Improvement of rAAV production by analysis of capsid modification, producer cell line generation and a novel affinity chromatography

Dissertation

zur Erlangung der naturwissenschaftlichen Doktorwürde

Technische Fakultät

Universität Bielefeld

vorgelegt von

Julian Teschner

geboren in Bielefeld

2019

Die vorliegende Arbeit wurde in der Zeit von Juli 2014 bis September 2019 in der Arbeitsgruppe Zelluläre und Molekulare Biotechnologie an der Technischen Fakultät der Universität Bielefeld unter der Leitung von Prof. Dr. Kristian M. Müller durchgeführt.

1. Gutachter: Prof. Dr. Kristian M. Müller
   Arbeitsgruppe Zelluläre und Molekulare Biotechnologie,
   Technische Fakultät, Universität Bielefeld

2. Gutachter: Prof. Dr. Thomas Noll
   Arbeitsgruppe Zellkulturtechnik,
   Technische Fakultät, Universität Bielefeld

~ Für meine Familie ~

„Wissenschaftliche Forschung läuft immer darauf hinaus, dass es plötzlich mehrere Probleme gibt, wo es früher ein einziges gegeben hat."

Norman Mailer, amerikanischer Schriftsteller, 1923 - 2007

„Zwei Dinge sind zu unserer Arbeit nötig. Unermüdliche Ausdauer und die Bereitschaft, etwas, in das man viel Zeit und Arbeit gesteckt hat, wieder wegzuwerfen."

Albert Einstein, deutscher Physiker, 1879 - 1955

„Das wahre Geheimnis des Erfolgs ist die Begeisterung."

Walter Percy Chrysler, amerikanischer Automobilpionier, 1875 - 1940

„Irgendwo tief in mir bin ich ein Kind geblieben. Erst dann, wenn ich's nicht mehr spüren kann, weiß ich es ist für mich zu spät, zu spät, zu spät."

Peter Maffay, deutscher Sänger, * 1949

„Viel zu lernen du noch hast."

Yoda, dagobahnischer Jedi Großmeister, 896 VSY - 4 NSY

Danksagung

Bevor nun der Hauptteil dieser Ausarbeitung beginnt, halte ich es für angebracht, zunächst auf die vergangenen gut fünf Jahre zurückzublicken und dabei denjenigen zu danken, die am Gelingen dieser Arbeit maßgeblich beteiligt waren. Wissenschaftliche Arbeit ist immer eine Form von Teamwork, bei der jeder von den Erfahrungen und Kompetenzen des Anderen profitieren und daraus lernen kann. Dies zeigt was für einen unschätzbaren Wert ein gutes und funktionierendes Team hat. Die Liste an Leuten, die in den letzten Jahren Teil dieses Teams geworden sind, ist erfreulicherweise sehr lang geworden.

Zuerst möchte ich die Betreuung durch Prof. Dr. Kristian M. Müller ganz besonders hervorheben. Ich fühlte mich immer gut beraten und habe davon sehr profitiert, darüber hinaus blieb mir genug Freiraum, um meine Ideen zu verwirklichen. Unsere Arbeitsgruppe „Zelluläre und Molekulare Biotechnologie“ ist in den letzten Jahren stetig gewachsen und viele Leute haben dort gearbeitet. Unser technischer Assistent Philipp Borchert half mir bei zahlreichen Experimenten und hat mir somit meine Arbeit deutlich erleichtert. Meine Frau Kathrin Teschner und auch der Rest meiner Familie standen mir immer unablässig bei, haben mich unterstützt und waren immer für mich da. Der wissenschaftliche Austausch und freundliche Umgang mit allen Kollegen hat immer für eine angenehme Arbeitsatmosphäre gesorgt. Dafür möchte ich mich ganz besonders bei Rebecca Feiner, Marco Radukic und Dinh To Le bedanken. Zusätzlich möchte ich mich bei allen beteiligten Masteranden, Bacheloranden und Projektanden für ihre Mitarbeit bedanken. Auch über unsere Arbeitsgruppe hinaus wurde ich von vielen Leuten unterstützt, daher möchte ich mich bei allen Mitarbeitern der Arbeitsgruppen für Zellkulturtechnik, der Fermentationstechnik und der Biochemie herzlich bedanken. Besonders hervorheben möchte ich Prof. Dr. Thomas Noll für die Begutachtung dieser Arbeit, Dominik Cholewa für die Mittelbauvertretung und Prof. Dr. Karl Friehs für den Vorsitz bei meiner Disputation. Darüber hinaus möchte ich mich bei Dr. Alexander Sczyrba für die bioinformatische Unterstützung, bei Sebastian Hanke von der Universität Potsdam für die methodischen Einarbeitung und bei Jessica Schlicht für die Korrektur diese Arbeit, bedanken. Einige Materialien wurden mir in Kooperationen mit der Xell AG und Thermo Fisher Scientific zur Verfügung gestellt. Letztlich möchte ich mich beim Bielefelder Nachwuchsfond für die finanzielle Unterstützung in der Fertigstellungsphase dieser Arbeit bedanken.

Danke!

Abbreviations

Beside the SI base and derived units, the following abbreviations were used in this work:

(v/v)	(volume/volume)
(w/v)	(weight/volume)
×g	multiple of the earth gravity
3'UTR	three prime untranslated region
Å	Ångström
A_{260}	absorption at 260 nm
A_{280}	absorption at 280 nm
AAP	assembly activating protein
AAV	adeno-associated virus
AAV2	adeno-associated virus serotype 2
AAVR	Adeno-associated virus receptor
AAVS1	Adeno-associated virus integration site 1 (AAV safe-harbor locus)
ABTS	2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid)
ad.	adherent
AFM	Atomic force microscopy
AG	Arbeitsgruppe (engl. working group)
approx.	approximate
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BEV	baculovirus expression vectors
BHK	baby hamster kidney (cell line)
bla (gene)	β-lactamase (gene)
bp	base pairs
BSA	bovine serum albumin
CAP	catabolite activator protein
CAT (gene)	Chloramphenicol acetyltransferase (gene)
CBh (promoter)	CMV early enhancer fused to modified chicken β-actin promoter
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate
CHO K1	Chinese hamster ovary K1 (cell line)
CMV	Cytomegalovirus
CpG	Two-base sequence motif cytosine and guanine
CRISPR/Cas9	Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated protein 9
Ctl.	Control
CV	column volume
Da	Dalton
DARPin	designed ankyrin repeat proteins
DMSO	Dimethyl sulfoxide

DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DSC	differential scanning calorimetry
DSF	differential scanning fluorimetry
DTT	Dithiothreitol
e.g.	*exempli gratia* (for example)
EDTA	Ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EMA	Europäische Arzneimittelagentur
engl.	english
Fc	fragment crystallizable
FCS	Fetal Calf Serum
FDA	(U.S.) Food and Drug Administration
FRT	Flippase Recognition Target (flipase recombinase recognition site)
FSC	forward scatter diode
FT	flowthrough
gc	genomic copies
GC (content)	guanine-cytosine content
GFP	green fluorescent protein
GOI	gene of interest
HBS (buffer)	HEPES buffered saline (buffer)
HBSS (buffer)	Hanks bufferd salt solution (buffer)
HEK293	human embryonic kidney 293 (cell line)
HeLa	Henrietta Lacks (cell line)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hGH	human growth hormone
His_6	polyhistidine-tag
HRP	horseradish peroxidase
HSPG	heparan sulfate proteoglycan
HSV	herpes simplex virus
HT-1080	A human cell line (cell line)
Hypert. SN.	Hypertonic supernatant
Hypot. SN.	Hypotonic supernatant
iGEM	international Genetically Engineered Machine
IgG-Fc	Immunoglobulin G-fragment crystallizable
IMAC	immobilized metal ion affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITR	inverted terminal repeats
kan^R (gene)	Kanamycin resistance (gene)
kb	Kilo bases
kDa	Kilo Dalton
lac	(part of the lac operon) lactose (operon)
lacI	lac repressor

LB (liquid/solid) medium	lysogeny broth (liquid/solid) medium
Len.	Length
LLC	limited liability company
Ltd.	Limited
mAU	milli absorbance units
MBP	Maltose-binding protein
MCS	Multiple cloning site
miRNA	micro RNA
mRNA	messenger RNA
Ni-NTA	nickel (-charged) nitrilotriacetic acid (resin)
NMWL	nominal molecular weight limit
nt	nucleotides
NTA	nitrilotriacetic acid
OD_{600}	optical density at 600 nm
ORF	Open reading frame
ori	origin of replication
PAM	protospacer adjacent motif
PBS (buffer)	phosphate buffer saline (buffer)
PCR	Polymerase chain reaction
PDB ID	Protein Data Bank Identity
PEI_{max}	Polyethylenimine (max)
PKD2	polycystic kidney disease 2
P_{lac}	lac promoter
PMSF	phenylmethylsulfonyl fluoride
pZMB	Plasmid from the working group of Cellular and Molecular Biotechnology
qPCR	quantitative Polymerase chain reaction
rAAV	recombinant adeno-associated virus
rAAV2	recombinant adeno-associated virus serotype 2
RAC	regenerated amorphous cellulose
RBE	minimal Rep binding element
RFC 10	assembly standard 10
RFC 25	assembly standard 25
RFP	red fluorescent protein
rHSV	recombinant herpes simplex virus
RNA	ribonucleic acid
rpm	revolutions per minute
scFv	single chain variable fragment
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
Sf9	*Spodoptera frugiperda* 9 (cell line)
SI	International System of Units (from French)
SLiCE	Seamless Ligation Cloning Extract
SMN1	survival motor neuron 1

SOC	Super Optimal broth with Catabolite repression
SSC	side scatter
ssDNA	single-stranded DNA
susp.	suspension
T7	Phage T7
T_A	annealing temperature
TAE (buffer)	Tris-acetate-EDTA (buffer)
T_d	disintegration temperature
T_{digest}	Digestion temperature
t_E	elongation time
TEMED	Tetramethylethylenediamine
T_m	melting temperature
TRS	terminal resolution site
U	unit
ULR	ultra low range
UV	Ultraviolet
V_H	heavy chain variable domain
V_L	light chain variable domain
WCE	whole cell extract
wt	wild type
ZMB	Zelluläre und Molekulare Biotechnologie

Publications

Articles:

Feiner RC, Teschner J, Teschner KE, Radukic MT, Baumann T, Hagen S, Hannappel Y, Biere N, Anselmetti D, Arndt KM, Müller KM, "rAAV engineering for capsid-protein enzyme insertions and mosaicism reveals resilience to mutational, structural and thermal perturbations", submitted to "International Journal of Molecular Sciences".

Feiner RC, Teschner KE, Teschner J, Müller KM, "HEK293-KARE1, a cell line with stably integrated adenovirus helper sequences simplifies rAAV production", submitted to "BMC Biotechnology".

Feiner RC, Teschner K, Schierbaum I, Teschner J, Müller KM," AAV production in suspension: evaluation of different cell culture media and scale-up potential", *BMC Proceedings* 12 (Suppl 1):P-349, (2018) doi: 10.1186/s12919-018-0097-x.

Poster:

Teschner J, Feiner RC, Teschner KE, Radukic MT, Hertle Y, Biere N, Anselmetti D, Müller KM, "rAAV2 capsid protein modification, expression and stability", DECHEMA, Frankfurt am Main, "Gene Therapy – Ready for the Market?", 30.-31. January 2019

Feiner RC, Teschner K, Teschner J, Scheiner O, Müller KM, "Recombinant adeno-associated virus for tumor therapy – capsid and genetic engineering", 4th Global Synthetic Biology & Gene Editing, London, 04.-05. December 2017.

Feiner RC, Teschner K, Schierbaum I, Teschner J, Müller KM, "AAV production in suspension: Evaluation of different cell culture media and scale-up potential", 25th ESACT Meeting: Cell technologies for innovative therapies, Lausanne, 14.-17. May 2017.

Feiner RC, Schlicht K, Teschner J, Arndt KM, Müller KM, "Recombinant Adeno-associated virus (rAAV) for tumor therapy: engineering of capsid and genetic modifications", 67. Mosbacher Kolloquium – "Protein Design: From First Principles to Biomedical Applications", Mosbach, 30. March - 02. April 2016.

Table of Contents

1 Abstract

Recombinant adeno-associated viruses (rAAV) represent the basis of modern gene therapy. They can be modified extensively at the genomic and capsid level enabling personalized therapies. The high potential of this technology and the great successes that have been achieved in this field require methods that allow for the development of modified rAAV variants, large production scales of viral vectors and finally the generation of high purity products for medical therapy. This thesis dealt with the expansion of a plasmid system for recombinant adeno-associated virus serotype 2 (rAAV2) production, the generation of rAAV producer cell lines and the development of a novel rAAV2 purification method based on affinity chromatography.

For a deeper understanding of the existing plasmid-based expression system, a stability assay was developed which aims at the biologically relevant release of rAAV DNA under increasing temperature exposure. Using this technique, it was discovered that, the thermal stability of capsid-modified rAAV2 particles bearing glycine-serine amino acid linker insertions at position 587, decreases with increasing linker length. In addition, a rAAV2 variant was generated which presents 60 active β-lactamase enzymes on its capsid. This variant had a similar thermal stability as the wild type variant and proved the enormous modifiability of the rAAV2 capsid.

Within the scope of cell line generation experiments an attempt was made to insert fluorescent reporter genes into suitable genomic target sites for the later insertion of the adenoviral E1a/b genes into the CHO K1 (chinese hamster ovary K1) genome using a CRISPR/Cas9 mediated rational approach. These experiments did not lead to the desired success, but it could be demonstrated that it is possible to produce rAAV2 particles using CHO K1 cells and that the production can be improved by the addition of the E1a/b genes. Later, monoclonal HEK293 cell lines could be generated, providing the helper genes required for rAAV production through random genomic integration, thus reducing the cost and effort of rAAV production.

In order to improve rAAV2 production, an attempt was made to generate a novel affinity chromatography based on a single-chain Fv (scFv) antibody designed using the murine A20 antibody structure. The construct was expressed in the periplasm of *E. coli.* Silica material was used as base material for the chromatography and the protein was immobilized via a fused silica tag. The principal functionality of the construct was verified. However, it turned out that silica is not suitable for rAAV purification due to its non-specific binding properties. A second construct, in which the scFv was fused to a human IgG-Fc part was produced in HEK293 suspension cells and has since been used for AAV capsid ELISAs. In the search for another suitable affinity ligand that can be expressed bacterially, the natural AAV receptor (AAVR) came into focus. A fusion protein was designed, which

consists of the PKD2 domain (polycystic kidney disease 2) from the AAVR in combination with a cellulose binding domain. The suitability of this construct in the context of rAAV2 affinity chromatography was proven and the production could be significantly improved by the genetic fusion with a maltose binding protein. In further studies, a fast, cost-effective and very efficient purification process for rAAV2 particles based on PKD2 was developed.

Ultimately, this thesis led to an improvement of rAAV production and modification, resulting in a contribution to rAAV-based gene therapy.

2 Zusammenfassung

Rekombinante adeno-assoziierte Viren (rAAV) bilden die Grundlage der modernen Gentherapie. Sie können auf genomischer und Kapsidebene umfassend modifiziert werden, was personalisierte Therapien ermöglicht. Das hohe Potenzial dieser Technologie und die großen Erfolge, die auf diesem Gebiet erzielt wurden, verlangen nach Methoden, die die Entwicklung modifizierter rAAV-Varianten, große Produktionsmaßstäbe viraler Vektoren und schließlich die Herstellung von hochreinen Produkten für die medizinische Therapie ermöglichen. Diese Arbeit beschäftigte sich mit der Erweiterung eines Plasmid-Systems für die rAAV2-Produktion, der Erzeugung von rAAV-Produktionszelllinien und der Entwicklung einer neuartigen Aufarbeitungsmethode für rAAV2 auf Affinitätschromatographie-basis.

Für ein tieferes Verständnis des bestehenden plasmidbasierten Expressions-systems wurde ein Stabilitätsassay entwickelt, der auf der biologisch relevanten Freisetzung von rAAV-DNA unter zunehmender Temperaturexposition aufbaut. Mit dieser Technik wurde herausgefunden, dass die thermische Stabilität von kapsid-modifizierten rAAV2-Partikeln, die Glycin-Serin-Aminosäure-Linker-Insertionen an Position 587 tragen, mit zunehmender Linkerlänge abnimmt. Zusätzlich wurde eine rAAV2-Variante generiert, die 60 aktive β-Lactamase Enzyme auf ihrem Kapsid präsentiert. Diese Variante wies eine vergleichbare thermische Stabilität auf wie die Wildtypvariante und bewies damit die enorme Modifizierbarkeit des rAAV2-Kapsids.

Im Rahmen der Experimente zur Zellliniengenerierung wurde versucht, Fluoreszenz-Reportergene an geeigneten genomischen Sequenzen für die spätere Insertion der adenoviralen E1a/b-Gene im CHO K1-Genom (Chinese hamster ovary K1) mit einem CRISPR/Cas9 vermittelten rationalen Ansatz einzubringen. Diese Experimente führten nicht zum gewünschten Erfolg, aber es konnte gezeigt werden, dass es möglich ist, rAAV2-Partikel mit Hilfe von CHO K1-Zellen zu produzieren und dass die Produktion durch den Zusatz der E1a/b-Gene verbessert werden kann. Später konnten monoklonale HEK293-Zelllinien durch genomische Zufallsintegration erzeugt werden, die Helferfunktionen aus dem so genannten pHelper-Plasmid tragen, was die Kosten und den Aufwand der rAAV-Produktion reduziert.

Um die rAAV2-Produktion zu verbessern, wurde versucht, eine neuartige Affinitäts-Chromatographie basierend auf einem single-chain Fv (scFv) Antikörper zu erzeugen, der auf der murinen A20 Antikörperstruktur basiert. Das Konstrukt wurde im Periplasma von *E. coli* exprimiert. Silica-Material wurde als Basismaterial für die Chromatographie verwendet und das Protein konnte über einen fusionierten Silica Tag immobilisiert werden. Die prinzipielle Funktionalität des Konstrukts konnte bestätigt werden. Es stellte sich jedoch heraus, dass Silica aufgrund seiner unspezifischen Bindungseigenschaften nicht für die rAAV-Reinigung geeignet ist. Ein zweites Konstrukt, bei dem der scFv mit einem

menschlichen IgG-Fc-Teil fusioniert wurde, wurde in HEK293-Suspensionszellen exprimiert und wird seitdem für AAV-Kapsid-ELISAs verwendet. Auf der Suche nach einem weiteren geeigneten Affinitätsliganden, der sich bakteriell exprimieren lässt, rückte der natürliche AAV Rezeptor (AAVR) in den Fokus. Ein weiteres Fusionsprotein wurde entwickelt, das aus der PKD2-Domäne (polycystic kidney disease 2) des AAVR in Kombination mit einer Cellulosebindedomäne besteht. Die Eignung dieses Konstrukts im Rahmen der rAAV2-Affinitätschromatographie wurde nachgewiesen und die Produktion konnte durch die genetische Fusion mit einem Maltosebindeprotein signifikant verbessert werden. In weiteren Studien wurde ein schneller, kostengünstiger und sehr effizienter Reinigungsprozess für rAAV2-Partikel auf Basis von PKD2 entwickelt.

Letztendlich führte diese Arbeit zu einer Verbesserung der rAAV-Produktion und Modifikation, was einen Beitrag zur rAAV-basierten Gentherapie lieferte.

3 Introduction

3.1 The biology of AAV

The adeno-associated virus (AAV), was first discovered in 1965 as contamination of simian adenovirus preparations [1]. It is a non-enveloped virus, which belongs to the genus of *Dependoparvovirus*, which in turn belongs to the family of *Parvoviridae* (see Figure 1A). It has a single-stranded DNA genome with a size of about 4.7 kb. The genome is mainly composed of two open reading frames (ORF) for the rep and cap genes, which are flanked by inverted terminal repeats (ITRs) [2]. The rep genes encodes for four non-structural proteins (Rep78, Rep68, Rep52 and Rep40) that have been reported to be involved in viral replication, packaging, and genomic integration [2, 3]. Their expression is driven, in mutual regulatory interdependence, by the two viral promoters p5 and p19 [4, 5]. The cap gene codes for the three structural proteins (VP1, VP2 and VP3), which assemble to the viral capsid. The capsid consists of 60 proteins, whereby the VP1, VP2 and VP3 proteins occur in a 1:1:10 ratio. The capsid proteins have the identical C-terminus but different N-termini, which is caused by alternative splicing and leaky scanning [6]. The expression of the VP protein is driven by the p40 promoter [2]. The structure of the AAV serotype 2 capsid has been determined to a 3 Å resolution [7]. The assembled capsid with its 20 nm diameter, serves as the delivery vehicle, harboring the viral DNA. In addition, an alternative ORF encoding the assembly activating protein (AAP) is located within the ORF of the cap genes. As its name suggests, this protein is involved in the assembly of capsid proteins [8, 9]. In 2018 the dependence of some AAV serotypes on AAP was demonstrated and by this led to a better understanding of viral evolution [10]. In addition, the same ORF contains the sequence of the so-called x gene, which probably plays a role in viral DNA replication [11]. Further protein coding sequences have been found whose functions are currently unknown (see Figure 1B) [12].

Recently Gao *et al.* obtained more than 120 novel primate AAVs [13]. However, AAV serotype 2 is the currently best characterized and studied serotype and therefore all investigations of this thesis are based on it. It shows a natural tropism towards skeletal muscles, neurons, vascular smooth muscle cells and hepatocytes [14–17]. Although serotype 2 is able to transduce neuronal cells efficiently, *in vivo* experiments showed that it is unable to cross the blood-brain barrier [18]. Numerous studies have discovered the heparan sulfate proteoglycane (HSPG) receptor as the primary receptor for the interaction of virus and target cell. The interaction can be applied to the residues R484, R487, K532, R585 and R588, whereby already two mutations (R585A and R588A) in cell culture experiments led to a significant reduction of infection ability [19–22]. Other coreceptors include the fibroblast growth factor 1, the hepatocyte growth factor, the integrins α5β1 and αVβ5, and the 37/67 kDa laminin receptor [23–27]. Pillay *et al*. identified another

receptor by the generation of a library of mutagenized haploid HAP1 cells using retroviral gene-trap vectors. This library consisted of almost all non-essential genes of the human genome. Through the infection of this library with rAAV2, the subsequent selection of all uninfected cells and their genomic characterization, 46 essential genes for rAAV2 infection were identified. A very prominent hit was the gene KIAA0319L, which codes for a type I transmembrane protein [28]. At this time, this gene was only known to be linked to dyslexia and that it is involved in neural migration and axon guidance [29, 30]. Since then, this receptor is known under the name AAV receptor. In the context of this work, the binding of rAAV2 to the AAVR could be narrowed down to the second polycystic kidney disease domain (PKD2) of this receptor and it was discovered that it is involved in the binding of many other serotypes [28].

Once the AAV has bound to the target cell, the viral vector is transported into the cell via receptor-mediated endocytosis, which ranges from the formation of clathrin-coated pits to vesicles [31]. Following internalization and trafficking to the nucleus, the virus triggers a slightly acidic environment which is sufficient to allow penetration into the cytosol. After this so-called endosomal escape, the AAVs accumulate perinuclearly and slowly pass through the nuclear pore complex into the nucleus. After uncoating and thus the release of the genomic ssDNA of the AAV particles, the synthesis of the second DNA strand takes place. The virus can then enter its replication cycle, also known as the lytic cycle, in the presence of an helper virus [15].

Suitable helper virus are the Adenovirus, Herpes simplex virus, Vaccinia virus or human papillomaviruses [1, 32–36]. In case of coinfection with the adenovirus, the following gene information supports the formation of AAV2 particles: The adenoviral E1A and E2A gene products are involved in transcriptional activation [37–43]. The gene products of the genes E2A as well as the product of the sixth ORF of E4 (E4(orf6)), the 55 kDa product of the E1B gene and the VAI RNA are related to the AAV2 RNA transport, viral stability and translation [43–49]. In addition, E2A and E4(orf6) are involved in DNA replication [44, 50–54]. At this time, the exact mechanisms involved in nuclear translocation and RNA splicing are not fully understood and cannot be assigned to any helper gene [55, 56].

If the host cell is not infected with a helper virus, the AAV genomes can persist as episomes or integrate into host cell chromosomal DNA, this process is also called the lysogenic cycle [57, 58]. The genomic integration takes place at a defined position in the so-called AAV safe-harbor locus (AAVS1) in human chromosome 19 [57, 59–61].

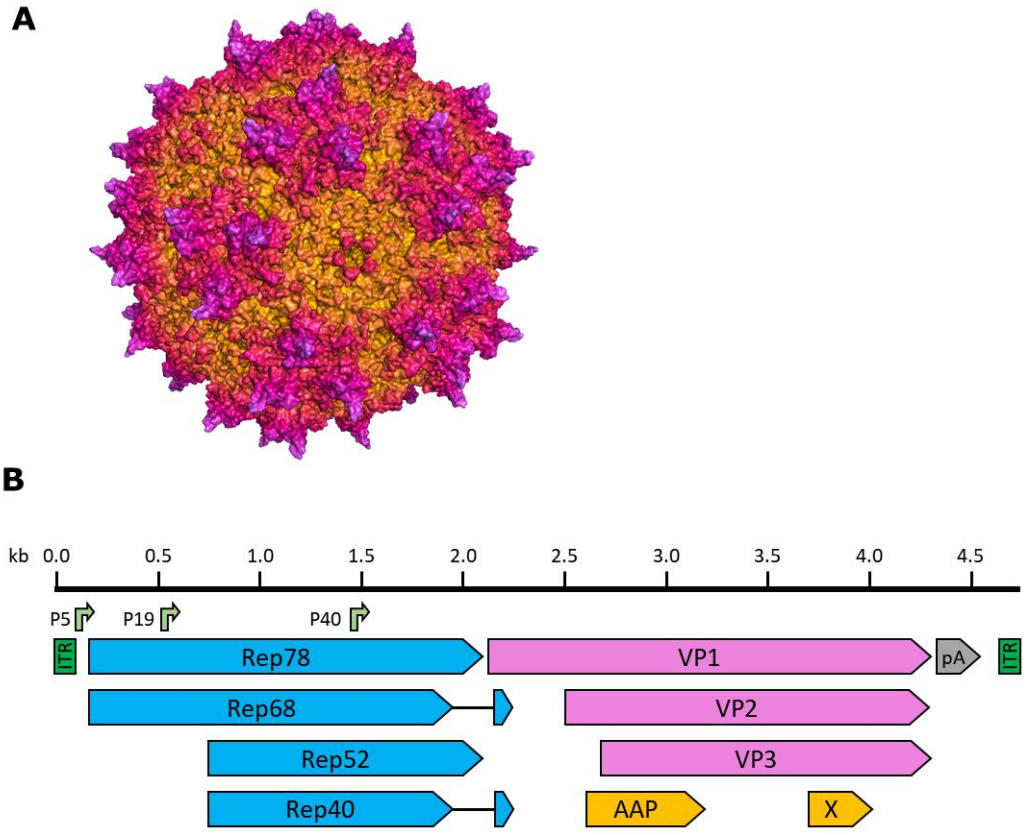


Figure 1: Surface and genomic structure of AAV2. A: Surface structure of AAV2 based on PDB ID: 1LP3. B: Genomic structure of AAV2. Genes of the rep ORF are indicated in blue and genes of the cap OFR are indicated in pink. Further gens like the AAP and X genes are shown in yellow. Promoters and poly-adenylation site are marked by green arrows or a gray box. ITR regions are indicated by green boxes.

3.2 **rAAV in context of gene therapy**

According to the "Bundesministerium für Gesundheit", cancer is the second most frequent cause of death in Germany after cardiovascular diseases. The most current data from the year 2013 for all types of cancer except non-melanotic skin cancer show that a total of 482'470 people (229'920 women and 252'550 men) were newly diagnosed with cancer that year. The total number of cancer-related deaths in the same year was 223'093 (101'779 women and 121'314 men) [62]. Existing therapeutic approaches such as surgery, thermotherapy, chemotherapy and radiotherapy often lead to serious side effects such as cytotoxicity to normal cells and strong immune responses. However, some types of cancer hardly respond to these therapies [63]. In addition to the frequent cancer types, there is also a long list of hereditary diseases, for which no meaningful therapeutic options exist due to their rare occurrence and genetic causes. Gene therapy might be a novel approach to cure all these diseases. Gene therapy means the therapeutic delivery of nucleic acid into a patient's cells as a drug to treat disease [64].

Essentially, gene therapy can be divided into three steps. The first is the construction of a suitable gene carrying vector. The second step describes the transfer of the gene into the target cells using this vector. And under the last step the expression of the corresponding gene product with the aim to cure the disease is meant. The first step is therefore of central importance. The gene carrying vectors can generally be divided into two categories: the non-viral and the viral vectors.

Non-viral vectors are naked plasmids, microbubbles, nanoparticles, liposomes, and various polymers. They are often introduced into the target cells by chemical or biophysical methods [65–71]. The advantages of high safety, low costs with low immunogenicity and the possibility of enabling large gene insertions have some disadvantages. These include the undirected and usually inefficient transfection and the resulting low expression of the target genes [72].

The most frequently used viral vectors include adenoviruses, adeno-associated viruses, retroviruses and lentiviruses. They enable a very effective and, in many cases, highly specific transduction as well as a strong, long-term expression of the target genes [73]. These advantages come from the fact that in the course of their evolution, viruses have been geared towards efficiently introducing their genetic material into certain target cells and proliferating it there. Viruses are therefore ideally suited as gene therapy vectors. However, they do have some disadvantages as well including high immune rejection, possible tumorigenicity, uncertain insertional mutagenesis, and limited sizes for gene insertions [74].

In addition to the type of vectors, gene therapy also distinguishes between two general forms of application: *ex vivo* gene therapy and *in vivo* gene therapy (see also Figure 2).

In *ex vivo* gene therapy, donor cells are first taken (usually from the patient himself) and then treated outside the body in cell culture using gene therapeutic methods. Following cell proliferation, these modified cells are reintroduced to the patient. This form of application is particularly suitable for the treatment of local diseases [64].

In *in vivo* gene therapy, the required gene vectors are usually injected directly to the patient's vascular system or, in the case of local tumors, the vector can also be injected directly into the tumor. This form of application offers not only the advantage of less effort but also that in case of a viral cancer therapy also metastasizing tumors can be cured simultaneously [64].

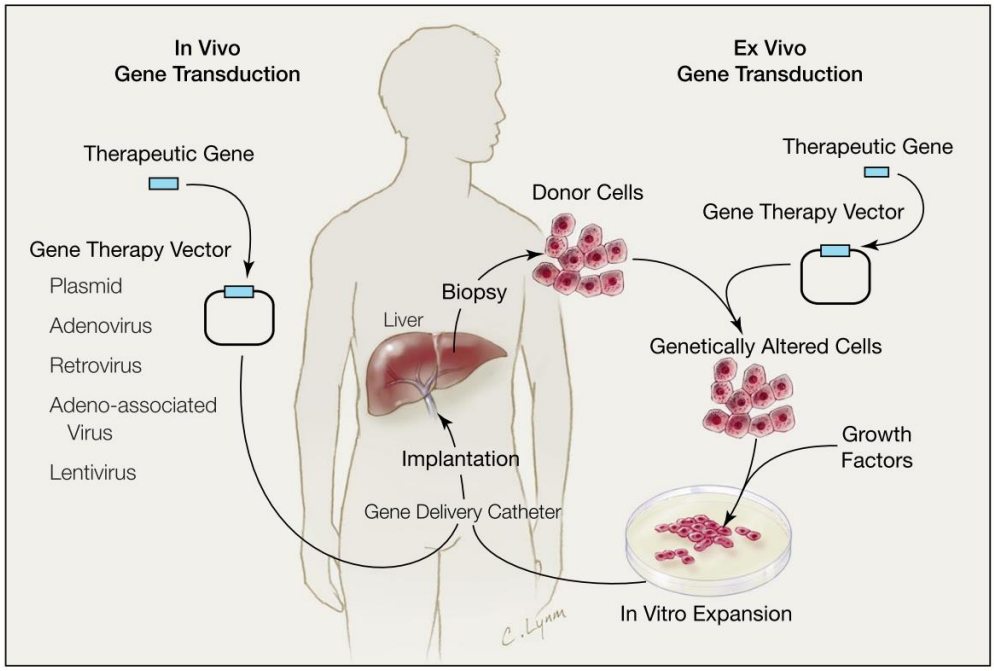


Figure 2: In vivo and ex vivo gene therapy approaches (from [64]*).*

The undesirable properties of some viral vectors, e.g. leading to immune reactions or cancer, have made clinical use much more difficult and led to a limitation to certain applications, such as vaccines and oncolytic strategies [75]. AAV based gene delivery systems have significant advantages due to their high safety profile. They are safer due to their lack of pathogenicity, the variety of different serotypes allows a wide range of target cells, genomic integration can lead to long-term expression of the introduced genes and they are able to transduce both dividing and non-dividing cells [15]. In addition, several studies have been published in recent years that deal with the modification of the capsid shell. This indicates the strong modifiability of the AAV and will open up new target possibilities and new applications in the future [76–81].

Nevertheless, an important point in the design of rAAV vectors for gene therapy is the packaging size of the expression cassette which is very restricted. In general, the size of this construct, including the viral ITR sequences, should not exceed approx. 5 kb [82]. Attempts have already been made to package larger expression cassettes in rAAV, resulting in a significant reduction of particle yield or truncation of the expression cassette [83]. The coding sequence of a full-length dystrophin e.g. exceeds this size, in this context strategies are meaningful in which overlapping sequences are transduced with the help of two vectors [84]. Another important point in this context is that the DNA transduced by the rAAV particles is initially single-stranded and must first be converted into its double-stranded form in the nucleus, which is a limiting step in the process of transgene expression. To avoid this, self-complementary DNA strands can also be packaged in rAAV, which however reduces the packaging capacity to about 3.3 kb [85, 86].

In addition to the gene to be expressed itself, regulatory sequences could also be introduced. The choice of a suitable promoter can, for example, influence the tissue/cell specificity as well as the choice of AAV serotype. Thus, the expression in certain tissues can be upregulated, or the expression in unwanted locations can be suppressed, e.g. by choosing the creatin kinase promoter high expression levels in skeletal muscle cells was reached, whereas the α-myosin heavy chain promoter is strongly suppressed in cardiac muscle cells [87, 88]. In addition, other promoters such as the survivin, the cyclooxygenase 2 and also the CXCR4 promoter, are known to have a high expression level in many tumor cells, which makes them interesting for possible tumor therapies [89–93].

Next to promoter-regulated expression, the selection of a suitable terminator sequence, the inclusion of post-transcriptional regulatory elements as well as messenger RNA (mRNA) stability elements and the additional use of microRNA (miRNA) target sequences can increase the specificity of gene expression [94]. The introduction of miRNA target sequences into the three prime untranslated region (3'UTR) of an AAV-delivered gene, made it available for miRNA-122-driven suppression in the liver [95]. The use of a let-7 miRNA target sequence could also generate tumor specificity, as reports have shown that this miRNA is downregulated in certain tumor cells while it is more present in healthy cells [96].

The tissue/cell specificity can be further increased by the selection of suitable serotypes. Recently, many new serotypes from human and primate tissue have been identified, that will be used for cloning to generate novel serotypes [97]. These recombinant techniques include capsid shuffling, directed evolution, and random peptide library insertions, which are used to generate AAVs with new properties from existing variants [77, 78, 98]. This way, it was already possible to generate novel AAVs that are able to transduce specific retinal cells or cells of the central nervous system [79, 98, 99]. Not only the specificity of transduction can thus be increased, mutations have also been discovered resulting in rAAVs which were less sensitive to neutralizing antibodies, and protected against proteolytic degradation and by this increased their transduction efficiency [100–106].

The mentioned methods pursue modification variants, which are mostly based on random events, but other approaches exist, where rational designs led to a change of the target cell by the insertion of larger binding molecules into the capsid. For example, DARPins (designed ankyrin repeat proteins) or fragment of protein A were presented on the capsid surface through N-terminal addition or intramolecular loop integration [76, 80, 81, 107].

Due to these excellent properties, rAAV plays an important role in gene therapy clinical studies worldwide. From currently 2930 studies, a total of 238 (8.1 %) relate exclusively to adeno-associated viruses (update December 2018) [108]. In 1984, the first AAV-based vector for an *in vivo* gene therapy for the treatment of cystric fibrosis was developed [109].

In 2012, the world's first gene therapy was approved by the "Europäische Arzneimittelagentur" (EMA). The drug named Glybera, which is based on AAV serotype 1, is used for the treatment of lipoprotein lipase deficiency. Between 27 and 60 injections, depending on patients' weight, with $1.5 \cdot 10^{12}$ vg each (in 500 µl) are injected in thigh muscle. Glybera was removed from the market again in October 2017 due to its high costs [110]. Luxturna is an AAV2 vector-based gene therapy indicated for the treatment of patients with confirmed biallelic RPE65 mutation-associated retinal dystrophy. The treatment is performed by a single subretinal injection of a total of $1.5 \cdot 10^{11}$ vg in 300 µl suspension [111]. In 2019, the drug Zolgensma was approved by the FDA ((U.S.) Food and Drug Administration) for the treatment of pediatric patients (less than 2 years) with spinal muscular atrophy with bi-allelic mutations in the survival motor neuron 1 (SMN1) gene. The drug is based on AAV serotype 9, which contains a double-stranded expression cassette for the SMN1 gene [112]. Zolgensma is currently the most expensive drug of all time, with costs of over 2 million dollars per single dose.

3.3 **Production of rAAV**

All currently available production methods for rAAVs are still expensive and often complex. The most widely used method for rAAV production is based on the simultaneous transfection of HEK293 cells with two to four (usually three) plasmids (see Figure 3) [76, 113, 114]. The necessary genes are divided on the plasmids as follows. On the so-called GOI-ITR plasmid the information for the expression of the gene of interest (GOI) is coded. This sequence is flanked by the ITR sequences so that it can later be packaged in the viral capsid. The Rep and Cap genes are located on a second plasmid, which later define the serotype or codes for a modified capsid variant. The Rep and Cap genes does not necessarily originate from the same serotype, it is common to combine the Rep genes of serotype 2 with the Cap genes of other serotypes [113]. The separation of the Rep/Cap gene from the DNA to be packed in the capsid on two plasmids, in this type of production, is intended to prevent the production of "wild type" AAV variants [115]. A third plasmid contains the adenoviral helper sequences, which are not already integrated in the genome of the HEK293 cells (E1a and E1b). Also, helper sequences of the herpes simplex virus could be used for production [116]. In some dual plasmid systems, the helper functions are located on one of the other two plasmids [113]. A system based on four plasmids is used for the production of mosaic rAAV variants, in which the capsids of the produced rAAVs are composed of differently modified VP proteins. In this case, the additional capsid gene variant is coded on the fourth plasmid [76, 114]. In the following some of the most common rAAV production platforms are outlined, an overview can be found in Table 1.

In many laboratories, HEK293 cells are transfected with these plasmids in adherent growing cultures. Due to the high space requirements of this form of cultivation,

applications have been developed based on HEK293 cells adapted to suspension culture. This allows a higher cell density to be achieved, which ultimately leads to a higher yield. However, transient transfection of suspension cells is somewhat more difficult due to the media composition. In suspension culture, an average of 10^5 gc/cell can be achieved, resulting in a final titer of 10^{14} gc/l [117]. These production quantities are usually sufficient to guarantee early clinical test phases, but the scalability of this method is not given due to the high demand for plasmid DNA and the related high costs [118–122]. More recent approaches include the integration of some required sequences into the genome of the production cell line to save costs, e.g. the helper functions were genomically integrated into HEK293 cells, reducing the cost of providing plasmid DNA by approximately 50% by using the resulting HEK293 KARE cell line [123]. A disadvantage of the transient delivery of the Rep/Cap gene is the heterogeneity of the resulting particles, of which a considerable percentage does not carry the transgene [118, 119].

Figure 3: Triple Transfection of HEK293 cells for rAAV production.

The cost reduction through the genomic integration of the genes required for production was also considered in the following approach. The genes Rep and Cap as well as the transgene sequences flanked by the ITRs were genomically integrated into HeLa S3 cells, which offer the possibility to be cultivated in suspension. Since HeLa cells do not carry any viral helper functions, the Rep proteins, which are toxic for mammalian and insect cells, cannot be expressed [124, 125]. The rAAV production is induced by the infection of the cells with wild type adenoviruses. This production platform offers the possibility to produce rAAVs in large batches, even though the yield per cell is comparatively low. However, the stable integration of all rAAV-specific sequences means that a new production cell line must be generated for each new project. In addition, the helper viruses must be removed as part of the downstream process [118, 119, 126].

The system that currently offers the highest rAAV yields (up to $5{\cdot}10^5$ gc/cell) is based on the insect cell line Sf9. This cell line can be cultivated in large-scale and high-density suspension culture. The required sequences for rAAV production are introduced by infections of the cells with baculovirus expression vectors (BEV). The rAAV production system exists in different variants. The oldest variant contains the Rep and Cap gene as well as the ITR-flanked GOI on three different BEVs. In another variant, the ITR-flanked

gene of interest sequence is integrated into the Sf9 genome and production is induced by the infection with only two BEVs [117, 118]. In the latest version of this production system, the Rep and Cap genes are genomically integrated and controlled by a promoter which is induced by the subsequent baculovirus infection with only one BEV containing the ITR-flanked GOI [127, 128]. In addition to the excellent scalability, this production platform offers a high degree of biological safety, since neither the insect cells nor the baculoviruses show human pathogenic potential [117–119, 127, 128].

Another system, that operates without the generation of a stable cell line, is based on the hamster cell line BHK (baby hamster kidney). This cell line can also be cultivated in suspension and is infected with two replication-defective rHSVs (recombinant herpes simplex viruses), which carry separately the rep and cap genes as well as the ITR-flanked GOI sequence, to induce rAAV production. The helper functions are provided by the HSV (herpes simplex virus). This system also has good scalability and high production yields. HSV contaminations have to be removed in the downstream process [129, 130].

Table 1: Different rAAV production platforms (from [131]*)*

	Triple transfection (ad.)	**Triple transfection (susp.)**	**Baculovirus-infected producer cell line**	**Herpes virus coinfection**	**Adenovirus-infected producer cell line**
Cell line	HEK293 (adherent)	HEK293 (suspension)	Sf9 (suspension)	BHK (suspension)	HeLa S3 (suspension)
Rep/Cap	Plasmid	Plasmid	Genomically integrated	(first) rHSV	Genomically integrated
ITR-transgene	Plasmid	Plasmid	BEV	(second) rHSV	Genomically integrated
Helper genes	Plasmid	Plasmid	BEV (same as above)	rHSV (both)	Wild type adenovirus
Production system	Cell factory, Roller bottle	Wave reactor	200 l stirred tank reactor	10 l wave reactor	250 l stirred tank reactor
Efficiency of DNA delivery	good	not so good	very good	very good	very good
Scalability	poor	good	very good	very good	very good
Yield (vg/cell)	up to $3.5 \cdot 10^5$	up to $2.1 \cdot 10^5$	up to $5 \cdot 10^5$	up to $1 \cdot 10^5$	up to $5 \cdot 10^4$
Safety concerns	None	None	None	HSV contamination	Adenovirus contamination
Advantages	Quick production in small-scale Helper virus free	Quick production in small-scale Helper virus free	Efficient large-scale production Safety from insect cells and virus	Efficient large-scale production No stable cell line required	Efficient large-scale production Same helper virus for all productions

	Triple transfection (ad.)	Triple transfection (susp.)	Baculovirus-infected producer cell line	Herpes virus coinfection	Adenovirus-infected producer cell line
Challenges	Low scalability	Low scalability	Low BEV stability	Two HSV helper needed HSV sensitive to production conditions	Stable producer cell line for every project
References	[120, 121]	[120, 122]	[127, 128]	[129, 130]	[126]

ad.: adherent, susp.: suspension

3.4 **Purification of rAAV**

Following the rAAV production the purification takes place, which is necessary to generate a clean product of the produced rAAV particles for the intended application. The purpose of purification is not only to remove contaminating cell components from the rAAVs, such as proteins, lipids or nucleic acids, but also to separate two different rAAV variants from each other. The functional rAAVs and the rAAVs, which consist only of the capsid shell and do not contain any packaged transgene have to be separated from each other. Although the influence of these “empty” capsids on the effect of a potential therapy has not yet been clarified, their proportion should at least be kept very low or be accurately controlled [119].

The first step of purification consists of the separation of the cells carrying the largest proportion of rAAV from the remaining culture medium. Usually centrifugation or filtration processes are used. Tangential flow filtration could also be used to clarify the whole cell lysate and to reduce the working volume [132]. Depending on the production system, the proportion of rAAV released into the culture medium varies greatly. In case of a high proportion of rAAV in the medium, usually precipitations are carried out using e.g. PEG or ammonium sulphate with the aim of volume reduction for the subsequent processing steps. The following step is cell disruption, whereby processes based on the physical disruption of the cells, such as cell disruption by microfluidization or by several freezing and thawing cycles, are generally used. Microfluidization is an efficient process that involves passage of the cells through a small-diameter fluid path under high pressure, leading to cell disruption. Contrary to what is often assumed, cell disruption by microfluidization does not lead to an increased loss of rAAV particles compared to the apparently gentler freezing and thawing process [133]. During the following step, DNase or benzonase is added to digest the DNA and finally generate a clear lysate.

Thereafter, the main part of the purification takes place. Chromatography and ultracentrifugation methods dominate this process. The number of processing steps and their sequence is strongly dependent on the intended application of the final product [134].

Ultracentrifugation is one of the most commonly used purification processes, especially on a laboratory scale. Due to the usually very small sample quantities and relatively long

processing times, this method is not well applicable for larger commercial scales. However, it is still a good and comparatively inexpensive way to separate rAAV particles with and without packaged DNA due to their difference in density. Here discontinuous step gradients are used, which are composed of iodixanol or cesium chloride solutions with different concentrations [135]. Ultracentrifugation is used as the first step in purification during processing or as a subsequent step after affinity chromatography [119, 136].

Many other rAAV purification methods are based on different chromatography methods. Affinity chromatography methods are widely used, which utilize the fact that many AAV serotypes have a high affinity for different carbohydrates, which serve as receptors for their transduction. Thus, heparin columns are often used for the purification of AAV [134, 137]. Since these columns do not have a high specificity, increasing numbers of affinity chromatography columns based on AAV-specific binding proteins were marketed in recent years. In 1998 Grimm *et al*. were the first who coupled the AAV2 specific antibody A20 to a sepharose column material and thus developed the first AAV2 specific affinity chromatography [138]. In recent years, the AAV chromatography market has been dominated by materials based on specially selected nanobody constructs of cameloid origins. Examples are the "AVB Sepharose" from GE Healthcare and the "POROS CaptureSelect AAV Resins" from ThermoScientific. These materials are capable of binding up to 10^{14} rAAVs per milliliter of resin and are also stable over a wide pH, ionic strength and pressure range [139, 140]. Depending on the serotype, further affinity partners suitable for purification were identified [141].

Within product polishing, various ion exchange materials are often used. These materials separate charged molecules based on their electrostatic interaction between the AAV capsids and the ionized groups incorporated into the column matrix. The rAAVs are eluted by increasing the ionic strength through an increase of the salt concentration [142–145]. According to some reports it is possible that rAAVs can be purified by only using ion exchange chromatography. For example sulfopropyl resin could be used in this case [146]. Ion exchange chromatography can also be used to separate rAAV particles that contain a transgene from particles that do not carry a transgene [147].

The last step of the actual purification is to buffer the samples in a suitable storage buffer and sterile filtering of the rAAV containing suspension.

Especially for use in clinical trials or therapies, particular quality controls at the end of purification are necessary in order to guarantee a clean, safe, stable and high-quality product [119].

4 **Aims**

Recombinant variants of adeno associated viruses play an increasingly important role in medicine in the context of gene therapy. The main challenges in the widespread use of AAVs are the large quantities of vectors required for therapy and the time-consuming and expensive production and purification. The long-term goal of this thesis is to improve the production of rAAV2. This is going to be achieved by simplifying the production methods and reducing the purification effort by more efficient methods. In order to achieve this, several projects were started within the scope of this thesis.

As the topic of rAAV was completely new at the University of Bielefeld at the beginning of this work, the establishment of an rAAV production process was the priority. My PhD supervisor Prof. Dr. Kristian M. Müller was involved in the realization of a modular plasmid system for rAAV2 production at his previous research institutions at the Universities of Freiburg and Potsdam. This system was developed within the iGEM project at the University of Freiburg in 2010 and later extended at the University of Potsdam and served as a starting point. At a later stage, attempts on expansion of this plasmid system were conducted, as well as investigations on the limits of the modifiability of the rAAV capsid.

In order to optimize the rAAV production, which relies on standard triple transfection of adherent growing HEK293 cells, production in the most frequently used hamster cell line CHO K1 was attempted as extensive knowledge was available at the University of Bielefeld. For enabling the process, an attempt was made to genomically integrate the adenoviral helper functions already integrated in HEK293 into CHO K1 cells. Another approach aimed at simplification of the standard manufacturing process by integrating the helper gene sequences of the pHelper plasmid in HEK293 cells.

For the purification of rAAV, combinations of many, usually poorly scalable and expensive, methods such as ultracentrifugation and various chromatography methods are generally used. In addition, the development of a novel downstream process based on affinity chromatography with two different approaches was aimed. The rAAV2 affinity ligands selected were a single-chain variant of the known AAV2-binding antibody A20 and PKD 2 domain of the natural AAV receptor. The main focus here was on the simple and inexpensive provision of the affinity ligand, as well as the use of a cheap carrier material together with a straightforward rAAV2 purification protocol.

5 **Material**

In the following subchapters all material used in this thesis are listed with their respective specifications and the corresponding manufacturers. This should enable the best reproducibility of all experiments presented. If required by the experimental procedure, sterile materials and reagents were always used. The degree of purity of the chemicals can be found in the respective designations or information provided by the manufacturers under the corresponding catalogue numbers.

5.1 **Laboratory equipment**

Table 2: Laboratory equipment

Device	**Device designation (Specification)**	**Manufacturer**
Agarose electrophoresis chamber	PerfectBlue Gelsystem Mini S (different combs with 1.5 mm thickness)	Peqlab
Agarose electrophoresis chamber	PerfectBlue Gelsystem Mini M (different combs with 1.5 mm thickness)	Peqlab
Agarose electrophoresis chamber	PerfectBlue Gelsystem Mini L (different combs with 1.5 mm thickness)	Peqlab
Analytical camera system	Fusion Fx7	Vilber
Atomic force microscope	Multimode 8 AFM with Tap300Al-G cantilevers	Bruker
Autoclave	GE6612	Getinge
Autoclave	FVS 2	Integra Bioscience
Automated cell counter	LUNA	Logos Biosystems
Blue light table	Serva blue light table	Serva Electrophoresis GmbH
Camera	EOS 600D (EFS 18-55 mm)	Canon
Centrifuge	Megafuge 1.0 (swing out rotor: 2704)	Heraeus Instruments
Centrifuge	Pico17 (rotor: 75003424)	Thermo Scientific
Chromatography empty columns	C 10/10 series (2 ml)	GE Healthcare Life Science
Clean bench	LaminAir Model 1.2	Holten
Clean bench	LaminAir HB 2448	Heraeus Instruments
CO_2 Incubator	CB E6	Binder GmbH
Column chromatography system	Äkta start (with fraction collector)	GE Healthcare Life Science
Cultivation tubes	(glass, 15 ml, with aluminum cap)	-

Device	Device designation (Specification)	Manufacturer
Flow cytometer	BD FACSCalibur	Becton Dickinson
Fluorescence microscope	DMI 6000 SD Objectives: HC FL PLAN 10×/0.25 Dry HC PL FLUOTAR 20×/0.40 Dry HC PL FLUOTAR 40×/0.46 Dry HC PL APO 40×/0.85 Dry HCX APO 63×/1.40 Oil Filters: 405: EX 375-435, DC 455, EM 445-495; YFP: EX 490-510, DC 515, EM 520-550; TX2: EX 540-580, DC 595, EM 607-683	Leica Microsystems
Freezer (-150 °C)	MDF2156VAN	Panasonic
Freezer (-20 °C)	(NoFrost)	Liebherr
Freezer (-80 °C)	KM-DU 73Y1	Panasonic
French press	SLM AMINCO	SLM Instruments
French Pressure Cell	FA-032 (max 40’000 PSI, 35 ml max)	Thermo Scientific
Fridge (4 °C)	-	Bosch
Glass pipettes	(1 ml, 5 ml, 10 ml, 25 ml)	Brand
Glass shake flask	(250 ml, 500 ml, 1 l and 2 l with baffles and aluminum cap)	-
Heating block	TB1 Thermoblock	Biometra
Heating block	Mixing Block MB-102 (with orbital shaking function)	BIOER
Ice machine	UBE 50-35	Ziegra Eismaschienen
Incubator	B 6120	Heraeus
Light microscope	Axiovert 25	ZEISS
Magnetic stirrer	RCT	Ikamag
Microplate Spectrophotometer	PowerWave HT	BioTek
Microwave	HF 22043	Siemens
Milli-Q water system	Milli-Q Water Purification System	Millipore
Orbital shaker	ES-X	Kühner
PCR Workstation	Ultraviolet Sterilizing	peqlab
pH electrode	InLab Expert Pt1000	Mettler Toledo
pH meter	Seven compact S220	Mettler Toledo
Pipette (automated)	Multipette stream	Eppendorf
Pipette (variable)	Research plus 0.1 – 2.5 µl	Eppendorf
Pipette (variable)	Research plus 0.5 – 10 µl	Eppendorf

Device	Device designation (Specification)	Manufacturer
Pipette (variable)	Research plus 10 – 100 µl	Eppendorf
Pipette (variable)	Research plus 100 – 1000 µl	Eppendorf
Pipette (variable)	Research plus 500 – 5000 µl	Eppendorf
Pipetting device	Pipetboy Comfort	Integra
Power supply	EV231	Consort
Precision scale	CP 224 S	Sartorius
Precision scale	XA205 DualRange	Mettler Toledo
Pump	WP6122050	Merck Millipore
Quartz cuvette	Blacked for NanoDrop	Hellma Analytic
Real time PCR system	Lightcycler 480 II	Roche
Refrigerated centrifuge	Multifuge X1R (fixed angle rotor: F15-8 50cy, swing out rotor: TX-400 4×400)	Thermo Scientific Heraeus
Refrigerated centrifuge	Centrifuge 5424R (FA-45-24-11 rotor)	Eppendorf
Refrigerated centrifuge	RC5C (SS-34 and GS-3 rotors)	Sorvall Instruments
Rocking shaker	Duo Max 1030	Heidolph
Roller tumbler mixer	Tube Roller N2400-7010	Star Lab
Scale	BP 2100 S	Sartorius
SDS PAGE electrophoresis chamber	Small Format Vertical Electrophoresis System SE260	Hoefer
SDS PAGE gel caster	Multiple Gel Caster	Hoefer
Semi-dry blotting apparatus	Semi-Dry-Blotter (20 x 20 cm)	VWR
Shaking incubator	SI-600 R (deflection: 40 mm)	Lab. Companion
Shaking platform	Sky Line	ELMI
Spectrophotometer	NanoDrop 2000c UV/Vis	Thermo Scientific
Thermocycler	Peqstar 96x Universal gradient	Peqlab
Ultra-centrifuge	Ultra Pro 80 (T-880 rotor)	Sorvall
Vacuum Manifold	QIAvac 24 Plus	Qiagen
Vortex	Vortex Genie 2	Scientific Instruments
Water bath	-	GFL
White light table	LED, (LP-400N)	Universal Electronics Industries

5.2 **Consumable material**

Table 3: Consumable materials

Name	**Specification**	**Manufacturer**
Adhesive film for real-time PCR-Plates	Ultra clear (A26979)	GeneOn
Amicon Ultra - 4 centrifugal filters	4 ml with 10 kDa or 100 kDa NMWL	Merck
Bijou sample containers	7 ml, with screw cap, sterile, PS	Sci Labware Limited
Canula needle	21G, 0.6x80 mm	B. Braun
Cell scraper	25 cm	Sarstedt
Cryogenic plastic vessel	2.0 ml	Star Lab
Cryogenic plastic vessel (cell culture)	1.8 ml, Internal thread, silicone seal in lid	Star Lab
Lightcycler Plate	96-well, white	Sarstedt
LUNA Cell Counting Slide	Two chamber slide	Logos Biosystems
MaxiSorp plate	96-well, flat-bottom plate	Nunc
Mica for AFM	(50x76 mm, 0.2 mm thick)	PLANO
Multipette Combitips	5 ml	Eppendorf
Multi-well plates	6-, 12-, 96-well, standard TC for adherent cells, flat base, sterile	Sarstedt
Nitrocellulose membrane	0.45 micron	Thermo Scientific
Petri dishes	92x16 mm, with ventilation cams	Sarstedt
Petri dishes (cell culture)	Diameter 100 mm	Sarstedt
pH test strips	MColorpHast, pH 5.0 – 10.0	Merk
Pipette tips	10 µl, 100 µl, 1000 µl, 5000 µl	Star Lab
Pipette tips (cell culture)	TipOne Filter Tips (10/20 µl, 200 µl, 1000µl, sterile, graduated)	Star Lab
Reaction tube	15 ml, PP	Sarstedt
Reaction tube	50 ml, PP	Sarstedt
Reaction tube	50 ml, plug style caps, 25k×g	VWR
Reaction vessels	0.5 ml, 1.5 ml, 2.0 ml	Sarstedt
Reaction vessels (for PCR)	0.2 ml, flat cap, thin wall	Star Lab
Scalpel blades	S123	Hartenstein
Shaker flask	250 ml, 500 ml,	Sarsted
Sterile filter	0.2 µm, 0.45 µm, PTFE	Sarstedt/Sartorius
Syringe	1.0 ml, 5 ml, 10 ml, 50 ml	B. Braun
T-flask	25 cm^2, 75 cm^2, 150 cm^2,	Sarstedt

Name	Specification	Manufacturer
Ultracentrifuge tubes	Open-top, PA, 16x76 mm	Science Service
Blotting paper	2 mm, 100 % cotton fiber	Bio-Rad

5.3 **Chemicals**

Table 4: Chemicals

Name	Manufacturer	Catalogue number
02TCHO	Xell AG	- (now CHO TF)
ABTS	AppliChem	A1088,0001
Acetic acide	Chemical storage University Bielefeld	5000022
Acrylamid-Bisacrylamid 30 % solution (29:1)	AppliChem	A4983,0500
Agar-Agar	Sigma Aldrich	9002-18-0
Agarose	AppliChem	A8963
Ammonium sulphate	Sigma Aldrich	A4418-1KG
Ampicillin sodium salt	Carl Roth	69-52-3
APS	AppliChem	7727-54-0
ATP	NEB	P0756S
Avicel PH-101 (particle size ~50 µm)	Sigma Aldrich	11365-1KG
BH_2NaO_4	Merck	10486-00-7
Blasticidin S hydrochloride	Carl Roth	CP14.2
Brilliant Blue R-250	Sigma Aldrich	27816-25G
Bromphenol blue	AppliChem	A2331,0025
BSA	Sigma Aldrich	9048-46-8
$CaCl_2$	Merck	10035-04-8
CelLytic B Cell Lysis Reagent	Sigma Aldrich	C8740-10ML
CHAPS	Carl Roth	75621-03-3
Chloramphenicol	Carl Roth	56-75-7
Citric acid	Carl Roth	X863.1
Divinyl sulfone	Sigma Aldrich	V3700
DMEM	Sigma Aldrich	D5796
DMEM F12 Ham	Sigma Aldrich	D6434
DMSO	Carl Roth	A994.2
DTT	Thermo Scientific	3483-12-3
EDTA	Carl Roth	8043.2
Ethanol (absolute)	VWR	64-17-5
Ethanol (denatured)	Chemical storage University Bielefeld	5003033
FCS (LOT number: BCBT0730)	Sigma Aldrich	P4333

Name	Manufacturer	Catalogue number
Glucose	VWR	24369.290
Glutamine solution (cell culture)	Sigma Aldrich	G7513-100ML
Glycerol	Carl Roth	7530.1
Glycine	Carl Roth	3187.2
HBSS buffer	Sigma Aldrich	H8264-500ML
HCl	VWR	20252.290
HEK-TF	Xell AG	861-0001
HEPES	Carl Roth	6763.2
HiTrap Protein A HP	GE Healthcare	10146184
Imidazole	Sigma Aldrich	I2399
IPTG	Carl Roth	2316.4
Kanamycin	Carl Roth	8063-07-8
KCl	VWR	26764.298
KH_2PO_4	Fluka	60220
LB medium (Lennox)	Carl Roth	X964.3
Liquid nitrogen	Chemical storage University Bielefeld	-
$MgCl_2$	Carl Roth	A537.1
$MgSO_4$	Fluka	63140
Milli-Q-water	-	-
Na_2CO_3	Carl Roth	A135.1
NaCl	VWR	7647-14-5
NaH_2PO_4	VWR	444425M
NaOH	VWR	28244.295
Nitrocefin	Sigma Aldrich	484400-5MG
Non-fat milk powder	AppliChem	A0830,0500
OptiPrep	Progen	1114542
Ortho-phosphoric acid	Chemical storage University Bielefeld	5000314
PEI_{max}	Polysciences, Inc	24765
Penicillin/Streptomycin solution (cell culture)	Sigma Aldrich	P4333-100ML
Phenolred	Sigma Aldrich	P3532-5G
PMSF	Sigma Aldrich	329-98-6
POROS CaptureSelect AAVX Affinity Resin	Thermo Fisher Scientific	A36739
Protino Ni-NTA Agarose	Macherey-Nagel	745400.25
Puromycin dihydrochloride	Sigma Aldrich	P8833-10MG
Roti-GelStain	Carl Roth	3865.1
RPMI-1640	Sigma Aldrich	D8758
SDS	Carl Roth	4360.2
Silica 60 (0.04 – 0.063 mm)	Merck Millipore	7631-86-9

Name	Manufacturer	Catalogue number
Sodium citrate	Carl Roth	HN13.2
Sucrose	Sigma Aldrich	S0389-500G
TC42	Xell AG	510-0001
TEMED	Sigma Aldrich	T9281-50ML
Triethanolamine	Carl Roth	6300.1
Tris	Carl Roth	77-86-1
Tris-HCl	Carl Roth	1185-53-1
Trypan blue solution	Sigma Aldrich	T8154
Trypsin/EDTA solution	Sigma Aldrich	T4049-100ML
Tryptone/Peptone	Carl Roth	8952.2
Tween 20	Carl Roth	9005-64-5
Xylen cyanol FF	Sigma Aldrich	X4126-10G
Yeast extract	Carl Roth	2363.1
Zeocin	Thermo Fisher Scientific	R25001

5.4 **Buffer and solutions**

The compositions of the buffers and solutions presented in the following subchapter are mainly taken from the method collection of the working group and some of them were adapted to the respective requirements. Some of the compositions have also been developed from scratch.

Table 5: Buffer and solutions

Name	Composition
AAV lysis buffer	50 mM Tris, 150 mM NaCl, 2 mM $MgCl_2$, pH 7.5 – 8.0
ABTS buffer with ABTS	3.25 mM BH_2NaO_4, 40 mM citric acid, 60 mM Na_2HPO_4, pH 4.5, 1 $g \cdot l^{-1}$ ABTS (added directly before use)
Agarose gel solution with Roti-GelStain	1 % (w/v) Agarose, 0.005 % (v/v) Roti-GelStain, in TAE buffer
Ampicillin stock solution (1000×)	100 $mg \cdot ml^{-1}$ Ampicillin sodium salt
APS solution (1.5 %)	1.5 % (w/v) APS
Blocking buffer (rAAV2 capsid ELISA)	0.8 % (w/v) BSA, in PBS
$CaCl_2$ solution (100 mM)	100 mM $CaCl_2$
$CaCl_2$ solution (85 mM) with glycerol	85 mM $CaCl_2$,

Name	Composition
	15 % (v/v) Glycerol
$CaCl_2$ solution (for HEK293 transfection)	0.3 M $CaCl_2$
CHAPS stock solution	10 % (w/v) CHAPS (make directly before use)
Chloramphenicol stock solution (1000×)	20 $mg \cdot ml^{-1}$ Chloramphenicol, in Ethanol (absolute)
Coomassie staining solution	0.1 % (w/v) Brilliant Blue R-250, 40 % (v/v) Ethanol, 10 % (v/v) Acetic acid
Destaining solution (for Coomassie)	10 % (v/v) Acetic acid
DNA loading buffer (10×)	10 mM Tris-HCl, 50 % (v/v) Glycerol, 0.2 % (w/v) Bromphenol blue, 0.2 % (w/v) Xylen cyanol FF, pH 7.5
DNaseI buffer (10×)	100 mM Tris-HCl, 25 mM $MgCl_2$, 5 mM $CaCl_2$, pH 7.6
DVS buffer	10 % (v/v) Divinyl sulfone, 0.1 M Na_2CO_3, pH 11.0
Elution buffer (for POROS)	100 mM Sodium citrate pH 2.5
HBS buffer (2×)	50 mM HEPES, 1.5 mM NaH_2PO_4, 280 mM NaCl, pH 7.05
HBSS buffer	see Table 4
Hypertonic buffer	30 mM Tris, 20 % (w/v) Sucrose, 1 mM EDTA
Hypotonic buffer	5 mM $MgSO_4$
IMAC elution buffer	50 mM NaH_2PO_4, 300 mM NaCl, 250 mM Imidazole, pH 8.0
IMAC equilibration buffer	50 mM NaH_2PO_4, 300 mM NaCl, 10 mM Imidazole, pH 8.0
IMAC wash buffer	50 mM NaH_2PO_4, 300 mM NaCl, 20 mM Imidazole, pH 8.0
IPTG stock solution	1 M IPTG

Name	Composition
Kanamycin stock solution (1000×)	50 mg·ml^{-1} Kanamycin
Lämmli buffer (5×)	0.05 % (w/v) Bromophenol blue, 20 % (v/v) Glycerol, 10 % (w/v) SDS, 0.2 M Tris-HCl, 250 mM DTT (DTT is added fresh prior to use)
$MgCl_2$ solution (100 mM)	100 mM $MgCl_2$
Neutralization buffer (for POROS)	1 M Tris, pH 9.0
Nitrocefin buffer	2 mM Nitrocefin, 500 mM KH_2PO_4, 5 % (v/v) DMSO, pH 7.0
PBS buffer (1×)	500 mM NaCl, 100 mM KCl, 10 mM Na_2HPO_4, 10 mM KH_2PO_4, pH 7.5
PBS buffer (10×)	5 M NaCl, 1 M KCl, 100 mM Na_2HPO_4, 100 mM KH_2PO_4, pH 7.5
PBS with FCS	10 % (v/v) FCS, in PBS buffer
PBS-MK (10×)	10 mM $MgCl_2$, 25 mM KCl, in 10×PBS
PBS-MK + 2 M NaCl (1×)	2 M NaCl, in PBS-MK
PBST	0.05 % (v/v) Tween-20, in PBS
PEI_{max} stock solution	1 g·l^{-1} PEI_{max}
PKD2 elution buffer	100 mM Sodium citrate pH 2.5
PMSF stock solution (100×)	100 mM PMSF, in Ethanol (absolute)
Protein A elution buffer	100 mM Citric acid, pH 3.0
Protein A equilibration buffer	20 mM NaH_2PO_4, pH 7.0
Protein A neutralization buffer	1 M Tris-HCl, pH 9.0
SDS solution (10 %)	10 % (w/v) SDS

Name	Composition
SDS-running buffer (1×)	200 mM Glycerol, 25 mM Tris, 0.1 % (w/v) SDS
Semi-dry transfer buffer	25 mM Tris-HCl, 192 mM Glycine, 20 % (v/v) Ethanol, pH 8.2
Separating gel buffer	1.5 M Tris, pH 8.8
Silica elution buffer(s)	100 mM Glycine HCl pH 3.0 or 100 mM Citric acide pH 3.0 or 100 mM Triethanolamine pH 11.5 or 50 mM Tris-HCl with 2.0 M, 2.5 M or 3.0 M $MgCl_2$ pH 7.0
Silica wash buffer(s)	25 mM Tris-HCl, 0.15 M, 0.3 M, 0.6 M or 1.0 M NaCl, 0.05 % or 0 % (v/v) Tween-20, pH 8.0
SLiCE buffer (10×)	500 mM Tris-HCl, 100 mM $MgCl_2$, 10 mM ATP, 10 mM DTT, pH 7.5
Stacking gel buffer	500 mM Tris, pH 6.8
TAE buffer (1×)	40 mM Tris, 1 mM EDTA, 20 mM Acetic acid, pH 8.0
TBS blocking buffer	10 % (w/v) Non-fat milk powder, in TBS buffer
TBS buffer	50 mM Tris-HCl, 150 mM NaCl pH 7.6
TBST buffer	0.05 % (v/v) Tween 20, in TBS buffer
Trypan blue solution	see Table 4
Trypsin/EDTA solution	see Table 4
Wash buffer (rAAV2 capsid ELISA)	0.8 % (w/v) BSA in PBST

5.5 **Media**

The following table lists all culture media used in this thesis and their compositions. Media for bacterial cultivation as well as for animal/human cell culture are listed together. The compositions of media presented in the following subchapter are mainly taken from the method collection of the working group and some of them were adapted to the respective requirements.

Table 6: Media

Name	**Composition**
02TCHO transfection medium	8 mM Glutamine, optional: 1 % (v/v) Penicillin/Streptomycin solution, in 02TCHO
2×YT media	16.0 $g \cdot l^{-1}$ Tryptone/Peptone, 10.0 $g \cdot l^{-1}$ Yeast extract, 5.0 $g \cdot l^{-1}$ NaCl
Cryo conservation Media (cell culture)	10 % (v/v) DMSO, in respective cultivation media
DMEM F12 Ham medium	8 mM Glutamine, 10 % (v/v) FCS, in DMEM F12 Ham
DMEM medium	8 mM Glutamine, 10 % (v/v) FCS, in DMEM
HEK-TF medium	8 mM Glutamine, optional: 1 % (v/v) Penicillin/Streptomycin solution, in HEK-TF
LB agar medium with ampicillin	2 % (w/v) LB medium (Lennox), 1.5 % (w/v) Agar-Agar, 0.1 % (v/v) Ampicillin stock solution (1000×)
LB agar medium with chloramphenicol	2 % (w/v) LB medium (Lennox), 1.5 % (w/v) Agar-Agar, 0.1 % (v/v) Chloramphenicol stock solution (1000×)
LB agar medium with kanamycin	2 % (w/v) LB medium (Lennox), 1.5 % (w/v) Agar-Agar, 0.1 % (v/v) Kanamycin stock solution (1000×)
LB liquid medium	2 % (w/v) LB medium (Lennox)
LB liquid medium with ampicillin	2 % (w/v) LB medium (Lennox), 0.1 % (v/v) Ampicillin stock solution (1000×)
LB liquid medium with chloramphenicol	2 % (w/v) LB medium (Lennox),

Name	Composition
	0.1 % (v/v) Chloramphenicol stock solution (1000×)
LB liquid medium with kanamycin	2 % (w/v) LB medium (Lennox), 0.1 % (v/v) Kanamycin stock solution (1000×)
RPMI-1640 medium	8 mM Glutamine, 10 % (v/v) FCS, in RPMI-1640
SOC media	2 % (w/v) Tryptone/Peptone, 0.5 % (w/v) Yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgSO_4$, 10 mM $MgCl_2$, 20 mM Glucose
TC42 medium	8 mM Glutamine, optional: 1 % (v/v) Penicillin/Streptomycin solution, in TC42

5.6 **Enzymes**

5.6.1 **Restriction enzymes**

Table 7: Restriction enzymes and their specifications

Name	Restriction site	Buffer	T_{digest} in °C	Cat. Number
AatII	**5′-GACGT|C-3′** **3′-C|TGCAG-5′**	CutSmart	37	R0117S
AgeI-HF	**5′-A|CCGGT-3′** **3′-TGGCC|A-5′**	CutSmart	37	R3552S
ApaI	**5′-GGGCC|C-3′** **3′-C|CCGGG-5′**	CutSmart	25	R0114S
AscI	**5′-GG|CGCGCC-3′** **3′-CCGCGC|GG-5′**	CutSmart	37	R0558S
BamHI-HF	**5′-G|GATCC-3′** **3′-CCTAG|G-5′**	CutSmart	37	R3136S
BbsI	**5′-GAAGAC(N)$_2$|-3′** **3′-CTTCTG(N)$_6$|-5′**	NEB 2.1	37	R0539S
BsaHI	**5′-GGTCTC(N)$_1$|-3′** **3′-CCAGAG(N)$_5$|-5′**	CutSmart	37	R0535S

Name	Restriction site	Buffer	T_{digest} in °C	Cat. Number
EcoRI-HF	5′-G|AATTC-3′ 3′-CTTAA|G-5′	CutSmart	37	R3101S
KpnI-HF	5′-GGTAC|C-3′ 3′-C|CATGG-5′	CutSmart	37	R3142S
MlyI	5′-GAGTC(N)$_5$|-3′ 3′-CTCAG(N)$_5$|-5′	CutSmart	37	R0610S
NdeI	5′-CA|TATG-3′ 3′-GTAT|AC-5′	CutSmart	37	R0111S
NgoMIV	5′-G|CCGGC-3′ 3′-CGGCC|G-5′	CutSmart	37	R0564S
NheI-HF	5′-G|CTAGC-3′ 3′-CGATC|G-5′	CutSmart	37	R3131S
NotI-HF	5′-GC|GGCCGC-3′ 3′-CGCCGG|CG-5′	CutSmart	37	R3189S
PvuII-HF	5′-CAG|CTG-3′ 3′-GTC|GAC-5′	CutSmart	37	R3151S
SalI	5′-G|TCGAC-3′ 3′-CAGCT|G-5′	CutSmart	37	R3138S
SpeI-HF	5′-A|CTAGT-3′ 3′-TGATC|A-5′	CutSmart	37	R3133S
XbaI	5′-T|CTAGA-3′ 3′-AGATC|T-5′	CutSmart	37	R0145S
XhoI	5′-C|TCGAG-3′ 3′-GAGCT|C-5′	CutSmart	37	R0146S

5.6.2 **Other enzymes**

Table 8: Enzymes

Name	Manufacturer	Catalogue Number
Antarctic Phosphatase	NEB	M0289S
Benzonase	Sigma Aldrich	E1014-25KU
DNaseI	NEB	M0303S
DNaseI	AppliChem	A3778
Phusion High-Fidelity DNA Polymerase	NEB	M0530S
T4 DNA Ligase	Thermo Fisher Scientific	EL0011
T4 Polynucleotide Kinase	NEB	M0201S

5.7 **Antibodies**

The listed antibody dilutions refer to the reconstitution solutions recommended by the manufacturer.

Table 9: Antibodies

Name	**Dilution**	**Manufacturer**
Tetra-His Antibody, BSA-free, mouse anti-$(H)_4$	1:2000	Qiagen (Cat.No.: 34670)
anti-AAV VP1/VP2/VP3 mouse monoclonal, B1	1:100	Progen (Cat.No.: 65171)
anti-AAV2 (intact particle) mouse monoclonal, A20	1:250 (ELISA)	Progen (Cat.No.: 61055)
Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP conjugate	1:2500	Thermo Scientific (Cat.No.: 31430)
Anti Human IgG1 (gamma 1 chain specific), mouse antibody, HRP	1:2500	Arcris (Cat.No.: AM08151HR-N)

5.8 **Oligonucleotides**

All oligonucleotides used in this thesis were purchased from Sigma Aldrich and delivered in dry, desalt purity. All data concerning the melting temperature (T_m) and the annealing temperature (T_A) have been calculated with the "Tm calculator" of NEB. In the case of oligonucleotides with overhangs, the calculated data refer only to the initial binding part.

Table 10: Oligonucleotides

Name	**Sequence (5'→ 3')**	**Len. /nt**	**GC /%**	**T_m /°C**	**T_A /°C**
A20_GenSynth_CHO_for	AAAAAGCTAGCGCCACCATG	20	50	61	64
A20_GenSynth_CHO_rev	ATTTTGGGCCCTTATTGGTACC	22	45	61	64
A20_GenSynth_EColi_for	AAAAACATATGAAATATCTGCTGCC	25	32	58	61
A20_GenSynth_EColi_rev	TTTTTCTCGAGTTAACGACCAC	22	41	58	61
A20-scFv-Linker-Ecoli-SliCE-for	GCACCAATCTGGAAATTAAAGGTGGTGGTAGCGGTGGTGGCTCTGGTGGTGGTTCAGGTGGCGGTAGTGATGTTCAGCTGCAAGA	85	54	-	-
A20-scFv-Linker-Ecoli-SliCE-rev	TCTTGCAGCTGAACATCACTACCGCCACCTGAACCACCACCAGAGCCACCACCGCTACCACCACCTTTAATTTCCAGATTGGTGC	85	54	-	-

Name	Sequence (5'→ 3')	Len. /nt	GC /%	T_m /°C	T_A /°C
AgeI-His-S-XhoI-for	CCGGTCATCATCACCATCATCATTAAC	27	44	-	-
AgeI-His-S-XhoI-rev	TCGAGTTAATGATGATGGTGATGATGA	27	37	-	-
A-HRL-for	AAAAAGACGTCGGTAGAACTGTATAGTGGGAAATG	35	40	58	61
A-HRL-rev	AAAAAGGATCCTTCTTGAATTCTCTGTCTGGTG	33	39	58	61
A-HRR-for	AAAAAACCGGTATAACTTCGTATAGCATACATTATACGAAGTTATCAGGGGGGAGTGAGTGTAC	64	39	61	62
A-HRR-rev	AAAAAACTAGTCCTGTCTGTCTGTTGTATTAATCTG	36	33	59	62
BamHI_wt_bla-for	AAAGGATCCGTATCTACCAACCTCCAGAGAGGCAACCACCCAGAAACGCTGGCGAAAG	58	53	-	-
BamHI-MBP-for	AAAAAGGATCCGGTGGTGGCTCTGGTGGTGGTTCAGGCGGTAAAATCGAAGAAGGTAAACTGG	63	51	56	60
BamHI-mCherry-for	AAAAAGGATCCGGTGGTGGCTCTGGTGGTGGTTCAGGCGGTGTGTCCAAGGGCGAAGAGG	60	60	62	64
B-HRL-for	AAAAAGACGTCCTCTCTATTGCTGTCATAAAACACCACAAAC	42	38	64	67
B-HRL-rev	AAAAAGGATCCTGGTCAAAGAGCCGCCCA	29	52	64	67
B-HRR-for	AAAAAACCGGTATAACTTCGTATAGCATACATTATACGAAGTTATCGGTGGCACTGTTTTCTTCAG	66	36	62	65
B-HRR-rev	AAAAAACTAGTCTAGGACTGGGAGGTCAGGA	31	45	62	65
CBD-His-S-XhoI-rev	TTTTTCTCGAGTTAATGATGATGGTGATGATGACCACCGCTACCAACGGTACACGGGGTG	60	48	67	67
C-HRL-for	AAAAAGACGTCTACTCCTTTAGTCCCAACATTTG	34	38	58	61
C-HRL-rev	AAAAAGGATCCGCCTCACCTTTAAGAACATTG	32	41	57	61
C-HRR-for	AAAAAACCGGTATAACTTCGTATAGCATACATTATACGAAGTTATTAACGGCTGGATAAGCGCTT	65	35	62	64
C-HRR-rev	AAAAAACTAGTTCAACTGTTTAAACTCCAGTTCCAG	36	33	62	64

Name	Sequence (5'→ 3')	Len. /nt	GC /%	T_m /°C	T_A /°C
CMV-eGFPd2-pA-AscI-rev	AAAAAGGCGCGCCCCATAGAGCCCACCGCATC	32	63	62	58
CMV-eGFPd2-pA-BamHI-for	AAAAAGGATCCACATTGATTATTGACTAGTTATTAATAG	39	26	54	58
D-HRL-for	AAAAAGACGTCATGCCTTACCTGGAAGCAGC	31	48	62	65
D-HRL-rev	AAAAAGGATCCACTCGTGGTACACAGACAAACATG	35	43	63	65
D-HRR-for	AAAAAACCGGTATAACTTCGTATAGCATACATTATACGAAGTTATGCTTGGTGCCTCAGAGGGTC	65	40	65	68
D-HRR-rev	AAAAAACTAGTCCTGATTGTCCTGGAACTCTCTGTAGAC	39	41	66	68
Fc-AgeI-His6-Stop-ApaI-rev	TTTTTGGGCCCTCAGTGATGGTGGTGGTGATGACCGGTCTTGCCGGGGCTCAG	53	60	59	62
His-AAVR_PKD2-PT-CBD-for	AAAAACATATGCATCATCACCATC	24	33	58	61
His-AAVR_PKD2-PT-CBD-rev	TTTTTCTCGAGTTAACCAACGGTA	24	38	60	61
KpnI-Fc-for	AAAAAGGTACCGAACCTAAGTCTTGCGACAAGACC	35	46	63	62
KpnI-His6-S-XhoI-for	CCATCATCACCATCATCATTAAC	23	39	-	-
KpnI-His6-S-XhoI-rev	TCGAGTTAATGATGATGGTGATGATGGGTAC	31	42	-	-
KpnI-MBP-rev	TTTTTGGTACCACCGCTACCGCCACCTGAACCACCACCGCTACCAGTCTGCGCGTCTTTCAG	62	58	60	60
KpnI-PT-CBD-for	AAAAAGGTACCACACCAACACCTACGCCGAC	31	52	65	67
mCherry-KpnI-rev	TTTTTGGTACCACCGCTACCGCCACCTGAACCACCACCGCTACCCTTGTACAGCTCATCCATGCC	65	57	61	64
NdeI-PKD2-for	AAAAACATATGAATCGTCCGCCTATTG	27	37	52	58
pJET1.2-for	CGACTCACTATAGGGAGAGCGGC	23	61	-	-
pJET1.2-rev	AAGAACATCGATTTTCCATGGCAG	24	42	-	-
PKD2-BamHI-Linker-KpnI-rev	TTTTTGGTACCACCGCTACCGCCACCTGAACCACCACCGGATCCCGGATAATCAACTGCTTTATTAACGG	70	51	55	58
PvuII_wt_bla-rev	AAACAGCTGTAGCTGCTTGTCTCCAATGCTTAATCAGTGAGGCACC	46	48	-	-
qPCR-hGH-for	CTCCCCAGTGCCTCTCCT	18	67	-	-

Name	Sequence (5'→ 3')	Len. /nt	GC /%	T_m /°C	T_A /°C
qPCR-hGH-rev	ACTTGCCCCTTGCTCCATAC	20	55	-	-
SEQ-A20-scFv-CHO-for	CTACGAGTGGTTTACCTATTGG	22	45	-	-
SEQ-Fc-rev	GTAGGTGGAGTTGTACTGTTC	21	48	-	-
SEQ-lITR-3	ATGAACTAATGACCCCGTAATTG	23	39	-	-
SEQ-lITR-5	GAAATGTTGAATACTCATACTCTTCC	26	35	-	-
SEQ-MBP-for	GCTACGCTCAATCTGGCCTG	20	60	67	-
SEQ-mCherry-for	CCTGTCCCCTCAGTTTATG	19	53	60	-
SEQ-pcDNA5FRT-for	CGCAAATGGGCGGTAGGCGTG	21	67	-	-
SEQ-pcDNA5FRT-rev	TAGAAGGCACAGTCGAGG	18	56	-	-
SEQ-pcDNA5-hinter-pA-rev	CTTAATGCGCCGCTACAGGG	20	60	-	-
SEQ-pcDNA5-vor-CMV-for	TTAAGCTACAACAAGGCAAG	20	40	-	-
SEQ-pET24-for	TAATACGACTCACTATAGG	19	37	-	-
SEQ-pET24-rev	GCTAGTTATTGCTCAGCGG	19	53	-	-
SEQ-pSB1C3-for	AATACGCCCGGTAGTGATCT	20	50	-	-
SEQ-pSB1C3-rev	GTATTACCGCCTTTGAGTGA	20	45	-	-
SEQ-rITR-3	AACGCCTGGTATCTTTATAGTCC	23	43	-	-
SEQ-rITR-5	CCTAATCTCAGGTGATCTACC	21	48	-	-
Target-A-Down-rev	CCCAGAGTTTGGATCAGGTGCC	22	59	65	68
Target-A-for	CACCGACAGAGAATTCAAGAACAG	24	46	-	-
Target-A-rev	AAACCTGTTCTTGAATTCTCTGTC	24	38	-	-
Target-A-Up-for	GAGTGCTGTCAGGAGCAGTTGTGC	24	58	68	68
Target-B-Down-rev	CCGCAAGGAGAACCGGCAG	19	68	65	67
Target-B-for	CACCGGGCGGCTCTTTGACCACGG	24	71	-	-
Target-B-rev	AAACCCGTGGTCAAAGAGCCGCCC	24	63	-	-
Target-B-Up-for	GTTCGGATGGCTTTGACATGTGTC	24	50	64	67
Target-C-Down-rev	GTCAGTCTGAGTTCTCATTCTTAAACCG	28	43	63	66
Target-C-for	CACCGTTCTTAAAGGTGAGGCTAA	24	46	-	-
Target-C-rev	AAACTTAGCCTCACCTTTAAGAAC	24	38	-	-
Target-C-Up-for	CAAGGGAAATGGCAAGACACCTG	23	52	64	66
Target-D-Down-rev	CCGAGACAGGGTTTCTCTGTGG	22	59	64	66
Target-D-for	CACCGTCTGTGTACCACGAGTGCT	24	58	-	-
Target-D-rev	AAACAGCACTCGTGGTACACAGAC	24	50	-	-
Target-D-Up-for	GTGGTAGTACATTTGGGTAATTCTCGTAC	29	41	63	66

Name	Sequence (5'→ 3')	Len. /nt	GC /%	T_m /°C	T_A /°C
Zeo-for	AAAAAGGCGCGCCATAACTTCGTATAGCATACATTATACGAAGTTATCTGTGGAATGTGTGTCAGTTAG	69	39	58	61
Zeo-rev	AAAAAACCGGTCAGGCTTTACACTTTATGCTTC	33	39	57	61

Len.: Length

5.9 **Plasmids**

5.9.1 **Cloning vectors**

5.9.1.1 **pSB1C3**

pSB1C3 is a 2070 bp long, high copy number plasmid and the designated registry shipping plasmid backbone for the iGEM competition. It was originally generated by Austin Che in 2008 and has the official registry number BBa_J04450. It has a pUC19-derived ori (origin of replication) and it carries a chloramphenicol resistance. pSB1C3 has a MCS which was designed to allow the idempotent cloning strategy RFC 10 (see Chapter 6.1.18). Its MCS is bracketed by terminator sites which should prevent transcription from the inside of the MCS into the vector's backbone. It is shipped in a 3139 bp variant with an insert coding for mRFP1 under the control of the bacterial, constitutive promoter LacI [148]. In the ZMB lab a variant of this plasmid called pSB1C3_001_LacIprom_RFP_Term (pZMB0084) is used. This 3140 bp long variant differs from the original by the following seven mutations (based on the original numbering): 34+A, G35C, C414T, T819C, A1293G (avoid SspI digest), C1479T and C1704G (avoid PvuII digest).

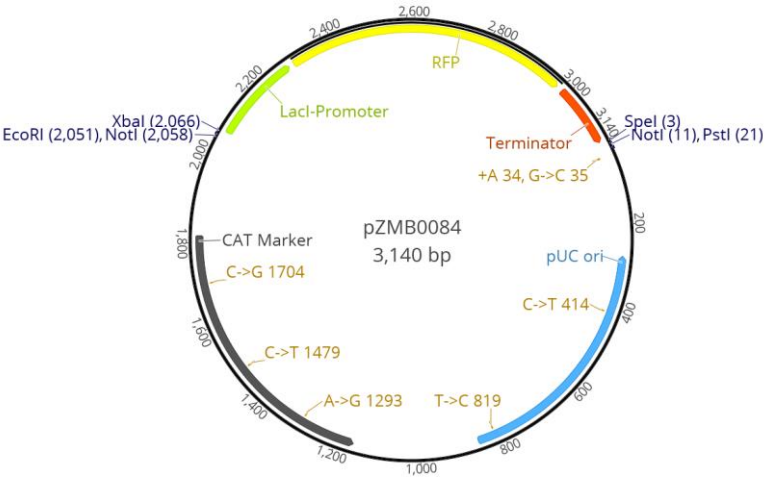


Figure 4: Vector map of pSB1C3_001_LacI_promoter-RFP-Terminator (pZMB0084, with pSB1C3_001 backbone)

5.9.1.2 **pJET1.2/blunt**

This 2947 bp long vector from Thermo Scientific is ideal for subcloning PCR products or linearized gene syntheses. It is available as a linearized blunt vector. The blunt site within the MCS is located in the ORF of the eco47IR gene, which encodes a lethal restriction endonuclease. An insertion of a DNA fragment interrupts the reading frame of the gene, which results in bacterial colony formation after transformation. The pJET1.2/blunt vector carries an ampicillin resistance. The following images show the vector map and the MCS [149].

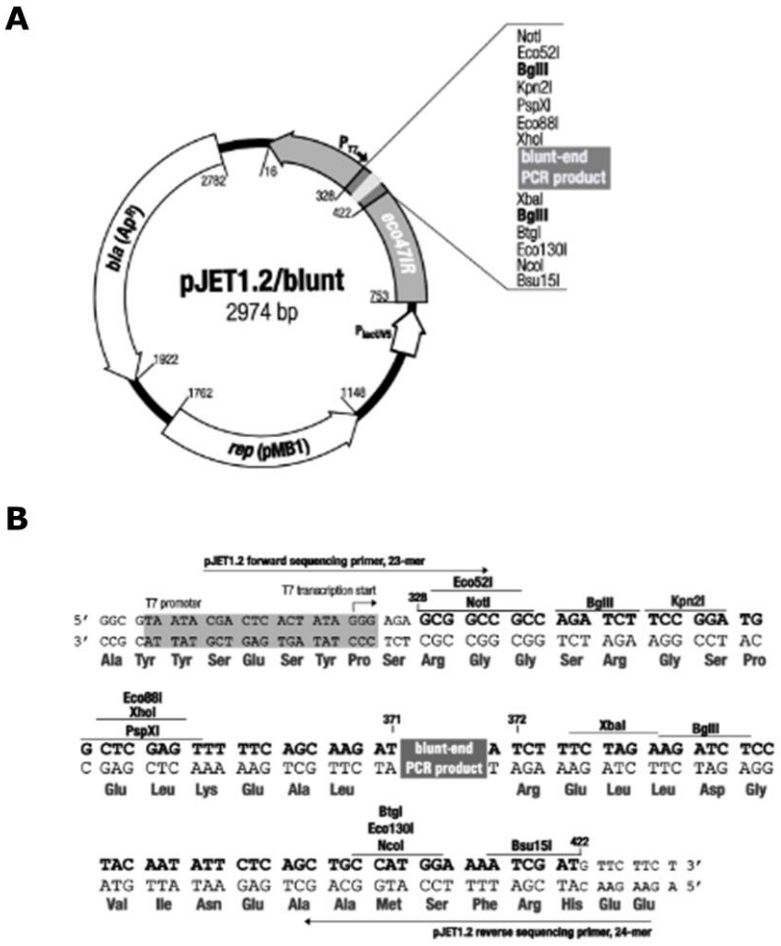


Figure 5: Vector map (A) and MCS (B) of pJET1.2/blunt vector [149].

5.9.1.3 **pUC19**

pUC19 is a 2686 bp long high copy number plasmid from Thermo Scientific. The high copy number of all pUC plasmids is a result of the lack of the rop gene and a single point mutation in the replicon rep of pMB1, which originally came from pBR322. pUC19 carries the bla gene encoding for β-lactamase, which leads to ampicillin resistance in the host organism, it differs from the bla gene of pBR322 by two point mutations. The plasmid carries the *E. coli* lac operon, which contains a CAP (catabolite activator protein) binding

site, the P_{lac} promoter, the lac repressor binding site and the 5'-terminal part of the lacZ gene encoding for the N-terminal fragment of β-galactosidase. This fragment, whose synthesis can be induced by IPTG, is capable of intra-allelic complementation with a defective form of β-galactosidase encoded by the host (mutation Δ(lacZ)M15). In the presence of IPTG, bacteria synthesize both fragments of the enzyme and form blue colonies on media with X-Gal. This can be used for a color selection, because DNA insertions into the MCS located within the lacZ gene (codons 6-7 of lacZ are replaced by MCS) interrupts the N-terminal fragment of β-galactosidase and inhibit the α-complementation. This leads to bacteria which form white colonies if they are carrying recombinant plasmids. The MCS of pUC19 is 54 bp long and contains the unique sites of 13 different restriction endonucleases. Its MCS is inverted compared to those of pUC18, which is the only difference between these plasmids. The following image show the vector map of pUC19 [150].

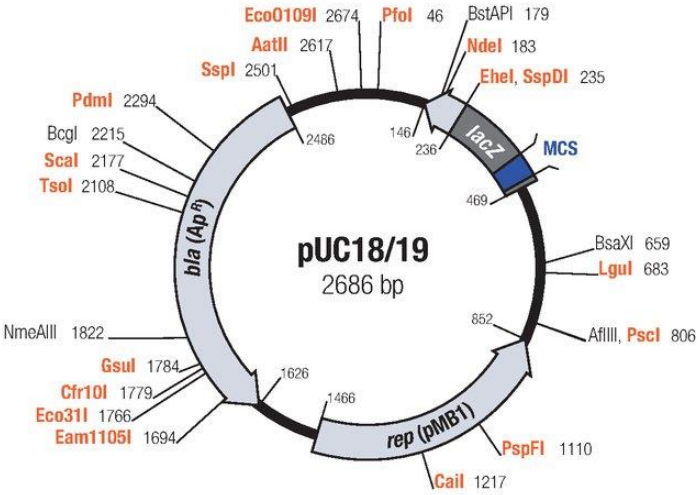


Figure 6: Vector map of pUC19 [150]

5.9.2 **Expression vectors**

5.9.2.1 **pcDNA5/FRT**

The vector pcDNA5/FRT of Life Technologies has a length of 5070 bp. The CMV (Cytomegalovirus) promoter makes it suitable for strong expression of the target gene in animal cells. Furthermore, it can be stably be integrated into the genome and selected by the flipase recombinase recognition site (FRT) and a hygromycin resistance gene, if appropriate cell lines are used. It can be selected and amplified in *E. coli* cells by its ampicillin resistance gene and the pUC ori. The following figures show a vector map and the MCS [151].

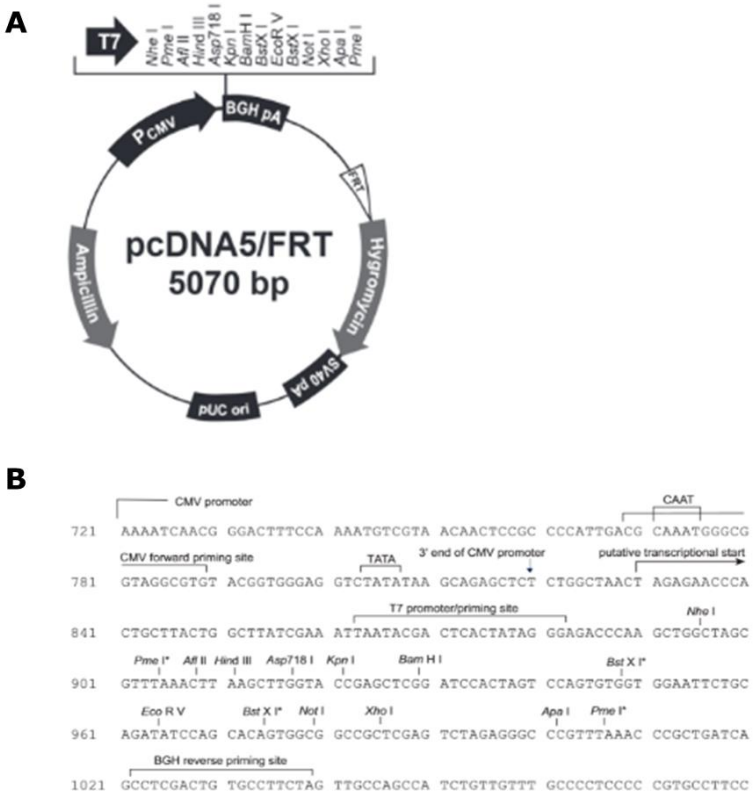


Figure 7: Vector map (A) and MCS (B) of pcDNA5/FRT vector [151].

5.9.2.2 **pET21a**

pET21a is a 5443 bp long, low copy, bacterial expression vector which uses the T7 expression system. Bacterial expression is possible in strains such as *E. coli* BL21(DE3). The DE3 appendix means that this strain contains the λDE3 lysogen that carries the gene for T7 RNA polymerase under control of the lacUV5 promoter. By induction with IPTG, the expression of the T7 RNA polymerase can be induced. The T7 RNA polymerase is necessary for the expression of the gene of interest as it is under the control of the T7 promoter on the pET21a vector. The pET21a vector also carries the natural promoter and coding sequence of the lac repressor (lacI). The lac repressor acts at the lacUV5 promoter in the host chromosome to repress transcription of the T7 RNA polymerase gene by the host polymerase, because of the lac operator sequence of the promoter. Furthermore, for the same reason, the transcription of the gene of interest is repressed, thus reducing the early production of the target protein. The pET21a Vector carries an ampicillin resistance gene and the MCS is terminated at the 5'-end by a T7-tag sequence and at the 3'-end by a His_6-tag sequence. The following figures show the vector map and the MCS [152].

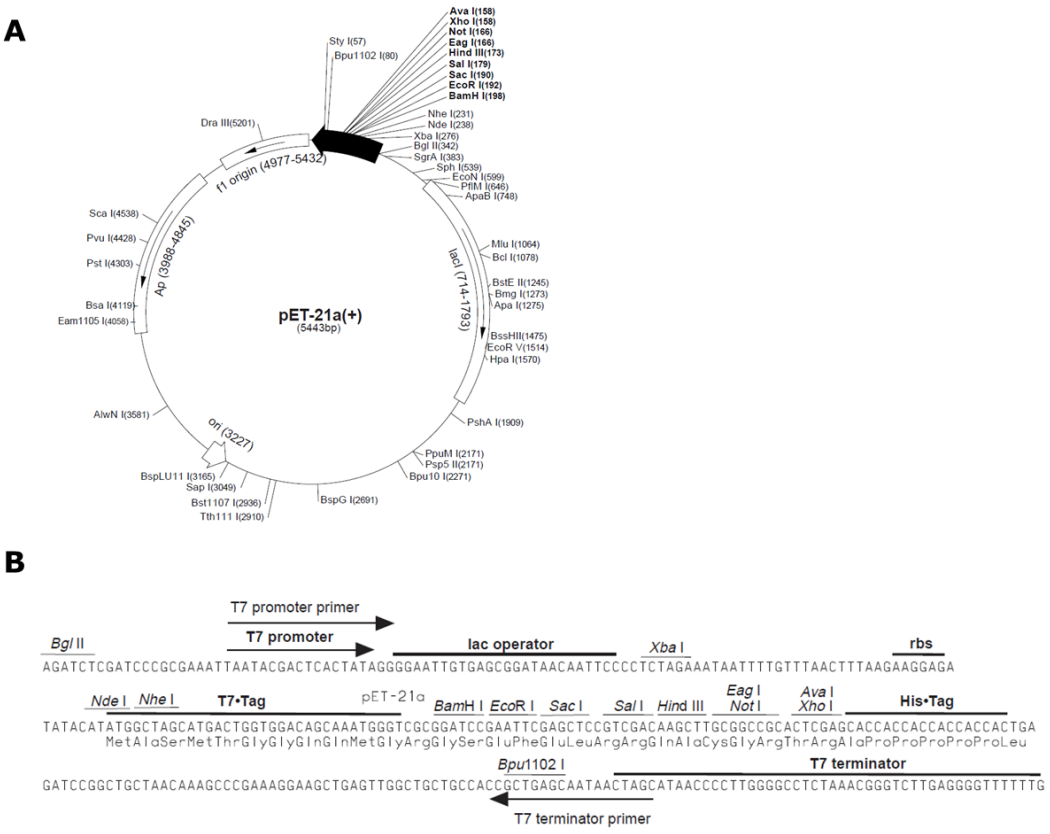


Figure 8: Vector map (A) and MCS (B) of pET21a [152].

5.9.2.3 **pET24b**

pET24b is a 5309 bp long, low copy, bacterial expression vector which uses the T7 expression system. The vector system works the same as the pET21a vector, which is why it is not discussed in detail here (see Chapter 5.9.2.2). The most important difference of both vectors is that the pET24b vector carries a kanamycin resistance gene instead of an ampicillin resistance gene. The MCS differs only by a deletion of the base C directly in front of the BamHI restriction site (position 198), so that the His_6-tag at the 3'-end of the MCS is in frame with the ATG (start codon) of the NdeI restriction site (see Figure 8B) [153].

5.9.3 **Plasmids for rAAV production**

One of the central aspects of this thesis is the production of recombinant AAV2 variants. This is achieved by using a special plasmid system. This plasmid system is based on Agilent's helper-free rAAV production system and was modified in 2010 as part of an iGEM competition project at the University of Freiburg. The students succeeded in expanding the system in such a way that capsid modifications and the modular assembly of the GOI sequence were possible by simple cloning steps [154]. Later, at the University of Potsdam, the system was extended by the possibility of mosaic virus production [76]. Ultimately, it was improved and expanded in this and other theses at the University of Bielefeld. It would

go beyond the scope of this work to go into all the possibilities of this system in detail at this point. Therefore, a selection of the most important plasmids necessary to produce rAAV2 will be discussed in the following subchapters. Further information on the plasmid system can be found in the manuscript of Feiner and Teschner *et al.* in the appendix.

5.9.3.1 **pHelper**

The 11635 bp large pHelper plasmid comes from the "AAV Helper-Free System" of Agilent and was used unmodified in this thesis. It carries an ampicillin resistance and a ColE1 ori. Within the used system, rAAV production in HEK293 cells was only possible by co-transfection of this plasmid, as it provides the adenoviral DNA sequences of the genes E2A and E4, as well as the sequence of the RNA VA. It can be found in the plasmid reserve of the working group under the number pZMB0088. The complete DNA sequence is shown in the appendix under Chapter 10.3.1.1. The following figure shows the pHelper vector map [155].

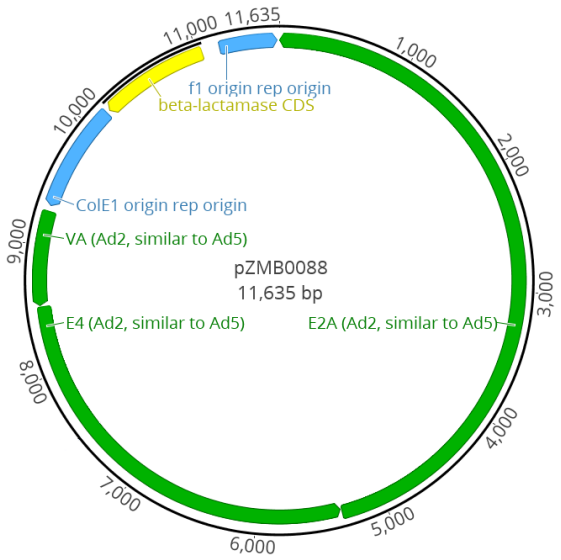


Figure 9: Vector map of pHelper (pZMB0088)

5.9.3.2 **pAAV-RC – pZMB0216**

For the rAAV production beside the pHelper and the GOI-ITR plasmid another plasmid is needed, which codes for the rep- and capsid proteins (usually abbreviated as RC- or RepCap-plasmid). The plasmid pSB1C3_001_Rep_VP123_453_587wt_p5tataless (pZMB0216) provides this function during the rAAV production, where the wild type capsid of serotype 2 should be generated. The 6455 bp plasmid is based on the pSB1C3_001 backbone and provides the possibility for modifications in the 587 and 453 loop regions. Via the singular restriction sites BamHI and PvuII the 587 loop region can be modified and via the singular restriction sites SSpI and SalI the 453 loop region can be modified. In this

thesis numerous capsid-modified rAAV2 were generated, usually by manipulation of this plasmid.

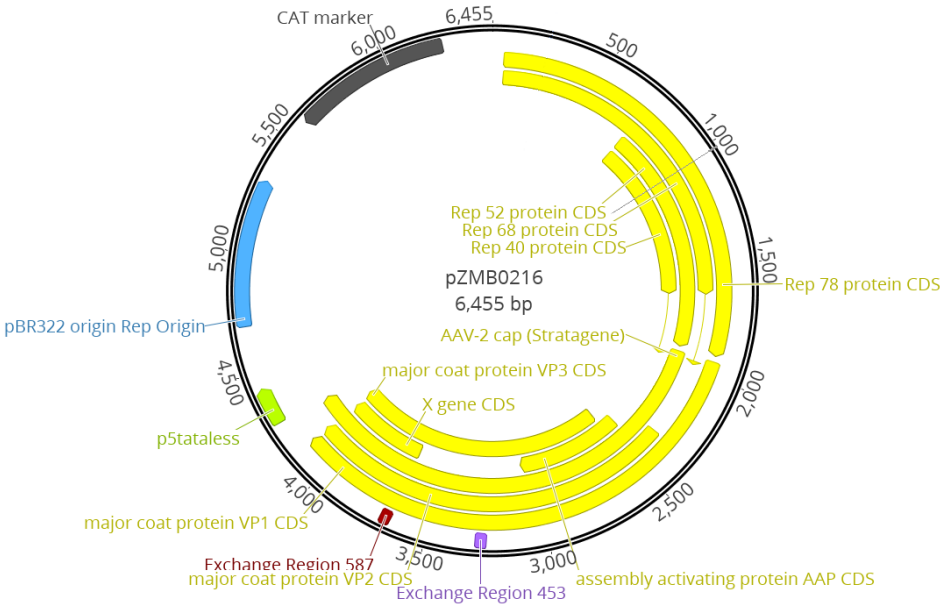


Figure 10: Vector map of pSB1C3_001_Rep_VP123_453_587wt_p5tataless (pAAV-RC, pZMB0216).

5.9.3.3 **GOI-ITR plasmid – pZMB0522**

The 4014 bp large plasmid pUC19bb_ITR_EXS_pCMV_mVenus_hGHpolyA (pZMB0522) was mainly used as a so-called GOI-ITR plasmid for rAAV production. The plasmid is based on the pUC19 backbone, which was extended by the left and right AAV2 ITR sequences. The right ITR shows the 11 bp deletion as described in Chapter 7.1.1. Via the single restriction sites EcoRI and XbaI at the 5'-end and SpeI at the 3'-end it is possible to exchange the sequence, which is located between the ITR sequences and is finally packaged in the rAAV particles. In this case, the coding sequence of the reporter protein mVenus is located there, which expression is regulated by the CMV promotor. A human growth hormone (hGH) polyadenylation signal completes this sequence. In the context of rAAV production, e.g. with the help of HEK293 cells, the cells begin to fluoresce during production, which is a side reaction of the rAAV production triggered by this plasmid, which can be used as transfection control.

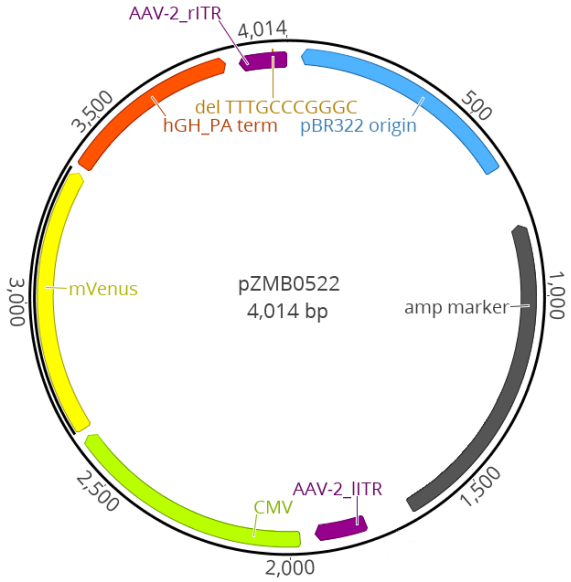


Figure 11: Vector map of pUC19bb_ITR_EXS_pCMV_mVenus_hGHpolyA (GOI-ITR plasmid, pZMB0522)

5.9.4 **CRISPR/Cas9 plasmids**

The CRISPR/Cas9 plasmids used in this thesis originate from the "ZhangLab" and were purchased via the plasmid database Addgene. This includes a variant containing the original *S. pyogenes* Cas9 (pSpCas9(BB)-2A-Puro (PX459), pZMB0065) and a construct containing the Cas9 D10A nickase variant (pSpCas9n(BB)-2A-Puro (PX462), pZMB0068). Both plasmids are 9200 bp long. They lead to the expression of the respective Cas9 variant and a Puromycin resistance protein linked via a 2A site for the selection of positive clones under the control of a CBh promoter. The insertion of the target site sequences into the sequence used for the transcription of the guide RNA can be performed by digestion using the type IIs restriction enzyme BbsI. The guide RNA is transcribed by the U6 promoter. A schematic overview of the guide sequence insertion site and the design of the corresponding target oligonucleotides is shown in Figure 12 [156].

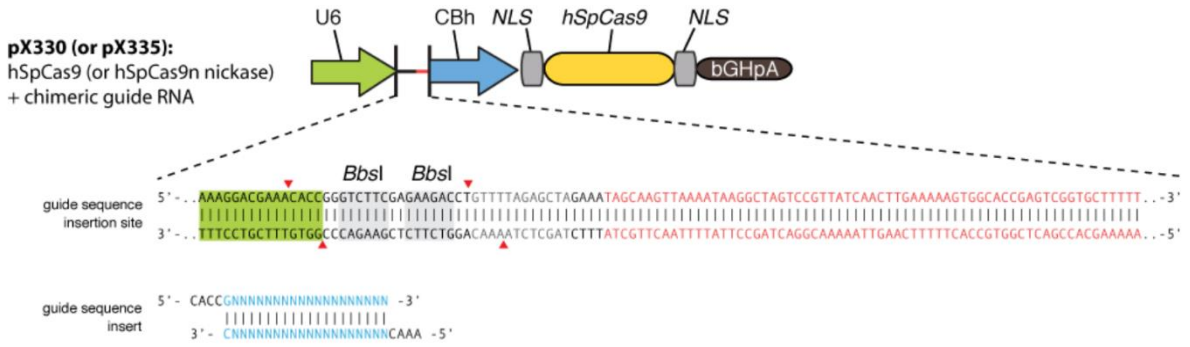


Figure 12: Schematic overview of the CRISPR/Cas9 guide RNA and target sequence design. Here pX330 an older version of the respective plasmid is shown [156].

5.9.5 **Generated and used plasmids**

The following table contains all plasmids generated and used in this thesis. It is not possible to concentrate on the generation of the self-produced plasmids in detail here. They were basically produced using standard molecular genetic methods (see Chapter 6.1) and gene syntheses (see Chapter 6.1.16). The planning and design of the most important plasmids is described in Chapter 7. All necessary plasmid and insert sequences can be found in the appendix under Chapter 10.3. The complete plasmid sequences are stored in the plasmid database of the working group. The correctness of all listed plasmids was confirmed by sequencing using the respective sequencing primers (see Chapter 6.1.17 and Table 10). All plasmids are stored under a specific "pZMB number" in plasmid reserve at -80 °C. In addition, an *E. coli* glycerol culture is deposited under the same number in the bacterial strain reserve in the respective cloning strain. Furthermore, *E. coli* BL21(DE3) glycerol cultures carrying plasmids intended for bacterial expression are stored in a separate bacterial strain reserve.

Table 11: Generated and used plasmids

pZMB	Name	*E. coli*	Creator
0001	pcDNA5/FRT	TOP10	LifeTech.
0034	IgG-Fc in pSB1C3	TOP10	JTE
0064	pJKME_bla_wt	TOP10	GFA
0065	pSpCas9(BB)-2A-Puro (PX459)	Stbl3	Addgene
0068	pSpCas9n(BB)-2A-Puro (PX462)	Stbl3	Addgene
0070	homologe Region Target A left in pJet1.2	TOP10	JTE
0071	homologe Region Target A right in pJet1.2	TOP10	JTE
0072	homologe Region Target B left in pJet1.2	TOP10	JTE
0073	homologe Region Target B right in pJet1.2	TOP10	JTE
0074	homologe Region Target C left in pJet1.2	TOP10	JTE
0075	homologe Region Target C right in pJet1.2	TOP10	JTE
0076	homologe Region Target D left in pJet1.2	TOP10	JTE
0077	homologe Region Target D right in pJet1.2	TOP10	JTE
0078	Zeocin Resistenzkassete in pJet1.2	TOP10	JTE
0083	LacIp-RFP-pSB1C3 (Bba_J04450)	TOP10	iGEM
0084	pSB1C3_001_LacI_promoter-RFP-Terminator	TOP10	JTE
0085	E1Ad5 from HEK293T in pJET1.2	TOP10	JTE
0086	eGFPd2 in pcDNA5/FRT	TOP10	JTE
0087	eGFPd2 in pJET1.2	TOP10	JTE
0088	pHelper	TOP10	Agilent
0092	pSB1C3_001_Rep_VP13_453_587ko_p5tataless	XL1 Blue	iGEM
0099	pSB1C3_001_MCS_AL_eGFPd2_Zeocin_AR	TOP10	JTE
0100	pSB1C3_001_MCS_BL_eGFPd2_Zeocin_BR	TOP10	JTE
0101	pSB1C3_001_MCS_CL_eGFPd2_Zeocin_CR	TOP10	JTE

pZMB	Name	*E. coli*	Creator
0102	pSB1C3_001_MCS_DL_eGFPd2_Zeocin_DR	TOP10	JTE
0115	pSB1C3_001_MCS	TOP10	JTE
0116	pSB1C3_001_MCS_E1_Zeocin	TOP10	JTE
0117	pSB1C3_001_MCS_eGFPd2_Zeocin	TOP10	JTE
0118	pSB1C3_001_MCS_A-left	TOP10	JTE
0119	pSB1C3_001_MCS_B-left	TOP10	JTE
0120	pSB1C3_001_MCS_C-left	TOP10	JTE
0121	pSB1C3_001_MCS_D-left	TOP10	JTE
0122	pSB1C3_001_MCS_AL_E1_Zeocin_AR	TOP10	JTE
0123	pSB1C3_001_MCS_BL_E1_Zeocin_BR	TOP10	JTE
0124	pSB1C3_001_MCS_CL_E1_Zeocin_CR	TOP10	JTE
0125	pSB1C3_001_MCS_DL_E1_Zeocin_DR	TOP10	JTE
0130	pSB1C3_001_MCS_Zeocin	TOP10	JTE
0131	pSB1C3_001_MCS_BL_E1_Zeocin	TOP10	JTE
0132	pSB1C3_001_MCS_CL_E1_Zeocin	TOP10	JTE
0133	pSB1C3_001_eGFPd2_RFC10	TOP10	JTE
0135	pSB1C3_hGHpA	TOP10	iGEM
0136	pSB1C3_leftITR	TOP10	iGEM
0138	pSB1C3_hGH_rITR	TOP10	iGEM
0139	pSB1C3_rightITR	TOP10	iGEM
0141	pSB1C3_001_VP23_453_587ko	TOP10	iGEM
0143	pSB1C3_pCMV	TOP10	JTE
0154	pSB1C3_001_Rep_VP123_453_587ko_p5tataless	TOP10	RFE
0155	pSB1C3_001_Rep_VP13_453_587ko_p5tataless	TOP10	RFE
0156	pSB1C3_001_VP23_453_587ko	TOP10	RFE
0160	pSB1C3_001_pCMV_VP23_453_587_wt	TOP10	RFE
0169	pUC19	DH5α	ThermoS
0196	pSB1C3_001_VP23_453_587koHis	DH5α	JTE
0200	pGolden AAV	DH5α	Addgene
0203	Target A in pX459	DH5α	JTE
0204	Target B in pX459	DH5α	JTE
0205	Target C in pX459	DH5α	JTE
0206	Target D in pX459	DH5α	JTE
0207	Target A in pX462	DH5α	JTE
0208	Target B in pX462	DH5α	JTE
0209	Target C in pX462	DH5α	JTE
0210	Target D in pX462	DH5α	JTE
0211	pSB1C3_001_mVenus_hGHpA	TOP10	JTE
0212	pSB1C3_001_pCMV_mVenus_hGHpA	TOP10	JTE
0216	pSB1C3_001_Rep_VP123_453_587wt_p5tataless	TOP10	JTE
0217	pSB1C3_001_Rep_VP123_453_587wtGG_p5tataless	TOP10	JTE

pZMB	Name	*E. coli*	Creator
0218	pSB1C3_001_Rep_VP123_453_587wtGGSG_p5tataless	TOP10	JTE
0219	pSB1C3_001_Rep_VP123_453_587wt2xGGSG_p5tataless	TOP10	JTE
0220	pSB1C3_001_Rep_VP123_453_587wt4xGGSG_p5tataless	TOP10	JTE
0221	pSB1C3_001_Rep_VP123_453_587wtbla_p5tataless	TOP10	JTE
0222	pSB1C3_001_CMV_VP23_453_587wtbla	TOP10	JTE
0223	iGEM 309 Freiburg (psb1c3_VCK_bla)	TOP10	JTE
0224	pSB1C3_001_RC_IRCK_p5tataless_HSPG-wt_VP2ko	TOP10	JTE
0230	pET21a	DH5α	Novagen
0246	pSB1C3_001_CMV_VP1up_NLS_mVenus_VP23_453_587koHis	DH5α	iGEM
0259	pHelper-Bsd	TOP10	MRA
0262	pET24b-PYP	TOP10	Potsdam
0295	pSpCas9(BB)-2A-Puro (PX459)	TOP10	Addgene
0296	pSpCas9n(BB)-2A-GFP (PX461)	TOP10	Addgene
0297	pSB1C3_001_pCMV_VP2_453_587ko	TOP10	KSC
0298	pSB1C3_001_pCMV_VP2_453_587wtHis	TOP10	KSC
0299	pSB1C3_001_Rep_VP13_453_587ko_p5tataless	TOP10	iGEM
0300	pSpCas9(BB)-2A-GFP (PX458)	TOP10	Addgene
0307	pSB1C3_001_pCMV_Kozak_VP23_453_587wt	TOP10	KSC
0311	pAAV_ ITR_EXS_pCMV_mVenus_hGHpolyA	TOP10	JTE
0315	pSB1C3_001_pCMV_Kozak_VP2_453_587wtHis	TOP10	KSC
0348	pET21a_A20-scFv_ohne Linker_Ecoli_His6_Silica-Tag	DH5α	JTE
0349	pET21a_A20-scFv_Ecoli_His6_Silica-Tag	DH5α	JTE
0381	pcDNA5/FRT IgG-SigP-A20scFv /w Stop /w hFc	DH5α	JTE
0382	pcDNA5/FRT IgG-SigP-A20scFv-hFc-His	DH5α	JTE
0401	pMal_c5x	DH5α	NEB
0415	pET21a_A20-scFv_Ecoli_His6	DH5α	JTE
0417	pET24b_A20-scFv_Ecoli_His6_Silica-Tag	DH5α	JTE
0421	pET24b_A20-scFv_Ecoli_His6	DH5α	JTE
0505	pET24b_His6_AAVR-PKD2_PT_CBDcex	DH5α	JTE
0522	pUC19bb_ITR_EXS_pCMV_mVenus_hGHpolyA	DH5α	KSC
0528	pGro7	DH5α	Takara
0547	pet-24b_AAVR_PKD2-shortPT-CBD_cex-His6	TOP10	JTE
0548	pet-24b_AAVR_PKD2-mCherry-shortPT-CBD_cex-His6	TOP10	JTE
0549	pet-24b_AAVR_PKD2-MBP-shortPT-CBD_cex-His6	TOP10	JTE
0577	pSB1C3_001_pCMV_Kozak_VP2_453_587wtbla	TOP10	KSC
0578	pet-24b_AAVR_PKD2-mCherry-His6	DH5α	JTE
0579	pet-24b_AAVR_PKD2-MBP-His6	DH5α	JTE
0600	pSB1C3_001_Rep_VP13_453_587wt_p5tataless	DH5α	JTE
0601	pSB1C3_001_pCMV_Kozak_VP2_453_587wt4xGGSG	DH5α	JTE

5.10 **DNA and protein ladder**

The DNA and protein length standards used in this thesis are listed below. In the case of agarose gel electrophoresis DNA ladders were used to determine the size of linearized DNA fragments, in the case of SDS-PAGE the protein standard was used to determine the size of the respective protein samples and as a visual control.

