VILNIUS UNIVERSITY

LIFE SCIENCES CENTER

EMBL PARTNERSHIP

INSTITUTE

2nd year master's student of molecular biotechnology study program

Daria PONOMARIOVA

Establishment of type IV CRISPR-Cas for genome editing in human cells

Master thesis

Supervisor

Dr. Patrick PAUSCH

Reviewer

Dr. Stephen Knox Jones Jr

Vilnius, 2023

CONTENTS

CONTENTS 2
ABBREVIATIONS 4
INTRODUCTION 5
1. LITERATURE REVIEW. 8
1.1 CRISPR-Cas system overview 8
1.2 Class 1 and 2 CRISPR-Cas systems. 9
1.3 Type IV CRISPR-Cas systems 11
1.4 CRISPR-Cas mediated genome editing. 13
1.5 Next generation genome editing 14
1.6 Summary. 16
2. MATERIALS AND METHODS 17
2.1 Materials 17
2.1.1. Equipment 17
2.1.2 Reagents 18
2.1.3 Enzymes and buffers 19
2.1.4 Stains and Ladders 20
2.1.5 Commercial kits 20
2.1.6 Plasmids 21
2.1.7 Bacterial strains 22
2.1.8 Media 23
2.1.9 Solutions 24
2.1.10 Software 26
2.2 Methods 26
2.2.1 Design of plasmids 26
2.2.2 Oligonucleotide annealing 28
2.2.3 PCRs 28
2.2.4 Colony PCRs 31
2.2.5 Removal of template DNA 32
2.2.6 DNA gel electrophoresis 33
2.2.7 Plasmid assembly 33
2.2.8 Golden Gate plasmid assembly 33
2.2.8 Plasmid DNA purification 36
2.2.9 Plasmid DNA purification without a kit 36

2.2.10 Restriction analysis 37
2.2.11 DNA sequencing 37
2.2.12 Bacterial glycerol stocks 37
2.2.13 Electro-competent *E. coli* cells 37
2.2.14 Calcium competent *E. coli* cells 38
2.2.15 Transformation of electrocompetent *E. coli* cells 38
2.2.16 Transformation of calcium competent *E. coli* cells 39
2.2.17 Efficiency of transformation assay 39
2.2.18 Statistical analysis 39
3. RESULTS 40
3.1 Construction of a type IV eukaryotic expression vector 40
3.2 Construction of compact type IV CRISPR-Cas system 44
3.2.1 Construction of plasmids for prokaryotic expression of Cas6e-deficient type IV systems 44
3.2.2 Construction of plasmids for prokaryotic expression of Cas6e and DinG -deficient type IV systems 45
3.2.3 Construction of a guide plasmid carrying a Hammerhead ribozyme sequence 46
3.3 Assay of the activity of engineered type IV CRISPR-Cas systems 47
3.3.1 Design of gRNAs for the efficiency of transformation assay 47
3.3.2 Analysis of the activity of CRISPR-Cas type IV system 48
DISCUSSION 50
CONCLUSIONS 53
SUMMARY 54
LITERATURE 56
ACKNOWLEDGMENTS 61

ABBREVIATIONS

AAV – Adeno-associated virus

Cas – CRISPR-associated (enzyme)

Cascade – CRISPR-associated complex for antiviral defense

CMV – Cytomegalovirus

CRISPR – Clustered regularly interspaced short palindromic repeats

CRISPRa – CRISPR-mediated gene activation

CRISPRi – CRISPR interference

crRNA – CRISPR RNA

DSB – Double-strand break

dsDNA – Double-stranded DNA

gRNA – Guide RNA

HDR – Homology-directed repair

HHR – Hammerhead ribozyme

MGE – Mobile genetic element

NGS – Next generation sequencing

NHEJ – Non-homologous end joining

NLS – Nuclear localization sequence

PAM – Protospacer-adjacent motif

PCR – Polymerase chain reaction

RNAi – RNA interference

RNP – Ribonucleoprotein

sgRNA – Single-guide RNA

ssDNA – Single-stranded DNA

INTRODUCTION

CRISPR-Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated protein) is a prokaryotic nucleoprotein system that facilitates adaptive immunity against mobile genetic elements such as bacteriophages and plasmids. The system includes a CRISPR array and CRISPR-associated genes encoding for Cas effector proteins. The immune process can be divided into three stages: spacer acquisition, crRNA biogenesis, and interference [1].

During spacer acquisition, DNA fragments, derived from mobile genetic elements, are cut into fragments and inserted into the CRISPR array as spacers. These spacers, located between short palindromic repeats, form a record of previous infections. Fragments located upstream from a protospacer adjacent motif (PAM) are chosen as spacers. This strategy helps to prevent self-targeting by the CRISPR-Cas system [4; 5]. During the second stage, CRISPR array is transcribed into precursor crRNA. Precursor crRNA is then processed into mature guiding RNA [6]. In the final stage guiding RNA assembles with Cas effector proteins to form a surveillance complex that can recognize and degrade foreign DNA or RNA sequences [7].

Two major classes can be identified in CRISPR-Cas systems, class 1 and class 2. Class 1 has a multi-subunit RNA-guided nucleoprotein effector complex and includes three types, Type I, III, and IV, defined based on signature genes and locus organization. Class 1 CRISPR-Cas systems rely on an effector nuclease and/or helicase for interference. Class 2 has a single, multi-functional effector protein, such as commonly used for gene editing Cas9 or Cas12, guided by RNA [9; 10].

Type IV CRISPR-Cas systems belong to class 1 and have a compact interference complex, similar in structure to type I Cascade that consists of five small proteins bound to a crRNA. The complex can bind fragments of DNA, located close to an appropriate PAM (protospacer adjacent motif) in a crRNA-dependent manner. Upon binding, an R-loop structure forms and a DinG helicase is recruited for interference. Type IV CRISPR-Cas systems are the least researched amongst the known CRISPR types [17].

CRISPR-Cas systems can be targeted to specific DNA sequences for cleavage and have been adapted for a variety of uses in gene editing [25]. Additionally, to the use of catalytically active CRISPR-Cas to cleave target genes, deactivated CRISPR-Cas complexes can regulate gene expression in systems such as CRISPRi, CRISPRa or CRISPR-mediated epigenetic editing. In CRISPR activation (CRISPRa) systems deactivated Cas effector is used to recruit transcription activators to transcription start sites to induce gene overexpression [36]. CRISPR interference

(CRISPRi) can be used to achieve targeted silencing of gene transcription by fusing the inactivated Cas effector with transcription repressor domains [37]. CRISPR-Cas systems can also be used for site-specific recruitment of chromatin-modifying epigenetic enzymes for sequence-specific epigenome editing [40].

Despite the abundance of potential applications, use of CRISPR editing in human cells is limited by the size of the system (sequence of spCas9, commonly used for gene editing, has a size of 4.2 kb), compared to the cargo capacity of AAV (a commonly used viral vector), which is <4.7 kb [46]. Since a major challenge for the widespread use of CRISPR-Cas systems in eukaryotic cells has been the delivery of these large complexes, development of minimal CRISPR-Cas editors, such as type IV, is important for future use in therapy.

Multi-subunit nature of type IV CRISPR-Cas complexes can also be used in CRISPRi, CRISPRa or CRISPR epigenetic editing, since it allows for the development of different Cas protein–effector domain fusion strategies.

In general, applying type IV CRISPR-Cas for use in eukaryotic cells has potential to advance genome editing. So, the goal of this study is to establish the functioning of CRISPR-Cas type IV system in eukaryotic cells to be used for therapeutic purposes, development of a minimal CRISPR-Cas system for use with low-cargo capacity vectors, as well as the development of type IV CRISPR editors for regulation of gene expression.

Research Objectives:

Establishment of type IV CRISPR-Cas system for gene editing in human cells.

Research tasks:

1. Establish the functioning of a type IV CRISPR-Cas system in eukaryotic cells.
2. Develop a hypercompact CRISPR-Cas type IV system.
3. Develop type IV CRISPR editors for epigenetic regulation of gene expression.

1. LITERATURE REVIEW.

1.1 CRISPR-Cas system overview

CRISPR-Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated proteins) is a naturally occurring nucleoprotein system that facilitates adaptive immunity of bacteria and archaea against mobile genetic elements (MGEs) such as bacteriophages and plasmids. CRISPR-Cas systems consist of a CRISPR array and CRISPR-associated genes that encode Cas effector proteins. [1]

CRISPR-Cas loci localize on the prokaryotic chromosome, as well as on some megaplasmids. [2] They consist of an AT-rich leader sequence, followed by an array of short, frequently ranging in size from 28 to 37 base pairs, palindromic repeats. The repeats are intervened by “spacers”, sequences acquired from the genomes of viruses or other MGEs during a previous infection. Spacers are usually 32 to 38 base pairs long, depending on the array. Some CRISPR repeats can form a secondary structure, such as a stem loop, judging by their dyad symmetry. [3]

The CRISPR-Cas mediated immune process is divided into three stages: spacer acquisition, crRNA biogenesis, and interference. During spacer acquisition the DNA sequences of foreign MGEs are cut into fragments and inserted as new spacers into the CRISPR array, downstream of the leader. These spacer sequences are thus a record of prior infections. [4; 5] The fragments of foreign genome chosen to be acquired as spacers, termed protospacers, tend to be located adjacent to 3-5 bp DNA sequences called PAMs (protospacer adjacent motifs). The presence of a PAM upstream or downstream of the potential crRNA-complementary region is recognized by the surveillance complex first and is integral for the beginning of the target strand-crRNA pairing during interference. This mechanism is needed to eliminate the possibility of self-targeting since PAMs are not present in the CRISPR array. [5; 6]

In the second stage, cRNA biogenesis, the CRISPR array is transcribed into a precursor crRNA, and is subsequently processed by cleavage into mature guiding crRNAs. Mature crRNAs contain the spacer sequence, complementary to one located in the foreign DNA molecule, flanked by partial CRISPR repeat sequences on one or both sides that are needed for binding to Cas proteins. [6]

The last stage of the process is the formation of the surveillance complex, assembled from one or several Cas proteins bound to a crRNA guide, giving the complex the ability to sequence-specifically recognize and degrade foreign DNA or RNA sequences. [7]

In brief, during DNA-interference, recognition of PAM initiates localized DNA unwinding and the creation of a "seed loop" intermediate needed for the target strand recognition process to begin. [7; 8] Hybridization of crRNA and the target strand displaces the nontarget strand and causes the seed loop to expand into a three-stranded R-loop structure, supported by Cas proteins. After this, the nucleic acid can be cut by the effector complex using either its intrinsic nuclease activity or a separately recruited nuclease. [8]

1.2 Class 1 and 2 CRISPR-Cas systems.

CRISPR-Cas systems can be divided into two major classes, class 1 and class 2, based on the architecture of their effector molecules. Class 1 is defined by a multi-unit effector complex, while class 2 effectors contain a single protein. [9]

CRISPR-Cas systems can additionally be classified into 6 types (from type I to type VI) and several subtypes, with types I, III, and IV belonging to class 1 (figure 1.2.1a). The distinction between the CRISPR-Cas types and subtypes is based on each type's signature genes and the organization of their respective loci [10].

Most of the CRISPR types target DNA, but some types, such as Type VI can specifically target RNA, while Type III CRISPR systems exhibit both RNA- and DNA- interference.

In type 1 surveillance complexes, multiple Cas proteins are assembled around a crRNA. The type I "Cascade" (CRISPR-associated complex for antiviral defense) interference complex has a seahorse-like architecture and consists of subunits Cas8, Cas5, Cas7, Cas11, Cas6, and multiple Cas7 subunits. Cas7 assembles along the crRNA spacer forming the helical "backbone" of the structure, Cas11 forming a "belly" filament, single copies of Cas8 (large subunit) and Cas5 positioned at the 5'end of the crRNA at the "foot" of the complex, and a Cas6 subunit capping the backbone of the complex by binding to the 3' end of the crRNA (figure 1.2.1b). [10; 11] Cas6 endonuclease is needed by most class 1 systems to process crRNA by cutting a pre-crRNA transcript to form individual crRNA molecules that consist of a 5′-tag derived from a palindromic repeat, a spacer region, and a 3′-stem loop hairpin. [13]

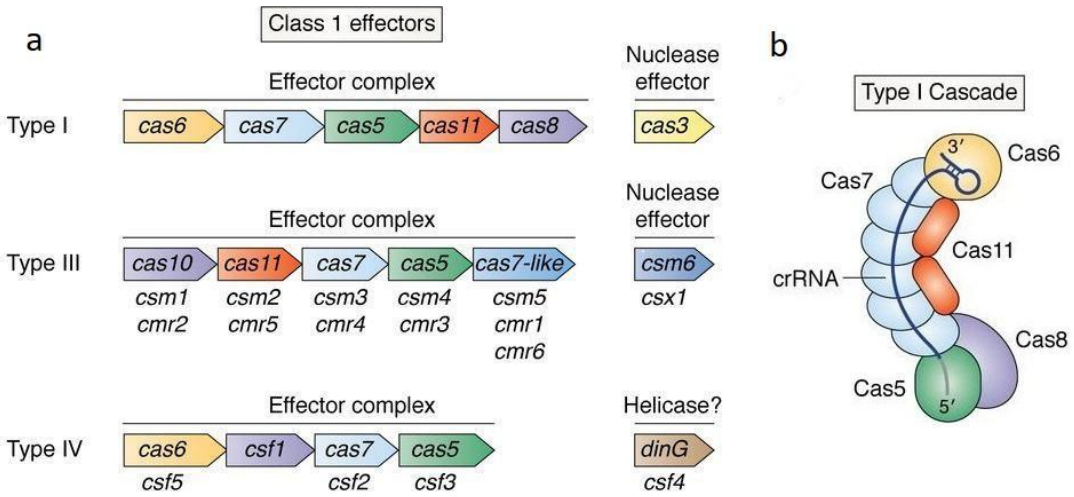

Figure 1.2.1 *Class 1 CRISPR-Cas effector complexes. a. Effector proteins characteristic of three known class 1 CRISPR-Cas types. Subunits that are analogous between the different types are shown with the same color. b. Type I crRNA-guided effector complex architecture. (Figure adapted from Liu, Doudna, 2020)*

A type III surveillance complex is characterized by architecture similar to that of a type I complex, but with a more extended shape. Cas10 protein replaces Cas8 as the large subunit of the "foot" of the complex in type III systems. The 3′ end of the crRNA in type III systems is bound by a specialized Cas7-like subunit. Type III surveillance complexes are also distinct in their lack of a Cas6 protein. [14]

Type IV CRISPR-Cas complexes include three subtypes, from IV-A to IV- E, and have a large subunit, Csf1 (Cas8), as well as subunits homologous to Cas7, Cas6, and Cas5. (figure 1.2.2)

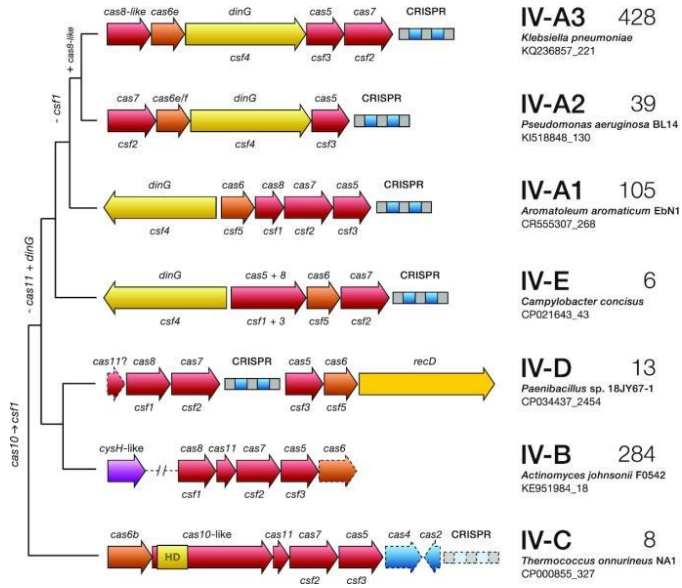

Figure 1.2.2 *Type IV CRISPR-Cas systems, organized according to genome loci, effector complex architectures and evolutionary relationships. Typical organization of operons of type IV system subtypes. Genes are colour-coded and labeled according to the protein families they encode. Cas nomenclature is represented at the top of each gene, Csf nomenclature is represented at the bottom. (R. Pinilla-Redondo, 2019)*

