Aus der Kinderklinik und Kinderpoliklinik im Dr. von Haunerschen Kinderspital
der Ludwig-Maximilians-Universität München

Direktor: Prof. Dr. Dr. med. Christoph Klein

Minimalistic 5'-UTRs and segmented poly(A) tails – a step towards increased potency of transcript therapies

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften an der Medizinischen Fakultät der Ludwig-Maximilians-Universität zu München

Željka Trepotec
aus Koprivnica, Kroatien

2019

Mit Genehmigung der Medizinischen Fakultät
der Universität München

Betreuer:	PD. Dr. Carsten Rudolph
Zweitgutachter:	Prof. Dr. Andreas Ladurner
Prüfungsausschuss:	Prof. Dr. Peter B. Becker
	Prof. Dr. Alexander Faussner
Dekan:	Prof. Dr. med. dent. Reinhard Hickel

Tag der mündlichen Prüfung: September 27, 2019

Eidesstattlicher Versicherung

Trepotec, Željka

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Dissertation mit dem Thema

Minimalistic 5'-UTRs and segmented poly(A) tails – a step towards increased potency of transcript therapies

selbständig verfasst, mich außer der Angegebenen keine weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

München, 16.10.2019

_____

Željka Trepotec

To my beloved parents.

“Da mi je samo znati, Draš, kaj ti zaprav misliš, gda si kaniš poštenoga posla najti?“

-“Ah, kaj ja znam Regica… Sem čuda put probal al sigde za kvalifikacije pitaju.”

Božja vola, prva priča

Gruntovčani

SUMMARY

Studies in the past decade have brought to light the therapeutic potential of transcript therapies as an alternative to classical gene therapy. In contrast to DNA, which has to be delivered to the nucleus of the cell to assure the production of the mRNA transcript and its translation in the cytoplasm, mRNA's target delivery location is the cytoplasm. Thus, unlike in DNA-based gene therapies, there is no risk of insertional mutagenesis when using mRNA-based transcript therapies. Transient translation of delivered mRNA is associated with its attributes such as degradation through physiological pathways. Apart from that, intrinsic immunogenicity is another attribute normally associated with mRNA. These challenges have been overcome by optimizing the structural elements (e.g. untranslated regions, poly(A) tail and/or usage of chemically modified nucleotides). Chemically modified mRNAs, therefore, emerge as a safer promising tool for transcript therapy.

The first part of the thesis is focused on the design of synthetic minimalistic 5'-UTRs containing sequence elements essential for mRNA production using T7 RNA polymerase and their efficient translation. Such synthetic UTRs were benchmarked against some of the conventionally used UTR sequences, such as human alpha-globin. To achieve the goal of finding the very minimal synthetic 5'-UTR, harboring features of sufficient expression but resistant to structural changes upon chemical modification, following approaches were investigated. Several luciferase-encoding unmodified and chemically modified mRNAs harboring minimal synthetic 5'-UTRs were investigated upon transfection in cell culture and *in vivo* in mice. A particularly short 5'-UTR, consisting of only 14 nucleotides between the T7 promoter and Kozak consensus sequence, yielded similar or even higher expression than a 37 nucleotides long human alpha-globin 5'-UTR containing mRNA in HepG2 and A549 cells. The functionality of translation regulators, namely Kozak and TISU element, were affected by the choice of modified nucleotides. Minimal 5'-UTRs identified in the *in vitro* experiments also performed better than the human alpha-globin 5'-UTR *in vivo*, what was confirmed by bioluminescence imaging of luciferase expression in mouse livers at 6h post-intravenous injection of a lipidoid nanoparticle-formulated RNA in female Balb/c mice. To rule out sequence-specific effects,

promising UTRs were also combined with hEPO. Minimal TISU-containing 5'-UTR hEPO mRNA reached elevated levels in supernatants from A549 cells post-transfection when compared to commonly used 5'-UTR benchmarks. Taken together, two promising candidates of synthetic minimalistic 5'-UTRs for use in transcript therapies were identified.

In the second study, the focus was set on poly(A) tail, a common structural feature of all cellular messenger RNAs (with the exception of histone protein-coding mRNAs). As therapeutic mRNA is transcribed *in vitro*, poly(A) can either be brought *via* the DNA template (PCR or plasmid-based) or added to the mRNA enzymatically, in a separate post-transcriptional step. A plasmid containing poly(A) tail enables easier and cheaper template production, but such homopolymeric regions recombine in *E.coli*, causing extensive shortening of the poly(A) tail. The segmented poly(A) approach, defined as at least two A-containing elements, each defined as a nucleotide sequence consisting of 40 – 60 adenosines, separated by a spacer element of different length, was investigated with respect to recombination of plasmids in *E.coli* and the potency of the poly(A)-containing mRNAs. Incorporation of those resulted in significant reduction of plasmid recombination in *E.coli,* without any negative effect on either mRNA half-life or protein expression. The effect was confirmed for sequences of varying length: d2EGFP (0.9 kb), firefly luciferase (1.7 kb), hEPO (0.9 kb) and CFTR (4.5 kb), proving the adaptability of the segmented poly(A) approach with 6 nucleotides as a spacer for different therapeutic targets (intracellular proteins: d2EGFP and luciferase; secretory protein: hEPO and transmembrane protein: CFTR). Promising results with physiological genes led to further investigation of the effect of the increased spacer length on the performance of segmented $poly(A)_{2x60}$.

Luciferase constructs with longer spacers of 12 and 24 nucleotides revealed higher translation efficiencies compared to widely used homogeneous poly(A) tail of 120 adenosines, preceded with reduced recombination rates in *E.coli*. Reducing the spacer length to a single nucleotide (either C, G or T) resulted in higher luciferase expression, and particularly the incorporation of a single G further reduced poly(A) recombination to zero. Significantly reduced recombination in *E.coli* with either

comparable or higher translation efficiencies of the resulting mRNA presents segmented poly(A) as an attractive alternative to the classical homogeneous poly(A)>100 bp, thereby facilitating plasmid-based template production. Results obtained in the current work allow the conclusion that segmented poly(A)$_{2x60}$ with either 6 or 1 nucleotide spacer, in plasmid vectors, significantly reduces recombination in *E.coli* without negatively affecting translation and mRNA stability when compared to the widely used poly(A)$_{120}$.

ZUSAMMENFASSUNG

Studien im letzten Jahrzehnt haben das therapeutische Potential der Transkript-Therapie als Alternative zur klassischen Gentherapie deutlich gemacht. Im Gegensatz zum DNA-basierten Ansatz, bei dem für die mRNA Produktion der Zellkern erreicht werden muss, müssen mRNAs lediglich ins Zytoplasma transportiert werden. Demzufolge besteht bei der Verwendung von mRNA Transkripten, im Gegensatz zu DNA-basierten Therapien, kein Risiko einer Insertionsmutagenese. Herkömmliche Probleme die der mRNA zugeordnet werden sind die intrinsische Immunogenität und kurzlebige Translation der gelieferten mRNA aufgrund ihres physiologischen Abbaus. Diese Herausforderungen können durch Optimierung der mRNA Strukturelemente (z.B. nicht translatierte Regionen, poly(A)-Schwanz und/oder Verwendung chemisch modifizierter Nukleotide) überwunden werden. Zum Beispiel sind chemisch modifizierte mRNAs ein sichereres und vielversprechenderes Werkzeug für die Transkript-Therapie.

Der erste Teil dieser Doktorarbeit konzentriert sich auf das Design und die effiziente Translation synthetischer minimalistischer 5'-UTRs, welche Sequenzelemente für die mRNA-Produktion unter Verwendung von T7-RNA-Polymerase enthalten. Diese synthetischen UTRs wurden mit herkömmlich verwendeten UTR-Sequenzen wie der humanem Alpha-Globin-UTR verglichen. Um minimale synthetische 5'-UTR mit ausreichender Expression zu finden, welche durch chemischen Modifikationen gegen strukturelle Veränderungen resistent sind, wurden folgende Ansätze untersucht. Mehrere Luciferase-kodierende, unmodifizierte und chemisch modifizierte mRNAs mit minimalen synthetischen 5'-UTRs wurden nach Transfektion *in vitro* und *in vivo* in Mäusen untersucht. Eine besonders kurze 5'-UTR, die aus nur 14 Nukleotiden zwischen dem T7-Promotor und der Kozak-Konsensussequenz besteht, ergab eine ähnliche oder sogar höhere Expression als eine humane alpha-Globin-5'-UTR mRNA mit 37 Nukleotiden in HepG2- und A549-Zellen. Die Funktionalität der Translationsregulatoren, nämlich der Kozak- und TISU-Elemente, wurde durch die Wahl der modifizierten Nukleotide beeinflusst. Minimale 5'-UTRs, die in den *in vitro*-Experimenten identifiziert wurden, zeigten ebenfalls *in vivo* eine Verbesserung gegenüber der humanen Alpha-

Globin-5'-UTR, was durch die Luciferase-Expression in Mausleber 6 Stunden nach intravenöser Injektion einer Lipidoid-Nanopartikel-formulierten mRNA in weiblichen Balb/c-Mäusen bestätigt wurde. Um sequenzspezifische Effekte auszuschließen, wurden auch vielversprechende UTRs mit hEPO kombiniert. Im Vergleich zu üblicherweise verwendeten 5'-UTR-Benchmarks wurde mit minimaler TISU, 5'-UTR-hEPO mRNA ein höherer Gehalt in Überständen von transfizierten A549-Zellen festgestellt. Zusammenfassend kann festgehalten werden, dass zwei vielversprechende Kandidaten synthetisch minimalistischer 5'-UTRs zur Verwendung in Transkript-Therapien identifiziert wurden.

Im zweiten Teil lag der Fokus auf dem poly(A)-Schwanz, einem Strukturmerkmal aller zellulären mRNA (mit Ausnahme der mRNA-codierenden Histonproteine). Da therapeutische mRNA *in vitro* transkribiert wird, kann der poly(A)-Schwanz entweder über die DNA-Matrize (auf PCR- oder Plasmidbasis) oder enzymatisch in einem separaten posttranskriptionellen Schritt zur mRNA hinzugefügt werden. Plasmide, die einen poly(A)-Schwanz enthalten ermöglichen eine einfachere und billigere Matrizenherstellung, jedoch neigen diese homopolymeren Regionen zur Rekombination in *E. coli*, was zu einer starken Verkürzung des poly(A)-Schwanz führt. Ein segmentierter poly(A)-Ansatz (mindestens zwei A-haltige Elemente, jeweils definiert als Nukleotidsequenz bestehend aus 40 - 60 Adenosinen, getrennt durch ein Spacerelement unterschiedlicher Länge) wurde hinsichtlich der Rekombination von Plasmiden in *E. coli* und die Aktivität der poly(A)-haltigen mRNAs untersucht. Der Einbau von diesen führte zu einer signifikanten Reduktion der Plasmidrekombination in *E. coli*, ohne dass die mRNA-Halbwertszeit oder die Proteinexpression negativ beeinflusst wurden. Die Wirkung wurde für Sequenzen unterschiedlicher Länge bestätigt: d2EGFP (0,9 kb), Firefly-Luciferase (1,7 kb), hEPO (0,9 kb) und CFTR (4,5 kb), was die Anpassungsfähigkeit des segmentierten poly(A)-Ansatzes mit 6 Nukleotiden als ein Spacer für verschiedene therapeutische Ziele zeigt (intrazelluläre Proteine: d2EGFP und Luciferase; sekretorisches Protein: hEPO und Transmembranprotein: CFTR). Vielversprechende Ergebnisse mit physiologischen Genen führten zu weiteren Untersuchungen des Einflusses der erhöhten Spacerlänge auf die Leistung von segmentiertem poly(A)2x60. Luciferase-

Konstrukte mit längeren Spacern von 12 und 24 Nucleotiden zeigten höhere Translationseffizienzen gegenüber einem weit verbreiteten homogenen poly(A)-Schwanz von 120 Adenosinen, denen in *E. coli* reduzierte Rekombinationsraten vorausgingen. Das Reduzieren der Spacer-Länge auf ein einzelnes Nukleotid (entweder C, G oder T) führte zu einer höheren Luciferase-Expression, insbesondere wurde durch den Einbau eines einzelnen G die poly(A)-Rekombination vermieden. Durch die signifikant reduzierte Rekombination in *E. coli* mit vergleichbarer oder höherer Translationseffizienz der resultierenden mRNA stellt ein segmentierter poly(A)-Schwanz eine attraktive Alternative zu dem klassischen homogenen poly(A)> 100 bp dar und erleichtert so die Plasmid-basierte Matrizen-Produktion. Die in der aktuellen Arbeit erzielten Ergebnisse lassen den Schluss zu, dass segmentiertes $poly(A)_{2x60}$ mit entweder 6 oder 1 Nukleotid-Spacer im Vergleich zu einem weit verbreiteten $poly(A)_{120}$ die Rekombination der Plasmidvektoren in *E. coli* signifikant reduziert ohne die Translation und die mRNA-Stabilität negativ zu beeinflussen.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A	adenine
ACH	poly(A) – poly(C) - histone stem loop
AMP	adenosine 5'-monophosphate
AUG	start codon
C	cytosine
CFTR	cystic fibrosis transmembrane (conductance) regulator
CMP	cytidine 5'-monophosphate
CMV	cytomegalovirus
CYBA	cytochrome b-245 alpha chain
d2EGFP	destabilized enhanced green fluorescent protein
DMEM	Dulbecco`s Modified Essential Medium
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
EGFP	enhanced green fluorescent protein
eIF	eukaryotic initiation factor
ELISA	enzyme-linked immunosorbent assay
FA	fragment analyzer
FACS	fluorescent-activated cell sorting
FBS	fetal bovine serum
G	guanine
GMP	guanosine 5'-monophosphate
GMP	good manufacturing practice
GOI	gene of interest
h	hours
hACE2	human angiotensin-converting enzyme part 2
HCl	hydrochloric acid
hEPO/EPO	human erythropoietin / erythropoietin
IRES	internal ribosome entry site
IVT	*in vitro*
IVT	*in vitro* transcribed/*in vitro* transcription
kb	kilobase
LNP	lipid nanoparticles

Luc2	luciferase
m7G	methyl-7-guanosine / 7-methylguanosine
MCS	multiple cloning sites
min	minutes
miRNA	micro RNA
mRNA	messenger ribonucleic acid
nt	nucleotides
ORF	open reading frame
PABP	poly(A) binding proteins
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
Poly(A)	polyadenosine
qPCR	quantitative PCR
RIG-I	retinoic acid inducible gene I
RLU	relative light units
RNA	ribonucleic acid
RPMI	Roswell Park Memorial Institute
SD	standard deviation
siRNA	small interfering RNA
T	thymine
TISU	translation initiator of short UTRs
TLR	Toll-like receptors
U	uridine
UMP	uridine 5’-monophosphate
UTR	untranslated region
°C	degree Celsius
β-gal	beta-galactosidase
ΔG	Gibbs free energy

1 INTRODUCTION

1.1. The RNA world and the origins of life

Although the function of modern organisms is orchestrated by biomolecules, namely proteins, lipids, carbohydrates, DNA and RNA, there is a compelling hypothesis that the earliest scions of life stored and propagated genetic information only *via* RNA molecules.[1] In this conceptual idea called RNA world[2–4], the primordial RNA genome independently replicated in a cooperative manner that set up the basics for the complexity during the early evolution of life, in a distant evolutionary past, approximately 4 billion years ago (Figure 1)[4]. The key molecule that could have supported such an evolving fundamental genetic system with the defining hallmarks of life, heredity, and evolution, was the RNA. With the ground-breaking discovery of ribonuclease P catalysts[5], self-splicing introns[6] in the 1980s and later RNA-catalyzed RNA polymerization[7], it was proved that RNA also has an additional, enzymatic property. Being able to store genetic information and catalyze peptide bond in the ribosome[8,9] as well as the transcription of an active ribozyme[10], it is a legitimate candidate for such a hypothesis.

Later, as the evolution of new biomolecules and selective pressure at the molecular level took place, there was a role splitting for features that solely RNA harbored. DNA took over the role of a template molecule of genetic material[11], due to a higher replication fidelity, greater stability under hydrolytic conditions[12] and rigidness to prevent self-folding[13]. Proteins, however, found a niche in driving enzymatic and chemical reactions, due to the whole spectrum of their structural diversity.[14] However, the role of an intermediary of genetic information transfer still belongs to the RNA. As it

used to be that the RNA gene is instructed for the synthesis of a corresponding ribozyme, now the same mechanism is adapted in the process of genome replication.[14] Therefore, the RNA molecules can be considered as molecular fossils.

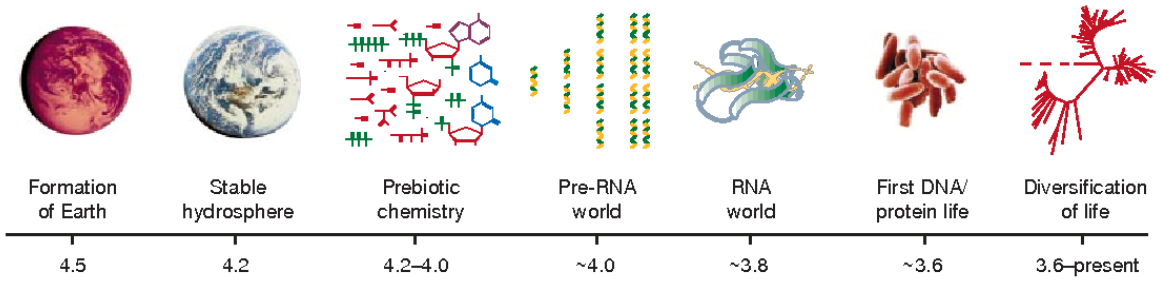

Figure 1. Timeline of the early events in Earth's history, dating billions of years ago.

The RNA world supposedly appeared ~3.8 billion years ago. Figure adapted from Joyce, 2002.[4]

1.2. Messenger RNA

This molecular fossil, the molecule of RNA which persevered from the far past till today due to its properties, is still the central molecule of all forms of life.[15] It is the core of the central dogma of molecular biology, enunciated by F. Crick in 1958.[16] Accepting the premise that DNA provides the coding information for genes transcribed into intermediary molecule, which is then translated into protein, the conclusion that follows is that the genes are the "blueprint for life".[17,18] That intermediary molecule, a portion of the total cellular RNA carrying genetic information from nuclear DNA to cytoplasmic ribosomes, was named "messenger RNA" (mRNA) by Francis Jacob and Jacques Monod in 1961.[19]

mRNA carries its role as a single-stranded molecule, with a possibility to form great structural diversity just by folding onto itself.[20] Therefore, it harbors many cellular functions, although the most important one is the transmission of information. It consists of four building blocks: adenosine 5'-monophosphate (AMP), guanosine 5'-monophosphate (GMP), cytidine 5'-monophosphate (CMP) and uridine 5'-monophosphate (UMP).[21] Eukaryotic mRNAs, with some exceptions, harbor cap at 5' end, a residue of 7-methylguanosine linked to the 5'-terminal residue of the molecule by 5'-5'-

triphosphate bond. It protects the molecule from ribonuclease cleavage but also enables recognition of the ribosome which leads to initiation of the translation.[22–24] Cap also plays an important role in the export of mRNA from the nucleus.[24,25]

At the 3' end, there is a string of 80 to 120 adenosines, forming the protective poly(A) tail. Among the protective role of the poly(A) against enzymatic destruction, it also helps to stabilize the molecule[26], exporting it from the nucleus[27], localizing it intracellularly and initiating the translation[28,29]. Both protective structures at each end of the naturally occurring mRNAs are attached enzymatically, during or shortly after transcription. The coding region is a central part of the mRNA molecule; composed of codons that are in turn translated into functional proteins. This core of the molecule is flanked by an untranslated region (UTR) from each side. As the name suggests, they are not translated, but their sequence defines their function in the molecule: altering the translation efficiency, mRNA stability and localization.[30–35] The schematic representation of the structural elements of an mRNA molecule is shown in Figure 2.

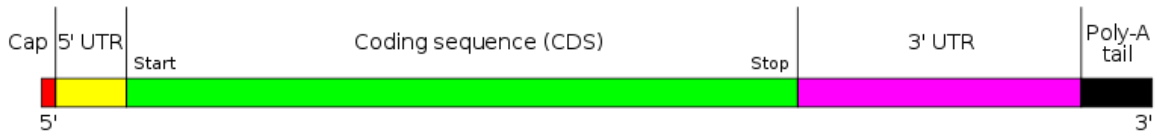

Figure 2. The structural elements of a mature mRNA transcript.

It is composed of the m7G cap, followed by 5'-UTR, coding sequence to which 3'-UTR is continuing and at the end, poly(A) tail. Figure adapted from Daylite 2008. [36]

Events leading to gene expression are highly regulated at the cellular level. Besides transcriptional control, a fundamental mechanism controlling gene expression is at the post-transcriptional level mediated through the multiple interactions between the regulatory elements of the mRNA and RNA-binding proteins. With respect to the mRNA structure, the numerous control points are the elements found within the 5'-UTR, where regulatory proteins can bind and inhibit ribosomal scanning.

As one of the first examples was that highly stable secondary structure can inhibit translation, but also structures with lower thermal energy if positioned proximally to m7G cap.[37,38] Besides, translational efficiency was reduced with increasing GC content in 5'-UTR, if the distance to the cap and predicted thermal stability were held constant.[39]

In the conventional cap-dependent translation, sequence elements have been described that support cap-independent translation of their respective genes.[40,41] Another structural element found more commonly in viral genomes is the internal ribosome entry site (IRES element). It is widely believed that inclusion of such elements at the 5'-UTR, specifically IRES, initiates cap-independent mRNA translation by a direct ribosomal recruitment.[31,42] However, due to the lack of proves to determine genuine IRES elements, it is impossible to recognize common features that would explain how internal translation is achieved.[43,44]

The other end of the mRNA molecule, namely the 3'-UTR is also a target for regulation. Sequence and/or structural elements within are recognized both by proteins (e.g. AU-rich element binding protein)[45] and RNAs (e.g. microRNAs)[46–49]. Depending upon the RNA binding partner, it can lead to either activation or repression of translation. RNA-binding proteins often alter translation *via* poly(A) tail, or other proteins that bind at the cap structure (e.g. initiation factor eIF4E). A small, micro RNA (miRNA) control gene expression either by repression of translation or by supporting mRNA degradation.[50] On a global cellular level, gene translation is achieved through phosphorylation of the initiation factors, e.g. eIF2 and eIF4E, and by modulating their binding affinities.[51] The overview is schematically shown in Figure 3. With that cognition, a whole new world of research into causes of diseases resulting from translational dysregulation could be established.

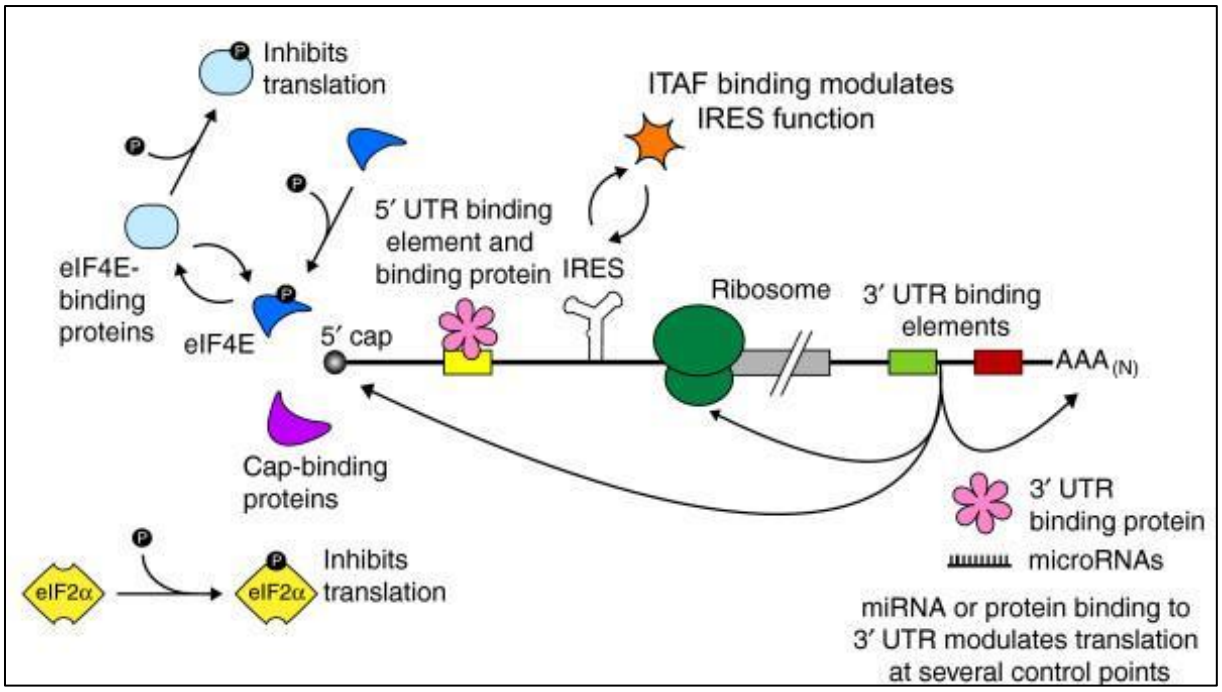

Figure 3. Regulation of eukaryotic mRNA translation occurs at numerous control points.

At the 5'-UTR, protein expression is regulated by binding elements and proteins. At the 3' end, poly(A) is also participating in the regulation of translation by PABP binding, which also participates in an interaction with cap-binding proteins. Figure adapted from Baker 2006. [51]

1.3. Transcript therapy

With the knowledge about the structure and function of the mRNA gained over time, we are now approaching its translation into a new class of therapeutics. Considering their great potential as a method of gene delivery, it could be used to cure genetic deficiencies and a whole spectrum of inherited diseases by the expression of therapeutic proteins. Transcript therapy, by definition, utilizes RNA as a therapeutic molecule, whether it be mRNA or small interfering RNA (siRNA), performing its function directly in the cytoplasm. This is just one of the conceptual advantages of transcript therapies over classical DNA-based gene therapy; the delivery method must overcome just one barrier instead of two (cytoplasm and nuclear lamina) to release the content. As the niche of DNA-based gene therapy is indeed nucleus, there is also a risk of integration into the chromosomes *via* random insertional mutagenesis[52,53], which has resulted in cancerous outcome in the past studies[54,55]. Moreover, mRNA transcripts directly utilize cellular ribosomes as bioreactors for establishing the

expression of the protein of interest.[56] As the cell itself produces the protein of interest, correct post-translational modifications are ensured, thereby reducing the immunogenicity of the therapeutic protein. This is the advantage compared to recombinant protein approach, where recombinant proteins expressed in heterologous systems can be recognized as foreign antigens (due to the lack of desirable post-translational modification(s)), therefore causing a systematic immune response by the host.[57] Moreover, some proteins, e.g. membrane proteins or intra-organelle proteins may be difficult/impossible to produce in heterologous systems or their delivery to target cells/organelles may be limiting. Protein expression in mRNA transcript therapy is transient, due to enzymatic mRNA degradation over time, which makes it more attractive as a therapeutic agent, especially when producing antigenic proteins for vaccinations or cancer immunotherapeutics.[58–62] mRNA-based genome editing tools are not only potent but offer safety advantages over protein-based or virus-encoded alternatives.[63]

The very first proof of concept that mRNA encoding a reporter protein has a potential therapeutic application was demonstrated by Wolf et al in 1990, with an intramuscular injection of mRNA into the mouse.[64] However, only during the recent years, the field of mRNA therapeutics has progressed to clinical applications. From then until now, it moved from the research to the clinical trials, especially in the field of cancer vaccines.[65–67] Improvements in mRNA delivery and the reduction of the mRNA's immunogenicity enabled its diverse therapeutic applications. Examples include, but are not limited to induced immunity against H10N8 and H7N9 influenza viruses[68], protection against Zika virus-induced congenital disease[69], regenerative treatment for heart failure, diabetic wound healing, and vascular diseases[70]. The latter has progressed to a first clinical trial.[71] As for the bone healing, promising options are being developed.[72,73] With the mRNA technology platform, there is a possibility to induce pluripotent stem cells by delivering mRNA-encoded transcription factors.[52] In preclinical models of human hereditary metabolic diseases, such as Crigler-Najjar syndrome, alpha-1-antitrypsin deficiency, thrombotic thrombocytopenic purpura, urea cycle disorder, mRNA transcript therapy is being proved as a lifesaving alternative to existing symptom-relief therapeutics.[56,74–76] The timeline of key discoveries and clinical advances in the mRNA therapeutic field is shown in Figure 4.

