DISSERTATION

Titel der Dissertation

"The *Sulfolobus solfataricus* CRISPR-Cas system"

Verfasser

Andrea Manica

angestrebter akademischer Grad

Doktor der Naturwissenschaften (Dr. rer. nat.)

Wien, 2013

Studienkennzahl lt. Studienblatt:	A 091 490
Dissertationsgebiet lt. Studienblatt:	Molekulare Biologie
Betreuerin:	Prof. Dr. Christa Schleper

“EVERYBODY IS A GENIUS. BUT IF YOU JUDGE A FISH BY ITS ABILITY TO CLIMB A TREE,
IT WILL LIVE ITS WHOLE LIFE BELIEVING THAT IT IS STUPID.”

-A. Einstein -

TABLE OF CONTENTS

1 ZUSAMMENFASSUNG

Das Vorkommen von Viren in allen Habitaten der Erde und ihre Koevolution mit Mikroorganismen führt zu einem hohen Selektionsdruck auf die Wirte. Dies führte zu der Entwickung einer großen Bandbreite von Resistenzmechanismen in Archaeen und Bakterien um viralen Infektionen vorzubeugen oder sie zu reduzieren. Erst kürzlich wurde ein neuer Mechanismus entdeckt, der mikrobielle Zellen vor Viren und anderen eindringenden genetischen Elementen schützt: das CRISPR-Cas System. Seine Aktivität beruht auf kleinen RNAs (crRNAs), die in einen Proteinkomplex (Cas Proteine) inkorporiert werden, welche die fremden DNA- oder RNA-Sequenzen, bei gegebener Komplementarität zwischen der crRNA und der Zielsequenz, degradiert.

Während meiner Doktorarbeit habe ich das CRIPSR-Cas System in dem hyperthermophilen Archaeon *Sulfolobus solfataricus* untersucht. Dieser Organismus besitzt ein umfangreiches CRISPR-Cas Set mit sechs CRISPR-Loci und fünf Cas-interferierenden Komplexen, die zusammen beinahe 10% des Genoms ausmachen.

Ein rekombinanter Shuttle-Vektor basierend auf dem *Sulfolobus* Virus 1 (SSV1) wurde verwendet um das CRISPR-Cas System von *Sulfolobus* herauszufordern und dessen Aktivität über die Anzahl der bei der Infektion entstehenden Virus-Plaques zu quantifizieren. Mit diesem Ansatz wurde gezeigt, dass Virussequenzen mit Komplementarität zu den CRISPR- Spacer Sequenzen (sogenannte Protospacer) die Fähigkeit des Viruses die Wirtszelle zu infizieren beeinträchtigen. Bei einer perfekten Komplementarität zwischen der Protospacer-Sequenz und einem bestimmten CRISPR-Locus war der Virus nicht mehr in der Lage seinen Wirt zu infizieren. Die verschiedenen CRISPR-Loci variierten in ihrer Effizienz, was wiederum mit der Transkription des entsprechenden Locus korrelierte. Der offene Leserahmen der die Protospacer-Sequenz auf dem Virusgenom kodiert war nicht essentiell für die Virusinfektion und beinhaltete keinen Transkriptionspromotor, was daraufhin weist, dass eher die DNA als die

RNA als Zielsequenz für das CRISPR-Cas System in diesem Assay dient. Der gleiche Assay wurde benutzt, um die Fähigkeit des Systems zu quantifizieren Protospacer mit geringerer Sequenzähnlichkeit als Zielsequenz zu erkennen. Die Anzahl der Mutationen die zwischen crRNA und Protospacer toleriert werden können, ist abhängig von der Position der Mutation innerhalb der Protospacer-Region. Die Mutationen, die die Hybridisierung mit der 5´seitigen Hälfte des Spacers beeinflussen, hatten einen starken Effekt auf die Zielerkennung, wobei sechs Fehlpaarungen innerhalb der ersten acht Nukleotide die DNA Interferenz-Aktivität beinahe vollständig verhinderten. Trotzdem war das System noch in der Lage den viralen Vektor effizient zu erkennen und zu degradieren, selbst wenn 14 Mutationen in die andere Hälfte der Hybridisierungsregion eingeführt wurden. Eine Mutation auf beiden Seiten des crRNA-Protospacer-Hybrids wirkten synergetisch und führten zu einer Abschwächung der DNA-Interferenz-Aktivität. Wir untersuchten ebenfalls den potentiellen Einfluss von Sequenzen außerhalb der Protospacer-Region auf die CRISPR-Aktivität. Es wurde berichtet das diese, dem Protospacer-benachbarten Motive (protospacer adjacent motifs, PAM), neben der Funktion des Erwerbs neuer Spacer auch einen Einfluss auf die Interferenz-Aktivität in einigen bakteriellen Systemen haben. Es konnte jedoch kein Hinweis für diese Rolle der Sequenzen im *Sulfolobus*-Assay gefunden werden. Nukleotide, die am 3´Ende des Protospacers lokalisiert waren, verhinderten eine CRISPR-Aktivität, sofern sie komplementär zum 8 nt-Anhang der crRNA waren. Die drei Positionen -3, -4 und -5 auf dem crRNA-Protospacer-Hybrid verhinderten eine Interferenz. Damit sind sie für das System wichtig zur Unterscheidung zwischen eigener und fremder DNA, ähnlich wie es bereits in bakteriellen Systemen nachgewiesen wurde.

Zusammenfassend wurde in dieser Arbeit ein archaeales CRISPR-Cas System detalliert im hyperthermophilen *S. solfataricus* untersucht. Das System weist einige Gemeinsamkeiten aber auch klare Unterschiede zu den bereits beschriebenen bakteriellen Systemen auf. Im Besonderen weist die Toleranz gegenüber einer großen Anzahl von Mutationen in der Zielerkennung darauf hin, dass die CRISPR-vermittelte Virusverteidigung

nicht nur als Verteidigung der ersten Stufe angesehen werden sollte, sondern als allgemeiner Mechanismus, der den Wirtszellen ermöglicht die Virenzahl längerfristig niedrig zu halten. Diese Resultate bezüglich der Erkennung der DNA als Fremd oder Eigen ebnen den Weg für zukünftige Studien an RNA-Interferenz in Archaea, da sie einen Mechanismus demonstrieren mit dem DNA-Interferenz in einem Organismus verhindert werden kann.

2 SUMMARY

The persistence of viruses in all environments on Earth and their co-evolution with microorganisms imposes high selective pressures on the hosts. This has led to the development of a portfolio of resistance mechanisms in Archaea and Bacteria to prevent or reduce viral infections. Recently, a new mechanism was discovered which protects microbial cells from viruses and other invading genetic elements: the CRISPR-Cas system. Its activity is based on small RNAs (crRNAs) that are incorporated into a protein complex (Cas proteins) which degrades foreign DNA or RNA sequences, respectively, when complementarity between the crRNA and the target is given.

During my PhD thesis I studied the CRISPR-Cas system in the hyperthermophilic archaeon *Sulfolobus solfataricus*. This organism possesses an extensive CRISPR-Cas set, with six CRISPR loci and five Cas "interference complexes" which together occupy almost 10% of the chromosome.

Using a recombinant shuttle vector based on the *Sulfolobus* virus SSV1, the CRISPR-Cas system of *Sulfolobus* was challenged and its activity was quantified by counting virus plaques upon transfection. With this approach it was demonstrated that viral sequences with complementarity to CRISPR spacer sequences (so-called protospacers) affect the ability of the virus to infect the host cell. When perfect complementarity between protospacer sequence and spacers of certain CRISPR loci was given, the virus completely lost its ability to infect the host. The different CRISPR loci varied in their efficiency which correlated with the transcription of the respective locus. The open reading frame carrying the protospacer sequence on the virus genome was not essential for virus infection and did not carry a transcriptional promoter, both indicating that DNA rather than RNA was a target of the CRISPR-Cas system in this assay. The same genetic tool was used to quantify the ability of the system in targeting less similar protospacers. The amount of mutations tolerated between crRNA and protospacer varied depending on the

location of the mutations along the protospacer region. Mutations influencing the hybrid formation with the 5´half of the spacer had a strong effect on target recognition, with six mismatches within the first 8 nucleotides almost completely abolishing the DNA interference activity. But the system was still able to efficiently recognize and degrade the viral vector even when 14 mutations were introduced in the other half of the hybridization region. A synergistic effect of mutations occurring simultaneously at both sides of the crRNA-protospacer hybrid was observed in attenuating the DNA interference activity. We also investigated the potential role of sequences located outside the protospacer region on CRISPR activity. Protospacer adjacent motifs (PAM), besides having a function in spacer recruitment, had been reported to influence interference activity in some bacterial systems. However, no indication of a role of this sequence was found in the *Sulfolobus* assay. However, nucleotides located 3' to the protospacer abolished the CRISPR activity when they were complementary to the 8nt-tag of the crRNA. The three positions -3, -4 and -5 of the crRNA-protospacer hybrid abolished interference. They were thus important for the system to discriminate self from non-self DNA in a similar way as found earlier in bacterial system.

In conclusion, an archaeal CRISPR-Cas system was studied in detail in the hyperthermophile *S. solfataricus*. It exhibits some commonalities but also clear differences to the currently described bacterial systems. In particular, the tolerance of a large number of mismatches in target recognition indicates that CRISPR-mediated virus defense should not only be considered a "first-barrier-defense-system" but may be seen as a broader mechanism that allows host cells to cope with viruses keeping them at reduced levels. The results on self/non-self recognition pave the way for future studies on RNA interference in Archaea, because they demonstrate a mechanism how DNA interference can be circumvented in the organism.

2.1 List of publication included in this thesis

The following publications and manuscripts are included in this thesis. Their text was reformatted but the original content of the different papers and figures has been maintained.

Andrea Manica and Christa Schleper

CRISPR-Mediated Defense Mechanisms in the Hyperthermophilic Archaeal Genus *Sulfolobus*. RNA Biol. 2013 Mar 27;10(5).

Brief description: This review paper describes in details the progress made in the understanding of the CRISPR-Cas system of the genus S*ulfolobus*, the most studied archaeal genus regarding this new defense mechanism. Insights into the diversity of the *Sulfolobales* CRISPR-Cas system, crRNA biogenesis and targeting are discussed in detail, as well the *in vitro* and *in vivo* activity of the Cas complexes.

My contribution: Structuring and writing of the first draft of the manuscript, including preparation of all figures.

Andrea Manica, Ziga Zebec, Daniela Teichmann, and Christa Schleper

***In vivo* Activity of CRISPR-Mediated Virus Defense in a Hyperthermophilic Archaeon**. Mol Microbiol. 2011 Apr;80(2):481-91.

Brief description: In this paper we describe the first example of CRISPR-Cas-mediated viral immunity in archaea. Viral DNA carrying a protospacer sequence is unable to transfect *S. solfataricus* cells expressing a homologous crRNA. Our experiments also demonstrate that the CRISPR-Cas system of *S. solfataricus* targets DNA and not RNA

My contributions: Conceptually and experimentally establishing the system, performing most of the experiments, preparing most of the figures for the manuscript.

A. Manica*, Z. Zebec*, J. Steinkellner, and C. Schleper
(*co-first authorship)

Unexpectedly Broad Target Recognition Ability of the CRISPR-mediated Virus Defense System in *Sulfolobus solfataricus*.

Manuscript accepted for publication in Nucleic Acid Research

Brief description: In this paper we analyzed in detail the requirements of the *S. solfataricus* CRISPR-Cas system to trigger immunity against DNA viruses. Our assay demonstrates that the system is very promiscuous allowing the targeting of highly degenerated protospacers and that additional sequences outside the protospacer sequence are not required for its activity. Furthermore, the requirements of the system to discriminate self from non-self DNA were identified.

My contributions: Experimental design and execution of about half of all experiments, preparation of figures and writing of the first manuscript draft.

3 AIM OF THE STUDY

In 2009 when I started working on the CRISPR system of *S. solfataricus* only few publications were available on CRISPR activity in Archaea. At that time CRISPR activity had been demonstrated *in vivo* only in two bacteria: *E. coli* and *S. thermophilus* (Barrangou *et al.*, 2007, Brouns *et al.*, 2008). Although several bioinformatic analyses showed the presence of such systems also in Archaea, no data were available about the *in vivo* properties of CRISPR-Cas in Archaea. The only publication from Archaea that became available at the beginning of my thesis was an *in vitro* study in *P. furiosus* (Hale *et al.*, 2009) in which RNA degradation mediated by one particular Cas module (Cmr) was demonstrated. This result opened the big question if the CRISPR-Cas systems in Archaea would in general only target RNA, similar to the Eukaryotic RNAi systems, or instead would target DNA as found in the bacterial counterparts or both.

The first part of my PhD work was centred to answer this question by establishing a system that allows to study CRISPR-Cas activity *in vivo* in *S. solfataricus*. A genetic system based on recombinant variants of the virus SSV1 (*S. shibatae* virus 1) had been developed in this organism (Jonuscheit *et al.*, 2003), which was a perfect tool for establishing an *in vivo* test system for CRISPR-Cas activities and to gain detailed insights into its function by challenging the host with recombinant viral variants as targets. In addition, a mini-CRISPR cassette targeting the host chromosome was constructed and introduced into the virus genome in order to study its activity and persistence in the host cells.

After establishing the system and demonstrating DNA interference in *Sulfolobus,* the second major goal was to gain additional insights into the peculiarities of the activity. For this purpose a multitude of virus variants were constructed based on a newly implemented gateway cloning cassette. They were used to understand: (i) the minimal similarity required between crRNA and protospacer which would allow interference, (ii) the "weight" of the different mutations along the protospacer region; (iii) the role of

protospacer flanking sequences during the DNA interference process; (iv) the mechanism of discrimination between self and non-self and (v) ways to abolish DNA interference which would allow to further study the RNAi activity of the Cmr module *in vivo*.

3.1 Reference

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

Brouns, S. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin & J. van der Oost, (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.

Hale, C. R., P. Zhao, S. Olson, M. O. Duff, B. R. Graveley, L. Wells, R. M. Terns & M. P. Terns, (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945-956.

Jonuscheit, M., E. Martusewitsch, K. M. Stedman & C. Schleper, (2003) A reporter gene system for the hyperthermophilic archaeon Sulfolobus solfataricus based on a selectable and integrative shuttle vector. *Mol Microbiol* **48**: 1241-1252.

4 INTRODUCTION

4.1 Archaea and the environment

For a long time during the 20th century prokaryotes were considered to represent a uniform group of bacteria. Only in 1977 through the analysis of ribosomal RNA libraries performed by Carl Woese and colleagues, the "prokaryotic world" was further divided into two distinct domains of life: Bacteria and Archaebacteria (later Archaea) (Woese & Fox, 1977; Woese *et al.*, 1990).

This seminal finding of Woese was subsequently supported by molecular and biochemical analyses which revealed further characteristics specific for the archaeal cell clearly separating the Archaea from the bacterial domain. On the cellular level Archaea are very similar to Bacteria lacking a nucleus and organelles. Both often possess a single circular chromosome and their genes are organized in operon-like structures (Brown and Doolittle, 1997). However, Archaea share striking similarities with the Eukarya. In particular their DNA-replication, transcription and translation machineries resemble those of eukaryotes (Garrett and Klenk, 2007).

Studies conducted on the archaeal RNA polymerase have demonstrated early on how this enzyme is highly similar with respect to complexity and sequence similarity to its eukaryotic counterparts (Huet *et al.*, 1983; Bell and Jackson, 1998).

Furthermore, homologous proteins to the eukaryotic Transcription Factor B (TFB), TATA binding protein, ubiquitin, ribosomal proteins and many others were found in Archaea. Although all these proteins are often less complex in Archaea they indicate a common evolution between Archaea and Eukarya clearly distinct from Bacteria (Bell and Jackson, 1998; Leipe *et al.*, 1999; Grabowski and Kelman, 2003; Londei, 2005; Ishino and Ishino, 2012).

Archaea also possess characteristic features which distinguish them from all the other domains of life. For example the archaeal membrane lipids are not composed of fatty acids like in Bacteria or Eukarya, but of isoprenoids, which

are ether-linked to glycerol (Kates, 1978). Instead of a double-layered lipid membrane typical for Bacteria many archaea have a single layer membrane composed exclusively or mostly by glycerol-dialkyl-glycerol tetraether lipids (GDGT). Additionally, penta- and hexacyclic ring structures were identified in the side chains (Weijers *et al.*, 2006). Those unique membrane-characteristics make the archaeal cell resistant to harsh conditions such as low pH and high temperature and often the membrane composition can vary depending on various environmental conditions (Matsumi *et al.*, 2011).

When the separation of Bacteria and Archaea into two domains of life was first proposed by Woese the only species placed in the domain Archaea were methanogens. But shortly after it was recognized that previously known hyperthermophile and halophiles also belong to this domain (Woese *et al.*, 1990). Many new archaeal species were isolated from hot springs and solfataric fields thanks to the work of Wolfram Zillig and Karl Stetter (Stetter, 2006a; Stetter, 2006b). Due to the ability of these first isolates to survive extreme conditions such as high salt concentrations, heat, low pH and low oxygen concentrations the Archaea were thought to be confined to extreme environments and often this is still propagated in textbooks. This view has, however, dramatically changed over the past two decades, based on molecular and physiological analyses, that led to the discovery of ammonia oxidizing and anaerobic methane oxidizing archaea as well as several other lineages (Konneke *et al.*, 2005; Hallam *et al.*, 2006; Brochier-Armanet *et al.*, 2008).

4.2 The hyperthermophilic model organism *Sulfolobus solfataricus*

Sulfolobus solfataricus, the organism used in this study was isolated from a hot mud pool (solfatara) of volcanic origin in the South of Italy by Zillig and co-workers in 1980 (Zillig, 1980; Zillig *et al.*, 1981). It belongs to the hyperthermophilic Crenarchaeota and its order Sulfolobales is represented by the genera Sulfolobus, Stygiolobus, Metallosphaera, Acidianus and Sulfurisphaera, which have been isolated from geothermally active regions all

over the world, such as Yellowstone National Park (USA), New Zealand, El Salvador, Iceland, Italy and Kamchatka (Russia) (Rice *et al.*, 2001).

The members of these genera show a characteristic coccoid morphology, low genomic G+C content (between 30-45 %) and are all (hyper)thermophiles (optimal growth temperature between 75 and 88°C) (Stetter, 1990).

Within this genus, *S. solfataricus* is one of the best studied species. Its cells are 1-2 µm in diameter (Zillig, 1980) and it can be easily cultivated and maintained under laboratory conditions. *S. solfataricus* grows aerobically at 80°C, pH 2-4 using different sources of carbohydrates and amino acids with a generation time of ~8 hours (Grogan, 1989). Furthermore, isolation of single colonies is possible after plating cell suspensions on gelrite plates (Grogan, 1989).

Sulfolobales, and in particular *S. solfataricus*, have attracted many researchers to study the physiology and molecular biology of this genus, due to its easy cultivation, its ability to grow in harsh conditions and the similarity of its rather simple processing machinery compared to the eukaryotic counterpart (Silvian *et al.*, 2001; Boudsocq *et al.*, 2001; Kelman and White, 2005; Shuck *et al.*, 2008).

4.3 Genetic elements of Sulfolobales species

Hot springs are not only the habitat of hyperthermophilic bacteria and archaea, but also many viruses and genetic elements propagate with their hosts in these environments. Several virus types were isolated from hot pools all over the world, which possess very diverse and characteristic morphologies (Reiter *et al.*, 1988; Zillig *et al.*, 1996; Prangishvili and Garrett, 2004; Prangishvili and Garrett, 2005; Prangishvili *et al.*, 2006a; Prangishvili *et al.*, 2006b). Although they are very diverse in shape (rigid rods, filaments, spindle-shaped and spherical virions), almost all virions isolated contain a double stranded (ds) DNA genome. Many of the viruses which infect Sulfolobales species are not lytic. However, their infection can impose strong growth retardation on the host (Prangishvili *et al.*, 2006a).

The discovery of viruses and plasmid-like genetic elements enabled the development of genetic tools to study the physiology and molecular biology of this genus.

The first genetic element exploited as a genetic tool was the S. shibatae virus 1 (SSV1) (Marin *et al.,* 1984). SSV1 can infect several Sulfolobales species and the infection can be easily visualized by a plaque assay (Schleper *et al.*, 1992). SSV1 has a circular genome of about 15 kb containing 34 predicted ORFs (Palm *et al.*, 1991). After infection, the circular viral genome integrates site specifically into the host genome within a tRNA gene (Reiter *et al.*, 1989). Virus propagation can be induced by UV-light, radiation or replication inhibitors such as mitomycin C (Martin *et al.*, 1984; Reiter *et al.*, 1987; Schleper *et al.*, 1992). The induction of the SSV1 genes after UV treatment was in depth studied by Froels and colleagues (2008) (Frols *et al.*, 2008). Due to its small circular dsDNA genome and the ability to be maintained in culture, SSV1 was the first genetic element used to develop a viral vector to express exogenous genes and to be used for the quantification of gene expression in *S. solfataricus* (Jonuscheit *et al.*, 2003; Albers *et al.*, 2006). The *Sulfolobus-E. coli*-shuttle vector pMJ was developed inserting a pUC18 plasmid carrying the *S. solfataricus* pyrEF gene within the SSV1 genome (Martusewitsch *et al.*, 2000; Jonuscheit *et al.*, 2003; Albers *et al.*, 2006). This integrative shuttle vector together with the development of the uracil auxotrophic mutant PH1-M16 (Martusewitsch *et al.*, 2000) allowed the first gene expression study in Sulfolobales. The shuttle vector, maintains the ability to form viable viral particles, although its UV-induced proliferation is reduced. Furthermore, naked viral DNA can be easily extracted from a growing culture and can be electroporated into competent *S. solfataricus* cells. Plaque assay protocols are also available to study the infectivity of the viral vector (Schleper *et al.*, 1992). During my studies the SSV1-based shuttle vector was further improved by adding a gateway recombination cassette (Invitrogen) allowing easier and faster genetic manipulation of the vector (Manica *et al.* submitted).

The genome sequences of *S. solfataricus* P2 and *S. solfataricus* 98/2 strains are deposited in the NCBI database. In addition, closely related strains P1 and PBL2025 are also used (and partially sequenced), allowing transformation and production of knock-out mutants in those species (She *et al.*, 2001; Worthington *et al.*, 2003; Albers and Driessen, 2008).

More recently, genome sequences and new transformation systems (using smaller genetic elements) as well as gene knock-out protocols were also developed in other *Sulfolobus* species widening the tool box and possibilities for molecular biological studies in Crenarchaeota (Allers and Mevarech, 2005; Berkner *et al.*, 2007; Reno *et al.*, 2009; Wagner *et al.*, 2009; Grogan, 2009; Guo *et al.*, 2011; Leigh *et al.*, 2011; Wagner *et al.*, 2012; Jaubert *et al.*, 2013; Rockwood *et al.*, 2013). Today, in particular *S. solfataricus* and *S. islandicus* are the best suited candidates to study viral-host interactions due to their sensitivity to several isolated viruses and the possibility to be transformed and infected by a viral-based shuttle vector.

During my PhD studies I used *S. solfataricus* and its virus SSV1 to investigate the activity and the underlying features of an archaeal CRISPR-Cas system *in vivo*.

4.4 Viruses and the environment

Viruses that infect and kill bacteria and archaea, are key players in food chains in all environments (Breitbart *et al.*, 2004). They can be found everywhere, from sea to deserts, from hypersaline environment to hot springs where they play a key role in controlling microbial populations. Virus particles outnumber microbial cells with an estimated ratio of 10 particles *per* cell in aquatic environments, placing them among the most abundant biological entities on Earth (Bergh *et al.*, 1989).

The release of nutrients from lysed microorganisms due to viral infection is one of the most important factors in the transfer of energy and nutrients to higher trophic levels. Extensive research conducted on the marine environment has demonstrated how viral presence significantly influences global carbon and nutrients cycles (Pomeroy, 1974; Azam, 1983; Kirchman,

1994). Viruses not only influence the carbon cycle by releasing nutrients in the environment, but also play a central role in controlling overall microbial populations, thus maintaining microbial diversity within the environment. There is a tight dynamic between phage and bacterial populations (or more general between viruses and microbial populations) which can be explained by the "killing the winning population" theory. This concept describes how phage populations grow by infecting the fastest growing bacterial population in a given ecological setting (Thingstad *et al.*, 1997). The phage epidemic ceases when the diminished host population no longer supports phage replication (Chibani-Chennoufi *et al.*, 2004). Several environmental studies as well as laboratory observations support this theory. In fact, it has been shown that virus production does not take place when host concentrations reach less than $1x10^4$ cells/ml and that starving cells are poorly infected. Moreover, infected stationary-phase cells release progeny virus only after the resumption of cell growth (Chibani-Chennoufi *et al.*, 2004). All these observations underline the natural predisposition of the virus community in controlling the faster growing microbial population. By targeting the winning population, viruses maintain the balance between different species, avoiding hyper-proliferation of single taxa and thus increasing the overall microbial diversity (Chibani-Chennoufi *et al.*, 2004).

The presence of a virus and its host in a given environment is the condition "*si ne qua non"* for infection. Despite the high viral density detected in most environments, the possibility for a particular virus to find a suitable host for infection is limited. In fact, each virus possesses a narrow host-range and is able to infect only cells in defined physiological states.

To maintain themselves in the bacterial population during unfavourable conditions bacteriophages have developed an important strategy: lysogeny. During lysogeny the phage integrates its DNA into the host chromosome or can form a circular replicon in the bacterial cytoplasm. The lysogenic form, called prophage, can then be transmitted to the daughter cells during subsequent cell division (Canchaya *et al.*, 2003). During lysogeny the viral replication is reduced to a minimum if not totally abolished (Canchaya *et al.*,

2003). Lysogeny will continue until the environmental conditions or external inducing factors (e.g. UV light) favour the reactivation of virus replication inducing, at its final stage, the release of new viral particles into the environment by host cell lysis. In contrast to replication during lytic infection cycles, the viral DNA is exposed to different selective pressures during lysogeny (Canchaya *et al.*, 2003). The prophage decreases the fitness of its host by at least two processes: the increased metabolic burden from replicating additional DNA and the lysis of the host (lysogen) after prophage induction (Canchaya *et al.*, 2003). To compensate this disadvantage the prophage often encodes particular functions that increase the fitness of the lysogen. Indeed, the presence of super-infection exclusion genes on prophages that protect the host from further infections is one of the most studied selective advantages for the lysogen (Canchaya *et al.*, 2003). Nevertheless, in environments where the selective pressure is not strong enough, other prophage genes need to increase the fitness of the lysogenic host to ensure prophage conservation (Canchaya *et al.*, 2003). Prophages (or proviruses) are not only found in bacterial but also in archaeal genomes. Beside bacterial-like lysogeny described for haloarchaea and few other archaea, a slightly different interaction has been described for most virus-host interactions in hyperthermophilic archaea: Here virus production is often not accompanied by cell lysis, instead buddying of virus particles from the host cells leads to (reversible) growth retardation (Martin *et al.*, 1984; Reiter *et al.*, 1987; Schleper *et al.*, 1992). Beside this form of "lysogeny", other less well understood interactions, often referred to as "carrier states", that are reminiscent of some virus-eukaryote relationships have been described (for reviews on archaeal-virus relationships see: Prangishvili *et al.*, 2001). The virus host interactions in hyperthermophiles might thus represent another strategy to maintain and vertically transmit phage genes in extreme environments thereby avoiding host lysis.

The transfer of genomic DNA between different microbial cells through the incorporation of host DNA sequences in the viral capsid during particle packaging is another important ecological aspect of virus-host interaction

(Chibani-Chennoufi *et al.*, 2004), referred to as specialized or general transduction. It was estimated that under laboratory conditions, incorporation of some host DNA occurs once every $1x10^8$ infections. Considering the total Earth viral population ($>1x10^{30}$ viral particles) viral infection makes a major contribution to the overall horizontal gene transfer that occurs between bacterial cells (Chibani-Chennoufi *et al.*, 2004).

Most of the viruses infecting prokaryotes use the host cell for reproduction and then kill the cell to release the newly formed viral particles or they impose severe growth retardations. This process, although of key importance for the environment, results in a high selective pressure on the host to develop resistance strategies in order to avoid viral invasion and replication. In fact, bacteria and archaea have developed a portfolio of mechanisms to avoid, block and prevent viral infection. These mechanisms range from blocking the injection of viral DNA/RNA, blocking the transcription of viral genes, to the avoidance of phage DNA replication and morphogenesis (Chibani-Chennoufi *et al.*, 2004).

One of the most studied mechanisms to avoid viral infection is the Restriction-Methylation system (Bingham and Atkinson, 1978). This defence mechanism uses restriction endonucleases to target the non-methylated viral DNA at particular DNA sites. The host DNA is protected from degradation by the presence of a methylase enzyme which adds methyl groups to the genomic DNA sequences which block the activity of the endonuclease enzyme (Bingham and Atkinson, 1978; Kessler and Holtke, 1986; Chibani-Chennoufi *et al.*, 2004).

Only recently a new defence mechanism was discovered in bacteria and archaea: the CRISPR-Cas system (Deveau *et al.*, 2008). The CRISPR-Cas defence mechanism is based on the incorporation of small viral DNA (or plasmid) sequence in a particular chromosomal location (called CRISPR) during viral challenge (Erdmann and Garrett, 2012; Yosef *et al.*, 2012). Once incorporated, this DNA fragment is transcribed and processed as small RNAs by a particular Cas protein (Brouns *et al.*, 2008; Hale *et al.*, 2008; Lintner *et al.*, 2011). Finally, the mature small RNA (crRNA) guides a protein complex

formed by other Cas proteins to the viral DNA target, blocking viral infection (Brouns *et al.*, 2008; Marraffini and Sontheimer, 2008; Manica *et al.*, 2011; Gudbergsdottir *et al.*, 2011; Semenova *et al.*, 2011)

4.5 The CRISPR system: brief introduction

In the 1980s, Ishino and colleagues identified a new peculiar type of short repeats in the genome of *E. coli* (Ishino *et al.*, 1987; Nakata *et al.*, 1989). Later, similar loci were detected in the genomes of many other bacteria and archaea (Hermans *et al.*, 1991; Groenen *et al.*, 1993; Mojica *et al.*, 1995; Bult *et al.*, 1996; Masepohl *et al.*, 1996; Klenk *et al.*, 1997; Sensen *et al.*, 1998; Kawarabayasi *et al.*, 1999; She *et al.*, 2001). It was not until the first years of the century that Mojica, Jansen and their co-workers defined this repeat region as CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) (Mojica *et al.*, 2000; Jansen *et al.*, 2002). In their experiments they not only characterized the CRISPR loci but also identified a set of proteins called "Cas" (CRISPR-Associated proteins), which were present only in genomes carrying CRISPR loci. Four of these proteins (Cas 1-4*)* were defined as core proteins due to their presence in all the genomes analysed at that time. Subsequently, further analysis of the *cas* genes were conducted on different genomes (Bolotin *et al.*, 2005; Pourcel *et al.*, 2005; Haft *et al.*, 2005; Godde and Bickerton, 2006; Lillestol *et al.*, 2006; Makarova *et al.*, 2006; Kunin *et al.*, 2007; Grissa *et al.*, 2007) adding the proteins Cas5 and Cas6 to the already existing core set. This allowed a more detailed classification of the Cas proteins into different families: As the CRISPR-Cas system was detected and analysed in many additional organisms, Makarova and co-workers proposed a new evolution-based classification of the CRISPR system in which Cas1 and Cas2 are considered as the only "true" ubiquitous core proteins (Makarova *et al.*, 2011b) (see below).

The CRISPR locus is constituted by 3 distinct units: the repeats, the spacers and the leader sequence. Repeats are usually 21 to 47 nt long (Godde and Bickerton, 2006; Karginov and Hannon, 2010) and all repeats in the same CRISPR locus have the same sequence and orientation which eliminates the

possibility of base pairing between repeats. Nevertheless several CRISPR arrays carry repeats with internal partial palindromic sequences that could form a stable stem-loop structure upon transcription (Kunin *et al.*, 2007). However, the ability of the repeats to form stable mRNA secondary structures does not seem to be a characteristic of all CRISPR loci (Makarova *et al.*, 2006; Kunin *et al.*, 2007). A particular repeat is often associated with a set of core Cas proteins. When a genome carries only a set of Cas genes but multiple CRISPR loci, the repeat sequences in the different CRISPR loci are usually the same. However, when different CRISPR loci are present in the same genome carrying different repeat types, multiple sets of Cas core genes are also present (Lillestol *et al.*, 2006).

Between two direct repeats a short unique DNA sequence is situated. This sequence is defined as spacer and is usually between 20 and 72nt long (Grissa *et al.*, 2007; Karginov and Hannon, 2010). The number of spacers located within a CRISPR locus is very variable. The average CRISPR length is of 20 spacer-repeat units but it is not uncommon to identify CRISPR with more than 100 spacer-repeat units (Lillestol *et al.*, 2009; Shah and Garrett, 2011).

When different CRISPR loci (with the same repeat type) are located in a genome, they share another important sequence region: the leader sequence. Leader sequences are DNA sequences of hundreds of bases that are always located immediately upstream of a CRISPR repeat cluster. They lack coding potential (Karginov and Hannon, 2010) and they often share a conserved motif in their sequences, which could represent putative transcriptional regulatory sites of the CRISPR locus (Lillestol *et al.*, 2009). The function of the leader sequence is not yet well understood. In those organisms in which it was studied, CRISPR transcription initiated between 60 bp (*E. coli*) and 25 bp (*S. acidocaldarius*) or immediately upstream of the first CRISPR repeat (Lillestol *et al.*, 2009; Pougach *et al.*, 2010; Pul *et al.*, 2010). In *E. coli* and *S. acidocaldarius* a promoter-like region within the leader sequence was found located 25 nt upstream from the transcription start site (Lillestol *et al.*, 2006; Lillestol *et al.*, 2009; Pougach *et al.*, 2010). Several conserved motifs were identified within the leader sequence of *S. acidocaldarius, S. islandicus and*

S. solfataricus (Lillestol *et al.*, 2006) but their function is still unknown. Furthermore, in their recent work, Yosef and co-workers demonstrated that the presence of a 20 nt sequence, located between position -60 and the first repeat in *E. coli,* is fundamental for the acquisition of new spacers (Yosef *et al.*, 2012).

4.6 CRISPR associated proteins (Cas)

The CRISPR-Cas system is present in more than 90% of archaeal and 50% of bacterial genomes and is very diverse. Several classifications of the system were proposed, either based on the type of repeat (Kunin *et al.*, 2007), *cas* genes located near the locus (Haft *et al.*, 2005) or based on a multidisciplinary approach which integrated phylogeny, sequence type, locus organization and content (Makarova *et al.*, 2011b).

The first classification based on the sequence of the Cas proteins has divided the different CRISPR systems into more than 45 different families (Haft *et al.*, 2005; Makarova *et al.*, 2006), which are now reduced to three major CRISPR types (Type I, Type II and Type III) (Makarova *et al.*, 2011a; Makarova *et al.*, 2011b).

The Type I CRISPR is found in Bacteria as well as Archaea and is the most diverse with six different subtypes. Its principal characteristic is the presence of the Cas3 protein, a helicase-nuclease protein involved in the targeting and degradation of foreign DNA (Makarova *et al.*, 2011a). The system was intensively studied in *E. coli* and in *P. aeruginosa* (Brouns *et al.*, 2008; Haurwitz *et al.,* 2010; Jore *et al.*, 2011; Wiedenheft *et al.*, 2011). In those systems it was demonstrated that a long precursor CRISPR RNA (pre-crRNA) is processed to crRNA by a particular Cas gene, Cas6. Cas6 encodes an endonuclease which recognizes the CRISPR repeat and cleaves 8´nt within the repeat. Once formed, the crRNA guides a complex composed of several Cas proteins called CASCADE (CRISPR associated complex for antiviral defence) to the target DNA and blocks the invasion of genetic elements carrying complementary DNA sequences (protospacer) to the crRNA (Brouns *et al.*, 2008).