Table 12: DNA and protein ladders

Name	**Manufacturer**	**Catalogue number**
GeneRuler 1 kb DNA Ladder	Thermo Scientific	SM0311
GeneRuler 100 bp DNA Ladder	Thermo Scientific	SM0241
GeneRuler Ultra Low Range DNA Ladder	Thermo Scientific	SM1211
Color Prestained Protein Standard, Broad Range (11–245 kDa)	New England Biolabs	P7712S

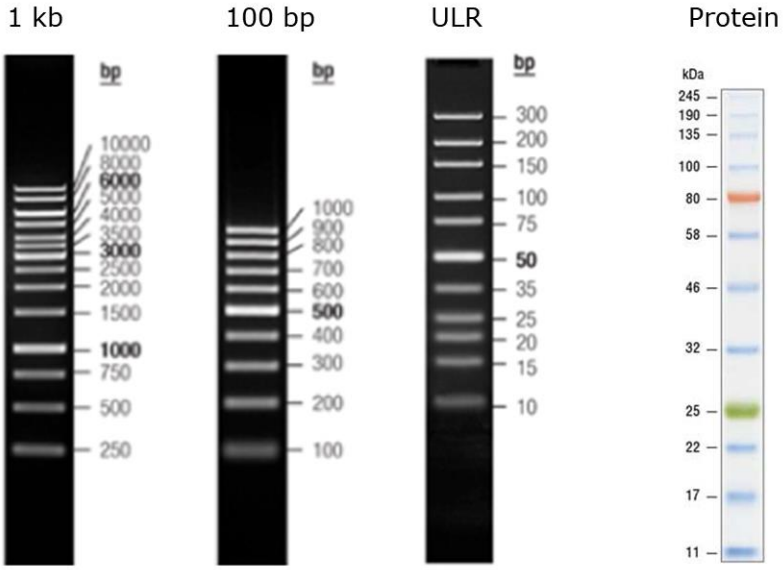


Figure 13: Used DNA and protein ladders. GeneRuler 1 kb DNA Ladder (1kb), GeneRuler 100 bp DNA Ladder (100 bp), GeneRuler Ultra Low Range DNA Ladder (ULR) and Color Prestained Protein Standard (Protein)[157, 158]

5.11 **Kits**

Table 13: Kits

Name	**Manufacturer**	**Catalogue number**
CloneJET PCR Cloning Kit	Thermo Scientific	K1231
GoTaq qPCR Master Mix	Promega	A6001
NucleoBond Xtra Midi	Macherey-Nagel	740410.50

Name	Manufacturer	Catalogue number
NucleoSpin Plasmid	Macherey-Nagel	740588.250
NucleoSpin Tissue	Macherey-Nagel	740952.50
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel	740609.250
SuperSignal West Pico PLUS Chemiluminescent Substrate	Thermo Scientific	34580

5.12 **Organisms and cell lines**

5.12.1 **Bacterial strains**

Table 14 Bacterial strains

Species	Strain	Manufacturer/Source
Escherichia coli	BL21(DE3)	AG Fermentationstechnik
Escherichia coli	DH5α	AG Fermentationstechnik
Escherichia coli	JM109	AG Fermentationstechnik
Escherichia coli	Stbl3	Thermo Fisher Scientific
Escherichia coli	TOP10 (DH10β)	Invitrogen

5.12.2 **Human and animal cells**

5.12.2.1 **HEK293**

HEK293 is a human embryonal kidney cell line. It was generated from primary embryonic tissue by the transformation with sheared adenovirus type 5 DNA. The cells contain four to five copies of the left end of the viral DNA (about 12 % of the viral genome), this includes the E1a and E1b genes and one copy of the right end of the viral DNA (10 % of the viral genome), this includes the E4 gene. The cell line grows as an adherent fibroblastoid monolayer in DMEM medium supplemented with 10 % (v/v) FCS. The doubling time is 24 to 30 hours. The cells are cultivated at 37 °C with 10 % CO_2. The confluent culture is sub cultivated every two to three days in a splitting ratio of 1:5 to 1:6. In Germany it is classified under biosafty level 1. The cell line was obtained from the DSMZ with the catalogue number ACC 305. In addition, a variant of this cell line (Freestyle 293F) was obtained from Thermo Scientific, which was adapted to growth in suspension (Cat.No.: R79007). The cells grew in HEK-TF medium and were diluted approximately every three to four days to a cell titer of $2 \cdot 10^5$ ml^{-1} [159–161].

5.12.2.2 **HT-1080**

HT-1080 is a human fibrosarcoma cell line derived from a biopsy of a 35-year-old Caucasian man from 1972. The cell line grows as an adherent epithelial-like monolayer in DMEM medium supplemented with 10 % (v/v) FCS. The doubling time is about 30 hours. The cells are cultivated at 37 °C with 10 % CO_2. The confluent culture is sub cultivated every three days in a splitting ratio of 1:10. In Germany it is classified under biosafty level 1. The cell line was obtained from the DSMZ with the catalogue number ACC 315. The cell line is known for its high HSPG presentation on the cell surface [162, 163].

5.12.2.3 **CHO-K1**

CHO-K1 is a hamster ovary cell line which is a subclone of a parental CHO cell line derived from a biopsy of a Chinese hamster (*Cricetulus griseus*) from 1957. The cell line grows as an adherent fibroblastoid monolayer in DMEM F12 Ham's medium supplemented with 10 % (v/v) FCS. The doubling time is about 24 hours. The cells are cultivated at 37 °C with 10 % CO_2. The confluent culture is sub cultivated every four to seven days in a splitting ratio of 1:3 to 1:10. In Germany it is classified under biosafty level 1. The cell line was a gift from the AG Zellkulturtechnik and has the DSMZ catalogue number ACC 110. In addition, a variant of this cell line was adapted to growth in suspension. The cells grew in TC42 medium and were diluted approximately every three to four days to a cell titer of $2 \cdot 10^5$ ml^{-1} [164, 165].

5.13 **Software and online applications**

Table 15: Software and online applications

Title	**Developer**	**Version number**
ApE	Wayne Davis	V2.0.55 (May 4, 2018)
Cell Quest	BD Biosciences	8-1996
DeepL Pro	DeepL GmbH	(2019)
Flowing Software	Perttu Terho, Turku Centre for Biotechnology – University of Turku, Finland	2.5.1
FUSION-CAPT	Vilber	-
GeneArt online tool	Thermo Fisher Scientific Inc.	-
Geneious	Biomatters Ltd.	9.1.8
Gwyddion	David Nečas and Petr Klapetek	2.48
ImageJ	Wayne Rasband, National Institutes of Health, USA	1.52a
KC4	BioTek	3.3
LightCycler 480 Software	Roche	1.5
Mendeley Desktop	Mendeley Ltd.	1.19.4

Title	Developer	Version number
NanoDrop 2000/2000c Operating Software	Thermo Fisher Scientific Inc.	1.5
Office 365 Pro Plus	Microsoft	Version 1908
OriginPro 2019	OriginLab Corporation	9.6.0.172
paint.net	dotPDN LLC	4.2.4
Tm Calculator	New England Biolabs	1.9.13
UCSC Genome Browser	UCSC Genome Informatics Group	Version 2014
UCSF Chimera	University of California	1.13.1
Unicorn start	GE Healthcare Life Science	1.0

6 **Methods**

The methods presented in the following subchapters mainly originate from the method collection of the working group, which is based among other sources on e.g. Sambrook *et al*.'s standard work and were adapted to the respective requirement [166]. Some novel methods were also developed.

6.1 **Molecular genetic methods**

6.1.1 **Generation of chemically competent *E. coli* cells by $CaCl_2$ method**

Competence refers to the ability of cells to take up DNA from the surrounding medium. For the experiments of this thesis, *E. coli* cells were used which have been made chemically competent according to the $CaCl_2$ method. For this purpose, a preculture consisting of 30 ml LB liquid medium and about 20 µl *E. coli* cell solution picked with a sterile pipette tip from a glycerol stock were mixed. This culture was then incubated overnight in a 250 ml shaker flask at 37 °C and 200 rpm. The preculture generated this way was then used to inoculate 500 ml LB liquid media with an OD_{600} of 0.1. This culture was cultivated in a 2 l baffled shaker flask at 37 °C and 180 rpm in a shaking incubator up to an OD_{600} of approximately 0.5 to 0.6. The cells were then incubated on ice for ten minutes and pelleted using a refrigerated centrifuge at 4 °C and 4000×g for ten minutes. The pelleted cells were resuspended in 130 ml ice-cold $MgCl_2$ solution (100 mM) and incubated on ice for five minutes. The suspension was then centrifuged again under the same conditions and the resulting pellet was resuspended in 26 ml ice-cold $CaCl_2$ solution (100 mM). This was followed by a 20-minute incubation on ice. After this incubation, the cells were pelleted again under the conditions mentioned above and finally mixed up with 5.2 ml ice-cold $CaCl_2$ solution (85 mM) with 15 % (v/v) glycerol. After that 50 µl aliquots were made from the cell solution and aliquoted using 1.5 ml plastic reaction vessels. Liquid nitrogen was used to freeze the aliquots which were finally stored at -80 °C [166].

The competence of the *E. coli* cells was tested (see Chapter 6.1.2) using a standard plasmid (here pUC19). The number of colonies which came up was then counted and compared to older stocks. The plasmid pBR322 produces up to $5 \cdot 10^8$ transformants per microgram of plasmid DNA used for transfection [167].

6.1.2 **Heat shock transformation**

Heat shock transformation is a method of improving the uptake of DNA by *E. coli* cells. This can be used to generate an *E. coli* strain for plasmid DNA amplification or a special *E. coli* production strain to produce a recombinant protein.

In case of purified plasmid DNA, 1 µl of the plasmid DNA (approx. 500 ng·µl^{-1}) was pipetted to the cells from a 50 µl aliquot of chemically competent *E. coli* cells thawed on ice. If a ligation batch was transformed, 5 µl (approx. 100 ng·µl^{-1}) were added. The mixes were incubated on ice for 30 minutes. A heat shock was then carried out for 45 seconds at 42 °C in the heating block. Immediately after that the cells were incubated on ice for four minutes. Meanwhile 1 ml ice-cold SOC medium was added to the cells. The transformation batch was then incubated in a 1.5 ml plastic reaction vessel at 37 °C and 200 rpm on a shaking incubator. The incubation time depended on the resistance gene provided by the used plasmid. If the resistance gene induced resistance to a bacteriostatic antibiotic such as ampicillin (bla gene), the cells were incubated for 30 minutes. If the resistance gene was directed against a bactericidal antibiotic such as chloramphenicol (CAT gene) or kanamycin (kan^{R} gene), the cells were incubated for 60 minutes.

If a purified plasmid was used for the transformation, 50 µl of the transformation batch were plated onto an antibiotic-containing LB agar plate after the incubation. In case of a ligation batch, 100 µl of it were plated onto antibiotic-containing LB agar plates. The rest of the transformation batch was centrifuged for one minute at 5000×g and the supernatant was decanted. The cell pellet was resuspended in the return flow and fully plated onto a second LB agar plate. The LB agar plates were incubated upside-down, overnight at 37 °C.

6.1.3 **Inoculation of an *E. coli* overnight culture and glycerol stock preparation**

E. coli overnight cultures were used either to amplify plasmid DNA or to expand cells of an *E. coli* production strain for subsequent day protein expression. Cultivation was carried out in 5 ml in a sterile 15 ml glass tube with an aluminum cap for plasmid mini preparation or in 50 ml in a 250 ml baffled shaker flask with an aluminum cap for plasmid midi preparation or protein production. If the *E. coli* cells came from a frozen glycerol stock, a sterile pipette tip was used to pick a small amount (approx. 20 µl) of the frozen cell suspension and the tip was transferred into the cultivation vessel. If an overnight culture should be prepared of an *E. coli* colony from a transformation, the colony was picked from the LB agar plate using a pipette tip which was being held using tweezers. In order to prevent contamination, the agar plate was held upside down for this. Cultivation was performed overnight at 37 °C and 200 rpm in a shaking incubator. The glass tubes were tilted at an angle of 45° for better mixing. On the following day (if necessary) 1 ml of the overnight culture were used to produce a permanent glycerol stock. Therefore 1 ml of the overnight culture was

transferred into a cryogenic 2.0 ml plastic vessel and the cell suspension was mixed with 0.5 ml of a sterile 87 % glycerol solution. The stocks were stored at -80 °C.

6.1.4 **Plasmid preparation**

Plasmid preparations play an important role in cloning and supplying plasmid DNA to produce rAAV and proteins. In this thesis two kits were used for the preparation of plasmid DNA, which differ basically by their scale. Both variants are described in the following two subchapters.

6.1.4.1 **Plasmid mini preparation**

For the isolation of plasmid DNA on a mini scale the "NucleoSpin Plasmid" kit from Macherey-Nagel was used. Plasmid isolation on this scale was normally performed during cloning work for the analysis of pecked *E. coli* clones or for tasks that did not require large amounts of plasmid DNA. The preparation was mainly carried out according to the manufacturer's instructions.

The starting material was approx. 4 ml of an *E. coli* overnight culture prepared the day before (see 6.1.3). After pelleting the cells, the plasmid DNA was released from the cells by alkaline lysis. The following neutralization step created suitable conditions for binding the plasmid DNA to the silica membrane of the column. Previously, proteins, genomic DNA, and cell debris were pelleted by a centrifugation step. If plasmid DNA was isolated from *E. coli* cells with high levels of nucleases like *E. coli* Stbl3, an additional washing step using wash buffer AW was performed. Contrary to the manufacturer's specifications, the plasmid DNA was not eluted using the included low-salt elution buffer but using nuclease-free water, since the produced plasmid DNA was usually used in a timely manner for further experiments. If the plasmid DNA was a sample intended for long-term storage (e.g. for the plasmid reserve), the buffer AE (5 mM Tris-HCl, pH 8.5) was used for elution [168].

6.1.4.2 **Plasmid midi preparation**

For the isolation of plasmid DNA on a midi scale the "NucleoBond Xtra Midi" kit from Macherey-Nagel was used. Plasmid isolation on this scale was normally performed for experiments that required large amounts of plasmid DNA e.g. the production of rAAV or protein production in cell culture. The preparation was mainly carried out according to the manufacturer's instructions. The starting material was approx. 150 ml of an *E. coli* overnight culture prepared the day before (see Chapter 6.1.3). The plasmid DNA was reconstituted by dissolving in 500 µl nuclease-free water [169].

6.1.5 **Preparation of genomic DNA**

In some cases, it was necessary to isolate the genomic DNA of human/animal cells for cloning work or genomic analysis. The "NucleoSpin Tissue" kit from Macherey-Nagel was used for this purpose. All steps were performed according to the "Standard protocol for human or animal cultured cells". The used starting material were 10^7 cells taken from the present cell culture. At the end the genomic DNA was eluted under low ionic strength conditions using the recommended slightly alkaline elution buffer. The produced eluate was aliquoted to 10 µl aliquots and stored at -20 °C to avoid damage to the genomic DNA caused by multiple freezing and thawing [170].

6.1.6 **Determination of concentration and purity of DNA and proteins**

The concentrations of the DNA and protein samples used and produced in this thesis were measured with the NanoDrop 2000c UV/Vis spectrophotometer from Thermo Scientific. For the measurement of DNA, the instrument was blanked with 1.5 µl of the solution in which the DNA sample was dissolved (usually Milli-Q-water). 1.5 µl of sample were used for the measurement. From the absorption quotient at 260 nm (A_{260}) and at 280 nm (A_{280}) (A_{260}/A_{280}) the purity of the measured sample could be determined based on the absorption maxima of the nucleic acids and proteins. A value in the range of 1.8 to 2.0 represents a purity which was desirable. Only samples with purity within this range were used for further experiments.

In the case of protein samples, the concentration can only be determined exactly from clean samples. 1.5 µl samples were used for the blank and the measurement. To determine the protein concentration, the molar extinction coefficient and the molecular weight of the protein needed to be mentioned. The protein concentration was measured automatically according to Lambert-Beer's law. If larger sample quantities are available, a blacked quartz cuvette should be used for the measurement since this increases the accuracy [171].

6.1.7 **Design of single-stranded DNA oligonucleotides**

Single-stranded oligonucleotides were used as primers for PCR (polymerase chain reaction) and sequencing or as starting materials for the hybridization of short double-stranded DNA fragments. All designed oligonucleotides were synthesized by the company Sigma-Aldrich. If possible, the following properties were considered within the manual design of the oligonucleotides. The primers should have a length of about 17 to 30 nucleotides and their complementary sequence should only occur once on the template DNA. The GC content should be in the range of 40 % to 60 % and the melting temperature (T_m) should not exceed 70 °C. The melting temperatures of the primers, which were used together, should not be more than 3 °C apart from each other. Furthermore, it had to be ensured that there were no sequence homologies within the primers and between the primers used together

in order to prevent primer dimers or hairpin formation. The Geneious software was used for this purpose. Base repeats exceeding the length of five nucleotides within the primer sequences should be avoided. In addition, PCR primers should end at the 3' end on bases C or G to ensure a stronger terminal binding. Some of the PCR primers had terminal sequence extensions that added sequence segments to the corresponding PCR product including restriction enzyme recognition sites or other sequences. In the case of terminal recognition sequences for digestion enzymes, the primer sequence was extended at this end by five additional nucleotides (usually A or T) in order to guarantee the activity of the respective digestion enzymes [172]. In case of the design of oligonucleotides intended for hybridization, the same properties should be considered if possible. In addition, the terminal overhangs for insertion into the target vector were considered in the design so that a restriction digestion of the hybridized product could be omitted [166]. All primers used in this thesis are listed in Table 10.

6.1.8 **Hybridization of single-stranded DNA oligonucleotides**

If small DNA fragments (≤120 bp) should be inserted into a vector during cloning work, the oligo hybridization method was ideal for the generation of the insert fragments. For this purpose, complementary single-stranded DNA oligonucleotides were designed and synthesized (Sigma-Aldrich). When designing the oligonucleotides, it had to be ensured that the nucleotides which were used in pairs were complementary and that sequences were generated which do not form unnecessary secondary structures in single-stranded form, considering the respective codon usage of the target organism. In addition, the sequences were designed in such a way that the overhangs required for ligation were present after hybridization without digestion. This way, nearly any (short) sequence could be cloned into a vector opened by digestion. These hybridized sequence segment could carry the desired insertion if only a small manipulation of the vector was desired, or it could contain recognition sites for additional restriction enzymes, enabling further (larger) insertions.

The hybridization was performed at a 10 µl scale in a PCR tube. For this, 1 µl of each oligonucleotide (100 mM) to be hybridized was mixed and 8 µl nuclease-free water was added. The following hybridization program was performed in a thermo cycler. First, the sample was heated to 95 °C for ten minutes. Then a temperature gradient starting at 95 °C decreasing to 20 °C at a speed of -1 °C·min^{-1} was carried out. The hybridization solution could then be phosphorylated if dephosphorylation was performed in the digestion reaction of the target vector or it could be used directly for ligation (see Chapters 6.1.12, 6.1.13 and 6.1.14).

6.1.9 **Polymerase chain reaction**

The PCR (polymerase chain reaction) is a molecular biological method for the *in vitro* amplification of nucleic acids. In this thesis the Phusion High-Fidelity DNA Polymerase from NEB was used for the amplification of DNA fragments by PCR. This DNA polymerase is characterized by its high processability (approx. 4 $kb \cdot min^{-1}$ on plasmid DNA) and low error rate. The starting material was either plasmid DNA (diluted to approx. 100 $ng \cdot \mu l^{-1}$), DNA from gene synthesis (20 $ng \cdot \mu l^{-1}$) or genomic DNA (approx. 250 $ng \cdot \mu l^{-1}$). Because most used templates had a high GC content (>50 %) or a complex sequence structure, the special 5× Phusion GC buffer was used in combination with a DMSO addition to increase the PCR product yield. The composition of a routine PCR reaction is shown in Table 16.

Table 16: Composition of 50 µl routine PCR reaction

Component	**Volume /µl**
5× Phusion GC buffer	10
dNTP mix (10 mM)	1
Forward Primer (10 µM)	2.5
Backward Primer (10 µM)	2.5
DMSO	1.5
Template DNA	variable
Phusion DNA Polymerase (2 $U \cdot \mu l^{-1}$)	0.5
Nuclease-free water	to 50
Total volume	**50**

If necessary, a negative control was carried out using Milli-Q water instead of template.

Table 17 shows the routine PCR program. The annealing temperature (T_A) depended on the melting temperatures determined by using the online application "NEB Tm Calculator" and the respective primer combinations (see Table 10). The needed elongation time (t_E) was calculated by dividing the size of the fragment which should be amplified by the processing activity of the DNA polymerase. Since processability is highly dependent on the template and the chosen PCR conditions, the calculated t_E was increased by 25 %.

Table 17: Thermocycling conditions for a routine PCR

Step	**Temperature /°C**	**Time /s**	**Cycles /-**
Initial Denaturation	98	30	1
Denaturation	98	10	30
Annealing	T_A	30	
Elongation	72	t_E	
Final Elongation	72	$2 \cdot t_E$ (min. 300 s)	1
Hold	8	∞	1

6.1.10 **PCR clean up and gel extraction**

The Macherey-Nagel "NucleoSpin Gel and PCR Clean-up" kit was used for the purification and re-buffering of PCR and restriction preparations as well as for the isolation of DNA from agarose gels (see 6.1.9, 6.1.11 and 6.1.15). For simplification, this method will further be called "clean-up". Most of the work was carried out according to the manufacturer's instructions. The elution was carried out under low salt conditions, deviating from the manufacturer's specifications, nuclease-free water was used for this instead of the elution buffer. The elution volume varied between 15 µl and 30 µl depending on the maximum possible volume of the following working steps [173].

6.1.11 **DNA digestion**

The restriction digestion of DNA by sequence-specific restriction endo-nucleases was used to generate DNA fragments with compatible ends during cloning, which could be linked together in a subsequent ligation reaction (see Chapter 6.1.14). Both plasmid DNA as well as DNA generated by a PCR reaction could be used as starting material.

Another important application of restriction digestion was the analysis of plasmid DNA obtained from apparently positive *E. coli* colonies from a cloning experiment. For this analysis variant, a restriction test was performed using the isolated plasmid DNA in combination with several (low-cost) restriction enzymes. The restriction enzymes were selected based on the DNA sequence which was expected. The resulting cleavage fragments should create a specific band pattern after separation by agarose gel electrophoresis (see Chapter 6.1.15), which should differ from the band pattern of incorrect cloned plasmids by size and/or count of the DNA fragments. Only plasmids which showed the correct band pattern were then sequenced (see Chapter 6.1.17) to check the entire sequence.

Nearly all restriction enzymes used in this thesis came from the company NEB. They could be used individually or in different combinations. Table 18 shows the composition of a 50 µL digestion reaction which were used for cloning with plasmid DNA or PCR products. Besides a 10 µL reaction composition is listed, as it was used in test digestions.

Table 18: Composition of a 50 µL and 10 µL digestion reaction

Component	**Volume /µl**	
Plasmid DNA or PCR product	variable (2 to 4 µg)	variable (~ 1 µg)
10× Digestion buffer (CutSmart)*	5	1
Digestion enzyme(s)	1 (each)	0.2 (each)
Nuclease-free water	to 50	to 10
Total volume	**50**	**10**

*: CutSmart was the most common used digestion buffer, but some enzymes needed other buffers.

Table 7 lists all restriction enzymes used in this thesis and information on their use. When restriction digestions were performed as described in Table 18 using HF enzymes, the incubation was about one hour at 37 °C to achieve a complete digestion. If combinations of enzymes, that did not use the same digestion buffer, were used, digestions were performed one after the other and a clean-up reaction was performed between every digestion step to exchange the buffer. After all digestion reactions, heat inactivation was carried out for 20 min at the specified temperatures.

If the used restriction enzyme could not be heat inactivated, a DNA purification in form of a PCR clean-up or gel extraction (see Chapter 6.1.10) was required before ligation. Then a preparative agarose gel electrophoresis was performed or, if the respective digestion produced only small fragments (<30 bp), a clean-up was performed to get rid of these fragments. In case of an analytical digestion, a normal agarose gel electrophoresis was performed (see Chapter 6.1.15).

6.1.12 **Dephosphorylation of DNA fragments**

For the dephosphorylation of DNA fragments, NEB's Antarctic Phosphatase was used to prevent the relegation of cut DNA fragments. This enzyme can remove 5'- and 3'-phosphates from DNA, RNA and dNTPs. It is active in all NEB restriction enzyme buffers when Antarctic Phosphatase Reaction Buffer is added during or after restriction digestion, which provides Zn^{2+} required for enzyme activity. To dephosphorylate 1 µg DNA in a total volume of 20 µl, 5 U (1 µl) Antarctic Phosphatase and 2 µl 10× Antarctic Phosphatase Reaction Buffer were added to the reaction mixture. Subsequently, a 30-minute incubation at 37 °C was performed, which could also be performed parallel to the restriction digestion. To stop the reaction, a two minute heat inactivation at 80 °C was performed [174].

6.1.13 **5' phosphorylation of DNA fragments**

For the 5' phosphorylation of DNA fragments, NEB's T4 polynucleotide kinase was used to promote subsequent ligation of the DNA fragments. This method was only used to catalyze the 5' phosphorylation of hybridized oligonucleotides (see Chapter 6.1.8) when the digested target vector was also dephosphorylated. The T4 polynucleotide kinase is capable to catalyze the transfer and exchange of P_i from the γ-position of ATP to the 5´-hydroxyl terminus of double-stranded and single-stranded DNA and RNA and nucleoside 3´-monophosphates. For the 5' phosphorylation of 1 µg DNA in a total volume of 50 µl, 10 U (1 µl) T4 polynucleotide kinase, 5 µl ATP solution (10 mM) and 5 µl 10× T4 PNK reaction buffer were added to the reaction mixture. Then an incubation of 30 minutes at 37 °C was performed. To stop the reaction, a heat inactivation at 65 °C for 20 min was required

[175]. After that, a clean-up procedure (see Chapter 6.1.10) followed by the ligation reaction (see Chapter 6.1.14) was performed.

6.1.14 **Ligation of DNA fragments**

During a ligation reaction, DNA fragments are linked by the formation of a phosphodiester bond. The used T4 DNA Ligase from Thermo Scientific is able to ligate fragments with blunt ends as well as 5'- and 3'-sticky ends.

The ligation reaction was carried out in a 20 µL reaction mixture. The incubation time was at least one hour at 22 °C. If a vector fragment had to be ligated with only one insert fragment, it needed to be considered that the insert had to be used in a 5-fold molar quantity in relation to the vector quantity. The molar mass was determined approximately through the length (in bp) of the used fragments.

If more than two fragments had to be ligated with each other, experience showed that it is better to ligate all insert fragments one after the other in groups of two. In these ligation reactions, equimolar quantities of DNA fragments were used. Following these ligations, the expected DNA fragments were separated by agarose gel electrophoresis and isolated by gel extraction (see Chapter 6.1.10) before these fragments were used for further ligations, in order to avoid unwanted by-products. After incubation, the T4 DNA Ligase was inactivated by heating to 70 °C for ten minutes [176]. The ligation reaction batch could be used directly for the transformation of *E. coli* cells (see Chapter 6.1.2). The following table shows the composition of a routine ligation reaction.

Table 19: Composition of 20 µl routine ligation reaction

Component	**Volume /µl**
10× T4 DNA Ligase Buffer	2
Linearized Vector DNA	variable (50 ng to 250 ng)
Insert DNA	variable (5-fold molar ratio of vector DNA)
50 % PEG 4000 solution*	2
T4 DNA Ligase (5 $U \cdot \mu l^{-1}$)	1
Nuclease-free water	to 20
Total volume	**20**

*: 50 % PEG 4000 solution was only used for the ligation of DNA fragment with blunt ends.

6.1.15 **Agarose gel electrophoresis**

Since DNA molecules carry a negative charge, they migrate in an electric field towards the anode. Agarose gels consist of a branched matrix of different sugar molecules, resulting in small DNA fragments migrating through the gel matrix faster than larger fragments when an electric field is applied, this is called the molecular sieve effect. This method can be used for both analytical and preparative purposes. The separation range of a gel can be influenced by the variation of the agarose concentration. For most applications, a 1 % (w/v) agarose gel is suitable, which separates DNA fragments in a size range from approx. 100 bp to 10 kb in an acceptable way [177, 178]. In this thesis the "PerfectBlue Gelsystem" of the company PeqLab was used for agarose gel electrophoresis. The generation of a 1 % agarose gel and the subsequent gel electrophoresis are briefly described below.

First, a 1 % (w/v) agarose TAE solution was prepared. The quantity depended on the size of the gel electrophoresis apparatus to be used. 40 ml, 60 ml or 80 ml agarose solution were used for a small, medium or large agarose gel. The agarose TAE buffer solution was briefly boiled in a microwave at 900 watts several times until the agarose was completely dissolved. The solution was then cooled down to about 60 °C under cold running water before adding 0.5 µl of Roti-GelStain per 10 ml gel solution [179]. After gently mixing the gel solution was poured into the gel carriage and the desired comb was placed in its intended position. Once the gel solution was completely cured, the electrophoresis apparatus was set up according to the manufacturer's instructions, the comb was carefully removed from the gel and the gel was top coated with TAE buffer. The resulting pockets could be filled with up to 25 µl DNA sample solution. The sample was first mixed with 10× loading buffer. The contained bromophenol blue was used to estimate the walking distance during gel electrophoresis. For subsequent estimations of the DNA fragments lengths, 6 µl of a DNA length marker were applied into a pocket. The electrophorese process was performed at 120 V for approx. 50 min. At the end of the run, the agarose gel was placed on a LED blue light table and a photo of the gel was taken with the camera system and the corresponding orange light filter, revealing the DNA fragments.

6.1.16 **Design of gene synthesis**

All gene syntheses of this thesis were produced by Invitrogen GeneArt. The online application of this company offers several opportunities to modify and optimize the desired DNA fragment. One of the most important functions here was the codon optimization, considering restriction sites or other motives to be avoided. The synthesizability of the fragment was always considered. It is possible to synthesize fragments with a length from 150 bp to 3000 bp and a GC content between 10 % and 80 %. In addition, longer repetitive sequences and areas that could form strong secondary structures due to homologies should be avoided. Most of the gene design work was done using the software Geneious. The

synthesized fragments were delivered as linearized, double-stranded DNA fragments in dried form. To rehydrate, nuclease-free water was added to the synthesis to form a 20 ng·µl^{-1} solution. Since the amounts of delivered DNA were usually very small, the solutions prepared this way served as a template for a PCR reaction to amplify the synthesized fragments. Therefore, a pair of primers suitable for the amplification was ordered for each gene synthesis (see Table 10).

6.1.17 **DNA sequencing**

A DNA analysis using DNA digestion (see Chapter 6.1.11) provides information on whether a plasmid has the expected size and the used restriction sites. A final confirmation whether the produced plasmid or DNA construct contains the desired sequence could only be confirmed by sequencing.

The sequencing work of this thesis was carried out by members of the "AG für Genomforschung" at the "Sequencing Core Facility" of the "CeBiTec" at the University of Bielefeld. Each sequencing reaction required 10 µL of a sequencing primer (10 µM) and 10 µl of the plasmid or DNA construct to be sequenced (approx. 250 ng·µl^{-1} to 500 ng·µl^{-1}). For the sequencing of a plasmid or a linear DNA fragment, special primers were designed, which can be identified by the prefix "SEQ" in their name (see Table 10). The design of the sequencing primers is described in Chapter 6.1.7. A sequencing reaction usually delivered 600 bp to 900 bp long readable sequences that could be used for verification. The sequences were stored as FASTA-, PHERO- and AB1- files. The last-mentioned were used for sequence verification by comparison with the theoretical sequence, using the alignment methods of the Geneious software (version 9.1.8).

6.1.18 **RFC[10] and RFC[25] cloning strategy**

The BioBrick cloning standard RFC[10] was originally developed by Tom Knight, in the context of the iGEM competition in 2003. This standard for interchangeable parts is based on an idempotent assembly strategy using the four restriction sites EcoRI, XbaI, SpeI and PstI, which are part of every BioBrick. The majority of parts in the registry database of the iGEM competition are RFC[10] compatible and were available in plasmids allowing cloning using the RFC[10] standard (e.g. pSB1C3 see Chapter 5.9.1.1). This system allows users a lot of freedom in the design of their bioparts. A central mechanism of this cloning system is the generation of compatible overhangs by digesting DNA using the restriction enzymes XbaI and SpeI. Therefore, these recognition sites are part of each prefix and suffix sequence of a BioBrick. An exact overview of the resulting prefix and suffix structure of each part is shown in Figure 14. Figure 14A shows the prefix and suffix structure of a non-coding part while Figure 14B shows that of a coding part starting with ATG as its start codon (A of ATG is part of the XbaI restriction site). An assembly of two parts can be

performed in two ways. If an insert part should be cloned to the 5'-end of an existing part, the plasmid of this part had to be opened via EcoRI and XbaI and the insert part had to be cut out of its vector via EcoRI and SpeI to make these two fragments compatible for a ligation. Another possibility is to attach an insert part to the 3'-end of an existing part. Therefore, the vector had to be opened via SpeI and PstI and the insert had to be cut out using XbaI and PstI in order to generate both parts for the ligation. After ligation, a scar would be formed at the ligation position, which do not allow further digestion at this position. But at the terminal ends of the assembled part, the RFC[10] restriction sites were still present, so that it was possible to add further parts. A disadvantage of this cloning system is that it is not possible to assemble coding sequences for fusion proteins [180]. For this reason, the RFC[10] system was expanded at the University of Freiburg in 2010 resulting in the RFC[25] cloning system. The exact prefix and suffix structures are shown in Figure 14C. In general, the existing system was extended by the two recognition sequences for the restriction enzymes NgoMIV and AgeI in frame at the 5'- and 3'-ends of the part sequence, respectively. The AgeI recognition sequence is followed by the stop code sequence TAA. Digestion with NgoMIV and AgeI also leads to complicated overhangs, making it easy to assemble sequences of fusion proteins from parts that are available in this standard [181]. All plasmids with the pSB1C3 backbone shown and used in this thesis are RFC[10] compatible and were generated this way.

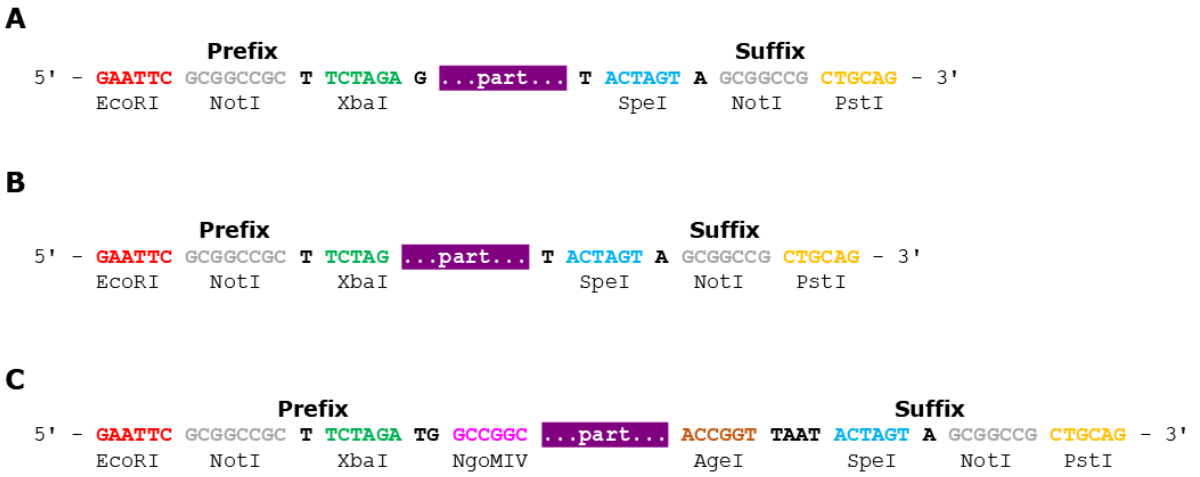


Figure 14: Prefix and suffix structure of RFC[10] and RFC[25] cloning standard. A: Prefix and suffix structure of non-coding RFC[10] parts, B: Prefix and suffix structure of coding RFC[10] parts, C: Prefix and suffix structure of RFC[25] parts allowing the generation of fusion proteins [180, 181].

6.1.19 **Blunt-end subcloning using pJET1.2/blunt-vector**

Using the "CloneJET PCR Cloning Kit" from Thermo Scientific it was possible to clone linear DNA fragments such as PCR products or string gene synthesis fragments, from 6 bp up to 10 kb length, into the linearized pJET1.2/blunt vector. This method could be used as a subcloning procedure to generate intermediate products in vector form, which made it easier to amplify the vector using *E. coli* cells. Within this work, only the blunt-end-cloning variant was used. The following table shows the composition of the 20 µl reaction batch.

Table 20: Composition of a 20 µl blunt end subcloning batch with pJET1.2/blunt vector

Component	**Volume /µl**
2× Reaction Buffer	10
Linearized DNA	approx. 0.15 pmol final
pJET1.2/blunt Vector (50 ng·µl^{-1})	1
T4 DNA Ligase (5 U·µl^{-1})	1
Nuclease-free water	to 20
Total volume	**20**

The reaction mixture listed in Table 20 was incubated at 22 °C for 20 minutes and then 5 µl of this mixture were used directly for the transformation of *E. coli* cells (see 6.1.2). The cells were then plated onto ampicillin-containing LB agar plates [149].

6.1.20 **SLiCE cloning**

SLiCE stands for "Seamless Ligation Cloning Extract" and describes a cloning method based on an easily generated bacterial cell extract. This cell extract contains all necessary enzymes to assemble several DNA fragments into a recombinant DNA molecule in a single *in vitro* recombination reaction. This requires terminal homologous sequence segments of 15 bp to 52 bp length, which could be added to the fragments e.g. by PCR using overhang primers. In contrast to traditional restriction enzyme-based cloning methods, the researcher is not dependent on the presence of singular restriction sites at the desired position for insertion, since the recombination reaction is also possible, if the homologous sequence section is flanked by small heterologous sequences [182, 183]. So, the SLiCE cloning method is a good alternative and addition to conventional cloning methods. In the context of this thesis, the SLiCE cloning was used as an alternative to the Gibson Assembly and was adapted and established for the laboratory as well as for a practical course experiment. In the following subchapters, the production of the required SLiCE extract and the general cloning strategy are described.

6.1.20.1 **Generation of the SLiCE extract**

The method for generating the SLiCE extract is based on the method described in Motohashi *et al.* 2015. The recommended *E. coli* strain PPY for the production of the extract was not available, so the strain *E. coli* JM109 which had similar characteristics was used. On the day before production a 5 ml overnight culture of the *E. coli* strain JM109 in antibiotic-free LB medium was prepared and incubated at 37 °C and 160 rpm. Since antibiotic-free medium was used during the entire process (*E. coli* JM109 has no resistance gene), special attention should be paid to sterile working. On the following day 1 ml of the *E. coli* overnight culture was used to inoculate the 50 ml main culture. The main culture was cultivated in 2×YT medium in a 250 ml round-bottom shaker flask at 37 °C and 160 rpm. During this process the growth of the cells was controlled by OD_{600} measurements. The cultivation was stopped after approx. 4 h to 6 h when an OD_{600} of 3.0 had been reached. The cells were then harvested by centrifugation at 5000×g for 5 min at 4 °C. The cell pellet was washed with 50 ml ice-cold sterilized water and then centrifuged again under the same conditions. The cell pellet was resuspended in 1.2 ml "CelLytic B Cell Lysis Reagent" and incubated for 10 min at room temperature. The resulting cell lysate was then centrifuged at 20000×g for 2 min at 4 °C. All subsequent working steps were carried out on ice. The supernatant of the centrifuged lysate was collected and 80 % (v/v) glycerol was added to the clear supernatant so that the resulting glycerol content was 40 % (v/v). The cell extract produced this way was aliquoted to 10 µl samples each in PCR tubes and frozen in liquid nitrogen. The aliquots were then stored at - 80 °C for a maximum of six months [183].

6.1.20.2 **SLiCE cloning strategy**

As already mentioned in Chapter 6.1.20, SLiCE cloning can be chosen as a cost-effective alternative to a Gibson assembly. In the context of regular cloning, this method replaces the ligation step, as both recombination and ligation of the vector and the insert fragment take place. The starting material was linearized vector DNA and a linear insert fragment. The vector DNA could originate from PCR amplification or restriction digestion. If the vector DNA originates from a restriction digest, then it was not absolutely necessary that the planned recombination takes place exactly at the restriction point. Empirically, recombination could tolerate a non-homologous region of up to 100 bp until the homologous region started. The used insert fragment usually originates from a PCR amplification in which the required 15 bp to 52 bp long terminal homology arms were added using PCR with overhang primers. If the insert was a rather short fragment, hybridized oligonucleotides could also be used. The SLiCE reaction was performed on a 10 µl scale. The components of the reaction mixture are listed in Table 21.

Table 21: Composition of a 10 µl SLiCE reaction mix

Component	**Volume /µl**
SLiCE buffer (10×)	1
SLiCE extract	1
Vector DNA	variable (100 to 300 ng)
Insert DNA	variable (5-fold molar ratio of vector DNA)
Nuclease-free water	to 10
Total volume	**10**

The reaction was carried out for 90 min at 37 °C. Then a 50 µl *E. coli* cell aliquot was transformed using the whole SLiCE reaction mixture (see Chapter 6.1.2)[182, 183].

6.2 **Cell culture methods**

6.2.1 **General cell culture**

All work presented below was carried out using sterile materials and media. As far as it was necessary for the preservation of sterility, work was carried out in the working area of a clean bench. Regarding the cell lines used, a general differentiation could be made between two types of cultivation - adherent cultivation and cultivation in suspension culture. The tissue culture flasks (T-flask) used for adherent cultivation were made of plastic and had a surface that had been hydrophilized by gas ionization, enabling cell growth on it. Tissue culture flasks with different growth surfaces of 25 cm^2, 75 cm^2 and 175 cm^2 were used, which could be filled with 7 ml, 20 ml or 40 ml medium (see Chapter 5.5). These T-bottles could be reused if the same cells were used. Cell culture trays with a diameter of 100 mm filled up with 10 ml medium were used for rAAV production. The cultivation of suspension cells usually took place in 50 ml plastic cultivation tubes on a shaking platform. The tubes were shaken tilted at an angle of approx. 45° at 185 rpm with a maximum volume of 15 ml. Larger volumes were cultivated in 250 ml or 500 ml culture flasks which can be filled up to 75 ml or 150 ml medium respectively. All cells were cultivated at 37 °C in a CO_2 incubator in a 5 % CO_2 atmosphere with a humidity of 90 %. The gas exchange was enabled via a sterile filter in the lid of the culture vessels.

6.2.2 **Thawing of cryopreserved cells**

The method of thawing cryopreservation cells was identical for adherent cells and cells growing in suspension culture. For cryopreservation, the cells were stored in DMSO-containing medium in 2 ml cryo vessel at -150 °C. The thawing process should be carried out as quickly as possible so that the cells were no longer in contact with the DMSO-containing medium as necessary, because of its cytotoxic effect. The cryo vessel coming

from the -150 °C cryo storage cabinet was thawed in a water bath preheated to 37 °C. The entire cell suspension was then transferred from the cryo vessel into a 50 ml centrifuge tube filled up with 20 ml PBS buffer preheated to 37 °C. After the suspension was gently mixed, it was centrifuged for 7 min at 900 rpm (145×g). The supernatant was decanted, and the cell pellet was resuspended in preheated medium. The type and amount of medium depends on the cell line and cultivation form. Always this type of medium was used in which the cells were cultivated before cryo conservation. Adherent growing cells were transferred into a T-flask with 75 cm^2 growth area (20 ml medium), suspension cells were transferred into a 50 ml cultivation tube (15 ml medium). The cells were then cultivated as described in Chapter 6.2.1.

6.2.3 **Subculture of adherent cells**

The adherent cell lines used in this thesis grew as monolayers. It was important that the available growth area of the T-flask was covered only up to a maximum of approx. 70 % with cells. If the maximum confluence was reached, which was estimated using a light microscope, the following steps were carried out. First, the used culture medium was poured off and the culture was washed twice with approx. 0.2 ml·cm^2 PBS buffer. The buffer solution had to be poured over the lid of the T-flask to avoid the detachment of the cells from the growing surface. To detach the cells, approx. 4 µl·cm^2 Trypsin/EDTA solution was added and an incubation at 37 °C for five minutes was performed. The detachment of the cells was supported by lightly tapping the T-flask on the working surface of the clean bench. If the cells were not detached, about 4 µl·cm^2 Trypsin/EDTA solution were added to the cells again. The protease reaction was stopped by adding at least the 10-fold volume of serum-containing medium. To separate the cells, they were pipetted up and down several times using a glass pipette until no more cell aggregates were visible in the suspension. It was important to avoid creating unnecessary foam bubbles, because the cells could be damaged by their bursting. If the cells should be passaged routinely, the corresponding amount of the cell suspension was returned to the T-flask, considering the recommended split ratio (see Chapter 5.12.2), and the remaining cell suspension was discarded. The cell suspension was then filled up to the required filling volume with fresh culture medium. The cells were further cultivated as described in Chapter 6.2.1. The routine passage of the cells was carried out two to three times per week, depending on the growth rate of the respective cell line.

6.2.4 **Titer and viability determination of a cell suspension**

The determination of the cell titer and the viability of a cell suspension was performed using the device “LUNA Automated Cell Counter” from Logos Biosystems. For this purpose, a cell suspension was required. Therefore, in case of adherent growing cells, the measurement was performed after separation of the cells during passaging (see Chapter

6.2.3). A 10 µl sample was taken from the cell suspension and mixed with 10 µl of a trypan blue solution. 10 µl of this cell suspension were pipetted into a "LUNA Cell Counting Slide" for measurement. By treating the cell suspension with trypan blue, dead cells were stained dark. The LUNA Automated CELL Counter took a photo of the sample and automatically determined the total and viable cell count and other parameters of the sample on the basis of an image analysis. This measurement was carried out twice to get precise data. The software settings that were used for the measurements are listed in Table 22.

Table 22: LUNA Automated Cell Counter software settings

Parameter	**Parameter setting**
Dilution Factor	2
Noise Reduction	5
Live Detection Sensitivity	5
Roundness	60 %
Min. Cell Size	7 µm
Max. Cell Size	60 µm
Declustering Level	Medium

6.2.5 **Subculture of suspension cells**

A certain cell density should not be exceeded in the case of suspension cell cultivation, e.g. to prevent an undersupply of nutrients. To dilute a cell suspension and exchange the medium, first a 10 µl sample was taken from the suspension cells and the living cell titer was determined (see Chapter 6.2.4). Subsequently, a certain volume was taken from the suspension culture which contains as many living cells as were required to achieve the desired cell titer in the final cultivation volume. The cell suspension was centrifuged for seven min at 900 rpm (145×g) and then the supernatant was discarded. The cell pellet was resuspended with fresh culture medium and filled up to the final cultivation volume. Depending on the cell line, a cell titer of $2 \cdot 10^5$ ml^{-1} to $1 \cdot 10^6$ ml^{-1} was adjusted in the desired final volume. This routine passage was performed two to three times per week. The cells were then cultivated as described in Chapter 6.2.1.

6.2.6 **Cryopreservation of human and animal cells**

For a long-term storage of human and animal cells, the cells were cryopreservation at -150 °C. The freezing process could lead to the formation of ice crystals that destroy the cell membranes. To prevent this dimethyl sulfoxide (DMSO) was added to the cell suspension. Since DMSO is cytotoxic, all steps had to be carried out quickly from the moment at which the cells came into contact with DMSO. The freezing medium was the respective cell culture medium, which was supplemented with 10 % DMSO (v/v) and was

pre-cooled to 4 °C. The cells had to be present as a cell suspension to be cryopreserved. For suspension cells, the living cell titer was first determined (see Chapter 6.2.4). The cell suspension was then centrifuged at 900 rpm (145×g) for 7 min, the supernatant was discarded, and the cells were resuspended in DMSO-containing freezing medium, resulting in a cell density of $1 \cdot 10^7$ ml^{-1}. Afterwards, the resulting cell suspension was quickly aliquoted to 1.0 ml aliquots each in 1.8 ml cryo vessels. In order to ensure a gentle freezing of the cell suspension, the cryogenic vessels were placed in a freezing chamber filled with isopropanol, which was then stored at -80 °C. This special freezing chamber leads to a cooling rate of approx. -1 °C·min^{-1}. The cells were stored for at least four hours at -80 °C and then transferred to the -150 °C cryo storage cabinet for long-term storage. In case of adherently growing cells, this procedure was completely analogous after the cells had been detached from the growth surface (see Chapter 6.2.3). The only difference was the number of cells that were used per cryo vessel. As many cells were aliquoted per cryo vessel as usually had to be seed in a T-flask with a growth area of 75 cm^2 plus 20 % more. This should avoid an unexpected overgrowing of the cells after thawing.

6.2.7 **Generation of a master and working cell banks**

In cell culture technology, reproducible work is only possible if master and working cell banks can be used for experiments, since the behavior of human and animal cells depends strongly on the previous cultivation conditions and the age of the cells. In order to create a master cell bank the cells, which were supplied by the respective source (usually DMSZ or a friendly working group) in the form of a cryo vessel, had to be expanded. For this purpose, the cells were thawed and expanded as far as possible within the first two passages. Most of these cells were then frozen as aliquots of the master bank as described in Chapter 6.2.6. Some of the cells were then further cultivated for one additional passage on the same large scale and subsequently cryopreserved as a working cell bank. Thus, almost identical cells could be used for many experiments. If all aliquots of the working cell bank were used up, a new working cell bank could be generated from the respective master cell bank.

6.2.8 **Generation of monoclonal cell lines by limited dilution**

The method of limited dilution was used to select single cells from a heterogeneous cell culture in order to generate monoclonal cell lines. In the context of this thesis this method was used, because at that time no flow cytometry device with cell sorting function was available. This method was used for cell cultures, after transfection in the context of genome editing experiments. First, the cell titer of the respective culture was determined in triplicates (see Chapter 6.2.4). Starting from the determined cell titer, the culture was diluted in small steps (maximum 1 to 10) to a final cell titer of 10 cells ml^{-1}. A special media composition for dilution and subsequent cultivation was used. Fresh growth medium was

mixed 1 to 1 with “used” medium from a parallel (untreated) culture of the same cell type. This used medium was removed from the culture at the time of cell passage and the cells were removed by centrifugation (7 min, 900 rpm (145×g) at room temperature) and sterile filtration (0.45 µm sterile filter). Whether the culture was an adherent or a suspension culture was irrelevant. This treatment was necessary, because most cells release signal substances into their environment, which are necessary for cell growth of the culture and a lack of these substances would negatively influence the further cultivation of single cells. Depending on the respective culture, all other supplements that were also present in the growth medium before, were added to the medium, assuming that the used medium contained its original concentration of the respective supplements. A total penicillin/streptomycin content of 1 % (v/v) (see Table 6) was adjusted for all cultures. The amount of added FCS was doubled and cultures previously cultured without FCS (e.g. CHO K1 suspension cells) were supplemented with a total of 10 % (v/v) FCS. Then, depending on the experimental setup, one or several 96-well cell culture plates were prepared for the seeding of the cells by filling the wells located at the edges (rows A and G, as well as columns 1 and 12) with 200 µl PBS each. This should prevent the samples from drying out. All inner wells of the 96-well plate were filled with 100 µl of the diluted cell culture (100 µl should statistically contain one cell). The 96-well plates were then cultivated in a cell incubator at 37 °C for 3 h in a special holder which tilted the culture plates at an angle of approx. 30°. The cells sedimented onto the lower edge of the wells and could be identified more easily and quickly during the subsequent microscopic evaluation. Within this microscopic analysis, all wells were marked and listed that contained only one cell without any doubt. Afterwards the 96-well plates were cultivated under normal conditions. Subsequently, the growth of the identified monoclonal cultures was observed microscopically until the cell-covered area of the well reached about 50 %. Once this was achieved, the complete culture was transferred into a well of a 12-well plate and 500 µl fresh growth medium were added. As soon as about 50 % of the growth area were covered again, the complete culture was transferred to a well of a 6-well plate with the addition of 1 ml fresh growth medium. When this well was covered to about 50 % the culture was further cultivated in a T-25 flask as described in Chapter 6.2.3. Thereafter, the cultures were analyzed further and cryo cultures were prepared at this stage of cultivation (see Chapter 6.2.6).

6.3 **Protein biochemical methods**

6.3.1 **Transient transfection of CHO K1 suspension cells for protein production**

The method of transient transfection of CHO K1 suspension cells was used in this thesis for the production of proteins, which could not be produced using *E. coli* cells. For this purpose, the cell titer and viability of the CHO K1 cells were determined 24 hours prior to the transfection (see Chapter 6.2.4). The viability of the cells should be above 98 %. A cell suspension with a cell titer of $1.5 \cdot 10^6$ ml^{-1} was then prepared using TC42 growth medium. The total volume depended on the desired production scale. It had to be taken into account that the final culture volume after transfection will be about 125 % larger than the culture volume produced in this step. In this thesis, transfection batches with final volumes of approx. 50 ml to 100 ml were mainly carried out. Scaling of this batches was easily possible, because the cultivation conditions could be kept the same (see Chapter 6.2.1). The following instructions refer to a transfection volume of 1 ml. The cells were cultivated for 24 hours as described in Chapter 6.2.1. On the day of transfection, the following transfection complex was prepared per milliliter of transfection batch: First, 125 µl of the 02TCHO transfection medium were added in a plastic reaction vial and mixed with 8 µl of the PEI_{max} solution (1 $g \cdot l^{-1}$). The mixture was incubated at room temperature for at least one minute. At the same time, 125 µl of the 02TCHO transfection medium mixed up with 2 µg of the plasmid DNA needed for transfection were incubated for at least one minute at room temperature in a second plastic reaction tube. After incubation, the PEI-medium and the DNA-medium mixtures were mixed together and incubated again for 15 minutes at room temperature. In the meantime, the cells were counted again to ensure that the live cell count was at least $3.0 \cdot 10^6$ ml^{-1} and the viability was higher than 98 %. The culture medium was then removed from the cells using a double washing step. The cells were centrifuged for 7 minutes at 900 rpm (145×g), the supernatant was removed, and the pellet was washed with PBS buffer. The amount of PBS buffer used was equal to the used volume of cell suspension. After the last washing step, the cell pellet was resuspended in the same volume of 02TCHO transfection medium. The incubated transfection complex was added to the cell suspension. The cells were then cultivated for 4 h at 37 °C in a CO_2 incubator at 185 rpm. After this incubation 1 ml TC42 medium was added to the cell suspension. The cells were then cultivated for five to nine days, during which the cell titer and the viability of the cells were controlled daily (see Chapter 6.2.4). As soon as the viability of the cells dropped below 60 %, they were harvested. Since the produced proteins had secretory signal peptides, they were expected to accumulate in the supernatant of the cell suspension. The cell suspension was centrifuged for ten minutes at 900 rpm (145×g), the supernatant was removed and sterile filtered by using a 0.22 µm sterile filter. The

filtered supernatant was stored at -20 °C until further processing, e.g. using IMAC (immobilized metal ion affinity chromatography) (see Chapter 6.3.5) or, in the case of antibody constructs, protein A purification (see Chapter 6.3.6) [184].

6.3.2 *E. coli* cultivation for protein production (pET system)

In this thesis the *E. coli* strain BL21(DE3) in combination with the pET vector system was used for bacterial production of proteins. Thereby, it was possible to induce the expression of the target proteins by IPTG induction using the underlying T7 expression system, which leads to high levels of bacterial protein production (see Chapter 5.9.2.2 and 5.9.2.3). For this purpose, the following procedure was used here on the example of a 1 l *E. coli* culture. A 50 ml overnight culture of the *E. coli* BL21(DE3) strain carrying the respective pET expression plasmid was prepared (see Chapter 6.1.3). The next day, the OD_{600} of the *E. coli* culture was determined and a 1 l main culture with an OD_{600} of 0.1 was started. Therefore, the *E. coli* suspension was divided to 500 ml each into two 2 l baffled shaker bottles, which were then cultivated at 37 °C and 180 rpm. The OD_{600} was determined regularly and once it had reached a value of 0.7, 500 µl of a 1M IPTG stock solution were added to each of the two cultures to induce the expression of the target protein (final 1 mM IPTG). After that, the cultivation was continued under the same conditions for four hours. After cultivation, the cell suspension was cooled on ice for ten minutes and then centrifuged at 6000×g and 4 °C for 15 minutes to harvest the cells. The supernatant of cultivation was discarded. The wet mass of the *E. coli* cell pellet was determined, and 10 ml of buffer were used per gram wet mass to resuspend the *E. coli* cells. The buffer, which was used for resuspending, depended on the later processing of the target protein. If, for instance, IMAC should be used for the purification, the IMAC equilibration buffer was used (see Table 5), otherwise PBS was used. Additionally, PMSF from a 100 mM PMSF stock solution was added (final 1 mM PMSF) to reduce proteinase activity. The cell suspension was then processed directly by French Press (see Chapter 6.3.3) or stored at -20 °C.

6.3.3 *E. coli* cell disruption using a French Press

The French Press was used for the cell disruption of *E. coli* cells. A cell suspension was compressed using a pressure of 1000 PSI to 2000 PSI in a pressure chamber with a piston pump. The pressure chamber had a needle valve with a small opening diameter. By opening this valve, the previously compressed liquid expanded, and the resulting shear forces led to the disruption of the respective cells' bio membranes.

The French Press of the company SLM AMINCO was used, which had a maximum filling volume of 30 ml. Before use, all parts of the pressure chamber were pre-cooled on ice. The pressure chamber was then assembled according to the manufacturer's specifications and the (thawed) cell suspension was filled in (see Chapter 6.3.2). The assembled pressure

chamber was inserted into the apparatus and a pressure of approx. 2000 psi was built up with the piston pump. The needle valve was then carefully opened, and the slowly dripping cell suspension was collected in an ice-cooled tube. When the suspension had completely drained out, the pressure chamber was disassembled again, and the described procedure was repeated two more times. The resulting suspension was then centrifuged for 30 min at 4 °C and 20000×g to get rid of the cell debris. To the resulting clear supernatant, the required quantity of 10× DNaseI buffer and 0.1 µl DNaseI solution (min 3000 $U \cdot ml^{-1}$) per milliliter were added, depending on the total volume and incubated at 37 °C for 20 minutes. The suspension was then sterile filtered using a 0.22 µm sterile filter and further processed, e.g. by IMAC purification (see Chapter 6.3.5) or could be used directly for suitable experiments. Depending on the produced protein the suspension could also be stored at -20 °C or -80 °C.

6.3.4 *E. coli* cultivation for periplasmatic protein production and cell disruption by osmotic shock

The periplasmatic protein expression in *E. coli* was basically the same as for the normal cytosolic protein expression, since a pET system was also used. Only some of the cultivation parameters, the cultivation time, and the subsequent processing of the cells were different. The generation of the overnight culture and the creation of the pre-culture were identical to the method shown in Chapter 6.3.2. The *E. coli* main culture was cultivated at 37 °C until an OD_{600} of 0.8 was reached, after which the protein expression was induced by the addition of IPTG (final concentration 0.1 mM from 1 M stock solution). Immediately after the induction, the cultivation temperature was reduced to 25 °C and the culture was further cultivated for 18 hours. The cells were harvested by centrifugation at 6000×g, for 10 min and 4 °C. The supernatant was discarded. Subsequently, the periplasmic proteins were released via osmotic shock by incubation with a hypertonic and a hypotonic buffer. This type of periplasmic cell disruption had the advantage that the resulting protein suspensions were almost free of cytosolic proteins, making the suspension much cleaner. In addition, species of the target protein that have not yet formed disulfide bonds were also nearly eliminated. The wet mass of the *E. coli* cell pellet was determined and 10 ml of the hypertonic buffer (see Table 5) were used per gram wet mass to resuspend the *E. coli* pellet. The suspension was incubated on ice for 30 min. Then centrifuged at 11600xg for 20 min at 4 °C. The resulting supernatant (hypertonic cell supernatant) was collected. Subsequently, the cell pellet was resuspended in the same volume of hypotonic buffer (see Table 5) as before. The suspension was incubated on ice for 30 min followed by a centrifugation step at 11600×g for 20 min and 4 °C. The collected supernatants contained a large amount of the periplasmatic proteins. Depending on the respective target protein, higher amounts were present in one or the other of them. The supernatants were

therefore used separately or combined for further processing (e.g. IMAC purification or silica purification, see Chapter 6.3.5 or 6.3.7) (adapted from [185]).

6.3.5 **IMAC column chromatography purification**

Immobilized metal ion affinity chromatography (IMAC) enable a fast purification method of recombinant polyhistidine-tagged (e.g. His_6-tag) proteins. Binding of protein is based on the interaction between the His_6-tag of the recombinant protein and immobilized Ni^{2+} ions. Due to the addition of imidazole, the target protein can then be eluted [186].

For IMAC protein purification, the "Protino Ni-NTA Agarose" from Macherey-Nagel, which was filled into self-packable chromatography columns from GE healthcare, was used in this thesis [186, 187]. Depending on the expected protein quantity, Ni-NTA columns with column volume (CV) between 0.5 ml and 2 ml were used and prepared as mentioned in the manufacturer's instructions. For chromatography the "ÄKTA start" with a connected fraction collector from GE healthcare was used, which allowed an automation of the purification and control of the process [188]. For the purification, the IMAC equilibration buffer (see Table 5) and the IMAC elution buffer (see Table 5) were connected to the buffer feeds A and B of the ÄKTA. The required wash buffer was mixed via the mixing unit of the ÄKTA out of the equilibration buffer and 7.0 % (v/v) elution buffer. After the buffers, the sample to be processed and the chromatography column had been connected to the ÄKTA, all necessary tubings were flooded with the respective buffers. Then the IMAC standard program was performed as described in the following section. During the entire process the flow rate was set to 1 $ml \cdot min^{-1}$, the maximum pressure should not exceed 0.45 mPa. First the column was equilibrated with a total of 10 CV equilibration buffer, the UV baseline at 280 nm should then be stable. Then an autozero of the measuring units was carried out. The clarified *E. coli* lysate (see Chapter 6.3.3) was then completely applied to the column via the sample feeder. The column was then washed with a total of 10 CV of the mixed washing buffer. Elution was performed with 5 CV of the elution buffer in total. The elution was collected in 500 µl fractions with the help of the fraction collector. After this process, the elution fractions were stored on ice and the protein concentration of each individual fraction was determined as described in Chapter 6.1.6. The fractions with the highest protein concentration were combined. The combined elution fractions were then re-buffered several times with the desired end buffer/storage buffer using Amicon Ultra 4 centrifugal filters and then stored in aliquots of 50 µl to 100 µl at -20 °C or -80 °C.

6.3.6 **Protein A column chromatography purification**

Protein A sepharose was designed for purification and isolation of monoclonal and polyclonal IgG from several sources, e.g. cell culture supernatants from CHO K1 cells (see Chapter 6.3.1). Protein A can bind IgG over a wide pH range, elution is mainly achieved by a decrease of the pH value [189].

For antibody purification, the "HiTrap Protein A HP" prepacked 1 ml columns from GE healthcare were used [189]. The "ÄKTA start" with connected fraction collector from GE healthcare allowed an automation of the purification process [188]. A protein A equilibration buffer, a protein A elution buffer and a neutralization buffer were prepared (see Table 5). The equilibration buffer and the elution buffer were connected to the buffer feeds A and B of the ÄKTA. There was no separate wash buffer. After the buffers, the supernatant sample and the chromatography column had been connected to the ÄKTA, all necessary tubings were flooded with the respective buffers. Then the protein A standard program was performed as described in the following section: During the entire process the flow rate was set to 1 $ml \cdot min^{-1}$, the maximum pressure should not exceed 0.45 mPa. First the column was equilibrated with a total of 10 CV equilibration buffer, afterwards the UV baseline at 280 nm should be stable. Then an autozero of the measuring units was carried out. The CHO K1 supernatant (see Chapter 6.3.1) was then completely applied to the column via the sample feeder. The column was then washed with a total of 10 CV of the equilibration buffer. Elution was performed with 5 CV of the elution buffer. The elution was collected in 500 µl fractions with the help of the fraction collector. Prior to the elution, 200 µl protein A neutralization buffer was added to each collection tube to reduce the exposure of the produced antibody constructs to the acidic elution buffer. After this process, the elution fractions were stored on ice and the protein concentration of each individual fraction was measured as described in Chapter 6.1.6. The fractions with the highest protein concentration were combined. The combined elution fractions were then re-buffered several times with the desired buffer or a storage buffer using an Amicon Ultra 4 centrifugal filter and then stored in aliquots of 50 µl to 100 µl at -20 °C or -80 °C.

6.3.7 **Binding of fusion proteins to silica material and further processing**

First, the silica material (particle size: 0.04 mm to 0.063 mm) was washed with silica wash buffer. Therefore 25 ml silica wash buffer were added to one gram silica material. The suspension was vortexed for ten seconds and then centrifuged for two min at 5000×g using a swing-out rotor. The supernatant was decanted, and the washing step was repeated three times. The washed silica pellet was filled up with silica wash buffer to a final volume of 4 ml (250 mg silica material per ml suspension). This washed silica suspension could be stored at 4 °C for up to two weeks. It served as a starting material for the subsequent

binding of fusion proteins with silica-tag. It was important that the silica suspension was thoroughly vortexed directly before each pipetting step and that the suspension was swiveled further during pipetting to avoid the silica material to settle down. Pipette tips with a minimum volume of 100 µl were used for pipetting, because the silica material could not be pipetted evenly using smaller pipette tips (pipette tips with smaller opening). The binding of the fusion protein to the silica material was performed in a 25 mg silica material per milliliter suspension. Therefore, the required amount of washed silica suspension was centrifuged for 2 min at 5000×g and the supernatant was removed. Subsequently, a protein-containing suspension was added to the silica pellet, using the hypotonic cell extract produced by the osmotic shock of *E. coli* cells (see Chapter 6.3.4). The suspension was briefly vortexed to mix the silica material with the protein solution and then incubated for 15 min at room temperature on a roller tumbler mixer. After centrifugation for two minutes at 5000×g the supernatant was discarded. By adding PBS buffer, a 250 mg silica per ml suspension was generated. This suspension served as the starting material for further experiments.

Whenever the experimental description mentions that a defined amount of silica material was used, it was taken by pipetting from a previously generated 250 $mg \cdot ml^{-1}$ stem suspension (as described above). The silica suspension was then centrifuged for 2 min at 5000×g and the liquid supernatant was removed. The resulting wet silica pellet consisted of the desired amount of silica material and some of the liquid from the slurry that remained inside the porous material. A weighting-experiment, using a defined amount of silica material and Milli-Q-water, revealed a factor of 195.3 ±0.5 % by which the weight of the dry silica material was increased in its wet form. Based on the density of the water (approx. 1.0 $g \cdot ml^{-1}$), this volume increase had to be considered as a correction factor for quantitative experiments (e.g. in context of AAV binding experiments).

Different quantities and different buffers were used for these experiments. Details will be discussed in the results chapter of this thesis. Further working steps were carried out under comparable conditions. All washing steps were carried out as described above, adding 5 times the volume of wash buffer (based on 250 mg per ml silica suspension), followed by ten seconds of vortexing and a centrifugation steps for two minutes at 5000×g for a total of three times. Elutions were performed by adding twice the volume of elution buffer (based on the 250 mg per ml silica suspension) and incubating for 15 minutes on the roller tumbler mixer at room temperature. The eluate was separated from the silica material by centrifugation (2 min, 5000×g). Usually two elution steps were performed.

6.3.8 **SDS-PAGE**

For the detection of proteins, samples were analyzed by denaturing SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). Within this thesis acrylamide gels with a thickness of 1.0 mm and a separation distance of approx. 10 cm were used for protein separation. The compositions of 10 %, 12 % and 15 % separating gels and 4 % stacking gel are listed in Table 23.

Table 23: Compositions of 10 %, 12 % and 15 % separating gels and 4 % stacking gel (sufficient for 4 gels)

Component	**Volume /ml**			
	10 %	**12 %**	**15 %**	**4 %**
Milli-Q water	16.33	13	8	8.17
Separating gel buffer	12.5	12.5	12.5	-
Stacking gel buffer	-	-	-	3.78
10 % SDS solution	0.5	0.5	0.5	0.15
Acrylamid-Bisacrylamid 30 % solution (29:1)	16.65	20	25	3.35
TEMED	0.025	0.025	0.025	0.015
1.5 % APS solution	4	4	4	0.9
Total volume	**50**	**50**	**50**	**15**

First, the Hoefer Multiple Gel Caster unit was assembled as described in the manufacturer's description [190]. The acrylamide separating gel solution (see Table 23) was then poured approximately 10 cm high in between the plates and then each gel was covered with approx. 300 µl butanol. After polymerization (about 20 minutes) the butanol was removed by pouring and the acrylamide stacking gel solution (see Table 23) was added to each gel. Then combs were inserted into the still liquid gel solution and the stacking gels were polymerized (about 20 min).

For electrophoresis the "Hoefer mini vertical gel electrophoresis unit" was assembled as described in the manufacturer's description [191]. Depending on the size of the comb used, different sample quantities could be loaded into a gel pocket. If a comb with ten teeth was used, then up to 25 µl sample per pocket could be applied. In the case of a comb with 15 teeth, 15 µl sample was used.

For sample preparation 5× Laemmli buffer was added to the samples and incubated at 95 °C for five minutes. The samples were then shortly centrifuged and filled into the pockets of the acrylamide gel with the use of a Hamilton microliter syringe. 5 µL of NEB's "Color Prestained Protein Standard" were filled into a free pocket. Gel electrophoresis was performed at 100 V for 30 minutes and was then continued at 200 V for additional 70 min to 80 min. The gel sandwich was then removed from the device and carefully opened. The

SDS gel was briefly washed with dH_2O and prepared for the subsequent staining (see Chapter 6.3.9). The unstained SDS gel could also be used for a Western blot (see Chapter 6.3.10).

6.3.9 **Coomassie staining**

For visualizing the protein bands on the SDS gel Coomassie staining was used. The SDS gel was briefly washed with dH_2O after the gel run and then coated with the Coomassie staining solution. The overcoated gel was heated up using a microwave set to 600 W for about two minutes but boiling should be avoided. The staining solution was then poured off and the gel was washed several times with Coomassie distaining solution and heated using the microwave for about one minute each time until most of the background staining of the gel was removed. The gel was then shaken for about two hours or overnight in the distaining solution to get a clear gel background and thus a high-contrast gel image. Afterwards the gel was placed on an LED white light stage and the camera system was used for documentation.

6.3.10 **Western blot**

The Western Blot is a method of transferring proteins which have previously been separated by gel electrophoresis using an SDS-PAGE (see Chapter 5.2.2) according to their molecular weight, onto a membrane. For the transfer of the proteins from the acrylamide gel to the nitrocellulose membrane the semi-dry blotting system of the company VWR was used. For this, the nitrocellulose membrane and six pieces of blotting paper were cut to the size of the acrylamide gel (approx. 8.5 cm × 9.0 cm). The nitrocellulose membrane, the blotting paper and the acrylamide gel were then separately equilibrated for five minutes with the semi-dry transfer buffer. After this equilibration, the materials required for the blot were stacked in the following order on top of the anode (red) of the blotting apparatus: three blotting papers, followed by the nitrocellulose membrane, followed by the acrylamide gel and three blotting papers on top of the acrylamide gel. With the use of a reaction tube, possible air bubbles were pushed out of the stacked layers by rolling movements. The cathode plate was then placed on top of the stack and a dish of ice water was placed on top of the apparatus for cooling. The transfer was performed at 240 mA for 20 minutes (corresponds to 3.1 mA/cm^2). Following the transfer, the membrane was placed in a closable dish that fits as accurately as possible, with the blotted side facing upwards. The visible bands of the protein standard were marked using a soft pencil. The membrane was then washed twice for ten minutes with approx. 20 ml TBS buffer by shaking it gently. The membrane was then incubated through gentle shaking, for one hour at room temperature or overnight at 4 °C with approx. 10 ml TBS blocking buffer. After this blocking step, the membrane was washed once with 20 ml TBS buffer for ten minutes before incubating with the primary antibody. The primary antibody solution consisted of 10 ml TBS blocking buffer

with a certain amount (see Table 9) of the antibody solution. The incubation of the membrane with the primary antibody was performed for one hour at room temperature under gentle shaking. The membrane was then washed twice with 20 ml TBST buffer each time for ten minutes, followed by a further washing step with TBS buffer under the same conditions. Afterwards the incubation with the secondary antibody was performed. Due to the conjugating of the secondary antibody with the horseradish peroxidase (HRP) enzyme, the target proteins could be detected by a chemiluminescence reaction. The secondary antibody was diluted as described in Table 9 in 10 ml TBS blocking buffer, whereupon the membrane was incubated for one hour at room temperature under gentle shaking with this secondary antibody solution. After the final antibody incubation, the nitrocellulose membrane was washed four times using 20 ml TBST buffer for ten minutes.

The "SuperSignal West Pico PLUS Chemiluminescent Substrate" kit from Thermo Scientific was used for the detection of the target proteins labelled by the antibodies [192]. In deviation to the manufacturer's recommendations, 10 µl per square centimeter of membrane of the "Stable Peroxide Solution" and the "Luminol/Enhancer Solution" were first mixed with each other and then evenly applied to the nitrocellulose membrane. After a five-minute incubation at room temperature, the wetted membrane was placed in between two clear plastic sheets and any trapped air bubbles were pushed out. The detection of the emitted light was done using the "Fusion analysis camera system" of the company Vilber. The exposure time of the sensor was varied between ten seconds and 15 minutes depending on the expected signal strength. In addition, a reflected-light image of the membrane had to be captured to ensure that the size of the detected bands could be determined. This was done by subsequent digital overlaying of the two images.

6.4 **Special rAAV methods**

6.4.1 **rAAV production with adherent HEK293 cells using $CaCl_2$ transfection method**

In general, rAAVs were produced using adherent growing HEK293 cells. In this thesis only the helper-free plasmid system was used, which had also been further developed. Below the $CaCl_2$ transfection of HEK293 cells is shown, which was performed using 100 mm dishes due to the adherent character of the HEK293 cells. Experience had shown that, timing and fast as well as concentrated working played an important role in this method in order to achieve high transfection efficiencies. Therefore, the preparation of a transfection batch will be discussed here in detail, which is sufficient for the transfection of (only) four 100 mm dishes. This scale was the maximum that should be prepared at the same time. If the rAAVs should be purified by ultracentrifugation, then a production scale of twelve to 40 dishes was used, whereby a multiple of four plates was recommended, which also simplified

subsequent working steps. A few days before transfection, HEK293 cells were expanded in T-175 flasks in order to get enough cells available for the transfection. On the day before transfection $3 \cdot 10^6$ cells per 100 mm dish in a total volume of 10 ml medium were seeded. Therefore, the cells were detached with trypsin as in normal passages, the cell number was determined and the cells were diluted in fresh growth medium (see Chapters 6.2.3 and 6.2.4). 24 hours later, the growth of the cells was checked by light microscopy. The growth surface should be evenly covered with HEK293 cells showing a confluence of about 50 %, furthermore the cells should be well attached and not clumped together. All further working steps were also performed under sterile conditions. Two Bijou sample containers were required for the transfection of four 100 mm dishes. 2 ml 2×HBS buffer were added in one sample container and 2 ml $CaCl_2$ solution (0.3 M) in another Bijou container. The required plasmid DNA with a total amount of 60 µg was added into the $CaCl_2$ solution, which was then mixed. If the wild-type rAAV2 or a fully modified variant were produced, the three plasmids RC-plasmid, GOI-ITR plasmid and pHelper were used in a molar ratio of 1:1:1. For the production of VP2-modified mosaic rAAV2s the four plasmids CMV_VP2, Rep2Cap2_VP13, GOI-ITR plasmid and pHelper were used in a molar ratio of 1:4:5:5. The molar mass was approximately determined by about the size of the plasmids (in bp). A more detailed explanation of the plasmid system for the production of rAAV can be found in Chapter 5.9.3. A list of all plasmids is given in Table 11. The DNA-$CaCl_2$ solution was then slowly added to the 2×HBS solution drop by drop. Then the mixture of both solutions was vortexted for about ten seconds. Immediately after vortexing, 1 ml of this transfection solution was applied evenly, drop by drop, to each of the 100 mm HEK293 dishes. The plates were then slightly swiveled for a better distribution of the transfection solution. The HEK293 cells treated this way were further incubated for 72 hours at 37 °C. In the context of this thesis GOI-ITR plasmids (e.g. pZMB0311 or pZMB0522) were mostly used for the production of rAAV, which caused the expression of a fluorescent protein (mVenus) under the control of a strong promoter (CMV promotor), thus this fluorescent protein was produced as a by-product of rAAV production. This could be used to estimate the transfection efficiency of the rAAV production before further processing steps would be done. For this purpose, some samples of the production were analyzed by using fluorescence microscopy. The percentage of fluorescent cells and thus also the percentage of probably rAAV-producing cells should be in the range of 80 % to 100 %. The cells were then detached from the surface using a cell scraper and the entire contents of four dishes were combined in a 50 ml tube. The samples were then centrifuged at 2000 rpm (~300×g) for ten minutes to separate the supernatant from the cell pellet. At this point, the rAAV-containing supernatant and the rAAV-containing cell pellet could be stored separately at -80 °C until further processing. If the processing work should be continued directly, then all samples were stored on ice in the meantime. For further processing, a crude cell lysate was produced out of the cell pellet (see Chapter 6.4.2) and the rAAVs from the supernatant

were precipitated by ammonium sulphate precipitation (see Chapter 6.4.3). Both samples were then combined and further processed by ultracentrifugation (see Chapter 6.4.4).

6.4.2 **Preparation of a crude cell lysate from HEK293 cells**

The produced cell pellet (see Chapter 6.4.1) was thawed in a water bath at 37 °C and AAV lysis buffer was added. The amount of AAV lysis buffer depended on the number of pellet samples that should be purified together in one ultracentrifugation tube. Since the maximum sample volume in an ultracentrifugation tube was approx. 4 ml, the total amount of AAV lysis buffer in this step should therefore not exceed 3 ml. Each pellet sample was vortexed for about one minute. Then the samples were frozen again at -80 °C and thawed in a water bath two additional times (three times in total). After every thawing step, the samples were vortexed for one minute. Following these freeze-thaw cycles, 0.8 µl Benzonase (≥250 $U \cdot \mu l^{-1}$) was added to each pellet sample. The samples were incubated for one hour at 37 °C under light shaking. A CHAPS solution was then added from a freshly prepared CHAPS stock solution (10 % (w/v) CHAPS in H_2O) to achieve a final CHAPS concentration of 0.5 % (w/v). This solution was incubated at 37 °C for 30 min and should prevent a clumping of the rAAVs. The suspension was then centrifuged at 3000×g and 4 °C for ten minutes. The resulting supernatant (HEK293 cell lysate) was used to resuspend the precipitated AAV-containing pellet from the processed supernatants (see Chapter 6.4.3).

6.4.3 **Ammonium sulfate precipitation of rAAV from HEK293 supernatant**

After the transfection of the HEK293 cells with the necessary plasmids for rAAV production and subsequent three-day incubation, not only the HEK293 cells contained rAAV but the generated supernatant of the culture. Since a lot of media supernatant was generated in larger productions, it was necessary to precipitate the rAAVs from the supernatant for further processing, e.g. by ultracentrifugation, to achieve a volume reduction. An ammonium sulphate precipitation was used. After harvesting the cells, the supernatant was separated from the cells by centrifugation and stored at -80 °C (see Chapter 6.4.1). This supernatant was thawed, and 12,52 g ammonium sulphate were added per 40 ml supernatant. The mixture was vortexed until the salt had been completely dissolved. The tube was then incubated on ice for one hour. After this incubation the tube was centrifuged for ten minutes at 4 °C and 8300×g to obtain the pellet with the precipitated rAAVs. The resulting supernatant was discarded. The rAAV-containing pellet was resuspended with the HEK293 cell lysate which was produced parallel (see Chapter 6.4.2). It was important to not exceed a total volume of 4 ml as this was the maximum sample volume for an ultracentrifugation tube. The resulting sample could be stored overnight at 4 °C but should

be further processed not later than on the following day to avoid rAAV damage (see Chapter 6.4.4).

6.4.4 **rAAV purification using ultra centrifugation**

For the purification of produced prepurified rAAV precipitation (see Chapter 6.4.2 and 6.4.3) an ultracentrifugation with a discontinuous iodixanol gradient was performed. Table 24 lists the iodixanol solutions and their composition required for the stacked gradients for two ultracentrifugation tubes. The basis for this was the Optiprep solution, which is a 60 % (v/v) solution of iodixanol. First the 54 % (v/v) iodixanol solution was produced from the Optiprep solution and then the other iodixanol solutions were made out of this solution.

Table 24: Composition of iodixanol solutions for two ultracentrifugation tubes

Component	**Volume /ml**				
	60 %	**54 %**	**40 %**	**25 %**	**15 %**
Optiprep	3.0	8.0	-	-	-
54 % solution	-	-	4.0	2.5	1.5
10× PBS-MK	-	0.89	-	-	-
1× PBS-MK	-	-	1.4	2.9	1.2
1× PBS-MK + 2 M NaCl	-	-	-	-	2.7
Phenolred*	+	-	-	+	-

*: Only a very small spatula spite was added to the respective solution in order to distinguish the individual layers better after ultracentrifugation due to the staining.

After the individual solutions for the step gradient had been prepared, the centrifuge lids were attached to the ultra-centrifuge plastic tubes and then loaded.

For loading a 5 ml syringe with a 0.6×80 mm needle was used. First 1.25 ml of the bubble-free 60 % (v/v) iodixanol solution were used. The needle of the syringe was inserted through the opening in the lid of the tube and the solution was pushed out of the syringe very slowly. For the second layer 1.25 ml of the 40 % (v/v) iodixanol solution was used. This layer had to be added very slowly and drop by drop so that the gradient layers would not mix. In order to achieve this, the flat side of the needle was pressed against the inner wall of the tube while the tip was held directly above the surface of the solution. For the following 25 % (v/v) iodixanol layer a total of 1.5 ml of the corresponding solution was used and for the last iodixanol layer 2.25 ml of the 15 % (v/v) solution was used. The layering procedure was identical to that of the second layer. Finally, approx. 4 ml of the rAAV-containing sample (see Chapter 6.4.3) was applied to the top of the last layer. Two centrifuge tubes were balanced against each other using the Mettler Toledo XA205 DualRange precision scale. PBS buffer was used as the balancing solution. The maximum weight difference between two tubes should be 1 mg. Subsequently, the tubes were

inserted into the T-880 rotor of the Sorval Ultra Pro 80 ultracentrifuge and ultracentrifugation was performed with the following settings. A vacuum was built up in the centrifuge chamber and the running speed was set to 60000 rpm. Centrifugation was performed at 18 °C for two hours. When decelerating, the brake was switched off at a speed of 2000 rpm. After centrifugation, the sample tubes were clamped in a holding device. The outside of the tube was wiped off with a 70% (v/v) ethanol solution. A 21G needle was used with a 1 ml syringe. To collect the layer that contained the rAAVs, the needle was pricked into the tube at an angle of 90° near the border between the 60 % (v/v) and the 40 % (v/v) iodixanol layers (this layer could be identified by their phenol red coloration). The intermediate layer that contains rAAV could thus be collected completely. Alternatively, the needle could be guided down to the bottom of the tube and 100 µl fractions (starting with the lowest layer) were taken up to about half of the 40 % (v/v) iodixanol layer. After the rAAV-containing layer had been collected, the solution was diluted to a total volume of 4 ml with sterile buffer (e.g. PBS). This suspension was sterile filtered using a small 0.45 µm sterile filter and transferred to an Amicon Ultra-4 100 K filter concentrator. The sample was centrifuged at 2500×g until it was concentrated to a volume of about 500 µl. Then the Amicon concentrator was filled up again to 4 ml with sterile buffer. The concentration process was repeated as described above, in order to dilute the remaining iodixanol out of the sample. This procedure was repeated six times in total. The AAV sample was then aliquoted to 50 µl each and stored at -80 °C. In addition, a PCR tube containing a total of 5 µl of the rAAV sample was also frozen separately at -80 °C. This sample was used for subsequent titer determination by qPCR (see Chapter 6.4.5) [134].

6.4.5 **Determination of rAAV genomic titer with qPCR**

The genomic titer of an rAAV sample was determined using qPCR (quantitative polymerase chain reaction) by the comparison of the signal from the sample with a plasmid standard series. The primers for the qPCR reaction (qPCR-hGH-for (5'-CTCCCCAGTGCCTCTCCT-3') and qPCR-hGH-rev (5'-ACTTGCCCCTTGCTCCATAC-3')) were chosen to amplify a 150 bp long region within the hGH poly-A sequence. The hGH poly-A sequence was part of all GOI-ITR plasmids used (e.g. pZMB0311 and pZMB0522), so it was packaged in all correct produced rAAVs and was therefore universally suitable as a target region for rAAV quantification. For this purpose, the separately frozen rAAV sample (5 µl) was first digested using DNaseI to remove any DNA contamination caused by broken rAAV capsids. The composition of a DNaseI digestion for the qPCR sample is shown in Table 25.

Table 25: Composition of a DNaseI digestion for qPCR sample preparation

Component	**Volume /µl**
DNaseI buffer (10×)	5
DNaseI (2 U·µl^{-1})	5
rAAV sample	5
Nuclease-free water	35
Total volume	**50**

This mixture was incubated at 37 °C for 30 min and then DNaseI was inactivated at 75 °C for 20 min. After the DNaseI digestion, the samples were diluted by adding 200 µl of nuclease-free water (1:5 dilution). These samples were vortexed for ten seconds and then used as the sample material for the qPCR reaction. The qPCR reactions were performed in 96-well format using the LightCycler 480 II from Roche [193], all needed components were mixed under the PCR hood using sterile low retention pipette tips. Every diluted sample was mixed with 2.5 µl of primer qPCR-hGH-for and 2.5 µl of primer qPCR-hGH-rev, each with a stock concentration of 4 µM, and 10 µl of 2× GoTaq qPCR Mastermix from Promega [194]. Beforehand, a mastermix solution was prepared from the primer solutions and the GoTaq qPCR Mastermix and was added to the wells of the sample plate. Then 5 µl each of the previously generated sample solutions were added. All samples were analyzed as duplicates, nuclease-free water served as a negative control and a sample of the previously generated plasmid standard series served as a reference point for later quantification (see Chapter 6.4.6). After the samples had been applied to the plate, the plate was sealed with an adhesive foil and then briefly centrifuged to combine the samples and the master mix solutions. The qPCR reaction was carried out using the LightCycler 480 II according to the program sequence shown in Table 26.

Table 26: Thermocycling conditions of a qPCR reaktion

Step	**Temperature /°C**	**Time /s**	**Cycles /-**
Initial Denaturation	95	600	1
Denaturation	95	15	40
Annealing + Elongation	60	30	
Melting Curve			-

Following the qPCR reaction, an alignment of the reference sample with a previously generated plasmid standard series (see Chapter 6.4.6) could be used by the LightCycler 480 software to draw conclusions on the rAAV titers of the analyzed samples. For this previous dilutions (50-fold) of the samples and the double-stranded nature of the plasmid DNA compared to the single-stranded DNA packaged in the rAAVs (which led to a halving

of the rAAV signal compared to the plasmid standard at equal particle count) must be taken into account (adapted from [195]).

6.4.6 **Preparation of qPCR plasmid standard curve for rAAV genomic titration**

As already mentioned in Chapter 6.4.5, a plasmid standard series was used as a reference for the exact determination of the genomic rAAV titer using qPCR. This plasmid standard series formed the basis for the rAAV quantification and should therefore be prepared with the highest possible precision. For the plasmid standard series, a plasmid was used which served as the GOI-ITR plasmid (usually pZMB0311 or pZMB0522). Both plasmids differed only in their backbones among other sequences, they carry the hGH pA sequence, which contained the target sequences for the primers used in the qPCR. A fresh midi plasmid preparation of the GOI-ITR plasmid was used, which was eluted with nuclease-free water (see Chapter 6.1.4.2). The concentration of this plasmid sample was determined in triplicates by Nanodrop measurements using a quartz glass cuvette (see Chapter 6.1.6). The plasmid solution was pre-diluted to a round number of plasmids per µl (e.g. 10^{11} plasmids per µl) using the molecular mass of the plasmid determined by the Geneious software. All dilution steps were performed under the PCR hood using low retention pipette tips. The dilutions were carried out within the highest accuracy range of the pipettes (upper third of the adjustable scale) and no dilutions higher than 1:10 were made. The concentrations of the pre-dilutions were controlled by Nanodrop. A dilution series was generated starting at 10^{10} plasmids per µl down to 10^{1} plasmids per µl. Each individual solution had a total volume of 0.9 ml to 1.0 ml. After each pipetting step, the resulting plasmid solution was vortexed for about one minute to achieve a good mixing of the solution. The produced plasmid solutions were stored at -20 °C. From the dilution steps with a plasmid concentration of 10^{4} plasmids per µl and 10^{5} plasmids per µl 12 µl aliquots were prepared. These aliquots were used as references for later qPCR titrations. To generate the plasmid standard series, qPCR reactions of all dilution levels were performed in triplicates (see Chapter 6.4.5). The results are saved in the LightCycler 480 software as a standard series for further analysis (adapted from [195]).

6.4.7 **Determination of transducing titer by flowcytometry**

The transducing titer indicates the infectivity of the rAAV variant in an *in vivo* experiment. As this thesis mainly deals with rAAV2 wildtype viruses or variants based on this wildtype tropism, the cell line HT-1080 was ideal for the determination of the transducing titer. The cell line HT-1080 is generally considered the standard cell line for this type of titer determination. As already described in Chapter 5.12.2.2, the cell line HT-1080 was known

for an increased presence of HSPG on the cell surface, which is the natural primary target of AAV serotype 2.

For this purpose, 10^4 HT-1080 cells per well, in a total volume of 500 µl of the corresponding medium, were seeded in a 12-well plate and were incubated at 37 °C and 5 % CO_2 for one hour, so the cells can adhere. In the meantime, dilutions of the rAAV samples to be analyzed were prepared in PBS, which cover the desired range of MOIs, based on the genomic titers determined by qPCR. Depending on the experimental setup, MOIs in a range from 10 to 10^5 were interesting. To avoid influencing the cell growth by the added rAAV sample, its volume should be kept as low as possible and the same total volume (e.g. 50 µl) should be added every well. Then 50 µl of each rAAV sample and dilution were added to the cells, 50 µl PBS serves as a negative control. Duplicates were made of every sample. The plate was then carefully swiveled so that the rAAV samples were distributed more evenly in every well. The cells were then cultivated at 37 °C for further 72 hours. 24 hours after rAAV addition, 500 µl medium were added to each well. If one of the GOI-ITR plasmids pZMB0311 or pZMB0522 was used to produce the used rAAVs, the cell infection by this rAAVs led to the expression of the yellow fluorescent reporter protein mVenus. The amount of HT-1080 cells transduced this way could be determined by flowcytometric analysis of the resulting fluorescence.

Following the 72-hour incubation, the cells were prepared for the flowcytometric measurement. For this purpose, the medium was carefully removed from each well and the cells were washed twice with 1 ml PBS. Then 50 µl trypsin solution were added to each well and the plates were incubated at 37 °C for five minutes. The trypsin reaction was stopped by adding 1 ml PBS supplemented with 10 % (v/v) FCS per well. The cells were separated by gently pipetting up and down and then transferred into a sample tube. The flow cytometer (BD FACSCalibur) was operated as described in the manufacturer's manual [193]. The forward scatter diode (FSC) and the side scatter PMT (SSC) (photomultiplier) were used for a size correlation and determination of the granularity of the cells. The fluorescence detector BP530/30 nm (FL1) was used for the detection of mVenus or GFP (green fluorescent protein) signals. The device was controlled via the BD CellQuest Pro software. 10^4 cells were recorded from each sample, without the use of a gate. The first step was the measurement of the negative control, the gain of the fluorescence signal was adjusted according to this sample so that the signal peak was lower than 10. The data was then evaluated using the "Flowing Software". Cells were regarded as successfully transduced if their fluorescence signal was stronger than the fluorescence limit value given by the negative control. This limit value defines the fluorescence signal intensity, which was as high that 99 % of all cells from the negative control had a lower fluorescence value. Thus, the maximum false positive error was 1 %. For the calculation of the transducing titer, the two MOIs were used for every sample, whose transduced cell amount was less

than 100 % but greater than 1 %. The calculation was done according to the following formula:

$$TU = \frac{n \cdot F_p \cdot d}{V}$$

TU: Transducing Units (ml^{-1}); n: number of cells in transduction batch; F_p: Amount of positive fluorescent cells; d: dilution factor of rAAV sample; V: volume of transduction batch (ml)

6.4.8 **rAAV2 capsid ELISA**

The rAAV2 capsid ELISA (enzyme-linked immunosorbent assay) was normally used to quantify rAAV2 capsids using an rAAV capsid standard. In the context of this thesis, a slightly modified variant was used to determine the functionality of generated antibody fragments (A20-scFv). First rAAV2 wt was immobilized on a Nunc MaxiSorp 96-well flat-bottom plate. Based on a rAAV2 wt sample purified by ultracentrifugation (see Chapter 6.4.4), a rAAV2 wt suspension with a titer of $8 \cdot 10^8$ vg·ml^{-1} was diluted using PBS. Per well 100 µl of this rAAV2 wt suspension were used for immobilization (total $8 \cdot 10^7$ vg per well). All samples were performed in triplicates. The MaxiSorp plate was covered with aluminum foil and ws incubated at 4 °C over night. The rAAV2 suspension was then removed and each well was blocked for one hour at room temperature with 200 µl blocking buffer (PBS + 0.8 % (w/v) BSA). After blocking, each well was washed three times with 200 µl wash buffer (PBST + 0.8 % (w/v) BSA). The primary antibody was then diluted in wash buffer. At this point different dilutions and combinations of antibodies or antibody fragments were used. Based on the complete A20 antibody, a 1 to 250 dilution of the 50 µg·ml^{-1} stock solution in wash buffer was used. Depending on the experimental question, the quantity of A20-scFv_hIgG1-Fc_His_6 used was adjusted accordingly. Each well was incubated with 100 µl of the primary antibody solution (or mixture) for 1.5 h at room temperature. Following this incubation, the primary antibody solution was removed, and each well was washed three times with 200 µl wash buffer. Subsequently, an incubation with a 1 to 2500 dilution of the secondary HRP-coupled antibody in wash buffer was performed. Should the complete A20 be detected, an anti mouse Fc antibody was used. In the case of A20-scFv_hIgG1-Fc_His_6, an anti human Fc antibody was used. The incubation with the secondary antibody was performed for one hour at room temperature. Each well was then washed again three times using 200 µl wash buffer and finally tapped carefully to get rid of liquid remains. Then 150 µl of the ABTS buffer mixed with ABTS were added per well. Two untreated wells served as blanks. For the incubation at room temperature the plate was covered with aluminum foil. After about 15 to 45 min, positive samples should show a noticeable green color change of the originally colorless ABTS solution. For quantification,

absorption was measured at 405 nm using a plate reader (PowerWave HT by BioTek) ([196, 197]).

6.4.9 **Determination of ITR sequences**

About 60 µg of GOI-ITR plasmid pZMB0522 were digested with 120 U MlyI in a total volume of 240 µl at 37 °C for 1.5 h (see Chapter 6.1.11) and separated on an agarose gel (see Chapter 6.1.15). DNA fragments including lITR and rITR sequences were isolated using gel extraction method (see Chapter 6.1.10). About 18 µg of each fragments were digested with 60 U BsaHI in a total volume of 120 µl at 37 °C for 2 h to split the sequences potentially forming the major stem loop of the ITR sequence in half. Resulting fragments were separated on an agarose gel and 5'- and 3'-parts of the lITR and rITR were isolated. Sufficient material for several sequencing reactions was recovered. The obtained four DNA fragments were analyzed by Sanger DNA-sequencing (see Chapter 6.1.17). The sequencing primers were SEQ-lITR-5: 5'-GAAATGTTGA ATACTCATAC TCTTCC, SEQ-lITR-3: 5'-ATGAACTAAT GACCCCGTAA TTG, SEQ-rITR-5: 5'-CCTAATCTCA GGTGATCTACC and SEQ-rITR-3: 5'-AACGCCTGGT ATCTTTATAG TCC.

6.4.10 **Atomic force microscopy characterization of rAAV**

Method was carried out by Marco T. Radukic.

Atomic force microscopy (AFM) measurements of rAAV2 wt and rAAV2_587_bla were performed on a Multimode 8 AFM (Bruker, Santa Barbara, CA, USA) with Tap300Al-G cantilevers (BudgetSensors, Sofia, Bulgaria) in tapping mode in air. 2 µl of sample in PBS were spotted onto freshly cleaved mica and incubated for one minute. The mica was then briefly rinsed with water and dried under a gentle nitrogen flow. Data analysis was performed with Gwyddion 2.48. Obtained images were treated with offset and plane correction algorithms and the size of visualized particles was measured at half maximum particle height. Statistical analysis of size measurements was performed using Excel 2016.

6.4.11 **rAAV stability assay**

An interesting property of rAAV variants is their thermal stability, which allows conclusions on the biological activity of the investigated particles to be drawn. The thermal stability of rAAV was often described using various methods, such as differential scanning calorimetry (DSC) or differential scanning fluorimetry (DSF). All these methods focus on the integrity of the capsid, the viral DNA, which has a particularly important function for biological activity, was normally not considered [198].

Based on methods already established within this thesis, a stability assay has been developed which had a special focus on the release of DNA from the AAV capsid when exposed to high temperatures. The samples examined were rAAV samples purified by

ultracentrifugation and whose genomic titer was determined (see Chapter 6.4.4 and 6.4.5). Technical duplicates were made for all samples. The rAAV samples were diluted to a genomic titer of $1.5 \cdot 10^9$ vg·ml^{-1} in PBS. Then 16 aliquots with 10 µl each were filled into PCR tubes from this rAAV stock solution. These aliquots were stored on ice before and after incubation. The temperature range selected for incubation laid between 37.5 °C and 75.0 °C in 2.5 °C increments. The incubation was performed using the thermocycler "peqSTAR 96 Universal Gradient" of the company peqlab, starting at a temperature of 8 °C. The thermocycler was heated up to the desired temperature at maximum speed, held the temperature for five minutes and then cooled the sample down to 8 °C at maximum speed. After this procedure was performed for all samples and at all temperatures, DNaseI digestions were prepared. These were performed in a total volume of 50 µl each, using the entire incubated rAAV sample (10 µl). The samples were then diluted with nuclease-free water as described and the genomic titers were determined by qPCR (see Chapter 6.4.5). Data analysis was performed using Origin2018. The total rAAV amount per reaction was normalized to 100 % and a logistic regression was performed (Logistic5: Amin = 0, Amax = 100, no weighting). The disintegration temperature (T_d) was determined as the temperature at which 50 % of rAAVs released its DNA.

6.4.12 **β-lactamase activity assays**

In the context of this thesis, rAAV variants were created that presented complete β-lactamase enzymes on their capsids. The activity of these β-lactamases was investigated quantitatively using a nitrocefin assay and qualitatively using a bacterial assay. Both methods are described in the following subchapters.

6.4.12.1 **Nitrocefin β-lactamase activity assay**

A nitrocefin assay was used to make a quantitative prediction about the β-lactamase activity of some rAAV variants. Nitrocefin is a substance belonging to the chemical group of cephalosporins, it changes its color from yellow to red when cleaved by β-lactamase. This spectrometric method was performed in 96-well format with a microplate spectrophotometer PowerWave HT of BioTek. For this purpose, 91 µl per well of the rAAV sample in PBS with a total number of $3.7 \cdot 10^8$ vg, or PBS as a negative control, were added per well. Then 9 µl of nitrocefin buffer were added to each well followed by mixing. The change in absorption was observed at 486 nm over a period of 40 min with a measurement interval of one minute at room temperature. All subsequent calculations were based on the Lambert-Beer' rule, which is part of the following formula:

$$\frac{\Delta A}{\Delta t} = \Delta\varepsilon_{486} \cdot c_{bla} \cdot k_{cat_{bla}} \cdot d$$

$\frac{\Delta A}{\Delta t}$: slope of linear regression from nitrocefin assay; $\Delta\varepsilon_{486}$: molar extinction coefficient of nitrocefin shift (here 16000 l·mol·$^{-1}$·cm^{-1}); c_{bla}: concentration of β-lactamase; $k_{cat_{bla}}$: catalytic constant of β-lactamase; d: distance (here 0.267 cm), either the turnover number (here 746 s^{-1}) or the concentration of active lactamases was assumed to be known.

6.4.12.2 **Bacterial β-lactamase activity assay**

A bacterial assay was established for the qualitative detection of the β-lactamase activity of some rAAV variants. In the bacterial assay an aliquot of about 200 µl of an *E. coli* DH5α culture transfected with plasmid pZMB0084 harboring a constitutively expressed chloramphenicol acetyltransferase and a red fluorescent protein (RFP) was plated on a LB agar dish supplemented with ampicillin (final 100 µg·ml^{-1}) and chloramphenicol (final 20 µg·ml^{-1}). After the bacteria suspension had been aspirated into the agar, 3 µl of rAAV samples and some dilutions were spotted on the dish, a PBS sample served as a negative control. After the sample drops had been aspirated into the agar, the dish was incubated at 37 °C overnight. Due to the lack of an ampicillin resistance, the *E. coli* cells were only able to grow, when ampicillin was hydrolyzed by active β-lactamase from the rAAV samples. Growing *E. coli* colonies could be easily detected by eye via their red appearance, because of the RFP expression.

6.4.13 **rAAV purification by PKD2 affinity chromatography**

6.4.13.1 **Preparation of regenerated amorphous cellulose (RAC)**

The starting material for the preparation of regenerated amorphous cellulose (RAC) was microcrystalline cellulose (Avicel PH-101, ~50 µm particle size, Sigma Aldrich). The used cellulose was dissolved in ortho-phosphoric acid and later precipitated with water. The hydroxy groups of the cellulose were first esterified with the phosphoric acid, whereby the cellulose was dissolved. By adding water, the reaction was reversed so that the cellulose precipitated. Compared to microcrystalline cellulose, RAC had a significantly increased surface area and thus potentially a higher binding capacity for fusion proteins with cellulose binding domains. The preparation of RAC was performed as shown below:

For the preparation of RAC, 200 mg Avicel were wetted with 600 µl Milli-Q-water before adding 10 ml ice-cold 85 % (v/v) ortho-phosphoric acid slowly in steps of 2 ml while stirring. The milky suspension was incubated on a roller tumbler mixer for 1 h at 4 °C, within a few minutes the suspension became clear. Subsequently, 40 ml of ice-cold Milli-Q water were added slowly in steps of 10 ml while stirring, resulting in a white precipitate. Following a centrifugation step (4600×g, 20 min, 4 °C, using a swing-out rotor), the

supernatant was discarded and the white RAC pellet was resuspended in approx. 45 ml ice-cold Milli-Q water. This washing step was carried out a total of four times. To neutralize any residual acid, the RAC pellet was resuspended in 45 ml ice-cold Milli-Q water and 500 µl 2.0 M Na_2CO_3 were added. Subsequently, a centrifugation step was performed again (4600×g, 20 min, 4 °C, with swing-out rotor), the supernatant was removed, and the pellet was resuspended in 45 ml ice-cold Milli-Q-water. This washing step was repeated a total of two times, after the last washing step the pH value of the supernatant was determined using test strips. At a pH value in the range of pH 5.0 to 7.0, a 10.0 $g \cdot l^{-1}$ suspension was prepared which was stored at 4 °C until further use, or washed again if the pH value deviated [199].

6.4.13.2 **Binding of CBDcex fusion proteins to cellulose**

To bind fusion proteins, which had a cellulose binding domain (e.g. CBDcex), to cellulose, the following procedure was used.

In a suitable vessel a 10 $g \cdot l^{-1}$ suspension was prepared out of the cellulose material e.g. Avicel or RAC together with the protein containing solution (or whole cell extract) in PBS. Since both RAC and the washed Avicel solution were present in a 10 $g \cdot l^{-1}$ stock solution in Milli-Q water, the required amount had to be measured off beforehand and the supernatant was removed by centrifugation (11000×g, 1 min, at room temperature). In this case the large exclusion volume of the RAC pellet had to be taken into account, so the suspension was filled up with the protein-containing solution to the intended volume. The suspension was incubated over night at 4 °C on a roller tumble mixer. Following this incubation, the suspension was centrifuged at 11000×g for one minute at room temperature and the supernatant was discarded. The pellet was washed three times with PBS. In each case, twice the volume of PBS compared to the initial suspension volume was used. After buffer addition the suspension was briefly vortexed and then centrifuged again (11000×g, 1 min, at room temperature). After the last washing step, the resulting chromatography material could be used directly as a 10 $g \cdot l^{-1}$ suspension (final) for further experiments. A possible storage of the fresh material was not investigated, as it was prepared fresh for each experiment.

6.4.13.3 **Binding of rAAV2 to cellulose affinity chromatography material**

In this thesis, binding experiments of rAAV2 particles to cellulose-based affinity chromatography materials were performed. For this purpose, the whole cell extract (WCE) containing His_6_PKD2_CBDcex from *E. coli* was bound to Avicel (see Chapter 6.4.13.2). Deviating from this method, a 20 $g \cdot l^{-1}$ suspension of loaded Avicel was prepared in PBS, after the last washing step. 25 µl of this suspension were mixed with 25 µl of a previously

per ultracentrifugation purified rAAV2 sample (approx. $1 \cdot 10^{10}$ vg·ml^{-1} in PBS) in a 2 ml reaction vessel. The samples were then incubated on a heat block with orbital shaking function at room temperature (20 °C) and 1000 rpm. The high shaking frequency in combination with the round bottom of the 2 ml reaction vessel ensured a good mixing of the rather small sample. Several batches were incubated simultaneously for different incubation periods between 5 min and 120 min. After incubation, the suspension was immediately centrifuged for one min at room temperature at 11000×g and the supernatant was removed and stored on ice. The cellulose pellet was briefly washed by adding 500 µl PBS and was then centrifuged again under the same conditions. The supernatant was then removed, and the cellulose pellet was stored on ice. To quantify the unbound rAAV2 particles, DNaseI digests from the supernatant samples and qPCR quantifications were performed as described in Chapter 6.4.5. For the quantification of the bound rAAV2 particles, the washed cellulose pellet was completely absorbed in 50 µl DNaseI-master-mix (5 µl DNaseI (1 U·µl^{-1}), 5 µl 10× DNaseI buffer and 40 µl Milli-Q-water per sample) and subsequently treated and quantified as described in chapter 6.4.5.

6.4.13.4 **Covalent binding of proteins to blotting paper**

In order to create a covalent bond between proteins and cellulose (in this case blotting paper), these had to be activated first using divinyl sulfone. This cellulose activation was carried out as follows:

First, round pieces with a diameter of 10 mm were punched out of a sheet of blotting paper. The paper had a thickness of 2 mm and an average weight of about 25 mg. About 40 to 50 pieces were incubated in a 50 ml reaction tube together with 10 ml DVS buffer (10 % divinyl sulfone, 0.1 M sodium carbonate, pH 11.0) for two hours at room temperature on a roller tumbler mixer. Following this incubation, the buffer was poured off. The paper pieces were then washed five times by adding 20 ml Milli-Q water. After a short shake, the Milli-Q water was poured off. The pieces of blotting paper were then separately dried on a plastic petri dish at room temperature for about two hours [200].

The protein-containing solution was then added to the paper pieces. In this case 100 µl of a 5 mM PKD2_MBP_GSPT_CBDcex_His_6 solution in PBS were pipped onto each piece of blotting paper, so that it was completely soaked. The petri dish was closed with the corresponding lid and placed in a larger petri dish filled with water, so that a moisture chamber was made. The moisture chamber was also closed and covered with a wet towel. After an overnight incubation, the paper pieces were transferred to a 50 ml reaction tube, which was then filled with 40 ml blocking buffer (PBST). An incubation of about one hour at room temperature was carried out on a roller tumbling mixer. Subsequently, the blocking solution was discarded, and the samples were washed five times with approx.

40 ml PBS each time. Until further use, the pieces of paper loaded this way were stored on ice in PBS buffers and were used fresh.

6.4.13.5 Binding and elution of rAAV to PKD2 affinity chromatography material

This method was developed to compare the affinity chromatography material based on blotting paper whose production is described in Chapter 6.4.13.4, with the POROS AAVX affinity chromatography material. Therefore, the conditions were adjusted as closely as possible.

All incubation steps were carried out at room temperature, gently shaking the respective mixtures on a roller tumbler mixer. Since the loaded blotting paper was washed with PBS and the used rAAV samples were available in PBS, no equilibration was necessary.

Then 500 µl of the rAAV sample to be tested were added to one piece of activated blotting paper in a 2 ml reaction vessel and were incubated for 1.5 h. The rAAV samples were already pre-cleaned by ultracentrifugation and buffered in PBS (see Chapter 6.4.4). The used genomic titers ranged from $5 \cdot 10^9$ vg·ml^{-1} to $8 \cdot 10^{10}$ vg·ml^{-1}. The flowthrough (FT) was collected by pipetting. This was followed by a total of three washing steps. For this, the samples were incubated with 500 µl PBS for five minutes each. The three washing fractions were pooled and collected. The subsequent elution was performed by adding 500 µl PKD2 elution buffer (100 mM sodium citrate, pH 2.5) followed by an incubation time of ten minutes. The elution was repeated two times. Also, in this case the elution fractions were pooled and collected. Subsequently, the elution fractions were re-buffered against HBSS buffer using an Amicon Ultra 4 centrifuge filter and concentrated to a final volume of approx. 250 µl. Following this procedure, samples of the rAAV starting solution, the flowthrough, the pooled wash fraction and the pooled eluate were digested by DNaseI and titrated by qPCR as described in Chapter 6.4.5.

6.4.14 POROS AAVX static binding protocol by ThermoFisher Scientific

The following protocol was used to determine the binding capacity and binding ability of affinity chromatography materials, especially the POROS CaptureSelect AAVX Affinity Resin from ThermoFisher Scientific, for different AAV variants. It was provided in cooperation with ThermoFisher Scientific [140].

The test was performed in static mode using 50 µl resin in a 1.5 ml plastic reaction vessel. All experiments were carried out in duplicates. All incubation steps were carried out at room temperature, gently shaking the respective mixtures. After each incubation step, the resin material was separated from the supernatant by centrifugation (three cycles each for

30 seconds at 12000×g). First, 50 µl resin slurry were equilibrated by adding 500 µl PBS. The equilibration was performed for ten minutes and was repeated three times. Then 1 ml of the rAAV sample to be tested was added to the chromatography material and was incubated for 15 minutes. The rAAV samples were already pre-cleaned by ultracentrifugation and buffered in PBS (see Chapter 6.4.4). The used genomic titers ranged from $5 \cdot 10^9$ vg·ml^{-1} to $8 \cdot 10^{10}$ vg·ml^{-1}. After centrifugation the flowthrough was collected. This was followed by a total of three washing steps. For this, the samples were incubated with 500 µl PBS for ten minutes each. The three washing fractions were pooled and collected. The subsequent elution was performed by adding 500 µl elution buffer (100 mM sodium citrate, pH 2.5) followed by an incubation time of ten minutes. The elution was repeated three times, each elution fraction was neutralized directly by adding 166.7 µl neutralization buffer (1 M tris, pH 9.0). Also, in this case the elution fractions were pooled and collected. Following this procedure, samples of the rAAV starting solution, the flowthrough, the pooled wash fraction and the pooled eluate were digested by DNaseI and titrated by qPCR as described in Chapter 6.4.5.

7 Results and Discussion

7.1 Establishing rAAV2 production in the working group

At the beginning of the work on this thesis, the whole rAAV thematic was completely new for the labs of the University of Bielefeld. Prof. Dr. Kristian M. Müller provided the necessary knowledge and material from previous projects at the Universities of Freiburg and Potsdam. Nevertheless, it took an extensive familiarization phase, dominated by numerous unsuccessful attempts, until it was possible to achieve the first useful results. A detailed discussion of the establishment of all the necessary methods, starting with the production and purification of rAAV up to all analytical methods for the detection of the particles, is not discussed here. In the methods section of this thesis, emphasis was therefore placed on a comprehensible presentation of all established methods (see Chapter 6). At this point, a method should be pointed out which has given some interesting results and was therefore also included in the manuscript of the paper entitled “rAAV engineering for capsid-protein enzyme insertions and mosaicism reveals resilience to mutational, structural and thermal perturbations” - the sequencing of the ITR sequences.

7.1.1 ITR sequencing

As already mentioned, a helper-free plasmid system for the production of rAAV was used in the context of this thesis. The GOI-ITR plasmid plays an important role, because it delivers the DNA sequence, which is finally packaged in single-stranded form inside the rAAV capsid. The sequence to be packed is terminated by the left and right ITR sequence. The most commonly used GOI-ITR plasmid here was plasmid pZMB0522 (pUC19bb_ITR_EXS_pCMV_mVenus_hGHpolyA), which codes for the yellow fluorescent reporter protein mVenus under the control of the CMV promoter (see Figure 15A). All GOI-ITR plasmids carry the ITR sequences, which play an important role in the packaging of viral DNA. The correctness of these 145 bases long, identical but inverted sequences is essential for a successful rAAV production. However, parts of these sequence segments tend to be deleted, probably due to recombination events [201]. Due to long complementary sequence segments in this region, when present in single-stranded form, the ITR segments tend to form a Y-shaped stem-loop structure, which is graphically shown in Figure 15B for the left ITR sequence according to the energy model by Matthews *et al.* at a temperature of 50 °C (using Geneious 9.1.8 software). Due to this strong secondary structure formation, it was not feasible to amplify this DNA segment by PCR or to sequence it by conventional molecular genetic methods. This chapter provides a strategy which enables sequencing of the ITRs on the basis of common molecular genetic methods.

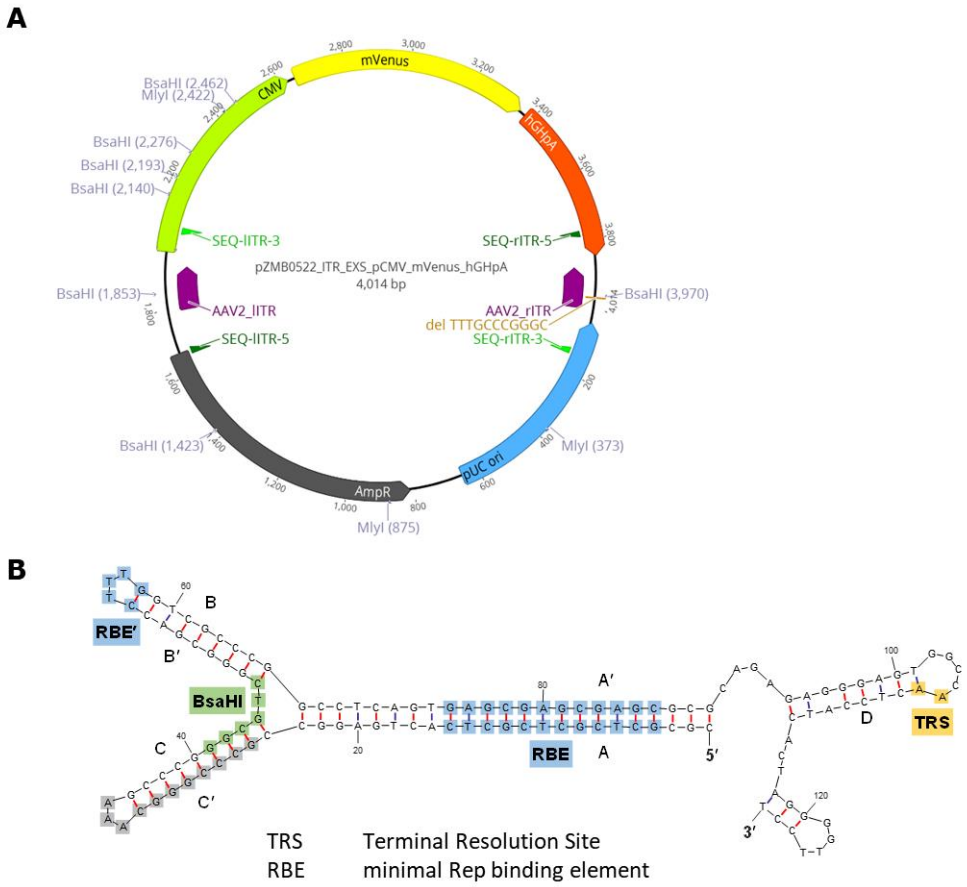


Figure 15: Schematic overview of the GOI-ITR plasmid pZMB0522 and an energy model of the left ITR sequence. A: pZMB0522 plasmid map with marked restriction sites BsaHI and MlyI; B: Sequence of the left ITR as present in the GOI-ITR plasmid pZMB0522. Regions, as well as the minimal Rep binding element (RBE) and the terminal resolution site (TRS) of the ITR are highlighted. BsaHI restrictions site is highlighted. Structure was computed using Mfold webserver (150 mM NaCl, 5 mM $MaCl_2$, 37°C).