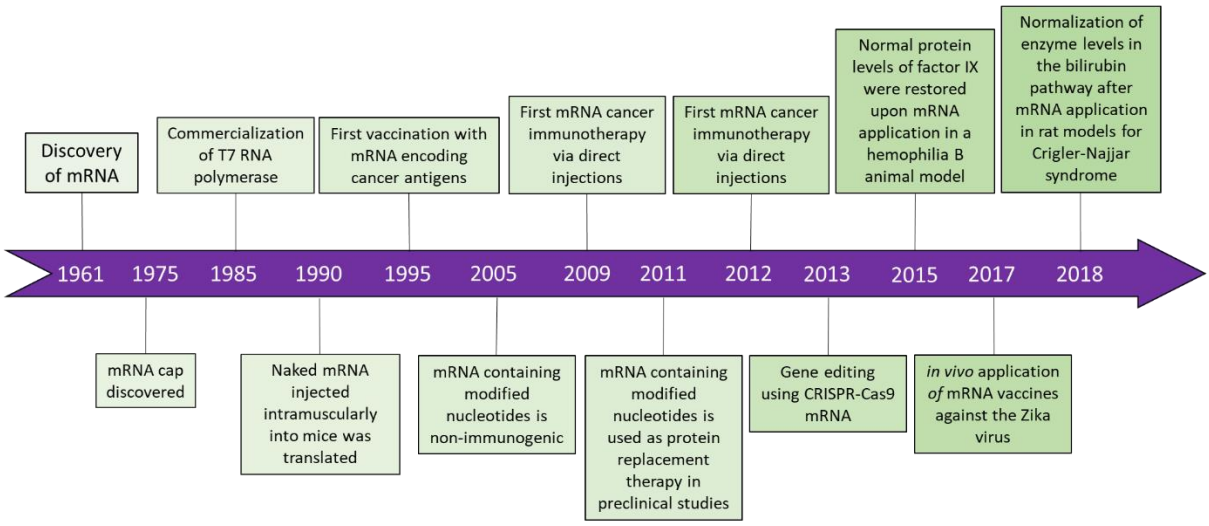

Figure 4. Key discoveries and advances in the development of mRNA as a drug technology.

The first demonstration of translation of naked mRNA injected into muscle tissue of mice was shown in 1990. A discovery that IVT mRNA with incorporated modified nucleosides is non-immunogenic happened in 2005. In 2012, a protective vaccination with flu-specific mRNAs entered the preclinical studies. In vivo application against virus infections was shown in 2017.

Despite the number of preclinical and clinical trials in progress, there is a need to overcome barriers related to mRNAs therapeutic potential and safety. To enhance its therapeutic potential, numerous approaches have been undertaken to reduce the innate immune activation upon delivery.[58,77] Significant reduction of mRNAs immunogenicity has been obtained with the incorporation of modified nucleotides.[58] Kariko showed that the incorporation of pseudoU into the RNA reduces the immunogenicity of the RNA molecule.[78] Following this idea, it was shown that synergistic decrease of mRNA binding to TLR3, LR7, TLR8 and RIG-I in human peripheral blood mononuclear cells could be achieved by the replacement of only 25% of uridines and cytidines with 2-thiouridine and 5-methylcytidine, respectively.[77] In contrast to the incorporation of modified nucleotides, Thess et al demonstrated the potential of sequence optimization wherein, the optimized sequences resulted in higher transgene expression and lower immune response when delivered as unmodified mRNA compared to its modified counterpart.[79] To improve the translational efficiency and intracellular stability of the delivered mRNA, research efforts are focused on every structural feature of the molecule. These are presented schematically as Figure 5[80].

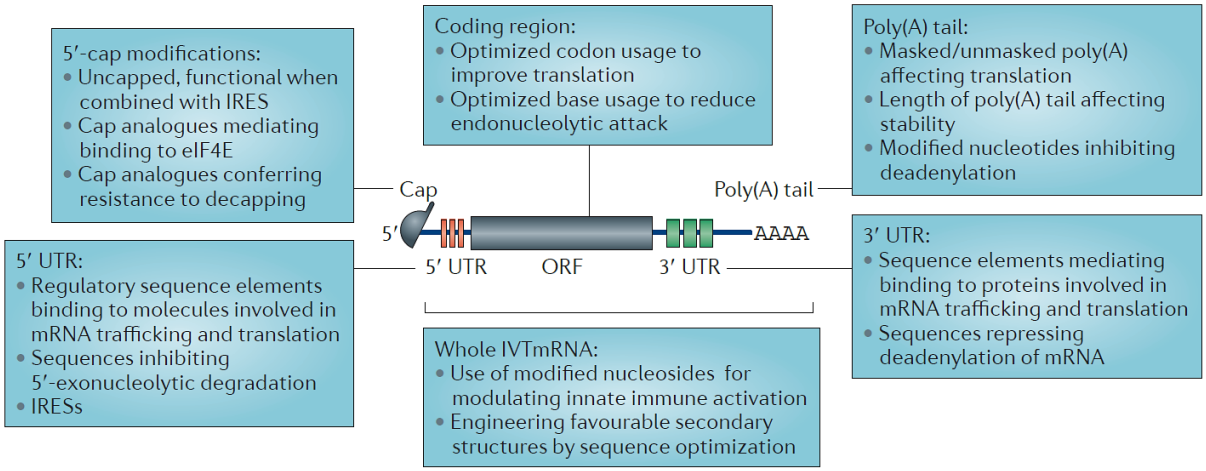

Figure 5. Structural modifications for tuning mRNA pharmacokinetics.

The 5′-UTR region can be modulated by incorporating IRES elements, other regulatory sequence elements which attract molecules involved in translation and trafficking or inhibiting RNA degradation. By changing the length of poly(A) tail, one can regulate mRNA stability and translation efficiency. Figure adapted from Sahin et al, 2014. [80]

2 AIMS AND PERSPECTIVES

The current thesis was aimed at designing potent mRNA molecules with a minimalistic approach. To this end, investigations were made into two of the critical structural elements, namely 5'-UTR and poly(A) tail. The first part of the dissertation focuses on 5'-UTR. Series of synthetic minimalistic UTRs were designed based on a rational approach to include only the essential elements needed for *in vitro* transcription by T7 RNA polymerase and efficient intracellular translation of RNA. A universal synthetic UTR, regardless of the encoding sequences, ideally has characteristics of being non-immunogenic and without bearing any secondary structures, which could theoretically inhibit scanning mechanism. The application of minimalistic UTRs is described in Chapter 4. This part of the thesis has been published as "*Maximizing the translational yield of mRNA therapeutics by minimizing 5'-UTRs*" in Tissue Engineering Part A, in April 2018 (DOI: 10.1089/ten.TEA.2017.0485).

The 3'-end of the mRNA is elaborated in Chapter 5. The major aim of the undertaken work was to reduce recombination of plasmid vectors containing homopolymeric poly(A) regions in *E.coli* without affecting mRNA stability and translation efficiency. For this, novel poly(A) tail structures, implemented by creating segments of different adenosine lengths and spacers between the regions were investigated. It could be shown that certain segmentation of the poly(A) tail provide either equal or higher levels of protein production, but significantly reduce recombination of plasmid DNA in *E.coli,* without negatively affecting mRNA half-life. This part of the thesis has been accepted for publication as "*Segmented poly(A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life*" in RNA, in December 2019 (ID: RNA/2018/069286).

In Chapter 6, achieved goals and future possible approaches for effective mRNA expression in more complex applications are discussed. We show that, in case of therapeutic applications, synthetic minimalistic 5'-UTR sequences outperform some of the standard, widely used 5'-UTRs such as human alpha-globin. Our results on segmented poly(A) tails demonstrate their superior potential compared to the conventionally used homogeneous poly(A) tails with respect to recombination of the plasmids and the resulting mRNA performance (half-life and translational efficiency). These results taken together represent some of the future possibilities achievable through rational design of mRNA.

3 MATERIALS

In the following chapter, all chemicals, devices, and software, which were used for both projects, are listed and described.

3.1. mRNA sequences

A whole list of coding sequences and UTRs is provided in Appendix, Chapter 9.

3.2. Cell lines

Table 1. Cell lines used

Product	***Supplier***	***Number***
A549 cells	DSMZ	ACC-107
HepG2 cells	DSMZ	ACC-180
HEK293 cells	DSMZ	ACC-305
CFBE cells	University of Münster	n/a
16HBE14o-	University of Münster	n/a

3.3. Cell culture media and supplements

Table 2. Cell culture media and supplements used

Product	*Supplier*
Dulbecco`s Modified Essential Medium (DMEM)/F-12, L-Glutamine, 15 mM Hepes	Gibco Life Technologies
Fetal calf serum (FCS), heat-inactivated	Gibco Life Technologies
Gentamycin	Sigma Aldrich
Minimum Essential Medium (MEM)	GlutaMAX Gibco Life Technologies

Phosphate-buffered saline (PBS)	Gibco Life Technologies
Penicillin/Streptomycin solution	Gibco Life Technologies
Roswell Park Memorial Institute (RPMI) 1640 Medium (1X)+ GlutaMAX™	Gibco Life Technologies
Trypsin/EDTA solution	Sigma
TrypLE™	Gibco Life Technologies

3.4. Transfection Reagents

Table 3. Transfection reagents used

Product	*Supplier*
Lipofectamine® MessengerMAX™	Thermo Fisher Scientific
Lipofectamine® 2000™	Thermo Fisher Scientific

3.5. Kits

Table 4. Assay kits used

Product	*Supplier*
NucleoSpin® Plasmid Miniprep kit	Machery-Nagel
NucleoBond® Xtra Maxi prep kit	Machery-Nagel
NucleoSpin® Gel and PCR Clean-up	Machery-Nagel
Single shot Cell Lysis kit	Bio-Rad
iScript Select cDNA Synthesis kit	Bio-Rad

3.6. Primers

A whole list of primers is provided in Appendix, Chapter 9.

3.7. Antibodies and detection kits

Table 5. Antibodies and detection kits used for Western Blot and staining reagent used for FACS

Antigen	*Catalog #*	*Supplier*
Primary antibodies		
α-CFTR	P13569	R&D Systems
α-Hsp90	TA500494	Origene
Secondary antibodies		
donkey-α-mouse IgG (HRP)	ab6820	abcam
Staining reagents and detection kits		
SuperSignal™ West Femto	34095	Thermo Fisher Scientific
Luminata™ Crescendo Western HRP substrate	WBLUR0500	EMD Millipore / Merck
Luminata™ Classico Western HRP substrate	WBLUC0500	EMD Millipore / Merck
Luminata™ Forte Western HRP substrate	WBLUF0500	EMD Millipore / Merck
Propidium iodide	11348639001	Sigma-Aldrich

3.8. Chemicals

Table 6. Chemicals used for cell culture

Chemical	*Supplier*
Dimethyl sulfoxide (DMSO)	Sigma Aldrich
Ethanol	Roth
GTP	New England Biolabs
HEPES	Gibco Life Technologies
L-Glutamine	Biochrome
Potassium dihydrogen phosphate	Sigma Aldrich
Sodium chloride	Roth
Water for injection (WFI)	B. Braun

Table 7. Chemicals used for in vitro transcription

Chemical	*Supplier*
2-Propanol	Sigma Aldrich
Adenosine triphosphate (ATP)	Sigma Aldrich
Ammonium acetate	Applichem
Bsp119I/BstBI	Thermo Fisher Scientific
Buffer Tango	Thermo Fisher Scientific
Chloroform	Sigma Aldrich
DNAse I	Thermo Fisher Scientific
Ethanol	Roth
Guanosine triphosphate (GTP)	New England Biolabs
Inorganic Pyrophosphatase	Thermo Fisher Scientific
mRNA Cap 2'-o-Methyltransferase	New England Biolabs
Poly(A) Polymerase (E.coli)	NEB
Poly(A) Polymerase Buffer	NEB
RiboLock RNase Inhibitor	Thermo Fisher Scientific
rATP	Jena Biosciences
rGTP	Jena Biosciences
rCTP	Jena Biosciences
rUTP	Jena Biosciences
5-Iodo-rUTP	Jena Biosciences
5-Iodo-rCTP	Jena Biosciences
2-thio-rUTP	Jena Biosciences
5-methyl-rCTP	Jena Biosciences
S-Methyladenosine (SAM)	New England Biolabs
Sodium acetate	Roth
T7 polymerase	Thermo Fisher Scientific
Water for injection (WFI)	B. Braun

Table 8. Other chemicals

Chemical	*Supplier*
1,4-Dithiothreitol (DTT)	Roth
2-Propanol	Sigma Aldrich
2-Mercaptoethanol	Sigma Aldrich
Bovine serum albumin	Sigma Aldrich
NuPage® Antioxidant	Thermo Fisher Scientific

NuPage® Sample Reducing Agent	Thermo Fisher Scientific
NuPAGE® SDS Running Buffer (20X)	Thermo Fisher Scientific
Tris-HCl	Roth
Triton X-100	Triton X-100 Sigma Aldrich
Tween®	Roth

3.9. Consumables

Table 9. Consumables used in this work

Product	*Supplier*
Bolt® 4-12 % Bis-Tris Plus gels (10, 12 and 15 pockets)	Thermo Fisher Scientific
Centrifuge Tube 15 and 50 mL	Corning Incorporated
Corning® 96 Well Black Flat Bottom Polystyrene Not Treated Microplate	Corning Incorporated
Costar™ 96-Well White Plates	Thermo Fisher Scientific
Costar™ cell culture plate, flat bottom, 96, 24 and 6 wells	Corning Incorporated
Costar Microcentrifuge Tube: 0.65 mL, 1.50 mL, 2.00 mL, 5.00 mL	Corning Incorporated
Costar™ Stripette 5, 10, 25, 50 mL	Corning Incorporated
Countess™ Counting Slides	Thermo Fisher Scientific
Diamond® Tipack™ D1200ST, D300ST, D200ST and DL10ST	Gilson
Eppendorf Safe-Lock Tubes 0,65; 1,7; 2,0 mL	Eppendorf
LightCycler®480 Multiwell Plate 96, white	Roche
Luminata Western HRP	Merck Chemicals
T75 Corning® cell culture flasks, angled neck, cap (vented)	Sigma Aldrich
T175 Corning® cell culture flasks, angled neck, cap (vented)	Sigma Aldrich
Trans-Blot® Turbo Transfer Pack Midi 0.2 µm PVDF	Bio-Rad
Vivaspin 20,000 kDa MWCO PES Filter	Sartorius

3.10. Technical Equipment

Table 10. List of technical equipment used in this work

Name	*Device*	*Supplier*
15-300 µL (12 channels), I39816B	Multichannel pipette	Eppendorf
15-300 µL (8 channels), L29846B	Multichannel pipette	Eppendorf
Attune® acoustic focusing cytometer	Flow cytometer	Thermo Fisher Scientific
BoltR® Mini Gel Tank	Gel electrophoresis tank	Thermo Fisher Scientific
Centrifuge 5810R	RNA-Centrifuge	Eppendorf
ChemiDoc™ XRS System	Molecular imager	BIO-RAD
Duomax 1030	Plate shaker	Heidolph Instruments
Fragment Analyzer	Fragment Analyzer	Advanced Analytical Technologies
Hettich Mikro 220	Centrifuge	Hettich Zentrifugen
Invitrogen™ countess automated cell counter	Cell counter	Invitrogen™

IR Sensor MCO-17AIC CO2 Incubator	Cell incubator	Sanyo
IVIS® Lumina XR	Imaging System	Caliper Life Sciences
KDS-210-CE	Dual Syringe Infusion/Withdrawal Pump	KD Scientific
Laminar Flow	Flow BDK	Luft- und Reinraumtechnik GmbH
Leica DMi8	Inverted Light Microscope	Leica Mikrosysteme
Leica DM2000 LED	Light Microscope	Leica Mikrosysteme
LightCycler® 96 Real-Time PCR System	qPCR machine	Roche
Mastercycler® gradient	Thermocycler	Eppendorf
Microcentrifuge 5415R/5415D	Microcentrifuge	Eppendorf
NanoDrop2000 UV-Vis	NanoDrop2000 UV-Vis	Thermo Fisher Scientific
New Brunswick™ Innova® Upright Freezers	Freezer	New Brunswick
Pipetboy acu	Pipette	INTEGRA Biosciences AG
PIPETMAN Classic™: 2, 20, 200 and 1000 µl	Pipettes	Gilson
PowerPac3000	Power supply	BIO-RAD
Reacti-Therm™ III	Heating/Stirring module	Pierce
Refrigerated centrifuge 3K15	Refrigerated centrifuge	Sigma
Tecan Infinite® 200 PRO	Plate reader spectrometer	Tecan
Thermomixer® compact	Thermomixer C	Eppendorf
Trans-BlotR Turbo™ Transfer System	Western blotting transfer system	BIO-RAD
Vortex Genie 2	Vortexer	Scientific Industries
Wallac Victor2 1420 Multilabel counter	Plate reader	PerkinElmer Inc
Water bath model 1003	Water bath	GFL

3.11. Software

Table 11. Software used for measurements and analysis

Product	*Supplier*
Attune® Cytometric Software V2.1	Thermo Fisher Scientific
ChemiDoc™XRS System	BIO-RAD
FlowJo® V10	FlowJo
GraphPad Prism® Version 7	GraphPad Software Inc.
Image Lab™ Software	BIO-RAD
LightCycler® 96 System	Roche
SnapGene Viewer	Snapgene
PROSize 3.0	Advanced analytical

4 DESIGN OF SHORT SYNTHETIC 5'-UTR SEQUENCES

4.1. Background

After interpretation of the results obtained by the huge "human genome project" in 2001, there was a cognition that only a small fraction of genetic material, roughly about 1.5%, is coding for proteins[81]. Realizing that most of the non-coding portion of the genome of higher eukaryotes is involved in regulation of gene expression and not actually coding for proteins, increased research efforts were focused on identifying the key players and their underlying mechanisms of gene regulation.

A systematic and comprehensive study has revealed that the length of 5'-UTR, which are much shorter than 3'-UTR, does not vary greatly among taxa. It ranges between 100 and 200 nucleotides. Its composition of G+C nucleotides is higher in warm-blooded vertebrates, where it accounts for approximately 60% of the total nucleotides. In general, a G+C content of 5'-UTR is higher compared to 3'-UTR.[31]

A strategy that affects the translational efficiency[42], stability[82] of IVT mRNA and their subcellular localization[83], is the incorporation of 5'-UTR regulatory sequences, especially those that have been previously found to modulate these features of endogenous mRNA. In one of the previous studies, five different cellular UTRs for enhanced translation and mRNA stability were compared[84]. Among the compared candidates, human cellular CYBA UTR sequences were reported to increase mRNA translation without affecting the half-life of recombinant RNA transcripts. The functionality of these

regulatory UTRs is embedded not only in their primary sequence but also involves the assembly of the secondary structure which is, in turn, recognized by RNA-binding proteins[31]. As shown in experimental studies, moderately stable 5'-UTR secondary structures (ΔG>-30 kcal/mol), directly preceding start codon, do not stall the scanning of the 40S ribosomal subunit. If, on the contrary, very stable structures are inserted at this strategic position, (ΔG<-30 kcal/mol), inhibitory effects on translation initiation have been reported.[38]

In the case of endogenous mature mRNA, already transported from the nucleus to the cytoplasm, the following event is recognition of the m7G cap structure by the initiation factor eIF4F. This protein complex consists of three subunits: the cap-binding protein (eIF4E), ATP-dependent RNA helicase (eIF4A) and a subunit that binds to other polyadenylate-binding proteins (eIF4G). Firstly, the helicase eIF4A unwinds the RNA, and possibly all secondary structures within, so that the small ribosomal subunit (40S) can easily attach to it.[85] Poly(A) binding protein bound to the poly(A) tail assists in the ribosomal attachment by creating a circular-loop structure together with m7G cap, to facilitate the physical interaction with initiation factor eIF4F.[86] Subsequently, the ribosome scans the mRNA for the first initiation codon (AUG). This ribosomal scanning is the most widely accepted mechanism for translation initiation of eukaryotic mRNA (*Figure 6*). As the initiation takes place at the first AUG codon, 40S ribosomal subunit, after attaching to the mRNA, moves along the mRNA and scans it forward. The context of nucleotides surrounding start codon AUG actually form a consensus, GCCRCCaugG, often called "Kozak consensus".[87] The most conserved nucleotides are positioned at -3 (purine, usually an A) with respect to the AUG start codon, and at +4 (guanine). These preferred nucleotides are conserved in the majority of the 5'-UTR region of animals, plants, and fungi, and may modulate recognition of AUG start codon.[88] Regulation of gene expression through 5'-UTR is tightly controlled, and mutations altering it can have disastrous pathological outcome.[45]

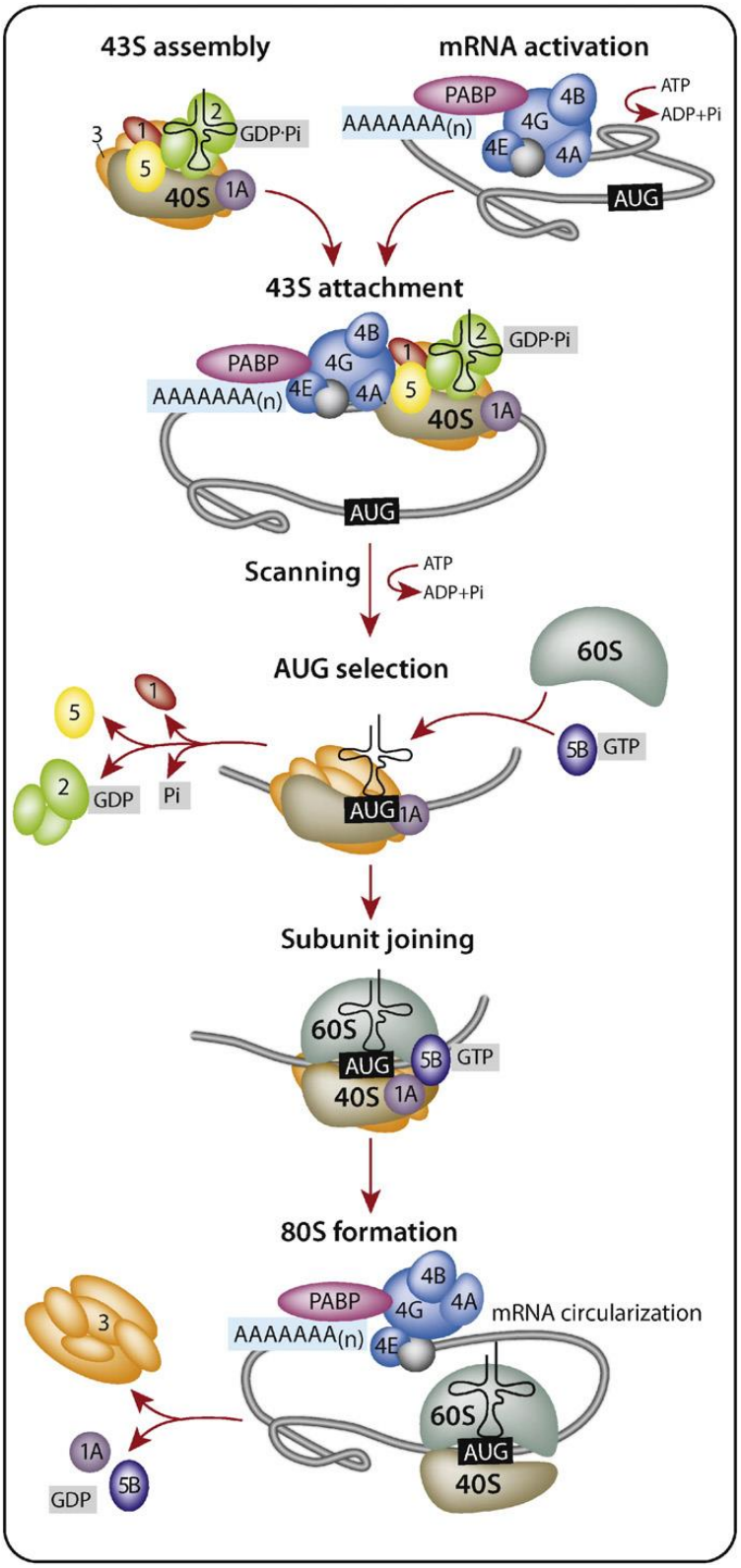

Figure 6. Scheme of the eukaryotic cap-dependent translation initiation mechanism.

The first step of the initiation (on the left) involves assembly of 43S ribosomal complex, consisting of the 40S ribosomal subunit, eIF2, GTP, Met-tRNAi, eIF1A, eIF3, and eIF5. At the same time, mRNA is activated by the unwinding of 5'-UTR region by eIF4A subunit. By positioning a 43S complex at the 5' cap, a 48S pre-initiation complex is formed. A poly(A) tail, containing PABPs interacts with eIF4G1 to circulate the mRNA. An ATP-dependent scanning process in a 5' to 3' direction starts until an AUG codon is found. Then, codon and anticodon are base-paired and Met-tRNAi is incorporating the first amino acid of the polypeptide chain, Met. Upon AUG recognition, a conformational change causes eIF5 to hydrolyze eIF2-GTP, thereby releasing Pi and eIF1. Larger ribosomal subunit, 60S, joins smaller 48S complex mediated by eIF5B and eIF1A, what releases eIF2-GDP, eIF3, eIF4 and eIF5. The newly composed 80S ribosome is now ready for elongation of the polypeptide chain after eIF5B hydrolyzed GTP and released eIF1A. Figure adapted from Haimov et al, 2015. [89]

As already mentioned in chapter 1.3, one of the ways to reduce immunogenicity of the IVT mRNA for therapeutic purposes is to incorporate modified nucleotides. However, this has been shown to alter the secondary structure of the mRNA molecule thereby resulting either in complete loss of function[79], or reduced binding to pattern recognition receptors[77]. Having that in mind, the rational design of beneficial 5'-UTRs should result in sequences that are not susceptible to secondary structure changes due to use of different modifications, and functional in a cell-type independent manner. Minimalistic 5'-UTR based mRNA design was combined with the use of chemically modified nucleotides with the aim of improving the translational efficiency of the resulting mRNAs. The focus was also laid on testing different translation regulators since most previous studies have been performed using the traditional Kozak element. Elfakees et al.[90] reported that approximately 5% of the protein-coding genes with short UTRs harbor a unique translation initiation element (translation initiator of short UTRs: TISU). TISU-mediated translation, though cap-dependent, is initiated through a non-canonical scanning mechanism.[90] The translation is facilitated by eIF1, indicating that it interacts with eIF4G1. Upon release of eIF4F, AUG is recognized by the 48S ribosomal subunit[40]. The detachment of the cap complexes enables translation from such short 5'-UTR, at the same time by preventing a clash between the eIF4F cap complex and 48S (Figure 7). It is assumed that TISU supports translation even under stressed conditions what is supported by their high prevalence among genes associated with high energy metabolism and mitochondrial activities.[89]

Some of the naturally occurring 5'-UTRs in cellular genes and UTR elements from viruses were used as benchmarks.[91–93] Besides being cell-type and modification independent, the UTR should ideally also work with a broad range of sequences. For this, minimal UTR design was tested with two different classes of proteins: intracellular protein (firefly luciferase) and secretory protein (human erythropoietin). To test for cell type dependency, experiments were performed in two cell types of different tissue origin: human lung alveolar epithelial cell line (A549) and human hepatocellular carcinoma cells (HepG2). Finally, the minimal UTR concept was also tested in mice after intravenous injection of mRNA-lipid nanoparticles.

