriminate viral DNA from self DNA (Lintner, 2011).

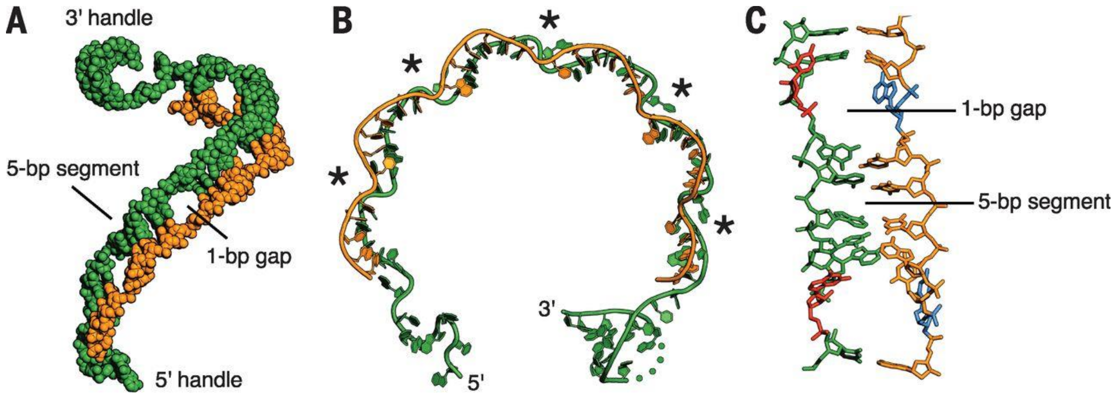

Figure 23. *Mulepati, S., Héroux, A. & Bailey, S. (2014). Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. Science, 345 (6203): 1479-1484. Ribbon-like structure formed when CASCADE complex binds to viral DNA.*

When PAM sequence on a viral DNA sequence has been detected, the complex will initiate binding and eventually forming of an unidirectional R-loop for further target cleavage. CASCADE surveillance complex forms a bond with the viral DNA, but the two strands of guide-target hybrid do not form helix when the bonds of R-loop are formed. Instead, an underwound ribbon-like structure is being formed (figure 23B): A so-called kink occurs every sixth base pair in the backbone of both the target and guide strands, and the nucleotides are then rotated approximately 90° in opposing directions, forming a five base pair segment and one base pair gap (figure 23 A and C).

Mutations in disrupted nucleotides (one base pair gap) does not affect binding efficiency of the complex, while mutations at positions one to five of the five base pair segment greatly reduce affinity (Mulepati, 2014).

→ Cse1

Cse1, found as a four-helix bundle in CASCADE, is often located near PAM sequence. It is suggested that Cse1 may have a role in stabilizing target DNA strand by making direct bonds with the phosphate backbone of the DNA. Mutation of the Cse1 bundle results in negative consequences for cleavage of the target DNA by Cas3 and CASCADE. It is critical for Cas3 recruitment and DNA cleavage (Jiang & Doudna, 2015).

→ Cse2

Cse2 is a small α-helical protein that forms head-to-tail-dimer with another copy of Cse2 in CASCADE complex. The protein has RNA recognition motif and shows affinity for nucleic acids (Ebihara, 2006; Agari et.al, 2008).

→ Cse5

Makes contact with 5’ end of crRNA. Together with Cas7 Cse5 forms six β-hairpins that interrupt crRNA-ssDNA bindings. Each sixth base pair of the unwound viral DNA is left unpaired and flipped outwards; Cse5 prevents crRNA from binding to viral DNA to maintain the stability of the complex (Jiang & Doudna, 2015).

Csy surveillance complex

Csy surveillance complex is subtype I-F RNP. All CRISPR-Cas systems use RNA guides, or so-called crRNA, which forms crRNA-guided surveillance complexes in combination with CRISPR proteins (figure 24). Csy1-4 Cas-proteins function by recruiting crRNA to perform complimentary base pairing with protospacers – or invading DNA sequences. In case viral DNA is detected, and the complex possess a complimentary crRNA the Csy surveillance complex will bind to the viral DNA and recruit the Cas3 nuclease-helicase for phage genome degradation. (Bondy-Denomy et al., 2015)

Complex can also recruit Cas6 endoribonuclease thru Csy3-Cas6 interactions, most probably for guide RNA maturation. All four Csy proteins have shown *in vivo* interactions without requirement of any other Cas, Csy or mature crRNA (Richter et al., 2012).

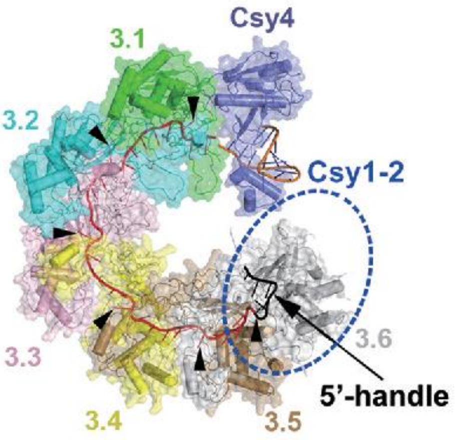

Figure 24. *Peng, R., Xu, Y., Zhu, T., Li, N., Qi, J., Chai, Y., Wu, M., Zhang, X., Shi, Y. & Wang, P. (2017). Alternate binding modes of anti-CRISPR viral suppressors AcrF1/2 to Csy surveillance complex revealed by cryo-EM structures. Cell research, 27 (7): 853.*

Structural composition of Csy surveillance complex.

→ Csy1

Csy1 is type I-F large subunit. The protein interacts with DNA and/or RNA, and might be a polymerase. Csy1 might be homologous to Cse1 (see Cse1). Together with Csy2, Csy1 mediates Csy3-Cas6 interactions (Makarova et al., 2011b; Richter et al., 2012).

→ Csy2

Csy2 and Csy1 form a subcomplex that has a role in binding Csy3 for further stabilization of the Csy surveillance complex. They might have an additional role in distinguishing target from non-target genomes (Marraffini & Sontheimer, 2010; Richter et al., 2012).

→ Csy3

Csy3, present in a dimer form in the complex, is responsible for making interactions with Csy1 and the nuclease. Csy3's role is stabilizing the backbone of the complex (Richter et al., 2012).

→ Csy4

In type I-F CRISPR–Cas system, the Csy4 protein is a CRISPR-specific endoribonuclease. Csy4 can bind to and cleave repeat sequences in the pre-crRNA, and is associated with the 3' end of the mature crRNA. Csy4 interacts with Csy1, Csy2 and Csy3 proteins to form a Csy1-Csy2-Csy3-Csy4 surveillance complex (Bondy-Denomy et al., 2015).

Csm surveillance complex

Csm complex is a subtype III-A surveillance complex, analog to CASCADE. Csm complex is composed of a multisubunit protein and small RNAs adapted for alteration of nucleic acids (both RNA and DNA) (figure 25). It has been used to perform studies of phage genes of unknown functions by performing blocking of plasmid transfer and phage infection (Tamulaitis et al., 2017).

Five Cas proteins – Cas10 (Csm1), Csm2, Csm3, Csm4 and Csm5 form an interference complex that has ability to target invading DNA (figure 25). Csm complex has shown signs of plasmid conjugation abilities by targeting and degrading DNA with help of crRNA. Cas10 (Csm1), Csm3 and Csm4 form a stable sub complex, presumably targeting RNA and/or DNA (Rouillon et al., 2013).

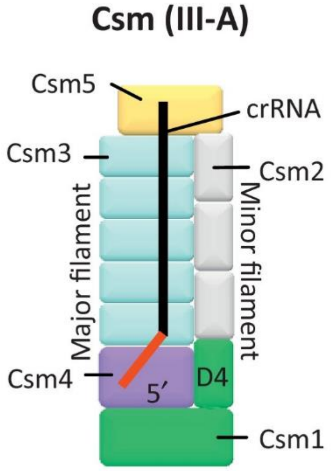

Figure 25. *Tamulaitis, G., Venclovas,*

CRISPR-Cas immunity: major differences brushed aside. Trends in microbiology, 25 (1): 49-61.

Structural composition of Csm surveillance complex.

