

Production and analysis of synthetic Cascade variants

Dissertation

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer.nat.)

des Fachbereichs Biologie der Philipps-Universität Marburg

Vorgelegt von

Daniel Gleditsch

Aus Halle (Saale)

Marburg, 2020

Die vorliegende Dissertation wurde von Februar 2016 bis November 2019 am Max-Planck-Institut für terrestrische Mikrobiologie in Marburg unter Leitung von Prof. Dr. Lennart Randau angefertigt.

Vom Fachbereich Biologie der Philipps-Universität Marburg (Hochschulkennziffer 1180) als Dissertation angenommen am _____

Erstgutachter(in): Prof. Dr. Lennart Randau

Zweitgutachter(in): Prof. Dr. Gert Bange

Tag der Disputation: _____

Erklärung

Ich versichere, dass ich meine Dissertation mit dem Titel „Production and analysis of synthetic Cascade variants“ selbstständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfsmittel bedient habe.

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

Marburg, den 13. Januar 2020

Daniel Gleditsch

Teile dieser Arbeit wurden in folgenden Artikeln veröffentlicht:

- **Gleditsch D, Randau L.** Means and Methods for the selective stabilization of RNA via Cas5 and Cas7 proteins, Patent issued Feb 7/2019, eu WO2019025444
- **Pausch P, Müller-Esparza H, Gleditsch D, et al.** Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. *Molecular Cell*. 2017, Aug 17;67(4):622-632.e4.
- **Gleditsch D, Müller-Esparza H, Pausch P, et al.** Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Research*. 2016, 2016;44(12):5872-5882

Weitere Veröffentlichungen:

- **Gleditsch D, Pausch P, Müller-Esparza H, et al.** PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biology*. 2018, Sep 18:1-14, doi: 10.1080/15476286.2018.1504546.

Contents

List of abbreviations	3
Summary	4
Zusammenfassung	5
1. Introduction	6
1.1 The CRISPR-Cas adaptive immune system	6
1.2 Classification of CRISPR-Cas systems	8
1.3 Interference mechanism of type I CRISPR-Cas systems	9
1.4 Variations in type I-F CRISPR-Cas systems	14
1.5 The minimal type I-Fv CRISPR-Cas system and its synthetic variants	16
2. Results	21
2.1 <i>In vitro</i> analysis of the minimal I-Fv CRISPR-Cas system	21
2.1.1 Optimized purification of I-Fv Cascade	21
2.1.2 <i>In vitro</i> analysis of target binding	22
2.1.3 3D Structure of small synthetic I-Fv Cascade	25
2.1.4 3D structure of I-Fv Cascade bound to target DNA	28
2.1.5 Requirement of AH and WL domains for complex formation	31
2.1.6 Investigation of the Cas3fv nuclease activity	32
2.2 Synthetic Cascade assembly and RNA wrapping	38
2.2.1 <i>In vitro</i> RNA wrapping with I-Fv Cas proteins	38
2.2.2 Directed <i>in vivo</i> RNA wrapping by I-Fv CRISPR-Cas repeat sequences	43
2.2.3 Additional applications of directed RNA wrapping	73
3. Discussion	82
3.1 <i>In vitro</i> analysis of type I-Fv Cascade	82
3.2 Investigation of the nuclease activity of Cas3fv and the Cas1-Cas2/3 super complex	86
3.3 Synthetic Cascade assembly and directed RNA-wrapping	90
3.4 Nature of filaments and rRNA contamination	94
3.5 Natural reasons for limitations in size of Cascade assembly	98
3.6 Applications	99
4. Material and Methods	105
4.1 Materials, instruments and source of supplies	105
4.1.1 Chemicals, Kits and enzymes	105
4.1.2 Instruments	107
4.1.3 Buffers and solutions	107
4.2 Strains and culture conditions	108
4.3 Plasmids and oligonucleotides	109
4.3.1 Plasmids	109
4.3.2 Oligonucleotides	110

4.4	Working with DNA	112
4.4.1	Preparation of plasmid DNA from <i>E. coli</i>	112
4.4.2	Sanger sequencing	112
4.4.3	Quantification of DNA	112
4.4.4	Electrophoresis of DNA	113
4.4.5	Purification of DNA	113
4.4.6	Polymerase chain reaction (PCR)	114
4.4.7	Modification of DNA	115
4.4.8	Transformation	116
4.4.9	Radioactive labeling	116
4.5	Working with RNA	117
4.5.1	Treatment of solutions, glassware and equipment	117
4.5.2	RNA extraction	117
4.5.3	Quantification of RNA	117
4.5.4	Electrophoresis of RNA	118
4.5.5	Northern Blotting	118
4.5.6	Illumina Sequencing	119
4.5.7	Nanopore sequencing	120
4.5.8	Mapping of sequencing reads	120
4.5.9	<i>In vitro</i> transcription	120
4.6	Biochemical Methods	121
4.6.1	Cell lysis	121
4.6.2	Affinity purification	121
4.6.3	Size-exclusion chromatography	121
4.6.4	Anion-exchange chromatography	122
4.6.5	Protein quantification by Bradford	122
4.6.6	Production and purification of recombinant proteins	122
4.6.7	<i>In vitro</i> RNA wrapping	124
4.6.8	RNA protection assays	124
4.6.9	Electrophoretic mobility shift assays (EMSA)	125
4.6.10	Nuclease assays	125
4.6.11	Crystallization and 3D structure analysis of I-Fv Cascade	126
4.6.12	Electron Microscopy	126
4.7	Cell biological methods	127
4.7.1	Fluorescence Microscopy	127
4.7.2	Fluorescence-activated cell sorting	127
5.	References	128
6.	Supplementary Material	136

List of abbreviations

aa	amino acid(s)	nt	nucleotides
Amp	ampicillin	rev	reverse
APS	ammonium persulfate	RBS	ribosome binding site
ATP	adenosine triphosphate	RNA	ribonucleic acid
bp	basepair(s)	RNase	ribonuclease
BSA	bovine serum albumin	RNA-Seq	high-throughput RNA sequencing
C-terminal	carboxy-terminal	rRNA	ribosomal RNA
Cam	chloramphenicol	NTP	nucleoside triphosphate
Cas	CRISPR-associated	OD ₆₀₀	optical density at 600 nm
Cascade	CRISPR-associated complex for antiviral defense	PAGE	polyacrylamide gel electrophoresis
cDNA	complementary DNA	PCR	polymerase chain reaction
cpm	counts per minute	pH	potential of hydrogen
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats	PIPES	piperazine-N,N'-bis(2-ethanesulfonic acid)
crRNA	CRISPR RNA	qRT-PCR	quantitative real-time PCR
crRNP	CRISPR ribonucleoprotein complex	RT	room temperature
Da	Dalton	s	second(s)
DAP	2,6-diaminopimelic acid	SEC	size-exclusion chromatography
DEPC	diethylpyrocarbonate	SDS	sodium dodecyl sulphate
DMSO	dimethyl sulfoxide	sfGFP	super folder green fluorescent protein
DNA	deoxyribonucleic acid	Spec	spectinomycin
dNTP	deoxyribonucleotide triphosphate	ssDNA	single-stranded DNA
dsDNA	double-stranded DNA	ssRNA	single-stranded RNA
DTT	dithiothreitol	TAE	Tris-acetate EDTA-buffer
e.g.	for example (exempli gratia)	TBE	Tris-borate EDTA-buffer
EDTA	ethylenediaminetetraacetic acid	TEMED	tetramethylethylene diamine
EMSA	electrophoretic mobility shift assay	Tris	tris (hydroxymethyl) aminomethane
EtBr	ethidium bromide	rpm	rounds per minute
<i>et al.</i>	and other (et alteri)	V	Volts
FACS	Fluorescence activated cell sorting	v	volume
FPLC	Fast Protein Liquid Chromatography	UV	ultraviolet
fwd	forward	W	Watt
g	gravitational acceleration	wt	wild-type
h	hour(s)	U	unit (enzyme activity)
HEPES	4-(2-hydroxyethyl)-1-piperazine- ethanesulfonic acid	%(v/v)	percent by volume
i.e.	that is (id est)	%(w/v)	percent by weight
IPTG	isopropyl β -D-1-thiogalactopyranoside	>	higher than
Kan	kanamycin	<	lower than
kb	kilo basepairs	Δ	deletion
kDa	kilo Dalton		
l	liter		
LB	lysogeny broth		
M	Molar (mol/l)		
m	meter		
min	minute(s)		
μ	micro (10 ⁻⁶)		
n	nano (10 ⁻⁹)		
Ni-NTA	nickel-nitrilotriacetic acid		
N-terminal	amino-terminal		

Summary

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR associated) is an adaptive immune system of Archaea and Bacteria. It is able to target and destroy foreign genetic material with ribonucleoprotein complexes consisting of CRISPR RNAs (crRNAs) and certain Cas proteins. CRISPR-Cas systems are classified in two major classes and multiple types, according to the involved Cas proteins. In type I systems, a ribonucleoprotein complex called Cascade (CRISPR associated complex for antiviral defence) scans for invading viral DNA during a recurring infection and binds the sequence complementary to the incorporated crRNA. After target recognition, the nuclease/helicase Cas3 is recruited and subsequently destroys the viral DNA in a step termed interference.

Multiple subtypes of type I exist that show differences in the Cascade composition. This work focuses on a minimal Cascade variant found in *Shewanella putrefaciens* CN-32. In comparison to the well-studied type I-E Cascade from *Escherichia coli*, this complex is missing two proteins usually required for target recognition, yet it is still able to provide immunity. Recombinant I-Fv Cascade was previously purified from *E. coli* and it was possible to modulate the complex by extending or shortening the backbone, resulting in synthetic variants with altered protein stoichiometry.

In the present study, I-Fv Cascade was further analyzed by *in vitro* methods. Target binding was observed and the 3D structure revealed structural variations that replace the missing subunits, potentially to evade viral anti-CRISPR proteins. The nuclease/helicase of this system, Cas2/3fv, is a fusion of the Cas3 protein with the interference-unrelated protein Cas2. A standalone Cas3fv was purified without the Cas2 domain and *in vitro* cleavage assays showed that Cas3fv degrades both free ssDNA as well as Cascade-bound substrates. The complete Cas2/3fv protein forms a complex with the protein Cas1 and was shown to reduce cleavage of free ssDNA, potentially as a regulatory mechanism against unspecific cleavage.

Furthermore, we established a process termed "RNA wrapping". Synthetic Cascade assemblies can be created by directing the general RNA-binding ability of the characteristic Cas7fv backbone protein on an RNA of choice such as reporter gene transcripts. Specific complex formation can be initiated *in vivo* by including a repeat sequence from the crRNA upstream a given target sequence and binding of the Cas5fv protein. The created complexes contain the initial 100 nt of the tagged RNA which can be isolated afterwards. While incorporated in complexes, RNA is stabilized and protected from degradation by RNases. Complex formation can be used to silence reporter gene transcripts. Furthermore, we provided initial indications that the backbone of synthetic complexes can be modified by addition of reporter proteins.

Zusammenfassung

CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR assoziiert) ist ein adaptives Immunsystem in Archaeen und Bakterien, das fremdes genetisches Material mit Hilfe von Ribonukleoprotein-Komplexen erkennt und zerstört. Diese Komplexe bestehen aus einer CRISPR RNA (crRNA) und Cas Proteinen. CRISPR-Cas Systeme sind in zwei Hauptklassen und mehrere Typen unterteilt, abhängig von den beteiligten Cas Proteinen. In Typ I Systemen sucht ein Komplex namens Cascade (CRISPR associated complex for antiviral defence) nach eingedrungener viraler DNA während einer Folgeinfektion und bindet die zu der eingebauten crRNA komplementäre Sequenz. Anschließend wird die Nuklease/Helikase Cas3 rekrutiert, welche die virale DNA degradiert (Interferenz).

Das Typ I System wird in mehrere Subtypen unterteilt, die Unterschiede im Aufbau von Cascade vorweisen. Im Fokus dieser Arbeit steht eine minimale Cascade-Variante aus *Shewanella putrefaciens* CN-32. Im Vergleich zur gut untersuchten Typ I-E Cascade aus *Escherichia coli* fehlen in diesem Komplex zwei Untereinheiten, die gewöhnlicher Weise für die Zielerkennung benötigt werden. Dennoch ist der Komplex aktiv. Rekombinante I-Fv Cascade wurde bereits aus *E. coli* aufgereinigt und es war möglich, den Komplex zu modifizieren, indem das Rückgrat entweder verlängert oder verkürzt wurde. Dadurch wurden synthetische Varianten mit veränderter Protein-Stöchiometrie erzeugt.

In der vorliegenden Arbeit wurde I-Fv Cascade weiter mit *in vitro* Methoden untersucht. So wurde die Bindung von Ziel-DNA beobachtet und die 3D Struktur zeigt, dass strukturelle Veränderungen im Komplex die fehlenden Untereinheiten ersetzen, möglicherweise um viralen Anti-CRISPR Proteinen zu entgehen. Die Nuklease/Helikase dieses Systems, Cas2/3fv, ist eine Fusion des Cas3 Proteins mit dem Interferenz-unabhängigen Protein Cas2. Ein unabhängiges Cas3fv ohne Cas2 Untereinheit wurde aufgereinigt und *in vitro* Assays zeigten, dass dieses Protein sowohl freie ssDNA als auch Cascade-gebundene Substrate degradiert. Das komplette Cas2/3fv Protein bildet einen Komplex mit dem Protein Cas1 und zeigt eine reduzierte Aktivität gegenüber freier ssDNA, möglicherweise als Regulationsmechanismus zur Vermeidung von unspezifischer Aktivität.

Weiterhin wurde ein Prozess namens „RNA wrapping“ etabliert. Synthetische Cascade-Komplexe wurden erzeugt, in denen die grundlegende RNA-Bindung des charakteristischen Cas7fv Rückgrat-Proteins auf eine ausgewählte RNA gelenkt wird. Diese spezifische Komplexbildung kann *in vivo* durch eine Repeat-Sequenz der crRNA stromaufwärts der Zielsequenz und durch Bindung des Cas5fv Proteins initiiert werden. Die erzeugten Komplexe beinhalten die ersten 100 nt der markierten RNA, die anschließend isoliert werden kann. Innerhalb der Komplexe ist die RNA stabilisiert und geschützt vor Degradation durch RNasen. Komplexbildung kann außerdem genutzt werden, um Reporter-Transkripte stillzulegen. Zusätzlich wurden erste Hinweise geliefert, dass das Rückgrat der synthetischen Komplexe durch Fusion mit weiteren Reporterproteinen modifiziert werden kann.

1. Introduction

1.1 The CRISPR-Cas adaptive immune system

Bacteria and Archaea are under constant attack by foreign genetic material. These mobile genetic elements (MGE) are most commonly introduced by viral infections but can also be transmitted by conjugation, transformation, transduction or transposition (Koonin & Dolja, 2013, Moreira & Lopez-Garcia, 2009). Consequently, Bacteria and Archaea have evolved numerous defence mechanisms that are either based on preventing the entry of invading DNA into the cell, inactivating foreign DNA after entry or induced cell death to protect the population (Koonin *et al.*, 2017). Most of these mechanisms are described as innate immune systems, such as restriction-modification-systems, abortive infection systems and the modification of virus receptors (Samson *et al.*, 2013).

Among the prokaryotic defence mechanisms, CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR associated) has been characterized as an adaptive immune system, capable of storing genetic information of previously encountered MGEs. This defence system utilizes short RNA molecules, called CRISPR RNAs (crRNAs), to degrade foreign DNA or RNA from invading viruses (Mojica *et al.*, 2005, Barrangou *et al.*, 2007). These crRNAs are stored in the CRISPR locus, which consists of arrays of unique sequences called spacers that are flanked by short palindromic repeat sequences. Additionally, a cluster of *cas* genes is usually located in close proximity (Makarova *et al.*, 2011). CRISPR-Cas systems are widely distributed and were found in 45 % of bacterial and in 84 % of archaeal genomes (Grissa *et al.*, 2007).

Fragments of viral genomes termed protospacers, can be inserted into the CRISPR locus in a process called adaptation. This process is carried out by the universal Cas proteins Cas1 and Cas2 and depends on a short sequence of 2-5 bp, the PAM sequence (protospacer adjacent motif). If this sequence is present and recognized, a complex of Cas1 and Cas2 binds and cleaves the neighbouring protospacer sequence to insert it as a new spacer in the extended CRISPR locus with the addition of an upstream repeat region (Nunez *et al.*, 2014).

The CRISPR locus is first transcribed into a long precursor crRNA (pre-crRNA), consisting of multiple spacer and repeat sequences. Subsequently, this transcript is processed into mature crRNAs by either endogenous RNase III or specific Cas proteins with an endoribonuclease function, depending on the type of CRISPR-Cas system. Mature crRNAs consist of one spacer sequence flanked by the remnants of the repeats and form a CRISPR ribonucleoprotein complex (crRNP) with Cas proteins. This complex can target foreign nucleic acids and mediate interference during a recurring infection. The crRNPs are able to distinguish between self and non-self DNA by recognition of the PAM sequence (Westra *et al.*, 2013). Target binding is followed by degradation of the foreign genetic material either by the crRNP itself or by

recruitment of an additional Cas protein with nuclease function (Barrangou *et al.*, 2007, Brouns *et al.*, 2008).

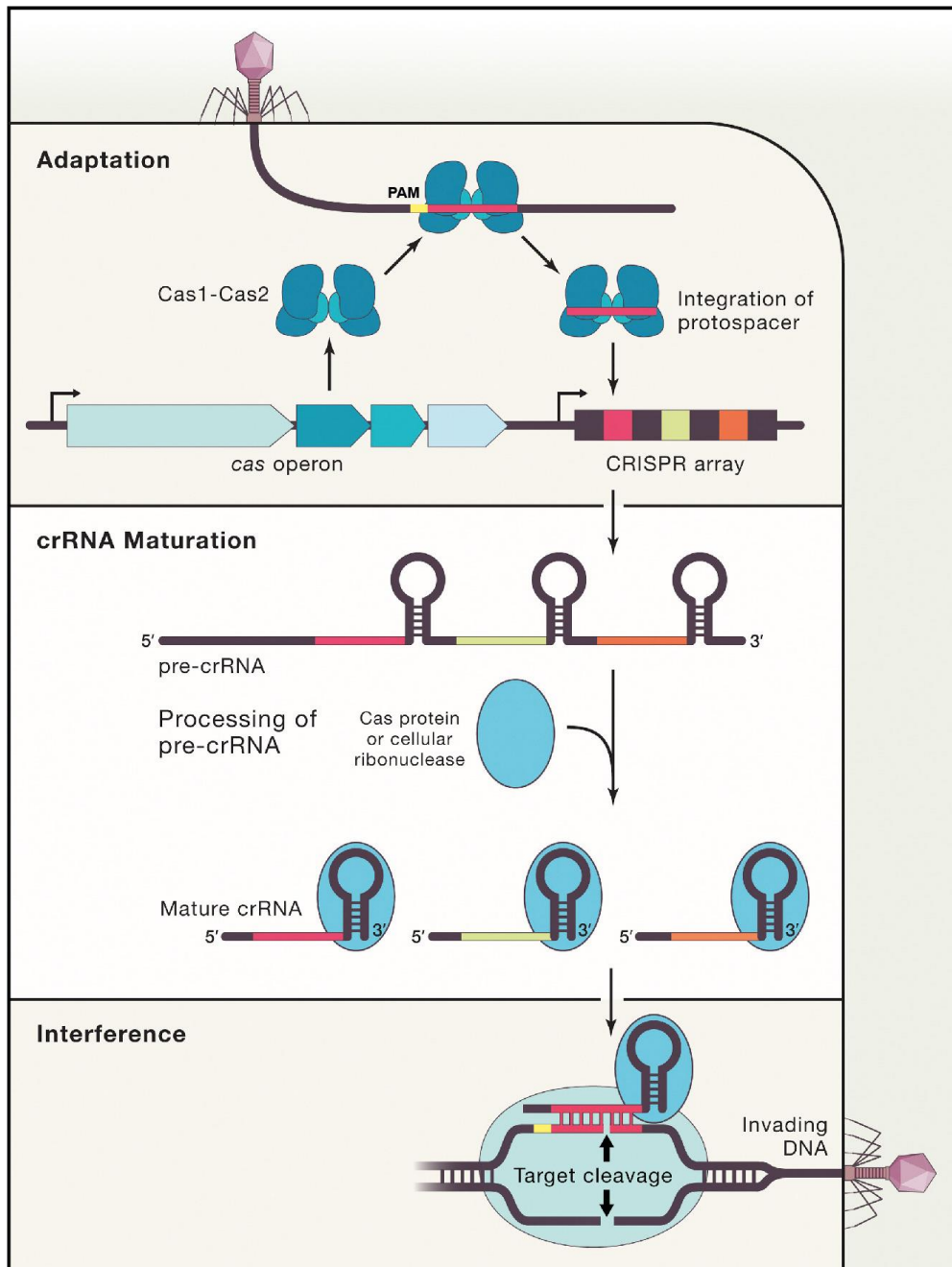


Figure 1.1: The three stages of CRISPR-Cas interference. During adaptation, the Cas1-Cas2 complex selects a part of the foreign DNA and integrates it into the host's CRISPR array. In the next stage (crRNA maturation), the CRISPR array is transcribed into a long pre-crRNA that is further processed by Cas proteins or, in some cases, by cellular RNases. In the interference stage, the mature crRNAs guide Cas nucleases to the cognate foreign DNA. The Cas proteins cleave the foreign nucleic acid upon binding of the crRNA to the matching target sequence. Interference and a daptation depend on recognition of a PAM sequence (in yellow) to distinguish self from non-self DNA. In class 1 systems, the interference machinery is a multi-Cas-protein complex, whereas class 2 systems utilize a single Cas protein for target cleavage. Figure modified from Hille *et al.*, 2018.

The arms race between viruses and prokaryotes promotes the evolution of viral counter-measures against CRISPR-Cas systems (Koonin & Dolja, 2013). Recently, small viral Anti-CRISPR (Acr) proteins were discovered that are able to inhibit CRISPR-Cas systems by blocking various positions in the effector complexes (Bondy-Denomy *et al.*, 2013, Pawluk *et al.*, 2014, Pawluk *et al.*, 2016). Alternatively, viruses are capable of mutating their PAM sequence to escape CRISPR-Cas interference which in turn is required to take up more spacers (Cady *et al.*, 2012).

1.2 Classification of CRISPR-Cas systems

Multiple types of CRISPR-Cas systems have been discovered, which are defined by the Cas proteins involved. To this date, two classes of CRISPR-Cas systems with multiple types and subtypes are described. The two classes are defined by a multisubunit protein complex (class 1) or a single protein (class 2) as effector units and are further separated into six main types with different signature Cas proteins responsible for target cleavage (Figure 1.2). Multiple subtypes exist in these types that have evolved different ways of crRNA processing, effector complex formation and PAM recognition (Koonin *et al.*, 2017).

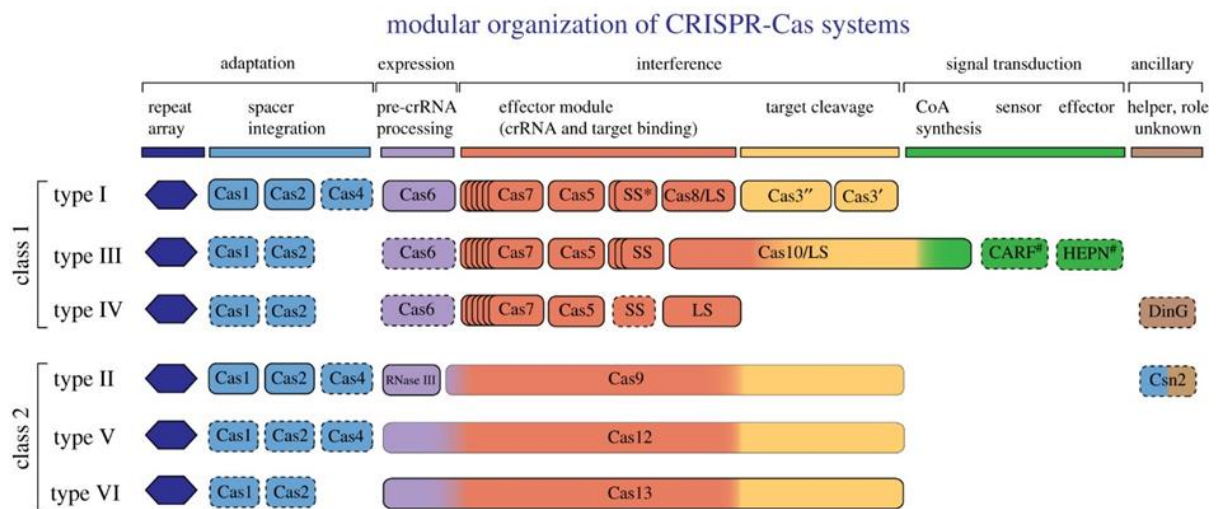


Figure 1.2: The two classes and six main types in CRISPR-Cas classification. CRISPR-Cas systems are classified based on the involved Cas proteins and their function. CRISPR arrays and the adaptation proteins Cas1 and Cas2 are conserved in all systems while in some types, the additional Cas4 protein is involved in trimming of protospacers. In class 1 systems, multiple Cas proteins (namely Cas6 for crRNA processing, Cas7, Cas5 and small (SS) and large subunits (LS)) are forming the effector complex on a crRNA for target binding. Target cleavage is performed by an additional nuclease Cas3 (sometimes split into two proteins or part of the large subunit in type III systems). In class 2 systems, a single Cas protein achieves interference. For the CRISPR-Cas9 system, endogenous RNase III is required for crRNA processing. In some systems, associated ancillary proteins have a helper function such as the DinG helicase or the Csn2 for coordination of adaptation and interference. Class 1 type III systems also encode proteins for signal transduction, further explained in the following section. Genes stated to be dispensable for immediate immunity are indicated by dashed outlines. Figure adapted from Koonin & Makarova, 2019.

The CRISPR-Cas9 system is the best characterized among class 2 systems and relies on the single effector protein Cas9 in combination with an additional trans-activating crRNA (tracrRNA) for target interference. In synthetic systems, tracrRNA and crRNA can be fused to a single-guide RNA (sgRNA) construct. Since its discovery, the CRISPR-Cas9 system has revolutionized genome-editing, CRISPR interference, and transcription regulation approaches (Jinek *et al.*, 2012, Ran *et al.*, 2013, Qi *et al.*, 2013, Larson *et al.*, 2013, Maeder *et al.*, 2013, Cheng *et al.*, 2013).

Due to the high diversity of CRISPR-Cas systems, novel effector Cas proteins are continuously evaluated for their applicability. Examples are class 2 systems other than type II, such as the type V system, defined by its signature protein Cas12 (Zetsche *et al.*, 2015) and the type VI system with the single-effector RNA-guided RNase Cas13 (Abudayyeh *et al.*, 2016, Smargon *et al.*, 2017).

In class 1 systems, DNA interference is achieved by complexes of multiple proteins (Koonin *et al.*, 2017). Type I systems utilize a crRNP called Cascade (Crispr associated complex for antiviral defence) (Brouns *et al.*, 2008) and a separate Cas3 helicase/nuclease protein for target degradation after recruitment by the complex (Huo *et al.*, 2014). This subtype will be discussed in more detail in the next section.

The effector complexes of type III systems are termed Csm or Cmr and are capable of targeting ssRNA in a transcription-coupled and PAM-independent manner, resulting in non-specific degradation of proximal DNA (Elmore *et al.*, 2016, Estrella *et al.*, 2016, Kazlauskienė *et al.*, 2016, Samai *et al.*, 2015). Interestingly, type III systems have also been shown to be involved in a cyclic oligoA (cOA) signalling pathway with allosteric regulation. Target binding of the effector complex stimulates the polymerase activity of the signature protein Cas10 for cOA synthesis. The produced cOA activates the promiscuous RNase activity of the Csm6 protein that indiscriminately degrades both target RNA and other random RNA molecules in proximity (Niewoehner *et al.*, 2017, Athukoralage *et al.*, 2018).

Much less is known about the type IV system, in which the absence of the usually conserved adaptation module and the apparent lack of an associated nuclease suggest novel functions of this system (Koonin & Makarova, 2019). Effector complex formation was shown by recombinant expression and purification in *E. coli* but its biological function remains uncertain (Ozcan *et al.*, 2019).

1.3 Interference mechanism of type I CRISPR-Cas systems

Type I systems are the most widespread in nature and while the general architecture of the Cascade effector complex is shared, its composition differs between the known eight subtypes (A-F, I-Fv, U) (Koonin *et al.*, 2017).

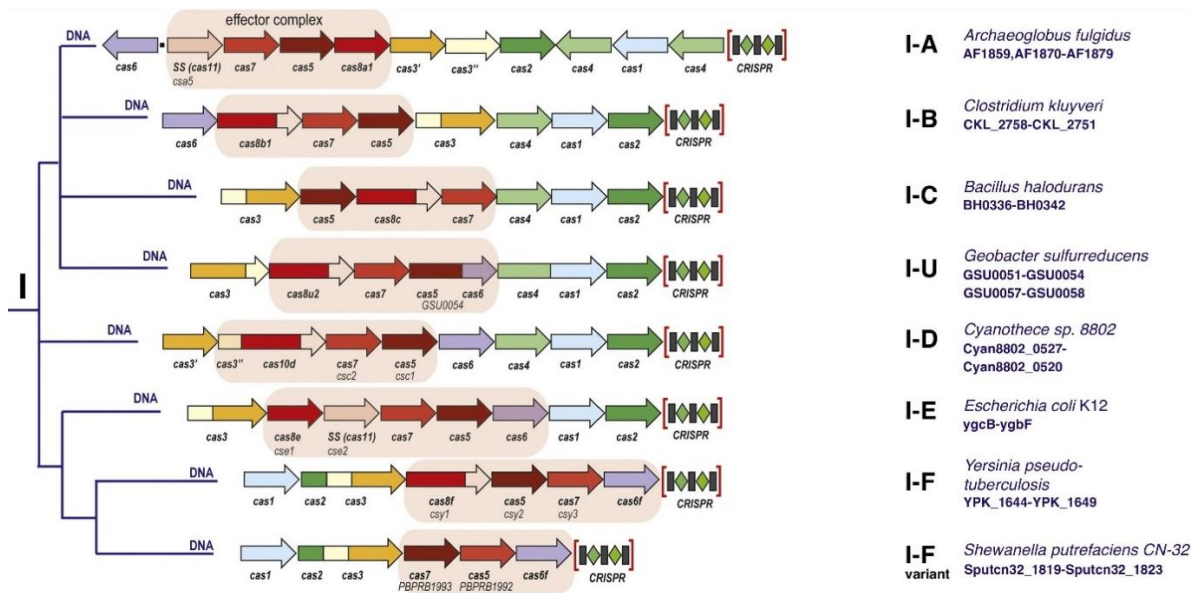


Figure 1.3: Schematic overview of type I CRISPR-Cas systems. *cas* genes encoding proteins forming the effector complex are marked, including *cas6*, *cas7*, *cas5* as well as genes encoding small (SS) and large subunits (LS). Adaptation proteins Cas 1 and Cas 2 are conserved in all types. Type I systems are defined by the signature protein Cas 3, an additional helicase/nuclease that is recruited for target degradation. Helicase and nuclease domains of *cas3* are split into multiple genes and/or fused to other *cas* genes in some subtypes. Figure adapted from Koonin *et al.*, 2017.

The earliest characterized subtype is the type I-E system from *E. coli* with multiple available crystal structures of the Cascade complex (Jackson *et al.*, 2014, Mulepati *et al.*, 2014, Zhao *et al.*, 2014). Type I-E Cascade consists of a 61 nt mature crRNA and five Cas proteins with an uneven stoichiometry ((Cse1)₁-(Cse2)₂-(Cas5)₁-(Cas7)₆-(Cas6)₁) and is described to have a “seahorse-like” shape with a mass of 405 kDa (Jore *et al.*, 2011). The mature crRNA consists of a 32 nt long spacer sequence flanked by an 8 nt long handle-region and a 21 nt long hairpin-region at the 5'- and 3'-end, respectively. These regions are generated from the repeats during crRNA maturation by Cas6. The endonuclease Cas6 stays tightly associated with the 3'-hairpin after processing while another Cas protein, Cas5, binds to the 5'-handle (Jore *et al.*, 2011, Carte *et al.*, 2010). The backbone of the structure is formed by the addition of multiple subunits of the protein Cas7, which binds in increments of 6 nt along the spacer sequence. The intertwined interaction of the Cas proteins in this backbone is due to distinct domains termed “fingers”, “palm” and “thumb” in an overall right-hand analogy. The palm contains a modified RNA recognition motif (RRM) plus two small loops and is responsible for crRNA binding, splaying out of every sixth base. Various interactions of the thumb domains with finger and palm domains of adjacent subunits connect all Cas proteins of the backbone (Mulepati *et al.*, 2014). The full complex is formed by the two additional proteins Cse1 (also known as the large subunit) and Cse2 (or small subunit). The former is responsible for PAM recognition and recruitment of the target nuclease Cas3, while the latter one forms a dimer that stabilizes the non-target strand of the foreign DNA during target binding (Jore *et al.*, 2011, Sashital *et al.*, 2012).

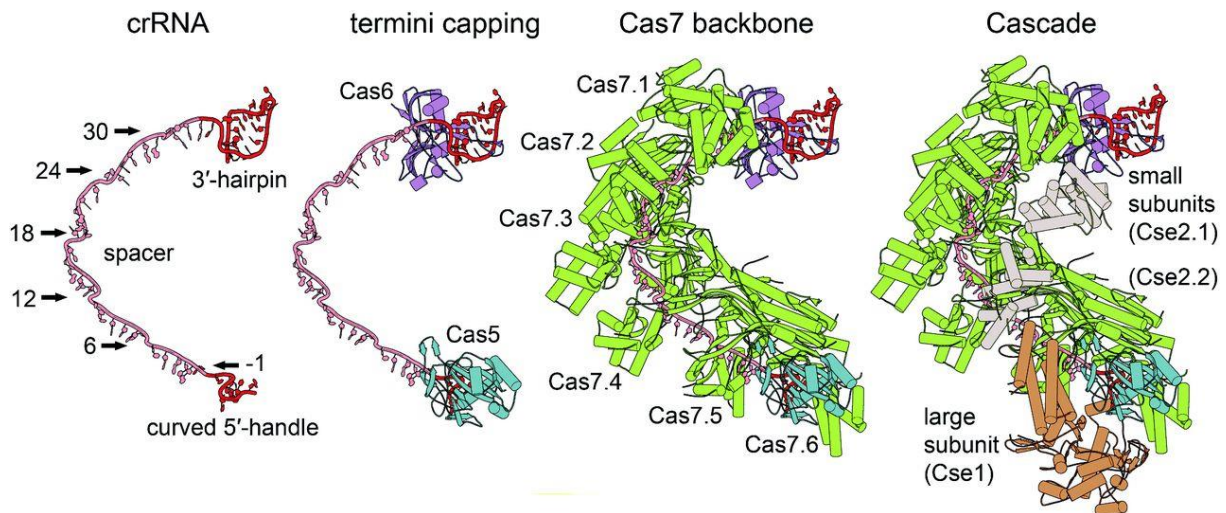


Figure 1.4: Assembly and structure of type I-E Cascade from *E. coli*. Type I-E Cascade consists of 11 protein subunits binding alongside the mature crRNA. Cas6 binds the 3'-hairpin after crRNA-processing while Cas5 binds the 5'-handle. Six subunits of Cas7 bind the spacer sequence in increments of 6 nt starting with position -1 in the 5'-handle as indicated by arrows. The large subunit and a dimer of the small subunit bind alongside the "belly" of the structure, stabilizing it (Jackson *et al.*, 2014, Zhao *et al.*, 2014, Mulepati *et al.*, 2014). PDB: 4TVX, Figure from Plagens *et al.*, 2015.

PAM recognition by the large subunit leads to destabilization of the DNA duplex and crRNA-directed strand invasion (Sashital *et al.*, 2012, Tay *et al.*, 2015). The complex forms a crRNA:target hybrid with a ribbon-like structure, termed R-loop. The displaced non-target strand is stabilized and guided along a groove at the belly of the complex by the large and small subunits upon structural rearrangements of these proteins (Mulepati *et al.*, 2014, Tay *et al.*, 2015).

Notable variants in this general Cascade composition exist that include less subunits, such as the minimal type I-C system in which the Cas5 protein also functions as an endoribonuclease for crRNA processing (Hochstrasser *et al.*, 2016) or the I-F system in which small subunits are compensated by structural variations in Cas7 (Cady *et al.*, 2012, Rollins *et al.*, 2015). The type I-F system will be further discussed in a later section.

R-loop formation is accompanied by further structural rearrangements that enable the recruitment of the signature protein Cas3 for target degradation (Hochstrasser *et al.*, 2014, Sinkunas *et al.*, 2011, Westra *et al.*, 2012). Type I-E Cas3 has been extensively studied with multiple available 3D structures in different states. Cas3 consist of an N-terminal metal-dependent histidine-aspartate (HD) nuclease domain, a C-terminal superfamily 2 helicase domain and a C-terminal accessory domain (CTD) (Gong *et al.*, 2014, Huo *et al.*, 2014, Jackson *et al.*, 2014, Mulepati & Bailey, 2013, Sinkunas *et al.*, 2011).

The helicase domain contains highly conserved residues of superfamily 2 (SF2) helicases including NTP-binding Walker A and B motifs (Makarova *et al.*, 2002, Jansen *et al.*, 2002). These domains consist of a tandem RecA-like fold, forming a channel with coordinated amino acids responsible for the binding of

NTP, divalent metal cations and nucleic acid substrates (Cordin *et al.*, 2006, Fairman-Williams *et al.*, 2010). The helicase enables the ATP-dependent unwinding of duplex DNA in 3'-5' direction (Mulepati & Bailey, 2013, Sinkunas *et al.*, 2011).

The HD nuclease is characterized as an exo- and endonuclease in the presence of divalent cations which are coordinated by the active site HD motif (Beloglazova *et al.*, 2011, Mulepati & Bailey, 2013, Sinkunas *et al.*, 2011). In some subtypes, the HD nuclease domain is separated from the helicase in an extra gene or fused to another *cas* gene (types I-A, I-B and I-D) (Koonin *et al.*, 2017). The ssDNase activity of Cas3 is commonly observed in the presence of a broad range of divalent cations (Mulepati & Bailey, 2011, Sinkunas *et al.*, 2011, Gong *et al.*, 2014, Huo *et al.*, 2014).

The CTD contacts both RecA-like domains, forming a closed channel for ssDNA and is suggested to be involved in loading of the helicase (Huo *et al.*, 2014, Gong *et al.*, 2014). The CTD domain is also suggested to connect Cas3 and Cascade (Gong *et al.*, 2014, Huo *et al.*, 2014). In fact, deletion of the CTD domain showed decreased Cascade binding affinity (Huo *et al.*, 2014).

Thermobifida fusca I-E Cas3

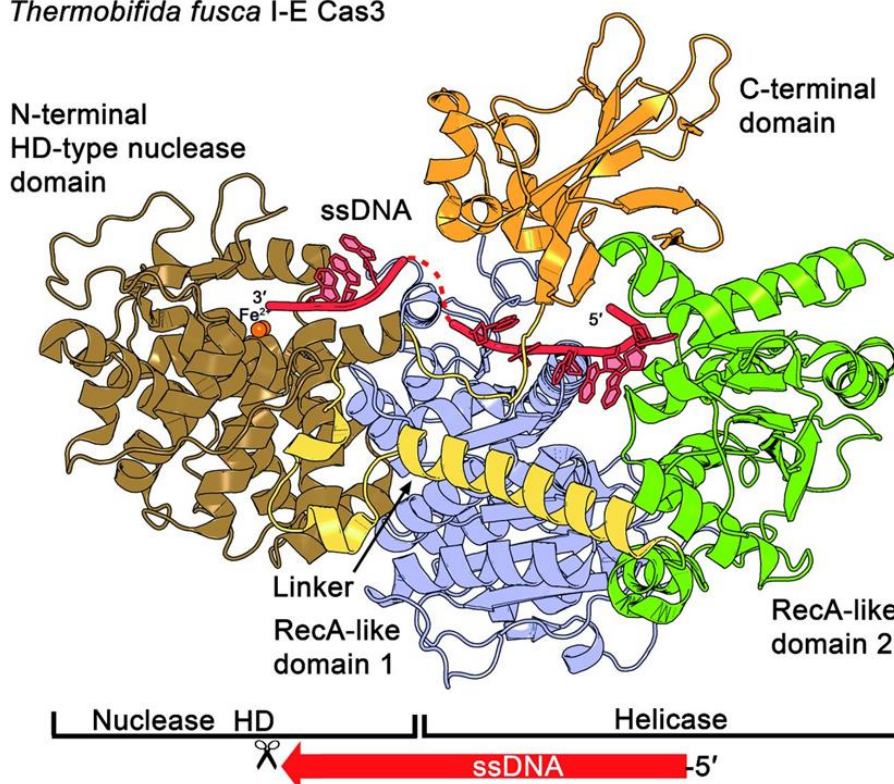


Figure 1.5: Structure of the DNA nuclease/helicase Cas3. The type I-E Cas3 crystal structure of *Th. fusca* (PDB: 4QQW) contains two tandem RecA-like domains, one HD-type nuclease domain and a CTD on top of the structure. The helicase core, consisting of the two RecA-like domains, forms a cleft that locates the residues for the binding of NTP, Mg^{2+} ions and the ssDNA substrate. Two Fe(II) ions are present at the catalytic centre's HD motif in this structure. The 5' end of the ssDNA enters Cas3 from the RecA2 side and is further threaded to RecA1 and the HD-type nuclease domain (indicated by a scissor). The CTD is proposed to close the ssDNA channel and to contact the Cascade complex. Figure from: Plagens *et al.*, 2015.

Cas3 specifically recognizes the fully formed Cascade/R-loop complex instead of partially bound substrates to avoid mistargeting and partial cleavage (Xiao *et al.*, 2017). Target degradation starts with nicking of the displaced non-target strand in the R-loop by the nuclease domain, followed by ATP-dependent unwinding of the remaining dsDNA in 3'-5' direction by the helicase domain of Cas3 and processive degradation of the produced ssDNA. Recent cryo-EM structures have captured the Cascade/R-loop/Cas3 complex in pre- and post-nicking states. The nuclease domain recruits the non-target strand at a flexible bulged region for nicking of single-stranded DNA, bypassing the helicase domain (Xiao *et al.*, 2018). Single-molecule fluorescence analysis has been used to describe the helicase unwinding as a repetitive DNA-reeling mechanism due to the reeling of the target DNA 3 bp at a time, underlined by three successive unwinding events of individual nucleotides (Loeff *et al.*, 2018).

Viruses are able to mutate the PAM or protospacer sequence to escape CRISPR-Cas immunity (Deveau *et al.*, 2008, Vercoe *et al.*, 2013). Cas3 cleavage generates products close to spacer length and enriched for PAM-like sequences that are suitable for integration into the CRISPR-locus as new spacers. This interference-driven adaptation, also called primed adaptation allows the host to quickly restore immunity against viral escape mutants (Kunne *et al.*, 2016, Fineran *et al.*, 2014). Remaining ssDNA can also be degraded by a standalone Cas3 (Mulepati & Bailey, 2013, Sinkunas *et al.*, 2013).

Overall, the type I interference mechanism consists of the following steps: (1) Cascade assembly, (2) target screening and R-loop formation, (3) Cas3 recruitment and (4) target DNA cleavage (Figure 1.6).

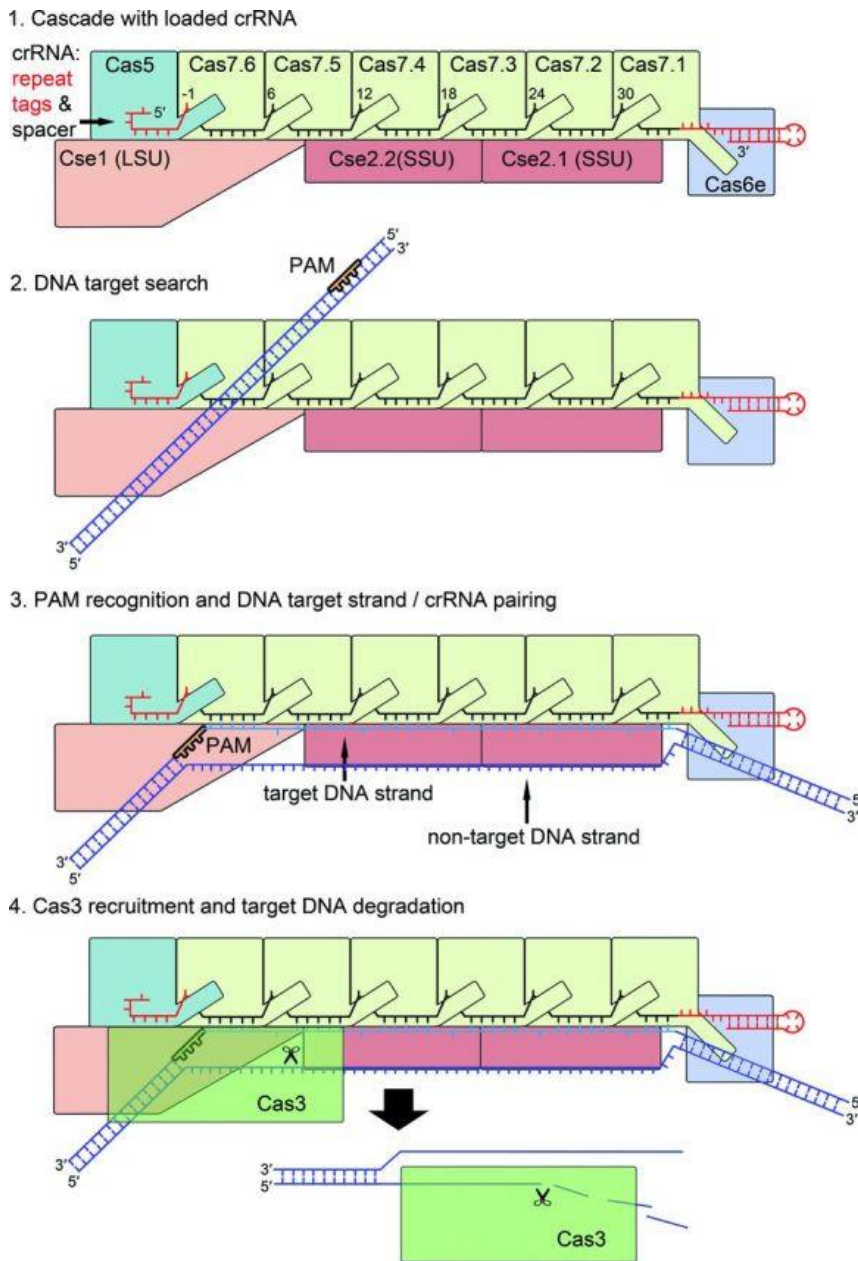


Figure 1.6: Mechanism of type I Cascade-mediated DNA interference. After the assembly of the crRNA-loaded Cascade, the surveillance complex (SSU: small subunits, LSU large subunit) scans DNA sequences. Potential DNA targets are identified via PAM recognition. This event triggers the destabilization of the DNA duplex and allows for the crRNA to pair with the target strand, while the non-target strand is displaced and spanned via the large and small subunit. Following R-loop formation, interaction sites at the base of the large subunit enable a stable interaction with Cas3. The HD domain of Cas3 nicks the DNA strand downstream of the PAM and the duplex is further unwound in 3'-5' direction and degraded. The remaining single-stranded target DNA can be cleaved by the stand-alone Cas3 enzyme. Figure modified: (Plagens *et al.*, 2015).

1.4 Variations in type I-F CRISPR-Cas systems

Another closely studied Cascade belongs to the type I-F system (*e.g.* present in *Pseudomonas aeruginosa*) which also targets foreign DNA in a PAM-dependent manner. In contrast to type I-E

Cascade, this complex is missing the small subunits and thus consists of only four proteins (Cady *et al.*, 2012, Rollins *et al.*, 2015).

The large subunit, here termed Cas8f, is responsible for recognition of a GG PAM sequence by specific amino acid interactions from the minor groove of the DNA. Opening of the dsDNA at this position is achieved by employing a “lysine wedge” of Cas8f, leading to hybridization of crRNA and target strand. The non-target strand of the opened dsDNA is stabilized with the help of Cas7f instead of small subunits, presumably by additional loops, termed “extended web”, forming a prominent positively charged channel (Chowdhury *et al.*, 2017, Rollins *et al.*, 2019, Guo *et al.*, 2017). Cas8f adopts a conformational change during full R-loop/Cascade complex formation and rotates by 180°, which exposes the recruitment site for the Cas3 protein (Rollins *et al.*, 2019).

Additionally, the Cas3 nuclease/helicase is fused to the adaptation protein Cas2 in this system. This Cas2/3 fusion was shown to form a complex with the Cas1 adaptation protein, resulting in a supercomplex that is capable of integrating new spacers during adaptation but inhibits Cas3 nuclease activity unless it is recruited by a target-bound Cascade complex (Figure 1.7) (Rollins *et al.*, 2017, Fagerlund *et al.*, 2017). In contrast to type I-E Cas3, Cas3f degrades both strands of the target DNA efficiently instead of primarily the non-target strand (Rollins *et al.*, 2017). This leads to spacer uptake from both foreign DNA strands in type I-F systems (Vorontsova *et al.*, 2015, Richter *et al.*, 2014, Staals *et al.*, 2016).

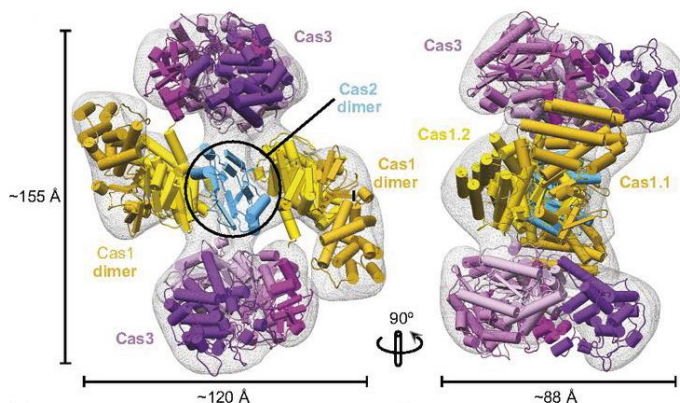


Figure 1.7: 3D structure of the "propeller-shaped" Cas1-Cas2/3 supercomplex from type I-F. Negative stain EM reconstruction of Cas 1–2/3 complex (EMD 8558). A pseudoatomic model was generated by docking crystal structures of Cas 1 (PDB ID code 3GOD) and Cas2/3 (PDB ID code 5B7I) into the EM density using Chimera (CC = 0.9). Figure modified from Rollins *et al.*, 2017.

14 Acr proteins are known to inhibit interference in type I-F systems while the mechanism of counter-defence has been discovered for three of them (AcrF1-3). AcrF1 and AcrF2 inhibit DNA recognition by

interacting with Cas7f and Cas8f (Chowdhury *et al.*, 2017, Guo *et al.*, 2017), while AcrF3 blocks Cas2/3 recruitment by mimicking a domain of Cas8f (Rollins *et al.*, 2019).

1.5 The minimal type I-Fv CRISPR-Cas system and its synthetic variants

Recent work in our group has studied a minimal variant of the type I-F system from *Shewanella putrefaciens* CN-32, in which not only the small but also the large subunit is missing. Furthermore, Cas7 and Cas5 are replaced by two new proteins which display no sequence similarity to other proteins of the type I systems and were initially uncharacterized. These proteins, now termed Cas7fv and Cas5fv, were confirmed to be the functional homologs of Cas7 and Cas5 in this variant system (Dwarakanath *et al.*, 2015) (Figure 1.8).

The Cas3 protein of the I-F variant system is fused to the Cas2 adaptation protein and bioinformatical predictions have identified an HD nuclease and a helicase domain (Dwarakanath, 2015). However, the lack of sequence similarity between the Cas3fv and its I-F counterpart suggests structural variations. *In vivo* assays have shown that the I-Fv system is active and confers interference (Dwarakanath *et al.*, 2015, Gleditzsch *et al.*, 2016). However, it was unclear how this system is able to substitute the missing large and small subunits for target recognition and how Cas2/3fv is recruited and achieves interference.

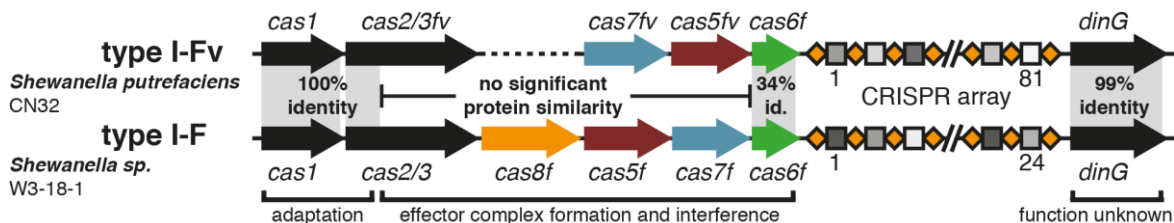


Figure 1.8: Schematic comparison of the type I-Fv and type I-F CRISPR-Cas systems from *Shewanella*. Arrows indicate the respective *cas* genes with the Cascade genes *cas5*, *cas7*, *cas6f* and *cas8f* colored in dark red, blue, green and orange, respectively. The CRISPR arrays are indicated as alternating rhombi (orange, crDNA repeat) and squares (grey, crDNA spacer). Grey emphasized areas indicate sections of high protein sequence similarity. No significant protein sequence similarity is found between the components, which are required for effector formation and Cascade mediated interference, except for Cas6f. Figure from Pausch *et al.*, 2017.

Previous work in our group has shown that only Cas7fv, Cas5fv and Cas6f are required to form a complex with mature crRNA (Dwarakanath *et al.*, 2015). The minimal recombinant Cascade complex can be produced in *E. coli* and purified (Figure 1.9 A). Initial structural analyses by transmission electron microscopy (TEM) showed that the complex adopts a similar crescent shape compared to other related Cascade complexes but with a more open configuration, likely due to the absence of additional subunits (Figure 1.9 B right). Unique filamentous structures of Cas7fv with a length of multiple hundred nm are consistently purified as byproducts (Figure 1.9 B left).