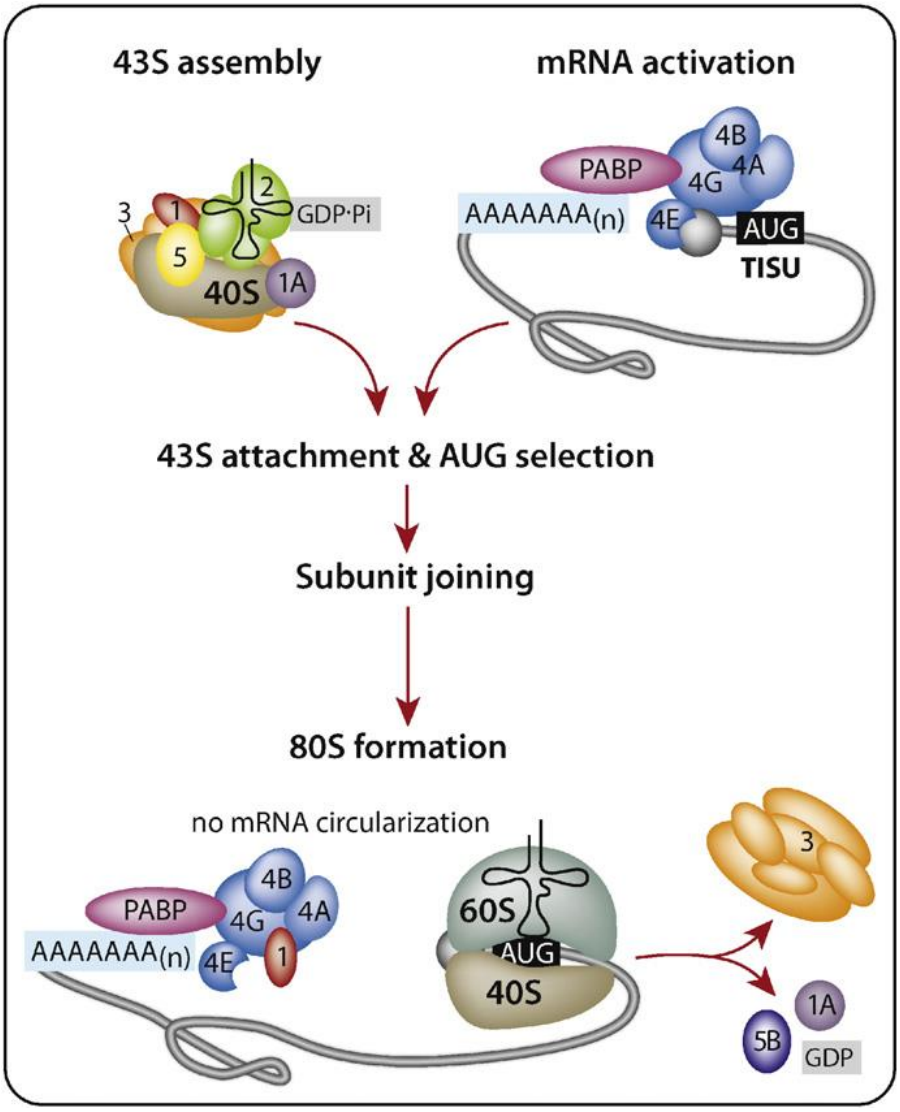

Figure 7. A schematic view on cap-dependent and scanning independent translation initiation of constructs harboring the TISU element.

When mRNA contains a very short 5'-UTR, AUG is located in a proximity of cap. Cap-binding complex recruits pre-initiation complex and recognize AUG, resulting in eIF1 displacement and eIF4F detachment from the cap. Then, AUG is placed at the P-site. In this model, RNA circulation is prevented by the release of eIF4F from the cap. Figure adapted from Haimov et al, 2015. [89]

4.2. Methods

4.2.1. Plasmid preparation

The synthetic 5'-UTR sequences 1-7, as well as reference UTRs, were cloned by a PCR-based strategy. The coding sequence for firefly luciferase was amplified from a pGL4.10 plasmid (Promega). For each UTR, a specific set of primers was designed. PCR reaction was performed using *Pfu* DNA polymerase (Promega) following the manufacturer's instructions. Correct PCR products (desired UTR with the gene of choice) were cloned into pUC57-Kana vector (GenScript).

4.2.2. Generation of mRNA

To generate *in vitro* transcribed mRNA, plasmids were linearized by *BstBI* digestion and purified by chloroform extraction and ethanol precipitation[94]. Purified linear plasmids were used as a template for *in vitro* transcription. Plasmid templates (0.5 µg/µl) were subjected to *in vitro* transcription using 3U/µl T7 RNA polymerase (Thermo Fisher Scientific), with a defined choice of natural and chemically modified ribonucleotides (Jena Biosciences). The modification set 1, involving 25% of 5-methylcytidine and 25% of 2-thiouridine was synthesized as described previously[77,95]. As for modification set 2, instead of 5-methylcytidine (25%) and 2-thiouridine (25%), 5-iodouridine (35%) and 5-iodocytidine (7.5%) were used. The complete IVT-mix was incubated at 37°C for 2h. Afterward, 0.01 U/µl DNase I (Thermo Fisher) was added for an additional 45 min at 37°C to remove the plasmid DNA template. RNA was precipitated with ammonium acetate at a final concentration of 2.5 mM, followed by two washing steps with 70% ethanol. The pellet was re-suspended in *aqua ad injectabilia*. A C1-m7G cap structure was added enzymatically to the 5' end of the transcript using Vaccinia Virus Capping Enzyme (NEB) following manufacturer's instructions. The 3' end of the transcript was subjected to enzymatic polyadenylation of ~ 120 nucleotides, using E. *coli* poly(A) Polymerase (NEB). RNA quality and concentration were measured spectrophotometrically on a

NanoDrop2000C (Thermo Fisher Scientific). Its correct size and purity were determined *via* automated capillary electrophoresis (Fragment Analyzer, Advanced Analytical).

4.2.3. Cell culture

A549 (ACC-107) and HepG2 (ACC-180) cells were purchased from DSMZ. They were cultured in Minimum Essential Media (MEM) with Glutamax (Gibco/Life Technologies) and RPMI 1640 plus GlutaMAX (Gibco/Life Technologies) respectively. Both media were supplemented with 10% heat-inactivated fetal bovine serum (FBS; Gibco/Life Technologies) and 1% penicillin/streptomycin (Gibco/Life Technologies). Cells were cultured in a humidified 5% CO_2 incubator at 37°C.

4.2.4. *In vitro* transfection

Both cell lines, A549 and HepG2 were transfected with two different doses of mRNA (125 ng/well and 250 ng/well). A549 and HepG2 cells were seeded at the density of $2x10^4$ cells/well and $4x10^4$ cells/well respectively in a 96 well plate. 24 hours post-seeding, cells were transfected using the commercial transfection reagent Lipofectamine®2000 (Thermo Fischer Scientific). Complexes were prepared at a ratio of 2 µl Lipofectamine®2000 per 1 µg mRNA. The mRNA was diluted 1:20 in water, and Lipofectamine®2000 1:10 separately in a serum-free MEM. mRNA was added to the Lipofectamine®2000 solution followed by 20 min incubation time at RT. The concentration of the final mRNA/Lipofectamine®2000 solution was 25 ng/µl, and a serial dilution 1:2 was performed. 10 µl of the complex solution was added to the cells and cells were incubated for 24 and 48 h, respectively. For every mRNA construct, replicates of three or six were prepared.

4.2.5. Firefly Luciferase Assay

For detection of firefly luciferase activity, cells were lysed for 30 min at RT in lysis buffer (25 mM Tris-HCl, 0.1% TritonX-100, pH 7.4). Luciferase assay was performed as described previously[96–98]. Photon luminescence emission was measured for 5 s using Tecan InfiniteR 200 PRO.

4.2.6. Rabbit reticulocyte lysate, the cell-free translation system

Nuclease treated rabbit reticulocyte lysate (Promega) were used for investigating translation from luciferase-encoded RNA transcripts. The experiment was set up following the manufacturer's procedure. The reaction was analyzed after luciferase buffer application, by measuring photon luminescence emission for 5 s using Tecan InfiniteR 200 PRO.

4.2.7. Lipid formulation

For *in vivo* experiments, mRNA was formulated in lipid nanoparticles (LNPs) as previously described.[99]

4.2.8. *In vivo* studies

For *in vivo* studies, Balb/c mice (Charles River Laboratories) at the age of six to eight weeks, were used. All animal experiments were carried out according to the guidelines of the German law of protection of animal life and reviewed by the local ethics committee. Mice were injected intravenously with 20 µg LNP formulated mRNA. *In vivo* imaging of luciferase expression was performed at 6 h post-delivery using an IVIS Lumina XR Imaging System (Caliper Life Sciences) and organs were harvested and analyzed for *ex vivo* luciferase measurements as described previously[99].

4.2.9. Enzyme-linked immunosorbent assay

Quantification of hEPO protein was performed using human Erythropoietin Quanktine IVD ELISA kit (R&D Systems) following manufacturer's instructions.

4.2.10. Statistical analysis

Each experiment was performed with at least three technical replicates per sample. Results are shown as means ±SD unless otherwise stated. Statistical analysis was performed using GraphPad Prism software (version 6). Data were tested for normal distribution using D'Agostino-Pearson omnibus normality test. Multiple comparisons were conducted by two-way ANOVA, followed by Sidak's test (pairwise comparison) or Dunnett's test (many-to-one comparison). A p-value ≤ 0.05 was considered statistically significant.

4.3. Results

Upstream of the gene of interest, recombinant DNA sequences making short synthetic 5'-UTRs were cloned into the plasmid. Their sequence and position of each base relative to the start codon is listed in Table 12.

Table 12. Sequences of synthetic 5'-UTR with annotation of each nucleotide position.

Each sequence comprised of the T7 promoter (T7), the desired 5'-UTR and a luciferase coding sequence. Listed are the respective sequences, each starting with a GGG as a transcription-starting site and ending with an ATG, a start codon. The upper row of numbers indicates the position of those nucleotides with respect to the transcription start site, as GGG is the first triplet transcribed into mRNA. The lower row of numbers indicates coordinates of nucleotides with respect to the start codon (ATG).

minimal 5'-UTR	**Promoter**[34,35]	**Transcription start site**[34,35] →						**Extra nt**		**Kozak consensus**[87]						**Start codon** ←		
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**									
										-6	**-5**	**-4**	**-3**	**-2**	**-1**	**1**	**2**	**3**
UTR1	T7	G	G	G	A	G	A			G	C	C	A	C	C	A	T	G
UTR2	T7	G	G	G	A	G	A	C		G	C	C	A	C	C	A	T	G
UTR3	T7	G	G	G	A	G	A	C	T	G	C	C	A	C	C	A	T	G
UTR4	T7	G	G	G	A	G	A	C	A	G	C	C	A	C	C	A	T	G
UTR5	T7	G	G	G	A	G	A	C	C	G	C	C	A	C	C	A	T	G
UTR6	T7	G	G	G	A	G	A	C	G	G	C	C	A	C	C	A	T	G
UTR7	T7	G	G	G	A	G	A	C	T	G	C	C	A	A	G	A	T	G

All sequences combined T7 promoter (TAATACGACTCACTATA) with Kozak consensus element (GCCACC), forming very short 5'-UTR upstream of the start codon ATG (minimal UTR; *Table 12*). The last 6 nucleotides of the T7 promoter, directly downstream of the TATA box, are needed to ensure the homogeneity of the 5' end, and to enhance the yields of *in vitro* transcribed mRNA[100–104]. Comprising only these necessary elements for transcription and translation, a short synthetic 5'-UTR was formed (UTR1), with the sequence: GGGAGAGCCACC. A UTR2 sequence harbored an additional C nucleotide at base position +6 in mRNA, inserted between the T7 promoter and Kozak consensus. With this insertion, the G at position +5 aligned with -9 position (with respect to start codon), thereby resulting in complete Kozak consensus (GCCGCCACC)[105].

4.3.1. Effect of spacer nucleotides between the T7 promoter and Kozak element

The first set of experiments were designed to compare translation efficiencies of UTR1 and UTR2 containing luciferase-coding transcripts. The efficiency of *in vitro* transcribed transcripts, using either unmodified or modified sets of nucleotides (modification 1, modification 2) was tested in two different cell lines, A549 and HepG2. As seen from Figure 8, a very minimal synthetic 5’-UTR, namely UTR1, expressed less luciferase than its counterpart harboring an extra nucleotide of C. A significant increase of protein levels for UTR2 was observed in both cell types and time points (24h and 48h). Since UTR2 performed better, its minimalistic design was used in all the following constructs.

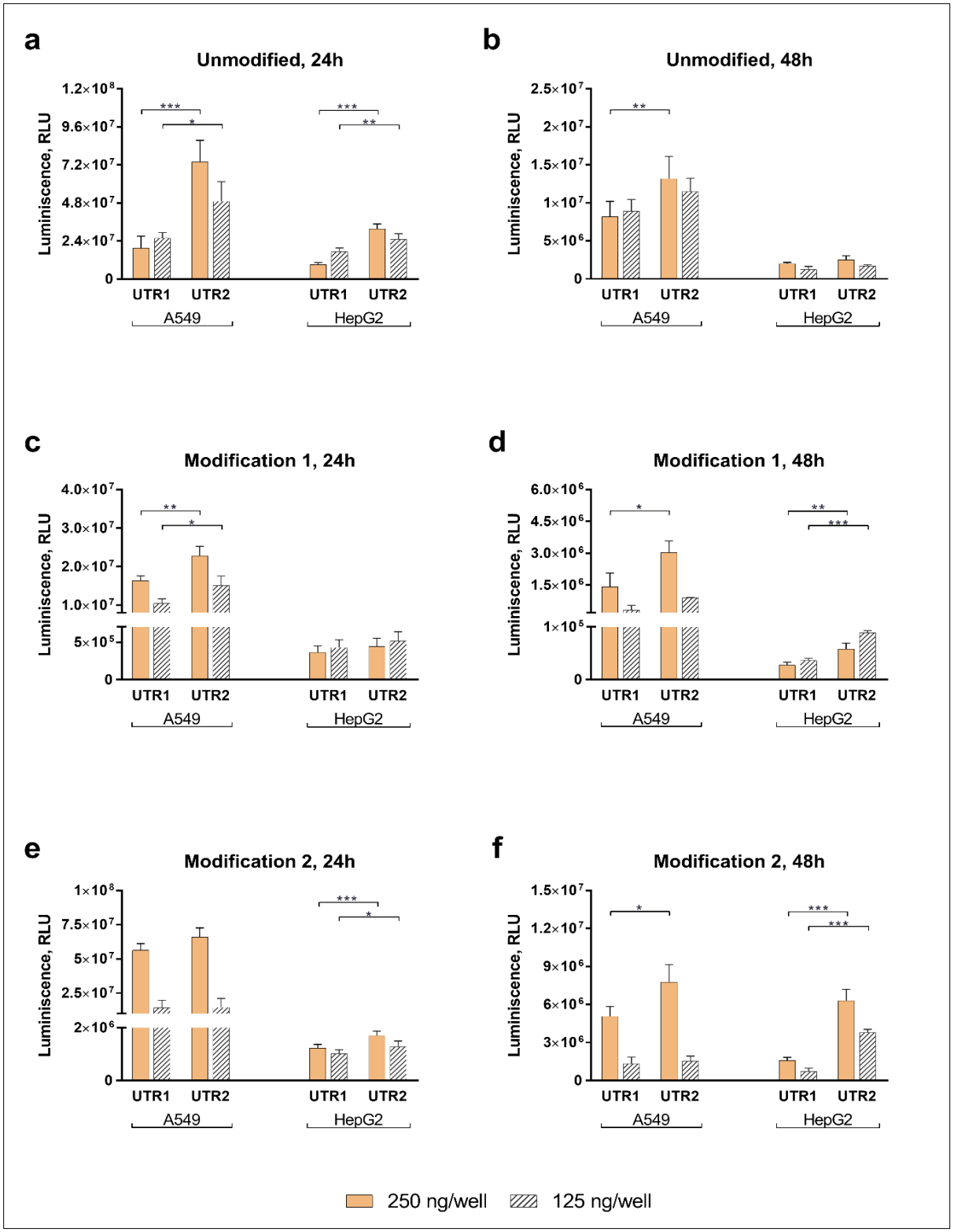

Figure 8. Luciferase activity in protein lysates from A549 and HepG2 cells transfected with minimal UTR1 and minimal UTR2 (Table 1) containing Luc mRNAs.

*Luciferase activity was measured 24h (a, c, e) and 48h (b, d, f) post-transfection. Statistical significance was assessed by 2-way ANOVA test with p-values: *p≤0.05, **p≤0.01, ***p≤0.001, n=6.*

Using UTR2 as a template, the assortment of short synthetic 5'-UTR was expanded by increasing the distance between the T7 promoter and Kozak consensus. The distance was increased by only 1 extra nucleotide, and the choice involved all 4 possibilities: T, A, C or G at the -7 position. This way, four new constructs were created, namely UTR3-6 (*Table 12*). In the subsequent experiment, 5'-UTR from human alpha-globin was inserted into luciferase-encoding mRNA. Since this combination of 5'-UTR was shown to increase protein translation in cellular systems[39], it was used as a benchmark for the short synthetic UTRs. As for previous experiment (Figure 8), all IVT mRNA constructs were produced in both, unmodified and modified forms and they were further used for A549 cell transfection. With the construct harboring an extra T between the T7 promoter and Kozak consensus (UTR3), the highest protein expression was observed, compared to other three constructs with the different extra single nucleotide at position -7 (UTR4-6). This translation enhancing effect was either comparable to or better than the effect seen with human alpha-globin 5'-UTR containing mRNA (refUTR1). Similar trends were observed at both time points (Figure 9).

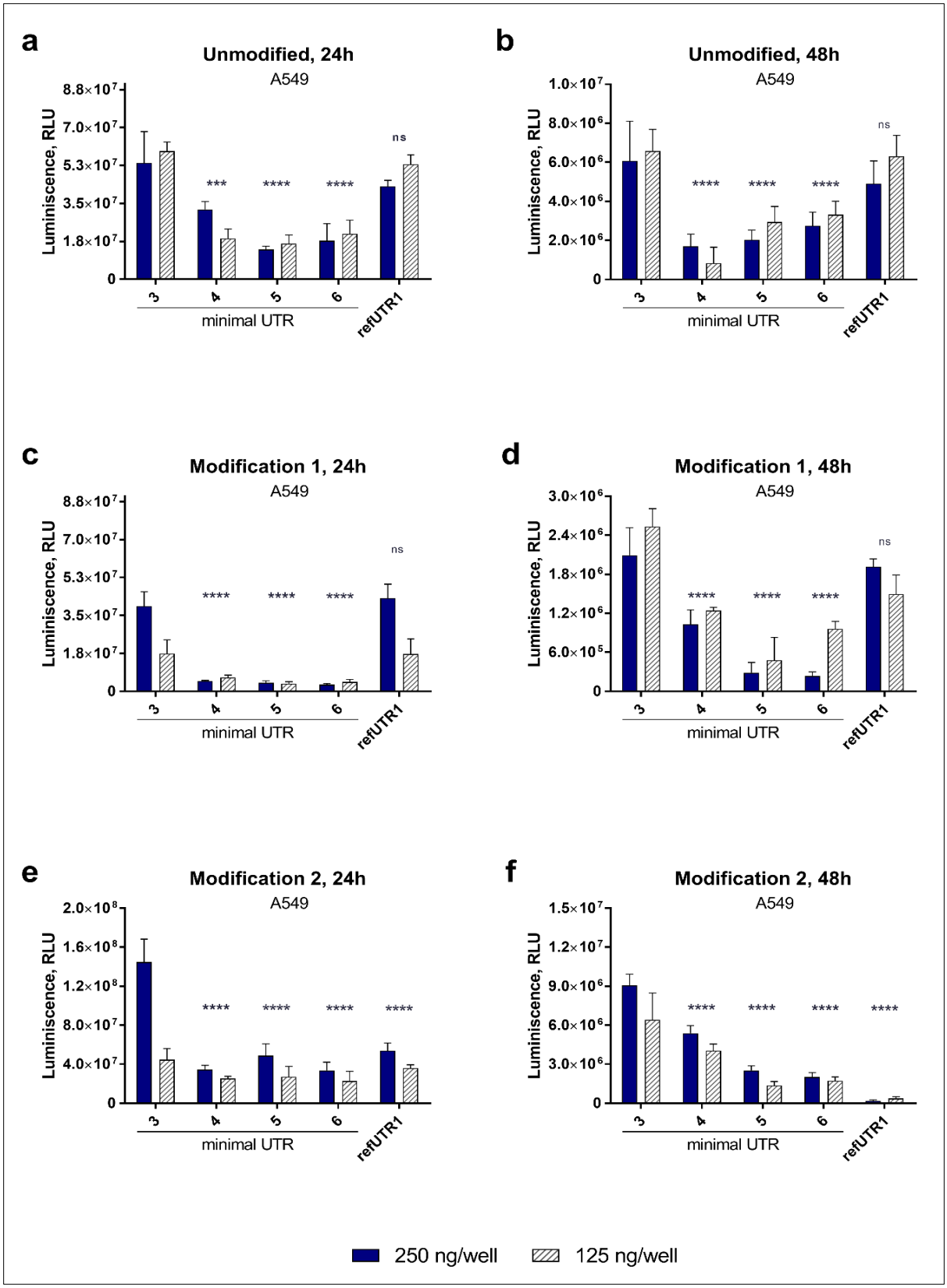

Figure 9. Luciferase activity in protein lysates from A549 cells transfected with different UTRs containing Luc mRNA.

*Luciferase activity was measured 24h (a, c, e) and 48h (b, d, f) post-transfection. Statistical significance was assessed by 2-way ANOVA test, comparing each mean of the column with the minimal UTR3. P-values indicate: *p≤0.05, **p≤0.01, ***p≤0.001, n=6.*

4.3.2. Effect of translation regulator

From published literature, it can be expected that the choice of translation regulator will influence mRNA translation. Due to their specific, conserved nucleotide sequence, subsequent experiments were designed to investigate their effect on translation as influenced by the choice of nucleotide modifications and cell types. TISU element was previously found to be a specific translation regulator for sequences with very short 5'-UTRs[106]. As UTR3 resulted in highest expression (comparable or better than benchmark refUTR1), minimal UTR3 was chosen as a template for creating minimal UTR7 *(Table 12)*. In UTR7, Kozak element is substituted with TISU element, and as a gene of interest, luciferase was further used. Both cell lines, A549 and HepG2 were transfected with those mRNAs, including unmodified and modification sets 1-2. The functionality of the regulator of translation was shown to be affected upon incorporation of modified nucleotides during *in vitro* transcription in both cell lines. In the case of transfection with unmodified IVT mRNA, luciferase expression levels were comparable. A significantly higher luciferase expression was observed in construct furnished with the TISU element (UTR7) while using IVT modification set 1 mRNA when compared to the construct with Kozak element (Figure 10). However, when observing expression from modification set 2, an opposite trend was observed. These observations were consistent at both time points.

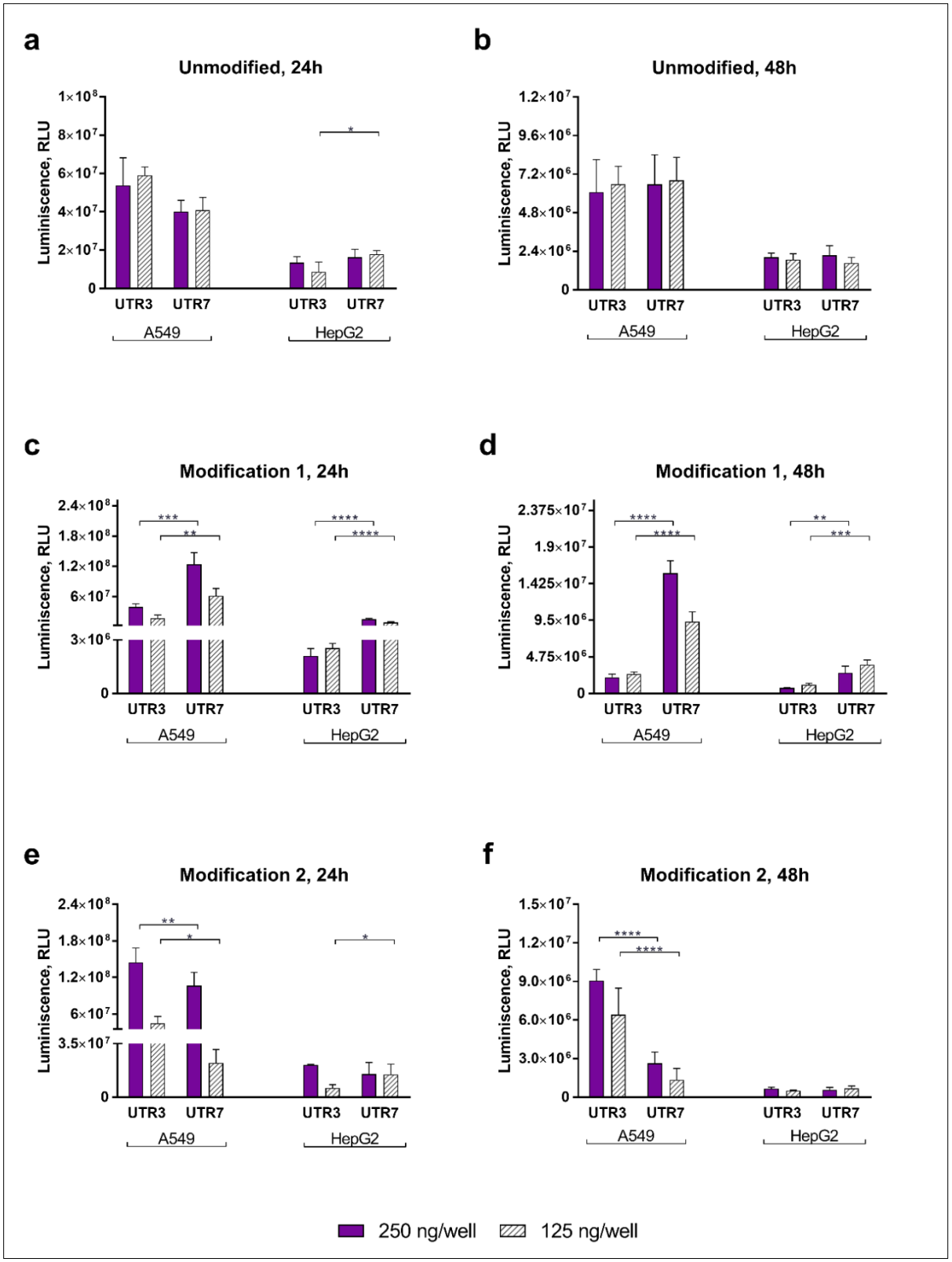

Figure 10. Luciferase activity in protein lysates from A549 and HepG2 cells transfected with either Kozak (UTR3) or TISU (UTR7) element within Luc mRNA.

