

# Spacer integration by the Cas1-Cas2 protein complex in different CRISPR-Cas systems

---

Noëlle ten Boer

S3291405

Bachelor thesis

Pre-master Molecular Biology and Biotechnology

Supervisor: prof. dr. J. Kok

Molecular Genetics

16-08-2018

words: 5476

## Abstract

Bacteriophages are small viruses that can infect a large variety of bacteria. Bacteria can protect themselves against these phages at different stages of their life cycle. The last decade, a lot of research has been done on a new bacterial defense mechanism, represented by the so-called clustered regularly interspaced short palindromic repeats (CRISPR). CRISPRs provide an adaptive immune system in bacteria and archaea that provides acquired immunity against invading DNA, such as from phages. CRISPRs were first reported in *Escherichia coli* and consists of palindromic spacers with short repeats in between them. Together they form the CRISPR array, which is flanked by a leader and several CRISPR associated (*cas*) genes.

Different types and classes of CRISPR-Cas systems have been reported. Cas1-Cas2 is a protein complex that is specified by almost all types of CRISPR-Cas loci. Cas1-Cas2 is functional in spacer acquisition, the integration of new spacers into the CRISPR-array. These spacers are derived from a foreign DNA molecule and give the bacterial host resistance against the bacteriophage or conjugative plasmids with this particular nucleotide motif.

Performing more research and developing more know-how on CRISPR-Cas systems would allow developing new genetic techniques. These could be used for biomedical purposes as well as biotechnological purposes, such as the recording of molecular events in genomes or replacing genes therein. Unfortunately, CRISPR-Cas received some negative feedback last months too in the popular news.

This thesis gives an overview about the CRISPR-Cas9 system and, more explicitly, on the Cas1-Cas2 protein complex. Its structural and mechanical basis, specificity, possible applications in future research and industrial applications will be discussed, since the system is important in all three main types of CRISPR-Cas systems known to date.

---

**Abbreviations.** Abi, abortive infection; aTc, anhydrotetracycline; bp, base pairs; cas, CRISPR-associated; Cascade, CRISPR-associated complex for antiviral defense; CRISPR, clustered regularly interspaced short palindromic repeats; crRNA, CRISPR RNA; ds, double stranded; EU, European union; GMO, genetically modified organism; IHF, integration host factor; IPTG, isopropyl- $\beta$ -D-1-thiogalactopyranoside; LAS, leader-anchoring sites; nt, nucleotide; PAM, protospacer adjacent motif; PICIs, phage-inducible chromosomal islands; Ppi, phage packaging interference; pRec, recording plasmid; pre-crRNA, precursor CRISPR RNA; R-M, restriction-modification; Sie, superinfection exclusion; TRACE, Temporal Recording in Arrays by CRISPR Expansion; tracrRNA, trans-activating CRISPR RNA; pTrig, trigger plasmid.

## Introduction

Bacteriophages are severe small viruses that can infect a large variety of bacteria. Phages are about 20-200 nm in size and were discovered nearly exactly 100 years ago. These bacterial predators have been very important in genetic and biological research. Researchers in 1950s identified DNA as genetic material by using phages in their research and besides that, they addressed a phenomenon called “enzyme induction”. This is the transcription of genes only in presence of a corresponding substrate. In the late 1960s and early 1970s researchers described restriction enzymes and their ability to cleave DNA. Bacteria produced these enzymes as a defense system against bacteriophage attack [1]. The restriction enzymes were at the basis of the genetic engineering revolution that started in the early 1970s.

Phages can have a big negative impact on food and biotechnology industries. In the worst case they can cause total production loss [2]. However, phages have some good aspects as well. They tend to be successful as a therapeutic against pathogenic bacterial infections. Phage-therapy was discovered and pioneered in 1919, but unfortunately the interest in phage-therapy in the West decreased when the industrial production of antibiotics increased [4]. However, antibiotic resistance is a big problem these days. This caused the interest in phages, and thus phage-therapy, to increase again [5,6].

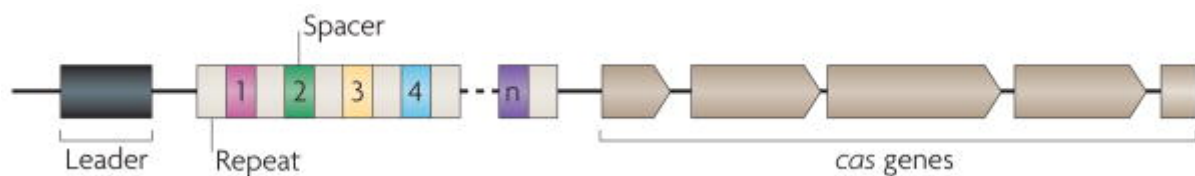
Bacteria can protect themselves in many ways against phage attack at and their different stages the phage’s life cycle (*Figure 1*). The first way is to prevent attachment of the phage to the bacterial cell and/or to specific surface receptors. This can be countered by a decrease in receptor availability, through mutations that modify the receptor structure or by concealing receptors with an additional physical barrier. The latter can be a capsule or, for example, other extracellular polymers. If the phage is already attached, the bacteria can still block the entry of the phage DNA. Superinfection exclusion (Sie) systems can block injection of phage DNA into the host cells. By removing the DNA from the infecting bacteriophage, the bacterial population has a strong advantage: it does not only protect the specific host cell, but also the surrounding cells. When phage DNA has entered the host cell, bacteria can use restriction-modification (R-M) systems to get rid of the phage DNA. The R-M system consists of a restriction endonuclease and a methyltransferase. First, the methyltransferase methylates the host DNA, whereafter the endonuclease cleaves the non-methylated phage DNA into smaller fragments. These fragment are harmless to the infected bacterium. Four different R-M systems have been described in literature. A fourth protective mechanism is called abortive infection (Abi), a system that leads to death of the host cell, preventing the phage to propagate inside the cell, as a sacrifice to protect surrounding bacteria [7]. More than 20 Abi systems have been described in the bacterium *Lactococcus Lactis*, which is confronted with a lot of phages during its use in cheese factories [8]. Each Abi system acts differently to get the same result, namely death and lysis of the host cell. Finally, bacteria can defend themselves against phages by interfering with the assembly of bacteriophage particles. In the latter defense mechanism, phage-inducible chromosomal islands (PICIs) inhibit the production of new phages in different ways. PICIs can generate capsules for the phage that are too small to carry the phage DNA, but they can also inhibit the phage terminase small subunit through the action of phage packaging interference (Ppi) proteins, suppressing the phage DNA recognition and packaging [7].



