

Engineering of modified linear and circular messenger RNA and study of the influence on protein expression tuning

Dissertation zur Erlangung des Grades

“Doktor der Naturwissenschaften”

am Fachbereich Chemie, Pharmazie, Geographie und Geowissenschaften
der Johannes Gutenberg-Universität Mainz

vorgelegt von

Kaouthar Slama

geboren am 12.01.1991 in Paris

Mainz, January 2021



Dekan:

1. Berichterstatter:
2. Berichterstatter:

Datum der mündlichen Prüfung:

Prüfungskommission der mündlichen

1. Prüfer
2. Prüfer
3. Prüfer

D77 (Dissertation Mainz)

The present dissertation was composed at the Institute of Pharmaceutical and Biomedical Sciences at the Johannes Gutenberg-University of Mainz.

Faculty of Chemistry, Pharmaceutical Sciences,
Geography and Geosciences Institute
of Pharmaceutical and Biomedical Sciences
Johannes Gutenberg-University Mainz

I hereby declare that I wrote the dissertation submitted without any unauthorized external assistance and used only sources acknowledged in the work. All textual passages which are appropriated verbatim or paraphrased from published and unpublished texts as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. In carrying out this research, I complied with the rules of standard scientific practice as formulated in the statutes of the Johannes Gutenberg-University Mainz to insure standard scientific practice.

Place, Date

Kaouthar Slama

“Everybody is a genius. But if you judge a fish by its ability to climb a tree, it will live its whole life believing that it is stupid.” Albert Einstein.

Acknowledgment

Contents

Abstract	xi
Zusammenfassung	xiii
List of figures	xiv
List of tables	xvi
Abbreviations.....	xvii
1 Introduction.....	1
1.1 Messenger RNA	1
1.1.1 mRNA biogenesis.....	2
1.1.2 The pre-mRNA processing and functional activity	3
1.2 Natural modifications in RNA.....	9
1.2.1 Modifications in eukaryotic mRNA	11
1.2.2 Modification detection <i>via</i> chromatography	18
1.3 Synthetic mRNA design and applications	19
1.3.1 Basic structural elements	19
1.3.2 Modified nucleosides insertion into synthetic mRNA	21
1.3.3 Applications of modified mRNA	22
1.4 Circular RNA.....	23
1.4.1 Biogenesis of circular RNA.....	23
1.4.2 Detection and characterization of circular RNA	25
1.4.3 Biological implications of circular RNA	26
1.4.4 Synthetic circular RNA	27
1.4.5 Translation of circular RNA.....	28
2 Motivation and objectives	31
3 Results and discussion	33
3.1 Engineering of nucleoside-modified mRNA.....	33
3.1.1 Generation of randomly-modified mRNA	33
3.1.2 Design and engineering of large point-modified mRNA	39
3.2 Establishment of mRNA functionality analysis <i>via</i> an in-gel fluorescence scan.....	50
3.3 Translation-tuning by modified nucleosides	52
3.3.1 Influence of single point-modified nucleosides on mRNA translation.....	53
3.3.2 Influence of random-modification incorporation on mRNA translation	58
3.4 Circular RNA.....	68

3.4.1	Design and engineering of circular mRNA	68
3.4.2	Design and synthesis of circular RNA	69
3.4.3	Circularization of modified mRNA precursor	70
3.4.4	Circularity validation	71
3.4.5	Protein synthesis from circular RNA	76
3.4.6	Translation of HPLC-purified circRNA	79
3.4.7	Translation of modified-circRNA	80
3.5	Application of randomly-modified mRNA: Determination of enrichment factors for modified RNA in MeRIP experiments	82
3.5.1	Method design and analysis	83
3.5.2	Determination of the enrichment factor of modified RNA	85
4	Conclusion and perspectives	90
4.1	Translation-tuning by modified mRNA	90
4.2	Circular RNA	93
5	Materials and methods	96
5.1	Materials	96
5.1.1	Instruments	96
5.2	Chemicals & Consumables	97
5.3	Disposables	99
5.4	Enzymes, reagents and kits	100
5.5	Buffers and solutions	100
5.6	Softwares	102
5.7	Oligonucleotides (all nucleotides were supplied from IBA (Göttingen))	102
5.8	Molecular biology methods and techniques	107
5.8.1	From DNA to RNA	107
5.8.2	Purification & concentration methods for oligonucleotides	109
5.8.3	Design and synthesis of ligated mRNA and analysis	110
5.8.4	PAGE electrophoresis and circular RNA purification	111
5.8.5	Liquid chromatography (LC)	113
5.8.6	Liquid chromatography-Tandem Mass spectrometry (LC-MS/MS)	114
5.8.7	<i>In vitro</i> translation and protein analytics	115
5.8.8	Immunoprecipitation and samples preparation (for section 3.5)	116
6	Appendix	119
	References	129
	Publications and Collaborations	155

Oral Presentations	155
Curriculum Vitae	156

Abstract

RNA modifications created the epitranscriptome that shaped our understanding of the regulation of biological processes. Modifications were mapped in mRNAs and little is known about their role in modulating their functionality.

