DIPLOMARBEIT

Required base complementarity
between guide RNAs and target sequences
in the CRISPR/Cas-based virus defense mechanism
of the hyperthermophilic archaeon *Sulfolobus solfataricus*

Verfasserin

Julia Steinkellner

angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag.rer.nat.)

Eigenständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen nicht benützt und die den benützten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Ort, Datum..Unterschrift...

Acknowledgement

First of all, I would like to express my deepest gratitude to my supervisor Dr. Christa Schleper who gave me the chance to be part of her CRISPR/Cas-research group. I am very happy that you gave me the possibility to study and learn more about this fascinating and relatively young field in microbiology. Many thanks that you made this work possible, even though I started as a student being quite inexperienced in the application of molecular techniques in microbiology!

Additionally, I am heartily thankful to Ziga Zebec. This thesis would not have been possible for me without your invaluable help at any times. Thank you for providing me with some additional SSV1 viral-vector constructs that helped me to complete this work and your great support in performing lots of plaque assays. Moreover, I am very grateful for showing me how to perform all the lab techniques and especially for being so patient with me and trying to answer all of my questions! You always had an open ear for me, encouraged me and also helped me to cope with negative results. It was really great to work with you not only for making me laugh which made "lab-life" easier for me! Thank you soooo much, Ziga!!!!!!

I am also very thankful to other members of the research-group- Romana, Dani, Michaela, Anja, Andrea, Pierre, Ricardo, Andi, Clarissa, Lisa, Nica, Amir, Melina and Nico who were always willing to help and give valuable advice.
Special thanks to my lab colleagues Anna, Chrisi and Ina for all the nice food you shared with me and having lunch with me! All of you made my "lab-life" so much easier!

Ein ganz herzlicher Dank gilt natürlich meinen Eltern, die mich all die Jahre während meines Studiums unterstützt haben. Ihr hattet immer ein offenes Ohr für mich, auch wenn ihr nicht immer ganz verstehen konntet, was ich denn da eigentlich den ganzen Tag im Labor so mache. Ihr habt mich immer ermuntert und an mich geglaubt, auch wenn ich oftmals den Glauben an mich selbst verloren hatte. Ohne euch wäre der

Wunsch vom Biologiestudium nie möglich gewesen! Danke für Alles (und die vielen Telefonate, die mir die Zeit in Wien sehr erleichtert haben)!!!!!

Einen ganz lieben Danke möchte ich zum Schluss an Markus aussprechen, der mich während meiner Zeit im Labor immer unterstützte und meine vielen Jammereien über sich „ergehen lassen" musste. Auch wenn es nicht immer leicht mit mir war, hast du sehr viel Geduld gezeigt und mir geholfen auf andere Gedanken zu kommen. Die gemeinsame Zeit mit dir hat mir geholfen neue Energie zu tanken und manches nicht ganz so ernst zu nehmen. Danke, dass du für mich da warst!!!

Wien, 2012

Julia Steinkellner

Table of contents

1 Abstract

In natural environments microorganisms are constantly exposed to a great diversity of viruses and other invasive genetic elements. Consequently, microbes have evolved various defense strategies in order to prevent viral attack and invasion of foreign DNA, respectively. Clustered regularly interspaced short palindromic repeats (in short: CRISPR) characterize one of those defense mechanisms and are found in many bacterial and most archaeal genomes. CRISPR/Cas systems confer acquired resistance against viruses and plasmids by specifically targeting invasive nucleic acids via sequence complementarity between small spacer-derived CRISPR RNAs (crRNAs) and so called protospacer sequence of the invading (targeted) nucleic acids. Conversely, viruses can find a way to escape CRISPR/Cas resistance by acquisition of mutations within their protospacer sequence. Interestingly, not all protospacer mutations lead to viral escape. So far the impact of protospacer mutations on CRISPR/Cas-mediated interference has mainly been studied in bacterial CRISPR/Cas systems. To learn more about the requirements of crRNA-protospacer interaction during CRISPR/Cas-mediated interference in Archaea, the hyperthermophilic archaeon *Sulfolobus solfataricus* and its virus SSV1 were used as a model system in this study. Various protospacer mutants of the recombinant SSV1 virus were tested in transfection experiments. The engineered mutants carried increasing numbers of mutations at the two different ends of the protospacer sequence, i.e. six up to 18 consecutive mutations at both ends, respectively. The *S. solfataricus* CRISPR/Cas system conferred almost 100 % immunity against the recombinant virus containing six mutations at the "down"-end of the protospacer, i.e. at the 3' end with respect to transcription of the spacer RNA. In contrast, viral resistance was severely reduced (by 75 %) when six mutations were introduced at the other end ("up"-end) of the protospacer. Still 50 % viral resistance was observed when 15 mutations were present at the protospacer "down"-end. Only 18 mutations at the "down"-end of the protospacer led to almost 100 % viral escape. Similar to observations recently made in *E. coli*, the CRISPR/Cas system was significantly more tolerant towards mutations present at the "down"-end of the protospacer than at the other end, suggesting that CRISPR/Cas interference in *S.*

solfataricus might be initiated by the recognition of a short sequence at the "up"-end of the protospacer. In contrast to the bacterial system however, the required base complementarity between the target sequence and crRNA during CRISPR/Cas interference seemed to be very low. This might indicate an adaptation to fast evolving target sequences, thus providing the host organism with a more flexible recognition of related invading elements.

2 Introduction

2.1 From the very beginnings of microbiology to the evolutionary tree of life

Even though microorganisms are the smallest form of life, they play central roles in many geochemical processes and are known to be major drivers of nutrient recycling and degradation of organic matter. The study of microorganisms does not only help to understand cellular processes but deals also with many important issues in medicine, agriculture and industry (e.g. production of antibiotics and human proteins). Although microorganisms are the most abundant entities on earth, microbial cells have not been discovered until the invention of the microscope. In 1676 Antoni van Leeuwenhoeck recognized microorganisms in a simple microscrope for the first time. Nevertheless further progress was only made in the mid- to late nineteenth century primarily driven by Louis Pasteur and Robert Koch. Pasteur's work enabled the development of sterilization methods not only for applications in microbiological research but also for the purpose of food preservation (see Madigan and Brock 2009). Robert Koch was the first to prove the theory that specific microorganisms are linked to the occurrence of specific diseases. Besides, he was able to grow organisms in pure culture by the usage of solid media (Madigan and Brock 2009). The elucidation of the genetic code by James D. Watson and Francis Crick represents another landmark in microbiology (Watson and Crick 1953). In the following years the field of molecular microbiology developed. New technologies like nucleic acid sequencing techniques and polymerase chain reaction discovered by Kary B. Mullis (Mullis and Faloona 1987) revolutionized microbiology and introduced molecular phylogeny (Madigan and Brock 2009). This method tries to resolve evolutionary relationships among organisms by the comparison of homologous genes and led to the discovery of a new domain of life named the Archaea (formerly Archaebacteria) by Carl Woese (Woese and Fox 1977). By using *rRNA*-based (*ribosomal RNA*) phylogenetic analysis Carl Woese identified the three domains of life: *Eukarya*, *Bacteria* and *Archaea* (Woese and Fox 1977; Woese *et al.,* 1990). Due to this finding all cellular organisms known to date can be divided into the three domains. The

combination of advanced molecular technologies and the recognition of phylogenetic relationships enabled the identification of microorganisms without the need for cultivation and revealed an enormous microbial diversity in all environments on Earth. These insights also gave rise to a new subdiscipline of microbiology, the microbial ecology, making us aware of the substantial impact that microorganisms have on the functioning of ecosystems.

2.2 *Archaea*-the third domain of life

2.2.1 Phylogeny of *Archaea*

Considering the phylogenetic tree of life, it seems that two main lineages diverged from a last common ancestor suggesting that all cellular organisms can be traced back to one common origin (Woese *et al.,* 1990). One of those two lineages includes the domain of *Bacteria* whereas the second main lineage most likely split again to yield the two domains: *Eukarya* and *Archaea*.

Even though *Archaea* are often associated with extreme environments continuously increasing sequence information of uncultured organisms indicate their widespread occurrence both in extreme and moderate habitats. Based on today's state of knowledge *Archaea* can be divided into at least five major phyla: the *Crenarchaeota*, the *Euryarchaeota*, the *Korarchaeota*, the *Nanoarchaeota* and the *Thaumarchaeota*. The hypothesis of the existence of the more recently discovered archaeal phylum *Thaumarchaeota* (Brochier-Armanet *et al.,* 2008) could eventually be supported with genome sequencing of the ammonium-oxidizing archaea *Nitrosopumilus maritimus* and *Nitrososphaera gargensis* (Spang *et al.,* 2010). The *Crenarchaeota* mainly comprise hyperthermophilic organisms often found in terrestrial hot springs and hydrothermal vents as for example the model organism *Sulfolobus solfataricus* (Zillig *et al.,* 1980). The Euryarchaeota are quite diverse and include thermophilic, halophilic as well as methanogenic organisms. *Korarchaeota* and *Nanoarchaeota* belong to two more phyla that have been proposed. So far the *Nanoarchaeota* comprise only one cultured representative, the hyperthermophilic *Nanoarchaeum equitans* (Huber *et al.,* 2002).

4

Beside the enrichment culture of an organism representing a member of the *Korarchaeota* (Elkins *et al.,* 2008) no cultivated member of this group is known so far.

2.2.2 Major archaeal characteristics-What separates *Archaea* from *Bacteria* and *Eukarya*?

Similar to *Bacteria*, members of *Archaea* predominantly consist of single cells that do not contain a nucleus or other cell compartments which are characteristic of eukaryotic cells. Unlike eukaryotic DNA which is present in linear form within the nucleus, the DNA of *Archaea* and *Bacteria* is predominantly circular. Both, archaeal and bacterial genes are arranged in clusters or operons. Genes of *Archaea* involved in diverse metabolic pathways often reveal greater similarity to those of *Bacteria,* which might reflect the need for similar metabolic capabilities of organisms inhabiting the same habitat. However, DNA packaging and information processing (including genes and mechanisms involved in replication, transcription and translation) resemble those of *Eukarya* indicating that *Archaea* are probably more closely related to *Eukarya* than to *Bacteria* (Forterre *et al.,* 2002).

The composition of the archaeal cytoplasmatic membrane differs from both *Bacteria* and *Eukarya*. While bacterial and eukaryotic lipids are esters of fatty acids and glycerol, lipids of *Archaea* contain phytanyl or biphytanyl chains that are linked to glycerol by an ether bond (De Rosa *et al.,* 1986). Beside the cytoplasmatic membrane, the composition of the archaeal cell wall differs from that of *Bacteria* as they lack murein which is present in the cell wall of most *Bacteria*. *Archaea* exhibit a great diversity of cell wall compounds including proteins, polysaccharides or *pseudomurein*, a polysaccharide that is very similar to peptidoglycan (Kandler and Konig 1998).

2.3 Viruses

2.3.1 General properties of viruses

Viruses do not carry out independent metabolic activity and thus are often not considered as living organisms. Nonetheless, they contain nucleic acid primarily

encoding for genes that are essential for successful propagation. Viruses are dependent on the infection of cellular organisms providing energy and enzymatic "equipment" needed for transcribing and replicating the viral genome. Once several copies of the virus genome as well as virus-derived proteins have been produced, the replicated genomes are packed into protein coats which finally lead to the formation of virions. These are released in order to infect new host cells. Viruses have the ability to infect members of all three domains of life whereas a specific virus is typically restricted to infect only a small group of closely related organisms. Contrary to cellular organisms, viruses use DNA, RNA or even both as genetic material that can be either single- or double-stranded. They are very diverse in terms of size and shape whereas their genomes are generally smaller than those of cellular organisms. With respect to their life cycle viruses can either be virulent or temperate. Virulent means that host cells are lysed upon propagation of the virus. Temperate viruses can enter the lysogenic pathway in which their genome integrates into the host genome and is passed as a so called prophage from one generation to the next. Dependent on host performance and environmental conditions, lysogenic viruses can become lytic finally leading to host destruction. Even though all viral infections exert stress on their host cells to various extents, some of them turn out to be even advantageous and the relationship between virus and host might better be described as mutualistic (Roossinck 2011).

2.3.2 Ecological and evolutionary impact

Viral numbers significantly exceed the abundance of cellular organisms while their numbers vary strongly between environments. Various observations have indicated the importance of viruses on the functioning and dynamic of diverse ecosystems. For example the release of nutrient matter driven by viral induced cell lysis can have great impact on microbial food webs and biogeochemical cycles (Weinbauer 2004). Fast spreading of viruses and phages in dominant populations are even considered as driving force to prevent low abundant species from being outcompeted known as 'killing the winner' hypothesis and by this, might help to sustain high species richness (Winter *et al.,* 2010). Additionally, viruses substantially contribute to horizontal gene transfer between prokaryotic cells. Virus-mediated gene transfer called transduction

means that parts of the host chromosome are accidentally packed inside the mature virion and are subsequently transferred into another organism after infection. When recombination with the host chromosome occurs, the foreign DNA is integrated into the host genome (Madigan and Brock 2009). Rapidly increasing information on whole-genome sequences revealed the importance of virus-mediated gene transfer and raised the hypothesis that viruses exhibit important drivers for the evolutionary change of prokaryotic organisms (Weinbauer and Rassoulzadegan 2004).

2.3.3   Archaeal viruses

Even though Archaea are widely distributed on Earth, their associated viruses have mainly been studied in extreme hydrothermal and hypersaline habitats. All archaeal viruses known so far have dsDNA genomes but are highly diverse in terms of their morphologies. In contrast to viruses of the phylum *Euryarchaeota* which predominantly resemble the common non-enveloped head-tail morphotype of bacteriophages, viruses infecting *Crenarchaeota* exhibit very diverse and even unique morphologies. Those include for example spherical, rod-shaped, bottle-shaped, spindle-shaped and filamentous structures (Prangishvili *et al.,* 2006). The majority of archaeal viruses that have been isolated from acidothermophilic environments seem to be non-lytic and rather persist stably in the host cell by multiplying and releasing virus particles continuously.

2.4   The CRISPR/Cas system - A newly discovererd viral defence mechanism in bacteria and archaea

To our knowledge viruses present the most abundant entities on earth, whose attacks microorganisms are facing continuously. Considering this, defence strategies against virus infections being often detrimental to their host seem to be crucial to assure efficient proliferation of bacterial and archaeal species, respectively. The discovery of the so-called CRISPR mechanism was an important progress in this respect and describes the acquisition of phage resistance in both bacteria and archaea. CRISPR is

the abbreviation for **c**lustered **r**egularly **i**nterspaced **s**hort **p**alindromic **r**epeats and designates an array of highly similar short direct repeats that are separated by non-repetitive spacer sequences widely spread in microbial genomes.

2.4.1   The history of CRISPR is comparatively short

CRISPR-research represents a relatively young field in microbiology. Even though the existence of a CRISPR sequence was first described in 1987 (Ishino *et al.,* 1987), its function was only studied very recently. Increasing genome information together with computational analysis revealed that CRISPR sequences are widely distributed in bacterial and archaeal genomes indicating that CRISPR-arrays might provide an important biological function (Mojica *et al.,* 2000). Findings of spacer sequences resembling sections of various phage and plasmid DNA (Bolotin *et al.,* 2005; Mojica *et al.,* 2005; Pourcel *et al.,* 2005) eventually raised the hypothesis that CRISPR loci might be involved in acquired immunity against invasive elements such as viruses and plasmids. Moreover, it was suspected that the underlying mechanism might resemble the RNA interference system in eukaryotes (Makarova *et al.,* 2006). Rodolphe Barrangou and co-workers were able to show experimentally that the acquisition of a new phage-derived spacer upon virus infection indeed confers resistance against the virus, thereby supporting initial hypothesis about CRISPR-based acquired viral resistance (Barrangou *et al.,* 2007). Whereas the *in vivo* activity of CRISPR-mediated immune response could already be demonstrated in various studies for bacteria (Brouns *et al.,* 2008; Marraffini and Sontheimer 2008; Garneau *et al.,* 2010), Manica *et al* (from our laboratory) have only recently been able to show *in vivo* activity in archaea for the first time (Manica *et al.,* 2011). As their work served as the basis for this diploma thesis the most important findings of this study as well as some background information are described in section 2.5.2.

