Aus dem Institut für Virologie
des Fachbereichs Veterinärmedizin
der Freien Universität Berlin

Molecular mechanism of virus attenuation by codon pair deoptimization

Inaugural-Dissertation

zur Erlangung des Grades eines

PhD of Biomedical Sciences

an der Freien Universität Berlin

vorgelegt von

Nicole Groenke

aus Berlin, Deutschland

Berlin 2020

Journal-Nr.: 4187

Gedruckt mit Genehmigung

des Fachbereichs Veterinärmedizin

der Freien Universität Berlin

Dekan:	Univ.-Prof. Dr. Jürgen Zentek
Erster Gutachter:	Univ.-Prof. Dr. Nikolaus Osterrieder
Zweiter Gutachter:	PD Dr. Michael Veit
Dritter Gutachter:	Univ.-Prof. Dr. Jörg Aschenbach

Deskriptoren (nach CAB-Thesaurus): influenza viruses, virulence, attenuation, codons, polymerase chain reaction, western blotting, SDS-PAGE

Tag der Promotion: 24.02.2020

1 Table of contents

2 List of figures

3 List of tables

4 Abbreviations

A/WSN/1933	Human influenza A H1N1 strain from the UK in 1933
A549	Human alveolar basal epithelial cell line
AB	Antibody
Amp	Ampicillin
BrU	Bromouridine
BSA	Bovine serum albumin
CAI	Codon adaptation index
CPD	Codon pair deoptimization
CPE	Cytopathic effect
CpG	Cytosine nucleotide followed by guanine nucleotide
CPS	Codon pair score
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
dpi	Days post infection
E. coli	Escherichia coli
eGFP	Enhanced green fluorescent protein
ELISA	Enzyme-linked immunosorbent assay
GEO	Gene Expression Omnibus
h	Hour
HA	Hemagglutinin
HEK 293T	Human embryonic kidney cells 293 containing the SV40 Large T-antigen
HIV	Human immunodeficiency virus
HPRT1	Hypoxanthine phosphoribosyltransferase 1
IAV	Influenza A virus
kb	Kilobases
L	Litre

M1, M2	Matrix protein 1 and 2
MDCK	Madin-Darby canine kidney cell line
MERS	Middle East respiratory syndrome
MOI	Multiplicity of infection
mRNA	Messenger RNA
NA	Neuraminidase
NEP	Nuclear export protein
NP	Nucleoprotein
NS1	Non-structural protein 1
nt	Nucleotide
P/S	Penicillin/streptomycin
PA	Polymerase acidic protein
PB1	Polymerase basic protein 1
PB2	Polymerase basic protein 2
PBS	Phosphate saline buffer
PCR	Polymerase chain reaction
PEI	Polyethylenimin
PFU	Plaque forming unit
PNK	Polynucleotide kinase
RNA	Ribonucleic acid
RP	Ribosome Profiling
RT-qPCR	Quantitative reverse transcription PCR
SARS	Severe acute respiratory syndrome
SAVE	Synthetic attenuated virus engineering
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SPF	Specific-pathogen-free
TpA	Thymine nucleotide followed by adenine nucleotide
tRNA	Transfer RNA
UpA	Uracil nucleotide followed by adenine nucleotide

UTR	Untranslated regions
vRNA	Viral RNA
vRNP	Viral Ribonucleoprotein
WB	Western blot
WT	Wild type

5 Introduction

5.1 Influenza A virus

Influenza viruses are rapidly evolving viruses that cause annual epidemics worldwide. Despite of the ongoing improvement and use of influenza vaccines, influenza is still a substantial threat to the global public health. Influenza viruses cause, typically in the winter months, mild to very severe respiratory illnesses that can lead to death. Influenza infections cause about 300,000 deaths every year[1]. Influenza viruses belong to the virus family Orthomyxoviridae. The family has a broad host range and contains four virus genera: influenza viruses A, B, C and D[2]. Human infections are predominantly caused by influenza viruses A and B.

The virus particle (virion) of an influenza virus contains an outer lipid membrane, an envelope, and encapsulate seven to eight negative-sense, single-stranded RNA segments. The virion adopts either spherical, elliptical or filamentous shapes with a diameter of 80-120 nm[3]. In this work, I only focus on Influenza A virus (IAV), which contains eight segments that encode ten essential proteins, and a few strain specific proteins[4,5]. The eight viral RNA (vRNA) segments encode proteins of the polymerase complex (PA, PB1 and PB2), the surface proteins neuraminidase (NA) and hemagglutinin (HA), the nucleoprotein (NP), the matrix proteins (M1 and M2) and the non-structural protein (NS1 and NS2, also known as NEP) (*Figure 1*). Viral segments PB2, HA, NP and NA encode one and segments PA, PB1, M and NS encode two or more proteins. Matrix proteins M1 and M2 and non-structural proteins NS1 and NEP are splice products, which are produced by the host cell spliceosome.

The viral envelope contains viral proteins HA, NA and M2, as well as some host membrane proteins[6]. It exhibits a lipid bilayer that is derived from the host plasma membrane. The envelope is enriched in sphingolipids and cholesterol, which are acquired as the virus buds from specific apical plasma membrane regions so-called rafts[7]. Based on their surface glycoproteins HA and NA, IAVs are classified in 18 HA and 11 NA subtypes, for example, H1N1 or H3N2[8,9]. HA is the most abundant viral envelope protein (around 80 %), followed by NA, and only a few molecules of M2 are present. The M1 protein is located below the viral membrane and forms a matrix to stabilize the viral ribonucleoproteins (vRNPs). The eight vRNPs form the core of the virion and consist of one negative-stranded vRNA wrapped around several NP proteins (*Figure 1*). One terminus of the vRNP contains one copy of the heterotrimeric polymerase complex (PA, PB1 and PB2)[10] that is responsible for the transcription and replication of the viral RNA.

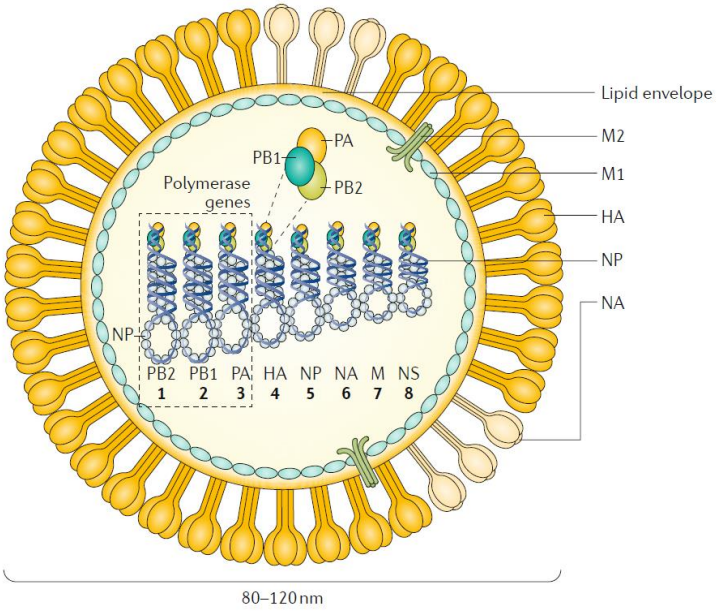

Figure 1: Influenza A virus particle. *Shown is an enveloped influenza A virus particle with its eight negative-sense RNA segments. It contains three RNA polymerases (PB1, PB2 and PA), the surface proteins HA, NA and M2, as well as the matrix protein M1 and the nuclear protein NP. The nonstructural protein (NS1) and the nuclear export protein (NEP) are not shown. The image is reused with permission from Krammer, F. et al. (2018)[11].*

5.2 Influenza virus lifecycle

Influenza is transmitted through the respiratory route in humans[11]. The viral surface protein HA is recognized by cellular sialic acid receptors. After the HA precursor (HA0) is cleaved by a host cell protease into the two active subunits (HA1 and HA2) which are linked by disulphide bonds[12], the cell imports the virus by endocytosis. Once inside the cell, the low pH of 5 in the endosome induces a conformational change of the HA protein. This conformational change exposes the fusion peptide (HA2), which inserts into the endosomal membrane. The receptor-binding domain (H1) anchors HA2 in the viral membrane to bring the viral and the endosomal membranes in close proximity enabling the formation of a fusion pore. The low pH in the endosome also opens up the M2 proton channels, which acidifies the viral particle and releases the vRNPs from M1 into the cytoplasm[5,13]. The vRNPs are then transported to the nucleus where the viral RNA transcription and replication occurs (*Figure 2*). For RNA replication, the viral RNA-dependent RNA polymerase transcribes the negative-sense vRNA initially in a positive sense complementary RNA (cRNA) by an unprimed process[14]. This complementary RNA is then transcribed into new vRNA copies.

The transcription from vRNA into mRNA is performed by the viral polymerase using capped primers derived from host transcripts, a process called cap snatching[15]. The PB2 subunit of

the viral polymerase binds to 5' caps of nascent host transcripts[16] and the PA subunit cleaves the host mRNA 10 to 13 nucleotides downstream of the 5' cap[17]. Using vRNA as a template, the viral polymerase extends these capped primers to synthesize mRNA transcripts. These transcripts are afterwards polyadenylated at the 3' end by the viral polymerase. The polymerase synthesizes the poly-A tail by repeated copying the five to seven uracil bases at the 5' end of the vRNA in several dissociation and reannealing cycles[18].

The translation of viral mRNA is dependent on the host cell and occurs in the cytoplasm. Newly synthesized membrane proteins HA, NA and M2 migrate directly to the plasma membrane into which they are inserted. The NP proteins and the viral polymerase (PA, PB1 and PB2) are targeted into the nucleus. Inside the nucleus, they support the synthesis of vRNA, viral mRNA and form new vRNPs. At later stages of the infection, the viral M1 and NEP protein are transported to the nucleus and bind the vRNPs to mediate their transport to the plasma membrane[11]. The vRNPs are bundled at the budding sites of the plasma membrane where new virions bud. To release the virus particle from the cell surface of the host cell, NA cuts off the binding between the viral surface protein HA and the cellular sialic acid receptors.

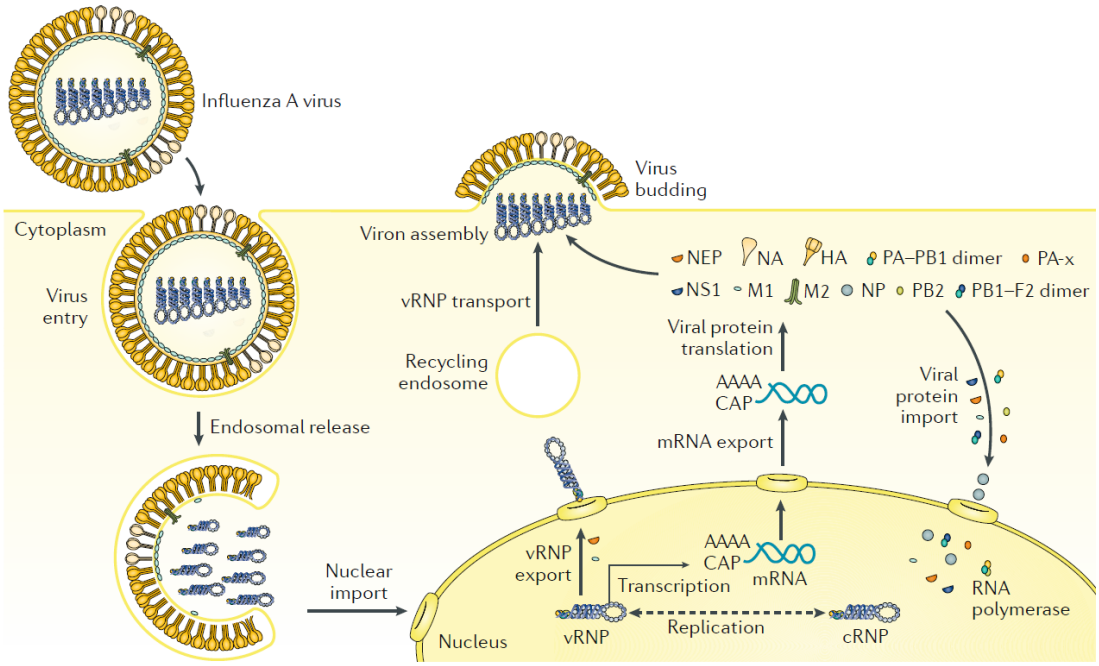

Figure 2: Influenza A virus life cycle.

The Influenza virus enters the cell by endosomal uptake. The vRNPs are released into the cytoplasm and imported in the nucleus. Therefrom the vRNA is transcribed into mRNA and replicated by means of complementary ribonucleoprotein (cRNPs). The viral proteins are synthesized from viral mRNA in the cytoplasm. All viral components assemble at the plasma membrane to virions and new viruses bud. Image is reused with permission from Krammer, F. et al. (2018)[11].

5.3 Viral neuraminidase

The neuraminidase protein (NA) is located in the virus envelope either as a local cluster or as single spikes that are surrounded by HA molecules[19] (*Figure 1*). Around 40-50 NA spikes and 300-400 HA spikes are located in the envelope of a single virion[19,20]. The NA is a homotetramer that contains around 470 amino acids in each of the four peptides. One subunit contains four domains: a short, six amino acids long N-terminal cytoplasmic sequence, a transmembrane domain, a stalk and an enzymatic head domain[21] (*Figure 3*).

NA plays a significant role in the virus release. It catalyses the hydrolysis of the glycosidic linkage that attaches the virus to the host cell[22]. The highly conserved active site is located in the head domain and interacts directly with sialic acids. It cleaves sialic acids that are α2-3-linked and α2-6-linked to galactose.

Removing the sialic acids from the cell surface prevents the nascent virion to bind back to the host cell. However, NA removes not only the sialic acids from the host cell but also from the glycosylated NA and HA in the viral envelope to prevent aggregation of virions[22]. A precise balance between the NA sialidase activity and the HA binding efficiency is necessary for the efficient entry and exit from cells[23]. An excess of NA on the surface of the virion causes a depletion of the sialic acid receptors and the virion cannot efficiently bind and enter the cell. In addition, if an excess of HA is existing, the virions remain bound to the sialic acid receptors and cannot release the cell since NA does not remove the sialic acid receptors efficiently.

Some studies showed that without NA activity, IAV can infect cells and produce virus, but the newly produced virus particles remain aggregated to the host cell and cannot spread to neighbouring cells[22]. Recent publications provide evidence that the function of NA is not only restricted to the virus release and spread. NA plays additionally an essential role in different stages of the infection cycle, like in the virus entry[24,25]. NA activity facilitates the binding and release of virions from the cell surface. Therefore, the virus can move across the cell surface to active sites of the cell and increases the chance of cell entry. The mucus in the respiratory epithelium is rich in sialylated glycoproteins. To enter a cell, NA activity is needed to enhance the movement through the mucus by removing decoy sialic acids from the mucus layer, and to access functional receptors on target cells[26].

NA is the main target for antivirals. As the enzymatic active site of NA is highly conserved, it is a suitable target for drugs. The NA inhibitors oseltamivir (Tamiflu™), zanamivir (Relenza™), peramivir (Rapivab™) and Laninamivir (Inavir™) are currently licensed. However, in 2008, a now circulating H1N1 virus emerged that is resistant to oseltamivir due to a single amino acid change[27].

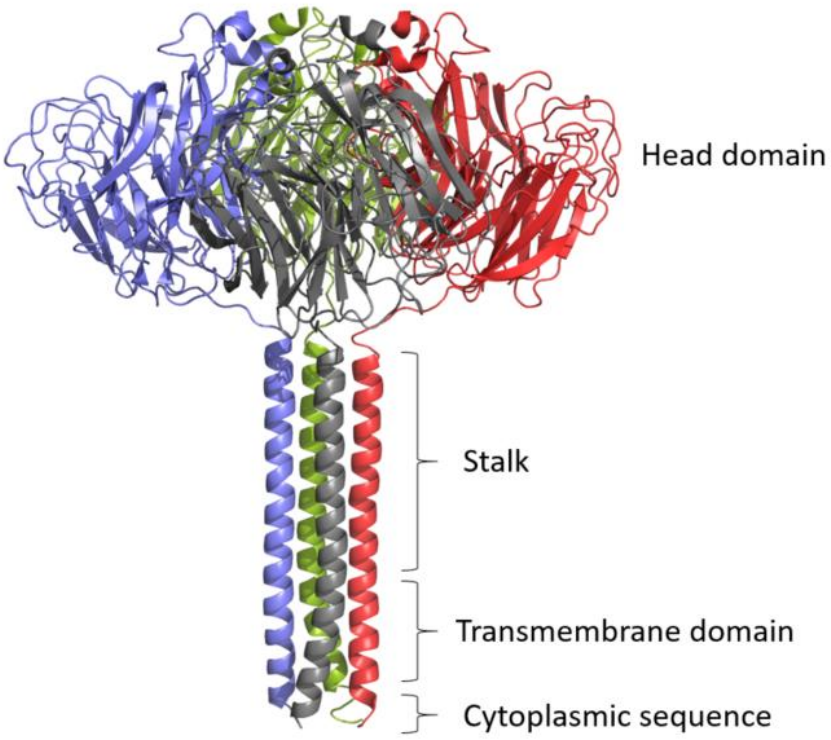

Figure 3: Influenza A neuraminidase. *The Influenza neuraminidase contains of four identical monomers, shown in blue, grey, red and green. Each monomer consists of the four domains: the head, stalk, transmembrane domain and the cytoplasmic tail. The structure was created with Pymol using PDB protein data bank information 6CRD (A/Tern/Australia/G70C/1975 H11N9).*

5.4 Vaccine development

The first successful vaccine was developed in 1796 by Edward Jenner against smallpox[28], which enabled the eradication of smallpox in 1977[29]. This milestone opened the door for new vaccine developments. The discovery of other vaccines like polio in 1950[30] or measles in 1963[31] drastically reduced the incidence of these diseases. To date, vaccines are the most efficient tools to fight infections[32]. Currently, vaccines are against around 35 infectious diseases available[33]. The most common types of vaccinations are live-attenuated, inactivated or split and subunit vaccines.

Live-attenuated vaccines contain altered pathogens that are less virulent than their parental counterparts. They are the most efficacious vaccines used in children as they induce a robust and long-term immune response[34]. The attenuation of the pathogens can be obtained by repeated passaging or by changing their genome sequences. Live-attenuated virus vaccines are available against yellow fever, oral poliovirus, influenza virus, rotavirus, herpes zoster, measles, mumps, rubella and varicella[33]. They do have the risk to revert to a more virulent

form and cause disease. Live-attenuated vaccines are not suitable for immunocompromised people as they could lead to severe infections.

Conversely, non-live vaccines as inactivated, split and subunit vaccines do not contain infectious particles and therefore cannot cause disease. However, the immunogenicity and the duration of protection might be less than for live-attenuated vaccines[35] and adjuvants or several doses are needed[36]. Inactivated vaccines contain whole pathogens that are inactivated. The inactivated pathogens are non-infectious, but their immunogenicity remains active. Thus, they provide a high safety. The inactivation is performed by chemicals, heat or radiation. Inactivated polio vaccine, whole-cell pertussis, rabies and hepatitis A are currently used as inactivated vaccines[35]. Split and subunit vaccines contain only parts of the pathogen, for example, selected proteins. They might have a lower immunogenicity, because they contain only fewer antigens. Split and subunit vaccines are available against hepatitis B, influenza virus, acellular pertussis, human papillomavirus, meningococcal B and herpes zoster[35].

A variety of novel strategies are emerging to facilitate the development of better vaccines that would have high safety and also induce strong immune responses. The emerging trends in vaccine development include vector-based vaccines, nucleic acid-based vaccines or vaccines designed by large-scale recoding of virus genomes. Vector-based vaccines derive from non-pathogenic but infectious viruses that express antigens of a pathogen[35]. In nucleic acid-based vaccines DNA or RNA is injected. The DNA or RNA encodes and expresses antigens from pathogens inducing a robust immune response[37]. The use of large-scale recoding I will discuss in detail in section '*5.6 Large-scale recoding of viruses*'.

In the last decades many devastating viruses like Ebola, severe acute respiratory syndrome (SARS) virus, Middle East respiratory syndrome (MERS) virus or avian influenza virus emerged[38]. Despite an ongoing development of vaccines, we still lack efficient vaccines for many threatening diseases for human and animals. Very recently, the first Ebola vaccine was approved[39]. Efforts are also made in the development of tetravalent dengue vaccines, where several are currently in clinical trials[40,41] and one (Dengvaxia®) with around 60 % efficacy was approved by the Food and Drug Administration (FDA)[42]. However, efficient vaccines against Zika, hepatitis C, the human immunodeficiency virus (HIV) and many other viruses are not available yet[33]. For the development of new vaccines and to improve vaccination efficiencies, more new, innovative concepts are urgently needed.

5.5 Influenza vaccines

To control annual seasonal epidemics, influenza vaccinations with inactivated, live-attenuated or recombinant vaccines are administered[43]. A big challenge in controlling IVA is the constant evolution of the surface proteins NA and HA (antigenic drift). Mutations in these proteins enable reinfection of the same host[44]. Due to the rapid evolution of IAV, available vaccines do not offer protection against the newly arising strains. Drastic changes in the antigenicity of IAV can be also acquired through reassortment of viral segments from diverse strains (antigenic shift). This phenomenon can be associated with devastating pandemics.

To select the vaccine strain for the next season, the World Health Organization collects data across the world and determines the sequences and the antigenicity of the seasonal virus[45]. Seasonal influenza infections are mainly caused by two influenza A subtypes (H1N1 and H3N2) and two influenza B lineages (Victoria and Yamagata). Trivalent and quadrivalent vaccines are available against these predominantly circulating strains. The trivalent vaccine contains both influenza A strains but only one of the influenza B strains. Since 2013, quadrivalent vaccines are available that contain both influenza A and B strains[1,45,46]. The quadrivalent influenza virus vaccine for this season (2019-2020) in the northern hemisphere contains the following components[1]:

- A/Brisbane/02/2018 (H1N1)pdm09-like virus
- A/Kansas/14/2017 (H3N2)-like virus
- B/Colorado/06/2017-like virus (B/Victoria/2/87 lineage)
- B/Phuket/3073/2013-like virus (B/Yamagata/16/88 lineage)

This year's trivalent vaccine contains both influenza A strains and the influenza B Victoria lineage. The in 2009 emerging H1N1 strain that caused a pandemic across the world, is now circulating as a seasonal influenza virus and is part of the influenza vaccine[1].

The most commonly used vaccines are inactivated influenza viruses. They are injected intramuscularly and are licensed from 6 months of age[47]. These virus vaccines are mainly produced in eggs and afterwards inactivated. They are treated to make subvirions or split into the individual components and are partially purified for the enrichment of the surface proteins HA and NA[11]. Adjuvants are often used in combination to increase the immunogenicity of vaccines. This year's season (2019/20) there is for the first time a quadrivalent influenza vaccine available in Europe that was produced in mammalian cells (Flucelvax® Tetra), which is free of egg components[48]. In the USA, this vaccine was already approved for the season 2016/2017.

For children between 2 and 18 years, STIKO (Standing Committee on Vaccination in Germany) recommends as well, a live-attenuated vaccine (FluMist®)[47]. Advantages of this vaccine over

the inactivated vaccine are the needle-free intranasal administration, as well as a more robust induction and durable immune response[47]. This seasonal vaccine is produced similar to the inactivated vaccines in eggs. The donor virus (A/Ann Arbor/6/1960) is cold adapted and shows efficient replication at lower temperatures (25°C to 33°C) but is temperature sensitive and attenuated. The cold-adapted phenotype is based on five mutations in three different segments[49]. Its replication at higher temperatures from 37°C to 39°C, as it is present in the lower respiratory tract, is limited[50]. Thus, replication is restricted to the colder upper respiratory tract where the virus does not cause disease. To obtain the vaccine strain, reassorted viruses are selected to contain the six internal RNA fragments of the cold-adapted virus and the NA and HA segment of the currently circulating strain[51].

In 2013, the FDA approved Flublok®, a third production technology for manufacturing recombinant HA subunit influenza vaccines[52,53]. This method uses recombinant baculoviruses to produce HAs of the recommended influenza virus vaccine strains in insect cells. The main advantage of this technology is that there is no need for eggs and in a vaccine shortage, the production can be scaled up quickly[53]. Additionally, the recombinant baculoviruses produce similar high titres of different HA proteins.

Depending if vaccines match the circulating viruses, the effectiveness of influenza vaccination in the last 15 years ranged from only 10 % to 60 %[45]. In Germany, the vaccination efficiency in the last two seasons (2017/18 and 2018/19) was with 15 % and 21 % very low. The main reason for the low efficiency in 2017/18 was that the Influenza B Yamagata stain, which was only included in the quadrivalent and not in the trivalent vaccine, was responsible for a majority of infections[54]. In the season 2018/19 the majority of cases was caused by Influenza A strains. The efficiency against the A(H1N1)pdm09 was relatively high (61 %), but the A(H3N2) vaccination strain had no effect[55].

Currently, attempts are made to develop universal vaccines that provide better cross-protection against different strains. Different technologies, including vector-based vaccines, virus-like particles, recombinant vaccines and broadly protective antibodies, are used to develop such vaccines[56,57]. These vaccines target the stalk or other highly conserved epitopes from different influenza protein[45]. The goal is that these new vaccines stimulate T-cells and induce a broad antibody response against diverse influenza strains.

5.6 Large-scale recoding of viruses

In 2002, an infectious poliovirus of 7.5 kb was the first virus generated completely synthetically[58]. Recent progress and decreased costs of de-novo synthesis of DNA enable large-scale recoding of genes, a systematic introduction of hundreds of synonymous mutations in the DNA sequence[59]. The aim of recoding is to change the dinucleotide content, the codon

usage or the codon pair bias in the recoded gene. Recoded genes produce less protein due to a series of not completely understood processes during the gene expression. The generated recoded viruses are attenuated and are therefore promising live-attenuated vaccine candidates[60]. In contrast, traditionally, live-attenuated virus vaccines were generated by serial passaging, which is time-consuming. Additionally, the outcome is unpredictable as passaging often introduces many mutations, whereas only a few might be responsible for the attenuation[34].

Importantly, the recoding of large-scale recoded genes, affects only the DNA or RNA sequence, leaving the amino acid sequence unchanged. Therefore, the recoded viruses are antigenically identical with their original non-recoded form and induce the same immune response as their pathogenic parents[61]. Large-scale recoding of virus genomes offers rapid design and production of live-attenuated vaccines: computer algorithms perform the recoding, the gene is chemically synthesized and subsequently the attenuated virus can be recovered. The entire procedure of the generation of a new live-attenuated vaccine can be completed within a few days or weeks and is universal applicable to a large variety of viruses[62]. This fast vaccine development and production are crucial for newly emerging virus infections or constantly evolving viruses like IAV. Currently, used live-attenuated vaccines for IAV contain the backbone of an attenuated cold-adapted donor virus and only the fast evolving surface glycoproteins NA and HA originate from the actual circulating strain. A fully recoded virus, for example by codon pair deoptimization, is antigenically identical to the target strain. Therefore, an immune response against all virus proteins and not only against glycoproteins NA and HA can develop. This increases the cross-protection against other strains[62].

The basic principles underlying attenuation due to the recoding remains largely unknown. Understanding of these principles is necessary, before live-attenuated vaccines could be produced in an automated fashion. One important aspect for future vaccine candidates is the genetic stability of the modified virus. The vaccine backbone currently used (FluMist®), contains five amino acid changes that are responsible for the temperature sensitivity and therefore for the attenuation[49]. Hence, it is theoretically possible that the live-attenuated vaccine reverts back to higher virulence. Through the introduction of a large number of mutations, viruses seem to be genetically stable[63,64]. It is hypothesized that an additive effect of numerous nucleotide changes is responsible for the overall attenuation of viruses[65]. This makes the attenuation adjustable[66]. As we do not know the effect of each dinucleotide, codon or codon pair change, we have to empirically test the efficiency of virus attenuation and the possible reversion of the virus back to a virulent virus.

5.7 Codon optimization and deoptimization

Codon (de)optimization is a method to recode genes and is used to adjust the protein level of a gene due to changes in the DNA sequence[67]. Recoding changes the DNA sequence of the recoded gene, but does not affect the amino acid sequence of the encoded protein. Codon optimization is typically used to improve the level of protein production[68]. Conversely, codon deoptimization is used for virus attenuation, for example, it was used to attenuate IAV[69] or poliovirus[70].

The genetic code defines how information encoded in genetic material (RNA or DNA) is translated into proteins. It specifies how a 4-letter code of DNA is translated into the 21-letter code of proteins. The genetic code consists of 64 triplets of nucleotides, which are called codons. Each codon either encodes one of the 21 amino acids or a stop signal. All amino acids, except for methionine, tryptophan and selenocysteine are encoded by several codons. Different codons that encode the same amino acid are called synonymous codons. However, synonymous codons do not encode amino acids with equal frequencies, a phenomenon known as codon (usage) bias. For instance, valine is encoded by four different codons (GUG, GUU, GUC, and GUA). However, in humans, the GUG codon is used more than three times as often as the GUA codon (*Table 1*).

Codon optimization involves recoding of a gene of interest with the most frequently used codon of each amino acid. In a pure sense, it means that codon optimized genes are encoded only with a set of 21 codons. However typically, commercial gene optimization tools also reduce known complexities that interfere with DNA manufacturing and protein expression, such as repeats, hairpins, and extreme GC content. Codon optimization has been widely used to optimize protein production and for diverse applications in synthetic biology[67]. Conversely, during codon deoptimization, genes are recoded using the least frequently used codons.