This strategy is based on the recognition sequence for the restriction enzyme BsaHI (5'-GR/CGYC-3') which is located in the ITR sequence. By digesting the plasmid pZMB0522 with the enzyme BsaHI, those sequence segments can be separated, that make up the largest part of the stem structure, when the DNA sequence is single-stranded (see Figure 15B). By a previous digestion with the enzyme MlyI, sequence fragments can be generated, which contain the left and right ITR sequences which can be separated clearly from each other. In the following this procedure is described in detail for the plasmid pZMB0522.

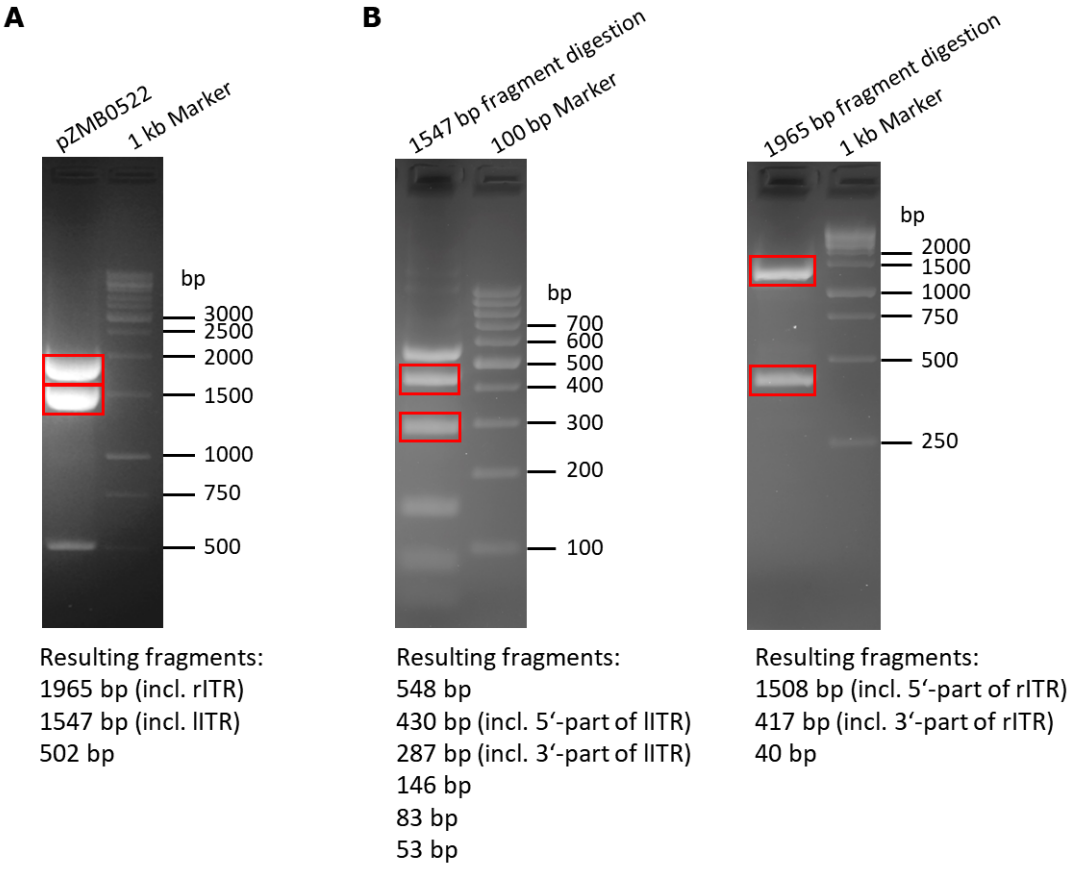


Figure 16: Agarose gel electrophoresis to separate digested DNA fragments of Plasmid pZMB0522 for ITR sequencing. A: The restriction digest of pZMB0522 with MlyI results in three fragments and both ITRs could be separated from the agarose gel as marked with red boxes; B: The larger fragments from the first digest are further digested with BsaHI to break up the Y-shaped stem-loop structure to enable Sanger DNA-sequencing after isolation of the marked fragments.

As the generation of the linearized DNA fragments for sequencing required two consecutive digestions with subsequent agarose gel extraction and with some of the target fragments are being quite small (smallest fragment is 287 bp), it was necessary to start with a relatively large initial amount of plasmid DNA. Therefore, in a first step 60 µg pZMB0522 plasmid DNA were digested with MlyI. An exact description of the entire procedure and the used quantities can be found in Chapter 6.4.9. The agarose gel on which the digested plasmid DNA was applied is shown in Figure 16A. During digestion, three DNA fragments with the sizes 1965 bp, 1547 bp and 502 bp were generated. The 1965 bp DNA fragment contained the sequence of the right ITR and the 1547 bp fragment the sequence of the left ITR. After an agarose gel extraction of these two fragments the DNA was digested using BsaHI and then separated again by agarose gel electrophoresis. These separations are shown in Figure 16B. During the BsaHI digestion of the 1547 bp DNA fragment, a total of six DNA fragments were formed, of which the 430 bp fragment contained the 5'-part of the left ITR and the 287 bp fragment contained the 3'-part. The corresponding digestion of the 1965 bp fragment yielded three DNA fragments, of which the 1508 bp fragment

contained the 5’-part of the right ITR and the 417 bp fragment the 3’-part. These four fragments were again isolated by gel extraction and then sequenced as linearized fragments using the corresponding sequencing primers (SEQ-lITR-5, SEQ-lITR-3, SEQ-rITR-5 and SEQ-rITR-3) (see Chapter 6.1.17). Alignments of the resulting sequences are shown in Figure 17 for all four fragments.

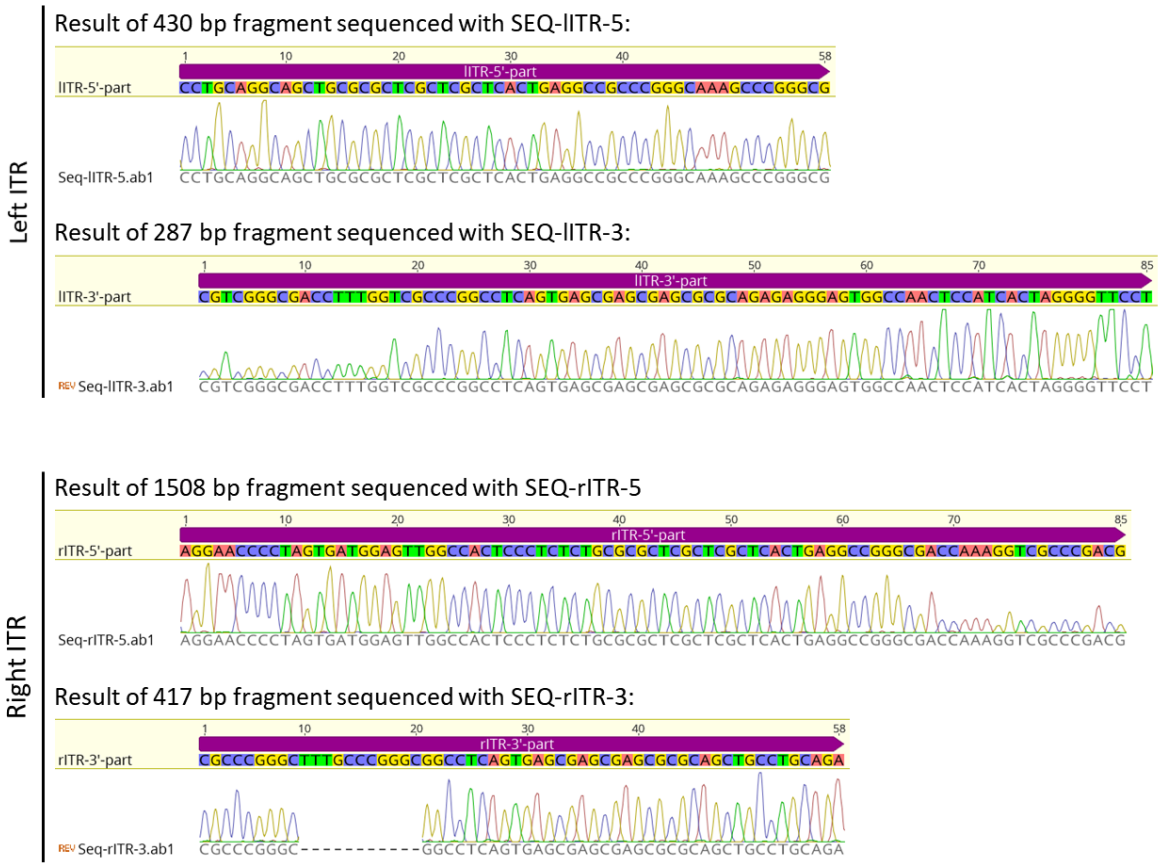


Figure 17: Results of Sanger-DNA sequencing aligned to original ITR sequences. Fragments were sequenced using the specified primer (see Table 10). The left ITR is completely intact. The 3’-part of the right ITR shows a deletion of 11 bp (5’-TTTGCCCGGGC-3’).

The alignments show the corresponding ITR sections up to the BsaHI restriction site. The left ITR is completely intact and has no errors. The 5’-part of the right ITR is also correct. Only the 3’-part of the right ITR shows an 11 bp deletion of the sequence 5’-TTTGCCCGGGC-3’. Due to the repetition of the sequence in this area, it could also be a deletion of the sequence 5’-GCCCGGGCTTT-3’. This 11 bp deletion is widely known and is present in many plasmids distributed by Addgene [202]. It also occurs in some commercial plasmids (e.g. pAAV_MCS from Cell Biolabs) [203]. Usually only one ITR is affected by this deletion. A possible explanation for this could be the position of the ITR on the plasmid in relation to the origin of replication. Most plasmids for ITR gene production use a pMB1-derived ori whitch works unidirectional with respect to RNA II elongation. About 200 bp downstream of the replication start there is a switch between PolI and PolII [204]. Often it is just this ITR located downstream from the ori that is affected by the observed deletion.

In several ITR plasmids including the pZMB0522 the distance between the ori-start (end of RNAII) and the deletion is about 90 bp and PolI might be more prone to induce deletions as well as the polymerase switch, which might take place near the observed deletion site. So, the distance of an ITR to the plasmid origin, the type of origin and the occurrence of Okazaki fragment processing sites may affect the genetic ITR stability during DNA amplification in *E. coli*. Nevertheless, within the context of this thesis it was possible to produce functional rAAV particles, although the 11 bp deletion was present in the GOI-ITR plasmid. It has even been reported that these or even larger deletions may have a negative effect on the amount of rAAV particles produced, but the expression of the transgene could be improved by this deletions [205].

7.2 **Systematic variation of loop modifications to determine rAAV2 engineering capacity**

In the context of this thesis and other projects of the working group, the plasmid system for the production of rAAV was continuously developed and optimized. The system offers, among many other features, the possibility to modify the capsid proteins both N-terminal and within the loop regions at positions 453 and 587. In particular, the insertion point in the loop region at position 587 is attractive for modifications in the context of rAAV2 capsid retargeting. Here the arginine's, which are responsible for the natural HSPG tropism of the serotype, are located in close proximity at the positions 585 and 588 [20]. In addition, this position is especially suitable for modifications due to its special exposure on the capsid structure. In the context of the plasmid system development, the question arose how much modifications could be made at this position without affecting the producibility of the rAAV and biological factors such as capsid stability. To answer this question, new plasmid constructs were generated for the production of rAAV2 variants, which carried different modifications at position 587. These modifications included flexible glycine-serine amino acid linkers with different lengths up to 16 amino acids or the insertion of a 29 kDa β-lactamase enzyme. Regarding the β-lactamase modification, two rAAV2 variants were created. One of them contains the β-lactamase sequence only in the VP2 capsid proteins. The other variant carries the β-lactamase insertion in all capsid proteins. The well-studied TEM β-lactamase was chosen, because of its small distance between the N- and C-terminus, which approximately matches the distance of the β-hairpin residues in the variable region of the VP protein [206]. Nevertheless, the enzyme β-lactamase offers the possibility to easily measure its activity, which opened new possibilities for detection.

The results of this project were published in shared authorship with Rebecca Feiner in the paper titled "rAAV engineering for capsid-protein enzyme insertions and mosaicism reveals resilience to mutational, structural and thermal perturbations" in the "International Journal of Molecular Sciences". The manuscript and related information are listed in Chapter 10.1.1 of the Appendix and its subchapters. In the following, the results that have contributed to this publication and further details are presented and discussed.

7.2.1 **Generation of DNA constructs for rAAV2 with loop insertions**

To generate the RepCap plasmids to produce the fully modified glycine-serine amino acid linker variant and the fully modified β-lactamase variant, the plasmid pZMB0154 (pSB1C3_001_Rep_VP123_453_587ko_p5tataless) was used as the starting plasmid. This plasmid originally came from the iGEM Freiburg 2010 plasmid collection and was listed as number P687. It contains, besides the modified RepCap CDS, the HSPG knock-out variant of the 587 insertion site (R585A and R588A). In addition, the complete backbone sequence, which is the 001-variant of the pSB1C3 backbone with all listed mutations (see Geneious

Database e.g. pZMB0084), was confirmed by sequencing. As already described in Chapter 5.9.3.2, the plasmid system allows insertions at loop position 587 on DNA level via the singular restriction sites BamHI and PvuII. Thus, the corresponding wild type variant of the plasmid and the glycine-serine linker variants were generated by oligonucleotide hybridization of the following oligonucleotide pairs: 587_wt-for/rev, 587_wt-GG-for/rev, 587_wt-GGSG-for/rev, 587_wt-2xGGSG-for/rev, 587_wt-4xGGSG-for/rev (see Table 10 and Chapter 6.1.8). In order to prevent recombination or errors in the DNA sequence, the designed DNA sequences for the glycine-serine linker were varied in deviation from the optimal human codon usage, avoiding longer sequence repetitions. The hybridized oligonucleotides were ligated via the restriction sites BamHI and PvuII into the plasmid pZMB0154. The resulting plasmids were stored under the numbers pZMB0216 to pZMB0220 (see Table 11). The cloning of the RepCap plasmid containing the β-lactamase coding sequence at insertion position 587 was based on the iGEM plasmid BBa_K404250 (pZMB0223). Amongst others BBa_K404250 codes for a thermostabilized variant of β-lactamase (TEM β-lactamase variant 14FM, see Table 27 for more details) which was amplified by PCR using the BamHI_wt_bla-for and PvuII_wt_bla-rev primers. The resulting BamHI and PvuII digested PCR product was inserted at position 587 of pZMB0154 like the hybridized oligonucleotides before and was finally stored as pZMB0221.

Table 27: Characteristics of TEM β-lactamase variant 14FM [207]

Characteristics	**TEM β-lactamase 14FM**
Mutations compared to wild-type TEM-116 β-lactamase (GenBank ID: AY425988)	V31A, A36L, L51I, R120G, E147G, H153R, V159T, M182T, L201P, I208M, E212K, A224V, A249V, T264M
Melting temperature T_M determined using far-UV circular dichroism in phosphate buffer	71.6 °C (2 µM) 72.7 °C (20 µM)
Catalytic constant k_{cat} for nitrocefin conversion in s^{-1}	746 ±76 s^{-1}

To generate mosaic viruses, which only carry insertions in the loop region 587 of the VP2 capsid proteins, the plasmid pZMB0315 (pSB1C3_001_pCMV_Kozak_VP2_453_587wtHis) served as the starting plasmid. This plasmid lacks the CDS of the Rep proteins and the coding sequence that distinguishes VP1 from VP2. VP2 is expressed under the control of the CMV promoter. The start codon of the VP3 CDS was knocked out and the transcription of the VP2 coding sequence is enhanced by the use of a strong Kozak sequence (GCC ACC) upstream of the VP2 start codon. The generation of this plasmid and the importance for the mosaic virus production is explained in the attached paper. As before, for the insertion of the β-lactamase sequence and the glycine-serine linker sequence (here only 4x(GGSG)), the corresponding PCR product or the hybridized oligonucleotides were inserted via the

restriction sites BamHI and PvuII into the plasmid pZMB0315. The resulting plasmids were stored under the numbers pZMB0577 and pZMB0601.

In order to produce mosaic viruses with the plasmids described above (pZMB0577 or pZMB0601), the gene sections that have been removed or knocked out, compared to the originating plasmids, had to be supplemented by a fourth complementing plasmid. The original system uses the plasmid pZMB0299 (pSB1C3_001_Rep_VP13_453_587ko_p5tataless) for mosaic virus production, which is very similar to the plasmid pZMB0154, because it also carries the mutations R585A and R588A. In this plasmid the VP2 starting codon ACG was mutated by a silent mutation, which suppresses the VP2 expression. For experiments with mosaic rAAV2 variants the wild type tropism of the VP1 and VP3 capsid proteins was desired, so the required complementing plasmid was generated based on pZMB0299 as follows. The DNA oligonucleotides 587_wt-for/rev which were already used for the generation of pZMB0216 were hybridized and inserted via the restriction sites BamHI and PvuII into the plasmid pZMB0299. The resulting plasmid was stored under the number pZMB0600.

An overview of the respective plasmids is shown in Figure 18.

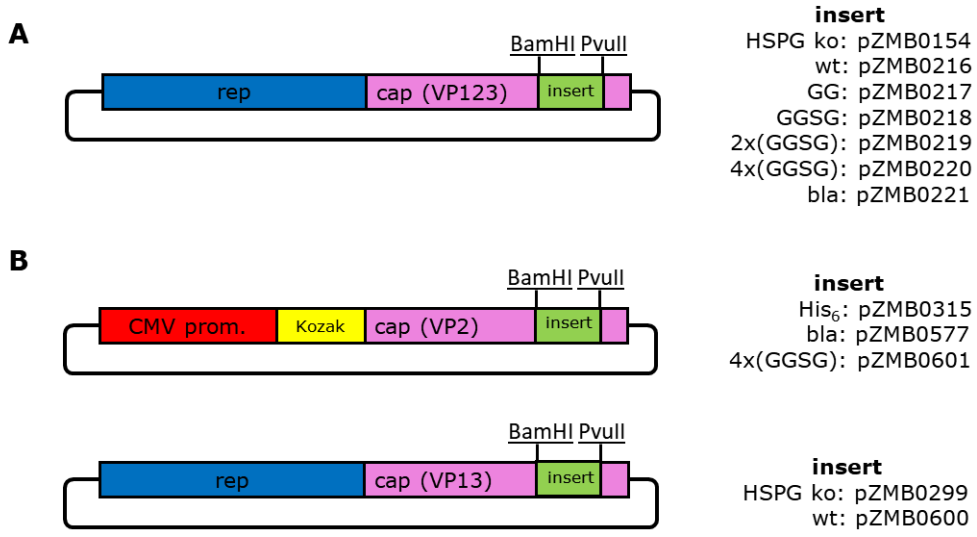


Figure 18: Schematic overview of plasmid constructs for rAAV loop insertions. A: plasmid constructs for full capsid modification; B: Plasmid constructs for rAAV mosaic production. Insert motifs for different plasmid numbers are indicated on the right site. All amino acid linker and bla constructs got the wild type tropism at residue number 586 and 588 (both arginine).

7.2.2 **rAAV2 variants with amino acid-linker loop insertions**

7.2.2.1 **Producibility and infectivity of rAAV2 with amino acid-linker loop insertions**

First the rAAV2 variants were produced, which carried the amino acid linker insertions at position 587, on all capsid proteins. For this purpose, adherent HEK293 cells were transfected with the corresponding plasmids using ten 100 mm dishes for each variant, as described in Chapter 6.4.1. For all batches the pHelper plasmid pZMB0088 and the GOI-ITR plasmid pZMB0522 were used. The plasmids pZMB0216 to pZMB0220 were used as RepCap plasmids. Thus, the four amino acid linker variants were generated as well as the wild type variant. Following production, the rAAVs were purified by ultracentrifugation and finally rebuffered into a total volume of 0.5 ml in PBS (see Chapters 6.4.2, 6.4.3 and 6.4.4). The resulting genomic titers were determined by qPCR (see Chapter 6.4.5). In addition, HT-1080 cells were transduced using a MOI of 50000 as described in Chapter 6.4.7 in order to analyze the transduction ability of the different rAAV2 variants. The transduction ability was determined by flow cytometry analysis of the resulting mVenus expression. The results of these are listed in Table 28. In the following, the produced rAAV2 variants are named as follows: rAAV2_wt, rAAV2_587_GG, rAAV2_587_GGSG, rAAV2_587_$(GGSG)_2$ and rAAV2_587_$(GGSG)_4$.

Table 28: Production titer and transduction ability of different rAAV2 587 amino acid-linker variants.

Sample	**Titer in vg·ml^{-1}**	**Transduction ability in %**
rAAV2 wt	$3.1 \cdot 10^{10}$	96.7 ±0.1
rAAV2_587_GG	$7.1 \cdot 10^{9}$	31.3 ±0.1
rAAV2_587_GGSG	$4.0 \cdot 10^{10}$	20.4 ±0.6
rAAV2_587_$(GGSG)_2$	$2.4 \cdot 10^{10}$	28.0 ±0.9
rAAV2_587_$(GGSG)_4$	$4.7 \cdot 10^{9}$	10.9 ±0.2

Genomic titers shown in Table 28 demonstrated that all genetic constructs lead to rAAV production and that the integration of flexible linkers only affected the titer for large integrations, apart from construct rAAV2_587_GG, which had a lower rAAV yield than comparable designs such as rAAV2_wt or rAAV2_587_GGSG.

The resulting rAAV titer rank within a potency of ten with the lowest titer of $4.7 \cdot 10^{9}$ vg·ml^{-1} (rAAV2_587_$(GGSG)_4$) and the highest titer of $4.0 \cdot 10^{10}$ vg·ml^{-1} (rAAV2_587_GGSG), experience had shown that this corresponds to a normal fluctuation range of a standard rAAV production. That the constructs rAAV2_587_$(GGSG)_4$ and rAAV2_587_GG showed the lowest titer in this experiment was not necessarily related to the constructs, because

there was no logical correlation between rAAV yield and linker length, but it could be caused e.g. by preparative fluctuations during ultracentrifugation purification.

The transduction ability of the different constructs was not the focus of the original experimental planning, as steric changes in a region that is important for the infection would, as expected, lead to a reduction of the transduction ability. The corresponding experiment, nevertheless, revealed an interesting and unexpected behavior of the transduction ability.

The insertion of only two amino acids (GG) resulted in a decrease of transduction ability from 96.7 ±0.1 % (wild type capsid) to 31.3 ±0.1 %. Interestingly, increasing the linker length to four amino acids (rAAV2_587_GGSG) resulted in a further decrease of transduction to 20.4 ±0.6 %, but rAAV with an insertion of eight amino acids (rAAV2_587_$(GGSG)_2$) showed an improved transduction of 28.0 ±0.9 %. Additional linker increasing decreased the transduction to 10.9 ±0.2 % for the rAAV2_587_$(GGSG)_4$ variant with 16 inserted amino acids (see Table 28). Figure 19A illustrates the statistical significance of this relationship.

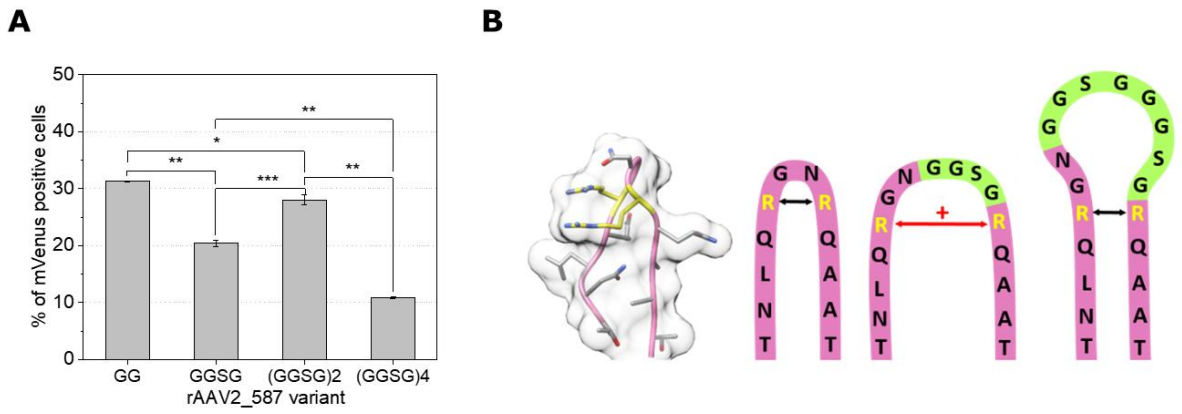


Figure 19: Comparison of different 587 loop variants. A: Transduction efficiencies of different serine-glycine linker rAAV variants. Statistical analysis was performed from triplicates using an unpaired student t-test. B: Model of the 587 loop region, from left to right: Structural model of the wild-type 587 loop with residue R585 and R588 highlighted in yellow (PDB: 1LP3); schematic model of 587 wild-type loop region (pink), 587 loop region with GGSG linker (light green) insertion leading to an increased arginine-arginine-distance; 587 loop region with $(GGSG)_2$ linker (light green) insertion leading to a normal arginine-arginine-distance but steric shielding of HSPG binding motif.

So, the inserted linker size had a strong impact on the transduction ability. Variable region VIII harbors residues R585 and R588, which mediate the primary interaction of AAV2 with the cell via HSPG [20, 21]. Insertions in this region might shield or misalign these important amino acids and thus reduce the interaction with the target cell. The observed decrease of the transduction ability did not behave like a monotonously falling curve as the insertion of eight amino acids $(GGSG)_2$ compared to four amino acids (GGSG) showed a significantly higher transduction efficiency (see Figure 19A). This might be explained by the spatial

separation of the two arginines, which is known to interfere with cell binding and internalization [22], this effect is explained graphically in Figure 19B. The insertion of small motifs such as two or four amino acids, presumably increases the tension within the β-hairpin and hence the distance between the arginines. Larger insertions (here eight amino acids) could lead to an increased flexibility of the linker sequence, restoring the natural distance of the arginines in the β-hairpin. However, the increasing length of the insertions could simultaneously lead to an enhanced steric shielding of the arginine by the growing linker loop.

7.2.2.2 **Thermal stability of rAAV2 with amino acid-linker loop insertions**

An interesting property of rAAV variants is their thermal stability, which allows conclusions on its biological activity. The thermal stability of rAAV is often described using various methods, such as DSC or DSF [198]. All these methods focus on the integrity of the capsid, but the viral DNA, which has a particularly important function for biological activity, is not considered within these methods. Building up on methods which were already established within this thesis, a stability assay has been developed which has a special focus on the release of DNA from the AAV capsid when exposed to high temperatures. For this stability assay, aliquots from purified samples of the different rAAV variants were incubated for five minutes in the thermocycler at different temperatures. The released DNA was then digested by DNaseI. The subsequent genomic titration by qPCR and the comparison of the samples previously incubated at different temperatures allowed conclusions about the temperature-dependent DNA release of the samples (see Chapter 6.4.11). This assay was performed for the rAAV2 wild type variant as well as for the fully modified 587 loop variants with different amino acid linker insertions. The normalized qPCR results and data analysis are shown in Figure 20 A to E. Therefore, the total rAAV amount per reaction was normalized to 100 % and a logistic regression was performed (Logistic5: A_{min} = 0, A_{max} = 100, no weighting) using the software Origin 2019. In order to differentiate the determined data from the "traditionally" known melting temperatures, the term "disintegration temperature" (T_d) was introduced, which describes the temperature at which 50 % of rAAVs released its DNA.

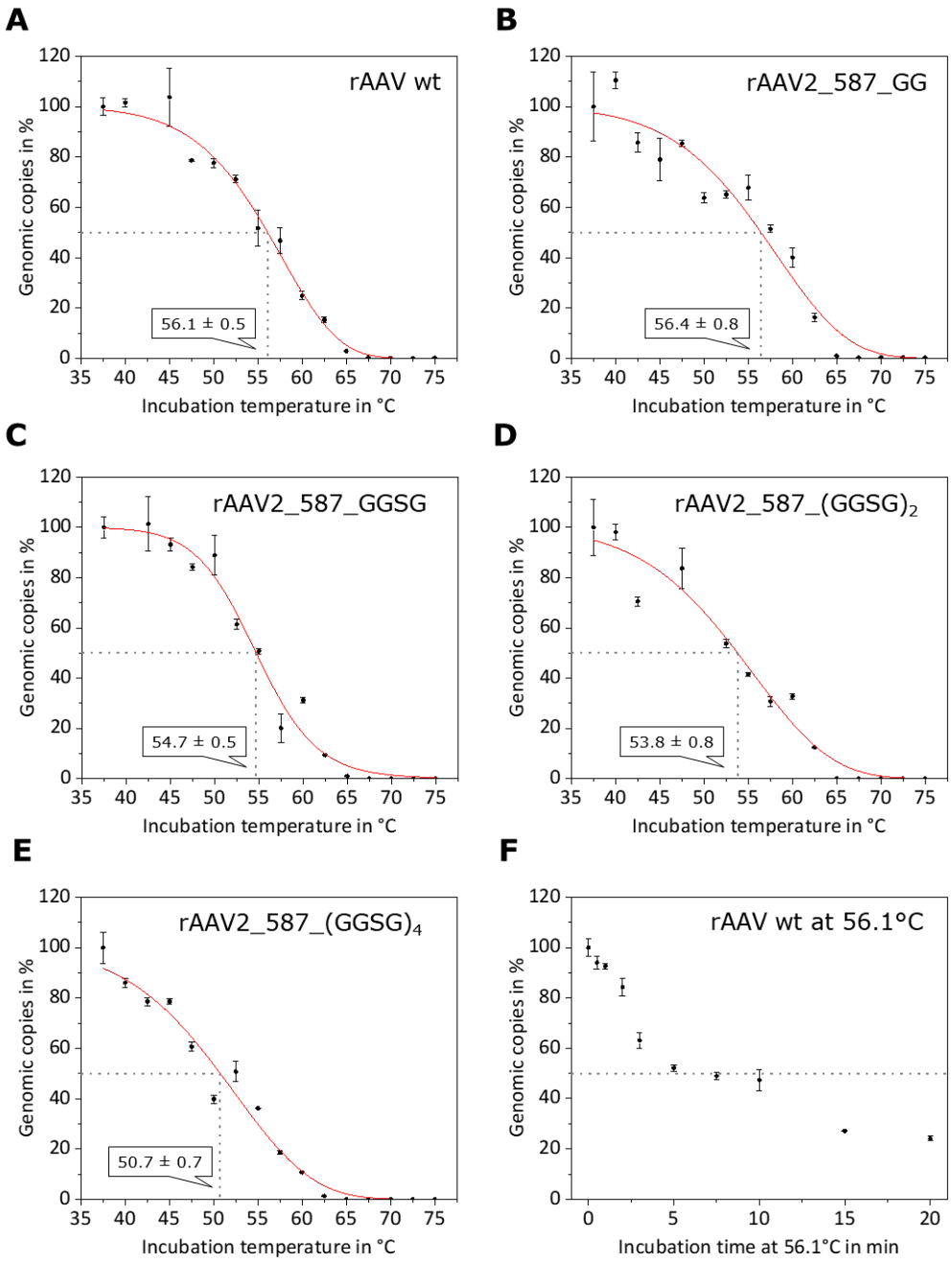


Figure 20: Results of thermal stability assay for different 587 loop variants of rAAV2 particles measured in PBS. A-E: The percentage of intact genomic copies is plotted against the incubation temperature in °C. Each point represents the standard deviation of a technical duplicate. Fitting curves (red) were calculated using a logistic function to determine the disintegration temperature (T_d) for all rAAV variants. T_d values are shown in black boxes. F: Disintegration kinetics of rAAV2 wt samples incubated at T_d = 56.1 °C for different time periods. A dashed line indicates the point where 50 % of genomic copies were released from the capsid.

The results of the determined T_d values are listed in the black boxes in Figure 20. The T_d value of the rAAV2 wild type variant was 56.1 ±0.5 °C. While the insertion of two amino acids at position 587 (GG) did not lead to a significant change of the T_d value, the insertion of the glycine-serine linker led to a gradual reduction of the T_d value with increasing linker length. Finally, the insertion of 16 amino acids $(GGSG)_4$ led to a reduction of the T_d value

by 5.4 °C to 50.7 ±0.7 °C compared to the wild type. This shows that the capsid shell is increasingly destabilized by inserts of increasing size, but stability remains assured at physiological temperatures, as the effect of the changes was relatively small.

In order to investigate the influence of the incubation time of the performed thermal stability assay, the rAAV2 wild type variant was incubated isothermally at the previously determined T_d value of 55.1 °C for different time periods ranging from 0 to 20 min. As in the previous stability assay, the samples were quantified by qPCR. The results are shown in Figure 20F. As expected for such a large complex, the rAAV2 particles were not in thermal equilibrium at elevated temperatures, which can make the thermodynamic interpretation more difficult. The determined T_d values are sensitive to the applied incubation time. A shorter incubation time leads to less destabilization than an extended incubation time. Interestingly, Figure 20F shows a multiphase behavior of the thermal destabilization, which has a plateau in the range of five to ten minutes incubation time. This behavior could be explained by different compositions or structural states of the capsid ensemble in the process of incubation. Since most experimentally determined thermal stability changes within variants were minor, it could be concluded that the capsid proteins form a stable framework that is highly resilient to sequence insertions.

As the results in this chapter have shown, the method presented for the determination of the T_d values is very well suited to describe the physical property of rAAV stability in the context of biological activity. It has been described that viral particles transfer into a metastable state when they are heated, which is defined by the ejection of the encapsulated ssDNA [208, 209]. A further increase of the temperature beyond this point ultimately leads to the complete breakdown of the capsid structure. By conventional stability determinations like DSC and DSF only this status is determined which leads to T_m values which are higher compared to the T_d values determined here. This observation is in agreement with already published melting temperatures for rAAV2, which, depending on the determination method, are about 15 °C higher than the corresponding T_d values [198, 210]. Therefore, it can be said that the disintegration temperature has a higher biological relevance with regard to the description of the viral stability than conventionally determined melting temperatures.

7.2.3 **rAAV2 variants with different β-lactamase loop insertions**

7.2.3.1 **Producibility, infectivity and thermal stability or rAAV2 with β-lactamase loop insertions**

The results of the last two chapters have shown that unstructured linker insertions up to 16 amino acids did not lead to problems with capsid assembly and stability but have a strong influence on transduction ability (see Chapter 7.2.3). In order to tie in and investigate the size limit for insertions at residue position 587, it should now be analyzed how the system behaves when inserting complete enzymes.

Thus, rAAV2 variants were produced, which carry different numbers of β-lactamase insertions at residue position 587. A fully modified variant was generated which presented the enzyme on all capsid proteins and a mosaic variant where only the VP2 capsid proteins showed modifications. Figure 21 illustrates the theoretical structure models of these rAAV2 variants in which the degree of modification is clearly visible, the capsid proteins are shown in purple and the β-lactamase enzymes in blue.

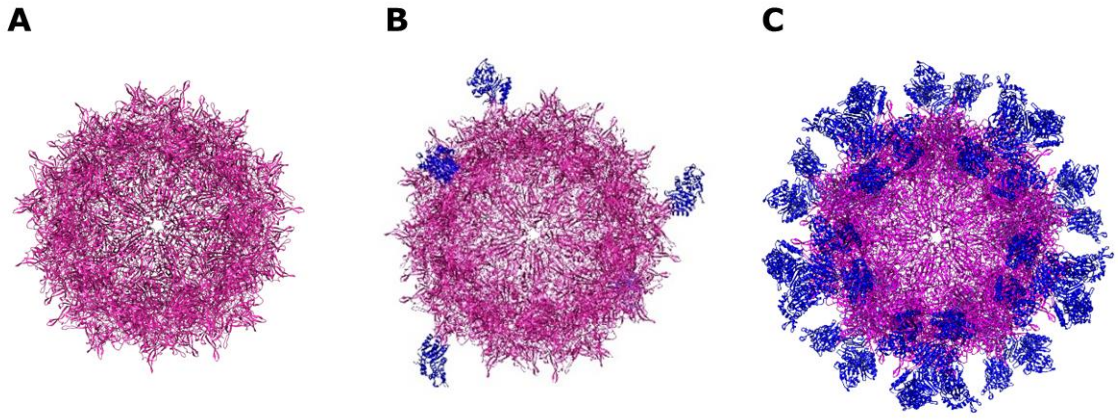


Figure 21: Theoretical structures of rAAV2 wt particle with and without several β-lactamase insertions. AAV2 wt structure (purple, PDB: 1LP3) and β-lactamase (blue, PDB: 3DTM). Generated with UCSF Chimera. A: rAAV2 wt with an estimated diameter of about 25 nm, B: rAAV2_587_bla_mosaic (mosaic variant), C: rAAV2_587_bla (fully modified) with an estimated diameter of about 35 nm.

For the production of these variants, adherent HEK293 cells were transfected with the corresponding plasmids using ten 100 mm dishes for each variant, as described in Chapter 6.4.1. For both batches the pHelper plasmid pZMB0088 and the GOI-ITR plasmid pZMB0522 were used. The plasmid pZMB0221 was used as the RepCap plasmid for the fully modified variant, whereas the plasmids pZMB0577 and pZMB0600 were used for the mosaic variant. Following production, the rAAVs were purified by ultracentrifugation and finally rebuffered into a total volume of 0.5 ml in PBS (see Chapters 6.4.2, 6.4.3 and 6.4.4). The resulting genomic titers were determined by qPCR (see Chapter 6.4.5). In addition,

HT-1080 cells were transduced using a MOI of 50000 as described in Chapter 6.4.7 in order to analyze the transduction ability of the different rAAV2 variants. The transduction ability was determined by flowcytometry analysis of the resulting mVenus expression. The results of these tests are listed in Table 29. In the following, the produced rAAV2 variants are named as follows: rAAV2_587_bla (fully modified) and rAAV2_587_bla_mosaic (mosaic variant).

Table 29: Production titer and transduction ability of the different rAAV2_587_β-lactamase variants.

Sample	**Titer in vg·ml^{-1}**	**Transduction ability in %**
rAAV2_587_bla	$1.3 \cdot 10^{10}$	1.2 ±0.1
rAAV2_587_bla_mosaic	$6.3 \cdot 10^{10}$	57.0 ±2.0

As shown in Table 29 it was possible to produce both the fully modified β-lactamase rAAV2 variant and the less modified mosaic β-lactamase variant. The genomic titers finally generated ranged from $1.3 \cdot 10^{10}$ vg·ml^{-1} for the fully modified variant to $6.3 \cdot 10^{10}$ vg·ml^{-1} for the mosaic variant. This lies within the range of the production of the rAAV2 wild type ($3.1 \cdot 10^{10}$ vg·ml^{-1}, see Table 28), although the four plasmid system had to be used for the mosaic variant. A greatly reduced transduction efficacy was observed with only 1.2 ±0.1 % of the cells in case of the fully modified variant. The maximum error which can be assumed for this determination method is 1.0 % of false positive cells. Since the transduction value is only slightly above this error value, it can be assumed that no significant transduction of HT-1080 cells with the completely modified variant took place. However, this was expected, since in this variant all residues required for the transfection were shielded by the presented enzymes. Thus, the rAAV2_587_bla variant has no biological activity. In contrast, a transduction of 57.0 ±2.0 % could be detected in the corresponding mosaic variant. Compared to the wild type (96.7 ±0.1 %) this is a significant reduction, however, this value is also significantly higher than for all investigated completely modified variants with amino acid-linker insertions (see Table 28). A reason for this value might be that the interaction of the cells with the rAAV2_587_bla_mosaic variant was affected by steric hindrances, but the unmodified VP1 and VP3 capsid proteins leaved enough interaction possibilities.

That it was even possible to produce the rAAV2_587_bla variant, despite the high degree of modification, was unexpected. Since this variant was not able to transduce HT-1080 cells, there is no biological relevance of this rAAV2 construct. Nevertheless, further experiments were performed to determine some interesting characteristics of this rAAV2 variant.

First, the different sizes of capsid proteins compared to the wild type were confirmed by western blot analysis. For this purpose, samples of the rAAV2 wild type variant and rAAV2_587_bla (each approx. $2.5 \cdot 10^8$ vg) were applied to an SDS-PAGE and separated. Subsequently, as described in Chapter 6.3.10, a western blot was performed using the B1 anti-AAV VP1/VP2/VP3 antibody. This primary antibody specifically binds a sequence at the N-terminus of VP proteins, which is identical for all VP proteins. The result of the Western blot can be seen in Figure 22A. In both lanes, a band is visible at the expected running height of the non-modified or modified VP3 protein. In the case of the wild type sample, the band runs at about 60 kDa, which corresponds to the expected molecular weight of the VP3 capsid protein. In the lane of the rAAV2_587_bla variant, a band at about 89 kDa can be detected. This corresponds to the expected increase of the VP3 band by 29 kDa due to the insertion of β-lactamase. Therefore, at least the complete modification of VP3 capsid proteins at the protein level could be demonstrated. The amount of rAAV2 particles used was probably not sufficient to detect the modified or unmodified VP1 and VP2 proteins. Due to the stochastic 1:1:10 distribution of the VP1, VP2 and VP3 capsid proteins, the VP1 and VP2 proteins occur around the factor of ten less frequently in the assembled rAAV, whereby the corresponding signal was probably below the detection limit of the western blot.

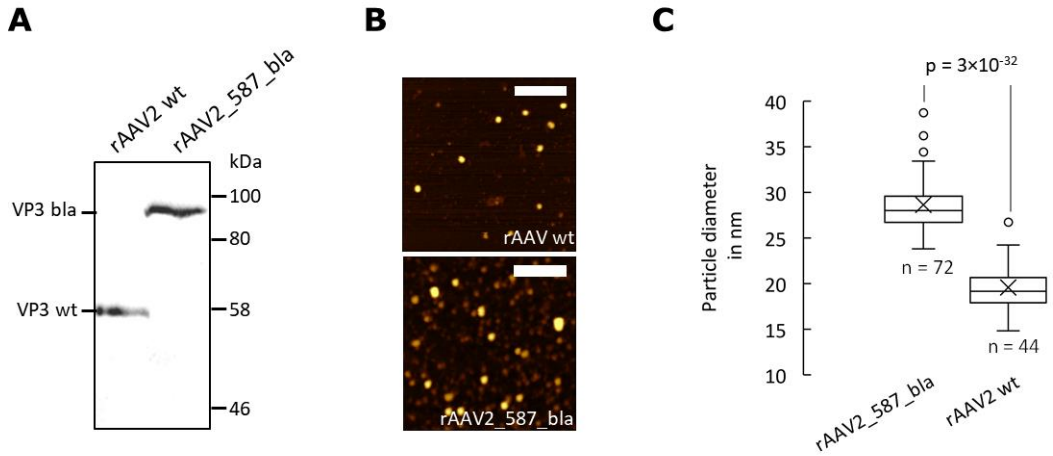


Figure 22: Comparison of rAAV2 wt and rAAV2_587_bla with respect to structure. A: Western blot of purified rAAV2 wt and rAAV2_587_bla with B1 anti-AAV VP1/VP2/VP3 antibody showing the mass shift of VP3 capsid protein ($VP3_{wt}$ = 60.1 kDa and $VP3_{587\_bla}$ = 88.9 kDa). B and C: AFM micrographs and particle size analysis of rAAV2 wt and rAAV2_587_bla. The calculated mean diameter of rAAV2 wt is 20 nm compared to 29 nm for rAAV_587_bla.

Next, the expected size change of the modified viral capsids should be determined. For this purpose, Marco T. Radukic made atomic force microscopy images of purified rAAV2_wt and rAAV2_587_bla particles according to the method described in Chapter 6.4.10. The recorded images were analyzed and average diameters of the two rAAV2 variants were determined. The images and results of the data analysis are shown in Figure 22 B and C.

A diameter of approx. 20 nm for rAAV2 wt was found by the AFM measurement. The deviation of the particle diameter of approx. 5 nm compared to rAAV2 wt particles, which were analyzed by TEM measurements can be attributed to the type of method used and different sample preparations. In contrast, a significantly higher viral particle diameter of approx. 29 nm was found for rAAV2_587_bla. From the capsid model shown in Figure 21C a diameter of 35 nm was estimated for a fully modified rAAV2 capsid and estimations for the wild-type capsid resulted in an average of 25 nm, which is in agreement with the observation (see Figure 21A). The box whisker plot and the statistical analysis from Figure 22C prove the high significance of the differences in diameter between the two rAAV2 variants.

Next, as with the completely modified amino acid-linker variants (see Chapter 7.2.2.2), the disintegration temperature should be investigated by the thermal stability assay. The assay was performed in the same way with a purified rAAV2_587_bla sample in PBS. The normalized qPCR data is shown in Figure 23.

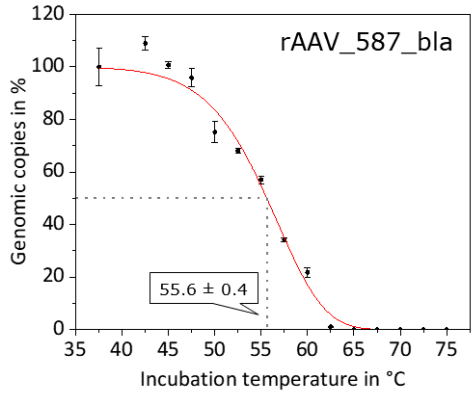


Figure 23: Results of thermal stability assay based on qPCR data for rAAV_587_bla. Each point represents the standard deviation of a technical duplicate. T_d is estimated to be 55.6 ±0.4 °C.

Based on the data from Figure 23, a T_d value of 55.6 ±0.4 °C was found for the rAAV2_587_bla variant. This value is only 0.5 °C below the T_d value determined for the wild type variant. This small reduction of the T_d value was unexpected considering the T_d values determined in Chapter 7.2.2.2 for the amino acid-linker variants, as insertions of only 16 amino acids resulted in a T_d reduction of 5.4 °C. The insertions of β-lactamase, which is approx. 29 kDa (263 amino acids) in size, therefore seems to disturb the integrity of the viral capsid less strong.

Therefore, it can be assumed that the folding of the capsid proteins and the capsid assembly were not significantly affected by larger insertions if they have a certain structure. The β-lactamase variant used in this thesis, as well as the previous variants [211], are known for their good folding behavior and high thermodynamic stability [207].

In contrast to the unstructured amino acid-linker, these characteristics could contribute positively to the robustness of the modified capsid proteins and the corresponding capsid assembly. Possibly the robust β-lactamase enzymes could even protect the capsid from unfolding.

7.2.3.2 **Enzyme activity of rAAV2 variants with β-lactamase loop insertions**

Since interesting thermodynamic properties have already been determined for the rAAV2_587_bla variant, the question remains whether the β-lactamase molecules presented on the capsid surface possess enzymatic activity. To answer this question, a simple bacterial β-lactamase activity assay was developed. The exact procedure is described in Chapter 6.4.12.2. In general, this method is based on platting a non-ampicillin-resistant *E. coli* culture on an ampicillin-containing LB-agar plate and then dropping samples of purified rAAV2_587_bla onto the LB-agar plate. After the samples have been drawn into the plate, it was incubated at 37 °C overnight. If the presented β-lactamase enzymes are active, the ampicillin would be catabolized at the corresponding spot and bacterial growth would occur. For better visualization, an *E. coli* culture was used, which carried the plasmid pZMB0084 (pSB1C3_001_LacI_promoter-RFP-Terminator). This plasmid mediates besides a chloramphenicol resistance (the LB-agar plate contains both ampicillin and chloramphenicol) the expression of RFP under the control of the constitutive LacI promoter.

Photographed sections of the LB agar plate of this bacterial assay are shown in Figure 24A. It can be seen that *E. coli* colonies grew at the spots where samples of rAAV2_587_bla, up to a dilution of 1:3, were dripped. At the spots where a sample of the rAAV2 wild type with an identical genomic titer was dripped and at the spot of the puffer control, no bacterial growth could be observed. Thus, this simple test proved that the β-lactamase enzymes presented on the capsid surface possess enzymatic activity.

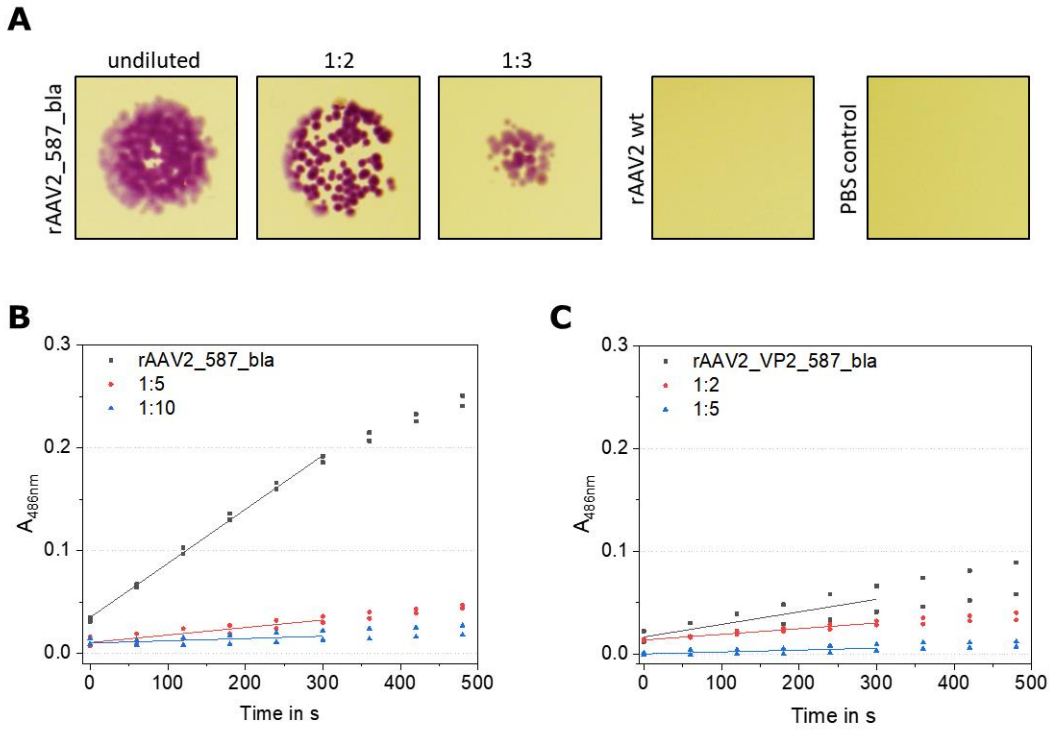


Figure 24: β-lactamase activity assays. A: Images of a bacterial growth assay on LB Agar plates for β-lactamase activity. Samples with rAAV2_587_bla show a β-lactamase activity up to 1:3 dilutions, whereas for rAAV2 wt and PBS no colony growth was observed. B: Results of nitrocefin assay, probing the β-lactamase activity of rAAV2_587_bla in different dilutions. C: Nitrocefin assay probing rAAV2_VP2_587_bla mosaic viral particles in different concentrations for lactamase activity. Slopes of linear regressions of the first 300 s were used to calculate the concentration of active β-lactamases.

Due to the enzymatic activity of the β-lactamases presented, it was possible to draw conclusions about the β-lactamase concentration being used, since the catalytic constant of this variant is known and the enzymatic turnover could be observed calorimetrically by a nitrocefin assay (see Chapter 6.4.12.1). Such a nitrocefin assay was performed for both the completely modified rAAV2_587_bla variant and the mosaic variant (here called rAAV2_VP2_587_bla), more detailed information can be found in Chapter 6.4.12.1. Three different dilutions were measured in duplicates. The change in absorption at 486 nm, measured in 60 seconds intervals, is shown in Figure 24 B and C for both variants. For the calculation of the slopes of the linear regressions resulting from the absorption data, the measuring points of the first 300 s were used. On the basis of Lambert-Beer's law (formula see Chapter 6.4.12.1) the following mean β-lactamase concentrations were determined: rAAV2_587_bla: $1.65 \cdot 10^{-10}$ mol·l^{-1} and rAAV2_VP2_587_bla: $1.73 \cdot 10^{-11}$ mol·l^{-1}. When these values were offset against the number of rAAV2 particles used, the average number of β-lactamases presented per capsid was: rAAV2_587_bla: 26.9 fully activ β-lactamase per capsid and rAAV2_VP2_587_bla: 5.6 fully activ β-lactamase per capsid. These numbers resulted from the assumption that all β-lactamases presented were fully active. In the case of a completely modified variant, 60 β-lactamases would be expected and in the case of

the VP2 mosaic variant five. For the mosaic variant, the measured data met expectations quite well. The assumption that only half of the expected β-lactamases were present on the surface of the completely modified variant is not realistic due to the capsid assembly, rather it could be assumed that the β-lactamases presented did not reach their full activity. This can be explained, for example, by the strength of the hindrances, which are more likely to be expected with the more densely packed capsid surface of the rAAV2_587_bla variant than with the less densely packed mosaic variant (see Figure 21 B and C). Consequently, it could be assumed that the presented enzymes, in the case of the rAAV2_587_bla variant, achieve only about 44.8 % of their possible activity. However, whether the values determined in this assay are absolutely trustworthy depends on other factors that are rather difficult to determine. One of these factors was the rAAV2 particle count used per sample. The data here referred only to the genomic titers previously determined by qPCR. However, since the rAAV2 samples were only purified by ultracentrifugation, it cannot be absolutely excluded that the samples also contain viral particles that did not contain DNA. As a result, the actual amount of used rAAV2 particles could be underestimated, which would ultimately increase the number of presented β-lactamases. A more precise determination of the capsid titer, e.g. per capsid ELISA, was not possible due to the strong modifications. Furthermore, the assumed maximum activity of β-lactamases was also doubtful as it based on measurements performed with free enzyme in solution. The integration of the enzymes into the viral capsid results in a kind of immobilization of which the influence on activity cannot be predicted. In industry, immobilized enzymes are often used to simplify procedures, which can lead to a change (positive or negative) in absolute enzyme activity [212]. As the viral capsids, due to their small size, do not cause immobilization in the classical sense, this influence is difficult to quantify.

Finally, it can be said that with the rAAV plasmid production system established in the working group, it is possible to produce such strongly modified rAAV2 particles and that they even show enzymatic activity. rAAV capsids with larger modifications have not been published elsewhere at this time.

7.3 **Improvement of rAAV2 production by generation of producer cell lines**

rAAVs are produced using a helper-free plasmid system. This includes the transfection of the respective production cell line with three to four different plasmids. Depending on the production cell line, the number of required plasmids and the transfection method, the rAAV yield can fluctuate strongly. A reduction of the required plasmids could not only reduce the costs of rAAV production but could also increase the rAAV yield. In the context of this thesis, experiments were performed on different genomic manipulations on the Chinese hamster cell line CHO K1 and the human cell line HEK293 with the aim to improve the rAAV2 production. The techniques and results of these experiments are presented in the following subchapters.

7.3.1 **Generation of CHO K1 producer cell lines for rAAV2 using CRISPR/Cas9**

For the production of rAAV using the helper-free system three plasmids are required. In addition, the production cell line provides additional adenoviral elements. The genes E1A, E1B are stably integrated into HEK293 cells which need to be delivered if using another producer cell line. As approximately 70 % of the biopharmaceuticals produced worldwide are currently produced with the aid of cells derived from the Chinese hamster cell line CHO it is regarded as a well characterized industry standard [213]. For this reason, an attempt was made to enable the production of rAAV2 in CHO K1 cells. The missing helper functions should therefore be integrated into the CHO K1 genome by a rational approach using the CRISPR/Cas9 technology. In order to anticipate the result of this subproject, it must be said that this goal was not achieved, and that this subproject was cancelled due to the immense effort of time and the low probability of success. Nevertheless, in the following chapters, the rational approach and execution of the experiments will be shown and discussed. This should provide a basis for further projects in this field and point to potential mistakes during this process.

7.3.1.1 **Finding suitable target sites for genome editing in the CHO K1 genome**

For the insertion of the desired gene sequences into the CHO K1 genome, suitable CRISPR target sites had to be found first. The requirements that these CRISPR target sites and the surrounding genomic region had to fulfill were defined using a profile list with hard criteria that had to be fulfilled in any case and soft criteria that should be fulfilled if possible. Before starting this project, Xu *et al.* published an assembly of the CHO K1 genome with a total size of 2.45 Gb including 24383 predicted genes, which formed the bases for the target site search [214].

One of the hard criteria that the CRISPR target sequences themselves had to fulfill was that they were unique in the entire CHO K1 genome. In addition, target sites with a low singe mistake off target rate were preferred, which means that there should be only few similar target sites in the genome that differ by only one base. This should reduce the possibility of misintegration. However, since these criteria apply to many thousands of target sites, the initial focus was to find suitable genomic regions in which CRISPR target sites should later be found. The criteria that these genomic regions had to fulfill were, that they had to be intergenic regions, since it was not intended that the insertion would damage any predicted gene sequences. The prediction of promoter regions and their associated regulatory elements is particularly difficult in eukaryotic organisms. In order to reduce the probability of damaging such elements, only intergenic regions were selected, on which the reading directions of the two surrounding predicted CDS move towards each other, because promoter regions are mostly located upstream of the coding sequence. In addition, the distance between the two surrounding genes should not be too large or too small. For this, intergenic regions with a size of about 5 kb to 50 kb were preferred. Furthermore, the following soft criteria should be fulfilled. The intergenic regions should have no sequencing gaps and no or only very few repeating elements, if possible, so that the cloning of the required homology arms for the integration constructs would be easier. Additionally, regions with a high number of CpG islands were avoided, as such regions are increasingly affected by silencing processes [215].

One of the most important criteria was that the intergenic region was located in a very active genomic environment. This means that the surrounding genes must have a high transcription activity. In order to be able to utilize this criterion, this project was supported by Dr. Alexander Sczyrba (head of “Bielefeld University Bioinformatics Services” and head of the working group “Computational Metagenomics” at Bielefeld University). With his help, a transcriptome data set of an RNA sequencing of CHO K1 (Accession: PRJNA214684) was mapped against the aforementioned CHO K1 genome assembly. Moreover, the respective contigs of this assembly were divided into sections (loci) of 50 kb each. For each of these loci a coverage was determined. The coverage describes the frequency with which sections of the loci were covered by RNA sequencing data. It was assumed that highly active or easily accessible areas of the genome are more frequently transcribed, and that the corresponding RNA occurs more often in the transcriptome. A modest disadvantage of this strategy is, that further transcriptional active areas, of which the produced RNA was degraded quickly by different regulatory processes, are not considered this way. A table is shown in Chapter 10.4.1 of the appendix (Table 33) in which the 75 loci with the highest coverage are listed in descending order. Using the UCSC Genome Browser and the data set of the RNA mapping, it was manually searched for intergenic regions within these loci and neighboring regions, which had the highest coverage and to which all previously mentioned criteria matched [216]. A list of the twelve “best” areas is shown in Chapter

10.4.2 of the appendix in Table 34. Finally, four CRISPR target sites were found that satisfy all criteria and therefore appeared to be suitable for the planned integrations (see Table 30).

Table 30: List of selected CRISPR target sites

Name	**Target sequence (G-(N)$_7$-(N)$_{12}$-NGG)**	**Contig**
Target A	G ACAGAGA ATTCAAGAACAG GGG	KE379448
Target B	G GGCGGCT CTTTGACCACGG TGG	KE379359
Target C	G TTCTTAA AGGTGAGGCTAA CGG	KE379248
Target D	G TCTGTGT ACCACGAGTGCT TGG	KE379184

CRISPR target site A is located on contig KE379448 between the genes Ctr9 (Paf1/RNA polymerase II complex component) and Eif4g2 (eukaryotic translation initiation factor 4 gamma, 2), target site B is located on contig KE379359 between the genes Ppp1r7 (protein phosphatase 1, regulatory subunit 7) and Hdlbp (high density lipoprotein binding protein), target site C lies on contig KE379248 between the genes Rnd2 (Rho family GTPase 2) and Brca1 (breast cancer 1, early onset) and target site D lies on contig KE379184 between the genes Rara (retinoic acid receptor, alpha) and Top2a (topoisomerase (DNA) II alpha 170 kDa).

7.3.1.2 **Generation of DNA constructs for CHO K1 genome editing**

After selecting four potential CRISPR target sites, several plasmid constructs were created for the genome editing process. In addition to the required CRISPR plasmids, further constructs were generated to characterize the target sites. These included constructs that cause the expression of a destabilized GFP variant. The idea was that after a successful genomic integration of these expression cassettes and the generation of monoclonal CHO K1 cell lines, the respective target sites could be characterized by their expression strength. As this process was expected to take a lot of time, further plasmid constructs were generated that could enable the rAAV2 production in CHO K1 cells. These included constructs containing the 4344 bp long adenoviral DNA fragment, which is genomically integrated into the HEK293 genome, in the context of its cell line generation. The CDS contained on his DNA fragment codes for the genes E1A and E1B, which are required for AAV production. In order not to go beyond the scope of this chapter, only some aspects will be discussed here.

For CRISPR/Cas9 genome editing, plasmids from the “ZhangLab” were purchased via the plasmid database Addgene, including a variant containing the original *S. pyogenes* Cas9 variant (pSpCas9(BB)-2A-Puro (PX459), pZMB0065) and a construct containing the Cas9 D10A nickase variant (pSpCas9n(BB)-2A-Puro (PX462), pZMB0068). Both plasmids lead

to expression of the respective Cas9 variant and a Puromycin resistance protein linked via a 2A site. The insertion of the target site sequences into the sequence used for the transcription of the guide RNA was carried out via the two times existing recognition site of the type IIs restriction enzyme BbsI. For this purpose, the DNA oligonucleotides Target-A-for/-rev, Target-B-for/-rev, Target-C-for/-rev and Target-D-for/-rev (see Table 10) were hybridized in pairs and then 5’ phosphorylation was performed. The plasmids pZMB0065 and pZMB0068 were digested using BbsI, dephosphorylated and subsequently ligated with the hybridized DNA. This resulted in the plasmids “Target A-D in pX459” (pZMB0203 to pZMB0206) and “Target A-D in pX462” (pZMB0207 to pZMB0210).

The design of the plasmid constructs needed for the various genomic integrations was more complex. Figure 25 shows schematic illustrations of this plasmid constructs and some intermediate constructs.

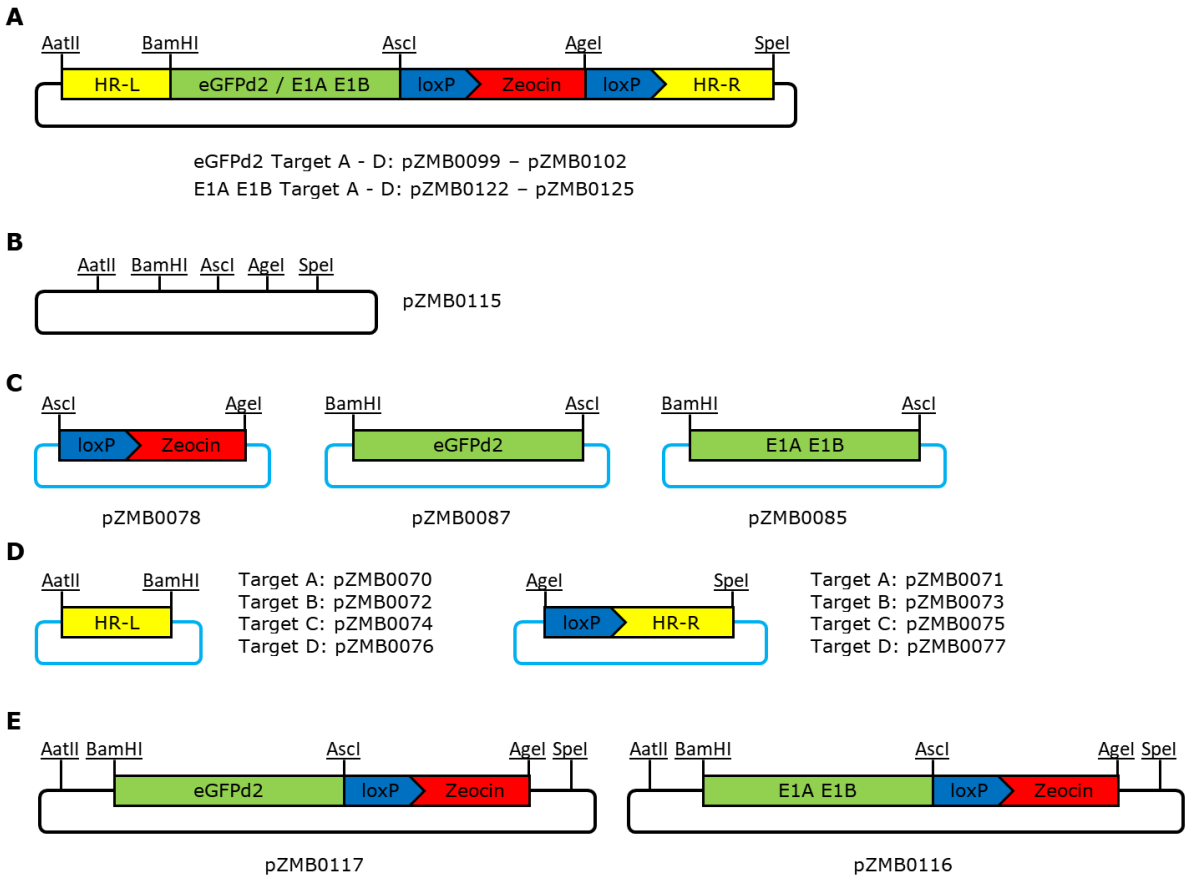


Figure 25: Schematic illustrations of plasmid constructs and some intermediate constructs for genomic insertions. A: Final plasmid constructs for the genomic integrations, HR-L: left homologous region, HR-R: right homologous region, eGFPd2/E1A E1B: destabilized eGFP expression cassette or adenoviral DNA fragment conteining E1a and E1B CDS, Zeocin: Zeocin resistenz cassette surrounded by loxP sites. B: Starting plasmid backbone based on psb1c3_001 with new introduced multiple cloning site. C: Expression cassettes for Zeocin resistenz, eGFPd2 and adenoviral E1A/E1B all with pJET1.2 backbone. D: Intermediate constructs containing left and right homologous regions for all target sites subcloned in pJET1.2. E: Intermediate constructs before adding homologous regions.

Figure 25A shows the schematic structure of the insertion constructs. Due to the modular design of the constructs, with each individual element being separated by a specific restriction site it was possible to combine the individual parts. The sequence to be integrated starts and ends with the 800 bp long homologous regions, called HR-L and HR-R. It is defined by the cutting position of the CRISPR target sites and the corresponding regions located upstream (HR-L) or downstream (HR-R) of this position. The second module consists of the expression cassette for the destabilized variant of eGFP (eGFPd2) or for the adenoviral DNA fragment containing the CDS for the genes E1A and E1B. This module forms the actual insertion fragment. The third part is a Zeocin resistance cassette, which is flanked by loxP sites. This resistance cassette was used for the selection of the clones and provides the possibility to remove the Zeocin resistance cassette through the loxP sites, if this would be necessary for later experiments.

Assembly of these individual parts in a plasmid was accomplished by generation of a starting plasmid with an individual multiple cloning site. AatII, BamHI, AscI, AgeI and SpeI were selected and integrated into plasmid pZMB0084 by hybridized oligonucleotides (MCS-for/-rev) which resulted in pSB1C3_001_MCS (pZMB0115).

The Zeocin resistance cassette, which is controlled by the SV40 promotor and terminated by the SV40 polyA, was amplified using the primer Zeo-for/-rev from the plasmid pSecTag2 (a gift from AG Fermentationstechnik). The used primer added the recognition sequence for AscI and the loxP sequence (ATAACTTCGTATAGCATACATTATACGAAGTTAT) at the 5'-end of the PCR product and the recognition sequence for AgeI at the 3'-end. The resulting PCR product was subcloned in pJET1.2 (see Chapter 6.1.19) and resulted in plasmid pZMB0078.

The destabilized GFP (eGFPd2) sequence was ordered as gene synthesis, based on the AddGene Plasmid "pCAG-GFPd2" (#14760), optimized for *Cricetulus griseus* codon usage and cloned into pcDNA5/FRT (pZMB0001) by restriction digestion with NheI and ApaI resulting in pZMB0086. Primer CMV-eGFPd2-pA-BamHI-for and CMV-eGFPd2-pA-AscI-rev introduced the restriction recognition sequences for BamHI and AscI and were used to amplify the CMV-eGFPd2-hGHpA sequence from the resulting plasmid which was then subcloned into pJET1.2 leading to pZMB0087.

The module consisting of the 4344 bp long adenoviral DNA fragment was isolated from genomic DNA of HEK293 cells (see Chapter 6.1.5). Since PCR amplification of the whole fragment in one piece, using the isolated genomic DNA as template, did not work, a total of three fragments were generated by PCR, which were ligated one after the other via existing restriction sites. Details of this cloning process will not be discussed here. The complete fragment, which was flanked by the restriction sites BamHI and AscI like the eGFPd2 expression cassette, was finally subcloned in pJET1.2 resulting in pZMB0085.

The homologous regions were amplified from genomic DNA isolated from CHO K1 cells. When planning the primers for the amplifications of these genomic regions, it was important that the position of the respective primers, which were partially located on the CRISPR target site, were defined by the cutting/nicking position, which is located between the third and fourth base upstream of the PAM sequence. The corresponding partner primers were placed approximately 800 bp upstream or downstream, according to the rules presented in Chapter 6.1.7. The restriction sites for the enzymes AatII and BamHI were added to the overhang of the primers, which amplify the genomic region upstream of the CRISPR target site ((A-D)-HR-L-for/-rev). The restriction sites for the enzymes AgeI and SpeI were added to the corresponding primer pairs that amplify the genomic region downstream of the CRISPR target sites ((A-D)-HR-R-for/-rev). In addition, the overhangs of the forward primers also contain the sequence of the loxP site. The PCR products of the eight reactions were isolated by agarose gel electrophoresis followed by DNA extraction and were then subcloned in pJET1.2. Sequencing of these genomic sections showed that the cloned sequence parts match the expected sequences. The plasmids were stored under the numbers pZMB0070 to pZMB0077.

After all submodules for the integration constructions were generated, the submodules were inserted one after the other into the plasmid pZMB0115, which contains the constructed (multiple cloning site) MCS. Intermediate constructs such as pZMB0116 or pZMB0117 are shown in Figure 25E. Finally, the insertion constructs containing the eGFPd2 expression cassettes were stored under the numbers pZMB0099 to pZMB0102 (Target A to D) and the insertion constructs containing the adenoviral DNA fragment with the CDS of E1A and E1B, were stored under the numbers pZMB0122 to pZMB0125 (Target A to D).

7.3.1.3 **Generation and analysis of monoclonal CHO K1 cell lines by CRISPR/Cas9 genome editing**

To determine the needed Zeocin concentration for the genome editing experiments of CHO K1 cells different Zeocin concentration were tested in eight cultivation batches of CHO K1 suspension cells with a starting cell titer of $1.5 \cdot 10^6$ ml^{-1}. Each of these batches had a volume of 5 ml and was cultivated in a 50 ml shaking tube. Directly at the beginning of the cultivation, Zeocin from a 100 $mg \cdot ml^{-1}$ stock solution was added to the normal TC42 growth media, resulting in the following total Zeocin concentrations: 0, 50, 100, 200, 300, 400, 500 and 600 $\mu g \cdot ml^{-1}$. The viable cell titers and viabilities were observed over a period of seven days (see Chapter 6.2.4). The resulting data are shown in Figure 26.

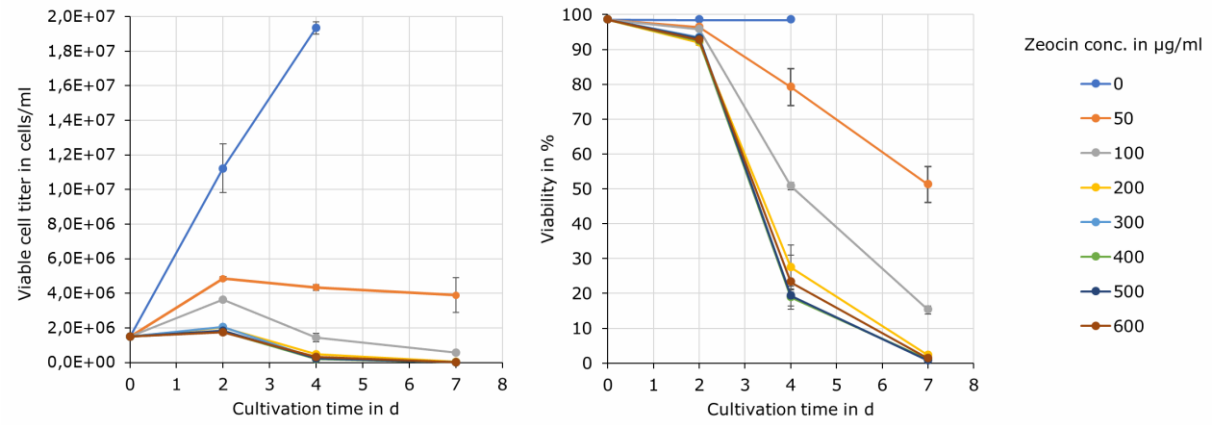


Figure 26: Viable cell titers and viabilities of CHO K1 suspension cell cultures supplemented with different Zeocin concentrations.

Figure 26 shows that the cell growth was strongly influenced by the addition of Zeocin, even at lower concentrations. A significant reduction in cell viability could first be seen after four days for all approaches (except negative control). After four days, the viability of the cells treated with 200 µg·ml^{-1} Zeocin decreased to 27.5 %. However, higher Zeocin concentrations barely showed a further decrease in viability at this time point and varied within the range of the 20 % mark. After seven days under Zeocin impact the viability of all batches treated with Zeocin concentrations above 200 µg·ml^{-1} decreased below 2 %. This experiment shows that a Zeocin concentrations above 200 µg·ml^{-1} in the culture medium, after about seven days of cultivation, leads to a sufficient killing of non-resistant cells. Finally, a working concentration of 300 µg·ml^{-1} was selected for further experiments in order to slightly increase the selection pressure.

Since at this time the cloning work for the constructs containing the adenoviral genes E1A and E1B had not yet been completed and the different CRISPR target sites should be characterized beforehand, eight different approaches for the genomic editing of CHO K1 were prepared using the eGFPd2 constructs. For each target site, two batches were made, which varied by the used CRISPR/Cas9 plasmid on the basis of pSpCas9(BB)-2A-Puro or pSpCas9n(BB)-2A-Puro (pZMB0203 to pZMB0206 or pZMB0207 to pZMB0210). The plasmid used combinations are shown in Table 31, for the transfections the needed total amount of plasmid DNA was kept the same, whereby equimolar amounts of each plasmid were used (see Chapter 6.3.1). Each of the transfection batches had a final volume of 5 ml and were cultivated in a 50 ml shaking tube. 24 h after transfection the medium was replaced by TC42 medium supplemented with 300 µg·ml^{-1} Zeocin and 5 µg·ml^{-1} Puromycin. Another 48 h later the medium was exchanged again, but at this time the Puromycin selection pressure was removed. The viability of the approaches varied between 33% and 42% after 72 hours. In the following period of time, every two days the medium was replaced by fresh medium supplemented with Zeocin. The cell density of the respective cultures were adjusted to 1.5·10^{6} ml^{-1} by adjusting the culture volume, but the volume

was not reduced below 1 ml or increased above 15 ml. Seven days after transfection, the viability of the cultures varied between 41% and 48%. This indicated that the transfections were mostly successful. After a total of 21 days, a viability of more than 95 % could be measured in all batches.

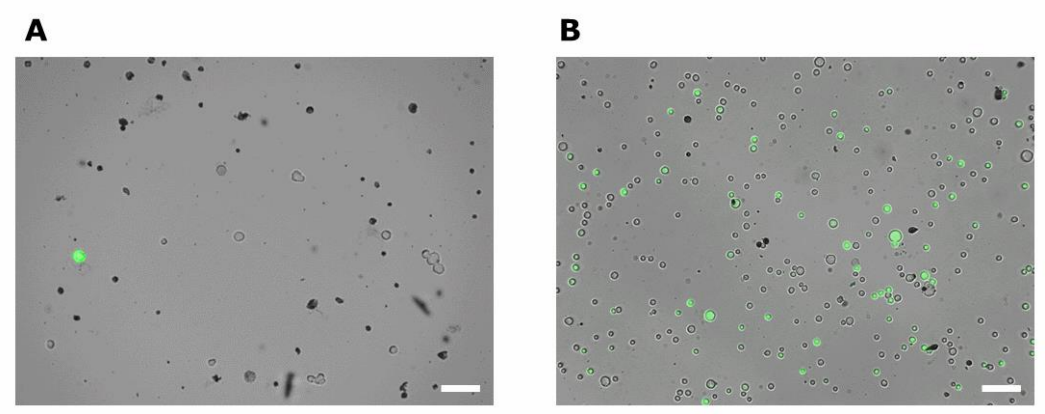


Figure 27: Microscopic images of CHO K1 cells transfected with plasmids pZMB0099 and pZMB0203. Overlay of transmitted light image with GFP fluorescence image. White bar indicates 100 µm. A: seven days after transfection, B: 19 days after transfection.

Figure 27 shows an example of two microscopic images (transmitted light merged with GFP fluorescence) taken from the culture transfected with plasmids pZMB0099 (pSB1C3_001_MCS_AL_eGFPd2_Zeocin_AR) and pZMB0203 (Target A in pX459). The images were taken from cell samples previously treated with trypsin solution during the cell count. The first image shows the cell sample seven days and the second image 19 days after the transfection. In both cases, the cell staining reflects the described change in viability. In addition, besides some (or single) fluorescent cells, a large proportion of living cells did not show GFP expression. In the case of the image taken 19 days after the transfection, it can be assumed that the plasmid DNA is not present extrachromosomally anymore, thus non-fluorescent cells could only be explained by the integration of a damaged eGFPd2 expression cassette or due to silencing mechanisms but seems unlikely since the resistance genes were intact. 21 days after the transfection, limited dilutions of all batches were made as described in Chapter 6.2.8 to generate monoclonal cell colonies.

Table 31: Generation of monoclonal CHO K1 cell lines after genome editing

Target site	Plasmids	Single cells (Day 1)	Cell colonies (Day 7)	Cell colonies on 12-well plates	Cell colonies in T-flasks
A	pZMB0099 + pZMB0203	18	15	5	2
	pZMB0099 + pZMB0207	15	1	0	0
B	pZMB0100 + pZMB0204	21	20	7	3
	pZMB0100 + pZMB0208	24	3	0	0
C	pZMB0101 + pZMB0205	22	13	5	2
	pZMB0101 + pZMB0209	22	2	0	0
D	pZMB0102 + pZMB0206	11	8	0	0
	pZMB0102 + pZMB0210	17	2	0	0

The data from Table 31 shows that the limited dilution of the transfection batches was successful. Three 96-well plates were used for the separation per transfection batch (corresponds to 180 wells per batch). Between 11 and 24 wells with fluorescent single cells could be identified immediately after separation. Seven days after separation, the number of remaining colonies that were still fluorescent and showed cell proliferation decreased significantly. Most of the observed cell colonies showed no further proliferation after two or three days or there was no fluorescence of the respective cells detectable, which was apparently a sign of cell death. Overall, it is remarkable that this cell death mainly affected batches based on pX462 CRISPR/Cas9 plasmids (pZMB0207 to pZMB0210). By the time the cell colonies were transferred to 12-well plates, no cell growth could be observed in all remaining pX462-based batches. The reason why the pX462 attempts were affected in particular could not be clarified conclusively. It is possible that this was due to the altered DNA cutting mechanism (only nicking instead of cutting), but genomic random integrations would still be possible leading to cell colonies. Also, no cell colony of target site D could be expanded. Finally, seven monoclonal cell lines (two of target A, three of target B and two of target C) could be generated. From these cell lines, master stocks were cryopreserved, and genomic analyses were performed.

For genomic analysis of each monoclonal cell line, genomic DNA was isolated from the cells. Subsequently, PCR reactions were performed on the isolated DNA using pairs of primers that bind approximately 300 bp upstream and downstream of the homologous

regions of the integration sites (Target-A-D-Up-for, Target-A-D-Down-rev). Figure 28 shows an agarose gel on which the products of these PCR reactions were applied. Exemplary PCR products, which were generated from the monoclonal cell lines A-P2-G4 (CHO K1 target A, plate 2, well G4) and A-P3-G9 (CHO K1 target A, plate 3, well G9), are shown.

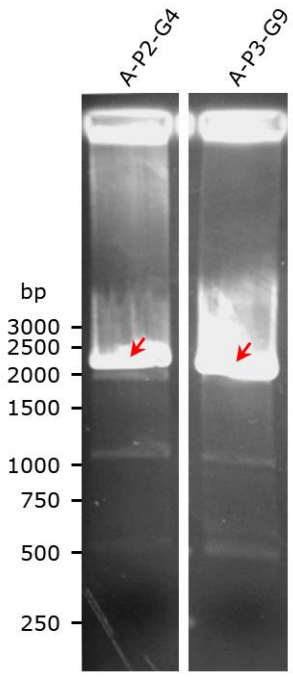


Figure 28: Agarose gels showing PCR products from genomic analysis of monoclonal cell lines A-P2-G4 and A-P3-G9. Primer "Target-A-Up-for" and "Target-A-Down-rev" were used for the PCR reaction.

For both clones a clear product band can be seen between the 2000 bp and 2500 bp markings (red arrows). In case of an integration of the intended construct in the target site, a PCR product band with a size of 5318 bp was expected. Without integration, the amplified genomic area should have a size of 2277 bp. The last point refers to the two generated PCR products, so in both cases no integration of the intended construct, at the target site, took place. The green fluorescence of the cells as well as the Zeocin resistance were probably caused in both cases by one (or more) random integrations of the insertion construct in the CHO K1 genomes. For the other monoclonal cell lines, no integration at the intended genomic position could be detected either (data not shown). Thus, no conclusions about the usefulness of the selected target sites could be drawn from this experiment.

After the insertion constructs for the adenoviral DNA fragment (pZMB0122 to pZMB0125) had been successfully cloned, analogous to the eGFPd2 constructs, transfection batches for the genomic insertions were performed. Since at this point it was already known that in previous experiments no cell lines could be generated using the CRISPR/Cas9 plasmids

derived from the plasmid pX462, only the pX459 based CRIISPR/Cas9 plasmids (pZMB0203 to pZMB0206) were used. In addition, a fifth transfection approach was performed using only the plasmid pZMB0122, which should lead to random integration without adding a CRISPR/Cas9 plasmid (in this case, supplementation of puromycin to the culture medium was not done). The experimental procedure corresponded to that of the previous series of experiments. No living cells could be detected in any of these five transfection approaches, within 14 days after transfection. The adenoviral DNA fragment containing the CDS for the genes E1A and E1B seems to have a lethal effect on CHO K1 cells.

Consequently, no further attempts were made to edit the genome of CHO K1 cells.

7.3.1.4 **Trials to produce rAAV2 using CHO K1 cells**

The next experiment should investigate whether it is possible to produce rAAV2 using CHO K1 cells and whether it is necessary to co-transfect the adenoviral DNA fragment coding for the genes E1A and E1B. In order to verify this, three different transfection batches were carried out in duplicates. rAAV2 particles were produced using plasmids pZMB0088 (pHelper), pZMB0216 (pSB1C3_001_Rep_VP123_453_587wt_p5tataless) and pZMB0311 (pAAV_ ITR_EXS_pCMV_mVenus_hGHpolyA) in CHO K1 suspension cells as well as in adherent growing HEK293 cells for control. In a third approach, the fourth plasmid pZMB0085 (pJET1.2_E1-full_from_Ad5) containing the adenoviral DNA was additionally transfected into the CHO K1 cells. The transfection of the HEK293 cells was performed using the $CaCl_2$ standard transfection method in two 100 mm dishes (see Chapter 6.4.1). The transfection of the CHO K1 suspension cells was performed using the PEI_{max} transfection method, starting with 5 ml transfection batches. This method was actually optimized for the transfection of only one plasmid construct for protein production [184]. Since higher concentrations of PEI could be toxic and the DNA to PEI ratio has a large impact on transfection efficiency, the total amount of DNA used for this experiment was retained as described in the protocol, but equimolar amounts of the three or four plasmids were used [217]. 72 h after the transfections, fluorescence microscopic images were made to control the transfection efficiency and the batches were harvested. Therefore the supernatants were first separated from the cells (see Chapter 6.4.2) and the cells were disrupted by adding 500 µl lysis buffer followed three freeze-thaw cycles. After the cell lysates were treated with Benzonase (see Chapter 6.4.2), they were combined with the respective supernatants (Resulting total volumes: 11.75 ml for each CHO K1 batch and 11.5 ml for each HEK293 batch). These solutions served as templates for the subsequent titration by qPCR (see Chapter 6.4.5). The results are shown in Figure 29.

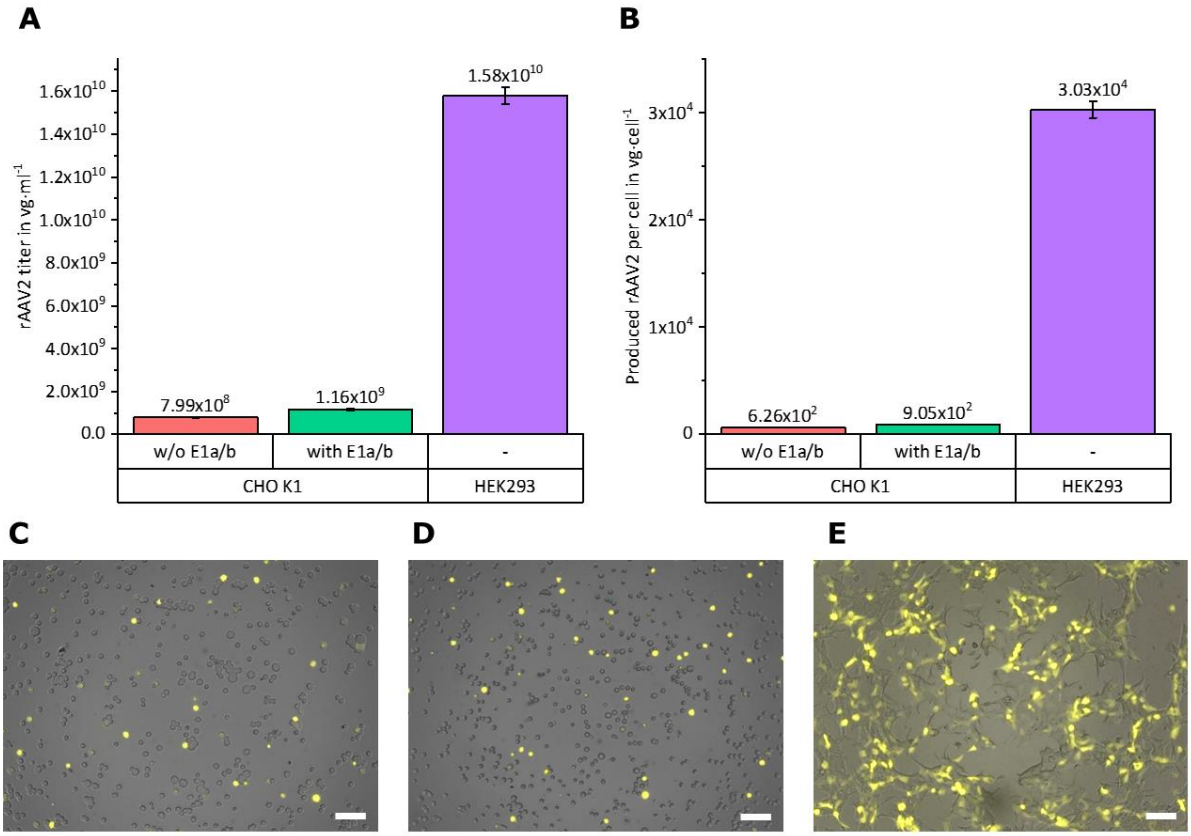


Figure 29: qPCR results of the rAAV2 wt production in CHO K1 cells with E1a/b genes and without E1a/b genes. A: qPCR results of rAAV2 titer in vg·ml^{-1}, B: Produced rAAV2 per cell in transfection batch in vg·cell^{-1}, C-E: Overlays of transmitted light images with mVenus fluorescence images, from different batches, for estimation of transfection efficiency. White bar indicates 100 µm. C: CHO K1 transfection batch without E1a/b (10.2 % transfection efficiency). D: CHO K1 transfection batch with E1a/b (9.8 % transfection efficiency). E: HEK293 transfection batch (about 85 % transfection efficiency).

Figure 29A shows the calculated rAAV2 titers in the generated suspensions. It can be seen that the rAAV2 titer with a value of $1.58 \cdot 10^{10}$ vg·ml^{-1} generated from HEK293 batches was about 13.6 or 19.8 times higher than the titers from the two CHO K1 batches, with approximately the same total volume. Nevertheless, it was possible to produce rAAV2 particles using CHO K1 cells. The addition of the plasmid containing E1A/B led to an increase in production of about 45.2 %. Thus, this DNA fragment was not essential for rAAV2 production, but it did increase production, although the gene dose used in this batch was reduced for all other genes, as already mentioned. For a better evaluation of the respective rAAV2 production performance of the two cell lines, the number of produced rAAV2 particles against the number of cells present during transfection is shown in Figure 29B. The resulting cell-related production performance is a better indicator for comparing the different approaches, as different cultivation methods and thus different cell titers were used in this experiment. Figure 29B shows the difference in rAAV2 production between HEK293 cells and CHO K1 cells even more clearly. Per HEK293 cell used, about $3.03 \cdot 10^4$ rAAV2 vg were produced, which is 33.5 or 48.4 times more than the CHO K1 cells were

able to produce. However, the relatively low production performance of the CHO K1 cells could be explained in part by the transfection method, which was not optimized for this system. In Figure 29 C to E, fluorescence microscopic images of the respective batches were made before the harvest, which were overlaid with transmitted light images. Due to the addition of the plasmid pZMB0311, the fluorescent protein mVenus was expressed as a byproduct of the rAAV2 production, under control of the CMV promoter on this plasmid. The resulting fluorescence signal can be used as a scale for the respective transfection efficiency. While the HEK293 transfection batches achieved a transfection efficiency of approx. 85 %, the CHO K1 transfection efficiency was only 10.2 % (without E1A/B) and 9.8 % (with E1A/B), respectively. In the context of protein production, CHO K1 cells achieve transfection efficiencies of over 80 % [184]. Thus, a successful optimization of the transfection method could increase the yield by a factor of about 8. Even then production on a CHO K1 basis could not keep up with HEK293 based rAAV2 production. However, the rAAV2 production was performed in CHO K1 suspension culture. Suspension cultures have some advantages over adherent growing cultures, such as significantly easier handling and, due to the usually higher cell titer, less space requirement. This could lead to a higher product output compared to an adherent culture even with a lower production yield per cell. In addition, serum-free medium is usually used for the cultivation of suspension cells. Due to its defined composition, this medium has advantages in the subsequent purification process. However, it might also be possible to adapt a HEK293 cell line to the growth in suspension culture, which could also result in such a cell line having the aforementioned advantages.

Since the optimization of a rAAV2 production process based on CHO K1 would take an enormous amount of time and probably would not have a great chance of improvement, compared to a HEK293 based system, no further work was done on the production of rAAV2 in CHO K1 cells within this thesis.

7.3.2 **Generation of HEK293 producer cell lines for rAAV2 using random integration**

After efforts to rationally generate rAAV producer cell lines on the basis of CHO K1 cells failed, the idea that rAAV production can be improved by a special production cell line was still kept in mind. As the pHelper plasmid is needed in every production, independent from the type of rAAV to be produced, this was a starting point to be considered for cell line optimization. The pHelper plasmid contains the genes E2A and E4 as well as the non-coding RNA VA, which are required for the production of rAAV. By integrating these sequences into the HEK293 genome, the transfection of the resulting production cell line with the pHelper plasmid could possibly be avoided during rAAV production. Because of the size of the pHelper plasmid, it accounts for about 50 % of the total plasmid DNA to be used in

usual productions. A reduction or even a total omission of this plasmid could significantly reduce the workload and costs of rAAV production.

Since the effects of genomic integration of further adenoviral sequences on the behavior of HEK293 cells are not foreseeable and probably strongly depend on the respective integration site, a rational approach, as for the CHO K1 cells, was not used for this project.

7.3.2.1 **Generation of monoclonal HEK293 cell lines with helper function for rAAV2 production**

For the integration of the pHelper plasmid sequence in the HEK293 genome, a eukaryotic selection marker was first added to the plasmid sequence, this work was done by Marco Radukic. For this purpose, the expression cassette of Blasticidin S deaminase from *Aspergillus terreus* was selected, which is located on the plasmid pcDNA6/myc-His from Thermo Fisher Scientific [218]. With the help of the primers bsd-SalI-for and bsd-NdeI-rev, the resistance cassette was amplified by PCR and the resulting PCR product was terminally extended by the restriction sites for the enzymes SalI and NdeI. The same recognition sequences are singular on the pHelper plasmid and are located adjacent to each other on an unannotated part of the plasmid sequence (see plasmid sequence in Chapter 10.3.1.1). The resistance marker was inserted into the pHelper plasmid via the appropriate digestion and finally the resulting plasmid pHelper-Bsd was stored under the number pZMB0259.

For genomic integration, about 50 µg of the plasmid pHelper-Bsd were linearized by digestion with SalI and then dephosphorylated (see Chapter 6.1.11 and 6.1.12). The successful linearization was confirmed by agarose gel electrophoresis in comparison to a non-digested plasmid sample (data not shown). $9 \cdot 10^6$ HEK293 cells were seeded in a T-150 flask 24 hours before transfection. 45 µg of the linearized plasmid were used for the $CaCl_2$ transfection of these cells (see Chapter 6.4.1). Three days after transfection, the growth medium was replaced by fresh growing medium supplemented with 10 $\mu g \cdot ml^{-1}$ Blasticidin. From this point onwards, the medium supplemented with antibiotics was exchanged every two days. During the media changes dead cells were removed by washing with 20 ml PBS. After a few days, small cell colonies were visible. 14 days after the first addition of Blasticidin, the cells were detached from the growth surface by a treatment with trypsin (see Chapter 6.2.3) and the resulting cell suspension was diluted for a limited dilution to 10 cells per ml for one 96-well plate (see Chapter 6.2.8). For cultivation a one to one mixture of fresh and sterile-filtered “used” growing medium was used, which was mixed with penicillin/streptomycin and the double amount of FCS (here 20 % (v/v)). The Blasticidin concentration was reduced to a total of 5 $\mu g \cdot ml^{-1}$ from this point on. Following the limited dilution, 23 wells were identified containing one single cell. The growth of these cells was observed microscopically for another two weeks. Four of the observed cell

colonies showed increased growth in comparison to the other colonies. These four monoclonal cell lines were selected and transferred into the wells of a 6-well plate. Ultimately, these clones were expanded, and experiments were made to characterize these cell lines. The four cell lines were named HEK293-KARE1a to HEK293-KARE1d.

7.3.2.2 **Characterization of monoclonal HEK293 cell lines with helper function for rAAV2 production**

After generating the four monoclonal cell lines HEK293-KARE1a to HEK293-KARE1d, further experiments were performed to characterize them. These works were carried out by Kathrin Teschner and Rebecca Feiner. Within this project the paper manuscript with the title "HEK293-KARE1, a cell line with stably integrated adenovirus helper sequences simplifies rAAV production" was created, which was submitted to the journal BMC Biotechnology. The manuscript, illustrations and supplementary information are attached in Chapter 10.1.2 and its subchapters. Overall, the stability of the integration was demonstrated for a period of more than four months. The new cell lines were compared with regard to growth and cell cycle to the original HEK293 cells, whereby no significant influences of the adenoviral genes could be detected. The RNA expression of E2A and E4 could be detected using RT-PCR at a stable low level comparable to the E1A gene. In addition, the rAAV2 production ability of these cell lines was analyzed and it was found that the cotransfection of the pHelper plasmid was not necessary when using these cell lines for the rAAV2 production. Even without the pHelper plasmid, rAAV2 yields were reached that are comparable with the original production system. The newly established cell lines offer a new alternative for rAAV production that saves both time and money.

Side note:

After completion of this project, the following unintended error was noticed in the construction of the pHelper-Bsd plasmid. Within the framework of the primer planning for the amplification of the expression cassette of Blasticidin S deaminase from the plasmid pcDNA6/myc-His, the primer binding sites were arranged in such a way that only a part of the Blasticidin S deaminase expression cassette was amplified. This part consists of the complete Blasticidin S deaminase CDS and the prokaryotic promoter and terminator sequences directly surrounding it. The eukaryotic expressions regulatory sequences framing this section were not included in the corresponding construct. Thus, after transfection of HEK293 cells with the SalI linearized plasmid (SalI cuts the construct upstream of the 5'-end of the Blasticidin S deaminase CDS), no direct expression of the Blasticidin S deaminase should be possible. An expression of the resistance protein can therefore only occur through a genomic random integration of the construct, which takes

place at sites in the genome with transcriptional activity. Furthermore, no or no important cellular function should be affected by this integration. This unintended circumstance probably led to the fact that relatively few HEK293 “colonies” developed after the transfection, since such an integration would be clearly more unlikely than a “normal” integration. However, this planning error probably also led to the fact that only clones were generated in which the integration took place at a transcriptionally suitable position, where the resistance gene as well as the viral helper genes were transcribed with such an intensity that no noticeable negative effects on cell viability were caused.

7.4 **Improvement of the rAAV2 purification by the development of new affinity chromatography materials**

In addition to the modification of the production cell line, an increase of the rAAV production could also be made by an improved downstream process. Common rAAV purification protocols are mostly based on complex and poorly scalable methods. This chapter focuses on two major subprojects of this thesis, which deal with the development of affinity chromatography materials for rAAV2 purification. First, an attempt was made to develop a single-chain construct based on the known AAV2-binding antibody A20, which could be used in different setups for rAAV2 purification. Later it became known that the AAV2 infection is also mediated by the receptor called AAVR. In the second project, an attempt was made to process rAAV2 particles based on an AAVR domain, which provides the required binding.

7.4.1 **Generation of single-chain antibody fragments based on the A20 antibody**

7.4.1.1 **Design and cloning of single-chain antibody fragments based on the A20 antibody**

The production of rAAV2 could not only be increased by manipulating the production cell line, but the subsequent purification of the rAAV2 particles also offers many potential targets for increasing the yield. Within the context of an AAV2 assembly study, hybridoma cell lines were generated to produce monoclonal antibodies against AAV capsid proteins in 1995. The subclone A20 showed no binding to VP proteins in western blot experiments [219]. Later it was shown that the monoclonal antibody A20 had a high affinity to assembled AAV2 particles [220]. Grimm *et al.* took advantage of this property to develop an AAV2 purification method using the A20 antibody for affinity chromatography in 1998 [138]. In 2000, a cryo-electron microscopy analysis further clarified the binding between an AAV2 particle and the Fab' fragment of the A20 antibody [221]. Based on this structure (PDB ID: 3J1S) a single-chain antibody (scFv) was generated within this thesis, which could be used for affinity chromatography purification of rAAV2 particles.

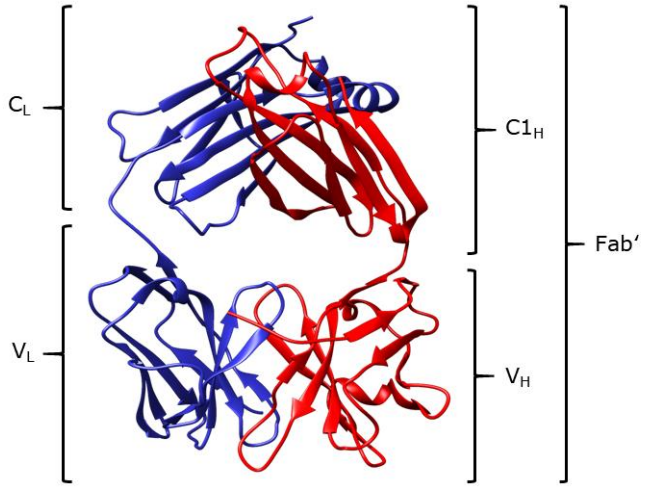


Figure 30: Ribbon diagram of A20 Fab' fragment based on PDB ID: 3J1S. Light chain is marked in blue, heavy chain is marked in red.

The protein design of this A20 scFv is discussed in detail here. The PBP structure file provides information that was used to design the scFv fragment. The crystal structure with the PDB ID 3J1S shows the Fab' fragment of the A20 antibody, the corresponding ribbon diagram is shown in Figure 31. Since the Fab' fragment consists of the variable and constant domain of the light chain and the variable and mainly the first constant domain of the heavy chain, it was necessary to decide up to which amino acid position the variable domains would be used to design the scFv.

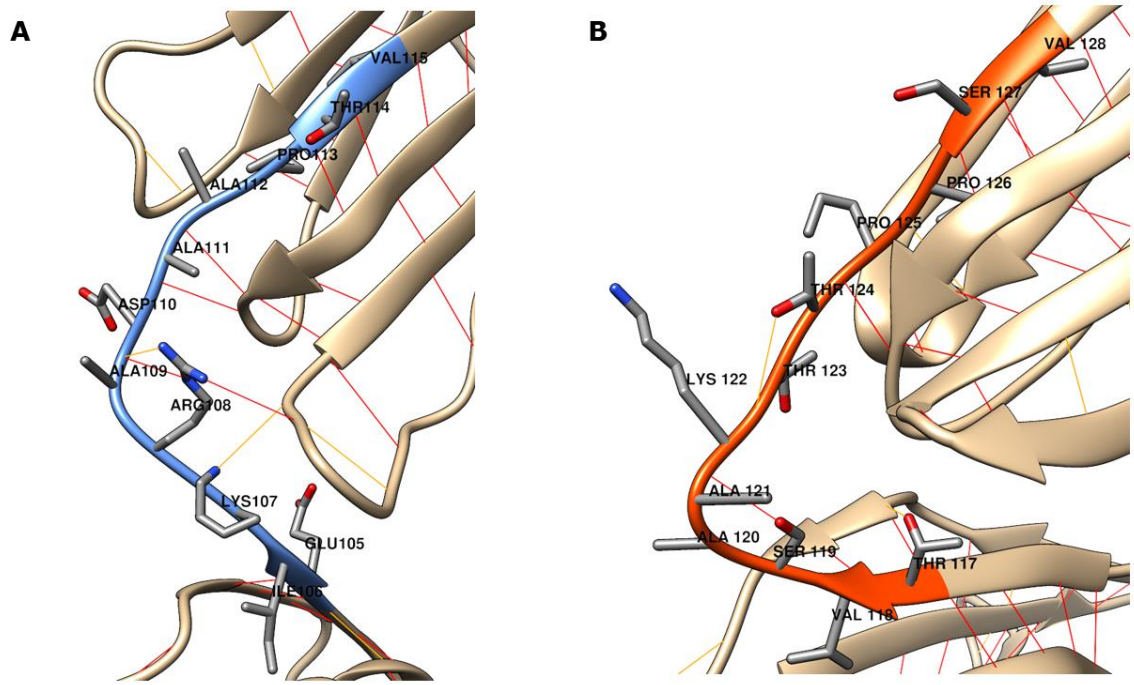


Figure 31: Information on the construction of the A20 scFv. A: Detailed section of the region between the variables and constant domain of the light chain, hydrogen bonds in red (strong) and orange (weak), B: Detailed section of the region between the variables and constant domain of the heavy chain.

Figure 31A shows a detailed section of the ribbon diagram of the area between the variable and constant domain (Glu105 to Val115) of the light chain. In addition, theoretical hydrogen bonds calculated with the software UFCS Chimera are shown as lines (red: strong hydrogen bond, orange: weak hydrogen bond). The figure shows that Glu105 and Ile106 still belong to a β-sheet of the variable domain. The nitrogen atom of the side chain of Lys107 forms a hydrogen bond to the oxygen atom of the side chain of Ser171. However, this potential hydrogen bond is large with a distance of approx. 3.5 Å and thus rather weak. The side group of Arg108, on the other hand, forms several hydrogen bonds to Ala109 and Asp170, making Arg108 probably more important for maintaining the structure of the constant domain than Lys107. As a result, it was decided to end the amino acid sequence of the variable light chain with the amino acid Lys107. In a similar way it was decided which amino acids might belong to the variable domain of the heavy chain. Figure 31B shows a detailed section of the ribbon diagram of the area between the variable and constant domain (Thr117 to Val128) of the heavy chain. Theoretical hydrogen bonds calculated with the software UFCS Chimera are shown as lines as before. The figure shows that the amino acids Thr117 and Val118 belong to a β-sheet of the variable domain. The subsequent amino acid Ser119 forms a hydrogen bond to Ala121, creating the “curved” structure of this region. The side groups of Lys122 and Thr123 seem to have no major structural significance. The amino acid Thr124, on the other hand, forms hydrogen bridge bonds to Lys122 and Phe153 and thus possesses structural properties. Two prolines and the amino acids Ser127 and Val128 follow this segment. Ser127 and Val128 are already part of a β-sheet in the constant domain. As proline is not suitable for flexible linker regions due to its “stiff” structure, they were not chosen. Finally, it was decided to end the amino acid sequence of the variable heavy chain with the amino acid Thr124. Although Thr124 has structuring elements related to the constant domain, the amino acids threonine as well as glycine and serine are frequently used in the construction of flexible linker sequences [222]. This way, a linker sequence was added at this point, partially keeping the natural sequence.

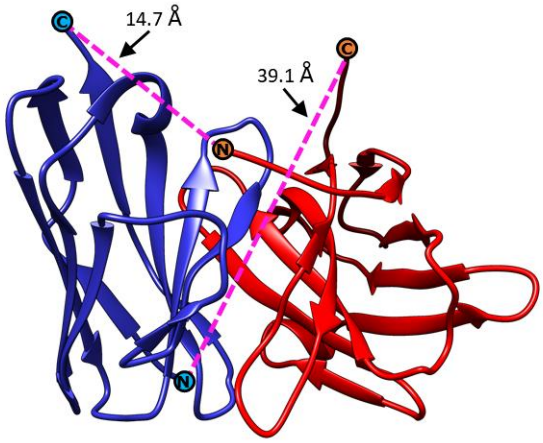


Figure 32: Distance calculation between linkable termini of variable domains. Distances are calculated with UFCS Chimera between the Cα-atoms of the terminal amino acids.

After defining the variable domains of the light and heavy chains of the A20 antibody, the next step was to decide in which order the domains will be linked via a flexible linker. Figure 32 shows the corresponding ribbon diagram for this purpose. The N- and C-termini of the light and heavy chain are highlighted. In addition, the UFCS Chimera software was used to determine the direct distances between the Cα-atoms of the linkable termini. Between the C-terminus of the light chain and the N-terminus of the heavy chain a distance of approx. 14.7 Å was calculated. The distance between the C-terminus of the heavy chain and the N-terminus of the light chain is 39.1 Å. In most scFvs, the C-terminus of the heavy chain is linked to the N-terminus of the light chain [223]. Due to the shorter distance between the C-terminus of the light chain and the N-terminus of the heavy chain, a linkage of these ends was more obvious than a linkage of the other two ends. Consequently, a shorter protein linker could be used, which usually leads to a more stable protein construct than a much longer protein linker would do. The direct connection of the C-terminus of the light chain with the N-terminus of the heavy chain does not sterically interfere with any part of the remaining protein. This is different with the straight line between the C-terminus of the heavy chain and the N-terminus of the light chain, here the direct path passes through the inner part of the protein, which means that a possible linker would need to be much longer than calculated. Another reason against the linkage of these ends was the fact that the N-terminus of the light chain is close to the CDS regions. If a linker sequence would be added to this end, it could influence the interaction of the scFv with the AAV2 particle. Finally, a $(GGGS)_4$ linker was chosen to link the C-terminus of the light chain with the N-terminus of the heavy chain, where the last serine of the linker is also the first amino acid of the variable heavy chain. This linker can reach a maximum distance of 50.8 Å in "stretched" form (calculates using UCSF Chimera). This corresponds to about 3.5 times of the required distance, which gives the protein linker the required flexibility. Therefore, the design of the protein sequence of the A20 scFv was final. The complete protein sequence is listed in Chapter 10.3.2.8 of the Appendix.

The cysteines Cys23 and Cys88 of the light chain and the cysteine Cys23 and Cys97 of the heavy chain, form disulfide bonds with each other. Therefore, a eukaryotic production system was chosen first. CHO cells are widely used for industrial antibody production [213]. Since good experience has already been made with CHO cells as a protein production system in previous works, the production of this A20 scFv should also be performed using CHO cells. For this reason, the designed A20 scFv had to be extended by some components. In order to enable secretory expression of the scFv, the protein construct was extended by the 22 amino acid long signal peptide sequence of the murine variants of the Ig-κ chain (UniProt: P01601) at its N-terminus. To be able to detect the construct later e.g. by ELISA and to be able to purify it, the C-terminal was extended by the sequence of the human IgG1-Fc part and a His_6-tag. A gene synthesis of this section could be omitted, since the IgG-Fc sequence was already present on the plasmid pZMB0034 of a previous work. This sequence encodes the 232 C-terminal amino acids of the human Ig gamma-1 chain C region, including the hinge region and the C_H2 and C_H3 domains of the antibody (UniProt: P01857). The corresponding DNA sequence was optimized for the *Cricetulus griseus* codon usage. Therefore, the A20 scFv sequence had been extended by the sequence of a 12 amino acid long glycine-serine linker, which additionally contains recognition sequences for the restriction enzymes KpnI and ApaI at the DNA level, in order to enable the subsequent addition of the Fc part. pcDNA5/FRT (pZMB0001) was selected as the production vector. Within the gene synthesis, the DNA sequence was optimized for the *Cricetulus griseus* codon usage and additionally extended by a Kozak sequence and the NheI recognition site at the 5'-end. The complete gene synthesis sequence is listed in Chapter 10.3.2.9 of the Appendix and a schematic illustration can be found in Figure 33A. After successful synthesis, the DNA fragment was amplified by PCR using the primers A20_GenSynth_CHO_for and A20_GenSynth_CHO_rev and inserted into the plasmid pZMB0001 via the restriction sites NheI and ApaI. The resulting intermediate construct got the number pZMB0381. To add the IgG Fc part and the His_6-tag sequences to the intermediate construct, a PCR was performed on the plasmid pZMB0034 using the primers KpnI-Fc-for and Fc-AgeI-His6-Stop-ApaI-rev. In addition to the amplification of the Fc, the recognition sites for the restriction enzymes KpnI and ApaI were added to the construct by primer overhangs. Furthermore, the recognition site for the restriction enzyme AgeI, a His_6-tag sequence and a stop codon (TGA) were added at the 3'-end of the construct. The resulting PCR product was inserted via the restriction sites KpnI and ApaI into the plasmid pZMB0381. With the additional recognition sequences for the restriction enzymes KpnI, AgeI and ApaI, both the IgG-Fc part and the His_6-tag could later be replaced by other sequences. The resulting final construct was stored under the number pZMB0382, a schematic illustration can be found in Figure 33B.

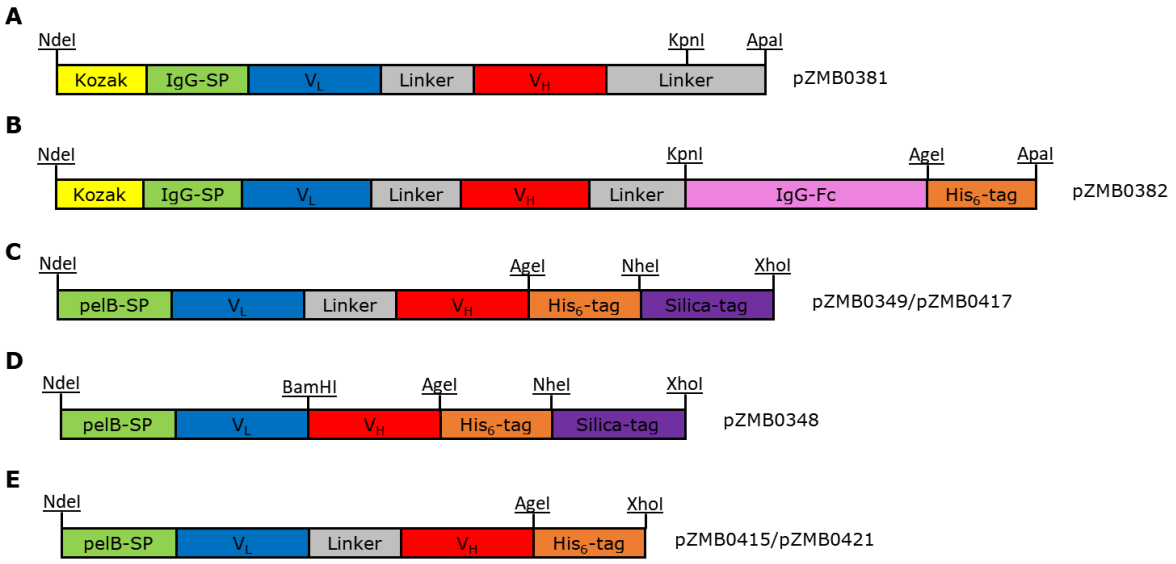


Figure 33: Schematic illustrations of A20 scFv constructs and intermediate constructs. A: Gene synthesis construct for the production of CHO K1 cells, B: Final construct for the production CHO K1 cells, C: Final construct and first gene synthesis for the production in E. coli cells, D: Second gene synthesis for production in E. coli cells, E: Variant of the E. coli construct without a silica-tag.

Since a bacterial expression system is generally faster, cheaper and usually more productive than a eukaryotic expression system, the production of the A20 scFv using *E. coli* should also be investigated. A second gene synthesis was planned for this purpose, which includes a version suitable for the expression using *E. coli* cells. The protein sequence of the A20 scFv should remain unchanged. Due to the disulfide bonds contained in the scFv construct, cytosolic production in *E. coli* cells alone is not possible. By exporting the expressed protein into the periplasm of gram-negative bacteria, the formation of disulfide bonds would be possible, due to the reducing environment. A commonly used signal peptide that promotes the export of the expressed protein into the periplasm is the 22 amino acid long pelB sequence which refers to pectate lyase B of *Erwinia carotovora* CE [224]. This leader sequence was N-terminal attached to the protein sequence of the A20 scFv. In order to be able to process the A20 scFv better and to be able to detect it later e.g. by Western blot, a His_6-tag sequence was added to the C-terminus. Furthermore, the protein construct should be able to be bound to a low-priced chromatography material for its later use. In 2010 a peptide sequence was discovered by Ikeda *et al.* which had a high affinity to silica material [225]. Silica is mainly used in chemistry as a cheap chromatography column material. In 2014, a variant of the originally 273 amino acid long silica-binding peptide, which was shortened to 14 amino acids, was published, which is supposed to be very suitable as a purification tag [226]. This silica tag sequence was added to the C-terminus sequence of the His_6-tag. On DNA level, the scFv sequence was optimized for the codon usage of *E. coli*. At the 5'-end, the sequence begins with a NdeI recognition sequence that overlaps with the ATG start codon. After the stop codon (TAA) the

recognition sequence for the restriction enzyme XhoI was added at the 3'-end of the DNA sequence. Via the two restriction sites for NdeI and XhoI, the DNA could be inserted into the expression vectors pET21a and pET24b. In order to be able to make subsequent changes to the tag sequences, recognition sites for the restriction enzyme AgeI was inserted in between the sequences of the variable heavy chain and the His_6-tag sequence and a NheI recognition sequence was inserted in between the His_6-tag sequence and the sequence of the silica tag. Both recognition sequences overlap with the surrounding DNA sequences so that only one amino acid was added to the protein sequence (glycine and alanine). The complete DNA sequence of the gene synthesis can be found in the appendix under Chapter 10.3.2.8. This DNA sequence could not be synthesized by GeneArt. The reason for this was "too many repetitive sequence sections". These repetitive DNA segments appeared especially within the linker sequence, which links the light with the heavy chain. Thus, a further gene synthesis was planned, in which the corresponding $(GGGS)_4$ linker sequence (the last serine is actually the first amino acid of the heavy chain) was completely removed and replaced by the sequence 5'-GGATTC-3' (Gly-Ser), which is the recognition sequence for the restriction enzyme BamHI (see Figure 33D). This DNA construct was successfully synthesized. It was amplified by PCR using the primers A20_GenSynth_EColi_for/rev and the PCR product was inserted into the plasmid pZMB0230 (pET21a) via the restriction sites NdeI and XhoI. The resulting intermediate construct was stored under the number pZMB0348. For the reconstruction of the originally intended sequence, the SLiCE cloning method was particularly suitable, since in this case the initial plasmid could be opened using BamHI and the original linker sequence could be inserted at this point by homologous recombination. The insert fragment, which contains the linker sequence in addition to the 20 bp long, terminal homologous arms, was generated by oligo hybridization using the two oligonucleotides A20-scFv-Linker-Ecoli-SliCE-for/rev. The non-homologous bases of the BamHI recognition site were removed by recombination. The final construction was stored under number pZMB0349. In addition, the insert was cut out of the plasmid pZMB0349 via the restriction sites NdeI and XhoI and inserted into the plasmid pZMB0262 (pET24b backbone) to exchange the plasmid backbone and thus the resistance gene. This plasmid was stored under the number pZMB0417.

Later in this thesis, the oligo hybridization of the oligonucleotides AgeI-His-S-XhoI-for and AgeI-His-S-XhoI-rev and the subsequent insertion of the resulting DNA fragment into the plasmids pZMB0349 and pZMB0417 via the restriction sites AgeI and XhoI, led to the plasmids pZMB0415 and pZMB0421. This modification generated constructs that had no silica tag but a C-terminal His_6-tag. A schematic representation of this constructs can be found in Figure 33E.