For efficient DNA targeting in the CRISPR Type I, the pairing between crRNA and its target as well as the presence of a PAM site seems to be a necessary condition (Mojica *et al.*, 2005; Semenova *et al.*, 2011; Gudbergsdottir *et al.*, 2011).

Furthermore *in vivo* and *in vitro* studies have demonstrated the importance of the first 8 nt at the 5´end of the crRNA ('seed' sequence) during the recognition process (Semenova *et al.*, 2011; Wiedenheft *et al.*, 2011). Mutations in the seed sequence are not tolerated and lead to the loss of immunity.

Type II CRISPR activity was demonstrated *in vivo* in *Streptococcus thermophilus* (Barrangou *et al.*, 2007) and was involved in the protection of the cell from viral and plasmid DNA that contained a protospacer that matched perfectly to a CRISPR encoded crRNA. A single mutation in the protospacer region as well as within the PAM sequence abolishes the interference in this organism (Barrangou *et al.*, 2007; Garneau *et al.*, 2010; Deltcheva *et al.*, 2011). In this CRISPR type the CRISPR pre-mRNA is not processed by Cas6 but by an endogenous RNase III enzyme whose activity is mediated by an RNA duplex formation between repeat sequence, tracrRNA and Cas9 protein (Deltcheva *et al.*, 2011). The Cas protein Cas9 is the Type II CRISPR signature (Makarova *et al.*, 2011a), and is a large multifunctional protein with the ability to generate crRNA as well as to target phage and plasmid DNA for degradation (Garneau *et al.*, 2010).

Type III CRISPR was extensively studied in *Staphylococcus epidermis* (CRISPR type IIIB) (Marraffini and Sontheimer, 2008), *Pyrococcus furiosus* (Hale *et al.*, 2009) and recently in *Sulfolobus solfataricus* (Zhang *et al.*, 2012) (CRISPR Type IIIA) . This system is characterized by the presence of a Cas10 protein which is possibly involved in crRNA processing and DNA targeting. Furthermore it (often) contains a Cas6 protein which is important for crRNA maturation. Interestingly, the CRISPR type IIIB (also known as Csm module) system targets DNA *in vivo* but not RNA (Marraffini and Sontheimer, 2008), whereas the type IIIA (Cmr module) system cleaves RNA *in vitro*, (Hale *et al.*,

2009; Zhang *et al.*, 2012). Both type IIIA and IIIB have shown no need for a PAM sequence to cleave RNA or DNA.

4.7 Activity of the CRISPR locus

Since its discovery in the 80's several hypotheses were postulated on the function of CRISPR. In 1995 Mojica and co-workers observed that an increase in the number of repeats in *Haloferax volcanii* resulted in altered replicon segregation, and suggested that the repeats could be involved in replicon partitioning (Mojica *et al.*, 1995; Sorek *et al.*, 2008). Later, due to the presence of multiple CRISPR loci in many organisms, Jansen and co-workers suggested the possibility that CRISPRs were a new type of mobile genetic elements (Jansen *et al.*, 2002). A few years later Makarova and co-workers postulated a DNA repair mechanism for the CRISPR locus due to the presence of many DNA manipulating domains in several Cas proteins (Makarova *et al.*, 2002; Sorek *et al.*, 2008) . It was only in 2005, when different research groups discovered that several spacers in the CRISPR loci possessed sequences with high similarity to invading genetic elements (phages and/or plasmids), that the CRISPR-Cas system was defined as the "new immune system of Archaea and Bacteria" (Bolotin *et al.*, 2005; Mojica *et al.*, 2005; Pourcel *et al.*, 2005). This observation was supported by the correlation between spacer presence and resistance to particular phages. Indeed, strains harbouring a CRISPR locus with a spacer that matched a particular virus were immune to the corresponding foreign invader (Brouns *et al.*, 2008; Bolotin *et al.*, 2005; Barrangou *et al.*, 2007; Karginov and Hannon, 2010).

These results demonstrated that the CRISPR-Cas is a new defence mechanism against invading genetic elements. In this system the Cas proteins are the enzymatic machinery and the spacers, that can be rapidly acquired from invading genetic elements, are conferring the system specificity (Karginov and Hannon, 2010).

With this system, new spacers need to be acquired after infection to block the genetic element proliferation. Acquiring new spacer units in the CRISPR

system not only confers protection against extra-chromosomal DNAs but also constitutes at the same time a "memory" of past infections. CRISPR loci evolve rapidly and the addition of new spacers was used for the classification of closely related strains (Groenen *et al.*, 1993; Kamerbeek *et al.*, 1997; Jansen *et al.*, 2002). Comparison of CRISPR sequences between closely related strains indicate that the addition of new spacers take place at the leader sequence side of the CRISPR locus. The distal part of the cluster contains "older" spacers that are often shared between closely related strains (Pourcel *et al.*, 2005; Lillestol *et al.*, 2009; Horvath and Barrangou 2010).

4.8 Acquisition of new CRISPR spacers

The first *in vivo* evidence that demonstrated the ability of the CRISPR system to prevent viral infection was reported by Barrangou and colleagues (Barrangou *et al.*, 2007). They were able to isolate phage-insensitive *S. thermophilus* mutants after challenging the strain with a bacteriophage.
Remarkably all the mutants showed the incorporation of new spacers within the CRISPR locus which matched perfectly with the genomic sequence of the invading bacteriophage (called protospacer). Furthermore the host resistance could be revoked in genetically-engineered mutants lacking the spacer.
All the spacers were found integrated at the leader proximal side of the CRISPR locus as earlier indicated by bioinformatic analysis (Mojica *et al.*, 2005; Lillestol *et al.*, 2006; Lillestol *et al.*, 2009). Knock out experiments in *S. thermophilus* demonstrated the importance of one of the Cas genes, *csn2*, whose inactivation lead to the inability of the system to incorporate new viral sequences into the CRISPR locus (Barrangou *et al.*, 2007). The insertion of new spacers takes place after a viral challenge, but the scientific community is still debating how the CRISPR sequence recognizes the extra-chromosomal DNA. What is clear now is that the incorporation of new spacers into the CRISPR locus is not random but driven by short sequences located at the 5'end of the DNA sequence that will be incorporated in the CRISPR locus. These short motifs, usually 2-3 nt long, are defined as PAM (protospacer adjacent motif). They indicate not only the sequences that will be

incorporated but also the orientation of the spacer that will be integrated into the CRISPR locus (Mojica *et al.*, 2005).

Interestingly PAM sequences seem to depend on the locus type: When different CRISPR loci with different repeat types are located within the same chromosome they have a preference for different PAMs. In fact, specific PAM sequences for the different CRISPR families were found in *S. solfataricus* when the different CRISPR-Cas loci of this organism were analysed (Lillestol *et al.*, 2009).

When an organism carries CRISPRs which possess different sets of core proteins (Cas1-Cas2), each CRISPR-Cas locus must be considered as a single defence entity with its own peculiarities, at least with respect to the integration of new spacers. Although the interaction between different Cas proteins of various Cas subtypes cannot be excluded, it seems likely that subtype-specific Cas proteins are involved in the incorporation of new spacers during the adaptation stage (Makarova *et al.*, 2006; Barrangou *et al.*, 2007; Yosef *et al.*, 2012).

Although the inactivation of the Cas gene *csn2* in *S. thermophilus* results in the inability to incorporate new spacers, another candidate Cas protein, Cas1, seems to have a central role in the acquisition of new spacers (Yosef *et al.*, 2012). This Cas core protein has a putative DNA nuclease/integrase domain (Makarova *et al.*, 2002; Makarova *et al.*, 2006) and is part of any active CRISPR locus. Its activity *in vitro* was recently validated in *P. aeruginosa, S. solfataricus* and *E. coli.* In *P. aeruginosa*, Cas1 is a metal-dependent DNase that is able to digest ssDNA and dsDNA yielding fragments of ~80 bp. Interestingly, the DNA cleavage requires Mn^{2+} or Mg^{2+}, while ssDNA cleavage is supported only by Mn^{2+}. As tested so far, the DNase activity is not sequence-specific or methylation-specific (Wiedenheft *et al.*, 2009). However, no nuclease activity was found for Cas1 protein SSO1450 from *S. solfataricus* in the presence of Ca^{2+} and Mg^{2+} but the protein retained the ability of binding ss/dsDNAs and ss/dsRNAs (Han *et al.*, 2009). A recent study has characterized the biochemical activity of the *E. coli* Cas1 protein *YgbT* (Babu *et al.*, 2011). This enzyme, as for the previously described Cas1 of

P. aeruginosa, has a prominent divalent metal cation-dependent nuclease activity against single stranded DNA and a lower activity on ssRNAs. The *YbgT* protein also cleaves short linear (34 nt) double stranded DNA fragments but no cleavage activity was detected with dsRNAs (16-39 nt) or longer dsDNAs (60 nt) (Babu *et al.*, 2011).

Moreover it was discovered that the *YbgT* protein has cleavage activity on cruciform like structures and CF-structures and solves holiday junction (HJs). HJs are formed in all organisms during DNA integration, recombination and recombinatorial DNA repair. CF-like structures can instead be formed on the repeat DNA sequence due to the palindromic sequence of the repeats (Babu *et al.*, 2011). Together, these results make Cas1 the most-likely candidate involved in the incorporation of new spacers, although the detailed mechanism of spacer incorporation remains unknown.

Recently, Yosef and co-workers have shown the incorporation of new spacers in *E. coli* cells by over-expressing only Cas1 and Cas2 genes (Yosef *et al.*, 2012). Their experiments demonstrated that the expression of Cas1 and Cas2 is sufficient for the incorporation of new spacers and no other Cas proteins appeared to be involved in the process (at least in *E. coli*). Furthermore, they confirmed the presence of the PAM sequence AWG as previously detected by bioinformatic analysis for this organism. The authors also addressed the importance of the leader sequence in the incorporation of new spacers. In fact, the absence of the leader nucleotide sequence between the putative promoter (position -60 nt) and the first repeat is necessary for spacer incorporation (Yosef *et al.*, 2012).

Interestingly, incorporation of plasmid sequence as a spacer in the CRISPR locus was preferred with respect to genomic DNA although the interference process was abolished in this strain due to the lack of all other Cas proteins. This observation opens new interesting possibilities where additional mechanisms could be involved in the specific recognition of foreign genetic elements (Yosef *et al.*, 2012).

4.9 CRISPR expression

The CRISPR clusters are transcribed as a multiunit precursor that is subsequently cleaved into smaller units called crRNA. The crRNA consists of one spacer flanked by two partial repeats (Tang *et al.*, 2002; Brouns *et al.*, 2008; Hale *et al.*, 2008). Each active CRISPR locus is preceded by a conserved promoter sequence called a leader. From a promoter located within the leader the CRISPR locus is unidirectionally transcribed (Pougach *et al.*, 2010), whereas evidence of bi-directional CRISPR transcription was detected in some Sulfolobales species (Lillestol *et al.*, 2009). In *E. coli* the most upstream region of the leader sequence is similar to the extended -10 promoter element consensus sequence (TGxTATAAT). Recently it was shown by Pul *et al.* (2010) that this region functions as a s70 RNA polymerase promoter (Pul *et al.*, 2010). Indeed, Pougach demonstrated that the two *E. coli* CRISPR cassettes are transcribed by a weak extended -10 class s70 promoter located within the leader regions (Pougach *et al.*, 2010). In *E. coli*, leader promoters appear to be weak, although with differences in expression between the two CRISPR promoters. Furthermore, as the CRISPR transcript is not translated, it is likely that a special anti-termination mechanism exists to ensure efficient transcription of full length CRISPR mRNAs (Pougach *et al.*, 2010).

In *E. coli* the expression of the CRISPR arrays and Cas genes is tightly regulated. During normal growth the system seems to be silenced by H-NS repression which regulates the transcription of CRISPR-Cas promoters (Pul *et al.*, 2010). The system repression can be relieved by the action of the transcriptional activator LeuO (Westra *et al.*, 2010).

The *hns* (histone-like protein) gene in *E. coli* acts as negative regulator of CRISPR expression. In Δhns mutants the abundance of processed CRISPR transcripts was comparable to the expression of *E. coli* cells overproducing *CasE* (Pougach *et al.*, 2010; Pul *et al.*, 2010). Moreover, Δ*hns* mutants are resistant to phages carrying protospacers matching a chromosomal CRISPR spacer, unlike the wild-type cells.

Over-expression of the enzyme CasE (Cas6) also leads to a significant increase in the abundance of processed CRISPR I transcripts (crRNA),

despite the very low steady-state levels of full sized CRISPR transcripts observed in wild-type cells (Westra *et al.*, 2010). Such increases in crRNA abundance is related to the increased stability of the crRNAs in the presence of CasE, which binds to the crRNAs and protects them from degradation by unknown cellular RNAses (Pougach *et al.*, 2010).

A feedback loop regulation is likely to be active on the CRISPR locus assuring the presence of crRNAs in the cell and avoiding the high energy cost of full length CRISPR RNA transcription in the absence of infection. No influence on expression level of the different crRNAs was detected based on their location within the CRISPR arrays. Due to this reason the "older" spacers, located more distally from the leader sequence in the *E. coli* CRISPR locus, and might be as functional as the leader proximal spacers. However, it is possible that a gradient in abundance of processed transcripts exists in organisms with long CRISPR arrays (Pougach *et al.*, 2010). In *S. solfataricus* high expression of leader distal spacers were detected by Northern blot analysis probably due to the presence of a promoter-like sequence lying within spacers that stimulates the transcription of the locus at internal sites. Another CRISPR regulator was found in *S. solfataricus* (Deng *et al.*, 2011). In this archaeon the protein Cbp1 binds the repeat DNA of the different CRISPR arrays increasing CRISPR pre-crRNA expression. In microarray experiments, a marked increase in CRISPR pre-crRNA levels was detected after over-expression of Cbp1 protein. In support of this result a strong decrease in CRISPR pre-crRNA was detected in Δ*Cbp1* knock-out mutants. Interestingly, Deng and co-workers also suggested the possibility that the Cbp1 protein reduces the expression of aberrant transcripts generated by induced internal spacer promoters by causing their premature termination (Deng *et al.*, 2011).

Regulation of the CRISPR-Cas system after viral challenge was studied in detail by Agari and co-workers in the thermophilic bacterium *Thermus thermophilus* strain HB8 (Agari *et al.*, 2009). When analysing the reaction of its CRISPR-Cas system by microarray analysis, they found that most of the Cas genes were significantly up-regulated when phage progeny began to be

produced (Agari *et al*., 2009). Interestingly, the expression of the system was not simply regulated by common regulatory factors such as cAMP. In fact, whereas the induction of several Cas genes decreased in ΔCRP (cAMP receptor protein) knock-out mutants, other Cas genes were not affected by the absence of CRP, demonstrating the presence of additional regulation factors.

.

4.10 Processing of CRISPR RNA

Once the CRISPR locus is transcribed, the pre-crRNA is processed by particular Cas proteins into small RNA (crRNA). Different CRISPR types use different strategies for the CRISPR mRNA processing. Type I and Type III CRISPR use a particular Cas protein called Cas6 to process the RNA within the repeats. CRISPR type II uses cellular RNAse III enzymes and a short CRISPR encoded tracrRNA to guide specific CRISPR RNA processing within the repeats (Hale *et al*., 2008; Brouns *et al*., 2008; Carte *et al*., 2008; Makarova *et al*., 2011a; Makarova *et al*., 2011b; Deltcheva *et al*., 2011; Lintner *et al*., 2011).

CRISPR type I pre-crRNA processing was intensively studied in several organisms. In *E. coli* a palindromic sequence within the direct repeat produces a stem loop which is recognized and cleaved by CasE (Cas6) (Brouns *et al*., 2008). The resulting mature crRNAs carry an 8nt 5' tag, a full spacer sequence followed by a 3' handle of variable length which corresponds to the successive repeat. Also in *S. solfataricus* and *P. furiosus* the protein Cas6 is responsible for binding and cleaving the direct repeat (Carte *et al*., 2008). In these two hyperthermophilic archaea the CRISPR repeats lack internal palindromic sequences. The absence of the formation of a stem loop is consistent with the observation made by Wang and colleagues (Wang *et al*., 2012). Analysing the crystal structure of *P. furiosus* Cas6 they predicted its interaction with single stranded RNA.

Furthermore, *in vitro* activity studies on Cas6 genes in *P. furiosus* and *S. solfataricus* have shown the ability of these enzymes to cleave the direct repeat at particular internal sites (Carte *et al*., 2008, Lintner *et al*., 2011).

Furthermore, *S. solfataricus* Cas6 (sso2004) shows metal-independent ribonuclease activity and cleaves the pre-crRNA within the repeat (AGGA/AUUG) at a single site (Lintner *et al.*, 2011).

CRISPR type II systems lack the Cas6 endonuclease protein. In the absence of Cas6 the organisms which carry the CRISPR type II system use an ingenious mechanism to cleave their direct repeats which involves a trans-encoded ncRNA (tracrRNA), host encoded RNAse III enzymes and a particular Cas protein, csn1 (Deltcheva *et al.*, 2011). In *S. pyogenes* where the system was studied, tracrRNA overcomes the deficiency of repeat stem-loop structures (Deltcheva *et al.*, 2011). The binding of the tracrRNA to the direct repeat determines the intramolecular RNA structure which is necessary for pre-crRNA processing (Deltcheva *et al.*, 2011).

Mature co-purified crRNAs from the interference complex (CASCADE) in *E. coli*, *S. solfataricus* and *P. aeruginosa* have all shown the same peculiarities. They can be divided in three functionally distinct parts: an 8nt 5' tag, a full length spacer, and a 3' handle composed of the remaining part of the successive repeat. Presumably, the 3' handle provides the sequence for charging the crRNA into the CASCADE complex, the spacer sequence guides the interference complex to the target DNA and the 8nt 5' tag prevents the complex from targeting its own DNA (Brouns *et al.*, 2008; Wiedenheft *et al.*, 2011; Lintner *et al.*, 2011).

It is interesting that in CRISPR type IIIA, the mature crRNA co-purified from the Cmr complex (Hale *et al.*, 2009; Zhang *et al.*, 2012) has apparently undergone further processing at its 3' end. In *S. solfataricus* co-precipitated crRNA shows a very short or completely absent 3' handle which would favour the processing of the 3' end of the crRNA from an unknown exonuclease. In *P. furiosus* the mature crRNA found in the Cmr complex was instead found in a distinct size range of 39 and 45nt (31nt/37+8nt 5'tag). This crRNA favours an enzymatic ruler-like mechanism which is involved in the trimming of the mature crRNAs.

4.11 CRISPR and DNA/RNA interference

CRISPR pre-crRNA is expressed and successively processed into mature crRNA. The crRNA guides the interference complex (e.g. CASCADE, Cmr etc.) formed by different Cas proteins to the target nucleic acids. Analysing the orientation of the spacers in the different CRISPR loci and their direction with respect to their putative target sequence, Mojica and co-workers speculated that the target of the system should be DNA and not RNA as previously hypothesized (Mojica *et al.*, 2005; Makarova *et al.*, 2006; Mojica *et al.*, 2009). Within the last few years, several groups have demonstrated that the primary target of most CRISPR systems is indeed DNA and not RNA. The first evidence rose from the work of Brouns and co-workers (Brouns *et al.*, 2008) where spacers of both orientations against the phage lambda lead to a reduction of sensitivity to phage infection. To exclude any possibility that CRISPR would act also on RNAs, Maraffini and Sontheimer performed an experiment with a self-splicing intron demonstrating that the crRNA-Cas complex is not able to target mRNA in *S. epidermidis* (CRISPR type IIIB) (Marraffini and Sontheimer, 2008). CRISPR activity leading to degradation of extrachromosomal DNA has now been demonstrated in several different organisms: *S. thermophilus, S. epidermis, E. coli*, *S. solfataricus and P. aeruginosa* (Barrangou *et al.*, 2007; Marraffini and Sontheimer, 2008; Brouns *et al.*, 2008; Manica *et al.*, 2011; Gudbergsdottir *et al.*, 2011; Semenova *et al.*, 2011). Although in many organisms DNA interference was demonstrated *in vivo*, in two hyperthermophilic archaea, *in vitro* studies have demonstrated the ability of a peculiar Cas interference complex (Cmr) to target RNA (CRISPR type IIIA) (Hale *et al.*, 2009; Zhang *et al.*, 2012).

4.11.1 Target recognition

Experimental data have shown how all the CRISPR systems so far investigated (at least those targeting DNA) include the CRISPR pre-crRNA processing enzyme (CasE, Cas6 or Cys4) associated with the CASCADE complex (Brouns *et al.*, 2008; Lintner *et al.*, 2011; Wiedenheft *et al.*, 2011). This fact indicates that once the crRNA is formed it might be directly charged

into the complex. Once processed and incorporated into the complex, the crRNA guides the CASCADE to the target DNA. The crRNA is essential to identify and target foreign nucleic acids in all CRISPR systems (Wiedenheft *et al.*, 2011). Wiedenheft and co-workers tested the target recognition properties of the *P. aeruginosa* CASCADE complex (cys complex) performing electrophoretic mobility shift assays. Their experiment demonstrated that the binding affinity of crRNA to the target DNA was greatly enhanced in the presence of the Cas proteins. Indeed, the cys complex had high affinity in binding the target ssDNA although no binding affinity was found when crRNA and cys complex were incubated with ssDNA carrying sequences with no similarity with the crRNA. Identical results were obtained by Lintner and colleagues using the *S. solfataricus* CASCADE complex (Lintner *et al.*, 2011). Structural and biochemical studies have shown that argonaute proteins facilitate target recognition by preordening the first 8 nt of the guide in an elican configuration (Parker *et al.*, 2009). To test if the target recognition by the Cys complex acts in a similar manner, experiments were conducted incubating the cys complex with short 8 nt oligonucleotides spanning the entire spacer sequence. Indeed, this test demonstrated that the first 8 nt of the spacer (seed sequence) are also of fundamental importance in the target recognition for the CASCADE complex. Oligos complementary to regions outside the seed sequence do not bind with measurable affinity (Wiedenheft *et al.*, 2011) and the same applies for the nucleotides located 3´ of the seed sequence, which have no detectable binding affinity.

The CRISPR/Cas system finds the target through seed interaction between the crRNA-Cas complex and the foreign DNA. In *S. thermophilus* a single mutation in the protospacer or in the protospacer adjacent motif (PAM) leads to virus escape. This observation corroborated the hypothesis that the PAM sequence was not only important for protospacer acquisition and incorporation but also during the interference process (Barrangou *et al.*, 2007).

Similar results were obtained in *E. coli* by Semenova and co-workers who identified several viruses that have the ability to escape the CRISPR/Cas

system carrying a single mutation in the PAM sequence or within the seed sequence, after challenging the *E. coli* CRISPR system with M13 phages (Semenova *et al*., 2011). The PAM function in the interference process remains puzzling: on one side the data from *S. thermophilus, E.coli* and *S. islandicus* reinforce the idea of a contribution of the PAM sequence in DNA interference (Barrangou *et al*., 2007; Gudbergsdottir *et al*., 2011; Semenova *et al*., 2011), while on the other hand, data from *S. epidermis* and *E. coli* (overexpressing the CASCADE genes) show no function of the PAM sequence during the interference process (Marraffini and Sontheimer, 2010). Interestingly, Semenova and co-workers observed a reduction in the binding affinity of the cys4-crRNA complex with the target DNA in the absence of the PAM sequence, whereas in *S. solfataricus* the binding activity of the CASCADE-crRNA complex to the target RNA was not influenced by the presence or absence of the PAM sequence (Brouns *et al*., 2008; Semenova *et al*., 2011; Lintner *et al*., 2011).

Another important aspect of crRNA target recognition is the degree of similarity between crRNA and DNA target which allows efficient interference. As described previously, in *S. thermophilus*, a single mutation in the spacer region is sufficient to abolish the interference. Other studies conducted on different systems have shown that recognition and interference can take place even in the presence of a high number of mutations between crRNA and the protospacer sequence (Semenova *et al*., 2011; Manica *et al*., 2011; Gudbergsdottir *et al*., 2011). This phenomenon was first recognized in archaea where the *S. solfataricus* CRISPR system was able to trigger interference also in the presence of several mismatches between crRNA and its protospacer (Gudbergsdottir *et al*., 2011; Manica *et al*., 2011). Subsequently, the same observation was made in *E. coli* were mutations allocated at the 5´ end of the crRNA were "silent" and not interfering with the CASCADE activity (Semenova *et al*., 2011). Remarkably, the presence of perfectly matching sequences between crRNA and the CRISPR locus DNA does not lead to any DNA degradation that would result in chromosomal cleavage and be highly deleterious for cell fitness.

4.11.2 Self non-self Target Discrimination

To understand how the CRISPR locus is protected from degradation by the self-originated spacers, Maraffini and co-workers performed a tedious mutagenic-based assay where the spacer flanking sequence (repeat) was mutated and replaced with a different nucleotide sequence (Marraffini and Sontheimer, 2010). Using this approach they were able to determine that the CRISPR interference activity is abolished when homology was shared between an 8 nt tag located 5' to the crRNA (which stems from the repeat sequence) and the protospacer adjacent sequences. Within the 8 nt 5' tag, complete identity with the protospacer adjacent sequence was not required to block interference. A minimum similarity of three nucleotides at preferential sites was sufficient to discriminate between self and non-self DNA, at least in *S. epidermidis*. No other experimental data is yet available for identifying self/non-self discrimination strategies in other bacteria or archaea (but see this work).

4.12 CRISPR mediated RNA interference

It has been demonstrated in the archaea *P. furiosus* and *S. solfataricus* that a particular CRISPR interference complex called Cmr (PfuCMR and ssoCMR, respectively) has the ability to cleave target RNAs *in vitro* once loaded with a mature crRNA (Hale *et al*., 2009; Zhang *et al*., 2012). Interestingly, the crRNAs co-precipitated from these two Cmr complexes differ from those isolated from the CASCADE complex by the presence of a shorter (or absent) 3' handle. In *P. furiosus* the 3' end processing of the repeat involves an exonuclease which produces two distinct crRNAs with sizes of 45 and 37 nt, respectively (Hale *et al*., 2009). In *S. solfataricus,* however, only a single species of crRNA was co-purified with the Cmr complex which lacks completely (or almost completely) the 3' handle (Zhang *et al*., 2012). No further processing of the 8 nt 5' tag has been observed in these organisms, demonstrating that this tag is a common trait of every crRNA, independent of the interference complex into which it is loaded.

Single stranded RNA cleavage activity of these two Cmr was demonstrated *in vitro* using reconstituted complexes, but no activity against dsRNA, ssDNA or dsDNA was detected (Hale *et al*., 2009; Zhang *et al*., 2012). In *P. furiosus,* targeted cleavage takes place 14 nt from the 3' end of the crRNA, whereas in *S. solfataricus* the cleavage takes place in the presence of a UA di-nucleotide. Due to this reason, the absence of UA sites in the target sequence prevents the Cmr RNA cleavage. Furthermore, RNA cleavage by PfuCMR and ssoCMR does not require any PAM sequence in the target mRNA. In contrast to the CASCADE complex, PfuCMR activity is not affected by the presence of complementarity between the target sequence and the 8nt 5' tag indicating that the complex lacks self/non-self discrimination (Hale *et al*., 2009; Zhang *et al*., 2012). The presence of the 8nt 5` tag in the guide crRNA is essential for the activity of the complex, and cannot be substituted with any other nucleotide sequence (Hale *et al*., 2009; Zhang *et al*., 2012). Interestingly, target recognition by the PfuCMR-crRNA complex is primarily affected by nucleotide pairing at the crRNA 3' end, and not by interaction within the seed sequence (5´ end) as found for the CASCADE-crRNA complex (Hale *et al*., 2009; Wiedenheft *et al*., 2011; Semenova *et al*., 2011).

4.13 References

Agari, Y., K. Sakamoto, M. Tamakoshi, T. Oshima, S. Kuramitsu & A. Shinkai, (2009) Transcription Profile of Thermus thermophilus CRISPR Systems after Phage Infection. *J Mol Biol*.

Albers, S. V. & A. J. Driessen, (2008) Conditions for gene disruption by homologous recombination of exogenous DNA into the Sulfolobus solfataricus genome. *Archaea* **2**: 145-149.

Albers, S. V., M. Jonuscheit, S. Dinkelaker, T. Urich, A. Kletzin, R. Tampe, A. J. Driessen & C. Schleper, (2006) Production of recombinant and tagged proteins in the hyperthermophilic archaeon Sulfolobus solfataricus. *Appl Environ Microbiol* **72**: 102-111.

Allers, T. & M. Mevarech, (2005) Archaeal genetics - the third way. *Nat Rev Genet* **6**: 58-73.

Azam, F., T. Fenchel, J. G. Field, J. S. Gray, L. A. Meyer-Reil, and F. Thingstad, (1983) The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser* **10:257-263**.

Babu, M., N. Beloglazova, R. Flick, C. Graham, T. Skarina, B. Nocek, A. Gagarinova, O. Pogoutse, G. Brown, A. Binkowski, S. Phanse, A. Joachimiak, E. V. Koonin, A. Savchenko, A. Emili, J. Greenblatt, A. M. Edwards & A. F. Yakunin, (2011) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol Microbiol* **79**: 484-502.

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

Bell, S. D. & S. P. Jackson, (1998) Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol* **6**: 222-228.

Bergh, O., K. Y. Borsheim, G. Bratbak & M. Heldal, (1989) High abundance of viruses found in aquatic environments. *Nature* **340**: 467-468.

Berkner, S., D. Grogan, S. V. Albers & G. Lipps, (2007) Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for Sulfolobus acidocaldarius and Sulfolobus solfataricus, model organisms of the (cren-)archaea. *Nucleic Acids Res* **35**: e88.

Bingham, A. H. & T. Atkinson, (1978) Restriction endonucleases and modification methylases in bacteria. *Biochem Soc Trans* **6**: 315-324.

Bolotin, A., B. Quinquis, A. Sorokin & S. D. Ehrlich, (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551-2561.

Boudsocq, F., S. Iwai, F. Hanaoka & R. Woodgate, (2001) Sulfolobus solfataricus P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic poleta. *Nucleic Acids Res* **29**: 4607-4616.

Breitbart, M., L. Wegley, S. Leeds, T. Schoenfeld & F. Rohwer, (2004) Phage community dynamics in hot springs. *Appl Environ Microbiol* **70**: 1633-1640.

Brochier-Armanet, C., B. Boussau, S. Gribaldo & P. Forterre, (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6**: 245-252.

Brouns, S. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin & J. van der Oost, (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.

Brown, J. R. & W. F. Doolittle, (1997) Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev* **61**: 456-502.

Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. Geoghagen & J. C. Venter, (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. *Science* **273**: 1058-1073.

Canchaya, C., C. Proux, G. Fournous, A. Bruttin & H. Brussow, (2003) Prophage genomics. *Microbiol Mol Biol Rev* **67**: 238-276, table of contents.

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

Chibani-Chennoufi, S., A. Bruttin, M. L. Dillmann & H. Brussow, (2004) Phage-host interaction: an ecological perspective. *J Bacteriol* **186**: 3677-3686.

Deltcheva, E., K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel & E. Charpentier, (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602-607.

Deng, L., C. S. Kenchappa, X. Peng, Q. She & R. A. Garrett, (2011) Modulation of CRISPR locus transcription by the repeat-binding protein Cbp1 in Sulfolobus. *Nucleic Acids Res* **40**: 2470-2480.

Deveau, H., R. Barrangou, J. E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath & S. Moineau, (2008) Phage response to CRISPR-

encoded resistance in Streptococcus thermophilus. *J Bacteriol* **190**: 1390-1400.

Erdmann, S. & R. A. Garrett, (2012) Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms. *Mol Microbiol* **85**: 1044-1056.

Frols, S., M. Ajon, M. Wagner, D. Teichmann, B. Zolghadr, M. Folea, E. J. Boekema, A. J. Driessen, C. Schleper & S. V. Albers, (2008) UV-inducible cellular aggregation of the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by pili formation. *Mol Microbiol* **70**: 938-952.

Garneau, J. E., M. E. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadan & S. Moineau, (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67-71.

Garrett, R. A. & H.-P. Klenk, (2007) *Archaea : evolution, physiology, and molecular biology,* p. xii, 388 p., 328 p. of plates. Blackwell Pub., Malden, MA.

Godde, J. S. & A. Bickerton, (2006) The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol* **62**: 718-729.

Grabowski, B. & Z. Kelman, (2003) Archeal DNA replication: eukaryal proteins in a bacterial context. *Annu Rev Microbiol* **57**: 487-516.

Grissa, I., G. Vergnaud & C. Pourcel, (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172.

Groenen, P. M., A. E. Bunschoten, D. van Soolingen & J. D. van Embden, (1993) Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. *Mol Microbiol* **10**: 1057-1065.

Grogan, D. W., (1989) Phenotypic characterization of the archaebacterial genus Sulfolobus: comparison of five wild-type strains. *J Bacteriol* **171**: 6710-6719.

Grogan, D. W., (2009) Homologous recombination in Sulfolobus acidocaldarius: genetic assays and functional properties. *Biochem Soc Trans* **37**: 88-91.

Gudbergsdottir, S., L. Deng, Z. Chen, J. V. Jensen, L. R. Jensen, Q. She & R. A. Garrett, (2011) Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol Microbiol* **79**: 35-49.

Guo, L., K. Brugger, C. Liu, S. A. Shah, H. Zheng, Y. Zhu, S. Wang, R. K. Lillestol, L. Chen, J. Frank, D. Prangishvili, L. Paulin, Q. She, L. Huang & R. A. Garrett, (2011) Genome analyses of Icelandic strains of Sulfolobus islandicus, model organisms for genetic and virus-host interaction studies. *J Bacteriol* **193**: 1672-1680.

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

Hale, C., K. Kleppe, R. M. Terns & M. P. Terns, (2008) Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA* **14**: 2572-2579.

Hale, C. R., P. Zhao, S. Olson, M. O. Duff, B. R. Graveley, L. Wells, R. M. Terns & M. P. Terns, (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945-956.

Hallam, S. J., K. T. Konstantinidis, N. Putnam, C. Schleper, Y. Watanabe, J. Sugahara, C. Preston, J. de la Torre, P. M. Richardson & E. F. DeLong, (2006) Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. *Proc Natl Acad Sci U S A* **103**: 18296-18301.

Han, D., K. Lehmann & G. Krauss, (2009) SSO1450--a CAS1 protein from Sulfolobus solfataricus P2 with high affinity for RNA and DNA. *FEBS Lett* **583**: 1928-1932.

Haurwitz, R. E., M. Jinek, B. Wiedenheft, K. Zhou & J. A. Doudna (2010), Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**: 1355-1358.

Hermans, P. W., D. van Soolingen, E. M. Bik, P. E. de Haas, J. W. Dale & J. D. van Embden, (1991) Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. *Infect Immun* **59**: 2695-2705.

Horvath, P. & R. Barrangou, (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**: 167-170.

Huet, J., R. Schnabel, A. Sentenac & W. Zillig, (1983) Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. *EMBO J* **2**: 1291-1294.

Ishino, Y. & S. Ishino, (2012) Rapid progress of DNA replication studies in Archaea, the third domain of life. *Sci China Life Sci* **55**: 386-403.

Ishino, Y., H. Shinagawa, K. Makino, M. Amemura & A. Nakata, (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., J. D. van Embden, W. Gaastra & L. M. Schouls, (2002b) Identification of a novel family of sequence repeats among prokaryotes. *OMICS* **6**: 23-33.

Jaubert, C., C. Danioux, J. Oberto, D. Cortez, A. Bize, M. Krupovic, Q. She, P. Forterre, D. Prangishvili & G. Sezonov, (2013) Genomics and genetics of Sulfolobus islandicus LAL14/1, a model hyperthermophilic archaeon. *Open biology* **3**: 130010.