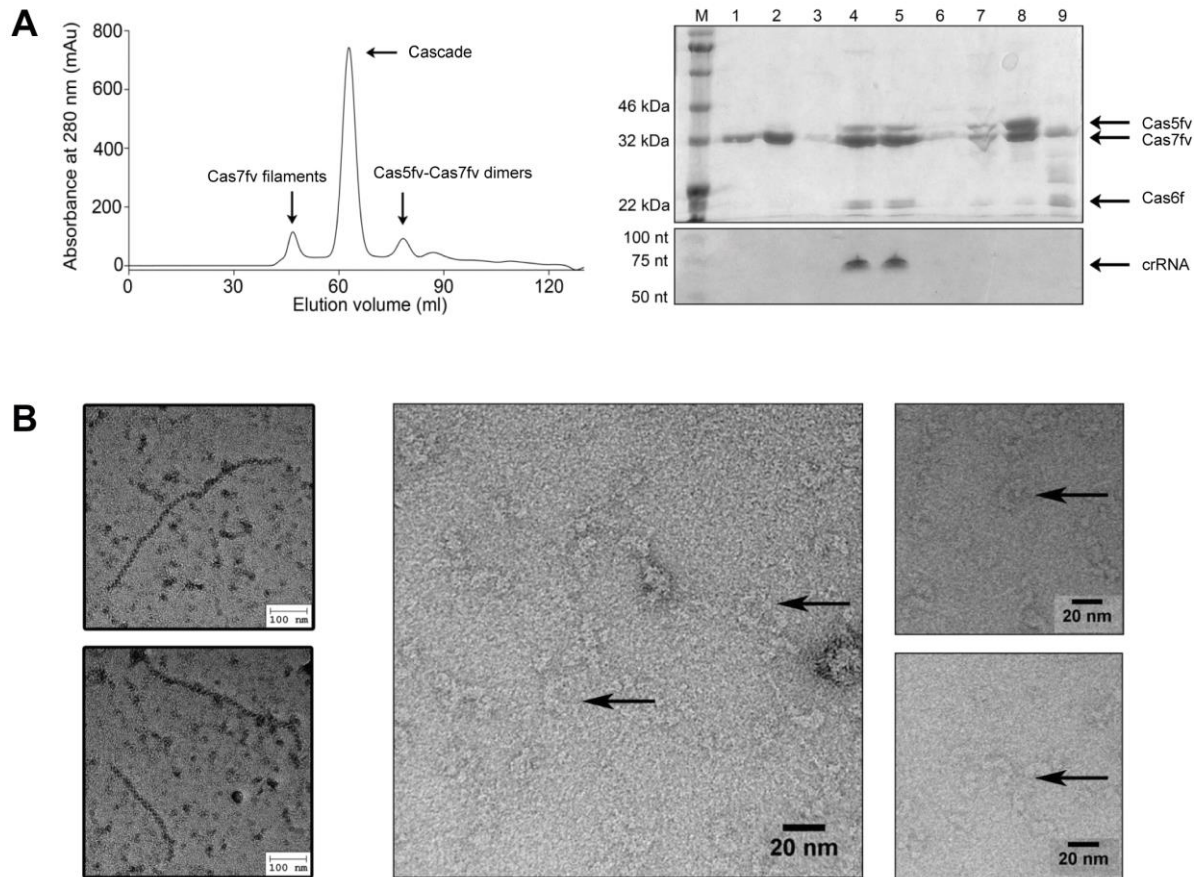


Figure 1.9: Purification of type I-Fv Cascade and Cas7fv filaments in *E. coli*. (A) Purification of recombinant type I-F variant Cascade complex. SDS-PAGE (top right) and 8M urea PAGE (bottom right) were used to separate the protein and RNA content of the peak fractions from size-exclusion chromatography (left). His-tagged Cas7fv co-eluted with Cas5fv, Cas6f and mature crRNA (fractions 4 and 5) which verified Cascade complex formation. Cas7fv filaments were observed in the void volume (fractions 1 and 2) and dimers of Cas7fv and Cas5fv (fraction 8) were identified. (B) TEM analysis verified filamentous structures of Cas7fv (left) and the crescent-shape of I-Fv Cascade (middle and right). Figure modified from Dwarakanath *et al.*, 2015 and Dwarakanath, 2015.

During my master thesis preceding this work, we were able to show that the backbone of I-Fv Cascade can be altered by co-producing the Cas proteins with crRNAs featuring elongated or truncated spacer sequences. Subsequent purification by size-exclusion yielded stable synthetic complexes with altered mass due to the binding of more or fewer subunits of the backbone-forming protein Cas7fv along the altered spacer sequence (Figure 1.10) (Gleditsch *et al.*, 2016).

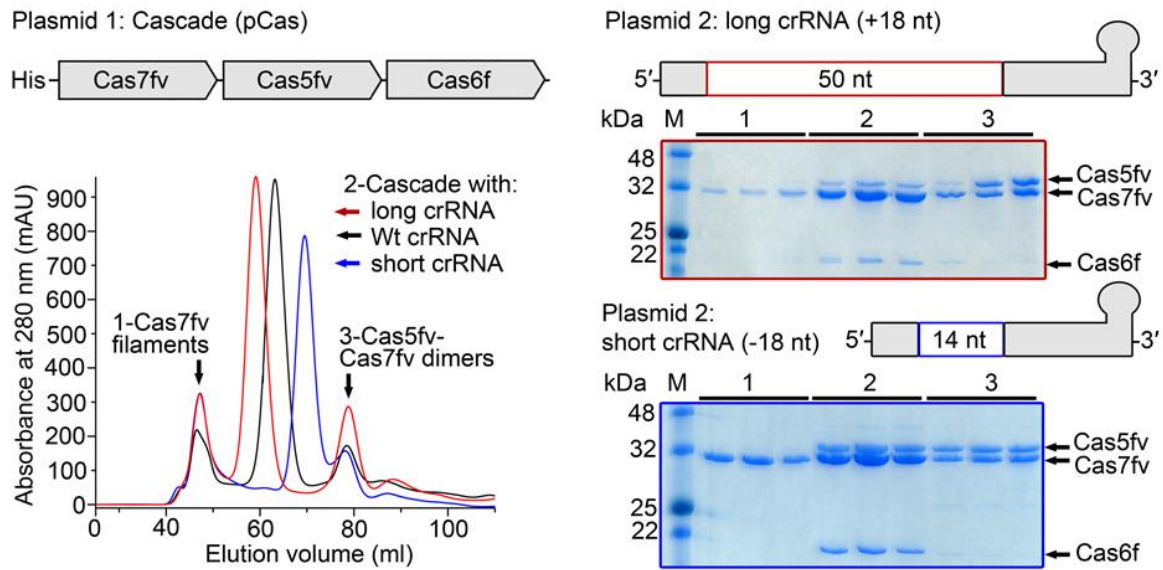


Figure 1.10: Recombinant production and purification of wild-type and synthetic Cascade variants. Variants of crRNAs with wild type (WT) spacer (32 nt), short spacer length (14 nt) and long spacer length (50 nt) were designed and co-produced with the Cas proteins in *E. coli*. Recombinant Cascade complexes were purified via Ni-NTA and size-exclusion chromatography. Cas7fv filaments (peak 1) and Cas5fv-Cas7fv dimers (peak 3) were observed and the middle peak corresponded to fully assembled Cascade ribonucleoproteins (peak 2). The relative shift of this peak during identical size-exclusion chromatography runs and SDS-PAGE revealed that additional spacer nucleotides result in additional Cas7fv subunits whereas a shorter crRNA results in fewer Cas7fv subunits in the Cascade complex. Figure from Gleditsch *et al.*, 2016.

The structure of these synthetic Cascade variants was analyzed by small-angle X-ray scattering (SAXS) which revealed that they retain the characteristic crescent shape but differ in terms of flexibility (Figure 1.11). While the short Cascade variant with a small spacer was contracted and less flexible, the long Cascade with an elongated spacer exhibited more flexibility than the wild-type as indicated by the random coil likeliness, which is a measure for the degree of freedom of the different proteins within the complex (Figure 1.11 A).

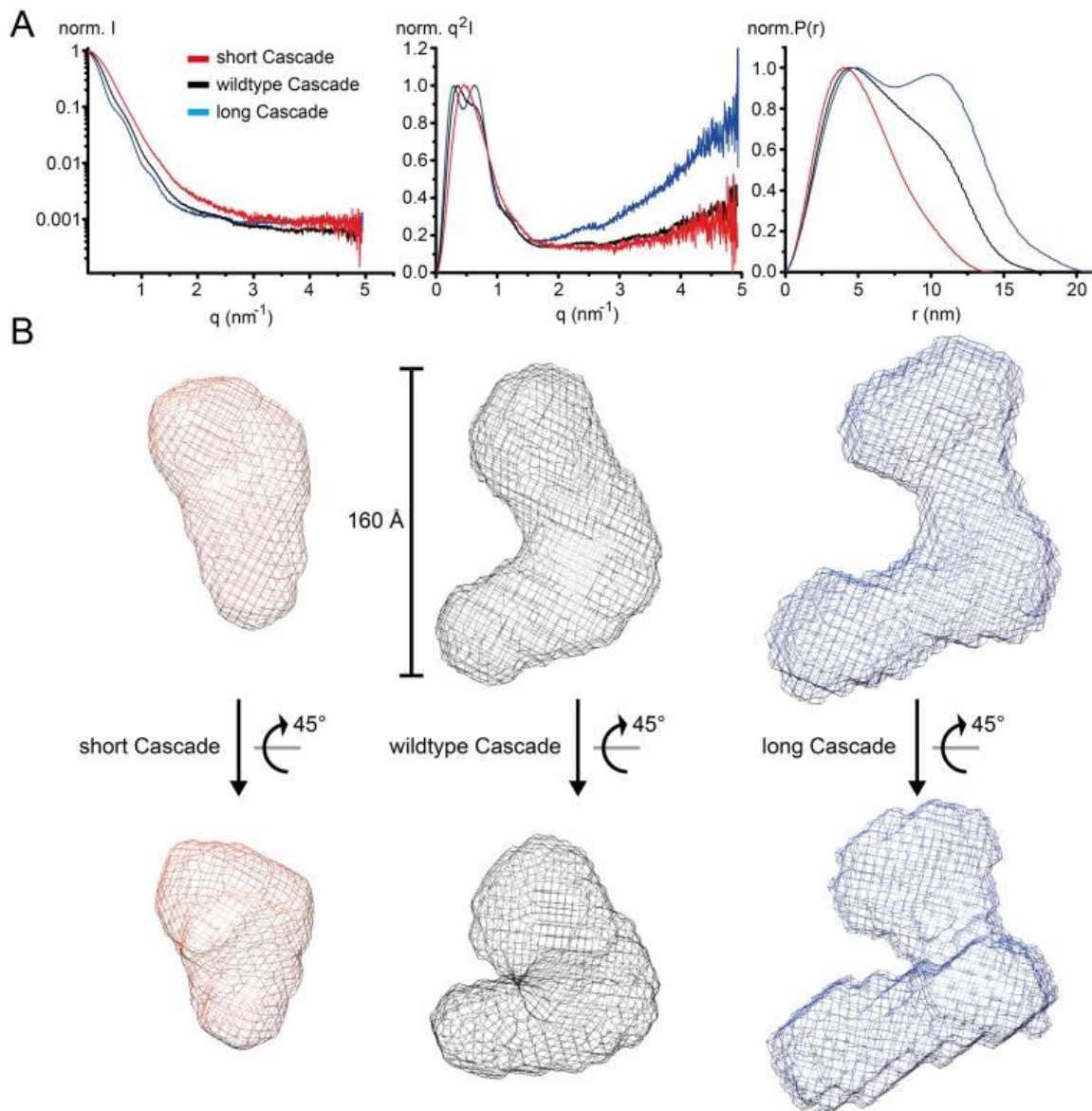


Figure 1.11: Small-angle X-ray scattering (SAXS) analyses of synthetic Cascade assemblies. (A) SAXS of the short (red), WT (black) and long (blue) Cascade constructs. Left: scattering curve, normalized to max I ; middle: Kratky plot, normalized to max q^2I , illustrating the 'random coil likeness' differences in the high q -range; right: $P(r)$ distance distribution curve, normalized to max $P(r)$, highlighting the different domain organization. (B) Surface grid representations of a veraged and filtered *ab initio* bead models calculated by Dammif (Konarev *et al.*, 2003). Figure from Gleditzsch *et al.*, 2016.

Overall, these experiments gave insight into the general structure of the minimal I-Fv Cascade. Wild-type and synthetic variants of I-Fv Cascade retain the typical crescent-shape known from other Cascade structures but exhibit increased flexibility due to the absence of additional subunits. The minimal complex backbone consisting of various Cas7fv subunits can be easily modified by increasing or decreasing the length of the spacer sequence in the provided crRNA. The formation of filament structures appears to be a consequence of this flexibility and Cas7fv filaments are consistently obtained as byproducts during Cascade purification. It was assumed that these structures originate from the

inherent RNA-binding ability of Cas7fv that enables the CRISPR-Cas system to work with varying spacer sequences in the first place.

In continuation of the described research on the type I-F variant systems, this thesis aims to follow two objectives. The first objective is to further analyze the recombinant minimal type I-Fv CRISPR-Cas system *in vitro*. Specifically, it is of interest (i) how I-Fv Cascade replaces the large subunit (and small subunits) present in other subtypes, (ii) how the PAM sequence is recognized and (iii) how target DNA is bound. *In vitro* binding assays will be performed to study the target binding ability of the complex in the presence or absence of a correct PAM.

We will also attempt to crystallize the recombinant I-Fv Cascade to obtain the 3D structure and elucidate the function of the novel Cas proteins in this effector complex. To study the novel Cas3fv protein on R-loop substrates and free DNA, it is required to optimize the previously used purification protocols for this protein. Purified Cas3fv will be used to study the cleavage mechanism of this novel protein. Additionally, we will investigate if the fusion of Cas3fv to Cas2 in the I-Fv system leads to the creation of a similar supercomplex known from type I-F and if this influences DNA cleavage.

The second part of this work focuses on the further synthetic modulation of I-Fv Cascade. In addition to the relative backbone-flexibility that has been shown by producing synthetic Cascade complexes with longer crRNA, the observation of filament structures on seemingly random RNA has led to the theory that the Cascade backbone can be formed on even longer RNA molecules, creating filaments in the process. We wondered if this assembly can be specifically directed on RNA and investigated this by placing a repeat region for initiation of backbone formation upstream of desired reporter gene sequences. It is anticipated that Cas7fv filament formation can be utilized for specific gene silencing and target RNA stabilization.

2. Results

2.1 *In vitro* analysis of the minimal I-Fv CRISPR-Cas system

The first part of this work focused on the detailed analysis of recombinant I-Fv CRISPR-Cas interference complexes *in vitro*. The initial purification and analysis of the I-Fv Cascade complex allowed for the characterization of the Cas5fv and Cas7fv proteins as functional homologs of Cas5 and Cas7 (Dwarakanath *et al.*, 2015). Further experiments were required to investigate how this minimal complex provides PAM-dependent interference without large and small subunits. Additionally, the novel Cas3fv nuclease is investigated because it shares no significant sequence similarity with the related Cas3 proteins of other systems and might provide interference by a different mechanism. The only similarity to the I-F system hereby is the fusion of Cas3 nuclease to the conserved adaptation protein Cas2.

2.1.1 Optimized purification of I-Fv Cascade

For *in vitro* analyses and later crystallization attempts, large amounts of pure Cascade were required, which necessitated optimization of the purification protocols. Therefore, we first switched the His-tag from the Cas6f protein and fused it to the C-terminus of the Cas5fv protein. The purification of wild-type I-Fv Cascade with this construct yielded fewer by-products than previous variants with the His-tag on the Cas7fv or Cas6f, especially in aggregated form, and Cascade eluted as a single distinct peak at an elution volume corresponding to the correct molecular weight of the complex during size-exclusion chromatography (Figure 2.1). This fraction containing Cascade was then used for all further experiments.

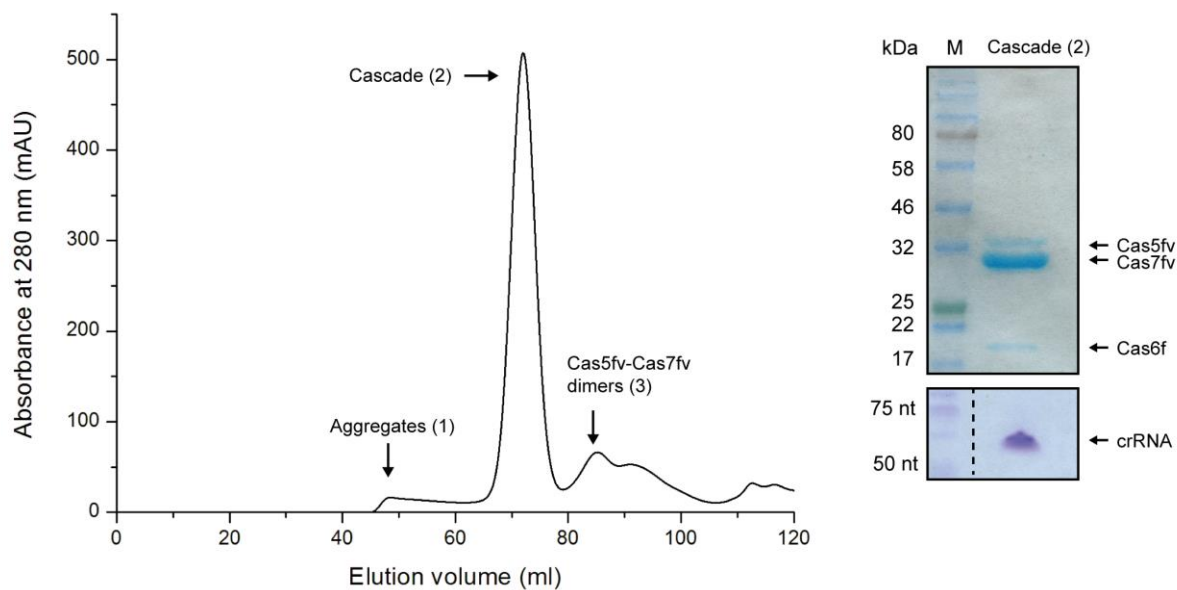


Figure 2.1: Purification of recombinant type I-Fv Cascade interacting with His-Cas5fv. UV chromatogram of size-exclusion purification of I-Fv Cascade with a His-tag on Cas5fv (left). The Cascade components Cas 7fv, Cas5fv, Cas6f as well as mature crRNA eluted as a single peak (peak 2). Dimers of Cas7fv and Cas5fv (peak 3) were observed as well as minimal amounts of aggregates of Cas7fv in the void volume (peak 1). SDS-PAGE (top right) confirmed the protein content and 8 M Urea -PAGE with toluidine blue staining (bottom right) was used to confirm the presence of a wild-type crRNA in the Cascade peak. For by-products see Figure 1.9 in the introduction.

2.1.2 *In vitro* analysis of target binding

Initial purification and characterization of the I-Fv Cascade complex already provided hints at the function of the novel Cas5fv and Cas7fv proteins. *In vivo* interference analysis by our group also shows that the complex is active and enables interference against phages and plasmids when produced in *E. coli*. To better understand this mechanism, it was necessary to show *in vitro* interference. As a first step, we attempted *in vitro* binding of the complex to a target DNA. Target binding by hybridization of a matching sequence to the spacer in the crRNA would form an R-loop structure necessary for subsequent degradation of target DNA. Target binding was performed for both ssDNA and dsDNA by incubating the complex with radioactively labeled target DNA molecules. These samples were then subjected to electrophoretic mobility shift assays (EMSAs) (Figure 2.2). Recombinant type I-Fv Cascade was capable of binding ssDNA with a complementary sequence to the spacer in the crRNA of the complex but no PAM-dependency or significant binding of a completely hybridized dsDNA was observed.

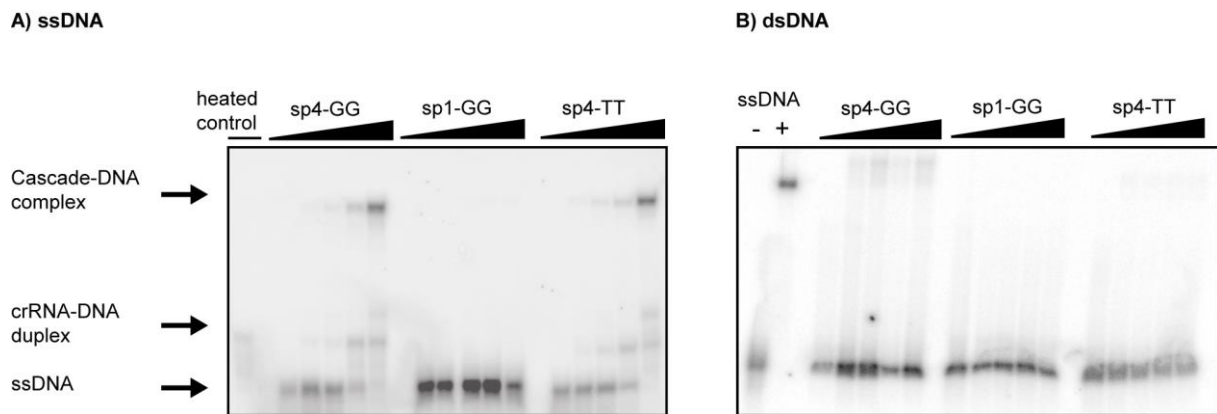


Figure 2.2: Electrophoretic mobility shift assays of type I-Fv Cascade with radioactively labeled ssDNA (A) or dsDNA (B). Increasing amounts of recombinant I-Fv Cascade were incubated with radioactively labeled DNA molecules. Target DNA contained either a sequence matching the sequence in the crRNA (sp4-GG), a non-complementary sequence (sp1-GG), or a matching sequence with a wrong PAM (sp4-TT). For ssDNA, a control was performed in which samples were heated at 95°C for 5 min. For B, an ssDNA-control was included with (+) or without (-) Cascade to confirm Cascade stability.

EMSA analysis shows that recombinant Cascade binds to complementary ssDNA *in vitro*. There seems to be no special target recognition in this reaction and binding only occurs due to Watson-Crick-base pairing during incubation. For this reason, it was also possible to see a duplex of crRNA and ssDNA without the complex which is especially apparent after a heated control. PAM-recognition could not be shown since the same gradient shift was observed for target ssDNA with a matching spacer sequence but a wrong PAM (TT). As target binding is based on hybridization of the crRNA to the protospacer, no target binding is possible for Cascade containing a non-matching crRNA (sp1).

Binding of dsDNA could not be clearly identified by EMSA. Only a faint band can be seen for the dsDNA substrate with the matching spacer sequence and the correct PAM which could indicate dsDNA binding with lower efficiency. A possible explanation for this difference is that type I-Fv Cascade evolved to exclusively target ssDNA, meaning that it might be coupled to processes in which ssDNA is formed.

To test this hypothesis and to otherwise provide a fully bound dsDNA target, a construct was designed that mimics a dsDNA strand unwound by a helicase or by R-loop formation during interference of other Type I systems (Figure 2.3 A). Specifically, two DNA oligonucleotides (i.e. target and non-target) were designed with the target containing the protospacer sequence matching the spacer of the crRNA. The upstream region of the 3'-sequence of the target strand was complementary to the radioactively labeled non-target strand to allow duplex formation. The remaining nucleotides of the non-target strand were not complementary to the target strand to not allow hybridization. Increasing amounts of Cascade were incubated with this molecule and analysed by EMSA (Figure 2.3 B).

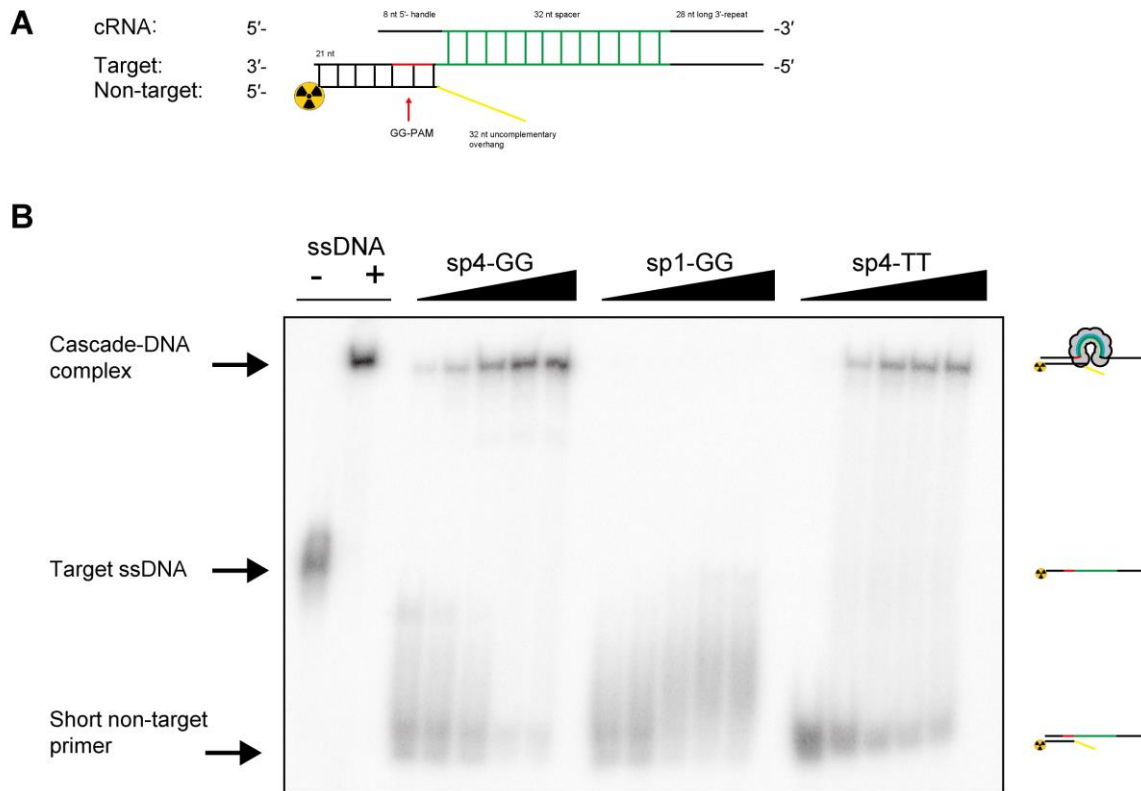


Figure 2.3: Electrophoretic mobility shift assays of type I-Fv Cascade with a radioactively labeled target mimicking opened dsDNA. (A) Schematic representation of the target constructs. The upper sequence shows the crRNA in the complex, below is the target sequence used in previous EMSAs. The bottom sequence shows the radioactively labeled non-target strand hybridized by the first nucleotides. Nucleotides in yellow are non-complementary to the spacer sequence (green, with an actual length of 32 nt) of the target strand. The PAM sequence is shown in red. **(B)** EMSA analysis of increasing amounts of type I-Fv Cascade with either matching sequence (sp4-GG), a non-complementary sequence (sp1-GG) in the target strand or a matching sequence with a wrong PAM (sp4-TT). The non-target strand was labeled instead of the target-strand to rule out ssDNA binding. An ssDNA-control was included with (+) or without (-) Cascade to confirm Cascade stability.

It is possible to see that target binding still occurs when a complementary sequence in the target strand is available, even if the upstream sequence is dsDNA. This suggests that binding of dsDNA molecules is possible if an opening is provided e.g. by a helicase. The potential connection to ssDNA-generating processes and the question of how the complex achieves R-loop formation without the missing subunits remains to be investigated. Target binding was again not observed for a non-complementary target sequence. PAM discrimination could also not be observed for this construct, as the band shift indicating target binding was still observed for the construct with the matching sequence and a wrong PAM (TT) as was the case for ssDNA. It remains to be shown how the complex is able to differentiate between PAMs *in vivo*.

2.1.3 3D Structure of small synthetic I-Fv Cascade

The major differences of type I-Fv Cascade compared to other Cascades from type I systems are the absence of large or small subunits and the lack of sequence similarity between Cas5fv and Cas7fv with their functional homologues (Dwarakanath *et al.*, 2015). To understand how this minimal type I-Fv Cascade and the highly divergent Cas5fv and Cas7fv proteins function and still provide interference, we aimed to determine the 3D structure of the complex by crystallization and X-ray diffraction in collaboration with Dr. Patrick Pausch from the research group of Prof. Dr. Bange. Crystals of wild-type Cascade diffracted poorly and we were unable to solve the 3D structure of the complex. As an alternative, we used the small synthetic Cascade variant that was previously analysed by size-exclusion chromatography and small-angle X-ray scattering (Gleditsch *et al.*, 2016). In this synthetic variant, the complex assembles around the reduced spacer with three instead of six subunits of the backbone-forming protein Cas7fv as well as one subunit of Cas5fv and Cas6f, respectively. Crystals of this condensed and less flexible variant were of sufficient quality to solve the 3D structure of the complex at a resolution of 3 Å (Table 6.1). The structure of type I-Fv Cascade revealed an elongated and crescent-shaped complex with a length of 130 Å along the crRNA axis (Figure 2.4).

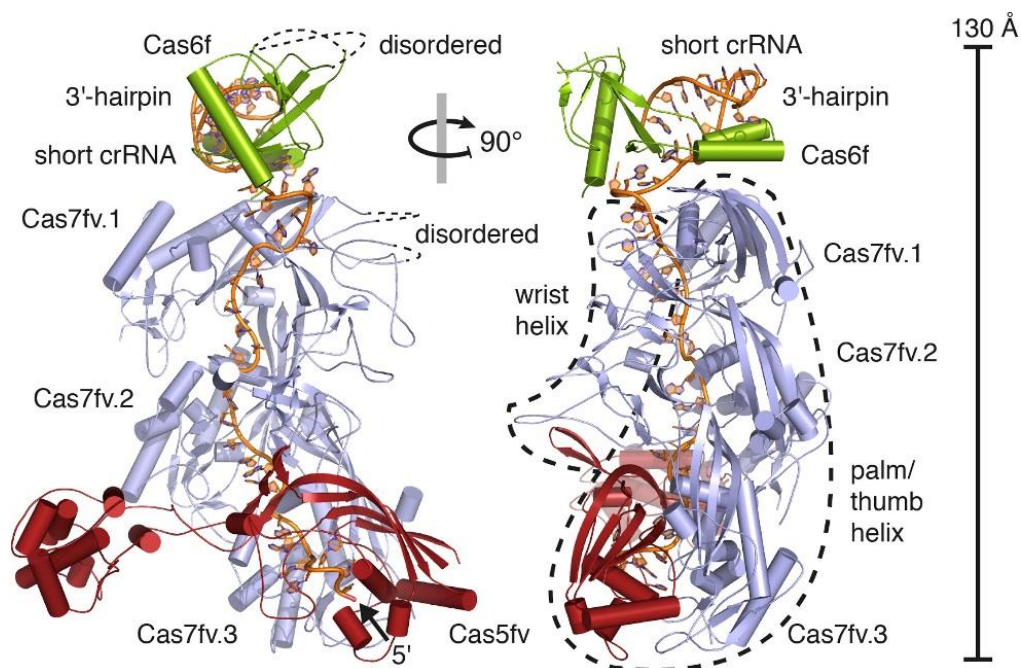


Figure 2.4: 3D structure of short I-Fv Cascade from *S. putrefaciens* CN-32. Cartoon representation of the short I-Fv Cascade X-ray crystal structure from *S. putrefaciens* CN-32 in two, 90° rotated orientations. Short crRNA, Cas5fv, Cas7fv and Cas6f are colored in orange, dark red, blue and green, respectively. The crRNA 3'-hairpin and 5'-end are indicated and the Cascade subunits are labeled. Disordered sections are labeled and indicated by thin dotted lines. The two parallel right-handed wrist and palm/thumb helices are labeled accordingly and are indicated by thick dotted lines. Right: scale bar illustrating the total height of 130 Å. Figure from Pausch *et al.*, 2017.

As expected, Cas6f binds one end of the structure by recognizing the 20 nt long crRNA 3'-hairpin tag. The backbone of Cas7fv assembles along the spacer sequence of the shortened crRNA with three subunits binding in increments of 6 nt. Cas5fv caps the other end of the structure at the 5'-end of the crRNA by interacting with the 8 nt long 5'-handle.

While the overall shape of I-Fv Cascade appears similar to related structures from other systems, certain aspects are drastically different. For reference, we compared the obtained structure of I-Fv Cascade to the related I-F Cascade from *Pseudomonas aeruginosa* (also termed crRNA-guided surveillance complex (Csy) complex) that was published a few weeks earlier (Chowdhury *et al.*, 2017). The absence of the additional subunits at the belly of the complex results in a more open configuration compared to its I-F counterpart (Figure 2.5 A) that is more reminiscent to the further related type I-E systems (Supplementary Figure 1).

The Cas6f protein of I-Fv Cascade is highly similar to its I-F counterpart as both are bound to the 20 nt long hairpin at the 3'-end of the crRNA and interact with the adjacent Cas7 protein (Figure 2.5 B & C). While the ferredoxin-like domain of Cas6f interacts with the palm domain of Cas7f in the type I-F system, the crRNA-binding α -helical hairpin establishes a similar interaction in type I-Fv (Figure 2.5 B & C). In comparison, this rearrangement leads to an approximately 90° tilted reorientation for the 3'-hairpin (Figure 2.5 D).

Both the palm and the thumb domain are present in the type I-Fv Cas7 protein but the fingers are strongly reduced and instead, two extensive loops (aa 25-77) are present next to thumb at bottom of palm (Figure 2.5 B). In accordance with the right-hand analogy of these domains, we termed them wrist-loops (WL1 and 2). WL1 and 2 are connected with each other and form a unique helical filament at the concave side of the complex, where small subunits are located in type I-E Cascade. This wrist helix runs parallel to the helix formed by the palm and thumbs, the palm/thumb helix (Figure 2.4, Figure 2.5 B). Cas5fv interacts with the 8 nt long S-shaped 5'-handle at the opposite end of the crRNA and encases it via the RRM domain in the palm (Figure 2.5 B & D), as is the case for type I-E and I-F (Chowdhury *et al.*, 2017). However, Cas5fv also contains a wrist that connects to the wrist helix formed by the Cas7fv backbone (Figure 2.4 and Figure 2.5 B). The most striking difference is the presence of an additional domain that consists of six α -helices (AH, residues: 110–266; Figure 2.5 A & B) and extrudes from the convex side of the complex, where the tip of the thumb pins the crRNA against the palm of the adjacent Cas7fv (Figure 2.5 B). The position of this domain at the location of the large subunit in type I-F and I-E suggests that it might compensate for its absence.

Overall, type I-Fv drastically differs from its homologues in type I-F. Structural alterations are present in the Cas7 and Cas5 proteins at positions where the small and large subunits are located in other Cascades of type I. Thus, large and small subunits that are usually essential for DNA recruitment and interference, are replaced.

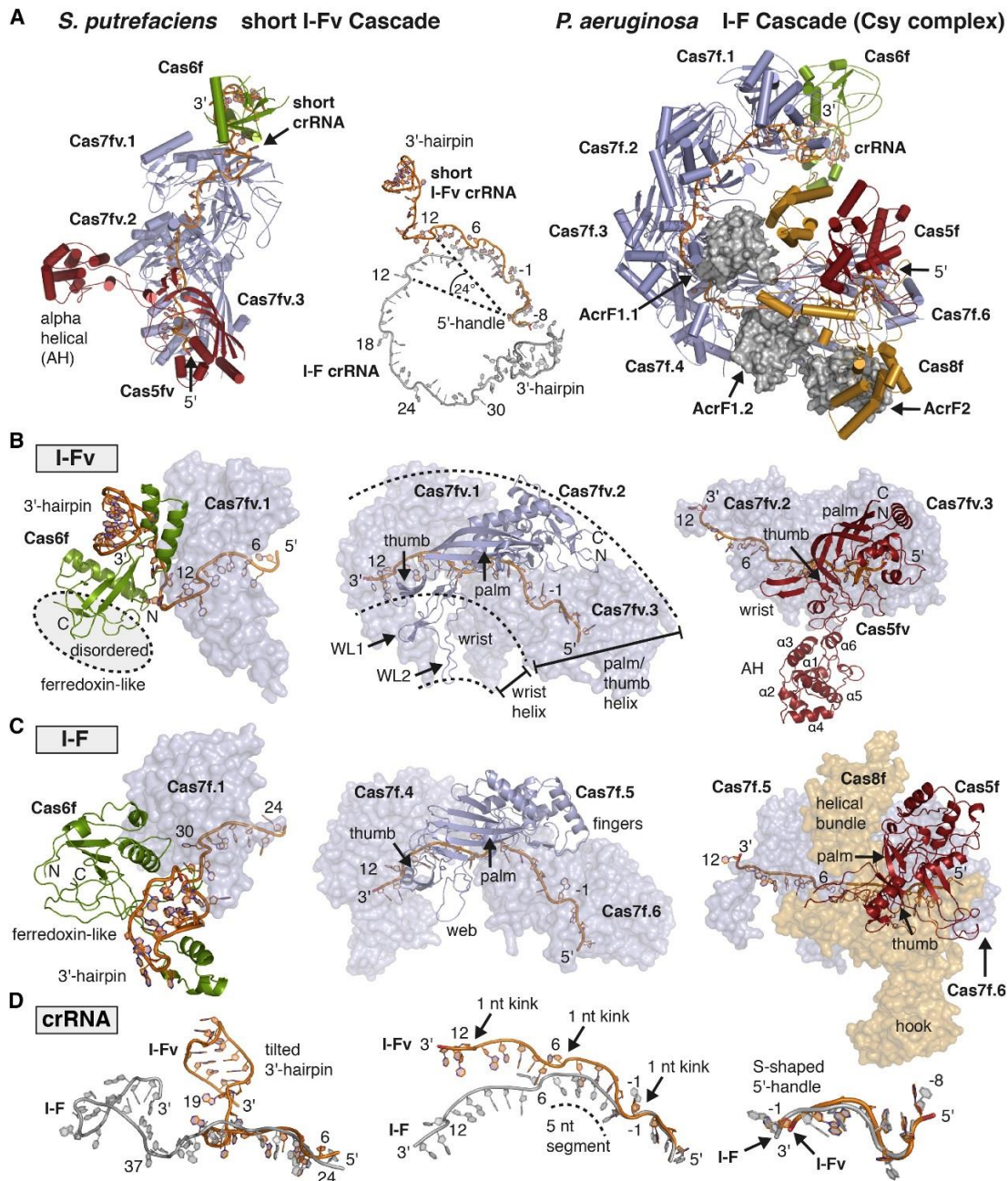


Figure 2.5: Structural comparison of type I-F and type I-Fv Cascade. (A) Left: X-ray crystal structure of the short *S. putrefaciens* I-Fv Cascade shown in a cartoon representation. Color scheme and labeling are as in Figure 2.4. Middle: crRNA spine superimposition of the short I-Fv (orange) and I-F (grey) crRNA on the basis of a 5'-handle alignment. Nucleotide positions upstream of the first spacer nucleotide are labeled with negative values and the positions of downstream nucleotides are indicated by positive values. The angle of 24° between nucleotide position -6 and 12 illustrates the different crRNA spine pitch. Right: cryo-EM structure of the AcrF1/2 bound *P. aeruginosa* I-F Cascade (PDB ID: 5ZU9; (Chowdhury *et al.*, 2017)). Components are labeled according to the current nomenclature for type I-F. Color scheme of the I-F Cas homologs is according to I-Fv. The additionally present large subunit protein Cas8f is shown in yellow and the activity inhibiting AcrF1/2 proteins are shown in a grey surface representation. (B) and (C): Side-by-side comparison of the I-Fv (B) and I-F (C) Cascade subunits. Color is according to Figure 2.4. The Cas6f proteins are compared in the left panel, Cas7 homologs in the middle panel and Cas5 homologs in the right panel. Adjacent Cascade subunits are shown as transparent surfaces and labeled respectively. Grey circles indicate disordered regions. N and C indicate N- and C-termini, respectively. (D) crRNA spine comparison of I-F (orange) and I-Fv (grey). Left: crRNA arrangement at the tilted head structure. Superimposition according to I-Fv nucleotides 6 to 12. Middle: crRNA architecture of the Cas7 backbone bound segment. Superimposition according to I-Fv nucleotides -1 to -4. At every sixth position, the nucleotide is splayed out from the base stacking segments ('kink'). Right: Superimposition of the similar S-shaped 5'-handles. Figure from Pausch *et al.*, 2017.

2.1.4 3D structure of I-Fv Cascade bound to target DNA

Next, we aimed to understand how the minimal I-Fv Cascade achieves PAM-dependent recognition of foreign DNA without small and large subunits. For this, we reconstituted and co-crystallized I-Fv Cascade bound to the previously used DNA target duplex (Figure 2.3) with minor modifications in terms of nucleotide length (Figure 2.6 B).

The target bound structure (3.25 Å) shows that Cas5fv directly recognizes the GG-PAM motif via the AH domain. The dsDNA section downstream of the GG-PAM is pinched in between the RRM fold and the AH domain of Cas5fv (Figure 2.6 A). In the section upstream of the PAM, the split target and non-target strand are guided along the Cas7fv backbone in two different routes (Figure 2.6 A) with a maximum distance of approximately 25 Å (non-target strand T22/ target strand A7). The target strand protospacer region is hybridized with the crRNA spacer while the non-target strand is aligned to the wrist helix.

Further upstream, the 5'-region of the target DNA protospacer pinches the thumb of Cas7fv.2 in between the crRNA and the target strand (Figure 2.6 A & C). The thumb of Cas7fv.2 is stabilized by this and establishes a salt bridge interaction of R155 to E17 of the ferredoxin-like domain of Cas6f, which rotates it by approximately 7 Å and stabilizes it in turn (Figure 2.6A, Figure 2.7 A). These conformational rearrangements might be relevant for target DNA association and Cascade stalling by R-loop retention. The separated target and non-target strands are recruited by the two parallel helices on each side of I-Fv Cascade. Association of the target strand in the positively charged central channel by the palm/thumb helix is similar to type I-E Cascade (Hayes *et al.*, 2016, Mulepati *et al.*, 2014) and relies on a set of sequence-independent DNA interactions. The thumbs of Cas7fv splay out every sixth nucleotide of the protospacer, while the nucleotides in between are hybridized with the crRNA. Aromatic residues emanating from the Cas7fv thumb (Y149, F160, F161) further stabilize this interaction and stack the nucleobases in place that lie 5'-adjacent to crRNA and DNA kinks (Figure 2.6 D). R155 at the thumb tip forms a salt-bridge with the adjacent Cas7fv D192, similar to the ferredoxin-like fold of Cas6f and the thumb of Cas7fv.2 (Figure 2.6 D). Surprisingly, Cas7fv not only interacts with the target strand but also guides the non-target strand in parallel via the wrist helix at the opposite side of the complex (Figure 2.6 A, E). The non-target strand passes along a path formed by WL1 and WI2 and is stabilized by sequence unspecific interactions via tyrosine 62 and 64 (Figure 2.6 E). Thus, the wrist helix establishes the trench route for the non-target strand and compensates for the loss of the large and small subunits.

The PAM containing DNA duplex is pinched between the AH domain and a small helix (SH) at the wrist of Cas5fv and recruited by a set of polar interactions (Figure 2.6A, F). Contrary to type I-E Cascade, the GG-PAM in type I-Fv is recognized in duplex form from the major groove side by the N-terminal linker and α -helix 6 of the AH domain (Figure 2.6 A, F). In the centre, E113 distorts the first PAM bases guanosine (G15) of the target and the corresponding cytosine of the non-target strand (C14) (Figure 2.6 F, G).

Guanosine G15 interacts with lysine 252, while the second base pair of the PAM, guanosine/cytosine (G16/C13), interacts with lysine 252 and aspartate 253 of the C-terminal helix of the AH domain, respectively (Figure 2.6 F, G). These central residues recognizing the GG-PAM are flanked at the AH domain by the side-chains threonine 251 and aspartate 254 of α -helix 6, possibly also contributing to PAM recognition (Figure 2.6 F, G).

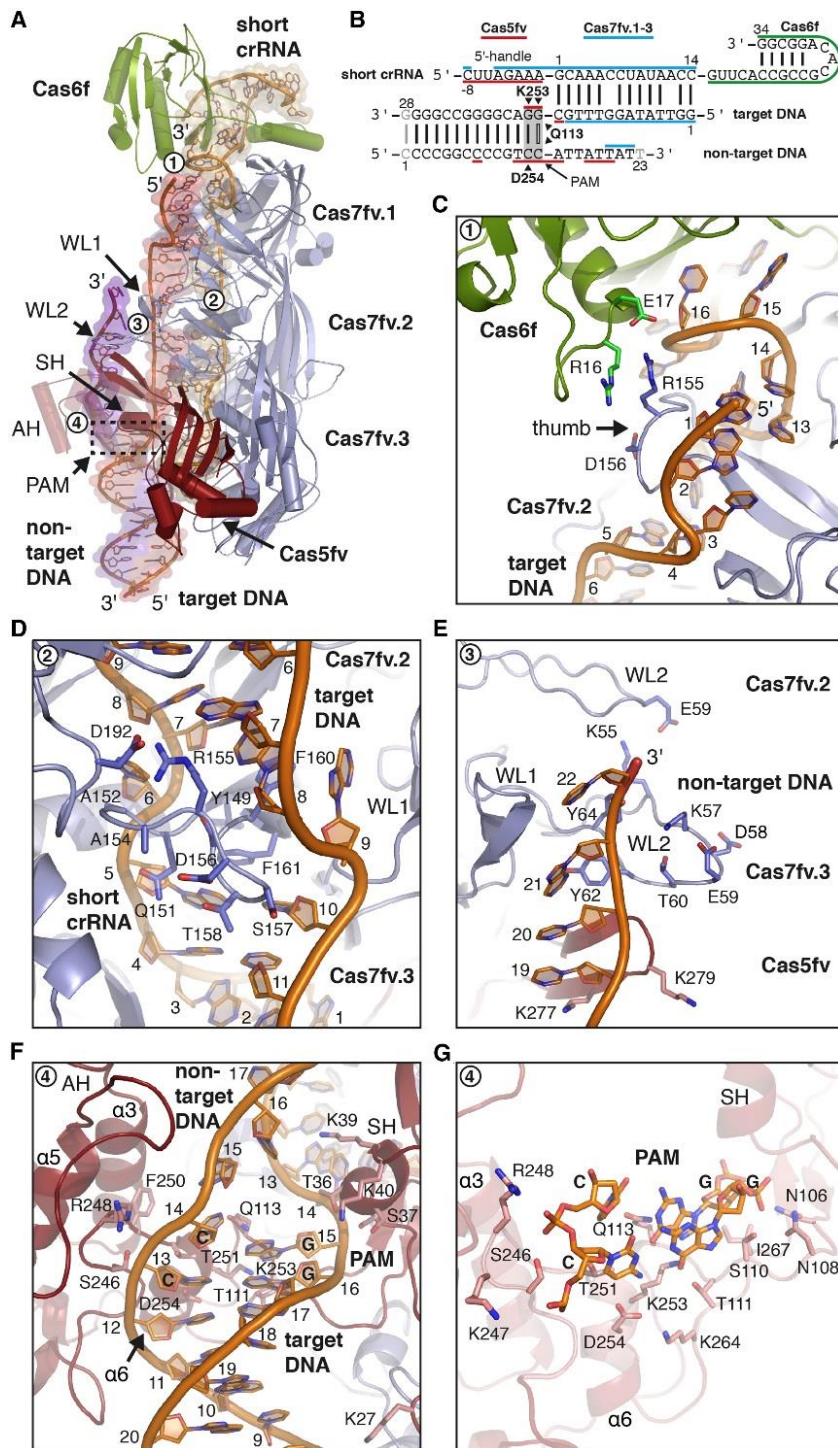


Figure 2.6: X-ray crystal structure of the short I-Fv Cascade R-loop complex. (A) Overview of the R-loop Cascade crystal structure. Components are shown in cartoon representation and colored and labeled according to Figure 2.4 and Figure 2.5. Nucleic acid components are highlighted for clarity by transparent surfaces (orange: short crRNA; red: DNA target strand; violet: non-target strand). Important regions for nucleic acid interaction, detailed subfigures C to G, are tagged with numbers in white circles (1-4) for orientation. (B) Design of target and non-target primers for the reconstitution of the R-loop/ I-Fv Cascade complex. The blue, red and green lines indicate the interface between Cas 7fv, Cas 5fv and Cas 6f and the nucleic acids as observed in the R-loop/ I-Fv Cascade structure. Arrows indicate amino acids of Cas 5fv interacting with the GG-PAM. Grey letters indicate disordered nucleotides. (C) to (F) Close up view of the DNA interacting regions close to the Cas 6f head structure (C), at the thumb of Cas 7fv.3 (D), the base of the wrist helix (E) and the PAM recognition site in between the AH and SH of Cas 5fv (F). Amino acid side chains in close proximity to nucleic acids are shown as sticks and are labeled according to their identity and position. Nucleic acids are labeled according to subfigure B. (G) Detailed view on the GG-PAM, shown in stick representation. Adjacent nucleotides were removed for clarity. Figure from Pausch *et al.*, 2017.

Superimposition of apo-Cascade and target-bound Cascade reveals a conformational shift of the AH domain by 12.5 Å towards the complex body, which inserts α -helix 3 against the first G-C base pair of the PAM (Figure 2.7 A). This “wedge” helix locks the target in position by not allowing re-association of the target and non-target strand at the seed region, as shown by an ideal B-form dsDNA aligned to the dsDNA PAM downstream region (Figure 2.7 B). The polar side-chains (N178-K178) might assist in DNA strand separation (Figure 2.7 C). Residues T251, K253 and D254 are predicted to interact directly with the PAM either at the target (K253) or at the non-target strand (T251). *In vivo* assays conducted in our laboratory by Dr. Hanna Müller have shown that single exchanges of these amino acids to alanine have only a mild effect on interference, while a triple exchange completely abolished interference.

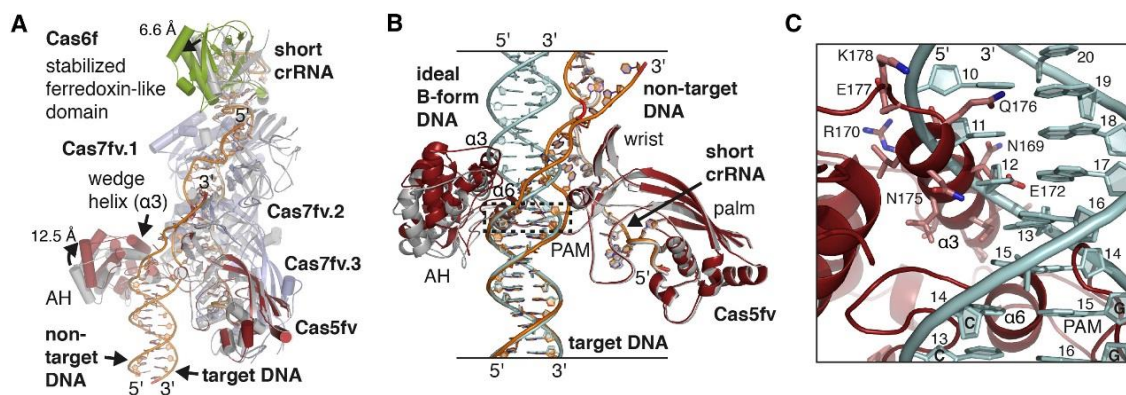


Figure 2.7: Structural reorganization of type I-F Cascade upon R-loop formation. (A) Superimposition of the short Apo Cascade (grey cartoon) and the R-loop associated short Cascade (colored cartoon). I-Fv Cascade undergoes structural rearrangements at the Cas6f head and Cas5fv AH domain, indicated by arrows and distances. (B) Superimposition along dsDNA segment of the aligned apo (grey) and R-loop bound (colored) short Cascades with an ideal B-form DNA (light blue). The PAM region is highlighted with a dashed line. (C) Close up on the superimposition with an ideal B-form DNA (shown in subfigure B) upstream of the GG-PAM, emphasizing steric clashes that would occur between the target strand DNA and α -helix 3 (wedge helix) of the Cas5fv AH domain upon target and non-target strand association. Residues are labeled according to their identity and position. Figure from Pausch *et al.*, 2017.

In conclusion, type I-Fv Cascade recognizes target DNA from the major groove side by Cas5fv. Cas5fv is not only responsible for capping the 5'-end of the crRNA but also for target recognition, compensating for the loss of the large subunit.

2.1.5 Requirement of AH and WL domains for complex formation

The impact of the described wrist helix domain of Cas7fv and the AH domain of Cas5fv on the formation and stability of I-Fv Cascade was investigated by replacing these domains with a flexible linker (GGSGGS). Both truncated constructs were co-expressed with the wild-type crRNA and used for Cascade production in *E. coli*. Cas proteins were purified by Ni-NTA followed by size-exclusion chromatography.

Recombinant I-Fv Cascade was still able to assemble with a deleted AH domain (Figure 2.8 A) but not without the wrist helix (Figure 2.8 B).

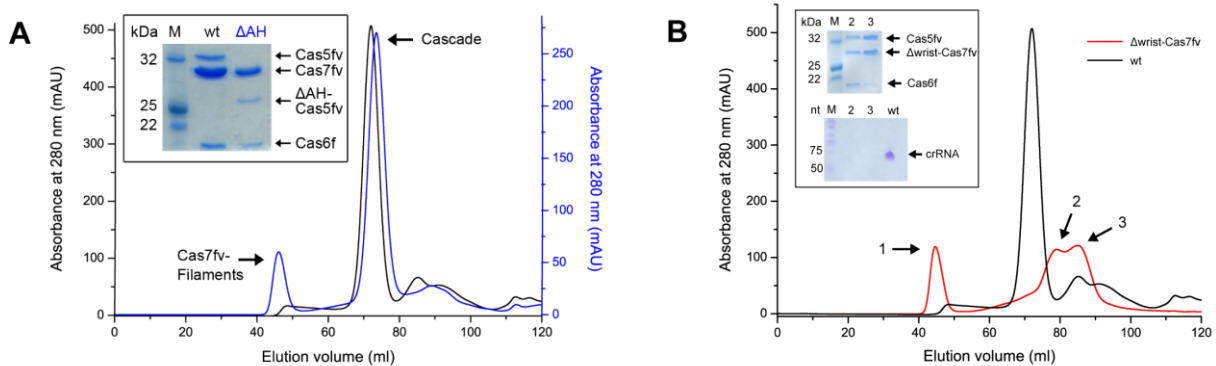


Figure 2.8: Requirements of Cas5fv and Cas7fv domains for I-Fv Cascade. (A) Truncated Cas5fv, missing the AH domain is incorporated into a stable Cascade complex. His-tagged Δ AH-Cas5fv co-elutes with Cas7fv and Cas6f as a single peak during size-exclusion chromatography (blue) but with a later elution volume, corresponding to the smaller size, in comparison to wt-Cascade (black). (B) No stable Cascade complex is formed with Δ wrist-Cas7fv. All three Cas proteins still elute in one peak (red), but without overrepresentation of Cas7fv and without incorporation of crRNA.

This further supports the hypothesis that the AH domain evolved to directly replace the large subunit and rules out misassembled complexes as a reason for the loss of interference in the conducted *in vivo* assays. No stable Cascade complex was obtained in the construct where the wrist loops of Cas7fv were deleted and only minor peaks were visible in the size-exclusion chromatogram (Figure 2.8 B). All three Cas proteins, Cas5fv, Cas6f and a truncated Cas7fv eluted at the same position and were detected by SDS-PAGE. However, the Δ wrist-Cas7fv is not overrepresented in comparison to Cas5fv and Cas6f and separation on Urea-PAGE did not reveal crRNA to be incorporated in these samples, further arguing against Cascade formation. Separating both loops individually might produce a stable Cascade complex without the wrist helix in future attempts. Otherwise, this result suggests that the wrist helix is required for Cascade assembly or stabilization.

2.1.6 Investigation of the Cas3fv nuclease activity

Previous analysis of the Cas2/3 fusion protein of the I-Fv system with bioinformatics tools suggested that this protein is a typical metal-dependent ssDNA nuclease responsible for target degradation. Previous work in our group showed that the deletion of these proteins both in the native host *S. putrefaciens*, as well as in *E. coli* abolished the activity of the CRISPR-Cas system *in vivo*. Recombinant production of Cas2/3fv turned out to be problematic and the fusion of a SUMO-tag increased solubility.

To optimize the purification procedure, we first attempted to purify the protein fused to an MBP-tag (maltose-binding protein) as this was described to assist in Cas3 protein production for related CRISPR-Cas systems. While good amounts could be purified via Ni-NTA, size-exclusion purification revealed that the protein eluted exclusively in the void volume and thus most likely in aggregated form. Indeed, no nuclease activity of these samples was observed on ssDNA.

For easier crystallization of the Cas3fv protein, we deleted the Cas2 domain from the Cas2/3fv fusion protein and were able to purify a stable standalone Cas3fv protein. This purification has been established by Dr. Patrick Pausch for higher yield by using a wash buffer with higher salt concentration (0.75 M NaCl) and immediate size-exclusion chromatography after Ni-NTA elution. With this, Cas3fv elutes both in the void volume, presumably in either an aggregated form or bound to nucleic acids, but also in monomeric form at the appropriate elution volume during size-exclusion chromatography (Figure 2.9 A).

The activity of the standalone Cas3fv protein was first investigated with ssDNA substrates by incubation for various time points in a buffer containing Mg^{2+} and Ca^{2+} ions as well as ATP. It was not possible to purify the catalytically dead HD-mutant of Cas3fv for control purposes because it remained insoluble. Instead, EDTA was added in a control reaction to quench the metal-dependent reaction. This assay showed that Cas3fv was able to degrade ssDNA over time (Figure 2.9 B).

During natural immunity in the host organism, Cas3 is recruited by the target-bound Cascade complex where an R-loop structure is formed. Thus, we tested Cas3fv activity on Cascade-bound target DNA. R-loop formation by Cascade was investigated by incubating Cascade with a dsDNA construct containing a 10 nt bubble adjacent to the PAM sequence. By EMSA analysis, Cascade was shown to be able to bind to this construct as increasing amounts of Cascade resulted in a band shift of the dsDNA substrate to a full R-loop substrate (Figure 2.9 C). The small 10 nt opening of the target dsDNA was apparently sufficient for the unwinding of the following dsDNA by Cascade and allowed full binding of the crRNA spacer to the matching sequence. The requirements for R-loop formation, the opening of dsDNA targets, and target recognition are open questions that require further investigation.

