Characterization of the Type IV CRISPR-Cas system of *Aromatoleum aromaticum* EbN1

Dissertation der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

Ahsen Özcan

München, 2019

Characterization of the Type IV CRISPR-Cas system of *Aromatoleum aromaticum* EbN1

Dissertation der Fakultät für Biologie

der Ludwig-Maximilians-Universität München

Ahsen Özcan

München, 2019

Diese Dissertation wurde angefertigt

unter der Leitung von Prof. Dr. Kai Papenfort

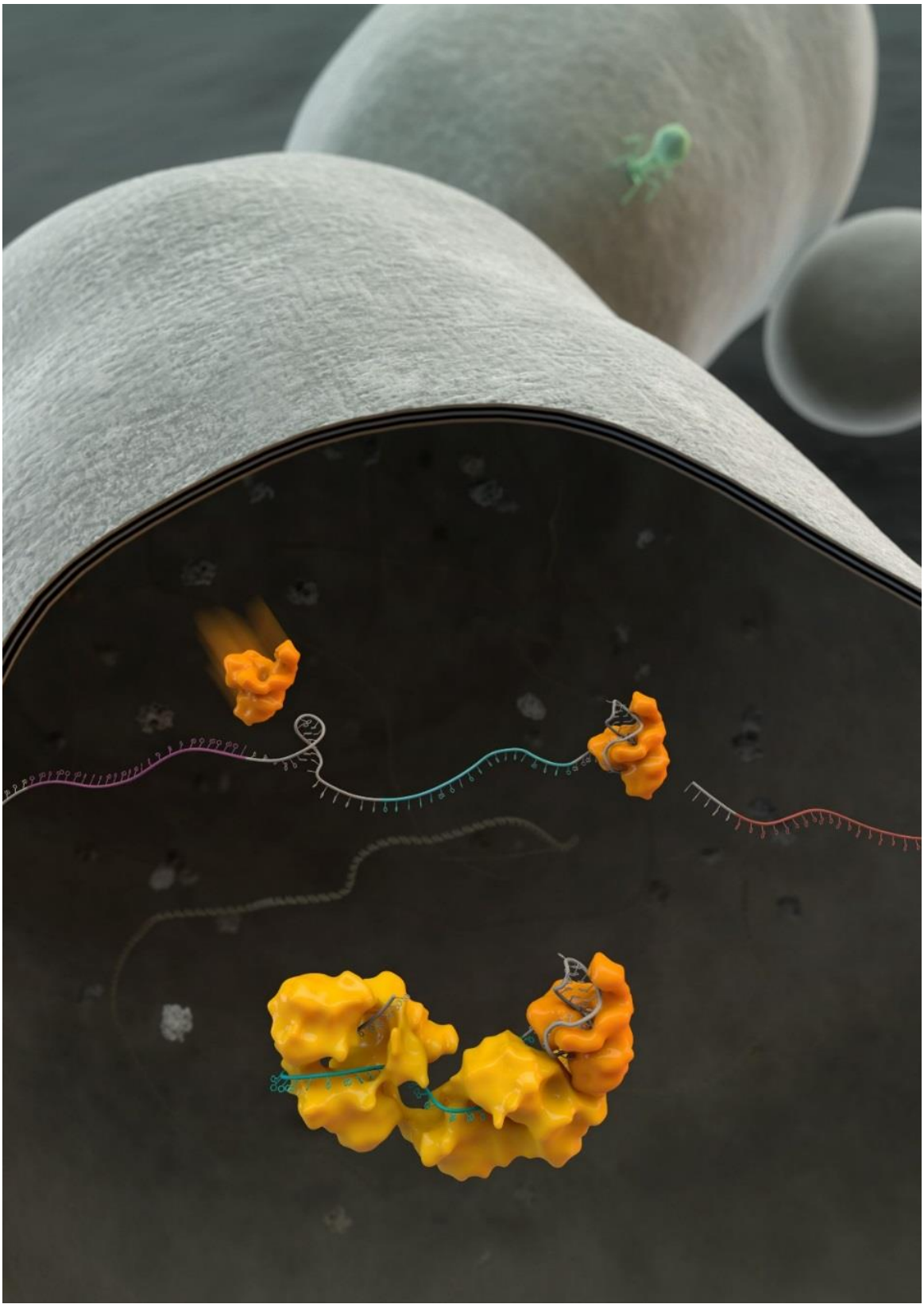
im Bereich von Department Biologie I

an der Ludwig-Maximilians-Universität München

Erstgutachter :	Prof. Dr. Kai Papenfort
Zweitgutachter:	Prof. Dr. Heinrich Jung
Tag der Abgabe:	11.06.2019
Tag der Mündlichen Prüfung:	10.09.2019

Parts of this work were published in:

Özcan A, Pausch P, Linden A, Wulf A, Schühle K, Heider J, Urlaub H, Heimerl T, Bange G, Randau, L. (2019). Type IV CRISPR RNA processing and effector complex formation in *Aromatoleum aromaticum*. *Nature Microbiology, 4*(1), 89-96.



Csf5 processes type IV pre-crRNA. Mature crRNAs are then incorporated into a type IV effector complex. Artistic Credit: Thomas Splettstößer

Table of Contents

Abstract **1**

1. Introduction **2**
1.1 Defense Systems in Prokaryotes 2
1.2 Prokaryotic Adaptive Immunity: The CRISPR-Cas System 6
1.3 CRISPR-Cas Diversity and Classification 9
1.4 The Type IV CRISPR-Cas System 13
1.5 Type IV associated DinG helicase (Csf4) 15
1.6 CRISPR-Cas Systems Beyond Defense Function 16
1.7 Aims of the Study 22
2. Results **23**
2.1 Gene organization and biogenesis of crRNAs in *Aromatoleum aromaticum* EbN1 23
2.2 Csf5 is resposible for type IV crRNA maturation 25
2.3 Structure of Csf5 in complex with a processed crRNA 3`-tag 27
2.4 Identification of active site residues of the Csf5 29
2.5 Csf2 (Cas7) is an RNA wrapping backbone protein 30
2.6 Formation of a type IV crRNP protein 31
2.7 Protein-protein and protein-RNA interaction sites of the type IV crRNP 33
2.8 Investigation of the nuclease activity of the type IV crRNP complex 36
2.9 Type IV CRISPR-Cas protospacers suggest PAM sequence conservation 38
2.10 Type IV CRISPR-Cas PAM identification in *A.aromaticum* EbN1 39
2.11 DinG (Csf4) helicase is not a stable component of the type IV crRNP complex 41
2.12 DinG (Csf4) helicase interactions with type IV crRNP components 43
2.13 Type IV associated DinG helicase (Csf4) retains ssDNA in vivo 46
3. Discussion **49**
3.1 Type IV crRNA biogenesis and maturation 49
3.2 Formation of the type IV crRNP complex and its features 51
3.3 Investigation of PAM requirements for the type IV CRISPR-Cas system 53
3.4 Putative functional roles of type IV CRISPR-Cas systems 54

3.5 Putative functional roles of type IV CRISPR-Cas associated DinG helicases 55
4. Materials and Methods 57
4.1 Material and sources of supply..... .. 57
4.2 Strains and culture conditions 62
4.3 Oligonucleotides, plasmids and constructed recombinant vectors..... 63
4.4 Working with DNA 71
4.5 Working with RNA 75
4.6 Biochemical methods 76
5. References 81
Contributions 88
Abbreviations 91
Curriculum Vitae 94
Acknowledgements 98
Salutary Declaration 99

Abstract

CRISPR-Cas systems provide prokaryotic adaptive immunity against invading agents, which also stimulated the development of indispensable tools in biological research. CRISPR-Cas systems show remarkable diversity and encompass two classes, six types and 16 subtypes. Among these six computationally classified CRISPR-Cas types, type IV CRISPR-Cas systems remained the only one without experimental data. In this study, we provide the first experimental characterization of a type IV CRISPR-Cas system using *Aromatoleum aromaticum* EbN1 as a model organism.

The *cas* genes and a minimal CRISPR array of the *A. aromaticum* EbN1 (type IV CRISPR-Cas system) were transferred into *Escherichia coli* BL21 AI to uncover the RNA and protein components of this CRISPR-Cas type. Type IV crRNAs were shown to yield unusually short 7 nucleotide 5′-repeat tags and stable 3′ hairpin structures. A unique Cas6 variant (Csf5) was identified that generates crRNAs that are specifically incorporated into type IV CRISPR–ribonucleoprotein (crRNP) complexes. Structures of RNA-bound Csf5 were obtained and the active site of the enzyme was determined. Recombinant production and purification of the type IV Cas proteins, together with electron microscopy, revealed that Csf2 acts as a helical backbone for type IV crRNPs that include Csf5, Csf3 and a large subunit (Csf1). Mass spectrometry analyses identified the cRNPs' protein–protein and protein–RNA contact sites. A possible PAM-sequence dependent DNA targeting mechanism of this complex and the involvement of a type IV CRISPR-Cas associated DinG helicase are discussed.

1. Introduction

1.1 Defense Systems in Prokaryotes

The most abundant biological entities on earth are viruses, which outnumber prokaryotes by a magnitude of 10-100. Prokaryotes have evolved diverse and elaborate antiviral defense systems in order to protect themselves from these parasitic agents. The co-evolution of host and viral genomes promotes their genetic diversification and the emergence of efficient defense strategies in prokaryotes. Beside parasitic agents, the compact genomes of bacteria and archaea are also under constant attack by mobile genetic elements (MGE), such as transposons. Therefore, defense mechanisms also support the survival of the cell by maintaining host genomes [2] [3] [4, 5] [6] [7].

To date, identified prokaryotic defense systems can be grouped into four categories based on targeting stage of the phage infection. These categories include (i) immunity by surface modification, (ii) blockage of phage genome injection, (iii) inactivation of injected phage genomes and [8] dormancy induction or programmed cell death (figure 1). In addition, recent and new discoveries about prokaryotic defense systems are also discussed.

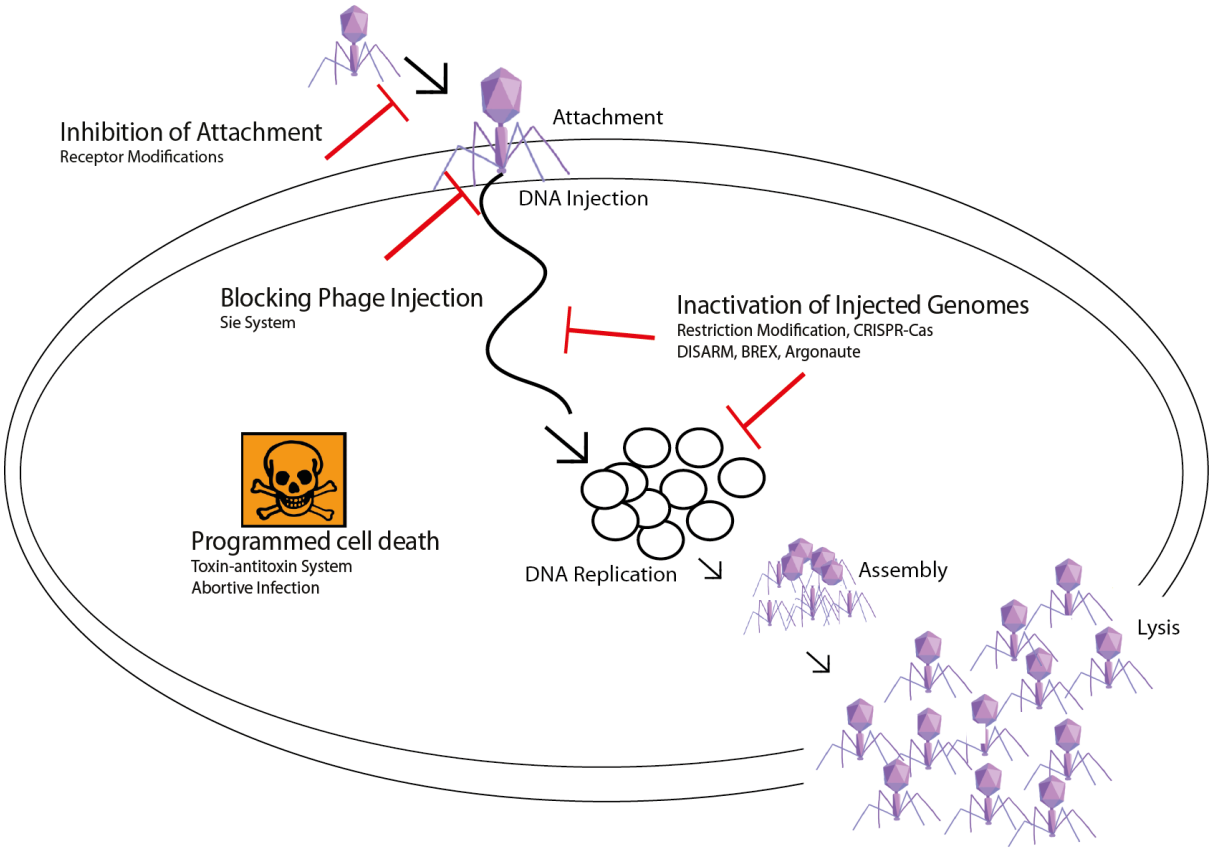


Figure 1: Overview of bacterial defense strategies. Defense strategies target different stages of the phage life cycle and can be classified into four groups: inhibition of attachment, blocking of phage genome injection, inactivation of injected phage genomes and cell dormancy or programmed cell death.

1.1.1 Immunity by surface modification and blocking of phage genome injection:

The first step in infection is the attachment of the phage to bacterial surface receptors. Bacteria have developed several strategies to prevent phage absorption which includes the physical masking of receptors by biofilm formation and the loss or downregulation of host receptors to block phage entry points. Another similar strategy is the blocking of the phage genome injection after attachment. An example is the superinfection exclusion systems which utilize membrane-associated proteins to block phage DNA injection after virus attachment [4, 6, 9] [10, 11] [12] [13] [8].

1.1.2 Inactivation of injected phage genomes:

Phage genomes that were injected into cells can be cleaved or modified by intracellular immune mechanisms. These defense mechanisms aim to prevent phage replication or release by inactivating its genetic material. Prime examples in this category are the CRISPR-Cas systems (detailed description in section 1.2), restriction-modification (R-M), DISARM, BREX and Argonaute proteins (pAgos) [9] [14, 15].

1.1.2.1 Restriction-Modification, DISARM and BREX

Restriction-modification (R/M) systems consist of a restriction endonuclease that cleaves specific sequences within foreign DNA molecules and a DNA methyltransferase that modifies the target sequences in the host genome to prevent self-cleavage. Invading nucleic acids are degraded by restriction enzymes as they lack methylated sequences. Some viruses evade this process by encoding their own methyltransferase or blocking active sites of the restriction enzymes necessitating the evolution of diversified R/M systems [16-21] [22].

A recently discovered defense system named DISARM (Defense island system associated with restriction-modification) includes restriction-modification systems, which deviate from other known R/M systems by their module composition and additional putative components. DISARM is suggested to provide protection against diverse phages while allowing phage absorption [14, 23]. However, its mechanism of action still remains to be elucidated. Another newly discovered phage defense system, named BREX (Bacteriophage Exclusion), also acts after phage absorption by targeting the phage replication pathways. The BREX system includes a set of genes, which exhibit protease, alkaline phosphatase, RNA binding, DNA methylase, and ATPase domains. Unlike R/M systems, the BREX system does not degrade the phage nucleic acids suggesting a novel type of activity. Self/non-self-discrimination was

found to exist via methylation of specific sequence patterns (TAGGAG) of the bacterial genome. However, the exact mechanism of BREX system is not understood [15] [8].

Interestingly, a recent study from Doron et al., 2018 revealed new defense systems. In this study, over 45.000 bacterial and archaeal genomes and more than 120 million genes were analyzed in the microbial pan-genomes. Using a guilt-by-association approach, nine previously unknown anti-phage systems and one anti-plasmid system were detected.These systems were named based on Greek mythology (Thoeris, Hachiman, Shedu, Gabija, Septu, Lamassu, Zarya, Kiwa, and Drunantia). However, their complete mechanism of action is not yet characterized [23].

1.1.2.2 Prokaryotic Argonaute proteins

Argonaute proteins are present in all domains of life and are key enzymes in the RNA interference pathways in eukaryotes. It has been found that genes encoding for Argonaute proteins also exist in the defense islands of prokaryotic genomes and play a role in innate immunity. Many prokaryotic Argonaute proteins preferentially bind DNA guides and direct the degradation of nucleic acids by their catalytic domains. In rare cases, DNAse activity without a DNA guide has been reported and the self-genome is protected by chromatinization of DNA. Plasmids have been shown to be potential targets of prokaryotic Argonaute proteins [24-27] [28] [29].

1.1.3 Dormancy induction and programmed cell death

Suicidal response to infection is an altruistic behavior following phage infection in cells. Cell death prevents phage replication and inhibits the spread of phages to neighboring cells. The toxin-antitoxin (TA) systems and abortive infection (Abi) systems cause cell suicide or dormancy in response to a virus infection. In TA systems, a toxin remains reversibly inactivated by an antitoxin component under normal growth conditions. However, the

antitoxin is deactivated under stress conditions. Unlike the toxins, antitoxins are metabolically unstable, thereby allowing sufficient amount of free toxin to accumulate and resulting in cell death [30, 31] [32].

Another phage resistance based cell suicide strategy is abortive infection system (Abi). Upon phage infection, the intracellular sensor molecule RexA is activated in the host. RexA subsequently activates the membrane-anchored ion channel protein RexB, which results in a sudden drop in cellular ATP levels. This aborts ATP-dependent cellular processes and leads to cellular suicide [33, 34] [35, 36].

1.2 Prokaryotic Adaptive Immunity: The CRISPR-Cas System

Unusual repetitive sequences of the CRISPR-Cas system were initially recognized by Ishino et al. in the late 1980s, but their function in the system was described six years later by F. Mojica in *Haloferax mediterranei* [37] [38]. The discovery of the CRISPR-Cas system as an adaptive immune system led to the realization that adaptive immunity is not only a feature of higher organisms but potentially an ancient mode of defense system found in prokaryotes [39] [40, 41] [42]. Compared to other defense systems, the most remarkable feature of the CRISPR-Cas system is its ability to store information of past infections and the capacity to use them when necessary. Stored infection memory hereby enables the detection and degradation of foreign nucleic acids of invading agents (such as viruses, plasmids, transposons) in order to defend the integrity of the prokaryotic genomes [43] [44].

CRISPR-Cas systems are found in around 90% of all archaea and in about 50% of all bacterial species. Almost all CRISPR-Cas systems share a two component architecture on their genomes, a CRISPR-array (Clustered Regularly Interspaced Short Palindromic Repeats) and a set of *cas* genes (CRISPR associated genes) that is usually found in close proximity. CRISPR arrays consist of an AT-rich leader sequence, a series of palindromic DNA segments

(repeats), which are interspaced by variable DNA sequences (spacers). The source of the spacers are often previously encountered MGEs (virus, plasmids, transposons) [45-47].

CRISPR-Cas mediated adaptive immunity consists of three connected stages: adaptation (acquisition), expression and interference. In adaptation, a new spacer derived from an invading sequence is inserted into CRISPR array (figure 2). PAM (protospacer adjacent motif: 2-7 nucleotide long motifs) initiate recognition of protospacers in foreign DNA by Cas proteins that catalyze their excision and integration in a CRISPR array between the leader and the first repeat sequence of the CRISPR array. The proteins Cas1, Cas2, and Cas4 were shown to be the main proteins involved in this adaptation process. Two dimers of the Cas1 integrase are complexed by a central Cas2 dimer, creating a hetero-hexameric complex that catalyzes (proto) spacer integration [47] [48-53] [54]. Recent studies demonstrated that the protein Cas4 also plays a role in adaptation by adding PAM specificity during protospacer selection by the Cas1-Cas2 complex. Therefore, the presence of Cas4 ensures the acquisition of functional spacers [55] [54, 56] [57, 58].

During expression, the CRISPR array is transcribed into long precursor CRISPR RNAs (pre-crRNAs) and subsequently processed by different RNA processing enzymes that vary between CRISPR-Cas types (see section 1.3). The pre-crRNA transcript can be processed by cellular RNAses (RNase III in type II) or dedicated Cas proteins (Cas6 family endoribonucleases in type I). Mature crRNAs, consist of a single spacer flanked by partial repeats [59] [60].

In the interference stage, mature crRNAs interact with Cas proteins and form a CRISPR ribonucleoprotein complex (crRNP). Depending on the class of the CRISPR-Cas system, single or multiple Cas proteins interact with a single crRNA. Even within types of the same family, crRNP arrangements vary. However, they exhibit a common mechanistic role, which is the identification and degradation of foreign genetic elements [61]. Individual crRNPs scan

DNA strands for the presence of PAM sequences and interrogate adjacent sequences for complementary to the crRNA spacer which defines a bona fide target. Binding of crRNPs to the targeted region creates a DNA-RNA hetero-duplex, termed R-loop. R-loop formation subsequently triggers either internal nuclease activity (Cas9) or the recruitment of Cas nucleases (Cas3) in order to destroy foreign mobile genetic elements [62-65] (figure 2).

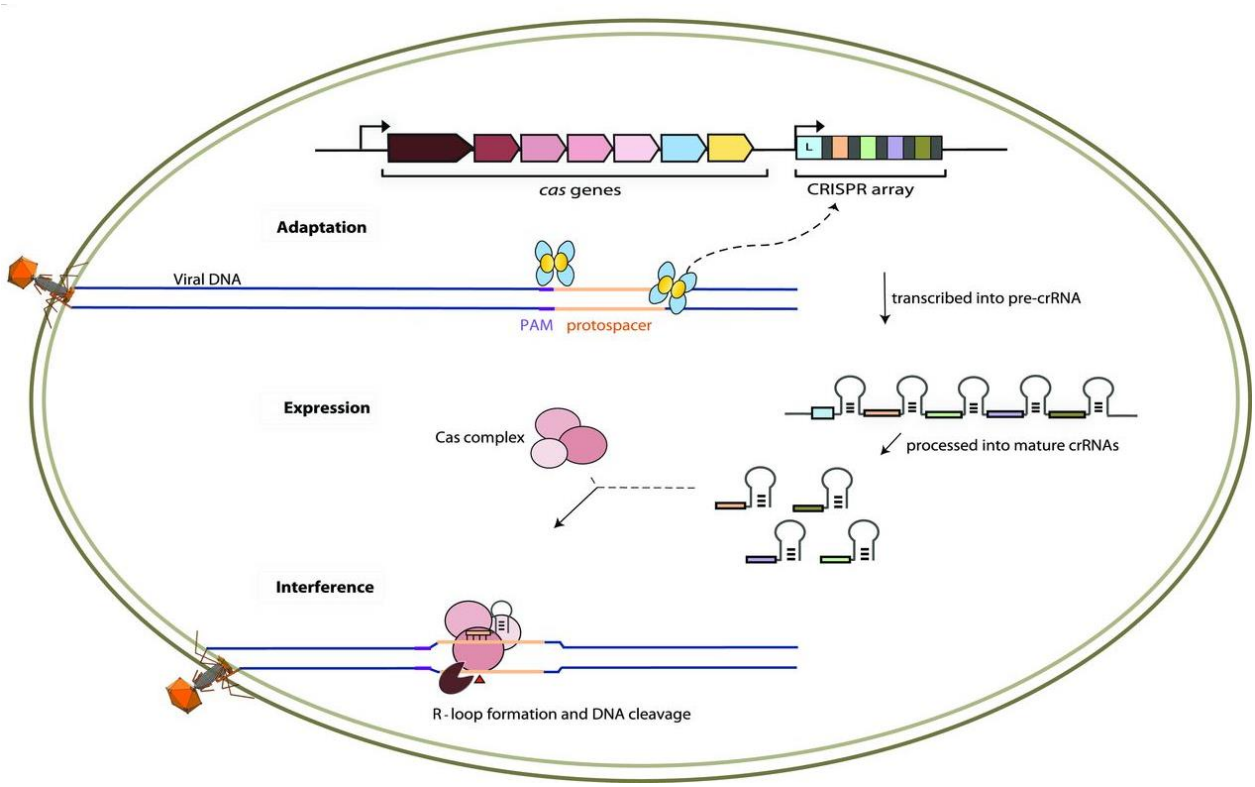


Figure 2: Three stages of the CRISPR-Cas mediated adaptive immunity. At the stage of adaptation, the proteins Cas1 (blue) and Cas2 (yellow) recognize the PAM sequence (purple) of foreign DNA and excise the protospacer to integrate it as a new spacer into the growing CRISPR array. During expression, the CRISPR array is transcribed, processed into small crRNAs, loaded onto Cas complex proteins (pink) which results in the formation of a crRNP complex. At the stage of interference, the crRNP complex scans DNA for PAM sequences (purple). Adjacent dsDNA sequences are interrogated for complementarity to the crRNA spacer, which creates an R-loop structure. The formation of the R-loop structure recruits cellular nucleases, which then degrade the foreign DNA. Figure modified from [68].

1.3 CRISPR-Cas Diversity and Classification

In the current classification scheme, CRISPR-Cas systems are divided into two classes, six types, and 33 subtypes. This classification is based on the genome architecture of the CRISPR-Cas loci and the presence of signature proteins that are associated with each type [66] [67].

The two major classes are established based on the architecture of the effector complex. Class I features multi-domain protein crRNPs, whereas Class II is distinguished by a crRNP containing a single effector protein (figure 3). Class I is further subdivided into types I, III and IV and their effector complexes are composed of four to seven different Cas proteins assembled around a central crRNA skeleton with a varying stoichiometry (figure 4). Class II effector complex functions are carried out by a single multi-domain protein (e.g. Cas9) and further subdivided into types II, V and IV [68] [66] [67] [58].

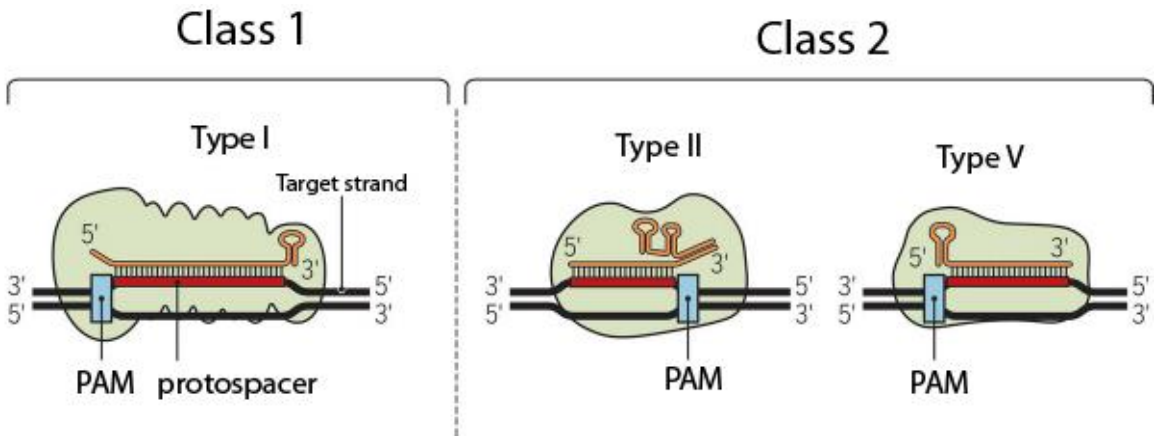


Figure 3: Simplified representation of the two major CRISPR-Cas classes based on the architecture of the effector complexes. Class I (Type I, III, IV effector complexes are formed by multiple proteins, whereas Class II complexes (Type II, V, VI) are only composed of a single effector protein. Similar functional roles of the effector complexes are achieved by varying target strand and PAM sequence location and the different structure of the RNA guide. Figure modified from [66].