→ Csm2

Csm2 is a small subunit of the Csm complex. Csm2 can exist in both monomeric and dimeric form *in vivo*, but is a dimer *in vitro* as a part of Csm complex. Possibly participating in in the binding of target oligonucleotides. Crystal structure analysis shows that Csm2 might be Cse2 and Cmr5 functional analogue, despite low structural and sequence similarity (Venclovas, 2016).

→ Cas10-Csm3-Csm4 subcomplex

Together Cas10-Csm3-Csm4 form a subcomplex (figure 26) that has been analyzed and showed ssRNA binding abilities in a non-sequence-specific manner (Walker et al., 2016). Cas10-Csm3-Csm4 complex can bind ssRNA; Both Csm3 and Csm4 show ssRNA binding activity and have been identified as orthologues (both proteins that have same ancestor gene). Additionally, it has been proven that Csm4 is responsible for the single strand RNA binding functions of the complex. Cas10 has a binding role in the complex, making interactions with Csm3 and Csm4 (Numata et al., 2015).

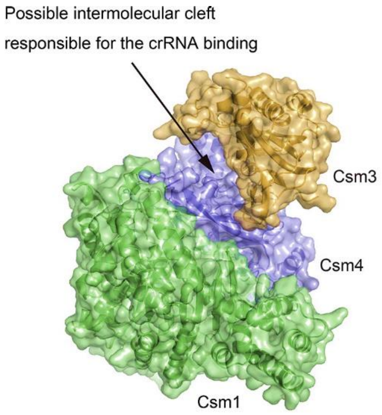
Possible intermolecular cleft responsible for the crRNA binding
Csm3
Csm4
Csm1

Figure 26. *Numata, T., Inanaga, H., Sato, C. & Osawa, T. (2015). Crystal Structure of the Csm3–Csm4 Subcomplex in the Type III-A CRISPR–Cas Interference Complex. Journal of molecular biology, 427 (2): 259-273.*

Crystal structure of Cas10-Csm3-Csm4 subcomplex.

→ Csm3

Csm3 is involved in Cas10-Csm3-Csm4 subcomplex, and structurally resembles Csm4 protein (despite low sequence similarity). Csm3 is a ruler protein that measures six-nucleotide increments and is involved in formation of helical backbone in the complex. Csm3 can bind ssRNA in sequence-independent manner, and at multiple sites. Csm3 may be involved in spacer binding, since it can form a bond with 5’ region of spacer sequence. Additionally, crystal structure of Csm3 shows structural resemblance to Cas7, which shares many of the same functions as backbone formation and crRNA maturation. Csm3 has a ferrodoxin-like fold, which has a stabilizing role in the complex structure (Numata et al., 2015).

→ Csm4

Csm4 is a part of Cas10-Csm3-Csm4 subcomplex, and is responsible for ssRNA binding of the complex. Csm4 can interact with 5’-end of crRNA in the context of Cas10-Csm3-Csm4 complex. Additionally, Csm4 has two ferrodoxin-like folds, and together with Csm3 plays a stabilizing role in the complex structure by making interactions with Cas10 (Numata et al., 2015).

→ Csm5

Csm5 is a large subunit of the Csm complex, required for crRNA maturation (Rouillon et al., 2013).

→ Csm6

Studies of Csm6 show that the protein is an ssRNA-specific endoribonuclease that forms a dimer. Csm6 performs RNA cleavage by two nuclease domains – N-terminal CARF, and C-terminal HEPN. Csm6 works in collaboration with Csm complex, and is probably recruited by the complex to perform RNA cleavage (Niewoehner & Jinek, 2016).

Csf surveillance complex

Csf surveillance complex is type IV crRNP. Özcan *et.al.* (Özcan et al., 2019) study of complex revealed functions of Csf1-5 proteins, but the group was unable to determine Csf crRNP function as a complex.

→ Csf1

Csf1 is large subunit of Csf surveillance complex. Cas8 analog (Özcan et al., 2019).

→ Csf2

Csf2 is a type IV CRISPR-associated protein that acts as a helical backbone in Csf surveillance complex, and is a paralogue of Cas7 (see Cas7) (Özcan et al., 2019).

→ Csf3

Csf3 is a Cas5 paralogue and has the same functions as Cas5 (see Cas5) (Özcan et al., 2019).

→ Csf4

Little is known about Csf4, other than Csf4 possess helicase properties, and is a signature gene for type IV CRISPR-Cas systems (Özcan et al., 2019).

→ Csf5

Csf5 is a type IV CRISPR-Cas protein, that is responsible for crRNA maturation in type IV CRISPR-Cas system. Furthermore, Csf5 is a crRNA endonuclease that generates an unusual 5’-terminal repeat tag of seven nucleotides, and has been proven to be a Cas6 homologue (see Cas6 for functions) (Özcan et al., 2019).

Cmr surveillance complex

Cmr surveillance complex is a subtype III-B CRISPR RNP complex. This 12-subunit complex (figure 27) targets ssRNA with help of Cmr1-6 Cas proteins. Complex mediates transcription-dependent silencing *in vivo* and RNA-activated cleavage *in vitro* (Taylor et al., 2015).

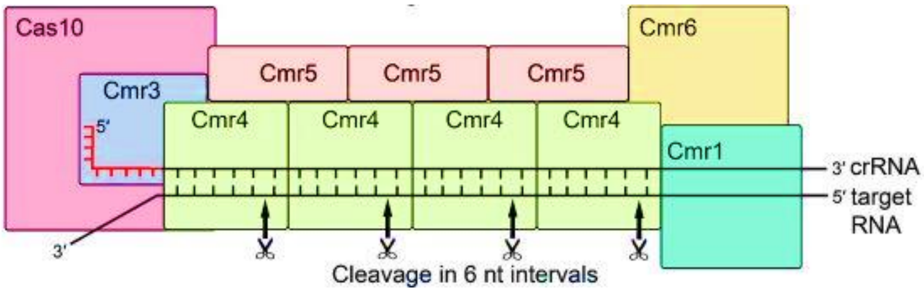

Figure 27. *Plagens, A., Richter, H., Charpentier, E. & Randau, L. (2015). DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. FEMS microbiology reviews, 39 (3): 442-463.*
Structural composition of Cmr surveillance complex.

→ Cmr1

Cmr1 is an activation module important for backbone RNA cleavage by Cmr complex. Cmr1 does not interact with other proteins in the complex unless crRNA is present in the complex, pointing to its crRNA-mediated activity. Cmr complex RNase activity is triggered when Cmr1 is activated by presence of crRNA (Li et al., 2017).

→ Cmr2-Cmr3 subcomplex

Cmr2 is another homologue of Cas10. This complex has two RAMP-like RRM domains with a G-rich loop each. G-rich loops play significant role in RNA binding and catalysis in the crRNA processing endoribonucleases. Together with Cmr3, it forms a Cmr2-Cmr3 complex that is responsible for recognition of 5'-handle of crRNA (figure 27) (Shao et al., 2013).

→ Cmr3

Cmr3 protein resembles Cas6, which is a crRNA processing endonuclease and a RAMP family protein. This protein is required for RNA-guided RNA cleavage and is critical for function of the Cmr2-Cmr3 subcomplex (Shao et al., 2013).

→ Cmr4

Cmr4 is a backbone unit of the Cmr complex. Cmr4 takes part in RNA binding properties of Cmr complex and functions as slicer protein (Zhu & Ye, 2014). Cmr4 performs RNA cleavage in six nucleotide intervals (figure 27) (Plagens et al., 2015).

→ Cmr5

Cmr5 is a globular α-helical protein involved in complex stability. Three adjacent Cmr5, together with Cas10 protein form double-helical body of Cmr complex (Plagens et al., 2015).

→ Cmr6

Cmr6 protein role in Cmr complex is disrupting base-pairing between crRNA and target RNA (Taylor et al., 2015).

CRISPR-Cas patents

Table 5. *Patents.google.com. (2019). Google Patents. Available at: https://patents.google.com/ (Accessed 14 May 2019).*
An overview of patents for CRISPR-Cas proteins.