## 2 CRISPR-Cas

CRISPRs were first reported in 1987, when Ishino et al. cloned a gene in *Escherichia coli* and discovered a set of 29 nucleotide (nt) repeats downstream of it. These repeats were separated by similar short sequences. These short sequences in between the repeats are now called spacers. It was noted that these spacers had homology to phage and plasmid DNA. [22] Indeed, spacers are derived from several different genomes of phages and conjugative plasmids and can be integrated in antisense (non-coding) or in sense (coding) orientation in bacterial chromosomes [16]. After more gene and whole-genome sequencing projects, Mojica et al. were able to recognize CRISPRs in many different bacterial species and archaea. In 2002 the abbreviation CRISPR was given to these genetic loci containing spacers and repeats [17].

As mentioned above, the *cas* genes are always near a CRISPR locus (*Figure 2*). The CRISPR-Cas array consists of 21-48 base pairs (bp) direct repeats that are interspaced by non-repetitive spacers of 26-72 bp. The repeats are palindromes that form hairpins in the derived RNA structure and they have a conserved 3'terminus, such as GAAA. These non-repetitive spacers are commonly preceded or followed by 4 to 20 different *cas* genes [10]. The six core genes, *cas1* until *cas6*, were first described in 2008. An universal marker for the presence of CRISPR-Cas in a bacterial or archaeal genome is *cas1*. Other *cas* genes are not always present in the bacterial or archaeal genome [16]. Besides the *cas*-genes, an AT-rich leader is present upstream of the repeats and spacers. The leader contains promoter elements and binding sites for regulatory proteins that are important for the spacer integration, done by Cas1-Cas2 as we shall discuss shortly [18]. It is suspected that the leader helps spacer integration because a new spacer is always added to the CRISPR site near the leader [16].



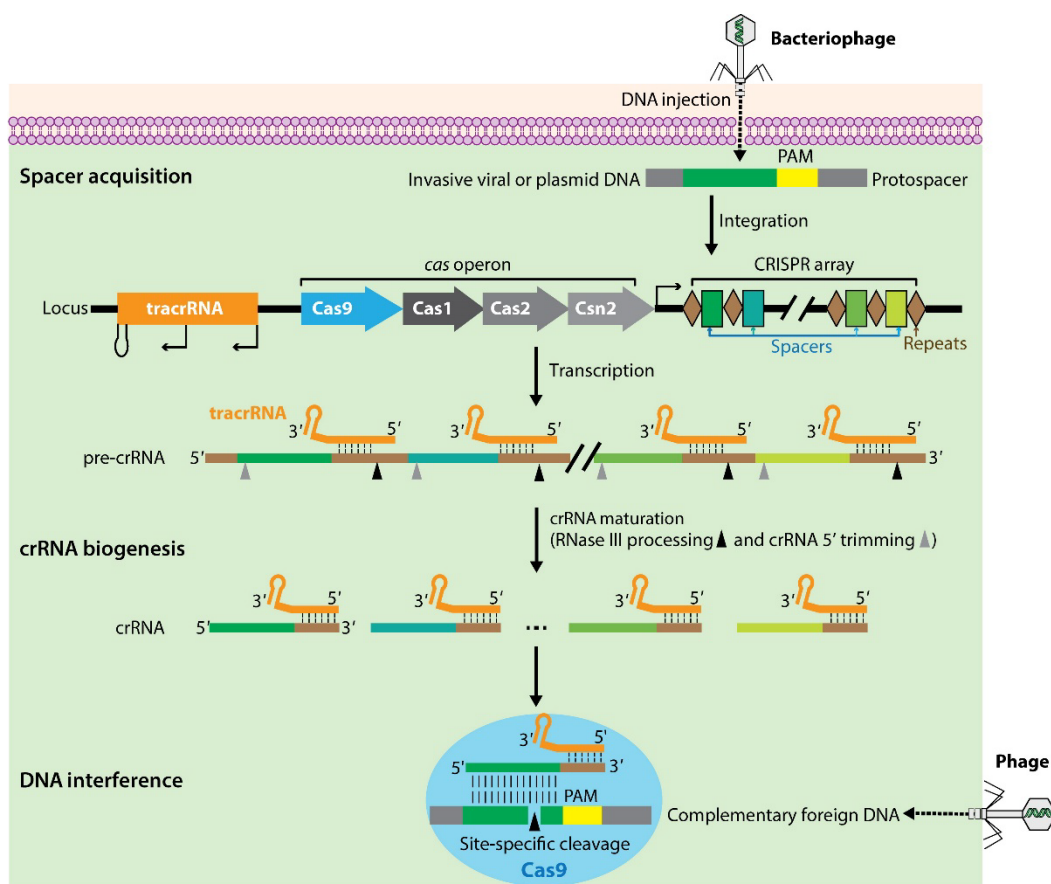
**Figure 2: Overview of a typical CRISPR locus.** The CRISPR repeats are shown in grey with different coloured spacers in between the repeats. The number of spacers can vary, as indicated by the *n* in the purple box. The repeats and spacers are preceded by a leader sequence, the black box. A set of five *cas* genes is following the repeats in this example [17].