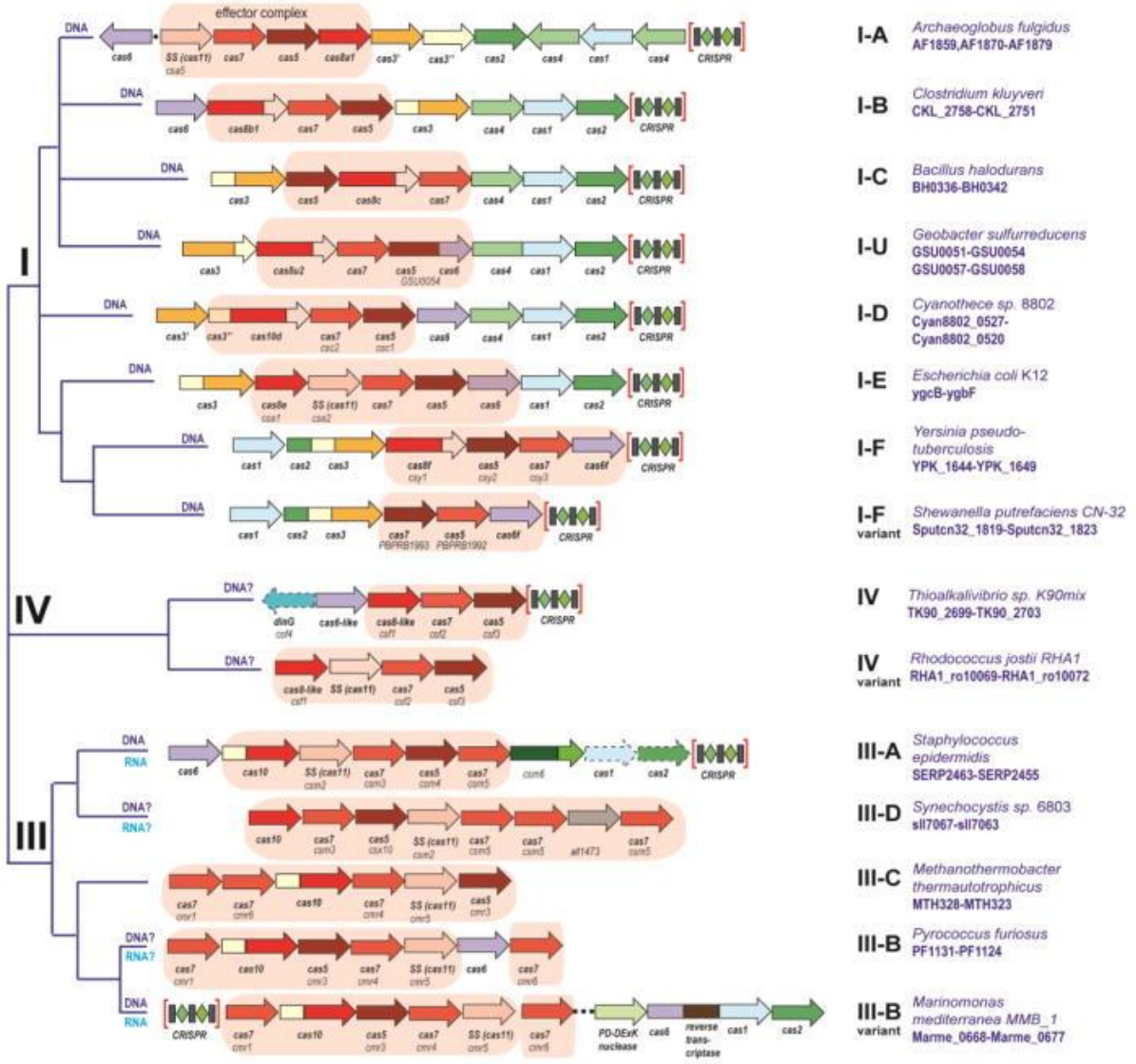


Figure 4: Classification of the Class I CRISPR-Cas systems. Class1 systems contains three types (type I, III and IV) and various subtypes. Effector complex modules are shaded in light red. Figure taken from [67].

The structural composition of the Class I effector complexes, which mediate the interference stage of the CRISPR-Cas defense, features common similarities among different types (type I, type III and type IV). In type I systems, the effector complex is known as CRISPR-associated complex for antiviral defense (Cascade), whereas in type III-A and type III-B systems effector complexes are known as Csm and Cmr complexes. The core Cas proteins found in class I effector complexes share a common structural feature, which is the presence of an

RNA recognition motif. In type I systems, the core complex proteins are Cas7, Cas5, and Cas6.

Cas5 and Cas7 are non-catalytic RNA binding proteins. Cas5 is responsible for 5'-repeat tag capping of the crRNA and Cas7 is responsible for multi-subunit backbone formation. Cas6 is an endoribonuclease that is responsible for crRNA processing with an exception of the type I-C system where crRNA processing is catalyzed by Cas5 [61, 69]. In addition to core effector complex components, two additional proteins named large subunit and small subunit are designated to the crRNP complexes. The large subunit is mostly present in type I and type III crRNPs (exception for type I-fv) and responsible for PAM recognition, whereas the presence of a small subunit, which is responsible for non-target strand stabilization, is rarer [67] [61].

Type I Cascade complexes identify foreign DNA and subsequently recruit the signature protein Cas3 nuclease/helicase in order to degrade foreign DNA. In type III systems, the interference process is coupled to transcription in a PAM independent manner. The multiprotein type III complexes target the nascent mRNA and then degrades DNA nearby [70] [71] [72-74] [61] [75] [76].

Electron microscopy images of type I-E *Escherichia coli* Cascade and type III *Thermus thermophilus* complexes show remarkable similarities based on their shape and Cascade architecture (figure5). Therefore, it is suggested that multi-subunit effector complexes evolved before the separation of Type I and Type III CRISPR-Cas branches [77] [78].

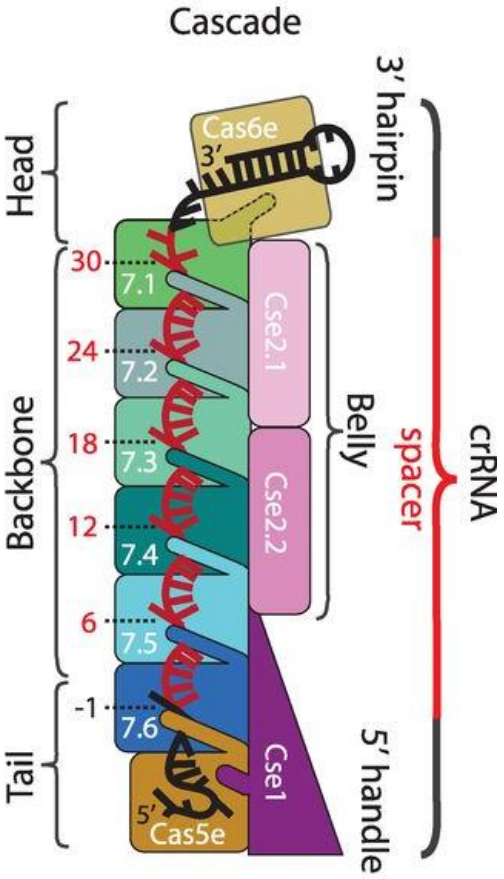


Figure 5: Representation of the basic architecture of the *E.coli* Type I-E Cascade based on its crystal structure. Cas proteins assemble around a single crRNA with an uneven stoichiometry of five different proteins. The backbone protein (Cas7e) wraps spacer of the crRNA with 6-nt intervals. Cas5e and Cas6e proteins hold the 5' handle and 3' hairpin structures of the crRNA, respectively. Figure adapted from [79].

1.4 The Type IV CRISPR-Cas System

Among 47 % of analyzed bacterial and archaeal genomes possessing CRISPR-Cas loci, less than 1% carry the rare type IV system [67] [80]. The minimalistic architecture of type IV loci is distinct from all type I and type III subtypes by lacking adaptation and target cleavage genes, thereby justifying its classification as a new type. The protein Csf1 (large subunit) was assigned as a signature protein for this distinct type [67] [80, 81] (Figure 6).

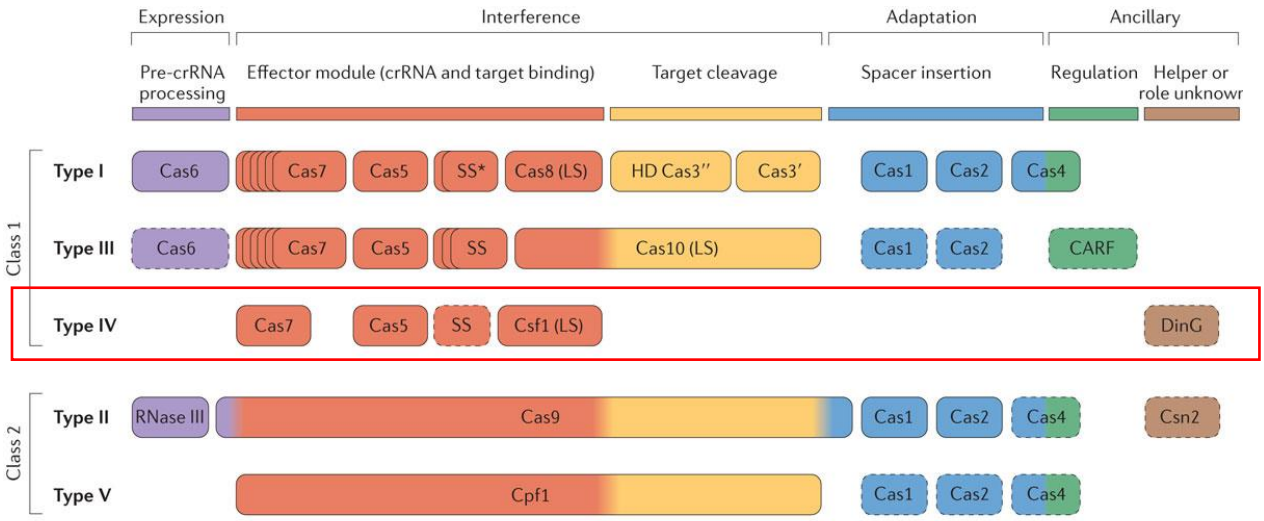


Figure 6: General scheme of the organization of the CRISPR-Cas systems: Cas proteins are classified based on their general function. Type IV CRISPR-Cas (indicated in red) interference complexes are proposed to be composed of Cas7, Cas5, small subunits, and a large subunit. Adaptation proteins Cas1, Cas2, and target cleavage proteins (e.g.Cas3) are missing in this system. The presence of a small subunit (SS) and a DinG protein are optional (shaded line). Figure adapted from [68].

There are unique features of the type IV system which deviate from other known CRISPR-Cas systems. These include its nearly universal localization on plasmids, the apparent loss of Cas1, and Cas2 adaptation genes and the presence of signature large subunit protein. Type IV systems are also often not found in close proximity to a CRISPR array or even encoded in a genome entirely lacking CRISPR arrays [67].

Type IV systems are predicted to encode minimal multi-subunit crRNA-effector complexes that consist of a large subunit protein [82], Cas5 (Csf3), Cas7 (Csf2) and occasionally a

putative small subunit or Cas6 family endoribonuclease (Csf5). There are two variants of type IV CRISPR-Cas systems, which are distinguished by the presence of the DinG helicase. The genomic architecture of type IV-A CRISPR-Cas systems contains additional putative genes for DinG and a type IV repeat specific Cas6 homolog (Csf5). However, type IV-B systems lack additional genes and a uniform CRISPR locus. Our model organism *Aromatoleum aromaticum* EbN1 possesses a type IV-A CRISPR-Cas locus on one of the two megaplasmids with a complete set of *cas* genes (*csf1, csf2, csf3, csf4,* and *csf5*) and a regular CRISPR-array. Additionally, *A. aromaticum* EbN1 contains a type I-C CRISPR-Cas system on the genome and DinG (Csf4); which is localized adjacent to the type IV CRISPR-Cas locus [67] [82].

1.5 Type IV associated DinG helicase (Csf4)

DinG (damage-inducible helicase G) is classified into the superfamily 2 (SF2) helicases in the sub-branch of the XPD family, which feature non-ring-formation, ATP dependency and translocation on single stranded or double stranded nucleic acids [83-85]. Most of the XPD family DNA helicases comprise an iron-sulfur cluster, which is represented as four cysteine residues located between two conserved Walker A and Walker B motifs. However, not all DinG proteins possess iron-sulfur clusters [86].

DinG helicase activity has been studied in various organisms. The deletion of *dinG* in *E. coli* did not create a severe phenotype, but a slight reduction of UV resistance was reported. The enzyme was characterized as a DNA dependent ATPase and 5'-3' DNA helicase.[87]. Further characterization suggests that DinG unwinds DNA-RNA hybrid duplexes, is active on D-loops and R-loops, and on forked structures. It was also shown that 5'-ss tails 11-15 nucleotide length are sufficient to initiate DNA duplex unwinding [88]. Similar results were also observed in *Mycobacterium tuberculosis,* where DinG exhibited an unwinding activity of substrates with 3'-overhangs [89].

Interestingly, the DinG helicase in *Staphylococcus aureus* was shown to lack an iron-sulfur cluster and to exhibit fused N-terminal exonuclease activity in the 3'-5' direction, but lacks unwinding activity [90]. The biological function of DinG in bacterial lineages is not fully understood. It is proposed that its function is dissolving R-loops in during the removal of the RNA transcripts from (stalled) replication forks. Interestingly, DinG helicases appear in bacterial lineages in association with CRISPR-Cas systems. The function of the DinG helicases in type IV CRISPR-Cas system is proposed to provide a helper role, however, experimental and functional data related to type IV CRISPR-Cas associated DinG helicases is lacking to date [81].

1.6 CRISPR-Cas Systems with Functions Beyond Defense

The main known function of the CRISPR-Cas system is RNA-guided nucleic acid detection and degradation, i.e. its role as an adaptive immune system. However, some CRISPR-Cas systems suggest to have functions beyond their canonical adaptive immune function (e.g. oxidative stress tolerance, DNA repair, host-microbe interactions or transposon guidance) [91, 92] [93] [94].

The Cas9 protein of type II systems was shown to modulate virulence and pathogenicity of the bacterial cells by regulating bacterial lipoprotein (BLP) expression at the mRNA level in *Francisella novicida*. Reduced BLP expression at the membrane protects pathogens against host immunity [95]. Similarly, in a separate study, the absence of the *cas9* gene resulted in increased swarming and reduced cytotoxic activity during infection of *Campylobacter jejuni* in human cells [96]. Another example of function beyond immunity is the involvement of CRISPR-Cas systems in DNA repair and recombination pathways. The type I-E CRISPR-Cas system of *E. coli* has been reported to be involved in DNA repair. Here, Cas1 nuclease of *E. coli* was shown to interact with various DNA recombination and DNA repair enzymes: RecB, RecC, RuvB, and UvrC. Deletion of the *cas1* was shown to increase sensitivity to DNA damage and defects in chromosome segregation [92]. Several studies have shown that abiotic stress also activates the transcription of *cas* genes in prokaryotes (*Thermoproteus tenax, Methanococcus jannaschii, Pyrococcus furious* and *Sulfolobus solfataricus*) [97, 98] [99] [100] [101].

Comparative genomic and phylogenetic analysis of minimal CRISPR-Cas systems proposed an association with distinct families of transposable elements. It was hypothesized that minimal CRISPR-Cas systems could be repurposed as a mechanism that helps transposable elements to propagate via RNA-guided transposition [102]. A minimal subtype of a type I-F CRISPR-Cas systems, which lacks the adaptation module or target cleavage Cas3 protein,

consists of only three effector proteins: a fusion protein of the large subunit (LS) with Cas5f, Cas6f, and a transposon-associated Cas7f. This minimal system is predicted to process generate mature crRNAs but lack the *cas3* target cleavage gene, which prevents a conventional RNA guided adaptive immune function. Comprehensive *in-silico* analysis shows that site-specific specialized group of Tn7 transposon family is linked to these minimal CRISPR-Cas systems (figure 6).

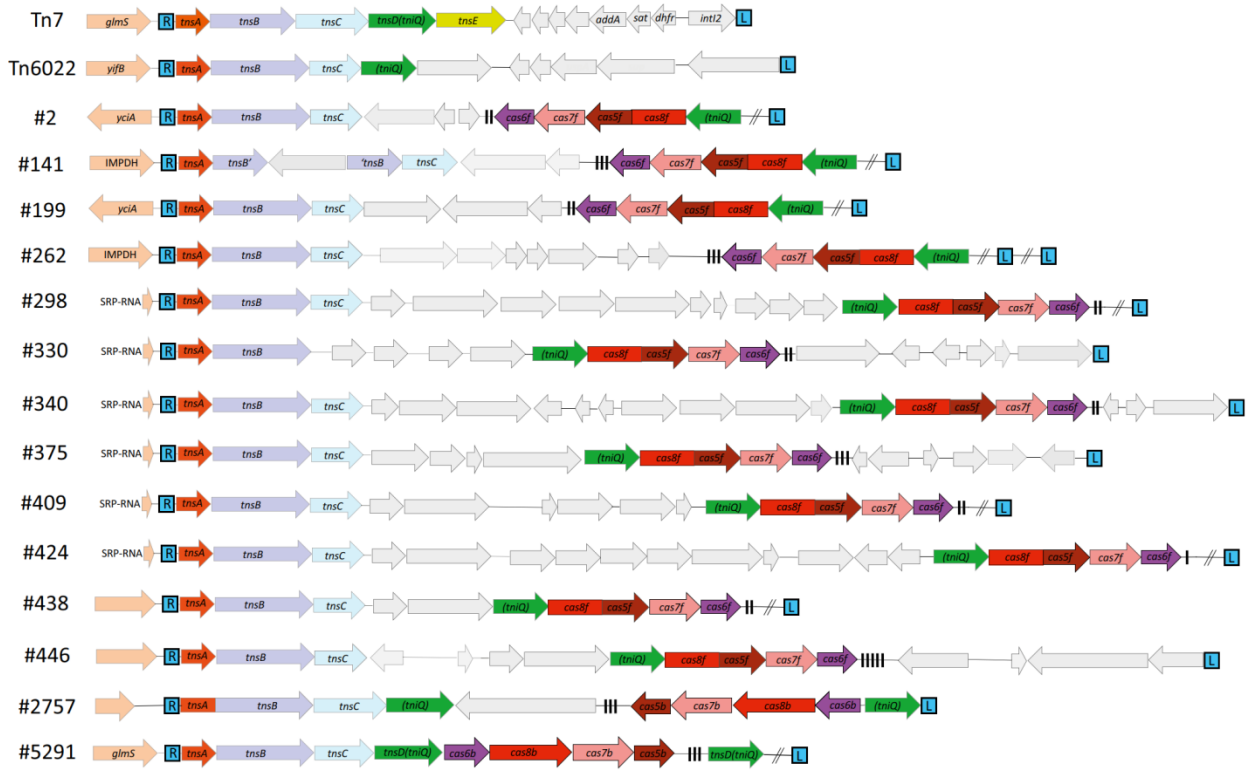


Figure 6: Schematic representation of Tn7-like transposons associated with *cas* genes (*cas6*, *cas5*, *cas7*, *cas8*). Components of the Tn7-like transposons (TnsA, TnsB, TnsC, TnsD, TniQ) are linked to *cas* genes (in a distance of 10 kb up or downstream in a genomic neighborhood). L: left flanking region, R: right flanking region. Figure adapted from [102].

1.6.1 Tn7 Based Transposition Mechanism

Bacterial Tn7 based transposition is distinguished from other transposons by having tight control over target site selection. Due to the specific site selection mechanism, Tn7 transpositions are usually not harmful to the hosts, as transposition is directed toward specific neutral sites (attTn7) which do not affect survival. A second pathway directs transposition into conjugal plasmids which helps facilitating the horizontal transfer of the elements to new host cells [103-105].

Mechanistically, transposition is carried out by five proteins; TnsA, TnsB, TnsC, TnsD, and TnsE. TnsA and TnsB provide the main transposase function when an appropriate site for insertion is identified. TnsD and Tns E are target site specifying proteins and TnsC functions as a regulator protein for communication between the transposase and targeting proteins. TnsD is a DNA binding protein and recognizes attTn7 sites, whereas TnsE recognizes specific structures on the DNA. These include replication forks which directs the core complex to replication-associated conjugal plasmids. The orientation of Tn7 transpositions is controlled by an unknown mechanism [105] [104] [106] [107] [108] (figure 7).

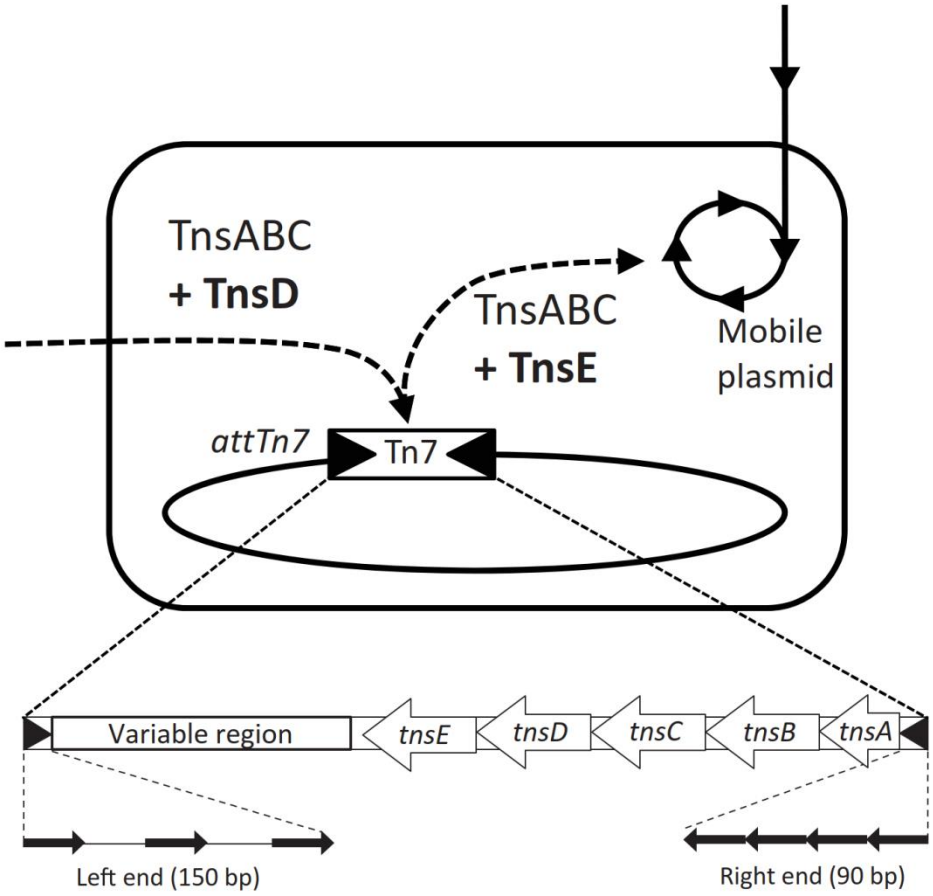


Figure 7: Schematic representation of the Tn7 transposon gene structure and transposition mechanisms. The genomic architecture of the Tn7 transposon includes left (L) and right (R) flanking sequences (depicted as triangles), a variable region which contains genes that likely benefit the host and *tnsA, B, C, D, E* site-specific transposition genes. Transposition is catalyzed by the TnsABC core complex and either TnsD or TnsE. The first pathway directs transposition to neutral attTn7 sites by Tns ABC + TnsD complex. Contrarily, the second pathway directs transposition to mobile plasmids and is accomplished by Tns ABC + TnsE. Neither pathway is likely to disrupt host genes. Figure adapted from [102].

1.6.2 Proposed Model for Tn7 Based Transposition with CRISPR-Cas Systems

Minimal subtype I-F and I-B CRISPR-Cas systems lacking adaptation and target nuclease genes (*cas3*), were identified to be associated with Tn7 transposons. These CRISPR-Cas associated transposons encode the protein TnsD which drives the transposon to attTn7 specific sites, but they were found to lack TnsE-like proteins. These proteins mediate transposition to specific structures of DNA during replication and direct Tn7 insertion into plasmids and bacteriophages. Therefore, it was proposed that minimal CRISPR-Cas systems can take the role of TnsE to enable crRNA-guided Tn7 transposition (Figure 8). Furthermore, CRISPR array analysis of Tn7 associated CRISPR-Cas systems also supported the proposed hypothesis. Some spacer matches were found inside the transposon boundaries.

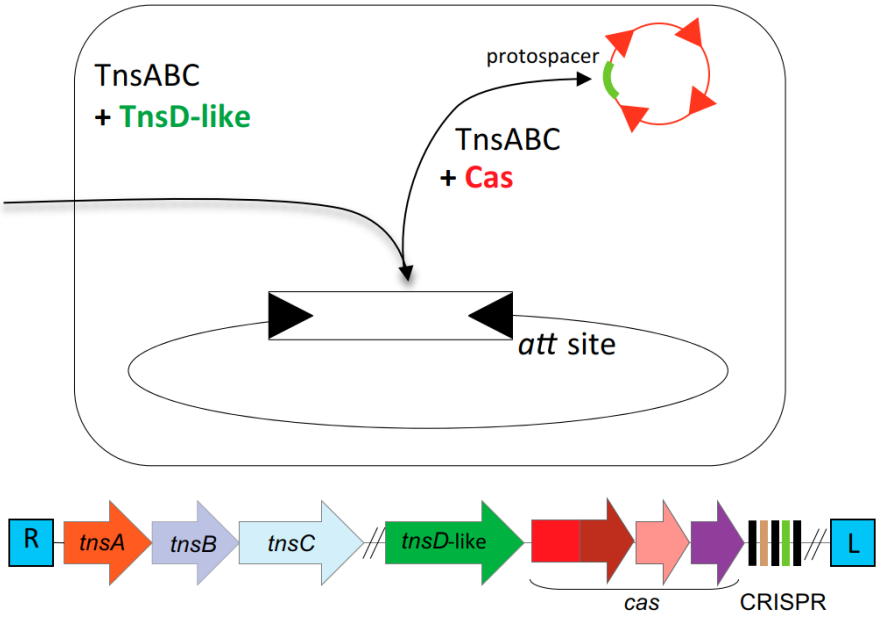


Figure 8: Proposed model for CRISPR-Cas system guided transposition. *Cas* genes and a short CRISPR array link to Tn7 transposon components. The CRISPR-Cas system replaces the TnsE function and drives transposons to mobile genetic elements by crRNA guided targeting. Figure adapted from [102].