7.4.1.2 **Production of the A20-scFv_His$_6$_Silica using *E. coli* BL21(DE3)**

The optimization of the cultivation conditions for the *E. coli* BL21(DE3) strain, which carries the plasmid pZMB0349 (coding for A20-scFv_His$_6$_Silica), will be discussed first. The cultivation and subsequent generation of protein extracts by osmotic lysis was performed as described in Chapter 6.3.4. Cultures with a total volume of 500 ml were prepared in LB medium. To optimize the cultivation conditions, only the temperature of the main cultivation (after IPTG induction) and the following cultivation time were varied. A total of five different conditions were evaluated. For a cultivation time of 18 hours, the product yields were investigated at cultivation temperatures of 37 °C, 25 °C and 15 °C. For a cultivation time of six hours, temperatures of 37 °C and 25 °C were analyzed.

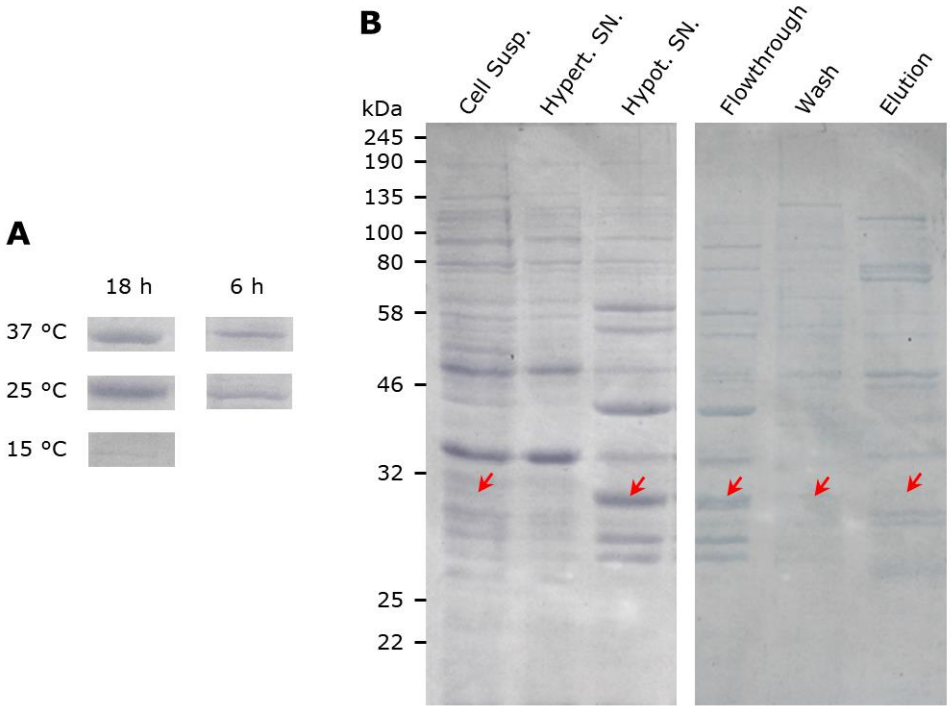


Figure 34: Sections from SDS PAGEs showing the optimization of A20-scFv_His$_6$_Silica production, corresponding E. coli cultivation and subsequent IMAC purification. A: SDS PAGE sections showing the optimization of A20-scFv_His$_6$_Silica production at different cultivation temperatures and cultivation times. B: SDS PAGE showing samples of E. coli cultivation for the production of A20-scFv_His$_6$_Silica. Cell Susp.: Cell suspension at harvest, Hypert. SN.: hypertonic supernatant, Hypot. SN.: hypotonic supernatant, Flowthrough, wash and elution of IMAC purification.

Figure 34A shows the bands that were at the level of the expected target protein (29.2 kDa) on an SDS PAGE. Only the respective sections from the hypotonic lysis extracts are shown. A comparison of the individual product bands shows that for a cultivation time of 18 hours at both 37 °C and 25 °C, the same amount of product was produced. By reducing the cultivation time to six hours, the quantity of product in both cases decreased in approximately equal proportions. In the case of cultivation at 15 °C for 18 hours, the lowest

amount of product was visible. Several studies had shown that the folding of proteins usually works better at lower temperatures than at higher temperatures. Because of this and because the product yield at the 37 °C and the 25 °C cultivations were about the same, cultivations at 25 °C for 18 hours were performed in further experiments. As part of the optimization experiments, an *E. coli* BL21(DE3) strain carrying the plasmid pZMB0417 was also cultivated. This plasmid differed from the plasmid pZMB349, which was used for the other optimization experiments, only by the backbone. Instead of a pET21a backbone this one had a pET24b backbone. The main difference between these two backbones is the resistance gene. pET24b has a kanamycin resistance gene instead of an ampicillin resistance gene. Since previous experiments had shown that 18-hour cultivation yields more product than 6-hour cultivations, it was advisable to use a selection system that can maintain the selection pressure over a longer period of time. A kanamycin-based selection system is more suitable for this, than an ampicillin-based selection system, since ampicillin has a significantly lower temperature stability than kanamycin [227]. However, an expression experiment using the *E. coli* BL21(DE3) strain carrying the plasmid pZMB0417 under the previously optimized cultivation conditions (25 °C cultivation temperature after induction, 18 h cultivation time) showed no increase in product output (data not shown).

Figure 34B shows an SDS PAGE on which samples from an *E. coli* cultivation, subsequent osmotic lysis and IMAC purification of the corresponding target protein were applied. This cultivation was carried out under the optimized conditions. The total volume of the culture was 1 l. The cell pellet obtained at the end of cultivation had a wet mass of 2.5 g, so 25 ml of hypertonic and hypotonic buffer were used for osmotic lysis. On the left side of Figure 34B, the first lane shows a sample of the cell suspension at the time of cell harvest. The expected target protein A20-scFv_His_6_Silica has a molecular weight of 29.2 kDa and is marked with a red arrow. When comparing the lane in which the cell suspension was applied with the lane in which the sample of the hypertonic supernatant was applied, a high consistency of the band pattern is noticeable. Most of the bands of both lanes were running at the same height and have a comparable intensity. This suggests that cytosolic proteins were also released by this type of lysis. A band of the target protein is not visible in the lane of the hypertonic supernatant, whereas a very weak band of the target protein is visible in the lane of the cell suspension. The band pattern of the lane from the hypotonic supernatant differs in the type of bands as well as in the respective intensities from the hypertonic supernatant lane. A clearly visible band running at the expected height of the A20-scFv_His_6_Silica protein can be identified here (red arrow). This band was also the most pronounced band on the lane of the hypotonic supernatant. Although the band of the target protein was rather weak compared to other cytosolic protein expressions using *E. coli*, it was sufficiently large due to the fact that the protein had to be exported into the periplasm.

The right side of Figure 34B shows samples from the subsequent IMAC purification (flowthrough, wash fraction and elution), for which the hypotonic supernatant was used as starting material. Prior to this, the extracted hypotonic supernatant (approx. 25 ml) was diluted with twice the volume of IMAC equilibrating buffer (approx. 50 ml) to create better binding conditions. The lane of the flowthrough had the same band pattern as that of the hypotonic supernatant, but the intensity of the bands was lower, because of the dilution. The band of the A20-scFv_His_6_Silica was also clearly visible (red arrow). In the two following lanes (wash fraction and elution) the bands of the target protein (red arrow) are only very slightly recognizable beside some more prominent bands of contaminating proteins. This indicates that the target protein did not adequately bind to the chromatography material. The UV signal of the elution fraction had a maximum amplitude at approx. 130 mAU, which indicated a low total protein concentration. Due to the fact that the band of the A20-scFv_His_6_Silica of the elution fraction was hardly recognizable, and the total protein concentration was low, it was not useful to continue working using this sample. One possible reason, for the failure of the IMAC purification, could be an inaccessibility of the His_6-tag. The His_6-tag of the A20-scFv_His_6_Silica construct is located intramolecularly near the C-terminus. As shown in Figure 32, the C-terminus of the A20-scFv heavy chain protrudes from the compact structured part of the scFv. Due to this naturally exposed structure, an extension of the A20-scFv_His_6_Silica sequence by a linker sequence at this position was omitted, because an accessibility of the His_6-tag should be possible. However, it could not be excluded, that the C-terminal silica tag had a negative effect on the binding of the His_6-tag to the IMAC chromatography material. To exclude this, a 1 l *E. coli* BL21(DE3) culture was cultivated, which carries the plasmid pZMB0421 (coded for A20-scFv_His_6), osmotically lysed and purified by IMAC. Also, in this case, the subsequent SDS PAGE shows a comparable picture as in Figure 34 (data not shown). The band of the target protein was only visible in the flowthrough but not in the elution fraction so the C-terminal silica tag of the construct A20-scFv_His_6_Silica seems not to be the reason or not the only reason for the previously failed purification.

7.4.1.3 **Experiments to establish an affinity chromatography method based on A20-scFv_His$_6$_Silica**

7.4.1.3.1 **Binding of A20-scFv_His6_Silica to silica material**

Chapter 7.4.1.2 describes that the fusion protein A20-scFv_His$_6$_Silica could be produced periplasmatic using *E. coli* BL21(DE3), but subsequent purification via IMAC failed. Compared to a complete cell disruption, osmotic lysis results only in the release of periplasmic proteins, which already provides a kind of cell extract, which has an increased purity, due to the non-disruption of the cytosol. Considering the aim of this project, which was to generate a fast and inexpensive purification method based on affinity chromatography for AAV2, it would be desirable if the IMAC purification could be avoided anyway. For this purpose, a sample of 100 mg silica was incubated with hypotonic supernatant, which came from a A20-scFv_His$_6$_Silica production and subsequent osmotic lysis of the cells. An SDS PAGE, shown in Figure 35, was performed on which samples taken during this experiment were applied. In the first lane (Hypot. SN.) a sample of the hypotonic supernatant was applied. The band of the 29.2 kDa A20-scFv_His$_6$_Silica proteins is clearly visible (red arrow). The second lane (Unbound) shows the supernatant that had formed after incubation with the silica material and thus contained proteins that had not bound to the silica material. It can be clearly seen that the band of A20-scFv_His$_6$_Silica protein decreased in intensity, which suggests a binding of the A20-scFv_His$_6$_Silica proteins to the silica material. In addition, the intensities of some bands of contaminating proteins had also decreased slightly. Samples of the three wash fractions were applied in the three subsequent lanes (Wash 1-3). It was remarkable that the bands of the A20-scFv_His$_6$_Silica protein in all three lanes of the washing fractions showed approximately the same, however weak, intensity. In contrast to that, the bands of impurity proteins that were most present in the second lane were only visible in the first washing fraction lane (yellow arrows). Due to the porous structure of the silica material and the washing steps carried out one after the other in batches, a step-by-step dilution of the impurities was expected. This explains the weak bands of impurity proteins in the first wash fraction. The constant intensity of the A20-scFv_His$_6$_Silica bands in all wash fractions could therefore only be explained by the fact that under the chosen washing conditions, a fraction of the bound A20-scFv_His$_6$_Silica proteins was detached from the silica material. The last lane (Bound) of the SDS PAGE shows a sample of the silica material boiled in Lämmli buffer. There was a significant increase in the intensity of the A20-scFv_His$_6$_Silica band (red arrow) detectable, which was also the thickest band in this lane. In addition, several bands of contaminating proteins could be identified which, like A20-scFv_His$_6$_Silica protein, had also bound to the silica material. In this lane A20-scFv_His$_6$_Silica represents the largest amount of bound protein. Due to many impurities

and the fact that a fraction of A20-scFv_His_6_Silica was detached from the silica material by the washing process, it could be assumed that the maximum binding capacity for proteins with silica tag has not yet been reached in this setup. A corresponding experiment was also performed using the hypertonic supernatant from an A20-scFv_His_6_Silica production (data not shown). As shown in Figure 34A, no A20-scFv_His_6_Silica protein could be detected in the hypertonic supernatant. The binding experiment did not result in any enrichment of A20-scFv_His_6_Silica to the silica material.

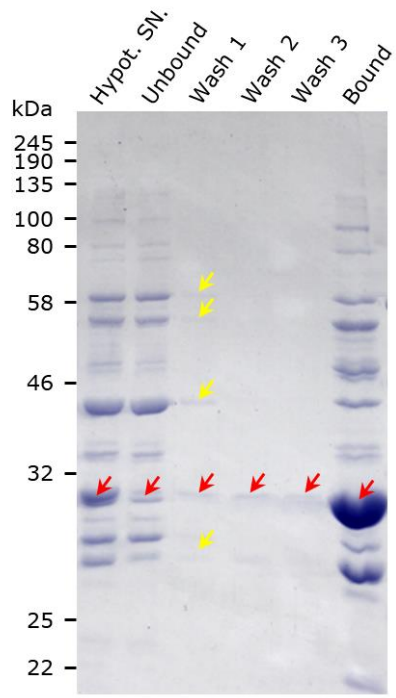


Figure 35: SDS PAGE showing A20-scFv_His_6_Silica binding experiment to silica material. Hypot.SN.: hypotonic supernatant from A20-scFv_His_6_Silica production, Unbound: unbound supernatant after silica incubation, Wash 1–3: supernatants of wash fractions 1-3, Bound: silica material after protein binding.

In subsequent experiments, it should be investigated whether the chromatography material, generated this way, was already suitable for the purification of rAAV2. However, since quantitative statements regarding possible binding capabilities were not possible due to this production method, an attempt was first made to remove the contaminating proteins from the chromatography material by optimizing the wash conditions.

7.4.1.3.2 **Experiments to remove impurities from A20-scFv_His$_6$_Silica chromatography material**

The following experiment should reveal whether it was possible to remove the contamination proteins from the hypotonic supernatant, which bind together with the A20-scFv_His$_6$_Silica, to the silica material. Eight wash buffers containing different sodium chloride concentrations and Tween-20 amounts were tested for this purpose. All wash buffers consisted of 25 mM Tris-HCl at pH 8.0. Two buffers each contained 150 mM, 300 mM, 600 mM or 1 M sodium chloride, of which one was added with 0.05 % (v/v) Tween-20 (see Table 5) were used. A gram silica material was incubated for this purpose with hypotonic supernatant from an A20-scFv_His$_6$_Silica production. Based on this, nine batches each containing 100 mg of silica material were generated, eight of which were washed three times each using one of the different wash buffers. One sample served as a control and was not washed. Samples were taken from the different wash fractions and from the resulting silica material and were applied on SDS PAGEs. A representative section of such an SDS PAGE is shown in Figure 36A. The first two lanes show the hypotonic supernatant (Hypot. SN.), which served as starting material for silica binding, and the unbound supernatant after silica incubation (Unbound). The band pattern was identical to the band pattern shown in Figure 35A and indicated that the binding of A20-scFv_His$_6$_Silica to the silica material worked well. The sixth lane (Bound Ctl.) from Figure 36A shows the silica sample of the unwashed control. As before, a clear enrichment of the A20-scFv_His$_6$_Silica on the silica material could be detected (red arrow). Lanes 3 to 5 show samples of the first and third wash fractions (Wash 1, Wash 3) as well as the silica sample after the washing procedure (Bound). In this case the samples came from the washing procedure using the buffer containing 1 M sodium chloride and 0.05 % (v/v) Tween-20, which was the most stringent washing buffer. In the samples of the two washing fractions, the protein A20-scFv_His$_6$_Silica and some weaker bands of contaminating proteins, which were washed off the material, were detectable. The intensity of all the bands was rather weak, indicating that low amounts of protein, both from A20-scFv_His$_6$_Silica and from the impurities, were removed from the material. A direct comparison of the resulting silica sample (Bound) and the unwashed silica reference sample (Bound Ctl.) shows that the washing procedure had hardly any effect, the intensities of all protein bands were nearly identical. In Figure 36A only a section of the total SDS PAGE is shown, because all other washing buffers led to the same result. It was therefore not possible to remove the impurities from the silica material after binding, using the washing buffers mentioned above.

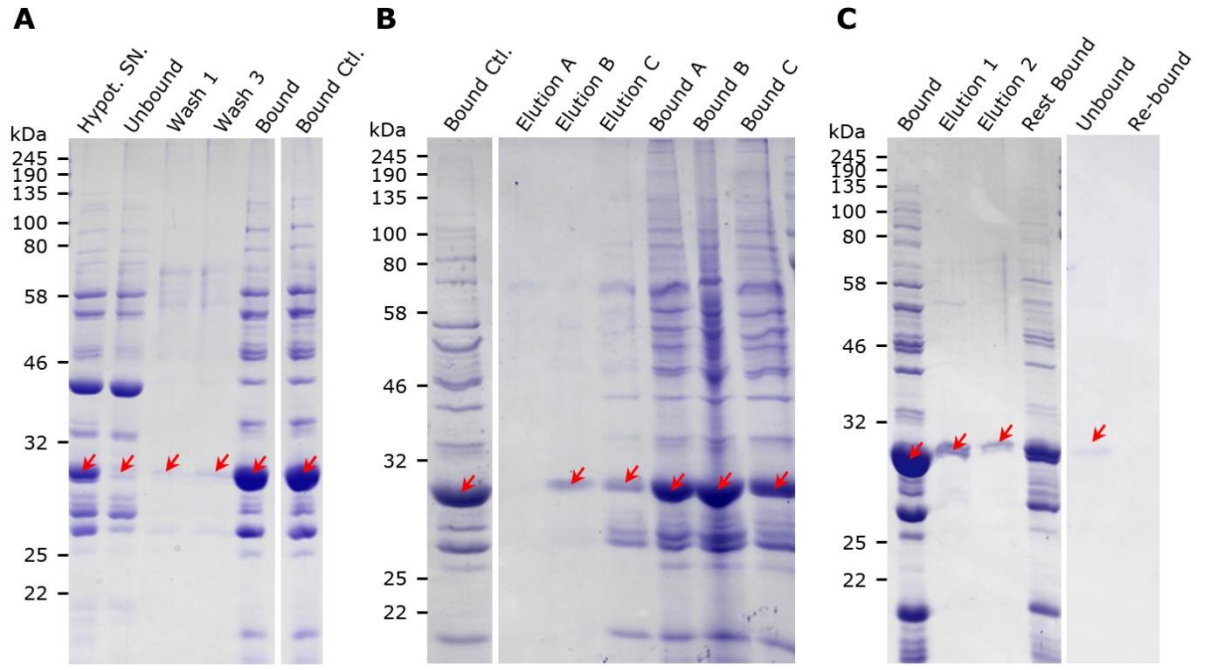


Figure 36: SDS PAGEs of wash-, elution-, and rebound-experiments of A20-scFv_His$_6$_Silica chromatography material. A: SDS PAGE of washing experiment. Hypot.SN.: hypotonic supernatant from A20-scFv_His$_6$_Silica production, Unbound: unbound supernatant after silica incubation, Wash 1 and 3: supernatants of wash fractions 1 and 3, Bound: silica material after protein binding, Bound Ctl.: silica material after protein binding without prior washing. B: SDS PAGE of elution experiment. Bound Ctl.: silica material after protein binding without elution, Elution A–C: elution fraction using elution buffer A-C, Bound A–C: silica material after elution procedure using elution buffer A - C. C: SDS PAGE of rebound experiment. Bound: silica material after protein binding, Elution 1 and 2: elution fraction 1 and 2, Rest Bound: remaining protein on silica material after elution procedure, Unbound: unbound supernatant after second silica incubation, Re-bound: silica material after second binding attempt.

An experiment was performed to investigate whether it was possible to elute the A20-scFv_His$_6$_Silica protein bound to the silica material and/or the bound contaminating proteins by using different elution buffers. Acidic and alkaline elution buffers should be investigated first, since treatments with salt-rich buffers as described above had hardly shown any effect. At first silica material (total 500 mg) was incubated with the hypotonic supernatant from a A20-scFv_His$_6$_Silica production. Afterwards the loaded silica material was washed three times using PBS and four samples of 100 mg silica material each were taken from the washed material. One sample served as reference and was not treated with elution buffer. The other three samples were treated with one of the following elution buffers. Elution buffer A: 100 mM glycine HCl pH 3.0, Elution buffer B: 100 mM citric acid pH 3.0 and Elution buffer C: 100 mM triethanolamine pH 11.5. The elutions were carried out twice using the double initial volume. Samples of the combined elution fractions, the silica materials after elution treatment and an uneluted reference sample of the silica material were applied to an SDS PAGE which is shown in Figure 36B. A direct comparison of the three elution samples (Elution A-C) shows that the elutions were overall not so

strong, which indicated that the binding of the different proteins to the silica material was rather strong. The elution sample using elution buffer A (Elution A) did not show any other visible proteins apart from a weak band at 70 kDa, so this elution did not work sufficiently. In the second elution sample (Elution B), in which the citric acid buffer was used, a clear band at the level of the A20-scFv_His_6_Silica protein could be recognized (red arrow). In the background of this lane only a few rather weak bands of contaminating proteins could be seen. It seemed as if the citric acid buffer was leading very selectively to the elution of the A20-scFv_His_6_Silica proteins. This elution buffer could therefore be a good candidate for a purification of A20-scFv_His_6_Silica using the silica tag. In the lane of the elution sample C (Elution C) an elution of A20-scFv_His_6_Silica (red arrow) could be seen, which showed a similar intensity as in elution B. However, the amount of contaminating proteins was significantly higher compared to elution B. Looking at the silica samples from all three batches (Bound A-C), it can be seen that a lot of protein remained on the silica material. Only in the case of elution C, a minor decrease of all attached proteins could be detected, which matches the corresponding elution lane. Although all investigated elutions were incomplete and rather weak, a further analysis of an elution using the citric acid buffer was performed, which had the aim to produce A20-scFv_His_6_Silica protein in purer quality.

In the next step, the A20-scFv_His_6_Silica protein from hypotonic extract should be bound to silica material, then the protein should be eluted using a citric acid buffer and the eluted protein should be rebounded to the silica material again, resulting in a cleaner chromatography material. Therefore, one gram of the silica was incubated with the hypotonic extract of an A20-scFv_His_6_Silica production. The silica material was washed three times using PBS and then two elution steps were performed using the citric acid elution buffer. In contrast to the method described in Chapter 6.3.7, the incubation time with the elution buffer was doubled (two times for 30 min). The eluates were combined and rebuffered against PBS a total of five times using an Amicon Spin Filter (1000 Da) before being concentrated to approx. 1 ml. The eluate treated this way, was then incubated with 25 mg silica material and subsequently washed three times using PBS. The results of the following SDS PAGE analysis are shown in Figure 36C.

The first lane (Bound) shows a sample of the silica material after incubation with the cell extract. As in the previous experiments, the binding of the A20-scFv_His_6_Silica protein (red arrow) is visible. The two following lanes (Elution 1 and 2) show samples of the first and second elution using the citric acid buffer. It can be seen that the elutions were quite clean. The amount of eluted protein decreased slightly in the second elution compared to the first elution. The fourth lane (Rest Bound) shows a sample of the silica material after elution. A significant reduction of all previously bound proteins could be observed, also the band of A20-scFv_His_6_Silica protein had decreased in intensity. In the following lane (Unbound), the supernatant was applied which had formed after the incubation of the

rebuffered eluate with the silica material. Only a weak band at the level of A20-scFv_His_6_Silica proteins could be recognized (red arrow). The next lane (Re-bound) was generated from the silica material sample which should bind the eluate again. No protein was detectable in this lane. Therefore, it was not possible to bind the eluted A20-scFv_His_6_Silica protein a second time to silica material, under these conditions. A possible reason could be different binding conditions. To check this, the experiment was repeated, but this time the eluted A20-scFv_His_6_Silica protein was not rebuffered with PBS, but with the hypotonic lysis buffer (5 mM $MgSO_4$), which was usually the basic agent for the first silica binding. The following binding experiment (same procedure as described here) showed also no binding (data not shown). It was also remarkable that after rebuffering, the concentration of A20-scFv_His_6_Silica decreased, although this method should increase the concentration. This could be determined by comparing the lanes of the first and second elution (Elution 1 and 2) with the lane of the unbound supernatant (Unbound), which contains almost the entire remaining amount of protein (see Figure 36C). Due to the fact that this small amount of A20-scFv_His_6_Silica protein remained from an *E. coli* culture with a total volume of about one liter, no further attempts were made to rebind A20-scFv_His_6_Silica to the silica material for further purification.

7.4.1.3.3 **Elution experiments using A20-scFv_His6_Silica chromatography material**

The next experiment should reveal whether the produced (unclean) silica-based affinity chromatography material could withstand specific elution conditions. In Grimm *et al.* it had already been shown, that the complete A20 antibody produced in this study, covalently bound to sepharose, was suitable for the purification of AAV2. A tris buffer containing 2.5 M $MgCl_2$ was used for the elution [138]. Since the currently existing chromatography material consists of silica material to which the A20-scFv_His_6_Silica and further contaminating proteins from the periplasmatic *E. coli* extract had bound, it should be investigated whether $MgCl_2$-rich elution buffers influences the binding of the proteins bound to the silica material. If *E. coli* proteins or the A20-scFv_His_6_Silica would be found in the eluate after using such a buffer, this could cause problems for subsequent applications.

To evaluate this, hypotonic *E. coli* cell extract was bound to a total of 400 mg silica material as before. This material was then washed three times using PBS and was divided into three batches of 100 mg each. Each of these batches would undergo a double elution process using a $MgCl_2$-rich elution buffer. Therefore, 50 mM Tris-HCl buffers at pH 7.0 which contained either 2.0 M, 2.5 M or 3.0 M $MgCl_2$ were used.

The following figure shows an SDS PAGE of samples taken from the elution experiment using the buffer containing 2.0 M $MgCl_2$. The first lane (Bound) shows a silica sample that was not treated with elution buffer. The usual band pattern showing the thick A20-

scFv_His_6_Silica band (red arrow) is clearly visible (compare e.g. to Figure 35 or Figure 36). In the two elution fractions (Elution 1 and 2) no protein is detectable on the SDS PAGE. The fourth lane (Rest Bound) shows the silica sample after the elution treatment. A clear reduction of the band intensities can be seen. It was noticeable, that after adding the elution buffer to the silica material, a slight crystallization occurred, which made it difficult to pipette the silica material. This could be an explanation for the decrease of the band intensities of the fourth sample. The corresponding samples of the two other elution experiments are not shown in Figure 37, as they show the same result apart from slight variations of the band intensity of the silica samples after elution caused by the crystallization problem. Whether a $MgCl_2$-rich elution buffer was generally suitable could not be clarified at this point, since a precipitation of the silica-bound proteins could not be ruled out. The crystallization of the elution buffer, which occurred in all three cases, would probably cause problems in the corresponding applications. However, the clear eluates contained no protein impurities.

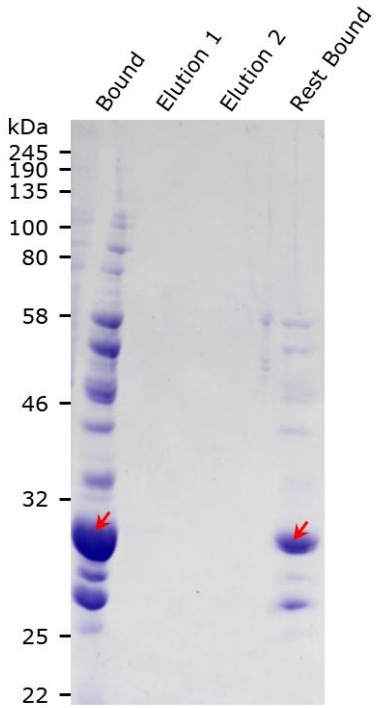


Figure 37: SDS PAGE of $MgCl_2$ elution experiment using 50 mM Tris-HCl with 2.0 M $MgCl_2$ pH 7.0 buffer for elution. Bound: silica material after protein binding, Elution 1 and 2: elution fraction 1 and 2 using Tris buffer with 2.0 M $MgCl_2$, Rest Bound: remaining protein on silica material after elution procedure.

7.4.1.3.4 **Binding of rAAV2 to A20-scFv_His$_6$_Silica chromatography material**

The recent results showed that the A20-scFv_His$_6$_Silica protein could be bound to silica material (see Chapter 7.4.1.3.1). At this point, it should be analyzed whether it is possible to bind rAAV2s to this material. For this purpose, A20-scFv_His$_6$_Silica chromatographic material was generated by binding the hypotonic *E. coli* extract to the silica material. 25 mg of this chromatography material were incubated with 1 ml HEK293 culture supernatant from a rAAV2 wt production for 15 min shaking at room temperature. For comparison, 1 ml of the same HEK293 culture supernatant was incubated with 25 mg silica material to which A20-scFv_His$_6$_Silica *E. coli* supernatant was not bound before. Following the rAAV2 incubation, the supernatants were separated from the silica materials by centrifugation (5000×g for 2 min) and the rAAV2 titer were determined from these samples by qPCR analysis. The results of this qPCR quantification are shown in Figure 38.

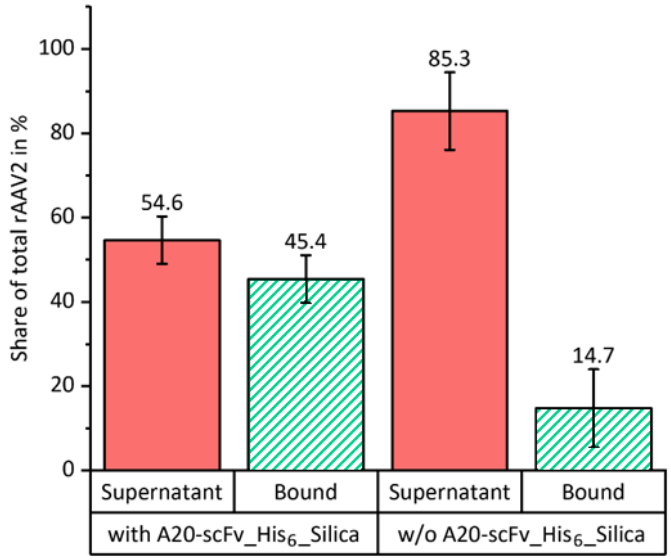


Figure 38: Binding experiment of rAAV2 from HEK293 supernatant to silica material incubated with or without E. coli lysate containing A20-scFv_His$_6$_Silica. Supernatant: relative rAAV2 titer in supernatant after incubation, Bound: calculated amount of bound rAAV2 relative to total starting amount. Data based only on calculations are displayed with hatches.

The results of the qPCR quantification shown in Figure 38 were normalized to the total amount of rAAV2 from the HEK293 culture supernatant (100.0 ±1.4 %). In addition, the liquid content of the wet silica pellets was taken into account for the quantifications. In the supernatant of the sample containing the silica material to which A20-scFv_His$_6$_Silica was previously bound only 54.6 ±5.6 % of the initially used rAAV2s were detectable. In the negative control, which did not contain A20-scFv_His$_6$_Silica, 85.3 ±9.2 % of the rAAV2 could be found in the supernatant. This means that about 45.4 ±5.6 % of rAAV2 bound to the silica material based on A20-scFv_His$_6$_Silica and about 14.7 ±9.2 % of rAAV2 bound to the "naked" silica material. Because the rAAV2 starting material in this case was HEK293

culture supernatant and thus the binding took place under real conditions as it would occur in the context of a purification method, binding of about half of the used rAAV2s is quite acceptable. The proportion of bound rAAV2s in the negative control was unexpectedly high with 14.7 ±9.2 %. However, as these qPCR samples had a larger error bar (±9.2 %) and pipetting imprecisions had not been considered, that could be an explanation for the deviation.

Next, an attempt was made to elute the bound rAAV2s. For this purpose, the experiment described above was up scaled by a factor of ten, to prevent problems at the detection of rAAV2 using qPCR. Then, 250 mg A20-scFv_His_6_Silica chromatographic material were generated. This chromatography material was incubated with 10 ml HEK293 culture supernatant from a rAAV2 wt production for 15 min shaking at room temperature. After that the supernatant was removed by centrifugation (5000×g for 2 min), three washing steps were performed using 5 ml PBS each time. Afterwards 2 ml elution buffer were added to the sample and the suspension was incubated for 15 min on a roller tumbler mixer. The elution was performed two times. Samples for rAAV2 quantification by qPCR were taken from all steps of this procedure, the corresponding results are shown in the following figure.

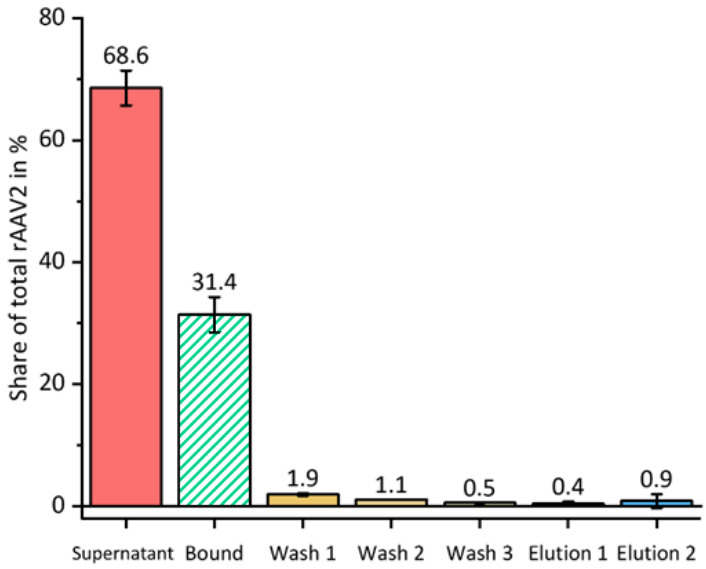


Figure 39: Big scale binding and elution experiment of rAAV2 from HEK293 supernatant to silica material incubated with E. coli lysate containing A20-scFv_His_6_Silica. Supernatant: relative rAAV2 titer in supernatant after incubation, Bound: calculated amount of bound rAAV2 relative to total starting amount, Wash 1-3: amount or rAAV2 found in wash fraction 1–3, Elution 1 and 2: Amount or rAAV2 found in elution fractions 1 and 2. Data based only on calculations are displayed with hatches.

The results of the qPCR quantification shown in Figure 39 were normalized to the total amount of rAAV2 from the HEK293 culture supernatant (100.0 ±2.8 %). Additionally, the liquid content of the wet silica pellets was taken into account for the quantifications (see Chapter 6.3.7). The supernatant of the incubated sample contained 68.6 ±2.9 % of the total rAAV amount. Therefore 31.4 ±2.9 % of the rAAV2s were bound to the

chromatography material. This was 14.0 % less than in the experiment using a lower scale (see Figure 38). However, both the rAAV2-containing HEK293 culture supernatant and the hypotonic supernatant of *E. coli* lysis, came from different production batches than those used in the previous experiment, which could explain this deviation. The three wash fractions contained between 1.9 ±0.2 % and 0.5 ±0.1 % of the total rAAV2 used. The value of the first wash fraction with 1.9 ±0.2 % was unexpectedly low because of the residual liquid in the porous structure of the silica material, the first wash fraction should be a dilution of the remaining flowthrough and should therefore contain at least 3.2 % of the originally used rAAV2. That the measured value was lower than expected may also be due to the fact the qPCR reaction reacts very sensitively to different sample compositions as experience had shown. The flowthrough fraction consisted mainly of the HEK293 culture supernatant, while the wash fractions consist mainly of PBS. The two elution fractions contain 0.4 ±0.3 % and 0.9 ±1.1 % of the rAAV2 amount and therefore a total of about 1.3 % of the initial rAAV2s. This corresponds to 4.1 % of the rAAV2 bound to the chromatography material in the first step. Such a low elution rate was not expected in this experimental setup, since comparable conditions were used for the elution as described in Grimm *et al.'s* work and the amount of elution was significantly higher with a share of >65 % [138]. The low elution rate could not be explained by the possibility that the A20-scFv_His_6_Silica has a higher affinity to rAAV2 than the original A20 antibody. If this was true, a higher percentage of the rAAV2s from the HEK293 culture supernatant should bind to this material. It could be possible, that the elution may not proceed completely as the addition of the elution buffer resulted in a slight crystallization of the silica material, which could have a negative effect on the elution result. In the quantification of the elution fractions, the result of the qPCR could also be influenced by the varying sample solutions. For this reason, following experiments will be performed with more defined starting materials, such as rAAV2 samples that had already been purified e.g. by ultracentrifugation and were present in clean solutions. This way, more concentrated samples could also be used, and the scale of the processing experiments could be reduced, which would increase the feasibility and comparability of the results.

The last experiment presented raised doubts about the correct binding of rAAV2 to the A20-scFv_His_6_Silica chromatography material. In addition, the percentage of rAAV2 bound to the chromatography material varies significantly between the individual experiments and other experimenters not described here (see Figure 38 and Figure 39). For all previous experiments rAAV2-containing HEK293 culture supernatant was used as starting material. In order to prevent further variations in the quality of the starting material and to achieve more reproducible results, the experiment presented at the beginning of this chapter should be carried out once again using rAAV2 purified by ultracentrifugation instead of rAAV2-containing HEK293 culture supernatant. Therefore, a rAAV2 wt production from a total of 20 100 mm dishes was purified by ultracentrifugation

(see Chapter 6.4.4). The resulting rAAV2 suspension was rebuffered against PBS and had a genomic rAAV2 wt titer of $1.36 \cdot 10^{10}$ vg·ml^{-1} (total volume approx. 500 µl). As before, two batches were carried out using 25 mg silica material per batch. One of the silica samples was first incubated with hypotonic lysate from an A20-scFv_His_6_Silica production in *E. coli* (see Chapter 6.3.4), the other sample consisted of untreated but washed silica material. To both samples 1 ml of a 1:10 dilution with PBS of the purified rAAV2 were added. Then both samples were incubated for 15 min shaking at room temperature. After incubation, the supernatants were separated from the silica material by centrifugation (5000×g, 2 min). The rAAV2 titers were then determined by qPCR analysis (see Chapter 6.4.5). The results of this analysis are shown in Figure 40.

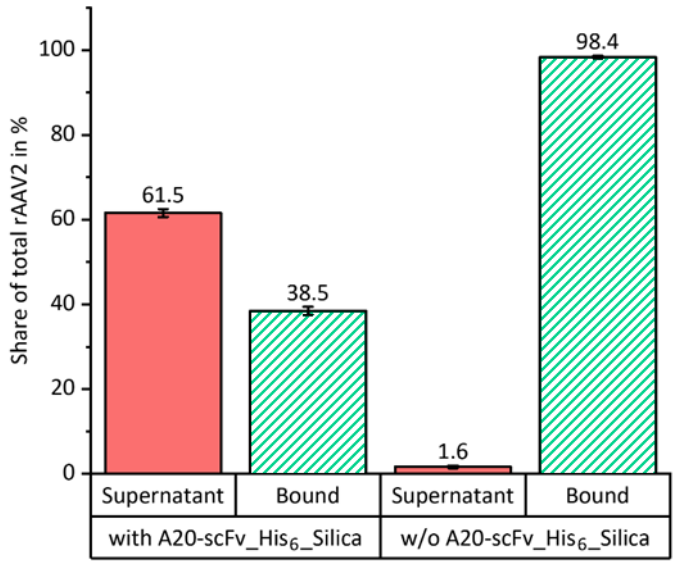


Figure 40: Binding experiment of rAAV2 purified by ultracentrifugation to silica material incubated with or without E. coli lysate containing A20-scFv_His_6_Silica. Supernatant: relative rAAV2 titer in supernatant after incubation, Bound: calculated amount of bound rAAV2 relative to total starting amount. Data based only on calculations are displayed with hatches.

The qPCR results from Figure 40 show that 38.5 ±1.0 % of the rAAV2 bound to the silica material treated with A20-scFv_His_6_Silica containing *E. coli* lysate. This lies in the middle of the bond proportion (about 31.4 % to 45.4 %) achieved in the two experiments shown above (see Figure 39 and Figure 38). This means that the use of clean raw material did not increase or significantly reduce the rAAV2 bond to the chromatography material. The second batch, on the other hand, showed a 98.4 ±0.3 % fraction of rAAV2, which was bound to the untreated silica material. This clear result indicated that a very strong unspecific binding of the rAAV2 to the silica material existed, if the binding was not disturbed by cellular impurities or other proteins. Thus, the silica material used in the previous experiments was not suitable as a base material for an affinity chromatographic purification process for rAAV2. This result also leads to an alternative interpretation of the results in Figure 38. The untreated silica sample (negative control), which showed a rAAV2

binding of 14.7 ±9.2 %, probably only bound such a small percentage of the rAAV2, because potential binding sites on the surface of the silica material were blocked by components from the HEK293 culture supernatant. On the other hand, the sample from this experiment, which was previously treated with *E. coli* lysate containing A20-scFv_His_6_Silica, showed an about three times higher rAAV2 bond (45.4 ±5.6 %). Since in this case the used silica material was blocked twice during pretreatment using the *E. coli* lysate and during the incubation with the HEK293 culture supernatant, the resulting rAAV2 binding or the proportion of the binding which exceeds the proportion of the negative control could only be explained by the functional bonding behavior of the bound A20-scFv_His_6_Silica proteins.

Assuming that the *E. coli* lysate blocks the possible rAAV2 binding sites on the silica material as well as the HEK293 culture supernatant did, it could be expected that the 38.5 ±1.0 % binding rate of the *E. coli* lysate treated sample from Figure 40 was also attributable to the functionality of the A20-scFv_His_6_Silica protein and not due to unspecified binding of the rAAV2 to the silica material. Due to the numerous handling problems as well as the unspecific binding of many proteins and rAAV2 particles, it was decided that the silica material is not suitable material for affinity chromatography of rAAV2 particles.

7.4.1.4 **Production of A20-scFv_hIgG1-Fc_His_6 antibody in cell culture**

Chapter 7.4.1 dealt with the design of the A20-scFv_hIgG1-Fc_His_6 protein. This construct was designed for rAAV2 purification and detection. The resulting plasmid pZMB0382 codes for the protein A20-scFv_hIgG1-Fc_His_6 using the *Cricetulus griseus* codon usage and is suitable for expression in eukaryotic cells due to the used CMV promoter.

First, the A20-scFv_hIgG1-Fc_His_6 antibody fragment should be produced in CHO K1 suspension cells. As described in Chapter 6.3.1, a 40 ml transfection batch was generated, which was then transfected using PEI_{max} and the plasmid pZMB0382. The resulting production culture with a total volume of approx. 100 ml was harvested six days after transfection. The collected supernatant was purified by a protein A column. In the subsequent photometric analysis of the elution fractions (see Chapter 6.1.6), no significant amount of protein could be detected in the eluate. A repetition of the production, where transfection was carried out using the linear variant of PEI (linear PEI, molecular weight 25000 Da) instead of PEI_{max}, did also not result in any significant amounts of protein in the eluate from the subsequent purification using a protein A column. Later, HEK-F suspension cells adapted to the growth in HEK-TF medium (from Xell AG) were used for transfection instead of CHO K1 suspensions cells. The transfection method using PEI_{max} differs only in one point from the method used for CHO K1 cells. Here, the HEK-TF medium could be used

for both as a growth medium and as a transfection medium. Also, in this case a 40 ml transfection batch was generated resulting in a production batch of about 100 ml. After subsequent protein A purification and rebuffering of the eluate against PBS using an Amicon spin filter, a protein solution with a total volume of 3.0 ml and a concentration of 0.8 mg·ml^{-1} could be obtained (monomeric form of A20-scFv_hIgG1-Fc_His$_6$ without leader peptide: 54.6 kDa, E: 96'885). This quantity was sufficient for further experiments. An SDS PAGE on which a sample of the resulting A20-scFv_hIgG1-Fc_His$_6$ protein solution was applied, confirmed the high purity of the product (date not shown).

7.4.1.5 **Analysis of A20-scFv_hIgG1-Fc_His6's binding ability to rAAV2 wt particles**

After the successful production of A20-scFv_hIgG1-Fc_His$_6$ using HEK293 suspension cells and subsequent purification using protein A, the binding ability to rAAV2 wt particles should be investigated by ELISA. Therefore, rAAV2 wt particles were bound onto a MaxiSorp plate overnight. A total of three different batches was analyzed. For the first batch, a 1 to 250 dilution of the 50 µg·ml^{-1} A20-scFv_hIgG1-Fc_His$_6$ stem suspension was used for binding after the blocking step. To detect the binding of the A20-scFv_hIgG1-Fc_His$_6$ to the rAAV2 wt particles a HRP-coupled anti-human secondary antibody was used. Each of the other two attempts served as negative references. In the second batch no A20-scFv_hIgG1-Fc_His$_6$ was added and in the third batch, wells were used, on which no rAAV2 wt particles were immobilized before. Otherwise, all three approaches were treated equally. The results of the subsequent absorption measurement are shown in Figure 41.

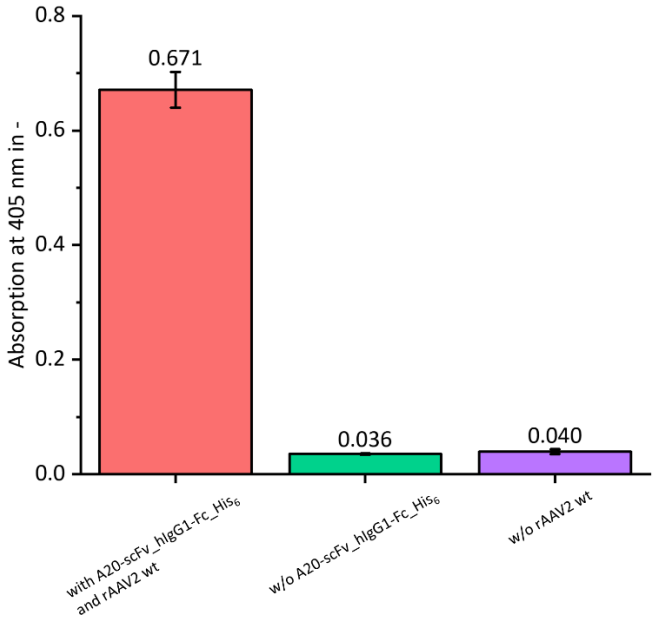


Figure 41: Results of rAAV2 wt capsid ELISA. Results of different approaches: 1. column: A20-scFv_hIgG1-Fc_His$_6$ and rAAV2 wt; 2. Column: without A20-scFv_hIgG1-Fc_His$_6$ 3. Column: without rAAV2 wt.

Figure 41 shows very clearly that an increase in the absorption signal had occurred in the first batch (0.671 ±0.032). Here, the A20-scFv_hIgG1-Fc_His_6 antibody was used as intended for the described capsid ELISA (see Chapter 6.4.8). In case of the other two setups, with 0.036 ±0.001 and 0.040 ±0.004, an increase of the absorption signal was hardly detectable. This shows that the binding of A20-scFv_hIgG1-Fc_His_6 to rAAV2 wt particles worked and that no cross-reactions between the secondary antibody and rAAV2 wt capsids or the A20-scFv_hIgG1-Fc_His_6 and BSA occurred in this ELISA arrangement. Thus, the binding ability of the A20-scFv_hIgG1-Fc_His_6 antibody produced in HEK293 cells to rAAV2 wt particles was proven.

In the following the binding capacity of the A20-scFv_hIgG1-Fc_His_6 antibody should be compared to that of the commercial A20 antibody in a competitive ELISA. However, since no experimental approach of a competition ELISA (see Chapter 6.4.8) has led to useful results, no results are presented here.

7.4.1.6 **Further experiments using the A20-scFv_hIgG1-Fc_His_6**

Since the successful production of the A20-scFv_hIgG1-Fc_His_6 antibody, it had been used in the laboratory for many experiments, because it is a cost-effective alternative to the commercial A20 antibody. Within the dissertation of Dinh To Le ELISA binding experiments were performed. Dinh To Le compared the binding behavior of the commercial A20 antibody with the A20-scFv_hIgG1-Fc_His_6 antibody with respect to conventionally produced rAAV2 wt particles and rAAV-VLPs produced with *E. coli* and assembled *in vitro,* which only consisted of the VP3 capsid proteins. The AAV capsid ELISA protocol corresponded to the method described in Chapter 6.4.8. Interestingly, he was able to observe clear differences in binding behavior. The absolute signal strength measured from the A20 antibody and the A20-scFv_hIgG1-Fc_His_6 with respect to rAAV2 wt particles was comparatively high. The A20-scFv_hIgG1-Fc_His_6 signal was 13.0 % lower than that of the A20 antibody. It was not possible to clarify at this point whether this difference was caused due to a different binding behavior. It was more likely that the signal strength difference was caused by the binding behavior of the used secondary antibodies, because for the A20 antibody an anti-mouse was used and for the A20-scFv_hIgG1-Fc_His_6 an anti-human HRP-coupled secondary antibody was used. Based on this slight difference, it was even more astonishing that the absolute signal differential of the *in vitro* assembled VLP was significantly higher for the A20-scFv_hIgG1-Fc_His_6. In this case, the measured mean signal was 85.6 % higher than that of the A20 antibody. The VLPs assembled *in vitro* were based only on VP3 and probably had a slightly different capsid structure than the rAAV2 wt variant. However, since statistically 50 of the 60 capsid proteins are VP3, this difference should not be very high. The higher signal strength in ELISA caused by the A20-scFv_hIgG1-Fc_His_6 could therefore only be explained by a higher tolerance of the A20-

scFv_hIgG1-Fc_His_6 antibody for slight variations of the capsid structure. Thus, the A20-scFv_hIgG1-Fc_His_6 antibody is more suitable for the detection of such VLPs than the commercial A20 antibody.

7.4.2 **Generation of affinity chromatography constructs based on PKD2 of AAVR**

7.4.2.1 **Design and cloning of affinity chromatography constructs based on PKD2 of AAVR**

During the work on the different A20-scFv constructions, it became clear that these constructions would be less suitable for a potential rAAV2 affinity chromatography due to their insufficient bacterial producibility. By improving production using human cells, it was possible to produce enough A20-scFv to have sufficient material for applications such as ELISA. However, the necessity of a eukaryotic production system missed the original purpose of this project, which was the generation of a simple, fast and inexpensive rAAV2 purification method. Thus, an alternative design which could satisfy the requirements was searched for.

In 2016, Pillay *et al.* generated a library of mutagenized haploid HAP1 cells, which had knockouts of almost all non-essential genes of the human genome, through the use of retroviral gene-trap vectors. After this library was infected with rAAV2, which mediated the expression of RFP, uninfected cells could be sorted from the library. The insertion sites of the gene-traps were then mapped against the genome and a total of 46 essential genes for the rAAV2 infection were identified. Besides genes that are involved for the heparan sulfate proteoglycan biosynthesis and other genes known to play a role in endocytic trafficking and retrograde transport systems, one hit was particularly prominent, a knockout of the gene KIAA0319L [28]. This gene encodes a type I transmembrane protein which contains a MANSC domain, five polycystic kidney disease (PKD) domains, and a C6 region near the N-terminus. At this time it was known that this gene had been linked to dyslexia and that it is involved in neural migration and axon guidance [29, 30]. This receptor has since been known as the AAV receptor (AAVR). Furthermore, in the context of the work of Pillay *et al*., the binding of rAAV2 to the AAVR could be narrowed down to the second PKD domain (PKD2).

In March 2019, after completion of the experimental work of this thesis, Zhang *et al*. published a high-resolution cryo-EM structure of the AAV2-AAVR complex, providing a detailed description of the interaction of the AAV2 capsid with the receptor. The interaction of the AAV2 capsids with the PKD2 domain mainly takes place via the variable regions I, III, VI and VIII of the VP proteins (see Figure 42) [228]. While the structural analysis of

the AAV2-A20 Fab fragment had shown that the epitope extends from the plateau to the side of the spike and into the conserved canyon of the fully assembled capsid [221]. This different binding behavior, which is not dependent on the fully assembled capsid but only on sequence segments of individual capsid proteins, could be taken into account in the future when generating chimeric rAAV2 serotypes in the context of retargeting and thus limit the applicability of a possible purification system not only to serotype 2 (and serotype 3, which is also bound) [229, 230].

Due to the excellent binding properties of the PKD2 domain to rAAV2 particles and the fact that the PKD2 domain has neither disulfide bridges nor glycosylation sites, this receptor fragment seems to be an ideal candidate for the generation of an affinity chromatography construct, which could also be produced using *E. coli*.

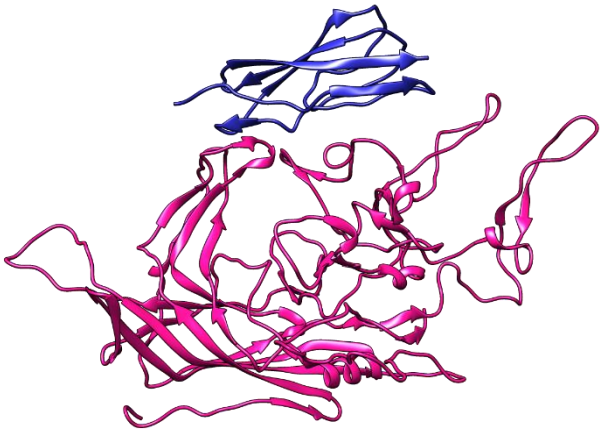


Figure 42: Structure model of the AAV2_VP1-AAVR_PKD2 complex (PDB ID: 6IHB). AAV2_VP1 is shown in pink and AAVR_PKD2 in blue.

The choice of silica gel as the base material for the A20-scFv based purification applications not successful. Thus, an alternative base material was sought. Finally, cellulose was chosen. Cellulose is available inexpensively in high purity and can be used in various forms (e.g. as a sheet in the form of blotting paper or in powder form as microcrystalline cellulose (Avicel)). In addition, cellulose is chemically very stable, has good mixing properties in the form of Avicel and pre-experiments have shown that cellulose does not bind rAAV2 particles unspecifically (see Chapter 7.4.2.2). Since 1993, the cellulose binding domain (CBD) of an exoglucanase from the organism *Cellulomonas fimi* (CBDcex) has been successfully used in fusion proteins to bind to cellulose [231, 232], this binding domain seemed to be a good choice as fusion partner in combination with the PKD2 domain for the generation of an affinity chromatography binding element for rAAV2 particles.

A string DNA synthesis was ordered from GeneArt for the cloning of the affinity chromatography fusion protein. This fusion protein was constructed as follows; a schematic representation of the sequence elements is shown in Figure 43A. The construct starts with

the restriction site NdeI, which simultaneously codes for the start codon (ATG). This is followed by the sequence coding for a His_6-tag. The next central element, the PKD2 domain sequence, is separated from the His_6-tag by a short linker sequence (GGSG). The PKD2 amino acid sequence corresponds to the PKD2 domain deposited under the UniProt number Q8IZA0. As also described in Pillay *et al.*, the 90 amino acid long sequence is terminally extended by the four surrounding amino acids (405-502), which are part of the corresponding intermediate sequences between PKD1 and PKD3 [233]. Next to this there is the natural 21 amino acids long proline-threonine linker and the 107 amino acids long sequence coding for the CBDcex as described in Ong *et al.* [232]. Following the stop codon (TAA) there is the restriction site XhoI. The entire DNA sequence was optimized for the codon usage of *E. coli.* After successful synthesis, the DNA fragment was amplified by PCR using the primers His-AAVR_PKD2-PT-CBD-for/-rev and inserted into the plasmid pZMB0262 (pET24b backbone) via the restriction sites NdeI and XhoI. The resulting plasmid construct got the number pZMB0505.

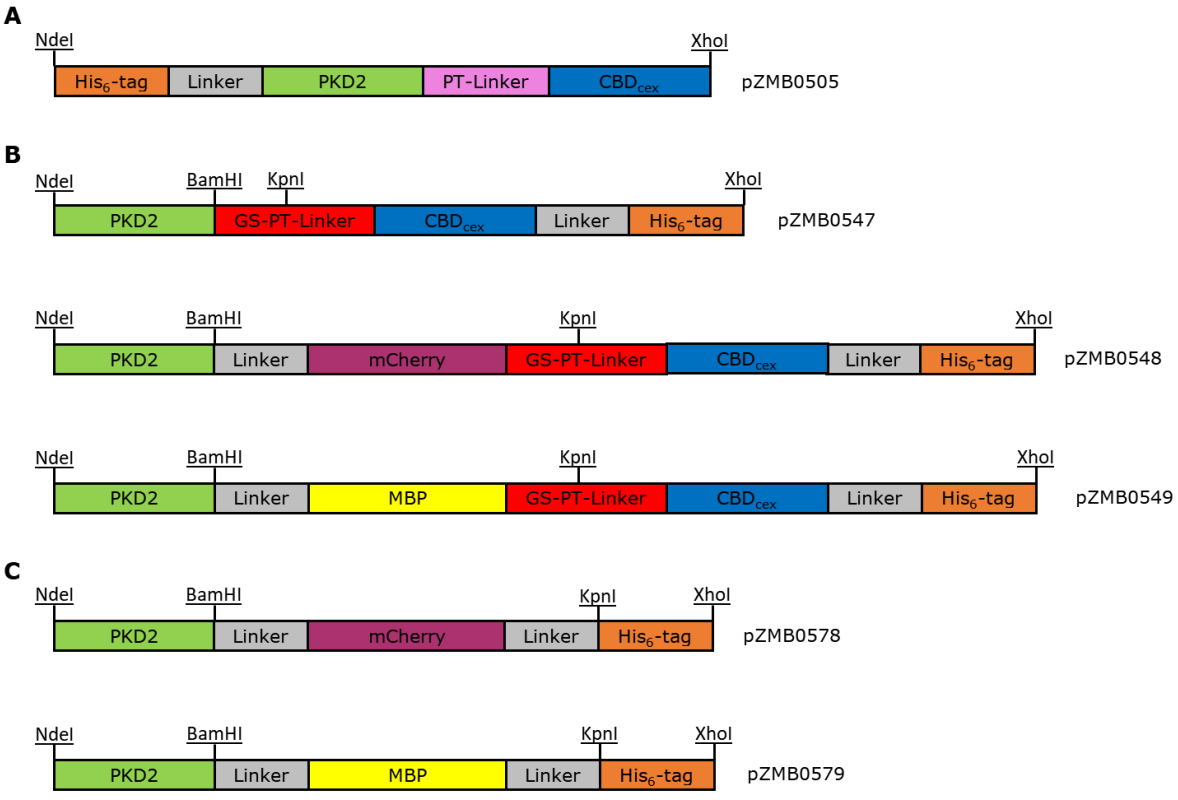


Figure 43: Schematic illustrations of the first (A), second (B) and third generation (C) of plasmid constructs for the generation of an affinity chromatography binding element for the purification of rAAV2 based on the PKD2 domain of AAVR.

After some experiments had shown that the fusion protein His_6_PKD2_CBDcex expressed with plasmid pZMB0505 was poorly soluble and leads to inclusion bodies during expression in *E. coli*, a second generation of this construct was designed. Essentially, the second generation of this construct differed from the first construct by inserting solubility

enhancing proteins between the PKD2 domain and the CBDcex, along with minor changes such as the C-terminal positioning of the His_6-tag and the shortening of the proline-threonine-linker including an addition of a glycine serine linker section. These solubility enhancing properties had already been described for the maltose binding protein (MBP) and for the fluorescent protein mCherry [234, 235]. In order to be able to generate such constructs, an intermediate construct based on the plasmid pZMB0505 was first created. First, the sequence of the PKD2 domain was amplified from the plasmid pZMB0505 using the primers NdeI-PKD2-for and PKD2-BamHI-Linker-KpnI-rev. The overhangs of the primers led to a PCR product being extended by the recognition sequence for the restriction enzyme NdeI at the 5'-end. At the 3'-end, the PKD2 sequence was extended by the recognition sequence for BamHI, a nine amino acid glycin-serine-linker and the recognition sequence for KpnI. By a second PCR reaction on the plasmid pZMB0505 using the primers KpnI-PT-CBD-for and CBD-His-S-XhoI-rev, the CBDcex sequence and the sequence of the last eight amino acids of the prolin-threonine linker were amplified. The overhangs of the primers lead to a PCR product being extended by the recognition sequence for the restriction enzyme KpnI at the 5'-end. At the 3'-end the CBDcex sequence was extended by the sequence for a His_6-tag, a stop codon and the recognition sequence for the restriction enzyme XhoI. Following these two PCR reactions, the products were cut with KpnI and ligated together after agarose gel extraction. The ligation product was cut using NdeI and XhoI and inserted into the plasmid pZMB0262 (pET24b backbone) via these sites. The resulting plasmid was stored under the number pZMB0547, a schematic representation of the plasmid is shown in Figure 43B.

To generate the intended plasmid constructs, the sequence coding for mCherry was first amplified by PCR from the plasmid pZMB0490 (pET21a-EGF-mCherry-His6) using the primers BamHI-mCherry-for and mCherry-KpnI-rev. The overhangs of the primers added parts of the required glycine-serine-linker sequences and the recognition sequences for the restriction enzymes BamHI (at the 5'-end) and KpnI (at the 3'-end) to the sequence coding for mCherry. The resulting PCR product was digested using BamHI and KpnI, purified by gel extraction and ligated into the previously generated pZMB0547 intermediate construct via the same restriction sites. The resulting plasmid was stored under number pZMB0548, a schematic representation of the sequence can be seen in Figure 43B. The plasmid pZMB0549, which contains the sequence coding for the maltose binding protein instead of the sequence coding for mCherry, was generated completely analogously. The insert sequence was amplified using the primers BamHI-MBP-for and KpnI-MBP-rev on the plasmid pZMB0401 (pMal_c5x). A schematic representation of the resulting plasmid pZMB0549 is also shown in Figure 43B and the insert DNA sequence is listed in the appendix under Chapter 10.3.2.10.

After completion of the experimental work on this thesis, this project was continued, and it turned out that it would be necessary to bind the expressed constructs covalently to cellulose using chemical coupling methods. In the context of these experiments, it was demonstrated that the binding mediated by the cellulose binding domain had a negative effect on the quality of the produced affinity chromatography material. Thus, a third generation of these constructs was designed, which differs from the second generation by the removal of the cellulose binding domain and the proline-threonine-linker sequence. Based on the plasmids pZMB0548 and pZMB0549 the following cloning work was performed. The two oligonucleotides KpnI-His6-S-XhoI-for/-rev were hybridized as described in Chapter 6.1.8 and then ligated into the plasmids pZMB0548 and pZMB0549, which were previously digested using KpnI and XhoI and purified by agarose gel extraction. The digestion removed the sequence coding for the proline-threonine-linker-, CBDcex- and His_6-tag-sequences. The hybridization product only reconstructed the His_6-tag and the stop codon in between the restriction sites KpnI and XhoI. The resulting plasmids were stored under the numbers pZMB0578 and pZMB0579 and schematics illustrations of the insert sequences are shown in Figure 43C.

7.4.2.2 **rAAV2 binding to different cellulose variants**

As already mentioned in Chapter 7.4.2, the applicability of cellulose as a base material for rAAV2 affinity chromatography was investigated prior to the start of this subproject. To be suitable as a base material, cellulose should not exhibit any unspecific binding of rAAV particles. To investigate this, two different cellulose variants were investigated with respect to their unspecific rAAV2 binding behavior. On the one hand, microcrystalline cellulose available under the trade name Avicel (Avicel PH-101, ~50 µm particle size, Sigma Aldrich) and on the other hand, regenerated amorphous cellulose (RAC) produced out of Avicel were used (see Chapter 6.4.13.1). Three batches were carried out in 2 ml reaction vessels each. 100 µl of a rAAV2 wild type sample in PBS (initial titer $7.5 \cdot 10^9$ vg·ml^{-1}) pre-cleaned by ultracentrifugation (see Chapter 6.4.4) were mixed with 100 µl cellulose suspension (10 g·l^{-1} in Milli-Q water), in this case Avicel or RAC, or Milli-Q water for control. The suspensions were incubated for 30 min at room temperature by gentle vortex mixing. Following this incubation, the samples were centrifuged for 1 min at 11000×g and 5 µl of the suspension supernatants were taken and used for the genomic rAAV titer determination by qPCR (see Chapter 6.4.5). The results of this titer determination are listed in Table 32.

Table 32: qPCR results of rAAV2 binding experiment to different cellulose variants.

Sample	**Titer in vg·ml^{-1}**	**Titer change in %**
Positive control	$3.74 \cdot 10^9$	-
Avicel supernatant	$3.78 \cdot 10^9$	+1.1
RAC supernatant	$5.17 \cdot 10^9$	+38.2

The results from Table 32 show that the incubation with the Avicel suspension resulted in an increase of the rAAV2 titer by 1.1 % and the incubation with the RAC suspension resulted in an increase of 38.2 %. Since no titer reduction of the supernatant occurred in both cases, it could be assumed that no unspecific rAAV2 binding to the cellulose materials occurred in the context of this binding experiment. The slight increase in the rAAV2 supernatant titer in the case of the Avicel sample or the higher increase in the supernatant titer in the case of the RAC sample can be explained, for example, by the exclusion volume of the cellulose sample. Both Avicel and RAC have a specific intrinsic volume, which would have led to a dilution of the rAAV sample in this experiment and therefore to deviating qPCR results. The intrinsic volume of regenerated amorphous cellulose is significantly higher than that of microcrystalline cellulose. Cellulose thus appears to be suitable as a base material for a rAAV2 affinity chromatography application.

7.4.2.3 **Expression of His_6_PKD2_CBDcex**

The first step was to express the fusion protein His_6_PKD2_CBDcex using an *E. coli* BL21(DE3) strain carrying the plasmid pZMB0505. A culture with a total volume of 200 ml was generated under standard conditions. Directly before induction with IPTG, a sample (250 µl) of the *E. coli* suspension was taken. Further samples were taken from the suspension in one-hour intervals until cultivation was completed six hours after induction. In order to check the expression of His_6_PKD2_CBDcex, an SDS-PAGE was prepared on which the *E. coli* suspension samples of this cultivation process were applied. This SDS-PAGE is shown in Figure 44A.

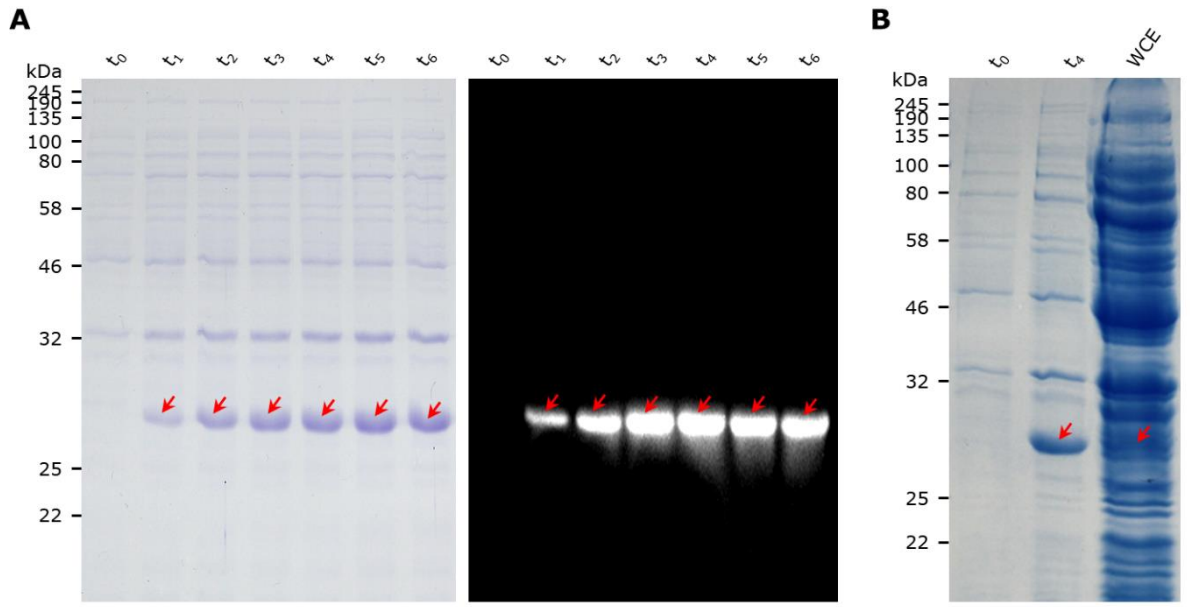


Figure 44: Expression of His_6_PKD2_CBDcex. A: SDS-PAGE and Western Blot of E. coli suspension samples taken from His_6_PKD2_CBDcex expression. Samples were taken in one-hour intervals starting right before induction (t_0) and ending six hours after induction (t_6). For the Western Blot the Tetra-His antibody in combination with the anti-mouse-HRP secondary antibody was used. B: SDS-PAGE of the expression of His_6_PKD2_CBDcex showing sample taken from the E. coli culture and the soluble phase after cell disruption (WCE).

Figure 44A shows that a protein band occurs after induction (red arrow), which is not visible before induction. This band increased in intensity over the cultivation process. The fusion protein His_6_PKD2_CBDcex has a molecular weight of 24.9 kDa (calculated from the protein sequence using Geneious software), but the protein occurring during expression ran on the SDS-PAGE clearly above the 25 kDa marker band. Size determination using ImageJ software showed that the labelled protein band had an estimated molecular weight of about 28.6 kDa. This was higher than expected. To ensure that this protein was the His_6_PKD2_CBDcex fusion protein, the SDS-PAGE was repeated using the same samples and a Western Blot was performed. The His_6_PKD2_CBDcex fusion protein has a His_6-tag and could therefore be detected by the Tetra-His antibody. Figure 44A shows the chemiluminescence signal of the corresponding Western Blot using the Tetra-His antibody in combination with the HRP-coupled anti-mouse secondary antibody (see Chapter 6.3.10). At the same height as on the SDS-PAGE, bands could be detected. As no further bands were visible, this confirms the assumption that this band was the target protein His_6_PKD2_CBDcex. A possible explanation for the abnormal running behavior could be the unusually high amount of proline (9.7 %), which is mostly caused by the proline-threonine linker but also by the high proline content of the PKD2 domain (7.1 %). So, the expression worked well. A closer look at the Western Blot signal reveals that the intensity of the bands had a maximum at time points t_3 and t_4. Thus, subsequent cultivations were stopped four hours after induction.

Next the aim was to purify the fusion protein after a larger production with a total volume of one liter. After the cultivation the cells were disrupted by French Press and the fusion protein was processed by IMAC. The IMAC purification did not deliver a useful result (data not shown). Figure 44B shows a section of an SDS-PAGE on which samples of this preparation were applied. In addition to the *E. coli* suspension samples from the cultivation at the time point t_0 (before induction) and t_4 (harvest), a sample of the clear whole cell extract (WCE), which was generated after the removal of the insoluble components by centrifugation following cell disruption, can be seen. In this case the WCE sample represents a 30-fold concentration compared to the *E. coli* suspensions sample (t_4). The *E. coli* suspension samples had a similar characteristic as in Figure 44A. However, the WCE sample indicates that the fusion protein band (red arrow) was hardly visible, despite a strong concentration. Maybe *E. coli's* own proteins cover the fusion protein up. This suggests that most of the fusion protein was removed by centrifugation together with insoluble *E. coli* components and that only a very small proportion of the produced fusion protein was present in the soluble phase. In order to solve this problem, Vivien Przybycin's bachelor thesis included experiments aimed at increasing the solubility of His_6_PKD2_CBDcex within different cultivations conditions. In addition, attempts were made to bind the fusion protein to cellulose, and it could be shown that the cellulose material was capable to bind rAAV2 particles afterwards. To increase the solubility of the fusion protein, different cultivation conditions, such as low cultivation temperatures and longer cultivation times, as well as the coexpression of chaperones (using the plasmid pGro7, TaKaRa, pZMB0528) were tested. None of the examined conditions has led to a significant improvement of the solubility of the fusion protein. Therefore, further experiments were carried out using the whole cell extract, which production is described in this chapter, as further purification attempts using IMAC failed as well.

7.4.2.4 **Binding of His_6_PKD2_CBDcex to cellulose**

Attempts were made to bind the whole cell extract containing His_6_PKD2_CBDcex to cellulose, as further purification trials had failed as described in Chapter 7.4.2.3. This could possibly omit the purification via IMAC. It was tried to bind the generated WCE (see Chapter 7.4.2.3), to Avicel and to regenerated amorphous cellulose (RAC). Both preparations had a total volume of 2 ml. Samples were taken from each individual step for an SDS-PAGE, which is shown in Figure 45.

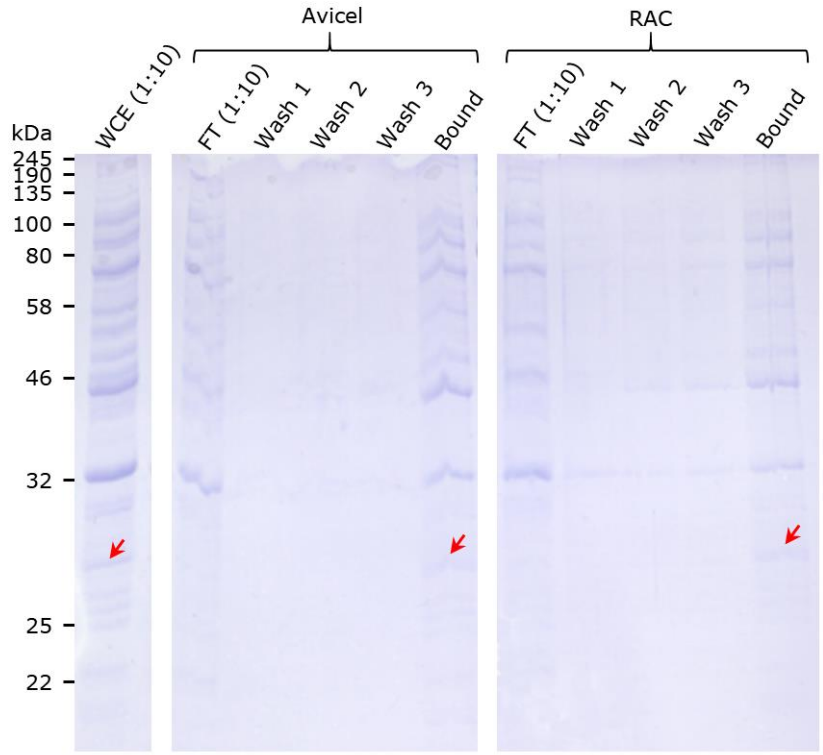


Figure 45: SDS-PAGE showing a binding experiment of His_6_PKD2_CBDcex in form of an E. coli whole cell extract (WCE), to different types of cellulose. Microcrystalline cellulose (Avicel) and regenerated amorphous cellulose (RAC) were investigated. WCE: whole cell extract, FT: flow through, Wash 1–3: wash fractions and Bound: cellulose sample with bound proteins. Red arrows pointing to the His_6_PKD2_CBDcex band.

Figure 45 shows that the His_6_PKD2_CBDcex band, which could still be seen in the dilution of the WCE sample (red arrow), no longer occurs in the flowthrough (FT) in both approaches. In the washing fractions, only individual bands of the previously probably loosely bound *E. coli* proteins could be identified. For both the Avicel sample and the RAC sample, the last lane shows that the His_6_PKD2_CBDcex fusion protein was bound. However, the band intensity of the bound fusion protein was slightly weaker than in the initial solution (this sample was diluted 1:10 for the SDS-PAGE), indicating that a larger proportion of His_6_PKD2_CBDcex proteins did not bind to the cellulose materials or was removed by the washing process. This could have been caused by a misfolding of the cellulose binding domain or the binding affinity to the selected materials was not as high as expected. The fact that no target protein could be detected in both the flowthrough fractions and the wash fractions could be explained by the respective dilutions or by the SDS-PAGE detection limit. Nevertheless, His_6_PKD2_CBDcex was bound to both cellulose materials. In addition to the target protein, several other *E. coli* proteins were non-specifically bound and could not be removed by the washing procedures. Therefore, this binding method was not suitable for the generation of a clean chromatography material. A difference between Avicel and RAC could not be discerned here, therefore only Avicel was used as the base material for further experiments, since it is easier to handle than RAC.

7.4.2.5 **Binding of rAAV2 wt to cellulose loaded with His$_6$_PKD2_CBDcex**

This chapter describes an experiment that was conducted to clarify whether the fusion protein His$_6$_PKD2_CBDcex bound to Avicel was able to bind rAAV2 wt particles. Since it was not possible to generate His$_6$_PKD2_CBDcex in pure form at this time, Avicel was loaded with *E. coli* whole cell extract, which came from a His$_6$_PKD2_CBDcex production. Samples of the chromatography material generated this way, were incubated with rAAV2 wt particles previously purified by ultracentrifugation. The binding was performed in PBS. A total of five batches were examined, in which the incubation times were varied between five and 120 minutes, in order to analyze a progression of the rAAV2 binding over time. The following figure shows the normalized amounts of rAAV2 particles that could be detected in the supernatant (unbound) or on the cellulose material (bound) by qPCR quantification depending on the incubation time.

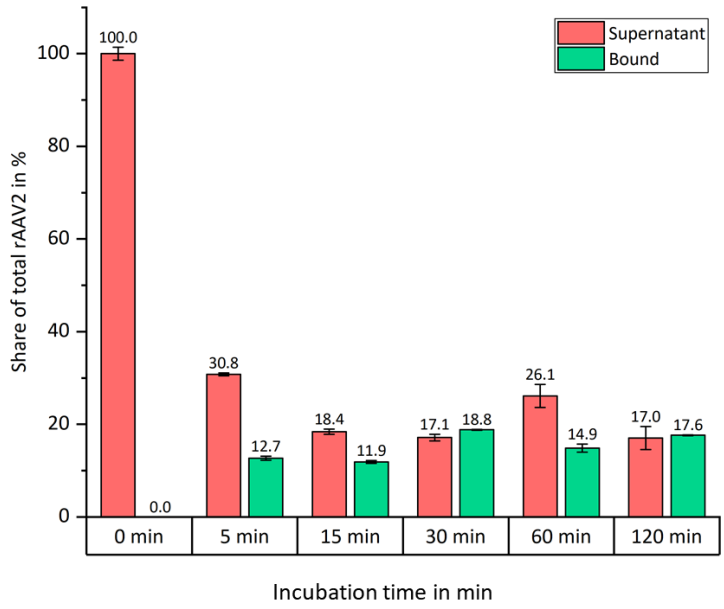


Figure 46: Normalized amounts of rAAV2 wt particles that could be detected in the supernatant (unbound) or on the cellulose chromatography material (bound) depending on the incubation time, in an binding experiment.

The quantity of rAAV2 particles initially used was set to 100% in this experiment. After a five-minute incubation the rAAV2 titer of the supernatant already decreased by 69.2 %, which shows that the generated resin was able to bind rAAV2 wt particles. A prolongation of the incubation time led to a further but relatively small decrease of the rAAV2 wt particles in the supernatant, whereby the remaining proportion fluctuated between 17.0 % and 26.1 % and showed no further trend depending on the incubation time. So, the question arose why approximately 20 % of the rAAV2 wt particles did not bind to the material. Although the progression shows a kind of saturation behavior, it could not be assumed that the material was overloaded due to the significant overdose of chromatography material

used compared to the used amount of rAAV2 particles. In publications in which AFM recordings were made of rAAV, it has already been reported that not all rAAV particles of a production are uniform in size [236]. This could be caused e.g. by an incorrect assembly or wrong folding of some capsid proteins. Since a fraction of the natural receptor was used here as a binding element, it could therefore be assumed that only (or mainly) correctly assembled and thus biologically active (infectious) rAAV2 particles were bound. This could be an advantage over conventional purification methods.

A further explanation for the unexpectedly high proportion of rAAV2 particles remaining in the supernatant could be that, under the buffer conditions used (use of PBS), some of the rAAV2 particles were already eluted or that even the His_6_PKD2_CBDcex fusion protein was only loosely bound to the cellulose material. But finally, an incubation time of about 15 minutes seemed to be sufficient to bind the maximum possible amount of rAAV2 particles from the suspension.

However, the proportion of bound rAAV2 particles did also not behave as expected. The bound fraction fluctuated independently of the incubation time (for incubation times >5 min) within a range of 11.9 % and 18.8 %, respectively at an average of approx. 15 %. On the other hand, this raised the question of where the remaining approx. 65 % of rAAV2 particles were. One possible explanation for the disappearance of these rAAV2 particles could be that through the binding of the rAAV2 particles to the PKD2 domain, a part of the viral DNA was released from the capsid. By DNaseI digestion prior to qPCR quantification this released DNA would be digested and could not be quantified. In order to investigate this theory, several different experiments were carried out, all of them did not provide clear results and are not shown here. Another possible explanation could be that the buffer conditions already elute rAAV2 particles or even release the fusion protein from the cellulose material, so they were lost during the washing process. The remaining bacterial proteins on the cellulose material could also have a negative effect on the qPCR reaction. Further experiments were carried out to investigate these possibilities.

First, the previous binding experiment was repeated in the same way, but the incubation time of the rAAV2 particles with the resin was set to 30 min. In addition to the unbound supernatant and the bound particles, the washing fraction generated after binding was also analyzed. In addition, two control approaches were generated. In the first control approach, the chromatography material consisted solely of Avicel, which was not loaded with *E. coli* whole cell extract. In this approach, no binding was to be expected. For the second control approach, an *E. coli* BL21(DE3) culture was generated that carried the plasmid pZMB0230 (pET21a empty vector) and was cultured as described in Chapter 6.3.2 as if a protein should be expressed (which was not the case due to the empty vector). Finally, as described in Chapter 6.3.3, an *E. coli* whole cell extract was generated from this culture by French Press cell disruption. This WCE, was used to load Avicel as described in

Chapter 6.4.13.2. The influence of the bound *E. coli* proteins on this setup should be investigated. The normalized qPCR data are shown in Figure 47A.

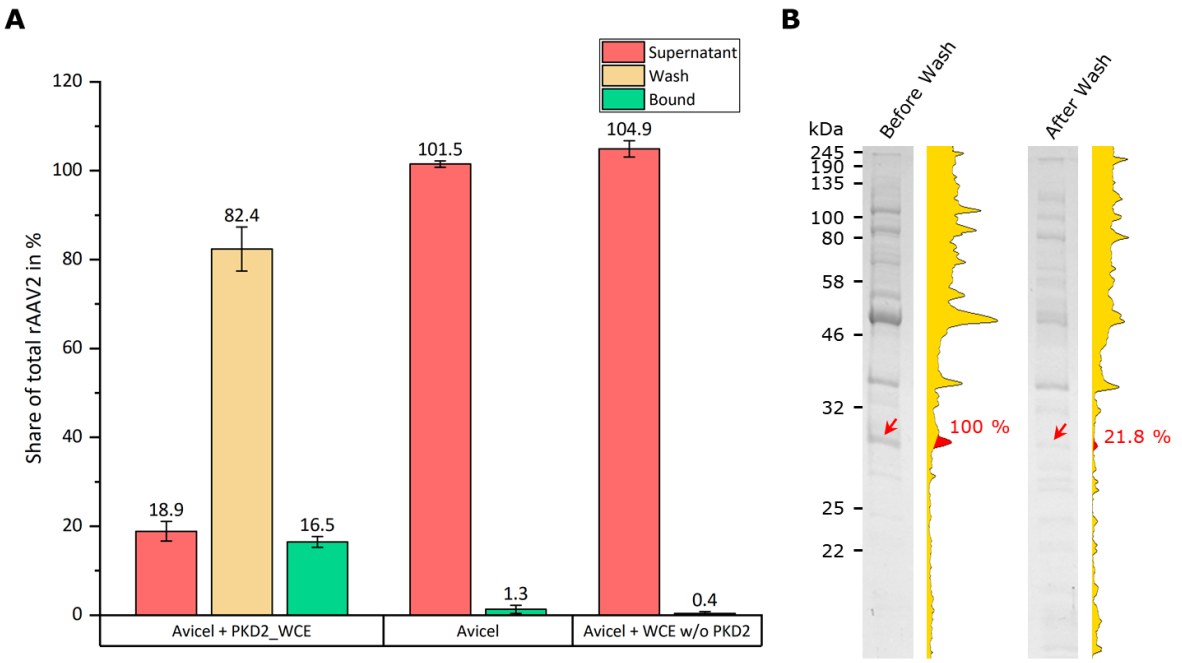


Figure 47: rAAV2 wt binding experiment to cellulose based affinity chromatography material. A: Normalized qPCR date from a control experiment with the standard setup, an untreated Avicel sample and Avicel loaded with E. coli WCE from a culture not expressing any target protein. B: SDS-PAGE lanes showing samples of Avicel loaded with His_6_PKD2_CBDcex containing E. coli WCE before and after a washing procedure. Band intensity was analyzed using ImageJ Gel analysis and is shown as a yellow histogram next to the lanes. His_6_PKD2_CBDcex band is marked in red.

The control approaches presented in Figure 47A both shows the expected results. When using unloaded Avicel as well as Avicel which was previously loaded with *E. coli* WCE, which does not contain His_6_PKD2_CBDcex fusion protein, virtually the entire amount of rAAV2 wt particles could be detected in the supernatants (101.5 % and 104.9 %), respectively. With 1.3 % and 0.4 %, of the initial used particles, almost no noteworthy binding could be substantiated to the materials. The small deviations could probably be attributed to the dilutions or pipetting inaccuracies that occurred. The repetition of the original initial experiment also showed comparable values for the titer of the unbound supernatant fraction (18.9 %) and the bound fraction (16.5 %) as shown in Figure 46. The crucial factor here was the proportion of rAAV2 wt particles occurring in the wash fraction. With 82.4 %, this was even higher than the expected maximum proportion of 64.6 %. This clarifies the question of where the remaining rAAV2 particles were, and the theory of the viral DNA release by the capsid binding to the PKD2 domain loses relevance. Similar observations were not described by other working groups until the completion of this thesis. The fact that the proportion of rAAV2 particles in the wash fraction was unexpectedly high may be due to the 1:10 dilution of the suspension caused by the washing process and the

associated higher probability of error in the qPCR quantification, since the values determined were closer to the detection limit.

Finally, it had to be investigated whether the rAAV2 particles were eluted from the resin or if there was even an elution of the His_6_PKD2_CBDcex fusion proteins. To answer this question, a washing experiment was performed in which a sample of (washed) Avicel was loaded with His_6_PKD2_CBDcex containing *E. coli* WCE. Subsequently, this sample was incubated in PBS on a roller tumble mixer for about 30 min, as it would happen during the binding process of rAAV2 particles. The supernatant was then removed by centrifugation and the sample was washed with the ten-fold volume of PBS before being adjusted to its initial concentration (10 $g \cdot l^{-1}$ loaded Avicel in PBS). The originally loaded sample and the sample which was created after this washing process were applied to an SDS-PAGE. The result of this process can be seen in Figure 47B.

When looking at the figure, it is obvious that the band intensity of the washed sample is significantly lower than those of the initial samples. This also illustrates the ImageJ gel analysis histogram. It shows for the His_6_PKD2_CBDcex band that the intensity of the washed sample was only 21.8 % of the original intensity (red mark). This corresponds to about one fifth of the initial quantity. Considering the qPCR data from Figure 47A, where the washed off rAAV2 fraction is about five times as high as the remaining fraction on the resin, this supports the thesis that although the rAAV2 particles were bound by the chromatography material, a substantial fraction (about 4/5) was washed off together with the fusion protein His_6_PKD2_CBDcex. Thus, the washing conditions of the investigated setup were not suitable or the binding of the fusion protein to the Avicel material mediated by the cellulose binding domain was not strong enough.

7.4.2.6 **Production of the second generation of PKD2 based fusion proteins**

As mentioned above, the second generation of fusion proteins for the PKD2-based affinity chromatography has already been cloned during the work described in the previous chapter. The aim was to increase the solubility of the fusion proteins in order to improve the yield by the expression using *E. coli*. As the results presented in the last section, which indicate that the CBDcex probably mediates a too weak binding of the fusion protein to the Avicel material, were not known at this time, this was not considered for the second generation of fusion proteins. After completion of the cloning work, *E. coli* BL21(DE3) strains carrying the plasmids pZMB0547, pZMB0548 and pZMB0549 were cultured in 200 ml scale and the fusion proteins were expressed. Following the expression (four hours after induction), the cells were harvested by centrifugation and, 10 ml PBS buffer were used per gram of wet cell mass to produce cell suspensions in order to be able to compare the productions. After cell disruption by French Press, insoluble components of the

respective whole cell extracts were removed by centrifugation. A comparative SDS-PAGE, on which samples taken during the expressions and samples of the clear WCEs, of the three productions and the previous production based on the plasmid pZMB0505, were applied, is shown in the figure below.

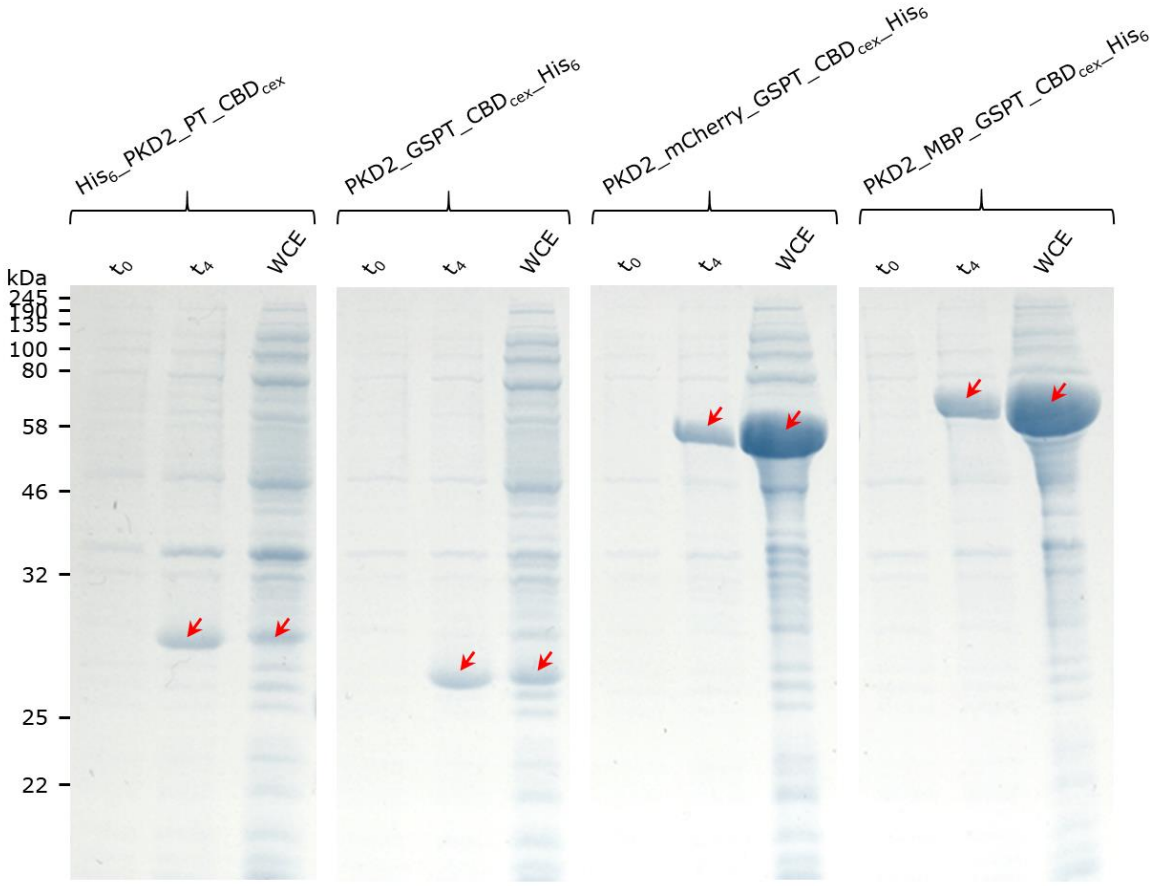


Figure 48: Comparative SDS-PAGE of four fusion protein expressions with E. coli. The expressions of the fusion proteins His_6_PKD2_PT_CBDcex (pZMB0505), PKD2_GSPT_CBDcex_His_6 (pZMB0547), PKD2_mCherry_GSPT_CBDcex_His_6 (pZMB0548) and PKD2_MBP_GSPT_CBDcex_His_6 (pZMB0549) are shown. t_0: E. coli suspensions sample directly before induction; t_4: E. coli suspension sample directly before cell harvest (four hours after induction); WCE: clear whole cell extract after cell disruption. The bands of the respective target proteins are marked with red arrows.

First the results of the constructs His_6_PKD2_PT_CBDcex and PKD2_GSPT_CBDcex_His_6 are compared with each other. Although PKD2_GSPT_CBDcex_His_6 with 24.4 kDa is only 0.4 kDa smaller than His_6_PKD2_PT_CBDcex, a comparison of the produced fusion proteins shows a clearer mass shift. According to ImageJ image analysis, PKD2_GSPT_CBDcex_His_6 has a calculated molecular weight of 27.2 kDa. The main difference at protein level was the shortening of the proline-threonine-linker. As already suspected in Chapter 7.4.2.3, the unusually high proportion of prolines led to this changed running behavior on the SDS-PAGE. Since this mass shift was smaller by shortening the corresponding part, this theory could be regarded as confirmed. As expected, the solubility of the PKD2_GSPT_CBDcex_His_6 construct did not show much difference compared to the

corresponding pre-construct. The product band in the soluble WCE was also lower than the corresponding product band in the *E. coli* suspension sample, despite approx. 30-fold concentration. Thus, the largest proportion of target proteins should be insoluble. On the other hand, the WCE samples of the other two approaches showed the desired result. Both the 51.7 kDa sized target protein PKD2_mCherry_GSPT_CBDcex_His_6 and the 65.4 kDa sized target protein PKD2_MBP_GSPT_CBDcex_His_6 showed significantly thicker product bands in the WCE samples than the other two constructs. This suggested that in this approaches a much larger proportion of the target proteins was soluble than in the construct of the first generation. The product bands were in fact so large that they represent the main protein content in the cell extract samples. The WCE sample of the PKD2_mCherry_GSPT_CBDcex_His_6 production also shows a clear red color. Based on these expressions, it should now be possible to perform IMAC purification with the aim of generating clean starting material for a possible PKD2-based affinity chromatography. This was done for the construct PKD2_MBP_GSPT_CBDcex_His_6 in the context of a production with a total volume of 2 l. After French press the fusion protein was purified by IMAC using a bed volume of 2 ml (see Chapter 6.3.5). Important samples of the expression and purification were applied to an SDS-PAGE, which is shown below.

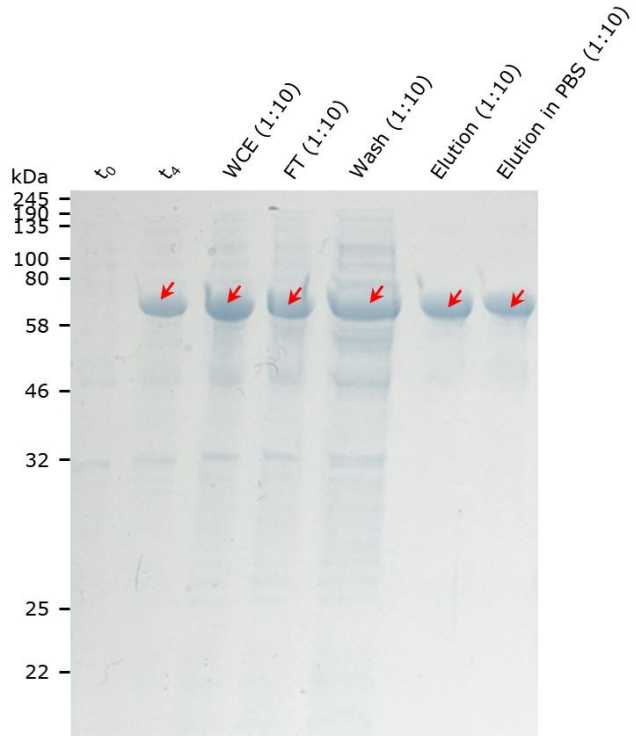


Figure 49: SDS-PAGE of samples from the PKD2_MBP_GSPT_CBDcex_His_6 expression and IMAC purification. t_0: E. coli suspensions sample directly before induction; t_4: E. coli suspension sample directly before cell harvest (four hours after induction); WCE: clear whole cell extract after cell disruption; FT: IMAC flow through; Wash: IMAC wash fraction; Elution: IMAC elution fraction; Elution in PBS: rebuffered sample of the elution fraction in PBS using an Amicon 10000 Da Spin Filter unit. The bands of the PKD2_MBP_GSPT_CBDcex_His_6 target protein are marked with red arrows.

The SDS-PAGE shown in Figure 49 illustrates that both production and IMAC purification had worked well. Only the fact that a large proportion of the target protein could still be found in the flowthrough and the wash fractions, showed that the selected IMAC bed volume was too small. However, both the elution sample and the elution sample re-buffered in PBS showed a clean target protein band with only few impurities. A total of 3.3 ml re-buffered protein solution with a concentration of 3.0 $g \cdot l^{-1}$ was produced. To enhance the total yield, the flowthrough could be reprocessed again. The high yield and great cleanliness of the sample should allow a wide range of applications.

Since my employment contract expired at this point of the project, Kathrin Teschner continued working on it. She further optimized the covalent binding of the PKD2_MBP_GSPT_CBDcex_His_6 protein to cellulose and developed a purification procedure based on the resulting affinity chromatography material.

7.4.2.7 **Binding ability of PKD2_MBP_GSPT_CBDcex_His_6-based cellulose affinity resin for different rAAV2 variants**

After Kathrin Teschner had further optimized the purification method, the potential of this method will be demonstrated here once again. In addition to the rAAV2 wild type variant, linker and β-lactamase variants with different degrees of modification should also be investigated.

For this purpose, a total of five different rAAV2 variants were produced, each production took place at a scale of twelve 100 mm dishes and was carried out according to the method described in Chapter 6.4.1. The following variants were selected:

rAAV2 wt (produced using the plasmids pZMB0216, pZMB0522 and pZMB0088) served as reference. rAAV2 587_4x(GGSG) mosaic (produced using plasmids pZMB0601, pZMB0600, pZMB0522 and pZMB0088) and rAAV2 587_4x(GGSG) full (produced using plasmids pZMB0220, pZMB0522 and pZMB0088) served as partially or fully modified representatives with flexible glycine-serine linker insertions at position 587. rAAV2 587_bla mosaic (produced using the plasmids pZMB0577, pZMB0600, pZMB0522 and pZMB0088) and rAAV2 587_bla full (produced using the plasmids pZMB0221, pZMB0522 and pZMB0088) served as partially or fully modified representatives with β-lactamase insertions at position 587. These five rAAV2 variants cover a broad spectrum of rAAV2 with different degrees of modifications. Following production, the samples were purified by ultracentrifugation and finally stored in PBS.

After production of PKD2_MBP_GSPT_CBDcex_His_6 (see Chapter 7.4.2.6), the fusion protein was covalently bound to cellulose (blotting paper) using divinyl sulfone as described in Chapter 6.4.13.4. For the subsequent purification, which was carried out according to the method described in Chapter 6.4.13.5, 500 µl of each of the five rAAV variants in PBS

were used together with one round piece of blotting paper (with a diameter of 10 mm). Subsequently, DNaseI digestions were prepared from all samples (starting solutions, flowthrough, combined wash fractions and the rebuffered eluate fractions), which were subsequently tittered by qPCR. The results are shown in the following figure.

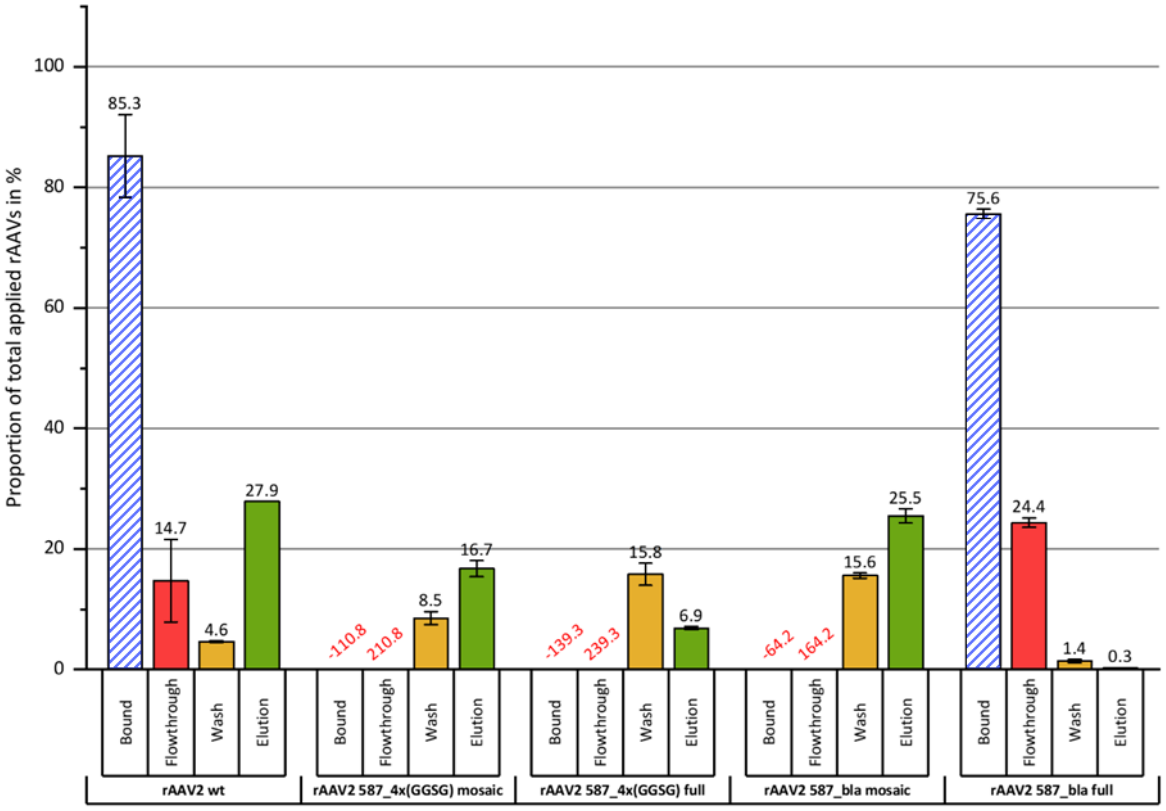


Figure 50: Normalized qPCR results for the "PKD2_MBP_GSPT_CBDcex_His$_6$-based cellulose affinity resin" purification test using different rAAV2 variants. rAAV2 wt (rAAV2 wild type variant); rAAV2 587_4x(GGSG) mosaic (rAAV2 with a 16 amino acid long insertion (4x(GGSG)) at position 587 in all VP2 capsid proteins); rAAV2 587_4x(GGSG) full ((rAAV2 with a 16 amino acid long insertion (4x(GGSG)) at position 587 in all capsid proteins); rAAV2 587_bla mosaic (rAAV2 with an approx. 29 kDa insertion of a β-lactamase at position 587 in all VP2 capsid proteins); rAAV2 587_bla full (rAAV2 with an approx. 29 kDa insertion of a β-lactamase at position 587 in all capsid proteins); The columns indicate the relative proportions of all DNaseI resistant rAAV particles detected by qPCR in the relevant fractions. Bonded to the chromatography material (blue, calculated using the total used amount and the flowthrough), flowthrough (red), wash fraction (yellow) and elution (green). Data based only on calculations are displayed with hatches.

Figure 50 shows that the purification for the rAAV2 wild type variant, for which the system was designed, worked best. From the initial amount of rAAV wt used, only 14.7 ±6.9 % could be found in the flowthrough, which means that 85.3 ±6.9 % have bound to the resin. The wash fraction contained only 4.6 % of rAAV2, which could have been caused by loose rAAV2 particles or by rAAV remaining in the supernatant, which were absorbed by the sponge-like structure of the paper sheet. About one third of the bound rAAV2 particles could ultimately be found in the eluate. This represented 27.9 ±0.0 % of the initial particles. This value is quite satisfying considering the fact that the method carried out in

this way was very quick and required only cheap materials. However, where about two thirds of the previously bound particle end up is not easy to clarify. Two possible assumptions were that the particles were not detached from the material by the elution condition applied or that e.g. some of the eluted particles were lost within buffering using Amicon spin filters prior to quantification. Such buffering often leads to the adsorption of rAAV particles onto the filter material. For the modified variants it was noticeable that purification seemed to be somewhat more difficult. For the fully modified β-lactamase variant, 75.6 ±0.8% of the rAAV initially used were bound by the resin. Only 1.4 ±0.3 % of the particles were detectable in the wash fraction and only 0.3 ±0.0 % could be found in the eluate. This shows that despite the strong degree of modification, a large proportion of the particles were accessible to the fusion protein. However, it remains unclear why only such a small proportion of rAAV could be found in the eluate. It is unlikely that the insertion of β-lactamase would in any sense increase the binding of the PKD2 domain to the capsid proteins, so that the chosen elution conditions would not be sufficient to release the rAAV particles from the resin. However, it is more likely that the rAAV2 587_bla full particle were destroyed during elution and released their DNA, so that it was digested by the DNaseI treatment and thus were not quantifiable. In the case of the other three modified variants (rAAV2 587_4x(GGSG) mosaic, rAAV2 587_4x(GGSG) full and rAAV2 587_bla mosaic), unexpectedly high genomic rAAV titers occurred for the titrations of the flowthroughs, which ranged between 64.2 ±10.0 % and 139.3 ±68.8 % above the respective initial titers. Thus, the calculation of the resin-bound rAAV proportions resulted in negative values (bound and flow-through columns were removed in Figure 50 for a better overview). One possible explanation could be that despite prior ultracentrifugation purification, impurities were present in the sample that had an inhibitory effect on the qPCR reaction and that these impurities bound to the resin or to the cellulose, which aided the quantification of the particles remaining in the flowthrough. If this is true, all initial titers may have been underestimated, as all samples were treated equally. Also, the estimation of the bound rAAV proportions for the wild type and the completely modified beta-lactamase variants would then be incorrect. Nevertheless, for all three variants between 6.9 ±0.3 % and 25.5 ±1.2 % of the rAAV could be detected in the eluate. The weaker modified mosaic variants showed slightly higher titers (rAAV2 587_4x(GGSG) mosaic 16.7 ±1.3 % and rAAV2 587_bla mosaic 25.5 ±1.2 %) than the completely modified amino acid linker variant (rAAV2 587_4x(GGSG) full 6.9 ±0.3 %). These yields are very useful in the context of such a simple, cost effective and fast purification procedure and indicate that this procedure could be used for the purification of rAAV2 wild type and less modified rAAV2 mosaic variants.

Within the optimization experiments carried out by Kathrin Teschner, it turned out that under the elution conditions used, a portion of the previously bound fusion proteins were found in the eluate of the purification process. It was assumed that this eluted fusion

protein was not covalently bound to the cellulose material via divinyl sulfone, but only via the cellulose binding domain, which was probably not able to maintain the binding under the used elution conditions. In order to avoid this, the third generation of fusion proteins was developed which, in contrast to the used second generation, do not contain a cellulose binding domain. Further experiments on these constructs are still pending at this point of time.

7.4.3 **Thermo Fisher Scientific Cooperation - Binding ability of POROS AAVX Affinity Resin for different rAAV2 variants**

The presentation of a poster of mine at the DECHEMA Conference in Frankfurt (January 30 - 31, 2019) on the topic “Gene Therapy - Ready for the Market?” (see Chapter 10.2.1) drew the attention of the Director of Ligand Application at ThermoFisher Scientific Dr. Frank Detmers to my work. This resulted in a cooperation in which I was to test the “POROS CaptureSelect AAVX Affinity Resin”, which was developed for the large-scale downstream purification of a broad range of naturally occurring and synthetic AAV serotypes, using some of the rAAV2 variants generated in this thesis.

For this purpose, a total of five different rAAV2 variants were selected: rAAV2 wt served as reference. rAAV2 587_4x(GGSG) mosaic and rAAV2 587_4x(GGSG) full served as partially or fully modified representatives with flexible glycine-serine linker insertions at position 587. rAAV2 587_bla mosaic and rAAV2 587_bla full served as partially or fully modified representatives with β-lactamase insertions at position 587. These variants were the same samples that were used in the purification experiment presented in Chapter 7.4.2.7, so both experiments are absolutely comparable.

The following purification experiments were carried out according to the protocol provided by Thermo Fisher Scientific using 50 µl “POROS CaptureSelect AAVX Affinity Resin” each in 1 ml scale as batches (see Chapter 6.4.14). Afterwards a DNaseI digestion was performed with all samples and the titers were determined by qPCR. The results are shown in the following figure.

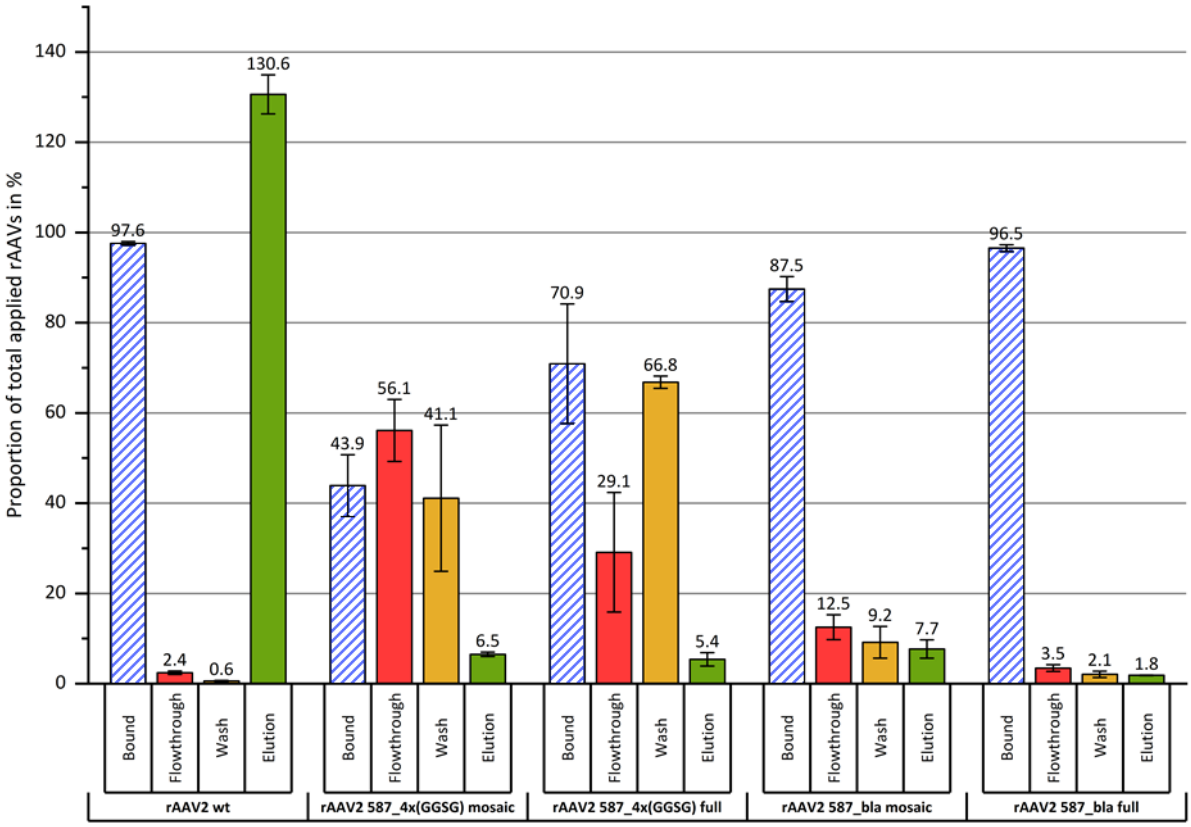


Figure 51: Normalized qPCR results for the "POROS CaptureSelect AAVX Affinity Resin" purification test using different rAAV2 variants. rAAV2 wt (rAAV2 wild type variant); rAAV2 587_4x(GGSG) mosaic (rAAV2 with a 16 amino acid long insertion (4x(GGSG)) at position 587 in all VP2 capsid proteins); rAAV2 587_4x(GGSG) full ((rAAV2 with a 16 amino acid long insertion (4x(GGSG)) at position 587 in all capsid proteins); rAAV2 587_bla mosaic (rAAV2 with an approx. 29 kDa insertion of a β-lactamase at position 587 in all VP2 capsid proteins); rAAV2 587_bla full (rAAV2 with an approx. 29 kDa insertion of a β-lactamase at position 587 in all capsid proteins); The columns indicate the relative proportions of all DNaseI resistant rAAV particles detected by qPCR in the relevant fractions. Bonded to the chromatography material (blue, calculated using the total used amount and the flowthrough), flowthrough (red), wash fraction (yellow) and elution (green). Data based only on calculations are displayed with hatches.

In Figure 51 it can be seen that the purification of the rAAV2 wt reference worked very well. A total of 97.6 ±0.4 % of the rAAVs used at the beginning were bound to the resin. Only 0.6 ±0.1 % of rAAV2 wt was found in the wash fraction. The binding of the rAAV2 particles to the chromatography material seems to be very stable under the tested conditions. In the pooled elution fraction, 130.6 ±4.3 % of the initially used rAAV2 wt were present. This value was not realistic and unexpected high. However, this effect had already been observed in comparable procedures with subsequent qPCR quantification. One possible explanation for this is that the elution fraction was not present in PBS but in a different buffer compared to the other samples (after neutralization about 75 mM sodium citrate, 250 mM tris, pH 7.0 to pH 7.5). These buffer differences could result in an influence on the very sensitive qPCR reaction. A further explanation for the unusually high value of the elution fraction could be the fact that the rAAV2 samples were cleaner in the elution

fraction than in the starting samples. Impurities in the starting samples could have an inhibitory effect on the qPCR reaction and thus led to reduced values in the titration of the starting samples, the flowthrough and the washing fractions. In the end, the rAAV2 wt reference fulfilled its purpose and confirmed the functionality of the material.

The purification of the four modified variants did not work sufficiently in any case. However, the two glycine-serine linker insertion variants showed a different behavior compared to the β-lactamase insertion variants. While both glycine-serine linker variants showed a clearly reduced affinity to the resin compared to the wild type (binding to the chromatography material: mosaic: 43.9 ±6.9 % and full: 70.9 ±13.3 %), the binding ability was only slightly influenced by the β-lactamase insertions (binding to the chromatography material: mosaic: 87.5 ±2.8 % and full: 96.5 ±0.7 %). This result was rather unexpected, since compared to the only 16 amino acids long linker variants, the β-lactamase variants present five or 60 complete approx. 29 kDa large enzymes on their capsid surfaces. Especially in the case of the completely modified β-lactamase variant, a significant part of the capsid surface is covered by the enzymes presented (see also Figure 21). Furthermore, it was noticeable that for both modification variants, the less-modified mosaic variants bonded weaker to the resin compared to their completely modified partner variants.

The binding of the glycine-serine linker variants to the resin appears to be so weak that it was almost totally destroyed by washing using PBS and a large part of the previously “loosely” bound rAAV2 was found in the pooled wash fraction (mosaic: 41.1 ±16.2 % and full: 66.8 ±1.4 %). Such an effect was not observed when using the β-lactamase variants, only 9.2 ±3.5 % (mosaic) or 2.1 ±0.7 % (full) of the rAAVs could be detected in the wash fractions.

The elution proportions of all modified variants in this test series were below 10 % in each case, which confirms that the POROS CaptureSelect AAVX Affinity Resin is not suitable for the downstream processing of these rAAV2 variants. In the case of the glycine-serine linker variants, these low elution values could be explained by the fact that most rAAVs were already eluted during the previous washing steps. However, the β-lactamase variants could not be eluted despite good binding to the resin. The presumption that, the binding of the β-lactamase rAAV2s to the resin was stronger than these of the wild type, and thus could not be detached by the usual elution conditions, seems to be an unrealistic explanation of this observation, considering the normally high specificity of affinity chromatography materials. One possible explanation for this result could be, that the rAAV2 587_bla variants become unstable due to the elution at acid pH and releases their DNA, which would, after routinely DNaseI digestion prior to qPCR quantification, lead to no (or few) rAAV2 particles being detected.

At this point it makes sense to draw a comparison between the "POROS CaptureSelect AAVX Affinity Resin" presented here and the affinity chromatography material based on PKD2, coupled to blotting paper discussed in the last chapter. Both test series were carried out under comparable conditions and were based on the same rAAV2 starting materials. Both materials have been developed for the purification of rAAV2 (the POROS material also binds a variety of other serotypes). The binding and subsequent elution of rAAV2 wild type particles works with the POROS material with nearly maximum binding and yield, whereas the self-produced affinity chromatography material could still be improved, especially at elution, since only about a quarter of the initially used particles could be found in the eluate. But the situation is different for the modified rAAV2 particles, where hardly any rAAV2 could be found in the eluates after purification using the POROS material. In the case of the self-made resin, at least for the rAAV2 mosaic variants, useful results were achieved. In view of the clear price differences between the two materials (POROS: 2655 € per 25 ml (September 2019), self-produced resin only a few Euro per mg fusion protein) and the applicability for rAAV2 mosaic viruses, the self-made affinity chromatography material is quite competitive.

8 Conclusions and Outlook

At this point a critical conclusion shall be drawn on the projects of the last years, which are described in this thesis.

The main topic of this work was the improvement of recombinant AAV2 production, which was handled in several subprojects. The beginning was characterized by the establishment of all necessary standard methods which are required for the production, processing and analysis of rAAV particles. Since this topic had not yet found any application in the laboratories of the University of Bielefeld at this time, this familiarization phase was characterized by numerous drawbacks. However, through the collaboration of all members of the working group, the rAAV production system was continuously improved and expanded over the last few years. Nevertheless, from the very beginning, it had been assumed that this plasmid-based production system could not compete with other systems in terms of viral particle yields per production cell. Since most publications lack information on the used production scales and finally achieved yields, this assumption could never be confirmed. It would have been advisable to procure additional plasmid systems early on in these projects in order to compare them more closely with the plasmid systems used here. Even an increase in rAAV production by ten or more at constant production scale would have resulted in more raw material being available for all subsequent experiments and less time being needed to invest in the time-consuming manufacturing and purification of the rAAV.

As part of the improvement and expansion of the plasmid system, a stability assay was developed which, unlike previous methods, aimed at the biologically relevant release of rAAV DNA under increasing temperature exposure. Interesting insights were gained into the relationship between the type and size of a capsid modification and the thermal stability of the respective rAAV particles. In addition, a rAAV2 variant was generated that presents 60 active β-lactamase enzymes on its capsid surface. Although this rAAV2 variant has no biological relevance, it was shown that the established production system allows such strong modifications. A manuscript on this topic was submitted in shared authorship with Rebecca Feiner to the “International Journal of Molecular Sciences”.