The nuclease activity of Cas3fv was analysed on radioactively labeled substrate bound to Cascade or the empty “bubble” construct. Cleavage on both substrates was observed when they were further incubated with Cas3fv for 2h. To observe the cleavage pattern, samples were separated via denaturing Urea-PAGE (Figure 2.9 D). Incubation with Cas3fv (500 nM) removed the 90 nt substrate completely and resulted in a prominent band with a size smaller than 10 nt, corresponding to shredded DNA. It should be noted that the substrate quality on the nuclease assays is improvable considering the smear it produces on the gel unless degraded by Cas3fv. Additionally, the HD-mutant would be the best control to rule out unspecific cleavage. Due to a control in which substrate alone was incubated for the full-time

period without protein and EDTA, additional contaminants in the Cas3 purification sample that are able to cleave DNA can be ruled out.

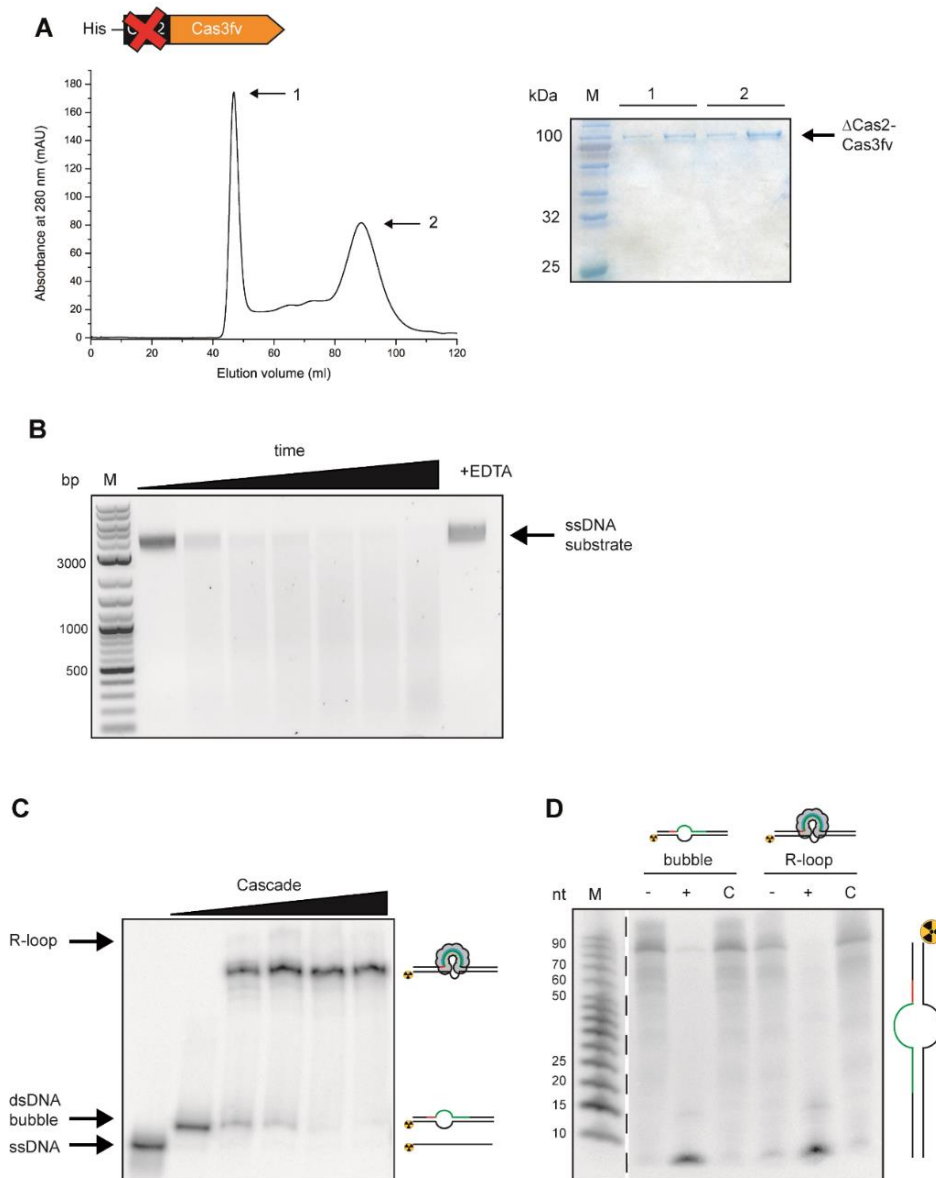


Figure 2.9: Purification and nuclease activity of standalone Cas3fv. (A) Purification of standalone Cas3fv without the Cas2 domain. Schematic representation in the top. UV chromatogram of the final size-exclusion purification step shows two peaks of Cas3fv. SDS-PAGE analysis confirms the presence of Δ Cas2-Cas3fv with a reduced size of 100 kDa in both peaks. (B) ssDNA cleavage activity of Cas3fv observed on agarose gel electrophoresis. 500 nM of Cas3 were incubated with substrate for various time points (0, 2, 5, 10, 20, 30 and 60 min). Increased incubation time results in degradation of the ssDNA unless the reaction was inhibited by EDTA. (C) EMSA of radioactively labeled dsDNA containing a small 10 nt opening (“bubble”) with increasing amounts of Cascade. Cascade is able to bind the substrate and unwind the following dsDNA section, forming an R-loop structure in the process. (D) Nuclease activity of Cas3fv on empty bubble dsDNA and R-loop substrates. Substrates were either incubated without Cas3fv (-) with Cas3fv (+) or with Cas3fv but also EDTA (C) for 2 h at 30 °C. Incubation with Cas3fv resulted in degradation of both labeled non-target strands to small fragments below 10 nt.

With the optimized purification, enough sample could be obtained to initiate structural analysis by HDX-MS and crystallization. Preliminary HDX-MS data suggests that Cas3fv binds the non-target strand as it exhibits increased protection from hydrogen deuterium exchange (data not shown). So far, crystallization of Cas3fv was not successful.

2.1.6.1 Interaction of Cas1 and the Cas2/3fv fusion protein

Another characteristic of type I-F Cas2/3 effector proteins is the fusion of the adaptation protein Cas2 to the nuclease Cas3 and the assembly of two Cas2/3 proteins with two dimers of Cas1, the second protein of the acquisition machinery (Rollins *et al.*, 2017).

Structural analysis of the Cas1-Cas2/3 complex from the type I-F system of *Pseudomonas aeruginosa* showed that the complex adopts a four-lobed propeller-shaped structure (Rollins *et al.*, 2017, Fagerlund *et al.*, 2017). This complex was shown to have a regulatory role in interference as Cas3 nuclease activity is significantly decreased for the Cas1-Cas2/3 complex and only restored on R-loop substrates due to recruitment by Cascade.

To study if this regulatory feature also exists in the I-Fv system, we purified the Cas1-Cas2/3 complex of this system by fusing a Strep-tag to the N-terminus of Cas1 and co-producing Cas1 with Cas2/3fv followed by size-exclusion chromatography (Figure 2.10 A). The complex elutes at a volume corresponding to a molecular weight of ~ 400 kDa, which indicates the presence of two subunits of Cas2/3 and two dimers of Cas1. In accordance with this, SDS-PAGE revealed similar band intensities for both Cas2/3fv and Cas1. Cas1 also elutes in the void volume as well as during late elution with a size that corresponds to a dimer of Cas1.

The ssDNA cleavage activity of the complex was investigated and compared to ssDNA cleavage by Cas3fv (Figure 2.10 B). While Cas3fv effectively processed and completely degraded ssDNA, only a much higher concentration of the Cas1-2/3 complex was able to degrade the substrate. This might be due to partial sample inhomogeneity by complex breakdown or a generally reduced activity. Overall, ssDNA cleavage activity of Cas2/3fv is reduced by the complex formation with Cas1 which is in agreement with studies performed with type I-F Cascade (Rollins *et al.*, 2017).

The nuclease activity of the Cas1-Cas2/3fv complex was then investigated with the radioactively labeled bubble dsDNA and R-loop substrates previously used to study Cas3fv nuclease activity (Figure 2.10 C). Unlike nuclease activity by standalone Cas3fv, cleavage by the Cas1-Cas2/3fv complex produced a prominent band at ~ 40 nt for both open dsDNA as well as R-loops, although not all substrate was processed. Incubation of Cas1-Cas2/3fv with ssDNA in this assay also partially removed the substrate and produced the prominent 40 nt band as well as completely degraded DNA below 10 nt. A control in

which EDTA has been added shows that Cas1-Cas2/3fv activity depends on divalent metal cations, which is in agreement with observations for Cas3fv nuclease activity.

Full cleavage of the substrates might be achieved by a higher input concentration of Cas1-2/3fv or longer incubation time. The distinct band of ~ 40 nt could represent the position at which Cas3fv initially nicks the target DNA before unwinding and degrading the adjacent dsDNA segments. It is unclear why this nicking is more prominent than for the Cas3fv experiment and repetition of this experiment is necessary to confirm these results. The cleavage products of the empty bubble construct were unexpected in general, considering that the complex was shown to have a decreased cleavage activity on ssDNA and the complex from the I-F system inhibited activity unless it was recruited by Cascade. This activity could be due to a mechanistic difference in Cas3fv compared to Cas3f or due to complex breakdown during the long incubation time in this assay. More experiments are required to confirm these results and investigate the function of the Cas1-Cas2/3fv supercomplex.

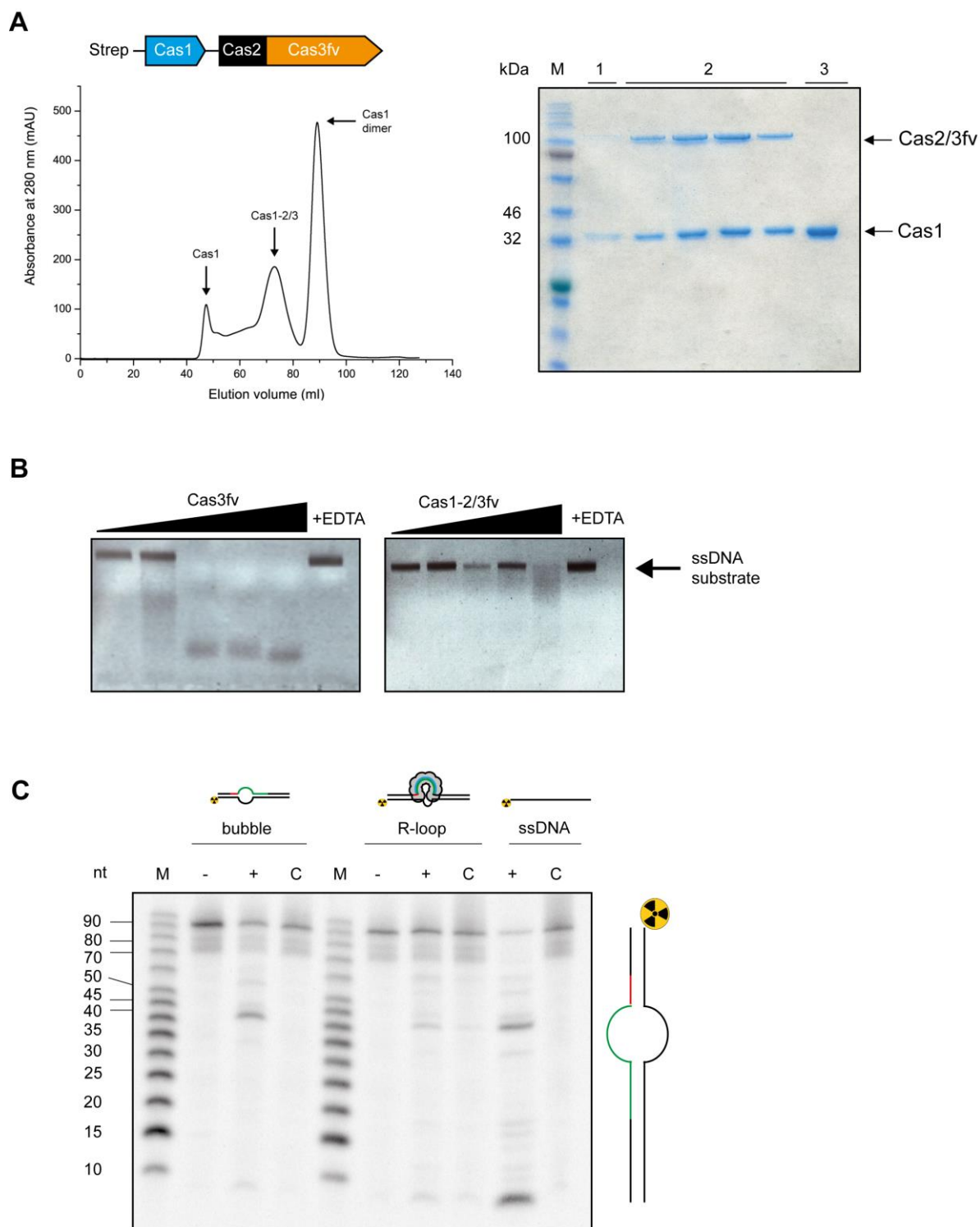


Figure 2.10: Purification and nuclease activity of the Cas1-Cas2/3fv complex. (A) UV chromatogram of Cas1-2/3fv complex purification by size-exclusion (left) and a schematic view of the proteins (top). The Cas1-Cas2/3fv complex was purified by Strep-tagged Cas1 and all purified proteins were separated by size-exclusion chromatography. Cas1-2/3fv complex elutes in a peak at ~400 kDa, flanked by two peaks of aggregated Cas1 and a Cas1 dimer. SDS-PAGE analysis of peak fractions from size-exclusion shows Cas1 and Cas2/3fv proteins (right). **(B)** Comparison of ssDNA cleavage activity of Cas3fv and Cas1-Cas2/3fv complex. M13mp18 ssDNA target substrate was incubated with increasing concentration of either Cas3fv or Cas1-2/3fv complex (0, 50, 100, 200 and 500 nM) and separated by agarose gel electrophoresis. Only the highest concentration (500 nM) of Cas1-2/3fv removed the substrate in comparison to much less concentration needed for just Cas3fv (100 nM). EDTA inhibited the cleavage for both proteins. **(C)** Nuclease assay of Cas1-2/3fv with dsDNA substrates containing a bubble in the spacer sequence, a full R-loop substrate by incubation with Cascade and a ssDNA substrate. Cleavage products were analysed by separation on Urea-PAGE, after incubation of Cas1-2/3 complex with the 5'-end non-target labeled substrates (+) or only substrate (-). EDTA was added as a negative control (C) to quench the reaction.

To compare the structure of the type I-Fv complex to the published I-F “propeller” shaped complex, we initiated SAXS analysis to model the general shape of the structure. High amounts of sample were purified and scattering data was recorded but the obtained data did not provide the structure and these experiments need to be repeated. Structural analysis of the Cas1-2/3fv complex and comparison to the I-F system could be useful to understand the general interference mechanism of this system, in addition to solving the 3D structure of Cas3fv.

2.2 Synthetic Cascade assembly and RNA wrapping

Previous experiments have investigated how the minimal I-Fv CRISPR-Cas system achieves interference by purification and analysis of the involved Cas proteins. A curious observation during these experiments was the formation of extended filaments by the Cascade backbone protein Cas7fv. These long helical structures are always by-products during purification when utilizing a His-tag on Cas7fv (Dwarakanath *et al.*, 2015). It thus seems that Cas7fv has the ability to bind RNA in general, which is essential for the formation of Cascade complexes assembled on varying spacer sequences from different sources in nature. If no crRNA is present, Cas7fv seems to bind unspecific RNAs present in the cell instead. The flexible backbone formed by Cas7fv subunits without any small or large subunits also makes it possible to cover longer sequences of RNA. We utilized this ability to artificially extend natural occurring spacers in the crRNA to produce synthetic Cascade variants.

With these observations in mind, we proposed to utilize the general RNA binding ability of Cas7fv for complex formation on specific RNA of choice, potentially providing a variety of useful applications that will be investigated in the following part of this work. We have termed this complex formation “RNA wrapping” and will refer to this term from this point on.

2.2.1 *In vitro* RNA wrapping with I-Fv Cas proteins

We first attempted to form complexes on provided RNA with purified Cas7fv *in vitro*. This process could be useful for the specific stabilization of extracted RNA. To achieve this *in vitro* RNA wrapping, it was first required to obtain the Cas7fv protein in an RNA-free state and not assembled to a complex. In theory, complexes could then be created by mixing this protein with RNA provided by extraction or *in vitro* transcription (Figure 2.11).

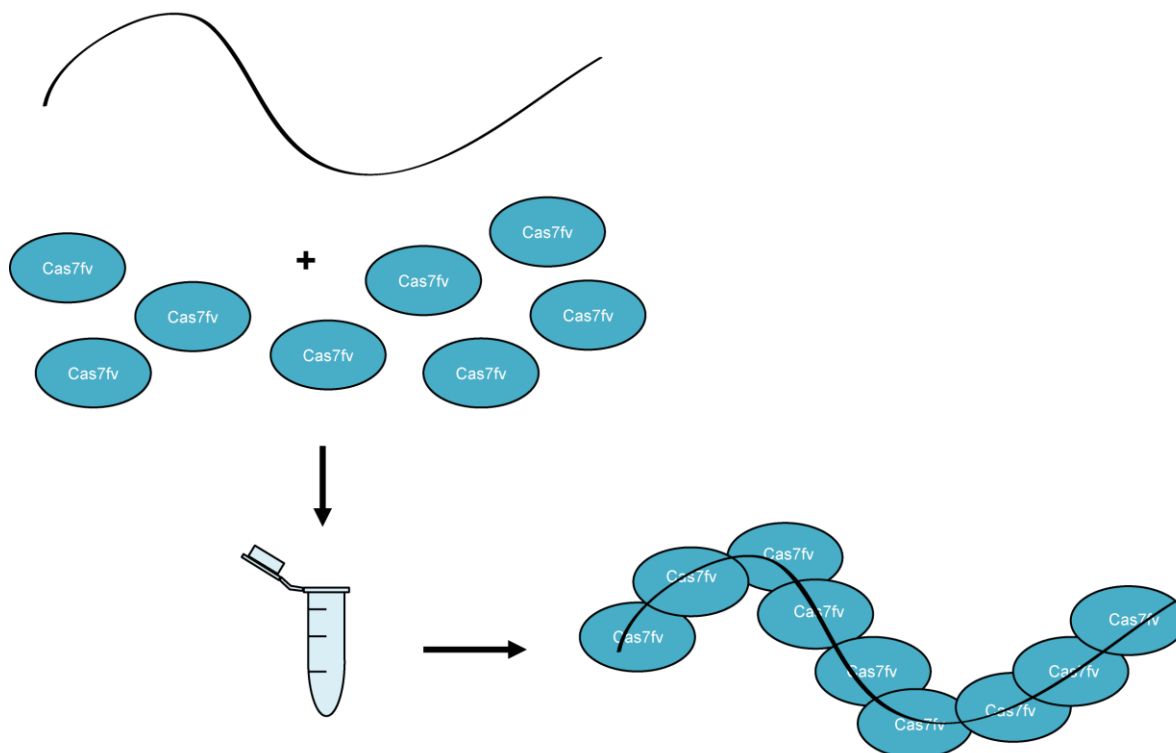


Figure 2.11: Schematic of the *in vitro* RNA wrapping process. Unbound RNA (extracted RNA or *in vitro* transcribed) is mixed and incubated with unassembled apo-Cas7fv, forming a Cas7fv-RNA complex in the process.

To obtain apo-Cas7fv, we used a variety of methods. First, we used the Cas5fv-Cas7fv dimer that is a consistent by-product during purifications because this dimerization stabilizes the otherwise insoluble proteins. We also theorized that the dimer form is required to deliver Cas7fv to the crRNA during natural Cascade assembly. To mimic this assembly process as closely as possible, we produced a target RNA with a 5'-handle sequencing by *in vitro* transcription and dephosphorylated it to create an OH-group at the 5'-end as is the case after processing by Cas6f. This was also considered to be potentially necessary, in case that assembly does require Cas5fv binding for initiation. In theory, the Cas5fv-Cas7fv dimer could also be used on a random RNA without repeat-tag for general RNA-binding without Cas5fv for target specificity but the properly processed repeat-tag should work for both eventualities if this process works in general. Fresh Cas5fv-Cas7fv dimer was taken from a I-Fv Cascade purification and mixed with the target RNA. After incubation for 1h at RT precipitation was noticeable. The supernatant was then loaded on an analytic size-exclusion column and separated to look for formed complexes. Unfortunately, no assembled complexes could be detected and the fractions corresponding to the only size-exclusion peak contains the Cas5fv-Cas7fv dimer as shown by SDS-PAGE (Figure 2.12).

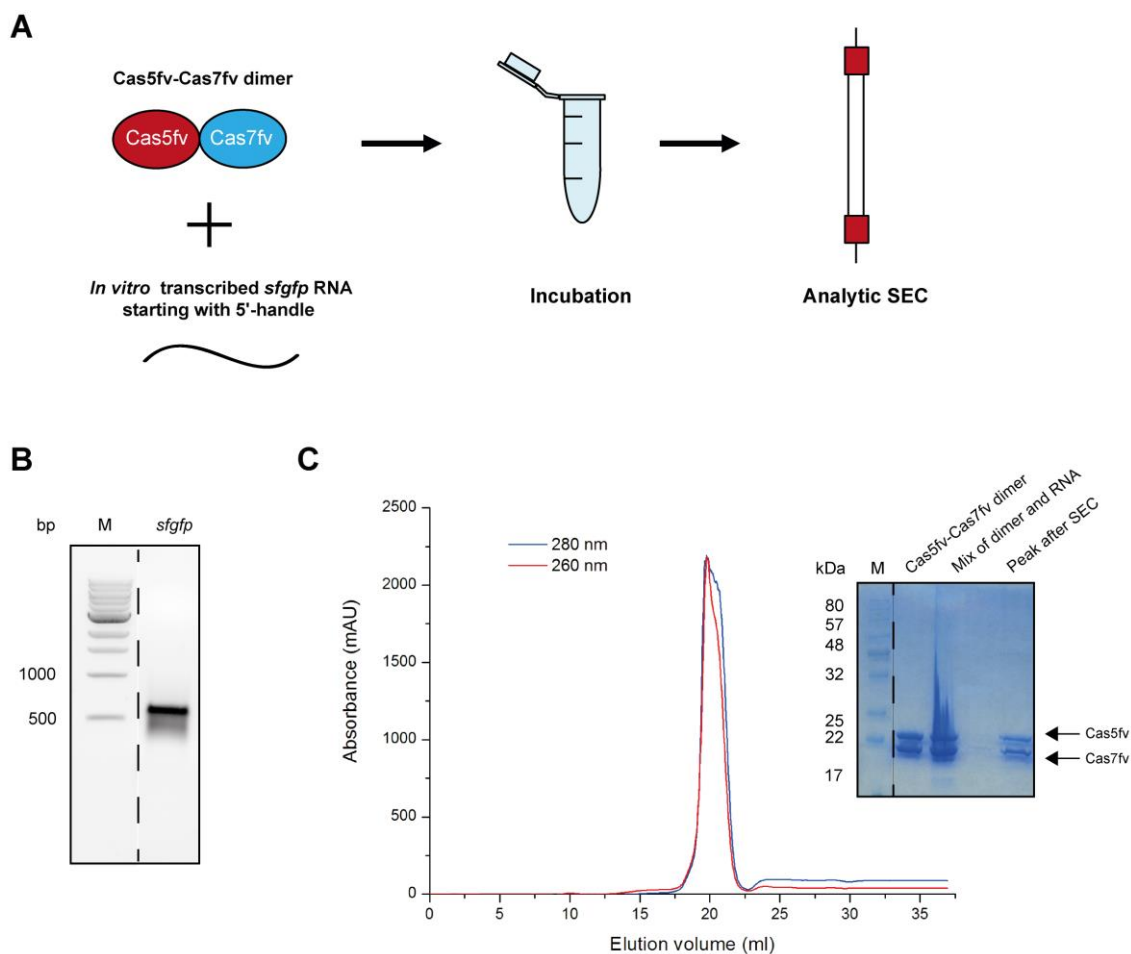


Figure 2.12: *In vitro* RNA wrapping with the Cas5fv-Cas7fv dimer. (A) Schematic principle of the experiment. Purified Cas5fv-Cas7fv dimer is incubated with *in vitro* transcribed *sfGFP* for 1h at RT. Precipitated protein is removed and the supernatant is then separated by aalytic size-exclusion chromatography. (B) Gel electrophoresis of *in vitro* transcribed *sfGFP* RNA used as template for RNA binding. (C) UV chromatogram of aalytic size-exclusion chromatography (left) and analysis by SDS-PAGE (right) of RNA incubated with the dimer. Complex assembly was not detected, and only a peak of the Cas5fv-Cas7fv dimer is eluting.

A possible explanation for the inability of the dimer to form a complex on the provided RNA could be that the interaction of both proteins in the dimer is too strong. In addition, precipitation was noticeable after incubation, presumably of the entire Cas5fv-Cas7fv dimer. This is commonly observed for the Cas5fv-Cas7fv dimer and indicates it could not bind to RNA which would have stabilized the protein.

As an alternative, we decided to use Cas7fv alone for *in vitro* RNA wrapping instead of the Cas5fv-Cas7fv dimer. The first investigated way to produce Cas7fv was to repeat the crystallisation of I-Fv Cascade which has been shown to produce Cas7fv crystals as a by-product under some conditions. Crystals were created in drop format by addition of a screening solution and various concentrations of purified I-Fv Cascade. After overnight incubation at 18 °C, crystals were visible under light microscopy. To obtain sufficient amounts, we attempted to switch from the drop format for crystallization to crystallization in

a batch format (i.e. a micro-tube). Under all tested conditions, including those that worked for the drop format, protein precipitated overnight and no crystals were obtained.

Because high-yield crystallization proved to be problematic and beyond the scope of this work, we purified monomeric RNA-free Cas7fv fused to a SUMO-tag for solubility following a previously established protocol that includes a high salt wash step to remove all bound RNA during Ni-NTA. Afterwards, monomeric Cas7fv was separated by size-exclusion chromatography and incubated with RNA to form complexes (Figure 2.13 A). SUMO protease was added in the mixture to remove the SUMO-tag and to exclude possible steric clashes during complex formation. We performed this experiment with both small RNAs and *in vitro* transcribed full length *sfgfp* RNA (Figure 2.13 B). Unfortunately, a second size-exclusion chromatography purification after incubation with RNA did not result in an additional peak at an earlier elution volume (Figure 2.13 D).

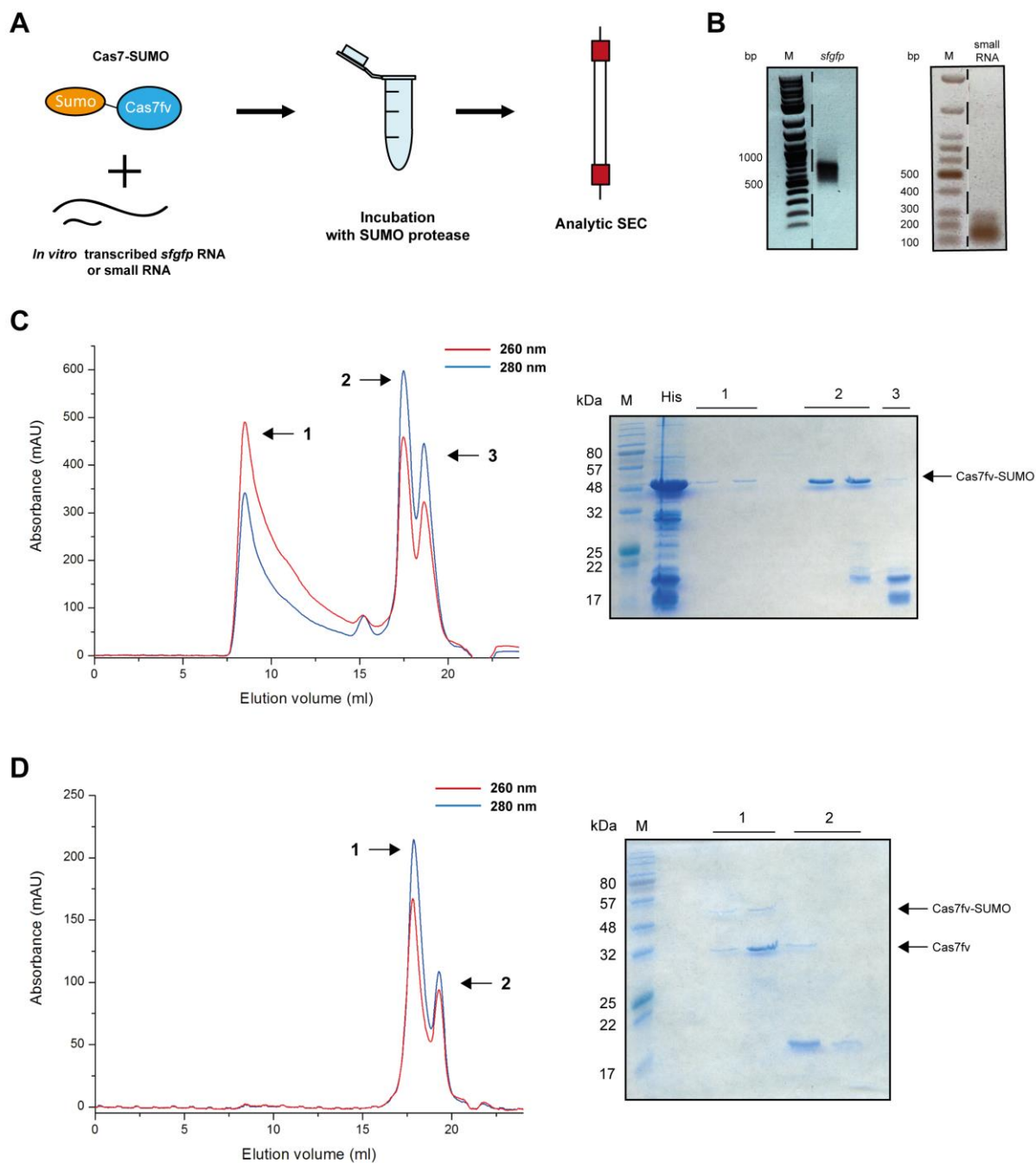


Figure 2.13: *In vitro* RNA wrapping with SUMO-tagged Cas7fv. (A) Schematic principle of the experiment. Purified SUMO-Cas7fv is incubated with either small RNA or *in vitro* transcribed *sfgfp* as well as SUMO protease for 1h at RT followed by overnight at 4°C. The sample is then separated by analytic size-exclusion chromatography. (B) Gel electrophoresis of either *in vitro* produced RNA or extracted small RNA used for experiments. (C) Purification of SUMO-Cas7fv by size-exclusion chromatography (left) and analysis by SDS-PAGE (right). SUMO-Cas7fv was eluted at the corresponding elution volume and showed a matching signal on SDS-PAGE, although some partially degraded SUMO-Cas7fv is present. (D) UV chromatogram of an analytic size-exclusion chromatography (left) and analysis by SDS-PAGE (right) of incubated protein and the extracted small RNA. Complex assembly was not detected, considering the same peaks of SUMO-Cas7fv are present.

Some aggregated SUMO-Cas7fv, presumably still bound to nucleic acids, elutes in the void volume of the first size-exclusion chromatography step. This is supported by the high 260 nm UV absorbance in this peak and a faint signal matching the size of Cas7fv during SDS-PAGE. Monomeric SUMO-Cas7fv elutes

either intact in peak 2 or mostly degraded in peak 3. The bands of these smaller products are also visible in the sample from the Ni-NTA step before size-exclusion chromatography and could be due to either unspecific interactions or by fragments created by the 1M NaCl wash step. The UV absorbance ratio of 280 nm compared to 260 nm indicates that mostly protein is eluted in this peak and RNA was indeed removed. In accordance with this, no RNA was detected by Urea-PAGE. All fractions of peak 2 and 3 with a signal of SUMO-Cas7fv were concentrated and mixed with the small RNA (3.5 µg of small RNA and 1 mg of SUMO-Cas7fv corresponding to a 160x molar excess of protein).

In vitro assembly and RNA wrapping by this method was apparently not successful and thus the same peaks are present in the UV chromatogram of the second size-exclusion purification step. SUMO protease treatment was partially effective considering that the intensity of the band of SUMO-Cas7fv was drastically reduced and a band at the size of monomeric Cas7fv was present instead. Due to the small size of the SUMO-tag, a shifted elution volume in this size-range is not distinguishable. The same result was obtained with the *in vitro* transcribed *sfgfp* RNA.

Overall, no method proved successful for *in vitro* RNA wrapping. The most promising method seems to be the purification of SUMO-tagged Cas7fv. While good amounts of monomeric Cas7fv were produced via SUMO-tag and the high salt wash removed all bound RNA, *in vitro* RNA binding was not observed. Complex formation might need to occur much faster as is the case for *in vivo* conditions. It could also be possible that still not enough RNA was provided (even though multiple µg should be in the detectable range).

2.2.2 Directed *in vivo* RNA wrapping by I-Fv CRISPR-Cas repeat sequences

As an alternative approach to *in vitro* wrapping of RNA, we attempted to create complexes *in vivo* by mimicking the natural Cascade assembly in the cell. During Cascade assembly, Cas5fv and Cas6f serve as roadblocks to limit RNA binding and to form a stable and specific Cascade complex. This assembly is likely initiated by crRNA-processing of Cas6f, which makes the 5'-handle available for binding of Cas5fv. The binding of this protein possibly initiates the rapid backbone formation of the Cascade complex instead of minor unspecific interaction on random RNA.

To investigate if this process works *in vivo*, we fused the 8 nt repeat sequence of the I-Fv CRISPR-Cas system upstream of a ribosome-binding site and the sequence of a reporter gene. Producing the repeat-tagged transcript in combination with the Cas proteins should lead to the initial stages of Cascade assembly and specific RNA wrapping by Cas7fv. A schematic representation of this process is shown in Figure 2.14.

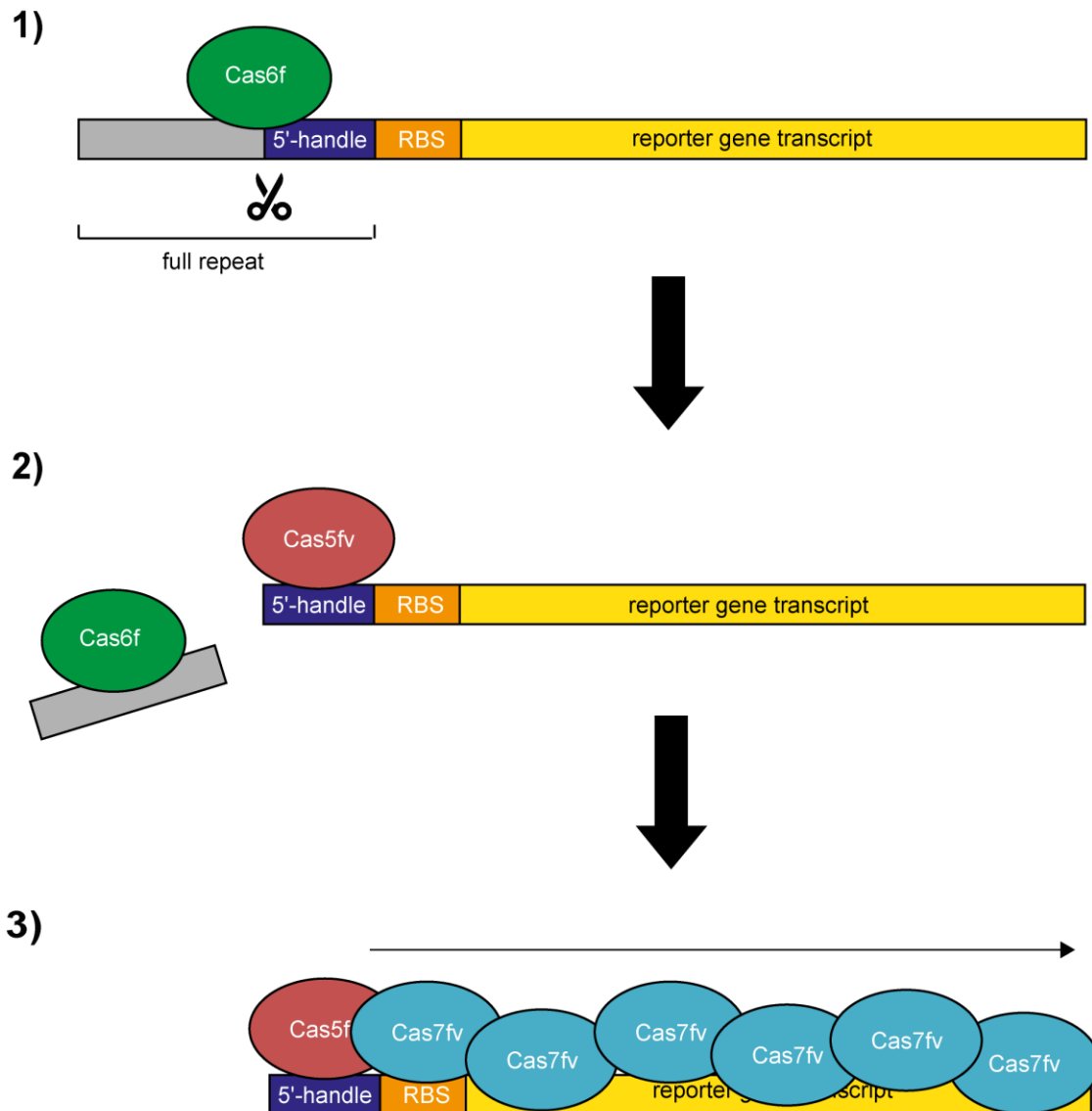


Figure 2.14: Schematic representation of directed RNA wrapping by I-Fv repeat sequences and Cas proteins. (1) Targeted RNA wrapping starts with crRNA-processing as is the case in the assembly of type I-Fv Cas cascade. Cas6f produces the 5'-handle by cleaving the full repeat sequence. **(2)** Cas5fv interacts with the 5'-handle which leads to **(3)** the subsequent backbone formation on the following sequence (in this case the sequence of a reporter gene instead of a crRNA spacer sequence). During Cas cascade assembly, this process is stopped by Cas6f bound to the 3'-hairpin, serving as a roadblock for backbone formation. Without this, long helical structures, possibly covering the entire transcript, are formed.

2.2.2.1 Purification of directed RNA wrapping complexes

To test if the hypothetical process works, we first attempted to purify the formed complexes and study the wrapped RNA for specificity. For the first experiments, superfolder GFP (sfGFP) was chosen as a reporter construct. In the experimental setup, the sequence of sfGFP tagged with a repeat sequence was cloned in the pBAD vector under control of the arabinose promoter. Thus, independent expression of the *cas* genes under control of the T7 promoter is possible. To study the specificity of this RNA wrapping process, a control construct was utilized that contained only *sfgfp* and the necessary RBS sequence but no upstream repeat sequence.

Both the transcript and Cas proteins were produced in *E. coli* expression cultures by overnight growth at 18 °C after induction at ~ OD_{600nm} 0.6. The next day, cells were lysed and proteins were purified via Ni-NTA utilizing the His-tag on Cas5fv. Due to an additional His-tag on sfGFP, this protein was co-purified from both cultures. Production of sfGFP was already visibly reduced in case of the repeat-tagged construct as indicated by the color of the cell pellet when compared to the control construct. Purified complexes were then analysed by SDS-PAGE, which visualized the Cas proteins and sfGFP (Figure 2.15 A & B). Afterwards, protein samples were pooled and RNA was extracted via phenol/chloroform followed by ethanol precipitation (Figure 2.15 C).

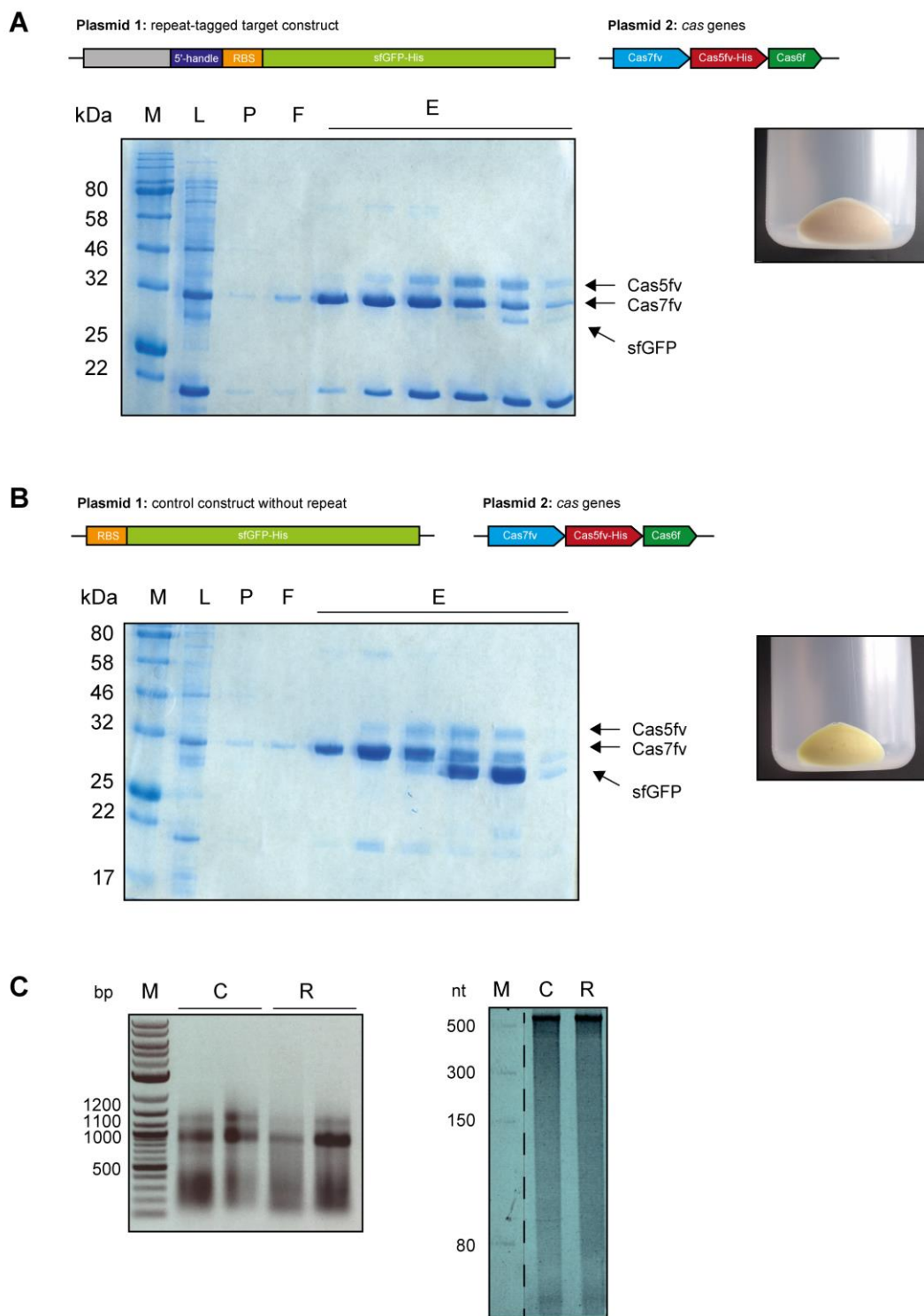


Figure 2.15: Ni-NTA purification of *sfGFP* constructs wrapped by Cas proteins. Purification and SDS-PAGE analysis of the repeat-tagged construct (**A**) and the control construct (**B**). Schematic representation of used plasmids (top): The sequence of *sfGFP* and the RBS was either tagged with a I-Fvr repeat sequences or not tagged. In both constructs, a C-terminal His-tag was fused to the sequence of *sfGFP* for co-purification. Plasmid 2 used in both setups contains the *cas* genes, with a His-tag fused to the C-terminus of Cas5fv. Cell pellets of harvested expression cultures (right side) show a clear color difference due to reduced *sfGFP* production for the repeat-tagged construct. SDS-PAGE analysis (bottom) shows the presence of all three Cas proteins as well as *sfGFP*. (**C**) RNA was extracted from Ni-NTA purified samples of the control (C) and the repeat construct (R) and visualized by gel electrophoresis. Left: separation on 1% agarose 1xTBE gel and ethidium bromide staining, Right: separation on 10% PAA-gel with 8 M Urea and SYBR-Gold staining.

Cas proteins were purified in both cases, showing a prominent band corresponding to Cas7fv, indicating the presence of numerous subunits of this protein. A weaker band corresponding to Cas5fv was purified as well, most likely by the one subunit starting the initiation of RNA wrapping as well as Cas5fv-Cas7fv dimer by-products which could be removed via size-exclusion chromatography.

Due to the His-tag on sfGFP, this protein was co-purified in both cases. SDS-PAGE analysis showed a significantly more intense sfGFP band for the control than for the repeat construct. This reduced sfGFP production by blocking translation on the transcript serves as another indication for specific complex formation on the tagged RNA construct. The extracted RNA from both the repeat as well as the control construct show multiple bands as well as a smear in both cases. Two clear bands with an approximate size corresponding to 1200 and 1000 bp of the dsDNA ladder are visible as well as an accumulation of signals below the 500 bp band of the ladder. An entire repeat-tagged *sfgfp* transcript would have a length of 800 nt (or 1000 nt including the transcription terminator). This estimated length does not fit completely to the observed. It should be noted, however, that the marker used for size comparison is dsDNA and single-stranded RNA runs usually lower on these gels. The extracted RNA was closely examined for small sizes, by loading the sample on a high percentage and denaturing PAA-gel (10%, 8M Urea) which shows a consistent smear due to the large size of the extracted RNA. The two distinct bands at a higher position were later identified as rRNA (see section 2.2.2.5).

2.2.2.2 RNA-seq analysis reveals specificity of RNA wrapping

To analyse the specificity of the RNA wrapping process, we subjected the extracted RNA of the purified complexes to Next-Generation Sequencing by Illumina. Prior to this, the RNA samples were treated with DNase I to remove potential DNA contamination. The extracted RNA was then fragmented by 10 min incubation at 95 °C and addition of 100 mM ZnCl₂ and libraries were created for Illumina sequencing with the NEBNext Small RNA Library Prep Set.

The obtained reads were then mapped against the genome of *E. coli* BL21 (DE3) as well as the pBAD plasmid harbouring the sequence of the repeat-tagged *sfgfp* (Figure 2.16 A). The mapped reads were almost exclusively mapping to *sfgfp* as compared to other parts of the plasmid and the host genome, confirming the specificity of this reaction. Reads originating from the sequenced RNA started directly with the Cas5fv-binding site upstream of the RBS followed by *sfgfp*, resulting in a peak in the mapping profile with a maximum of ~ 800,000 reads from a total of ~ 1,300,000 reads. However, not the entire sequence of *sfgfp* is covered as the peak has a clear edge and significantly fewer reads are present after ~ 100 nt of the *sfgfp* sequence.

RNA molecules extracted from the control construct (without the repeat-tag upstream of *sfgfp*) were analysed with RNA-seq in parallel (Figure 2.16 B). In this case, minor amounts of reads (a maximum of

25,000 reads compared to 4 million reads in total) could be mapped over the entire genome and some parts of the plasmid. The mapping profile thus resembles an overall transcriptome representation, with the highest coverage originating from transcripts of highly expressed genes.

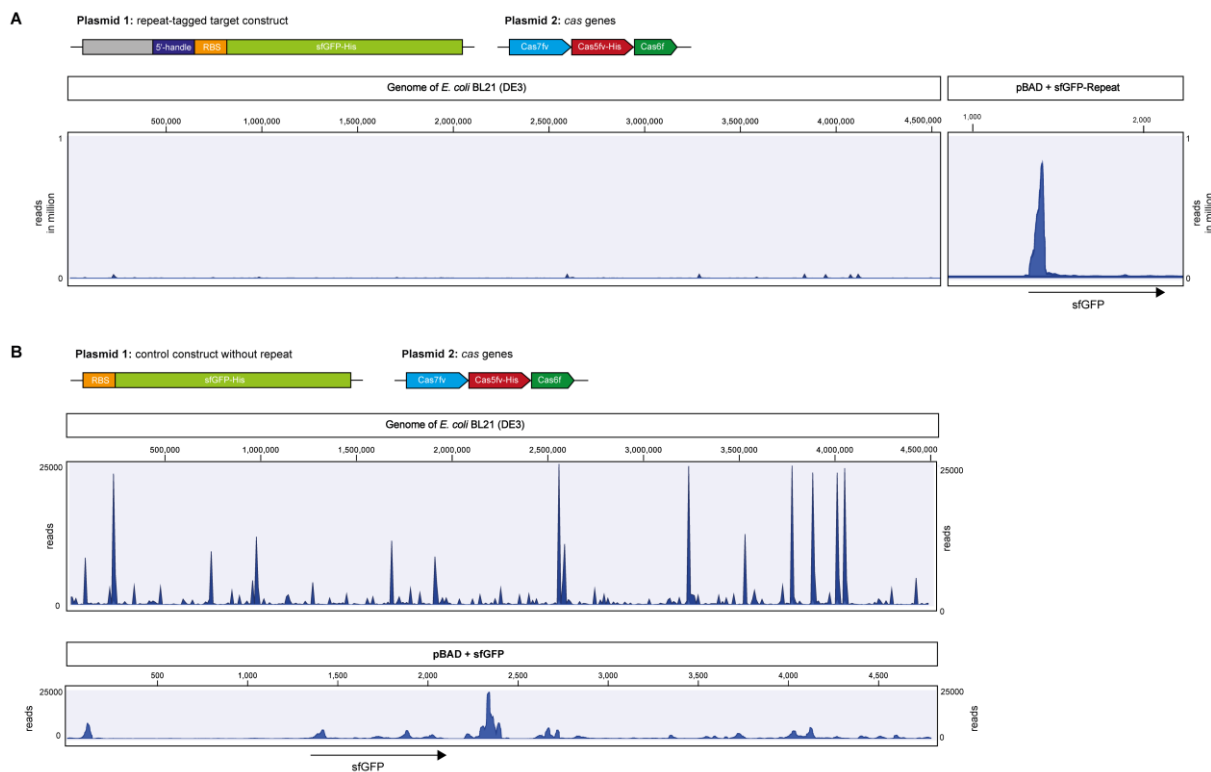


Figure 2.16: RNA-seq analysis of RNA wrapping of sfGFP-Repeat and sfGFP-Control. (A) RNA wrapping of sfGFP-Repeat with a schematic representation of the expressed constructs from both plasmids (top). The obtained reads from RNA-seq were mapped against the genome of *E. coli* BL21 (DE3) as well as the plasmid with the sfGFP-Repeat sequence (bottom). The reads were mapped almost exclusively to the start of the sequence of sfGFP, confirming specificity. Only the first part of the sequence of sfGFP is covered (arrow). **(B)** RNA wrapping on sfGFP-Control with a schematic representation of the expressed constructs from both plasmids (top). Mapping of the obtained reads from RNA-Seq against the genome of *E. coli* BL21 (DE3) (middle) and the plasmid with the sfGFP sequence (bottom). Matching reads were detected from all over the genome and the plasmid.

Overall, the RNA-seq analysis and the vast overrepresentation of reads mapping to the tagged *sfgfp* shows that the wrapping process can be directed by utilizing a I-Fv repeat sequence. However, it was apparently not possible to wrap and extract an entire *sfgfp* transcript as only minor amounts of reads are present beyond approximately 100 nt. These results are in contrast to the clearly visible two bands on agarose gel electrophoresis indicating the presence of larger nucleic acids.

To investigate why not the entire sequence of sfGFP is wrapped and to prove the applicability of this process on different kinds of reporter genes, we created a construct with a I-Fv repeat tagged gene encoding the lacZ- α subunit. Directed RNA wrapping complexes were produced and purified as before and the extracted RNA was sequenced. SDS-PAGE analysis of the purification and the subsequent RNA

extracted looked similar to sfGFP-Repeat (Supplementary Figure 2). The obtained reads were mapped against the host genome and the lacZ-Repeat plasmid (Figure 2.17). Similar to RNA wrapping of the sfGFP-Repeat construct, a significant peak in the mapping profile was only present for the initial sequence of lacZ- α . This peak starts again with the repeat sequence upstream of lacZ- α but falls off drastically after ~ 100 nt. A smaller additional peak is present after this, but no significant amount of reads could be mapped after ~ 150 nt of the lacZ- α sequence.

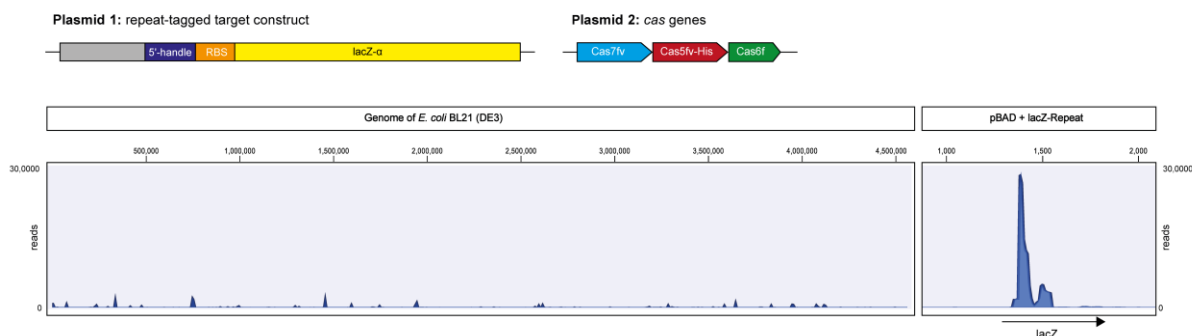


Figure 2.17: RNA wrapping of lacZ-Repeat constructs. RNA wrapping of the lacZ-Repeat construct with a schematic representation of the expressed constructs from both plasmids (top). The obtained reads from RNA-seq were mapped against the genome of *E. coli* BL21 (DE3) and the plasmid with the lacZ-Repeat sequence (bottom). The vast majority of reads were mapped to the first part of the sequence of lacZ (arrow).

The similar limits of read coverage for both investigated reporter genes speak against intrinsic factors, such as secondary structures, that might terminate the wrapping process. The apparent limit of RNA stabilization is thus more likely a limitation of the wrapping process itself. This is plausible, considering no regions leading to strong secondary structures were found on either sfGFP or lacZ- α sequences. Nonetheless, this result confirms the possibility to wrap the initial 5'-terminal region of a tagged RNA.

2.2.2.3 Structural analysis of filament structures

To analyse the structure of the purified complexes and to compare them to the previously observed Cas7fv filaments, we attempted to visualize these structures by transmission electron microscopy (TEM). To do this, the Ni-NTA purified complexes with the sfGFP-repeat construct were further purified by size-exclusion chromatography (Figure 2.18 A). The UV chromatogram shows two major peaks with the first one in the void volume of the column and the latter at the position of the Cas5fv-Cas7fv dimers. The protein content of these peaks was visualized by SDS-PAGE and revealed a band corresponding to Cas7fv while Cas6f or Cas5fv were not observed. Cas5fv is perhaps not visible due to its underrepresentation compared to Cas7fv considering that only one subunit is necessary for initiation. Alternatively, an overlap with the band of Cas7fv is often observed during SDS-PAGE. A complex of the

first 100 nt transcript plus approximately 16 subunits of Cas7fv (1 subunit per 6 nt) would be eluted in the void volume of the HiLoad Superdex 200 SEC column that was used for this experiment.

Sample fractions of the void volume of size-exclusion purification were then handed to analysis by transmission electron microscopy (TEM) by Dr. Thomas Heimerl. Samples contained long helical filaments with a size of approximately 100-200 nm with a turn at every ~ 10 nm. (Figure 2.18 B). Similar structures were also purified from the lacZ-Repeat construct (Supplementary Figure 3).