Jonuscheit, M., E. Martusewitsch, K. M. Stedman & C. Schleper, (2003) A reporter gene system for the hyperthermophilic archaeon Sulfolobus solfataricus based on a selectable and integrative shuttle vector. *Mol Microbiol* **48**: 1241-1252.

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

Kamerbeek, J., L. Schouls, A. Kolk, M. van Agterveld, D. van Soolingen, S. Kuijper, A. Bunschoten, H. Molhuizen, R. Shaw, M. Goyal & J. van Embden, (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *J Clin Microbiol* **35**: 907-914.

Karginov, F. V. & G. J. Hannon, (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* **37**: 7-19.

Kates, M., (1978) The phytanyl ether-linked polar lipids and isoprenoid neutral lipids of extremely halophilic bacteria. *Progress in the chemistry of fats and other lipids* **15**: 301-342.

Kawarabayasi, Y., Y. Hino, H. Horikawa, S. Yamazaki, Y. Haikawa, K. Jin-no, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, H. Nakazawa, M. Takamiya, S. Masuda, T. Funahashi, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, H. Kikuchi & *et al.*, (1999) Complete genome sequence of an aerobic hyper-

thermophilic crenarchaeon, Aeropyrum pernix K1. *DNA Res* **6**: 83-101, 145-152.

Kelman, Z. & M. F. White, (2005) Archaeal DNA replication and repair. *Curr Opin Microbiol* **8**: 669-676.

Kessler, C. & H. J. Holtke, (1986) Specificity of restriction endonucleases and methylases--a review. *Gene* **47**: 1-153.

Kirchman, D. L., (1994) The uptake of inorganic nutrients by heterotrophic bacteria. *Microb. Ecol.* : 28:255-271.

Klenk, H. P., R. A. Clayton, J. F. Tomb, O. White, K. E. Nelson, K. A. Ketchum, R. J. Dodson, M. Gwinn, E. K. Hickey, J. D. Peterson, D. L. Richardson, A. R. Kerlavage, D. E. Graham, N. C. Kyrpides, R. D. Fleischmann, J. Quackenbush, N. H. Lee, G. G. Sutton, S. Gill, E. F. Kirkness, B. A. Dougherty, K. McKenney, M. D. Adams, B. Loftus, S. Peterson, C. I. Reich, L. K. McNeil, J. H. Badger, A. Glodek, L. Zhou, R. Overbeek, J. D. Gocayne, J. F. Weidman, L. McDonald, T. Utterback, M. D. Cotton, T. Spriggs, P. Artiach, B. P. Kaine, S. M. Sykes, P. W. Sadow, K. P. D'Andrea, C. Bowman, C. Fujii, S. A. Garland, T. M. Mason, G. J. Olsen, C. M. Fraser, H. O. Smith, C. R. Woese & J. C. Venter, (1997) The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. *Nature* **390**: 364-370.

Konneke, M., A. E. Bernhard, J. R. de la Torre, C. B. Walker, J. B. Waterbury & D. A. Stahl, (2005) Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543-546.

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

Leigh, J. A., S. V. Albers, H. Atomi & T. Allers, (2011) Model organisms for genetics in the domain Archaea: methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiol Rev* **35**: 577-608.

Leipe, D. D., L. Aravind & E. V. Koonin, (1999) Did DNA replication evolve twice independently? *Nucleic Acids Res* **27**: 3389-3401.

Lillestol, R. K., P. Redder, R. A. Garrett & K. Brugger, (2006) A putative viral defence mechanism in archaeal cells. *Archaea* **2**: 59-72.

Lillestol, R. K., S. A. Shah, K. Brugger, P. Redder, H. Phan, J. Christiansen & R. A. Garrett, (2009) CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Mol Microbiol* **72**: 259-272.

Lintner, N. G., M. Kerou, S. K. Brumfield, S. Graham, H. Liu, J. H. Naismith, M. Sdano, N. Peng, Q. She, V. Copie, M. J. Young, M. F. White & C. M. Lawrence, (2011) Structural and functional characterization of an archaeal CASCADE complex for CRISPR-mediated viral defense. *J Biol Chem*.

Londei, P., (2005) Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol Rev* **29**: 185-200.

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

Makarova, K. S., N. V. Grishin, S. A. Shabalina, Y. I. Wolf & E. V. Koonin, (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., D. H. Haft, R. Barrangou, S. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost & E. V. Koonin, (2011b) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9**: 467-477.

Manica, A., Z. Zebec, D. Teichmann & C. Schleper, (2011) In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Mol Microbiol* **80**: 481-491.

Marraffini, L. A. & E. J. Sontheimer, (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**: 1843-1845.

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

Martin, A., S. Yeats, D. Janekovic, W. D. Reiter, W. Aicher & W. Zillig, (1984) SAV 1, a temperate u.v.-inducible DNA virus-like particle from the archaebacterium Sulfolobus acidocaldarius isolate B12. *EMBO J* **3**: 2165-2168.

Martusewitsch, E., C. W. Sensen & C. Schleper, (2000) High spontaneous mutation rate in the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by transposable elements. *J Bacteriol* **182**: 2574-2581.

Masepohl, B., K. Gorlitz & H. Bohme, (1996) Long tandemly repeated repetitive (LTRR) sequences in the filamentous cyanobacterium Anabaena sp. PCC 7120. *Biochim Biophys Acta* **1307**: 26-30.

Matsumi, R., H. Atomi, A. J. Driessen & J. van der Oost, (2011) Isoprenoid biosynthesis in Archaea--biochemical and evolutionary implications. *Res Microbiol* **162**: 39-52.

Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez & C. Almendros, (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733-740.

Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez & E. Soria, (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174-182.

Mojica, F. J., C. Diez-Villasenor, E. Soria & G. Juez, (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **36**: 244-246.

Mojica, F. J., C. Ferrer, G. Juez & F. Rodriguez-Valera, (1995) Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Mol Microbiol* **17**: 85-93.

Nakata, A., M. Amemura & K. Makino, (1989) Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. *J Bacteriol* **171**: 3553-3556.

Palm, P., C. Schleper, B. Grampp, S. Yeats, P. McWilliam, W. D. Reiter & W. Zillig, (1991) Complete nucleotide sequence of the virus SSV1 of the archaebacterium Sulfolobus shibatae. *Virology* **185**: 242-250.

Parker, J. S., E. A. Parizotto, M. Wang, S. M. Roe & D. Barford, (2009) Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol Cell* **33**: 204-214.

Pomeroy, L. R., (1974) The ocean's food web, a changing paradigm. *BioScience* **24:499-504**.

Pougach, K., E. Semenova, E. Bogdanova, K. A. Datsenko, M. Djordjevic, B. L. Wanner & K. Severinov, (2010) Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Mol Microbiol*.

Pourcel, C., G. Salvignol & G. Vergnaud, (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**: 653-663.

Prangishvili, D., P. Forterre & R. A. Garrett, (2006a) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* **4**: 837-848.

Prangishvili, D. & R. A. Garrett, (2004) Exceptionally diverse morphotypes and genomes of crenarchaeal hyperthermophilic viruses. *Biochem Soc Trans* **32**: 204-208.

Prangishvili, D. & R. A. Garrett, (2005) Viruses of hyperthermophilic Crenarchaea. *Trends Microbiol* **13**: 535-542.

Prangishvili, D., R. A. Garrett & E. V. Koonin, (2006b) Evolutionary genomics of archaeal viruses: unique viral genomes in the third domain of life. *Virus Res* **117**: 52-67.

Prangishvili, D., K. Stedman & W. Zillig, (2001) Viruses of the extremely thermophilic archaeon Sulfolobus. *Trends Microbiol* **9**: 39-43.

Pul, U., R. Wurm, Z. Arslan, R. Geissen, N. Hofmann & R. Wagner, (2010) Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Mol Microbiol* **75**: 1495-1512.

Reiter, W. D., P. Palm & S. Yeats, (1989) Transfer-Rna Genes Frequently Serve as Integration Sites for Prokaryotic Genetic Elements. *Nucleic Acids Research* **17**: 1907-1914.

Reiter, W. D., P. Palm, S. Yeats & W. Zillig, (1987) Gene expression in archaebacteria: physical mapping of constitutive and UV-inducible transcripts from the Sulfolobus virus-like particle SSV1. *Mol Gen Genet* **209**: 270-275.

Reiter, W. D., W. Zillig & P. Palm, (1988) Archaebacterial viruses. *Adv Virus Res* **34**: 143-188.

Reno, M. L., N. L. Held, C. J. Fields, P. V. Burke & R. J. Whitaker, (2009) Biogeography of the Sulfolobus islandicus pan-genome. *Proc Natl Acad Sci U S A* **106**: 8605-8610.

Rice, G., K. Stedman, J. Snyder, B. Wiedenheft, D. Willits, S. Brumfield, T. McDermott & M. J. Young, (2001) Viruses from extreme thermal environments. *Proc Natl Acad Sci U S A* **98**: 13341-13345.

Rockwood, J., D. Mao & D. W. Grogan, (2013) Homologous Recombination in the Archaeon Sulfolobus acidocaldarius: Effects of DNA Substrates and Mechanistic Implications. *Microbiology*.

Schleper, C., K. Kubo & W. Zillig, (1992) The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. *Proc Natl Acad Sci U S A* **89**: 7645-7649.

Semenova, E., M. M. Jore, K. A. Datsenko, A. Semenova, E. R. Westra, B. Wanner, J. van der Oost, S. J. Brouns & K. Severinov, (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**: 10098-10103.

Sensen, C. W., R. L. Charlebois, C. Chow, I. G. Clausen, B. Curtis, W. F. Doolittle, M. Duguet, G. Erauso, T. Gaasterland, R. A. Garrett, P. Gordon, I. H. de Jong, A. C. Jeffries, C. Kozera, N. Medina, A. De Moors, J. van der Oost, H. Phan, M. A. Ragan, M. E. Schenk, Q. She, R. K. Singh & N. Tolstrup, (1998) Completing the sequence of the Sulfolobus solfataricus P2 genome. *Extremophiles* **2**: 305-312.

Shah, S. A. & R. A. Garrett, (2011) CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. *Res Microbiol* **162**: 27-38.

She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C. Chan-Weiher, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. Gordon, I. Heikamp-de Jong, A. C. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. L. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A. Garrett, M. A. Ragan, C. W. Sensen & J. Van der Oost, (2001) The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc Natl Acad Sci U S A* **98**: 7835-7840.

Shuck, S. C., E. A. Short & J. J. Turchi, (2008) Eukaryotic nucleotide excision repair: from understanding mechanisms to influencing biology. *Cell Res* **18**: 64-72.

Silvian, L. F., E. A. Toth, P. Pham, M. F. Goodman & T. Ellenberger, (2001) Crystal structure of a DinB family error-prone DNA polymerase from Sulfolobus solfataricus. *Nat Struct Biol* **8**: 984-989.

Sorek, R., V. Kunin & P. Hugenholtz, (2008) CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6**: 181-186.

Stetter, K. O., (2006a) History of discovery of the first hyperthermophiles. *Extremophiles* **10**: 357-362.

Stetter, K. O., (2006b) Hyperthermophiles in the history of life. *Philos Trans R Soc Lond B Biol Sci* **361**: 1837-1842; discussion 1842-1833.

Stetter, K. O., Fiala, G., Huber, G., Huber, R. and Segerer, A., (1990) Hyperthermophilic microorganisms. *FEMS Microbiol. Rev.* **75**: 117-124.

Tang, T. H., T. S. Rozhdestvensky, B. C. d'Orval, M. L. Bortolin, H. Huber, B. Charpentier, C. Branlant, J. P. Bachellerie, J. Brosius & A. Huttenhofer, (2002) RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res* **30**: 921-930.

Thingstad, T., R. Lingell & (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology* **60** 3167-3174.

Wagner, M., S. Berkner, M. Ajon, A. J. Driessen, G. Lipps & S. V. Albers, (2009) Expanding and understanding the genetic toolbox of the hyperthermophilic genus Sulfolobus. *Biochem Soc Trans* **37**: 97-101.

Wagner, M., M. van Wolferen, A. Wagner, K. Lassak, B. H. Meyer, J. Reimann & S. V. Albers, (2012) Versatile Genetic Tool Box for the Crenarchaeote Sulfolobus acidocaldarius. *Front Microbiol* **3**: 214.

Wang, R., H. Zheng, G. Preamplume, Y. Shao & H. Li, (2012) The impact of CRISPR repeat sequence on structures of a Cas6 protein-RNA complex. *Protein Sci* **21**: 405-417.

Weijers, J. W., S. Schouten, E. C. Hopmans, J. A. Geenevasen, O. R. David, J. M. Coleman, R. D. Pancost & J. S. Sinninghe Damste, (2006) Membrane lipids of mesophilic anaerobic bacteria thriving in peats have typical archaeal traits. *Environ Microbiol* **8**: 648-657.

Westra, E. R., U. Pul, N. Heidrich, M. M. Jore, M. Lundgren, T. Stratmann, R. Wurm, A. Raine, M. Mescher, L. Van Heereveld, M. Mastop, E. G. Wagner, K. Schnetz, J. Van Der Oost, R. Wagner & S. J. Brouns, (2010) H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. *Mol Microbiol* **77**: 1380-1393.

Wiedenheft, B., E. van Duijn, J. B. Bultema, S. P. Waghmare, K. Zhou, A. Barendregt, W. Westphal, A. J. Heck, E. J. Boekema, M. J. Dickman & J. A. Doudna, (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* **108**: 10092-10097.

Wiedenheft, B., K. Zhou, M. Jinek, S. M. Coyle, W. Ma & J. A. Doudna, (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17**: 904-912.

Woese, C. R. & G. E. Fox, (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* **74**: 5088-5090.

Woese, C. R., O. Kandler & M. L. Wheelis, (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* **87**: 4576-4579.

Worthington, P., V. Hoang, F. Perez-Pomares & P. Blum, (2003) Targeted disruption of the alpha-amylase gene in the hyperthermophilic archaeon Sulfolobus solfataricus. *J Bacteriol* **185**: 482-488.

Yosef, I., M. G. Goren & U. Qimron, (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res*.

Zhang, J., C. Rouillon, M. Kerou, J. Reeks, K. Brugger, S. Graham, J. Reimann, G. Cannone, H. Liu, S. V. Albers, J. H. Naismith, L. Spagnolo & M. F. White, (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* **45**: 303-313.

Zillig, W., D. Prangishvilli, C. Schleper, M. Elferink, I. Holz, S. Albers, D. Janekovic & D. Gotz, (1996) Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic Archaea. *FEMS Microbiol Rev* **18**: 225-236.

Zillig, W., Stetter, K.O., Wunderl, S., Schulz, W., Priess, H. and Scholz, I., (1980) The Sulfolobus "Caldariella" group: taxonomy on the basis of the structure of DNA-dependet RNA polymerases. *Arch. Microbiol.* **125**: 259-269.

Zillig, W., J. Tu & I. Holz, (1981) Thermoproteales--a third order of thermoacidophilic archaebacteria. *Nature* **293**: 85-86.

5 REVIEW: CRISPR-MEDIATED DEFENSE MECHANISMS IN THE HYPERTHERMOPHILIC ARCHAEAL GENUS *SULFOLOBUS*

Andrea Manica* and Christa Schleper

University of Vienna, Department of Genetics in Ecology, Althanstr. 14, A-1090 Vienna, Austria

*corresponding author,

Andrea Manica, e-mail: andrea.manica@univie.ac.at

Keywords: Sulfolobales, archaea, virus defense, CRISPR-Cas system, small RNAs.

5.1 Abstract

CRISPR (clustered regularly interspaced short palindromic repeats)-mediated virus defense based on small RNAs is a hallmark of archaea and also found in many bacteria. Archaeal genomes, and in particular organisms of the extremely thermoacidophilic genus *Sulfolobus* carry extensive CRISPR loci each with dozens of sequence signatures (spacers) able to mediate targeting and degradation of complementary invading nucleic acids. The diversity of CRISPR systems and their associated protein complexes indicates an extensive functional breadth and versatility of this adaptive immune system. *Sulfolobus solfataricus* and *S. islandicus* represent two of the best characterized genetic model organisms in the archaea not only with respect to the CRISPR system. Here we address and discuss in a broader context particularly recent progress made in understanding spacer recruitment from foreign DNA, production of small RNAs, *in vitro* activity of CRISPR-associated protein complexes and attack of viruses and plasmids in *in vivo* test systems.

5.2 Introduction

Small RNAs play a key role in cell regulation, stability and defense in all three domains of life (Huttenhofer *et al.* 2005; Matera *et al.* 2007; Bartel 2009; Ghildiyal and Zamore 2009; Waters and Storz 2009; Marchfelder *et al.* 2012). Many different types of small RNAs were identified and their mode of action in gene regulation, epigenetics and defense were extensively studied in eukaryotes and prokaryotes over the past years (Alvarado and Scholthof 2009; Zhou and Xie 2010; Storz *et al.* 2011). Recently a new type of small RNAs called crRNAs (from CRISPR RNAs, Clustered Regularly Interspaced Short Palindromic Repeats) was discovered in Bacteria and Archaea which has been shown to be active against invading genetic elements (Barrangou *et al.* 2007; Brouns *et al.* 2008; Marraffini and Sontheimer 2008; Hale *et al.* 2009; Gudbergsdottir *et al.* 2011; Manica *et al.* 2011). This defense system is found in almost every genome of archaea and in about half of the bacterial genomes (http://crispi.genouest.org/) (Rousseau *et al.* 2009). Although CRISPR-mediated defense is a fundamentally new process, its mode of action resembles formally the basic principles of siRNA activities in Eukaryotes (Brodersen and Voinnet 2006; Tolia and Joshua-Tor 2007; Marraffini and Sontheimer 2010), including i) recognition of the invading genetic element, (ii) amplification of the effector molecule, (iii) processing of the precursor RNA, and (iv) targeting of the invader through base-pair recognition. However, there are also many remarkable differences between the two systems. For example, crRNAs being processed from a long precursor RNA (pre-crRNA) transcribed from a CRISPR locus, i.e. a DNA region with dozens of unique DNA sequences of 30-45 nt (called spacers) that are separated by identical direct repeats (Jansen *et al.* 2002; Bolotin *et al.* 2005; Haft *et al.* 2005; Godde and Bickerton 2006; Lillestol *et al.* 2006; Makarova *et al.* 2006; Grissa *et al.* 2007; Kunin *et al.* 2007). Upon viral or plasmid invasion new short DNA sequences from the invader can be incorporated as spacers in the leader-proximal part of a CRISPR locus (Barrangou *et al.* 2007; Erdmann and Garrett 2012; Yosef *et al.* 2012). The mechanism of recognition of the extra-chromosomal element as invader is still poorly understood. However, based

on the enzymatic properties of the proteins involved in the recognition/integration of new spacers (Babu, *et al*. 2011; Makarova *et al*. 2011; Datsenk *et al*. 2012; Erdmann and Garrett 2012; Swarts, Mosterd *et al*. 2012; Yosef *et al*. 2012), this process unlike the eukaryotic RNAi system, should not involve the presence of dsRNAs (Fire *et al*. 1998; Hamilton and Baulcombe 1999; Bernstein *et al*. 2001). The CRISPR Cas system activity can be divided into three temporally and functionally distinct processes: Adaptation, Expression/processing and Interference. All these steps involve different sets of Cas proteins (CRISPR-associated proteins) and in some systems endogenous non-cas gene products (Deltcheva *et al*. 2011).

During adaptation two proteins defined as Cas1 and Cas2 ("core proteins"), ubiquitous to all CRISPR systems, recognize the invading genetic element and drive the incorporation of new DNA sequences into the CRISPR locus (Makarova *et al*. 2011). In this way the invasion of an extra-chromosomal element is "recorded" in the host chromosome allowing the system to be prepared for future infections of the same invader. Once integrated, the new spacer is co-transcribed with all other spacers located in the same CRISPR locus. Transcription initiates at a promoter located within the leader sequence (upstream of the repeat cluster) and allows the CRISPR-Cas system to amplify the signal by producing a multitude of crRNAs from a single spacer. CRISPR and Cas expression is often under the control of cellular regulators (Agari *et al*. 2009; Pul *et al*. 2010; Westra *et al*. 2010) and seems to be inducible by abiotic and/or biotic stress (Agari *et al*. 2009; Perez-Rodriguez *et al*. 2011; Plagens *et al*. 2012).

Upon transcription the long CRISPR pre-RNA is processed by other Cas proteins (i.e. Cas6) or by endogenous proteins, to the short mature crRNAs (Brouns *et al*. 2008; Carte *et al*. 2008; Lintner *et al*. 2011). A mature crRNA is composed of a spacer derived from the extra-chromosomal genetic element, an 8nt 5´tag derived from the preceding repeat and a 3´end handle of variable length that stems from the downstream repeat (Brouns *et al*. 2008; Carte *et al*. 2008; Hale *et al*. 2009; Lintner *et al*. 2011; Zhang *et al*. 2012). Once mature, the crRNA is incorporated into the "interference" complex formed by

several CRISPR-associated proteins and guides the enzymatic machinery to the targeted foreign nucleic acid *via* base pairing interaction.

Several classifications of the highly divergent CRISPR-Cas systems were proposed based on the type of repeat (Kunin *et al*. 2007), and/or Cas proteins encoded near the locus. Haft and colleagues initially identified several CRISPR type-specific proteins which allowed the classification into more than 45 families (Haft *et al*. 2005). Recently, a new classification merged the different CRISPR families into three major types: type I, type II, type III, which can be further subdivided (Makarova *et al*. 2011). Several CRISPR types can be present simultaneously on the chromosome of the same organism, indicating that each type could have different activities against various genetic elements or nucleic acids. CRISPR systems were extensively studied in different organisms and many molecular details of their activities were elucidated. CRISPR type I *(Escherichia coli, S. solfataricus, Pseudomonas aeruginosa),* type II *(Streptococcus thermophilus)* and type IIIA *(Staphylococcus epidermidis)* seem to target principally DNA whereas type IIIB CRISPR (*Pyrococcus furiosus, S. solfataricus*) might be involved in the targeting of invading RNAs (Barrangou *et al*. 2007; Brouns *et al*. 2008; Marraffini and Sontheimer 2008; Hale *et al*. 2009; Gudbergsdottir *et al*. 2011; Manica *et al*. 2011; Semenova *et al*. 2011; Zhang *et al*. 2012). Interestingly, type II CRISPR are only present in bacterial genomes (Makarova *et al*. 2011) while type I and III CRISPR are widespread within Bacteria and Archaea (Makarova *et al*. 2011). Based on comparative studies it has been inferred that intra- as well as interdomain transfer of CRISPR loci between archaea and bacteria must have occurred (Makarova *et al*. 2011; Shah and Garrett 2011). The latter is particularly remarkable, considering the fundamental differences of transcriptional and translational mechanisms in the two domains, as well as of their cell wall structures and transferable genetic elements (Cavicchioli 2007; Garrett and Klenk 2007). *S. solfataricus and S. islandicus* currently represent the most intensely studied archaea regarding the CRISPR-Cas system. Expression and genomic arrangements of the CRISPR loci have been analysed (Brugger *et al*. 2004; Lillestol *et al*. 2006;

Han and Krauss 2009; Han *et al*. 2009; Lillestol *et al*. 2009; Shah *et al*. 2009; Garrett *et al*. 2011; Shah and Garrett 2011), *in-vivo* and *in-vitro* activity of several complexes and proteins have been characterized and crystal structures of Cas proteins have been solved (Han and Krauss 2009; Han *et al*. 2009; Lintner *et al*. 2011; Zhang *et al*. 2012; Zhang *et al*. 2012). *S. solfataricus* possesses an extensive and complex CRISPR system which includes six different CRISPR loci, two "adaptation" Cas cassettes and five "interference" Cas protein complexes, the latter belonging to CRISPR type I (three clusters), CRISPR type IIIA and type IIIB, respectively. The complexity of the CRISPR-Cas systems might be correlated to the high diversity and abundance of genetic elements and viruses in archaea: Viruses from eight out of twelve structurally and genomically highly divergent virus families infect *Sulfolobales* or closely related thermoacidophilic organisms (Pina *et al*. 2011). In the 3 Mbp genome of *S. solfataricus* P2 a total of 344 putatively mobilizable genetic elements can be found that make up about 10% of the whole genome (Brugger *et al*. 2004). Besides the large numbers of mobile genetic elements, genetic tools have been developed for several *Sulfolobus* species (She *et al*. 2009; Wagner *et al*. 2009) which allow to manipulate and functionally study the virus-host systems (Iverson and Stedman 2012; Maaty *et al*. 2012) as well as the CRISPR-based defense mechanism (Gudbergsdottir *et al*. 2011; Manica *et al*. 2011; Erdmann and Garrett 2012).

5.3 CRISPR Locus Organisation in *Sulfolobales*

Within the Sulfolobales more than 80 repeat clusters were identified, with a total of ~4500 unique spacer sequences (Held *et al*. 2010; Garrett *et al*. 2011; Garrett *et al*. 2011). The length of individual CRISPR spacers is fairly conserved within the different CRISPR loci ranging between 34 and 44 nucleotides. The repeat-clusters can be several kilobases in length carrying more than 100 different spacers. CRISPR loci are usually conserved with respect to their repeat sequence and spacer length although in some CRISPR loci repeat sequences are not always uniform (Shah and Garrett 2011). Interestingly, also two self-transmissible plasmids (pNOB8 and pKEF9)

which can spread among Sulfolobales carry small repeat-clusters in their genomes. However they lack the adjacent Cas genes (Lillestol *et al*. 2009).

The majority of the spacers in a CRISPR locus are unique, although duplicate spacers can also be found in large repeat clusters (Shah *et al*. 2009). Different strains of *S. solfataricus* share many identical spacers in the leader distal part of the CRISPR locus, whereas different strains of *S. islandicus* isolated from geographically distant hot-springs show no spacer conservations (Shah and Garrett 2011). Moreover, only low CRISPR spacer conservation was found within a population of *S. islandicus* strains isolated from the same hot-spring in Kamchatka (Held *et al*. 2010).

Many *Sulfolobus* species carry several CRISPR loci on their chromosome which do not always share the same repeat type. A particular type of repeat is often associated with a particular cluster of core proteins (Cas1/Cas2). Also, the upstream region (leader) is often conserved in CRISPR loci that share the same repeat type (Lillestol *et al*. 2006; Lillestol *et al*. 2009).

5.4 CRISPR transcription and regulation in *Sulfolobales*

The leader sequence provides the DNA elements for CRISPR transcription. In *S. acidocaldarius* transcriptional CRISPR start sites were mapped directly before or 17-21 nt upstream of the first repeat sequence and they were always preceded by a typical archaeal BRE/TATA box promoter motif (Torarinsson *et al*. 2005; Lillestol *et al*. 2009). Along the leader sequence additional motifs were found to be conserved between the different CRISPR families although the role of these motifs remains unknown (Lillestol *et al*. 2009).

The long pre-crRNA transcript covering the full length of the locus is successively processed by Cas6 into small mature crRNAs (Lintner *et al*. 2011). Full length transcripts were detected by Northern blot assay for each CRISPR locus of *S. acidocaldarius* using oligonucleotide probes complementary to different CRISPR spacers and repeats (Lillestol *et al*. 2006; Lillestol *et al*. 2009). Small RNAs of 50-70nt were also detected, compatible with the processing of the pre-crRNA within the repeat sequence (Lillestol *et*

al. 2009; Lintner *et al*. 2011). Furthermore, RNAs of shorter size, i.e. ~40nt in length were detected using spacer-specific probes, but not with repeat-based probes, demonstrating that some crRNAs undergo a further processing of their repeat by an unknown exonuclease (Lillestol *et al*. 2009). These results were supported by sequencing of crRNAs that co-purified with the Cmr complex of *S. solfataricus (*Zhang *et al. 2012).*

Northern blot analysis of CRISPR pre-RNA revealed the presence of antisense transcripts for each of the *S. acidocaldarius* CRISPR loci (Lillestol *et al*. 2009). Bi-directional transcription was also experimentally verified in *S. solfataricus* CRISPR loci (Tang *et al*. 2002; Tang *et al*. 2005; Lillestol *et al*. 2006; Lillestol *et al*. 2009) and was confirmed by analyzing publicly available transcriptome data (Wurtzel *et al*. 2009). However, the loci were primarily transcribed from the leader strand (Tang *et al*. 2002; Tang *et al*. 2005; Lillestol *et al*. 2009). Additionally, internal transcription start sites have also been identified for some *Sulfolobus* CRISPR loci, probably caused by the incorporation of spacer sequences carrying promoter-like motifs (Garrett *et al*. 2011). The presence of bi-directional transcripts has been found in few archaeal species: *Sulfolobus spp., P. furiosus and Methanococcus maripaludis* (Hale *et al*. 2009; Richter *et al*. 2012). In bacteria, with the exception of *Clostridium thermocellum,* no antisense transcripts were detected and CRISPR transcription seems to be mostly uni-directional (Brouns *et al*. 2008; Pul *et al*. 2010; Westra, Richter *et al*. 2012). Although several hypotheses have been formulated, the biological importance of bidirectional transcription remains unclear and no experimental data have so far elucidated the role of antisense RNAs in CRISPR regulation or interference (Lillestol *et al*. 2009; Deng *et al*. 2011).

The presence of several putative regulatory motifs within the leader sequence, and the high energetic cost related to the possible constitutive expression of a long non-coding RNA (pre-crRNA) in the absence of an invading genetic element, speaks for a complex transcriptional regulation of the CRISPR locus. In *E. coli* CRISPR locus and associated Cas gene

transcription is under the control of the repressor H-NS which binds particular DNA motifs within the leader sequence and in the promoter of some Cas genes silencing their expression (Pul *et al*. 2010). Transcription can be restored only in the presence of the transcriptional activator LeuO (Westra *et al*. 2010). Within *Sulfolobales* (or other archaea) no such type of regulation has so far been demonstrated. However, Deng and co-workers identified a protein, Cbp1 (SSO0454) which directly interferes with the CRISPR locus transcription by binding to the CRISPR DNA repeats of *S. solfataricus* and of the conjugative plasmid pNOB8 (Peng *et al*. 2003; Deng *et al*. 2011). Mutation and expression studies performed in *S. solfataricus* and *S. islandicus*, respectively, imply a role of this protein in the transcriptional regulation of the CRISPR locus (Deng *et al*. 2011). Cpb1 knock out mutants of *S. islandicus* showed strong reduction in CRISPR pre-crRNA levels (>600nt) compared to the wild type strain and a Cbp1 complemented strain (Deng *et al*. 2011). Conversely, overexpression of the Cbp1 protein in *S. solfataricus* strain P2 revealed higher pre-crRNA (>600nt long) transcription levels compared to the wild-type strain (Deng *et al*. 2011).

Cbp1 showed different binding affinity depending upon the repeat type, with higher affinity to the repeat sequence of locus C and D. In contrast to its enhancing activity on the leader transcript, no increase of CRISPR antisense transcripts and a decrease in transcription starting at internal sites was detected (Deng *et al*. 2011). Interestingly, microarray analysis of *Sulfolobus* P2 cells over-expressing Cbp1 showed significant reductions of the SSO1101 transcript, encoding one of the few proteins found to be involved in biofilm formation in *Sulfolobales sp*. (Deng *et al*. 2011; Koerd *et al*. 2011) . As postulated by the authors the reduction of expression of proteins involved in biofilm formation could be a synergistic mechanism to block invader colonization of the neighboring cells (Deng *et al*. 2011). If correct, this would imply another difference to the eukaryotic RNAi system in which the interference signal can spread from cell to cell through the whole organism helping the recognition and the targeting of the invaders even in cells located very distantly from the infection site (Hyun *et al*. 2010).

Taken together the above results indicate a role of Cbp1 in the transcriptional regulation of *Sulfolobus* CRISPR loci, although its mode of action is still not fully understood.

5.5 Complexity and Adaptation of the *Sulfolobus* CRISPR-Cas System

Locus A and B of the six CRISPR loci (termed A through F) in *S. solfataricus* share the same repeat and similar leader sequences (Lillestol *et al.* 2006; Lillestol *et al.* 2009) which are distinct from those shared among loci C, D, and E. In the proximity of locus A, B C, and D two "adaptation" cassettes (Cas1-Cas2) are present (Fig.1).

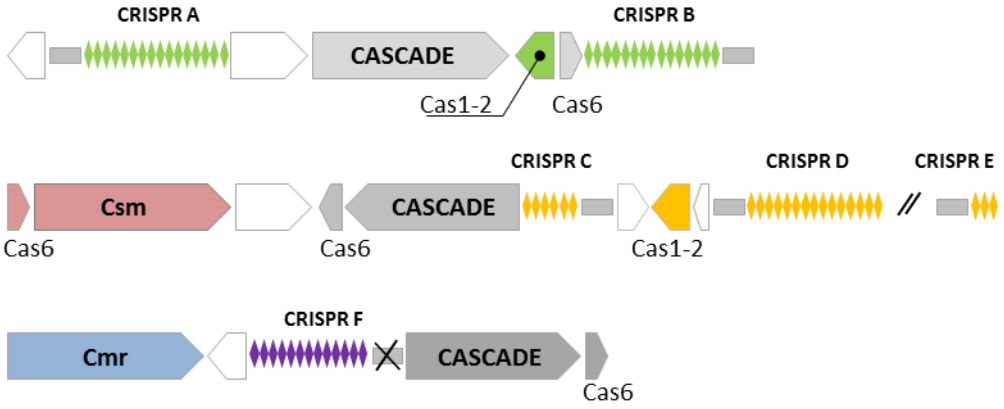


Fig.1: **Schematic representation of the CRISPR-Cas systems in Sulfolobus solfataricus**. The six CRISPR loci are named from A through F; the small gray rectangle in front of each locus represents the leader sequence. Loci with the same repeats share the same colour. The putative "integration cassette" Cas1-Cas2 is highlighted with the same colour as its adjacent locus. In dark grey the three CASCADE cassettes located adjacent to a CRISPR locus, and the Cmr/Csm cassettes near the CRISPR F (light blue) and C (light red), respectively. In white additional CRISPR related proteins.

The originally suggested classification by Lillestol and co-workers (2009), which comprised eight different CRISPR families for archaea was based on leader sequences, repeat type and Cas1 gene phylogeny (Lillestol, Shah *et al.* 2009; Shah *et al.* 2009; Shah and Garrett 2011). Since it has been shown that Cas1 is involved in the integration of new spacers (Babu *et al.* 2011; Yosef *et*

al. 2012) this first classification could also reflect different spacer acquisition mechanisms of the different archaeal CRISPR families, whereas the classification of Makarova *et al*. [30] is based primarily on the phylogeny of Cas proteins involved in the processing and interference process.

Although it is still unknown how the CRISPR system discriminates the extra-chromosomal element from the host chromosome during acquisition of new spacers, this process has been demonstrated under laboratory conditions for *E. coli* (Swarts *et al*. 2012; Yosef *et al*. 2012) , *S. thermophilus* (Barrangou *et al. 2007)*, and recently *S. solfataricus* (Erdmann and Garrett 2012). The process was shown to be dependent on a small recognition motif called PAM (protospacer adjacent motif) (Barrangou *et al*. 2007; Mojica *et al*. 2009; Gudbergsdottir *et al*. 2011; Semenova *et al*. 2011) located in the vicinity of the protospacer DNA in the foreign genetic elements, i.e. adjacent to the segment that is incorporated as a new spacer.

