Philipps Universität Marburg

Characterization of a minimal Type I CRISPR-Cas system found in *Shewanella putrefaciens* CN-32

DISSERTATION

zur
Erlangung des Doktorgrades
der Naturwissenschaften
(Dr. rer. nat.)

Dem Fachbereich Biologie
der Philipps-Universität Marburg
vorgelegt von

Srivatsa Dwarakanath
aus Bengaluru, India

Marburg/Lahn
Juni 2015

Die Untersuchungen der vorliegenden Arbeit wurden von Oktober 2011 bis Juni 2015 unter Betreuung von Dr. Lennart Randau in Marburg am Max-Planck-Institut für terrestrische Mikrobiologie in der Arbeitsgruppe Prokaryotic Small RNA Biology durchgeführt.

Vom Fachbereich Biologie

der Philipps-Universität Marburg als Dissertation

angenommen am: 04.08.2015

Erstgutachter: Dr. Lennart Randau

Zweitgutachter: Prof. Dr. Kai Thormann

Tag der mündlichen Prüfung: 07.09.2015

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel “Characterization of a minimal Type I CRISPR-Cas system found in *Shewanella putrefaciens* CN-32” selbstständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfsmittel bedient habe.

Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Srivatsa Dwarakanath

Marburg, den 24.06.2015

Declaration

I hereby declare that the dissertation entitled “Characterization of a minimal Type I CRISPR-Cas system found in *Shewanella putrefaciens* CN-32” submitted to the Department of Biology, Philipps-Universität Marburg, is the original and independent work carried out by me under the guidance of the PhD supervisors, and the dissertation is not formed previously on the basis of any award of Degree, Diploma or other similar titles.

Srivatsa Dwarakanath

Marburg, on 24.06.2015

Dedicated to my
loving parents

"karmany evadhikaras te
ma phalesu kadachana
ma karma-phala-hetur bhur
ma te sango 'stv akarmani"

"Sri Krishna said: You have a right to perform your prescribed duty, but you are not entitled to the fruits of action. Never consider yourself the cause of the results of your activities, and never be attached to not doing your duty."

(Bhagavad Gita: Chapter Two verse 47)

Table of contents

List of abbreviations

aa	amino acid(s)
Amp	ampicillin
APS	ammonium persulfate
ATP	adenosine triphosphate
bp	basepair(s)
BSA	bovine serum albumin
C-terminal	carboxy-terminal
Cas	CRISPR-associated
Cascade	CRISPR-associated complex for antiviral defense
cDNA	complementary DNA
cpm	counts per minute
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Da	Dalton
DAP	2,6-diaminopimelic acid
DEPC	diethylpyrocarbonate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dsDNA	double-stranded DNA
DTT	dithiothreitol
e.g.	for example (exempli gratia)
EDTA	ethylendiamintetraacetic acid
EMSA	electrophoretic mobility shift assay
EtBr	ethidium bromide
et al.	and other (et alteri)
fw	forward
g	gravitational acceleration
Gua-HCl	guanidinie hydrochloride
h	hour(s)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
i.e.	that is (id est)
IPTG	isopropylβ-D-1-thiogalactopyranoside
Kan	kanamycin
kb	kilo basepairs
kDa	kilo Dalton
l	liter
LB	lysogeny broth
M	Molar (mol/l)
m	meter
min	minute(s)
µ	micro (10^{-6})
n	nano (10^{-9})
N-terminal	amino-terminal
nt	nucleotides
NTP	nucleoside triphosphate
OD_{600}	optical density at 600 nm
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pH	negative logarithm of the hydrogen ion (H^+) concentration
PIPES	piperazine-N,N′-bis(2-ethanesulfonic acid)
qRT-PCR	quantitative real-time PCR
rv	reverse
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal RNA
rpm	rounds per minute
RT	room temperature
s	second(s)
SDS	sodium dodecyl sulfate
Spec	spectinomycin
ssDNA	single-stranded DNA
ssRNA	single-stranded RNA
TAE	Tris-acetate ETDA-buffer
TBE	Tris-borate ETDA-buffer
TEMED	tetramethylethylene diamine
Tris	tris (hydroxymethyl) aminomethane
U	unit (enzyme activity)
V	Volts
v	volume
UV	ultraviolet
W	Watt
%(v/v)	percent by volume
%(w/v)	percent by weight
>	higher than
Δ	deletion

Summary

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated) is an adaptive immune system in prokaryotes that uses small CRISPR RNA (crRNA) to detect and degrade foreign DNA or RNA. CRISPR-Cas systems consist of a CRISPR array and a set of *cas* genes. Individual crRNAs are generated from CRISPR array transcripts and are incorporated into CRISPR ribonucleoprotein (crRNP) surveillance complexes that include different Cas proteins. The Type I CRISPR-Cas systems are identified by the presence of the signature protein Cas3 and utilize a crRNP surveillance complex termed Cascade. The Cas protein content of Cascade differs between the CRISPR-Cas subtypes. A minimal variant of the Type I-F subtype was identified in *Shewanella putrefaciens* CN-32. The genome of *S. putrefaciens* CN-32 contains only five Cas proteins (Cas1, Cas3, Cas6f, Cas1821 and Cas1822) and a single CRISPR array with 81 spacers. This system contains Cas1, an integrase that mediates spacer acquisition, Cas3, the target DNA nuclease and Cas6f, the endonuclease that generates mature crRNAs. Two additional proteins, Cas1821 and Cas1822, show no apparent sequence similarity to any known Cas protein families. A large subunit protein responsible for target recognition in other subtypes is absent. RNA-Seq analyses showed that the CRISPR array is transcribed and that mature crRNAs are generated. *In vivo* interference activity was demonstrated for this minimal system using a plasmid conjugation assay. The interference activity was dependent on the recognition of a dinucleotide GG sequence, termed Protospacer Adjacent Motif (PAM). The deletion of *cas1821* and *cas1822* in *S. putrefaciens* CN-32 abolished the *in vivo* interference activity and resulted in the loss of a stable crRNA pool. A minimal Cascade complex was isolated which contained multiple Cas1821 copies, Cas1822, Cas6f and mature crRNAs. Recombinant Cas1821 formed helical filaments upon RNA binding and the analysis of mutant strains demonstrated that DNA interference depended on conserved arginine residues of Cas1822. Cas1822 and Cas3, which contains a conserved Cas2-like N-terminus, are discussed to compensate for the absence of the large and small subunits present in other Cascade assemblies. These results provide insights into the evolution of reduced CRISPR-Cas systems and demonstrate Cascade functionality without a large subunit.

Zusammenfassung

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-assoziiert) ist ein adaptives Immunsystem in Prokaryoten, das fremde DNA oder RNA mit Hilfe von kleinen CRISPR-RNAs (crRNAs) erkennt und diese anschließend degradiert. Das CRISPR-Cas System besteht dabei aus einem CRISPR-Lokus und einer Gruppe von *cas* Genen. Individuelle crRNA Moleküle werden aus einem langen Transkript des CRISPR-Lokus generiert und im Anschluss in CRISPR-Ribonukleoproteinkomplexen inkorporiert. Diese bestehen aus verschiedenen Cas-Proteinuntereinheiten bestehen und übernehmen die Funktion der zellulären Kontrolle der Immunantwort. Die Typ I CRISPR-Cas Systeme sind gekennzeichnet durch die Anwesenheit des Signatur-Proteins Cas3 und des CRISPR- Ribonukleoproteinkomplexes mit der Bezeichnung Cascade. Dabei variieren die Cas-Proteine in den jeweiligen Cascade-Komplexen der unterschiedlichen Subtypen von Typ I CRISPR-Cas Systemen. Eine minimale Variante eines Typ I-F Systems konnte in *Shewanella putrefaciens* CN-32 identifiziert werden. Das Genom von *S. putrefaciens* CN-32 enthält dabei nur fünf Cas Proteine (Cas1, Cas3, Cas6f, Cas1821 und Cas1822) und einen einzelnen CRISPR-Lokus mit 81 Spacer-Sequenzen. Dieses System beinhaltet Cas1, eine Integrase welche die Integration neuer Spacer vermittelt, Cas3, eine Nuklease welche die Ziel-DNA degradiert und Cas6f, eine Endonuklease die reife crRNAs generiert. Die zwei weiteren Proteine, Cas1821 und Cas1822, zeigen keine Sequenzähnlichkeit zu allen bisher bekannten Cas-Protein Familien. Die große Untereinheit des Cascade-Komplexes, die in anderen Subtypen verantwortlich für die Identifizierung der Ziel-DNA ist, fehlt. RNA-Seq Analysen wiesen die Transkription des CRISPR-Lokus und die Herstellung der reifen crRNAs nach. Interferenz-Aktivität für dieses minimale System konnte *in vivo* mit Hilfe eines Plasmid Konjugations-Assays gezeigt werden. Die Interferenz-Aktivität war dabei abhängig von der Erkennung des Dinukleotids „GG“ in unmittelbarer Nähe zur Ziel-Sequenz. Dieses Motiv wird PAM (Protospacer Adjacent Motif) genannt. Die Deletion von *cas1821* und *cas1822* in *S. putrefaciens* CN-32 führte zur Aufhebung der *in vivo* Interferenz Aktivität und resultierte gleichzeitig in dem Verlust einer stabilen crRNA-Population in der Zelle. Ein minimaler Cascade Komplex konnte isoliert werden, der aus Cas1822, Cas6f, mehreren Kopien von Cas1821 und reifer

crRNA besteht. Rekombinant hergestelltes Cas1821 bildete durch Bindung von RNA lange helikale Filamente aus und die Analyse von Stämmen mit Cas1822 Mutationen demonstrierte die Abhängigkeit der DNA Interferenz-Aktivität von konservierten Arginin-Resten in Cas1822. Es wird diskutiert, ob Cas1822 oder Cas3, welches einen konservierten Cas2-ähnlichen Bereich am N-terminalen Ende enthält, die Funktion der kleinen und großen Untereinheiten ersetzen zu können. Diese Ergebnisse geben Einblicke in die Evolution von reduzierten CRISPR-Cas Systemen und demonstrieren Cascade-Funktionalität ohne die Anwesenheit einer großen und kleinen Untereinheit.

1 Introduction

Prokaryotes, i.e. Archaea and Bacteria, are ubiquitous in nature and thrive in the harshest of environments (1). However, recent estimates show that, depending on the environment, the number of viruses exceeds that of prokaryotes by a factor of 5 to 25 (2,3). In most environments, viruses and prokaryotes are locked in a battle for supremacy, which has led to the evolution of prokaryotic virus resistance mechanisms (4,5). Defense strategies targeting various stages of a viral infection have evolved in prokaryotes. Some defense mechanisms block the viral attachment to the cell surface whereas others prevent the injection of a viral genome into the host. Mechanisms that (i) degrade viral genomes (restriction modification systems (6), Bacteriophage Exclusion (BREX) (7)) or (ii) lead to programmed cell death (toxin-antitoxin systems (8)) after infection have also been identified. Most of these defense strategies in prokaryotes are innate in nature. In contrast, CRISPR-Cas is a recently identified adaptive immune system, which is discussed in detail below.

1.1 CRISPR-Cas systems

The only adaptive immune system that has been characterized in prokaryotes is called CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated) (9-11). The hallmark of this system is the CRISPR locus that contains a series of direct repeats that are separated by sequences that can be of viral origin (12-14). Ishino *et al.* reported the first identification of the CRISPR locus in *Escherichia coli* in 1987 (15). They described five homologous sequences of 29 nt length that were separated by 32 nt long variable sequences. Initially, the biological significance of these unusual sequences was not realized. However, since 2007 it is known that the CRISPR loci and the Cas proteins provide immunity against foreign nucleic acids (16). A brief description about the mechanism, distribution and classification of these diverse systems is provided below.

1.2 Organization and mechanism

CRISPR-Cas systems require two components to mediate an immune response. The first component is the CRISPR array, which contains a series of repeat sequences interspaced by unique DNA fragments called spacers that are

derived from conjugative plasmids or viral genomes (Fig. 1.1) (12-14). The second is a set of *cas* genes that are found in the vicinity of the CRISPR array. These genes encode for proteins that perform essential functions during each step of the immune response (17,18).

Figure 1.1: General organization of a CRISPR-Cas system. The CRISPR locus consists of a series of direct repeats (black) that are interspaced by short unique DNA fragments called spacers (colored). A set of *cas* genes is associated with the locus and is essential for the immune response (modified from (19)).

The immune response of the CRISPR-Cas system is triggered during a viral infection of prokaryotic cells, which can result in high mortality of the cells (Fig. 1.2) (3). However, a small fraction of surviving cells can incorporate a fragment of the viral genome (i.e. the protospacer) into the CRISPR array (20). This stage of the immune response during which a new spacer is incorporated into the CRISPR array is called acquisition. Although this mechanism is not fully understood, it has been found that the Cas proteins, Cas1 and Cas2 and the recognition of a short signature sequence called the Protospacer Adjacent Motif (PAM) in the viral genome are essential for this process (21-23). After acquiring a new spacer, the CRISPR array is transcribed into a long precursor-CRISPR RNA (pre-crRNA) that gets processed to CRISPR RNA (crRNA). This stage of the immune response is called processing and is performed by a Cas endoribonucleases or RNase III. The Cas endoribonuclease e.g. Cas6, recognizes the repeat sequence in the pre-crRNA transcript and cleaves within this sequence to generate short crRNAs that contain a hydroxyl-group at their 5′ ends and a 2′, 3′ cyclic phosphate group at their 3′ ends (24-29). The generated crRNAs contain a spacer sequence that is flanked by partial repeat sequences at both ends and become an integral part of a ribonucleoprotein complexes (crRNP) composed of multiple Cas proteins. During a subsequent viral attack, the crRNA-loaded crRNP complex scans the viral genome for the protospacer sequence. Recognition of the target occurs via Watson-Crick base pairing between the crRNA and the protospacer, which results in the degradation of the viral genome (30-34). This stage of the immune response is termed interference.

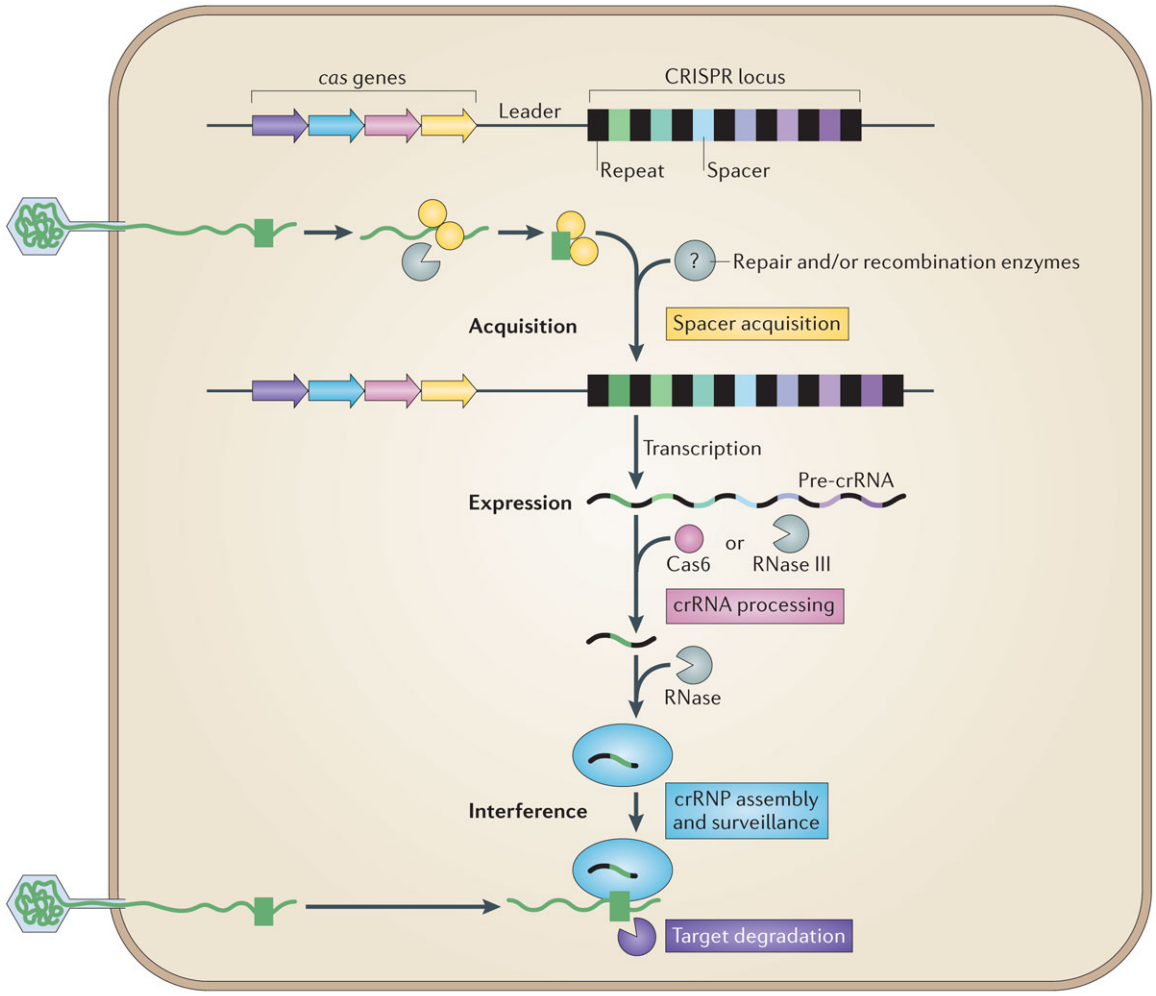


Figure 1.2: Schematic representation of the CRISPR-Cas based immune response. The three stages of CRISPR-Cas based immunity are acquisition, processing and interference (5). During acquisition, Cas1 and Cas2 recognize a viral fragment (protospacer) and integrate it into the extending CRISPR locus. The CRISPR locus is then transcribed and processed into crRNAs that are incorporated into crRNP complexes. These complexes scan and cleave target nucleic acids during a subsequent viral infection. (modified from (33))

1.3 Classification of CRISPR-Cas systems

The current classification scheme, which is based on phylogenetic analyses of Cas proteins, places the CRISPR-Cas systems of bacteria and archaea into three main types and eleven subtypes (Fig. 1.3). The three types of CRISPR-Cas systems are defined based on the presence of the conserved signature proteins Cas3 (Type I), Cas9 (Type II) and Cas10 (Type III), respectively (17,35). Only Cas1 and Cas2, which have been found to be essential for acquisition of new spacers, are conserved in all CRISPR-Cas systems (21,22,36,37).

The three types of CRISPR-Cas systems differ mainly in their crRNP complex composition and assembly and the mechanism involved in target recognition and cleavage. Type I crRNP complexes are termed Cascade (CRISPR-associated complex for antiviral defense) and are constituted by 4 to 7 Cas proteins (30,38,39). Similarly, the Type III crRNP complexes are composed of multiple Cas proteins (32,40). However, unlike the Type I systems that target DNA in a PAM-dependent manner, the Type III systems are PAM-independent and target both RNA and DNA (41-44). In contrast, the Type II crRNP complex contains a single large nuclease, termed Cas9, that performs the dual role of PAM-dependent target DNA recognition and degradation (45,46). Another unique characteristic of the Type II systems is the utilization of RNase III for the generation of mature crRNAs (29). A similar function is generally performed by the Cas6 endoribonuclease in Type I and Type III systems.

1.4 Distribution of CRISPR-Cas systems

CRISPR-Cas systems are a widespread in prokaryotes and are found in approximately 40% of all sequenced bacterial genomes and nearly all of the archaeal genomes (47). Recent reports have also identified CRISPR-Cas systems in genomes of bacteriophages which can help in the evasion of the immune response of the host (48). The Type I and Type III systems are found in both, bacterial and archaeal lineages. However, the Type I systems are widespread among bacteria and the Type III systems are more common in archaeal genomes. The Type I subtypes I-C, I-E and I-F are usually present in bacteria whereas the Type I subtypes I-A, I-B and I-D are more prevalent in archaea. The Type III systems are the only CRISPR-Cas system that can target RNA and hence usually occur together with subtypes of Type I CRISPR-Cas system. Thus, by combining two systems that target different nucleic acids, the versatility of the CRISPR-Cas based immune response is increased. The Type II systems are not widespread and are exclusive found in bacteria. (17,35,49). In the following sections, the different types of CRISPR-Cas systems are briefly described.

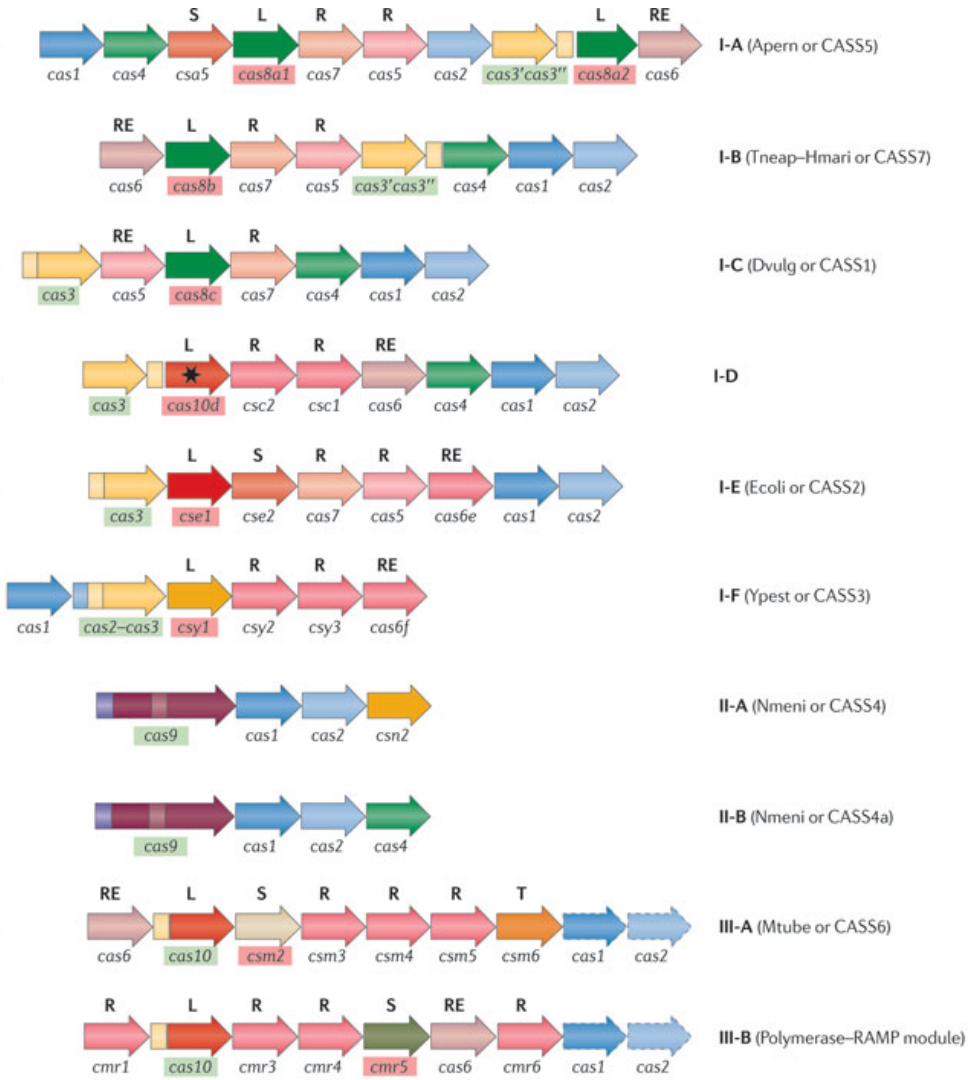


Figure 1.3: Current classification of the CRISPR-Cas systems. CRISPR-Cas systems are classified into three main types based on the presence of a signature gene and into different subtypes based on phylogenetic analyses of the *cas* genes. The signature genes for each type are highlighted in green boxes. The letters above the genes indicate major categories of Cas proteins: large subunits (L), small subunits (S), crRNP backbone subunits (R), RNases involved in crRNA processing (RE) and transcriptional regulators (T) (modified from (17)).

1.5 Type I CRISPR-Cas systems

CRISPR-Cas systems are classified as Type I based on the presence of the conserved signature protein Cas3 and are further divided into six subtypes (I-A to I-F) (17,35). All Type I crRNP complexes are termed Cascade and are composed of a single crRNA and multiple Cas proteins. In most Type I subtypes, the endoribonuclease Cas6 cleaves within the repeat sequences of the pre-crRNA to generate crRNAs. Cleavage activity of the enzyme is dependent on the recognition

of the sequence and structure of the repeat and results in crRNAs that carry an 8 nt repeat tag at their 5′ end (27,50,51). Due to the catalytic activity of the enzyme, all crRNAs carry a hydroxyl group at the 5′ end and a 2′-3′ cyclic phosphate at the 3′ end (31,52). In a few subtypes, further trimming at the 3′ end of the crRNA is observed (25,34,53). The Cascade complex is assembled along this mature crRNA and recognizes a DNA target based on the presence of a unique PAM sequence and the base paring between the crRNA and its complementary DNA strand (Fig. 1.4) (31,38,39,54). Binding of the crRNA to its complementary sequence on the target DNA displaces the non-complementary strand and results in the formation of an R-loop structure. This triggers the recruitment of Cas3, a single stranded DNA nuclease and ATP-dependent helicase, by the Cascade complex to degrade the dsDNA target (55). The metal-dependent HD nuclease domain of Cas3 cleaves the non-target strand while its helicase domain unwinds the dsDNA in 3′ to 5′ direction. Subsequently, a second Cas3 subunit is recruited to degrade the target strand. Cascade complexes from different Type I subtypes have been studied (19) and some of them are briefly described below.

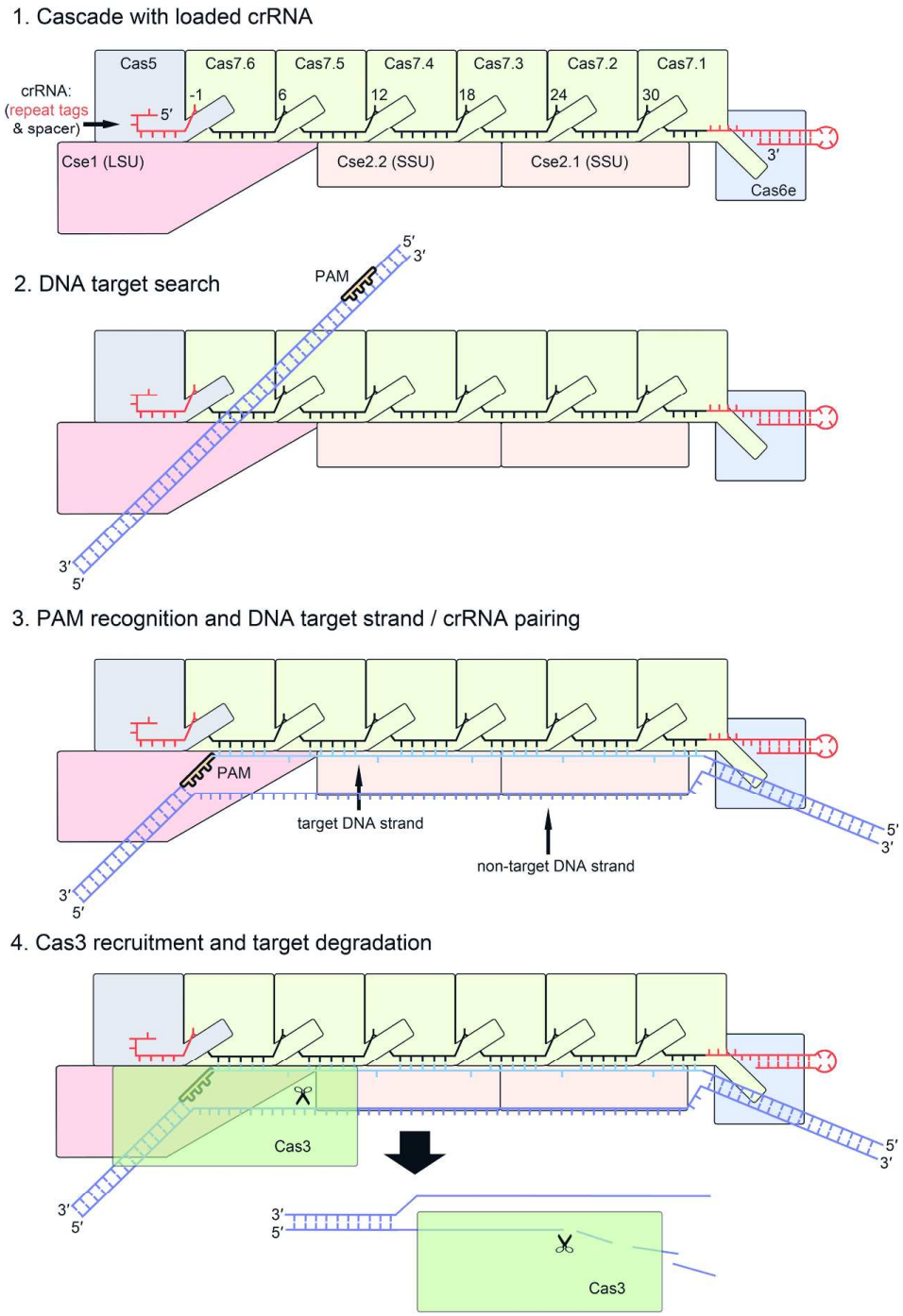


Figure 1.4: Representation of the mechanism involved in target DNA interference by Cascade. The Cascade complex assembles along the crRNA and constantly scans for potential targets. Identification of a DNA target is dependent on the recognition of the PAM sequence and results in the formation of an R-loop structure due to base pairing between crRNA and the target strand. Formation of a stable R-loop triggers the recruitment of Cas3, which cleaves the DNA in the 3′ to 5′ direction (modified from (19)).

1.5.1 Subtype I-E

The best studied Type I-E system is found in *Escherichia coli* and its crRNP complex termed Cascade consists of a 61 nt crRNA bound by five Cas proteins. The mature 61 nt crRNAs contains a 32 nt long spacer flanked by 8 nt and 21 nt long repeat sequences at their 5′ and 3′ ends, respectively. They are generated via specific Cas6e cleavage of the repeat sequences of a pre-crRNA transcript (31,56). After cleavage, Cas6e remains associated with the 3′ hairpin of the mature crRNA and initiates the assembly of Cascade around it. The 405 kDa Cascade complex has a sea horse-like architecture and is formed by Cas5, Cas6e, Cas7 and two subtype specific proteins Cse1 and Cse2, with an uneven protein stoichiometry of $(Cas7)_6$-$(Cse2)_2$-$(Cse1)_1$-$(Cas5)_1$-$(Cas6e)_1$ (38,39,51). Six copies of Cas7, which contains a groove to bind and protect the crRNA, form the helical backbone of the Cascade. Additionally, Cas6e and Cas5 bind and protect the 3′ end and the 5′ end of the crRNA, respectively. Cse1, the large subunit protein of the Cascade complex, is involved in PAM recognition and recruits Cas3 after target recognition, whereas the small subunit protein Cse2 functions in stabilizing the R-loop by binding to the displaced strand (55).

1.5.2 Subtype I-F

Many similarities are observed between the subtype I-F and the subtype I-E systems (57). Similar to Cas6e, the endoribonuclease Cas6f (also known as Csy4) cleaves pre-crRNA in a repeat specific manner and remains tightly associated with the 3′ end of the crRNA product (26,58). The generated crRNAs are 60 nt long with characteristic 8 nt 5′ repeat handles and a 20 nt 3′ repeat handle containing a stable stem loop structure. The tight association of Cas6f to the 3′ handle of the crRNA product is essential for RNA protection and I-F Cascade assembly (59,60). The crescent-shaped Type I-F Cascade from *Pseudomonas aeruginosa* has a mass of 350 kDa and is composed of four Cas proteins with a stoichiometry of $(Csy1)_1$-$(Cas5)_1$-$(Cas7)_6$-$(Cas6f)_1$ (61). The Cascade backbone is formed by six copies of a Cas7 family protein (known as Csy3) that binds to the crRNA. The 5′ and 3′ termini are protected by Cas5 (known as Cys2) and Cas6f, respectively (61). Csy1 is proposed to perform the functions of target recognition and DNA binding (35,62).