### 2.1 CRISPR-Cas9

The most studied of the CRISPR-Cas loci and their mechanisms of action is that of CRISPR-Cas9. CRISPR-Cas9 has ultimately been transformed into a powerful tool for engineering genomes of different organisms [19].

The mechanism of action of CRISPR-Cas9 consists of three different stages (*Figure 3*), namely adaptation; expression; and interference. A bacteriophage can infect the bacteria or a plasmid can get in by conjugation. Both involve foreign DNA that is recognized as non-self by the bacteria. This foreign DNA contains so-called protospacers and a preceding protospacer adjacent motif (PAM). This short motif plays an important role in targeting and degrading foreign DNA. The PAM sequence is crucial to distinguish between self and non-self DNA. First the spacer acquisition takes place, which means adding a new spacer into the CRISPR array in the bacterial genome. Cas1 and Cas2 proteins perform this action. The spacer is needed in a process that ultimately should protect the bacteria against a second exposure to the same foreign DNA (e.g. another attack by the same phage).

The new CRISPR array with the newly inserted spacer is then transcribed into a long precursor CRISPR RNA (pre-crRNA), containing all repeats and spacers in one transcript. In CRISPR-Cas9 systems extra genes are located upstream of the CRISPR spacer-repeat array encoding trans-activating CRISPR RNA (tracrRNA) as seen in *Figure 3*. TracrRNA is transcribed from the CRISPR locus and anneals to the pre-crRNA repeats to form an unique hybrid structure. This structure contains a hairpin because of the palindrome sequence in the *tracr* gene. Subsequently, the biogenesis of CRISPR-RNA (crRNA) starts. The enzyme RNase III cleaves the pre-crRNA into its separate crRNAs while unknown endonucleases trim the 5' end of the crRNA to ultimately result in sequences of 20 nt. During these enzymatic actions, tracrRNA holds the crRNA in place. The crRNA-tracrRNA hybrid complex binds to the Cas9 protein and interferes with the incoming foreign phage or plasmid DNA through specific cleavage by the "RNA-guided" Cas9 endonuclease. Cas9 is only able to cleave the incoming DNA at a nucleotide sequence that is complementary to the 20 nt crRNA to which it is bound and only so when that sequence is immediately preceded by a PAM sequence [9,20].

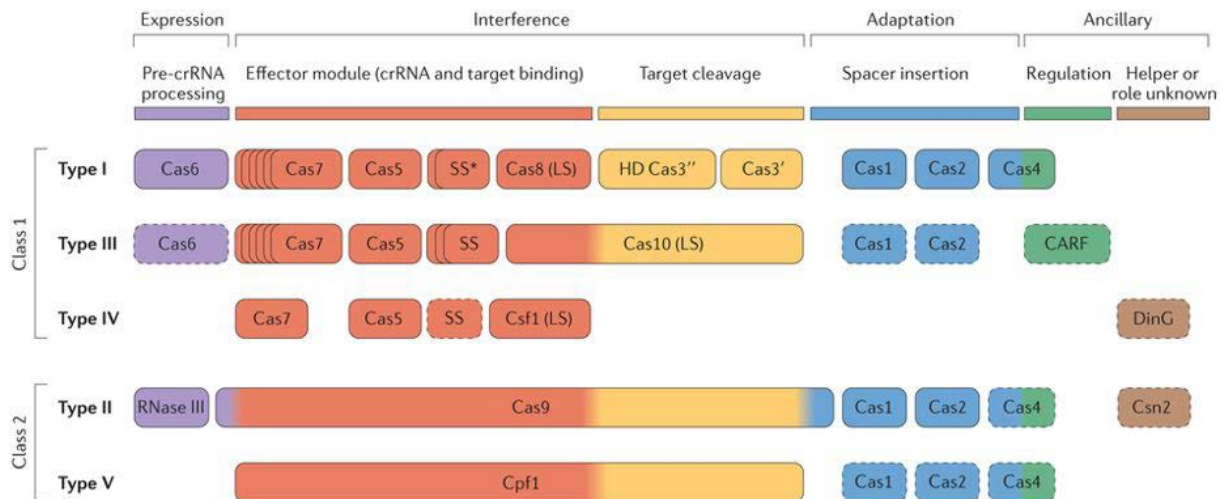


**Figure 3: CRISPR-Cas9 mediated DNA interference in bacterial adaptive immunity.** The three stages are shown on the left. Foreign DNA was derived from a phage that is integrated into the genome. When the bacterium is exposed to the same foreign DNA again it transcribes pre-crRNA and tracrRNA complementary to the foreign protospacer. TracrRNA anneals to the pre-crRNA before crRNA maturation was done with RNase III and crRNA 5' trimming. Cas9 catalysis the DNA interference and is shown in the blue oval [18].

The mechanism of action of the CRISPR-Cas systems is studied by many research groups. There are different CRISPR systems that all give the same result, adaptive protection against invading phage/plasmide DNA, but they may contain different *cas* genes. The systems have been classified in different classes and types. To understand the core of CRISPR-Cas and the different types, it is important to know more about the enzymes that are present in most of the systems and, on the other hand, the differences between them.