A subproject that was worked on over a long period of time was the generation of rAAV production cell lines. The aim was to introduce the required adenoviral E1a/b genes into CHO K1 cells using a CRISPR/Cas9 mediated rational approach. For this purpose, on the basis of RNA sequencing data, suitable genomic target sites were selected, which should first be further characterized. All attempts regarding these characterizations failed due to random insertions of the generated GFP constructs. In addition, it was shown that a permanent insertion of the E1a/b genes into CHO K1 cells had lethal consequences, which is why further experiments of this kind were not continued. Nevertheless, it could be shown

that rAAV production using CHO K1 cells is possible. Later, monoclonal HEK293 cell lines were generated by random integration, which carry the sequences, present on the pHelper plasmid, which are necessary for rAAV production. This project was continued by Kathrin Teschner and Rebecca Feiner who were able to further characterize the generated cell lines and finally demonstrate that the genomic integration of the helper functions allows to dispense the addition of the pHelper plasmids during rAAV2 production, thus saving a part of the costs. This project also ended with the submission of a manuscript to the journal BMC Biotechnology, with my co-authorship.

In addition, the production of rAAV2 should be increased by improving downstream processing. To this end, single-chain constructs based on the A20 antibody were designed which, should form the basis for an affinity chromatography. A construct was designed that could be expressed in *E. coli* and exported to the periplasm. The insertion of a DNA sequence for a Silica tag opened up the possibility of binding the scFv to silica material. The binding ability of this construct for rAAV2 particles could be confirmed, but it turned out that silica is not suitable as a base material for rAAV purification due to its non-specific binding properties. In addition, production volumes were insufficient to produce useful affinity columns. Another construct, which was intended for eukaryotic expression and which contained a human IgG-Fc part, could later be produced using HEK293 suspensions cells and has since been used for AAV capsid ELISAs. In the later phase of the thesis another fusion protein was developed which used the PKD2 domain of AAVR in combination with a cellulose binding domain to process rAAV2 particles. The functionality of this construct was demonstrated and the initially insufficient solubility during expression was significantly increased by the fusion with a maltose binding protein. This construct formed the basis for the further development of an affinity chromatography material which was carried out by Kathrin Teschner. In a final series of assays, it was shown that this resin is suitable for the purification of rAAV2 wild-type particles as well as modified mosaic rAAV2. This resin therefore has advantages over the commercially available POROS AAVX affinity resin. However, further optimizations are conceivable in the future.

9 References

1. Atchison RW, Casto BC, Hammon WM. Adenovirus-Associated Defective Virus Particles. Science (80-). 1965;149:754–5.

2. Srivastava A, Lusby EW, Berns KI. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J Virol. 1983;45:555–64.

3. Mendelson E, Trempe JP, Carter BJ. Identification of the trans-acting Rep proteins of adeno-associated virus by antibodies to a synthetic oligopeptide. J Virol. 1986;60:823–32.

4. Labow MA, Hermonat PL, Berns KI. Positive and negative autoregulation of the adeno-associated virus type 2 genome. J Virol. 1986;60:251–8.

5. Beaton a, Palumbo P, Berns KI. Expression from the adeno-associated virus p5 and p19 promoters is negatively regulated in trans by the rep protein. J Virol. 1989;63:4450–4.

6. Trempe JP, Carter BJ. Alternate mRNA splicing is required for synthesis of adeno-associated virus VP1 capsid protein. J Virol. 1988;62:3356–63.

7. Xie Q, Bu W, Bhatia S, Hare J, Somasundaram T, Azzi A, et al. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. Proc Natl Acad Sci U S A. 2002;99:10405–10.

8. Sonntag F, Schmidt K, Kleinschmidt JA. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. Proc Natl Acad Sci U S A. 2010;107:10220–5.

9. Sonntag F, Köther K, Schmidt K, Weghofer M, Raupp C, Nieto K, et al. The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes. J Virol. 2011;85:12686–97.

10. Maurer AC, Pacouret S, Cepeda Diaz AK, Blake J, Andres-Mateos E, Vandenberghe LH. The Assembly-Activating Protein Promotes Stability and Interactions between AAV's Viral Proteins to Nucleate Capsid Assembly. Cell Rep. 2018;23:1817–30.

11. Cao M, You H, Hermonat PL. The X gene of adeno-associated virus 2 (AAV2) is involved in viral DNA replication. PLoS One. 2014;9:e104596.

12. Stutika C, Gogol-Döring A, Botschen L, Mietzsch M, Weger S, Feldkamp M, et al. A Comprehensive RNA Sequencing Analysis of the Adeno-Associated Virus (AAV) Type 2 Transcriptome Reveals Novel AAV Transcripts, Splice Variants, and Derived Proteins. J Virol. 2016;90:1278–89.

13. Gao G, Vandenberghe LH, Alvira MR, Lu Y, Calcedo R, Zhou X, et al. Clades of Adeno-

Associated Viruses Are Widely Disseminated in Human Tissues. 2004;78:6381–8.

14. Manno CS, Chew AJ, Hutchison S, Larson PJ, Herzog RW, Arruda VR, et al. AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B. Blood. 2003;101:2963–72.

15. Bartlett JS, Samulski RJ, Mccown TJ, Al BET. Selective and rapid uptake of adeno-associated virus type 2 in brain. Hum Gene Ther. 1998;20:1181–6.

16. Richter M, Iwata A, Nyhuis J, Nitta Y, Miller AD, Halbert CL, et al. Adeno-associated virus vector transduction of vascular smooth muscle cells in vivo. Physiol Genomics. 2000;2:117–27.

17. Koeberl DD, Alexander IE, Halbert CL, Russell DW, Miller AD. Persistent expression of human clotting factor IX from mouse liver after intravenous injection of adeno-associated virus vectors. Proc Natl Acad Sci. 1997;94:1426–31.

18. Merkel SF, Andrews AM, Lutton EM, Mu D, Hudry E, Hyman BT, et al. Trafficking of AAV vectors across a model of the blood-brain barrier; a comparative study of transcytosis and transduction using primary human brain endothelial cells. J Neurochem. 2016;:216–30.

19. Summerford C, Samulski RJ. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. J Virol. 1998;72:1438–45.

20. Kern A, Schmidt K, Leder C, Müller OJ, Wobus CE, Bettinger K, et al. Identification of a heparin-binding motif on adeno-associated virus type 2 capsids. J Virol. 2003;77:11072–81.

21. Opie S, Jr KW, Agbandje- M. Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding. J Virol. 2003;77:6995–7006.

22. Perabo L, Goldnau D, White K, Endell J, Boucas J, Humme S, et al. Heparan sulfate proteoglycan binding properties of adeno-associated virus retargeting mutants and consequences for their in vivo tropism. J Virol. 2006;80:7265–9.

23. Zhou S, Hansen J, Srivastava A, Mah C, Dwarki V, Qing K. Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. Nat Med. 2002;5:71–7.

24. Iwabuchi K, Daida H, Matsumoto K, Oshimi K, Watanabe M, Kashiwakura Y, et al. Hepatocyte Growth Factor Receptor Is a Coreceptor for Adeno-Associated Virus Type 2 Infection. J Virol. 2004;79:609–14.

25. Summerford C, Bartlett JS, Samulski RJ. αVβ5 integrin: a co-receptor for adeno-

associated virus type 2 infection. Nat Med. 1999;5:78–82.

26. Asokan A, Hamra JB, Govindasamy L, Agbandje-McKenna M, Samulski RJ. Adeno-Associated Virus Type 2 Contains an Integrin α5β1 Binding Domain Essential for Viral Cell Entry. J Virol. 2006;80:8961–9.

27. Akache B, Grimm D, Pandey K, Yant SR, Xu H, Kay MA. The 37/67-Kilodalton Laminin Receptor Is a Receptor for Adeno-Associated Virus Serotypes 8, 2, 3, and 9. J Virol. 2006;80:9831–6.

28. Pillay S, Meyer NL, Puschnik AS, Davulcu O, Diep J, Ishikawa Y, et al. An essential receptor for adeno-associated virus infection. Nature. 2016;530:108–12.

29. Poon MW, Tsang WH, Chan SO, Li HM, Ng HK, Waye MMY. Dyslexia-associated Kiaa0319-Like protein interacts with axon guidance Receptor Nogo Receptor 1. Cell Mol Neurobiol. 2011;31:27–35.

30. Poelmans G, Buitelaar JK, Pauls DL, Franke B. A theoretical molecular network for dyslexia: Integrating available genetic findings. Mol Psychiatry. 2011;16:365–82.

31. Bartlett JS, Wilcher R, Samulski RJ. Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors. J Virol. 2000;74:2777–85.

32. Georg-Fries B, Biederlack S, Wolf J, Zur Hausen H. Analysis of proteins, helper dependence, and seroepidemiology of a new human parvovirus. Virology. 1984;134:64–71.

33. Parks WP, Melnick JL, Rongey R, Mayor HD. Physical assay and growth cycle studies of a defective adeno-satellite virus. J Virol. 1967;1:171–80.

34. Schlehofer JR, Ehrbar M, zur Hausen H. Vaccinia virus, herpes simplex virus, and carcinogens induce DNA amplification in a human cell line and support replication of a helpervirus dependent parvovirus. Virology. 1986;152:110–7.

35. Ogston P, Raj K, Beard P. Productive Replication of Adeno-Associated Virus Can Occur in Human Papillomavirus Type 16 (HPV-16) Episome-Containing Keratinocytes and Is Augmented by the HPV-16 E2 Protein. J Virol. 2002;74:3494–504.

36. Walz C, Deprez A, Dupressoir T, Dürst M, Rabreau M, Schlehofer JR. Interaction of human papillomavirus type 16 and adeno-associated virus type 2 co-infecting human cervical epithelium. J Gen Virol. 1997;78:1441–52.

37. Chang LS, Shi Y, Shenk T. Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor. J Virol. 1989;63:3479–88.

38. Laughlin CA, Jones N, Carter BJ. Effect of deletions in adenovirus early region 1 genes upon replication of adeno-associated virus. J Virol. 1982;41:868–76.

39. Richardson WD, Westphal H. Requirement for either early region 1a or early region 1b adenovirus gene products in the helper effect for adeno-associated virus. J Virol. 1984;51:404–10.

40. Shi Y, Seto E, Chang LS, Shenk T. Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. Cell. 1991;67:377–88.

41. Tratschin JD, West MH, Sandbank T, Carter BJ. A human parvovirus, adeno-associated virus, as a eucaryotic vector: transient expression and encapsidation of the procaryotic gene for chloramphenicol acetyltransferase. Mol Cell Biol. 1984;4:2072–81.

42. Chang LS, Shenk T. The adenovirus DNA-binding protein stimulates the rate of transcription directed by adenovirus and adeno-associated virus promoters. J Virol. 1990;64:2103–9.

43. Janik JE, Huston MM, Rose JA. Locations of adenovirus genes required for the replication of adenovirus-associated virus. Proc Natl Acad Sci. 1981;78:1925–9.

44. Carter BJ, Antoni BA, Klessig DF. Adenovirus containing a deletion of the early region 2A gene allows growth of adeno-associated virus with decreased efficiency. Virology. 1992;191:473–6.

45. Janik JE, Huston MM, Cho K, Rose JA. Efficient synthesis of adeno-associated virus structural proteins requires both adenovirus DNA binding protein and VA I RNA. Virology. 1989;168:320–9.

46. Jay FT, Laughlin CA, Carter BJ. Eukaryotic translational control: adeno-associated virus protein synthesis is affected by a mutation in the adenovirus DNA-binding protein. Proc Natl Acad Sci U S A. 1981;78:2927–31.

47. Myers MW, Carter BJ. Adeno-associated Virus Replication - The effect of L-canavanine or a helper virus mutation on accumulation of viral capsids and progeny single-stranded DNA. J Biol Chem. 1981;256:567–70.

48. Myers MW, Laughlin CA, Jay FT, Carter BJ. Adenovirus helper function for growth of adeno-associated virus: Effect of temperature-sensitive mutations in adenovirus early gene region 2. J Virol. 1980;35:65–75.

49. Samulski RJ, Shenk T. Adenovirus E1B 55-Mr polypeptide facilitates timely cytoplasmic accumulation of adeno-associated virus mRNAs. J Virol. 1988;62:206–10.

50. Quinn CO, Kitchingman GR. Functional analysis of the adenovirus type 5 DNA-binding

protein: Site-directed mutants which are defective for adeno-associated virus helper activity. J Virol. 1986;60:653–61.

51. Ward P, Dean FB, O'Donnell ME, Berns KI. Role of the adenovirus DNA-binding protein in in vitro adeno-associated virus DNA replication. J Virol. 1998;72:420–7.

52. Carter BJ, Marcus-Sekura CJ, Laughlin CA, Ketner G. Properties of an adenovirus type 2 mutant, Ad2d/807, having a deletion near the right-hand genome terminus: Failure to help AAV replication. Virology. 1983;126:505–16.

53. Huang MM, Hearing P. Adenovirus early region 4 encodes two gene products with redundant effects in lytic infection. J Virol. 1989;63:2605–15.

54. Richardson WD, Westphal H. A cascade of adenovirus early functions is required for expression of adeno-associated virus. Cell. 1981;27:133–41.

55. Xiao W, Warrington KH, Hearing P, Hughes J, Muzyczka N. Adenovirus-Facilitated Nuclear Translocation of Adeno-Associated Virus Type 2. J Virol. 2002;76:11505–17.

56. Mouw MB, Pintel DJ. Adeno-Associated Virus RNAs Appear in a Temporal Order and Their Splicing Is Stimulated during Coinfection with Adenovirus. J Virol. 2000;74:9878–88.

57. Kotin RM, Siniscalco M, Samulski RJ, Zhu XD, Hunter L, Laughlin CA, et al. Site-specific integration by adeno-associated virus. Proc Natl Acad Sci. 1990;87:2211–5.

58. Duan D, Sharma P, Yang J, Yue Y, Dudus L, Zhang Y, et al. Circular Intermediates of Recombinant Adeno-Associated Virus Have Defined Structural Characteristics Responsible for Long-Term Episomal Persistence in Muscle Tissue. J Virol. 1998;72:8568–77.

59. Samulski R, Zhu X, Xiao X, Brook J, Housman D, Epstein N, et al. Targeted integration of adeno-associated virus (AAV) into human chromosome 19. EMBO J. 1991;10:3941.

60. Kotin RM, Linden RM, Berns KI. Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination. EMBO J. 1992;11:5071–8.

61. Linden RM, Ward P, Giraud C, Winocour E, Berns KI. Site-specific integration by adeno-associated virus. Proc Natl Acad Sci U S A. 1996;93:11288–94.

62. Zentrum für Krebsregisterdaten im Robert Koch-Institut. Bericht zum Krebsgeschehen in Deutschland 2016. 2016.

63. Furman RR, Cheng S, Lu P, Setty M, Perez AR, Guo A, et al. Ibrutinib Resistance in Chronic Lymphocytic Leukemia. N Engl J Med. 2014;370:2349–52.

64. Kaji EH, Leiden JM. Gene and stem cell therapies. J Am Med Assoc. 2001;285:545–50.

65. Wolff JA, Williams P, Chong W, Acsadi G, Jani A, Malone RW, et al. Direct gene transfer into mouse muscle in vivo. Science (80-). 1990;247:1465–8.

66. Felgner PL, Gadek TR, Holm M, Roman R, Chan HW, Wenz M, et al. Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. Proc Natl Acad Sci U S A. 1987;84:7413–7.

67. Liu G, Molas M, Grossmann GA, Pasumarthy M, Perales JC, Cooper MJ, et al. Biological Properties of Poly-L-lysine-DNA Complexes Generated by Cooperative Binding of the Polycation. J Biol Chem. 2001;276:34379–87.

68. Boussif O, Lezoualc'h F, Zanta M a, Mergny MD, Scherman D, Demeneix B, et al. A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. Proc Natl Acad Sci U S A. 1995;92:7297–301.

69. Coster HGL. A Quantitative Analysis of the Voltage-Current Relationships of Fixed Charge Membranes and the Associated Property of "Punch-Through." Biophys J. 1965;5:669–86.

70. Shohet R V, Chen S, Zhou Y, Wang Z. Echocardiographic Destruction of Albumin Microbubbles Directs Gene Delivery to the Myocardium. Circulation. 2000;101:2554–6.

71. Yang NS, Burkholder J, Roberts B, Martinell B, McCabe D. In vivo and in vitro gene transfer to mammalian somatic cells by particle bombardment. Proc Natl Acad Sci. 1990;87:9568–72.

72. Su C-H, Wu Y-J, Wang H-H, Yeh H-I. Nonviral gene therapy targeting cardiovascular system. Am J Physiol Circ Physiol. 2012;303:H629–38.

73. Petrs-Silva H, Linden R. Advances in recombinant adeno-associated viral vectors for gene delivery. Curr Gene Ther. 2013;13:335–45.

74. Luo J, Luo Y, Sun J, Zhou Y, Zhang Y, Yang X. Adeno-associated virus-mediated cancer gene therapy: Current status. Cancer Lett. 2015;356:347–56.

75. Cotter MJ, Muruve DA. The induction of inflammation by adenovirus vectors used for gene therapy. Front Biosci. 2005;10:1098–105.

76. Hagen S, Baumann T, Wagner HJ, Morath V, Kaufmann B, Fischer A, et al. Modular adeno-associated virus (rAAV) vectors used for cellular virus-directed enzyme prodrug therapy. Sci Rep. 2014;4:3759.

77. Büning H, Huber A, Zhang L, Meumann N, Hacker U. Engineering the AAV capsid to

optimize vector–host-interactions. Curr Opin Pharmacol. 2015;24:94–104.

78. Kotterman MA, Schaffer D V. Engineering adeno-associated viruses for clinical gene therapy. Nat Rev Genet. 2014;15:445–51.

79. Dalkara D, Byrne LC, Klimczak RR, Visel M, Yin L, Merigan WH, et al. In Vivo-Directed Evolution of a New Adeno-Associated Virus for Therapeutic Outer Retinal Gene Delivery from the Vitreous. Sci Transl Med. 2013;5:189ra76-189ra76.

80. Warrington KH, Gorbatyuk OS, Harrison JK, Opie SR, Zolotukhin S, Muzyczka N. Adeno-Associated Virus Type 2 VP2 Capsid Protein Is Nonessential and Can Tolerate Large Peptide Insertions at Its N Terminus. J Virol. 2004;78:6595–609.

81. Ried MU, Girod A, Leike K, Buning H, Hallek M. Adeno-Associated Virus Capsids Displaying Immunoglobulin-Binding Domains Permit Antibody-Mediated Vector Retargeting to Specific Cell Surface Receptors. J Virol. 2002;76:4559–66.

82. Dong B, Nakai H, Xiao W. Characterization of genome integrity for oversized recombinant AAV vector. Mol Ther. 2010;18:87–92.

83. Wu Z, Yang H, Colosi P. Effect of genome size on AAV vector packaging. Mol Ther. 2010;18:80–6.

84. Chamberlain K, Riyad JM, Weber T. Expressing Transgenes That Exceed the Packaging Capacity of Adeno-Associated Virus Capsids. Hum Gene Ther Methods. 2016;27:1–12.

85. McCarty DM, Monahan PE, Samulski RJ. Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. Gene Ther. 2001;8:1248–54.

86. McCarty DM. Self-complementary AAV vectors; advances and applications. Mol Ther. 2008;16:1648–56.

87. Powell SK, Rivera-Soto R, Gray SJ. Viral expression cassette elements to enhance transgene target specificity and expression in gene therapy. Discov Med. 2015;19:49–57.

88. Wang B, Li J, Fu FH, Chen C, Zhu X, Zhou L, et al. Construction and analysis of compact muscle-specific promoters for AAV vectors. Gene Ther. 2008;15:1489–99.

89. Yang L, Cao Z, Li F, Post DE, Van Meir EG, Zhong H, et al. Tumor-specific gene expression using the survivin promoter is further increased by hypoxia. Gene Ther. 2004;11:1215–23.

90. Chen J-S, Liu J-C, Shen L, Rau K-M, Kuo H-P, Li YM, et al. Cancer-specific activation of the survivin promoter and its potential use in gene therapy. Cancer Gene Ther.

2004;11:740–7.

91. Sobolewski C, Cerella C, Dicato M, Ghibelli L, Diederich M. The role of cyclooxygenase-2 in cell proliferation and cell death in human malignancies. Int J Cell Biol. 2010;2010:215158.

92. Wang Z-X, Bian H-B, Yang J-S, De W, Ji X-H. Adenovirus-mediated suicide gene therapy under the control of Cox-2 promoter for colorectal cancer. Cancer Biol Ther. 2009;8:1480–8.

93. Zhu Z, Makhija S, Lu B, Wang M, Kaliberova L, Liu B, et al. Transcriptional targeting of adenoviral vector through the CXCR4 tumor-specific promoter. Gene Ther. 2004;11:645–8.

94. Gray JT, Zolotukhin S, Snyder RO, Moullier P. Design and Construction of Functional AAV Vectors. In: Adeno-Associated Virus: Methods and Protocols. Humana Press; 2011. p. 25–46.

95. Geisler A, Fechner H. MicroRNA-regulated viral vectors for gene therapy. World J Exp Med. 2016;6:37–54.

96. Wang XXX, Cao L, Wang Y, Wang XXX, Liu N, You Y. Regulation of let-7 and its target oncogenes (Review). Oncol Lett. 2012;3:955–60.

97. Limberis MP, Vandenberghe LH, Zhang L, Pickles RJ, Wilson JM. Transduction efficiencies of novel AAV vectors in mouse airway epithelium in vivo and human ciliated airway epithelium in vitro. Mol Ther. 2009;17:294–301.

98. Bartel MA, Weinstein JR, Schaffer D V. Directed evolution of novel adeno-associated viruses for therapeutic gene delivery. Gene Ther. 2012;19:694–700.

99. Castle MJ, Turunen HT, Vandenberghe LH, Wolfe JH. Controlling AAV Tropism in the Nervous System with Natural and Engineered Capsids. In: Gene Therapy for Neurological Disorders: Methods and Protocols. New York, NY: Springer New York; 2016. p. 133–49.

100. Tseng YS, Agbandje-McKenna M. Mapping the AAV capsid host antibody response toward the development of second generation gene delivery vectors. Front Immunol. 2014;5 JAN:1–11.

101. Li C, Diprimio N, Bowles DE, Hirsch ML, Monahan PE, Asokan A, et al. Single Amino Acid Modification of Adeno-Associated Virus Capsid Changes Transduction and Humoral Immune Profiles. J Virol. 2012;86:7752–9.

102. Maersch S, Huber A, Büning H, Hallek M, Perabo L. Optimization of stealth adeno-associated virus vectors by randomization of immunogenic epitopes. Virology. 2010;397:167–75.

103. Bartel M, Schaffer D, Büning H. Enhancing the clinical potential of aav vectors by capsid engineering to evade pre-existing immunity. Front Microbiol. 2011;2 OCT:1–10.

104. Aslanidi G V., Rivers AE, Ortiz L, Song L, Ling C, Govindasamy L, et al. Optimization of the Capsid of Recombinant Adeno-Associated Virus 2 (AAV2) Vectors: The Final Threshold? PLoS One. 2013;8.

105. Zhong L, Li B, Mah CS, Govindasamy L, Agbandje-McKenna M, Cooper M, et al. Next generation of adeno-associated virus 2 vectors: Point mutations in tyrosines lead to high-efficiency transduction at lower doses. Proc Natl Acad Sci. 2008;105:7827–32.

106. Mays LE, Wang L, Lin J, Bell P, Crawford A, Wherry EJ, et al. AAV8 induces tolerance in murine muscle as a result of poor APC transduction, T cell exhaustion, and minimal MHCI upregulation on target cells. Mol Ther. 2014;22:28–41.

107. Münch RC, Janicki H, Völker I, Rasbach A, Hallek M, Büning H, et al. Displaying high-affinity ligands on adeno-associated viral vectors enables tumor cell-specific and safe gene transfer. Mol Ther. 2013;21:109–18.

108. Ltd. JW and S. Gene Therapy Clinical Trials Worldwide. http://www.abedia.com/wiley/vectors.php. Accessed 29 Jul 2019.

109. Tratschin JD, West MH, Sandbank T, Carter BJ. A human parvovirus, adeno-associated virus, as a eucaryotic vector: transient expression and encapsidation of the procaryotic gene for chloramphenicol acetyltransferase. Mol Cell Biol. 1984;4:2072–81.

110. EMA. Glybera. https://www.ema.europa.eu/en/documents/product-information/glybera-epar-product-information_de.pdf. Accessed 29 Jul 2019.

111. FDA. Luxturna. https://www.fda.gov/media/109906/download. Accessed 29 Jul 2019.

112. FDA. Zolgensma. https://www.fda.gov/vaccines-blood-biologics/zolgensma. Accessed 29 Jul 2019.

113. Grimm D, Kay MA, Kleinschmidt JA. Helper virus-free, optically controllable, and two-plasmid-based production of adeno-associated virus vectors of serotypes 1 to 6. Mol Ther. 2003;7:839–50.

114. Feiner R, Teschner J, Teschner K, Radukic M, Baumann T, Hagen S, et al. rAAV engineering for capsid-protein enzyme insertions and mosaicism reveals resilience to mutational, structural and thermal perturbations. unpublished. 2019.

115. Allen JM, Debelak DJ, Reynolds TC, Miller AD. Identification and elimination of replication-competent adeno-associated virus (AAV) that can arise by nonhomologous recombination during AAV vector production. J Virol. 1997;71:6816–22.

116. Samulski RJ, Muzyczka N. AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. Annu Rev Virol. 2014;1:427–51.

117. Kotin R. Large-scale recombinant adeno-associated virus production. Hum Mol Genet. 2011;:1–23.

118. Clément N, Grieger JC. Manufacturing of recombinant adeno-associated viral vectors for clinical trials. Mol Ther - Methods Clin Dev. 2016;3 November 2015:16002.

119. Wright JF. Manufacturing and characterizing AAV-based vectors for use in clinical studies. Gene Ther. 2008;15:840–8.

120. Wright JF. Transient transfection methods for clinical adeno-associated viral vector production. Hum Gene Ther. 2009;20:698–706.

121. Lock M, Alvira M, Vandenberghe LH, Samanta A, Toelen J, Debyser Z, et al. Rapid, Simple, and Versatile Manufacturing of Recombinant Adeno-Associated Viral Vectors at Scale. Hum Gene Ther. 2010;21:1259–71.

122. Grieger JC, Soltys SM, Samulski RJ. Production of Recombinant Adeno-associated Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector. Mol Ther. 2016;24:287–97.

123. Feiner R, Teschner K, Teschner J, Müller K. HEK293-KARE1, a cell line with stably integrated adenovirus helper sequences simplifies rAAV production. unpublished. 2019.

124. Batchu RB, Shammas MA, Wang JY, Munshi NC. Interaction of adeno-associated virus Rep78 with p53: Implications in growth inhibition. Cancer Res. 1999;59:3592–5.

125. Schmidt M, Afione S, Kotin RM. Adeno-Associated Virus Type 2 Rep78 Induces Apoptosis through Caspase Activation Independently of p53. J Virol. 2000;74:9441–50.

126. Thorne BA, Takeya RK, Peluso RW. Manufacturing Recombinant Adeno-Associated Viral Vectors from Producer Cell Clones. Hum Gene Ther. 2009;20:707–14.

127. Mietzsch M, Grasse S, Zurawski C, Weger S, Bennett A, Agbandje-McKenna M, et al. OneBac: Platform for Scalable and High-Titer Production of Adeno-Associated Virus Serotype 1–12 Vectors for Gene Therapy. Hum Gene Ther. 2013;25:212–22.

128. Mietzsch M, Hering H, Hammer E-M, Agbandje-McKenna M, Zolotukhin S, Heilbronn R. OneBac 2.0: Sf9 Cell Lines for Production of AAV1, AAV2, and AAV8 Vectors with Minimal Encapsidation of Foreign DNA. Hum Gene Ther Methods. 2017;28:15–22.

129. Thomas DL, Wang L, Niamke J, Liu J, Kang W, Scotti MM, et al. Scalable recombinant adeno-associated virus production using recombinant herpes simplex virus type 1 coinfection of suspension-adapted mammalian cells. Hum Gene Ther.

2009;20:861–70.

130. Clément N, Knop DR, Byrne BJ. Large-Scale Adeno-Associated Viral Vector Production Using a Herpesvirus-Based System Enables Manufacturing for Clinical Studies. Hum Gene Ther. 2009;20:796–806.

131. Naso MF, Tomkowicz B, Perry WL, Strohl WR. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. BioDrugs. 2017;31:317–34.

132. Chulay JD, Ye G-J, Thomas DL, Knop DR, Benson JM, Hutt JA, et al. Preclinical Evaluation of a Recombinant Adeno-Associated Virus Vector Expressing Human Alpha-1 Antitrypsin Made Using a Recombinant Herpes Simplex Virus Production Method. Hum Gene Ther. 2010;22:155–65.

133. Zen Z, Espinoza Y, Bleu T, Sommer JM, Wright JF. Infectious titer assay for adeno-associated virus vectors with sensitivity sufficient to detect single infectious events. Hum Gene Ther. 2004;15:709–15.

134. Zolotukhin S, Byrne BJ, Mason E, Zolotukhin I, Potter M, Chesnut K, et al. Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. Gene Ther. 1999;6:973–85.

135. Fraser Wright J, Wellman J, High K. Manufacturing and Regulatory Strategies for Clinical AAV2-hRPE65. Curr Gene Ther. 2012;10:341–9.

136. Arden E, Metzger JM. Inexpensive, serotype-independent protocol for native and bioengineered recombinant adeno-associated virus purification. J Biol Methods. 2016;3:38.

137. Clark KR, Liu X, McGrath JP, Johnson PR. Highly purified recombinant adeno-associated virus vectors are biologically active and free of detectable helper and wild-type viruses. Human gene therapy. 1999;10:1031–9.

138. Grimm D, Kern A, Rittner K, Kleinschmidt JA. Novel tools for production and purification of recombinant adenoassociated virus vectors. Hum Gene Ther. 1998;9:2745–60.

139. Healthcare GE, Sciences L. Custom designed media AVB Sepharose ™ High Performance. https://www.gelifesciences.co.jp/catalog/pdf/28920754ab.pdf. Accessed 1 Aug 2019.

140. Thermo Fisher Scientific. POROS CaptureSelect AAVX Affinity Resin. https://www.thermofisher.com/order/catalog/product/A36739. Accessed 14 May 2019.

141. Auricchio A, O'Connor E, Hildinger M, Wilson JM. A single-step affinity column for purification of serotype-5 based adeno-associated viral vectors. Mol Ther. 2001;4:372–4.

142. Brument N, Morenweiser R, Blouin V, Toublanc E, Raimbaud I, Chérel Y, et al. A versatile and scalable two-step ion-exchange chromatography process for the purification of recombinant adeno-associated virus serotypes-2 and -5. Mol Ther. 2002;6:678–86.

143. Davidoff AM, Ng CYC, Sleep S, Gray J, Azam S, Zhao Y, et al. Purification of recombinant adeno-associated virus type 8 vectors by ion exchange chromatography generates clinical grade vector stock. J Virol Methods. 2004;121:209–15.

144. Kaludov N, Handelman B, Chiorini JA. Scalable Purification of Adeno-Associated Virus Type 2, 4, or 5 Using Ion-Exchange Chromatography. Hum Gene Ther. 2002;13:1235–43.

145. Okada T, Nonaka-Sarukawa M, Uchibori R, Kinoshita K, Hayashita-Kinoh H, Nitahara-Kasahara Y, et al. Scalable purification of adeno-associated virus serotype 1 (AAV1) and AAV8 vectors, using dual ion-exchange adsorptive membranes. Hum Gene Ther. 2009;20:1013–21.

146. Potter M, Lins B, Mietzsch M, Heilbronn R, Van Vliet K, Chipman P, et al. A simplified purification protocol for recombinant adeno-associated virus vectors. Mol Ther Methods Clin Dev. 2014;1 March:14034.

147. Qu G, Bahr-Davidson J, Prado J, Tai A, Cataniag F, McDonnell J, et al. Separation of adeno-associated virus type 2 empty particles from genome containing vectors by anion-exchange column chromatography. J Virol Methods. 2007;140:183–92.

148. iGEM. Registry of Standard Biological Parts - pSB1C3. http://parts.igem.org/Part:pSB1C3. Accessed 8 Feb 2019.

149. Thermo Fisher Scientific. CloneJET PCR Cloning Kit. https://www.thermofisher.com/order/catalog/product/K1231. Accessed 8 Feb 2019.

150. Thermo Fisher Scientific. pUC19 DNA. https://www.thermofisher.com/order/catalog/product/SD0061. Accessed 11 Feb 2019.

151. Thermo Fisher Scientific. pcDNA™5/FRT Mammalian Expression Vector. https://www.thermofisher.com/order/catalog/product/V601020. Accessed 8 Feb 2019.

152. Merck. pET-21a(+) DNA - Novagen. http://www.merckmillipore.com/DE/de/product/pET-21a+-DNA-Novagen,EMD_BIO-69740#anchor_VMAP. Accessed 11 Feb 2019.

153. Merck. pET-24b(+) DNA - Novagen. http://www.merckmillipore.com/DE/de/product/pET-24b+-DNA-Novagen,EMD_BIO-69750. Accessed 12 Feb 2019.

154. Freiburg Bioware. Virus Construction Kit for Therapy.

http://2010.igem.org/Team:Freiburg_Bioware. Accessed 12 Feb 2019.

155. Agilent. AAV Helper-Free System. https://www.agilent.com/en/product/protein-expression/protein-expression-vectors-kits/viral-mediated-delivery-systems/aav-helper-free-system-232989. Accessed 12 Feb 2019.

156. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013;339:819–23.

157. Thermo Fisher Scientific. GeneRuler DNA Ladders. https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/thermo-scientific-nucleic-acid-electrophoresis-purification/dna-electrophoresis-thermo-scientific/dna-ladders-thermo-scientific/generuler-dna-ladders.html. Accessed 6 Feb 2019.

158. New England Biolabs. Color Prestained Protein Standard. https://international.neb.com/products/p7712-color-prestained-protein-standard-broad-range-11-245-kda#Product Information_Notes. Accessed 6 Feb 2019.

159. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol. 1977;36:59–72.

160. DSMZ. 293 - DSMZ. https://www.dsmz.de/collection/catalogue/details/culture/ACC-305. Accessed 10 Sep 2019.

161. Thermo Fisher Scientific. Freestyle 293F. https://www.thermofisher.com/order/catalog/product/R79007. Accessed 10 Sep 2019.

162. Rasheed S, Nelson-Rees WA, Toth EM, Arnstein P, Gardner MB. Characterization of a newly derived human sarcoma cell line (HT-1080). Cancer. 1974;33:1027–33.

163. DSMZ. HT-1080 - DSMZ. https://www.dsmz.de/collection/catalogue/details/culture/ACC-315. Accessed 10 Sep 2019.

164. Kao FT, Puck TT. Genetics of somatic mammalian cells, VII. Induction and isolation of nutritional mutants in Chinese hamster cells. Proc Natl Acad Sci. 1968;60:1275–81.

165. DSMZ. CHO-K1 - DSMZ. https://www.dsmz.de/collection/catalogue/details/culture/ACC-110. Accessed 10 Sep 2019.

166. Sambrook J, Russell DW. Molecular Cloning: A Laboratory Manual, Volume 1. CSHL Press; 2001.

167. Hanahan D, Harbor CS. Studies on Transformation of Escherichia coli with Plasmids.

J Mol Biol. 1983;166:557–80.

168. Macherey-Nagel. NucleoSpin® Plasmid. https://www.mn-net.com/ProductsBioanalysis/DNAandRNApurification/PlasmidDNApurificationeasyfastreliable/NucleoSpinPlasmidplasmidMiniprepkit/tabid/1379/language/en-US/Default.aspx. Accessed 15 Feb 2019.

169. Macherey-Nagel. NucleoBond® Xtra Midi / Maxi. https://www.mn-net.com/ProductsBioanalysis/DNAandRNApurification/PlasmidDNApurificationeasyfastreliable/NucleoBondXtraplasmidMidiprepMaxiprepkits/tabid/1479/language/en-US/Default.aspx. Accessed 15 Feb 2019.

170. Macherey-Nagel. NucleoSpin® Tissue. https://www.mn-net.com/ProductsBioanalysis/DNAandRNApurification/DNA/DNAfromtissueandcells/NucleoSpinTissue/tabid/1353/language/en-US/Default.aspx. Accessed 15 Feb 2019.

171. Glasel JA. Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. Biotechniques. 1995;18:62–3.

172. New England Biolabs. Cleavage Close to the End of DNA Fragments. https://international.neb.com/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments. Accessed 19 Jun 2019.

173. Macherey-Nagel. NucleoSpin® Gel and PCR Clean-up. https://www.mn-net.com/Products/DNAandRNApurification/Cleanup/NucleoSpinGelandPCRCleanup/tabid/1452/language/en-US/Default.aspx. Accessed 15 Feb 2019.

174. New England Biolabs. Antarctic Phosphatase. https://international.neb.com/products/m0289-antarctic-phosphatase#Protocols & Manuals. Accessed 14 Feb 2019.

175. New England Biolabs. T4 Polynucleotide Kinase. https://international.neb.com/products/m0201-t4-polynucleotide-kinase#Protocols & Manuals. Accessed 14 Feb 2019.

176. Thermo Fisher Scientific. T4 DNA Ligase. https://www.thermofisher.com/order/catalog/product/EL0011. Accessed 14 Feb 2019.

177. Tiselius A. A new apparatus for electrophoretic analysis of colloidal mixtures. In: Transactions of the Faraday Society. 1937. p. 524–31.

178. Serwer P. Agarose gels: Properties and use for electrophoresis. Electrophoresis. 1983;4:375–82.

179. Carl Roth. Roti®-GelStain. https://www.carlroth.com/downloads/ba/en/3/BA_3865_EN.pdf. Accessed 18 Feb 2019.

180. iGEM. Help:Standards/Assembly/RFC10. http://parts.igem.org/Help:Standards/Assembly/RFC10. Accessed 21 Feb 2019.

181. iGEM. Help:Standards/Assembly/RFC25. http://parts.igem.org/Help:Standards/Assembly/RFC25. Accessed 21 Feb 2019.

182. Zhang Y, Werling U, Edelmann W. SLiCE: A novel bacterial cell extract-based DNA cloning method. Nucleic Acids Res. 2012;40:1–10.

183. Motohashi K. A simple and efficient seamless DNA cloning method using SLiCE from Escherichia coli laboratory strains and its application to SLiP site-directed mutagenesis. BMC Biotechnol. 2015;15.

184. Klausing S. Cationic-polymer (PEI) mediated transfection. 2013.

185. Neu HC, Heppel LA. The release of enzymes from Escherichia coli by osmotic shock and during the formation of spheroplasts. J Biol Chem. 1965;240:3685–92.

186. Macherey-Nagel. Purification of His-tag proteins - Protino Ni-NTA Agarose. https://www.mn-net.com/Portals/8/attachments/Redakteure_Bio/Protocols/Protino/UM_ProtinoNi-NTA.pdf. Accessed 27 Feb 2019.

187. GE Healthcare. Columns Instructions. https://cdn.gelifesciences.com/dmm3bwsv3/AssetStream.aspx?mediaformatid=10061&destinationid=10016&assetid=12797. Accessed 27 Feb 2019.

188. GE Healthcare. ÄKTA start documents. https://www.gelifesciences.com/en/us/shop/chromatography/chromatography-systems/akta-start-p-05773#related-documents. Accessed 27 Feb 2019.

189. GE Healthcare. HiTrap Protein A HP column. https://cdn.gelifesciences.com/dmm3bwsv3/AssetStream.aspx?mediaformatid=10061&destinationid=10016&assetid=14403. Accessed 27 Feb 2019.

190. Hoefer. Hoefer Multiple Gel Casters. https://www.hoeferinc.com/media/PDFs/Manuals/SE215-IML0-English.pdf. Accessed 25 Feb 2019.

191. Hoefer. Hoefer™ Mighty Small™ II Mini Vertical Electrophoresis Systems. https://www.fishersci.com/shop/products/hoefer-mighty-small-ii-mini-vertical-electrophoresis-systems-5/p-3004140. Accessed 25 Feb 2019.

192. Thermo Scientific. SuperSignal West Pico PLUS Chemiluminescent Substrate. https://www.thermofisher.com/document-connect/document-connect.html?url=https%3A%2F%2Fassets.thermofisher.com%2FTFS-

Assets%2FLSG%2Fmanuals%2FMAN0015920_2162617_SuperSigWestPicoPLUS_Chemil_Substr_UG.pdf&title=VXNlciBHdWlkZTogU3VwZXJTaWduYWwgV2VzdCBQaWNvIFB. Accessed 26 Feb 2019.

193. Roche. LightCycler 480 Instrument II. https://lifescience.roche.com/en_de/products/lightcycler14301-480-instrument-ii.html#documents. Accessed 5 Mar 2019.

194. Promega. GoTaq qPCR Master Mix. https://www.promega.de/-/media/files/resources/protocols/technical-manuals/101/gotaq-qpcr-master-mix-protocol.pdf?la=de-de. Accessed 5 Mar 2019.

195. Aurnhammer C, Haase M, Muether N, Hausl M, Rauschhuber C, Huber I, et al. Universal Real-Time PCR for the Detection and Quantification of Adeno-Associated Virus Serotype 2-Derived Inverted Terminal Repeat Sequences. Hum Gene Ther Methods. 2012;23:18–28.

196. Grimm D, Kern a, Pawlita M, Ferrari F, Samulski R, Kleinschmidt J. Titration of AAV-2 particles via a novel capsid ELISA: packaging of genomes can limit production of recombinant AAV-2. Gene Ther. 1999;6:1322–30.

197. Kuck D, Kern A, Kleinschmidt J a. Development of AAV serotype-specific ELISAs using novel monoclonal antibodies. J Virol Methods. 2007;140:17–24.

198. Rayaprolu V, Kruse S, Kant R, Venkatakrishnan B, Movahed N, Brooke D, et al. Comparative Analysis of Adeno Associated Virus Capsid Stability and Dynamics. J Virol. 2013;87:13150–60.

199. Percival Zhang YH, Cui J, Lynd LR, Kuang LR. A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: Evidence from enzymatic hydrolysis and supramolecular structure. Biomacromolecules. 2006;7:644–8.

200. Yu A, Shang J, Cheng F, Paik BA, Kaplan JM, Andrade RB, et al. Biofunctional Paper via the Covalent Modification of Cellulose. Langmuir. 2012;28:11265–73.

201. Wang XS, Srivastava A. Rescue and autonomous replication of adeno-associated virus type 2 genomes containing Rep-binding site mutations in the viral p5 promoter. J Virol. 1998;72:4811–8.

202. Samulski RJ, Srivastava A, Berns KI, Muzyczka N. Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. Cell. 1983;33:135–43.

203. CellBiolabs. pAAV-MCS expression vector VPK-410. 2016.

204. Troll C, Yoder J, Alexander D, Hernández J, Loh Y, Camps M. The mutagenic

footprint of low-fidelity Pol I ColE1 plasmid replication in E. coli reveals an extensive interplay between Pol I and Pol III. Curr Genet. 2014;60:123–34.

205. Xie J, Mao Q, Tai PWL, He R, Ai J, Su Q, et al. Short DNA Hairpins Compromise Recombinant Adeno-Associated Virus Genome Homogeneity. Mol Ther. 2017;25:1–12.

206. Speck J, Hecky J, Tam HK, Arndt KM, Einsle O, Müller KM. Exploring the molecular linkage of protein stability traits for enzyme optimization by iterative truncation and evolution. Biochemistry. 2012;51:4850–67.

207. Baumann T. Stability and interconnected protein properties studied with TEM β-lactamase. University of Potsdam; 2013.

208. Ros C, Baltzer C, Mani B, Kempf C. Parvovirus uncoating in vitro reveals a mechanism of DNA release without capsid disassembly and striking differences in encapsidated DNA stability. Virology. 2006;345:137–47.

209. Bernaud J, Rossi A, Fis A, Gardette L, Aillot L, Büning H, et al. Characterization of AAV vector particle stability at the single-capsid level. J Biol Phys. 2018;44:181–94.

210. Bennett A, Patel S, Mietzsch M, Jose A, Lins-Austin B, Yu JC, et al. Thermal Stability as a Determinant of AAV Serotype Identity. Mol Ther - Methods Clin Dev. 2017;6 September:171–82.

211. Hecky J, Müller KM. Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of β-lactamase. Biochemistry. 2005;44:12640–54.

212. Rehm FBH, Chen S, Rehm BHA. Enzyme engineering for in situ immobilization. Molecules. 2016;21.

213. Jayapal K, Wlaschin K, Hu W, Yap M. Recombinant protein therapeutics from CHO cells - 20 years and counting. Chem Eng. 2007;103:40–7.

214. Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, et al. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat Biotechnol. 2011;29:735–41.

215. Bird A. Epigentic Memory. Genes Dev. 2002;16:16–21.

216. UCSC Genome Bioinformatics group. UCSC Genome Browser. https://genome.ucsc.edu/index.html. Accessed 8 Apr 2019.

217. Florea BI, Meaney C, Junginger HE, Borchard G. Transfection efficiency and toxicity of polyethylenimine in differentiated Calu-3 and nondifferentiated COS-1 cell cultures. AAPS PharmSci. 2002;4.

218. Thermo Fisher Scientific. pcDNA™6/myc-His A, B, and C.

https://assets.thermofisher.com/TFS-Assets/LSG/manuals/pcdna6mychis_man.pdf. Accessed 16 Sep 2019.

219. Wistuba a, Weger S, Kern A, Kleinschmidt JA. Intermediates of adeno-associated virus type 2 assembly: identification of soluble complexes containing Rep and Cap proteins. J Virol. 1995;69:5311–9.

220. Wistuba A, Kern A, Weger S, Grimm D, Kleinschmidt RA, Krebsforschungszentrum D, et al. Subcellular Compartmentalization of Adeno-Associated Virus Type 2 Assembly. 1997;71:1341–52.

221. McCraw DM, O'Donnell JK, Taylor KA, Stagg SM, Chapman MS. Structure of adeno-associated virus-2 in complex with neutralizing monoclonal antibody A20. Virology. 2012;431:40–9.

222. Chen X, Zaro JL, Shen W-C. Fusion protein linkers: property, design and functionality. Adv Drug Deliv Rev. 2013;65:1357–69.

223. Khantasup K, Chantima W, Sangma C, Poomputsa K, Dharakul T. Design and Generation of Humanized Single-chain Fv Derived from Mouse Hybridoma for Potential Targeting Application. Monoclon Antib Immunodiagn Immunother. 2015;34:404–17.

224. Lei S-P, Wang S-S, Higaki P, Wilcox G, Lin H-C. Characterization of the Erwinia carotovora pelB Gene and Its Product Pectate Lyase. J Bacteriol. 1998;169:4379–83.

225. Ikeda T, Ninomiya K, Hirota R, Kuroda A. Single-step affinity purification of recombinant proteins using the silica-binding Si-tag as a fusion partner. Protein Expr Purif. 2010;71:91–5.

226. Abdelhamid MAA, Motomura K. Affinity purification of recombinant proteins using a novel silica-binding peptide as a fusion tag. 2014;:5677–84.

227. Peteranderl R, Shotts EB, Wiegel J. Stability of antibiotics under growth conditions for thermophilic anaerobes. Appl Environ Microbiol. 1990;56:1981–3.

228. Zhang R, Cao L, Cui M, Sun Z, Hu M, Zhang R, et al. Adeno-associated virus 2 bound to its cellular receptor AAVR. Nat Microbiol. 2019;4:675–82.

229. Mietzsch M, Grasse S, Zurawski C, Weger S, Bennett A, Agbandje-McKenna M, et al. OneBac: platform for scalable and high-titer production of adeno-associated virus serotype 1-12 vectors for gene therapy. Hum Gene Ther. 2014;25:212–22.

230. Wobus CE, Hügle-Dörr B, Girod A, Petersen G, Hallek M, Kleinschmidt J a. Monoclonal antibodies against the adeno-associated virus type 2 (AAV-2) capsid: epitope mapping and identification of capsid domains involved in AAV-2-cell interaction and neutralization of AAV-2 infection. J Virol. 2000;74:9281–93.

231. Tomme P, Boraston A, McLean B, Kormos J, Creagh AL, Sturch K, et al. Characterization and affinity applications of cellulose-binding domains. J Chromatogr B Biomed Sci Appl. 1998;715:283–96.

232. Ong E, Gilkes NR, Miller RC, Warren RAJ, Kilburn DG. The cellulose-binding domain (CBDCex) of an exoglucanase from Cellulomonas fimi: Production in Escherichia coli and characterization of the polypeptide. Biotechnol Bioeng. 1993;42:401–9.

233. Pillay S, Zou W, Cheng F, Puschnik AS, Meyer NL, Ganaie SS, et al. AAV serotypes have distinctive interactions with domains of the cellular receptor AAVR. J Virol. 2017;91:JVI.00391-17.

234. Sun P, Tropea JE, Waugh DS. Enhancing the Solubility of Recombinant Proteins in Escherichia coli by Using Hexahistidine-Tagged Maltose-Binding Protein as a Fusion Partner. In: Methods in molecular biology (Clifton, N.J.). 2011. p. 259–74.

235. Mestrom L, Marsden SR, Dieters M, Achterberg P, Stolk L, Bento I, et al. Artificial fusion of mCherry enhances trehalose transferase solubility and stability. Appl Environ Microbiol. 2019;85:1–16.

236. Chen H. Comparative observation of the recombinant adeno-associated virus 2 using transmission electron microscopy and atomic force microscopy. Microsc Microanal. 2007;13:384–9.

237. FDA Advisory Committee. Spark Therapeutics Briefing Document. 2017.

238. Rose JA, Maizel J V, Inman JK, Shatkin AJ. Structural proteins of adenovirus-associated viruses. J Virol. 1971;8:766–70.

239. Matsushita T, Elliger S, Elliger C, Podsakoff G, Villarreal L, Kurtzman GJ, et al. Adeno-associated virus vectors can be efficiently produced without helper virus. Gene Ther. 1998;5:938–45.

240. Xiao X, Li J, Samulski RJ. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. J Virol. 1998;72:2224–32.

241. Feiner RC, Müller KM. Recent progress in protein-protein interaction study for EGFR-targeted therapeutics. Expert Rev Proteomics. 2016;13:817–32.

242. Grimm D, Kay M a. From virus evolution to vector revolution: use of naturally occurring serotypes of adeno-associated virus (AAV) as novel vectors for human gene therapy. Curr Gene Ther. 2003;3:281–304.

243. Wu Z, Asokan A, Samulski RJ. Adeno-associated virus serotypes: vector toolkit for human gene therapy. Mol Ther. 2006;14:316–27.

244. Lux K, Goerlitz N, Schlemminger S, Perabo L, Goldnau D, Endell J, et al. Green Fluorescent Protein-Tagged Adeno-Associated Virus Particles Allow the Study of Cytosolic and Nuclear Trafficking. J Virol. 2005;79:11776–87.

245. Girod, Ried, Wobus, Lahm, Leike, Kleinschmidt, et al. Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. Nat Med. 1999;5:1438.

246. Shi W, Bartlett JS. RGD inclusion in VP3 provides adeno-associated virus type 2 (AAV2)-based vectors with a heparan sulfate-independent cell entry mechanism. Mol Ther. 2003;7:515–25.

247. Liu Y, Fang Y, Zhou Y, Zandi E, Lee C-L, Joo K-I, et al. Site-specific modification of adeno-associated viruses via a genetically engineered aldehyde tag. Small. 2013;9:421–9.

248. Falck G, Müller KM. Enzyme-Based Labeling Strategies for Antibody–Drug Conjugates and Antibody Mimetics. Antibodies. 2018;7:4.

249. Ried MU, Girod A, Leike K, Büning H, Hallek M. Adeno-associated virus capsids displaying immunoglobulin-binding domains permit antibody-mediated vector retargeting to specific cell surface receptors. J Virol. 2002;76:4559–66.

250. Nieto K, Weghofer M, Sehr P, Ritter M, Sedlmeier S, Karanam B, et al. Development of AAVLP(HPV16/31L2) particles as broadly protective HPV vaccine candidate. PLoS One. 2012;7:e39741.

251. Judd J, Wei F, Nguyen PQ, Tartaglia LJ, Agbandje-McKenna M, Silberg JJ, et al. Random Insertion of mCherry Into VP3 Domain of Adeno-associated Virus Yields Fluorescent Capsids With no Loss of Infectivity. Mol Ther Nucleic Acids. 2012;1 September:e54.

252. Freiburg iGEM Team. Virus Construction Kit For Therapy. http://2010.igem.org/Team:Freiburg_Bioware. Accessed 24 Jan 2019.

253. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003;31:3406–15.

254. Guo P, El-Gohary Y, Prasadan K, Shiota C, Xiao X, Wiersch J, et al. Rapid and simplified purification of recombinant adeno-associated virus. J Virol Methods. 2012;183:139–46.

255. Horowitz ED, Rahman KS, Bower BD, Dismuke DJ, Falvo MR, Griffith JD, et al. Biophysical and ultrastructural characterization of adeno-associated virus capsid uncoating and genome release. J Virol. 2013;87:2994–3002.

256. Zeltner N, Kohlbrenner E, Clément N, Weber T, Linden RM. Near-perfect infectivity

of wild-type AAV as benchmark for infectivity of recombinant AAV vectors. Gene Ther. 2010;17:872–9.

257. Ellis BL, Hirsch ML, Barker JC, Connelly JP, Steininger RJ, Porteus MH. A survey of ex vivo/in vitro transduction efficiency of mammalian primary cells and cell lines with Nine natural adeno-associated virus (AAV1-9) and one engineered adeno-associated virus serotype. Virol J. 2013;10:74.

258. Trempe JP, Carter BJ. Alternate mRNA splicing is required for synthesis of adeno-associated virus VP1 capsid protein. J Virol. 1988;62:3356–63.

259. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. J Comput Chem. 2004;25:1605–12.

260. Warrington KH, Gorbatyuk OS, Harrison JK, Opie SR, Zolotukhin S, Muzyczka N. Adeno-associated virus type 2 VP2 capsid protein is nonessential and can tolerate large peptide insertions at its N terminus. J Virol. 2004;78:6595–609.

261. Shi W, Arnold GS, Bartlett JS. Insertional mutagenesis of the adeno-associated virus type 2 (AAV2) capsid gene and generation of AAV2 vectors targeted to alternative cell-surface receptors. Hum Gene Ther. 2001;12:1697–711.

262. Zhang H, Xie J, Dmitriev I, Kashentseva E, Curiel DT, Hsu H, et al. Addition of six-His-tagged peptide to the C terminus of adeno-associated virus VP3 does not affect viral tropism or production. J Virol. 2002;76:12023–31.

263. Kozak M. Initiation of translation in prokaryotes and eukaryotes. Gene. 1999;234:187–208.

264. Knight T. Idempotent Vector Design for Standard Assembly of Biobricks Standard Biobrick Sequence Interface. BBF RFC. 2007. http://hdl.handle.net/1721.1/45138. Accessed 24 Jan 2019.

265. Müller KM, Arndt KM, IGEM_Freiburg, Grünberg R. Fusion Protein (Freiburg) Biobrick assembly standard. http://hdl.handle.net/1721.1/45140. Accessed 24 Jan 2019.

266. Agilent Technologies. AAV Helper-Free System Instruction Manual. https://www.agilent.com/cs/library/usermanuals/Public/240071.pdf. Accessed 13 Aug 2019.

267. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;9:671–5.

10 **Appendix**

10.1 **Manuscripts**

During the work on this thesis, I was involved in the writing and experimental design of three manuscripts for publications. In two cases manuscripts were submitted to the respective journals, but due to some shortcomings they were rejected, and the manuscripts were corrected and extended and then submitted again. The correction of the manuscript entitled “rAAV engineering for capsid-protein enzyme insertions and mosaicism reveals resilience to mutational, structural and thermal perturbations” (short title: “rAAV system evaluation”) was submitted to the “International Journal of Molecular Sciences”. I was involved in the writing and correction of the manuscript as well as in the experimental design and execution. Specifically, I performed and analyzed the systematic variation of rAAV loop modifications. Furthermore, I performed the experiments on production, characterization and enzymatic activity detection of the rAAV β-lactamase variants. Ultimately, I was responsible for the experimental design and analysis of the ITR sequencing experiments.

The correction of the manuscript entitled “HEK293-KARE1, a cell line with stably integrated adenovirus helper sequences simplifies rAAV production” (short title: “HEK293-KARE1”) was resubmitted in the journal “BMC Biotechnology”. My part in this project was the experimental design and generation of the HEK293-KARE1 cell lines as well as the correction of the manuscript.

The submitted manuscripts and supporting information are shown in the following subchapters. The third manuscript, which deals with PKD2-based AAV purification, had not yet been completed at the time of printing of this thesis.

10.1.1 rAAV system evaluation

10.1.1.1 Manuscript of "rAAV system evaluation"

Article

rAAV engineering for capsid-protein enzyme insertions and mosaicism reveals resilience to mutational, structural and thermal perturbations

Rebecca C. Feiner[1,†], Julian Teschner[1,†], Kathrin E. Teschner[1], Marco T. Radukic[1], Tobias Baumann[2], Sven Hagen[3], Yvonne Hannappel[4], Niklas Biere[5], Dario Anselmetti[5], Katja M. Arndt[6], Kristian M. Müller[1,*]

[1] Cellular and Molecular Biotechnology, Faculty of Technology, Bielefeld University, Bielefeld, Germany; rebecca.feiner@uni-bielefeld.de (R.C.F.); julian.teschner@uni-bielefeld.de (J.T.); kathrin.schlicht@uni-bielefeld.de (K.E.T.); marco.radukic@uni-bielefeld.de (M.T.R.)
[2] present address: Biocatalysis group, Department of Chemistry, Technische Universität Berlin, Berlin, Germany; tobias.baumann@tu-berlin.de (T.B.)
[3] present address: CO.DON AG, Berlin, Germany; sven.hagen@outlook.com (S.H.)
[4] Physical and Biophysical Chemistry (PCIII), Department of Chemistry, Bielefeld University, Germany; yvonne.hannappel@uni-bielefeld.de (Y.H.)
[5] Experimental Biophysics and Applied Nanoscience, Physics Department, Bielefeld University, Germany; nbiere@physik.uni-bielefeld.de (N.B.); dario.anselmetti@physik.uni-bielefeld.de (D.A.)
[6] Molecular Biotechnology, Institute for Biochemistry and Biology, University of Potsdam, 14476 Potsdam, Germany; arndtk@uni-potsdam.de (K.M.A.)

† Both authors contributed equally.
* Correspondence: kristian@syntbio.net; Tel.:+49-521-106-6323

Received: date; Accepted: date; Published: date

Abstract: Recombinant adeno-associated viruses (rAAV) provide outstanding options for customization and superior capabilities for gene therapy. To access their full potential, facile genetic manipulation is pivotal, including capsid loop modifications. Therefore, we assessed capsid tolerance to modifications of the VP proteins in terms of stability and plasticity. Flexible glycine-serine linkers of increasing sizes were, at the genetic level, introduced into the 587 loop region of the VP proteins of serotype 2, the best studied AAV representative. Analyses of biological function and thermal stability with respect to genome release of viral particles revealed structural plasticity. In addition, insertion of the 29 kDa enzyme β-lactamase into the loop region was tested with a complete or a mosaic modification setting. For the mosaic approach, investigation of VP2 trans expression revealed that a Kozak sequence was required to prevent leaky scanning. Surprisingly, even the full capsid modification with β-lactamase allowed for the assembly of capsids with a concomitant increase in size. Enzyme activity assays revealed lactamase functionality for both rAAV variants, which demonstrates the structural robustness of this platform technology.

Keywords: Adeno-associated-virus; β-lactamase; inverted terminal repeat (ITR); loop modification; capsid stability

1. Introduction

Recombinant adeno-associated viruses (rAAV) are frequently used as a basic research tool and are emerging as therapeutic agents. For example, the US FDA recently approved voretigene neparvovec (Luxturna), which is based on rAAV serotype 2 (rAAV2) and delivers a gene to supplement biallelic RPE65 mutation-associated retinal dystrophy [1]. In addition, products based on AAV1 (Alipogene tiparvovec / Glybera) and very recently AAV9 (Onasemnogene abeparvovec / Zolgensma) obtained approval in Europe or the United States. The increasing impact of rAAV on gene therapy relies on a high safety profile resulting from the inability to replicate autonomously and on long-term target gene expression [2]. Wild-type AAVs have a non-enveloped, icosahedral capsid formed by 60 subunits of VP1, VP2 and VP3 proteins in an approximate molar ratio of 1:1:10 [3,4]. In the AAV wild-type setting, the single-stranded DNA genome of about 4.7 kb includes two main reading frames (Rep and Cap) and is flanked by inverted terminal-repeat (ITR) sequences, which provide the encapsidation signal. Genetic engineering enables decoupling of the capsid coding genes from the encapsidated DNA. In the recombinant setting, *rep* and *cap* genes are provided *in trans* on a RepCap plasmid whereas a transgene expression cassette, frequently named gene of interest (GOI), is provided between the ITRs on the ITR plasmid. AAV needs additional 'helper' functionality from other viruses for production, which is provided on a separate pHelper plasmid [5,6]. For this reason, a three-plasmid system is often used, wherein the pHelper delivers the essential adenoviral elements E2A, E4 and the non-coding RNA VA. As host, HEK293 cells provide further adenoviral elements (E1A, E1B) and allow for high-titer production[5,6]. Alternative versions are also used such as a two-plasmid system combining the genetic information of the adenoviral helper sequences with AAV serotype specific *rep* and *cap* genes [7].

For diverse applications, e.g. virus-directed enzyme prodrug therapy (VDEPT) [8], viral targeting of specific cells is desired. AAV serotypes differ in their tropism and thus provide a first choice to achieve target specificity [9]. Deeper control over the target tropism requires genetic intervention. For this purpose, directed randomization and selection or rational engineering have been applied. Chimeric rAAV capsids are composed of proteins, which originate from different serotypes, and are often identified by evolutionary methods [10]. On the rational side, N-terminal fusions to e.g. the VP2 protein in rAAV2 have been studied [11–13]. In these cases, the addition of larger proteins, e.g. GFP and DARPins was compatible with capsid assembly and targeting.

A further and early adopted rational approach, which is extended in this publication, is the integration of motifs in previously identified loop positions of the VP proteins. Two groups demonstrated that capsid formation and gene packaging are only slightly influenced by integration of peptide sequences in VP proteins at various residue positions [14,15]. Insertions in these positions has also been used for biorthogonal labelling of capsids [16,17]. Capsid accommodation capacity was shown for the integration of larger moieties such as the minimal F_c-binding motif Z34C (34 amino acids) into the 587 loop region [18]. Production of such a Z34C rAAV2 with subsequent binding of an antibody was shown and transduction of target cells was observed. For vaccination via viral particle display, peptides up to 35 or 31 amino acids were integrated in the 453 or 587 position, respectively [19]. To our knowledge, the largest reported insertion to date is the fluorescent protein mCherry, which was functionally included in variable region IV at the 453 position of VP1, and allowed for the production of mosaic particles [20].

Our aim was to expand the loop modification strategy in combination with a systematic analysis of the engineering capacity of rAAV. For the construction of viruses, we extended an existing plasmid toolbox for rAAV2 manipulation and production [13,21]. Insertion of peptides in capsid proteins was studied with regard to rAAV productivity and transduction capability. The impact of capsid protein modifications on thermal stability has, to our knowledge, not been investigated. Thus, we first tested rAAV stability with glycine-serine insertions of varying length at residue position 587. These experiments confirmed that larger insertions are tolerated and we opted to insert the enzyme β-lactamase. As the introduction of an entire protein could interfere with capsid assembly, we tested partial insertions only in VP2 proteins. This required adaptation of the plasmid system for the production of mosaic rAAVs exclusively bearing VP2 loop modifications. Resulting mosaic particles were found to tolerate the insertion of a full-length β-lactamase in VP2 proteins. Finally, we set up a

complete β-lactamase modification of all VP proteins. Production of these fully decorated rAAVs was possible and allowed for further characterization. In summary, our analyses demonstrated resilience of the virus to modifications at the genetic and protein level. We believe that the plasmid system combines facile genetic manipulation with a broad range of rAAV capsid engineering options.

2. Results

2.1. Modifications of ITR and RepCap plasmids are compatible with rAAV production.

Despite the availability of different ITR and RepCap plasmids for rAAV2 production, options for facile loop capsid modification are sparse. In this work, previously modularized versions of the RepCap and ITR plasmids were used, which are largely compatible with the BioBrick RFC[10] cloning strategy (SI Figure 1) [21]. The RepCap plasmid (pZMB0216_Rep_VP123_453_587wt_ p5tataless) was retained but the ITR plasmid was redesigned to reduce cloning steps and serve user expectations outside the synthetic biology community. The new ITR plasmid (pZMB0522_ITR_EXS_ CMV_mVenus_hGHpA) contains the viral ITRs as part of a pUC19-based backbone, provides restriction sites for insertions based on BioBrick RFC[10] flanking the fluorescent reporter mVenus, and serves as a final destination plasmid. Plasmid generations are described in SI Method 1, final constructs used during this work are given in Table 1 and selected plasmids of the plasmid toolbox are given in SI table 1.

For sequence analysis of the ITRs, which commonly defy Sanger cycle sequencing, we cleaved the ITR DNA with BsaHI, generating two halves that are amenable to standard sequencing protocols (SI Figure 2, SI Method 2). Our plasmid pZMB0522 carries one complete ITR (in this plasmid notation referred to as 5′-ITR) and an ITR shortened by 11 bp (Figure 1).

Table 1. Overview of plasmids submitted to Addgene including a short description of their features. Information on the cloning of the plasmids is available in the supplementary information.

Plasmid name with description	Length	Backbone
pZMB0522_ITR_EXS_CMV_mVenus_hGHpA AAV2 ITR flanking a CMV promoter expressing the fluorescent protein mVenus	4014 bp	pUC19
pZMB0216_Rep_VP123_453_587wt_p5tataless expression of VP1/2/3 of AAV2 with cloning ready 453 and 587 loop regions, arginines in 587 loop region are intact, p5 promoter at end of expression cassette	5455 bp	pSB1C3_001
pZMB0600_Rep_VP13_453_587wt_p5tataless expression of VP1 and VP3 with cloning ready 453 and 587 loop regions Arg in 587 loop region are intact	6455 bp	pSB1C3_001
pZMB0315_CMV_Kozak_VP2_453_587wtHis expression of VP2 with Kozak sequence to prevent leaky scanning and VP3 start knock out, cloning ready 453 and 587 loop regions, Arg in 587 intact, His-tag in 587 loop	4705 bp	pSB1C3_001

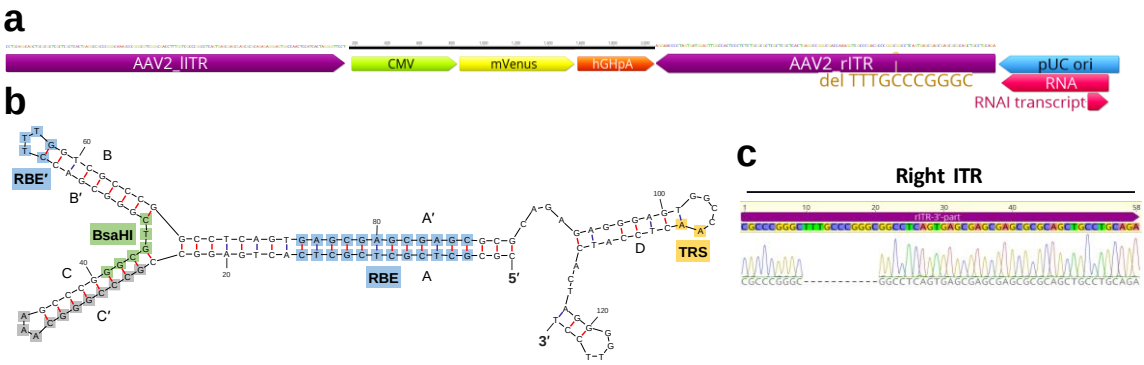


Figure 1. Overview of the ITR sequences of rAAV2. (a) Scheme of ITR plasmid components. The ITRs are enlarged compared to the interchangeable GOI (CMV promoter, mVenus fluorescent protein gene and hGHpolyA sequence). The deleted 11 bp sequence is highlighted in the right ITR adjacent to the pUC ori. (b) Structure of one ITR computed by Mfold webserver (150 mM NaCl, 5 mM $MaCl_2$, 37°C) [22]. Standard Sanger DNA-sequencing of ITRs was enabled by digestion with BsaHI whose recognition site is highlighted. (c) Results of Sanger-DNA sequencing aligned to original ITR sequences. Fragments were sequenced with oligonucleotides given in SI Method 2. The 3'-part of rITR shows a deletion of 11 bp (5'-TTTGCCCGGGC-3'). Sequencing results of the remaining ITR fragments are shown in SI Figure 2 c, d.

Production of DNaseI-resistant particles and thus functionality of the ITR plasmid (pZMB0522) in combination with either the RepCap plasmid (pZMB0216) or a commercially available counterpart (pAAV-RC, GenBank: AF369963.1) was assayed using small-scale transfections and quantitative real-time PCR (qPCR). Genomic titers in crude lysates with both plasmids showed no significant differences (Figure 2a). A larger preparation (pZMB0522, pZMB0216) was purified by precipitation [23] and imaged by transmission electron microscopy. Capsid diameter measurements resulted in an average of 23.7 ± 1.2 nm (SI Figure 3), which is in good agreement with the expected value of 25 nm [24]. Manually counting over 500 particles yielded a fraction of filled capsid between 60% and 80% (Figure 2b, SI Figure 3).

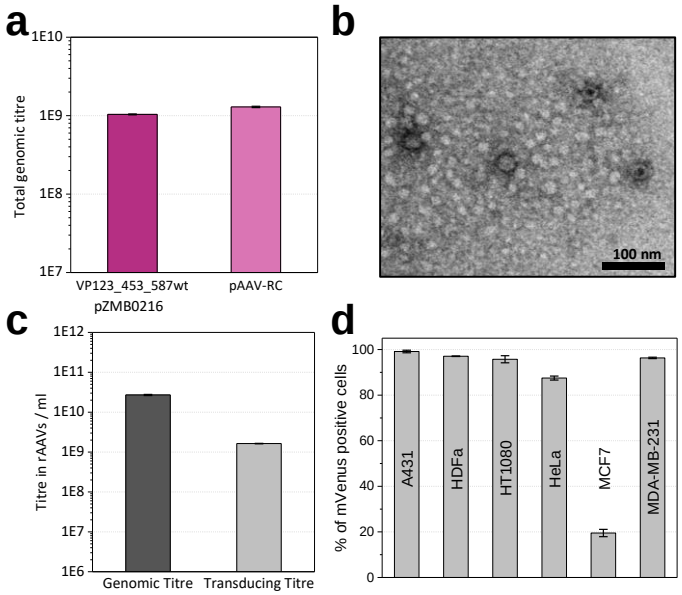


Figure 2. Production of rAAVs packaging the fluorescent reporter mVenus gene with plasmids pZMB0522 and pZMB0216. (a) Production yields of rAAV samples determined by qPCR from crude cell lysate from one 100 mm cell culture dish. Comparable values were obtained either using the new RepCap plasmid (left bar) or the commercial pAAV-RC plasmid (right bar). Standard deviations of three biological and two technical replicates were calculated for each sample type. (b) Transmission electron

microscopy image analysis of precipitation-purified viral samples at 39,000-fold magnification revealed viral particles with a size of about 24 nm. (c) Comparison of viral genomic and transducing titers. Genomic titers were determined via qPCR from three technical replicates of ultracentrifugation-purified viral samples. Transducing titers were analyzed by mVenus expression from biological and technical duplicates after flow cytometry analysis of HT1080 cells. (d) Transduction of different cell lines. Cells were incubated with a MOI of 10,000. Biological duplicates at two time points were analyzed by flow cytometry, measuring mVenus fluorescence and counting 10,000 events. Histograms of flow cytometry analysis can be found in SI Figure 4.

To obtain higher purity samples, all further preparations were purified by iodixanol gradient ultracentrifugation. For unmodified rAAV2 wt (pZMB0522, pZMB0216) genomic titers were determined by qPCR (Figure 2c). Using a comparable setup but depending on culture conditions and transfection efficiencies, titers between 1×10^{10} to 1×10^{12} vg/ml have been obtained.

Functionality of gene delivery was investigated with transduction assays. As common for AAV2, HT1080 cells were used, which express high levels of the rAAV2 primary receptor heparan sulfate proteoglycane (HSPG). Successful transduction was detected by the expression of the delivered transgene mVenus using flow cytometry. Based on a dilution series, the transducing titer was calculated. Genomes to infectious units yielded a specific infectivity of 16:1 (Figure 2c). This is in agreement with previous values, since for wild-type AAV2 a ratio of 1:1 and for rAAV2 ratios between 55:1 and 124:1 have been observed [25]. In addition, also transduction of the cancer cell lines A431, HeLa, MCF7, MDA-MB-231 and normal adult human dermal fibroblasts (HDFa) was tested with a multiplicity of infection (MOI) of 10,000. Flow cytometry analysis (Figure 2d) revealed that the rAAV2 wt preparation was able to transduce a variety of different cells with high efficiencies. In agreement with previous reports, only the breast cancer cell line MCF7 showed low transduction [26]. These results demonstrated production and function of the rAAV plasmid system and provided the basis for further investigations regarding the tolerance of the viral capsid to insertions in the 587 loop region.

2.2. Systematic variation of loop modifications shows a complex pattern of stability and transduction efficiency.

Previous experiments showed that, VP proteins tolerate peptide insertions at residue positions 453 and 587 [14,15]. We aimed at systematically analyzing the impact of increasing insertion length on rAAV thermal stability and biological function. Since a protein had already been inserted in the 453 loop region [20] and the 587 loop region is even more frequently used for modification, we opted to integrate linkers with increasing size at residue position 587 and investigated production and function of the resulting rAAVs. Glycine-serine linkers were chosen because of their flexibility and solubility, resulting in incremental changes correlating with length. Specifically, amino-acid linkers with the sequences GG, GGSG, $(GGSG)_2$ and $(GGSG)_4$ were integrated at the genetic level in all VP proteins to yield homogeneously decorated viral particles as presented in a schematic overview in Figure 3a.

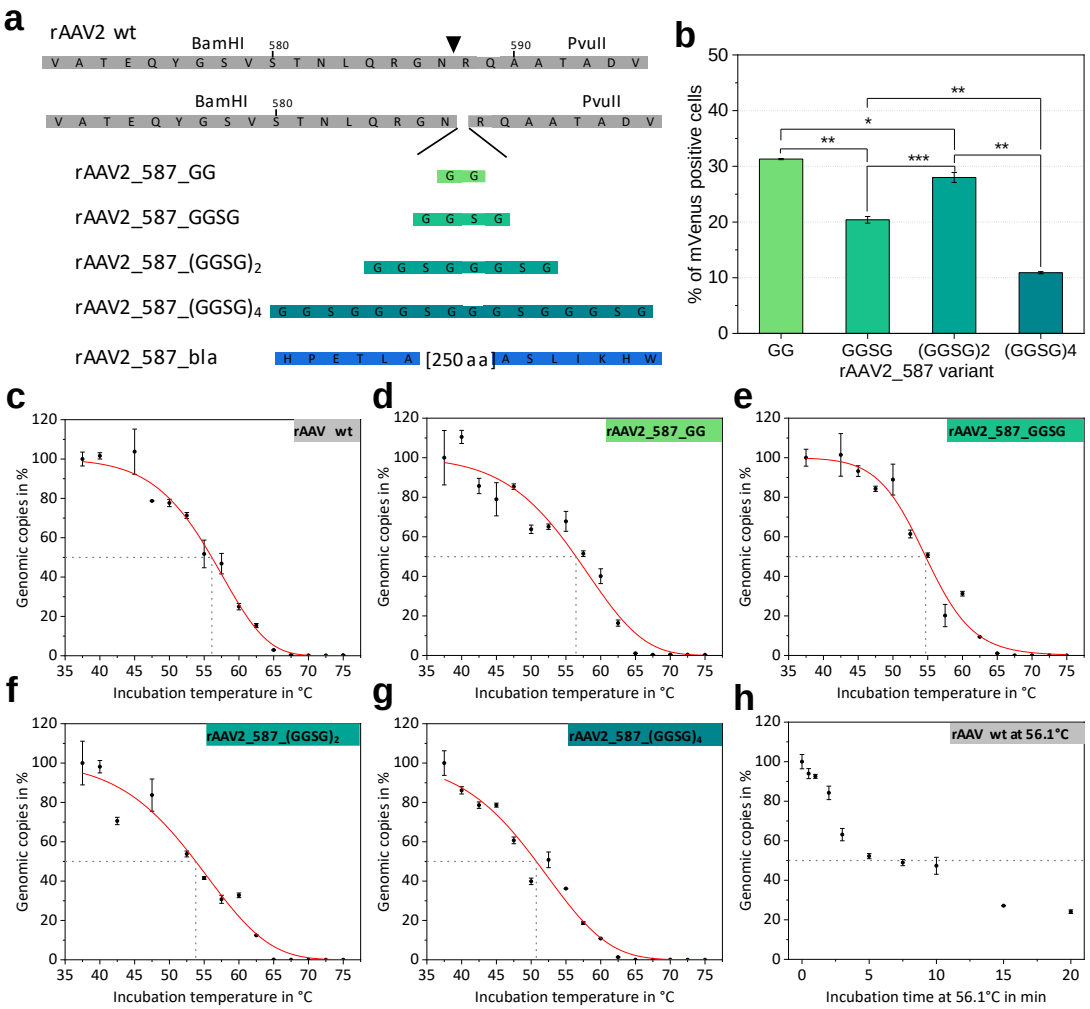


Figure 3. Comparison of different 587 loop variants and results of thermal stability assays. (a) Schematic overview on the coding sequence of the rAAV wt VP sequence. Restrictions site positions enabling the introduction of motifs flanking the 587 position (triangle) are indicated and amino-acid sequences of inserted linkers are shown. For β-lactamase only the first and the last amino acids are given. (b) Transduction efficiencies of different serine-glycine linker rAAV variants. Histograms of flow cytometry analysis can be found in SI Figure 7. Statistical analysis was performed as described in the Material and Methods section. (c-g) Thermal stability assays of rAAV particles measured in PBS. The percentage of qPCR detected genomic copies is plotted against the incubation temperature. Each point represents a technical duplicate with the standard deviation. Fitted curves (red) were calculated using a logistic function to determine the disintegration temperature for all rAAV variants. Table 2 lists the T_d values. (h) Disintegration kinetics of rAAV2 wt samples incubated at T_d = 56.1°C for different time periods. A dashed line indicates the point where 50% of genomic copies were released from the capsid.

Cloning was facilitated by the unique restriction sites of RepCap plasmid pZMB0216 flanking the loop region (SI Figure 1). Genomic titers of iodixanol-purified preparations demonstrated that all genetic constructs lead to rAAV production and that integration of flexible linkers only affected the titer for large integrations (Table 2). rAAV transduction ability was determined based on mVenus expression in HT1080 cells incubated with a MOI of 50,000 viral genomes. Increasing linker length impeded transduction (Table 2), but not in a linear fashion. The integration of only two amino acids resulted in a decrease of transduction ability from 97% (rAAV2 wt) to 31% (rAAV_587_GG) in the given experiment. Interestingly, increasing the linker length to four amino acids (GGSG) resulted in a further decrease of transduction to 20%, but rAAV with an insertion of eight amino acids ($(GGSG)_2$) showed a relatively improved transduction of 28% followed by a further drop in transduction to 11% for 16 amino acids (Figure 3b).

Table 1. Overview of plasmids submitted to Addgene including a short description of their features. Information on the cloning of the plasmids is available in the supplementary information.

Sample	Titer in vg ml^{-1} [a]	Transduction ability in % [b]	$T_{d,\ 5\ min}$ in °C[c]
rAAV2 wt	3.1×10^{10}	96.7 ± 0.1	56.1 ± 0.5
rAAV2_587_GG	7.1×10^{9}	31.3 ± 0.1	56.4 ± 0.8
rAAV2_587_GGSG	4.0×10^{10}	20.4 ± 0.6	54.7 ± 0.5
rAAV2_587_(GGSG)2	2.4×10^{10}	28.0 ± 0.9	53.8 ± 0.8
rAAV2_587_(GGSG)4	4.7×10^{9}	10.9 ± 0.2	50.7 ± 0.7
rAAV2_587_bla	1.3×10^{10}	1.2 ± 0.1	55.6 ± 0.4
rAAV2_VP2_587_bla	6.3×10^{10}	57.0 ± 2	n.d.

[a] Genomic titers are given in viral genomes (vg) per ml as determined by qPCR. Each value corresponds to a production with 10× 10 cm cell culture dishes and a final purified volume of 0.5 ml.

[b] Transduction ability was assayed with flow cytometry of HT1080 cells using a MOI of 50,000 and is given as percentage of mVenus expressing cells. The error is based on biological triplicates from one viral preparation.

[c] Disintegration temperatures T_d were determined in the qPCR-based stability assay.

Thermal stability of AAV capsids is an interesting biological and biophysical parameter and various methods to assess capsid stability have been described, such as differential scanning calorimetry (DSC), differential scanning fluorimetry (DSF) and electron microscopy[27], all of which monitor capsid breakdown but do not detect DNA release. We propose that the point of DNA release during heat treatment is a biological relevant event to describe capsid stability and that rAAV capsid integrity can be monitored by the DNase accessibility of the encapsidated DNA. Consequently, we repurposed the standard assay for genomic copy number determination and incubated rAAV samples at different temperatures before the treatment with DNase I. Subsequent analysis via qPCR yielded genomic copies of the rAAV sample plotted against the incubation temperature (Figure 3c-g). The disintegration temperature after a five-minute incubation ($T_{d,\ 5\ min}$) was determined as the temperature at which 50% of rAAVs have released their DNA. The term disintegration temperature T_d was chosen to distinguish the value from the melting temperature (T_m) reported by other methods (e.g. DSF, DSC). The results for all glycine-serine linker variants are listed in Table 2. With an increase in linker size T_d decreases slightly, showing that the capsid is destabilized by large insertions, but also that despite significant structural intervention, stability is maintained at physiological temperatures.

To estimate the influence of the incubation time and the decay kinetics in our thermal release assay, rAAV2 wt was isothermally incubated at the previously determined T_d (56.1 °C) for different time points. The percentage of intact genomic copies decreased in a hyperbolic fashion albeit displaying different phases (Figure 3h).

2.3. Mosaic rAAVs with a 29 kDa β-lactamase at position 578 in VP2 require a Kozak consensus sequence.

To illuminate the possibly even larger insertion capacity of rAAV capsids, we chose the well-studied TEM β-lactamase [28] as a protein with a small distance between the N- and C-terminus, which approximately matches the distance of the β-hairpin residues in the variable region of the VP protein. The protein was inserted at the genetic level at residue position 587 of the VP proteins. Lactamase offers the possibility to easily measure enzyme activity. In order to avoid maximum interference with capsid assembly, β-lactamase was in this experiment incorporated only into the 587 loop of solely VP2.

The modular rAAV plasmid system, which we used as a starting point, had been deployed to produce N-terminal VP2 protein fusions and the respective modified rAAV particles [13]. A mutation in the RepCap plasmid eliminated the VP2 start codon (pZMB0600_Rep_VP13_453_587wt_p5tataless) and a fourth plasmid expressing VP2 and VP3 (here named shortly CMV_VP23 plasmid) was provided *in trans*. The regulatory mechanisms for expression of the three VP proteins is complex [29,30]. Briefly, splicing results in two mRNA transcripts that code for VP1 or for VP2/3, respectively. VP2 and VP3 are coded by the same mRNA and expression is controlled by a leaky-scanning mechanism at the VP2 start, thus the VP3 sequence is always part of a VP2 gene. For the previous modifications leaky scanning did

not pose a problem, because they were located at the unique N-terminus (SI Figure 1c). However, since the 587 loop region lies within the coding sequence of both VP2 and VP3 proteins, loop modifications desired for VP2 only require suppression of concomitant expression of a likewise modified VP3.
In order to understand and insure the sole expression of VP2, we conducted an expression and mutation analysis starting with the RepCap and the CMV_VP23 plasmid. Transient transfection of HEK293 cells with this plasmid and subsequent Western blot analysis of VP expression was performed. Expression of all three VP proteins after transfection with the unmodified RepCap plasmid showed approximately the expected molar ratio between the three VP proteins (1:1:10) (Figure 4a, lane 1). A strong expression of both VP2 and VP3 proteins was observed for the CMV_VP23 plasmid (Figure 4a, lane 2). Apparently the leaky scanning mechanism is still active in the context of the CMV promoter and the cloning context (iGEM RFC[10]). To prevent undesired VP3 expression, a new plasmid abbreviated as CMV_VP2 was constructed, in which the VP3 start codon was removed by an exchange from ATG to ATC (Ile) (Figure 4b). As seen in the third lane of Figure 4a, expression of VP3 is unexpectedly still observed, which might be due to a second start codon located 24 bp downstream. To suppress leaky scanning, a strong Kozak sequence (GCC ACC) was introduced upstream of the VP2 start codon resulting in plasmid CMV_Kozak_VP2 (pZMB0315). Finally, solely the expression of VP2 (Figure 4a, lane 4) was detected with an expected increase in chemiluminescence intensity, indicating a higher level of expression.

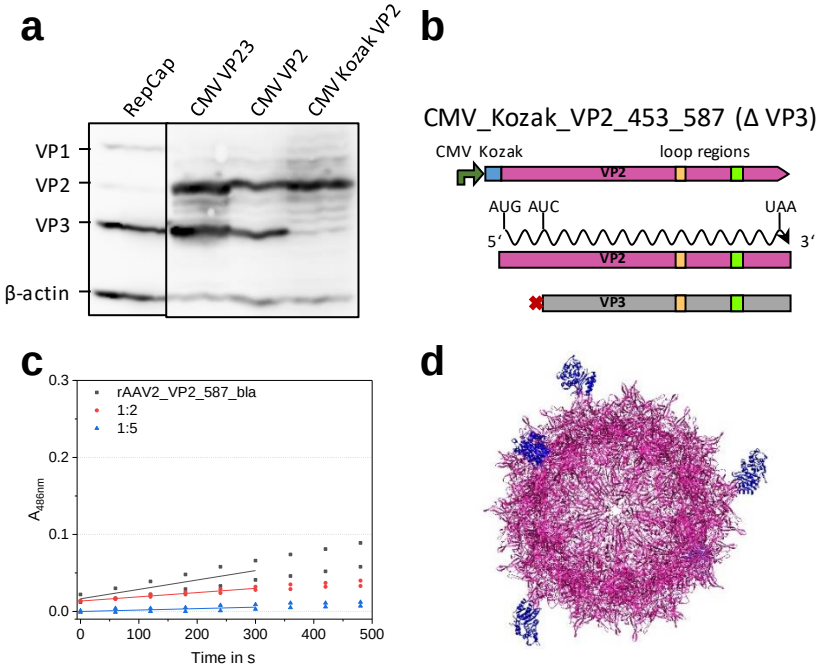


Figure 4. Establishment of mosaic 587 rAAV production using a modified four-plasmid system. (a) Analysis of *cap* protein expression after transient transfection in Western blots of crude cell lysates. Transfection of pZM0216_RepCap showed the expected ratio of 1:1:10 for VP1, VP2 and VP3 (lane 1). Three plasmid versions for VP2 expression were analyzed containing i) the VP2 and VP3 cassette (lane 2, CMV_VP23, pZMB0160), ii) the cassette with VP3 start codon removal (lane 3, CMV_VP2, pZMB0298), and iii) the cassette with an upstream Kozak sequence and VP3 start knock out (lane 4, CMV_Kozak_VP2 (pZMB0315). Expression of VP3 after removal of the start codon can be explained with a second start codon 24 bp downstream and persistent leaky scanning. SI Figure 5 shows full length images. (b) Scheme of the final expression construct. (c) Nitrocefin assay probing rAAV2_VP2_587_bla mosaic viral particles in different concentrations for lactamase activity. Slopes of linear regressions of the first 300 s were used to calculate the concentration of active β-lactamases. (d) Theoretical structure of the mosaic particle (UCSF Chimera [31]) composed of the AAV2 wt structure (purple, PDB: 1LP3) and five copies of β-lactamase (blue, PDB: 3DTM).

The gene of the stabilized β-lactamase variant 14FM was cloned into plasmid CMV_Kozak_VP2. The resulting plasmid (pZMB0577_pSB1C3_001_CMV_Kozak_VP2 _453_587wtbla) was used for rAAV

production in combination with Rep_VP13 plasmid (pZMB0600), the mVenus bearing ITR plasmid (pZMB0522) and pHelper. The molar ratio of these plasmids, particularly of the VP expressing plasmids, needed to be controlled in order to provide the right amounts of each VP protein (1:1:10 for VP1:VP2:VP3) for desired capsid assembly and relative amount of modified VP protein. The AAV promoter p40 is weaker compared to CMV and the Kozak sequence additionally enhances expression. Two molar plasmid ratios of 5:5:1:4 and 5:5:4:1 (pHelper : ITR : Rep_VP13 : CMV_VP2_587_bla) were tested for protein expression and production. The ratio of 5:5:1:4 (higher amount of modified plasmid) was associated with a high proportion of VP2 protein (SI Figure 6), which potentially leads to a higher portion of modified VP2 in the assembled capsid compared to the wild-type ratio. Next to Western blot, crude cell lysate samples from one 100 mm cell culture dish were analyzed regarding their genomic titers. The sample with the lower plasmid dose of the modified VP2_587_bla protein resulted in roughly three times higher amounts of viral particles (2.48×10^{10} vg/ml for a 1:4 VP13 : VP2_587_bla ratio compared to 7.98×10^{10} vg/ml for a 4:1 VP13 : VP2_587_bla (pZMB0600:pZMB0577)).

Based on the higher particle yield, mosaic rAAV2-VP2_587_bla particles were subsequently produced using the 5:5:4:1 ratio. A genomic titer of iodixanol-purified particles of 6.3×10^{9} vg was obtained per 100 mm dish. Thus, production using the four-plasmid system yielded rAAVs in the same range as the triple transfection production (Table 2). Incubation of HT1080 cells with a MOI of 50,000 resulted in about 57 ± 2% mVenus-positive cells (Table 2, SI Figure 7h). Comparison with transduction values of rAAV2 wt shows a reduction of transduction ability for the enzyme-bearing particles (SI Figure 7b). However, it should be noted that fully modified glycine-serine linker variants showed a much stronger reduction of the transduction ability.

We were interested if the enzymes presented on the capsid surface retained activity and thus, a colorimetric nitrocefin assay was performed. Enzymatic activity was evaluated from the linear correlation of absorbance against incubation time (Figure 4c). From this data we were able to estimate the number of active β-lactamases on the capsid surface. The β-lactamase variant 14FM which we used is a semi-rational combination of mutations described in literature (Hecky, Baumann unpublished data)[32]. Characteristics of this enzyme are presented in SI Table 2. Combining the genomic copy number and the known turnover number of the free lactamase k_{cat} allows for the estimation of the total number of active lactamases in the sample. In this experiment a β-lactamase concentration of 1.73×10^{-11} mol l^{-1} was calculated. This value is equivalent to 5.6 enzymes per DNaseI-resistant particle. Such particles with five β-lactamases might look as illustrated in Figure 4d. In conclusion, mosaic rAAV2s (rAAV2-VP2_587_bla) with an incorporated full-length protein were produced with unaltered efficiency - thus capsid assembly was not posing a problem. Functionality for β-lactamase was proven for enzymes presented on the viral particle surface.

2.4. Fully-lactamase decorated rAAV capsids can be produced and show enzyme activity.

As a limit for the integration of motives into the capsid was not found in the mosaic experiment, we aimed at the production of rAAV with all VP proteins modified with a lactamase named rAAV2_587_bla. The β-lactamase gene was thus cloned in the RepCap plasmid (pZMB0216) at the 587 position (yielding pZMB0221_Rep_VP123_453_587bla_p5tataless). A fully modified capsid model is depicted in Figure 5a. The genomic titer after ultracentrifugation was comparable to those of glycine-serine linker insertions (Table 2). To our surprise, the insertion with the size of approximately 29 kDa in every VP protein did not abrogate capsid assembly.

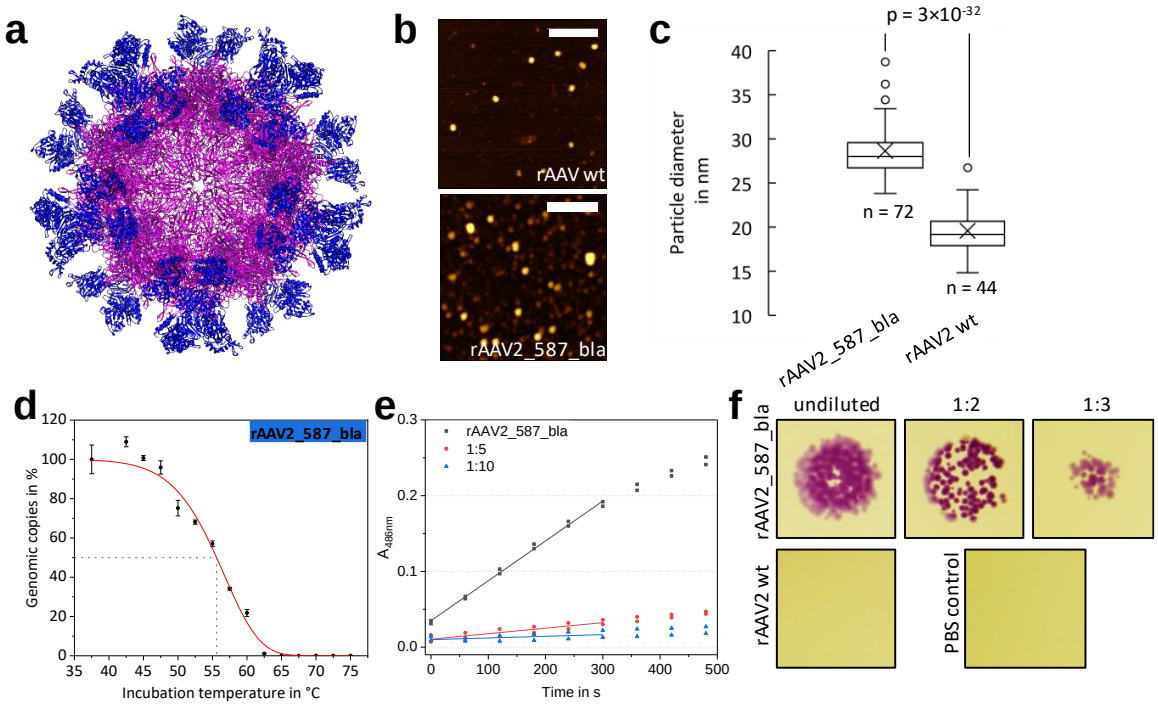


Figure 5. Comparison of rAAV2 wt and rAAV2_587_bla with respect to structure and β-lactamase activity. (a) Theoretical structure of rAAV2_587_bla constructed from the AAV2 wt structure (purple, PDB: 1LP3) and 60 copies of β-lactamase (blue, PDB: 3DTM). (b, c) AFM micrographs and particle size analysis of rAAV2 wt and rAAV2_587_bla. The calculated mean diameter of rAAV2 wt is 20 nm and 29 nm for rAAV_587_bla. SI Figure 8 gives AFM raw data. (d) Results of thermal stability assay based on qPCR data for rAAV_587_bla. Each point represents the standard deviation of a technical duplicate. $T_{d,5\ min}$ was calculated with 55.6 ± 0.4 °C. **(e)** Nitrocefin assay, probing the β-lactamase activity of rAAV2_587_bla in different dilutions. (f) Images of a bacterial growth assay on LB Agar plates for β-lactamase activity. Samples with rAAV2_587_bla show a β-lactamase activity up to 1:3 dilutions, whereas for rAAV2 wt and PBS no colony growth was observed.

We were interested to see how the 60 insertions would affect the overall size of the capsid. Measuring the particle width by atomic force microscopy (Figure 5b) a diameter of approx. 20 nm for rAAV2 wt was found. Note that the 5 nm deviation in particle size compared to the TEM measurements can be attributed to the type of method used and sample preparation. In contrast, a significantly higher viral particle diameter of approx. 29 nm was found for rAAV2_587_bla (Figure 5d). From the capsid model in Figure 5a a diameter of 35 nm was estimated for a fully enzyme-modified rAAV and estimations for the wild-type capsid resulted in an average of 25 nm, which agrees with the observation.

When rAAV2_587_bla samples were assayed for transduction ability on HT1080 cells, almost no transduction was measurable (Table 2). The thermal stability assay revealed a disintegration temperature Td for rAAV2_587_bla of 55.6 °C. Surprisingly, this value corresponds approximately to that of rAAV2 wt. (Figure 5d, Table 2).

Besides testing the effect of the enzyme decoration on capsid integrity, functionality of the integrated β-lactamases was evaluated. From the nitrocefin assay a β-lactamase concentration of 1.65×10-10 mol l^{-1} was calculated, equivalent to 26.9 lactamases per DNaseI-resistant particle (Figure 5e). Looking from a different angle and assuming all 60 β-lactamase are active, yields a catalytic rate of 20,056.2 s^{-1} per capsid and a k_{cat} value of 334.3 s^{-1} for each β-lactamase. The presence of active enzyme was furthermore studied in a simplistic bacterial growth assay. *E. coli* cells lacking the corresponding resistance gene are not able to survive on agar plates supplemented with ampicillin. The presence of functional β-lactamase, however, leads to degradation of the antibiotic and thus bacterial growth is possible. Mixing samples of rAAV2 wt with *E. coli* does not allow for bacterial growth, whereas for

samples of rAAV2_587_bla growth of bacterial colonies was observed (Figure 5f) - again indicating the presence of active enzyme in the viral particle preparation.

3. Discussion

At the outset of the project, which relies on modular RepCap and ITR plasmids that are largely compatible with the BioBrick cloning standard, we first constructed an ITR plasmid. In this context we reconsidered the very old problem, that the ITR sequences in plasmids are not stable during plasmid propagation in E. coli and that specifically one ITR is prone to deletions despite the symmetry of their sequences [33]. To our knowledge no explanation has been put forward. We hypothesize that the mechanism of plasmid replication of the prevalently used pMB1-derived origin of replication, which works unidirectional with respect to RNA II elongation and result in a switch between PolI and PolIII about 200 bp downstream of the replication start contributes to this genetic instability (SI Figure 2) [34]. In several ITR plasmids including our version, the distance between ori-start (end of RNAII) and the ITR with the more frequent deletion is about 74 bp. PolI and the polymerase switch, which might take place around the observed deletion site, might be more prone to induce deletions compared to the PolIII mediated replication. Since the most often observed deletion of 11 bp is also present in commercial and Addgene plasmids (e.g. pAAV_MCS [35]) and since it has been reported that even larger ITR deletions reduce production but increase transgene expression [36], we maintained the deletion in one ITR and placed it strategically next to the ori. Due to the pUC numbering scheme and expected GOI orientation of our plasmid, the deletion is located in the right or 3´ ITR, respectively (Figure 2e). Future experiments with various ITR to ori distances with high throughput ITR sequencing will test our hypothesis.

We were pleased to see that production of rAAV2 wt with the modified RepCap plasmid and the ITR plasmid (Table 2) enabled production in useful quantity and quality. Previous work has shown that insertion of motifs at various sites of VP proteins is possible [37,38]. We investigated glycine-serine linkers incorporated at residue position 587 by assaying capsid assembly, functionality and thermal stability. Production yields equal to that of rAAV wt showed compatibility with capsid assembly. Regarding transduction, we observed a significant impact of increasing insertion size with a non-linear decrease. The insertion of eight amino acids $(GGSG)_2$ showed a significantly higher transduction compared to four amino acids (GGSG). Variable region VIII harbors residues R585 and R588, which mediate the primary interaction of AAV2 with the cell via HSPG [39,40]. Spatial separation of the two arginines is known to interfere with cell binding and internalization [41]. The insertion of small motifs presumably increases the tension within the β-hairpin and hence the distance between the arginines (Figure 6). Larger flexible insertions probably are compatible with the correct arginine positioning but with increasing size a shielding effect becomes more and more prevalent.

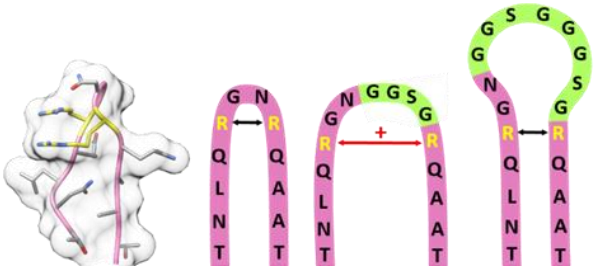


Figure 1. Model of the 587 loop region, from left to right: Structural model of the wild-type 587 loop with residue R585 and R588 highlighted in yellow (PDB: 1LP3); schematic model of 587 wild-type loop region (pink), 587 loop region with GGSG linker (light green) insertion leading to an increased arginine-arginine-distance; 587 loop region with $(GGSG)_2$ linker (light green) insertion leading to a regular arginine-arginine-distance but steric shielding of HSPG binding motif.

Next, we analyzed thermal stability using a qPCR-based method. We propose that the temperature of viral DNA release given by a disintegration temperature named T_d is a physical value that is on a par or even better corresponding to the biological activity of rAAV as values obtained by DSF or DSC experiments. It was hypothesized that viral particles transfer to a *metastable* state upon heating, which is defined by ejection of the encapsulated ssDNA [42,43]. Further temperature increases then completely

rupture the capsid. DSC and DSF are more likely to capture the complete disassembly of the capsid, and thus are expected to result in higher T_m values compared to the T_d values. This is in agreement with our data which yielded disintegration temperatures about 15 °C lower than the melting temperatures published so far [27,44].

As seen in the time course of isothermal disintegration (Figure 3h), the particles are not in thermal equilibrium at elevated temperatures discouraging thermodynamic interpretations. Thus, the T_d values obtained are sensitive to the incubation time, which is expected for the megadalton complex. Interestingly, Figure 2h reveals a multi-phase behavior, which could be either interpreted by different composition or structural states of the initial capsid ensemble or by two unfolding pathways. Longer incubation times (> 5 min) most likely result in even lower T_d values. Since most experimentally determined thermal stability changes within variants are minor, one can conclude that the capsid proteins form a stable framework highly resilient to sequence insertions.

Since peptide insertions of up to 16 amino acids were well tolerated, a β-lactamase was integrated at residue position 587. In one experiment β-lactamase was to be integrated only within the VP2 protein and not the VP3, which results from the same reading frame. An expression analysis of our existing CMV_VP2 plasmid revealed VP3 expression despite the CMV promoter for VP2 and even after deletion of the VP3 start codon. This can be explained by persistent leaky scanning and by a second AUG codon present 24 bp downstream of the mutated VP3 start. Mutating the first AUG of the VP3 protein was described before [45], but no Western blot was provided suggesting that a retained expression of VP3 could have occurred. That the VP2/3 gene retains expression of two proteins even if taken out of context and after a start codon removal indicates a deeply engrailed resilience to mutations and points to high robustness of the complex viral genome. We found that an additional strong Kozak sequence in front of VP3 resulted in the expression of solely modified VP2 [46].

Initial experiments of the mosaic lactamase approach indicated that the molar ratio (i.e. gene dose) of the transfected plasmids plays a crucial role for the amount of modified VP2 proteins in the final capsid. Optimizing the quadruple setting, mosaic viral particles displaying lactamases were obtained with titers only slightly lower than those of wild-type preparations. Transduction efficiency of mosaic β-lactamase particles was reduced, but not as dramatically as in case of the 16 amino acids insertion into the homogenously modified rAAV_587_$(GGSG)_4$. Most likely the lactamases pose a steric hindrance to the interaction with the cell, but the wild-type capsid proteins can still engage with HSPG. To our knowledge, this presents the first example of a successful enzyme incorporation at position 587 in the AAV capsid.

From a nitrocefin enzyme assay the number of active β-lactamases on the mosaic rAAV2_VP2_587bla capsid was estimated to be 5.6. Our calculations are based on the genomic titer and the proportion of empty capsids has not been taken into account. An alternative analysis of the full capsid titer using ELISA techniques is probably equally misleading as the modification of the capsid surface might interfere with antibody binding. Another uncertainty for the exact determination of the lactamase concentration lies within the assumption of a constant k_{cat} value, which was experimentally determined for the enzyme in solution. Due to immobilization on the capsids surface, the enzymes possess less degrees of freedom and substrate access is hindered, which might result in lower turnover numbers. Taken together, the true number of capsids is higher than estimated by genomic titer, because of empty capsids, but the true k_{cat} on the surface is lower than the one taken from solution measurements. It is conceivable that a heterogeneous mixture of active, partially active and inactive enzymes is presented, but due to the highly cooperative unfolding of lactamase this is less likely. Therefore, the deduced enzyme concentration should only be taken as an estimate. Still, the calculated number of 5.6 enzymes per viral capsid is very close to the expected number of modified VP2 proteins in the wild-type setting where five of the 60 capsid-forming proteins would be VP2. In combination, the data for a rAAV2 mCherry modification in the VP1 453 position [20] and for a lactamase modification in the VP2 587 position support the conclusion of a high structural plasticity of the AAV2 capsid.

In order to further explore the robustness of AAV assembly, the integration of β-lactamase was expanded to all VP proteins. We were intrigued to see that introduction of the enzyme into position 587 is possible while viral integrity and thermal stability are maintained. AFM measurements showed an

increase in capsid diameter of 10 nm, which corresponds well to our estimations. Genomic titers were in the same range compared to the one of rAAV2 wt, which is remarkable considering the size of the β-lactamase. This suggests that VP folding, core capsid assembly and DNA encapsidation are not significantly affected by bulky insertions. Like the originating β-lactamase variants [32], the enzyme variant used herein is known to fold well and to possess high thermodynamic stability (Hecky, Baumann unpublished data), which might contribute to the robustness of the modified VP protein and the corresponding capsid assembly. The high thermal stability of the inserted enzyme might even protect the VP proteins from unfolding. In contrast, glycine-serine linkers are intrinsically disordered. As expected, they were found to destabilize the VP proteins and the overall capsid. Enzyme activity of the sterically confined β-lactamases was measured. Calculations on the basis of the genomic titer and the turnover number show that approximately the equivalent of 26.9 active enzymes are presented. As discussed above, enzyme activity could be reduced and deduced numbers of active enzymes per intact capsid present only an estimation.

In summary we provide plasmids for facile genetic manipulation of AAV2 capsids and illuminate the intriguing resilience of AAV2 to a breadth of genetic and structural modifications.

4. Materials and Methods

4.1. Construction of plasmids.

All constructs were made by standard cloning techniques mainly using idempotent cloning strategies according to RFC[10] or RFC[25], respectively [47,48]. AAV plasmids listed in Table 1 and SI Table 1 were cloned as described in SI Method 1. ITR sequencing is described in SI Method 2. The CMV and hGHpA containing plasmids are from the iGEM parts registry (parts.igem.org). Resulting vectors were analyzed for their correctness by Sanger DNA-sequencing (Sequencing Core Facility, CeBiTec, Bielefeld, Germany).

4.2. Cell Culture.

HDFa (Thermo Fisher Scientific), HEK293, HeLa, HT1080 (DSMZ) cells were cultured in Dulbecco's Modified Eagle Medium supplemented 10% (v/v) fetal calf serum and 1% (v/v) penicillin/streptomycin (Sigma Aldrich). MCF7, A431 and MDA-MB-231 (DSMZ) were cultured in RPMI supplemented 10% (v/v) fetal calf serum and 1% (v/v) penicillin/streptomycin. Cells were maintained at 37 °C and 5% CO_2.

4.3. Viral particle production.

HEK293 cells were seeded at a density of 3×10^6 cells per 100 mm dish the day before transfection. A total amount of 15 µg DNA per 100 mm dish was transfected using calcium phosphate. RepCap plasmid, ITR-containing plasmid and pHelper plasmid were used in a 1:1:1 molar ratio [49]. For mosaic viral particles four plasmids were used in a 5:5:4:1 molar ratio of pHelper:ITR:RepCap_VP13:CMV_VP2. After 72 h of incubation at 37 °C, cells were harvested and pelleted by centrifugation (2000×g, 5 min).

4.4. Purification of viral particles.

Cells were resuspended in lysis buffer (50 mM Tris, 150 mM NaCl, 2 mM $MgCl_2$, pH 7.5) and viral particles were released from cells with three freeze-thaw cycles. Remaining DNA contamination was degraded by incubation with benzonase nuclease (final 100 U/ml, Sigma Aldrich) at 37 °C prior to addition of CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propane sulfonate, 0.5% w/v final). The crude lysate was cleared from cell debris by centrifugation (3,000×g, 10 min). This crude viral stock was further purified with a discontinuous iodixanol gradient[50]. Briefly, the lysate was transferred onto a gradient of 60%, 40%, 25% and 15% iodixanol in an open top polyallomer 16 x 76 mm tube (Science Services). Tubes were sealed and centrifuged in a T-880 rotor (Sorvall) at 340,000×g for 2 h at 18 °C. The rAAV containing fraction was collected with a 21G x 1 1/2" injection needle and the buffer was exchanged to 1× PBS (137 mM NaCl, 2.6 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.2) via Amicon Ultra-4 100K centrifugal filter units (Merck Millipore).

4.5. SDS-PAGE and Western blot analysis.

Cell pellets from rAAV production (1× 100 mm dish) were resuspended in 100 µl PBS and 5× SDS loading buffer. Samples were incubated at 95 °C for 10 min, centrifuged and 20 µl per lane were loaded

on a 10% SDS-polyacrylamide gel (Hoefer SE260). Samples were blotted onto a 0.45 µm nitrocellulose membrane (Thermo Fisher Scientific) using semi-dry electrophoretic transfer (V20-SDB, Sci Plas). After blocking the membrane with 10% (w/v) non-fat milk in TBS, the membrane was incubated simultaneously for 1.5 h with the B1 antibody (mouse monoclonal, supernatant, 1:100, Progen) and an anti β-Actin antibody (8H10D10, mouse monoclonal, 1:1000, Cell Signaling Technology). After incubation with an anti-mouse IgG, HRP-linked antibody (1:5000, Cell Signaling Technology), blots were imaged by luminescence detection (Pierce ECL Western Blot Substrate, Thermo Fisher Scientific).

4.6. Determination of genomic titers.

Before determination of genomic titers via qPCR, samples were treated with 10 U DNase I (New England Biolabs) in 10× DNaseI buffer in a final volume of 50 µl at 37 °C for 30 min before heat inactivation of the DNase I (75 °C, 20 min). Crude lysate samples were additionally incubated with 0.8 U Proteinase K (New England Biolabs) for 50 min at 37 °C before heat inactivation (95 °C, 10 min). Dilutions of the DNase I digest were used as template in the qPCR reaction. The sample was mixed with 2.5 µl primer qPCR-hGH-for (5'-CTCCCCAGTG CCTCTCCT-3') and 2.5 µl primer qPCR-hGH-rev (5´-ACTTGCCCCT TGCTCCATAC-3´), each at a stock concentration of 4 µM, and 10 µl of 2× GoTaq qPCR Mastermix (Promega). The qPCR reaction was carried out as described in the manual (TM318 6/14, Promega) with an increased time interval for the first denaturation step (95 °C, 10 min) using a LightCycler 480 II (Roche). The genomic titer was calculated from a standard curve of 10^2 to 10^7 copies of the ITR plasmid (pZMB0522) with an efficiency between 90-110% and an R value less than 0.1. Genomic titers in crude lysates were estimated from a standard curve mixed with the same amount of a non-transfected cell lysate.

4.7. Transducing titer assay.

10,000 cells per well were seeded in 500 µl of the corresponding media on a 12-well plate, settled for 1 h at 37 °C and then ultracentrifugation-purified rAAV samples were added. After 12 h incubation, 500 µl fresh medium was added. After further 72 h incubation, cells were detached with 0.25% Trypsin/EDTA, resuspended in PBS and analyzed using a FACSCalibur and counting 10,000 events. All experiments were performed as biological duplicates from two independent rAAV preparations.

4.8. Transmission electron microscopy.

Carbon-coated copper grids, 200 mesh (Electron Microscopy Science) were treated with oxygen plasma (Zepto, Diener electronic GmbH). After this, 3 µl of precipitation-purified rAAV sample [23] was applied to the grid and incubated for 2 min. Excess liquid was drained off, the grid was dried at room temperature and washed with three drops of distilled water. Negative staining was performed using 3 µl 2% (v/v) uranyl acetate replacement stain (Electron Microscopy Sciences) for 30 s. Excess liquid was drained off and grids were dried before channeling the sample into the microscope. rAAVs were visualized with a Philips CM100 (PW6021) instrument with an acceleration voltage of 80 kV. Images were analyzed using the Soft Imaging Viewer (Olympus) and ImageJ [51].

4.9. Atomic force microscopy.

AFM measurements of rAAV2 wt and rAAV2_587_bla were performed on a Multimode 8 AFM (Bruker) with Tap300Al-G cantilevers (BudgetSensors) in tapping mode in air. 2 µl of sample in PBS were spotted onto freshly cleaved mica and incubated for one minute. The mica was then briefly rinsed with water and dried under a gentle nitrogen flow. Data analysis was performed with Gwyddion 2.48. Obtained images were treated with offset and plane correction algorithms and the size of visualized particles was measured at half maximum particle height. Statistical analysis of size measurements was performed using Excel 2016.

4.10. AAV stability assay.

Thermal stability was analyzed using iodixanol-purified rAAV2 variants in two technical replicates. rAAV2 samples were diluted to 1.5×10^9 vg ml^{-1} with PBS. 10 µL aliquots of rAAV2 samples were incubated at temperatures ranging from 37.5 °C to 75.0 °C in 2.5 °C steps for 5 min in a thermocycler (peqSTAR 96 Universal Gradient, peqlab). To determine the disintegration time kinetics rAAV2 wt samples were incubated at 56.1 °C for different time points. Immediately after incubation,

samples were stored on ice. Genomic titers of each sample (10 µl) were determined by qPCR as described above. Due to the conceptual identity to the qPCR titration method, sensitivity and reliability are from this point on well established. A complete digest of the not encapsidated DNA was assured by additional tests. We evaluated the minimum titer of the starting material for a successful evaluation of vector particle stability. This needed to be higher than 1×10^9 genomic copies per ml in order to result in a data set that covers the sigmoidal-shaped disintegration curve. Our qPCR standard curve shows a linear correlation between 10^2 and 10^7 genomic copies per sample which represents the dynamic range. Note that our qPCR primers hybridize in the hGHpolyA region and that we detect completeness of this region. Probing of different parts or the whole genome is possible. Data analysis was performed using Origin2018. The total rAAV amount per reaction was normalized to 100% and a logistic regression $y = A_{min} + \frac{(A_{max}-A_{\min})}{\left(1+\left(\frac{x_0}{x}\right)^h\right)^s}$ was performed with A_{min} = 0, A_{max} = 100, no weighting). The disintegration temperature (T_d) was determined as the temperature at which 50% of rAAVs released its DNA.

4.11. β-lactamase activity assays.

Activity of β-lactamase presented on the rAAV capsid was determined using a nitrocefin and a bacterial assay. For the spectrophotometric nitrocefin assay 91 µl of rAAV sample containing either rAAV2_587_bla or rAAV2_VP2_587_bla (each 3.7×10^8 vg total) or PBS (negative control) were mixed with 9 µl of nitrocefin buffer (2 mM nitrocefin, 500 mM KH_2PO_4 and 5% (v/v) DMSO at pH 7.0) in a 96-well plate. Absorption at 486 nm was measured with a microplate spectrophotometer (PowerWave HT, BioTek) in 1 min intervals for 40 min at room temperature. Calculations are based on Lambert-Beer' rule as given in the formula: $\frac{\Delta A}{\Delta t} = \Delta\varepsilon_{486} \cdot c_{bla} \cdot k_{cat_{bla}}$ with $\frac{\Delta A}{\Delta t}$ = slope of linear regression from nitrocefin assay; $\Delta\varepsilon_{486}$ = molar extinction coefficient of nitrocefin shift (here 16000 l mol^{-1} cm^{-1}); c_{bla} = concentration of β-lactamase; $k_{cat_{bla}}$ = catalytic constant of β-lactamase; d = distance (here 0.267 cm). Either the turnover number (here 746 s^{-1}) or the concentration of active lactamases was assumed to be known.

In the bacterial assay an aliquot of about 200 µl of an E. coli DH5α culture transfected with plasmid pSB1C3_BBa_J04450 harboring a constitutively expressed chloramphenicol acetyltransferase and a red fluorescent protein was plated on LB agar supplemented with ampicillin (100 µg ml^{-1} final) and chloramphenicol (20 µg ml^{-1} final). rAAV samples (3 µl) were spotted (rAAV2_587_bla undiluted, 1:2, 1:3 and rAAV2 wt undiluted). The dish was incubated at 37 °C overnight. Due to lacking ampicillin resistance, growth and colony formation of E. coli cells only occur, when ampicillin is hydrolyzed by active β-lactamase of the rAAV sample. Growing E. coli colonies could be easily detected by eye via their red appearance.

4.12. Statistical analysis and reproducibility.

Standard deviation was calculated for all biological and technical replicates. To test whether differences were statistically significant based on a 0.05 significance level, data were checked for a normal distribution by a Shapiro-Wilk-test, then an independent Student's t-test was performed and p-values are given according to: * $p \le 0.05$; ** $p \le 0.01$; *** $p \le 0.001$. Variation of transduction efficiency determination was exemplarily tested in fully independent experiments with independent viral rAAV2 wt preparations by two authors. These data agree nicely with 95.7% (Figure 2d, SI Figure 4a) and 96.7% (Table 2, SI Figure 7b). Variation in the qPCR stability assay was tested with two independent viral preparations in two settings as shown in Figure 3c and 3h. In both assays 50% of genomic copies are detected after 5 min incubation.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/xxx/s1.

Author Contributions: RCF, JT and KET contributed to the design and generation of the plasmids. RCF, JT, KET and KMM conceived and designed the experiments. RCF and JT prepared the figures, wrote and edited the manuscript. KMM helped with writing and edited the manuscript. YH performed electron microscopy sample preparation and data acquisition. MTR generated the AFM images and performed data analysis. NB and DA contributed to AFM measurements. TB characterized and provided the β-lactamase for integration into the viral

capsid. SH, TB, KMA and KMM were former members of the iGEM team and provided information on design and generation of the plasmid system. KMM secured funding and supervised all experiments. All authors critically read and approved the final manuscript.

Funding: We acknowledge the financial support of the German Research Foundation (DFG) and the Open Access Publication Fund of Bielefeld University for the article processing charge.