Codon choice affects protein synthesis in its translational speed and accuracy. This is caused by a series of factors like mRNA stability[71,72], efficiency of the translation process[73] or protein folding[74] . Codon choice is therefore an essential factor regarding the efficiency of successful translation. Recent literature also stresses the importance of codon optimality, which describes the ability of a given codon to affect mRNA stability in a translation-dependent manner[71]. There is a notable impact of codon choice and the protein expression level, whereas codon bias is more distinct in highly expressed genes. mRNAs enriched in non-optimal codons show lower RNA levels, are unstable and have shorter poly(A)-tails. Additionally, regulatory elements located in the 3' untranslated region can affect mRNA stability. Non-optimal codons have lower tRNA levels and lower ratios of charged tRNAs[75]. It was shown in yeast that there is a correlation between high tRNA abundance fast traverse of ribosomes on codons[76,77]. The tRNA

abundance is not the only factor determining elongation rates. The sequence context of codons is an important determinant for the extent of codon effects like mRNA stability[78]. Adjacent codons – codon pairs – seem to play also an important role in the translation speed[78].

As there is not always a clear correlation between mRNA levels and protein expression, post-transcriptional regulations are important to regulate the amount of protein expression. Another factor influencing ribosome elongation dynamics is wobble base pairing between the mRNA codon and the tRNA anticodon. Watson-Crick base pairing is more favourable[79] and ribosome move faster off codons compared to wobble-base pairing[77].

Overall codon (de)optimization is an efficient method to regulate protein expression and is applicable for any gene, although the whole mechanism behind is not fully understood yet.

Table 1**:** *Amino acids with their corresponding codons and the frequency of usage of these codons[80]. Note that the dinucleotides CG and UA (depicted in bold) occur in the least frequent codons.*

Amino acid	Codon	Usage [%]	Amino acid	Codon	Usage [%]	Amino acid	Codon	Usage [%]
Alanine	G**CG**	8.9	**Glycine**	GGU	17.1	**Serine**	U**CG**	4.7
	GCA	25.2		GGG	24.3		AGU	16.2
	GCU	27.8		GGA	27.1		UCA	16.4
	GCC	38.1		GGC	31.5		UCU	19.6
Arginine	**CG**U	8.1	**Isoleucine**	A**UA**	18.8		UCC	20.1
	CGA	11.4		AUU	38.1		AGC	22.9
	CGC	15.7		AUC	43.1	**Stop codons**	**UA**G	21.9
	CGG	19.1	**Leucine**	C**UA**	7.6		**UA**A	27.8
	AGG	21.8		U**UA**	9.0		UGA	50.3
	AGA	23.8		UUG	13.7	**Threonine**	A**CG**	10.4
Asparagine	AAC	49.8		CUU	14.4		ACU	26.3
	AAU	50.2		CUC	18.2		ACA	30.5
Aspartic acid	GAU	49.7		CUG	37.0		ACC	32.8
	GAC	50.3	**Lysine**	AAA	46.5	**Tryptophan**	UGG	100.0
Cysteine	UGU	49.0		AAG	53.5	**Tyrosine**	**UA**U	47.3
	UGC	51.0	**Methionine**	AUG	100.0		**UA**C	52.7
Glutamic acid	GAA	45.9	**Phenyl-alanine**	UUU	49.4	**Valine**	G**UA**	13.1
	GAG	54.1		UUC	50.6		GUU	19.9
Glutamine	CAA	28.5	**Proline**	C**CG**	9.8		GUC	22.9
	CAG	71.5		CCA	29.7		GUG	44.1
Histidine	CAU	44.8		CCC	30.2			
	CAC	55.2		CCU	30.3			

5.8 Codon pair deoptimization

Aside from codon deoptimization, codon pair deoptimization is a new powerful method for recoding genes. It was first described in 2008 and is mainly applied to attenuate viruses. Codon pair deoptimization was successfully used to recode several viruses like Influenza A virus[60], Zika[81] or poliovirus[65] and can lead to dramatic attenuation[65]. As codon pair deoptimization affects the virulence of the viruses, it is now used in the development of live-attenuated vaccines for humans and animals[62]. Currently, the first codon pair deoptimized influenza vaccine (CodaVax-H1N1) is in clinical trials[82].

Numerous studies were performed to analyse codon optimality, but only a small number of studies focused on the understanding of codon pair bias[66]. Codon pair deoptimization involves changing the positions of existing synonymous codons within the recoded region without changing the amino acid sequence or the codon usage (*Figure 4*). The usage of codon pairs is not random. There are $4^3 \times 4^3$ (4,096) codon pair combinations, excluding stop codons still 3,721 possibilities. Independent on the frequency of certain codons, some of these codon pairs occur significantly more or less frequent than expected in open reading frames. A particular bias in codon pair usage was found in every investigated species and is highly different between species[83,84]. Closely related species, such as mouse and human, have similar codon pair preferences, whereas more distantly related species, for example mosquitoes and humans, show a completely different codon bias[60,85]. To evaluate the degree of deoptimization, a codon pair score (CPS) was developed.

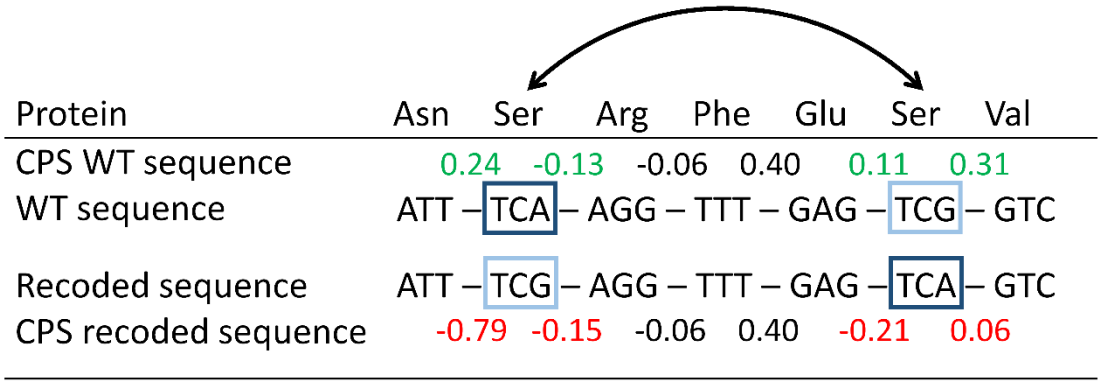

Figure 4: *Codon pair deoptimization of a DNA sequence. Shown is the WT sequence and a recoded, codon pair-deoptimized sequence, both with their corresponding codon pair scores (CPS) and amino acid sequence. Recoding was performed by changing the positions of two synonymous codons, which encode serine (Ser). The recoding affected the CPS of the recoded sequences. The recoded sequence contains lower CPS (red) than the WT sequence (green).*

The CPS is independent of the codon and amino acid usage and can be calculated for each individual codon pair[65]. It is the proportion of the observed frequency (f) of the codon pair AB relative to the expected frequency based on the number of occurrences of the codons A and B and the for these two codons encoding amino acids X and Y.

$$\mathrm{CPS}_{AB} = \ln\left(\frac{\mathrm{f(AB)}}{\frac{\mathrm{f(A)\ x\ f(B)}}{\mathrm{f(X)x\ f(Y)}}\ \mathrm{x\ f(XY)}}\right)$$

Overrepresented codon pairs have positive CPS, underrepresented codon pairs have negative CPS value. To calculate the CPS for the entire recoded gene or gene fragment with k codon pairs, an average CPS of each single codon pair (i) is calculated[65].

$$\text{average CPS} = \sum_{i=1}^{k} \frac{\mathrm{CPS}_i}{k-1}$$

For a maximal codon pair deoptimization, thousands of potential swaps of synonymous codons are evaluated. A computer algorithm finds the optimal (or near optimal) arrangement of codons to obtain the lowest overall CPS for the selected gene. Consequently, hundreds of nucleotide changes are introduced in the recoded sequence without changing the amino acid sequence.

The exact mechanism of attenuation by codon pair deoptimization remains unknown. Multiple factors including translation[65], protein folding[86], mRNA stability[62,87], RNA secondary structures[87] or dinucleotide content[88] might be to be involved. Originally, it was proposed that underrepresented codon pairs create disadvantageous conditions for efficient protein production. During translation, charged tRNAs bind initially to the aminoacyl site (A-site) of the ribosome, move to the peptidyl site (P-site) where they hold the peptide chain and are finally released from the exit site (E-site). New peptide bonds are formed between the amino acid of the charged tRNA in the A-site and the peptide held by the P-site tRNA. The peptide is transferred from the P-site tRNA to the A-side tRNA and afterwards the tRNAs are translocated from the A to the P site and from the P to the E-site. It is assumed that different geometric structures of tRNAs lead to altered interactions between adjacent tRNAs[89]. The translation rate and accuracy depends on the optimal fit of the tRNAs in the ribosome A- and P-site. Of particular importance, seem to be thereby the identity of the A-side codon and the third nucleotide of the P-site codon[89].

As a consequence suboptimal translation of codon pair deoptimized mRNA might lead to mistranslation or stalled translation which often results in premature termination[87]. This could be linked to the observed reduced mRNA levels in codon pair deoptimized gens[62,87]. Increased mRNA degradation is potentially coupled with the no-go decay of stalled ribosomes during translation[90]. Thereby stalled ribosomes or damaged mRNAs are recognized by a control system and an endonuclease cleaves the mRNA upstream of the stalled ribosome. Subsequently the mRNA fragments are degraded, at the 3' end from the exosome and at the 5' end from the Xrn1 exonuclease[91].

Another groundbreaking study found 17 codon pairs that inhibit translation efficiency in yeast *Saccharomyces cerevisiae*[78]. Therefore, it might be possible that codon pairs with a strong inhibitory effect on translation also exist in vertebrates, and that not the additive effect of many suboptimal codon pairs lead to a hampered translation but rather the introduction of a few codon pairs.

A second theory suggests that not the underrepresented codon pairs are responsible for the attenuation of viruses but the change in the CpG and UpA dinucleotide content that occurs during codon pair deoptimization (see section '*5.9 Dinucleotide preferences among different species*')[88].

5.9 Dinucleotide preferences among different species

The four nucleotides adenine, guanine, cytosine and thymine form 16 different dinucleotide combinations in the DNA. It was discovered that different species have different dinucleotide compositions[92]. CpG and TpA dinucleotides are highly repressed in vertebrates. Synonymous codons containing CpG or TpA dinucleotides occur less frequently than expected, as can been seen for example for the amino acid alanine (*Table 1*). The dinucleotide bias is similar for mammals but very different for other species as for example for insects. Non-vertebrates do not show CpG suppression, however TpA dinucleotides are underrepresented[85].

Additionally, CpG dinucleotides located at the codon boundary of a codon pair (NNC-GNN) are suppressed and are among the most underrepresented codon pairs in vertebrates[85] (*Figure 5*). 97 % of the 100 rarest codon pairs contain a CpG dinucleotide at their codon boundary[93]. Similarly but to a lesser degree, TpA dinucleotides are repressed (*Figure 5*). Therefore, it is inevitable during codon pair deoptimization that CpG (and TpA) dinucleotides are introduced at the codon boundaries in recoded genes. As a result, it was hypothesized that the CpG (and TpA) dinucleotides are the cause of virus attenuation.

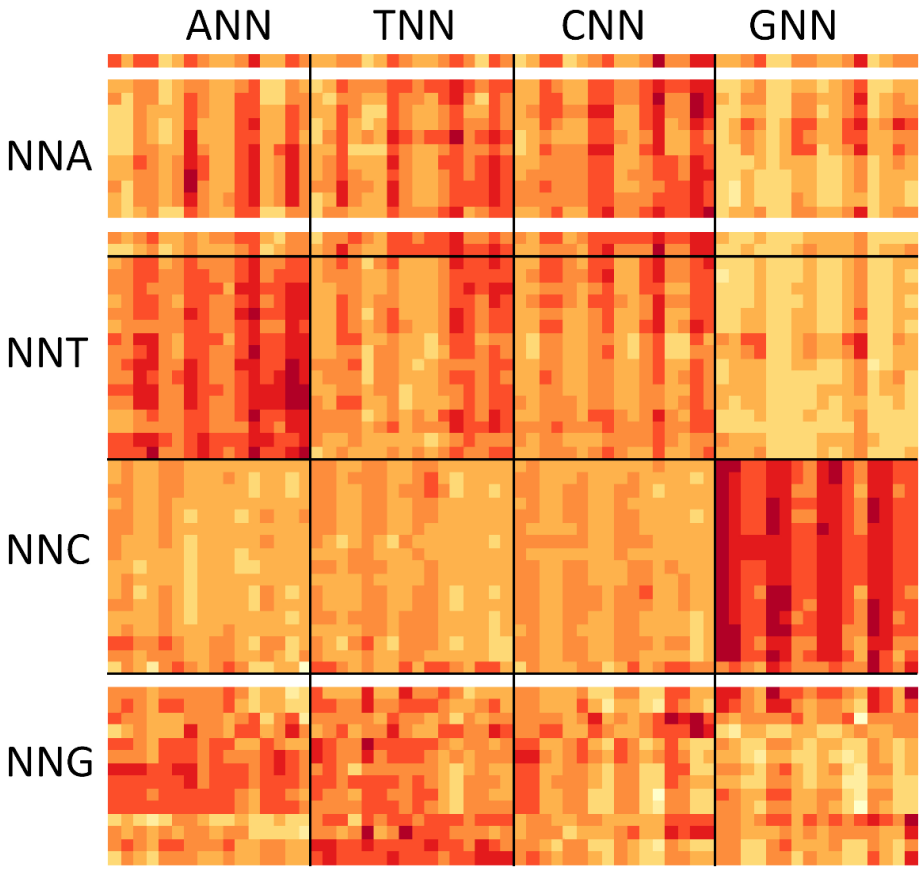

Figure 5: *Heat map of codon pair scores (CPS) of all possible 4,096 codon pair combinations. Low CPS are shown in red, high scores in yellow. Labels on the left and on top give the nucleotide at the last or first position of the codon respectively, of all 16-pictured codon pairs. Codon pairs of the type NNC-GNN have overall the lowest scores (dark red). White lines in the heat map result from codon pairs containing a stop codon in the first codon.*

The reason for suppression of CpG dinucleotides in vertebrates could be due to a frequent methylation of cytosine in CpG dinucleotides. 5-methylcytosine tends to mutate through a deamination to thymine, which increases the mutation probability from CpG to TpG[94-96]. It was observed that most vertebrate RNA viruses also show a reduced number of CpG dinucleotides in their genomes, even though methylation of CpG does not occur in RNA.

One hypothesis is that CpG-rich sequences might act as motifs that stimulate the innate immune response. The zinc-finger antiviral protein (ZAP) was recently discovered to bind RNAs rich in CpG dinucleotides[97,98] and to target them for degradation[99]. To escape the recognition of the host, viruses might have adapted their dinucleotide content[100].

5.10 Ethical concerns

In 2002, the first de-novo synthesised virus was produced[58]. Nowadays, also bacteria[101] and whole eukaryotic chromosomes can be chemically synthesized[102]. Even though the synthesis of poliovirus had no potential to threat the global human health, the public response was strong and the work was called dangerous and irresponsible[103]. The fast development of the *de novo* synthesis of DNA provides a potential risk of this technology to misuse such as bioterrorism. However, the responsibility of a scientist is to consider between risk and potential use of new technologies. Today, we see the diverse benefits of the chemical synthesis of DNA in medical applications, such as in the development of new vaccine strategies. One of these strategies is the codon pair deoptimization of viral genes, applied in this work.

The best way to control a disease is to eradicate it, but since we know the genome sequences, virus genomes can be made synthetically at any time in a laboratory. This includes not only poliovirus but also smallpox. To protect us from misuse, new strategies need to be developed[103].

5.11 Project outline

Large-scale recoding is a promising tool to develop rapidly efficient vaccines for diverse viruses. Some codon pair deoptimized viruses have been shown to serve as highly protective vaccines in animals like mice and ferrets[62,87]. These days the first codon pair deoptimized influenza viruses in humans are tested in clinical trials (Coda-Vax)[104]. Although the understanding of the mechanistic background of attenuation is essential for the use as safe and efficient vaccine, the analysis of the attenuation mechanism has not been solved. In this study, I analyse the underlying mechanism of virus attenuation by codon pair deoptimization.

It is discussed controversially if virus attenuation through codon pair deoptimization is based on disfavoured codon pairs or if it is an artefact of increased CpG (or UpA) dinucleotide content[88,105,106]. Even though it is from great importance, the attenuation mechanism is widely unknown. Additionally, different recoding strategies have never been compared to identify the most efficient virus attenuation method.

To shed light onto this lack of knowledge, I aimed to answer these open questions in this project with three specific aims:

1) Are underrepresented codon pairs or the CpG dinucleotides major determinants of codon pair deoptimization? (Part I)
2) What factors affect transcription and translation of recoded genes, and thus contribute to the virus attenuation? (Part I)

3) Is codon deoptimization or codon pair deoptimization a more efficient virus attenuation method? (Part II)

To answer these questions, eight different influenza A virus NA genes were designed and synthesized. The genes contained different CpG content or different degree of codon pair deoptimization, but encoded identical proteins. This enabled us to disentangle the effects of these two possible factors on attenuation of viruses (Aim 1). I rescued the viruses and compared their properties *in vitro* and *in vivo*. For further analysis of the mechanism that leads to reduced protein expression and virus attenuation, I detected transcription and translation efficiencies as well as RNA stability (Aim 2). To compare the recoding strategies, I recoded the NA gene using different optimization and deoptimization methods (Aim 3).

6 Material and methods

6.1 Material

6.1.1 General Equipment

Device	Manufacturer
AB StepOnePlus Real-time PCR system	Invitrogen Life Technologies, Grand Island
Axiovert S 100 fluorescence microscope	Carl Zeiss MicroImaging GmbH, Jena
Bacterial incubator, 07-26860	Binder, Turtlingen
Bacterial incubator/shaker Innova 44	New Brunswick Scientific, New Jersey
CASY cell counter	Roche, Mannheim
Cell incubators Excella ECO-1	New Brunswick Scientific, New jersey
Centrifuge 5424, rotor FA-45-24-11	Eppendorf, Hamburg
Centrifuge 5804R, rotors A-4-44 and F45-30-11	Eppendorf, Hamburg
Chemiluminescence imaging system - Fusion SL	Peqlab, Erlangen
Electrophoresis power supply Power Source	VWR, Darmstadt
Galaxy mini centrifuge	VWR, Darmstadt
Gel electrophoresis chamber Mini	VWR, Darmstadt
HiSeq4000	Illumina, San Diego
Low binding mirco tube	Sarstedt, Nümbrecht
LTQ Orbitrap Elite mass spectrometer	Thermo Fisher, Darmstadt
Microscope AE20	Motic, Wetzlar
Nanodrop 1000	Peqlab, Erlangen
Potima XPN-80 Ultracentrifuge	Beckman Coulter, Brae
Professional TRIO Thermocycler	Analytik Jena, Jena
Semidry membrane blotting machine	Peqlab, Erlangen
StepOnePlus Real-Time PCR System	Thermo Fisher, Darmstadt
Synergy HT plate reader (ELISA)	BioTek, Vermont
Thermocycler T-Gradient	Biometra, Göttingen

6.1.2 Consumables

Name	Manufacturer
Biosphere plus SafeSeal Micro Tube, RNase free	Sarstedt, Nümbrecht
Cell culture consumes (plates, dishes, pipettes)	Sarstedt, Nümbrecht
Cell Strainers	BD Falcon, San Jose
Cryotubes 1.8 ml	Nunc, Kamstrupvej
Eppendorf tubes 1.5 and 2 ml	Sarstedt, Nümbrecht
Petri dishes for bacterial culture	Sarstedt, Nümbrecht
Roti-PVDF, pore size 0.45 µm	Carl Roth, Karlsruhe
Sterile syringe filters PVDF 0,45 µm	VWR, Darmstadt
Whatmann blotting paper	GE Healthcare, Freiburg

6.1.3 Reagents

Name	Manufacturer
30 % acrylamide and bisacrylamide	Carl Roth, Karlsruhe
35 % Bovine serum albumin (BSA)	Sigma-Aldrich, St. Louis
5-Bromouridine (BrU) [850187]	Sigma-Aldrich, St. Louis
Acetic acid	VWR, Darmstadt
Adenosine triphosphate (ATP)	Thermo Fisher, Darmstadt
Ammoniumchlorid (NH_4Cl)	Carl Roth, Karlsruhe
Ammoniumpersulfate (APS)	Merck, Darmstadt
Anti-FLAG M2 magnetic beads	Sigma-Aldrich, St. Louis
Avicel	FMC BioPolymer, Sandvika
Chloroacetamide	Sigma-Aldrich, St. Louis
Chloroform	Sigma-Aldrich, St. Louis
Complete, protease inhibitor cocktail	Roche, Mannheim
Crystal Violet	Carl Roth, Karlsruhe
Cyclohexaminde	Sigma-Aldrich, St. Louis
DEAE-Dextran	Sigma-Aldrich, St. Louis
Dimethyl sulfoxide (DMSO)	Merck, Darmstadt
Dithiothreitol (DTT)	Carl Roth, Karlsruhe
dNTP Mix (10 mM total)	Bioline, Luckenwalde
Ethanol	VWR, Darmstadt
Ethidium bromide 1 %	Carl Roth, Karlsruhe

Ethylenediaminetetraacetic acid (EDTA)	Applichem, Darmstadt
Fetal bovine serum (FBS)	Biochrom AG, Berlin
Flavopiridol	MedChem Express, Sollentuna
Formaldehyde 37 %, stabilized with 10 % methanol	Applichem, Darmstadt
Glycerol	Applichem, Darmstadt
Glycine	Applichem, Darmstadt
Glycoblue	Thermo Fisher, Darmstadt
Hematoxylin	Sigma-Aldrich, St. Louis
Isopropyl alcohol (2-propanol)	Applichem, Darmstadt
L-glutamine	PAN-Biotech GmbH, Aidenbach
Methanol	VWR, Darmstadt
Neufuchsin	VWR, Darmstadt
NP-40	Sigma-Aldrich, St. Louis
Polyethylene glycol 8000 (PEG-8000)	Sigma-Aldrich, St. Louis
Polyethylenimine, linear (PEI)	Sigma-Aldrich, St. Louis
Potassium hydrogen carbonate ($KHCO_3$)	Carl Roth, Karlsruhe
Protein G Dynabeads [10004D]	Thermo Fisher, Darmstadt
Random hexamers	Thermo Fisher, Darmstadt
RLT buffer	Qiagen, Hilden
RNase free PBS	Thermo Fisher Scientific, Darmstadt
SDS (sodium dodecyl sulphate)	Sigma-Aldrich, St Louis
Sodium chloride (NaCl)	Carl Roth, Karlsruhe
Sodium hydrogen carbonate ($NaHCO_3$)	Carl Roth, Karlsruhe
Sucrose	Sigma-Aldrich, St. Louis
SYBR Gold	Thermo Fisher Scientific, Darmstadt
TEMED	Carl Roth, Karlsruhe
Tris(hydroxymethyl)-aminomethane (Tris)	Carl Roth, Karlsruhe
Triton X-100	Thermo Fisher Scientific, Darmstadt
TRIzol LS Reagent	Thermo Fisher Scientific, Darmstadt
Trypsin	Biochrom AG, Berlin
Trypsin (TPCK treated)	Sigma-Aldrich, St. Louis
Tween-20	Carl Roth, Karlsruhe
Uridine	Sigma-Aldrich, St. Louis
Water, ultrapure, DEPC treated	Carl Roth, Karlsruhe
Xylene	Sigma-Aldrich, St. Louis
β-mercaptoethanol	Serva, Heidelberg

6.1.4 Buffers and solutions

Name	Composition
1 % Agarose gel	1 % w/v Agarose, TAE buffer, 0.5 µg/ ml Ethidium bromide
17 % urea gel	Rotiphorese Sequencing Gel System (40.8 ml concentrate, 7.2 ml diluter, 6 ml buffer concentrate, 6 ml H_2O), 10 % APS, 30 µl TEMED
ACK buffer	150 mM NH_4Cl, 0.1 mM $KHCO_3$ and 0.1 mM Na_2EDTA, pH 7.2
BrU-IP buffer	RNase free PBS, 0.05 % Tween
Coomassie solution	45 % (v/v) ethanol, 10 % (v/v) acetic acid, 0.25 % (w/v) Coomassie Brilliant Blue G-250
Laemmli, SDS loading buffer	0.1% ß-Mercaptoethanol, 0.0005 % Bromophenol blue, 10 % Glycerol, 2 % SDS, 63 mM Tris-HCl, pH 6.8
LB agar (1 L)	10 g Tryptone, 5 g Yeast Extract, 15 g Agar, 10 g NaCl
LB medium (1 L)	10 g Tryptone, 5 g Yeast Extract, 10 g NaCl
PBS	2 mM KH_2PO_4, 10 mM Na_2HPO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4
PBS-T	PBS with 0.1 % Tween-20
RIPA I	20 mM Tris, pH 7.5; 150 mM NaCl; 1 % Nonidet P-40; 0.5 % sodium deoxycholate; 0.1 % SDS, protease inhibitors
RIPA II	RIPA I + 1 mM EDTA
SDS - polyacrylamide gel electrophoresis (PAGE) running buffer	25 mM Tris HCl, 190 mM Glycine, 0,1 % SDS
SDS-Separating-gel	10 % (w/v) acrylamide/bisacrylamide stock solution, 0.1 % (w/v) SDS, 375 mM Tris·HCl (pH 8.8), 0.05 % (w/v) APS, 0.1 % (v/v) TEMED
SDS-Stacking-gel	5 % (w/v) acrylamide/bisacrylamide, 0.1 % SDS, 125 mM Tris·HCl (pH 6.8), 0.075 % (w/v) APS, 0.15 % (v/v) TEMED
SOC medium (1 L)	20 g Tryptone, 5 g Yeast Extract, 0.6 g NaCl, 0.2 g KCl, 20 mM, Glucose, pH to 7.0
TAE buffer	40 mM Tris, 1 mM $Na_2EDTA{\cdot}2H2O$, 20 mM Acetic acid 99 %, pH 8.0
TNE buffer	10 mM Tris, 100 mM NaCl und 1 mM EDTA, pH 7.4
Towbin WB transfer buffer	1 L: 25 mM Tris = 3.03 g, 192 mM Glycine = 14.4 g, 20 % methanol = 200 ml, 0.1 % SDS = 1 g
Western blot blocking buffer	3 % skimmed-milk powder in PBS-T

6.1.5 Cell culture media

Name	Manufacturer/content
2X Eagle's Minimum Essential Medium (EMEM)	Lonza, Basel
Complete RPMI (cRPMI)	RPMI 1640, 10 % FBS, 2 mM L-glutamine, penicillin (100 units/mL) /streptomycin (100 µg/ml)
Dulbecco's Modified Eagle Medium (DMEM)	PAN-Biotech GmbH, Aidenbach
Growth medium	DMEM, 10 % (v/v) FBS, penicillin (100 units/mL) /streptomycin (100 µg/ml)
Opti-MEM	Thermo Fisher Scientific, Darmstadt
Plaque assay overlay medium	1X EMEM, 0.6 % avicel, 0.2 % BSA, 0.1 % FBS, penicillin (100 units/mL) /streptomycin (100 µg/ml), 2 µg/ml TPCK-Trypsin, 0.01 % DEAE-Dextran, 0.05 % $NaHCO_3$
RPMI 1640	PAN-Biotech GmbH, Aidenbach
Virus infection medium	DMEM, 0.1 % FBS, penicillin (100 units/mL)/ streptomycin (100 µg/ml), 2 mM Na-pyruvat, 0.2 % BSA and 2 µg/ ml TPCK-Trypsin

6.1.6 Antibiotics

Name	Manufacturer
Ampicillin (100 µg/ml)	Carl Roth, Karlsruhe
Hygromycin B solution	Carl Roth, Karlsruhe
Penicillin G	Carl Roth, Karlsruhe
Streptomycin	Carl Roth, Karlsruhe

6.1.7 Enzymes and markers

Name	Manufacturer
Alkaline Phosphatase CIP	New England Biolabs, Ipswich
Ambion RNase I [100 U/µl]	Thermo Scientific, Darmstadt
AspN	Sigma-Aldrich, St Louis
BamHI	New England Biolabs, Ipswich
Benzonase	Sigma-Aldrich, St Louis
BglII	New England Biolabs, Ipswich
Chymotrypsin	Sigma-Aldrich, St Louis
DpnI	New England Biolabs, Ipswich
Elastase	Sigma-Aldrich, St Louis
GeneRuler 1 kb Plus DNA Ladder	Thermo Scientific, Darmstadt
HOT FIREPol DNA Polymerase	Solis Biodyne, Tartu
MluI	New England Biolabs, Ipswich
M-MLV reverse transcriptase	Promega, Mannheim
PageRuler Prestained plus marker	Thermo Scientific, Darmstadt
Phusion High-Fidelity DNA Polymerase	New England Biolabs, Ipswich
RNasin Ribonuclease Inhibitors	Promega, Mannheim
SalI-HF	New England Biolabs, Ipswich
ScaI	New England Biolabs, Ipswich
SepI	New England Biolabs, Ipswich
Superase-Inhibor	Thermo Scientific, Darmstadt
T4 DNA ligase	New England Biolabs, Ipswich
T4 Polynucleotide Kinase (PNK)	Thermo Scientific, Darmstadt
T4 RNA ligase 1	Thermo Scientific, Darmstadt
Taq DNA-Polymerase	PeqLab, Erlangen
Trypsin	Thermo Scientific, Darmstadt
Turbo DNase	Thermo Scientific, Darmstadt