## 2.2 CRISPR systems

CRISPR systems have been grouped into five different types on basis of the CRISPR-loci. Using an evolutionary classification, Marakova et al. defined two Classes, five Types and 16 subtypes of CRISPR-Cas systems identified. At the top of the hierarchy are three main Types of CRISPR-Cas systems, as defined by the presence of one of three different proteins. Type I has Cas3, Type II contains Cas9 and Type III is recognizable by Cas10. Besides the Classes and Types, the Cas proteins encoded by the various systems can be divided into functional modules as well, since Cas proteins are involved in the different stages of CRISPR action described above: Adaptation; Expression and Interference; a fourth group entails the Ancillary ones (Figure 4).



**Figure 4: Functional classification of Cas proteins.** The nomenclature and classification is followed by the protein names. The Classes and Types are shown on the left. The color of the protein shows matches the functional module. SS, LS: small large subunit, respectively. An asterisk indicates that the SS is fused to Cas8 in multiple subtypes of type I CRISPR-Cas. Cas3', Cas3'': two variants of Cas3. HD stands for an domain with nuclease activity against ssDNA and ssRNA [20].

Cas proteins are divided into two classes, as characterized and defined by different effector modules. Class 1 systems contain crRNA-effector protein complexes, which are responsible for crRNA and target binding and degradation. These complexes are called CRISPR-associated complex for antiviral defense (Cascade) complex. The Class 2 systems work with a single protein that carries out all functions of the effector complexes. An example is Cas9, which is encoded by a Type II system. Type II and Type V are both Class 2 CRISPR-Cas systems. The effector complex is a small part of an activity module called Interference. The interference module describes the genes involved in crRNA processing, target binding and target cleavage. In Type IV the Csf1 protein was described as an effector module protein. This protein is a Cas8 family protein, of which is by itself active in Type I systems.

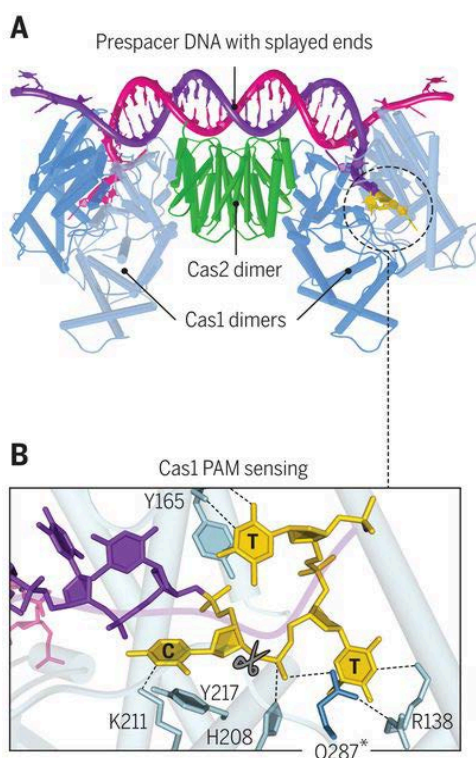
The Expression module refers to the processing of pre-crRNA. This is mostly done by Cas6 or Cas9 and, in Class 2 Type II systems, by RNase III. The Cas6 protein is typical for Type I and III systems. The Adaptation module stands for the insertion of new spacers into the CRISPR array in other words, spacer acquisition. The Cas1-Cas2 protein complex is mainly involved in this process. In Type I, II and V systems, Cas1 and Cas2 are supported by Cas4. The two different colors for Cas4 and the three different colors for Cas9 in Figure 4 indicate that these proteins are active in different stages of CRISPR-Cas immunity responses. The expression module "Ancillary" describes proteins with an unknown, helper or regulatory function [20].

### 3 Cas1 – Cas2

The proteins Cas1 and Cas2, which are both functional and required for spacer acquisition, together form a so-called integration complex. Spacer acquisition includes three main processes: substrate capture, recognition of the CRISPR array, and spacer integration in bacterial chromosome [21]. Cas1 can be divided into two domains (e.g. Cas1a and Cas1b) with both different conformations. The protein forms an asymmetrical dimer containing a N-terminal  $\beta$ -sheet and a C-terminal  $\alpha$ -helical domain. Cas2 also forms a dimer, but in this case it is a symmetrical dimer. The Cas2 dimer forms a bridge between the two Cas1 dimers (*Figure 5a*). Nunez et al. observed contacts of Cas2 with Cas1a, but contacts were not identified between Cas2 and Cas1b. [14] Similar contacts are present on both sides between Cas2 and Cas1a, making the complex symmetrical [15].

#### 3.1 The mechanism of Cas1-Cas2 mediated spacer acquisition

The two different conformations of the two Cas1 domains imply two different biological functions. The Cas1a domains are catalytic subunits responsible for recognition of the PAM sequence and cleaving of the double stranded (ds)-DNA protospacer with the PAM sequence out of the foreign invading DNA [22]. One example of a PAM sequence of Type I is 5'-NTT-3', where N can be any nucleotide. This sequence can differ per CRISPR-Cas Type. The PAM is recognized by one of the PAM-sensing sites in Cas1. Specific amino acid residues of Cas1 form hydrogen bonds with nucleotides of the PAM sequence. Lys211 of Cas1a forms a hydrogen bond with the N nucleotide, in this example a C (*Figure 5b*). Tyr217 of Cas1a and Ile291 of Cas1b help stabilizing the N-nucleotide. His208, Tyr165 and Arg138 of Cas1a form hydrogen bonds with the T's of the PAM. Besides Cas1a, Cas1b Gln287 forms hydrogen bonds with the middle nucleotide of the PAM sequence [21].



**Figure 5: Cas1-Cas2 structure and PAM sensing** **A)** 3D structure of *E. coli* Cas1-Cas2 loaded with a protospacer substrate. **B)** The Cas1 PAM sensing site. The PAM sequence (CTT) is shown in yellow. Light blue: amino acids residues of Cas1a, darker blue (with asterisk): amino acid residue from Cas1b. H, histidine; K, lysine; Q, glutamine; R, arginine; Y, tyrosine [21].