2.4.2   CRISPR-array composition

The characteristic CRISPR-array exhibits a leader sequence followed by multiple identical or nearly identical direct repeats interspersed with non-repetitive spacer sequences ranging from about 26-72 bp's in seize. The direct repeats are 28-40 bp's in

length. The number of spacer-repeat units can vary strongly from one CRISPR-locus to another. Even though the repeat sequences are highly similar within the same CRISPR-locus, they can vary between distinct loci within one organism. The spacer sequences are highly variable but exhibit similar lengths within the same CRISPR-locus. Considering the huge numbers of viruses found on earth, the fact that only a small portion of spacers match extrachromosomal sequences from phages or plasmids (Shah *et al.,* 2009) currently represented in genome databases, is hardly surprising. CRISPR-loci can occur multiplicate in archaeal as well as bacterial genomes, whereas they often differ in both repeat sequence and associated *cas* genes The presence of conserved clusters of *cas* (CRISPR-associated) genes that are often found in close proximity to the CRISPR region (Jansen *et al.,* 2002) constitutes another conserved feature within the highly diverse CRISPR systems. Those protein clusters again can be extremely divergent between both distinct species and different CRISPR-loci in one species, while certain *cas* gene clusters might be linked to certain types of repeat sequences (Kunin *et al.,* 2007). Based on multiple sequence alignments and computational analysis Haft and co-workers proposed eight Cas subtypes (Haft *et al.,* 2005) of which only two (Cas1 and Cas2) out of 45 identified *cas* gene families (Haft *et al.,* 2005) are present in all Cas subtypes. Besides, a ninth subtype called CRISPR RAMP (repeat associated mysterious proteins) module usually associated with other Cas subtypes has been identified (Haft *et al.,* 2005; Makarova *et al.,* 2006; van der Oost *et al.,* 2009). However, growing genomic data sets as well as advanced computational analyses continue to reveal previously unknown relationships between CRISPR/Cas systems and Cas proteins. Hence, Makarova *et al.* recently proposed an updated classification dividing the CRISPR/Cas systems into three major types (Cas1 and Cas2 is found in all three types) each of which can be further classified into system subtypes (Makarova *et al.,* 2011). Cas proteins that belong to the so-called RAMP superfamily, which can further be split into three major groups (Cas5, Cas6 and Cas7) are found in both type I and III systems (Makarova *et al.,* 2011). Several Cas proteins were predicted to have nuclease, helicase, integrase or polymerase property, additionally indicating that they are linked to processes of CRISPR related immune response (Haft *et al.,* 2005; Makarova *et al.,* 2006). Even though the crucial biological role of many Cas proteins remain to be resolved some of their functions could already be supported

experimentally (Al-Attar *et al.,* 2011). Interestingly, the same CRISPR-locus referring to repeat sequence and associated *cas* gene cluster can be present in different even distantly related species implying that CRISPR-loci frequently spread by horizontal gene transfer events during microbial evolution (Godde and Bickerton 2006).

2.4.3 The CRISPR/Cas mechanism

The CRISPR/Cas system can be divided into three main stages (van der Oost *et al.,* 2009): a) CRISPR-adaptation relating to the incorporation of a new spacer sequence within the CRISPR-locus, b) CRISPR-expression that involves transcription of the repeat-spacer array known as pre-crRNA that is followed by further processing into smaller RNA stretches (crRNAs), c) CRISPR-interference meaning the recognition/binding of invading nucleic acid molecules and their subsequent degradation. Figure 1 depicts a simplified model for the CRISPR/Cas mechanism.
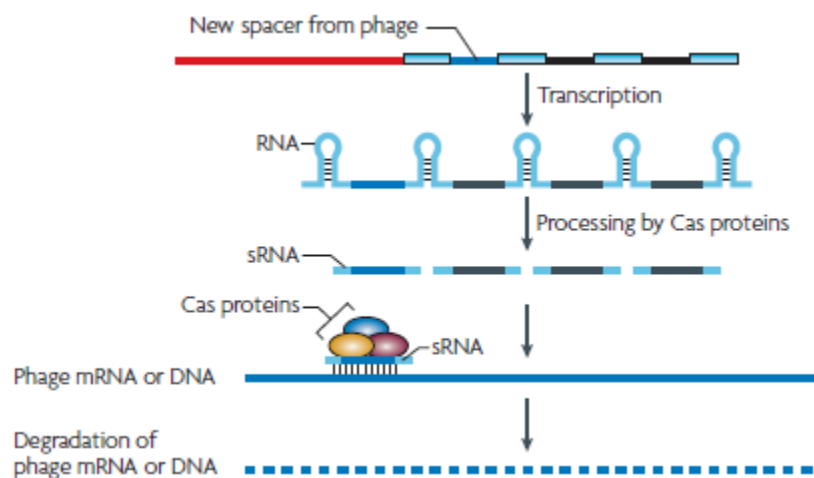
Figure 1: Simplified model for CRISPR/Cas mechanism (Sorek *et al.,* 2008). The repeat-spacer array is transcribed as a long precursor CRISPR-mRNA (pre-crRNA) that is subsequently processed by Cas proteins to obtain small CRISPR RNAs (sRNAs), each of which correspond to a single spacer. A complex of CRISPR sRNA and Cas proteins binds invading DNA or RNA molecules, respectively via base pairing eventually resulting in their degradation. This process is supposed to be mediated by CRISPR-associated proteins and is known as CRISPR interference. **Source**: Review, (Sorek *et al.,* 2008).

a) CRISPR-adaptation

Experimental studies on *Streptococcus thermophilus* demonstrated that viral infection triggered the integration of new phage-derived spacers, in turn conferring resistence to the organism previously challenged by the virus (Barrangou *et al.,* 2007). The acquisition of novel spacers observed so far appeared to happen at the very beginning of the repeat-spacer sequence (Pourcel *et al.,* 2005; Barrangou *et al.,* 2007; Deveau *et al.,* 2008) consistent with observations of various strains of *Streptococcus thermophilus* exhibiting greater spacer variability at the leader end (Horvath *et al.,* 2008). Beside the addition of new spacers, spacer loss has also been reported to happen within CRISPR-loci indicating the system's potential to actively respond to rapidly changing viral pressure, thus creating highly variable CRISPR regions (Tyson and Banfield 2008). Despite its essential role regarding CRISPR/Cas activity, the process of both spacer selection and acquisition is barely understood so far. The presence of conserved nucleotide motifs flanking protospacer sequences termed PAMs (protospacer adjacent motif) (Mojica *et al.,* 2009) led to the hypothesis that PAMs might contribute to spacer recognition and uptake processes. Nevertheless, the existence of so called PAMs as well as their actual function is highly speculative and requires additional experimental support.

b) CRISPR-expression

Transcription of the repeat-spacer array allows the system to access the spacer-encoded information whereas initially a long precursor CRISPR-mRNA (pre-crRNA) is transcribed from a promoter that most probably lies within the leader sequence (Hale *et al.,* 2009; Lillestol *et al.,* 2009). The full-length pre-crRNA is subsequently cleaved within the repeat sequences producing small CRISPR RNAs each corresponding to a single spacer flanked by partial repeat sequences (Lillestol *et al.,* 2006; Brouns *et al.,* 2008; Carte *et al.,* 2008). Northern analysis of CRISPR transcripts from *Pyrococcus furious* an extremophilic archaeon suggested even further processing of the small CRISPR RNAs whose mature forms are finally composed of one spacer and a partial (8 nucleotides) repeat sequence at its 5' end known as psi-tag (Hale *et al.,* 2009).

c) CRISPR-interference

The mature crRNAs guide a complex of Cas proteins towards invading DNA molecules allowing for their recognition via base-pairing between spacer and target sequence.

CRISPR/Cas interference eventually results in degradation of targeted invading molecules. Interestingly, spacers turned out to be complementary to both viral DNA strands, implying that rather dsDNA than RNA serves as a target for some CRISPR/Cas systems. However, contrary to studies on *S. epidermidis, S. thermophilus* and *S. solfataricus* (Marraffini and Sontheimer 2008; Garneau *et al.,* 2010; Manica *et al.,* 2011) providing evidence that DNA serves as target for the CRISPR/Cas machinery, *in vitro* studies on *P. furiosus* demonstrated cleavage of invader ssRNA (Hale *et al.,* 2009). These studies suggest that both DNA and RNA might act as potential targets, thereby providing the host organism with a more flexible immune system. Considering the mechanistic principle of the CRISPR/Cas system the host encounters the problem of potential autoimmune responses, which is a challenge for all immune systems. Thus, the cell has to find a way to prevent targeting of the host's chromosome by base-pairing of crRNA with spacer DNA encoded within the CRISPR-locus during CRISPR/Cas interference. Experimental data derived from tests on *Staphilococcus epidermis* strongly indicated that the flanks corresponding to CRISPR-repeat sequences in the crRNA are essential for self/non-self discrimination. The authors were able to demonstrate that matches between the flanking region upstream of the spacer and target sequence conferred protection to the target whereas the absence of a flanking repeat enabled interference (Marraffini and Sontheimer 2010).

2.4.4 Evading the CRISPR/Cas-based immune response

Virus-host interactions resemble a never-ending arms race primarily triggered by selective pressure that is constantly present within virus-host populations. Viruses able to circumvent CRISPR/Cas-induced immunity arise in response to ongoing adapting CRISPR/Cas systems. This was for example observed for virus particles that specifically accumulated mutations into the protospacer sequence leading to loss of resistance against the virus. The protospacer describes the nucleotide sequence within the virus genome that is complementary to the spacer. Studies focusing on bacterial systems initially indicated that even a single point mutation within the protospacer sequence permits viruses to evade CRISPR/Cas resistance (Deveau *et al.,* 2008; Heidelberg *et al.,*

2009). However, *in vivo* experiments on the hyperthermophilic crenarchaeon *S. solfataricus* demonstrated that immune response was still detectable to some degree upon infection with a virus exhibiting multiple mutations within its protospacer sequence (Manica *et al.,* 2011). Consistent with these findings, recent investigations based on the *E. coli* as well as the *Pseudomonas aeruginosa* CRISPR/Cas system suggest that especially mutations within a defined protospacer region allow for evasion whereas mutations outside this region can be tolerated more easily by the CRISPR/Cas complex (Semenova *et al.,* 2011; Wiedenheft *et al.,* 2011). Besides, mutations within 'quasi'-conserved sequences adjacent to protospacers (called PAMs) (Mojica *et al.,* 2009) have also been claimed to be involved in the loss of phage resistance (Deveau *et al.,* 2008; Mojica *et al.,* 2009; Semenova *et al.,* 2011). Moreover the deletion of protospacers as well as genome recombination events may demonstrate additional strategies to prevent interference (Deveau *et al.,* 2008; Heidelberg *et al.,* 2009).

The last few years provided substantial insights in terms of CRISPR/Cas functionality and mechanistic processes, respectively. Nevertheless, many aspects that are involved in CRISPR/Cas induced bacterial and archaeal immune response remain obscure and need to be clarified by future research work.

2.5   Demonstration of *in vivo* CRISPR/Cas activity in the archaeon *Sulfolobus solfataricus*

2.5.1   The model organism *Sulfolobus solfataricus* and its associated virus SSV1

Sulfolobus solfataricus is a hyperthermophilic Crenarchaeote of the order *Sulfolobales* and was first isolated from a solfataric field in Pisciarelli, Italy in 1980 (Zillig *et al*., 1980). Since then, *S. solfataricus* has emerged into an intensely studied model organism in archaeal research that has revealed previously unknown insights into cellular processes and gene functions of hyperthermophilic organisms.

S. solfataricus grows in acidic, sulfur-rich hot springs; its optima of temperature and pH range from 75-78 °C and 2-3, respectively. *S. solfataricus* can be cultured relatively easily in liquid as well as on solid media and requires both aerobic and heterotrophic growth conditions (Brock *et al.,* 1972; Grogan 1989). The genome of strain *S. solfataricus* P2 was fully sequenced and exhibits a size of 2.9 Mb (She *et al.,* 2001). As the genomic DNA of *S. solfataricus* comprises high abundance of IS elements, transposon mutagenesis is proposed to be the dominant mechanism of mutation events and enabled the isolation of beta galactosidase- (*lac*S) as well as uracil-auxotrophic mutants containing a transposable element in the

Figure 2: Transmission electron micrograph of wild-type SSV1 particles attached to a membrane fragment (Stedman *et al.,* 1999). **Source**: Stedman *et al.*

respective genes (Schleper *et al.,* 1994; Martusewitsch *et al.,* 2000). The isolation of the virus SSV1 (*Sulfolobus shibatae* virus 1) eventually provided the basis for the development of an efficient transfection procedure (Schleper *et al.,* 1992) and the selectable shuttle vector pMJ0305 (Jonuscheit *et al.,* 2003) in order to genetically manipulate *S. solfataricus*. SSV1 exhibits a lemon-shaped morphology (Zillig *et al.,* 1996) (compare Figure 2) and contains a circular ds-genome of 15.5 kbp. Even though SSV1 was originally isolated from *Sulfolobus shibatae*, the application of plaque assays demonstrated that SSV1 can also infect *S. solfataricus*. The virus DNA is found both as plasmid and integrated into the host chromosome after infection whereby the release of virus particles results in evident growth retardation of *S. solfataricus* without inducing cell lysis (Schleper *et al.,* 1992).

2.5.2 *In vivo* study of CRISPR-mediated virus defence in *Sulfolobus solfataricus*

Manica *et al*. managed to establish an *in vivo* test system for the hyperthermophilic archaeon *S. solfataricus* to allow for the study of CRISPR/Cas related activity in the archaeal domain of life. In order to obtain a virus-based target that is recognized by the CRISPR/Cas machinery, ORF406 an open reading frame of the conjugative plasmid pNOB8 (She *et al.,* 1998) containing a short nucleotide sequence highly similar to spacer 53 within CRISPR locus 3 of *S. solfataricus* strain P2 as well as ORF406 mutants were cloned into the recombinant SSV1 virus (pMJ0305). Subsequently, those SSV1

mutants exhibiting either 7 bp, 3 bp, 0 bp mutations or 1 bp deletion with respect to spacer 53 were tested in plaque assays. The 7 bp mutations (construct 406-7M) correspond to those found in the ORF406 wild-type gene. Construct 406-3M (3 bp mutations) only contains three out of seven mutations (with respect to the spacer sequence) initially found in the wild-type gene. The 3 bp mutations are positioned at the protospacer end being complementary to the 5'-end of spacer RNA. Construct 406-1D has 1 bp deletion in the centre of the protospacer sequence. The respective DNA sequences of the different versions of ORF406 are shown in Figure 3. Transfection efficiencies of *S. solfataricus* P2 with 406-mutants whose protospacer sequence perfectly (without mismatch) corresponded to spacer 53 were dramatically lower than those with seven mismatches-containing ORF406-constructs (see Figure 4). Unexpectedly, the 3 bp mismatch as well as the 1 bp deletion constructs yielded comparatively low transfection efficiencies (see also Figure 4) implying that CRISPR/Cas immunity is still given, albeit to reduced extent, when mutations occur within the protospacer sequence. In contrast, equally high transfection efficiencies for each of the different constructs were obtained in *S. solfataricus* strain M18 (see Figure 4) that does not carry the ORF406-specific spacer within its CRISPR loci (Manica *et al.,* 2011).
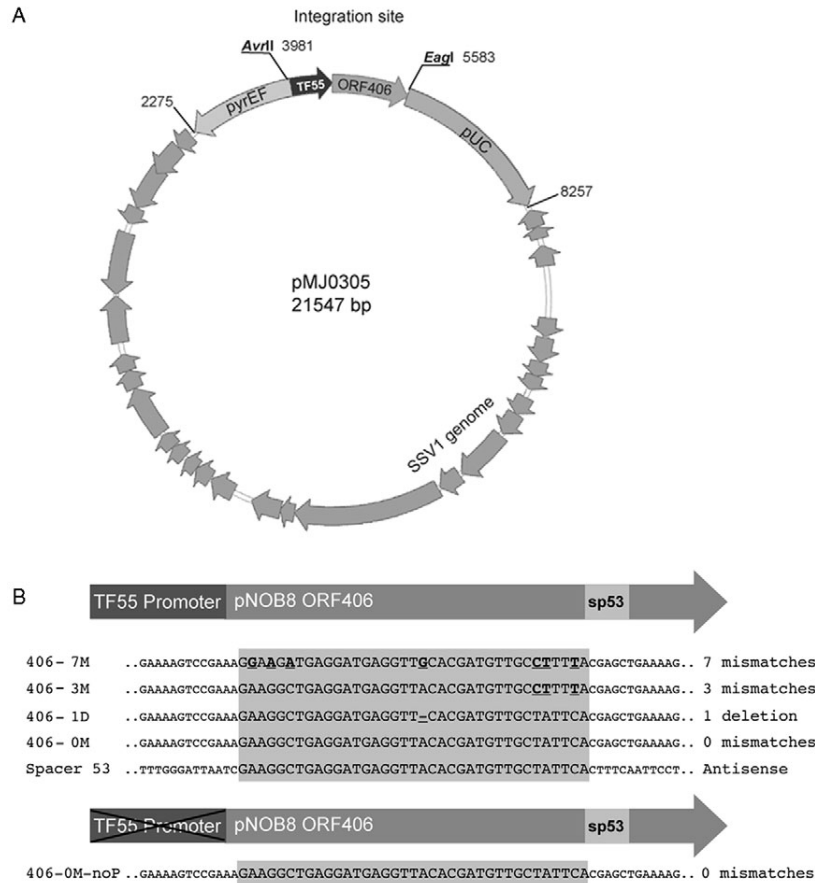
Figure 3: Overview of recombinant shuttle vector and protospacer constructs.

A. *E. coli*/S. *solfataricus* shuttle vector pMJ0305. The different ORF406 gene variants have been incorporated into the insertion site.