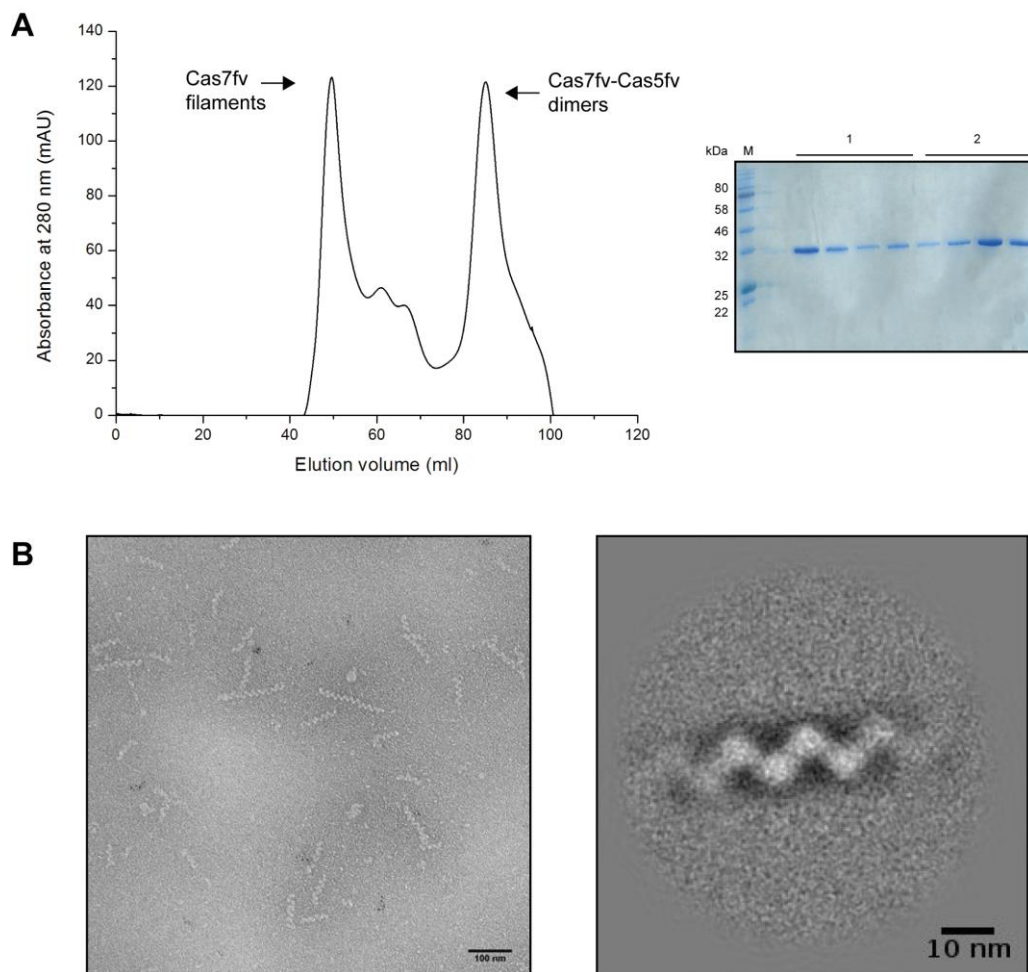


Figure 2.18: Structural analysis of filamentous Cas7fv structures with sfGFP-repeat constructs. (A) UV chromatogram of size-exclusion chromatography (left) and SDS-PAGE analysis (right). The UV chromatogram reveals two major peaks, in the void volume and at the elution volume of the Cas 5fv-Cas 7fv dimer that both show a distinct band of Cas 7fv on SDS-PAGE. **(B)** Transmission electron microscopy of peak sample in the void volume (right) and 2D class averaging of filamentous complexes.

This indicates that the formed structures in the expression cultures are similar to the observed filament by-products of Cascade purification. It should be noted that these filaments seem smaller compared to the sometimes extremely long filaments observed as by-products during earlier Cascade purification that exhibited a size of multiple hundreds of nm (Dwarkanath *et al.*, 2015).

While attempting to solve the 3D structure of I-Fv Cascade, Dr. Patrick Pausch was able to crystallize and solve the structure of Cas7fv. These crystals were produced as fragments during crystallization of I-Fv Cascade and did not contain RNA. In this crystal structure, the Cas7fv molecules form a helix similar to the observed filaments (Figure 2.19). This 3D structure highlights the helical nature of the filament in which 8 Cas7fv subunits are required for one full rotation. Surface charge visualization highlights the positively charged wrist helix on the side of the structure (Figure 2.19 B, in blue).

A model of a segment of helices seen by TEM was also created by 2D class averaging (Supplementary Figure 4) that matches the 3D structure.

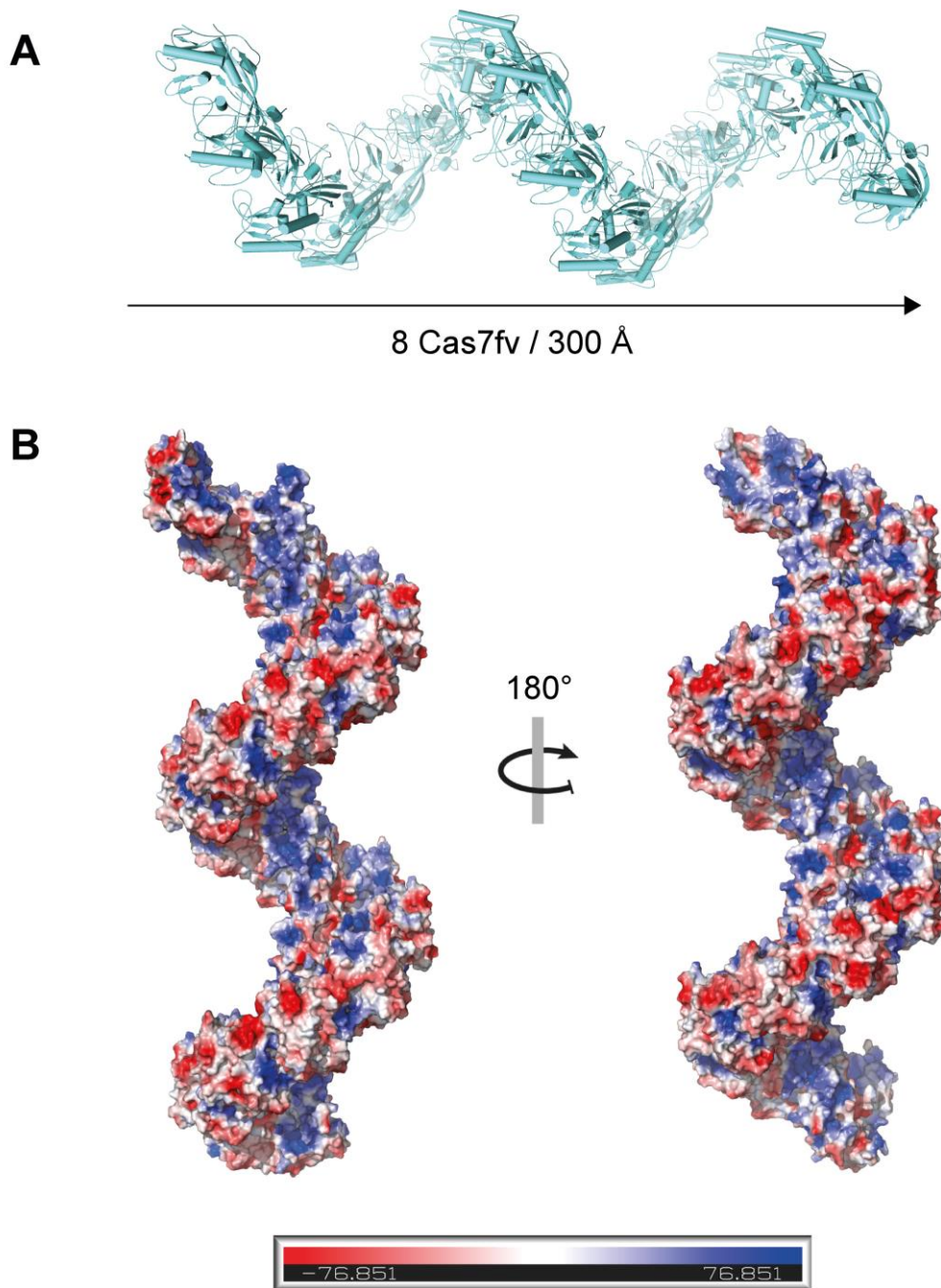


Figure 2.19: 3D structure of Cas7fv filaments in (A) Ribbon cartoon representation and **(B)** surface charge representation with color according to electric charge (electrostatic surface potential calculated in arbitrary units). The structure was created by elongation of consecutive Cas7fv dimer subunits. 8 Cas7fv subunits are required for one full rotation over a distance of 300 Å.

With the structural data, a calculated estimation can be made about how much of the transcript would be incorporated in the filaments. A full rotation consisting of 8 Cas7fv subunits has a length of ~ 30 nm (300 Å) which equals 3.75 nm per bound Cas7fv subunit. Each Cas7fv subunit binds 6 nt in the crystal structure of I-Fv Cascade, which calculates as ~ 0.625 nm/nt. If the ~ 800 nt long *sfgfp* transcript is completely covered by subunits of Cas7fv, this would generate a length of up to 550 nm which is longer

than the observed structures ranging from approximately 100 to 200 nm. If smaller fragments of the targeted RNA are wrapped by Cas proteins, such as the ~100 nt 5'-terminal portion that is overrepresented in RNA-seq, they could fit inside the observed filament structures. By this calculation, a 100 nt RNA wrapped by the appropriate number of Cas7fv subunits would have a length of only 62.5 nm. Most observed structures are 100-200nm in size and by calculation, these should encase 160-320 nt. While some smaller fragments are indeed visible on TEM, they are almost all longer than 60 nm. It remains to be investigated how much RNA is indeed bound by Cas7fv in these structures.

2.2.2.4 Attempts to increase the length of wrapped RNA

2.2.2.4.1 Attempted purification of complexes on entire reporter gene transcripts with two repeat sequences

While investigating the extent of complex formation and the maximum length of RNA wrapping, an alternative approach was pursued that involved putting a second repeat region at the 3'-end of the transcript. After processing, Cas6f would bind this region as it does in the natural Cascade complex. The created structure would resemble a massively elongated Cascade complex, with the reporter gene transcript as its crRNA. By putting a second affinity tag (e.g. a Strep-tag) on Cas6f, it would be possible to separate this complex wrapping the entire RNA molecule with two adjacent purification steps (Figure 2.20 A). The plasmid for *cas* gene expression was co-produced with the repeat-tagged *sfgfp* construct in *E. coli* in the same fashion as for previous cultures. The produced proteins were purified by Strep-tag affinity and the elution fractions were analysed by SDS-PAGE (Figure 2.20 B). Only Cas6f could be purified in this purification step as visible on the gel. Bands of Cas5fv and Cas7fv could be detected in the flow-through of the purification. This result further indicates that the complex is not formed on the entire transcript, as indicated by previous experiments.

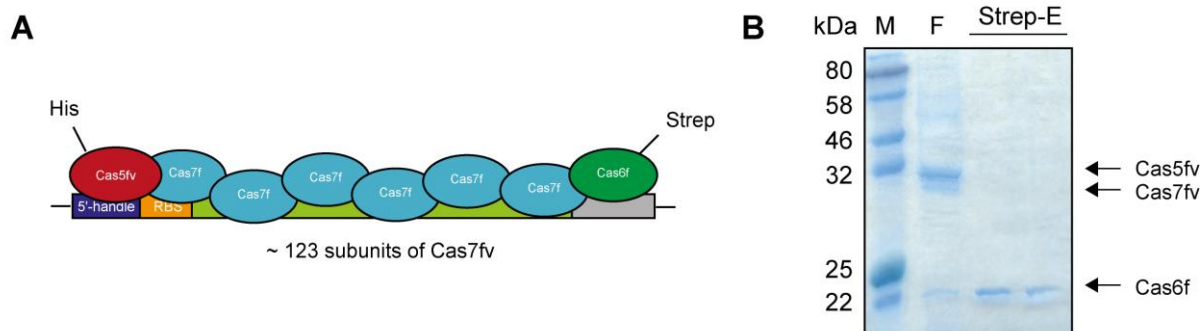


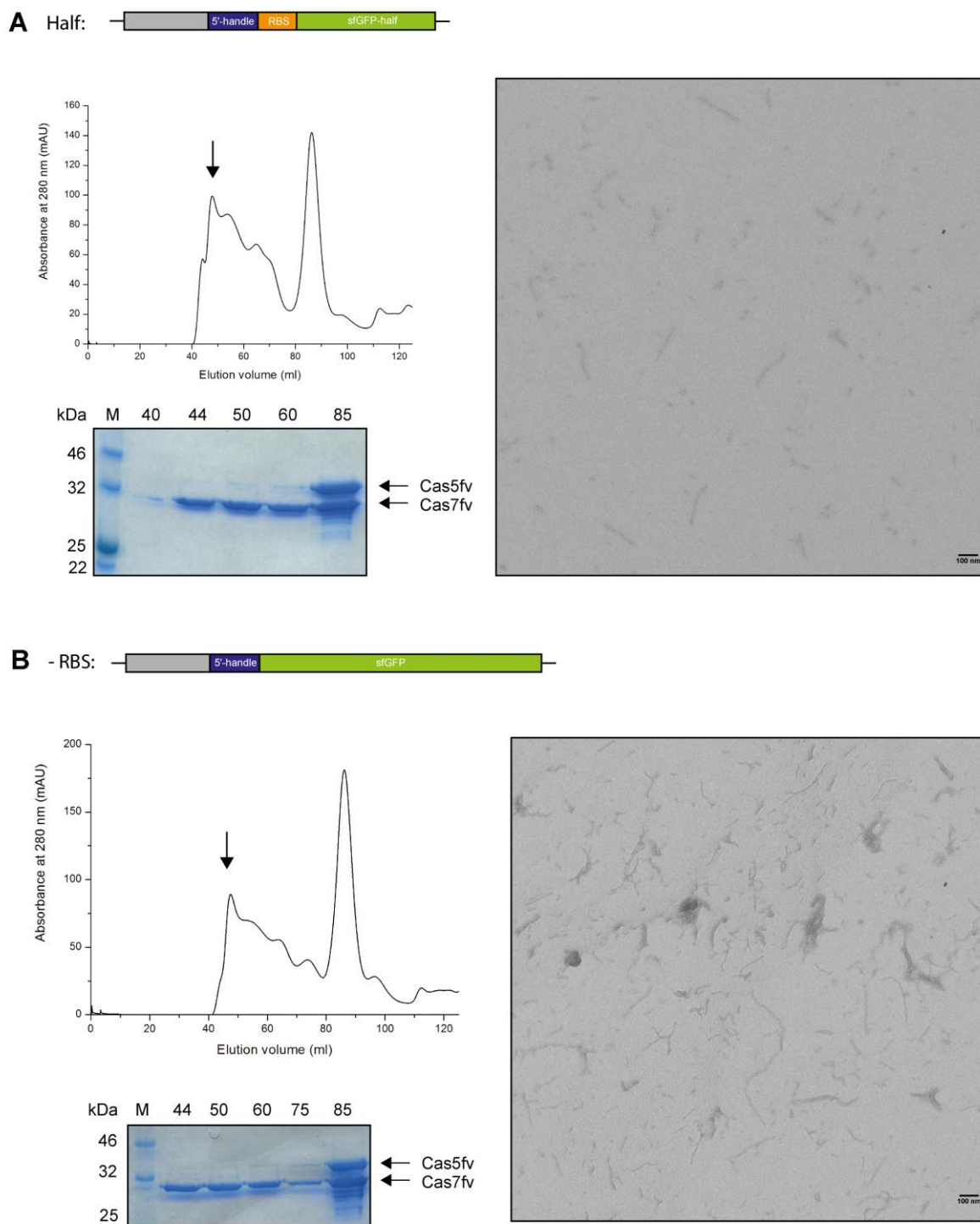
Figure 2.20: RNA wrapping of entire reporter gene transcripts flanked by two repeat sequences. (A) Schematic principle of RNA wrapping including a second repeat sequence at the 3'-end of the reporter gene. Cas6fv binds the repeat sequence after processing, as is the case in a natural crRNA. By fusing a C-terminal Strep-tag with Cas6fv in addition to the His-tag on Cas5fv, the formed structure could be purified with two adjacent affinity purification steps. The number of Cas7fv subunits on the backbone is simplified for visualization and not representative of the number of subunits required (~ 123). **(B)** SDS-PAGE analysis after Strep-tag affinity purification. Only Cas6fv can be observed in elution samples (Strep-E), while Cas7fv can be observed in the flow-through (F).

2.2.2.4.2 Investigation of expression condition factors for improved RNA wrapping

All results so far indicated that filament structures can be specifically formed on a repeat-tagged RNA. We presumed that this RNA is somehow located in the observed filaments. While RNA-seq results seemed to confirm that only the initial 5'-terminal portion of the tagged transcript can be wrapped and isolated, additional rRNA was always present in the RNA extractions and it remained unclear where this RNA is located or more specifically if it is also located in the filaments.

To study if the length of the filaments changes on different reporter constructs, we created an sfGFP-repeat construct in which the second half of the coding sequence was removed (sfGFP-Half). The transcribed RNA up to the transcription terminator should be thus approximately half the size of the normal sfGFP-Repeat transcript. We then purified protein complexes wrapping this RNA and analysed the void volume fractions from size-exclusion chromatography with TEM to see if the typical 100-200 nm filaments are still created or if they are decreased in size (Figure 2.21 A).

Another possible explanation for the incomplete wrapping was that translation could impact the wrapping process. The limited wrapping would be dependent on the time of Cas7fv production and the wrapping of ~ 100 nt might reflect the distance the RNA polymerase allows for binding. To disable translation on the produced transcript, we removed the RBS from the sfGFP-repeat construct (-RBS). No ribosome binding and translation should be possible on these transcripts. Protein complexes wrapping this RNA construct were purified in the same way and void volume fractions were analysed by TEM (Figure 2.21 B).



Both purifications delivered a similar UV chromatogram with a very broad peak starting in the void volume that contains mostly Cas7fv as well as some minor amount of Cas5fv. A second and more distinct peak of the Cas5fv-Cas7fv dimer is present as well.

Fractions of the void volume were analysed with TEM and both fractions contained the observed filament structures. The structures produced on the GFP-half construct had a similar length to the previously purified filaments, so the length of the construct did not influence the length of the filament structures. This further confirms the RNA-seq results which show only the first ~ 100 nt are wrapped. On the GFP construct without RBS, some structures with a size of more than 200 nm are visible, with some going up to ~ 500 and even 800 nm. However, it is not clear if these structures are formed on multiple RNA molecules or if they are overlapping smaller filaments. Additionally, most structures still show the usual size of ~ 200 nm. Nonetheless, this could indicate an effect of translation on RNA wrapping. RNA extracted from the void volume fractions was analysed via Urea-PAGE and showed a smear through the entire lane of the gel.

So far, for all expression cultures, the three *cas* genes were produced in equal amounts regardless of the massive overrepresentation of Cas7fv in a potential filament structure that wraps an entire *sfgfp* transcript (>100 Cas7fv subunits compared to one subunit of Cas5fv). Even when only the first 100 nt of the 5'-terminal portion is covered, this would require at least 16 Cas7fv subunits compared to one Cas5fv. To address this, we recloned *cas5fv* and *cas6f* and placed them on another plasmid with a lower copy number (~20 copies of pACYC compared to >100 of pRSF). Additionally, we created one plasmid with one *cas7fv* in each of the two multiple cloning sites to increase the production of Cas7fv. We produced these new constructs in *E. coli* expression cultures with initial expression conditions and overnight incubation at 18 °C after induction. Complexes were purified as usual via the His-tag on Cas5fv followed by RNA extraction. The extracted RNA was fragmented and RNA-seq analysis by Illumina was performed (Figure 2.22). For both constructs, again, only the initial 5'-terminal portion could be detected.

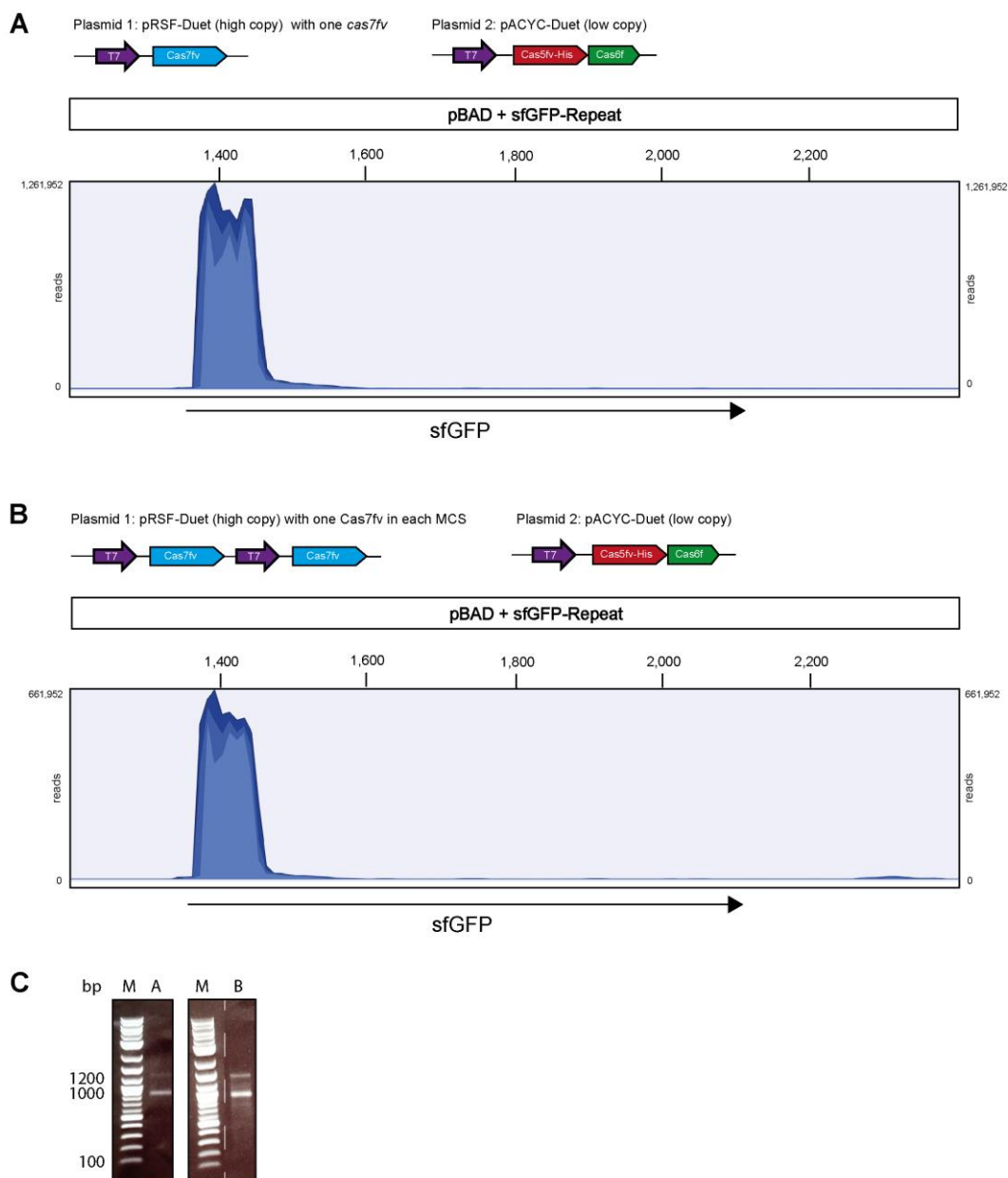


Figure 2.22: RNA-seq analysis of RNA wrapping with *Cas7fv* overproduction compared to *Cas5fv* and *Cas6f*. *Cas7fv* was either expressed from one MCS (**A**) or by one copy in each of the two multiply cloning of the high-copy plasmid (**B**). *Cas5fv* and *cas6f* were expressed from the low copy plasmid. Schematic representations in the top show the plasmids used for *cas* gene expression. Mapping of the obtained reads to the target plasmid reveals an overabundance of reads matching the first ~100 nt of the repeat-tagged *sfGFP*. (**C**) RNA extraction on separated on an agarose gel electrophoresis shows the typical bands of rRNA for both constructs.

To confirm the previous TEM analysis of complexes on repeat-tagged RNA without ribosome binding site, we performed RNA-seq. For this, we created a construct where we not only removed the RBS between the repeat region and the reporter gene but also exchanged the reporter gene itself with a sequence containing no start-codon. On this non-coding construct, absolutely no translation should be possible that could interfere with the wrapping process. Considering the newly used target sequence, we also created a control construct where the RBS was re-added upstream of the sequence as well as an

ATG start codon. Expression cultures for both constructs were grown as before with the *cas* genes split into two plasmids and proteins were purified via His-tagged Cas5fv. The extracted RNA was then sequenced with Illumina (Figure 2.23).

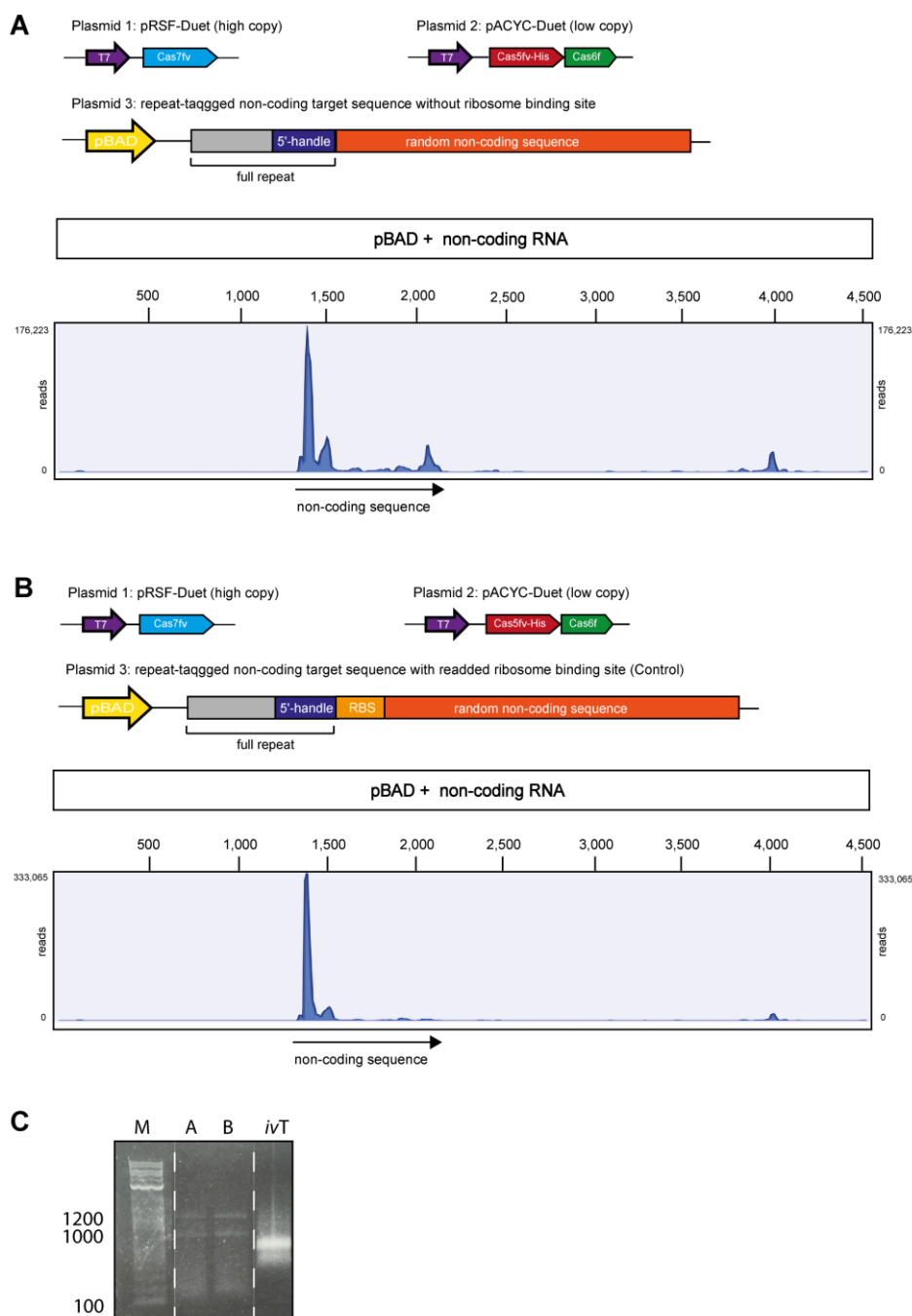


Figure 2.23: RNA-seq analysis of RNA wrapping on a repeat-tagged non-coding construct and the control construct. (A) Schematic representation of the used plasmid for RNA wrapping of the repeat-tagged non-coding construct (top) and mapping of the obtained reads from to the sequence of the target plasmid (bottom). Similar to previous experiments, the majority of reads were mapped to the initial 5'-terminal portion of the tagged construct. **(B)** Schematic representation of the used plasmid for RNA wrapping of the repeat-tagged control construct (top) and mapping of the obtained reads from to the sequence of the target plasmid (bottom). **(C)** RNA extraction on separated on a agarose gel electrophoresis shows the typical band pattern including rRNA for both the non-coding as well as the control construct. A control sample of *in vitro* produced *sfGFP* RNA was loaded for size differentiation (ivT).

Overall, the RNA-seq results show no drastic difference to the previous experiments. Again, only the initial 5'-terminal portion is detected with significant reads resulting in a clear peak. A smaller peak is visible adjacent to the first one with a maximum read count of ~40,000 reads for the non-coding sequence and covering a sequence of up to 200 nt. For the complete non-coding construct, a higher baseline of reads mapped to the rest of the sequence can be seen that reaches the end of the transcription terminator, but the overall read count in this area is much lower than for the initial peak. This could be explained by an increased presence of transcript due to it not being converted to protein and is either wrapped or somehow co-purified with actually wrapped RNA. The few larger filament structures observed with TEM for the GFP-construct without RBS could be related to this.

2.2.2.5 Identification of co-purified ribosomal RNA by Nanopore Sequencing

The second major problem with the wrapping process so far was the inconsistency of the observed band pattern of extracted RNA and the RNA-seq results detecting only a small RNA. To identify what the higher sized extracted RNA really is, the newly extracted RNA samples were analysed by Nanopore sequencing. This technique allows the sequencing of full strands without fragmentation. The obtained reads were mapped against the sequence of the *sfGFP* target plasmid (Figure 2.24 B). Unmapped reads were mapped against the genome of the *E. coli* expression strain (Figure 2.24 C). This mapping revealed that about 80% of the total reads belong to 16S rRNA. Upon closer inspection, a very minor amount of reads (~30) were mapped to 23S rRNA.

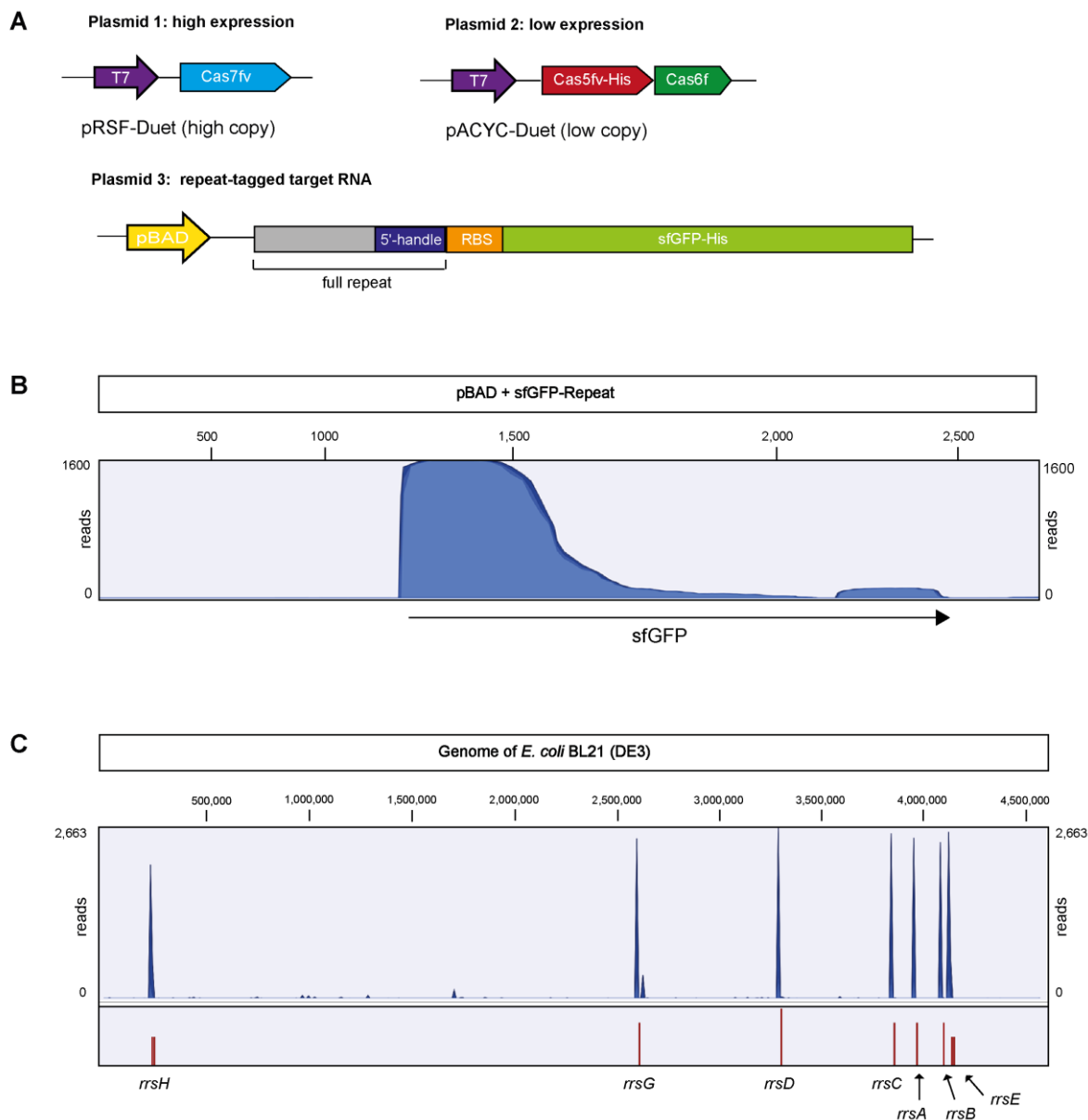


Figure 2.24: Nanopore-sequencing of extracted RNA. (A) Schematic representation of the plasmids used for expression cultures prior to Ni-NTA purification and RNA extraction. **(B)** Mapping of the obtained reads to the sequence of the target plasmid reveals an overabundance of reads matching the first part of the target sequence. **(C)** Mapping of the obtained reads to the genome of the *E. coli* expression strain reveals that a majority of reads match the genes encoding ribosomal RNA. Reference track with the seven rRNA gene clusters in the bottom.

Mapping of reads to the plasmid containing the sequence of the repeat-tagged *sfGFP* presents the same pattern as seen in previous data. However, fewer reads were obtained (~1,600 reads) due to the generally lower read output of this sequencing compared to Illumina sequencing (only ~114,000 reads). The peak of the 5'-terminal portion in the mapping profile appears to be slightly broader and stretches to up to ~250 nt after the 5'-handle sequence. Few reads are mapped that match the whole *sfGFP* sequence (~150 reads). This is in contrast to the previous Illumina sequencing in which no significant reads were mapped after ~100 nt, possible due to the absence of fragmentation in Nanopore

sequencing. In this case, Nanopore sequencing would suggest that the extracted repeat-tagged RNA is moderately longer than previously thought.

The mapping of reads to the genome of *E. coli* delivered a clear explanation for the unexpected bands close to the 1000 and 1200 bp band of the DNA ladder, seen in all previous RNA extractions on agarose gel electrophoresis (e.g. Figure 2.15). Since the majority of reads correspond to 16S rRNA, the minor amount of purified and extracted repeat-tagged target RNA is less visible when analysed on an agarose gel. In the RNA extraction that was used for this sequencing, mostly the 1000 bp band is visible on gel electrophoresis and only a faint band running close to the 1200 bp ladder band can be seen (see Figure 2.22 C), which explains the presence of more reads matching 16S rRNA compared to 23S rRNA. The reason for this contamination of rRNA remains unclear. Sequence analysis did not show any sequences similar to repeats that could trigger RNA wrapping. Additionally, proper 5'-ends on repeat-similar sequences by Cas6f processing would be necessary for initiation of backbone formation as well.

2.2.2.6 Increased target RNA production by T7 RNA polymerase

At this point, we investigated the possibility that coupling of transcription and translation is the reason for rRNA contamination. In the current set up, the target construct was produced by *E. coli* RNA polymerase (RNAP) under control of the arabinose promoter, contrary to the Cas proteins produced by T7 RNAP. The Cas proteins could potentially “bump” into ribosomes attached to *E. coli* RNAP during the wrapping process which would lead to their co-purification. To counter this, we exchanged the promoter for control of the target construct to a T7 promoter (Figure 2.25 A). Target RNA produced by T7 RNAP would ensure that no coupling of *E. coli* RNAP and rRNA interferes with wrapping of the target construct and leads to co-purification of rRNA.

Complexes were then created as previously described and RNA was extracted from Ni-NTA purified protein samples. Extracted RNA was visualized by agarose gel electrophoresis and compared to the previously purified RNA with the transcript produced by *E. coli* RNA polymerase (Figure 2.25 B). The extracted RNA was sequenced with Illumina and the obtained reads mapped against the sequence of the target plasmid (Figure 2.25 C).

Although rRNA was still co-purified with small RNA produced with T7 RNAP, more full-length transcript was produced.

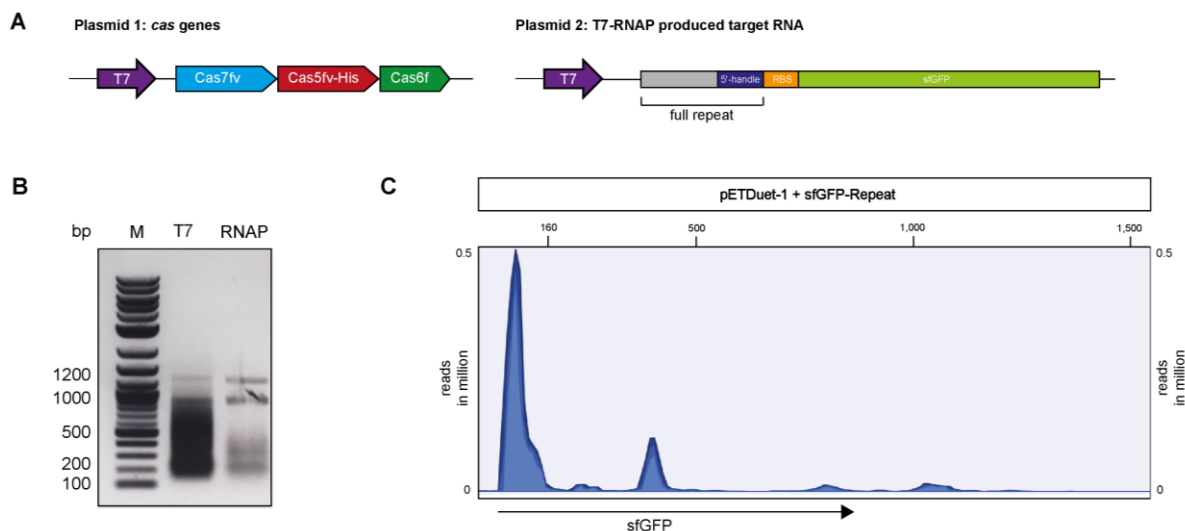


Figure 2.25 RNA wrapping of repeat-tagged RNA produced by T7 RNAP. (A) Schematic representation of the used expression plasmid. The repeat-tagged *sfGFP* construct is expressed from the pETDuet-1 vector under control of a T7 promoter. (B) RNA extraction after Ni-NTA was visualized by a agarose gel electrophoresis. Samples produced with the T7 promoter were compared to previously purified samples where the target construct was purified with *E. coli* RNA polymerase (RNAP). (C) Illumina RNA-seq analysis of the T7 produced target RNA. The obtained reads were mapped against the sequence of the target plasmid containing the repeat-tagged *sfGFP*.

The newest RNA extraction of transcript produced by T7 shows a strong smear below the 1000 bp band of the ladder as well as the usual bands at a position of approximately 1000 and 1200 bp. For previously extracted RNA produced by RNA polymerase, a moderate and smeared signal can be seen from 200 to 400 bp that is similar to early extractions. The strong increase in RNA with a size below 1000 bp indicates a positive effect of increased target transcription rate by T7 RNA polymerase. RNA-seq analysis, again, shows only a strong peak corresponding to the initial 5'-terminal portion with approximately 500,000 reads. For the remaining sequence corresponding to the rest of the transcript, only ~12,000 to 15,000 reads were mapped. One additional peak of up to ~100,000 reads in the middle of the sequence is also visible which might be explained by the “CTTCAAA” sequence at the beginning of this peak, which is partially similar to the “CTTAGAAA” sequence of the 5'-handle. This sequence is only present on the slightly different coding sequence for sfGFP on the new vector used for T7 RNAP target production. Additional binding of Cas5fv might lead to the wrapping of the following ~100 nt, resulting in the observed peak. Perhaps, this is only noticeable due to the higher transcript production by T7 RNAP in this case. In previously used transcripts, we could not find a sequence similar to the 8 nt 5'-handle. Sequences with two nucleotide mismatches did not result in an extra peak. While not necessarily visible by RNA-seq analysis due to the vast overabundance of the 5'-terminal portion, the RNA extraction indicates that T7 RNAP produces more transcripts. However, it is unclear if Cas proteins wrap this RNA. The synthetic RNA wrapping complexes still seem to only contain the first ~100 nt of the repeat-tagged target due to the clear peak on the RNA-seq mapping.

2.2.2.7 Purification of small synthetic Cascade assemblies for RNA wrapping and removal of co-purified rRNAs

When it became clear that the purified samples were contaminated by rRNA, we focused on trying to eliminate this problem. It was also necessary to investigate where exactly the rRNA is located and thereby co-purified. One possibility was that the rRNA is unspecifically wrapped and thus located inside of the complex. The other possibility was that these highly abundant nucleic acids are sticking to the outside of the complexes and are pulled down with them, perhaps by the positively charged wrist helix. For the latter possibility, we tried a variety of methods to clean the wrapping complexes during or after purification (Table 2.1).

Table 2.1: Various methods used for clean-up of purified complexes to remove rRNA.

Clean-up Methods
RNase I _f treatment of protein samples before RNA extraction
Washing samples bound to the Ni-NTA column with a high molarity salt buffer
Purification by Strep-affinity instead of Ni-NTA
Size-exclusion purification
Neutralization of the positively charged wrist helix by amino acid exchange to alanine

None of these methods effectively removed rRNA, considering that after purification and RNA extraction, both corresponding bands were highly visible on gel electrophoresis.

In the end, the contamination problem was apparently solved by simply omitting MgCl₂ from the wash buffer for affinity purification. The repeat-tagged non-coding RNA was produced together with the Cas proteins and the isolated RNA after Ni-NTA and size-exclusion purification contained mostly small RNA around ~100 nt or full transcript that is most likely co-purified (Figure 2.26).

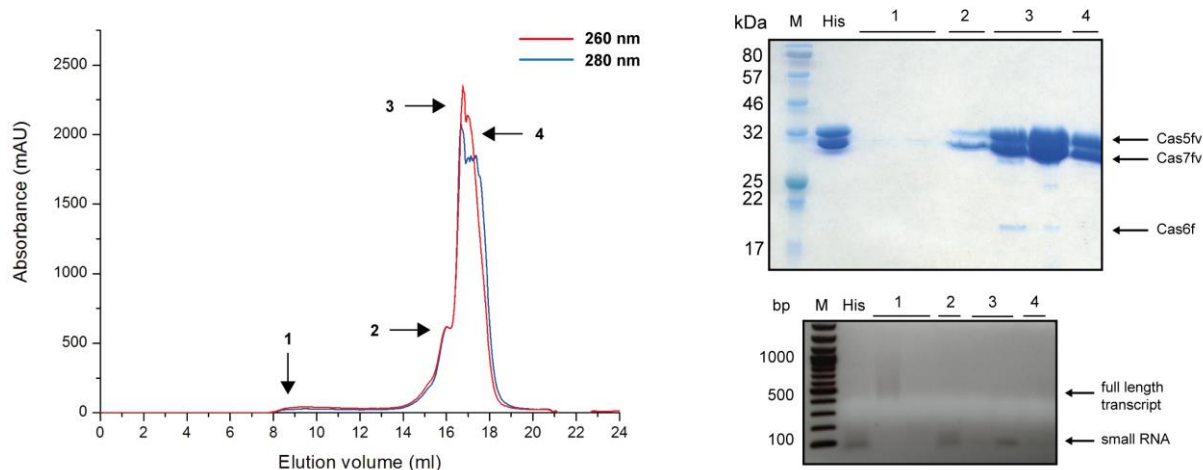


Figure 2.26: Purification of Cas proteins with repeat-tagged non-coding RNA without $MgCl_2$ in the purification buffer. UV chromatogram of size-exclusion after Ni-NTA (left) shows one broad peak and minimal aggregates in the void volume. SDS-PAGE analysis (top right) of a fraction from the void volume (1) as well as various fractions of the broad peak (2-4) show the presence of mostly Cas7fv and Cas5fv. His: Sample of Ni-NTA prior to size-exclusion. Mostly small-sized RNA can be detected by a gel electrophoresis of the extracted RNA from the elution fractions, except for the sample of the void volume where a faint and smeared signal band in the range of full-length transcript can be detected.

Separation of the Ni-NTA purified sample by size-exclusion chromatography features no distinct peak in the void volume contrary to previous purifications. For this purification, we used a newly acquired Superose 6 Increase column capable of separating proteins and complexes up to a molecular weight of 1000 kDa. Compared to the fractionation range of the before used Superdex column of 600 kDa, this means that complexes and aggregates with high molecular weight are more spread out on the UV chromatogram instead of eluting as one single peak in the void volume.

For the very initial eluted fraction (1), only a faint signal of Cas7fv can be seen on SDS-PAGE. However, the minor amount of RNA extracted from this sample seems to match a full-length transcript of the repeat-tagged non-coding RNA. Considering that in the RNA extraction prior to Ni-NTA purification, only the small RNA can be clearly seen, the relative amount of full-length transcript seems to be very minor. This observation fits previous RNA-seq analyses with a massive overabundance of reads for the initial 5'-terminal portion. The remaining purified protein complexes are eluted at a later position in a very broad peak with an estimated molecular weight of ~ 400 kDa to as low as ~ 50 kDa. From a sample on the side of this peak (2), small RNA can be extracted. Protein content and elution volume of this fraction could indicate a smaller complex that consists mostly of Cas7fv and one Cas5fv subunit. A similar small-sized RNA can be extracted from samples eluting later (3). Here, Cas7fv and Cas5fv can be detected in an equal ratio on SDS-PAGE, with some Cas6f pulled down additionally. This by-product could represent something similar to a Cascade complex but without a fully formed backbone. The fraction at the left side of the peak (4) contains only Cas7fv and Cas5fv in an equal ratio but no RNA, which most likely represents the Cas5fv-Cas7fv dimer in accordance to the elution volume.

However, in no RNA extraction sample of this purification, the typical 16S rRNA or sometimes 23S rRNA bands can be observed. The effect of Mg^{2+} ions on structured RNA could have had a stabilizing effect on these and led to their co-purification. It also seems that no filaments were co-purified in this case as no distinct peak appeared in the void volume of this purification.

In addition, we attempted to use anion-exchange chromatography to remove additional bound nucleic acids such as rRNA. While the omission of $MgCl_2$ from the purification buffer apparently removed the rRNA contamination, we were able to remove full-length transcript from other Cascade-like assemblies containing small RNA by this method. The most efficient purification protocol now included the omission of $MgCl_2$, use of T7 RNAP for production of the repeat-tagged target RNA and loading of the concentrated sample after the Ni-NTA purification on a MonoQ column. The bound sample was then separated along a salt gradient followed by SDS-PAGE analysis and RNA extraction (Figure 2.27 A). Complexes free of unspecifically bound nucleic acids were then further purified and analysed by size-exclusion purification (Figure 2.27 B).

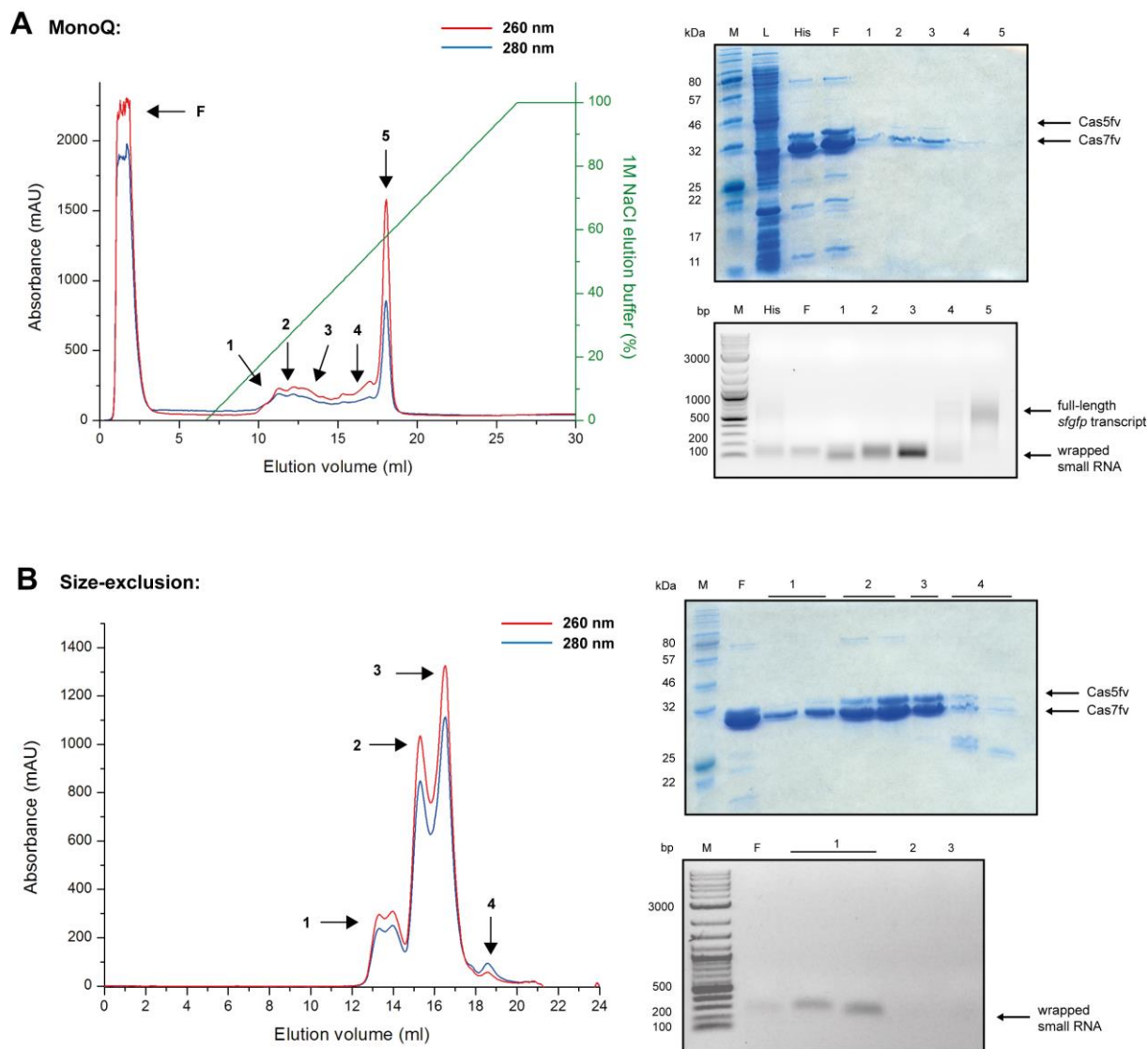


Figure 2.27: Anion-exchange chromatography separates directed Cascade assemblies from full-length transcript. (A) Anion-exchange with MonoQ column. Concentrated sample from Ni-NTA was loaded on a MonoQ column, then eluted in a gradient by an increase of the salt concentration (from 0.3-1M NaCl). Protein complexes with mostly Cas7fv and minor amount of Cas5fv are eluted during the gradient (1-4) that contain small RNA exclusively as shown by SDS-PAGE analysis and RNA extraction (right). Pure RNA matching full-length transcript is eluted closer to the end of the salt gradient (5). **(B)** Size-exclusion chromatography of the flow-through (F) of the MonoQ column, reveals multiple peaks of complexes with varying amount of Cas proteins bound to the small RNA. Proteins were analysed by SDS-PAGE and extracted RNA was analysed with agarose gel electrophoresis.

Because we used T7 RNAP to produce the target RNA instead of the weakly transcribed non-coding target from the previous purification, noticeably more full-length transcript eluted as a sharp peak at the end of the salt gradient in the anion-exchange chromatography. No clear protein band was detected for this peak on SDS-PAGE but a relatively clear signal fitting to full-length transcript could be seen on gel electrophoresis after RNA extraction. Additionally, the peak also features a high UV absorbance ratio of 260 nm to 280 nm, indicating the content of this peak to be mostly RNA. Other protein complexes eluted during the salt gradient (1-4) contained a relatively high amount of Cas7fv and a smaller amount

of Cas5v. RNA extracted from these samples as well as from the flow-through of the MonoQ column contained small RNA as indicated by a clear signal during gel electrophoresis. No full-length transcript could be detected in these samples. Considering that a relatively weak signal corresponding to full-length transcript can be seen in the sample of the Ni-NTA purification prior to anion-exchange chromatography, it can be assumed that the full-length transcript was bound to the MonoQ column and subsequently washed off.

The flow-through samples of the MonoQ purification step were further analysed by size-exclusion chromatography (Figure 2.27 B). The following UV chromatogram shows three major peaks and a very minor fourth one. Most notable is the first peak eluting at a very early elution volume that corresponds to an approximate molecular weight of 700-400 kDa. The SDS-PAGE analysis of these samples shows that this peak contains mostly Cas7fv as well as a very minor amount of Cas5fv which fits for Cascade assemblies on small RNA with the appropriate number of Cas7fv subunits. The relatively broad elution volume of this peak could indicate a variety of different complex sizes.

For the second peak, the elution volume at ~ 400 - 150 kDa as well as the protein content visualized by SDS-PAGE speaks for the presence of smaller complexes with less Cas7fv involved and which are perhaps more similar in size to a real Cascade complex. The third peak most likely contains the typical Cas5fv-Cas7fv dimer that is commonly purified. The first broad peak contains the small RNA already overserved for MonoQ purified samples, further indicating this to contain the specific RNA wrapping complexes. The latest and very small peak apparently contains only artefacts.

2.2.2.8 Structural analysis of RNA wrapping complexes on small RNA

With a better understanding of how the repeat-tagged RNA is wrapped as well as having removed the co-purified rRNA and unwrapped transcript, we analysed the structure of the newly purified complexes with TEM. Samples from the first peak of the size-exclusion after anion-exchange chromatography revealed numerous “shrimp”-like complexes with a length of approximately 20 nm. However, clear structures were only identified in the fraction of the right side from the peak (Supplementary Figure 5). The visualized structures appear to be quite flexible and lie in various different orientations when observed by TEM (Figure 2.28 A & B). By 2D class averaging we were able to obtain a rough 3D model that shows the general shape of this complex. Approximately seven Cas7fv subunits and one additional Cas5v (or Cas7fv) fit in this structure.