Cas3, which is involved in target DNA degradation, contains a HD nuclease domain and an ATP-dependent helicase domain and at the N-terminal end is fused with a Cas2-like domain (62). This Cas2-Cas3 fusion protein was shown to be involved in target DNA degradation and spacer acquisition (37,62).

1.5.3 Other Type I subtypes (I-A, I-B, I-C and I-D)

In addition to subtypes I-E and I-F, four other Type I subtypes (I-A, I-B, I-C and I-D) are found in prokaryotes. As observed for subtypes I-E and I-F, the Cas6 endoribonuclease is responsible for crRNA generation in all of these subtypes except for subtype I-C (51). This subtype does not encode for a Cas6 enzyme and a homolog of Cas5 is responsible for crRNA generation (52,63). In most of these subtypes, the mature crRNAs contain an 8 nt repeat tag at the 5′ end and a 3′ repeat tag that is gradually trimmed. Similar to subtypes I-E and I-F, these subtypes form a multi-Cas protein Cascade complex composed of Cas5, Cas7, Cas6 and the large and the small subunit proteins (64). In all of these subtypes, Cas7 is proposed to bind to the crRNA as a multimer to form the backbone of the Cascade complex (64).

1.5.4 Structure and assembly of Cascade

Low-resolution EM structures of Cascade complexes are available for most Type I subtypes (64,65). Although the Cas protein composition of Cascade differs between the Type I subtypes, the general crescent-shape was found to be conserved. Recently, crystal structures of the Type I-E Cascade were obtained, which provide insights into its architecture and assembly (38,39,66). The structures verified that Type I-E Cascade is composed of Cas5, Cas6, Cas7, the two subtype specific proteins Cse1 and Cse2 and a 61 nt long crRNA (31). Cas6 endoribonuclease remains tightly associated with the 3′ end of the crRNA, and is thought to initiate the assembly of Cascade (56). Oligomerization of Cas7 along the crRNA and the binding of Cas5 to the crRNA's 5′ repeat tag complete the assembly of the basic Cascade scaffold. The comparison of the crystal structures of Cas5 and Cas7 suggests that the crRNA binding mechanism is similar for these proteins. Superimposition of the structures reveled similarities in a flexible thumb domain and a palm domain of the two proteins (38,39,66). In addition to the palm and thumb

domains, a finger domain was identified only in Cas7. The binding of Cas5 to the 5′ end of the crRNA results in the formation of a hook-like structure in the repeat tag. The thumb domain of Cas5 folds over the last nucleotide in the repeat tag and interacts with the finger domain of the adjacent Cas7 subunit, creating a kink at this point of the crRNA backbone (38,39,66). A similar mechanism is employed during the formation of the Cas7 hexamer along the crRNA. The thumb domain of one Cas7 subunit interacts with both, the palm and the finger domains of the adjacent Cas7 subunit, resulting in a kink at every sixth nucleotide of the crRNA. Due to these interactions, the position of the thumb domain for all Cas7 subunits is comparable except for the Cas7 subunit that interacts with Cas6. This Cas7 subunit, upon interaction with Cas6, rotates its thumb domain outwards (38,39,66). The intertwined Cas protein arrangement separates the crRNA into six segments, in which one nucleotide is splayed and the following five nucleotides are accessible for coordination with target DNA (Fig. 1.5). Upon target DNA hybridization, the thumb domain of Cas7 also causes a kink in the target DNA backbone, which destabilizes the dsDNA and initiates the R-loop formation (38,39,66). The assembly of an active Cascade is completed by binding of the small subunit (Cse2) dimer and the large subunit (Cse1) to the Cascade backbone at the 3′ and 5′ ends of the crRNA, respectively. Both proteins form a basic groove along the belly of Cascade that is necessary for the binding of the non-target strand after R-loop formation (38,39,66). Additionally, Cse1 undergoes structural changes upon target DNA binding which results in the recruitment of the Cas3 nuclease (67). This arrangement of the Type I-E Cas proteins along the crRNA results in an active Cascade that is constantly screening for potential targets. Crystal structures of Cas5 and Cas7 proteins are available for other Type I subtypes. The comparison of these structures revealed a similar domain organization to *E. coli* Cas5 and Cas7 from Type I-E (64). These observations suggest that the mechanism involved in the basic Cascade backbone assembly might be conserved for all Type I subtypes.

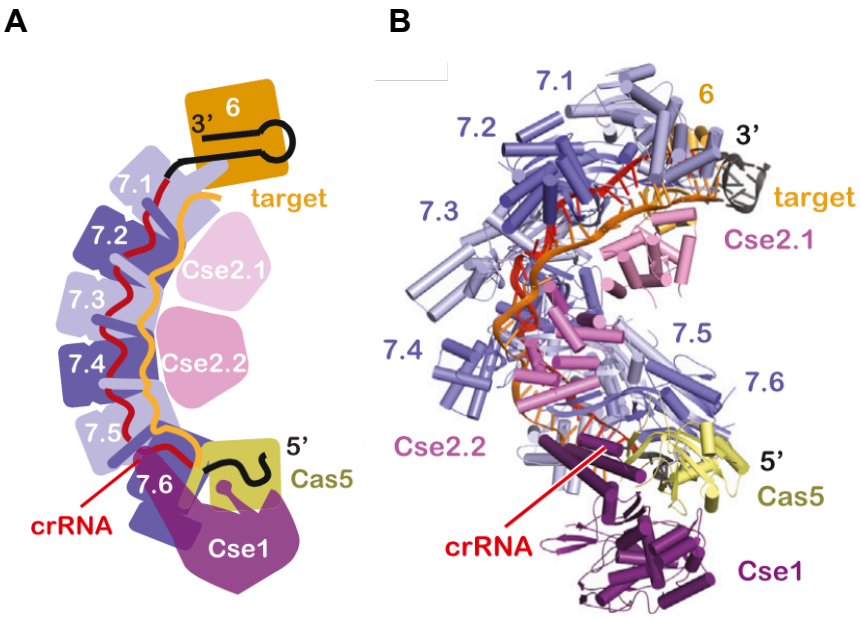


Figure 1.5: Structure of the Type I-E Cascade. Depicted are the (A) Cartoon structure and the (B) crystal structure of type I-E Cascade (68). The complex contains 11 protein subunits with the following stoichiometry $(Cse1)_1$-$(Cse2)_2$-$(Cas5)_1$-$(Cas7)_6$-$(Cas6)_1$ and a 61 nt long crRNA. Binding of Cas5 and Cas6 to the crRNA protects the 5′ end and 3′ end, respectively. The crescent shaped structure of Cascade is a result of the kinking of the crRNA by the thumb domains of Cas5 and Cas7. The two Cse2 subunits bind at the 3′ end of the crRNA via protein:protein interactions with Cas7. The large subunit Cse1 is positioned at the 5′ end of the crRNA and interacts with Cas5, Cas7 and Cse2 (38,39,66) (modified from (68)).

1.6 Type II CRISPR-Cas systems

Type II CRISPR-Cas systems are exclusively found in bacteria and employ the stand-alone signature protein Cas9 to perform DNA target recognition and degradation (46,69). One of the interesting features of the Type II systems is the absence of the endoribonuclease Cas6. Instead, the generation of crRNAs in Type II systems involves Cas9, the bacterial host enzyme RNase III and a tracrRNA (trans-activating crRNA). Biogenesis of crRNA begins with the formation of a RNA duplex between the anti-repeat sequence of the tracrRNA and the repeats of the pre-crRNA. This duplex, upon stabilization by Cas9, is recognized and cleaved by RNase III (28). The obtained crRNAs contain the spacer sequence at the 5′ terminus and the repeat sequence at the 3′ terminus and remain associated with the tracrRNA (28,29).

Although variation in the sequences of Cas9 among Type II containing organisms were observed, three conserved features were detected: (i) the HNH and

the RuvC nuclease domains, (ii) an arginine rich motif and (iii) a general globular architecture (46,70,71). Cas9 utilizes the tracrRNA:crRNA duplex to recognize and bind its target in a PAM dependent manner during interference (Fig. 1.6). On target binding, the HNH nuclease domain cleaves the DNA strand that is complementary to the spacer region of the crRNA, while the RuvC nuclease domain cleaves the non-target DNA strand (72,73). Target recognition is dependent on base paring between the crRNA and the target and the presence of a PAM on the DNA (72,73). Although, few mutations in the protospacer sequence are tolerated by this system, mutations in the PAM sequence result in the target evading CRISPR-Cas mediated immunity. In *Streptococcus pyogenes*, it was shown that two conserved arginine residues R1333 and R1335 are essential for the recognition of the PAM sequence 'NGG' (45). Mutations of these residues abolished the targeting ability of the Type II CRISPR-Cas system.

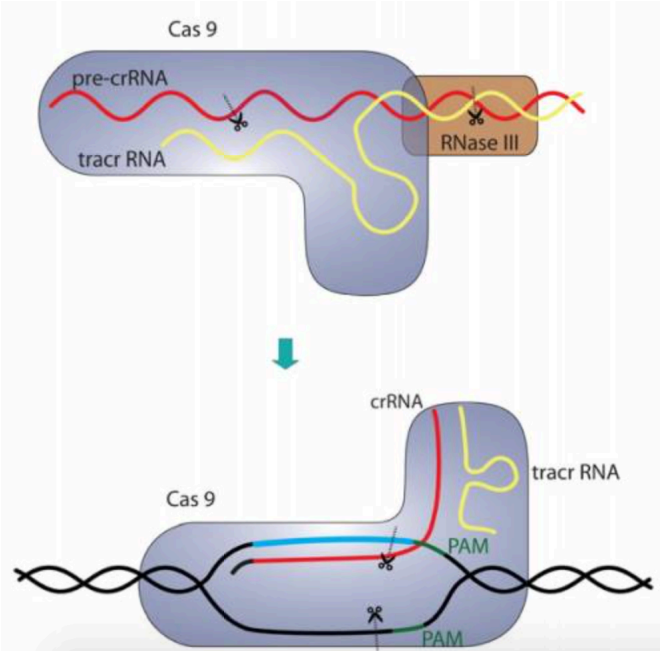


Figure 1.6: Interference mechanism of a Type II CRISPR-Cas system. The system contains a stand-alone Cas9 enzyme that performs the functions of DNA target binding and degradation. The crRNAs are generated with the help of a tracrRNA and the host enzyme RNase III. The crRNA:tracrRNA duplex guides Cas9 to recognize the DNA target in a PAM-dependent manner. Two separate nuclease domains of the protein cleave the DNA target (modified from (74)).

1.7 Type III CRISPR-Cas systems

The Type III CRISPR-Cas systems are characterized by the presence of the signature protein Cas10, but share significant similarities with Type I systems (17,35). The crRNAs in Type III systems are also generated by the Cas6 endoribonuclease and form an integral part of a multi-Cas protein crRNP complex that is necessary for the recognition and degradation of target nucleic acids. The mature crRNAs contain the standard 8 nt 5′ repeat tag. However, trimming of the crRNA at the 3′ ends is observed and results in two crRNA variants that differ by 6 nt in length (75,76). The Type I and Type III systems are clearly distinguished by their mode of recognition and selection of the targets. Type I systems exclusively target dsDNA and rely on the PAM sequences to distinguish between self and non-self targets. In contrast, Type III systems target both ssRNA and dsDNA (40,76-78). It was recently shown that transcription across dsDNA is necessary for cleavage of the target DNA by Cas10 of the Type III Csm complex (78). This activity was PAM-independent and instead relied on the mismatches between the 5′ repeat tag of the crRNA and the 3′ region of the target DNA to prevent autoimmunity. Csm3, the backbone protein of the Csm complex is responsible for ssRNA cleavage, for which a self-targeting prevention mechanism was not detected (44).

1.8 The CRISPR-Cas system of *Shewanella putrefaciens* CN-32

Shewanella putrefaciens is a gram-negative bacterium that belongs to the class of gammaproteobacteria. The rod shaped organism contains a polar flagellum for motility and is usually found in marine environments. *S. putrefaciens* is not considered to be a typical human pathogen. However, cases of benign skin and soft tissue infections, peritonitis, biliary tract infection and bacteremia in immune-compromised patients have been linked to the presence of *S. putrefaciens* (79). In the absence of oxygen, *S. putrefaciens* has been reported to use a wide variety of terminal electron acceptors that include nitrate, nitrite, fumarate, thiosulphate, elemental sulphur, trimethylamine N-oxide and, the particulate metal oxides of Mn(IV) and Fe(III) (80). Furthermore, the strain *S. oneidensis* MR-1 is used as a model organism for anaerobic respiration and electron transport linked metal reduction (81,82).

S. putrefaciens is a facultative anaerobe and can be cultured both aerobically and anaerobically, with an optimal growth temperature of 30°C. Few species of *Shewanella* have been reported to grow at temperatures ranging from 3°C to 35°C (83). Additionally, genetic manipulation tools and gene expression systems for *S. putrefaciens* are available (84).

The genome of *Shewanella putrefaciens* CN-32 contains a single CRISPR locus. The locus contains 81 spacers and pseudo-palindromic repeats that are 28 nt long. Five *cas* genes were found associated with this CRISPR locus. This system was classified as a variant of the Type I-F system and was first described in *Photobacterium profundum* SS9 (PBPRB1995 - PBPRB1991) (Fig. 1.7) (35). The system contains genes that encode for Cas1, the integrase that mediates spacer acquisition, Cas3, the target DNA nuclease and Cas6f, the endonuclease that generates mature crRNAs. Two additional genes showed no detectable sequence similarity to known *cas* gene families and a gene encoding for the large subunit could not be detected (35). The large number of spacers in the CRISPR locus and the conservation of the gene operon in several proteobacterial species suggest that this variant system is active.

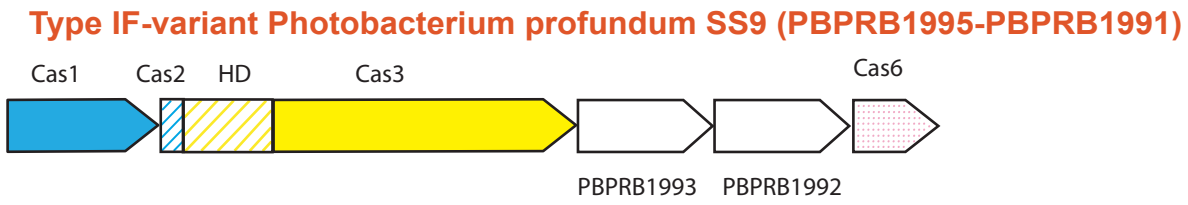


Figure 1.7: Genome organization of the type I-F variant system identified in *Photobacterium profundum* SS9. The system contains five Cas proteins including Cas1, Cas3 and Cas6. The two additional proteins share no obvious homology to known Cas protein families (modified from (35)).

1.9 Aim of this work

A variant Type I-F CRISPR-Cas system was identified In *Shewanella putrefaciens* CN-32. The system consists of a CRISPR locus containing 81 spacers and five associated *cas* genes. Three of these genes, *Sputcn32_1819*, *Sputcn32_1820* and *Sputcn32_1823* encode proteins that belong to Cas1, Cas3 and Cas6 protein families, respectively. The two additional genes *cas1821* (*Sputcn32_1821*) and *cas1822* (*Sputcn32_1822*) encode proteins that share no detectable homology with known Cas protein families. Additionally, a large Cascade subunit protein is absent in this system. Nevertheless, this minimal *cas* gene operon organization is conserved in several proteobacterial species. One aim of this study was to verify that this system is active *in vivo*. Subsequently, the interference activity of this system was investigated with a focus on the functional characterization of the two unclassified Cas proteins, Cas1821 and Cas1822.

The Cascade crRNP complexes responsible for DNA target recognition in Type I systems are composed of a crRNA and 4 to 7 Cas proteins. The second aim of this study was the analysis of a minimal Cascade of *S. putrefaciens* CN-32 that is proposed to contain only three proteins: Cas1821, Cas1822 and Cas6. Finally, the elucidation of the minimal Cascade architecture can be used to define general assembly and functionality concepts for Type I Cascade complexes.

2 Results

2.1 Identification of a minimal CRISPR-Cas subtype in *Shewanella putrefaciens* CN-32

A minimal Type I-F variant CRISPR-Cas subtype was first described for *Photobacterium profundum* SS9 (PBPRB1995-PBPRB1991) (35). A similar system was found in *Shewanella putrefaciens* CN-32 (Fig. 2.1) and it harbors a single CRISPR array with 81 spacers and only five *cas* genes: *cas1* (*Sputcn32_1819*), *cas3* (*Sputcn32_1820*), *cas6f* (*Sputcn32_1823*), *cas1821* (*Sputcn32_1821*) and *cas1822* (*Sputcn32_1822*). The comparison of this system with the subtype I-F system present in the *Shewanella* sp. strain W3-18-1 illustrates the loss of the large subunit and the diversification of the Csy1 – Csy3 proteins (Fig. 2.1). These CRISPR-Cas systems share 100% amino acid identity for the flanking Cas1 and a CRISPR-array adjacent DinG-helicase. In addition, the Cas6f endonuclease and the N-terminal portion of the Cas3 nuclease show significant sequence identity. However, the other Cas proteins that constitute the subtype I-F Cascade (Cas5 (Csy2), Cas7 (Csy3) and the large subunit (Csy1)) diverged drastically from the two Cas proteins Cas1821 and Cas1822. A large subunit candidate is absent in this minimal variant. Homology searches of Cas1821 and Cas1822 revealed the presence of similar I-F variant systems in few other beta- and gammaproteobacteria. These include the human pathogens *Legionella pneumophila* (strain 2300/99 Alcoy) and *Vibrio cholerae* (strain TM 11079-80).

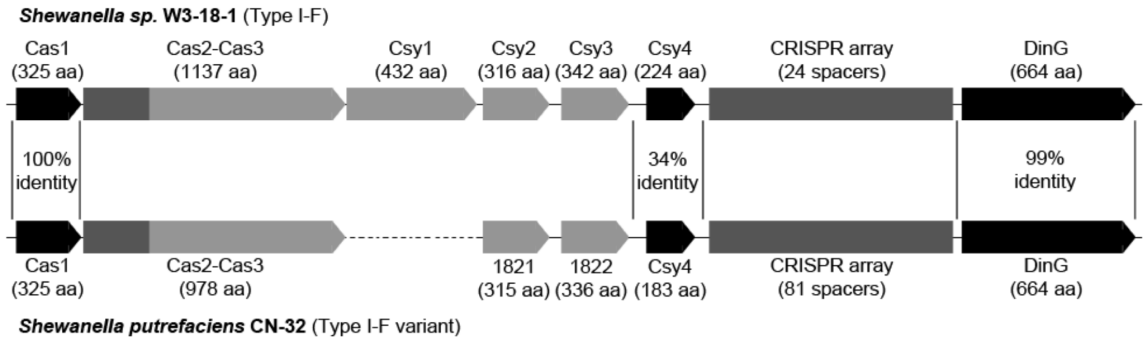


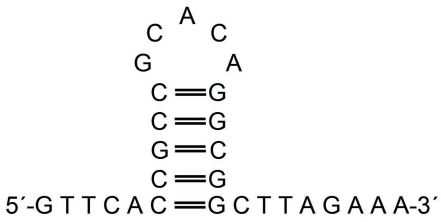
Figure 2.1: Comparison of CRISPR-Cas systems in *S. putrefaciens* strains. Schematic diagrams show the single CRISPR-Cas systems from *Shewanella* sp. W3-18-1 (Sputw3181_2191 - Sputw3181_2185, Type I-F) and *S. putrefaciens* CN-32 (Sputcn32_1819 - Sputcn32_1824, Type I-F variant). Cas1, DinG and the N-terminal 94 amino acids of the Cas3 proteins share 100% aa identity. The C-terminal Cas3 portion and the putative Cascade components Cas1821 and Cas1822 do not reveal significant protein similarity.

2.2 *In silico* characterization of the minimal Type I-F variant system

2.2.1 Characterization of Sputcn32_1823

The *S. putrefaciens* CN-32 CRISPR array contains partially palindromic repeats of 28 nt length that are separated by ~32 nt long spacers (only 3 of the 81 spacers are 33 nt in length). The minimum free energy structure of the repeat as predicted by RNAplot (85) revealed a five base-pair stem-loop formed by the palindromic region (Fig. 2.2A). A similar pseudo palindromic repeat was observed for the Type I-F system found in *Pseudomonas aeruginosa* (26). It was shown that Cas6f generates crRNAs by recognizing the stem-loop of the repeat and cleaving at the base of the stem loop at the 3′ end (26). BLAST analyses identified Sputcn32_1823 as a member of the Cas6 protein family with 37% sequence identity to Cas6f from *P. aeruginosa* (Fig. 2.2B). Additionally sequence alignment of the two proteins identified a conserved histidine residue at position 29, which was reported to be essential for the endoribonucleolytic activity of Cas6f from *P.aeruginosa* (Fig. 2.2B) (26). Therefore, Sputcn32_1823 (Cas6f) was predicted to be the enzyme responsible for the biogenesis of crRNA in *S. putrefaciens* CN-32.

A



B

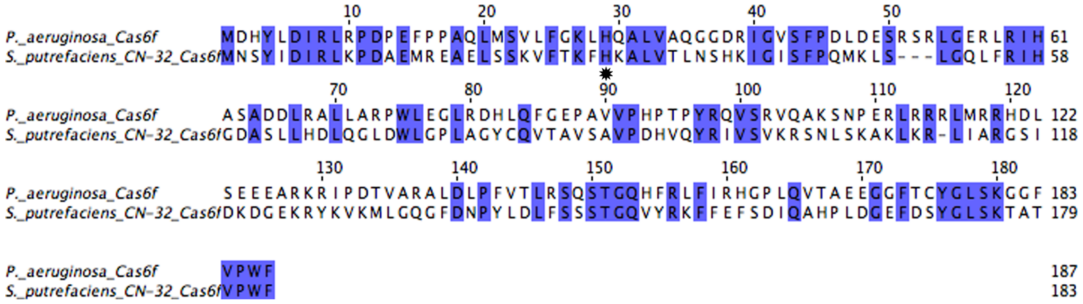


Figure 2.2: Sputcn32_1823 is a Cas6f homolog. (A) Predicted minimum free energy structure of the repeat sequence of the *S. putrefaciens* CN-32 CRISPR array. (B) Sequence alignment of Cas6f from *S. putrefaciens* and *P. aeruginosa*. The identical residues are colored in blue and the catalytic histidine residue is marked by an asterisk.

2.2.2 Characterization of Sputcn32_1820

A Type I-F system, similar to the system present in *Shewanella* sp. W3-18-1 was identified in *Pectobacterium atrosepticum*. The Cas3 protein of this system was reported to contain an HD nuclease domain, a helicase domain and a Cas2 like domain at the N-terminus of the protein (62). Although the overall sequence similarity between Sputcn32_1820 and this protein was very low, high identity was observed at the N-terminal region. The tertiary structure prediction of Cas3 protein from *S. putrefaciens* CN-32 using Phyre2 and I-Tasser identified a Cas2-like fold for this region (86,87) (Fig. 2.3A). Additionally, the domain prediction tool Prosite identified a HD nuclease domain and a helicase domain (belonging to the Superfamily 2 helicases) at the C-terminal end of the protein (88). Hence, as observed in Type I-F systems, Sputcn32_1820 (Cas3) is a fusion protein that contains both Cas2 and Cas3 domains (Fig. 2.3B). The HD nuclease domain of Cas3 from *Streptococcus thermophilus* was shown to cleave ssDNA in the presence of transition metals ions like Mg^{2+} or Mn^{2+} whereas the helicase domain unwound dsDNA in the 3′ to 5′ direction by utilizing ATP (89). Therefore, *S. putrefaciens* CN-32 Cas3 is predicted to have a similar function and is proposed to degrade foreign DNA during interference.

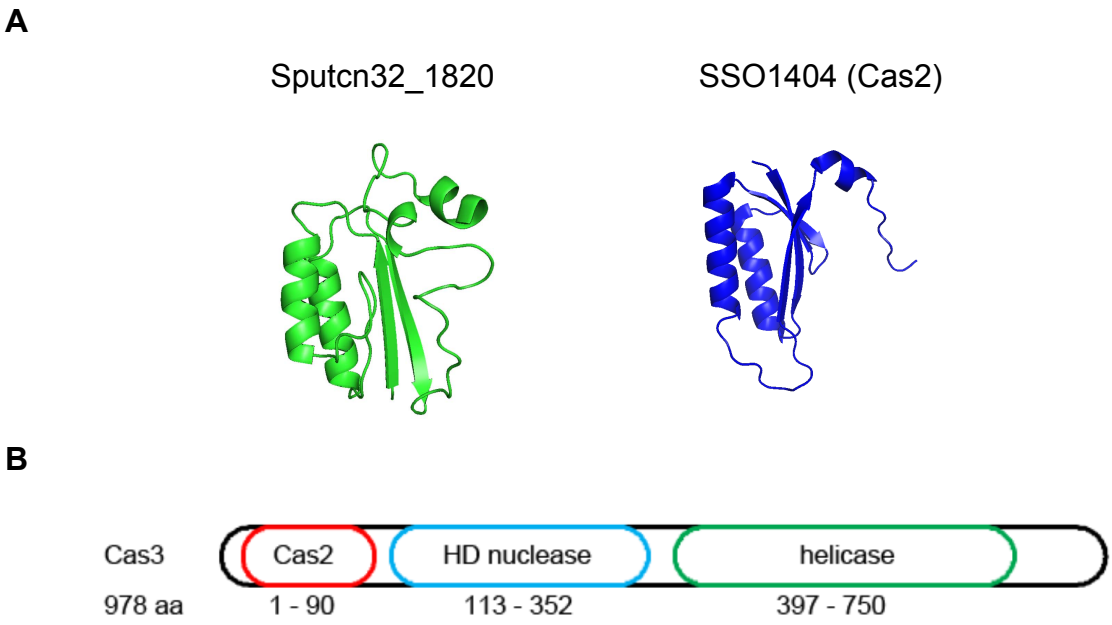


Figure 2.3: Organization of Cas3 from *S. putrefaciens* CN-32. (A) 3D predicted structure of the N-terminal 100 aa of Cas3. The predicted structure was similar to Cas2 from *Sulfolobus solfataricus* (SSO1404, in blue) with which an 11% aa identity was shared. (B) Schematic representation of the predicted domains of Cas3.

2.2.3 Characterization of Sputcn32_1821 and Sputcn32_1822

The variant Type I-F system harbors two hypothetical genes in the *cas* operon. The two proteins Cas1821 and Cas1822 appear to replace the Type I-F Cascade forming Cas proteins Csy1 (large subunit), Csy5 (Cas5) and Csy3 (Cas7). Makarova *et al.* classified these two proteins as Cas7-like and Cas5-like based on their length and the position of the respective genes in the operon (35). However, psi-BLAST analyses could not identify similarities between these proteins and the Cas7 and Cas5 protein families. Accordingly, phylogenetic analyses with Cas7 and Cas5 protein families grouped these proteins and its homologs into a separate cluster. Interestingly, Cas1821 shared an evolutionary relationship with the proteins from the Type I-F system, whereas Cas1822 was related to the proteins from the Type I-C system (Appendix 1 and 2). Further bioinformatic analyses using the domain prediction tool, Prosite and the tertiary structure prediction tools, Phyre2 and I-Tasser, could not identify conserved domains and folds in both Cas1821 and Cas1822. Therefore, the biochemical functions of these unique Cas proteins could not be predicted using bioinformatic tools and analyses.

The variant Type I-F system is characterized by the absence of the large subunit protein. This variant system was identified in nine mesophilic beta- and

gamma-proteobacteria: *S. putrefaciens* (strains CN-32 & 200), *Oligella ureolytica* (strain DSM 18253), *Pseudoalteromonas tunicata* (strain D2), *L. pneumophila* (strain 2300/99 Alcoy), *Oxalobacter formigenes* (strain OXCC13), *P. profundum* (strain SS9), *V. cholerae* (strain TM 11079-80) and *Methylophaga nitratireducenticrescens*. Phylogenetic profiling of these genomes was performed using the Phylogenetic profiler tool offered by the JGI platform, to identify proteins that could substitute for the large subunit (90). *S. oneidensis* MR-1 genome, which lacks a CRISPR-Cas system, was included in the analyses as an outgroup. Therefore, genes in *S. putrefaciens* CN-32 that had homologs in *S. oneidensis* MR-1 were not considered during the analyses. Out of the 596 genes that were not shared between the two genomes, the 5 *cas* genes and 16S and 23S ribosomal RNA modification enzymes were identified to be conserved with > 95% identity (Table 2.1) in the nine Type I-F variant system containing genomes. On repeating the analyses with *Shewanella* sp. W3-18-1 genome, which encodes the subtype I-F CRISPR-Cas system, as an outgroup, only *cas1821*, *cas1822* and *cas6* were conserved in all Type I-F variant containing organisms. However, a potential candidate responsible for the PAM recognition function could not be identified. Additionally, in *S. putrefaciens* CN-32, downstream of the CRISPR array an ATP-dependent DNA helicase DinG (Sputcn32_1824) was observed. Makarova *et al*. predicted this protein to be a functional component of the Type I-U system (35). Therefore, DinG could be a potential enzyme that took over the role of the large subunit in the variant Type I-F system.

Table 2.1: List of conserved proteins obtained from phylogenetic profiling of the nine genomes.

Locus Tag	Annotation	Percentage identity
Sputcn32_0834	23S rRNA pseudouridine746 synthase RluA	99.6%
Sputcn32_0922	RNA pseudouridylate synthase family protein	97.1%
Sputcn32_0952	Transcriptional regulator LysR family	100%
Sputcn32_1000	16S rRNA pseudouridine synthase RluA family	99.7%
Sputcn32_1019	16S rRNA pseudouridine synthase RsuA family	99.6%
Sputcn32_1022	23S rRNA pseudouridine1911/1915/1917 synthase RluD	100.0%
Sputcn32_1339	tRNA pseudouridine65 synthase TruC	98.9%
Sputcn32_1574	23S rRNA pseudouridine955/2504/2580 synthase RluC	99.7%
Sputcn32_1819	Cas1	98.8%
Sputcn32_1820	Cas3-HD	99.6%
Sputcn32_1821	Conserved hypothetical protein	100.0%

Sputcn32_1822	Conserved hypothetical protein	100.0%
Sputcn32_1823	Csy4	100.0%
Sputcn32_2013	Alanine dehydrogenase Ald	100.0%
Sputcn32_2231	23S rRNA pseudouridine2457 synthase RluE	97.4%
Sputcn32_2398	23S rRNA pseudouridine2605 synthase RluB	100.0%
Sputcn32_2441	tRNA pseudouridine38-40 synthase TruA	100.0%
Sputcn32_2488	23S rRNA pseudouridine synthase RluA family	100.0%
Sputcn32_2832	tRNA pseudouridine55 synthase TruB	99.7%
Sputcn32_2938	16S rRNA pseudouridine synthase RluA family	98.7%

2.3 Characterization of the crRNAs of the Type I-F variant CRISPR-Cas system

2.3.1 The CRISPR array in *S. putrefaciens* CN-32 is transcribed and processed *in vivo*

Activity of the variant Type I-F system has not been observed. One of the hallmarks of an active CRISPR-Cas system is the presence of crRNAs. To identify crRNAs small RNA molecules were isolated from *S. putrefaciens* CN-32 and subjected to RNA-Seq analyses. Sequencing of crRNAs required T4 PNK treatment of the isolated small RNA molecules during Illumina Hiseq2000 library preparation, which indicates the presence of crRNA 5′ hydroxyl termini. More than 11 million sequence reads were mapped to the *S. putrefaciens* CN-32 reference genome. The presence of all 81 crRNA transcripts could be verified and precursor transcript processing generated the expected 8 nt repeat tags at the crRNAs' 5′-ends (5′-CUUAGAAA-3′) and a 20 nt repeat tag at the 3′-termini (Fig. 2.4 Insert). 73,377 reads mapped to the CRISPR locus, indicating a lower crRNA abundance in comparison to very similar RNA-Seq studies for type I-A and type I-B systems (25,34). A minimum free-energy five base pair stem-loop was predicted in the 3′-repeat tags of all crRNAs, which could explain the absence of additional 3′-terminal degradation for most crRNAs (Fig. 2.4). The abundance of the individual crRNAs varied significantly.