*Luciferase activity was measured 24h (a, c, e) and 48h (b, d, f) post-transfection. Statistical significance was assessed by 2-way ANOVA test. P-values indicate ns>0.05, *p≤0.05, **p≤0.01, ***p≤0.001, n=6.*

4.3.3. *In vivo* comparison of different Luciferase coding mRNA constructs

As minimal UTR3 showed potential supported by results with comparable expression levels to those from commonly used human alpha-globin 5'-UTR and UTR7, it was selected as a lead candidate to be tested *in vivo*. Apart from the human alpha-globin UTR, luciferase construct proposed by Guild et al[91], composed of 5'-UTR from the human CMV enhancer and 3'-UTR from the human growth hormone (refUTR2), was selected as a second benchmark. Many other short UTRs have been published previously[39,91,92], which were tested as additional benchmarks for the short synthetic UTR constructs from the current study. For this, minimal UTR7 was compared to, the additional reference UTRs (refUTR-5). Their sequences are listed in Table 13. These UTR sequences were previously reported to enhance protein expression[91]. Listed mRNAs were transcribed incorporating chemically modified nucleotides as defined by set 1 and afterward formulated in a previously described lipid formulation[99]. Experiments were performed in 6-8 weeks old Balb/c mice by intravenously injecting 20 µg LNPs per mouse. Bioluminescence imaging was performed 6h post-injection and luciferase expression was as photons/sec/cm^2/sr. As IV delivery of LNPs results in maximal expression in the liver, luciferase expression in isolated organs (liver and spleen) was also measured *ex vivo*. Similar to the results obtained *in vitro*, minimal UTR3 was as effective as the benchmark (refUTR1). Comparing minimal UTR7 harboring TISU translation regulator to Kozak containing counterpart (minimal UTR3), the levels of expression were significantly higher. The second benchmark, refUTR2, was significantly lower than the minimal UTR7 in liver and spleen. Similar trends were observed irrespective of the imaging (whole animal *in vivo* vs. organs ex vivo) (Figure 11).

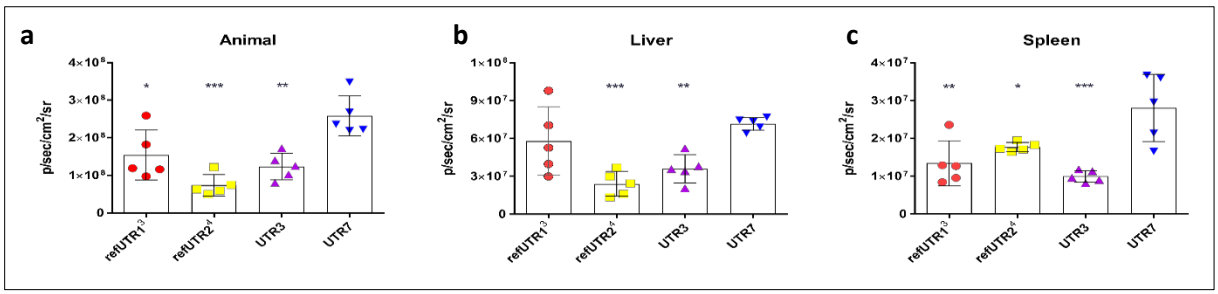

Figure 11. Luciferase expression in mice injected with 20µg of different luciferase coding RNAs produced using modification1.

Luciferase expression was measured at 6 h post i.v. injection using In Vivo Imager (animal - a). Liver (b) and spleen (c) were harvested post measurement for ex vivo luciferase measurements.

4.3.4. Comparison of short synthetic UTRs to naturally occurring viral and human UTRs

Many other short UTRs have been published previously[39,91,92], which were tested as additional benchmarks for the short synthetic UTR constructs from the current study. For this, minimal UTR7 was compared to, the additional reference UTRs (refUTR3-5). Their sequences are listed in Table 13.

Table 13. Sequences of 5' and 3'-UTRs references used in the present study.

Listed are five 5’-UTR sequences that were used as benchmarks, with their gene source and literature reference.

Reference UTR	5’-UTR sequence	5’-UTR source	3'-UTR sequence	5’-UTR source	Reference
refUTR1	GGGAGACTCTTCTGGTCCC CACAGACTCAGAGAGAAC	hαg	/	/	[39]
refUTR2	GGAGACGCCATCCACGCT GTTTTGACCTCCATAGAAG ACACCGGGACCGATCCAG CCTCCGCGGCCGGGAACG GTGCATTGGAACGCGGAT TCCCCGTGCCAAGAGTGA CTCACCGTCCTTGACACG	human cytomegalovirus (hCMV)	CGGGTGGCATCCCTGT GACCCCTCCCCAGTGC CTCTCCTGGCCCTGGA AGTTGCCACTCCAGTG CCCACCAGCCTTGTCC TAATAAAATTAAGTTG CATC	human growth hormone (hGH)	[91]
refUTR3	GGGAGACAAAAAAAAAAA AAAAAAAAA	orthopoxvirus	/	/	[92]
refUTR4	GGGAGACCATATTGAAGA GACAGAGTGATATATAAA ACTGCTAA	deep sequencing of human genome	/	/	[93]
refUTR5	GGGAGACAAAACTGCTAA	deep sequencing of human genome	/	/	[93]

In vitro transfection was done in A549 cells, with above-mentioned IVT mRNAs. Again, both unmodified and chemically modified transcripts were tested. Synthetic UTR7 achieved comparable expression to refUTR3, regardless of the modification used (Figure 12). Higher expression with minimal UTR7 was achieved when compared to refUTR4 (unmodified, modification 2) and reference UTR5 (modification 2) after 24h, but the effect diminished after 48h.

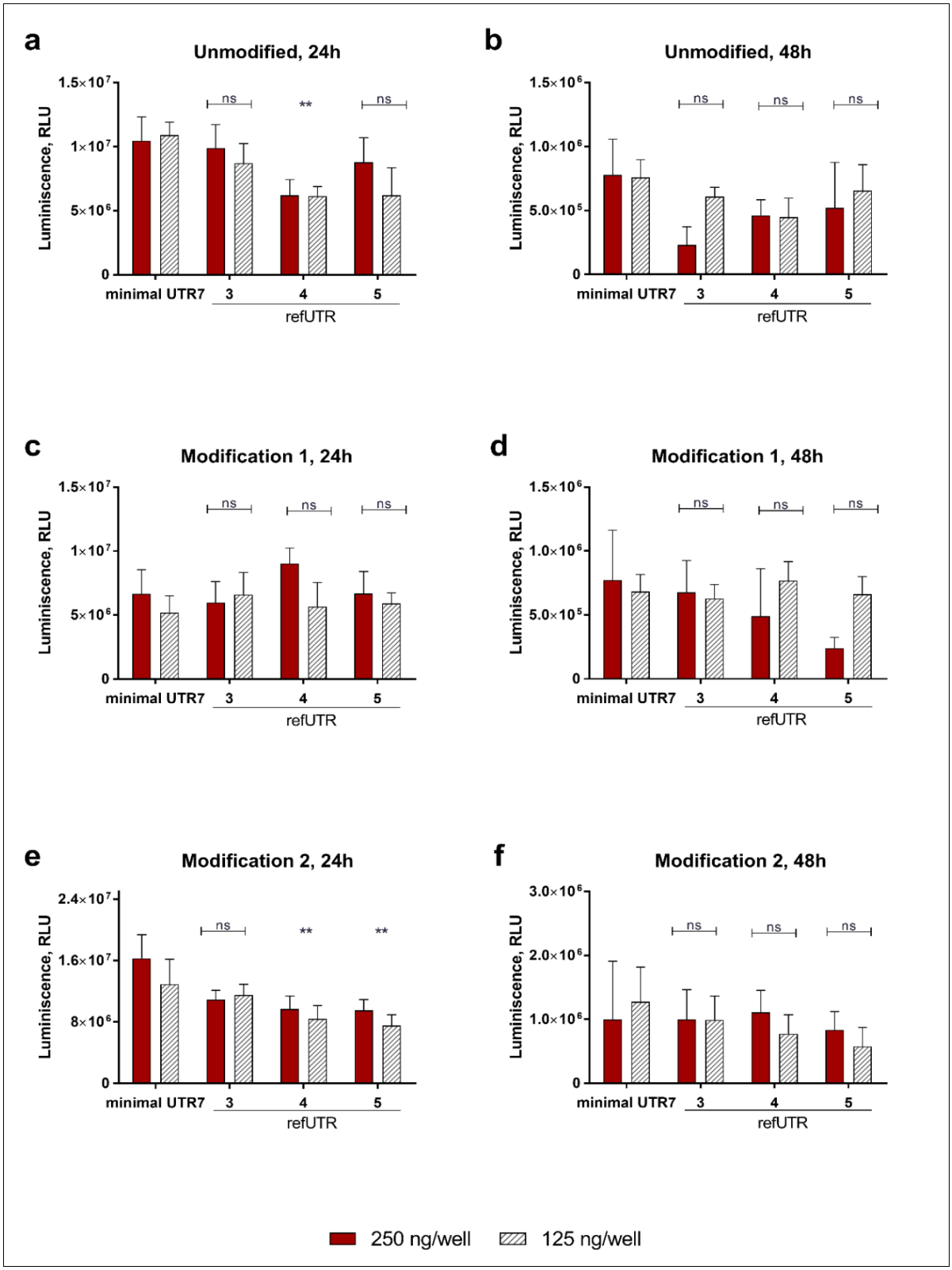

Figure 12. Luciferase activity in protein lysates from A549 cells transfected with different short transcription regulator elements.

*Luciferase activity was measured 24h (a, c, e) and 48h (b, d, f) post-transfection. Statistical significance was assessed by 2-way ANOVA test. P-values indicate: ns>0.05, **p≤0.01, n=4.*

To rule out the possibility that different UTR sequences were affecting critical mRNA attributes (e.g. capping, a proportion of full-length transcripts etc.) or differential transfection efficiencies contributing to the above-observed differences, translation experiment was carried out in nuclease-treated rabbit reticulocyte lysates. Equal amounts of IVT mRNA of minimal UTR7 and refUTR3-7 were applied to the cell lysates. For each mRNA type (unmodified, modification 1 and modification 2), a comparable level of luciferase expression was observed among the different UTRs (Figure 13). These results confirm that the observed differences in translation efficiency can be attributed to the cellular performance of particular UTR elements.

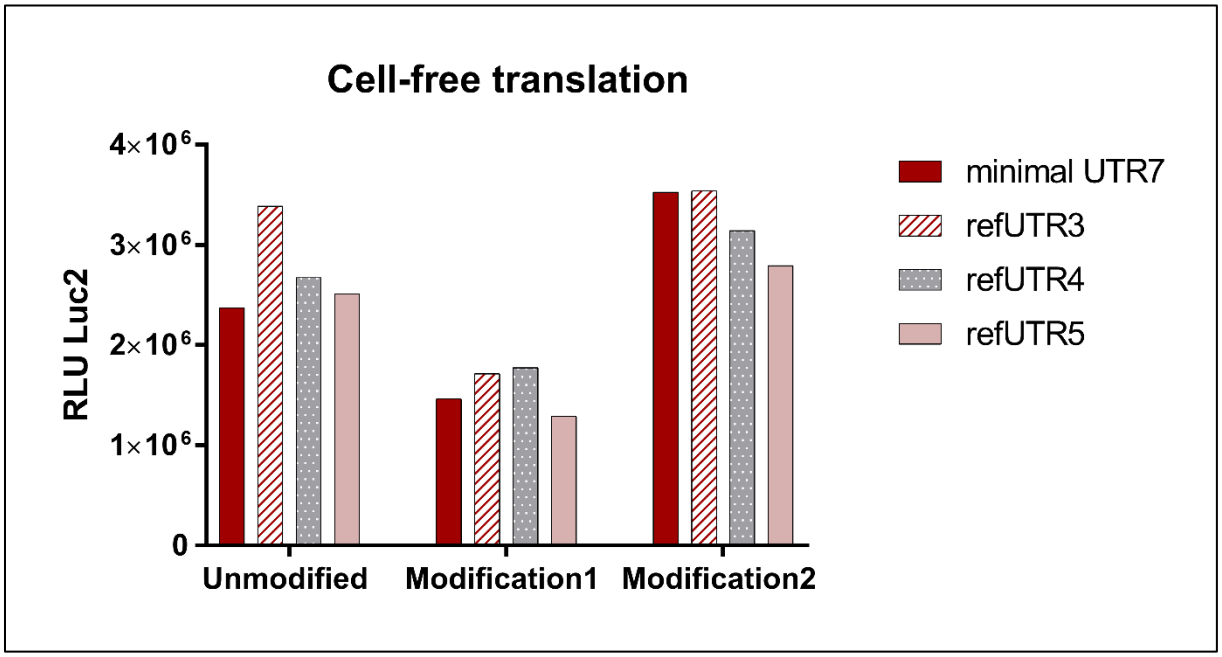

Figure 13. Luciferase activity in a cell-free translation system of rabbit reticulocyte lysates with different short transcription regulator elements.

Luciferase activity was measured 60 min post-transfection.

4.3.5. A sequence-independent function of minimal UTR7

To test if results observed with luciferase, an intracellular protein, were transferable to another class of protein e.g. secreted protein, additional constructs were made with UTR7 and human EPO. As a reference, refUTR2 was combined with the EPO coding sequence. mRNA was *in vitro* transcribed using modified nucleotides as defined by set 1. A549 and HepG2 cells were transfected, and EPO expression was quantified *via* ELISA at 24h post-transfection. Expression of EPO in A549 cells revealed that UTR7-containing mRNA produced more protein than refUTR2 (Figure 14). Similar results were obtained in HepG2 cells as well, although the difference was not statistically significant.

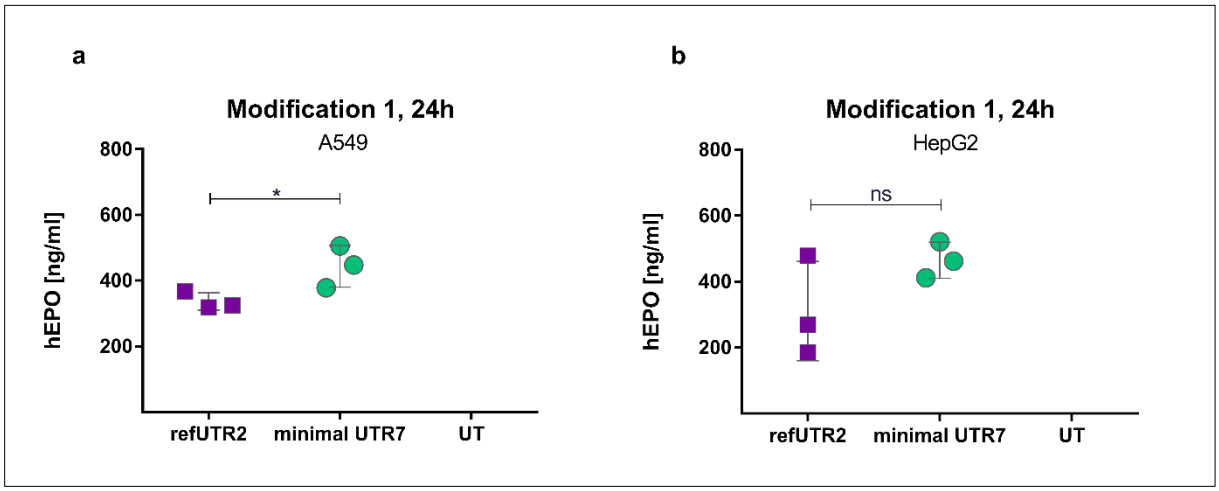

Figure 14. Quantification of secreted human erythropoietin protein levels as measured via ELISA in A549 (a) and HepG2 (b) cell supernatants at 24 hours post-transfection.

*Values represent mean ± SD of 3 replicates. Statistical significance was assessed by 1-way ANOVA test with p-values: ns>0.05, *p≤0.05. UT = untransfected.*

4.4. Discussion

The goal of this study was to design minimalistic 5'-UTRs and compare their functionality to naturally occurring 5'-UTRs with respect to their effect of protein translation after mRNA delivery to the cells. Synthetic minimal UTRs were comprised only of elements necessary for *in vitro* transcription (T7 promoter) and translation in the cells (AUG start codon), with several different versions of connecting nucleotides between the elements.

A T7 promoter is recognized by a phage T7 polymerase, which catalyzes the formation of RNA in the 5' → 3' direction. A T7 polymerase is a convenient tool based on its low error rate and extremely promoter-specific expression[107]. In the earlier studies, it has been proposed that T7 promoter contains two functional domains: the binding region, from -17 to -5, and the catalytic domain downstream of -5 position[103,104]. In the study by Imburgio et al, it was reported that the position of nucleotides in the binding region of mRNA affects the promoter recognition and, therefore, initiation of the reaction.[102] A selection of promoter variants was designed, and it was found that any base substitution at position +1 to +6 from the wild-type sequence (GGGAGA) negatively affected the strength of promotor. Based on this finding, short synthetic minimal 5'-UTRs in the current study were designed with respect to conserved T7 promoter elements.

The major factor determining the ATG recognized as the start codon is the Kozak consensus sequence, in vertebrates determined as (GCC)GCCRCCATGG, where R is a purine (A or G).[108–111] The data were compiled by Kozak et al, from 699 vertebrate mRNAs, where 33% contained a G at the -9 position. At the position -7, 37% contained a C.[87] The experimental manipulation of the nucleotides within a certain position in the sequence revealed that some nucleotide positions are more important than the others are. This was especially true for -3 (R) and +4 (G) positions, where their mismatch caused poor gene expression.[112] In line with these details, minimal UTR2 actually followed the perfect consensus Kozak rule (GCCGCCACC), by fusing T7 promoter and Kozak element, with an extra C in between. By inserting an additional C nucleotide at position -7 (seven nucleotides upstream from the ATG codon), a G residue previously placed at the -8 position (in minimal UTR1) is

shifted to the -9 position (in minimal UTR2), with respect to the start codon. Therefore, bringing in an extra C into the very minimal UTR1 sequence, significantly increased protein expression in cells from delivered mRNA.

Further experiments, where each possible nucleotide was fused into the sequence at the base position +8 in the mRNA among the elements, showed that particularity an extra T (minimal UTR3) enhanced protein expression. This observation was independent of the use of modified or unmodified nucleotides in the mRNA.

The potential of minimal UTR3 was later compared to 5'-UTR from human alpha-globin gene, which has been shown to affect mRNA stability and its translation. As an example, in a study by Schrom et al[113], incorporation of the 5'-UTR from human alpha-globin prior to the gene encoding human angiotensin-converting enzyme part 2 (hACE2), increased enzyme translation and mRNA stability, when compared to naturally occurring hACE2 5'-UTR, or even the CYBA UTR.[84,113] Moreover, 5'-UTR from human alpha-globin has been reported to have higher translation efficiency compared to beta-globin 5'-UTR[39]. The reference sequence for human alpha-globin from the NCBI database (NM_000517.4) was used as a 5'-UTR, but only from positions 31 to 60, although its full length is 66 nucleotides.[114] The reason for selecting this section is that nucleotides prior to those are not transcribed into the mRNA. In our constructs, nucleotides from 55 to 60 in the wild-type 5'-UTR from human alpha-globin (GGGAGA) were replaced by standard Kozak sequence (GCCACC).

The performance of different UTRs was also affected by the choice of nucleotide modifications used, what was especially seen in the case of the benchmark human alpha-globin 5'-UTR (refuUTR1) when used with modification 2. The functionality of refUTR1 in luciferase expression was reduced, what was especially seen at 48h post-transfection. However, while using it either with the unmodified or with modification 1 mRNA, no significant difference in expression was observed. These findings highlight the sensitivity of natural UTR sequences to the choice of modification used in mRNA production. Some of the chemically modified nucleotides might alter the secondary structure of UTR region in such a manner that it may lose its functionality. This, in turn, might result in loss of function

and, therefore, reduced translation.[79,84] An example of such an impairment of the 5' element function was shown by Thess et al, 2015. By the incorporation of pseudouridine, IRES element completely lost its function as a recruiter of translation initiation in a cap-independent manner.[79]

As some of the previous studies imply, the minimal length of 5'-UTR to ensure proper binding of the ribosome and therefore translation initiation at the start codon is 32 nucleotides. Having fewer nucleotides than that initiates the translation at some of the downstream ATG instead of the first start codon[111,115]. Considering that the short synthetic UTRs in the current study consisted of a maximum of 15 nucleotides, TISU element was tested as a translational regulator in a direct comparison with Kozak element to support the translation from these short 5'-UTRs. Moreover, a majority of previous studies involving usage of chemically modified nucleotides in mRNA transcripts did not investigate its effect on any other translation regulator but Kozak element.[72,73,77,79,84,95,113,116] Defined as a translation initiator of short UTRs, TISU was first discovered in the 5% of short 5'-UTRs of non-coding human gene region.[106] With a unique sequence of AAGAUGG (minimal UTR7), it is distinguished from the Kozak sequence RCCAUGG (minimal UTR3). In the direct side-by-side comparison of these two translation regulators, significant translation differences were observed when using modification 1 containing mRNA. Minimal UTR7 mRNA translated more efficiently at 24h and 48h post-transfection, what can be directly contributed to the TISU element since this is the only difference between the two sequences. Since both modification sets (modification 1 and 2) contain modified C (5-methylcytidine in modification 1 and 5-iodocytidine in modification 2), it is possible that either all or some of the Cs preceding the ATG codon in the Kozak context (ACCATG) get modified. In the case of TISU-containing sequences, this scenario is theoretically not possible since there is no C located in the three bases upstream from ATG (AAGATG). This could be the explanation of how specific chemical modification(s) could alter and affect the performance of regulatory elements, whether it be UTRs or translation regulators.

Due to the encouraging results obtained *in vitro*, the potential of minimal UTR3 and UTR7 luciferase-encoding constructs was also confirmed *in vivo* in mice. The refUTR2 from a previous study[91] was

used as an additional benchmark in these *in vivo* experiments. Similar to the *in vitro* results, minimal UTR7 resulted in highest translation efficiency compared to minimal UTR3 and refUTR2. As the same trend was observed in liver and spleen, the activity of minimal UTRs is likely to be cell-type independent.

Data with human EPO-mRNA in A549 and HepG2 cells confirm that translation-enhancing effect of minimal UTR7 is the sequence- and cell-type- independent. Minimal UTR2 had a positive impact on translation compared to CYBA 5'-UTR[84] or natural 5'-UTR from hACE2[113] in previous studies, what confirms a conceptual advantage of minimalistic UTRs (minimal UTR7 is 113 nucleotides shorter than refUTR2).

Encouraged by the *in vitro* and *in vivo* results, minimal UTR7 was further benchmarked to several other naturally occurring 5'-UTRs. A short homopolymeric stretch of 20 adenosines at the 5'-end is a characteristic of orthopoxviruses, enabling them to repress mRNA decay (refUTR3)[92]. Based on the metagenomics study of the human genome that revealed a 37-nucleotide long conserved sequence, refUTR4 was created. The sequence was reported to promote expression in a cytoplasmic expression system that is based on Vaccinia virus (VACV)[93]. Based on the length of a functional promoter in the reported system[93], two new constructs were designed: refUTR4 containing 37 nucleotides of promotor with translation enhancer, and refUTR5 containing only the translation enhancer. In a direct comparison, minimal UTR7 was either comparable to or better than these new benchmarks (refUTR3-5).

Although the mechanism of the performance of short synthetic 5'-UTRs was not investigated, one of the underlying factors for their expression effect is unlikeliness of forming secondary structures, either themselves or with downstream sequences. That also makes them less susceptible to the changes upon chemical modification of nucleotides. Synthetic, minimal but effective UTRs as such will reduce the amount of mRNA to achieve therapeutic efficiency, at the same time reducing the costs of such therapies.

4.5. Conclusion

It could be shown that synthetic sequences, designed on a rational sequence-design approach, outperform some of the most widely used 5'-UTRs, both *in vitro* and *in vivo*. Based on these data, minimal UTR3 and minimal UTR7 present themselves as promising candidates for the future development and use in mRNA-based therapeutics. None of these two sequences had any negative effects on RNA yield during *in vitro* transcription. Translation of these UTR containing mRNAs was a sequence- and cell-type- independent and either comparable or better than some of the widely used UTRs. Due to their short length, UTR3 and UTR7 are less prone to the structural change upon incorporation of chemically modified nucleotides, or cell- and target sequence- specific effects.

5 SEGMENTATION OF THE POLY(A) TAIL

5.1. Background

A distinct feature of the cellular mRNAs is a long chain of adenosines, located at the 3'-end of the molecule. Regardless of the protein that it encodes for or the tissue where it is expressed, it is a common structural feature of all cellular mRNAs, with some exceptions, e.g. histones. In the nucleus of the cell, shortly after the mRNA is synthesized by transcription, it undergoes several processes altering both ends, known as RNA processing. The posttranscriptional modifications are essential for a multitude of processes starting from export from the nucleus to successful translation including mRNA stability. One of such processes, occurring at the 3'-end is polyadenylation, a two-step process of adding poly(A) tail to the nascent molecule downstream of the gene-encoded polyadenylation signal (AATAAA). However, the poly(A) tail is not encoded in the genes; it is an end-product of a polyadenylation reaction, which follows site-specific cleavage[117]. It varies in length among eukaryotes, ranging from 90 adenosines in yeast to ~250 in mammals.[118,119]

Poly(A) tail also plays a role in mRNA maturation, by making it more stable and preventing degradation by nucleases[28]; it is an essential part for the nuclear export of mature mRNA[27]; it initiates the translation by the formation of the closed-loop structure. A formation of the closed-loop state is initiated by association of poly(A) binding proteins (PABP) with the poly(A) tail, which then interacts with the eIF4G of the translation initiation complex, attached to the cap [118,120] (Figure 15). Sachs et al

have shown that such a formation promotes 40S ribosomal recruitment towards mRNA, resulting in enhancement of translation in *S.cerevisiae.*[29]

Although there were many debates on whether the P-bodies play a role in mRNA decay or storage, recent findings show that these small cytoplasmic granules are the place of continued storage of mRNAs. An important parameter for the storage is the size of the poly(A) tail; the data suggest that there is a tendency towards shorter poly(A) tails, but there is no confirmation yet.[121]

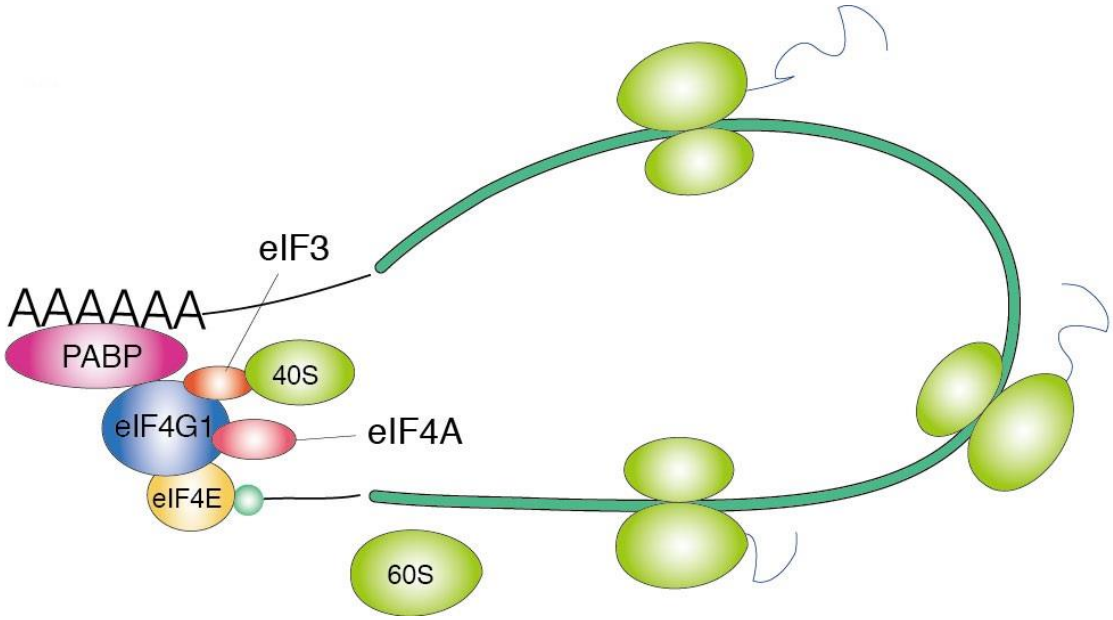

Figure 15. Closed-loop structure of the mRNA. It is initiated by the interactions between eIF4E, a cap-binding protein, and PABP, a poly(A)-binding protein through eIF4G1 as a linker.