In *S. solfataricus* active uptake of new spacers was demonstrated after challenging the host with environmental samples. In her study, Erdmann found that only locus C, D and E (CRISPR family I) incorporated new spacers. While the inactivity was expected for locus F (family I), which lacks the Cas1 and Cas2 genes and the leader sequence upstream of the CRISPR locus (Lillestol *et al*. 2009; Shah and Garrett 2011), it was more surprising that no spacers were acquired by CRISPR loci A and B (family II) which show a complete integration cassette formed by Cas4, Cas1 and Cas2 and which were already shown to be active during the interference process (Gudbergsdottir *et al*. 2011; Manica *et al*. 2011). The newly acquired spacers were incorporated after the first repeat in locus C and D similar to the adaptation mode found in *E. coli* (Yosef *et al*. 2012). Unexpectedly, new spacers were also integrated at different positions within the CRISPR locus E (Erdmann and Garrett 2012). All these three CRISPR loci (C, D and E) carry the same direct repeat, underlining the possibility that different integration mechanisms could be active in the different CRISPR families (Erdmann and Garrett 2012). Remarkably, almost all of the newly integrated spacers were found to match plasmid-like sequences and no viral spacers; although the

presence of plasmids and viruses in the same culture was demonstrated (Erdmann and Garrett 2012). This result suggests a bias towards the incorporation of plasmid-like sequences in the CRISPR family I, although previous bioinformatic analyses of the spacers located within *S. solfataricus* and *S. islandicus* CRISPR loci do not indicate preferential integration of plasmid-derived sequences (Lillestol *et al*. 2006; Held and Whitaker 2009).

The incorporation of "exclusively" plasmid-derived spacers could be explained by the presence of a CRISPR-resistant virus and an aggressive conjugative plasmid in the enrichment culture (Lillestol *et al*. 2006; Held and Whitaker 2009; Deng *et al*. 2011; Erdmann and Garrett 2012). It is not difficult to think that the co-evolutionary arms-race between host and invaders could have led to the selection of mechanisms to avoid CRISPR interference or/and to block the incorporation of new spacers (Bondy-Denomy *et al*. 2013). Integration of the viral genome into the host chromosome or modification of the nucleic acid via e.g. methylation could represent simple but effective mechanisms to avoid CRISPR spacer integration and interference.

In her experiment Erdmann reported the integration of hundreds of new spacers into the CRISPR loci increasing our understanding of the protospacer adjacent motif (PAM) in *Sulfolobus*. The mapping of the newly acquired protospacer on the plasmid genome has confirmed the short di-nucleotide sequence GG for loci C, D, and E as PAM, the same di-nucleotide sequence previously identified by bioinformatic analyses (Lillestol *et al*. 2006; Held and Whitaker 2009; Lillestol *et al*. 2009; Mojica *et al*. 2009; Deng *et al*. 2011; Erdmann and Garrett 2012).

5.6 CAS protein complexes of *Sulfolobus*

Clusters of genes encoding CRISPR-associated proteins that provide the enzymatic machinery of the system are present in the proximity of CRISPR loci. On the *S. solfataricus* chromosome, two "integration cassettes" five "interference complexes" as well as other CRISPR-related proteins are encoded near the different loci. Loci A-B, C-D and F carry in their proximity a Cas cassette with high similarity to the CASCADE complex found in *E.coli*

(Shah and Garrett 2011), additionally a Csm and Cmr cassette (the latter formerly also referred to as RAMP proteins and according to Makarova (Makarova *et al.* 2011) now referred to as CRISPR type III A or B module) can be found in the vicinity of CRISPR C and F, respectively (Garrett *et al.* 2011; Manica *et al.* 2011; Shah and Garrett 2011) (Fig.2).

The Type I CRISPR is the most diverse type with six different subtypes and its principle characteristic is the presence of the Cas3 protein, a helicase-nuclease involved in the targeting and degradation of foreign ssDNA (Sinkunas *et al.* 2011). CRISPR type I was intensively studied in *E. coli* and *P. aeruginosa* (Brouns *et al.* 2008; Jore *et al.* 2011; Wiedenheft *et al.* 2011; Cady *et al.* 2012; Haurwitz *et al.* 2012). In those systems a long precursor CRISPR RNA is processed to crRNA by Cas6, an endonuclease which recognizes and cleaves the pre-crRNA transcript at a particular site within the repeat. Once formed, the crRNA guides a complex of several Cas proteins called CASCADE (CRISPR associated complex for antiviral defense) (Brouns, Jore *et al.* 2008) to the target DNA, blocking the invasion of genetic elements that carry DNA sequences (protospacer) complementary to the guide crRNA. In the CRISPR type I the pairing between crRNA and its target and the presence of a PAM site (Mojica *et al.* 2009) are essential conditions for an efficient DNA degradation (Gudbergsdottir *et al.* 2011; Semenova *et al.* 2011). The presence of PAM originally identified as being essential for *de-novo* spacer acquisition was demonstrated to be an essential feature also for the interference process in CRISPR type I systems (Semenova *et al.* 2011). Furthermore *in vivo* and *in vitro* studies have demonstrated the importance of the first 7-8 nucleotides at the 5´end of the crRNA ("seed" sequence) during the recognition process (Semenova *et al.* 2011; Wiedenheft *et al.* 2011). Mutations in the seed sequence are not tolerated and lead to a loss of immunity. In *S. solfataricus* three different CASCADE complexes are located in the proximity of CRISPR locus A-B, C-D and F.

Two of the Cas proteins located near the CRISPR locus C and D show high similarity to those of the *E. coli* CASCADE complex. This Cas cluster encodes proteins homologous to CasC (Csa2; SSO1442 also known as Cas7) and CasD

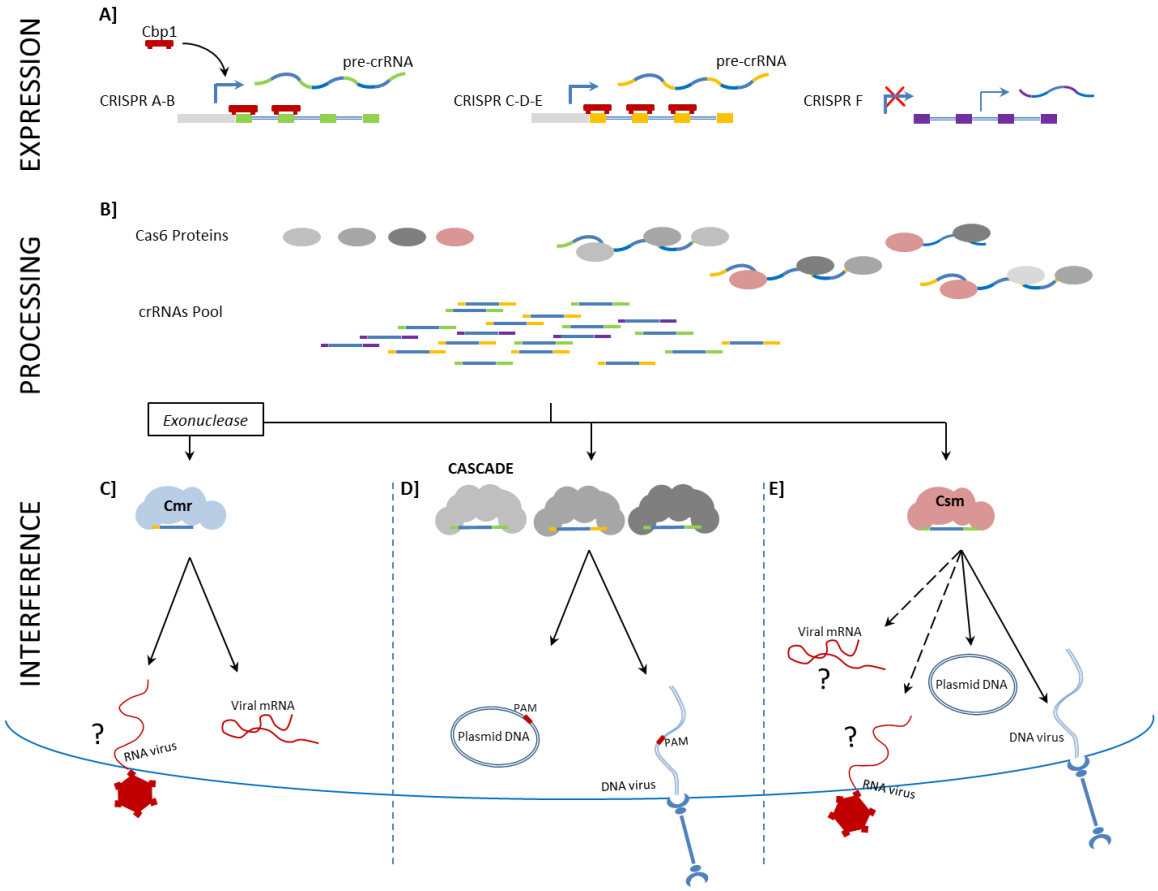


Fig.2: Overview of the CRISPR-Cas system from *Sulfolobus solfataricus* (SSO). CRISPR expression and processing in Fig. A-B and CRISPR interference in Fig. C-E.

CRISPR EXPRESSION: A] Representation of the different CRISPR loci in *Sulfolobus solfataricus*. Leader transcriptional start sites are indicated. In dark red the CRISPR regulator Cbp1 which binds specifically CRISPR repeats influencing CRISPR transcription (Deng *et al.* 2011). Putative internal transcriptional start sites are reported for locus F. **B]** The pre-crRNA is recognized by Cas6 and processed (Lintner *et al.* 2011). The four orthologous Cas6 proteins are coloured based on their proximity to the respective Cas module. An unknown exonuclease could be involved in the processing of the crRNAs before these are integrated into the Cmr complex.

CRISPR INTERFERENCE: different crRNAs are loaded into the different protein complexes. **C]** CRISPR type IIIA: *Csm* module, due to its similarity with the *S. epidermidis* system its most probable target is dsDNA (Marraffini and Sontheimer 2008). Question marks denote possible, but still unknown functions of the complex. **D]** CRISPR type IA: The three CASCADE complexes interact with crRNAs coming from different loci targeting extra-chromosomal DNA (Lintner *et al.* 2011). **E]** CRISPR type IIIB: the Cmr module binds crRNAs lacking the 3´end of the repeat (Zhang *et al.* 2012) cleaving RNAs. Question marks denote putative, but still not identified archaeal RNA viruses.

(Cas5a, SSO1441) which co-purify in *S. solfataricus* with the processed crRNA and can bind specifically crRNAs and recognize single-stranded DNA when expressed and purified from *E. coli* (Lintner *et al*. 2011). The complex binds crRNAs of 60-70nt in length and sequencing of the cloned small RNA fragments revealed the presence of full repeat-spacer units with 8nt of the repeat at the 5´-end of the spacer sequence followed by the unique spacer with the remaining 16-17nt of the repeat at its 3´end (Lintner *et al*. 2011). Interestingly, the crRNAs that co-purified with the archaeal aCASCADE originated from all the different *S. solfataricus* CRISPR loci and not only from the adjacent loci C and D, indicating that this complex might be able to exert its activity with all or most crRNAs of the organism.

The Cas7 (Csa2)-Cas5a complex (CASCADE) also co-purified with Cas6 (SSO1437), Csa5 (SSO1443) and possibly with Csa4 (Lintner *et al*. 2011). Cas6 even seemed to be physically linked with the Cas7-Cas5a complex and was directly involved in the processing of the direct repeat, as previously described for Cas6 proteins from the archaeon *P. furiosus* and from *E. coli* (Brouns *et al*. 2008; Carte *et al*. 2008; Lintner *et al*. 2011).

S. solfataricus possesses four Cas6 homologues located in the proximity of the three putative CASCADE complexes and one in the vicinity of the Csm module (Haft *et al*. 2005). One of those (heterologously expressed) Cas6 proteins (SSO2004) which is encoded adjacent to the CASCADE complex of locus F showed the ability to cleave specifically within the direct repeats in a sequence-specific manner (Lintner *et al*. 2011). The presence of four different Cas6 genes in *S. solfataricus*, primarily located at the beginning or at the end part of the different Cas cassettes, reinforces the observation that each Cas6 protein is linked to one of the different CASCADE and/or Csm complexes. But the presence of crRNAs derived from all the different CRISPR loci in the CASCADE complex analysed from Lintner and co-workers raises the possibility that a single Cas6 gene can recognize repeats from all different CRISPR loci (Lintner *et al*. 2011). Remarkably, a recent publication has shown that in *S. islandicus* REY15A three different Cas interference cassette complexes (CRISPR type I and two type IIIB) coexist with a single Cas6

protein. This observation suggests that in some CRISPR-Cas systems a single Cas6 protein can interact with different Cas interference complexes or that the pre-crRNA processing by Cas6 takes place without a physical interaction between Cas6 and the interference complexes (Deng *et al*. 2013).

Additionally, during the isolation of the *Sulfolobus* CASCADE complex, paralogous proteins of Cas7 (Csa2; SSO1399) and Cas5a (SSO1400) genes were co-purified (Lintner *et al*. 2011). The ability of a single CASCADE complex to interact with crRNAs coming from all different CRISPR loci and the possibility of the formation of "mixed" CASCADE complexes, underline the complexity of the *Sulfolobus* CRISPR-Cas system. This complexity could be important to add further versatility and flexibility to the system. The possibility that the different interference complexes could use any spacers located on the CRISPR chromosome might increase the flexibility of the immune system. Furthermore, the presence of multiple putative CASCADE cassettes and Cmr/Csm modules, not only constitutes an extended protection for the cell, but is also a fertile source for the evolution of new CRISPR-Cas variants.

In vitro, the *Sulfolobus* CASCADE showed the ability to bind single stranded DNA upon interaction with a crRNA, but did not convey DNA cleavage (Lintner *et al*. 2011). However, *in vivo* studies in *Sulfolobus* demonstrate that CRISPR crRNA-based DNA cleavage indeed occurs (Gudbergsdottir *et al*. 2011; Manica *et al*. 2011).

The presence of several CRISPR types suggests a specialization of the different complexes in targeting different extra-chromosomal elements and/or nucleic acids (Fig.2). The type III CRISPR system has been intensively studied from *Staphylococcus epidermis* (CRISPR type IIIA) (Marraffini and Sontheimer 2008), *P. furiosus* (Hale *et al*. 2009) and recently *S. solfataricus* (Zhang *et al*. 2012) (CRISPR Type IIIB). It is characterized by the presence of Cas10 proteins possibly involved in crRNA processing and nucleic acid targeting. Furthermore, it usually contains a Cas6 protein shown to be important for crRNA maturation. Interestingly, the CRISPR type IIIA system targets DNA *in vivo* but not RNA (Marraffini and Sontheimer 2008), whereas

the type IIIB system of the archaea *Pyrococcus* and *Sulfolobus* (Cmr module SSO1986-92) cleaves RNA *in vitro* (Hale *et al*. 2009; Zhang *et al*. 2012). Both, type IIIA and IIIB have shown no need for PAM sequences to cleave RNA or DNA, respectively (Hale *et al*. 2009; Marraffini and Sontheimer 2010). Nevertheless, a recent study on one of the two Cmr complexes of *S. islandicus* REY15A (Cmr-α), which is phylogenetically very distant from the previous Cmr complexes analysed *in-vitro*, demonstrates that type IIIB CRISPR-Cas systems can also target DNA *in vivo*, although its activity seems to be dependent on protospacer transcription (Deng *et al*. 2013).

The Cmr complex of *Sulfolobus*, like the CASCADE, showed the ability to interact with spacers from many CRISPR loci, but it interacted preferentially with crRNAs transcribed from CRISPR A and D (Zhang *et al*. 2012). Additionally, the crRNAs that co-precipitated with the Cmr complex showed the characteristic 8nt 5´tag but a shorter, often completely absent 3´ handle (Zhang *et al*. 2012), as previously observed in *Pyrococcus furiosus,* although the trimming of the crRNA handle seems to involve a different mechanism in that archaeon (Hale *et al*. 2009). This result fits with Cas6 cleavage of the repeat followed by exonucleolytic digestion of the spacer 3´ handle from an unknown nuclease (Zhang *et al*. 2012).

The presence of the crRNA 8nt 5´tag and an unpaired sequence at the 3´ end of the protospacer were pre-requisites for the Cmr-mediated RNA cleavage *in vitro*. Furthermore, no need of PAM sequences in the targeted RNA was necessary. Not only cleavage of the target RNA was observed in *Sulfolobus* but also a concomitant cleavage of crRNA albeit at lower levels. RNA cleavage sites were mapped on AU dinucleotides (Zhang *et al*. 2012) and the absence of AU sites in the target RNA abolished cleavage. Interestingly, cleavage of the Cmr complex in the archaeon *Pyrococcus* follows a different mechanism, as the cutting site is located at a fixed position measured from the 3'end of the cognate crRNA (Hale *et al*. 2009).

5.7 Targeting the invader: *in vivo* studies

CRISPR-mediated immunity was recently demonstrated and characterized *in vivo* in *Sulfolobus cells* using recombinant viruses (Manica, Zebec *et al*. 2011) and plasmids (Gudbergsdottir *et al*. 2011), respectively, that contained protospacer sequences compatible with crRNAs of the chromosome.

Spacers of CRISPR locus A, and D conferred immunity against extra-chromosomal elements carrying homologous protospacers whereas spacers located in the CRISPR locus F did not (Gudbergsdottir *et al*. 2011; Manica *et al*. 2011). The absence of interference activity of *S. solfataricus* CRISPR F can be correlated with its lower pre-RNA transcription that might in turn be caused by the absence of a leader sequence (Lillestol *et al*. 2009; Shah *et al*. 2009; Wurtzel *et al*. 2009; Shah and Garrett 2011). Nevertheless, processed crRNAs of this potentially "inactive" locus were found in the co-purified CASCADE crRNA pool (Lintner *et al*. 2011).

The DNA targeting activity of the *Sulfolobus* CRISPR system was quantified and characterized in detail. Both viruses and plasmids carrying perfectly matching protospacers showed transformation/transfection efficiencies that were orders of magnitudes lower than those of control vectors or were even not detectable (Gudbergsdottir *et al*. 2011; Manica *et al*. 2011). Furthermore, challenging the CRISPR system with perfectly matching protospacers under selective pressure, led to partial deletion of the respective CRISPR locus (Gudbergsdottir *et al*. 2011). Similarly, an artificial miniCRISPR locus encoded on the incoming virus and targeting an endogenous non-essential gene was not stable in culture and underwent extensive recombination that led to loss of the targeting spacers (Manica *et al*. 2011). Taken together these results indicate that DNA was targeted by the CRISPR-Cas system. Interestingly, few nucleotide mutations between crRNA and protospacers did not strongly abolish the DNA interference (Gudbergsdottir *et al*. 2011; Manica *et al*. 2011). The system tolerated up to 4 mutations at the 3´half of the crRNA and a more recent study in our laboratory demonstrates that many more mutations are tolerated by the system (Manica *et al*., submitted).

Despite the high tolerance of the CRISPR-Cas system for mutations located at the crRNA 3´half, only six mutations in the crRNA 5´half ("seed" sequence region) are sufficient to abolish the interference (Manica *et al*., submitted). These findings support the observation made by Semenova and co-workers (Semenova *et al*. 2011) in *E. coli*, where the first 8nt of the crRNA 5´half (seed) were crucial for crRNA-target recognition.

5.8 Conclusion

Viruses, plasmids and other mobile genetic elements have been described in most archaea, but occur in extraordinary numbers and diversity in members of the genus *Sulfolobus*. It might therefore not be too surprising that species of this genus also harbor extensive and diverse CRISPR loci that help to defend against or cope with these elements. While it seems *ad hoc* unfavourable to study specific functions of the archaeal immune system in such a complex genomic context, the studies made in *Sulfolobus* clearly demonstrate the advantages of this complexity: the activity of clusters from different CRISPR families can be directly compared and their interactions studied, as demonstrated for spacer recruitment, recombination of loci, CASCADE complex formation and crRNA sorting. Specificity of proteins from different CRISPR families can be investigated in parallel and both DNA degradation (so far demonstrated only *in vivo*) and RNA degradation (so far only demonstrated *in vitro*) of invading nucleic acids can be studied in the same system. It will be important to obtain more mechanistic details on these functions in the future and to analyse the ecological relevance of the virus defense system for the host organisms and their invading genetic elements. Equally important will be the investigation of CRISPR systems in other genetically and/or biochemically tractable archaea in order to obtain a comprehensive picture of the function of diverse and abundant CRISPR RNAs in archaea.

5.9 References

Agari, Y., K. Sakamoto, *et al.* (2009). "Transcription Profile of Thermus thermophilus CRISPR Systems after Phage Infection." J Mol Biol.

Alvarado, V. and H. B. Scholthof (2009). "Plant responses against invasive nucleic acids: RNA silencing and its suppression by plant viral pathogens." Semin Cell Dev Biol **20**(9): 1032-1040.

Babu, M., N. Beloglazova, *et al.* (2011). "A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair." Mol Microbiol **79**(2): 484-502.

Barrangou, R., C. Fremaux, *et al.* (2007). "CRISPR provides acquired resistance against viruses in prokaryotes." Science **315**(5819): 1709-1712.

Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." Cell **136**(2): 215-233.

Bernstein, E., A. A. Caudy, *et al.* (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." Nature **409**(6818): 363-366.

Bolotin, A., B. Quinquis, *et al.* (2005). "Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin." Microbiology **151**(Pt 8): 2551-2561.

Bondy-Denomy, J., A. Pawluk, *et al.* (2013). "Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system." Nature **493**(7432): 429-432.

Brodersen, P. and O. Voinnet (2006). "The diversity of RNA silencing pathways in plants." Trends Genet **22**(5): 268-280.

Brouns, S. J., M. M. Jore, *et al.* (2008). "Small CRISPR RNAs guide antiviral defense in prokaryotes." Science **321**(5891): 960-964.

Brugger, K., E. Torarinsson, *et al.* (2004). "Shuffling of Sulfolobus genomes by autonomous and non-autonomous mobile elements." Biochem Soc Trans **32**(Pt 2): 179-183.

Cady, K. C., J. Bondy-Denomy, *et al.* (2012). "The CRISPR/Cas adaptive immune system of Pseudomonas aeruginosa mediates resistance to naturally occurring and engineered phages." J Bacteriol **194**(21): 5728-5738.

Carte, J., R. Wang, *et al.* (2008). "Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes." Genes Dev **22**(24): 3489-3496.

Cavicchioli, R. (2007). Archaea: molecular and cellular biology, ASM Press.

Datsenko, K. A., K. Pougach, *et al.* (2012). "Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system." Nat Commun **3**: 945.

Deltcheva, E., K. Chylinski, *et al.* (2011). "CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III." Nature **471**(7340): 602-607.

Deng, L., R. A. Garrett, *et al.* (2013). "A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus." Mol Microbiol.

Deng, L., C. S. Kenchappa, *et al.* (2011). "Modulation of CRISPR locus transcription by the repeat-binding protein Cbp1 in Sulfolobus." Nucleic Acids Res **40**(6): 2470-2480.

Erdmann, S. and R. A. Garrett (2012). "Selective and hyperactive uptake of foreign DNA by adaptive immune systems of an archaeon via two distinct mechanisms." Mol Microbiol **85**(6): 1044-1056.

Fire, A., S. Xu, *et al.* (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-811.

Garrett, R. A. and H. P. Klenk (2007). Archaea: evolution, physiology, and molecular biology, Blackwell.

Garrett, R. A., S. A. Shah, *et al.* (2011). "CRISPR-based immune systems of the Sulfolobales: complexity and diversity." Biochem Soc Trans **39**(1): 51-57.

Garrett, R. A., G. Vestergaard, *et al.* (2011). "Archaeal CRISPR-based immune systems: exchangeable functional modules." Trends Microbiol **19**(11): 549-556.

Ghildiyal, M. and P. D. Zamore (2009). "Small silencing RNAs: an expanding universe." Nat Rev Genet **10**(2): 94-108.
Godde, J. S. and A. Bickerton (2006). "The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes." J Mol Evol **62**(6): 718-729.
Grissa, I., G. Vergnaud, *et al.* (2007). "The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats." BMC Bioinformatics **8**: 172.
Gudbergsdottir, S., L. Deng, *et al.* (2011). "Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers." Mol Microbiol **79**(1): 35-49.
Haft, D. H., J. Selengut, *et al.* (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.
Hale, C. R., P. Zhao, *et al.* (2009). "RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex." Cell **139**(5): 945-956.
Hamilton, A. J. and D. C. Baulcombe (1999). "A species of small antisense RNA in posttranscriptional gene silencing in plants." Science **286**(5441): 950-952.
Han, D. and G. Krauss (2009). "Characterization of the endonuclease SSO2001 from Sulfolobus solfataricus P2." FEBS Lett **583**(4): 771-776.
Han, D., K. Lehmann, *et al.* (2009). "SSO1450--a CAS1 protein from Sulfolobus solfataricus P2 with high affinity for RNA and DNA." FEBS Lett **583**(12): 1928-1932.
Haurwitz, R. E., S. H. Sternberg, *et al.* (2012). "Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA." EMBO J **31**(12): 2824-2832.
Held, N. L., A. Herrera, *et al.* (2010). "CRISPR associated diversity within a population of Sulfolobus islandicus." PLoS One **5**(9).
Held, N. L. and R. J. Whitaker (2009). "Viral biogeography revealed by signatures in Sulfolobus islandicus genomes." Environ Microbiol **11**(2): 457-466.
Huttenhofer, A., P. Schattner, *et al.* (2005). "Non-coding RNAs: hope or hype?" Trends Genet **21**(5): 289-297.
Hyun, T. K., M. N. Uddin, *et al.* (2010). "Cell-to-cell trafficking of RNA and RNA silencing through plasmodesmata." Protoplasma **248**(1): 101-116.
Iverson, E. and K. Stedman (2012). "A genetic study of SSV1, the prototypical fusellovirus." Front Microbiol **3**: 200.
Jansen, R., J. D. van Embden, *et al.* (2002). "Identification of a novel family of sequence repeats among prokaryotes." OMICS **6**(1): 23-33.
Jore, M. M., M. Lundgren, *et al.* (2011). "Structural basis for CRISPR RNA-guided DNA recognition by Cascade." Nat Struct Mol Biol **18**(5): 529-536.
Koerdt, A., A. Orell, *et al.* (2011). "Macromolecular fingerprinting of sulfolobus species in biofilm: a transcriptomic and proteomic approach combined with spectroscopic analysis." J Proteome Res **10**(9): 4105-4119.
Kunin, V., R. Sorek, *et al.* (2007). "Evolutionary conservation of sequence and secondary structures in CRISPR repeats." Genome Biol **8**(4): R61.
Lillestol, R. K., P. Redder, *et al.* (2006). "A putative viral defence mechanism in archaeal cells." Archaea **2**(1): 59-72.
Lillestol, R. K., S. A. Shah, *et al.* (2009). "CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties." Mol Microbiol **72**(1): 259-272.
Lintner, N. G., K. A. Frankel, *et al.* (2011). "The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system." J Mol Biol **405**(4): 939-955.

Lintner, N. G., M. Kerou, *et al.* (2011). "Structural and functional characterization of an archaeal CASCADE complex for CRISPR-mediated viral defense." J Biol Chem.

Maaty, W. S., J. D. Steffens, *et al.* (2012). "Global analysis of viral infection in an archaeal model system." Front Microbiol **3**: 411.

Makarova, K. S., L. Aravind, *et al.* (2011). "Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems." Biol Direct **6**: 38.

Makarova, K. S., N. V. Grishin, *et al.* (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., D. H. Haft, *et al.* (2011). "Evolution and classification of the CRISPR-Cas systems." Nat Rev Microbiol **9**(6): 467-477.

Manica, A., Z. Zebec, *et al.* (2011). "In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon." Mol Microbiol **80**(2): 481-491.

Marchfelder, A., S. Fischer, *et al.* (2012). "Small RNAs for defence and regulation in archaea." Extremophiles **16**(5): 685-696.

Marraffini, L. A. and E. J. Sontheimer (2008). "CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA." Science **322**(5909): 1843-1845.

Marraffini, L. A. and E. J. Sontheimer (2010). "CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea." Nat Rev Genet **11**(3): 181-190.

Marraffini, L. A. and E. J. Sontheimer (2010). "Self versus non-self discrimination during CRISPR RNA-directed immunity." Nature **463**(7280): 568-571.

Matera, A. G., R. M. Terns, *et al.* (2007). "Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs." Nat Rev Mol Cell Biol **8**(3): 209-220.

Mojica, F. J., C. Diez-Villasenor, *et al.* (2009). "Short motif sequences determine the targets of the prokaryotic CRISPR defence system." Microbiology **155**(Pt 3): 733-740.

Peng, X., K. Brugger, *et al.* (2003). "Genus-specific protein binding to the large clusters of DNA repeats (short regularly spaced repeats) present in Sulfolobus genomes." J Bacteriol **185**(8): 2410-2417.

Perez-Rodriguez, R., C. Haitjema, *et al.* (2011). "Envelope stress is a trigger of CRISPR RNA-mediated DNA silencing in Escherichia coli." Mol Microbiol **79**(3): 584-599.

Pina, M., A. Bize, *et al.* (2011). "The archeoviruses." FEMS Microbiol Rev **35**(6): 1035-1054.

Plagens, A., B. Tjaden, *et al.* (2012). "Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon Thermoproteus tenax." J Bacteriol **194**(10): 2491-2500.

Pul, U., R. Wurm, *et al.* (2010). "Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS." Mol Microbiol **75**(6): 1495-1512.

Richter, H., J. Zoephel, *et al.* (2012). "Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis." Nucleic Acids Res **40**(19): 9887-9896.

Rousseau, C., M. Gonnet, *et al.* (2009). "CRISPI: a CRISPR interactive database." Bioinformatics **25**(24): 3317-3318.

Semenova, E., M. M. Jore, *et al.* (2011). "Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence." Proc Natl Acad Sci U S A **108**(25): 10098-10103.

Shah, S. A. and R. A. Garrett (2011). "CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems." Res Microbiol **162**(1): 27-38.
Shah, S. A., N. R. Hansen, *et al.* (2009). "Distribution of CRISPR spacer matches in viruses and plasmids of crenarchaeal acidothermophiles and implications for their inhibitory mechanism." Biochem Soc Trans **37**(Pt 1): 23-28.
She, Q., C. Zhang, *et al.* (2009). "Genetic analyses in the hyperthermophilic archaeon Sulfolobus islandicus." Biochem Soc Trans **37**(Pt 1): 92-96.
Sinkunas, T., G. Gasiunas, *et al.* (2011). "Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system." EMBO J **30**(7): 1335-1342.
Storz, G., J. Vogel, *et al.* (2011). "Regulation by small RNAs in bacteria: expanding frontiers." Mol Cell **43**(6): 880-891.
Swarts, D. C., C. Mosterd, *et al.* (2012). "CRISPR interference directs strand specific spacer acquisition." PLoS One **7**(4): e35888.
Tang, T. H., N. Polacek, *et al.* (2005). "Identification of novel non-coding RNAs as potential antisense regulators in the archaeon Sulfolobus solfataricus." Mol Microbiol **55**(2): 469-481.
Tang, T. H., T. S. Rozhdestvensky, *et al.* (2002). "RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing." Nucleic Acids Res **30**(4): 921-930.
Tolia, N. H. and L. Joshua-Tor (2007). "Slicer and the argonautes." Nat Chem Biol **3**(1): 36-43.
Torarinsson, E., H. P. Klenk, *et al.* (2005). "Divergent transcriptional and translational signals in Archaea." Environ Microbiol **7**(1): 47-54.
Wagner, M., S. Berkner, *et al.* (2009). "Expanding and understanding the genetic toolbox of the hyperthermophilic genus Sulfolobus." Biochem Soc Trans **37**(Pt 1): 97-101.
Waters, L. S. and G. Storz (2009). "Regulatory RNAs in bacteria." Cell **136**(4): 615-628.
Westra, E. R., U. Pul, *et al.* (2010). "H-NS-mediated repression of CRISPR-based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO." Mol Microbiol **77**(6): 1380-1393.
Wiedenheft, B., E. van Duijn, *et al.* (2011). "RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions." Proc Natl Acad Sci U S A **108**(25): 10092-10097.
Wurtzel, O., R. Sapra, *et al.* (2009). "A single-base resolution map of an archaeal transcriptome." Genome Res **20**(1): 133-141.
Yosef, I., M. G. Goren, *et al.* (2012). "Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli." Nucleic Acids Res.
Zhang, J., T. Kasciukovic, *et al.* (2012). "The CRISPR associated protein Cas4 Is a 5' to 3' DNA exonuclease with an iron-sulfur cluster." PLoS One **7**(10): e47232.
Zhang, J., C. Rouillon, *et al.* (2012). "Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity." Mol Cell **45**(3): 303-313.
Zhou, Y. and J. Xie (2010). "The roles of pathogen small RNAs." J Cell Physiol **226**(4): 968-973.

6 *IN VIVO* ACTIVITY OF CRISPR-MEDIATED VIRUS DEFENSE IN A HYPERTHERMOPHILIC ARCHAEON

Andrea Manica, Ziga Zebec, Daniela Teichmann, and Christa Schleper*

University of Vienna, Department of Genetics in Ecology, Althanstr. 14, 1090-Vienna, Austria

*corresponding author:
Christa Schleper, e-mail: christa.schleper@univie.ac.at
Phone: +43 1 4277 57800, Fax: +43 1 4277 9578

6.1 Summary

CRISPR/Cas systems are found widespread in bacterial and archaeal genomes and exhibit considerable diversity. However, closer insights into the action of most of the CRISPR modules have remained elusive in particular in Archaea due to the lack of suitable *in vivo* test systems.

Here we demonstrate CRISPR/Cas-based immune defense in the hyperthermophilic archaeon *Sulfolobus solfataricus*. Recombinant variants of the SSV1 virus containing a gene of the conjugative plasmid pNOB8 that represents a target for a corresponding CRISPR spacer in the chromosome were tested in transfection experiments. Almost 100% immunity against the recombinant virus was observed when the chromosomal CRISPR spacer matched perfectly to the protospacer. Different from bacterial systems immunity was still detected, albeit at decreased levels, when mutations distinguished target and spacer. CRISPR/Cas targeting was independent of the transcription of the target gene. Furthermore, a mini-CRISPR locus introduced on the viral DNA with spacers targeting the (non-essential) chromosomal beta-galactosidase gene was unstable in host cells and triggered recombination with the indigenous CRISPR locus.

Our experiments demonstrate *in vivo* activity of CRISPR/Cas in archaea for the first time and suggest that - unlike the recently demonstrated *in vitro* cleavage of RNA in *Pyrococcus* - DNA is targeted in this archaeon.