Here, selected modifications which are naturally occurring in RNA were inserted in a targeted manner in an enhanced green fluorescence protein (EGFP) mRNA, called point-modified, to assess their influence on protein expression. In analogy to a piece of a puzzle, the mRNA design included sequence breakdown into three fragments, where the middle fragment harbours the modification at the desired site. Subsequently, the full point-modified mRNA was reconstituted by a 3-way-one-pot ligation, and then purified by real-time gel elution. Such a ligation technique, previously applied for shorter size length, provided the possibility for application on much longer mRNA. Essential sites for translation initiation (the Kozak motif and the start codon), which were not yet investigated, as well as internal codons from the coding sequence were modified. The obtained point-modified mRNA was validated by LC-MS/MS and RiboMeth-seq. The presence of 2'-O-ribose-methylated or pseudouridine in point-modified mRNA, revealed a codon-position-dependent effect, stimulating or hampering the overall protein production. Findings obtained here in the context of IRES-driven translation expanded prior work on the influence of modifications on the canonical-translation system.

In a comparative study, the random incorporation of modifications in mRNA was also explored under different conditions, by adding or omitting structural elements (namely IRES, cap and poly(A)). Considering the remarkable structure and differences between circular and linear RNA, enzymatic-based strategy was used for RNA circularization. The obtained product was *via* RP-HPLC. Protein expression levels of circular RNA relative to the linear RNA form depended on (i) the purification method and (ii) the used translation system. In the last project, randomly modified mRNA was used to outline the limited antibody specificity upon enrichment with immunoprecipitation methods, otherwise required for a significant analysis at the transcriptome-wide level.

Zusammenfassung

RNA-Modifikationen schufen das Epitranskriptom, das unser Verständnis der Regulation biologischer Prozesse prägte. Modifikationen wurden in mRNAs kartiert und es ist wenig über ihre Rolle bei der Modulation ihrer Funktionalität bekannt.

Hier wurden ausgewählte Modifikationen, die natürlicherweise in der RNA vorkommen, gezielt in eine Enhanced Green Fluorescence Protein (EGFP) mRNA eingebaut, als punktmodifiziert bezeichnet, um ihren Einfluss auf die Proteinexpression zu untersuchen. In Analogie zu einem Puzzle wurde die mRNA in drei Fragmente (bzw. Puzzleteile) unterteilt, wobei das mittlere Fragment die Modifikation an der gewünschten Stelle beherbergt. Anschließend wurde die vollständige, punktmodifizierte mRNA durch eine sogenannte „3-Way-One-Pot“ Ligation rekonstituiert und anschließend durch „Real-Time“-Gel-Elution aufgereinigt. Diese Ligationsmethode, die zuvor für kürzere Sequenzlängen angewandt wurde, bot die Möglichkeit der Anwendung für wesentlich längere mRNA. Essentielle Stellen für die Translationsinitiation (das Kozak-Motiv und das Startcodon), die bisher noch nicht untersucht wurden, sowie interne Codons aus der kodierenden Sequenz wurden modifiziert. Die erhaltene punktmodifizierte mRNA wurde mittels LC-MS/MS und RiboMeth-seq validiert. Das Vorhandensein von 2'-O-Ribose-methylierter oder mit Pseudouridin punktmodifizierter mRNA zeigte einen von der Codon-Position abhängigen Effekt, der die gesamte Proteinproduktion stimuliert oder beeinträchtigt. Die hier im Kontext der IRES-gesteuerten Translation gewonnenen Erkenntnisse erweiterten frühere Arbeiten zum Einfluss von Modifikationen auf das kanonische Translationssystem.

In einer vergleichenden Studie wurde auch der zufällige Einbau von Modifikationen in mRNA unter verschiedenen Bedingungen durch Hinzufügen oder Weglassen von Strukturelementen (nämlich IRES, cap und poly(A)) untersucht. In Anbetracht der bemerkenswerten Struktur und der Unterschiede zwischen zirkulärer und linearer RNA wurde eine auf Enzymen basierende Strategie zur RNA-Zirkularisierung verwendet. Das erhaltene Produkt wurde über RP-HPLC aufgereinigt. Das Proteinexpressionsniveau der zirkulären RNA im Vergleich zur linearen RNA-Form war abhängig von (i) der Aufreinigungsmethode und (ii) dem verwendeten Translationssystem. Im letzten Projekt wurde zufällig modifizierte mRNA verwendet, um die limitierte Antikörperspezifität bei der Anreicherung mit Immunpräzipitationsmethoden aufzuzeigen, die für signifikante Analysen auf transkriptomweiter Ebene erforderlich ist.