Previously, it was shown that the gain of function mutations of TnsC regulator protein directs transposition to specific DNA structures (distorted B-form DNA). Furthermore, TnsABC+E based transposition was proposed to be induced by distortions in the target DNA. In the model, crRNP complexes induce DNA distortions, by forming R-loops during crRNA and protospacer duplex formation [102].

1.7 Aims of the Study

The functionality of type IV CRISPR-Cas systems has not been characterized experimentally and the possible roles of type IV Cas proteins have only been predicted using computational methods. Therefore, the aim of this study is to biochemically characterize the components of the type IV CRISPR-Cas system of the model organism *A. aromaticum* EbN1. Recombinant production of the *A. aromaticum* EbN1 Cas proteins Csf1, Csf2, Csf3, Csf5 (Cas6-like) and Csf4 (DinG helicase) is enabled in *E. coli* strains in the presence of minimal CRISPR arrays. This approach allows for insights into the mechanisms of (i) crRNA maturation and (ii) crRNP formation. First, the crRNA maturation pathway and the role of Csf5 should be identified. Second, it should be investigated if crRNAs are forming complexes with Cas proteins to generate Type IV-specific crRNPs that facilitates RNA-guided target searching. The presence of PAM motifs and an association with DinG proteins will be investigated for known Type IV CRISPR-Cas systems.

2. Results

2.1 Gene organization and biogenesis of crRNAs in *Aromatoleum aromaticum* EbN1

To study the composition and functionality of a model type IV CRISPR-Cas system, two model organisms were selected: *Methylobacterium extorquens* AM1 and *Aromatoleum aromaticum* EbN1. Due to observed limitations of *M. extorquens* AM1 Cas protein production and the apparent absence of a uniform CRISPR array, most of the data presented in this thesis focuses on the type IV system of *A. aromaticum* EbN1.

The model organism *A. aromaticum* EbN1 possesses a regular CRISPR array with associated Type IV cas genes (*csf1, csf2, csf3, csf5*) and DinG helicase (*csf4*) on one of its megaplasmids. In addition to the type IV CRISPR-Cas system on the megaplasmid, *A. aromaticum* possesses a type I-C CRISPR-Cas system on its chromosome.

Illumina RNA-sequencing was used to investigate the biogenesis of crRNAs in a type IV CRISPR-Cas system. Small RNAs (< 200 nt) were isolated from *A. aromaticum* EbN1 and subjected to 5′-phosphorylation and 3′-dephosphorylation established for compatibility with Illumina library preparation. Over 11 million reads were sequenced and mapped to the *A. aromaticum* EbN1 genome. Mapped reads suggest that the type IV CRISPR array was successfully transcribed and processed into small crRNAs. Interestingly, type IV crRNAs were shown to contain an unusual 5´-terminal 7-nt tag with the sequence 5´- GUUGAAG-3´. The 3´-termini of mature crRNAs were not trimmed and contained a sequence that can form a hairpin structure (figure 9).

Additionally, RNA-Seq verified crRNA processing in the subtype I-C CRISPR-Cas cluster on the chromosome. It was observed that the I-C CRISPR array is also transcribed and processed into small crRNAs. In agreement with previous studies on subtype I-C systems, crRNAs with an 11 nt 5´- terminal tag with a sequence of 5´-UGGAUUGAAAC-3` were observed. This tag is generated by an unusual, catalytically active Cas5 enzyme variant [68] [113]. RNA-Seq analysis of *Aromatoleum aromaticum* EbN1 also revealed few reads indicating low expression of *cas* genes. Transcriptional activation of the *cas* genes might require specific conditions such as stress, virus infection or external signals.

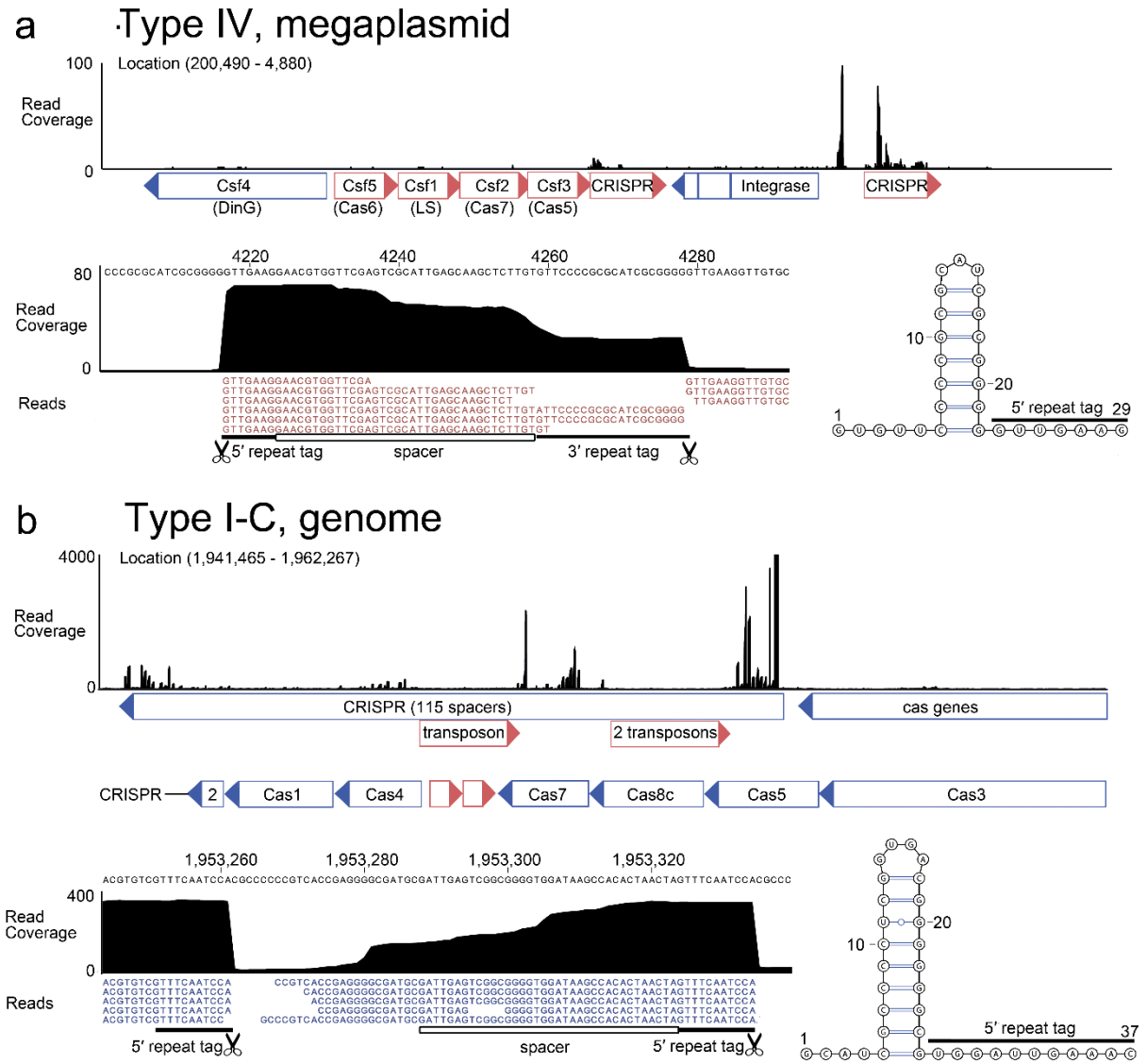


Figure 9: Maturation of *A. aromaticum* EbN1 crRNAs. Depicted are gene organization and RNA-seq coverage of the two CRISPR-Cas systems of *A. aromaticum*. Gene names and suggested Cas gene

families (LS (large subunit) of type IV CRISPR-Cas systems are given. **a.** The type IV CRISPR-Cas system contains two split CRISPR arrays with 9 and 15 spacers, respectively, which are separated by a mobile genetic element. RNA-Seq reads indicate crRNA maturation and generation of a 5′-terminal 5′-GUUGAAG-tag. Cleavage occurs at the base of the hairpin structure of the 29 nt repeats (bottom right). **b.** The type I-C CRISPR-Cas systems contain three CRISPR arrays that are interspersed with three antisense transposon sequences. RNA-Seq reads covering the indicated genomic repeat region reveal that type I-C crRNAs exhibit a 5′-terminal 5′-UGGAUUGAAAC-tag and cleavage occurs at the base of the hairpin structure of the 37 nt repeats (bottom right).

2.2 Csf5 is responsible for type IV crRNA maturation

Observation of crRNA processing yielding a unique 7 nt 5′-terminal tag raised the question of which Cas protein is responsible for crRNA maturation in type IV systems. To address this question, *in vivo* cleavage assays were conducted. All type IV *cas* genes (*csf1, csf2, csf3* & *csf5*) from *A. aromaticum* were codon optimized and expressed in *E.coli*. Beside the *cas* genes, a mini CRISPR array, consisting of a single repeat-spacer-repeat unit was expressed in the same strain. Pre-crRNA production and crRNA maturation was detected by northern blot analysis using a probe against the spacer sequence of the mini-array (figure 10).

Different Cas proteins were provided with the mini-CRISPR array in *in-vivo* cleavage assays, to identify protein responsible for crRNA maturation in type IV. A 62 nt mature and stable crRNA were observed only in the presence of Csf5 and the mini CRISPR array this suggests that Csf5 acts as crRNA endonuclease of the Cas6 family, responsible for generating the unusual 7 nt 5'-terminal repeat tag. Additionally, the hypothesis of whether Csf5 interferes with the crRNA maturation of the type I-C CRISPR array was tested. Therefore, a subtype I-C mini CRISPR array was provided with Csf5 *in vivo*. However, mature crRNA formation from the type I-C mini CRISPR array was not observed. The *in vivo* cleavage assays suggest that Csf5 is a Cas6-like endoribonuclease, which specifically processes type IV CRISPR-Cas system associated CRISPR arrays.

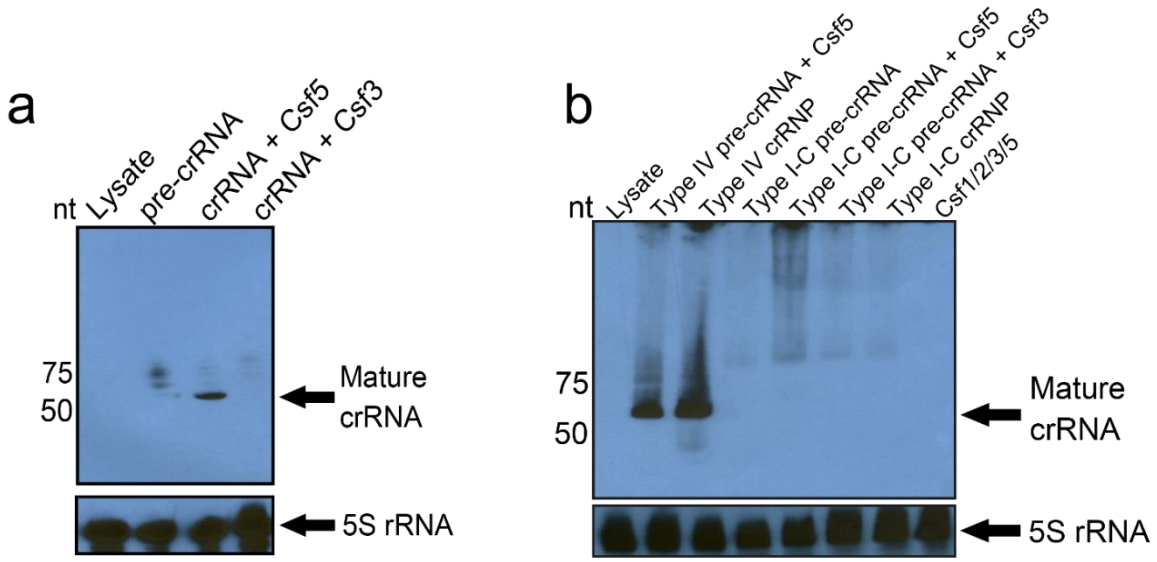


Figure 10: In-vivo cleavage assays via northern blot. Respective Cas proteins and a type IV repeat-spacer-repeat-array were produced in *E. coli* BL21-AI. Northern blots were performed with a probe against the spacer. **(a)** Mature crRNA is detected in the presence of Csf5. **(b)** Respective Cas proteins (Csf3, Csf5) and a Type IV repeat-spacer-repeat-array were produced in *E. coli* BL21-AI (Wild type (Wt)). Northern blots were performed with a probe against the spacer. Repeat-spacer-repeat transcripts with identical spacers were produced either with type IV or type I-C repeats in *E. coli* BL21-AI together with the Cas proteins (indicated in figure). Mature crRNAs were only detected with type IV repeats together with Csf5 or a crRNP complex (Csf1, Csf2, Csf3 and Csf5).

2.3 Structure of Csf5 in complex with a processed crRNA 3`-tag

Additionally, *in vitro* maturation assays were attempted with Csf5 and crRNA. However, purification of soluble Csf5 was not achieved. Soluble Csf5 could only be produced in the presence of bound crRNA, and purification by affinity and gel filtration chromatography yielded milligram amounts of recombinant Csf5 proteins in complex with a crRNA component. Purified Csf5 was therefore subjected to structure determination by X-ray crystallography, using single anomalous dispersion (SAD) due to lack of suitable search models for molecular replacements. The crystal structure revealed two monomers of the Csf5 protein associated with the crRNA 3`- repeat tag with a 1.75 Å resolution. Electron density for nucleotides 14 and 15 of the crRNA hairpin was missing in the electron density map, suggesting a degraded hairpin loop. The degraded hairpin structure was assumed to be the result of RNases during the 20°C purification procedure. This changed the purification conditions to 4°C, in order to reduce potential RNase activity. Similarly, Csf5-crRNA complex revealed two monomers of Csf5 in complex with the intact crRNA repeat tag at a resolution of 2.3 Å (figure 11).

The crystal structure of the Csf5 showed that it is not only a functional but also a structural homolog of the Cas6 superfamily. In brief, features of Csf5 exhibit common similarities to the Cas6 family endoribonucleases. Two positively charged RNA recognition motifs (i.e. the N-terminal RRM and C-terminal RRM') largely accommodate RNA contacts and enclose the endonuclease active site. Additionally, Csf5 exhibits a glycine-rich loop (G-loop) and shows a characteristic β2' and β3' hairpin structure (β-HP) that assists in positioning the scissile phosphate. The groove-binding element (GBE) was identified, which recognizes the major groove of the crRNA hairpin. Apart from highly conserved Cas6 features, Csf5 uniquely features an α-helical finger domain that connects β1 and β2 of the N-terminal RRM and

interacts with the crRNA hairpin via the crRNA repeat hairpin minor groove. This likely contributes to the specificity and stability of crRNA binding.

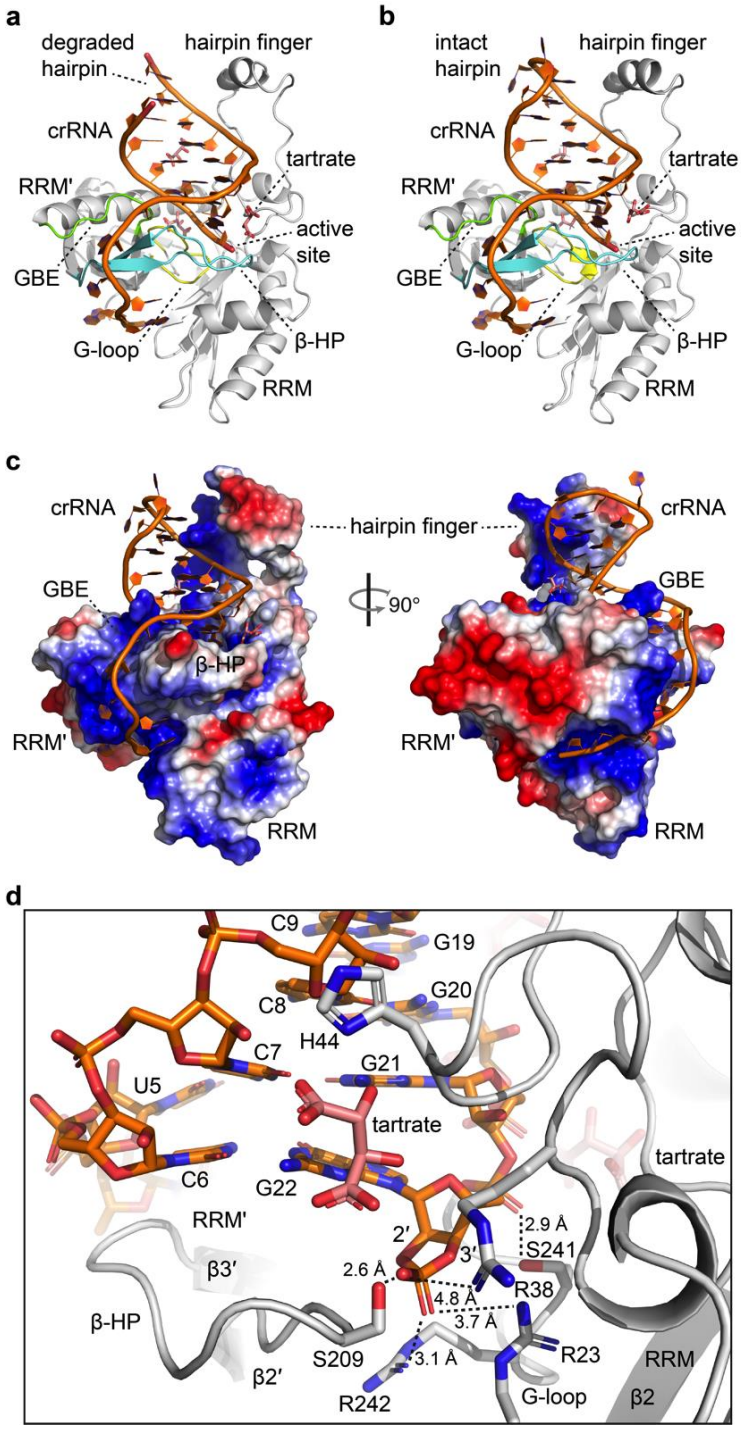


Figure 11: Crystal structures of Csf5. Crystal structures of Csf5 (grey cartoon) associated with a degraded **(a)** or an intact **(b)** crRNA 3'-repeat tag hairpin (orange cartoon). The conserved elements of the Cas6 family, GBE [92], G-loop (yellow) and β-HP (turquoise) are labeled accordingly. **(c)** Electrostatic charge distribution of Csf5 surface (intact hairpin). Blue surfaces indicate positively

charged and red surfaces indicate negatively charged regions. **(d)** Close-up view of the active site. The crRNA is shown in an orange stick representation. Csf5 is represented as a grey cartoon. Sidechains in proximity to the 2', 3' cyclic phosphate are shown in stick representation and the distances are indicated by dotted lines and labeled accordingly. The tartrate ligand is shown in a salmon-colored stick representation.

2.4 Identification of active site residues of the Csf5

The structure of the Csf5-crRNA also revealed a 2', 3' cyclic phosphate as the reaction product of the endonucleolytic crRNA repeat cleavage. Therefore, the nucleophilic attack of the scissile phosphate is more likely to be facilitated by the 2' hydroxyl group of the 5' upstream nucleotide (G22) as observed for other Cas6 endonucleases. Alanine scanning assays (figure 12a-c) revealed that the active site is mainly shaped by three arginine residues (i.e. R23, R38, and R242) that surround the 2', 3' cyclic phosphate and might participate in the

i. activation of the 2' hydroxyl group,
ii. protonation of the leaving group and
iii. stabilization of the transition state.

Among the three single mutants, R23A demonstrated the strongest effect, where double mutants of R23A with either of the other Arg residues, resulted in a complete loss of pre-crRNA processing. Serine 209 and 241 of the β-HP and the G-loop, respectively, might further contribute to the positioning of the scissile phosphate and binding of the crRNA repeat at the active site. We further identified tartrate, originating from the crystallization solution, bound in close proximity to G22 and histidine 44 (figure 11d). Hence, the tartrate ligand might affect the position of histidine 44. However, Csf5 with an H44A mutation still produces mature crRNA *in vivo* which suggest that a catalytic histidine is not required for cleavage activity of Csf5 (figure 12b).

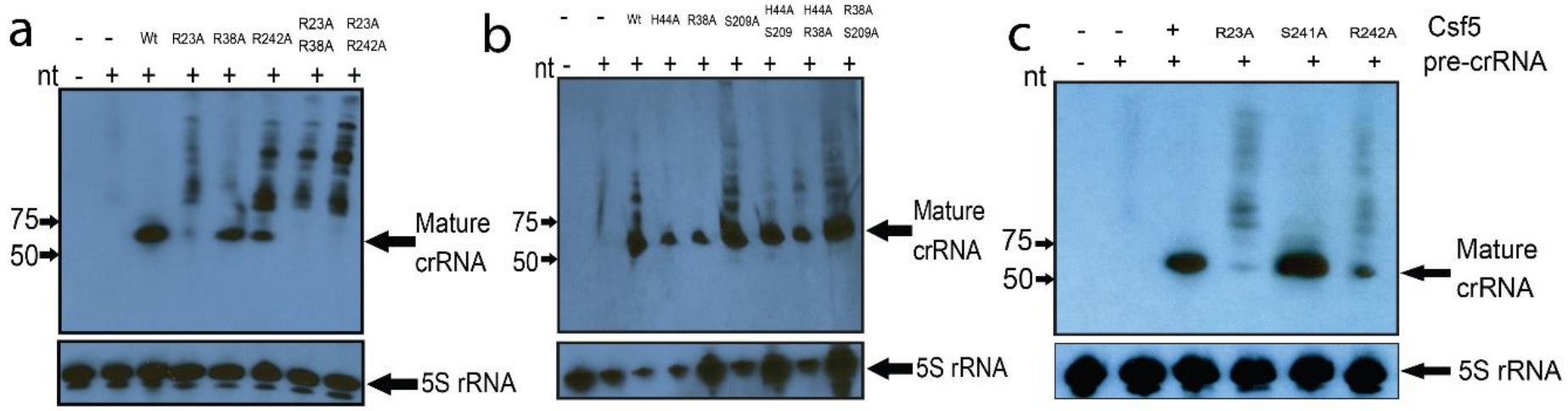


Figure 12. Csf5 crRNA maturation activity by alanine scanning assay. (**a)** R23A, R38A and R242A variants and double mutation variants of Csf5 were screened for crRNA maturation. The mature crRNA formation was disrupted when double mutations of arginine residues were applied. (**b)** H44A, R38A and S209A variants Csf5 were screened for crRNA maturation. (**c)** R23A, S241A and R242A variants of Csf5 were assayed for crRNA production. The results in a – c are representatives of three experiments.

2.5 Csf2 (Cas7) oligomerizes and binds RNA

The most abundant protein in type I and type III effector complexes are the Cas7 family backbone proteins. Cas7 family proteins unspecifically oligomerize around mature crRNA and constitute a central backbone of the surveillance complexes [109, 110]. To investigate the backbone forming protein in the type IV CRISPR-Cas system, recombinant Csf2 protein was over-expressed in *E.coli* and purified by affinity chromatography followed by size exclusion chromatography. In size exclusion chromatography (SEC), high molecular weight structures were observed at the void volume and analyzed by various methods: RNase and DNase assays, SDS and Urea-PAGE and transmission electron microscopy (TEM). These void volume fractions were also observed to contain RNA bound Csf2 oligomers. Transmission electron microscopy revealed long helical filament structures of variable length (figure 13a). In type I systems, Cas7 proteins were shown build similar unspecific RNA wrapped helical filaments and to bind to the spacer component of the crRNA in the Cascade (e.g. analyzed for type I-E, type I-A, type I-Fv) complexes[109, 110] [79].

Therefore, it is possible that Csf2 is a Cas7-family protein which exhibits unspecific binding to the crRNA spacer regions and forms the backbone of the type IV crRNP complex.

2.6 Formation of a type IV crRNP Complex

Observation of crRNA transcription and maturation suggests crRNA guided protein complex formation. In order to test formation of type IV crRNP complex formation, all Cas proteins produced from the type IV loci (Csf1, Csf2, Csf3 and Csf5) were co-produced with the mini-CRIPSR array in *E. coli* BL21-AI and purified by affinity purification of the his-tagged Csf5 followed by SEC. A central peak fraction was observed to contain a stable complex with all Cas proteins with increased abundance of Csf2 (Cas7-like) protein and a single mature crRNA (figure 13b). Our results indicate that the type IV cas7-like variant is represented by several

subunits in a type IV crRNP, which is in an agreement with its proposed role as a backbone-forming protein that mediates crRNA spacer binding. Protein and RNA components of the purified complexes were identified via mass spectrometry (Cas proteins) and northern blot analyses (crRNA). Negative staining transmission electron microscopy analysis of these samples revealed crescent-shaped structures with dimensions similar to the size of previously identified Cascade complexes (figure 13b) (ie. type I-E, type I-C) [111].

Additionally, the first peak was identified as Csf2 (Cas7-like) filaments and the third peak was observed to represent the Csf5 endoribonuclease. It was attempted to produce type IV crRNP complex with crRNA of the *A. aromaticum* type I-C mini-CRISPR array. Csf5 and crRNP stability was observed to rely on the presence of the type IV crRNAs. Type I-C pre-crRNA appeared to be not processed by Csf5 (figure 10b). Therefore, type IV CRISPR-Cas activity is proposed to rely on Cascade-like complexes and guidance by a specific type IV-associated crRNA component.