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 with 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 spacer53. Nucleotides depicted on the left and right side to the spacer53 represent the repeat region of the locus. **Source**: Manica *et al.*, 2011.
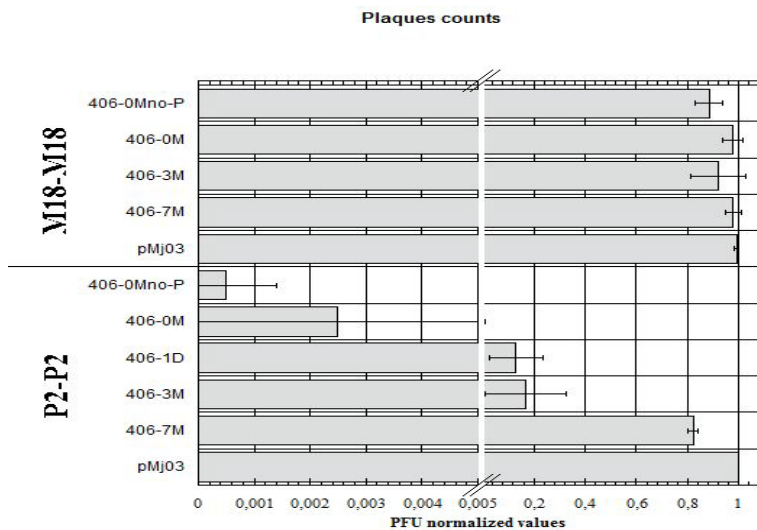
Figure 4: Obtained transfection efficiencies for different SSV1 shuttle-vector constructs.
Source: modified from table 1, Manica *et al.*, 2011.

2.6   The goal of this study

The widespread occurrence of CRISPR/Cas based immune systems in both bacterial and archaeal organisms indicate their ecological relevance in shaping viral populations, thereby also manipulating microbial population dynamics and composition. Even though several studies revealed many new aspects about CRISPR/Cas functionality and related mechanistic processes the complete mechanism as well as its contribution to ecological processes is not fully understood yet. As many investigations are primarily focusing on bacterial CRISPR/Cas-mediated immune systems, this study aimed to learn more about CRISPR/Cas related functioning in the archaeal domain.

More precisely, the primary task of the study was to explore the base-pairing requirement for crRNA-protospacer DNA interaction using the hyperthermophilic crenarchaeote *Sulfolobus solfataricus* as a model organism. The study involved the gradual introduction of increasing numbers of mutations at the two different ends of the protospacer DNA sequence. The generation of protospacer mutants included polymerase chain reaction mutagenesis as well as the usage of Gateway® Cloning technology. Consequently, the goal was to test the impact of both increasing number and differently positioned protospacer mutations on CRISPR/Cas-mediated immune activity via the application of plaque assays. The tested mutants included protospacer variants carrying from six up to ten consecutive mutations at the "up"-end of the protospacer (complementary to the 5'-end of spacer RNA) as well as variants containing consecutive mutations ranging from six up to 18 at the other end of the protospacer sequence.

The investigation should serve to elucidate how tolerant the system may be towards protospacer mutations and which mutation positions may have the strongest impact on CRISPR/Cas related DNA interference.

3 Materials

3.1 Microorganisms

Escherichia coli Top10   Invitrogen

Sulfolobus solfataricus P2 (DSM 1617)   Zillig *et al.*, 1980

Sulfolobus solfataricus M18   Martusewitsch *et al.*, 2000
(*pyr*EF mutant of strain P1; DSM 1616)

3.2 Oligonucleotides

ORF406-specific primer pair	Primer sequence 5' - 3'
pNOB8_ORF406_fw	ATACCATGGACAGCATAGGATTTTGTTTTCGAG
pNOB8_ORF406_rw	TATGGGCCCCTATGCTAGCTTAGTGGAGTGTGAG

Mutagenesis primers	Primer sequence 5' – 3'
Rw1-406-6M-up	GCAACATCGTGTAACCTCATCC
Fw1-406-6M-over-up	CTGATCCGAGCTGAAAAGCATCTTGAAG
Fw-8M-UP	GATCAGAAAACATCGTGTAACCTC
Fw-WOP	CGAGCTGAAAAGCATCTTGAA
Fw-10M-up	GATCAGAATTCATCGTGTAACCTC
Fw-WOP	CGAGCTGAAAAGCATCTTGAA
Fw2-406-6M-down	TGAGGATGAGGTTACACGATGTT
Rw2-406-6M-over-down	GTACGTTTCGGACTTTTCCACCAACT
Rw-9M-down	CGTACAAATGGATGAGGTTACAC
Rw-WOP	TTTCGGACTTTTCCACCAACT
Fw4-406-13M-down	GAGGTTACACGATGTTGCTATTCA
Rw4-406-13M-specific	TCATATGTGTACGTTTCGGACTT
Fw3-406-14M-down	AGGTTACACGATGTTGCTATTCAC
Rw3-406-14M-over-down	ATCATATGTGTACGTTTCGGACTTTTCCACCAACT

Fw5-406-15M-down	TTGGTTACACGATGTTGCTATTC
Rw4-406-13M-specific	TCATATGTGTACGTTTCGGACTT
Fw-18-down	CGTACACATATGATTAACTACACGATGT
Rw-WOP	TTTCGGACTTTTCCACCAACT
Fw-WOP	CGAGCTGAAAAGCATCTTGAA
Rw-WOP	TTTCGGACTTTTCCACCAACT

3.3 Media for cultivation of *E. coli*

Luria Bertani Medium (LB-Medium)

	for 1 litre:
Tryptone	10 g
Yeast extract	5 g
NaCl	10 g
solid medium for plating:	
Agar	15 g

Antibiotics for selection

Ampicillin	50 µg/ml
Spectinomycin	100 µg/ml

SOB-Medium

Bacto Trypton	2 %
Yeast extract	0.5 %
NaCl	10 mM
KCl	2.5 mM

SOC-Medium

SOB-Medium	
Add: Glucose	20 mM
Mg_2	20 mM

3.4 Media for cultivation of *S. solfataricus*

Strain P2:

Brock salts basal medium (per litre)

100x brock solution	
$(NH_4)_2SO_4$	130 g
$MgSO_4$ x 7 H_2O	25 g
$FeCl_3$ x 6 H_2O	2 g

200x brock salt solution

KH_2PO_4	56 g
$MnCl_2$	360 mg
$ZnSO_4$	44 mg
$CuCl_2$	10 mg
$VOSO_4$	6 mg
Na_2MoO_4	6 mg
$Na_2B_4O_7$	0.9 mg

1000x bock salt solution

$CaCl_2$ x 2 H_2O	70 g

carbon sources

tryptone	0.1% [w/v]
sucrose	0.2% [w/v]

→ pH was adjusted to 3 with 50 % H_2SO_4

Strain M18:

Brock salts basal medium (per litre)

100x brock solution

$(NH_4)_2SO_4$	130 g
$MgSO_4$ x 7 H_2O	25 g
$FeCl_3$ x 6 H_2O	2 g

200x brock salt solution

KH_2PO_4	56 g
$MnCl_2$	360 mg
$ZnSO_4$	44 mg
$CuCl_2$	10 mg
$VOSO_4$	6 mg
Na_2MoO_4	6 mg
$Na_2B_4O_7$	0.9 mg

1000x bock salt solution

$CaCl_2$ x 2 H_2O	70 g

carbon sources

tryptone	0.1% [w/v]
sucrose	0.2% [w/v]

→ pH was adjusted to 3 with 50 % H_2SO_4

for complementation of uracil auxotrophy

Uracil	0.01 mg/ml

Solid medium for plaque assays

Brock salts basal medium (see above)

carbon sources

yeast extract	0.1% [w/v]
sucrose	0.2% [w/v]

for solidification

MgSO$_4$	10 mM
gelrite	0.6 % [w/v]

→ pH was adjusted to 3 with 50 % H$_2$SO$_4$

3.5 Buffers

Tris Acetate buffer (TAE-buffer)

Tris base	242 g
EDTA 0.5M (pH8)	100 ml
Acetic acid	57.1 ml
add to 1liter ddH2O	

TE-buffer

Tris/HCl, pH7.5	10 mM
EDTA	1 mM

3.6 List of chemicals, enzymes, size markers, cloning vectors

Vectors

pCR$^®$8/GW/TOPO$^®$	Invitrogen
pMJ0305	Jonuscheit *et al.*, 2003

Size markers

GeneRuler™ 1kb Plus bp DNA Ladder	Fermentas
GeneRuler™ 1kb bp DNA Ladder	Fermentas
GeneRuler™ 100 Plus bp DNA Ladder	Fermentas

Enzymes

T4 polynucleotide kinase	Fermentas
T4 DNA Ligase	Fermentas
Taq DNA Polymerase	Fermentas
Phusion DNA Polymerase	Finnzymes

Gateway® LR Clonase™ Enzyme mix   Invitrogen
BigDye® Terminator   Applied biosystems
→ all reactions were performed in buffers supplied with the enzymes

Kits

NucleoSpin® Gel and PCR Clean-up	Macherey-Nagel
E.Z.N.A® Plasmid Miniprep Kit	Omega Bio-Tek Inc

Chemicals

Agarose LE	Biozym
KH2PO4	Carl Roth GmbH & Co
MnCl2	Carl Roth GmbH & Co
ZnSO4	Carl Roth GmbH & Co
CuCl2	Carl Roth GmbH & Co
VOSO4	Carl Roth GmbH & Co
Na2MoO4	Carl Roth GmbH & Co
Na2B4O7	Carl Roth GmbH & Co
(NH4)2SO4	Carl Roth GmbH & Co
MgSO4 x 7 H2O	Carl Roth GmbH & Co
FeCl3 x 6 H2O	Carl Roth GmbH & Co
CaCl2 x 2 H2O	Carl Roth GmbH & Co
Spectinomycin dihydrochloride pentahydrat	Sigma
Ampicillin	Sigma
Agar Agar Kobe I	Carl Roth GmbH & Co
Bacto Yeast Extract	BD
Glycerol	Carl Roth GmbH & Co
Sucrose	Serva
Tryptone	Carl Roth GmbH & Co
H_2SO_4	Carl Roth GmbH & Co
NaCl	AppliChem
GelRite	Carl Roth GmbH & Co
Uracil	Sigma-Aldrich

Other materials

MF-Millipore™ Membrane Filters	Millipore

4 Methods

4.1 Construction of shuttle vectors using Gateway® Cloning

4.1.1 Vectors used for cloning

Entry vector pEnter-406 (Figure 5) is based on pCR®8/GW/TOPO® and was used to generate mutations on the protospacer sequence in the ORF406. Resulting protospacer mutants were used for subsequent cloning into the viral shuttle-vector pMJ0305.

The vector contains, beyond others, a pUC origin of replication (*ori*) allowing high-copy replication and maintenance in *E. coli*, spectinomycin promoter and spectinomycin resistance gene (*aad*A1) that confers resistance to the cells that have taken up the plasmid under transformation. Since spectinomycin inhibits the protein synthesis, cells that are not resistant to spectinomycin are unable to divide on selective media containing this antibiotic. For the purpose of CRISPR studies the *Taq*-amplified PCR product of the ORF406 was cloned into the TOPO® Cloning site of pCR®8/GW/TOPO®. Additionally the entry vector contains Bacteriophage λ-derived recombination sequences (*att* sites) that allow recombinational cloning of the inserted open reading frame ORF406 with a Gateway® destination vector (Landy 1989).
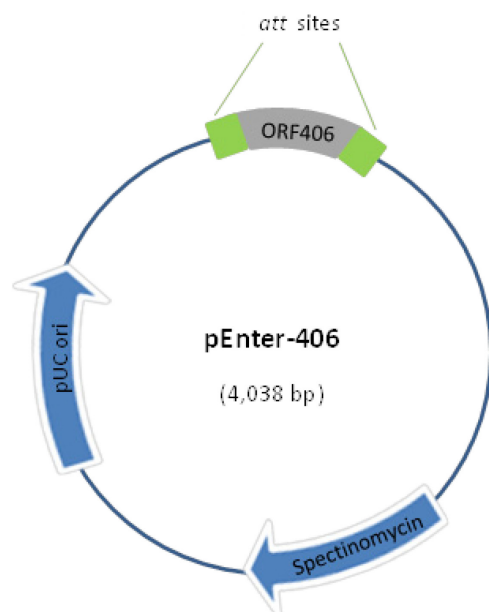
Figure 5: Simplified picture of the entry vector pEnter-406. The vector is based on the cloning vector pCR®8/GW/TOPO® and contains pNOB8 open reading frame ORF406 that has been varied to obtain different protospacer mutants. The protospacer containing ORF406 is flanked by two recombination sequences (*att* sites). **Source** of the vector: Manica *et al*., unpublished

The destination vector (Figure 6) is based upon the *E. coli*/S. *solfataricus* shuttle vector pMJ0305 comprising the complete genome of virus SSV1 that has originally been isolated as a 15 kbp plasmid from *Sulfolobus shibatae*. Besides the sequence of SSV1 the vector contains pUC18 allowing for propagation in *Escherichia coli* and the genes *pyrEF* (coding for orotidine-5'-monophosphate pyrophosphorylase and orotidine-5'-monophosphate decarboxylase) of *Sulfolobus solfataricus* as selectable marker to complement pyrimidine auxotrophic mutants. Additionally, the shuttle vector encodes for beta lactamase providing ampicillin resistance as selection marker after transformation of *E. coli*. In order to enable gene transfer from entry to destination vector two *att* sites flanking a cassette that contains a *ccd*B and chloramphenicol resistance gene were introduced into the vector's insertion site. The CcdB protein interferes with *E. coli* DNA gyrase (Bernard *et al.,* 1993), thereby inhibiting growth of most *E. coli* strains (*e.g.* DH5α™, TOP10). When recombination between destination and entry vector occurs the *ccd*B gene is replaced by the gene that is located between the two *att* sites on the entry vector. Thus, cells that take up unreacted vectors carrying the *ccd*B gene or by-product molecules retaining the *ccd*B gene will fail to grow.
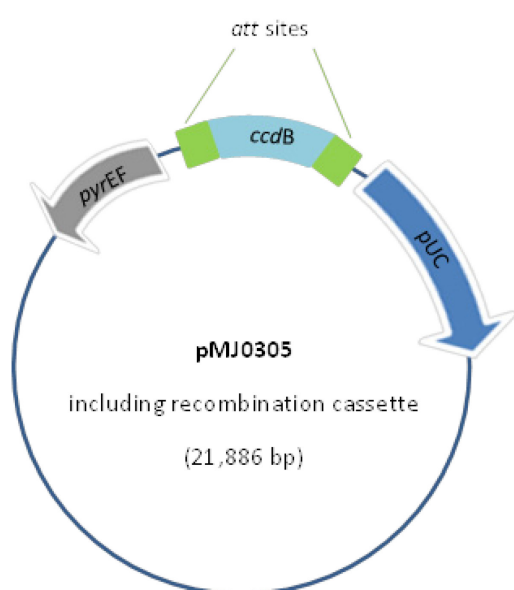
Figure 6: Simplified picture of the destination vector pMJ0305 including a recombination cassette. The vector was developed by Manica *et al.*, (unpublished).

24

4.1.2   Inverse PCR (Polymerase Chain Reaction) Mutagenesis

Inverse PCR constitutes a modified form of PCR where the primers in contrast to common PCR are not oriented towards each other but are directed in inverse orientation on a circular template. This method was used to generate site-directed mutations on the pre-vector DNA of approximately 4 kb. For each reaction one sequence-specific and one primer with overhang not matching the template DNA to introduce the desired mutation were designed. During PCR amplification the mutation is integrated in the vector DNA. As Phusion DNA Polymerase provides higher fidelity than commonly used Taq DNA Polymerase, Phusion was used for mutagenesis PCR.

The standard reaction mix included:

- 5 µl of 5x Phusion buffer (+ MgCl$_2$)

- 0.2 mM dNTP mix

- each 0.2 µM forward and reverse primers

- 2 units Phusion DNA Polymerase

- 1 µl template (1-10 ng/µl vector DNA)

- ddH$_2$O added to a total volume of 25 µl

The reactions were performed in a thermocycler by use of the following program:

Initial denaturation for 2 min at 98 °C

- Denaturation for 15s at 98°C

- Annealing for 20s at 56/57/60°C

- Elongation for 4 min at 72°C

} 30 cycles

- Final elongation was done for 5 min at 72°C.

The denaturation step causes melting of template DNA yielding ssDNA molecules. Subsequent temperature lowering allows hybridisation between primer and ss template DNA during the annealing step. Annealing temperature is primary dependent on GC-content of the primer sequence and has to be calculated before reaction procedure. During elongation the polymerase binds to the primer-template hybrid and starts the synthesis of the new DNA. This step was performed at 72°C in which the enzyme catalyses the polymerization of approximately 1,000 nt per minute. The final elongation step is performed after the last PCR cycle to ensure full extension of remaining ssDNA.

Primer-pairs as well as templates used for generation of the respective protospacer mutants are listed in Table 1. For primer sequences see section 3.2.

Table 1: Templates and corresponding primer pairs used to create modified versions of the ORF406 protospacer sequence.