Protein name	Patent code	Current assignee	Patent status	Application filled date	Application expiration date
Cas1	US10087431B2	University of California	Granted	2011-03-02	2033-01-06
Cas2	EP2825654B1	DuPont Nutrition Biosciences APS	Granted	2006-08-25	2026-08-25
Cas3	EP2336362B1	DuPont Nutrition Biosciences APS	Granted	2006-08-25	2026-08-25
Cas4	US10125361B2	Caribou Biosciences Inc	Granted	2016-05-19	2034-06-15
Cas5	US9410198B2	Caribou Biosciences Inc	Granted	2015-06-25	2034-03-12
Cas6	US9404098B2	University of Georgia Research Foundation Inc (UGARF)	Granted	2009-11-05	2029-11-05
Cas7	CN106834323A	Anhui University	Pending	2016-12-01	-
Cas8	EP2931898B1	Massachusetts Institute of Technology Broad Institute Inc	Granted	2013-12-12	2033-12-12
Cas9	US008697359B1	Massachusetts Institute of Technology, Broad Institute Inc	Granted	2013-10-15	2033-10-15
Cas10	US20180251787A1	University of Alabama (UA)	Pending	2018-03-02	-
Cas11	-	-	-	-	-
Cpf1	KR20180107155A	Benson Hill Biosystems Inc	Pending	2017-02-15	-
Cas13	WO2018170333A1	Massachusetts Institute Of Technology, Broad Institute Inc	Pending	2018-03-15	-
Csc1	-	-	-	-	-
Csc2	-	-	-	-	-
Csb1	-	-	-	-	-
Csb2	-	-	-	-	-
Csb3	-	-	-	-	
Csx1	US20170191047A1	University of Georgia Research Foundation Inc (UGARF)	Abandoned	2016-11-16	-
Csx3	-	-	-	-	-
Csx10	-	-	-	-	-
Csx14	-	-	-	-	-
Csx15/20	-	-	-	-	-
Csx16	-	-	-	-	-
Csx17	-	-	-	-	-
Csx18	-	-	-	-	-
Csx19	-	-	-	-	-
Csx24	-	-	-	-	-
Csx26	-	-	-	-	-
Csx27	US20170211142A1	Massachusetts Institute of Technology, Broad Institute Inc	Pending	2016-10-21	-
Csx28	WO2018191388A1	Massachusetts Institute of Technology, Broad Institute Inc	Pending	2018-04-11	-
CsaX	-	-	-	-	-
Cse1	EP3091072B1	Caribou Biosciences Inc	Granted	2012-12-21	2032-12-21
Cse2	JP6408914B2	Caribou Biosciences, Inc.	Granted	2012-12-21	2032-12-21
Cse5	-	-	-	-	-
Csy1	US20170283779A1	Charles Stark Draper Laboratory Inc	Pending	2017-03-27	-

Csy2	-	-	-	-	-
Csy3	-	-	-	-	-
Csy4	US9115348B2	University of California	Granted	2012-11-07	2032-01-01
Csm2	-	-	-	-	-
Csm3	US20170198286A1	Vilniaus Universitetas	Pending	2017-03-03	-
Csm4	US20170198286A1	Vilniaus Universitetas	Pending	2017-03-03	-
Csm5	-	-	-	-	-
Csm6	-	-	-	-	-
Csn2	-	-	-	-	-
Csf1	-	-	-	-	-
Csf2	-	-	-	-	-
Csf3	-	-	-	-	-
Csf4	-	-	-	-	-
Csf5	-	-	-	-	-
Cmr1	US8546553B2	University of Georgia Research Foundation Inc (UGARF)	Granted	2009-07-24	2029-07-24
Cmr2	US8546553B2	University of Georgia Research Foundation Inc (UGARF)	Granted	2009-07-24	2029-07-24
Cmr3	US8546553B2	University of Georgia Research Foundation Inc (UGARF)	Granted	2009-07-24	2029-07-24
Cmr4	US8546553B2	University of Georgia Research Foundation Inc (UGARF)	Granted	2009-07-24	2029-07-24
Cmr5	US8546553B2	University of Georgia Research Foundation Inc (UGARF)	Granted	2009-07-24	2029-07-24
Cmr6	US8546553B2	University of Georgia Research Foundation Inc (UGARF)	Granted	2009-07-24	2029-07-24

Table 5 is showing patent situation for all known CRISPR-Cas proteins. Majority of essential CRISPR-Cas proteins is either patented, or patent application has been applied for. The only exception so far is Cas11. Regarding additional CRISPR-Cas proteins and crRNPs – apart from CASCADE and Cmr surveillance complex the majority of those Cas proteins is not patented yet, except for Csx27 and Csx28 accessory proteins associated with Cas13, Csy1 subunit of Csy surveillance complex and Csm3-Csm4 subcomplex (Patents.google.com, 2019).

Discussion

This thesis aims to make a simple overview of functions and patent situation around known CRISPR-Cas proteins, as well as analyzing alternatives to the widely used CRISPR-Cas9 complex for genome editing purposes.

As shown in the results, a huge amount of CRISPR-associated proteins have been mapped since the discovery of CRISPR-Cas systems, and new CRISPR-Cas proteins or alternatives are being reported regularly. For a long time there has been a lot of confusion around CRISPR-Cas proteins, due to lack of a common nomenclature for the CRISPR-Cas systems. Before Makarova *et. al.* (Makarova et.al, 2011) proposed their now widely used classification of CRISPR-Cas systems, nomenclature of CRISPR-Cas proteins was more or less non-existing, leaving it up to each research group to name Cas-proteins they discovered. This led to a lot of confusion, and as shown in the overview of CRISPR-Cas proteins in Appendix I and II a large number of CRISPR-Cas proteins have several alternative names which have been assigned by different research groups.

After discovery of highly conserved CRISPR-Cas1, it was attempted to set up a classification based on Cas1 phylogeny. Later on, after more CRISPR-Cas locus were analyzed and genetically mapped, researchers found out than Cas1 is not as preserved as estimated, and several CRISPR-Cas locus lack Cas1 genes. When Makarova *et.al*. (Makarova et. al, 2011) published their work on “Evolution and classification of the CRISPR–Cas systems” the situation around CRISPR-Cas systems classification and nomenclature became much better. Proposed nomenclature of CRISPR systems, now based on locus architecture and composition rather than Cas1 phylogeny revealed how different CRISPR classes and types are connected and can be applied to *in vivo* and *in vitro* genome editing purposes. Later on, several nomenclature corrections have been made, and updated, modern classification based on Makarova *et. al.* work (Makarova et.al 2011,2015a,2015b,2018; Koonin et.al., 2017; Koonin & Makarova, 2019) is presented in the “Major types of CRISPR-Cas systems and their subtypes” chapter of this thesis.

This classification reveals two distinct CRISPR classes – class I and class II: The main difference between class I and class II CRISPR-Cas systems lies in the way CRISPR-Cas proteins are composed in these classes. While class I relies on cooperation of a set of single domain CRISPR-Cas proteins, class II contains multidomain proteins that often can achieve the same result.

Class II, being the most studied CRISPR systems thanks to discovery of many biotechnological appliances of CRISPR type II systems, has given rise to a set of modern, powerful genome editing tools commonly known as CRISPR-Cas9. CRISPR-Cas9 many alternatives have been widely used in different studies of human, animal and plant genomes and have shown an incredible potential for *in vivo* and *in vitro* appliances, and got a lot of attention in both research environment and media. CRISPR-Cas9 indisputable advantages over other programmable nucleases such as ZFNs and TALENs have brought attention to other CRISPR systems.

Quite recently discovered, type V CRISPR-Cas systems have shown that other CRISPR-Cas systems may have potential as programmable genome editing tools. The CRISPR-Cas9 analogue CRISPR-Cpf1 have shown even better results compared to CRISPR-Cas9, due to its intolerance for mismatches in crRNA-target-DNA bindings. Mismatch intolerance results in lower cytotoxicity and reduced risks for off-target activity which have in some cases been a huge problem for CRISPR-Cas9 genome editing. It seems reasonable to assume that CRISPR-Cpf1 is going to follow the same path as CRISPR-Cas9 has done since its discovery, and a series of modified Cpf1 proteins will be used to study CRISPR-Cpf1 potential.