Class 1 CRISPR-Cas systems require the presence of an effector nuclease and/or helicase that works with the surveillance complex to cleave target DNA or RNA. Most type I Cascade as well as type IV effector complexes have no intrinsic enzymatic activity, which causes reliance on recruitment of helicase-nuclease Cas3 to degrade dsDNA in the case of type I system, and a DNAse-free DinG family helicase Csf4, in the case of a type IV system. [10, 12] Type III CRISPR-Cas Csm and Cmr possess intrinsic DNase and RNase activities but still utilize an accompanying nuclease effector (Csm6/Csx1), the activity of which is stimulated by a secondary messenger, cOA. [14]

In contrast to class 1 systems, class II systems utilize a single multi-functional protein, such as Cas9 or Cas12, that can target and degrade DNA, guided either by an RNA heteroduplex consisting of a crRNA and a trans-activating crRNA (tracrRNA), or a single guide RNA. [15]

1.3 Type IV CRISPR-Cas systems.

Type IV CRISPR-Cas is a subclass of class 1 systems and, therefore, shares several key characteristics of class 1, such as a multi-subunit interference complex with a seahorse-like architecture. [16] Type IV CRISPR-Cas systems can be divided into several subtypes: IV-A (1; 2; 3), IV-B, IV-C, IV-D and IV-E. Type IV CRISPR-Cas is the least researched of the currently known CRISPR types. [16; 17]

The interference complex of the system is compact and consists of 3–5 small proteins and a crRNA. The arrangement of genes observed in a type IV-A CRISPR locus tends to follow the csf4 (dinG), csf5 (cas6-Like), csf1 (cas8-Like), csf2 (cas7), csf3 (cas5) scheme [18] The loci also contain a conserved leader sequence succeeded by a CRISPR array (figure 1.3).

Type IV systems tend to lack proteins needed for acquisition of spacers from foreign DNA, i.e., Cas1 and Cas2, but some still contain adjacent CRISPR arrays. These are mostly located close to subtype IV-A systems. In the absence of adaptation genes, type IV-A systems can co-localize with some type I systems and utilize these system's adaptation proteins for the incorporation of new spacers into their CRISPR-array. [18]

Expression of different variants of a pre-crRNA processing endonuclease has been observed in different type IV systems. These endonucleases are denoted Cas6, Cas6e, and Csf5. Histidine and tyrosine active site residues perform the cleavage in Cas6 and Cas6e enzymes, while the Csf5 active site employs arginine residues [18; 19]. Loci of subtype IV-A systems encode for a Csf5 crRNA endonuclease which performs cleavage of pre-crRNA repeats at the 3′ side of a predicted stem loop, ensuring crRNA maturation during the biogenesis stage. At the same time subtype IV-B systems lack these proteins [20; 21], it is unclear what crRNAs are used by IV-B systems and how they function for nucleic acid interference. Some type IV systems utilize a Cas6 homolog from a different system. E.g., type I-E. In type IV-A systems, the mature crRNA retains the 3′ hairpin structure.

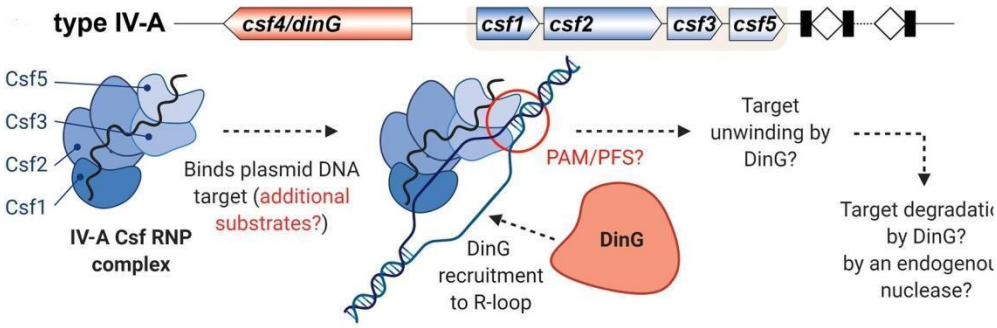

Figure 1.3.1 *Model of type IV ribonucleic complex structure. Type IV-A CRISPR-Cas locus consists of genes encoding for Csf1, Csf2, Csf3, Csf5 and DinG, as well as a CRISPR array. The effector complex consists of Csf1, Csf2, Csf3 and Csf5, DinG helicase is recruited separately. IV-A RNP complexes likely bind DNA targets and recruit DinG for target unwinding and degradation. (Figure adapted from Taylor, 2021)*

For interference, type IV systems form a crRNA-guided effector complex similar to type I Cascade, composed of Csf1, Csf3, Csf5, and multiple copies of Csf2. In the complex Csf2 proteins form a helical scaffold around the crRNA. Meanwhile Csf1 and Csf5 are located close to the 3′-end of the crRNA [22] (figure 1.3.1)

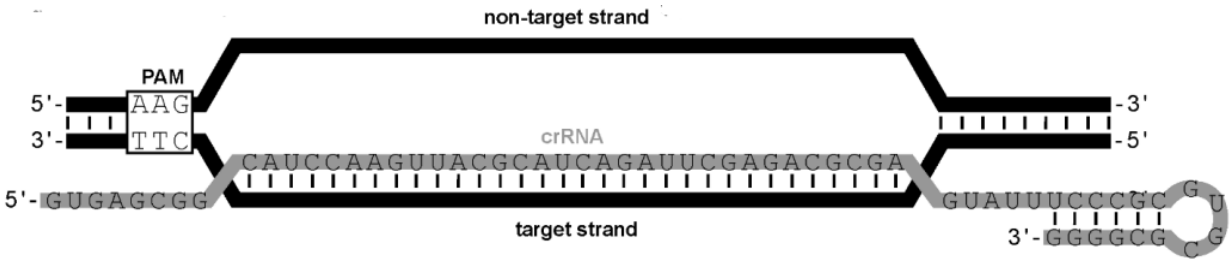

Figure 1.3.2 *Schematic representation of a Type IV-A R-loop. Type IV-A crRNA paired with the complementary DNA strand upstream of a PAM, non-target strand is displaced. (Figure adapted from Guo, 2022)*

The complex functions by binding fragments of nucleic acid sequences complementary to the guiding crRNA, which causes a formation of an R loop (figure 1.3.2). A DinG helicase that is not part of the complex is later recruited to the R loop for interference an appropriate PAM is needed to begin the DNA recognition process. IV-A protospacers were observed to be flanked with a 5′-TTC-3′ PAM similar to I-E systems. [23]

1.4 CRISPR-Cas mediated genome editing.

One of the core characteristics of CRISPR-Cas systems is their ability to recognize a specific DNA/RNA sequence in the target genome or RNome in a guide-RNA programmable manner. This feature allowed for efficient adaptation of CRISPR-Cas for gene editing.

Gene editing is a developing set of techniques, targeted towards changing the genetic makeup of cells and organisms for research and therapeutic applications [24]

The CRISPR-Cas system most frequently used for genome editing belongs to class 2 type II CRISPR systems and is composed of two components: Cas9 nuclease and an artificial single guide RNA, a fusion of a (crRNA) and (tracrRNA). [25] The most widely used variant of Cas9 is derived from Streptococcus pyogenes., Upon PAM recognition by the spCas9-sgRNA complex, the target DNA strand located upstream of the PAM is paired with the spacer portion of the sgRNA. The target region is then cleaved by the Cas9 nuclease, creating a blunt-end double-stranded break. Class 2 CRISPR-Cas systems have been mostly used in genome editing due to their precision, simplicity and the small number of components required for DNA targeting. [26] This contrasts class 1 systems, which are characterized by effector complexes consisting of multiple proteins that usually introduce large genomic deletions in gene editing applications [27].

Several strategies for gene editing using CRISPR-Cas exist. The most common use for the CRISPR-Cas systems is gene knockout. In this strategy, the system, equipped with a sgRNA sequence matching that of a gene that needs to be inactivated is delivered into the cell. SgRNA then guides the Cas effector to the target DNA, after target strand DNA-sgRNA complex formation Cas nuclease will form a double-stranded break in the protospacer region. [28, 29, 30]

The break in the DNA can be repaired through either non-homologous end joining (NHEJ) or homology-directed repair (HDR). Activation of the NHEJ pathway happens in the absence of a template and causes insertions or deletions, disruptive to the targeted loci. Targeted insertions

into the genome are possible with the use of a homology-directed repair pathway. If a donor DNA fragment, homologous to the targeted locus, is present after the Cas protein created a double-stranded break the HDR pathway will be activated, which makes it possible to insert a new sequence. [31]

1.5 Next generation genome editing

Beyond conventional gene engineering, CRISPR-Cas systems can be applied for control of gene expression, utilizing the ability of the complex to bind specific, predetermined regions of the genome, but not its catalytic activity. The deactivation of the functional domain of a Cas protein or removal of the putative helicase from a class 1 complex does not affect the ability of the system to bind DNA in the target site, which makes the use of inactive CRISPR-Cas in next generation genome editing possible. [31]

Posttranslational DNA modifications can help guide the activity of the cell by regulating rates of expression of certain genes. [32] The use of novel epigenetic editing techniques based on CRISPR-Cas systems can help program gene expression in the cell and manipulate the architecture of native chromatin. [33, 36] Epigenome editing can be used for the activation and repression of certain genes as it offers a broad spectrum of opportunities for graded control of gene expression. CRISPR-Cas systems can be used to recruit effector molecules to a specific site, allowing for control over modifications to DNA, histones, and chromatin architecture. [34]

DNA binding platforms able to perform chromatin reorganizations and to alter the epigenome at a specific sequence by recruiting epigenetic effectors have already been created based on deactivated class 2 CRISPR-Cas systems. [35]

CRISPR systems can be used for CRISPR activation (CRISPRa) systems, in which sgRNAs and deactivated, or dead (dCas9), are used to recruit transcription activators to transcription start sites (TSS) to induce gene overexpression. Multiple transcriptional activators can be recruited to a TTS with a single sgRNA to induce higher levels of overexpression. Several strategies were developed to achieve this, activator domains can be fused to dCas9, recruited to a protein scaffold fused to dCas9, or recruited to an RNA scaffold fused to the sgRNA. [36]

Targeted silencing of gene transcription can be achieved with CRISPR interference (CRISPRi). Like in CRISPRa, the system consists of a dCas9 protein and an sgRNA. RNA polymerase-mediated transcript elongation can be stopped when the Cas9-sgRNA complex binds to the DNA region complementary to the sgRNA, creating a steric hindrance for the polymerase and causing repression of the target gene. To improve the efficiency of the system in cases where the dCas9-

sgRNA complex is not sufficient to fully block the RNA polymerase dCas9 can be fused with transcription repressor domains. [37]

CRISPR-Cas systems can also be used for site-specific recruitment of chromatin-modifying epigenetic enzymes for sequence-specific epigenome editing. [40]

Although the use of class 1 CRISPR-Cas systems is less widespread in gene editing than the use of class 2 proteins such as Cas 9, they have the potential to be used for development of next generation editors, in part, due to their multi-subunit effector complexes that do not possess intrinsic catalytic activity. Targeting a class 1 CRISPR-Cas complex to a DNA sequence, in the absence of Cas3/DinG can block transcriptional machinery from binding to the sequence and lead to the repression of the transcription in the targeted gene without altering its nucleic acid composition.

An important characteristic of epigenetic editing is the necessity to create fusions of enzymes that take part in epigenetic regulation in vivo, and DNA-binding CRISPR-Cas proteins or complexes that can target the epigenetic effector to a specific sequence. [39, 41] If a transcription activator or repressor domain is fused to one of the proteins in the class 1 complex a more precise and efficient control of gene expression may be achieved.

Type IV CRISPR-Cas systems possess multi-subunit effector complexes, and the functionality of the type IV system may be engineered by adding regulatory domain-fusions. The system has multiple subunits for different Cas protein–epigenetic effector domain fusion strategies. By fusing activator and repressor proteins to the complex and targeting it to the promotor region of the gene upregulation or downregulation of its expression, respectively, can be possible without the need to induce double-stranded breaks in the target sequence.

A major drawback to the use of class 2 systems is their size. Cas9 variants discovered in the wild tend to be large, which can hinder the use of such editors because of the limited packaging capacity of common gene editing vectors, such as Adeno Associated viruses (AAV), needed for the delivery of the system to the cell. Packaging SpCas9 (~4.2 kb) into an AAV vector is challenging due to its packaging capacity of ~4.5 kb. [42] While this approach is feasible, it leaves little room for additional regulatory elements. Cas9 and multiple gRNAs can be packaged into separate AAV vectors, increasing overall packaging capacity but necessitating co-infection of two AAVs which limits dose and efficacy [43]. Development of smaller CRISPR-Cas systems can have great therapeutic potential.

1.6 Summary.

In conclusion, CRISPR-Cas is an adaptive bacterial immunity system that can bind to and destroy nucleic acids in a guide-RNA dependent manner. In the wild, the system consists of a CRISPR array and Cas proteins needed for spacer acquisition, pre-crRNA processing and effector complex formation. As a genome editing tool it is represented by a complex of guide RNA and a Cas effector that targets the system to the DNA fragment of interest.

CRISPR-Cas systems include two classes, class 1, which is characterized by a multi-subunit effector complex, and class 2, with a single effector protein. Most CRISPR editors currently in use are based around class 2 systems. Type IV4 CRISPR-Cas systems belong to class 1 and share several features with other class 1 systems, such as a multi subunit, crescent-shaped effector complex. Out of all the currently known CRISPR-Cas systems, type IV systems are the least researched, not all type IV mechanisms of action and biological functions are known [17].

CRISPR-Cas systems can be targeted to specific DNA fragments for cleavage., CRISPR-Cas can also be used to introduce new genes into the genome in combination with an HDR mechanism. Deactivated CRISPR-Cas complexes can regulate gene expression in systems such as CRISPRi, CRISPRa or CRISPR-mediated epigenetic editing.

Potential use of CRISPR editing in human cells is severely limited, in part by the decreased cargo capacity of AAV (a commonly used viral vector)., This problem may be solved by development of compact class 1 editors, such as type IV-A since the DNA-targeting components could be reduced to about 2.5 - 3 kB.

The development of minimal CRISPR-Cas editing systems is especially important for future use in therapy since a major challenge for the widespread use of these systems in eukaryotic cells has been the delivery of these large complexes into the cell. Moreover the multisubunit nature of Csf2/Cas7 in class 1 CRISPR-Cas complexes allows recruiting several effector domains, encoded on a short stretch of DNA.

The goal of this study is to improve the understanding of type IV CRISPR-Cas system, establish the functioning of this system in eukaryotic cells, development of a minimal CRISPR-Cas system as well as the development of type IV CRISPR editors for epigenetic regulation of gene expression.