Template pEnter	Construct name	Primer name
406-0M	406-6Mup	Rw1-406-6M-up Fw1-406-6M-over-up
406-6M-up	406-8M-up	Fw-8M-UP Fw-WOP
406-6M-up	406-10M-up	Fw-10M-up Fw-WOP
406-0M	406-6Mdown	Fw2-406-6M-down Rw2-406-6M-over-down
406-6M-down	406-9M-down	Rw-9M-down Rw-WOP
406-14Mdown	406-13Mdown	Fw4-406-13M-down Rw4-406-13M-specific
406-0M	406-14Mdown	Fw3-406-14M-down Rw3-406-14M-over-down
406-14Mdown	406-15Mdown	Fw5-406-15M-down Rw4-406-13M-specific
406-15M-down	406-18M-down	Fw-18-down Rw-WOP
406-0M	406-WOP	Fw-WOP Rw-WOP

After amplification the PCR product was investigated by loading a small amount of the sample onto a 0.8 % ethidium bromide stained agarose gel. In order to obtain the PCR product of the right size the band was excised from the agarose gel and purified before *Dpn*I digestion. The reaction mixture was incubated at 37°C for 2 hours to remove template DNA from amplified mutated DNA. *Dpn*I only hydrolyses methylated template DNA whereas the amplified DNA remains unmethylated during PCR reaction

and cannot be cut by the *Dpn*I enzyme. Afterwards the PCR product was purified from remaining reaction components that might interfere with the ligation procedure.

4.1.3   Ligation of vector DNA

In order to get the circular form of the vector DNA it had to be ligated after PCR amplification. For the purpose of blunt-end ligation T4 DNA ligase and 25-35 ng of phosphorylated vector DNA were used. As the ligation procedure requires a 5'-phosphate at the free ends of DNA the primers were phosphorylated before PCR.

The whole reaction mixture included:

- 5 µl of 10x T4 DNA ligase buffer

- 5 µl of 50 % PEG4000 solution (for blunt end ligation)

- 1 µl of T4 DNA ligase

- 25-35 ng of linear DNA

- ddH$_2$O added to a total volume of 50 µl.

The DNA ligase catalyses the formation of a covalent bond between 3'-hydroxyl ends with 5'-phosphate ends under ATP consumption. The reaction mixture was incubated overnight at 22°C.

4.1.4   Primer phosphorylation

Previous to inverse mutagenesis PCR, 100 pmol of primer were phosphorylated by T4-polynucleotide kinase. The enzyme transfers the phosphate group from ATP to the 5'-hydroxyl terminus of the DNA. Additionally the 20 µl reaction included 10x T4 polynucleotide kinase buffer, 1 µl of 10 mM ATP, ddH$_2$O and it was incubated at 37 °C for 1 hour. After incubation the enzyme was inactivated by freezing at -20°C.

4.1.5   Agarose gel electrophoresis

To check the quality of DNA and to separate DNA fragments of different size agarose gel electrophoresis was performed. For this purpose the appropriate amount of agarose was added to 1xTAE buffer and heated in a microwave to allow to completely dissolve the agarose. Before adding about 3 drops of ethidium-bromide (final

concentration of approximately 0.1-0.5 µg/ml) the suspension was cooled down and finally poured into a gel rack containing a comb. When the agarose gel was stiff the comb was removed and the chamber was filled with 0.5xTAE. After mixing the DNA samples with loading dye in a proper ratio (about 6:1) electrophoresis was performed at 80-120 Volt. When an electric current is applied, DNA molecules move through the gel from the negative to the positive pole as their phosphate backbone is negatively charged. The migrational velocity of the molecules is dependent on their number of base pairs. By using proper DNA markers containing nucleic acid fragments of known size, the length of the investigated DNA-molecules can be estimated. Ethidium bromide intercalates into double-stranded DNA-molecules and is activated by UV-light enabling the visualisation of nucleic acids by a specific scanning device which detects fluorescence signals.

4.1.6 Purification of PCR products by PCR-clean up kit (NucleoSpin^{TM})

After PCR amplification the PCR products were purified to avoid interference of remaining reaction compounds when performing subsequent reaction procedures. The instructions provided by the product manual were closely followed and the final elution step was done with 40 µl of provided EB buffer. The samples were stored at 4°C as freezing may result in fragmentation of the PCR products.

4.1.7 Excision of bands from agarose gel

When specific DNA bands had to be excised from agarose gel the sample was loaded on a 1 % agarose gel to allow for initial electrophoretic separation of the designated DNA bands. Under UV exposition the right bands were cut out from the respective sample. In order to prevent DNA damage the UV exposition should be kept as short as possible. Subsequently the excised DNA fragments were extracted by using NucleoSpin^{TM} kit. The respective protocol was closely followed and the extracted DNA was eluted in 60-80 µl EB buffer.

4.1.8 Quantification of nucleic acids

To quantify the extracted DNA the nanodrop spectrophotometer, which calculates the concentration of nucleic acids by measuring UV-absorbance of DNA molecules at 260 nm (absorption maxima of nucleic acids), was used. Only 1 µl of the respective sample is required for measuring. The purity of the sample is obtained from the absorption curve and the absorbance ratios at 260nm/280nm and 260nm/230nm. The first ratio gives you information about protein contamination and should ideally have a value of 1.6-2.0 whereas the latter one gives information about contamination by other compounds such as phenol (absorbing at wavelength 230) and should be in the range of 1.8.

4.1.9 Transformation of chemically competent *E.coli* cells

After the designated mutations were integrated the circular vector DNA was used to transform chemically competent cells of *E.coli* strain Top10 (F-) (Invitrogen). Each time, 25 µl of competent cells were incubated on ice with 1-5 µl of the respective ligation product for 15 minutes. Subsequently, the cells were heat-shocked for 30 s in a thermo-cycler with 42°C and immediately transferred to ice. Cell recovery was done by adding 200 µl of S.O.C.-Medium and regenerating at 32-37°C for at least 45 minutes during which the cells were shaken at 300 rpm.

4.1.10 Plating of *E.coli*

After regeneration the transformation mix was plated on LB plates containing either spectinomycin [100 $^{\mu g}/_{ml}$] or ampicillin [100 $^{\mu g}/_{ml}$] and grown over night. In order to prevent recombination events between vector and chromosomal DNA of *E.coli*, likely to occur when transforming big vectors, the plates were incubated at 28-32°C. According to vector size different concentrations of antibiotics were used, e.g. for big vectors (> 20 kb) only half of the concentration which means [50 $^{\mu g}/_{ml}$] was added to the plates. The addition of antibiotics to the culture medium allows for the selection of transformed *E.coli* cells. As there is an antibiotic resistance gene located on the plasmid only those cells carrying the plasmid will survive.

4.1.11 Plasmid purification from transformed *E.coli* cells

Plasmids were purified from single clones of transformed *E.coli* cells by use of the E.Z.N.A. Miniprep Kit. Prior to plasmid purification, single colonies were picked and inoculated in 5-7 ml of liquid LB-medium supplied with either spectinomycin [100 $^{\mu g}/_{ml}$] or ampicillin [50 $^{\mu g}/_{ml}$] and grown over night at 28-32°C. The cells were harvested by centrifuging at 4,000 rpm for 10 min and the supernatant was discarded. Plasmid purification was performed as described in the manufacturer's manual.

4.1.12 Polymerase Chain Reaction (PCR)

The polymerase chain reaction provides the opportunity to amplify specific DNA regions such as genes or non-coding DNA sequences. In order to check for the presence of the mutated ORF406 a specific primer pair (pNOB8_ORF406_fw, pNOB8_ORF406_rw; see section 3.2) targeting the 1.2 kb DNA fragment located on the transformation vector was used for amplification.

The standard reaction mix included:

- 5 µl of 5xTaq Polymerase buffer

- 0.2 mM dNTP mix

- 2 mM $MgCl_2$

- each 0.2 µM forward and reverse primers

- 2 units Taq DNA Polymerase

- 1 µl template (1-10 ng/µl vector DNA)

- ddH$_2$O added to a total volume of 25 µl.

The corresponding PCR program was as follows:

- Initial denaturation for 2 min at 95 °C

- Denaturation for 15s at 95 °C

- Annealing for 20s at 52 °C } 34 cycles

- Elongation for 1.20 min at 72°C

- Final elongation was done for 5 min at 72°C.

When PCR was directly performed on *E.coli* transformants (colony PCR), single colonies were picked with a sterile pipette tip, transferred in PCR tubes and used as template

for PCR amplification. In the case of colony PCR initial heating was extended to ensure cell destruction and by this to free enough template for the reaction.

4.1.13 Sequence reaction and determination

To determine defined regions of DNA sequences and/or to check for the accuracy of inserted mutations the purified plasmids were used for sequencing. For this purpose the dye termination sequencing procedure was used. In order to obtain sequence fragments of diverse length the dNTP mix also included fluorescently labelled 3'-didesoxyribonucleosidtriphosphates (ddNTP). As the ddNTPs lack the hydroxylgroup on the 3' C-atom strand synthesis is interrupted each time a fluorescently labelled nucleoside is attached. The reaction mix included the following components: 1.5 µl Big Dye Chemistry (ABI, Applied Biosystems), 1 µl 5x Sequencing Buffer (Applied Biosystems), 5 pmol primer (either forward or reverse), 20-25 ng template DNA (purified PCR product) and ddH2O to a final volume of 10 µl. When more than one clone had to be analysed, a master mix was prepared. The reaction was conducted in a thermo cycler by use of the following program: Initial heating step at 96°C for 1 min and 30 cycles of [denaturation at 96°C for 20 s, annealing at 52°C for 15 s, elongation at 60°C for 4 min]. The sequence identification was carried out by use of an ABI capillary DNA Sequencer.

4.1.14 Sequence analysis

The DNA sequence was determined for three to five randomly chosen clones per vector construct. After proofreading the sequences by use of the chromatogram, a sequence alignment was performed with ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The less well resolved ends of the sequences were removed previous to sequence alignment. Only those vectors whose DNA sequence revealed the accurate mutations in the protospacer region were chosen for subsequent Gateway® recombination.

4.1.15 Gateway® Recombination (LR reaction)

In order to transfer the *att*-flanked gene from the entry clone to the destination vector the so called LR reaction was performed. The reaction is mediated by an enzyme mixture called LR Clonase®, which contains the necessary protein activity to excise the gene from the entry clone and integrate it into the destination vector.

Gateway® LR Clonase™ II enzyme mix contains a proprietary blend of Int (Integrase), IHF (Integration Host Factor) and Xis (Excisionase) enzymes that catalyze the *in vitro* recombination between entry clone and destination vector.

The reaction mixture included following components:

- 150 ng of entry vector

- 680 ng of destination vector

- 2 µl of Gateway® LR Clonase™ II enzyme mix

- TE-buffer added to a total volume of 10 µl

The mixture was incubated overnight at 25°C.

2 µl of the reaction mixture were used for subsequent transformation into *E. coli* cells (*E. coli* TOP10).

4.1.16 Glycerol stocks of transformed *E.coli* colonies

To be able to conserve positive clones that have taken up plasmids glycerol stocks were made. For this purpose the overnight *E.coli* culture was mixed with sterile 100 % glycerol in the ratio of either 1:1 or 1:1.5 (for pMJ0305 constructs). The glycerol stocks were stored at -80°C.

4.2 Growth of *Sulfolobus solfataricus*

4.2.1 Liquid culture

Strain *S. solfataricus* P2 was grown in 50-200 ml Brock's medium (Grogan, 1989) containing 0.1% (w/v) tryptone and 0.2% sucrose at pH 3 adjusted with 50 % H_2SO_4. Long-neck bottles were filled with basal media and a defined volume of pre-culture or

300 µl of 10x concentrated *S. solfataricus* glycerol stocks were inoculated and incubated at 78°C in a shaking oil-bath providing highly aerobic conditions. The optical density of liquid cultures was monitored at 600 nm.

4.2.2 Conservation of *Sulfolobus solfataricus*

4.2.2.1 Glycerol stocks

Exponentially grown cells were cooled down and centrifuged at 4,000 rpm for 10 min at 4°C. The supernatant was discarded and the cell pellet was resuspended in Brock's basal medium supplemented with 25% glycerol. The cell suspension was stored at -80 °C.

4.2.3 Transfection of *Sulfolobus solfataricus*

4.2.3.1 Dialysis of vector DNA

As increased conductivity due to high salt concentrations (e.g. buffer salts) may lead to disturbances during the electroporation procedure vector DNA had to be dialysed first. For this purpose ddH$_2$O was supplemented with 0.05mM Tris-HCl (final concentration) and a filter disc (Millipore, pore size 0.025 µm) was put on the liquid surface. 20-40 µl of vector DNA were pipetted on the filter disc and dialysed for 2 h at room temperature. Subsequently DNA was quantified with a nano-Drop spectrometer (Peqlab) and used for electroporation of *S. solfataricus* cells.

4.2.3.2 Electroporation of *Sulfolobus solfataricus*

Beside some modifications due to optimization reasons the electroporation of *S. solfataricus* cells was performed as described in Kurosawa and Grogan (Kurosawa and Grogan 2005).

Exponentially grown *S. solfataricus* P2 cells were used for the preparation of competent cells previous to electroporation. 150 ng of dialysed viral DNA were mixed

with 50 µl competent cells and pipetted into cooled cuvettes (1mm gap). Subsequent electroporation was performed by the use of Gene Pulser Xcell (BioRad). Instrument adjustments were as follows: 1250V, 25µF, 1000Ω. The cells were regenerated in 50 µl pre-warmed recovery solution (Berkner *et al.,* 2007) for 45-60 min at 75°C immediately after electroporation. A defined volume of transformation mixture was used for subsequent plating (plaque assay).

4.2.3.3 Preparation of electrically competent *Sulfolobus solfataricus* cells

For the preparation of competent P2 cells an overnight culture having an optical density (OD_{600}) of approximately 0.18 was cooled down in ice water and centrifuged for 10 min at 4,000 rpm and 4°C. In order to remove the salts the cells were washed by gently resuspending the cell pellet in 20 ml of ice-cold 20mM D-sucrose. This step was repeated three times. A final concentration of 10^{10} cells/ml was adjusted by adding the appropriate volume of 20mM D-sucrose and this was used for subsequent electroporation.

4.3 Plaque assay

4.3.1 Media preparation

Solid medium was directly prepared before plating and consisted of 0.6% gellan gum (gelrite), 10 mM $MgSO_4$, 0.1% (w/v) yeast extract, 0.2% sucrose and Brock's medium. As the plates were incubated at 80 °C the more heat resistant gelrite instead of agar was used for solid media. Gelrite is a polysaccharide composed of uronic acids and requires an increased fraction of bivalent cations in the medium for crosslinking.

4.3.2 Plating

For the *S. solfataricus* P2 lawn 280 µl of 10 times concentrated exponentially growing cells were added to 2.8 ml of pre-warmed Brock salts supplemented with 0.4 % gelrite.

After adding 5-12 µl of transfected cells the mixture was quickly poured on pre-warmed solid gellan gum plates and incubated in humid atmosphere at 78°C. When *S. solfataricus* M18 cells were used in overlays, the 10 times concentrated cell suspension was supplemented with uracil to ensure sufficient formation of cell lawn. Each 280 µl of 10x concentrated cells as well as 1 µl of transfected M18 cells were used for subsequent plating. The volume was reduced to 1 µl of transfected cells (compared to 5-12 µl of transfected P2 cells) as resulting plaques were clearly bigger in size than those formed when using *S. solfataricus* P2 cells in plaque tests. Plaques were formed after 2 days of incubation.

4.3.3 Determination of transfection efficiency

Transfection efficiencies were determined in duplicate and triplicate respectively 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 M18 and lawns of strain M18, respectively. Maximal transfection efficiencies for strain P2 varied in the different experiments. Therefore, transfection efficiencies were normalized in Table 2 and Table 3 to the highest plaque count obtained in each single experiment.

5 Results

5.1 Design of protospacer variants for the recombinant virus SSV1

In order to determine the requirement for base-pairing between crRNA and protospacer in *S. solfataricus* various protospacer mutants were constructed and tested in the *S. solfataricus*/SSV1 system in transfection experiments. Initially two constructs both carrying six consecutive nucleotide substitutions but differently located within the protospacer were generated. Construct 406-6Mup contained six mutations at positions +1 to +6 whereas construct 406-6Mdown carried mutations at positions +32 to +37 within the protospacer sequence. Since these constructs yielded divergent results, mutations were gradually increased at both sides of the protospacer obtaining constructs 406-8Mup, 406-10Mup, 406-9Mdown, 406-13Mdown, 406-14Mdown, 406-15Mdown and 406-18Mdown. All protospacer mutations refer to spacer 53 that is located within the active CRISPR locus 3 of *S. solfataricus* P2. Construct 406-WOP (without protospacer; obtained from Ziga Zebec, unpublished) contained ORF406 but lacked the spacer 53 corresponding protospacer sequence and served as a control to check whether the ORF406 sequence itself can affect transfection efficiency with respect to the native recombinant viral-vector pMJ0305. DNA sequence of each protospacer variant is depicted in Figure 7. For a better understanding the single nucleotide positions within the protospacer are numbered from +1 up to +37. Protospacer position +1 refers to the first crRNA position immediately followed by the eight nucleotide repeat-derived flanking sequence. Subsequent positions are referred to as 2, 3, 4, etc. ending with 37 being the last position of the protospacer. The designation "up" describes the protospacer corresponding end of the crRNA region that is adjacent to the repeat-derived flanking sequence whereas "down" indicates the opposite end of the protospacer.