6.2 Introduction

Sequence analyses of complete bacterial and archaeal genomes have led to the discovery of Clustered Regularly Interspaced Short Palindromic Repeats (in short CRISPR) (Jansen *et al.*, 2002). The potential function of these repeats and their intervening short spacer sequences as well as the function of their associated (Cas-) proteins as constituents of an immune defense system against viruses and other genetic elements, has only recently been recognized (for reviews see van der Oost *et al.*, 2009; Horvath & Barrangou, 2010; Marraffini and Sontheimer, 2010; Karginov and Hannon, 2010, Deveau *et al.*, 2010). The finding, that some of the intervening spacers match to viral and plasmid sequences (Mojica *et al.*, 2005; Pourcel *et al.*, 2005; Bolotin *et al.*, 2005) triggered the initial hypothesis that a nucleic-acid mediated immune function, reminiscent of the RNA interference mechanism of eukaryotes might be active in bacteria and archaea (Makarova *et al.*, 2006). *In vivo* activity of this system was demonstrated in *Streptococcus thermophilus* by Barrangou *et al.* (Barrangou *et al.*, 2007). Upon exposure to bacteriophages, incorporation of spacers with homology to the phage genome into the host chromosome occurred, which in turn led to immunity against the invader. Later it was shown by Maraffini and Sontheimer (Marraffini and Sontheimer, 2008) that CRISPR-mediated interference by the Cas protein family present in *Staphylococcus epidermidis* targets DNA and not RNA. In *S. thermophilus* rapid cleavage of invading double-stranded DNA was demonstrated *in vivo* (Garneau *et al.*, 2010). Using *Escherichia coli* and bacteriophage lambda, Brouns *et al.* (Brouns *et al.*, 2008) studied the activity of the Cas proteins that cleave the CRISPR RNA precursors in the repeat sequences, as a prerequisite of the activity of mature CRISPR RNAs in interference with virus DNA. While test systems to study CRISPR/Cas-mediated immunity *in vivo* have been established for these three bacteria and progress has been made in the

biochemical characterization of CRISPR/Cas in archaea (Carte *et al.*, 2008; Beloglazova *et al.*, 2008; Hale *et al.*, 2009; Han *et al.*, 2009; Gudbergsdottir *et al.,* 2011) no assay for directly studying *in vivo* activity exists yet for archaea. However, CRISPR/Cas loci are found in the majority of archaeal genomes compared to about half of the sequenced bacterial genomes (http://crispi.genouest.org/) (Godde and Bickerton, 2006; Grissa *et al.*, 2007; Rousseau *et al.*, 2009, Karginov and Hannon, 2010), indicating that it plays an important role in this domain. This is supported by the fact that CRISPR loci of archaea tend to be larger and occur in greater numbers than in bacterial genomes (Lillestol *et al.*, 2006). They have been shown to constitute up to 1% of the chromosome in a methanogenic archaeon (Lillestol *et al.*, 2006). Furthermore, Archaea, in particular all crenarchaeota, (the archaeal kingdom to which *S. solfataricus* belongs), have unique and very special viruses, many of which are not related to invading genetic elements of the bacterial or eukaryotic world (Prangishvili *et al.*, 2006; Bize *et al.*, 2009; Haring *et al.*, 2005). The central DNA- and RNA-related informational processing systems of archaea (e.g. replication, transcription, DNA binding proteins etc.) are fundamentally different from bacteria and rather show similarities to the eukaryotic systems (Garrett and Klenk, 2006), which make it even more relevant to study virus defense of archaea. Although the distribution of the different types of CRISPR/Cas systems in the bacteria and archaea suggests that these are transferred horizontally among organisms within and between the two domains (Makarova *et al.*, 2006; Godde and Bickerton, 2006; Chakraborty *et al.*, 2010) many CRISPR-associated proteins in archaea are highly divergent from their bacterial counterparts (Haft *et al.*, 2005) as are the promoter structures in the CRISPR leader sequences (Lillestol *et al.*, 2009). Interestingly, in the hyperthermophilic archaeon *Pyrococcus furiosus*, Hale *et al.* (2009) recently demonstrated site-specific cleavage of target RNAs *in vitro* that was dependent on complementarity of CRISPR-derived small RNAs and on the Cas RAMP module (Cmr) proteins. Their findings suggest that an RNA silencing mechanism, similar to the eukaryotic RNAi system functions in invader defense in this archaeon and

perhaps in other organisms that have the Cmr module of *cas* genes (Hale *et al.*, 2008; van der Oost and Brouns, 2009). However it is not clear, whether other archaeal systems are targeting RNA or DNA, i.e. whether the different activities correlate with the type of organisms (bacterial versus archaeal), the type of test systems (*in vivo* or *in vitro*) or the types of CRISPR/Cas modules that have been studied.

In order to explore CRISPR/Cas-mediated activities in archaea, we have developed an *in vivo* system for the hyperthermophilic crenarchaeon *Sulfolobus solfataricus* and its virus SSV1. This well-studied virus is non-lytic and replicates its circular 15 kb DNA as a plasmid in the cells, but also integrates site-specifically into the host chromosome (Martin *et al.*, 1984; Reiter *et al.*, 1989; Schleper *et al.*, 1992). *S. solfataricus* grows optimally at around 80°C and pH 3 and transfection/infection and plaque assays have been optimized for this virus-host system (Schleper *et al.*, 1992). Additionally, a recombinant SSV1-based shuttle vector has been established for *in vivo* gene expression studies (Jonuscheit *et al.*, 2003). *S. solfataricus* P2 contains a complex array of CRISPR/Cas loci, with six different CRISPR repeat arrays, and different sets of (more or less) associated *cas* genes (Fig. 1).

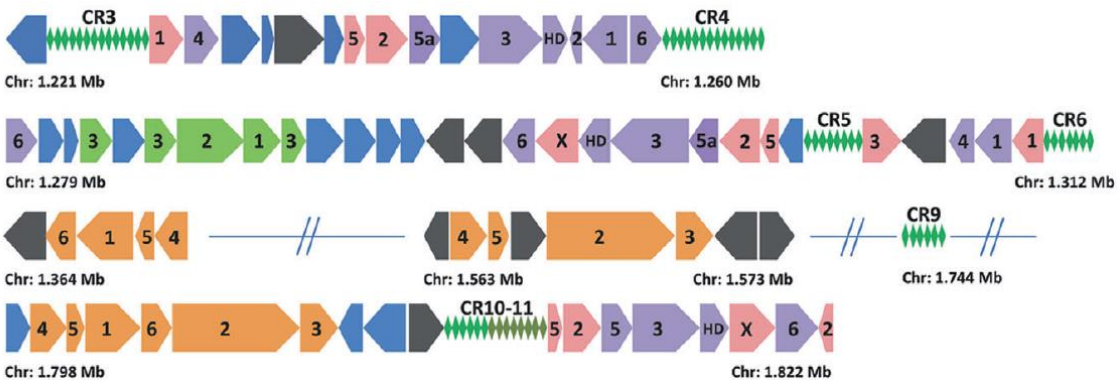


Fig.1: **Schematic representation of the different CRISPR/Cas loci in *S. solfataricus P2*.** Different colors report different categories of CRISPR-associated proteins (Cas) classified according to Haft *et al.* (2005), purple: "core" *cas* genes (numbers representing *cas*1 through *cas*6), pink: *Apern subtype,* (*csa*1 through *csa*5a and *csa*X), green: *cas* genes with high similarity to *Mtube subtype* (*csm*1 through *csm*3), orange: Cmr Module *(cmr*1 through *cmr*6), blue: proteins annotated as CRISPR-related proteins (UCSC Archaeal Genome Browser http://archaea.ucsc.edu/) but with no or low similarity to other proteins with known function in the databases, grey: IS elements and putative transposase-related genes in the proximity of the CRISPR/Cas loci. The CRISPR locus names refer to the UCSC Archaeal Genome browser annotation.

One of the *cas*-gene sets, located in the proximity of CRISPR loci 3 and 4 (UCSC Archaeal Genome Browser database) has been classified to belong to *Sulfolobus* CRISPR family II, according to Lillestol *et al*. (2009). The CRISPR repeat areas have 102 (CR3) and 94 (CR4) spacers, respectively. These two CRISPR sets differ from the closely related isolate *S. solfataricus* P1 by their number and nature of spacers, in particular at the end located towards the leader sequences (Lillestol *et al*., 2009). Only 52 spacers in CR3 and 35 spacers in CR4 are identical to those in strain P1. These differences indicate that the locus is actively modified by the acquisition of novel spacers, which are preferentially incorporated at the side next to the leader-sequence in bacteria (Barrangou *et al*., 2007; Deveau *et al*., 2008).

Here, we demonstrate CRISPR-mediated defense against a recombinant SSV1 virus in *S. solfataricus* P2, based on a CRISPR spacer of CR3-4 that is not present in the closely related strain P1. Furthermore, we show that a CRISPR module carried on an extrachromosomal element is not tolerated by the host cells when it contains spacers (in sense or antisense orientation) that target a non-essential gene in the chromosome, but that it is tolerated when the module contains unrelated sequences. Our results imply that the CRISPR/Cas defense mechanism involves attack of invader DNA in *S. solfataricus*.

6.3 Results

6.3.1 CRISPR-mediated immunity against recombinant SSV1 viruses

The 37 nt long spacer on the 53rd position in the CRISPR locus CR3 of *S. solfataricus* P2 (see Fig. 1 and Fig. S1) has high similarity (30/37 nt) to a region in ORF406 of the conjugative plasmid pNOB8 (She *et al.*, 1998), while the closely related strain *S. solfataricus* P1 does not contain this spacer. Expression of the spacer under various growth conditions and growth phases was confirmed by Northern analyses using a complementary probe specific to

the spacer sequence (suppl. material Fig. S2). Upon transcription from the leader sequence of CR3, the RNA of spacer53 is complementary (with 7 mismatches) to the transcript of ORF406. The DNA sequence of this open reading frame (ORF) was amplified from pNOB8 DNA and cloned into the *E.coli*/S.*solfataricus* shuttle vector pMJ0305 (Fig. 2A) to yield construct 406-7M in Fig.2B) under the control of a promoter of the thermophilic factor TF55 (alpha subunit of the major chaperonin of *S. solfataricus*). This promoter has a strong basic activity even without heat induction (Jonuscheit *et al*., 2003). In addition, a mutated version of ORF406 perfectly complementary to spacer53 was cloned (406-0M) as well as two other versions that had a 1 bp deletion in their center with respect to the spacer sequence (406-1D, see Fig. 2B) and 3 bp mutations at positions found in the wild-type gene at the right end of the protospacer (406-3M in Fig. 2B), respectively.

A fifth construct (406-0MnoP) was made, again with the gene perfectly complementary to the spacer region but without any preceding promoter sequence. These constructs were used to transfect the two strains, *S. solfataricus* P1 and P2, respectively, and transfection efficiencies were determined in plaques assays (Schleper *et al*., 1992) using cells of the respective strain (P1 or P2) in overlays. While equally high transfection efficiencies were obtained in strain P1 that does not contain the ORF406-specific spacer (see table 1), the transfection efficiency in P2 was severely reduced when the spacer matched the protospacer (target) sequence of ORF406.

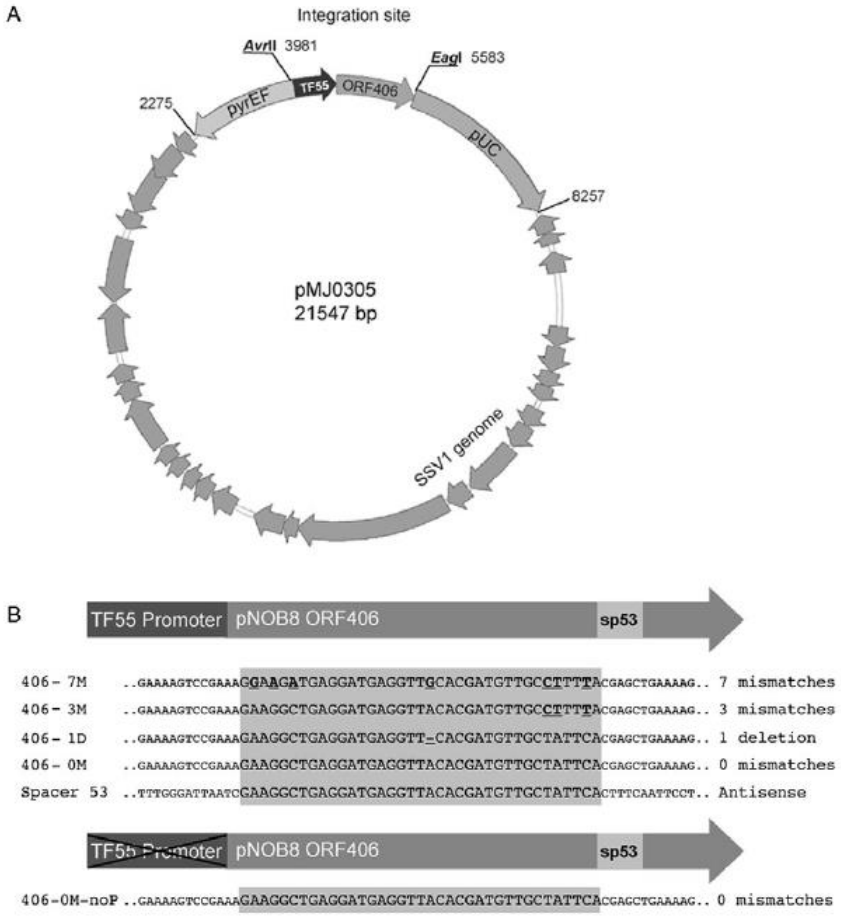


Fig.2: Overview of recombinant shuttle vector and protospacer constructs.
A) E.*coli*/S.*solfataricus* shuttle vector pMJ0305, based on the complete sequence of virus SSV1 of *S. shibatae*, the *pyr*EF genes of *S. solfataricus*, pUC18 of *E. coli* and an insertion site, into which the different gene variants and mini-CRISPR of this study have been incorporated.
B) DNA sequences (protospacers) of ORF406 that have been varied in the different constructs (highlighted in bold and underlined are the mismatches and deletion compared to the spacer sequence). The construct 406-0MnoP is identical to 406-0M, but without the TF55 promoter in front of the gene. The grey shaded region is similar/identical to the sequence of spacer 53; the flanking nucleotides in ORF 406 do not show a PAM according to Mojica *et al.* (2009) and Kunin *et al.* (2007). Nucleotides depicted on the left and right side to the spacer53 represent the repeat region of the locus.

Between two to three orders of magnitude less plaques were obtained (per µg of DNA) for the 406-variants that were complementary to the spacer (406-0M and 406-0MnoP), as compared to transfection efficiencies obtained with the 7 mismatch containing ORF (406-7M). Interestingly, even transfection efficiencies of the latter (406-7M) were still slightly lower than those of the recombinant vector without protospacer (pMJ0305). The transfection efficiencies of the 1bp-deletion (406-1D) and the 3 mismatch construct (406-3M) were also severely reduced (by about 85%, see table 1), but not as much

as the perfectly complementary constructs (406-0M and 406-0M-noP). Similar to the other CRISPR/Cas pathways studied in bacteria and archaea (Barangou *et al.*, 2007; Brouns *et al.*, 2008; Hale *et al.*, 2009), our results give thus evidence for dependence of interference on effector RNA produced from the spacer, because mutations in the target DNA (ORF406) decreased the effect.

Strain	Construct	Average*	SD§
P2	pMJ03	1	0
	406-7M	8.21^{-1}	1.80^{-2}
	406-3M	1.67^{-1}	1.58^{-1}
	406-1D	1.32^{-1}	1.07^{-1}
	406-0M	2.46^{-3}	4.51^{-3}
	406-0Mno-P	5.40^{-4}	9.36^{-4}
P1	pMJ03	9.92^{-1}	7.17^{-3}
	406-7M	9.79^{-1}	2.89^{-2}
	406-3M	9.21^{-1}	1.08^{-1}
	406-0M	9.76^{-1}	4.11^{-2}
	406-0Mno-P	8.83^{-1}	5.25^{-2}

* >5 independent transfection experiments for each construct and strain

§ Standard deviation

Tab.1: Efficiencies of plating for recombinant SSV1 virus with different protospacer inserts.

6.3.2 CRISPR-mediated immunity is independent of protospacer transcription

Both constructs, 406-0M and 406-0M-noP, yielded comparable low transfection efficiencies, showing that immunity was independent of the presence of a promoter sequence in front of the target gene ORF406 (Table 1). In order to compare the amount of transcripts of both constructs in the cells, single plaques of transfected *S. solfataricus* P1 cells were picked and inoculated and samples were taken from exponential and stationary growth phase, respectively. The presence and comparable abundance of the plasmid in the transfectants was verified by semi-quantitative PCR (not shown). As expected, there was no signal for ORF406 mRNA in the transfectant of 406-

0MnoP detectable whereas a strong band hybridized from the promoter-containing construct (Fig. 3). This means that CRISPR mediated interference most probably targeted directly DNA of the protospacer (or alternatively very low, i.e. undetectable amounts of mRNA).

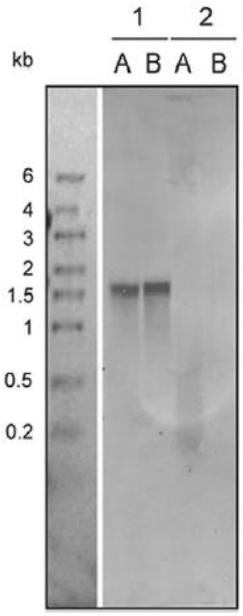


Fig.3: Northern blot with RNA of transfectants.
Five µg of RNA of *S. solfataricus* P1 in exponential (A) and stationary (B) growth phase, transfected with the construct 406-0M (1) and 406-0MnoP (2) were loaded on a denaturing 1.2% agarose gel, followed by transfer to a nylon membrane. Hybridization was performed with a DNA probe specific for ORF406. RNA size marker was co-electrophoresed and blotted, and excised from the membrane before the washing procedure.

6.3.3 Visualizing immunity in plaque assays

Plaques assays were routinely performed for quantification of transfectants (as shown in table 1). In Fig. 4 we display a different plaque experiment in which the two *Sulfolobus* strains were used to directly visualize the defense acting in a 'natural' infection of virus particles instead of acting upon incoming DNA that was artificially introduced by electroporation. For this purpose strain P1 (without matching spacer) was transfected with recombinant SSV1 virus and the transfected cells were plated on a lawn of the spacer-containing strain P2 that exhibited immune response against construct 406-0M but not against 406-7M (as shown in table 1). Big turbid plaques for virus 406-7M can be clearly seen, that are indistinguishable in

appearance from those of the wild-type virus SSV1 (Schleper *et al.*, 1992). The same amount of transfected P1 cells was plated for construct 406-0M on strain P2 in Fig.4B. However, plaques were barely visible, more turbid, formed only a tiny clear spot in the inner center and exhibited halos of variable sizes. This demonstrates that growth of the lawn cells was only affected in the very center of the plaque, where the density of expelled virus particles is highest.

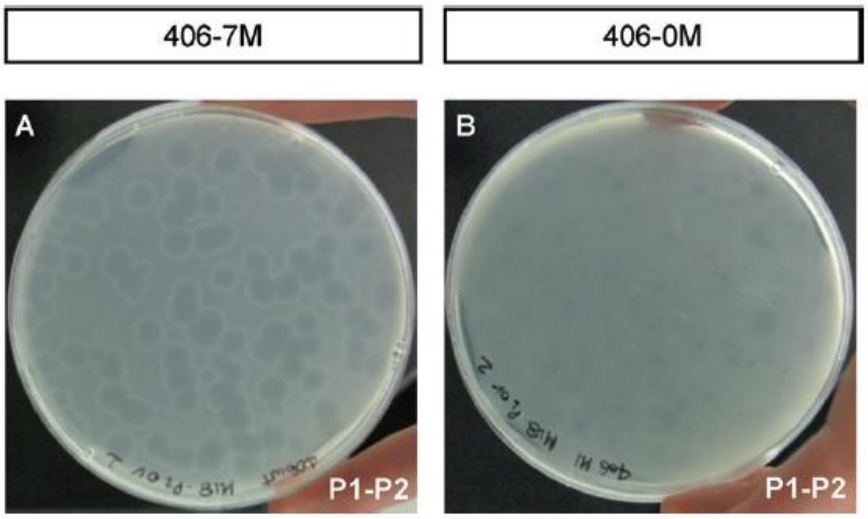


Fig.4: Effects of CRISPR/Cas mediated immunity on plaque morphologies.
Plaques were obtained after plating cells of *S. solfataricus* P1 transfected with virus construct 406-0M (B) and 406-7M (A) on lawns of cells of *S. solfataricus* P2 that contains the complementary spacer in locus CR3-4.

6.3.4 Virus-encoded CRISPR locus targeting the chromosome

We have constructed an artificial mini-CRISPR locus that was cloned into the recombinant SSV1-shuttle vector (Fig. 2A). It was originally meant for investigating if a CRISPR/Cas-based RNA-targeted effect can be obtained by silencing the transcription of a chromosomal gene. The mini-CRISPR contained the leader sequence, four spacers (grey arrows labeled 1 through 4 in Fig 5A) and repeats (light grey boxes), as well as the downstream sequence of CRISPR locus CR4 of *S. solfataricus* P1. Additionally three spacers, each separated by the respective repeats of the locus were introduced that were complementary to the mRNA of the indigenous non-essential beta-

galactosidase gene of *S. solfataricus* (a1 through a3 in Fig. 5A, construct LAT). Two additional spacers with random sequence (x1, x2) flanking these, were part of the mini-CRISPR. In a second construct, the beta galactosidase-targeting spacers were cloned in sense direction, such that their transcripts would not be able to pair with the mRNA of the gene (Fig. 5A, construct LST, spacers s1 through s3). A third control construct contained the mini-CRISPR with one random spacer, but without any beta-galactosidase-targeting spacers (LM in Fig. 5A). Upon transfection of *S. solfataricus* M18 (*pyr*EF mutant of strain P1, Martusewitsch *et al.* 2000) cells were inoculated into liquid media and grown for three days in order to allow spreading of the virus throughout the culture by infection of viruses produced from the primary transfectants. We were unable to detect any significant effect of introduction of construct LAT relative to LST or LM on beta-galactosidase activity in transformed cells. However, compared to control cultures harbouring other shuttle-vectors, the cells with constructs LAT and LST grew considerably slower (delayed by several days or even 2 weeks) when transferred into selective medium, i.e. to media without uracil, selecting for the presence of the recombinant vector with *pyr*EF genes (Jonuscheit *et al.*, 2003). In order to analyse transfectants, the transformation mixture was plated and single colonies were isolated. These were grown in two successive rounds in selective media of which DNA samples were prepared. PCR analyses using primers targeting sequences of the leader and terminator of the chromosomal (and plasmid-encoded) CRISPR locus (Fig. 5B) or targeting specifically the mini-CRISPR on the shuttle vector (Fig. 5C, D), respectively, were performed. Fig. 5B and C show PCR products obtained from slowly growing transformants of a first selection round (lanes 3 through 8 in 5B and 1 through 6 in 5C), and products of five of these cultures after a second transfer into selective media, in which the cultures had resumed regular growth rates (lanes 9 through 13 and lanes 7 through 11, respectively). A series of mostly smaller fragments than those expected for the introduced mini-CRISPR (control lane 12 in Fig. 5C) were amplified, indicating that the locus on the introduced viral DNA apparently underwent considerable changes, probably

due to recombination events. Fragments of different sizes appeared even within the same cultures (Fig. 5C lanes 1 through 7 and 10). Especially in the 2nd selection round a larger 5 kb fragment dominated, that had already appeared in the cultures of the 1st selection round.

By contrast, construct LM, that did not contain spacers complementary to the chromosomal beta-galactosidase gene did not exhibit any changes in its CRISPR locus neither after the 1st nor after successive selection rounds (Fig. 5D). Also, when the constructs LAT and LST were used to transform strain PBL2025, that lacks the chromosomal beta-galactosidase gene (Schelert *et al.*, 2004), they appeared stable and no recombination was observed (data not shown).

In order to analyse the changed virus DNA of LAT and LST transformants that had apparently all undergone changes, extrachromosomal DNA was isolated from 20 cultures of the second selection round and was used to transform *E. coli*. Two colonies of each of the 12 successful transformations were picked and plasmids analysed. 21 of these showed the same 5 kb insert as found in the transformants of *S. solfataricus* (Fig. 5C, lanes 1 through 5, 9, 10) and the same characteristic restriction pattern (Suppl. Fig.3A and B). Two did not contain plasmid and one showed a different pattern. Three plasmids (of the same RFLP pattern) were sequenced. The DNA of the CRISPR locus encoded on these plasmids was identical to the sequence of the native CRISPR locus CR04 (DQ831675) that resides in the chromosome of *Sulfolobus solfataricus* P1. The size of this locus matched exactly the size of the bigger PCR product in Fig. 5C (lanes 9 and 10). Thus, it appeared that the presence of the mini-CRISPR locus was deleterious for the cell and resulted in slower growth rates and recombination events. The cells that recovered to normal growth rates were those that had undergone recombination with the chromosomally-located CRISPR locus. Other plasmids with smaller inserts (Fig. 5C, lanes 1-5, 7) were apparently not replicable in *E. coli*, as they have not been recovered by retransformation. The exchange of the artificial mini-CRISPR locus of the plasmid with the

chromosomally located locus was possible because the up- and downstream flanking regions of the repeat structure were identical.

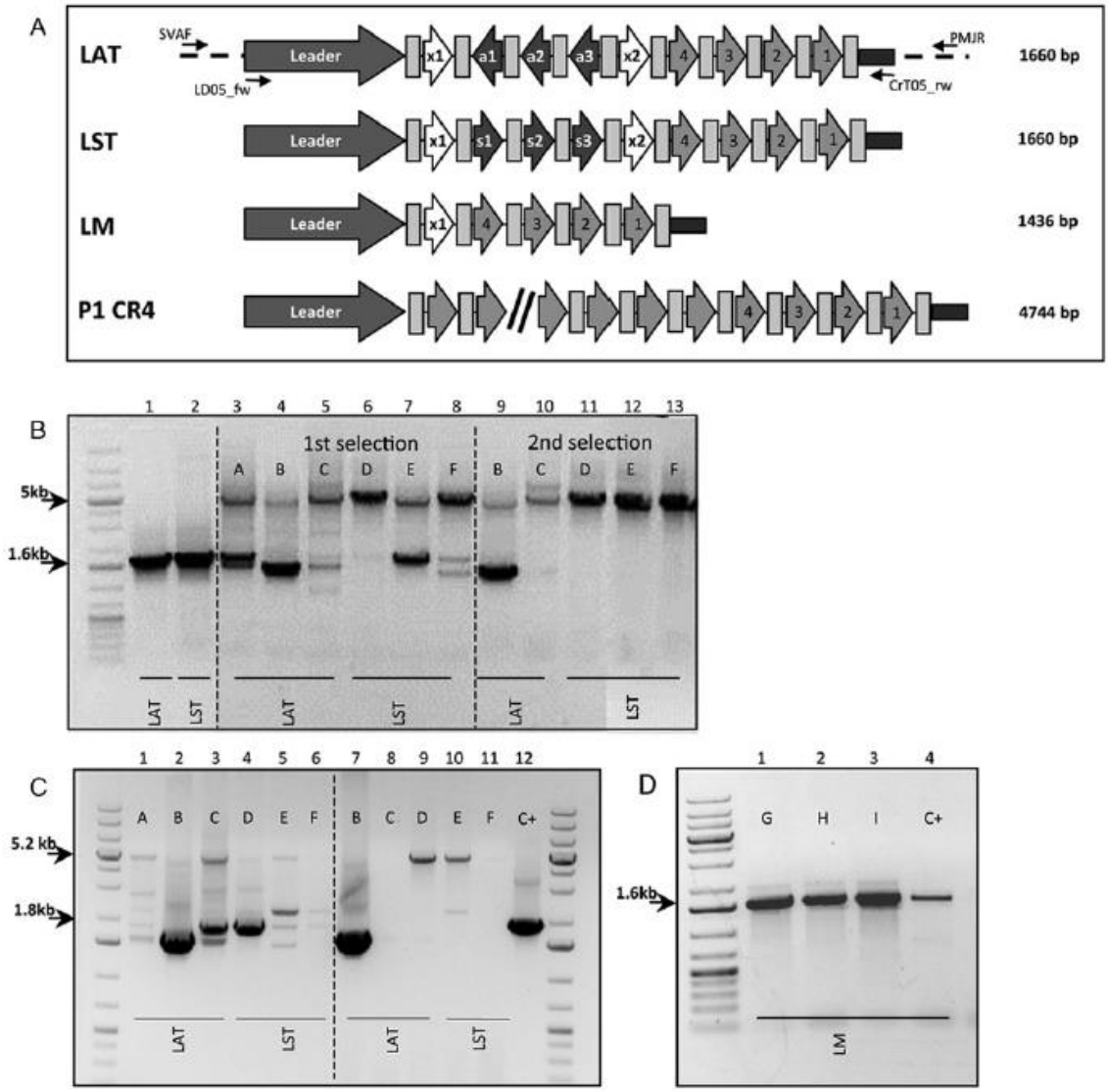


Fig.5: Transfection of *S. solfataricus* M18 (*pyr*EF mutant of P1) with mini CRISPR-locus targeting a non-essential gene in the chromosome.

A) Scheme of the artificial locus that contains three spacers each, directed against the beta-galactosidase gene (SSO3019) in sense (LST) and antisense direction (LAT). Control construct (LM) and the rest of the mini-CRISPR loci in LAT and LST are identical to chromosomal locus CR4 of strain P1 (Ssol94 in Lillestol *et al*., (2006)) as depicted on the bottom of Fig. 5A (except for two artificially introduced spacers x1, x2 that were included for cloning purposes). **B)** Mini-CRISPR amplification using total DNA extracted from single transfectants that contained LAT and LST plasmid, respectively, after the first (lanes 3 to 8) and second (lanes 9 to 13) selection round, using CRISPR specific primers LD05_fw and CrT05_rw. Lane 1 and 2: amplification from input plasmid as positive control. Identical letters (A through F) indicate cultures of the same transfectants. **C)** Amplification from the same templates using the vector-specific primers SVAF and PMJR to avoid amplification of the chromosomal CRISPR. **D)** PCR Amplification of DNA from 3 single colonies transfected with the LM control plasmid with the vector specific primers SVAF and PMJR (lane 1-3). Lane 4: DNA from LM vector as control template.

However, the viral DNA without beta-gal-specific spacers, but with the identical flanking regions did not show any recombination (see Fig. 5D), demonstrating that instability was dependent on the presence of self-targeting spacers. The *pyr*EF genes used as selection marker on the shuttle vector (Jonuscheit *et al.*, 2003) constitute a 1707 bp long region with 100% sequence identity to the chromosome, but it has never been observed to recombine. We thus conclude that cells containing a CRISPR locus on an extrachromosomal element with spacers against a chromosomally located gene (i.e. beta-galactosidase) can only survive when they eliminate the respective CRISPR locus by recombination. This would also explain why freshly transfected cells grew poorly compared to controls with other shuttle vectors (e.g. constructs 406-0M or pMJ0305, see above), indicating that most primary transfectants died. Since both constructs, LST and LAT underwent recombination, the effect appears to be similar for both sense and antisense oriented spacers implying that the observed interference mechanism acts on the DNA level.

6.4 Discussion

Different from the bacterial systems, in which it was demonstrated that DNA is targeted; CRISPR RNA-guided cleavage was shown to act on RNAs in an *in vitro* system of the hyperthermophilic archaeon *Pyrococcus furiosus* (Hale *et al.*, 2009). In the study presented here, we have established the first *in vivo* system for archaea that allows analyzing CRISPR RNA-guided activity. Our results indicate that in this *in vivo* system of the archaeon *S. solfataricus*, interference with DNA is the dominant activity, for the following three reasons: 1) The target gene ORF406 that conferred immunity and was complementary to the chromosomal spacer in the host chromosome did not stem from the virus SSV1 itself but from another genetic element and is thus not essential for propagation of the virus. An RNA interference mechanism acting on the mRNA of this (non-essential) gene should not have an effect on virus propagation. 2) Immunity was not impeded when the promoter region for the transcription of the invading target gene was deleted (see table 1) and

transcription abolished (Fig. 3), showing that the effect was apparently not dependent on mRNA production. 3) The introduced mini-CRISPR with spacers targeting a non-essential gene (beta-galactosidase, see Schelert *et al.,* 2004; Schleper *et al.*, 1994), was unstable in the host cells and triggered recombination events. If only mRNA were targeted we would not expect to see such a strong effect of constructs LST and LAT, because the amount of beta-galactosidase mRNA should not have a deleterious or recombinogenic effect. Its expression has been varied in many experiments before (Jonuscheit *et al*., 2003).

Our results give evidence for dependence of direct interference on effector RNA produced from the spacer, because mutations in the target DNA (ORF 406) decreased the effect. However, different from the CRISPR/Cas systems studied so far (Barangou *et al*., 2007; Brouns *et al*., 2008), interference in *Sulfolobus* was still effective (albeit lower) even when up to 3 mismatches were discriminating spacer from protospacer. This might indicate that the archaeal system is able to exert immunity against mutated invading elements, who would otherwise escape the CRISPR/Cas mediated interference. Furthermore, CRISPR-mediated virus defense in *Sulfolobus* does not seem to depend on specific interactions with a PAM sequence as shown for *Streptococcus thermophilus* (Deveau *et al*., 2008) as we were not able to detect such a sequence adjacent to our protospacer. According to Koonin *et al.* (2007), CRISPR locus CR3-4 belongs to cluster 7 and the corresponding protospacers of this cluster should have a PAM sequence (NGG) at zero to three nucleotides distance (Mojica *et al*. 2009).

CRISPR/Cas systems are very complex and diverse with respect to their associated proteins and spacer sequences and different systems for classification have been suggested (Makarova *et al*., 2006; Haft *et al*., 2005; Koonin and Makarova, 2009). Considering the variations of the systems, one can assume that different modules of Cas proteins may exhibit different mechanisms of action including their nucleic acid target (DNA versus RNA) and the nature of the invading genetic element they are attacking. Because CRISPRs and *cas* gene modules are often tightly linked, it is hypothesized

that the proteins and RNAs encoded by physically linked modules may function together (Kunin *et al.*, 2007). However, it is also possible but currently unclear, whether Cas proteins of different modules interact with each other or with different mature RNAs that are generated upon action of Cas nuclease (Carte *et al.*, 2008). In *Pyrococcus furiosus*, three modules of Cas proteins are encoded in two gene loci and seven CRISPR loci are distributed in the genome. CRISPR RNA-guided cleavage of target RNAs was shown to be mediated by the P. *furiosus* Cmr or Cas RAMP module proteins *in vitro* (Hale *et al.*, 2009). The Cmr complex associates with RNAs from all 7 CRISPR loci, but apparently not with other Cas proteins in *P. furiosus*. *S. solfataricus* also encodes several modules of *cas* genes, including the Cmr genes, the latter of which are located adjacent to the CRISPR locus CR10-11 (Ssol88 in Lillestol *et al.*, 2009, see Fig. 1) but not to the locus studied in our *in vivo* system (CR3-4). It remains to be shown if not only DNA or RNA, but both can be targeted in archaea and bacteria within the same cells, and if distinct activities can be assigned to the different CRISPR/Cas modules in *Sulfolobus*, as well as in *Pyrococcus*.

The activity mediated by our mini-CRISPR introduced by an extrachromosomal element was dependent on the presence of the beta-galactosidase gene in the chromosome and on the presence of the beta-galactosidase targeting spacers. We therefore assume that the chromosome of *S. solfataricus* has been directly attacked and the construct was thus deleterious for those cells that did not eliminate the mini-CRISPR. The effect of a potential RNA interference was impossible to analyze, because of the high instability of the construct. Our results strongly support the conclusions made by (Stern *et al.*, 2010) based on a bioinformatics study, that self-targeting CRISPR are rather a part of autoimmunity than an indication for regulatory mechanisms. This was also recently supported by Aklujkar and Lovley (2010), who demonstrated that the presence of a self-targeting spacer can inhibit growth in a bacterium and might contribute to evolutionary changes.

The effective recombination that we observed between the incoming mini-CRISPR locus and the chromosomal partner locus might reflect a situation that could occur in natural environments. A deleterious CRISPR-locus brought into the cell by a genetic element could be frequently replaced by a chromosomal CRISPR copy and this would help to quickly propagate CRISPR loci and newly acquired spacers among cells in a natural population. Such mobilization, at least in one direction, has been demonstrated here, i.e. the movement of a CRISPR locus from the chromosome to an invading element. In line with our finding of dynamic recombination among CRISPR loci, spontaneous dynamic recombination events have also been demonstrated in *Sulfolobus* recently, when the cells were challenged with invading matching protospacers (Gudbergsdottir *et al.,* 2011)

6.5 Conclusions

The development of an *in vivo* test system for CRISPR-mediated immunity in archaea opens new avenues for the study of virus-host interactions in this domain of life and will allow for detailed comparisons to the bacterial systems. Furthermore, it enables us to study the activity and targets of all CRISPR loci in *Sulfolobus* systematically. The investigation of this system with other *Sulfolobus* strains for which the technique of chromosomal knockouts has been demonstrated (Deng *et al.*, 2009), will allow studying the activity of specific *cas* genes *in vivo*. Additionally, we believe that mini-CRISPR loci, like the one employed here, are an effective tool to help distinguish features that determine CRISPR-mediated RNA versus DNA interference.