Figure adapted from "An emerging interest in RNA world" [Internet]. MBL Life Sci. [122]

mRNA molecules intended for therapeutic purposes of protein synthesis are generated *by in vitro* transcription. Considering the procedure, the poly(A) tail can be either added enzymatically by poly(A) polymerase from *E.coli* in a post-transcriptional step or encoded already in the DNA template (plasmid or PCR template). Both methods suffer from certain limitation. Enzymatic addition of the poly(A) tail to the *in vitro* transcribed mRNA ensures sufficiently long poly(A) tails[123,124], but the

product is heterogeneous as the same length of the poly(A) for all mRNA molecules cannot be ensured; i.e. the length remains difficult to optimize. This means, that the end product is actually a heterogeneous mixture of different length of the poly(A) tails, which could in turn cause problems meeting the FDA regulatory requirements.[58] mRNA is susceptible to alkaline hydrolysis and as poly(A) polymerase requires alkaline conditions (pH > 7.5) to be fully functional, post-polyadenylated mRNAs are often of lower quality when compared to their controls without polyadenylation.[125]

Thus, template-encoded poly(A) is a preferred way of adding a poly(A) tail to the mRNA over enzymatic post-polyadenylation of mRNA. Reproducibility of the defined poly(A) length, which in turn results in a homogeneous product are some of the advantages of this approach.[126] Moreover, the procedure involves fewer steps of RNA production than enzymatic post-polyadenylation, reducing the costs of production.

As for the small-scale mRNA production, a PCR-based approach offers advantages such as high throughput. However, its high production costs and the risk of mutagenesis during PCR amplification that each polymerase carries, restrict its application for large-scale mRNA production for pre-clinical and clinical applications.

Considering large-scale production of the DNA template, a plasmid-based approach is a preferred way, due to its ease of scalability, and a well-established procedure that can be implemented under good manufacturing practice (GMP) conditions. The risk of mutations in the coding sequence is also lower compared to a PCR approach. However, encoding long homopolymeric stretches, such as poly(A), into the plasmid DNA causes unpredictable recombination during the bacterial amplification of the plasmid DNA.[127] In such cases, the poly(A) encoding sequence is shortened over time[127], what was reported in previous studies. When encoded poly(A) tail sequence was 70 nucleotides long, its length remained the same during bacterial amplification.[127] After increasing its length to at least 100 nucleotides, the spontaneous deletion mutants started appearing during the plasmid amplification.[128] Longer poly(A) tails, comprising 150 nucleotides were so unstable that no positive clones could be selected.[127]

Instability of poly(A) in circular plasmids led to the development of the linear plasmid-based system, namely pEVL. Such vectors offer the stability of the homopolymeric sequences during the cloning procedures and stretches up to 500 base pairs have been successfully clones[127]. Such linear plasmid-based system is commercially available as BigEasy® v2.0 Linear Cloning System (Lucigen), but it is a relatively large (>12 kb), very low copy vector and therefore not suitable for a large-scale production. Regardless of its large size, it offers only a limited choice of restriction sites; either *SmaI* or *NotI* which limit the cloning/linearization choices

An ideal solution, addressing the limitations mentioned by each of the above-described approaches for adding a stable, well-defined and homogeneous poly(A) sequence in the mRNA, would be to have a template-encoded poly(A) tail in the plasmid, which is not undergoing recombination during the plasmid amplification in bacteria. A poly(A) tail encoded in such a way would be able to deliver templates for *in vitro* transcribed mRNA production on a large-scale.

The current study aimed at investigation of segmented poly(A) tail and its effect on bacterial recombination during a high-copy plasmid amplification in *E.coli*. As a starting point, the most widely used poly(A) tail of 120 As in length ($poly(A)_{120}$) was used. To achieve the segmentation effect, but also taking note of the functional PABP footprint, the length of 120 As in total was split into either 2 or 3 segments of 60A's ($poly(A)_{2x60}$) or, 40A's ($poly(A)_{3x40}$) respectively. Earlier it was shown that PABP requires a minimum of 12 As to functionally bind to the poly(A) tail, but a single PABP molecule is not sufficient to successfully promote translation.[128] However, if oligomers of PABP are formed, they occupy approx. 27-30 nucleotides in a repetitive manner of the same poly(A) stretch.[129–131] In line with these publications, constructs were designed such that at least one oligomeric stretch of PABP can bind to each segment of the poly(A) tail. All variations of the poly(A) segmentation were compared to $poly(A)_{120}$, as it was shown to support the high expression of *in vitro* transcribed mRNA.[126,132] Initial read out for comparison was the recombination rate of the plasmid in *E.coli* among the constructs. From the selected positive clones, IVT mRNA was produced and transfected

into A549, HEK293 and 16HBE14o- cells: both translation efficiency and mRNA half-life measurements were made at different time points post-transfection.

It could be shown that segmented poly(A) significantly reduces recombination of plasmids in *E.coli* without negatively affecting mRNA half-life or translation efficiency.

5.2. Methods

5.2.1. Plasmid preparation

The synthetic poly(A) sequences were introduced to the vector backbone either as annealed complementary oligonucleotides or fragments created by PCR (Table 1). For sequences comprising of 2x60, 3x40 and ACH, specific sets of complementary oligonucleotides were synthesized and annealed. The synthetic poly(A) fragments of A120, 2x60_1, 2x60_12 and 2x60_24 were created by PCR.

Annealing of complementary oligonucleotides was performed as follows: 100 µM of each oligonucleotide were mixed with 40 µl annealing buffer (10 mM Tris HCl, 50 mM NaCl, 1 mM EDTA, pH 7.5) and incubated for 5 min at 95°C. Subsequently, the mixture was let to cool down to room temperature before proceeding with restriction digestion (*Bgl*II-*BstB*I).

For the high performance of PCR reaction, Phusion High-fidelity PCR master mix (Thermo Fisher Scientific) was used. To the mastermix, which contains 2x Phusion DNA Polymerase, nucleotides and optimized reaction buffer including $MgCl_2$, 0.5 µM of forward and reverse primer, 3% DMSO and 1 ng of template DNA were added to the reaction. The total reaction volume of 25 µl was initially denatured at 98°C for 30 sec, following by 30 cycles at 98°C for 10 sec, annealing at 72°C for 30 sec and extension at 72°C for 30 sec/kb. The final extension was performed at 72°C for 10 min. The size of the PCR product was confirmed on 1% agarose gel and the desired band was purified using NucleoSpin Gel and PCR clean-up kit (Macherey Nagel). Purified PCR product was digested with *Nhe*I-*BstB*I and stored at -20°C until further use.

Digested products of annealed oligonucleotides and PCR products were cloned into accordingly digested pUC57-Kana vector (GenScript) containing the desired coding sequences (firefly luciferase, d2EGFP, human EPO and human CFTR).

Table 14. Segmented poly(A) sequences and their corresponding cloning strategy using either PCR primer sets or oligonucleotides.

Construct	***Strategy***	***PCR primer forward / Oligo I***	***PCR primer reverse / Oligo II***
2x60_6	Oligonucleotides	GTGACTAGATCTAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAATGCATAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAATTCGAAGTGACT	AGTCACTTCGAATTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TATGCATTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAGATC TAGTCAC
3x40_6	Oligonucleotides	GTGACTAGATCTAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAATGCATAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAT ATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAATTCGAAGTGACT	AGTCACTTCGAATTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTGATATCTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTATGCATTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TAGATCTAGTCAC
A120	PCR	GTGACTGCTAGCTAATACGACTCACTATAGGGAG	AGTCACTTCGAATTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTAGATCTTTAC ACGGCGATCTTGCCGCCCTTC
ACH[79]	Oligonucleotides	AGATCTAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAATGCATCCCCCCCCCCCCCCCCCCCCCCCCCCC CCCAAAGGCTCTTTTCAGAGCCACCAGAATTC**TTC** **GAA**GTGACT	AGTCACTTCGAAGAATTCTGGTGGCTCTGAAA AGAGCCTTTGGGGGGGGGGGGGGGGGGGG GGGGGGGGGGGATGCATTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTAGATCT
2x60_C	PCR	GTGACTGCTAGCTAATACGACTCACTATAGGGAG	AGTCACTTCGAATTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTAGATCTTTA CACGGCGATCTTGCCGCCCTTC
2x60_G	PCR	GTGACTGCTAGCTAATACGACTCACTATAGGGAG	AGTCACTTCGAATTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTAGATCTTTA CACGGCGATCTTGCCGCCCTTC
2x60_T	PCR	GTGACTGCTAGCTAATACGACTCACTATAGGGAG	AGTCACTTCGAATTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTAGATCTTTA CACGGCGATCTTGCCGCCCTTC
2x60_12nt	PCR	GTGACTGCTAGCTAATACGACTCACTATAGGGAG	AGTCACTTCGAATTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TCCTCTAATGGCGTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTAGATCTTTACACGGCGATCTTGCCGCCCTTC
2x60_24nt	PCR	GTGACTGCTAGCTAATACGACTCACTATAGGGAG	AGTCACTTCGAATTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TAGATGCGGACACAATCAGGGGTTGTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTAGATCTTTACACGGCGA TCTTGCCGCCCTTC

5.2.2. Generation of mRNA

To generate *in vitro* transcribed mRNA, plasmids were linearized by *BstBI* (Thermo Fisher Scientific) digestion and purified by chloroform extraction and ethanol precipitation. Purified linear plasmids were used as a template for *in vitro* transcription. Plasmid templates (0.5 µg/µl) were subjected to *in vitro* transcription using 3 U/µl T7 RNA polymerase (Thermo Fisher Scientific), transcription buffer II (Ethris GmbH), 1 U/µl RiboLock RNase inhibitor (Thermo Fisher Scientitic), 0.015 U/µl inorganic pyrophosphatase 1 (Thermo Fisher Scientific) with a defined choice of natural and chemically modified ribonucleotides (Jena Biosciences).

RNA with modification set 1 was synthesized using 5-methylcytidine (25%) and 2-thiouridine (25%), in addition to unmodified nucleotides. For modification set 2, instead of 5-methylcytidine (25%) and 2-thiouridine (25%), 5-iodouridine (35%) and 5-iodocytidine (7.5%) were used. The complete IVT-mix was incubated at 37°C for 2h. Afterward, 0.01 U/µl DNase I (Thermo Fisher) was added for an additional 45 min at 37°C to remove the plasmid template. RNA was precipitated with ammonium acetate at a final concentration of 2.5 mM, followed by two washing steps with 70% ethanol. The pellet was re-suspended in *aqua ad injectabilia*. A C1-m7G cap structure was added enzymatically by 0.5 mM *Vaccinia* Virus Capping Enzyme (NEB) to the 5' end of the previously denatured transcript (1 mg/ml) at 80°C for 5 min. The capping reaction mix contained 1x capping buffer (NEB), 0.5 mM GTP (NEB), 0.2 mM S-methyladenosine (NEB), 2.5 U/µl mRNA Cap 2'-o-Methyltransferase (NEB) and 1 U/µl RiboLock RNase Inhibitor (Thermo Fisher Scientific). The capping mixture was incubated for 60 min at 37°C, followed by RNA precipitation with ammonium acetate at a final concentration of 2.5 mM and two washing steps with 70% ethanol. The pellet was re-suspended in *aqua ad injectabilia*.

RNA quality and concentration were measured spectrophotometrically on a NanoDrop2000C (Thermo Fisher Scientific). Its correct size and purity were determined *via* automated capillary electrophoresis (Fragment Analyzer, Advanced Analytical).

5.2.3. Cell culture

A549 (ACC-107) and HEK293 (ACC-305) cells were purchased from DSMZ. 16HBE14o- cells were kindly provided by Prof. Weber (University of Münster, Germany).

All cells were cultivated in Minimum Essential Media (MEM) with Glutamax (Gibco/Life Technologies). Media were supplemented with 10% heat-inactivated fetal bovine serum (FBS; Gibco/Life Technologies) and 1% penicillin/streptomycin (Gibco/Life Technologies). Cells were cultured in a humidified 5% CO_2 incubator at 37°C.

5.2.4. *In vitro* transfection

A549 and HEK293 cells were seeded at the density of $2x10^4$ cells/well and $4x10^4$ cells/well respectively in a 96 well plate, for the purpose of firefly luciferase, FACS measurements, and EPO ELISA assay. 16HBE14o- cells were seeded in a 6-well plate at the density of $7.5x10^5$ cells/well, for the purpose of Western blot analysis. At 24 hours post-seeding, cells were transfected with specific mRNAs using the commercial transfection reagent Lipofectamine®2000 (Thermo Fischer Scientific). Complexes were prepared at a ratio of 2 µl Lipofectamine®2000 per 1 µg mRNA. A549 cells were transfected with a dose of 250 ng/well. In HEK cells, transfections were performed using two doses: 250 and 125 ng/well mRNA. For experiments in A549 and HEK293 cells, required amounts of mRNA were diluted in water and the needed amounts of Lipofectamine®2000 in serum-free MEM. mRNA was added to the Lipofectamine®2000 solution followed by 20 min incubation at RT. The concentration of the final mRNA/Lipofectamine®2000 solution was 25 ng/µl and 12.5 ng/µl. 10 µl of the complex solution was added to the cells and cells were incubated for 24 h. For every mRNA construct, replicates of three or six were prepared. For 16HBE14o- cells, Lipofectamine®MessengerMax was used due to its superior transfection efficiency (data not shown). For transfection, 7.5 µg mRNA was diluted in 125 µl water, and 11.25 Lipofectamine®MessengerMax separately in 125 µl serum-free MEM. The mRNA solution was added to the

Lipofectamine®MessengerMax solution followed by 5 min incubation time at RT. A total volume of 250 µl of the lipoplex solution was added to the cells containing 2 ml normal growth media. The media was changed 4h after transfection.

5.2.5. Flow cytometry analysis for d2EGFP

Cells were washed with PBS, detached with TrypLE (Gibco/Life Technologies), and re-suspended in flow cytometry buffer (PBS supplemented with 10% FBS). Shortly before measurement, cells were stained with propidium iodide for discrimination between live and dead cells (1 µg/mL; Sigma Aldrich). The analysis was performed on an Attune Acoustic Focusing Cytometer (Life Technologies) with Attune Cytometric Software (version 2.1; Life Technologies) and FlowJo (version 10).

5.2.6. Firefly Luciferase Assay

For detection of firefly luciferase activity, the assay was performed 24h post-transfection. Cells were washed with PBS, followed by addition of 100 µl lysis buffer (25 mM Tris-HCl, 0.1% TritonX-100, pH 7.4). Cells were shaken for 20 min at room temperature. After lysis, 50 µl of the cell lysate was used to measure luciferase activity *via* photon luminescence emission for 5 s using InfiniteR 200 PRO (Tecan). The protein amount in each sample was quantified in 5 µl of the cell lysate with BioRad protein assay (Bio-Rad), using bovine serum albumin as a standard. Luciferase values were normalized to the protein concentrations.

5.2.7. Western Blot Analysis for human CFTR

Cells, washed with PBS and collected into a tube were lysed with RIPA buffer for 30 min on ice (50 mM Tris, pH 8.0, 150 mmol/L NaCl, 1% Triton X-100, 0.5% sodium deoxycholate, and 0.1% sodium dodecyl sulfate), followed by a 10 min centrifugation at 4°C. Total protein amount from the

supernatant was determined by the bicinchoninic acid (BCA) assay, following the manufacturer's instructions (Thermo Fisher Scientific). Cell lysates were separated on 3%-8% TRIS-Acetate gels (Thermo Fisher Scientific) and transferred to polyvinylidene fluoride (PVDF) blotting membranes (Bio-Rad). Membranes were blocked in Western Breeze blocking buffer (Thermo Fisher Scientific) and probed with antibodies against CFTR (R&D Systems MAB25031; 1:2.000) and Hsp90 (Origene TA500494; 1:15.000). Horseradish peroxidase (HRP)-conjugated anti-mouse antibody (1:10.000; ab6820; Abcam) was used as secondary antibody. Blots of CFTR were developed using Super Signal West Femto (Thermo Scientific), and of Hsp90 using Luminata Forte Western HRP Substrate (Millipore).

5.2.8. Enzyme-linked immunosorbent assay for hEPO

Quantification of hEPO protein in cell supernatants was performed using human Erythropoietin Quanktine IVD ELISA kit (R&D Systems) following manufacturer's instructions.

5.2.9. RNA isolation and reverse transcription

RNA was isolated at different time points post-transfection using Single Shot Cell Lysis kit (Bio-Rad) following manufacturer's protocol. From the lysates (1 μg of RNA), cDNA was synthesized using iScript Select cDNA Synthesis kit (Bio-Rad) with oligo(dT) primers following the manufacturer's instructions. The synthesized cDNA was stored at -20°C.

5.2.10. Quantitative real-time Polymerase Chain Reaction (qPCR)

Real-time qPCR was performed with short hydrolysis probes for d2EGFP and Luciferase targets (Universal Probe Library #37 and #29; Roche) on a Roche Light Cycler 96 (Roche Diagnostics). For d2EGFP, the following primers were used: 5'-cctgaagttcatctgcacca-3'and 5'-ctcgtgaccaccctgacc-3'.

Luciferase mRNA was quantified using the following primer, 5'-acgccgagtacttcgagatg-3' and 5'-attcagcccatagcgcttc-3'. Absolute mRNA values were calculated by interpolation from the standard curve. The dilutions for the standard curve were made using IVT mRNA in 1:10 ratio, starting with the concentration of 1 ng/µl to $1x10^{-5}$ ng/µl. 1 µl of each standard was mixed with 3 µl of untransfected cells, collected at the last time point, and reversely transcribed.

5.2.11. Statistical analysis

Each experiment was performed with at least three technical replicates per sample. Results are shown as means ±SD unless otherwise stated. Statistical analysis was performed using GraphPad Prism software (version 6). Data were tested for normal distribution using D'Agostino-Pearson omnibus normality test. Multiple comparisons were conducted by two-way ANOVA, followed by Sidak's test (pairwise comparison) or Dunnett's test (many-to-one comparison). A p-value ≤ 0.05 was considered statistically significant.

5.3. Results

5.3.1. Design of modified / segmented poly(A) tails

For the purpose of producing recombinant RNA with segmented poly(A) tails, the corresponding DNA sequences were cloned downstream of the gene of interest into a circular plasmid vector. Segmented poly(A) constructs including their separator attributes, tested in the current study are schematically presented as Figure 16. The most conventionally used poly(A) tail, comprising of 120 adenosines ($poly(A)_{120}$) was used as a benchmark[72,73,77,84,126,133]. It was split into either two or three segments of As. Two equal segments separated by 6 nucleotides contained 60 As each ($poly(A)_{2x60\_6}$). In the case of three equal segments, each segment comprised of 40 As and a separator of 6 nucleotides among each of the segment ($poly(A)_{3x40\_6}$). With an aim of exploring the effect of the spacer between the two segments of 60 As, five additional constructs were made with variable length of the spacer: 12 nucleotides ($poly(A)_{2x60\_12}$), 24 nucleotides ($poly(A)_{2x60\_24}$) or a single nucleotide ($poly(A)_{2x60\_C}$, $poly(A)_{2x60\_G}$, $poly(A)_{2x60\_T}$).

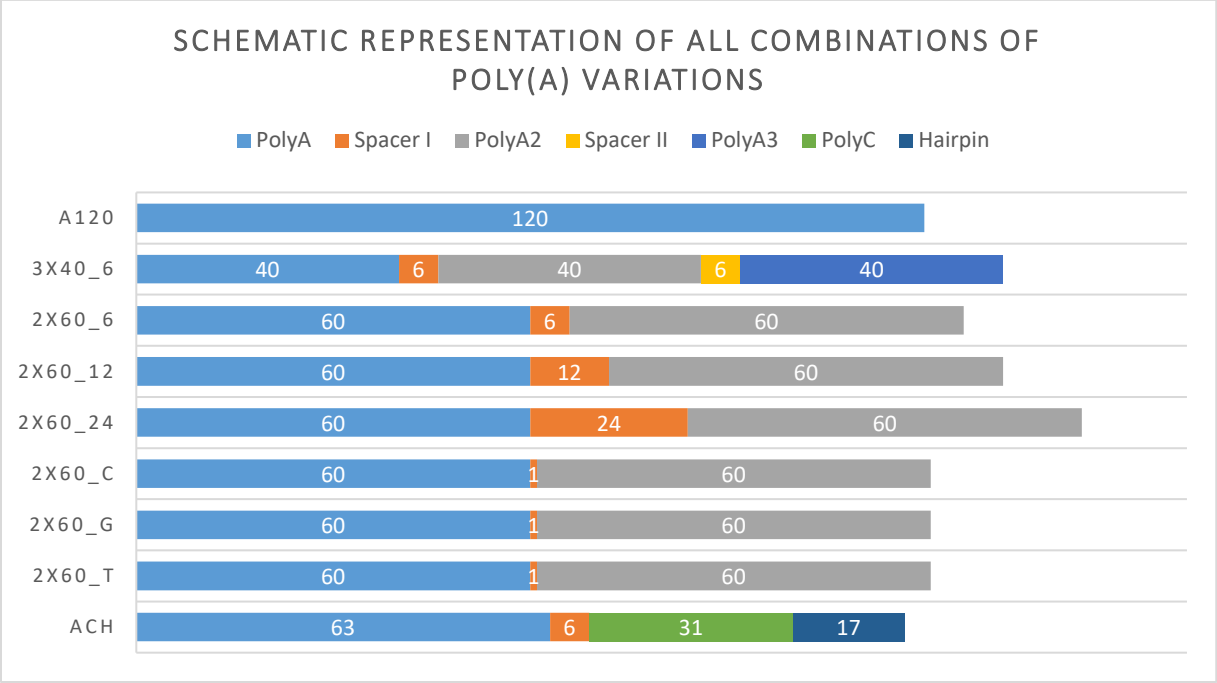

Figure 16. Schematic representation of all combinations of Poly(A) modifications tested in the current study.

Different homo- or hetero-polymeric poly(A) stretches were inserted downstream of the gene of interest (GOI).

5.3.2. Segmented poly(A) tails drastically reduce recombination of poly(A) containing plasmids in *E.coli*

The main reason for technical difficulties during a large-scale DNA template production with encoded poly(A) tail of sufficient length is instability of poly(A) containing sequences in circular plasmids[134,135]. Here, it was investigated if the application of mRNA containing segmented poly(A) tails offered any advantages with regard to plasmid recombination after transformation into *E.coli*, compared to a benchmark $Poly(A)_{120}$. The purpose was to see if the results of reduced recombination are sequence-dependent or can be widely applied to different sequences. Coding sequences (ORF) of three different genes (hEPO: 0.7 kb, d2EGFP: 0.9 kb, Luc2: 1.7 kb) were combined with each of the three poly(A) formats, $poly(A)_{120}$, $poly(A)_{2x60\_6}$, and $poly(A)_{3x40\_6}$ and cloned into a pUC57-Kanamycin (Genescript) vector. Clones with insert were screened for the stability of the poly(A) region *via* restriction digestion of the poly(A) region and resolving its size *via* capillary gel electrophoresis (fragment analyzer). Figure 17 shows the distribution of recombinant clones for each of the three tested poly(A) constructs. More than 50% of the clones containing homologous poly(A) tail, $poly(A)_{120}$, recombined. Use of segmented poly(A) significantly reduced recombination of plasmids with the strongest effect seen for $poly(A)_{2x60\_6}$ with less than 20% recombined clones. As a similar trend was observed with all three ORF sequences, this effect appears to be sequence independent.

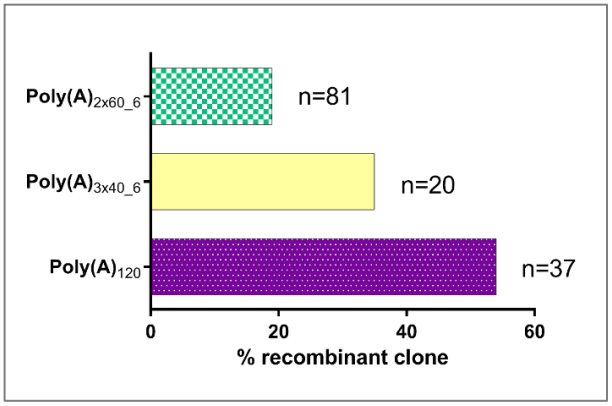

Figure 17. Quantification of poly(A) tail recombination for A120 and segmented poly(A) tails of poly(A)3x40_6 and poly(A)2x60_6.

Significantly reduced recombination of the poly(A) region was observed using segmented poly(A) tails. n: a total number of clones of d2EGFP, luciferase and hEPO sequences tested with a particular poly(A) format.

5.3.3. Effects of poly(A) segmentation on mRNA productivity

Reduced recombination of plasmids with segmented poly(A)s was the driver to further explore these constructs with respect to stability and translation efficiency of the resulting mRNAs. Destabilized EGFP with a relatively short protein half-life (d2EGFP) was chosen as a reporter protein for these initial studies. ORF sequence of d2EGFP was cloned into vectors containing different poly(A) formats. IVT mRNA was produced using modification set 1, using previously described protocols.[77,136] A549 cells were transfected with the mRNA, and at four different time points post-transfection (4, 24, 48 and 72h), the protein, as well as mRNA levels, were quantified. Since d2EGFP is a fluorescent protein, its quantitation was accessed using FACS, and mRNA was quantified with quantitative real-time reverse transcriptase PCR (qPCR). Comparable levels of d2EGFP protein were observed at all four time-points (Figure 18, a) for the compared poly(A) formats. mRNA decay kinetics were comparable for the three poly(A) formats at all time points, except at 24h post-transfection (Figure 18, b) where significantly higher amounts of mRNA could be quantified for the A120 format. Protein and mRNA amounts were used to calculate mRNA productivity, defined as the amount of protein (d2EGFP fluorescence intensity) normalized to the amount of mRNA (quantified *via* qPCR), for the compared constructs. At earlier time points (4h, 24h), segmented poly(A) mRNA was more productive than the one furnished with homogeneous poly(A) tail (Figure 18, c) resulting in 50-70% more protein per mRNA molecule.