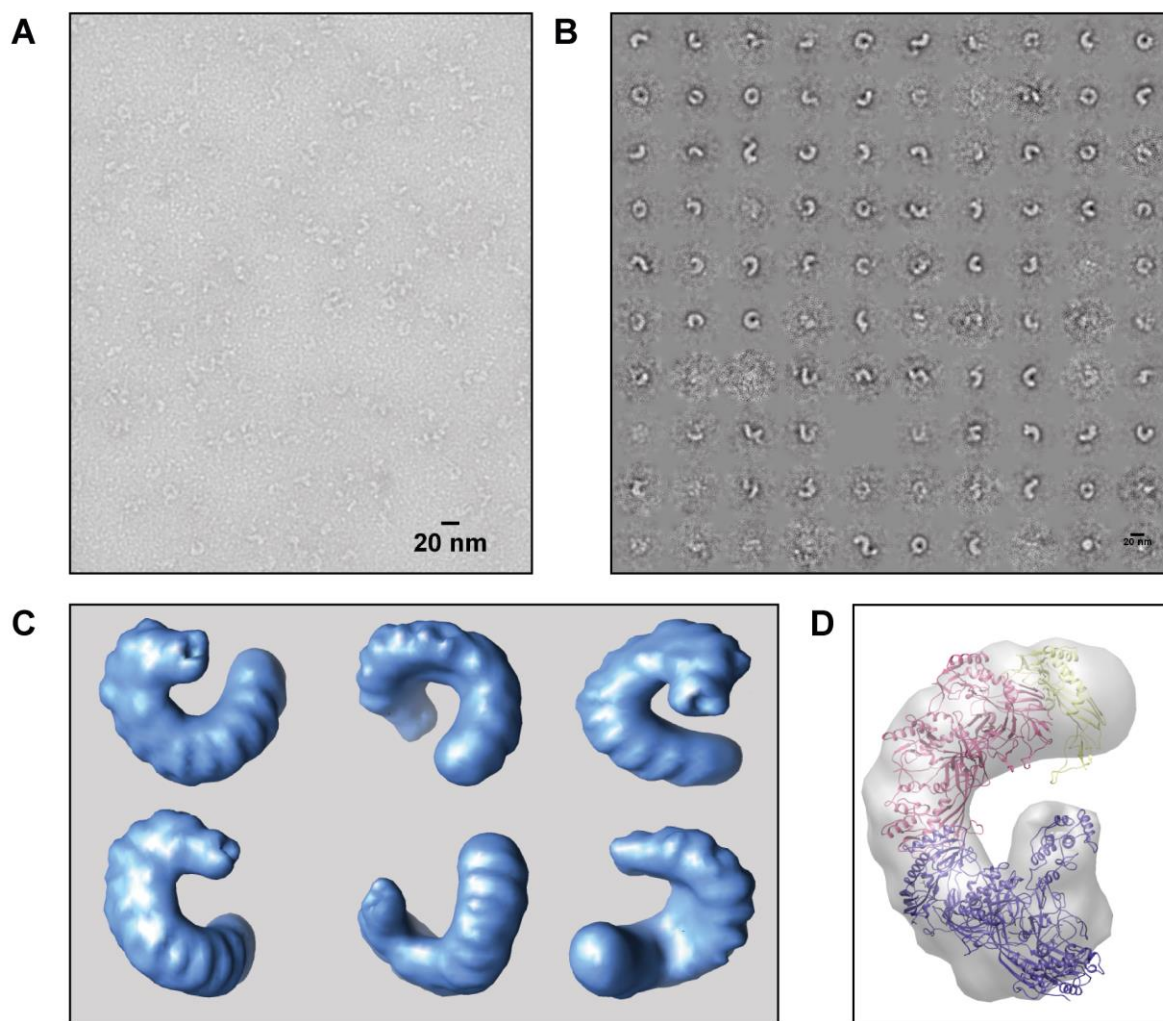


Figure 2.28: TEM analysis and 3D structure of RNA wrapping complexes by 2D class averaging. (A) Original TEM view of purified structures by a ion-exchange and size-exclusion chromatography. (B) Collection of various shrimp-like structure from TEM in various different orientations. (C) 2D class averaging model by merging of all structures from different viewpoints. (D) Superimposition of Cas protein structures from the I-Fv Cascade model in the 2D class averaging model. Approximately 8 Cas proteins can fit in the structure. Purple: 1 Cas5fv plus 3 Cas7fv, pink: 3 Cas7fv, yellow: 1 Cas7fv monomer.

The optimal purification and the following TEM analysis showed the shape of the synthetic Cascade assemblies on the initial 5'-terminal portion of the repeat-tagged RNA. The general shape of this specific complex is not very different in comparison to I-Fv Cascade. In fact, the obtained 3D model by 2D class averaging suggests that approximately the same number of Cas5fv and Cas7fv subunits as are present in the natural interference complex. It should be noted though that this model is not comparable to e.g. electron density data from crystallography and only a rough estimation can be made about how many proteins potentially fit in this structure. Additionally, the observed structures on TEM itself also look very similar to I-Fv Cascade complexes previously analysed by TEM. This is contradictory to the

previously performed RNA-seq analysis which shows that the wrapped RNA should be at least 100 nt long and definitely longer than a 60 nt crRNA.

Similarly shaped structures of about the same size were visualized by TEM in the elution fractions of salt gradient during the MonoQ purification step (Supplementary Figure 6). However, these structures were a bit more aggregated and the size was not easily identifiable.

2.2.2.9 Backbone-modification of synthetic Cascade complexes

In the following experiments, we investigated how the synthetic Cascade assemblies we created can be modified. For this, Hamrithaa Shanmuganathan fused the fluorescent reporter sfGFP to the backbone-forming protein Cas7fv during her internship.

To see if backbone modification is possible in general, we aimed to modify the natural I-Fv Cascade complex instead of the artificial RNA wrapping complexes. I-Fv Cascade with the sfGFP-Cas7fv fusion protein was created by co-expression of the *cas* genes and with a naturally occurring crRNA (Figure 2.29 A). Proteins were purified by Ni-NTA and size-exclusion chromatography. Afterwards, the fractions were analysed by SDS- and Urea-PAGE for protein and RNA content respectively. I-Fv Cascade was still formed and eluted as a stable peak that contained a crRNA as well as all Cas proteins including the sfGFP-Cas7fv fusion protein that exhibited noticeable fluorescence (Figure 2.29 B).

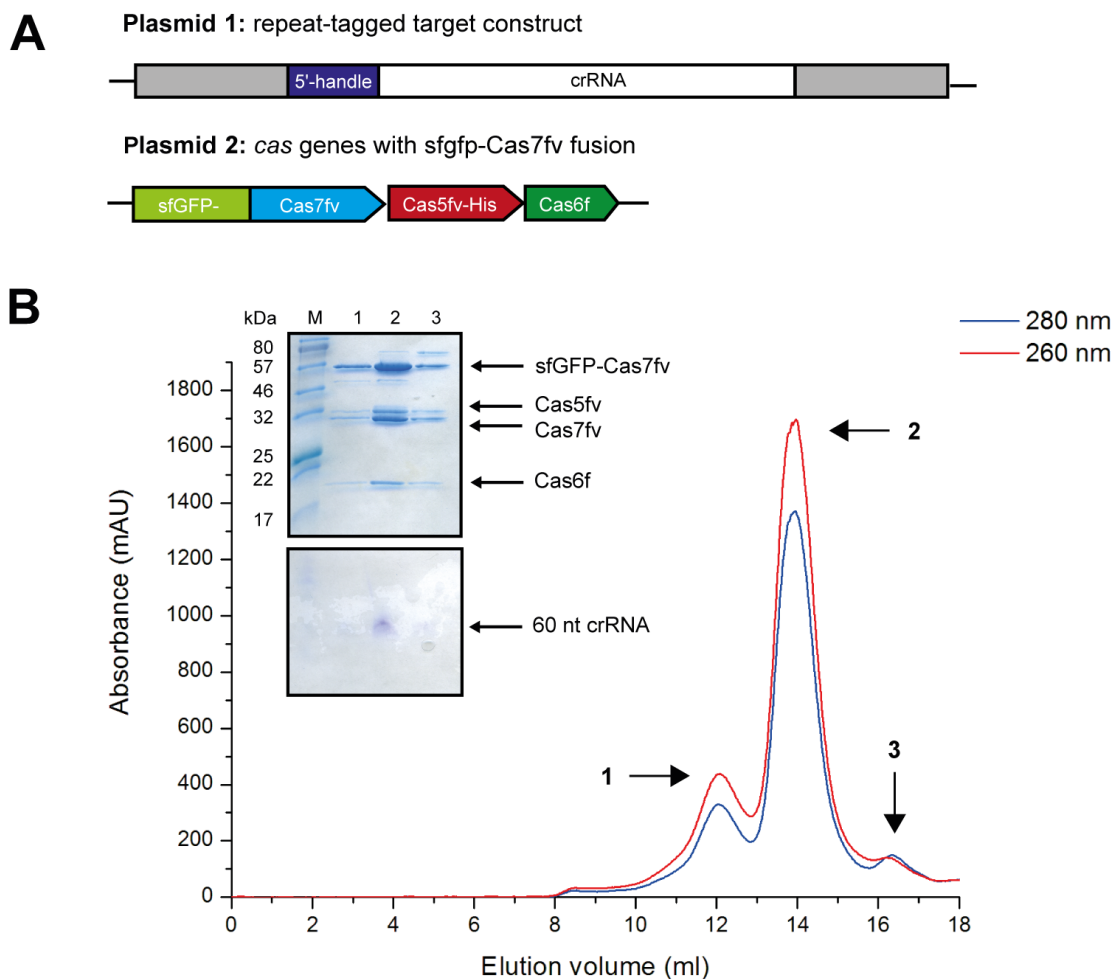


Figure 2.29: Backbone modification of I-Fv Cascade with sfGFP fusion. (A) Schematic overview of the expressed genes for assembly of I-Fv Cascade with sfGFP-Cas7fv backbone. The typical wt crRNA used for I-Fv Cascade production was co-expressed with the *cas* genes including the *sfgfp-cas7fv* fusion. **(B)** Size-exclusion chromatography and SDS- and Urea-PAGE analysis confirms the stable formation of I-Fv Cascade with all Cas proteins and a crRNA (2).

The UV chromatogram shows two significant peaks as well as a minor third one. The first peak contains all Cas proteins but not RNA is visible on Urea-PAGE. The early elution volume that corresponds to a high molecular weight speaks for aggregates in this fraction. The second peak apparently contains a modified I-Fv Cascade considering all Cas proteins including the sfGFP-Cas7fv fusion protein and crRNA can be detected in this fraction. Interestingly, a band corresponding to unmodified Cas7fv is detected on SDS-PAGE as well. The presence of both the sfGFP-Cas7fv fusion protein as well as the unmodified Cas7fv could indicate that the backbone contains a mixture of both these proteins. This is perhaps due to unspecific cleavage of the small GSGS linker that connects sfGFP and Cas7fv. Co-purification of unspecific proteins with 80 and 50 kDa is also apparent by extra bands on SDS-PAGE.

After we established that backbone-modification is possible in general, we then attempted to assemble the RNA wrapping complexes on the repeat-tagged reporter construct with the sfGFP-Cas7fv fusion

protein. Size-exclusion chromatography after Ni-NTA purification yielded multiple peaks including one at an elution volume containing the previously observed complex (Figure 2.30).

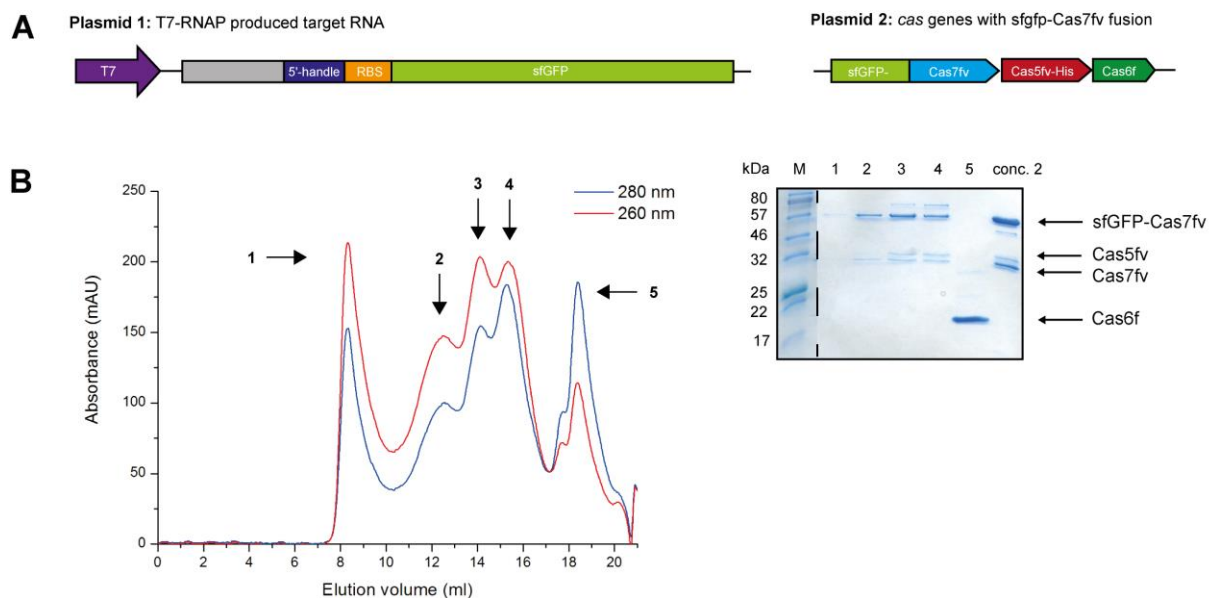


Figure 2.30: Backbone modification of synthetic Cascade assembly with repeat-tagged RNA. (A) Schematic overview of the expressed constructs for a assembly of complexes on repeat-tagged RNA with the sfGFP-Cas7fv fusion protein. (B) Size-exclusion chromatography (left) and SDS-PAGE analysis (right) reveals the formation of multiple complexes with varying size.

The UV chromatogram contains a wide range of peaks that are partially overlapping. The very first one is located at the earliest point in the void volume and only a faint band of Cas7fv was detected on SDS-PAGE in this fraction. This peak contains again most likely aggregates. Interestingly, this purification contains more aggregates compared to the purification without Cas7fv-modification which may be due to the omission of the anion-exchange chromatography step in this experiment. The second peak is located at the elution volume of the previous synthetic RNA wrapping complex but contains mostly the sfGFP-Cas7fv-fusion protein and minor amount of Cas7fv but also Cas5fv. This is most clearly visible after concentration of these fractions due to the low amount of protein purified in general. After concentration, the sample is also notably green due to the sfGFP-Cas7fv fusion protein. Peak 3 and 4 at a later elution volume contain the Cas proteins but with the additional 80kDa protein contaminant. The fifth peak at a very late elution volume contains what is most likely co-purified Cas6 or artefacts of the purification. Some Cas6f, presumably bound to RNA is eluted at the very end of the elution volume (5). At this point, we did not have another repeat-tagged target besides *sfGFP* under control of the T7-promoter available. While it is obviously not ideal to use the transcript of sfGFP as a target for RNA wrapping while also modifying the backbone of the complex with the sfGFP protein, these initial experiments serve as a proof-of-principal and indicate backbone-modification to some extent. The

presence of both unmodified Cas7fv and sfGFP-Cas7fv fusion could indicate an unevenly modified backbone that should be further studied. TEM should be performed to confirm the shape of these structures.

2.2.2.10 Analysis of wrapped small RNA in unmodified and modified complexes

To ensure that the newest complex with the sfGFP-Cas7fv fusion still encases the same RNA as the complex without backbone modification and that it is not an unspecific by-product, we extracted the RNA and loaded it on a 10% Urea-PAGE with subsequent SYBR-Gold staining for small RNA analysis with high intensity. We also loaded the RNA extraction sample of the unmodified complex from earlier to obtain a higher separation of the small RNA than on the previous agarose gel electrophoresis. For exact size identification, we also loaded a wild type crRNA as well as the elongated crRNA with a +18 nt spacer from previous extractions.

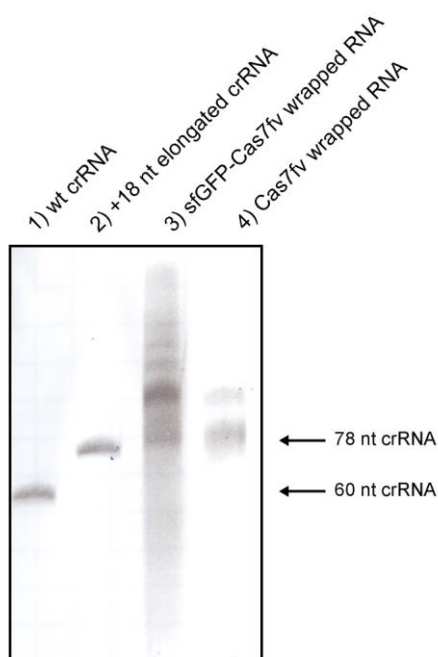


Figure 2.31: Comparison of small RNA by Urea-PAGE. A variety of RNA extraction samples was loaded on a 10% PAA-gel with 8 M Urea for direct size comparison in low range. (1) Wild type crRNA from I-Fv Cascade, (2) +18 nt elongated crRNA from the previously created synthetic Cascade, (3) RNA extracted from RNA wrapping complexes containing the sfGFP-Cas7fv fusion backbone and (4) RNA extracted from the directed RNA wrapping complexes with unmodified backbone.

The RNA extracted from the complex wrapping RNA with unmodified Cas7fv shows two distinct bands on this gel. The lower band runs on the same position as the +18 nt elongated crRNA with a total length of ~ 80 nt. The upper band, however, would fit the approximate 100 nt in accordance to the RNA-seq analysis. The same RNA can be detected in the sample of the complex with the sfGFP-Cas7fv fusion,

suggesting that this is still the same complex but with modified backbone. Considering that no anion-exchange chromatography step was included in the purification of this sample, more bands on a higher position as well as a smear going through the entire lane are visible on the gel. While this might indicate the presence of complexes wrapping around longer RNA, it could also highlight the importance of this additional purification step in removing co-purified nucleic acids.

The direct comparison of both these samples with a 60 nt wild type and a 78 nt elongated crRNA allows to effectively judge the size of this wrapped RNA and proves that longer RNA is present in the directed wrapping complexes even though we only obtained a surprisingly small structure by 2D class averaging. In the +18 nt elongated crRNA from the synthetic Cascade complex, the second repeat region at the 3'-end for Cas6f binding is present in the RNA sequence as well. If the distinct band from the RNA wrapping complexes has the same size, this would mean that an additional 20 nt for three Cas7fv subunits are available in the 5'-repeat-tagged transcript.

2.2.3 Additional applications of directed RNA wrapping

2.2.3.1 Silencing of reporter transcripts by directed RNA wrapping

While investigating the possibility and limitations of directed RNA wrapping by CRISPR repeat sequences in general, we also attempted to establish potential applications of this process for other purposes besides specifically isolating stabilized RNAs.

One of these applications was to silence reporter genes on the level of translation. An initial proof of principle was made for application by drastically reducing sfGFP production in expression cultures.

To further study the effect of silencing, we first performed fluorescence microscopy of *E. coli* after expression of the repeat-tagged target construct or the un-tagged control construct as well as the *cas* genes (Figure 2.32).

Fluorescence was not noticeable in cells producing the repeat-tagged construct in contrast to cells expressing the untagged *sfGFP* or not expressing the *cas* genes, indicating effective silencing. It should be noted that some elongated and deformed cells were observed, possibly as a stress response due to high expression.

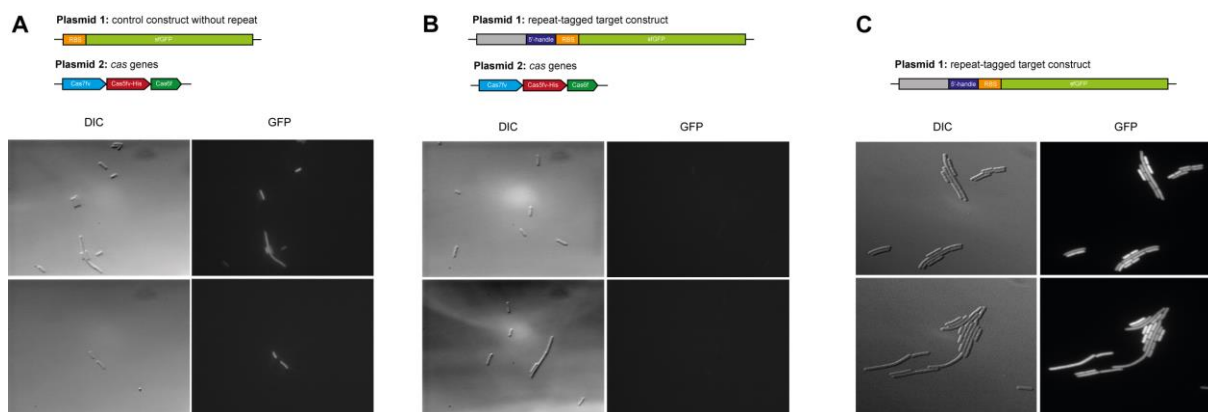


Figure 2.32: Fluorescence microscopy of *E. coli* producing (A) sfGFP and the Cas proteins, (B) sfGFP with an upstream repeat sequence and the Cas proteins, or (C) only sfGFP with an upstream repeat sequence but without Cas proteins. Cells were either excited by normal light (DIC) or for GFP. Cas gene expression completely abolished fluorescence by sfGFP in cells producing the transcript with the repeat. In contrast, fluorescent cells were detected in the control without the repeat sequence or when no Cas proteins were present. RBS: ribosome binding site

To quantify this apparent silencing effect, cells were analysed by fluorescence-activated cell sorting (FACS). In the first experiment, it was also investigated if the time of induction for either the *cas* genes or the sfGFP construct has an effect on fluorescence. Cas genes were either induced before *sfgfp*, at the same time or after *sfgfp* (Figure 2.33).

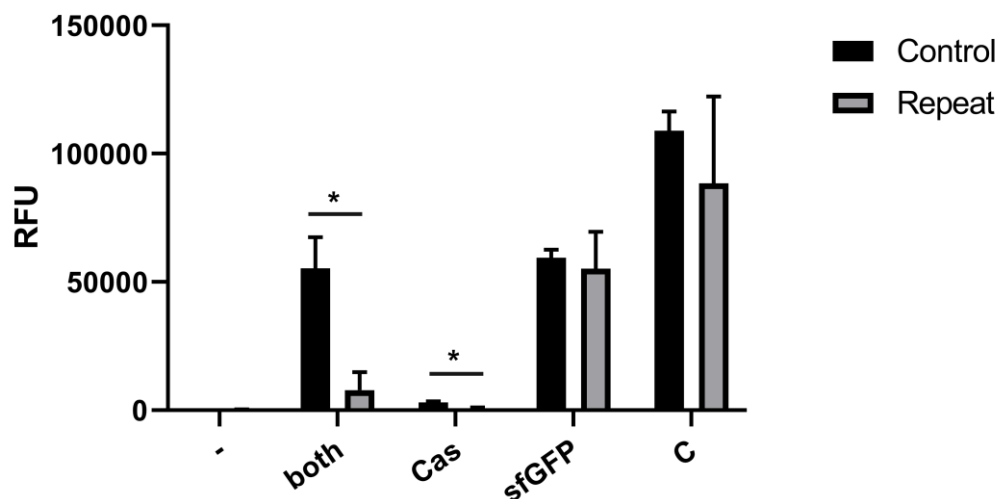


Figure 2.33: FACS analysis of *E. coli* expressing cas genes as well as the control or the repeat-tagged construct. Fluorescence was measured in relative fluorescence units (RFU). Different induction time points of *cas* gene expression were investigated. Cas genes were induced 1h before *sfgfp* (Cas), at the same time (both) or 1h after *sfgfp* (sfGFP). Cells not producing sfGFP at all (-) showed no fluorescence at all and a control expressing only *sfgfp* (C) showed the maximum fluorescence. Significance was measured by a Student's t-test and is indicated as by a star.

Effective silencing under these conditions and over a short time was most visible when *sfgfp* and *cas* genes were expressed at the same time (both). While cells expressing the control construct exhibited noticeable fluorescence, all values were significantly lower for cells expressing the repeat-tagged construct under the same conditions. The lower fluorescence for the repeat-tagged construct further confirms the specificity and that directed silencing on RNA is possible.

When *sfgfp* expression was induced one hour before the *cas* genes, both constructs produced notable fluorescence. However, the values are significantly lower compared to the control in which no *cas* genes were expressed (C). Additionally, when *cas* gene expression was induced one hour before *sfgfp*, both constructs showed no significant amount of fluorescence. While this would be expected for the repeat-construct, it is not for the control and could indicate that this overproduction of Cas proteins can lead to unspecific interaction and completely inhibit later *sfgfp* translation. Otherwise, this could indicate a toxic effect on the cell or stress and growth inhibition. Active transcription of *sfgfp* provides a target for Cas proteins countering this effect. This experiment was also performed with simultaneous induction of *cas* genes and *sfgfp* overnight at 18 °C, which showed the same effect of silencing (Supplementary Figure 7). These experiments have been performed with the target constructs produced under control of the pBAD promoter by arabinose, to study the effect of separate induction of the target and *cas* genes. However, as stated before, the target is hereby produced by *E. coli* RNAP which leads to significantly less target transcript produced. In the following experiments, we aimed to investigate the silencing effect on target transcripts produced by T7 RNAP that was used for purification of the RNA wrapping complexes. Silencing was also noticeable on the repeat-tagged construct produced by T7 RNAP after overnight growth at 18 °C in comparison to the untagged control construct (Figure 2.34).

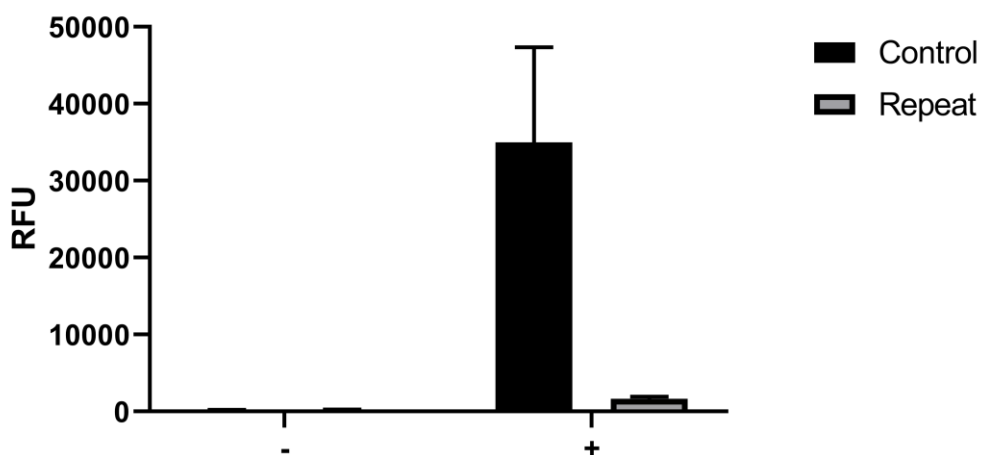


Figure 2.34: FACS analysis of *E. coli* expressing *cas* as well as either the control or the repeat-tagged construct by T7 RNAP. Since both constructs were under control of the T7 promoter, expression of both *cas* genes and *sfgfp* was either induced (+) or not induced (-) before overnight growth at 18°C of both cultures.

These measurements confirm the possibility to use this process for silencing on an RNA level by inhibiting translation on wrapped RNA in complexes which was first observed by the color of pelleted expression cultures not able to produce sfGFP. Silencing works for target production with *E. coli* RNAP as well as T7 RNAP. The drastic difference in fluorescence of the control and the repeat-construct highlights the specificity of this process.

2.2.3.2 Protection of RNA molecules by wrapping in Cas proteins

Another potential application is the use of directed RNA wrapping for protection of encased RNAs. In theory, while bound to Cas proteins, the RNA would be protected from degradation as long as the proteins are intact.

To study the stability of the produced complexes on RNA over a longer time period, we incubated fresh Ni-NTA purified proteins in complex with the repeat-tagged lacZ- α construct for numerous time points with RNase I_r (Figure 2.35 A). After incubation, RNA was extracted and loaded on Urea-PAGE for initial analysis of stability (Figure 2.35 B). In an extended experiment, the stability of the small repeat-tagged RNA was confirmed by Northern Blot analysis (Figure 2.35 C).

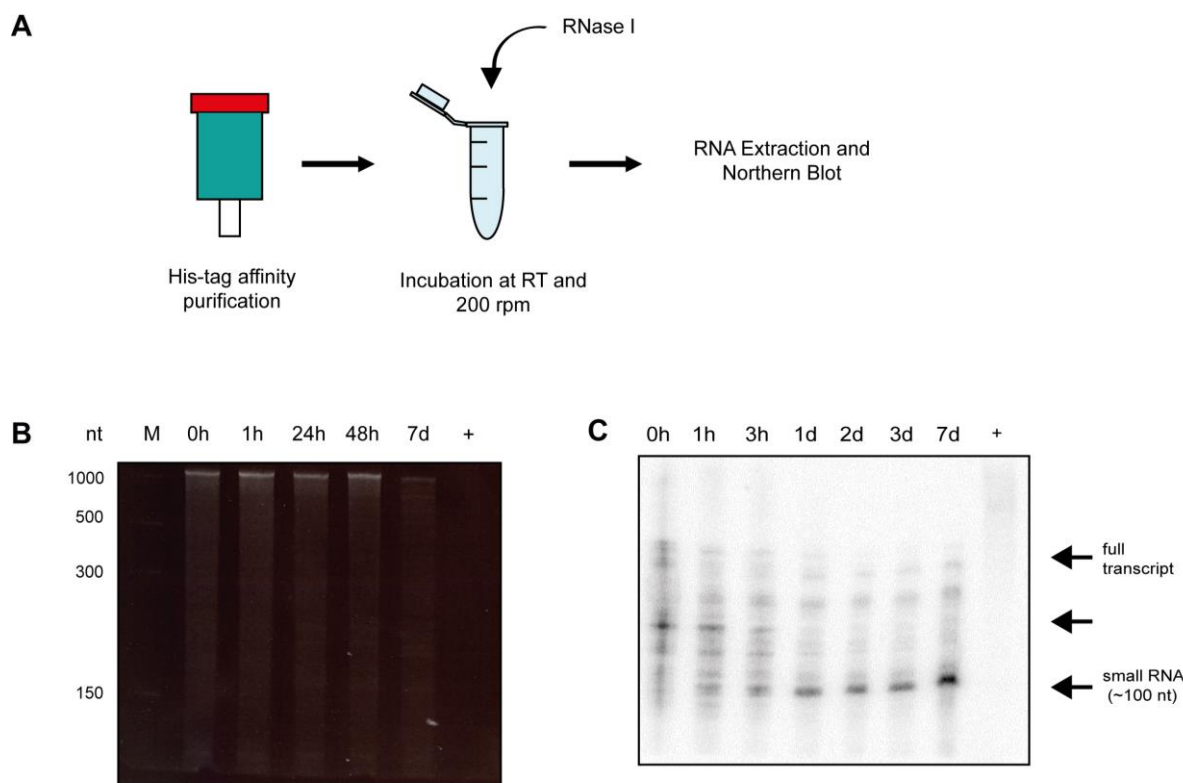


Figure 2.35: Stability of wrapped RNA. (A) Schematic principle of the experimental design. Ni-NTA purified protein samples were incubated for different time points with RNase I_r at RT and 200 rpm. RNA was then extracted and separated on Urea-PAGE (B) for subsequent Northern Blot analysis with a probe against the start of the sequence (C). Extracted RNA was incubated with RNase I_r and loaded as a control (+).

Because the experiment was performed when unspecific RNA was not yet removed, initial analysis by separation on Urea-PAGE only showed they typical smear going through the entire lane and small wrapped RNA was not identifiable on this gel. Nonetheless, RNA was present in all samples even after 7 days of incubation. Already extracted RNA, not protected by the complex, was completely degraded, confirming that RNase I is active (+). In the Northern Blot analysis, clear signals can be observed for directly extracted RNA (0h) as well as samples incubated with RNase I. No clear band can be seen in the control sample with unbound RNA, in which all RNA was degraded. On the other hand, a smear can be detected in the sample of directly extracted RNA. This is most likely due to the presence of full-length transcript or intermediates in the sample. Bands related to larger RNA decrease over time and a smaller band becomes more apparent. Even in the sample in which RNA was extracted after 7 days of incubation, this band is still clearly visible.

Overall, we showed that specific RNAs can be protected by Cas proteins for long periods of time. The exact size of the RNA is not completely defined by this method because we did not provide an *in vitro* transcribed full-length transcript and this experiment was performed before we removed unspecific RNA and full-length transcript. Nonetheless, the position of the signal on the Northern Blot fits to the size of

small RNA in the extractions of the latest structures (Figure 2.31) and that was confirmed by RNA-seq analysis.

2.2.3.3 Induced release of RNA from complexes

Another potential application would be the induced release of specifically wrapped RNA from the created complexes. While the repeat-tagged RNA can already be isolated by *in vitro* phenol/chloroform extraction and ethanol precipitation, the specific unpacking of the “cargo” RNA from the created complexes could prove useful, especially for *in vivo* applications. We investigated two potential ways to achieve this.

First, we attempted to remove Cas proteins from the wrapped RNA by an Anti-CRISPR protein (Acr). As mentioned before, these viral proteins serve as a countermeasure against CRISPR-Cas systems of the host, usually by occupying critical positions for DNA binding in the effector complex. The recently discovered AcrF15 from *Alcanivorax xenomutans*, is able to remove the backbone-forming proteins from the crRNA in I-F Cascade. In order to validate that AcrF15 is able to effectively disrupt our packing system, our group performed *in vivo* plasmid transformation assays. However, this assay showed no activity of Acr in the type I-Fv system (data not published yet).

As an alternative, we attempted to create complexes on repeat-tagged RNA with the I-F Cas proteins that are known to be affected by this Acr. For this, it was first required to investigate if the synthetic complexes can be assembled on repeat-tagged RNA with the type I-F CRISPR-Cas system. If possible, these complexes could then be targeted by the Acr protein to release the wrapped RNA (Figure 2.36 A). In the course of this, the directed RNA wrapping capabilities of the I-F system could be studied and compared to the I-Fv system.

To produce and purify potential I-F complexes directed on repeat-tagged RNA, the I-Fv repeat of previously used non-coding RNA construct was exchanged with the sequence of a I-F repeat and the new construct was co-expressed with the I-F Cas proteins from a commercially available plasmid (Figure 2.36 B). The Ni-NTA purified and concentrated samples were then separated by size-exclusion chromatography (Figure 2.36 C). No distinct peak was present on the following UV chromatogram but protein eluted along the entire elution volume. SDS-PAGE revealed the presence of the Cas7f protein in these fractions at earlier elution volumes. However, no clear structure could be identified with TEM in the void volume or the later fractions.

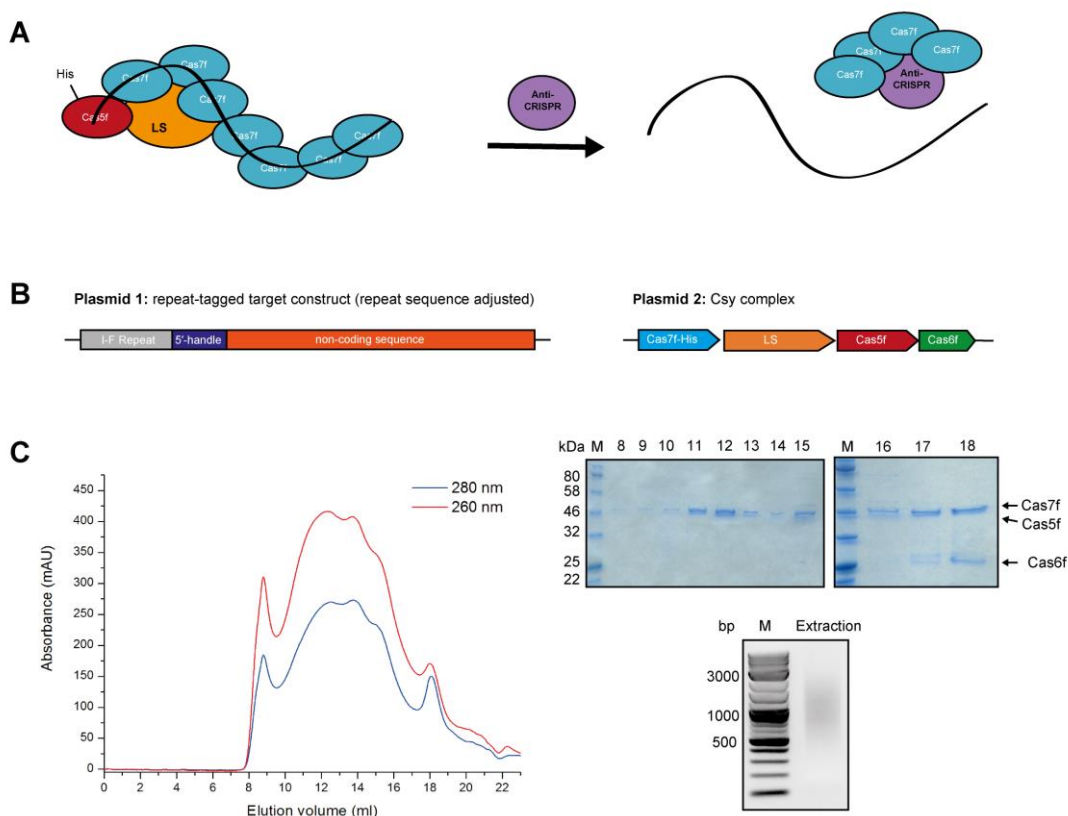


Figure 2.36: Unpacking by Acr. (A) Schematic principle for theoretical unpacking of RNA wrapped by I-F Cas proteins. A large subunit (LS) should be part of the hypothetical complex besides Cas5f and numerous Cas7f proteins. If complexes are formed and similar to Cascade, the addition of the Anti-CRISPR could potentially remove the Cas proteins from the complex. (B) Schematic representation of the expressed constructs for complex assembly. The targeted non-coding sequence was tagged with a repeat sequence from the I-F CRISPR-Cas system and cas genes were expressed from a commercially available plasmid including all I-F Cas proteins Cas7f, Cas5f, Cas6f and the large subunit (LS). (C) UV chromatogram of size-exclusion after Ni-NTA purification via His-tagged Cas7f shows a broad peak along the entire elution volume (left). SDS-PAGE analysis of the fractions shows the presence of the Cas7f protein and a minor amount of Cas5f and Cas6f (top right). RNA extraction of all fractions pooled and concentrated did not contain distinct RNA when loaded on agarose gel electrophoresis (bottom right).

As Cas7f was present in all elution fractions, it can be assumed that complexes with a wide variety of sizes were created. A faint band of Cas5f can also be seen close below Cas7f. In the later elution volumes, Cas6f is co-eluted with Cas7f due to the interaction of both proteins. However, no real specific complex was purified in a distinct peak, especially since no RNA was detected by Urea-PAGE in these samples. RNA extraction of a concentrated sample with all the protein yielded barely any RNA and only a faint smear was visible on agarose gel electrophoresis. Together with the fact that neither Cascade-similar nor filamentous structures were observed with TEM, we concluded that Cas7f is not ideal for RNA wrapping and we did not continue studies with the Acr protein.

As RNA wrapping with the I-F CRISPR-Cas system was not possible in this work, we chose to investigate the release of wrapped RNA by specific degradation of the I-Fv Cas proteins. To achieve this, we used the established synthetic system for inducible protein degradation in *E. coli* based on the *Mesoplasma*

florum transfer-messenger RNA system (Cameron & Collins, 2014). The *ssrA* tag from *M. florum* is specifically degraded by the endogenous Lon protease (*mf*-Lon) but not by *E. coli* proteases. Likewise, *mf*-Lon is not targeting *E. coli* *ssrA* (Gur & Sauer, 2008). The system can thus be used to specifically target and degrade proteins marked with this protein degradation tag (pdt).

Initially, we fused the 27 aa pdt to the C-terminus of Cas5fv and Cas7fv, with the goal to specifically degrade the formed complexes on repeat-tagged RNA with the induction of the *mf*-Lon protease (Figure 2.37 A). For complex formation in *E. coli* expression cultures, we used the optimal plasmids with T7 RNAP production of the repeat-tagged *sfgfp* RNA (Figure 2.37 B).

Unfortunately, it was not possible to create directed RNA complexes with protein degradation tags. The absence of complex formation and RNA wrapping with these constructs was immediately apparent by the green color of the pelleted expression cells due to unhindered sfGFP production.

In the following Ni-NTA purification, Cas5fv and Cas7fv were only detected by SDS-PAGE in the pellet after ultracentrifugation (Figure 2.37 C). A likely explanation for this is that the fusion of pdt inhibited proper folding of the Cas proteins. This is surprising, considering that entire proteins have been fused to the C-terminus of these proteins before (such as sfGFP for Cas7fv and Dendra for Cas5fv).

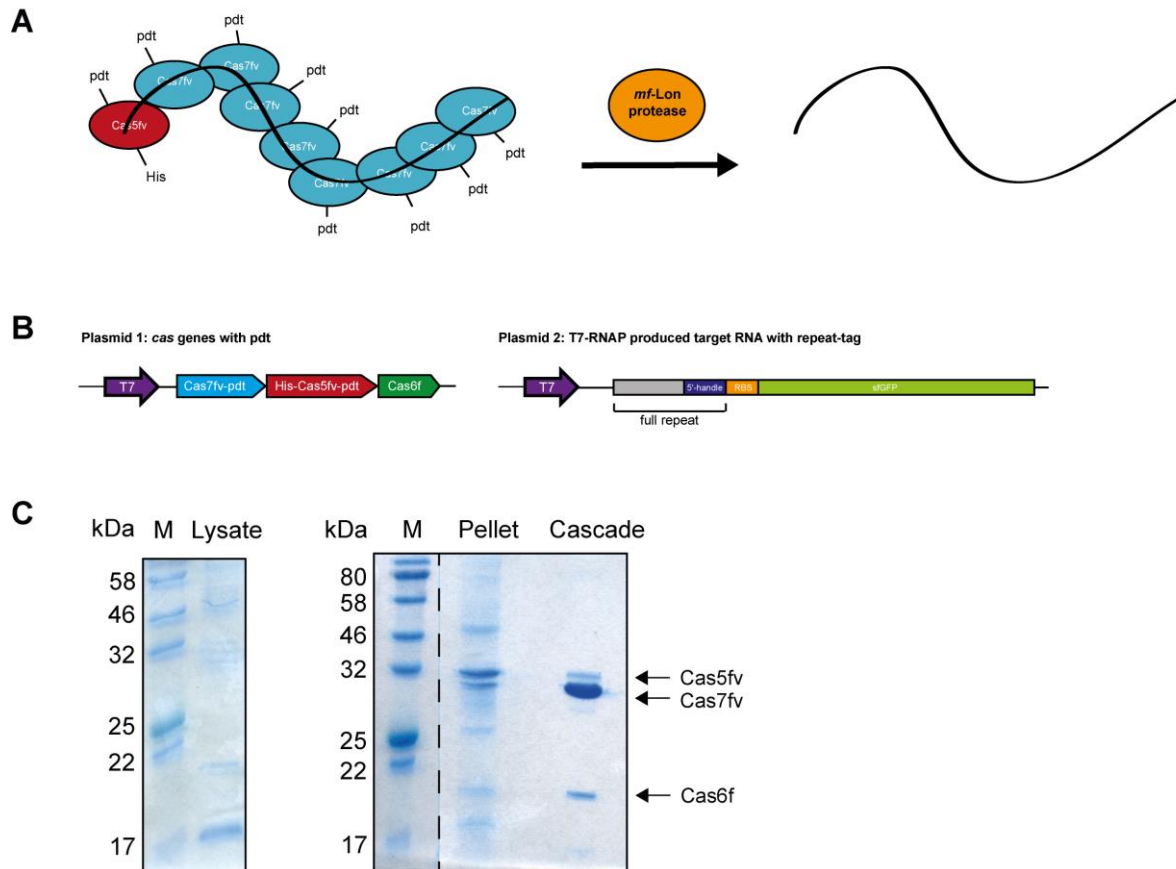


Figure 2.37: Production of Cas proteins with protein degradation tags for induced unpacking of RNA. (A) Schematic principle of the theoretical process for release of RNA from complexes by protein degradation. Protein degradation tags (pdt) are fused to the C-terminus of His-Cas5fv and Cas7fv that form a complex on the targeted RNA. After addition of the specific *mf*-Lon protease, the proteins are degraded and the RNA is released from the complex. (B) Plasmids used in the expression culture for production of the complex with protein degradation tags. Cas proteins with pdt and repeat-tagged *sfGFP* RNA were produced by T7 RNAP. (C) SDS-PAGE analysis of lysate and pellet after cell harvest and homogenisation. Cas5fv and Cas7fv were detected in the pellet but not in the lysate. Cascade was loaded for size comparison.

Overall, though the creation of specific RNA wrapping complexes was successful, we found no way to release the wrapped RNA from synthetic complexes *in vivo* during the scope of this work. Future studies will have to investigate why the addition of the small 27 aa pdt on the C-terminus of both Cas7fv and Cas5fv is problematic for protein folding and how to circumvent this problem. Furthermore, alternative ways for the release of wrapped RNA could be examined.

3. Discussion

3.1 *In vitro* analysis of type I-Fv Cascade

The research that was conducted before this PhD thesis established that the proteins Cas5fv and Cas7fv form the Cascade complex of the type IFv system together with Cas6f and a processed crRNA. Previously, I-Fv Cascade was purified with a His-tag on Cas7fv and Cas6f (Dwarakanath *et al.*, 2015, Gleditzsch *et al.*, 2016). For the research described in this thesis, we switched the His-tag to the Cas5fv protein. This purification procedure allows for limited amounts of co-purified byproducts, such as Cas7fv filaments in case of purification by His-tagged Cas7fv or Cas6f interacting with Cas7fv or RNA in case of His-tagged Cas6f. In general, purification of I-Fv Cascade is possible with a His-tag on all proteins. The later solved 3D structure revealed that all termini are accessible.

In vitro assays have shown that the I-Fv CRISPR-Cas system enables interference when produced in *E. coli*. To study the interference mechanism, it was first necessary to prove *in vitro* target binding which was analyzed in a step-wise fashion. First, ssDNA binding was observed which served as an initial proof after purification that this complex is a functional CRISPR-Cas interference complex and interference is due to crRNA hybridization with the target DNA. However, it was not possible to show PAM dependency in these experiments.

Additionally, no clear shift indicating dsDNA target binding could be observed by EMSA. Upon close inspection, a faint shift can be identified which could indicate dsDNA binding but with much lower efficiency compared to ssDNA. One possible explanation would be that type I-Fv Cascade evolved to exclusively target ssDNA and relies on processes in which ssDNA is formed similar to transcription-coupled type III systems (Jia *et al.*, 2019, You *et al.*, 2019).

Because no clear dsDNA binding and thus R-loop formation was detected, the next target construct was designed to mimic an opened dsDNA. Cascade was still able to bind this target, although again without PAM-dependency. A similar construct was also used for structural analysis of type I-E Cascade (Hayes *et al.*, 2016). We later used this approach to successfully solve the 3D structure of I-Fv Cascade bound to target DNA. This highlighted the target recognition mechanism of Cas5fv and showed direct amino acid interaction with the GG-PAM confirming PAM-dependency *in vivo*.

Earliest cleavage assays were not able to show Cas3fv recruitment to the R-loop mimic construct. When we optimized purification of Cas3fv and purified a stable standalone variant without the Cas2 domain, we adopted target constructs containing a small 10 nt “bubble” opening that were already successfully used for Cas3 cleavage assays of the type I-F system (Rollins *et al.*, 2017). Our EMSA assays revealed that this small opening of dsDNA is enough for I-Fv Cascade to unwind the adjacent dsDNA. These substrates that allow full R-loop formation were used to study Cas3fv recruitment. Cascade mediated DNA target

binding efficiency was also measured with Biolayer Interferometry (BLI) in our group (Müller-Esparza, 2019). These experiments showed that Cascade is able to bind a full complementary dsDNA strand, but with a much weaker dissociation constant (K_D) compared to the opened bubble opening. This small opening, preferably in proximity to the PAM, is enough to increase binding efficiency and allow R-loop formation.

For dsDNA, constructs with the correct GG-PAM also showed a much stronger binding efficiency compared to constructs with a wrong PAM. PAM dependency is thus apparent for dsDNA, while binding efficiency with ssDNA is so strong that PAM recognition does not play an observable effect. The 3D structure confirmed that the PAM sequence is only recognized in the double-stranded form. With these results in mind, the weak signal in EMSAs with dsDNA could represent a minimal amount of full R-loop structure products. It remains unclear, why we could not observe complete dsDNA binding with EMSA, although Cascade was added in excess to target DNA. BLI assays proved to be more sensitive and thus suitable for this purpose.

The 3D structure of I-Fv Cascade provided a significant amount of information about the complex. It showed how structural variations in the novel Cas7fv and Cas5fv proteins substitute for the missing subunits of other type I CRISPR-Cas systems. Additionally, it explained the mechanism of PAM recognition.

In the type I-E system, PAM recognition is based on minor groove DNA interactions of the large subunit Cas8e (Hayes *et al.*, 2016) which is required for a more promiscuous PAM recognition mechanism that allows the recognition of several PAM sequences. The type I-C system has a very similar mechanism for Cas8c which was investigated by cryo-EM (Hochstrasser *et al.*, 2016). The PAM recognition mechanism by minor groove interactions was thus suggested to be a conserved feature in type I systems. In contrast, type I-Fv PAM recognition proceeds from the major groove side of DNA.

The trench route is here formed by the wrist-loops of Cas5fv and Cas7fv. In type I-F, the extended web is expected to serve the same function and to substitute the small subunits. Contrary to type I-E PAM recognition, which employs a glutamine wedge, type I-Fv uses Cas5fv to insert α -helix 3 as a wedge, leading to seed bubble formation.

In comparison to established R-loop formation by type I-E (Figure 1.6), we propose the following model for R-loop formation by type I-Fv Cascade (Figure 3.1): I. Cascade scans the major groove of dsDNA for correct GG PAM sequences via the AH and wrist of Cas5fv. II. PAM recognition leads to conformational rearrangements of AH, which inserts α -helix 3 as a wedge into the major groove resulting in seed bubble formation. III. Cascade samples for protospacers complementary to the crRNA spacer forming crRNA/target heteroduplex. IV. Simultaneous association of the non-target strand to the trench route formed by the wrist helix. Full R-loop formation and locking are required for Cas3fv recruitment.

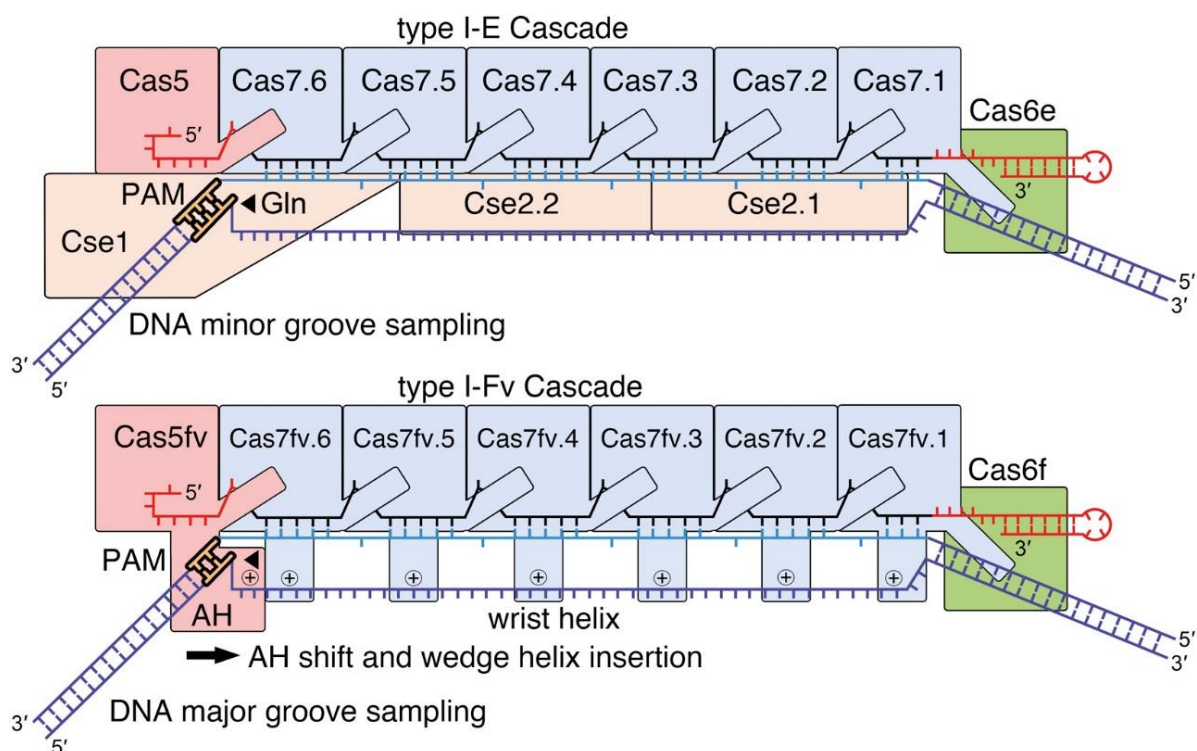


Figure 3.1: R-Loop formation by type I-E and I-Fv Cascade. Side-by-side comparison of components and mechanism of target DNA recognition by type I-E and I-Fv Cascade as deduced from structural analyses. Mechanism of surveillance and R-loop formation by the *E. coli* type I-E Cascade (compare to (Hayes *et al.*, 2016)) are shown in the upper panel. A glutamine wedge (arrow) of the large subunit Cse1 (Cas8e) opens the R-loop, and the non-target strand is guided by small subunits Cse2. The minimal *S. putrefaciens* type I-Fv Cascade structure revealed a shift of the AH domain upon target DNA binding, which results in wedge helix insertion (arrow), as shown in the lower panel. The non-target strand is guided along a trench route formed by Cas5fv and Cas7fv WLs. Figure from Pausch *et al.*, 2017

Why did the observed structural variations of type I-F Cascade complexes evolve? We speculate that they occurred due to evolutionary pressure by viral anti-CRISPR proteins. The arms race between prokaryotes and viral predators is considered to be a key factor that drives evolution (Koonin *et al.*, 2017). Our structural comparison with the structure of I-F Cascade bound to AcrF1 and AcrF2 shows that all targets for these viral proteins are either removed (LS) or modified (Cas7 and Cas5) (Figure 3.2).

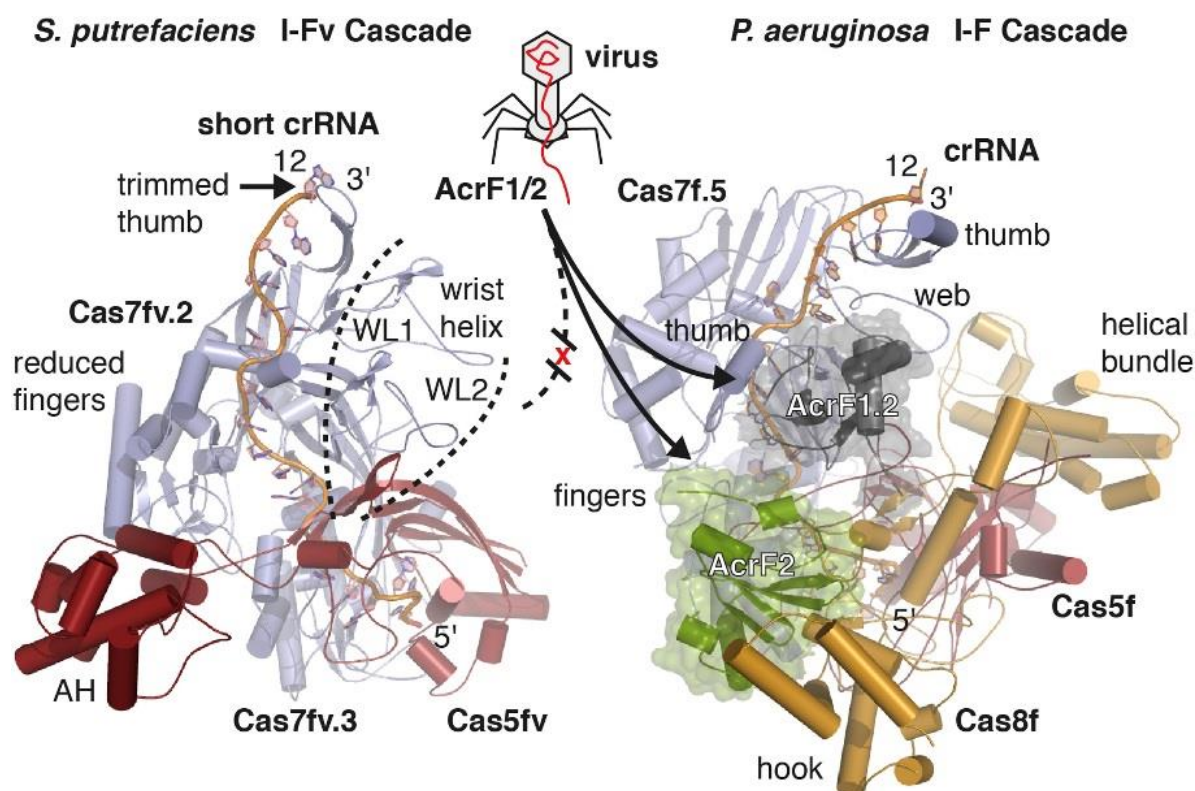


Figure 3.2: Structural Cascade variation in response to AcrF1 and AcrF2. Side-by-side comparison of the I-Fv Cascade base structure and the homologous and AcrF1/2 vulnerable I-F Cascade base structure (PDB ID: 5ZU9; (Chowdhury *et al.*, 2017)). Both structures are shown in cartoon representation and the Cascade body and head structures were removed for clarity. The Cas proteins were colored according to Figure 1.8. The anti-CRISPR proteins AcrF1.2 (grey) and AcrF2 (green) are highlighted by transparent surfaces. Regions to which AcrF1/2 associate in I-F Cascade (i.e. Cas7f thumb and web; Cas8f) differ drastically in their structure to I-Fv, not allowing binding of AcrF1/2 to the I-Fv Cascade. Figure from Pausch *et al.*, 2017.