6.1.8 Antibodies

Name	Manufacturer
Alkaline phosphatase-conjugated rabbit anti-goat antibody	Vector Laboratories, Burlingame
Goat anti-Influenza A H1N1 antibody [5315-0064]	BIO RAD, Marnes-la-Coquette
Horseradish peroxidase goat anti-rabbit antibody	Cell Signaling, Boston
Horseradish peroxidase-conjugated goat anti-mice IgG antibody	Sigma-Aldrich, St Louis
Mouse anti-BrdU antibody [555627]	BD Pharmingen, San Jose
Mouse anti-NA antibody [GT288]	GeneTex, Irvine
Mouse anti-NP antibody [GT778]	GeneTex, Irvine
Rabbit anti-beta-actin antibody	Cell Signaling, Boston
Rabbit anti-FLAG antibody	Sigma-Aldrich, St Louis
Rabbit anti-GFP polyclonal antibody	Cell Signaling, Boston
Rabbit anti-NS1 antibody	Schierhorn et al. (2017) [107]

6.1.9 Kits for Molecular Biology

Name	Manufacturer
AmbiClean PCR/Gel Purification Kit	Vivantis, USA
ECL Prime Western Blotting System	GE Healthcare, Solingen
Hi Yield Gel/PCR DNA Fragments Extraction Kit	SLG, Gauting
MicroSpin S-400 HR columns	GE Healthcare, Solingen
NEB Monarch Gel	New England Biolabs, Ipswich
Plasmid Midi kit	Qiagen, Hilden
PureLink Pro 96 total RNA Purification Kit	Thermo Scientific, Darmstadt
Ready-Set-Go Elisa Kits	eBioscience, San Diego
RiboZero Kit [MRZH11124]	Illumina, San Diego
RNeasy Plus Mini Kit	Qiagen, Hilden
Roti-Quant Universal Kit (BSA assay)	Carl Roth, Karlsruhe
Roti-Quanti Universal	Carl Roth, Karlsruhe
RTP DNA/RNA Virus Mini Kit	Stratec Molecular GmbH, Berlin
TruSeq Stranded mRNA Kit	Illumina, San Diego

6.1.10 Cells and animals

Name	Reference
A549 (human adenocarcinomic alveolar basal epithelial cells)	ATCC CCL-185
BALB/cJRj mice	Janvier, France
Escherichia coli Top10	Invitrogen, Carlsbad
HEK-293T (Human epithelial kidney cell line, SV-40 T-antigen)	ATCC CRL-3216
MDCK II (Madin-Darby canine kidney cells)	ATCC CCL- 2936

6.1.11 Primers

6.1.11.1 Primers for cloning and sequencing

Name	Sequence 5'-3'
DK132	AAACCACCGCTAATTCAAAGCAACCG
DK133	CACGTTTTGTGTCATTGGGGAAACC
MS307	TATAGAATCAGAGTAGCAGATGC
MS310	TATGAAGAGTGCATGTAAGAGC
NG03	GAATATGCAACCAAGGCAGC
NG04	TTGCGTCCTCCTCAGGTAG
NG19	ATATATAGATCTGCCACCATGAATCCAAACCAGAAAAT
NG20	ATATATGTCGACCTACTTGTCAATGGTGAACG
NG27	ATATACGCGTGCCACCATGGATTACAAGGATGACGACGA TAAGAATCCAAACCAGAAAATAATAAC
NG28	ATATACGCGTGGTGGCAGATCCGATATCACCGGT

6.1.11.2 Primers and probes for qPCR

Name	Sequence 5'-3'
NA 3' end forward	CTGGACTAGTGGGAGCATCA
NA 3' end reverse	ATGGTGAACGGCAACTCAG
NA 3' end probe	6-FAM-CACCGTCTGGCCAAGACCAATC-TAMRA
NA 5' end forward	TGGGTCAATCTGTATGGTAGTC
NA 5' end reverse	GCTGCCTTGGTTGCATATT
NA 5' end probe	6-FAM-TGGATTAGCCATTCAATTCAAACCGGA-TAMRA
eGFP forward	CCACAAGTTCAGCGTGTCC
eGFP reverse	GAACTTCAGGGTCAGCTTGC
eGFP probe	6-FAM-TGGCATCGCCCTCGCCCTCG-TAMRA
NP forward	CCAAAGGCACCAAACGATCT
NP reverse	AGTGGCATTCTGGCGTTCTC
NP probe	6-FAM-ACGAACAGATGGAGACTG-TAMRA

6.1.11.3 Primers for ribosome profiling

Name	Sequence
NN-RA3	P–NNTGGAATTCTCGGGTGCCAAGG-InvdT
OR5-NN	5′–GUUCAGAGUUCUACAGUCCGACGAUCNN
RTP	5′–GCC–TTGGCACCCGAGAATTCCA
RP1	5′–AATGATACGGCGACCACCGAGATCTACACGTTCAGAG TTCTACAGTCCGA
RPI1–RPI27	5'–CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA

6.1.12 Plasmids

Name	Purpose
pHW181-PB2	Reverse genetics WSN plasmids
pHW182-PB1	Reverse genetics WSN plasmids
pHW183-PA	Reverse genetics WSN plasmids
pHW184-HA	Reverse genetics WSN plasmids
pHW185-NP	Reverse genetics WSN plasmids
pHW186-NA	Reverse genetics WSN plasmids
pHW187-M	Reverse genetics WSN plasmids
pHW188-NS	Reverse genetics WSN plasmids
pVitro2-Hygro-MCS	Dual expression plasmid for NA and eGFP gene, Invitrogen, Carlsbad

6.1.13 Sequences of recoded NA segment

>WT

CAGGACTCAACTTCAGTGATATTAACCGGCAATTCATCTCTTTGTCCCATCCGTGGGTGGGCTATACACAGCAAA
GACAATGGCATAAGAATTGGTTCCAAAGGAGACGTTTTTGTCATAAGAGAGCCTTTTATTTCATGTTCTCACTTG
GAATGCAGGACCTTTTTTCTGACTCAAGGCGCCTTACTGAATGACAAGCATTCAAGGGGGACCTTTAAGGACAGA
AGCCCTTATAGGGCCTTAATGAGCTGCCCTGTCGGTGAAGCTCCGTCCCCGTACAATTCAAGGTTTGAATCGGTT
GCTTGGTCAGCAAGTGCATGTCATGATGGAATGGGCTGGCTAACAATCGGAATTTCTGGTCCAGATGATGGAGCA
GTGGCTGTATTAAAATACAACCGCATAATAACTGAAACCATAAAAAGTTGGAGGAAGAATATATTGAGAACACAA
GAGTCTGAATGTACCTGTGTAAATGGTTCATGTTTTACCATAATGACCGATGGCCCAAGTGATGGGCTGGCCTCG
TACAAAATTTTCAAGATCGAGAAGGGGAAGGTTACTAAATCGATAGAGTTGAATGCACCTAATTCTCACTACGAG
GAATGTTCCTGTTACCCTGATACCGGCAAAGTGATGTGTGTGTGCAGAGACAATTGGCACGGTTCGAACCGACCA
TGGGTGTCCTTCGACCAAAACCTAGATTATAAAATAGGATACATCTGCAGTGGGGTTTTCGGTGACAACCCGCGT
CCCAAAGATGGAACAGGCAGCTGTGGCCCAGTGTCTGCTGATGGAGCAAACGGAGTAAAGGGATTTTCATATAAG
TATGGCAATGGTGTTTGGATAGGAAGGACTAAAAGTGACAGTTCCAGACATGGGTTTGAGATGATTTGGGATCCT
AATGGATGGACAGAGACTGATAGTAGGTTCTCTATGAGACAAGATGTTGTGGCAATAACTAATCGGTCAGGGTAC
AGCGGAAGTTTCGTTCAACATCCTGAGCTAACAGGGCTAGAC

>72Min/NA-CPD

CAGGATAGTACTAGCGTGATACTAACCGGTAATTCGTCACTATGTCCGATTAGGGGGTGGGCTATACACTCTAAG
GATAATGGCATACGTATAGGGTCTAAAGGCGATGTGTTCGTTATACGCGAACCTTTTATATCATGTTCGCATTTA
GAGTGTAGAACTTTTTTTCTGACACAAGGCGCCCTGTTAAACGATAAGCATAGTAGGGGGACTTTTAAGGATAGG
TCACCATATAGGGCACTAATGTCATGTCCCGTAGGCGAAGCACCTAGTCCATACAATAGTAGATTCGAATCCGTC
GCATGGTCCGCTAGCGCATGTCACGACGGTATGGGGTGGTTGACAATCGGTATATCCGGACCTGACGACGGAGCC
GTTGCCGTGCTTAAATACAATCGGATTATAACCGAGACTATTAAGTCATGGCGTAAAAATATACTGAGAACCCAA
GAGTCTGAGTGTACATGCGTTAACGGATCATGTTTTACAATTATGACTGACGGACCTAGCGACGGATTGGCATCA
TACAAAATTTTTAAAATCGAGAAAGGTAAGGTGACAAAATCGATCGAGTTGAACGCTCCTAATTCGCATTACGAA
GAGTGTTCATGTTATCCCGATACCGGTAAGGTTATGTGCGTTTGTAGGGATAATTGGCACGGATCTAATAGGCCA
TGGGTGTCATTCGATCAAAATCTAGACTATAAGATAGGGTACATCTGCAGCGGCGTATTCGGCGATAATCCTAGA
CCGAAAGACGGAACCGGTAGTTGCGGACCGGTTAGCGCAGACGGCGCTAATGGCGTTAAGGGGTTTTCATACAAA
TACGGTAACGGAGTGTGGATAGGCAGAACTAAATCCGATAGTTCTAGACACGGATTCGAAATGATATGGGATCCT
AACGGATGGACCGAAACCGATAGTAGGTTTAGTATGCGACAAGACGTAGTCGCTATAACTAATAGATCCGGGTAT
TCTGGATCTTTTGTGCAACATCCAGAGTTAACCGGGTTAGAC

>46-Med

CAGGATTCAACATCAGTCATATTAACCGGTAATTCCTCATTGTGTCCGATACGCGGATGGGCAATACACTCTAAG
GATAACGGCATTAGGATAGGGTCTAAGGGAGACGTTTTTGTGATAAGGGAGCCTTTTATCTCATGCTCTCACCTA
GAATGCAGAACATTTTTTCTGACTCAAGGCGCCTTGCTGAATGACAAACATTCAAGGGGGACTTTTAAAGACAGA
TCACCGTATAGGGCCTTAATGTCTTGCCCAGTCGGGGAAGCCCCTAGCCCTTACAATAGTAGATTCGAAAGCGTA
GCTTGGTCCGCTTCCGCATGTCATGACGGAATGGGATGGCTAACAATCGGAATTAGCGGCCCAGATGACGGAGCA
GTGGCAGTGTTAAAGTACAATAGGATAATAACCGAAACCATAAAGTCCTGGCGTAAAAATATACTTAGAACCCAA
GAGTCCGAATGTACTTGTGTTAACGGTAGTTGTTTTACCATTATGACCGATGGCCCATCTGACGGTCTGGCATCG
TATAAAATTTTCAAGATTGAGAAAGGTAAGGTTACTAAATCGATAGAGTTGAATGCTCCCAATTCTCATTACGAA
GAGTGTAGTTGTTACCCTGATACAGGCAAAGTGATGTGTGTATGTAGAGACAATTGGCACGGTTCGAACCGACCT
TGGGTGTCATTCGACCAAAATCTAGACTACAAAATCGGGTATATCTGCAGTGGGGTTTTCGGTGATAATCCGAGA
CCCAAAGATGGAACAGGCTCGTGTGGCCCAGTGAGTGCTGATGGAGCTAACGGAGTTAAGGGTTTTTCATACAAA
TACGGCAATGGCGTTTGGATAGGGAGGACTAAGAGCGATAGTAGTAGGCACGGGTTTGAGATGATATGGGATCCT
AACGGATGGACCGAGACTGATAGTAGATTCTCTATGCGTCAAGACGTTGTGGCAATAACTAATCGGTCAGGGTAT
AGCGGATCATTTGTACAACATCCTGAGTTAACCGGACTAGAT

>27-Min

CAAGACTCAACCTCAGTGATACTAACCGGCAATTCTAGCTTATGCCCTATTAGGGGGTGGGCTATACACTCCAAA
GACAATGGCATACGCATAGGTAGTAAGGGAGACGTTTTTGTGATACGTGAACCCTTTATATCCTGTTCGCATCTA
GAGTGTCGTACATTTTTTTTGACCCAAGGCGCCTTACTGAATGACAAACACTCTAGGGGAACCTTTAAGGATAGA
TCCCCATACAGAGCCCTAATGTCATGTCCGGTCGGTGAGGCACCTAGCCCATACAACTCTAGGTTTGAATCGGTT
GCATGGTCAGCTAGTGCATGCCATGATGGTATGGGATGGTTGACAATCGGAATCTCAGGCCCTGATGATGGGGCA
GTGGCAGTCCTTAAATACAATCGGATAATCACTGAGACTATTAAGAGTTGGCGAAAAAACATACTGAGAACCCAA
GAGTCTGAATGCACTTGTGTTAATGGATCCTGTTTTACAATTATGACCGATGGACCCTCTGATGGCCTAGCTAGC
TATAAGATATTCAAAATCGAGAAAGGTAAGGTAACCAAATCGATAGAACTGAATGCCCCTAACTCCCATTACGAG
GAATGCTCATGTTACCCTGATACCGGCAAAGTTATGTGTGTGTGTAGAGACAATTGGCACGGTAGCAATAGGCCA
TGGGTTAGTTTCGACCAAAACTTAGACTATAAGATAGGGTACATATGTTCGGGAGTGTTCGGTGACAATCCTAGA
CCGAAAGATGGTACAGGGTCATGTGGACCAGTGAGTGCTGATGGAGCTAACGGAGTTAAGGGATTCTCATACAAA
TATGGCAATGGGGTATGGATAGGCAGAACTAAGTCTGACTCTAGTAGGCATGGGTTTGAGATGATTTGGGATCCG
AATGGGTGGACTGAGACTGATAGTAGGTTTAGTATGAGACAGGATGTAGTGGCAATTACTAATAGGTCAGGGTAT
AGCGGATCATTCGTTCAACACCCTGAATTGACAGGCTTAGAC

>72H-Max

CAAGATTCAACTTCAGTAATATTAACTGGAAATTCTTCTCTTTGTCCTATTCGTGGTTGGGCAATTCATTCAAAG
GACAACGGAATAAGGATCGGAAGTAAAGGAGATGTTTTCGTGATTCGCGAGCCTTTTATTTCTTGTTCTCATTTA
GAATGTAGGACTTTTTTTCTAACTCAAGGTGCTTTGCTAAATGATAAACATTCAAGAGGAACATTTAAAGATAGA
AGTCCTTACAGAGCTTTGATGTCGTGTCCTGTTGGCGAGGCTCCTTCTCCTTATAATTCAAGGTTCGAGAGCGTC
GCATGGAGCGCTTCCGCATGTCACGACGGGATGGGATGGCTAACTATTGGAATATCCGGGCCCGACGACGGCGCC
GTGGCTGTTCTAAAGTACAATAGAATAATAACCGAGACAATAAAGTCGTGGAGGAAGAATATATTAAGAACTCAA
GAAAGTGAATGTACTTGCGTGAACGGCTCGTGTTTTACAATAATGACCGACGGGCCGTCCGACGGGCTGGCAAGT
TATAAAATATTTAAGATCGAGAAAGGAAAAGTAACAAAATCAATAGAATTAAATGCACCAAATTCACACTACGAA
GAATGTTCTTGTTACCCAGATACCGGGAAAGTAATGTGCGTGTGTAGAGATAATTGGCACGGAAGTAATCGGCCG
TGGGTTTCTTTCGATCAAAATTTGGATTATAAAATAGGTTACATCTGCAGCGGCGTTTTCGGCGATAATCCAAGG
CCGAAGGACGGGACCGGCTCGTGCGGGCCCGTGAGCGCCGACGGCGCCAACGGCGTGAAAGGATTTAGTTACAAG
TACGGAAACGGCGTGTGGATCGGTCGAACCAAGAGCGACAGTTCAAGGCACGGTTTCGAGATGATATGGGATCCA
AACGGTTGGACCGAGACCGATTCAAGATTTTCAATGAGGCAGGACGTCGTTGCAATAACAAATCGTTCCGGTTAT
TCCGGAAGTTTTGTTCAACATCCTGAGCTGACCGGGCTGGAT

>72L-Max

CAAGATTCAACAAGCGTGATATTAACCGGTAATTCATCTCTTTGTCCTATTCGAGGCTGGGCAATTCATTCAAAA
GATAATGGGATAAGAATTGGGTCAAAGGGCGATGTGTTCGTTATTCGCGAACCTTTTATAAGTTGTTCTCATTTG
GAGTGTAGGACTTTTTTTCTAACTCAAGGCGCCTTGCTAAACGATAAACATTCAAGAGGGACTTTTAAGGACAGG
AGTCCTTACAGAGCTTTGATGAGTTGTCCCGTAGGCGAAGCACCGTCGCCGTATAATTCTCGTTTCGAATCCGTC
GCATGGTCCGCTAGCGCATGTCACGACGGTATGGGCTGGCTGACCATCGGTATATCCGGACCTGACGACGGAGCC
GTTGCCGTGCTAAAATATAATAGAATAATAACCGAGACAATAAAATCATGGAGGAAGAATATATTAAGAACTCAA
GAGTCGGAGTGTACTTGCGTTAACGGAAGTTGTTTTACTATAATGACAGACGGACCAAGCGACGGATTAGCAAGT
TATAAAATATTTAAGATCGAGAAGGGGAAAGTTACTAAAAGTATCGAGCTAAACGCTCCAAATTCACACTACGAA
GAGTGTTCTTGTTACCCCGATACCGGTAAAGTGATGTGCGTTTGTAGAGATAATTGGCACGGAAGTAATCGGCCG
TGGGTGTCTTTCGATCAAAATTTAGATTATAAAATAGGTTACATCTGCAGCGGCGTATTCGGCGATAATCCAAGG
CCAAAGGACGGAACCGGTTCGTGCGGACCTGTGAGCGCAGACGGCGCTAATGGCGTTAAAGGATTTTCTTACAAG
TACGGTAACGGAGTGTGGATTGGGAGGACAAAGTCCGATTCTTCAAGGCACGGATTCGAAATGATATGGGATCCT
AACGGATGGACCGAAACCGATTCAAGATTTTCAATGAGGCAGGACGTAGTCGCTATAACAAATCGTTCCGGGTAC
TCGGGGAGTTTTGTTCAACATCCTGAGCTGACCGGGCTGGAC

>CD

CAAGATTCGACGTCGGTAATACTAACGGGTAATTCGTCGCTATGTCCGATACGTGGTTGGGCGATACATTCGAAA
GATAATGGTATACGTATAGGTTCGAAAGGTGATGTATTTGTAATACGTGAACCGTTTATATCGTGTTCGCATCTA
GAATGTCGTACGTTTTTTCTAACGCAAGGTGCGCTACTAAATGATAAACATTCGCGTGGTACGTTTAAAGATCGT
TCGCCGTATCGTGCGCTAATGTCGTGTCCGGTAGGTGAAGCGCCGTCGCCGTATAATTCGCGTTTTGAATCGGTA
GCGTGGTCGGCGTCGGCGTGTCATGATGGTATGGGTTGGCTAACGATAGGTATATCGGGTCCGGATGATGGTGCG
GTAGCGGTACTAAAATATAATCGTATAATAACGGAAACGATAAAATCGTGGCGTAAAAATATACTACGTACGCAA
GAATCGGAATGTACGTGTGTAAATGGTTCGTGTTTTACGATAATGACGGATGGTCCGTCGGATGGTCTAGCGTCG
TATAAAATATTTAAAATAGAAAAAGGTAAAGTAACGAAATCGATAGAACTAAATGCGCCGAATTCGCATTATGAA
GAATGTTCGTGTTATCCGGATACGGGTAAAGTAATGTGTGTATGTCGTGATAATTGGCATGGTTCGAATCGTCCG
TGGGTATCGTTTGATCAAAATCTAGATTATAAAATAGGTTATATATGTTCGGGTGTATTTGGTGATAATCCGCGT
CCGAAAGATGGTACGGGTTCGTGTGGTCCGGTATCGGCGGATGGTGCGAATGGTGTAAAAGGTTTTTCGTATAAA
TATGGTAATGGTGTATGGATAGGTCGTACGAAATCGGATTCGTCGCGTCATGGTTTTGAAATGATATGGGATCCG
AATGGTTGGACGGAAACGGATTCGCGTTTTTCGATGCGTCAAGATGTAGTAGCGATAACGAATCGTTCGGGTTAT
TCGGGTTCGTTTGTACAACATCCGGAACTAACGGGTCTAGAT

>CPO

CAAGACAGCACCTCAGTAATATTAACTGGAAATTCATCTCTTTGTCCTATTCGTGGCTGGGCAATTCATTCAAAA
GATAATGGAATAAGAATTGGAAGTAAAGGAGATGTTTTTGTAATAAGAGAGCCATTTATTTCTTGTTCACACCTG
GAGTGCAGGACCTTCTTTCTAACTCAAGGTGCTTTGCTAAATGACAAGCACAGTCGTGGCACCTTCAAGGACAGA
AGTCCTTATCGAGCTTTGATGAGCTGTCCTGTTGGAGAGGCCCCGTCGCCTTATAATTCAAGATTTGAGTCGGTG
GCCTGGTCAGCATCTGCATGTCATGATGGGATGGGCTGGCTAACAATAGGAATATCAGGGCCTGATGATGGTGCT
GTGGCTGTGCTGAAGTACAACAGAATAATAACTGAAACAATAAAATCCTGGAGGAAGAATATATTAAGAACTCAA
GAAAGTGAATGTACCTGTGTAAATGGAAGTTGTTTCACCATCATGACAGATGGGCCTTCAGATGGGCTGGCTTCC
TACAAAATATTTAAAATAGAAAAAGGAAAAGTCACCAAGTCAATAGAATTAAATGCACCAAATTCACACTATGAG
GAGTGCAGCTGCTACCCGGACACTGGGAAGGTGATGTGTGTCTGCAGGGACAACTGGCATGGCAGCAACCGGCCC
TGGGTTTCTTTTGATCAAAATTTGGACTACAAAATAGGCTACATCTGTTCTGGTGTTTTTGGGGACAACCCCAGG
CCAAAAGATGGAACTGGCAGCTGTGGGCCTGTTTCTGCAGATGGTGCAAATGGTGTGAAGGGCTTCTCCTACAAA
TATGGAAATGGTGTTTGGATTGGGAGGACCAAGAGTGACTCGTCCAGGCATGGATTTGAGATGATCTGGGACCCA
AATGGCTGGACAGAAACAGACTCGCGCTTCTCCATGAGGCAAGATGTGGTGGCCATCACCAACAGAAGTGGCTAC
AGTGGTTCTTTTGTTCAGCACCCGGAGCTAACTGGATTAGAT

>CO

CAGGACAGCACCAGCGTGATCCTGACCGGCAACAGCAGCCTGTGCCCCATCAGAGGCTGGGCCATCCACAGCAAG
GACAACGGCATCAGAATCGGCAGCAAGGGCGACGTGTTCGTGATCAGAGAGCCCTTCATCAGCTGCAGCCACCTG
GAGTGCAGAACCTTCTTCCTGACCCAGGGCGCCCTGCTGAACGACAAGCACAGCAGAGGCACCTTCAAGGACAGA
AGCCCCTACAGAGCCCTGATGAGCTGCCCCGTGGGCGAGGCCCCCAGCCCCTACAACAGCAGATTCGAGAGCGTG
GCCTGGAGCGCCAGCGCCTGCCACGACGGCATGGGCTGGCTGACCATCGGCATCAGCGGCCCCGACGACGGCGCC
GTGGCCGTGCTGAAGTACAACAGAATCATCACCGAGACCATCAAGAGCTGGAGAAAGAACATCCTGAGAACCCAG
GAGAGCGAGTGCACCTGCGTGAACGGCAGCTGCTTCACCATCATGACCGACGGCCCCAGCGACGGCCTGGCCAGC
TACAAGATCTTCAAGATCGAGAAGGGCAAGGTGACCAAGAGCATCGAGCTGAACGCCCCCAACAGCCACTACGAG
GAGTGCAGCTGCTACCCCGACACCGGCAAGGTGATGTGCGTGTGCAGAGACAACTGGCACGGCAGCAACAGACCC
TGGGTGAGCTTCGACCAGAACCTGGACTACAAGATCGGCTACATCTGCAGCGGCGTGTTCGGCGACAACCCCAGA
CCCAAGGACGGCACCGGCAGCTGCGGCCCCGTGAGCGCCGACGGCGCCAACGGCGTGAAGGGCTTCAGCTACAAG
TACGGCAACGGCGTGTGGATCGGCAGAACCAAGAGCGACAGCAGCAGACACGGCTTCGAGATGATCTGGGACCCC
AACGGCTGGACCGAGACCGACAGCAGATTCAGCATGAGACAGGACGTGGTGGCCATCACCAACAGAAGCGGCTAC
AGCGGCAGCTTCGTGCAGCACCCCGAGCTGACCGGCCTGGAC

6.1.14 Software and programs

Software	Purpose
Chemi-Capt	Western blot images
Felxbar, tophat2, bowtie2, quasR, RiboDiff	Ribosome profiling analysis
GraphPad Prism 8.1.2	Graphs
Image J 1.41	Plaque size measurement
Microsoft Excel 2016	Calculations
Microsoft PowerPoint 2016	Graphs
pDRAW32	Cloning
Perl, programming language	Design of recoded NA genes
Pymol	Protein structure
R, version 3.5.3	Plots
StepOne software v2.3	RT-qPCR

6.2 Methods

6.2.1 Virus design

Eight mutant influenza A viruses were designed, which differ in their RNA sequences of the NA gene but encode the same amino acid. Only the middle part of the NA gene was recoded, the first 183 nucleotides and the last 162 nucleotides were unchanged to ensure correct packaging and replication of the NA segment. Custom perl scripts were used to design recoded NA genes by reshuffling the synonymous codons to create codon pairs with high or low codon pair scores[65] and changes in the CpG dinucleotide content without altering the codon bias. Scores for individual codon pairs we calculated based on the frequency of codon pairs in the human genome in the CCDS database (CCDS 15, Homo sapiens GRCh37.p13). For codon optimization or deoptimization, two recoded NA genes were constructed which contain only the most over or underrepresented codons. Recoded NA genes were ordered as linear DNA fragments (gBlocks) from IDT Technologies, Inc and cloned into the pHW186-NA plasmid of the 8-segment reverse genetics system of the IAV strain A/WSN/1933. The recoded NA fragments are listed in section *6.1.13 Sequences of recoded NA segment*. The sequences are also deposited in NCBI GenBank under accession numbers MN176602 to MN176607.

6.2.2 Cloning

Cloning of recoded NA genes into NA-WT plasmid pHW186-NA

The linear NA DNA fragment was amplified by PCR using the primers MS307 and MS310. Amplified PCR product and pHW186-NA plasmid were digested with restriction endonucleases SalI-HF and SepI and ligated with T4 DNA ligase. 50 µl chemical competent cells were added to ligated plasmids, incubated for 30 min on ice, followed by a heat shock for 30 seconds at 42 °C and cooled down on ice. Cells were incubated for 1 h at 37 °C with 800 µl SOC medium and plated on agar plate containing ampicillin. Plates were incubated over night at 37 °C. Clones were picked and grown in LB medium overnight. Plasmids were purified from *E. coli* culture and sequenced with primers NG03 and NG04 to examine the outcome of ligation.

Cloning recoded NA genes in dual expression plasmids pVitro2-eGFP

The WT and recoded NA genes were amplified from pHW plasmids with NG19 and NG20. The PCR product was digested with BglII, SalI and ScaI and purified. The pVitro2-eGFP plasmid was digested with BamHI and SalI, dephosphorylated and the 7 kb fragment gel purified. The PCR product and the linearized plasmids were ligated and clones were selected on hygromycin

plates. To confirm the correct sequences, plasmids were sequenced with DK133 and DK132 primers.

Cloning N-terminal flag tag on NA genes in dual expression plasmid pVitro-NA-eGFP

NA genes were amplified from pVitro-NA-eGFP plasmids with NG27 and NG28 primers. The 8.4 kb PCR product was gel purified and digested with DpnI and MluI, ligated and selected on hygromycin plates. To confirm correct sequences, plasmids were sequenced with DK132.