List of figures

Figure 1. 1: Structure of Nucleic Acids.	1
Figure 1. 2: The step-wise assembly of pre-initiation complex assembly in eukaryotes.	3
Figure 1. 3: The eukaryotic mRNA 5'cap structure.....	5
Figure 1. 4: Models of translation initiation.	7
Figure 1. 5: Schematic diagram of the EMCV IRES secondary structure main features.....	9
Figure 1. 6: Examples of RNA modifications.....	11
Figure 1. 7: Overview of the distribution of chemical modifications in eukaryotic mRNA.	12
Figure 1. 8: Translation initiation driven by m ⁶ A.	15
Figure 1. 9: Structures of cap analog.....	20
Figure 1. 10: Overview of circular RNA biogenesis.	25
Figure 1. 11: Translation of circular RNA.....	30
 Figure 2.1: Global overview of the present work and possible perspectives.....	 32
 Figure 3. 1: Synthesis and analysis of modified IRES-EGFP mRNA.....	 35
Figure 3. 2: Engineering and synthesis of modified, capped and polyadenylated mRNA.....	38
Figure 3. 3: The puzzle concept.....	39
Figure 3. 4: 3-way-one-pot synthesis of large mRNA harbouring single-point modification.....	43
Figure 3. 5: Modification detection in ligated RNA.	44
Figure 3. 6: Purification of linear ligated mRNA on RP-HPLC.	47
Figure 3. 7 : Purification of RNA ligation by 3-way-one-pot ligation.....	48
Figure 3. 8: Gel particle characterization after RNA real-time gel elution.	49
Figure 3. 9: Monitoring protein expression by in-gel fluorescence detection.	52
Figure 3. 10: Translation of ligated point-modified IRES-EGFP mRNA.	54
Figure 3. 11: Kinetics of the EGFP protein production from point-modified mRNA.	55
Figure 3. 12: Translation inhibition effect's rescue.....	56
Figure 3. 13: Translation efficiency of randomly modified IRES-EGFP mRNA.	61
Figure 3. 14: Translation of pseudouridine modified EGFP mRNA.	64
Figure 3. 15: Translation of m ⁶ A modified EGFP mRNA.	66
Figure 3. 16: m ⁶ A modified-RNA requires translation initiation factor eIF4E.....	68
Figure 3. 17: In vitro synthesis of circular RNAs and purification.....	70
Figure 3. 18: Circularity verification by different biochemical assays.....	72
Figure 3. 19: Identification of RNA circularity via HPLC.	75
Figure 3. 20 : Translation of circular RNA in rabbit reticulocyte lysate (RRL).	77
Figure 3. 21: Translation of circular RNA purified via HPLC or by gel-elution.	80
Figure 3. 22: Translation of m ⁶ A-modified circRNA.	81
Figure 3. 23: Size comparison (Å) between potential binding surface of an IgG molecule and m ⁶ A residue.....	83
Figure 3. 24: TLC analysis and enrichment factor determination.	85
Figure 3. 25: Quantification of enrichment factor after MeRIP.....	86
Figure 3. 26: Determination of enrichment factor for optimized IP.....	88
 Figure 6. 1: Cap mRNA analysis.....	 119
Figure 6. 2: Analysis of polyadenylated EGFP mRNA.....	119

Figure 6. 3: Analysis of 2'-O-Me point-modified mRNA.....	120
Figure 6. 4: Detection of unexpcted modifications in m⁶A-point modified mRNA.....	121
Figure 6. 5: Generation of 5' IRES-lig block and internal-lig block.....	122
Figure 6. 6: HPLC purification of 5'IRES-lig block.....	123
Figure 6. 7: Purification of ligated RNA.	124
Figure 6. 8: Set-up tool for real-time gel elution.....	124
Figure 6. 9: Optimization of RNA concentration for translation in RRL system.	125
Figure 6. 10: Circularization reaction optimization.....	125
Figure 6. 11: Analysis of circular RNA migration on increasing cross-linking PAGE.....	126
Figure 6. 12: RNA treatment with RNase R.	126
Figure 6. 13: Translation of HPLC-purified circRNA.....	127
Figure 6. 14: Schematic overview of the m⁶A immunoprecipitation protocol for enrichment factor determination.....	128