2. MATERIALS AND METHODS

2.1 Materials

2.1.1. Equipment

Table 2.1.1.1: Equipment used in the study

Device	**Manufacturer**
Analytical balances	Fisherbrand Analytical Balances
Benchtop centrifuge	Fisherbrand GT4 Benchtop Centrifuge
Benchtop pH meter	Fisherbrand™ accumet™ AE150 Benchtop pH Meter
Biological safety cabinet	Airstream Gen 3 Class II Type A2 Biological Safety Cabinet
Block heater	Thermo Scientific block heater
Electrophoresis system	multiSUB Mini, Mini Horizontal Electrophoresis System
Electroporator	Bio-Rad MicroPulser Electroporator
Fumehood	SCALA
Gel documentation system	GelDoc Go Imaging System Gel documentation system
Magnetic stirrer	Fisherbrand Magnetic Stirrer
Incubating shaker	New Brunswick Innova 4000 Incubating Shaker
	INCUCELL laboratory incubator Incubating shaker
Microwave oven	Severin Microwave oven 20 liters 700 watts
PCR thermocycler	ProFlex™ 3 x 32-well PCR System thermocycler
Pipette filler	Thermo Scientific S1 Pipette Filler
Power source for electrophoresis	Invitrogen PowerEase Touch 120W Power Supply DNA electrophoresis system
Refrigerated centrifuge	Fisher Scientific accuSpin Micro 17R Refrigerated Centrifuge
Spectrophotometer	NanoDrop ND-1000 Spectrophotometer
	Fisherbrand Cell Density Meter

	Spectrophotometer
Variable volume pipettes	Finnpipette F2 Variable Volume Pipettes
Vortex mixer	Fisherbrand Mini Vortex Mixer
Water purification system	Millipore Milli-Q Direct Water Purification System

2.1.2 Reagents

Table 2.1.2.1: Reagents used in the study

Chemical.	**Manufacturer**	**Article**
Acetic Acid (USP, BP, Ph. Eur.) pure, pharma grade	Roth	6755.1
Agar, for molecular genetics, granulated	ThermoFisher Scientific	BP9744500
Agarose broad separation range for DNA/RNA, DNase- and RNase-Free	Fisher bioreagents	BP1356-500
Ampicillin Sodium Salt (Crystalline Powder)	ThermoFisher Scientific	BP1760-5
Calcium chloride dihydrate	Roth	5239.1
Carbenicillin disodium salt (White Powder)	Roth	6344.4
Chloramphenicol	Roth	3886.2
D(+)-Glucose anhydrous	Panreac	1B014276
Ethylenediaminetetraacetic acid (EDTA)	Panreac	A1103,1000
Glycerol, for Molecular Biology	Fisher bioreagents	BP229-4
Hydrochloric acid, ACS reagent	Sigma-Aldrich	320331
Isopropyl ß-D-1-thiogalactopyranoside (IPTG)	Formedium	R036
Kanamycin	Roth	T832.5

Magnesium chloride hexahydrate, 99%	Acros organics	197530010
Ethanol, 99%,	Honeywell	32221
Potassium Acetate	Roth	HN10.3
Potassium Chloride	Fisher Chemical	P/4240/60
Sodium dodecyl sulfate (SDS) pellets	Roth	CN30.2
Sodium chloride	ThermoFisher Scientific	J21618.A1
Sodium hydroxide	Honeywell	10094491
Super Optimal broth with Catabolite repression (SOC) medium	MP Biomedicals	11347689
TRIS base	Fisher bioreagents	BP152-1
Tryptone	Roth	8952.1
Yeast Extract	Roth	2904.3

2.1.3 Enzymes and buffers

Table 2.1.3.1: Enzymes and buffers used in the study

Reagent.	**Manufacturer**	**Article**
AarI (2 U/µL)	ThermoFisher Scientific	ER1581
50X Oligonucleotides for AarI, BvaI	ThermoFisher Scientific	ER1581
10X Buffer for AarI	ThermoFisher Scientific	ER1581
dNTP set	ThermoFisher Scientific	R0181
FastDigest DpnI	ThermoFisher Scientific	FD1703
FastDigest BsaI/Eco31I	ThermoFisher Scientific	ER0291

FastDigest Value Pack	ThermoFisher Scientific	K1991
FastAP thermosensitive alkaline phosphatase	ThermoFisher Scientific	EF0654
Phusion Plus DNA Polymerase	ThermoFisher Scientific	F630XL
Phusion Plus DNA Polymerase buffer	ThermoFisher Scientific	F630XL
Phusion Plus high GC enhancer	ThermoFisher Scientific	F630XL
Phusion DNA Polymerase	ThermoFisher Scientific	F530XL
Phusion DNA Polymerase buffer	ThermoFisher Scientific	F530XL
Q5 DNA polymerase	NEB	M0491S
T4 DNA Ligase	ThermoFisher Scientific	EL0011
T4 Polynucleotide Kinase	ThermoFisher Scientific	EK0031

2.1.4 Stains and Ladders

Table 2.1.4.1: Stains and ladders used in the study

Reagent	**Manufacturer**	**Article**
Ethidium bromide	Roth	HP47.1
GeneRuler DNA Ladder Mix	ThermoFisher Scientific	SM0331
GeneRuler 1-kb DNA ladder	ThermoFisher Scientific	SM0311
GeneRuler 100-bp DNA ladder	ThermoFisher Scientific	SM0241
SYBR® Green DNA dye, 9.0 ml, 5 x 1.8 ml	Roth	1CN2.2
SYBR™ Gold Nucleic Acid Gel Stain, 500 µL	Invitrogen	S11494

2.1.5 Commercial kits

Table 2.1.5.1: Commercial kits used in the study

Kit	**Manufacturer**	**Catalog number**
GeneJET Gel Extraction Kit	ThermoFisher Scientific	K0691
GeneJET PCR Purification Kit	ThermoFisher Scientific	K0702
GeneJET Plasmid Miniprep Kit	ThermoFisher Scientific	K0502

2.1.6 Plasmids

Table 2.1.7.1: Plasmids used in the study

Plasmid	Description	Manufacturer/supplier/reference source
pACYCDuet-1	Plasmid carrying a chloramphenicol resistance gene, used as a target during the efficiency of transformation assay.	Obtained from the plasmid collection of Dr Patrick Pausch.
pX330-U6-Chimeric_BB-CBh-hSpCas9	High-copy plasmid for mammalian expression, used for the construction of the eukaryotic type IV expression vector. It is a human codon optimized SpCas9 and chimeric crRNA guide expression plasmid. Plasmid carries an ampicillin resistance gene, a U6 promoter, a CMV enhancer sequence and a gRNA scaffold.	Provided by the laboratory of Dr Urtė Neniškytė.
pYTK095	High copy plasmid, used as a	Obtained from the plasmid

	backbone to construct crRNA guide encoding vectors targeting the chloramphenicol resistance gene. Plasmid carries resistance to ampicillin, a ColE1 origin and an sfGFP dropout	collection of Dr Patrick Pausch.
pRSF-Duet-1	Bacterial expression vector that carries the RSF1030 replicon, lacI gene and kanamycin resistance gene. Used as a backbone for construction of the type IV system prokaryotic expression vector.	Obtained from the plasmid collection of Dr Patrick Pausch.

2.1.7 Bacterial strains

2.1.7.1 Mach1

Mach1 (str. W ΔrecA1398 endA1 fhuA Φ80Δ(lac)M15 Δ(lac)X74 hsdR(rK–mK+)) chemically competent *E. coli* cells used as cloning hosts for prokaryotic expression plasmids. Amongst the most frequently used *E. coli* strains for routine cloning applications. Mach1 cells are resistant to T1 phage and are one of the fastest-growing chemically competent strains. Obtained from ThermoFisher Scientific.

2.1.7.2 DH5α

DH5α (F– endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK–mK+), λ–) chemically competent *E. coli* cells were used as cloning hosts for prokaryotic expression plasmids. DH5α cells are endonuclease A deficient and are frequently used for routine cloning applications, especially for cloning large plasmids. Insert stability is increased due to RecA1 and endA1 mutations. The mutations also improve the quality of plasmid DNA prepared from minipreps. Obtained from the Department of Protein - DNA Interactions, VU LSC.

2.1.7.3 TOP10

TOP10 (F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1 λ-) electrocompetent *E. coli* cells were also used as cloning hosts for prokaryotic expression plasmids. TOP10 cells have hsdR for transformation of unmethylated DNA from PCR amplifications. Obtained from the Laboratory of Dr Stephen Knox Jones.

2.1.7.4 NEB Stable

NEB Stable endonuclease I free electrocompetent *E. coli* cells (F' proA+B+ lacIq Δ(lacZ)M15 zzf::Tn10 (TetR)/Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- Φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC)) were used for long-term storage of the type IV eukaryotic expression vector. NEB Stable enables the isolation of plasmid clones containing repeat elements and unstable inserts. Obtained from New England Biolabs (NEB).

2.1.7.5 BL21 Star (DE3)

BL21 Star (DE3) (B F– ompT gal dcm lon hsdSB(rB–mB–) λ (DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB+] K-12(λS)) *E. coli* cells were used for the expression of type IV system proteins during the plasmid curing assay. BL21 Star (DE3) competent cells are suitable for the expression of non-toxic heterologous genes. The strain contains the lambda DE3 prophage that carries the gene for T7 RNA polymerase under control of a lacUV5 promoter, allowing expression of the T7 RNA polymerase to be induced with IPTG. Obtained from the Department of Protein - DNA Interactions.

2.1.8 Media

2.1.8.1 LB liquid media

10 g of tryptone, 5 g of yeast extract 10 g of NaCl were dissolved in 1 L of MilliQ water and autoclaved at 121 °C, for 20 minutes. Autoclaved LB liquid media was stored at room temperature.

2.1.8.2 LB solid media.

10 g of tryptone, 5 g of yeast extract 10 g of NaCl were dissolved in 1 L of MilliQ water, 16 g of Agar was added, the mixture was then autoclaved (121 °C, 20 min) and stored at room temperature.

2.1.8.3 SOC recovery media.

31 g of SOC powder were dissolved in 1 L of autoclaved Milli Q water, pH was adjusted to 6.99-7 with 1 N NaOH. SOC recovery media was filter-sterilized with a 0.22 µm PVDF filter and stored at 4 °C.

2.1.9 Solutions

2.1.9.1 Ampicillin stock solution (100 mg/mL).

1 g of ampicillin sodium salt was dissolved in 10 mL of MilliQ and filter-sterilized with a 0.22 µm PVDF filter. The solution was aliquoted into 1.5 -mL microcentrifuge tubes and stored at -20 °C.

2.1.9.2 $CaCl_2$ (100 mM).

14.702 g of $CaCl_2$ dihydrate were dissolved in 1 L of MilliQ water. The solution was autoclaved at 121°C, for 20 minutes and kept at 4 °C.

2.1.9.3 $CaCl_2$ (85 mM), 15% glycerol.

12.49 g of $CaCl_2$ dihydrate were dissolved in 700 mL of MilliQ water, 300 mL of 50% glycerol were added to the solution. The solution was autoclaved at 121 °C, for 20 minutes and kept at 4 °C.

2.1.9.4 Chloramphenicol stock solution (20 mg/mL).

0.2 g of chloramphenicol were dissolved in 10 mL of 100 % EtOH and filter-sterilized with a 0.22 µm PVDF filter. The solution was aliquoted into 1.5 -mL microcentrifuge tubes and stored at -20 °C.

2.1.9.5 Denaturing solution for kit-free isolation of plasmid DNA from bacterial culture (0.2 N NaOH, 1.0 % SDS).

8 g of NaOH and 0.576 g of SDS were dissolved in 200 mL MilliQ water. The solution was stored at room temperature.

2.1.9.6 Glycerol (10%).

100 mL of glycerol were dissolved in 900 mL of MilliQ water. The solution was autoclaved at 121 °C for 20 minutes. Glycerol solution was stored at 4 °C.

2.1.9.7 Glycerol (50%).

500 mL of glycerol were dissolved in 500 mL of MilliQ water. Glycerol solution was stored at room temperature.

2.1.9.8 IPTG (0.05 M).

0.11 g of IPTG were dissolved in 10 mL of MilliQ water and filter-sterilized with a 0.22 µm PVDF filter. The solution was aliquoted into 1.5-mL microcentrifuge tubes and stored at -20 °C.

2.1.9.9 Kanamycin stock solution (50 mg/mL).

0.5 g of kanamycin were dissolved in 10 mL of MilliQ water and filter-sterilized with a 0.22 µm PVDF filter. The solution was aliquoted into 1.5 -mL microcentrifuge tubes and stored at -20 °C.

2.1.9.10 MgCl_2 (100 mM).

20.33 g of MgCl_2 hexahydrate were dissolved in 1 L of MilliQ water. The solution was autoclaved at 121 °C, for 20 minutes and kept at 4 °C.

2.1.9.11 NaOH (1 N).

40 g of NaOH were dissolved in 1 L MilliQ water. The solution was stored at room temperature.

2.1.9.12 Renaturing solution for kit-free isolation of plasmid DNA from bacterial culture (3 M potassium acetate, 2 M acetic acid).

120 mL of 5M Potassium acetate and 23 mL of glacial acetic acid were dissolved in 57 mL of MilliQ water. The solution was stored at 4 °C

2.1.9.13 Resuspension buffer for kit-free isolation of plasmid DNA from bacterial culture (25 mM TRIS-HCl (pH 8), 50 mM glucose, 10 mM EDTA).

0.6 g of TRIS were dissolved in 200 mL MilliQ water and pH was adjusted until it reached 8, 1.8 g of glucose and 0.58 g of EDTA were then added. The solution was stored at 4 °C.

2.1.9.14 TAE 50x electrophoresis buffer (2 M TRIS base, 1 M acetic acid, 0.05 M EDTA).

242 g of TRIS base, 100 ml of 0.5 M EDTA and 57.1 ml of glacial acetic acid were dissolved in 1 L of MiliQ water. The solution was stored at room temperature.

2.1.9.15 TAE 1x electrophoresis buffer (40 mM TRIS base, 20 mM acetic acid, 1 mM EDTA).

20 mL of 50x TAE buffer were dissolved in 980 mL of MiliQ water. The solution was stored at room temperature.

2.1.9.16 TE buffer (1 mM EDTA, 10 mM Tris-Cl).

0.2 mL of EDTA (0.5 M, pH 8.0) and 1 mL of Tris-Cl (1 M, pH 8.0) were added to 100 mL of MiliQ water.

2.1.10 Software

2.1.10.1 Benchling.

Benchling is a cloud-based software platform for molecular biology. Benchling was used to plan experiments, design plasmids, primers and crRNA guides, and analyze sequencing results.

2.1.10.2 GraphPad Prism version 5.01.

GraphPad Prism is a Scientific Graphing Software used for scientific graphing, comprehensive curve fitting (nonlinear regression), statistics, and data organization. GraphPad Prism was used for statistical analysis and visualization of results.

2.2 Methods

2.2.1 Design of plasmids

2.2.1.1 Design of crRNA guides targeting the pACYCDuet-1 chloramphenicol resistance gene

Benchling crRNA guide design tool was used to design crRNA guides targeting the chloramphenicol resistance gene of the pACYCDuet-1 plasmid. pACYCDuet-1 plasmid sequence was imported into Benchling and sequences downstream from AAG PAMs in the gene and the gene promoter were selected as crRNA guides. 32 bp-long crRNA guides (table 2.2.1.1) were used in the study.