6.2.3 Cell culture and virus recovery

MDCK, A549 and HEK 293T cells were grown in growth medium at 37°C and 5 % CO_2. To recover recombinant influenza A viruses the eight-plasmid reverse genetics system was used[108]. The seven original pHW plasmids and the recoded pHW186-NA plasmids were used to generate the mutant viruses. 300 ng of each plasmid were transfected in HEK 293T cells in 6-well dishes using 100 µl OptiMEM and 10 µl PEI (1 mg/ml). After 2 h, DMEM medium containing 10 % FBS was changed to infection medium and incubated at 37 °C and 5 % CO_2 for 48 h. Supernatant was harvested, centrifuged at 4 °C for 5 min at 2,000 × g, to clear of from cell debris, and used to infect 6-well dishes of MDCK cells in infection medium. Supernatant with viruses were harvested 2-3 days after infection at cytopathic effect (CPE) of 80-90 %. Supernatant containing influenza virus was aliquoted and stored at -80 °C. To confirm the correct sequence of the NA segment, viral RNA was isolated (RTP DNA/RNA Virus Mini Kit), reverse transcribed with M-MLV reverse transcriptase and Sanger sequenced using primers NG03 and NG04.

6.2.4 Virus titration and plaque size measurement

For virus titration and plaque size measurement, plaque assays were performed on MDCK cells in 6-wells plates according to standard procedure. Cells were infected with 500 µl using serial 10-fold dilutions of viral stocks in infection medium. After adsorption for 1h, the inoculum was removed and cells were overlaid with overlay medium. Two to three days after infection, cells were washed with PBS, fixed with 4 % formaldehyde and stained with 0.1 % crystal violet in 25 % methanol for 30 min. Plaques were counted to determine the titre and plaque sizes were measured with ImageJ.

6.2.5 Multi-step growth kinetics

To assess growth kinetics, 100 % confluent MDCK or A549 cells were infected in 12-well plates with 500 µl recombinant virus at a MOI of 0.0001. Infection medium contained 2 µg/ml TPCK-Trypsin for MDCK cells and 1 µg/ml for A549 cells. After 1 h adsorption, the medium was changed to 1 ml infection medium. Every eight hours, at indicated time points, supernatant was taken and viral titres were determined by plaque assays.

6.2.6 SDS-PAGE and Western blot analysis

For detection of protein production in recombinant viruses, 6-well plates of 100 % confluent MDCK cells were infected with 5 MOI of different viruses. After adsorption for 1 h, infectious medium was removed and cells were incubated with infection medium. Cells were incubated for additional 5 h, then were harvested and analysed for protein expression.

For detection of NA and eGFP protein from plasmids, 6-well dishes of 70 % confluent HEK 293T cells were transfected with pVitro2-eGFP-NA plasmids (0.5 µg plasmid, 100 µl OptiMEM and 10 µl PEI (1mg/ml)). Recoded pVitro2-eGFP-NA plasmids contained non-recoded eGFP and different NAs. After 24 h, cells were harvested and cell pellet was lysed.

Cell pellets were resuspended any lysed in 80 µl Ripa I, 0.5 µl benzonase was added, vortexed and incubated for 15 min on ice. 80 µl Ripa II was added, vortexed and incubated for another 15 min on ice. Samples were centrifuged at 15,000 g at 4 °C for 5 min and supernatant was taken. Supernatant was mixed with Laemmli loading dye, 2 min incubated at 98 °C and cooled down on ice. Afterwards proteins were separated by SDS-PAGE followed by semi-dry Western blotting for 1 h at 12 V on a PVDF membrane. The membranes were blocked for 1 h in 3 % milk in PBS-T and incubated over night with a primary antibody (1:2,000 dilution) in 3 % milk. After several washing steps with PBS-T, horseradish peroxidase-coupled secondary antibody was used (1:2,000 dilution) for chemiluminescence detection with ECL prime Western blotting detection kit.

6.2.7 Mass spectrometry

To identify the truncated NA protein fragment, mass spectrometry was performed. Therefore, FLAG-tagged 27-Min NA protein was purified with beads, coupled to an anti-FLAG antibody and precipitated. From beads eluted proteins were separated by SDS-PAGE. Gel slice containing the truncated NA protein were cut out and its proteins were reduced with DTT, ankylated with chloroacetamide and digested *in situ* with trypsin, chymotrypsin, elastase or AspN[109]. The digested peptides were extracted from the gel matrix and analysed by nano-

liquid chromatography tandem mass spectrometry on a LTQ Orbitrap Elite mass spectrometer[109]. Amino acid sequence of protein fragment was identified by searching the UniProt virus database with Mascot search engine.

6.2.8 RT-qPCR

RNA from cell lysate was isolated with the RNeasy Plus Mini Kit, followed by reverse transcription with random hexamers using M-MLV reverse transcriptase according to the manufacturer's instructions. Primers and TaqMan probes for NA, NP and eGFP listed in *'Primers and probes for qPCR'* were used to quantify cDNA. For all NA genes, the same primer and probe sets were used as the probe binds at the 3' or 5' end in the non-recoded region of the NA gene. The cDNA quantification was performed on a StepOnePlus Real-Time PCR System using 3 µl of cDNA, 0.2 µl (10 µM) probe, 0.12 µl (10 µM) primers and 4 µl of 5x HOT FIREPol DNA polymerase in a total volume of 20 µl. The cycling conditions were the following: 95 °C, 15 min; 40 × (95°C, 15 s; 60 °C, 30 s). Calibration curves were generated using a serial dilution of known concentration of the pVitro2-eGFP-NA plasmids or the IAV pHW186-NA and pHW185-NP plasmids.

6.2.9 RNA degradation measurement based on transcription inhibition

HEK 293T cells were transfected with dual expression pVitro2-eGFP-NA plasmids containing eGFP and recoded NA genes. After 24 h transcription inhibitor flavopiridol was added to culture medium to a final concentration of 1 µM. Cells were incubated with the inhibitor for additional 0, 1, 2 or 5 h, then RLT buffer from RNeasy Plus Mini Kit was added to stop all reactions and cells were harvested. The RNA was isolated using RNeasy Plus Mini Kit, reverse transcribed and qPCR was performed as described above in section '6.2.8 RT-qPCR'.

6.2.10 RNA synthesis and degradation measurements based on BrU labelling

The rates of RNA synthesis and degradation were determined with assays that are based on metabolic labelling of nascent RNA using BrU[110]. 293T cells grown in 6-well plates were transfected with pVitro2-eGFP-NA plasmids. 24 h after transfection cell culture medium was removed and cells were incubated with DMEM containing 10 % FBS and 2 mM BrU. After 1 h, excess of BrU was removed. For 0 h time point RLT buffer was directly added, all other cells were washed 3 times with PBS. 1.2 ml uridine (20 mM) in DMEM containing 10 % FBS was added for 0.5, 1 or 2 h and incubated under normal growth conditions. At indicated time points

cells were washed with PBS, RLT buffer was added and cells were harvested. RNA was isolated with RNeasy Plus Mini Kit.

For the preparation of the purification beads, 15 µl of magnetic protein G Dynabeads per sample were transferred to a 1.5 ml microfuge Protein Low binding tube. Beads were captured with magnetic stand and storage buffer removed. For 6 samples, 90 µl beads were resuspended in 600 µl BrU-IP buffer (0.05 % Tween in RNase free PBS), beads captured on the magnetic stand and supernatant aspirated. Beads were washed once with BrU-IP buffer and resuspended in 300 µl (6 samples) BrU-IP buffer supplemented with 1 µl RNase inhibitor RNasin. For each sample 1 µl (0.5 µg) anti-BrdU antibody was add solution was incubated for 30 min at room temperature with gentle rotation. Beads were washed three times as above and resuspended in 100 µl BrU-IP buffer per sample supplemented with RNase inhibitor.

For isolation of BrU-labelled RNA, 100 µl of the magnetic beads conjugated with the anti-BrdU antibody and 135 µl 1x BrU-IP buffer were added to the isolated RNA, followed by 1 h of incubation at room temperature with gentle rotation in protein low binding tubes. Beads were washed twice with BrU-IP buffer for 5 min, then captured on the magnetic stand and supernatant was aspirated. Three additional washes with BrU-IP buffer were performed and the bead pellet was resupended in 40 µl DEPC-water and incubate for 10 min in a 95°C heat block to elute BrU-RNA from the beads. Beads were captured in the magnetic stand and supernatant containing BrU RNA collected into a clean 1.5 ml microfuge tube. Afterwards NA and eGFP copies from isolated BrU RNA was reverse transcribed and measured by qPCR as described in section ‘6.2.8 RT-qPCR’.

For the comparison of the RNA synthesis rates transfected HEK 293T cells are labelled for 30 min with BrU and labelled RNA is measured by qPCR following the procedure described.

6.2.11 Ribosome profiling

HEK 293T cells were transfected with dual expression plasmids pVitro2-eGFP-NA for 24 h. Ribosome profiling was performed as previously described[111], with the difference that the mRNA fragments were excised between 18 and 35 nucleotide marker oligos, instead of 27 and 30. Briefly, cells were flash-frozen in liquid nitrogen, lysed and scraped of the plates. Ribosome-protected RNA fragments were purified with MicroSpin S-400 HR columns and RNA was extracted using TRIzol LS. The ribosomal RNA was removed using RiboZero Kit. The size range between 18 and 35 nucleotides was cut out from a 17 % denaturing urea-PAGE gel and RNA fragments were used for library preparation using 3’ NN-RA3 and 5’ OR5-NN adapters. RTP primer was used for reverse transcription and for PCR amplification, primers RP1 and RPI1 to RPI27 (containing barcodes) were used.

Total non-fragmented RNA extracted from cell lysates was converted into sequencing libraries using the Illumina TruSeq Stranded mRNA kit. Samples were pooled and sequenced on a HiSeq4000 (Illumina) device to generate 1x51 single-end reads. Read sequences and analyses are available under Gene Expression Omnibus (GEO) accession number GSE134752.

Sequencing reads were demultiplexed and adapter-trimmed using flexbar[112] and mapped to the hg19 version of the human genome using tophat2[113] or to the NA construct sequences using bowtie2[114]. Only reads with mapping quality 24 or higher were kept, which also filtered out reads mapping to the constant regions at the beginning and end of the NA constructs. Sequencing reads per gene as annotated in the RefSeq database were counted using quasR[115]. Translation efficiencies from read counts were calculated using RiboDiff[116].

6.2.12 Animal experiments

In a blinded experiment, 6-week old female Balb/c mice were purchased from Janvier. Animals were housed under standard specific pathogen-free (SPF) conditions. Food and water were given ad libitum. Eight groups (27-WT, 72-Min/CPD, 46-Med, 72H-Max, 72L-Max, 27-Min, CD, mock) with each nine animals were infected intranasally with 30 µl (5 x 10^4 PFU) of the different constructed mutant influenza viruses virus or with DMEM medium as control group. The body weight was measured every day over the 18 days of the experiment. Mice, which showed severe signs of disease or lost more than 20 % of their body weight, were euthanized to minimize suffering. On day 2, 5 and 18, three mice from each group were killed and blood and lungs were taken for further analysis.

In a second experiment mice were vaccinated with 30 µl (5 x 10^4 PFU) mutant viruses which did not harm the mice (72-Min, 46-Med, 27-Min, CD) or a mock control. 28 days post vaccination mice were challenged with 5 x 10^5 PFU the 27-WT virus and again morbidity and mortality were monitored. Additionally, to determine antibody titres, venous blood was taken on day 18 and 28 after initial infection and on day 38, at the termination of the experiment.

All animal experiments were conducted in accordance with the EU legislation and German law (paragraph 8 Tierschutzgesetz). Animal experiments were approved by the Landesamt für Gesundheit und Soziales in Berlin, Germany (approval G0291/17).

6.2.13 Viral load and copy number in the lungs

The right lung bulb was homogenized in 1 ml DMEM medium and plaque assays were performed to determine virus load as described in section *'6.2.4 Virus titration and plaque size measurement'*. To measure RNA copy numbers of the NA and NP fragment, 200 µl of the homogenized lung was used to isolate RNA with RNeasy Plus Mini Kit, reverse transcripted with random primers using M-MLV reverse transcriptase and the virus copy number of the NP and the deoptimized NA fragment was determined by qPCR as described in section '*6.2.8 RT-qPCR*'.

6.2.14 Antibody titres in mouse serum

For the detection of the antibody titers in the mouse serum a hemagglutination inhibition (HAI) and a neutralisation assay were performed of all serum samples as described in the World Health Organization manual on animal influenza diagnosis and surveillance[117]. For neutralization assay heat activated serial 2-fold dilutions of mouse serum was incubated for 2 h with 200 PFU of WT IAV. This serum-virus mixture was incubated for 1.5 h on 100 % confluent MDCK cells. Afterwards, medium was removed and infection medium was added. After 48 h, CPE was observed to determine neutralization titre.

For HAI test, inactivated serum was serial diluted in 96-well round bottom plates and chicken red blood cells were added. After 30 min, the HAI titre, the highest dilution of serum that prevents hemagglutination was determined.

6.2.15 IFN-γ detection by ELISA

Spleens were taken from mice, homogenized and to obtain single cell suspensions filtered through a 70 µm cell strainers. Erythrocytes in spleen cell suspension were lysed for 3 min using ACK buffer. Cells were washed, resuspended in cRPMI and counted using a CASY automated cell counter. 1×10^6 spleen cells were stimulated for 3 days with 10 µg/ml purified 27-WT virus, or 10 µg/ml control protein that was purified from cell supernatant. Cells were cultured in RPMI 1640 medium containing 10 % FCS, and Penicillin [100 U/ml]/Streptomycin [100 µg/ml]. Supernatants were analysed for IFN-γ using Ready-Set-Go Elisa Kits according to the manufacturer's instructions with Synergy HT ELISA reader.

To purify virus antigen, MDCK cells were infected with a low MOI of 27-WT virus. After 48 h, supernatant was collected and cleared from cell debris by 10 min centrifugation at 5,000 × g. 35 ml virus was purified using 2 ml of a 20 % sucrose cushion in TNE buffer by centrifugation for 2 h at 100,000 × g. For control protein supernatant of uninfected MDCK cells was used.

The resulting virus pellet was dissolved in 150 µl TNE buffer and protein concentration was determined with BCA assay. For BCA assay 50 µl of the sample was incubated for 30 min at 37 °C with 100 µl reaction mix (Roti Quant reagent 1 and 2 from Roti-Quant universal kit in a ratio of 15:1) and analysed with an ELISA reader. The standard was performed with serial dilutions of a BSA solution with known concentration.

6.2.16 Histopathology and Immunohistochemistry

The left lobe of the mouse lung was removed and fixed by immersion in formalin pH 7.0 for 48 h, embedded in paraffin, and cut into 2 µm sections. Lung sections were stained with hematoxylin and eosin for histological evaluation in a blinded fashion. The following pathological features were assessed: damage to the airway epithelium (necrosis of bronchi, bronchioli and alveoli) and inflammation (perivascular, interstitial, bronchus associated lymphoid tissue (BALT)) as well as pneumocyte type II hyperplasia (repair). Each feature was scored: 0 (no lesion), 1 (mild), 2 (moderate), 3 (severe).

For confirmation of influenza infection, immunohistochemistry was performed using polyclonal anti-Influenza A H1N1 antibodies on selected slides. Antigen retrieval was accomplished in citrate buffer for 12 minutes (600 Watt) in a microwave. Polyclonal goat anti-Influenza A H1N1 antibody (dilution 1:250) was incubated for 1 h at room temperature. Subsequently, slides were incubated with a secondary, alkaline phosphatase-conjugated rabbit anti-goat antibody (1:200) for 30 min at room temperature. The alkaline chromogen triamino-tritolyl-methanechloride (neufuchsin) was used as phosphatase substrate for colour development. All slides were counterstained with hematoxylin, dehydrated through graded ethanol, cleared in xylene and coverslipped.

7 Results

7.1 Part I: Attenuation mechanism of codon pair deoptimization

7.1.1 Design and recoding of viruses

To decipher the molecular basis of codon pair deoptimization, different recoded NA genes were designed. The mouse-adapted influenza A strain A/WSN/1933 (H1N1) was used as a model virus. The NA gene was chosen because it was shown to be amenable to deoptimization and does not contain alternative reading frames[62]. The first 183 N-terminal and the last 162 C-terminal nucleotides of the 454 amino acid long coding sequence, were not recoded to prevent disruption of *cis* signals that are essential for the virus packaging (*Figure 6A*).

Recoded genes were designed using custom computer algorithms, which reshuffle existing codons and calculate the resulting CPS. Genes were synthesized *de novo* and cloned in an eight-plasmid reverse genetics system[108]. Mutant viruses were recovered after cotransfection of 8 plasmids in susceptible HEK 293T cells.

The first aim of my research project was to design viruses that would enable to separate the contribution of underrepresented codon pairs from CpG dinucleotides on virus attenuation. To dissect these two factors, the number of CpG dinucleotides and the level of codon pair deoptimization was independently varied *(Figure 6B*).

Five recoded NA genes were constructed based on the WT sequence, followed by recovery of the mutant viruses (*Table 2*). The nomenclature for the recoded viruses is as follows: The first part of the name indicates the number of CpG dinucleotides in the recoded NA part, the second part indicates the degree of codon pair deoptimization. For example, the 27-Min gene has 27 CpG dinucleotides and the minimal CPS (maximal deoptimization) that could be reached through recoding without increasing the number of CpG dinucleotides in recoded sequences (CPS= -0.22). An exception to this rule is the name of the parental sequence, which contains 27 CpG dinucleotides and is named 27-WT.

Three different strategies were used to recode the NA sequence. In the first strategy, the sequence was recoded using the original codon pair deoptimization method[65]. Applying this method, the number of CpG dinucleotides is inadvertently increased. A maximal deoptimization with 72 CpG dinucleotides and a minimum codon pair score of -0.45 (**72-Min**) or a medium deoptimization (CPS= -0.23) with 46 CpG dinucleotides (**46-Med**) was reached (*Figure 6B*).

In the second approach, the number of CpG dinucleotides was artificially kept low to 27, by keeping the NNC-GNN codon pair the same as in the WT, but the codon pair score was

reduced as much as possible (**27-Min,** CPS= -0.22). A lower CPS for 27-Min is not attainable without increasing the number of CpG dinucleotides.

In the third approach, two different genes were designed to have a high number of CpG dinucleotides but without codon pair deoptimization. One construct **72H-Max** (H stands for high) has a high average CPS (-0.05), similar to the score of wild type (27-WT, CPS = 0.0), but 72 CpG dinucleotides are introduced. Only CpG dinucleotides at the codon boundary (NNC-GNN) can be introduced, because the codon usage is preserved. Codon pairs of the NNC-GNN type have typically very low CPS. To increase the average CPS of this construct, only codon pairs of the NNC-GNN type that have relatively high CPS were used. The other construct (**72L-Max,** L stands for low, CPS= -0.21) contains exactly the NNC-GNN codon pairs as 72-Min, which are the codon pairs with the lowest CPS. The remaining codon pairs were selected to contain high CPS (*Figure 6C*). The beneficial information that can be obtained from 72L-Max is, if only a few codon pairs with very low scores have a decisive role in codon pair deoptimization.

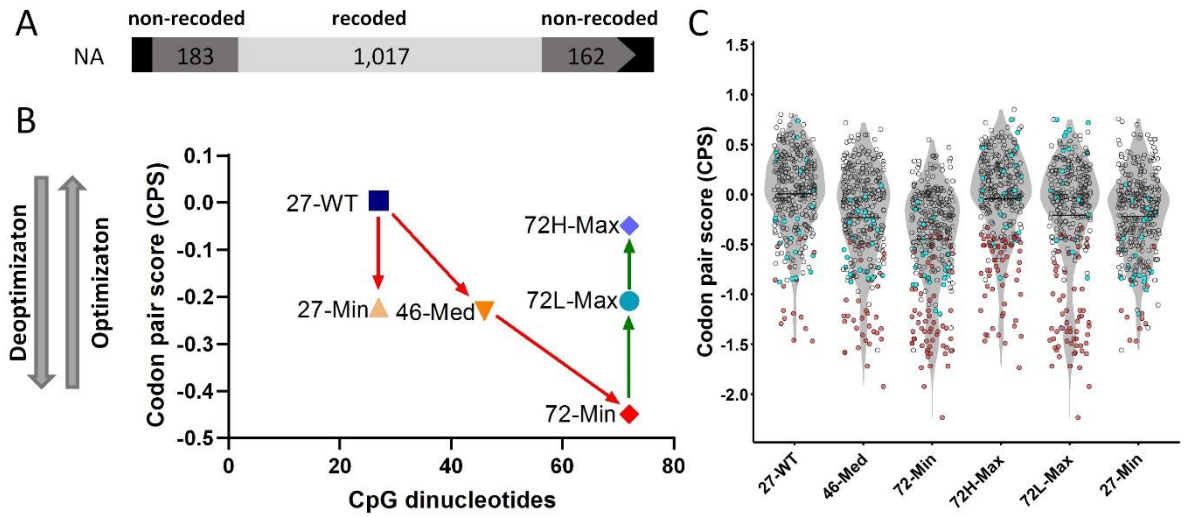

Figure 6: Properties of recoded NA genes. *(A) Structure of recoded IAV NA gene segment. Terminal non-coding (black) and coding regions (grey) were not recoded due to packaging signals. Only the central part of the NA gene was recoded (light grey). Length of individual NA gene fragments are given in nucleotides. (B) Recoded NA sequences with their average codon pair scores (CPS) and number of CpG dinucleotides. (C) Distribution of codon pairs with their corresponding CPS in recoded NA genes. The grey violin plots show the CPS distribution, with a black line indicating mean CPS. Each black circle represents the CPS of a codon pair. Red filled circles represent codon pairs that contain a CpG at the codon boundary (NNC-GNN). All codon pairs of the NNC-GNN type have negative CPS. Blue filled circles represent codon pairs, which contain CpG dinucleotides within the codons (CGN, or NCG).*

Table 2. Properties of recoded IAV H1N1 strain A/WSN/1933 NA genes.

Name	Description	CPS	CAI	Codon identity (%)	Nucleotide identity (%)	CpG	UpA
27-WT	original CPS	0.00	0.74	100.0	100.0	27	54
72-Min	minimal CPS	-0.45	0.74	37.8	75.6	72	100
46-Med	reduced CPS	-0.23	0.74	61.4	83.6	46	77
72H-Max	same number of NNC-GNN as 72-Min, maximal CPS	-0.05	0.74	39.2	76.3	72	57
72L-Max	same NNC-GNN as 72-Min, maximal CPS	-0.21	0.74	38.9	76.4	72	71
27-Min	same NNC-GNN as 27-WT, minimal CPS	-0.22	0.74	52.5	80.2	27	82

CPS – average codon pair score; CAI – codon adaptation index (a measure of codon bias); Codon/Nucleotide identity - percentage of codons/nucleotides that occupy the same position in WT and recoded genes; CpG – number of CpG dinucleotides in recoded sequences; UpA - number of UpA dinucleotides in recoded sequences

7.1.2 Characterization of recoded viruses in cell culture

The viral fitness of all mutant viruses was analysed, depending on their codon pair bias and their CpG dinucleotide content. I determined growth kinetics, plaque sizes and protein expression of the recoded NA genes (*Figure* 7).

To analyse the replication kinetics of different mutant viruses, Madin-Darby canine kidney (MDCK) and human alveolar basal epithelial (A549) cell were infected with a low MOI and virus titres were determined by plaque assay every eight hours over 56 h or 96 h, respectively. All viruses replicated well on MDCK cells and grew to titres of up to 10^8 PFU/ml comparable to 27-WT (*Figure* 7*A*). On A549 cells, virus replication is around 32 h delayed compared to MDCK cells. Viruses with high numbers of CpG dinucleotides (72H-Max and 72L-Max) grew comparable to 27-WT virus on A549 cell. The growth of codon pair deoptimized viruses instead (72-Min, 46-Med and 27-Min), was severely hampered with 1,000- to 10,000-fold lower titres observed than for 27-WT virus (*Figure* 7*B*).

All mutant viruses show similar growth on MDCK cells, however the plaque sizes on MDCK cells substantially differed from each other. All codon pair deoptimized viruses (72-Min, 46-Med and 27-Min) formed significantly smaller plaques compared to 27-WT (*Figure* 7*C and 7D*). Plaque sizes from 72H-Max and 72L-Max, which both have a high number of CpG dinucleotides, were undistinguishable from 27-WT.

To analyse NA protein production of the respective viruses, MDCK cells were infected with 5 MOI of recoded viruses, cells were lysed, and proteins in lysates were detected via Western blot analysis (*Figure* 7*E*). The viral genes NP and NS1 that were not recoded show similar protein levels in all mutant viruses as wells as the cellular control beta-actin. In contrast, the recoded NA genes show strongly reduced protein levels in codon pair deoptimized viruses (72-Min, 46-Med and 27-Min). The 72-Min showed the lowest, almost undetectable expression of NA. In contrast, viruses with an increased number of CpG dinucleotides (72H-Max and 72L-Max) show similar protein expressions of NA as 27-WT. In general, NA protein production correlates very well with growth kinetics of recoded viruses on A549 cells and their plaque sizes. This indicates that diminished NA production is responsible for reduced growth of codon pair deoptimized viruses (72-Min, 46-Med and 27-Min).

Altogether, these results indicate that not the number of CpG dinucleotides but the number or the nature of underrepresented codon pairs is responsible for virus attenuation in cell culture.

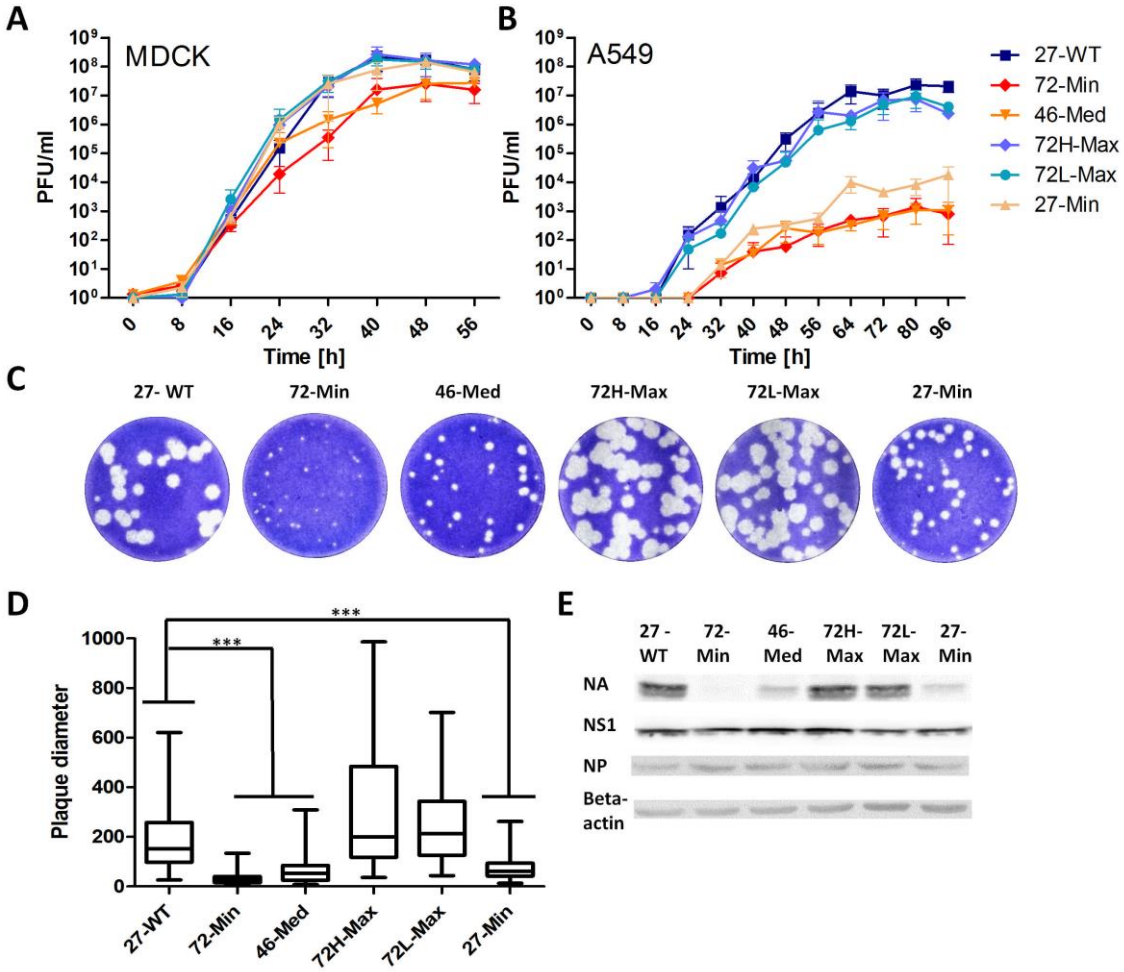

Figure 7: Biological properties of recoded viruses in cell culture. *Growth kinetics of 27-WT and recoded viruses on (A) MDCK and (B) A549 cells. Data show the means ± SD of three biological replicates. (C) Representative images of plaques on MDCK cells. (D) Relative plaque diameters of recoded viruses and 27-WT on MDCK cells. 72-Min, 46-Med and 27-Min produce significant smaller plaques than 27-WT. P-values (*** P<0.001) were calculated using one-way ANOVA Bonferroni's multiple comparison test. (D) Protein production of NA, NP and NS1 in MDCK cells 5 h after infection, detected by Western blot analysis. Beta-actin was used as cellular loading control.*

7.1.3 Pathogenicity of recoded viruses *in vivo*

Experiments in tissue culture showed that codon pair deoptimized viruses (72-Min, 46-Med and 27-Min) were severely attenuated, whereas 72H-Max and 72L-Max viruses grew with similar kinetics like 27-WT.