CRISPR-Cas type VI system has not yet established itself as a genome editing tool in the same way as type II and V CRISPR systems. Unlike Cas9 and Cpf1, Cas13 – the signature protein of type VI systems, is able to solely process RNA molecules. Despite the fact of CRISPR-Cas type VI systems not being able to alternate DNA molecules they still may have a major role in genome editing. RNA processing power of Cas13 may have uses in post transcriptional epigenetic modifications of genes, such as expression regulations. It is possible that CRISPR-Cas13 has much of potential seen in RNA interference (RNAi) processes.

For a long time, compositionally more advanced class I has been existing in the shadow of class II systems. Until recently it was assumed that class I CRISPR systems are not suitable for genome editing as there were no evidence of successful use of class I for that purpose. However, in April 2019, a group of researchers led by Dolan A.E. (Dolan, 2019) have shown that genome editing by class I is possible. In their study of "Introducing a Spectrum of Long-Range Genomic Deletions in Human Embryonic Stem Cells Using Type I CRISPR-Cas" researchers have managed to get promising results by performing genomic alterations by CASCADE surveillance complex, fused with type I CRISPR-Cas3'' HD nuclease.
This discovery of class I CRISPR systems potential might indicate a new milestone in CRISPR guided genome editing with crRNPs – CRISPR ribonucleoproteins, or so-called CRISPR surveillance complexes. To this date five crRNPs are available in CRISPR systems – CASCADE surveillance complex, Csy surveillance complex, Csm surveillance complex, Csf surveillance complex and Cmr surveillance complex.

Even though it seems that further studies of those complexes are needed to unlock their full potential, it might look like they have a promising role for further development of CRISPR-Cas guided genome engineering. Functions of many additional Cas proteins is based on either sequence or structure similarity to other known proteins rather than *in vivo* or *in vitro* activities. Those assumption might be incorrect, taking in account that protein structure is more conserved than protein sequence.
Although there is a large amount of scientific papers regarding crRNP and the rest of additional class I CRISPR-Cas proteins it might still be quite challenging to estimate, discover and describe their functions. This challenging work needs to be done to gather full understanding of CRISPR class I proteins.

Regarding patent situation around CRISPR-Cas’es – it might seem like the majority of essential CRISPR-Cas proteins is either patented, or is under the process of patent acquisitions. Despite the fact that patent applications might limit and influence the process of CRISPR-Cas systems function discoveries in a negative way to a certain degree it seems understandable that researchers and research group get acknowledgments for their work on the matter. On the other hand, considering additional Cas proteins and crRNPs, apart from Cmr and CASCADE surveillance complexes and a few other Cas-proteins of class I the majority is patent free. That would suggest that those proteins have to undergo more research, making study of crRNPs a logical step in CRISPR systems studies.

Conclusion and future

This thesis aims to make a simple overview of functions and patent situation around known CRISPR-Cas proteins, as well as analyzing alternatives to the widely used CRISPR-Cas9 complex for genome editing purposes.

Complete overview with short description of functions is presented in Appendix I and II. The overview presents over 56 CRISPR-Cas proteins, divided into two classes, six types and 30 subtypes.

Analysis of CRISPR-Cas systems reveals that many CRISPR-associated proteins are nuclease enzymes that are able to process RNA and/or DNA. When it comes to CRISPR-Cas9 alternatives, CRISPR-Cas type V has already been proven to be able to perform at even higher rate than CRISPR-Cas9. CRISPR-Cas system type VI seems to be another good alternative for targeting RNA sequences.

Discovery of CASCADE-Cas3 potential as a programmable nuclease indicates that CRISPR-Cas class I should be studied for genome editing purposes. Furthermore, one can speculate that knowing functions of essential CRISPR-Cas proteins researchers may someday be able to compose their own programmable nuclease complexes like CASCADE using single domain CRISPR-Cas proteins. It looks like crRNPs may play a big role in achieving such purposes.

When it comes to patent situation around CRISPR-Cas, almost all essential CRISPR-Cas proteins seem to be either patented, or under the process of patent pending as shown in “CRISPR-Cas patents” chapter. On the other hand, most of the CRISPR-Cas proteins involved in Cse, Csm, Csy, CASCADE and Csf surveillance complexes do not seem to be patented. This may again indicate the need for further studies of those complexes, something that possibly will draw researchers’ attention in near future.

CRISPR-Cas systems has brought an enormous potential for modern genome engineering. With the constantly growing need to adapt to ever evolving needs for modern world problem solutions like development of crop defense mechanism against pests, parasites and diseases, treatment of human and animal diseases, antibiotic resistance et cetera, programmable nucleases might help the humanity in the race against the clock.

It might seem like there is still a long way to go before we can fully understand and adapt the bacterial and archaeal mechanism of self-defense, before we can unlock CRISPR-Cas systems full potential. CRISPR-Cas systems are an extremely difficult field in modern biotechnology and sometimes even the researchers seem to have trouble understanding the metabolism and pathways of this constantly growing field.

Even though it still seems like there is a long way to go, discovery of CRISPR-Cas systems like CRISPR-Cas9, CRISPR-Cpf1 and CASCADE-Cas3 seem to have started an extremely important process potentially leading to human engineered processes of disease treatment in living organisms, programmed genome editing and better understanding of nature itself.

References

Abudayyeh, O., Gootenberg, J. (2019). Cas13 — Zhang Lab. *Zhang Lab*. Available at: https://zlab.bio/cas13 (Accessed 13 May 2019).

Agari, Y., Yokoyama, S., Kuramitsu, S. & Shinkai, A. (2008). X-ray crystal structure of a CRISPR-associated protein, Cse2, from Thermus thermophilus HB8. 73 (4): 1063-1067.

Aryan, A., Anderson, M. A., Myles, K. M. & Adelman, Z. N. (2013). TALEN-based gene disruption in the dengue vector Aedes aegypti. *PLoS One*, 8 (3): e60082.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315 (5819): 1709-1712.

Barrangou, R. (2015). Diversity of CRISPR-Cas immune systems and molecular machines. *Genome Biology*, 16 (1): 247.

Bibikova, M., Carroll, D., Segal, D. J., Trautman, J. K., Smith, J., Kim, Y.-G. & Chandrasegaran, S. (2001). Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Molecular and cellular biology*, 21 (1): 289-297.

Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. (2002). Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics*, 161 (3): 1169-1175.

Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. & Bonas, U. (2009). Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, 326 (5959): 1509-1512.

Bondy-Denomy, J., Garcia, B., Strum, S., Du, M., Rollins, M. F., Hidalgo-Reyes, Y., Wiedenheft, B., Maxwell, K. L. & Davidson, A. R. (2015). Multiple mechanisms for CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature*, 526 (7571): 136.

Carlson, D. F., Tan, W., Lillico, S. G., Stverakova, D., Proudfoot, C., Christian, M., Voytas, D. F., Long, C. R., Whitelaw, C. B. A. & Fahrenkrug, S. C. (2012). Efficient TALEN-mediated gene knockout in livestock. *Proceedings of the National Academy of Sciences*, 109 (43): 17382-17387.

Carroll, D. (2017). Focus: Genome Editing: Genome Editing: Past, Present, and Future. *The Yale journal of biology medicine,* 90 (4): 653.

Carte, J., Wang, R., Li, H., Terns, R. M. & Terns, M. P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes development,* 22 (24): 3489-3496.

Cass, S. D., Haas, K. A., Stoll, B., Alkhnbashi, O. S., Sharma, K., Urlaub, H., Backofen, R., Marchfelder, A. & Bolt, E. L. (2015). The role of Cas8 in type I CRISPR interference. *Bioscience reports,* 35 (3): e00197.

Chandrasegaran, S. & Carroll, D. (2016). Origins of programmable nucleases for genome engineering. *Journal of molecular biology*, 428 (5): 963-989.

Chen, K. & Gao, C. (2013). TALENs: customizable molecular DNA scissors for genome engineering of plants. *Journal of Genetics and Genomics*, 40 (6): 271-279.

Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature biotechnology*, 31 (3): 230.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W. & Marraffini, L. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science,* 339 (6121): 819-823.

Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J. & Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471 (7340): 602.