Table 2.2.1.1: crRNA guide sequences

CrRNA guide	**Sequence (5' – 3')**	**Strand, complementary to crRNA guide**
CrRNA guide 1	uauccagugauuuuuuucuccauuuuagcuuc	crRNA guide 1 is complementary t the bottom strand of the targeted sequence.
CrRNA	ucgauaacucaaaaaauacgcccgguagugau	crRNA guide 2 is

guide 2		complementary to the bottom strand of the targeted sequence.
CrRNA guide 3	aucuuauuucauuauggugaaaguuggaaccu	crRNA guide 3 is complementary to the bottom strand of the targeted sequence.

2.2.1.2 Design of a eukaryotic vector for mammalian expression of type IV CRISPR-Cas system

Type IV CRISPR-Cas system genes encoding for Csf1, Csf2, Csf3, Csf4 and DinG proteins, containing C-terminal SV40 nuclear localization signals (NLS), FLAG tags and P2A sequences were designed and codon-optimized for expression in *Homo sapiens*. FLAG-tag artificial tags, with the sequence DYKDDDDK, were used to tag type IV system proteins for capture and detection. P2A sequences, 24 amino-acid-long viral oligopeptides were used to induce separationn of polypeptides into separate type IV proteins during translation in eukaryotic cells (Figure 2.2.1.2.1). Synthetic DNA sequences were ordered from Twist Biosciences.

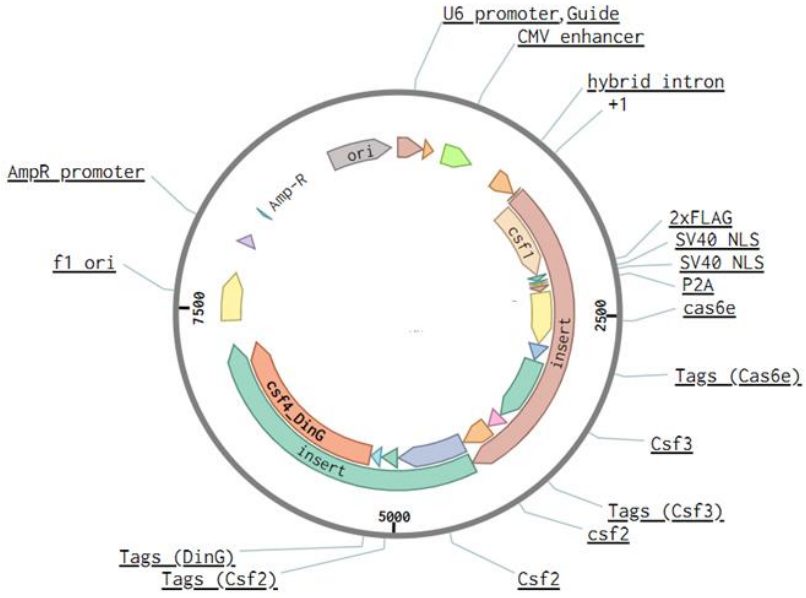

Figure 2.2.1.2.1: *Map of eukaryotic expression vector. Positions of DNA sequences encoding for Csf1, Csf2, Csf3, Cas6e and DinG proteins, as well as NLS sequences, FLAG-tags, P2A sites (referred to as Tags) are shown. U6 promoter, CMV enhancer and crRNA guide sequence upstream from the type IV genes are also shown.*

Table 2.2.1.2: CrRNA guide oligonucleotide sequences for eukaryotic expression plasmid. The crRNA guide is located downstream from the U6 promoter.

Oligonucleotide	Sequence (5’ – 3’)	Description
Oligo 1 F	gtgtttcccctgcgtgcgggggttatcggcggtcttcgg	Oligonucleotides 1 and 3 assemble with each other
Oligo 3 R	tcgatccgaagaccgccgataacccccgcacgcagggga	
Oligo 2 F	tgcacctcccgaagaagaccggtgtttcccctgcgtgcg	Oligonucleotides 2 and 4 assemble with each other
Oligo 4 R	ccgataacccccgcacgcaggggaaacaccggtcttctt	

2.2.2 Oligonucleotide annealing

Oligonucleotides were phosphorylated and annealed by incubation with T4 polynucleotide kinase for one hour at 37 °C, then heated to 95 °C for two minutes in a thermoblock (table 2.1.1.1). After that the block was removed from the heater and the reaction mixture gradually cooled down until it reached room temperature.

Table 2.2.2.1: Oligonucleotide annealing reaction mixture

Reagent	Volume
Forward oligo (100 µM)	1 µL
Reverse oligo (100 µM)	1 µL
T4 PNK	1 µL
10x T4 DNA ligase buffer	0.5 µL
MilliQ water	To 5 µL

2.2.3 PCRs

All PCR reactions were performed using Phusion Plus Polymerase in Phusion Plus Polymerase Buffer with addition of 10 mM dNTPs and a GC enhancer. Concentration of primers in the final reaction mixture was 0.5 µM. Preparation of reaction mixtures was done on ice to prevent the dimerization of primers.

Table 2.2.3.1: PCR reaction composition

Reagent	Volume
5X phusion buffer	4 µL
High GC enhancer	4 µL
dNTPs (10 mM)	0.4 µL
Forward primer x10	1 µL
Reverse primer x10	1 µL
Template DNA	1 µL

Phusion DNA polymerase	0.2 µL
MilliQ water	To 20 µL

The PCRs were performed in a thermal cycler with initial denaturation temperature set at 98 °C for 30 seconds, denaturation at 98 °C (10 seconds), primer annealing at 55 °C (10 seconds), extension at 72 °C (30 s/kb), final extension at 72 °C (5 minutes). Denaturation, annealing and extension steps were repeated for 35 cycles.

PCRs were performed to introduce type IIS enzyme recognition sites into the plasmid regions being amplified (table 2.2.3.2).

Table 2.2.3.2: Primers used in the study.*

Primer	Sequence (5' – 3')	Description	Experiment
pX330_R_BsaI	aaggcacgGTCTCcccatccatggtggcaccgg	pX330. Addition of BsaI restriction site.	PCR of the pX330 eukaryotic plasmid was carried out to introduce BsaI recognition sites for subsequent Golden Gate assembly of the eukaryotic expression vector and removal of the spCas9 sequence. Due to complexity of the eukaryotic sequence, the PCR had to be carried out in three separate reactions, with a >1.5kb fragment of the required section of the plasmid being amplified in each reaction. Primers were designed to introduce BsaI recognition sites into the fragments of pX330 plasmid.
pX330_F_BsaI	aagcacgGTCTCcgcagaggaaccccta gtgatggag	pX330. Addition of BsaI restriction site.	
pX330_crRNA guide_R	tatagGTCTCgacacggtgtttcgtcctttcc ac	pX330. Addition of BsaI restriction site.	
pX330_crRNA guide_F	tatagGTCTCgtcggttttttgttttagagcta gaaatagcaag	pX330. Addition of BsaI restriction site.	
pX330_AmpR_R	tatagGTCTCgcatcttacggatggcatga c	pX330. Addition of BsaI restriction site.	
pX330_AmpR_F	tatagGTCTCtaggccaccacttcaagaac	pX330. Addition of BsaI restriction site.	

	tc		
pYTK095_HHR_F	tataCACCTGCtatacgagaccgggatgcac	pYTK095. Addition of AarI restriction site	PCR of pYTK095 plasmid was carried out to introduce AarI recognition sequences for insertion of the hammerhead ribozyme sequence.
010-T7P-R	gggaattatgctgagtgatatcctacccCACCTGatat	pYTK095. Addition of AarI restriction site	
IVA3-dDinG_Cas6del_R	tataCACCTGCtataactgggtcatacctttttcaagaggtgc	IVA3-dDinG. Cas6e removal with addition of AarI restriction site.	PCRs were carried out with the prokaryotic expression plasmids carrying type IV genes and type IV genes without DinG helicase as templates, to obtain plasmids carrying minimal type IV system sequences that do not contain Cas6e protein as well as non-His-tagged sequences. AarI cut sites were introduced for the circularization of the PCR product.
IVA3-dDinG_Cas6del_F	tataCACCTGCtatacagttcatgttaaatttaaaccgtatcgggtc	IVA3-dDinG. Cas6e removal with addition of AarI restriction site.	
IVA3-Chis-delCas6e-R	ataCACCTGCtatacagtagaatacatcagcggtatt	IVA3-Chis. Cas6e removal with addition of AarI restriction site.	
002-Duet-MCS1-F-AarI	attaCACCTGCtgcttaagtcgaacagaaagtaatcg	IVA3-Chis. Chis removal with addition of AarI restriction site.	
005-RPCsf2-R	tataCACCTGCccatcttagcgggcagtgaaatc	IVA3-Chis. Chis removal with addition of AarI restriction site.	

* Capital letters denote sequences of introduced cut sites. The annealing temperatures of all the designed primers were within the 55-57 °C range.

The tubes were briefly centrifuged before opening. Oligonucleotides and primers were resuspended in MilliQ water to achieve concentration of 100 µM.

2.2.4 Colony PCRs

Phusion Polymerase in Phusion Polymerase Buffer with addition of 10 mM dNTPs was used for the colony PCRs. Concentration of primers in the final reaction mixture was 0.5 µM.

Figure 2.2.4.1: Colony PCR reaction composition

Reagent	Volume
5X phusion buffer	4 µL
dNTPs (10 mM)	0.4 µL
Forward primer x10	1 µL
Reverse primer x10	1 µL
Template DNA	1 µL
Phusion DNA polymerase	0.2 µL
MilliQ water	To 20 µL

Samples of *E. coli* colonies were transferred into the PCR reaction mixtures with sterilized wooden toothpicks.

The PCRs were performed in a thermal cycler with initial denaturation temperature set at 98 °C for 30 seconds, denaturation at 98 °C (10 seconds), primer annealing at 55 °C (30 seconds), extension at 72 °C (30 s/kb), final extension at 72 °C (10 minutes). Denaturation, annealing and extension steps were repeated for 35 cycles. The outcome of PCRs was evaluated by performing gel electrophoresis of the amplified DNA.

To verify the sequences of constructed plasmids after ligation colony PCRs were performed (table 2.2.4.2).

Table 2.2.4.2: Primers used for colony PCRs*

Primer	Sequence (5’ – 3’)	Description	Experiment
pX330_AAV2_ ITR_R	cgcacagatgcgta aggag	2nd intron presence	Sequence verification for type IV eukaryotic expression construct.
pX330_DinG_F	gctgagtcactgtat ggctac	2nd intron presence	
pX330_gRNA_	ggctagtccgttatc	G scaffold presence	

scaffold_F	accttg		
pX330_CMV_R	ggtacctctagagccatttgtc	G scaffold presence	
pX330_hybrid_intron_F	cctggagcacctgcc	Type IV eukaryotic expression construct. 1st intron presencce	
pX330_Csf1_R	gctggccagaagtcgc	1st intron presencce	
pX330_U6_F	ggactatcatatgcttaccgtaacttg	CrRNA guide presence	
IVA3-Chis_F_seq2	cgagcttgttcgctcactg	Cas6e presence.	Sequence verification for type IV minimal system expression construct.
007-RPCas6e-R	tatacacctgcccatatgactactcccctgccag	Cas6e presence.	
IVA3-Chis_F_seq2	cgagcttgttcgctc	Cas6e presence.	Sequence verification for type IV minimal system expression construct without DinG.
RC_IVA3_DinG_Chis_R	tatacacctgcatgcatggtggtgatgatggccacactggatatcctcttcaaaatacg	Cas6e presence.	
T7_terminator_R	ctcaagacccgttagaggc	HHR presence.	Sequence verification for pYTK095 guide plasmid.

*The annealing temperatures of all the designed primers were within the 50-55 °C range.

2.2.5 Removal of template DNA

PCR products were digested with DpnI restriction enzyme (1 µL per 50 µL of reaction mixture) for three hours at 37 °C to remove the leftover template, the enzyme was deactivated by incubation of the reaction mixture at 80 °C for 20 minutes. The plasmid DNA was purified from the reaction mixture using GeneJet PCR Purification Kit. If unspecific PCR products were

detected, the product would be purified from the electrophoretic gel with a GeneJet Gel Purification Kit.

2.2.6 DNA gel electrophoresis

1% agarose gels were used to analyze 0.5 – 10 kb long sequences. 2 % gels were used to analyze 0.1 – 0.5 kb long sequences. Gels were prepared by mixing 1 or 2 % (w/v) of agarose powder with 1xTAE buffer (40 mM TRIS base, 20 mM acetic acid, 1 mM EDTA) and heating the mixture until the agarose melts. Ethidium bromide was added to the melted agarose to achieve the concentration of 0.5 µg/mL. GeneRuler DNA Ladder Mix was used to evaluate sizes of the DNA bands. Electrophoresis was performed at a voltage of 120 V for 30 minutes. Results were then visualized by illuminating the gels with UV light.

2.2.7 Plasmid assembly

2.2.7.1 Assembly of guide sequences into pYTK095-tIX-crArray plasmid.

For assembly of RNA crRNA guides targeting the chloramphenicol resistance gene, pYTK095-tIX-crArray plasmid, obtained from Rimvyde Cepaite, was digested with BsaI (Eco31I) restriction endonuclease. The annealed crRNA guide oligonucleotides were diluted 100 times and assembled with the plasmid using a T4 ligase in T4 ligase buffer.

Ligation mixtures were incubated at room temperature for 15 minutes and then transformed into calcium-competent *E. coli* Mach1 strain cells. After overnight incubation at 37 °C on selective solid LB agar containing ampicillin the colonies that formed were screened by colony PCR.

2.2.8 Golden Gate plasmid assembly

2.2.8.1 Golden Gate protocol

For the assembly of plasmids Golden Gate ligation protocol was used. Golden Gate cloning uses a type IIS restriction enzyme and ligase in a restriction-ligation to assemble several DNA fragments in a defined order in a vector in a single reaction. T4 ligase and type IIS restriction enzyme were used in a 5 µL reaction. The vector to insert ratio was 1:1.

2.2.8.2 Assembly of plasmids carrying genes for the expression of minimal type IV expression systems

Plasmids carrying genes for the expression of minimal type IV expression systems were assembled from PCR products.

Table 2.2.5.2: Protocol for assembly of minimal type IV system expression plasmids.

Reagent	Volume
T4 DNA ligase buffer	0.5 µL
T4 DNA ligase	0.3 µL
AarI	0.16 µL
Oligos for AarI	0.1 µL
PCR product	200 ng
MilliQ water	To 5 µL

The mix was incubated in a thermal cycler, the reaction conditions are provided in table 2.2.7.1.

Table 2.2.8.2.1: Golden Gate assembly reaction conditions

Temperature	Time	
37 °C	2 min	15 cycles
16 °C	5 min	
37 °C	5 min	
80 °C	10 min	

The ligated plasmid was transformed into Top10 electrocompetent cells. After overnight incubation at 37 °C on selective solid LB agar containing kanamycin, colonies that formed were screened by colony PCR.

2.2.8.3 Assembly of a guide plasmid containing a hammerhead ribozyme sequence

To assemble a guide plasmid containing a hammerhead ribozyme sequence the pYTK095 PCR product (table 2.2.3.2) was incubated with T4 ligase, AarI and an HHR oligonucleotide (table 2.2.8.3.2) to circularize the DNA and insert the HHR sequence.

Table 2.2.8.3.1: Protocol for assembly of guide plasmid with HHR.

Reagent	Volume
T4 DNA ligase buffer	0.5 µL
T4 DNA ligase	0.3 µL
AarI	0.16 µL
Oligos for AarI	0.1 µL
pYTK PCR product	160 ng
Hammerhead ribozyme annealed oligonucleotide	28 ng
MilliQ water	To 5 µL

The mix was incubated in a thermal cycler, the reaction conditions are provided in table 2.2.7.1.