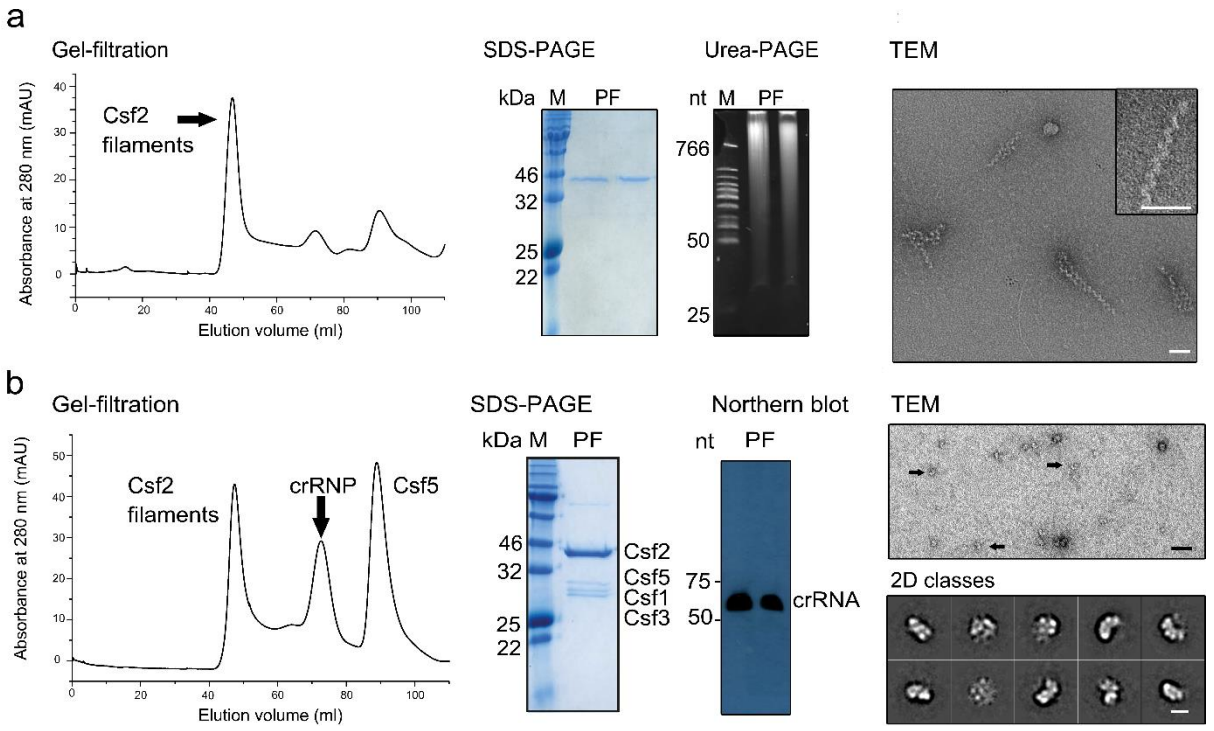


Figure 13: *A. aromaticum* Type IV Cas protein production in *E. coli*. Cas proteins were purified via affinity and gel-filtration chromatography. The peak fractions (PF, arrow) of the gel-filtration chromatograms were loaded onto an SDS-PAGE to detect purified Cas proteins and onto Urea-PAGE to identify RNA components. (**a)** Csf2 purification depicted large helical filaments in TEM (scale bars, 50 nm). Co-purified RNAs were detected with urea–PAGE and ethidium bromide staining (**b)** A crRNP complex was purified and visualized by TEM (scale bar: 50 nm). Mature crRNAs were detected via Northern blot analyses. TEM revealed crescent-shaped complexes (arrows). Approximately 5400 particles were used for 2D averaging 10 classes (scale bar: 20 nm). The results are representatives of three experiments.

2.7 Protein-protein and protein-RNA interaction sites in Type IV crRNPs

To gain more insights into the type IV crRNP complex formed by Csf1, Csf2, Csf3, Csf5 and crRNA, protein-protein crosslinking and protein-RNA crosslinking approaches followed by mass spectrometry was used to identify interaction sites based on structural composition of the complex. Protein-protein interactions were identified by applying the BS3 chemical cross-linker, which creates covalent bonds between adjacent lysine residues of proteins located in close proximity (figure 14). BS3 cross-linked complexes were then tryptically digested and crosslinked peptides were analyzed via mass spectrometry. Highest enrichment of intra-protein (homo-multimeric links) crosslinks in Csf2 was observed suggesting that Csf2 occurs in multiples in type IV surveillance complexes. Csf2 was also the only identified protein in the complex to interact with all other Cas proteins (Csf 1,2,3,5). This observation further strengthens the observation that Csf2 acts backbone in the complex. Surprisingly, no inter-crosslinks between Csf1 and Csf3 were detected, which suggests that these proteins are not in close proximity with other in the assembled type IV crRNP complex architecture. However, crosslinks between Csf1 and Csf5 what suggests a localization near the 3′ end of the crRNA in the complex. In summary, protein-protein crosslinks further verify the structural composition of the type IV crRNP complex *in vitro*.

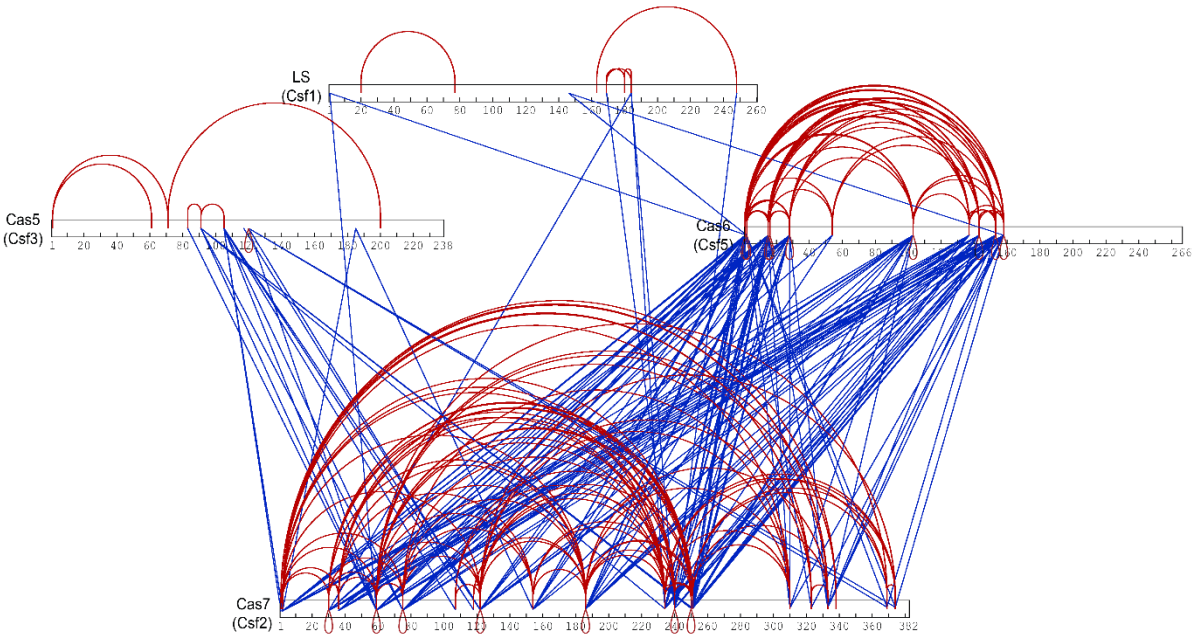


Figure 14: Type IV crRNP complex protein-protein crosslinking. Purified type IV crRNP complexes were crosslinked by a BS3 chemical cross-linker followed by mass spectrometry. Observed intra-protein (red), inter-protein (blue) and homomultimeric (loops) crosslinks are indicated. The results are from a single HPLC-MS measurements.

Interaction sites between crRNA and Cas protein components were further identified. Purified type IV crRNP complexes were identified by UV irradiation of type IV crRNP complexes. Tryptically, digested peptide-oligonucleotide crosslinks were enriched and analyzed by HPLC-mass spectrometry (Table 1). Csf2, Csf3 and Csf5 were found to interact with the crRNA. Among crosslinked residues, the crystal structure indicated that the amino acid K157 interacts with U2 of the repeat sequence. Additional crosslinking sites of Csf5, S134, K142, Y144, K157 are located within the RRM' and RRM motif, suggesting Csf2 and Csf5 interactions with the spacer (figure 15).

Peptide	Protein	Adduct
kTNPYDIAR	Csf3 (Cas5)	U-H2O
GLDDPkR	Csf5 (Cas6)	U-H2O
GLTkTAENTTQVASR	Csf5 (Cas6)	U-H2O
FALPAGKk	Csf5 (Cas6)	U-H2O
ALPLSNGSEAVESGDGyK	Csf2 (Cas7)	U
ASkHVYM(ox)GLFGGGTR	Csf2 (Cas7)	U-H2O
GVGVELEQHTLsIK	Csf5 (Cas6)	U
GLLLcALR	Csf2 (Cas7)	U
cELAALTVPSLMEFFTGK	Csf2 (Cas7)	U
RyFVR	Csf5 (Cas6)	CU
hIvRVDDVTR	Csf2 (Cas7)	U
GyIEQHGHEPPPPTEEQR	Csf5 (Cas6)	U
IIADHpAMSTGR	Csf5 (Cas6)	U
GLDDPkR	Csf5 (Cas6)	U-H2O
LTEVrHIVR	Csf2 (Cas7)	U
m(Ox)GQQHLLR	Csf5 (Cas6)	U-H2O
IADEIAkAAEK	Csf1 (LS)	U-H2O
LTAPM(ox)HVADSIQYSLDDKGyVVR	Csf2 (Cas7)	U
GyWFAGNLTSR	Csf5 (Cas6)	AU
VkIPYFPANDLR	Csf2 (Cas7)	U-H2O
YkPLLEAATLPAIQIVTQR	Csf5 (Cas6)	U-H2O
kSDLANVVSVQSIIAGTELYVR	Csf2 (Cas7)	U-H2O

Table 1: Type IV crRNA interacting Cas protein residues. Recombinant Type IV crRNP complexes were subjected to ultraviolet light crosslinking, followed by HPLC-MS. Specific amino acid residues (red) of the Cas proteins were found to interact with the indicated crRNA bases (adduct).

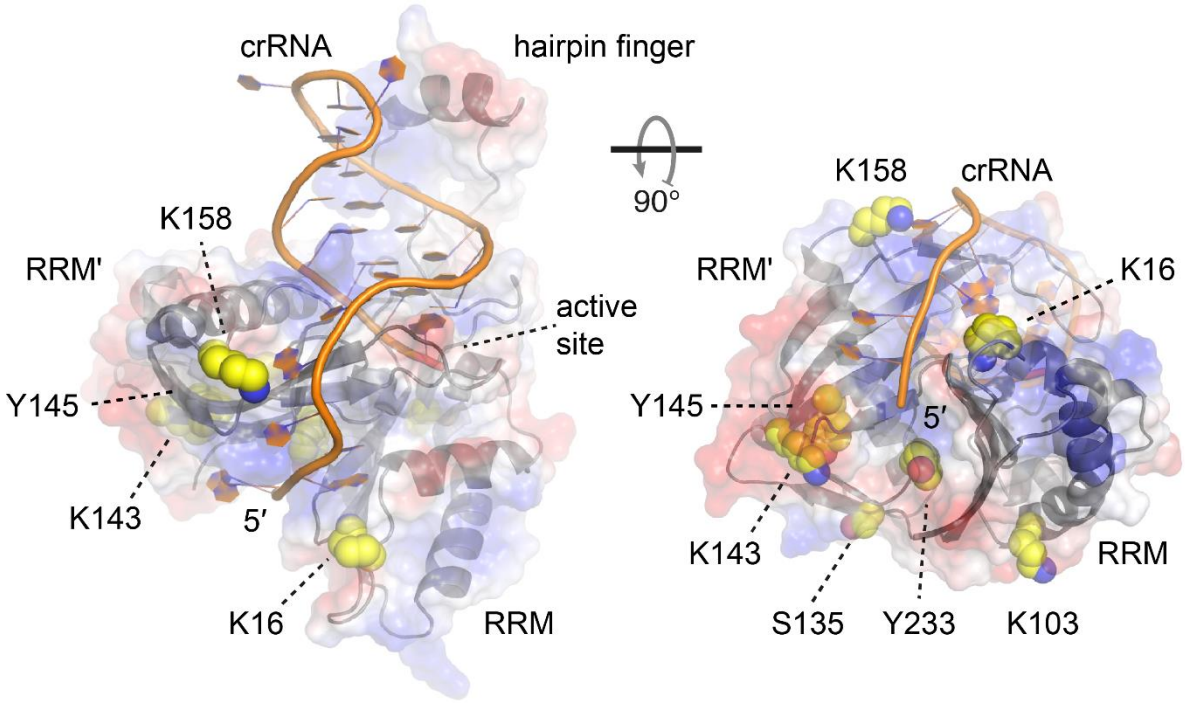


Figure 15: Csf5 – crRNA interactions. Mapping of elucidated cross-links (yellow) in UV irradiation of crRNP-crRNA onto the crystal structure of Csf5.

2.8 Investigation of the nuclease activity of the type IV crRNP complex

Nuclease activity of the type IV ribonucleoprotein complex was tested via *in-vitro* nuclease assays. Various concentrations (0-200 nM) of the purified type IV crRNP complex were incubated with 43 nt long 5′-terminal radiolabeled complementary ssDNA, RNA and non-complementary ssDNA substrates. The nuclease activity of substrates was then checked with a 10% native denaturing gel and visualized using a Phosphor-Imager. Under *in-vitro* conditions, no nuclease activity of the native type IV crRNP complex was detected (figure 16). Different metal ions were also tested for the activation of nucleases (Mg^{+2}, Ca^{+2}, Mn^{+2}). However, cleavage activity remained undetectable in the presence of these metal ions.

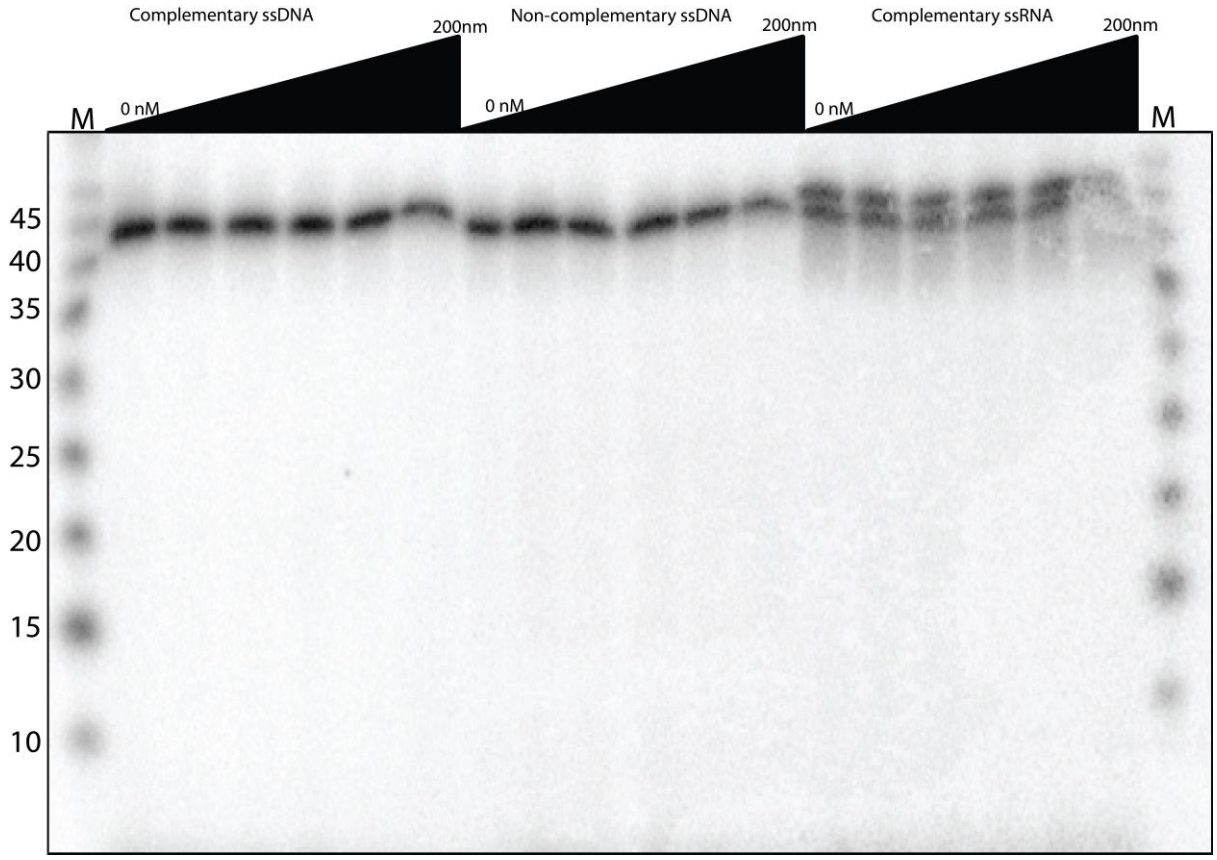


Figure 16: Type IV crRNP complex nuclease activity test *in vitro*. Nuclease activity of the crRNP complex (increasing amounts) with 5´-terminal radiolabeled complementary ssDNA, RNA, and non-complementary DNA was tested. Cleavage activity of the type IV crRNP complex was not observed under the tested conditions *in vitro*.

2.9 Type IV CRISPR-Cas protospacers suggest PAM sequence conservation

Initially, the conservation of the PAM sequence in the type IV system was addressed in-situ via computational detection. Type IV spacers from various microbial organisms which have matching homologous sequences, were previously identified [112]. Matching spacers of type IV systems from various organisms were kindly shared by Sergei Shmakov. Identified spacer hits were mainly located in viral genomes and ORFs of prokaryotic genomes. In order to detect the PAM sequence from matching protospacers, the computation programs CRISPRDetect and CRISPRTarget were used [113]. Surrounding sequences from matching protospacers were then plotted in a sequence logo structure (figure 17). The combined analysis demonstrated no consensus PAM sequence. However, a species-based analysis identified a species-specific type IV PAM sequence conservation (Detailed PAM analysis of all type IV organisms were investigated in the thesis of Marcus Ziemann).

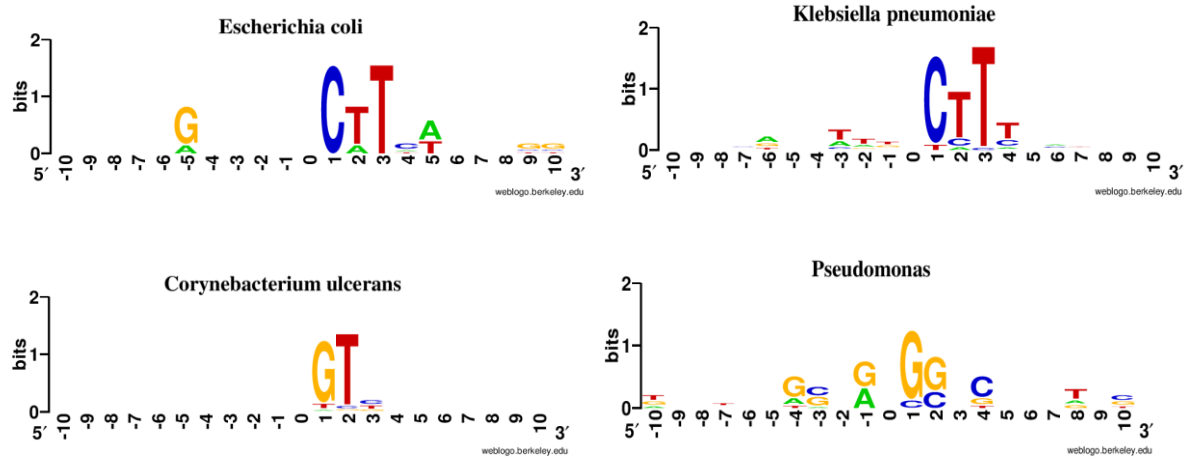


Figure 17: Conservation of PAM sequences in type IV CRISPR-related protospacers. Protospacer targets (mainly phage genomes) of various species with Type IV-associated CRISPR arrays were analyzed computationally. Sequence logos of the protospacer-adjacent sequences were constructed, depicting conservation of nucleotides adjacent to the protospacers (position 0). (Bioinformatic analyses were performed by Marcus Ziemann).

2.10 Type IV CRISPR-Cas PAM identification in *A. aromaticum* EbN1

Different screening methods were applied to identify possible PAMs for the type IV CRISPR-Cas system of *A. aromaticum* EbN1. Both computational and experimental methods were applied to detect the type IV PAM sequence from *A. aromaticum* EbN1. Similar to the previous approach, type IV spacer targets of *A. aromaticum* EbN1 were detected by CRISPRTarget [113]. Spacer 10 was identified to target a plasmid from *Thauera* sp. (figure 18).

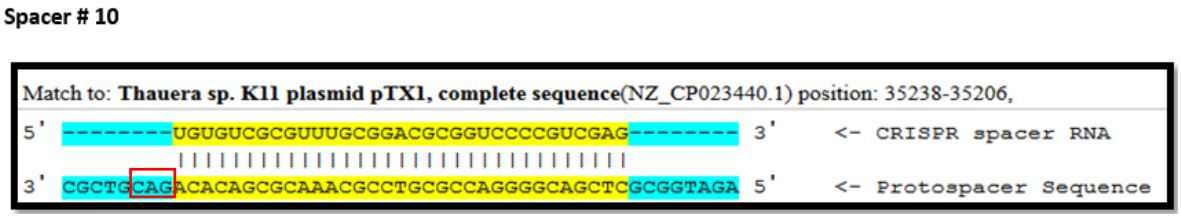


Figure 18: ***A. aromaticum EbN1*** **protospacer targets**. Spacer 10 was found to have matching sequence to a plasmid of *Thauera* sp. The three nucleotide motif at the location of the 5′ crRNA handle (5′-GUUGAAG3′) is observed to be 5′GAC-3′ indicated in red.

PAM depletion assays were designed to detect a variety of PAM sequences of *A. aromaticum* EbN1 *in vivo*. The assay aimed to deplete the correct PAM containing plasmids based upon crRNP targeting, which could interfere plasmid replication. In this respect, the first spacer of *A. aromaticum* EbN1 was inserted between the ribosome binding site (RBS) and promoter of the chloramphenicol resistance gene (CmR) with 5 nucleotides of randomized upstream and downstream sequences. Two plasmids pACYC-X5 (upstream) and pACYC-X3 (downstream) were designed as plasmid libraries to test different experimental conditions (figure 19) .These libraries were verified by Illumina HiSeq3000 sequencing and the original distribution of randomized nucleotides was then used for normalization of the depletion assays.

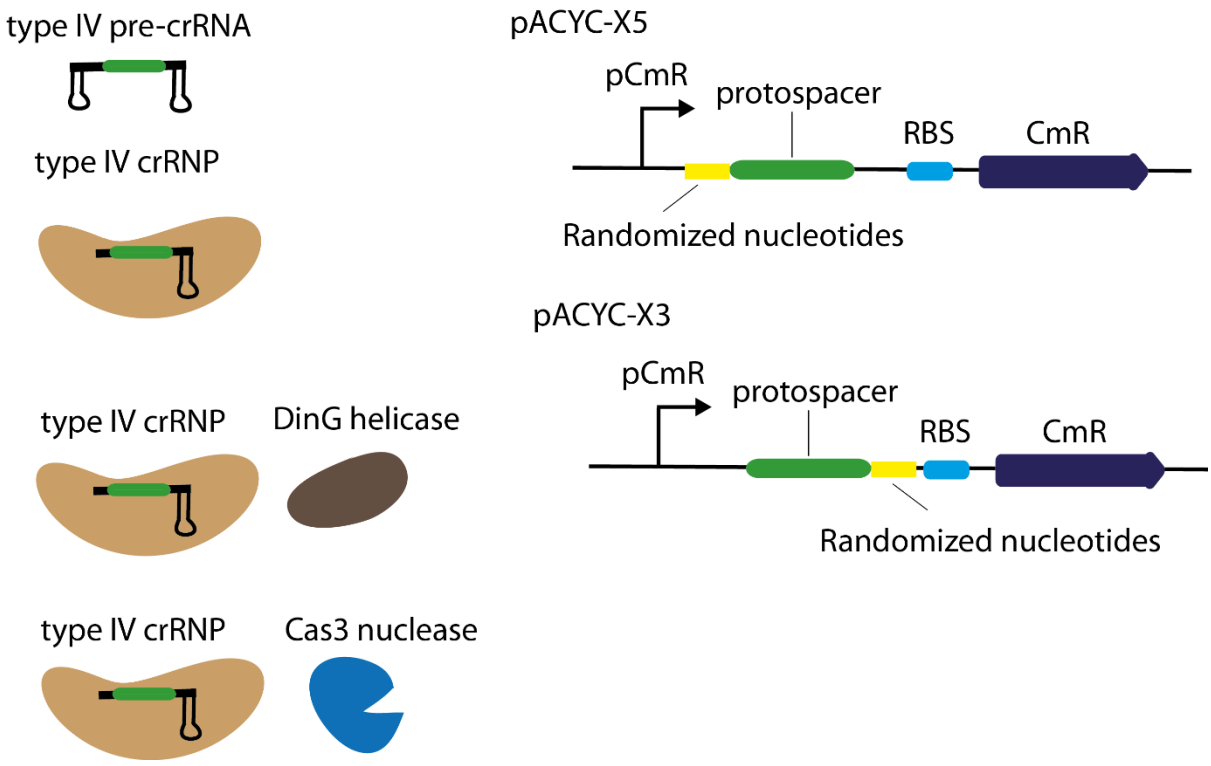


Figure 19: PAM depletion assay experimental design. A 5 nucleotide randomized library with a targeting protospacer was cloned between the promoter and the ribosome binding site of the chloramphenicol resistance gene on the pACYC duet vector. Randomized libraries were incubated with type IV pre-crRNA, type IV crRNP complexes and type IV crRNP complexes in the presence of either DinG or Cas3 nuclease of the type I-C system of *A. aromaticum* EbN1. Depleted motifs were detected by Illumina HiSeq3000 sequencing.

Verified libraries were then co-transformed into *E.coli* BL21-AI with the following experimental conditions: type IV pre-crRNA, type IV crRNP, type IV crRNP with DinG helper protein, type IV crRNP with type I-C Cas3 from the genome. The transformants were harvested after 6 hours of incubation. Strong PAM depletion was only observed for the individual type IV crRNP complex in presence of associated DinG helicase or Cas3c of the type I-C CRISPR-Cas locus. Significant reduction of certain motifs (GvR: V: not thymine; R: purine) was observed downstream of protospacers which correspond to opposite strand of the 5´ handle of crRNA (GUUGAAG). A strong selection on the G at the first position was observed.

2.11 DinG (Csf4) helicase is not a stable component of the type IV crRNP complex

PAM depletion assays suggest that the presence of DinG helicase affects type IV CRISPR-Cas activity, indicating a potential functional role of the DinG helicase. To investigate its potential role with type IV crRNP, his-tagged DinG was heterologously co-expressed in *E. coli* BL21-AI with type IV crRNP components (Csf1, Csf2 his-tagged, Csf3 & Csf5 his-tagged and CRISPR-miniarray), and purified via affinity chromatography and SEC. Without DinG, a central peak fraction was identified to contain stable type IV crRNP. It was observed that the location of the central peak did not move to higher molecular weight fractions when DinG helicase was co-expressed (figure 20). For this reason, DinG helicase is suggested to be an unstable component of the type IV crRNP complex.