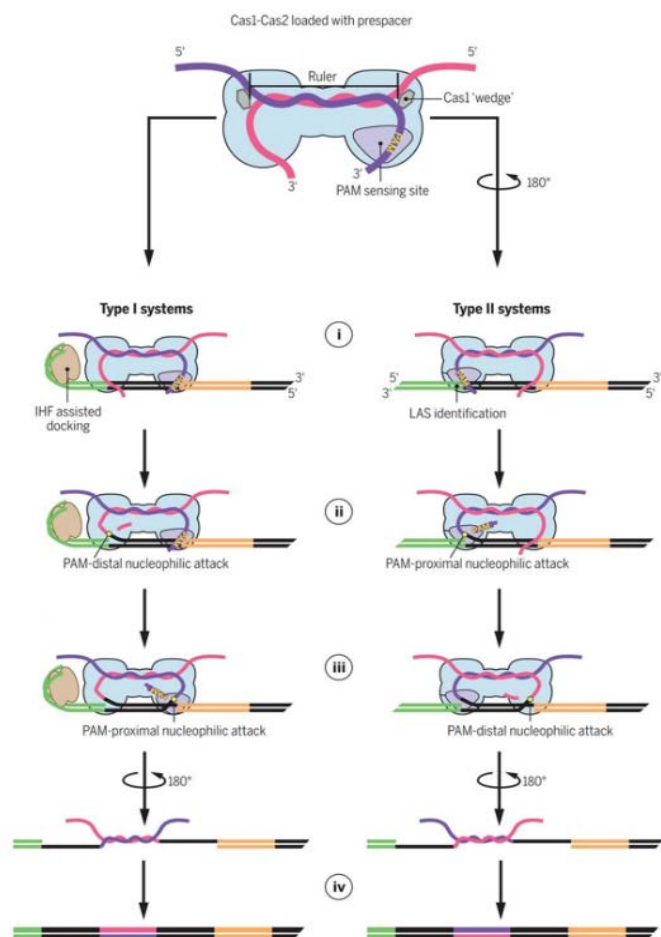
Cas1b is responsible for the formation of the Cas1-Cas2 complex. The interaction between the two monomers involved Arg245 and Arg248 of Cas1b.

As mentioned, Cas2 forms a bridge between the two Cas1 dimers. The bridge made by Cas2 provides a binding surface for the protospacer and the PAM sequence. Cas2 acts like a space holder and ensures that the Tyr22 residues of the two Cas1a domains are exactly 23 nt from each other. This automatically means that the protospacer has a length of 23 nt. The Tyr22 residues are called “wedges”. Cas2 also binds the protospacer of the target DNA by forming hydrogen bonds with the backbone of the DNA duplex. This stabilizes the protospacer DNA when it is bound to the Cas1-Cas2 complex [22].

After recognition of the protospacer with the PAM, the Cas1-Cas2 complex binds to the CRISPR leader and first repeat on the chromosome of the host cell. Subsequently, the Cas1-Cas2 complex integrates the protospacer into the bacterial genome (*Figure 6*). There is a difference in the integration process between Type I and Type II CRISPR systems. Type I Cas1-Cas2 complexes are assisted by the host encoded protein integration host factor (IHF). Type II CRISPR systems use leader-anchoring sites (LAS). Details about these accessory proteins will be discussed later. The Cas1-Cas2 protein complex catalyzes the integration by breaking the chromosomal DNA at two sites on opposing strands at the CRISPR locus. One nucleophilic attack occurs between the leader and repeat while the other takes place between the repeat and the spacer. The first nucleophilic attack most likely occurs at the leader-repeat border, forming a half-integrated intermediate. After the second nucleophilic attack, a ds-DNA break is formed and the protospacer can fully integrate into the CRISPR array [21]. The PAM sequence is cut off in one of the two nucleophilic attacks, depending on which Type of CRISPR system [21]. On both sides, a 5 nt 3'-overhang is generated after cleaving of the PAM sequence, so that ultimately a protospacer of 23 nt with a 3'OH-group was generated [22]. This group is important for the protospacer to integrate into the bacterial chromosome. DNA repair enzymes and polymerases will fill up the gaps and breaks that emerged during the integration, after which the spacer is functional as immunological memory [21].

**Figure 6: Cas1-Cas2-mediated spacer acquisition**

At the top a schematic representation of the Cas1-Cas2 protein complex is shown. It is loaded with the spacer and an active PAM sensing site is colored light purple. The DNA protospacer is colored purple and pink. The Cas1 'wedge' is shown in grey and spacer integration in steps i till iv. The differences in Type I and Type II systems and the accessory proteins IHF and LAS are shown in the steps. The first nucleophilic attack most likely occurs at the boundary between leader (green) and the repeat (black) site and the second nucleophilic attack occurs between the repeat and the spacer (orange) [21].



### 3.2 Accessory proteins in Cas1-Cas2 mediated spacer acquisition

The integration of new spacers into the bacterial genome needs to be very specific to avoid deleterious insertions. The Cas1-Cas2 complex in Type I CRISPR systems is aided by IHF to obtain specific integration [15]. IHF, which contains two subunits  $\alpha$  and  $\beta$ , binds to a site in the CRISPR leader sequence and induces a bend in the DNA of approximately 160 ° directly upstream of the first CRISPR repeat. The bend induced by IHF does not only contribute to the CRISPR-Cas mechanism of action, but in many other biological processes as well, such as DNA replication and transcription. The bend created by IHF suppresses off-target integration. IHF recognizes the sequence 5'-WATCAANNNTTTR-3' (W is A or T, R is A or G and N can be any nucleotide). If this consensus sequence is absent IHF will not bind to the CRISPR leader. Nunez et al. have shown that IHF is required for the spacer acquisition as no spacer was integrated in *E.coli* with a mutated IHF. It is not expected that IHF has any protein-protein contacts with Cas1-Cas2; its primary function is bending DNA, an activity already described in 1988 by Friedman [23].