Table 2.2.8.3.2: Hammerhead ribozyme oligonucleotide sequence

Oligonucleotide	Sequence (5' – 3)
pYTK095 HHR oligo F	taggcgataactgatgaggccttcgggccgaaacggtgaaagccgtattatcgg
pYTK095 HHR oligo R	ctcgccgataatacggctttcaccgtttcggcccgaaggcctcatcagttatcg

Ligated plasmid was transformed into Top10 electrocompetent cells. After overnight incubation at 37 °C on selective solid LB agar containing ampicillin, colonies that formed were screened by colony PCR.

2.2.8.4 Assembly of eukaryotic expression vector

Type IV CRISPR-Cas system genes and crRNA guide sequences were cloned into a pX330 backbone under the control of a U6 promoter. Csf2-Csf4-DinG and Csf1-Cas6e-Csf3-Csf2 synthetic gene sequences, fragments of the crRNA guide sequence and three fragments of the pX330 plasmid backbone were assembled into a type IV eukaryotic expression vector in one reaction (table 2.2.8.4.1).

Table 2.2.8.4.1: Golden Gate assembly reaction mix

Reagent	Volume
T4 DNA ligase	1,5 µL
10X ligation buffer	2,5 µL
BsaI	1 µL
1.2 kb pX330 fragment (7,9 ng/ µL: 1200 bp)	1,99 µL
1.5 kb pX330 fragment (14,5 ng µL; 1500 bp)	1,35 µL
1.4 kb pX330 fragment (12,27 ng/ µL; 1400 bp)	1,5 µL
Csf 1, 3, 2 Cas6e (38,4 ng/ µL; 5002 bp)	1,7 µL
Csf 2, 4 DinG (33,9 ng/ µL; 4939 bp)	1,91 µL
Oligo 1 (2,8 ng/ µL; 44 bp)	0,2 µL
Oligo 2 (2,8 ng/ µL; 50 bp)	0,2 µL
MilliQ water	To 25 µL

The volume of each fragment was calculated according to the formula:

$$\frac{\frac{20\ (\text{fmol}) \times \text{size (bp)}}{\text{concentration (ng/µl)}}}{1520}$$

Table 2.2.8.4.1: Golden Gate assembly with BsaI restriction endonuclease reaction conditions for >3 fragments. The mix was incubated in a thermal cycler (table 2.1.1.1).

Temperature	Time	
37 °C	2 min	50 cycles
16 °C	5 min	
50 °C	5 min	
80 °C	10 min	

After incubation the assembled plasmid was transformed into electrocompetent cells from the NEB Stbl *E. coli* strain. After overnight incubation at 37 °C on selective solid LB agar containing ampicillin, colonies that formed were screened by colony PCR.

2.2.8 Plasmid DNA purification

Cells derived from colonies selected after colony PCR were incubated overnight at 37 °C and 220 rpm in cultivation tubes containing 4 mL of liquid LB media supplemented with an appropriate antibiotic. Plasmid samples were purified from overnight colonies with a GeneJet Plasmid Miniprep Kit according to the provided instruction manual. The Nanodrop 1000 spectrometer (table 2.1.1.1) was used to measure the concentration of extracted DNA.

2.2.9 Plasmid DNA purification without a kit

1.5 mL of overnight culture was added into a 2 mL microfuge tube and centrifuged at 10,000 g for 30 seconds. After pouring off the supernatant, the pellet was resuspended in 100 µL of cold resuspension solution (25 mM TRIS-HCl (pH 8), 50 mM glucose, 10 mM EDTA) and vortexed for 2 minutes. 200 µL of denaturing solution (0.2 N NaOH, 1.0 % SDS) was added and the contents of the tube were gently mixed. The solution was incubated on ice for 5 minutes, then 150 µL of cold renaturing solution (3 M potassium acetate, 2 M acetic acid) was added and the tube was again incubated on ice for 5 minutes, then centrifuged at 12,000 g for 5 minutes. The supernatant was carefully pipetted into a new tube. The mixture was incubated with 1 µL of 10 mg/mL RNAse A at 37 °C for 5 minutes. After the incubation 2 volumes (900 µL) of 100% ethanol and 1/10 volume (45 µL) of 3 M Na-acetate (pH 4.8) were added to the solution. The tube was placed at -80 °C for 30 minutes, centrifuged at 12,000 g for 30 minutes at 4°C, then the supernatant was poured off and the open, inverted tubes were drained out on a paper towel. The pellet was washed with 500 µL of cold 70 % ethanol. After centrifugation at 12,000 g for 5 minutes the ethanol was poured out and the tube was dried by inverting on a paper towel for 20

minutes. TE buffer (1 mM EDTA, 10 mM Tris-Cl) was added, and the tube was left overnight at room temperature for the resuspension to take place.

2.2.10 Restriction analysis

The sequences of plasmid constructs selected after colony PCR were verified before sequencing using restriction analysis.

Benchling software was used to predict the location of the bands on a gel after electrophoresis of the target plasmid digested with restriction enzymes. The lengths of the products obtained after digestion were then compared to the lengths expected.

ThermoFisher FastDigest enzymes were used for the restriction analysis. To evaluate the sequence of type IV eukaryotic expression vector digestion with XhoI was used, as well as double digestion with AarI and XbaI to verify the integrity of the GC rich region adjacent to the plasmid's U6 promoter. Double digestion with XbaI and BamHI was used to analyze the minimal type IV system expression vector sequence.

2.2.11 DNA sequencing

DNA sequences of the resultant constructs were verified by sequencing performed by SeqVision, MB.z

2.2.12 Bacterial glycerol stocks

Cultivation tubes containing 4 mL of liquid LB media supplemented with 4 µL of ampicillin were inoculated with NEB Stb *E. coli* cells carrying the eukaryotic type IV expression vector. The cells were grown overnight at 37 °C and 220 rpm. On the next day, 1 mL of overnight culture was transferred to a 2 mL screw top tube, 1 mL of 50 % glycerol was added to the tube and the contents were gently mixed. The stock tubes were kept in a freezer at -80 °C.

2.2.13 Electro-competent *E. coli* cells

With a sterile toothpick a colony of *E. coli* cells was inoculated into a cultivation tube with 4 mL of LB media and an appropriate antibiotic and grown overnight. 100 mL of LB media was then inoculated with 2 mL of the overnight culture to achieve a starting OD_{600} of 0.05. The cultures were grown in a shaker-incubator (table 2.1.1.1) at 37 °C and 220 rpm until OD_{600} reached 0.5-

0.6. The cultures were rapidly cooled in an ice bath and the cells were harvested by centrifugation at 4 °C and 4,000 g for 10 minutes in conical 50 mL centrifugation bottles. The cells were then washed with 50 mL ice-cold MilliQ water for 10 minutes at 4,000 g twice and with 50 mL ice-cold 10 % glycerol for 10 minutes at 4,000 g twice. After the last wash cells were resuspended in 200 µL of ice-cold 10% glycerol, aliquoted into 1.5 mL micro centrifuge tubes, flash-frozen in liquid nitrogen and stored at -80 °C.

2.2.14 Calcium competent *E. coli* cells

1 L of LB media was inoculated with 10 mL of *E. coli* overnight culture and grown at 37 °C and 220 rpm in a shaker-incubator until the OD_{600} reached 0.35-0.4. The culture flasks were rapidly cooled on an ice bath. The 1 L culture was split into 50 mL ice-cold conical centrifuge bottles. The cells were harvested by centrifugation for 15 minutes at 3,000 g. After the supernatant was removed the pelleted cells were resuspended in ice-cold 100 mM $MgCl_2$ and combined into one 50 mL centrifugation bottle. The cells were then washed with ice-cold 100 mM $MgCl_2$ by centrifugation for 15 minutes at 2,000 g. The pellet was resuspended in 50 mL of ice cold 100 mM $CaCl_2$ and kept on ice for 20 minutes. The cells were centrifuged for 15 minutes at 2,000 g. The supernatant was decanted with the pellet being resuspended in 50 mL of ice cold 85 mM $CaCl_2$, 15% glycerol. The cells were centrifuged for 15 minutes at 1,000. All centrifugation steps were performed at 4 °C. After decanting the supernatant, the pellet was resuspended in 2 mL of ice-cold 85 mM $CaCl_2$, 15% glycerol and the suspension was aliquoted into 1.5 mL microcentrifuge tubes and frozen in liquid nitrogen. The cells were then stored at -80 °C.

2.2.15 Transformation of electrocompetent *E. coli* cells

Tubes with electrocompetent *E. coli* cells that were kept at -80 °C were thawed on ice. 20 µL of the thawed electrocompetent cells were mixed with 20 ng of plasmid DNA and incubated on ice for 2 minutes. The competent cells were then added into an ice-cold 1-mm electroporation cuvette (Fisherbrand), bubble formation was avoided by flicking the tube. The cuvette was placed into the cuvette chamber of the MicroPulser electroporator (table 2.1.1.1), and a bacterial protocol (1.8 kV, 1 pulse) was used to electroporate the cells. The cells were then immediately resuspended in 1 mL of 37 °C SOC recovery media (pH = 6.9), transferred into a sterile cultivation tube and incubated in a shaker-incubator at 37 °C and 220 rpm for one hour. The culture was then spread on selective LB agar plates and incubated at 37 °C overnight.

2.2.16 Transformation of calcium competent *E. coli* cells

A 1.5 mL microcentrifuge tube containing 50 µL of calcium-competent *E. coli* cells was removed from the -80 °C freezer and thawed on ice. 20 ng of plasmid DNA was added to 50 µL of calcium-competent *E. coli* cells and incubated on ice for 30 minutes. The cells were then heated for 45 seconds at 42 °C, and incubated on ice for 2 minutes. After the incubation 500 µL of pre-warmed LB media was introduced into the microcentrifuge tube and the cells were kept at 37 °C for 45 minutes for outgrowth. The culture was then spread on selective LB agar plates and incubated at 37 °C overnight.

2.2.17 Efficiency of transformation assay

During the efficiency of transformation assay, calcium-competent *E. coli* BL21 Star (DE3) cells were transformed with chloramphenicol resistant target plasmids, as well as ampicillin resistant plasmids encoding for one of three selected crRNA guide sequences, or ampicillin resistant non-targeting control plasmids. Electrocompetent cells, carrying both crRNA guide and target plasmids, were prepared. The cells were then transformed with kanamycin resistant plasmids carrying the sequences of type IV systems, activity of which was evaluated. Six 10x dilution steps were performed with suspensions of transformed cells, 5 µL of each dilution was pipetted onto selective media containing chloramphenicol, kanamycin and ampicillin. The cell density was determined by the amount of colony forming units in the lowest dilution where colonies were still present.

2.2.18 Statistical analysis

The significance of data obtained during the plasmid curing assay was evaluated using a t-test performed in GraphPad Prism.

3. RESULTS

3.1 Construction of a type IV eukaryotic expression vector

The main objective of the study is the establishment of expression of functional type IV CRISPR-Cas system in human cells for gene editing purposes. To achieve the expression of a type IV CRISPR-Cas system in eukaryotic cells, an expression vector was constructed. This expression vector consists of a geneblock, which contains type IV CRISPR-Cas sequences and a guide sequence in a pX330 eukaryotic expression backbone plasmid. To construct the expression vector a gene block containing sequences, encoding for type IV proteins was designed (figure 3.3.1).

First, DNA sequences of prokaryotic CRISPR-Cas type IV proteins were codon-optimized for expression in *Homo sapiens*. Codon optimization was performed to achieve an efficient expression in *Homo sapiens* HEK cells.

P2A ribosome-skipping sequences were located between the codon-optimized genes, encoding for type IV proteins. Since the study is aimed at establishing joint expression of several proteins from one ORF in eukaryotic cells, and unlike prokaryotes, polycistronic expression is rare in eukaryotes, P2A sequences had to be located between the genes so that several independent polypeptide chains can be encoded from the same mRNA. During translation P2A sequences positioned between genes cause the skipping of a ribosome. As the ribosome reaches the P2A sequence, peptide chain is released after which translation is resumed, resulting in an absent peptide bond and individual release of the proteins.

Additionally, NLS sequences, followed by a FLAG tag were positioned at the C-terminal end of each type IV protein except DinG, which was tagged at the N-terminus. NLS sequences were required to facilitate nuclear import of synthesized and folded proteins back into the nucleus. SV40 Nuclear localization signals were used. A terminal FLAG-tag at the end of each protein was used to label the proteins for detection. During future studies, FLAG-tags will be used for Western blot analysis to detect expression of target proteins in the cells, as well as to determine their stoichiometry. The evaluation of the processing and stoichiometry of type IV proteins expressed in eukaryotic cells is needed to determine the effectiveness of P2A sequences.

Figure 3.1.1 *Arrangement of Csf1, Csf2, Csf3, Cas6e and DinG. Tags (FLAG-tag (red), NLS (pink), P2A (grey)) are located at the C-terminal ends of Csf1 (green), Csf2 (brown), Csf3 (grey), Cas6e (yellow) and the N-terminal end of DinG (blue).*

Sequence, encoding for the type IV proteins was assembled from two geneblocks.

The constructed expression vector contains a non-targeting crRNA guide sequence, since crRNA is required for the assembly of a type IV CRISPR-Cas interference complex, but we did not need the complex to target a gene at this stage of the study. Guide sequence was constructed from two oligonucleotides during the Golden Gate assembly.

As a plasmid backbone a pX330 plasmid for expression of CRISPR-Cas sequences in eukaryotic cells was used (figure 3.1.2). The plasmid is an established eukaryotic expression vector [44]. The plasmid contains a gRNA scaffold, so the crRNA sequence is encoded by the same plasmid, without need for cotransfection of several plasmids.

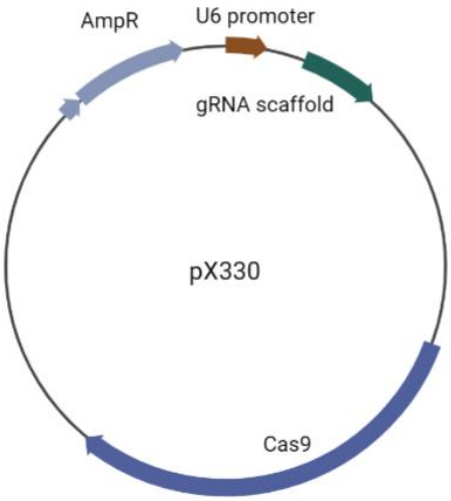

Figure 3.1.2 *pX330 eukaryotic expression vector used as a backbone for the eukaryotic expression of type IV proteins. The plasmid contains a U6 promoter and a gRNA scaffold.*

In a pX330 plasmid the protein encoding sequences are placed under control of a CMV-CAG promoter, the guide sequence is placed under control of a U6 promoter, which is a type III RNA polymerase III promoter with a tight control over the length of the transcribed sequence. The plasmid encodes for Cas9 protein.

For the construction of the expression vector Golden Gate ligation was used. We assembled the vector backbone with the Cas-proteins encoding ORF and the guide sequence, with Cas-proteins encoding ORF replacing Cas9 in the vector.

Type IIS enzyme recognition sequences had to be added to the plasmid for the Golden Gate assembly. PCR amplifications were carried out to introduce AarI enzyme recognition sites into the pX330 plasmid.

The efficiency of PCR can be impacted by the GC content of the template sequence. DNA templates with an >65% GC content can form complex secondary structures. Guanine and cytosine bases form more hydrogen bonds, resulting in the DNA being resistant to melting.