Following this hypothesis and considering the low sequence similarity to other Cas3 subunits, Cas3fv most likely evolved as well in response to the changes in the I-Fv interface or by AcrF3 proteins. Structural data of the Cas3fv protein will be required to elucidate these scenarios.

So far, no Acr has been found that targets the type I-Fv system and all 14 Acrs that target the *P. aeruginosa* I-F Cascade are inactive against I-Fv Cascade (Müller-Esparza, 2019). A potential Acr that targets I-Fv Cascade and perhaps has evolved to counter these countermeasures in the continuing arms race has yet to be identified and requires more sequencing data from phages or MGEs targeting the type I-Fv system.

The unique AH domain of Cas5fv can be deleted without affecting complex stability of I-Fv Cascade, further suggesting the independent evolution of this domain. On the other hand, deletion of the wrist loops from Cas7fv destroys the wrist helix that turns out to be essential for complex stability. The loss of small subunits appears to be a more conserved feature. Small subunits are also absent in type I-F Cascade where they are substituted by the extended web, serving the same function. If small subunits have been indeed replaced and did not evolve from a common ancestor, a possible explanation for this

might be the existence of an Acr protein that targets small subunits directly. Another possible and more general advantage might be that fewer proteins allow for faster complex formation and in turn a faster immune response.

Although minimization provides an apparent advantage for the type I-Fv system, it is not common in other type I systems (Makarova *et al.*, 2015). Interestingly, interference levels of the minimal type I-Fv system were calculated to be 10-fold lower than for the type I-F system of *Pseudomonas aeruginosa* PA14 by *in vivo* assays performed in our group (Steube, 2018). In accordance with this, the K_D values for dsDNA binding calculated with BLI were much lower for type I-Fv Cascade (46.65nM (Müller-Esparza, 2019)) compared to for type I-F Cascade from *Pseudomonas aeruginosa* (1nM (Rollins *et al.*, 2015)). A possible explanation for this could be that the PAM recognition mechanism from the major groove leads to higher target recognition times. Otherwise, the minimized architecture of I-Fv Cascade could result in lower stability of Cascade/R-loop complexes. Indeed, mismatches between both target DNA strands were shown to increase binding affinity, indicating that Cas7fv cannot efficiently stabilize the non-target strand (Müller-Esparza, 2019). Decreased interference might also be a result of a different Cas3fv recruitment mechanism that still needs to be investigated. Overall, the minimal type I-Fv might have sacrificed binding affinity for Acr resistance, which might explain why this system is not as widespread among prokaryotes as e.g. the type I-E system.

3.2 Investigation of the nuclease activity of Cas3fv and the Cas1-Cas2/3 super complex

The structure and activity of the type I-Fv-specific Cas3fv DNA nuclease remained largely unexplored. In previous experiments, various expression conditions and purifications protocols were tested, including the use of different purification tags. Among these, expression with a SUMO-tag was found to be successful and small amounts of soluble protein could be purified by which it was possible to establish the ssDNA cleavage ability of this protein in presence of divalent cations (Dwarakanath, 2015).

In this work, we have tried to optimize purification protocols for Cas3fv and study the nuclease activity of the protein. In the meantime, usage of an MBP-tag (Maltose binding protein) has been established in purifications of related Cas3 proteins, including Cas3 from type I-F (Hochstrasser *et al.*, 2014). While the MBP-tag seemed to improve the amount of purified protein by Ni-NTA, size-exclusion revealed that all protein eluted in the void volume, indicating either aggregation or binding to large nucleic acids. Additionally, ssDNA cleavage activity was not observed. Future experiments could investigate nucleic acids co-eluting with this protein and include an ion-exchange purification step to remove them. Additionally, a GST-tag could be utilized that has been successful for purification of different Cas3 variants (Wang *et al.*, 2016).

Decent amounts of soluble Cas3fv were purified when we also investigated the fusion of the Cas3fv protein to the Cas2 domain. The successful purification and relative stability of standalone Cas3 indicate that the Cas2 domain is not required for interference.

Activity of standalone Cas3fv, independent of Cas2, was confirmed on ssDNA substrates. Cas3fv cleaved the fixed amount of substrate with increasing time in the presence of divalent Mg^{2+} and Mn^{2+} ions and is thus capable of repeated degradation. In radioactively labeled nuclease assays, standalone Cas3fv was also able to fully degrade bubble substrates with a small 10 nt opening in the presence of ATP, indicating a functioning helicase capable of unwinding the remaining ssDNA and subsequent DNA cleavage.

Full degradation of R-loop substrates by Cas3fv confirmed that the R-loop/Cascade complex is able to recruit Cas3fv. The small cleavage products below 10 nt correspond to shredded DNA.

Additional assays should attempt to quantitatively measure nuclease activity after this general proof of *in vitro* cleavage. In the provided nuclease assays with R-loop substrates, activity appeared quite low and conditions should be optimized to show progressive degradation in form of a gradient of cleavage with increasing incubation time.

In general, a nuclease-deficient HD-mutant of Cas3fv would be the most ideal control for these assays. Unfortunately, the HD-mutant we created remained insoluble. Further optimization of purification or the addition of solubility-tags might help to overcome this issue. The effect of different divalent cations on cleavage activity could be studied as well. In some systems, Fe^{2+} ions were shown to inhibit cleavage (Wang *et al.*, 2016). If the same effect can be achieved with any metal ions, these could be used as a control in assays or even for stabilization in purification. With a purified but catalytically inactive Cas3fv, future assays could be performed to study recruitment of Cas3fv for example by observing a supershift on EMSA with Cascade/R-loop substrates. Additionally, the helicase activity might be studied with appropriate assays, including AMP-PN to block the helicase domain for control purposes.

In the end, structural analysis would also enable a clearer understanding of how Cas3fv is recruited and how target cleavage is performed. Future attempts at solving its 3D structure will focus on crystallization or cryo-EM.

Structural analysis should also be performed on the Cas1-Cas2/3fv supercomplex, whose formation in the type I-Fv system was proven in this work. This would confirm the architecture of this supercomplex but also help to solve the 3D structure of Cas3fv because the complex proved to be highly soluble during purification in comparison to standalone Cas3fv or Cas2/3fv which allows for easy purification. On the other hand, the size and complexity of this supercomplex provide challenges for crystallization, further arguing for the use of cryo-EM.

Initial TEM analysis did not show any clear structure of the supercomplex. We then attempted SAXS analysis, which we successfully used before to analyze the shape of the synthetic Cascade variants.

Unfortunately, the data quality of the performed measurements was not sufficient to obtain a 3D model. This could be due to complex breakdown during freezing/thawing and subsequent sample inhomogeneity. Structural analysis should be further attempted to confirm and compare the structure of the supercomplex. Nonetheless, the general assembly of this systems complex is expected to be similar to type I-F, considering that the estimated molecular weight of the elution volume fits a complex of two Cas1 dimers and two Cas2/3fv proteins.

The fusion of *cas3* to *cas2* and the formation of the Cas1-Cas2/3 supercomplex was first investigated in the I-F system of *Pseudomonas aeruginosa* (Rollins *et al.*, 2017, Fagerlund *et al.*, 2017). EM analysis and reconstruction revealed this complex to have a four-lobed “propeller-shaped” structure consisting of two Cas2/3 proteins and two dimers of Cas1. Cas1 and Cas2 were known to form a complex for adaptation (Nunez *et al.*, 2014) and the fusion of Cas2 with Cas3 integrates the nuclease into this complex. Radioactively labeled cleavage assays with the Cas1-Cas2/3 were shown to inhibit nuclease activity unless this complex is recruited by target-bound Cascade, in which case activity is restored (Rollins *et al.*, 2017).

The Cas1-Cas2/3fv complex from the I-F variant system investigated in this work seems to inhibit Cas3 ssDNA cleavage activity as well, as shown by incubation of complex with ssDNA substrate. The substrate was only cleaved with very a high concentration of added protein, presumably either due to complex breakdown overtime or sample inhomogeneity. However, in radioactively labeled nuclease assays, Cas1-Cas2/3fv was shown to partially cleave empty bubble dsDNA substrates without Cascade as well as R-loop substrates. The cleavage was in both cases not complete and a strong band of intact substrate remained. The partial cleavage seen in these assays differs in comparison to nuclease assays with R-loop substrates both in the type I-Fv and type I-F system, because of the very prominent band at around 40 nt, representing approximately half the size of the 90 nt substrate.

A possible explanation for the intermediate band is that it represents the nicking position of Cas3fv and that the following dsDNA could not be degraded. Likewise, small ssDNA was not fully degraded in this assay and an intermediate band was produced in addition to fully degraded DNA. In the case of the I-F system, a similar but faint signal has been observed for “bubble” substrates without bound Cascade when the target was almost completely degraded. Here, this was thought to be a sign of imprecise positioning of the HD-active site to the ssDNA bubble (Rollins *et al.*, 2017). The size of the intermediate cleavage products is also in a range similar to a typical spacer of the type I-Fv system which could be an indication for production of prespacers by Cas3fv in the primed adaptation process known from other CRISPR-Cas systems (Kunne *et al.*, 2016).

HDX-MS experiments performed by our collaborators have already shown that Cas3 recruitment takes places on the non-target strand as this position was more protected from H/D exchange (data not yet published). Future nuclease assays could be performed with a different radioactive labeling method

such as 3'-end or body labeling instead of the typical 5'-end labeling which could confirm the position of Cas3 nicking. The position of the band on cleavage assays might indicate the exact cleavage position of Cas3fv on the target substrate. Additionally, AMP-PNP could be used to “freeze” the helicase domain and determine the exact position of Cas3fv nicking.

The impaired exonuclease activity and generally minimal cleavage in these assays could be due to inhibition by Cas1-Cas2/3fv that is perhaps only able to nick substrates in the type I-F variant system but this difference compared to the I-F system would be surprising. Another explanation for this band is that assay conditions require further optimization.

The presumed structural alterations of Cas3fv could have occurred after fusion to the adaptation machinery. Alternatively, it is possible that Cas3 was replaced by a different protein that evolved into Cas3fv. As mentioned before, the driving force behind this change could have been viral Acrs such as AcrF3 known to inhibit recruitment of Cas3 to I-F Cascade (Rollins *et al.*, 2019).

The reason for the fusion of the adaptation machinery with the nuclease/helicase is a current subject of discussion. One initial suggestion of Richter *et al.* is that the fusion aids in the interference driven acquisition of new spacers that requires Cas3 in addition to Cas1 and Cas2 (Richter *et al.*, 2012). However, Cas3 nuclease and helicase activity were not involved in spacer capture studied by *in vitro* integration assays (Fagerlund *et al.*, 2017).

Rollins *et al.* also argue that the loss of activity without R-loop substrates would inhibit degradation of unrelated ssDNA in the cell and Cas1 would thus function as an anti-toxin to toxic Cas2/3 activity. As for why this is not conserved in all type I systems, they speculate that anti-CRISPR activity might have forced this diversification, considering that AcrF3 is capable of blocking adaptation and interference of type I-F CRISPR-Cas systems which might present an Achilles heel that forced the loss of this fusion (Rollins *et al.*, 2017). This theory would be in line with the structural alterations in I-Fv Cascade shown in this work.

Fusions of other *cas* genes are fairly common; *cas4* can be fused to *cas1* (in types I-B, I-U, and V-B) and *cas3* can also be split or fused to other *cas* genes (Koonin *et al.*, 2017). This highlights the evolution of ancestral proteins into dedicated Cas proteins during evolution. For example, Koonin and colleagues propose that Cas1 and Cas2 originally evolved from a toxin/anti-toxin pair (Makarova *et al.*, 2013). Besides the fusion of Cas2 and Cas3, the general mechanism of Cas3 interference is conserved in type I systems and Cas3 was suggested to have evolved from a generic nuclease/helicase to target foreign DNA (Nimkar & Anand, 2019).

3.3 Synthetic Cascade assembly and directed RNA-wrapping

The process of “RNA wrapping” by I-Fv Cas proteins and a repeat sequence was established and confirmed in this work. It is possible to form a complex of Cas proteins on an RNA molecule of choice by adding the 5'-repeat sequence upstream of the desired sequence. The repeat sequence on the produced transcript is processed by Cas6f which enables Cas5fv binding that in turn initiates Cas7fv backbone assembly on the tagged transcript. A second repeat sequence at the 3'-end as found in natural crRNAs is not required. The formed structures can be purified with an affinity tag fused on Cas5fv and the wrapped RNA later be extracted. The process works with various RNAs, including the here tested *sfGFP*, *lacZ-alpha* and a non-coding RNA construct.

While RNA-seq analysis and the decrease in sfGFP production confirm the specificity of this process, it is not possible to create complexes on an entire reporter gene transcript and there seems to be an apparent maximum length of ~ 100 nt of the directed Cascade assemblies. Additional confirmation of this length limitation was obtained by the inability to purify a complex with a second repeat tag on the 3'-end of the transcript for Cas6f binding, which would theoretically create a highly extended Cascade complex.

If no repeat region is present and Cas proteins are overproduced, Cas proteins and RNA are still co-purified with His-tagged Cas5fv. Sequencing of this RNA shows no-sequence bias and reveals only general RNA from the *E. coli* transcriptome. Fluorescence microscopy in our laboratory indicates that Cas5fv can bind RNA in general, when RNA is massively produced and no repeat is available (Müller-Esparza, 2019). This would not be an issue under wild-type conditions because Cas proteins are only produced in small amounts.

The *E. coli* genome was also investigated for natural 5'-handle sequences that could lead to unspecific RNA wrapping and 17 sequences were found. Upon closer inspection, a minor peak can be found in the mapping profile of obtained reads to the genome which starts directly at one of these 5'-handle sequences located in a non-coding area. However, this peak only reaches a maximum of ~ 100 reads in comparison to other unspecific peaks with a maximum of up to 1,000 reads. Additionally, in this sequencing, only ~ 40,000 reads were mapped to the entire genome from a total of ~ 900,000 reads in the entire sequencing and most of the reads match the 5'-terminal portion of the repeat-tagged *sfGFP*. The majority of other unspecific reads were thus mapped to unrelated positions in the genome. On top of this, no sequences can be found that match the full 28 nt long repeat sequence required for 5'-end processing and initiation of RNA wrapping. It can be assumed that the absence of a full repeat sequence limits the availability for unspecific initiation even further in addition to the generally lower transcription of this random sequence compared to the repeat-tagged target.

In the sequencing of the T7 RNAP produced *sfgfp* target, a minor additional peak is noticeable in the mapping profile that started with a two amino acid difference to the wild-type 5'-handle. Considering that this construct is highly produced, it could be sufficient for a second Cas5fv binding site and initiate RNA wrapping although with much lower efficiency compared to the wild-type repeat sequence. It should be noted that this peak is not present in previous sequencing of other repeat-tagged *sfgfp* constructs due to a different coding sequence on these plasmids compared to T7 RNAP produced *sfgfp*.

One theory for the length limitation was that the expression conditions were not optimal or more specifically, that it is necessary to provide individual proteins and RNAs in an optimal ratio. To compensate for the necessary overabundance of Cas7fv in a complex spanning an entire reporter gene transcript, we re-cloned *cas5fv* and *cas6f* on a plasmid with a lower copy number. However, RNA-seq results did not show an increased length of wrapped RNA, even when *cas7fv* was expressed from both MCS of the high copy expression plasmid.

Another investigated reason for the length limitation was that the beginning translation on the transcript would interfere with the assembly process. Transcription and translation are coupled processes that occur consecutively (Proshkin *et al.*, 2010, Landick *et al.*, 1985). Under special conditions, this can even lead to the formation of a transcribing and translation “expressome” complex (Kohler *et al.*, 2017). The ~ 100 nt space could thus represent the space between RNA polymerase and the ribosome. We first investigated this by removing the RBS but filaments observed on TEM showed the same typical length. We later optimized the experimental design by also removing every possible start-codon but RNA-seq analysis confirmed no positive effect on RNA wrapping and excluded this reason. To completely remove the possibility that *E. coli* RNAP is coupled to ribosomes, even in the absence of translation, we changed the promoter for target expression to the T7 promoter. While this didn't remove rRNA contamination, an improvement in the amount of RNA wrapping complexes was observed. We realized that standard Cascade production was most ideal when the crRNA was produced with T7 RNAP from the particularly high copy number vector, pUC19. This setup ensures that enough crRNA is available to form a complex with the Cas proteins. The limited amount of transcript produced by *E. coli* RNAP was apparently not sufficient for assembly to be detected as a peak during size-exclusion chromatography. Instead, the massively overproduced Cas proteins formed filaments and other byproducts. We initially chose the original setup for independent expression for the silencing experiment but it seems that this was counterproductive to complex formation and purification. It should be noted that the process itself worked even with the low amount of transcript, so silencing of *sfgfp* was immediately apparent and wrapped RNA could be isolated and detected by RNA-seq.

When the highly produced Cas proteins were able to form more complexes on the now appropriately available repeat-tagged transcript, the Cascade-like structures became detectable during purification

and with TEM. This also seemed to decrease the production of filaments which were not visible in the latest purifications. This could also be related to $MgCl_2$ omission, because a clear peak in the void volume disappeared when the repeat-tagged non-coding construct from previous setups with *E. coli* RNAP production was co-purified with Cas proteins. Future experiments will need to investigate if filaments are still produced in a very minor amount, or if only some aggregated full-length transcript is in the void volume of this purification.

The purified RNA wrapping complexes exhibit a form that is more similar to the “shrimp”-like type I-Fv Cascade, in contrast to the strict helical shape of filament structures. This shape of these structures is highlighted by the obtained 3D model from 2D class averaging of TEM pictures. Both, the 3D model and the original structures on TEM are not distinctively bigger than wild-type I-Fv Cascade. However, this size does not completely match the wide range of molecular weight calculated for the peak during size-exclusion purification.

It is not clear yet, how the wrapped RNA fits in this artificial Cascade-sized complex. A possible explanation for this is that not the entire structure was captured by TEM. After all, the peak in the elution volume of size-exclusion chromatography was very broad which indicates inhomogeneity and the image quality was only sufficient to see clear structures in the fraction at the right side of this peak which corresponds to an elution volume at the lower end up to 400 kDa. The estimated molecular weight of complexes eluting in the first half of this peak would be greater, with a maximum of 700 kDa, which correlates to ~ 19 subunits of Cas7fv/Cas5fv and one ~ 100 nt RNA molecule. This would be still smaller than the typical 100-200 nm filament structures from previous purifications with $MgCl_2$. Also, no clear filament structure like from previous purifications was identifiable in this fraction (Supplementary Figure 5). Still, even the visualized complex from the right side elutes at an elution volume corresponding to a higher minimal molecular weight than the natural I-Fv Cascade complex with less than 300 kDa. Even if the maximum of 8 Cas proteins can truly be fitted in the model, this complex would have a size of slightly more than 300 kDa.

Similarly shaped structures of about the same size were visualized by TEM in the elution fractions of salt gradient during the MonoQ purification step (Supplementary Figure 6). However, these structures were a bit more aggregated and the size was not easily identifiable. It is also possible, that these are fragments of larger filaments or Cascade-like structures that were broken apart by the high salt conditions.

Another possibility is that the model generated by 2D class averaging also did not present us the maximum length of these complexes considering the many orientations observed. In comparison to the previously observed filament structures, these smaller Cascade-like complexes are more flexible. On TEM pictures itself, they lie in various orientations resulting in different shapes in this 2D view, ranging

from an S-like structure to a closed ring (Figure 2.28 B). In the model created by 2D class averaging, this visualizes itself as a sideways turn, compared to the strict helical nature of filaments and the more compact form. This is the case for filaments from previous purifications that were analyzed with TEM as well as for the RNA-free Cas7fv helix obtained during crystallization (Figure 2.18, Figure 2.19). RNA-binding by the Cas proteins could potentially loosen the strict helical form and bend the structure to the typical “shrimp”-like form of Cascade. Without capping at the 3'-end and due to their longer backbone, these flexible structures can lie in various orientations. The most likely explanation for the sudden stop in Cascade assembly in the range of the first ~ 100 nt is a steric clash of one end of this complex with the existing structure (for example in the closed ring form).

Although the full and definite structure of the directed RNA wrapping complexes is most likely not obtained yet, it can be assumed that the full 100 nt wrapping complexes feature the same architecture. Additionally, RNA purified from these complexes and their analysis on high percentage gels suggests a variety of complexes with different sizes. A first distinct band was confirmed to have a size of ~ 80 nt, while a second distinct band corresponds to RNA with a size of ~ 100 nt. Both these fractions match the peak in RNA-seq results and this also explains the broadness of the peak during size-exclusion chromatography. Nanopore sequencing also suggests a moderately longer length in the total pool of available transcripts with a decreasing abundance the longer the length of the transcript. Synthetic Cascade assemblies might be formed on a repeat-tagged RNA of up to ~ 250 nt, although with a continuously lower chance for the assembly process to cover this size of RNA.

The few full-length transcripts produced were apparently bound to the MonoQ column and separated by ion-exchange chromatography. The UV absorbance ratio suggests that this is not bound to protein complexes from Cascade-like structures. Most likely, by increased target production rate with T7 RNAP, some transcript escaped complex formation and translation. Another possibility for the presence of full-length transcript is that while complex assembly already started, it breaks off perhaps by crashing into a roadblock for example by ribosomes. In this case, a long RNA strand would only be bound at the 5'-end by Cas5fv and perhaps some Cas7fv and thus co-purified. In combination with the slightly larger transcripts identified by Nanopore sequencing, a possibility is that the remaining part of larger RNA molecules sticks out of the end of the fully formed complexes and thus not wrapped but still co-purified.

A remaining possibility to cover a large RNA molecule would be to provide this molecule in very low numbers before the *cas* genes are expressed and not inhibit translation. It could then potentially be covered by Cas proteins completely. It cannot be guaranteed that an ordered assembly reaction would take place on this large construct.

To confirm the maximum length of RNA in the synthetic Cascade-like complexes, it will be necessary to purify full Cascade complexes with Cas6f at the 3'-end that include crRNAs with a spacer length close to the observed ~ 100 nt limit. By step-wise increasing the spacer length by 6 nt for one Cas7fv subunit, it would be possible to confirm the maximum number of Cas7fv subunits in the Cascade backbone (Figure 3.3). The addition of Cas6f at the 3'-end would guarantee a defined end of the complex, ensure stability and also allow for purification with a second affinity tag.

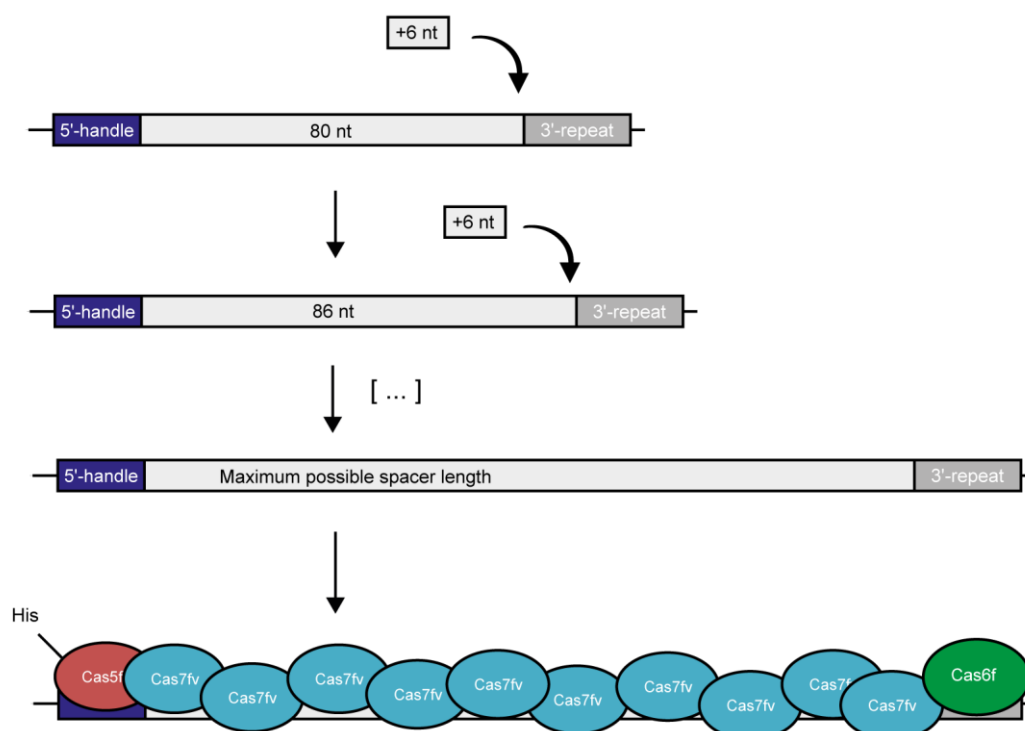


Figure 3.3: Schematic representation of step-wise crRNA spacer extension. Starting from an ensured possible spacer length such as 80 nt (a approximately the position of the highest count of mapped reads during RNA-seq analysis and size of a defined band during small RNA Urea-PAGE), the crRNA length is increased in steps of 6 nt for 1 Cas7fv subunit at a time. Multiple constructs with varying length are created and used for Cascade assembly. Purification is then performed to identify formed Cascade complexes. The first construct that does not form a Cascade complex marks the definite end of possible spacer extension. A 3'-repeat sequence for Cas6f binding and full Cascade assembly is included to ensure stability of these complexes.

3.4 Nature of filaments and rRNA contamination

The discovery of Cascade-like assemblies containing small RNA and the removal of ribosomal RNA elucidated how RNA is wrapped in this process. However, the nature of filament structures is not fully clear yet.

Filament structures have been previously purified with a His-tag fused to Cas7fv and the general co-purification of large RNA molecules has led to the assumption that they are formed on random RNA (Dwarakanath, 2015). We still obtained these structures when we switched the His-tag from Cas7fv to Cas5fv, so some form of interaction of Cas5fv with these structures is apparent and we assumed that

they are starting with a Cas5fv subunit. Cas5fv was never really detected on SDS-PAGE though, which we thought to be due to overlapping bands or due to the overrepresentation of Cas7fv in these structures.

The 3D structure of a small Cas7fv helix obtained during crystallization of I-Fv Cascade Cas7fv matched the purified filamentous structures observed on TEM and the later created model by 2D class averaging. This crystallized helix also does not contain RNA and while it is possible that RNA was pulled out during crystallization, it is also possible that the purified filaments are empty and simply oligomers caused by overproduction of Cas7fv. Recombinant overexpression in *E. coli* compared to minimal wild-type expression in *S. putrefaciens* CN-32 might have also increased the amount of unspecific interaction and by-product formation.

A recent crystallographic study elucidated the 3D structure of the Cas7f protein from *Zymomonas mobilis* (ZmCsy3), forming a molecular helix and filamentous structure in the crystalline state (Gu *et al.*, 2019). The model of this helix looks generally similar to the Cas7fv helix provided in this work, exhibiting a hollow cleft through the structure by the concave palms and with a positively charged cleft to the solvent by the extended regions (here “extended web” instead of wrist-loops). In contrast to the Cas7fv helix in this work, the ZmCsy3 helix requires seven instead of eight subunits for one full rotation and is thus slightly shorter in comparison (Figure 3.4). In addition, the ZmCsy3 helix also appears moderately compressed with a distance of 82 Å between coils in comparison to ~130 Å. However, this might also be due to the packing in the crystal form.

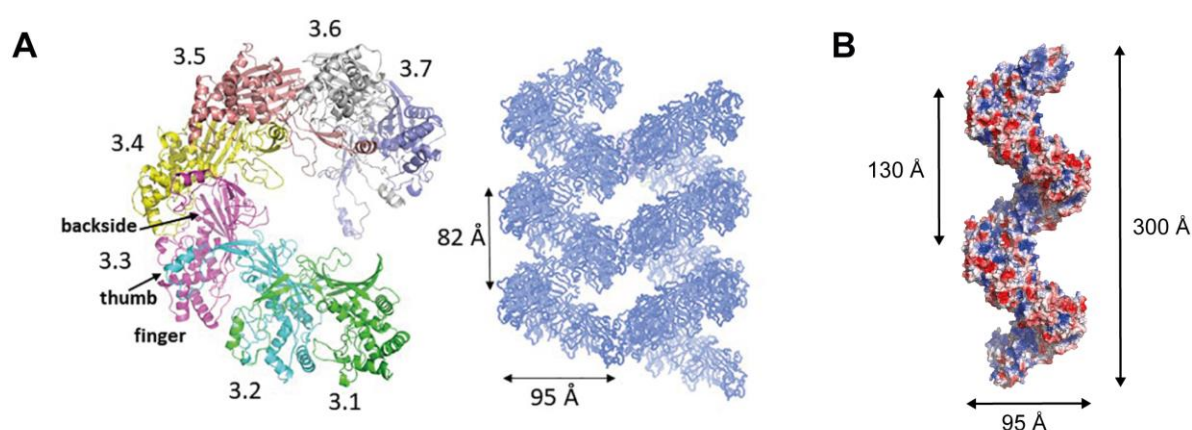


Figure 3.4: Comparison of filament structures. (A) Crystal structure of ZmCsy3 (Cas7f from *Zymomonas mobilis*). Left: side view of the molecular helix formed by seven Csy3 molecules in the asymmetric unit of the crystalline state. Each molecule is differentiated by colors and labeled from 3.1 to 3.7 to indicate the first to 7th Csy3 molecules. Right: The filamentous structures formed by Csy3 molecules. The symmetry-related molecules are displayed with coils. Figure from Gu *et al.*, 2019. (B) Crystal Structure of Cas7fv helix from this work (see 2.2.2.3 and Figure 2.19).

Gu *et al.* also claim that the molecular helix is formed in the absence of crRNA in the crystalline state and suggest the possibility for ZmCsy3 to aggregate at high concentrations to form a molecular backbone without crRNA. In turn, this might be an indication for the backbone to self-assemble before binding crRNA and other subunits. This research further indicates that the filament structures we always observed do not contain the specifically wrapped RNA and they are only an unspecific product of the purification. In the first performed RNA wrapping experiments, filaments would only be formed in case of low amounts of available repeat-tagged targets compared to massive overproduction of Cas7fv, which is not a native condition. In this case, instead of forming a Cascade-like complex, Cas7fv would aggregate to the long helical filament structures.

Besides the RNA-free helices from type I-F and type I-Fv, the length of all filaments visualized with TEM was mostly consistent and did not match the length of any repeat-tagged RNA construct potentially inside RNA. Specifically, filaments purified from cultures expressing the *sfgfp*-half construct looked identical to filaments obtained when full repeat-tagged *sfgfp* was expressed. The few observed filaments with a longer length could be overlapping structures. Small repeat-tagged RNA was only extracted from Cascade-like structures. Identical looking filament structures with the same lengths were also co-purified during recombinant production of type IV crRNP complex in *E. coli* (Ozcan *et al.*, 2019). All this combined, highly suggests that filaments are not formed on RNA.

Another direct comparison can be made with somewhat filamentous structures of Cas7 from the type I-C system (Hochstrasser *et al.*, 2016). These structures also appear to be filamentous, with a similar size compared to the Cas7fv filaments in this work. However, they feature a much more open and less compressed configuration (Figure 3.5). Filaments of Cas7fc are stated to be formed on a 44 nt RNA after incubation and presumably by bridging adjacent RNA molecules together which would explain their large size compared to a type I-C Cascade. Either these structures are also a product of oligomerization of Cas7 without RNA or they are more similar to the synthetic Cascade assemblies produced in this work, perhaps presenting something closer to the larger complexes we couldn't visualize yet.

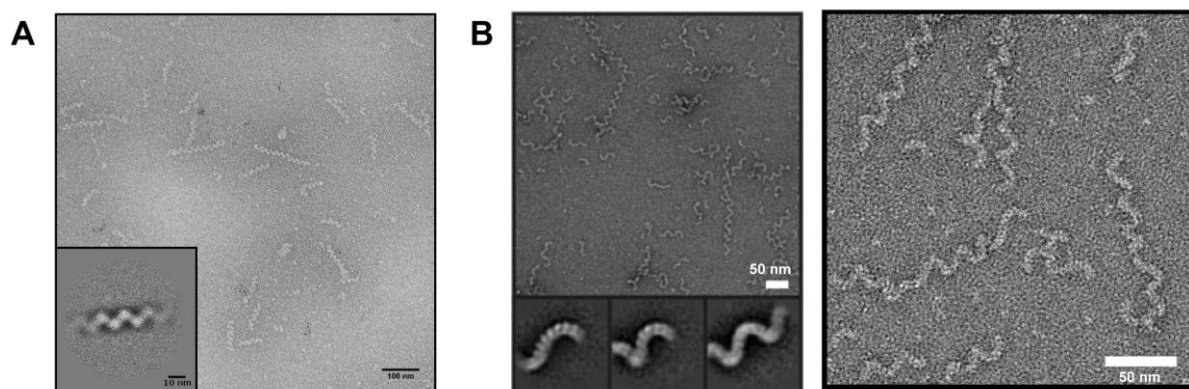


Figure 3.5: Comparison of filament structures from type I-Fv and type I-C visualized with TEM. (A) Filamentous structures from the void volume of Cas protein purifications in this work (see section 2.2.2.3 and Figure 2.18). **(B)** TEM analysis of structures from oligomerized type I-C Cas7 protein, obtained after incubation of monomeric protein with RNA. Figure modified from Hochstrasser *et al.*, 2016.

Because RNA was always co-purified with these proteins, we were previously unable to locate the position of this RNA, specifically if it is bound by the central RRM of Cas7fv or sticking on the outside. Only when rRNA (and other additional co-purified RNA) was removed, we were able to separate small RNA containing complexes and full-length transcript.

Ribosomal RNA was commonly purified before we removed MgCl_2 from the purification buffer which indicates that this component was essential for rRNA contamination. Mg^{2+} ions are essential co-factors for ribosomes and an increased MgCl_2 concentration can lead to their stabilization and co-purification (Nierhaus, 2014). While Mg^{2+} is naturally occurring in the cell, their addition to the wash buffer could have pulled out rRNA or entire ribosomes. However, ribosomal proteins were not detected on SDS-PAGE arguing that only rRNA was co-purified. Otherwise, it might have been possible that ribosome binding to the repeat-tagged transcript might be the reason for co-purification.

Ribosomal RNA proved difficult to remove and no method (RNase I treatment, size-exclusion, strep-affinity or salt-wash) showed any noticeable effect which speaks for a strong interaction between rRNA and Cas proteins. We also hypothesized that rRNA was in the filament structures due to the unique positively charged wrist loops of Cas7fv. However, the co-purification of repeat-tagged RNA and Cas proteins with mutated and neutralized amino acids still contained rRNA. It is still unclear, why rRNA was not detected by Illumina RNA-seq and only by Nanopore sequencing. It is likely that this was a problem with fragmentation, which is circumvented by Nanopore sequencing. These results are intriguing because rRNA usually is a common contaminant during RNA-seq with Illumina.

3.5 Natural reasons for limitations in size of Cascade assembly

The hypothesis that Cascade assembly can be specifically directed with a repeat sequence on an RNA of choice, was made due to the previous research on modulation of the I-Fv Cascade backbone in which its size was increased or decreased by including differently sized crRNAs (Gleditsch *et al.*, 2016). This modulation is based on the general RNA binding ability of Cas7fv that enables complex formation on random RNA sequences in the first place.

Recently, this possible modulation has also been shown for the type I-E system. Altered length is here achieved by including less or more of the small subunits in the complex, in addition to the backbone forming protein Cas7 (Luo *et al.*, 2016, Kuznedelov *et al.*, 2016, Songailiene *et al.*, 2019). In these experiments, spacers were extended with up to 24 additional nt, corresponding to 4 additional subunits of Cas7e and 2 additional small subunits. However, the maximum length was not disclosed.

In the newest research that provided the ZmCsy3 helix from a type I-F system, it was shown that monomeric ZmCsy3 forms different oligomeric states depending on the length added crRNA which is in line with other research that investigates the Cascade assembly and size of complexes. However, no other Cas proteins were added in these experiments.

In the minimal type I-C systems, where the Cas5 protein is additionally responsible for crRNA maturation, substituting Cas6, Cas7 oligomers have been identified on RNA without 5'-handle and/or 3'-hairpin (Figure 3.5) (Hochstrasser *et al.*, 2016).

Type III systems are known to bind crRNAs of different length by Csm3 or Cmr4 that act as ruler proteins spanning the crRNA in 6 nt increments and effective crRNAs of varying length are found in nature (Hatoum-Aslan *et al.*, 2013, Hale *et al.*, 2009).

The general question arises, why it is possible to modulate these crRNP backbones based on crRNA length. The multi-subunit arrangement of class 1 systems is different from single-effector class 2 systems that are not able to bind extended RNA spacers. However, it was shown that the length of guide RNA can be moderately reduced for up to 3 nt in *Streptococcus pyogenes* Cas9, which increased specificity and reduced off-target effects (Fu *et al.*, 2014).

Kuznedelov *et al.* argue that type I CRISPR effectors might have evolved from much simpler ancestral complexes with as little as one Cas7 subunit to the most optimal arrangement of six Cas7 subunits for tighter interaction with target DNA (Kuznedelov *et al.*, 2016). The length of spacer CRISPR RNA is also influenced by the preferred size of protospacers of the Cas1-Cas2 acquisition complex. Songailiene *et al.* suggest that the optimal length despite the modular architecture of these complexes might be the result of their coevolution along Cas1 and Cas2 (Songailiene *et al.*, 2019).

Luo *et al.* have shown that type I-E Cascade complexes with spacers extended up to 24 additional nt exhibited increased silencing levels in a CRISPR interference assay (Luo *et al.*, 2016). However, elongated complexes were shown still form R-loop structures on target DNA of the same length and while target DNA binding affinity was increased, Cas3-mediated cleavage did not require extended R-loops (Songailiene *et al.*, 2019). Likewise, no positive effect on interference was observed for extended I-Fv Cascade complexes with up to 18 additional nucleotides during *in vivo* interference assays (Gleditzsch *et al.*, 2016).

3.6 Applications

A variety of applications for the established RNA wrapping process were investigated. Silencing translation of reporter gene transcripts and protecting RNA from degradation while encased in the Cas7fv backbone was established and additionally, we have provided indications that the backbone of synthetic Cascade-assemblies can be modified. Research on additional potential applications, namely packing of RNA in complexes *in vitro* and the induced unpacking of complexes was initiated. These applications will be discussed in greater detail in the following sections.

Silencing of reporter gene transcripts:

Cas5fv-Cas7fv-mediated silencing of reporter genes was confirmed by the color expression cultures, fluorescence microscopy and FACS. Silencing takes place on the level of translation after transcription. The transcript becomes inaccessible for ribosome binding and translation due to Cas7fv binding and blocking of the RBS and the following sequence.

Quantification of the silencing measured by FACS revealed a ~ 95.3 % sfGFP reduction when comparing the control to the repeat-tagged construct produced by T7 RNAP overnight at 18 °C. Comparison of the repeat-tagged control expressing no *cas* genes to the same construct expressing only *sfgfp* and *cas* genes simultaneously results in sfGFP reduction of only ~ 91.1 %. This indicates that some transcript will always escape wrapping by Cas proteins and will be translated.

Silencing is less strong when *sfgfp* was expressed before *cas* genes. This further indicates direct coupling of transcription and translation. If transcripts are produced ahead of time, they will be converted to sfGFP and Cas proteins will only form complexes on freshly produced mRNA.

High levels of Cas protein production seem to cause major stress as apparent by the elongated shape of cells detected with fluorescence microscopy and FACS. This might be caused by *cas* gene overexpression in general or by a toxic effect due to unspecific RNA binding of Cas7fv. The general ability of Cas7fv to bind RNA *in vitro* without Cas5fv initiation has been shown by EMSA analysis (Dwarakanath, 2015).

Sequencing of RNA extracted from the control construct without repeat-tag shows a variety of RNA. Even though the addition of the repeat-tag shifts this relation to the 5'-terminal portion of the tagged construct, this unspecific binding should be investigated further.

In contrast to established CRISPR interference (CRISPRi) assays on a DNA level, the level of silencing is moderately weaker. For example, a 99.9 % silencing efficiency was obtained with catalytically dead dCas9 (Larson *et al.*, 2013). CRISPRi was also established with DNA targeting of type I-E Cascade in absence of Cas3, producing similar levels of silencing (Rath *et al.*, 2015). It is also predicted that Cascade-mediated CRISPRi is accompanied by less off-target effects compared to dCas9.

Silencing experiments with RNA-targeting single effectors such as Cas13 have focused on knockdown by degradation of the RNA target and managed a silencing rate of up to 95 %. For these *in vivo* assays, reporter production can be simply reversed by stopping Cas protein production, reinitiating transcription and translation. A strong focus in the applied research in these systems is RNA-editing (Cox *et al.*, 2017) and nucleic-acid detection (Gootenberg *et al.*, 2017).

The well-established eukaryotic RNA interference mechanisms also achieve silencing by degrading foreign RNA but are characterized by low efficiency compared to CRISPRi (Hannon, 2002, Zamore *et al.*, 2000).

A topic of current research is the degree of off-target effects with the Cas9 and CRISPR-Cas systems in general. In type I interference, R-loop formation depends on the presence of PAM and the seed sequences while the specificity of the presented RNA wrapping process is based on the presence of a repeat-tag. Mismatches in the repeat-tag were shown to not be tolerated in other type I systems and even a single mismatch was shown to disable Cascade formation (Beloglazova *et al.*, 2015). As mentioned before, a single of natural 5'-handle sequence in the genome of *E. coli* resulted in an extremely small peak in the mapping profile while a comparably minor peak was present in the sequencing of the T7 RNAP produced repeat-tagged transcript with a two amino acid difference to the wild-type 5'-handle. These peaks indicate very minor off-target activity that should be further investigated. In general, this system has proven to be very specific for silencing.

To summarize, most CRISPRi methods are based on silencing transcription while the presented method of RNA wrapping inhibits translation. In contrast to other RNA targeting systems, RNA wrapping in the Cascade complex leaves the RNA intact, which can be a potential advantage such as faster reversibility and return to standard conditions. While degradation could also be avoided with an inactivated dCas13, Cascade complexes or Cascade-assemblies without Cas6f can encase longer sequences.

RNA Protection:

In addition to the basic application for specific RNA isolation by providing a repeat, we have shown that RNA content is also protected from degradation even when incubated with RNase I.

These experiments were performed before rRNA contamination was removed, so only a smear was visible on normal RNA gels. This entire smear remained after incubation with RNase I, which indicates that this RNA is highly structured and inaccessible for RNase I degradation as well. However, it is unlikely that this unspecific rRNA is protected by being wrapped in Cascade-assemblies in the same fashion as the specific repeat-tagged RNA. In future assays from purifications without MgCl₂, this option should be excluded.

No exact size determination was possible for the Northern Blot analysis but the obtained signal fits in the range of the usual 5'-terminal portion that was observed by RNA-seq and in extractions with T7 RNAP produced constructs.

Noticeable precipitation was visible in the incubated samples, likely by the unstable co-purified byproducts such as Cas5fv-Cas7fv dimers that were not removed from the Ni-NTA fraction in this experiment. Northern Blot confirms that the RNA that is wrapped in a Cascade-like complex is protected for numerous days. The missing capping of Cas6f does not seem to affect stability and the sample is apparently as stable as normal Cascade complex, although this would need to be confirmed directly for comparison. In type I-C Cascade, crRNA processing is performed by Cas5c in addition to 5'-end capping and the 3'-end is only capped by the crRNA stem-loop (Hochstrasser *et al.*, 2016). In general, Cascade complexes evolved to be highly stable as they are required to constantly scan the cellular DNA content for foreign DNA. With these observations, the specific RNA wrapping can be used as a tool to stabilize the specifically isolated repeat-tagged RNA.

***In vitro* RNA packing:**

Because of the stabilizing effect on RNA, it would be especially useful to be able to pack RNA in Cascade-like complexes *in vitro*. Ideally, mixing monomeric apo-Cas7fv with RNA would create these complexes. We focused on this approach due to its envisioned simplicity. Alternatively, *in vitro* complex formation could require Cas5fv and a repeat tag which might also be generally desirable even if random RNA cannot be stabilized in an *in vitro* setting.

Unfortunately, it was not possible to assemble synthetic Cascade assemblies *in vitro*. Previously, we were also not able to use the Cas5fv-Cas7fv dimer for *in vitro* assembly of a Cascade complex (Gleditsch, 2015) and likewise, we were not able to create Cascade-like complexes in this work. It can be assumed that the dimer of these two proteins is not ideal for this application because the interaction

is too strong for later assembly. We did not use small RNA for this *in vitro* wrapping approach, which would represent a remaining option. The inability to form complexes with the Cas5fv-Cas7fv dimer argues against the hypothesis that the dimer transports Cas7fv to the crRNA.

A helix of Cas7fv subunits was obtained by chance during crystallization of I-Fv Cascade. This presented an intriguing approach for obtaining apo-Cas7fv, especially because I-Fv Cascade can be purified easily. However, while crystals were obtained in small-scale drop format, upscaling was not possible and we discontinued this approach to focus on purification of soluble Sumo-Cas7fv. Nonetheless, future attempts could focus on improving crystallization conditions with the aim of obtaining crystalline Cas7fv in higher amounts. After all, this approach of purifying protein on a large scale is commonly used in existing biotechnological protocols such as purification of insulin (Mirsky *et al.*, 1963).

While purified Sumo-Cas7fv can bind RNA *in vitro* as shown by EMSAs in previous work (Dwarakanath, 2015), we were not able to obtain clear structures with this approach with both small and long RNA. Purification of Sumo-Cas7fv with an included high salt wash seems to be the most ideal approach to obtain soluble protein. A problem might be the need for coordinated Cas7fv addition on the backbone that is most likely happening *in vivo* directly after transcription. Protein needs to be directly provided on RNA for coordinated backbone formation and to avoid oligomerization, byproducts like Cas5fv-Cas7fv and perhaps inhomogeneous RNA plus Cas7fv complexes.

One benefit of the *in vitro* approach compared to *in vivo* conditions we hoped for, was to use *in vitro* incubation to wrap longer RNA molecules. However, so far, it cannot be confirmed that this is possible. The obtained structure of purified Cascade-like assemblies indicates that the process only works effectively with small RNAs up to ~ 100 nt. If *in vitro* RNA wrapping can be achieved, it is more likely to work with small RNA and further attempts should focus on this. Even more ideally, for proof of principle, a specifically designed small RNA without secondary structures should be used.

In vitro packing of RNA in Cas proteins should be further investigated due to the usefulness of this potential application. Monomeric Cas7 from the type I-C system has been purified and claimed to be used for the formation of complexes on RNA (Hochstrasser *et al.*, 2016). Additionally, there has been successful research performed on *in vitro* assembly of related Cascade structures with existing protocols that are based on isolation and solubilization of Cas proteins from inclusion bodies with subsequent refolding on RNA (Plagens & Randau, 2015, Plagens *et al.*, 2014). Alternatively, Beloglazova *et al.* examined crRNA loading of empty Cascade purified from *E. coli* (Beloglazova *et al.*, 2015). These approaches could be transferred to the I-Fv system and synthetic Cascade-assemblies.

Induced unpacking of RNA from Cas protein complexes:

Another potential application we started to investigate is the induced unpacking of RNA from the created complexes *in vivo*. While complexes can be degraded *in vitro* by standard RNA extraction protocols such as phenol/chloroform, the induced release of RNA from stabilized complexes in the cell could prove immensely useful e.g. for therapeutic purposes.

Our first attempt to achieve this was to use a unique Acr protein that has been shown to remove the backbone of type I-F Cascade (unpublished). The benefit of this full CRISPR-method would be that the backbone complex does not need to be further modified and that the Acr protein can be easily induced and expressed for unpacking of the RNA from the complexes. Because *in vivo* transformation assays, showed that this Acr does not target the type I-F variant system, we attempt to create RNA wrapping complexes with the type I-F Cas proteins and repeat, something which had not been investigated yet.

During purification, we were not able to detect a specific complex on the repeat-tagged non-coding RNA. Even though this experiment failed, our improvements in the co-expression of the target RNA and I-Fv Cas proteins, performed in the meantime, could also be applied to type I-F. Future configurations of this co-expression could be helpful to rule out I-F complexes as potential candidates for a packing/unpacking system.

The second investigated approach was based on the *Mesoplasma florum* transfer-messenger RNA system that utilizes protein degradation tags in combination with a specific and separately inducible protease that recognizes these tags (Gur & Sauer, 2008). This system has been successfully used for specific and tunable protein degradation (Cameron & Collins, 2014). This approach for unpacking of repeat-tagged RNA remains promising but more research is required. Complexes could not be purified because Cas5fv and Cas7fv became insoluble and were only detected in the pellet of homogenized cells. It is unexpected that the addition of the protein degradation tag poses a problem for the Cas proteins because we have created RNA wrapping complexes with sfGFP fused to Cas7fv and Cascade has been purified with the Dendra-protein fused to the C-terminus of Cas5fv (Müller-Esparza, 2019). The 3D structure of I-Fv Cascade also shows that the C-terminus, which is required for the protein degradation tag, is accessible. The solubility of the pdt-Cas proteins might improve by increasing the linker length. Additionally, the fusion of protein degradation tags to Cascade instead of synthetic wrapping complexes should be investigated.

Backbone modification:

Finally, the backbone of the directed Cascade assemblies can be modified, at least partially, by fusing a reporter on the Cas7fv backbone. It is unclear why this seems to be a partial modification and both

unmodified Cas7fv and sfGFP-Cas7fv were detected in the backbone. One explanation might be limited space on the backbone, especially when the flexible complex has an unfavorable orientation such as a closed ring in which both ends of the complex clash. To confirm this, we need to further investigate the exact position of the reporter in the modified backbone. Backbone modification of the synthetic complexes could enable the use of a variety of proteins, comparable to e.g. CRISPR activation (Maeder *et al.*, 2013, Perez-Pinera *et al.*, 2013). We could also envision backbone-modification to help for transport in the cell on a directed RNA of choice. Comparable to single-effector CRISPR-Cas systems, the multi-subunit type I-Fv system allows greater flexibility in modulation.

In conclusion, the presented RNA wrapping process provides a variety of useful applications. Foremost is the specific isolation of RNA by the addition of a repeat-tag. The wrapping process can be used to block translation and protect the incorporated RNA. The simplified Cascade backbone provides advantages such as greater flexibility and straightforward potential modification. More potential applications remain and some have been initiated but were beyond the scope of this work. However, they provide a direction for further research on this synthetic method.

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 the companies listed in Table 4.1.

Table 4.1: List of special chemicals and reagents used in this work.

Label	Manufacturer
2-log DNA ladder (0.1-10.0 kb)	New England Biolabs, Frankfurt
Acrylamide, N,N-methylenebisacrylamide	Roth GmbH, Karlsruhe
Amicon [®] Ultra Centrifugal Filters	Merck Millipore KGaA, Darmstadt
Ammonium Persulfate (APS)	Roth GmbH, Karlsruhe
Antarctic Phosphatase	New England Biolabs GmbH, Frankfurt
Antibiotics (kanamycin, ampicillin, spectinomycin, chloramphenicol)	Sigma-Aldrich, Taufkirchen ; Roth GmbH, Karlsruhe
ATP	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
ATP [γ - ³² P]	Hartmann Analytic GmbH, Braunschweig
Bacterial protease inhibitor	Roth GmbH, Karlsruhe
β -Mercaptoethanol	Sigma-Aldrich, Taufkirchen
Bradford Reagent	BioRad, München
Bovine serum albumin (BSA)	Sigma-Aldrich, Taufkirchen
Coomassie Instablu	Expedeon, Cambridge
Color prestained protein standard (10-250 kDa)	New England Biolabs GmbH, Frankfurt
Diethyl Pyrocarbonate (DEPC)	AppliChem GmbH, Darmstadt
Dimethyl Sulfoxide (DMSO)	Sigma-Aldrich, Taufkirchen
DNA Oligonucleotides	Eurofins MWG operon, Ebersberg
DNase I	Sigma-Aldrich, Taufkirchen
dNTP Mix	New England Biolabs, Frankfurt
Ethidium Bromide	Roth GmbH, Karlsruhe
Gel Breaker Tubes	IST Engineering, Milpitas, CA, USA
Gel Filtration Markers Kit for Protein Molecular Weights 12,000-200,000 Da	Sigma-Aldrich, Taufkirchen
Gelpilot DNA Loading Dye (5x)	Qiagen GmbH, Hilden
Glycogen	Roche Diagnostics GmbH, Mannheim
Illustra MicroSpin G-25 columns	GE Healthcare
Isopropyl- β -thiogalactopyranoside (IPTG)	Roth GmbH, Karlsruhe
Low Molecular Weight Marker	Affymetrix/USB [™]
Low Molecular Weight DNA Ladder	New England Biolabs GmbH, Frankfurt
Low Range ssRNA Ladder	New England Biolabs GmbH, Frankfurt

Lysozyme	Sigma-Aldrich, Taufkirchen
Millex® AP20 Syringe Filter, pore size 20 µm	Merck KGaA, Darmstadt
NEBNext® Small RNA Library Prep Set	New England Biolabs GmbH, Frankfurt
NEB Builder HiFi DNA Assembly Master Mix	New England Biolabs GmbH, Frankfurt
NTPs (ATP/GTP/CTP/UTP)	Jena Bioscience GmbH, Jena
Phenol/Chloroform Mix (acidic)	Ambion, Darmstadt
Phusion DNA polymerase	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
QIAprep Spin Miniprep Kit	Qiagen GmbH, Hilden
QIAquick Gel Extraction Kit	Qiagen GmbH, Hilden
QIAquick PCR Purification Kit	Qiagen GmbH, Hilden
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Qubit RNA HS Assay Kit	Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Quick-Load® 2-Log DNA Ladder (0.1-10.0 kb)	New England Biolabs GmbH, Frankfurt
Quick-Load® pBR322 DNA-MspI Digest	New England Biolabs GmbH, Frankfurt
Restriction endonucleases	New England Biolabs GmbH, Frankfurt
RNase I _f	New England Biolabs GmbH, Frankfurt
RNase Inhibitor (murine)	New England Biolabs GmbH, Frankfurt
Roti®-Nylon plus, pore size 0.45 µm	Roth GmbH, Karlsruhe
Sodium dodecyl sulfate (SDS)	Roth GmbH, Karlsruhe
SigmaPrep™ Spin Column Kit	Sigma-Aldrich, Taufkirchen
Sumo Protease	Own production
SYBR Gold® Nucleic acid stain	Thermo Fisher Scientific 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
Tetramethylethylenediamine (TEMED)	Sigma-Aldrich, Taufkirchen
ULTRAhyb-Oligo hybridization buffer	Ambion, Darmstadt
Whatman GB 004, 3MM	Schleicher & Schuell GmbH, Dassel
X-Ray CEA RP New film screen	CEA GmbH, Hamburg
ZelluTrans Dialysis tubes (6,000-8,000 MWCO)	Roth GmbH, Karlsruhe

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

4.1.2 Instruments

Table 4.2: List of Instruments used in this study

Instrument	Model and Company
Agarose gel electrophoresis	Chambers and Casting trays produced by 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, NL
Bioanalyzer	Agilent 2100 Bioanalyzer, Agilent, Santa Clara, CA, USA
Centrifuges	Centrifuge 5424, Eppendorf AG, Hamburg; Sigma 3-30K, Sigma Laborzentrifugen GmbH, Osterode am Harz; Sorvall Lynx 4000, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Chromatography columns	HisTrap HP 1 ml and 5 ml, StrepTrap HP 5 ml, HiLoad® 16/600 Superdex® 200 pg, Superose® 6 10/300 GL, Mono Q 5/50 GL anion exchange chromatography column; GE Healthcare Europe GmbH, Freiburg
Flow cytometer	BD LSRFortessa, BD Biosciences, Heidelberg
FPLC	ÄKTA purifier™ 10: Pump P-900, Monitor UV-900, Monitor UPC-900, Valve INV-907, Mixer M-925; ÄKTA pure L1: Pump P9 Cpl, Mixer M9, Injection valve V9-Inj, UV-Monitor U9-L Cpl; GE Healthcare Europe GmbH, Freiburg
Denaturing polyacrylamide gel electrophoresis	PROTEAN II Electrophoresis Chamber, BioRad Laboratories GmbH, Munich
Hybridization oven	Hybrid Shake 'n' Stack, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Incubators	KB53, Binder GmbH, Tuttlingen
Magnetic stirrer	IKA® RCT Standard, IKA®-Werke GmbH & Co. KG, Staufen
Magnetic Separation Rack	2-Tube Magnetic Separation Rack; New England Biolabs GmbH, Frankfurt
Microscope	Axioplan 2, Carl Zeiss Microscopy GmbH, Göttingen; CoolSnap HQ camera, Visitron Systems GmbH, Puchheim; FluoArc HBO Lamp, Carl Zeiss Microscopy GmbH, Göttingen
Microfluidizer	Microfluidics LM10, Sysmex Deutschland GmbH, Norderstedt.
MinIon	Oxford Nanopore Technologies, UK
MiniSeq	Illumina, Inc. USA
Nanodrop	NanoDrop® ND-1000 Spectrometer, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
PCR-Cycler	C1000™ Thermal Cycler, BioRad Laboratories GmbH, Munich
Peristaltic pump	Peristaltic Pump P-1, GE Healthcare Europe GmbH, Freiburg
pH-meter	INOLAB pH level 1, WTW, Weilheim
Phosphorimager	Storm 840 phosphorimager, Molecular Dynamics, GE Healthcare Europe GmbH, Freiburg
Qubit Fluorometer	Qubit 2.0, Thermo Fisher Scientific Ltd. & Co. KG, Bonn
Rocker	Gyrorocker SSL3, Sigma-Aldrich, Taufkirchen
Scintillation counter	Beckmann LS 6500, Beckman Coulter GmbH, Krefeld
Semi-dry transfer cell	Trans-Blot® SD Semi-Dry Transfer Cell, BioRad Laboratories GmbH, Munich
SDS polyacrylamide gel electrophoresis	Mini-PROTEAN Tetra Cell, Bio-Rad Laboratories GmbH, Munich; Power supply PowerPac Basic, Bio-Rad Laboratories GmbH, Munich
Sonicator	Branson Sonifier 250, Branson Ultrasonix, Danbury, CT, USA
Spectrophotometer	Ultrospec 3000 <i>pro</i> , GE Healthcare Europe GmbH, Freiburg
Thermomixer	Thermomixer Comfort 5350, Eppendorf AG, Hamburg
Thermoshaker	HT Thermotron, Infors AG, Bottmingen, Switzerland
UV-Crosslinker	UV Stratalinker® 1800, Stratagene, La Jolla, USA
UV-Transilluminator	BioDocd-IT system, UVP, Upland, CA, USA
Vortex Mixer	Vortex Genie 2, Scientific Industries, Bohemia, NY, USA

4.1.3 Buffers and solutions

Buffers, solutions and media are mentioned in the chapter of the respective method. All media, solutions and buffers were, if necessary, autoclaved for 20 min at 121°C prior to usage.