A

```
8 nt repeat-
  derived
flanking seq
        +1        +10       +20       +30
   ┌──┴──┐ |         |         |         |
5'-AUUGAAAGUGAAUAGCAACAUCGUGUAACCUCAUCCUCAGCCUUC-3' spacer RNA
   |      ||||||||||||||||||||||||||||||||||||||
3'-AGTCGAGCACTTATCGTTGTAGCACATTGGAGTAGGAGTCGGAAGAAAGCCTG-5'
```

protospacer DNA (antisense)

B

```
AGTCGAGCACTTATCGTTGTAGCACATTGGAGTAGGAGTCGGAAGAAAGCCTG 406-0M
........CTAGTC...................................... 406-6M-up
........CTAGTCTT...................................... 406-8M-up
........CTAGTCTTAA.................................... 406-10M-up

...............................................ACATGC........ 406-6M-down
..........................................TACACATGC........ 406-9M-down
.....................................AGTATACACATGC........ 406-13M-down
....................................TAGTATACACATGC........ 406-14M-down
...................................TTAGTATACACATGC........ 406-15M-down
................................CAATTAGTATACACATGC........ 406-18M-down
     |    |    |         |         |   |   |
    +1   +6  +10       +20       +28 +32 +37
```

C

```
AGTCGAGCACTTATCGTTGTAGCACATTGGAGTAGGAGTCGGAAGAAAGCCTG 406-WOP
```
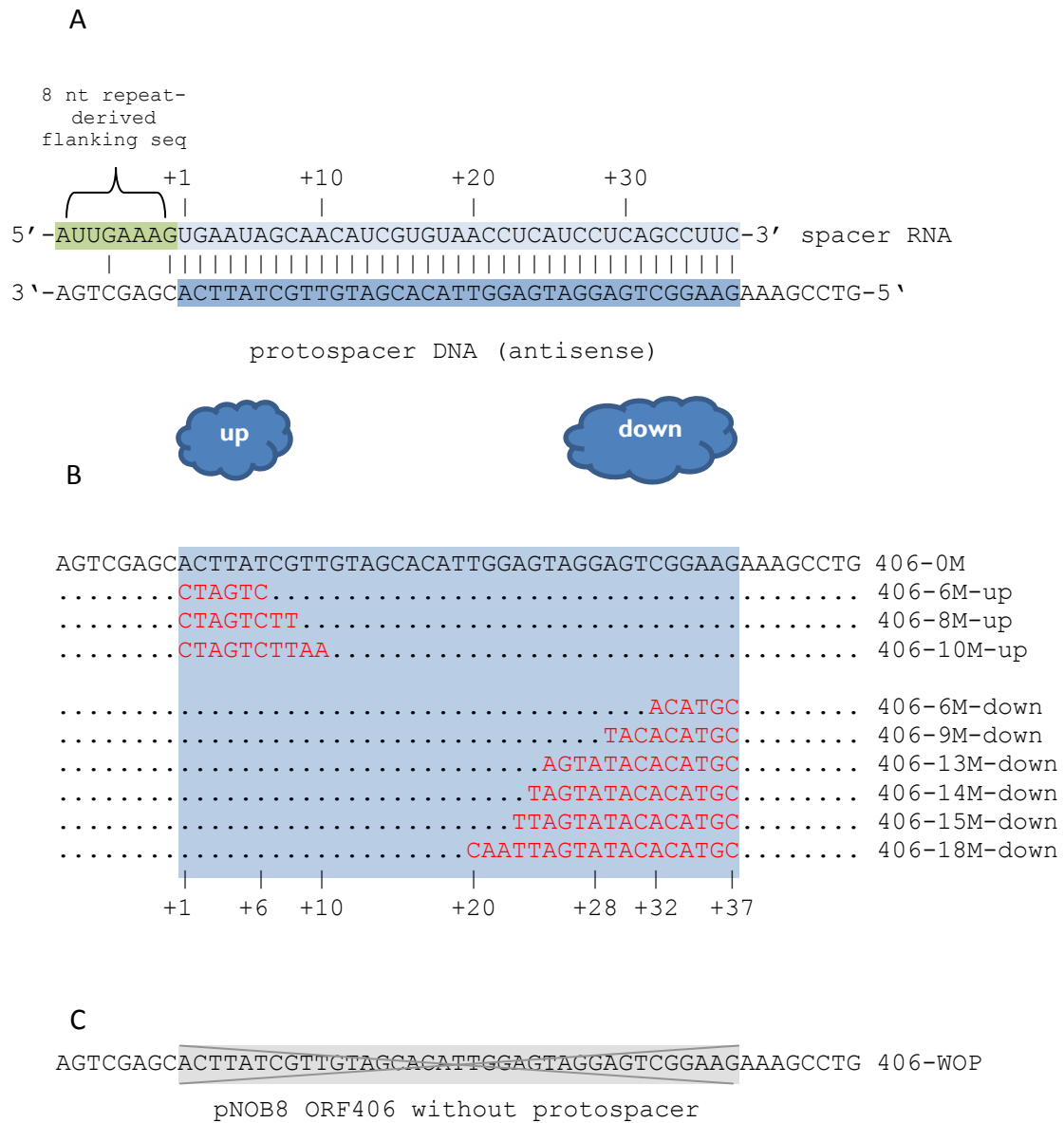
pNOB8 ORF406 without protospacer

Figure 7: A) RNA sequence of spacer 53 (highlighted in light blue). The green region represents the 8 nt repeat-derived flanking sequence that is predicted to stay attached to the processed crRNA. The sequence below represents the corresponding shuttle virion SSV1 encoded protospacer (-) DNA sequence. "Up" describes the protospacer region adjacent to the 8 nt repeat-derived flanking sequence. "Down" corresponds to the opposite protospacer region. B) DNA sequence of sp53 corresponding protospacer (406-0M) as well as DNA sequences of engineered protospacer mutants (below). Introduced mutations are indicated in red. The 37 nt long protospacer region within ORF406 is highlighted in blue. C) Control construct 406-WOP lacks the sp53 corresponding protospacer sequence.

5.2 Generation of protospacer variants

5.2.1 Site-directed mutagenesis via inverse PCR (Polymerase Chain Reaction)

The application of inverse polymerase chain reaction mutagenesis enabled the introduction of mutations at defined regions in the protospacer sequence by amplifying the entire entry vector carrying the corresponding open reading frame, ORF406 at the same time. For the purpose of PCR mutagenesis two complementary primers, one of them carrying the desired mutations were used for amplification, in which 1-10 ng of ORF406 containing entry vector (pEnter-406) was used as a template. The extension of mutagenic primers results in the incorporation of the desired mutations (Figure 8).
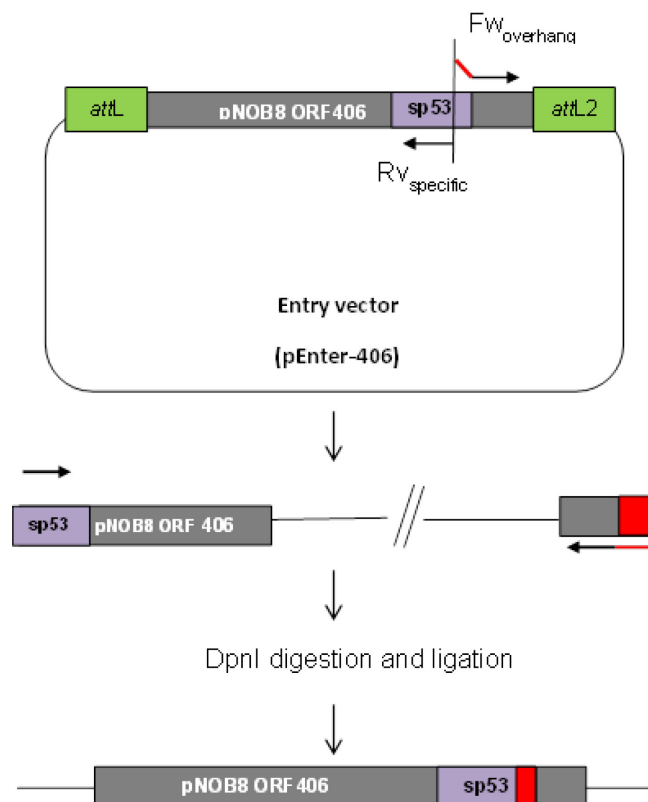

Figure 8: Scheme of applied polymerase chain reaction (PCR) mutagenesis. Red indicates mutations introduced via the usage of mutagenic primers. Fw=forward primer, Rv=reverse primer

Subsequent to PCR mutagenesis, the amplification products were checked for their nucleotide lengths expected to be 4 kb in size. As expected, the application of the respective mutagenic primers resulted in the amplification of the 4 kb vector DNA. However, in addition to an intense DNA band exhibiting the correct length of 4,000 bp's, faint unspecific bands of various sizes could be detected on the agarose gel. Thus, the PCR reaction mixtures were loaded on a preparative gel and only those bands being 4 kb in size were excised from the gel (Figure 9) and used for subsequent vector re-closure.

Figure 9 shows obtained amplification products generated by PCR mutagenesis that were selected for the subsequent ligation and transformation procedure. Applied annealing temperatures are cited next to the corresponding amplification product name.
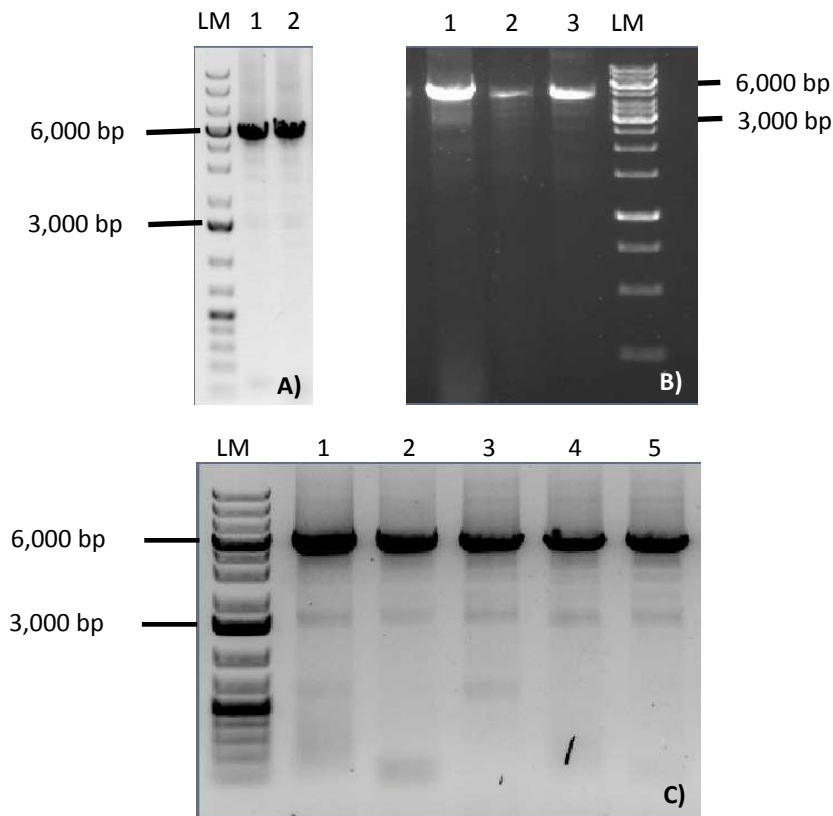
Figure 9: 0.8 % agarose gels (stained with ethidium-bromide) showing products of mutagenesis PCR. Entry vector was used as a template. Amplification was done with specifically designed mutagenic primers using two different annealing temperatures for each primer pair. **A**: 1) pEnter-406-15Mdown; t=56 °C 2) pEnter-406-13Mdown; t=56 °C. **B**: 1) pEnter-406-14Mdown; t=60.2 °C 2) pEnter-406-6Mup; t=60.2 °C 3) pEnter-406-6Mdown; t=60.2 °C **C**: 1) pEnter-406-WOP; t=57 °C 2) pEnter-406-8Mup; t=57 °C 3) pEnter-406-10Mup; t=57 °C. 4) pEnter-406-9Mdown; t=57 °C 5) pEnter-406-18Mdown; t=57 °C. **LM**: DNA ladder mix; 1 kb. **t**: annealing temperature.

In order to enable ligation of the mutated linear vector DNA, primers used for PCR mutagenesis were phosphorylated prior to the amplification procedure. After *Dpn*I digestion the amplified vector was blunt-end ligated by the addition of T4 DNA-ligase and used to transform chemically competent *E. coli* TOP10 cells.

5.2.2   Verification of the mutated protospacer regions

As exemplified by the agarose gel in Figure 10 , three to five single colonies per protospacer construct were selected and initially screened for the presence of ORF406. For this purpose primer pair pNOB8_ORF406_fw and pNOB8_ORF406_rw targeting the respective vector region (ORF406) being 1.2 kb in size was used for colony PCR on obtained *E. coli* clones. The amplification products were then visualized on a 1 % agarose gel. In most cases, a distinct band of the expected size (1.2 kb) was visible on the agarose gel. Each three of the investigated clones that yielded an amplification product of 1.2 kb were chosen for plasmid purification and subsequent Sanger sequencing. The obtained sequences were analyzed by use of ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/) a freely available internet-based alignment tool. Each one of the analysed vectors revealing to carry the desired protospacer mutations was used for cloning of the different ORF406 variants into the SSV1 virus-derived shuttle vector pMJ0305.
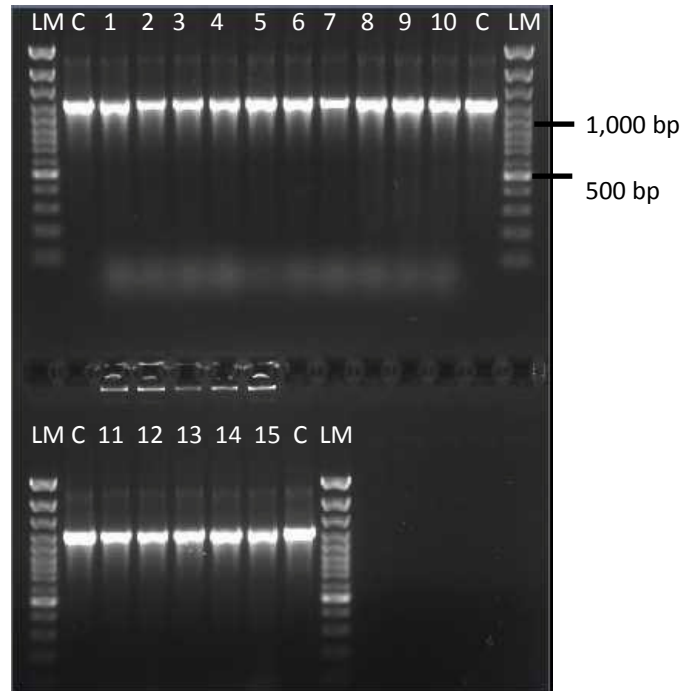
Figure 10: 1 % ethidium-bromide stained agarose gel showing products of colony PCR using ORF406-specific primer pair pNOB8_ORF406_fw and pNOB8_ORF406_rw. Each three to four clones of the different constructs were used as template. **1**) pEnter-406-14Mdown_C1 **2**) pEnter-406-14Mdown_C2 **3**) pEnter-406-14Mdown_C3 **4**) pEnter-406-14Mdown_C4 **5**) pEnter-406-14Mdown_C5 **6**) pEnter-406-6Mup_C1 **7**) pEnter-406-6Mup_C2 **8**) pEnter-406-6Mup_C3 **9**) pEnter-406-6Mup_C4 **10**) pEnter-406-6Mup_C5 **11**) pEnter-406-6Mdown_C1 **12**) pEnter-406-6Mdown_C2 **13**) pEnter-406-6Mdown_C3 **14**) pEnter-406-6Mdown_C4 **15**) pEnter-406-6Mdown_C5. **LM**) DNA ladder mix; 100 bp plus. **C**) control.