The discovery of the CRISPR/Cas systems in bacteria and archaea has provided a new research field and demonstrates once more the importance of viruses for microbial populations and genome evolution (Forterre, 2010). Like the research on viruses of archaea has led to unexpected discoveries, the study of CRISPR activity in this domain will undoubtedly reveal more surprises.

6.6 Experimental Procedures

6.6.1 Construction of shuttle-vectors

The primers pNOB8_ORF406_fw (ATACCATGGACAGCATAGGATTTTG-TTTTCGAG) and pNOB8_ORF406_rw (TATGGGCCCCTATGCTAGCTTAG-TGGAGTGTGAG) were used to amplify pNOB8-ORF406 from strain NOB8H2 via PCR. The product was cut with *Nco*I and *Apa*I and ligated to the pre-vector pSVA11 to add the heat inducible promoter TF55 to the PCR fragment. The construct TF55-ORF406 was then subcloned into the shuttle vector pMJ0305 (Jonuscheit *et al.*, 2003) via the *Avr*II and *Eag*I restriction sites. pMJ0305 contains the full genome of the virus SSV1 (15.465 bp), the pUC18 of E.*coli*, a 1707 bp region of *pyr*EF genes from *S. solfataricus* that can be used to complement uracil auxotrophic mutants and a 380bp region with the TF55 promoter of *S. solfataricus* (Jonuscheit *et al.*, 2003).

PCR mutagenesis was performed on ORF406 cloned into the pre-vector pSVA11 (Lubelska *et al.*, 2006) using two mutagenic phosphorylated primers pNOB8_ORF406_M3rw (TAACCTCATCCTCAGCCTTCTTTCGGAC) and pNOB8_ORF406_M3fw (CACGATGTTGCTATTCACGAGCTGA) in reverse PCR with Phusion® DNA polymerase (Finnenzyme). The resulting amplified plasmid was digested with *Dpn*I, self-ligated using Fast Ligation Kit (NEB) and transformed into Top10® (Invitrogen) competent cells. After isolation the resulting plasmids were sequenced and three different constructs were selected for further analysis: 1) 406-0M with 100% complementarity to the spacer53 of the *Sulfolobus* chromosome; 2) 406-1D ORF with a 1bp deletion at the center of the spacer (Fig.1) and 3) 406-3M with 3 mismatches. The mutated constructs were ligated into the pMJ0305 shuttle vector as described above.

The miniCRIPSRs were designed to resemble the *S. solfataricus* CR4 locus (archaeal genome browser). The putative leader sequence region "L" (600bp) and downstream region "T" of the CRISPR locus was amplified by PCR using the primers LD05_fw (CCTAGGCCGATACGTCCCCAGCAATGTAA) and LD05_rev (CCATGGTACGATATTATACCGGTTGAGCCTGCA) as well as the

primers CrT05_fw (GGGCCCGTACGGGTTGGAAGAGACTCTGA) and CrT05_rev (CGGCCGGTGTTATTCCCTTACGTTCCACTCC). Two DNA fragments called "A" (antisense) and "S" (sense) were designed to carry 4 repeats and 3 spacer sequences with perfect homology to the sense and antisense strand of the beta-galactosidase gene (SSO3019) at the positions 38-76; 441-476; 1054-1093, respectively, from the translational start site. Two additional artificial spacers "x1" and "x2" were introduced for cloning purposes (x1: ACATTTTTTGCAGGCTCAACCGGTATAATATCGTCCATGG; x2: GGCCCGTGTACGGGTTGGAAGAGACTCTGAATGAGTTTCTATTACT). The fragments were synthesized by Mr.Gene (Regensburg). Finally, the constructs LAT and LST were assembled using fragment A and S, the leader region and the terminal part of the CRISPR locus 4, by multistep cloning using the restriction sites NcoI and ApaI and were then inserted in the pMJ0305 shuttle vector, as described above. All constructs were verified by sequencing.

6.6.2 Growth of S. solfataricus and plaque assay

Strains *S. solfataricus* P1 and P2 (DSM 1616, 1617) were grown at 78°C and pH 3 in Brock's medium (Grogan, 1989), with 0.1% (w/v) tryptone and 0.2% sucrose, using long necked flasks in a shaking incubator. The optical density of liquid cultures was monitored at 600 nm. Solid media was prepared by adding gellan gum to a final concentration of 0.6% and Mg^{2+} and Ca^{2+} to 0.3 and 0.1 M, respectively. Plates were incubated for five days at 80°C. For plaque assays, transfected cells were mixed with 300 µl of 10x concentrated, logarithmically grown cells (for the lawn) and with 2 ml of prewarmed growth medium supplemented with 0.3% gelrite (final concentrations) and sucrose/yeast extract. The mixture was quickly poured on solid gellan gum (Gelrite; Kelco Biopolymers) plates. Plaques formed after incubation for 2 days at 80°C.

6.6.3 Transfection of S. solfataricus

DNA was quantified with a nano-Drop spectrometer (Peqlab) and loaded on a 0,8% agarose–gel to check for purity and DNA topology. 100 to 150 ng of viral DNA were used for electroporation of exponentially grown *S. solfataricus* cells as described in (Kurosawa and Grogan, 2005). Cells were regenerated for 1 hour in regeneration solution (Berkner *et al.*, 2007) immediately after electroporation and were subsequently used for plating (plaque assay) or for inoculation of liquid cultures. Transfection efficiencies were determined in triplicate by counting plaque forming units (pfu) on gelrite plates with transformants of strain *S. solfataricus* P2 and cell lawns of P2 or with transformants of strain P1 and lawns of strain P1. Maximal transfection efficiencies for strain P2 were usually 12,000 PFU/µg DNA and 6,600 PFU/µg DNA for P1. Therefore transfection efficiencies were normalized in table 1 to the highest plaque count obtained in each single experiment.

6.6.4 Analysis of transformants

The mini-CRISPR constructs in *S. solfataricus* transformants were checked by PCR using two vector-specific primers, flanking the insertion region: SVAF (AGGTGCTGATGTGATAGGATTTGC) and PMJR (AGCGGATAACAATTTCACACAGGA) as well as two primers specific for the CRISPR locus (LD05_fw and CrT05_rev). Total DNA was extracted from exponentially grown cells of transformation mixtures or single colonies, respectively, and 100ng of DNA was used as template for the polymerase chain reaction.

For sequence analysis, total DNA was extracted from LAT and LST single colonies and 1µg of DNA was used to re-transform chemically-competent *E. coli* TOP10 cells. Single colonies of LAT and LST were selected and used to isolate the recombinant shuttle vector. The sequence of the inserts was determined by standard Sanger sequencing. For beta-galactosidase assays, crude extracts of logarithmically grown cells were prepared and beta-galactosidase activity was determined as previously described in Jonuscheit *et al*. (2003).

6.6.5 Northern Analysis

Total RNA was extracted from 20-50 ml of culture using mirVana™ miRNA Isolation Kit (Ambion). Five µg of RNA were denatured for 10 min at 65°C before separation on a 12% polyacrylamide / 6M urea gel (Fig. S2) or a 1.2% denaturing formaldehyde-containing agarose gel (Fig. 3), followed by transfer to nylon membranes (Membrane Hybond-XL (Amersham) or Biodyne® A 0.2 µm (PALL) by electro- or capillary-blotting, respectively. After cross-linking, the membrane was hybridized with a denatured 363 nt DIG-labelled dsDNA of ORF406 (Fig. 3) or a 24 nt long [^{32}P] 5´-end-labeled oligonucleotide probe specific for the antisense transcript of ORF406 (spacer region), (406antisense GAAGGCTGAGGATGAGGTTACACG) and a sense probe (with the same sequence in reverse complementation) at 42°C overnight (Suppl. Fig. S2). The washing was done twice for 10 min at 42°C with buffer I (2xSSC/ 0.1%SDS) and buffer II (0.1xSSC/ 0.1%SDS), or 20 min and twice 30 min at 55°C and 60/65°C with 0.2xSSC/0.1% SDS-buffer respectively, and the hybridization signals were visualized using a Phosphor Imager (Molecular Dynamics) or exposure to film.

6.7 Acknowledgements

We thank Rebecca and Michael Terns, Renee Schröder and Udo Blaesi for inspiring discussions and encouragement and Anja Spang for critically reading the manuscript. This project was partly financed through the Norwegian Research Council in the frame of the Era-Net Project "SulfoSYS" (SysMo P–N-01-09-23) and by the EU-RTN project SOLAR (MCRTN-2005-033499-2).

6.8 References

Aklujkar, M. & D. R. Lovley, (2010) Interference with histidyl-tRNA synthetase by a CRISPR spacer sequence as a factor in the evolution of Pelobacter carbinolicus. *BMC Evol Biol* **10**: 230.

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

Beloglazova, N., G. Brown, M. D. Zimmerman, M. Proudfoot, K. S. Makarova, M. Kudritska, S. Kochinyan, S. Wang, M. Chruszcz, W. Minor, E. V. Koonin, A. M. Edwards, A. Savchenko & A. F. Yakunin, (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J Biol Chem* **283**: 20361-20371.

Berkner, S., D. Grogan, S. V. Albers & G. Lipps, (2007) Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for Sulfolobus acidocaldarius and Sulfolobus solfataricus, model organisms of the (cren-)archaea. *Nucleic Acids Res* **35**: e88.

Bize, A., E. A. Karlsson, K. Ekefjard, T. E. Quax, M. Pina, M. C. Prevost, P. Forterre, O. Tenaillon, R. Bernander & D. Prangishvili, (2009) A unique virus release mechanism in the Archaea. *Proc Natl Acad Sci U S A* **106**: 11306-11311.

Bolotin, A., B. Quinquis, A. Sorokin & S. D. Ehrlich, (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551-2561.

Brouns, S. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin & J. van der Oost, (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.

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

Chakraborty, S., A. P. Snijders, R. Chakravorty, M. Ahmed, A. M. Tarek & M. A. Hossain, (2010) Comparative network clustering of direct repeats (DRs) and cas genes confirms the possibility of the horizontal transfer of CRISPR locus among bacteria. *Mol Phylogenet Evol* **56**: 878-887.

Deng, L., H. Zhu, Z. Chen, Y. X. Liang & Q. She, (2009) Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon Sulfolobus islandicus. *Extremophiles* **13**: 735-746.

Deveau, H., R. Barrangou, J. E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath & S. Moineau, (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol* **190**: 1390-1400.

Deveau, H., J.E. Garneau & S. Moineau (2010) CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol.* **64**:475-93.

Forterre, P., (2010) Defining life: the virus viewpoint. *Orig Life Evol Biosph* **40**: 151-160.

Garneau, J.E., M.E. Dupuis, M. Villion, D.A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A.H. Magadán & S. Moineau (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**:67-71.

Garrett, R. A. & Klenk, H.P., (2006) Archaea: Evolution, Physiology, and Molecular biology *ed. John Wiley and Sons Ldt Blackwell Publishing Ltd illustrated edition*.

Godde, J. S. & A. Bickerton, (2006) The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. *J Mol Evol* **62**: 718-729.

Grissa, I., G. Vergnaud & C. Pourcel, (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172.

Grogan, D. W., (1989) Phenotypic characterization of the archaebacterial genus Sulfolobus: comparison of five wild-type strains. *J Bacteriol* **171**: 6710-6719.

Gudbergsdottir, S., L. Deng, Z. Chen, J.V. Jensen, L.R. Jensen, Q. She & R.A. Garrett (2011) Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol Microbiol.* **79**:35-49.

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

Hale, C., K. Kleppe, R. M. Terns & M. P. Terns, (2008) Prokaryotic silencing (psi)RNAs in Pyrococcus furiosus. *RNA* **14**: 2572-2579.

Hale, C. R., P. Zhao, S. Olson, M. O. Duff, B. R. Graveley, L. Wells, R. M. Terns & M. P. Terns, (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945-956.

Han, D., K. Lehmann & G. Krauss, (2009) SSO1450--a CAS1 protein from Sulfolobus solfataricus P2 with high affinity for RNA and DNA. *FEBS Lett* **583**: 1928-1932.

Haring, M., G. Vestergaard, R. Rachel, L. Chen, R. A. Garrett & D. Prangishvili, (2005) Virology: independent virus development outside a host. *Nature* **436**: 1101-1102.

Horvath, P. & R. Barrangou, (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**: 167-170.

Jansen, R., J. D. Embden, W. Gaastra & L. M. Schouls, (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**: 1565-1575.

Jonuscheit, M., E. Martusewitsch, K. M. Stedman & C. Schleper, (2003) A reporter gene system for the hyperthermophilic archaeon Sulfolobus solfataricus based on a selectable and integrative shuttle vector. *Mol Microbiol* **48**: 1241-1252.

Karginov, F. V. & G. J. Hannon, (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell* **37**: 7-19.

Koonin, E. V. & K. S. Makarova, (2009) CRISPR/Cas: an adaptive immunity system in prokaryotes. *F1000 Biol Rep* **1**: 95.

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

Kurosawa, N. & D. W. Grogan, (2005) Homologous recombination of exogenous DNA with the Sulfolobus acidocaldarius genome: properties and uses. *FEMS Microbiol Lett* **253**: 141-149.

Lillestol, R. K., P. Redder, R. A. Garrett & K. Brugger, (2006) A putative viral defence mechanism in archaeal cells. *Archaea* **2**: 59-72.

Lillestol, R. K., S. A. Shah, K. Brugger, P. Redder, H. Phan, J. Christiansen & R. A. Garrett, (2009) CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Mol Microbiol* **72**: 259-272.

Lubelska, J., M. Jonuscheit, C. Schleper, S. Albers & A. J. Driessen, (2006) Regulation of expression of the arabinose and glucose transporter genes in the thermophilic archaeon Sulfolobus solfataricus. *Extremophiles* **10**: 383-391.

Makarova, K. S., N. V. Grishin, S. A. Shabalina, Y. I. Wolf & E. V. Koonin, (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.

Marraffini, L. A. & E. J. Sontheimer, (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**: 1843-1845.

Marraffini, L. A. & E. J. Sontheimer, (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**: 181-190.

Martin, A., S. Yeats, D. Janekovic, W. D. Reiter, W. Aicher & W. Zillig, (1984) SAV 1, a temperate u.v.-inducible DNA virus-like particle from the archaebacterium Sulfolobus acidocaldarius isolate B12. *EMBO J* **3**: 2165-2168.

Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez & E. Soria, (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174-182.

Pourcel, C., G. Salvignol & G. Vergnaud, (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**: 653-663.

Prangishvili, D., P. Forterre & R. A. Garrett, (2006) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* **4**: 837-848.

Reiter, W. D., P. Palm & S. Yeats, (1989) Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic Acids Res* **17**: 1907-1914.

Rousseau, C., M. Gonnet, M. Le Romancer & J. Nicolas, (2009) CRISPI: a CRISPR interactive database. *Bioinformatics* **25**: 3317-3318.

Schelert, J., V. Dixit, V. Hoang, J. Simbahan, M. Drozda & P. Blum, (2004) Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon Sulfolobus solfataricus by use of gene disruption. *J Bacteriol* **186**: 427-437.

Schleper, C., K. Kubo & W. Zillig, (1992) The particle SSV1 from the extremely thermophilic archaeon Sulfolobus is a virus: demonstration of infectivity and of transfection with viral DNA. *Proc Natl Acad Sci U S A* **89**: 7645-7649.

Schleper, C., R. Roder, T. Singer & W. Zillig, (1994) An insertion element of the extremely thermophilic archaeon Sulfolobus solfataricus transposes into the endogenous beta-galactosidase gene. *Mol Gen Genet* **243**: 91-96.

She, Q., H. Phan, R. A. Garrett, S. V. Albers, K. M. Stedman & W. Zillig, (1998) Genetic profile of pNOB8 from Sulfolobus: the first conjugative plasmid from an archaeon. *Extremophiles* **2**: 417-425.

Stern, A., L. Keren, O. Wurtzel, G. Amitai & R. Sorek, (2010) Self-targeting by CRISPR: gene regulation or autoimmunity? *Trends Genet* **26**: 335-340.

van der Oost, J. & S. J. Brouns, (2009) RNAi: prokaryotes get in on the act. *Cell* **139**: 863-865.

van der Oost, J., M. M. Jore, E. R. Westra, M. Lundgren & S. J. Brouns, (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* **34**: 401-407.

6.9 Supplementary material

6.9.1 Supplementary figures

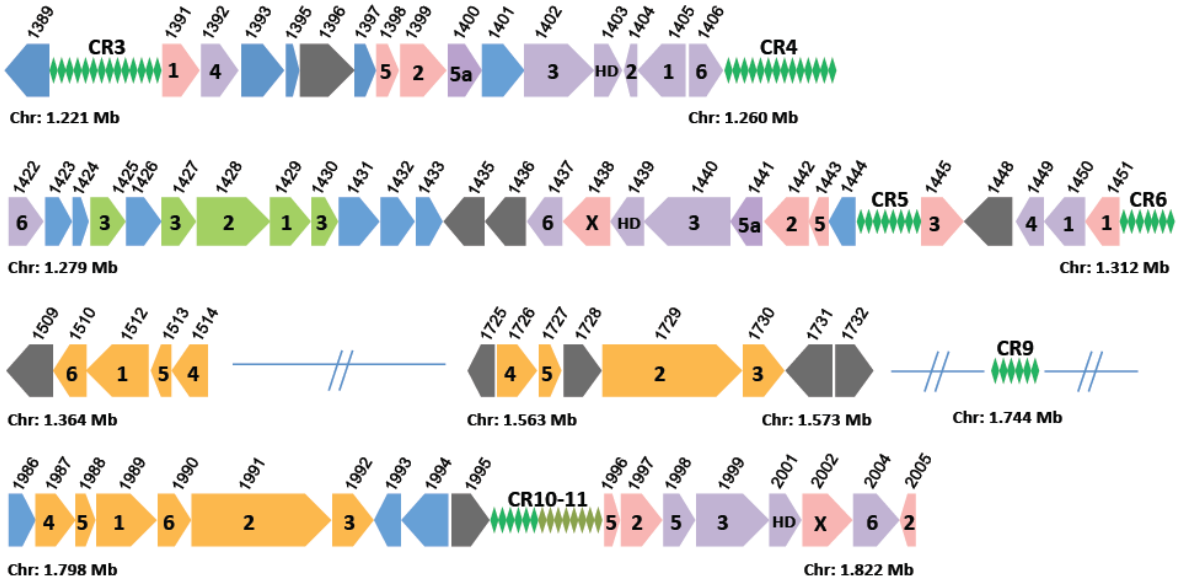


Fig.S1: **Schematic representation of the different CRISPR/Cas loci in *S. solfataricus P2*.** Same as in Fig. 1, main paper, but SSO numbers given for each gene.

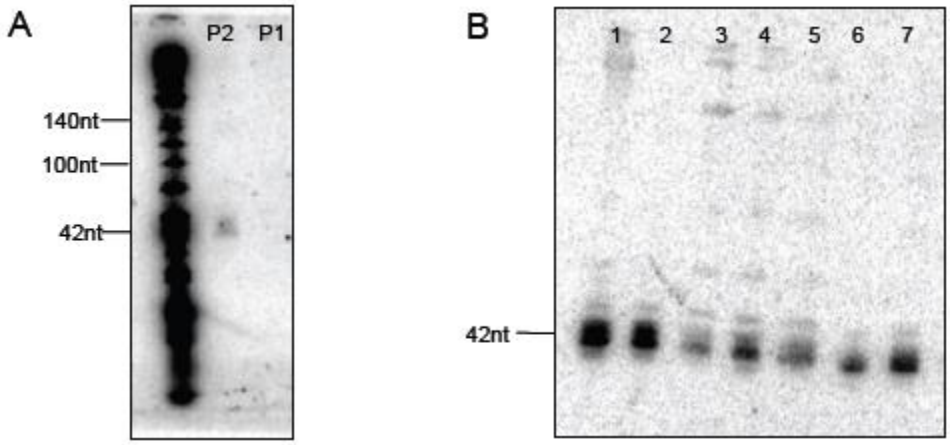


Fig.S2: Northern blot using an oligonucleotide probe complementary to spacer 53 in locus CR3-4 of *S.solfataricus* strain P2. Five µg of total RNA were loaded in each lane, **A)** RNA of strain P1 and P2,respectively, RNA marker (left) for size determination, **B)** strain P2, in lane 1-3 RNA from cells grown at 65°C, 70°C and 80°C, respectively, with RNA of exponentially grown culture in tryptone (lane 4)and yeast media (lane 5), respectively, with RNA of a culture in the stationary growth phase in tryptone (lane 6) and yeast media (lane 7). When hybridizing with an inversely oriented sense-probe to the same spacer, we did not obtain any signals, indicating that transcription of the CRISPR module proceeds only (or mostly) from the side of the leader sequence (data not shown).

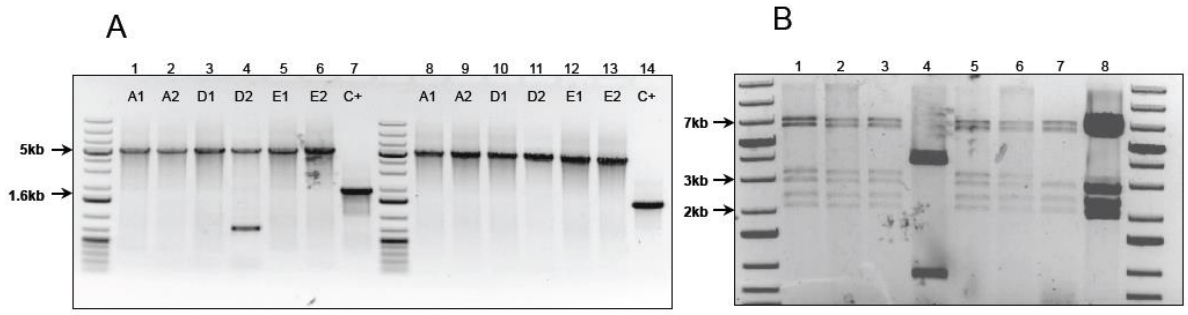


Fig.S3: **Analysis of transfectants from mini-CRISPR experiment as shown in Fig. 5** (main paper): **A)** PCR amplification of the mini-CRISPR locus with CRISPR-specific primers from plasmids recovered from *E.coli* cells after retransformation of the DNA from the *Sulfolobus* transformants A through F, after 2nd selection (see Fig. 5 part A,B). A1 and A2 design *E.coli* colonies derived after transforming DNA from colony A of Fig. Part A,B). D1 and D2 are descendants of colony D in Fig. 5 Part A, B), and E1, E2 stem from E. Lane C+: control, template LAT construct. **B)** *Eco*RI restriction digests of plasmid DNA recovered from the single transformants A1 through E2 (lanes 1-6) as depicted in Fig. Part A. 5 out of the 6 plasmids showed the same pattern. These are representatives of a total of 22 plasmids analysed (21 showed the same pattern as those in lanes 1,2,3,5 and 6. lane 7, 8: plasmid with LAT and LST construct (controls)

7 UNEXPECTEDLY BROAD TARGET RECOGNITION OF THE CRISPR-MEDIATED VIRUS DEFENSE SYSTEM IN THE ARCHAEON *SULFOLOBUS SOLFATARICUS*

*Andrea Manica, *Ziga Zebec, Julia Steinkellner and Christa Schleper§
University of Vienna, Department of Genetics in Ecology, Althanstr. 14, A-1090 Vienna, Austria
*these authors contributed equally to the work

§corresponding author:
Christa Schleper, e-mail: christa.schleper@univie.ac.at
phone: +43 14277 57800, fax: +43 142779578

7.1 Abstract

The hyperthermophilic archaeon *Sulfolobus solfataricus* carries an extensive array of CRISPR (clustered regularly interspaced short palindromic repeats) systems able to mediate DNA degradation of invading genetic elements when complementarity to the small CRISPR-derived (cr)RNAs is given. Studying virus defense *in vivo* with recombinant viral variants, we demonstrate here that an unexpectedly high number of mutations is tolerated between the CRISPR-derived guide RNAs (crRNAs) and their target sequences (protospacer). Up to 15 mismatches in the crRNA still led to approx. 50% of DNA degradation, when these mutations were outside the "seed"-region. More than 15 mutations were necessary to fully abolished interference. Different from other CRISPR systems investigated *in vivo*, mutations outside the protospacer region indicated no need for a PAM sequence (protospacer

adjacent motif) to confer DNA interference. However, complementarity of only 3 nucleotides between the repeat-derived 5'handle of the crRNA and nucleotides adjacent to the protospacer enabled self recognition, i.e. protection of the host locus. Our findings show commonalities and differences among the various CRISPR-mediated defense systems and suggest that they should not merely be perceived as a "first-barrier-defense system" but may be considered to have a broader mechanism that allows host cells to cope with viruses keeping them at reduced levels.

7.2 Introduction

Recently a new defence mechanism that protects bacteria and archaea from invading genetic elements through an RNA-mediated DNA interference mechanism was discovered. Upon challenge of a virus or plasmid the host cells integrate small DNA sequences of 30 to 45 nt of the invader genome into their chromosome (the spacers), in a region called CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats). After transcription of the CRISPR region, the full length RNA is processed into small CRISPR RNAs (crRNAs) that mediate degradation of a complementary invading DNA with the help of protein complexes encoded adjacent to the CRISPR locus, called Cas (CRISPR associated proteins) (Barrangou *et al.*, 2007; Brouns *et al.*, 2008; Marraffini & Sontheimer, 2008; Gudbergsdottir *et al.*, 2011). This defense mechanism is present in most archaea (87%) and in more than 45% of bacteria (CRISPRdb http://crispr.u-psud.fr/crispr/) and shows a great diversity (Grissa *et al.*, 2007). CRISPR loci can be divided into more than 45 different families based on the presence of particular Cas protein sets, which are now classified into three major types: I, II, III (Bolotin *et al.*, 2005; Haft *et al.*, 2005; Makarova *et al.*, 2011).

The CRISPR locus is transcribed in unique orientation, from a promoter region within the leader sequence (Brouns *et al.*, 2008), although the presence

of antisense transcripts has also been detected (Lillestol *et al.*, 2009; Deng *et al.*, 2011; Richter *et al.*, 2012). Processing of the full length CRISPR transcript has been studied in several archaeal and bacterial systems (Lillestol *et al.*, 2009; Pougach *et al.*, 2010; Pul *et al.*, 2010). It is performed by a Cas protein belonging to the Cas6 family in the CRISPR types I and III, or by a mechanism involving non-coding antisense RNA (tracrRNAs) and cellular RNAseIII in type II CRISPR systems (Deltcheva *et al.*, 2011). In the type I and III CRISPR-Cas systems studied so far the crRNA contains an 8nt-tag at its 5´terminus (5´handle) corresponding to 8 nucleotides of the repeat preceding the spacer, followed by the unique spacer sequence, then followed by a 3´sequence corresponding to the remaining part of the next direct repeat (Carte *et al.*, 2008; Lintner *et al.*, 2011). The 5´tag was shown to be responsible for self-nonself discrimination of the target sequence in the bacterium *Staphylococcus epidermidis*, protecting the genomic CRISPR locus from degradation by the self-originated crRNA (Marraffini and Sontheimer, 2010). The authors identified the positions important in the protospacer adjacent sequence (which we refer to as PAS here) that can block CRISPR activity when matching the central 4nt of the 5´handle of the spacer, demonstrating how the CRISPR system can protect its own locus by the targeting of self-originated spacers.

New insights into the nature of the interaction between the crRNA and its DNA target sequence was recently gained by Wiedenheft *et al.* in *Pseudomonas aeruginosa* and by Semenova *et al.* in *E. coli* who provided evidence that in bacterial type I systems the first 8nt of the 5´region of the spacer are critical for target recognition (acting like a "seed" sequence) and that mutations in the corresponding region of the protospacer have a major impact on the DNA interference activity (Semenova *et al.*, 2011; Wiedenheft *et al.*, 2011), while four mutations between spacer and target outside the seed region were tolerated (Semenova *et al.*, 2011). This observation differs from the conclusion drawn by Deveau and colleagues in *Streptococcus thermophilus* where even a single mutation in the protospacer sequence, abolished completely the CRISPR defence mechanism (Deveau *et al.*, 2008).

The PAM (protospacer adjacent motif) sequences were first detected by Mojica *et al.* as short dinucleotide sequences that have an important role in new spacer acquisition (Mojica *et al.*, 2009). These short motifs are often located between the first nucleotides at the 3´end of the protospacer and determine in which orientation the protospacer will be inserted into the CRISPR locus. Successive studies have also demonstrated the influence of this region during the interference process in some systems, such that even a single mutation in the PAM sequence can abolish DNA interference (Deveau *et al.*, 2008; Gudbergsdottir *et al.*, 2011; Semenova *et al.*, 2011).

We and others have recently established an *in vivo* study system for the hyperthermophilic archaeon *Sulfolobus solfataricus* (Gudbergsdottir *et al.*, 2011, Manica *et al.*, 2011). This organism possesses a rather complex and extended CRISPR-Cas assembly with several Cas modules classified into the CRISPR-Cas system type I, type IIIA and IIIB (Makarova *et al.*, 2011). The activity of three of the six *S. solfataricus* CRISPR-loci (Lillestol *et al.*, 2006; Lillestol *et al.*, 2009) were previously investigated *in vivo* (Gudbergsdottir *et al.*, 2011; Manica *et al.*, 2011). While CRISPR E-F were inactive due to the lack of transcription and the absence of correctly processed crRNAs (Gudbergsdottir *et al.*, 2011), loci A-B and C-D produced active crRNAs and these were able to reduce transfection efficiency of a virion (Manica *et al.*, 2011) or plasmid (Gudbergsdottir *et al.*, 2011) carrying protospacers with high similarity to the respective crRNA. From those studies it appeared that in *S. solfataricus* the CRISPR-Cas system is able to trigger immunity even when several mismatches between crRNA and its target sequence are present, indicating that the archaeal system might be more promiscuous than the currently studied bacterial systems.

In order to compare the archaeal system to other bacterial CRISPR systems and to dissect potential differences, we investigate here the minimal requirement of sequence similarity between crRNAs and protospacers in *S. solfataricus,* and analyse if a potential SEED sequence can be dissected. We explore the basis for self/non-self discrimination by crRNAs and we also shed light on the need of a PAM sequence in the interference process.

7.3 Material and Methods

7.3.1 Strain growth and plaque assay

Strains *S. solfataricus* P1 and P2 (DSM 1616, 1617) were grown at pH 3 and 78°C in Brock's medium (Grogan, 1989) using long necked flasks, with 0.1% (w/v) tryptone and 0.2% sucrose, in a shaking incubator. The optical density of liquid cultures was monitored at 600 nm. Plates were prepared by adding gellan gum (Gelrite; Kelco Biopolymers) to the Brock´s solution to a final concentration of 0.6% and Mg^{2+} and Ca^{2+} to 0.3 and 0.1 M, respectively. For the plaque assays, 5µl transfected cells were mixed with 300 µl of 10x concentrated, logarithmically grown cells (for the lawn) and with 2 ml of pre-warmed growth medium supplemented with 0.3% gellan gum (final concentrations) without carbon source. The mixture was quickly poured on plates. Plaques formed after incubation for 2 days at 78°C.

7.3.2 Preparation of vector constructs

The ORF 406 of pNOB8 plasmid (She *et al.*, 1998) was amplified with primers 406Fw and 406Rw using *Phusion DNA polymerase* (Finnzymes) with the following PCR condition: 96°C 2´, 30 cycles of 96°C 15", 52°C 20", 72°C 1´30" followed by 72°C 5´. The PCR product was gel purified and cloned into the pCR8-GW cloning vector (Invitrogen). Plasmid DNA was extracted (E.Z.N.A. Plasmid Miniprep KitI, Omega Bio-tek) and 1ng of plasmid was used as template for inverse PCR with mutagenic primers (see suppl. material Tab.S1 for primer sequences). PCR was performed in 1x GC Buffer, 1U Phusion polymerase 0.4mM dNTPs, and 400 pmol mutagenesis primer mix using the following PCR conditions: 96°C 2´, 33 cycles of 96°C 15", 58°C 15", 72° 3´ with a final extension of 5´ at 72°C. Primers were phosphorylated prior to use, mixing 2µl 100µM FW and RW mutagenic primer with 2µl T4 kinase buffer and 1U of T4 DNA kinase in 20µl reaction volume and incubating at 37°C 1h. In suppl. table S1 the different primer combinations for each construct are

listed. The linear plasmids amplified by inverse PCR were cleaned using PCR clean up kit (EZNA) and digested with 2U of *Dpn*I (to digest the non-amplify wt plasmid) for 2h at 37°C. After digestion the plasmids were column purified and 50ng were used in a ligation reaction with 1U T4 Ligase (New England Biolabs) and 2µl of 10x T4 ligase Buffer in a volume of 50µl and incubation for 3h at 21°C. Two µl of re-circularized plasmids were then used to transform TOP10® cells (Invitrogen). Positive colonies were screened by PCR and the mutated protospacer sequence was analysed by Sanger sequencing.

Correct clones were amplified and purified by plasmid preparation and 100ng of it was used in a gateway LR-recombination reaction according to the manufacturer's instructions (Invitrogen), with the pDEST gateway vector pMJ-GW (see suppl Fig.S1). This vector was designed by placing the gateway cassette A (Invitrogen) between the restriction sites AvrII and EagI of the *Sulfolobus* shuttle vector pMJ03-05 (Jonuscheit *et al.*, 2003).

7.3.3 Transfection assay and PFU normalization

DNA of the recombinant shuttle vectors was extracted from transformed *E. coli* cells using E.Z.N.A. Plasmid Miniprep KitI Omega Bio-tek and quantified in a Nano-Drop spectrometer (PeqLab), then an aliquot was loaded on a 0.8% agarose gel to analyse its purity and topology. 150ng of viral DNA was used to transfect electrocompetent *Sulfolobus* cells (Kurosawa and Grogan, 2005) using 1mm cuvettes (PeqLab) with the following electroporation condition: 1250 V/25vµF/ 1000vΩ in a gene pulser (Biorad). After transfection the cells were incubated for 1h in recovery solution (Berkner *et al.*, 2007) at 75°C. The transfected cells were then used for plating (plaque assay). Transfection efficiencies were determined in triplicate by counting plaque forming units (PFU) on plates and normalizing to the viral vector with the highest plaque count.