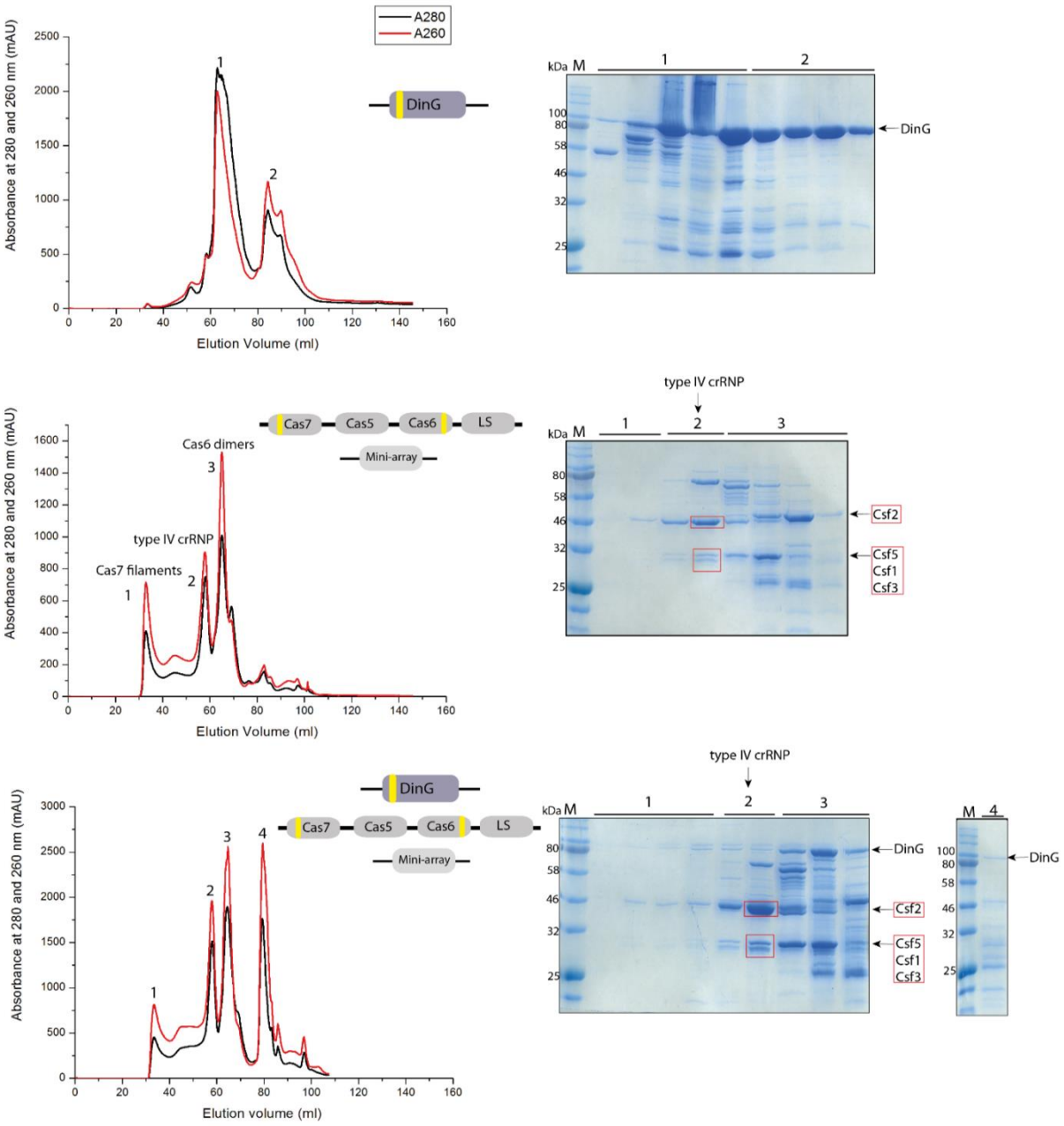


Figure 20: *A. aromaticum* Type IV Cas protein production with DinG helicase in *E. coli*. Cas proteins were purified and separated via gel-filtration chromatography. The peak fractions of the gel-filtration chromatograms were analyzed via SDS-PAGE to detect purified Cas proteins and DinG helicase. **(a)** N-terminally His-tagged (yellow) DinG helicase purification. **(b)** Type IV crRNP complex purification. Type IV crRNP was observed in fraction of the middle peak (peak 2). **(c)** Type IV crRNP complex with DinG helicase purification. The middle peak (peak 2) did not shift to higher molecular weight fractions.

2.12 DinG (Csf4) helicase interactions with type IV crRNP components

To test possible transient interactions of DinG with type IV CRISPR-Cas components, untagged DinG was co-expressed with his-tagged type IV crRNP complex components followed by gradient affinity chromatography (Ni-NTA). Untagged DinG was mainly observed in the fractions containing type IV crRNP components (figure 21).

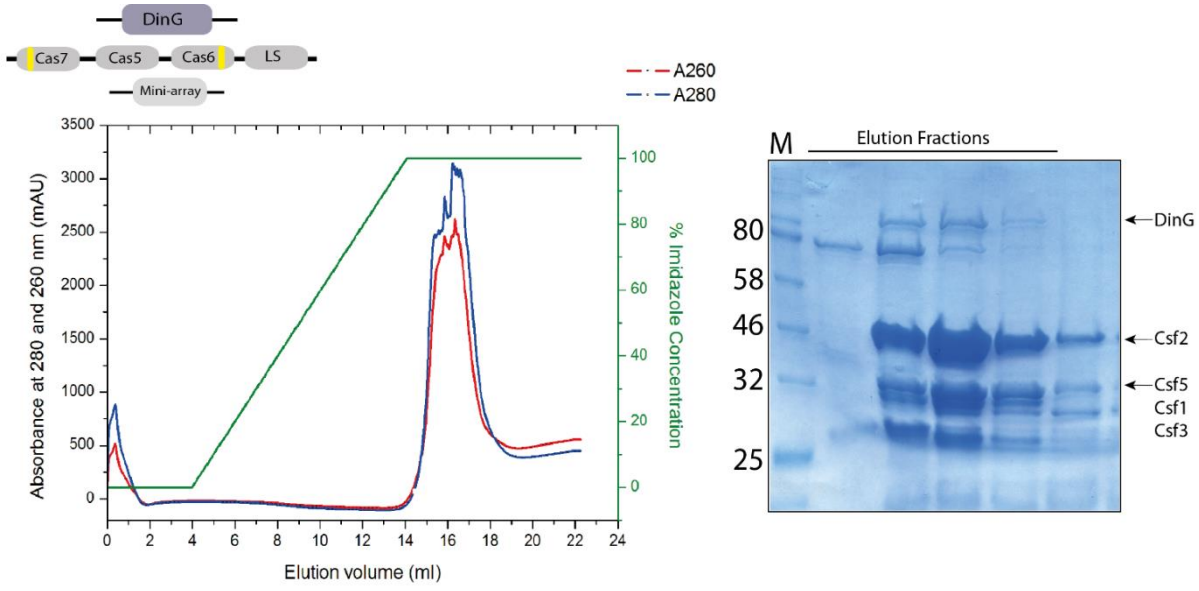


Figure 21: DinG (Csf4) helicase pulled down with type IV crRNP components during Ni-NTA affinity purification. Untagged DinG helicase was co-expressed with type IV crRNP components and purified with gradient Ni-NTA purification. The fractions were analyzed via SDS-PAGE. Hereby, DinG helicase observed to co-migrate with type IV crRNP components. The results are representative of three independent purifications.

It is important to note that Ni-NTA affinity chromatography not only elutes the pure type IV crRNP complex, but also Csf2-filaments and Csf5 dimers. Therefore, it is possible that the helicase DinG interacts with sub-complexes of the type IV crRNP. Elution fractions of the affinity chromatography were concentrated and subjected to chemical cross-linking with disuccinimidyl dibutyric urea (DSBU), followed by HPLC-MS. Fragmentation patterns of the mass spectrometry demonstrated that the DinG helicase exclusively interacts with Csf5 in the presence of all type IV crRNP components (figure 22). In comparison to previous

crosslinking-MS experiments (figure 5) with BS3, interacting crRNP proteins demonstrated a similar cross-linking architecture, including a lack of interactions with Csf1. Interacting amino acid residues were mapped onto the crystal structure of Csf5 and the computationally constituted structure of the DinG helicase (figure 23). Mapped peptide fragments suggest that interaction points of Csf5 are closer to both 3´ and 5´ endpoints of the crRNA, whereas identified peptides on DinG are distributed all over the protein and more peptides observed at the DNA funneling hole (Table 2).

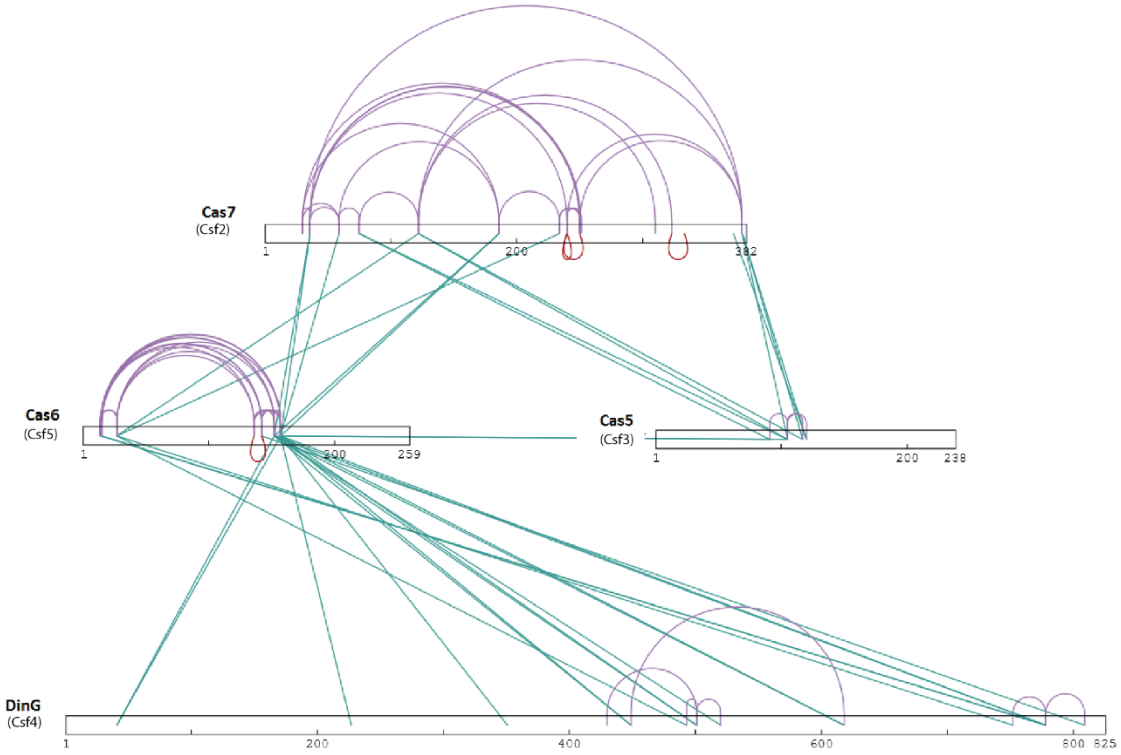


Figure 22: Protein-protein crosslinking of type IV crRNP complex and DinG helicase. Affinity purified Type IV crRNP components together with DinG were crosslinked by a DSBU chemical cross-linker followed by HPLC-MS. Observed intra-protein (purple), inter-protein (blue) and homo-multimeric (loops) crosslinks are indicated (Cross-linking and mass spectrometry analysis were performed by Lyle Kroell).

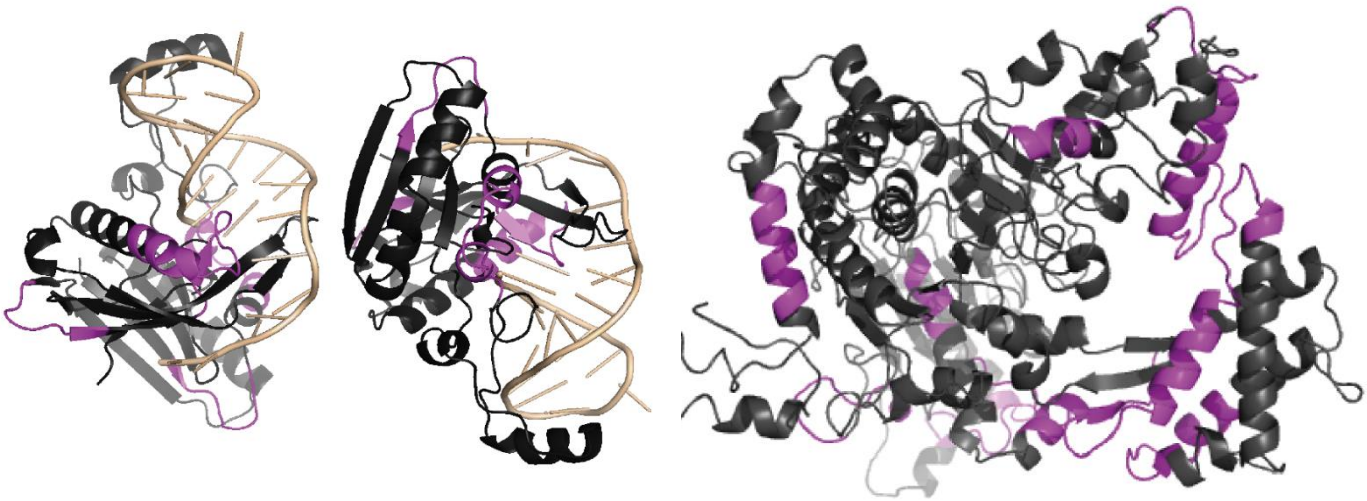


Figure 23: Csf5 – DinG helicase interactions by protein-protein crosslinking. Amino acids that were identified interact are indicated in violet on the crystal structure of Csf5 (left) and the structure model of DinG (Csf4) (DinG structure computationally built with Phyre 2 [51].

Peptide 1	Protein 1	Peptide2	Protein 2
[AEENGKEK]	DinG (Csf4)	[FALPAGKK]	Cas6 (Csf5)
[WIELDGEPLPDSPFQKR]	DinG (Csf4)	[GLDDPKR]	Cas6 (Csf5)
[ELLKNPK]	DinG (Csf4)	[GLTKTAENTTQVASR]	Cas6 (Csf5)
[LKQGMGR]	DinG (Csf4)	[GLTKTAENTTQVASR]	Cas6 (Csf5)
[AISLFAKGETIALSER]	DinG (Csf4)	[GLTKTAENTTQVASR]	Cas6 (Csf5)
[SAHAKLR]	DinG (Csf4)	[GLTKTAENTTQVASR]	Cas6 (Csf5)
[NPKYFDAMQAVER]	DinG (Csf4)	[GLTKTAENTTQVASR]	Cas6 (Csf5)
[QHALVKFSPVR]	DinG (Csf4)	[GLTKTAENTTQVASR]	Cas6 (Csf5)
[STVAKALEEK]	DinG (Csf4)	[NLVMKR]	Cas6 (Csf5)
[GDKLTSAALQEAER]	DinG (Csf4)	[NLVMKR]	Cas6 (Csf5)
[LPFGLNKSVTHEYR]	DinG (Csf4)	[NLVMKR]	Cas6 (Csf5)
[IHRPHFKMIDER]	DinG (Csf4)	[NLVMKR]	Cas6 (Csf5)
[EVEYAKIILR]	DinG (Csf4)	[NLVMKR]	Cas6 (Csf5)
[EVEYAKIILR]	DinG (Csf4)	[EALAKHDLPPLFFSR]	Cas6 (Csf5)
[AISLFAKGETIALSER]	DinG (Csf4)	[EALAKHDLPPLFFSR]	Cas6 (Csf5)

Table 2: Csf5-DinG interacting protein residues. Recombinant type IV crRNP complex components and DinG helicase were subjected to DSBU chemical cross-linker, followed by mass spectrometry. Specific amino acid residues of the DinG helicase with Csf5 were found to interact with the indicated peptides.

2.13 Type IV CRISPR-Cas associated DinG helicase (Csf4) retains ssDNA *in vivo*

Recombinant type IV associated DinG (Csf4) helicase was soluble and it was possible to purify milligram amounts of DinG. Therefore, different tagged versions of the DinG helicase were created to obtain pure DinG for biochemical and structural analyses. The employed two-step purification strategy included affinity (Ni-NTA) chromatography followed by SEC purification. It was observed that C-terminally his-tagged DinG retained high amounts of nucleic acids after purification. Eluted fractions were analyzed by SDS and Urea-PAGE. DinG was co-purified with transcription and translation coupled proteins and nucleic acids (figure 24). To obtain insights about the composition of the nucleic acids that were bound to DinG, they were purified by phenol-chloroform extraction and ethanol precipitation and subjected to nuclease assays using DNase-free RNase and RNase-free DNase.

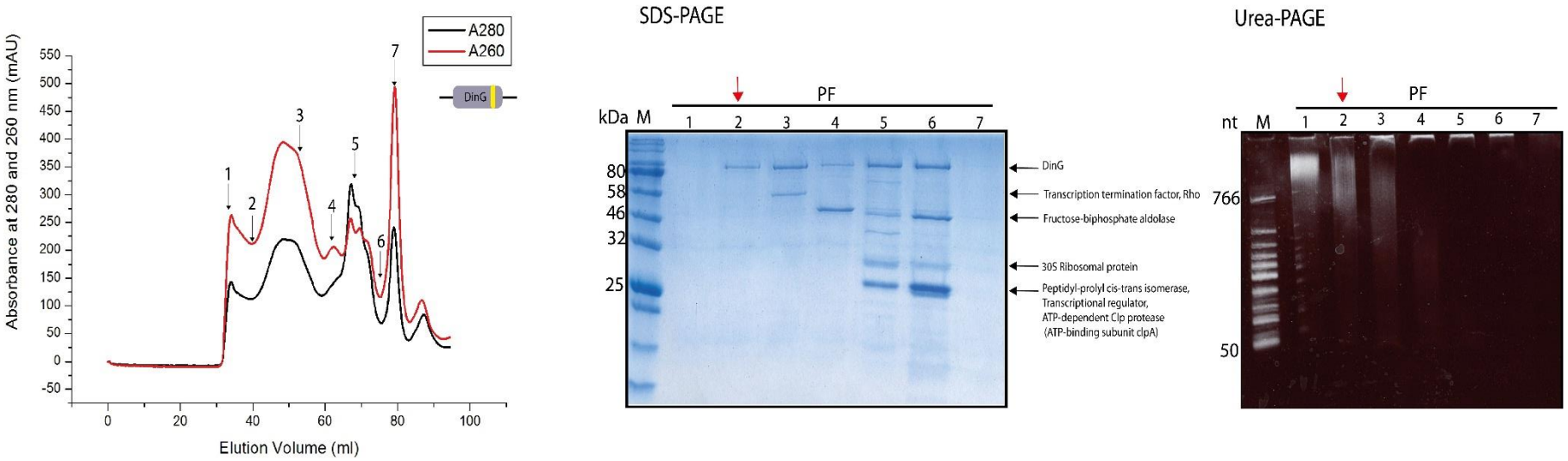


Figure 24: DinG helicase purification retains nucleic acids. C-terminally his-tagged DinG purified by Ni-NTA followed by SEC. Purification fractions [1] analyzed by SDS and Urea-PAGE. Retained proteins were detected by SDS-PAGE and HPLC-MS. Bound nucleic acids of various size were detected (Urea-PAGE). The results were from two replicates. (SDS and Urea-PAGE was generated together with Eva Grümpel)

Degradation was only observed with DNase1, confirming that DNA was co-purified. Further analysis included S1 nuclease digestion to specifically target ssDNA. S1 nuclease was observed to degrade most of the eluted DNA, suggesting that the nucleic acid content was largely single stranded DNA (figure 25).

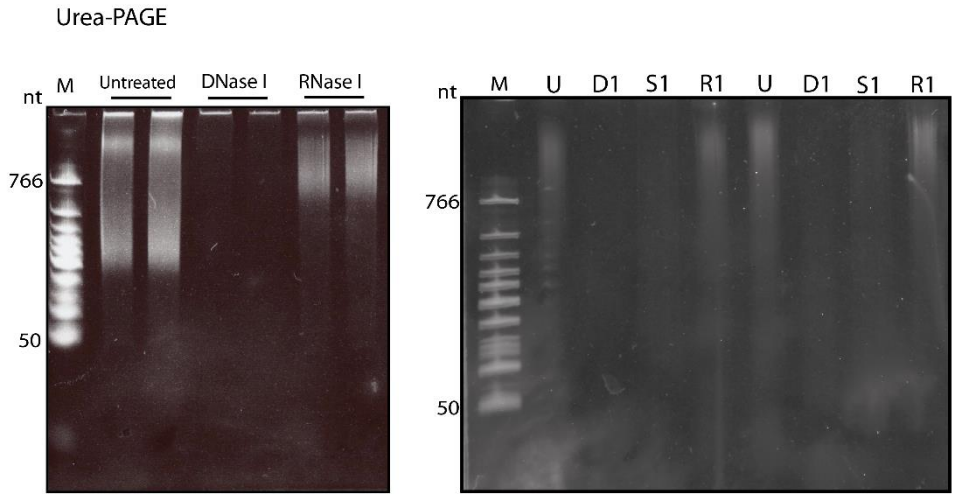


Figure 25: DinG helicase bound nucleic acid detection assays. a. Purified DinG bound nucleic acids were treated with DNase1 and RNase1 overnight at 37°C. Degradation was detected only with DNase1. **b.** Purified DinG bound nucleic acids were treated with DNase1 (D1), S1 nuclease (S1) and RNase1 (R1). DNase1 and S1 nuclease degrade completely. (U) Untreated sample. (The second Urea-PAGE was generated by Eva Grümpel)

3.Discussion

Among the six computationally classified CRISPR-Cas types, members of the type IV CRISPR-Cas system remained without experimental support [80].In this thesis, we provide the first experimental characterization of a model type IV CRISPR-Cas system that is located on a megaplasmid of *Aromatoleum aromaticum* EbN1. The system is characterized by conserved *cas* genes, encoding effector module forming proteins, a uniform CRISPR array and a *dinG* gene upstream of the *cas* gene operon.

3.1 Type IV crRNA biogenesis and maturation

The maturation of crRNAs is critical for the activity of CRISPR-Cas systems. Different types of CRISPR-Cas systems have evolved distinct crRNA biogenesis pathways that suggest diversified processing mechanisms. In type I and type III systems, Cas6 family endoribonucleases (or alternatively Cas5d) cleave the pre-crRNA at the stem-loop structured repeat regions. In type II systems, the trans-activating CRISPR RNA (tracrRNA) base pairs with each repeat of the pre-crRNA to form a RNA duplex region that is cleaved by RNase III in the presence of the Cas9 protein [59, 114] [60]. CRISPR RNA processing was not observed for Type IV CRISPR arrays. Our observation of a uniform CRISPR array in the type IV-A system of *A. aromaticum* suggested the presence of a dedicated crRNA processing mechanism for this system. Therefore, we analyzed its crRNA maturation process. RNA-Seq analysis of the small RNome of *A. aromaticum* revealed that crRNA processing occurs in its repeat regions, creating a unique 7 nt-5´-terminal tag (5´- GUUGAAG-3´) and a stable stem-loop structure at the 3´-terminus of the crRNA. In this regard, crRNA maturation of the type IV-A system of *A. aromaticum* is similar to most type I CRISPR-Cas system`s crRNA maturation process. One major difference is that the common 8 nt tag was shortened by one nucleotide. Cleavage occurs at the base of a hairpin structure within the repeat, which is also commonly

observed for type I pre-crRNA processing. In addition, crRNA processing was also investigated for the second, subtype I-C CRISPR-Cas system of *A. aromaticum*. 11 nt-5´-terminal repeat tags were observed, which is in agreement with other well characterized type I-C systems [114].

The observation of type IV crRNA processing activity stimulated our search for the responsible RNA nuclease. We identified that Csf5 acts as a Cas6-family endoribonuclease. A crystal structure of this enzyme was obtained which provided further evidence for its cleavage activity at the base of the stem-loop of the crRNA repeat. Additionally, crystal structures of RNA-bound Csf5 suggest a single turnover cleavage mechanism, during which Csf5 stays bound to the crRNA and assembles into a complex with other Cas proteins for downstream targeting. This mode of cleavage is also common in type I-B, I-C, I-E and I-F systems [115] [60, 114].

Despite considerable sequence variations, all Cas6 enzymes show similar structural features that are important for pre-crRNA binding. Structural features of Csf5 were shown to coincide with Cas6 family endoribonucleases with very limited sequence homology. Moreover, most of the Cas6 enzymes utilize catalytic histidine residues for crRNA cleavage. However, alanine scanning assays of Csf5 revealed that the mutation of the closest histidine at the active site does not abolish cleavage activity, supporting the absence of a canonical catalytic histidine residue. Instead, the active site of Csf5 is rather composed of three arginine residues that are important for cleavage to occur [114].

In vivo cleavage assays of Csf5 with a subtype I-C repeat did not yield mature crRNA, showing that Csf5 is specific for the repeats of the type IV CRISPR-Cas system. A crosstalk between type I-C and type IV crRNA maturation processes does not exist. This observation further supports the notion of co-evolution between diversified Cas6 enzymes and cognate repeat sequences [116].

Csf5 was observed to occur as a homodimer with its bound crRNA. During size exclusion chromatography, Csf5 eluted in a fraction that corresponds to a molecular weight of 80 kDa, suggesting a dimerization of proteins with a bound RNA component. The obtained crystal structures also indicated dimerization. Similar dimerization patterns were observed in other Cas6 enzymes of mostly thermophilic organisms. The Cas6 enzyme of *Sulfolobus solfataricus* was also shown to form dimers in order to structure crRNA before cleavage. However, in the case of Csf5, it remains unclear whether dimerization has a functional role during the structuring of RNA [117] [1, 118-121].

In summary, we elucidated the crRNA maturation of the type IV CRISPR-Cas system and identified Csf5 as a Cas6 family protein. However, several questions still remain. It is not clear how Csf5 achieves the delivery of mature crRNAs to the type IV interference machinery while protecting the crRNA during assembly. In addition, it is of interest to determine an explanation as to why Csf5 proteins dimerize in the presence of a structured crRNA substrate and how dimerization might affect crRNA processing.