Dolan, A. E., Hou, Z., Xiao, Y., Gramelspacher, M. J., Heo, J., Howden, S. E., Freddolino, P. L., Ke, A. & Zhang, Y. (2019). Introducing a Spectrum of Long-Range Genomic Deletions in Human Embryonic Stem Cells Using Type I CRISPR-Cas. *Molecular cell*, 74: 1-15.

Ebihara, A., Yao, M., Masui, R., Tanaka, I., Yokoyama, S. & Kuramitsu, S. (2006). Crystal structure of hypothetical protein TTHB192 from Thermus thermophilus HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein science*, 15 (6): 1494-1499.

Feng, Z., Zhang, B., Ding, W., Liu, X., Yang, D.-L., Wei, P., Cao, F., Zhu, S., Zhang, F. & Mao, Y. (2013). Efficient genome editing in plants using a CRISPR/Cas system. *Cell research*, 23 (10): 1229.

Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. (2012). Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, 109 (39): E2579-E2586.

Genovese, P., Schiroli, G., Escobar, G., Di Tomaso, T., Firrito, C., Calabria, A., Moi, D., Mazzieri, R., Bonini, C. & Holmes, M. C. (2014). Targeted genome editing in human repopulating haematopoietic stem cells. *Nature*, 510 (7504): 235.

Ghorbal, M., Gorman, M., Macpherson, C. R., Martins, R. M., Scherf, A. & Lopez-Rubio, J.-J. (2014). Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system. *Nature biotechnology*, 32 (8): 819.

Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology,* 1(6): e60.

Hatoum-Aslan, A., Samai, P., Maniv, I., Jiang, W. & Marraffini, L. A. (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *Journal of Biological Chemistry*, 288 (39): 27888-27897.

Hatoum-Aslan, A., Maniv, I., Samai, P. & Marraffini, L. A. (2014). Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *Journal of bacteriology*, 196 (2): 310-317.

Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., Retterath, A., Stoddard, T., Juillerat, A. & Cedrone, F. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant biotechnology journal*, 12 (7): 934-940.

Hickman, A. B. & Dyda, F. (2015). The casposon-encoded Cas1 protein from Aciduliprofundum boonei is a DNA integrase that generates target site duplications. *Nucleic Acids Research*, 43 (22): 10576-10587.

Holkers, M., Maggio, I., Liu, J., Janssen, J. M., Miselli, F., Mussolino, C., Recchia, A., Cathomen, T. & Goncalves, M. A. (2012). Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic acids research*, 41 (5): e63-e63.

Hsu, P. D., Scott, D. A., Weinstein, J. A., Ran, F. A., Konermann, S., Agarwala, V., Li, Y., Fine, E. J., Wu, X. & Shalem, O. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature biotechnology*, 31 (9): 827.

Hsu, P. D., Lander, E. S. & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell,* 157 (6): 1262-1278.

InvivoGen. (2019). *RNA Interference.* Available at: https://www.invivogen.com/review-rna-interference. (Accessed: 26.03.2019).

Jiang, F. & Doudna, J. A. (2015). The structural biology of CRISPR-Cas systems. *Current opinion in structural biology,* 30: 100-111.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337 (6096): 816-821.

Joung, J. K. & Sander, J. D. (2013). TALENs: a widely applicable technology for targeted genome editing. *Nature reviews Molecular cell biology,* 14 (1): 49.

Kandavelou, K., Ramalingam, S., London, V., Mani, M., Wu, J., Alexeev, V., Civin, C. I. & Chandrasegaran, S. (2009). Targeted manipulation of mammalian genomes using designed zinc finger nucleases. *Biochemical and biophysical research communications*, 388 (1): 56-61.

Karvelis, T., Gasiunas, G., Miksys, A., Barrangou, R., Horvath, P. & Siksnys, V. (2013). crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. *RNA biology*, 10 (5): 841-851.

Kim, Y.-G., Cha, J. & Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences*, 93 (3): 1156-1160.

Kim, T.-Y., Shin, M., Yen, L. H. T. & Kim, J.-S. (2013). Crystal structure of Cas1 from Archaeoglobus fulgidus and characterization of its nucleolytic activity. *Biochemical and biophysical research communications*, 441 (4): 720-725.

Kim, D., Kim, J., Hur, J. K., Been, K. W., Yoon, S.-h. & Kim, J.-S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nature biotechnology,* 34 (8): 863.

Kim, H. K., Song, M., Lee, J., Menon, A. V., Jung, S., Kang, Y.-M., Choi, J. W., Woo, E., Koh, H. C. & Nam, J.-W. (2017). In vivo high-throughput profiling of CRISPR–Cpf1 activity. *Nature methods,* 14 (2): 153.

Kleinstiver, B. P., Tsai, S. Q., Prew, M. S., Nguyen, N. T., Welch, M. M., Lopez, J. M., McCaw, Z. R., Aryee, M. J. & Joung, J. K. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nature biotechnology,* 34 (8): 869.

Klug, A. (2010). The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annual Review of Biochemistry*, 79: 213-231.

Komor, A. C., Badran, A. H. & Liu, D. R. (2017). CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell*, 168 (1-2): 20-36.

Koonin, E. V., Makarova, K. S. & Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current Opinion in Microbiology*

Koonin, E. V. & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B*, 374 (1772): 20180087.

Lee, H., Zhou, Y., Taylor, D. W. & Sashital, D. G. (2018). Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. *Molecular cell,* 70 (1): 48-59. e5.

Leenay, R. T., Maksimchuk, K. R., Slotkowski, R. A., Agrawal, R. N., Gomaa, A. A., Briner, A. E., Barrangou, R. & Beisel, C. L. (2016). Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Molecular cell,* 62 (1): 137-147.

Li, Y., Zhang, Y., Lin, J., Pan, S., Han, W., Peng, N., Liang, Y. X. & She, Q. (2017). Cmr1 enables efficient RNA and DNA interference of a III-B CRISPR–Cas system by binding to target RNA and crRNA. *Nucleic acids research*, 45 (19): 11305-11314.

Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q. & Copié, V. (2011). Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *Journal of Biological Chemistry,* 286 (24): 21643-21656.

Ma, N., Liao, B., Zhang, H., Wang, L., Shan, Y., Xue, Y., Huang, K., Chen, S., Zhou, X. & Chen, Y. (2013). Transcription activator-like effector nuclease (TALEN)-mediated gene correction in integration-free β-thalassemia induced pluripotent stem cells. *Journal of Biological Chemistry*, 288 (48): 34671-34679.

Majumdar, S., Zhao, P., Pfister, N. T., Compton, M., Olson, S., Glover, C. V., Wells, L., Graveley, B. R., Terns, R. M. & Terns, M. P. (2015). Three CRISPR-Cas immune effector complexes coexist in Pyrococcus furiosus. *RNA,* 21 (6): 1147-1158.

Majumdar, S. & Terns, M. P. (2019). CRISPR RNA-guided DNA cleavage by reconstituted Type IA immune effector complexes. *Extremophiles*, 23 (1): 19-33.

Makarova, K. S., Aravind, L., Wolf, Y. I. & Koonin, E. V. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology direct,* 6(1): 38.

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I. & Yakunin, A. F. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, 9 (6): 467.

Makarova, K. S., Wolf, Y. I. & Koonin, E. V. (2013). The basic building blocks and evolution of CRISPR–Cas systems. *Biochemical Society Transactions*, 41 (6): 1392-1400.

Makarova, K. S. & Koonin, E. V. (2015). Annotation and classification of CRISPR-Cas systems. In *CRISPR*, pp. 47-75: Springer.

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J., Charpentier, E. & Haft, D. H. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology,* 13 (11): 722.

Makarova, K. S., Zhang, F. & Koonin, E. V. (2017). SnapShot: class 1 CRISPR-Cas systems. *Cell*, 168 (5): 946-946. e1.

Makarova, K. S., Wolf, Y. I. & Koonin, E. V. (2018). Classification and nomenclature of CRISPR-Cas systems: where from here? *The CRISPR journal*, 1 (5): 325-336.

Mani, M., Kandavelou, K., Dy, F. J., Durai, S. & Chandrasegaran, S. (2005). Design, engineering, and characterization of zinc finger nucleases. *Biochemical and biophysical research communications*, 335 (2): 447-457.