5.2.3 Cloning strategy for the generation of SSV1 protospacer mutants

In order to generate different SSV1 protospacer mutants the mutated ORF406 sequences had to be cloned into the *S. solfataricus*/*E. coli* shuttle vector pMJ0305. This was achieved by the application of Gateway® Cloning which is based on the site-specific recombination system of bacteriophage lambda. Both entry (pEnter-406) and destination vector (pMJ0305) carry bacteriophage λ-derived recombination sequences (*att* sites) that allow for recombinational transposition of ORF406 that is situated between two recombination sequences. Based on the recombination procedure initiated by enzymatic activity, the mutated ORF406 sequences were integrated into the modified destination vector pMJ0305 containing the corresponding recombination cassette. Unreacted destination vectors still carry the *ccd*B gene that prevents growth of cells taking those vectors. This enabled the selection for recombined destination

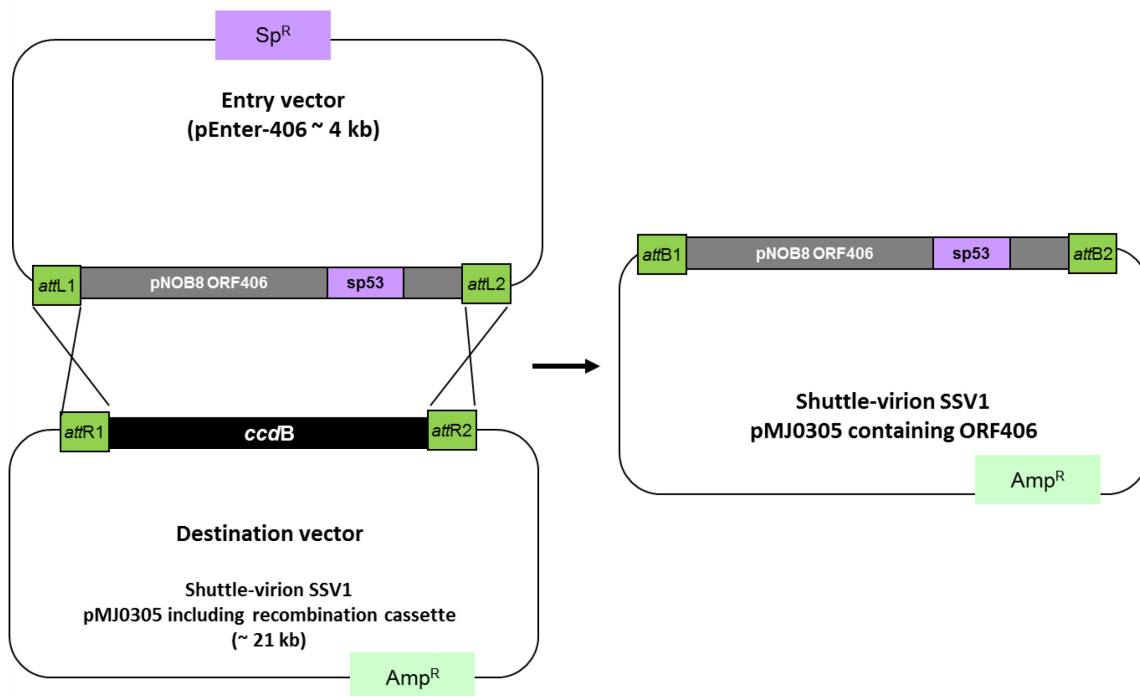vectors. Figure 11 depicts a simplified scheme of the applied cloning strategy based on Gateway® Cloning.

Figure 11: Principle of applied Gateway® Cloning. Mutated versions of ORF406 were transferred from entry vector pEnter-406 to *E*. *coli*/S. *solfataricus* shuttle vector pMJ0305 via recombinational cloning.

Entry and destination vector were added to the reaction mixture by applying a ratio of about 1:1. The mixture was incubated overnight and used for subsequent transformation of *E. coli*. In order to select for the desired transformants the cells were grown on ampicillin [50 $^{\mu g}/_{ml}$] containing LB-plates. In each case a large amount of clones could be obtained, some of which were selected for plasmid purification.

5.2.4 Detection of ORF406 inserts and determination of vector DNA topology

The obtained *E. coli* transformants were examined for the presence of the mutated ORF406 sequence within the SSV1-based destination vector pMJ0305. Hence, three to four clones each were chosen for colony PCR using primer pair pNOB8_ORF406_fw and pNOB8_ORF406_rw. Each two to three positive clones were finally utilized to isolate the DNA of the respective SSV1 shuttle-vector constructs.

As DNA topology affects the efficiency of the transformation/transfection procedure, each construct was checked for its topology prior to transfection of *S. solfataricus*. For this purpose the isolated vector DNA was visualized on a 0.8 % agarose gel and examined for both purity and topology (Figure 12). Only SSV1 shuttle vectors showing correct topology were chosen for transfection experiments.

Figure 12 shows DNA samples of isolated SSV1 vector constructs, some of which exhibiting different DNA topology. Only those samples framed in yellow were used for transfection of *S. solfataricus*. The different bands per lane illustrate distinct topological forms of vector DNA which are explained in Figure 13. The different topological forms of DNA affect its migration velocity during agarose gel electrophoresis.
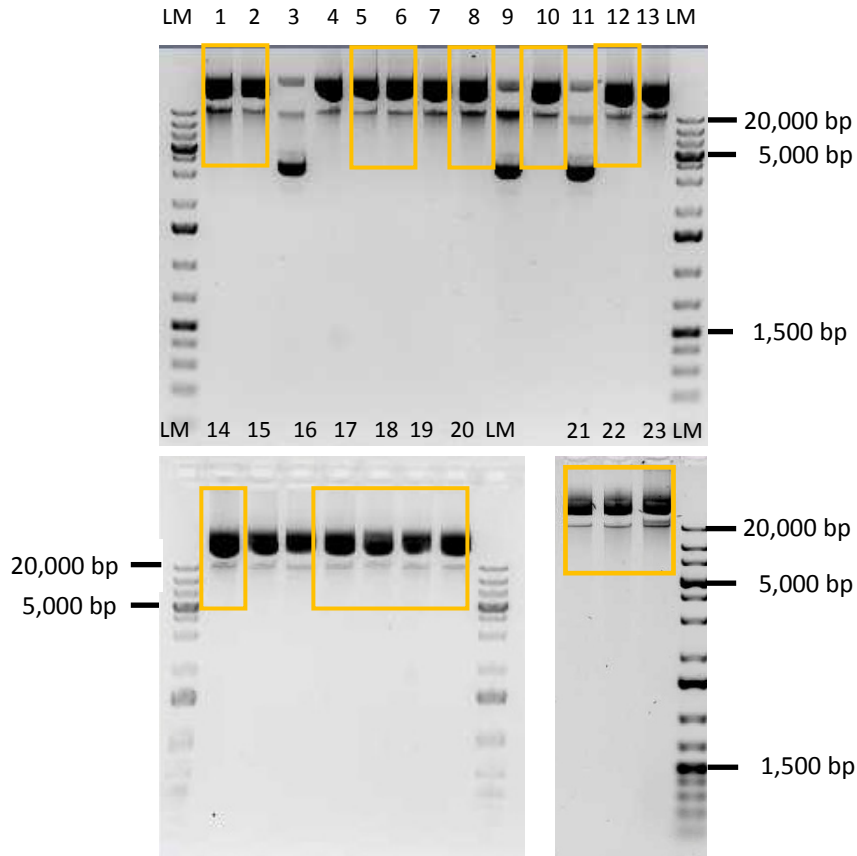
Figure 12: 0.8 % agarose gel showing SSV1 shuttle-vector constructs. 1) pMJ0305-406-9Mdown_C1 2) pMJ0305-406-9Mdown_C2 3) pMJ0305-406-9Mdown_C3 4) pMJ0305-406-18Mdown_C1 5) pMJ0305-406-18Mdown_C2 6) pMJ0305-406-18Mdown_C3 7) pMJ0305-406-WOP_C1 8) pMJ0305-WOP_C2 9) pMJ0305-406-WOP_C3 10) pMJ0305-406-8Mup_C1 11) pMJ0305-406-8Mup_C2 12) pMJ0305-406-8Mup_C3 13) pMJ0305-406-10Mup_C1 14) pMJ0305-406-10Mup_C2 15) pMJ0305-406-10Mup_C3 16) pMJ0305-406-13Mdown_C1 17) pMJ0305-406-13Mdown_C2 18) pMJ0305-406-14Mdown_C1 19) pMJ0305-406-6Mup_C1 20) pMJ0305-406-6Mdown_C1 21) pMJ0305-406-15Mdown_C1 22) pMJ0305-406-15down_C2 23) Control: pMJ0305. LM) DNA ladder mix; 1 kb plus. Orange squares: SSV1 shuttle-vector constructs used for transfection experiments.

Figure 13: 0.8 % agarose gel visualizing **1**) open circle (upper band) and linear form of SSV1 shuttle-vector DNA. **2**) open circle (upper band), linear (middle band) and supercoiled (lower band) form of SSV1 shuttle-vector DNA. **LM**) DNA ladder mix; 1 kb plus.

5.3 *In vivo* CRISPR/Cas mediated interference studies in *Sulfolobus solfataricus*

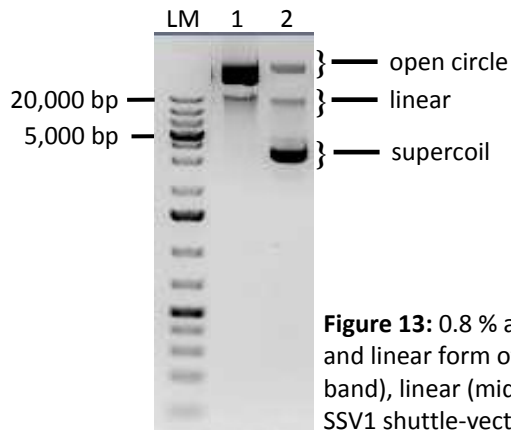### 5.3.1 Transfection of strain *S. solfataricus* P2 carrying the corresponding spacer

Sulfolobus solfataricus strain P2 contains several clusters of CRISPR-associated repeats. One of these loci (CRISPR locus CR3) includes spacer 53 whose sequence is corresponding, more or less (dependent on the respective construct) to the ORF406 protospacer region of the different SSV1 shuttle-vector constructs. Thus, strain P2 carrying spacer 53 was used as a host in order to provide potential immunity against the engineered SSV1 protospacer mutants. Exponentially grown overnight cultures of *S. solfataricus* P2 were used as recipients in electroporation experiments with the various vector constructs. The constructs were tested by using either the same (for constructs 406-6Mup, 406-6Mdown, 406-13Mdown, 406-14Mdown and 406-15Mdown; see also Figure 12) or two different (for constructs 406-WOP, 406-8Mup, 406-10Mup, 406-9Mdown, 18Mdown; see also Figure 12) vector DNA preparations each, in various transfection experiments. Each construct was tested in at least four independent experiments. In order to supply equal amounts of vector DNA when electroporating P2 competent cells, the DNA concentration of each construct was

previously determined by spectrophotometry. Subsequent to regeneration of the electroporated cells, the cell suspension was used in plaque assays.

5.3.2 Visualization of transfectants in plaque assays

To allow for the detection of successful viral infection, transfected *S. solfataricus* P2 cells were used in plaque tests. For this purpose, a defined volume (see section 4.3.2) of the transfected cell suspension was mixed with spacer-containing strain P2 host cells in melted gelrite and plated on the surface of a Brock's sucrose-yeast medium containing gelrite plate. The transfection mixtures were plated in duplicates or triplicates, respectively. If the introduced viral shuttle-vector successfully proliferates within a host cell, it will initiate infection of surrounding host cells growing on the plate. In consequence of virus infection, those cells are retarded in growth and can be seen as a light area, called plaque in the lawn of P2 overlay cells. Plaques obtained from different viral vector constructs were comparable in form and size whereas constructs yielding comparatively low transfection efficiencies generally exhibited plaques being smaller in size (compare Figure 14).
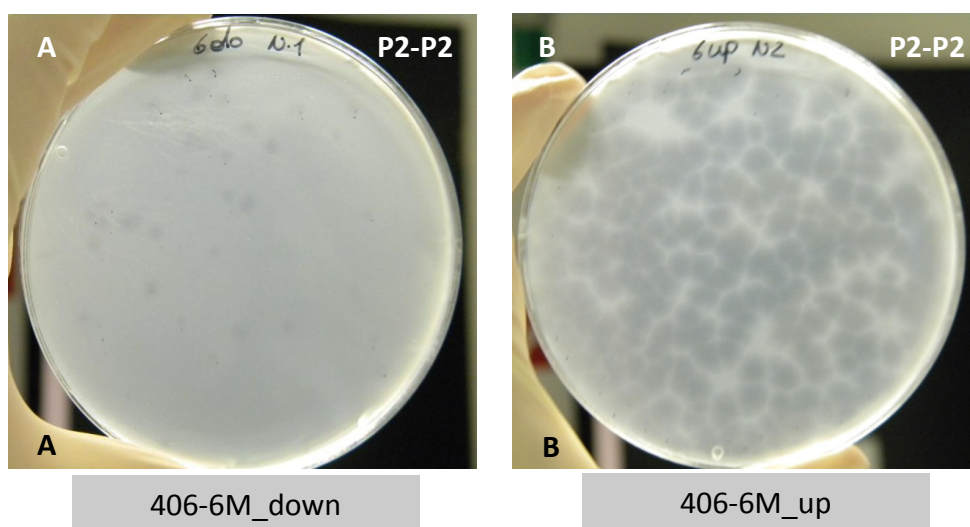

Figure 14: Effects of CRISPR/Cas-mediated immunity on plaque formation. Shown plaques were obtained after plating cells of *S. solfataricus* P2 transfected with virus construct 406-6M-down (A) or 406-6M-up (B) on cell lawns of *S. solfataricus* P2 carrying the respective spacer.

5.3.3 Quantification of transfectants by determining plaque forming units (PFU's)

To obtain a measure of efficient transfection of P2 cells the number of plaque forming units (PFU) had to be determined for each experiment. For this purpose, plaques obtained from the various plaque assay experiments were counted and transfection efficiencies of the different SSV1 shuttle-vector constructs were calculated per µg of applied vector DNA. As maximal transfection efficiencies for strain P2 varied in the separate plaque tests, the calculated transfection efficiencies were normalized to the highest plaque count obtained in each single experiment.

In order to explore if differently positioned mutations within the protospacer may entail differences in CRISPR/Cas-mediated immune capacity, initially two different SSV1 shuttle-vector constructs namely 406-6Mup (mutations at positions +1 to +6; see Figure 7), 406-6Mdown (mutations at positions +32 to +37; see Figure 7.) were generated and tested in plaque assays. As depicted in Figure 15 the two constructs showed significant differences when comparing transfection efficiencies. Construct 406-6Mup reached 75 % (see Table 2) transfection efficiency whereas 406-6Mdown showed only about 4 % (Table 2) transfection efficiency. To further investigate if there is indeed a difference in the need of base complementarity between the two sides, six additional protospacer mutants were generated. These included two constructs carrying eight mutations at protospacer "up" positions +1 to +8 (406-8Mup) and ten mutations at positions +1 to +10 (406-10Mup), respectively as well as four constructs exhibiting a stepwise increase of consecutive mutations at the opposite side ("down") of the protospacer sequence. These "down" side mutations were situated at positions +29 to +37 (406-9Mdown), +25 to +37 (406-13Mdown), +24 to +37 (406-14Mdown), +23 to +37 (406-15Mdown) and at positions +20 to +37 (406-18Mdown). The respective DNA sequences of the various protospacer constructs are depicted and specified in Figure 7.

When constructs 406-8Mup and 406-10Mup were tested in plaque assays, both caused a massive reduction in CRISPR/Cas interference obtaining only slightly lower transfection efficiencies (Figure 15, Table 2) than the viral shuttle-vector without protospacer (406-WOP).

However, as evident from Figure 15 only 18 mutations (406-18Mdown) in the protospacer "down" region totally restrained (90 % transfection efficiency) CRISPR/Cas -mediated immune response. When 15 consecutive mutations (construct 406-15down) were introduced at the "down" region of the protospacer, the system was still able to trigger immunity in about 50 % of transfection events (Figure 15). Moreover the obtained values reveal that the increase of mutations referred to both sites of the protospacer gradually lowers CRISPR interference, albeit considerably more effective within the protospacer "up" region.

Unexpectedly, the construct 406-13Mdown yielded significantly higher transfection efficiencies than constructs 406-14Mdown and 406-15Mdown despite its lower degree of mutation. Even though transfection efficiencies obtained from construct 406-15Mdown were higher than those obtained from 406-14Mdown, they still tended to fall below those of 406-13Mdown.
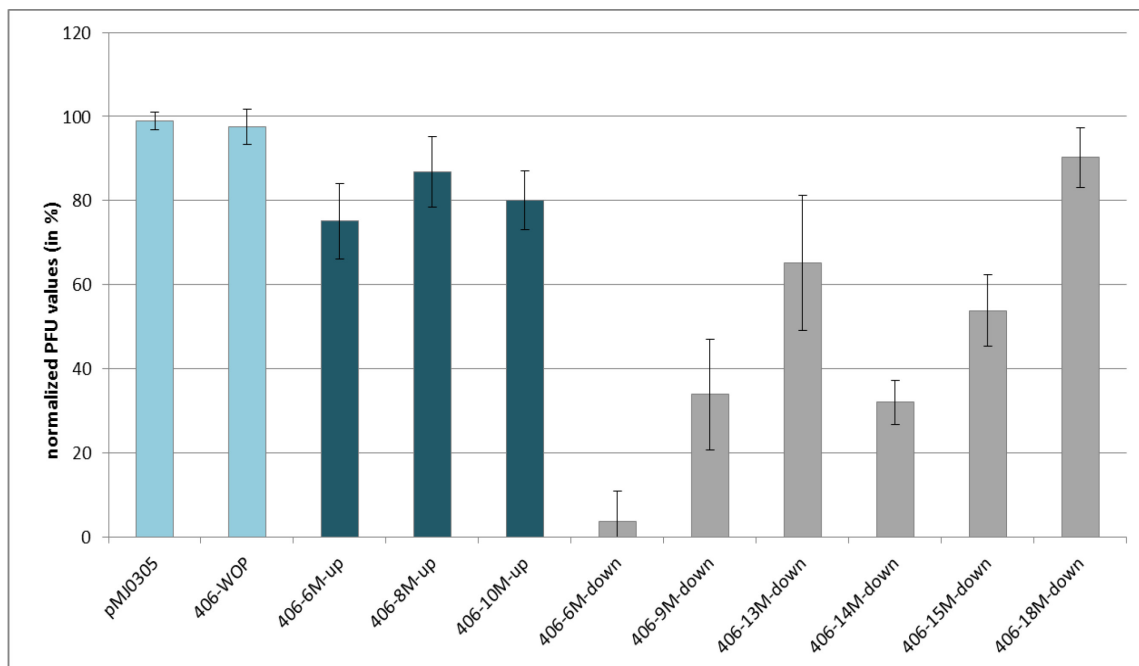

Figure 15: Bar chart showing efficiencies of plating (in %) for different recombinant SSV1 protospacer mutants. Transfection efficiencies were normalized to the highest plaque count obtained in each single experiment. Each construct was tested in at least four independent transfection experiments. Error bars indicate standard deviations.