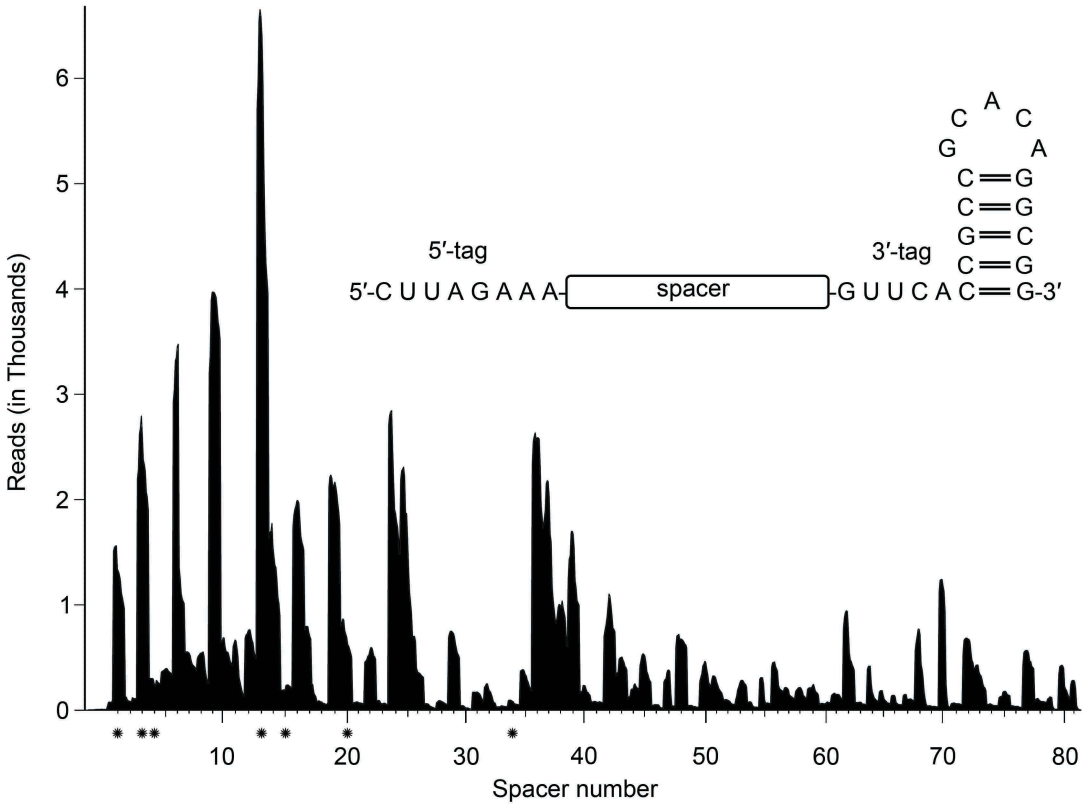


Figure 2.4: Analysis of CRISPR RNA abundance and termini. The coverage plot of *S. putrefaciens* CN-32 sequence reads (Illumina HiSeq2000) indicates the variable abundance of the individual crRNAs. Mature crRNAs contained 5′-terminal 8 nt tags (5′-CUUACAAA-3′) and 20 nt 3′-ends without further trimming events (inset). Spacers that are marked with an asterisk were tested for interference activity (see Fig. 2.7 and Fig. 2.9).

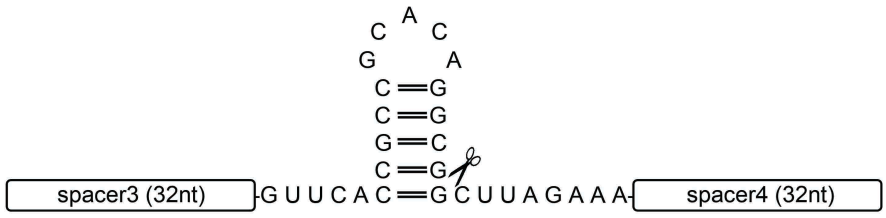
2.3.2 Cas6f is responsible for crRNA biogenesis

Bioinformatic analyses identified Sputcn32_1823 as a homolog of Cas6f in this variant Type I-F system. To verify its responsibility for crRNA processing, the recombinant protein was produced in *E. coli* with a C-terminal 6x His-tag. Inclusion body formation was observed when the protein was produced at 37°C. However, protein production at 18°C resulted in soluble protein, which was purified to >95% purity as evaluated via SDS-PAGE after Ni-NTA chromatography. The purified Cas6f was highly unstable and precipitated within 12h. Nevertheless, it showed endonuclease activity and cleaved a 100 nt long spacer3-repeat–spacer4 transcript (Fig. 2.5). Comparison of the obtained fragment sizes with the RNA-Seq analyses revealed the point of cleavage to be at the base of the stem-loop (Fig. 2.5A).

S. putrefaciens CN-32 Cas6f has a 37% amino acid identity with Csy4 (Cas6f) from *P. aeruginosa*. A sequence alignment of the two proteins verified a conserved

histidine residue at position 29 that was previously reported to be essential for Cas6f activity (Fig. 2.2B) (26). Mutation of this His29 residue to Ala abolished the endonuclease activity (Fig. 2.5B).

A



B

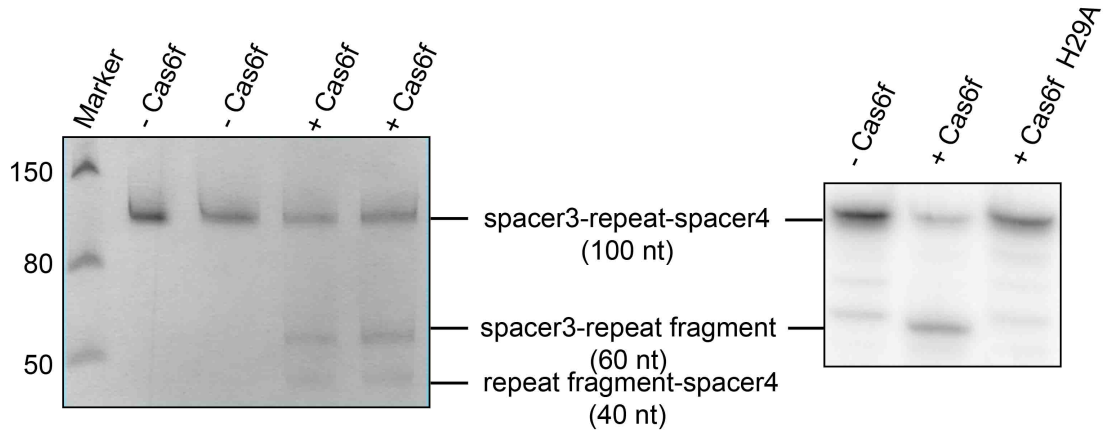


Figure 2.5: Cas6f endonuclease activity. (A) Schematic representation of the RNA transcript used in the Cas6f nuclease assay. The site of cleavage was determined based on the fragment sizes obtained from the nuclease assay and RNA-Seq analyses. (B) Recombinant Cas6f endonuclease cleaved spacer3-repeat-spacer4 RNA transcripts into two fragments of ~60 nt and ~40 nt length (left panel, toluidine blue stained urea-polyacrylamide gel). Mutation of the catalytic His29 abolished the endonuclease activity of the protein (right panel) for RNA transcripts labeled at 5′ end with γ-[^{32}P]-ATP.

2.4 The Type I-F variant CRISPR-Cas system show *in vivo* interference activity

2.4.1 Identification of the PAM sequence

Spacer sequences are known to match viral genomes and conjugative plasmids. Therefore, the tool CRISPRTarget was used to search for potential targets of the spacers from *S. putrefaciens* CN-32 (91). Of the 81 spacers, only spacer 33 (~97% identity between spacer and target) and spacer 34 (100% identity between

spacer and target) were found to potentially target *Shewanella* sp. W3-18-1 and *Shewanella* sp. MR7, respectively (Fig. 2.6). In both cases, the di-nucleotide GG was found at the 3′ end of target strand match. Hence, GG was considered to be the potential PAM sequence, which is in agreement with computational predictions by Mojica *et al.* (23).

Figure 2.6: Identification of the potential PAM. Potential targets for crRNA34 and crRNA33 (repeat sequence are indicated in grey and spacer sequence in black) were identified by the CRISPRtarget tool in prophages found in *Shewanella* sp. W3-18-1 and *Shewanella* sp. MR7, respectively. In both cases, the di-nucleotide ′GG′ PAM sequence (in red) was identified at the 3′ end of the target strand.

2.4.2 *In vivo* interference activity

Due to the low crRNA abundance, detection of CRISPR-Cas-mediated DNA interference activity in *S. putrefaciens* CN-32 required the establishment of a sensitive conjugation-based *in vivo* assay (Figure 2.7A). Two conjugative plasmids were used: (i) a non-target control plasmid (pNT = pBBR1MCS2) and (ii) a target plasmid (pT) containing the 13th spacer of the *S. putrefaciens* CN-32 CRISPR array (spacer13) and the PAM sequence ′GG′ at the 3′ end of the sequence complementary to spacer13 on the target strand. The two plasmids pNT and pT were conjugated from *E. coli* WM3064 into *S. putrefaciens* CN-32. In addition, markerless gene deletions were constructed for the genes *cas1*, *cas3*, *cas1821* and *cas1822* and the two plasmids were conjugated into these deletion strains. The colonies carrying pNT or pT were counted and the conjugation efficiency was calculated as pT over pNT. Relative conjugation efficiency (pT/pNT) below 1 indicates interference activity. Interference was observed in wild-type *S. putrefaciens* CN-32 cells (Fig. 2.7B) and the Δ*cas1* strain, which is in agreement with an exclusive role of Cas1 in spacer acquisition. In contrast, DNA interference activity was not

apparent in the Δ*cas1821* and Δ*cas1822* strains and significantly reduced in the Δ*cas3* strain. To investigate the reason for this loss of interference activity, the crRNA pool in these deletion strains was assayed. Total RNA was isolated from the different *S. putrefaciens* CN-32 cells and Northern blot analyses were performed using a radioactively labeled probe against the crRNA repeat sequence (Fig. 2.7C). The production of crRNAs was observed in wild-type cells and the Δ*cas1* and Δ*cas3* deletion strains. However, crRNAs were absent in the Δcas*1821* and Δcas*1822* strains. These results suggest that Cas6f generates mature crRNAs, which are then bound by Cas1821 and Cas1822 to maintain a stable crRNA pool in the cell. Thus, the absence of interference activity for the Δcas*1821* and Δcas*1822* strains correlates with the absence of crRNAs in these cells.

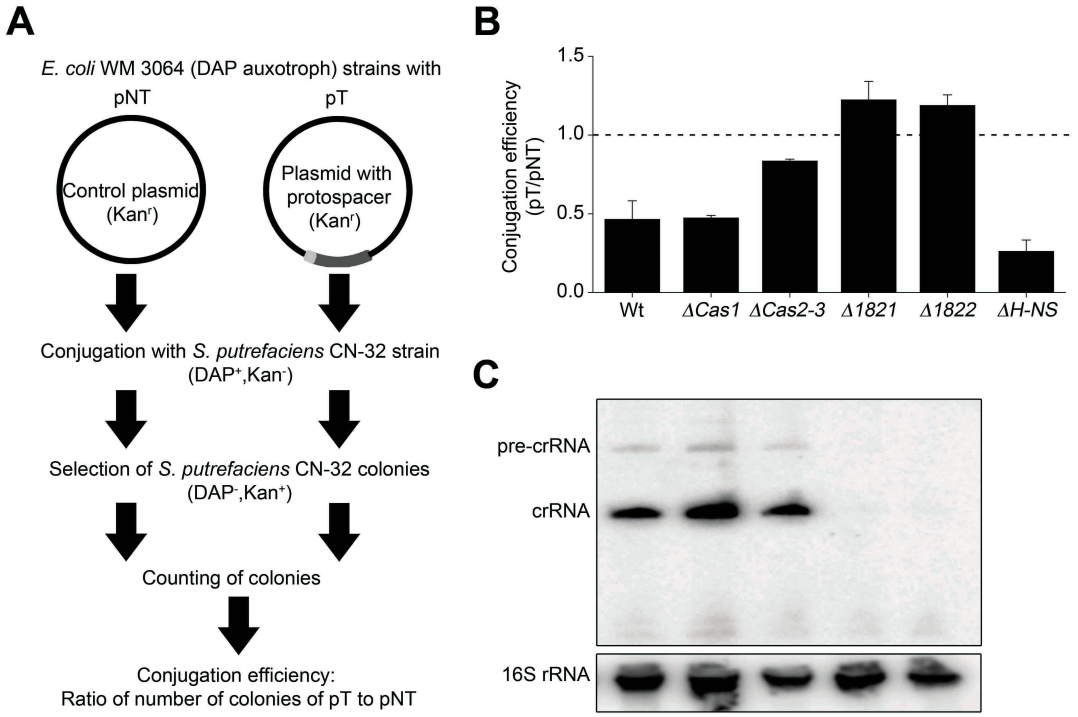


Figure 2.7: Conjugation assays reveal DNA interference activity of the *S. putrefaciens* CN-32 CRISPR-Cas system. (A) The experimental set-up of the conjugation assay is depicted. Two plasmids were used: (I) a control plasmid (pNT, plasmid non-target) and (II) a plasmid containing a PAM sequence (light grey) and a sequence matching spacer 13 (dark grey) of the *S. putrefaciens* CN-32 CRISPR array (pT, plasmid target). The number of *S. putrefaciens* CN-32 colonies carrying each plasmid was counted and the conjugation efficiency was determined. Interference is observed if pT/pNT is lower than 1 in triplicate assays. (B) The pT/pNT ratio is calculated for conjugation assays into *S. putrefaciens* CN-32 wild-type cells and strains containing deletions of genes encoding Cas1, Cas3, Cas1821, Cas1822 and H-NS. (C) Northern blot analyses were performed with extracted total

RNA from *S. putrefaciens* CN-32 wild-type and *cas* gene knockout strains using a 5′-γ-[^{32}P]-ATP labeled probe complementary to the repeat sequence. A stable crRNA pool was absent in the *S. putrefaciens* CN-32 Δ*cas1821* and Δ*cas1822* strains.

2.4.3 H-NS regulates the variant Type I-F CRISPR-Cas system

Previous studies of the type I-E CRISPR-Cas system in *E. coli* revealed that the heat-stable nucleoid-structuring (H-NS) protein, a global transcriptional repressor, can regulate the expression of CRISPR-Cas systems (92). A homolog of this protein was identified in *S. putrefaciens* CN-32. To assay its effect on the interference activity, a Δ*H-NS S. putrefaciens* CN-32 strain was constructed. A slight increase in the interference activity was observed for this strain in the conjugation-based assay (Fig. 2.7B). Although an increase in crRNA abundance could not be observed via northern blot analysis, a two-fold increase in *cas* gene transcripts was detected via qRT-PCR (Fig. 2.8), which could explain the observed increase in the interference activity.

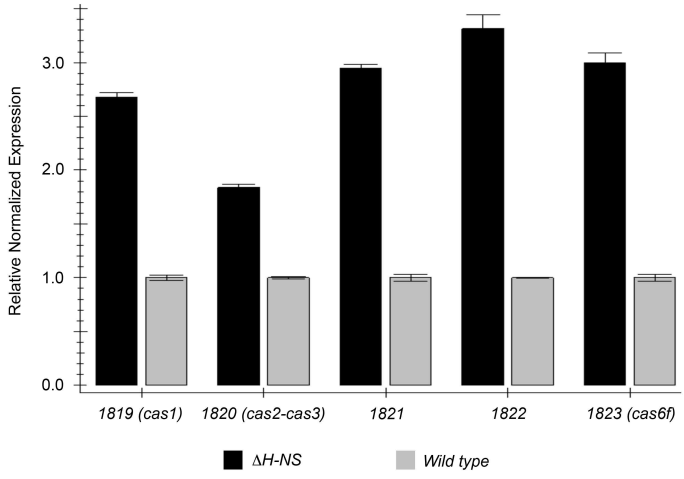


Figure 2.8: Analysis of *cas* gene transcript abundance. qRT-PCR analysis was performed to compare the *cas* gene transcript abundance between *S. putrefaciens* wild-type and Δ*H-NS* strains.

2.4.4 The CRISPR-Cas system recognizes PAM sequences and targets dsDNA

Conjugation between the donor and the recipient cell requires the transfer of one strand of plasmid DNA through an intercellular cytoplasmic bridge, beginning at the origin of transfer and progressing in 5′ to 3′ direction. The transferred strand is converted into circular double-stranded plasmid DNA in the recipient and a new

strand is synthesized in the donor to replace the transferred strand. In our experimental setup, the transferred strand can be directly targeted by the crRNA as it contains the sequence complementary to spacer13. To address if the minimal I-F variant system targets ssDNA or dsDNA, a plasmid was constructed with the sequence of spacer13 on the transferred strand. Thus, the complementary target of crRNA 13 exists only after synthesis of the second strand in the recipient cell. The relative conjugation efficiency (pT/pNT) for this plasmid in both, wild-type and Δ*H-NS* strains was comparable to the previous construct (Fig. 2.9A). This confirms that the I-F variant CRISPR-Cas system targets dsDNA.

To investigate if the absence of a large subunit had an effect on the recognition of PAM sequences, variants of the targeted plasmid were constructed that contained mutations in the PAM sequence. These variants contained the dinucleotides AG, GA, AA or CC at the 3′ end of the spacer13 match on the target strand. In addition, a plasmid was constructed with the first 10 nt of the target being replaced by a poly-A stretch. The relative conjugation efficiency (pT/pNT) of these constructs highlighted a loss of interference activity and demonstrated that PAM sequences are recognized in *S. putrefaciens* CN-32 (Fig. 2.9A).

The RNA-Seq analysis of crRNA production revealed a highly variable crRNA abundance profile (Fig. 2.4). Therefore, the effect of the abundance of individual crRNAs on interference activity was investigated. Thus, targeted plasmids with GG PAM sequences in which the spacer13 target was replaced with spacer1, spacer3 (abundant crRNAs) or spacer15, spacer20 and spacer34 targets (low crRNA abundance) were generated. The conjugation assays verified that highly abundant crRNAs (spacers1, 3, 13) yield efficient interference activity (Fig. 2.9B). In agreement, the observed absence of stable crRNAs with spacer34 in *S. putrefaciens* CN-32 correlated with a complete loss of interference activity. However, interference activity was observed for crRNAs with spacer15 and spacer20 even though only minimal amounts of these crRNAs were detected in our RNA-Seq studies (Fig. 2.9B). These results indicate that variable crRNA abundance is a factor that can influence the efficiency of DNA targeting.

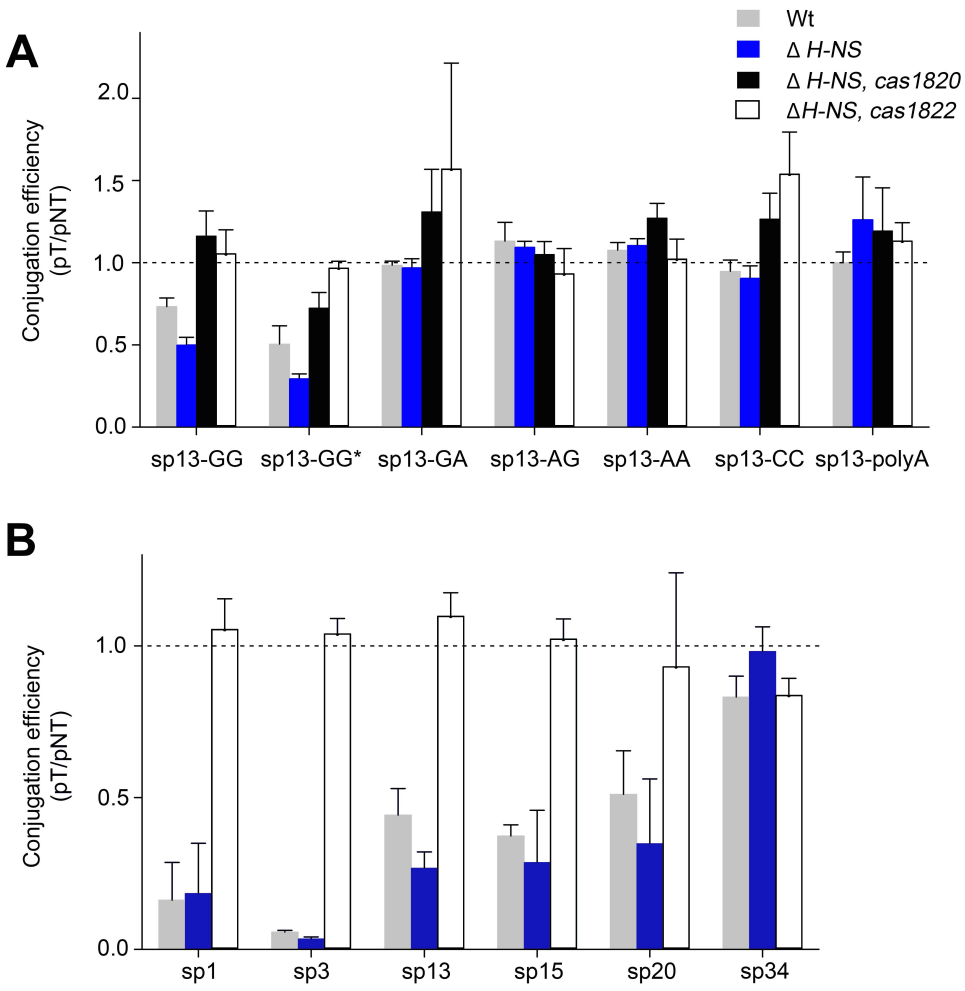


Figure 2.9: Analysis of the DNA target specificity of the *S. putrefaciens* CN-32 CRISPR-Cas system. Conjugation assays (see Fig. 2.7) were used to analyze plasmid DNA target variants for DNA interference in *S. putrefaciens* CN-32 cells. (A) Interference was observed if the transferred DNA strand contained a sequence complementary to the crRNA13 (sp13-GG) or a sequence identical with spacer13 (sp13-GG*). Interference was evident for the wild-type and Δ*H-NS* strains, which contained all *cas* genes. However, the deletion of *cas3* (*Sputcn32_1820*) or cas*1822* abolished interference activity. The presence of the PAM sequence GG at the 3′ end of the DNA target was essential for the interference activity and PAM mutants (sp13-AG, sp13-GA, sp13-AA or sp13-CC) and a 10 nt poly-A sequence disrupting crRNA/DNA complementarity (sp13-polyA) were not targeted. (B) DNA interference activity was dependent on spacer sequence or crRNA abundance (Spacers 1, 3, 13, 15, 20 and 34, see Fig. 2.4).

2.5 *In vitro* characterization of Cas proteins

2.5.1 Characterization of recombinant Cas3 as an ssDNA nuclease

Bioinformatic predictions indicated Cas3 to be a fusion protein with an N-terminal Cas2 like domain fused to the nuclease and the helicase domains of Cas3. In other Type I systems, Cas3 was reported to contain metal-dependent ssDNA nuclease activity responsible for target DNA degradation. The reduction in the interference activity for the Δcas*3 S. putrefaciens* strain suggests a similar role for the *S. putrefaciens* Cas3 protein. Thus, the recombinant protein was produced in *E. coli* to test its activity *in vitro*.

Soluble recombinant Cas3 was produced as a SUMO-tag fusion construct in cells grown in NZ-Amine media and purified via Ni-NTA affinity chromatography (Fig. 2.10A). Protein purity was relatively low and could not be improved by including additional purification strategies like ion exchange chromatography and size-exclusion chromatography. Additionally, the removal of the SUMO-tag from Cas3 using SUMO-protease was unsuccessful. Hence, to improve the yield and purity of the protein several different expression conditions were tested (Table 2.2). However, in all of the tested conditions, the protein was not produced or formed inclusion bodies. Therefore, the obtained SUMO-Cas3 was assayed for its ssDNA nuclease activity. When incubated in the presence of the transition metal ions Mg^{2+} or Mn^{2+}, Cas3 cleaved 5′-[γ-P^{32}]–ATP labeled ssDNA. The absence of these metal ions or the addition of EDTA abolished the nuclease activity (Fig. 2.10B). For *S. thermophilus* Cas3, the His76 and Asp77 residues in the HD nuclease domain were shown to be essential for ssDNA hydrolysis (89). These residues were also identified in the HD domain of *S. putrefaciens* CN-32 Cas3. The His156 and Asp157 residues were mutated to alanine residues. However the variant proteins could not be produced in *E. coli*.

Table 2.2: List of expression conditions tested for the production of soluble recombinant Cas3.

Construct	Medium	Temperature	Induction of protein production
Cas3 with N-terminal maltose binding protein (MBP)-tag	LB	37°C	1 mM IPTG, 3h at 37°C
Cas3 with N-terminal MBP-tag	NZA	37°C	1 mM IPTG, 3h at 37°C
Cas3 with N-terminal MBP-tag	LB	37°C -18°C	1 mM IPTG, overnight at 18°C

Cas3 with N-terminal MBP-tag	NZA	37°C -18°C	1 mM IPTG, overnight at 18°C
Cas3 with C-terminal 6xHis-tag	LB	37°C	1 mM IPTG, 3h at 37°C
Cas3 with C-terminal 6xHis-tag	NZA	37°C	1 mM IPTG, 3h at 37°C
Cas3 with C-terminal 6xHis-tag	LB	37°C -18°C	1 mM IPTG, overnight at 18°C
Cas3 with C-terminal 6xHis-tag	NZA	37°C -18°C	1 mM IPTG, overnight at 18°C
Cas3 with C-terminal cysteine protease domain (CPD)-tag	NZA	37°C	1 mM IPTG, 3h at 37°C
Cas3 with C-terminal CPD-tag	NZA	37°C -18°C	1 mM IPTG, overnight at 18°C

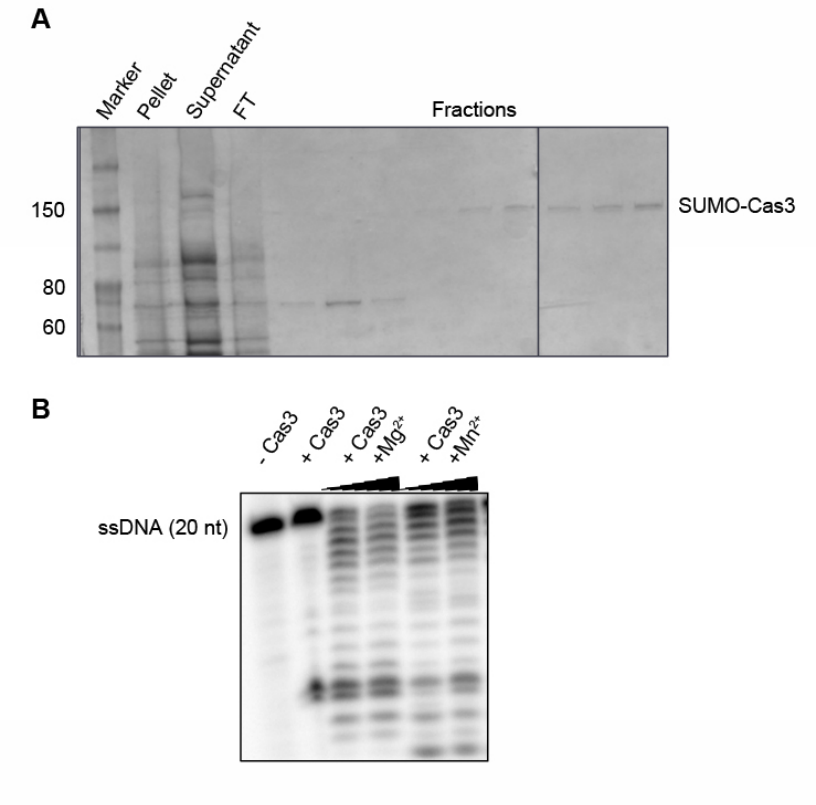


Figure 2.10: ssDNA nuclease activity of Cas3. (A) 6% SDS-PAGE of SUMO-Cas3 after Ni-NTA purification. SUMO-Cas3 co-purified with low molecular weight contaminant proteins. (B) Increasing concentrations (250 nM and 500 nM) of recombinant SUMO-Cas3 were incubated with 50 nM of 5′-γ-[^{32}P]-ATP labeled 20 nt ssDNA in the presence of 10 mM Mg^{2+} or Mn^{2+}. The cleavage products were separated by 20% UREA-PAGE and visualized via phosphorimaging.

2.5.2 Characterization of recombinant Cas1821 as a helical Cascade backbone protein

The absence of observable crRNAs in the Δcas*1821* and Δcas*1822* strains suggests that the encoded Cas proteins are required for maintaining a stable crRNA pool in the cell. Thus, these two proteins Cas1821 and Cas1822 could fulfill the roles of Cas7 and Cas5 proteins even though sequence similarity calculations did not indicate them as members of these Cas protein families (35). Hence, recombinant *S.*

putrefaciens CN-32 Cas1821 and Cas1822 proteins were individually produced in *E. coli*.

Soluble recombinant Cas1821 was produced as a SUMO-tag fusion construct. The SUMO-Cas1821 was purified via Ni-NTA chromatography and co-eluted with bound nucleic acid contaminants. Therefore, a high-salt washing step with a buffer including 1 M NaCl was added before protein elution from the Ni-NTA column, and a second cation exchange chromatography step was included in the purification protocol. Before cation exchange chromatography, SUMO-Cas1821 was treated with SUMO-protease to remove the tag. This procedure yielded Cas1821 with a purity of >95 % and without nucleic acid contaminants. Gel-elution chromatography of Cas1821 revealed a ~35 kDa monomer (Fig. 2.11C). Electrophoretic mobility shift assays of Cas1821 with radiolabeled crRNA transcripts revealed a slow migrating band indicating the crRNA-binding potential of Cas1821 (Fig. 2.11A). This band disappeared upon addition of 1 µg of competitor yeast RNA, suggesting a non-specific interaction. This behavior is in agreement with the observed non-specific binding of *E. coli* contaminant RNA. Recombinant Cas1821 bound to these contaminants eluted near the void volume during gel-elution chromatography (Fig. 2.11C). These fractions were visualized by electron microscopy and revealed the formation of long helical filament-structures of varying length (Fig. 2.11B). These observations suggest that Cas1821 could fulfill the role of Cas7, which was shown to form the helical Cascade backbone filament after unspecific RNA binding in other CRISPR-Cas subtypes (e.g. type I-A, type I-E) (34,93).

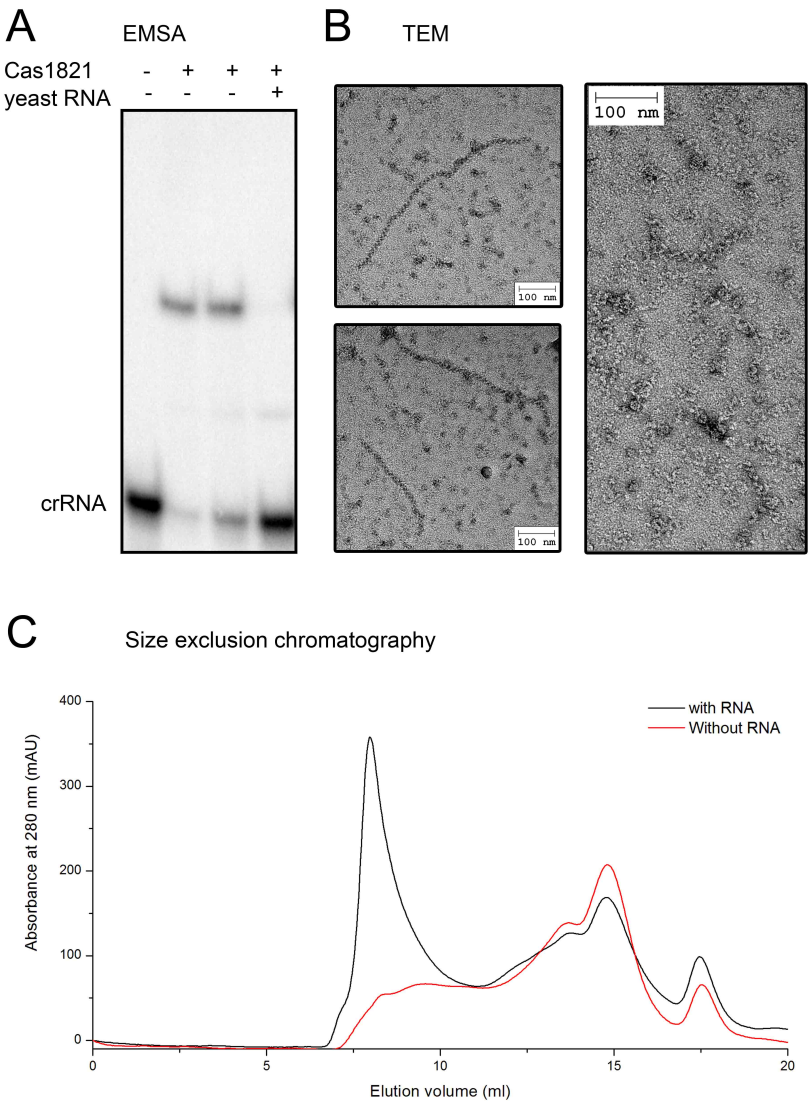


Figure 2.11: Characterization of the recombinant Cas1821 protein. (A) EMSA assays indicate the binding of 5′-γ-[^{32}P]-ATP labeled crRNA by 250 nM Cas1821 (Lane 2) or 500 nM Cas1821 (Lane 3). Bands were separated by 6% native PAGE. The RNA binding is unspecific and outcompeted by yeast total RNA. (B) Transmission electron micrographs of negatively stained helical structures formed by RNA-bound recombinant Cas1821 (provided by Prof. Dr. Andreas Klingl). (C) Size exclusion chromatograms of Cas1821 with and without bound RNA. The elution profile of apo-Cas1821 measured at 280 nm (in red) shows a single major peak corresponding to a mass of ~35 kDa was observed. When bound to RNA, most of the protein was found to elute in the void volume (in black).