The pathogenicity of all mutant viruses was examined by experiments in mice. In this Part I will describe only the pathogenicity of viruses that have same codon bias as 27-WT (72-Min, 46-Med, 27-Min, 72L-Max and 72H-Max). The pathogenicity of CD viruses is described in the separate sections '*7.2.3 Pathogenicity of codon and codon pair deoptimized viruses in vivo*' and '*7.2.4 Vaccination experiment of codon and codon pair deoptimized viruses*', because they were designed by a different recoding strategy.

All the animal experiments were performed in a blinded manner to prevent experimental and operational bias. Eight groups with each nine six-week-old Balb/c mice were infected intranasally with 50,000 PFU of the respective virus or mock-control.

The body weight was recorded, as one health indicator, over the 18-day experiment (*Figure 8 and 9A*). At 2, 5 and 18 dpi (days post infection) three mice per group were euthanized and blood, lungs and spleens were taken for further analysis (*Figure 9A*). If mice showed severe clinical symptoms or lost more than 20 % of their body weight, they were immediately euthanized. All animals infected with 27-WT virus showed severe clinical signs and lost more than 20 % of their body weight between day four and six. Likewise, most animals from the groups 72L-Max and 72H-Max, which have a high CpG dinucleotide content, but no additional codon pair deoptimization, were euthanized within five to seven days, as they developed severe pneumonia. However, one mouse in the 72H-Max group and three in the 72L-Max group did not show drastically body weight loss compared to mice infected with 27-WT (*Figure 8*). In contrast, mice infected with codon pair deoptimized viruses 72-Min, 46-Med and 27-Min, as well as the mock control group did not show any clinical signs of an influenza infection.

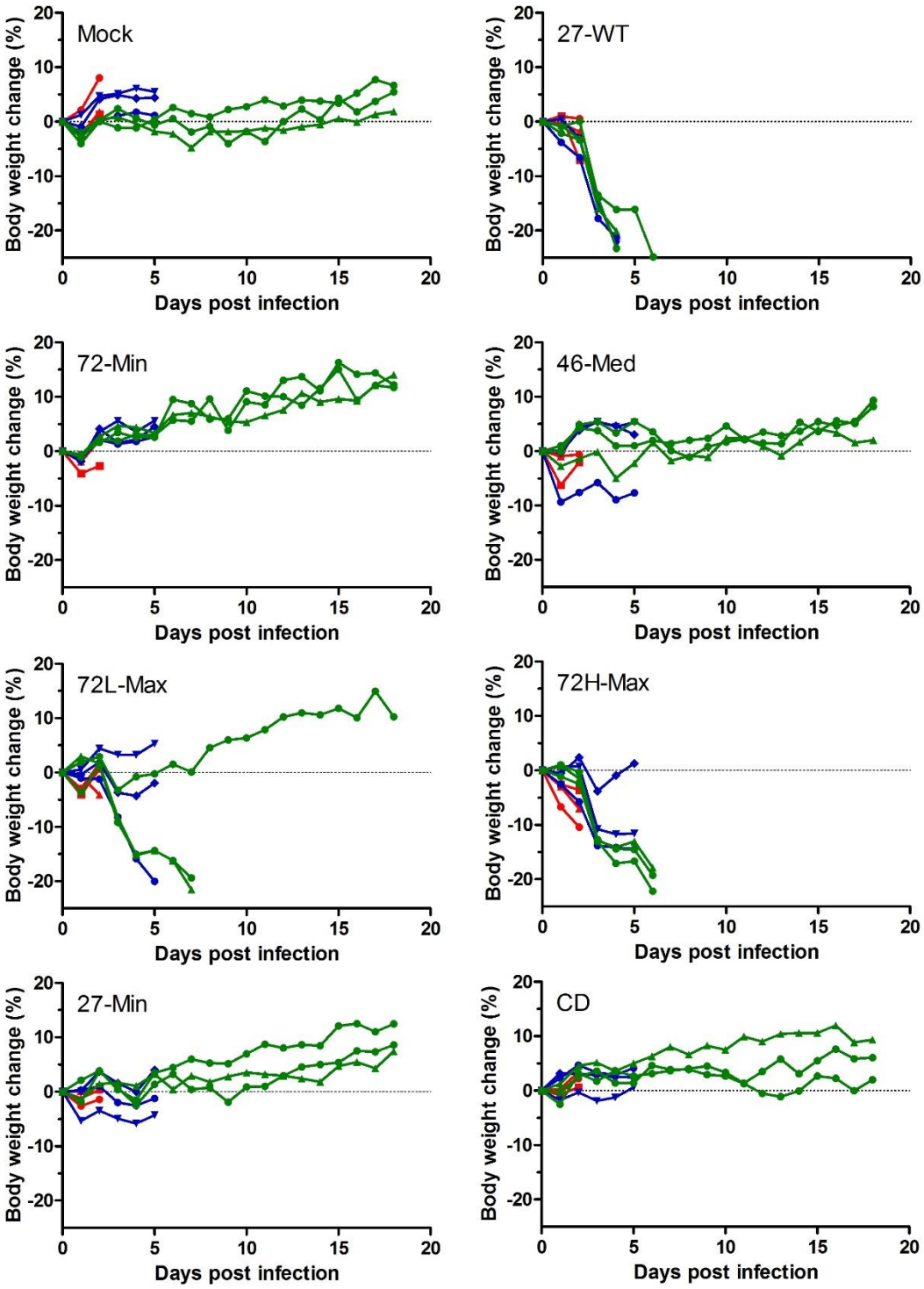

Figure 8: Body weight change of individual mice after IAV infection. *Body weight of individual mice after infection with different recoded influenza viruses or mock-control. On day 2 (red), 5 (blue) and 18 (green), three animals of each group were euthanized to analyse virus titres and the immune response in lungs and blood of mice.*

To compare the growth and pathogenesis of different mutant viruses, I determined the virus titres by plaque assays and RT-qPCR. Additionally, I compared the pathological changes in lungs of all mice. The 27-WT virus and the mutant viruses, 72L-Max and 72H-Max, replicated efficiently in the lungs and replicated to high titres (*Figure 9B*) and showed high copy numbers of the viral NP (*Figure 9C*) and the NA fragment (*Figure 9D*). In contrast, the virus titres in mice, infected with the codon pair deoptimized viruses were 10- to 1,000-fold lower than in the 27-WT infected group. The 72-Min group showed the lowest virulence in mice with no detectable infectious virus obtained at 5 dpi. On day 18, no infectious virus or viral RNA could be detected in any mouse. To evaluate the immune response in these mice, I quantified the antibody titres of IAV in the sera of mice. Samples from codon pair deoptimized mice (72-Min, 46-Med and 27-Min) and the single surviving mouse in the 72L-Max group showed high antibody titres in microneutralization (*Figure 9E*) and hemagglutination inhibition assay (*Figure 9F*), confirming a strong immune response. Additionally, spleen cells were re-stimulated with purified 27-WT virus on 18 dpi. The production of high levels of IFN-γ from spleen cells confirmed a robust antiviral T-cell response (*Figure 9G*).

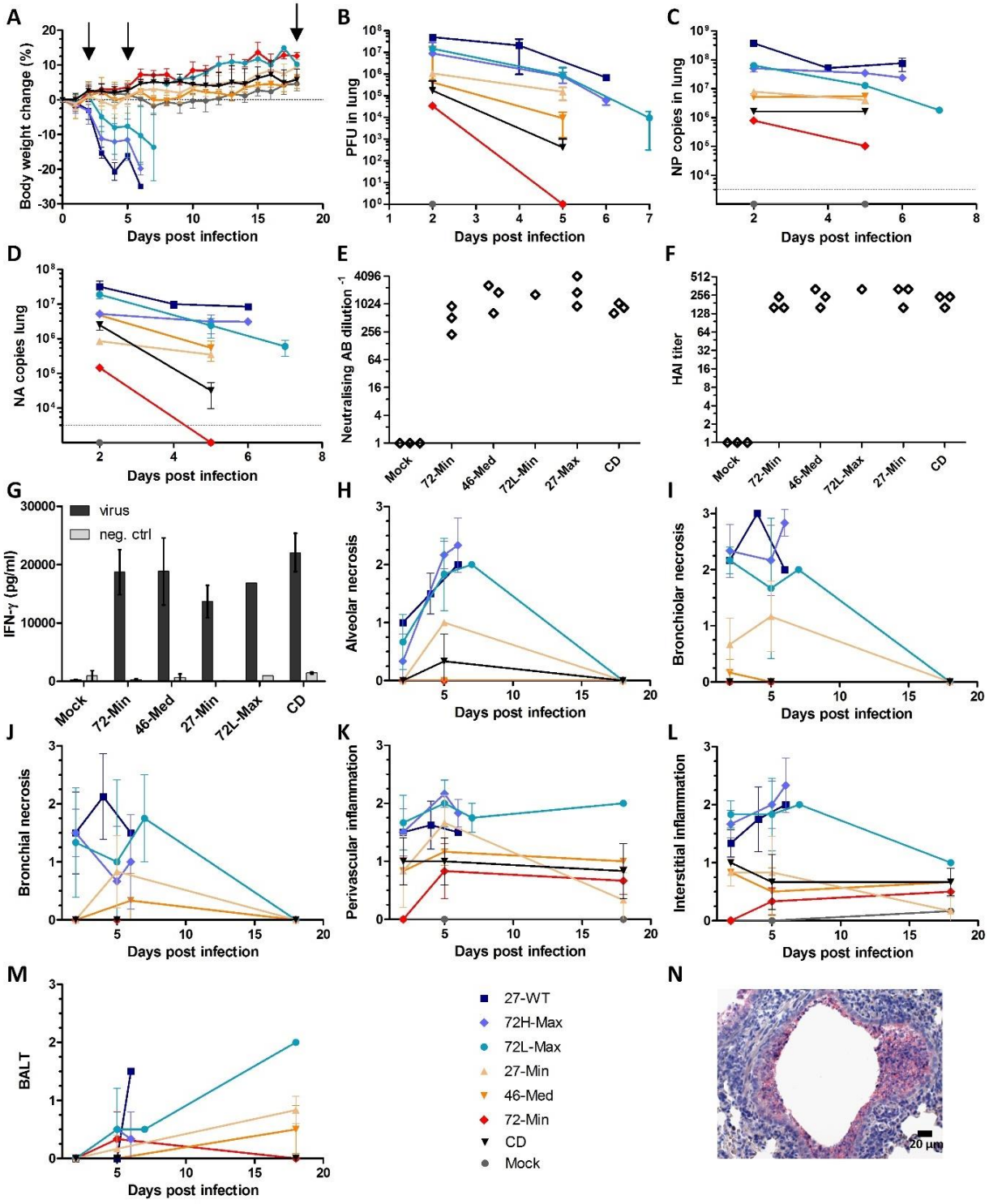

Figure 9: Immune responses in infected mice. *(A) Relative body weight of 6-week-old Balb/c mice after infection with different recoded influenza viruses. 2, 5 and 18 dpi, three animals of each group were euthanized (indicated by arrows) and blood and organs were analysed. Data show the mean ± SD. (B-D) The right lung lobes of mice were homogenized to quantify virus load in lungs. (B) Viral titres were determined by plaque assay and copy number of NP (C) and NA (D) segment was quantified by RT-qPCR. (E) Anti-IAV neutralizing antibody titres were measured 18 dpi from mice serum, as well as HAI titres (F). (G) Measurement of IFN-γ production. Mouse spleen cells were isolated 18 dpi, stimulated with purified IAV virus*

or control protein for 3 days and produced IFN-γ was quantified by ELISA. Data show the mean ± SD. (H-M) Histopathological analysis of mice lungs. The severity and distribution of alveolar necrosis (H), bronchiolar necrosis (I), bronchial necrosis (J), perivascular inflammation (K), interstitial inflammation (L) and bronchus-associated lymphoid tissue (BALT) (M) in lungs of infected animals were assessed and scored from 0 (no lesion) to 3 (severe lesions). Data show the mean ± SD. (N) Detection of IAV antigens in the necrotic bronchiolar epithelium of infected mice (red) confirmed by immunohistochemical staining.

Next, the level of necrosis and inflammation in mice lungs was quantified histologically (*Figure 10*). Mice infected with 27-WT, 72H-Max and 72L-Max showed pneumonia with strong necrosis in alveoli (*Figure 9H*), bronchioles (*Figure 9I*) and bronchi (*Figure 9J*) as well as inflammation in perivascular (*Figure 9K*) and interstitial spaces (*Figure 9L*). Bronchus-associated lymphoid tissue (BALT) was detected in 27-WT and in 72L-Max mice (*Figure 9M*). Mice infected with codon pair deoptimized viruses 72-Min, 46-Med and 27-Min showed none to only moderate inflammation and necrosis. The 72-Min virus was the most attenuated virus causing no signs of a respiratory disease (*Figure 9H-M*). The widely distributed IAV antigen in mice lungs detected by immunohistochemical staining provided an additional confirmation that IAV infection caused the pneumonia (*Figure 9N*).

Examination of mouse blood, organs and observed symptoms indicated that codon pair deoptimized viruses (72-Min, 46-Med and 27-Min) induced strong immune response in mice without causing any clinical symptoms whereas 27-WT, 72L-Max and 72H-Max caused severe pneumonia.

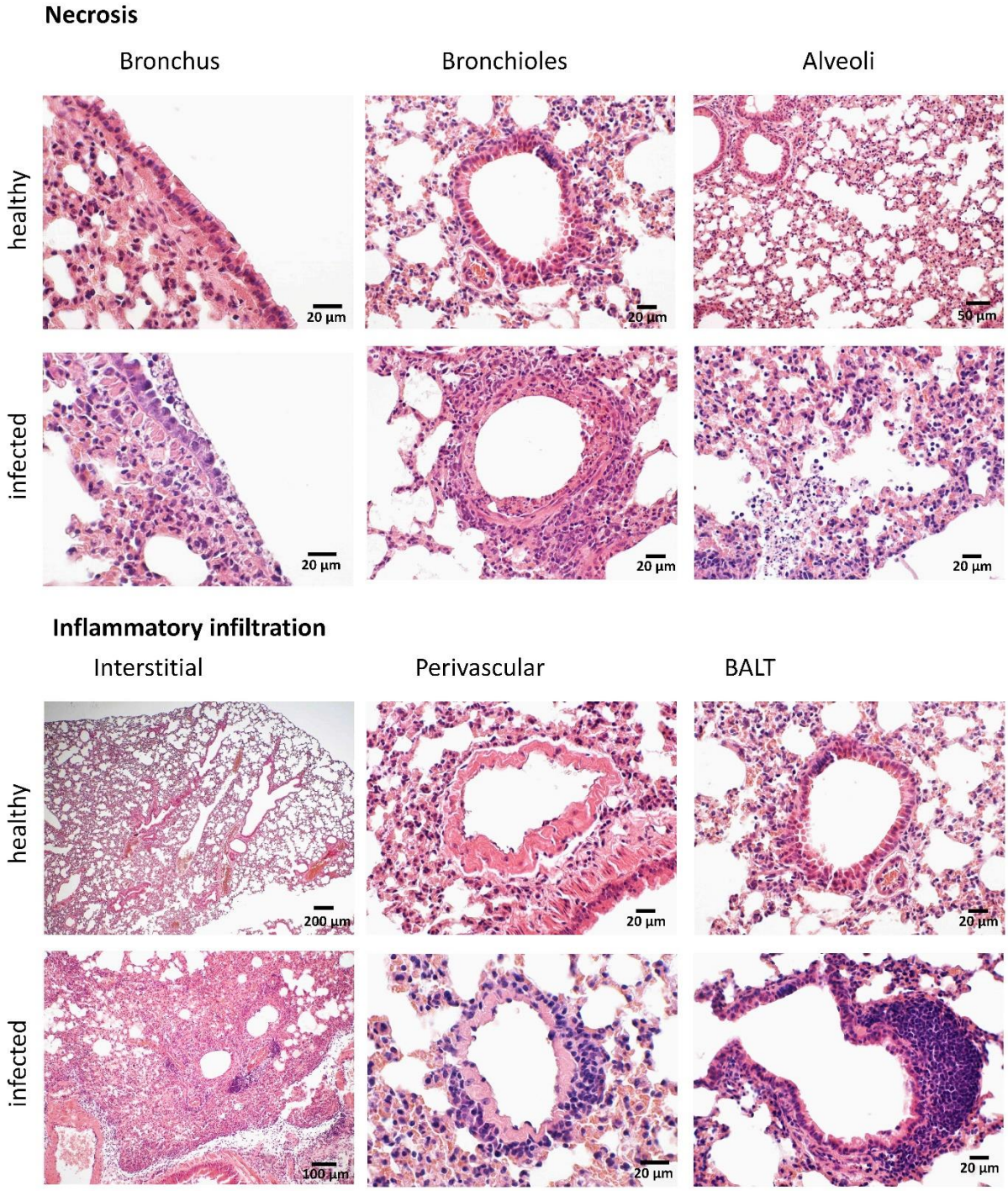

Figure 10: Representative images of histopathological changes that were assessed in the lungs of infected mice. *Exemplary images of infected and uninfected mice lung sections are shown to depict the different necrosis and inflammation parameter that were used to score the lungs.*

7.1.4 Vaccination experiment

To evaluate the protective immunity induced by mutant viruses against 27-WT challenge, a vaccination experiment was performed. Eight six-week-old mice were vaccinated with 50,000 PFU of attenuated viruses that did not cause any clinical symptoms in mice in the first experiment (72-Min, 46-Med and 27-Min) or were mock vaccinated. After 28 days, mice were challenged with 500,000 PFU of 27-WT virus (equivalent to 10 × LD100). All mice immunized with attenuated viruses, showed a comprehensive protection against 27-WT virus challenge and did not show any body weight loss (*Figure 11A*). High antibody titres against IAV could be detected 18 or 28 dpi in serum, which increased only slightly after challenge (*Figure 11B*). At day 10 after challenge, virus was cleared in vaccinated mice and no viral copies could be detected by RT-qPCR in lungs. Nevertheless, 10 days after challenge, spleen cells from mice produced high levels of IFN-γ after stimulation with 27-WT virus, indicating the strong adaptive immune response (*Figure 11C*). No necrosis (*Figure 11D-F*) and only mild inflammation (*Figure 11G-I*) was detected in lungs from these mice vaccinated with codon pair deoptimized viruses (72-Min, 46-Med and 27-Min).

The mice of the mock control group quickly lost weight (*Figure 11A*) and showed mild to severe inflammation (*Figure 11G-I*) and necrosis (*Figure 11D-F*). Only two mice from the mock group recovered after notable loss of body weight. Survival of these two mice can be explained with age dependent immunity, as mice were older in this experiment at the time point of WT-challenge (all mice succumbed to the WT-infection in the first animal experiment). However, the lungs of these two mice were severely damaged and showed significant degree of repair (pneumocyte type II hyperplasia) (*Figure 11J and K*). Unvaccinated mice showed 10^7 to 10^8 NA and NP copies three to four days after challenge and around 10^4 NP copies in their lungs ten days after challenge (*Figure 11L*).

All analysed codon pair deoptimized viruses (72-Min, 46-Med and 27-Min), analysed in this study, are excellent, live-attenuated vaccine candidates and provide a robust protection against 27-WT challenge through a strong adaptive immune response.

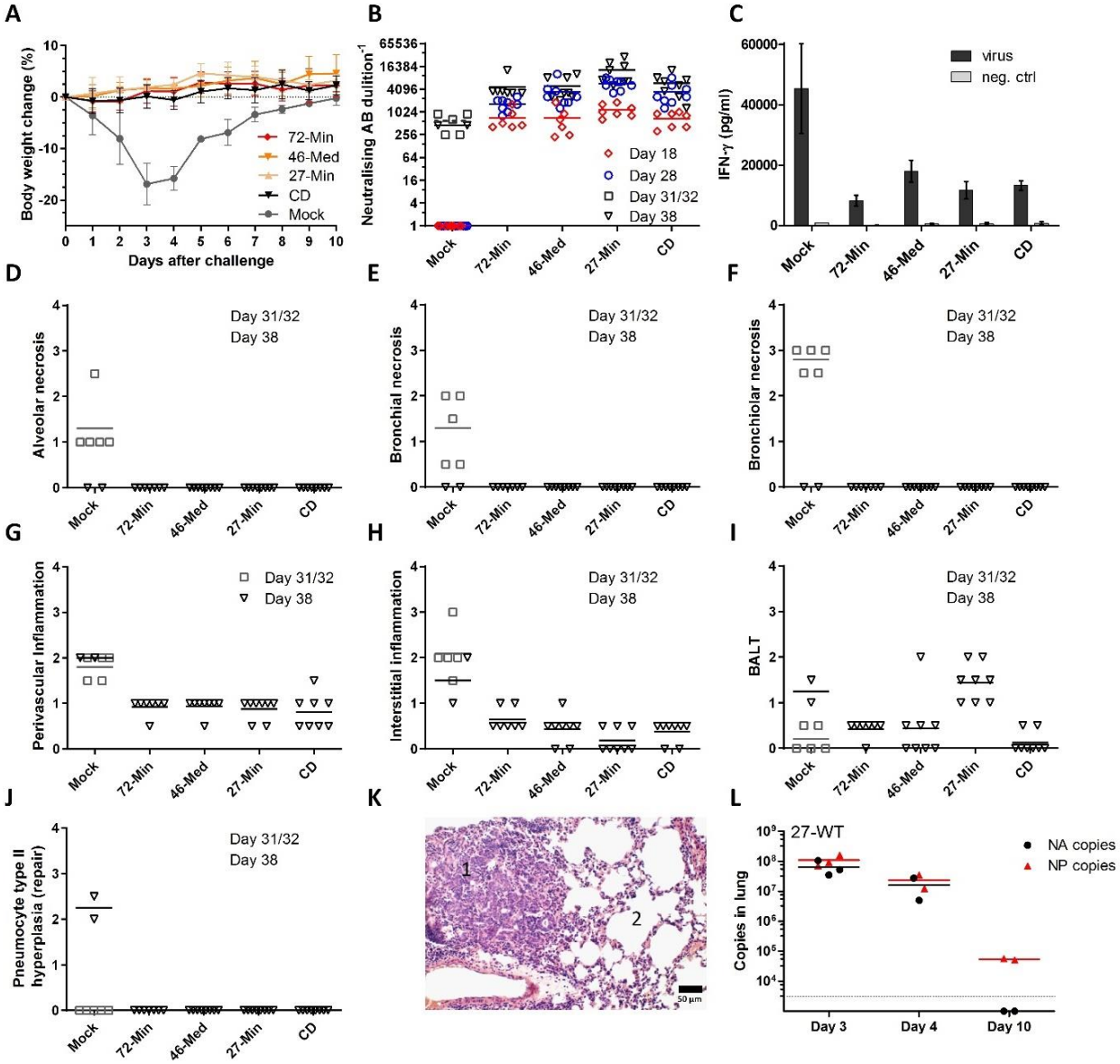

Figure 11: Immune protection of vaccinated mice. *(A) Relative body weight of vaccinated Balb/c mice after challenge with 500,000 PFU of 27-WT IAV. Mice were vaccinated with indicated live-attenuated influenza viruses. Data are shown as mean ± SD. (B) Anti-IAV neutralizing antibody titres from mice serum, 18, 28 (day of challenge) and 38 days after vaccination. (C) IFN-γ production of mouse spleen cells. Spleen cells were isolated 38 days after vaccination (10 days after challenge) and stimulated with purified IAV virus or control protein for 3 days. Produced IFN-γ was quantified by ELISA. Mock control shows the two unvaccinated mice, which survived the 27-WT challenge. Data show the mean ± SD. (D-J) Histopathological analysis of mice lungs. The severity and distribution of alveolar necrosis (D), bronchial necrosis (E), bronchiolar necrosis (F), perivascular inflammation (G), interstitial inflammation (H), bronchus-associated lymphoid tissue (BALT) (I) and pneumocyte type II hyperplasia (J) in lungs of infected animals were assessed and scored from 0 (no lesion) to 3 (severe lesions). Data show the mean ± SD. (K) Histopathological image of pneumocyte type II hyperplasia (repair). The image shows on the left side affected tissue (1) and on the right side unaffected tissue (2). (L) Copy numbers of NA and NP segment of mock-vaccinated mice after challenge measured by RT-qPCR.*

7.1.5 Molecular mechanism of attenuation by codon pair deoptimization

Codon pair deoptimized genes produce less protein than the corresponding non-recoded genes[65]. Viruses, containing such codon pair deoptimized genes, are attenuated. However, it is still unknown, which sequence features and which steps in the transcription or translation process cause the reduced protein production. Here, I analysed the mechanism underlying the reduced protein production in codon pair deoptimized genes, to gain insights in the molecular basis of codon pair deoptimization.

The recoded NA genes were cloned into a dual expression plasmid, expressing an eGFP as control protein. NA and the eGFP expression are driven from two different independent promoters. The respective plasmids were transiently transfected in HEK 293T cells, and RNA kinetics and protein levels were measured in cell lysates 24 h post transfection.

The RNA expression levels of recoded NA genes were detected by RT-qPCR and normalized to eGFP expression. The NA mRNA levels were significantly reduced in codon pair deoptimized genes 72-Min, 46-Med and 27-Min (*Figure 12A*). In contrast, mRNA level of the constructs having a high number of CpG dinucleotides show only slight reduction (72L-Max) or even an increase (72H-Max).

The relative protein expression levels of recoded NA genes in transfected cells from plasmids (*Figure 12B*) were comparable to NA expression of recoded viruses in infected cells (*Figure* 7*E*). Western blot analysis revealed that codon pair deoptimized genes (72-Min, 46-Med and 27-Min) produced less NA protein, whereas 72H-Max and 72L-Max showed comparable NA expression to 27-WT.

Knowing the steady state levels of NA RNAs and protein, I compared the RNA synthesis and degradation rates. To analyse synthesis and degradation, newly synthesized RNA was labelled with a ribonucleotide analog BrU. Additionally, for RNA degradation measurement, BrU labelling was followed by addition of an excess of uridine (pulse-chase labelling) to stop labelling. BrU labelled RNA was immunoprecipitated and quantified by RT-qPCR. The RNA synthesis rates of was similar in all recoded NA genes (*Figure 12C*). Surprisingly, RNA degradation rates showed notable differences between recoded NA genes. The NA mRNA produced by 72H-Max and 72L-Max, which both have a high number of CpG dinucleotides, are relatively stable, similarly to the mRNA of 27-WT (*Figure 12D*). All codon pair deoptimized NA genes (72-Min, 46-Med and 27-Min) produced RNA that rapidly degraded within two hours.

To confirm that codon pair deoptimized genes produce unstable RNA transcripts, a second assay was performed to measure RNA stability. HEK 293T cells were transiently transfected with the dual expression plasmids containing different recoded NA genes. At 24 h post transfection, RNA synthesis was stopped with 1 µM of the transcription inhibitor flavopiridol

and degradation of NA RNA was determined over time. The 72-Min, 46-Med and 27-Min showed again rapid RNA degradation, while 27-WT and genes with a high CpG content (72H-Max and 72L-Max) produced more stable RNA (*Figure 12E*). This assay confirms that mRNAs produced by codon pair deoptimized genes are faster degraded compared the non-recoded gene.