Table 2: Transfection efficiencies obtained for recombinant virus SSV1 carrying different protospacer inserts.

Strain	Construct	Average	SD
P2	pMJ0305	9,89E-01	2,11E-02
	406-WOP	9,76E-01	4,24E-02
	406-6M-up	7,51E-01	9,04E-02
	406-8M-up	8,68E-01	8,43E-02
	406-10M-up	8,01E-01	7,01E-02
	406-6M-down	3,62E-02	7,29E-02
	406-9M-down	3,39E-01	1,32E-01
	406-13M-down	6,52E-01	1,60E-01
	406-14M-down	3,20E-01	5,24E-02
	406-15M-down	5,39E-01	8,51E-02
	406-18M-down	9,02E-01	7,01E-02

a. ≥ 4 independent transfection experiments for each construct.
SD, standard deviation

5.3.4 Transfection of construct 406-13Mdown, 406-14Mdown and 406-15Mdown using *S. solfataricus* strain M18 as host cells

Contrary to expectation construct 406-13Mdown obtained higher transfection efficiencies than construct 406-14Mdown and even 406-15Mdown. Based on that observation we initially speculated that arising transfection efficiencies of these three constructs might have been biased by differences in viral vector DNA quality or inaccurate DNA quantification. To exclude the possibility that CRISPR/Cas function unrelated factors affected calculated transfection efficiencies, constructs 406-13Mdown, 14M-down and 15Mdown were checked for their capacity to infect *S. solfataricus* strain M18 as host cells. *S. solfataricus* M18 is a close relative of strain P2 but differs in spacer composition including the absence of ORF406-specific spacer 53. For subsequent plaque tests again cells of *S. solfataricus* strain M18 instead of P2 were used in overlays.

Nearly equally high values were obtained for constructs 406-13Mdown, 14Mdown, 15Mdown and the ORF406-missing viral vector pMJ0305, used as a control in the transfection experiment (Figure 16, Table 3). Due to comparatively large plaque sizes and difficulties with transfection of M18 cells, respectively proper PFU values could

only be calculated for a single (out of four) transfection experiment. Consequently, standard deviations for transfection efficiencies of *S. solfataricus* M18 are missing in Figure 16 and Table 3.
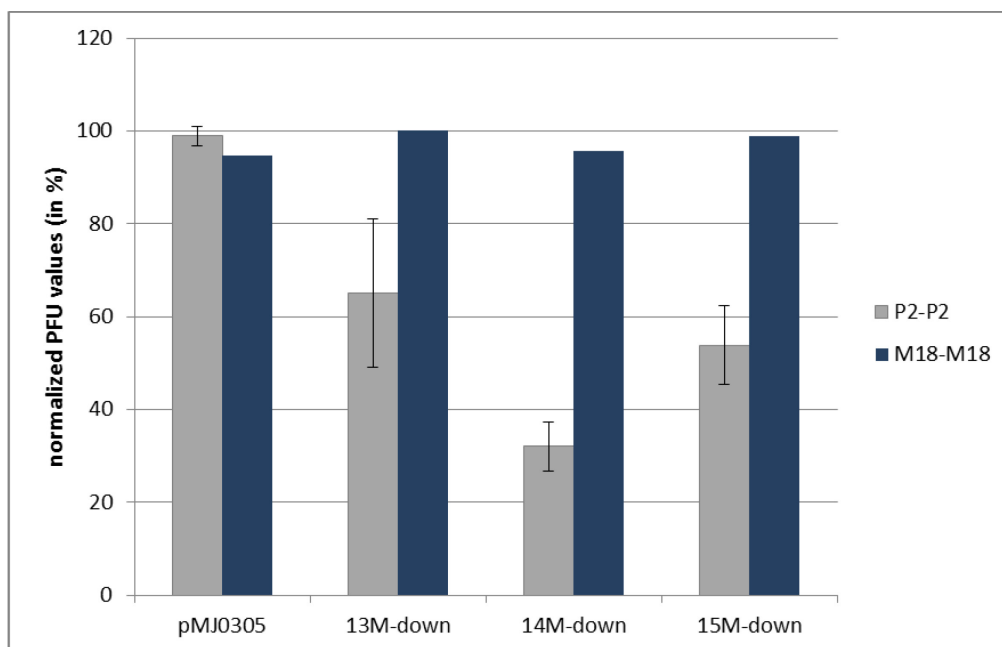
Figure 16: Bar chart showing efficiencies of plating (in %) for constructs 406-13Mdown, -14Mdown and -15Mdown when using either *S. solfataricus* strain P2 (grey bars) or strain M18 (blue bars) that is missing the complementary spacer for transfection experiments. pMJ0305 was used as a control. Error bars indicate standard deviations.

Table 3: Transfection efficiencies obtained for recombinant virus SSV1 carrying different protospacer inserts.

Strain	Construct	Average	SD
P2	pMJ03	9,89E-01	2,11E-02
	13M-down	6,52E-01	1,60E-01
	14M-down	3,20E-01	5,24E-02
	15M-down	5,39E-01	8,51E-02
M18	pMJ03	9,48E-01	
	13M-down	1	
	14M-down	9,56E-01	
	15M-down	9,88E-01	

a. ≥ 4 independent transfection experiments for strain P2.
SD, standard deviation

6 Discussion

This study aimed to closer investigate the requirements for base-pairing between crRNA and protospacer during CRISPR interference in the hyperthermophilic archaeon *S. solfataricus*.

The development of an *in vivo* test system allowed to confirm CRISPR activity in *S. solfataricus* and has been achieved by the implementation of an engineered protospacer carrying viral vector (Manica et al., 2011). That system is based on the SSV1 viral shuttle-vector pMJ0305 carrying an artificially introduced protospacer sequence and served in this study to test several different protospacer mutants for their effect on CRISPR-mediated immune response in transfection experiments.

The previous study of Manica *et al*. has already revealed that up to three mutations within the engineered pNOB8-derived ORF406 protospacer sequence of the recombinant SSV1 virus are tolerated and still allow triggering CRISPR/Cas immunity (Manica *et al.*, 2011). Within the scope of this diploma thesis, additional ORF406 variants carrying different mutations within the corresponding target protospacer region were generated to gain a better understanding of CRISPR/Cas interference functioning and requirements, respectively. For this purpose polymerase chain reaction mutagenesis was performed to obtain desired protospacer mutants. The resulting ORF406 constructs were cloned into the viral shuttle-vector pMJ0305 by the application of Gateway® Cloning and used for subsequent transfection of *S. solfataricus* strain P2 that contains the corresponding spacer within one of its CRISPR loci. The application of plaque assays finally enabled to detect and compare *in vivo* CRISPR activity that has been triggered by the different viral protospacer mutants.

When constructs 406-6Mup and 406-6Mdown, carrying six mutations each but on different sides of the protospacer, were tested in transfection experiments they revealed an obvious variation in efficiency of transformation. While 406-6Mup (mutations at positions +1 to +6) obtained 75 % transfection efficiency, construct 406-6Mup (mutations at positions +32 to +37) hardly caused plaque formation showing only 4 % transfection efficiency. The fact that mutations at positions +32 to +37 allow substantially less viral escape compared with mutations occurring at the opposite end

of the protospacer suggests that CRISPR interference is strongly dependent on base-pair formation at crRNA positions that are immediately followed at the repeat-derived flanking sequence. This initial hypothesis was supported when mutations were gradually increased at both sides of the protospacer. Contrary to mutations occurring at positions +1 to +10, the system appeared to be more permissive towards mutations present at opposite positions ("down") within the protospacer sequence. While eight mutations at positions +1 to +8 hardly triggered CRISPR interference activity, fifteen mutations at positions +23 to +37 still conferred viral resistance in 50 % of transfection events. Surprisingly, thirteen mutations (positions +25 to +37) at the "down" end conferred significantly less resistance (65 compared to 32 % transfection efficiency; see Figure 15 and Table 2) than fourteen mutations (positions +24 to +37) at the same end of the protospacer. Transfection of *S. solfataricus* strain M18 (see section 5.3.4) that does not contain the ORF406-specific spacer supported that the obtained, unexpected high transfection efficiency for the thirteen mutations carrying construct (406-13Mdown) has not been affected by some CRISPR/Cas function unrelated factors. Thus suggesting that the unexpected (compared to construct 406-14Mdown and 406-15Mdown) high number of plaques (per µg of DNA) obtained when strain P2 was transfected with construct 406-13Mdown, may be ascribed to a characteristic of the CRISPR/Cas mechanism that is currently unknown.

Similar to observations made in this study, Semenova *et al.* have only recently reported that single mutations within the first seven nucleotides of the protospacer called "seed" (corresponding to crRNA positions immediately following the 5'-repeat-derived flanking sequence) led to viral escape while multiple protospacer mutations (up to four) outside the "seed" region did not affect CRISPR/Cas interference in *E. coli.* Consequently, they suggest that target recognition is initially based on a seven-nucleotide protospacer "seed" region only then followed by full-length crRNA-protospacer interaction (Semenova *et al.*, 2011). Thus, similar to *E. coli*, CRISPR/Cas interference in *S. solfataricus* might be initiated by the recognition of a short sequence at the "up"-end (complementary to the 5'-end of spacer RNA) of the protospacer. This strategy may reflect a general characteristic of both archaeal and bacterial CRISPR/Cas systems facilitating a rapid detection of invading nucleic acid molecules. Additionally supportive is the work on a CRISPR associated protein complex isolated from

Pseudomonas aeruginosa showing high binding affinity when crRNA binds oligonucleotides matching the first 8 nucleotides from the 5'-end of RNA spacer sequence compared to weak affinity interaction between crRNA and nucleotides located outside the "seed" sequence (Wiedenheft *et al.,* 2011).

However, the indication that only a partial spacer sequence may already be sufficient for CRISPR interference raises the question why comparatively long DNA sequences are selected for spacer incorporation. One possible explanation would be the need for a minimal crRNA length to ensure stable complex formation of required Cas proteins during the interference process. Besides, the obtained data suggest that once the target has been recognized via sequence similarity at the 5' end of spacer RNA the interference process also requires a minimal degree of base pairing at the crRNA 3' proximal end. Lintner *et al*. achieved to isolate and characterize an archaeal CRISPR-associated protein complex from *S. solfataricus* showing a helical structure of variable length that may result from progressive aggregation of various complex components during CRISPR/Cas interference (Lintner *et al.,* 2011). Even though several mutations at the spacer RNA 3' end are accepted without affecting CRISPR/Cas interference (compare construct 406-6Mdown) too many of them obviously weaken (compare constructs 406-9Mdown, -13Mdown, -14Mdown, 15Mdown) or even abolish (compare construct 406-18Mdown) interference. This might have been induced by resulting nucleic acid distortions preventing the correct positioning of processing Cas proteins despite target recognition in the first place. The existence of relatively "long" spacers exceeding the sequence information that is apparently sufficient for target recognition might additionally increase the probability to find additional cutting sites within protospacers, provided these are sequence specific. So far CRISPR/Cas-induced cleavage of invading DNA or RNA, respectively has only been found to occur sequence-unspecific. Garneau *et al* were able to demonstrate CRISPR/Cas-mediated *in vivo* cutting of invading DNA in *Streptococcus thermophilus* indicating cleavage at consistent, sequence-unspecific positions within analyzed protospacers (Garneau *et al.,* 2010). *In vitro* studies from *Pyrococcus* suggest that protospacer RNA instead of DNA underlies cleavage that occurs at a fixed 14nt distance from the 3' end of crRNA (Hale *et al.,* 2009). In contrast, the studies shown in this work might suggest that sequence-specific cleavage occurs: the unexpected low immune response (compared

to construct 406-14Mdown and -15Mdown) obtained from construct 406-13Mdown might imply that cutting happens at "AT"-sequences within the protospacer. Comparing these three viral vector constructs, 406-13Mdown exhibits one "AT" less than constructs 406-14Mdown and -15Mdown both of which contain 5 times "AT" (Figure 7) thus reducing the probability of actual cleavage within the protospacer sequence. Nevertheless, this hypothesis is highly speculative and needs to be verified by further experimental data.

Various researchers suggested that mutations within so called PAM (protospacer adjacent motif) sequences may allow viruses to avoid CRISPR/Cas interference (Deveau *et al*; 2008; Mojica *et al*; 2009; Semenova *et al*; 2011). PAMs have primarily been mentioned by Mojica *et al*. and describe conserved nucleotide motifs that precede protospacer sequences (Mojica *et al.,* 2009). Within this study substitutions at flanking positions upstream the ORF406-derived protospacer sequence have not been checked for their effect on CRISPR/Cas mediated viral defense. The tested protospacer mutants carry identical flanking sequences indicating that obtained transfection efficiencies have not been affected or even caused by mutations in protospacer proximal positions. Viral escape induced by mutations occurring at protospacer adjacent positions may rather be ascribed to growing complementarity between the repeat-derived flanking sequence of spacer RNA and corresponding protospacer flanking regions than to absent PAMs. Those repeat-derived flanking sequences have been shown to play a crucial role in the avoidance of autoimmunity (Marraffini and Sontheimer 2010).

As already reported by Manica *et al* (Manica et al., 2011) CRISPR/Cas based immunity of *S. solfataricus* was not dependent on target transcription. As the tested recombinant SSV1 variants in this study carried no promoter sequence in front of the protospacer containing gene this strongly implies that immunity resulted from CRISPR-mediated degradation of invading viral DNA.

Most notably, compared to *E. coli* the CRISPR/Cas complex in *S. solfataricus* seems to be remarkably tolerant towards mutations distinguishing target and spacer. An increased tolerance towards occurring protospacer mutations may reflect an adaptation to fast evolving target sequences providing the host organism a more flexible recognition of related invading elements without the need of incorporating

additional spacers. If the ability to successfully target an extraordinary broad spectrum of related viruses by a single spacer is restricted to the archaeal domain of life remains to be clarified by future research work.

Regardless of applied viral protospacer mutants, the plaques showed no visible differences with respect to their turbidity. Comparing plaques that were e.g. obtained from constructs 406-6Mup and 406-6Mdown (Figure 14) varieties could only be seen in the sizes of their halos. Considering the experimental set-up of applied plaque tests, plaque formation results from a huge number of infections initially caused by propagation of an artificially introduced viral DNA molecule able to escape CRISPR/Cas interference. Hence, acquired transfection efficiency data derive from CRISPR/Cas activity upon the first infection event during electroporation. Once, the incoming viral DNA escapes the first CRISPR/Cas confrontation, replication results in the release of viral particles. Even though each subsequent infection event might be challenged by the CRISPR/Cas system the consistency in plaque turbidity suggests that most surrounding host cells face an infection. Probably due to the high dosage of viral particles, the CRISPR/Cas system cannot cope with infections anymore. The probability of interference might therefore primarily be represented by the number of resulting plaque forming units than by differences in plaque appearance.

Varying levels of CRISPR/Cas immune response may allow for suppression and survival of viral populations that might under certain conditions be advantageous for the host organism. Reduced CRISPR/Cas-mediated interference might allow for controlled propagation of viral populations without substantial decrease of host density. Hence, viruses may be maintained over several generations within microbial populations allowing the conservation of foreign DNA even though the viral development is affected by the CRISPR/Cas system. As viral infection may result in the introduction of new genes that can under certain environmental conditions provide beneficial functions to the host (Madigan and Brock 2009), such strategy might additionally help to adapt to challenging environments and to invade new niches (reviewed in Roossinck, 2011). Besides, the maintenance of viral populations might also be advantageous when competing with other organisms not being resistant to the virus. However, to support these hypotheses, it needs to be verified if a viral population

suppressed to a certain extent by CRISPR/Cas could indeed be sustained over several generations of host populations.

These hypotheses suggest that CRISPR/Cas strongly affects viral-host interactions determining besides other factors the intensity of new infections and the fitness of the host. Thus, the system might play a central role concerning developments and changes in microbial population structures and dynamics. However, future studies that help to better understand the system's role within complex microbial communities are required to clarify the ecological relevance of the highly diverse CRISPR/Cas systems.

7 References

Al-Attar, S., E. R. Westra, J. van der Oost and S. J. Brouns (2011). "Clustered regularly interspaced short palindromic repeats (CRISPRs): the hallmark of an ingenious antiviral defense mechanism in prokaryotes." Biol Chem **392**(4): 277-289.