3.2 Formation of the type IV crRNP complex and its features

In type I CRISPR-Cas systems, individually processed crRNAs are incorporated into multi-protein complexes to carry out anti-viral defense [108] [45]. In diverse type I CRISPR-Cas families, the composition and the architecture of the crRNA bound multiprotein effector complexes vary, whereas their biological function is conserved. Multiple copies of Cas7 proteins assemble around the crRNA during or after Cas6 family mediated crRNA maturation [108]. Structural and mass spectrometry-based studies have shown that type I and type III CRISPR-Cas interference complexes employ conserved building blocks. The presence of a protein backbone remains central to all multi-subunit crRNP protein assemblies [79, 110, 122]. Structural studies of type I-E Cascade demonstrated the formation of Cas7 multimers, which non-specifically wrap around RNA [78]. Consistent with this observation, the protein

Csf2 of the type IV system was observed to form helical filaments around RNA molecules. Mass spectrometry based crosslinking analyses further strengthen its backbone function due to the presence of multiple copies of Csf2 in type IV crRNP assemblies.

Previous studies on type I Cascade revealed that Cas proteins bound to a crRNA formed a crescent-shaped structure [79, 110, 122]. In this study, we demonstrated that the type IV crRNP complex is similarly composed of a mature crRNA and four different types of Cas proteins (Csf1, Csf2, Csf3, Csf5). Mass-spectrometric and biochemical analyses provided first insights into the stoichiometry and the arrangement of the proteins in a type IV effector complex. To obtain detailed insights, we attempted to crystallize the type IV crRNP complex. However, this attempt failed due to a low yield of purified proteins. Cryogenic electron microscopy (cryo-EM) analysis may provide a viable alternative method for further structural analysis of the type IV crRNP complex.

3.3 Investigation of PAM requirements for the type IV CRISPR-Cas system of *A. aromaticum*

The activity of most CRISPR-Cas systems requires recognition of short sequence motifs (PAM, PFS). DNA targeting CRISPR-Cas systems rely on the recognition of a PAM sequence prior to crRNA mediated target binding. In this study, we provide a first description of the involvement of PAM sequences for type IV CRISPR-Cas systems. Previous computational analyses identified protospacer targets for type IV CRISPR-arrays from nine organisms [112].Using these identified protospacers, we analyzed upstream and downstream sequences and demonstrated the presence and conservation of species-specific PAMs for type IV CRISPR-Cas systems (Ziemann et al., unpublished).

In addition, *in-vivo* PAM depletion assays were conducted to gain a deeper understanding of the type IV PAM sequence of our model organism *A. aromaticum*. Thereby, randomized PAM depletion libraries were created and tested with type IV CRISPR-Cas components in heterologous *E.coli* BL21-AI system. Significant reduction of certain motifs (GVR: V: not thymine; R: purine) was obtained downstream of the protospacers, which corresponds to the opposite strand of the 5′-handle of the crRNA. The lack of complementarity for PAM regions and 5′-crRNA repeat handles for this type IV crRNP is in agreements with other type I CRISPR-Cas systems that rely on perfect pairing of the handle to avoid self-targeting [123].

To our surprise, the most significant effects were obtained when DinG or Cas3c was present. The depletion of specific PAM sequences suggests that the type IV crRNP complex, together with either DinG or Cas3c, is able to reduce the transformation efficiency of targeted plasmids. It remains to be investigated how increased helicase activity, provided either by DinG or Cas3c aids in type IV crRNP functionality.

3.4 Putative functional roles of type IV CRISPR-Cas systems

Our results show that the type IV CRISPR-Cas system of *A. aromaticum* generates mature crRNAs and consequently forms a crRNP complex which has PAM sequence conservation. Based on our current data and previous computational studies on nuclease and adaptation unit free CRISPR-Cas systems, we hypothesize several possibilities related to the functional role of the type IV CRISPR-Cas system.

First, type IV CRISPR-Cas systems might function as adaptive immune function like other CRISPR-Cas systems. This possibility is supported by a previous study which showed the highest protospacer hits on viral genomes [112]. In addition, PAM depletion assays provided further proof that the type IV crRNP complex, together with either DinG or Cas3c, could provide a defense against mobile genetic elements. In the case of an adaptive immune function, the type IV CRISPR-Cas system of *A. aromaticum* highlights a unique mechanism of interference due to the absence of any assigned nucleases. Therefore, the involvement of additional host proteins is possible.

A second hypothesis is that some crRNPs are capable of transposon guidance. Minimal CRISPR-Cas systems without adaptation unit and DNA nucleases were suggested to guide Tn7 transposons[114] [102]. In this regard, type IV CRISPR-Cas systems might function in a similar way, which is supported by a high number of mobile genetic elements in the *A. aromaticum* EbN1 genome.

Third, a crRNA-guided protein complex can be used for the specific control of gene expression. As an example, the mutation of Cas9 to generate a nuclease-deficient deadCas9 (dCas9) variant was used to control of gene expression (CRISPRi). Our system might constitute a natural "dead" Cascade system that can be employed to control of gene expression [124].

In future experiments, we need to do phage plaque assays, transposition assays and GFP protein based gene repression assays to investigate if any of these hypothesis is true. It is also noteworthy that type IV CRISPR-Cas systems localize mainly on plasmids thereby, it is possible to have a function in maintenance of the plasmids (Newire et al.,2019, Bioarchive) .

3.5 Putative functional roles of type IV CRISPR-Cas associated DinG helicases

Finally, the function of the DinG helicase, encoded by a conserved gene upstream of the cas gene operon, remained unclear and its association with the type IV crRNP complex was analyzed. Initially, we investigated if DinG is a stable component of the crRNP complex. *In-vitro* purification results demonstrated that the DinG helicase is neither a component of the type IV crRNP complex nor does it interfere with the complex stability. Cross-linking coupled mass spectrometry analysis demonstrated that DinG exclusively interacts with Csf5. This observation suggests that transient interactions of DinG with the type IV crRNP are possible. In addition, expression of the DinG helicase with the type IV crRNP complex resulted in a strong depletion of functional PAM sequences. Therefore, we hypothesize multiple possible functions of DinG together with type IV crRNP complex.

The first hypothesis of DinG is to provide effective interference activity. The type IV crRNP complex accomplishes PAM specific DNA interactions and creates an R-loop based on sequence complementarity at the target region. R-loop formation could consequently recruit DinG and resulting in defense against MGEs. This hypothesis is supported by our experimental PAM depletion assays. On the contrary to our proposed hypothesis, type I-Fv CRISPR-Cas associated DinG helicase from *Shewanella putrefaciens* strain CN32 demonstrated that DinG does not have an effect on interference activity [125]. However, the I-Fv CRISPR-Cas system possesses an assigned target cleavage nuclease (Cas3) which results in efficient interference activity. Considering that type IV CRISPR-Cas systems are deficient of any assigned target cleavage protein, it is possible that the type IV CRISPR-Cas specific

DinG has evolved to fulfil a similar function which is also exemplified in DinG of S. aureus [90].

DinG might also play a helper role for transposon guidance. Considering that DinG is active on R-loops, a similar function to TnsE can be attributed to DinG. The type IV crRNP complex is responsible for R-loop formation and might recruit DinG to the target region. This process can result in distortions on double stranded DNA which, in turn, could create hot spots for the dissemination of transposable elements.

In addition, type IV CRISPR-Cas systems are categorized into two groups based on the presence of *dinG*. In type IV-A CRISPR-Cas systems, a regular CRISPR array, *dinG* and *csf5* occurrence are noteworthy as compared to type IV-B gene cluster. Hence, it is possible that DinG might indirectly play a role in adaptation of the type IV-A CRISPR-Cas systems.

Finally, DinG was proposed to dissolve R-loops by removal of the RNA transcripts to restore collapsed replication forks. Therefore, a similar role could be suggested in the context of R-loops created by crRNP complexes. Thus, DinG might remove crRNA-DNA interactions to recycle crRNP complexes. Regarding a CRISPRi function, such scenario would suggests a tunable gene regulation effect. A detailed understanding of DinG functionality will require (i) understanding of the functional role of type IV crRNP (ii) knowledge of DinG substrate specificity and (iii) structural insights.

4. Material and Methods

4.1 Materials, instruments and source of supplies

4.1.1 Chemicals, kits and enzymes

The chemicals, kits and enzymes used in this work were obtained from companies listed in Table 4.1.

Table 4.1: List of chemicals, enzymes, kits and consumables used in this study.

Chemicals	**Company**
Acrylamid/Bisacrylamid (29:1 und 37.5:1)	Carl Roth GmbH, Karlsruhe
Amicon ® Ultra Centrifugal Filters	Merck Millipore KGaA, Darmstadt
Antarctic Phosphatase	New England Biolabs GmbH, Frankfurt
Antibiotics (Ampicillin, Chloramphenicol, Kanamycin, Spectinomycin)	Carl Roth GmbH, Karlsruhe; Sigma-Aldrich, Taufkirchen
Bovine Serum Albumin (BSA)	Sigma-Aldrich, Taufkirchen
Bradford Reagent	BioRad Laboratories GmbH, Munich
ColorPlus Prestained Protein Ladder, Broad Range (10-230 kDa)	New England Biolabs GmbH, Frankfurt
Diethylpyrocarbonate (DEPC)	AppliChem GmbH, Darmstadt
DNaseI	Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
dNTPs	New England Biolabs GmbH, Frankfurt
Gelrite	Carl Roth GmbH, Karlsruhe

Gene Pulser ® Cuvette, 0.1 cm gap	BioRad Laboratories GmbH, Munich
Glycogen	Roche Diagnostics GmbH, Mannheim
Instant Blue	Sigma-Aldrich, Taufkirchen (Expedeon)
Isopropyl-β-D-thiogalactopyranosid (IPTG)	Carl Roth GmbH, Karlsruhe
Low Molecular Weight Ladder	New England Biolabs GmbH, Frankfurt
Low Range ssRNA Ladder	New England Biolabs GmbH, Frankfurt
Lysozym	Sigma-Aldrich, Taufkirchen
mirVana™ miRNA Isolation Kit	Applied Biosystems, Darmstadt
NTPs	Jena Bioscience GmbH, Jena
Phenol (Roti-Phenol & Roti-Aqua-Phenol)	Carl Roth GmbH, Karlsruhe
Phusion ® High-Fidelity DNA Polymerase	Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
QIAGEN Plasmid Plus Maxi Kit	Qiagen GmbH, Hilden
QIAprep Spin Miniprep Kit	Qiagen GmbH, Hilden
QIAquick Gel Extraction Kit	Qiagen GmbH, Hilden
Quick-Load ® 2-Log DNA Ladder (0.1-10.0 kb)	New England Biolabs GmbH, Frankfurt
Restriction endonucleases	New England Biolabs GmbH, Frankfurt
RNase Exitus Plus™	Applichem GmbH, Darmstadt
RNase Inhibitor	New England Biolabs GmbH, Frankfurt

RNaseR	Epicentre, Madison, USA
Roti ®-Nylon plus, pore size 0.45 µm	Carl Roth GmbH, Karlsruhe
SDS	Carl Roth GmbH, Karlsruhe
SYBR Gold ® Nucleic acid stain	Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
T4-DNA-Ligase	New England Biolabs GmbH, Frankfurt
T4-Polynucleotide-Kinase	New England Biolabs GmbH, Frankfurt
T7-RNA-Polymerase	New England Biolabs GmbH, Frankfurt
Taq DNA Polymerase	New England Biolabs GmbH, Frankfurt
Topo ® TA cloning ®	Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
TRIzol Reagent	Ambion, Darmstadt
ULTRAhyb-Oligo hybridization buffer	Ambion, Darmstadt
Whatman GB004, 3MM	Schleicher & Schuell, Dassel

4.1.2. Instruments

Table 4.2: Instruments used in this study.

Agarose gel electrophoresis	Chambers and Casting tray: company technician Philipps-University Marburg; Power supply: Consort E835; MS Laborgeräte, Dielheim

Aqua bidest. water system	PURELAB Plus, ELGA LabWater, Celle
Autoclave	5075 EL, Tuttnauer Europe B.V., Breda
Semi-dry transfer cell	Trans-Blot® SD Semi-Dry Transfer Cell, BioRad Laboratories GmbH, Munich
Centrifuges	Centrifuge 5424, Eppendorf AG, Hamburg; Sorvall RC5B Plus, Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
Chromatography columns	HisTrap HP 1 ml/ MonoQ 5/50 GL/ Superdex 200 10/300 GL/HiTrap Heparin HP 1 ml/ MonoS 5/50 GL; GE Healthcare Europe GmbH, Freiburg
FPLC	Äkta-System: Pump P-900, Monitor UV-900, Monitor UPC-900, Valve INV-907, Mixer M-925; GE Healthcare Europe GmbH, Freiburg
Denaturing polyacrylamide gel chambers	PROTEAN II Electrophoresis Chamber, BioRad Laboratories GmbH, Munich
Gene Pulser	Gene Pulser ® Electroporation System, Pulse Controller Plus, Capacitance Extender Plus; BioRad Laboratories GmbH, Munich
Hybridization oven	Hybaid Shake 'n' Stack, Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
Incubators	Thermotron, Infors AG, Bottmingen,

	Switzerland
Magnetic stirrer	IKA® RCT Standard, IKA®-Werke GmbH & Co. KG, Staufen
Nanodrop	NanoDrop® ND-1000 Spectrometer, Thermo Fisher Scientific Germany Ltd. & Co. KG, Bonn
PCR-Cycler	C1000™ Thermal Cycler, BioRad Laboratories GmbH, Munich
Phosphorimager	Storm 840 phosphorimager, Molecular Dynamics
Polyacrylamide gel electrophoresis	Mini-PROTEAN Tetra Cell,Bio-Rad Laboratories GmbH, Munich; Power supply PowerPac Basic, Bio-Rad Laboratories GmbH, Munich
Scintillation counter	Beckmann LS 6500, Beckman Coulter, Krefeld
Spectrophotometer	Ultrospec 3000 *pro*, GE Healthcare Europe GmbH, Freiburg
Thermomixer	Thermomixer Comfort 5350, Eppendorf AG, Hamburg
UV-Crosslinker	UV Stratalinker ® 1800, Stratagene, La Jolla, USA

UV-Transilluminator	BioDocd-IT system, UVP, Upland, USA

4.2 Strains and culture conditions

4.2.1. Strains

Table 4.3: Bacterial and archaeal strains used in this study.

Strain	Relevant genotype	Reference
***Escherichia coli* K12 DH5α**	F⁻Φ80*lac*ZΔM15Δ(*lac*ZYA-*arg*F)U169*rec*A1*end*A1*hsd*R17(r_K^-, m_K^+)*pho*A *sup*E44λ–*thi*-1*gyr*A96*rel*A1	**[126]**
***Escherichia coli* BL21-AI**	F⁻ *omp*T *hsd*S_B (r_B^-, m_B^-) *gal* *dcm ara*B::T7RNAP-*tet*A	**Thermo Fisher Scientific Germany Ltd. &Co. KG, Bonn**
Aromatoleum aromaticum EbN1		**AGHeider-Philipps University, Marburg**

4.2.2. *Escherichia coli* growth conditions

E. coli cultures were grown in LB medium (1 % tryptone (w/v), 0.5 % yeast extract, 1 % NaCl (w/v), pH 7.2) in a rotatory shaker at 200 rpm at 37°C or on solid medium plates (LB medium containing 1.5 % (w/v) agar-agar). Single colonies were inoculated with a pre-culture (2% (v/v)) which contain LB medium with appropriate antibiotics (spectinomycin, kanamycin 50 μg/ml and ampicillin 100 μg/ml) based on plasmid encoded antibiotic resistance gene. Expression of recombinant proteins were induced at growth phase (OD_{600}:0.6) with 1 mM IPTG and 20% L-arabinose and grown for 3-4h. Cells were subsequently harvested by centrifugation (6000g, 4°C). The *E. coli* strain K12 DH5α was used for cloning, storage and preparation of plasmid DNA.

4.2.3 *Aromatoleum aromaticum* EbN1 growth conditions

Aromatoleum aromaticum strain EbN1 was grown anaerobically in a 30 liters fed-batch fermenter culture on minimal medium with benzoate (3 mM) as sole carbon source and nitrate (10 mM) as electron acceptor. Cells were harvested in exponential growth phase at optical densities (OD: 578 nm) of 3.5-4 [127].

4.3 Oligonucleotides, plasmids and constructed recombinant vectors

4.3.1 Plasmids and constructed recombinant vectors

Table 4.4: Plasmids used in this study.

Vector	Features	Application	Source
pUC19	*lacZ*, Amp^r	Transcription of pre-crRNA	NEB
pET-Duet1	Amp^r	Heterologous gene expression	Novagen
pRSF-Duet1	Kan^r	Heterologous gene expression	Novagen
pCDF-Duet1	$Spec^r$	Heterologous gene expression	Novagen

Table 4.5: Constructed recombinant plasmids for recombinant protein production

Plasmid + Insert	Description of the insert
pCDF + Csf1	*E.coli* codon optimized large subunit (Csf1) protein from *A.aromaticum EbN1*.Csf1 cloned in MCS2 (Nde1 cut site).

pCDF+Csf5-Csf1	*E.coli* codon optimized large subunit and Csf5 (C-terminal histag).Csf5 cloned in MCS1 (Nco1 cut site), Csf1 cloned in MCS2 (Nde1 cut site).
pRSF+Csf3	*E.coli* codon optimized Csf3 protein from *A.aromaticum EbN1*. Csf3 cloned in MCS2 (Nde1 cut site).
pRSF+Csf2-Csf3	*E.coli* codon optimized Csf2 (N-terminal his-tag) and Csf3 protein from *A.aromaticum EbN1*. Csf2 cloned in MCS1 (Nco1 cut site), Csf3 cloned in MCS2 (Nde1 cut site).
pRSF +Csf3	*E.coli* codon optimized Csf3 with (C-terminal his-tag) protein from *A.aromaticum EbN1*. Csf3 cloned in MCS2 (Nde1 cut site).
pRSF +Csf2	*E.coli* codon optimized Csf2 protein from *A.aromaticum EbN1*. Csf2 cloned in MCS1 (Nco1 cut site).
pCDF +Csf5	*E.coli* codon optimized Csf5 (C-terminal his-tag) protein from *A.aromaticum EbN1*. His-tagged Csf5 cloned in MCS1 (Nco1 cut site).
pRSF +Csf2	*E.coli* codon optimized Csf2 protein from *A.aromaticum EbN1*. Csf2 cloned in MCS1 (Nco1 cut site). His-tag was removed.
	E.coli codon optimized Csf2 and Csf3 protein from *A.aromaticum*

pRSF+Csf2+Csf3	*EbN1*. Csf2 cloned in MCS1 (Nco1 cut site), Csf3 cloned in MCS2 (Nde1 cut site). His-tags were removed.
pCDF +Csf1	*E.coli* codon optimized large subunit protein (Csf1 with C-terminal his-tag) from *A.aromaticum EbN1*.Csf1 cloned in MCS2 (Nde1 cut site).
pCDF+Csf5-Csf1	*E.coli* codon optimized large subunit (Csf1 with C-terminal his-tag) and Csf5 protein from *A.aromaticum EbN1*. Csf5 cloned in MCS1 (Nco1 cut site), large subunit cloned in MCS2 (Nde1 cut site).
pCDF+Csf5-Csf1	*E.coli* codon optimized large subunit Csf1 and Csf5 protein from *A.aromaticum EbN1*. Csf5 cloned in MCS1 (Nco1 cut site), large subunit cloned in MCS2 (Nde1 cut site). His-tags were removed.
pRSF+Csf2-Csf3	*E.coli* codon optimized Csf2 and Csf3 (C-terminal his-tag) protein from *A.aromaticum EbN1*. Csf2 cloned in MCS1 (Nco1 cut site), Csf3 cloned in MCS2 (Nde1 cut site).
pCDF+Csf5-Csf2-Csf3-Csf1	*E.coli* codon optimized Csf5 (C-terminal his-tag), Csf2 (N-terminal his-tag), Csf3 and Csf1 (large subunit) proteins from *A.aromaticum EbN1*. Csf5 cloned in MCS1 (Nco1 cut site) and rbs sequence inserted before Csf2. Csf3 cloned in MCS2 (Nde1 cut site) and rbs sequences were inserted before Csf1. This

	plasmid was created together with Xiohan Guo.
pRSF + DinG	*E.coli* codon optimized Csf4 (DinG) protein from *A.aromaticum EbN1*. Csf4 (DinG) cloned in MCS1 (Nco1 cut site). Affinity tags were removed.
pRSF+DinG -FLAG	*E.coli* codon optimized Csf4 (DinG) protein from *A.aromaticum EbN1*. N-terminal his-tag, C-terminal FLAG tag was inserted. Csf4 (DinG) cloned in MCS1 (Nco1 cut site)
pRSF+DinG	*E.coli* codon optimized Csf4 (DinG with C-terminal his-tag) protein from *A.aromaticum EbN1*. Csf4 (DinG) cloned in MCS1 (Nco1 cut site).

Table 4.6: Constructed recombinant plasmids for pre-crRNA production

Plasmid + Insert	**Description of the insert**
pUC19 + type IV miniCRISPR	type IV repeat-spacer-repeat from *A.aromaticum* was cloned with T7 promoter, restriction sites BamHI/HindIII. Based on RNA-Seq data.
pUC19 + type I-C miniCRISPR	Type I-C repeat-spacer-repeat from *A.aromaticum* was cloned with T7 promoter, restriction sites BamHI/HindIII. Based on RNA-Seq data

Table 4.7: Constructed recombinant plasmids for alanine scanning assays

Plasmid + Insert & Mutation	**Description of the insert**
pCDF + Csf5 H44A	*E.coli* codon optimized Csf5 with H44A replacement at MCS1 (Nco1 cut site).
pCDF + Csf5 S209A	*E.coli* codon optimized Csf5 with S209A replacement at MCS1 (Nco1 cut site).
pCDF + Csf5 R38A	*E.coli* codon optimized Csf5 with R38A replacement at MCS1 (Nco1 cut site).
pCDF + Csf5 R23A	*E.coli* codon optimized Csf5 with R23A replacement at MCS1 (Nco1 cut site).
pCDF + Csf5 S241A	*E.coli* codon optimized Csf5 with S241A replacement at MCS1 (Nco1 cut site).
pCDF + Csf5 R242A	*E.coli* codon optimized Csf5 with R242A replacement at MCS1 (Nco1 cut site).
pCDF + Csf5 R23A+ R38A	*E.coli* codon optimized Csf5 with R23A+ R38A double replacement at MCS1 (Nco1 cut site).
pCDF + Csf5 R23A+ R242A	*E.coli* codon optimized Csf5 with R23A+ R242A double replacement at MCS1 (Nco1 cut site).
	E.coli codon optimized Csf5 with H44A+S209A double

pCDF + Csf5 H44A+S209A	replacement at MCS1 (Nco1 cut site).
pCDF + Cas6 H44A+R38A	*E.coli* codon optimized Csf5 with H44A+ R38A double replacement at MCS1 (Nco1 cut site).
pCDF + Cas6 R38A+S209A	*E.coli* codon optimized Csf5 with R38A+S209A double replacement at MCS1 (Nco1 cut site).

4.3.2 Oligonucleotides

All oligonucleotides were synthesized from Eurofins MWG Operon. RNA oligonucleotide was synthesized by Sigma.-Aldrich.

Table 4.8: Oligonucleotides used for northern blot, nuclease assays and mini-CRISPR array formation.

Name	Sequence (5′-3′)
Type IV mini-CRISPR array F	GATCCTAATACGACTCACTATAGGGGTGTTCCCCGCGCATCGCG GGGGTTGAAGCGTCAGGTCTGCAACAAAGATCAACCCTACTCG GTGTTCCCCGCGCATCGCGGGGGGTTGAAGA
Type IV mini-CRISPR array R	AGCTTCTTCAACCCCCGCGATGCGCGGGGAACACCGAGTAGGG TTGATCTTTGTTGCAGACCTGACGCTTCAACCCCCGCGATGCGC GGGGAACACCCCTATAGTGAGTCGTATTAG
TypeI-C mini-CRISPR array F	GAT CCT AAT ACG ACT CAC TAT AGG GGC ATC GCC CCT CGG TGA CGG GGG GCG TGG ATT GAA ACC GTC AGG TCT GCA ACA AAG ATC AAC
TypeI-C miniCRISPR array R	AAT CCA CGC CCC CCG TCA CCG AGG GGC GAT GCC CCT ATA GTG AGT CGT ATT AG
TypeI-C mini-CRISPR array F	CCT ACT CGG CAT CGC CCC TCG GTG ACG GGG GGC GTG GAT TGA AAC A
TypeI-C	AGC TTG TTT CAA TCC ACG CCC CCC GTC ACC GAG GGG CGA

mini-CRISPR-array R	TGC CGA GTA GGG TTG ATC TTT GTT GCA GAC CTG ACG GTT TC
5SrRNA compl. sequence	DIG - GGGGTCAGGTGGGACCACCGCGCTACGGCCGCC
TypeIV compl. spacer	DIG - CGTCAGGTCTGCAACAAAGATCAACCCTACTCG

Table 4.9: Oligonucleotides used for site directed mutagenesis.

Agilent site Quick change primer design tool was used for site directed mutagenesis primer design.

Name	**Sequence 5' → 3'**
Csf5 H44A F	CGG GTG ATC GCA GCA CCG GTC TGC GGA TCA C
Csf5 H44A R	GTG ATC CGC AGA CCG GTG CTG CGA TCA CCC G
Csf5 S209A F	GGT CGC ACC GGT GGC GGT CAC CAG ACG C
Csf5 S209A R	GCG TCT GGT GAC CGC CAC CGG TGC GAC C
Csf5 R38A F	CCG GTC TGC GGA TCA GCG CTA AAG AAC AGC GG
Csf5 R38A R	CCG CTG TTC TTT AGC GCT GAT CCG CAG ACC GG
Csf5 R23A F	CGC CAG CGC CTC AGC CAG ATC ATT CGG C
Csf5 R23A R	GCC GAA TGA TCT GGC TGA GGC GCT GGC G
Csf5 S241A F	GAC CGT AGC CAC GGG CGG TCA GGT TAC CCG
Csf5 S241A R	CGG GTA ACC TGA CCG CCC GTG GCT ACG GTC
Csf5 R242A F	ATA CGA CCG TAG CCA GCG CTG GTC AGG TTA CC

Csf5 R242A R	GGT AAC CTG ACC AGC GCT GGC TAC GGT CGT AT

4.4 Working with DNA

4.4.1 Preparation of plasmid DNA from *E. coli*

E. coli (2-5 ml) overnight cell cultures were pelleted and plasmid DNA isolation performed by using the QIAprep Spin Miniprep Kit or the QIAGEN Plasmid Plus Maxi Kit according to the manufacturer's instructions.