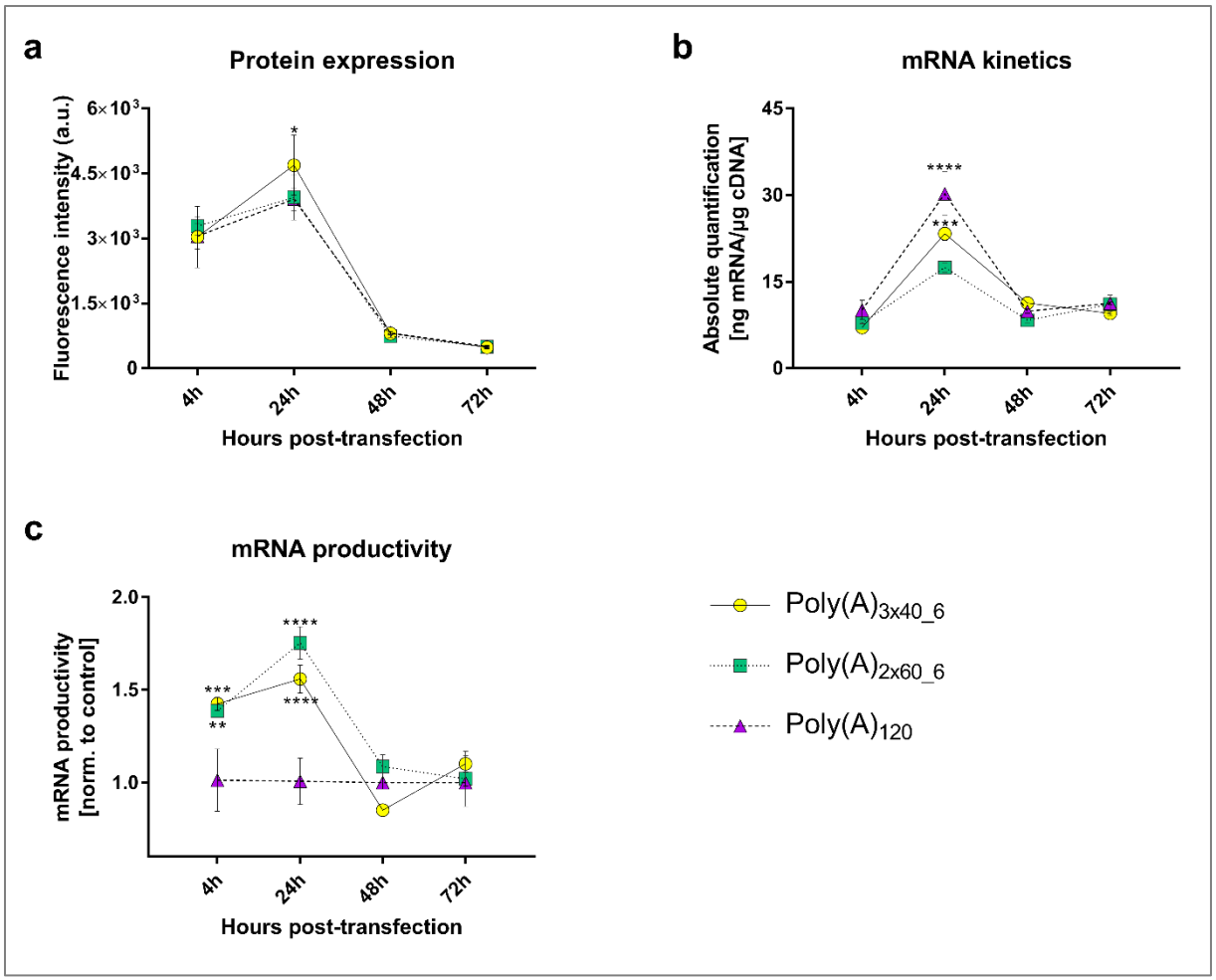

Figure 18. Determination of d2EGFP protein expression and mRNA decay kinetics of different poly(A) containing d2EGFP mRNAs post-transfection in A549 cells.

*(a) Median fluorescence intensity at 4, 24, 48 and 72h post-transfection, measured by FACS in A549 cells. (b) d2EGFP mRNA decay kinetics in A549 cells. (c) mRNA productivity was calculated by dividing the median fluorescence intensity (FACS data; a) by the mRNA amounts (real-time PCR data; b) and normalizing these ratios to those observed with poly(A)$_{120}$ construct. Statistical significance was assessed by two-way ANOVA test with p-values: *p<0.5, **p<0.01, ***p<0.001, ****p<0.0001, n=3.*

To rule out sequence dependency of the observed data with d2EGFP, the different poly(A) formats (poly(A)$_{2x60\_6}$, poly(A)$_{3x40\_6}$ vs. poly(A)$_{120}$) were also investigated using luciferase. Besides sequence, nucleotide modifications have also been shown to influence the functionality of non-coding sequence elements in mRNA[79]. For this reason, in subsequent experiments with luciferase both unmodified and modified (using modification set 1 and modification set 2) mRNAs were tested. A549 cells were transfected with corresponding luciferase-encoding mRNA, transcribed in all three

different nucleotide modifications; unmodified, modification 1 and modification 2.[136] Luciferase expression and mRNA decay kinetics for each construct were determined at 24h post-transfection. The time point was chosen based on the d2EGFP data where the maximum difference in mRNA half-life and productivity were seen at 24 h post mRNA delivery. In this experiment, in addition to A120, the performance of segmented poly(A) constructs was additionally compared to a previously published construct[79] comprising homopolymeric A stretch (A_{63}), homopolymeric C stretch (C_{31}) and histone stem-loop (Figure 19). Luciferase activity was significantly increased in a modification-independent manner when the cells were transfected with segmented poly$(A)_{2x60_6}$, compared to homogeneous poly$(A)_{120}$ and ACH benchmark (Figure 19). qPCR revealed that the differences between the mRNA amounts for the different poly(A) tail containing luciferase mRNAs, including different modifications, was only moderate. These data taken together demonstrated that the most productive construct was segmented poly$(A)_{2x60_6}$ if transcribed in modification sets 1-2. ACH construct resulted in significantly lower luciferase protein and mRNA productivity across all three types of mRNAs (unmodified and modified) compared to poly(A) containing constructs.

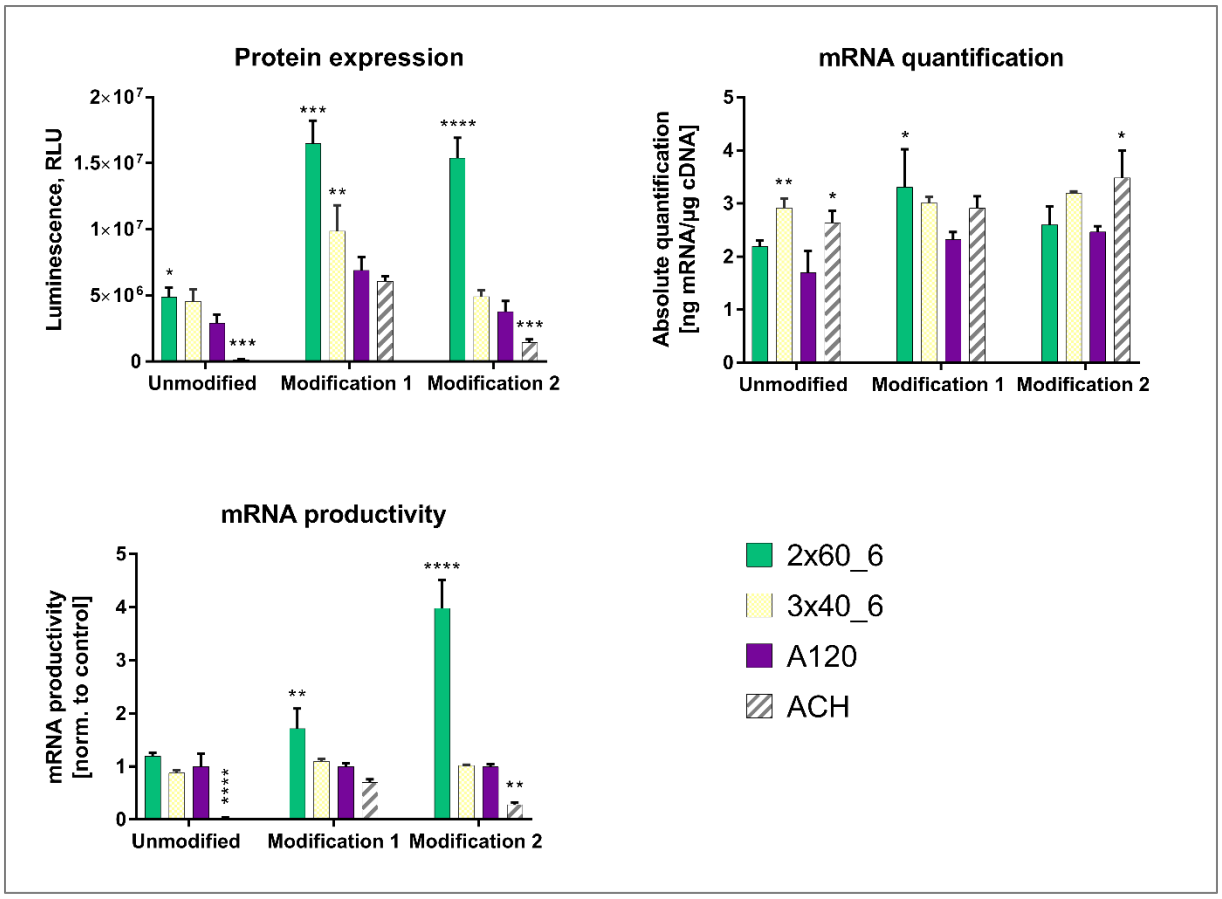

Figure 19. Determination of luciferase expression and mRNA decay kinetics of different poly(A) containing luciferase mRNA 24h post-transfection in A549 cells.

*(a) Luciferase activity in protein lysates from A549 cells transfected with different poly(A) containing luciferase RNA measured 24h post-transfection. (b) Luciferase mRNA quantification in A549 cells. (c) mRNA productivity was calculated by dividing the luciferase expression values (RLU; a) by the mRNA amounts (real-time PCR data; b) and normalizing these ratios to those observed with Poly(A)$_{120}$ construct. Statistical significance was assessed by two-way ANOVA test with p-values: *p<0.5, **p<0.01, ***p<0.001, ****p<0.0001, n=6.*

5.3.4. Effects of poly(A) segmentation on the translation of physiological targets

Due to encouraging results from previous experiments using segmented poly(A) tails in a combination with intracellular reporter proteins, such as d2EGFP and luciferase, the concept of poly(A)$_{2x60\_6}$ was further tested in physiological targets. Human EPO was selected as an example of a secretory protein (length of a coding region of 0.9 kb). As an example of a membrane protein, human cystic fibrosis transmembrane conductance regulator (CFTR) with a coding region of 4.5 kb was used.

Codon-optimized sequence (optimized using GeneOptimizer) for hEPO was cloned upstream of the poly(A)$_{120}$ or poly(A)$_{2x60\_6}$ version of the tail, in the pUC57-Kanamycin vector. The resulting plasmids were used as a template for mRNA production *via in vitro* transcription incorporating unmodified, modification 1 or modification 2 sets of nucleotides. Transfection experiments were performed in human HEK293 cells with two different doses of mRNA (250ng/well and 125ng/well). At three time points post-transfection (24h, 48h, 72h), EPO protein concentrations were determined *via* ELISA. As Figure 20 shows, among the compared poly(A) formats at any of the applied doses, time points or modifications, no significant difference in protein expression was observed. An exception of this observation is the expression of unmodified mRNA at 24h and 72h post-transfection and mRNA in modification set 1 at 72h.

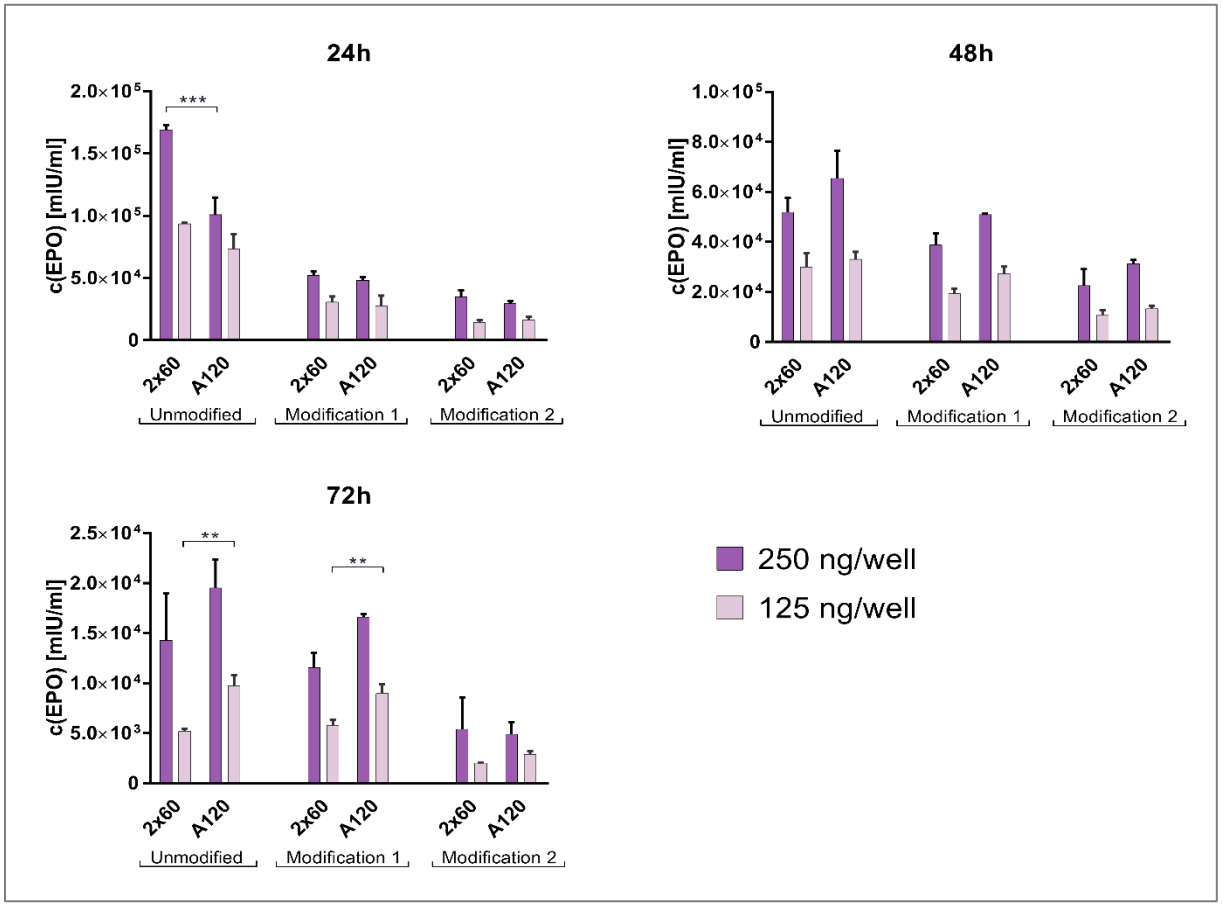

Figure 20. Quantification of secreted human erythropoietin protein levels as measured via ELISA in supernatants from HEK293 cells transfected either with poly(A)120 or poly(A)2x60 containing EPO mRNA at 24h (a), 48h (b) and 72h (c) post-transfection.

*Values represent mean ± SD of three replicates. Statistical significance was assessed by two-way ANOVA test with p-values: **p<0.01, ***p<0.001, n=3.*

As an example of a transmembrane protein, codon optimized sequence for human CFTR coding sequence was cloned upstream of either poly(A)$_{2x60\_6}$ or poly(A)$_{120}$ variant of the poly(A) tail. 16HBE14o- cells were transfected with unmodified IVT mRNA, and cells were lysed for the further procedure of Western blotting after 24h and 48h. The expression of CFTR was compared to expression of a housekeeper gene Hsp90, to obtain a normalized value of the expression. This experiment lead to similar observation as already seen with previous protein targets; comparable protein expression levels from both the compared poly(A) formats (Figure 21).

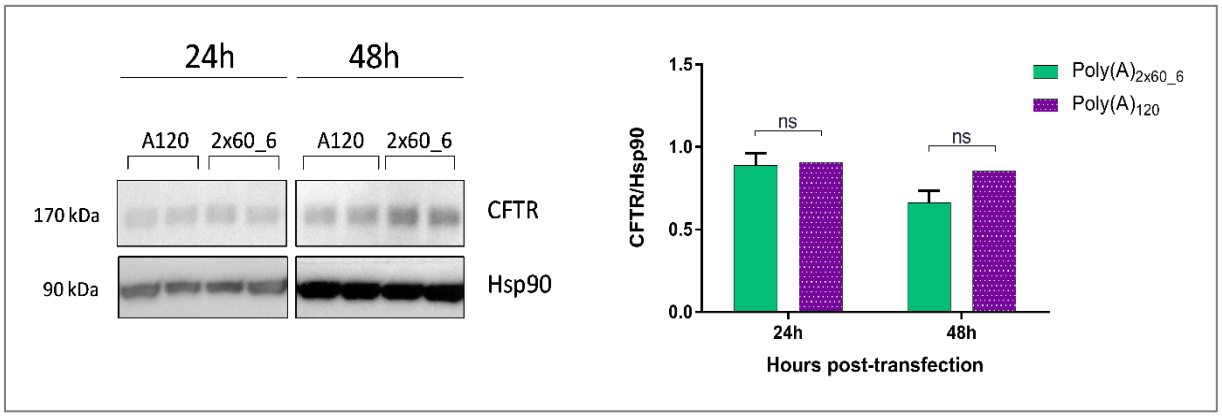

Figure 21. Relative quantification of CFTR protein in 16HBE14o- lysates as measured via Western blot.

16HBE14o- cells were transfected either with poly(A)$_{120}$ or poly(A)$_{2x60\_6}$ containing hCFTR mRNA and protein lysates analyzed at 24 and 48h post-transfection. (b) Densitometry analysis of western blot images. Values represent mean ± SD of two replicates. Statistical significance was assessed by paired t-test with p-values: ns = p>0.5.

5.3.5. Spacer region expansion in poly(A)$_{2x60}$

After getting confirmation from the results with physiological genes, that segmented poly(A) with 6 nucleotides as a spacer between the segments does not negatively affect translation, the effects of different spacer lengths were tested. As the starting point, the spacer was expanded from 6 nucleotides (poly(A)$_{2x60\_6}$), to 12 (poly(A)$_{2x60\_12}$) and 24 nucleotides (poly(A)$_{2x60\_24}$). Three constructs harboring luciferase coding region upstream of poly(A) tail were cloned, and further *in vitro* transcribed into mRNAs as unmodified and in modification set 1-2. Transfection was done in A549 cells, and measurements were performed 24h post-transfection. Observing the results at

unmodified- and modification 1- contacting mRNA, longer spacers express significantly lower amounts of luciferase (Figure 22). Comparable levels of luciferase mRNA could be quantified in the cells for all constructs regardless of modification, with a single exception. The same applies to mRNA productivity. Segmented poly(A) tails with an increased spacer length of more than 6 nucleotides did not result in any significant advantage, neither in translation efficiency nor in mRNA stability.

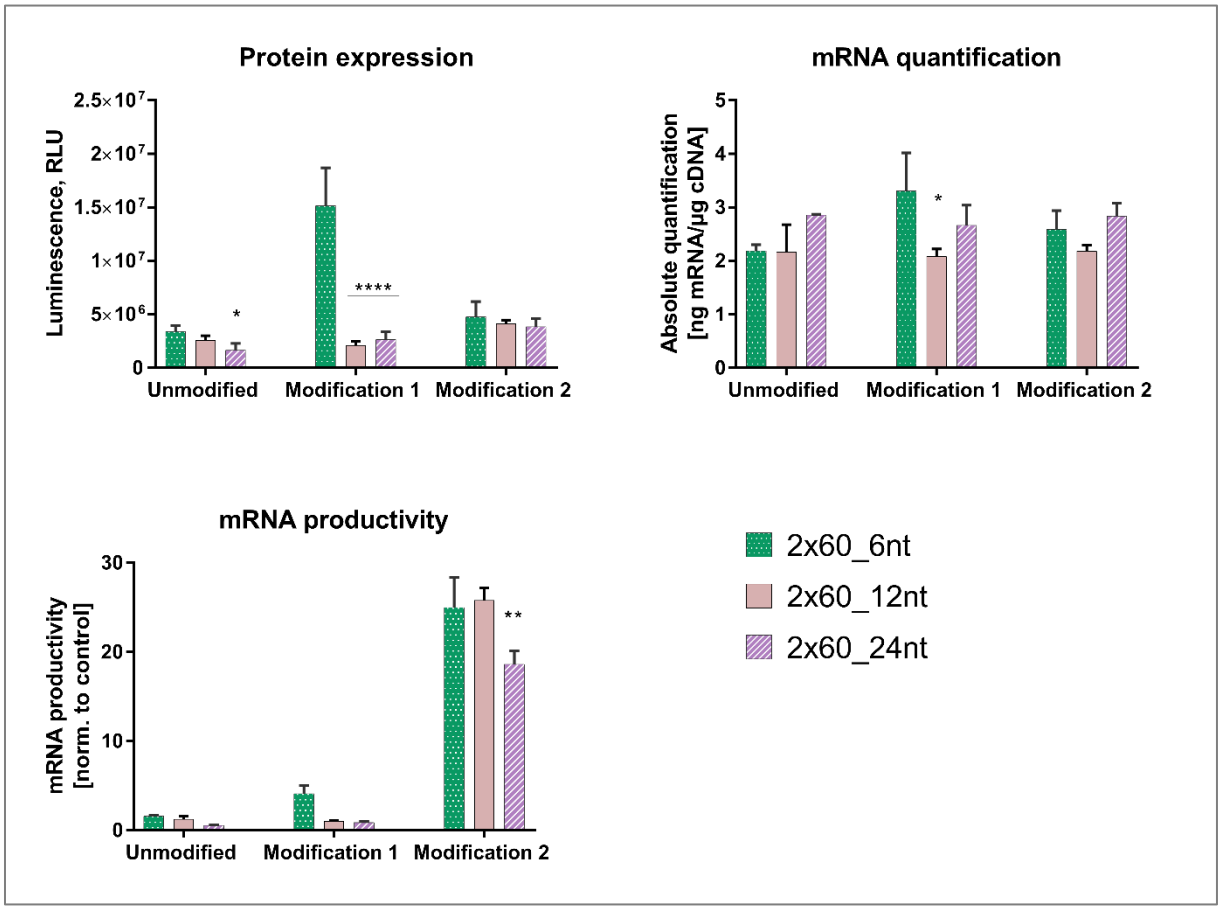

Figure 22. Determination of luciferase expression and mRNA quantification of different poly(A) containing luciferase mRNA at 24h post-transfection in A549 cells.

*(a) Luciferase activity in protein lysates from A549 cells transfected with different poly(A) containing luciferase mRNA. (b) Luciferase mRNA quantification in A549 cells transfected with different poly(A) containing luciferase mRNA. (c) mRNA productivity was calculated by dividing the luciferase expression values (RLU; a) by the mRNA amounts (real-time PCR data; b) and normalizing these ratios to those observed with Poly(A)$_{120}$ construct. Statistical significance was assessed by two-way ANOVA test with p-values: *p<0.5, **p<0.01, ****p<0.0001, n=6.*

5.3.6. Spacer region reduction in poly(A)$_{2x60}$

With a similar idea in mind as described above, the effect of reducing the spacer length to a single nucleotide in segmented poly(A)$_{2x60}$ tail on protein expression and mRNA productivity was examined. As the variability of options is limited by the number of nucleotides available in naturally occurring constructs, we created three new constructs, each harboring either C, T or G as a spacer nucleotide between two poly(A) segments in the tail. Also, mRNAs in unmodified, and both modification versions (modification 1 and 2) were produced by IVT and used for transfection of A549 cells. Expression levels, measured at 24h post-transfection were benchmarked to a standard, homogeneous poly(A)$_{120}$. As seen from Figure 23, all three constructs resulted in higher luciferase expression when compared to poly(A)$_{120}$, regardless of the spacer nucleotide or modification choice. Most of the mRNAs quantified from each construct were at comparable levels. Considering the mRNA productivity, all the constructs with segmented poly(A) were more productive than homogeneous poly(A)$_{120}$.

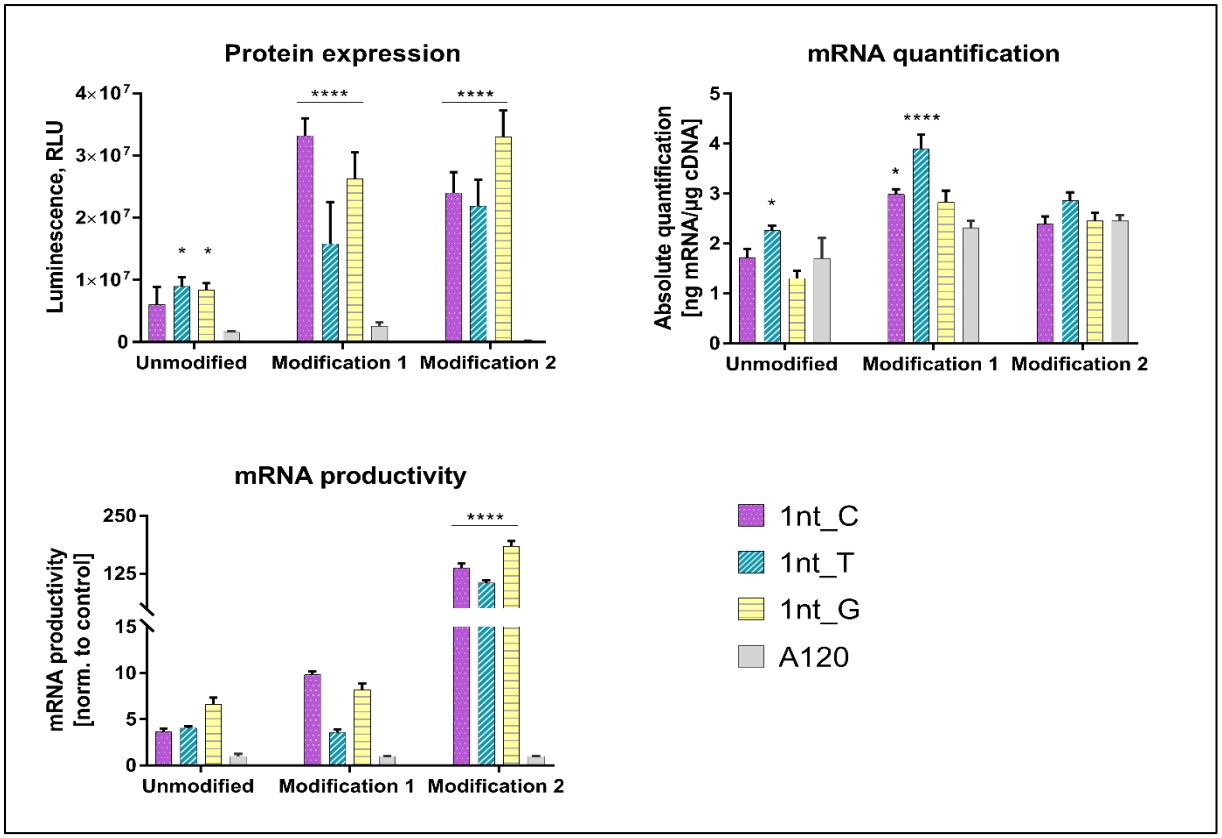

Figure 23. Determination of luciferase expression and mRNA quantification of different poly(A) containing luciferase mRNA at 24h post-transfection in A549 cells.