4.2 Strains and culture conditions

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

Strain	Description	Source
<i>Escherichia coli</i> K12 DH5 α	F- Φ 80 <i>lacZ</i> Δ M15 Δ (<i>lacZYA-argF</i>) U169 <i>recA1 endA1 hsdR17</i> (rK-, mK+) <i>phoA supE44</i> λ - <i>thi-1 gyrA96 relA1</i>	(Hanahan, 1983)
<i>Escherichia coli</i> Rosetta2 (DE3) pLysS	F- <i>ompT hsdSB</i> (rB- mB-) <i>gal dcm</i> (DE3) pLysSRARE2 (CamR)	Novagen, Darmstadt
<i>Escherichia coli</i> BL21(DE3) pLysS	F- <i>ompT hsdSB</i> (rB- mB-) <i>gal dcm</i> (DE3) pLysS (SpecR)	Novagen, Darmstadt

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 100 μ g/ml, kanamycin 50 μ g/ml, ampicillin 100 μ g/ml and chloramphenicol 34 μ g/ml) based on plasmid encoded antibiotic resistance gene.

E. coli DH5 α was used for cloning procedures. This strain transforms with high efficiency and has a number of features useful for cloning. *E. coli* BL21 (DE3) was used for expression cultures. This strain features the gene for expression of the T7 polymerase as well as spectinomycin resistance. Overexpression by the T7 promoter on transformed plasmids is repressed until IPTG induction from a lac promoter. *E. coli* BL21 (DE3) pLysS cells were used for the production of recombinant Cascade proteins, while Rosetta (DE3) pLysS cells were used for Cas1 and Cas2/3fv. Cultures were grown in LB media with respective antibiotics until an OD₆₀₀ of ~ 0.6 was reached. Expression of *cas* genes and crRNA or repeat-tagged target RNA from plasmids under control of the T7 promoter was induced by addition of 1 mM IPTG. Expression of target RNA from the pBAD plasmid was induced with 0.2 % arabinose. Cultures were grown overnight at 18 °C after induction. Cultures expressing SUMO-Cas7fv were grown for 3 h at 37 °C after induction and then harvested immediately.

Cells were harvested by centrifugation (8.000 rpm, 30 min, 4 °C) and either stored at -80 °C or immediately lysed for protein purification.

4.3 Plasmids and oligonucleotides

4.3.1 Plasmids

Table 4.4: Plasmids used in this work.

Vector	Resistance	Application	Source
pUC19	Amp ^R	crRNA production	NEB
pRSFDuet	Kan ^R	Protein production	Novagen
pACYDuet	Ca m ^R	Protein production	Novagen
pET21d	Kan ^R	Protein production	Novagen
petM-43	Kan ^R	Protein production	(Dwarakanath, 2015)
pEC-His-A	Kan ^R	Protein production	(Dwarakanath, 2015)
pBAD	Amp ^R	Target RNA production and protein production	Invitrogen

Table 4.5: Recombinant plasmids for protein production

Plasmid + Insert	Description
pRSFDuet + <i>cas7fv</i> + <i>cas5fv</i> -His (in MCS1) + <i>cas6f</i> (in MCS2)	C-terminal His-tagged <i>cas5fv</i> and <i>cas7fv</i> in MCS1 <i>cas6f</i> in MCS2
pRSFDuet + Δ AH- <i>cas5fv</i> , <i>cas7fv</i> , <i>cas6f</i>	deletion of aa K121-Y259 of <i>Cas5fv</i> and insertion of 6x Glycine/Serine linker
pRSFDuet + Δ Cas7fv-loop	deletion of aa 27-77 of <i>Cas7fv</i> and insertion of 6x Glycine/Serine linker
pEC-His-A + SUMO-His- <i>cas7fv</i>	<i>Cas7fv</i> with a 6x His-SUMO-tag (N-terminal) (Dwarakanath, 2015)
prsfDuet + <i>cas7fv</i> + <i>cas5fv</i> -His + <i>cas6f</i> -Strep	C-terminal His-tagged <i>Cas5fv</i> , <i>Cas7fv</i> and C-terminal Strep-tagged <i>Cas6f</i>
petM-43 + <i>cas2/3fv</i>	<i>Cas2/3fv</i> with N-terminal His-MBP-tag (Dwarakanath, 2015) with frame shift corrected
pet24d + <i>cas3fv</i>	<i>cas3fv</i> with <i>cas2</i> deleted (by Dr. Patrick Pausch)
prsfDuet + Strep- <i>cas1</i> + <i>cas2/3fv</i>	N-terminal Strep-tagged <i>Cas1</i> and <i>Cas2/3fv</i>
pACYCDuet + His- <i>cas5fv</i> + <i>cas6f</i>	C-terminal His-tagged <i>Cas5fv</i> and <i>Cas6f</i> from low copy pACYCDuet vector
pRSFDuet + <i>cas7fv</i>	One copy of <i>cas7fv</i> in MCS1 (untagged)
pRSFDuet + <i>cas7fv</i> + <i>cas7fv</i>	Two copies of <i>cas7fv</i> (one in each MCS)
pRSFDuet + WL-neutral- <i>cas7fv</i> + <i>cas5fv</i> -His + <i>cas6f</i>	wrist loops in <i>Cas7fv</i> neutralized (aa 62-67 exchanged with alanine)
pRSFDuet + <i>cas7fv</i> -pdt + His- <i>cas5fv</i> -pdt + <i>cas6f</i>	For pdt-tagged <i>Cas5fv</i> and <i>Cas7fv</i> proteins and untagged <i>Cas6f</i>
pBAD + <i>sfgfp</i> - <i>cas7fv</i> + <i>cas5fv</i> + <i>cas6f</i>	For sGFP- <i>Cas7fv</i> fusion protein, <i>Cas5fv</i> and <i>Cas6f</i>
pCsy_complex	Type I-F Cascade: <i>Csy1</i> , <i>Csy2</i> , <i>Csy3</i> and <i>Csy4</i> (from addgene ID 89232)

Table 4.6: Recombinant plasmids for RNA production

Plasmid + Insert	Description
puc19 + wt crRNA	T7 promoter + pre-crRNA with a 32 nt spacer4 from <i>Shewanella putrefaciens</i> CN-32 (Dwarakanath, 2015)
pBAD + RBS + repeat- <i>sfgfp</i>	Repeat-tagged <i>sfgfp</i> RNA including RBS
pBAD + RBS + <i>sfgfp</i>	Control construct without repeat-tag on <i>sfgfp</i>
pBAD + RBS + repeat-lacZ- α	Repeat-tagged lacZ- α RNA including RBS
pBAD + repeat + <i>sfgfp</i> + repeat	pBAD + RBS + repeat- <i>sfgfp</i> with a additional repeat (partial, no 5'-handle) for <i>Cas6f</i> binding
pBAD + repeat-sfGFP	pBAD + RBS + repeat- <i>sfgfp</i> with RBS removed
pBAD + RBS + repeat- <i>sfgfp</i> -half	pBAD + RBS + repeat- <i>sfgfp</i> the latter half of <i>sfgfp</i> deleted
pBAD + repeat-non-coding	500 nt complete non-coding region from pRSFDuet (without ATG) cloned in pBAD with repeat but no RBS
pBAD + Non-coding-repeat-control	pBAD + repeat-non-coding with and added RBS + 7 nt spacer sequenz + ATG
pETDuet1 + repeat + RBS + <i>sfgfp</i>	Repeat-tagged <i>sfgfp</i> RNA including RBS

4.3.2 Oligonucleotides

Table 4.7: Primers used for cloning

Name	Sequence 5'-3'	Description
ΔAH-Cas5fv-invPCR-fwd ΔAH-Cas5fv-invPCR-rev	GGCGGCAGCGGCGGCAGCACAAACGGGACCCAAA AAA GATCATTCCAGTAAATGCAT	Inverse PCR primer to delete alpha-helical domain of <i>cas5fv</i> and replacement with 6xGlycine/Serine linker
ΔCas7fvl loop-invPCR-fwd ΔCas7fvl loop-invPCR-rev	GGCGGCAGCGGCGGCAGCCTGTATATAAGTCAA AAT CCCATTCCAATCACTACGC	Inverse PCR primer to delete <i>cas7fv</i> WL-loops regions and replacement with 6xGlycine/Serine linker
sfGFP2xRepeat-invPCR-fwd sfGFP2xRepeat-invPCR-rev	TAAGTTCACCGCCGCACAGGCGGCTTAGAAAAGC TCGAGATCTGCAGCTG GCTGCCTTTATACAGTTCATCCATACC	Inverse PCR primer for addition of second repeat downstream of pBAD + RBS + repeat- <i>sfgfp</i>
Cas75-His-6-Strep-invPCR-fwd Cas75-His-6-Strep-invPCR-rev	TGGAGCCATCCGCAGTTTGAAAAATAACTCGAGT CTGGTAAAGAAACC AAACCAAGGACTGTAGCGG	Inverse PCR primer for addition of Strep-tag on C-terminus of <i>cas6f</i> in pRSFDuet + <i>cas7fv</i> + <i>cas5fv</i> -His (in MCS1) + <i>cas6f</i> (in MCS2)
lacZ-Repeat-fwd lacZ-Repeat-rev	GCACAGGCGGCTTAGAAAGAAGGAGATACCATG GCATGACCATGATTACGCCAAG GAATTCATATGGTACCAGCTGCAGATCTCGAG CTCTATGCGGCATCAGAGCAGA	Primer for cloning of lacZ-α in pBAD-Repeat, cut with NcoI and XhoI
minusRBS-fwd minusRBS-rev	ATGGTTAGCAAAGGTGAA GGTTTCTAAGCCGCTGT	Primer for removal of RBS from pBAD + RBS + repeat- <i>sfgfp</i> by inverse PCR
gfp-half-fwd gfp-half-rev	CGGCTGATAAAACAGAATT TTCAATGCGGTTACCAGGG	Primer for removal of the latter half of <i>sfgfp</i> by inverse PCR
iPCR-Cas1-2/3-fwd iPCR-Cas1-2/3-rev	GCGGCCGCATAATGCTTA GATTTATTCCTCATCTTC	Primer for removal of Cascade genes from pRSFDuet + <i>cas7fv</i> + <i>cas5fv</i> -His (in MCS1) + <i>cas6f</i> (in MCS2)
Cas7-BamHI-fwd Cas7-HindIII-rev	CCAGGATCCATGCAAAAAGTAACGG CCGCAAGCTTCTATTTTGCATAAAAATACTG	Primer for cloning <i>cas7fv</i> in first MCS of pRSFDuet, cut by BamHI and HindIII
Cas7-NdeI-fwd Cas7-XhoI-rev	TACATATGCAAAAAGTAACGGG GACTCGAGCTATTTTGCATAAAAATAC	For cloning of <i>cas7fv</i> in second MCS of pRSFDuet (NdeI/XhoI)
Cas5-BamHI-fwd Cas6-NotI-rev	CAGGATCCGATGAAAATAATCATAG ATGCGGCCGCTTAAAACCAAGGACTGTAG	For cloning of <i>cas5fv</i> and <i>cas6f</i> in pACYCDuet restricted with BamHI and NotI (extra G for N-terminal His-tag)
random-pBAD-fwd random-pBAD-rev	ACGTTACCGCCGCACAGGCGGCTTAGAAAGCAA AAAGCAAAGCACCG CGAATTCATATGGTACCAGCTGCAGATCCAAC TCTTTGAACCAAGG	For cloning of 500 nt random sequence (without RBS and any ATG) from pRSFDuet backbone in pBAD
invPCR-5-Strep-fwd invPCR-5-Strep-rev	CAGTTTAAAAAAGCCAGGATCCGATGAAA CGGATGGCTCCAGCTGCTGCCATGGTATA	Exchange of His-tag with Strep-tag in pACYCDuet + His- <i>cas5fv</i> + <i>cas6f</i>
NC-Control-fwd NC-Control-rev	CGGGAGTATGCAAAAAGCAAAGCACCG TATCTCTTCTTTCTAAGCCGCTGTGC	For cloning of RBS + 7nt spacing distance + ATG upstream of pBAD + repeat-non-coding

minus-His-Cas5-fwd	TAGTAAGCGGCCGCATAATGCT	For removal of C-terminal His-tag on <i>cas5fv</i>
minus-His-Cas5-rev	AAGCTTAATGTTTGATACATAG	
N-His-Cas5-fwd	CATCACCATCATCACCACAAAATAATCATAGAATATG	For addition of N-terminal His-tag on <i>cas5fv</i>
N-His-Cas5-rev	CATGCAACCTCCTATTTTG	
pdt1-fwd	GCGGCGAACAAAAACGAAGAAAAACCAACGAA GTGCCGACCTTTATGCTGAACGCGGGCCAGGCGA ACCGCCGCCGCTG	Synthesis von pdt#3 from (Cameron & Collins, 2014) for placement on C-terminus of <i>cas7fv</i>
pdt-1-revcomp	CACGCGGCGGCGGTTTCGCTGCCCCGTTTCAGC ATAAAGGTCGGCACTTCGTTGGTGTTCCTTCGTT TTTGTTCGCCGC	
pdt2-fwd	GCAGCCAATAAGAATGAGGAGAATACGAATGAG GTTCTACGTTATGCTCAATGCCGACAAGCTA ATCGTCGACGGGTC	Synthesis of pdt#3 from (Cameron & Collins, 2014) for placement on C-terminus of <i>cas5fv</i> (varied codon sequence)
pdt-2-revcomp	GACCCGTCGACGATTAGCTTGTCCGGCATTGAGC ATGAACGTAGGAACCTCATTCTCTCTCATT CTTATTGGCTGC	
Cas7-wl-neutral-fwd	GCCGCCGCCGCGGCACAAGCAACTGACATTA	For exchange of aa 62-66 to alanine to neutralize wrist loops by inverse PCR
Cas7-wl-neutral-rev	GCCCGTTTCATCTTTAC	
I-F-Repeat-fwd	AGAAAGAAGGAGATACC	Inverse PCR to change I-Fv Repeat sequence to I-F Repeat in pBAD + repeat-non-coding
I-F-Repeat-rev	TAGCTGCCTATACGGCA	
pdt1+linker-fwd	GCGGCGAACAAAAACGAAG	addition of GSGS linker between pdt and <i>cas7fv</i>
pdt1+linker-rev	CGACCCCCCTTTTGATAAAAAACTG	
pdt2+linker-fwd	GCAGCCAATAAGAATGAG	addition of GSGS linker between pdt and <i>cas5fv</i>
pdt2+linker-rev	CGACCCCCCAAGCTTAATGTTTGATAC	

Table 4.8: Primers used for in vitro assays

Name	Sequence 5'-3'	Description
T7sfGFP-Repeat-fwd	GAAATTAATACGACTCACTATAGGGAGAGTTC ACCGCCGCACAGGCGG	Fwd Primer for amplification of <i>in vitro</i> transcription template, including a T7 promoter sequence
T7-5'handle-gfp-fwd	GAAATTAATACGACTCACTATAGTTAGAAAGA AGGAGATAC	Fwd Primer for amplification of <i>in vitro</i> transcription template, including a T7 promoter sequence and directly starting with processed 5'-handle (first nt C instead of G)
T7sfGFP-Repeat-rev	TTAATGGTGATGATGATGGTG	Rev primer for amplification of <i>in vitro</i> transcription template
IacZ-probe	TTGTAACGACGCGCCAGTGAATTCGAGCTCG GTA	Northern Blot probe for detection of IacZ- α (in initial 100 nt)
Sp4-GG-tar	AAGCTTGAGGGCCCAAGCCGTTATGCTAGGGT TATAGGTTTGC GCGTCTTGCTGGGCGATAGGA CTCCCTATAGTGAGTCGTATTAGGATCC	EMSA substrate including complementary sequence to crRNA spacer and GG-PAM
Sp1-GG-tar	AAGCTTGAGGGCCCAAGCCGTTATGCTAGCAA TGTGGTCGCGCAATTTATGATTTGGTTGAGGA CTCCCTATAGTGAGTCGTATTAGGATCC	EMSA substrate including non-complementary sequence to crRNA spacer and GG-PAM
Sp4-GG-non-target	GGATCCTAATACGACTCACTATAGGGAGTCCT ATCGCCAGCAAGACGCGCAACCTATAACCC TAGCATAACGGCTTGGGCCCTCAAGCTT	Complementary sequence to Sp4-GG-tar for dsDNA constructs

Sp1-GG-non-target	GGATCCTAATACGACTCACTATAGGGAGTCCT CAACCAAATCATAAATTGCGCGACCACATTGCT AGCATAACGGCTTGGGCCCTCAAGCTT	Complementary sequence to Sp1-GG-tar for dsDNA constructs
Sp4-AA-tar	AAGCTTGAGGGCCCAAGCCGTTATGCTAGGGT TATAGGTTTGC GCGTCTTGCTGGGCGATAAAA CTCCCTATAGTGAGTCGTATTAGGATCC	EMSA substrate including complementary sequence to crRNA spacer and a wrong PAM sequence
Sp4-AA-ntar	GGATCCTAATACGACTCACTATAGGGAGTTTTA TCGCCAGCAAGACGCGCAAACCTATAACCT AGCATAACGGCTTGGGCCCTCAAGCTT	Complementary sequence to Sp4-AA-tar for dsDNA constructs
Rloop-mimic-tar	GGTTATAGGTTTGC GCGTCTTGCTGGGCGATA GGACTCCCTATAGTGAG	EMSA substrate including complementary sequence to crRNA spacer and GG-PAM
Rloop-mimic-ntar	CTCACTATAGGGAGTCCATTATTATT	Partially complementary sequence to Rloop- mimic-tar
Rloop-Sp4-TT-tar	GGTTATAGGTTTGC GCGTCTTGCTGGGCGATA AAACTCCCTATAGTGAG	EMSA substrate including complementary sequence to crRNA spacer and wrong PAM
Rloop-Sp4-TT-ntar	CTCACTATAGGGAGTTTATTATTATT	Partially complementary sequence to Rloop- Sp4-TT-tar
Rloop-bubble-ntar	GGATCCTAATACGACTCACTATAGGGAGTCCA TAGCGGGTCCAAGACGCGCAAACCTATAACCC TAGCATAACGGCTTGGGCCCTCAAGCTT	Complementary sequence to Sp4-GG-tar with small 10 nt opening adjacent to PAM

4.4 Working with DNA

4.4.1 Preparation of plasmid DNA from E. coli

Plasmids were isolated with the Qiagen plasmid mini kit or Qiagen plasmid maxi kit according to the manufacturer's instructions.

4.4.2 Sanger sequencing

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

4.4.3 Quantification of DNA

4.4.3.1 Spectrophotometric quantification

The concentration of plasmid DNA and precipitated DNA oligonucleotides was determined by spectrophotometry by measuring the absorbance at 260 and 280 nm. The purity was determined by an A₂₆₀/A₂₈₀ ratio of 1.8 - 2.0.

4.4.3.2 Fluorometric quantification

The Qubit fluorometer was used for high sensitivity quantification of low-yield DNA preparations, e.g. after cDNA library preparation (see section 4.5.6). The Qubit utilizes fluorescent dyes (for ss/dsDNA, RNA or protein), which emit a signal only when bound to the specific target. By calibration with

DNA/RNA/protein standards, the concentration of the utilized sample can be determined. DNA samples were quantified via the Qubit dsDNA HS Assay Kit according to the manufacturer's instructions.

4.4.4 Electrophoresis of DNA

4.4.4.1 Agarose gel electrophoresis of DNA

To confirm the linearization of plasmid DNA and correct amplification by PCR, DNA molecules were separated by agarose gel electrophoresis. Agarose gels with 1 % to 2 % (w/v) agarose in TAE buffer (40 mM Tris-acetate, 1 mM EDTA pH 8) and 0.5 µg/mL ethidium bromide were prepared depending on the size of the analyzed DNA fragments. Before the DNA samples were applied into the sample wells of the gels, they were mixed with loading dye (6x stock: 0.2 % bromphenol blue, 0.2 % xylene cyanol FF, 60 % (v/v) glycerol, 60 mM EDTA pH 8). 5 µl of 2-Log DNA ladder (New England Biolabs) was loaded on each gel to determine the size of the DNA. Electrophoresis was performed at 80-120 V at RT in TAE buffer. The DNA was visualized by UV irradiation at 254 nm.

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

Electrophoretic separation of smaller DNA fragments (< 300nt) was performed under non-denaturing conditions using polyacrylamide gels. Depending on the size of the DNA fragment, the concentration of polyacrylamide (acrylamide /bisacrylamide, 40%, ratio 29:1) in the gel (90 mM Tris pH 8.0, 90 mM boric acid, 2 mM EDTA, 0.03% (v/v) APS, 0.005% (v/v) TEMED) was varied between 4% and 12% (v/v). The DNA samples mixed with 6x DNA loading dye were applied on to the gel. DNA marker containing a mixture of DNA fragments of known size was also applied onto the gel to size the fragments. The gel run was performed in 1x TBE at 10 W for 1 h. After electrophoresis, gels were stained for 5 min in SybrGold dissolved in TBE buffer and visualized by UV light at 254 nm.

4.4.5 Purification of DNA

4.4.5.1 PCR Purification

PCR reactions showing a single distinct band after gel electrophoresis were purified with the QIAquick PCR Purification Kit according to the manufacturer's instructions.

4.4.5.2 Gel extraction from agarose gels

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

4.4.5.3 Gel extraction from polyacrylamide gels

DNA was extracted from polyacrylamide gels by cutting the respective bands. The gel pieces were then transferred to a Gel Breaker tube (centrifuge tube with small holes) and centrifuged (14,600 rpm, 2 min, RT) into a 2 ml collection tube. 500 µl gel elution buffer (20 mM Tris-HCl pH 7.5, 250 mM sodium acetate, 1 mM EDTA, 0.25 % SDS) were added on the gel debris and the mixture was incubated overnight on ice while shaking (300 rpm). Following this, the DNA containing gel elution buffer was transferred to a Costar® centrifuge filter tube and centrifuged (14,600 rpm, 2 min, RT) to remove remaining gel debris. The DNA was subsequently purified with EtOH precipitation (see section 1.5.2).

4.4.6 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was used for amplification of DNA fragments. Two primers were designed, flanking the sequence of interest to be amplified. The elongation of these primers was carried out by Phusion polymerase, with an included proofreading ability and a reduced mutation rate. A standard PCR reaction included the following main steps: I) Denaturation: Heating the reaction at 95°C results in the melting of dsDNA into ssDNA (template). II) Primer annealing: Annealing or binding of the primers to their complementary DNA. III) Elongation: Extension or elongation of the primer in the 5' to 3' direction. DNA polymerase catalyzes the elongation by addition of complementary nucleotides. The above-listed steps were repeated to achieve sufficient amplification (25-30x).

Inverse PCR was performed for deletions or small insertions in existing plasmids. In this variation, the entire plasmid is with the forward primer directed downstream and the reverse primer directed upstream. Primers either included the sequence overhangs with the sequence to be inserted or were flanking the sequence to be deleted. The elongation time during PCR was adjusted accordingly.

PCR amplifications from genomic or plasmid DNA were performed using the following reaction mixture: ~ 50 ng template DNA, 250 µM dNTPs, 0.2 µM of each primer, 1 x concentrated GC buffer 3% (v/v) DMSO, 1 U Phusion polymerase and adjusted to 50 µl with water. The reaction was performed in a thermal cycler (BioRad) using the following program:

Step 1) 95°C – 60 sec

Step 2) 95°C – 30 sec

Step 3) 55-65°C – 30 sec x 30 – 35

Step 4) 72°C – 30 sec/kb

Step 5) 72°C – 5 min

The primers used for PCR are listed in Table 4.7.

4.4.7 Modification of DNA

4.4.7.1 Restriction

The restriction digestion of plasmid DNA and PCR products was achieved with appropriate restriction endonucleases in respective buffers according to the manufacturer's instructions. The reaction mixture containing 5-10 U enzyme/ μ g DNA was incubated at 37 °C for 2 h to digest the DNA.

4.4.7.2 Ligation

T4 DNA ligase was used for ligation of restricted plasmid DNA in the appropriate buffer according to the manufacturer's instructions. In a standard ligation reaction, 0.02 pmol vector DNA was mixed with 0.06 pmol insert DNA (ratio 1:3) and 4 U T4 DNA ligase in the recommended DNA ligase reaction buffer containing ATP. Phosphorylated inverse PCR products were self-ligated by addition of 10 U T4 DNA ligase to the phosphorylation reaction. The reactions were incubated overnight at 16°C and subsequently used for transformation with *E. coli* DH5 α .

4.4.7.3 Phosphorylation

Inverse PCR products were phosphorylated by T4 Polynucleotide Kinase (PNK) to allow self-ligation. For this, 200 ng of inverse PCR product was mixed with 10 U T4 PNK in 1x DNA ligase reaction buffer containing ATP. The reaction was incubated for 1 h at 37 °C.

4.4.7.4 Dephosphorylation

The 5'-ends of restricted plasmid DNA were dephosphorylated with Antarctic Phosphatase to avoid self-ligation during the ligation reaction. A standard dephosphorylation reaction included 1 μ g restricted plasmid DNA and 5-10 U Antarctic Phosphatase in the recommended reaction buffer of the manufacturer. The reaction was incubated at 37 °C for 1-2 h followed by heat inactivation at 65°C for 15 min.

4.4.7.5 Gibson Assembly

The Gibson Assembly technique was used to clone most of the plasmids for *cas* gene expression. This technique requires primers including overlapping regions (15-20 nt) to assemble with the backbone

(Gibson *et al.*, 2009). During isothermal conditions, a T5 exonuclease degrades dsDNA in 5' to 3' direction, resulting in long 3'-overhangs which bind to the complementary overhangs of the neighboring DNA fragment. DNA polymerase subsequently fills up the single-stranded DNA by incorporating the complementary nucleotides. The resulting gaps are afterwards filled up by a DNA ligase. A self-made Gibson Assembly reaction mix was used, which did not contain the DNA ligase making use of *E. coli* own ligase instead. This Hot Fusion reaction mix was proved to contain a higher assembly efficiency than the original Gibson Assembly mix (Fu *et al.*, 2014). A typical reaction contained 100 fmol of PCR product, 1.5 U T5 exonuclease and 20 U Phusion DNA polymerase in pre-assembly buffer (100 mM Tris pH 8.0, 10 mM MgCl₂, 200 μM dNTPs, 10 mM DTT and 5% (v/v) PEG-8000) and was incubated for 1 h at 50°C. The reaction was afterwards transformed into *E. coli*.

4.4.8 Transformation

Competent cells of *E. coli* DH5α or *E. coli* expression strains were chemically prepared with rubidium chloride (RbCl) and calcium chloride (CaCl₂) (Inoue *et al.*, 1990). 100 ml of LB medium was supplemented with 10 mM MgC₂ and 10 mM MgSO₄ and inoculated with 2 ml of an overnight culture of *E. coli* and grown until an OD_{600nm} of 0.6. The culture was cooled on ice for 30 min and cells were harvested by centrifugation (3.000xg, 10 min, 4 °C). Subsequently, the pellet was resuspended in 33 ml cold RF1 solution (30 mM potassium acetate pH 5.8, 100 mM RbCl, 50 mM MnCl₂, 10 mM CaCl₂ and 15 % glycerol) and incubated on ice for 30 min. Cells were again centrifuged (3.000xg, 10 min, 4 °C) and the pellet was gently resuspended in 5 ml cold RF2 solution (10 mM RbCl, 10 mM MOPS pH 5.8, 75 mM CaCl₂, 15 % glycerol). Cells were incubated again for 30 min on ice and 100 μl aliquots were created and stored at -80 °C.

For transformation, plasmid DNA was gently mixed with one aliquot of competent cells and incubated on ice for a minimum of 15 min. Cells were then head-shocked by incubations for 45 s at 42 °C and placed on ice again for 1 min. Following this, 900 μl of LB medium was added and the mixture was incubated for 45-60 min at 37 °C and 300 rpm. 100 μl of transformed cells were plated on LB plates with the appropriate antibiotics. The remaining cells were pelleted by (8000 rpm, 30 s, RT) and resuspended in 100 μl to be plated as well. Plates were incubated overnight at 37 °C until visible colonies were formed that could be screened for positive clones containing the plasmid.

4.4.9 Radioactive labeling

DNA oligonucleotides were radioactively labeled to create probes for Northern Blot or substrates for EMSA or Nuclease assays. A total of 100 pmol of each oligonucleotide was 5'-labeled with [γ-³²P]-ATP

(5000Ci/mmol, Hartmann Analytic) and T4 PNK (NEB) for 1 h at 37°C. The reaction was stopped by addition of formamide loading buffer and substrates were separated by denaturing-PAGE (10% polyacrylamide, 8 M Urea, 1x TBE). After autoradiographic exposure, bands were cut from the gel, eluted and EtOH precipitated (see section 4.5.2). Low Molecular Weight Marker (Affymetrix) radioactively labeled with illustra MicroSpin G-25 columns (GE healthcare) for size determination in nuclease assays.

4.5 Working with RNA

4.5.1 Treatment of solutions, glassware and equipment

All applied buffers and solutions were treated with 0.1% (v/v) DEPC and autoclaved to remove traces of DEPC after overnight incubation at RT. Non-disposable plastic ware was treated with RNase Exitus Plus (AppliChem) whereas disposable plastic such as pipette tips was purchased RNase-free conditions and autoclaved before use. Glassware was sterilized by incubation at 210°C for at least 2h before use.

4.5.2 RNA extraction

RNA was extracted from purified proteins via phenol/chloroform extraction followed by EtOH precipitation. Samples were thoroughly mixed with 1 volume of acidic phenol/chloroform mix (Ambion) and centrifuged (12,000 x g, 10 min and 4 °C). The upper phase was transferred to a new tube and the process was repeated. This upper phase was then mixed with 1 volume of chloroform and centrifuged again. The resulting upper phase was then transferred subjected to EtOH precipitation.

2 volumes of ethanol and 0.3 M Na-acetate were added to the solution and the mixture was incubated at -20°C for 1 h. The sample was then centrifuged (12,000 x g, 10 min and 4 °C), the supernatant removed and the pellet washed by addition of 1 volume of EtOH and repeated centrifugation. After removal of the supernatant, the pellet was air-dried and then resuspended in DEPC-ddH₂O.

4.5.3 Quantification of RNA

4.5.3.1 Spectrophotometric quantification of RNA

Quantification and quality control of extracted RNA was performed by Nanodrop spectrophotometer as described for DNA (see section 4.4.3).

4.5.3.2 Fluorometric quantification

Low-yield RNA preparations were quantified with the Qubit fluorometer (see section 4.4.3.2). Quantification was performed using the Qubit RNA HS Assay Kit according to the manufacturer's instructions.

4.5.4 Electrophoresis of RNA

4.5.4.1 Agarose gel electrophoresis of RNA

Larger RNA molecules (>100 nt) were separated on agarose gels (1 %, 1x TBE) as described in section 4.4.4.1.

4.5.4.2 Denaturing polyacrylamide gel electrophoresis (Urea-PAGE)

RNA preparations or protein samples co-eluting with RNA were separated and visualized by urea-polyacrylamide gel electrophoresis. 8 M urea was added in these gels to resolve the secondary structure of RNA and guarantee separation based on size. Urea-gels otherwise consisted of 8 M urea, 90 mM Tris pH 8.0, 90 mM boric acid, 2 mM EDTA, 1.0 % (v/v) ammonium persulfate (APS) and 0.1 % (v/v) tetramethylethylenediamine (TEMED), and 4 % (v/v) polyacrylamide (acrylamide / bisacrylamide, 40 %, ratio 29:1). Gels with a total volume of 10 ml were prepared for normal RNA analysis while bigger gels with approximately 40 ml total volume were made for the extraction of *in vitro* transcribed RNA and Nuclease assays. Fully polymerized gels were run in 1x TBE buffer (90 mM Tris pH 8.0, 90 mM boric acid and 2 mM EDTA) at 200 V till sufficient separation. The RNA samples were mixed with 2x formamide loading dye (80% formamide, 10 mM EDTA, 0.05% (w/v) bromophenol blue and 0.05 % (w/v) xylene cyanol) and incubated for 10 min at 95 °C before loading them on the urea-gel. The RNA bands were visualized by toluidine blue staining and destaining in ddH₂O overnight or by SybrGold staining for 5 min and UV transillumination at 254 nm.

4.5.5 Northern Blotting

A semi-dry electrophoretic transfer system was used to transfer the RNA that was separated on denaturing polyacrylamide gels onto a positively charged nylon membrane (Roti®-Nylon plus, pore size 0.45 µm). Prior to the transfer, the membrane, the gels as well as Whatman GB004, 3MM Paper were equilibrated in 1x TBE buffer for 5 min. The blot was assembled in the order 6x Whatman paper, nylon membrane, polyacrylamide gel, 6x Whatman paper and the transfer was performed for 2 h at 20 V. Subsequently, the RNA was UV-crosslinked to the membrane.

The membrane was pre-hybridized for 30 min at 42°C in ULTRAhyb-Oligo Hybridization Buffer (1 ml/10 cm² membrane) to block non-specific binding sites. The 5'-terminal radiolabeled probes (10⁶ cpm/ml hybridization buffer) were applied to the hybridization buffer after incubation at 95°C for 5 min. The hybridization was performed overnight at 42°C. The blot was washed twice, with 15 ml low stringency buffer (2x SSC, 0.1 % SDS) and with 15 ml high stringency buffer (1x SSC, 0.1 % SDS) for 30 min at 42°C each to remove unbound probe. The membranes were exposed to phosphor screens overnight and the bands on the phosphor screen were visualized with a phosphorimager.

4.5.6 Illumina Sequencing

RNA was extracted from purified proteins via phenol/chloroform extraction (acidic) followed by EtOH precipitation. Pellets were resuspended in DEPC-ddH₂O. For Illumina sequencing, RNA had to be fragmented to create cDNA libraries. Fragmentation was achieved by addition of a ZnCl₂ fragmentation buffer (final concentration: 10 mM Tris-HCl pH 6.8, 10 mM ZnCl₂) and heating for 10 min at 95°C. Fragmentation was immediately stopped by putting the sample on ice and addition of 2 µl 500 mM Na₂-EDTA pH 8. The fragmented RNA was separated by denaturing Urea-PAGE and extracted from the gel (see section 4.5.4.2 and 4.4.5.3). Fragmentation by bivalent metal ions presumably results in 2',3'-cyclic phosphate and 5'-OH RNA termini (Forconi & Herschlag, 2009), so fragments were first dephosphorylated (for 2',3'-cyclic phosphate curation) and subsequently phosphorylated (for 5'-OH curation) with T4 PNK. 15 µl of fragmented RNA was mixed with 6 µl of 5x dephosphorylation buffer (500 mM Tris-HCl pH 6.5, 500 MgAc, 25 mM β-mercaptoethanol) and 1 µl T4 PNK (Ambion) in a total volume of 30 µl and incubation for 6 h at 37 °C for dephosphorylation. For phosphorylation, 1 mM ATP and 1 µl T4 PNK were added and the mixture was further incubated for 1 h at 37 °C. The treated RNA was afterwards purified by EtOH precipitation including 1 % (v/v) glycogen.

cDNA libraries were created with the NEBNext® Multiplex Small RNA Library Pret Set for Illumina according to the manufacturer's instructions. 100 ng of input RNA was used and amplified cDNA libraries were separated by native PAGE for size selection of 120-250 bp. The selected sizes were extracted from the gel and EtOH precipitated

Initial Illumina sequencing of *sfgfp* and *lacZ* constructs was performed by the Max Planck Genome Centre (Köln) by Illumina HiSeq2500. The performed sequencing had a length limitation of 150 nt. Paired-end sequencing was performed to eliminate the change of not sequencing the end of cDNAs, in case they have a too large size due to insufficient fragmentation. With Paired-end sequencing, cDNA pieces are sequenced from both ends.

Newer sequencing (non-coding construct, *sfgfp* co-produced with *cas* genes on multiple plasmids and *sfgfp* produced with T7 RNAP) was performed with the MiniSeq Sequencing System in our own laboratory.

4.5.7 Nanopore sequencing

Nanopore sequencing was performed to sequence extracted RNA without the need for fragmentation. Extracted RNA was treated with *E. coli* Poly(A) Polymerase (NEB) for 30 min 37°C to add a poly(A) tail and then processed for Nanopore sequencing with the Direct RNA Sequencing Kit according to the protocol provided by Oxford Nanopore Technologies. Sequencing was performed with The MinIon Sequencing Device.

4.5.8 Mapping of sequencing reads

Mapping of the sequencing data was performed using CLC Genomics Workbench 9.5.3 (Qiagen, Germany). The sequencing data was processed by (i) removal of sequences of low quality (quality score limit, 0.05; maximum number of ambiguities, 2), (ii) trimming of adapter sequences, and (iii) filtering by length (15-nt cutoff). The trimmed sequences were mapped to the reference genome of BI21 (DE3) and the sequence of the target plasmid using default settings.

4.5.9 *In vitro* transcription

In vitro transcription was performed to obtain *sfgfp* for *in vitro* RNA wrapping experiments. To generate a template, primers were produced that contained the sequence of the T7 promoter upstream of a region flanking the *sfgfp* gene. PCR amplification was confirmed by agarose gel electrophoresis and the product was extracted. *In vitro* transcription was performed in a total volume of 1 ml (40 mM HEPES/KOH pH 8, 22 mM MgCl₂, 5 mM DTT, 1 mM spermidine, 4 mM of each NTP, 10 µg DNA template, and 30 nM T7 RNA polymerase). All ingredients for the reaction were produced fresh with DEPC-ddH₂O to guarantee stability of the produced RNA. The reaction mix was incubated for 3-5 h at 37 °C to produce RNA. An additional Formed pyrophosphate was removed by centrifugation. *In vitro* produced RNA was identified by loading 10 µl on a Urea-PAGE. After identification of the included RNA, the rest of the reaction mix was loaded on a bigger Urea-PAGE and the RNA completely separated. The produced RNA was extracted by cutting the band from the gel and incubating it with 500 µl gel elution buffer (20 mM Tris/HCl pH 7.5, 250 mM sodium acetate, 1 mM EDTA and 0.25 % SDS) overnight at 4 °C. On the following day, the gel pieces were removed by centrifugation (1 min, 13,000 rpm, RT) and the

supernatant was transferred to a new reaction tube. RNA was then obtained by EtOH precipitation (see section 4.5.2).

4.6 Biochemical Methods

4.6.1 Cell lysis

After the heterologous expression of recombinant proteins (see section 2.3.3), the harvested cells were resuspended in the according wash buffer with a ratio of 5 ml buffer per 1 g cell mass. Furthermore, 0.25 ml of protease inhibitor was added per 1 g cell mass for purification of Cas3fv. Lysis was achieved by the addition of lysozyme and incubation on ice for 15 min followed by cell disruption in an LM10 Microfluidizer® at 18,000 psi. The lysate was centrifuged at 4 °C and 38,500 × g for 45 min to remove cell debris. The supernatant was subsequently filtered using a syringe filter (pore size: 0.22 µm) for purification.

4.6.2 Affinity purification

All proteins and protein complexes were initially purified with affinity purification. For most experiments, a 6xHis-tag was used (Ni-NTA) while a Strep-tag was used for purification of the Cas1-Cas2/3fv complex and for the RNA wrapping complex including a second repeat sequence at the 3'-end of the target RNA. Cell lysate was continuously applied to a 5 ml HisTrap or StrepTrap column for 1 h. Bound protein was eluted after connecting the column to the FPLC and washing with approximately 20 ml washing buffer (50 mM Tris/HCl, 20 mM imidazole, 300 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 10 % glycerin, pH 7). Proteins bound to the column were eluted by a stepwise gradient of imidazole by raising the concentration of elution buffer (50 mM Tris/HCl, 500 mM imidazole, 300 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 10 % glycerine, pH 7). In Strep-tag affinity purifications, a wash buffer without imidazole was used and the elution buffer contained 2.5 mM Desthiobiotin.

Eluted samples during the gradient were collected in fractions of 1 ml. The absorbance at 260 and 280 nm was continuously measured by a built-in UV detector. Fractions with a high absorbance rate, indicating proteins and RNA were analyzed by SDS-PAGE and Urea-PAGE.

4.6.3 Size-exclusion chromatography

After general purification by affinity chromatography, proteins were further purified by size-exclusion chromatography. Samples from the His-Tag affinity chromatography containing protein were pooled and concentrated to a final volume of 2 ml by using Amicon centrifugal concentrators and centrifugation at 6,000 g at 4 °C. Concentrated protein samples were further separated by size with a gel filtration column

(HiLoad 16/600 Superdex 200 pg or Superose 6 Increase 10/300 GL) connected to an FPLC. Separation was performed with a total volume of 120 ml HEPES buffer (50 mM HEPES, 150 mM NaCl, 1 mM DTT, pH 7). The absorbance at 280 nm was continuously measured by a built-in UV detector. Proteins in the eluted samples were identified by SDS-PAGE, while RNA was detected by Urea-PAGE. The total amount of purified protein was determined using Bradford assay.

4.6.4 Anion-exchange chromatography

Anion-exchange chromatography was performed for small RNA wrapping complexes formed on repeat-tagged *sfgfp* RNA produced T7 RNAP following the protocol of Jahn *et al.* (Jahn *et al.*, 1991). The concentrated sample was loaded on a MonoQ column and the flow-through was collected. The bound sample was gradually eluted in a linear gradient over 20 column volumes and an increasing concentration of NaCl (50 mM Tris/HCl, 300-1000 mM NaCl, 1 mM DTT, 10 % glycerin, pH 7). To remove the remaining bound protein, the column was further washed with 5 CV of wash buffer (50 mM Tris/HCl, 300 mM NaCl, 1 mM DTT, 10 % glycerin, pH 7).

4.6.5 Protein quantification by Bradford

The amount of purified protein was measured by Bradford assay (Bradford, 1976). 200 μ l Bradford reagent was added to a dilution of the protein sample and the mix was incubated for 15 min. Afterwards, the OD_{595nm} was measured and the amount of protein was calculated by a fresh calibration curve with BSA.

4.6.6 Production and purification of recombinant proteins

4.6.6.1 Purification of recombinant I-Fv Cascade and truncated variants

For production and purification of recombinant I-Fv Cascade, *cas* genes and crRNA were co-purified. The *cas* genes *cas7fv*, *cas5fv*, *cas6f* were provided in the vector pRSFDuet-1 which allows for the simultaneous production of all three proteins with Cas5fv fused to an N-terminal His-tag. This plasmid was co-transformed into *E. coli* BL21 (DE3) pLys with a second pUC19 vector containing the repeat-spacer4-repeat sequence of the single *S. putrefaciens* CN-32 CRISPR array downstream of a T7 RNA polymerase promoter. Cultures were grown and harvested as described (see section 4.2). Cell pellets were lysed and proteins first purified via Ni-NTA purification in a buffer containing 50 mM Tris/HCl, 20-500 mM imidazole, 300 mM NaCl, 10 mM MgCl₂, 1 mM DTT, 10 % glycerine, pH 7. Samples containing protein as detected by UV were then pooled, concentrated to 2 ml and subjected to size-exclusion

chromatography (HiLoad 16/600 Superdex 200 pg) in a buffer containing 50 mM HEPES-NaOH pH 7.0, 150 mM NaCl.

Truncated Cascade, missing the AH domain of Cas5fv or missing the wrist loops from Cas7fv as well as Cascade with the sfGFP-Cas7fv fusion purified were by co-production of the respective genes from pRSFDuet with a crRNA from pUC19 in the same fashion as for wt-Cascade.

4.6.6.2 Purification of recombinant Cas3fv and Cas1-Cas2/3fv complex

The target nuclease Cas2/3fv was purified as a fusion construct with an N-terminal His-tag and a C-terminal MBP-Tag using the vector pETM-43. Standalone Cas3 with the Cas2 portion removed was provided on a pet24(+) vector by Dr. Patrick Pausch. Cas3 was produced and purified via Ni-NTA chromatography and size exclusion chromatography (Superose 6 Increase 10/300 GL) in a buffer containing 20 mM HEPES/KOH pH 7.5, 750 mM NaCl, 10 mM MgCl₂ and 2% Glycerin at 4 °C. The Cas1-Cas2/3 complex was purified by co-expression of *cas1* and *cas2/3fv* from pRSFDuet-1 with an N-terminal Strep-tag fused on Cas1. The final size-exclusion chromatography was performed with HiLoad 16/600 Superdex 200 pg.

4.6.6.3 Purification of filaments and RNA wrapping complexes

For directed RNA wrapping a repeat-tagged sequence, either *sfgfp*, *lacZ-α* or a non-coding sequence (from the backbone of the pRSFDuet vector) were transcribed from a pBAD vector for arabinose induction or pETDuet-1 for induction with IPTG. *Cas* genes were produced from pRSFDuet or pCsy_complex (addgene). In a later experiment, *cas* genes were split on different vectors. In this setup, *cas7fv* was expressed from pRSFDuet (either with a copy in the first MCS or with a copy in both MCS of the plasmid) and *cas5fv* and *cas6f* were expressed from pACYCDuet. Repeat-tagged target transcripts were expressed from pBAD as before.

Formed complexes were purified via Ni-NTA chromatography in a standard purification buffer (50 mM Tris/HCl pH 7.0, 300 mM NaCl, 20-500 mM imidazole, 10 mM MgCl₂, 10% glycerol, 1 mM DTT) as for Cascade purification. For further clean-up and TEM analysis, size-exclusion chromatography was performed in the same fashion as for I-Fv Cascade and samples from the fractions of the void volume were taken for analyses. For double affinity purification, a C-terminal Strep-tag was fused to Cas6f in addition to the N-terminal His-tag of Cas5fv on the pRSFDuet plasmid and Cas proteins were produced and purified via Strep-tag affinity chromatography.

To remove rRNA, Purification was later repeated by Ni-NTA under the same conditions as before but without MgCl₂ in the wash buffer (50 mM Tris/HCl, 300 mM NaCl, 1 mM DTT, 10 % glycerin, pH 7).

Samples were concentrated and loaded on an analytic size-exclusion (Superose 6 Increase 10/300 GL) in the standard SEC buffer of I-Fv Cascade.

For purification of small RNA wrapping complexes, repeat-tagged *sfgfp* was produced from pETDuet with T7 RNAP and co-purified by Ni-NTA with Cas proteins via His-tagged Cas5fv. Samples were concentrated to a final volume of 500 μ l and subjected to Anion-exchange chromatography. Fractions were collected in 700 μ l and analyzed by SDS- and Urea-PAGE. The flow-through was pooled and subjected to size-exclusion chromatography with a small semi-preparative and analytic column (Superose 6 Increase 10/300 GL). Samples were eluted in standard SEC buffer.

4.6.6.4 Purification of SUMO-Cas7fv

His-SUMO-Cas7fv was purified with the standard purification buffer for Cascade but an extra wash step with a high salt buffer (50 mM KCl, 1 M NaCl, 10 mM MgSO₄, 2 mM ATP) was applied for 20 column volumes while protein was bound to the column. Protein was then eluted with standard Cascade elution buffer containing imidazole and eluted protein was concentrated before being loaded on size-exclusion chromatography at 4 °C to separate monomeric SUMO-Cas7fv from aggregated protein.

4.6.7 *In vitro* RNA wrapping

500 μ l RNA-free SUMO-Cas7fv or Cas5fv-Cas7fv dimer (~ 1 mg) were mixed with either *in vitro* transcribed *sfgfp* RNA (~ 5 μ g) or small RNA extracted from *pseudomonas oleovorans* in our laboratory (3.5 μ g, heated up at 95 °C and cooled down on ice before added to protein) and incubated for 1 h at RT followed by incubation overnight at 4 °C (with the addition of SUMO protease (1:500) in case of Cas7fv-SUMO). The following day, precipitated protein was removed by centrifugation and the supernatant was loaded on size-exclusion chromatography (Superose 6 Increase 10/300 GL) to separate potential complexes from monomeric protein.

Crystallization of Cas7fv was performed in hanging-drop format or in microtubes following a protocol established by Dr. Patrick Pausch. Freshly purified I-Fv Cascade (concentrated to 54 mg/l) was mixed in various ratios (2:1, 1:1, 1:2) with screening solution (0.2 M MgCl, 0.1 M Tris pH 7.2, 2 M NaCl and 3 % fructose) and incubated at 18 °C until crystals were formed.

4.6.8 RNA protection assays

Ni-NTA purified RNA wrapping complexes formed with on the lacZ-Repeat construct were incubated at 23 °C, 200 rpm and 10 U of RNase I for varying time points of up to 7 days. Following this, RNA was extracted and separated on Urea-PAGE. 1 μ g of directly extracted RNA was further incubated under the

same conditions to ensure the activity of RNase I and then separated on Urea-PAGE. Gels were either stained with Sybr-Gold for nucleic acid detection or used for Northern Blot analysis. The 50 bp DNA ladder (NEB) was used for orientation.

4.6.9 Electrophoretic mobility shift assays (EMSA)

The recombinant Cascade complex was tested for its ability to bind target DNA constructs in electrophoretic mobility shift assays (EMSAs). Utilized target oligonucleotides contained a sequence complementary to the spacer sequence in the crRNA (spacer4) with a correct PAM (GG) or a wrong PAM (TT) at the 3'-end adjacent to the spacer. A non-complementary spacer sequence (spacer1) was used as a control. For dsDNA constructs, the non-target strand was labeled to ensure that a shift in EMSA is due to fully hybridized construct (see section 4.4.9). Target and non-target strand were hybridized by incubation for 10 min at 95 °C and cooling down to RT over 2 hours. A total of 20 nM (~20,000 cpm) of labeled substrate was incubated with varying concentrations of Cascade (0–60 nM) for 30 min at 30 °C in 50 mM HEPES-NaOH pH 7.0, 150 mM NaCl and 1 mM DTT. Samples were mixed with Gel Pilot loading dye (Qiagen) and separated via non-denaturing TBE-PAGE (6% polyacrylamide, 1x TBE).

4.6.10 Nuclease assays

Nuclease assays were performed to study *in vitro* interference of Cas3fv and the Cas1-Cas2/3fv complex. To study activity of Cas3fv on ssDNA, a consistent amount of protein (500 nM) was incubated with 1 µg ssDNA substrate (M13mp18 ssDNA, NEB) over increasing time points at 30 °C in a buffer containing 20 mM HEPES pH 7, 100 mM NaCl, 5 mM MgCl₂, 5 mM MnCl₂ and 1 mM ATP. To compare ssDNA activity of Cas3fv and Cas1-Cas2/3fv, increasing amounts of protein (0, 50, 100, 200 and 500 nM) were incubated with 1 µg ssDNA substrate with the same buffer. 10 mM EDTA was added to stop the reaction in control samples. Remaining ssDNA substrate was separated and visualized by gel electrophoresis in 1x TBE (see section 4.4.4.1).

To study the activity on dsDNA substrates containing a small 10 nt “bubble” opening and Cascade-bound R-loop substrates (created by incubation with Cascade as in 4.6.9), radioactively labeled substrates were used. In these assays, ~20,000 cpm of substrate was incubated without protein or with 500 nM Cas3fv or Cas1-2/3fv for 2 h at 30 °C. Samples were separated via denaturing PAGE (see section 4.5.4.2) and gels were visualized via phosphoimaging. Radioactively labeled Low Molecular Weight Marker (Affymetrix) was used for size determination.

4.6.11 Crystallization and 3D structure analysis of I-Fv Cascade

Crystallization of I-Fv Cascade and solving of the 3D structure was performed by Dr. Patrick Pausch in the laboratory of Prof. Dr. Bange of the Philipps-University Marburg (Pausch *et al.*, 2017). Purified Cascade was concentrated to an absorbance at 280 nm of 35 AU (NanoDrop Lite Spectrophotometer) and crystals were formed by hanging drop vapor-diffusion at 20 °C. Crystallization of the Se-Met labeled Cascade was conducted in a two-step protocol, based on the initial formation of seed-crystals derived from native Cascade complexes. Seed crystals were generated by mixing equal volumes (1 µl) of Cascade sample and crystallization buffer (16% w/v PEG6000, 0.1 M Tris pH 8.0 and 20 mM 5-amino-2,4,6-triiodoisophthalic acid). Sword-shaped seed crystals grew overnight and were subsequently used for streak seeding with a cat whisker into crystallization drops containing Se-Met labeled Cascade. Crystals grew within days in drops containing 1 µl of Se-Met labeled Cascade sample and 1 µl crystallization buffer (18% w/v PEG6000, 0.1 M Tris pH 7.6 and 20 mM 5-amino-2,4,6-triiodoisophthalic acid). Crystals were transferred into crystallization buffer containing 20% v/v glycerol as cryo-protectant, subsequently flash frozen and stored in liquid nitrogen. R-loop/Cascade samples were concentrated to an absorbance at 280 nm of 30 AU (NanoDrop Lite Spectrophotometer) and crystals were formed by hanging drop vapor-diffusion at 20 °C. Needle shaped crystals grew within days in drops containing 1 µl of R-loop/Cascade and 1 µl crystallization buffer (22.5% w/v PEG4000, 15% v/v glycerol, 153 mM ammonium acetate and 85 mM sodium citrate pH 5.6). R-loop/Cascade crystals were flash-frozen and stored in liquid nitrogen.