2.5.3 Purification of recombinant Cas1822

The production of recombinant Cas1822 as a SUMO fusion construct or with a C-terminal 6x His-tag resulted in the formation of inclusion bodies. Several different expression conditions were tested before solubilized Cas1822 was produced as a

fusion protein containing a C-terminal cysteine protease domain (CPD)-tag (Table 2.3). A low concentration of Cas1822 was purified via Ni-NTA chromatography and subsequently treated with Inositol-6-phosphate to induce autocatalytic cleavage of the CPD-tag. The purity of the obtained Cas1822 was low and could not be improved by including additional purification steps (Fig 2.12). Hence, *in vitro* biochemical characterization of the protein was not possible.

Table 2.3: List of expression conditions tested for the production of soluble recombinant Cas1822.

Construct	Medium	Temperature	Induction of protein production
Cas1822 with N-terminal SUMO-tag	LB	37°C	1 mM IPTG, 3h at 37°C
Cas1822 with N-terminal SUMO-tag	NZA	37°C	1 mM IPTG, 3h at 37°C
Cas1822 with N-terminal SUMO-tag	LB	37°C -18°C	1 mM IPTG, overnight at 18°C
Cas1822 with N-terminal SUMO-tag	NZA	37°C -18°C	1 mM IPTG, overnight at 18°C
Cas1822 with C-terminal 6x His-tag	LB	37°C	1 mM IPTG, 3h at 37°C
Cas1822 with C-terminal 6xHis-tag	NZA	37°C	1 mM IPTG, 3h at 37°C
Cas1822 with C-terminal 6xHis-tag	LB	37°C -18°C	1 mM IPTG, overnight at 18°C
Cas1822 with C-terminal 6xHis-tag	NZA	37°C -18°C	1 mM IPTG, overnight at 18°C

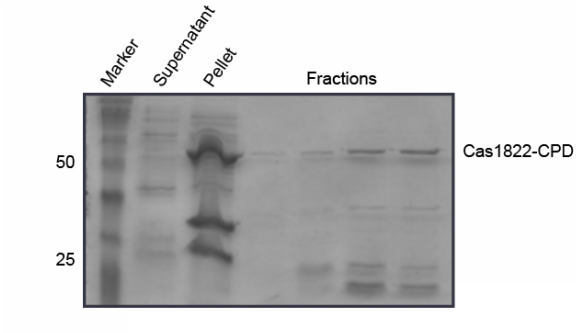


Figure 2.12: Characterization of recombinant of Cas1822 protein. SDS-PAGE of Ni-NTA purified Cas1822-CPD. The eluted fractions contained multiple contaminant proteins co-eluting with Cas1822-CPD.

2.6 PAM recognition mechanism

2.6.1 Identification of host proteins involved in PAM recognition

The loss of interference activity for target DNA molecules with mutations of the PAM sequence suggests that the Type I-F variant Cascade contains a PAM recognition mechanism. The absence of a *cas* gene encoding for the large subunit protein suggests that either the other Cas proteins of the Cascade or other host proteins perform the PAM recognition function. A potential host protein candidate

involved in target recognition by Cascade was not identified by bioinformatic approaches. Therefore, potential candidates were identified using co-immunoprecipitation. Three *S. putrefaciens* CN-32 strains were constructed that produced Cas1821, Cas1822 and Cas6f with an N-terminal 6x His-HA-tag, respectively. To determine the effect of the tags on the functionality of the proteins, interference activity of the three strains was tested via the conjugation-based assay. The relative conjugation efficiency of *Cas1822-His-HA* and *Cas6f-His-HA* strains suggested a loss of interference activity. In addition, the *in vivo* crRNA abundance was probed in these strains. The addition of a His-HA tag on Cas6f or Cas1822 resulted in a loss of the crRNA pool in the cell, which is consistent with the loss of interference activity in these strains. In contrast, the *Cas1821-His-HA* strain carried a stable crRNA pool and showed interference activity, as its conjugation efficiency was comparable to the background Δ*H-NS* strain (Fig. 2.13). Hence, this strain was used to perform Co-IP studies.

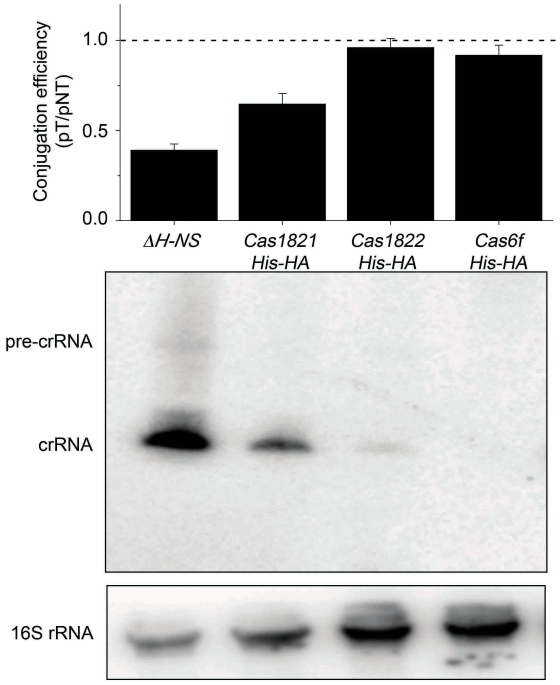


Figure 2.13: Interference activity of strains producing His-HA-tagged Cas proteins. (A) *S. putrefaciens* CN-32 strains producing His-HA-tagged Cas proteins were tested for DNA interference using the conjugation assay (see Fig. 2.7). Interference was not observed in *Cas1822-His-HA* and *Cas6f-His-HA* strains. Northern blot analyses performed with a radiolabeled probe complementary to the repeat sequence associated the loss of interference in these strains to the absence of a stable crRNA pool.

The *Cas1821-His-HA S. putrefaciens* CN-32 strain produces Cas1821 with an N-terminal His-HA-tag. Therefore, purification of this protein using Ni-NTA chromatography was performed. The eluted fractions were subjected to SDS-PAGE and the presence of His-HA-tagged Cas1821 was detected by western blot analysis using anti-His antibodies. The fractions containing His-HA-tagged Cas1821 were pooled and incubated with magnetic Dynabeads coupled to anti-HA antibodies. After incubation, the beads were thoroughly washed and the bound proteins were eluted using SDS loading buffer. Following separation by SDS-PAGE, the eluted proteins were identified via mass spectrometry. On comparing the proteins eluting with His-HA-Cas1821 to the Δ*H-NS* strain, potential Cas1821 interacting partners could be identified. Non-specific binding of *S. putrefaciens* CN-32 proteins to the beads in both the background and *Cas1821-His-HA* strains was observed (Fig. 2.14). Using mass spectrometry, a large list of identified proteins was obtained for the elution fraction from the beads for both of the strains (Appendix 3). Specific interaction partners of His-HA-Cas1821 could not be observed.

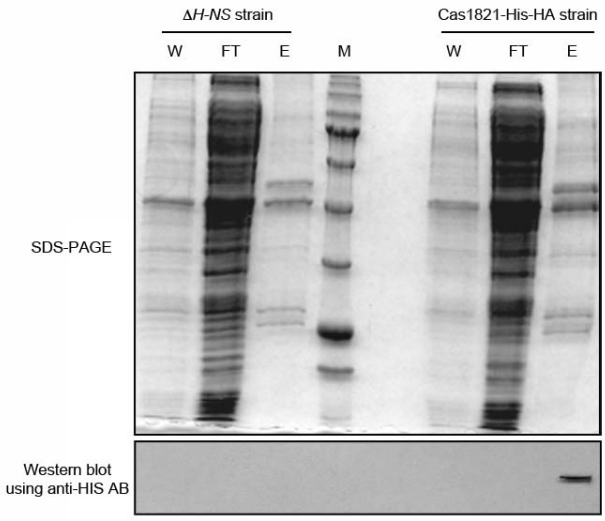


Figure 2.14: Co-IP analysis using *Cas1821-His-HA S. putrefaciens* strain. Proteins samples eluted (Lane E) from Dynabeads after performing Co-IP on Δ*H-NS* and *Cas1821-His-HA* strains were separated via SDS-PAGE. The presence of Cas1821-His-HA was detected by western blot analyses using anti-His antibodies. The lanes indicate the wash fraction (W), flow-through after applying the sample (FT) and the elution fraction.

2.6.2 Investigation of DinG participation

The *S. putrefaciens* CN-32 CRISPR-Cas system is flanked by a homolog of the *DinG* gene (*Sputcn32_1824*), an ATP-dependent DNA helicase that was predicted to be a functional component of Type I-U CRISPR-Cas systems (35). To test if this protein plays a role in DNA interference, a Δ*DinG* strain was constructed. A loss of interference activity for this strain was not observed in the conjugation assays, indicating that this DinG homolog does not play a direct role in DNA interference and is not a component of subtype I-F variant Cascade (Fig. 2.15). This is in agreement with the observation, that this DinG homolog is conserved in *S. putrefaciens* strains without CRISPR-arrays or other CRISPR-Cas subtypes.

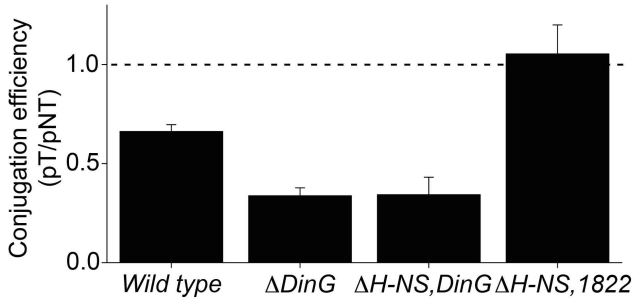


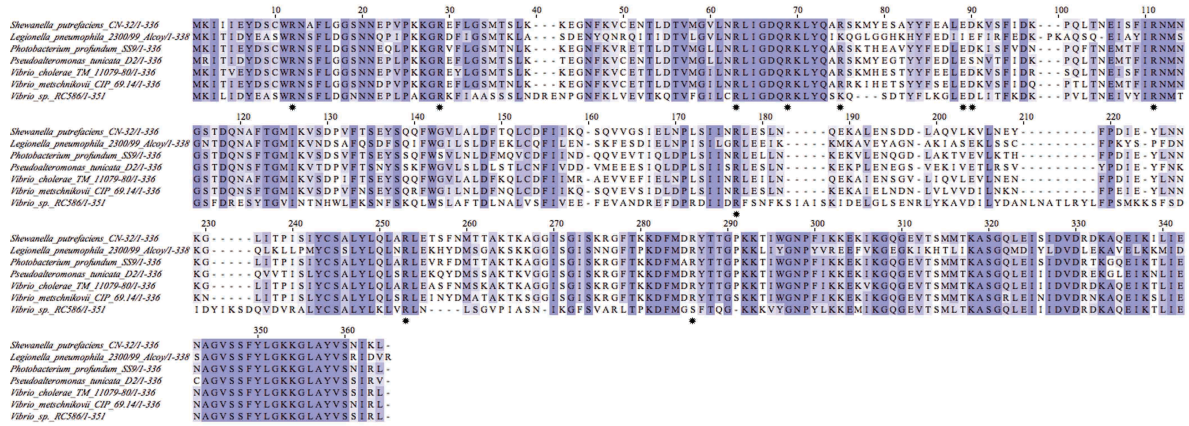
Figure 2.15: Influence of DinG on DNA interference activity. A deletion strain of *DinG* was generated in *S. putrefaciens* CN-32 wild-type and Δ*H-NS* strains and tested for interference using the conjugation-based assay (see Fig. 2.7). The interference-inactive *ΔH-NS,Cas1822* strain served as control.

2.6.3 *In vivo* characterization of a potential DNA target interaction partner

The absence of a large subunit suggests that other Cas proteins can substitute for its role(s) during DNA target interaction. In *S. putrefaciens*, Cas6f is responsible for crRNA biogenesis and Cas1821 is proposed to form the backbone of the Cascade. In contrast, the role of Cas1822 is unclear and it is a potential candidate for performing the role of large subunit. Therefore, the effect of mutations of conserved arginine residues in Cas1822 on DNA interference was tested. Basic residues like arginine stabilize protein-nucleic acid interactions by forming salt bridges with the phosphodiester backbone and were thus chosen to be mutated. A multiple alignment of Cas1822 proteins revealed conserved arginine residues and a potential Asp/Glu motif (Fig. 2.16A). We constructed point mutations of the *cas1822*

gene at the appropriate positions to generate corresponding Ala substitutions in Cas1822 *in vivo*. The Cas1822 variants were tested for interference activity. Interference was observed for Δ*H-NS/R170A*, Δ*H-NS/R225A* and Δ*H-NS/E86A,D87A* strains (Fig. 2.16B). A Northern blot analysis verified the presence of mature crRNAs in these strains (Fig. 2.16B). In contrast, the two strains Δ*H-NS/R60A* and Δ*H-NS/R66A* lacked interference activity and stable crRNAs. We propose that these arginine residues are required for Cascade formation and crRNA protection. Finally, the two strains Δ*H-NS/R29A* and Δ*H-NS/R258A* did not show interference activity even though a stable crRNA pool was observed in the cell. Thus, these arginine residues are involved in either the direct interaction with target DNA or with the Cas component that recognizes the PAM motif.

A



B

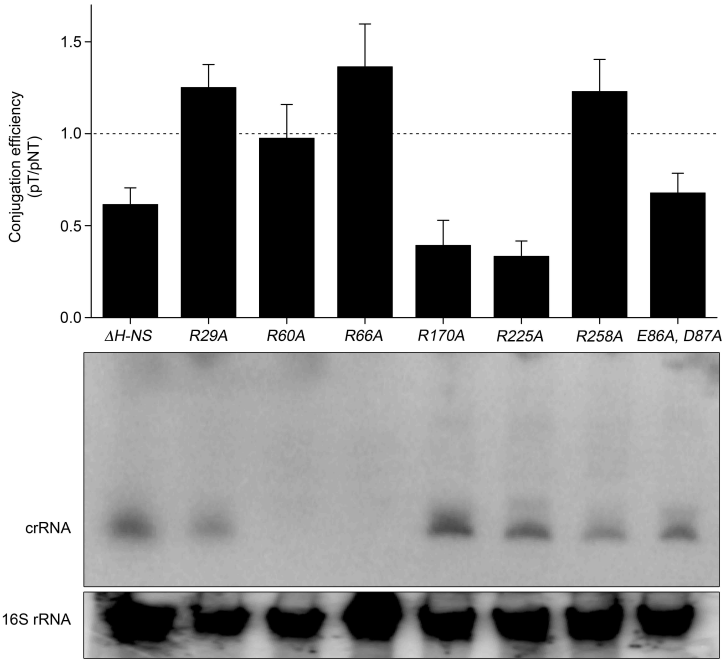


Figure 2.16: *In vivo* characterization of Cas1822 point-mutants. (A) Multiple sequence alignment of Cas1822 from *S. putrefaciens* CN-32 and its homologs identified conserved arginine residues and a partially conserved Asp/Glu motif (marked by an asterisk). (B) *S. putrefaciens* CN-32 strains carrying Cas1822 proteins with the indicated point mutations were tested for DNA interference using the conjugation-based assay (see Fig.2.7). Northern Blot analyses with anti-repeat probes revealed the cellular crRNA pool of these strains.

2.7 Characterization of the minimal Cascade complex

2.7.1 Reconstitution of the subtype I-F variant Cascade complex from inclusion bodies

The soluble production of the individual Cas proteins was problematic and formation of inclusion bodies was observed for Cas1, Cas3, Cas1822 and Cas6. In a recent publication, the reconstitution of an active Type I-A Cascade complex from inclusion bodies of six Cas proteins was described (34). Therefore, a similar approach for obtaining a reconstituted variant Type I-F Cascade was considered. Inclusion bodies of the Cas proteins Cas1, Cas3, Cas1822 and Cas6 were purified from *E. coli* in high purity and solubilized by the addition of the denaturing agent, guanidine hydrochloride (Gua-HCl) (Fig. 2.17A). The solubilized inclusion bodies of the four Cas proteins and soluble Cas1821 were mixed in small quantities (20 µg/protein) and added to a large volume of buffer (2 ml) to rapidly dilute the concentration of the denaturing agent. The reduction of the denaturing agent results in the potential refolding of the denatured proteins into their active form. To determine the optimum reconstitution procedure, various buffer conditions and temperatures were tested and chosen based on the amount of proteins obtained in soluble fraction observed via SDS-PAGE (Fig. 2.17B and Table 2.4). By this procedure the best refolding condition was determined to be in a buffer containing 50 mM TRIS/HCl pH 7.0, 300 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT and 10% glycerol at 4°C.

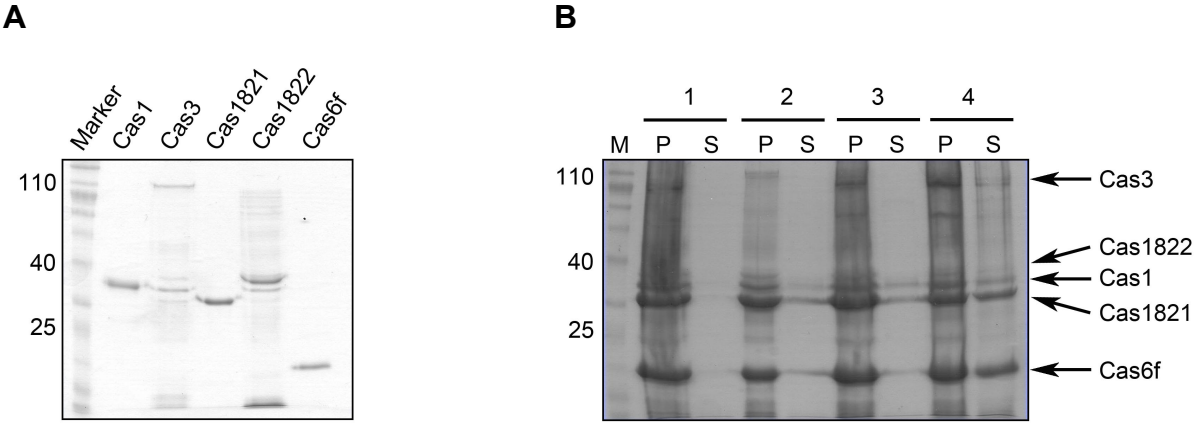


Figure 2.17*: In vitro* refolding of Cas proteins. (A) SDS-PAGE of the solubilized inclusion bodies of the individual Cas proteins from *S. putrefaciens*. (B) The solubilized (S) and precipitated (P) fractions of the tested reconstitution condition

were analyzed via SDS-PAGE. To determine the optimum reconstitution procedure several different reconstitution conditions were tested (buffer condition listed in Table 2.4).

Table 2.4: List of the reconstitution buffer conditions that were tested.

	Buffer conditions
1	50 mM MES pH 6.0, 10% glycerol, 300 mM NaCl
2	50 mM TRIS/HCl pH 7.0, 10% glycerol, 300 mM NaCl
3	50 mM TRIS/HCl pH 8.0, 10% glycerol, 300 mM NaCl
4	50 mM MES pH 6.0, 10% glycerol, 300 mM NaCl, 1 mM DTT
5	50 mM TRIS/HCl pH 7.0, 10% glycerol, 300 mM NaCl, 1 mM DTT
6	50 mM TRIS/HCl pH 8.0, 10% glycerol, 300 mM NaCl, 1 mM DTT
7	50 mM MES pH 6.0, 10% glycerol, 300 mM NaCl, 0.03 mM GSSG, 0.03 mM GSH, 1 mM EDTA
8	50 mM TRIS/HCl pH 7.0, 10% glycerol, 300 mM NaCl, 0.03 mM GSSG, 0.03 mM GSH, 1 mM EDTA
9	50 mM TRIS/HCl pH 8.0, 10% glycerol, 300 mM NaCl, 0.03 mM GSSG, 0.03 mM GSH, 1 mM EDTA
10	50 mM MES pH 6.0, 10% glycerol, 300 mM NaCl, 10 mM $MgCl_2$
11	50 mM TRIS/HCl pH 7.0, 10% glycerol, 300 mM NaCl, 10 mM $MgCl_2$
12	50 mM TRIS/HCl pH 8.0, 10% glycerol, 300 mM NaCl, 10 mM $MgCl_2$

An alternative method was used to increase the yield of the reconstituted proteins. Inclusion bodies of the proteins were mixed at a higher concentration (300 µg/protein) in 50 mM TRIS/HCl pH 7.0, 300 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT and 10% glycerol containing 4M Gua-HCl. Instead of rapidly diluting this mixture, the denaturing agent was removed by step-wise dialysis. In this procedure, the denatured proteins were first brought to equilibrium with a high denaturant concentration. Subsequently, the concentration of the denaturant was decreased and brought to equilibrium at a medium concentration. This procedure was continued until the equilibrium at a very low concentration of denaturant was achieved. Although the refolding efficiency of this procedure is better, the formation of inactive or mis-folded proteins and protein aggregates is higher. Therefore, after the refolding procedure, the dialyzed mixture was centrifuged to separate the reconstituted complex from the precipitated proteins and used for further analyses. Following this procedure, Cas3 and Cas1 were found to not be essential for the refolding procedure, as omission of these proteins did not affect the efficiency of the refolding of the other proteins. Therefore two different combinations of reconstituted proteins were obtained, Combi-1, which contained only Cas1821, Cas1822 and Cas6f and Combi-2, which contained all five Cas proteins (Fig. 2.18).

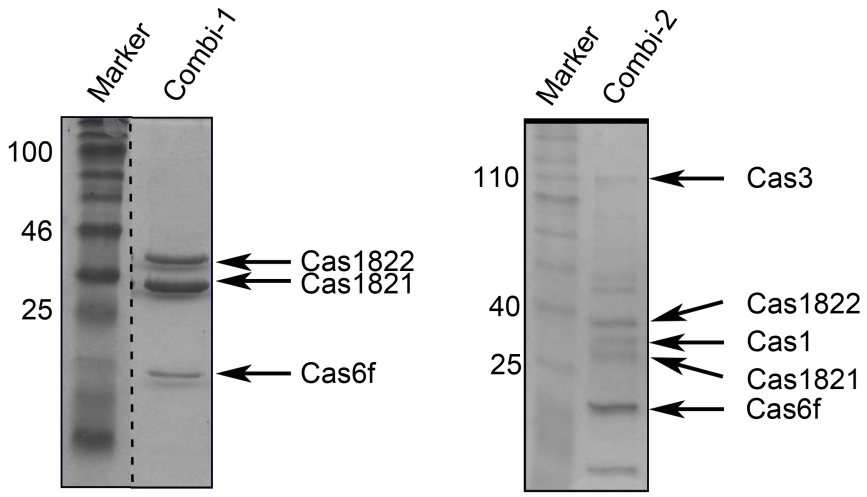


Figure 2.18: Large-scale reconstitution of Cas proteins. SDS-PAGE of solubilized Cas proteins obtained from the large-scale reconstitution procedure. Two different combinations of Cas proteins could be reconstituted, Combi-1 contained Cas1821, Cas1822 and Cas6f where as Combi-2 contained all 5 Cas proteins.

To verify that the reconstituted proteins were folded correctly, the activity of individual proteins was tested. First, the endoribonuclease activity of Cas6 was tested. A cleavage product was obtained by incubating a radiolabeled spacer3-repeat-spacer4 transcript with the reconstituted Combi-1 or Combi-2 complexes. The endonuclease activity depended on the presence of the catalytic His29 of Cas6. Although the use of the His29Ala Cas6 mutant did not affect the reconstitution efficiency, a loss of the ribonuclease activity for Combi-1 was observed (Fig. 2.19A). Since Cas1821, Cas1822 and Cas6f were proposed to form the Cascade complex, the specificity of Combi-1 towards crRNA was tested. While soluble Cas1821 bound crRNA nonspecifically, the interaction between the proteins of Combi-1 and crRNA could not be disrupted by the addition of total yeast RNA (Fig. 2.19B). Additionally, when subjected to size-exclusion chromatography the reconstituted Combi-1 revealed a major peak at ~75 kDa corresponding to a dimer of Cas1821 and Cas1822 (Fig. 2.20). Since Cas3 was included in Combi-2, the ssDNA nuclease activity of the complex was tested and found to cleave ssDNA in a metal-dependent manner (Fig. 2.19C).

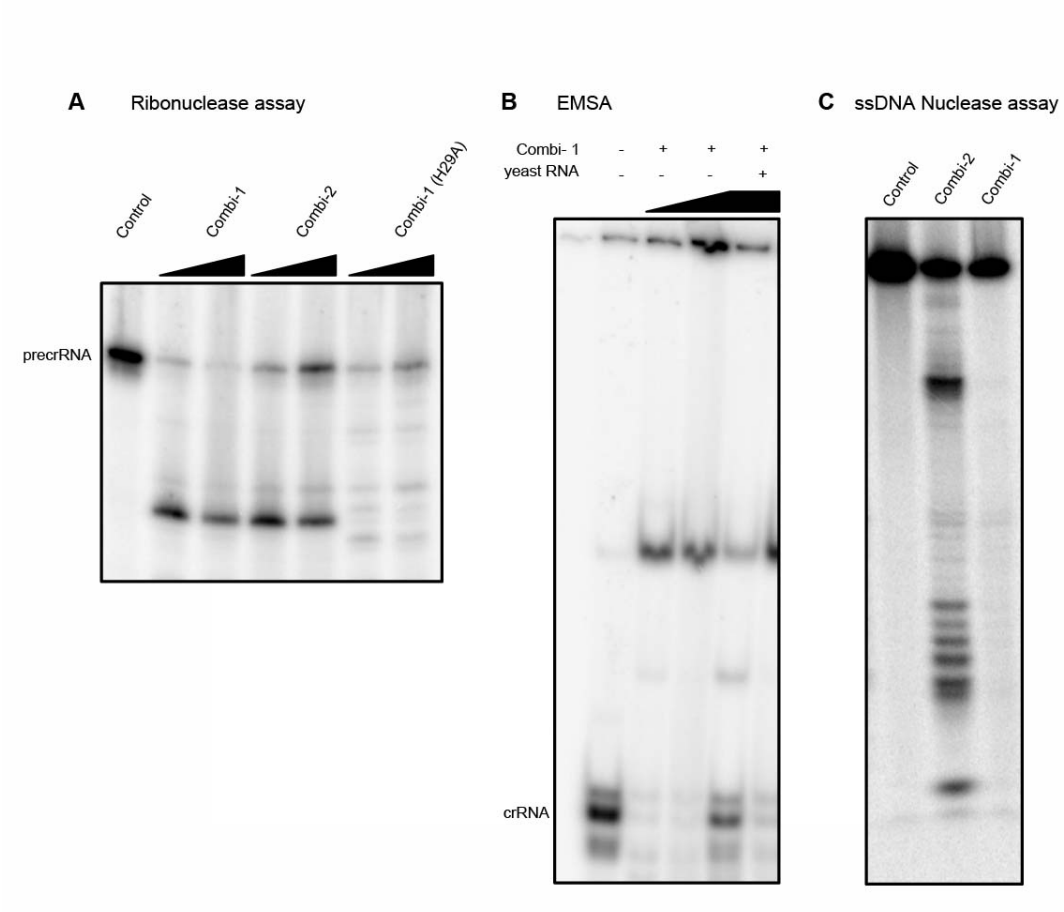


Figure 2.19: Activity of reconstituted Cas proteins (A) The endoribonuclease activity of Combi-1 and Combi-2 was tested as these complexes contained reconstituted Cas6. 250 nM and 500 nM Combi-1 and Combi-2 were incubated with 50 nM of 5′ radiolabeled spacer-repeat-spacer RNA transcript at 30 °C for 30 min. Both Combi-1 and Combi-2 cleaved the RNA and the cleavage product was visualized by phosphorimaging after 20% UREA-PAGE. This activity was specific, as the use of the Cas6f H29A mutant during reconstitution abolished the endoribonuclease activity. (B) Reconstituted Combi-1 specifically bound to the crRNA. A slow migration band could be visualized after 6% non-denaturing PAGE on incubating 250 nM and 500 nM Combi-1 with 50 nM of crRNA. This band was not lost but reduced when 1µg of yeast RNA was added as competitor. (C) The ssDNA nuclease activity of Cas3, which was reconstituted in Combi-2 was tested. 500 nM of Combi-2 could cleave 50 nM of ssDNA in the presence of 10 mM Mg^{2+} ions. Under similar condition, Cas3 lacking Combi-1 did not show ssDNA nuclease activity.

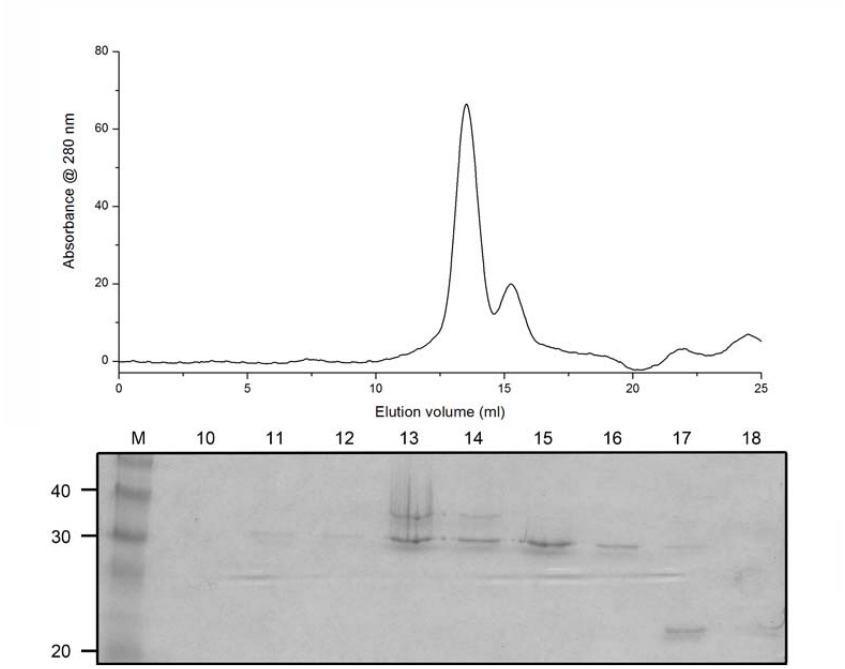


Figure 2.20: Size-exclusion chromatography of reconstituted Combi-1. Dimers of Cas1821 and Cas1822 (Lanes 13 and 14) were observed in the major peak of the elution profile from Combi-1. After SDS-PAGE, the second peak was found to contain monomers of Cas1821 (Lanes 15 and 16) and Cas6f (Lane 17).