In Type II systems, a LAS and the first six bp of the repeat are essential for spacer integration and, thus, CRISPR adaption. The LAS is present next to the first repeat of the CRISPR array. It has been shown that shifting of the LAS can lead to double spacer integration or ectopic integration, the integration at another locus in the genome. Changing the first six bp of the repeat will cause the same effect [21]. Type II CRISPR systems, thus, depend on the location of LAS for proper spacer acquisition.

Fusions of the Cas1-Cas2 protein with Cas4 have been observed in several organisms and CRISPR systems. Cas4 is a ring forming protein that is present in Type I, II and V systems. It has nuclease activity and, therefore, it is probably involved with protospacer generation. Cas4 consists of a RecB-like domain [21]. RecB is part of the RecBCD protein complex that fulfills a nuclease activity in homologous recombination [24]. Besides the RecB like domain, Cas4 has four conserved cysteine residues that are probably involved in iron-sulfur cluster coordination [21]. Until a few months ago it was not yet clear what Cas4 does in spacer integration. Last march Knieper et al. published a report in which they show that Cas4 facilitates the integration of spacers, by shortening the spacer length and assisting Cas1-Cas2 to select the correct PAM sequence [25].

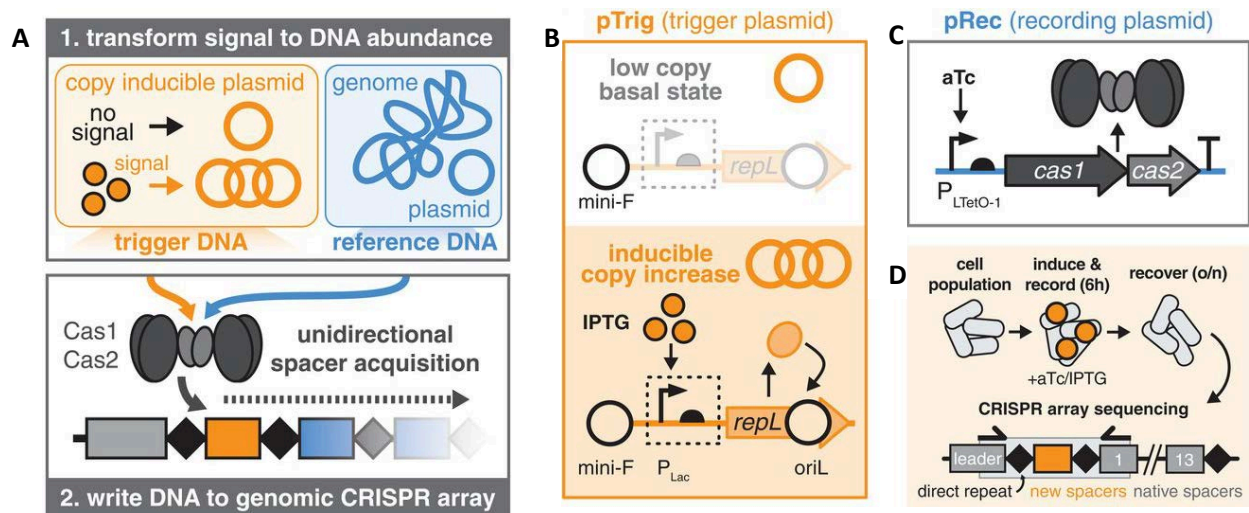
## 4 Applications

The development of recombinant DNA technologies marked a new era in biology and way beyond. Molecular biologists can now engineer and manipulate DNA molecules. This can be and is already used in biotechnology, medical research, drug development and studies in genetic variation. Since the 1970s multiple genome editing technologies have been developed. The immunological memory discussed in this report, CRISPR-Cas, is also useful as a gene editing technique. The most developed CRISPR-cas gene editing technique uses Cas9. This system is most useful as Cas9, as a single protein, mediates target DNA cleavage in contrary to other Cascade complexes. Cascades only denature the DNA before its cleavage. DNA cleavage catalyzed by Cas9 is very specific because of tracrRNA hybridized with crRNA together facilitate the RNA-guided targeting of Cas9 to the proper site on the DNA to be cleaved. The crRNA bound by Cas9 can in principle be matched with any DNA, also in living eukaryotic cells, after the Cas9 complex recognizes a PAM sequence. Subsequently, the DNA of the cell will be cleaved at that position. This double strand DNA break can be repaired by the cell, but mutations can occur. This technique could be used in gene therapy, by replacing mutated genes with correct ones. This is done by adding a template DNA carrying the desired gene sequence to the CRISPR-Cas9 reaction that takes place in the (diseased) host cell. This template will pair up with the cut ends of the Cas9 endonuclease reaction and introduces the template sequence into the genome. This way of repairing double stranded breaks is called homology-directed repair [13]. In similar ways genome changes could be made in plants to make them e.g., resistant against diseases. Besides that it could be used in microbiology research and hopefully in the future in human cells to cure genetic diseases or treat cancer.

Instead of engineering the DNA of an organism with Cas9 and thus its genetics, it is also a perception to add whole (immunity) genes or gene fragments to organisms, which is the key feature of CRISPR-Cas mediated immunity. Cas1-Cas2 protein complex is an interesting complex that could be used to integrate foreign genes with an adjacent PAM sequence into host DNA. A complete molecular understanding of the specificity rules of CRISPR-Cas spacer acquisition and integration into the target site could definitely contribute to the development of new genome engineering technologies and provide immunity to organisms. Changing and/or evolving the target specificity of Cas1-Cas2 protein complex and expressing it may help in the development of different gene editing methods [26].

Since more knowledge was obtained of the Cas1-Cas2 protein complex structure, its specificity and the mechanism by which spacers are acquired, further research has allowed developing new techniques. A recent study showed how Cas1-Cas2 could be used as a system to record molecular events into the genome [27]. Recording the molecular and environmental changes has many possibilities, for example in determining the presence of metabolites and regulatory changes. Sheth et al. reported an *E. coli* strain that recorded and stored information in its DNA on the presence of biological signals over time. This system, called Temporal Recording in Arrays by CRISPR Expansion (TRACE), is a combination of the CRISPR-Cas mediated immunity and a trigger-induced copy number increase of a so-called trigger plasmid (Figure 7). The trigger plasmid (pTrig) contains an inducible promoter driving the gene encoding a lytic bacteriophage replication protein. In presence of the trigger, in this case isopropyl- $\beta$ -D-1-thiogalactopyranoside (IPTG), transcription from the Lac promoter increases and results in expression of the replication protein. In addition, an inducible plasmid expressing Cas1 and Cas2 upon addition of anhydrotetracycline (aTc), the recording plasmid (pRec), was also present. The CRISPR-Cas spacer acquisition machinery is used to record the amount of trigger plasmid DNA that was added to the CRISPR-array.

The environmental trigger initiates replication of the trigger plasmid, the incorporation of trigger plasmid DNA into the CRISPR array and the expression of Cas1-Cas2 protein. The length of the CRISPR-arrays determines the level of trigger and temporal information, which can be determined with CRISPR array sequencing. Sixteen different signal profiles were assessed and half of them were perfectly classified. Although technical improvements are needed, the TRACE system sets high expectations for massive parallel recording of different biological signals [28].



**Figure 7: Temporal recording in arrays by CRISPR expansion.** **A)** TRACE systems function by transforming a biological signal to DNA abundance of trigger DNA. This trigger DNA differs from the reference DNA and is then recorded as spacers into CRISPR arrays of a cell **B)** pTrig is a trigger plasmid with mini-F origin for stable maintenance.  $P_{Lac}$ , Lac promotor is inducible by IPTG.  $repl$ , replication system of phage P1 that causes copy number increase **C)** pRec is a recording plasmid.  $P_{LTetO-1}$  is an aTc inducible promotor, which is responsible for the expression of Cas1 and Cas2 **D)** Experimental overview of the induction and CRISPR array sequencing. o/n, overnight [28].

## 5 Discussion and conclusion

CRISPR-Cas has been developing fast in recent years. It is an adaptive bacterial and archaeal immune system and memory of origin. The adaptive immunological memory reviewed here, was discussed earlier in a lot of reviews and reports, since CRISPR-Cas became one of the most promising genome editing techniques. Although our understanding of all facets of CRISPR-Cas is not yet complete, more and more information about the proteins, the different systems and mechanisms is published through time. In this thesis Cas9 and the Cas1-Cas2 protein complex, which catalyzes the addition of new spacers into the host genome, were discussed. This protein is also called the “workhorse” of the system [21].

As said, Cas1-Cas2 is an important protein complex, since it is present in almost all CRISPR-Cas systems, but a lot more research needs to be done to use Cas1-Cas2 in possible editing applications. It can integrate foreign gene fragments with an PAM sequence into the host organism. This protein complex could be very important in gene therapy and gene editing techniques. It is important to conduct more research on spacer integration by the Cas1-Cas protein complex. After integration, the fragments are probably not directly functional and side effects are not yet investigated. Which promoters need to be used, or are integrated fragments transcribed, are important questions if Cas1-Cas2 will be used as a genetically editing method. Another thing to take into account is the evolution of bacteriophages. Phages can overcome CRISPR-Cas interference by only acquiring one single mutation in or around the protospacer sequence [17].

Research to CRISPR-Cas is ongoing and a lot of positive feedback in the popular news. Unfortunately, this promising system also had some negative feedback the last year. Genome editing with Cas9 is not that specific as people thought it would be. The use of CRISPR-Cas could lead to deletions extending over many kilobases and complex rearrangements. Kosicki et al. showed this phenomenon in mouse embryonic stem cells, but in a human differentiated cell line as well. These observed deletions and rearrangements have many pathogenic consequences for organisms [29]. Haapaniemi et al. reported that genome editing with CRISPR-Cas causes a p53 DNA-damage response that, in theory, can initiate tumors [30]. The Scientific American published an article about the side effects of CRISPR-Cas gene editing and reported the initiation of tumors as well. Besides this they mention that after publishing Haapaniemi et al. their article in Nature Biotechnology, the market value of three CRISPR-Cas companies was reduced by 300 million dollars [31].

A big discussion is continuing about the genetically modified organisms (GMO) produced with CRISPR-Cas editing techniques. The European court decided last month that CRISPR-Cas techniques are GMO producing methods. Proponents of the CRISPR-Cas technology think that this technique should not fall under current European Union (EU) GMO legalization. After all, the gene mutations that occur are naturally and no foreign DNA was introduced by CRISPR-Cas9. This could be very important to e.g., develop plant variants that are resistant against diseases [32]. Environmental organizations, such as Greenpeace, are against the legalization and usage of CRISPR-Cas gene editing. The use of Cas9 does not insert foreign DNA, but any future insertions with Cas1-Cas2 does. The usage of this protein complex will bring up more discussions and will probably be labeled as a GMO producing technique. This is not favorable for the technique itself and the usage of it, since the modified organisms most of the time cannot be used or consumed in the EU.