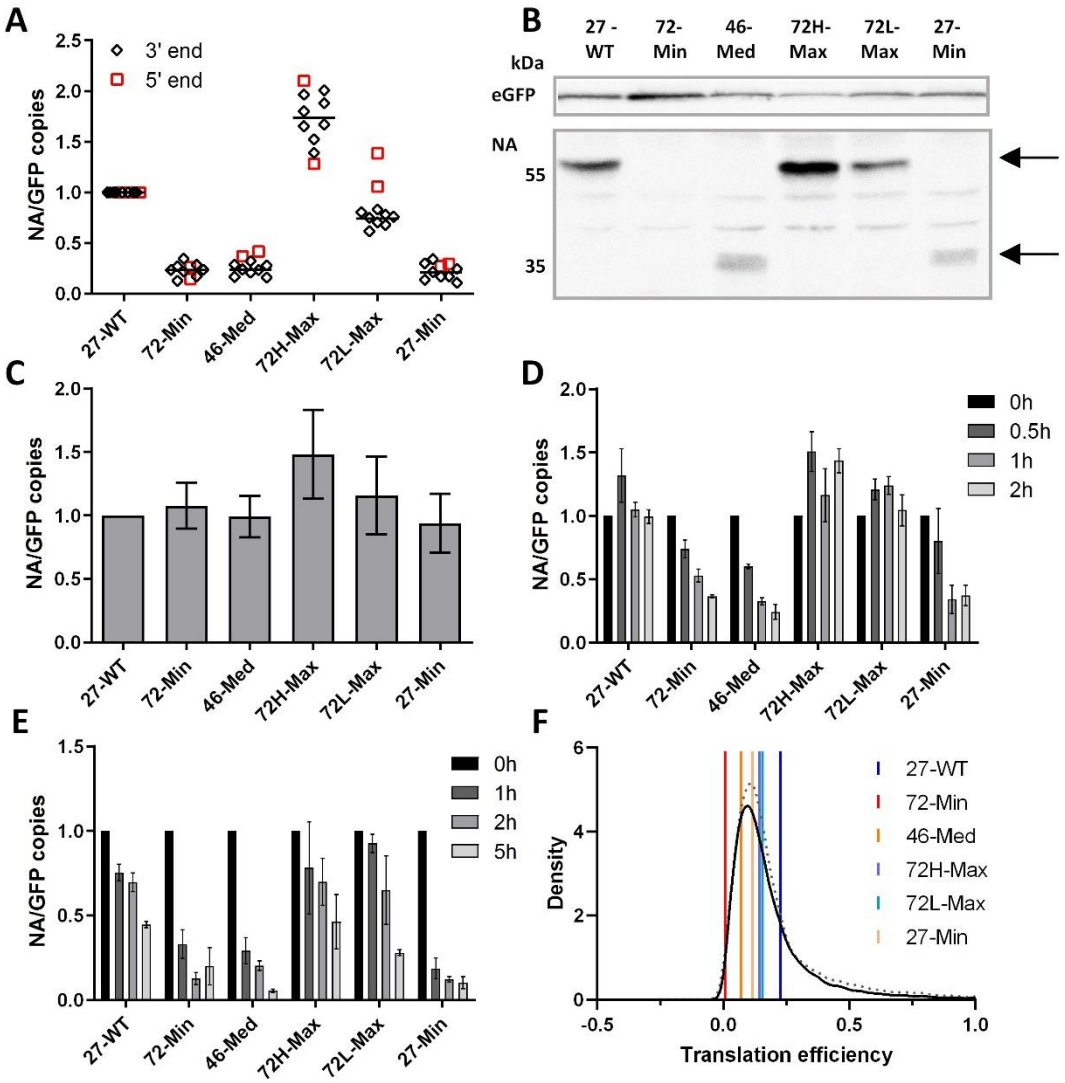

Figure 12: RNA kinetics and protein production of recoded NA genes. *(A-F) HEK 293T cells were transiently transfected with dual expression plasmid containing recoded NA genes and eGFP. (A) Steady state RNA level of recoded NA genes in transiently transfected cells. RT-qPCR measurements were performed with a probe at the 3' end (black) or at the 5' end (red) of the NA genes. EGFP was used for normalization. (B) Western blot analysis of NA protein from recoded genes. EGFP was used for normalization. (C) Transcription rates of NA genes determined by RT-qPCR of BrU labelled RNA. (D-E) RNA degradation rates, (D) measured by BrU labelled RNAs or (E) by transcription inhibition with flavopiridol. Data are the means of three biological replicates ± SD. (F) Translation efficiency of 29 nt footprints of recoded NA genes from ribosome profiling experiments. Translation efficiency of all cellular genes of HEK 293T cells are given as a density plot with a dotted line as SD for different experiments.*

After I showed that codon pair deoptimization leads to decreased RNA stability (*Figure 12D-E*), I evaluated if the translation efficiency is additionally involved in the mechanism that leads to reduced protein production. Therefore, I performed ribosome profiling experiments and analysed the positions and frequency of ribosome protected mRNA fragments (footprints). Translation initiation rates are the rate-limiting step in the synthesis of proteins[77]. Due to the construction of the NA genes, which all have identical 5' and 3' ends, the impact of any terminal sequence motif or secondary structure that influences the translation initiation could be excluded. 72H-Max and 72L-Max show only slightly reduced translation efficiency. However, I detected a strong reduction in the translation efficiency of the codon pair deoptimized NA genes (72-Min, 46-Med and 27-Min) (*Figure 12F*), additionally to the reduced mRNA stability (*Figure 12D-E*). Consistent with the highest attenuation in a virus background (*Figure* 7), 72-Min showed the lowest translation efficiency, with very little detected footprints over the whole gene (*Figure 13*). In the 46-Med and 27-Min almost no footprints could be detected beyond 760 nucleotides, indicating an abrupt stop of translation at this position (*Figure 13*). Total RNA reads were detected over the whole gene, excluding truncated mRNAs (*Figure* 13). Additionally, I confirmed the presence of the full-length mRNA transcript with RT-qPCR primers and probes at the 5' and the 3' termini. The termination of translation that was detected by ribosome footprints (46-Med and 27-Min) could be also detected as a truncated protein fragment by Western blot analysis (*Figure 12B*). This protein fragment was analysed by mass spectrometry confirming the premature translation stop by identification of the 240 N-terminal amino acids.

My results show that codon pair deoptimization leads to reduced protein levels due to rapid RNA degradation and reduced translation efficiency. An increase of CpG dinucleotides instead, did influence neither the protein production nor the RNA stability.

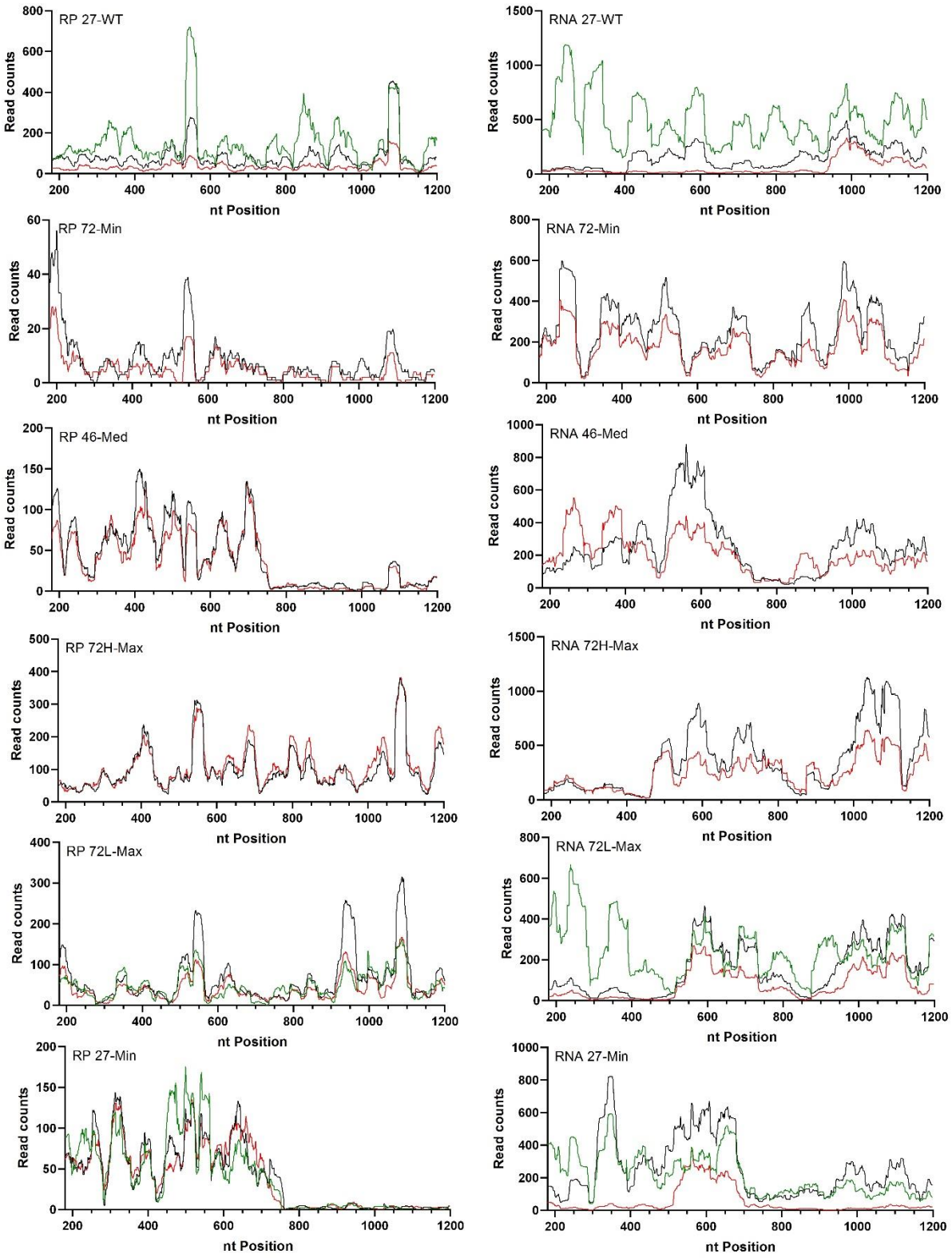

Figure 13: Read density of ribosome footprints and total RNA of recoded NA genes. *Individual replicates of the read density by next generation sequencing in the recoded part of the NA gene are shown in red, black and green.*

7.1.6 Summary obtained results Part I

The obtained results from Part I are graphically summarized in *Figure 14*. I showed that codon pair deoptimized viruses (72-Min, 46-Med and 27-Min) have strong growth deficits in cell culture, whereas viruses with a high number of CpG dinucleotides (72H-Min and 72L-Min) behave like 27-WT. In addition, the codon pair deoptimized viruses are – in contrast to the viruses with a high CpG content – attenuated in mice. Despite strong attenuation, the codon pair deoptimized viruses provide a strong adaptive immune response in mice and are therefore suitable vaccine candidates. Additionally, I showed that the reduced protein production that leads to attenuation in codon pair deoptimized viruses is caused by increased mRNA decay and reduced translation efficiency.

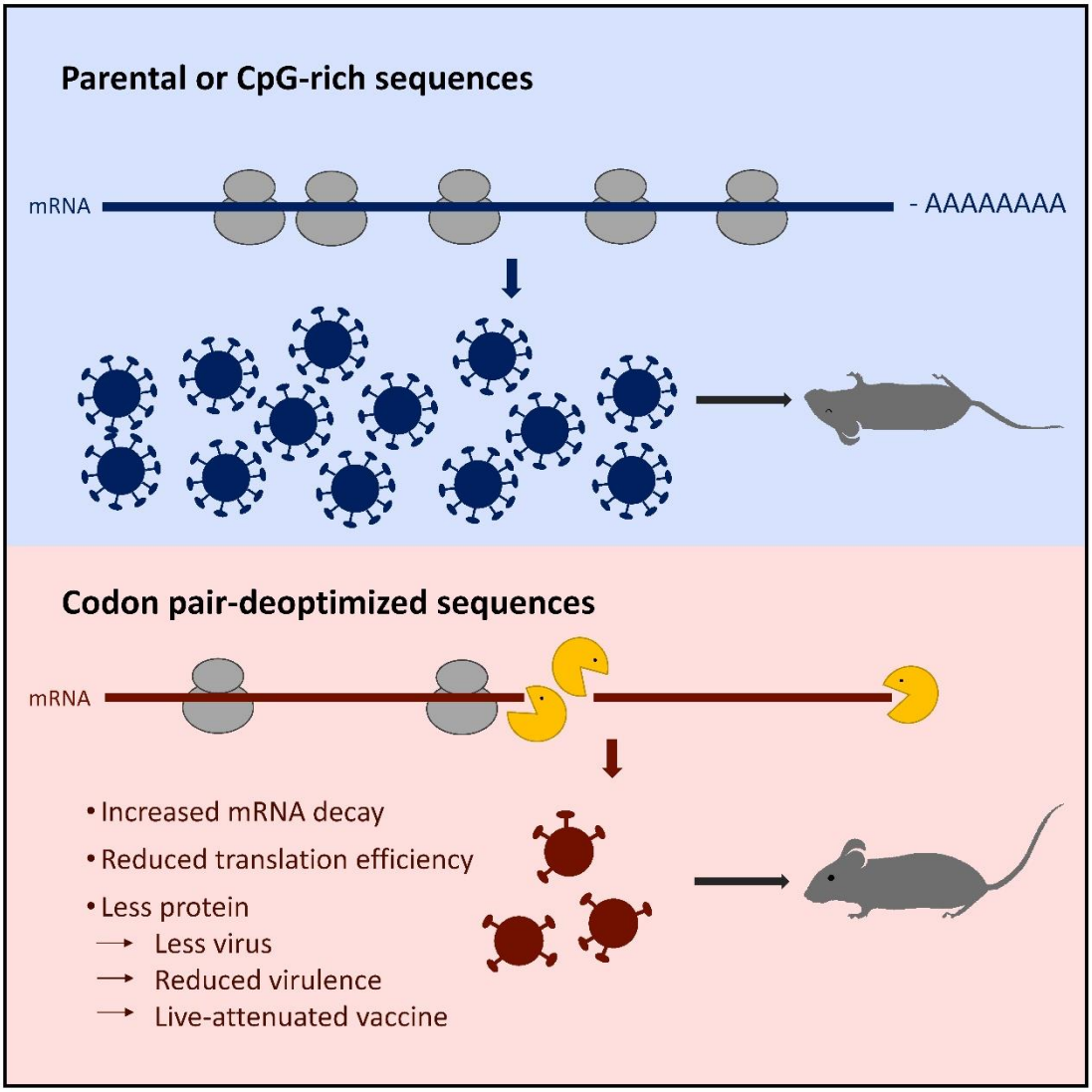

Figure 14: Graphical summary of Part I. *The reasons for attenuation of codon pair deoptimized viruses are increased mRNA decay and reduced translation efficiency. The attenuated viruses provide good vaccine candidates. Viruses with high numbers of CpG dinucleotides show the same properties as WT virus.*

7.2 Part II: Codon versus codon pair optimization/deoptimization – two tools with great potential

7.2.1 Virus recoding (Part II)

The second main goal of this project was to compare the efficiency of codon and codon pair optimization and deoptimization as a recoding strategy based on the NA gene of IAV. Similar to the recoding in Part I (*7.1.1 Design and recoding of viruses*), the middle part of the NA gene from the IAV strain WSN was recoded (*Figure 6A*). The first 183 nucleotides and the last 162 nucleotides of the NA segment were omitted from recoding to ensure efficient packaging of viral RNA into a virion and to preserve the same level of translation initiation from different NA constructs. The NA gene was recoded with two different recoding strategies, however the amino acid sequence of all recoded NA genes remained unchanged. The first method is **codon pair** optimization or deoptimization. As described in section *'5.8 Codon pair deoptimization'*, with this strategy the recoded genes contain the same codons as the parental sequences, but the existing codons are reshuffled to create codon pairs that are overrepresented or underrepresented in the human host. The second recoding method is **codon** optimization or deoptimization. Here, the target sequence is recoded only with synonymous codons that are the most or least frequently used in the host species.

The RNA and protein levels of four different NA genes were compared. The genes CD, CO, CPD and CPO were designed by codon deoptimization/optimization, or codon pair deoptimization/optimization, respectively (*Table 3*). The parental WT and the codon pair deoptimized genes are identical with the NA genes that were studied in Part I. For the sake of simplicity and clarity, the 27-WT NA gene is henceforth called WT, and the codon pair deoptimized 72-Min NA gene CPD.

Table 3. Properties of recoded IAV H1N1 strain A/WSN/1933 NA genes Part II.

Name	Description	CPS	CAI	Codon identity (%)	Nt identity (%)	CpG	UpA
WT	Original sequence	0.00	0.74	100.0	100.0	27	54
CPD	Codon pair deoptimized, identical to 72-Min	-0.45	0.74	37.8	75.6	72	100
CD	Codon deoptimized	-0.097	0.46	36.3	74.5	108	99
CPO	Codon pair optimized	0.286	0.74	46.0	78.6	12	45
CO	Codon optimized	0.047	1.00	33.9	72.6	73	11

CPS – average codon pair score; CAI – codon adaptation index (a measure of codon bias); Codon/Nucleotide identity - percentage of codons/nucleotides that occupy the same position in WT and recoded genes; CpG – number of CpG dinucleotides in recoded sequences; UpA - number of UpA dinucleotides in recoded sequences

7.2.2 Characterization of recoded viruses in cell culture (Part II)

I recovered the four recoded viruses with optimized or deoptimized NA genes and compared their growth abilities to WT. In MDCK cells, codon deoptimized (CD) and codon pair deoptimized (CPD) viruses showed significant smaller plaque sizes, with the size of only around 10 % of the WT virus (*Figure 15A & B*). Nevertheless, the deoptimized viruses CPD and CD showed only slightly reduced growth in MDCK cells with 10- to 100-fold lower titres than the WT (*Figure 15C*), which grew up to 10^8 PFU/ml. In human lung A549 cells, the WT grew to titres of up to 2×10^7 PFU/ml. In A549 CPD and CD virus mutants had a clear growth disadvantage (*Figure 15D*). Their growth was 24 h delayed and the maximum titres that were detected were only 2×10^2 PFU/ml, which is 100,000 lower than WT. Overall, CPD and CD show a very similar growth kinetics, they have a major growth defect and are highly attenuated compared to WT.

Codon and codon pair optimized viruses (CO and CPO) in contrast showed comparable plaque sizes to WT (*Figure 15A and B*). Similarly, the growth of CPO on MDCK and A549 cells wassimilar to WT (*Figure 15C and D*). The CO also displayed similar plaque sizes as WT (*Figure 15A and B*), but grew to 10- to 100-fold times lower titres in MDCK cells (*Figure 15C*) and even 1000-fold lower titres in A549 cells (*Figure 15D*). These results are surprising, because I expected that the codon optimized virus to grow better than the parental virus.

Next, I compared the NA protein levels of different mutant viruses. MDCK cells were infected with MOI of 5 and 5 h after infection protein expression was evaluated by Western blot analysis.

The NA protein levels correlate well with the growth abilities on A549 cells. As expected, NA-protein levels were highly reduced in CPD and CD viruses. The CPO virus showed a similar to increased NA production, whereas codon optimized virus CO showed a very low expression of NA, comparable to its poor growth capabilities. The non-recoded viral proteins NP, NS1 and beta-actin as a cellular control, showed similar protein production in cells infected by different viruses (*Figure 15E*).

Overall, the deoptimized viruses CPD and CD have a strongly reduced NA production compared to WT, they showed hampered growth and formed smaller plaques. Both viruses showed a similar attenuation level and no differences in the fitness of the viruses can be detected. CPO shows similar behaviour in virus fitness as WT. Unexpectedly, even though CO has similar plaque sizes as WT, the growth and NA protein production of the codon optimized virus is severely reduced.

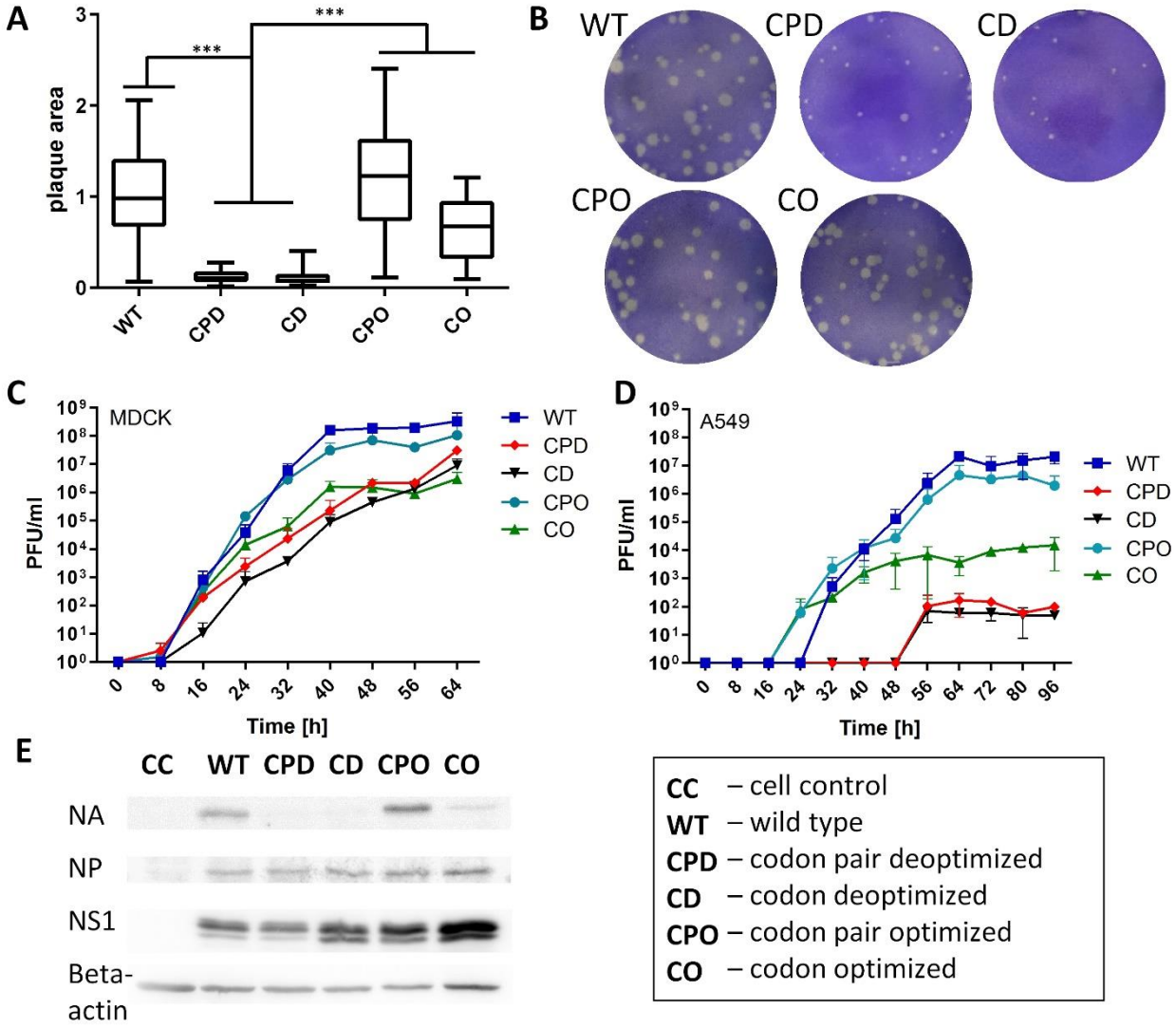

Figure 15: Biological properties of recoded viruses in cell culture (Part II). *(A) Relative plaque diameter of WT and recoded viruses. CPD and CD produce significant smaller plaques than WT, CPO and CO. P-values (*** P<0.001) were calculated using one-way ANOVA Bonferroni's multiple comparison test. (B) Representative images of plaques. (C-D) Growth kinetics on (C) MDCK and (D) A549 cells. Data show the means ± SD. (E) Protein production detected by Western blot analysis. Viral NA, NP and NS1 and cellular protein beta-actin were analysed 5 h after infection of MDCK cells.*

7.2.3 Pathogenicity of codon and codon pair deoptimized viruses *in vivo*

I determined the pathogenesis of the codon (CD) and codon pair deoptimized viruses (CPD/72-Min) *in vivo* and their potential use as vaccines in a mouse model. The mouse experiments were performed together with the mutant viruses from Part I of my work to reduce the number of control animals and to reuse the results from the codon pair deoptimized virus 72-Min now called CPD/72-Min. The pathogenic potential of optimized viruses (CPO, CO) was

not assessed in this study in the mouse model, because they are not designed to be potential live-attenuated vaccine candidates. In the first experiment, the virulence of the two differently deoptimized viruses was compared. Nine mice were infected with 50,000 PFU of WT, CD or CPD/72-Min virus or a medium control (mock). On days 2, 5 and 18 three mice were euthanized and lungs, spleens and blood were collected for further analysis.

After infection, the pathogenicity of mutant viruses was determined by monitoring clinical signs and daily weights of infected mice (*Figure 8*). All WT-infected mice lost significant amount of weight between day 2 and 6 post infection. CPD/72-Min, CD and mock-infected mice did not lose any weight and did not show any clinical signs associated with an influenza infection (*Figure 9A*). The virus titres in lungs were determined by plaque assays (*Figure 9B*). WT-infected mice had high virus titres, up to 10^8 PFU, in the right lobe of the lung between days 2 and 6. On day 2, CPD/72-Min and CD showed substantial lower titres of 3×10^4 PFU and 2×10^5 PFU in the right lung, respectively. On day five, all mice in CPD (72-Min) and two out of three mice in CD cleared the virus. Additionally, copy numbers of NP (*Figure 9C*) and NA segment (*Figure 9D*) were determined in mouse lungs by RT-qPCR. In WT-infected mice, the high virus titre (*Figure 9B*) correlated to the high copy number of NP and NA segment in lungs (*Figure 9C*). The lungs of CD- and CPD/72-Min-infected mice showed 100-1,000-fold lower copy numbers of NP and NA compared to WT-infected mice. In CPD/72-Min-infected mice, no NA copies could be detected on day 5. On day 18, no virus could be detected in any mouse by plaque assay nor by RT-qPCR. Overall, mice infected with CPD/72-Min virus have slightly lower NA and NP copy numbers and lower virus titres in the lung than mice infected with CD.

To evaluate the immune response, HAI titres (*Figure 9F*) and neutralising antibody titres (*Figure 9E*) against IVA were determined 18 days post infection. I detected a high antibody titre in serum of mice that were infected with CPD/72-Min and CD viruses. Additionally, INF-γ production of isolated spleen cells was quantified after the stimulation with purified IAV to analyse the antiviral T-cell response. Spleen cells isolated 18 days post infection with CPD/72-Min or CD, produced high levels of INF- γ after stimulation with virus antigen (*Figure 9G*).

Finally, mice lungs were evaluated according to their necrosis and inflammation. Over the whole duration of the experiment, no necrosis in lungs from mice infected with CPD/72-Min virus was detected and only very low alveolar necrosis in CD-infected mice lungs (*Figure 9H-J*). Mild inflammation was present in CPD/72-Min infected mice on day 5 and 18 and in CD-infected mice additionally on day 2 (*Figure 9K-M*). On the contrary, WT-infected mice show severe alveolar, bronchial and bronchiolar necrosis, as well as acute interstitial and perivascular inflammation (*Figure 9H-L*).

In summary, the codon and codon pair deoptimized viruses caused no clinical symptoms in mice but induced high adaptive immune responses against IAV.

7.2.4 Vaccination experiment of codon and codon pair deoptimized viruses

In a second animal experiment, eight mice were vaccinated with 50,000 PFU of CPD/72-Min or CD virus. After 28 days, vaccinated mice and a mock control group were challenged with 500,000 PFU WT virus. All mice in the mock-control group, inoculated with medium, lost rapidly weight within the first 3 days (*Figure 11A*). Only two recovered after strong weight loss. The CPD/72-Min- and CD-vaccinated mice neither lost any body weight nor showed any disease symptoms. However, virus infection induced high titres of neutralising antibodies in serums, 18 and 28 days after vaccination (*Figure 11B*). The antibody titres increased slightly, 10 days post WT challenge. Spleen cells of vaccinated mice showed high expression of INF-γ after stimulation with WT virus (*Figure 11C*). These assays confirmed the strong immune response of mice vaccinated with the live-attenuated CPD/72-Min and CD viruses. The pathological examination of lungs from vaccinated mice showed no necrosis (*Figure 11D-F*) and very mild inflammation after WT challenge (*Figure 11G-I*), however high necrosis and inflammation in non-vaccinated control mice.

Taken together, these results show that both methods, codon deoptimization and codon pair deoptimization provided excellent vaccine candidates (CD, CPD/72-Min). They have the ability to provide a strong immune response and to protect against WT challenge.

7.2.5 Protein expression and RNA stability of codon and codon pair (de)optimized genes in absent of viral background

The CO virus is, despite codon optimization, attenuated in cell culture. To determine the reason for attenuation, I investigated the effect of RNA levels, amount of protein expression and the translation efficiency of recoded genes, in the absent of a viral background. I studied these genes *in vitro*, by utilizing dual-expression plasmids that express both, the recoded NA gene and the eGFP reporter.

CD and CPD NA genes show very low steady-state mRNA level as detected by RT-qPCR (*Figure 16A*). To investigate if this low observed RNA levels are caused by reduced RNA stability, RNA degradation rates were assessed. Therefore, RNA synthesis was inhibited with RNA polymerase II inhibitor flavopiridol and the half-lives of mRNA were followed. Faster turnover of mRNA was detected in CD- and CPD-NA genes (*Figure 16B*). Consistent with low RNA levels, the NA protein of CPD and CD were low, below detection limit of Western blot

analysis (*Figure 16C*). As recoding did not affect other genes, I observed high expression of eGFP control protein (*Figure 16D*). An addition to the reduced RNA stability, very inefficient translation was observed as determined with ribosome profiling experiments (*Figure 16E*). These findings strongly suggests that attenuation of codon and codon pair deoptimized viruses (*Figure 15*) are caused by reduced RNA stability and translation efficiency.

The codon optimized (CO) and codon pair optimized (CPO) NA gene produced increased RNA levels (*Figure 16A*) and showed higher protein expression (*Figure 16C*) than the WT NA gene. CO, consisting of only codons that are used most frequently in the human host, expresses even higher RNA and protein levels compared to CPO. In contrast, to all other NA genes, the CO gene showed different characteristics on a plasmid level compared to viral level. The CO gene on virus level showed reduced growth abilities and protein expression (*Figure 15*).

The RNA stability of CO and CPO optimized genes are similar to WT (*Figure 16B*). The translation efficiency of CPO is equally good as in WT, but surprisingly, despite high protein expression, reduced in CO (*Figure 16E*). Suggesting that high protein expression is a result of high RNA levels.

Overall, codon and codon pair optimization are both highly valuable tools to increase the protein expression of a target gene by recoding.