7.4 Result

7.4.1 Minimal requirement of sequence complementarity between crRNA and protospacer

In our previous work we have demonstrated that the CRISPR system in *S. solfataricus* can recognize and trigger degradation of a protospacer also when up to four mutations are introduced into the protospacer region, and that the efficiency of transfection decreased with the increase of the number of mismatches between spacer and protospacer (Manica *et al*., 2011).To understand the minimal requirement of mutations that can completely abolish the DNA interference, we constructed 14 new viral variants carrying a mutated protospacer and used those to transfect *S. solfataricus*. Each virus variant was designed to carry the ORF 406 of the conjugative plasmid pNOB8 (She *et al*., 1998) in which the protospacer sequence matching spacer n°53 of CRISPR locus A was mutated to a different degree (Fig.1A). Each viral DNA was successively analysed for its ability to persist in *S. solfataricus* strain P2 and the transfection rate was quantified in a plaque assay. The viral vectors were divided into two groups: the "5P "constructs and the "3P" constructs (see Fig.1B). The "5P" constructs contained mismatches between the 5´half of the spacer and the protospacer, while the "3P" constructs were designed to study the importance of the interaction between the 3´-half of the spacer and the targeted sequence (see Fig.1B). Whereas construct 5P-3M (3 mismatches with respect to the spacer) had a transfection efficiency of ~20% compared to the control vector NS (no interference) carrying the full ORF406 without any protospacer, transfection efficiency rose to 75% for construct 5P-6M, and even to ~90% for 5P-8M and 5P-10M. These results demonstrate that more than six mutations in the interaction region between the 3´end of the protospacer and 5´half of the spacer strongly affect and eight and more mutations almost completely abolish the DNA interference in *S. solfataricus*. A different outcome was obtained when testing the region of interaction on the other side, i.e. between the protospacer and the spacer 3´half. Seven

different protospacer sequences with different numbers of mutations (Fig.1B) were cloned into the viral vector, with 6 (3P-6M), 9 (3P-9M), 11 (3P-11M) 13 (3P-13M), 14 (3P-14M), 15 (3P-15M) and 18 (3P-18M) nucleotides being replaced at the 5'end of the 0M protospacer. In addition to the above constructs also vector 3P-4M was tested, as previously done (Manica *et al.*, 2011) which carries four mutations in positions +36, +34, +32 and +18 (in positions 36 and 18 G-U pairing between spacer and protospacer was possible in the DNA-RNA hybrid). The transfection efficiency of the constructs with respect to the control is reported in Fig.1B. Four mutations (3P-4M) showed no effect on DNA degradation. The construct was still cleaved to the same degree as the perfect matching protospacer 0M. Six consecutive mutations of 3P-6M increased the ability to form plaques (PFU) only to ~5%, and only nine mutations increased the efficiency of transfection to ~25%. Eleven and thirteen mutations were still not sufficient to fully abolish DNA degradation; in fact the viral vectors 3P-11M and 3P-13M caused only ~40 and ~65% of plaques when electroporated into *Sulfolobus* cells, respectively. Surprisingly, plaques formation dropped again from ~65% to ~30% with construct 3P-14M, whereas construct 3P-15M formed ~55% of plaques compared to the control NS vector. The transfection efficiency rose to that of the control (NS) only when the viral vector 3P-18M was tested (>95%). To understand if the difference in transfection efficiency of the construct 3P-13M and 3P-14M was due to reasons not related to sequence complementarity but e.g. to plasmid quality, we used the same constructs to transfect *S. solfataricus* strain P1, a close relative of strain P2 which does not contain spacer A53 in its genome. No significant differences in infectivity were found demonstrating that the transfection capacity of all viral vectors was equally high.

Since the different transfection efficiencies of the viral vectors 3P-13M and 3P-14M could not be explained by differences in plasmid quality, we investigated the possibility that the replaced nucleotides in the mutated protospacer sequences could be directly involved in lowering or increasing the CRISPR activity. We therefore focused our attention on the single nucleotide difference between constructs 3P-13M and 3P-14M. In construct 3P-14M a

single G base had been mutated to a T base to increase the number of mismatches between spacer and its target by one nucleotide. This mutation had created an AT di-nucleotide sequence immediately in front of the region of homology between protospacer and spacer (Fig.1B). To understand if this AT di-nucleotide could be a putative CRISPR cutting site, (as previously seen for the *S. solfataricus* Cmr complex (Zhang *et al.*, 2012) which could increase protospacer degradation in construct 3P-14M, we replaced the AT sequence in pos. 24-25 with the di-nucleotide GC. Interestingly, the transfection efficiency of the newly designed vector was ~30%, as for construct 3P-14M. Thus, the reason of this different transfection ability between vectors 3P-13M and 3P-14M remained obscure. We also excluded the possibility that RNA secondary structure could be a trigger for higher infectivity of the viral vector 3P-13M due to the fact that the ORF406 is not transcribed in our viral vector and we are directly manipulating the target DNA and not the spacer A53.

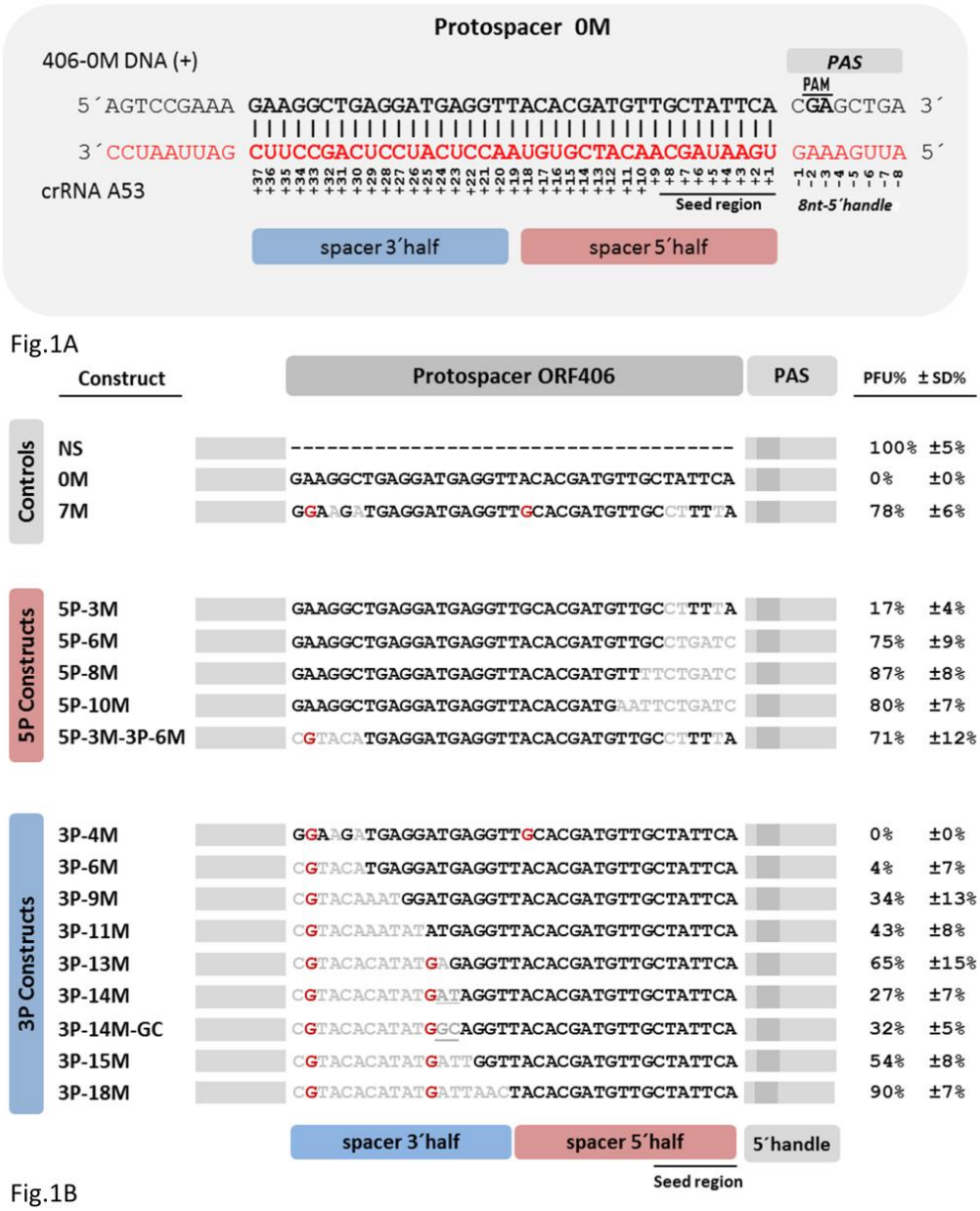


Figure 1: A) Schematic representation of the crRNA-protospacer interaction region, as well as flanking sequences, including 5'handle, protospacer adjacent motif (PAM) and protospacer adjacent sequence (PAS). **B)** CRISPR-mediated DNA interference as measured by the capability of plaque formation of different protospacer-containing virus DNA constructs. Controls were NS, which did not carry the protospacer sequence (negative control), 0M which carries a protospacer sequence in the ORF406 homologous to the crRNA53 of CR3 (positive control); and 7M the wild type pNOB ORF406 sequence with 7 mutations to the crRNA. 5-P and 3-P distinguish the constructs with respect to the location of mutations in the 5´half and 3´half, respectively, and further numbers indicate the number of mutations between crRNA and protospacer. Transfection efficiencies of each construct are given as percentage of that of the control construct NS. Bars represent standard deviations of ≥ 3 replicates.

7.4.2 Synergistic effect of mutations

The above results indicate that the ability of the different viral vectors to escape the CRISPR system is mostly proportional to the amount of mutations and that every mutation has a different "weight" depending on its localization along the interaction region between protospacer sequence and crRNA. Mutations towards the 5' end of the protospacer (= 3'half of spacer) do not or only minimally influence the ability of the virus to escape the CRISPR defense (see Fig.1B constructs 3P-4M, 3P-6Md, 3P-9Md). However, we have noticed in our earlier study, that these less important positions also influence interference when they occur in combination with others. In (21) constructs 406-7M and 406-3M (here reported as 7M and 5P-3M) had been tested. The constructs 5P-3M and 3P-4M (see above) are derivatives of construct 7M, where the mutations at the 5´half (5P-3M) and 3´half (3P-4M) of the 7M protospacer, respectively, were replaced to perfectly match the spacer A53. The transfection efficiency described for construct 5P-3M was ~20% and for construct 3P-4M was ~0%. Interestingly, construct 7M which had both sets of mutations (5P-3M and 3P-4M) escaped the CRISPR system in ~80% of the cases. These data show that those additional mutations at the opposite side of the protospacer triggered a synergistic effect, increasing the capacity of the virus to escape the CRISPR system. To confirm this result we designed a new protospacer called 5P-3M-3P-6M in which the mutations of construct 5P-3M were combined with those of the 3P-6M construct. The PFUs obtained with construct 5P-3M-3P-6M (~70%) were higher than those of both parental constructs: 5P-3M (~20%) and 3P-6M (~5%). These results indicate that base pairing in the first six nucleotides of the 3´half of the crRNA are also important for the interference process, as mutations decrease the CRISPR system efficiency when occurring together with mutations within the seed sequence of crRNA.

7.4.3 Self/non-self target recognition in *S. solfataricus*

In *Staphylococcus epidermidis* the presence of at least three complementary nucleotides in the 3' flanking region of the protospacer adjacent sequence

(which we term PAS here) is needed for "self-recognition", i.e. for protecting the protospacer from degradation and thus to avoid the targeting of the CRISPR locus by the self-originated spacers (Marraffini and Sontheimer, 2010). To increase the understanding of this process in archaea we mutated the PAS of the protospacer 0M on the viral vector pMJ to match the 5'handle of the corresponding crRNA, that stems from the last 8nt of the repeat present in the CRISPR locus (CRISPR A-B of *S. solfataricus*) (Fig.2).

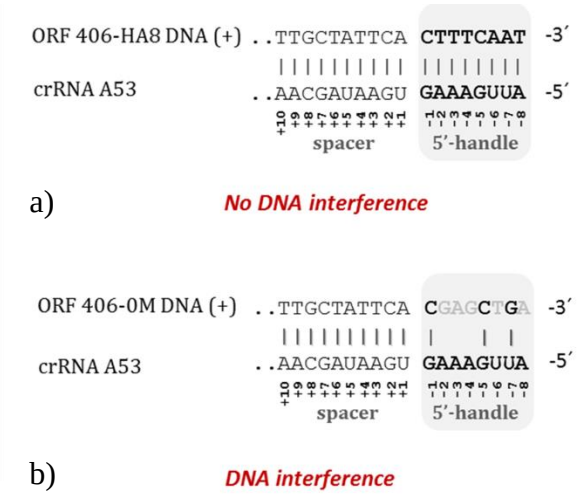


Fig.2: Protospacer adjacent region (PAS) of ORF406 (pos. -1 to -8) and the 8nt 5´handle of the crRNA A53 of CR3 in constructs HA8 [A] and 0M [B].

The results are reported in Fig.3. Construct HA8 (HA for HAndle) in which the PAS of the protospacer matched perfectly the 8nt long 5'handle of the crRNA, was no more a target of the CRISPR-mediated DNA interference, i.e. yielded as many plaques as a virus without protospacer.

We then reduced the number of matches between flanking sequence and 8nt 5´handle to characterize the minimal requirement for full protection. Six nucleotides complementarity (HA6) and even only 4 nt (HA4a and HA4b) fully protected the target, when the protospacer had at least 3 consecutive matches with the 5´handle between positions -3 to -5 (see Fig.3). Four matches, when covering different positions in the 5´handle (pos. -1 to -3 and pos. -6) did not confer full protection but partially protected the protospacer from degradation (30% for HA4c). To identify which positions in the flanking sequence are important for the protection we then positioned 3 consecutive

matching nucleotides at different places in the PAS (HA3a, HA3b and HA3c) (Fig.3). Construct HA3c which carried three consecutive nucleotides complementary to the last three nucleotides of the 5'handle (pos. -6,-7,-8) conferred only ~10% of immunity and construct HA3b with three nucleotides matching the first three position of the 5´handle (pos. -1, -2, -3) ~40% with respect to the control 0M construct. Full immunity was obtained with construct HA3a having the three consecutive nucleotide matches at the position -3,-4 and -5 between PAS and the 5'handle.

The plaque forming ability of two constructs with only two matching nucleotides in the same critical region (HA2a and HA2b) was around ~5%, indicating that at least three matches between 5'-handle and PAS are necessary to achieve self from non-self discrimination in archaea.

To verify if the self/non-self recognition process needs a full match between protospacer and crRNA to be active, we designed 3 new viral vectors called 3P-6M-HA, 3P-14M-HA and 5P-3M-HA. Those vectors carried the protospacer 3P-6M, 3P-14M and 5P-3M respectively, while their PAS was mutated to match perfectly the 5´handle. All three constructs independently of the amount of mutations and their positions along the protospacer were fully protected from degradation having a >95% capability of plaques formation compared to the control virus 0M (~ 0%).

These results demonstrate that the presence of a perfectly matching PAS sequence blocks CRISPR-mediated DNA degradation also of protospacers that do not have 100% homology to the respective crRNA.

Construct	PAS Pos. -1 to -8	PFU%	± SD %
HA8	CTTTCAAT |||||||| GAAAGUUA	100	± 3.4
HA6	TATTCAAT |||||| GAAAGUUA	>95	± 15.0
HA4a	TTTTCTTA |||| GAAAGUUA	>95	± 5.3
HA4b	TATTCATA |||| GAAAGUUA	>95	± 12.0
HA3a	TATTCTTA ||| GAAAGUUA	>95	± 5.4
HA4c	CTTCATA ||| | GAAAGUUA	70	±12.8
HA3b	CTTCATTA ||| GAAAGUUA	40	± 7.5
HA3c	TACCGAAT ||| GAAAGUUA	10	± 9.0
HA2a	TACTCTTA || GAAAGUUA	5	± 5.4
HA2b	TATTGTTA || GAAAGUUA	3	± 2.9
0M	CGAGCTGA | | | GAAAGUUA	0	± 0.3
BGM	GATGATC | | GAAAGUUA	4	± 4.0

Fig.3: Self versus non-self target discrimination

Numbers in constructs (first column) refer to the numbers of matching nucleotides between protospacer adjacent sequence (PAS) and crRNA´s 5´ handle. In column 2 the similarity between PAS and 5´ handle is shown and in column 3 the plaque formation of the different constructs upon transfection are given (with respect to the 100% control HA8). At least three consecutive nucleotide matches in positions -3,-4,-5 were necessary to block interference. The nucleotide sequences of positions -2,-3 (putative PAM position) of each construct is shown with a black line above the corresponding nucleotides. The different constructs are reported in the order of their ability to escape the CRISPR system. The transfection efficiency of the different constructs is reported in column 4 with the respective standard deviation of ≥3 replicates.

7.4.4 PAM sequence and DNA interference

To elucidate the importance of the PAM in *Sulfolobus*, we used the above constructs and their protospacer flanking sequence. As described above, the construct 0M carries a 37 nucleotide long protospacer perfectly matching spacer A53. Its upstream flanking sequence contains a previously described *S. solfataricus* PAM motif (GA) (Gudbergsdottir *et al*., 2011), and this construct is fully recognized and cleaved by the CRISPR system under our test conditions.

To analyse if the DNA interference activity is affected by the presence of the PAM di-nucleotide sequence, we introduced the same 37nt protospacer into a different genetic background (pos. +283/+320 of SSO3019 lacS). The new viral vector BGM showed the di-nucleotide sequence AT in position -2 and -3. Only 5% difference in transfection efficiency was noticed between the constructs 0M and BGM (Fig.3). Both constructs were unable to efficiently infect *S. solfataricus* P2 cells bearing an active CRISPR locus demonstrating that the DNA interference activity of the CRISPR system is independent of a PAM sequence. Furthermore, constructs HA4c and HA3b (Fig.3) that were used to demonstrate the self/non-self recognition, carry the nucleotide sequence TT in positions -2 and -3. Both triggered DNA interference in *S. solfataricus* P2 and were able to escape the CRISPR system in ~70% and ~40% of the cases, respectively (Fig.3). Both viral vectors should have had the same transfection efficiency, if the reason for escape of these virions were related to the absence of a PAM sequence. As already mentioned the difference in transfection efficiency found between these two constructs can be explained by the degree of similarity between PAS (protospacer flanking sequence) and 5´handle of the spacer. These results were confirmed also by the transfection efficiency of constructs HA2a and HA2b which carry the di-nucleotide sequence AC and AT in the PAM position -2,-3 and which are nevertheless recognized and cleaved by the CRISPR system.

7.5 Discussion

7.5.1 High permissiveness of the CRISPR activity in *S. solfataricus*

Our study demonstrates the extraordinary ability of the *S. solfataricus* CRISPR-Cas system in targeting protospacers with high numbers of mismatches to the complementary crRNA. Even up to 15 mismatches between protospacer and the 3'half of the crRNA can be tolerated by the system and still trigger 50% of DNA degradation. However, similar to the findings in *E. coli* (Semenova *et al.*, 2011) also in *S. solfataricus* mutations between positions +1 and +8 ("seed" sequence) determine a rapid decrement in the interference process. In our system, viruses carrying protospacers with three mutations in the seed sequence are able to escape in 20% of the cases. The DNA interference process is almost completely abolished when more than six mutations between spacer and protospacer are present in the first 8nt of the 5'half region of the spacer. We conclude that the first 6-8nt of the spacer are of crucial importance in target recognition also in archaea, and that the seed sequence might by a characteristic of all CRISPR systems, or at least of type I systems. In archaea, but maybe also in many bacteria the system is far less stringent with respect to mutations on the opposite side of the spacer. However, we found that mutations within this (less important) protospacer region can enhance protection when mutations in the seed sequence occur in addition, thus triggering a synergistic effect. This might explain (in parts), why the length of the spacers is maintained during evolution of the CRISPR systems. Future experiments on the interference mechanism will be needed in order to fully understand this synergistic effect.
It will be interesting to elucidate if a CRISPR system can eradicate a virus from a population when it is targeted with partially matching crRNAs or if this virus will be maintained at lower levels, e.g. intracellularly at lower copy numbers. If the latter is true, the view of the CRISPR system in virus defense should be extended from a mere "first barrier defense mechanism" to a "virus defense and maintenance" system. The ability to maintain viruses at lower

levels might be beneficial permitting invasion of genetic elements that carry genes that are favourable to the host. Those viruses would probably get slowly lost if the host does not gain a selective advantage, but would be kept at low levels as long as they are beneficial.

Furthermore, maintaining a virus alive and controlling its copy number through the CRISPR system, thus avoiding excessive replication and a lytic cycle might be an effective mechanism of defense as well. Recently Swarts and colleagues demonstrated that the presence of CRISPR spacers matching the invading genetic element seems to facilitate the incorporation of additional spacers (Swarts *et al.*, 2012). In the light of this finding, the presence of spacers targeting an invading genetic element (even with low similarity) would allow the system to readily respond against virus over-proliferation. Upon infection, the virus titre could reach a threshold at which the concentration of partially matching spacer and its target would allow pairing and a first (low efficiency) degradation of the target. This degradation of the protospacer could then induce the acquisition of new spacers, now perfectly matching the viral DNA. The newly acquired spacers will then block the viral proliferation allowing the CRISPR interference cassettes to target the invader with high efficiency and to eliminate it (Swarts *et al.*, 2012). Thus, high permissibility (i.e. tolerance of mutations between spacers and protospacers) would represent an effective system against a huge diversity of genetic elements as found in *Sulfolobales* (Prangishvili *et al.*, 2006).

It will be interesting to see if the extraordinary ability of the *Sulfolobus* CRISPR-Cas system to target highly diverse protospacers is also found in other bacteria or archaea or if it remains a special feature of *Sulfolobus* or hyperthermophiles in general. We cannot rule out the possibility that the high tolerance of mutations could be specific for these organisms, because RNA-DNA hybrids do not easily form at high temperatures and might have to be specifically supported through the action of protein complexes.

7.5.2 Autoimmunity

Despite the huge variety of CRISPR-Cas systems found in different organisms, the match between an (8nt) 5´handle of the crRNA and a protospacer adjacent sequence (PAS) seems to be a common trade to avoid self-degradation. Indeed, in *Sulfolobus* the complete match between 5'handle and PAS protospacer flanking sequence blocked the interference process completely and a high degree of protection was still maintained even with only three matches in positions -3,-4,-5. Also in bacteria (although at different positions) three matches between PAS and crRNA 5´handle were the minimal requirement to protect the CRISPR locus from its own degradation (Marraffini and Sontheimer, 2010). Our result are also in general compatible with the recent finding in *S. islandicus* REY15A where similarity between crRNA 5´handle and PAS sequence conferred protection against CRISPR DNA targeting, although different from that study we found only 40% of protection when homology between position -1 and -3 was present (Deng *et al.*, 2013).

7.5.3 Role of the PAM sequence

We have also evaluated the need of a PAM sequence for triggering DNA interference. Our results indicate, that it is not necessary for triggering DNA interference, at least not for the spacer of CRISPR locus A analysed in our study. Our data strongly support the hypothesis that mutations in the PAM region decrease transfection efficiency only when they increase the similarity between the crRNA 5´handle and PAS, triggering self/non-self discrimination. Previous results in *S. solfataricus* and *S. islandicus* (Gudbergsdottir *et al.*, 2011) have demonstrated the need of a PAM sequence to trigger DNA interference. These different results could be explained by the use of different protospacers targeted by crRNAs of different CRISPR loci sub-families. *S. solfataricus* carries seven different CRISPR-Cas systems with three different repeat types and three putative CASCADE cassettes, a Csm and a Cmr module (Lillestol *et al.*, 2009, Shah and Garrett, 2011; Manica and Schleper, 2013). So far only the *in vitro* targeting activity of one Cmr complex was

elucidated (Zhang *et al.*, 2012). Interestingly, Lintner and colleagues reported that no alteration in binding affinity between *Sulfolobus* CASCADE-crRNA complex and protospacer was found in the presence or absence of a PAM motif, contrarily to what was found in other systems (Lintner *et al.*, 2011). No *in vitro* DNA targeting activity has been tested yet for the CASCADE complex studied in this work. Furthermore, it should be mentioned that the crRNAs which co-purify together with one CASCADE module (Lintner *et al.*, 2011) originate from different CRISPR loci. This underlines the possibility that different spacers could associate with different CRISPR interference complexes within the cell which might have different requirements with respect to the PAM site. If the different interference complexes discriminate the different spacers is still under debate, but it is likely that the crRNA studied here could be recognized by a CRISPR interference complex different from the one used in Gudbergsdottir experiments (Gudbergsdottir *et al.*, 2011). Indeed, *S. solfataricus* possesses a CRISPR type III-A system (Csm module) which was found in bacteria to be insensitive to the presence of a PAM site for triggering its DNA degradation activity (Marraffini and Sontheimer, 2010).

Another possible explanation for the different outcome of the two experiments rises from the different approaches used to determine CRISPR activity. In our assay an engineered virus carrying a matching protospacer is transfected into the cells and needs to proliferate and spread to the neighbour cells to trigger plaque formation. This approach differs from the plasmid based approach used in the previous studies to determine PAM function. It is possible that the presence of our vector inside the cell increases Cas expression determining higher interference activity which would perhaps not require a PAM site.

In a mutational study of *Haloferax volcanii* Fischer *et al* have shown that 6 different variants (trinucleotides) can function as a PAM when the organism is challenged with a plasmid (Fischer *et al.*, 2012). We cannot completely rule out, that also in our system there is a requirement of a PAM for target recognition, but that we do not identify it because many variants of it can be

tolerated. However, we consider it unlikely as it would mean, that we have by chance picked the right nucleotides when exchanging the existing potential "PAM" from a GA (in construct 0M) to an AT (in BGM and HA2b) or an AC (HA2a), respectively.

More work will be needed to define the role of sequences lying adjacent to the protospacer in the invader and to distinguish it in the different CRISPR types. The sequence requirements of the PAM motif might also be different for the two different functional roles (acquisition and interference). Therefore, it has recently been suggested to differentiate between a spacer acquisition motif (SAM) and a target interference motif (TIM) (Shah *et al.*, 2013). In light of this our results indicate that a TIM does not seem to play a role in target interference of our system.

7.6 Acknowledgments

This work was supported by the SulfoSYS-project (SysMo P–N-01-09-23) and by FWF project 23000 to CS.

7.7 References

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

Berkner, S., D. Grogan, S. V. Albers & G. Lipps, (2007) Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for Sulfolobus acidocaldarius and Sulfolobus solfataricus, model organisms of the (cren-)archaea. *Nucleic Acids Res* **35**: e88.

Bolotin, A., B. Quinquis, A. Sorokin & S. D. Ehrlich, (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551-2561.

Brouns, S. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin & J. van der Oost, (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.

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

Deltcheva, E., K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel & E. Charpentier, (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602-607.

Deng, L., R. A. Garrett, S. A. Shah, X. Peng & Q. She, (2013) A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. *Mol Microbiol*.

Deng, L., C. S. Kenchappa, X. Peng, Q. She & R. A. Garrett, (2011) Modulation of CRISPR locus transcription by the repeat-binding protein Cbp1 in Sulfolobus. *Nucleic Acids Res* **40**: 2470-2480.

Deveau, H., R. Barrangou, J. E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D. A. Romero, P. Horvath & S. Moineau, (2008) Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J Bacteriol* **190**: 1390-1400.

Fischer, S., L. K. Maier, B. Stoll, J. Brendel, E. Fischer, F. Pfeiffer, M. Dyall-Smith & A. Marchfelder, (2012) An archaeal immune system can detect multiple protospacer adjacent motifs (PAMs) to target invader DNA. *J Biol Chem* **287**: 33351-33363.

Grissa, I., G. Vergnaud & C. Pourcel, (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**: 172.

Grogan, D. W., (1989) Phenotypic characterization of the archaebacterial genus Sulfolobus: comparison of five wild-type strains. *J Bacteriol* **171**: 6710-6719.

Gudbergsdottir, S., L. Deng, Z. Chen, J. V. Jensen, L. R. Jensen, Q. She & R. A. Garrett, (2011) Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol Microbiol* **79**: 35-49.

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

Jonuscheit, M., E. Martusewitsch, K. M. Stedman & C. Schleper, (2003) A reporter gene system for the hyperthermophilic archaeon Sulfolobus solfataricus based on a selectable and integrative shuttle vector. *Mol Microbiol* **48**: 1241-1252.

Kurosawa, N. & D. W. Grogan, (2005) Homologous recombination of exogenous DNA with the Sulfolobus acidocaldarius genome: properties and uses. *FEMS Microbiol Lett* **253**: 141-149.

Lillestol, R. K., P. Redder, R. A. Garrett & K. Brugger, (2006) A putative viral defence mechanism in archaeal cells. *Archaea* **2**: 59-72.

Lillestol, R. K., S. A. Shah, K. Brugger, P. Redder, H. Phan, J. Christiansen & R. A. Garrett, (2009) CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Mol Microbiol* **72**: 259-272.

Lintner, N. G., M. Kerou, S. K. Brumfield, S. Graham, H. Liu, J. H. Naismith, M. Sdano, N. Peng, Q. She, V. Copie, M. J. Young, M. F. White & C. M. Lawrence, (2011) Structural and functional characterization of an archaeal CASCADE complex for CRISPR-mediated viral defense. *J Biol Chem*.

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

Manica, A. & C. Schleper, (2013) CRISPR-mediated defense mechanisms in the hyperthermophilic archaeal genus Sulfolobus. *RNA Biol* **10**.

Manica, A., Z. Zebec, D. Teichmann & C. Schleper, (2011) In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Mol Microbiol* **80**: 481-491.

Marraffini, L. A. & E. J. Sontheimer, (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**: 1843-1845.

Marraffini, L. A. & E. J. Sontheimer, (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**: 568-571.

Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez & C. Almendros, (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733-740.

Pougach, K., E. Semenova, E. Bogdanova, K. A. Datsenko, M. Djordjevic, B. L. Wanner & K. Severinov, (2010) Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Mol Microbiol*.

Prangishvili, D., P. Forterre & R. A. Garrett, (2006) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* **4**: 837-848.

Pul, U., R. Wurm, Z. Arslan, R. Geissen, N. Hofmann & R. Wagner, (2010) Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Mol Microbiol* **75**: 1495-1512.

Richter, H., J. Zoephel, J. Schermuly, D. Maticzka, R. Backofen & L. Randau, (2012) Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. *Nucleic Acids Res* **40**: 9887-9896.

Semenova, E., M. M. Jore, K. A. Datsenko, A. Semenova, E. R. Westra, B. Wanner, J. van der Oost, S. J. Brouns & K. Severinov, (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**: 10098-10103.

Shah, S. A., S. Erdmann, F. J. Mojica & R. A. Garrett, (2013) Protospacer recognition motifs: Mixed identities and functional diversity. *RNA Biol* **10**: 891-899.

Shah, S. A. & R. A. Garrett, (2011) CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. *Res Microbiol* **162**: 27-38.

She, Q., H. Phan, R. A. Garrett, S. V. Albers, K. M. Stedman & W. Zillig, (1998) Genetic profile of pNOB8 from Sulfolobus: the first conjugative plasmid from an archaeon. *Extremophiles* **2**: 417-425.

Swarts, D. C., C. Mosterd, M. W. van Passel & S. J. Brouns, (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**: e35888.

Wiedenheft, B., E. van Duijn, J. B. Bultema, S. P. Waghmare, K. Zhou, A. Barendregt, W. Westphal, A. J. Heck, E. J. Boekema, M. J. Dickman & J. A. Doudna, (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc Natl Acad Sci U S A* **108**: 10092-10097.

Zhang, J., C. Rouillon, M. Kerou, J. Reeks, K. Brugger, S. Graham, J. Reimann, G. Cannone, H. Liu, S. V. Albers, J. H. Naismith, L. Spagnolo & M. F. White, (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* **45**: 303-313.

7.8 Supplementary material

7.8.1 Supplementary figure

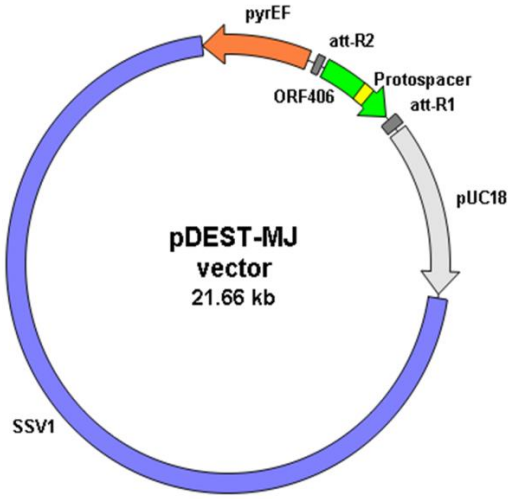
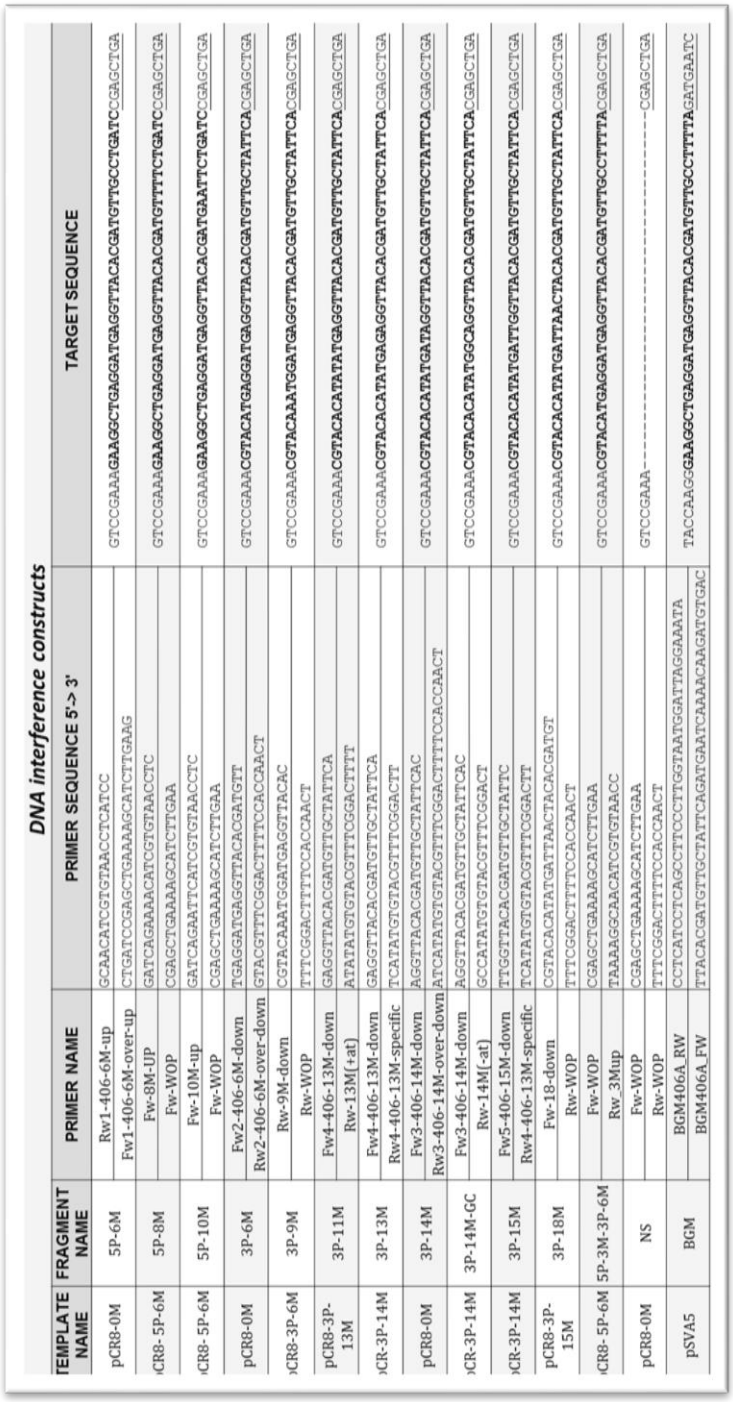


Fig.S1: Gateway destination vector pDEST-MJ, i.e. shuttle vector between *E. coli* and *S. solfataricus* based on the Sulfolobus SSV1 virus and pUC18.