4.4.2 Phenol/chloroform extraction of DNA and nucleic acid precipitation

To extract bound nucleic acids from DinG protein, phenol-chloroform extraction followed by nucleic acid precipitation was applied. C-terminally tagged purified DinG was mixed with 1 volume of phenol/chloroform (1:1) pH 8.0. After centrifugation (15,000 x g, 1 min, RT), the upper aqueous phase was transferred into a fresh tube and 1 volume of chloroform was added and mixed with the sample. After centrifugation (15,000 x g, 1 min, RT), the upper aqueous phase was transferred into a fresh tube and DNA precipitation was performed via ethanol precipitation. 0.3 M Na-acetate and two volumes of 100 % ethanol (v/v) were added to the sample and incubated for 30 min at -20°C. The sample was centrifuged at 12,000 x g for 10 min at RT. Afterwards, the supernatant was removed and the DNA pellet was washed with 70 % ethanol (v/v) (12,000 x g, 2 min, RT). The supernatant was discarded and the DNA pellet was dried and resuspended in ddH_2O.

4.4.3 Spectrophotometric and Fluorometric quantitation of DNA

Concentration and the purity of DNA was controlled with a spectrophotometer. The Qubit fluorometer was also used for high sensitivity quantification. Quantification of isolated small RNAs of *A.aromaticum* EbN1, cDNA libraries for RNA-Seq and DinG bound nucleic acids were quantified by Qubit fluorometer. Prior to measurements Qubit fluorometer was calibrated with corresponding kit.

4.4.4 Agarose gel electrophoresis of DNA

Different lengths of the DNA fragments were separated by agarose gels. Depending on the length of the DNA, different concentrations of agarose gels were made. 0.8-2 % (w/v) agarose in TAE buffer (40 mM Tris-acetate, 1 mM EDTA, 20 mM acetic acid pH 8.0) and 0.1 µl/ml ethidium bromide (10 mg/ml) were used. Before loading onto the gel, samples were mixed with 6x loading dye (40 % (v/v) sucrose, 0.25 % (w/v) bromophenol blue, 0.25 % (w/v) xylene cyanol. Electrophoresis was performed at 90-120 V in TAE buffer. As size marker the 2-log DNA Ladder was used. After the electrophoresis the DNA was visualized by UV irradiation at 254 nm.

4.4.5 Amplification of plasmid DNA, genomic DNA or cDNA

PCR reactions were performed by mixing 50-100 ng of template DNA (genomic, plasmid or cDNA). 0.5 µM of each oligonucleotide, 200 µM dNTPs and the 10x reaction buffer of the suitable polymerase. 1 U of the Phusion polymerase or 2 U of the Taq polymerase were used. The PCR reactions were carried out in a thermal cycler. Standard cycling conditions are shown in table 4.10.

Table 4.10: Standard-PCR program using Phusion polymerase

	Step 1	Step 2-35		Step 36
Denaturation	98°C, 30 s	98°C	10 s	
Annealing		45-72°C	30 s	
Elongation		72°C	30 s/kb	72°C, 10 min

4.4.6 Site-directed mutagenesis PCR

Site directed mutagenesis PCR was used to introduce mutations. Agilent Quick change primer design tool was used to design site directed mutagenesis DNA oligos. Based on manufacturer`s instructions, PCR reaction was performed to obtain mutagenized plasmids and subjected to subsequent kinase, ligase and DpnI treatments (Site-Directed Mutagenesis Kit, NEB). Transformation of mutagenized plasmids were performed to chemically competent *E.coli* DH5α cells and selected based on the appropriate antibiotics. Plasmids from growing colonies were isolated by mini-prep plasmid isolation kit (Qiagen). Mutagenized plasmids were verified by Sanger sequencing.

4.4.7 PCR purification and gel extraction from agarose gels

PCR reactions were purified using QIAquick Gel extraction kit according to manufacturer´s protocol. DNA fragments after PCR or treatment with restriction endonucleases were separated by agarose gel electrophoresis. Corresponding DNA band was excised and gel extracted by using QIAquick Gel extraction kit based on manufacturer`s protocol.

4.4.8 Phosphorylation and hybridization of DNA oligonucleotides

Prior to hybridization, 1 nmol of each oligonucleotide was 5'-phosphorylated in separate reactions. Reaction mixture contains 50 U of T4-Polynucleotide Kinase (PNK), 1x of the corresponding reaction buffer, and 20 mM ATP together with oligonucleotide in a 20 µl reaction volume. The reaction mixture was incubated at 37°C for 1 h. Phosphorylated complementary DNA oligos were hybridized by mixing 1 nmol of each oligonucleotide with T4 DNA ligase buffer (10x) and ddH_2O in a reaction volume of 10 µl. Heating the reaction for 5 min at 95°C and gradual cool down to RT for 1-2 h. trigger hybridization. After the hybridization, the mixture was ligated into appropriate plasmids.

4.4.9 Restriction enzyme modification of DNA

The DNA was mixed with 10-20 U restriction enzyme/µg DNA and incubated at 37°C for 2 h. or overnight with appropriate buffer based on the manufacturer`s protocol.

4.4.10 5'-dephosphorylation of linearized vector

In order to avoid re-ligation of the restriction enzyme digested plasmid DNA, antarctic phosphatase treatment was used to remove the 5'-end phosphoryl groups. 0.5 U/ pmol of antarctic phosphatase, DNA and the corresponding buffer were added to mixture and then further incubated at 37°C for 1 h. before ligation reaction.

4.4.11 Ligation

T4 DNA ligase was used to catalyze phosphodiester bond formation between 3' hydroxyl and 5' phosphate groups of DNA molecules. 3:1 molar ratio of PCR fragment to plasmid DNA was used for the ligation reaction in a volume of 10 µl. A reaction mixture was incubated at 45°C for 5 min to dissolve secondary structures then 1 µl T4 DNA ligase and the appropriate ligation buffer (10x) was added and incubated overnight at 16°C. T4 DNA ligase was inactivated by incubating at 70°C for 10 min. Reaction mixture were then used for transformation.

4.4.12 Sequencing

Cloned constructs were verified by Sanger sequencing performed by Eurofins MWG Operon (Ebersberg). For the sequencing reaction standard primers were used.

4.4.13 5'- terminal radioactive labeling

Single-stranded DNA and RNA oligonucleotides were synthesized by MWG or SIGMA-Aldrich. 100 pmol of the oligonucleotides were mixed with 5 pmol of $\gamma[^{32}P]$-ATP (Hartmann Analytic), 10 U of T4 PNK and the corresponding reaction buffer (10x) and then incubated at 37°C for 1 h.

4.4.14 Denaturing polyacrylamide gel electrophoresis of radiolabeled DNA

Electrophoretic separation of the radiolabeled DNA was performed with (8 -12 %) of the polyacrylamide stock solution containing acrylamide and bisacrylamide in a ratio of 29:1, depending on the size of the oligonucleotide. Additionally the gels contained 8 M urea, TBE buffer (90 mM Tris, 2 mM EDTA pH 8.0, 90 mM boric acid), 0.1 % (v/v) APS und 0.1 % (v/v) TEMED were prepared. Prior to loading on the gel the samples were mixed with 2x formamide loading buffer (95 % (v/v) formamide, 0.025 % (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol, 5 mM EDTA pH 8.0) and incubated at 95°C for 5 min. Separation was carried out at 12W.

4.4.15 Detection of radiolabeled DNA by phosphorimaging

Gels or blots were exposed to phosphor screens overnight. The bands on the phosphor screen were visualized with a phosphoimager.

4.4.17 Extraction of radiolabeled DNA from urea-polyacrylamide gels

The radiolabeled DNA visualized by phosphorimager and then excised from the gel. Excised gel pieces dissolved in 500 µl of gel elution buffer (20 mM Tris/HCl pH 7.5, 250 mM sodium acetate, 1 mM EDTA pH 8.0). Samples were placed at -20°C for 30 min and then shaked on ice overnight. The supernatant was ethanol precipitated (section 4.4.2) with glycogen addition (1:100 ratio). The radioactivity was measured in a scintillation counter.

4.5 Working with RNA

4.5.1 Treatment of solutions, glassware and equipment

RNases were inactivated from all buffers and solutions by treatment of 0.1 % (v/v) diethylpyrocarbonate (DEPC), stirred overnight at RT and autoclaved. Reusable plastic ware was treated with RNase Exitus Plus™.

4.5.2 Small RNA isolation from *A.aromaticum* EbN1 with mirVana™ miRNA Isolation Kit (Ambion)

For the isolation of small RNAs (< 200 nt), the mirVana™ Isolation Kit was used. The isolation was performed according to the manufacturer's instructions. The homogenization of 5 g cells of *A. aromaticum EbN1* was carried out in 1 ml lysis/binding buffer using a glass Teflon homogenizer for 3 min on ice. After the addition of 100 µl miRNA Homogenate Additive™ and 10 min incubation on ice, the RNAs were phenol/chloroform extracted. To enrich for small RNAs the sample was brought to an ethanol concentration of 25 % to immobilize large RNAs on a glass-fiber filter. Small RNAs were collected in the filtrate. The ethanol concentration in the filtrate was increased to 55 % and the filtrate was applied to a second glass-fiber filter to immobilize small RNAs. Both RNA fractions were washed and eluted in 100 µl elution solution.

4.5.3 TRIzol Reagent (Ambion)

For clean northern blot analysis, total RNA was isolated for cleavage and alanine scanning assays. *E.coli* cultures with corresponding plasmids were harvested at exponential growth phase (OD_{600nm} 0.4-0.6). The cell pellet was re-suspended in 1 ml TRIzol Reagent and lysed for 5 min at room temperature. After addition of 200 µl chloroform, mixing and incubation at room temperature for 10 min, the solution was centrifuged (16,000xg, 5 min, 4°C). The upper aqueous phase was transferred to a fresh tube and the RNA precipitated by the addition of 500 µl isopropanol and incubation at -20°C for 10 min. The RNA was pelleted by centrifugation (16.000g,10 min,4°C).Then washed two times with 1ml 70% ethanol, dried at room temperature and re-suspended in 50 µl of H_2O_{DEPC}.

4.5.4 Small RNA Sequencing

Total RNA isolated from *A. aromaticum* EbN1 5 g of cell pellets were purified using mirVana RNA extraction kit (Ambion) with an enrichment of small RNAs (<250 nt). 3 µg

of isolated small RNA from both *A. aromaticum* EbN1 were phosphorylated with T4 PNK to allow for termini ligation: RNA was incubated at 37°C for 6 h with 10 U of T4 PNK (Ambion) and 10 µl 5x T4 PNK buffer (0.5 M Tris-HCI pH 6.5, 0,5 m MgAc, 25 mM β-mercaptoethanol) in a total volume of 50 µl. 1 mM ATP was added and the reaction mixture was incubated for 1 h at 37°C to obtain mono-phosphorylated 5′-termini. Afterwards, RNA library constructions with adaptors were prepared with an Illumina TruSeq RNA Sample Prep Kit and sequencing on an Illumina HiSeq2000 sequencer was performed at the Max-Planck Genomecentre Cologne.

4.5.5 Northern Blot Analysis

2 µg of total RNA was loaded onto a 8 M urea 10% denaturing polyacrylamide gel. A semidry electrophoretic transfer system was used to transfer RNA bands onto a positively charged membrane (Roti-Nylon plus, pore size 0.45 µm) for 2 h at 20 V and crosslinked with UV cross linker (120 seconds). After the transfer, the membrane was pre-hybridized for 30 min at 55°C in ULTRAhyb-Oligo hybridization buffer and non-specific binding was blocked. The membrane was incubated with 30 to 50 ng/µl of a DIG labeled probes in 8 ml of hybridization buffer overnight at 55°C. DIG bound antibodies were detected by X-ray imager film processor (Agfa-CP1000).

4.6 Biochemical methods

4.6.1 Cloning of expression constructs

Type IV CRISPR-Cas loci *cas* genes and were codon optimized and synthesized by Genescript and cloned into the two multiple cloning sites of pCDFDuet-1 (*csf1* and *csf5*, *pCsf1-Csf5*) and pRSFDuet-1 (*csf2* and *csf3*, *pCsf2-Csf3*) vectors (Novagen). In addition, csf5 and csf1 were cloned individually into pCDFDuet-1 to yield pCsf5 and pCsf1 plasmids and *csf2* and *csf3* were cloned individually into pRSFDuet-1 to yield *pCsf2* and

pCsf3 plasmids. A hexa-histidine tag was included at the 5′ end of *csf5* for Ni-NTA affinity chromatography of recombinant complexes. The repeat-spacer1-repeat sequence was placed under control of a T7 RNA polymerase promoter cloned into the pUC19 vector (NEB) BamHI and HindIII restriction sites using two with annealed oligonucleotides (pminiCRISPR, Table 4.8). QuikChange site-directed mutagenesis (Agilent) was applied to obtain plasmids for the expression of Csf5 mutants. Wildtype and Csf5 mutant plasmids were verified by sequencing. *dinG* (*csf4*) gene was also codon optimized and synthesized by Genescript and cloned into multiple cloning site of pRSF-Duet1. Type IV *cas* genes were cloned into pCDFDuet-1 multiple cloning site by inserting rbs site in between the *cas* genes (pCsf2-rbs-Csf3, Csf5-rbs-Csf1).

4.6.2 Production and purification of recombinant proteins

E. coli BL21-AI (Invitrogen) was transformed with the indicated expression plasmids (pCsf2-Csf3, pCsf1-Csf5 and pminiCRISPR were used to obtain the crRNP, pCsf5 and pminiCRISPR were used to purify RNA-bound Csf5). The expression cultures were grown at 37 °C in 1 liter LB medium with appropriate antibiotics (50 μg/ml kanamycin and spectinomycin and 100 μg/ml ampicillin) under rigorous shaking (200 r.p.m.) to an OD_{600} of 0.6 - 0.8. Expression was induced with 1 mM IPTG and 20% L-arabinose for 3-4 h. Cells were subsequently harvested by centrifugation (6000g, 4°C). Cell pellets were re-suspended in lysis buffer (20 mM HEPES pH 7.5, 300 mM NaCI and 10 mM imidazole). Cells were lysed by addition of lysozyme for 30 min on ice and subsequently sonified (8 x 30 s: Branson Sonifier 250). The cell lysate was clarified by centrifugation (18.000 r.p.m., 30 min, 4°C) and the supernatant was applied to a 5 ml HisTrap HP column (GE Healthcare). The column was washed with 10 column volumes of washing buffer (20 mM HEPES pH 7.5, 300 mM NaCI, 20 mM imidazole) before elution of bound proteins by an imidazole gradient (20 mM-500 mM). Eluates were concentrated to a final volume of 2 ml and injected on a HiLoad 16/600

Superdex 200 column (GE Healthcare) equilibrated in SEC buffer (20 mM HEPES pH 7.5, 300 mM NaCl) for size exclusion chromatography. Peak fractions were analyzed by SDS and urea-polyacrylamide gel electrophoresis. Csf5-crRNA complexes were purified by the same protocol and size-exclusion chromatography was either performed at room temperature or 4°C yielding two different substrates for crystallization. DinG + type IV crRNP purifications were performed at 4°C. After cell lysis, supernatant was applied to 1 ml HisTrap HP column. Ni-NTA purification performed as described above. Eluates were concentrated to a final volume of 500 µl and injected to Superose 6 Increase 10/300 column equilibrated in SEC buffer (20 mM HEPES pH 7.5, 150 mM NaCl) for size exclusion chromatography. Untagged DinG + type IV crRNP cross-linking samples were only purified by Ni-NTA purification and concentrated to final volume of 1 ml before chemical cross-linker was applied.

4.6.3 Electron microscopy

Carbon coated copper grids (400 mesh) were hydrophilized by glow discharging (PELCO easiGlow, Ted Pella, USA). 5 µl of a 13.7 µg/ml protein suspension was applied onto the hydrophilized grids and stained with 2% uranyl acetate after a short washing step with double-distilled H_2O. Samples were analyzed with a JEOL JEM-2100 transmission electron microscope using an acceleration voltage of 120 kV. A F214 FastScan CCD camera (TVIPS, Gauting) was used for image acquisition. For 2D class averaging, a sample with a concentration of 0.4 mg/ml was diluted in H2Obidest to a final concentration of 20 µg/ml and prepared as described above. 30 montages (2x2 to 6x6 stitched single 2k images) were taken at 200 kV and processed with cisTEM. For particle picking the cisTEM ab initio algorithm was used before manual editing (i.e. adding/deleting particles from the dataset). A total of ~5400 particles were used for averaging 10 classes (20 cycles) [128].

4.6.4 Crystallization and vaporizing iodine labeling

Purified Csf5 bound to the recognized crRNA repeat was concentrated to an absorbance at 280 nm of 50 AU (NanoDrop Lite Spectrophotometer) and subjected to crystallization by vapor diffusion at 20 °C. Crystals for vaporizing iodine labeling (VIL) were generated by combining 1 µL protein solution, containing at 20 °C purified Csf5 preparations, with 1 µL crystallization solution (0.16 M di-ammonium tartrate, 16% w/v PEG3350, 200 mM NaCl). Elongated block shaped crystals appeared after 2-3 days and were subsequently transferred into crystallization solution containing 20% glycerol for cryoprotection. The crystals were then incubated for 1.5 h in the presence of a 0.5 µl drop of KI/I_2 solution (0.67 M KI and 0.47 M I_2) for VIL according to [129]. Crystals were subsequently flash frozen and stored in liquid nitrogen. Native crystals of the 20°C preparations were generated by combining 1 µL protein solution with 1 µL crystallization solution (0.18 M di-ammonium tartrate, 18% w/v PEG3350, 3% glycerol). Crystals appeared after 2-3 days and were subsequently transferred into crystallization solution containing 20% glycerol for cryoprotection. Crystals were then flash frozen and stored in liquid nitrogen. Crystals of the 4°C preparations were generated by combining 1 µL protein solution with 1 µL crystallization solution (0.2 M di-sodium tartrate, 20% w/v PEG3350). Block shaped crystals appeared after 4 days and were subsequently transferred into crystallization solution containing 20 % glycerol for cryoprotection. Crystals were then flash frozen and stored in liquid nitrogen.

4.6.5 Protein-protein and protein-RNA crosslinking

To identify amino acids that interact with the crRNAs, 100 µg of RNA-bound Type IV crRNPs was UV-irradiated at 254 nm for 10 min, while a second sample served as a non-irradiated control. The protein-RNA samples were loaded onto a pre-packed C18 column (Harvard Apparatus, Microspin C18 Column, Massachusetts, United States), mounted to a Dionex UltiMate 3000 UHPLC$^+$ focused (Thermo Scientific), equipped with an analytical

column (75 µm x 300 mm, ReproSil-Pur 120 C18-AQ, 1.9 µm, Dr. Maisch GmbH, packed in-house). Peptides were separated by reverse-phase chromatography on a 58 min multi-step gradient (flow rate of 0.3–0.4 µl/min) before entering the mass spectrometer (QExactive HF-X, Thermo Scientific). MS1 spectra were recorded in profile mode (resolution of 120k), and MS2 spectra were recorded in centroid mode (resolution of 30k); isolation window set to 1.6 m/z and dynamic exclusion set to 7 s. Raw data of RNA-protein heteroconjugates were analysed and manually validated with the OpenMS pipeline RNP xl as previously described by (Kramer et al., 2014). For protein-protein crosslinking, 10 µg-aliquots of the complex were incubated with either 0.05, 0.1, 0.25, 0.5 or 1 mM of bis (sulfosuccinimidyl) suberate (BS3, Thermo Scientific). For DinG interaction analysis, DSBU chemical cross-linker was applied. Each of the samples were incubated at room temperature for 30 min and subsequently quenched by a final concentration of 50 mM Tris. Proteins were then separated by PAGE using a 4-12% gradient gel (NuPAGE, Invitrogen). The upper quarter of the lanes containing the complex crosslinked with 0.5 mM and 1 mM BS3 were cut into four equally sized slices and proteins were in-gel digested. Briefly, proteins were reduced and alkylated by 10 mM dithiotreitol and 55 mM iodoacetamide, respectively, and finally digested with trypsin at 37°C for 18 h. Extracted and dried peptides were dissolved in 2% ACN/0.05% TFA and subjected to LC-MS using the above mentioned setup with the following changes: an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Scientific) was used with a dynamic exclusion of 10 s and technical duplicates were measured. Raw files were converted to mgf format with ProteomeDiscoverer 1.4 (Thermo Scientic) and analysed with the software pLink (v. 1.23, pFind group for identification of crosslinked peptides. Here, default settings were applied with carbamidomethylation of cysteines as fixed and oxidation of methionines as variable modification, FDR was set to 0.01. Results were filtered by excluding crosslinks supported

by only one spectrum and additionally by applying a score cut-off value of 3. Interaction networks were visualised by xiNET.

5. References

1. Carte, J., et al., *Binding and cleavage of CRISPR RNA by Cas6.* Rna-a Publication of the Rna Society, 2010. **16**(11): p. 2181-2188.
2. Koonin, E.V. and Y.I. Wolf, *Evolution of the CRISPR-Cas adaptive immunity systems in prokaryotes: models and observations on virus-host coevolution.* Mol Biosyst, 2015. **11**(1): p. 20-7.
3. Makarova, K.S., et al., *Defense islands in bacterial and archaeal genomes and prediction of novel defense systems.* J Bacteriol, 2011. **193**(21): p. 6039-56.
4. Koonin, E.V., K.S. Makarova, and Y.I. Wolf, *Evolutionary Genomics of Defense Systems in Archaea and Bacteria.* Annu Rev Microbiol, 2017. **71**: p. 233-261.
5. Iranzo, J., et al., *Evolutionary dynamics of the prokaryotic adaptive immunity system CRISPR-Cas in an explicit ecological context.* J Bacteriol, 2013. **195**(17): p. 3834-44.
6. tenOever, B.R., *The Evolution of Antiviral Defense Systems.* Cell Host Microbe, 2016. **19**(2): p. 142-9.
7. Koonin, E.V. and M. Krupovic, *Evolution of adaptive immunity from transposable elements combined with innate immune systems.* Nat Rev Genet, 2015. **16**(3): p. 184-92.
8. Gordeeva, J., et al., *BREX system of Escherichia coli distinguishes self from non-self by methylation of a specific DNA site.* Nucleic Acids Research, 2019. **47**(1): p. 253-265.
9. van Houte, S., A. Buckling, and E.R. Westra, *Evolutionary Ecology of Prokaryotic Immune Mechanisms.* Microbiol Mol Biol Rev, 2016. **80**(3): p. 745-63.
10. Labrie, S.J., J.E. Samson, and S. Moineau, *Bacteriophage resistance mechanisms.* Nat Rev Microbiol, 2010. **8**(5): p. 317-27.
11. Samuel, A.D., et al., *Flagellar determinants of bacterial sensitivity to chi-phage.* Proc Natl Acad Sci U S A, 1999. **96**(17): p. 9863-6.
12. Tan, D., S.L. Svenningsen, and M. Middelboe, *Quorum Sensing Determines the Choice of Antiphage Defense Strategy in Vibrio anguillarum.* MBio, 2015. **6**(3): p. e00627.
13. Hoyland-Kroghsbo, N.M., R.B. Maerkedahl, and S.L. Svenningsen, *A quorum-sensing-induced bacteriophage defense mechanism.* MBio, 2013. **4**(1): p. e00362-12.
14. Ofir, G., et al., *DISARM is a widespread bacterial defence system with broad anti-phage activities.* Nature Microbiology, 2018. **3**(1).
15. Goldfarb, T., et al., *BREX is a novel phage resistance system widespread in microbial genomes.* Embo Journal, 2015. **34**(2): p. 169-183.
16. Roberts, R.J., et al., *A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes.* Nucleic Acids Research, 2003. **31**(7): p. 1805-1812.
17. Korona, R. and B.R. Levin, *Phage-Mediated Selection and the Evolution and Maintenance of Restriction-Modification.* Evolution, 1993. **47**(2): p. 556-575.
18. Sneppen, K., et al., *Restriction modification systems as engines of diversity.* Frontiers in Microbiology, 2015. **6**.
19. Vasu, K. and V. Nagaraja, *Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense.* Microbiology and Molecular Biology Reviews, 2013. **77**(1): p. 53-72.
20. Kruger, D.H. and T.A. Bickle, *Bacteriophage Survival - Multiple Mechanisms for Avoiding the Deoxyribonucleic-Acid Restriction Systems of Their Hosts.* Microbiological Reviews, 1983. **47**(3): p. 345-360.

21. Tock, M.R. and D.T.F. Dryden, *The biology of restriction and anti-restriction.* Current Opinion in Microbiology, 2005. **8**(4): p. 466-472.
22. Butterer, A., et al., *Type III restriction endonucleases are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA.* Nucleic Acids Research, 2014. **42**(8): p. 5139-5150.
23. Doron, S., et al., *Systematic discovery of antiphage defense systems in the microbial pangenome.* Science, 2018. **359**(6379).
24. Willkomm, S., K.S. Makarova, and D. Grohmann, *DNA silencing by prokaryotic Argonaute proteins adds a new layer of defense against invading nucleic acids.* Fems Microbiology Reviews, 2018. **42**(3): p. 376-387.
25. Willkomm, S., et al., *Structural and mechanistic insights into an archaeal DNA-guided Argonaute protein.* Nature Microbiology, 2017. **2**(6).
26. Swarts, D.C., et al., *DNA-guided DNA interference by a prokaryotic Argonaute.* Nature, 2014. **507**(7491): p. 258-+.
27. Swarts, D.C., et al., *The evolutionary journey of Argonaute proteins.* Nature Structural & Molecular Biology, 2014. **21**(9): p. 743-753.
28. Makarova, K.S., et al., *Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements.* Biology Direct, 2009. **4**.
29. Ryazansky, S., A. Kulbachinskiy, and A.A. Aravin, *The Expanded Universe of Prokaryotic Argonaute Proteins.* Mbio, 2018. **9**(6).
30. Yamaguchi, Y., J.H. Park, and M. Inouye, *Toxin-Antitoxin Systems in Bacteria and Archaea.* Annual Review of Genetics, Vol 45, 2011. **45**: p. 61-79.
31. Makarova, K.S., N.V. Grishin, and E.V. Koonin, *The HicAB cassette, a putative novel, RNA-targeting toxin-antitoxin system in archaea and bacteria.* Bioinformatics, 2006. **22**(21): p. 2581-2584.
32. Koonin, E.V. and F. Zhang, *Coupling immunity and programmed cell suicide in prokaryotes: Life-or-death choices.* Bioessays, 2017. **39**(1).
33. Chopin, M.C., A. Chopin, and E. Bidnenko, *Phage abortive infection in lactococci: variations on a theme.* Current Opinion in Microbiology, 2005. **8**(4): p. 473-479.
34. Snyder, L., *Phage-Exclusion Enzymes - a Bonanza of Biochemical and Cell Biology Reagents.* Molecular Microbiology, 1995. **15**(3): p. 415-420.
35. Dy, R.L., et al., *A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism.* Nucleic Acids Research, 2014. **42**(7): p. 4590-4605.
36. Fineran, P.C., et al., *The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(3): p. 894-899.
37. Ishino, Y., et al., *Nucleotide-Sequence of the Iap Gene, Responsible for Alkaline-Phosphatase Isozyme Conversion in Escherichia-Coli, and Identification of the Gene-Product.* Journal of Bacteriology, 1987. **169**(12): p. 5429-5433.
38. Mojica, F.J.M., G. Juez, and F. Rodriguezvalera, *Transcription at Different Salinities of Haloferax-Mediterranei Sequences Adjacent to Partially Modified Psti Sites.* Molecular Microbiology, 1993. **9**(3): p. 613-621.
39. Mojica, F.J.M., et al., *Long Stretches of Short Tandem Repeats Are Present in the Largest Replicons of the Archaea Haloferax-Mediterranei and Haloferax-Volcanii and Could Be Involved in Replicon Partitioning.* Molecular Microbiology, 1995. **17**(1): p. 85-93.
40. Mojica, F.J.M., et al., *Biological significance of a family of regularly spaced repeats in the genomes of Archaea, bacteria and mitochondria.* Molecular Microbiology, 2000. **36**(1): p. 244-246.