The above experiments suggested that correctly folded and active Cas proteins were reconstituted from inclusion bodies. The next aim was to show that the reconstituted proteins could recognize and cleave target DNA in an *in vitro* interference assay. Therefore, crRNA4 bound Combi-2 complex was incubated with ds-spacer4-DNA target that carried a radiolabel at the 5′ end on one of its strands. Since ds-spacer1-DNA is not complementary to crRNA4, it was used as a control for the interference assay. Although activity of individual proteins could be observed, cleavage of a dsDNA by the reconstituted proteins loaded with crRNA could not be observed (data not shown).

2.7.2 Production of recombinant subtype I-F variant Cascade

(The production of the Type I-F variant Cascade was investigated by MSc. student Daniel Gleditzsch under my supervision.)

Due to the described problems associated with inclusion body refolding, the simultaneous production of all Cascade components in *E. coli* was considered. Therefore, two compatible plasmids, one for the production of the Cas proteins Cas1821, Cas1822 and Cas6f and one for the production of pre-crRNA were

constructed and transformed into *E. coli*. Only Cas1821 contained an N-terminal 6xHis-Tag and was used to successfully purify the protein from *E. coli*. Interestingly, co-purifying with Cas1821 were the Cas proteins Cas1822 and Cas6f and a crRNA of ~60nt in length. The purified proteins were subjected to size-exclusion chromatography and the RNA and protein content of the fractions was analyzed (Fig. 2.21). A distinct peak revealed the formation of a stable complex with an estimated size of approximately 215 kDa. The Cas1821 protein was overrepresented in the complex, which supports its role as Cascade backbone. Cas1822 and Cas6f were co-purified and are suggested to be the proteins that cap the 3′ hairpin tag (Cas6f) and the 8 nt 5′ tag (Cas1822) of the crRNA. Additionally a mature crRNA was observed to co-purify in the Cascade containing fractions. Thus confirming that Cas6f processed the pre-crRNA transcript in *E.coli* and loaded the Cascade complex. EM pictures of the eluted complex revealed a crescent shaped architecture similar to the I-F Cascade observed in *P. aeruginosa*. In addition, Cas1821-Cas1822 dimers were detected during size exclusion chromatography. Observation of similar dimers in the reconstituted proteins suggests that these units are similar to the Cas5-Cas7 dimers observed for the type I-A Cascade and involved in the assembly of the Cascade complex.

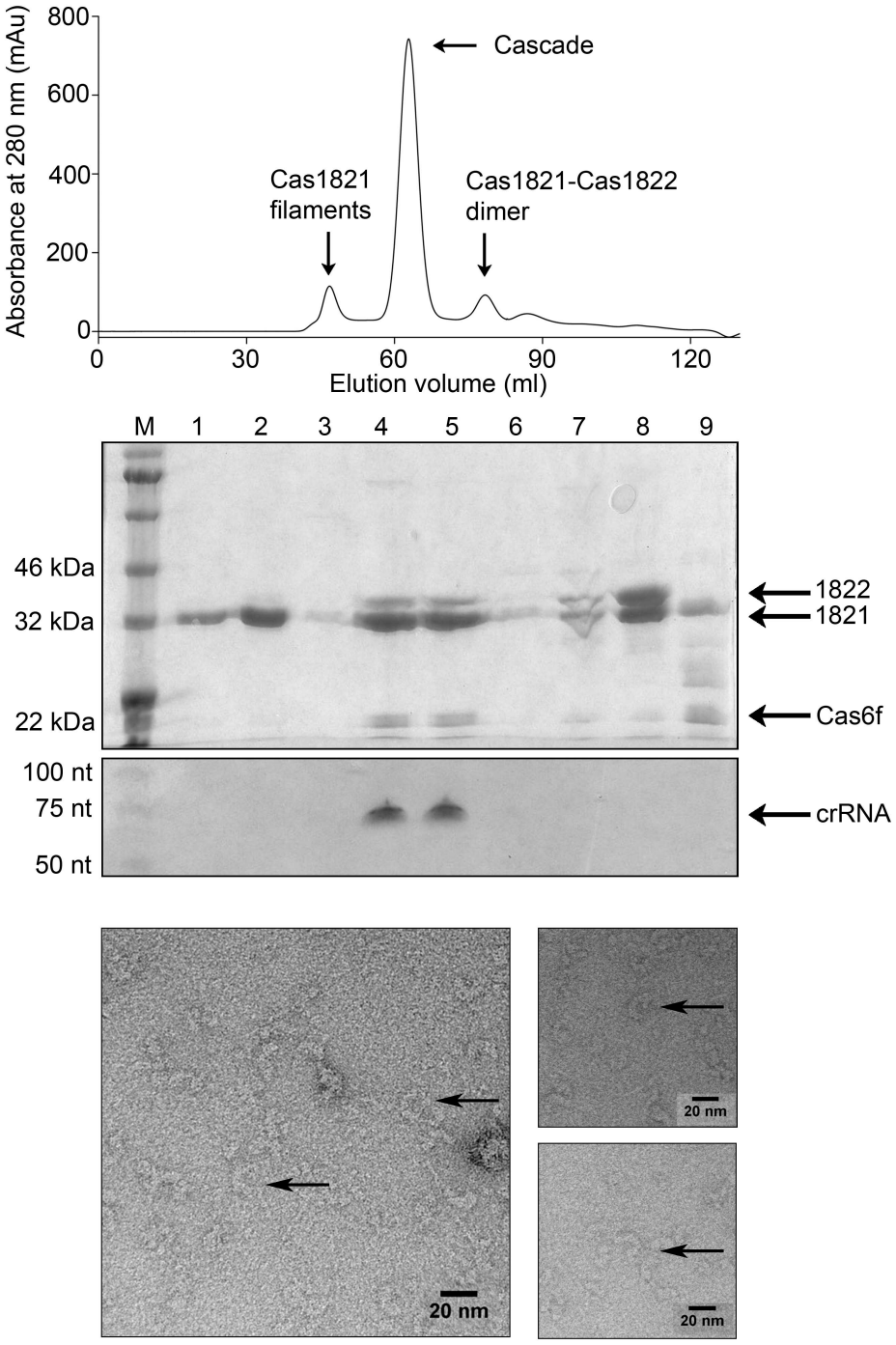


Figure 2.21: Reconstitution of a recombinant subtype I-F variant Cascade complex. His-tagged Cas1821 co-eluted with Cas1822, Cas6f and mature crRNA during gel-elution chromatography (fractions 4 and 5), which verified Cascade complex formation. A crescent-shaped structure for the complex was observed via TEM (indicated by arrows in the lower panel, provided by Dr. Thomas Heimerl). Cas1821 filaments were observed in the void volume (fractions 1 and 2). Dimers of Cas1821 and Cas1822 (fraction 8) could be identified. SDS-PAGE (middle) and UREA-PAGE (bottom) was used to separate the protein and RNA content of the fraction indicated in the gel-elution chromatogram (top).

3 Discussion

3.1 The minimal Type I-F variant CRISPR-Cas system is active

Shewanella putrefaciens CN-32 contains a minimal Type I-F variant CRISPR-Cas system that encodes a single CRISPR array with 81 spacers and five *cas* genes. Bioinfomatic analyses of this CRISPR-Cas system highlighted features that were shared with the Type I-F system. Only three of the five encoded Cas proteins have homologs in other Type I-F CRISPR-Cas systems: Cas1 (Sputcn32_1819), Cas3 (Sputcn32_1820) and Cas6f (Sputcn32_1820). The two additional protein Cas1821 (Sputcn32_1821) and Cas1822 (Sputcn32_1822) shared no detectable homology with other known Cas protein families. Accordingly, phylogenetic analyses with Cas7 and Cas5 protein families grouped these proteins and its homologs into a separate cluster. However, these analyses revealed an evolutionary relationship between Cas1821 and the Type I-F Cas7 protein family (Appendix 1). The repeat sequence of the *S. putrefaciens* CRISPR array is pseudo-palindromic and predicted to form a stem-loop structure, which was also observed for other Type I-F systems (26,58). However, the investigated *S. putrefaciens* CRISPR-Cas system displays significant differences to the standard Type I-F *cas* gene operon organization. First, the three essential Cascade subunits of the Type I-F system (Csy1-3) are absent. Instead, the two hypothetical proteins Cas1821 and Cas1822 are present. Second, a gene encoding the large subunit protein that is necessary for target DNA binding as well as PAM recognition could not be identified in the variant system. Nevertheless, the system was shown to be active *in vivo*. RNA-Seq analyses of isolated small RNAs identified all 81 crRNAs from the CRISPR array. The mature crRNAs contain the standard 8 nt 5′ repeat tag and an untrimmed 3′ repeat tag with a possible stable hairpin structure. The identified Cas6f enzyme processed pre-crRNA transcripts in *in vitro* cleavage assays. Bioinformatic analyses identified this protein as a Cas6f variant. In addition to sharing 37% identity with Cas6f from *Pseudomonas aeruginosa*, a conserved histidine at position 29 was identified and found to be essential for the endonuclease activity (59). The *P. aeruginosa* Cas6f was shown to remain associated with the hairpin at the 3′ end of the crRNA after cleavage (26,43,59). This observation was also reported for Cas6e and Cas5d from *Escherichia coli* and *Bacillus halodurans* (50,52,94). A similar association of *S.*

putrefaciens Cas6f with the crRNAs could also explain the absence of trimming at the 3′ terminal.

The Type I-F variant CRISPR-Cas system exhibited specific targeting of plasmid DNA. Using a conjugation-based *in vivo* interference assay, the targeting of a plasmid carrying a PAM sequence and protospacer was observed. The targeting ability of the system depended on *cas3*, *cas1821* and *cas1822*, as the DNA interference activity was abolished in the corresponding deletion strains. The absence of interference activity in *cas1821* and *cas1822* deletion strains corresponded with the loss of an observable cellular crRNA pool. Analyses of *E. coli* and *Haloferax volcanii* Cascade structures showed that Cas5 and Cas7 proteins are essential for crRNA maintenance and stability (95,96). Thus, Cas1821 and Cas1822 could potentially fulfill the roles of these two proteins. The non-specific RNA binding of Cas1821 led to the formation of oligomeric structures that could be visualized as helical filaments using electron microscopy. The formation of helical filaments was previously reported for Cas7 from *Thermoproteus tenax*, a sub-complex of Cas7 and Cas5a subunits from *Sulfolobus solfataricus* and for Type I-E Cascade in *E. coli* (34,38,93). Thus, it was concluded that Cas1821 subunits build the helical backbone of the Cascade structures and fulfill the role of Cas7 proteins. Accordingly, a recombinant variant Type I-F Cascade could be purified from *E. coli*, containing multiple copies of Cas1821, Cas1822, Cas6f and a mature crRNA.

The genes encoding small and large subunits proteins are missing in this Type I-F variant CRISPR-Cas system. The effect of the loss of these subunits on DNA interference was studied using the conjugation assay. It was observed that this system targets dsDNA and depends on the presence of a PAM sequence as expected for Type I CRISPR-Cas activity (23,97). The PAM sequence was found to be GG and mutations in this sequence resulted in the loss of interference activity. This PAM sequence is shared with subtype I-F CRISPR-Cas systems. Additionally, as observed for other Type I subtypes, mutations in the seed sequence were not tolerated and resulted in the loss of the interference activity (41,97). These results raised the question of orphan Cas proteins existing in the genomes of organisms containing the subtype I-F variant, which could fulfill the function of small and large subunits. It should be noted that an apparent small subunit is also missing for subtype I-F CRISPR-Cas systems. The subtype I-F variant CRISPR-Cas system was identified in nine mesophilic beta- and gamma-proteobacteria: *S. putrefaciens*

(strains CN-32 & 200), *O. ureolytica* (strain DSM 18253), *P. tunicata* (strain D2), *L. pneumophila* (strain 2300/99 Alcoy), *O. formigenes* (strain OXCC13), *P. profundum* (strain SS9), *V. cholerae* (strain TM 11079-80) and *M. nitratireducenticrescens*. Phylogenetic profiling of these genomes was performed to identify proteins that can substitute for the large subunit. This is a bioinformatics technique based on the idea that proteins that are functionally related are gained or lost together from genomes during evolution. Using *S. oneidensis* MR1 as an outgroup, the analyses identified the five Cas proteins to be conserved with > 95% identity in all identified genomes containing this variant Type I-F system. However, no additional Cas proteins could be identified that co-evolved with this minimal system. A homolog of a DinG helicase was identified downstream of the CRISPR array, which was reported to be part of Type I-U CRISPR-Cas systems (35). However, this enzyme is also found in related *Shewanella* strains without CRISPR systems or with different CRISPR-Cas subtypes. A deletion of the *DinG* gene did not abolish the interference activity indicating that the encoded protein is not required for subtype I-F variant Cascade activity. To date, RNase III and PNPase are the only non-Cas proteins that can be involved in conferring CRISPR-Cas based immunity (28,98). Both proteins play a key role in crRNA generation and are not involved in DNA interactions. In conclusion, no evidence for any additional Cascade components that are not encoded in the minimal *cas* gene operon was observed.

During interference, the large subunit of *E. coli* Type I-E Cascade interacts with the PAM motif and recruits Cas3 while both, the small and the large subunits are involved in the stabilization of the crRNA-target hybrid (38,67). The subtype I-F CRISPR-Cas systems lack a small subunit protein. A large subunit (Csy1) is present and contains eight predicted alpha helices at the C-terminus and four helices at the extreme N- terminus that is speculated to be a domain homologous to the small subunit (35). This protein was reported to form a dimer with Csy2 (Cas5) and bind to the 5′ end of the crRNA during the assembly of the Type I-F Cascade (62). The subtype I-F Cascade consists of Csy1 (large subunit), Csy2 (Cas5), Csy3 (Cas7) and Csy4 (Cas6f) and was shown to recognizes the double stranded G-C/G-C PAM (99). It is possible that either Cas1822 and/or Cas3 are involved in the recognition of the PAM sequence of the target DNA. The mutation of conserved arginine residues in Cas1822 influenced DNA interference activity even in the presence of a stable cellular crRNA pool. Phylogenetic analyses of Cas5 protein families revealed an

evolutionary relationship between Cas1822 and Cas5d (Appendix 2). Cas5d is the enzyme responsible for crRNA biogenesis in the Type I-C CRISPR-Cas system and was found to interact non-specifically with dsDNA (63). These results hint that Cas1822 might harbor additional functions. Alternatively, Cas3 could interact with the dsDNA during interference and could provide additional DNA target selectivity. It is interesting to note that the subtype I-F and subtype I-F variant Cas3 enzymes contain an N-terminal portion that resembles Cas2 (62). Only this region is conserved between Cas3 enzymes found in different *Shewanella* strains, while the larger C-terminal portion shows no apparent homology. Thus, novel Cas3 functions could be encoded in this part. It should be noted that the Type I-E and I-F Cascade, which were purified without Cas3 bind specifically to dsDNA targets that contain the type-specific PAM sequence (38,62,99). In both Types, Cas3 recruitment takes place after the recognition of a DNA target and requires structural changes in the large subunit (67). A recombinant I-F variant Cascade complex containing only Cas1821, Cas1822 and Cas6f can be purified with bound crRNA. If this I-F variant Cascade employs a DNA targeting mechanism similar to the I-E and I-F Cascade systems, wherein Cas3 is recruited after DNA target recognition, it is plausible that Cas1822 might be responsible for PAM recognition. The production of pure and soluble recombinant Cas1822 and Cas3 proteins was not successful and hence *in vitro* DNA binding experiments with the I-F variant Cascade are necessary to decipher the molecular details of PAM recognition mechanism in this variant I-F CRISPR-Cas system.

Previous publications on the Type III CRISPR-Cas systems describe *in vitro* assays that show crRNA-guided RNA cleavage and genetic analyses that demonstrated DNA targeting (40,100). However, a molecular mechanism that explains these different targeting activities is not obvious. In a recent publication, the cleavage of target DNA and its transcript during transcription was reported (78). It was shown that the cleavage activity was mediated by two independent active sites located on different Cas proteins of the Type III crRNP complex. The palm domain of Cas10 was shown to cleave only the target DNA strand of a dsDNA target while it is transcribed. Self and non-self discrimination was observed based on the complementarity of the 5′ repeat tag of the crRNA to the degraded target strand. In contrast, an autoimmune detection mechanism was not identified for the RNA target cleavage that was carried out by backbone protein Csm3 and not by Cas10. The

dependence of the DNA interference activity on the presence of PAM sequence eliminates the possibility of a similar target recognition mechanism in the I-F variant system. However, it is plausible that the system employs a similar target DNA cleavage mechanism. The *in vivo* assay used to demonstrate the interference activity was based on plasmid conjugation. During conjugation only one DNA strand of a plasmid is transferred from the donor to the recipient and replicated into circular dsDNA in the recipient. Although the Type I-F variant system targets dsDNA, the *in vivo* assay is limited and cannot distinguish between the cleavage of ssDNA or dsDNA. Therefore, it is possible that this variant system can only target the ssDNA that is formed during the replication or transcription of the plasmid, as observed for the Type III system. This question can be addressed by performing *in vitro* interference experiments with the recombinant Type I-F variant Cascade complex.

3.2 crRNA abundance in *S. putrefaciens* CN-32

The RNA-Seq analyses indicated highly variable crRNA abundance patterns. The general crRNA abundance was lower than found for other subtypes, e.g. subtype I-B (*M. maripaludis)* or subtype I-A (*T. tenax*) (25,34). Usually, the most abundant crRNAs are the ones closest to the promoter in the CRISPR leader regions (25,34,101). These crRNAs contain the most recently acquired spacers. The variability of crRNA abundance could be explained by the presence of anti-crRNA sequences and internal promoters, which were observed in other cases (48). These scenarios are not apparent for the *S. putrefaciens* CRISPR array. It is plausible that the Cascade loading efficiency differs for different crRNAs. Additionally, pre-crRNA could form internal structures that would influence Cas6f cleavage site availability.

3.3 Regulation of CRISPR-Cas systems

A possible regulation of the minimal CRISPR-Cas system in *S. putrefaciens* was identified. Studies in *E. coli* and *Salmonella typhii* reported the repression of transcription of the *cas* operon by a global transcriptional repressor, the Heat-stable Nucleoid Structuring protein (H-NS) (92,102). In both organisms, H-NS bound to the upstream and downstream regions of the transcription initiation site of Cse1 promoter, resulting in a repression of Cascade transcription. The deletion of the *H-*

NS gene in *S. putrefaciens* resulted in a two-fold change in the *cas* gene transcript levels. Accordingly, an increase in interference activity of the Δ*H-NS* strain compared to the wild-type strain was observed. It is interesting to note that complex and multilayered regulatory mechanisms are observed for CRISPR-Cas systems of many bacteria. For, *E. coli* and *S. typhii*, a complex regulatory mechanism of the CRISPR-Cas system involves H-NS, the global regulator Leucine-responsive Regulatory Protein (Lrp) and a global transcriptional activator LeuO (92,102,103). Additionally a secondary regulatory mechanism consisting of RcsB and BglJ was reported to influence the transcriptional activation of the Cascade operon in *E. coli* (104). A recent publication demonstrated the regulation of both the acquisition and interference mechanisms for the Type I-F system by CRP-cAMP and GalM in *Pectobacterium atrosepticium* (105). Regulation of CRISPR-Cas systems is however not limited only to bacteria. The CRISPR-associated proteins, Csa3 and Csx1 have been found to regulate the CRISPR-Cas systems in archaea (106). These types of complex regulatory mechanisms could be essential to guarantee an appropriate defense response with precise timing against viral attacks. In some bacteria, CRISPR-Cas systems perform additional unconventional functions other than the immune response. The production of Cas9 in *Campylobacter jejuni* lacking CRISPR loci has been demonstrated to increase virulence (107) whereas in *L. pneumophila* Cas2 is required for the infection of amoebae in a CRISPR-independent manner (108). In *P. aeruginosa* and *Myxococcus xanthus*, the corresponding CRISPR-Cas systems are involved in the inhibition of biofilm formation and in exopolysaccharide production, respectively (109,110). Although the exact mechanisms behind these additional functions are not always clear, they are driven by both environmental (in case of virulence, cell morphology) and cytoplasmic (in case of gene regulation) stimuli. Therefore, the understanding of these sophisticated regulatory mechanisms and stimuli should provide insights into the evolution and maintenance of CRISPR-Cas systems in prokaryotic genomes. In *S. putrefaciens CN-32*, genes encoding the small and the large subunit proteins could not be identified. For the other Type I CRISPR-Cas systems, these proteins play a crucial role during DNA interference and are responsible for the target DNA recognition and the stabilization of crRNA-DNA hybrid (34,38,39,66,67,111,112). The absence of these proteins in *S. putrefaciens* CN-32 may reduce the specificity of the DNA targeting ability of the system and could result in off-target cleavage events. Therefore, in addition to

triggering a precisely timed immune response, regulation of *S. putrefaciens* CN-32 CRISPR-Cas system could be necessary to avoid toxicity from off-target cleavage events.

3.4 Assembly of the Cascade complex

In Cascade complexes of the Type I-C, I-E and I-F CRISPR-Cas systems, the corresponding Cas6 enzymes also remain associated with the 3′ hairpin after crRNA generation. The stable association of Cas6 to the 3′ hairpin of the crRNA is suggested to serve as the starting point of Cascade assembly (51). However, the order of the Cascade backbone assembly is unclear. It is not known if Cas7 oligomerization begins or terminates at one of the two termini of the crRNA. Recent publications indicate that the *E. coli* and *H. volcanii* CRISPR-Cas systems can target DNA with crRNAs generated in a Cas6-independent manner (113,114). These reports, together with observation that the removal of the 3′ repeat handle has no effect on the target recognition and R-loop formation *in vitro*, suggests that Cascade assembly might initiate at 5′ end of the crRNA (115). Dimers of Cas7 and Cas5a were observed in the Type I-A system from *S. solfataricus* and suggested to form the initial building blocks of Cascade (93). The identification of Cas1821 and Cas1822 dimers during the purification and reconstitution of recombinant *S. putrefaciens* Cascade supports this hypothesis. The recombinant *S. putrefaciens* Cascade contains Cas1822, Cas6f, multiple copies of Cas1821 and a crRNA. The multiple copies of Cas1821 are proposed to form the backbone of Cascade and bind to the crRNA, while Cas1822 and Cas6f might protect the 5′ and 3′ end of the crRNA respectively. The recombinant I-F variant Cascade complex can be produced in *E. coli* using two compatible plasmids, one for the production of the Cas proteins and one for production of pre-crRNA. This can serve as the starting point for designing experiments that address the question of the order of Cascade assembly. Second, the constraints placed by the crRNA on the structure of Cascade can be tested. Finally, structural studies using X-ray crystallography techniques could be carried out for the Type I-F variant complex.

The visualization of the Type I-F variant minimal Cascade using electron microscopy revealed a crescent shaped architecture of the complex. The general shape is conserved in crRNP complexes from both Type I and Type III CRISPR-Cas

systems even though the constituting Cas proteins, as well as the nucleic acid targets vary significantly (64,68). This universal layout is a result of the binding of crRNAs by proteins of the Cas7 superfamily (e.g., Cas7, Cmr4, and Csm3). In all of these crRNP complexes, oligomerization of Cas7 results in a crRNA that is kinked at six-nucleotide intervals caused by the interaction of the thumb domain of one backbone subunit with the palm domain of the adjacent subunit (38,39,116). During the hybridization of the crRNA with a complementary target, five nucleotides of each interval base pair with the target, while the thumb domain of Cas7 blocks the base pairing of the sixth nucleotide. This results in the kinking of the target backbone and, in Type I systems, the destabilization of dsDNA and the initiation of R-loop formation (38,39). A similar dsDNA binding mechanism is probably observed in the Type I-F variant system. However, further structural and biochemical studies of the minimal Cascade are necessary to confirm this hypothesis. Although the targeting mechanism of the Type I and Type III systems are distinct, the structural and functional similarities observed for the surveillance complexes of these systems suggest an evolutionary connection (35,117). It was proposed that the modern CRISPR-Cas systems evolved form an ancestral Type III-A like system (17,35). This system probably consisted of six or seven genes including the two universal Cas proteins, Cas1 and Cas2, (for spacer acquisition) along with four or five additional proteins (crRNP complex) involved in crRNA processing and interference. The crRNP complex would include the large subunit (Cas10/Cas8), the small subunit and two or three RNA binding backbone proteins. The evolution of Type I systems might be a result of the inactivation of the Cas10 palm domain, which led to the evolution of Cas8 and the acquisition of Cas3 helicase. In contrast to the Type III crRNP functionality, the acquisition of Cas3 helicase equipped Type I systems to unwind dsDNA without depending on the host RNA polymerase activity. This event probably triggered the switch from the targeting of both DNA and RNA, as observed in Type III systems, to the exclusive targeting of dsDNA and the evolution of PAM sequences as autoimmune detection mechanisms. Compared to the Type I and Type III systems, the Type II CRISPR-Cas systems are distinct. This system is considered to have evolved recently, based on the presence of two unrelated nuclease domains within a single Cas9 protein that performs the function of both, CRISPR processing and interference. Nevertheless, experimental data on Cas9 activity indicates functional similarity to the crRNP complexes of Type I and Type III systems (46). The

loss of large and small subunits in the minimal Type I-F variant system might coincide with the evolution of specialized minimal type I Cascade systems that still rely on PAM motif recognition for DNA target specificity.

4 Material and methods

4.1 Chemicals, enzymes and consumables

Table 1: List of chemicals, enzymes, kits and consumable used in this study.

Chemicals	Company
Acrylamide, N,N-methylenebisacrylamide	Roth GmbH, Karlsruhe
Alkaline phosphatase, calf intestinal (HC)	Fermentas GmbH, St. Leon-Rot
Kanamycin, ampicillin, spectinomycin, chloramphenicol	Sigma-Aldrich, Taufkirchen ; Roth GmbH, Karlsruhe
Bradford Reagent	BioRad, München
Bovine serum albumin (BSA)	Sigma-Aldrich, Taufkirchen
Coomassie brilliant blue R and G	SERVA GmbH, Heidelberg
Dialysis tubes	Roth GmbH, Karlsruhe; Serva GmbH, Heidelberg
Diethyl pyrocarbonate (DEPC)	AppliChem GmbH, Darmstadt
DNA ladder Mix	New England Biolabs, Frankfurt
DNA ladder, ultra low range	New England Biolabs, Frankfurt
Phusion DNA polymerase	Fermentas GmbH, St. Leon-Rot
Taq DNA polymerase	New England Biolabs, Frankfurt
DNase I	Sigma-Aldrich, Taufkirchen
dNTP Mix	New England Biolabs, Frankfurt
Isopropyl-ß-thiogalactopyranoside (IPTG)	Roth GmbH, Karlsruhe
*mir*vana™ miRNA Isolation Kit	Applied Biosystems, Darmstadt
Ni-NTA agarose	Qiagen, Hilden ; Life Technologies, Karlsruhe
Oligonucleotides	Eurofins MWG operon, Ebersberg
Color prestained protein standard, 11-245 kDa	New England Biolabs, Frankfurt
Color prestained protein standard, 10-250 kDa	New England Biolabs, Frankfurt
Roti®-Nylon plus, nylon membrane	Roth GmbH, Karlsruhe
Qiagen Plasmid-Kit, QIAquick Gel Extraction Kit, QIAquick Nucleotide Removal Kit, QIAquick PCR Purification Kit	Qiagen, Hilden
Restriction endonucleases	New England Biolabs, Frankfurt
Reverse transcriptase	New England Biolabs, Frankfurt
RNA ladder, high range	New England Biolabs, Frankfurt
RNase-ExitusPlus™	AppliChem GmbH, Darmstadt
Sodium dodecyl sulfate	SERVA GmbH, Heidelberg
T4 DNA Ligase	New England Biolabs, Frankfurt
T7 RNA Polymerase	New England Biolabs, Frankfurt
TRIzol Reagent	Life Technologies, Karlsruhe
Whatman GB 004, 3MM	Schleicher & Schuell GmbH, Dassel
T4 polynucleotide kinase	New England Biolabs, Frankfurt; Life technologies, Karlsruhe
4-(2-aminoethyl)benzenesulfonylfluoride hydrochloride	Sigma-Aldrich, Taufkirchen
Protease cocktail	Roth GmbH, Karlsruhe
Lysozyme	Sigma-Aldrich, Taufkirchen
iQ SYBR Green Supermix Kit	Bio-Rad, München

The chemicals and consumables that are not listed above were purchased from AppliChem GmbH (Darmstadt), BioRad Laboratories GmbH (München), Biozym GmbH (Hessisch-Oldendorf), Difco Laboratories GmbH (Augsburg), Invitrogen (Karlsruhe), Merck KGaA (Darmstadt), Roche GmbH (Mannheim), Roth GmbH (Karlsruhe), VWR International (Darmstadt), SERVA GmbH (Heidelberg) and Sigma-Aldrich Co. (Deisenhofen).

4.2 Strains and culture conditions

4.2.1 Strains used

Table 2: List of bacterial strains used in this study.

Strain	Description	Source
Shewanella putrefaciens CN-32, ATCC BAA-453D-5		Fredrickson *et al.*, 1998 (118)
Escherichia coli DH5α	FhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ) M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	Novagen
Escherichia coli DH5α λpir	sup E44, ΔlacU169 (ΦlacZΔM15), recA1, endA1, hsdR17, thi-1, gyrA96, relA1, λpir phage lysogen	Miller *et al.*, 1988 (119)
Escherichia coli WM3064	thrB1004 pro thi rpsL hsdS lacZΔM15 RP4-1360 Δ(araBAD)567 ΔdapA1341::[erm pir]	W. Metcalf, University of Illinois
Escherichia coli BL21(DE3) pLsyS	F-ompT hsdSB(rB- mB-) gal dcm (DE3) pLysS ($Spec^R$)	Novagen

4.2.2 *S. putrefaciens* CN-32 strains

Table 3: List of *S. putrefaciens* CN-32 strains generated during this study.

Strain	Description	Source
Δ*cas1*	Δ*Sputcn32_1819*	AG Thormann
Δ*cas2-cas3*	Δ*Sputcn32_1820*	AG Thormann
Δ*1821*	Δ*Sputcn32_1821*	AG Thormann
Δ*1822*	Δ*Sputcn32_1822*	AG Thormann
Δ*H-NS*	Δ*Sputcn32_2512*	This work
Δ*DinG*	Truncation in *Sputcn32_1824*	This work
Δ*H-NS, cas2-cas3*	Δ*2512*, Δ*cas1820*	This work
Δ*H-NS, 1822*	Δ*2512*, Δ*cas1822*	This work
Δ*H-NS, DinG*	Δ*2512,* truncation of *Sputcn32_1824*	This work
Δ*H-NS, 1822(R29A)*	Δ*2512*, R29A mutation in Cas1822	This work
Δ*H-NS, 1822(R60A)*	Δ*2512*, R60A mutation in Cas1822	This work
Δ*H-NS, 1822(R66A)*	Δ*2512*, R66A mutation in Cas1822	This work
Δ*H-NS, 1822(R170A)*	Δ*2512*, R170A mutation in Cas1822	This work
Δ*H-NS, 1822(R225A)*	Δ*2512*, R225A mutation in Cas1822	This work

Δ*H-NS, 1822(R258A)*	Δ*2512*, R258A mutation in Cas1822	This work
Δ*H-NS, 1822(E86A, D87A)*	Δ*2512*, E86A and D87A mutations in Cas1822	This work
Δ*H-NS, Cas1821-His-HA*	Δ*2512*, Cas1821 produced with N-terminal His-HA tag	AG Thormann
Δ*H-NS, Cas1822-His-HA*	Δ*2512*, Cas1822 produced with N-terminal His-HA tag	AG Thormann
Δ*H-NS, Cas6f-His-HA*	Δ*2512*, Cas6 produced with N-terminal His-HA tag	AG Thormann

4.2.3 Culture condition for *S. putrefaciens* CN-32

Aerobic growth of *S. putrefaciens* CN-32 strains were carried out at 30°C using Lysogeny Broth (LB) media (1% bacto-tryptone, 0.5% yeast extract, 1% NaCl (w/v), pH 7.0) in Erlenmeyer flask with constant agitation in a rotary shaker (200 rpm) or on solid medium plates (LB containing 1.5% (w/v) agar). The medium was supplemented with 50 mg/ml kanamycin (kan), 300 µM 2,6-diaminopimelic acid (DAP) or 10 % sucrose (v/v) when required.