Mansour, S. L., Thomas, K. R. & Capecchi, M. R. (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature*, 336 (6197): 348.

Marraffini, L. A. & Sontheimer, E. J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, 463 (7280): 568.

Miller, J., McLachlan, A. & Klug, A. (1985). Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *The EMBO journal*, 4 (6): 1609-1614.

Mohanraju, P., Makarova, K. S., Zetsche, B., Zhang, F., Koonin, E. V. & Van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science*, 353 (6299): aad5147.

Moscou, M. J. & Bogdanove, A. J. (2009). A simple cipher governs DNA recognition by TAL effectors. *Science*, 326 (5959): 1501-1501.

Mulepati, S., Héroux, A. & Bailey, S. (2014). Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. *Science*, 345 (6203): 1479-1484.

Muller, H. J. (1927). Artificial transmutation of the gene. *Science*, 66 (1699): 84-87.

Nam, K. H., Kurinov, I. & Ke, A. (2011). Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity. *Journal of Biological Chemistry*, 286 (35): 30759-30768.

Niewoehner, O. & Jinek, M. (2016). Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *Rna*, 22 (3): 318-329.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., Ishitani, R., Zhang, F. & Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, 156 (5): 935-949.

Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J., Wang, L., Kang, Y., Zhao, X., Si, W. & Li, W. (2014). Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell*, 156 (4): 836-843.

Numata, T., Inanaga, H., Sato, C. & Osawa, T. (2015). Crystal Structure of the Csm3–Csm4 Subcomplex in the Type III-A CRISPR–Cas Interference Complex. *Journal of molecular biology*, 427 (2): 259-273.

Nuñez, J. K., Kranzusch, P. J., Noeske, J., Wright, A. V., Davies, C. W. & Doudna, J. A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature structural molecular biology,* 21 (6): 528.

Nuñez, J. K., Lee, A. S., Engelman, A. & Doudna, J. A. (2015). Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature*, 519 (7542): 193.

Patents.google.com. (2019). *Google Patents*. Available at: https://patents.google.com/ (Accessed 14 May 2019).

Pavletich, N. P. & Pabo, C. O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science*, 252 (5007): 809-817.

Plagens, A., Tjaden, B., Hagemann, A., Randau, L. & Hensel, R. (2012). Characterization of the CRISPR/Cas subtype IA system of the hyperthermophilic crenarchaeon Thermoproteus tenax. *Journal of bacteriology,* 194 (10): 2491-2500.

Plagens, A., Richter, H., Charpentier, E. & Randau, L. (2015). DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS microbiology reviews*, 39 (3): 442-463.

Ramalingam, S., Annaluru, N., Kandavelou, K. & Chandrasegaran, S. (2014). TALEN-mediated generation and genetic correction of disease-specific human induced pluripotent stem cells. *Current gene therapy*, 14 (6): 461-472.

Reeks, J., Graham, S., Anderson, L., Liu, H., White, M. F. & Naismith, J. H. (2013). Structure of the archaeal Cascade subunit Csa5: relating the small subunits of CRISPR effector complexes. *RNA biology*, 10 (5): 762-769.

Richter, C., Gristwood, T., Clulow, J. S. & Fineran, P. C. (2012). In vivo protein interactions and complex formation in the Pectobacterium atrosepticum subtype IF CRISPR/Cas System. *PloS one*, 7 (12): e49549.

Robertson, A., Klungland, A., Rognes, T. & Leiros, I. (2009). DNA repair in mammalian cells. *Cellular molecular life sciences,* 66 (6): 981-993.

Rothstein, R. J. (1983). One-step gene disruption in yeast. In vol. 101 *Methods in enzymology*, pp. 202-211: Elsevier.

Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C. V., Spagnolo, L. & White, M. F. (2013). Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Molecular cell*, 52 (1): 124-134.

Rusk, N. (2019). Spotlight on Cas12. *Nature methods*, 16 (3): 215.

Scherer, S. & Davis, R. W. (1979). Replacement of chromosome segments with altered DNA sequences constructed in vitro. *Proceedings of the National Academy of Sciences,* 76 (10): 4951-4955.

Segal, D. J. & Meckler, J. F. (2013). Genome engineering at the dawn of the golden age. *Annual review of genomics and human genetics*, 14: 135-158.

Shah, S. A., Alkhnbashi, O. S., Behler, J., Han, W., She, Q., Hess, W. R., Garrett, R. A. & Backofen, R. (2019). Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR-cas gene cassettes reveals 39 new cas gene families. *RNA biology*, 16 (4): 530-542.

Shao, Y., Cocozaki, A. I., Ramia, N. F., Terns, R. M., Terns, M. P. & Li, H. (2013). Structure of the Cmr2-Cmr3 subcomplex of the Cmr RNA silencing complex. *Structure*, 21 (3): 376-384.

Shen, B., Zhang, W., Zhang, J., Zhou, J., Wang, J., Chen, L., Wang, L., Hodgkins, A., Iyer, V. & Huang, X. (2014). Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nature methods*, 11 (4): 399.

Sheppard, N. F., Glover, C. V., Terns, R. M. & Terns, M. P. (2016). The CRISPR-associated Csx1 protein of Pyrococcus furiosus is an adenosine-specific endoribonuclease. *Rna*, 22 (2): 216-224.

Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. & Siksnys, V. (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *The EMBO journal,* 30 (7): 1335-1342.

Smargon, A. A., Cox, D. B., Pyzocha, N. K., Zheng, K., Slaymaker, I. M., Gootenberg, J. S., Abudayyeh, O. A., Essletzbichler, P., Shmakov, S. & Makarova, K. S. (2017). Cas13b is a type VI-B CRISPR-associated RNA-guided RNase differentially regulated by accessory proteins Csx27 and Csx28. *Molecular cell,* 65 (4): 618-630. e7.

Smith, J., Bibikova, M., Whitby, F. G., Reddy, A., Chandrasegaran, S. & Carroll, D. (2000). Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic acids research*, 28 (17): 3361-3369.

Sternberg, S. H. & Doudna, J. A. (2015). Expanding the biologist's toolkit with CRISPR-Cas9. *Molecular cell,* 58 (4): 568-574.

Streubel, J., Blücher, C., Landgraf, A. & Boch, J. (2012). TAL effector RVD specificities and efficiencies. *Nature biotechnology,* 30 (7): 593.

Strohkendl, I., Saifuddin, F. A., Rybarski, J. R., Finkelstein, I. J. & Russell, R. (2018). Kinetic basis for DNA target specificity of CRISPR-Cas12a. *Molecular cell,* 71 (5): 816-824. e3.

Tamulaitis, G., Venclovas, Č. & Siksnys, V. (2017). Type III CRISPR-Cas immunity: major differences brushed aside. *Trends in microbiology*, 25 (1): 49-61.

Tan, W., Carlson, D. F., Lancto, C. A., Garbe, J. R., Webster, D. A., Hackett, P. B. & Fahrenkrug, S. C. (2013). Efficient nonmeiotic allele introgression in livestock using custom endonucleases. *Proceedings of the National Academy of Sciences*, 110 (41): 16526-16531.

Taylor, D. W., Zhu, Y., Staals, R. H., Kornfeld, J. E., Shinkai, A., van der Oost, J., Nogales, E. & Doudna, J. A. (2015). Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science*, 348 (6234): 581-585.

Thomas, K. R., Folger, K. R. & Capecchi, M. R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell,* 44 (3): 419-428.

Townsend, J. A., Wright, D. A., Winfrey, R. J., Fu, F., Maeder, M. L., Joung, J. K. & Voytas, D. F. (2009). High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature*, 459 (7245): 442.

UniProtKB. (2019). UniProtKB - G2FJ81 (G2FJ81_9GAMM), *cas11 - CRISPR-associated endonuclease*. Availible at: *Uniprot.org.* (Accessed: 22.04.19).

Urnov, F. D., Miller, J. C., Lee, Y.-L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D. & Holmes, M. C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, 435 (7042): 646.

Venclovas, Č. (2016). Structure of Csm2 elucidates the relationship between small subunits of CRISPR-Cas effector complexes. *FEBS letters*, 590 (10): 1521-1529.