*(a) Luciferase activity in protein lysates from A549 cells transfected with different poly(A) containing luciferase mRNA. (b) Luciferase mRNA quantification in A549 cells transfected with different poly(A) containing luciferase mRNA. (c) mRNA productivity was calculated by dividing the luciferase expression values (RLU; a) by the mRNA amounts (real-time PCR data; b) and normalizing these ratios to those observed with Poly(A)$_{120}$ construct. Statistical significance was assessed by two-way ANOVA test with p-values: *p<0.5, ****p<0.0001, n=6.*

In contrast to luciferase expression where no difference was observed between the three single nucleotide spacers, choice of the spacer nucleotide influenced recombination efficiency of the poly(A) tail in *E.coli*. Highest recombination rate, comparable to values observed with poly(A)$_{120}$ (approx. 50%) was seen when C was used as a spacer in contrast to no recombination with G as single nucleotide spacer between the two segments (Figure 24).

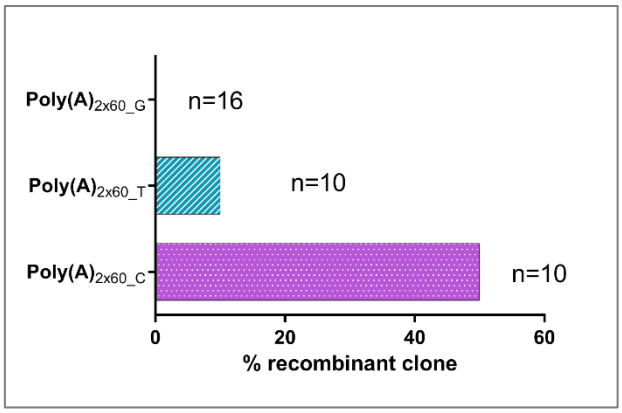

Figure 24. Quantification of poly(A) tail recombination rate for segmented poly(A) tails with a single nucleotide spacer.

n: a total number of clones of luciferase tested with a particular poly(A) format.

Taken together, these results show that segmented poly(A)$_{2x60}$ with either a 6 or a single nucleotide (G/T) spacer offers advantages for use in RNA production by plasmid-based vectors. Due to their reduced recombination rate compared to conventionally used poly(A)$_{120}$, they could significantly simplify the DNA template production with template-encoded poly(A) tails. Moreover, they did not show any negative effect on protein expression and mRNA half-life.

5.4. Discussion

The motivation for segmenting the poly(A) tail into smaller fragments separated by spacer elements arose after technical difficulties during the DNA template production. As the most common poly(A) tail is composed of 120 adenosines, its length served as the basis for segmenting the poly(A) tail. The constructs were segmented into either two or three parts, each comprising an equal number of nucleotides. In the case of 2 segments, each part contained 60 adenosines ($poly(A)_{2x60\_6}$), and each of the 3-part poly(A) contained 40 adenosines ($poly(A)_{3x40\_6}$). Restriction site of *Nsi*I was used as a spacer of 6 nucleotides. With this approach, a physical distance between two polyadenosine sequences and the restriction site could still be used to check the size of each fragment. Another reason for keeping the minimum length of 40 nucleotides per segment is to enable the binding of multimers of PABP. It has been reported that a minimum of 12 adenosines is needed for the binding of a single PABP molecule[129–131], but also that a monomer of PABP, although binding to eIF4 complex, was not enough to support translation[128]. Therefore, $poly(A)_{3x40\_6}$ and $poly(A)_{2x60\_6}$ segments were designed to be long enough to host at least three PABPs per segments.

OFRs of target proteins, namely luciferase, d2EGFP, hEPO and CFTR were cloned upstream of different poly(A) variations. Purified ligations were transformed into a DH10B bacterial strain of *E.coli* and tested for insert and poly(A) length. Screening of the insert positive clones containing the $poly(A)_{120}$ format revealed that approximately 50% had shorter poly(A) than 120 adenosines implicating recombination in the homopolymer poly(A). Comparable rates of poly(A) recombination have been reported by Grier et al 2016.[127] The recombination was also shown to be sequence independent. However, constructs with segmented $poly(A)_{2x60\_6}$ had a 2-fold reduced rate of recombination. Considering the translation efficiency of the $poly(A)_{2x60\_6}$ in d2EGFP, it resulted in comparable values to those observed with $poly(A)_{120}$, and the stability of the mRNA was not affected. Apart from the conventionally used $poly(A)_{120}$ format, the segmented poly(A) was additionally compared to ACH, published by Thess et al.[79] It is composed of a homopolymeric A stretch, a homopolymeric C stretch, and a histone stem-loop. The functions of poly(A) are overtaken by the

conserved stem-loop [137,138]. This approach was tested in the experiments with luciferase, and significantly higher expression was observed with the usage of poly(A)$_{2x60\_6}$ than either ACH or poly(A)$_{120}$. There were no significant effects on mRNA stability.

Encouraged with the results observed with reporter proteins (d2EGFP, luciferase), the next step was to investigate the segmented poly(A)$_{2x60\_6}$ effect on translation of physiological genes. A different class of proteins was chosen since both luciferase and d2EGFP are intracellularly localized. As an example of a secretory protein, EPO was chosen. It is normally secreted by the kidney and stimulates the production of red blood cells. CFTR, which serves as an ion channel conducting chloride and thiocyanate ions was selected as an example of a transmembrane protein. Its function is impaired in patients suffering from cystic fibrosis. Regarding the translated proteins detected post-transfection in relevant cell lines with respective mRNAs differing in the poly(A) format, no significant differences could be observed with any of the physiological targets. The same trend was observed regardless of the time point and the usage of modifications when EPO was used as a target protein. In the case of CFTR, experiments were performed only using unmodified mRNA, because a recent study[132] has demonstrated functional restoration of CFTR in human CF airway epithelia after transfection with unmodified CFTR mRNA.

Further, the varying length of the spacer positioned between the A segments and its effect on protein expression and recombination was investigated with luciferase as a reporter gene. When the length of the spacer was increased from 6 nucleotides to 12 or 24 nucleotides, no specific advantages could be observed. Rather, if using modification 1 and prolonged spacers, protein expression was lower than with segmented poly(A)$_{2x60\_6}$. A possible explanation for such an outcome with modified mRNA is that incorporation of the modified nucleotides within the spacer region could affect the binding of poly(A) binding protein to the two segments of poly(A). Reduction of the spacer length to a single nucleotide significantly increased protein expression, regardless of chosen nucleotide (T, C or G), when compared to poly(A)$_{120}$. mRNA quantification confirmed comparable levels for all the three spacers. This implies that segmented poly(A) with single nucleotide spacer positively affected mRNA

translation. Considering recombination rates of constructs with single nucleotides, the reduction of it was observed. As recombination with poly(A)$_{2x60\_6}$ was already lower (20% compared to 50% with poly(A)$_{120}$), incorporation of a single G as a spacer further reduced it to zero.

All these results taken together imply that the use of segmented poly(A)$_{2x60}$ with either 6 or 1 nucleotide spacer (G/T), cloned into plasmid vectors, significantly reduces recombination in *E.coli* without negatively affecting translation and mRNA stability when compared to the relatively unstable and widely used poly(A)$_{120}$.

5.5. Conclusion

A general procedure of mRNA production for therapeutic purposes involves either a PCR or a plasmid DNA template containing all of the necessary elements, such as the coding sequence, UTRs, and poly(A) tail. When a homopolymeric stretch such as poly(A) is cloned into a conventional plasmid vector, a common phenomenon is its recombination and therefore shortening. This is a major risk which limits the use of plasmid DNA vectors containing poly(A) tails form large-scale RNA production. A promising approach to target this issue is splitting the poly(A) tail into 2 segments, each consisting of 60 adenosines, separated by a spacer element. Using such a segmented poly(A) significantly reduced recombination of the plasmid in *E.coli* and mRNAs containing such segmented poly(A)s were comparable in terms of their translation efficiency and intracellular stability. These effects were independent of sequence and nucleotide modification(s) used in the mRNA production reactions and as such present segmented poly(A) as a promising variation of the poly(A) tail for the ease of plasmid production.

6 FUTURE PERSPECTIVES

In the first part of the work, it was shown that use of minimal 5'-UTRs yielded to higher protein expression. These UTRs were designed by taking only those sequence elements which were needed for transcription (T7 promoter), and translation (translation regulator). The results from this section highlight the superior performance of such short synthetic 5'-UTRs over naturally occurring, long 5'-UTRs. The latter, due to their longer sequence are often susceptible to structural changes upon incorporation of chemically modified nucleotides, are cell-specific and their performance is often influenced by the downstream sequence[79] which was not the case with minimal 5'-UTRs.

Segmented poly(A) tails significantly reduce recombination in *E.coli* during plasmid amplification without affecting either the half-life or translation of the *in vitro* produced mRNA using such plasmids as template. Based on these findings, its incorporation into the standard DNA-template design would ease everyday procedure and significantly reduce the time invested in screening and finding the clones of the right length.

The research presented in this thesis has opened a number of research-based and pragmatic questions that should be addressed in the future. In Chapter 4, a high performance of short synthetic 5'-UTRs were demonstrated by both, *in vitro* and *in vivo* experiments. The actual potential of these synthetic minimalistic UTR sequences for transcript therapies would be confirmed in disease animal models using physiological genes impaired in metabolic diseases. The innovative approach towards plasmid-based templates for *in vitro* RNA production by segmenting the poly(A) tails was described in Chapter 5. With respect to that, a future focus should be put into understanding the molecular

mechanism behind reduced recombination of segmented poly(A) tails. In line with a broader scope of transcript therapies, such approaches may provide an attractive platform to simplify 5'-UTR mRNA design and to facilitate DNA-based template production with encoded poly(A) tail.

7 REFERENCES

1. Orgel LE. Evolution of the genetic apparatus. J. Mol. Biol. **38**(3), 381, 1968;
2. Crick FHC. The origin of the genetic code. J. Mol. Biol. **38**(3), 367, 1968;
3. Higgs PG, Lehman N. The RNA World: Molecular cooperation at the origins of life. Nat. Rev. Genet. **16**(1), 7, 2015;
4. Joyce GF. The antiquity of RNA-based evolution. Nature. **418**(6894), 214, 2002;
5. Guerrier-Takada C, Gardiner K, Marsh T, Pace N, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell. **35**, 849, 1983;
6. Kruger K, Grabowski PJ, Zaug AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. Cell. **31**(1), 147, 1982;
7. Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. Science. **292**(5520), 1319, 2001;
8. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosomal activity in peptide bond synthesis. Science. **289**(5481), 920, 2000;
9. Rodnina M V., Beringer M, Wintermeyer W. How ribosomes make peptide bonds. Trends Biochem. Sci. **32**(1), 20, 2007;
10. Wochner A, Attwater J, Coulson A, Holliger P. Ribozyme-catalyzed transcription of an active ribozyme. Science. **332**(6026), 209, 2011;
11. Ma W, Yu C, Zhang W, Wu S, Feng Y. The emergence of DNA in the RNA world: An in silico simulation study of genetic takeover. BMC Evol. Biol. BMC Evolutionary Biology; **15**(1), 1, 2015;

12. Lindahl T. Instability and decay of the primary structure of DNA. Nature. **362**(6422), 709, 1993;

13. Bashkin JK. DNA enzymes: new-found chemical reactivity. Curr. Biol. **7**(5), R286, 1997;

14. Ma W. What Does “the RNA World” Mean to “the Origin of Life”? Life. **7**(4), 49, 2017;

15. Weissman D. mRNA transcript therapy. Expert Rev. Vaccines. **14**(02), 1, 2014;

16. Crick F. Central dogma of molecular biology. Nature. **227**(5258), 561, 1970;

17. Franklin S, Vondriska TM. Genomes, Proteomes, and the Central Dogma. Circ. Cardiovasc. Genet. **4**(5), 576, 2011;

18. Lieberman J. Unveiling the RNA World. N. Engl. J. Med. **379**(13), 1278, 2018;

19. Nelson D. Lehninger Principles of Biochemisty. 5th ed. New York: W.H. Freeman, 2005;

20. Tinoco I, Bustamante C. How RNA folds. J. Mol. Biol. **293**(2), 271, 1999;

21. Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, et al. Structure of a ribonucleic acid. Science. **147**(3664), 1462, 1965;

22. Ghosh A, Lima CD. Enzymology of RNA cap synthesis. Wiley Interdiscip. Rev. RNA. **1**(1), 152, 2010;

23. Svitkin Y V, Ovchinnikov LP, Dreyfuss G, Sonenberg N. General RNA binding proteins render translation cap dependent. EMBO J. **15**(24), 7147, 1996;

24. Lewis JD, Izaurralde E. The role of the cap structure in RNA processing and nuclear export. Eur. J. Biochem. **247**(2), 461, 1997;

25. Hamm J, Mattaj IW. Monomethylated cap structures facilitate RNA export from the nucleus. Cell. **63**(1), 109, 1990;

26. Oliveira CC, McCarthy JEG. The relationship between eukaryotic translation and mRNA stability. A short upstream open reading frame strongly inhibits translational initiation and greatly accelerates mRNA degradation in the yeast Saccharomyces cerevisiae. J. Biol. Chem. 270(15), 8936, 1995;

27. Fan J, Kuai B, Wang K, Wang L, Wang Y, Wu X, et al. mRNAs are sorted for export or degradation before passing through nuclear speckles. Nucleic Acids Res. **46**(16), 8404, 2018;

28. Sachs A. The role of poly(A) in the translation and stability of mRNA. Curr. Opin. Cell Biol. **2**(6), 1092, 1990;

29. Wells SE, Hillner PE, Vale RD, Sachs AB. Circularization of mRNA by eukaryotic translation

initiation factors. Mol. Cell. **2**(1), 135, 1998;

30. Mignone F, Gissi C, Liuni S, Pesole G. Untranslated regions of mRNAs. Genome Biol. **3**(3), 2002;

31. Pesole G. Structural and functional features of eukaryotic mRNA untuntranslated regions. Gene. **276**, 73, 2001;

32. Dvir S, Velten L, Sharon E, Zeevi D, Carey LB, Weinberger A, et al. Deciphering the rules by which 5 ’-UTR sequences affect protein expression in yeast. Proc. Natl. Acad. Sci. U. S. A. **110**(30), E2792, 2013;

33. Wegrzyn JL, Drudge TM, Valafar F, Hook V. Bioinformatic analyses of mammalian 5’-UTR sequence properties of mRNAs predicts alternative translation initiation sites. BMC Bioinformatics. **9**, 232, 2008;

34. Asrani KH, Farelli JD, Stahley MR, Miller RL, Cheng CJ, Subramanian RR, et al. Optimization of mRNA untranslated regions for improved expression of therapeutic mRNA. RNA Biol. **15**(6), 756, 2018;

35. Ji X, Kong J, Carstens RP, Liebhaber S a. The 3’ untranslated region complex involved in stabilization of human alpha-globin mRNA assembles in the nucleus and serves an independent role as a splice enhancer. Mol. Cell. Biol. **27**(9), 3290, 2007;

36. Daylite. Diagramatic structure of a typical human protein-coding mRNA including the untranslated regions (UTRs).It is drawn approximately to scale. The cap is only one modified base, average en:5’ UTR length 170, en:3’ UTR 700 [Internet]. Wikipedia. 2008 [cited 2018 Sep 14]. Available from: https://commons.wikimedia.org/wiki/File:MRNA_structure.svg

37. Kozak M. Influence of mRNA secondary structure on binding and migration of 40S ribosomal subunits. Cell. **19**(1), 79, 1980;

38. Kozak M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. Proc. Natl. Acad. Sci. U. S. A. **83**(9), 2850, 1986;

39. Babendure JR, Babendure JL, Ding J, Tsien RY. Control of mammalian translation by mRNA structure near caps Control of mammalian translation by mRNA structure near caps. RNA. **12**, 851, 2006;

40. Sinvani H, Haimov O, Svitkin Y, Sonenberg N, Tamarkin-Ben-Harush A, Viollet B, et al. Translational tolerance of mitochondrial genes to metabolic energy stress involves TISU and eIF1-eIF4GI cooperation in start codon selection. Cell Metab. **21**(3), 479, 2015;

41. Wellensiek BP, Larsen AC, Stephens B, Kukurba K, Waern K, Briones N, et al. Genome-wide profiling of human cap-independent translation-enhancing elements. Nat. Methods. **10**(8), 747, 2013;

42. van der Velden AW, Thomas AA. The role of the 5' untranslated region of an mRNA in translation regulation during development. Int. J. Biochem. Cell Biol. **31**(1), 87, 1999;

43. Kozak M. New ways of initiating translation in eukaryotes? Mol. Cell. Biol. **21**(6), 1899, 2001;

44. Kozak M. A second look at cellular mRNA sequences said to function as internal ribosome entry sites. Nucleic Acids Res. **33**(20), 6593, 2005;

45. Conne B, Stutz A, Vassalli JD. The 3' untranslated region of messenger RNA: A molecular "hotspot" for pathology? Nat. Med. **6**(6), 637, 2000;

46. Ebert MS, Sharp PA. Roles for MicroRNAs in Conferring Robustness to Biological Processes. Cell. **149**(3), 515, 2012;

47. Jain R, Frederick JP, Huang EY, Burke KE, Mauger DM, Andrianova EA, et al. MicroRNAs Enable mRNA Therapeutics to Selectively Program Cancer Cells to Self-Destruct. Nucleic Acid Ther. **28**(5), 285, 2018;

48. Wahid F, Khan T, Kim YY. MicroRNA and diseases: Therapeutic potential as new generation of drugs. Biochimie. **104**(1), 12, 2014;

49. Bartel DP. MicroRNAs: Target Recognition and Regulatory Functions. Cell. **136**(2), 215, 2009;

50. Cai Y, Yu X, Hu S, Yu J. A Brief Review on the Mechanisms of miRNA Regulation. Genomics, Proteomics Bioinforma. **7**(4), 147, 2009;

51. Baker KE, Coller J. The many routes to regulating mRNA translation. Genome Biol. **7**(12), 1, 2006;

52. Bernal J a. RNA-based tools for nuclear reprogramming and lineage-conversion: Towards clinical applications. J. Cardiovasc. Transl. Res. **6**(6), 956, 2013;

53. Cannon G, Weissman D. RNA Based Vaccines. DNA Cell Biol. **21**(12), 953, 2002;

54. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, et al. LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. Science. **302**(5644), 415, 2003;

55. Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, McIntyre E, et al. A Serious Adverse Event after Successful Gene Therapy for X-Linked Severe Combined

Immunodeficiency. N. Engl. J. Med. **348**(3), 255, 2003;

56. Liu-Chen S, Connolly B, Cheng L, Subramanian RR, Han Z. MRNA treatment produces sustained expression of enzymatically active human ADAMTS13 in mice. Sci. Rep. **8**(1), 1, 2018;

57. Baker MP, Reynolds HM, Lumicisi B, Bryson CJ. Immunogenicity of protein therapeutics: The key causes, consequences and challenges. Self/Nonself - Immune Recognit. Signal. **1**(4), 314, 2010;

58. Weissman D. mRNA transcript therapy. Expert Rev. Vaccines. **14**(2), 265, 2015;

59. Wang Y, Su HH, Yang Y, Hu Y, Zhang L, Blancafort P, et al. Systemic delivery of modified mRNA encoding herpes simplex virus 1 thymidine kinase for targeted cancer gene therapy. Mol. Ther. **21**(2), 358, 2013;

60. John S, Yuzhakov O, Woods A, Deterling J, Hassett K, Shaw CA, et al. Multi-antigenic human cytomegalovirus mRNA vaccines that elicit potent humoral and cell-mediated immunity. Vaccine. **36**(12), 1689, 2018;

61. Lindgren G, Ols S, Liang F, Thompson EA, Lin A, Hellgren F, et al. Induction of Robust B cell responses after influenza mRNA vaccination is accompanied by circulating hemagglutinin-specific ICOS+ PD-1+ CXCR3+ T follicular helper cells. Front. Immunol. **13**(8), 1539, 2017;

62. Richner JM, Jagger BW, Shan C, Fontes CR, Dowd KA, Cao B, et al. Vaccine Mediated Protection Against Zika Virus-Induced Congenital Disease. Cell. **170**(2), 273, 2017;

63. Yin H, Song CQ, Suresh S, Wu Q, Walsh S, Rhym LH, et al. Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. Nat. Biotechnol. **35**(12), 1179, 2017;

64. Wolff J, Malone R, Williams P, Chong W, Acsadi G, Jani A, et al. Direct gene transfer into mouse muscle in vivo. Science. **247**(4949), 1465, 1990;

65. Circelli L, Petrizzo A, Tagliamonte M, Heidenreich R, Tornesello ML, Buonaguro FM, et al. Immunological effects of a novel RNA-based adjuvant in liver cancer patients. Cancer Immunol. Immunother. **66**(1), 103, 2017;

66. Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature. **547**(7662), 222, 2017;

67. Kreiter S, Diken M, Selmi A, Türeci Ö, Sahin U. Tumor vaccination using messenger RNA: Prospects of a future therapy. Curr. Opin. Immunol. **23**(3), 399, 2011;

68. Bahl K, Senn JJ, Yuzhakov O, Bulychev A, Brito L a., Hassett KJ, et al. Preclinical and Clinical Demonstration of Immunogenicity by mRNA Vaccines against H10N8 and H7N9 Influenza Viruses. Mol. Ther. **25**(6), 1316, 2017;

69. Richner JM, Himansu S, Dowd K a., Butler SL, Salazar V, Fox JM, et al. Modified mRNA Vaccines Protect against Zika Virus Infection. Cell. **169**(1), 176, 2017;

70. Kaczmarek JC, Kowalski PS, Anderson DG. Advances in the delivery of RNA therapeutics: from concept to clinical reality. Genome Med. **9**(1), 60, 2017;

71. Mullard A. mRNA-based drug approaches Phase I milestone. Nat. Rev. Drug Discov. **15**, 595, 2016;

72. Balmayor ER, Geiger JP, Aneja MK, Berezhanskyy T, Utzinger M, Mykhaylyk O, et al. Chemically modified RNA induces osteogenesis of stem cells and human tissue explants as well as accelerates bone healing in rats. Biomaterials. **87**, 131, 2016;

73. Balmayor ER, Geiger JP, Koch C, Aneja MK, van Griensven M, Rudolph C, et al. Modified mRNA for BMP-2 in Combination with Biomaterials Serves as a Transcript-Activated Matrix for Effectively Inducing Osteogenic Pathways in Stem Cells. Stem Cells Dev. **26**(1), 25, 2017;

74. Apgar JF, Tang JP, Singh P, Balasubramanian N, Burke J, Hodges MR, et al. Quantitative Systems Pharmacology Model of hUGT1A1-modRNA Encoding for the UGT1A1 Enzyme to Treat Crigler-Najjar Syndrome Type 1. CPT Pharmacometrics Syst. Pharmacol. **7**(6), 404, 2018;

75. Connolly B, Isaacs C, Cheng L, Asrani KH, Subramanian RR. SERPINA1 mRNA as a Treatment for Alpha-1 Antitrypsin Deficiency. J. Nucleic Acids. **2018**, 1, 2018;

76. Brandt G, Prieve M, Bell E, Blevins T, Galperin A, Harvie P, et al. Messenger RNA (mRNA) delivery to the liver corrects ornithine transcarbamylase deficiency in a mouse disease model. Mol. Genet. Metab. **120**(1), S31, 2017;

77. Kormann MSD, Hasenpusch G, Aneja MK, Nica G, Flemmer AW, Herber-Jonat S, et al. Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. Nat. Biotechnol. **29**(2), 154, 2011;

78. Karikó K, Buckstein M, Ni H, Weissman D. Suppression of RNA recognition by Toll-like receptors: The impact of nucleoside modification and the evolutionary origin of RNA. Immunity. **23**(2), 165, 2005;

79. Thess A, Grund S, Mui BL, Hope MJ, Baumhof P, Fotin-Mleczek M, et al. Sequence-engineered mRNA Without Chemical Nucleoside Modifications Enables an Effective Protein Therapy in

Large Animals. Mol. Ther. **23**(9), 1456, 2015;

80. Sahin U, Karikó K, Türeci Ö. MRNA-based therapeutics-developing a new class of drugs. Nat. Rev. Drug Discov. **13**(10), 759, 2014;

81. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The Sequence of the Human Genome. Science. **291**(5507), 1304, 2001;

82. Bashirullah A, Cooperstock RL, Lipshitz HD. Spatial and temporal control of RNA stability. Proc. Natl. Acad. Sci. U. S. A. **98**(13), 7025, 2001;

83. Jansen RP. mRNA localization: Message on the move. Nat. Rev. Mol. Cell Biol. **2**(4), 247, 2001;

84. Ferizi M, Aneja MK, Balmayor ER, Badieyan ZS, Mykhaylyk O, Rudolph C, et al. Human cellular CYBA UTR sequences increase mRNA translation without affecting the half-life of recombinant RNA transcripts. Sci. Rep. **6**(1), 39149, 2016;

85. Maitra U, Stringer EA, Chaudhuri A. Initiation Factors in Protein Biosynthesis. Annu. Rev. Biochem. **51**(1), 869, 1982;

86. Michel YM, Poncet D, Piron M, Kean KM, Borman AM. Cap-poly(A) synergy in mammalian cell-free extracts. Investigation of the requirements for poly(A)-mediated stimulation of translation initiation. J. Biol. Chem. **275**(41), 32268, 2000;

87. Kozak M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger rNAS. Nucleic Acids Res. **15**(20), 8125, 1987;

88. Pesole G, Gissi C, Grillo G, Licciulli F, Liuni S, Saccone C. Analysis of oligonucleotide AUG start codon context in eukariotic mRNAs. Gene. **261**(1), 85, 2000;

89. Haimov O, Sinvani H, Dikstein R. Cap-dependent, scanning-free translation initiation mechanisms. Biochim. Biophys. Acta - Gene Regul. Mech. **1849**(11), 1313, 2015;

90. Elfakess R, Sinvani H, Haimov O, Svitkin Y, Sonenberg N, Dikstein R. Unique translation initiation of mRNAs-containing TISU element. Nucleic Acids Res. **39**(17), 7598, 2011;

91. Guild BC, Derosa F HM. Lipid nanoparticle compositions and methods for mrna delivery: WO 2012/170930 A1. US: World Intellectualy Property Organization; 2012.