Due to the high GC content of the pX330 backbone, especially the sequences adjacent to the promoter, affecting efficiency of PCRs, the plasmid backbone was divided into three fragments, 1.5 kb, 1.4 kb, and 1.2 kb in length, to be amplified in separate reactions (figure 3.1.3).

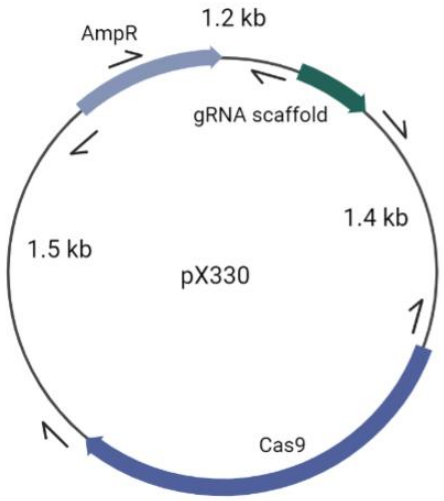

Figure 3.1.3 *pX330 expression vector. Binding positions of the primers used to amplify the pX330 plasmid are shown on the scheme.*

Type IIS enzyme recognition sites were inserted into the sequences for sticky end ligation during the Golden Gate assembly.

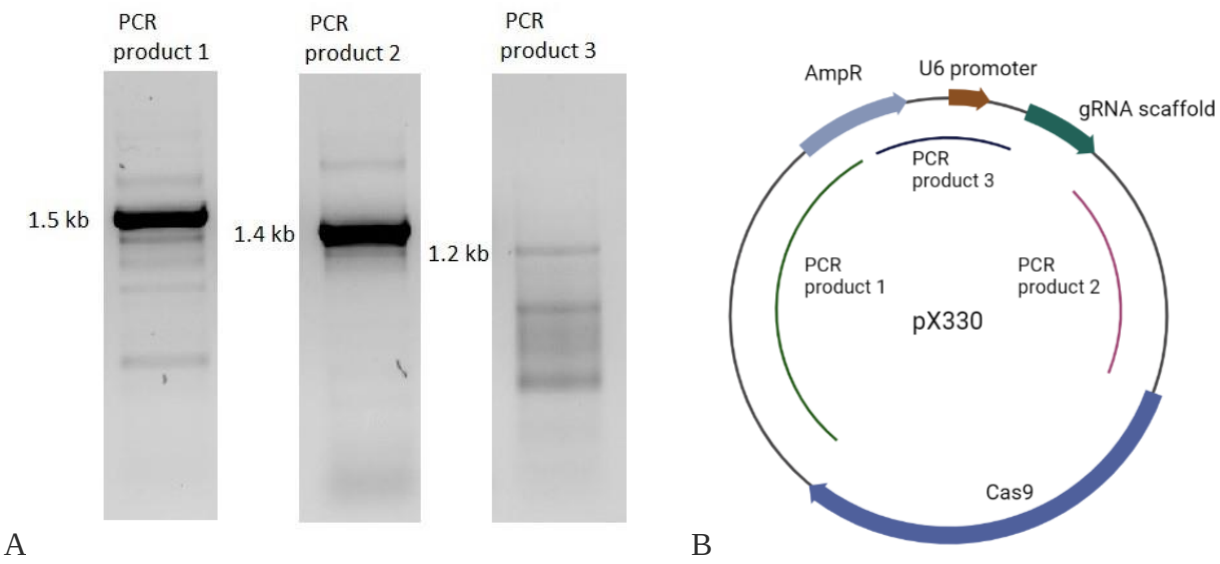

Figure 3.1.4 *Three separate PCR reactions performed during PCR amplification of pX330 plasmid. A Electrophoretic gel, showing the three PCR products obtained during the amplification, PCR product 1 is 1.5 kb, PCR product 2 is 1.4 kb, and PCR product 3 is 1.2 kb long. B Locations of the three sequences in the plasmid.*

Three fragments of the pX330 plasmid obtained after PCRs were assembled with the guide sequence and two geneblocks carrying the type IV sequences in a Golden Gate reaction (figure 3.1.5).

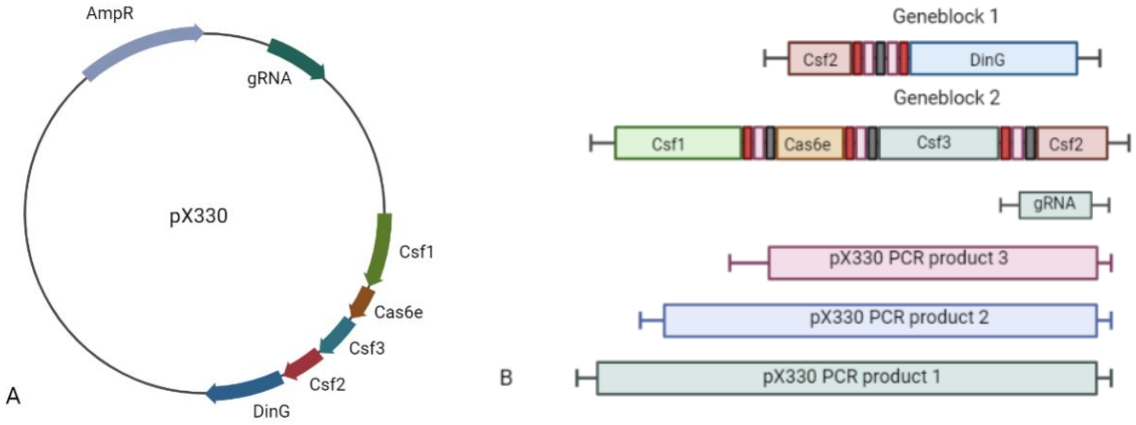

Figure *3.1.**5** Assembly of the eukaryotic expression vector. A Scheme of the constructed type IV eukaryotic expression vector. B Sequences, used for the assembly of the eukaryotic expression vector. The three fragments of pX330 plasmid, two type IV geneblocks and a guide oligonucleotide are shown.*

Sequencing results for the constructed vector for the eukaryotic expression of CRISPR-Cas type IV system matched the expected sequence and indicated the absence of mutations.

3.2 Construction of compact type IV CRISPR-Cas system

Hypercompact CRISPR-Cas systems for genetic engineering can be used with vectors with a low cargo capacity, such as AAV [45]. The maximum cargo capacity of an AAV is <4.7 kb, spCas9 is 4.2 kb. type IV CRISPR-Cas system can be reduced to 2.5 - 3 kb So, the possibility of developing a compact, functional CRISPR-Cas type IV system through the removal of Cas6e from the interference complex is being investigated in the study.

Type IV CRISPR-Cas Cas6e protein is responsible for the cleavage of pre-crRNA repeats, ensuring the maturation of crRNA during the stage of crRNA biogenesis. Cas6e enzymes use a metal-ion-independent mechanism to cleave crRNAs on the 3′-side of stem-loops formed within the palindromic CRISPR repeat. The crRNA is cleaved on the 3′ side of the predicted stem-loop structure, with nucleophilic attack on the scissile phosphate coming from the 2′ hydroxyl of base G22 of the repeat. Cas6 cleavage activity generates products with 5′-hydroxyl and 2′–3′ cyclic phosphate ends.

The Hammerhead ribozyme is a compact RNA motif that can perform self-cleavage through endonucleolytic activity. It consists of a catalytic core that is surrounded by three helices, two of which are crucial for tertiary interactions required for self-cleavage in physiological conditions. When the HHR undergoes self-cleavage, it generates 5′-OH and 2′, 3′-cyclic phosphate termini, which are the same as those produced by Cas6e.

If type IV system is supplemented with mature unit-sized crRNAs in a Cas6e-independent manner, it can be possible to establish functioning of type IV CRISPR-Cas without Cas6e protein.

To provide mature gRNAs to a Cas6e-deficient type IV system for binding to Cas5, a Hammerhead ribozyme sequence, located upstream from the gRNA on a guide plasmid may be used to cleave pre-crRNA with formation of 2′, 3′-cyclic phosphate termini necessary for recognition of the crRNA by Cas5.

During this part of the study, prokaryotic expression plasmids, carrying a type IV system that lacks Cas6e were constructed. A guide plasmid with an HHR sequence upstream from the crRNA sequence was constructed to be used with the Cas6e-deficient type IV systems.

3.2.1 Construction of plasmids for prokaryotic expression of Cas6e-deficient type IV systems

A pRSF vector plasmid, encoding for Cas6e deficient type IV system was constructed using a plasmid carrying a complete type IV system as a template.

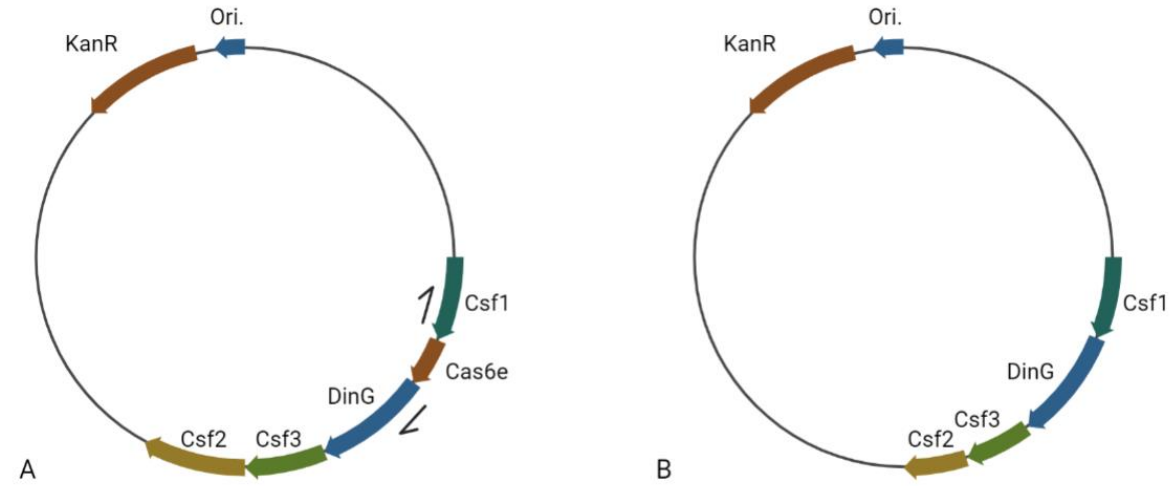

Figure 3.2.1.1 *A Template plasmid used for the construction of a Cas6e deficient type IV expression plasmid. Primer binding sites are shown. B Scheme of the constructed Cas6e deficient type IV expression plasmid*

PCR was performed to remove Cas6e from the sequence and insert BsaI recognition sites (figure 3.2.1.1). A Golden Gate ligation was performed to recircularize the plasmid.

Sequencing of the constructed expression vector indicated that it matched the expected sequence.

3.2.2 Construction of plasmids for prokaryotic expression of Cas6e and DinG - deficient type IV systems

DinG is a DNA helicase, separately recruited by the type IV interference complex upon binding target DNA. A DinG deficient type IV system can be used in the development of next-generation gene editors since it can bind DNA but does not possess catalytic activity.

A pRSF vector plasmid, encoding for the type IV system that lacks Cas6e and DinG was constructed, using a plasmid carrying a type IV system without DinG as a template.

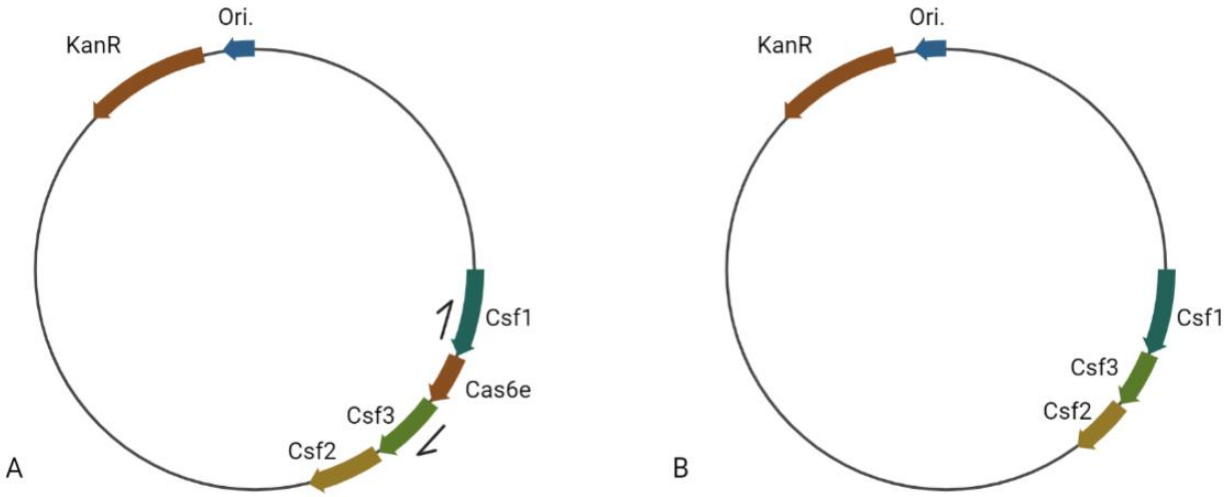

Figure 3.2.2.1 *A Template plasmid used for the construction of a Cas6e and DinG deficient type IV expression plasmid. Primer binding sites are shown. B Scheme of the constructed Cas6e and DinG deficient type IV expression plasmid*

PCR was performed to remove Cas6e and insert BsaI recognition sites (figure 3.2.2.1). A Golden Date ligation was performed to recircularize the plasmid.

Sequencing of the constructed expression vector indicated that it matched the expected sequence and lacked significant mutations.

3.2.3 Construction of a guide plasmid carrying a Hammerhead ribozyme sequence

A modified guide plasmid, carrying a self-cleaving sequence is required to ensure maturation of pre-crRNAs in a type IV CRISPR-Cas system that lacks Cas6e. In this study, an HHR sequence was introduced into the guide plasmid (figure 3.2.3.1) to maintain the formation of mature crRNAs in a Cas6e independent manner by self-cleavage of the HHR sequence.

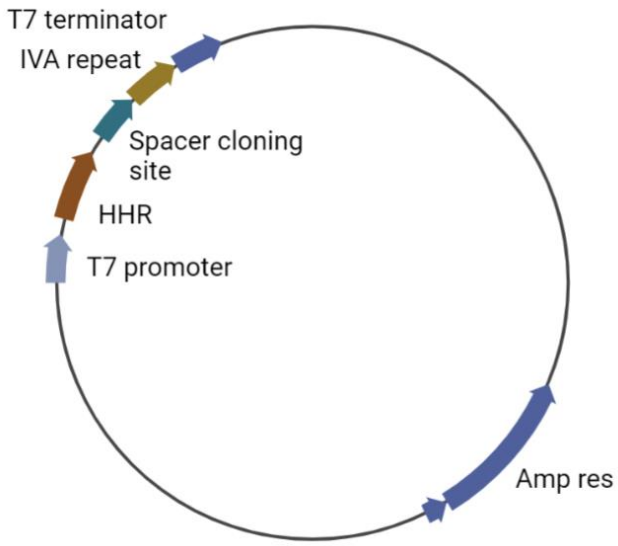

Figure 3.2.3.1 *Guide plasmid with a HHR sequence located upstream from the guide RNA sequence*

Since cleavage by a Hammerhead ribozyme creates 5′-OH and 2′, 3′-cyclic phosphate termini, an HHR sequence, located upstream from the guide RNA can form mature guide RNAs, identical to those produced by Cas6e.

In this study, a self-cleaving ribozyme module was attached upstream of a guide RNA encoding sequence, to create a ribozyme-gRNA cassette.