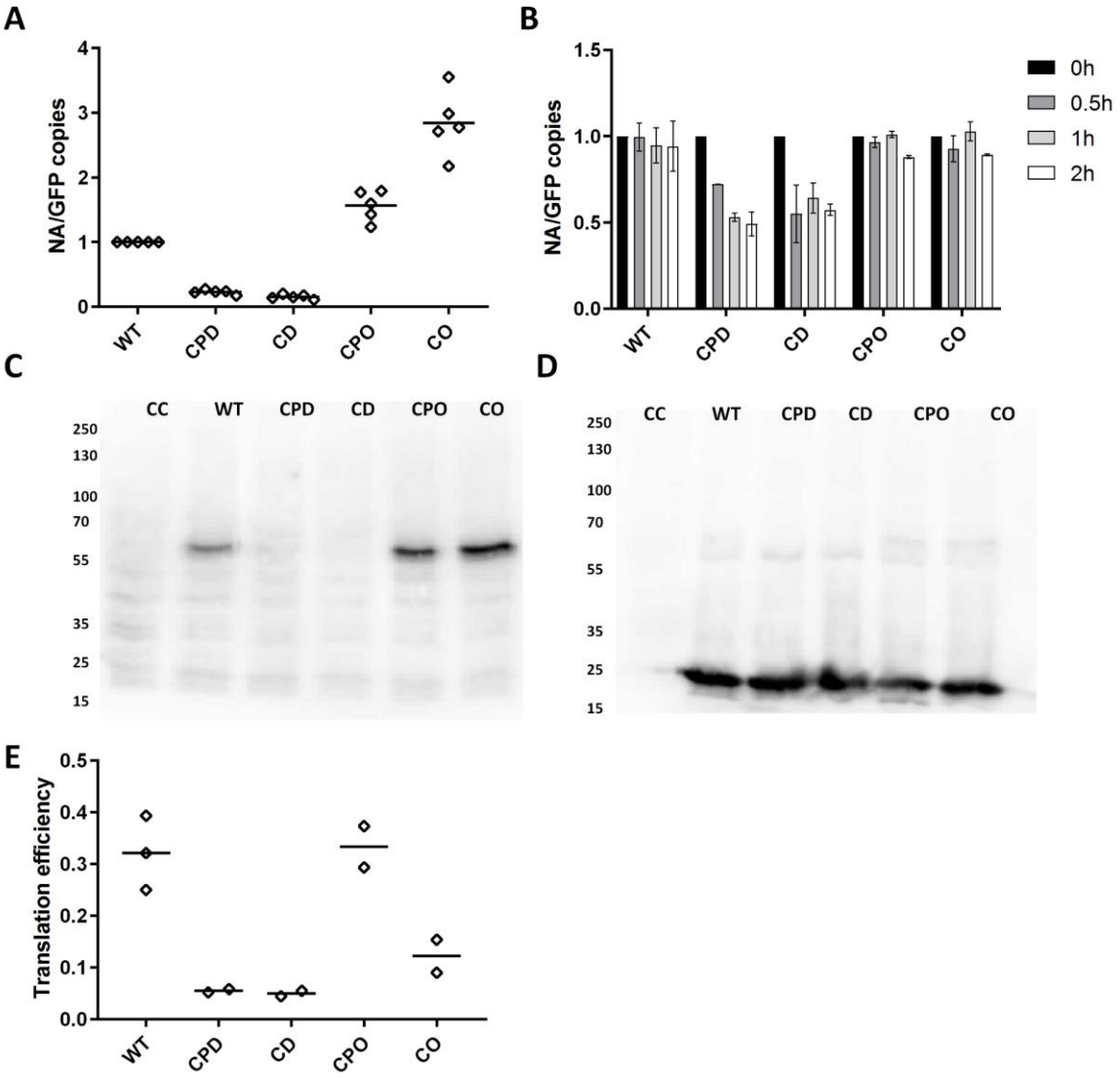

Figure 16: RNA kinetics and protein production of recoded NA genes in transfected cell (Part II). *(A-E) HEK 293T cells were transiently transfected with dual expression plasmids containing recoded NA genes and non-recoded eGFP. (A) Steady state RNA levels of recoded NA genes in transiently transfected cells. RT-qPCR measurements were performed with a probe at the 3' end of the NA genes. EGFP was used for normalization. (B) Stability of NA RNA transcripts determined by transcription inhibitor flavopiridol. Data are the means of three biological replicates ± SD. (C-D) Western blot analysis of (C) recoded NA protein and non-recoded eGFP. (E) Translation efficiency of 29 nt footprints of recoded NA genes from ribosome profiling experiments normalized to housekeeping gene HPRT1.*

8 Discussion

8.1 Recoding to attenuate viruses

Due to the advances in DNA synthesis, full-length genes containing hundreds of nucleotide changes can be chemically synthesized. Such recoding methods are excellent tools to develop live-attenuated vaccines. During the recoding, the amino acid composition remains unchanged and only synonymous mutations are introduced, so that the mutant virus is antigenically identical to the pathogenic, parental virus. Maintaining the identical antigenicity is very important to induce a strong immune response in the host. However, the complete molecular mechanisms behind the attenuating effects of large-scale recoding remains unclear. This knowledge is of enormous importance to efficiently design rapid and safe vaccines for example in case of a virus outbreak. To improve the understanding of the molecular processes, I analysed the effects of dinucleotide content, translation efficiency, RNA stability and multiple other factors during transcription and translation that might lead to the virus attenuation.

8.2 Effects of CpG dinucleotides on virus attenuation

CpG dinucleotides are repressed in vertebrates and in vertebrate viruses[92,118]. My first aim was to analyse the contribution of CpG dinucleotides to virus attenuation in the context of codon pair deoptimization. During codon pair deoptimization, CpG dinucleotides are inadvertently introduced into the genes, because codon pairs with CpG dinucleotides at the codon pair boundary (NNC-GNN) have low CPS. It is difficult to separate the two phenomena of CpG dinucleotides and underrepresented codon pairs in codon deoptimized sequences since they are directly connected. Attempts to address this question failed due to no clean separation of codon usage and codon pair bias[88,106,119]. Attenuation of viruses in these previous studies could be due to changes in the codon usage and not because of their dinucleotide pattern[119]. Additionally, these viruses with an increased number of CpG dinucleotides do not show such a strong attenuation as codon pair deoptimized viruses in other studies[62].

In contrast to previous studies[119,120], the codon bias in my constructs remained unchanged and I completely separated the possible effects of codon usage, codon pair usage and CpG dinucleotides. By recoding the NA gene, I was able to introduce an additional 45 CpG dinucleotides into the coding sequence. The CpG dinucleotide content was solely changed by reshuffling existing codons creating additional CpG dinucleotides at the codon boundary.

The recoded NA genes with an increased number of CpG dinucleotides have either a similar CPS to 27-WT, such as 72H-Max, or a lower average CPS, such as 72L-Max that resulted from CpG containing codon pairs (NNC-GNN) with very low scores (*Figure 6C*). This

introduction of additional CpG dinucleotides into 72H-Max and 72L-Max NA genes did not result in virus attenuation in cell culture (*Figure* 7).

DNA sequences that are CpG-rich were shown to induce innate immune responses[100,121]. Therefore, 72H-Max and 72L-Max were tested as vaccine candidates in mice despite no detection of any growth deficit in cell culture. Only a slightly reduced virulence could be observed in mice infected with recoded viruses that have an increased number of CpG dinucleotides (*Figure 9*). However, this attenuation was much weaker than the attenuation of codon pair deoptimized viruses and most infected mice showed severe morbidity, manifested as weight loss (*Figure 8*). These viruses, containing a large number of CpG dinucleotides, but having a preserved codon usage and no codon pair deoptimization, are therefore unsuitable to be used as vaccine candidates. This leads to the conclusion that CpG dinucleotides do not contribute significantly to the attenuation of influenza A virus in cell culture or in mice.

8.3 Effects of UpA dinucleotides on virus attenuation

TpA (UpA) dinucleotides, even if to a lesser degree than CpG dinucleotides, are also repressed in vertebrates and vertebrate viruses. Previous studies suggested that increasing the frequency of TpA (UpA) dinucleotides in recoded sequences, in addition to CpG dinucleotides, could lead to attenuation of recoded viruses[121]. Nevertheless, attenuating potential of TpA (UpA) dinucleotides is much weaker than those of CpG dinucleotides[121]. Therefore, I also analysed the sequences for other dinucleotide patterns that might influence virus attenuation. Since the codon bias in the sequences remained unchanged and solely existing codons were reshuffled, the only positions capable for dinucleotide changes are the codon boundaries. I did not specifically focus on TpA (UpA) dinucleotides in my work, however, analysing the TpA (UpA) and other dinucleotide frequencies in my constructs, I could not find any clear correlation between dinucleotide content and virus attenuation. Indeed, constructs showing an attenuated phenotype contain a higher TpA (UpA) content – the genes 27-Min, 46-Med, 72-Min have 82, 77 and 100 UpA dinucleotides, respectively, compared to 54 UpA dinucleotides of 27-WT NA gene. However, 72L-Max having also a high number of 71 UpA dinucleotides shows no virus attenuation.

This clearly shows that underrepresented codon pairs and not the increase of CpG or TpA (UpA) dinucleotides are responsible for virus attenuation (*Table 2*).

8.4 Reasons for CpG dinucleotide suppression in RNA viruses

Avian influenza viruses contain a higher percentage of guanine and cytosine nucleotides in their genomes whereas human viruses have a higher adenine and uracil content[122]. The H1N1 influenza subtype jumped approximately 100 years ago from avian to human. Since then the cytosine and guanine content of the H1N1 subtype decreased to adapt to the human nucleotide bias[122]. Especially the CpG dinucleotide content has decreased during H1N1 evolution[100]. The reasons for the underrepresentation of CpG dinucleotides in RNA viruses remains unresolved. I showed that CpG dinucleotides do not play important role in attenuation of the Influenza virus in cell culture (*Figure* 7). In Balb/c mice, I found only a minimal attenuation (*Figure 9, Figure 8*). One reason for the CpG suppression might be that these dinucleotides alone stimulate the innate immune response. It has been shown recently that the zinc-finger antiviral protein ZAP binds to CpG-rich RNAs[8] and recruits the exosome for degradation of the viral RNA[99]. Viruses with a large number of CpG dinucleotides their genomes might be therefore targeted by ZAP. Escaping the recognition from ZAP, might have promoted the reduction of the CpG dinucleotide content in RNA viruses. However, some viruses acquired the ability to develop resistances against ZAP. For example, it has been shown that the influenza NS1 protein and the PB1 polymerase inhibit ZAP activity[123,124]. Therefore, it is likely that ZAP is not the only factor that drove RNA viruses to mimic the dinucleotide content of to their hosts.

8.5 Effects of underrepresented codon pairs on virus attenuation

In my results, I show that codon pair deoptimization – without increasing the number of CpG dinucleotides (27-Min) – leads to a strong attenuation of influenza virus in cell culture (*Figure* 7) and in mice (*Figure 9*), while still inducing high antibody titres and a strong protection against 27-WT virus challenge (*Figure 11*).

I hypothesize that underrepresented codon pairs (codon pairs with very low frequencies in the host) are responsible for the reduced protein expression and virus attenuation. Fine-tuning the degree of attenuation, by choosing deliberately the number of unfavourable codon pairs, might be an important feature for the vaccine development. In a recent study, a codon pair deoptimized dengue virus vaccine was too attenuated in rhesus macaques, to provide protection against dengue virus viremia after challenge[66]. I could confirm that the introduction of a greater number of underrepresented codon pairs (72-Min) lead to a stronger virus attenuation and lower protein production than introduction of a smaller number of underrepresented codon pairs (46-Med) (*Figure* 7)[65,66].

However, we still do not know which codon pairs contribute to which extend to the attenuation. I could exclude codon pairs with very low scores containing CpG dinucleotides from being

responsible for attenuation. However, the CPS is also for viruses having a similar number of CpG dinucleotides not a precise predictor for the attenuation level[66]. As long as the reasons for attenuation are unknown, the exact attenuation level is still unpredictable[66].

My main aim was to focus on the mechanistic processes of codon pair deoptimization and to find the molecular basis of this recoding approach that leads to attenuation. I detected reduced protein levels from the recoded NA genes that underwent codon pair deoptimization (*Figure 12B*). In addition, I could also show reduced RNA levels for these genes (*Figure 12A*). Further investigations showed that the transcription rate of these genes is unaffected (*Figure 12C*), however the RNA degradation was significantly faster in codon and codon pair deoptimized genes (*Figure 12D-E*). This accelerated RNA turnover that leads to overall lower RNA levels, is important for the regulation of gene expression.

It was previously shown that **codon** optimality is a major determinant of mRNA stability[71]. Remarkably, by reshuffling existing codons of the influenza NA gene, I was able to show that **codon pair** choice is another key determinant of mRNA stability.

In addition, I could also show that the translation efficiency of codon pair deoptimized sequences is impaired (*Figure 12F*). Since the used codons in the codon pair deoptimized genes are exactly the same as in WT, this shows that the translation efficiency is greatly dependent on the order of codons or rather on adjacent codons. Alongside the reduced mRNA stability, the reduced translation efficiency is the second major factor causing decreased protein production and attenuation in codon pair deoptimized viruses.

8.6 Genetic stability of live-attenuated vaccines

One important aspect for the security of live-attenuated vaccines is the genetic stability. In contrast to inactivated vaccines, live-attenuated vaccines contain still infectious virions that can replicate. Therefore, it is of great importance that the vaccine does not change to higher virulence. The currently used influenza vaccine FluMist® contains only five amino acid changes, which are responsible for the cold adaptation and therefore causing the restricted replication in the upper respiratory tract[49,50]. Thus, even if very unlikely, the theoretic possibility of adaptation, back to replication at higher temperatures, exist.

In large-scale recoding, it is assumed that hundreds of mutations are responsible for the overall attenuation. If the theory that virus attenuation is caused by an accumulative effect of nucleotide changes is true[66], new vaccine candidates produced with large-scale recoding might have a great stability due to their huge number of genetic changes. It was shown that large-scale recoded viruses are genetically stable but under strong selective pressure, they

can revert back to virulence if not designed carefully[63]. It is unknown whether all underrepresented codon pairs contribute equally to the virus attenuation or whether some codon pairs are the major cause for the reduced protein production. Because the contribution of individual codon pairs to attenuation is not known, the current scoring system for codon pair deoptimization is incapable to predict the level of attenuation accurately (see '*8.7 Limitation of the scoring system for codon pair deoptimization' below*).

In addition, I found an unexpected termination of translation at the same position in the two independently recoded NA genes 46-Med and 27-Min (*Figure 13*). This could be due to an accumulative effect of many unfavourable codon pairs. However, another possible explanation is that one codon pair or a specific sequence motif in these sequences leads to the premature stop of the translation process. This could be a severe problem for the vaccine development based on large-scale recoding, because one specific sequence motif is prone to revert back to the WT phenotype. If attenuation is based on a single motif, would this contradict the currently believed hypothesis, which assumes that an additive effect of many nucleotide changes leads to the virus attenuation. This needs to be further explored. Nevertheless, it is also possible that an accumulation of changes in the first half of the genes leaded to the abrupt stop during translation. Overall, it is crucial that we need a better understanding of the attenuation process. Due to this lacking information, we cannot predict the reversion of live-attenuated vaccines and we cannot guarantee the production of safe genetically stable vaccines.

8.7 Limitation of the scoring system for codon pair deoptimization

To evaluate the degree of deoptimization in codon pair deoptimized sequences, a scoring system – the codon pair score (CPS) – was introduced. The codon pair score calculates the observed frequency relative to its expected frequency of each codon pair, independent of the codon and amino acid usage (see section *'5.8 Codon pair deoptimization'*).

CpG dinucleotides have a strong influence on the average codon pair score[85], as codon pairs with CpG dinucleotides have extremely low scores due to their rare occurrence. Surprisingly, my results indicate that CpG dinucleotides do not have an influence on virus attenuation and therefore pointing out the limitation of the actual scoring system.

The 72L-Max contains many codon pairs of the NNC-GNN type, which contain a CpG at the codon boundary and have very low scores (*Figure 6C*). The remaining codon pairs were selected to have high scores. Caused by the NNC-GNN codon pairs, the average CPS of 72L-Max is as low (CPS = -0.21) as well the average CPS of the codon pair deoptimized constructs 46-Med or 27-Min (CPS = -0.23 and -0.22). My results show indisputably that the

72L-Max virus is not attenuated in contrast to 46-Med and 27-Min, which are attenuated, despite the fact that all three viruses have a similar average CPS (*Figure* 7). This demonstrates that the virus attenuation does not necessarily correlate with the average CPS, if the low CPS is primarily caused by underrepresented codon pairs containing CpG dinucleotides. The cause for the evolutionary reduction of CpG dinucleotides in RNA viruses might have arisen independently from other underrepresented codon pairs (see section *'8.4 Reasons for CpG dinucleotide suppression in RNA viruses'*).

A recent publication showed as well that the CPS does not always perfectly correlate with the actual degree of attenuation. The envelope protein (E) of dengue virus was split in two independent recoded parts each having a very similar CPS, length and number of CpG dinucleotides. They found that one of the constructs was significantly stronger attenuated than the other, indicating the unreliability of CPSs as a predictor of attenuation[66].

Due to the lack of knowledge of the contribution of every single codon pair to the attenuation level of a virus, we observe only an imperfect correlation of attenuation level and CPS and therefore an insufficient evaluation of recoded genes by the CPS. To successfully fine-tune virus attenuation and to understand the complete attenuation mechanism, analysing the contribution of the 3,721 possible codon pairs will be a challenging task but might be inevitable.

8.8 Codon and codon pair deoptimization – two tools for efficient virus attenuation

I discussed above codon pair deoptimization as a strategy for virus attenuation[65] and the knowledge gaps in understanding this methods. Beside codon pair deoptimization, a similar method – codon deoptimization – was develop to design live-attenuated vaccines[70,125]. These two methods both involve the recoding of the DNA/RNA sequences without changing the amino acid sequence and provide great vaccine candidates. Both methods were applied for a series of different viruses, including influenza virus[62,69], poliovirus[65,125] or dengue virus[126,127].

Even though codon usage was studied a lot more than codon pair usage, the exact molecular mechanisms underlying the virus attenuation need to be further explored. Knowing – for both methods – the exact molecular mechanism leading to reduced protein production and therefore to virus attenuation would improve the vaccine development and would allow fine-tuning the protein level.

A comparison of codon and codon pair deoptimized viruses has never been done before. Therefore, my aim was to compare these two deoptimization methods to find the better approach to create live-attenuated viral vaccines. As expected, codon deoptimization and codon pair deoptimization of the influenza NA gene showed a strong reduction of virulence *in*

vitro (*Figure 15*) and *in vivo* (*Figure 9*). Both viruses showed reduced growth and formed smaller plaque sizes than the parental virus. Nevertheless, the viruses showed efficient protection against a WT challenge and were capable of inducing a strong immune response. Therefore, both attenuation methods have a great potential for the use as vaccines (*Figure 11*). However, the viruses were extremely attenuated, so that I could not distinguish which virus was stronger attenuated. The only small tendency I found while comparing the two viruses was that the codon pair deoptimized virus had a lower titre in mice lungs and caused less necrosis in lungs of mice as well as less inflammation on day 2 after infection (*Figure 9*). This leads to the conclusion that my codon pair deoptimized viruses might be slightly stronger attenuated than codon deoptimized viruses, but still possess full immune protection against WT challenge. However, to get to a comprehensive and universal conclusion, which attenuation strategy is more valuable, additional genes and different viruses need to be analysed.

8.9 Optimization - a tool to improve protein expression

A common application of the **codon** optimization technology is to improve protein expression, for example for diverse fluorescent proteins. To obtain higher expression in mammalian cells, the codon usage of the green fluorescent protein (GFP) was optimized to the now commonly used enhanced GFP (eGFP)[68]. **Codon pair** optimization on the other hand, has not been used as a tool to increase protein expression of target genes. So far, the structural region of poliovirus was codon pair optimized and showed higher translatability compared to the WT gene[65]. Additionally, a recent work showed that codon pair optimization of the UL30 gene in Marek's disease herpesvirus induced in cell culture bigger plaques than the parental virus[61].

In my study, the NA gene of the influenza virus was optimized – not to obtain higher virulent viruses – but rather to find universal mechanistic rules valid not only for deoptimization, but also for optimization to for example improve protein expression of target genes.

The codon and codon pair optimization of influenza NA gene did not lead to an improved growth of viruses in cell culture (*Figure 15*). The codon pair optimized virus showed similar behaviour as the parental virus according to its growth behaviour and spread despite the insertion of 218 nucleotide changes in the NA gene sequence. The codon optimized virus was, on the contrary, attenuated.

Knowing that the codon and codon pair optimized NA does not increase virus growth, I analysed the protein expression in a dual expression plasmid containing the NA genes and eGFP under control of two independent promoters. Surprisingly, I found that the codon optimized NA that produces attenuated viruses showed high RNA levels and protein

expression in transfected cells. Therefore, the reduced fitness of the codon optimized virus seems to be a virus specific feature and independent of the successful optimization of the NA gene (see section '*8.10 Why is the NA- codon optimized virus* attenuated?*' below*).

Similarly, to the codon optimized gene, the codon pair optimized gene showed significantly higher protein expression from a dual expression plasmid. Here I can demonstrate that not only the widely used method of **codon** optimization but also **codon pair** optimization is a powerful tool to increase the protein expression. This new recoding tool of changing codon pair biases might be a good alternative to the actually used codon optimization with the difference that the codon usage is preserved.

8.10 Why is the NA- codon optimized virus attenuated?

Despite an efficient protein production of the codon optimized NA gene on a plasmid level, the fitness of virus containing this gene is significantly reduced (*Figure 15*). This attenuation occurs likely through a virus specific feature during virus assembly or replication.

It was shown in a cell-free translation system that codon optimized mRNA of a luciferase gene had faster translation elongation and also produced more protein, however most of the protein was non-functional and therefore the production of functional protein was much slower using the optimized mRNA compared to the parental one[128]. The possibility that produced protein of the optimized NA is non-functional and therefore leading to reduced virus replication, was excluded by performing a Western blot analysis 5 h after infection with a MOI of 5. A life cycle of influenza takes 6 h, so by Western blot analysis, I detected the total amount of NA protein this includes the non-functional protein that is, for example incorrectly folded (*Figure 15E*).

The CO virus formed plaques with similar sizes as WT virus, showing that the virus spread is not hampered. However, CO showed severe growth deficits on A549 and MDCK cells. One reason for reduced production of infectious virus can be a change in specific binding patterns of the recoded NA-vRNA to the nuclear protein NP resulting in insufficient packaging. Another possibility could be impaired production of NA-vRNA in CO virus.

9 Concluding remarks and outlook

I discovered that reduced RNA stability and reduced translation efficiency are major factors for reduced protein levels and virus attenuation of codon pair deoptimized viruses. These findings explain how recoded, codon pair deoptimized viruses become attenuated on a molecular level.

However, the obtained results also highlight the necessity of further studies to gain more information about the mechanism underlying codon pair deoptimization. I show that the CPS is not an accurate predictor for the attenuation level of codon pair deoptimized viruses. Furthermore, my results question the current hypothesis that virus attenuation is based on an additive effect of many unfavourable codon pairs. As I observe a premature translation stop at one certain position in recoded RNA, there is a possibility that only one specific sequence pattern might be responsible for the reduced protein expression in codon deoptimized sequences. If this is the case, a rethinking in the vaccine development is necessary, as a single sequence pattern is prone to revert back to the WT phenotype.

To evaluate the virus attenuation level with an accurate score but also to guarantee the genetic stability of codon pair deoptimized viruses, the impact of every single codon pair is thereby of huge importance. Further decoding of the mechanism behind codon pair deoptimization would advance the development of safe, live-attenuated vaccines of a broad range of viruses.

Finally, I showed here that codon pair optimization is a new method to improve protein expression by large-scale recoding. This new tool could be used for similar applications as the nowadays-used codon optimization with the advantage that the codon usage stays identical to the original gene.

10 Summary

Molecular mechanism of virus attenuation by codon pair deoptimization

The chemical synthesis of nucleic acids enables the generation of recoded genes with novel properties. The synonymous genome recoding introduces a large number of mutations into recoded genes without changing the amino acid composition of encoded proteins. The goal of the recoding is to alter codon usage, codon pair usage or dinucleotide frequencies of the recoded genes and thereby change different biological properties. Codon optimization and codon pair optimization are used to improve protein production from recoded genes. On the contrary, the goal of codon deoptimization and codon pair deoptimization is to reduce the protein expression of recoded genes, which can be used for the production of experimental live-attenuated virus vaccine candidates. While codon and codon pair deoptimization are capable of producing attenuated viruses, the molecular basis behind the attenuation is not well understood.

In my work, I confirmed that codon and codon pair deoptimization are both excellent tools for virus attenuation, and therefore suitable for the development of viral vaccines. These live-attenuated virus vaccines, prepared by both methods, induced in mice model a remarkable protection against WT virus challenge. The main goal of this project was to elucidate the molecular mechanism of virus attenuation by codon pair deoptimization. Due to low frequencies of CpG dinucleotides in the human genome, codon pair deoptimization unintentionally increases the number of CpG dinucleotides in the recoded genes. Therefore, it was assumed that CpG dinucleotides have a substantial impact on virus attenuation. I showed that increased numbers of CpG dinucleotides do not influence influenza A virus. However, I discovered that RNAs produced by codon pair deoptimized genes have higher degradation rates. Additionally, I found out that codon pair deoptimization reduces protein production of recoded genes also by decreasing their translation efficiency. These findings elucidate the highly complex molecular mechanism underlying the synonymous genome recoding.

In the second part of my work, I showed that the commonly used codon optimization is a suitable method to increase the protein expression of target genes. More importantly, I demonstrated that also codon pair optimization has a great potential to improve RNA stability and protein production. Therefore, codon pair optimization might be a new powerful tool to enhance protein expression for diverse applications in biotechnology.

11 Zusammenfassung

Molekularer Mechanismus der Codonpaardeoptimierung zur Attenuierung von Viren

Die chemische Synthese von Nukleinsäuren ermöglicht die Erzeugung von neu kodierten Genen mit veränderten Eigenschaften. Die synonyme Recodierung fügt eine große Anzahl von Mutationen in Gene ein, ohne die Aminosäurezusammensetzung der gebildeten Proteine zu verändern. Das Ziel der Recodierung von Genen ist, die Codonverwendung, Codonpaarverwendung oder die Häufigkeit von verschiedenen Dinukleotiden zu verändern und damit verbunden auch biologische Eigenschaften. Codonoptimierung und Codonpaaroptimierung werden verwendet um die Proteinproduktion zu verbessern. Codondeoptimierung und Codonpaardeoptimierung hingegen, haben das Ziel, die Proteinproduktion der umcodierten Gene zu verringern. Dies findet Anwendung in der Erzeugung von attenuierten Viren, die dann für die Entwicklung von Lebendimpfstoffen verwendet werden. Der molekulare Hintergrund, auf dem die Attenuierung der Viren basiert, ist jedoch bislang noch unklar.

Ich zeige in meiner Arbeit, dass Codon- und Codonpaardeoptimierung beides exzellente Methoden sind um Viren abzuschwächen und um Lebendimpfstoffe zu entwickeln. Die mit beiden Methoden entwickelten Lebendimpfstoffe induzierten einen bemerkenswerten Schutz gegen eine Wildtyp Infektion. Das Hauptziel dieses Projekts war es, den molekularen Mechanismus der Virenattenuierung durch Codonpaardeoptimierung zu ergründen. Aufgrund des seltenen Auftretens von CpG Dinukleotiden im menschlichen Genom, erhöht die Codonpaardeoptimierung ungewollt die Anzahl der CpG Dinukleotide im recodierten Gen. Daher wurde bisher angenommen, dass die CpG Dinukleotide einen großen Einfluss auf die Attenuierung von Viren haben. Ich zeige jedoch in meiner Arbeit, dass eine erhöhte Anzahl von CpG Dinukleotiden das Influenza-A-Virus nicht beeinflusst. Hingegen zeigten die RNAs, die von codonpaardeoptimierten Genen gebildet wurden, eine höhere Degradationsrate. Zusätzlich reduzierte die Codonpaardeoptimierung die Proteinproduktion durch eine geringere Translationseffizienz. Meine Ergebnisse erläutern den hoch komplexen molekularen Mechanismus, welcher der Codonpaardeoptimierung von Genen zugrunde liegt.

In dem zweiten Teil meiner Arbeit zeigte ich, dass die gängige Methode der Codonoptimierung geeignet ist um die Proteinproduktion eines Genes zu erhöhen. Wichtiger jedoch, ich konnte zeigen, dass Codonpaaroptimierung auch ein großes Potential hat um die RNA Stabilität zu verbessern und die Proteinproduktion zu erhöhen. Daher scheint die Codonpaaroptimierung eine neue, geeignete Methode zu sein um die Proteinproduktion zu steigern, was diverse Anwendungen z.B. in der Biotechnologie finden kann.

12 References

1 World Health Organization. Influenza. https://www.who.int/influenza/vaccines/en/. Accessed November 11, 2019.

2 Hause, B. M. *et al.* Characterization of a novel influenza virus in cattle and Swine: proposal for a new genus in the Orthomyxoviridae family. *mBio* **5**, e00031-00014, doi:10.1128/mBio.00031-14 (2014).

3 Noda, T. Native morphology of influenza virions. *Frontiers in microbiology* **2**, 269, doi:10.3389/fmicb.2011.00269 (2011).

4 McGeoch, D., Fellner, P. & Newton, C. Influenza virus genome consists of eight distinct RNA species. *Proceedings of the National Academy of Sciences of the United States of America* **73**, 3045-3049, doi:10.1073/pnas.73.9.3045 (1976).

5 Dou, D., Revol, R., Ostbye, H., Wang, H. & Daniels, R. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. *Frontiers in immunology* **9**, 1581, doi:10.3389/fimmu.2018.01581 (2018).

6 Shaw, M. L., Stone, K. L., Colangelo, C. M., Gulcicek, E. E. & Palese, P. Cellular proteins in influenza virus particles. *PLoS pathogens* **4**, e1000085, doi:10.1371/journal.ppat.1000085 (2008).