Walker, F. C., Chou-Zheng, L., Dunkle, J. A. & Hatoum-Aslan, A. (2016). Molecular determinants for CRISPR RNA maturation in the Cas10–Csm complex and roles for non-Cas nucleases. *Nucleic acids research,* 45 (4): 2112-2123.

Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. (2011). Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure,* 19 (2): 257-264.

Wang, L., Mo, C. Y., Wasserman, M. R., Rostøl, J. T., Marraffini, L. A. & Liu, S. (2019). Dynamics of Cas10 Govern Discrimination between Self and Non-self in Type III CRISPR-Cas Immunity. *Molecular cell*, 73 (2): 278-290. e4.

Woo, J. W., Kim, J., Kwon, S. I., Corvalán, C., Cho, S. W., Kim, H., Kim, S.-G., Kim, S.-T., Choe, S. & Kim, J.-S. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature biotechnology*, 33 (11): 1162.

Yan, X., Guo, W. & Yuan, Y. A. (2015). Crystal structures of CRISPR-associated Csx3 reveal a manganese-dependent deadenylation exoribonuclease. *RNA biology*, 12 (7): 749-760.

Zaidi, S. S.-e.-A., Mahfouz, M. M. & Mansoor, S. (2017). CRISPR-Cpf1: a new tool for plant genome editing. *Trends in plant science*, 22 (7): 550-553.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., Van Der Oost, J. & Regev, A. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell,* 163 (3): 759-771.

Zhang, J., Kasciukovic, T. & White, M. F. (2012). The CRISPR associated protein Cas4 Is a 5′ to 3′ DNA exonuclease with an iron-sulfur cluster. *PLoS One,* 7(10): e47232.

Zhu, X. & Ye, K. (2014). Cmr4 is the slicer in the RNA-targeting Cmr CRISPR complex. *Nucleic acids research*, 43 (2): 1257-1267.

Özcan, A., Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G. & Randau, L. (2019). Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nature microbiology*, 4 (1): 89.

A. Appendix

Appendix I

Table A-1. *Essential CRISPR-Cas proteins.*

Cas-protein name	System type or subtype	Alternative names and homologues	Function/type of protein
Cas1	• Type I • Type II • Subtype III-A •Type V •Subtype VI-A •Subtype VI-D	-	Casposase, DNase, spacer integration
Cas2	• Type I • Type II • Subtype III-A • Subtype V-A • Subtype V-E • Subtype V-B • Subtype VI-A • Subtype VI-D	-	RNase, specific to U-rich regions, DNase, spacer integration
Cas3	• Type I	Cas3', Cas3''	'Helicase, ''HD endonuclease, viral DNA degradation
Cas4	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-D • Subtype I-U • Subtype II-B	Csa1	Exonuclease, reverse transcriptase, spacer integration
Cas5	• Type I • Type III • Type IV	Cas5a, Cas5d, Cas5e, Cas5h, Cas5p, Cas5t, Cmx5, CasD, COG1688, GSU0054	Nuclease, CASCADE complex, crRNA maturation
Cas6	• Subtype I-A • Subtype I-B • Subtype I-U • Subtype I-D • Subtype I-E • Subtype I-F • Subtype III-A • Subtype III-B • Subtype IV-A	Cmx6, Csf5 Cse3, CasE Csy4, Cas6a, Cas6b, Cas6c, Cas6d, Cas6e, Cas6f	Endoribonuclease, crRNA maturation
Cas7	• Type I • Type III • Type IV	Csa2, Csd2, Cse4, Csh2, Csp1, Cst2, CasC	CASCADE stabilization protein
Cas8	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-U • Subtype I-E • Subtype I-F • Type IV	Cmx1, Cst1, Csx8, Csx13, CXXC-CXXC, Csa4, Csx9, Csh1 ,TM1802, Csd1, Csp2, Cas8a,Cas8a1, Cas8a2, Cas8b, Cas8c	RNase, PAM recognition
Cas9	• Type II	Csn1, Csx12	Endonuclease, viral DNA degradation
Cas10	• Subtype I-D • Type III	Cmr2, Csm1, Csx1, Csc3, MTH326, Csx11	Cyclase polymerase, crRNA biogenesis
Cas11	• Type I • Type III • Subtype IV-B	Csa5	Endodeoxyribonuclease, Small Subunit protein
Cas12	• Type V	Cpf1	RNase, DNase, viral DNA/RNA degaradation
Cas13	•Type VI	Cas13a, Cas13b, Cas13c, Cas13d	Ribonuclease

Appendix II

Table A-2. *Additional CRISPR-Cas proteins*

Cas-protein name	System type or subtype	Alternative names	Function/type of protein
Csy1	• Subtype I-F	-	Csy RNP large subunit, possibly a polymerase
Csy2	• Subtype I-F	-	Csy RNP stabilization, possibly target recognition protein
Csy3	• Subtype I-F	-	Csy RNP backbone stabilizing protein
Csy4	• Subtype I-F	-	Csy RNP endoribonuclease
Cse1	• Subtype I-E	CasA	CASCADE RNP stabilization protein, Cas3' recruitment
Cse2	• Subtype I-E	CasB, TTHB189, RoseRS_0649, Ppro_2341, Pmen_3759	CASCADE RNP protein
Cse5	• Subtype I-E	-	CASCADE RNP stabilization protein, prevents crRNA binding to viral DNA
Csc1	• Subtype I-D	-	Possibly RNase
Csc2	• Subtype I-D	-	Unknown
Csn2	• Subtype II-A	-	Metal regulated DNase
Csm2	• Subtype III-A	Cse2(?)	Csm RNP small subunit, target binding
Csm3	• Subtype III-A	-	Csm RNP ruler protein, RNase
Csm4	• Subtype III-A	COG1567	Csm RNP stabilization protein, Cas10 recruitment
Csm5	• Subtype III-A	-	Csm RNP large subunit, crRNA maturation
Csm6	• Subtype III-A	APE2256	Csm RNP endoribonuclease
Cmr1	• Subtype III-B	-	Cmr RNP activation module protein
Cmr2	• Subtype III-B	CasiO, Csml	Cmr RNP protein, cyclase polymerase, crRNA biogenesis
Cmr3	• Subtype III-B	COG1768	Cmr RNP protein, endonuclease
Cmr4	• Subtype III-B	-	Cmr RNP backbone stabilizing protein
Cmr5	• Subtype III-B	-	Cmr RNP backbone stabilizing protein
Cmr6	• Subtype III-B	-	Cmr RNP protein, prevents crRNA binding to viral DNA
Csb1	• Subtype I-U	GSU0053	Unknown
Csb2	• Subtype I-U	-	Unknown
Csb3	• Subtype I-U	-	Unknown
Csx1	• Subtype III-B	csa3,csx2,DXTHG, NE0113, TIGR02710	Temperature-dependent RNase, specific to A-rich regions
Csx3	• Subtype III-B	-	RNase
Csx10	• Subtype I-U	all1473, Cas5-Cas7 fusion	Nuclease
Csx14	• Subtype I-U	-	Unknown
Csx15	• Type III	Csx20	Peptidase, crRNA maturation
Csx16	• Subtype III-U	VVA1 548	Unknown
Csx17	• Subtype I-U	-	Unknown
Csx18	• Subtype I-U	-	Unknown
Csx19	• Type III	-	Unknown
Csx24	• Type III	-	Unknown
Csx26	• Type III	-	Putative SS protein
Csx27	• Subtype VI-B1	-	Accessory protein, represses Cas13
Csx28	• Subtype VI-B2	-	Accessory protein, enhances Cas13
CsaX	• Subtype III-U	-	Unknown
Csf1	• Type IV	RHA1_ro10070	Csf RNP protein, target recognition
Csf2	• Type IV	-	Csf RNP protein, helical backbone stabilization protein
Csf3	• Type IV	-	Csf RNP protein, nuclease, crRNA maturation
Csf4	• Type IV	DinG	Csf RNP protein, helicase
Csf5	• Type IV	-	Csf RNP protein, endonuclease generating unusual 5'-terminal 7nt tag repeat , crRNA maturation