4.2.4 Culture conditions for *E. coli*

Aerobic growth of *E. coli* was carried out at 37°C using LB medium or NZA medium (1% NZ-Amine, 0.5% yeast extract, 1% NaCl (w/v), pH 7.0) in Erlenmeyer flasks accompanied by shaking in a rotary shaker or on solid LB media plates. Growth was monitored at OD_{600} after inoculating the media supplemented with appropriate antibiotics with an overnight pre-culture (2% v/v). Protein production was induced by the addition of IPTG (1 mM final concentration) when OD_{600} of 0.6-0.8 was achieved. Cells were harvested by centrifugation (4000 rpm, 20 min, 4°C) after overnight incubation at 18°C or incubation for 3 h at 37°C and stored at -80°C. *E. coli* BL21(DE3) pLysS strain was used for heterologous production of *S. putrefaciens* CN-32 proteins using the plasmids listed in Table 5. *E. coli* strains DH5α and DH5α λpir were used for cloning and storage of plasmids whereas *E. coli* WM3064 was used during conjugation.

4.3 Primers and constructed recombinant vectors

Table 4: Plasmids used in this study.

Vector	Resistance	Application
pUC19	Amp^R	*In vitro* transcription
pBRR1MCS2	Kan^R	Cloning
pET-20b	Amp^R	Protein production
pRSF-Duet	Kan^R	Protein production
pETM-43	Kan^R	Protein production
pEC-His-A	Amp^R	Protein production
pCPD	Amp^R	Protein production
pCDF-Duet	$Spec^R$	Protein production
pNPTS138-R6KT	Kan^R	In-frame insertion or gene deletion

Table 5: List of recombinant vectors constructed in this study.

Plasmid + Insert	Description
pET-20b + *cas1*	Cas1 with a 6x His-tag (C-terminal)
pET-20b + *cas3*	Cas3 with a 6x His-tag (C-terminal)
pET-20b + *cas1821*	Cas1821 with a 6x His-tag (C-terminal)
pET-20b + *cas1822*	Cas1822 with a 6x His-tag (C-terminal)
pET-20b + *cas6*	Cas6f with a 6x His-tag (C-terminal)
PETM-43 + *cas3*	Cas3 with a 6x His-maltose binding protein-tag (N-terminal)
pEC-His-A + *cas3*	Cas3 with a 6x His-SUMO-tag (N-terminal)
pCPD + *cas3*	Cas3 with a cysteine protease domain (CPD)-6x His- tag (C-terminal)
pEC-His-A + *cas1821*	Cas1821 with a 6x His-SUMO-tag (N-terminal)
pET20b + *cas1821* (no tag)	Cas1821 without a tag
pEC-His-A + *cas1822*	Cas1822 with a 6x His-SUMO-tag (N-terminal)
pCPD + *cas1822*	Cas1822 with a CPD 6x His- tag (C-terminal)
pET20b + *cas1822* (no tag)	Cas1822 without a tag
pCDF-Duet + *cas1821, cas1822*	Cas1822 with a 6x His-tag (N-terminal) and Cas1821 without a tag
pEC-His-A + *cas3* (H156A)	Cas3 with a 6x His-SUMO-tag (N-terminal) carrying a H156A mutation
pCPD + *cas3* (H156A)	Cas3 with a CPD-6x His- tag (C-terminal) carrying a H156A mutation
pCPD + *cas3* (D157A)	Cas3 with a CPD-6x His- tag (C-terminal) carrying a D157A mutation
pET-20b + *cas6* (H29A)	Cas6f with a 6x His-tag (C-terminal) carrying a H29A mutation
pRSF-Duet 567	pRSFduet1 carrying *cas1821*, *cas1822* and *cas6* as an operon with an N-terminal 6x His-tag on Cas1821
pUC19 +spacer3-repeat-spacer4	pUC19 carrying a spacer3-repeat-spacer4 fragment
pUC19 + crRNA4	pUC19 carrying a repeat-spacer4-repeat (partial) fragment
pUC19 + hammerhead crRNA1	pUC19 carrying a hammerhead ribozyme and repeat-spacer1-repeat (partial) fragment
pUC19 + hammerhead crRNA4	pUC19 carrying a hammerhead ribozyme and repeat-spacer4-repeat (partial) fragment
pNPTS138-R6KT-Δ*H-NS*	*H-NS* deletion fragment in pNPTS138-R6KT
pNPTS138-R6KT-Δ*DinG*	*DinG* deletion fragment in pNPTS138-R6KT

pNPTS138-R6KT-*cas1822*	*cas1822* and 500 bp upstream and downstream fragment (3011 bp) in pNPTS138-R6KT
pNPTS138-R6KT- *cas1822*-R29A	Fragment carrying the R29A mutation of *cas1822* in pNPTS138-R6KT
pNPTS138-R6KT- *cas1822*-R60A	Fragment carrying the R60A mutation of *cas1822* in pNPTS138-R6KT
pNPTS138-R6KT- *cas1822*-R66A	Fragment carrying the R66A mutation of *cas1822* in pNPTS138-R6KT
pNPTS138-R6KT- *cas1822*-R170A	Fragment carrying the R170A mutation of *cas1822* in pNPTS138-R6KT
pNPTS138-R6KT- *cas1822*-R225A	Fragment carrying the R225A mutation of *cas1822* in pNPTS138-R6KT
pNPTS138-R6KT- *cas1822*-R258A	Fragment carrying the R258A mutation of *cas1822* in pNPTS138-R6KT
pNPTS138-R6KT- *cas1822*-E86A,D87A	Fragment carrying the E86A and D87A mutations of *cas1822* in pNPTS138-R6KT
pBBR1MCS2-sp1-GG	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 1 with the ´GG´ PAM on the 3´end of the target strand
pBBR1MCS2-sp3-GG	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 3 with the ´GG´ PAM on the 3´end of the target strand
pBBR1MCS2-sp13-GG	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 13 the ´GG´ PAM on the 3´end of the target strand
pBBR1MCS2-sp15-GG	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 15 with the ´GG´ PAM on the 3´end of the target strand
pBBR1MCS2-sp20-GG	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 20 with the ´GG´ PAM on the 3´end of the target strand
pBBR1MCS2-sp34-GG	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 34 with the ´GG´ PAM on the 3´end of the target strand
pBBR1MCS2-sp13rev-GG	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 13 in the opposite orientation than before
pBBR1MCS2-sp13-AA	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 13 with the ´AA´PAM on the 3´end of the target strand
pBBR1MCS2-sp13-GA	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 13 with the ´GA´PAM on the 3´end of the target strand
pBBR1MCS2-sp13-AG	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 13 with the ´AG´PAM on the 3´end of the target strand
pBBR1MCS2-sp13-CC	pBBR1MCS2 containing *S. putrefaciens* CN-32 spacer 13 with the ´CC´PAM on the 3´end of the target strand
pBBR1MCS2-sp13-polyA	pBBR1MCS2 containing *poly A* as the first 10 nucleotides of *S. putrefaciens* CN-32 spacer 13

Table 6: List of primers used in this study.

Primers	**Sequence 5′→3′**
Cloning of cas1 into pET-20b	
Forward	GGTACAGAACATATGATGGTGACATTC
Reverse	CATAAGCTTTTCCTCATCTTCAGTGG
Cloning of cas3 into pET-20b	
Forward	GTACAGAAGCTCATATGATGGTGACATTCA
Reverse	TACTTTTTGCATCTCGAGTTCCTCATCTTC
Cloning of cas1821 into pET-20b	
Forward	AGATGAGGAATAACATATGCAAAAAGTAAC

Reverse	TTCTTGCAACCTCGAGTTTTGCAT
Cloning of cas1822 into pET-20b	
Forward	GCAAAATAGGAGGTCATATGAAAATAATC
Reverse	CATCATTGGTCGACAAGCTTAATGTTT
Cloning of cas6 into pET-20b	
Forward	TTTAGGACATATGATGAACTCATATATTGATA
Reverse	AAGAACAAAAATAAACTCGAGAAACCAAGG
Cloning of cas3 into pETM-43	
Forward	AGAATTCATGATGGTGACATTCATCAGTCAGTGCGAG
Reverse	GCTCGAGTTCCTCATCTTCAGTGGTTTGATTAATT
Cloning of cas3 into pEC-His-A	
Forward	ACCAGGAACAAACCGGCGGCCGCTCGATGATGGTGACA TTCATCAGTCAGTGCGAG
Reverse	GCAAAGCACCGGCCTCGTTATTATTCCTCATCTTCAG
Cloning of cas3 into pCPD	
Forward	GTACAGAAGCTCATATGATGGTGACATTCA
Reverse	TACTTTTTGCATCTCGAGTTCCTCATCTTC
Cloning of cas1821 into pEC-His-A	
Forward	ACCAGGAACAAACCGGCGGCCGCTCGATGCAAAAAGTA ACGGGAATTAAAAGTGT
Reverse	GCAAAGCACCGGCCTCGTTACTATTTTGCATAAAAATACT GTGCAAATGGC
Cloning of cas1821 into pET-20b without a tag	
Forward	ATCATATGCAAAAAGTAACGGGAATTAAAA
Reverse	TTCTCGAGCTATTTTGCATAAAAATACTGTGC
Cloning of cas1822 into pEC-His-A	
Forward	ACCAGGAACAAACCGGCGGCCGCTCGATGAAAATAATC ATAGAATATGACTCTTGCTGGCG
Reverse	GCAAAGCACCGGCCTCGTTACTAAAGCTTAATGTTTGAT ACATAGGCCAG
Cloning of cas1822 into pCPD	
Forward	GCAAAATAGGAGGTCATATGAAAATAATC
Reverse	CATCATTGGTCGACAAGCTTAATGTTT
Cloning of cas1822 into pET-20b without a tag	
Forward	CCCATATGAAAATAATCATAGAATATGACTC
Reverse	TTCTCGAGCTAAAGCTTAATGTTTGATACATAGG
Cloning of cas1821 and cas1822 into pCDF-duet	
1822-Forward	GCAAAATAGGAGAATTCATGAAAATAATC
1822-Reverse	CATCATTGGTCGACAAGCTTAATGTTT
1821-Forward	AGATGAGGAATAACATATGCAAAAAGTAAC
1821-Reverse	TTCTTGCAACCTCGAGTTTTGCAT
Cloning of cas1821, cas1822 and cas6 as an operon into pRSF-duet	
Forward	AGCGGATCCGATGCAAAAAGTAACGGGAATTAAAAGTGT T
Reverse	GATTGCGGCCGCTTAAAACCAAGGTACTGTAGCGGTTTT GCT
Cloning of sp3-rep-sp4	
Forward	GGATCCTAATACGACTCACTATAGGGAGATATTGAGTTTG CATCAAGCA
Reverse	AAGCTTGGTTATAGGTTTGCGCGTCTTGCTGGGCGATA
In frame deletion of H-NS	
H-NS KO up fw	CGAATTCATACTGACAACAACCACTTCGGTA
H-NS KO up OL rv	CCTTTTACTGCGAGTTTGTATGGAATCTATTTTCGTCCCA CTTTAATTTTTAAT
H-NS KO dwn OL fw	ATTAAAAATTAAAGTGGGACGAAAATAGATTCCATACAAA CTCGCAGTAAAAGG

H-NS KO dwn rv	GGTTGGATCCGGTAAAGGGATCGATAAATCCGGG
H-NS check fw	TGCAGTACGTAGCTGTTCTTGTA
H-NS check rv	TAATATCATAAACAGGGCGTGTCAGT
In frame truncation of DinG	
DinG KO up fw	GCGGAATTCTATCTTAAATCAGCATAATGCTTCGAACTC
DinG KO up OL rv	TTGTCCGCCAATATTATTAACAAGATGTTTTCTTGCAGATT TTTACTGCCCGTACTAATG
DinG KO dwn OL fw	CATTAGTACGGGCAGTAAAAATCTGCAAGAAAACATCTTG TTAATAATATTGGCGGACAA
DinG KO dwn rv	CGCGGATCCACCCAACAATAAAGAAGGAAAATAATATGC
DinG check fw	GTAAATCTGAATAGCCGCACACGCGGCTTTAA
DinG check rv	ATACCTACATCATATTCACACAACATGCGC
Construction of in vivo His-HA tagged Cas1821	
1821 up BamHI fw	CGGATCCGAACACTTATACCTTAGCTGTTCC
1821 up OL rv	GCATTATCGATCAGATGGTGATGGTGATGGTGCATGATT TATTCCTCATCTTCAGTGG
1821 up OL fw	CCATCTGATCGATAATGCATACCCATACGATGTTCCAGAT TACGCTCAAAAAGTAACGGGAATTAAAAGTG
1821 dwn EcoRI rv	CGAATTCGCGTTCTAATGTTTCAGCAACC
Construction of in vivo His-HA tagged Cas1822	
1822 up BamHI fw	CGGATCCATTGCATTCGCCATCATTTATTTCG
1822 up OL rv	GCATTATCGATCAGATGGTGATGGTGATGGTGCATGCAA CCTCCTATTTTGCATAAAAATAC
1822 up OL fw	CCATCTGATCGATAATGCATACCCATACGATGTTCCAGAT TACGCTAAAATAATCATAGAATATGACTCTTGCTG
1822 dwn EcoRI rv	CGAATTCGTCCCGTTGTATATCGGTCC
Construction of in vivo His-HA tagged Cas6	
Cas6 up BamHI fw	CGGATCCTTATAGATAAGCCGCAGTTGACC
Cas6 up OL rv	GCATTATCGATCAGATGGTGATGGTGATGGTGCATGCAT CATTGCTCCTAAAGCTTAATG
Cas6 up OL fw	CCATCTGATCGATAATGCATACCCATACGATGTTCCAGAT TACGCTAACTCATATATTGATATACGGCTG
Cas6 dwn EcoRI rv	CGAATTCTGGTCGCGCAATTTATGATTTGG
In vivo mutations in Cas1822	
1822 fw	GAGGAATTCGGAGCAAGAGATAGTACATCATTTTTTTCA
1822 rv	AGCGGATCCTGAGCTTGAATGTCGCTAAACTCAAAGAAT
R29A fw	CATTGAGCCTAAAAATTCGGCGCCTTTTTTCGGTACAGGT
R29A rv	ACCTGTACCGAAAAAAGGCGCCGAATTTTTAGGCTCAAT G
R60A fw	TTCGTTGGTCGCCAATTAAAGCATTTAGTACTCCCATTAC GG
R60A rv	CCGTAATGGGAGTACTAAATGCTTTAATTGGCGACCAAC GAA
R66A fw	CGCGCCTGATAAAGCTTTGCTTGGTCGCCAATTAAACG
R66A rv	CGTTTAATTGGCGACCAAGCAAAGCTTTATCAGGCGCG
R170A fw	GATTAAGGCTCTCCAAGGCATTGATAATACTTAACGGGTT TAACTCAATA
R170A rv	TATTGAGTTAAACCCGTTAAGTATTATCAATGCCTTGGAG AGCCTTAATC
R225A fw	TATTAAAGCTAGTTTCTAAGGCGGCAAGCTGAAGGTAAA GCG
R225A rv	CGCTTTACCTTCAGCTTGCCGCCTTAGAAACTAGCTTTAA TA
R258A fw	TTTTTTGGGTCCCGTTGTATATGCGTCCATGAAGTCTTTT TTAGTG
R2258A rv	CACTAAAAAAGACTTCATGGACGCATATACAACGGGACC CAAAAAA

E86A D87A fw	GCTTATCTATAAAGCTGACTTTAGCTGCCAGTGCTTCGAAGTAATAAGCG
E86A D87A rv	CGCTTATTACTTCGAAGCACTGGCAGCTAAAGTCAGCTTTATAGATAAGC
Generation of Cas3 H156A mutant	
Forward	CGAATTTTATTGCAGGTTGCTTGGCTGATTTAGGGAAAATTGACCCAA
Reverse	TTGGGTCAATTTTCCCTAAATCAGCCAAGCAACCTGCAATAAAATTCG
Generation of Cas3 D157A mutant	
Forward	TTGCAGGTTGTTTGCATGCTTTAGGGAAAATTGACCC
Reverse	GGGTCAATTTTCCCTAAAGCATGCAAACAACCTGCAA
Generation of Cas6 H29A mutant	
Forward	GCTGAGCTCAGTAGTAAAGTATTTACTAAATTTGCGAAAGCGTTAGTGACGCTC
Reverse	GAGCGTCACTAACGCTTTCGCAAATTTAGTAAATACTTTACTACTGAGCTCAGC
Cloning of crRNA4	
Forward	GATCCTAATACGACTCACTATAGGGAGAGTTCACCGCCGCACAGGCGGCTTAGAAATATCGCCCAGCAAGACGCGCAAACCTATAACCGTTCACCGCCGCACAGGCGGA
Reverse	AGCTTCCGCCTGTGCGGCGGTGAACGGTTATAGGTTTGCGCGTCTTGCTGGGCGATATTTCTAAGCCGCCTGTGCGGCGGTGAACTCTCCCTATAGTGAGTCGTATTAG
Cloning of hammerhead crRNA1	
Forward1	GATCCTAATACGACTCACTATAGGGAGATAAGCTGATGAGTCCGTGAGGACGAAACGGTACCCG
Forward2	GTACCGTCCTTAGAAATCAACCAAATCATAAATTGCGCGACCCACATTGGTTCACCGCCGCACAGGCGGCCA
Reverse1	GACGGTACCGGGTACCGTTTCGTCCTCACGGACTCATCAGCTTATCTCCCTATAGTGAGTCGTATTAG
Reverse2	AGCTTGGCCGCCTGTGCGGCGGTGAACCAATGTGGTCGCGCAATTTATGATTTGGTTGATTTCTAAG
Cloning of hammerhead crRNA4	
Forward1	GATCCTAATACGACTCACTATAGGGAGATAAGCTGATGAGTCCGTGAGGACGAAACGGTACCCG
Forward2	GTACCGTCCTTAGAAATATCGCCCAGCAAGACGCGCAAACCTATAACCGTTCACCGCCGCACAGGCGGCCA
Reverse1	GACGGTACCGGGTACCGTTTCGTCCTCACGGACTCATCAGCTTATCTCCCTATAGTGAGTCGTATTAG
Reverse2	AGCTTGGCCGCCTGTGCGGCGGTGAACGGTTATAGGTTTGCGCGTCTTGCTGGGCGATATTTCTAAG
Cloning of sp1 with the ´GG´ PAM	
Forward	AATTCCCTCAACCAAATCATAAATTGCGCGACCACATTGG
Reverse	GATCCCAATGTGGTCGCGCAATTTATGATTTGGTTGAGGG
Cloning of sp3 with the ´GG´ PAM	
Forward	AATTCCCTATTGAGTTTGCATCAAGCACGGTATGGCGCTG
Reverse	GATCCAGCGCCATACCGTGCTTGATGCAAACTCAATAGGG
Cloning of sp13 with the ´GG´ PAM	
Forward	AATTCCCCGTTATGTCTGTTACACTGTCAACTTCCAATAG
Reverse	GATCCTATTGGAAGTTGACAGTGTAACAGACATAACGGGG
Cloning of sp15 with the ´GG´ PAM	
Forward	AATTCCCGTTAACAGTTGCTGGCAACGTCTTTTGCACCTG
Reverse	GATCCAGGTGCAAAAGACGTTGCCAGCAACTGTTAACGGG

Cloning of sp20 with the ´GG´ PAM	
Forward	AATTCCCGCAGGGACTCGCGCAAAGGCTGGCGTTATGAGG
Reverse	GATCCCTCATAACGCCAGCCTTTGCGCGAGTCCCTGCGGG
Cloning of sp34 with the ´GG´ PAM	
Forward	AATTCCCTGTTGGATTTGGCTCTTGGTTCGGCGCTTCGGG
Reverse	GATCCCCGAAGCGCCGAACCAAGAGCCAAATCCAACAGGG
Cloning of sp13 with the ´GG´ PAM in reverse	
Forward	AATTCTATTGGAAGTTGACAGTGTAACAGACATAACGGGG
Reverse	GATCCCCCGTTATGTCTGTTACACTGTCAACTTCCAATAG
Cloning of sp13 with the ´AA´ PAM	
Forward	AATTCTTCGTTATGTCTGTTACACTGTCAACTTCCAATAG
Reverse	GATCCTATTGGAAGTTGACAGTGTAACAGACATAACGAAG
Cloning of sp13 with the ´GA´ PAM	
Forward	AATTCTCCGTTATGTCTGTTACACTGTCAACTTCCAATAG
Reverse	GATCCTATTGGAAGTTGACAGTGTAACAGACATAACGGAG
Cloning of sp13 with the ´AG´ PAM	
Forward	AATTCCTCGTTATGTCTGTTACACTGTCAACTTCCAATAG
Reverse	GATCCTATTGGAAGTTGACAGTGTAACAGACATAACGAGG
Cloning of sp13 with the ´CC´ PAM	
Forward	AATTCGGCGTTATGTCTGTTACACTGTCAACTTCCAATAG
Reverse	GATCCTATTGGAAGTTGACAGTGTAACAGACATAACGCCG
Cloning of polyA containing sp13	
Forward	AATTCCCAAAAAAAAAAGTTACACTGTCAACTTCCAATAG
Reverse	GATCCTATTGGAAGTTGACAGTGTAACTTTTTTTTTTGGG
qPCR primers	
16S-fw	GCCTTATCTGTGACGTTACC
16S-rv	GCTTTACGCCCAGTAATTCC
1819-Cas1-fw	AAGACGCCATCGTATTACCC
1819-Cas1-rv	GGTCGGTAAAGGCTTGTAAC
1820-Cas2,3-fw	GGCATAGTGCCAGTGAATAG
1820-Cas2,3-rv	GTGGCTGTAATTGGGTGTTG
1821-fw	GGCGTAGTGAATTGGAATGG
1821-rv	CCGGTCAAGTTTGTGTAGC
1822-fw	TGGCAGTACGGATCAGAATG
1822-rv	GTGAAATCGAGAGCCAACAC
1823-fw	TATACGGCTGAAGCCTGATG
1823-rv	GGCTATTGAGCGTCACTAAC
Probes for northern blots	
Repeat sequence	GTCTTGCTGGGCGATATTTC
16S control	GCTTTACGCCCAGTAATTCC
In vitro interference assay primers	
Sp1 target	AAGCTTGAGGCCCCAAGGGGTTATGCTAGCAATGTGGTCGCGCAATTTATGATTTGGTTGAGGACTCCCTATAGTGAGTCGTATTAGGATCC
Sp1 non-target	GGATCCTAATACGACTCACTATAGGGAGTCCTCAACCAAATCATAAATTGCGCGACCACATTGCTAGCATAACCCCTTGGGGCCTCAAGCTT
Sp4 target	AAGCTTGAGGCCCCAAGGGGTTATGCTAGGGTTATAGGTTTGCGCGTCTTGCTGGGCGATAGGACTCCCTATAGTGAGTCGTATTAGGATCC

Sp4 non-target	GGATCCTAATACGACTCACTATAGGGAGTCCTATCGCCC AGCAAGACGCGCAAACCTATAACCCTAGCATAACCCCTT GGGGCCTCAAGCTT

4.4 Working with RNA

4.4.1 Treatment of solutions, glassware and equipment

For protection against RNases, all applied buffers and solutions were treated with 0.1% (v/v) diethylpyrocarbonate (DEPC) and autoclaved to remove traces of DEPC after an overnight incubation at RT. Non-disposable plastic ware were treated with RNase Exitus Plus (AppliChem) whereas pipette tips and reaction tubes were autoclaved twice before use. Glassware was heat sterilized by incubation at 210°C for at least 2 hours before use.

4.4.2 Isolation of total RNA and small RNA from *S. putrefaciens* CN-32

4.4.2.1 Isolation of RNA using isopropanol

Total RNA from *S. putrefaciens* CN-32 was extracted using TRIzol reagent (Invitrogen), a solution containing phenol and guanidinium isothiocyanate. Two ml of an overnight culture was harvested and lysed by the addition of one ml TRIzol. Following the addition of 200 µl chloroform and centrifugation (16,000 x *g*, 5 min, 4°C), the upper aqueous phase was transferred to a fresh tube. The RNA in the aqueous phase was precipitated by the addition of 500 µl of isopropanol and pelleted by centrifugation (16,000 x *g*, 10 min, 4°C). The RNA pellet was washed with one ml of 70% ethanol, dried at 37°C and resuspended in 50 µl of DEPC treated water.

4.4.2.2 Isolation of RNA using *mir*Vana™ miRNA Isolation Kit (Ambion)

For the isolation of small RNA species (< 200 nt) the *mir*Vana™ miRNA Isolation Kit (Ambion) was used. Using a hand-held glass Teflon homogenizer, 0.1 g *S. putrefaciens* CN-32 cells were homogenized in 1 ml lysis buffer (containing guanidinium isothiocyanate) for 3 min on ice. After the addition of 100 µl miRNA Homogenate™ additive and incubation for 10 min on ice, the sample was phenol/chloroform extracted. To separate the total RNA into two fractions and enrich the small RNA species, the sample was brought in a first step to 25% ethanol to immobilize large RNAs on the glass-fiber filter. The small RNA species were

collected in the filtrate. The ethanol concentration of the filtrate was then increased to 55% and passed through a second glass-fiber filter to immobilize small RNAs. Both RNA fractions were washed and eluted in 100 µl Elution Solution pre-heated at 95°C.

4.4.3 Quantitative and qualitative analyses of RNA

The concentration of extracted RNA was determined spectrophotometrically, by measuring sample absorbance at 260 nm. The A_{260}/A_{280} ratio was used as an indicator for RNA purity, as pure RNA preparations showed a ratio higher than 1.8, whereas ratios lower than 1.8 indicate possible contamination with protein or phenol. The integrity of purified RNA was checked by Urea-PAGE (S 4.4.4) or by northern blotting (S 4.4.5).

4.4.4 Gel electrophoresis of RNA (Urea-PAGE)

Gel electrophoresis of RNA was performed using urea-polyacrylamide gels. 8 M urea was used to resolve the secondary structures of RNA and obtain separation based on the size of the RNA. To 8 M urea, 90 mM Tris pH 8.0, 90 mM boric acid, 2 mM EDTA, 1.0% (v/v) ammonium persulfate (APS) and 0.1% (v/v) tetramethylethylenediamine (TEMED), varying concentrations (6% to 20%) of polyacrylamide (acrylamide / bisacrylamide, 40%, ratio 29:1) was added depending on the size of the applied RNA. The polymerized gel was assembled in the gel chamber (Mini Protean Tetra System, Biorad) and run in 1x TBE buffer (90 mM Tris pH 8.0, 90 mM boric acid and 2 mM EDTA) at 200 V for 1h. The RNA samples were mixed with 2x formamide loading dye (80% formamide, 10 mM EDTA, 0.05% (w/v) bromophenol blue and 0.05% (w/v) xylenecyanol) and incubated for 5 min at 95°C before applying them to the urea-polyacrylamide gel. The RNA bands were visualized by toluidine blue staining.

4.4.5 Northern blotting

4.4.5.1 Transfer of RNA to a nylon membrane

Using a semi dry electrophoretic transfer system, the RNA separated based on sizes on a urea-polyacrylamide gel were transferred to a positively charged nylon membrane (Roti®-Nylon Plus, Roth). As a preparation for transfer, the membrane was soaked for 1 min with DEPC treated water and equilibrated along with the gel in 1x TBE buffer for 5 min. The transfer was performed for 2h at 25 V in a Biorad semidry apparatus. Following the transfer, the blot was washed in DEPC treated water to remove salts and the RNA was UV-crosslinked to the membrane.

4.4.5.2 Labeling of oligonucleotide probes

RNA probes for EMSAs and nuclease assays and DNA oligonucleotide probes for northern blotting (Table 6) were radiolabeled at the 5′ end using γ-[^{32}P]-ATP. 100 pmoles of the probe was 5′ labeled with γ-[^{32}P]-ATP (5000 Ci/mmol, Hartmann Analytic) using 10U of T4 polynucleotide kinase (NEB, Ambion) at 37°C for 1h. The reaction was stopped by the addition of 2x formamide loading dye and separated on a 10% urea-polyacrylamide gel. The labeled probe was cut out of the gel after phosphorimaging, gel eluted and precipitated using ethanol. After resuspending the probe in 50 µl of DEPC treated water, its radioactivity was measured using a scintillation counter.

4.4.5.3 Hybridization with labeled probes and visualization

Upon hybridization with radiolabelled probes, specific RNAs were detected on a blot containing 10 µg of total RNA from *S. putrefaciens* CN-32. Therefore, the blot was initially pre-incubated with ULTRAhyb-Oligo hybridization buffer for 30 min at 42°C to block non-specific binding sites. The probe (10^6 cpm/ml of hybridization buffer) was applied to the ULTRAhyb-Oligo buffer after incubation at 95°C for 5 min. After overnight incubation at 42°C, the blot was washed twice, with low stringency buffer (2X SSC + 0.1% (w/v) SDS) and with high stringency buffer (00.1X SSC + 0.1% (w/v) SDS) at 42°C for 15 min to remove unbound probe. Overnight exposure of the blot to a phosphorimaging screen followed by phosphorimaging using a storm 840 phosphorimager resulted in visualization of the bands.

4.4.6 Generation of specific RNA by *in vitro* transcription

Fragments for RNA substrates (fragment size ~60-150 bp) were amplified via PCR using plasmid DNA listed in Table 5, purified after agarose gel electrophoresis and used as templates for RNA *in vitro* transcription.

In vitro transcription to generate crRNAs was performed using T7 RNA polymerase in a standard reaction assay (20 µl) containing 500-1,000 ng purified PCR product. The reaction buffer contained 40 mM HEPES/KOH pH 8.0, 22 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 1 mM spermidine, 4 mM NTP mix and 30 nM of T7 RNA polymerase. The samples were mixed, centrifuged briefly and incubated for 2h at 37°C. Addition of 2x formamide loading dye stopped the reaction and the products were separated on a 10% urea-polyacrylamide gel and visualization by toluidine blue staining. The RNA was gel eluted, precipitated using ethanol and resuspended in 50 µl of DEPC-treated water. Final concentration of the RNA was determined spectrophotometrically (S. 4.4.3). For higher concentrations of RNA, *in vitro* transcription at a larger volume was performed.

4.4.7 Nuclease assays

The endoribonuclease activity of recombinant Cas6 and *in vitro* reconstituted Cas proteins were assayed using radiolabeled RNA as substrates. Labeling of the *in vitro* generated RNA substrate was performed as described in S 4.4.5.2. A standard nuclease assay contained 50 nM radiolabeled RNA substrates (CRISPR substrates or RNA controls) and 100-500 nM proteins in a reaction buffer of 50 mM Tris/HCl pH 7.0, 100 mM NaCl, and 1 mM DTT. The reaction was initiated by incubating the sample at 30°C for 30 min and stopped by adding 2x formamide loading dye and boiling at 95°C for 5 min. Separation of the cleavage products was achieved via 10% or 20% Urea-PAGE and the bands were visualized using a storm 840 phosphorimager.

4.4.7 Electrophoretic Mobility Shift Assay (EMSA)

Electrophoretic mobility shift assays were employed to investigate the nucleic acid-binding properties of proteins. It is a procedure that utilizes the slow migratory

behavior of a nucleic acid bound protein complex compared to free nucleic acid through a non-denaturing polyacrylamide gel (S. 4.5.3.2). To perform the assay, 50 nM radiolabeled RNA was mixed with 100 -500 nM protein in a buffer containing 50 mM Tris/HCl pH 7.0, 100 mM NaCl, and 1 mM DTT and incubated for 30 min at 30°C. The sample was mixed with GelPilot loading dye (Qiagen GmbH) and applied to a 6% non-denaturing polyacrylamide gel. The gel was run at 10 W for 2 h and exposed to a phosphorimaging screen overnight. The bands were visualized using the storm 840 phosphorimager.