7.8.2 Supplementary tables



DNA interference constructs

TEMPLATE NAME	FRAGMENT NAME	PRIMER NAME	PRIMER SEQUENCE 5'-> 3'	TARGET SEQUENCE
pCR8-0M	5P-6M	Rw1-406-6M-up	GCAACATCGTGTAACCTCATCC	GTCCGAAA**GAAGGCTGAGGATGAGGTTACACGATGTTGCCTGATC**CGAGCTGA
		Fw1-406-6M-over-up	CTGATCCGAGCTGAAAAGCATCTTGAAG	
pCR8- 5P-6M	5P-8M	Fw-8M-UP	GATCAGAAAACATCGTGTAACCTC	GTCCGAAA**GAAGGCTGAGGATGAGGTTACACGATGTTTTCTGATC**CGAGCTGA
		Fw-WOP	CGAGCTGAAAAGCATCTTGAA	
pCR8- 5P-6M	5P-10M	Fw-10M-up	GATCAGAATTCATCGTGTAACCTC	GTCCGAAA**GAAGGCTGAGGATGAGGTTACACGATGAATTCTGATC**CGAGCTGA
		Fw-WOP	CGAGCTGAAAAGCATCTTGAA	
pCR8-0M	3P-6M	Fw2-406-6M-down	TGAGGATGAGGTTACACGATGTT	GTCCGAAA**CGTACATGAGGATGAGGTTACACGATGTTGCTATTCA**CGAGCTGA
		Rw2-406-6M-over-down	GTACGTTTCGGACTTTTCCACCAACT	
pCR8-3P-6M	3P-9M	Rw-9M-down	CGTACAAATGGATGAGGTTACAC	GTCCGAAA**CGTACAAATGGATGAGGTTACACGATGTTGCTATTCA**CGAGCTGA
		Rw-WOP	TTTCGGACTTTTCCACCAACT	
pCR8-3P-13M	3P-11M	Fw4-406-13M-down	GAGGTTACACGATGTTGCTATTCA	GTCCGAAA**CGTACACATATATGAGGTTACACGATGTTGCTATTCA**CGAGCTGA
		Rw-13M(+at)	ATATATGTGTACGTTTCGGACTTTT	
pCR-3P-14M	3P-13M	Fw4-406-13M-down	GAGGTTACACGATGTTGCTATTCA	GTCCGAAA**CGTACACATATGAGAGGTTACACGATGTTGCTATTCA**CGAGCTGA
		Rw4-406-13M-specific	TCATATGTGTACGTTTCGGACTT	
pCR8-0M	3P-14M	Fw3-406-14M-down	AGGTTACACGATGTTGCTATTCAC	GTCCGAAA**CGTACACATATGATAGGTTACACGATGTTGCTATTCA**CGAGCTGA
		Rw3-406-14M-over-down	ATCATATGTGTACGTTTCGGACTTTTCCACCAACT	
pCR-3P-14M	3P-14M-GC	Fw3-406-14M-down	AGGTTACACGATGTTGCTATTCAC	GTCCGAAA**CGTACACATATGGCAGGTTACACGATGTTGCTATTCA**CGAGCTGA
		Rw-14M(-at)	GCCATATGTGTACGTTTCGGACT	
pCR-3P-14M	3P-15M	Fw5-406-15M-down	TTGGTTACACGATGTTGCTATTC	GTCCGAAA**CGTACACATATGATTGGTTACACGATGTTGCTATTCA**CGAGCTGA
		Rw4-406-13M-specific	TCATATGTGTACGTTTCGGACTT	
pCR8-3P-15M	3P-18M	Fw-18-down	CGTACACATATGATTAACTACACGATGT	GTCCGAAA**CGTACACATATGATTAACTACACGATGTTGCTATTCA**CGAGCTGA
		Rw-WOP	TTTCGGACTTTTCCACCAACT	
pCR8- 5P-6M	5P-3M-3P-6M	Fw-WOP	CGAGCTGAAAAGCATCTTGAA	GTCCGAAA**CGTACATGAGGATGAGGTTACACGATGTTGCCTTTTA**CGAGCTGA
		Rw_3Mup	TAAAAGGCAACATCGTGTAACC	
pCR8-0M	NS	Fw-WOP	CGAGCTGAAAAGCATCTTGAA	GTCCGAAA----------CGAGCTGA
		Rw-WOP	TTTCGGACTTTTCCACCAACT	
pSVA5	BGM	BGM406A_RW	CCTCATCCTCAGCCTTCCCTTGGTAATGGATTAGGAAATA	TACCAAGG**GAAGGCTGAGGATGAGGTTACACGATGTTGCCTTTTA**GATGAATC
		BGM406A_FW	TTACACGATGTTGCTATTCAGATGAATCAAAACAAGATGTGAC	

Tab.S1: List of all primers and template plasmids used for the preparation of the different protospacer constructs for the DNA interference experiments. In the first two columns the name of the inverse PCR template plasmid and the name of the final construct are given. In the third column the primer name followed by the primer sequence in column 4. The sequence 5´-> 3´ of the different protospacers is reported in column 5. In bold the nucleotides which correspond to the protospacer sequence, underlined the 8 nucleotides of the protospacer adjacent sequence.

Self non-self discrimination constructs

TEMPLATE NAME	FRAGMENT NAME	PRIMER NAME	PRIMER SEQUENCE 5'-> 3'	TARGET SEQUENCE
pCR8-0M	HA8	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCTATTCACTTTCAAT
		Fw1	CTTTCAATCACCCTTAGTGCCTGTGGAG	
pCR8-0M	HA6	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCTATTCATATTCAAT
		Fw5	TATTCAATCACCCTTAGTGCCTGTGGAG	
pCR8-0M	HA4a	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCTATTCATTTCTTA
		Fw6	TTTCTTACACCCTTAGTGCCTGTGGAG	
pCR8-0M	HA4b	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCTATTCATATTCATA
		Fw9	TATTCATACACCCTTAGTGCCTGTGGAG	
pCR8-0M	HA4c	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCTATTCACTTCAATC
		Fw2	CTTCATTACACCCTTAGTGCCTGTGGAG	
pCR8-0M	HA3a	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCTATTCATATTCTTA
		Fw7	TATTCTTACACCCTTAGTGCCTGTGGAG	
pCR8-0M	HA3b	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCTATTCACTTCATTA
		Fw2	CTTCATTACACCCTTAGTGCCTGTGGAG	
pCR8-0M	HA3c	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCTATTCATACCGAAT
		Fw3	TACCGAATCACCCTTAGTGCCTGTGGAG	
pCR8-0M	HA2a	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCTATTCATACTCTTA
		Fw-F12	TACTCTTACACCCTTAGTGCCTGTGGAG	
pCR8-0M	HA2b	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCTATTCATATTGTTA
		Fw-F14	TATTGTTACACCCTTAGTGCCTGTGGAG	
pCR8- 3P-6M	3P-6M-F8	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAACGTACATGAGGATGAGGTTACACGATGTTGCTATTCACTTTCAAT
		Fw1	CTTTCAATCACCCTTAGTGCCTGTGGAG	
CR8-3P-14M	3P-14M-F8	Rw1	TGAATAGCAACATCGTGTAACCT	GTCCGAAACGTACACATATGATAGGTTACACGATGTTGCTATTCACTTTCAAT
		Fw1	CTTTCAATCACCCTTAGTGCCTGTGGAG	
pCR8-0M	5P-3M-F8	Rw_3Mup	TAAAAGGCAACATCGTGTAACC	GTCCGAAAGAAGGCTGAGGATGAGGTTACACGATGTTGCCTTTTACTTTCAAT
		Fw1	CTTTCAATCACCCTTAGTGCCTGTGGAG	
pCR8-0M	NS	Fw-WOP	CGAGCTGAAAAGCATCTTGAA	GTCCGAAA----------CGAGCTGA
		Rw-WOP	TTTCGGATCTTTTCCGATCCGAATCT	

Tab.S2: List of all primers and template plasmids used for the preparation of the different protospacer constructs for the self/non-self discrimination experiment. In the first two columns the names of the inverse PCR template plasmid and the name of the final construct are given. In the third column the primer name is given, followed by the primer sequence in column 4. The sequence 5´-> 3´ of the different protospacers is reported in column 5. In bold the nucleotides which correspond to the protospacer sequence, underlined the mutated 8 nucleotides of the protospacer adjacent sequence.

8 GENERAL DISCUSSION

A multitude of mechanisms have been developed by bacteria and archaea to defend themselves against invading genetic elements. Recently, a new defence mechanism widespread within the prokaryotic world was discovered: the CRISPR Cas system. This system differs from all the other defence mechanisms so far studied in its ability to evolve rapidly and to remember past infections.

The CRISPR-Cas system is present in almost 90% of archaea so far analysed although the knowledge about its activity is still scarce in this domain of live. My PhD thesis was centred on the understanding of the peculiarities of the CRISPR-Cas system in the hyperthermophilic archaeon *S. solfataricus*. To date, this organism ranges among the most studied archaea with respect to biochemical and genetic investigations. Detailed bioinformatic analysis of its CRISPR loci, the development of virus-host assays to test CRISPR reaction, expression studies and characterization of *in vitro* activity of different Cas protein complexes, locate *S. solfataricus* as one of the best "model organisms" regarding the CRISPR-Cas system. Despite many favourable aspects of this organism it presents two major challenges: the impossibility to knockout specific genes in the chromosome and the presence of many and different Cas complexes encoded on the same chromosome, which could interfere with each other in the analyses. Fortunately, in closely related species knock-out protocols are available (*S. acidocaldarius* and *S. islandicus*) which will allow unravelling the peculiarities of the different modules of the *Sulfolobus* CRISPR-Cas system in the future.

S. solfataricus CRISPR-Cas systems protect the host cell from invading genetic elements carrying protospacer sequences with high similarity to a crRNA located within its CRISPR locus (Gudbergsdottir *et al.*, 2011; Manica *et al.*, 2011). On its chromosome there are six different CRISPR loci (named A to F), which carry two different types of direct repeats (Lillestol *et al.*, 2006; Lillestol *et al.*, 2009). All CRISPR loci are transcribed although CRISPR F shows very low expression compared to the other loci; probably

due to the absence of a leader sequence upstream of the repeat cluster (Lillestol *et al.*, 2009; Gudbergsdottir *et al.*, 2011; Manica *et al.*, 2011). We and others (Gudbergsdottir *et al.*, 2011) tested the activity of CRISPR A, D and F *in vivo*. CRISPR A and D express mature crRNAs and are active against extra-chromosomal elements (Gudbergsdottir *et al.*, 2011; Manica *et al.*, 2011). Spacers located within locus F, instead, are unable to confer protection, probably due to the low transcription level detected for this locus. Despite its low expression, co-purification experiments of crRNA with the CASCADE complex have shown the presence of crRNAs originating from locus F, indicating that at least some activity is to be expected (Lintner *et al.*, 2011).

To study in detail the defense reaction against infecting genetic elements, we modified the existing viral shuttle vector pMJ03 (Jonuscheit *et al.*, 2003) which was based on the *Sulfolobus* virus SSV1 and a replicon of an *E. coli* plasmid, by equipping it with a gateway recombination cassette to facilitate the addition and exchange of different protospacers. Using this vector we designed and created a series of recombinant SSV1 viruses to test the influence of protospacer sequences and their adjacent sequences on the activity in CRISPR-cas mediated protection of the host cells.

In this way we were able to demonstrate that the *S. solfataricus* CRISPR system (i) targets dsDNA; (ii) is the most promiscuous CRISPR system regarding the number of mismatches that can be tolerated between crRNA and protospacer; (iii) operates independently of a protospacer adjacent motif (PAM) and (iv) discriminates self from non-self targets by the homology between crRNA 5´handle and protospacer 3´ flanking sequence (PAS).

8.1 The Sulfolobus CRISPR-Cas system targets DNA

Until 2009, only few papers addressing the activity of the CRISPR-Cas system were available. Most of the *in-vivo* tests were performed in bacteria (*E. coli* and *S. thermophilus*) which were indicative of a crRNA-based DNA targeting activity of the CRISPR system (Barrangou *et al.*, 2007; Brouns *et al.*, 2008). At that time a study on the *in vitro* activity of one Cas module, the

Cmr, in the archaeon *P. furiosus* was published. In this publication Hale and colleagues demonstrated the ability of the Cmr module in targeting RNA *in vitro* (Hale *et al.*, 2009). This result opened a new, very speculative scenario in which the CRISPR-Cas system of archaea would target RNA instead of DNA as its bacterial counterpart.

A first objection against the exclusive RNA targeting activity of the CRISPR system in archaea arose from bioinformatic analyses of the spacers incorporated into the CRISPR loci of different archaea and bacteria. Many spacers were found to be oriented in sense with respect to their putative targets, so they would have been unable to target the respective mRNA (Mojica *et al.*, 2009). Despite this observation, Lillestol and co-workers demonstrated that in *S. acidocaldarius* and *S. solfataricus* CRISPR loci were transcribed in both orientations albeit in lower amounts in the opposite direction as would be directed by the leader sequence. In addition the locus was not only transcribed but small, probably mature antisense crRNAs were also detected (Lillestol *et al.*, 2009). Bidirectional transcription was at that time detected only in archaea (in *Sulfolobus* sp.), and this fact was corroborating the hypothesis of a CRISPR-based RNA interference mechanism in archaea.

S. solfataricus carries in proximity of one of its CRISPR loci a complete Cmr cassette (Hale *et al.*, 2009; Zhang *et al.*, 2012) although four other Cas "interference" cassettes are located on its chromosome (Shah and Garrett, 2010; Manica *et al.*, 2011). Three Cas proteins have high similarity to proteins of the CASCADE complexes analysed in *E. coli* and one shares similarity with the Csm complex previously studied in *Staphylococcus epidermidis* (Marraffini & Sontheimer, 2010). Due to the complex array of Cas proteins it was obvious to speculate that more than one mechanism operates and probably more than one type of nucleic acid could be targeted by the CRISPR system in *S. solfataricus*. Using our *in vivo* assay we were able to show, that dsDNA is a target of the CRISPR-Cas system. A viral vector carrying the ORF406 of the conjugative plasmid pNOB8 perfectly matching the crRNA 53 of locus A (CRISPR 3) was not able any more to

infect the host cells, and the targeting was independent of the transcription of the ORF. Our work, together with the results obtained by Gudbergsdottir and colleagues, who demonstrated DNA interference with a plasmid system, demonstrated for the first time that the CRISPR-system operates as an active defence mechanism in archaea and that DNA is targeted in this organism.

With our system we were also able to demonstrate that the DNA targeting activity is maintained even if the protospacer is not identical to the respective crRNA. At the time of our discovery this was very surprising due to the fact that in previous studies (Barrangou *et al*., 2007; Brouns *et al*., 2008) a single mutation abolished the DNA targeting activity of the system. In a recently published study it was demonstrated that also in *E. coli* up to four mutations can be tolerated by the system (Semenova *et al*., 2011). Different from bacteria, *S. solfataricus* possesses an unexpected ability to target protospacers with large numbers of mismatches between protospacer and the respective crRNA. We demonstrated that the CRISPR-Cas system maintains a DNA interference level of around 50% even in the presence of as many as 14 mutations between crRNA and protospacer. The effect of the different mutations with respect to their position along the protospacer was also tested. As reported for *E. coli* the homology region between protospacer and spacer 5´half was critical for the target recognition and degradation. Indeed mutations in the first six to eight nucleotides within the spacer 5´half ('seed' sequence) almost completely abolished the DNA interference activity, adding one more similarity between the archaeal CRISPR-system type I and one of its bacterial counterparts (Semenova *et al*., 2011). Mutations at the 3´ half were more tolerated by the system, its activity decreased constantly with the increment of the number of mutations. As much as eighteen mutations at the 3´half were necessary to completely abolish CRISPR targeting in *Sulfolobus*. This indicates that our system might act with one of the most promiscuous mechanisms regarding the number of mutations necessary to block DNA interference activity.

Another important characteristic of the CRISPR-Cas systems is the ability to discriminate self from non-self targets. So far it was only demonstrated for *S. epidermidis* that the CRISPR system can discriminate between invader DNAs and its own DNA based on the similarity between crRNA 5´end and protospacer flanking sequence (Marraffini & Sontheimer, 2010). In the bacterium, the presence of three complementary nucleotides at particular positions was a sufficient condition to avoid protospacer targeting and DNA degradation. In our study we were able to demonstrate that the same principle works also in *Sulfolobus*. Indeed, the presence of three matches between crRNA 5´tag and protospacer flanking sequence was a sufficient condition to block CRISPR-mediated DNA cleavage activity.

Although we demonstrated activity of the system against dsDNA molecules, RNA interference activity of *Sulfolobus* CRISPR cannot be excluded. A recent *in vitro* study of the *S. solfataricus* Cmr module has shown its ability to cleave mRNA (Zhang *et al.*, 2012). This discovery underlines the importance to develop new assays to characterize the RNA interference ability of the CRISPR system, by avoiding, at the same time, CRISPR-mediated DNA interference activity. Recent progress in our laboratory demonstrates that this is indeed possible (Zebec, Manica *et al.*, manuscript in preparation).

Now that the *in vitro* or *in vivo* activities of different Cas "interference cassettes" were demonstrated in *Sulfolobus*, one can speculate that a single crRNA could be active against different types of nucleic acids based on the Cas module in which it is loaded. In fact co-purification studies showed that both the *Sulfolobus* CASCADE and the Cmr complex can incorporate crRNA coming from different CRISPR loci, even when showing different repeat types (Lintner *et al.*, 2011; Zhang *et al.*, 2012). The presence of three different types of interference cassettes located on the *Sulfolobus* chromosome (CASCADE, Csm, and Cmr) indicates potential functional diversity of the CRISPR complexes. Based on the comparison between *Sulfolobus* Cas complexes and their bacterial counterpart we can speculate that both CASCADE and Csm complexes could be involved in the recognition

and targeting of DNA while the Cmr module is implicated in the control of RNA invaders.

Acid hot springs are rich in mobile genetic elements, such as viruses and conjugative plasmids (Prangishvili *et al.*, 2006) and it is probably for this reason that *S. solfataricus* possesses a very extensive CRISPR-Cas array which occupies almost 10% of its genome. But what makes the CRISPR-Cas system so special as a defence mechanism? What is the advantage of possessing a CRISPR-Cas System instead of restriction-modification systems? And what are the disadvantages?

The ability of the system to evolve fast could be the key of CRISPR success. Integrating a new spacer after each viral invasion has the potential to block the infection, destroy the invaders and add a memory in the host chromosome to avoid subsequent infection of other genetic elements carrying the same (or a similar) protospacer sequence. Nevertheless, the integration of a new spacer increases the fitness cost for the host. Furthermore, the integration of new spacers must be preceded by the recognition of the invading genetic element. If the cells are recognizing the invader why don´t they target it immediately without integrating a new spacer? Increasing fitness cost?

A first explanation might lie in the scarce ability of the host to recognize the invader. The CRISPR-Cas system can be divided temporally into two phases: the acquisition phase and the interference phase. During the acquisition phase a small fragment of the invader DNA is integrated into the host chromosome as a spacer that is successively transcribed and used as "probe" to sense and target the invading genetic element. In order to integrate a new DNA fragment of foreign origin the system must recognize, cleave and integrate the new DNA fragment into the chromosome. The first two steps of acquisition (recognition and cleavage of the target) are the same (conceptually) steps of the interference phase. Therefore, if the system would recognize and cleave the invading genetic element with high efficiency, a second energy expensive phase (interference) would not be necessary. The low ability to recognize the invading genetic element during the acquisition

phase could be related to the potential danger of the integration of chromosomal DNA, which would lead to the dead of the host cell.

A possibility to reduce the spacer acquisition efficiency of the CRISPR system could lie in the inability of the protein involved in the acquisition phase to cleave intact double stranded DNA.

This hypothesis is supported by the recent finding that the integration of new spacers is stimulated by the presence of truncated invaders matching spacers in the host chromosome (Swarts *et al.*, 2012), despite the fact that *E. coli* is able to integrate new spacers also when the CRISPR-Cas interference was abolished (Yosef *et al.*, 2012).

A second explanation could be the impossibility of the host to amplify the interference signal without integrating a new spacer. No data are available to support the possibility of multiple DNA being targeted by a single crRNA-Cas complex. It is still not clear if crRNAs are cleaved during the interference process together with the target DNA. The impossibility of multiple targeting loops for a single crRNA-Cas complex would decrease the ability of the system to defend the cell from the invading genetic element without a continued production of new crRNAs. The third possibility is the least explored: The presence of spacers with low similarity to an invader could not only stimulate the integration of new spacers but could also trigger a decrease in the intracellular titre of the invading genetic element (Swarts *et al.*, 2012). Perez-Rodriguez and co-workers demonstrated that in *E.coli* the CRISPR-Cas system is able to decrease the copy number of a plasmid carrying protospacers with a high degree of mutations (only 12-15 matches at the 5´half of the crRNA) to their respective crRNA when the CRISPR activity is stimulated by envelope stress. This result underlines a property of the CRISPR system that so far was left aside: the ability to control and not only to destroy an invader. Horizontal gene transfer is a key process for many organisms to acquire “new genes” and new properties to survive in difficult conditions. Genetic elements like plasmids and viruses are very important for this natural process. For this reason the ability of the CRISPR-Cas system to control the intracellular titre of a genetic element could be a key

aspect for the high diffusion of this defense mechanism within prokaryotes. It is easy to imagine a scenario where an invading genetic element could infect a cell and its intracellular titre would be "controlled" by low similarity crRNAs located within a CRISPR locus. In this way the system would not block the acquisition of beneficial genes which could be present on the chromosome of the invaders. At the same time the CRISPR-Cas system would be able to avoid hyper-proliferation of the genetic element. If environmental conditions increased viral proliferation above a critical threshold the CRISPR-Cas system could recognize the invader and integrate a new perfectly matching spacer in its chromosome. The new crRNA produced will then be able to target the invaders with high efficiency thus blocking its proliferation. By avoiding the immediate incorporation of a new spacer at each infection, i.e. avoiding the eradication of an invader, the system would also lower the selective pressure which exerts on the invaders. At the same time the fitness costs associated with the incorporation of new spacers in the host chromosome would decrease.

8.2 CRISPR vs. Eukaryotic RNAi: a brief comparison

The CRISPR-Cas system can be divided into three fundamental steps: spacer acquisition, CRISPR expression and processing, and interference. Those three steps are closely linked to each other, although they could express their function independently. Indeed spacer acquisition is independent from the expression of the CRISPR locus as well as from the presence of interference complexes (Yosef *et al.*, 2012). At the same time CRISPR expression and processing is carried out by a different type of proteins which are not involved in the acquisition or interference process (at least in the CRISPR type I and III) (Makarova *et al.*, 2011). Only the interference process seems to be functionally linked with the CRISPR expression and processing, although it cannot be excluded that certain interference complexes could interact with other ncRNAs or produce ncRNAs directly targeting the invader as seen for the Eukaryotic RNAi mechanism.

Several papers have underlined the difference and similarity between the CRISPR-Cas system and the Eukaryotic RNAi machinery (van der Oost & Brouns, 2009; van der Oost *et al.*, 2009; Makarova *et al.*, 2009). The two systems are remarkably different when comparing their enzymatic machinery whereas they share striking similarities at the functional level. There are three crucial points in the activity of these two defence mechanisms in blocking viral infection and spreading: (i) recognition of the invading genetic element, (ii) amplification of the silencing signal, and (iii) targeting of the extra-chromosomal element. In the Eukaryotic RNAi (EuRNAi) the recognition of the invading genetic element is mediated by the formation of double stranded RNA (dsRNA), which is recognized and processed by a special enzyme (Dicer) into small RNA (siRNAs)(Ghildiyal and Zamore, 2009). Once produced, the siRNA guides the RISC complex against the invading genetic elements. Furthermore the siRNA can be used as a primer for an RNA-dependent RNA polymerase which produces additional dsRNA (from the invaders) (Bouche *et al.*, 2006). In addition it will be involved in the production of new siRNA and in this way in the amplification of the interference signal.

So far it is unknown how the CRISPR system senses the invading genetic element. Due to the *in vitro* activity of the proteins involved in the acquisition process it is difficult to speculate that dsRNA could be the trigger for viral recognition. After incorporation, the CRISPR spacer is transcribed by a promoter-like sequence located upstream of the CRISPR locus. The CRISPR pre-crRNA produced is then processed into mature crRNA, and many crRNAs are produced from a single spacer, amplifying the interference signal. Interestingly, the most recognized difference between EuRNAi and CRISPR reside in this step. In my opinion, the primary function of the incorporation of new spacers lies in amplification of the silencing signal. Clearly the acquisition of new spacers gives also the possibility to remember past infections which is without doubt a key aspect of the CRISPR system. Nevertheless, I would not consider this as a key difference to the EuRNAi. My hypothesis rises from the fact that incorporation of spacers in the host

genome would be disadvantageous to higher eukaryotes, increasing the fitness cost in maintaining additional genome sequences in a single cell without any beneficial effect for the "whole organism". When considering multicellular organisms, we should take into account that only few cells give progeny. Due to this reason the incorporation of viral sequences into the cell will not necessarily transfer the resistance to successive generations, but it will only be confined in the previous infected cell. Nevertheless, the EuRNAi had to maximize its ability to rapidly sense and respond to the infection, instead of incorporating new DNA sequences and to remember past infection. Due to the impossibility to transmit the resistance trait to the progeny, eukaryotic cells have developed mechanisms to provide the systematic spread of the silencing signal through the whole organism. On the contrary the integration of new spacers in the host genome of a unicellular organism would give a selective advantage to the whole population derived from that single cell, protecting the progeny from the attack of similar genetic elements.

Furthermore, the crRNAs were recently compared also to Piwi RNAs (PiRNA) (Ghildiyal and Zamore, 2009; Kumar and Chen, 2012). Piwi RNAs and proteins were found highly expressed in animal germ lines where they are involved in the control of transposable genetic elements. Whereas the proteins involved do not share similarity with the CRISPR counterpart, there is a logic similarity between the two systems (Kumar and Chen, 2012): In both cases nucleic acid of the invading genetic elements are incorporated into particular loci after the transcription and these guide the interference machinery to the invading genetic element in *trans* (Kumar and Chen, 2012). The incorporation of new DNA sequences into particular genomic loci that can be transmitted to the next generation might be a good example for Lamarkian evolution. The addition of new sequences in the Piwi locus was only found in animal germlines underlining the advantage of incorporating new sequences only when the beneficial character can be transmitted to the next generations.

Taken together these observations could favour the hypothesis of a common evolution of CRISPR, EuRNAi and Piwi, where the common CRISPR ancestor was then evolving into Piwi or EuRNAi systems. A different selective pressure could have been active on the CRISPR system based on the cell type. Cells with the ability of transmitting new characters to the progeny have maintained the trait of incorporating new DNA sequences into their genomes. Differentiated cells, instead, have developed a different mechanism to amplify and spread the signal through the whole organism.

8.3 References

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

Bouche, N., D. Lauressergues, V. Gasciolli & H. Vaucheret, (2006) An antagonistic function for Arabidopsis DCL2 in development and a new function for DCL4 in generating viral siRNAs. *EMBO J* **25**: 3347-3356.

Brouns, S. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. Slijkhuis, A. P. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin & J. van der Oost, (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.

Ghildiyal, M. & P. D. Zamore, (2009) Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10**: 94-108.

Gudbergsdottir, S., L. Deng, Z. Chen, J. V. Jensen, L. R. Jensen, Q. She & R. A. Garrett, (2011) Dynamic properties of the Sulfolobus CRISPR/Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol Microbiol* **79**: 35-49.

Hale, C. R., P. Zhao, S. Olson, M. O. Duff, B. R. Graveley, L. Wells, R. M. Terns & M. P. Terns, (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**: 945-956.

Jonuscheit, M., E. Martusewitsch, K. M. Stedman & C. Schleper, (2003) A reporter gene system for the hyperthermophilic archaeon Sulfolobus solfataricus based on a selectable and integrative shuttle vector. *Mol Microbiol* **48**: 1241-1252.

Kumar, M. S. & K. C. Chen, Evolution of animal Piwi-interacting RNAs and prokaryotic CRISPRs. *Brief Funct Genomics* **11**: 277-288.

Kumar, M. S. & K. C. Chen, (2012) Evolution of animal Piwi-interacting RNAs and prokaryotic CRISPRs. *Briefings in functional genomics* **11**: 277-288.

Lillestol, R. K., P. Redder, R. A. Garrett & K. Brugger, (2006) A putative viral defence mechanism in archaeal cells. *Archaea* **2**: 59-72.

Lillestol, R. K., S. A. Shah, K. Brugger, P. Redder, H. Phan, J. Christiansen & R. A. Garrett, (2009) CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties. *Mol Microbiol* **72**: 259-272.

Lintner, N. G., M. Kerou, S. K. Brumfield, S. Graham, H. Liu, J. H. Naismith, M. Sdano, N. Peng, Q. She, V. Copie, M. J. Young, M. F. White & C. M.

Lawrence, (2011) Structural and functional characterization of an archaeal CASCADE complex for CRISPR-mediated viral defense. *J Biol Chem*.

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

Makarova, K. S., Y. I. Wolf, J. van der Oost & E. V. Koonin, (2009) Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol Direct* **4**: 29.

Manica, A., Z. Zebec, D. Teichmann & C. Schleper, (2011) In vivo activity of CRISPR-mediated virus defence in a hyperthermophilic archaeon. *Mol Microbiol* **80**: 481-491.

Marraffini, L. A. & E. J. Sontheimer, (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**: 568-571.

Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez & C. Almendros, (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733-740.

Prangishvili, D., P. Forterre & R. A. Garrett, (2006) Viruses of the Archaea: a unifying view. *Nat Rev Microbiol* **4**: 837-848.

Semenova, E., M. M. Jore, K. A. Datsenko, A. Semenova, E. R. Westra, B. Wanner, J. van der Oost, S. J. Brouns & K. Severinov, (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci U S A* **108**: 10098-10103.

Shah, S. A. & R. A. Garrett, (2010) CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. *Res Microbiol* **162**: 27-38.

Swarts, D. C., C. Mosterd, M. W. van Passel & S. J. Brouns, (2012) CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**: e35888.

van der Oost, J. & S. J. Brouns, (2009) RNAi: prokaryotes get in on the act. *Cell* **139**: 863-865.

van der Oost, J., M. M. Jore, E. R. Westra, M. Lundgren & S. J. Brouns, (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* **34**: 401-407.

Yosef, I., M. G. Goren & U. Qimron, (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res*.

Zhang, J., C. Rouillon, M. Kerou, J. Reeks, K. Brugger, S. Graham, J. Reimann, G. Cannone, H. Liu, S. V. Albers, J. H. Naismith, L. Spagnolo & M. F. White, (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* **45**: 303-313.

9 CURRICULUM VITAE

Personal Information:

Name:	Andrea Manica
Date of Birth:	23th December 1982
Place of Birth:	Rovereto (TN) Italy
Private address:	Engerthstrasse 213/23 1020 Wien

Education:

2007-2013	PhD studies at the University of Vienna
2004-2006	Master Degree "Nutritional Biotechnology" University of Padova, Italy Marks: 110/110 cum laude
2001-2004	Bachelor Degree "Plant Biotechnology" University of Padova, Italy Marks: 110/110 cum laude
1996-2001	High school degree "Farm Advisor" Istituto di S.Michele a/A (TN) Italy Marks: 95/100

Stays abroad and other Education:

2006-2007	Student assistant Umea Plant Science Centre, Umea, Sweden
2004-2005	Erasmus students Universidad Politecnica de Valencia, Spain

Workshop, Talks and Poster Presentation:

2013	*International conference*: "CRISPR evolution, mechanisms and infection" 17–19/06/2013 St. Andrews, Scotland UK. Poster: Unexpectedly Broad Target Recognition in CRISPR-Mediated DNA Interference of the Archaeon *Sulfolobus solfataricus*

2012 *International conference*: "Molecular Biology of Archaea 3".
02-04/07/2012 Marburg, Germany.
Poster: CRISPR Activity and Viral Defence: increasing the understanding of Sulfolobus solfataricus crRNA target recognition.

2010 *International Workshop*: "Biodiversity, Molecular Biology and Biogeochemistry of Thermophiles".
10-18/08/2012 Petropavlovsk-Kamchatsky, Russia.
Talk: The Archaeal CRISPR-Cas system.

2010 *International workshop for sampling archaea from terrestrial hot springs.*
12-16/06/2010 Iceland.

2009 *EMBO international workshop*: "New Functions of Regulatory RNAs in Pro- and Eukaryotes".
13-15/01/2009 Vienna, Austria.

Language skills:

Italian: mother tongue
English: very good
Spanish: good
German: basic

Publications:

1. **SulfoSYS (Sulfolobus Systems Biology): towards a silicon cell model For the central carbohydrate metabolism of the archaeon Sulfolobus solfataricus under temperature variation.**
Albers SV, Birkeland NK, Driessen AJ, Gertig S, Haferkamp P, Klenk HP, Kouril T, **Manica A**, Pham TK, Ruoff P, Schleper C, Schomburg D, Sharkey KJ, Siebers B, Sierocinski P, Steuer R, van der Oost J, Westerhoff HV, Wieloch P, Wright PC, Zaparty M.
Biochem Soc Trans. 2009 Feb;37 (Pt 1):58-64. doi: 10.1042/BST0370058.

2. **"Hot standards" for the thermoacidophilic archaeon Sulfolobus solfataricus.**
Melanie Zaparty; Dominik Esser; Susanne Gertig; Patrick Haferkamp; Theresa Kouril; **Andrea Manica**; Pham Trong K; Julia Reimann; Kerstin Schreiber; Pawel Sierocinski; Daniela Teichmann; Marleen van Wolferen; Mathias von Jan; Patricia Wieloch; Sonja V Albers; Arnold J M Driessen; Hans-Peter Klenk; Christa Schleper; Dietmar Schomburg; John van der Oost; Phillip C Wright; Bettina Siebers.

Extremophiles. 2010 Jan;14(1):119-42. doi: 10.1007/s00792-009-0280-0. Epub 2009 Oct 4.

3. **In vivo Activity of CRISPR-Mediated Virus Defense in a Hyperthermophilic Archaeon Sulfolobus solfataricus.**
   Andrea Manica, Ziga Zebec, Daniela Teichman, Christa Schleper.
   Mol Microbiol. 2011 Apr; 80(2):481-91. doi: 10.1111/j.1365-2958.2011.07586.x. Epub 2011 Mar 8.

4. **Antisense regulation by transposon-derived RNAs in the hyperthermophilic archaeon Sulfolobus solfataricus.**
   Märtens B, Manoharadas S, Hasenöhrl D, **Manica A**, Bläsi U.
   EMBO Rep. 2013 Jun; 14(6):527-33. doi: 10.1038/embor.2013.47. Epub 2013 Apr 12.

5. **CRISPR-mediated defense mechanisms in the hyperthermophilic archaeal genus Sulfolobus.**
   Manica A, Schleper C.
   RNA Biol. 2013 May 1; 10(5):671-8. doi: 10.4161/rna.24154. Epub 2013 Mar 27.

6. **Unexpectedly Broad Target Recognition Ability of the CRISPR-mediated Virus Defense System in *Sulfolobus solfataricus*.**
   A. Manica*, Z. Zebec*, J. Steinkellner, and C. Schleper
   (*co-first authorship)
   Manuscript under revision

7. **CRISPR-mediated targeted mRNA degradation in vivo**
   Ziga Zebec*, **Andrea Manica***, Jing Zhang, Malcolm White, Christa Schleper.
   (*co-first authorship)
   Manuscript submitted

10 ACKNOWLEDGEMENT

My PhD started in November 2007 and now after six years I can say that one of the most challenging chapters of my life is coming to an end.
During this time I met amazing people who have helped me to grow as scientist and as a person. I would therefore like to take the chance to thank each of them for their support.

I thank Prof. Christa Schleper to have chosen me as one of her PhD students and for the possibility to work on such an exciting project like the CRISPR, for the sampling trips, and for our discussions.
I would like to thank Michael and Rebecca Terns to introduce me to the CRISPR world and for the support given during the first part of my PhD.
Udo Bläsi and Birgit Martens for the fruitful collaboration on Sulfolobus small RNAs, it was a pleasure to collaborate with you.
I thank Prof. Bettina Siebers, I am grateful to you for convincing me to go back to university again. Without your call, I would not be here today.
A special thank goes to Ziga, it was really a pleasure to work next to you for the last four years. Thanks for all your help, discussion and ideas; we have been a great team.
A big thank goes to my friends and (ex)colleagues of the Dep. of Genetics in Ecology and especially to Gabriel, Clarissa and Michaela, who were always "happy" to listen to me when I needed a friend to discuss about life and work.
I would like to thank Daniela and Rita for their friendship, for their professional inputs, for the endless hours spent discussing about work and philosophizing about life, for the coffee breaks and the late "lab-night pizzas". Thanks to all of you!
Big thanks I owe to my family, who always gave me the support needed and the space to follow my dreams.
Most of all I would like to thank you Beate, it was not easy to stand my moods and my fears during the last 5 years. I am grateful to have you by my side.