Acknowledgments: We thank Philipp Borchert for experimental assistance. We acknowledge the iGEM team Freiburg 2010 for making their plasmid system available for further research and analysis. The team designed and cloned the plasmids, which form the basis of the presented work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

rAAV	Recombinant Aden-associated virus
TEM	Transmission electron Microscopy
AFM	Atomic force microscopy
bla	β-lactamase

References

1. FDA Advisory Committee *Spark Therapeutics Briefing Document*; 2017;

2. Clément, N.; Grieger, J.C. Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Mol. Ther. - Methods Clin. Dev.* **2016**, *3*, 16002.

3. Xie, Q.; Bu, W.; Bhatia, S.; Hare, J.; Somasundaram, T.; Azzi, A.; Chapman, M.S. The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 10405–10.

4. Rose, J.A.; Maizel, J. V; Inman, J.K.; Shatkin, A.J. Structural proteins of adenovirus-associated viruses. *J. Virol.* **1971**, *8*, 766–770.

5. Matsushita, T.; Elliger, S.; Elliger, C.; Podsakoff, G.; Villarreal, L.; Kurtzman, G.J.; Iwaki, Y.; Colosi, P. Adeno-associated virus vectors can be efficiently produced without helper virus. *Gene Ther.* **1998**, *5*, 938–45.

6. Xiao, X.; Li, J.; Samulski, R.J. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J. Virol.* **1998**, *72*, 2224–32.

7. Grimm, D.; Kay, M.A.; Kleinschmidt, J.A. Helper virus-free, optically controllable, and two-plasmid-based production of adeno-associated virus vectors of serotypes 1 to 6. *Mol. Ther.* **2003**, *7*, 839–50.

8. Feiner, R.C.; Müller, K.M. Recent progress in protein-protein interaction study for EGFR-targeted therapeutics. *Expert Rev. Proteomics* **2016**, *13*, 817–32.

9. Grimm, D.; Kay, M. a From virus evolution to vector revolution: use of naturally occurring serotypes of adeno-associated virus (AAV) as novel vectors for human gene therapy. *Curr. Gene Ther.* **2003**, *3*, 281–304.

10. Wu, Z.; Asokan, A.; Samulski, R.J. Adeno-associated virus serotypes: vector toolkit for human gene therapy. *Mol. Ther.* **2006**, *14*, 316–27.

11. Lux, K.; Goerlitz, N.; Schlemminger, S.; Perabo, L.; Goldnau, D.; Endell, J.; Leike, K.; Kofler, D.M.; Finke, S.; Hallek, M. Green Fluorescent Protein-Tagged Adeno-Associated Virus Particles Allow the Study of Cytosolic and Nuclear Trafficking. *J. Virol.* **2005**, *79*, 11776–11787.

12. Münch, R.C.; Janicki, H.; Völker, I.; Rasbach, A.; Hallek, M.; Büning, H.; Buchholz, C.J. Displaying high-affinity ligands on adeno-associated viral vectors enables tumor cell-specific and safe gene transfer. *Mol. Ther.* **2013**, *21*, 109–18.

13. Hagen, S.; Baumann, T.; Wagner, H.J.; Morath, V.; Kaufmann, B.; Fischer, A.; Bergmann, S.; Schindler, P.; Arndt, K.M.; Müller, K.M. Modular adeno-associated virus (rAAV) vectors used for cellular virus-directed

enzyme prodrug therapy. *Sci. Rep.* **2014**, *4*, 3759.

14. Girod; Ried; Wobus; Lahm; Leike; Kleinschmidt; Deléage; Hallek Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. *Nat. Med.* **1999**, *5*, 1438.

15. Shi, W.; Bartlett, J.S. RGD inclusion in VP3 provides adeno-associated virus type 2 (AAV2)-based vectors with a heparan sulfate-independent cell entry mechanism. *Mol. Ther.* **2003**, *7*, 515–25.

16. Liu, Y.; Fang, Y.; Zhou, Y.; Zandi, E.; Lee, C.-L.; Joo, K.-I.; Wang, P. Site-specific modification of adeno-associated viruses via a genetically engineered aldehyde tag. *Small* **2013**, *9*, 421–9.

17. Falck, G.; Müller, K.M. Enzyme-Based Labeling Strategies for Antibody–Drug Conjugates and Antibody Mimetics. *Antibodies* **2018**, *7*, 4.

18. Ried, M.U.; Girod, A.; Leike, K.; Büning, H.; Hallek, M. Adeno-associated virus capsids displaying immunoglobulin-binding domains permit antibody-mediated vector retargeting to specific cell surface receptors. *J. Virol.* **2002**, *76*, 4559–66.

19. Nieto, K.; Weghofer, M.; Sehr, P.; Ritter, M.; Sedlmeier, S.; Karanam, B.; Seitz, H.; Müller, M.; Kellner, M.; Hörer, M.; et al. Development of AAVLP(HPV16/31L2) particles as broadly protective HPV vaccine candidate. *PLoS One* **2012**, *7*, e39741.

20. Judd, J.; Wei, F.; Nguyen, P.Q.; Tartaglia, L.J.; Agbandje-McKenna, M.; Silberg, J.J.; Suh, J. Random Insertion of mCherry Into VP3 Domain of Adeno-associated Virus Yields Fluorescent Capsids With no Loss of Infectivity. *Mol. Ther. Nucleic Acids* **2012**, *1*, e54.

21. Freiburg iGEM Team Virus Construction Kit For Therapy Available online: http://2010.igem.org/Team:Freiburg_Bioware (accessed on Jan 24, 2019).

22. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **2003**, *31*, 3406–15.

23. Guo, P.; El-Gohary, Y.; Prasadan, K.; Shiota, C.; Xiao, X.; Wiersch, J.; Paredes, J.; Tulachan, S.; Gittes, G.K. Rapid and simplified purification of recombinant adeno-associated virus. *J. Virol. Methods* **2012**, *183*, 139–146.

24. Horowitz, E.D.; Rahman, K.S.; Bower, B.D.; Dismuke, D.J.; Falvo, M.R.; Griffith, J.D.; Harvey, S.C.; Asokan, A. Biophysical and ultrastructural characterization of adeno-associated virus capsid uncoating and genome release. *J. Virol.* **2013**, *87*, 2994–3002.

25. Zeltner, N.; Kohlbrenner, E.; Clément, N.; Weber, T.; Linden, R.M. Near-perfect infectivity of wild-type AAV as benchmark for infectivity of recombinant AAV vectors. *Gene Ther.* **2010**, *17*, 872–9.

26. Ellis, B.L.; Hirsch, M.L.; Barker, J.C.; Connelly, J.P.; Steininger, R.J.; Porteus, M.H. A survey of ex vivo/in vitro transduction efficiency of mammalian primary cells and cell lines with Nine natural adeno-associated virus (AAV1-9) and one engineered adeno-associated virus serotype. *Virol. J.* **2013**, *10*, 74.

27. Rayaprolu, V.; Kruse, S.; Kant, R.; Venkatakrishnan, B.; Movahed, N.; Brooke, D.; Lins, B.; Bennett, A.; Potter, T.; McKenna, R.; et al. Comparative Analysis of Adeno Associated Virus Capsid Stability and Dynamics. *J. Virol.* **2013**, *87*, 13150–13160.

28. Speck, J.; Hecky, J.; Tam, H.K.; Arndt, K.M.; Einsle, O.; Müller, K.M. Exploring the molecular linkage of protein stability traits for enzyme optimization by iterative truncation and evolution. *Biochemistry* **2012**, *51*, 4850–4867.

29. Trempe, J.P.; Carter, B.J. Alternate mRNA splicing is required for synthesis of adeno-associated virus VP1 capsid protein. *J. Virol.* **1988**, *62*, 3356–63.

30. Stutika, C.; Gogol-Döring, A.; Botschen, L.; Mietzsch, M.; Weger, S.; Feldkamp, M.; Chen, W.; Heilbronn, R. A Comprehensive RNA Sequencing Analysis of the Adeno-Associated Virus (AAV) Type 2 Transcriptome Reveals Novel AAV Transcripts, Splice Variants, and Derived Proteins. *J. Virol.* **2016**, *90*, 1278–89.

31. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF

Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

32. Hecky, J.; Müller, K.M. Structural perturbation and compensation by directed evolution at physiological temperature leads to thermostabilization of β-lactamase. *Biochemistry* **2005**, *44*, 12640–12654.

33. Samulski, R.J.; Srivastava, A.; Berns, K.I.; Muzyczka, N. Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. *Cell* **1983**, *33*, 135–43.

34. Troll, C.; Yoder, J.; Alexander, D.; Hernández, J.; Loh, Y.; Camps, M. The mutagenic footprint of low-fidelity Pol I ColE1 plasmid replication in E. coli reveals an extensive interplay between Pol I and Pol III. *Curr. Genet.* **2014**, *60*, 123–34.

35. CellBiolabs *pAAV-MCS expression vector VPK-410*; 2016;

36. Xie, J.; Mao, Q.; Tai, P.W.L.; He, R.; Ai, J.; Su, Q.; Zhu, Y.; Ma, H.; Li, J.; Gong, S.; et al. Short DNA Hairpins Compromise Recombinant Adeno-Associated Virus Genome Homogeneity. *Mol. Ther.* **2017**, *25*, 1–12.

37. Warrington, K.H.; Gorbatyuk, O.S.; Harrison, J.K.; Opie, S.R.; Zolotukhin, S.; Muzyczka, N. Adeno-associated virus type 2 VP2 capsid protein is nonessential and can tolerate large peptide insertions at its N terminus. *J. Virol.* **2004**, *78*, 6595–609.

38. Shi, W.; Arnold, G.S.; Bartlett, J.S. Insertional mutagenesis of the adeno-associated virus type 2 (AAV2) capsid gene and generation of AAV2 vectors targeted to alternative cell-surface receptors. *Hum. Gene Ther.* **2001**, *12*, 1697–1711.

39. Kern, A.; Schmidt, K.; Leder, C.; Müller, O.J.; Wobus, C.E.; Bettinger, K.; Von der Lieth, C.W.; King, J.A.; Kleinschmidt, J.A. Identification of a heparin-binding motif on adeno-associated virus type 2 capsids. *J. Virol.* **2003**, *77*, 11072–81.

40. Opie, S.; Jr, K.W.; Agbandje-, M. Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding. *J. Virol.* **2003**, *77*, 6995–7006.

41. Perabo, L.; Goldnau, D.; White, K.; Endell, J.; Boucas, J.; Humme, S.; Work, L.M.; Janicki, H.; Hallek, M.; Baker, A.H.; et al. Heparan sulfate proteoglycan binding properties of adeno-associated virus retargeting mutants and consequences for their in vivo tropism. *J. Virol.* **2006**, *80*, 7265–9.

42. Ros, C.; Baltzer, C.; Mani, B.; Kempf, C. Parvovirus uncoating in vitro reveals a mechanism of DNA release without capsid disassembly and striking differences in encapsidated DNA stability. *Virology* **2006**, *345*, 137–147.

43. Bernaud, J.; Rossi, A.; Fis, A.; Gardette, L.; Aillot, L.; Büning, H.; Castelnovo, M.; Salvetti, A.; Faivre-Moskalenko, C. Characterization of AAV vector particle stability at the single-capsid level. *J. Biol. Phys.* **2018**, *44*, 181–194.

44. Bennett, A.; Patel, S.; Mietzsch, M.; Jose, A.; Lins-Austin, B.; Yu, J.C.; Bothner, B.; McKenna, R.; Agbandje-McKenna, M. Thermal Stability as a Determinant of AAV Serotype Identity. *Mol. Ther. - Methods Clin. Dev.* **2017**, *6*, 171–182.

45. Zhang, H.; Xie, J.; Dmitriev, I.; Kashentseva, E.; Curiel, D.T.; Hsu, H.; Mountz, J.D. Addition of six-His-tagged peptide to the C terminus of adeno-associated virus VP3 does not affect viral tropism or production. *J. Virol.* **2002**, *76*, 12023–12031.

46. Kozak, M. Initiation of translation in prokaryotes and eukaryotes. *Gene* **1999**, *234*, 187–208.

47. Knight, T. Idempotent Vector Design for Standard Assembly of Biobricks Standard Biobrick Sequence Interface Available online: http://hdl.handle.net/1721.1/45138 (accessed on Jan 24, 2019).

48. Müller, K.M.; Arndt, K.M.; IGEM_Freiburg; Grünberg, R. Fusion Protein (Freiburg) Biobrick assembly standard Available online: http://hdl.handle.net/1721.1/45140 (accessed on Jan 24, 2019).

49. Agilent Technologies AAV Helper-Free System Instruction Manual Available online: https://www.agilent.com/cs/library/usermanuals/Public/240071.pdf (accessed on Aug 13, 2019).

50. Zolotukhin, S.; Byrne, B.J.; Mason, E.; Zolotukhin, I.; Potter, M.; Chesnut, K.; Summerford, C.; Samulski,

R.J.; Muzyczka, N. Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Ther.* **1999**, *6*, 973–85.

51. Schneider, C.A.; Rasband, W.S.; Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **2012**, *9*, 671–5.

1. Author 1, A.B. Title of Thesis. Level of Thesis, Degree-Granting University, Location of University, Date of Completion.
2. Title of Site. Available online: URL (accessed on Day Month Year).

10.1.1.2 **Supplementary information of "rAAV system evaluation"**

Supplementary information

rAAV engineering for capsid-protein enzyme insertions and mosaicism reveals resilience to mutational, structural and thermal perturbations

Rebecca C. Feiner[1†], Julian Teschner[1†], Kathrin E. Teschner[1], Marco T. Radukic[1], Tobias Baumann[2], Sven Hagen[3], Yvonne Hannappel[4], Niklas Biere[5], Dario Anselmetti[5], Katja M. Arndt[6], Kristian M. Müller[1*]

1 Methods

1.1 SI Method 1. Plasmid construction

1.1.1 Plasmids constructed in this work

pZMB0522_ITR_EXS_CMV_mVenus_hGHpA. This is the streamlined version of the rAAV ITR plasmid. To allow BioBrick compatible integration of expression casettes into the ITR plasmid, the region in-between the ITRs of vector pGolden-AAV (Addgene plasmid # 51424 was a gift from Yonglun Luo [47]) was changed and EcoRI, XbaI as prefix and SpeI as suffix sites were integrated by hybridized oligonucleotides through the unique BsmFI and AgeI sites. The backbone of this new generated RFC[10]-compatible ITR plasmid was digested with PstI and ligated with PCR-amplified pUC19 backbone (pUC19_PstI_for 5´-CTGCAGAAAA GGCCAGCAAA AGGC and pUC19_PstI_rev 5´-CTGCAGGCAC TTTTCGGGGA AATG) yielding pUC19_ITR_EXS. In the last step, the previously assembled BioBrick CMV_mVenus_hGHpA was cloned via EcoRI and SpeI into pUC19_ITR_EXS to generate pZMB0522_ITR_EXS_CMV_mVenus_hGHpA. This plasmid can be used to test GOI expression and at the same time allows integration of other genes. Note that XbaI and SpeI generate compatible cohesive ends. Facilitating new assemblies, a BioBrick CMV promoter (pZMB0143-CMV) and BioBrick hGHpA (pZMB0135_hGHpA) are also available.

pZMB0315_CMV_Kozak_VP2_453_587wtHis. This vector was constructed in two steps starting from pZMB0156_VP23_453_587wt. After RFC[10] integration of the CMV promoter, the VP3 initiation codon was mutated from ATG (Met) to ATC (Ile) using site-directed mutagenesis primers SDM_VP3ko_for 5´-CTAATACGAT CGCTACAGGC AGTGGC and SDM_VP3ko_rev 5´-CCTGTAGCGA TCGTATTAGT TCCCAGAC. In the second step, a strong Kozak sequence was introduced via PCR primers XbaI_CMV-for 5´-CTTCTAGAGC GATGTACGGG and CMV_Kozak_NgoMIV-rev 5´-ATAATGCCGG CCATGGTGGC CTAGTAATTT CGATAAGCCA GTAAG and recloning of the fragment. The HisTag in 587 position was integrated by hybridization of oligonucleotide as described in **SI Figure 11**.

Rep_VP123_453_587wtGG_p5tataless. This plasmid was constructed based on pZMB0216_Rep_VP123_453_587wt_p5tataless. A sequence coding for a Gly-Gly linker was integrated in position 587 by hybridization of oligonucleotides as described in **SI Figure 11** upon digestion with BamHI and PvuII. The cloning strategy for other GGSG linker constructs is similar to this one.

Rep_VP123_453_587wtbla_p5tataless. This plasmid was constructed based on iGEM plasmid BBa_K404250. Amongst others BBa_K404250 codes for a thermostabilised variant of β-lactamase which was amplified by PCR using BamHI_wt_bla-for 5´- AAAGGATCCG TATCTACCAA CCTCCAGAGA GGCAACCACC CAGAAACGCT GGCGAAAG and PvuII_wt_bla-rev 5´- AAACAGCTGT AGCTGCTTGT CTCCAATGCT TAATCAGTGA GGCACC primers. The BamHI and PvuII digested PCR product was inserted in 587 position of pZMB0216_Rep_VP123_453_587wt_p5tataless by standard cloning techniques.

1.1.2 Plasmids previously generated and deposited

pZMB0216_Rep_VP123_453_587wt_p5tataless. Base pair numbering given here refers to the start codon of the RepCap coding region. Starting with the pAAV-RC plasmid (GenBank: AF369963.1) silent mutations of two PstI restriction sites (nucleotide substitutions G177C and G3940C) in the Rep coding region were introduced by site-directed mutagenesis (QuikChange II Site-Directed Mutagenesis Kit, Agilent Technologies). Two EcoRI sites (substitutions A1449G and T1668C) and a PstI site (substitution C1641T) were removed by cloning a synthesized DNA fragment (GeneArt, Darmstadt, Germany) into rep via BstEII/SwaI. To enable restriction-based modification of the 453 or 587 loop coding sequences, the unique restriction sites SspI/SalI (G3202A, C3205T, C3206T; T3254A, C3255G, A3256T, A3257C, G3259A) and BamHI/PvuII (T3613A, T3616C; C3658A, A3661T) were introduced adjacent to the 453 and 587 loop of the viral capsid protein (VP), respectively. This was achieved by cloning a second synthesized DNA fragment (GeneArt) into the respective region via XcmI and BsiWI. BamHI (C729T) and SalI (C1110G) within the Rep coding region were removed by silent site-directed mutagenesis. The final RepCap construct was cloned in pSB1C3_001 via PCR using the primer Prefix_Rep68_78-ex (5´-GGAATTCGCG GCCGCTTCTA GATGGCGGGG TTTTACGAGA TTGTGATTAA G) and Suffix VP123ex RFC_25 5´-GCTACTAGTA TTAACCGGTG TAGTTAATGA TTAACCCGCC ATGCTACTTA TC) yielding pSB1C3_001_Rep_VP123_453_ 587wt. The promoter p5 TATA-less was converted into a BioBrick using p5_primer_for 5´-GCTCTAGAGG GAGGGGTGGA GTCGTGACGT G and p5_primer_rev 5´-TTCTGCAGCGG CCGCTACTAG TAGTTCAAAC CTCCCGCTTC AAAATGG and was cloned into pSB1C3_001_Rep_VP123_453_587wt via RFC[10] to generate the final construct pZMB0216_Rep_VP123_453_587wt_p5tataless.

pZMB0600_Rep_VP13_453_587ko_p5tataless. This plasmid was obtained by silent site-directed mutagenesis of the ACG initiation codon of VP2 in pZMB0216_Rep_VP123_453_587wt_p5tataless with the primers: VP2-ko_for 5´-GGTTGAGGAACCTGTTAAGACCGCTCCGGGAAAAAAGAGG and VP2-ko_rev 5´-CCTCTTTTTTCCCGGAGCGGTCTTAACAGGTTCCTCAACC to ACC.

pZMB0091_CMV_DARPinE01_mli_VP23 _453_587koHis. This vector is the product of two BioBrick cloning steps. First DARPinE01 with a GGSGGGSG linker sequence was cloned via RFC[25] into pZMB0156_VP23_453_587wt and then the CMV promoter from pZMB0143_CMV was added via RFC[10]. The HSPGko and the His-tag were integrated into the 587 loop region via hybridization of oligonucleotides as described in **SI Figure 11**.

pZMB0246_CMV_VP1up_NLS_mVenus_VP23_453_587koHis. This plasmid was generated by integrating the mVenus gene in pZMB0156_VP23_453_587wt fusing the VP1up_NLS part from pZMB0503_VP1up_NLS using the RFC[25] standard and adding the CMV promoter of pZMB0143_CMV. Finally, the 587 loop region was modified by hybridization of oligonucleotide coding for a His-Tag and the HSPGko as described in **SI Figure 11**.

pZMB0156_VP23_453_587wt. This vector is a BioBrick plasmid for subcloning purposes and was constructed by PCR amplification of the cap part from pZMB0216_Rep_VP123_453_587wt_p5tataless with the primers Prefix_VP2ex 5‘-ATGGCCGGCG CTCCGGGAAA AAAGAGGCCG and Suffix_VP123ex_RFC_25 5´-GCTACTAGT ATTAACCGGT GTAGTTAATG ATTAACCCGC CATGCTACTT ATC followed by cloning the product into pSB1C3_001 via RFC[10] to generate pZMB0156_VP23_453_587wt. This plasmid serves for further N-terminal modifications. Thus, the VP3 initiation codon was not mutated. R585 and R588 (587ko) was introduced by hybridization of oligonucleotide as described in **SI Figure 11**.

pZMB0503_VP1up_NLS. A fragment of VP1 named VP1up was converted to a RFC[25] Biobrick by PCR amplification from pAAV-RC using the primers Prefix_VP1ex 5´-ATGGCCGGCG CTGCCGATGG TTATCTTCCA G and VP1up_Suffix_rev 5´-GGACTAGTA TTAACCGGTC GCCTTAACAG GTTCCTCAAC CAGG and cloning into pSB1C3. The NLS was generated in RFC[25] standard with the oligonucleotides NLS_for 5´-AATTCGCGGC CGCTTCTAGA TGGCCGGCCC TGCAAGAAAA AGATTGAATA CCGGTTAATA CTAGTAGCGG CCGCTGCA and NLS_rev 5´-GCGGCCGCT ACTAGTATTA ACCGGTATTC AATCTTTTTC TTGCAGGGCC GGCCATCTAG AAGCGGCCGC G and cloned downstream of the VP1up sequence.

pZMB0143_CMV. The CMV promoter was kindly provided by the iGEM team Ljubljana 2007 (BBa_I712004). A sequence alignment of this BioBrick with the CMV promoter used in pAAV-MCS shows slight differences.

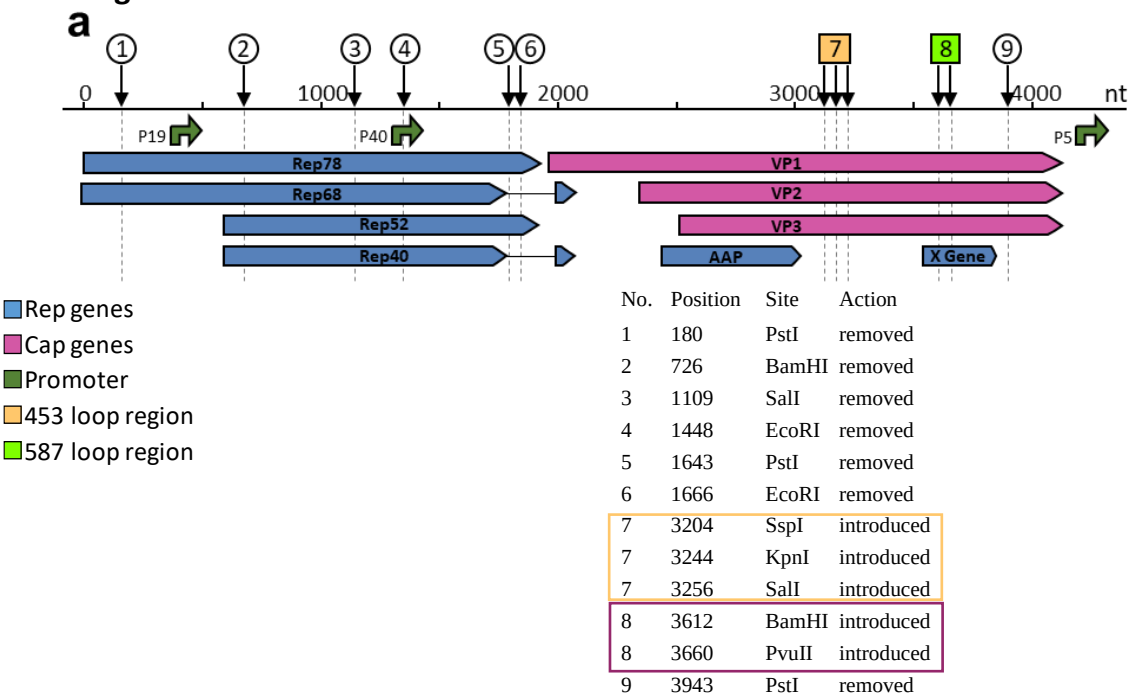
pZMB0135_hGHpA. The human growth hormone polyadenylation signal (hGHpA) was converted into RFC[10] by PCR using pAAV-MCS as template and hGH_primer_for 5´-GCTCTAGACG GGTGGCATCC CTGTGAC and hGH_primer_rev 5´-GAACTGCAGC GGCCGCTACT AGTAAGGACAG GGAAGGGAGC AG and cloning the fragment in a pSB1C3 backbone.

1.2 SI Method 2. Determination of ITR sequences using Sanger sequencing

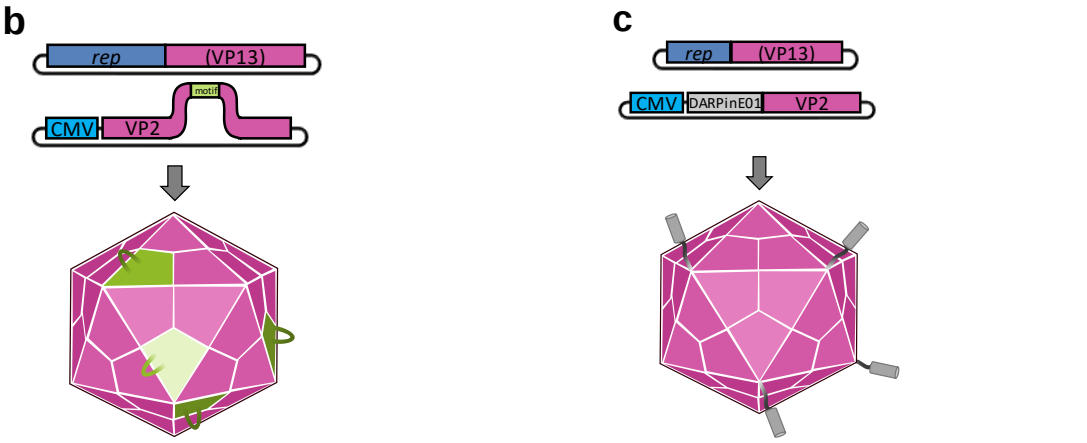
About 60 µg of ITR plasmid pZMB0522 were digested with 120 U MlyI in a total volume of 240 µl at 37 °C for 1.5 h and separated on an agarose gel. lITR and rITR fragments were isolated using NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL). About 18 µg of each fragments were digested with 60 U BsaHI in a total volume of 120 µl at 37 °C for 2 h to split the sequence forming the major stem loop of the ITR sequence in half. Resulting fragments were separated on an agarose gel and 5´- and 3´-parts of the lITR and rITR were isolated. Sufficient material for several sequencing reactions was recovered. The obtained four DNA fragments were analysed by Sanger DNA-sequencing (Sequencing Core Facility, CeBiTec, Bielefeld, Germany). The sequencing primers were SEQ-lITR-5: 5´-GAAATGTTGA ATACTCATAC TCTTCC, SEQ-lITR-3: 5´-ATGAACTAAT GACCCCGTAA TTG, SEQ-rITR-5: 5´-CCTAATCTCA GGTGATCTACC and SEQ-rITR-3: 5´-AACGCCTGGT ATCTTTATAG TCC. Results are shown in **SI Figure 2.**

2 Supplementary Information Figures

2.1 SI Figure 1

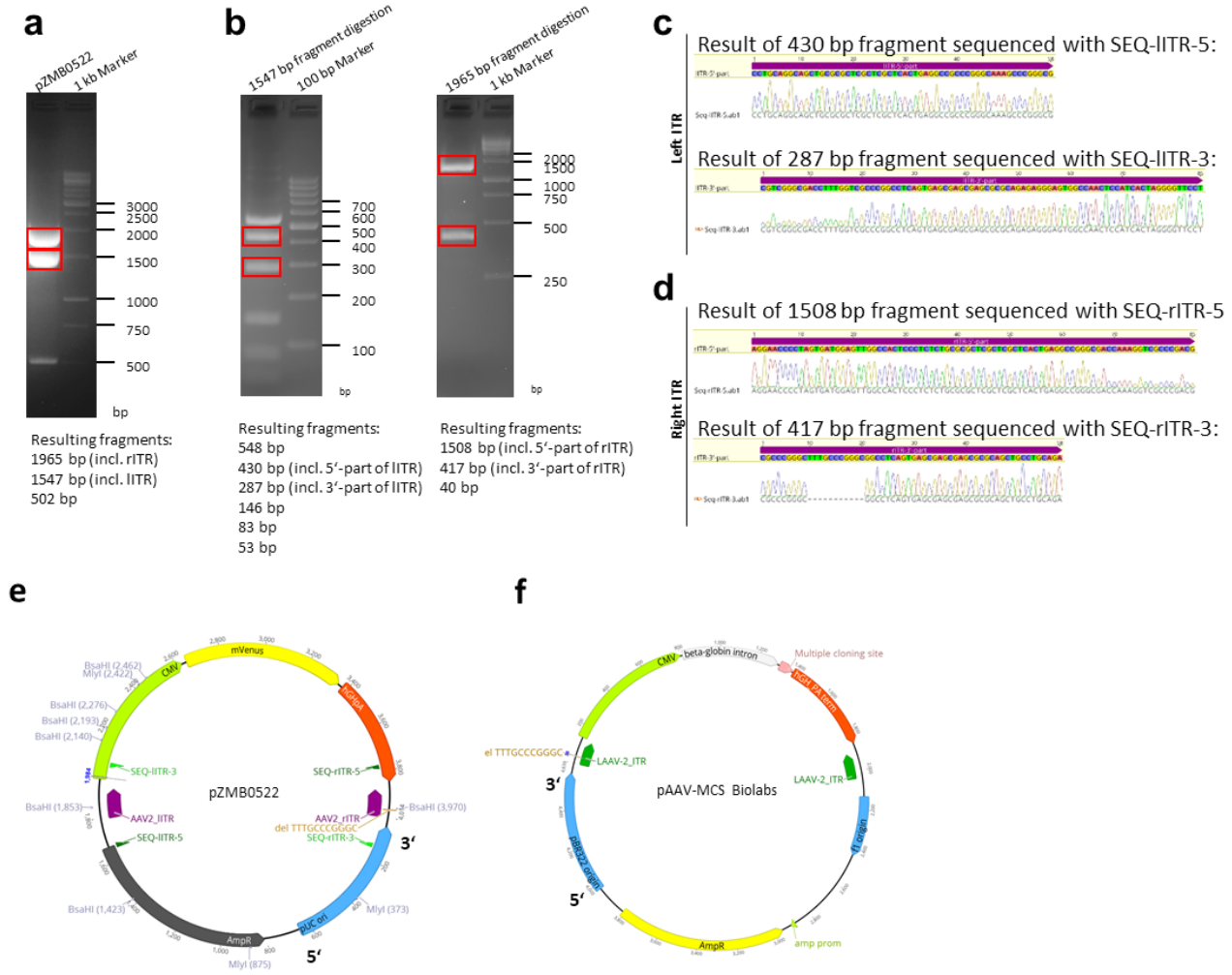


No.	Position	Site	Action
1	180	PstI	removed
2	726	BamHI	removed
3	1109	SalI	removed
4	1448	EcoRI	removed
5	1643	PstI	removed
6	1666	EcoRI	removed
7	3204	SspI	introduced
7	3244	KpnI	introduced
7	3256	SalI	introduced
8	3612	BamHI	introduced
8	3660	PvuII	introduced
9	3943	PstI	removed



SI Figure 1. Construction of the AAV2 RepCap plasmid (pZMB0216_Rep_VP123_453_587wt_p5tataless) and schematic overview of the rAAV plasmid production system. **a)** *Rep* and *cap* genes were integrated into the pSB1C3_001 backbone. Restriction sites not compatible with the RFC[10] standard were eliminated. Removal of restriction sites is marked with numbers and explained in the inset. For easy modification of 453 and 587 loop regions, the recognition sequences of SspI and SalI, as well as BamHI and PvuII were introduced. These allow for easy modification of the loop regions. Promoters are marked as green arrows. **b)** A forth plasmid allows for the production of mosaic rAAVs with e.g. modified 453 or 587 loop regions (green, e.g. His-tag) of the virion. Expressing a modified VP2 from a separate plasmid reduces the number of motifs on the capsid surface. **c)** Displaying proteins (grey, e.g. DARPinE01) on the capsid surface is also possible using a N-terminal fusion to VP2. **d)** Assembly of fusion proteins is carried out using the RFC[25] standard. The RFC[25] compatible multiple cloning site is visualized here.

2.2 SI Figure 2

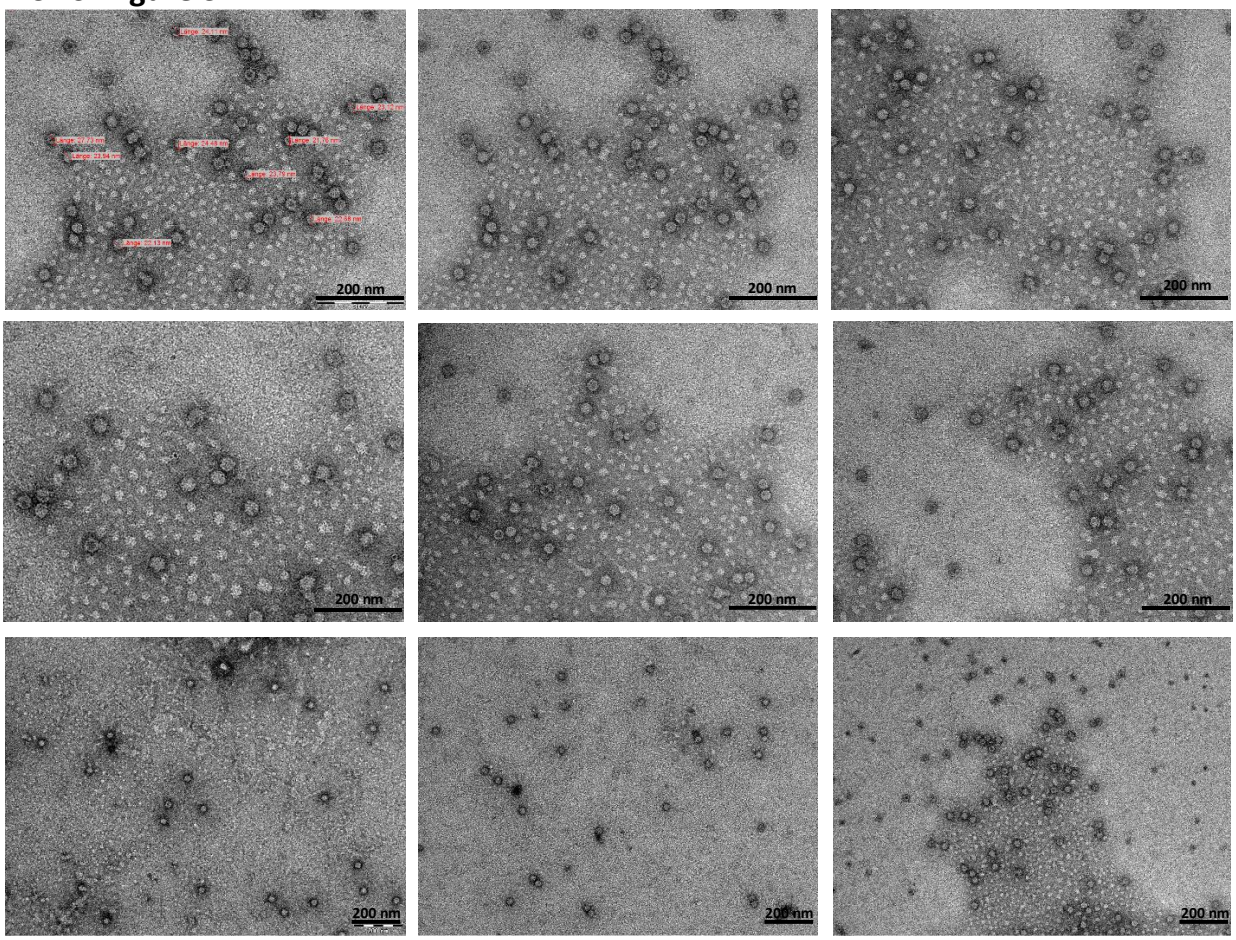


SI Figure 2: Preparation of fragments suitable for sequencing of ITRs**. a, b)** Agarose gel electrophoresis using a 1 % gel at a voltage of 120 V for 90 min. **a)** The restriction digest of pZMB0522 with MlyI results in three fragments and both ITRs are separated on an agarose gel as highlighted in red. **b)** The larger fragments are further digested with BsaHI to break up the T-shaped stem loop and to enable Sanger DNA-sequencing. **c, d)** Results of Sanger-DNA sequencing aligned to original ITR sequences. Fragments were sequenced with oligonucleotides given in the methods. **c)** The left ITR is completely intact. **d)** The 3'-part of rITR shows a deletion of 11 bp (5'-TTTGCCCGGGC-3'). **e)** Schematic overview of the ITR plasmid with used restriction sites. The 3' end of the ori is orientated towards the right ITR. **f)** Overview of the commercially available pAAV-MCS from Cell Biolabs with highlighted left and right ITR. The 3' end of the ori is oriented towards the left ITR which shows the 11 bp deletion. The distance between 3' end of the ori and the ITR with deletions is 74bp and 85 bp, respectively.

We hypothesize that the distance of an ITR to the plasmid origin, the type of origin and the occurrence of Okazaki fragment processing sites may affect genetic ITR stability. In particular we assume that the *E. coli* PolI itself and the switch to PolII, which takes place shortly downstream of pMB1 type ori, contribute to the genetic instability.

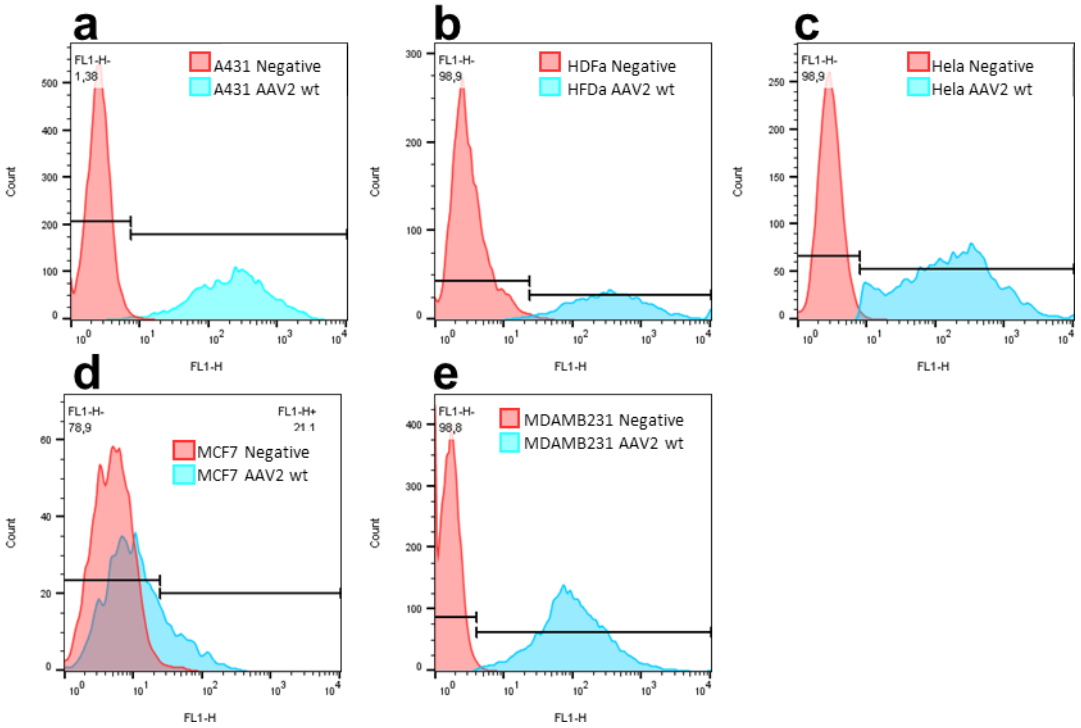
Note that the original modular ITR plasmid was numbered and presented as most ITR plasmids (panel f) and the deletion carrying ITR was the left or 5´ ITR. In the new ITR plasmid (pZMB0522) we kept the classical pUC numbering and thus the deletion is formally now in the right or 3´ ITR.

2.3 SI Figure 3



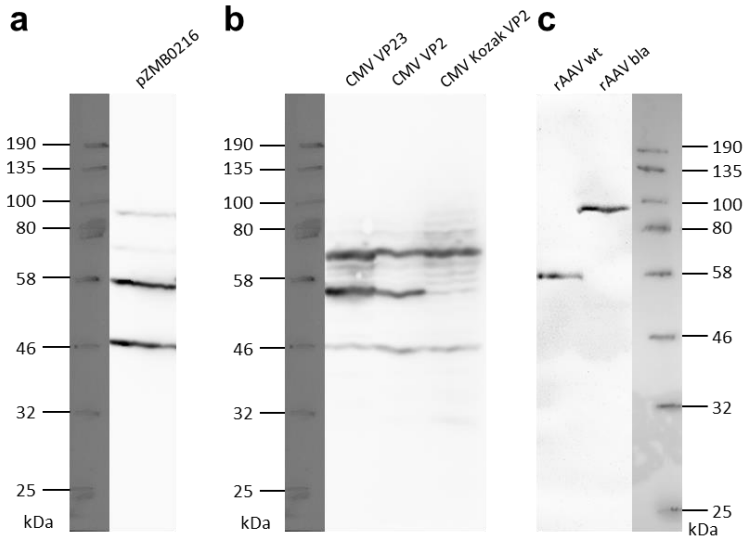
SI Figure 3. Transmission electron microscopy micrographs to calculate the full to empty capsid ratio of rAAV particles. An overview of images used for calculations is shown. Samples were applied to carbon-coated copper grids and stained with 2% uranyl acetate replacement stain (Science Services). rAAVs were visualized with a Philips CM100 (PW6021) with an acceleration voltage of 80 kV. Image analysis was performed using Soft Imaging viewer (Olympus). Scale bars indicate 200 nm. The first image shows diameter measurements to verify the diameter of rAAV particles.

2.4 SI Figure 4



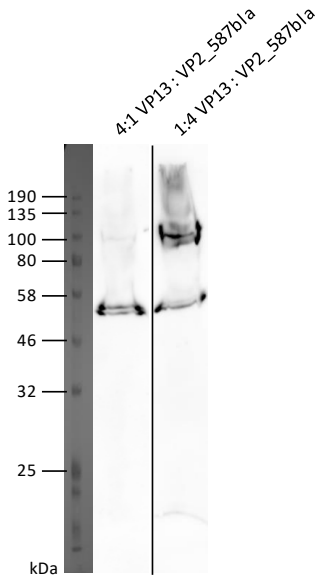
SI Figure 4. Flow cytometry data for rAAV variants in comparison to the negative buffer control. Cells were transduced with rAAVs at a MOI of 10,000 and analyzed via flow cytometry after incubation. Data analysis was performed using FlowJo. In a first step a population of live cells was gated. In a second step a gate of 1% false positive cells was selected in the sample of the negative control. This gate is visualized in each diagram.

2.5 SI Figure 5



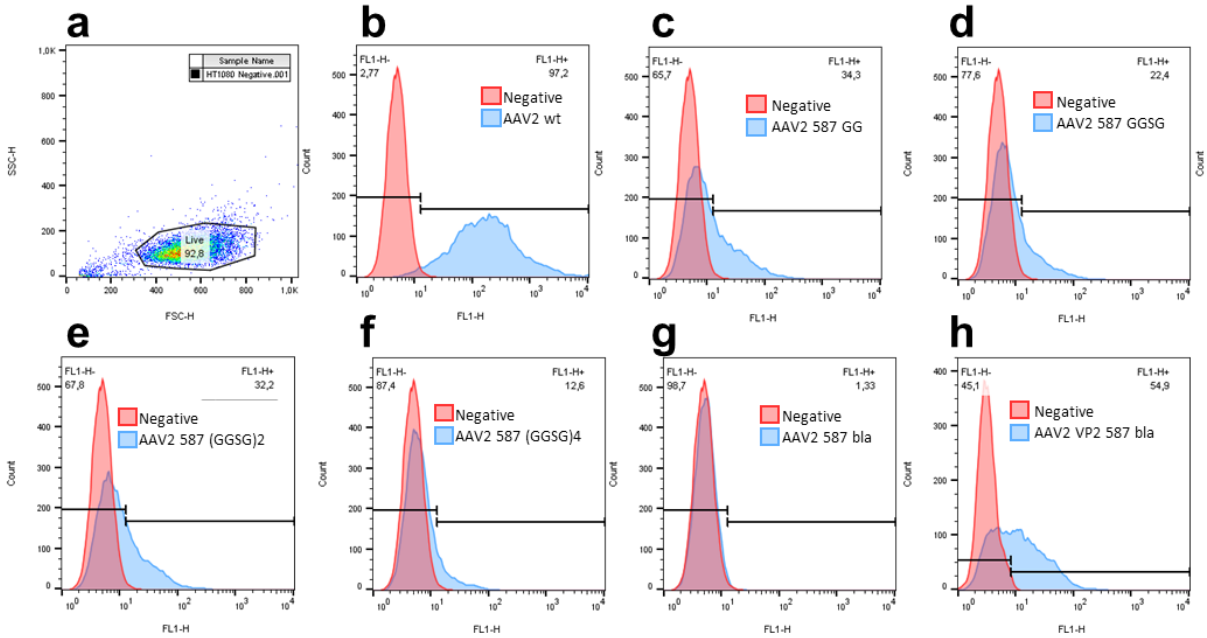
SI Figure 5. Full length western blot images. **a)** Western blot of crude cell lysate after HEK293 triple-transfection with pZMB0216, pZMB0522 and the pHelper plasmid. VP proteins were detected with the B1 antibody (Progen), a secondary HRP-coupled antibody and subsequent chemiluminescence imaging for 270 s. **b)** Western blot of crude cell lysate after transfection with the above mentioned plasmids. VP proteins were detected with the B1 antibody (Progen), a secondary HRP-coupled antibody and subsequent chemiluminescence imaging for 200 s. **c)** Western blot of a rAAV wt and rAAV bla from purified viral stocks. VP proteins were detected with the B1 antibody (Progen), a secondary HRP-coupled antibody and subsequent chemiluminescence imaging for 600 s.

2.6 SI Figure 6



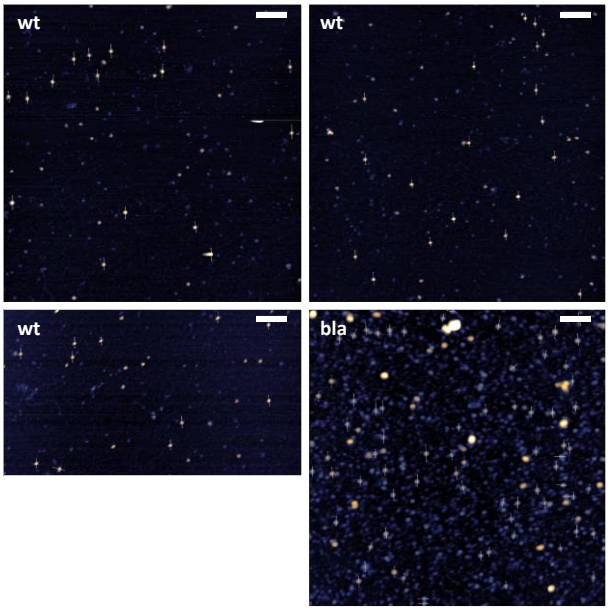
SI Figure 6. Western Blot analysis for rAAV2_mosaicVP2_587bla with respect to the gene dose of each plasmid. Crude cell lysate was applied to a 10 % SDS PAGE gel for electrophoresis. After semi-dry blotting onto a nitrocellulose membrane the VP proteins were detected with the B1 antibody (Progen). The plasmid ratio of the quadruple transfection was based on the ratio of the triple transfection. ITR, RC and pHelper plasmid have been used in a molar ratio of 1:1:1. To create a mosaic virus, CMV_Kozak_VP2_587bla and Rep2Cap2_VP13 have been transfected. The ratio of these plasmids was varied between 4:1 and 1:4.

2.7 SI Figure 7



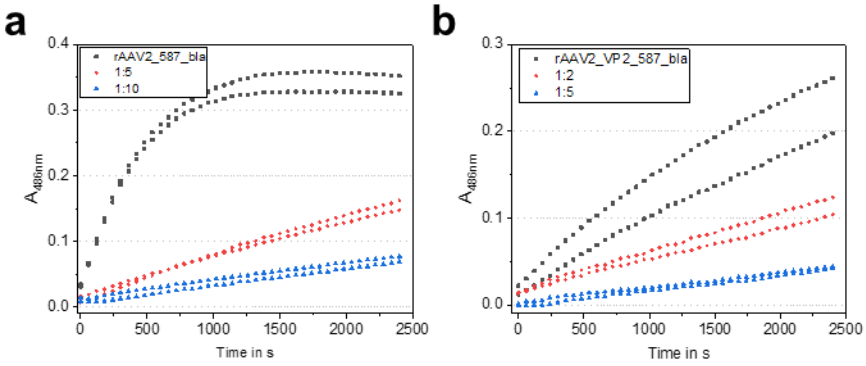
SI Figure 7. Flow cytometry data for rAAV variants in comparison the negative buffer control. HT1080 cells were transduced with rAAVs at a MOI of 50,000 and analyzed via flow cytometry after incubation. Data analysis was performed using FlowJo. In a first step a population of live cells was gated (**a**). In a second step a gate of 1% false positive cells was selected in the sample of the negative control. This gate is visualized in each diagram. Overlay histograms for the wild-type rAAV2 **b**), all glycine-serine linker variants **c-f**) and both bla variants **g-h**) are shown.

2.8 SI Figure 8



SI Figure 8. Atomic force micrographs of wild-type (wt) and fully β-lactamase modified rAAV (bla). The scale bars indicate 200 nm. Thin white lines with ticks show height profiles that were extracted to measure the individual particle diameter for particles that met a height threshold of 5 nm (wild-type capsids) and 6 nm respectively. Particle width was determined at half maximum height.

2.9 SI Figure 9



c

$$\frac{\Delta A}{\Delta t} = \Delta\varepsilon_{486} \cdot c_{bla} \cdot k_{cat_{bla}} \cdot d$$

Sample	Slope in s^{-1}	c(bla) in mol/L	c(rAAV) in mol/L	k_{Cat} in s^{-1}
AAV bla full	5.26×10^{-4}	1.65×10^{-10}	6.14×10^{-12}	334.3
AAV bla full 1:5	7.31×10^{-5}	2.29×10^{-11}	1.23×10^{-12}	232.1
AAV bla full 1:10	2.21×10^{-5}	6.95×10^{-12}	6.14×10^{-13}	140.6
AAV bla mosaic	1.22×10^{-4}	3.83×10^{-11}	6.14×10^{-12}	930.7
AAV bla mosaic 1:2	5.50×10^{-5}	1.73×10^{-11}	3.07×10^{-12}	838.2
AAV bla mosaic 1:5	1.88×10^{-5}	5.90×10^{-12}	1.23×10^{-12}	716.6

SI Figure 9. Nitrocefin assay for determination of β-lactamase activity. **a, b)** Absorption measurements of rAAV2_587_bla and rAAV2_VP2_587_bla in different concentrations at 486 nm. Data points were measured in 1 min intervals. **c)** The first six data point were included in linear regression calculations. Resulting values are given in the table and correlate with dilution. Since very low concentrations were used, there are variations which can be explained in the pipetting the dilutions.

3 SI Tables

3.1 SI Table 1

SI Table 1. Overview of plasmids provided by the iGEM competition in 2010 including a short description of their features.

Plasmid name with description	Length	Backbone
pZMB0246_CMV_VP1up_NLS_mVenus_VP23_453_587koHis expression of VP1 with N-terminal mVenus as well as VP2 and VP3, R585 and R588 in 587 loop regions are changed to Ala (HSPGko), His-tag in 587 loop	5863 bp	pSB1C3_001
pZMB0091_CMV_DARPinE01_mli_VP23 _453_587koHis expression of VP2 with N-terminal DARPin E01 fusion, Arg in 587 loop changed to Ala (HSPGko)	5191 bp	pSB1C3_001
pZMB0156_VP23_453_587wt construct for cloning, encodes VP2/3 with cloning ready loop regions, allows for the construction of a N-terminal VP2 fusion	4004 bp	pSB1C3_001 RFC[25]
pZMB0503_VP1up_NLS construct for cloning, encodes the unique upstream region of VP1 (VP1up) and a nuclear localization sequence (NLS), precursor for assembly of surface exposed VP1 integrations	2525 bp	pSB1C3_001 RFC[25]
pZMB0143_CMV CMV promoter in RFC[10] for construction of N-terminal VP2 fusion vectors or ITR plasmid assembly	2725 bp	pSB1C3 RFC[10]
pZMB0135_hGHpA hGH polyadenylation signal in RFC[10] for assembling of the ITR plasmid.	2550 bp	pSB1C3 RFC[10]

3.2 SI Table 2

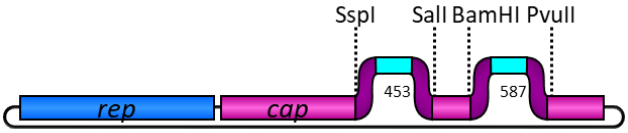
SI Table 2. Characteristics of TEM β-lactamase variant 14FM.

	β-lactamase 14FM
Mutations compared to wild-type TEM-116 β-lactamase (GenBank ID: AY425988)	V31A, A36L, L51I, R120G, E147G, H153R, V159T, M182T, L201P, I208M, E212K, A224V, A249V, T264M
Melting temperature T_M determined using far-UV circular dichroism in phosphate buffer	71.6 °C (2 µM) 72.7 °C (20 µM)
Catalytic constant k_{cat} for nitrocefin conversion in s^{-1}	746 ± 76 s^{-1}

Source: T. Baumann, 2013, PhD thesis, University of Potsdam, Germany, "Stability and interconnected protein properties studied with TEM β-lactamase"

4 Cloning strategy

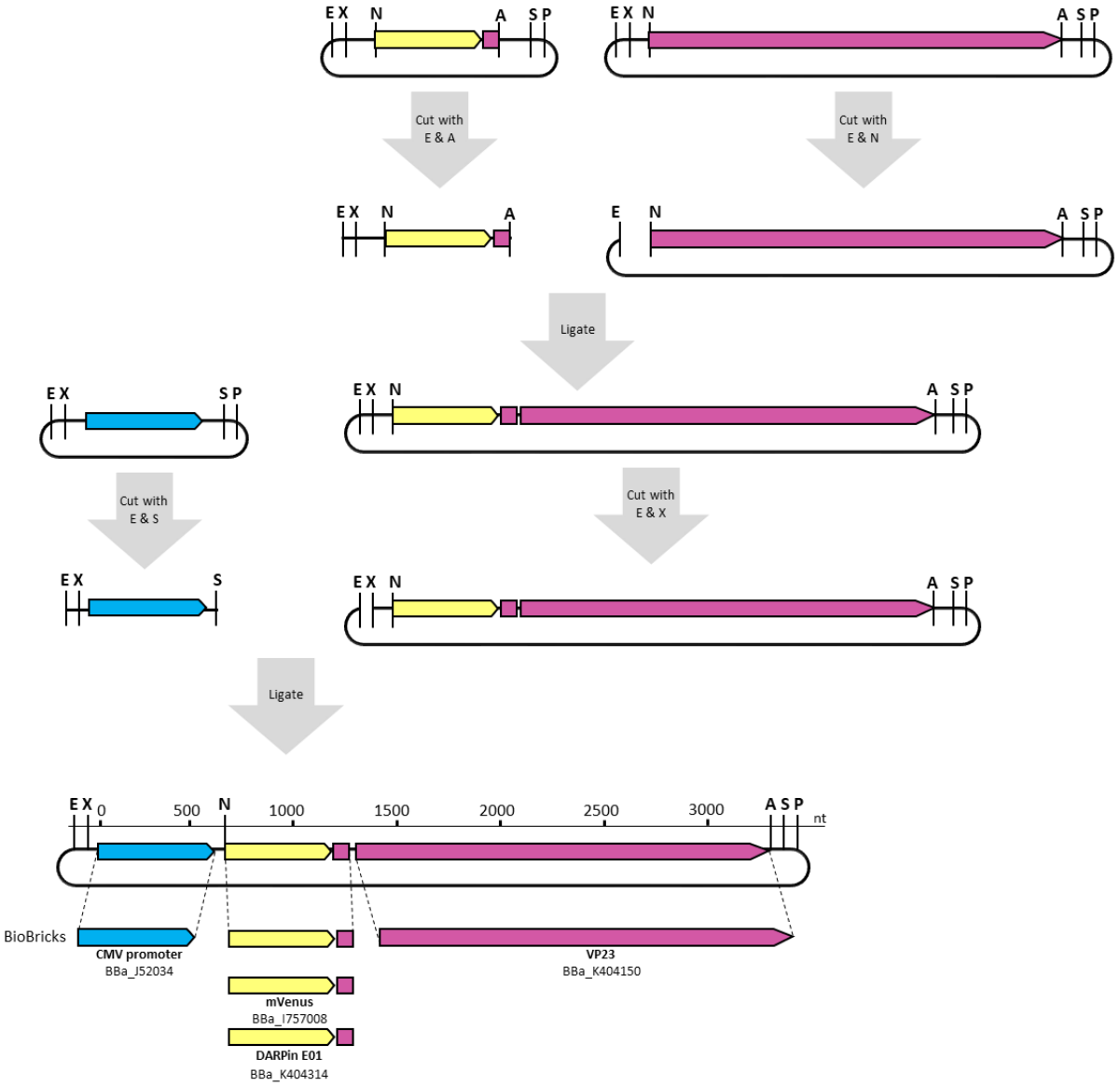
4.1 Cloning of loop modifications in pZMB0216




```
Residue 453 (position determined via Uniprot #P03135)
For 5´-ATTTGTATTACTTGAGCAGAACAAACACTCCAAGTGGTNNN…NNNACCACCACGCAGAG-3´
Rev 5´-TAAACATAATGAACTCGTCTTGTTTGTGAGGTTCACCANNN…NNNTGGTGGTGCGTCTCAGCT-3´

Residue 587 WT
For 5´-GATCCGTATCTACCAACCTCCAGAGAGGCAACNNN…NNNAGACAAGCAGCTACAG-3´
Rev 5´-GCATAGATGGTTGGAGGTCTCTCCGTTGNNN…NNNTCTGTTCGTCGATGTC-3´

Residue 587 KO
For 5´-GATCCGTATCTACCAACCTCCAGGCTGGCAACNNN…NNNGCCCAAGCAGCTACAG-3´
Rev 5´-GCATAGATGGTTGGAGGTCCGACCGTTGNNN…NNNCGGGTTCGTCGATGTC-3´

Viral Brick 587ko-empty
For 5´-GATCCGTATCTACCAACCTCCAGGCTGGCAACGCCCAAGCAGCTACAG-3´
Rev 5´-CTGTAGCTGCTTGGGCGTTGCCAGCCTGGAGGTTGGTAGATACG-3´

Viral Brick 587-GG
For 5´-GATCCGTATCTACCAACCTCCAGAGAGGCAACGGAGGCAGACAAGCAGCTACAG-3´
Rev 5´-CTGTAGCTGCTTGTCTGCCTCCGTTGCCTCTCTGGAGGTTGGTAGATACG-3´

Viral Brick 587-GGSG
For 5´-GATCCGTATCTACCAACCTCCAGAGAGGCAACGGAGGCTCTGGTAGACAAGCAGCTACAG-3´
Rev 5´-CTGTAGCTGCTTGTCTACCAGAGCCTCCGTTGCCTCTCTGGAGGTTGGTAGATACG-3´

Viral Brick 587-2x(GGSG)
For 5´-GATCCGTATCTACCAACCTCCAGAGAGGCAACGGAGGCTCTGGTGGCGGTTCAGGAAGACAAGCAGCTACAG-3´
Rev 5´-CTGTAGCTGCTTGTCTTCCTGAACCGCCACCAGAGCCTCCGTTGCCTCTCTGGAGGTTGGTAGATACG-3´

Viral Brick 587-4x(GGSG)
For 5´-GATCCGTATCTACCAACCTCCAGAGAGGCAACGGAGGCTCTGGTGGCGGTTCAGGAGGTGGAAGCGGCGGAGGTAGTGGCAGACAAGCAGCTACAG-3´
Rev 5´-CTGTAGCTGCTTGTCTGCCACTACCTCCGCCGCTTCCACCTCCTGAACCGCCACCAGAGCCTCCGTTGCCTCTCTGGAGGTTGGTAGATACG-3´
```

SI Figure 11: Cloning strategy for 453 and 587 loop regions. Sequences of oligonucleotides used for the generation of different loop modifications by hybridization are shown. The given oligonucleotides have been cloned via the above mentioned restriction enzymes.

4.2 Cloning of N-terminal VP2 modifications in rAAV2



SI Figure 12: Cloning strategy for pZMB0091_CMV_GOI_mli_VP23 _453_587koHis. This plasmid was not used during this work but shows the potential assembly of N-terminal fusion proteins. Assembly of BioBricks can be performed as shown. The following abbreviations are used for the restriction enzymes and their recognition sites, respectively: E= EcoRI, X= XbaI, N= NgoMIV, A= AgeI, S= SpeI, P= PstI.

10.1.2 **HEK293-KARE1**

10.1.2.1 **Manuscript of "HEK293-KARE1"**

HEK293-KARE1, a cell line with stably integrated adenovirus helper sequences simplifies rAAV production

Rebecca C. Feiner[1,†], rebecca.feiner@uni-bielefeld.de

Kathrin E. Teschner[1,†], kathrin.schlicht@uni-bielefeld.de

Julian Teschner[1], jteschner@uni-bielefeld.de

Kristian M. Müller[1,*], kristian@syntbio.net

† Both authors contributed equally to this work.

[1] Cellular and Molecular Biotechnology, Faculty of Technology, Bielefeld University, Bielefeld, Germany

[*]Corresponding author information (Address; Email; Phone)

Kristian M. Müller, Cellular and Molecular Biotechnology, Faculty of Technology, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany; Email: kristian@syntbio.net; Tel. +49-521-106-6323

Abstract

Background: Recombinant adeno-associated viruses (rAAV) bear great potential for gene therapy and their production with helper-free plasmid systems in HEK293 cells is well established. However, current approaches require the tedious preparation and co-transfection of the invariant adenoviral helper genes for each production cycle. Genomic integration of essential elements can simplify the manufacturing process.

Results: We engineered a HEK293 cell line by stably integrating adenoviral elements E2A, E4 and the non-coding VA RNA. After selection, single cell colonies were analyzed for sequence integration. Insert stability was maintained over more than four months of continuous culture. The novel HEK293-KARE1 cell line was characterized with regard to growth and cell cycle state demonstrating a very small influence of adenoviral genes. RNA Expression of E2A and E4 was verified using RT-PCR at a low level compared to already present E1A. Transduction and rAAV2 production in the new cell line by solely providing a RepCap and a ITR/vector plasmid was established with yields comparable to those obtained with the well-established three-plasmid system in the original HEK293 cells.

Conclusions: The results show that HEK293-Ad-helper cell lines have the potential to reduce upstream time and cost while ensuring consistent and safe rAAV manufacturability.

Keywords: AAV2, adenovirus, virus-like particle

Background

Recombinant adeno-associated virus (rAAV) approaches gained an outstanding reputation in gene therapy due to their beneficial characteristics. Among the numerous advantages over other virus-based gene therapy approaches is the attainable long-term gene expression and the high safety profile resulting from the inability to autonomously replicate without additional genes from a wild type and a helper virus. AAVs target dividing and non-dividing cells, possess low immunogenicity and are produced at larger scale for clinical applications [1]. The AAV toolbox for gene therapy is large: currently 13 different serotypes are characterized and over 100 AAV variants are known [2]. Serotype 2 was studied intensively and was the first serotype to be produced in mammalian cells [3, 4]. Mammalian cells and insect cells are well established as host systems. The best studied and most widely used system relies on the human embryonic kidney cell line 293 (HEK293) [5]. Formerly, a helper virus (e.g. adenovirus (Ad)) was used to facilitate production, but nowadays the biologically safer use of a plasmid system is the standard, in which the genes coding for supporting adenoviral proteins are encoded *in trans* [6]. The latter is named the helper-free plasmid system and handling is in many countries regarded as biosafety level 1 work. The helper-free production was independently published by two research groups in 1998, demonstrating that the presence of the Ad genes E1A, E1B, E2A and E4 and the non-coding RNA VA are required for replication of AAV [7, 8]. The production cell line HEK293 cell line was generated by transfection with fragments of mechanically sheared adenovirus 5 DNA and the presence of adenoviral E1A and E1B in chromosome 19 (19q13.2) was confirmed indicating random integration [9–11]. Besides their biological function for immortalization, transcripts from both genes were identified. The gene of E1A (early E1A protein) promotes host DNA synthesis and stabilizes the tumor suppressor p53 inducing apoptosis [12]. Two gene products generated from E1B (E1B19K and E1B55K) are able to protect cells from the cytotoxic effects of E1A [12]. In the helper-free system the remaining genes E2A and E4 and the non-coding RNA VA are provided by plasmids called pVAE2AE4-5 or pXX6 [7, 8], which are collectively referred to as 'helper plasmids'. More specifically, pHelper GenBank AF369965.1 contains the VA RNAs, the E2A part coding for the 72 kDa single stranded DNA binding protein, and the coding region of E4orf6 [13]. Undesired adenovirus late genes were omitted from those helper plasmids eliminating the possibility of a wild-type adenovirus contamination. Expression of E1A and E1B induces expression of E2A and E4. E1A also acts on the AAV P5 promoter and induces expression of AAV Rep and Cap genes *in trans* [14]. E2A codes mainly for the Ad single-stranded DNA binding protein (DBP), which is besides several other functions involved in augmentation of rAAV circular intermediate formations [15]. The gene E4 codes for several proteins but during rAAV production mainly E4orf6 is of importance. E4orf6 is able to associate to E1B55K and regulates p53 stability promoting apoptosis [16, 17]. The last important feature required for

helper-free rAAV production is VA I RNA. VA I RNA generates a small RNA transcript, that regulates gene expression at the level of translation in transfected cells [18]. In presence of DBP it enhances AAV cytosolic mRNA by stabilization [19].

For rAAV production, the two commonly used helper-free plasmid systems either employ a two- or a three-plasmid system [20, 21]. By distributing the genetic information of rAAV on different plasmids the likelihood of the emergence of a replicative competent virus becomes extremely low. The three-plasmid system consists of the pHelper plasmid coding for the required Ad5 genes, a RepCap plasmid and an ITR-containing plasmid often named vector plasmid. The RepCap plasmid provides the information for the four nonstructural proteins Rep 78, 68, 52, 40 and the three proteins VP1, VP2, VP3 forming the icosahedral capsid structure. Furthermore, the Cap open reading cassette codes, in an alternative reading frame, for the assembly activating protein (AAP) as well the X protein, which influences replication and vision production [22–25]. The third plasmid needed is the ITR-containing plasmid, which harbors the gene of interest flanked on both sites by an inverted terminal repeat (ITR) sequence. ITRs mediate DNA replication and function as packaging signal for the single-stranded DNA into the capsid shell [26]. The two-plasmid system, which was devised by Grimm et al in 2003, combines the genetic information of the adenoviral helper sequences E2A, E4 and VA with AAV serotype specific *rep* and *cap* genes on the pDG plasmid [20]. This approach has been further optimized by the use of minicircle DNA [27].

The widely adopted helper-free plasmid systems are reliable, but pose the inconvenience of the continuous production and co-transfection of the invariant adenoviral helper genes of 9279 bp. So far, even in cellular optimized processes for clinical rAAV production developed by leading research groups, the three-plasmid system is deployed [1].

Several groups have tried to establish production cell lines harboring either single parts of the adenoviral elements or combinations of different elements. First trials to create a production cell line started with genomic integration of rAAV RepCap and ITR parts into HeLa cells [28]. rAAV production starts upon infection with adenovirus and, thus does not need DNA transfection. Next, groups started with integration of RepCap and ITR parts into a E1A and E1B inducible cell line and went on by further introducing adenoviral elements E2, E4 and VA [29]. Nonetheless, integration of all necessary elements resulted in less stable cell lines with decreasing yields of viral particles. Other attempts to create AAV production cell lines relied on genomic integration of adenoviral elements or RepCap parts but also resulted in cell lines where either transfection of invariant plasmids or the use of the adenovirus are required [30, 31].

In order to ease and improve rAAV production, we generated a HEK293-Ad-helper cell line, that stably provides the Ad5 genes necessary for rAAV formation. As known from previous work, high levels of early adenoviral gene products can have toxic effects to cells [32]. Until now, the most effective gene dose for high-titer rAAV production is not known, which provides the opportunity that cells producing a non-toxic level of proteins are able to produce viral particles without addition of a helper plasmid. Such an engineered cell line was generated and named HEK293-KARE1. The integration of helper genes was verified in three clones. rAAV2 virus-particle production was monitored and the transfection protocol was optimized with regard to the amount of required DNA and the potential need for selection pressure for genomic integration. Furthermore, genomic and

transducing titers were determined and compared to the common three-plasmid production system. The presented methodology simplifies the well-established AAV production process and reduces the risks of undesired DNA packaging of pHelper DNA [33].

Results

Integration of pHelper sequences into HEK293

As introduced above, pHelper plasmid (GenBank Accession number: AF369965.1) harbors the three adenoviral elements E2A, E4 and VA and a BLAST comparison confirms 100 % sequence coverage with the corresponding parts of the genome of adenovirus 2 (GenBank: J01917.1). Integration of the entire pHelper plasmid into the genome of HEK293 was facilitated by introduction of a selectable marker into the pHelper plasmid. The gene for blasticidin S deaminase from *Aspergillus terreus* was amplified and cloned into pHelper via two unique restriction sites (NdeI, SalI) yielding pHelper-BSD (**Figure 1**). This plasmid was linearized, dephosphorylated and transfected in HEK293 cells. Three days post transfection selective pressure was applied and cell death of putatively non-transfected cells was observed after further incubation for two days. Medium was exchanged every second day and after two weeks dividing cells were observable, that potentially carried the desired construct. Consequently, a limiting dilution in 96-well plates was used to select single cells for further cultivation and characterization. Cells growing as monoclonal colonies were identified by microscopy and, after reaching confluence, four monoclonal cultures were transferred in a 6-well plate and named HEK293-KARE1a to HEK293-KARE1d. These clones were chosen, because they showed the highest growth rates. The growth of all clones was comparable to that of the parental HEK293 cell line, while maintaining the blasticidin concentration. Cells were passaged regularly, and since HEK293-KARE1d showed decreased viability and transfection efficiency this clone was discarded. For cell lines HEK293-KARE1a to 1c characteristic HEK293-like growth in monolayers, adherence and viability were confirmed beyond 30 passages. During subculturing no changes in morphology were seen in bright field microscopy at 40-fold magnification as exemplified by HEK293-KARE1c after 20 passages (**Figure 2A**) in comparison to the parental HEK293 (**Figure 2B**). Growth behavior of HEK293 and HEK293-KARE1a to 1c was analyzed for 72 h after passage three and is depicted in **Figure 2C**. Experimental results were fitted with an exponential function to determine the population doubling times. Clones HEK293-KARE1a and 1c showed a doubling time comparable to that of HEK293 cells. For HEK293-KARE1b a slower growth rate was calculated.

Verifying genomic integration of pHelper sequences

After isolation of monoclonal cultures and detection of normal cell growth, genomic integration of the helper genes was analyzed. Genomic DNA of the three clones was isolated using a Chelex100 extraction. The presence of the *bsd* and the E2A gene was verified using PCR with two different sets of primers. As seen in **Figure 1** primers Bsd-for and Bsd-rev as well as Ad5-for and Ad5-rev anneal at different regions of the plasmid and, in the case of a successful complete integration, amplification products with a size of 499 bp and 524 bp respectively for E2A

and *bsd* should emerge. Experimentally obtained fragments are shown in **Figure 2D** for all samples, demonstrating that both genes were integrated successfully into the genome of HEK293. Genetic integration of the plasmid sequence might not automatically result in successful gene expression. In a next step, the presence of the gene products was verified.

Gene expression of adenoviral E2A and E4 does not interfere with cell growth and cell cycle progression

The resistance of monoclonal cells to blasticidin indicated that integrated genes are not silenced and gene expression is likely. To verify the presence and expression of the integrated elements E2A and E4 the mRNA level was determined. RNA from HEK293 and HEK293-KARE1 cells was extracted and subsequently cDNA was generated. A qPCR was carried out with three different sets of primers for E1A, E2A and E4. Oligonucleotide sequences to amplify E2A were selected to bind in the coding sequences for the DNA-binding protein of human adenovirus type 2 (GenBank: AAA92217.1). The primer pair used to amplify E4 was selected to anneal at open reading frame 6 coding for E4 34k, mainly involved in the AAV production process [34]. Calculation of ΔΔCt values showed an elevated relative expression level of both adenoviral early genes in HEK293-KARE1 clones compared to HEK293 (**Figure 3A**). These calculations were based on E1A as a normalizing gene which is present in the parental and the novel HEK293 cell line. Since expression of E2A and E4 might influence the expression of E1A, we evaluated three independent RNA isolations and qRT-PCR reaction to show that cycle thresholds for E1A were constant during the experiments and E1A expression is not altered (SI Table 1). The expression level of E2A was higher compared to the E4 expression. Analysis of the PCR products on an agarose gel (**Figure 3B**) revealed the correct sizes with 68 bp, 65 bp and 143 bp for E1A, E2A and E4 respectively. Previously published work showed that expression of adenoviral parts in HEK293 cells can have a strong impact on cell growth and cell cycle. Integration of E4 in an inducible system showed that the progression through the cell cycle can be altered [35]. Furthermore, it is known that high levels of E2A and E4 in the presence of E1 gene products are toxic to cells [32]. To gain a deeper insight in the newly developed cell line, cell cycle progression was determined.

In literature, it was proposed that expression of the E4 ORF6 protein leads to an accumulation of the cells in the S-phase of the cell cycle [35]. To this end, we investigated the effect of adenoviral elements on the cell cycle under general AAV production conditions. A delay in the S-phase was not observed as seen in **Figure 3C** and **D**. Cells were seeded at the same high density as typically done for AAV production (3×10^6 per 100 mm dish), and thus, cell growth was limited. The observed distribution of cell cycle states over the typical production time of 72 h of the cultivation is expected. There is no significant difference between the HEK293 cell lines and the newly developed HEK293-KARE1 cell lines. As a control for adenoviral element expression, a rAAV producing HEK293 sample was used. This showed a high accumulation in G1 phase right after the transfection, which was reversed as the cells adapted to the conditions of the transfection. These experiments suggest that production of rAAVs in HEK293-KARE1 cells is possible, because necessary proteins are expressed and cell proliferation is largely unaltered upon genetic intervention.

rAAV production was confirmed for HEK293-KARE1

In order to assess the rAAV production ability and their long-term productivity, three HEK293-KARE1 clones were cultured similar to normal HEK293 with the only difference of maintaining antibiotic selection. First attempts for rAAV production started with adapting the three-plasmid co-transfection for high titer rAAV production in HEK293 cells. Here, pHelper plasmid, Rep2Cap2^{587wt}- and ITR-plasmid are transfected in a molar 1:1:1 ratio and a total plasmid DNA amount of 15 µg per 100 mm dish. We started with a double-transfection of the Rep2Cap2^{587wt} - and ITR-plasmid in a 1:1 ratio omitting the pHelper plasmid but keeping the total plasmid DNA amount of 15 µg per 100 mm dish. These experiments were carried out with the three cell lines HEK293-KARE1a to 1c after passage 3 and 9 and as a negative control with HEK293 cells and as a positive control with HEK293 cells and triple plasmid transfection. Genomic titers, or more precisely of DNase I-resistant particles, were determined directly without further purification from the crude cell extract using qPCR with a set of ITR-plasmid specific primers. To get meaningful results, the standard curve used for determination of the genomic titer was set up in equal amounts of crude cell extract from untransfected cells. Resulting genomic titers of this experiment are presented in **Figure 4A**. As expected, rAAV production was observed for HEK293-KARE1 clones when omitting the pHelper plasmid. HEK293 cells transfected without addition of pHelper show a dramatic reduction in production capacity of viral particles. Note that this background may be related to small impurities of transfected plasmid. Productivity of all clones is comparable to the parental HEK293 cell line after the well-known triple transfection. Optimization of the rAAV production process was performed only with clone HEK293-KARE1c which yielded in the highest genomic titer ($4.6x10^{10}$ vg per ml) in crude cell extract. Furthermore, no reduction of rAAV titer was observed when cells were cultivated for nine passages. Cryopreservation did not alter production capacity as these experiments were conducted from biological duplicates that were independently thawed. We additionally optimized the production process as the high DNA amount and addition of the antibiotic might not be desired during the production process. Lowering the required plasmid dose and omitting the antibiotic would reduce production costs and possible sources of product contaminations. We hypothesized that reduction of the total DNA amount is compatible with rAAV production. Thus, a reduction of the total plasmid DNA amount from 15 µg to 7.4 µg was tested. This value corresponds to the total DNA amount of Rep2Cap2^{587wt}- and ITR-plasmid in the standard triple-transfection protocol with just omitting the pHelper plasmid DNA. Previous attempts using the standard calcium-phosphate transfection protocol indicated a deterioration of transfection efficiency and associated lower genomic titer (data not shown). The calcium-phosphate transfection method is based on DNA-calcium phosphate precipitation which is dependent on a specific DNA to calcium ion ratio [36]. Simply reducing the DNA amount did not lead to satisfying precipitation, which was observed by a decreased turbidity of the transfection solution. We adapted the transfection protocol for a reduced amount of DNA which resulted in genomic titers comparable to the use of 15 µg total DNA per transfection (**Figure 4B,** first and second bar). In a second optimization step we evaluated the need for the selective pressure during the production process. We did not want to generally omit the use of the antibiotic as it is known that gene silencing might occur if the selective pressure is not maintained. We seeded cells for rAAV production without addition of blasticidin and performed transfections the day after. Comparison of genomic titers from crude cell extracts after 72 h proved rAAV production even without the addition of the antibiotic in comparable yields (**Figure 4B,** third and fourth

bar). The variations observed between separate biologic experiments (**Figure 4B**, second and third bar) can be attributed to passaging and handling. These differences might seem high but it should be considered that experiments were conducted in the course of cell line characterization.

rAAV2 was successfully produced in HEK293-KARE1c, purified and characterized

After reducing the amount of DNA and the concentration of antibiotic, the quality of the produced viral particles was characterized. We produced rAAV2 in the favorite HEK293-KARE1c clone. Cells were seeded on ten 100 mm dishes the day before transfection. Transfections were carried out using the previously adapted calcium-phosphate transfection protocol with a total plasmid DNA of 7.4 µg (2.5 pg per cell) and 1:1 molar ratio of Rep2Cap2^{587wt} to ITR-plasmid. Viral particles were purified using a standard iodixanol-gradient ultracentrifugation protocol. The transducing titer was determined for a serial dilution of viral particles on HT1080 cells. The ITR-delivered gene of interest coded for the fluorescent protein mVenus under the control of a CMV promoter and thus successful transduction resulted in yellow fluorescent cells, which were analyzed by fluorescence microscopy and flow cytometry. To calculate the transducing titer from flow cytometry data, the lowest serial dilution was used **(Figure 5A)** and the genomic titer was determined via qPCR. Both titers are presented in **Figure 5B**. The transducing titer is about 10-fold lower compared to the genomic titer which is in agreement with the literature [37]. The sample purified by ultracentrifugation was analyzed by atomic force microscopy (AFM) (**Figure 5C**). The calculated mean diameter of the marked particles is with 21.5 nm in the expected range of 20 to 24 nm **(Figure 5D**) [38].

Discussion

The demand for an effective, safe rAAV production system is high. A well-known DNA contaminant in rAAV preparations is, beside others, DNA from the Helper plasmid. Depending on the rAAV purification method, 0.01 % to 0.07 % contaminating Helper DNA has been found [33]. This can reduce vector potency and is therefore undesirable in rAAV preparations. Staying with the well-established production system in HEK293 cells, we describe the generation and characterization of a new mammalian cell line providing all necessary adenoviral elements. By eliminating the need of transfecting the Helper plasmid and simultaneously reducing the DNA amount for successful rAAV production, this cell line offers a significant time and cost advantage compared to the use of the parental HEK293 cell line.

rAAV production in HEK293 is classified as biosafety level 1 work in many countries. We assume that a stable HEK293-Ad-helper cell line does not change this assessment. The risk of the generation of a replicative adenovirus or a replicative AAV is not altered whether the adenoviral helper genes are maintained on plasmids or within the genome. Late adenoviral genes required for adenovirus capsid assembly are nonexistent in either system and the risk of unintended recombination or contamination of the culture with other viruses remains at the same level. One could even speculate that recombination events are rarer due to the genomic fixation.

Furthermore, the remaining two plasmid system is unchanged disallowing production of replication competent wtAAV, because ITRs and RepCap genes are separated.

The HEK293-KARE1 cell line described here was generated using linear, dephosphorylated DNA. This allowed for a random integration of the construct into the genome. Integration can cause dysregulation or impede the expression of endogenous genes leading to cell death. Surviving cells can show different levels of the transgene product as the expression is strongly influenced by the surrounding genomic structure. A cell line that highly expresses adenoviral genes could show an inhibited cell growth or high mortality as seen for HEK293-KARE1d. These potential problems might also explain why a directed integration via CRISPR/Cas9 that we previously attempted at a predetermined locus did fail.

Integration of adenoviral elements was proven by PCR for all clones. Analysis of cell growth and cell cycle progression revealed, that for all isolated clones the effect of transgene expression was not significant. The slower growth rate seen for HEK293-KARE1b in Figure 2C might be explained by the initial higher cell density limiting growth after 72 h or indeed by impeded growth. In comparison with the parental HEK293 cell line, growth characteristics of the other KARE clones were nearly equivalent suggesting that an adaptation to suspension and subsequent culturing without serum is possible. This would open new chances for a large-scale production of rAAVs, e.g. in a bioreactor.

The mRNA expression level of E2A and E4 was shown to be elevated compared to the parental HEK293 cell line. To our knowledge, the helper-gene dose required for highly efficient AAV production is not known. Expression of early adenoviral genes is controlled by several promoters starting from the E1A promoter. Upon E1A expression the promoters of E1B, E2, E3 and E4 become active [39]. The strength of the promoter and thus the transcription of mRNA can differ between adenoviral elements, even when all parts are integrated at the same genomic site. The observed relative expression levels were expected, as high E2A expression would compromise cell growth. Previous work showed that a high level of E2A in presence of E1 proteins can have toxic effects on cells [32].

All HEK293-KARE1 clones were analyzed for their ability to produce virus in a two-plasmid system. Transfecting a RepCap plasmid offers the opportunity to change parts of the capsid and the packaged DNA and thus provides the user with maximum flexibility. In comparison to the parental HEK293 cell line, all clones show a high capacity to produce rAAV. We observed during our experiments that the HEK293-KARE1 cell lines require more time to adapt to growth conditions after cryopreservation. For this reason, we believe that at later passages cells were better adapted and yielded in higher titer. To a lower extent we observed the influence on cell adaptation also for HEK293 cells. Adjusting transfection conditions and DNA ratios for HEK293-KARE1c increased the overall titer. These results demonstrate that the low mRNA levels of E2A and E4 support rAAV manufacturing. HEK293-KARE1c was chosen for a larger scale production with subsequent rAAV purification using a well-established purification protocol. These rAAV2 particles showed the expected ratio between genomic and transducing titer as it was known from the three-plasmid system [37]. The rAAV yield was in the range of 3.2×10^4 to 8.4×10^4 GC per cell (Fig. 4B). For similar small-scale experiments with HEK293 cells, albeit in a commercial context, a value of 8.3×10^4 GC per cell has been reported [40]. Other groups determined that the choice of primers and the reference

material for qPCR (circular, linear DNA) have a tremendous effect on the genomic titers [41, 42].Specifically, ITR primers compared to primers binding the gene of interest yielded up to ten times higher genomic copies when plasmids were used as reference. We would like to mention that we used primers binding the hGHpA region whereas in literature primers binding the ITR region have been applied and both experiments used plasmid DNA as reference [40]. Also size and shape of the produced viral particles, as detected by AFM, are in good agreement with traditionally produced rAAV. Hence, the quality of rAAVs produced using the novel cell line is equivalent.

This cell line also opens up the new possibility for experiments that require a wild-type intermediate step, e.g. directed evolution of rAAV. Directed evolution requires the presence of either a wild-type rAAV or the helper genes to allow for amplification of the viral particles. The HEK293-KARE cell line facilitates this method, since neither wild type virus nor transfections have to be used.

Conclusions

For efficient AAV vector production a self-sufficient packaging cell line was created. Based on previous work by others the adenoviral genes VA RNA, E2A and E4orf6 were stably integrated in HEK293 cells in addition to the already present genes E1A and E1B19K. We demonstrated that the resulting HEK293-KARE1c cell line provides a cost-efficient production platform for rAAV2, because a helper plasmid is no longer required. This means not only higher safety, but also the necessity of their time-consuming production and purification. It also offers a streamlined method for the directed evolution approaches.

. The simplified production and the final quality of the preparations show the potential of stably transfected cell lines. We assume that our approach can be combined with other stable transfection strategies such as the integration of the ITR plasmid or the RepCap plasmid open up avenues for inexpensive and safe large scale rAAV production. An adaption to suspension will allow large scale production of rAAVs, e.g. in a bioreactor

Methods

Plasmids. The gene for blasticidin S deaminase of *Aspergillus terreus* was amplified from pcDNA6/myc-His (Thermo Fisher Scientific, Cat. No. V22120; sequence available from the vendor's website), without designated promoter and poly(A) signal, using the primers bsd-SalI (5'-AAAAAAGTCG ACGTTTAAA*C GGATCTGATC AGCACGTGTT*-3') and bsd-NdeI (5'-AAAAAACATA TG*ATCTCGTA GCACGTGTCA GT*-3'), which introduced the underlined SalI and NdeI restriction sites. For better comprehensibility the hydrizinding area is italicized. The digested insert was ligated to the likewise opened pHelper plasmid (GenBank: AF369965.1) resulting in pHelper-BSD (see Additional File). Successful cloning was confirmed by Sanger DNA sequencing (Sequencing Core Facility, CeBiTec, Bielefeld, Germany).

Plasmids for viral particle production were based on a pSB1C3 backbone and were assembled using the iGEM RFC10 BioBrick assembly standard [43]. pZMB0216_Rep_VP123_453_587wt_p5tataless contains a genetically modified *rep/cap* similar to one encoded in pAAV-RC (Genbank Accession number: AF369963.1). In the ITR containing plasmid pZMB0522_ITR_EXS_CMV_mVenus_hGHpA, a CMV promoter expresses the fluorescent protein mVenus. Plasmids are described in detail by Feiner et. al [submitted 2019].

Cell Culture. HEK293 (ACC-305, DSMZ) and HT1080 (ACC-315, DSMZ) were cultured in DMEM (Sigma Aldrich) supplemented with 10 % fetal calf serum (v/v) (FCS, Sigma Aldrich) and 1 % penicillin/streptomycin (v/v) (P/S, Sigma Aldrich) further called DMEM growth medium. The genomic profile of these cells was frequently verified using STR analysis.

HEK293-Ad-helper cell line generation. The pHelper-BSD plasmid was linearized with SalI and dephosphorylated. 9×10^6 HEK293 cells were seeded on a 150 mm^2 T-flask one day before transfection. A total of 45 µg plasmid DNA of pHelper-BSD was transfected using calcium phosphate. Three days after transfection 10 µg/ml of the antibiotic blasticidin (BSD, Sigma Aldrich) was added for selective pressure. Culture medium (DMEM, 10 % FCS, 1% P/S, 10 µg/ml BSD) was changed every second day until cell growth was observable (approximately two weeks). Cells were trypsinized and subjected to limiting dilution on a 96-well plate. From this point on, cells were cultured in DMEM growth medium with 5 µg/ml BSD. Growth of single cell colonies was observed after about two additional weeks. Single cell colonies were subcultured and characterized for the integration of the adenoviral elements.

Analysis of genomic integration. Genomic DNA from subcultured monoclonal cell clones was isolated by Chelex 100 (Bio-Rad) extraction. 2.5×10^6 cells were centrifuged (1 min 10000×g) and washed with PBS. The pellet was resuspended in 100 µl Chelex suspension (5 % w/v), incubated for 30 min at 56 °C, vortexed for 10 s and incubated for 8 min at 100 °C. The solution was centrifuged briefly at approx. 17,000×g and the supernatant was transferred to a new tube. The primer pairs Ad5-for (5'-CAACTCCATG CTTAACAGTC CCCA-3')/Ad5-rev (5'-TCCAGTGCTG CAACCCTGTG TATC-3') and Bsd-for (5'-GGATCTGATC AGCACGTGTT-3')/Bsd-rev (5'-ATCTCGTAGC ACGTGTCAGT-3') are binding on E2A- and bsd-gene, respectively (Figure 1 A) and were used for the PCR reaction. After successful detection of integration four clones were selected and name HEK293-KARE1a to HEK293-KARE1d.

Growth curves. 1×10^5 cells of each clone were seeded in a 12-well plate and incubated under standard conditions. Cells were harvested by trypsinization and resuspended in 1 ml cell culture medium. Samples for viable cell density and viabilities were taken daily and analyzed using an automated cell counting system (CEDEX, Roche Diagnostics). Growth curves were analyzed by exponential fitting with Origin (OriginLab 2018) and population doubling times were calculated.

Cell cycle analysis. 2×10^6 cells were harvested by trypsinization and washed with ice-cold PBS. Briefly, cells were resuspended in 500 µl ice-cold PBS before 500 µl 2 % (w/v) paraformaldehyde was added to the solution. The mixture was stored 1 h on ice before washing with ice-cold PBS and resuspension in 1 ml ice-cold 70 % ethanol. Samples were taken daily and stored at -20 °C for 5-7 days. After permeabilization cells to be analyzed by flow cytometry were washed with PBS before treatment with 1 ml staining solution (100 µg/ml RNase A (AppliChem) and 30 µg/ml propidium iodide (AppliChem)) in PBS at 37 °C for 30 min in the dark. Measurements of samples

were performed on a Becton-Dickenson FACSCalibur. 10,000 counted events were analyzed with the software ModFit LT (Verity House Software).

RNA isolation and quantitative RT-PCR. Total cellular RNA from HEK293 and HEK293-KARE1 clones was isolated using Trizol reagent (Invitrogen) according to the instructions by the manufacturer. 1 µg of RNA was subsequently digested by 1 U DNase I (ThermoFisher Scientific) for 30 min at 37 °C. cDNA was synthesized by oligo(dT)$_{18}$ and Revert Aid M-MuLV reverse transcriptase (ThermoFisher Scientific). Quantitative real-time PCR using GoTaq® qPCR Master Mix (Promega) was performed with a LightCycler 480 II detection system (Roche) according to the manufacturer's protocol. The following primer pairs were used at a 4 µM concentration to detect E1A (forward primer: 5'-AACCAGTTGC CGTGAGAGTTG-3'; reverse primer: 5'-CTCGTTAAGC AAGTCCTCGA TACAT-3'), E2A (forward primer: 5'-TCAGGCACAA CCATCCGCGG-3'; reverse primer 5'-GGTCGGGCGC CGATATCTTGA-3') and E4 (forward primer: 5'-GAACCCTAGT ATTCAACCTG CCACCTCCC-3'; reverse primer: 5'- CCACACGGTT TCCTGTCGAGCC-3'). All samples were run in triplicate. Ct values of technical replicates were averaged and averages were used for further calculations. The relative expression ΔΔCt of E2A and E4 in HEK293-KARE1 cells was calculated from the differences in HEK293 and HEK293-KARE1a to HEK293-KARE1c after normalizing to E1A gene expression (ΔΔCt=Ct(Gene$^{HEK293\text{-}KARE}$)-Ct(E1A$^{HEK293\text{-}KARE}$)-[Ct(Gene^{HEK293})-Ct(E1A^{HEK293})]).

Viral particles production and purification. HEK293-KARE1 cells were seeded at a density of 3×10^6 cells per 100 mm dish the day before transfection. On the day of transfection, a total amount of 15 µg DNA per 100 mm dish of pSB1C3_001-RepCap^{587WT} (pZMB0216) and pUC19bb-CMV_mVenus_hGH-pA (pZMB0522) was transfected using the calcium-phosphate method. The DNA was added to 0.5 ml 0.3 M $CaCl_2$ and added dropwise to 0.5 ml of 2× HBS buffer (50 mM HEPES pH 7.05, 1.5 mM Na_2HPO_4, 280 mM NaCl). The transfection solution was mixed well and added immediately to the cells. After 72 h incubation at 37 °C the cells were harvested and separated from the medium by centrifugation (900×g, 5 min). The cell pellet was subjected to three freeze-thaw cycles by alternating between liquid nitrogen and a 37 °C water bath. After Benzonase (50 U/ml, Sigma Aldrich) treatment at 37 °C for 30 min, the lysate was clarified via centrifugation at 3,000×g, 10 min. The crude lysate was added to a discontinuous iodixanol gradient (Progen Biotechnik GmbH) of 60 %, 40 %, 25 % and 15 % iodixanol in an open top polyallomer 16 × 76 mm tube (Science Services) using a syringe equipped with a 23G x 3 1/8" injection needle as described in literature [44]. Tubes were sealed with cap assemblies and centrifuged in a T-880 rotor (Sorvall) at 340,000×g for 4 h at 18 °C. The fraction containing viral particles was collected with a 21G x 1 1/2" injection needle from the bottom of the tube and the buffer was exchanged to 1× HBSS (Sigma Aldrich) via an Amicon Ultra-4 100K ultrafiltration unit.

Optimized viral particles production. HEK293-KARE1c cells were seeded as described before in DMEM without BSD. On the day of transfection, a total amount of 7.4 µg DNA per 100 mm dish of pSB1C3_001-RepCap^{587WT} (pZMB0216) and pUC19bb-CMV_mVenus_hGH-pA (pZMB0522) in a molar ratio of 1:1 was used. The DNA was added to 0.25 ml 0.3 M $CaCl_2$ and added dropwise to 0.25 ml of 2× HBS buffer. The transfection solution was mixed well and added immediately to the cells. After a 72 h incubation at 37 °C cells were harvested and processed as described in viral particles production and purification.

Determination of genomic titers. Genomic titers of viral particles were determined via quantitative real-time PCR (qPCR). An aliquot of rAAV containing sample (5 µl) was incubated with 4 U DNase I (NEB) at 37 °C for 30 min followed by a 20 min heat inactivation at 75 °C. In the qPCR reaction 5 µl of a fivefold dilution of the DNase I digest was mixed with 5 µl primer-mix at a final concentration of 4 µM (qPCR-hGH-for: 5'-CTCCCCAGTG CCTCTCCT-3' / qPCR-hGH-rev: 5'-ACTTGCCCCT TGCTCCATAC-3') and 2x GoTaq qPCR Mastermix (Promega) to a final volume of 20 µl. The reaction was carried out as described by the manufacturers protocol using a LightCycler 480 II system (Roche Life Science). The genomic titer was calculated from a standard curve of 10 to 10^6 copies of the ITR plasmid (pZMB0522) with an efficiency between 90-110 % and an R value less than 0.1 (mean squared error of the single data points fit to the regression line). Genomic titers in crude lysates were estimated from a standard curve mixed with the same amount of a non-transfected cell lysate. Since DNase I is partially inhibited in the presence of cell debris a small plasmid background can remain. All samples were processed identically in technical triplicates.

Transduction assay. Functionality of purified rAAV2 preparations was analyzed by transduction of HT1080 cells with a vector plasmid coding for the fluorescence reporter mVenus (pZMB0522). 10,000 cells were seeded on a 12-well plate and after cells had settled a serial dilution of viral particles was applied to each well. Cells were detached by trypsinization for flow cytometry analysis. 10,000 events were counted using a Becton-Dickenson FACSCalibur instrument. Data sets were analyzed using Flowing Software 2.5.1. Positive cells were gated above the 99 % interval of the negative control.

Atomic force microscopy. AFM measurements were performed on a Multimode 8 AFM (Bruker, Santa Barbara, CA, USA) with Tap300Al-G cantilevers (BudgetSensors, Sofia, Bulgaria) in tapping mode in air. 2 µl sample of UC purified rAAV2 from HEK293-KARE1c was spotted onto freshly cleaved mica for 1 min before it was rinsed three times with distilled water and dried under a gentle nitrogen flow. Data analysis was performed with Gwyddion 2.48. Obtained images were treated with offset and plane correction algorithms and the size of visualized particles was measured at half maximum particle height.

List of abbreviations

rAAV: recombinant Adeno-associated virus

Ad5: Adenovirus type 5

qPCR: quantitative polymerase chain reaction

hGH: human Growth hormone

BSD: blasticidin

ORF: Open reading frame

ITR: Inverted terminal repeat

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Funding

We acknowledge support of the publication fee by Deutsche Forschungsgemeinschaft and the Open Access Publication Funds of Bielefeld University. This work was supported by Bielefeld University.

Authors' contributions

All authors were involved in the design of this study. JT and KT generated the cell line. RF performed the experiments to characterize the cell line. KT analyzed and optimized rAAV production. RF, KT and KM wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Marco Radukic for cloning of pHelper-BSD and Philipp Borchert for experimental assistance.

References

1. Grieger JC, Soltys SM, Samulski RJ. Production of Recombinant Adeno-associated Virus Vectors Using Suspension HEK293 Cells and Continuous Harvest of Vector From the Culture Media for GMP FIX and FLT1 Clinical Vector. Mol Ther. 2016;24:287–97.

2. Wu Z, Asokan A, Samulski RJ. Adeno-associated virus serotypes: vector toolkit for human gene therapy. Mol Ther. 2006;14:316–27.

3. Hagen S, Baumann T, Wagner HJ, Morath V, Kaufmann B, Fischer A, et al. Modular adeno-associated virus (rAAV) vectors used for cellular virus-directed enzyme prodrug therapy. Sci Rep. 2014;4:3759.

4. Srivastava A. In vivo tissue-tropism of adeno-associated viral vectors. Curr Opin Virol. 2016;21:75–80.

5. Miyake K, Miyake N, Yamazaki Y, Shimada T, Hirai Y. Serotype-independent method of recombinant adeno-associated virus (AAV) vector production and purification. J Nippon Med Sch. 2012;79:394–402.

6. Richardson WD, Westphal H. A cascade of adenovirus early functions is required for expression of adeno-associated virus. Cell. 1981;27 1 PART 2:133–41.

7. Matsushita T, Elliger S, Elliger C, Podsakoff G, Villarreal L, Kurtzman GJ, et al. Adeno-associated virus vectors can be efficiently produced without helper virus. Gene Ther. 1998;5:938–45.

8. Xiao X, Li J, Samulski RJ. Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. J Virol. 1998;72:2224–32.

9. Louis N, Evelegh C, Graham FL. Cloning and sequencing of the cellular-viral junctions from the human adenovirus type 5 transformed 293 cell line. Virology. 1997;233:423–9.

10. Lin Y-C, Boone M, Meuris L, Lemmens I, Van Roy N, Soete A, et al. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. Nat Commun. 2014;5:4767.

11. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol. 1977;36:59–74.

12. Lowe SW, Ruley HE. Stabilization of the p53 tumor suppressor is induced by adenovirus-5 E1A and accompanies apoptosis. Genes Dev. 1993;7:535–45.

13. Colosi P. Patent US 6,004,797 Adenovirus helper-free recombinant AAV Virion production. 1999.

14. Chang LS, Shi Y, Shenk T. Adeno-associated virus P5 promoter contains an adenovirus E1A-inducible element and a binding site for the major late transcription factor. J Virol. 1989;63:3479–88.

15. Duan D, Sharma P, Dudus L, Zhang Y, Sanlioglu S, Yan Z, et al. Formation of adeno-associated virus circular genomes is differentially regulated by adenovirus E4 ORF6 and E2a gene expression. J Virol. 1999;73:161–9.

16. Sarnow P, Hearing P, Anderson CW, Halbert DN, Shenk T, Levine AJ. Adenovirus early region 1B 58,000-dalton tumor antigen is physically associated with an early region 4 25,000-dalton protein in productively infected cells. J Virol. 1984;49:692–700.

17. Querido E, Marcellus RC, Lai A, Charbonneau R, Teodoro JG, Ketner G, et al. Regulation of p53 levels by the E1B 55-kilodalton protein and E4orf6 in adenovirus-infected cells. J Virol. 1997;71:3788–98.

18. Akusjärvi G, Svensson C, Nygård O. A mechanism by which adenovirus virus-associated RNAI controls translation in a transient expression assay. Mol Cell Biol. 1987;7:549–51.

19. Janik JE, Huston MM, Cho K, Rose JA. Efficient synthesis of adeno-associated virus structural proteins requires both adenovirus DNA binding protein and VA I RNA. Virology. 1989;168:320–9.

20. Grimm D, Kay MA, Kleinschmidt JA. Helper virus-free, optically controllable, and two-plasmid-based production of adeno-associated virus vectors of serotypes 1 to 6. Mol Ther. 2003;7:839–50.

21. Grieger JC, Choi VW, Samulski RJ. Production and characterization of adeno-associated viral vectors. Nat Protoc. 2006;1:1412–28.

22. Sonntag F, Schmidt K, Kleinschmidt JA. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. Proc Natl Acad Sci U S A. 2010;107:10220–5.

23. Sonntag F, Köther K, Schmidt K, Weghofer M, Raupp C, Nieto K, et al. The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes. J Virol. 2011;85:12686–97.

24. Cao M, Chiriva-Internati M, Hermonat PL. AAV2 X increases AAV6 rep/cap-driven rAAV production. Virology. 2015;482:84–8.

25. Cao M, You H, Hermonat PL. The X gene of adeno-associated virus 2 (AAV2) is involved in viral DNA replication. PLoS One. 2014;9:e104596.

26. Yan Z, Zak R, Zhang Y, Engelhardt JF. Inverted terminal repeat sequences are important for intermolecular recombination and circularization of adeno-associated virus genomes. J Virol. 2005;79:364–79.

27. Schnödt M, Schmeer M, Kracher B, Krüsemann C, Espinosa LE, Grünert A, et al. DNA Minicircle Technology Improves Purity of Adeno-associated Viral Vector Preparations. Mol Ther Acids. 2016;5:e355.

28. Clark KR, Voulgaropoulou F, Fraley DM, Johnson PR. Cell lines for the production of recombinant adeno-associated virus. Hum Gene Ther. 1995;6:1329–41.

29. Qiao C, Li J, Skold A, Zhang X, Xiao X. Feasibility of generating adeno-associated virus packaging cell lines containing inducible adenovirus helper genes. J Virol. 2002;76:1904–13.

30. Nakamura S, Nakamura R, Shibata K, Kobayashi M, Sahara N, Shigeno K, et al. Development of packaging cell lines for generation of adeno-associated virus vectors by lentiviral gene transfer of trans-complementary components. Eur J Haematol. 2004;73:285–94.

31. Yuan Z, Qiao C, Hu P, Li J, Xiao X. A versatile adeno-associated virus vector producer cell line method for scalable vector production of different serotypes. Hum Gene Ther. 2011;22:613–24.

32. Klessig DF, Grodzicker T, Cleghon V. Construction of human cell lines which contain and express the adenovirus DNA binding protein gene by cotransformation with the HSV-1 tk gene. Virus Res. 1984;1:169–88.

33. Lecomte E, Tournaire B, Cogné B, Dupont JB, Lindenbaum P, Martin-Fontaine M, et al. Advanced characterization of DNA molecules in rAAV vector preparations by single-stranded virus next-generation sequencing. Mol Ther - Nucleic Acids. 2015;4:e260.

34. Mohammadi ES, Ketner EA, Johns DC, Ketner G. Expression of the adenovirus E4 34k oncoprotein inhibits repair of double strand breaks in the cellular genome of a 293-based inducible cell line. Nucleic Acids Res. 2004;32:2652–9.

35. Grifman M, Chen NN, Gao G p, Cathomen T, Wilson JM, Weitzman MD. Overexpression of cyclin A inhibits augmentation of recombinant adeno-associated virus transduction by the adenovirus E4orf6 protein. J Virol.

1999;73:10010–9.

36. Jordan M, Wurm F. Transfection of adherent and suspended cells by calcium phosphate. Methods. 2004;33:136–43.

37. Zeltner N, Kohlbrenner E, Clément N, Weber T, Linden RM. Near-perfect infectivity of wild-type AAV as benchmark for infectivity of recombinant AAV vectors. Gene Ther. 2010;17:872–9.

38. Chen H. Comparative observation of the recombinant adeno-associated virus 2 using transmission electron microscopy and atomic force microscopy. Microsc Microanal. 2007;13:384–9.

39. Berk AJ. Adenovirus promoters and E1A transactivation. Annu Rev Genet. 1986;20:45–79.

40. Rego M, Hanley LM, Ersing I, Guerin K, Tasissa M, Haery L, et al. Improved yield of AAV2 and rAAV2-retro serotypes following sugar supplementation during the viral production phase. bioRxiv. 2018;:488585.

41. Furuta-Hanawa B, Yamaguchi T, Uchida E. 2D droplet digital PCR as a tool for titration and integrity evaluation of recombinant adeno-associated viral vectors. Hum Gene Ther Methods. 2019;:hgtb.2019.031.

42. Wang F, Cui X, Wang M, Xiao W, Xu R. A reliable and feasible qPCR strategy for titrating AAV vectors. Med Sci Monit Basic Res. 2013;19:187–93.

43. Shetty RP, Endy D, Knight TF. Engineering BioBrick vectors from BioBrick parts. J Biol Eng. 2008;2:5.

44. Zolotukhin S, Byrne BJ, Mason E, Zolotukhin I, Potter M, Chesnut K, et al. Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. Gene Ther. 1999;6:973–85.

Figures/Figure Legends

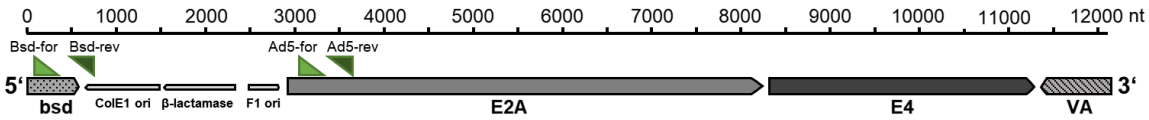


Figure 1: Integration of adenoviral elements into HEK293. Scheme of the linearized helper plasmid after restriction digest with SalI. The blasticidin resistance gene *bsd* is located at the 5'-end. PCR primer positions are marked with green triangles.

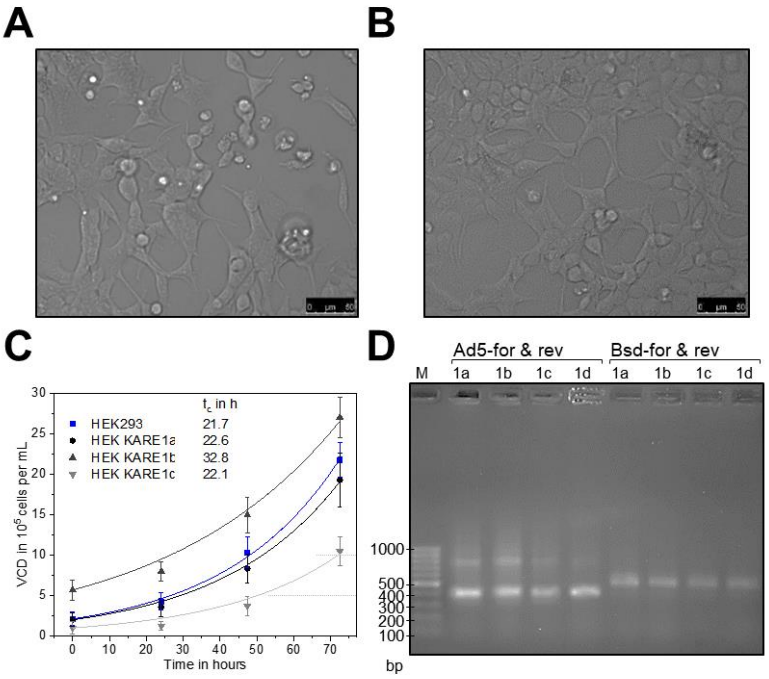


Figure 2: Analysis of cell morphology and growth. (A, B) Comparison of cell morphology in bright field images of (A) HEK293-KARE1c and (B) HEK293. The morphology of HEK293-KARE1c looks similar to the one of the parental HEK293 cell line. (C) Growth curves of HEK293-KARE1 clones and HEK293 were recorded. Cells were seeded on a 12-well plate and harvested after the indicated incubation time. Viable cell density and viabilities were analyzed using an automated cell counting system (CEDEX, Roche Diagnostics). Doubling times t_d were calculated from an exponential fit with Origin 2018 (OriginLab) and are given in the inset of the graph. (D) Agarose gel of the PCR analysis of pHelper-BSD integration. Genomic DNA of the four clones HEK293-KARE1a to HEK293-KARE1d probed by PCR using the primer pairs given above each lane.

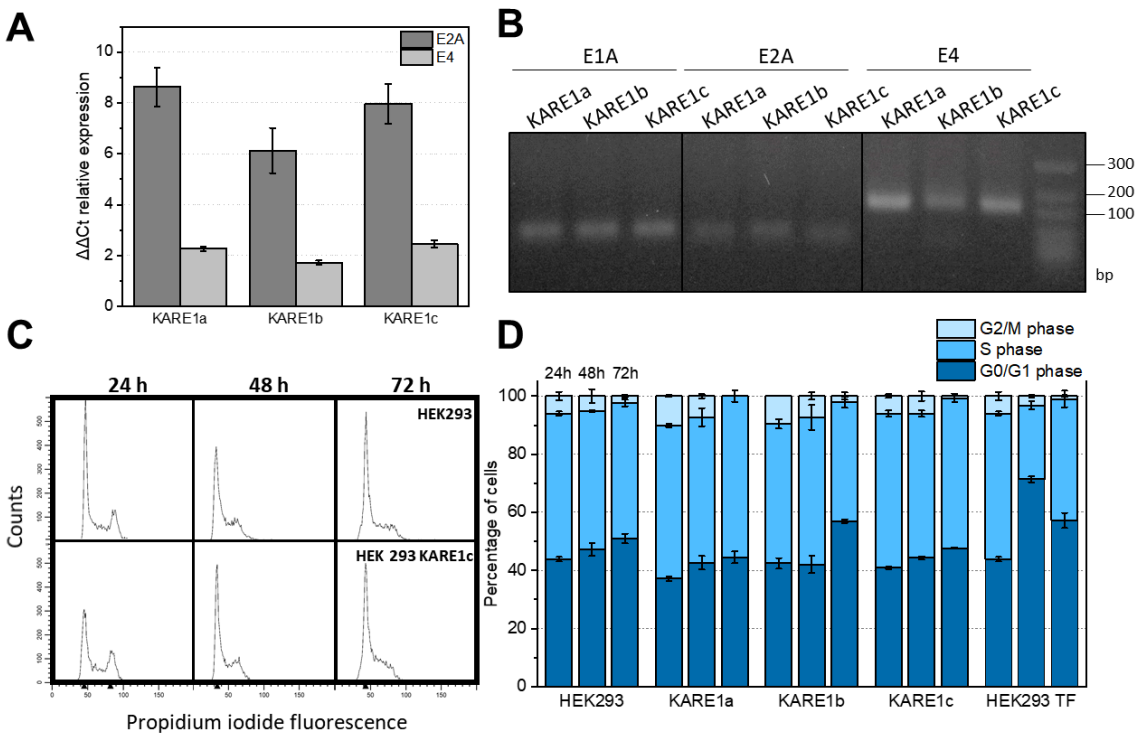


Figure 3: Analysis of adenoviral elements expression and their effect on the cell cycle of HEK293-KARE cell lines. (A) RNA was isolated from HEK293 and HEK293-KARE1 cells. The mRNA levels of E2A and E4 were determined by RT-qPCR, the error bars represent the standard deviation from three technical replicates. ΔΔCt values were calculated from the difference between expression in HEK293 and HEK293-KARE1 cells after normalizing to E1A gene expression. (B) Agarose gel from qPCR products for HEK293-KARE1-clones. (C, D) Effects of adenoviral early genes on cell cycle progression. (C) The DNA content of cells (exemplary for HEK293, HEK293-KARE1c) was measured after propidium iodide staining by flow cytometry at defined time points (24 h, 48 h, 72 h). In total 10,000 events were counted. (D) Percentage of cells in different cell cycle phases are presented for each cell line. HEK293 TF represent triple-transfected HEK293 cells for AAV production. Transfection was carried out 24 hours after seeding. This latter analysis was done based on cells gated for mVenus production present on the ITR plasmid. The flow cytometry data were analyzed with ModFit LT (Verity Software House).

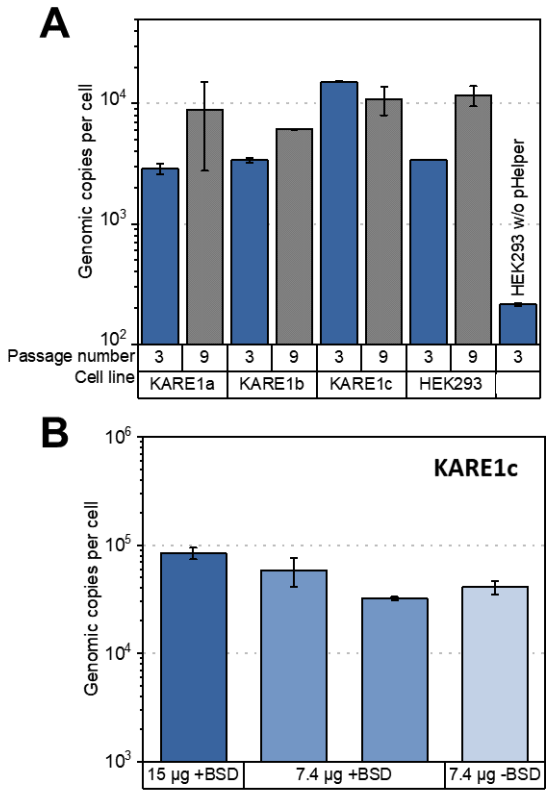


Figure 4: Analysis of rAAV2 production in HEK293-KARE1 and HEK293 cells at passage three and nine after cryopreservation. (A) HEK293-KARE1 cells (1×100 mm dish) were transfected with ITR-plasmid and RepCap-plasmid, for HEK293 the pHelper was included. Three days after transfection rAAV yield was determined from crude cell lysate after three freeze-thaw cycles by qPCR. A molar ratio of 1:1 of the plasmids RepCap:ITR was used during this experiment. For HEK293 cells experiments were performed also with a plasmid ratio of 1:1:1 for pHelper:RepCap:ITR. (B) Comparison of rAAV production in HEK293-KARE1c cells depending on15 µg transfection protocol and the optimized 7.4 µg transfection protocol. rAAV2 yield was quantified in crude lysates by qPCR. In general, RepCap and ITR plasmid were transfected in a molar 1:1 ratio.

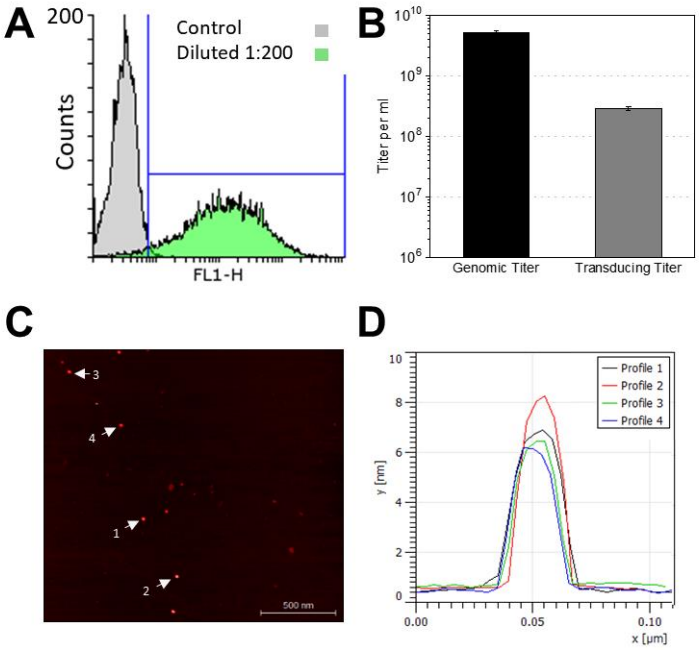


Figure 5: Characterization of rAAV2. Viral particles were produced in the HEK293-KARE1c cell line seeded in ten 100 mm dishes and subsequently purified via iodixanol ultracentrifugation. (A) Determination of transduction ability of rAAV (1:200 diluted sample) on HT-1080 cells using expression of the mVenus gene measured by flow cytometry of 10,000 cells in total. (B) The amount of DNase I-resistant particles was determined by qPCR from a standard curve of the ITR-plasmid. The total volume of the final viral sample was 2.5 ml. The transducing titer was determined based on percent of fluorescent cells as depicted in panel A (see methods). (C, D) rAAV were visualized by atomic force microscopy and resulting particle height profiles were determined. A mean diameter of 21.5 ± 1.8 nm was measured at half maximum height of 8 particles.

10.1.2.2 **Additional file of "HEK293-KARE1"**

Additional File

HEK293-KARE1, a cell line with stably integrated adenovirus helper sequences simplifies rAAV production

Rebecca C. Feiner[1,†], Kathrin E. Teschner[1,†,], Julian Teschner[1],Kristian M. Müller[1,*],

† Both authors contributed equally to this work.

[1] Cellular and Molecular Biotechnology, Faculty of Technology, Bielefeld University, Bielefeld, Germany

[*]Corresponding author information (Address; Email; Phone)

Kristian M. Müller, Cellular and Molecular Biotechnology, Faculty of Technology, Bielefeld University, Universitätsstraße 25, 33615 Bielefeld, Germany; Email: kristian@syntbio.net; Tel. +49-521-106-6323

SI Table 1: Raw data of E1A mRNA analysis from three independent qRT-PCR experiments. mRNA samples were isolated from cells directly before the experiment.