Appendix III

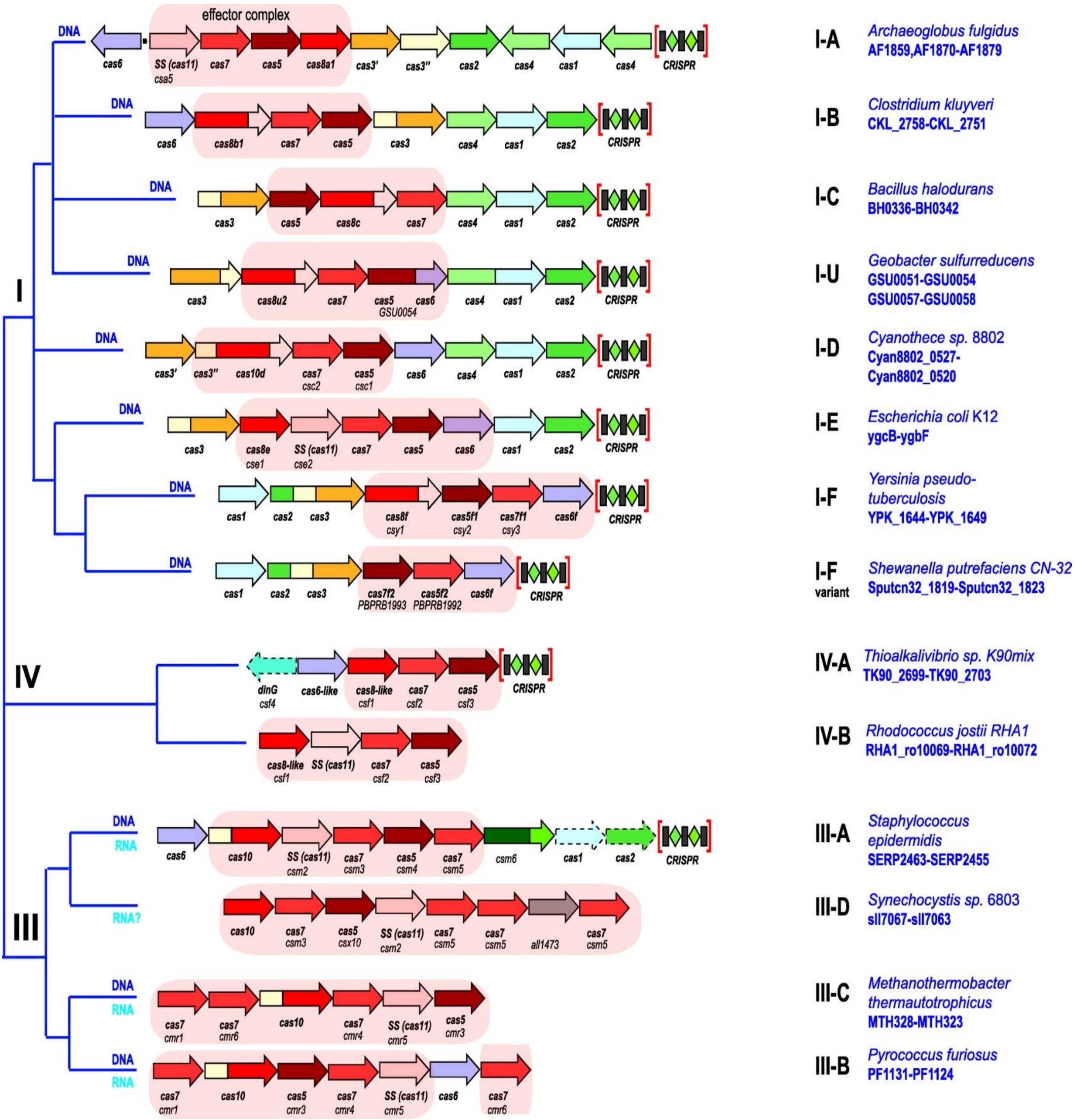

Figure A-3. *Koonin, E. V. & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. Philosophical Transactions of the Royal Society B, 374 (1772): 20180087.*

CRISPR-Cas class I systems overview.

Appendix IV

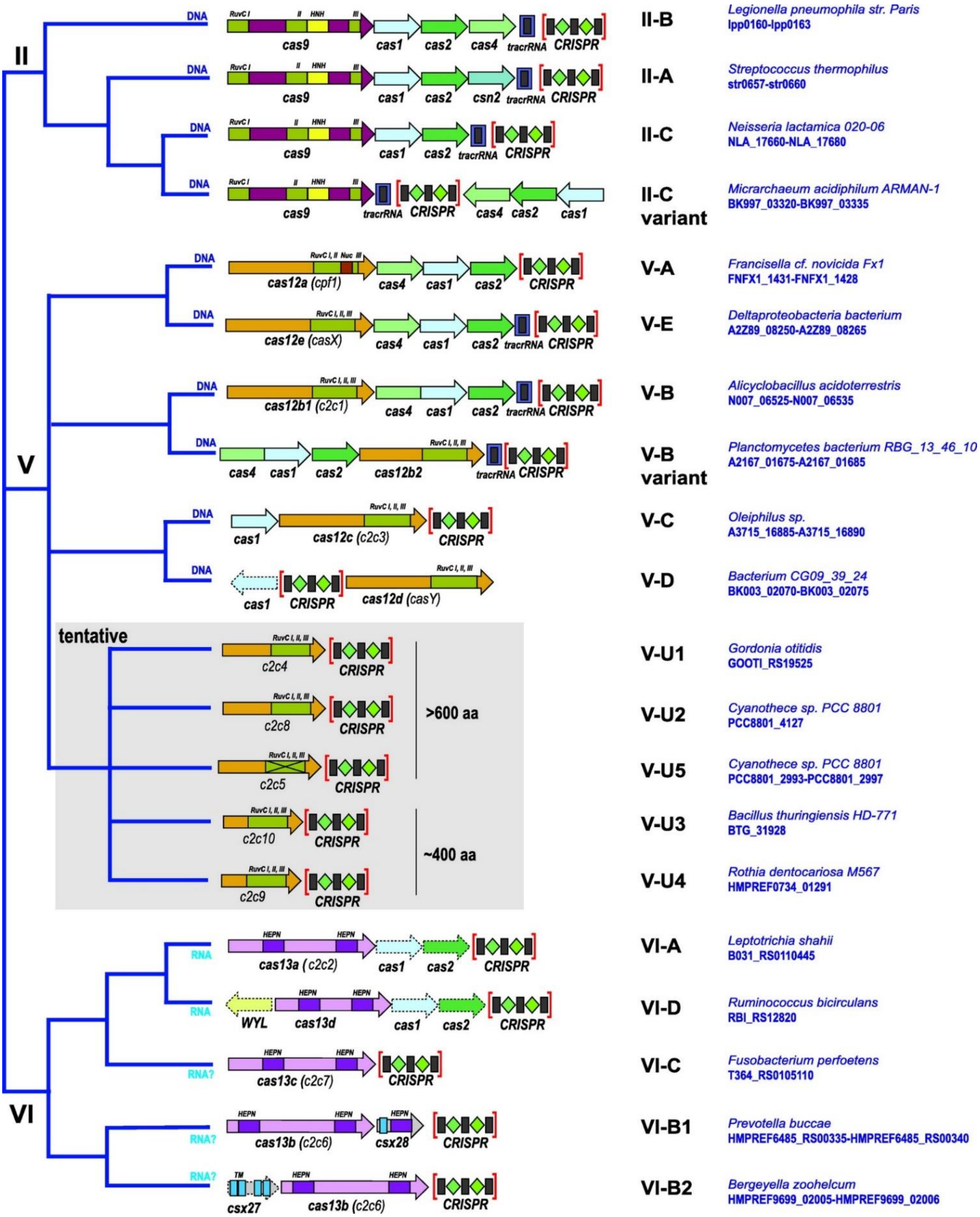

Figure A-4. *Koonin, E. V. & Makarova, K. S. (2019). Origins and evolution of CRISPR-Cas systems. Philosophical Transactions of the Royal Society B, 374 (1772): 20180087.*

CRISPR-Cas class II systems overview.

Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway