

Essay

**The prokaryotic viral defense system CRISPR/Cas – a source
for applications in research and biotechnology**

Master programme
Molecular biology and biotechnology

Molecular genetics

January 2014

Student: Irina Lucia Schmidt (S2339935)

Supervisor: Prof. Dr. Jan Kok

Table of contents

The prokaryotic viral defense system CRISPR/Cas – a source for applications in research and biotechnology

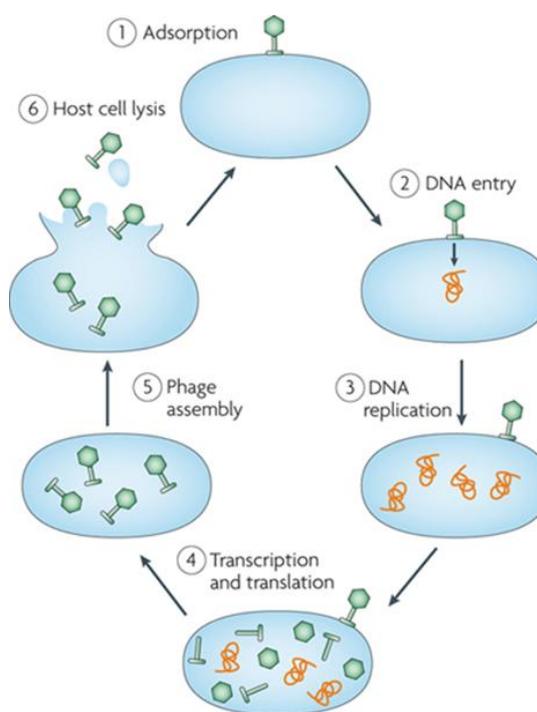
1. Abstract.....	1
2. Introduction.....	1
3. CRISPR/Cas components and mode of action	4
3.1. CRISPR/Cas components	5
3.2. CRISPR/Cas mode of action	6
4. Classification of CRISPR/Cas systems I-III.....	11
4.1. Common features of Type I CRISPR/Cas systems	13
4.2. Common features of Type II CRISPR/Cas systems	14
4.3. Common features of Type III CRISPR/Cas systems.....	15
5. Eukaryotic RNA silencing: a concept analogous to CRISPR/Cas	16
5.1. Employing eukaryotic RNA silencing in research	20
6. CRISPR/Cas applications in research.....	21
7. CRISPR/Cas in biotechnology.....	27
7.1. CRISPR/Cas as a protection mechanism for large-scale bacterial fermentation processes	27
7.2. Applications in the pharmaceutical industry	28
8. Outlook and conclusions – future perspectives for CRISPR/Cas-based applications	29
8.1. Next steps in CRISPR/Cas-based developments	29
8.2. Perspectives for genome editing and gene therapy treatment.....	30
8.3. Perspectives for the protection of large-scale fermentations	31
9. References.....	32

1. Abstract

There are several systems established to defend prokaryotic cells against invaders. Only recently an adaptive immune system against invading nucleic acids, such as viruses or conjugative plasmids, was discovered. This system is called CRISPR/Cas (Clustered regularly interspaced short palindromic repeats/CRISPR associated (proteins)). Although it was since then extensively genetically and biochemically studied, there are still many processes that we do not fully understand and the diversity of CRISPR/Cas makes research even more challenging. However in this short time many practical applications for the CRISPR/Cas system, or certain parts of it, have emerged. Some of them are going to be used in research and some in biotechnology. Some of these applications are already in use and some of them are only conceivable. An important question is if these techniques are going to be seminal and what advantages these methods have over these that were in use before (in case there were any) and what alternatives there are and how long these methods are going to be in use before another method is going to replace a CRISPR/Cas based method.

2. Introduction

Prokaryotes (Bacteria and Archea) have various defense mechanisms to protect themselves against invading DNA. Probably, the most dangerous kind of invading DNA is that of viruses. Viruses are found everywhere and are generally considered the most abundant and diverse biological entities on Earth (Bergh *et al.*, 1989). Bacteriophages are viruses that infect bacteria in order to replicate. When reproduction is completed they burst out of the host, killing the cell (Figure 1). Thus, from a bacterial perspective, phages pose a persistent lethal threat. Similar things happen in archea, although the terminology here is not clear: In the last 40 years, approximately 20 different names have been used to describe viruses that infect archea. (Abedon and Murray, 2013). For simplicity I



Nature Reviews | Microbiology

Figure 1 Bacteriophages in action. The phage attaches to the cell (1) and injects genetic information (2) The foreign DNA is replicated several times (3) and proteins for new phage particles are produced (4) . Once replication and production of phage parts is completed new phages are assembled in the cell (5) cell lysis causes the fully assembled phages to be released (6) (Taken from: Labrie *et al.*, 2010)

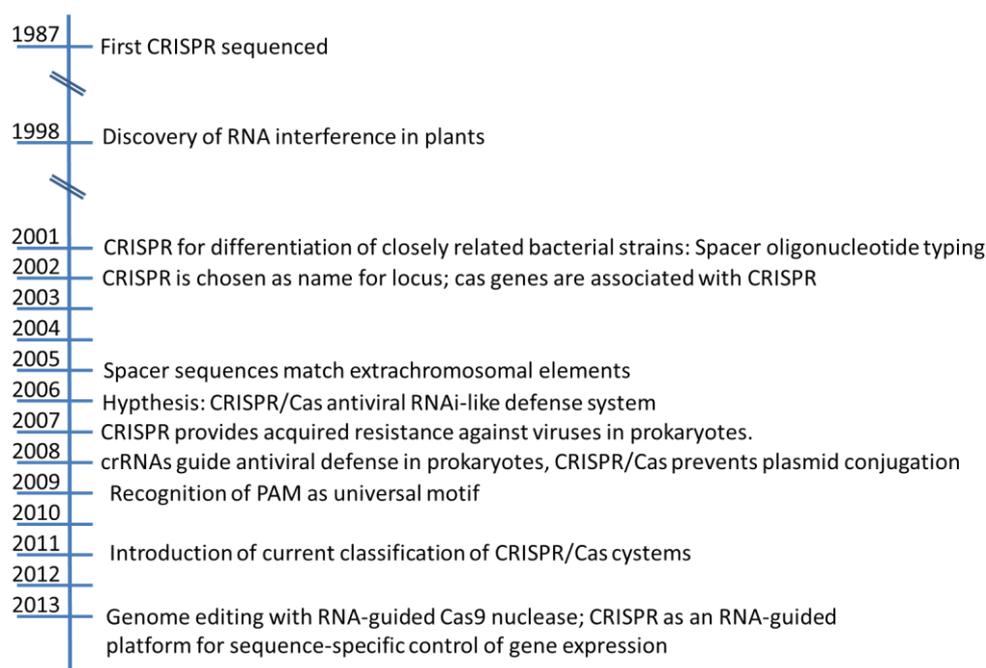


Figure 2 Timeline with important discoveries in CRISPR/Cas, all events are mentioned throughout the different chapters in this review

will mostly refer to the term “virus” as umbrella term for bacteriophages and viruses which infect archaea.

Most systems used in bacteria to fight phages are unspecific, such as the prevention of absorption or blocking of injection and restriction modification systems, which attacks foreign nucleic acids to name just a few (Horvath and Barrangou, 2010). Prokaryotes also have one specific defense mechanism that adapts to new infections and remembers old ones. This system is called CRISPR/Cas, an abbreviation for Clustered regularly interspaced short palindromic repeats/CRISPR-associated (proteins).

Box 1 For more up to date information about bacterial defense systems and how phages manage to by-pass read: Samson *et al.* (2013)

This clumsy abbreviation does cover the most important properties of the peculiar locus in the genome that was first described in 1987 by Ishino *et al.* who had sequenced the first CRISPR/Cas in *E. coli*. In old papers CRISPR/Cas systems are also named short regularly spaced repeats (SRSR). Ishino *et al.* (1987) described a cluster of short of repetitive sequences (repeats) that were intercepted by short differing sequences (called spacers) and did not seem to code for a protein. Later, protein coding sequences in close proximity to the CRISPR locus (*cas* genes) were identified by Jansen *et al.* (2002). Although similar clusters have been identified in various organisms in the meantime, the role of CRISPR/Cas remained elusive to researchers for over 20 years (Figure 2). Two major breakthroughs were achieved in 2005 and 2007 when it was

found out by different laboratories that the spacer sequences correspond to unique sequences in viral DNA (Bolotin *et al.*, 2005; Mojica *et al.*, 2005). These sequences could prevent viral infection if the spacer sequence corresponded to a sequence in the viral DNA (Figure 3). Consequentially, loss of the genetic element led to loss of immunity (Makarova *et al.*, 2006; Barrangou *et al.*, 2007). The same time the importance of noncoding RNA in eukaryotes was revealed and studied intensively and it was discovered that noncoding RNA is also present in prokaryotes. This newly acquired knowledge contributed equally to the identification of the role of CRISPR/Cas as did the increase in computational power and the improvements in DNA and RNA sequencing techniques.

Since the function of CRISPR loci in prokaryotes was solved in 2007 many studies were performed to unravel the molecular mechanisms of the system. In simple terms, CRISPR/Cas based immunity is achieved in three stages: acquisition of foreign DNA, CRISPR RNA (crRNA) biogenesis, and target interference (Figure 3). Short stretches of invasive nucleic acids are inserted into a CRISPR locus. These “spacers” serve as immune markers. Upon infection the whole CRISPR locus is transcribed. Interestingly it is still elusive, how an infection triggers transcription of the CRISPR locus. Subsequently the long RNA precursor is processed into small noncoding interfering CRISPR RNAs (crRNAs). The

crRNAs form a ribonucleoprotein complex with Cas proteins to guide them toward target nucleic acids for specific cleavage of homologous sequences. crRNA is used by Cas (CRISPR-associated) proteins to target and subsequently degrade invading nucleic acids of in a sequence-dependent manner.

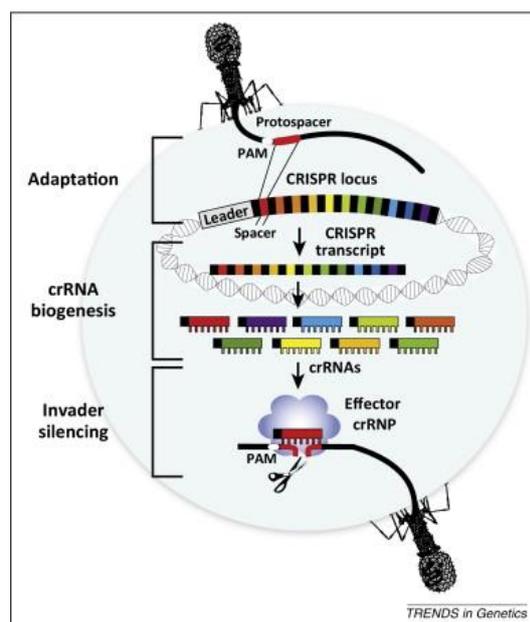


Figure 3 Overview of the three stages of CRISPR/Cas based immunity (Taken from: Terns and Terns, 2014)

Not all prokaryotes have a CRISPR/Cas system. Large sequencing projects suggest that about 90% of archeal strains and 50% of the bacterial strains have a CRISPR/Cas system. Identified CRISPR locus are stored in an online database (<http://crispr.u-psud.fr/crispr/CRISPRdatabase.php>).

Viruses on the other hand have developed strategies to bypass CRISPR/Cas systems (Box 1). This continuous co-evolution between viruses and their hosts has led to a great variety in CRISPR/Cas systems. CRISPR/Cas systems are classified into

three major Types: I, II and III. , based on the phylogeny and presence of particular Cas proteins (Makarova *et al.*, 2011b). Classes are further divided into subtypes. The large variety in CRISPR/Cas systems causes an almost complete lack of conserved elements in CRISPR/Cas systems, which still challenges researchers and is one of the reasons why many processes of the CRISPR/Cas system are still elusive, although CRISPR/Cas systems have been studied intensively during the last six years.

Nevertheless, many ideas have emerged as to how this system or some of its components could be used and the first steps have been taken towards in the development of several applications. Exciting papers about developments and new CRISPR/Cas applications are published every month. Some of these ideas could revolutionize the fields of biotechnology and medical research if they could indeed be applied and established successfully. CRISPR/Cas could be used to immunize bacterial strains in large-scale fermentations, it could be employed for genome engineering purposes and it could hold the key for future gene therapy treatment in humans.

In this essay I will focus on applications that are derived from CRISPR/Cas-based components. It is intended to explain how these applications work, what they are used for and to give an insight into future applications. Chapter 3 is designed to give general information on the CRISPR/Cas system . Therefore the CRISPR/Cas

components and how they are arranged is going to be described in more detail. Furthermore it will describe the state of the art of the CRISPR/Cas mode of action and the regulation of the genes involved. Chapter 4 deals with the classification of CRISPR/Cas systems. Similarities and differences in these systems will be the focus of this chapter but evolutionary aspects will also be covered. Chapter 5 is gives a brief insight into the broad field of small noncoding RNA ,ncRNA in eukaryotes. Although they are analogous to prokaryotic small noncoding RNAs some eukaryotic ncRNAs are surprisingly similar to crRNAs in their biogenesis and function. Another important aspect explained in this chapter is how these eukaryotic ncRNAs are applied in research. Chapter 3 to 5 are intended to give the basis that will allow to understand Chapter 6, 7 and 8. In these chapters CRISPR/Cas-based applications are will be described and discussed with a special focus on recent developments. Chapter 8 deals with advantages and disadvantages of the CRISPR/Cas-based methods. They will be compared to existing alternative methods and all will be critically discussed.

3. CRISPR/Cas components and mode of action

As already mentioned in the introduction, CRISPR/Cas components and mode of action vary depending on the system and the

set of Cas proteins used. This chapter is supposed to give a general overview. Exceptions in components and mode of action and the names of specific proteins are left out on purpose but are going to be further discussed in chapter 4.

3.1. CRISPR/Cas components

All CRISPR/Cas systems are composed of three vital elements, a leader, the CRISPR array, and Cas proteins, coded by genes upstream or downstream of the CRISPR array (Figure 4). The **leader** sequence is an AT-rich sequence of 200-500 base pairs and includes the promoter for the transcription of the CRISPR array (Pul *et al.*, 2010). Furthermore the leader sequence is important for the acquisition of new spacers into the CRISPR array (Yosef *et al.*, 2012).

The **CRISPR array** consists of repeat and spacer sequences. **Repeats** are arranged as direct repeats or as palindromes and are typically 23-47 base pairs long. After transcription of the CRISPR array the repeats allow the transcript to form stable secondary stem-loop structures. The repeat sequences are located in one locus and are

highly conserved sequences, thus repeats within one array are usually identical in length and sequence. If an organism contains several CRISPR arrays the repeats are different in every array. Despite the extreme diversity of CRISPR repeat sequences, most repeats have a conserved GAAA(C/G) motif at the 3' end, which may serve as a binding site for one or more of the conserved Cas proteins (Kunin *et al.*, 2007). **Spacers** are short variable sequences of 21-72 base pairs, which separate the repeats (Jansen *et al.*, 2002). Spacer sequences correspond to viral or plasmid sequences. CRISPRs serve as a memory of prior exposure to the cell of invading DNA (Pourcel *et al.*, 2005; Mojica *et al.*, 2005; Bolotin *et al.*, 2005). Therefore, the spacer sequences are mostly unique within one genome and they can be used as well to identify closely related species. The size of a CRISPR array can vary greatly. The largest CRISPR array up to now identified comes from in *Haliangium ochraceum* DSM 14365: It contains 587 repeats (Zhang *et al.*, 2005). However, the size of a CRISPR locus is limited. Occasional loss of repeat-spacer units has been observed. These deletions usually occur toward the trailer end

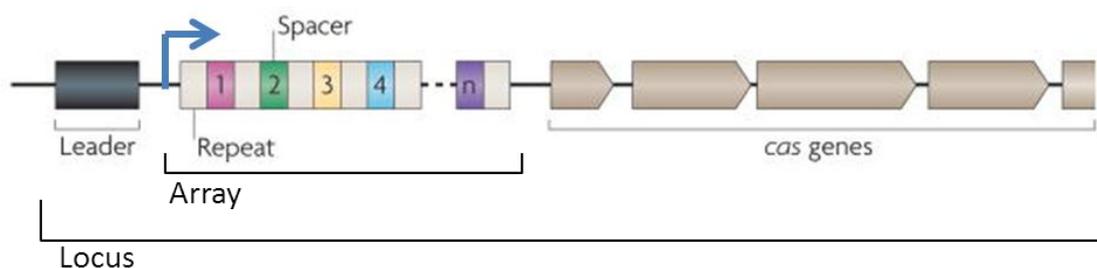


Figure 4 General features of CRISPR loci (Adapted from: Marraffini and Sontheimer, 2010)

(opposite the leader sequence) of the CRISPR locus, possibly supporting preferential elimination of “outdated” spacers that target ancient phages or plasmids while maintaining the more contemporary arsenal of spacers at the leader end. The efficiency of spacer acquisition differs between different arrays (Hovarth *et al.*, 2008). Furthermore it is possible that one organism harbors several CRISPR arrays of different sizes in its genome. In this case the efficiency of spacer acquisition is correlated to the size of the CRISPR arrays (Hovarth *et al.*, 2008).

Cas-genes are located in close proximity to the CRISPR array. The encoded Cas proteins provide the enzymatic machinery that is essential for acting upon invasion of foreign DNAs. Among other tasks they mediate the integration of new CRISPR elements and degradation of invasive DNA. The set of Cas proteins used to provide this machinery varies between the different CRISPR/Cas types and subtypes. Although the total number of proteins differs between the various systems, the functional principles of all CRISPR/Cas systems are similar. Typical functional domains of Cas proteins include those for helicase, polymerase, nuclease activity and for polynucleotide-binding (Haft *et al.*, 2005). Because most of the functional domains have nucleic acid related functions Cas proteins were falsely thought to be involved in DNA repair (Makarova *et al.*, 2002). Up to 45 different protein families have been identified to be

commonly associated with CRISPRs. It is possible that unexpected functions may still reside within Cas proteins that are yet to be characterized.

Six of *cas* genes (*cas1–cas6*) are widely conserved and are considered core *cas* genes, but only *cas1* and *cas2* are universally conserved in genomes that contain CRISPR loci (Haft *et al.*, 2005). Because of that Cas1 and Cas2 are important proteins for studying the evolution of the CRISPR/Cas types and subtypes (Makarova *et al.*, 2011b).

3.2. CRISPR/Cas mode of action

The response to a bacteriophage infection can be divided into three stages: **adaptation**, **expression** and **interference** (Figure 5).

During the **adaptation** process a new intruder DNA is being recognized as such and one or several new repeat-spacer elements complementary and unique to this phage are added to the CRISPR cluster. The first clues about the existence of the adaptation process came from a simple experiment where bacteria were deliberately infected with phages. Bacteria who survived the infection were screened for changes in the CRISPR locus. It was found that all bacteria who had survived the infection had acquired new sequences. Deletion of these newly acquired sequences caused loss of immunity against the phage (Barrangou *et al.*, 2007). The sequences in the foreign

genome from which spacers are derived are In *E. coli* it was recently shown that selection of the protospacer the intruding DNA does not occur randomly but seems to be biased to special motifs in the foreign DNA that are recognized by the proteins complex facilitating of new spacers acquisition (Yosef *et al.*, 2013). These motifs are sequences flanking the protospacer and are commonly referred to as protospacer adjacent motifs (PAMs). The PAM is 2-5 bp long and can be located upstream or downstream of the protospacer (Figure 6 A). The sequence itself is dependent on the type of CRISPR/Cas system and the organism in which it is located. Although the role of the PAM is still vague it is believed to be

termed protospacers (Deveau *et al.*, 2008).

important for motif recognition by the proteins which facilitate cleavage of the intruding DNA and integration of the protospacer into the host genome (Figure 6 B). The PAM is not only relevant for selection of the protospacer but also plays a crucial role at the interference stage discussed later in this chapter.

Comparative analyses have shown that spacer sequences nearest to the leader are most diverse, whereas repeats farthest from the leader are often more conserved among closely related species. This led to the conclusion that addition of a protospacer into the genome occurs in a polarized manner with new elements being integrated

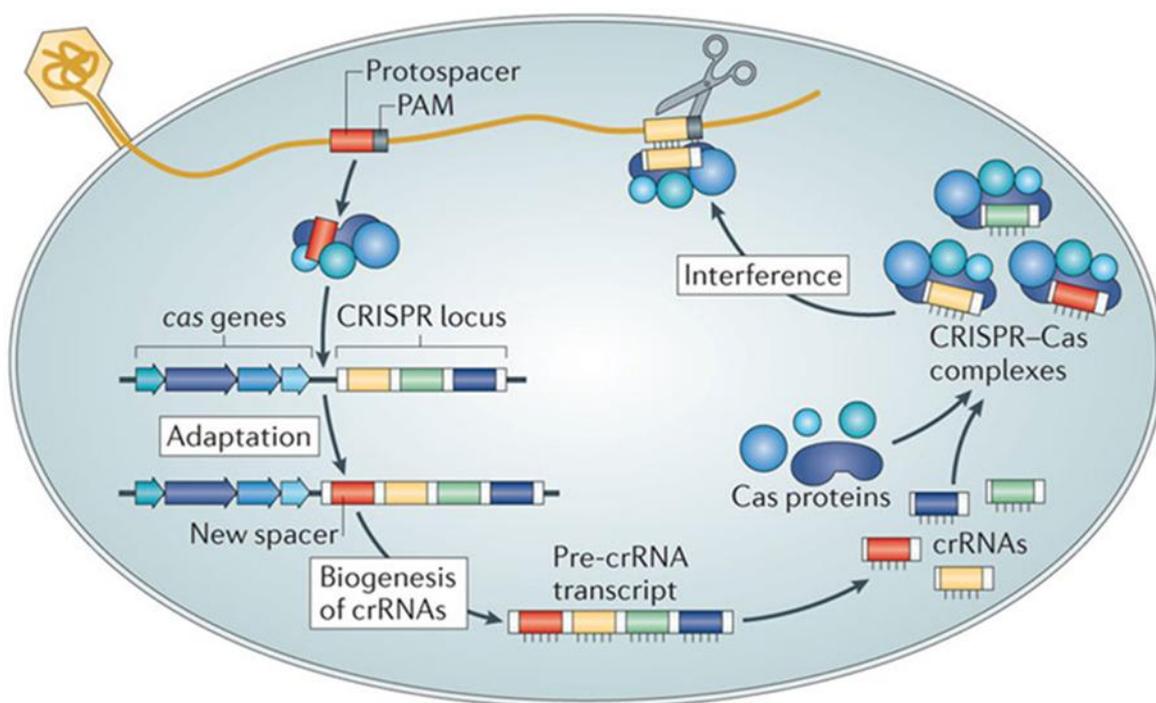


Figure 5 showing the three stages Adaptation, Expression (Biogenesis of crRNAs) and Interference. Adaptation: upon infection the phage releases its DNA, which is recognized by Cas proteins and included as a new spacer into the CRISPR array. During the expression stage the CRISPR array is being transcribed and mature crRNA is generated. During the interference stage the mature crRNAs guide the Cas proteins to identify the invading DNA and degrade it. (Taken from: Samson *et al.*, 2013)

at the AT-rich leader site of the CRISPR array (Pourcel *et al.*, 2005) (Figure 5). Consequently a CRISPR array is a chronological record of past phage infections. The role of the leader sequence is not exactly known, although it has been shown that its sequence is important for adaptation. Possibly the leader is required for repeat duplication and/or spacer integration. This step is essential to immunity and only cells that adapt (integrated a new spacer into the CRISPR array) will survive the infection (Barrangou *et al.*, 2007).

The adaptation process is supposed to be more conserved within the different CRISPR/Cas types than the expression and interference processes. This is supported by the fact that best conserved Cas proteins,

Cas1 and Cas2, are essential for spacer acquisition in the *E. coli* CRISPR/Cas system I-E. Still, many details of the adaptation process are still elusive.

During the **expression** stage small crRNAs (CRISPR RNAs) are generated (Figure 5). As mentioned in Chapter 3.1.. It was demonstrated that *cas* genes in *E. coli*, *S. thermophilus* and *Thermus thermophilus* are upregulated upon phage infection. However activation of CRISPR/Cas is induced by different factors. The heat-stable nucleoid-structuring protein H-NS was identified as a key regulator of the CRISPR/Cas system of *E. coli*. H-NS binds to the promoter of the CRISPR locus and suppresses the expression of *cas* genes and the CRISPR array (Pul *et al.*, 2010). Phage-encoded H-NS and conjugative plasmid-encoded H-NS paralogs

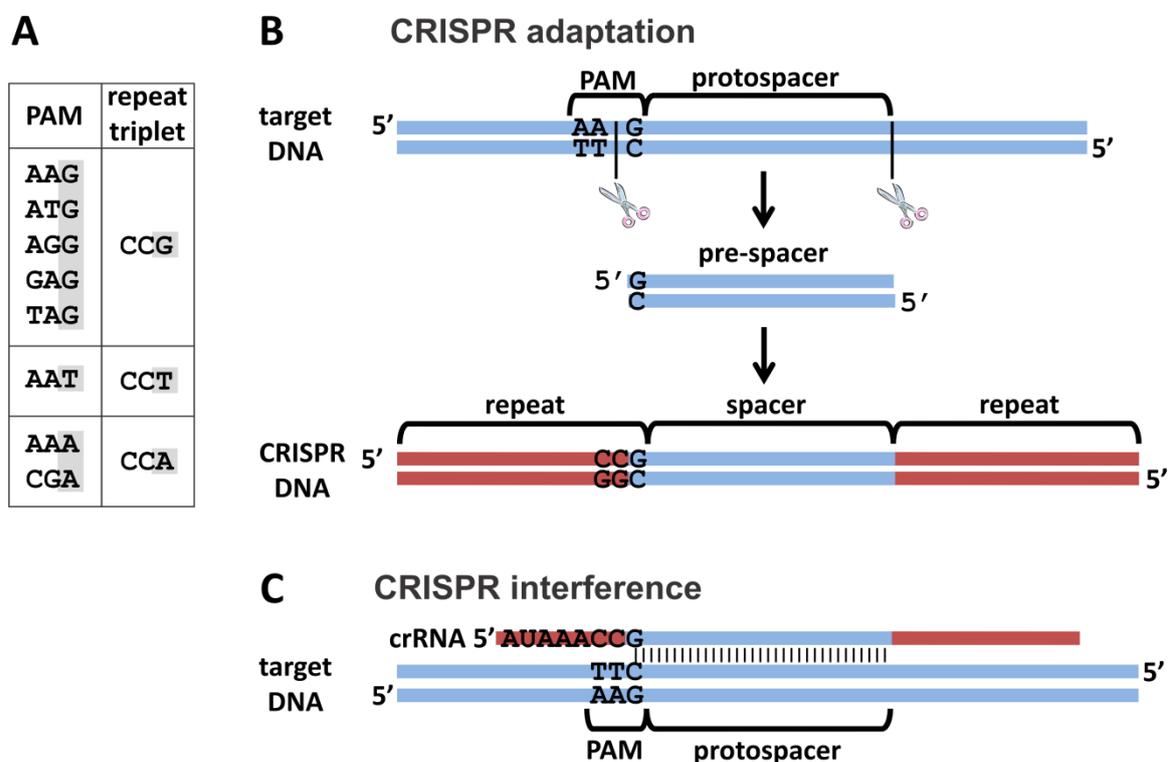


Figure 6 Pam sequences found in *E. coli* K12 and their prospective functions in adaptation and interference. (Taken from: Swarts *et al.*, 2012)

have been reported. They probably help the invading DNA to suppress the immune response (Doyle *et al.*, 2007). Expression of *cas* genes in *T. thermophilus* was reported to be regulated by the cyclic-AMP receptor protein (CRP) (Shinkai *et al.*, 2007).

Most studies have reported unidirectional transcription from the leader proximal side (Brouns *et al.*, 2008; Hale *et al.*, 2008; Marraffini and Sontheimer 2008; Semenova *et al.*, 2009). It is therefore anticipated that in most cases a single promoter at the leader side controls transcription of a CRISPR locus although in exceptional cases bidirectional transcription (two promoters) has been shown (Lillestol *et al.*, 2006; Lillestol *et al.*, 2009). The CRISPR cluster is transcribed as a long precursor RNA (pre-crRNA). Subsequently the precursor RNA is processed into small crRNA fragments by Cas proteins (Brouns *et al.*, 2008). The processing of the pre-crRNA into crRNA is commonly referred to as maturation. The matured crRNAs contain eight nucleotides of the repeat termed the 5' handle, the spacer, and a large part of the next repeat including a part of the stem-loop termed the 3' handle (Brouns *et al.*, 2008). It has been proposed that these handles, the conserved parts of the crRNAs, are recognized by subunits of Cas proteins that are involved in target DNA degradation (Brouns *et al.*, 2008). The mature crRNAs are usually associated with Cas proteins together they and form a ribonucleoprotein (crRNP) complex. The expression of *cas* genes and

formation of complexes from Cas proteins is continue during the interference stage, until the invading DNA is completely destroyed.

In the last phase of the process, the **interference** stage, the crRNAs, through their basepairing potential, allow the Cas proteins to find the invasive DNA (Figure 5). Initially it was thought that the CRISPR/Cas system mainly targets mRNAs, In a way similar to eukaryotic RNA silencing. However, there are strong indications that contradict this view. Spacer sequences correspond to viral DNA in the sense but also in antisense direction. If CRISPR/Cas would target RNA the crRNAs would have to be complementary to the mRNA of the virus. In the case of an antisense RNA mechanism only spacers from one strand should be incorporated.

One of the preconditions for all immune systems is to successfully discriminate between self and non self. For the CRISPR/Cas system this means to be able to differentiate between DNA sequences on the invader genome and the spacer sequences, which are derived from the invader but are located on the host genome. If discrimination between self and non-self DNA cannot be ensured the host would start cleaving its own genomic DNA. The exact mechanism of discriminating self from non-self DNA is dependent on the CRISPR/Cas type. In general discrimination is ensured by base-pairing potential. Each mature crRNA is composed of the spacer which perfectly matches the intruding DNA but the crRNA

Box 2 CRISPR/Cas in *E. coli*

Not surprisingly, the best understood CRISPR/Cas system is found in *E. coli* K12. *E. coli* K12 has a CRISPR/Cas system of the type IE, the same system can be found as well some actinobacteria, firmicutes, and methanogenic archaea as well as in many proteobacteria like *E. coli*. All three stages of CRISPR defense have been studied for the type I-E system. Although CRISPR adaptation is dormant in wild-type *E. coli* K12. The promoter that regulates the CRISPR array and most of the cas-genes in *E. coli* is tightly regulated so that usually there is no expression of the CRISPR array. This certainly is one of the reasons why CRISPR/Cas was not discovered earlier. Spacer acquisition could be observed in cells in which cas1 and cas2 expression levels were elevated.

also contains regions that do not match to the intruding DNA. These regions are important for self-recognition. In most CRISPR/Cas systems recognition of self-DNA is dependent on whether or not the PAM sequence forms base pairs with the DNA (Figure 5 C). The spacer region of a crRNA binds to DNA that is complementary to its own sequence. Invading DNA is identified by the fact that the PAM sequence on the crRNA does not form base pairs with the DNA entity to which it is bound. In case the PAM sequence forms base pairs with the DNA, the crRNA has bound to the host genome and degradation will be prevented. The molecular details of how base pairing at certain positions is sensed, and whether it

involves Cas proteins, is currently unknown. Interestingly, auto-immune reactions are not unknown for prokaryotes: approximately one in five CRISPR/Cas-containing organisms contains self-targeting spacers. The reasons and functions of this phenomenon are not clear but evidence is growing that this could be a form of gene regulation (Stern *et al.*, 2010).

Finally the invasive DNA is degraded by Cas proteins (Figure 5). The sets of Cas Cas9 proteins involved in DNA degradation varies greatly. The only Cas proteins conserved in almost all CRISPR/Cas systems are Cas1 and Cas2. The reason for the large variety of Cas proteins is probably the continuous co-evolution of viruses and the defense systems of their hosts, leading to a great variety of CRISPR/Cas systems (Bondy-Demomy *et al.*, 2013). The variety of CRISPR/Cas systems and their classification will be dealt with in the next chapter.

4. Classification of CRISPR/Cas systems

I-III

The first classification of CRISPR/Cas systems was proposed by Haft *et al.* in 2005. It was based on an analysis of 40 bacterial and archaeal genomes. The different CRISPR/Cas systems were named after a

representative organism, using a three letter code. For example, the CRISPR/Cas system in *E. coli* K12 was designated CSE (CRISPR system *E. coli*). The names of four core *cas* genes were adopted as originally proposed by Jansen *et al.* in 2002. Two other core genes, *cas5* and *cas6*, were

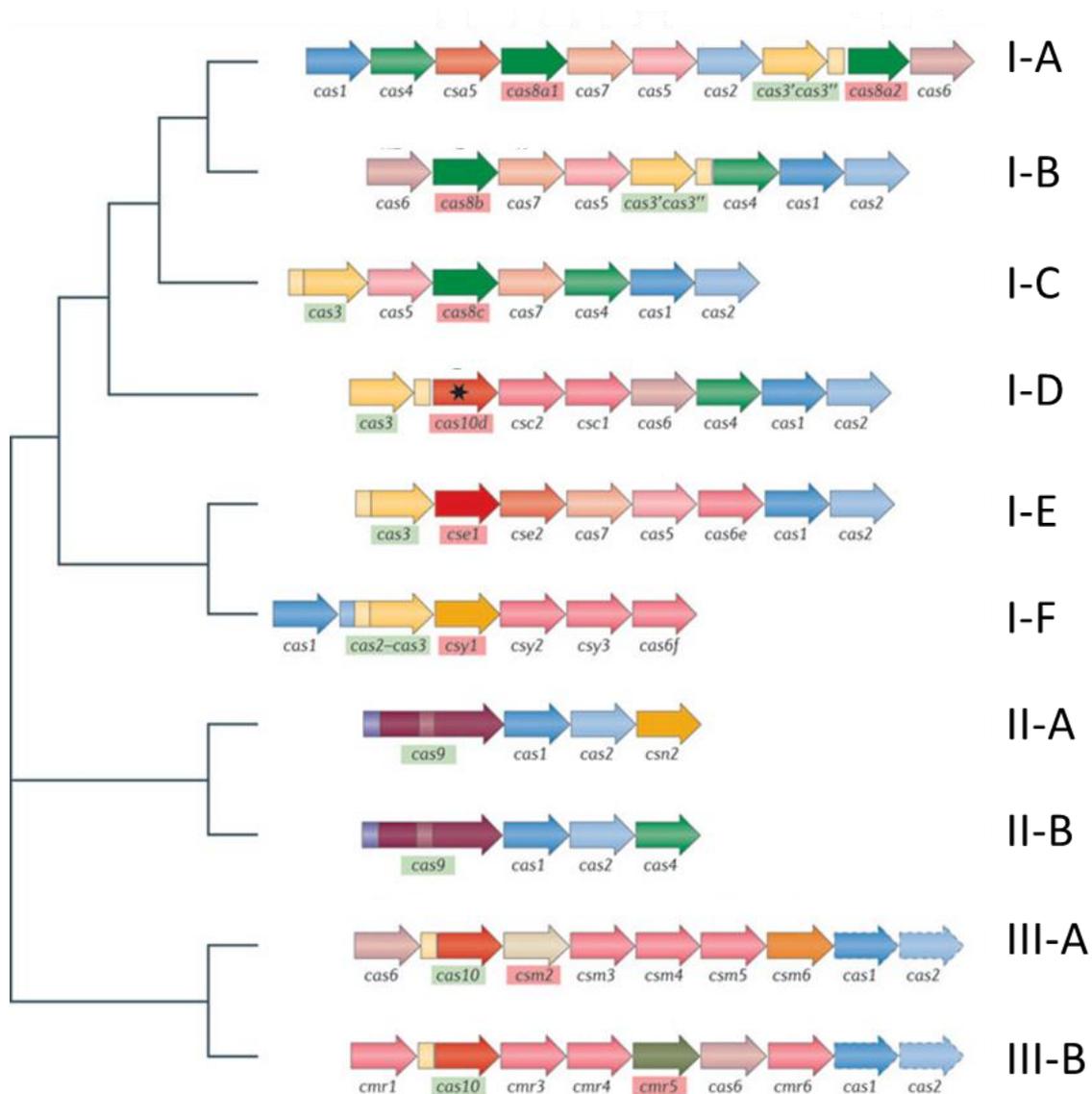


Figure 7 Classification and evolution of CRISPR/Cas systems according to the current classification system. Typical operon organisation of CRISPR/Cas subtype is shown. Orthologous genes are color coded. The signature genes of each CRISPR/Cas Type are highlighted in green and the signature proteins for subtypes are highlighted in red. The star indicates a predicted inactivated polymerase with an HD domain. The family tree on the left indicates how the different systems are related (Adapted from: Makarova *et al.* 2011b).

Table 1 Cas proteins, in which types of CRISPR/Cas systems they are found and their prospective functions

Cas protein	CRISPR/Cas Type	Function
Cas1	I-III	Integration of protospacer, DNA endonuclease (Wiedenheft <i>et al.</i> , 2009), RNA and DNA binding activity Hen <i>et al.</i> , 2008)
Cas2	I-III	Integration of protospacer, Ribonuclease activity (Beloglazova <i>et al.</i> , 2008)
Cas3	I	Helicase and nuclease, degradation of target DNA (Han and Krauss, 2009)
Cas4	I; II	RecB-like nuclease
Cas5	I	CASCADE core complex protein
Cas6	I; III	Endoribonuclease that cleaves the pre-crRNA transcript into crRNA units (Brouns <i>et al.</i> , 2008; Carte <i>et al.</i> , 2008)
Cas7	I	CASCADE core complex protein, involved in recognition of foreign nucleic acid and/or integration of novel spacer elements
Cas8	I	Part of CASCADE complex
Cas9	II; III	Targeting of invading DNA, cleavage of dsDNA maturation of crRNA and tracrRNA
Cas10	II; III	Likely to be involved in target interference, contains HD nuclease domain that is proposed to have similar function to Cas3

then added using the same principle, and names for genes encoding proteins specific to each of the eight CRISPR systems were proposed. Each gene was assigned a number according to its position in the *cas* gene cluster (e.g., in *E. coli*: *cse1*, *cse2* etc.). The *cas* genes in the other systems were named using a similar strategy resulting in 12 different CRISPR/Cas systems. Although the original approach seemed to be relatively simple, the weak spot of the classification was that it did not take into account the distant relationships that had later on been shown to exist between many Cas proteins. Thus some proteins present in the majority of CRISPR/Cas systems are clearly orthologous but were given several different names. Furthermore, the old classification did not take into account the complexity of the evolutionary relationships between the CRISPR–Cas systems in diverse bacteria

and archaea. Some prokaryotes were found to harbor several CRISPR/Cas systems in their genomes, which also caused trouble with the old classification system. To sum up, new results and the clear confusion between different names given to orthologous proteins required a new CRISPR/Cas classification system, which was introduced by Makarova *et al.* in 2011 (Makarova *et al.*, 2011b). Still, it is useful to be aware of the old classification systems since many papers still use the old names for Cas proteins.

Currently, CRISPR/Cas systems are classified into Type I, II and III, based on the phylogeny and presence of particular Cas proteins (Makarova *et al.*, 2011b) (Figure 7). In exceptional cases where classification of CRISPR/Cas systems into one of the types or subclasses is not possible, these systems are called Type U.

Proteins involved in the adaptation process are highly conserved while expression and interference systems vary greatly between different organisms. Unfortunately the knowledge about the adaptation process is only scarce. Since Cas1 and Cas2 are the most conserved proteins (Table 1) among all CRISPR/Cas systems it is likely that these proteins are involved in the adaptation process. Cas1 is a ds DNA endonuclease, involved in the maturation process of crRNAs while Cas2 is a sequence specific endoribonuclease, which cleaves uracil-rich single-stranded RNAs. However, it has been reported that some CRISPR/Cas subtypes lack any homologues of Cas1 and/or Cas2.

4.1. Common features of Type I CRISPR/Cas systems

Type I CRISPR/Cas systems are found in bacteria and archaea and are divided into six different subtypes (I-A to I-F) (Figure 7). The essential and most conserved protein of Type I systems is **Cas3**. Cas3 has two functional domains, a HD phosphohydrolase domain and a DExH-like helicase domain (Makarova *et al.*, 2011a). The helicase domain has been shown to unwind dsDNA while the phosphohydrolase domain cleaves single strand DNA (dsDNA) (Sinkunas *et al.*, 2011). In some Type I systems (subtypes A, B, and D), separate genes encode the nuclease and helicase domains, but in all of these systems these two domains are anticipated to work together by cleaving

(HD domain) and unwinding (helicase domain) dsDNA targets for processive degradation. However Cas3 lacks RNA binding activity and is therefore not able to interact with crRNA which is needed to find and bind invasive DNA.

To identify and bind a foreign DNA Cas3 interacts with a complex of other Cas proteins called **Cascade** (CRISPR-associated complex for antiviral defense) (Figure 8).

Cascade binds to the crRNA protects the crRNA from degradation, and recruits Cas3 for degradation of the invading viral DNA. The best studied Type I system is the CRISPR/Cas system I-E of *E. coli* (Brouns *et al.*, 2008; Jore *et al.*, 2011). The I-E Cascade complex has an atomic mass of 405 kDa and is composed of the five subunits,

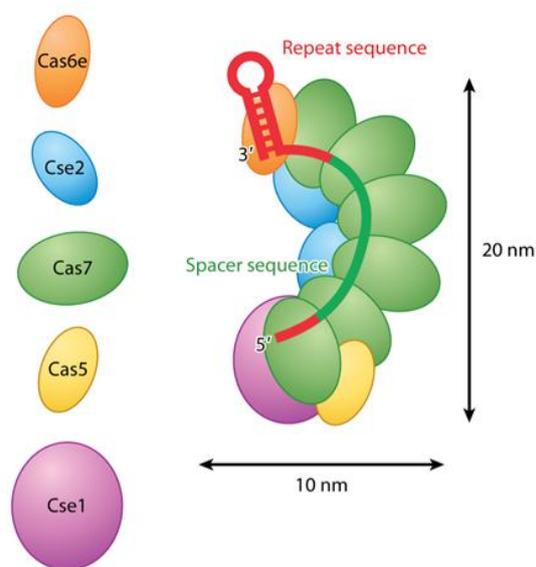


Figure 8 Schematic representation of CASCADE. Composition and structure of CASCADE and its bound crRNA (Taken from: Westra *et al.*, 2012)

Cas6e, Cse1, Cse2, Cas7 and Cas5 (Bourns *et al.*, 2008). **Cas6e** is a CRISPR-specific endoribonuclease that cleaves long CRISPR RNA into mature 61-nt crRNAs. Cas6e and the crRNAs are required for stable assembly of the other Cas proteins. A similar complex exists in *Sulfolobus solfataricus* (Type I-A). Experiments performed on **Cas7** in the Type I-A system suggest that it pre-positions crRNA in an unwound and stretched conformation which is optimal for strand invasion and exchange (transition state stabilization), similar to that described for RecA (Box3) (Lintner *et al.*, 2011, Chen *et al.*, 2008). However, unlike RecA, which catalyzes DNA repair, target recognition by Cascade may induce a conformational change that recruits Cas3 for destruction of the invading DNA (Wiedenheft *et al.*, 2011). Another important feature of the Type I

CRISPR/Cas system is the PAM sequence. It is 2-3 nt long and located upstream of the spacer (Figure 9).

Box3 Function of RecA

RecA is a protein which is essential for the repair and maintenance of DNA.

RecA binds together with ATP to single-stranded DNA (ssDNA), several RecA proteins form a helical filament. The filament is able to bind to double-stranded DNA (dsDNA) and search for homologous regions. Once a complementary region has been identified RecA catalyzes the DNA synapsis reaction between the dsDNA and the (damaged) ssDNA. (Chen *et al.*, 2008)

4.2. Common features of Type II CRISPR/Cas systems

Type II CRISPR/Cas systems are

Type	Subtype					
	A	B	C	D	E	F
I	CC ₂ -Proto	CC ₃ -Proto	IIIC-Proto	Not determined	AAG ₃ -Proto	CC ₃ -Proto
II	Proto-GGG ₃ Proto-AGAA ₃	Proto-GG ₂				
III	No PAM	No PAM				

Figure 9 Conserved elements of PAM sequences in CRISPR/Cas types and subtypes including the position of the protospacer relative to the PAM sequence (Adapted from: Sorek *et al.*, 2013)

exclusively found in bacteria and are divided into two subtypes (II-A and II-B). Characteristic for Type II systems are a minimal set of Cas proteins (Makarova *et al.*, 2011b) and the multifunctional protein **Cas9**. Type II CRISPR/Cas systems consist of only four Cas proteins: Cas9, Cas1, Cas2, and either Csn2 (Type II-A) or Cas4 (Type II-B) (Figure 7).

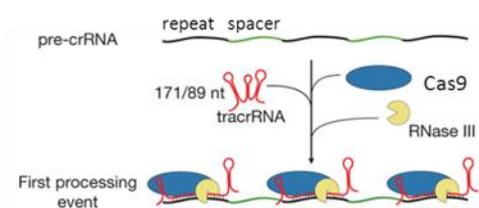


Figure 10 Model for tracrRNA-mediated crRNA maturation involving RNase III and Cas9. The tracrRNA binds to the repeat region in the pre-crRNA transcript which promotes processing by RNaseIII. The first processing event is followed by a second processing event (Adapted from: Deltcheva *et al.*, 2011).

Cas9 is involved in maturation of crRNA and destruction of invading DNA. Maturation of crRNA in Type II systems differs from that of other systems because it requires a trans-activating crRNA (tracrRNA). TracrRNA is expressed in two isoforms that are 89 and 171 nt in length (Figure 10). Both isoforms contain a 25 nt sequence that is, apart from one mismatch, perfectly complementary to the repeat sequence in the pre-crRNA (Figure 10). The tracrRNA is encoded on the host genome in close proximity to the crRNA. In *Streptococcus pyogenes*, for example, the tracrRNA is encoded upstream on the opposite strand of the CRISPR/Cas locus. Cas9 mediates the formation of a duplex

Box4 Ruler-based cleavage of RNA

Ruler-based cleavage is a processing mechanism, which is independent of the sequence, or secondary structure of the processed region. Instead the RNA is processed to a final length of the resulting RNA product. This is measured in basepairs. Ruler based cleavage is found in maturation of crRNAs and in the maturation of many ncRNAs (Hatoum-Aslan *et al.*, 2011).

RNA from tracrRNA and pre-crRNA (Deltcheva *et al.*, 2011). Subsequently the pre-crRNA is processed into mature crRNA by the cellular (non-Cas) RNaseIII. The Cas9:RNA complex is a sequence-specific endonuclease. It generates a blunt-ended double-stranded break 3 base pairs upstream of the 3' end of the protospacer. This process is mediated by two catalytic domains: an HNH nuclease domain that cleaves the strand complementary to the crRNA and a RuvC-like nuclease domain that cleaves the non-complementary strand. Recent results show that tracrRNA is not only required required for Cas9-mediated cleavage of target DNA, although the underlying details are still unclear (Jinek *et al.*, 2012). The PAM sequence of Type II systems is a 3 or 5 nt-long conserved sequence (5 nt: NGGNG; 3 nt: NGG). In contrast to Type I systems the PAM sequence in Type II systems is not located upstream but downstream of the spacer (Figure 9).

4.3. Common features of Type III CRISPR/Cas systems

Type III CRISPR/Cas systems are predominantly found in archea. They are

divided into the two subtypes III-A and III-B. Interestingly Type III-B systems could so far only be identified in combination with other CRISPR/Cas subtypes. Characteristic for Type III systems is the presence of Cas6 in combination with Cas10. Type III systems do not require a PAM sequence (Figure 9). Type III-A CRISPR-Cas systems consist of nine cas genes (*cas1*, *cas2*, *cas10*, *csm2*, *csm3*, *csm4*, *csm5*, *csm6*, *cas6*). For the Type III-A system of *Staphylococcus epidermidis* autoimmunity is prevented through a mechanism that relies on sensing base pairing between the 5'-handle (the repeat-derived sequence at the 5'-end of the crRNA, see Chapter 3.1) and the corresponding portion of the CRISPR repeat. After a primary processing step of the pre-crRNA, the resulting crRNAs are further matured through ruler-based cleavage (Box4) from the 3'-end, yielding crRNA species of 43 and 37 nt (Hatoum-Aslan *et al.*, 2011). These mature crRNAs guide one or more Cas proteins to target DNA, by base pairing between the crRNA spacer sequence and the complementary protospacer sequence. However, CRISPR-interference is inhibited when, in addition to base pairing over the spacer sequence, the 5'-handle also forms base pairs with the protospacer-flanking sequence of the target DNA. In this manner, self-targeting of the CRISPR locus is avoided by default, since self-targeting

inevitably leads to full base pairing of the 5'-handle of the crRNA with the CRISPR repeat sequence from which it is transcribed. Type III-B systems are the only systems that can target both DNA and RNA although the biological relevance of this “double-edged sword” is not clear yet. Similar evidence for PAM-independent discrimination of self and non-self was obtained for Type III-B systems.

5. Eukaryotic RNA silencing: a concept analogous to CRISPR/Cas

Eukaryotes have developed different methods to fight viral infections. Mammals combat viruses by apoptosis of infected cells, clearance of infected cells by natural killer cells and antibody production by the cell-based adaptive immune system. Plants also employ programmed cell death through the hypersensitive response, whereas insects use intracellular protein-mediated antiviral defense. Just like in bacteria, RNA also plays an important role in fighting viral infections in plants, insects and possibly in mammals. Eukaryotic RNA-based viral defense is based on RNA interference (RNAi), which does not target intruding DNA but effects the degradation of homologous mRNAs. RNAi is not only a mechanism to fight viruses but also plays a key role in eukaryotic gene regulation.

In 1990, Richard Jorgensen was the first one to observe the effects of RNAi (Napoli *et al.*, 1990). During the engineering of transgenic petunias for altered pigmentation to intensify the flower colour. The introduction of

transgenes unexpectedly resulted in variegated pigmentation, with some plants lacking pigment altogether (Napoli *et al.*, 1990). The basic RNAi mechanism was solved in 1998 (Fire *et al.*, 1998). In *C.*

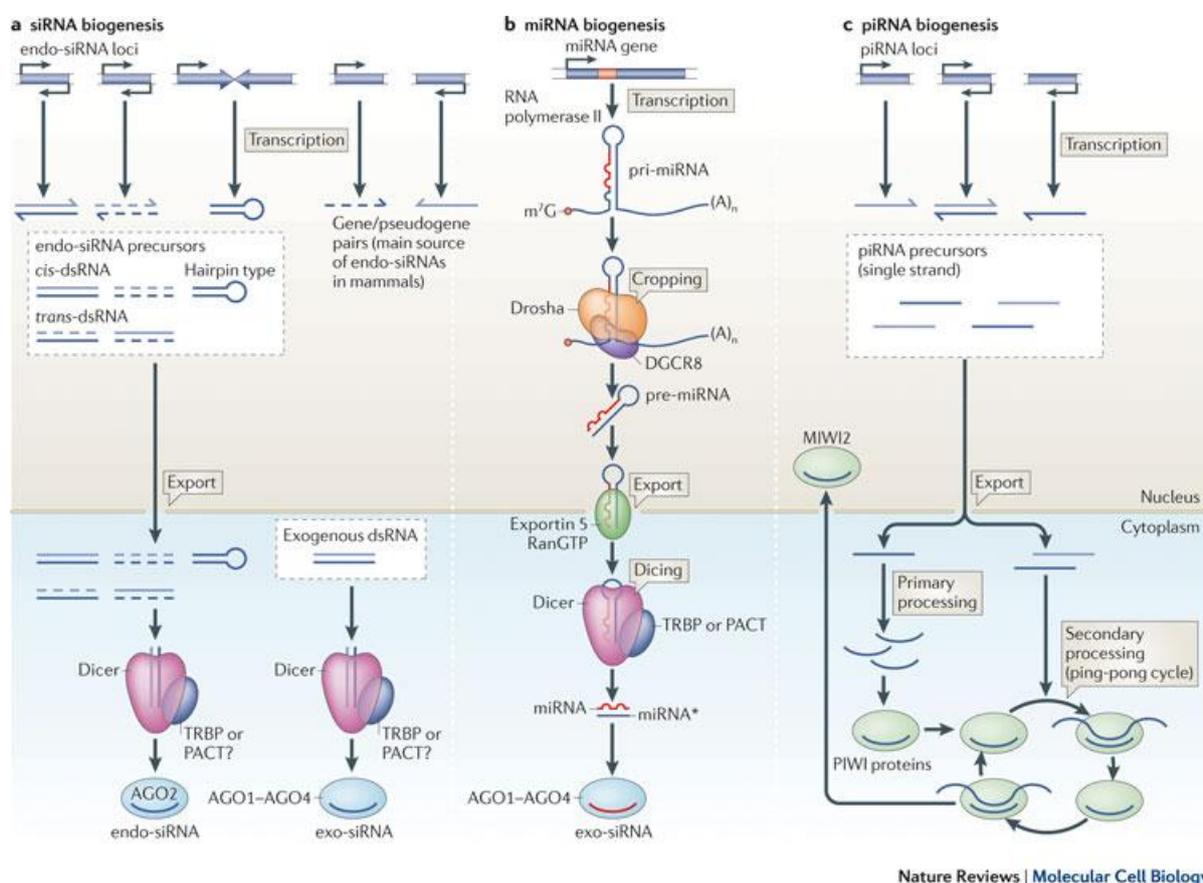


Figure 10 Overview of the most important eukaryotic small noncoding RNAs biogenesis and function. a | Biogenesis of small interfering RNA (siRNA). Transcripts that are able to form double-stranded RNA (dsRNA) or long stem-loop structures serve as endogenous siRNA (endo-siRNA) precursors. Pseudogenes transcribed in an antisense orientation produce RNA that pairs with cognate gene mRNAs as well as with transcripts that are derived from intergenic repetitive sequences on the genome, including transposons. endo-siRNA is processed by Dicer, whereas the role of Dicer-binding proteins — such as TAR RNA-binding protein (TRBP; also known as TARBP2) and PACT (also known as PRKRA) — remains undetermined. After maturation, endo-siRNAs are loaded onto Argonaute 2 (AGO2). Exogenous siRNAs (exo-siRNAs) are derived from exogenous dsRNAs by Dicer–TRBP (or Dicer–PACT). exo-siRNAs are loaded onto AGO1, AGO2, AGO3 and AGO4; however, only the AGO2–siRNA complex functions in RNA interference. **b | Biogenesis of MicroRNA (miRNA).** The primary transcripts of miRNAs (pri-miRNAs) are transcribed by RNA polymerase II from miRNA genes on the genome. pri-miRNAs form hairpin structures and are processed to ~60–70 nt miRNA precursors (pre-miRNAs) by Drosha–DGCR8 (DiGeorge syndrome critical region 8) complex in the nucleus. After being exported by exportin 5 and RanGTP, pre-miRNAs are further processed to ~22 nt miRNA–miRNA* duplexes (in which miRNA* is the passenger strand that is degraded) by the Dicer–TRBP (or Dicer–PACT) complex. Mature miRNAs are then loaded onto AGO1, AGO2, AGO3 and AGO4. **c | Biogenesis of PIWI-interacting RNA (piRNA).** piRNAs are processed from single-stranded RNA precursors that are transcribed largely from particular intergenic repetitive elements known as piRNA clusters. First, primary piRNAs are produced through the primary processing pathway and are amplified through the ping-pong pathway, which requires Slicer activity of PIWI proteins. Primary piRNA processing and loading onto mouse PIWI proteins might occur in the cytoplasm. MIWI2 (also known as PIWI-like protein 4) specifically associates with secondary piRNAs that are processed through the amplification loop, and is localized in the nucleus to exert its silencing function. (A)_n indicates the poly(A) tail formed during translation, and m7G indicates the 5'-terminal cap of the mRNA. (Taken from: Siomi *et al.*, 2011)

C. elegans silencing could be initiated by exposing cells to double-stranded RNA (dsRNA), which directed the destruction of mRNAs containing similar sequences. Although this will not be discussed here in detail, it should be noted that small noncoding RNAs (ncRNAs) in eukaryotes are not only involved in transcriptional gene regulation but also play important roles in epigenetics (Marchese and Huarte, 2013). RNA-based gene regulation in eukaryotes has been studied intensively. Different kinds of small ncRNAs were identified, especially by next-generation sequencing. Tools to study gene function have been developed by employing RNAi (see Chapter 5.1). Much of the early discoveries has become textbook knowledge in the meantime. Eukaryotes have developed different methods for RNA based gene regulation (Box 5). There are three major kinds of small regulatory RNAs in many eukaryotic cells: small interfering RNAs (siRNAs), micro RNAs (miRNAs) and piwi-interacting RNAs (piRNAs) (Figure 11).

siRNA can be derived from exogenous or endogenous origins. **Exogenous siRNA** enters the cell as long dsRNA. Exogenous siRNA can be derived from viral RNA. The dsRNA is cleaved into ds siRNAs by Dicer (Bernstein *et al.*, 2001) (Figure 11), a double-stranded RNA-specific ribonuclease from the RNase III protein family. In most species, cleavage of long ds RNAs by Dicer produces ds siRNAs of approximately 21-25 nt (Zamore *et al.*, 2000). These have a two-

nt overhang at their 3'-end, as well as a 5'-phosphate and a 3'-hydroxyl group (Elbashir *et al.*, 2001b). The strand that directs silencing is called the guide strand, whereas the other strand is the passenger strand (Schwarz *et al.*, 2003). Target regulation by siRNAs is mediated by the RNA-induced silencing complex (RISC) (Hammond *et al.*, 2000). The relative thermodynamic stabilities of the 5'-ends of the two siRNA strands in the duplex determines the identity of the guide and passenger strands (Schwarz *et al.*, 2003). Release of the passenger strand, which is ultimately destroyed, converts pre-RISC to mature RISC, which contains only single-stranded guide RNA. RISC is a complex composed of Argonaute proteins and RNA. The number of different Argonaute proteins involved in this complex varies among species. For example, there are more than 25 Argonautes in *C. elegans* but only five in *D. melanogaster*. Argonaute catalyses cleavage near the centre of the region of the mRNA that is bound by the siRNA (Ghildiyal and Zamore, 2009).

miRNAs are the best-studied kind of eukaryotic ncRNA. The miRNAs are involved in posttranscriptional silencing of host gene expression by interaction with mRNA. miRNAs derive from precursor transcripts called primary miRNAs (pri-miRNAs), which are typically formed by RNA polymerase II (RNA Pol II) (Lee *et al.*, 2004). The pri-miRNA is processed in the nucleus into a 60 to 70 nt pre-miRNA by Drosha, an enzyme which belongs to the

RNaseIII family of RNases (Lee *et al.*, 2004) (Figure 11). So, similar to the Type II CRISPR/Cas system, the precursors of siRNA and miRNA are cleaved by an RNaseIII type protein into their mature products. The resulting pre-miRNA has a hairpin structure. Pre-miRNAs have a two-nt overhang at their 3'-ends and a 5'-phosphate group, which are indicative of their production by an RNase III enzyme. The nuclear export protein Exportin 5 carries the pre-miRNA to the cytoplasm, while the latter is bound to Ran, a GTPase that moves RNA and proteins through the nuclear pore (Lund *et al.*, 2004). In the cytoplasm, Dicer cleaves the pre-miRNA. Dicer cleavage generates a duplex containing two strands, termed miRNA and miRNA* (Ghildiyal and Zamore, 2009).

Especially striking are the similarities between the prokaryotic crRNAs and the eukaryotic **piRNAs**. piRNAs were only recently discovered and many details about biogenesis and functions of piRNAs are still elusive. Mature piRNAs are 21-30 nt long and are associated with a multiprotein complex called piwi, in animal cells. The piwi complex represses “non-self” DNA sequences such as transposable elements to prevent these from moving or multiplying into new positions into the genome (Figure 11). Transposable elements are present in the genomes of all organisms and can constitute a large fraction of the genomic DNA (e.g. 45% of the human genome) so, not surprisingly, transposable elements often

Box 5 For more information about eukaryotic noncoding RNA read: Röther and Meister (2011) for a comprehensive review.

carry vital information. On the other hand, they are also a common cause of mutations and genome rearrangements and can thus be involved in e.g. cancer. Therefore, transposable elements need to be tightly controlled by the cell.

piRNA is encoded in the genome on piRNA loci or clusters. Biogenesis of piRNA differs from that of other eukaryotic RNAs. There are two different pathways for precursor transcription and biogenesis of piRNAs, one from *C. elegans* and one from *D. melanogaster*. As is the case in CRISPR/Cas Type I and Type III systems the single stranded RNA precursor is processed into mature piRNA that guides a protein complex (piwi). In *Drosophila* the piwi protein family consists of three proteins: Piwi, Aubergine and Ago3. Piwi and Aubergine form a complex that is associated to RNAs that are antisense to transposon RNA, while Ago3 is associated to piRNAs. Piwi cleaves the transposon derived targetRNA 10 nt downstream from the 5'-end of the guiding piRNA. The CRISPR/Cas system preferentially targets DNA while eukaryotic RNAi, including piRNA targets RNA only (Siomi *et al.*, 2011).

5.1. Employing eukaryotic RNA silencing in research

The conserved mechanism employed by eukaryotes to degrade dsRNAs such as miRNA or siRNA has been exploited by researchers to control gene expression in various organisms. The introduction of dsRNA into an organism results in sequence-specific gene silencing (Fire *et al.*, 1991). This methodology is generally known as RNA interference (RNAi) (Fire *et al.*, 1998). In analogy to the term knock-out for the physical deletion of genes from the genome, RNAi-mediated silencing of genes is often referred to as 'knock-down'. RNAi is widely used to study gene function in eukaryotes for several reasons. First of all, limitations of normal knock-out strains can be overcome by this method. For instance, it enables researchers to study essential genes that cannot be knocked out. It also reduces the relatively long time it takes to make a gene knockout, especially in higher organisms like mice. It also circumvents the restrictions associated with embryonic lethality. Furthermore, several genes can be silenced at the same time which saves time when compared to knocking out multiple genes in several transfection steps. Thus, using RNAi can be used for high throughput screening methodologies.

RNAi is also a tool in the study of the functions of non-coding RNA. For that dsRNA with a sequence complementary to the gene to be knocked down is chemically

synthesized. Different ways can be used to introduce the dsRNA into the organism. In *C. elegans* for example the dsRNA can be injected into the gonads, the dsRNA can be administered directly or bacteria engineered to express the dsRNA can be fed (Timmons and Fire, 1998).

The dsRNA molecule mimics Dicer substrates. Upon its introduction into the model organism Dicer cleaves the synthetic constructs into functional siRNA (Bernstein *et al.*, 2001; Elbashir *et al.*, 2001a). The siRNAs are then incorporated into RISC, which facilitates gene silencing in a sequence specific manner (Hammond *et al.*, 2000).

RNA induced silencing spreads systemic from cell to cell resulting in silenced gene expression in the whole organism (Fire *et al.*, 1998). The effects of RNAi in mammals are only temporary: they depend on the stability of the siRNA structure. The transient nature of externally introduced siRNA in mammals can be overcome by using another system. Mammalian cells can be transfected with a vector containing information for a so-called small-hairpin RNA. A promoter driving expression of this RNA molecule can be used that is tunable so that the effect of the silenced gene is only observed when needed. However, construction and transfection of such a vector is time consuming and therefore this technique is only used when necessary. Worms and plants on the other hand have

mechanisms with which they can amplify the siRNA and therefore gene silencing in these organisms is permanent. In several organisms, including plants, *Drosophila*, *C. elegans* and trypanosomes, RNAi has been made stable and heritable by the forced expression of the silencing trigger, usually an inverted repeat sequence forming a hairpin structure *in vivo*.

Uses other than functional genomics have also been suggested for RNAi. One of these ideas is to switch off gene expression of “trouble-making” genes, e.g. genes linked to certain diseases. However the efficient delivery of small RNAs and the targeting to

specific tissues *in vivo* has proven to be rather difficult so far (Ghildiyal and Zamore, 2009). Alternatives to the classical RNAi mechanisms have also been developed which may replace RNAi in many experimental setups. Interestingly, some of the alternatives are derived from the CRISPR/Cas system and will be discussed below.

6. CRISPR/Cas applications in research

Initially, the CRISPR/Cas system was used for easy identification of closely related bacterial species/strains. This method is known as spoligotyping (spacer

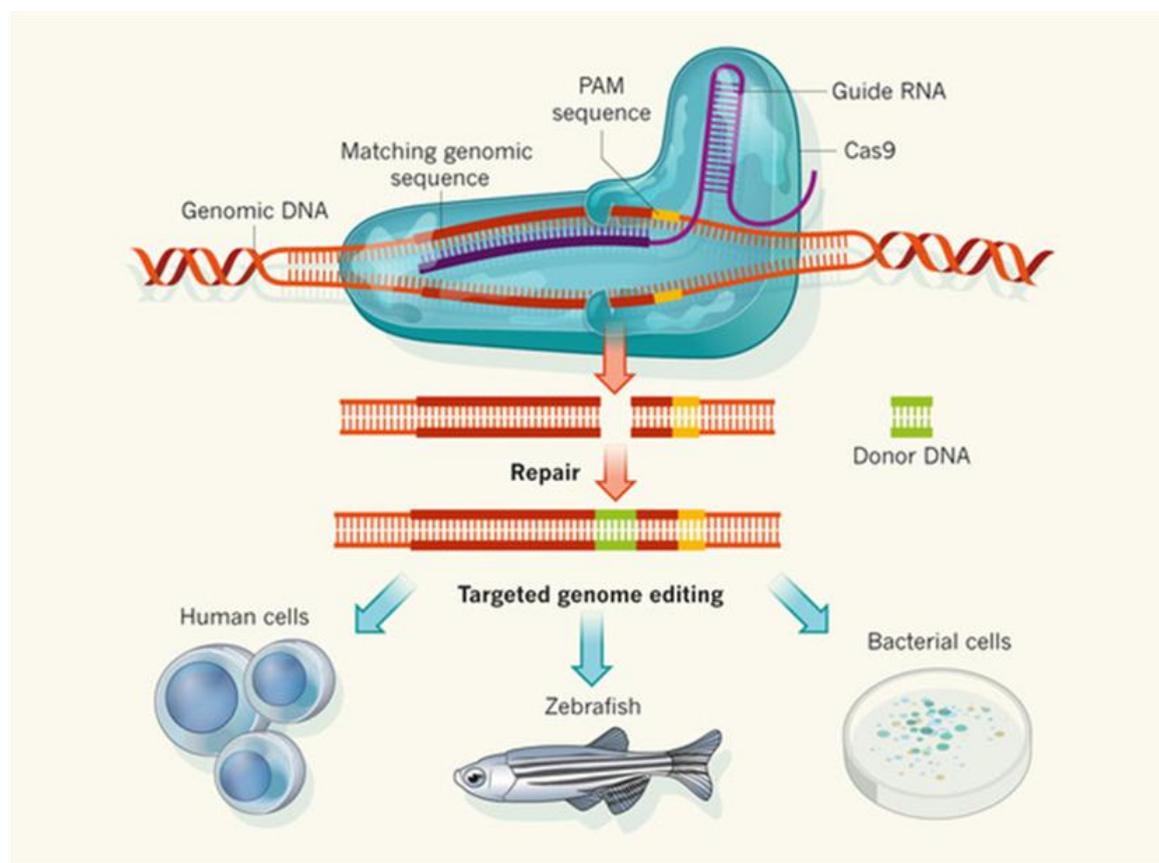


Figure 11 The principle of genome editing by RNA-guided nucleases. Cas9 which is bound to a guide RNA can introduce sequence specific double strand breaks, which can be repaired by homologous recombination if a donor DNA is provided. This technique can be applied in different organisms. Cas9 was used for targeted genome editing on important models such as human cells, zebrafish and bacteria. (Taken from: Charpentier and Doudna, 2013)

oligonucleotide typing) (Dale *et al.*, 2001). Since the discovery of the large and diverse functions of the Cas proteins more and more powerful functions of CRISPR/Cas systems as genetic tools have been developed. The Type II protein Cas9 takes a key role in these newly developed applications. This is because, in contrast to Type I and Type III Cas proteins that mostly rely on multi-protein complexes, the Type II Cas9 is only a single protein. Therefore it is easy to express in other organisms and easier to engineer and optimize. Usually optimization of Cas9 is not necessary in bacteria. A single-chain chimeric RNA produced by fusing crRNA and tracrRNA sequences can replace the two RNAs in the Cas9-RNA complex to form a single-guide-RNA:Cas9

endonuclease (sgRNA:Cas9) (Figure 12). In eukaryotes, however, codon optimization and addition of a nuclear localization signal (NLS) is required for successful use of Cas9. The best way to deliver the CRISPR/Cas components depends on the organism it is applied in. In mammalian cells, researchers have used, electroporation, nucleofection, Lipofectamine-mediated transfection of nonreplicating plasmid DNA to transiently express Cas9 and gRNAs. Another option in cultured mammalian cells is transfection with lentiviral vectors. The CRISPR/Cas components can be injected into an embryo or into the gonads of adult animals. In plants standard delivery methods including PEG-mediated transformation of protoplasts, Agrobacterium-mediated transfer in embryos

Box 6 Genome editing in a nutshell

Genome-editing or genome engineering is defined as the *in vivo* modification of genetic information. This includes insertion new genetic information, deletion of genetic information and modification of genetic information by replacing single basepairs. An important step for genome engineering is the introduction of sequence specific double strand breaks, which can be accomplished with several different techniques (Figure 13).

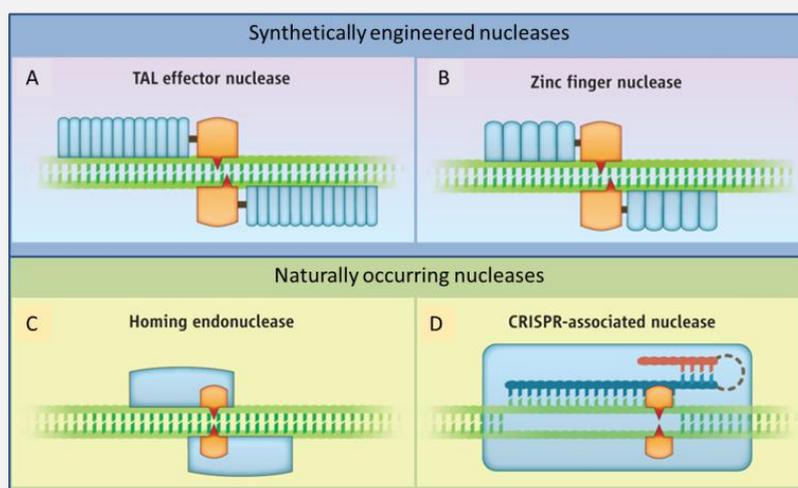


Figure 13 Schematic representation of nucleases, which are used for genome engineering. (Adapted from van den Oost, 2013)

Box 6 Genome editing in a nutshell

A Transcription activator–like effector nucleases (TALENs) were originally discovered by studying secreted proteins of the bacterial plant pathogen *Xanthomonas* and named for their innate ability to activate transcription of endogenous plant genes essential for pathogenicity. The DNA-binding domain of TAL proteins was found to comprise several 33 to 35 amino acid repeats with differences at residues 12 and 13, a position that was named the repeat variable di-residue (RVD). This TAL-effector domain can be designed to recognize a special DNA sequence of 24-59 bp. Since 2010 TALENs are used for genome editing (Christialn *et al.*, 2010). A fusion to the endonuclease domain FokI ensures cleavage of the site of interest, so a pair of TALENs has to be engineered. The targeting efficiency varies for different TALEN pairs and in different loci. The efficiency of different TALENs is therefore hard to predict (Wei *et al.*, 2013).

B Designer zinc finger nucleases (ZFNs) are engineered restriction enzymes based on the feature that different zinc fingers recognize different sets of nucleotide triplets. The recognition size of a zinc finger motif can be engineered to recognize a DNA sequence of 18-36 bp. Cleavage of the binding site is ensured by a fusion of the zinc finger motif to the endonuclease domain FokI. FokI only cleaves DNA upon dimerization, so a pair of different zinc finger motifs has to be engineered. Important limitations of ZFNs are that not all nucleotide triplets have got their corresponding zinc fingers discovered and that the production of ZFNs with high selectivity is costly, laborious, and time consuming (Urnov *et al.*, 2010).

C. Homing meganucleases are also referred to as rare cutters or homing endonucleases. They are restriction enzymes that have none or just a few natural recognition sites in the genomes of many eukaryotes. These nucleases are highly specific and show precise cleavage of their target. They are characterized by a nucleotide sequence recognition site (20-30 bp) which is larger than that of normal restriction enzymes (4-8 bp). Homing meganucleases occur naturally in various organisms, where they are often encoded in introns or inteins. Engineering homing meganucleases makes it possible to change the recognition site of the enzyme but has proven to be difficult in the past because of off-target activity problems. Screening tools to identify novel naturally homing meganucleases also exist. (Barzel *et al.*, 2011)

D RNA guided nucleases (RGENs) are currently based on CRISPR/Cas type II protein Cas9. The restriction site depends on the RNA that it is bound to Cas9. Typical cleavage recognition sizes are 20bp.

For up to date and comprehensive reviews about genome editing tools read Gasiunas and Siksnys, (2013) or van der Oost, (2013).

and leaf tissue, and/or bombardment of callus cells with plasmid DNA can be used to deliver the genetic information needed to express Cas9/RNA.

Probably the most exciting development based on Cas9 are **RNA-guided nucleases** (Figure 12). RNA-guided nucleases make use of the ability of Cas9 to introduce sequence specific double strand breaks

(Figure 14). The DNA target site can be customized by replacing a short synthetic RNA molecule without changing the protein component. It is also possible to introduce sequence specific nicks, if one of the nuclease domains of Cas9 is destroyed (mutated) (Figure 15). Furthermore, multiple guide sequences can be encoded into a single CRISPR array to enable simultaneous change of several sites within a genome.

This process is called multiplexing and is, so far exclusively possible with the CRISPR/Cas9 system. It has been shown that sgRNA:Cas9 can efficiently induce site-specific genome modifications (Cho *et al.*, 2013). These nucleases can be used for genome editing (Box 6). Applications developed on the CRISPR/Cas system are not only useful for prokaryotes but can be used in any eukaryotic species as well and they are generally thought to be affordable, and easy to engineer. In eukaryotes, successful genome editing has been performed in zebrafish, yeast, fruit fly, frog, mouse, rice, tobacco, silk worms and human cells, to name just a few examples (Sander and Joung, 2014). For most RGEN applications, transient expression of gRNAs and Cas9 is typically sufficient to induce

efficient genome editing. Although constitutive expression of RGEN components might potentially lead to higher on-target editing efficiencies, extended persistence of these components in the cell might also lead to increased frequencies of off-target mutations, a phenomenon that has been previously reported with ZFNs (Gaj *et al.*, 2012). In zebrafish it has been reported that the Cas9/gRNA system is an easier and more economic for the construction of genome editing tools, than with ZFN or TALENs, and it is comparable with respect to generating site-specific indels (insertions or deletions) (Figure 13). Nevertheless, the off-target effects, especially unspecific cleavage, of the Cas9/gRNA system remains to be addressed prior to its large-scale implementation in genome editing. The Cas9

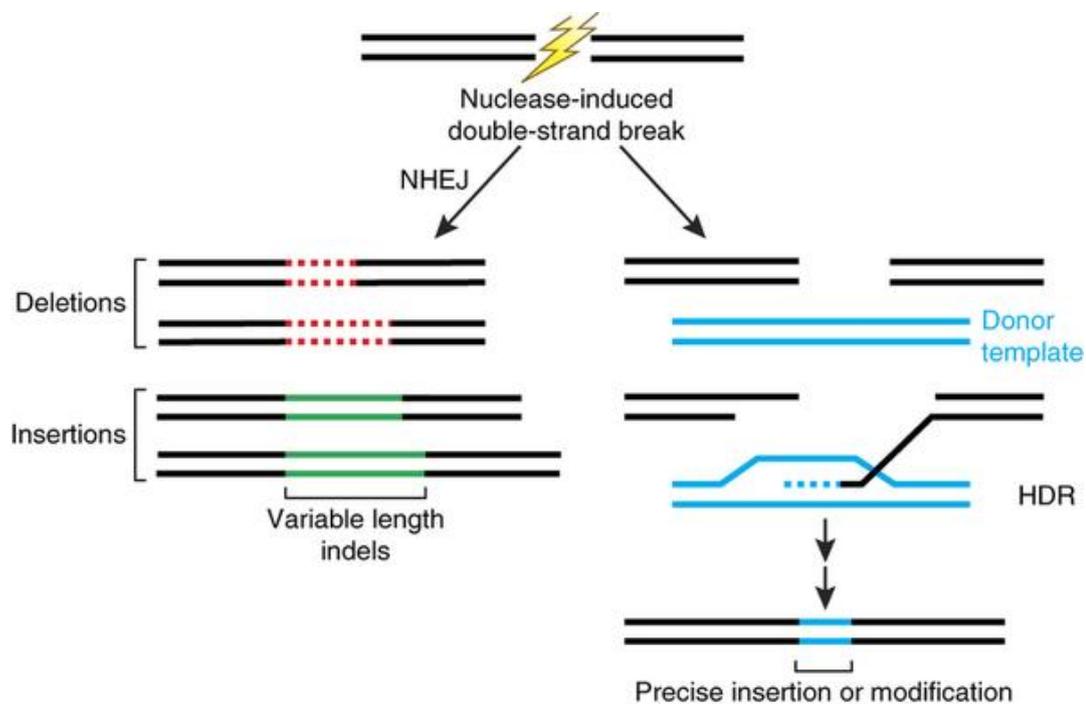


Figure 12 Repair strategies for nuclease-induced double-strand breaks (DSBs). **Left:** nonhomologous end joining (NHEJ). NHEJ-mediated repair is inexact and can produce insertion and/or deletion mutations of variable length at the site of the DSB. **Right:** homology-directed repair (HDR) pathways. HDR-mediated repair can introduce precise point mutations or insertions from a single-stranded or double-stranded DNA donor template. (Taken from: Sander and Joung, 2014)

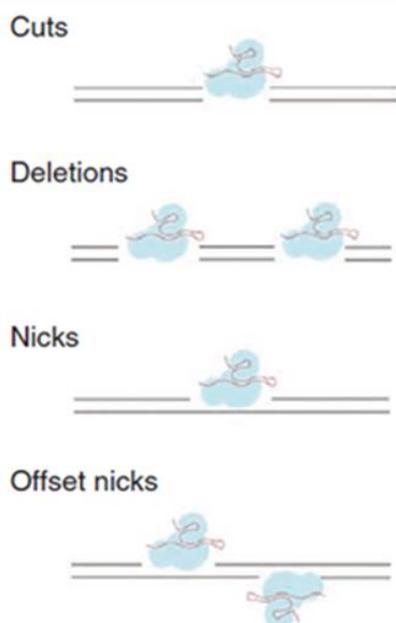
system is a promising tool for the generation of conditional knock-outs, which were not successful so far in zebrafish (Chang *et al.*, 2013).

In fact, the RNA-guided nucleases are so successful that the first parts have already been commercialized. Companies start to offer expression plasmids (Sigma Aldrich, System Biosciences) or even a full gene editing service to produce a genetically modified cell using any mammalian cell line and targeting any gene (Genscript, Toolgen). Custom made gRNA in combination with purified Cas9 as a “ready to use” for injection or electroporation kit is also

offered and can be delivered to the laboratory within two weeks according to the supplier (Toolgen) (status: March 2014).

Similarly to gene silencing methods derived from eukaryotic systems described in Chapter 5.2., gene silencing can also be achieved through **CRISPR interference (CRISPRi)**. The advantages of this system are its reversibility, the fact that it does not show off-target effects and its genome wide scale applicability. CRISPRi uses catalytically dead Cas9 (dCas9). This mutant protein lacks endonuclease activity and is co-expressed with a short guide RNA (sgRNA) (Figure 15). This generates a DNA

Cas9 nuclease



Cas9 without nuclease activity

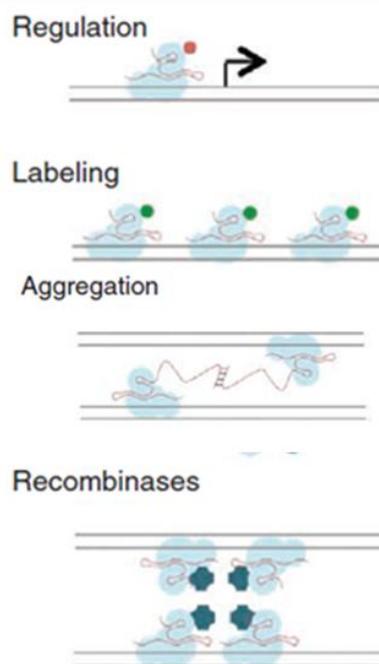


Figure 13 Possible applications based on different engineered Cas9 proteins. Cas9 proteins with nuclease activity can be used to introduce sequence specific nicks or double strand breaks for genome editing. Cas9 proteins without nuclease activity can be used to activate or suppress transcription of genes, label DNA sequence specific, tether DNA or support homologous recombination (Adapted from: Mali *et al.*, 2013).

binding complex, which can interfere with transcriptional elongation, RNA polymerase binding or transcription factor binding, depending on the recognition site on the genome of the sgRNA (Qi *et al.*, 2013). If the sgRNA is complementary to a the promoter region, binding of sgRNA:Cas9 blocks transcription initiation. Alternatively the elongation process can be stopped by introducing an sgRNA, which is identical to a region within the gene of interest. In this case, the sgRNA:Cas9 complex will bind to the non-template strand of the gene. Furthermore, CRISPRi can repress gene expression of several genes at the same time if more than one sgRNA is used (Qi *et al.*, 2013). The system is very effective in shutting down gene expression in bacteria (Bikard *et al.*, 2013). In eukaryotes dCas9 had to be fused to a chromatin modifier domain (KRAB, Mxi1) to improve the efficiency of the system (Gilbert *et al.*, 2013). With a few modifications the CRISPRi machinery can also be used to **activate gene expression**. In eukaryotes dCas9 can be fused to different transcription activator domains (Zhang *et al.*, 2012). In prokaryotes dCas9 was fused to the ω subunit of RNA polymerase (RNAP) to achieve sequence-dependent gene activation (Bikard *et al.*, 2013). The sgRNA used for transcription activation has to bind in the

promotor region to allow transcription initiation.

Other Cas proteins are used in the development of research-based applications as well. Cas6f is used in some applications concerning translation efficiency of mRNA and mRNA localization. To make translation efficiency more predictable, the cleavage site sequence of a CRISPR repeat to a gene of interest. Cas6f catalyzes cleavage of the repeat site, which physically separates genetic elements on the mRNA level. This reduces the amount of secondary structures and thereby makes gene expression more predictable (Qi *et al.*, 2012). Additionally the high RNA binding affinity of Cas6f was used to create a specific RNA-binding bait protein. For this application a version with conditional ribonuclease activity of Cas6f is used. Cas6f binds irreversibly to RNA which is engineered with a 16-nt hairpin sequence at their 5' ends. Consequentially Cas6f can be pulled down together with the RNA and potential binding partners. Contaminating proteins and other molecules can be removed by extensive washing. Upon addition of imidazole, Cas6f is activated to cleave the RNA. This removes the hairpin tag and releases the native transcript along with its specifically bound protein partners (Lee *et al.*, 2013).

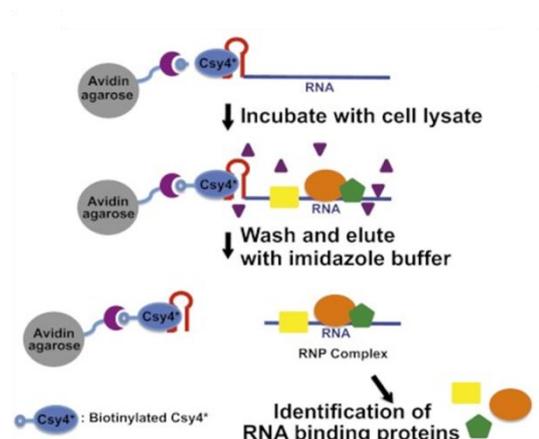


Figure 14 Mechanism of utilization of Cas6f (Csy4) in an RNA-binding bait strategy. (Taken from: Lee *et al.*, 2013)

A For some of the applications currently used in the research environment one could imagine further applications in biotechnology. Those are going to be discussed in the next chapter.

7. CRISPR/Cas in biotechnology

There are many prospective uses for the CRISPR/Cas-based applications in biotechnology. On two branches of biotechnology the CRISPR/Cas-based applications could have a major impact so they are going to be discussed in more detail.

7.1. CRISPR/Cas as a protection mechanism for large-scale bacterial fermentation processes

Microorganisms and especially bacteria are of special importance in biotechnology. Advantages of microorganisms are that they are fast growing, because they are easy to cultivate and most of all easy to manipulate for different purposes. Some important

examples for the utilization of bacteria in biotechnology are: production of cheese and yoghurt (*Lactococcus ssp.*, *Lactobacillus ssp*) and *E. coli* (production of human insulin, vitamin B, enzymes of all kinds etc.) CRISPR/Cas systems are more common in archea than in bacteria Only about 40% of sequenced bacteria possess one or more CRISPR/Cas systems (Grissa *et al.*, 2007). Especially well-established bacterial species used in research or biotechnology often lack the CRISPR/Cas system, probably due to their rare exposure to bacteriophages (Marraffini and Sontheimer, 2010). Bacterial strains used in big-scale biotechnology are very valuable and a bacteriophage infection could lead to total production loss in the worst case. Especially vulnerable to phage infections are bacteria used in food fermentation processes for the production of fermented dairy products because they are often grown in large vats and under non-sterile conditions. Phage infection of dairy starter cultures disrupts normal fermentation cycles, stalls the manufacturing chain and decreases the quality of the finished product. Protection of bacteria used in food fermentation is even more challenging since genetic modification is strictly regulated. From an economical point of view genetically modified food might not be as profitable as its non-genetically modified counterpart because a large proportion of the public is not willing to buy products which contains genetically modified ingredients. Therefore using techniques which would

require labelling the food as “genetically modified” should always be avoided.

The CRISPR/Cas system could give an effective protection against phage infections without require labeling of the resulting products as ‘genetically modified’. It is easy to screen for strains with a CRISPR/Cas system. It is conceivable to vaccinate bacterial strains used in biotechnology against common phages e.g. by transforming them with a plasmid that contains a CRISPR/Cas system already equipped with spacers of those phages. Alternatively one could provide the bacteria with a basic CRISPR/Cas system and expose them deliberately to certain phages and select for cells which adapt (integrate new spacers into their CRISPR/Cas systems) without further genetic modification. Nevertheless the number of CRISPRs would be limited by the maximum size of a plasmid.

7.2. Applications in the pharmaceutical industry

Some of the CRISPR/Cas components currently reported to be successful in research could have great perspectives in the pharmaceutical industry. The Cas9 RNA-

guided nucleases, described in Chapter 6, are promising tools that could be used in gene therapy.

Gene therapy is of the big dreams of medicine. With hopes of curing e.g. aids and cancer or find solutions for all kinds of hereditary diseases. Sometimes these diseases are caused by a single nucleotide. Changes in the human genome for the treatment of a such a disease it would be necessary to have a nuclease that specifically recognizes the sequence and introduces a single-strand nick or a double-strand break at the affected site. This of course means that the larger the genome of an organism is the larger the recognition site of the editing nuclease has to be. After the sequence specific cleavage and depending on the damage caused, DNA repair systems need to specifically delete or replace the sequence by homologous or non-homologous recombination (Figure 13). Another application described in Chapter 6 which could be profitably used in the pharmaceutical industry is CRISPRi. CRISPRi could be used as an antiviral drug by specifically preventing transcription of the viral genes, which are essential for viral

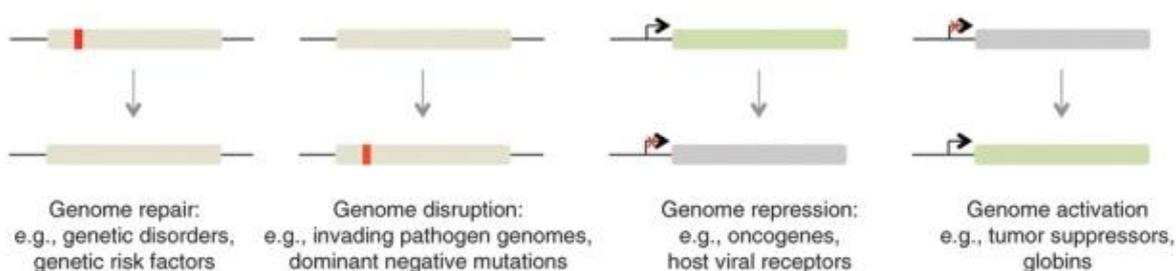


Figure 15 Abilities of Cas9 and its potential fields in therapeutic applications (Adapted from Mali *et al.*, 2013)

infection (Figure 17).

Of course the development of an efficient antiviral treatment is also of interest for other branches in biotechnology such as in crop plant development and in domesticated animals.

However, delivery of a CRISPRi-based antiviral drug might be challenging, especially since RNAi-based approaches for the treatment of viral infections have not been successful for this exact reason. Most of the methods applied in research to deliver the CRISPRi components into mammalian cells are only suitable for single cells (e.g. injection into the nucleus) or cell cultures (e.g. electroporation) but not for complex tissues or whole organisms.

8. Outlook and conclusions – future perspectives for CRISPR/Cas-based applications

Even though the CRISPR/Cas system is still a relatively new discovery it has already proven to be a source for the development of a large variety of applications in many different fields. Most of these are progressing fast but in the midst of this “hype” we should try to see what the current limitations are, what the potential drawbacks of certain applications and which advantages alternative applications could have. This critical discussion will be the last part of the current essay.

Box 7 Epigenetics

Epigenetics is defined as a heritable form of gene expression which are not caused by changes in the genetic information. Epigenetic regulation is caused by DNA modifications, post translational modifications (PTMs) on histones which alter function and accessibility to the DNA and RNA interference mechanism also influence epigenetics.

Examples for DNA modifications are: methylation, hydroxymethylation, formylation and carboxylation. Examples for histone PTMs are acetylation, methylation, ubiquitylation and phosphorylation.

Epigenetics are involved in numerous diseases and disease related processes such as obesity and aging. How DNA and histone modifications influence gene expression is often still unknown. (Egger *et al.*, 2004)

8.1. Next steps in CRISPR/Cas-based developments

CRISPRi already has proven to be a feasible tool to upregulate or repress gene expression. The next step would be to fine tune the control over endogenous transcription levels. Especially when upregulating genes, the transcription level is hard to predict. One limiting factor is genome architecture, which is partly influenced by epigenetics. Still in early development and without any published results yet, is the application of dCas9 to introduce epigenetic markers (Box 7) (Mali *et al.*, 2013; Sander and Joung, 2014). The principle is very simple: Similar to CRISPRi dCas9 is fused to functional domains that recruit chromatin remodeling complexes such as histone acetylases, methylases,

Although more difficult to express and engineer, RGENs based on the PAM sequence-independent Type III-A system could overcome this limitation.

During the last years protein engineering and prediction of protein functioning has improved greatly and it is going to improve even more with the development of more powerful computers. Engineering-dependent genome editing tools like ZNFs or TALENS could profit from this development. By contrast, generation of RGENs is very simple and they are easy to generate for any required sequence. Notwithstanding this, the sequence length is a limiting factor and the formation of secondary structures in the sgRNA might cause a change of activity and/or efficiency of RGENs. Efficiency, especially in mammalian cells, and specificity of RGENs could and still need to be improved, for instance by experimental

evolution.

For a successful disease treatment using RGENs it is vital that the system employed is safe and does not have any off-target activity. It should be efficient enough to successfully cure a disease and cost-effective enough to be considered as treatment. Realistically seen, it is still going to take quite some time until RGENs are developed enough to be used in gene therapy for disease treatment. Applying these genome surgery techniques to correct human disease-associated genetic mutations, resulting in functional gene therapy and in curing genetic disorders, are therefore not foreseen in the near future.

8.3. Perspectives for the protection of large-scale fermentations

Although the idea of being able to vaccinate bacteria used for biotechnology sounds great there are several drawbacks to this idea. The biggest part of the biotechnology industry is the food industry. As the use of GMOs in food is not allowed or heavily restricted in many countries getting permission for large-scale applications in food industry will probably be a long process if even successful: many people avoid buying food containing GMOs. Screening could be an alternative strategy to achieve phage immunity in bacteria exists. It should work very efficiently and would not require labelling the resulting product as GMO.

Box 8 The good phage – uses for phages

- Some phages have a very specific host spectrum and are used to **protect starter cultures** from potentially dangerous bacteria like *Listeria monocytogenes*. Some phages already have a GRAS status, and there are more to come (Henry and Debarbieux, 2012)
- Phages can be used as **biosensors** for fast identification of (multi-resistant) pathogens (Singh *et al.* 2012).
- Phages can be used as replacements for antibiotics in the **treatment of infections** with (multi-resistant) bacteria.

For further information about these and more virus-based applications read: Henry and Debarbieux, (2012)

In my opinion, a more urgent issue that should be considered is whether big-scale vaccinations could cause the spread of resistant phages.

Phages can be as well useful and they are not utilized to their full potential yet (Box 8). It is possible, that the CRISPR/Cas system, in a way similar to antibiotic resistance, could spread in (pathogenic) bacteria once it is used in big-scale applications. This is especially alarming since there is evidence for horizontal gene transfer in the evolution of different CRISPR/Cas systems (Portillo and Gonzales, 2009). These issues need to be clarified before CRISPR/Cas can be used in big-scale applications.

9. References

- Abedon, S.T., Murray, K.L.** (2013) Archaeal viruses, not archaeal phages: an archaeological dig. *Archaea*, 2013:251245.
- 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):1709-12.
- Barzel, A., Privman, E., Peeri, M., Naor, A., Shachar, E., Burstein, D., Lazary, R., Gophna, U., Pupko, T., Kupiec, M.** (2011) Native homing endonucleases can target conserved genes in humans and in animal models. *Nucleic Acids Res.*;39(15):6646-59.
- Bergh, O., Borsheim, K.Y., Bratbak, G., Heldal, M.** (1989) High abundance of viruses found in aquatic environments. *Nature* 340:467-68
- Bernstein, E., Caudy, A. A., Hammond, S. M., Hannon, G. J.** (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409, 363-366.
- Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., Marraffini, L.A.** (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* 41:7429-37.
- Bolotin, A., Quinquis, B., Sorokin, A., Ehrlich, S.D.** (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*;151(Pt 8):2551-61.
- Bondy-Denomy, J., Pawluk, A., Maxwell, K.L., Davidson, A.R.** (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, 493, 429-432.
- Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., van der Oost, J.** (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*; 321(5891):960-4.
- Chang, N., Sun, C., Gao, L., Zhu, D., Xu, X., Zhu, X., Xiong, J.W., Xi, J.J.** (2013) Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res.*; 23(4):465-72.
- Charpentier, E., Doudna, J.A.** (2013) Biotechnology: Rewriting a genome. *Nature.*;495(7439):50-1
- Chen, Z., Yang, H., Pavletich, N.P.** (2008) Mechanism of homologous recombination from the RecAssDNA/dsDNA structures. *Nature* 453:489-84
- Cho, S.W., Kim, S., Kim, J.M., Kim, J.S.** (2013) Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.*;31(3):230-2.
- Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., Voytas, D.F.** (2010) Targeting DNA double-strand breaks with

- TAL effector nucleases. *Genetics*;186(2):757-61.
- Dale, J.W., Brittain, D., Cataldi, A.A., Cousins, D., Crawford, J.T., Driscoll, J., Heersma, H., Lillebaek, T., Quitugua, T., Rastogi, N., Skuce, R.A., Sola, C., Van Soolingen, D., Vincent, V.** (2001) Spacer oligonucleotide typing of bacteria of the *Mycobacterium tuberculosis* complex: Recommendations for standardised nomenclature. *Int. J. Tuberc. Lung Dis.*, 5, 216–219.
- 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* 471:602–7
- Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., Moineau, S.** (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol*, 190: 1390–1400.
- Doyle, M., Fookes, M., Ivens, A., Mangan, M.W., Wain, J., Dorman, C.J.** (2007) An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science* 315:251–52
- Egger, G., Liang, G., Aparicio, A., Jones, P.A.** (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature*;429(6990):457-63.
- Elbashir SM, Lendeckel W, Tuschl T.** (2001a) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*; 15(2):188-200.
- Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., Tuschl, T.** (2001b) Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.*;20(23):6877-88.
- Fire, A., Albertson, D., Harrison, S.W., Moerman, D.G.** (1991) Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle. *Development*; 113(2):503-14.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C.** (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*; 391, 806-811.
- Gaj, T., Guo, J., Kato, Y., Sirk, S.J. & Barbas, C.F.** (2012). Targeted gene knockout by direct delivery of zinc-finger nuclease proteins. *Nat. Methods* 9, 805–807
- Gasiunas, G., Siksnys, V.** (2013) RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? *Trends Microbiol.*; 21(11):562-7.
- Ghildiyal, M., Zamore, P.D.** (2009) Small silencing RNAs: an expanding universe. *Nat Rev Genet.*;10(2)
- Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., Weissman, J.S., Qi, L.S.** (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154:442–51
- Grissa, I., Vergnaud, G., Pourcel, C.** (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*; 8:172.
- 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 Comput Biol.*;1(6):e60.
- Hammond, S.M., Bernstein, E., Beach, D., Hannon, G.J.** (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature*; 404(6775):293-6.
- Hatoum-Aslan, A., Maniv, I., Marraffini, L.A.** (2011) Mature clustered, regularly interspaced, short palindromic repeats RNA

(crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc Natl Acad Sci USA*;108(52):21218-22.

Henry, M., Debarbieux, L. (2012) Tools from viruses: Bacteriophage successes and beyond. *Virology*;434(2):151-61

Horvath, P., Romero, D.A., Côté-Monvoisin, A.C., Richards, M., Deveau, H.; Moineau, S., Boyaval, P., Fremaux, C., Barrangou, R. (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.*, 190, 1401–1412.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., Nakata, A. (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 169, 5429–5433.

Jansen, R., Embden, J.D., Gaastra, W., Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol.*;43(6):1565-75.

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* 337:816–21

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., Beijer, M.R., Barendregt, A., Zhou, K., Snijders, A.P., Dickman, M.J., Doudna, J.A., Boekema, E.J., Heck, A.J., van der Oost, J., Brouns, S.J. (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.*, 18, 529–536.

Jore, M.M., Brouns, S.J., van der Oost, J. (2012) RNA in defense: CRISPRs protect prokaryotes against mobile genetic elements. *Cold Spring Harb Perspect Biol*;4(6).

Jansen, R., Embden, J.D., Gaastra, W., Schouls, L.M. (2002) Identification of genes

that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.*, 43, 1565–1575.

Kunin, V., Sorek, R., Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 8:R61

Labrie, S.J., Samson, J.E., Moineau, S. (2010) Bacteriophage resistance mechanisms. *Nat Rev Microbiol*;8(5):317-27.

Lee, H.Y., Haurwitz, R.E., Apffel, A., Zhou, K., Smart, B., Wenger, C.D., Laderman, S., Bruhn, L., Doudna, J.A. (2013) RNA-protein analysis using a conditional CRISPR nuclease. *Proc Natl Acad Sci USA*;110(14):5416-21.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rådmark, O., Kim, S., Kim, V.N. (2003) The nuclear RNase III Drosha initiates microRNA processing. *Nature*; 425(6956):415-9.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., Kim, V.N. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.*;23(20):4051-60.

Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q., Copié, V., Young, M.J., White, M.F., Lawrence, C.M. (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.* 286:21643–56

Lord, C.J., Ashworth, A. (2012) The DNA damage response and cancer therapy. *Nature.* 2012;481(7381):287-94.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J. E., Kutay, U. (2004) Nuclear export of microRNA precursors. *Science*; 303, 95–98.

Makarova, K.S., Aravind, L., Grishin, N.V., Rogozin, I.B., Koonin, E.V. (2002) A DNA repair system specific for thermophilic

- Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res* 30: 482–496.
- Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., Koonin, E.V.** (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*; 1:7.
- Makarova, K.S., Aravind, L., Wolf, Y.I., Koonin, E.V.** (2011a) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct*, 6, doi:10.1186/1745-6150-6-38.
- 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., van der Oost, J., Koonin, E.V.**(2011b) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.*; 9, 467–477.
- Marchese, F.P., Huarte, M.** (2013) Long non-coding RNAs and chromatin modifiers: Their place in the epigenetic code. *Epigenetics*;9(1). [Epub ahead of print]
- Marraffini, L.A., Sontheimer, E.J.** (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genetics*, 11, 181–190.
- Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., Soria, E.**(2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol.*;60(2):174-82.
- Napoli, C. A., Lemieux, C., Jorgensen, R.** (1990) Introduction of a chimeric chalcone synthetase gene in *Petunia* results in reversible co-suppression of homologous genes in trans. *Plant Cell* 2, 279–289
- van der Oost, J.** (2013) Molecular biology. New tool for genome surgery. *Science*; 339(6121):768-70.
- Portillo, M.C., Gonzalez, J.M.** (2009) CRISPR elements in the Thermococcales: evidence for associated horizontal gene transfer in *Pyrococcus furiosus*. *J Appl Genet.*;50(4):421-30.
- Pourcel, C., Salvignol, G., Vergnaud, G.** (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*; 151(Pt 3):653-63.
- Pul, Ü., Wurm, R., Arslan, Z., Geissen, R., Hofmann, N., Wagner, R.** (2010) Identification and characterization of *E. coli* CRISPR-cas promoters and their silencing by H-NS. *Mol. Microbiol.*, 75, 1495–1512.
- Qi, L., Haurwitz, R.E., Shao, W., Doudna, J.A., Arkin, A.P.** (2012) RNA processing enables predictable programming of gene expression. *Nat. Biotechnol.*; 30, 1002–1006.
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A.** (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*;152(5):1173-83
- Richter, C., Chang, J.T., Fineran, P.C.** (2012) Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems. *Viruses*;4(10):2291-311.
- Richter, H., Randau, L., Plagens, A.** (2013) Exploiting CRISPR/Cas: interference mechanisms and applications. *Int J Mol Sci.*;14(7):14518-31.
- Röther, S., Meister, G.** (2011) Small RNAs derived from longer non-coding RNAs. *Biochimie*;93(11):1905-15.
- Samson, J.E., Magadán, A.H., Sabri, M., Moineau, S.,** (2013) Revenge of the phages: defeating bacterial defences. *Nat. Rev. Microbiol.*:11,675–687
- Sander, J.D., Joung, J.K.** (2014) CRISPR-Cas systems for editing, regulating and

- targeting genomes. *Nat Biotechnol.* doi: 10.1038/nbt.2842. [Epub ahead of print]
- Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., Zamore, P.D.** (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell*; 115(2):199-208.
- Shinkai, A., Kira, S., Nakagawa, N., Kashihara, A., Kuramitsu, S., Yokoyama, S.** (2007) Transcription activation mediated by a cyclic AMP receptor protein from *Thermus thermophilus* HB8. *J. Bacteriol.* 189:3891–901
- Singh, A., Arutyunov, D., Szymanski, C.M., Evoy, S.** (2012) Bacteriophage based probes for pathogen detection. *Analyst.*;137(15):3405-21
- 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. *EMBO J.*, 30, 1335–1342.
- Siomi, M.C., Sato, K., Pezic, D., Aravin, A.A.** (2011) PIWI-interacting small RNAs: the vanguard of genome defence. *Nat. Rev. Mol. Cell Biol.*; 12(4): 246-58
- Sorek, R., Lawrence, C.M., Wiedenheft, B.** (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem.*;82:237-66.
- Stern, A., Keren, L., Wurtzel, O., Amitai, G., Sorek, R.** (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet.* 26:335–40
- Swarts, D.C., Mosterd, C., van Passel, M.W., Brouns, S.J.** (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One.*7(4):e35888.
- Terns, R.M., Terns, M.P.** (2014) CRISPR-based technologies: prokaryotic defense weapons repurposed. *Trends Genet.*;30(3):111-118.
- Timmons, L., Fire, A.** (1998) Specific interference by ingested dsRNA. *Nature*; 395(6705):854.
- Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S., Gregory P.D.** (2010) Genome editing with engineered zinc finger nucleases *Nat. Rev. Genet.*, 11 (9): 636–646
- Wei, C., Liu, J., Yu, Z., Zhang, B., Gao, G., Jiao, R.** (2013) TALEN or Cas9 - rapid, efficient and specific choices for genome modifications *J Genet Genomics.*;40(6):281-9.
- Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A., Nogales, E.** (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* 477:486–89
- Westra, E.R., Swarts, D.C., Staals, R.H., Jore, M.M., Brouns, S.J., van der Oost, J.** (2012) The CRISPRs, they are a-changin': how prokaryotes generate adaptive immunity. *Annu Rev Genet.*;46:311-39.
- Yosef, I., Goren, M.G., Qimron, U.** (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.*, 40, 5569–5576.
- Yosef, I., Shitrit, D., Goren, M.G., Burstein, D., Pupko, T., Qimron, U.** (2013) DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array. *Proc Natl Acad Sci U S A*;110(35):14396-401
- Zamore, P. D., Tuschl, T., Sharp, P. A. & Bartel, D. P.** (2000) RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101, 25-33.
- Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S., Naismith, J.H., Spagnolo, L., White, M.F.** (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell.*, 45:303–13.

Zhang, Y.Q., Li, Y.Z., Wang, B., Wu ZH, Zhang, C.Y., Gong, X., Qiu, Z.J., Zhang, Y. (2005) Characteristics and living patterns of marine myxobacterial isolates". *Applied Environmental Microbiology* ;71(6):3331-3336.