41. Mojica, F.J.M., et al., *Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements.* Journal of Molecular Evolution, 2005. **60**(2): p. 174-182.
42. Pourcel, C., G. Salvignol, and G. Vergnaud, *CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies.* Microbiology-Sgm, 2005. **151**: p. 653-663.
43. Rath, D., et al., *The CRISPR-Cas immune system: biology, mechanisms and applications.* Biochimie, 2015. **117**: p. 119-28.
44. Koujah, L., D. Shukla, and A.R. Naqvi, *CRISPR-Cas based targeting of host and viral genes as an antiviral strategy.* Semin Cell Dev Biol, 2019.
45. Amitai, G. and R. Sorek, *CRISPR-Cas adaptation: insights into the mechanism of action.* Nature Reviews Microbiology, 2016. **14**(2): p. 67-76.
46. Plagens, A., et al., *DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes.* Fems Microbiology Reviews, 2015. **39**(3): p. 442-463.
47. Sternberg, S.H., et al., *Adaptation in CRISPR-Cas Systems.* Molecular Cell, 2016. **61**(6): p. 797-808.
48. Staals, R.H.J., et al., *Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system.* Nature Communications, 2016. **7**.
49. Hynes, A.P., et al., *Detecting natural adaptation of the Streptococcus thermophilus CRISPR-Cas systems in research and classroom settings.* Nature Protocols, 2017. **12**(3): p. 547-565.
50. Cooper, L.A., A.M. Stringer, and J.T. Wade, *Determining the Specificity of Cascade Binding, Interference, and Primed Adaptation In Vivo in the Escherichia coli Type I-E CRISPR-Cas System.* Mbio, 2018. **9**(2).
51. Nunez, J.K., et al., *Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity.* Nature Structural & Molecular Biology, 2014. **21**(6): p. 528-534.
52. Amitai, G. and R. Sorek, *CRISPR-Cas adaptation: insights into the mechanism of action.* Nat Rev Microbiol, 2016. **14**(2): p. 67-76.
53. Barrangou, R. and L.A. Marraffini, *CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity.* Mol Cell, 2014. **54**(2): p. 234-44.
54. McGinn, J. and L.A. Marraffini, *Molecular mechanisms of CRISPR-Cas spacer acquisition.* Nat Rev Microbiol, 2019. **17**(1): p. 7-12.
55. Hou, Z.G. and Y. Zhang, *Insights into a Mysterious CRISPR Adaptation Factor, Cas4.* Molecular Cell, 2018. **70**(5): p. 757-758.
56. Lee, H., Y. Dhingra, and D.G. Sashital, *The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation.* Elife, 2019. **8**.
57. Kieper, S.N., et al., *Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation.* Cell Reports, 2018. **22**(13): p. 3377-3384.
58. Yosef, I., M.G. Goren, and U. Qimron, *Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli.* Nucleic Acids Res, 2012. **40**(12): p. 5569-76.
59. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-7.
60. Charpentier, E., et al., *Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity.* FEMS Microbiol Rev, 2015. **39**(3): p. 428-41.
61. Plagens, A., et al., *DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes.* FEMS Microbiol Rev, 2015. **39**(3): p. 442-63.
62. Xiao, Y.B., et al., *Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System.* Cell, 2017. **170**(1): p. 48-+.

63. Rutkauskas, M., et al., *Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection.* Cell Reports, 2015. **10**(9): p. 1534-1543.
64. Szczelkun, M.D., et al., *Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(27): p. 9798-9803.
65. Sinkunas, T., et al., *Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system.* Embo Journal, 2011. **30**(7): p. 1335-1342.
66. Koonin, E.V., K.S. Makarova, and F. Zhang, *Diversity, classification and evolution of CRISPR-Cas systems.* Current Opinion in Microbiology, 2017. **37**: p. 67-78.
67. Makarova, K.S., et al., *An updated evolutionary classification of CRISPR-Cas systems.* Nature Reviews Microbiology, 2015. **13**(11): p. 722-736.
68. Mohanraju, P., et al., *Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems.* Science, 2016. **353**(6299).
69. Nam, K.H., et al., *Cas5d Protein Processes Pre-crRNA and Assembles into a Cascade-like Interference Complex in Subtype I-C/Dvulg CRISPR-Cas System.* Structure, 2012. **20**(9): p. 1574-1584.
70. Zhu, Y., et al., *Shooting the messenger: RNA-targetting CRISPR-Cas systems.* Biosci Rep, 2018. **38**(3).
71. Silas, S., et al., *Type III CRISPR-Cas systems can provide redundancy to counteract viral escape from type I systems.* Elife, 2017. **6**.
72. Kazlauskiene, M., et al., *A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems.* Science, 2017. **357**(6351): p. 605-+.
73. Han, W.Y., et al., *A type III-B CRISPR-Cas effector complex mediating massive target DNA destruction.* Nucleic Acids Research, 2017. **45**(4): p. 1983-1993.
74. Johnson, K., et al., *-Target sequence requirements of a type III-B CRISPR-Cas immune system.* J Biol Chem, 2019.
75. Ichikawa, H.T., et al., *Programmable type III-A CRISPR-Cas DNA targeting modules.* PLoS One, 2017. **12**(4): p. e0176221.
76. Peng, W., et al., *An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference.* Nucleic Acids Res, 2015. **43**(1): p. 406-17.
77. Zhao, H., et al., *Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli.* Nature, 2014. **515**(7525): p. 147-50.
78. Wiedenheft, B., et al., *Structures of the RNA-guided surveillance complex from a bacterial immune system.* Nature, 2011. **477**(7365): p. 486-489.
79. Jackson, R.N., et al., *Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli.* Science, 2014. **345**(6203): p. 1473-9.
80. Koonin, E.V., K.S. Makarova, and F. Zhang, *Diversity, classification and evolution of CRISPR-Cas systems.* Curr Opin Microbiol, 2017. **37**: p. 67-78.
81. Makarova, K.S. and E.V. Koonin, *Annotation and Classification of CRISPR-Cas Systems.* Methods Mol Biol, 2015. **1311**: p. 47-75.
82. Ozcan, A., et al., *Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum.* Nature Microbiology, 2019. **4**(1): p. 89-+.
83. Koonin, E.V., *Similarities in Rna Helicases.* Nature, 1991. **352**(6333): p. 290-290.
84. Cheng, K.Y. and D.B. Wigley, *DNA translocation mechanism of an XPD family helicase.* Elife, 2018. **7**.
85. White, M.F., *Structure, function and evolution of the XPD family of iron-sulfur-containing 5 '-> 3 ' DNA helicases.* Biochemical Society Transactions, 2009. **37**: p. 547-551.

86. Rudolf, J., et al., *The DNA repair helicases XPD and FancJ have essential iron-sulfur domains.* Molecular Cell, 2006. **23**(6): p. 801-808.
87. Voloshin, O.N., et al., *Characterization of the DNA damage-inducible helicase DinG from Escherichia coli.* Journal of Biological Chemistry, 2003. **278**(30): p. 28284-28293.
88. Voloshin, O.N. and R.D. Camerini-Otero, *The DinG protein from Escherichia coli is a structure-specific helicase.* Journal of Biological Chemistry, 2007. **282**(25): p. 18437-18447.
89. Thakur, R.S., et al., *Mycobacterium tuberculosis DinG Is a Structure-specific Helicase That Unwinds G4 DNA IMPLICATIONS FOR TARGETING G4 DNA AS A NOVEL THERAPEUTIC APPROACH.* Journal of Biological Chemistry, 2014. **289**(36): p. 25112-25136.
90. McRobbie, A.M., et al., *Staphylococcus aureus DinG, a helicase that has evolved into a nuclease.* Biochemical Journal, 2012. **442**: p. 77-84.
91. Gunderson, F.F., et al., *Nuclease Activity of Legionella pneumophila Cas2 Promotes Intracellular Infection of Amoebal Host Cells.* Infection and Immunity, 2015. **83**(3): p. 1008-1018.
92. Babu, M., et al., *A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair.* Mol Microbiol, 2011. **79**(2): p. 484-502.
93. Gunderson, F.F. and N.P. Cianciotto, *The CRISPR-Associated Gene cas2 of Legionella pneumophila Is Required for Intracellular Infection of Amoebae.* Mbio, 2013. **4**(2).
94. Westra, E.R., A. Buckling, and P.C. Fineran, *CRISPR-Cas systems: beyond adaptive immunity.* Nat Rev Microbiol, 2014. **12**(5): p. 317-26.
95. Sampson, T.R., et al., *A CRISPR/Cas system mediates bacterial innate immune evasion and virulence.* Nature, 2013. **497**(7448): p. 254-7.
96. Shabbir, M.A.B., et al., *The Involvement of the Cas9 Gene in Virulence of Campylobacter jejuni.* Front Cell Infect Microbiol, 2018. **8**: p. 285.
97. Williams, E., et al., *Microarray analysis of the hyperthermophilic archaeon Pyrococcus furiosus exposed to gamma irradiation.* Extremophiles, 2007. **11**(1): p. 19-29.
98. Boonyaratanakornkit, B.B., L.Y. Miao, and D.S. Clark, *Transcriptional responses of the deep-sea hyperthermophile Methanocaldococcus jannaschii under shifting extremes of temperature and pressure.* Extremophiles, 2007. **11**(3): p. 495-503.
99. Strand, K.R., et al., *Oxidative stress protection and the repair response to hydrogen peroxide in the hyperthermophilic archaeon Pyrococcus furiosus and in related species.* Arch Microbiol, 2010. **192**(6): p. 447-59.
100. Gotz, D., et al., *Responses of hyperthermophilic crenarchaea to UV irradiation.* Genome Biol, 2007. **8**(10): p. R220.
101. Veesenmeyer, J.L., et al., *NilD CRISPR RNA contributes to Xenorhabdus nematophila colonization of symbiotic host nematodes.* Mol Microbiol, 2014. **93**(5): p. 1026-42.
102. Peters, J.E., et al., *Recruitment of CRISPR-Cas systems by Tn7-like transposons.* Proc Natl Acad Sci U S A, 2017. **114**(35): p. E7358-E7366.
103. Craig, N.L., *Tn7: a target site-specific transposon.* Mol Microbiol, 1991. **5**(11): p. 2569-73.
104. Finn, J.A., A.R. Parks, and J.E. Peters, *Transposon Tn7 directs transposition into the genome of filamentous bacteriophage M13 using the element-encoded TnsE protein.* Journal of Bacteriology, 2007. **189**(24): p. 9122-9125.
105. Peters, J.E. and N.L. Craig, *Tn7: Smarter than we thought.* Nature Reviews Molecular Cell Biology, 2001. **2**(11): p. 806-814.

106. Rao, J.E., P.S. Miller, and N.L. Craig, *Recognition of triple-helical DNA structures by transposon Tn7.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(8): p. 3936-3941.
107. Mitra, R., et al., *Characterization of the TnsD-attTn7 complex that promotes site-specific insertion of Tn7.* Mobile DNA, 2010. **1**.
108. Choi, K.Y., J.M. Spencer, and N.L. Craig, *The Tn7 transposition regulator TnsC interacts with the transposase subunit TnsB and target selector TnsD.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(28): p. E2858-E2865.
109. Gleditzsch, D., et al., *Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system.* Nucleic Acids Res, 2016. **44**(12): p. 5872-82.
110. Pausch, P., et al., *Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance.* Mol Cell, 2017. **67**(4): p. 622-632 e4.
111. Nam, K.H., et al., *Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system.* Structure, 2012. **20**(9): p. 1574-84.
112. Shmakov, S.A., et al., *The CRISPR Spacer Space Is Dominated by Sequences from Species-Specific Mobilomes.* Mbio, 2017. **8**(5).
113. Biswas, A., et al., *CRISPRDetect: A flexible algorithm to define CRISPR arrays.* BMC Genomics, 2016. **17**: p. 356.
114. Hochstrasser, M.L. and J.A. Doudna, *Cutting it close: CRISPR-associated endoribonuclease structure and function.* Trends Biochem Sci, 2015. **40**(1): p. 58-66.
115. Brouns, S.J., et al., *Small CRISPR RNAs guide antiviral defense in prokaryotes.* Science, 2008. **321**(5891): p. 960-4.
116. Sokolowski, R.D., S. Graham, and M.F. White, *Cas6 specificity and CRISPR RNA loading in a complex CRISPR-Cas system.* Nucleic Acids Res, 2014. **42**(10): p. 6532-41.
117. Reeks, J., et al., *Structure of a dimeric crenarchaeal Cas6 enzyme with an atypical active site for CRISPR RNA processing.* Biochem J, 2013. **452**(2): p. 223-30.
118. Ebihara, A., et al., *Crystal structure of hypothetical protein TTHB192 from Thermus thermophilus HB8 reveals a new protein family with an RNA recognition motif-like domain.* Protein Sci, 2006. **15**(6): p. 1494-9.
119. Niewoehner, O., M. Jinek, and J.A. Doudna, *Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases.* Nucleic Acids Res, 2014. **42**(2): p. 1341-53.
120. Shao, Y. and H. Li, *Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6.* Structure, 2013. **21**(3): p. 385-93.
121. Richter, H., et al., *Comparative analysis of Cas6b processing and CRISPR RNA stability.* Rna Biology, 2013. **10**(5): p. 700-707.
122. Mulepati, S., A. Heroux, and S. Bailey, *Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target.* Science, 2014. **345**(6203): p. 1479-84.
123. Westra, E.R., et al., *Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition.* PLoS Genet, 2013. **9**(9): p. e1003742.
124. Larson, M.H., et al., *CRISPR interference (CRISPRi) for sequence-specific control of gene expression.* Nature Protocols, 2013. **8**(11): p. 2180-2196.
125. Dwarakanath, S., et al., *Interference activity of a minimal Type I CRISPR-Cas system from Shewanella putrefaciens.* Nucleic Acids Research, 2015. **43**(18): p. 8913-8923.
126. Hanahan, D., *Studies on Transformation of Escherichia-Coli with Plasmids.* Journal of Molecular Biology, 1983. **166**(4): p. 557-580.

127. Rabus, R. and F. Widdel, *Anaerobic degradation of ethylbenzene and other aromatic hydrocarbons by new denitrifying bacteria.* Arch Microbiol, 1995. **163**(2): p. 96-103.
128. Grant, T., A. Rohou, and N. Grigorieff, *cisTEM, user-friendly software for single-particle image processing.* Elife, 2018. **7**.
129. Miyatake, H., T. Hasegawa, and A. Yamano, *New methods to prepare iodinated derivatives by vaporizing iodine labelling (VIL) and hydrogen peroxide VIL (HYPER-VIL).* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 3): p. 280-9.

Contributions

Contributions are listed in the paper as following:

A.Ö. and P.P. purified proteins. P.P. determined the crystal structure. A.L., A.W. and H.U. performed mass spectrometry analyses. K.S. and J.H. cultured *A. aromatoleum* strains. T.H. performed transmission electron microscopy analyses. L.R and A.Ö. conceived the experiments. L.R. wrote the manuscript with support from A.Ö., P.P., G.B., J.H., T.H. and H.U.

Detailed collaborations are listed below:

I established recombinant production and purification of all Type IV crRNP components, identified the crRNA endonuclease using in vivo crRNA maturation assays and established crRNP formation. Furthermore, I worked with the following collaboration partners and their contributions are listed below:

Patrick Pausch (AG Bange, Philipps University, Marburg)

Patrick Pausch and I purified proteins. Patrick Pausch determined the crystal structure of Csf5.

Andreas Linden (AG Urlaub, MPI-Göttingen)

Protein-protein cross-linking: I purified the type IV crRNP complex. Mass spectrometry analyses were performed by Andreas Linden.

Alexander Wulf (AG Urlaub, MPI-Göttingen)

Protein-RNA cross-linking: I purified the type IV crRNP complex. Alexander Wulf carried mass spectrometry analysis of RNA-protein cross-links.

Thomas Heimerl (AG Bange, Philipps University, Marburg)

Electron Microscopy: Type IV crRNP complexes and Csf2 filaments were purified by me and Thomas Heimerl generated electron microcopy images, 2D classes and 3D construction of the type IV crRNP complex.

Karola Schühle (AG Heider) kindly shared *A.aromaticum EbN1* pellet with me**.** This pellet was used for small RNA isolation and further RNA-Seq analysis.

UNPUBLISHED RESULTS

The following section lists contribution by master students that worked under my supervision and collaboration partners for unpublished results that are described in this thesis:

Marcus Ziemann (Master Student)

Philipps University, Marburg, May 2017-March 2019

I supervised the work performed by Marcus Ziemann during his master thesis study on the `` Analysis of PAM recognition by the type IV CRISPR-Cas system of *Aromatoleum aromaticum*``. Marcus Ziemann established PAM depletion assays for type IV CRISPR-Cas system and performed most of the PAM associated bioinformatic analyses.

Sergei Shmakov (National Institute of Health-Koonin Group) kindly shared type IV protospacers with targets. These protospacers were used for computational determination of PAM sequence conservation of the Type IV CRISPR-Cas system by Marcus Ziemann.

Eva Grümpel (Master Student)

Philipps University, Marburg, since October 2018

I supervised the initial work performed by Eva Grümpel during her master studies on the ´´Biochemical & Functional Characterization of the Type IV CRISPR-Cas associated DinG helicase (Csf4) of *Aromatoleum aromaticum* EbN1´´.

I designed and cloned multiple constructs with different tags of the DinG helicase and performed initial purification and nucleic acid detection analyses. Constructs and initial purification conditions were established and provided to Eva Grümpel.

Lyle Kroell (Master Student,): I purified type IV CRISPR-Cas complexes and following protein-protein crosslinking titrations and HPLC-MS were performed by Lyle Kroell.

Jörg Kahnt (Mass Spectrometry and Proteomics Facility, MPI, Marburg)

Jörg Kahnt performed the mass spectrometry analysis of purified proteins.

Abbreviations

ad	add to the volume
APS	ammonium persulfate
ATP	adenosine triphosphate
β-Me	β-mercaptoethanol
bp	basepair
BSA	bovine serum albumin
cDNA	complementary DNA
C-terminal	carboxy-terminal
cpm	counts per minute
CV	column volume
kDa	Kilo Dalton
ddH2O	two times destilled water
DEPC	diethylpyrocarbonate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	desoxyribonuclease
dNTP	deoxyribonucleotide triphosphate
ds	downstream
dsDNA	double-stranded DNA
DTT	dithiothreitol
e.g.	for example
EDTA	ethylene-diamine-tetraacetic acid
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
et al.	*et alteri* = and others
HPLC	High-performance liquid chromatography
g	gram
x g	gravitational acceleration

h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	histidine tag
IPTG	isopropyl β-D-1-thiogalactopyranoside
kb	kilobase
l	liter
LB	lysogeny broth
M	molar (mol/l)
min	minutes
MW	molecular weight
μ	micro (10^{-6})
n	nano (10^{-9})
Ni^{2+}-NTA	nickel nitriloacetic acid
nt	nucleotides
N-terminal	amino-terminal
NTP	nucleoside triphosphate
OD_{600nm}	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
RNA	ribonucleic acid
RNase	ribonuclease
RNP	ribonucleoprotein complex
rpm	rounds per minute
rRNA	ribosomal RNA
RT	room temperature
s	seconds
SANS	small-angle neutron scattering
SAXS	small-angle X-ray scattering

SDS	sodium dodecyl sulfate
crRNA	CRISPR RNA
ssDNA	single-stranded DNA
ssRNA	single stranded RNA
TAE	tris-acetate-EDTA-buffer
TBE	tris-borate-EDTA-buffer
TEMED	N,N,N';N'-tetramethylethylenediamide
Tris	tris-(hydroxymethyl)-aminomethane
U	unit (enzyme activity)
us	upstream
UTR	untranslated region
UV	ultraviolet
W	watt
w/o	without
% (v/v)	percent by volume
% (w/v)	percent by weight
>	higher than
<	lower than
Δ	deletion

Curriculum vitae

Current position

March 2016/ expected May 2019 (3.2 Years)

Max Planck Institute for Terrestrial Microbiology
Doctoral studies at the Ludwig Maximilian University of Munich

Field: CRISPR-Cas Biology & Microbiology

Topic: Characterization of the type IV CRISPR-Cas system of the *Aromatoleum aromaticum EbN1*.

Supervision: Dr. Lennart Randau

Education

10/2014-09/2016 **Sabancı University, Istanbul**
Master of Science in Molecular Biology, Genetics and Bioengineering (CGPA:3.78/4.00)

Field: Immunology & Genome Editing

Topic: Glucocorticoid Related Drug Resistance and Investigation of Mechanism by Using Novel Gene Therapy Application

Supervision: Prof. Dr. Batu Erman

9/2012-07/2013 **Virginia Commonwealth University, Richmond**
Fulbright Scholar in Human Genetics Department

9/2006-07/2011 **Middle East Technical University, Ankara**
Bachelor of Science in Biology (CGPA: 3.50/4.00)

High Honor Student

*Top ranked university.
*English-medium education.

08/2002-06/2006 **Adana Anatolian High School, Adana**
High School Education (GPA: 5.00/5.00)
Specialization of Science, Mathematics and English Literature

Publications

2019 **Özcan, A.**, Pausch, P., Linden, A., Wulf, A., Schühle, K., Heider, J., Urlaub, H., Heimerl, T., Bange, G. and Randau, L. (2018) Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. **Nature Microbiol.** 4, 89-96. (First Author)

2018 Gleditzsch, D., Pausch, P., Müller-Esparza, H., **Özcan, A.**, Guo, X., Bange, G. and Randau L. (2018) PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. RNA Biol. 18, 1-14.

(In this review article, I wrote type II Cas9 part)

Awards

FINALIST in National Science Projects Competition for High School Students (2005)

Description: Recombinant production of amylase enzymes in *B. subtilis*
The Scientific and Technological Research Council of Turkey

SILVER MEDAL in IGEM (2009**)**
International Genetically Engineered Machine Competition (IGEM 2009), Massachusetts Institute of Technology (MIT)

Description: Designing wound dressing with fast healing mechanism (Cloning and expression of human epidermal growth factor with ABC transporter system and oxygen promoter into *E.coli*)

Bachelor Award
3.50/4.00 CGPA (High Honor)
Ranked in 3rd place in peers according to CGPA in top ranked university in Turkey.

Fulbright Scholarship
2012-2013-Graduate studies in Human Genetics Department at Virginia Commonwealth University

Scholarships and Fundings

2010-2011 Erasmus Scholarship-Czech Republic

2011 Radboud University Summer Research School Scholarship

Fulbright Scholarship
2012-2013-Graduate studies in Human Genetics Department at Virginia Commonwealth University

2014-2016- Sabancı University, Graduate and Research Assistant

2016-2017-Scientific and Technological Research Council of Turkey, 2213 Funding Program
Description: Covered first year of the doctoral studies.

2017-2019- Max-Planck Institute, IMPRS program

Language Skills

Turkish: Native Speaker

English: Proficient Speaker (TOEFL IBT: 95/120)

German: B1 level

Internships

29/06/2009 - 05/09/2009	**Arizona State University, USA** School of Life Sciences Undergraduate Research Assistant **Technical Skills:** Real Time PCR analysis and Functional Gene Expressions
28/01/2008 - 01/08/2010	**Middle East Technical University, Turkey** METU Central Laboratory Molecular Biology and Biotechnology Research Center Undergraduate Research Assistant **Technical Skills:** RNA isolations, Protein quantification, Western Blotting
16/11/2010 - 5/03/2011	**Institute of Chemical Technology, Czech Republic** Department of Biochemistry and Microbiology

Undergraduate Research Assistant

Technical Skills: Protein-protein interactions
yeast two hybrid system, transformation, Western Blotting, SDS PAGE, Yeast cultivation techniques.

15/06/2011 - 15/09/2011 **Radboud University, Netherlands**
Human Genetics Department
Undergraduate Research Assistant

Technical Skills: Chromatin Immunoprecipitation Sequencing (ChIP-Seq)

Acknowledgements

I would like to express my deep gratitude to Dr. Lennart Randau my research supervisor, for his patient guidance, scientific vision and useful critiques of this research work. I would also like to thank Prof. Dr. Kai Papenfort for his assistance in keeping my progress on schedule. My grateful thanks are also extended to my TAC committee members Prof. Dr. Tobias Erb and Prof. Dr. Johann Heider for inspirations and helpful comments.

I would like to offer my special thanks to two students who gave great contributions to progress of this work: Marcus Ziemann and Eva Grümpel.

I would also like to extend my thanks to all my colleagues from Randau group for their support and all scientific discussions: Dr. Alvaro Orell, Dr. Vanessa Tripp, Dr. Michael Daume, Hanna Müller Esparza, Daniel Gleditzsch, Xiohan Guo, Dr. Jose Vicente Gomez Filho, Ruth Beuer and finally our labmama Julia Wiegel.

Special thanks should be given to Dr. Patrick Pausch and Prof. Dr. Gert Bange for their valuable collaboration on this project. Additionally, I wish to thank various people for their contribution to this project; Prof. Dr. Henning Urlaub, Andreas Linden, Dr. Karola Schüle, Timo Glatter and Lyle Kroell.

I would thank to Max Planck Society, IMPRS program and TÜBiTAK for funding.

Finally, I wish to thank my parents and Matthias Stöhr for their support and encouragement throughout my study.

Eidesstattliche Erklärung
Statutory Declaration

Ich versichere hiermit an Eides statt, dass meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist. Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt. Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

I hereby declare, that the thesis entitled

Characterization of the Type IV CRISPR-Cas system of *Aromatoleum aromaticum* EbN1

was written single-handed and without forbidden help. Also I declare, that no references or resources other than those which are explicitly stated, were used. This work, in the present or a similar version, was not handed in to another university and did not serve for any other examination purposes prior to this date.

Berlin, den 12. 09.2019

Ahsen Özcan