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**(5819): 1709-1712.

Berkner, S., D. Grogan, S. V. Albers and 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**(12): e88.

Bernard, P., K. E. Kezdy, L. Van Melderen, J. Steyaert, L. Wyns, M. L. Pato, P. N. Higgins and M. Couturier (1993). "The F plasmid CcdB protein induces efficient ATP-dependent DNA cleavage by gyrase." J Mol Biol **234**(3): 534-541.

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

Brochier-Armanet, C., B. Boussau, S. Gribaldo and P. Forterre (2008). "Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota." Nat Rev Microbiol **6**(3): 245-252.

Brock, T. D., K. M. Brock, R. T. Belly and R. L. Weiss (1972). "Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature." Arch Mikrobiol **84**(1): 54-68.

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 and J. van der Oost (2008). "Small CRISPR RNAs guide antiviral defense in prokaryotes." Science **321**(5891): 960-964.

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

De Rosa, M., A. Gambacorta and A. Gliozzi (1986). "Structure, biosynthesis, and physicochemical properties of archaebacterial lipids." Microbiol Rev **50**(1): 70-80.

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

encoded resistance in Streptococcus thermophilus." J Bacteriol **190**(4): 1390-1400.

Elkins, J. G., M. Podar, D. E. Graham, K. S. Makarova, Y. Wolf, L. Randau, B. P. Hedlund, C. Brochier-Armanet, V. Kunin, I. Anderson, A. Lapidus, E. Goltsman, K. Barry, E. V. Koonin, P. Hugenholtz, N. Kyrpides, G. Wanner, P. Richardson, M. Keller and K. O. Stetter (2008). "A korarchaeal genome reveals insights into the evolution of the Archaea." Proc Natl Acad Sci U S A **105**(23): 8102-8107.

Forterre, P., C. Brochier and H. Philippe (2002). "Evolution of the Archaea." Theor Popul Biol **61**(4): 409-422.

Garneau, J. E., M. E. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadan and S. Moineau (2010). "The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA." Nature **468**(7320): 67-71.

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.

Grogan, D. W. (1989). "Phenotypic characterization of the archaebacterial genus Sulfolobus: comparison of five wild-type strains." J Bacteriol **171**(12): 6710-6719.

Haft, D. H., J. Selengut, E. F. Mongodin and 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**(6): e60.

Hale, C. R., P. Zhao, S. Olson, M. O. Duff, B. R. Graveley, L. Wells, R. M. Terns and M. P. Terns (2009). "RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex." Cell **139**(5): 945-956.

Heidelberg, J. F., W. C. Nelson, T. Schoenfeld and D. Bhaya (2009). "Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes." PLoS One **4**(1): e4169.

Horvath, P., D. A. Romero, A. C. Coute-Monvoisin, M. Richards, H. Deveau, S. Moineau, P. Boyaval, C. Fremaux and R. Barrangou (2008). "Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus." J Bacteriol **190**(4): 1401-1412.

Huber, H., M. J. Hohn, R. Rachel, T. Fuchs, V. C. Wimmer and K. O. Stetter (2002). "A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont." Nature **417**(6884): 63-67.

Ishino, Y., H. Shinagawa, K. Makino, M. Amemura and 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**(12): 5429-5433.

Jansen, R., J. D. Embden, W. Gaastra and L. M. Schouls (2002). "Identification of genes that are associated with DNA repeats in prokaryotes." Mol Microbiol **43**(6): 1565-1575.

Jonuscheit, M., E. Martusewitsch, K. M. Stedman and C. Schleper (2003). "A reporter gene system for the hyperthermophilic archaeon Sulfolobus solfataricus based on a selectable and integrative shuttle vector." Mol Microbiol **48**(5): 1241-1252.

Kandler, O. and H. Konig (1998). "Cell wall polymers in Archaea (Archaebacteria)." Cell Mol Life Sci **54**(4): 305-308.

Kunin, V., R. Sorek and P. Hugenholtz (2007). "Evolutionary conservation of sequence and secondary structures in CRISPR repeats." Genome Biol **8**(4): R61.

Kurosawa, N. and D. W. Grogan (2005). "Homologous recombination of exogenous DNA with the Sulfolobus acidocaldarius genome: properties and uses." FEMS Microbiol Lett **253**(1): 141-149.

Landy, A. (1989). "Dynamic, structural, and regulatory aspects of lambda site-specific recombination." Annu Rev Biochem **58**: 913-949.

Lillestol, R. K., P. Redder, R. A. Garrett and K. Brugger (2006). "A putative viral defence mechanism in archaeal cells." Archaea **2**(1): 59-72.

Lillestol, R. K., S. A. Shah, K. Brugger, P. Redder, H. Phan, J. Christiansen and R. A. Garrett (2009). "CRISPR families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties." Mol Microbiol **72**(1): 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 and C. M. Lawrence (2011). "Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE)." J Biol Chem **286**(24): 21643-21656.

Madigan, M. T. and T. D. Brock (2009). Brock biology of microorganisms. San Francisco, CA, Pearson/Benjamin Cummings.

Makarova, K. S., L. Aravind, Y. I. Wolf and 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., N. V. Grishin, S. A. Shabalina, Y. I. Wolf and 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 and E. V. Koonin (2011). "Evolution and classification of the CRISPR-Cas systems." Nat Rev Microbiol **9**(6): 467-477.

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

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). "Self versus non-self discrimination during CRISPR RNA-directed immunity." Nature **463**(7280): 568-571.

Martusewitsch, E., C. W. Sensen and C. Schleper (2000). "High spontaneous mutation rate in the hyperthermophilic archaeon Sulfolobus solfataricus is mediated by transposable elements." J Bacteriol **182**(9): 2574-2581.

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

Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez and E. Soria (2005). "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements." J Mol Evol **60**(2): 174-182.

Mojica, F. J., C. Diez-Villasenor, E. Soria and G. Juez (2000). "Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria." Mol Microbiol **36**(1): 244-246.

Mullis, K. B. and F. A. Faloona (1987). "Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction." Methods Enzymol **155**: 335-350.

Pourcel, C., G. Salvignol and 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**(Pt 3): 653-663.

Prangishvili, D., P. Forterre and R. A. Garrett (2006). "Viruses of the Archaea: a unifying view." Nat Rev Microbiol **4**(11): 837-848.

Roossinck, M. J. (2011). "The good viruses: viral mutualistic symbioses." Nat Rev Microbiol **9**(2): 99-108.

Schleper, C., K. Kubo and 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**(16): 7645-7649.

Schleper, C., R. Roder, T. Singer and W. Zillig (1994). "An insertion element of the extremely thermophilic archaeon Sulfolobus solfataricus transposes into the endogenous beta-galactosidase gene." Mol Gen Genet **243**(1): 91-96.

Semenova, E., M. M. Jore, K. A. Datsenko, A. Semenova, E. R. Westra, B. Wanner, J. van der Oost, S. J. Brouns and 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**(25): 10098-10103.

Shah, S. A., N. R. Hansen and R. A. Garrett (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., H. Phan, R. A. Garrett, S. V. Albers, K. M. Stedman and W. Zillig (1998). "Genetic profile of pNOB8 from Sulfolobus: the first conjugative plasmid from an archaeon." Extremophiles **2**(4): 417-425.

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, *et al.* (2001). "The complete genome of the crenarchaeon Sulfolobus solfataricus P2." Proc Natl Acad Sci U S A **98**(14): 7835-7840.

Sorek, R., V. Kunin and P. Hugenholtz (2008). "CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea." Nat Rev Microbiol **6**(3): 181-186.

Spang, A., R. Hatzenpichler, C. Brochier-Armanet, T. Rattei, P. Tischler, E. Spieck, W. Streit, D. A. Stahl, M. Wagner and C. Schleper (2010). "Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota." Trends Microbiol **18**(8): 331-340.

Stedman, K. M., C. Schleper, E. Rumpf and W. Zillig (1999). "Genetic requirements for the function of the archaeal virus SSV1 in Sulfolobus solfataricus: construction and testing of viral shuttle vectors." Genetics **152**(4): 1397-1405.

Tyson, G. W. and J. F. Banfield (2008). "Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses." Environ Microbiol **10**(1): 200-207.

van der Oost, J., M. M. Jore, E. R. Westra, M. Lundgren and S. J. Brouns (2009). "CRISPR-based adaptive and heritable immunity in prokaryotes." Trends Biochem Sci **34**(8): 401-407.

Watson, J. D. and F. H. Crick (1953). "The structure of DNA." <u>Cold Spring Harb Symp Quant Biol</u> **18**: 123-131.

Weinbauer, M. G. (2004). "Ecology of prokaryotic viruses." <u>FEMS Microbiol Rev</u> **28**(2): 127-181.

Weinbauer, M. G. and F. Rassoulzadegan (2004). "Are viruses driving microbial diversification and diversity?" <u>Environ Microbiol</u> **6**(1): 1-11.

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 and J. A. Doudna (2011). "RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions." <u>Proc Natl Acad Sci U S A</u> **108**(25): 10092-10097.

Winter, C., T. Bouvier, M. G. Weinbauer and T. F. Thingstad (2010). "Trade-offs between competition and defense specialists among unicellular planktonic organisms: the "killing the winner" hypothesis revisited." <u>Microbiol Mol Biol Rev</u> **74**(1): 42-57.

Woese, C. R. and G. E. Fox (1977). "Phylogenetic structure of the prokaryotic domain: the primary kingdoms." <u>Proc Natl Acad Sci U S A</u> **74**(11): 5088-5090.

Woese, C. R., O. Kandler and M. L. Wheelis (1990). "Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya." <u>Proc Natl Acad Sci U S A</u> **87**(12): 4576-4579.

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-dependent RNA-polymerases. <u>Arch. Microbiol</u>. 125: 259-269

Zillig, W., D. Prangishvilli, C. Schleper, M. Elferink, I. Holz, S. Albers, D. Janekovic and D. Gotz (1996). "Viruses, plasmids and other genetic elements of thermophilic and hyperthermophilic Archaea." <u>FEMS Microbiol Rev</u> **18**(2-3): 225-236.

Appendix

A) Abbreviations and definitions

°C	degree celcius
µF	microfarad
µg	microgram
µl	microlitre
µm	micromol
Amp	ampicillin
ATP	adenosine-5'-triphosphate
bp's	base pairs
c	concentration
Cas	CRISPR-associated
CRISPR	clustered regularly interspaced short palindromic repeat
crRNA	CRISPR RNA
ddH$_2$0	double-distilled water
DNA	desoxyribonucleic acid
dNTP	desoxynucleotide triphosphate
ds	double stranded
e.g.	for example
fw	forward
g	gram
kb	kilo bases
l	litre
LB	Luria Bertani
LM	DNA ladder mix
M	molar
min	minutes
ml	millilitre
mM	millimolar
ng	nanogram
nm	nanometer
nt	nucleotide
OD	optical density
ORF	open reading frame
ori	origin of replication
PAM	protospacer adjacent motif; nucleotide sequence flanking the protospacer
PCR	polymerase chain reaction
PFU	plaque forming unit
protospacer	nucleotide sequence in the phage genome corresponding to the

	spacer
RAMP	repeat-associated mysterious protein
RNA	ribonucleic acid
rpm	runs per minute
rw	reverse
Sp	Spectinomycin
sp53	spacer 53
spacer	small nucleotide sequence inserted between two repeats that corresponds to a foreign DNA sequence
ss	single stranded
Taq Polymerase	Thermus aquaticus polymerase
u	units
V	volt
w/v	weight per volume
WOP	without protospacer
x	times
λ	lamda
Ω	ohm

B) Zusammenfassung

In natürlichen Ökosystemen sind Mikroorganismen stets einer enormen Anzahl und Vielfalt an Viren und anderen invasiven Nukleinsäuren ausgesetzt. Infolge dessen haben Mikroben verschiedene Abwehrstrategien entwickelt, um virale Infektionen bzw. den Eintritt von Fremd-DNA zu vermeiden. Clustered regularly interspaced short palindromic repeats (kurz: CRISPR) beschreiben bestimmte sich wiederholende DNA-Abschnitte, die in vielen bakteriellen und den meisten archaealen Genomen auftreten und charakterisieren einen erst kürzlich bekannt gewordenen Abwehrmechanismus. Mittels CRISPR/Cas-Systemen erwerben Mikroorganismen Immunität gegen Infektion durch Viren und Eindringen anderer exogener Nukleinsäuren. Der CRISPR/Cas Mechanismus beruht auf der spezifischen Erkennung eindringender Fremd-DNA durch kleine, komplementäre RNAs, die als transkribierte Spacer-Sequenzen aus dem CRISPR locus als CRISPR RNAs (crRNAs) hervorgehen und einen so-genannten Protospacer erkennen, d.h. einen Sequenzabschnitt invasiver Nukleinsäuren, welcher Homologie zu einem Spacer aufweist. Im Gegenzug können Viren durch den Erwerb von Mutationen innerhalb ihrer Protospacer-Sequenz einer möglichen DNA Degradation entkommen und somit die erworbene Resistenz ihres Wirtsorganismus umgehen. Interessanterweise führt nicht jede Mutation, die innerhalb des Protospacers auftritt, zur Vermeidung von CRISPR/Cas-Interferenz. Bislang wurden die Auswirkungen von Protospacermutationen auf die CRISPR/Cas-basierende Interferenz vor allem in bakteriellen CRISPR/Cas Systemen untersucht. Um mehr über die Anforderungen für die Interaktion zwischen crRNA und Protospacer während der CRISPR/Cas-Interferenz in Archaea zu lernen, wurde der hyperthermophile Crenarchaeote *Sulfolobus solfataricus* und sein Virus SSV1 als Modellsystem in dieser Studie verwendet. Für die Arbeit wurden verschiedene Protospacermutanten des rekombinanten Virus SSV1 hergestellt und anschließend in mehreren Transfektionsexperimenten untersucht. Die gentechnisch-veränderten Mutanten trugen eine steigende Anzahl an Mutationen an den jeweiligen Enden der Protospacer-Sequenz; d.h. im Ausmaß von sechs bis 18 aufeinander folgenden Mutationen an beiden Protospacerenden. Trotz der eingeführten sechs Mutationen am „down"-Ende des Protospacers, welches dem 3'-Ende transkribierter Spacer RNA entspricht, vermittelte das *S. solfataricus* CRISPR/Cas

System beinahe 100% Immunität gegen das rekombinante Virus. Im Gegensatz dazu, wurde die Resistenz gegen das rekombinante Virus stark vermindert (um 75%), als sechs Mutationen am anderen Ende („up"-Ende) des Protospacers eingeführt wurden. Es konnten immer noch 50% CRISPR/Cas-basierende Immunität nachgewiesen werden, als 15 aufeinander folgende Mutationen am Protospacer „down"-Ende vorhanden waren. Erst die Einführung von 18 Mutationen am „down"-Ende führten zu einem 90 %igen Rückgang der erworbenen Immunität. Ähnlich den erst kürzlich angestellten Beobachtungen in *E. coli*, erwies sich das CRISPR/Cas System als deutlich toleranter gegenüber Mutationen, die am „down"-Ende des Protospacers auftraten. Dies deutet darauf hin, dass CRISPR/Cas-Interferenz in *S. solfataricus* durch die Erkennung einer kurzen Sequenz am „up"-Ende des Protospacers initiiert wird. Verglichen mit dem bakteriellen System, schien die erforderliche Basenkomplementarität zwischen Zielsequenz und crRNA während der CRISPR/Cas-Interferenz jedoch sehr gering zu sein. Dies könnte auf eine Anpassung an sich rasch evolvierende Zielsequenzen hindeuten, was dem Wirtsorganismus eine flexiblere und effektivere Erkennung verwandter/nicht identer, invasiver Nukleinsäuren ermöglichen würde.

Lebenslauf

Persönliche Daten

Name: Julia Steinkellner

Mobil: 0699/12129845

E-Mail: julia_steinkellner@gmx.at

Nationalität: Österreich

Muttersprache: Deutsch

Geburtsdatum: 31.01.1986

Schulausbildung/Studium

Seit Oktober 2005   Diplomstudium Biologie, Universität Wien
                         Dr.-Karl-Lueger-Ring 1
                         1010 Wien
                         - Studienzweig: Ökologie
                         - Schwerpunkt: Mikrobielle Ökologie
                         - Titel der Diplomarbeit:
                           "Required base complementarity between guide
                           RNAs and target sequences in the CRISPR/Cas-
                           based virus defense mechanism of the
                           hyperthermophilic archaeon *Sulfolobus
                           solfataricus";* Betreuerin: Univ.-Prof. Dipl.-Biol. Dr.
                           Christa Schleper

2000-2005   Bundeshandelsakademie, Gmunden

1996-2000   Bundesgymnasium, Gmunden

Berufliche Erfahrung

03/2011-06/2011   Im Rahmen der Diplomarbeit: Tätigkeit als studentische
                         Mitarbeiterin (Studienassistentin), Universität Wien

Sprachkenntnisse

Englisch in Sprache und Schrift

Französisch: Grundkenntnisse

Unterschrift: Ort, Datum: Wien, 2012

Sprachkenntnisse