PCR was performed to introduce BsaI recognition sequences. The HHR fragment was inserted into the plasmid using a Golden Gate protocol.

Sequencing of the constructed expression vector indicated that it matched the expected sequence.

3.3 Assay of the activity of engineered type IV CRISPR-Cas systems

CRISPR-Cas systems can be used to affect gene expression without altering the sequence of the gene in next-generation gene editing. To engineer a next-generation gene editing system it is necessary to create fusions of CRISPR-Cas proteins that can be targeted to a specific sequence with enzymes that can act as epigenetic effectors. To fuse a transcription repressor or activator domain to one of the proteins in the type IV complex, without the complex losing its DNA binding efficiency it is necessary to determine tolerance of the type IV system to introduction of additional regulatory domains. For this purpose, a method was established to detect and quantify activity of the type IV system in prokaryotic cells, with the goal of using this method to determine the effect of modifications on the activity of type IV CRISPR-Cas systems.

For the assay, E. coli cells that carry a chloramphenicol-resistance target plasmids and plasmids, encoding for the type IV CRISPR-Cas system evaluated in the experiment are used.

These cells were transformed with either a guide plasmid or a non-targeting control plasmid. The guide plasmid encodes for crRNAs, targeting the chloramphenicol resistance gene. Transformed cells are plated on selective media that contains chloramphenicol. The decreased number of colonies formed by the cells transformed with the targeting guide, compared to the control cells, is an indication of activity of the evaluated system (figure 3.3.1).

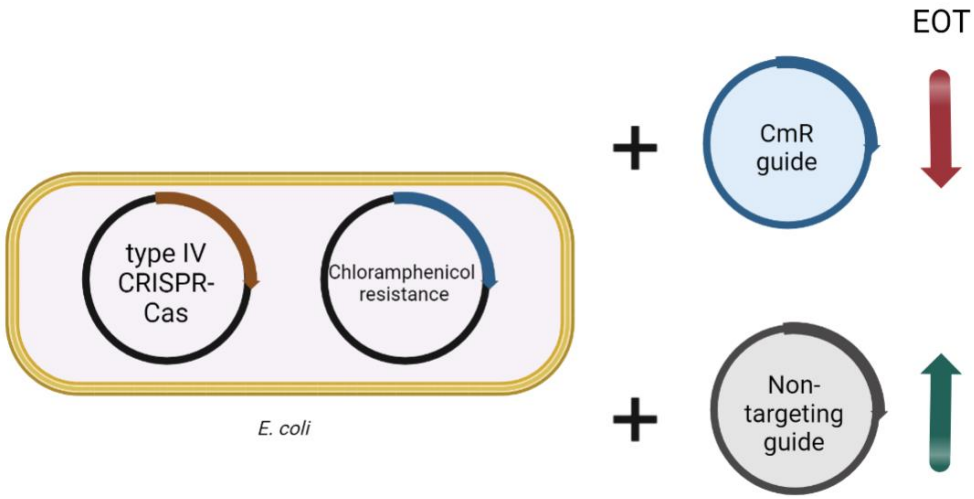

Figure 3.3.1 *Experimental setup for the efficiency of transformation (EOT) assay of a type IV CRISPR-Cas system. Cells, carrying a chloramphenicol-resistance target plasmids and plasmids, encoding for the type IV CRISPR-Cas system evaluated are transformed with either a guide plasmid or a non-targeting control plasmid*

3.3.1 Design of gRNAs for the efficiency of transformation assay

To perform the efficiency of transformation assay a selection of crRNAs, targeting the chloramphenicol resistance gene was designed. Three crRNA sequences were chosen for the assay according to the region of the gene they target. crRNA 1 targets the sequence of the gene that is close to the promoter, while crRNAs 2 and 3 target the gene promoter (figure 3.3.1.1). All crRNAs are complementary to the bottom strand of the targeted gene. The crRNA encoding sequences were cloned into a guide plasmid.

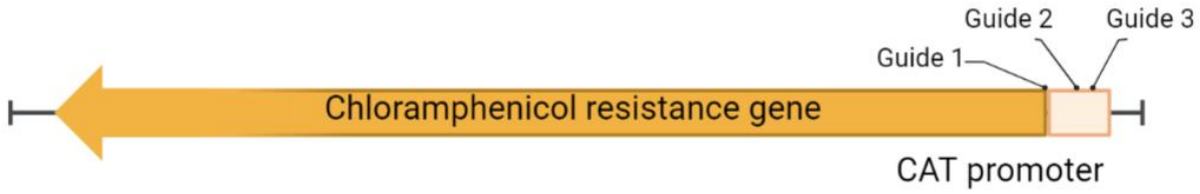

Figure 3.3.1.1 *Target gene that encodes chloramphenicol resistance. Sites complementary to the crRNA sequences used during the study are shown.*

3.3.2 Analysis of the activity of CRISPR-Cas type IV system

E. coli BL21(DE3)-Star cells, containing the chloramphenicol-resistance plasmids, plasmids, encoding for the type IV CRISPR-Cas system and guide plasmids that encode for crRNAs, targeting the chloramphenicol resistance gene, or a plasmid carrying a non-targeting guide crRNA were inoculated onto plates with selective media containing chloramphenicol. After the plated cells were incubated for 12 hours, activity of the type IV system was determined according to the number of colonies formed on selective media. Cells, transformed with guide plasmids forming fewer colonies than cells, transformed with control plasmids, indicate that the system being evaluated is active and cleaves DNA, while colony count similar to the control indicates that the system is inactive.

Activity of the system was quantified as the number of colony forming units per nanogram of guide plasmid. The number of colony-forming units per nanogram of plasmid was determined by the number of colonies formed on selective plates after 12 hours of incubation in the highest dilution of cells (figure 3.3.2.1). Colonies in the highest dilution of cells were counted, then, the number of colony forming units in the original sample was calculated. To determine the number of colony forming units in the original sample, the number of colony forming units in a dilution was multiplied by 1/ FDF (final dilution factor). The final dilution factor was calculated according to a formula:

$$FDF = SDF \times TSDF \times PDF$$

Here, SDF is Sample Dilution Factor, TSDF is Total Series Dilution Factor and PDF is Plating Dilution Factor.

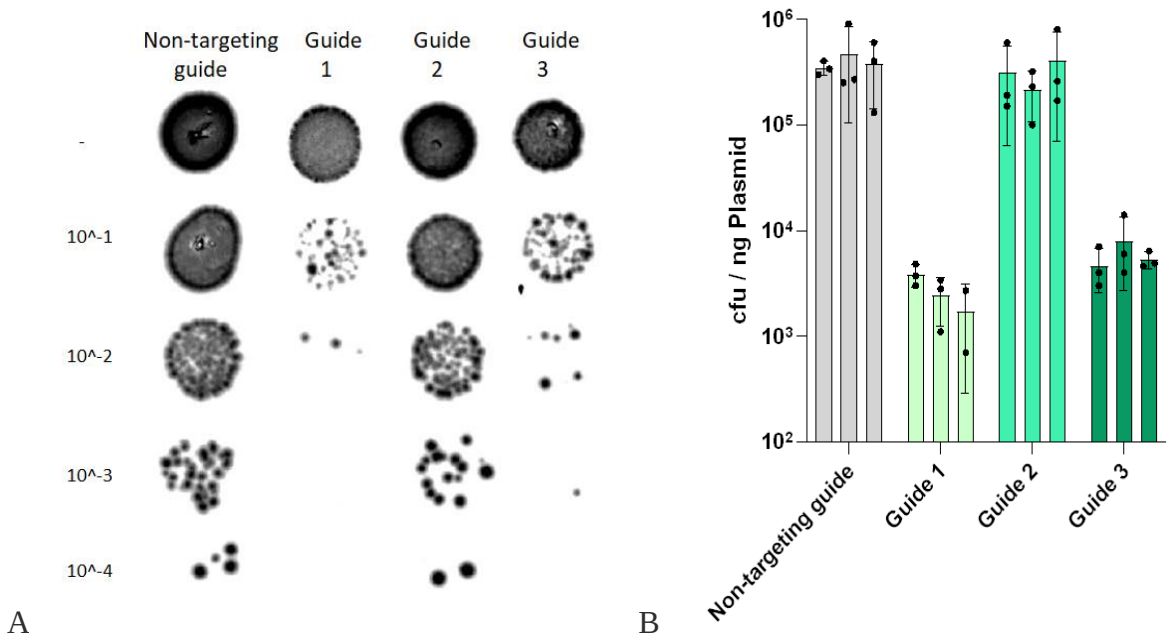

Figure 3.3.2.1 *Results of efficiency of transformation analysis of the CRISPR-Cas type IV system. A. Transformed E. Coli cells after 12-hour incubation on selective media, no IPTG. B Bar graphs representing the data obtained from three biological replicates performed for the experiment. Dots represent individual values obtained from technical replicates.*

Type IV system showed activity in the presence of guides 1 and 3. Activity of the system with guide 2 seems to match that of the non-targeting control.

DISCUSSION

CRISPR-Cas is a nucleoprotein system found in bacteria and archaea, which plays a crucial role in adaptive immunity against mobile genetic elements by recognizing and degrading nucleic acid sequences in a guide-RNA dependent manner. Due to their ability to bind and cleave specific nucleic acid sequences complementary to the crRNA, CRISPR-Cas systems are widely used in gene editing. At the same time, use of CRISPR-Cas systems in therapeutics can be limited by low cargo capacity of gene editing delivery systems, such as Adeno-associated virus (AAV).

CRISPR-Cas systems include two major classes and six types. Class 1 has a multi-subunit effector complex while class 2 has a single-protein effector complex. Type IV CRISPR-Cas systems belong to class 1 and have a multi-subunit effector complex that consists of several small proteins and a crRNA. Type IV CRISPR-Cas systems offer several advantages as potential gene editing systems, including their smaller size that can be useful in development of minimal CRISPR-Cas systems needed to overcome the problems presented by the low cargo capacity of AAV. Additionally, multi-subunit effector complex of type IV CRISPR-Cas allows for various effector protein-regulatory domain fusion strategies for design of epigenetic editors.

This project is aimed at establishing the use of class 1 type IV CRISPR-Cas systems as gene editors. To achieve this, a eukaryotic vector was constructed for the expression of type IV CRISPR-Cas system in mammalian cells. A hypercompact type IV CRISPR-Cas system was also engineered.

According to the aims of the project, a type IV CRISPR-Cas eukaryotic expression vector for gene editing in mammalian cells was constructed. To direct the gene editing tool to the nucleus of the eukaryotic cell enable targeting of the eukaryotic genome, editing proteins were outfitted with NLS (nuclear localization signal) tags. Sequences were also outfitted with FLAG-tags–to enable detection, purification, and localization within cells. P2A sequences–were positioned between the type IV proteins in the ORF. To cause a ribosomal skipping during translation, resulting in the production of two separate polypeptide chains and allowing for the simultaneous expression of multiple proteins without the need for separate vectors. The expression of guide RNA (gRNA) and Cas proteins from the single vector was established with the use of a dual-promoter system: U6 promoter was used to drive the expression of the gRNA, while the CMV-CAG promoter drove the expression of the Cas proteins. In future experiments the construct will be used to express the type IV system in eukaryotic cells.

Type IV CRISPR-Cas systems are compact but can possibly be minimized even further. A strategy for developing a hypercompact type IV CRISPR-Cas system was explored in this study.

The hypercompact type IV system was developed through the removal of Cas6e protein from the system and introduction of a hammerhead ribozyme sequence to perform its function. Cas6e is an endonuclease enzyme that cleaves the crRNA precursor into smaller, mature crRNAs that can guide the Cas effector proteins to their target sequences in the genome. During Cas6e-fascilitated cleavage, crRNA products with 5′-hydroxyl and 2′–3′ cyclic phosphate ends are generated. HHR is a catalytic RNA molecule able to perform self-cleavage while forming 5′-hydroxyl and 2′–3′ cyclic phosphate ends, identical to those produced by Cas6e. This suggests that it can be possible for HHR to perform the function of Cas6e in the type IV system, so prokaryotic expression vectors were constructed to carry compact type IV CRISPR-Cas systems that lack a Cas6e sequence. Plasmid carrying a guide sequence, supplemented with a hammerhead ribozyme sequence, where hammerhead ribozyme sequence was positioned upstream from the guide to establish crRNA maturation in a Cas6e independent manner was also constructed. In the future activity of the constructed hypercompact type IV CRISPR-Cas system will be evaluated to determine whether self-cleaving hammerhead ribozyme can be used to achieve maturation of pre-crRNA in the absence of Cas6e.

Before constructing type IV CRISPR-Cas-based next generation gene editors, it is important to establish a method of evaluation of the activity of the system. To do this, two crRNA guides that enable a type IV CRISPR-Cas system to target a chloramphenicol resistance gene were designed for the study. Engineered type IV systems can be transformed into cells that contain chloramphenicol resistance plasmids and designed chloramphenicol-resistance targeting guide plasmids. Activity of the engineered system can then be evaluated according to the number of colony forming units after incubation on chloramphenicol-containing media. The method will later be used to test the activity of engineered CRISPR-Cas sequences during development of next generation genome editors. This method can allow us to determine locations in which the regulatory domains can be fused to the type IV proteins without affecting proper folding of the proteins, assembly of the complex or its efficiency in binding target DNA.

Expression of type IV CRISPR-Cas systems in eukaryotic cells can be useful to study their function and potential applications in genome editing and gene regulation. Type IV CRISPR-Cas systems are relatively new and have unique features, such as their small size, that make them promising tools for genetic manipulation. Type IV CRISPR-Cas based genome engineering systems have potential benefits and can expand the CRISPR engineering toolbox, by expressing those systems in eukaryotic cells, their efficiency, specificity, and safety in vivo can be explored. If determined to be active, the constructed system can be used in AAV vectors for in vivo CRISPR editing. Development of compact CRISPR-Cas systems is vital because they have

several advantages over larger systems. Compact systems are easier to deliver into cells, which makes them more efficient for various gene editing applications, including in vivo gene therapy.

CONCLUSIONS

1. A type IV CRISPR-Cas eukaryotic expression vector for gene editing in mammalian cells was constructed. After its activity is evaluated, the constructed eukaryotic expression sequence can be used for gene editing in human cells.

2. To create a compact type IV CRISPR-Cas system that can function in absence of Cas6e several plasmids were engineered. Prokaryotic expression vectors were constructed to carry type IV CRISPR-Cas systems that lack a Cas6e sequence, as well as compact type IV CRISPR-Cas systems that lack both DinG and Cas6e sequences. A guide plasmid, supplemented with a self-cleaving Hammerhead ribozyme sequence located upstream from the crRNA-encoding sequence was constructed to be used with the compact type IV systems. If determined to be effective, the developed system can be used to help overcome the challenge presented by low cargo capacity of AAV

3. A method for evaluation of activity of a type IV system in prokaryotic cells was established. crRNA guides that enable a type IV CRISPR-Cas system to target a chloramphenicol resistance gene were designed. The system will be used to test the activity of engineered CRISPR-Cas effector complexes, helping in development of novel next-generation type IV genome editors.