	Ct (1)	Ct (2)	Ct (3)
Kare1a	19.22	15.01	15.78
Kare1b	18.98	14.55	15.80
Kare1c	19.70	14.95	15.76
HEK293	19.22	15.89	15.77

Sequence information for pHelper-BSD

```
  1 ggtacccaac tccatgctta acagtcccca ggtacagccc accctgcgtc gcaaccagga
 61 acagctctac agcttcctgg agcgccactc gccctacttc cgcagccaca gtgcgcagat
121 taggagcgcc acttcttttt gtcacttgaa aaacatgtaa aaataatgta ctaggagaca
181 ctttcaataa aggcaaatgt ttttatttgt acactctcgg gtgattattt acccccccacc
241 cttgccgtct gcgccgttta aaaatcaaag gggttctgcc gcgcatcgct atgcgccact
301 ggcagggaca cgttgcgata ctggtgttta gtgctccact taaactcagg cacaaccatc
361 cgcggcagct cggtgaagtt ttcactccac aggctgcgca ccatcaccaa cgcgtttagc
421 aggtcgggcg ccgatatctt gaagtcgcag ttggggcctc cgccctgcgc gcgcgagttg
481 cgatacacag ggttgcagca ctggaacact atcagcgccg ggtggtgcac gctggccagc
```

```
 541 acgctcttgt cggagatcag atccgcgtcc aggtcctccg cgttgctcag ggcgaacgga
 601 gtcaactttg gtagctgcct tcccaaaaag ggtgcatgcc caggctttga gttgcactcg
 661 caccgtagtg gcatcagaag gtgaccgtgc ccggtctggg cgttaggata cagcgcctgc
 721 atgaaagcct tgatctgctt aaaagccacc tgagcctttg cgccttcaga gaagaacatg
 781 ccgcaagact tgccggaaaa ctgattggcc ggacaggccg cgtcatgcac gcagcacctt
 841 gcgtcggtgt tggagatctg caccacattt cggccccacc ggttcttcac gatcttggcc
 901 ttgctagact gctccttcag cgcgcgctgc ccgttttcgc tcgtcacatc catttcaatc
 961 acgtgctcct tatttatcat aatgctcccg tgtagacact taagctcgcc ttcgatctca
1021 gcgcagcggt gcagccacaa cgcgcagccc gtgggctcgt ggtgcttgta ggttacctct
1081 gcaaacgact gcaggtacgc ctgcaggaat cgccccatca tcgtcacaaa ggtcttgttg
1141 ctggtgaagg tcagctgcaa cccgcggtgc tcctcgttta gccaggtctt gcatacggcc
1201 gccagagctt ccacttggtc aggcagtagc ttgaagtttg cctttagatc gttatccacg
1261 tggtacttgt ccatcaacgc gcgcgcagcc tccatgccct tctcccacgc agacacgatc
1321 ggcaggctca gcgggtttat caccgtgctt tcactttccg cttcactgga ctcttccttt
1381 tcctcttgcg tccgcatacc ccgcgccact gggtcgtctt cattcagccg ccgcaccgtg
1441 cgcttacctc ccttgccgtg cttgattagc accggtgggt tgctgaaacc caccatttgt
1501 agcgccacat cttctctttc ttcctcgctg tccacgatca cctctgggga tggcgggcgc
1561 tcgggcttgg gagaggggcg cttctttttc tttttggacg caatggccaa atccgccgtc
1621 gaggtcgatg gccgcgggct gggtgtgcgc ggcaccagcg catcttgtga cgagtcttct
1681 tcgtcctcgg actcgagacg ccgcctcagc cgcttttttg ggggcgcgcg gggaggcggc
1741 ggcgacggcg acggggacga cacgtcctcc atggttggtg gacgtcgcgc cgcaccgcgt
1801 ccgcgctcgg gggtggtttc gcgctgctcc tcttcccgac tggccatttc cttctcctat
1861 aggcagaaaa agatcatgga gtcagtcgag aaggaggaca gcctaaccgc ccccttttgag
1921 ttcgccacca ccgcctccac cgatgccgcc aacgcgccta ccaccttccc cgtcgaggca
1981 cccccgcttg aggaggagga agtgattatc gagcaggacc caggttttgt aagcgaagac
2041 gacgaggatc gctcagtacc aacagaggat aaaaagcaag accaggacga cgcagaggca
2101 aacgaggaac aagtcgggcg gggggaccaa aggcatggcg actacctaga tgtgggagac
2161 gacgtgctgt tgaagcatct gcagcgccag tgcgccatta tctgcgacgc gttgcaagag
2221 cgcagcgatg tgcccctcgc catagcggat gtcagccttg cctacgaacg ccacctgttc
2281 tcaccgcgcg taccccccaa acgccaagaa aacggcacat gcgagcccaa cccgcgcctc
2341 aacttctacc ccgtatttgc cgtgccagag gtgcttgcca cctatcacat ctttttccaa
2401 aactgcaaga taccctatc ctgccgtgcc aaccgcagcc gagcggacaa gcagctggcc
2461 ttgcggcagg gcgctgtcat acctgatatc gcctcgctcg acgaagtgcc aaaaatcttt
2521 gagggtcttg gacgcgacga gaaacgcgcg gcaaacgctc tgcaacaaga aaacagcgaa
2581 aatgaaagtc actgtggagt gctggtggaa cttgagggtg acaacgcgcg cctagccgtg
2641 ctgaaacgca gcatcgaggt cacccacttt gcctacccgg cacttaacct acccccaag
2701 gttatgagca cagtcatgag cgagctgatc gtgcgccgtg cacgacccct ggagagggat
2761 gcaaacttgc aagaacaaac cgaggagggc ctacccgcag ttggcgatga gcagctggcg
2821 cgctggcttg agacgcgcga gcctgccgac ttggaggagc gacgcaagct aatgatggcc
2881 gcagtgcttg ttaccgtgga gcttgagtgc atgcagcggt tctttgctga cccggagatg
2941 cagcgcaagc tagaggaaac gttgcactac acctttcgcc agggctacgt gcgccaggcc
3001 tgcaaaattt ccaacgtgga gctctgcaac ctggtctcct accttggaat tttgcacgaa
3061 aaccgcctcg ggcaaaacgt gcttcattcc acgctcaagg gcgaggcgcg ccgcgactac
3121 gtccgcgact gcgtttactt atttctgtgc tacacctggc aaacggccat gggcgtgtgg
3181 cagcaatgcc tggaggagcg caacctaaag gagctgcaga agctgctaaa gcaaaacttg
3241 aaggacctat ggacggcctt caacgagcgc tccgtggccg cgcacctggc ggacattatc
3301 ttccccgaac gcctgcttaa aaccctgcaa cagggtctgc cagacttcac cagtcaaagc
3361 atgttgcaaa actttaggaa ctttatccta gagcgttcag gaattctgcc cgccacctgc
3421 tgtgcgcttc ctagcgactt tgtgcccatt aagtaccgtg aatgccctcc gccgctttgg
3481 ggtcactgct accttctgca gctagccaac taccttgcct accactccga catcatggaa
3541 gacgtgagcg gtgacggcct actggagtgt cactgtcgct gcaacctatg caccccgcac
3601 cgctccctgg tctgcaattc gcaactgctt agcgaaagtc aaattatcgg tacctttgag
3661 ctgcagggtc cctcgcctga cgaaaagtcc gcggctccgg ggttgaaact cactccgggg
3721 ctgtggacgt cggcttacct tcgcaaattt gtacctgagg actaccacgc ccacgagatt
3781 aggttctacg aagaccaatc ccgcccgcca aatgcggagc ttaccgcctg cgtcattacc
3841 cagggccaca tccttggcca attgcaagcc atcaacaaag cccgccaaga gtttctgcta
3901 cgaaagggac ggggggttta cctggacccc cagtccggcg aggagctcaa cccaatcccc
3961 ccgccgccgc agccctatca gcagccgcgg gcccttgctt cccaggatgg cacccaaaaa
4021 gaagctgcag ctgccgccgc cgccacccac ggacgaggag gaatactggg acagtcaggc
4081 agaggaggtt ttggacgagg aggaggagat gatggaagac tgggacagcc tagacgaagc
4141 ttccgaggcc gaagaggtgt cagacgaaac accgtcaccc tcggtcgcat tcccctcgcc
4201 ggcgccccag aaattggcaa ccgttcccag catcgctaca acctccgctc ctcaggcgcc
4261 gccggcactg cctgttcgcc gacccaaccg tagatgggac accactggaa ccagggccgg
4321 taagtctaag cagccgccgc cgttagccca agagcaacaa cagcgccaag gctaccgctc
4381 gtggcgcggg cacaagaacg ccatagttgc ttgcttgcaa gactgtgggg gcaacatctc
4441 cttcgcccgc cgctttcttc tctaccatca cggcgtggcc ttcccccgta acatcctgca
4501 ttactaccgt catctctaca gcccctactg caccggcggc agcggcagcg gcagcaacag
4561 cagcggtcac acagaagcaa aggcgaccgg atagcaagac tctgacaaag cccaagaaat
4621 ccacagcggc ggcagcagca ggaggaggag cgctgcgtct ggcgcccaac gaacccgtat
4681 cgacccgcga gcttagaaat aggatttttc ccactctgta tgctatattt caacaaagca
```

```
4741 ggggccaaga acaagagctg aaaataaaaa acaggtctct gcgctccctc acccgcagct
4801 gcctgtatca caaaagcgaa gatcagcttc ggcgcacgct ggaagacgcg gaggctctct
4861 tcagcaaata ctgcgcgctg actcttaagg actagtttcg cgccctttct caaatttaag
4921 cgcgaaaact acgtcatctc cagcggccac acccggcgcc agcacctgtc gtcagcgcca
4981 ttatgagcaa ggaaattccc acgccctaca tgtggagtta ccagccacaa atgggacttg
5041 cggctggagc tgcccaagac tactcaaccc gaataaacta catgagcgcg ggaccccaca
5101 tgatatcccg ggtcaacgga atccgcgccc accgaaaccg aattctcctc gaacaggcgg
5161 ctattaccac cacacctcgt aataacctta atccccgtag ttggcccgct gccctggtgt
5221 accaggaaag tcccgctccc accactgtgg tacttcccag agacgcccag gccgaagttc
5281 agatgactaa ctcaggggcg cagcttgcgg gcggctttcg tcacagggtg cggtcgcccg
5341 ggcgttttag ggcggagtaa cttgcatgta ttgggaattg tagtttttt aaaatgggaa
5401 gtgacgtatc gtgggaaaac ggaagtgaag atttgaggaa gttgtgggtt ttttggcttt
5461 cgtttctggg cgtaggttcg cgtgcggttt tctgggtgtt ttttgtggac tttaaccgtt
5521 acgtcatttt ttagtcctat atatactcgc tctgtacttg gccctttta cactgtgact
5581 gattgagctg gtgccgtgtc gagtggtgtt ttttaatagg ttttttact ggtaaggctg
5641 actgttatgg ctgccgctgt ggaagcgctg tatgttgttc tggagcggga gggtgctatt
5701 ttgcctaggc aggagggttt ttcaggtgtt tatgtgtttt tctctcctat taattttgtt
5761 atacctccta tgggggctgt aatgttgtct ctacgcctgc gggtatgtat tcccccgggc
5821 tatttcggtc gctttttagc actgaccgat gttaaccaac ctgatgtgtt taccgagtct
5881 tacattatga ctccggacat gaccgaggaa ctgtcggtgg tgctttttaa tcacggtgac
5941 cagtttttt acggtcacgc cggcatggcc gtagtccgtc ttatgcttat aagggttgtt
6001 tttcctgttg taagacaggc ttctaatgtt taaatgtttt tttttttgtt attttatttt
6061 gtgtttaatg caggaacccg cagacatgtt tgagagaaaa atggtgtctt tttctgtggt
6121 ggttccggaa cttacctgcc tttatctgca tgagcatgac tacgatgtgc ttgctttttt
6181 gcgcgaggct ttgcctgatt ttttgagcag caccttgcat tttatatcgc cgcccatgca
6241 acaagcttac ataggggcta cgctggttag catagctccg agtatgcgtg tcataatcag
6301 tgtgggttct tttgtcatgg ttcctggcgg ggaagtggcc gcgctggtcc gtgcagacct
6361 gcacgattat gttcagctgg ccctgcgaag ggacctacgg gatcgcggta tttttgttaa
6421 tgttccgctt ttgaatctta tacaggtctg tgaggaacct gaatttttgc aatcatgatt
6481 cgctgcttga ggctgaaggt ggagggcgct ctggagcaga tttttacaat ggccggactt
6541 aatattcggg atttgcttag agacatattg ataaggtggc gagatgaaaa ttatttgggc
6601 atggttgaag gtgctggaat gtttatagag gagattcacc ctgaagggtt tagcctttac
6661 gtccacttgg acgtgagggc agtttgcctt ttggaagcca ttgtgcaaca tcttacaaat
6721 gccattatct gttctttggc tgtagagttt gaccacgcca ccggagggga gcgcgttcac
6781 ttaatagatc ttcattttga ggttttggat aatcttttgg aataaaaaaa aaaaaacatg
6841 gttcttccag ctcttcccgc tcctcccgtg tgtgactcgc agaacgaatg tgtaggttgg
6901 ctgggtgtgg cttattctgc ggtggtggat gttatcaggg cagcggcgca tgaaggagtt
6961 tacatagaac ccgaagccag ggggcgcctg gatgctttga gagagtggat atactacaac
7021 tactacacag agcgagctaa gcgacgagac cggagacgca gatctgtttg tcacgcccgc
7081 acctggtttt gcttcaggaa atatgactac gtccggcgtt ccatttggca tgacactacg
7141 accaacacga tctcggttgt ctcggcgcac tccgtacagt agggatcgcc tacctccttt
7201 tgagacagag acccgcgcta ccatactgga ggatcatccg ctgctgcccg aatgtaacac
7261 tttgacaatg cacaacgtga gttacgtgcg aggtcttccc tgcagtgtgg gatttacgct
7321 gattcaggaa tgggttgttc cctgggatat ggttctgacg cgggaggagc ttgtaatcct
7381 gaggaagtgt atgcacgtgt gcctgtgttg tgccaacatt gatatcatga cgagcatgat
7441 gatccatggt tacgagtcct gggctctcca ctgtcattgt tccagtcccg gttccctgca
7501 gtgcatagcc ggcgggcagg ttttggccag ctggtttagg atggtggtgg atggcgccat
7561 gtttaatcag aggtttatat ggtaccggga ggtggtgaat tacaacatgc caaaagaggt
7621 aatgtttatg tccagcgtgt ttatgagggg tcgccactta atctacctgc gcttgtggta
7681 tgatggccac gtgggttctg tggtccccgc catgagcttt ggatacagcg ccttgcactg
7741 tgggattttg aacaatattg tggtgctgtg ctgcagttac tgtgctgatt taagtgagat
7801 cagggtgcgc tgctgtgccc ggaggacaag gcgtctcatg ctgcgggcgg tgcgaatcat
7861 cgctgaggag accactgcca tgttgtattc ctgcaggacg gagcggcggc ggcagcagtt
7921 tattcgcgcg ctgctgcagc accaccgccc tatcctgatg cacgattatg actctacccc
7981 catgtaggcg tggacttccc cttcgccgcc cgttgagcaa ccgcaagttg gacagcagcc
8041 tgtggctcag cagctggaca gcgacatgaa cttaagcgag ctgcccgggg agtttattaa
8101 tatcactgat gagcgtttgg ctcgacagga aaccgtgtgg aatataacac ctaagaatat
8161 gtctgttacc catgatatga tgctttttaa ggccagccgg ggagaaagga ctgtgtactc
8221 tgtgtgttgg gagggaggtg gcaggttgaa tactagggtt ctgtgagttt gattaaggta
8281 cggtgatcaa tataagctat gtggtggtgg ggctatacta ctgaatgaaa aatgacttga
8341 aattttctgc aattgaaaaa taaacacgtt gaaacataac atgcaacagg ttcacgattc
8401 tttattcctg ggcaatgtag gagaaggtgt aagagttggt agcaaaagtt tcagtggtgt
8461 attttccact ttcccaggac catgtaaaag acatagagta agtgcttacc tcgctagttt
8521 ctgtggattc actagaatcg atgtaggatg ttgcccctcc tgacgcggta ggagaagggg
8581 agggtgccct gcatgtctgc cgctgctctt gctcttgccg ctgctgagga ggggggcgca
8641 tctgccgcag caccggatgc atctgggaaa agcaaaaaag gggctcgtcc ctgtttccgg
8701 aggaatttgc aagcggggtc ttgcatgacg gggaggcaaa cccccgttcg ccgcagtccg
8761 gccggcccga gactcgaacc gggggtcctg cgactcaacc cttggaaaat aaccctccgg
8821 ctacagggag cgagccactt aatgctttcg ctttccagcc taaccgctta cgccgcgcgc
8881 ggccagtggc caaaaaagct agcgcagcag ccgccgcgcc tggaaggaag ccaaaaggag
```

```
     8941 cgctccccg ttgtctgacg tcgcacacct gggttcgaca cgcgggcggt aaccgcatgg
     9001 atcacggcgg acggccggat ccggggttcg aaccccggtc gtccgccatg ataccttgc
     9061 gaatttatcc accagaccac ggaagagtgc ccgcttacag gctctccttt tgcacggtct
     9121 agagcgtcaa cgactgcgca cgcctcaccg gccagagcgt cccgaccatg gagcactttt
     9181 tgccgctgcg caacatctgg aaccgcgtcc gcgactttcc gcgcgcctcc accaccgccg
     9241 ccggcatcac ctggatgtcc aggtacatct acggattacg tcgacgttta aacggatctg
     9301 atcagcacgt gttgacaatt aatcatcggc atagtatatc ggcatagtat aatacgacaa
     9361 ggtgaggaac taaaccatgg ccaagccttt gtctcaagaa gaatccaccc tcattgaaag
     9421 agcaacggct acaatcaaca gcatccccat ctctgaagac tacagcgtcg ccagcgcagc
     9481 tctctctagc gacggccgca tcttcactgg tgtcaatgta tatcatttta ctgggggacc
     9541 ttgtgcagaa ctcgtggtgc tgggcactgc tgctgctgcg gcagctggca acctgacttg
     9601 tatcgtcgcg atcggaaatg agaacagggg catcttgagc ccctgcggac ggtgccgaca
     9661 ggtgcttctc gatctgcatc ctgggatcaa agccatagtg aaggacagtg atggacagcc
     9721 gacggcagtt gggattcgtg aattgctgcc ctctggttat gtgtgggagg gctaagcact
     9781 tcgtggccga ggagcaggac tgacacgtgc tacgagatca tatgatcagc tcactcaaag
     9841 gcggtaatac ggttatccac agaatcaggg gataacgcag gaaagaacat gtgagcaaaa
     9901 ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc
     9961 cgccccccctg acgagcatca caaaaatcga cgctcaagtc agaggtggcg aaacccgaca
    10021 ggactataaa gataccaggc gtttcccccct ggaagctccc tcgtgcgctc tcctgttccg
    10081 accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt ggcgctttct
    10141 catagctcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt
    10201 gtgcacgaac cccccgttca gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag
    10261 tccaacccgg taagacacga cttatcgcca ctggcagcag ccactggtaa caggattagc
    10321 agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa ctacggctac
    10381 actagaagaa cagtatttgg tatctgcgct ctgctgaagc cagttacctt cggaaaaaga
    10441 gttggtagct cttgatccgg caaacaaacc accgctggta gcggtggttt ttttgtttgc
    10501 aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag atcctttgat cttttctacg
    10561 gggtctgacg ctcagtggaa cgaaaactca cgttaaggga ttttggtcat gagattatca
    10621 aaaaggatct tcacctagat ccttttaaat taaaaatgaa gttttaaatc aatctaaagt
    10681 atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgaggc acctatctca
    10741 gcgatctgtc tatttcgttc atccatagtt gcctgactcc ccgtcgtgta gataactacg
    10801 atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga cccacgctca
    10861 ccggctccag atttatcagc aataaaccag ccagccggaa gggccgagcg cagaagtggt
    10921 cctgcaactt tatccgcctc catccagtct attaattgtt gccgggaagc tagagtaagt
    10981 agttcgccag ttaatagttt gcgcaacgtt gttgccattg ctacaggcat cgtggtgtca
    11041 cgctcgtcgt ttggtatggc ttcattcagc tccggttccc aacgatcaag gcgagttaca
    11101 tgatccccca tgttgtgcaa aaaagcggtt agctccttcg gtcctccgat cgttgtcaga
    11161 agtaagttgg ccgcagtgtt atcactcatg gttatggcag cactgcataa ttctcttact
    11221 gtcatgccat ccgtaagatg cttttctgtg actggtgagt actcaaccaa gtcattctga
    11281 gaatagtgta tgcggcgacc gagttgctct tgcccggcgt caatacggga taataccgcg
    11341 ccacatagca gaactttaaa agtgctcatc attggaaaac gttcttcggg gcgaaaactc
    11401 tcaaggatct taccgctgtt gagatccagt tcgatgtaac ccactcgtgc acccaactga
    11461 tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg aaggcaaaat
    11521 gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa tactcatact cttccttttt
    11581 caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt
    11641 atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccacctaaa
    11701 ttgtaagcgt taatattttg ttaaaattcg cgttaaattt ttgttaaatc agctcatttt
    11761 ttaaccaata ggccgaaatc ggcaaaatcc cttataaatc aaaagaatag accgagatag
    11821 ggttgagtgt tgttccagtt tggaacaaga gtccactatt aaagaacgtg gactccaacg
    11881 tcaaagggcg aaaaaccgtc tatcagggcg atggcccact acgtgaacca tcaccctaat
    11941 caagtttttt ggggtcgagg tgccgtaaag cactaaatcg gaaccctaaa gggagccccc
    12001 gatttagagc ttgacgggga aagccggcga acgtggcgag aaaggaaggg aagaaagcga
    12061 aaggagcggg cgctagggcg ctggcaagtg tagcggtcac gctgcgcgta accaccacac
    12121 ccgccgcgct taatgcgccg ctacagggcg cgatggatcc
//
```

10.2 **Poster**

10.2.1 **Poster – DECHEMA Frankfurt am Main – 30-31 January 2019**

Universität Bielefeld

synTbio

rAAV2 capsid protein modification, expression and stability

Julian Teschner[1]; Rebecca C. Feiner[1]; Kathrin E. Teschner[1]; Marco T. Radukic[1]; Yvonne Hertle[2]; Niklas Biere[3]; Dario Anselmetti[3]; Kristian M. Müller[1*]

[1] Department of Cellular and Molecular Biotechnology, Bielefeld University, Bielefeld, Germany,
[2] Physical and Biophysical Chemistry (PCIII), Bielefeld University, Germany
[3] Experimental Biophysics and Applied Nanoscience, Bielefeld University, Germany
[*] Corresponding author: kristian@syntbio.net

Recombinant adeno-associated viruses (rAAV) capsid modifications, specifically loop modifications, can endow the gene delivery vehicle with new and purposefully designed tropisms or functions. To rationalize the rAAV-2 design process, we systematically evaluated the influence of size of flexible loop insertions (2 to 16 aa at the 587 position) on the thermal capsid stability and transduction ability. Even the insertion of a 29 kDa enzyme with its N- and C-terminus in close proximity yielded particles. In this context, mosaic capsids comprising wild-type and altered proteins may be desired. We demonstrate that the reported VP3 start knock out is not sufficient to suppress VP3 expression in mosaic context and that an additional Kozak sequence does the trick while requiring a lower gene dosage. In this work, all genetic manipulations and viral particle productions were performed with our modular, BioBrick compatible plasmid kit. To simplify production and reduce costs, we also generated a HEK293 cell line with stably integrated helper functionality.

Viral construction kit

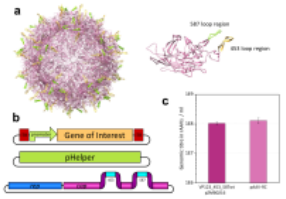


Fig. 1: (a) Crystal structure of the rAAV capsid and VP3 (PDB: 1LP3). (b) rAAV production using our helper-free system. (c) Production of rAAVs with our plasmid system compared to Agilents pAAV-RC plasmid system

- Our plasmid set consists of several plasmid for rAAV2 cap mutations and GOI assembly
- Motifs can be inserted into the loop sequences (loop 453 and loop 587) of the cap genes by singular restriction sites
- rAAV2 vectors are produced by triple transfection in HEK293 cells and purified using established methods (iodixanol gradient ultracentrifugation)
- The performance of our production system is comparable to the commercially available helper-free system by Agilent

ITR sequencing

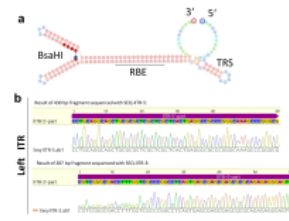


Fig. 2: Preparation of fragments suitable for sequencing of ITRs. (a) left ITR stem loop structure. (b) Sequencing results of digested fragments showing expected left-ITR sequence

- We established an ITR sequencing procedure to verify correct ITR sequences
- After separating lITR and rITR a BsaHI digest breaks up sequence parts which forms an Y-shaped stem loop structure if the sequence is single-stranded during Sanger DNA-sequencing

Capsid stability of rAAV with loop insertions

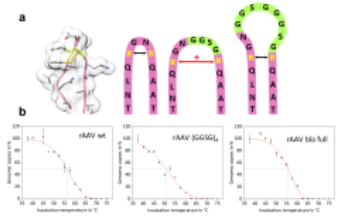


Fig. 3: Analysis of rAAV capsid stability. (a) Structural model of the 587 loop with yellow highlighted R585 and R588 and schematic model of 587 wild-type loop region (pink), 587 loop region with GGSG linker (light green). (b) Thermal stability assay of different rAAV particles measured in PBS. Fitting curves (red) were calculated using a logistic function to determine the disintegration temperature (T_d) for all rAAV variants.

Tab. 1: Transduction ability (of HT1080 cells) and thermal stability of rAAV2 variants with GGSG linker or β-lactamase insertion at residue position 587.

Loop variant	Transduction ability in %	T_d in °C
rAAV2 wt	97	56.1
rAAV2_587_GG	31	56.4
rAAV2_587_GGSG	20	54.7
rAAV2_587_$(GGSG)_3$	28	53.8
rAAV2_587_$(GGSG)_4$	11	50.7
rAAV2 bla full	1	55.6

- 587 loop regions with small GGSG linker insertions leading to an increased arginine-arginine-distance whereas big linker insertions leading to a normal arginine-arginine-distance but steric shielding of HSPG binding motifs
- Disintegration temperature (T_d): Temperature at which 50 % of all rAAVs release their DNA
- Td is calculated from qPCR data. Previously all samples were incubated separatly at different temperatures and then DNaseI digestions were performed
- An increased (GGSG) linker insertion decreases T_d but β-lactamase insertion is almost stable like rAAV wt
- rAAV2 bla full can not transduce HT1080 cells

rAAV presenting 60 β-lactamases on capsid surface

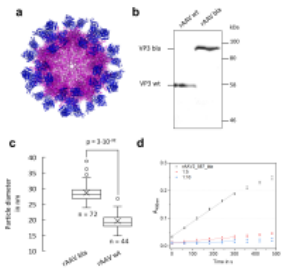


Fig. 4: (a) Theoretical structure of rAAV bla showing AAV2 wt structure (purple, PDB: 1LP3) and ß-lactamase (blue, PDB: 3DTM). (b) Western blot of purified rAAV wt and rAAV bla with B1 antibody. (c) AFM particle size analysis of rAAV wt and rAAV bla. (d) Nitrocefin assay, probing the β-lactamase activity.

- rAAV presenting 60 β-lactamases on its capsid surface are producible with our system
- Expected mass shift of VP3 capsid proteins ($VP3_{wt}$ = 60.1 kDa and $VP3_{bla}$ = 88.9 kDa) was confirmed by western blot
- The calculated mean diameter of rAAV wt is 20 nm compared to 29 nm for rAAV bla in atomic force microscopy
- Nitrocefin assay confirmed β-lactamase activity, calculated β-lactamase activity was equivalent to 26.9 per DNaseI resistant particle. In addition, the β-lactamase activity was also confirmed by a self-made bacterial assay

Optimizing mosaic rAAV production

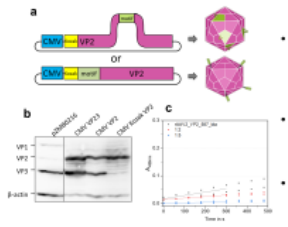


Fig. 5: (a) A fourth plasmid allows the production of mosaic rAAVs with modified loop regions (453 or 587) or displaying proteins on the capsid using a N-terminal fusion to VP2. (b) Western blot analysis of plasmids expressing VP2 with or without Kozak sequence. (c) Nitrocefin assay confirmed β-lactamase activity

- Mosaic rAAV production needs a fourth plasmid, which carries the motif sequence either in the loop region or N-terminal of the VP2 protein.
- Tight control of VP2 protein expression is needed to guarantee the correct mosaic rAAV production.
- Combination of a strong Kozak sequence and a VP3 start codon knock out enabled sole VP2 expression
- For rAAV2_VP2_587_bla the number of active β-lactamases was determined to be about 5 in nitrocefin assay

"HEK293 KARE" with stably integrated helper functionality

- Three cell lines based on HEK293 for the production of rAAV were generated by stable genomic integration of adenovirus genes from pHelper plasmid with blasticidin resistance cassette and named HEK KARE1a-c
- Elevated relative expression level of both adenoviral early genes (E2A and E4) in HEK293-KARE clones compared to HEK293 has been detected
- HEK KARE cells are able to produce rAAV over a long period of passages
- Total DNA amount used for transfection can be reduced by 50 % without significant loss of productivity, because the Helper plasmid is not longer needed

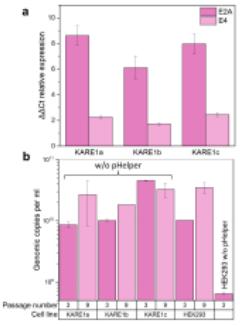


Fig. 6: (a) Identification of the expression of adenoviral elements E2A and E4 in HEK KARE1a-c cells. (b) Results of rAAV wt productions using HEK KARE1a-c and normal HEK293 cells, over a period of passages, with and without Helper plasmid.

Acknowledgement

The authors would like to thank Philipp Borchert for his experimental assistance and support in the cloning work.

References

• S. Hagen et al. Modular adeno-associated virus (rAAV) vectors used for cellular virus-directed enzyme prodrug therapy, *Sci. Rep.* 4., **2014**, 3759.

10.3 **DNA sequences**

In the following subchapters a selection of the most important DNA sequences of this thesis is shown. This should help to understand the described cloning steps and plasmid constructs more easily. A complete listing of all plasmid constructs is available in the Geneious database of the working group.

10.3.1 **Plasmid sequences**

10.3.1.1 **pHelper sequence**

```
            10        20        30        40        50        60        70        80        90        100
             |         |         |         |         |         |         |         |         |         |
  1 ggtacccaactccatgcttaacagtccccaggtacagcccaccctgcgtcgcaaccaggaacagctctacagcttcctggagcgccactcgccctacttc
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

            110       120       130       140       150       160       170       180       190       200
             |         |         |         |         |         |         |         |         |         |
 101 cgcagccacagtgcgcagattaggagcgccacttcttttttgtcacttgaaaaacatgtaaaaaataatgtactaggagacactttcaataaaggcaaatgt
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

            210       220       230       240       250       260       270       280       290       300
             |         |         |         |         |         |         |         |         |         |
 201 ttttatttgtacactctcgggtgattatttacccccacccttgccgtctgcgccgtttaaaaatcaaaggggttctgccgcgcatcgctatgcgccact
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

            310       320       330       340       350       360       370       380       390       400
             |         |         |         |         |         |         |         |         |         |
 301 ggcagggacacgttgcgatactggtgtttagtgctccacttaaactcaggcacaaccatccgcggcagctcggtgaagttttcactccacaggctgcgca
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

            410       420       430       440       450       460       470       480       490       500
             |         |         |         |         |         |         |         |         |         |
 401 ccatcaccaacgcgtttagcaggtcgggcgccgatatcttgaagtcgcagttggggcctccgccctgcgcgcgcgagttgcgatacacagggttgcagca
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

            510       520       530       540       550       560       570       580       590       600
             |         |         |         |         |         |         |         |         |         |
 501 ctggaacactatcagcgccgggtggtgcacgctggccagcacgctcttgtcggagatcagatccgcgtccaggtcctccgcgttgctcagggcgaacgga
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

            610       620       630       640       650       660       670       680       690       700
             |         |         |         |         |         |         |         |         |         |
 601 gtcaactttggtagctgccttcccaaaaagggtgcatgcccaggctttgagttgcactcgcaccgtagtggcatcagaaggtgaccgtgcccggtctggg
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

            710       720       730       740       750       760       770       780       790       800
             |         |         |         |         |         |         |         |         |         |
 701 cgttaggatacagcgcctgcatgaaagccttgatctgcttaaaagccacctgagcctttgcgccttcagagaagaacatgccgcaagacttgccggaaaa
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

            810       820       830       840       850       860       870       880       890       900
             |         |         |         |         |         |         |         |         |         |
 801 ctgattggccggacaggccgcgtcatgcacgcagcaccttgcgtcggtgttggagatctgcaccacatttcggccccaccggttcttcacgatcttggcc
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

            910       920       930       940       950       960       970       980       990      1000
             |         |         |         |         |         |         |         |         |         |
 901 ttgctagactgctccttcagcgcgcgctgcccgttttcgctcgtcacatccatttcaatcacgtgctccttatttatcataatgctcccgtgtagacact
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

           1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
             |         |         |         |         |         |         |         |         |         |
1001 taagctcgccttcgatctcagcgcagcggtgcagccacaacgcgcagcccgtgggctcgtggtgcttgtaggttacctctgcaaacgactgcaggtacgc
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)

           1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
             |         |         |         |         |         |         |         |         |         |
1101 ctgcaggaatcgccccatcatcgtcacaaaggtcttgttgctggtgaaggtcagctgcaacccgcggtgctcctcgtttagccaggtcttgcatacggcc
    <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
    E2A (Ad2, similar to Ad5)
```

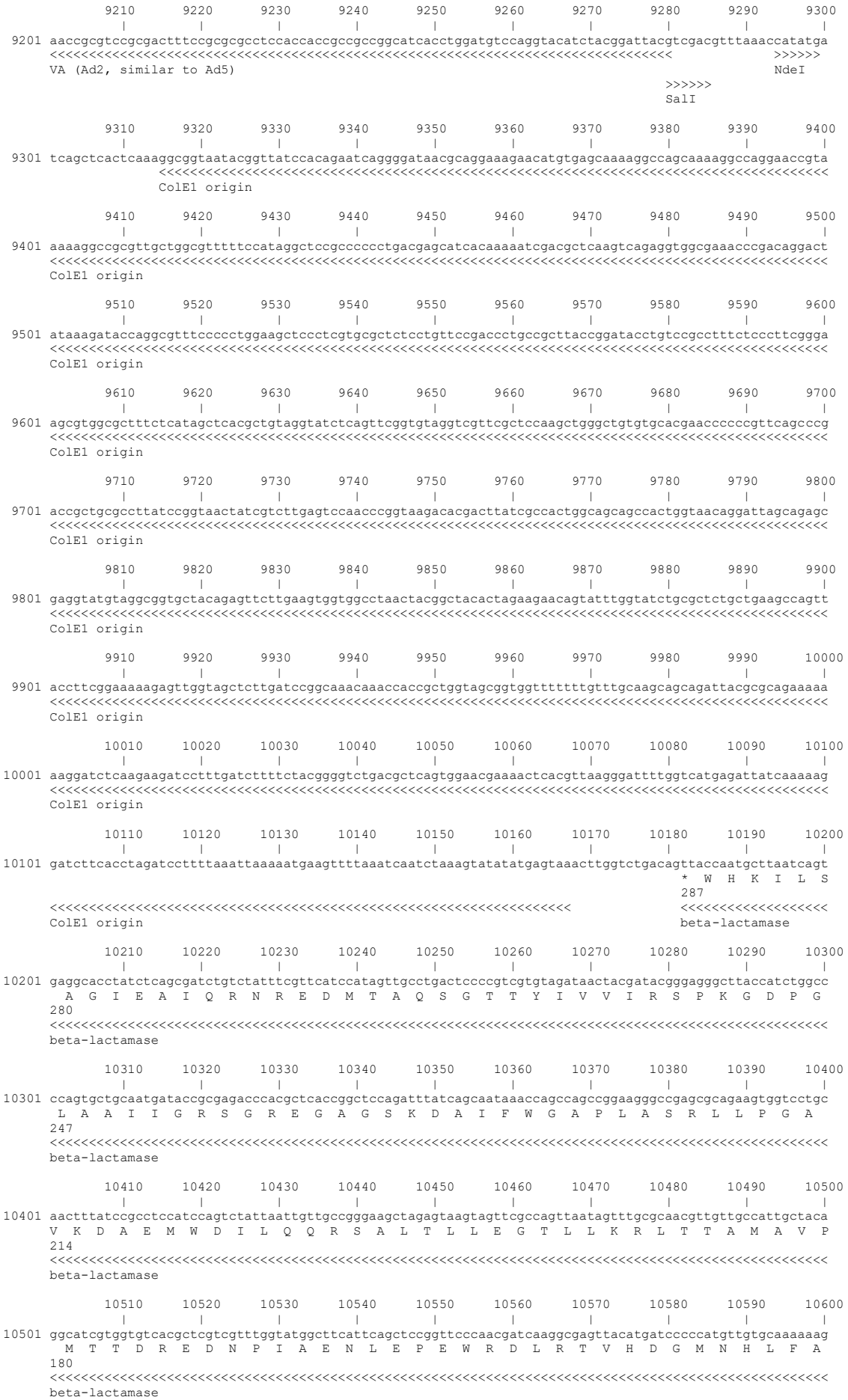
```
            1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
              |         |         |         |         |         |         |         |         |         |
1201 gccagagcttccacttggtcaggcagtagcttgaagtttgcctttagatcgttatccacgtggtacttgtccatcaacgcgcgcgcagcctccatgccct
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
              |         |         |         |         |         |         |         |         |         |
1301 tctcccacgcagacacgatcggcaggctcagcgggtttatcaccgtgctttcactttccgcttcactggactcttccttttcctcttgcgtccgcatacc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
              |         |         |         |         |         |         |         |         |         |
1401 ccgcgccactgggtcgtcttcattcagccgccgcaccgtgcgcttacctcccttgccgtgcttgattagcaccggtgggttgctgaaacccaccatttgt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
              |         |         |         |         |         |         |         |         |         |
1501 agcgccacatcttctctttcttcctcgctgtccacgatcacctctggggatggcgggcgctcgggcttgggagaggggcgcttcttttttcttttggacg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
              |         |         |         |         |         |         |         |         |         |
1601 caatggccaaatccgccgtcgaggtcgatggccgcgggctgggtgtgcgcggcaccagcgcatcttgtgacgagtcttcttcgtcctcggactcgagacg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
              |         |         |         |         |         |         |         |         |         |
1701 ccgcctcagccgctttttttgggggcgcgcggggaggcggcggcgacggcgacggggacgacacgtcctccatggttggtggacgtcgcgccgcaccgcgt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
              |         |         |         |         |         |         |         |         |         |
1801 ccgcgctcgggggtggtttcgcgctgctcctcttcccgactggccatttccttctcctataggcagaaaaagatcatggagtcagtcgagaaggaggaca
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
              |         |         |         |         |         |         |         |         |         |
1901 gcctaaccgccccctttgagttcgccaccaccgcctccaccgatgccgccaacgcgcctaccaccttccccgtcgaggcaccccgcttgaggaggagga
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
              |         |         |         |         |         |         |         |         |         |
2001 agtgattatcgagcaggacccaggttttgtaagcgaagacgacgaggatcgctcagtaccaacagaggataaaaagcaagaccaggacgacgcagaggca
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
              |         |         |         |         |         |         |         |         |         |
2101 aacgaggaacaagtcgggcgggggggaccaaaggcatggcgactacctagatgtgggagacgacgtgctgttgaagcatctgcagcgccagtgcgccatta
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
              |         |         |         |         |         |         |         |         |         |
2201 tctgcgacgcgttgcaagagcgcagcgatgtgcccctcgccatagcggatgtcagccttgcctacgaacgccacctgttctcaccgcgcgtacccccccaa
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
              |         |         |         |         |         |         |         |         |         |
2301 acgccaagaaaacggcacatgcgagcccaacccgcgcctcaacttctaccccgtatttgccgtgccagaggtgcttgccacctatcacatctttttccaa
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
              |         |         |         |         |         |         |         |         |         |
2401 aactgcaagataccctatcctgccgtgccaaccgcagccgagcggacaagcagctggccttgcggcagggcgctgtcatacctgatatcgcctcgctcg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            2510      2520      2530      2540      2550      2560      2570      2580      2590      2600
              |         |         |         |         |         |         |         |         |         |
2501 acgaagtgccaaaaatctttgagggtcttggacgcgacgagaaacgcgcggcaaacgctctgcaacaagaaaacagcgaaaatgaaagtcactgtggagt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            2610      2620      2630      2640      2650      2660      2670      2680      2690      2700
              |         |         |         |         |         |         |         |         |         |
2601 gctggtggaacttgagggtgacaacgcgcgcctagccgtgctgaaacgcagcatcgaggtcacccactttgcctacccggcacttaacctacccccaag
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            2710      2720      2730      2740      2750      2760      2770      2780      2790      2800
              |         |         |         |         |         |         |         |         |         |
2701 gttatgagcacagtcatgagcgagctgatcgtgcgccgtgcacgacccctggagagggatgcaaacttgcaagaacaaaccgaggagggcctacccgcag
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)
```

```
            2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
              |         |         |         |         |         |         |         |         |         |
2801 ttggcgatgagcagctggcgcgctggcttgagacgcgcgagcctgccgacttggaggagcgacgcaagctaatgatggccgcagtgcttgttaccgtgga
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
              |         |         |         |         |         |         |         |         |         |
2901 gcttgagtgcatgcagcggttctttgctgacccggagatgcagcgcaagctagaggaaacgttgcactacacctttcgccagggctacgtgcgccaggcc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            3010      3020      3030      3040      3050      3060      3070      3080      3090      3100
              |         |         |         |         |         |         |         |         |         |
3001 tgcaaaatttccaacgtggagctctgcaacctggtctcctaccttggaattttgcacgaaaaccgcctcgggcaaaacgtgcttcattccacgctcaagg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            3110      3120      3130      3140      3150      3160      3170      3180      3190      3200
              |         |         |         |         |         |         |         |         |         |
3101 gcgaggcgcgccgcgactacgtccgcgactgcgtttacttatttctgtgctacacctggcaaacggccatgggcgtgtggcagcaatgcctggaggagcg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            3210      3220      3230      3240      3250      3260      3270      3280      3290      3300
              |         |         |         |         |         |         |         |         |         |
3201 caacctaaaggagctgcagaagctgctaaagcaaaacttgaaggacctatggacggccttcaacgagcgctccgtggcccgcgcacctggcggacattatc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            3310      3320      3330      3340      3350      3360      3370      3380      3390      3400
              |         |         |         |         |         |         |         |         |         |
3301 ttccccgaacgcctgcttaaaaccctgcaacagggtctgccagacttcaccagtcaaagcatgttgcaaaactttaggaactttatcctagagcgttcag
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            3410      3420      3430      3440      3450      3460      3470      3480      3490      3500
              |         |         |         |         |         |         |         |         |         |
3401 gaattctgcccgccacctgctgtgcgcttcctagcgactttgtgcccattaagtaccgtgaatgccctccgccgctttggggtcactgctaccttctgca
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            3510      3520      3530      3540      3550      3560      3570      3580      3590      3600
              |         |         |         |         |         |         |         |         |         |
3501 gctagccaactaccttgcctaccactccgacatcatggaagacgtgagcggtgacggcctactggagtgtcactgtcgctgcaacctatgcaccccgcac
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            3610      3620      3630      3640      3650      3660      3670      3680      3690      3700
              |         |         |         |         |         |         |         |         |         |
3601 cgctccctggtctgcaattcgcaactgcttagcgaaagtcaaattatcggtacctttgagctgcagggtccctcgcctgacgaaaagtccgcggctccgg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            3710      3720      3730      3740      3750      3760      3770      3780      3790      3800
              |         |         |         |         |         |         |         |         |         |
3701 ggttgaaactcactccggggctgtggacgtcggcttaccttcgcaaatttgtacctgaggactaccacgcccacgagattaggttctacgaagaccaatc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            3810      3820      3830      3840      3850      3860      3870      3880      3890      3900
              |         |         |         |         |         |         |         |         |         |
3801 ccgcccgccaaatgcggagcttaccgcctgcgtcattacccagggccacatccttggccaattgcaagccatcaacaaagcccgccaagagtttctgcta
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            3910      3920      3930      3940      3950      3960      3970      3980      3990      4000
              |         |         |         |         |         |         |         |         |         |
3901 cgaaagggacgggggggttacctggacccccagtccggcgaggagctcaacccaatccccccgccgccgcagccctatcagcagccgcgggcccttgctt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            4010      4020      4030      4040      4050      4060      4070      4080      4090      4100
              |         |         |         |         |         |         |         |         |         |
4001 cccaggatggcacccaaaaagaagctgcagctgccgccgccgccacccacggacgaggaggaatactgggacagtcaggcagaggaggttttggacgagg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            4110      4120      4130      4140      4150      4160      4170      4180      4190      4200
              |         |         |         |         |         |         |         |         |         |
4101 aggaggagatgatggaagactgggacagcctagacgaagcttccgaggccgaagaggtgtcagacgaaacaccgtcaccctcggtcgcattcccctcgcc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            4210      4220      4230      4240      4250      4260      4270      4280      4290      4300
              |         |         |         |         |         |         |         |         |         |
4201 ggcgccccagaaattggcaaccgttcccagcatcgctacaacctccgctcctcaggcgccgccggcactgcctgttcgccgacccaaccgtagatgggac
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            4310      4320      4330      4340      4350      4360      4370      4380      4390      4400
              |         |         |         |         |         |         |         |         |         |
4301 accactggaaccagggccggtaagtctaagcagccgccgccgttagcccaagagcaacaacagcgccaaggctaccgctcgtggcgcgggcacaagaacg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)
```

```
            4410      4420      4430      4440      4450      4460      4470      4480      4490      4500
              |         |         |         |         |         |         |         |         |         |
4401 ccatagttgcttgcttgcaagactgtgggggcaacatctccttcgcccgccgctttcttctctaccatcacggcgtggccttcccccgtaacatcctgca
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            4510      4520      4530      4540      4550      4560      4570      4580      4590      4600
              |         |         |         |         |         |         |         |         |         |
4501 ttactaccgtcatctctacagcccctactgcaccggcggcagcggcagcggcagcaacagcagcggtcacacagaagcaaaggcgaccggatagcaagac
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            4610      4620      4630      4640      4650      4660      4670      4680      4690      4700
              |         |         |         |         |         |         |         |         |         |
4601 tctgacaaagcccaagaaatccacagcggcggcagcagcaggaggaggagcgctgcgtctggcgcccaacgaacccgtatcgacccgcgagcttagaaat
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            4710      4720      4730      4740      4750      4760      4770      4780      4790      4800
              |         |         |         |         |         |         |         |         |         |
4701 aggattttcccactctgtatgctatatttcaacaaagcaggggccaagaacaagagctgaaataaaaaacaggtctctgcgctccctcacccgcagct
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            4810      4820      4830      4840      4850      4860      4870      4880      4890      4900
              |         |         |         |         |         |         |         |         |         |
4801 gcctgtatcacaaaagcgaagatcagcttcggcgcacgctggaagacgcggaggctctcttcagcaaatactgcgcgctgactcttaaggactagtttcg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            4910      4920      4930      4940      4950      4960      4970      4980      4990      5000
              |         |         |         |         |         |         |         |         |         |
4901 cgccctttctcaaatttaagcgcgaaaactacgtcatctccagcggccacacccggcgccagcacctgtcgtcagcgccattatgagcaaggaaattccc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            5010      5020      5030      5040      5050      5060      5070      5080      5090      5100
              |         |         |         |         |         |         |         |         |         |
5001 acgccctacatgtggagttaccagccacaaatgggacttgcggctggagctgcccaagactactcaacccgaataaactacatgagcgcgggaccccaca
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            5110      5120      5130      5140      5150      5160      5170      5180      5190      5200
              |         |         |         |         |         |         |         |         |         |
5101 tgatatcccgggtcaacggaatccgcgcccaccgaaaccgaattctcctcgaacaggcggctattaccaccacacctcgtaataaccttaatccccgtag
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            5210      5220      5230      5240      5250      5260      5270      5280      5290      5300
              |         |         |         |         |         |         |         |         |         |
5201 ttggcccgctgccctggtgtaccaggaaagtcccgctcccaccactgtggtacttcccagagacgcccaggccgaagttcagatgactaactcaggggcg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)

            5310      5320      5330      5340      5350      5360      5370      5380      5390      5400
              |         |         |         |         |         |         |         |         |         |
5301 cagcttgcgggcggctttcgtcacagggtgcggtcgcccgggcgttttagggcggagtaacttgcatgtattgggaattgtagtttttttaaaatgggaa
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E2A (Ad2, similar to Ad5)
                                        <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
                                        E4 (Ad2, similar to Ad5)

            5410      5420      5430      5440      5450      5460      5470      5480      5490      5500
              |         |         |         |         |         |         |         |         |         |
5401 gtgacgtatcgtgggaaaacggaagtgaagatttgaggaagttgtgggtttttttggctttcgtttctgggcgtaggttcgcgtgcggttttctgggtgtt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            5510      5520      5530      5540      5550      5560      5570      5580      5590      5600
              |         |         |         |         |         |         |         |         |         |
5501 ttttgtggactttaaccgttacgtcatttttttagtcctatatatactcgctctgtacttggcccttttttacactgtgactgattgagctggtgccgtgtc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            5610      5620      5630      5640      5650      5660      5670      5680      5690      5700
              |         |         |         |         |         |         |         |         |         |
5601 gagtggtgtttttaataggttttttactggtaaggctgactgttatggctgccgctgtggaagcgctgtatgttgttctggagcgggagggtgctatt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            5710      5720      5730      5740      5750      5760      5770      5780      5790      5800
              |         |         |         |         |         |         |         |         |         |
5701 ttgcctaggcaggagggtttttcaggtgtttatgtgtttttctctcctattaattttgttatacctcctatgggggctgtaatgttgtctctacgcctgc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            5810      5820      5830      5840      5850      5860      5870      5880      5890      5900
              |         |         |         |         |         |         |         |         |         |
5801 gggtatgtattcccccgggctatttcggtcgctttttagcactgaccgatgttaaccaacctgatgtgtttaccgagtcttacattatgactccggacat
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            5910      5920      5930      5940      5950      5960      5970      5980      5990      6000
              |         |         |         |         |         |         |         |         |         |
5901 gaccgaggaactgtcggtggtgcttttttaatcacggtgaccagttttttttacggtcacgccggcatggccgtagtccgtcttatgcttataagggttgtt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)
```

```
            6010      6020      6030      6040      6050      6060      6070      6080      6090      6100
              |         |         |         |         |         |         |         |         |         |
6001 tttcctgttgtaagacaggcttctaatgtttaaatgttttttttttttgttatttttattttgtgtttaatgcaggaacccgcagacatgtttgagagaaaa
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            6110      6120      6130      6140      6150      6160      6170      6180      6190      6200
              |         |         |         |         |         |         |         |         |         |
6101 atggtgtctttttctgtggtggttccggaacttacctgcctttatctgcatgagcatgactacgatgtgcttgctttttgcgcgaggctttgcctgatt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            6210      6220      6230      6240      6250      6260      6270      6280      6290      6300
              |         |         |         |         |         |         |         |         |         |
6201 ttttgagcagcaccttgcattttatatcgccgcccatgcaacaagcttacataggggctacgctggttagcatagctccgagtatgcgtgtcataatcag
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            6310      6320      6330      6340      6350      6360      6370      6380      6390      6400
              |         |         |         |         |         |         |         |         |         |
6301 tgtgggttcttttgtcatggttcctggcggggaagtggccgcgctggtccgtgcagacctgcacgattatgttcagctggccctgcgaagggacctacgg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            6410      6420      6430      6440      6450      6460      6470      6480      6490      6500
              |         |         |         |         |         |         |         |         |         |
6401 gatcgcggtatttttgttaatgttccgcttttgaatcttatacaggtctgtgaggaacctgaatttttgcaatcatgattcgctgcttgaggctgaaggt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            6510      6520      6530      6540      6550      6560      6570      6580      6590      6600
              |         |         |         |         |         |         |         |         |         |
6501 ggagggcgctctggagcagatttttacaatggccggacttaatattcgggatttgcttagagacatattgataaggtggcgagatgaaaattatttgggc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            6610      6620      6630      6640      6650      6660      6670      6680      6690      6700
              |         |         |         |         |         |         |         |         |         |
6601 atggttgaaggtgctggaatgtttatagaggagattcaccctgaagggtttagcctttacgtccacttggacgtgagggcagtttgcctttttggaagcca
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            6710      6720      6730      6740      6750      6760      6770      6780      6790      6800
              |         |         |         |         |         |         |         |         |         |
6701 ttgtgcaacatcttacaaatgccattatctgttctttggctgtagagtttgaccacgccaccggaggggagcgcgttcacttaatagatcttcattttga
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            6810      6820      6830      6840      6850      6860      6870      6880      6890      6900
              |         |         |         |         |         |         |         |         |         |
6801 ggttttggataatctttttggaataaaaaaaaaaaaaacatggttcttccagctcttcccgctcctcccgtgtgtgactcgcagaacgaatgtgtaggttgg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            6910      6920      6930      6940      6950      6960      6970      6980      6990      7000
              |         |         |         |         |         |         |         |         |         |
6901 ctgggtgtggcttattctgcggtggtggatgttatcagggcagcggcgcatgaaggagtttacatagaacccgaagccaggggggcgcctggatgctttga
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            7010      7020      7030      7040      7050      7060      7070      7080      7090      7100
              |         |         |         |         |         |         |         |         |         |
7001 gagagtggatatactacaactactacacagagcgagctaagcgacgagaccggagacgcagatctgtttgtcacgcccgcacctggttttgcttcaggaa
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            7110      7120      7130      7140      7150      7160      7170      7180      7190      7200
              |         |         |         |         |         |         |         |         |         |
7101 atatgactacgtccggcgttccatttggcatgacactacgaccaacacgatctcggttgtctcggcgcactccgtacagtagggatcgcctacctccttt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            7210      7220      7230      7240      7250      7260      7270      7280      7290      7300
              |         |         |         |         |         |         |         |         |         |
7201 tgagacagagacccgcgctaccatactggaggatcatccgctgctgcccgaatgtaacactttgacaatgcacaacgtgagttacgtgcgaggtcttccc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            7310      7320      7330      7340      7350      7360      7370      7380      7390      7400
              |         |         |         |         |         |         |         |         |         |
7301 tgcagtgtgggatttacgctgattcaggaatgggttgttccctgggatatggttctgacgcgggaggagcttgtaatcctgaggaagtgtatgcacgtgt
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            7410      7420      7430      7440      7450      7460      7470      7480      7490      7500
              |         |         |         |         |         |         |         |         |         |
7401 gcctgtgttgtgccaacattgatatcatgacgagcatgatgatccatggttacgagtcctgggctctccactgtcattgttccagtcccggttccctgca
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            7510      7520      7530      7540      7550      7560      7570      7580      7590      7600
              |         |         |         |         |         |         |         |         |         |
7501 gtgcatagccggcgggcaggttttggccagctggtttaggatggtggtggatggcgccatgtttaatcagaggtttatatggtaccgggaggtggtgaat
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)
```

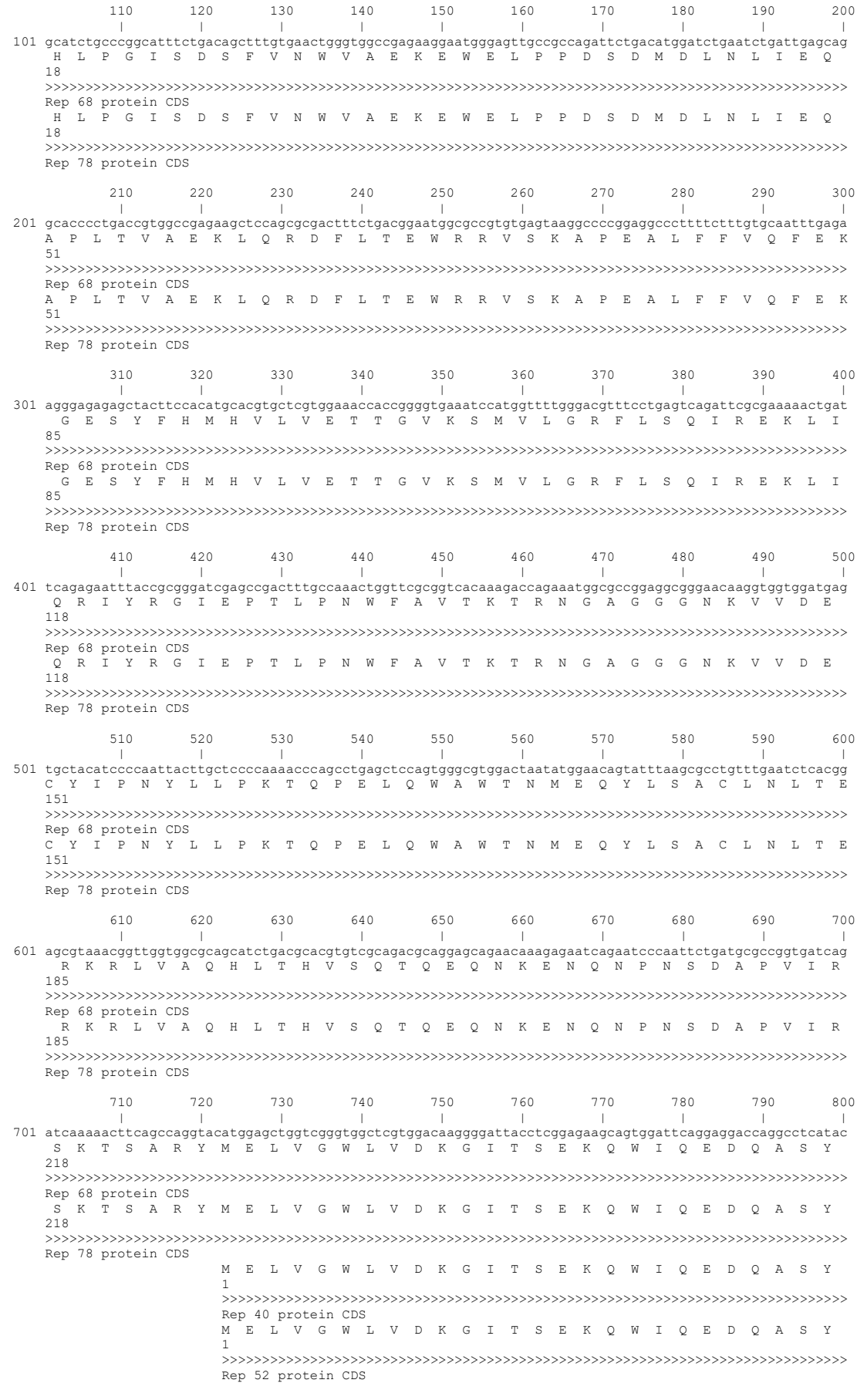
```
            7610      7620      7630      7640      7650      7660      7670      7680      7690      7700
              |         |         |         |         |         |         |         |         |         |
7601 tacaacatgccaaaagaggtaatgtttatgtccagcgtgtttatgaggggtcgccacttaatctacctgcgcttgtggtatgatggccacgtgggttctg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            7710      7720      7730      7740      7750      7760      7770      7780      7790      7800
              |         |         |         |         |         |         |         |         |         |
7701 tggtccccgccatgagctttggatacagcgccttgcactgtgggattttgaacaatattgtggtgctgtgctgcagttactgtgctgatttaagtgagat
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            7810      7820      7830      7840      7850      7860      7870      7880      7890      7900
              |         |         |         |         |         |         |         |         |         |
7801 cagggtgcgctgctgtgcccggaggacaaggcgtctcatgctgcgggcggtgcgaatcatcgctgaggagaccactgccatgttgtattcctgcaggacg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            7910      7920      7930      7940      7950      7960      7970      7980      7990      8000
              |         |         |         |         |         |         |         |         |         |
7901 gagcggcggcggcagcagtttattcgcgcgctgctgcagcaccaccgccctatcctgatgcacgattatgactctacccccatgtaggcgtggacttccc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            8010      8020      8030      8040      8050      8060      8070      8080      8090      8100
              |         |         |         |         |         |         |         |         |         |
8001 cttcgccgcccgttgagcaaccgcaagttggacagcagcctgtggcttcagcagctggacagcgacatgaacttaagcgagctgcccggggagtttattaa
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            8110      8120      8130      8140      8150      8160      8170      8180      8190      8200
              |         |         |         |         |         |         |         |         |         |
8101 tatcactgatgagcgtttggctcgacaggaaaccgtgtggaatataacacctaagaatatgtctgttacccatgatatgatgctttttaaggccagccgg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            8210      8220      8230      8240      8250      8260      8270      8280      8290      8300
              |         |         |         |         |         |         |         |         |         |
8201 ggagaaaggactgtgtactctgtgtgttgggagggaggtggcaggttgaatactagggttctgtgagtttgattaaggtacggtgatcaatataagctat
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            8310      8320      8330      8340      8350      8360      8370      8380      8390      8400
              |         |         |         |         |         |         |         |         |         |
8301 gtggtggtggggctatactactgaatgaaaaatgacttgaaatttttctgcaattgaaaaataaacacgttgaaacataacatgcaacaggttcacgattc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            8410      8420      8430      8440      8450      8460      8470      8480      8490      8500
              |         |         |         |         |         |         |         |         |         |
8401 tttattcctgggcaatgtaggagaaggtgtaagagttggtagcaaaagtttcagtggtgtattttccactttcccaggaccatgtaaaagacatagagta
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)

            8510      8520      8530      8540      8550      8560      8570      8580      8590      8600
              |         |         |         |         |         |         |         |         |         |
8501 agtgcttacctcgctagtttctgtggattcactagaatcgatgtaggatgttgcccctcctgacgcggtaggagaaggggagggtgccctgcatgtctgc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     E4 (Ad2, similar to Ad5)
                                     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
                                     VA (Ad2, similar to Ad5)

            8610      8620      8630      8640      8650      8660      8670      8680      8690      8700
              |         |         |         |         |         |         |         |         |         |
8601 cgctgctcttgctcttgccgctgctgaggaggggggcgcatctgccgcagcaccggatgcatctgggaaaagcaaaaaaggggctcgtccctgtttccgg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     VA (Ad2, similar to Ad5)

            8710      8720      8730      8740      8750      8760      8770      8780      8790      8800
              |         |         |         |         |         |         |         |         |         |
8701 aggaatttgcaagcggggtcttgcatgacggggaggcaaaccccccgttcgccgcagtccggccggcccgagactcgaaccgggggtcctgcgactcaacc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     VA (Ad2, similar to Ad5)

            8810      8820      8830      8840      8850      8860      8870      8880      8890      8900
              |         |         |         |         |         |         |         |         |         |
8801 cttggaaaataaccctccggctacagggagcgagccacttaatgctttcgctttccagcctaaccgcttacgccgcgcgcggccagtggccaaaaaagct
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     VA (Ad2, similar to Ad5)

            8910      8920      8930      8940      8950      8960      8970      8980      8990      9000
              |         |         |         |         |         |         |         |         |         |
8901 agcgcagcagccgccgcgcctggaaggaagccaaaaggagcgctcccccgttgtctgacgtcgcacacctgggttcgacacgcgggcggtaaccgcatgg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     VA (Ad2, similar to Ad5)

            9010      9020      9030      9040      9050      9060      9070      9080      9090      9100
              |         |         |         |         |         |         |         |         |         |
9001 atcacggcggacggccggatccggggttcgaaccccggtcgtccgccatgatacccttgcgaatttatccaccagaccacggaagagtgcccgcttacag
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     VA (Ad2, similar to Ad5)

            9110      9120      9130      9140      9150      9160      9170      9180      9190      9200
              |         |         |         |         |         |         |         |         |         |
9101 gctctccttttgcacggtctagagcgtcaacgactgcgcacgcctcaccggccagagcgtcccgaccatggagcacttttgccgctgcgcaacatctgg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     VA (Ad2, similar to Ad5)
```


9210 9220 9230 9240 9250 9260 9270 9280 9290 9300
9201 aaccgcgtccgcgactttccgcgcgcctccaccaccgccgccggcatcacctggatgtccaggtacatctacggattacgtcgacgtttaaaccatatga
VA (Ad2, similar to Ad5)
NdeI
SalI
9310 9320 9330 9340 9350 9360 9370 9380 9390 9400
9301 tcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgta
ColE1 origin
9410 9420 9430 9440 9450 9460 9470 9480 9490 9500
9401 aaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggact
ColE1 origin
9510 9520 9530 9540 9550 9560 9570 9580 9590 9600
9501 ataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcggga
ColE1 origin
9610 9620 9630 9640 9650 9660 9670 9680 9690 9700
9601 agcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaacccccgttcagcccg
ColE1 origin
9710 9720 9730 9740 9750 9760 9770 9780 9790 9800
9701 accgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagc
ColE1 origin
9810 9820 9830 9840 9850 9860 9870 9880 9890 9900
9801 gaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagtatttggtatctgcgctctgctgaagccagtt
ColE1 origin
9910 9920 9930 9940 9950 9960 9970 9980 9990 10000
9901 accttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaa
ColE1 origin
10010 10020 10030 10040 10050 10060 10070 10080 10090 10100
10001 aaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaag
ColE1 origin
10110 10120 10130 10140 10150 10160 10170 10180 10190 10200
10101 gatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagt
* W H K I L S
287
ColE1 origin
beta-lactamase
10210 10220 10230 10240 10250 10260 10270 10280 10290 10300
10201 gaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggcc
A G I E A I Q R N R E D M T A Q S G T T Y I V V I R S P K G D P G
280
beta-lactamase
10310 10320 10330 10340 10350 10360 10370 10380 10390 10400
10301 ccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgc
L A A A I I G R S G R E G A G S K D A I F W G A P L A S R L L P G A
247
beta-lactamase
10410 10420 10430 10440 10450 10460 10470 10480 10490 10500
10401 aactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctaca
V K D A E M W D I L Q Q R S A L T L L E G T L L K R L T T A M A V P
214
beta-lactamase
10510 10520 10530 10540 10550 10560 10570 10580 10590 10600
10501 ggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaag
M T T D R E D N P I A E N L E P E W R D L R T V H D G M N H L F A
180
beta-lactamase

```
            10610     10620     10630     10640     10650     10660     10670     10680     10690     10700
              |         |         |         |         |         |         |         |         |         |
10601 cggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctcttactgtcat
       T  L  E  K  P  G  G  I  T  T  L  L  L  N  A  A  T  N  D  S  M  T  I  A  A  S  C  L  E  R  V  T  M
      147
      <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
      beta-lactamase

            10710     10720     10730     10740     10750     10760     10770     10780     10790     10800
              |         |         |         |         |         |         |         |         |         |
10701 gccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaata
      G  D  T  L  H  K  E  T  V  P  S  Y  E  V  L  D  N  Q  S  Y  H  I  R  R  G  L  Q  E  Q  G  A  D  I  R
      114
      <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
      beta-lactamase

            10810     10820     10830     10840     10850     10860     10870     10880     10890     10900
              |         |         |         |         |         |         |         |         |         |
10801 cgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagat
        S  L  V  A  G  C  L  L  V  K  F  T  S  M  M  P  F  R  E  E  P  R  F  S  E  L  I  K  G  S  N  L  D
      80
      <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
      beta-lactamase

            10910     10920     10930     10940     10950     10960     10970     10980     10990     11000
              |         |         |         |         |         |         |         |         |         |
10901 ccagttcgatgtaacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgc
       L  E  I  Y  G  V  R  A  G  L  Q  D  E  A  D  K  V  K  V  L  T  E  P  H  A  F  V  P  L  C  F  A  A
      47
      <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
      beta-lactamase

            11010     11020     11030     11040     11050     11060     11070     11080     11090     11100
              |         |         |         |         |         |         |         |         |         |
11001 aaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttttcaatattattgaagcatttatcagggttattgtctcatgagcgga
      F  F  P  I  L  A  V  R  F  H  Q  I  S  M
      14
      <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
      beta-lactamase

            11110     11120     11130     11140     11150     11160     11170     11180     11190     11200
              |         |         |         |         |         |         |         |         |         |
11101 tacatatttgaatgtatttagaaaaataaacaaatagggqttccgcgcacatttccccgaaaagtgccacctaaattgtaagcgttaatattttgttaaa
                                                                          >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                                          f1 origin

            11210     11220     11230     11240     11250     11260     11270     11280     11290     11300
              |         |         |         |         |         |         |         |         |         |
11201 attcgcgttaaatttttgttaaatcagctcatttttttaaccaataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttg
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      f1 origin

            11310     11320     11330     11340     11350     11360     11370     11380     11390     11400
              |         |         |         |         |         |         |         |         |         |
11301 agtgttgttccagtttggaacaagagtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtg
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      f1 origin

            11410     11420     11430     11440     11450     11460     11470     11480     11490     11500
              |         |         |         |         |         |         |         |         |         |
11401 aaccatcaccctaatcaagttttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagcc
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      f1 origin

            11510     11520     11530     11540     11550     11560     11570     11580     11590     11600
              |         |         |         |         |         |         |         |         |         |
11501 ggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgcgtaaccaccacacccgcc
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      f1 origin

            11610     11620     11630
              |         |         |
11601 gcgcttaatgcgccgctacagggcgcgatggatcc
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      f1 origin
```

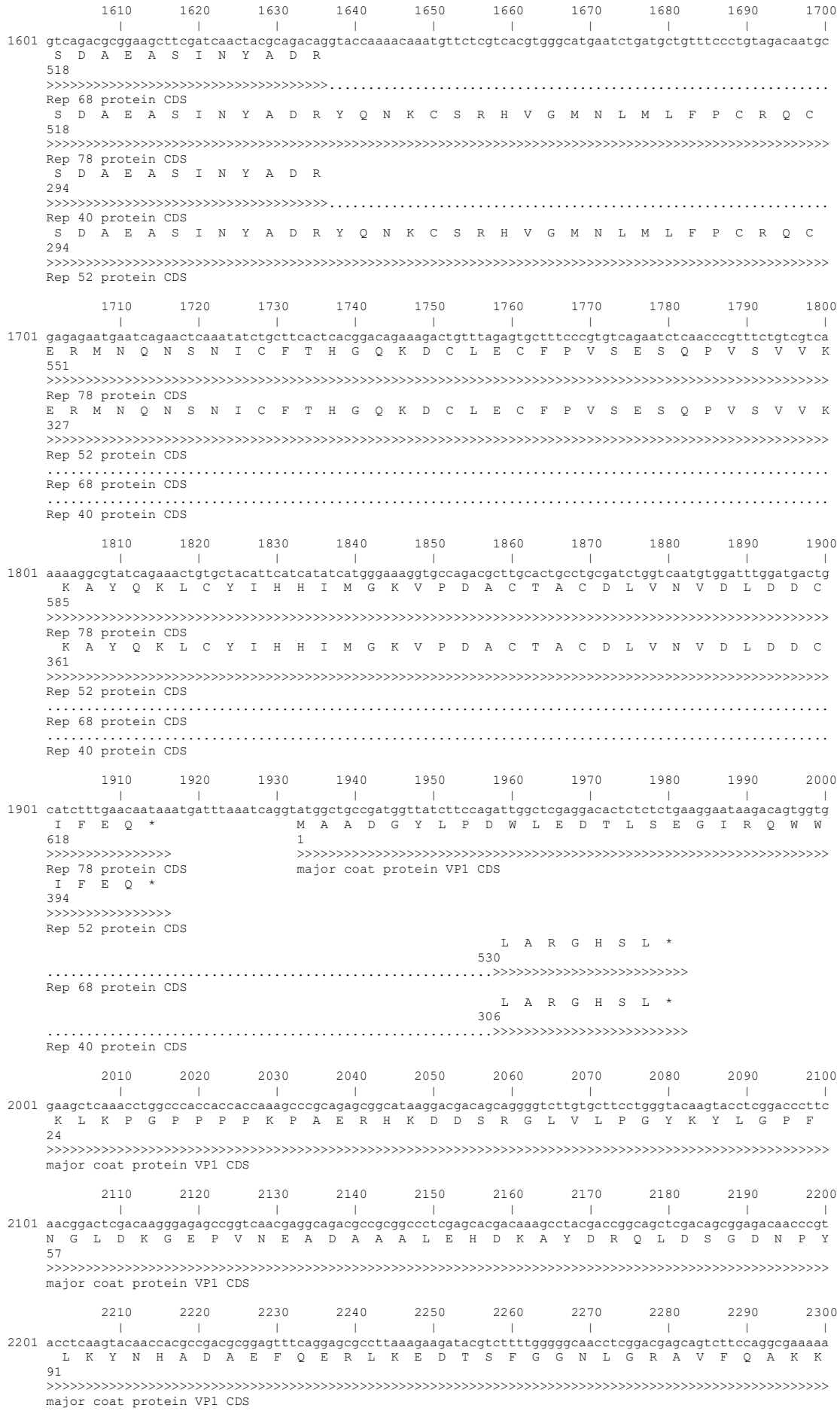
10.3.1.2 **pAAV-RC pZMB0216 sequence**

```
             10        20        30        40        50        60        70        80        90        100
              |         |         |         |         |         |         |         |         |         |
    1 atccttagctttcgctaaggatgatttctggaattcgcggccgcttctagatggcggggttttacgagattgtgattaaggtccccagcgaccttgacga
                                                M  A  G  F  Y  E  I  V  I  K  V  P  S  D  L  D  E
                                                1
                                                >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                Rep 68 protein CDS
                                                M  A  G  F  Y  E  I  V  I  K  V  P  S  D  L  D  E
                                                1
                                                >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                Rep 78 protein CDS
```


110 120 130 140 150 160 170 180 190 200
101 gcatctgcccggcatttctgacagctttgtgaactgggtggccgagaaggaatgggagttgccgccagattctgacatggatctgaatctgattgagcag
H L P G I S D S F V N W V A E K E W E L P P D S D M D L N L I E Q
18
Rep 68 protein CDS
H L P G I S D S F V N W V A E K E W E L P P D S D M D L N L I E Q
18
Rep 78 protein CDS
210 220 230 240 250 260 270 280 290 300
201 gcacccctgaccgtggccgagaagctccagcgcgactttctgacggaatggcgccgtgtgagtaaggccccggaggccctttttctttgtgcaatttgaga
A P L T V A E K L Q R D F L T E W R R V S K A P E A L F F V Q F E K
51
Rep 68 protein CDS
A P L T V A E K L Q R D F L T E W R R V S K A P E A L F F V Q F E K
51
Rep 78 protein CDS
310 320 330 340 350 360 370 380 390 400
301 agggagagagctacttccacatgcacgtgctcgtggaaaccaccggggtgaaatccatggttttgggacgtttcctgagtcagattcgcgaaaaactgat
G E S Y F H M H V L V E T T G V K S M V L G R F L S Q I R E K L I
85
Rep 68 protein CDS
G E S Y F H M H V L V E T T G V K S M V L G R F L S Q I R E K L I
85
Rep 78 protein CDS
410 420 430 440 450 460 470 480 490 500
401 tcagagaatttaccgcgggatcgagccgactttgccaaactggttcgcggtcacaaagaccagaaatggcgccggaggcgggaacaaggtggtggatgag
Q R I Y R G I E P T L P N W F A V T K T R N G A G G G N K V V D E
118
Rep 68 protein CDS
Q R I Y R G I E P T L P N W F A V T K T R N G A G G G N K V V D E
118
Rep 78 protein CDS
510 520 530 540 550 560 570 580 590 600
501 tgctacatccccaattacttgctccccaaaacccagcctgagctccagtgggcgtggactaatatggaacagtatttaagcgcctgtttgaatctcacgg
C Y I P N Y L L P K T Q P E L Q W A W T N M E Q Y L S A C L N L T E
151
Rep 68 protein CDS
C Y I P N Y L L P K T Q P E L Q W A W T N M E Q Y L S A C L N L T E
151
Rep 78 protein CDS
610 620 630 640 650 660 670 680 690 700
601 agcgtaaacggttggtggcgcagcatctgacgcacgtgtcgcagacgcaggagcagaacaaagagaatcagaatcccaattctgatgcgccggtgatcag
R K R L V A Q H L T H V S Q T Q E Q N K E N Q N P N S D A P V I R
185
Rep 68 protein CDS
R K R L V A Q H L T H V S Q T Q E Q N K E N Q N P N S D A P V I R
185
Rep 78 protein CDS
710 720 730 740 750 760 770 780 790 800
701 atcaaaaacttcagccaggtacatggagctggtcgggtggctcgtggacaaggggattacctcggagaagcagtggattcaggaggaccaggcctcatac
S K T S A R Y M E L V G W L V D K G I T S E K Q W I Q E D Q A S Y
218
Rep 68 protein CDS
S K T S A R Y M E L V G W L V D K G I T S E K Q W I Q E D Q A S Y
218
Rep 78 protein CDS
M E L V G W L V D K G I T S E K Q W I Q E D Q A S Y
1
Rep 40 protein CDS
M E L V G W L V D K G I T S E K Q W I Q E D Q A S Y
1
Rep 52 protein CDS

```
            810       820       830       840       850       860       870       880       890       900
             |         |         |         |         |         |         |         |         |         |
801 atctccttcaatgcggcctccaactcgcggtcccaaatcaaggctgccttggacaatgcgggaaagattatgagcctgactaaaaccgcccccgactacc
    I  S  F  N  A  A  S  N  S  R  S  Q  I  K  A  A  L  D  N  A  G  K  I  M  S  L  T  K  T  A  P  D  Y  L
    251
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 68 protein CDS
    I  S  F  N  A  A  S  N  S  R  S  Q  I  K  A  A  L  D  N  A  G  K  I  M  S  L  T  K  T  A  P  D  Y  L
    251
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 78 protein CDS
    I  S  F  N  A  A  S  N  S  R  S  Q  I  K  A  A  L  D  N  A  G  K  I  M  S  L  T  K  T  A  P  D  Y  L
    27
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 40 protein CDS
    I  S  F  N  A  A  S  N  S  R  S  Q  I  K  A  A  L  D  N  A  G  K  I  M  S  L  T  K  T  A  P  D  Y  L
    27
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 52 protein CDS

            910       920       930       940       950       960       970       980       990      1000
             |         |         |         |         |         |         |         |         |         |
901 tggtgggccagcagcccgtggaggacatttccagcaatcggatttataaaattttggaactaaacgggtacgatccccaatatgcggcttccgtctttct
      V  G  Q  Q  P  V  E  D  I  S  S  N  R  I  Y  K  I  L  E  L  N  G  Y  D  P  Q  Y  A  A  S  V  F  L
    285
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 68 protein CDS
      V  G  Q  Q  P  V  E  D  I  S  S  N  R  I  Y  K  I  L  E  L  N  G  Y  D  P  Q  Y  A  A  S  V  F  L
    285
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 78 protein CDS
      V  G  Q  Q  P  V  E  D  I  S  S  N  R  I  Y  K  I  L  E  L  N  G  Y  D  P  Q  Y  A  A  S  V  F  L
    61
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 40 protein CDS
      V  G  Q  Q  P  V  E  D  I  S  S  N  R  I  Y  K  I  L  E  L  N  G  Y  D  P  Q  Y  A  A  S  V  F  L
    61
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 52 protein CDS

           1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
             |         |         |         |         |         |         |         |         |         |
1001 gggatgggccacgaaaaagttcggcaagaggaacaccatctggctgtttgggcctgcaactaccgggaagaccaacatcgcggaggccatagcccacact
     G  W  A  T  K  K  F  G  K  R  N  T  I  W  L  F  G  P  A  T  T  G  K  T  N  I  A  E  A  I  A  H  T
    318
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 68 protein CDS
     G  W  A  T  K  K  F  G  K  R  N  T  I  W  L  F  G  P  A  T  T  G  K  T  N  I  A  E  A  I  A  H  T
    318
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 78 protein CDS
     G  W  A  T  K  K  F  G  K  R  N  T  I  W  L  F  G  P  A  T  T  G  K  T  N  I  A  E  A  I  A  H  T
    94
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 40 protein CDS
     G  W  A  T  K  K  F  G  K  R  N  T  I  W  L  F  G  P  A  T  T  G  K  T  N  I  A  E  A  I  A  H  T
    94
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 52 protein CDS

           1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
             |         |         |         |         |         |         |         |         |         |
1101 gtgcccttctacgggtgcgtaaactggaccaatgagaactttcccttcaacgactgtgtggacaagatggtgatctggtgggaggaggggaagatgaccg
    V  P  F  Y  G  C  V  N  W  T  N  E  N  F  P  F  N  D  C  V  D  K  M  V  I  W  W  E  E  G  K  M  T  A
    351
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 68 protein CDS
    V  P  F  Y  G  C  V  N  W  T  N  E  N  F  P  F  N  D  C  V  D  K  M  V  I  W  W  E  E  G  K  M  T  A
    351
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 78 protein CDS
    V  P  F  Y  G  C  V  N  W  T  N  E  N  F  P  F  N  D  C  V  D  K  M  V  I  W  W  E  E  G  K  M  T  A
    127
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 40 protein CDS
    V  P  F  Y  G  C  V  N  W  T  N  E  N  F  P  F  N  D  C  V  D  K  M  V  I  W  W  E  E  G  K  M  T  A
    127
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Rep 52 protein CDS
```

```
            1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
              |         |         |         |         |         |         |         |         |         |
1201 ccaaggtcgtggagtcggccaaagccattctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccagatagacccgactcccgtgat
       K  V  V  E  S  A  K  A  I  L  G  G  S  K  V  R  V  D  Q  K  C  K  S  S  A  Q  I  D  P  T  P  V  I
     385
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 68 protein CDS
       K  V  V  E  S  A  K  A  I  L  G  G  S  K  V  R  V  D  Q  K  C  K  S  S  A  Q  I  D  P  T  P  V  I
     385
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 78 protein CDS
       K  V  V  E  S  A  K  A  I  L  G  G  S  K  V  R  V  D  Q  K  C  K  S  S  A  Q  I  D  P  T  P  V  I
     161
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 40 protein CDS
       K  V  V  E  S  A  K  A  I  L  G  G  S  K  V  R  V  D  Q  K  C  K  S  S  A  Q  I  D  P  T  P  V  I
     161
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 52 protein CDS

            1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
              |         |         |         |         |         |         |         |         |         |
1301 cgtcacctccaacaccaacatgtgcgccgtgattgacgggaactcaacgaccttcgaacaccagcagccgttgcaagaccggatgttcaaatttgaactc
      V  T  S  N  T  N  M  C  A  V  I  D  G  N  S  T  T  F  E  H  Q  Q  P  L  Q  D  R  M  F  K  F  E  L
     418
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 68 protein CDS
      V  T  S  N  T  N  M  C  A  V  I  D  G  N  S  T  T  F  E  H  Q  Q  P  L  Q  D  R  M  F  K  F  E  L
     418
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 78 protein CDS
      V  T  S  N  T  N  M  C  A  V  I  D  G  N  S  T  T  F  E  H  Q  Q  P  L  Q  D  R  M  F  K  F  E  L
     194
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 40 protein CDS
      V  T  S  N  T  N  M  C  A  V  I  D  G  N  S  T  T  F  E  H  Q  Q  P  L  Q  D  R  M  F  K  F  E  L
     194
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 52 protein CDS

            1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
              |         |         |         |         |         |         |         |         |         |
1401 acccgccgtctggatcatgactttgggaaggtcaccaagcaggaagtcaaagacttttttccggtgggcaaaggatcacgtggttgaggtggagcatgagt
     T  R  R  L  D  H  D  F  G  K  V  T  K  Q  E  V  K  D  F  F  R  W  A  K  D  H  V  V  E  V  E  H  E  F
     451
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 68 protein CDS
     T  R  R  L  D  H  D  F  G  K  V  T  K  Q  E  V  K  D  F  F  R  W  A  K  D  H  V  V  E  V  E  H  E  F
     451
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 78 protein CDS
     T  R  R  L  D  H  D  F  G  K  V  T  K  Q  E  V  K  D  F  F  R  W  A  K  D  H  V  V  E  V  E  H  E  F
     227
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 40 protein CDS
     T  R  R  L  D  H  D  F  G  K  V  T  K  Q  E  V  K  D  F  F  R  W  A  K  D  H  V  V  E  V  E  H  E  F
     227
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 52 protein CDS

            1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
              |         |         |         |         |         |         |         |         |         |
1501 tctacgtcaaaaagggtggagccaagaaaagacccgcccccagtgacgcagatataagtgagcccaaacgggtgcgcgagtcagttgcgcagccatcgac
       Y  V  K  K  G  G  A  K  K  R  P  A  P  S  D  A  D  I  S  E  P  K  R  V  R  E  S  V  A  Q  P  S  T
     485
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 68 protein CDS
       Y  V  K  K  G  G  A  K  K  R  P  A  P  S  D  A  D  I  S  E  P  K  R  V  R  E  S  V  A  Q  P  S  T
     485
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 78 protein CDS
       Y  V  K  K  G  G  A  K  K  R  P  A  P  S  D  A  D  I  S  E  P  K  R  V  R  E  S  V  A  Q  P  S  T
     261
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 40 protein CDS
       Y  V  K  K  G  G  A  K  K  R  P  A  P  S  D  A  D  I  S  E  P  K  R  V  R  E  S  V  A  Q  P  S  T
     261
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Rep 52 protein CDS
```


1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
1601 gtcagacgcggaagcttcgatcaactacgcagacaggtaccaaaacaaatgttctcgtcacgtgggcatgaatctgatgctgtttccctgtagacaatgc
S D A E A S I N Y A D R
518
Rep 68 protein CDS
S D A E A S I N Y A D R Y Q N K C S R H V G M N L M L F P C R Q C
518
Rep 78 protein CDS
S D A E A S I N Y A D R
294
Rep 40 protein CDS
S D A E A S I N Y A D R Y Q N K C S R H V G M N L M L F P C R Q C
294
Rep 52 protein CDS
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
1701 gagagaatgaatcagaactcaaatatctgcttcactcacggacagaaagactgtttagagtgctttcccgtgtcagaatctcaacccgtttctgtcgtca
E R M N Q N S N I C F T H G Q K D C L E C F P V S E S Q P V S V V K
551
Rep 78 protein CDS
E R M N Q N S N I C F T H G Q K D C L E C F P V S E S Q P V S V V K
327
Rep 52 protein CDS
Rep 68 protein CDS
Rep 40 protein CDS
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
1801 aaaaggcgtatcagaaactgtgctacattcatcatatcatgggaaaggtgccagacgcttgcactgcctgcgatctggtcaatgtggatttggatgactg
K A Y Q K L C Y I H H I M G K V P D A C T A C D L V N V D L D D C
585
Rep 78 protein CDS
K A Y Q K L C Y I H H I M G K V P D A C T A C D L V N V D L D D C
361
Rep 52 protein CDS
Rep 68 protein CDS
Rep 40 protein CDS
1910 1920 1930 1940 1950 1960 1970 1980 1990 2000
1901 catctttgaacaataaatgatttaaatcaggtatggctgccgatggttatcttccagattggctcgaggacactctctctgaaggaataagacagtggtg
I F E Q * M A A D G Y L P D W L E D T L S E G I R Q W W
618 1
Rep 78 protein CDS major coat protein VP1 CDS
I F E Q *
394
Rep 52 protein CDS
L A R G H S L *
530
Rep 68 protein CDS
L A R G H S L *
306
Rep 40 protein CDS
2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
2001 gaagctcaaacctggcccaccaccaccaaagcccgcagagcggcataaggacgacagcaggggtcttgtgcttcctgggtacaagtacctcggaccccttc
K L K P G P P P P K P A E R H K D D S R G L V L P G Y K Y L G P F
24
major coat protein VP1 CDS
2110 2120 2130 2140 2150 2160 2170 2180 2190 2200
2101 aacggactcgacaagggagagccggtcaacgaggcagacgccgcggccctcgagcacgacaaagcctacgaccggcagctcgacagcggagacaacccgt
N G L D K G E P V N E A D A A A L E H D K A Y D R Q L D S G D N P Y
57
major coat protein VP1 CDS
2210 2220 2230 2240 2250 2260 2270 2280 2290 2300
2201 acctcaagtacaaccacgccgacgcggagtttcaggagcgccttaaagaagatacgtcttttgggggcaacctcggacgagcagtcttccaggcgaaaaa
L K Y N H A D A E F Q E R L K E D T S F G G N L G R A V F Q A K K
91
major coat protein VP1 CDS

```
          2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
            |         |         |         |         |         |         |         |         |         |
2301 gagggttcttgaacctctgggcctggttgaggaacctgttaagacggctccgggaaaaaagaggccggtagagcactctcctgtggagccagactcctcc
      R  V  L  E  P  L  G  L  V  E  E  P  V  K  T  A  P  G  K  K  R  P  V  E  H  S  P  V  E  P  D  S  S
     124
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
                                          T  A  P  G  K  K  R  P  V  E  H  S  P  V  E  P  D  S  S
                                          1
                                          >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                          major coat protein VP2 CDS

          2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
            |         |         |         |         |         |         |         |         |         |
2401 tcgggaaccggaaaggcgggccagcagcctgcaagaaaaagattgaatttggtcagactggagacgcagactcagtacctgacccccagcctctcggac
     S  G  T  G  K  A  G  Q  Q  P  A  R  K  R  L  N  F  G  Q  T  G  D  A  D  S  V  P  D  P  Q  P  L  G  Q
     157
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
     S  G  T  G  K  A  G  Q  Q  P  A  R  K  R  L  N  F  G  Q  T  G  D  A  D  S  V  P  D  P  Q  P  L  G  Q
     20
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
                                                      L  E  T  Q  T  Q  Y  L  T  P  S  L  S  D
                                                      1
                                                      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                      assembly activating protein AAP CDS

          2510      2520      2530      2540      2550      2560      2570      2580      2590      2600
            |         |         |         |         |         |         |         |         |         |
2501 agccaccagcagccccctctggtctgggaactaatacgatggctacaggcagtggcgcaccaatggcagacaataacgagggcgccgacggagtgggtaa
       P  P  A  A  P  S  G  L  G  T  N  T  M  A  T  G  S  G  A  P  M  A  D  N  N  E  G  A  D  G  V  G  N
     191
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
       P  P  A  A  P  S  G  L  G  T  N  T  M  A  T  G  S  G  A  P  M  A  D  N  N  E  G  A  D  G  V  G  N
     54
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
     S  H  Q  Q  P  P  L  V  W  E  L  I  R  W  L  Q  A  V  A  H  Q  W  Q  T  I  T  R  A  P  T  E  W  V  I
     15
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     assembly activating protein AAP CDS
                                       M  A  T  G  S  G  A  P  M  A  D  N  N  E  G  A  D  G  V  G  N
                                       1
                                       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                       major coat protein VP3 CDS

          2610      2620      2630      2640      2650      2660      2670      2680      2690      2700
            |         |         |         |         |         |         |         |         |         |
2601 ttcctcgggaaattggcattgcgattccacatggatgggcgacagagtcatcaccaccagcacccgaacctgggccctgcccacctacaacaaccacctc
      S  S  G  N  W  H  C  D  S  T  W  M  G  D  R  V  I  T  T  S  T  R  T  W  A  L  P  T  Y  N  N  H  L
     224
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
      S  S  G  N  W  H  C  D  S  T  W  M  G  D  R  V  I  T  T  S  T  R  T  W  A  L  P  T  Y  N  N  H  L
     87
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
       P  R  E  I  G  I  A  I  P  H  G  W  A  T  E  S  S  P  P  A  P  E  P  G  P  C  P  P  T  T  T  T  S
     49
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     assembly activating protein AAP CDS
      S  S  G  N  W  H  C  D  S  T  W  M  G  D  R  V  I  T  T  S  T  R  T  W  A  L  P  T  Y  N  N  H  L
     22
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS

          2710      2720      2730      2740      2750      2760      2770      2780      2790      2800
            |         |         |         |         |         |         |         |         |         |
2701 tacaaacaaatttccagccaatcaggagcctcgaacgacaatcactactttggctacagcacccctgggggtattttgacttcaacagattccactgcc
     Y  K  Q  I  S  S  Q  S  G  A  S  N  D  N  H  Y  F  G  Y  S  T  P  W  G  Y  F  D  F  N  R  F  H  C  H
     257
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
     Y  K  Q  I  S  S  Q  S  G  A  S  N  D  N  H  Y  F  G  Y  S  T  P  W  G  Y  F  D  F  N  R  F  H  C  H
     120
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
      T  N  K  F  P  A  N  Q  E  P  R  T  T  I  T  T  L  A  T  A  P  L  G  G  I  L  T  S  T  D  S  T  A
     82
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     assembly activating protein AAP CDS
     Y  K  Q  I  S  S  Q  S  G  A  S  N  D  N  H  Y  F  G  Y  S  T  P  W  G  Y  F  D  F  N  R  F  H  C  H
     55
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS
```

```
          2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
             |         |         |         |         |         |         |         |         |         |
2801 acttttcaccacgtgactggcaaagactcatcaacaacaactggggattccgacccaagagactcaacttcaagctctttaacattcaagtcaaagaggt
       F  S  P  R  D  W  Q  R  L  I  N  N  N  W  G  F  R  P  K  R  L  N  F  K  L  F  N  I  Q  V  K  E  V
     291
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
       F  S  P  R  D  W  Q  R  L  I  N  N  N  W  G  F  R  P  K  R  L  N  F  K  L  F  N  I  Q  V  K  E  V
     154
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
     T  F  H  H  V  T  G  K  D  S  S  T  T  T  G  D  S  D  P  R  D  S  T  S  S  S  L  T  F  K  S  K  R  S
     115
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     assembly activating protein AAP CDS
       F  S  P  R  D  W  Q  R  L  I  N  N  N  W  G  F  R  P  K  R  L  N  F  K  L  F  N  I  Q  V  K  E  V
     89
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS

          2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
             |         |         |         |         |         |         |         |         |         |
2901 cacgcagaatgacggtacgacgacgattgccaataaccttaccagcacggttcaggtgtttactgactcggagtaccagctcccgtacgtcctcggctcg
      T  Q  N  D  G  T  T  T  I  A  N  N  L  T  S  T  V  Q  V  F  T  D  S  E  Y  Q  L  P  Y  V  L  G  S
     324
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
      T  Q  N  D  G  T  T  T  I  A  N  N  L  T  S  T  V  Q  V  F  T  D  S  E  Y  Q  L  P  Y  V  L  G  S
     187
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
       R  R  M  T  V  R  R  R  L  P  I  T  L  P  A  R  F  R  C  L  L  T  R  S  T  S  S  R  T  S  S  A  R
     149
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     assembly activating protein AAP CDS
      T  Q  N  D  G  T  T  T  I  A  N  N  L  T  S  T  V  Q  V  F  T  D  S  E  Y  Q  L  P  Y  V  L  G  S
     122
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS

          3010      3020      3030      3040      3050      3060      3070      3080      3090      3100
             |         |         |         |         |         |         |         |         |         |
3001 gcgcatcaaggatgcctcccgccgttcccagcagacgtcttcatggtgccacagtatggatacctcaccctgaacaacgggagtcaggcagtaggacgct
     A  H  Q  G  C  L  P  P  F  P  A  D  V  F  M  V  P  Q  Y  G  Y  L  T  L  N  N  G  S  Q  A  V  G  R  S
     357
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
     A  H  Q  G  C  L  P  P  F  P  A  D  V  F  M  V  P  Q  Y  G  Y  L  T  L  N  N  G  S  Q  A  V  G  R  S
     220
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
      R  I  K  D  A  S  R  R  S  Q  Q  T  S  S  W  C  H  S  M  D  T  S  P  *
     182
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     assembly activating protein AAP CDS
     A  H  Q  G  C  L  P  P  F  P  A  D  V  F  M  V  P  Q  Y  G  Y  L  T  L  N  N  G  S  Q  A  V  G  R  S
     155
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS

          3110      3120      3130      3140      3150      3160      3170      3180      3190      3200
             |         |         |         |         |         |         |         |         |         |
3101 cttcattttactgcctggagtactttccttctcagatgctgcgtaccggaaacaactttaccttcagctacacttttgaggacgttcctttccacagcag
       S  F  Y  C  L  E  Y  F  P  S  Q  M  L  R  T  G  N  N  F  T  F  S  Y  T  F  E  D  V  P  F  H  S  S
     391
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
       S  F  Y  C  L  E  Y  F  P  S  Q  M  L  R  T  G  N  N  F  T  F  S  Y  T  F  E  D  V  P  F  H  S  S
     254
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
       S  F  Y  C  L  E  Y  F  P  S  Q  M  L  R  T  G  N  N  F  T  F  S  Y  T  F  E  D  V  P  F  H  S  S
     189
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS

          3210      3220      3230      3240      3250      3260      3270      3280      3290      3300
             |         |         |         |         |         |         |         |         |         |
3201 ctacgctcacagccagagtctggaccgtctcatgaatcctctcatcgaccaatatttgtattacttgagcagaacaaacactccaagtggtaccaccacg
      Y  A  H  S  Q  S  L  D  R  L  M  N  P  L  I  D  Q  Y  L  Y  Y  L  S  R  T  N  T  P  S  G  T  T  T
     424
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
      Y  A  H  S  Q  S  L  D  R  L  M  N  P  L  I  D  Q  Y  L  Y  Y  L  S  R  T  N  T  P  S  G  T  T  T
     287
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
      Y  A  H  S  Q  S  L  D  R  L  M  N  P  L  I  D  Q  Y  L  Y  Y  L  S  R  T  N  T  P  S  G  T  T  T
     222
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS
                                                                >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                                Exchange Region 453
```

```
          3310      3320      3330      3340      3350      3360      3370      3380      3390      3400
            |         |         |         |         |         |         |         |         |         |
3301 cagagtcgacttcagttttctcaggccggagcgagtgacattcgggaccagtctaggaactggcttcctggaccctgttaccgccagcagcgagtatcaa
     Q  S  R  L  Q  F  S  Q  A  G  A  S  D  I  R  D  Q  S  R  N  W  L  P  G  P  C  Y  R  Q  Q  R  V  S  K
     457
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
     Q  S  R  L  Q  F  S  Q  A  G  A  S  D  I  R  D  Q  S  R  N  W  L  P  G  P  C  Y  R  Q  Q  R  V  S  K
     320
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
     Q  S  R  L  Q  F  S  Q  A  G  A  S  D  I  R  D  Q  S  R  N  W  L  P  G  P  C  Y  R  Q  Q  R  V  S  K
     255
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS
     >>>>>
     Exchange Region 453

          3410      3420      3430      3440      3450      3460      3470      3480      3490      3500
            |         |         |         |         |         |         |         |         |         |
3401 agacatctgcggataacaacaacagtgaatactcgtggactggagctaccaagtaccacctcaatggcagagactctctggtgaatccgggcccggccat
       T  S  A  D  N  N  N  S  E  Y  S  W  T  G  A  T  K  Y  H  L  N  G  R  D  S  L  V  N  P  G  P  A  M
     491
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
       T  S  A  D  N  N  N  S  E  Y  S  W  T  G  A  T  K  Y  H  L  N  G  R  D  S  L  V  N  P  G  P  A  M
     354
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
       T  S  A  D  N  N  N  S  E  Y  S  W  T  G  A  T  K  Y  H  L  N  G  R  D  S  L  V  N  P  G  P  A  M
     289
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS

          3510      3520      3530      3540      3550      3560      3570      3580      3590      3600
            |         |         |         |         |         |         |         |         |         |
3501 ggcaagccacaaggacgatgaagaaaagtttttttcctcagagcggggttctcatctttgggaagcaaggctcagagaaaacaaatgtggacattgaaaag
      A  S  H  K  D  D  E  E  K  F  F  P  Q  S  G  V  L  I  F  G  K  Q  G  S  E  K  T  N  V  D  I  E  K
     524
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
      A  S  H  K  D  D  E  E  K  F  F  P  Q  S  G  V  L  I  F  G  K  Q  G  S  E  K  T  N  V  D  I  E  K
     387
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
      A  S  H  K  D  D  E  E  K  F  F  P  Q  S  G  V  L  I  F  G  K  Q  G  S  E  K  T  N  V  D  I  E  K
     322
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS

          3610      3620      3630      3640      3650      3660      3670      3680      3690      3700
            |         |         |         |         |         |         |         |         |         |
3601 gtcatgattacagacgaagaggaaatcaggacaaccaatcccgtggctacggagcagtatggatccgtatctaccaacctccagagaggcaacagacaag
     V  M  I  T  D  E  E  E  I  R  T  T  N  P  V  A  T  E  Q  Y  G  S  V  S  T  N  L  Q  R  G  N  R  Q  A
     557
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
     V  M  I  T  D  E  E  E  I  R  T  T  N  P  V  A  T  E  Q  Y  G  S  V  S  T  N  L  Q  R  G  N  R  Q  A
     420
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
     V  M  I  T  D  E  E  E  I  R  T  T  N  P  V  A  T  E  Q  Y  G  S  V  S  T  N  L  Q  R  G  N  R  Q  A
     355
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS
                                                          M  D  P  Y  L  P  T  S  R  E  A  T  D  K
                                                          1
                                                          >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                          X gene CDS
                                                             >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                             Exchange Region 587

          3710      3720      3730      3740      3750      3760      3770      3780      3790      3800
            |         |         |         |         |         |         |         |         |         |
3701 cagctacagctgatgtcaacacacaaggcgttcttccaggcatggtctggcaggacagagatgtgtaccttcaggggcccatctgggcaaagattccaca
       A  T  A  D  V  N  T  Q  G  V  L  P  G  M  V  W  Q  D  R  D  V  Y  L  Q  G  P  I  W  A  K  I  P  H
     591
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
       A  T  A  D  V  N  T  Q  G  V  L  P  G  M  V  W  Q  D  R  D  V  Y  L  Q  G  P  I  W  A  K  I  P  H
     454
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
       A  T  A  D  V  N  T  Q  G  V  L  P  G  M  V  W  Q  D  R  D  V  Y  L  Q  G  P  I  W  A  K  I  P  H
     389
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS
     Q  L  Q  L  M  S  T  H  K  A  F  F  Q  A  W  S  G  R  T  E  M  C  T  F  R  G  P  S  G  Q  R  F  H  T
     15
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     X gene CDS
     >>>>>>>>>
     Exchange Region 587
```

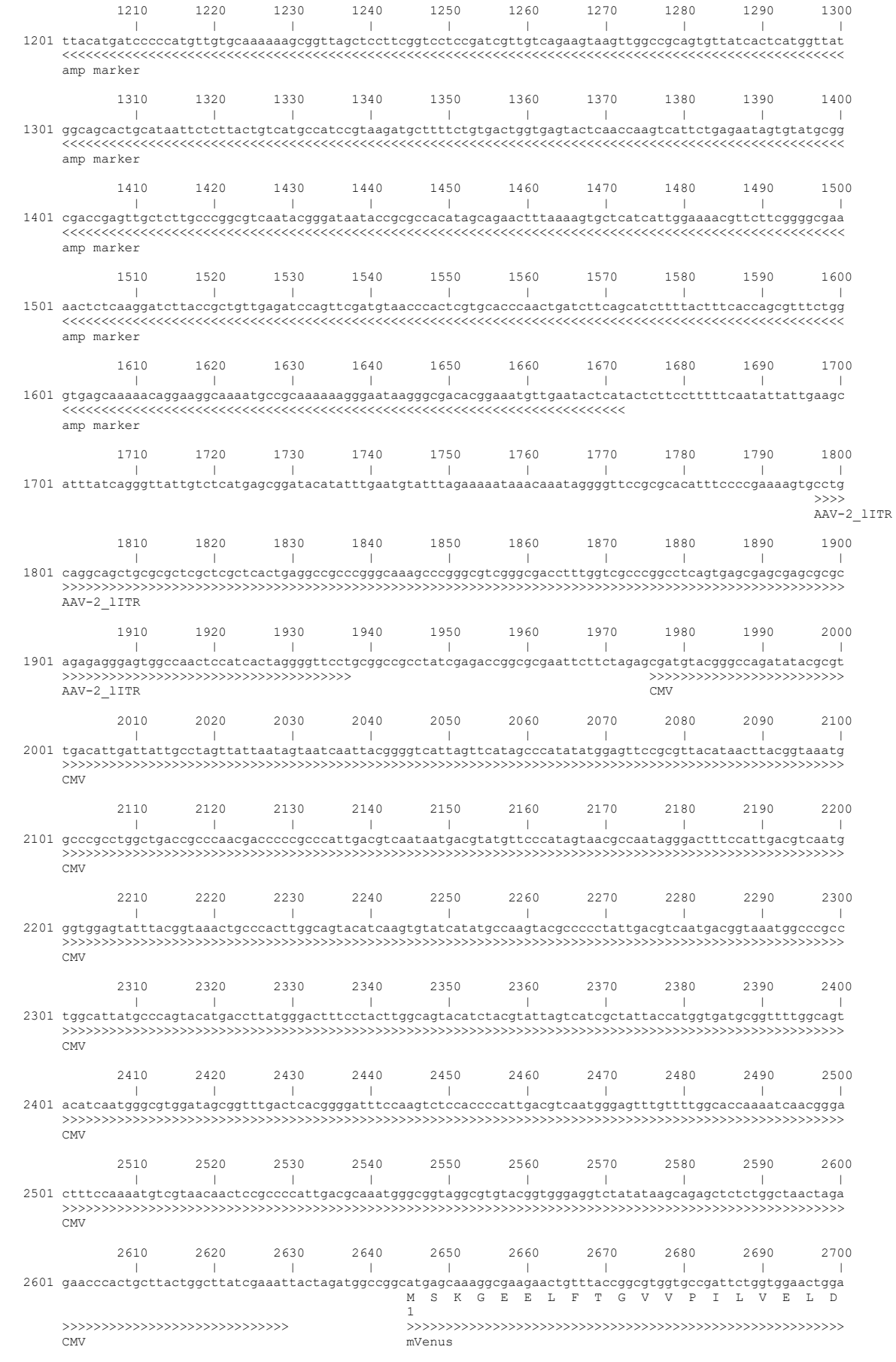
```
          3810      3820      3830      3840      3850      3860      3870      3880      3890      3900
            |         |         |         |         |         |         |         |         |         |
3801 cacggacggacattttcacccctctcccctcatgggtggattcggacttaaacaccctcctccacagattctcatcaagaacaccccggtacctgcgaat
      T  D  G  H  F  H  P  S  P  L  M  G  G  F  G  L  K  H  P  P  P  Q  I  L  I  K  N  T  P  V  P  A  N
     624
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
      T  D  G  H  F  H  P  S  P  L  M  G  G  F  G  L  K  H  P  P  P  Q  I  L  I  K  N  T  P  V  P  A  N
     487
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
      T  D  G  H  F  H  P  S  P  L  M  G  G  F  G  L  K  H  P  P  P  Q  I  L  I  K  N  T  P  V  P  A  N
     422
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS
       R  T  D  I  F  T  P  L  P  S  W  V  D  S  D  L  N  T  L  L  H  R  F  S  S  R  T  P  R  Y  L  R  I
     49
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     X gene CDS

          3910      3920      3930      3940      3950      3960      3970      3980      3990      4000
            |         |         |         |         |         |         |         |         |         |
3901 ccttcgaccaccttcagtgcggcaaagtttgcttccttcatcacacagtactccacgggacaggtcagcgtggagatcgagtgggagctccagaaggaaa
     P  S  T  T  F  S  A  A  K  F  A  S  F  I  T  Q  Y  S  T  G  Q  V  S  V  E  I  E  W  E  L  Q  K  E  N
     657
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
     P  S  T  T  F  S  A  A  K  F  A  S  F  I  T  Q  Y  S  T  G  Q  V  S  V  E  I  E  W  E  L  Q  K  E  N
     520
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
     P  S  T  T  F  S  A  A  K  F  A  S  F  I  T  Q  Y  S  T  G  Q  V  S  V  E  I  E  W  E  L  Q  K  E  N
     455
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS
      L  R  P  P  S  V  R  Q  S  L  L  P  S  S  H  S  T  P  R  D  R  S  A  W  R  S  S  G  S  S  R  R  K
     82
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     X gene CDS

          4010      4020      4030      4040      4050      4060      4070      4080      4090      4100
            |         |         |         |         |         |         |         |         |         |
4001 acagcaaacgctggaatcccgaaattcagtacacttccaactacaacaagtctgttaatgtggactttactgtggacactaatggcgtgtattcagagcc
       S  K  R  W  N  P  E  I  Q  Y  T  S  N  Y  N  K  S  V  N  V  D  F  T  V  D  T  N  G  V  Y  S  E  P
     691
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
       S  K  R  W  N  P  E  I  Q  Y  T  S  N  Y  N  K  S  V  N  V  D  F  T  V  D  T  N  G  V  Y  S  E  P
     554
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
       S  K  R  W  N  P  E  I  Q  Y  T  S  N  Y  N  K  S  V  N  V  D  F  T  V  D  T  N  G  V  Y  S  E  P
     489
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS
     T  A  N  A  G  I  P  K  F  S  T  L  P  T  T  T  S  L  L  M  W  T  L  L  W  T  L  M  A  C  I  Q  S  L
     115
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     X gene CDS

          4110      4120      4130      4140      4150      4160      4170      4180      4190      4200
            |         |         |         |         |         |         |         |         |         |
4101 tcgccccattggcaccagatacctgactcgtaatctgtaattgcttgttaatcaataaaccgtttaattcgtttcagttgaactttggtctctgcgtatt
      R  P  I  G  T  R  Y  L  T  R  N  L  *
     724
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP1 CDS
      R  P  I  G  T  R  Y  L  T  R  N  L  *
     587
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP2 CDS
      R  P  I  G  T  R  Y  L  T  R  N  L  *
     522
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     major coat protein VP3 CDS
       A  P  L  A  P  D  T  *
     149
     >>>>>>>>>>>>>>>>>>>>>>>>>>>
     X gene CDS

          4210      4220      4230      4240      4250      4260      4270      4280      4290      4300
            |         |         |         |         |         |         |         |         |         |
4201 tctttcttatctagtttccatggctacgtagataagtagcatggcgggttaatcattaactactactagagggaggggtggagtcgtgacgtgaattacg
                                                                        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                                        p5tataless

          4310      4320      4330      4340      4350      4360      4370      4380      4390      4400
            |         |         |         |         |         |         |         |         |         |
4301 tcatagggttagggaggtcctgtattagaggtcacgtgagtgttttgcgacattttgcgacaccatgtggtctcgctgggggggggggcccgagtgagca
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     p5tataless

          4410      4420      4430      4440      4450      4460      4470      4480      4490      4500
            |         |         |         |         |         |         |         |         |         |
4401 cgcagggtctccattttgaagcgggaggtttgaactactagtagcggccgctgcagtccggcaaaaaaacgggcaaggtgtcaccaccctgccctttttc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     p5tataless
```

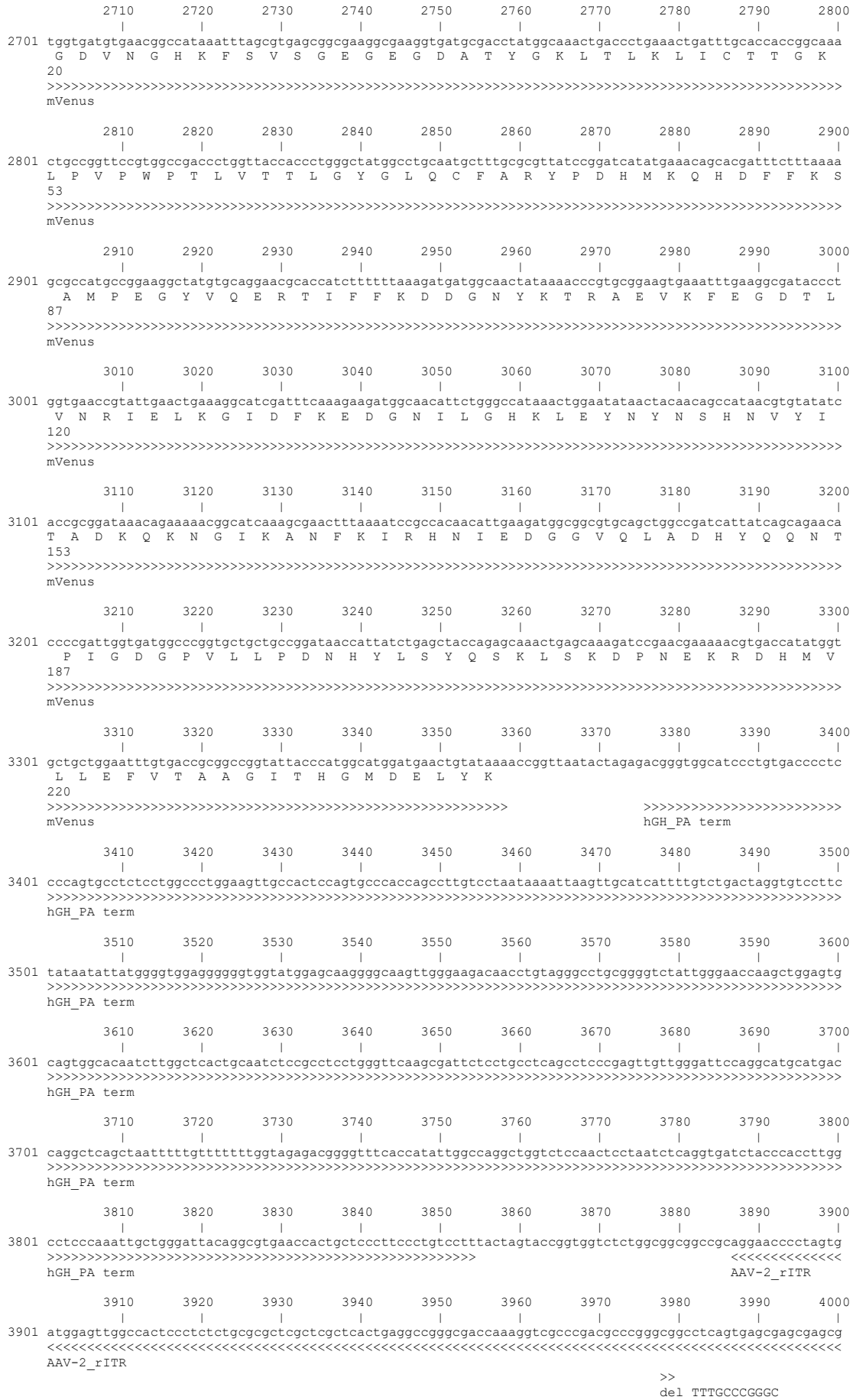
```
              4510       4520       4530       4540       4550       4560       4570       4580       4590       4600
                |          |          |          |          |          |          |          |          |          |
  4501 tttaaaaccgaaaagattacttcgcgttatgcaggcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactca

              4610       4620       4630       4640       4650       4660       4670       4680       4690       4700
                |          |          |          |          |          |          |          |          |          |
  4601 aaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgt
                                                                                                        >>>>>
                                                                                                        pBR322 origin
Rep Origin

              4710       4720       4730       4740       4750       4760       4770       4780       4790       4800
                |          |          |          |          |          |          |          |          |          |
  4701 tgctggcgtttttccataggctccgccccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagatacca
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
       pBR322 origin Rep Origin

              4810       4820       4830       4840       4850       4860       4870       4880       4890       4900
                |          |          |          |          |          |          |          |          |          |
  4801 ggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctt
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
       pBR322 origin Rep Origin

              4910       4920       4930       4940       4950       4960       4970       4980       4990       5000
                |          |          |          |          |          |          |          |          |          |
  4901 tctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgcct
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
       pBR322 origin Rep Origin

              5010       5020       5030       5040       5050       5060       5070       5080       5090       5100
                |          |          |          |          |          |          |          |          |          |
  5001 tatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtagg
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
       pBR322 origin Rep Origin

              5110       5120       5130       5140       5150       5160       5170       5180       5190       5200
                |          |          |          |          |          |          |          |          |          |
  5101 cggtgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaa
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
       pBR322 origin Rep Origin

              5210       5220       5230       5240       5250       5260       5270       5280       5290       5300
                |          |          |          |          |          |          |          |          |          |
  5201 agagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaag
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
       pBR322 origin Rep Origin

              5310       5320       5330       5340       5350       5360       5370       5380       5390       5400
                |          |          |          |          |          |          |          |          |          |
  5301 aagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaggatcttcaccta
       >>>>>>>>>>>>>>>>
       pBR322 origin Rep Origin

              5410       5420       5430       5440       5450       5460       5470       5480       5490       5500
                |          |          |          |          |          |          |          |          |          |
  5401 gatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagctcgaggcttggattctcaccaataaaaaac

              5510       5520       5530       5540       5550       5560       5570       5580       5590       5600
                |          |          |          |          |          |          |          |          |          |
  5501 gcccggcggcaaccgagcgttctgaacaaatccagatggagttctgaggtcattactggatctatcaacaggagtccaagcgagctcgatatcaaattac
                                                                                                        <<<<
                                                                                                        CAT marker

              5610       5620       5630       5640       5650       5660       5670       5680       5690       5700
                |          |          |          |          |          |          |          |          |          |
  5601 gccccgccctgccactcatcgcagtactgttgtaattcattaagcattctgccgacatggaagccatcacaaacggcatgatgaacctgaatcgccagcg
       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
       CAT marker

              5710       5720       5730       5740       5750       5760       5770       5780       5790       5800
                |          |          |          |          |          |          |          |          |          |
  5701 gcatcagcaccttgtcgccttgcgtatagtatttgcccatggtgaaaacgggggcgaagaagttgtccatattggccacgtttaaatcaaaactggtgaa
       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
       CAT marker

              5810       5820       5830       5840       5850       5860       5870       5880       5890       5900
                |          |          |          |          |          |          |          |          |          |
  5801 actcacccagggattggctgagacgaaaaacatattctcaataaaccctttagggaaatagqccaggttttcaccgtaacacgccacatcttgcgaatat
       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
       CAT marker

              5910       5920       5930       5940       5950       5960       5970       5980       5990       6000
                |          |          |          |          |          |          |          |          |          |
  5901 atgtgtagaaactgccggaaatcgtcgtggtattcactccagagcgatgaaaacgtttcagtttgctcatggaaaacggtgtaacaagggtgaacactat
       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
       CAT marker

              6010       6020       6030       6040       6050       6060       6070       6080       6090       6100
                |          |          |          |          |          |          |          |          |          |
  6001 cccatatcaccagctcaccgtctttcattgccatacgaaattccggatgagcattcatcaggcgggcaagaatgtgaataaaggccggataaaacttgtg
       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
       CAT marker
```

```
            6110       6120       6130       6140       6150       6160       6170       6180       6190       6200
              |          |          |          |          |          |          |          |          |          |
6101 cttatttttctttacggtctttaaaaaggccgtaatatcgagctgaacggtctggttataggtacattgagcaactgactgaaatgcctcaaaatgttct
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     CAT marker

            6210       6220       6230       6240       6250       6260       6270       6280       6290       6300
              |          |          |          |          |          |          |          |          |          |
6201 ttacgatgccattgggatatatcaacggtggtatatccagtgatttttttctccattttagcttccttagctcctgaaaatctcgataactcaaaaaata
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     CAT marker

            6310       6320       6330       6340       6350       6360       6370       6380       6390       6400
              |          |          |          |          |          |          |          |          |          |
6301 cgcccggtagtgatcttatttcattatggtgaaagttggaacctcttacgtgcccgatcaactcgagtgccacttgacgtctaagaaaccattattatca

            6410       6420       6430       6440       6450
              |          |          |          |          |
6401 tgacattaacctataaaaatagqcgtatcacgaggcagaatttcagataaaaaaa
```