Diffraction data was collected at beamline ID29 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France. Data was processed with the XDS program package for data reduction (Kabsch, 2010), Crank2 for experimental phasing of Se-Met labeled Cascade (CCP4 package, (Winn *et al.*, 2011), coot (Emsley & Cowtan, 2004) in combination with Refmac5 (CCP4 package) and phenix.refine (PHENIX package) for iterative model building and refinement (Adams *et al.*, 2010). The R-loop/Cascade dataset was solved by molecular replacement using the Cascade structure via the CCP4 implemented program Phaser (McCoy *et al.*, 2007). Figures were prepared in Pymol.

4.6.12 Electron Microscopy

Transmission Electron Microscopy was performed by Dr. Thomas Heimerl of the Philipps-University Marburg to visualize filament structures and small RNA wrapping complexes. 2D Class averaging was performed to obtain a 3D model by combining a total of 30 montages.

4.7 Cell biological methods

4.7.1 Fluorescence Microscopy

Fluorescence Microscopy was performed by Julia Wiegel. Cultures for fluorescence microscopy were grown for 1 h after induction for Cas proteins and sfGFP respectively. After incubation, 4 µl were fixated on a microscope slide covered with 2% agarose. Samples were analyzed with the Axioplan 2 fluorescence microscope (Carl Zeiss Microscopy GmbH) using an ocular (10X) and a Plan-Apochromat (100X), 1.4 oil DIC immersion objective. Pictures were taken with the GFP- and DIC-protocol from Metamorph (version 62r6). The exposure time was 60 ms for DIC and 800 ms for sfGFP.

4.7.2 Fluorescence-activated cell sorting

Fluorescence-activated cell sorting (FACS) analysis was performed by Hamrithaa Shanmuganathan as part of her Master's thesis. For the first experiment with target expressed from pBAD, cultures were grown for 1 h after induction of *cas* genes and *sfgfp* expression respectively. For the second experiment, both *sfgfp* and *cas* genes were expressed by T7 RNAP overnight at 18 °C after induction. After incubation, 30 µl of induced samples were taken and diluted with 1% PBS. Biological triplicates from different cultures were used for calculation of standard deviation and calculation of error bars. FACS analysis was performed at BD LSR Fortessa (BD Biosciences) with a blue laser (detector C) and configuration 4 in CTS setting. Data was processed with BD coherent connection and BD FACSDiva. Two-tailed, unpaired Student's t-tests were performed to calculate the significance of the data using a p-value of 0.05.

5. References

- Abudayyeh, O.O., J.S. Gootenberg, S. Konermann, J. Joung, I.M. Slaymaker, D.B. Cox, S. Shmakov, K.S. Makarova, E. Semenova, L. Minakhin, K. Severinov, A. Regev, E.S. Lander, E.V. Koonin & F. Zhang, (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* 353: aaf5573.
- Adams, P.D., P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger & P.H. Zwart, (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* 66: 213-221.
- Athukoralage, J.S., C. Rouillon, S. Graham, S. Gruschow & M.F. White, (2018). Ring nucleases deactivate type III CRISPR ribonucleases by degrading cyclic oligoadenylate. *Nature* 562: 277-280.
- Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D.A. Romero & P. Horvath, (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315: 1709-1712.
- Beloglazova, N., K. Kuznedelov, R. Flick, K.A. Datsenko, G. Brown, A. Popovic, S. Lemak, E. Semenova, K. Severinov & A.F. Yakunin, (2015). CRISPR RNA binding and DNA target recognition by purified Cascade complexes from *Escherichia coli*. *Nucleic acids research* 43: 530-543.
- Beloglazova, N., P. Petit, R. Flick, G. Brown, A. Savchenko & A.F. Yakunin, (2011). Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *Embo J* 30: 4616-4627.
- Bondy-Denomy, J., A. Pawluk, K.L. Maxwell & A.R. Davidson, (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* 493: 429-432.
- Bradford, M.M., (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* 72: 248-254.
- Brouns, S.J., M.M. Jore, M. Lundgren, E.R. Westra, R.J. Slijkhuis, A.P. Snijders, M.J. Dickman, K.S. Makarova, E.V. Koonin & J. Van Der Oost, (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321: 960-964.
- Cady, K.C., J. Bondy-Denomy, G.E. Heussler, A.R. Davidson & G.A. O'Toole, (2012). The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *Journal of bacteriology* 194: 5728-5738.
- Cameron, D.E. & J.J. Collins, (2014). Tunable protein degradation in bacteria. *Nature biotechnology* 32: 1276-1281.
- Carte, J., N.T. Pfister, M.M. Compton, R.M. Terns & M.P. Terns, (2010). Binding and cleavage of CRISPR RNA by Cas6. *Rna* 16: 2181-2188.
- Cheng, A.W., H. Wang, H. Yang, L. Shi, Y. Katz, T.W. Theunissen, S. Rangarajan, C.S. Shivalila, D.B. Dadon & R. Jaenisch, (2013). Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell research* 23: 1163-1171.
- Chowdhury, S., J. Carter, M.F. Rollins, S.M. Golden, R.N. Jackson, C. Hoffmann, L. Nosaka, J. Bondy-Denomy, K.L. Maxwell, A.R. Davidson, E.R. Fischer, G.C. Lander & B. Wiedenheft, (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell* 169: 47-57 e11.

- Cordin, O., J. Banroques, N.K. Tanner & P. Linder, (2006). The DEAD-box protein family of RNA helicases. *Gene* 367: 17-37.
- Cox, D.B.T., J.S. Gootenberg, O.O. Abudayyeh, B. Franklin, M.J. Kellner, J. Joung & F. Zhang, (2017). RNA editing with CRISPR-Cas13. *Science* 358: 1019-1027.
- Deveau, H., R. Barrangou, J.E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D.A. Romero, P. Horvath & S. Moineau, (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of bacteriology* 190: 1390-1400.
- Dwarakanath, S., (2015). Characterization of a minimal Type I CRISPR-Cas system found in *Shewanella putrefaciens* CN-32. Dissertation. Philipps-Universität Marburg.
- Dwarakanath, S., S. Brenzinger, D. Gleditsch, A. Plagens, A. Klingl, K. Thormann & L. Randau, (2015). Interference activity of a minimal Type I CRISPR-Cas system from *Shewanella putrefaciens*. *Nucleic acids research* 43: 8913-8923.
- Elmore, J.R., N.F. Sheppard, N. Ramia, T. Deighan, H. Li, R.M. Terns & M.P. Terns, (2016). Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR-Cas system. *Genes & development* 30: 447-459.
- Emsley, P. & K. Cowtan, (2004). Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* 60: 2126-2132.
- Estrella, M.A., F.T. Kuo & S. Bailey, (2016). RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes & development* 30: 460-470.
- Fagerlund, R.D., M.E. Wilkinson, O. Klykov, A. Barendregt, F.G. Pearce, S.N. Kieper, H.W.R. Maxwell, A. Capolupo, A.J.R. Heck, K.L. Krause, M. Bostina, R.A. Scheltema, R.H.J. Staals & P.C. Fineran, (2017). Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex. *Proceedings of the National Academy of Sciences of the United States of America* 114: E5122-E5128.
- Fairman-Williams, M.E., U.P. Guenther & E. Jankowsky, (2010). SF1 and SF2 helicases: family matters. *Curr Opin Struct Biol* 20: 313-324.
- Fineran, P.C., M.J. Gerritzen, M. Suarez-Diez, T. Kunne, J. Boekhorst, S.A. van Hijum, R.H. Staals & S.J. Brouns, (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *Proceedings of the National Academy of Sciences of the United States of America* 111: E1629-1638.
- Forconi, M. & D. Herschlag, (2009). Metal Ion-Based RNA Cleavage as a Structural Probe. *Method Enzymol* 468: 91-106.
- Fu, C.L., W.P. Donovan, O. Shikapwashya-Hasser, X.D. Ye & R.H. Cole, (2014). Hot Fusion: An Efficient Method to Clone Multiple DNA Fragments as Well as Inverted Repeats without Ligase. *PLoS one* 9(12): e115318.
- Fu, Y., J.D. Sander, D. Reyon, V.M. Cascio & J.K. Joung, (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology* 32: 279-284.
- Gibson, D.G., L. Young, R.Y. Chuang, J.C. Venter, C.A. Hutchison & H.O. Smith, (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods* 6: 343-345.
- Gleditsch, D., (2015). Characterization of a minimal Type I CRISPR-Cas interference complex. Master Thesis. Technische Universität Darmstadt.
- Gleditsch, D., H. Müller-Esparza, P. Pausch, K. Sharma, S. Dwarakanath, H. Urlaub, G. Bange & L. Randau, (2016). Modulating the Cascade architecture of a minimal Type I-F CRISPR-Cas system. *Nucleic Acids Research* 44: 5872-5882.

- Gong, B., M. Shin, J. Sun, C.H. Jung, E.L. Bolt, J. van der Oost & J.S. Kim, (2014). Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proceedings of the National Academy of Sciences of the United States of America* 111: 16359-16364.
- Gootenberg, J.S., O.O. Abudayyeh, J.W. Lee, P. Essletzbichler, A.J. Dy, J. Joung, V. Verdine, N. Donghia, N.M. Daringer, C.A. Freije, C. Myhrvold, R.P. Bhattacharyya, J. Livny, A. Regev, E.V. Koonin, D.T. Hung, P.C. Sabeti, J.J. Collins & F. Zhang, (2017). Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 356: 438-442.
- Grissa, I., G. Vergnaud & C. Pourcel, (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids research* 35: W52-57.
- Gu, D.H., S.C. Ha & J.S. Kim, (2019). A CRISPR RNA Is Closely Related With the Size of the Cascade Nucleoprotein Complex. *Frontiers in microbiology* 10: 2458.
- Guo, T.W., A. Bartesaghi, H. Yang, V. Falconieri, P. Rao, A. Merk, E.T. Eng, A.M. Raczkowski, T. Fox, L.A. Earl, D.J. Patel & S. Subramaniam, (2017). Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell* 171: 414-426 e412.
- Gur, E. & R.T. Sauer, (2008). Evolution of the *ssrA* degradation tag in *Mycoplasma*: specificity switch to a different protease. *Proceedings of the National Academy of Sciences of the United States of America* 105: 16113-16118.
- Hale, C.R., P. Zhao, S. Olson, M.O. Duff, B.R. Graveley, L. Wells, R.M. Terns & M.P. Terns, (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* 139: 945-956.
- Hanahan, D., (1983). Studies on Transformation of *Escherichia Coli* with Plasmids. *Journal of molecular biology* 166: 557-580.
- Hannon, G.J., (2002). RNA interference. *Nature* 418: 244-251.
- Hatoum-Aslan, A., P. Samai, I. Maniv, W. Jiang & L.A. Marraffini, (2013). A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *The Journal of biological chemistry* 288: 27888-27897.
- Hayes, R.P., Y. Xiao, F. Ding, P.B. van Erp, K. Rajashankar, S. Bailey, B. Wiedenheft & A. Ke, (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature* 530: 499-503.
- Hille, F., H. Richter, S.P. Wong, M. Bratovic, S. Ressel & E. Charpentier, (2018). The Biology of CRISPR-Cas: Backward and Forward. *Cell* 172: 1239-1259.
- Hochstrasser, M.L., D.W. Taylor, P. Bhat, C.K. Guegler, S.H. Sternberg, E. Nogales & J.A. Doudna, (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proceedings of the National Academy of Sciences of the United States of America* 111: 6618-6623.
- Hochstrasser, M.L., D.W. Taylor, J.E. Kornfeld, E. Nogales & J.A. Doudna, (2016). DNA Targeting by a Minimal CRISPR RNA-Guided Cascade. *Molecular cell* 63: 840-851.
- Huo, Y., K.H. Nam, F. Ding, H. Lee, L. Wu, Y. Xiao, M.D. Farchione, Jr., S. Zhou, K. Rajashankar, I. Kurinov, R. Zhang & A. Ke, (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nature structural & molecular biology* 21: 771-777.
- Inoue, H., H. Nojima & H. Okayama, (1990). High-Efficiency Transformation of *Escherichia-Coli* with Plasmids. *Gene* 96: 23-28.

- Jackson, R.N., S.M. Golden, P.B. van Erp, J. Carter, E.R. Westra, S.J. Brouns, J. van der Oost, T.C. Terwilliger, R.J. Read & B. Wiedenheft, (2014). Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* 345: 1473-1479.
- Jackson, R.N., M. Lavin, J. Carter & B. Wiedenheft, (2014). Fitting CRISPR-associated Cas3 into the Helicase Family Tree. *Curr Opin Struc Biol* 24: 106-114.
- Jahn, M.J., D. Jahn, A.M. Kumar & D. Soll, (1991). Mono-Q Chromatography Permits Recycling of DNA-Template and Purification of Rna Transcripts after T7-Rna Polymerase Reaction. *Nucleic Acids Research* 19: 2786-2786.
- Jansen, R., J.D.A. van Embden, W. Gaastra & L.M. Schouls, (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology* 43: 1565-1575.
- Jia, N., C.Y. Mo, C.Y. Wang, E.T. Eng, L.A. Marraffini & D.J. Patel, (2019). Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity. *Molecular Cell* 73: 264-+.
- Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna & E. Charpentier, (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816-821.
- Jore, M.M., M. Lundgren, E. van Duijn, J.B. Bultema, E.R. Westra, S.P. Waghmare, B. Wiedenheft, Ü. Pul, R. Wurm & R. Wagner, (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nature structural & molecular biology* 18: 529-536.
- Kabsch, W., (2010). Xds. *Acta crystallographica. Section D, Biological crystallography* 66: 125-132.
- Kazlauskienė, M., G. Tamulaitis, G. Kostiuk, C. Venclovas & V. Siksnys, (2016). Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Molecular cell* 62: 295-306.
- Kohler, R., R.A. Mooney, D.J. Mills, R. Landick & P. Cramer, (2017). Architecture of a transcribing-translating expressome. *Science* 356: 194-197.
- Konarev, P.V., V.V. Volkov, A.V. Sokolova, M.H.J. Koch & D.I. Svergun, (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36: 1277-1282.
- Koonin, E.V. & V.V. Dolja, (2013). A virocentric perspective on the evolution of life. *Current opinion in virology* 3: 546-557.
- Koonin, E.V. & K.S. Makarova, (2019). Origins and evolution of CRISPR-Cas systems. *Philos T R Soc B* 374.
- Koonin, E.V. & K.S. Makarova, (2019). Origins and evolution of CRISPR-Cas systems. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 374: 20180087.
- Koonin, E.V., K.S. Makarova & Y.I. Wolf, (2017). Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annual review of microbiology* 71: 233-261.
- Koonin, E.V., K.S. Makarova & F. Zhang, (2017). Diversity, classification and evolution of CRISPR-Cas systems. *Current opinion in microbiology* 37: 67-78.
- Kunne, T., S.N. Kieper, J.W. Bannenberg, A.I. Vogel, W.R. Mielliet, M. Klein, M. Depken, M. Suarez-Diez & S.J. Brouns, (2016). Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Molecular cell* 63: 852-864.
- Kuznedelov, K., V. Mekler, S. Lemak, M. Tokmina-Lukaszewska, K.A. Datsenko, I. Jain, E. Savitskaya, J. Mallon, S. Shmakov, B. Bothner, S. Bailey, A.F. Yakunin, K. Severinov & E. Semenova, (2016). Altered stoichiometry *Escherichia coli* Cascade complexes with shortened CRISPR RNA spacers are capable of interference and primed adaptation. *Nucleic Acids Research* 22: 10849-10861.

- Landick, R., J. Carey & C. Yanofsky, (1985). Translation Activates the Paused Transcription Complex and Restores Transcription of the Trp Operon Leader Region. *Proceedings of the National Academy of Sciences of the United States of America* 82: 4663-4667.
- Larson, M.H., L.A. Gilbert, X. Wang, W.A. Lim, J.S. Weissman & L.S. Qi, (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* 8: 2180-2196.
- Loeff, L., S.J.J. Brouns & C. Joo, (2018). Repetitive DNA Reeling by the Cascade-Cas3 Complex in Nucleotide Unwinding Steps. *Molecular cell* 70: 385-394 e383.
- Luo, M.L., R.N. Jackson, S.R. Denny, M. Tokmina-Lukaszewska, K.R. Maksimchuk, W. Lin, B. Bothner, B. Wiedenheft & C.L. Beisel, (2016). The CRISPR RNA-guided surveillance complex in *Escherichia coli* accommodates extended RNA spacers. *Nucleic Acids Research* 44: 7385-7394.
- Maeder, M.L., S.J. Linder, V.M. Cascio, Y. Fu, Q.H. Ho & J.K. Joung, (2013). CRISPR RNA-guided activation of endogenous human genes. *Nature methods* 10: 977-979.
- Makarova, K.S., L. Aravind, N.V. Grishin, I.B. Rogozin & E.V. Koonin, (2002). A DNA repair system specific for thermophilic Archaea and Bacteria predicted by genomic context analysis. *Nucleic Acids Research* 30: 482-496.
- Makarova, K.S., D.H. Haft, R. Barrangou, S.J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F.J. Mojica, Y.I. Wolf, A.F. Yakunin, J. van der Oost & E.V. Koonin, (2011). Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology* 9: 467-477.
- Makarova, K.S., Y.I. Wolf, O.S. Alkhnbashi, F. Costa, S.A. Shah, S.J. Saunders, R. Barrangou, S.J. Brouns, E. Charpentier, D.H. Haft, P. Horvath, S. Moineau, F.J. Mojica, R.M. Terns, M.P. Terns, M.F. White, A.F. Yakunin, R.A. Garrett, J. van der Oost, R. Backofen & E.V. Koonin, (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology* 13: 722-736.
- Makarova, K.S., Y.I. Wolf & E.V. Koonin, (2013). Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Research* 41: 4360-4377.
- McCoy, A.J., R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni & R.J. Read, (2007). Phaser crystallographic software. *J Appl Crystallogr* 40: 658-674.
- Mirsky, I.A., G. Perisutti & R. Jinks, (1963). Isolation and Crystallization of Human Insulin. *Journal of Clinical Investigation* 42: 1869-&.
- Mojica, F.J., C. Diez-Villasenor, J. Garcia-Martinez & E. Soria, (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60: 174-182.
- Moreira, D. & P. Lopez-Garcia, (2009). Ten reasons to exclude viruses from the tree of life. *Nature reviews. Microbiology* 7: 306-311.
- Mulepati, S. & S. Bailey, (2011). Structural and Biochemical Analysis of Nuclease Domain of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Protein 3 (Cas3). *Journal of Biological Chemistry* 286: 31896-31903.
- Mulepati, S. & S. Bailey, (2013). *In Vitro* Reconstitution of an *Escherichia coli* RNA-guided Immune System Reveals Unidirectional, ATP-dependent Degradation of DNA Target. *Journal of Biological Chemistry* 288: 22184-22192.
- Mulepati, S., A. Heroux & S. Bailey, (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* 345: 1479-1484.
- Müller-Esparza, H.C., (2019). Characterization of DNA interference by a minimal Type I-F CRISPR-Cas system. Dissertation. Philipps-Universität Marburg.

- Nierhaus, K.H., (2014). Mg²⁺, K⁺, and the Ribosome. *Journal of bacteriology* 196: 3817-3819.
- Niewoehner, O., C. Garcia-Doval, J.T. Rostol, C. Berk, F. Schwede, L. Bigler, J. Hall, L.A. Marraffini & M. Jinek, (2017). Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* 548: 543-548.
- Nimkar, S. & B. Anand, (2019). Cas3 Mediated Target DNA Recognition and Cleavage is Independent of the Composition and Architecture of Cascade Surveillance Complex. *bioRxiv*: 666776.
- Nunez, J.K., P.J. Kranzusch, J. Noeske, A.V. Wright, C.W. Davies & J.A. Doudna, (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nature structural & molecular biology* 21: 528-534.
- Ozcan, A., P. Pausch, A. Linden, A. Wulf, K. Schuhle, J. Heider, H. Urlaub, T. Heimerl, G. Bange & L. Randau, (2019). Type IV CRISPR RNA processing and effector complex formation in *Aromatoleum aromaticum*. *Nature microbiology* 4: 89-96.
- Pausch, P., H. Muller-Esparza, D. Gleditsch, F. Altegoer, L. Randau & G. Bange, (2017). Structural Variation of Type I-F CRISPR RNA Guided DNA Surveillance. *Molecular cell* 67: 622-632 e624.
- Pawluk, A., J. Bondy-Denomy, V.H. Cheung, K.L. Maxwell & A.R. Davidson, (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of *Pseudomonas aeruginosa*. *mBio* 5: e00896.
- Pawluk, A., R.H. Staals, C. Taylor, B.N. Watson, S. Saha, P.C. Fineran, K.L. Maxwell & A.R. Davidson, (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nature microbiology* 1: 16085.
- Perez-Pinera, P., D.D. Kocak, C.M. Vockley, A.F. Adler, A.M. Kabadi, L.R. Polstein, P.I. Thakore, K.A. Glass, D.G. Ousterout, K.W. Leong, F. Guilak, G.E. Crawford, T.E. Reddy & C.A. Gersbach, (2013). RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nature methods* 10: 973-976.
- Plagens, A. & L. Randau, (2015). *In Vitro* Co-reconstitution of Cas Protein Complexes. *Methods in molecular biology* 1311: 23-33.
- Plagens, A., H. Richter, E. Charpentier & L. Randau, (2015). DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS microbiology reviews* 39: 442-463.
- Plagens, A., V. Tripp, M. Daume, K. Sharma, A. Klingl, A. Hrle, E. Conti, H. Urlaub & L. Randau, (2014). *In vitro* assembly and activity of an archaeal CRISPR-Cas type I-A Cascade interference complex. *Nucleic acids research* 42: 5125-5138.
- Proshkin, S., A.R. Rahmouni, A. Mironov & E. Nudler, (2010). Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. *Science* 328: 504-508.
- Qi, L.S., M.H. Larson, L.A. Gilbert, J.A. Doudna, J.S. Weissman, A.P. Arkin & W.A. Lim, (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152: 1173-1183.
- Ran, F.A., P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott & F. Zhang, (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8: 2281-2308.
- Rath, D., L. Amlinger, M. Hoekzema, P.R. Devulapally & M. Lundgren, (2015). Efficient programmable gene silencing by Cascade. *Nucleic Acids Research* 43: 237-246.
- Richter, C., R.L. Dy, R.E. McKenzie, B.N. Watson, C. Taylor, J.T. Chang, M.B. McNeil, R.H. Staals & P.C. Fineran, (2014). Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic acids research* 42: 8516-8526.

- Richter, C., T. Gristwood, J.S. Clulow & P.C. Fineran, (2012). *In vivo* protein interactions and complex formation in the *Pectobacterium atrosepticum* subtype I-F CRISPR/Cas System. *PLoS one* 7: e49549.
- Rollins, M.F., S. Chowdhury, J. Carter, S.M. Golden, H.M. Miettinen, A. Santiago-Frangos, D. Faith, C.M. Lawrence, G.C. Lander & B. Wiedenheft, (2019) Structure Reveals a Mechanism of CRISPR-RNA-Guided Nuclease Recruitment and Anti-CRISPR Viral Mimicry. *Molecular cell* 74: 132-142 e135.
- Rollins, M.F., S. Chowdhury, J. Carter, S.M. Golden, R.A. Wilkinson, J. Bondy-Denomy, G.C. Lander & B. Wiedenheft, (2017). Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proceedings of the National Academy of Sciences of the United States of America* 114: E5113-E5121.
- Rollins, M.F., J.T. Schuman, K. Paulus, H.S.T. Bukhari & B. Wiedenheft, (2015). Mechanism of foreign DNA recognition by a CRISPR RNA-guided surveillance complex from *Pseudomonas aeruginosa*. *Nucleic Acids Research* 43: 2216-2222.
- Samai, P., N. Pyenson, W. Jiang, G.W. Goldberg, A. Hatoum-Aslan & L.A. Marraffini, (2015). Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell* 161: 1164-1174.
- Samson, J.E., A.H. Magadan, M. Sabri & S. Moineau, (2013). Revenge of the phages: defeating bacterial defences. *Nature reviews. Microbiology* 11: 675-687.
- Sashital, D.G., B. Wiedenheft & J.A. Doudna, (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. *Molecular cell* 46: 606-615.
- Sinkunas, T., G. Gasiunas, C. Fremaux, R. Barrangou, P. Horvath & V. Siksnys, (2011). Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *Embo J* 30: 1335-1342.
- Sinkunas, T., G. Gasiunas, S.P. Waghmare, M.J. Dickman, R. Barrangou, P. Horvath & V. Siksnys, (2013). *In vitro* reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *Embo J* 32: 385-394.
- Smargon, A.A., D.B.T. Cox, N.K. Pyzocha, K. Zheng, I.M. Slaymaker, J.S. Gootenberg, O.A. Abudayyeh, P. Essletzbichler, S. Shmakov, K.S. Makarova, E.V. Koonin & F. Zhang, (2017). Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Molecular cell* 65: 618-630 e617.
- Songailiene, I., M. Rutkauskas, T. Sinkunas, E. Manakova, S. Wittig, C. Schmidt, V. Siksnys & R. Seidel, (2019). Decision-Making in Cascade Complexes Harboring crRNAs of Altered Length. *Cell reports* 28: 3157-3166 e3154.
- Staals, R.H., S.A. Jackson, A. Biswas, S.J. Brouns, C.M. Brown & P.C. Fineran, (2016). Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nature communications* 7: 12853.
- Steube, N., (2018). Characterization of Anti-CRISPR proteins against three variants of subtype I-F CRISPR-Cas systems. Master Thesis. Philipps-Universität Marburg.
- Tay, M., S. Liu & Y.A. Yuan, (2015). Crystal structure of *Thermobifida fusca* Cse1 reveals target DNA binding site. *Protein science : a publication of the Protein Society* 24: 236-245.
- Vercoe, R.B., J.T. Chang, R.L. Dy, C. Taylor, T. Gristwood, J.S. Clulow, C. Richter, R. Przybilski, A.R. Pitman & P.C. Fineran, (2013). Cytotoxic Chromosomal Targeting by CRISPR/Cas Systems Can Reshape Bacterial Genomes and Expel or Remodel Pathogenicity Islands. *PLoS Genetics* 9.

- Vorontsova, D., K.A. Datsenko, S. Medvedeva, J. Bondy-Denomy, E.E. Savitskaya, K. Pougach, M. Logacheva, B. Wiedenheft, A.R. Davidson, K. Severinov & E. Semenova, (2015). Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. *Nucleic acids research* 43: 10848-10860.
- Wang, X., D. Yao, J.G. Xu, A.R. Li, J. Xu, P. Fu, Y. Zhou & Y. Zhu, (2016). Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nature structural & molecular biology* 23: 868-870.
- Westra, E.R., E. Semenova, K.A. Datsenko, R.N. Jackson, B. Wiedenheft, K. Severinov & S.J. Brouns, (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet* 9: e1003742.
- Westra, E.R., P.B. van Erp, T. Kunne, S.P. Wong, R.H. Staals, C.L. Seegers, S. Bollen, M.M. Jore, E. Semenova, K. Severinov, W.M. de Vos, R.T. Dame, R. de Vries, S.J. Brouns & J. van der Oost, (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Molecular cell* 46: 595-605.
- Winn, M.D., C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A. Vagin & K.S. Wilson, (2011). Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* 67: 235-242.
- Xiao, Y., M. Luo, A.E. Dolan, M. Liao & A. Ke, (2018). Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science* 361.
- Xiao, Y., M. Luo, R.P. Hayes, J. Kim, S. Ng, F. Ding, M. Liao & A. Ke, (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell* 170: 48-60 e11.
- You, L.L., J. Ma, J.Y. Wang, D. Artamonova, M. Wang, L. Liu, H. Xiang, K. Severinov, X.Z. Zhang & Y.L. Wang, (2019). Structure Studies of the CRISPR-Csm Complex Reveal Mechanism of Co-transcriptional Interference. *Cell* 176: 239-253.
- Zamore, P.D., T. Tuschl, P.A. Sharp & D.P. Bartel, (2000). RNAi: Double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* 101: 25-33.
- Zetsche, B., J.S. Gootenberg, O.O. Abudayyeh, I.M. Slaymaker, K.S. Makarova, P. Essletzbichler, S.E. Volz, J. Joung, J. van der Oost, A. Regev, E.V. Koonin & F. Zhang, (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163: 759-771.
- Zhao, H., G. Sheng, J. Wang, M. Wang, G. Bunkoczi, W. Gong, Z. Wei & Y. Wang, (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* 515: 147-150.

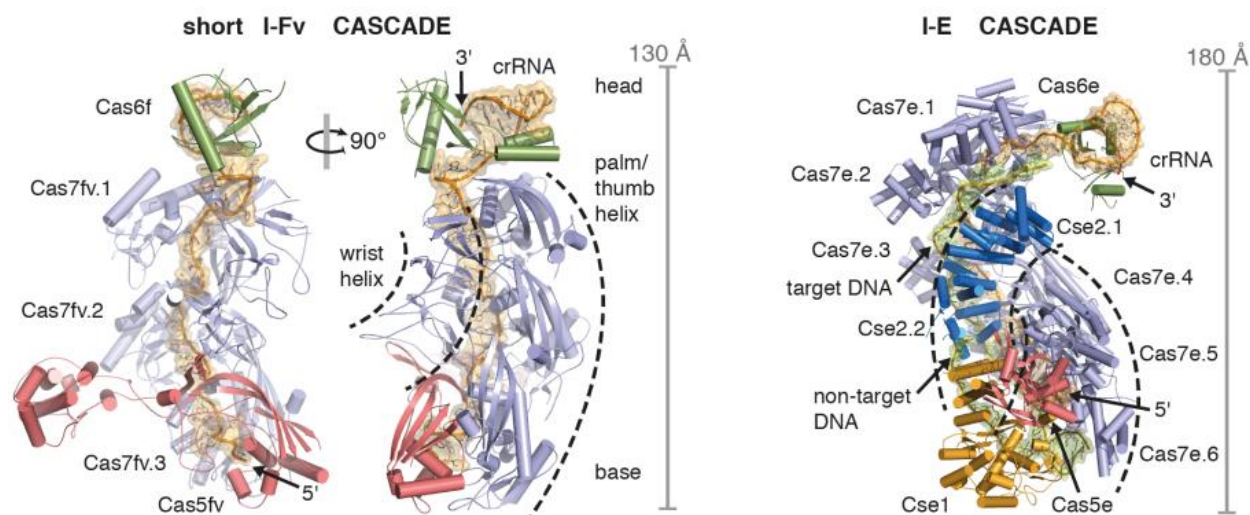
6. Supplementary Material

Table 6.1: Crystallographic Table

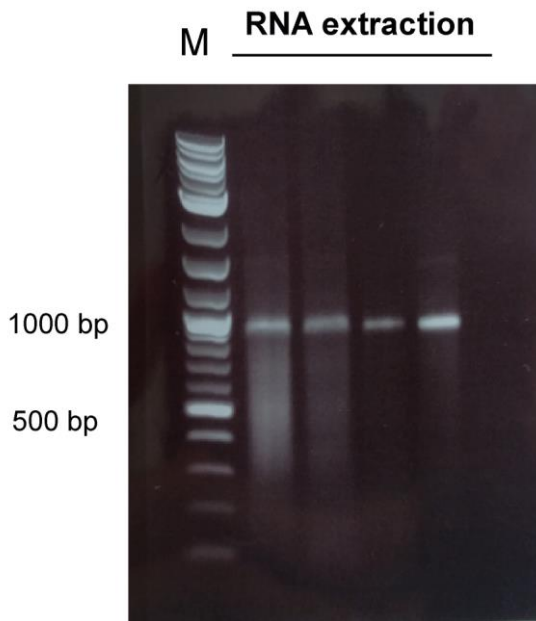
Data Collection ^a	SpCascade-I-Fv (SeMet)	SpCascade-I-Fv-R-Loop
Space group	I2	P 3 ₂ 21
Cell dimensions		
a, b, and c (Å)	157.002	143.316
-	65.894	143.316
-	160.682	172.698
α, β, and γ (°)	90.00	90.00
-	98.61	90.00
-	90.00	120.00
Energy (Å)	0.979	0.991
Resolution (Å)	47.97 - 3.00	45.27 - 3.25
-	(3.10 - 3.00)	(3.36 - 3.25)
R _{merge}	0.0547 (0.356)	0.168 (1.922)
I / σI	13.18 (2.05)	11.66 (1.80)
Completeness (%)	100.0 (100.0)	100.0 (100.0)
Redundancy	10.0 (9.8)	11.1 (11.4)
CC(1/2)	0.99 (0.72)	0.99 (0.85)
Anomalous completeness (%)	99.5 (99.8)	-
Anomalous redundancy	5.2 (5.3)	-

Refinement		
Resolution (Å)	49.28 - 3.00	46.91 - 3.25
No. reflections	32,875 (3,271)	32,783 (3,192)
R _{work} / R _{free}	19.4	21.8
-	24.9	27.7
No. atoms	11,803	13,131
Macromolecule	11,803	13,131
Ligand	0	0
Water	0	0
R.m.s deviations	-	-
Bond lengths (Å)	0.015	0.010
Bond angles (°)	1.34	1.29
Ramachandran (%)	-	-
Preferred	95.93	96.76
Allowed	3.92	3.1
Outliers	0.15	0.14

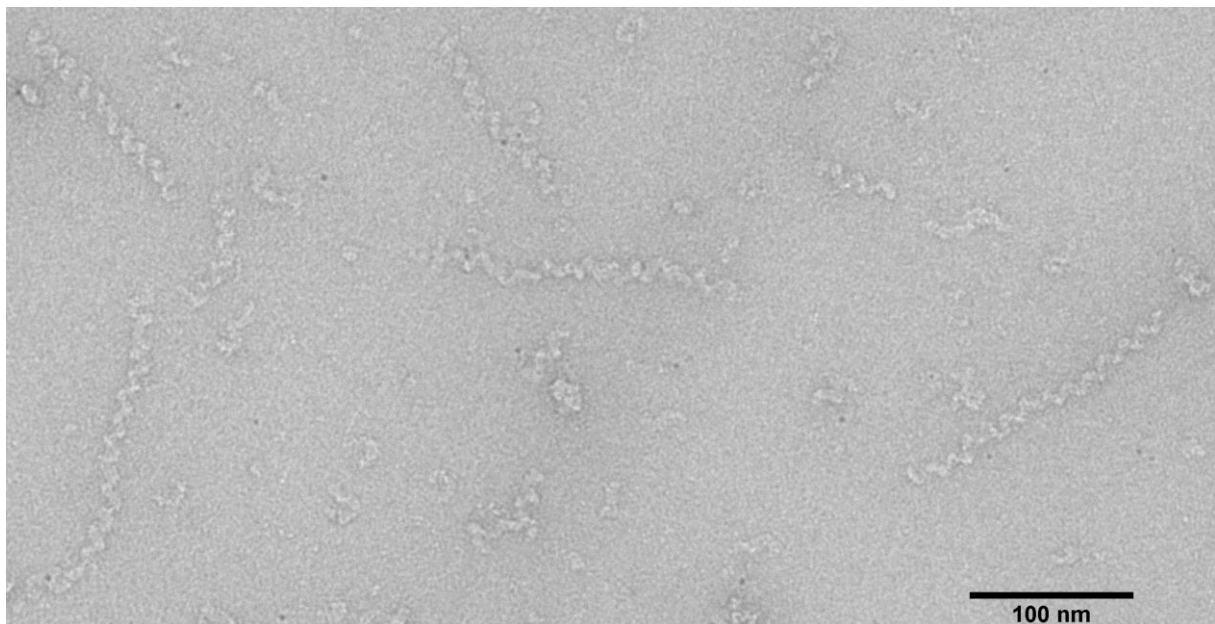
^a Statistics for the highest-resolution shell are shown in parentheses.



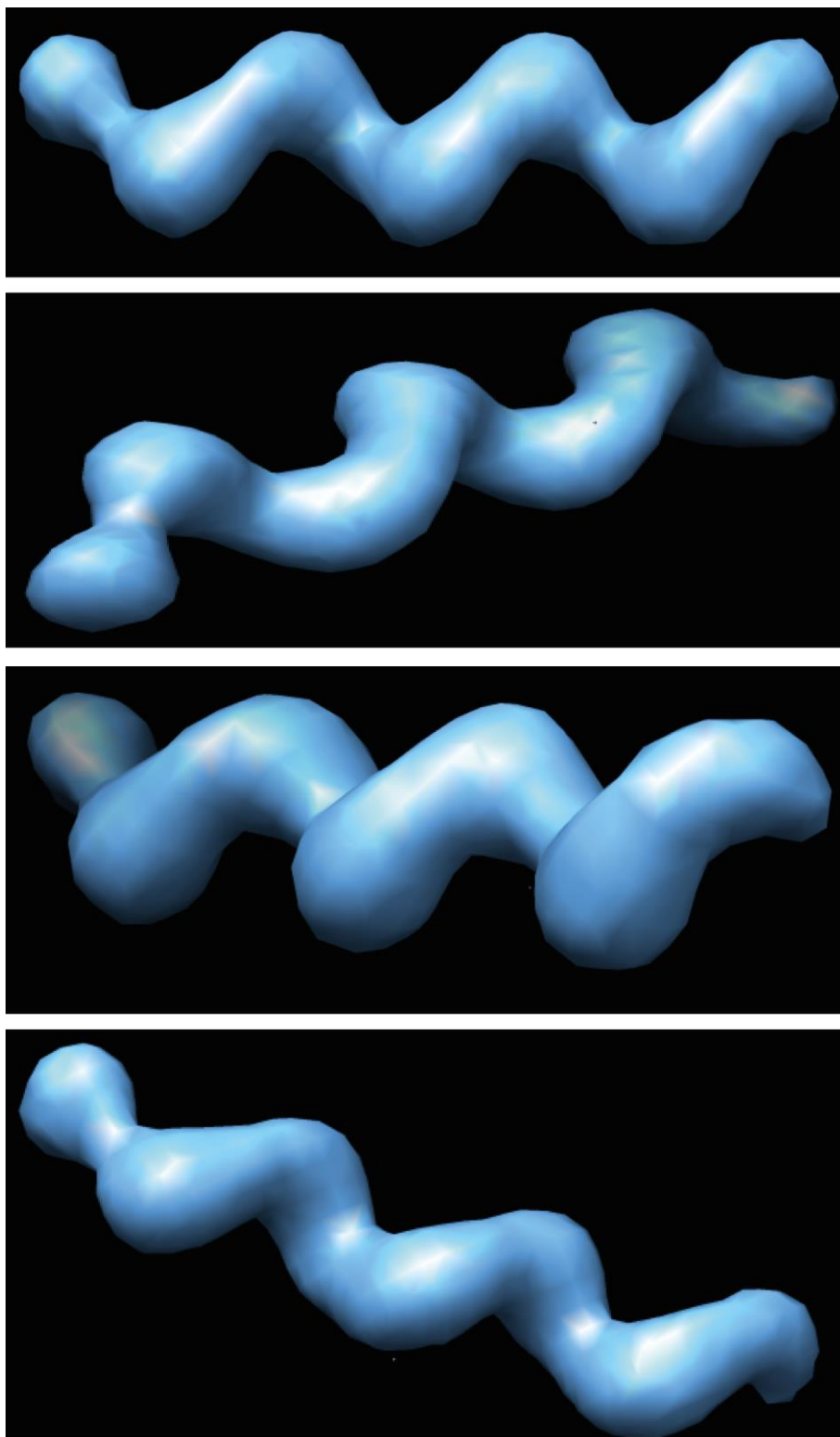
Supplementary Figure 1: Comparison of the short type I-Fv and type I-E Cascade. Left: The cartoon shows the crystal structure of the type I-Fv Cascade complex from *S. putrefaciens* CN32 in two different orientations. The crRNA, Cas6f, the three Cas7fv proteins and Cas5fv are shown in orange, green, blue and red, respectively. The 3' and 5' ends of the crRNA are indicated. Dotted lines indicate the wrist helix and palm/thumb helix formed by Cas7fv.1-3/Cas5fv. Right: Cartoon of the R-loop/type I-E Cascade from *E. coli*. The Cas7e.1-6, Cas6e, Cas5e, Cse1 (Cas8e) and Cse2.1-2 are in light blue, green, red, yellow and dark blue, respectively. The crRNA is represented as orange cartoon/surface. Target and non-target DNA are shown as light-green cartoon/surface. The coordinates are derived from PDB-ID: 5H9A, Figure from Pausch *et al.*, 2017.



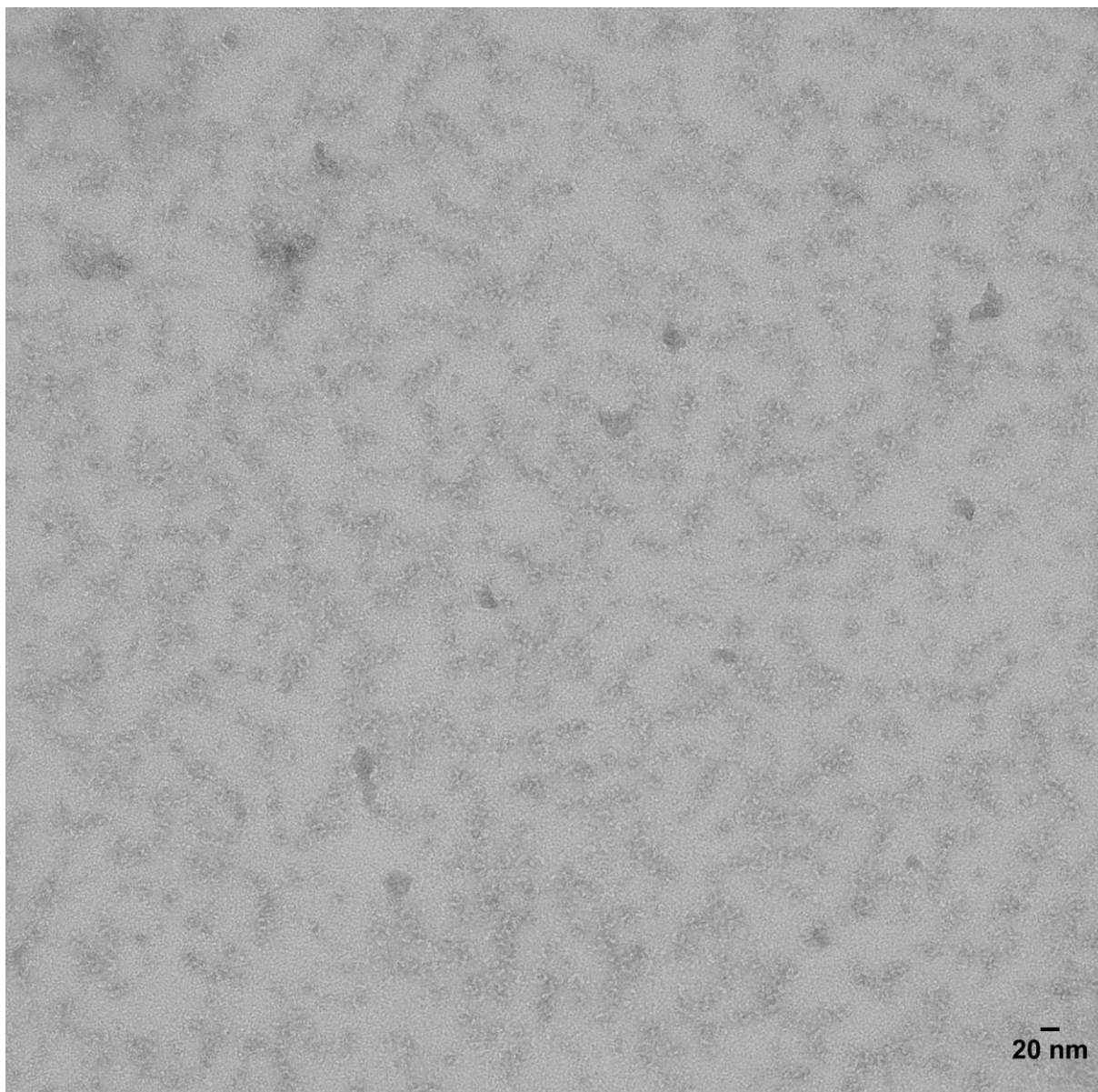
Supplementary Figure 2: LacZ-repeat RNA extraction. RNA was extracted from various samples of the Ni-NTA purification of Cas proteins and repeat-tagged RNA. Extracted RNA was separated by agarose gel electrophoresis in which 2-log DNA ladder (M) was loaded for size reference.



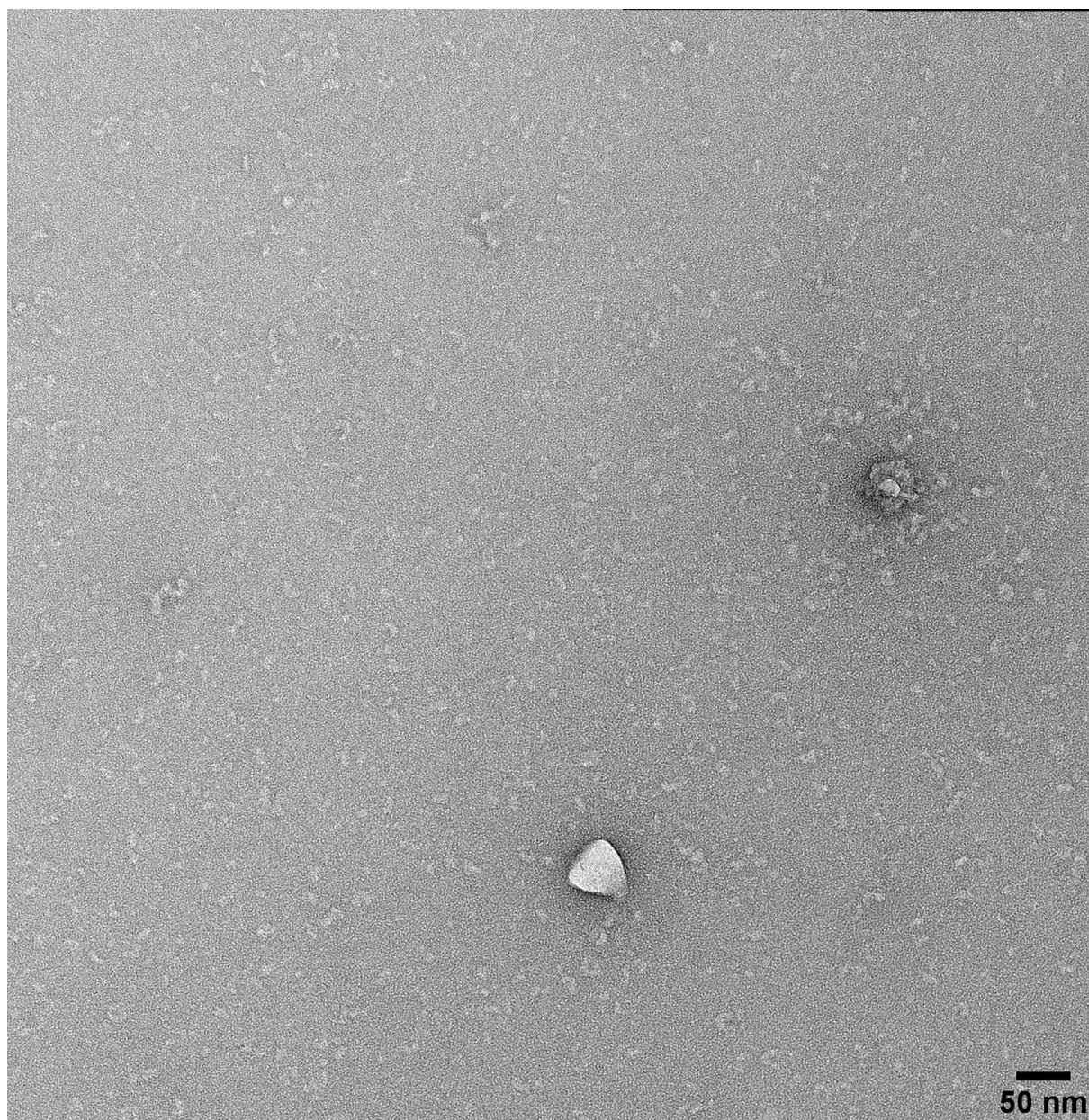
Supplementary Figure 3: TEM analysis of filaments from lacZ-Repeat.



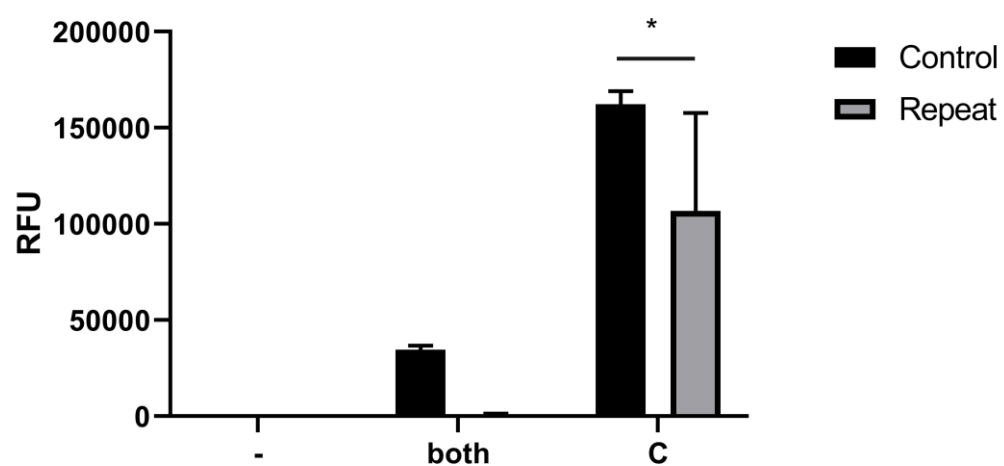
Supplementary Figure 4: 2D class averaging model of filaments structures from TEM analysis (see Figure 2.18).



Supplementary Figure 5: TEM analysis of first half of the peak (fraction of 13ml elution volume) after size-exclusion of MonoQ purification.



Supplementary Figure 6: TEM analysis of MonoQ purification sample (fraction of 13 ml elution volume during salt gradient).



Supplementary Figure 7: FACS analysis of *E. coli* expressing *cas* genes as well as either the control or the repeat-tagged construct overnight at 18 °C after induction. Fluorescence was measured in relative fluorescence units (RFU). Cells not producing sfgFP at all (-) showed no fluorescence at all and a control expressing only *sfgfp* (C) showed the maximum fluorescence.

Abgrenzung der Eigenleistung

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

Dr. Patrick Pausch:

Hat im Rahmen seiner Promotion, die Kristallstruktur von I-Fv Cascade und der Cas7fv Helix gelöst und die dazugehörigen Abbildungen erstellt (Figure 2.4 -2.7 und Supplementary Figure 1). Zusätzlich hat er das Cas3fv-Konstrukt kloniert und die Expressionsbedingungen für Cas3fv optimiert.

Dr. Thomas Heimerl:

Hat als Kollaborationspartner die elektronenmikroskopischen Aufnahmen der Filamente und der synthetischen Cascade-Varianten erstellt, sowie die Modelle durch 2D Class averaging erzeugt.

Julia Wiegel:

Hat die Fluoreszenzmikroskopie-Experimente für sfGFP-Silencing durchgeführt.

Hamrithaa Shanmuganathan:

Hat im Rahmen ihres Forschungspraktikums das sfGFP-Cas7fv Fusions-Konstrukt kloniert und im Rahmen ihrer Masterarbeit die FACS-Experimente für sfGFP-Silencing durchgeführt.

Danksagung

Mein ganz besonderer Dank gilt Herrn Prof. Dr. Lennart Randau für die Möglichkeit diese Doktorarbeit in seinem Labor durchzuführen, aber auch für die ständige Gesprächsbereitschaft, konsequente Beratung und den stetigen Enthusiasmus an meiner Arbeit.

Herzlicher Dank gilt meinem Thesis Advisory Committee, bestehend aus Prof. Dr. Gert Bange und Prof. Dr. Anke Becker für die hilfreichen Ratschläge und Anregungen während meiner Promotion. Zusätzlicher Dank geht an Prof. Dr. Bange für die Zweitkorrektur dieser Arbeit. Dem gesamten Komitee sowie Prof. Dr. Tobias Erb ist außerdem für die Teilnahme an meiner Prüfungskommission zu danken.

Ein elementarer Bestandteil dieser Arbeit war die Kooperation mit der Arbeitsgruppe Bange der Phillips-Universität Marburg, wofür ich nochmal Prof. Dr. Bange danken möchte. Besonderer Dank gilt hier Dr. Patrick Pausch, der nicht nur für die Kristall-Strukturen in dieser Arbeit verantwortlich ist, sondern mir auch stets mit gutem Rat und für wissenschaftliche Diskussionen zur Verfügung stand. Dank gilt ebenso seinem Nachfolger in diesem Projekt, Nils Mais, für Unterstützung und Rat nach Patricks Abschluss. Dr. Thomas Heimerl ist für sämtliche elektronenmikroskopische Aufnahmen und den daraus resultierenden Modellen zu danken, aber auch für die freundliche Beratung. Ohne diese Kooperation wäre diese Arbeit, in dieser Form nicht möglich gewesen.

Bei meiner Graduiertenschule IMPRS bedanke ich mich für diverse Förderung und für die Leitung durch die Promotion im Allgemeinen.

Zudem möchte ich mich bei der gesamten Arbeitsgruppe Randau für die gute Atmosphäre und die stetige Unterstützung bedanken. Dies gilt sowohl für alle derzeitigen als auch ehemaligen Mitglieder. Wichtig sind hierbei auch die ehemaligen Studenten, die ich betreut habe: Shivabalan Arun Prabha, Eva Grümpel und Hamrithaa Shanmuganathan. Sie haben einen wichtigen Beitrag für diese Arbeit geleistet und mich die Betreuung von Studenten gelehrt. Danken möchte ich aber auch meinem ehemaligen Betreuer Dr. Srivatsa Dwarakanath, der die Untersuchung des Typ I-Fv Systems etabliert hat.

Zuletzt möchte ich noch all meinen Freunden, meiner Familie und meiner Verlobten Iryna Salii danken, mit deren Unterstützung ich auch schwierige Zeiten überstehen konnte und ohne die diese Arbeit nicht möglich gewesen wäre.

Curriculum Vitae

Personal Information

Name: Daniel Gleditzsch
Date of Birth: 02.09.1990 in Halle (Saale), Germany

Education

02/2016 -11/2019 **PhD** at the Max-Planck-Institute for Terrestrial Microbiology, Marburg
Laboratory of Prokaryotic Small RNA Biology

- Protein purification and bioanalytics
- CRISPR/Cas biology
- RNA extraction and analysis
- Next-Generation Sequencing (Illumina & Nanopore)
- Operation, maintenance and validation of FPLC (Äkta)
- Working with radioactive Isotopes

Dissertation: "Production and analysis of synthetic Cascade variants"

2013 - 2015 **Master of Science in Technical Biology**, Technische Universität Darmstadt

Specialization:

- Technical Genetics, Radiation Biophysics, Cellular Biophysics, Microbiology

Master-Thesis:

- "Characterization of a minimal Type I CRISPR-Cas interference complex", at the Max-Planck-Institute for Terrestrial Microbiology, Marburg

2010 - 2013 **Bachelor of Science in Biology**, Technische Universität Darmstadt

Specialization:

- Microbiology, Genetics, Plant Biotechnology, Physiology

Bachelor-Thesis:

- "Expression of Genes for Sulfur Oxygenases (ETHE1) in *E. coli*"