4.4.8 qRT-PCR analysis

To compare mRNA transcript levels between wild-type and deletion strains of *S. putrefaciens* CN-32 quantitative real-time polymerase chain reactions (qRT-PCR) were performed. cDNA was synthesized from the extracted DNA-free total RNA using the First Strand cDNA Synthesis Kit (Thermo Scientific). qRT-PCR was performed with the iQ SYBR Green Supermix Kit (Bio-Rad, Munich, Germany) using 10ng of synthesized cDNA and gene specific primers (Table 6), designed with a specific annealing temperature of 59°C. The PCR was performed in a CFX96 Real-Time PCR Detection System (Bio-Rad) using the following protocol:

Step 1) 95°C – 3 min
Step 2) 95°C – 10 sec
Step 3) 59°C – 10 sec } x 40
Step 4) 72°C – 30 sec/kb

The PCR was also performed using primers against the 16S rRNA, which served as control. The specificity of the amplification was verified using the melting curve of the CFX96 Real-Time PCR Detection System. The calculation of the C_T-value (threshold cycle), which describes the cycle at which the fluorescence is higher than the background fluorescence, was done using the Bio-Rad CFX Manager Software. The expression levels of *cas* genes was depicted relative to the control 16S rRNA. Three independent experiments were conducted and the mean of the values was used for the analysis.

4.5 Working with DNA

4.5.1 Preparation of plasmid DNA from *E. coli*

Plasmids used for cloning or sequencing analyses were prepared using the Qiagen plasmid mini kit or the Qiagen plasmid maxi kit according to the manufacturer′s instructions.

4.5.2 Quantitative and qualitative analyses of DNA

The concentration of DNA was determined spectrophotometrically by measuring the absorption of the sample at 260 nm. The purity of DNA was determined by the ratio of absorbance at 260 nm to 280 nm. An A_{260}/A_{280} ratio of 1.8-2.0 was observed in pure DNA preparations whereas a ratio lower than 1.8 indicated that the presence of potential contaminants.

4.5.3 Electrophoresis of DNA

4.5.3.1 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis was performed to separate DNA fragments based on size. Depending on the size of the DNA fragments, agarose gels of varying concentrations (1% to 2% (w/v) of agarose) in TAE buffer (40 mM Tris pH 8.0, 40 mM acetic acid and 1 mM EDTA) were prepared. Ethidium bromide (final concentration: 0.5 µg/ml) was added to the gel mixture to visualize DNA, as it intercalates between base pairs of DNA or RNA molecules and emits fluorescence when excited by UV light. The DNA samples mixed with 6x DNA loading dye (10 mM Tris/HCl pH 7.6, 0.03% bromophenol blue, 0.03% xylene cyanol, 60% glycerol and 60 mM EDTA) were applied and electrophoresis were performed at 200 V in TAE buffer. 5 µl of a DNA marker containing a mixture of DNA fragments of known size and amount was also applied onto the gel. After the electrophoresis run, the DNA was visualized by exposing the gel to UV light using a Chemi-Doc-gel documentation system. The size of the separated DNA fragments was determined by comparing their relative positions to those of the DNA ladder.

4.5.3.2 Non-denaturing polyacrylamide gel electrophoresis of DNA (Native-PAGE)

Electrophoretic separation of smaller DNA fragments (< 200nt) was performed under non-denaturing conditions using polyacrylamide gels (120). Depending on the size of the DNA fragment, the concentration of polyacrylamide (acrylamide / bisacrylamide, 40%, ratio 29:1) in the gel (90 mM Tris pH 8.0, 90 mM boric acid, 2 mM EDTA, 0.03% (v/v) APS, 0.005% (v/v) TEMED) was varied between 4% and 12% (v/v). The DNA samples mixed with 6x DNA loading dye were applied on to the gel. DNA marker containing a mixture of DNA fragments of known size was also applied onto the gel to size the fragments. The gel run was performed in 1x TBE at 10 W for 1 h.

4.5.4 Purification of DNA fragments

The PCR amplification products were separated according to size on an agarose gel. The fragment of interest was cut out and extracted from the gel piece using the QIAquick gel extraction kit (Qiagen GmbH) following the instructions of the manufacturer.

4.5.5 Polymerase chain reaction (PCR)

The PCR technique was used for the amplification of specific DNA fragments *in vitro*. Two primers were designed such that they hybridize to the 5′ end of the coding and the noncoding strand, flanking the sequence of interest that has to be amplified. The elongation of these primers was catalyzed by the Phusion polymerase, which possesses proofreading capability and has a reduced mutation rate. A standard PCR reaction includes the following main steps:

1) Denaturation: Heating the reaction at 95°C results in the melting of dsDNA into ssDNA (template).
2) Primer annealing: Annealing or binding of the primers to their complementary DNA
3) Elongation: Extension or elongation of the primer in the 5′ to 3′ direction. DNA polymerase catalyzes the elongation by addition of complementary nucleotides.

Amplification of the sequence of interest was obtained by repetition of the above steps (25-30x). One important factor that affects the yield and purity of PCR is the annealing temperature of the primers. The primers used for PCR are listed in Table 6.

4.5.5.1 Amplification of genomic DNA and plasmid DNA

PCR amplifications from genomic or plasmid DNA was performed using the following reaction mixture:

~50 µg template DNA
250 µM dNTPs
0.2 µM Fw. primer
0.2 µM Rv. primer
1 x concentrated GC buffer
3% (v/v) DMSO
1 U Phusion polymerase
adjusted to 50 µl with water

The reaction was performed in a thermal cycler (BioRad) using the following program:

Step 1) 95°C – 60 sec
Step 2) 95°C – 30 sec
Step 3) 55-65°C – 30 sec } x 30 – 35
Step 4) 72°C – 30 sec/kb
Step 5) 72°C – 5 min

4.5.5.2 Site directed mutagenesis

Using the Agilent site-directed mutagenesis primer design platform, primer sets containing the desired mutations were designed (Table 6). Mutations were introduced in the plasmids by QuikChange site-directed mutagenesis (Stratagene) using the designed primers. The plasmids carrying the mutations were identified via sequencing.

4.5.5.3 Overlap extension PCR

In the overlap extension PCR (121), two PCR products with overlapping sequences are fused together and amplified to obtain a longer product. Two internal primers (b and c) and two external primers (a and d) are designed and in the first round of PCR, small DNA fragments are generated using, the primer combinations a-b and c-d. The PCR products were gel eluted (S 4.5.4), mixed together and using the external primer set a and d, a longer fused product was amplified. The primer

sets used for overlap extension PCR are listed in Table 6.

4.5.6 Modification of DNA

4.5.6.1 Restriction of DNA

The restriction digestion of plasmid DNA and PCR products was carried out with appropriate restriction endonucleases in respective buffers as recommended by the manufacturer. To obtain digested product, reaction mixture containing 5-10 U enzyme/µg DNA was incubated at 37°C for 2 h.

4.5.6.2 Ligation

T4 DNA ligase enzyme catalyzes the formation of a phosphodiester bond between 5′-phosphate and 3′-hydroxyl termini of duplex DNA in the presence of ATP and hence used in ligation of a restricted DNA fragment to a vector DNA (122-124). Restricted plasmid DNA and restricted insert in 1:3 molar ratio where taken in a ligation buffer (400 mM Tris/HCl pH 7.8, 100 mM $MgCl_2$, 100 mM DTT and 5 mM ATP). After the addition of 3 µl of T4 DNA ligase (1 Weiss-unit/µl), the reaction mixture was incubated overnight at 16°C and transformed into *E. coli* DH5α.

4.5.6.3 5′ labeling of DNA oligonucleotides

DNA oligonucleotide probes were radiolabeled using the procedure described in S. 4.4.5.2.

4.5.7 Electrophoretic Mobility Shift Assay (EMSA)

EMSA was performed to study the nucleic acid binding properties of a protein. To obtain the slow migration band on an EMSA, 50 nM of 5′ radiolabelled oligonucleotide primers was incubated with recombinant proteins or *in vitro* refolded complexes (100- 500 nM) in a buffer containing 50 mM HEPES/KOH pH 7.0, 50 mM NaCl, 5 mM $MgCl_2$ and 5 mM $MnCl_2$. The assay was performed at 30°C for 15 min, mixed with GelPilot loading dye (Qiagen GmbH) and applied to an 8% non-denaturing polyacrylamide gel. To ensure specificity of the interaction, 1 µg of yeast RNA as a competitor was added to the assay. The migration bands were visualized by phosphorimgaing after an electrophoretic run at a constant 10 W for 2 h.

4.5.8 Nuclease assays

The nuclease activity of recombinantly produced Cas3 and *in vitro* reconstituted Cas proteins were assayed using radiolabeled oligonucleotide substrates. Labeling of the oligonucleotide primers was performed as described in S 4.4.5.2. The nuclease assay contained 50 nM radiolabeled oligouleotide (CRISPR substrates or DNA controls) and 100-500 nM proteins in a reaction buffer containing 50 mM HEPES/KOH pH 7.0, 50 mM NaCl, 5 mM ATP, 5 mM $MgCl_2$, 5 mM $MnCl_2$, and 20 ng yeast RNA. The reaction was incubated at 30°C for 15 min, stopped by adding 2x formamide loading dye and boiling at 95°C for 5 min and separated via 20% Urea-PAGE. Phosphorimaging was employed for the visualization of the cleavage products.

4.6 Transformation

4.6.1 Transformation of *E. coli*

4.6.1.1 Preparation of competent *E. coli* DH5α and DH5α λpir cells

Competent cells of *E. coli* (S. 2.3) were chemically prepared by using rubidium chloride (RbCl) and calcium chloride ($CaCl_2$) (125). 100 ml 2x YT medium containing 10 mM $MgCl_2$ and 10 mM $MgSO_4$ was inoculated with 2 ml of an overnight culture of *E. coli*. The culture was grown at 37°C with constant agitation at 200 rpm until OD_{600} of ~ 0.6 was reached. After cooling the culture on ice for 30 min, the cells were harvested by centrifugation (3000 x *g*, 10 min, 4°C). Subsequently, the cell pellet was gently resuspended in 33 ml chilled solution A (30 mM Potassium acetate pH 5.8, 100 mM RbCl, 50 mM $MgCl_2$, 10 mM $CaCl_2$ and 15% glycerol) and incubated on ice for 30 min. After centrifugation (3000 x *g*, 10 min, 4°C) the pellet was again gently resuspended in 5 ml chilled solution B (10 mM MOPS pH 5.8, 75 mM $CaCl_2$, 10 mM RbCl and 15% glycerol) and incubated for 30 min on ice. 100 µl aliquots of the resuspended solution was made and was used for transformation (S. 4.6.1.3) or stored at -80°C.

4.6.1.2 Preparation of competent *E. coli* WM3064 cells

Chemically competent cells of *E. coli* WM3064 were prepared using the Inoue method (126). 250 ml of SOB medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl and 20 mM $MgCl_2$) supplemented with 300 µM DAP was inoculated with an overnight culture such that the starting OD_{600} of the culture was 0.05. The cells were grown at 18°C with constant agitation at 200 rpm till an OD_{600} of ~ 0.6 was reached. After cooling the culture on ice for 10 min, the cells were harvested by centrifugation (2,500 x *g*, 10 min, 4°C). The cells were gently resuspended in 75 ml of ice-cold TB-buffer (10 mM PIPES/KOH pH 6.7, 250 mM KCl and 15 mM $CaCl_2$) supplemented with DAP, incubated on ice for 10 min and pelleted (2,500 x *g*, 10 min, 4°C). The cells were again resuspended in 20 ml of ice-cold TB-buffer with DAP and 7% (v/v) DMSO and incubated for 10 min. After incubation, 200 µl aliquots were prepared and used for transformation. Long-term storage was carried out at -80°C.

4.6.1.3 Transformation of competent *E. coli* cells

Plasmid DNA was gently mixed with 100 µl of competent *E. coli* cells and incubated on ice for at least 45 min. The cells were given a heat shock by incubating at 42°C for 45 sec followed by placing on ice for 1 min. 1 ml of LB medium was added and incubated at 37°C for about 45 min accompanied by shaking. 100 µl of transformed cells were plated on LB agar plates containing the respective antibiotics. Whereas the remaining 900-1000 µl cell suspension was centrifuged, resuspended in about 100 µl LB-medium and plated. Overnight incubation of LB agar plates at 37°C resulted in colonies that were screened for positive clones carrying the recombinant plasmid using restriction digestion (S. 2.6.8.1) of isolated plasmid DNA (S. 2.6.2.1, S. 2.6.2.2). For the transformation of *E. coli* WM3064, media was every supplemented with DAP.

4.6.3 Conjugation of *S. putrefaciens* CN-32

E. coli WM3064 strain carrying the desired plasmid and *S. putrefaciens* CN-32 were grown in 20 ml of LB media overnight (medium for *E. coli* WM3064 was supplemented with DAP and antibiotics for the plasmid). Equal OD_{600} of both strains were harvested and washed once with LB supplemented with DAP (LB-DAP). Both

strains were resuspended in 100 µl of LB-DAP, mixed and spotted as a single drop on LB-DAP plates. After overnight incubation at 30°C, the spot was dislodged from the plate and washed three times using plain LB. 100 µl of this mating culture was platted on LB plates supplemented with antibiotics and incubated at 30°C to screen for transformed *S. putrefaciens* CN-32. The presence of plasmid was confirmed through colony PCR or plasmid isolation and sequencing.

4.7 Working with *S. putrefaciens* CN-32

4.7.1 In-frame insertions and deletions

4.7.1.1 Construction of in-frame insertion and deletion plasmids

For in-frame deletions, 500 bp upstream and downstream to the gene of interest were PCR amplified using genomic DNA and gel purified. Using overlap extension PCR, the two fragments were fused and the final product was restriction digested and ligated to the suicide plasmid, pNPTS138-R6KT (kan^R). Similarly for in-frame insertions, an insert with the gene of interest and 500 bp upstream and downstream of the locus of interest were generated via PCR and ligated to the suicide plasmid. The generated suicide plasmids are listed in Table 5.

4.7.1.2 In-frame insertion and deletion via homologous recombination

The above-generated plasmids were conjugated into *S. putrefaciens* CN-32 using *E. coli* WM3064 as a donor strain and due to homologous recombination, the insertion or deletion strains were generated. Single crossover integration was obtained by screening on LB kanamycin (LB-kan) plates lacking DAP. Single colonies were grown in plain LB to achieve the second homologous recombination and plated on LB plates containing 10% sucrose (v/v). Colonies lacking the plasmid were selected based on its kanamycin sensitivity and using colony PCR, were checked for target insertion or deletion.

4.7.2 Conjugation-based *in vivo* interference assay

4.7.2.1 Construction of plasmids for conjugation

Oligonucleotides that contained the protospacer and PAM sequences were designed such that on hybridization, the sticky ends of EcoRI and BamHI restriction sites flanked the ends of the protospacer sequence. These oligonucleotides were phosphorylated at the 5′ terminus (as described in S. 4.4.5.2 using non-labeled ATP), mixed and hybridized. The obtained DNA fragment was ligated to the linearized pBRR1MCS2 (kan^R) vector. Insertion of the fragment was into the vector was confirmed by sequencing. The oligonucleotides used for the construction of these plasmids are listed in Table 6.

4.7.2.2 *In vivo* interference assay

These plasmids were conjugated into *S. putrefaciens* CN-32 wild-type and deletion strains using *E. coli* WM3064 as donor as described in S. 4.6.3. The *S. putrefaciens* CN-32 colonies containing the plasmid were obtained on LB-kan plates lacking DAP and counted. As a control, conjugation of *S. putrefaciens* CN-32 strains was performed with the empty plasmid and the colonies counted. The conjugation efficiency for a plasmid containing a protospacer in a strain was determined by taking the ratio of number of colonies obtained for this plasmid to the empty plasmid. Interference in this strain was observed if its conjugation efficiency was found to significantly lower than 1.

4.8 Biochemical methods

4.8.1 Heterologous production of Cas proteins in *E. coli*

For the heterologous production of *S. putrefaciens* CN-32 proteins, the encoding genes were PCR amplified and cloned into expression vectors (Table 5). PCR was performed using Phusion DNA polymerase with genomic DNA of *S. putrefaciens* CN-32 serving as the template and the primers listed in Table 6. The sequences of the cloned genes were verified by sequencing and used for heterologous expression in *E. coli* BL21(DE3)pLysS.

4.8.2 Preparation, enrichment and purification of the recombinant enzymes

4.8.2.1 Purification of the recombinant Cas3 protein

Recombinant Cas3 was produced as a SUMO-tagged protein in *E. coli* BL21(DE3)pLysS. The cell pellet was resuspended in buffer containing 50 mM Tris/HCl pH 7.0, 10% glycerol and 300 mM NaCl (5 ml/g cells) and incubated on ice with lysozyme. The cells were lysed by sonication and the cell lysate was clarified by centrifugation (18,000 rpm, 20 min, 4°C). Clarified lysate was applied to a His-Trap Ni Sepharose column (GE Healtcare) and using a step gradient (20% imidazole steps) in imidazole concentration of the buffer on the ÄKTA protein purification system (GE Healtcare), the protein was eluted from the column. Fractions containing the protein were identified after SDS-PAGE and concentrated using vivaspin centrifugal concentrators. The concentration of the protein was determined using the Bradford assay (127).

4.8.2.2 Purification of the recombinant Cas1821 protein

Recombinant Cas1821 was produced as a SUMO-tagged protein in *E. coli* BL21(DE3) pLysS. The cell pellet was resuspended in buffer containing 100 mM KPO_4 pH7.5, 500 mM NaCl, 10% glycerol and 20 mM imidazole (5 ml/mg) and incubated on ice for 4°C with protease inhibitor (1 ml per 50 ml of lysate), 4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride (serine protease inhibitor, 75 µl/ 50 ml of lysate) and DNase I (0.05 mg/ml). The cells were lysed by sonication, clarified by centrifugation (18,000 rpm, 20 min, 4°C) and applied to a His-Trap Ni Sepharose column. Using a very steep gradient in imidazole concentration, the protein was eluted and dialyzed overnight at 4°C into 50 mM Tris/HCl pH 7.5, 300 mM NaCl, 10% glycerol, 1 mM DTT and 1 mM EDTA. Addition of SUMO-protease during dialysis resulted in the cleavage of the SUMO-tag from the protein. Following a second dialysis step into 50 mM Tris/HCl pH 7.0, 100 mM NaCl, 10% glycerol, 1 mM DTT and 1 mM EDTA for 3h at 4°C, the protein was applied to a Heparin Sepharose column and using a gradient in the salt concentration, the protein was eluted from the column. The purity of this protein was checked by SDS-PAGE and the concentration determined using the Bradford assay.

4.8.2.3 Purification of the recombinant Cas1822 protein

Recombinant Cas1822 was produced as a CPD-tagged protein in *E. coli* BL21(DE3)pLysS. The cell pellet was resuspended in buffer containing 100 mM KPO_4 pH7.5, 500 mM NaCl, 10% glycerol and 20 mM imidazole (5 ml/mg). The cells were lysed by sonication after incubation on ice for 30 min with lysozyme. After clarification of the cell lysate by centrifugation (18,000 rpm, 20 min, 4°C) the lysate was applied to a His-Trap Ni Sepharose column. A linear gradient in imidazole was used to elute the protein from the column and the eluted protein dialyzed into 50 mM Tris/HCl pH 7.5, 100 mM NaCl, 10% glycerol, 1 mM DTT and 1 mM EDTA at 4°C. During dialysis, 100 µM of phytic acid was added to induce autocatalysis of the CPD-tag from the protein. The cleavage of the tag was checked by SDS-PAGE and the protein concentration was determined using the Bradford assay.

4.8.2.4 Purification of the recombinant Cas6f protein

Recombinant Cas6 was produced with a C-terminal 6x His-tag in *E. coli* BL21(DE3)pLysS. The cell pellet was resuspended in buffer containing 50 mM Tris/HCl pH 8.0, 500 mM NaCl, 20 mM Imidazole, 5% glycerol, 0.01% triton X100 and 1 mM DTT (5 ml/mg) and lysed via sonication. Cell lysate was clarified by centrifugation (18,000 rpm, 20 min, 4°C) and applied to a His-Trap Ni Sepharose column. The column was washed with the above buffer containing 100 mM imidazole and using a linear gradient in imidazole concentration, the protein was eluted. The purity of the protein was checked by SDS-PAGE and the concentration was determined using the Bradford assay.

4.8.2.5 Purification of the variant Type I-F Cascade complex

E. coli BL21(DE3)pLysS cells producing the interference complex were resuspended in buffer containing 50 mM Tris/HCl pH 7.0, 500 mM NaCl, 10% glycerol, 20 mM Imidazole and 1 mM DTT (5 ml/mg) and lysed via sonication. Cell lysate was clarified by centrifugation (18,000 rpm, 20 min, 4°C), applied to a His-Trap Ni Sepharose column and eluted with a linear gradient in imidazole concentration. Fractions containing the complex were pooled, concentrated and further purified on a size-exclusion chromatography column equilibrated with 50 mM HEPES/NaOH, 150 mM NaCl and 1 mM DTT. The purity of the complex was verified by SDS-PAGE and the presence of RNA was detected by Urea-PAGE. The concentration of the complex was determined using the Bradford assay.

4.8.2.6 Inclusion body purification

Due to a very low amount of recombinant protein in the soluble fraction, the proteins were purified from inclusion bodies, which are insoluble aggregates that are often formed during high-level production of recombinant proteins in *E. coli*. Inclusion bodies contain the respective protein in a highly enriched form and can be isolated by *in vitro* refolding techniques (128,129). For that purpose, 5 g of BL21(DE3)pLysS cells producing Cas1, Cas3, Cas1822 and Cas6 were resuspended in 25 ml buffer 1 (100 mM Tris/HCl pH 7.0, 1 mM EDTA). Lysozyme (1.5 mg/g cells) was added to the cell suspension, mixed gently and incubated for 30 min at 4°C. The cells were lysed via sonication and incubated for 30 min at 25°C after the addition of 10 µg/ml DNase and 3 mM $MgCl_2$. Following the addition of 0.5 vol of buffer 2 (60 mM EDTA, 6% Triton X-100, 1.5 M NaCl pH 7.0) and incubation for 30 min at 4°C on a rocking platform, the inclusion bodies were pelleted by centrifugation (31,000 x *g*, 10 min, 4°C). The supernatant was discarded and the inclusion bodies were resuspended in 40 ml buffer 3 (100 mM Tris/HCl pH 7.0 and 20 mM EDTA) and again centrifuged (31,000 x *g*, 10 min, 4°C). The obtained inclusion bodies were stored at -20°C for a maximum period of two weeks. For solubilisation of inclusion bodies, 50 mg of the pellet was dissolved in 5 ml buffer 4 (100 mM Tris/HCl pH 8.0, 6 M guanidine hydrochloride (Gua-HCl), 100 mM DTT and 1 mM EDTA) and incubated for 2-4 h at 25°C on a rocking platform. Subsequently the pH was decreased to 3-4 with 250-300 µl 1M HCl and the sample centrifuged (10,000 x *g*, 10 min, 4°C) to remove cell debris. To eliminate DTT from the sample, two dialysis steps were conducted. First, the protein mixture was dialyzed for 2 h at RT against 500 ml buffer 5 (4 M Gua-HCl and 10 mM HCl), then dialyzed overnight at 4°C against 500 ml 4 M Gua-HCl. The protein concentration was determined using the Bradford assay. The solubilized protein was stored at -80°C.

4.8.2.7 Reconstitution of the Cas proteins from inclusion bodies

The protein complex was refolded by removal of the denaturing agent by stepwise dialysis against a native Gua-HCl free buffer (129). Equal amounts (300 µg) of each solubilized protein Cas1, Cas3, Cas1822 and cas6 (S. 4.8.2.6) and 300 µg of the purified Cas1821 were mixed with 10 ml of buffer 6 (50 mM Tris/HCl pH 7.0, 3.5 M Gua-HCl, 10% glycerol, 300 mM NaCl, 10 mM $MgSO_4$ and 1 mM DTT). The solution was mixed thoroughly and dialyzed at 4°C against 200 ml of buffer 7

(50 mM Tris/HCl pH 7.0, 3 M Gua-HCl, 10% glycerol, 300 mM NaCl, 10 mM $MgSO_4$ and 1 mM DTT). By decreasing the concentration of the denaturing agent in the dialyzing buffer in regular steps (by 0.5 M Gua-HCl every 2h), the complex was reconstituted. After dialysis, the protein sample was centrifuged (14,000 *x g*, 15 min, 4°C) and its concentration was determined using the Bradford assay. The purity of refolded complex was checked by SDS-PAGE.

4.8.3 Western blot analysis

4.8.3.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis

For the separation of proteins based on their molecular weight, discontinuous sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed (130). The SDS polyacrylamide gels are composed of a stacking gel (125 mM Tris/HCl pH 6.8, 5% (v/v) Acrylamide, 0.1% (w/v) SDS, 0.03% (v/v) APS and 0.005% (v/v) TEMED) and a separation gel (375 mM Tris/HCl pH 8.8, 15% (v/v) Acrylamide, 0.1% (w/v) SDS, 0.03 % (v/v) APS, 0.005% (v/v) TEMED). The acrylamide concentration in the separating gel was varied between 10 and 15% (v/v) depending on the desired separation of the protein samples. Samples mixed with 2x SDS loading dye (125 mM Tris/HCl pH 6.8, 20% glycerol, 4% (v/v) SDS, 10% (v/v) β-mercaptoethanol and 0.01% (w/v) bromophenol blue) were applied to the gel after boiling at 95°C for 5 min. The run was carried out in a gel chamber (Mini Protean Tetra System, Biorad) containing 1x SDS running buffer (25 mM Tris/HCl pH 8.0, 190 mM glycine and 0.1% (v/v) SDS) at 200V for 50 min. After separation, Coomassie based staining solution (Roth) was used to visualize the protein bands on the gel.

4.8.3.2 Transfer and detection of the proteins

Using a semi dry electrophoretic transfer system, the proteins separated on an SDS-polyacrylamide gel were transferred to a nitrocellulose membrane (Amersham Hybond ECL Nitrocellulose Membrane; GE Healthcare). As a preparation for transfer, the membrane was soaked for 1 min in methanol and then equilibrated along with the gel in transfer buffer (50 mM Tris/HCl pH 7.5, 39 mM Glycine, 0.0375% (w/v) SDS and 20% Methanol) for 5 min. The transfer was

performed for 90 min at 25 V in a Biorad semidry apparatus. Following the transfer, the membrane was washed in water and incubated with blocking solution (50 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.02% (v/v) Tween20 and 5% (w/v) Skimmed milk powder) at room temperature (RT) for 1 h and then the antibody solution containing the anti-His-HRP antibody (1: 10,000; Anti-His Mouse monoclonal IgG_{2b} antibody; Life Technologies) was added and incubated at RT for 45 h. The membrane was washed three times with the blocking buffer for 10 min and chemiluminescence reaction was conducted using the Roti®-Lumin plus reagent (Roth) according to the supplier's protocol. The film (Medical X-Ray Screen; CEA, Assamstadt) was exposed for 60 s – 6 h depending on the signal strength and developed in a developing machine (QX-60; Konica, Munich).

4.8.4 Co-immunoprecipitation

The interaction partners of His-HA-Cas1821 were identified using Dynabeads coupled to anti-HA antibodies. The coupling of the Dynabeads (Invitrogen, 50 µl/sample) to anti-HA antibodies (3 µl/ sample, anti-HA monoclonal mouse Ig_{2b} antibody, Life technologies) was achieved by incubating the mixture at 4°C for 1 h. Before the addition of the antibodies, the Dynabeads were equilibrated with 1xPBS buffer. To remove unbound antibodies, the Dynabeads were washed twice with 1xPBS buffer and resuspended in buffer containing 50 mM Tris/HCl pH 7.5, 300 mM NaCl and 10% glycerol. These beads were then added to *S. putrefaciens* cell lysate and incubated at 4°C for 2 h with gentle rotation. After incubation, the Dynabeads were pelleted and washed five times in buffer containing 50 mM Tris/HCl pH 7.5 and 300 mM NaCl to remove unbound proteins. The bound proteins were eluted by resuspended the beads in 2x SDS loading dye and boiling for 5 min at 95°C and separated via SDS-PAGE. Presence of the protein of interest was detected by western blot analysis and its interaction partners were identified by mass spectrometry.

4.8.5 Mass spectrometry

Interaction partners that were obtained during co-IP experiments were identified using mass spectrometry and were performed in collaboration with Dr. Jörg Kahnt at the MPI for Terrestrial Microbiology, Marburg.

4.8.6 Electron Microscopy

Prof. Dr. Andreas Klingl and Dr. Thomas Heimerl of the Philipps-Universität Marburg, performed the Electron Microscopy of the Cas1821 filaments and the variant Type I-F Cascade complexes.

4.8.7 Sequencing of RNA

Preparation of RNA-seq libraries and the sequencing of RNA libraries by Illumina HiSeq2000 approach were performed at the Max-Planck genome Centre, Cologne.

5 References

1. Whitman, W.B., Coleman, D.C. and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences*, **95**, 6578-6583.
2. Breitbart, M. and Rohwer, F. (2005) Here a virus, there a virus, everywhere the same virus? *Trends in microbiology*, **13**, 278-284.
3. Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. *Nature*, **399**, 541-548.
4. Labrie, S.J., Samson, J.E. and Moineau, S. (2010) Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, **8**, 317-327.
5. Samson, J.E., Magadán, A.H., Sabri, M. and Moineau, S. (2013) Revenge of the phages: defeating bacterial defences. *Nature Reviews Microbiology*, **11**, 675-687.
6. Wilson, G.G. and Murray, N.E. (1991) Restriction and modification systems. *Annual review of genetics*, **25**, 585-627.
7. Goldfarb, T., Sberro, H., Weinstock, E., Cohen, O., Doron, S., Charpak‐Amikam, Y., Afik, S., Ofir, G. and Sorek, R. (2015) BREX is a novel phage resistance system widespread in microbial genomes. *The EMBO journal*, **34**, 169-183.
8. Yamaguchi, Y., Park, J.-H. and Inouye, M. (2011) Toxin-antitoxin systems in bacteria and archaea. *Annual review of genetics*, **45**, 61-79.
9. Horvath, P. and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167-170.
10. Bhaya, D., Davison, M. and Barrangou, R. (2011) CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annual review of genetics*, **45**, 273-297.
11. Sorek, R., Lawrence, C.M. and Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual review of biochemistry*, **82**, 237-266.
12. Pourcel, C., Salvignol, G. and Vergnaud, G. (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, **151**, 653-663.
13. Mojica, F.J., García-Martínez, J. and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular evolution*, **60**, 174-182.
14. Grissa, I., Vergnaud, G. and Pourcel, C. (2009), *Molecular Epidemiology of Microorganisms*. Springer, pp. 105-116.
15. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of bacteriology*, **169**, 5429-5433.
16. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709-1712.
17. Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I. and Yakunin, A.F. (2011)

Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology*, **9**, 467-477.
18. Haft, D.H., Selengut, J., Mongodin, E.F. and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology*, **1**, e60.
19. Plagens, A., Richter, H., Charpentier, E. and Randau, L. (2015) DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS Microbiology Reviews*, fuv019.
20. Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A. and Severinov, K. (2013) High-throughput analysis of type IE CRISPR/Cas spacer acquisition in E. coli. *RNA biology*, **10**, 716-725.
21. Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M., Amitai, G., Edgar, R., Qimron, U. and Sorek, R. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, **520**, 505-510.
22. Nuñez, J.K., Lee, A.S., Engelman, A. and Doudna, J.A. (2015) Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature*.
23. Mojica, F., Diez-Villasenor, C., Garcia-Martinez, J. and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**, 733-740.
24. Gesner, E.M., Schellenberg, M.J., Garside, E.L., George, M.M. and MacMillan, A.M. (2011) Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nature structural & molecular biology*, **18**, 688-692.
25. Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R. and Randau, L. (2012) Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. *Nucleic acids research*, **40**, 9887-9896.
26. Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. and Doudna, J.A. (2010) Sequence-and structure-specific RNA processing by a CRISPR endonuclease. *Science*, **329**, 1355-1358.
27. Carte, J., Pfister, N.T., Compton, M.M., Terns, R.M. and Terns, M.P. (2010) Binding and cleavage of CRISPR RNA by Cas6. *Rna*, **16**, 2181-2188.
28. Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J. and Charpentier, E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, **471**, 602-607.
29. Chylinski, K., Le Rhun, A. and Charpentier, E. (2013) The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA biology*, **10**, 726-737.
30. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V. and Van Der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960-964.
31. Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, Ü., Wurm, R. and Wagner, R. (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology*, **18**, 529-536.
32. Rouillon, C., Zhou, M., Zhang, J., Politis, A., Beilsten-Edmands, V., Cannone, G., Graham, S., Robinson, C.V., Spagnolo, L. and White, M.F. (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Molecular cell*, **52**, 124-134.