10.3.1.3 **GOI-ITR plasmid (pZMB0522) sequence**

```
             10         20         30         40         50         60         70         80         90        100
              |          |          |          |          |          |          |          |          |          |
   1 gaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgccccccctgacgagcatcacaaaaatcgacgctc
     <<                                <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     AAV-2_rITR                        pBR322 origin

             110        120        130        140        150        160        170        180        190        200
              |          |          |          |          |          |          |          |          |          |
 101 aagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttacc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     pBR322 origin

             210        220        230        240        250        260        270        280        290        300
              |          |          |          |          |          |          |          |          |          |
 201 ggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     pBR322 origin

             310        320        330        340        350        360        370        380        390        400
              |          |          |          |          |          |          |          |          |          |
 301 gctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgccactggc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     pBR322 origin

             410        420        430        440        450        460        470        480        490        500
              |          |          |          |          |          |          |          |          |          |
 401 agcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaactacggctacactagaagaacagta
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     pBR322 origin

             510        520        530        540        550        560        570        580        590        600
              |          |          |          |          |          |          |          |          |          |
 501 tttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggttttttttg
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     pBR322 origin

             610        620        630        640        650        660        670        680        690        700
              |          |          |          |          |          |          |          |          |          |
 601 tttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgtta
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     pBR322 origin

             710        720        730        740        750        760        770        780        790        800
              |          |          |          |          |          |          |          |          |          |
 701 agggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaact

             810        820        830        840        850        860        870        880        890        900
              |          |          |          |          |          |          |          |          |          |
 801 tggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataa
               <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
               amp marker

             910        920        930        940        950        960        970        980        990       1000
              |          |          |          |          |          |          |          |          |          |
 901 ctacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagc
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     amp marker

            1010       1020       1030       1040       1050       1060       1070       1080       1090       1100
              |          |          |          |          |          |          |          |          |          |
1001 cggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaat
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     amp marker

            1110       1120       1130       1140       1150       1160       1170       1180       1190       1200
              |          |          |          |          |          |          |          |          |          |
1101 agtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgag
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     amp marker
```


amp marker
AAV-2_lITR
CMV
mVenus


2710 2720 2730 2740 2750 2760 2770 2780 2790 2800
2701 tggtgatgtgaacggccataaatttagcgtgagcggcgaaggcgaaggtgatgcgacctatggcaaactgaccctgaaactgatttgcaccaccggcaaa
G D V N G H K F S V S G E G E G D A T Y G K L T L K L I C T T G K
20
mVenus
2810 2820 2830 2840 2850 2860 2870 2880 2890 2900
2801 ctgccggttccgtggccgaccctggttaccaccctgggctatggcctgcaatgctttgcgcgttatccggatcatatgaaacagcacgatttctttaaaa
L P V P W P T L V T T L G Y G L Q C F A R Y P D H M K Q H D F F K S
53
mVenus
2910 2920 2930 2940 2950 2960 2970 2980 2990 3000
2901 gcgccatgccggaaggctatgtgcaggaacgcaccatctttttaaagatgatggcaactataaaacccgtgcggaagtgaaatttgaaggcgataccct
A M P E G Y V Q E R T I F F K D D G N Y K T R A E V K F E G D T L
87
mVenus
3010 3020 3030 3040 3050 3060 3070 3080 3090 3100
3001 ggtgaaccgtattgaactgaaaggcatcgatttcaaagaagatggcaacattctgggccataaactggaatataactacaacagccataacgtgtatatc
V N R I E L K G I D F K E D G N I L G H K L E Y N Y N S H N V Y I
120
mVenus
3110 3120 3130 3140 3150 3160 3170 3180 3190 3200
3101 accgcggataaacagaaaaacggcatcaaagcgaactttaaaatccgccacaacattgaagatggcggcgtgcagctggccgatcattatcagcagaaca
T A D K Q K N G I K A N F K I R H N I E D G G V Q L A D H Y Q Q N T
153
mVenus
3210 3220 3230 3240 3250 3260 3270 3280 3290 3300
3201 ccccgattggtgatggcccggtgctgctgccggataaccattatctgagctaccagagcaaactgagcaaagatccgaacgaaaaacgtgaccatatggt
P I G D G P V L L P D N H Y L S Y Q S K L S K D P N E K R D H M V
187
mVenus
3310 3320 3330 3340 3350 3360 3370 3380 3390 3400
3301 gctgctggaatttgtgaccgcggccggtattacccatggcatggatgaactgtataaaaaccggttaatactagagacgggtggcatccctgtgacccctc
L L E F V T A A G I T H G M D E L Y K
220
mVenus
hGH_PA term
3410 3420 3430 3440 3450 3460 3470 3480 3490 3500
3401 cccagtgcctctcctggccctggaagttgccactccagtgcccaccagccttgtcctaataaaattaagttgcatcattttgtctgactaggtgtccttc
hGH_PA term
3510 3520 3530 3540 3550 3560 3570 3580 3590 3600
3501 tataatattatggggtggaggggggtggtatggagcaaggggcaagttgggaagacaacctgtagggcctgcggggtctattgggaaccaagctggagtg
hGH_PA term
3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
3601 cagtggcacaatcttggctcactgcaatctccgcctcctgggttcaagcgattctcctgcctcagcctcccgagttgttgggattccaggcatgcatgac
hGH_PA term
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
3701 caggctcagctaatttttgtttttttggtagagacggggtttcaccatattggccaggctggtctccaactcctaatctcaggtgatctacccaccttgg
hGH_PA term
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
3801 cctcccaaattgctgggattacaggcgtgaaccactgctcccttccctgtccttactagtaccggtggtctctggcggcggccgcaggaacccctagtg
hGH_PA term
AAV-2_rITR
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
3901 atggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcccgggcggcctcagtgagcgagcgagcg
AAV-2_rITR
del TTTGCCCGGGC

```
            4010
              |
4001 cgcagctgcctgca
     <<<<<<<<<<<<<<
     AAV-2_rITR
```

10.3.2 **Insert sequences**

10.3.2.1 **GG-insertion in 587-loop region**

```
             10        20        30        40        50
              |         |         |         |         |
     1 BamHI                                             53 PvuII
     |                                                   |
   1 ggatccgtatctaccaacctccagagaggcaacggaggcagacaagcagctacagctg
     G  S  V  S  T  N  L  Q  R  G  N  G  G  R  Q  A  A  T  A
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      Exchange Region 587
                                      >>>>>>
                                      GG-insertion
```

10.3.2.2 **GGSG-insertion in 587-loop region**

```
             10        20        30        40        50        60
              |         |         |         |         |         |
     1 BamHI                                                   59 PvuII
     |                                                         |
   1 ggatccgtatctaccaacctccagagaggcaacggaggctctggtagacaagcagctacagctg
     G  S  V  S  T  N  L  Q  R  G  N  G  G  S  G  R  Q  A  A  T  A
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      Exchange Region 587
                                      >>>>>>>>>>>>
                                      GGSG-insertion
```

10.3.2.3 **2x(GGSG)-insertion in 587-loop region**

```
             10        20        30        40        50        60        70
              |         |         |         |         |         |         |
     1 BamHI                                                               71 PvuII
     |                                                                     |
   1 ggatccgtatctaccaacctccagagaggcaacggaggctctggtggcggttcaggaagacaagcagctacagctg
     G  S  V  S  T  N  L  Q  R  G  N  G  G  S  G  G  G  S  G  R  Q  A  A  T  A
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      Exchange Region 587
                                      >>>>>>>>>>>>>>>>>>>>>>>>
                                      2x(GGSG)-insertion
```

10.3.2.4 **4x(GGSG)-insertion in 587-loop region**

```
             10        20        30        40        50        60        70        80        90       100
              |         |         |         |         |         |         |         |         |         |
     1 BamHI                                                                                         95 PvuII
     |                                                                                               |
   1 ggatccgtatctaccaacctccagagaggcaacggaggctctggtggcggttcaggaggtggaagcggcggaggtagtggcagacaagcagctacagctg
     G  S  V  S  T  N  L  Q  R  G  N  G  G  S  G  G  G  S  G  G  G  S  G  G  G  S  G  R  Q  A  A  T  A
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      Exchange Region 587
                                      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                      4x(GGSG)-insertion
```

10.3.2.5 **bla-insertion in 587-loop region**

```
             10        20        30        40        50        60        70        80        90       100
              |         |         |         |         |         |         |         |         |         |
     1 BamHI
     |
   1 ggatccgtatctaccaacctccagagaggcaaccacccagaaacgctggcgaaagtaaaagatctggaagatcagttgggtgcacgagtgggttacatcg
     G  S  V  S  T  N  L  Q  R  G  N  H  P  E  T  L  A  K  V  K  D  L  E  D  Q  L  G  A  R  V  G  Y  I  E
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      587 Exchange Region
                                      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                      bla

            110       120       130       140       150       160       170       180       190       200
              |         |         |         |         |         |         |         |         |         |
 101 aactggatatcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtatt
       L  D  I  N  S  G  K  I  L  E  S  F  R  P  E  E  R  F  P  M  M  S  T  F  K  V  L  L  C  G  A  V  L
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     587 Exchange Region
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     bla
```

```
          210       220       230       240       250       260       270       280       290       300
           |         |         |         |         |         |         |         |         |         |
201 atcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtcacagaaaagcatcttacg
     S  R  I  D  A  G  Q  E  Q  L  G  R  R  I  H  Y  S  Q  N  D  L  V  E  Y  S  P  V  T  E  K  H  L  T
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    587 Exchange Region
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    bla

          310       320       330       340       350       360       370       380       390       400
           |         |         |         |         |         |         |         |         |         |
301 gatggcatgacagtaggagaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaagggcc
    D  G  M  T  V  G  E  L  C  S  A  A  I  T  M  S  D  N  T  A  A  N  L  L  L  T  T  I  G  G  P  K  G  L
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    587 Exchange Region
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    bla

          410       420       430       440       450       460       470       480       490       500
           |         |         |         |         |         |         |         |         |         |
401 taaccgctttttgcgtaacatgggggatcatacaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccac
      T  A  F  L  R  N  M  G  D  H  T  T  R  L  D  R  W  E  P  E  L  N  E  A  I  P  N  D  E  R  D  T  T
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    587 Exchange Region
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    bla

          510       520       530       540       550       560       570       580       590       600
           |         |         |         |         |         |         |         |         |         |
501 gacgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactccagcttcccggcaacaattaatggactggatgaaagcggat
     T  P  V  A  M  A  T  T  L  R  K  L  L  T  G  E  L  L  T  P  A  S  R  Q  Q  L  M  D  W  M  K  A  D
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    587 Exchange Region
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    bla

          610       620       630       640       650       660       670       680       690       700
           |         |         |         |         |         |         |         |         |         |
601 aaagttgcaggaccacttctgcgctcggtccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcattgcag
    K  V  A  G  P  L  L  R  S  V  L  P  A  G  W  F  I  A  D  K  S  G  A  G  E  R  G  S  R  G  I  I  A  V
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    587 Exchange Region
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    bla

          710       720       730       740       750       760       770       780       790       800
           |         |         |         |         |         |         |         |         |         |
701 tgctggggccagatggtaagccctcccgtatcgtagttatctacatgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagatagg
      L  G  P  D  G  K  P  S  R  I  V  V  I  Y  M  T  G  S  Q  A  T  M  D  E  R  N  R  Q  I  A  E  I  G
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    587 Exchange Region
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    bla

          810       820       830       840
           |         |         |         |
                                836 PvuII
                               |
801 tgcctcactgattaagcattggagacaagcagctacagctg
     A  S  L  I  K  H  W  R  Q  A  A  T  A
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    587 Exchange Region
    >>>>>>>>>>>>>>>>>>>>>>>>
    bla
```

10.3.2.6 **eGFPd2-Zeocin-cassettes without homologous arms**

```
          10        20        30        40        50        60        70        80        90        100
           |         |         |         |         |         |         |         |         |         |
    1 BamHI
    |
  1 ggatccgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttac
        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
        CMV Promoter

          110       120       130       140       150       160       170       180       190       200
           |         |         |         |         |         |         |         |         |         |
101 ggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccattga
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    CMV Promoter

          210       220       230       240       250       260       270       280       290       300
           |         |         |         |         |         |         |         |         |         |
201 cgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaat
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    CMV Promoter

          310       320       330       340       350       360       370       380       390       400
           |         |         |         |         |         |         |         |         |         |
301 ggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggtt
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    CMV Promoter
```

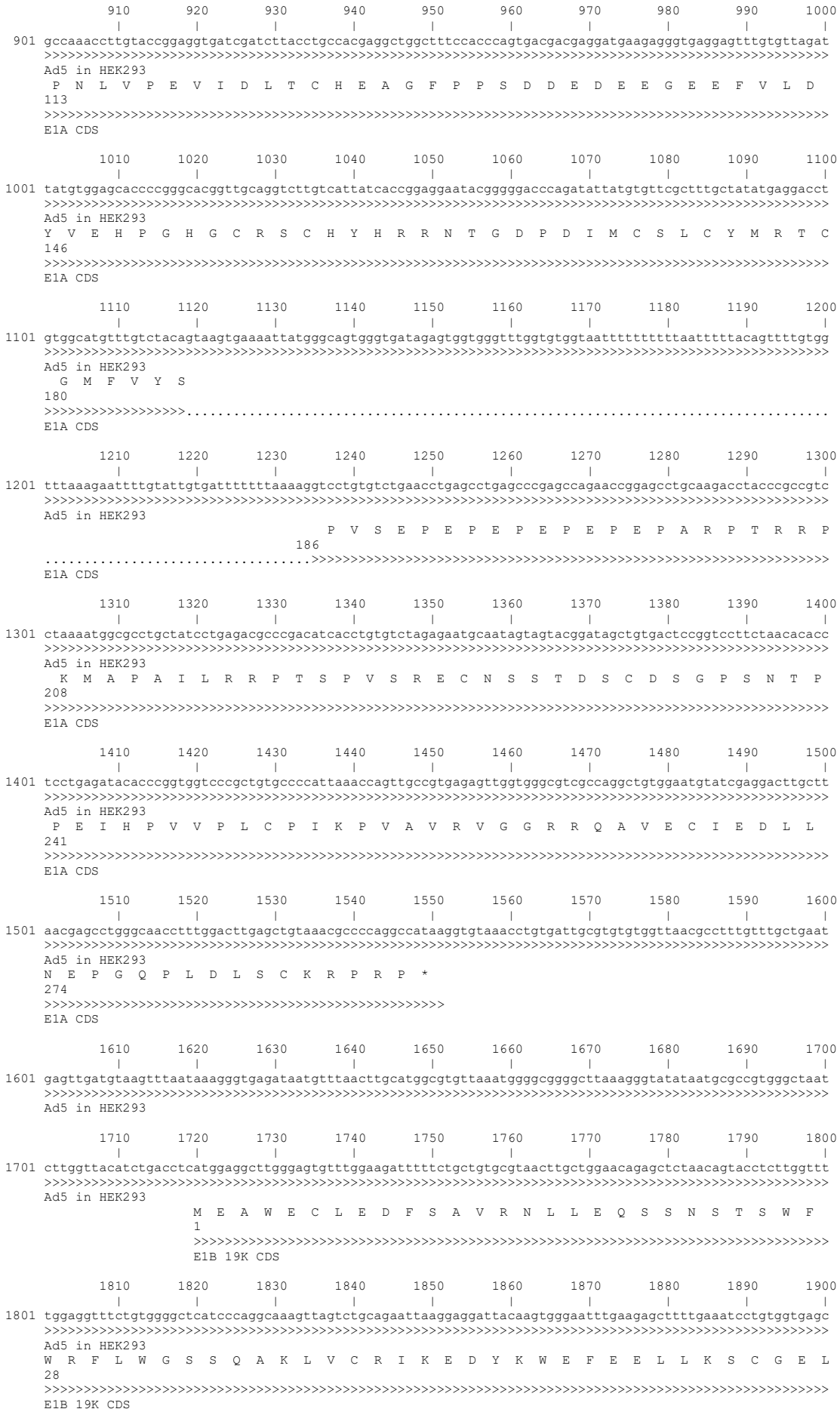
```
          410       420       430       440       450       460       470       480       490       500
           |         |         |         |         |         |         |         |         |         |
401 ttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaat
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    CMV Promoter

          510       520       530       540       550       560       570       580       590       600
           |         |         |         |         |         |         |         |         |         |
501 caacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggc
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    CMV Promoter

          610       620       630       640       650       660       670       680       690       700
           |         |         |         |         |         |         |         |         |         |
601 taactagagaacccactgcttactggcttatcgaaattaatacgactcactatagggagacccaagctggctagcatggtgtccaagggcgaggaactgt
                                                                                M  V  S  K  G  E  E  L  F
                                                                                1
                                                                                >>>>>>>>>>>>>>>>>>>>>>>>>
                                                                                eGFPd2

          710       720       730       740       750       760       770       780       790       800
           |         |         |         |         |         |         |         |         |         |
701 tcaccggcgtggtgcccatcctggtggaactggatggcgacgtgaacggccacaagttctccgtgtctggcgagggcgaaggcgacgctacctatggcaa
      T  G  V  V  P  I  L  V  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G  K
    10
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    eGFPd2

          810       820       830       840       850       860       870       880       890       900
           |         |         |         |         |         |         |         |         |         |
801 gctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccttggcctaccctcgtgaccaccctgacctacggcgtgcagtgcttctccagatac
     L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  T  Y  G  V  Q  C  F  S  R  Y
    43
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    eGFPd2

          910       920       930       940       950       960       970       980       990      1000
           |         |         |         |         |         |         |         |         |         |
901 cccgaccatatgaagcagcacgacttcttcaagtccgccatgcccgagggctacgtgcaggaacggaccatcttctttaaggacgacggcaactacaaga
    P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F  F  K  D  D  G  N  Y  K  T
    76
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    eGFPd2

         1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
           |         |         |         |         |         |         |         |         |         |
1001 ccagggccgaagtgaagttcgagggcgacaccctcgtgaaccggatcgagctgaagggcatcgacttcaaagaggacggcaacatcctgggccacaagct
      R  A  E  V  K  F  E  G  D  T  L  V  N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K  L
    110
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    eGFPd2

         1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
           |         |         |         |         |         |         |         |         |         |
1101 ggagtacaactacaactcccacaacgtgtacatcatggccgacaagcagaaaaacggcatcaaagtgaacttcaagatccggcacaacatcgaggacggc
     E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  Q  K  N  G  I  K  V  N  F  K  I  R  H  N  I  E  D  G
    143
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    eGFPd2

         1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
           |         |         |         |         |         |         |         |         |         |
1201 tccgtgcagctggccgaccactaccagcagaacacccctatcggcgacggccctgtgctgctgcctgacaaccactacctgtccacccagtccgccctgt
    S  V  Q  L  A  D  H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y  L  S  T  Q  S  A  L  S
    176
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    eGFPd2

         1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
           |         |         |         |         |         |         |         |         |         |
1301 ccaaggaccctaacgagaagcgggaccacatggtgctgctggaattcgtgaccgccgctggcatcaccctgggcatggacgagctgtacaagaagctgtc
      K  D  P  N  E  K  R  D  H  M  V  L  L  E  F  V  T  A  A  G  I  T  L  G  M  D  E  L  Y  K  K  L  S
    210
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    eGFPd2

         1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
           |         |         |         |         |         |         |         |         |         |
1401 ccacggcttcccacccgaggtggaagaacaggacgatggcaccctgcccatgtcctgcgctcaggaatccggcatggacagacaccctgccgcctgtgcc
     H  G  F  P  P  E  V  E  E  Q  D  D  G  T  L  P  M  S  C  A  Q  E  S  G  M  D  R  H  P  A  A  C  A
    243
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    eGFPd2

         1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
           |         |         |         |         |         |         |         |         |         |
1501 tctgcccggatcaatgtctgagggcccgtttaaacccgctgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctccccgtgc
    S  A  R  I  N  V  *
    276
    >>>>>>>>>>>>>>>>>>>>>>                          >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    eGFPd2                                         BGH polyA

         1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
           |         |         |         |         |         |         |         |         |         |
1601 cttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgggggg
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    BGH polyA
```

```
           1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
             |         |         |         |         |         |         |         |         |         |
                                                                                1778 AscI
                                                                                |
1701 tggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatggggcgcgccataacttcgtatagc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     BGH polyA                                                                                 Zeocin-cassette

           1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
             |         |         |         |         |         |         |         |         |         |
1801 atacattatacgaagttatctgtggaatgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaat
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette

           1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
             |         |         |         |         |         |         |         |         |         |
1901 tagtcagcaaccaggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgccctaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette

           2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
             |         |         |         |         |         |         |         |         |         |
2001 ctccgcccatcccgcccctaactccgcccagttccgcccattctccgccccatggctgactaatttttttttatttatgcagaggccgaggccgcctctgc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette

           2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
             |         |         |         |         |         |         |         |         |         |
2101 ctctgagctattccagaagtagtgaggaggctttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccatttttcggatctgatcag
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette

           2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
             |         |         |         |         |         |         |         |         |         |
2201 cacgtgttgacaattaatcatcggcatagtatatcggcatagtataatacgacaaggtgaggaactaaaccatggccaagttgaccagtgccgttccggt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
                                                             M  A  K  L  T  S  A  V  P  V
                                                             1
                                                             >>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                             Zeocin-R

           2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
             |         |         |         |         |         |         |         |         |         |
2301 gctcaccgcgcgcgacgtcgccggagcggtcgagttctggaccgaccggctcgggttctcccgggacttcgtggaggacgacttcgccggtgtggtccgg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
      L  T  A  R  D  V  A  G  A  V  E  F  W  T  D  R  L  G  F  S  R  D  F  V  E  D  D  F  A  G  V  V  R
     11
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-R

           2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
             |         |         |         |         |         |         |         |         |         |
2401 gacgacgtgaccctgttcatcagcgcggtccaggaccaggtggtgccggacaacaccctggcctgggtgtgggtgcgcggcctggacgagctgtacgccg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
     D  D  V  T  L  F  I  S  A  V  Q  D  Q  V  V  P  D  N  T  L  A  W  V  W  V  R  G  L  D  E  L  Y  A  E
     44
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-R

           2510      2520      2530      2540      2550      2560      2570      2580      2590      2600
             |         |         |         |         |         |         |         |         |         |
2501 agtggtcggaggtcgtgtccacgaacttccgggacgcctccgggccggccatgaccgagatcggcgagcagccgtgggggcgggagttcgccctgcgcga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
       W  S  E  V  V  S  T  N  F  R  D  A  S  G  P  A  M  T  E  I  G  E  Q  P  W  G  R  E  F  A  L  R  D
     78
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-R

           2610      2620      2630      2640      2650      2660      2670      2680      2690      2700
             |         |         |         |         |         |         |         |         |         |
2601 cccggccggcaactgcgtgcacttcgtggccgaggagcaggactgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
      P  A  G  N  C  V  H  F  V  A  E  E  Q  D  *
     111
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-R

           2710      2720      2730      2740      2750      2760      2770      2780      2790      2800
             |         |         |         |         |         |         |         |         |         |
2701 ggaatcgttttccgggacgccggctggatgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette

           2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
             |         |         |         |         |         |         |         |         |         |
2801 gttacaaataaagcaatagcatcacaaatttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatca
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
```

```
            2910       2920       2930       2940       2950       2960       2970       2980       2990       3000
              |          |          |          |          |          |          |          |          |          |
2901 tgtctgtataccgtcgacctctagctagagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacata
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette

            3010       3020       3030
              |          |          |
                                    3030 AgeI
                                    |
3001 cgagccggaagcataaagtgtaaagcctgaccggt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
```

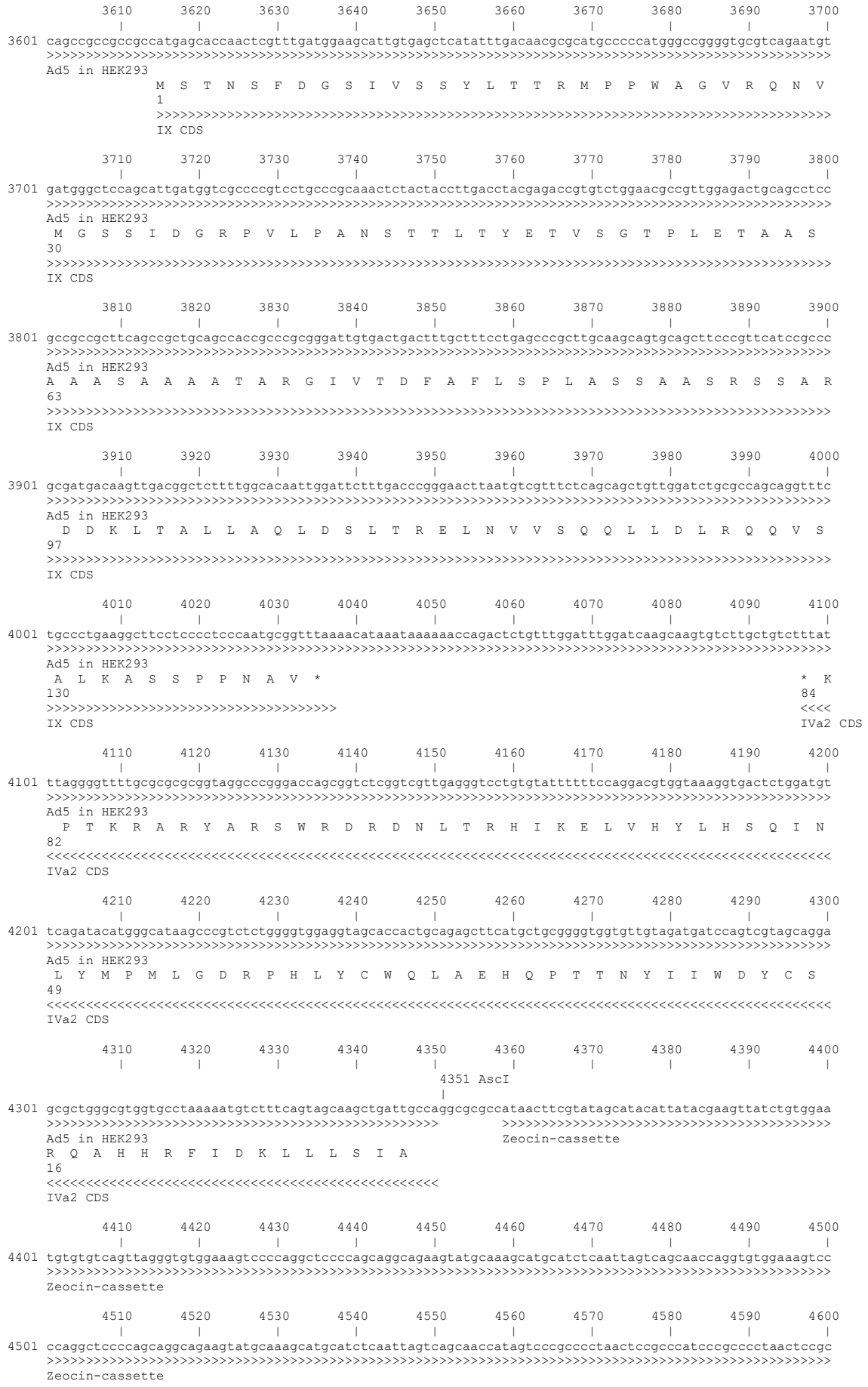
10.3.2.7 **E1a/b-Zeocin-cassettes without homologous arms**

```
             10         20         30         40         50         60         70         80         90         100
              |          |          |          |          |          |          |          |          |          |
     1 BamHI
     |
   1 ggatcccatcatcaataatatacctttatttggattgaagccaatatgataatgaggggtggagtttgtgacgtggcgcggggcgtgggaacggggcgg
         >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
         Ad5 in HEK293
         >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
         ITR repeat region

             110        120        130        140        150        160        170        180        190        200
              |          |          |          |          |          |          |          |          |          |
 101 gtgacgtagtagtgtggcggaagtgtgatgttgcaagtgtggcggaacacatgtaagcgacggatgtggcaaaagtgacgtttttggtgtgcgccggtgt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
     >>>>>>>>>
     ITR repeat region

             210        220        230        240        250        260        270        280        290        300
              |          |          |          |          |          |          |          |          |          |
 201 acacaggaagtgacaattttcgcgcggttttaggcggatgttgtagtaaatttgggcgtaaccgagtaagatttggccattttcgcgggaaaactgaata
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293

             310        320        330        340        350        360        370        380        390        400
              |          |          |          |          |          |          |          |          |          |
 301 agaggaagtgaaatctgaataatttttgtgttactcatagcgcgtaatatttgtctagggccgcggggactttgaccgtttacgtggagactcgcccaggt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293

             410        420        430        440        450        460        470        480        490        500
              |          |          |          |          |          |          |          |          |          |
 401 gtttttctcaggtgttttccgcgttccgggtcaaagttggcgttttattattatagtcagctgacgtgtagtgtatttatacccggtgagttcctcaaga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293

             510        520        530        540        550        560        570        580        590        600
              |          |          |          |          |          |          |          |          |          |
 501 ggccactcttgagtgccagcgagtagagttttctcctccgagccgctccgacaccgggactgaaaatgagacatattatctgccacggaggtgttattac
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
                                                                M  R  H  I  I  C  H  G  G  V  I  T
                                                                1
                                                                >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                                E1A CDS

             610        620        630        640        650        660        670        680        690        700
              |          |          |          |          |          |          |          |          |          |
 601 cgaagaaatggccgccagtcttttggaccagctgatcgaagaggtactggctgataatcttccacctcctagccattttgaaccacctacccttcacgaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
      E  E  M  A  A  S  L  L  D  Q  L  I  E  E  V  L  A  D  N  L  P  P  P  S  H  F  E  P  P  T  L  H  E
     13
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1A CDS

             710        720        730        740        750        760        770        780        790        800
              |          |          |          |          |          |          |          |          |          |
 701 ctgtatgatttagacgtgacggcccccgaagatcccaacgaggaggcggtttcgcagatttttcccgactctgtaatgttggcggtgcaggaagggattg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
     L  Y  D  L  D  V  T  A  P  E  D  P  N  E  E  A  V  S  Q  I  F  P  D  S  V  M  L  A  V  Q  E  G  I  D
     46
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1A CDS

             810        820        830        840        850        860        870        880        890        900
              |          |          |          |          |          |          |          |          |          |
 801 acttactcacttttccgccggcgcccggttctccggagccgcctcacctttcccggcagcccgagcagccggagcagagagccttgggtccggtttctat
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
       L  L  T  F  P  P  A  P  G  S  P  E  P  P  H  L  S  R  Q  P  E  Q  P  E  Q  R  A  L  G  P  V  S  M
     80
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1A CDS
```


910 920 930 940 950 960 970 980 990 1000
901 gccaaaccttgtaccggaggtgatcgatcttacctgccacgaggctggctttccacccagtgacgacgaggatgaagagggtgaggagtttgtgttagat
Ad5 in HEK293
P N L V P E V I D L T C H E A G F P P S D D E D E E G E E F V L D
113
E1A CDS
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
1001 tatgtggagcaccccgggcacggttgcaggtcttgtcattatcaccggaggaatacgggggacccagatattatgtgttcgctttgctatatgaggacct
Ad5 in HEK293
Y V E H P G H G C R S C H Y H R R N T G D P D I M C S L C Y M R T C
146
E1A CDS
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
1101 gtggcatgtttgtctacagtaagtgaaaattatgggcagtgggtgatagagtggtgggtttggtgtggtaatttttttttttaatttttacagttttgtgg
Ad5 in HEK293
G M F V Y S
180
E1A CDS
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
1201 tttaaagaattttgtattgtgatttttttaaaaggtcctgtgtctgaacctgagcctgagcccgagccagaaccggagcctgcaagacctacccgccgtc
Ad5 in HEK293
P V S E P E P E P E P E P E P E P A R P T R R P
186
E1A CDS
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
1301 ctaaaatggcgcctgctatcctgagacgcccgacatcacctgtgtctagagaatgcaatagtagtacggatagctgtgactccggtccttctaacacacc
Ad5 in HEK293
K M A P A I L R R P T S P V S R E C N S S T D S C D S G P S N T P
208
E1A CDS
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
1401 tcctgagatacacccggtggtcccgctgtgccccattaaaccagttgccgtgagagttggtgggcgtcgccaggctgtggaatgtatcgaggacttgctt
Ad5 in HEK293
P E I H P V V P L C P I K P V A V R V G G R R Q A V E C I E D L L
241
E1A CDS
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
1501 aacgagcctgggcaacctttggacttgagctgtaaacgccccaggccataaggtgtaaacctgtgattgcgtgtgtggttaacgcctttgtttgctgaat
Ad5 in HEK293
N E P G Q P L D L S C K R P R P *
274
E1A CDS
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
1601 gagttgatgtaagtttaataaagggtgagataatgtttaacttgcatggcgtgttaaatggggcggggcttaaagggtatataatgcgccgtgggctaat
Ad5 in HEK293
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
1701 cttggttacatctgacctcatggaggcttgggagtgtttggaagatttttctgctgtgcgtaacttgctggaacagagctctaacagtacctcttggttt
Ad5 in HEK293
M E A W E C L E D F S A V R N L L E Q S S N S T S W F
1
E1B 19K CDS
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
1801 tggaggtttctgtggggctcatcccaggcaaagttagtctgcagaattaaggaggattacaagtgggaatttgaagagcttttgaaatcctgtggtgagc
Ad5 in HEK293
W R F L W G S S Q A K L V C R I K E D Y K W E F E E L L K S C G E L
28
E1B 19K CDS

```
            1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
              |         |         |         |         |         |         |         |         |         |
1901 tgtttgattctttgaatctgggtcaccaggcgcttttccaagagaaggtcatcaagactttggatttttccacaccggggcgcgctgcggctgctgttgc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
       F  D  S  L  N  L  G  H  Q  A  L  F  Q  E  K  V  I  K  T  L  D  F  S  T  P  G  R  A  A  A  A  V  A
     62
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 19K CDS

            2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
              |         |         |         |         |         |         |         |         |         |
2001 ttttttgagttttataaaggataaatggagcgaagaaacccatctgagcgggggtacctgctggattttctggccatgcatctgtggagagcggttgtg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
      F  L  S  F  I  K  D  K  W  S  E  E  T  H  L  S  G  G  Y  L  L  D  F  L  A  M  H  L  W  R  A  V  V
     95
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 19K CDS
                           M  E  R  R  N  P  S  E  R  G  V  P  A  G  F  S  G  H  A  S  V  E  S  G  C  E
                           1
                           >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                           E1B 55K CDS

            2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
              |         |         |         |         |         |         |         |         |         |
2101 agacacaagaatcgcctgctactgttgtcttccgtccgcccggcgataataccgacggaggagcagcagcagcaggaggaagccaggcggcggcggc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
     R  H  K  N  R  L  L  L  L  S  S  V  R  P  A  I  I  P  T  E  E  Q  Q  Q  Q  Q  E  E  A  R  R  R  R  Q
     128
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 19K CDS
       T  Q  E  S  P  A  T  V  V  F  R  P  P  G  D  N  T  D  G  G  A  A  A  A  A  G  G  S  Q  A  A  A  A
     27
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

            2210      2220      2230      2240      2250      2260      2270      2280      2290      2300
              |         |         |         |         |         |         |         |         |         |
2201 aggagcagagcccatggaacccgagagccggcctggaccctcgggaatgaatgttgtacaggtggctgaactgtatccagaactgagacgcattttgaca
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
       E  Q  S  P  W  N  P  R  A  G  L  D  P  R  E  *
     162
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 19K CDS
      G  A  E  P  M  E  P  E  S  R  P  G  P  S  G  M  N  V  V  Q  V  A  E  L  Y  P  E  L  R  R  I  L  T
     60
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

            2310      2320      2330      2340      2350      2360      2370      2380      2390      2400
              |         |         |         |         |         |         |         |         |         |
2301 attacagaggatgggcaggggctaaaggggtaaagagggagcgggggcttgtgaggctacagaggaggctaggaatctagcttttagcttaatgacca
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
     I  T  E  D  G  Q  G  L  K  G  V  K  R  E  R  G  A  C  E  A  T  E  E  A  R  N  L  A  F  S  L  M  T  R
     93
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

            2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
              |         |         |         |         |         |         |         |         |         |
2401 gacaccgtcctgagtgtattacttttcaacagatcaaggataattgcgctaatgagcttgatctgctggcgcagaagtattccatagagcagctgaccac
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
       H  R  P  E  C  I  T  F  Q  Q  I  K  D  N  C  A  N  E  L  D  L  L  A  Q  K  Y  S  I  E  Q  L  T  T
     127
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

            2510      2520      2530      2540      2550      2560      2570      2580      2590      2600
              |         |         |         |         |         |         |         |         |         |
2501 ttactggctgcagccaggggatgattttgaggaggctattagggtatatgcaaaggtggcacttaggccagattgcaagtacaagatcagcaaacttgta
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
      Y  W  L  Q  P  G  D  D  F  E  E  A  I  R  V  Y  A  K  V  A  L  R  P  D  C  K  Y  K  I  S  K  L  V
     160
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

            2610      2620      2630      2640      2650      2660      2670      2680      2690      2700
              |         |         |         |         |         |         |         |         |         |
2601 aatatcaggaattgttgctacatttctgggaacggggccgaggtggagatagatacggaggatagggtggcctttagatgtagcatgataaatatgtggc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
     N  I  R  N  C  C  Y  I  S  G  N  G  A  E  V  E  I  D  T  E  D  R  V  A  F  R  C  S  M  I  N  M  W  P
     193
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS
```

```
          2710      2720      2730      2740      2750      2760      2770      2780      2790      2800
            |         |         |         |         |         |         |         |         |         |
2701 cgggggtgcttggcatggacggggtggttattatgaatgtaaggtttactggccccaattttagcggtacggttttcctggccaataccaaccttatcct
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
       G  V  L  G  M  D  G  V  V  I  M  N  V  R  F  T  G  P  N  F  S  G  T  V  F  L  A  N  T  N  L  I  L
     227
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

          2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
            |         |         |         |         |         |         |         |         |         |
2801 acacggtgtaagcttctatgggtttaacaatacctgtgtggaagcctggaccgatgtaagggttcggggctgtgccttttactgctgctggaaggggggtg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
      H  G  V  S  F  Y  G  F  N  N  T  C  V  E  A  W  T  D  V  R  V  R  G  C  A  F  Y  C  C  W  K  G  V
     260
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

          2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
            |         |         |         |         |         |         |         |         |         |
2901 gtgtgtcgccccaaaagcagggcttcaattaagaaatgcctctttgaaaggtgtaccttgggtatcctgtctgagggtaactccagggtgcgcccacaatg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
     V  C  R  P  K  S  R  A  S  I  K  K  C  L  F  E  R  C  T  L  G  I  L  S  E  G  N  S  R  V  R  H  N  V
     293
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

          3010      3020      3030      3040      3050      3060      3070      3080      3090      3100
            |         |         |         |         |         |         |         |         |         |
3001 tggcctccgactgtggttgcttcatgctagtgaaaagcgtggctgtgattaagcataacatggtatgtggcaactgcgaggacagggcctctcagatgct
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
       A  S  D  C  G  C  F  M  L  V  K  S  V  A  V  I  K  H  N  M  V  C  G  N  C  E  D  R  A  S  Q  M  L
     327
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

          3110      3120      3130      3140      3150      3160      3170      3180      3190      3200
            |         |         |         |         |         |         |         |         |         |
3101 gacctgctcggacggcaactgtcacctgctgaagaccattcacgtagccagccactctcgcaaggcctggccagtgtttgagcataacatactgacccgc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
      T  C  S  D  G  N  C  H  L  L  K  T  I  H  V  A  S  H  S  R  K  A  W  P  V  F  E  H  N  I  L  T  R
     360
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

          3210      3220      3230      3240      3250      3260      3270      3280      3290      3300
            |         |         |         |         |         |         |         |         |         |
3201 tgttccttgcatttgggtaacaggaggggggtgttcctaccttaccaatgcaatttgagtcacactaagatattgcttgagcccgagagcatgtccaagg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
     C  S  L  H  L  G  N  R  R  G  V  F  L  P  Y  Q  C  N  L  S  H  T  K  I  L  L  E  P  E  S  M  S  K  V
     393
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

          3310      3320      3330      3340      3350      3360      3370      3380      3390      3400
            |         |         |         |         |         |         |         |         |         |
3301 tgaacctgaacggggtgtttgacatgaccatgaagatctggaaggtgctgaggtacgatgagacccgcaccaggtgcagaccctgcgagtgtggcggtaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
       N  L  N  G  V  F  D  M  T  M  K  I  W  K  V  L  R  Y  D  E  T  R  T  R  C  R  P  C  E  C  G  G  K
     427
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

          3410      3420      3430      3440      3450      3460      3470      3480      3490      3500
            |         |         |         |         |         |         |         |         |         |
3401 acatattaggaaccagcctgtgatgctggatgtgaccgaggagctgaggcccgatcacttggtgctggcctgcacccgcgctgagtttggctctagcgat
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
      H  I  R  N  Q  P  V  M  L  D  V  T  E  E  L  R  P  D  H  L  V  L  A  C  T  R  A  E  F  G  S  S  D
     460
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     E1B 55K CDS

          3510      3520      3530      3540      3550      3560      3570      3580      3590      3600
            |         |         |         |         |         |         |         |         |         |
3501 gaagatacagattgaggtactgaaatgtgtgggcgtggcttaagggtgggaaagaatatataaggtgggggtcttatgtagttttgtatctgttttgcag
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Ad5 in HEK293
     E  D  T  D  *
     493
     >>>>>>>>>>>>>>>>
     E1B 55K CDS
```

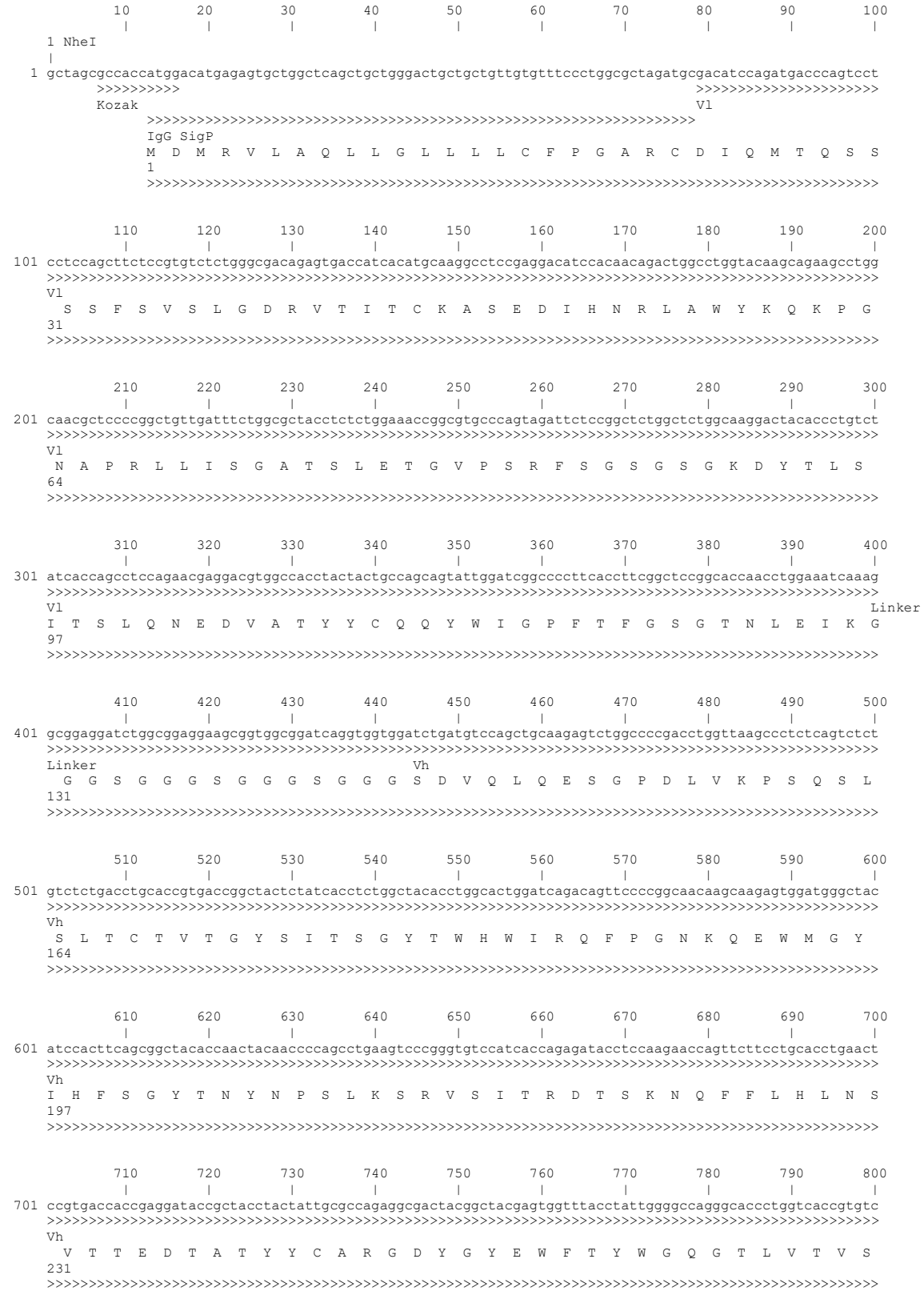

3610 3620 3630 3640 3650 3660 3670 3680 3690 3700
3601 cagccgccgccgccatgagcaccaactcgtttgatggaagcattgtgagctcatatttgacaacgcgcatgccccatgggccggggtgcgtcagaatgt
Ad5 in HEK293
M S T N S F D G S I V S S Y L T T R M P P W A G V R Q N V
1
IX CDS
3710 3720 3730 3740 3750 3760 3770 3780 3790 3800
3701 gatgggctccagcattgatggtcgccccgtcctgcccgcaaactctactaccttgacctacgagaccgtgtctggaacgccgttggagactgcagcctcc
Ad5 in HEK293
M G S S I D G R P V L P A N S T T L T Y E T V S G T P L E T A A S
30
IX CDS
3810 3820 3830 3840 3850 3860 3870 3880 3890 3900
3801 gccgccgcttcagccgctgcagccaccgcccgcgggattgtgactgactttgctttcctgagcccgcttgcaagcagtgcagcttcccgttcatccgccc
Ad5 in HEK293
A A A S A A A A T A R G I V T D F A F L S P L A S S A A S R S S A R
63
IX CDS
3910 3920 3930 3940 3950 3960 3970 3980 3990 4000
3901 gcgatgacaagttgacggctcttttggcacaattggattctttgacccgggaacttaatgtcgtttctcagcagctgttggatctgcgccagcaggtttc
Ad5 in HEK293
D D K L T A L L A Q L D S L T R E L N V V S Q Q L L D L R Q Q V S
97
IX CDS
4010 4020 4030 4040 4050 4060 4070 4080 4090 4100
4001 tgccctgaaggcttcctcccctcccaatgcggtttaaaacataaataaaaaaccagactctgtttggatttggatcaagcaagtgtcttgctgtctttat
Ad5 in HEK293
A L K A S S P P N A V *
130
IX CDS
* K
84
<<<<
IVa2 CDS
4110 4120 4130 4140 4150 4160 4170 4180 4190 4200
4101 ttaggggttttgcgcgcgcggtaggcccgggaccagcggtctcggtcgttgagggtcctgtgtatttttttccaggacgtggtaaaggtgactctggatgt
Ad5 in HEK293
P T K R A R Y A R S W R D R D N L T R H I K E L V H Y L H S Q I N
82
IVa2 CDS
4210 4220 4230 4240 4250 4260 4270 4280 4290 4300
4201 tcagatacatgggcataagcccgtctctggggtggaggtagcaccactgcagagcttcatgctgcggggtggtgttgtagatgatccagtcgtagcagga
Ad5 in HEK293
L Y M P M L G D R P H L Y C W Q L A E H Q P T T N Y I I W D Y C S
49
IVa2 CDS
4310 4320 4330 4340 4350 4360 4370 4380 4390 4400
4351 AscI
4301 gcgctgggcgtggtgcctaaaaatgtctttcagtagcaagctgattgccaggcgcgccataacttcgtatagcatacattatacgaagttatctgtggaa
Ad5 in HEK293
Zeocin-cassette
R Q A H H R F I D K L L L S I A
16
IVa2 CDS
4410 4420 4430 4440 4450 4460 4470 4480 4490 4500
4401 tgtgtgtcagttagggtgtggaaagtccccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccaggtgtggaaagtcc
Zeocin-cassette
4510 4520 4530 4540 4550 4560 4570 4580 4590 4600
4501 ccaggctccccagcaggcagaagtatgcaaagcatgcatctcaattagtcagcaaccatagtcccgcccctaactccgcccatcccgcccctaactccgc
Zeocin-cassette

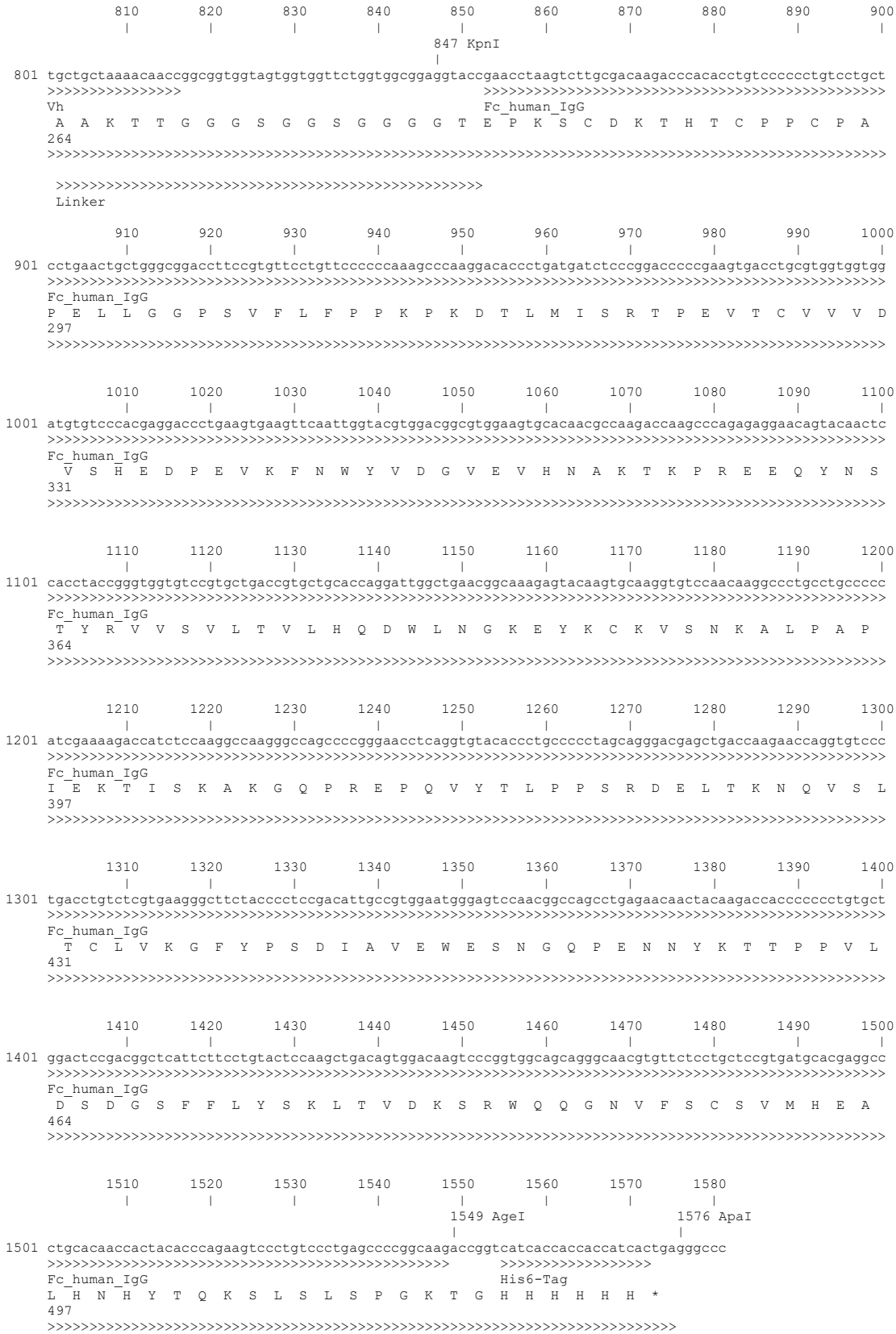
```
            4610      4620      4630      4640      4650      4660      4670      4680      4690      4700
              |         |         |         |         |         |         |         |         |         |
4601 ccagttccgcccattctccgccccatggctgactaattttttttatttatgcagaggccgaggccgcctctgcctctgagctattccagaagtagtgagg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette

            4710      4720      4730      4740      4750      4760      4770      4780      4790      4800
              |         |         |         |         |         |         |         |         |         |
4701 aggctttttttggaggcctaggcttttgcaaaaagctcccgggagcttgtatatccattttcggatctgatcagcacgtgttgacaattaatcatcggcat
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette

            4810      4820      4830      4840      4850      4860      4870      4880      4890      4900
              |         |         |         |         |         |         |         |         |         |
4801 agtatatcggcatagtataatacgacaaggtgaggaactaaaccatggccaagttgaccagtgccgttccggtgctcaccgcgcgcgacgtcgccggagc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
                                                M  A  K  L  T  S  A  V  P  V  L  T  A  R  D  V  A  G  A
                                                1
                                                >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                                Zeocin-R

            4910      4920      4930      4940      4950      4960      4970      4980      4990      5000
              |         |         |         |         |         |         |         |         |         |
4901 ggtcgagttctggaccgaccggctcgggttctcccgggacttcgtggaggacgacttcgccggtgtggtccgggacgacgtgaccctgttcatcagcgcg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
      V  E  F  W  T  D  R  L  G  F  S  R  D  F  V  E  D  D  F  A  G  V  V  R  D  D  V  T  L  F  I  S  A
     20
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-R

            5010      5020      5030      5040      5050      5060      5070      5080      5090      5100
              |         |         |         |         |         |         |         |         |         |
5001 gtccaggaccaggtggtgccggacaacaccctggcctgggtgtgggtgcgcggcctggacgagctgtacgccgagtggtcggaggtcgtgtccacgaact
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
     V  Q  D  Q  V  V  P  D  N  T  L  A  W  V  W  V  R  G  L  D  E  L  Y  A  E  W  S  E  V  V  S  T  N  F
     53
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-R

            5110      5120      5130      5140      5150      5160      5170      5180      5190      5200
              |         |         |         |         |         |         |         |         |         |
5101 tccgggacgcctccgggccggccatgaccgagatcggcgagcagccgtgggggcgggagttcgccctgcgcgacccggccggcaactgcgtgcacttcgt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
       R  D  A  S  G  P  A  M  T  E  I  G  E  Q  P  W  G  R  E  F  A  L  R  D  P  A  G  N  C  V  H  F  V
     87
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-R

            5210      5220      5230      5240      5250      5260      5270      5280      5290      5300
              |         |         |         |         |         |         |         |         |         |
5201 ggccgaggagcaggactgacacgtgctacgagatttcgattccaccgccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctgg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette
      A  E  E  Q  D  *
     120
     >>>>>>>>>>>>>>>>>>>
     Zeocin-R

            5310      5320      5330      5340      5350      5360      5370      5380      5390      5400
              |         |         |         |         |         |         |         |         |         |
5301 atgatcctccagcgcggggatctcatgctggagttcttcgcccaccccaacttgtttattgcagcttataatggttacaaataaagcaatagcatcacaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette

            5410      5420      5430      5440      5450      5460      5470      5480      5490      5500
              |         |         |         |         |         |         |         |         |         |
5401 atttcacaaataaagcatttttttcactgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctgtataccgtcgacctctagcta
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette

            5510      5520      5530      5540      5550      5560      5570      5580      5590      5600
              |         |         |         |         |         |         |         |         |         |
5501 gagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Zeocin-cassette


       5603 AgeI
       |
5601 tgaccggt
     >>
     Zeocin-cassette
```

10.3.2.8 **A20-scFv_His$_6$_Silica (from pZMB0349/pZMB0417)**

```
          10        20        30        40        50        60        70        80        90        100
           |         |         |         |         |         |         |         |         |         |
   1 NdeI
   |
 1 catatgaaatatctgctgccgaccgcagcagcgggtctgctgttgctggcagcacagcctgcaatggcagatattcagatgacccagagcagcagctcat
    M  K  Y  L  L  P  T  A  A  A  G  L  L  L  L  A  A  Q  P  A  M  A  D  I  Q  M  T  Q  S  S  S  S  F
    1
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    pelB                                                                Vl

          110       120       130       140       150       160       170       180       190       200
           |         |         |         |         |         |         |         |         |         |
101 ttagcgttagcctgggtgatcgtgttaccattacctgtaaagcaagcgaagatattcataatcgtctggcctggtataaacagaaaccgggtaatgcacc
     S  V  S  L  G  D  R  V  T  I  T  C  K  A  S  E  D  I  H  N  R  L  A  W  Y  K  Q  K  P  G  N  A  P
    34
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Vl

          210       220       230       240       250       260       270       280       290       300
           |         |         |         |         |         |         |         |         |         |
201 gcgtctgctgattagcggtgcaaccagcctggaaacaggtgttccgagccgttttagcggtagcggttcaggtaaagattataccctgagcattaccagc
     R  L  L  I  S  G  A  T  S  L  E  T  G  V  P  S  R  F  S  G  S  G  S  G  K  D  Y  T  L  S  I  T  S
    67
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Vl

          310       320       330       340       350       360       370       380       390       400
           |         |         |         |         |         |         |         |         |         |
301 cttcagaatgaagatgttgccacctattattgtcagcagtattggattggtccgtttacctttggtagcggcaccaatctggaaattaaaggtggtggta
    L  Q  N  E  D  V  A  T  Y  Y  C  Q  Q  Y  W  I  G  P  F  T  F  G  S  G  T  N  L  E  I  K  G  G  G  S
    100
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Vl                                                                                          Linker

          410       420       430       440       450       460       470       480       490       500
           |         |         |         |         |         |         |         |         |         |
401 gcggtggtggctctggtggtggttcaggtggcggtagtgatgttcagctgcaagaaagcggtccggatctggttaaaccgagccagagcctgagtctgac
     G  G  G  S  G  G  G  S  G  G  G  S  D  V  Q  L  Q  E  S  G  P  D  L  V  K  P  S  Q  S  L  S  L  T
    134
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Linker                            Vh

          510       520       530       540       550       560       570       580       590       600
           |         |         |         |         |         |         |         |         |         |
501 ctgtaccgttaccggctatagcattacctcaggttatacctggcattggattcgtcagtttccgggtaataaacaagaatggatgggctatattcatttc
     C  T  V  T  G  Y  S  I  T  S  G  Y  T  W  H  W  I  R  Q  F  P  G  N  K  Q  E  W  M  G  Y  I  H  F
    167
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Vh

          610       620       630       640       650       660       670       680       690       700
           |         |         |         |         |         |         |         |         |         |
601 agcggctataccaattataacccgagcctgaaaagccgtgtttcaattacccgtgataccagcaaaaaccagttttttctgcatctgaacagcgttacca
    S  G  Y  T  N  Y  N  P  S  L  K  S  R  V  S  I  T  R  D  T  S  K  N  Q  F  F  L  H  L  N  S  V  T  T
    200
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Vh

          710       720       730       740       750       760       770       780       790       800
           |         |         |         |         |         |         |         |         |         |
701 ccgaagataccgcaacctattactgtgcacgtggtgattatggttatgagtggtttacctattggggtcagggcaccctggttaccgttagcgcagccaa
     E  D  T  A  T  Y  Y  C  A  R  G  D  Y  G  Y  E  W  F  T  Y  W  G  Q  G  T  L  V  T  V  S  A  A  K
    234
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Vh

          810       820       830       840       850       860       870       880
           |         |         |         |         |         |         |         |
       805 AgeI              829 NheI                                  877 XhoI
       |                     |                                         |
801 aaccaccggtcatcatcaccatcatcatgctagcggtcgtgcacgtgcccagcgtcagagcagccgtggtcgttaactcgag
     T  T  G  H  H  H  H  H  H  A  S  G  R  A  R  A  Q  R  Q  S  S  R  G  R  *
    267
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>   >>>>>>>>>>>>>>>>>>>   >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Vh       His6 Tag              Silica Tag
```

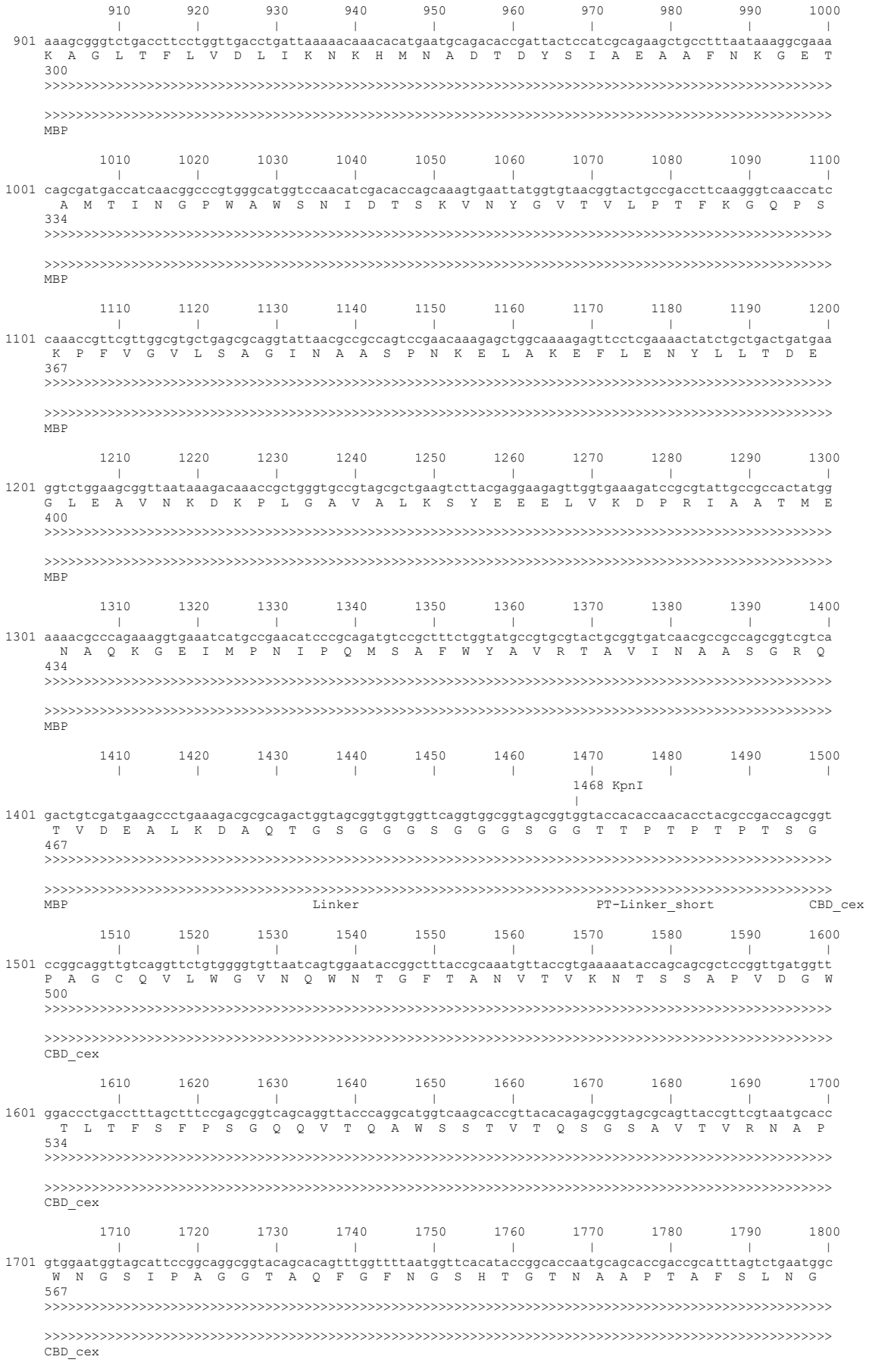
10.3.2.9 **A20-scFv_hIgG1-Fc_His$_6$**




810 820 830 840 850 860 870 880 890 900
847 KpnI
801 tgctgctaaaacaaccggcggtggtagtggtggttctggtggcggaggtaccgaacctaagtcttgcgacaagacccacacctgtccccctgtcctgct
Vh
Fc_human_IgG
A A K T T G G G S G G S G G G G T E P K S C D K T H T C P P C P A
264
Linker
910 920 930 940 950 960 970 980 990 1000
901 cctgaactgctgggcggaccttccgtgttcctgttccccccaaagcccaaggacaccctgatgatctcccggaccccgaagtgacctgcgtggtggtgg
Fc_human_IgG
P E L L G G P S V F L F P P K P K D T L M I S R T P E V T C V V V D
297
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
1001 atgtgtcccacgaggaccctgaagtgaagttcaattggtacgtggacggcgtggaagtgcacaacgccaagaccaagcccagagaggaacagtacaactc
Fc_human_IgG
V S H E D P E V K F N W Y V D G V E V H N A K T K P R E E Q Y N S
331
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
1101 cacctaccgggtggtgtccgtgctgaccgtgctgcaccaggattggctgaacggcaaagagtacaagtgcaaggtgtccaacaaggccctgcctgccccc
Fc_human_IgG
T Y R V V S V L T V L H Q D W L N G K E Y K C K V S N K A L P A P
364
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
1201 atcgaaaagaccatctccaagggccaagggccagccccgggaacctcaggtgtacaccctgcccctagcagggacgagctgaccaagaaccaggtgtccc
Fc_human_IgG
I E K T I S K A K G Q P R E P Q V Y T L P P S R D E L T K N Q V S L
397
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
1301 tgacctgtctcgtgaagggcttctaccctccgacattgccgtggaatgggagtccaacggccagcctgagaacaactacaagaccacccccccctgtgct
Fc_human_IgG
T C L V K G F Y P S D I A V E W E S N G Q P E N N Y K T T P P V L
431
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
1401 ggactccgacggctcattcttcctgtactccaagctgacagtggacaagtcccggtggcagcagggcaacgtgttctcctgctccgtgatgcacgaggcc
Fc_human_IgG
D S D G S F F L Y S K L T V D K S R W Q Q G N V F S C S V M H E A
464
1510 1520 1530 1540 1550 1560 1570 1580
1549 AgeI
1576 ApaI
1501 ctgcacaaccactacacccagaagtccctgtccctgagccccggcaagaccggtcatcaccaccaccatcactgagggccc
Fc_human_IgG
His6-Tag
L H N H Y T Q K S L S L S P G K T G H H H H H H *
497

10.3.2.10 PKD2_MBP_GSPT_CBDcex_His$_6$

```
              10        20        30        40        50        60        70        80        90       100
               |         |         |         |         |         |         |         |         |         |
    1 NdeI
    |
  1 catatgaatcgtccgcctattgcaattgttagtccgcagtttcaagaaatttcactgccgaccaccagtaccgttattgatggtagccagagcaccgatg
       M  N  R  P  P  I  A  I  V  S  P  Q  F  Q  E  I  S  L  P  T  T  S  T  V  I  D  G  S  Q  S  T  D  D
       1
       >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

                      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                      PKD2

             110       120       130       140       150       160       170       180       190       200
               |         |         |         |         |         |         |         |         |         |
101 atgataaaattgttcagtatcactgggaagaactgaaaggtccgctgcgtgaagaaaaaatcagcgaagataccgcaattctgaaactgagcaaactggt
      D  K  I  V  Q  Y  H  W  E  E  L  K  G  P  L  R  E  E  K  I  S  E  D  T  A  I  L  K  L  S  K  L  V
    34
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    PKD2

             210       220       230       240       250       260       270       280       290       300
               |         |         |         |         |         |         |         |         |         |
201 tccgggtaattatacctttagcctgaccgttgttgatagtgatggtgcaaccaatagcaccaccgcaaatctgaccgttaataaagcagttgattatccg
     P  G  N  Y  T  F  S  L  T  V  V  D  S  D  G  A  T  N  S  T  T  A  N  L  T  V  N  K  A  V  D  Y  P
    67
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    PKD2

             310       320       330       340       350       360       370       380       390       400
               |         |         |         |         |         |         |         |         |         |
    301 BamHI
    |
301 ggatccggtggtggctctggtggtggttcaggcggtaaaatcgaagaaggtaaactggtaatctggattaacggcgataaaggctataacggtctcgctg
    G  S  G  G  G  S  G  G  G  S  G  G  K  I  E  E  G  K  L  V  I  W  I  N  G  D  K  G  Y  N  G  L  A  E
    100
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    Linker                              MBP

             410       420       430       440       450       460       470       480       490       500
               |         |         |         |         |         |         |         |         |         |
401 aagtcggtaagaaattcgagaaagataccggaattaaagtcaccgttgagcatccggataaactggaagagaaattcccacaggttgcggcaactggcga
      V  G  K  K  F  E  K  D  T  G  I  K  V  T  V  E  H  P  D  K  L  E  E  K  F  P  Q  V  A  A  T  G  D
    134
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    MBP

             510       520       530       540       550       560       570       580       590       600
               |         |         |         |         |         |         |         |         |         |
501 tggccctgacattatcttctgggcacacgaccgctttggtggctacgctcaatctggcctgttggctgaaatcaccccggacaaagcgttccaggacaag
     G  P  D  I  I  F  W  A  H  D  R  F  G  G  Y  A  Q  S  G  L  L  A  E  I  T  P  D  K  A  F  Q  D  K
    167
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    MBP

             610       620       630       640       650       660       670       680       690       700
               |         |         |         |         |         |         |         |         |         |
601 ctgtatccgtttacctgggatgccgtacgttacaacggcaagctgattgcttacccgatcgctgttgaagcgttatcgctgatttataacaaagatctgc
    L  Y  P  F  T  W  D  A  V  R  Y  N  G  K  L  I  A  Y  P  I  A  V  E  A  L  S  L  I  Y  N  K  D  L  L
    200
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    MBP

             710       720       730       740       750       760       770       780       790       800
               |         |         |         |         |         |         |         |         |         |
701 tgccgaacccgccaaaaacctgggaagagatcccggcgctggataaagaactgaaagcgaaaggtaagagcgcgctgatgttcaacctgcaagaaccgta
      P  N  P  P  K  T  W  E  E  I  P  A  L  D  K  E  L  K  A  K  G  K  S  A  L  M  F  N  L  Q  E  P  Y
    234
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    MBP

             810       820       830       840       850       860       870       880       890       900
               |         |         |         |         |         |         |         |         |         |
801 cttcacctggccgctgattgctgctgacgggggttatgcgttcaagtatgaaaacggcaagtacgacattaaagacgtgggcgtggataacgctggcgcg
     F  T  W  P  L  I  A  A  D  G  G  Y  A  F  K  Y  E  N  G  K  Y  D  I  K  D  V  G  V  D  N  A  G  A
    267
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    MBP
```


910 920 930 940 950 960 970 980 990 1000
901 aaagcgggtctgaccttcctggttgacctgattaaaaacaaacacatgaatgcagacaccgattactccatcgcagaagctgcctttaataaaggcgaaa
K A G L T F L V D L I K N K H M N A D T D Y S I A E A A F N K G E T
300
MBP
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
1001 cagcgatgaccatcaacggcccgtgggcatggtccaacatcgacaccagcaaagtgaattatggtgtaacggtactgccgaccttcaagggtcaaccatc
A M T I N G P W A W S N I D T S K V N Y G V T V L P T F K G Q P S
334
MBP
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
1101 caaaccgttcgttggcgtgctgagcgcaggtattaacgccgccagtccgaacaaagagctggcaaaagagttcctcgaaaactatctgctgactgatgaa
K P F V G V L S A G I N A A S P N K E L A K E F L E N Y L L T D E
367
MBP
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
1201 ggtctggaagcggttaataaagacaaaccgctgggtgccgtagcgctgaagtcttacgaggaagagttggtgaaagatccgcgtattgccgccactatgg
G L E A V N K D K P L G A V A L K S Y E E E L V K D P R I A A T M E
400
MBP
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
1301 aaaacgcccagaaaggtgaaatcatgccgaacatcccgcagatgtccgctttctggtatgccgtgcgtactgcggtgatcaacgccgccagcggtcgtca
N A Q K G E I M P N I P Q M S A F W Y A V R T A V I N A A S G R Q
434
MBP
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
1468 KpnI
1401 gactgtcgatgaagccctgaaagacgcgcagactggtagcggtggtggttcaggtggcggtagcggtggtaccacaccaacacctacgccgaccagcggt
T V D E A L K D A Q T G S G G G S G G G S G G T T P T P T P T S G
467
MBP Linker PT-Linker_short CBD_cex
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
1501 ccggcaggttgtcaggttctgtggggtgttaatcagtggaataccggctttaccgcaaatgttaccgtgaaaaataccagcagcgctccggttgatggtt
P A G C Q V L W G V N Q W N T G F T A N V T V K N T S S A P V D G W
500
CBD_cex
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
1601 ggaccctgacctttagctttccgagcggtcagcaggttacccaggcatggtcaagcaccgttacacagagcggtagcgcagttaccgttcgtaatgcacc
T L T F S F P S G Q Q V T Q A W S S T V T Q S G S A V T V R N A P
534
CBD_cex
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
1701 gtggaatggtagcattccggcaggcggtacagcacagtttggttttaatggttcacataccggcaccaatgcagcaccgaccgcatttagtctgaatggc
W N G S I P A G G T A Q F G F N G S H T G T N A A P T A F S L N G
567
CBD_cex

```
            1810      1820      1830      1840      1850
              |         |         |         |         |
                                                    1849 XhoI
                                                    |
1801 accccgtgtaccgttggtagcggtggtcatcatcaccatcatcattaactcgag
     T  P  C  T  V  G  S  G  G  H  H  H  H  H  H  *
     600
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>

     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     CBD_cex           Linker    His6-tag
```

10.4 **CHO K1 genome editing**

10.4.1 **CHO K1 loci coverage**

Table 33: Loci in CHO K1 genome sorted by RNA sequencing coverage

Locus	Contig	Start	Stop	Coverage
KE379267:1-50000	KE379267	1	50000	662
KE379448:200001-250000	KE379448	200001	250000	496
KE376759:200001-250000	KE376759	200001	250000	485
KE382521:300001-350000	KE382521	300001	350000	483
KE376858:300001-350000	KE376858	300001	350000	477
KE379610:300001-350000	KE379610	300001	350000	464
KE380250:250001-300000	KE380250	250001	300000	443
KE379288:1250001-1300000	KE379288	1250001	1300000	422
KE381908:100001-150000	KE381908	100001	150000	406
KE378070:200001-250000	KE378070	200001	250000	380
KE379835:350001-400000	KE379835	350001	400000	377
KE378348:500001-550000	KE378348	500001	550000	377
KE379560:750001-800000	KE379560	750001	800000	360
KE379359:450001-500000	KE379359	450001	500000	346
KE377509:400001-450000	KE377509	400001	450000	328
KE380058:1-50000	KE380058	1	50000	321
KE379537:1700001-1750000	KE379537	1700001	1750000	313
KE381285:750001-800000	KE381285	750001	800000	308
KE380159:1350001-1400000	KE380159	1350001	1400000	289
KE378234:800001-850000	KE378234	800001	850000	280
KE377500:300001-350000	KE377500	300001	350000	280
KE377199:2050001-2100000	KE377199	2050001	2100000	278
KE376759:250001-300000	KE376759	250001	300000	274
KE379820:1-50000	KE379820	1	50000	272
KE377738:2150001-2200000	KE377738	2150001	2200000	266
KE379941:200001-250000	KE379941	200001	250000	262
KE379817:1700001-1750000	KE379817	1700001	1750000	262
KE379392:850001-900000	KE379392	850001	900000	260
KE378480:1-50000	KE378480	1	50000	260

Locus	Contig	Start	Stop	Coverage
KE379359:400001-450000	KE379359	400001	450000	258
KE378097:50001-100000	KE378097	50001	100000	257
KE378347:750001-800000	KE378347	750001	800000	256
KE382159:1850001-1900000	KE382159	1850001	1900000	250
KE379184:1000001-1050000	KE379184	1000001	1050000	248
KE383531:250001-300000	KE383531	250001	300000	244
KE379052:400001-450000	KE379052	400001	450000	244
KE376951:1900001-1950000	KE376951	1900001	1950000	239
KE382432:1-50000	KE382432	1	50000	238
KE380961:300001-350000	KE380961	300001	350000	237
KE376593:200001-250000	KE376593	200001	250000	237
KE377809:850001-900000	KE377809	850001	900000	236
KE379583:1-50000	KE379583	1	50000	228
KE378618:1-50000	KE378618	1	50000	228
KE383660:500001-550000	KE383660	500001	550000	221
KE383081:350001-400000	KE383081	350001	400000	220
KE379560:1000001-1050000	KE379560	1000001	1050000	219
KE379403:1400001-1450000	KE379403	1400001	1450000	219
KE379621:1550001-1600000	KE379621	1550001	1600000	217
KE381346:250001-300000	KE381346	250001	300000	216
KE377743:100001-150000	KE377743	100001	150000	214
KE381544:50001-100000	KE381544	50001	100000	213
KE376355:2000001-2050000	KE376355	2000001	2050000	212
KE376700:50001-100000	KE376700	50001	100000	211
KE383301:50001-100000	KE383301	50001	100000	209
KE380969:100001-150000	KE380969	100001	150000	208
KE380305:1-50000	KE380305	1	50000	208
KE379392:900001-950000	KE379392	900001	950000	205
KE378095:550001-600000	KE378095	550001	600000	205
KE379273:250001-300000	KE379273	250001	300000	203
KE377441:150001-200000	KE377441	150001	200000	203
KE381716:100001-150000	KE381716	100001	150000	202
KE379560:950001-1000000	KE379560	950001	1000000	201
KE381566:750001-800000	KE381566	750001	800000	199
KE380701:50001-100000	KE380701	50001	100000	198
KE379284:4500001-4550000	KE379284	4500001	4550000	198
KE378347:1100001-1150000	KE378347	1100001	1150000	198
KE379882:350001-400000	KE379882	350001	400000	196
KE379204:1650001-1700000	KE379204	1650001	1700000	195
KE379128:1-50000	KE379128	1	50000	195
KE380150:2200001-2250000	KE380150	2200001	2250000	193

Locus	Contig	Start	Stop	Coverage
KE379641:250001-300000	KE379641	250001	300000	191
KE379248:200001-250000	KE379248	200001	250000	191
KE376951:1850001-1900000	KE376951	1850001	1900000	190
KE379987:200001-250000	KE379987	200001	250000	189
KE379597:800001-850000	KE379597	800001	850000	189

Yellow marked loci were selected for further analysis

10.4.2 **CHO K1 intergenic regions for CRISPR target search**

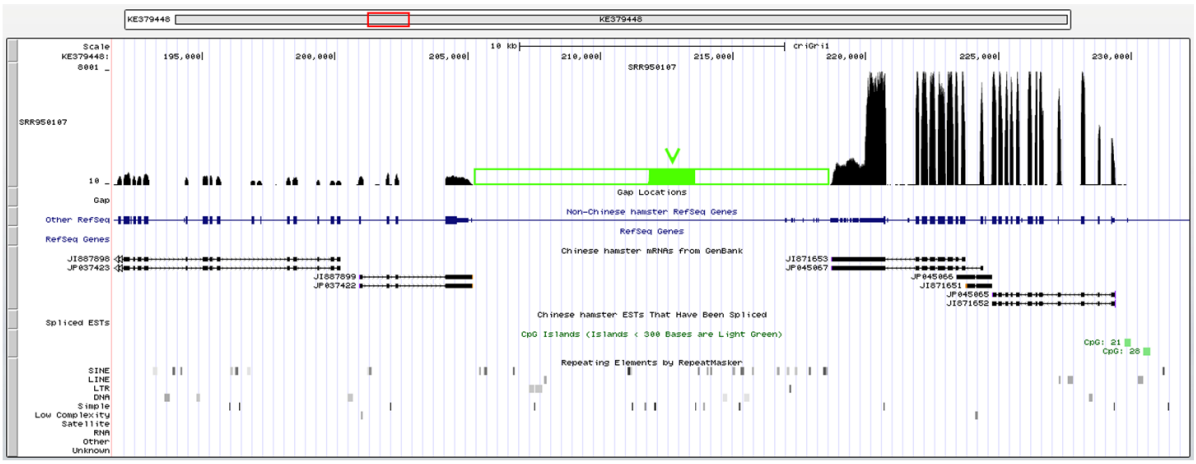
Table 34: Chosen CHO K1 intergenic regions for CRISPR target search

Contig	Start base	Stop base
KE377809	877277	883265
KE376951	1945212	1969988
KE379448	205175	218684
KE379359	363758	386309
KE378234	820260	823908
KE382432	17683	22256
KE381908	91237	101715
KE376759	231735	242331
KE380701	101368	107560
KE379248	234400	245673
KE377743	159117	172950
KE379184	956278	968708

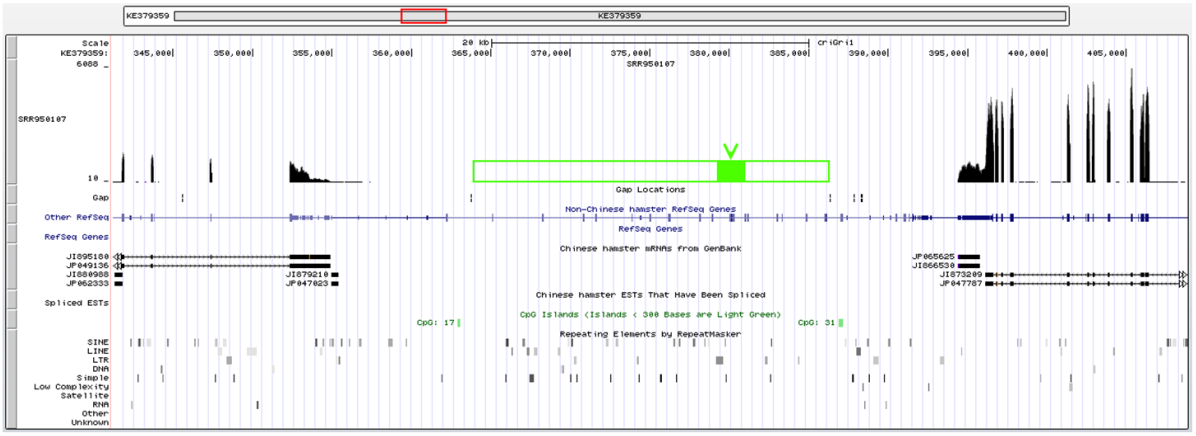
10.4.3 **Selected CHO K1 CRISPR target sites**

Figure 52A-D shows screenshots from the UCSC Genome Browser of the four selected CRISPR target sites. The rows marked with “SRR950107” show the respective coverages of the RNA sequencing.

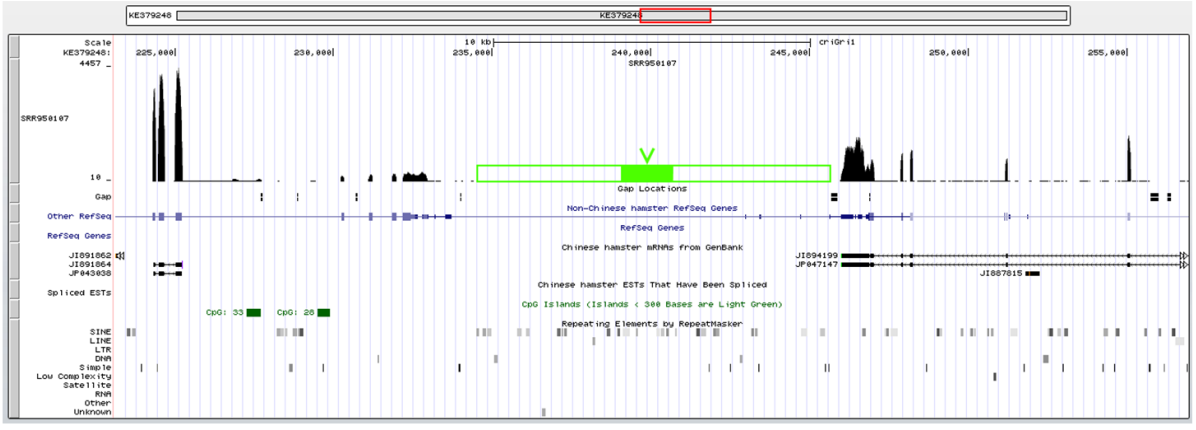
A


KE379448
Scale
SRR950107
Gap Locations
Non-Chinese hamster RefSeq Genes
RefSeq Genes
Chinese hamster mRNAs from GenBank
Chinese hamster ESTs That Have Been Spliced
CpG Islands (Islands < 300 Bases are Light Green)
Repeating Elements by RepeatMaster
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown

B


KE379359
Scale
SRR950107
Gap Locations
Non-Chinese hamster RefSeq Genes
RefSeq Genes
Chinese hamster mRNAs from GenBank
Chinese hamster ESTs That Have Been Spliced
CpG Islands (Islands < 300 Bases are Light Green)
Repeating Elements by RepeatMaster
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown

C


KE379248
Scale
SRR950107
Gap Locations
Non-Chinese hamster RefSeq Genes
RefSeq Genes
Chinese hamster mRNAs from GenBank
Chinese hamster ESTs That Have Been Spliced
CpG Islands (Islands < 300 Bases are Light Green)
Repeating Elements by RepeatMaster
SINE
LINE
LTR
DNA
Simple
Low Complexity
Satellite
RNA
Other
Unknown

D

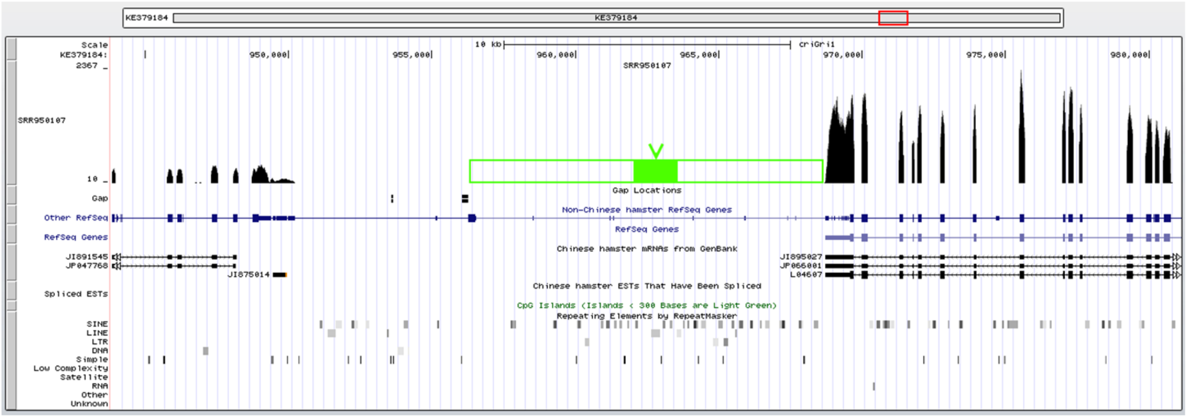


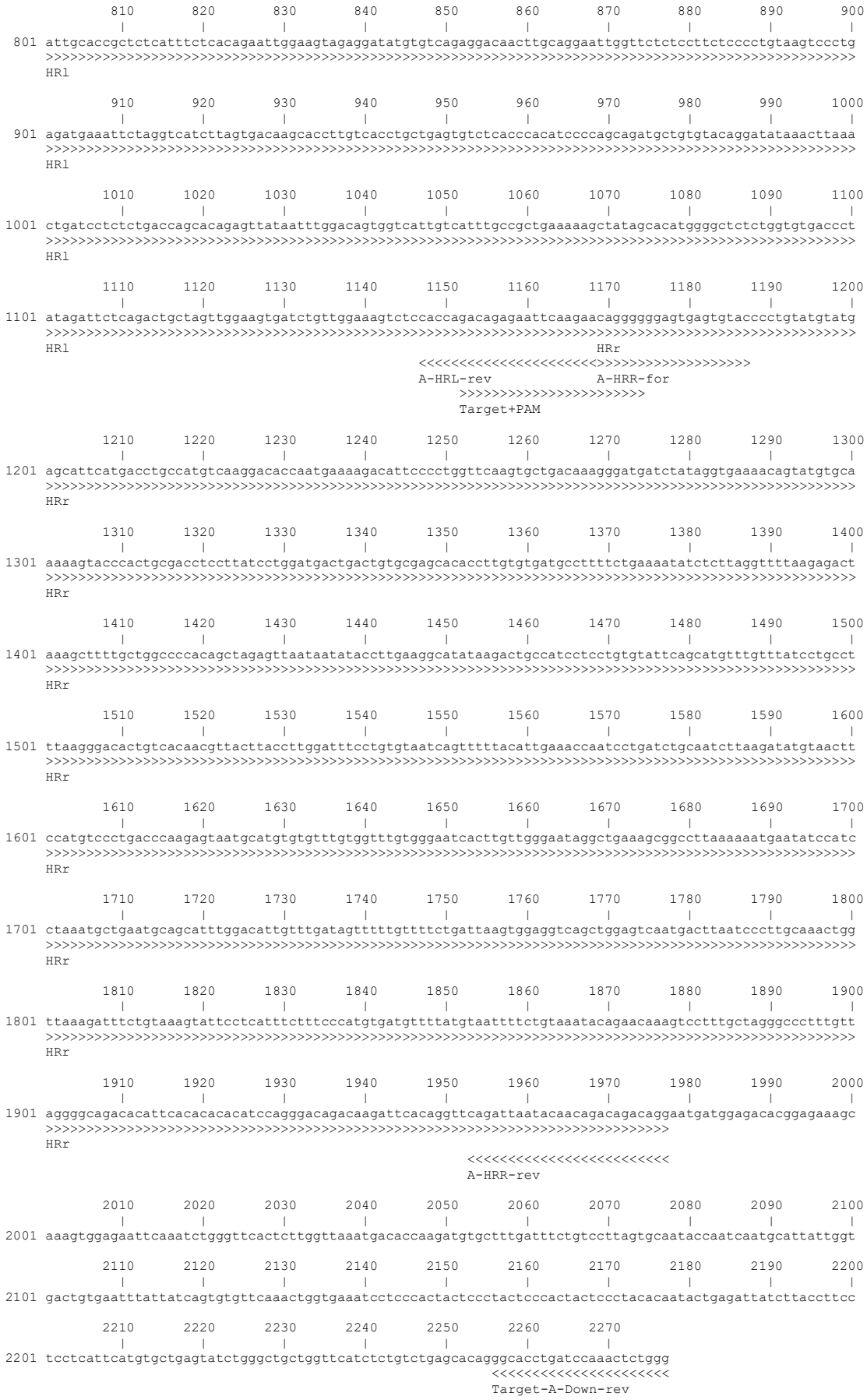
Figure 52: Screenshots from the UCSC Genome Browser of the four selected CRISPR target sites. A-D: Target site A-D. Green arrow: target site position, Green box: intergenic region, Solid green box: selected region for homologues arms.

10.4.4 **Genomic regions of CRISPR target sites**

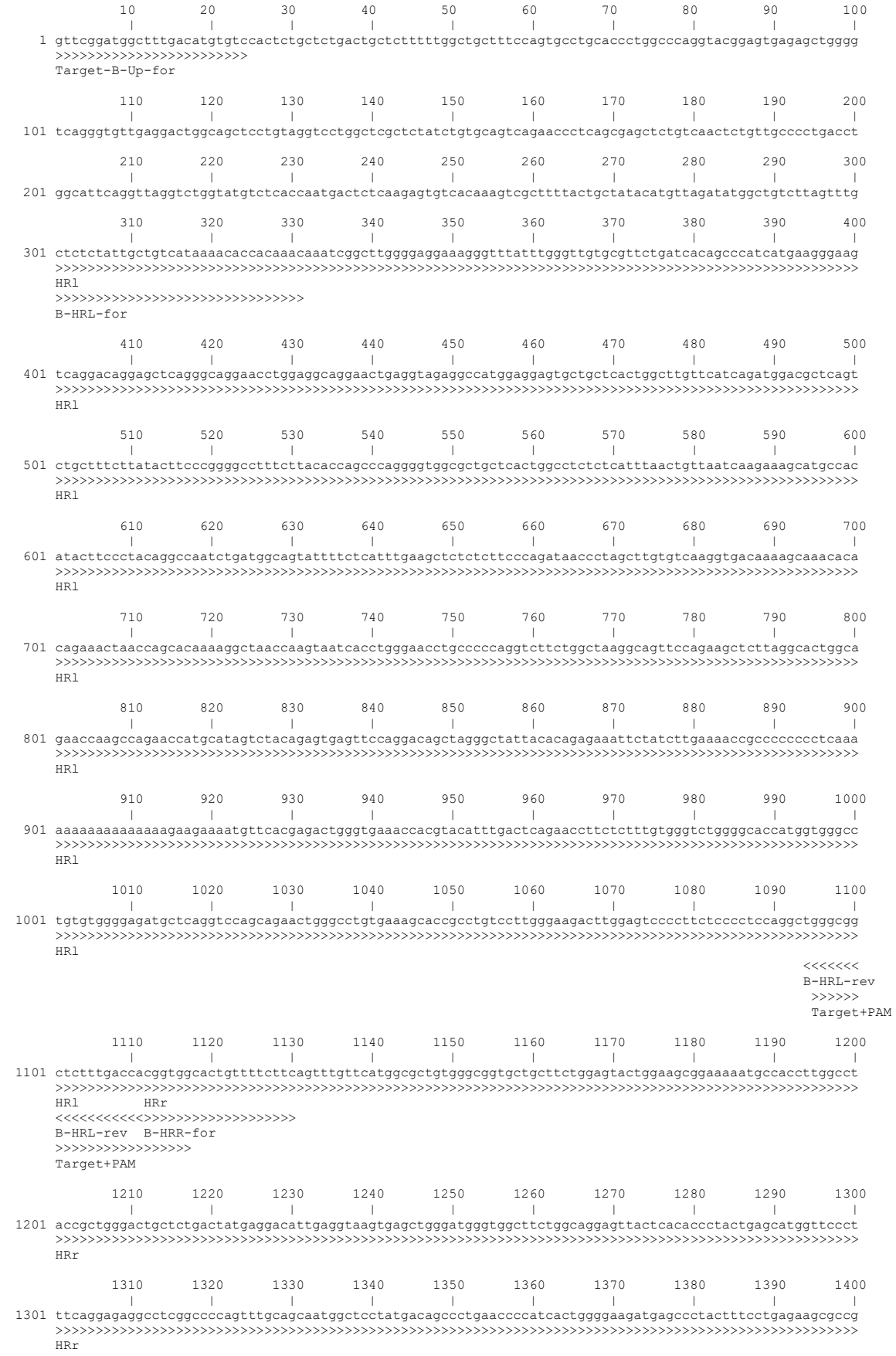
In the following subchapters the genomic regions of the selected CRISPR target sites A to D are presented. In addition to the actual target sequences including PAM, the exact homologous arms and surrounding genomic sequences, which were expected as results of the identification sequences, are shown.

10.4.4.1 **Genomic region of CRISPR target site A**

```
             10        20        30        40        50        60        70        80        90       100
              |         |         |         |         |         |         |         |         |         |
  1 gagtgctgtcaggagcagttgtgcctcatgccaacagagttctaaattattaaaacatttaagtgccatattctgtaggcctttgaagtgtttgaagatc
    >>>>>>>>>>>>>>>>>>>>>>>>
    Target-A-Up-for

            110       120       130       140       150       160       170       180       190       200
              |         |         |         |         |         |         |         |         |         |
101 aaatatctatgctgagtacatttctgtgcatctagagaacctgactaacaaagctaagtatgagaagcatgagtaggtacaacttgtaattctcaaaaac

            210       220       230       240       250       260       270       280       290       300
              |         |         |         |         |         |         |         |         |         |
201 acatattgcttactgactgaagatataacaatctttaaacagatttgcagcataaggacaatgacctcaaagttgtaactatgtacatatatatgtaaat

            310       320       330       340       350       360       370       380       390       400
              |         |         |         |         |         |         |         |         |         |
301 atcttgatcagaggtagaactgtatagtgggaaatgacagaaatactcttaaattgtatcatcatacaaaataacttaagcatatgtagaaacatataaa
                >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                HRl
                >>>>>>>>>>>>>>>>>>>>>>>>
                A-HRL-for

            410       420       430       440       450       460       470       480       490       500
              |         |         |         |         |         |         |         |         |         |
401 tatatatacaatatgataaaaataactacatttgtgtgcaaaaacattggaaacaaataaactatacaatatgacaaatataatttcaagtttgaaacaa
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    HRl

            510       520       530       540       550       560       570       580       590       600
              |         |         |         |         |         |         |         |         |         |
501 tatacaaactatgtaaaacaggaatacaaaattaattttatataaatatcagtatacaacaatctatatcaatgtaagttgtccatagctaatagacaac
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    HRl

            610       620       630       640       650       660       670       680       690       700
              |         |         |         |         |         |         |         |         |         |
601 ttacaaccaaagccacactctaaggccagtacctcgaatctattgatggaacaagatgccactacaaccaggcttgacagtaggtgcctttgtctgttga
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    HRl

            710       720       730       740       750       760       770       780       790       800
              |         |         |         |         |         |         |         |         |         |
701 gcatctcatgtacgtgcacacatttgcacacacgcatacacgcatacacacacacacacacacacacacacacacacacacacacacacagtctctta
    >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    HRl
```


HRl
HRr
A-HRL-rev
A-HRR-for
Target+PAM
A-HRR-rev
Target-A-Down-rev

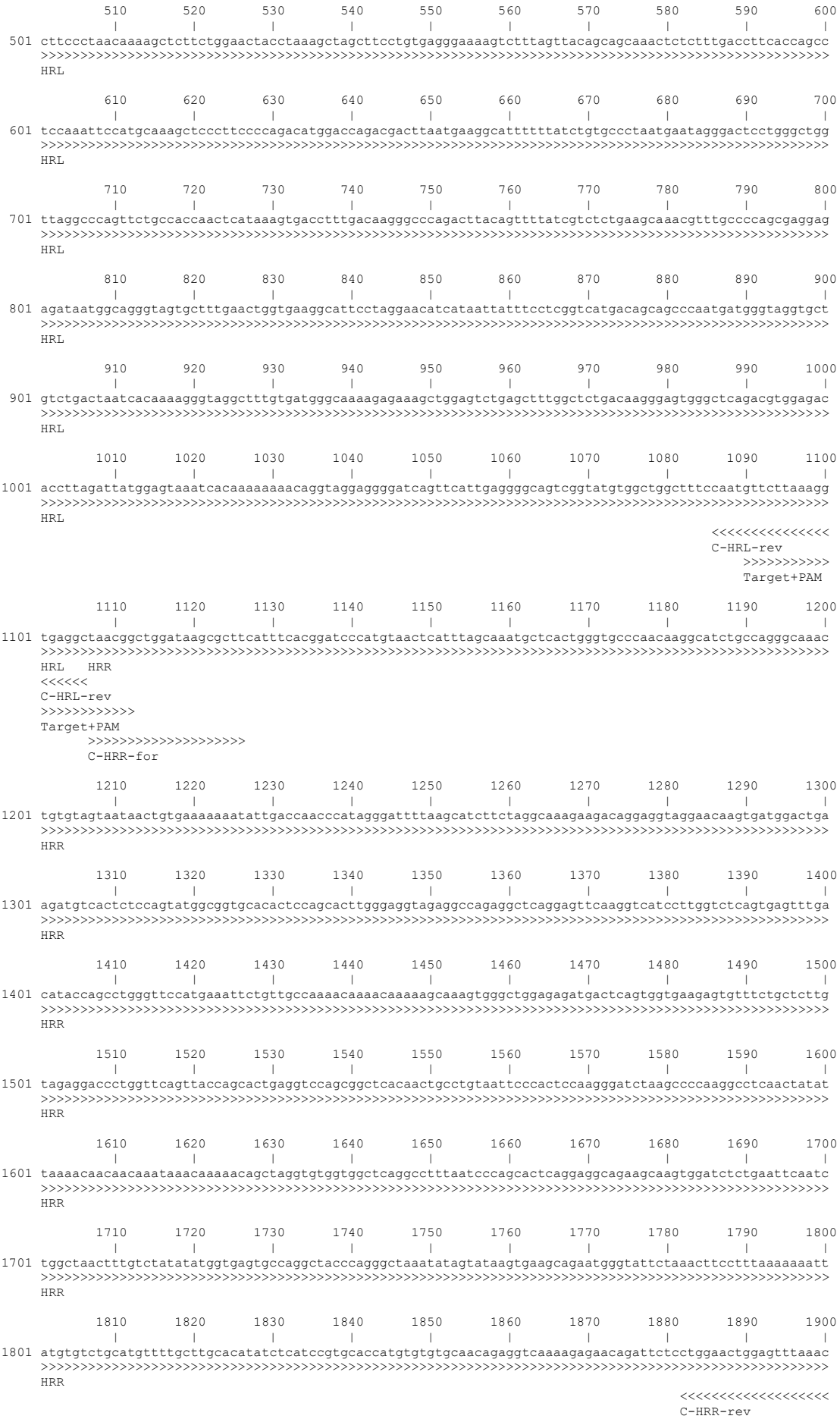
10.4.4.2 **Genomic region of CRISPR target site B**



```
            1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
              |         |         |         |         |         |         |         |         |         |
1401 tgtacaccgcattctggctggctctgtggtttctcatgatgatggtgaggggggtgggggtgggagggatcctgccaccctgcctttctcttctgtgtctgt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr

            1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
              |         |         |         |         |         |         |         |         |         |
1501 cttcctgagccacaggggccattccactcttgctatctcctccagacataagaccctcactgccctccagaaaggccctgtgctttgtgcctcagggaag
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr

            1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
              |         |         |         |         |         |         |         |         |         |
1601 gaaggaccttagcccagggccaagtgataggttttcatacggggtgtgtatgtgtatgtgccatgaggggctctggaagagcttctccctatgtgacctg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr

            1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
              |         |         |         |         |         |         |         |         |         |
1701 gggcccccattaaacaagggaggcacactaatccctttggtaccttctagccaaagactgctgaaaaggacccccacccctttccaggtactatggacccc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr

            1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
              |         |         |         |         |         |         |         |         |         |
1801 aatttagggctgtaaagaggtgggtgagtggggcacagttctccagcactaggtgtaaggacagcatacagtagctatctggcatcctgacctcccagtc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr
                                                                                  <<<<<<<<<<<<<<<<
                                                                                  B-HRR-rev

            1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
              |         |         |         |         |         |         |         |         |         |
1901 ctagcctgccccacctctgcagaacccaggggacacaggcaggttgcagatgggagcccctggaaaaggaacctgtggagctgctatggacccagaagcc
     >>>>
     HRr
     <<<<
     B-HRR-rev

            2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
              |         |         |         |         |         |         |         |         |         |
2001 tgtgctagcctacggcctgaaggtgtgggtgctgggcccatggcaggggactcaggacttctctggtcctgtaggtggcggtggtgattatgtgccttgt

            2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
              |         |         |         |         |         |         |         |         |         |
2101 gtctatcatcctgtaccgagccatcatggctgtaattgtgtccaagtccgacaatgccttcctctcggcctgggtgagtctctgctgccggttctccttg
                                                                                  <<<<<<<<<<<<<<<<
                                                                                  Target-B-Down-rev



2201 cgg
     <<<
     Target-B-Down-rev
```

10.4.4.3 **Genomic region of CRISPR target site C**

```
              10        20        30        40        50        60        70        80        90        100
               |         |         |         |         |         |         |         |         |         |
   1 caagggaaatggcaagacacctgcttacctgcagtctttgtatttacctgaccctcctccgggaccttgggaatgctttaaagaaagggaaggggctgag
     >>>>>>>>>>>>>>>>>>>>>
     Target-C-Up-for

             110       120       130       140       150       160       170       180       190       200
               |         |         |         |         |         |         |         |         |         |
 101 gagacagctcagttggtaaaggaattgcatgtaaacctgaggaccggagtttgacccccagtagtcatgaaaaaagcagggcatggtagtatgggaggtg

             210       220       230       240       250       260       270       280       290       300
               |         |         |         |         |         |         |         |         |         |
 201 gttcaagaagaatcctagggcttacggctggctaaagatcctggtgagagaccttgtcaccaaaaaagtgtttgtcaggtatggtaatatactccttttag
                                                                                                >>>>>>>>>>>
                                                                                                HRL
                                                                                                >>>>>>>>>>>
                                                                                                C-HRL-for

             310       320       330       340       350       360       370       380       390       400
               |         |         |         |         |         |         |         |         |         |
 301 tcccaacatttgggaggcagagggaaaggatctgtgattttgaggccagcctggtctgtatgatgggtcaggccagccatggccacacagtgagagtgaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRL
     >>>>>>>>>>>>
     C-HRL-for

             410       420       430       440       450       460       470       480       490       500
               |         |         |         |         |         |         |         |         |         |
 401 accctgtctaaaacgaccaaacaaaagccaaggtgtttggccacatgaagacttctggtcgccacatgcccagcacacatgtacacacacaaacatacgc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRL
```


510 520 530 540 550 560 570 580 590 600
501 cttccctaacaaaagctcttctggaactacctaaagctagcttcctgtgagggaaaagtctttagttacagcagcaaactctctttgaccttcaccagcc
HRL
610 620 630 640 650 660 670 680 690 700
601 tccaaattccatgcaaagctccctttccccagacatggaccagacgacttaatgaaggcatttttttatctgtgccctaatgaatagggactcctgggctgg
HRL
710 720 730 740 750 760 770 780 790 800
701 ttaggcccagttctgccaccaactcataaagtgacctttgacaagggcccagacttacagttttatcgtctctgaagcaaacgtttgccccagcgaggag
HRL
810 820 830 840 850 860 870 880 890 900
801 agataatggcagggtagtgctttgaactggtgaaggcattcctaggaacatcataattatttcctcggtcatgacagcagcccaatgatgggtaggtgct
HRL
910 920 930 940 950 960 970 980 990 1000
901 gtctgactaatcacaaaagggtaggctttgtgatgggcaaaagagaaagctggagtctgagctttggctctgacaagggagtgggctcagacgtggagac
HRL
1010 1020 1030 1040 1050 1060 1070 1080 1090 1100
1001 accttagattatggagtaaatcacaaaaaaaacaggtaggaggggatcagttcattgaggggcagtcggtatgtggctggctttccaatgttcttaaagg
HRL
<<<<<<<<<<<<<<
C-HRL-rev
>>>>>>>>>>>
Target+PAM
1110 1120 1130 1140 1150 1160 1170 1180 1190 1200
1101 tgaggctaacggctggataagcgcttcatttcacggatcccatgtaactcatttagcaaatgctcactgggtgcccaacaaggcatctgccagggcaaac
HRL HRR
<<<<<<
C-HRL-rev
>>>>>>>>>>>>
Target+PAM
>>>>>>>>>>>>>>>>>>>>>
C-HRR-for
1210 1220 1230 1240 1250 1260 1270 1280 1290 1300
1201 tgtgtagtaataactgtgaaaaaaatattgaccaaccccatagggattttaagcatcttctaggcaaagaagacaggaggtaggaacaagtgatggactga
HRR
1310 1320 1330 1340 1350 1360 1370 1380 1390 1400
1301 agatgtcactctccagtatggcggtgcacactccagcacttgggaggtagaggccagaggctcaggagttcaaggtcatccttggtctcagtgagtttga
HRR
1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
1401 cataccagcctgggttccatgaaattctgttgccaaaacaaaacaaaaagcaaagtgggctggagagatgactcagtggtgaagagtgtttctgctcttg
HRR
1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
1501 tagaggaccctggttcagttaccagcactgaggtccagcggctcacaactgcctgtaattcccactccaagggatctaagccccaaggcctcaactatat
HRR
1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
1601 taaaacaacaacaaataaacaaaaacagctaggtgtggtggctcaggcctttaatcccagcactcaggaggcagaagcaagtggatctctgaattcaatc
HRR
1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
1701 tggctaactttgtctatatatggtgagtgccaggctacccagggctaaatatagtataagtgaagcagaatgggtattctaaacttcctttaaaaaaatt
HRR
1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
1801 atgtgtctgcatgttttgcttgcacatatctcatccgtgcaccatgtgtgtgcaacagaggtcaaaagagaacagattctcctggaactggagtttaaac
HRR
<<<<<<<<<<<<<<<<<<<
C-HRR-rev

```
                1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
                 |         |         |         |         |         |         |         |         |         |
   1901 agttgacagcagtcaggtgagtactgggaattgaacctgagtcctctggaagagcagccagtgctcttagtcaccaagtcatctctccagctccaaattg
        >>>>>>
        HRR
        <<<<<<
        C-HRR-rev

                2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
                 |         |         |         |         |         |         |         |         |         |
   2001 ggtattctaagtggggagggattctgggctgggaaatcaggggaggctttgtgaaggacaggggtgaagcaagattcaggccttgcagagatggctgat

                2110      2120      2130      2140      2150      2160      2170      2180      2190      2200
                 |         |         |         |         |         |         |         |         |         |
   2101 cagtactcattacagcttgggcaaaatttgtgggttcctacacatggtagccagagcacccaaaagggatcactgggtgacatattaaaggattcggttt
                                                                                                      <<<<<<
                                                                                                      Target-C-Down-
rev

                2210      2220
                 |         |
   2201 aagaatgagaactcagactgac
        <<<<<<<<<<<<<<<<<<<<<<<
        Target-C-Down-rev
```

10.4.4.4 **Genomic region of CRISPR target site D**

```
                10        20        30        40        50        60        70        80        90       100
                 |         |         |         |         |         |         |         |         |         |
      1 gtggtagtacatttgggtaattctcgtactctggatgttgagacaggaggattgcaagctctaggccatcctggatcatagagcaagcctgtctcaaaac
        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>
        Target-D-Up-for

                110       120       130       140       150       160       170       180       190       200
                 |         |         |         |         |         |         |         |         |         |
    101 aaaagaaatctgcatctctttgtgtatctatgtatacaagcagtcacaatcactgtatttctgcctggtcctggccgccattaccccaggcttactttac

                210       220       230       240       250       260       270       280       290       300
                 |         |         |         |         |         |         |         |         |         |
    201 tacccggtacttggtttggtcttgagagcactctgaggtcttctccagcatctggccaggaggatgcagacagaaccacaatctggtcagcccaaaccaa

                310       320       330       340       350       360       370       380       390       400
                 |         |         |         |         |         |         |         |         |         |
    301 tggcagccatgccttacctggaagcagctctccatccaaaggactgagcttgaaactggtcctagaaagttatcggagtcactggagtcagaaggcaagg
                 >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                 HRl
                 >>>>>>>>>>>>>>>>>>>>
                 D-HRL-for

                410       420       430       440       450       460       470       480       490       500
                 |         |         |         |         |         |         |         |         |         |
    401 tcacctgcagaccaagctttgtcctttcctctagctaaagtgacatttactggtatactttttaatttgaactggggcaacctgaggtaagtataacatt
        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
        HRl

                510       520       530       540       550       560       570       580       590       600
                 |         |         |         |         |         |         |         |         |         |
    501 ttacaccaaagaaacacatctgattctcaaggaaaaatttaagtataaatgaaacctgtatttggaaagacagctcttcagtttgtagacttaaacatgg
        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
        HRl

                610       620       630       640       650       660       670       680       690       700
                 |         |         |         |         |         |         |         |         |         |
    601 gacctgatgagcccattgccaggaattccacccactgttgaaaaaaaaaaaaaacccaacgtgaaatgggtacatactgtacaattcacaacagctttac
        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
        HRl

                710       720       730       740       750       760       770       780       790       800
                 |         |         |         |         |         |         |         |         |         |
    701 tcaaaatagacccaagctggaaacaatccaattgtccatcagtaggtgaatggatcaacaaacccatatcaacaaactcaacaataataaaaagaaacta
        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
        HRl

                810       820       830       840       850       860       870       880       890       900
                 |         |         |         |         |         |         |         |         |         |
    801 tatttttgtctgttttgaaaaaggaactactgatacatgcaatgtgggtacattttaaatttatttaaaatttttattgcattttgaaggcttggcat
        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
        HRl

                910       920       930       940       950       960       970       980       990      1000
                 |         |         |         |         |         |         |         |         |         |
    901 ttaccaccatactgtgtgtgtgtgtgtgtgtcaaacaaaactacatatattttgatttgaaacagagttttactatgtagccctggctggcttagaactc
        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
        HRl

                1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
                 |         |         |         |         |         |         |         |         |         |
   1001 tgtagaccaaattgtcctcaaactcacagacttgcaccaccactgcctgggttgttttatggttttaattttatgtgtatgagtgttttgcctgcatgtt
        >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
        HRl

                                                                                                      <<<<<<
                                                                                                      D-HRL-rev
```

```
            1110      1120      1130      1140      1150      1160      1170      1180      1190      1200
              |         |         |         |         |         |         |         |         |         |
1101 tgtctgtgtaccacgagtgcttggtgcctcagagggtcagagaagatattagatccctggaaactagagttacaaatggatttgaactggagttatgggt
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRl                HRr
     <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
     D-HRL-rev          D-HRR-for
      >>>>>>>>>>>>>>>>>>>>>>>>>
      Target+PAM

            1210      1220      1230      1240      1250      1260      1270      1280      1290      1300
              |         |         |         |         |         |         |         |         |         |
1201 ggtagtgaattatcatgtaggtgctgggaattgaaccctggttctttgaaagagcaaaaaatgctcctaaccactaagccatctctctagcccctaatat
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr

            1310      1320      1330      1340      1350      1360      1370      1380      1390      1400
              |         |         |         |         |         |         |         |         |         |
1301 ttttatttttaaaagtcattactgtgggtgagaatgcatgtgtcatggcaaacaaattgaagtcagggcacaattttgtggagttggttctctcctaga
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr

            1410      1420      1430      1440      1450      1460      1470      1480      1490      1500
              |         |         |         |         |         |         |         |         |         |
1401 ttttcaagtttgatggttggggacaagctcttttacccactgagccatcctctagcccctaacaatatgaataaaatttaaagcatcatactaaataaaa
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr

            1510      1520      1530      1540      1550      1560      1570      1580      1590      1600
              |         |         |         |         |         |         |         |         |         |
1501 ggagctaatcacagaaacaatgacactgttattccattcctgggaggtttgtggaacaggaaaaagtgacctgtagctgtggaaagcagacccgtgactg
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr

            1610      1620      1630      1640      1650      1660      1670      1680      1690      1700
              |         |         |         |         |         |         |         |         |         |
1601 agcccagagactgtttaaagaggcacgctagaaacacagaggtgacacaaagccttctaacggccggatgtggtggcacatgtctgaggtaggaagatca
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr

            1710      1720      1730      1740      1750      1760      1770      1780      1790      1800
              |         |         |         |         |         |         |         |         |         |
1701 aggccaggctgggctaaaaagcagaaacccatctcaacaaacaagaaaaccaccttgctttcagttcctaagaacgtgatcagttcctgagtcacctctc
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr

            1810      1820      1830      1840      1850      1860      1870      1880      1890      1900
              |         |         |         |         |         |         |         |         |         |
1801 aagctgttaagtgtgtctgagcttcgtggttgtgctcgcctttaatctcaggcagaggcaggtggctctcttgtgagttccaggacagcctggtctacag
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     HRr
                                                                                                 <<<<<<<<
                                                                                                 D-HRR-rev

            1910      1920      1930      1940      1950      1960      1970      1980      1990      2000
              |         |         |         |         |         |         |         |         |         |
1901 agagttccaggacaatcaggactacacagtgaaacctgtctttaaaacaccaacaccctcccctcaaattgcttctgagccccacagttctggggtgtgc
     >>>>>>>>>>>>>>>>>>>>
     HRr
     <<<<<<<<<<<<<<<<<<<<
     D-HRR-rev

            2010      2020      2030      2040      2050      2060      2070      2080      2090      2100
              |         |         |         |         |         |         |         |         |         |
2001 tcaatactctgtatttttctttctttaaaaaaacaaaaaactggctgggcgttggtggggcacgcctttaatcccagcactcgggaggcagaggtaggtgga

            2110      2120      2130      2140      2150      2160      2170      2180
              |         |         |         |         |         |         |         |
2101 tctctgtgagttcgagaccagcctggtctacaagagctagttccaggacaggctccaaagccacagagaaaccctgtctcgg
                                                              <<<<<<<<<<<<<<<<<<<<<<
                                                              Target-D-Down-rev
```

11 **Eigenständigkeitserklärung**

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig und ohne unzulässige fremde Hilfe verfasst habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und unveröffentlichten Publikationen stammen, sind als solche kenntlich gemacht. Diese Arbeit hat in dieser oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Bielefeld, den

_____

(Julian Teschner)