7 Gerl, M. J. *et al.* Quantitative analysis of the lipidomes of the influenza virus envelope and MDCK cell apical membrane. *The Journal of cell biology* **196**, 213-221, doi:10.1083/jcb.201108175 (2012).

8 Gamblin, S. J. & Skehel, J. J. Influenza hemagglutinin and neuraminidase membrane glycoproteins. *The Journal of biological chemistry* **285**, 28403-28409, doi:10.1074/jbc.R110.129809 (2010).

9 Tong, S. *et al.* New world bats harbor diverse influenza A viruses. *PLoS pathogens* **9**, e1003657, doi:10.1371/journal.ppat.1003657 (2013).

10 Samji, T. Influenza A: understanding the viral life cycle. *The Yale journal of biology and medicine* **82**, 153-159 (2009).

11 Krammer, F. *et al.* Influenza. *Nature reviews. Disease primers* **4**, 3, doi:10.1038/s41572-018-0002-y (2018).

12 Huang, Q. *et al.* Early steps of the conformational change of influenza virus hemagglutinin to a fusion active state: stability and energetics of the hemagglutinin. *Biochimica et biophysica acta* **1614**, 3-13, doi:10.1016/s0005-2736(03)00158-5 (2003).

13 Pinto, L. H. & Lamb, R. A. The M2 proton channels of influenza A and B viruses. *The Journal of biological chemistry* **281**, 8997-9000, doi:10.1074/jbc.R500020200 (2006).

14 Newcomb, L. L. *et al.* Interaction of the influenza a virus nucleocapsid protein with the viral RNA polymerase potentiates unprimed viral RNA replication. *Journal of virology* **83**, 29-36, doi:10.1128/jvi.02293-07 (2009).

15 Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**, 847-858, doi:10.1016/0092-8674(81)90449-9 (1981).

16 Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nature structural & molecular biology* **15**, 500-506, doi:10.1038/nsmb.1421 (2008).

17 Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458**, 914-918, doi:10.1038/nature07745 (2009).

18 Poon, L. L., Pritlove, D. C., Fodor, E. & Brownlee, G. G. Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. *Journal of virology* **73**, 3473-3476 (1999).

19 Harris, A. *et al.* Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 19123-19127, doi:10.1073/pnas.0607614103 (2006).

20 Du, R., Cui, Q. & Rong, L. Competitive Cooperation of Hemagglutinin and Neuraminidase during Influenza A Virus Entry. *Viruses* **11**, doi:10.3390/v11050458 (2019).

21 Air, G. M. Influenza neuraminidase. *Influenza and other respiratory viruses* **6**, 245-256, doi:10.1111/j.1750-2659.2011.00304.x (2012).

22 Palese, P., Tobita, K., Ueda, M. & Compans, R. W. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology* **61**, 397-410, doi:10.1016/0042-6822(74)90276-1 (1974).

23 Guo, H. *et al.* Kinetic analysis of the influenza A virus HA/NA balance reveals contribution of NA to virus-receptor binding and NA-dependent rolling on receptor-containing surfaces. *PLoS pathogens* **14**, e1007233, doi:10.1371/journal.ppat.1007233 (2018).

24 Su, B. *et al.* Enhancement of the influenza A hemagglutinin (HA)-mediated cell-cell fusion and virus entry by the viral neuraminidase (NA). *PloS one* **4**, e8495, doi:10.1371/journal.pone.0008495 (2009).

25 Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. A. & Klenk, H. D. Neuraminidase is important for the initiation of influenza virus infection in human airway epithelium. *Journal of virology* **78**, 12665-12667, doi:10.1128/jvi.78.22.12665-12667.2004 (2004).

26 Cohen, M. *et al.* Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virology journal* **10**, 321, doi:10.1186/1743-422x-10-321 (2013).

27 Hurt, A. C. *et al.* Community transmission of oseltamivir-resistant A(H1N1)pdm09 influenza. *The New England journal of medicine* **365**, 2541-2542, doi:10.1056/NEJMc1111078 (2011).

28 Feery, B. J. The conquest of smallpox. *Australian family physician* **5**, 720-733 (1976).

29 Weiss, R. A. & Esparza, J. The prevention and eradication of smallpox: a commentary on Sloane (1755) 'An account of inoculation'. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **370**, doi:10.1098/rstb.2014.0378 (2015).

30 Koprowski, H. First decade (1950-1960) of studies and trials with the polio vaccine. *Biologicals : journal of the International Association of Biological Standardization* **34**, 81-86, doi:10.1016/j.biologicals.2006.03.009 (2006).

31 Griffin, D. E. Measles Vaccine. *Viral immunology* **31**, 86-95, doi:10.1089/vim.2017.0143 (2018).

32 Centers for Disease, C. & Prevention. Impact of vaccines universally recommended for children--United States, 1990-1998. *MMWR Morb Mortal Wkly Rep* **48**, 243-248 (1999).

33 Delany, I., Rappuoli, R. & De Gregorio, E. Vaccines for the 21st century. *EMBO molecular medicine* **6**, 708-720, doi:10.1002/emmm.201403876 (2014).

34 Lauring, A. S., Jones, J. O. & Andino, R. Rationalizing the development of live attenuated virus vaccines. *Nat Biotechnol* **28**, 573-579, doi:10.1038/nbt.1635 (2010).

35 Vetter, V., Denizer, G., Friedland, L. R., Krishnan, J. & Shapiro, M. Understanding modern-day vaccines: what you need to know. *Annals of medicine* **50**, 110-120, doi:10.1080/07853890.2017.1407035 (2018).

36 Shi, S. *et al.* Vaccine adjuvants: Understanding the structure and mechanism of adjuvanticity. *Vaccine* **37**, 3167-3178, doi:10.1016/j.vaccine.2019.04.055 (2019).

37 Deering, R. P., Kommareddy, S., Ulmer, J. B., Brito, L. A. & Geall, A. J. Nucleic acid vaccines: prospects for non-viral delivery of mRNA vaccines. *Expert opinion on drug delivery* **11**, 885-899, doi:10.1517/17425247.2014.901308 (2014).

38 Wang, L. F. & Crameri, G. Emerging zoonotic viral diseases. *Rev Sci Tech* **33**, 569-581 (2014).

39 Make Ebola a thing of the past: first vaccine against deadly virus approved. (2019) https://www.nature.com/articles/d41586-019-03490-8. Accessed November 15, 2019.

40 Biswal, S. *et al.* Efficacy of a Tetravalent Dengue Vaccine in Healthy Children and Adolescents. *The New England journal of medicine*, doi:10.1056/NEJMoa1903869 (2019).

41 Villar, L. *et al.* Efficacy of a tetravalent dengue vaccine in children in Latin America. *The New England journal of medicine* **372**, 113-123, doi:10.1056/NEJMoa1411037 (2015).

42 Prompetchara, E., Ketloy, C., Thomas, S. J. & Ruxrungtham, K. Dengue vaccine: Global development update. *Asian Pacific journal of allergy and immunology*, doi:10.12932/ap-100518-0309 (2019).

43 Barberis, I., Myles, P., Ault, S. K., Bragazzi, N. L. & Martini, M. History and evolution of influenza control through vaccination: from the first monovalent vaccine to universal vaccines. *Journal of preventive medicine and hygiene* **57**, E115-e120 (2016).

44 Bedford, T. *et al.* Integrating influenza antigenic dynamics with molecular evolution. *eLife* **3**, e01914, doi:10.7554/eLife.01914 (2014).

45 Rajao, D. S. & Perez, D. R. Universal Vaccines and Vaccine Platforms to Protect against Influenza Viruses in Humans and Agriculture. *Frontiers in microbiology* **9**, 123, doi:10.3389/fmicb.2018.00123 (2018).

46 Graaf, H. & Faust, S. N. Fluarix quadrivalent vaccine for influenza. *Expert review of vaccines* **14**, 1055-1063, doi:10.1586/14760584.2015.1057573 (2015).

47 Sridhar, S., Brokstad, K. A. & Cox, R. J. Influenza Vaccination Strategies: Comparing Inactivated and Live Attenuated Influenza Vaccines. *Vaccines* **3**, 373-389, doi:10.3390/vaccines3020373 (2015).

48 Buhler, S. & Ramharter, M. Flucelvax Tetra: a surface antigen, inactivated, influenza vaccine prepared in cell cultures. *ESMO open* **4**, e000481, doi:10.1136/esmoopen-2018-000481 (2019).

49 Jin, H. *et al.* Multiple amino acid residues confer temperature sensitivity to human influenza virus vaccine strains (FluMist) derived from cold-adapted A/Ann Arbor/6/60. *Virology* **306**, 18-24, doi:10.1016/s0042-6822(02)00035-1 (2003).

50 Murphy, B. R. & Coelingh, K. Principles underlying the development and use of live attenuated cold-adapted influenza A and B virus vaccines. *Viral immunology* **15**, 295-323, doi:10.1089/08828240260066242 (2002).

51 Gomez Lorenzo, M. M. & Fenton, M. J. Immunobiology of influenza vaccines. *Chest* **143**, 502-510, doi:10.1378/chest.12-1711 (2013).

52 Treanor, J. J. *et al.* Protective efficacy of a trivalent recombinant hemagglutinin protein vaccine (FluBlok(R)) against influenza in healthy adults: a randomized, placebo-controlled trial. *Vaccine* **29**, 7733-7739, doi:10.1016/j.vaccine.2011.07.128 (2011).

53 King, J. C., Jr. *et al.* Evaluation of the safety, reactogenicity and immunogenicity of FluBlok trivalent recombinant baculovirus-expressed hemagglutinin influenza vaccine administered intramuscularly to healthy children aged 6-59 months. *Vaccine* **27**, 6589-6594, doi:10.1016/j.vaccine.2009.08.032 (2009).

54 Robert Koch-Institut. Bericht zur Epidemiologie der Influenza in Deutschland, Saison 2017/18. https://influenza.rki.de/Saisonberichte/2017.pdf. Accessed November 12, 2019.

55 Robert Koch-Institut. Bericht zur Epidemiologie der Influenza in Deutschland, Saison 2018/19. https://edoc.rki.de/bitstream/handle/176904/6253/RKI_Influenzabericht_2018-19.pdf?sequence=1&isAllowed=y. Accessed November 12, 2019.

56 Kumar, A., Meldgaard, T. S. & Bertholet, S. Novel Platforms for the Development of a Universal Influenza Vaccine. *Frontiers in immunology* **9**, 600, doi:10.3389/fimmu.2018.00600 (2018).

57 Stadlbauer, D. *et al.* Broadly protective human antibodies that target the active site of influenza virus neuraminidase. *Science (New York, N.Y.)* **366**, 499-504, doi:10.1126/science.aay0678 (2019).

58 Cello, J., Paul, A. V. & Wimmer, E. Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template. *Science (New York, N.Y.)* **297**, 1016-1018, doi:10.1126/science.1072266 (2002).

59 Wimmer, E., Mueller, S., Tumpey, T. M. & Taubenberger, J. K. Synthetic viruses: a new opportunity to understand and prevent viral disease. *Nat Biotechnol* **27**, 1163-1172, doi:10.1038/nbt.1593 (2009).

60 Mueller, S. *et al.* Live attenuated influenza virus vaccines by computer-aided rational design. *Nat Biotechnol* **28**, 723-726, doi:10.1038/nbt.1636 (2010).

61 Eschke, K., Trimpert, J., Osterrieder, N. & Kunec, D. Attenuation of a very virulent Marek's disease herpesvirus (MDV) by codon pair bias deoptimization. *PLoS pathogens* **14**, e1006857, doi:10.1371/journal.ppat.1006857 (2018).

62 Yang, C., Skiena, S., Futcher, B., Mueller, S. & Wimmer, E. Deliberate reduction of hemagglutinin and neuraminidase expression of influenza virus leads to an ultraprotective live vaccine in mice. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 9481-9486, doi:10.1073/pnas.1307473110 (2013).

63 Le Nouen, C. *et al.* Genetic stability of genome-scale deoptimized RNA virus vaccine candidates under selective pressure. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E386-e395, doi:10.1073/pnas.1619242114 (2017).

64 Ni, Y. Y. *et al.* Computer-aided codon-pairs deoptimization of the major envelope GP5 gene attenuates porcine reproductive and respiratory syndrome virus. *Virology* **450-451**, 132-139, doi:10.1016/j.virol.2013.12.009 (2014).

65 Coleman, J. R. *et al.* Virus attenuation by genome-scale changes in codon pair bias. *Science (New York, N.Y.)* **320**, 1784-1787, doi:10.1126/science.1155761 (2008).

66 Stauft, C. B. *et al.* Extensive genomic recoding by codon-pair deoptimization selective for mammals is a flexible tool to generate attenuated vaccine candidates for dengue virus 2. *Virology* **537**, 237-245, doi:10.1016/j.virol.2019.09.003 (2019).

67 Quax, T. E., Claassens, N. J., Soll, D. & van der Oost, J. Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular cell* **59**, 149-161, doi:10.1016/j.molcel.2015.05.035 (2015).

68 Zhang, G., Gurtu, V. & Kain, S. R. An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochemical and biophysical research communications* **227**, 707-711, doi:10.1006/bbrc.1996.1573 (1996).

69 Nogales, A. *et al.* Influenza A virus attenuation by codon deoptimization of the NS gene for vaccine development. *Journal of virology* **88**, 10525-10540, doi:10.1128/JVI.01565-14 (2014).

70 Burns, C. C. *et al.* Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *Journal of virology* **80**, 3259-3272, doi:10.1128/JVI.80.7.3259-3272.2006 (2006).

71 Presnyak, V. *et al.* Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111-1124, doi:10.1016/j.cell.2015.02.029 (2015).

72 Harigaya, Y. & Parker, R. Analysis of the association between codon optimality and mRNA stability in Schizosaccharomyces pombe. *BMC genomics* **17**, 895, doi:10.1186/s12864-016-3237-6 (2016).

73 Gingold, H. & Pilpel, Y. Determinants of translation efficiency and accuracy. *Mol Syst Biol* **7**, 481, doi:10.1038/msb.2011.14 (2011).

74 Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature structural & molecular biology* **16**, 274-280, doi:10.1038/nsmb.1554 (2009).

75 Wu, Q. *et al.* Translation affects mRNA stability in a codon-dependent manner in human cells. *eLife* **8**, doi:10.7554/eLife.45396 (2019).

76 Gardin, J. *et al.* Measurement of average decoding rates of the 61 sense codons in vivo. *eLife* **3**, doi:10.7554/eLife.03735 (2014).

77 Weinberg, D. E. *et al.* Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell reports* **14**, 1787-1799, doi:10.1016/j.celrep.2016.01.043 (2016).

78 Gamble, C. E., Brule, C. E., Dean, K. M., Fields, S. & Grayhack, E. J. Adjacent Codons Act in Concert to Modulate Translation Efficiency in Yeast. *Cell* **166**, 679-690, doi:10.1016/j.cell.2016.05.070 (2016).

79 Curran, J. F. Decoding with the A:I wobble pair is inefficient. *Nucleic acids research* **23**, 683-688, doi:10.1093/nar/23.4.683 (1995).

80 Athey, J. *et al.* A new and updated resource for codon usage tables. *BMC bioinformatics* **18**, 391, doi:10.1186/s12859-017-1793-7 (2017).

81 Li, P. *et al.* Zika Virus Attenuation by Codon Pair Deoptimization Induces Sterilizing Immunity in Mouse Models. *Journal of virology* **92**, doi:10.1128/jvi.00701-18 (2018).

82 Stauft, C. B. *et al.* Live-attenuated H1N1 influenza vaccine candidate displays potent efficacy in mice and ferrets. *PloS one* **14**, e0223784, doi:10.1371/journal.pone.0223784 (2019).

83 Moura, G. *et al.* Comparative context analysis of codon pairs on an ORFeome scale. *Genome biology* **6**, R28, doi:10.1186/gb-2005-6-3-r28 (2005).

84 Moura, G. *et al.* Large scale comparative codon-pair context analysis unveils general rules that fine-tune evolution of mRNA primary structure. *PloS one* **2**, e847, doi:10.1371/journal.pone.0000847 (2007).

85 Kunec, D. & Osterrieder, N. Codon Pair Bias Is a Direct Consequence of Dinucleotide Bias. *Cell reports* **14**, 55-67, doi:10.1016/j.celrep.2015.12.011 (2016).

86 Wang, B. *et al.* Recoding of the vesicular stomatitis virus L gene by computer-aided design provides a live, attenuated vaccine candidate. *mBio* **6**, doi:10.1128/mBio.00237-15 (2015).

87 Broadbent, A. J. *et al.* Evaluation of the attenuation, immunogenicity, and efficacy of a live virus vaccine generated by codon-pair bias de-optimization of the 2009 pandemic H1N1 influenza virus, in ferrets. *Vaccine* **34**, 563-570, doi:10.1016/j.vaccine.2015.11.054 (2016).

88 Tulloch, F., Atkinson, N. J., Evans, D. J., Ryan, M. D. & Simmonds, P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *eLife* **4**, doi:10.7554/eLife.04531 (2014).

89 Buchan, J. R., Aucott, L. S. & Stansfield, I. tRNA properties help shape codon pair preferences in open reading frames. *Nucleic acids research* **34**, 1015-1027, doi:10.1093/nar/gkj488 (2006).

90 Doma, M. K. & Parker, R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**, 561-564, doi:10.1038/nature04530 (2006).

91 Passos, D. O. *et al.* Analysis of Dom34 and its function in no-go decay. *Molecular biology of the cell* **20**, 3025-3032, doi:10.1091/mbc.E09-01-0028 (2009).

92 Campbell, A., Mrazek, J. & Karlin, S. Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 9184-9189 (1999).

93 Osterrieder, N. & Kunec, D. Attenuation of Viruses by Large-Scale Recoding of their Genomes: the Selection Is Always Biased. *Current Clinical Microbiology Reports* **5**, 66-72, doi:10.1007/s40588-018-0080-3 (2018).

94 Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. Molecular basis of base substitution hotspots in Escherichia coli. *Nature* **274**, 775-780 (1978).

95 Bird, A. P. DNA methylation and the frequency of CpG in animal DNA. *Nucleic acids research* **8**, 1499-1504 (1980).

96 Cooper, D. N. & Gerber-Huber, S. DNA methylation and CpG suppression. *Cell differentiation* **17**, 199-205 (1985).

97 Takata, M. A. *et al.* CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* **550**, 124-127, doi:10.1038/nature24039 (2017).

98 Ficarelli, M. *et al.* KHNYN is essential for the zinc finger antiviral protein (ZAP) to restrict HIV-1 containing clustered CpG dinucleotides. *eLife* **8**, doi:10.7554/eLife.46767 (2019).

99 Guo, X., Ma, J., Sun, J. & Gao, G. The zinc-finger antiviral protein recruits the RNA processing exosome to degrade the target mRNA. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 151-156, doi:10.1073/pnas.0607063104 (2007).

100 Greenbaum, B. D., Rabadan, R. & Levine, A. J. Patterns of oligonucleotide sequences in viral and host cell RNA identify mediators of the host innate immune system. *PloS one* **4**, e5969, doi:10.1371/journal.pone.0005969 (2009).

101 Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science (New York, N.Y.)* **329**, 52-56, doi:10.1126/science.1190719 (2010).

102 Annaluru, N. *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science (New York, N.Y.)* **344**, 55-58, doi:10.1126/science.1249252 (2014).

103 Wimmer, E. The test-tube synthesis of a chemical called poliovirus. The simple synthesis of a virus has far-reaching societal implications. *EMBO reports* **7 Spec No**, S3-9, doi:10.1038/sj.embor.7400728 (2006).

104 Codagenix Inc. Vaccine pipeline. https://codagenix.com/vaccine-programs/pipeline/. Accessed September 09, 2019.

105 Simmonds, P., Tulloch, F., Evans, D. J. & Ryan, M. D. Attenuation of dengue (and other RNA viruses) with codon pair recoding can be explained by increased CpG/UpA dinucleotide frequencies. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E3633-3634, doi:10.1073/pnas.1507339112 (2015).

106 Futcher, B. *et al.* Reply to Simmonds et al.: Codon pair and dinucleotide bias have not been functionally distinguished. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E3635-3636, doi:10.1073/pnas.1507710112 (2015).

107 Schierhorn, K. L. *et al.* Influenza A Virus Virulence Depends on Two Amino Acids in the N-Terminal Domain of Its NS1 Protein To Facilitate Inhibition of the RNA-Dependent Protein Kinase PKR. *Journal of virology* **91**, doi:10.1128/JVI.00198-17 (2017).

108 Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G. & Webster, R. G. A DNA transfection system for generation of influenza A virus from eight plasmids. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6108-6113, doi:10.1073/pnas.100133697 (2000).

109 Morrison, E. *et al.* Quantitative analysis of the human T cell palmitome. *Scientific reports* **5**, 11598, doi:10.1038/srep11598 (2015).

110 Paulsen, M. T. *et al.* Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods (San Diego, Calif.)* **67**, 45-54, doi:10.1016/j.ymeth.2013.08.015 (2014).

111 Calviello, L. *et al.* Detecting actively translated open reading frames in ribosome profiling data. *Nature methods* **13**, 165-170, doi:10.1038/nmeth.3688 (2016).

112 Dodt, M., Roehr, J. T., Ahmed, R. & Dieterich, C. FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology* **1**, 895-905, doi:10.3390/biology1030895 (2012).

113 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).

114 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

115 Gaidatzis, D., Lerch, A., Hahne, F. & Stadler, M. B. QuasR: quantification and annotation of short reads in R. *Bioinformatics (Oxford, England)* **31**, 1130-1132, doi:10.1093/bioinformatics/btu781 (2015).

116 Zhong, Y. *et al.* RiboDiff: detecting changes of mRNA translation efficiency from ribosome footprints. *Bioinformatics (Oxford, England)* **33**, 139-141, doi:10.1093/bioinformatics/btw585 (2017).

117 Organization, W. H. WHO manual on animal influenza diagnosis and surveillance. (Geneva: World Health Organization, 2002).

118 Karlin, S., Doerfler, W. & Cardon, L. R. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *Journal of virology* **68**, 2889-2897 (1994).

119 Gaunt, E. *et al.* Elevation of CpG frequencies in influenza A genome attenuates pathogenicity but enhances host response to infection. *eLife* **5**, e12735, doi:10.7554/eLife.12735 (2016).

120 Burns, C. C. *et al.* Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *Journal of virology* **83**, 9957-9969, doi:10.1128/JVI.00508-09 (2009).

121 Atkinson, N. J., Witteveldt, J., Evans, D. J. & Simmonds, P. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic acids research* **42**, 4527-4545, doi:10.1093/nar/gku075 (2014).

122 Rabadan, R., Levine, A. J. & Robins, H. Comparison of avian and human influenza A viruses reveals a mutational bias on the viral genomes. *Journal of virology* **80**, 11887-11891, doi:10.1128/jvi.01414-06 (2006).

123 Liu, C. H., Zhou, L., Chen, G. & Krug, R. M. Battle between influenza A virus and a newly identified antiviral activity of the PARP-containing ZAPL protein. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 14048-14053, doi:10.1073/pnas.1509745112 (2015).

124 Tang, Q., Wang, X. & Gao, G. The Short Form of the Zinc Finger Antiviral Protein Inhibits Influenza A Virus Protein Expression and Is Antagonized by the Virus-Encoded NS1. *Journal of virology* **91**, doi:10.1128/jvi.01909-16 (2017).

125 Mueller, S., Papamichail, D., Coleman, J. R., Skiena, S. & Wimmer, E. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *Journal of virology* **80**, 9687-9696, doi:10.1128/JVI.00738-06 (2006).

126 Stauft, C. B. *et al.* Extensive recoding of dengue virus type 2 specifically reduces replication in primate cells without gain-of-function in Aedes aegypti mosquitoes. *PloS one* **13**, e0198303, doi:10.1371/journal.pone.0198303 (2018).

127 Manokaran, G., Sujatmoko, McPherson, K. G. & Simmons, C. P. Attenuation of a dengue virus replicon by codon deoptimization of nonstructural genes. *Vaccine* **37**, 2857-2863, doi:10.1016/j.vaccine.2019.03.062 (2019).

128 Yu, C. H. *et al.* Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular cell* **59**, 744-754, doi:10.1016/j.molcel.2015.07.018 (2015).

13 List of publications

Groenke, N., Trimpert, J., Merz, S., Wyler, E., Zhang, H., Hazapis, O., Rausch, S., Landthaler, M., Osterrieder, N., Kunec, D. (2020). Mechanism of virus attenuation by codon pair deoptimization. Cell Reports. Accepted.

Groenke, N., Trimpert, J., Merz, S., Wyler, E., Zhang, H., Hazapis, O., Rausch, S., Landthaler, M., Osterrieder, N., Kunec, D. (2020). Synthetic recoding of viral genomes - codon versus codon pair (de)optimization. Manuscript in preparation.

Diez, L., Wiglenda, T., Secker, C., Brusendorf, L., Trepte, P., **Groenke, N.**, Neuendorf, N., Purfürst, B., Reissmann, S., Eilitz, U., Schnoegl, S., Boeddrich, A., Wanker EE. (2020). The small molecule LT1 mediates the formation of a polymorphic amyloid-beta fibril variant with low stability and seeding activity. Submitted.

Wiglenda, T., **Groenke, N.**, Hoffmann, W., Manz, C., Diez, L., Buntru, A., Brusendorf, L., Neuendorf, N., Schnoegl, S., Haenig, C., et al. (2020). Sclerotiorin Stabilizes the Assembly of Nonfibrillar Abeta42 Oligomers with Low Toxicity, Seeding Activity, and Beta-sheet Content. Journal of molecular biology.

Trimpert, J., **Groenke, N.**, Kunec, D., Eschke, K., He, S., McMahon, D.P., and Osterrieder, N. (2019b). A proofreading-impaired herpesvirus generates populations with quasispecies-like structure. Nature microbiology 4, 2175-2183.

Kheimar, A., Trimpert, J., **Groenke, N.**, and Kaufer, B.B. (2019). Overexpression of cellular telomerase RNA enhances virus-induced cancer formation. Oncogene 38, 1778-1786.

Boeddrich, A., Babila, J.T., Wiglenda, T., Diez, L., Jacob, M., Nietfeld, W., Huska, M.R., Haenig, C., **Groenke, N.**, Buntru, A., et al. (2019). The Anti-amyloid Compound DO1 Decreases Plaque Pathology and Neuroinflammation-Related Expression Changes in 5xFAD Transgenic Mice. Cell chemical biology 26, 109-120.e107.

Dayaram, A., Tsangaras, K., Pavulraj, S., Azab, W., **Groenke, N.**, Wibbelt, G., Sicks, F., Osterrieder, N., and Greenwood, A.D. (2018). Novel Divergent Polar Bear-Associated Mastadenovirus Recovered from a Deceased Juvenile Polar Bear. mSphere 3(4).

Trimpert, J., **Groenke, N.**, Jenckel, M., He, S., Kunec, D., Szpara, M.L., Spatz, S.J., Osterrieder, N., and McMahon, D.P. (2017). A phylogenomic analysis of Marek's disease virus reveals independent paths to virulence in Eurasia and North America. Evolutionary applications 10, 1091-1101.

Groenke, N., Seisenbaeva, G.A., Kaminskyy, V., Zhivotovsky, B., Kost, B., and Kessler, V.G. (2012). Structural characterization, solution stability, and potential health and environmental effects of the Nano-TiO2 bioencapsulation matrix and the model product of its biodegradation TiBALDH. Rsc Adv 2, 4228-4235.

14 Acknowledgement

First of all, I want to thank Prof. Dr. Nikolaus Osterrieder who gave me the opportunity to work in his institute and who supported me constantly throughout my PhD.

A special thanks to Dusan Kunec who supervised me and invented my great project. He guided me successfully through my PhD and continuously supported me with valuable advice and discussions.

I want to acknowledge my colleagues Netti, Jakob, Chris, Timo, Luca, Andelé, Kathrin, Michaela, Ann, Kia, Darren, Walid, Viviane, Mohamed, Ahmed, Na, the animal caretakers and all the other members at the Institute of Virology who supported me, discussed experiments and assisted me with their expertise. I also want to thank Michael Veit and Benedikt Kaufer for their mentorship, their constructive criticism and suggestions. All of you created an environment that made me happy working here. I especially enjoyed working next to Netti, who spread positive energy in the whole lab and assured that my work never got boring. During my PhD, I developed not only professional skills but I also improved my social skills by learning a lot about different cultures in our international lab.

I am very pleased to have the opportunity to collaborate with the group from Markus Landthaler at the Max Delbrück Center for Molecular Medicine. His group supported me with their technical knowledge and experience in ribosome profiling. Therefore, a special thanks to all the helpful lab members, especially to Emanuel Wyler and Mathias Kaiser.

I am also very thankful for the groundbreaking meeting with Florian Heyd, who guided me in the right direction that led to the analysis of RNA synthesis and stability experiments.

Not to forget, I am grateful for the financial support from ZIBI/ IMPRS and the Elsa-Neumann-Stipendium. I am glad to be part of the ZIBI network and thankful for the opportunity of taking part in the variety of interesting courses and advanced trainings that were offered from ZIBI and the DRS.

Finally, big thanks to my whole family and friends for all your support and always being there in any situation of life.

15 Selbständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig und ausschließlich unter Zuhilfenahme der genannten Quellen und Hilfen angefertigt habe.

Berlin, den 24.02.2020 Nicole Groenke