VILNIUS UNIVERSITY

EMBL partnership

institute

Daria Ponomariova

Establishment of type IV CRISPR-Cas for genome editing in human cells

Master thesis

Vilnius University, EMBL

Partnership Institute

The Laboratory of Dr Patrick Pausch

SUMMARY

CRISPR-Cas (clustered regularly interspaced short palindromic repeats–CRISPR-associated protein) is a gene editing system derived from prokaryotic organisms that can recognize and degrade nucleic acid sequences in a guide-RNA dependent manner. The system is represented by a nucleoprotein effector complex that consists of Cas proteins and a guide RNA that targets the system to the DNA fragment of interest. Based on the structure of the effector complex, as well as signature proteins CRISPR-Cas systems can be divided into two major classes and a variety of types and subtypes. Class 1 systems are characterized by an effector complex that consists of multiple proteins, while class 2 systems have an effector complex with a single effector protein. Type IV CRISPR-Cas systems belong to class 1, and have a compact, crRNA-guided multi-subunit effector complex. Due to their ability to bind and cleave predetermined regions of DNA or RNA, CRISPR-Cas systems are often used in gene editing. Beyond nucleic acid cleavage, deactivated CRISPR-Cas complexes, sometimes fused to regulatory domains, can be used for regulation of gene expression in systems such as CRISPRi, CRISPRa or CRISPR-mediated epigenetic editing.

Despite the development of versatile tools for genome editing based on CRISPR-Cas systems, potential applications of CRISPR editing in human cells is limited by low cargo capacity of commonly used viral vectors, such as AAV. However, there is a possibility to adapt type IV CRISPR-Cas systems for use with low cargo capacity vectors, as their effector complexes are compact, compared to the common CRISPR-Cas effectors such as Cas9. Another advantage of type IV CRISPR-Cas over conventionally used CRISPR-Cas systems is their multi-subunit

effector complex. Type IV system subunits can be fused with regulatory domains and used in CRISPRi, CRISPRa or CRISPR epigenetic editing, since it allows for the development of different Cas protein–effector domain fusion strategies.

This research work describes the adaptation of a type IV CRISPR-Cas system for genome editing, which includes construction of a plasmid vector for expression of a type IV CRISPR-Cas system in human cells, development of a minimal CRISPR-Cas type IV system and development of next-generation genome editors based on type IV CRISPR-Cas. The research carried out can help to establish class 1 type IV CRISPR complexes as viable tools for eukaryotic gene editing.

LITERATURE

1. Makarova KS, Koonin EV. Annotation and Classification of CRISPR-Cas Systems. Methods Mol Biol. 2015;1311:47-75. doi: 10.1007/978-1-4939-2687-9_4. PMID: 25981466; PMCID: PMC5901762.

2. Kamruzzaman Muhammad, Iredell Jonathan R. CRISPR-Cas System in Antibiotic Resistance Plasmids in Klebsiella pneumonia. Frontiers in Microbiology vol 10 2020 https://www.frontiersin.org/articles/10.3389/fmicb.2019.02934 doi: 10.3389/fmicb.2019.02934

3. Samuel H. Sternberg, Hagen Richter, Emmanuelle Charpentier, Udi Qimron, Adaptation in CRISPR-Cas Systems. Molecular Cell, Volume 61, Issue 6, 2016, Pages 797-808, ISSN 1097 2765, https://doi.org/10.1016/j.molcel.2016.01.030.

4. Adrian G Patterson, Mariya S Yevstigneyeva, Peter C Fineran, Regulation of CRISPR–Cas adaptive immune systems, Current Opinion in Microbiology, Volume 37, 2017, Pages 1-7, ISSN 1369-5274, https://doi.org/10.1016/j.mib.2017.02.004.

5. Anaïs Le Rhun, Andrés Escalera-Maurer, Majda Bratovič & Emmanuelle Charpentier (2019) CRISPR-Cas in Streptococcus pyogenes, RNA Biology, 16:4, 380-389, DOI: 10.1080/15476286.2019.1582974

6. Koonin EV, Makarova KS. Origins and evolution of CRISPR-Cas systems. Philos Trans R Soc Lond B Biol Sci. 2019 May 13;374(1772):20180087. doi: 10.1098/rstb.2018.0087. PMID: 30905284; PMCID: PMC6452270.

7. Globyte, Viktorija, Lee, Seung Hwan, Bae, Taegeun, Kim, Jin-Soo, Joo, Chirlmin "CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion" The EMBO Journal vol. 38 is 4 SN 0261-4189 https://doi.org/10.15252/embj.201899466

8. Yan Zeng, Yang Cui, Yong Zhang, Yanruo Zhang, Meng Liang, Hui Chen, Jie Lan, Guangtao Song, Jizhong Lou, The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex, Nucleic Acids Research, Volume 46, Issue 1, 9 January 2018, Pages 350–361, https://doi.org/10.1093/nar/gkx1117

9. Eugene V Koonin, Kira S Makarova, Feng Zhang, "Diversity, classification and evolution of CRISPR-Cas systems", Current Opinion in Microbiology, Volume 37, 2017, Pages 67-78, ISSN 1369-5274, https://doi.org/10.1016/j.mib.2017.05.008.

10. Liu TY, Doudna JA. Chemistry of Class 1 CRISPR-Cas effectors: Binding, editing, and regulation. J Biol Chem. 2020 Oct 16;295(42):14473-14487. doi: 10.1074/jbc.REV120.007034. Epub 2020 Aug 14. PMID: 32817336; PMCID: PMC7573268.

11. Yanli Zheng, Jie Li, Baiyang Wang, Jiamei Han, Yile Hao, Shengchen Wang, Xiangdong Ma, Shihui Yang, Lixin Ma, Li Yi and Wenfang Peng "Endogenous Type I CRISPR-Cas: From Foreign DNA Defense to Prokaryotic Engineering", Frontiers in Bioengineering and Biotechnology, 2020, vol. 8

12. Jackson RN, Lavin M, Carter J, Wiedenheft B. Fitting CRISPR-associated Cas3 into the helicase family tree. Curr Opin Struct Biol. 2014 Feb;24:106-14. doi: 10.1016/j.sbi.2014.01.001. Epub 2014 Jan 28. PMID: 24480304; PMCID: PMC3984625.

13. Pickar-Oliver, A., Black, J.B., Lewis, M.M. et al. Targeted transcriptional modulation with type I CRISPR–Cas systems in human cells. Nat Biotechnol 37, 1493–1501 (2019). https://doi.org/10.1038/s41587-019-0235-7

14. Ning Jia, Charlie Y. Mo, Chongyuan Wang, Edward T. Eng, Luciano A. Marraffini, Dinshaw J. Patel, "Type III-A CRISPR-Cas Csm Complexes: Assembly, Periodic RNA Cleavage, DNase Activity Regulation, and Autoimmunity", Molecular Cell, Volume 73, Issue 2, 2019, Pages 264-277.e5, ISSN 1097-2765, https://doi.org/10.1016/j.molcel.2018.11.007.

15. Shmakov, S., Smargon, A., Scott, D. et al. Diversity and evolution of class 2 CRISPR–Cas systems. Nat Rev Microbiol 15, 169–182 (2017). https://doi.org/10.1038/nrmicro.2016.184

16. Hannah N. Taylor, Emily E. Warner, Matthew J. Armbrust, Valerie M. Crowley, Keith J. Olsen & Ryan N. Jackson (2019) Structural basis of Type IV CRISPR RNA biogenesis by a Cas6 endoribonuclease, RNA Biology, 16:10, 1438-1447, DOI: 10.1080/15476286.2019.1634965

17. Valerie M. Crowley, Adam Catching, Hannah N. Taylor, Adair L. Borges, Josie Metcalf, Joseph Bondy-Denomy, and Ryan N. Jackson "A Type IV-A CRISPR-Cas System in Pseudomonas aeruginosa Mediates RNA-Guided Plasmid Interference In Vivo" The CRISPR Journal 2019 2:6, 434-440

18. Koonin, Eugene V., Kira S. Makarova, and Feng Zhang. "Diversity, classification and evolution of CRISPR-Cas systems." Current opinion in microbiology 37 (2017): 67-78.

19. Taylor Hannah N., Laderman Eric, Armbrust Matt, Hallmark Thomson, Keiser Dylan, Bondy-Denomy Joseph, Jackson Ryan N. Positioning Diverse Type IV Structures and Functions Within Class 1 CRISPR-Cas Systems Frontiers in Microbiology Vol. 12 2021

20. Özcan, A., Pausch, P., Linden, A. et al. Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. Nat Microbiol 4, 89–96 (2019).

21. Bernheim, Aude, et al. "Atypical organizations and epistatic interactions of CRISPRs and cas clusters in genomes and their mobile genetic elements." Nucleic acids research 48.2 (2020): 748-760.

22. Pinilla-Redondo, Rafael, et al. "Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids." Nucleic acids research 48.4 (2020): 2000-2012.

23. Makarova, Kira S., et al. "Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants." Nature Reviews Microbiology 18.2 (2020): 67-83.

24. Khan SH. Genome-Editing Technologies: Concept, Pros, and Cons of Various Genome-Editing Techniques and Bioethical Concerns for Clinical Application. Mol Ther Nucleic Acids. 2019 Jun 7;16:326-334. doi: 10.1016/j.omtn.2019.02.027. Epub 2019 Apr 3. PMID: 30965277; PMCID: PMC6454098.

25. Asmamaw M, Zawdie B. Mechanism and Applications of CRISPR/Cas-9-Mediated Genome Editing. Biologics. 2021 Aug 21;15:353-361. doi: 10.2147/BTT.S326422. PMID: 34456559; PMCID: PMC8388126.

26. Hakim Manghwar, Keith Lindsey, Xianlong Zhang, Shuangxia Jin, CRISPR/Cas System: Recent Advances and Future Prospects for Genome Editing, Trends in Plant Science, Volume 24, Issue 12, 2019, Pages 1102-1125, ISSN 1360-1385, https://doi.org/10.1016/j.tplants.2019.09.006.

27. Adli M. The CRISPR tool kit for genome editing and beyond. Nat Commun. 2018 May 15;9(1):1911. doi: 10.1038/s41467-018-04252-2. PMID: 29765029; PMCID: PMC5953931.

28. Fellmann, C., Gowen, B., Lin, PC. et al. Cornerstones of CRISPR–Cas in drug discovery and therapy. Nat Rev Drug Discov 16, 89–100 (2017). https://doi.org/10.1038/nrd.2016.238

29. Heidenreich, M., Zhang, F. Applications of CRISPR–Cas systems in neuroscience. Nat Rev Neurosci 17, 36–44 (2016). https://doi.org/10.1038/nrn.2015.2

30. Zhang D, Zhang Z, Unver T, Zhang B. CRISPR/Cas: A powerful tool for gene function study and crop improvement. J Adv Res. 2020 Oct 21;29:207-221. doi: 10.1016/j.jare.2020.10.003. PMID: 33842017; PMCID: PMC8020163.

31. Wu X, Kriz AJ, Sharp PA. Target specificity of the CRISPR-Cas9 system. Quant Biol. 2014 Jun;2(2):59-70. doi: 10.1007/s40484-014-0030-x. PMID: 25722925; PMCID: PMC4338555.

32. Brezgin S, Kostyusheva A, Kostyushev D, Chulanov V. Dead Cas Systems: Types, Principles, and Applications. Int J Mol Sci. 2019 Nov 30;20(23):6041. doi: 10.3390/ijms20236041. PMID: 31801211; PMCID: PMC6929090.

33. Gangopadhyay SA, Cox KJ, Manna D, Lim D, Maji B, Zhou Q, Choudhary A. Precision Control of CRISPR-Cas9 Using Small Molecules and Light. Biochemistry. 2019 Jan 29;58(4):234-244. doi: 10.1021/acs.biochem.8b01202. Epub 2019 Jan 22. PMID: 30640437; PMCID: PMC6586488.

34. Goell JH, Hilton IB. CRISPR/Cas-Based Epigenome Editing: Advances, Applications, and Clinical Utility. Trends Biotechnol. 2021 Jul;39(7):678-691. doi: 10.1016/j.tibtech.2020.10.012. Epub 2021 May 7. PMID: 33972106.

35. Thakore PI, Black JB, Hilton IB, Gersbach CA. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. Nat Methods. 2016 Feb;13(2):127-37. doi: 10.1038/nmeth.3733. PMID: 26820547; PMCID: PMC4922638.

36. Brezgin S, Kostyusheva A, Kostyushev D, Chulanov V. Dead Cas Systems: Types, Principles, and Applications. Int J Mol Sci. 2019 Nov 30;20(23):6041. doi: 10.3390/ijms20236041. PMID: 31801211; PMCID: PMC6929090.

37. Casas-Mollano JA, Zinselmeier MH, Erickson SE, Smanski MJ. CRISPR-Cas Activators for Engineering Gene Expression in Higher Eukaryotes. CRISPR J. 2020 Oct;3(5):350-364. doi: 10.1089/crispr.2020.0064. PMID: 33095045; PMCID: PMC7580621.

38. Larson MH, Gilbert LA, Wang X, Lim WA, Weissman JS, Qi LS. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc. 2013 Nov;8(11):2180-96. doi: 10.1038/nprot.2013.132. Epub 2013 Oct 17. PMID: 24136345; PMCID: PMC3922765.

39. Anton T, Bultmann S. Site-specific recruitment of epigenetic factors with a modular CRISPR/Cas system. Nucleus. 2017 May 4;8(3):279-286. doi:

10.1080/19491034.2017.1292194. Epub 2017 Feb 23. PMID: 28448738; PMCID: PMC5499900.

40. Enríquez P. CRISPR-Mediated Epigenome Editing. Yale J Biol Med. 2016 Dec 23;89(4):471-486. PMID: 28018139; PMCID: PMC5168826.

41. Ding L, Schmitt LT, Brux M, Sürün D, Augsburg M, Lansing F, Mircetic J, Theis M, Buchholz F. DNA methylation-independent long-term epigenetic silencing with dCRISPR/Cas9 fusion proteins. Life Sci Alliance. 2022 Mar 14;5(6):e202101321. doi: 10.26508/lsa.202101321. PMID: 35288457; PMCID: PMC8921183.

42. Lara Marrone, Paolo M. Marchi & Mimoun Azzouz (2022) Circumventing the packaging limit of AAV-mediated gene replacement therapy for neurological disorders, Expert Opinion on Biological Therapy, 22:9, 1163-1176, DOI: 10.1080/14712598.2022.2012148

43. Matthew Nitzahn, Gabriella Allegri, Suhail Khoja, Brian Truong, Georgios Makris, Johannes Häberle, Gerald S. Lipshutz, Split AAV-Mediated Gene Therapy Restores Ureagenesis in a Murine Model of Carbamoyl Phosphate Synthetase 1 Deficiency, Molecular Therapy, Volume 28, Issue 7, 2020, Pages 1717-1730, ISSN 1525-0016, https://doi.org/10.1016/j.ymthe.2020.04.011.

44. Multiplex Genome Engineering Using CRISPR/Cas Systems. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F. Science. 2013 Jan 3. 10.1126/science.1231143 PubMed 23287718

45. Francesco Aulicino and others, Highly efficient CRISPR-mediated large DNA docking and multiplexed prime editing using a single baculovirus, Nucleic Acids Research, Volume 50, Issue 13, 22 July 2022, Pages 7783–7799, https://doi.org/10.1093/nar/gkac587

46. Wang D, Tai PWL, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. Nat Rev Drug Discov. 2019 May;18(5):358-378. doi: 10.1038/s41573-019-0012-9. PMID: 30710128; PMCID: PMC6927556.

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor, Dr. Patrick Pausch, for the opportunity to work in his laboratory, as well as for his guidance and countless consultations. His expertise, feedback, and constant availability have been invaluable in helping my research.

I would also like to give my thanks to the members of the laboratory of Dr Pausch, Rimvydė Čepaitė, Monika Jasnauskaitė and Rokas Grigaitis for their valuable contributions to my academic development. Their patience and assistance with mastering new methods as well as their help with improvig my presentation skills were instrumental in shaping my thesis.