92. Bergman N, Moraes KCM, Anderson JR, Zaric B, Kambach C, Schneider RJ, et al. Lsm proteins bind and stabilize RNAs containing 5' poly(A) tracts. Nat. Struct. Mol. Biol. **14**(9), 824, 2007;

93. Wellensiek BP, Larsen AC, Flores J, Jacobs BL, Chaput JC. A leader sequence capable of enhancing RNA expression and protein synthesis in mammalian cells. Protein Sci. **22**(10),

1392, 2013;

94. Zumbo P. Ethanol Precipitation. Weill Cornell Med. Coll. 1, 1, 2013;

95. Badieyan ZS, Berezhanskyy T, Utzinger M, Aneja MK, Emrich D, Erben R, et al. Transcript-activated collagen matrix as sustained mRNA delivery system for bone regeneration. J. Control. Release. **239**, 137, 2016;

96. Konishi M, Kawamoto K, Izumikawa M, Kuriyama H, Yamashita T. Gene transfer into guinea pig cochlea using adeno-associated virus vectors. J. Gene Med. **10**(6), 610, 2008;

97. Aneja MK, Geiger J, Imker R, Uzgun S, Kormann M, Hasenpusch G, et al. Optimization of Streptomyces bacteriophage phi C31 integrase system to prevent post integrative gene silencing in pulmonary type II cells. Exp. Mol. Med. **41**(12), 919, 2009;

98. Huth S, Lausier J, Gersting SW, Rudolph C, Plank C, Welsch U, et al. Insights into the mechanism of magnetofection using PEI-based magnetofectins for gene transfer. J. Gene Med. **6**(8), 923, 2004;

99. Jarzebinska A, Pasewald T, Lambrecht J, Mykhaylyk O, K?mmerling L, Beck P, et al. A Single Methylene Group in Oligoalkylamine-Based Cationic Polymers and Lipids Promotes Enhanced mRNA Delivery. Angew. Chemie - Int. Ed. **55**(33), 9591, 2016;

100. Ikeda, R. A., & Richardason CC. Interactions of the RNA polymerase of the bacteriophage T7 with its promoter during binding and initiation of transcription. Proc. Natl. Acad. Sci. U. S. A. **83**(June), 3614, 1986;

101. Überla K, Stang A SP. T7 Promoter variants and methods of using the same field of the invention: WO 2014/013067. World Intellectualy Property Organization; 2014.

102. Imburgio D, Rong M, Ma K, McAllister WT. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. Biochemistry. **39**(34), 10419, 2000;

103. Újvári A, Martin CT. Identification of a minimal binding element within the T7 RNA polymerase promoter. J. Mol. Biol. **273**(4), 775, 1997;

104. Li T, Ho HH, Maslak M, Schick C, Martin CT. Major groove recognition elements in the middle of the T7 RNA polymerase promoter. Biochemistry. **35**(12), 3722, 1996;

105. Kozak M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. J. Mol. Biol. **196**(4), 947, 1987;

106. Elfakess R, Dikstein R. A translation initiation element specific to mRNAs with very short 5´UTR that also regulates transcription. PLoS One. **3**(8), 2008;

107. Raskin C, Diaz G, McAllister W. T7 RNA polymerase mutants with altered promoter specificities. Proc Natl Acad Sci USA. **90**(8), 3147, 1993;

108. Kozak M. An analysis of vertebrate mRNA sequences: Intimations of translational control. J. Cell Biol. **115**(4), 887, 1991;

109. Kozak M. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**(2), 857, 1984;

110. Kozak M. Emerging links between initiation of translation and human diseases. Mamm. Genome. **13**(8), 401, 2002;

111. Kozak M. Structural features in eukaryotic mRNAs that modulate the initiation of translation. J. Biol. Chem. **266**(30), 19867, 1991;

112. Sakai H, Washio T, Saito R, Shinagawa A, Itoh M, Shibata K, et al. Correlation between sequence conservation of the 5' untranslated region and codon usage bias in Mus musculus genes. Gene. **276**(1–2), 101, 2001;

113. Schrom E, Huber M, Aneja M, Dohmen C, Emrich D, Geiger J, et al. Translation of Angiotensin-Converting Enzyme 2 upon Liver- and Lung-Targeted Delivery of Optimized Chemically Modified mRNA. Mol. Ther. - Nucleic Acids. **7**(June), 350, 2017;

114. Ghassemifar R, Forster L, Qadah T, Finlayson J. Identification and characterization of two novel and differentially expressed isoforms of human α2-and α1-globin genes. Hemoglobin. **36**(5), 421, 2012.

115. Schleiss MR, Degnin CR, Geballe a P. Translational control of human cytomegalovirus gp48 expression. J Virol. **65**(12), 6782, 1991;

116. Karikó K, Muramatsu H, Welsh F a, Ludwig J, Kato H, Akira S, et al. Incorporation of Pseudouridine Into mRNA Yields Superior Nonimmunogenic Vector With Increased Translational Capacity and Biological Stability. Mol. Ther. **16**(11), 1833, 2008;

117. Wahle E, Rüegsegger U. 3'-End processing of pre-mRNA in eukaryotes. FEMS Microbiol. Rev. **23**(3), 277, 1999;

118. Mangus DA, Evans MC, Jacobson A. Poly(A)-binding proteins: multifunctional scaffolds for the post- transcriptional control of gene expression. Genome Biol. **4**(7), 223, 2003;

119. Hershey JWB, Sonenberg N, Mathews MB. Principles of translational control: An overview. Cold Spring Harb. Perspect. Biol. **4**(12), 1, 2012;

120. Goldstrohm AC, Wickens M. Multifunctional deadenylase complexes diversify mRNA control. Nat. Rev. Mol. Cell Biol. **9**(4), 337, 2008;

121. Standart N, Weil D. P-Bodies: Cytosolic Droplets for Coordinated mRNA Storage. Trends Genet. **34**(8), 612, 2018;

122. An emerging interest in RNA world [Internet]. MBL Life Sci. [cited 2018 Oct 8]. Available from: http://ruo.mbl.co.jp/bio/g/product/epigenetics/RNAworld.html

123. Cao G-J, Sarkar N. Identification of the gene for an Escherichia coli poly(A) polymerase. Biochem. Commun. by Charles C. Richardson J. Bacteriol. **89**(171), 10380, 1992;

124. Martin G, Keller W. Tailing and 3'-end labeling of RNA with yeast poly(A) polymerase and various nucleotides. RNA. **4**, 226, 1998;

125. Voet, D, Voet J. Biochemistry. 4th ed. New York: J. Wiley & Sons; 2011.

126. Holtkamp S, Kreiter S, Selmi A, Simon P, Koslowski M, Huber C. Modification of antigen encoding RNA increases stability , translational efficacy and T-cell stimulatory capacity of dendritic cells Modification of antigen encoding RNA increases stability , translational efficacy and T-cell stimulatory capacity of dendr. Blood. **108**(13), 4009, 2006;

127. Grier AE, Burleigh S, Sahni J, Clough CA, Cardot V, Choe DC, et al. pEVL: A Linear Plasmid for Generating mRNA IVT Templates With Extended Encoded Poly(A) Sequences. Mol. Ther. Nucleic Acids. **5**(April), 306, 2016;

128. Preiss T, Muckenthaler M, Hentze MW. Poly (A)-tail-promoted translation in yeast : implications for translational control. RNA. **4**(11), 1321, 1998;

129. Baer B, Kornberg R. Repeating structure of cytoplasmic poly(A)-ribonucleoprotein. Proc Natl Acad Sci USA. **77**(4), 1890, 1980;

130. Baer BW, Kornberg RD. The protein responsible for the repeating structure of cytoplasmic poly(A)-ribonucleoprotein. J. Cell Biol. **96**(3), 717, 1983;

131. Wang Z, Day N, Trifillis P, Kiledjian M. An mRNA stability complex functions with poly(A)-binding protein to stabilize mRNA in vitro. Mol. Cell. Biol. **19**(7), 4552, 1999;

132. Bangel-Ruland N, Tomczak K, Fernández Fernández E, Leier G, Leciejewski B, Rudolph C, et al. Cystic fibrosis transmembrane conductance regulator-mRNA delivery: a novel alternative for

cystic fibrosis gene therapy. J. Gene Med. **15**(11–12), 414, 2013;

133. Vallazza B, Petri S, Poleganov MA, Eberle F, Kuhn AN, Sahin U. Recombinant messenger RNA technology and its application in cancer immunotherapy, transcript replacement therapies, pluripotent stem cell induction, and beyond. Wiley Interdiscip. Rev. RNA. **6**(5), 471, 2015;

134. Godiska R, Mead D, Dhodda V, Wu C, Hochstein R, Karsi A, et al. Linear plasmid vector for cloning of repetitive or unstable sequences in Escherichia coli. Nucleic Acids Res. **38**(6), 1, 2009;

135. Kühn U, Wahle E. Structure and function of poly(A) binding proteins. Biochim. Biophys. Acta - Gene Struct. Expr. **1678**(2–3), 67, 2004;

136. Trepotec Z, Aneja M, Geiger J, Hasenpusch G, Plank C, Rudolph C. Maximizing the translational yield of mRNA therapeutics by minimizing 5'-UTRs. Tissue Eng. Part A. **25**(1-2), 2019;

137. Williams AS, Marzluff WF. The sequence of the stem and flanking sequences at the 3' end of histone mRNA are critical determinants for the binding of the stem-loop binding protein. Nucleic Acids Res. **23**(4), 654, 1995;

138. Zanier K, Luyten I, Crombie C, Muller B, Schümperli D, Linge JP, et al. Structure of the histone mRNA hairpin required for cell cycle regulation of histone gene expression. RNA. **8**(1), 29, 2002;

8 ASSOCIATED PUBLICATIONS & PATENT

Publications:

1. Trepotec Z., Aneja M., Geiger J., Hasenpusch G., Plank C., Rudolph C. (2018): Maximizing the translational yield of mRNA therapeutics by minimizing 5'-UTRs. Tissue Engineering Part A, accepted 11.04.2018

2. Trepotec Z., Geiger J., Plank C., Aneja M., Rudolph C. (2018): Segmented poly(A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life. RNA, accepted 22.12.2018

Patent:

Trepotec Z., Aneja M., Plank C., Rudolph C., Weiss L. (2018): Plasmid containing a sequence encoding an mRNA with a segmented poly(A) tail. Anmeldetag: 11. Oct 2018, amtl. Aktenzeichen: 18 19 9857.6

9 APPENDIX

Table 15. 5'-UTR sequences used in this study

5'-UTR	*5'-UTR sequence*
UTR1	GGGAGAGCCACC
UTR2	GGGAGACGCCACC
UTR3	GGGAGACTGCCACC
UTR4	GGGAGACAGCCACC
UTR5	GGGAGACCGCCACC
UTR6	GGGAGACGGCCACC
UTR7	GGGAGACTGCCAAG
refUTR1	GGGAGACTCTTCTGGTCCCCACAGACTCAGAGAGAAC
refUTR2	GGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGCGGCCGGGAACGGTGCATTGGAACGCGGATTCCCCGTGCCAAGAGTGACTCACCGTCCTTGACACG
refUTR3	GGGAGACAAAAAAAAAAAAAAAAAAAAA
refUTR4	GGGAGACCATATTGAAGAGACAGAGTGATATATAAAACTGCTAA
refUTR5	GGGAGACAAAACTGCTAA

Table 16. Poly(A) sequences used in this study

Poly(A)	*Poly(A) sequence*
2x60_6	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
3x40_6	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGCATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGATATCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
A120	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
ACH[79]	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATGCATCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
2x60_C	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

2x60_G	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
2x60_T	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
2x60_12nt	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACGCCATTAGAGGAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
2x60_24nt	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAACAACCCCTGATTGT GTCCGCATCTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Table 17. GOI sequences used in this study

GOI	*Sequence*
Luc2	ATGGAAGATGCCAAAAACATTAAGAAGGGCCCAGCGCCATTCTACCCACTCGAAGACGGGACCGCCGGCGAGCAG CTGCACAAAGCCATGAAGCGCTACGCCCTGGTGCCCGGCACCATCGCCTTTACCGACGCACATATCGAGGTGGACA TTACCTACGCCGAGTACTTCGAGATGAGCGTTCGGCTGGCAGAAGCTATGAAGCGCTATGGGCTGAATACAAACCA TCGGATCGTGGTGTGCAGCGAGAATAGCTTGCAGTTCTTCATGCCCGTGTTGGGTGCCCTGTTCATCGGTGTGGCT GTGGCCCCAGCTAACGACATCTACAACGAGCGCGAGCTGCTGAACAGCATGGGCATCAGCCAGCCCACCGTCGTAT TCGTGAGCAAGAAAGGGCTGCAAAAGATCCTCAACGTGCAAAAGAAGCTACCGATCATACAAAAGATCATCATCAT GGATAGCAAGACCGACTACCAGGGCTTCCAAAGCATGTACACCTTCGTGACTTCCCATTTGCCACCCGGCTTCAACG AGTACGACTTCGTGCCCGAGAGCTTCGACCGGGACAAAACCATCGCCCTGATCATGAACAGTAGTGGCAGTACCGG ATTGCCCAAGGGCGTAGCCCTACCGCACCGCACCGCTTGTGTCCGATTCAGTCATGCCCGCGACCCCATCTTCGGCA ACCAGATCATCCCCGACACCGCTATCCTCAGCGTGGTGCCATTTCACCACGGCTTCGGCATGTTCACCACGCTGGGC TACTTGATCTGCGGCTTTCGGGTCGTGCTCATGTACCGCTTCGAGGAGGAGCTATTCTTGCGCAGCTTGCAAGACTA TAAGATTCAATCTGCCCTGCTGGTGCCCACACTATTTAGCTTCTTCGCTAAGAGCACTCTCATCGACAAGTACGACCT AAGCAACTTGCACGAGATCGCCAGCGGCGGGGCGCCGCTCAGCAAGGAGGTAGGTGAGGCCGTGGCCAAACGCT TCCACCTACCAGGCATCCGCCAGGGCTACGGCCTGACAGAAACAACCAGCGCCATTCTGATCACCCCCGAAGGGGA CGACAAGCCTGGCGCAGTAGGCAAGGTGGTGCCCTTCTTCGAGGCTAAGGTGGTGGACTTGGACACCGGTAAGAC ACTGGGTGTGAACCAGCGCGGCGAGCTGTGCGTCCGTGGCCCCATGATCATGAGCGGCTACGTTAACAACCCCGA GGCTACAAACGCTCTCATCGACAAGGACGGCTGGCTGCACAGCGGCGACATCGCCTACTGGGACGAGGACGAGCA CTTCTTCATCGTGGACCGGCTGAAGAGCCTGATCAAATACAAGGGCTACCAGGTAGCCCCAGCCGAACTGGAGAGC ATCCTGCTGCAACACCCCAACATCTTCGACGCCGGGGTCGCCGGCCTGCCCGACGACGATGCCGGCGAGCTGCCCG CCGCAGTCGTCGTGCTGGAACACGGTAAAACCATGACCGAGAAGGAGATCGTGGACTATGTGGCCAGCCAGGTTA CAACCGCCAAGAAGCTGCGCGGTGGTGTTGTGTTCGTGGACGAGGTGCCTAAAGGACTGACCGGCAAGTTGGACG CCCGCAAGATCCGCGAGATTCTCATTAAGGCCAAGAAGGGCGGCAAGATCGCCGTGTAA
hEPO	ATGGGCGTGCACGAATGTCCTGCTTGGCTGTGGCTGCTGCTGAGCCTGCTGTCTCTGCCTCTGGGACTGCCTGTGCT GGGAGCCCCTCCTAGACTGATCTGCGACAGCCGGGTGCTGGAAAGATACCTGCTGGAAGCCAAAGAGGCCGAGAA CATCACCACCGGCTGCGCCGAGCACTGCAGCCTGAACGAGAATATCACCGTGCCCGACACCAAAGTGAACTTCTAC GCCTGGAAGCGGATGGAAGTGGGCCAGCAGGCTGTGGAAGTGTGGCAGGGACTGGCCCTGCTGAGCGAAGCTGT GCTGAGAGGACAGGCTCTGCTCGTGAACAGCAGCCAGCCTTGGGAGCCTCTGCAGCTGCACGTGGACAAGGCCGT GTCTGGCCTGAGAAGCCTGACCACACTGCTGAGAGCCCTGGGGGCCCAGAAAGAGGCCATCTCTCCACCTGATGCC GCCTCTGCCGCCCCTCTGAGAACCATCACCGCCGACACCTTCAGAAAGCTGTTCCGGGTGTACAGCAACTTCCTGCG GGGCAAGCTGAAGCTGTACACAGGCGAGGCCTGCCGGACCGGCGATAGATAA
d2EGFP	ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGG CCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACC ACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACC CCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC AAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTC TATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGC GTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACC TGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGC CGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGAAGCTTAGCCATGGCTTCCCGCCGGAGGTGGAGGAGCA GGATGATGGCACGCTGCCCATGTCTTGTGCCCAGGAGAGCGGGATGGACCGTCACCCTGCAGCCTGTGCTTCTGCT AGGATCAATGTGTAG

CFTR

ATGCAGAGAAGCCCCCTGGAAAAGGCCAGCGTGGTGTCCAAGCTGTTCTTCAGCTGGACCCGGCCCATCCTGCGGA
AGGGCTACAGACAGAGACTGGAACTGAGCGACATCTACCAGATCCCCAGCGTGGACAGCGCCGACAACCTGAGCG
AGAAGCTGGAAAGAGAGTGGGACAGAGAGCTGGCCTCCAAGAAGAACCCCAAGCTGATCAACGCCCTGCGGCGG
TGCTTCTTCTGGCGGTTTATGTTCTACGGCATCTTCCTGTACCTGGGCGAAGTGACCAAGGCCGTGCAGCCTCTGCT
GCTGGGCAGAATCATTGCCAGCTACGACCCCGACAACAAAGAGGAACGGTCTATCGCCATCTACCTGGGCATCGGC
CTGTGCCTGCTGTTCATCGTGCGGACCCTGCTGCTGCACCCTGCCATCTTTGGCCTGCACCACATCGGCATGCAGAT
GCGGATCGCCATGTTCAGCCTGATCTACAAGAAAACCCTGAAGCTGAGCAGCCGGGTGCTGGACAAGATCAGCATC
GGACAGCTGGTGTCCCTGCTGAGCAACAACCTGAACAAGTTCGACGAGGGCCTGGCCCTGGCCCACTTCGTGTGGA
TTGCACCACTGCAGGTGGCACTGCTGATGGGCCTGATCTGGGAACTGCTGCAGGCCAGCGCCTTTTGCGGCCTGGG
CTTTCTGATTGTGCTGGCCCTGTTCCAGGCCGGACTGGGCCGGATGATGATGAAGTACCGGGACCAGAGAGCCGG
CAAGATCTCCGAGCGGCTCGTGATCACCAGCGAGATGATCGAGAACATCCAGAGCGTGAAGGCCTACTGCTGGGA
AGAGGCCATGGAAAAGATGATTGAGAATCTGCGGCAGACCGAGCTGAAACTGACCCGGAAGGCCGCCTACGTGCG
CTACTTCAACAGCAGCGCCTTCTTCTTCTCCGGCTTCTTCGTGGTGTTCCTGAGCGTGCTGCCCTACGCCCTGATCAA
GGGCATCATCCTGAGAAAGATTTTCACCACCATTTCTTTCTGCATCGTGCTGCGGATGGCCGTGACCAGACAGTTCC
CTTGGGCTGTGCAGACTTGGTACGACAGCCTGGGCGCCATCAACAAGATCCAGGACTTCCTGCAGAAGCAGGAGTA
CAAGACCCTGGAGTACAACCTGACCACCACCGAGGTCGTGATGGAAAACGTGACCGCCTTCTGGGAGGAAGGCTT
CGGCGAGCTGTTCGAGAAGGCCAAGCAGAACAACAACAACCGCAAGACCAGCAACGGCGACGACTCCCTGTTCTTC
TCCAACTTCTCCCTGCTGGGGACCCCCGTGCTGAAGGACATCAACTTCAAGATCGAGCGGGGACAGCTGCTGGCCG
TGGCTGGATCTACAGGCGCCGGAAAGACCAGCCTGCTGATGGTCATCATGGGCGAGCTGGAACCCAGCGAGGGCA
AGATCAAGCACAGCGGCCGGATCAGCTTCTGTAGCCAGTTCTCCTGGATCATGCCCGGCACCATCAAAGAGAACAT
CATCTTCGGCGTGTCCTACGACGAGTACAGATACCGCAGCGTGATCAAGGCCTGCCAGCTGGAAGAGGACATCAGC
AAGTTCGCCGAGAAGGACAATATCGTGCTGGGCGAGGGCGGCATCACACTGTCTGGCGGCCAGAGGGCCAGAATC
AGCCTGGCCAGAGCCGTGTACAAGGACGCCGATCTGTACCTGCTGGACAGCCCCTTCGGCTACCTGGACGTGCTGA
CCGAGAAAGAGATCTTCGAGAGCTGCGTGTGCAAGCTGATGGCCAACAAGACCCGGATCCTCGTGACCAGCAAGA
TGGAACACCTGAAGAAGGCCGACAAGATCCTGATCCTGCACGAGGGCAGCAGCTACTTTTACGGCACCTTCAGCGA
GCTGCAGAACCTGCAGCCCGACTTCAGCAGCAAACTGATGGGCTGCGACAGCTTCGACCAGTTCAGCGCCGAGCG
GCGGAACAGCATCCTGACCGAAACCCTGCACCGGTTCTCCCTGGAAGGCGACGCTCCTGTGTCCTGGACCGAGACA
AAGAAGCAGAGCTTCAAGCAGACCGGCGAGTTTGGCGAGAAGAGAAAGAACTCCATCCTGAACCCCATCAATAGC
ATCCGGAAGTTCAGCATCGTGCAGAAAACCCCCCTGCAGATGAACGGCATCGAAGAGGACTCCGACGAGCCCCTG
GAACGCAGACTGAGCCTGGTGCCTGATAGCGAGCAGGGCGAGGCCATCCTGCCCAGAATCTCCGTGATCAGCACC
GGCCCTACCCTGCAGGCTCGGAGAAGGCAGTCTGTGCTGAACCTGATGACCCACAGCGTGAACCAGGGACAGAAT
ATCCACAGAAAGACCACCGCCAGCACCAGAAAAGTGTCACTGGCCCCCCAGGCCAACCTGACAGAGCTGGACATCT
ACAGCAGACGGCTGAGCCAGGAAACAGGCCTGGAAATCAGCGAGGAAATCAACGAAGAGGATCTGAAAGAGTGC
TTTTTCGACGACATGGAATCCATCCCCGCCGTGACAACCTGGAATACCTACCTGCGGTACATCACCGTGCACAAGTC
CCTGATCTTCGTGCTGATCTGGTGTCTCGTGATCTTCCTGGCCGAGGTGGCCGCTTCTCTGGTGGTGCTGTGGCTGC
TGGGAAACACCCCTCTGCAGGACAAGGGCAACAGCACCCACAGCCGGAACAACAGCTATGCCGTGATCATCACAAG
CACCAGCTCCTACTACGTGTTCTACATCTACGTGGGCGTGGCCGACACCCTGCTGGCCATGGGCTTTTTCAGAGGCC
TGCCCCTGGTGCACACCCTGATCACCGTGTCCAAGATTCTGCACCATAAGATGCTGCACAGCGTGCTGCAGGCTCCC
ATGAGCACCCTGAATACCCTGAAGGCTGGCGGCATCCTGAATCGGTTCAGCAAGGACATTGCCATCCTGGACGACC
TGCTGCCTCTGACCATCTTCGACTTCATCCAGCTGCTGCTGATCGTGATCGGCGCCATTGCCGTGGTGGCTGTGCTG
CAGCCCTATATCTTCGTGGCCACCGTGCCCGTGATCGTGGCCTTCATTATGCTGCGGGCCTACTTTCTGCAGACCTCT
CAGCAGCTGAAACAGCTGGAATCCGAGGGCAGAAGCCCCATCTTCACCCACCTCGTGACAAGCCTGAAGGGCCTGT
GGACCCTGAGAGCCTTCGGCAGACAGCCCTACTTCGAGACACTGTTCCACAAGGCCCTGAACCTGCACACCGCCAA
CTGGTTTCTGTATCTGTCCACCCTGCGGTGGTTCCAGATGAGGATCGAGATGATTTTTGTGATCTTCTTTATCGCCGT
GACCTTCATCAGCATTCTGACCACCGGCGAGGGGGAGGGCAGAGTGGGCATTATTCTGACACTGGCCATGAACATC
ATGAGCACTCTGCAGTGGGCCGTGAACAGCTCCATCGACGTGGACTCCCTGATGCGGAGCGTGTCCCGGGTGTTCA
AGTTCATCGACATGCCCACAGAGGGCAAGCCCACCAAGAGCACCAAGCCCTACAAGAACGGCCAGCTGAGCAAAG
TGATGATTATCGAGAACTCCCACGTGAAGAAGGATGACATCTGGCCCAGCGGCGGACAGATGACCGTGAAGGATC
TGACCGCCAAGTACACCGAGGGCGGAAACGCCATTCTGGAAAACATCAGCTTTAGCATCTCTCCCGGCCAGCGCGT
GGGCCTGCTGGGACGGACAGGATCTGGCAAGTCTACCCTGCTGTCCGCCTTCCTGCGGCTGCTGAATACCGAAGGC
GAGATCCAGATCGACGGGGTGTCCTGGGACAGCATTACACTGCAGCAGTGGCGCAAGGCCTTTGGCGTGATCCCC
CAGAAGGTGTTCATTTTCAGCGGCACCTTTCGGAAGAACCTGGACCCCTACGAGCAGTGGAGCGACCAGGAAATCT
GGAAGGTGGCCGATGAAGTGGGACTGAGAAGCGTGATCGAGCAGTTTCCCGGCAAGCTGGATTTCGTGCTGGTGG
ACGGCGGCTGCGTGCTGTCTCACGGACACAAGCAGCTGATGTGCCTGGCTAGATCCGTGCTGTCCAAGGCCAAGAT
CCTGCTGCTGGACGAGCCTAGCGCCCACCTGGATCCCGTGACATACCAGATCATCAGACGGACACTGAAGCAGGCC
TTCGCCGATTGCACCGTGATCCTGTGCGAGCACCGGATCGAGGCCATGCTGGAATGCCAGCAGTTTCTCGTGATTG
AAGAGAACAAAGTGCGGCAGTACGACTCCATTCAGAAGCTGCTGAACGAGAGAAGCCTGTTCCGGCAGGCCATCTC
CCCCAGCGACAGAGTGAAGCTGTTCCCCCACCGGAACTCCAGCAAGTGCAAGTCCAAGCCCCAGATCGCCGCCCTG
AAAGAAGAAACCGAGGAAGAGGTGCAGGATACCCGGCTGTGA

Table 18. Primers used for qPCR

Target gene	*# UPL (Roche Diagnostics)*	*Sequence (forward)*	*Sequence (reverse)*
Luc2	29	5’-acgccgagtacttcgagatg-3’	5’-attcagcccatagcgcttc-3’
d2EGFP	37	5’-cctgaagttcatctgcacca-3’	5’-ctcgtgaccaccctgacc-3’

10 ACKNOWLEDGEMENTS

I would like to pay special thankfulness, warmth, and appreciation to the persons below who made my research successful and assisted me at every point to cherish my goal.

My greatest appreciation goes to Prof. Carsten Rudolph and Prof. Christian Plank, for the given opportunity to become a part of the Ethris family. With their professional support and immense knowledge, I was inspired to conduct my research and present it here as a doctoral dissertation.

I would like to extend my special thanks to Dr. Manish Aneja, who has supervised me all the way through with his bright mind, great ideas and limitless motivation. Besides, I could have not imagined having a better advisor and mentor, who has taught me to always look ahead and has provided his support in numerous ways. For all of these, forever I shall remain indebted to him and I am truly thankful for that.

Assistance and introduction to Ethris's "RNA world", as unique as the RNA world indeed is, provided by Dr. Johannes Geiger, and is greatly appreciated.

I am also thankful to all other members of the Ethris family, who provided a pleasurable working atmosphere and occasional meetups. It was a pleasure to work with you!

My closest family, mom, dad, and sister have once again proven their role with all their sacrifices made on my behalf, encouraging my education and explorative mind from day one. With their support and understandings, I became who I am today. I am also thankful to all other family members who not only assisted me financially but also extended their support morally and

emotionally. My greatest luck is to have so many good friends, whom I, although scattered all over the globe, always felt close by. For being my day-to-day support during this period, and for bringing so much happiness and cherishment to my life, a heartfelt thanks to Alexey.