Altogether, the discovery of gene editing methods, specifically CRISPR-Cas and its application, are very important for the future of biomedical and biotechnological techniques and work methods. Understanding CRISPR-Cas and his future potential is only in the initial phase. Researchers anticipate extensive further developments in this field. Knowing all the ins and outs of CRISPR-Cas and the Cas1-Cas2 protein complex will provide more insight into understanding different biological systems and eventually new genetic engineering techniques [33].

## References

- [1] Salmond GPC, Fineran PC. A century of the phage: past, present and future. *Nature Reviews Microbiology*. 2015;13:777.
- [2] Samson JE, Moineau S. Bacteriophages in food fermentations: new frontiers in a continuous arms race. *Annual review of food science and technology*. 2013;4:347-68.
- [3] Suttle CA. Marine viruses--major players in the global ecosystem. *Nature reviews Microbiology*. 2007;5(10):801-12.
- [4] Criscuolo E, Spadini S, Lamanna J, Ferro M, Burioni R. Bacteriophages and Their Immunological Applications against Infectious Threats. *Journal of immunology research*. 2017;2017:3780697.
- [5] Lin DM, Koskella B, Lin HC. Phage therapy: An alternative to antibiotics in the age of multi-drug resistance. *World journal of gastrointestinal pharmacology and therapeutics*. 2017;8(3):162-73.
- [6] Teesalu T, Sugahara KN, Ruoslahti E. Mapping of vascular ZIP codes by phage display. *Methods in enzymology*. 2012;503:35-56.
- [7] Seed KD. Battling Phages: How Bacteria Defend against Viral Attack. *PLoS pathogens*. 2015;11(6):e1004847.
- [8] Chopin MC, Chopin A, Bidnenko E. Phage abortive infection in lactococci: variations on a theme. *Current opinion in microbiology*. 2005;8(4):473-9.
- [9] Horvath P, Barrangou R. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science*. 2010;327(5962):167-70.
- [10] Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*. 2010;8:317.
- [11] Dibitetto D, La Monica M, Ferrari M, Marini F, Pellicoli A. Formation and nucleolytic processing of Cas9-induced DNA breaks in human cells quantified by droplet digital PCR. *DNA repair*. 2018;68:68-74.
- [12] Xiao Y, Ng S, Nam KH, Ke A. How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature*. 2017;550:137.
- [13] Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014;157(6):1262-78.
- [14] Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW, Doudna JA. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature structural & molecular biology*. 2014;21(6):528-34.
- [15] Wright AV, Liu JJ. Structures of the CRISPR genome integration complex. *Science*. 2017;357(6356):1113-8.
- [16] Sorek R, Kunin V, Hugenholtz P. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature reviews Microbiology*. 2008;6(3):181-6.
- [17] Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews Genetics*. 2010;11:181.
- [18] Rath D, Amlinger L, Rath A, Lundgren M. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie*. 2015;117:119-28.
- [19] Jiang F, Doudna JA. CRISPR-Cas9 Structures and Mechanisms. *Annual Review of Biophysics*. 2017;46(1):505-29.
- [20] Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*. 2015;13:722.
- [21] Jackson SA, McKenzie RE, Fagerlund RD, Kieper SN, Fineran PC, Brouns SJJ. CRISPR-Cas: Adapting to change. *Science*. 2017;356(6333).
- [22] Wang J, Li J, Zhao H, Sheng G, Wang M, Yin M, et al. Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell*. 2015;163(4):840-53.
- [23] Nuñez James K, Bai L, Harrington Lucas B, Hinder Tracey L, Doudna Jennifer A. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Molecular Cell*. 2016;62(6):824-33.

- [24] Ajiboye TO, Skiebe E, Wilharm G. Contributions of RecA and RecBCD DNA repair pathways to the oxidative stress response and sensitivity of *Acinetobacter baumannii* to antibiotics. *International journal of antimicrobial agents*. 2018.
- [25] Kieper SN, Almendros C, Behler J, McKenzie RE, Nobrega FL, Haagsma AC, et al. Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. *Cell Reports*. 2018;22(13):3377-84.
- [26] Moch C, Fromant M, Blanquet S, Plateau P. DNA binding specificities of *Escherichia coli* Cas1-Cas2 integrase drive its recruitment at the CRISPR locus. *Nucleic acids research*. 2017;45(5):2714-23.
- [27] Shipman SL, Nivala J, Macklis JD, Church GM. Molecular recordings by directed CRISPR spacer acquisition. *Science*. 2016;353(6298):aaf1175.
- [28] Sheth RU, Yim SS, Wu FL, Wang HH. Multiplex recording of cellular events over time on CRISPR biological tape. *Science*. 2017;358(6369):1457-61.
- [29] Kosicki M, Tomberg K, Bradley A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*. 2018;36:765.
- [30] Haapaniemi E, Botla S, Persson J, Schmierer B, Taipale J. CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nature Medicine*. 2018;24(7):927-30.
- [31] Begley S. Potential DNA Damage from CRISPR "Seriously Underestimated," Study Finds STAT2018 [cited 2018 8-8-2018]. Available from: <https://www.statnews.com/2018/07/16/crispr-potential-dna-damage-underestimated/>.
- [32] Bouwmeester R. Europees Hof: CRISPR-Cas valt onder gmo-wetgeving Nieuwe Oogst2018 [updated 7-8-2018]. Available from: <https://www.nieuweoogst.nu/nieuws/2018/07/25/europees-hof-crispr-cas-valt-onder-gmo-wetgeving>.
- [33] Schmidt F, Platt RJ. Applications of CRISPR-Cas for synthetic biology and genetic recording. *Current Opinion in Systems Biology*. 2017.