33. van der Oost, J., Westra, E.R., Jackson, R.N. and Wiedenheft, B. (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nature Reviews Microbiology*.
34. Plagens, A., Tripp, V., Daume, M., Sharma, K., Klingl, A., Hrle, A., Conti, E., Urlaub, H. and Randau, L. (2014) In vitro assembly and activity of an archaeal CRISPR-Cas type IA Cascade interference complex. *Nucleic acids research*, **42**, 5125-5138.
35. Makarova, K.S., Aravind, L., Wolf, Y.I. and Koonin, E.V. (2011) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct*, **6**, 38.
36. Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W. and Doudna, J.A. (2014) Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nature structural & molecular biology*.
37. Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H. and Fineran, P.C. (2014) Priming in the Type IF CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic acids research*, **42**, 8516-8526.
38. Mulepati, S., Héroux, A. and Bailey, S. (2014) Crystal structure of a CRISPR RNA–guided surveillance complex bound to a ssDNA target. *Science*, **345**, 1479-1484.
39. Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J. and Wiedenheft, B. (2014) Crystal structure of the CRISPR RNA–guided surveillance complex from Escherichia coli. *Science*, **345**, 1473-1479.
40. Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H. and Albers, S.-V. (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Molecular cell*, **45**, 303-313.
41. Maier, L.-K., Lange, S.J., Stoll, B., Haas, K.A., Fischer, S.M., Fischer, E., Duchardt-Ferner, E., Wöhnert, J., Backofen, R. and Marchfelder, A. (2013) Essential requirements for the detection and degradation of invaders by the Haloferax volcanii CRISPR/Cas system IB. *RNA biology*, **10**, 865-874.
42. Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H. and Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, **468**, 67-71.
43. Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J. and Dickman, M.J. (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proceedings of the National Academy of Sciences*, **108**, 10092-10097.
44. Staals, R.H., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N. and Varossieau, K. (2014) RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus. *Molecular cell*, **56**, 518-530.
45. Anders, C., Niewoehner, O., Duerst, A. and Jinek, M. (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*, **513**, 569-573.

46. Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K. and Lin, S. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, **343**, 1247997.
47. Jiang, F. and Doudna, J.A. (2015) The structural biology of CRISPR-Cas systems. *Current opinion in structural biology*, **30**, 100-111.
48. Seed, K.D., Lazinski, D.W., Calderwood, S.B. and Camilli, A. (2013) A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature*, **494**, 489-491.
49. Staals, R.H. and Brouns, S.J. (2013), *CRISPR-Cas Systems*. Springer, pp. 145-169.
50. Wang, R., Preamplume, G., Terns, M.P., Terns, R.M. and Li, H. (2011) Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure*, **19**, 257-264.
51. Niewoehner, O., Jinek, M. and Doudna, J.A. (2014) Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic acids research*, **42**, 1341-1353.
52. Nam, K.H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisa, M.P. and Ke, A. (2012) Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype IC/Dvulg CRISPR-Cas system. *Structure*, **20**, 1574-1584.
53. Scholz, I., Lange, S.J., Hein, S., Hess, W.R. and Backofen, R. (2013) CRISPR-Cas systems in the cyanobacterium Synechocystis sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. *PLoS One*, **8**, e56470.
54. Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K. and Brouns, S.J. (2013) Type IE CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS genetics*, **9**, e1003742.
55. Westra, E.R., van Erp, P.B., Künne, T., Wong, S.P., Staals, R.H., Seegers, C.L., Bollen, S., Jore, M.M., Semenova, E. and Severinov, K. (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular cell*, **46**, 595-605.
56. Sashital, D.G., Jinek, M. and Doudna, J.A. (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nature structural & molecular biology*, **18**, 680-687.
57. Makarova, K.S., Wolf, Y.I. and Koonin, E.V. (2013) The basic building blocks and evolution of CRISPR-Cas systems. *Biochem Soc Trans*, **41**, 1392-1400.
58. Przybilski, R., Richter, C., Gristwood, T., Clulow, J.S., Vercoe, R.B. and Fineran, P.C. (2011) Csy4 is responsible for CRISPR RNA processing in Pectobacterium atrosepticum. *RNA biology*, **8**, 517-528.
59. Haurwitz, R.E., Sternberg, S.H. and Doudna, J.A. (2012) Csy4 relies on an unusual catalytic dyad to position and cleave CRISPR RNA. *The EMBO journal*, **31**, 2824-2832.
60. Sternberg, S.H., Haurwitz, R.E. and Doudna, J.A. (2012) Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA*, **18**, 661-672.
61. Wiedenheft, B., Lander, G.C., Zhou, K., Jore, M.M., Brouns, S.J., van der Oost, J., Doudna, J.A. and Nogales, E. (2011) Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature*, **477**, 486-489.

62. Richter, C., Gristwood, T., Clulow, J.S. and Fineran, P.C. (2012) In vivo protein interactions and complex formation in the Pectobacterium atrosepticum subtype IF CRISPR/Cas System. *PLoS One*, **7**, e49549.
63. Koo, Y., Ka, D., Kim, E.-J., Suh, N. and Bae, E. (2013) Conservation and variability in the structure and function of the Cas5d endoribonuclease in the CRISPR-mediated microbial immune system. *Journal of molecular biology*, **425**, 3799-3810.
64. Judith, R., James, H.N. and Malcolm, F.W. (2013) CRISPR interference: a structural perspective. *Biochemical Journal*, **453**, 155-166.
65. Tsui, T.K.M. and Li, H. (2015) Structure Principles of CRISPR-Cas Surveillance and Effector Complexes. *Annual Review of Biophysics*, **44**.
66. Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z. and Wang, Y. (2014) Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. *Nature*, **515**, 147-150.
67. Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E. and Doudna, J.A. (2014) CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proceedings of the National Academy of Sciences*, **111**, 6618-6623.
68. Jackson, R.N. and Wiedenheft, B. (2015) A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses. *Molecular Cell*.
69. Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic acids research*, gkr606.
70. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F. and Nureki, O. (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, **156**, 935-949.
71. Chylinski, K., Makarova, K.S., Charpentier, E. and Koonin, E.V. (2014) Classification and evolution of type II CRISPR-Cas systems. *Nucleic acids research*, **42**, 6091-6105.
72. Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012) Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, **109**, E2579-E2586.
73. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, **337**, 816-821.
74. Richter, H., Randau, L. and Plagens, A. (2013) Exploiting CRISPR/Cas: interference mechanisms and applications. *International journal of molecular sciences*, **14**, 14518-14531.
75. Carte, J., Wang, R., Li, H., Terns, R.M. and Terns, M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes & development*, **22**, 3489-3496.
76. Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M. and Terns, M.P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*, **139**, 945-956.
77. Hale, C.R., Cocozaki, A., Li, H., Terns, R.M. and Terns, M.P. (2014) Target RNA capture and cleavage by the Cmr type III-B CRISPR–Cas effector complex. *Genes & development*, **28**, 2432-2443.

78. Samai, P., Pyenson, N., Jiang, W., Goldberg, G.W., Hatoum-Aslan, A. and Marraffini, L.A. (2015) Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell*, **161**, 1164-1174.
79. Pagani, L., Lang, A., Vedovelli, C., Moling, O., Rimenti, G., Pristera, R. and Mian, P. (2003) Soft tissue infection and bacteremia caused by Shewanella putrefaciens. *Journal of clinical microbiology*, **41**, 2240-2241.
80. DiChristina, T.J., Moore, C.M. and Haller, C.A. (2002) Dissimilatory Fe (III) and Mn (IV) reduction by Shewanella putrefaciens requires ferE, a homolog of the pulE (gspE) type II protein secretion gene. *Journal of bacteriology*, **184**, 142-151.
81. Myers, C.R. and Nealson, K.H. (1988) Microbial reduction of manganese oxides: interactions with iron and sulfur. *Geochimica et Cosmochimica Acta*, **52**, 2727-2732.
82. Moser, D.P. and Nealson, K.H. (1996) Growth of the facultative anaerobe Shewanella putrefaciens by elemental sulfur reduction. *Applied and environmental microbiology*, **62**, 2100-2105.
83. Abboud, R., Popa, R., Souza-Egipsy, V., Giometti, C.S., Tollaksen, S., Mosher, J.J., Findlay, R.H. and Nealson, K.H. (2005) Low-temperature growth of Shewanella oneidensis MR-1. *Applied and environmental microbiology*, **71**, 811-816.
84. Thormann, K.M., Saville, R.M., Shukla, S. and Spormann, A.M. (2005) Induction of rapid detachment in Shewanella oneidensis MR-1 biofilms. *Journal of bacteriology*, **187**, 1014-1021.
85. Lorenz, R., Bernhart, S.H., Zu Siederdissen, C.H., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, **6**, 26.
86. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, **10**, 845-858.
87. Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC bioinformatics*, **9**, 40.
88. De Castro, E., Sigrist, C.J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A. and Hulo, N. (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic acids research*, **34**, W362-W365.
89. Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011) Cas3 is a single‐stranded DNA nuclease and ATP‐dependent helicase in the CRISPR/Cas immune system. *The EMBO journal*, **30**, 1335-1342.
90. Wu, J., Kasif, S. and DeLisi, C. (2003) Identification of functional links between genes using phylogenetic profiles. *Bioinformatics*, **19**, 1524-1530.
91. Biswas, A., Gagnon, J.N., Brouns, S.J., Fineran, P.C. and Brown, C.M. (2013) CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA biology*, **10**, 817-827.
92. Pul, Ü., Wurm, R., Arslan, Z., Geißen, R., Hofmann, N. and Wagner, R. (2010) Identification and characterization of E. coli CRISPR‐cas promoters and their silencing by H‐NS. *Molecular microbiology*, **75**, 1495-1512.
93. Lintner, N.G., Kerou, M., Brumfield, S.K., Graham, S., Liu, H., Naismith, J.H., Sdano, M., Peng, N., She, Q. and Copié, V. (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short

palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *Journal of Biological Chemistry*, **286**, 21643-21656.
94. Garside, E.L., Schellenberg, M.J., Gesner, E.M., Bonanno, J.B., Sauder, J.M., Burley, S.K., Almo, S.C., Mehta, G. and MacMillan, A.M. (2012) Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *Rna*, **18**, 2020-2028.
95. Brendel, J., Stoll, B., Lange, S.J., Sharma, K., Lenz, C., Stachler, A.-E., Maier, L.-K., Richter, H., Nickel, L. and Schmitz, R.A. (2014) A complex of Cas proteins 5, 6, and 7 is required for the biogenesis and stability of clustered regularly interspaced short palindromic repeats (crispr)-derived rnas (crrnas) in Haloferax volcanii. *Journal of Biological Chemistry*, **289**, 7164-7177.
96. Pougach, K., Semenova, E., Bogdanova, E., Datsenko, K.A., Djordjevic, M., Wanner, B.L. and Severinov, K. (2010) Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Molecular microbiology*, **77**, 1367-1379.
97. Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J. and Severinov, K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proceedings of the National Academy of Sciences*, **108**, 10098-10103.
98. Sesto, N., Touchon, M., Andrade, J.M., Kondo, J., Rocha, E.P., Arraiano, C.M., Archambaud, C., Westhof, É., Romby, P. and Cossart, P. (2014) A PNPase dependent CRISPR system in Listeria. *PLoS genetics*, **10**, e1004065.
99. Rollins, M.F., Schuman, J.T., Paulus, K., Bukhari, H.S. and Wiedenheft, B. (2015) Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from Pseudomonas aeruginosa. *Nucleic acids research*, **43**, 2216-2222.
100. Hatoum-Aslan, A., Maniv, I., Samai, P. and Marraffini, L.A. (2014) Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *Journal of bacteriology*, **196**, 310-317.
101. Su, A.A., Tripp, V. and Randau, L. (2013) RNA-Seq analyses reveal the order of tRNA processing events and the maturation of C/D box and CRISPR RNAs in the hyperthermophile Methanopyrus kandleri. *Nucleic acids research*, gkt317.
102. Medina-Aparicio, L., Rebollar-Flores, J., Gallego-Hernandez, A., Vazquez, A., Olvera, L., Gutierrez-Rios, R., Calva, E. and Hernandez-Lucas, I. (2011) The CRISPR/Cas immune system is an operon regulated by LeuO, H-NS, and leucine-responsive regulatory protein in Salmonella enterica serovar Typhi. *Journal of bacteriology*, **193**, 2396-2407.
103. Westra, E.R., Pul, Ü., Heidrich, N., Jore, M.M., Lundgren, M., Stratmann, T., Wurm, R., Raine, A., Mescher, M. and Van Heereveld, L. (2010) H‐NS‐mediated repression of CRISPR‐based immunity in Escherichia coli K12 can be relieved by the transcription activator LeuO. *Molecular microbiology*, **77**, 1380-1393.
104. Stratmann, T., Pul, Ü., Wurm, R., Wagner, R. and Schnetz, K. (2012) RcsB‐BglJ activates the Escherichia coli leuO gene, encoding an H‐NS antagonist and pleiotropic regulator of virulence determinants. *Molecular microbiology*, **83**, 1109-1123.

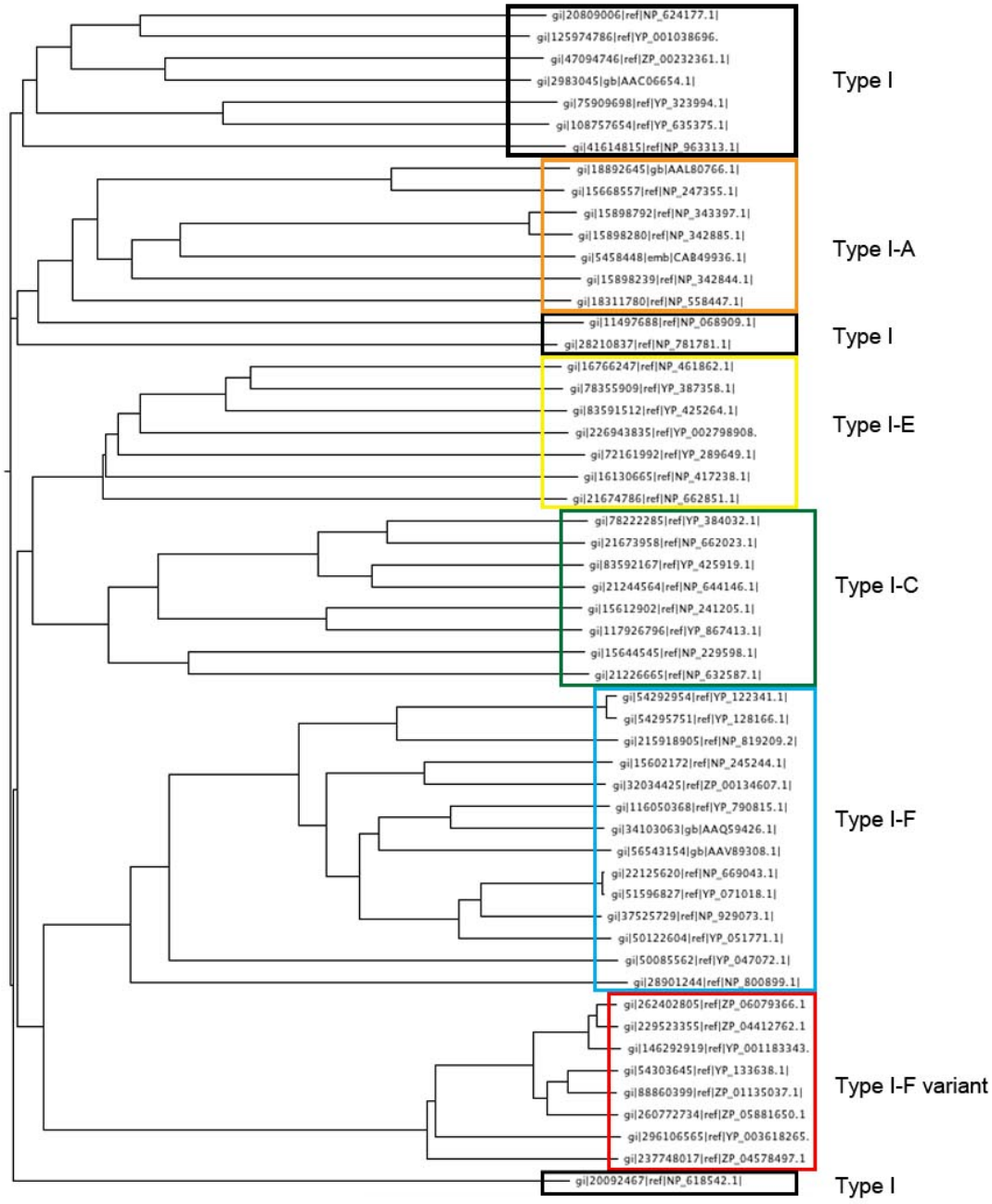
105. Patterson, A.G., Chang, J.T., Taylor, C. and Fineran, P.C. (2015) Regulation of the Type IF CRISPR-Cas system by CRP-cAMP and GalM controls spacer acquisition and interference. *Nucleic acids research*, gkv517.
106. Lintner, N.G., Frankel, K.A., Tsutakawa, S.E., Alsbury, D.L., Copié, V., Young, M.J., Tainer, J.A. and Lawrence, C.M. (2011) The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. *Journal of molecular biology*, **405**, 939-955.
107. Louwen, R., Horst-Kreft, D., De Boer, A., Van Der Graaf, L., de Knegt, G., Hamersma, M., Heikema, A., Timms, A., Jacobs, B. and Wagenaar, J. (2013) A novel link between Campylobacter jejuni bacteriophage defence, virulence and Guillain–Barre syndrome. *European journal of clinical microbiology & infectious diseases*, **32**, 207-226.
108. Gunderson, F.F. and Cianciotto, N.P. (2013) The CRISPR-associated gene cas2 of Legionella pneumophila is required for intracellular infection of amoebae. *MBio*, **4**, e00074-00013.
109. Viswanathan, P., Murphy, K., Julien, B., Garza, A.G. and Kroos, L. (2007) Regulation of dev, an operon that includes genes essential for Myxococcus xanthus development and CRISPR-associated genes and repeats. *Journal of bacteriology*, **189**, 3738-3750.
110. Zegans, M.E., Wagner, J.C., Cady, K.C., Murphy, D.M., Hammond, J.H. and O'Toole, G.A. (2009) Interaction between bacteriophage DMS3 and host CRISPR region inhibits group behaviors of Pseudomonas aeruginosa. *Journal of bacteriology*, **191**, 210-219.
111. Cass, S., Haas, K.A., Stoll, B., Alkhnbashi, O., Sharma, K., Urlaub, H., Backofen, R., Marchfelder, A. and Bolt, E.L. (2015) The role of Cas8 in type I CRISPR interference. *Bioscience reports*.
112. Daume, M., Plagens, A. and Randau, L. (2014) DNA binding properties of the small cascade subunit csa5. *PloS one*, **9**, e105716.
113. Semenova, E., Kuznedelov, K., Datsenko, K.A., Boudry, P.M., Savitskaya, E.E., Medvedeva, S., Beloglazova, N., Logacheva, M., Yakunin, A.F. and Severinov, K. (2015) The Cas6e ribonuclease is not required for interference and adaptation by the E. coli type IE CRISPR-Cas system. *Nucleic acids research*, gkv546.
114. Maier, L.-K., Stachler, A.-E., Saunders, S.J., Backofen, R. and Marchfelder, A. (2015) An Active Immune Defense with a Minimal CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) RNA and without the Cas6 Protein. *Journal of Biological Chemistry*, **290**, 4192-4201.
115. Beloglazova, N., Kuznedelov, K., Flick, R., Datsenko, K.A., Brown, G., Popovic, A., Lemak, S., Semenova, E., Severinov, K. and Yakunin, A.F. (2014) CRISPR RNA binding and DNA target recognition by purified Cascade complexes from Escherichia coli. *Nucleic acids research*, gku1285.
116. Osawa, T., Inanaga, H., Sato, C. and Numata, T. (2015) Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog. *Molecular cell*, **58**, 418-430.
117. Makarova, K.S. and Koonin, E.V. (2013), *CRISPR-Cas Systems*. Springer, pp. 61-91.
118. Fredrickson, J.K., Zachara, J.M., Kennedy, D.W., Dong, H., Onstott, T.C., Hinman, N.W. and Li, S.-m. (1998) Biogenic iron mineralization accompanying the dissimilatory reduction of hydrous ferric oxide by a groundwater bacterium. *Geochimica et Cosmochimica Acta*, **62**, 3239-3257.

119. Miller, V.L. and Mekalanos, J.J. (1988) A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in Vibrio cholerae requires toxR. *Journal of bacteriology*, **170**, 2575-2583.
120. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning*. Cold spring harbor laboratory press New York.
121. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51-59.
122. Engler, M.J. and Richardson, C.C. (1982) 1 DNA Ligases. *The enzymes*, **15**, 3-29.
123. Remaut, E., Tsao, H. and Fiers, W. (1983) Improved plasmid vectors with a thermoinducible expression and temperature-regulated runaway replication. *Gene*, **22**, 103-113.
124. Maniatis, T., Fritsch, E. and Sambrook, J. (1989) Strategies for cloning in plasmid vectors, p. 1.53–1.73. *Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY*.
125. Storts, D. (1993) Promega technical manual 24 fmol DNA sequencing system (TM024). *Promega Corporation, Madison*.
126. Inoue, H., Nojima, H. and Okayama, H. (1990) High efficiency transformation of Escherichia coli with plasmids. *Gene*, **96**, 23-28.
127. Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry*, **72**, 248-254.
128. Rudolph, R. and Lilie, H. (1996) In vitro folding of inclusion body proteins. *The FASEB Journal*, **10**, 49-56.
129. Lilie, H., Schwarz, E. and Rudolph, R. (1998) Advances in refolding of proteins produced in E. coli. *Current opinion in biotechnology*, **9**, 497-501.
130. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *nature*, **227**, 680-685.

Appendix

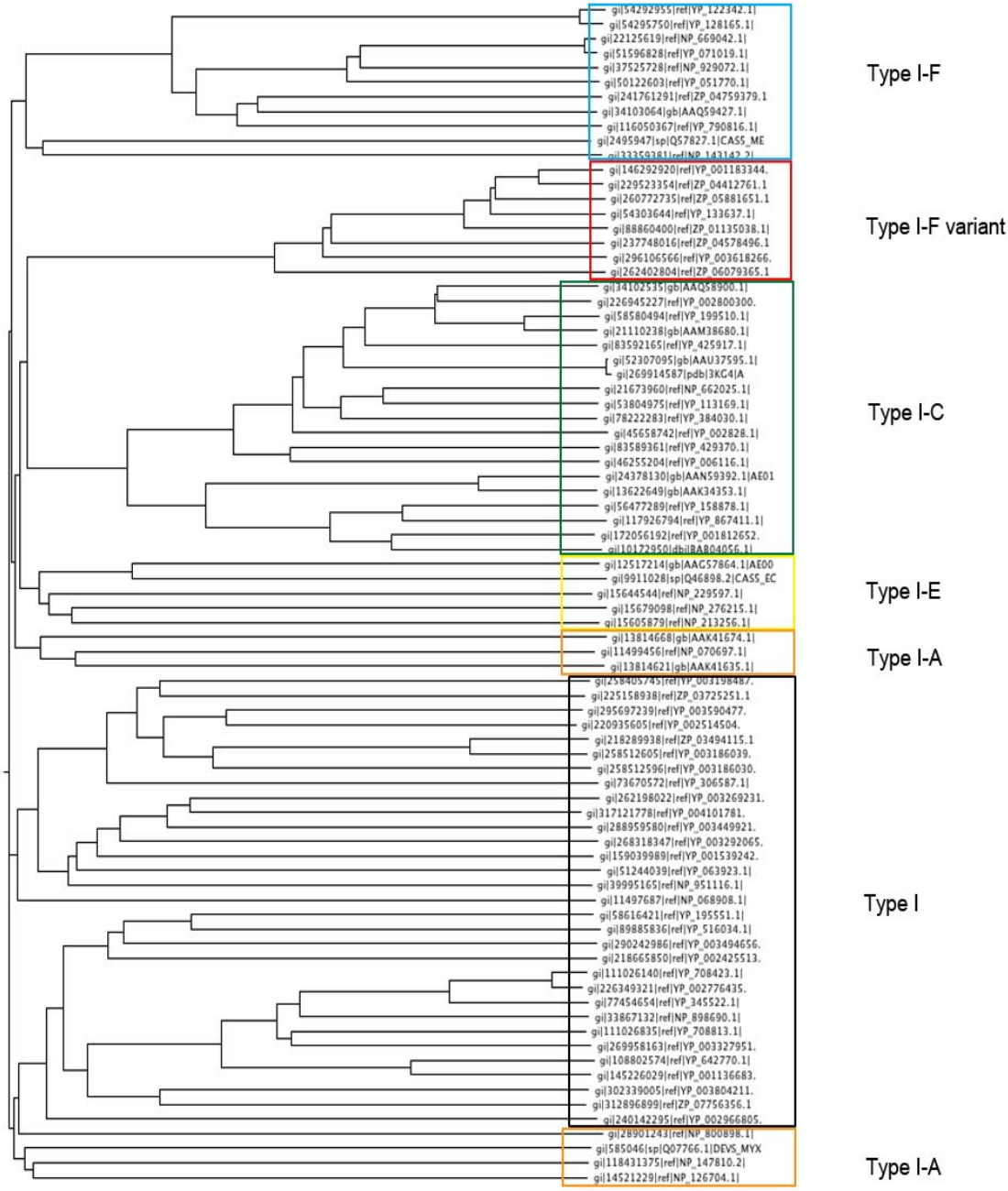
Appendix 1:

Phylogenetic analysis of Cas1821 and its homologs with proteins from the Cas7 family (protein sequences obtained from Makarova *et al*., 2011). Cas1821 and its homologs were found to cluster together and shared an evolutionary relationship with Cas7 from Type I-F.



Appendix 2:

Phylogenetic analysis of Cas1822 and its homologs with proteins from Cas5 protein family (protein sequences obtained from Makarova *et al*., 2011). Cas1822 and its homologs were found to cluster together and share an evolutionary relationship with Cas5 from Type I-C.



Appendix 3:

List of proteins identified by mass spectrometry after Co-IP studies of *Cas1821-His-HA* and *ΔH-NS* strains of *S. putrefaciens* CN-32. In yellow are the proteins that were identified for both the strains.

	Protein name
	List of proteins identified for *His-HA-Cas1821* strain
1	Signal recognition particle-docking protein FtsY
2	DNA gyrase subunit B
3	ATP-dependent protease La
4	Putative ABC transporter ATP-binding protein
5	Tonb-dependent receptor
6	30S ribosomal protein S1
7	Elongation factor G
8	Atpase
9	GTP-binding protein LepA
10	Preprotein translocase subunit SecA
11	Molecular chaperone DnaK
12	ATP-dependent helicase HepA
13	Amidophosphoribosyltransferase
14	Carbamoyl-phosphate synthase L chain, ATP-binding
15	Thiamine biosynthesis protein ThiI
16	DEAD/DEAH box helicase domain-containing protein
17	CTP synthetase
18	DNA mismatch repair protein MutS
19	1-deoxy-D-xylulose-5-phosphate synthase
20	Ribose-phosphate pyrophosphokinase
21	Hypothetical protein Sputcn32_0667
22	Molybdopterin oxidoreductase
23	Dihydrolipoamide acetyltransferase
25	Hypothetical protein Sputcn32_1449
26	GTP-binding protein TypA
27	DNA-dependent helicase II
28	Transcription termination factor Rho
29	MotA/TolQ/ExbB proton channel
30	Gtpase ObgE
31	Cell division protein FtsZ
32	Outer membrane channel protein
33	Aromatic hydrocarbon degradation membrane protein
34	FliI/YscN family atpase
35	Porin
36	30S ribosomal protein S3
37	Ribose-phosphate pyrophosphokinase
38	Flagellin domain-containing protein
39	Rod shape-determining protein MreB
40	Hypothetical protein Sputcn32_1821
41	Flagellin domain-containing protein
42	Ompa/MotB domain-containing protein

43	Recombinase A
	List of proteins identified for *ΔH-NS* strain
1	ATP-dependent protease La
2	Putative ABC transporter ATP-binding protein
3	30S ribosomal protein S1
4	DNA gyrase subunit B
5	Signal recognition particle-docking protein FtsY
6	Elongation factor G
7	Ribose-phosphate pyrophosphokinase
8	GTP-binding protein LepA
9	dihydrolipoamide acetyltransferase
10	Amidophosphoribosyltransferase
11	CTP synthetase
12	TonB-dependent receptor
13	Preprotein translocase subunit SecA
14	Prolyl-tRNA synthetase
15	ATP-dependent helicase HepA
16	GTPase ObgE
17	Cell division protein FtsZ
18	Transcription termination factor Rho
19	ATP-dependent protease ATP-binding subunit HslU
20	Signal recognition particle protein
21	Aromatic hydrocarbon degradation membrane protein
22	Aldehyde dehydrogenase
23	Na(+)-translocating NADH-quinone reductase subunit F
24	Na(+)-translocating NADH-quinone reductase subunit A
25	MotA/TolQ/ExbB proton channel
26	DEAD/DEAH box helicase domain-containing protein
27	Porin
28	Flagellin domain-containing protein
29	Ribose-phosphate pyrophosphokinase
30	Flagellin domain-containing protein
31	Rod shape-determining protein MreB
32	Recombinase A
33	OmpA/MotB domain-containing protein
34	Septum site-determining protein MinD

Abgrenzung der Eigenleistung

Die in dieser Arbeit präsentierten Ergebnisse wurden von mir selbständig ohne andere als die hier aufgeführte Hilfe durchgeführt. Im Folgenden werden weitere an dieser Arbeit beteiligten Personen sowie deren experimentellen Beiträge genannt:

Patrick Pausch
Hat im Rahmen seiner Bachelorarbeit die Konstrukte pET20b+*cas6*, pUC19+spacer3-repeat-spacer4 und pUC19+hammerhead crRNA1 erstellt. Bild 2.5B stammt aus seiner Bachelorarbeit.

Daniel Gleditzsch
Hat im Rahmen seiner Masterarbeit Bild 2.21 angefertigt.

Prof. Dr. Andreas Klingl
Hat als Kollaborationspartner elektronenmikroskopische Aufnahmen der Cas7-Filamente angefertigt und Bild 2.11 bereitgestellt.

Dr. Thomas Heimerl
Hat als Kollaborationspartner elektronenmikroskopische Aufnahmen des Typ-I-F Cascade-Komplexes angefertigt und Bild 2.21 bereitgestellt.

AG Thormann
Hat im Rahmen unserer Kollaboration die Plasmide pNPTS138-R6KT und pBBR1MCS2 sowie die *Escherichia coli* Stämme WM3064 und DH5α λpir und die *Shewanella putrefaciens* Stämme CN-32, Δ*Cas1*, Δ*Cas2-Cas3*, Δ*Cas1821*, Δ*Cas1822*, Δ*H-NS,Cas1821-His-HA*, Δ*H-NS,Cas1822-His-HA* und Δ*H-NS,Cas6-His-HA* angefertigt.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr. Lennart Randau for his encouragement, guidance and advice throughout the course of this work. I would also like to thank him for his patience and motivation during the time of research and writing of the thesis. I am grateful to Dr. André Plagens for his guidance, suggestion and the time we spent on all the wonderful scientific and non-scientific discussions. I also take this opportunity to thank all the past and present members of the Randau lab, for their help, support and the wonderful working environment during the course of this work. I would like to say a big thank you to all the wonderful friends and people I have met at the MPI and in Marburg. A special thank you needs to be mentioned to Judith Zöphel, Andreas Su, Bastian Kemmerich, Philipp Erchinger, Michael Daume and Hagen Richter for all the fun, help and support during the last few years. Lastly, I would like to thank all my friends, Dr. Adhip Mandal and family, Janardan Venkatrama and family, my brothers, Srinidhi Dwarakanth and Dr. Badri Rajagopal and my parents. Without their understanding and encouragement, this work would not have been possible.

Curriculum vitae