List of tables

Table 1 : Unmodified RNA oligonucleotides (middle fragment, fragment 2).	102
Table 2 : Modified RNA oligonucleotides (middle fragment, fragment 2).	103
Table 3: Modified RNA oligonucleotides discarded from the study (please refer to 3.1.2).	103
Table 4: DNA oligo.	104
Table 5: pDNA vectors.	104
Table 6: RNA sequences	105
Table 7 : Settings for gel scans based on spectroscopic properties of staining dyes/labels.	113
Table 8 : HPLC gradient applied for for circular RNA	114
Table 9 : HPLC gradient applied for ligated linear RNA purification	114

Abbreviations

A	Adenosine
ACN	Acetonitrile
AMP	Adenosine monophosphate
APS	Ammonium persulfate
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
A _m	2'-O-methyladenosine
C	Cytidine
CBP20	Cap-binding protein 20
CBP80	Cap-binding protein 80
CHCl ₃	Chloroform
C _m	2'-O-methylcytidine
CPSF	Polyadenylation specificity factor
CrPV	Cricket paralysis virus
CSTF	Cleavage stimulation factor
CTD	Carboxy-terminal domain
DAD	Diode array detector
DCP1	Decapping enzyme subunit 1
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
<i>E. coli</i>	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
eIF	Translation initiation factor
EMCV	Encephalomyocarditis virus
ESI	Electrospray ionization
EtOH	Ethanol
G	Guanosine
G _m	2'-O-methylguanosine
GMP	Guanosine monophosphate
GTF	Transcription factors
GTP	Guanosine triphosphate
HeLa	Henrietta Lacks cells
hm ⁵ C	5-hydroxycytosine
HPLC	High-performance liquid chromatography
IRES	Internal ribosome entry site
IVT	<i>In vitro</i> transcription
KCl	Potassium chloride
LB	Lysogeny broth (Luria-Bertani broth)
LC	Liquid chromatography
m ¹ A	N ¹ -methyladenosine
m ³ Um	3, 2'-O-dimethyluridine
m ⁵ C	5-methylcytosine
m ⁶ ₂ Am	N ⁶ , N ⁶ , 2'-O-trimethyladenosine
m ⁶ A	N ⁶ -methyladenosine
m ⁶ Am	2'-O-dimethyladenosine
m ⁶ AMP	N ⁶ -methyladenosine 5'-monophosphate

m ⁷ G	N7-methylguanosine
MDA5	Melanoma differentiation-associated protein 5
Met	Methionine
MgCl ₂	Magnesium chloride
mRNA	Messenger RNA
MS	Mass spectroscopy
MST	Microscale thermophoresis
NTP	Nucleoside triphosphate
NTP	Nucleoside triphosphate
PAGE	Polyacrylamide gel electrophoresis
PAMP	Pathogen associated molecular pattern
PAP	Poly(A) polymerase
pDNA	Plasmid-DNA
PIC	Pre-initiation complex
PKR	protein kinase RNA-activated
pol	Polymerase
PRR	Pattern recognition receptors
PRR	Pattern recognition receptors
QQQ	Triple quadrupole mass spectrometer
RIG-I	Retinoic acid-inducible gene I protein
RIG-I	Retinoic acid-inducible gene I protein
RNA	Ribonucleic acid
RP	Reversed phase
rpm	Rotation per minute
RRL	Rabbit Reticulocyte Lysate
rRNA	Ribosomal ribonucleic acid
rRNA	Ribosomal RNA
s ² U	2-thiouridine
SAM	S-adenosyl methionine
siRNA	Small interfering RNA
T	Thymidine
TBE	Tris-boronic acid-ethylenediaminetetraacetic acid
TEAA	Triethylammonium acetate
TEMED	Tetramethylethylenediamine
TLC	Thin layer chromatography
TLR	Toll-like receptor
Tris	Tris(hydroxymethyl)-aminomethane
tRNA	Transfer RNA
U2AF	Splicing auxiliary factor
Um	2'-O-methyluridine
UTR	Untranslated region
VCE	Vaccinia virus capping enzyme
Ψ	Pseudouridine/Psi
2'-O-Me	2'-O-methylation

1 Introduction

1.1 Messenger RNA

The classic view of the central dogma of biology described the flow of genetic information contained in the deoxyribonucleic acid (DNA), to ribonucleic acid (RNA), then to proteins ¹. DNA and RNA are mainly composed from different nucleobases which can be divided into purines (adenine, A, and guanine, G) and pyrimidines (Figure 1.1) ². While DNA contains a thymine (T) as a pyrimidine, RNA has an uracil (U). Both molecules differ also in the presence of a hydroxyl group on the RNA ribose which is absent in DNA. The discovery of an intermediate molecule mRNA, between DNA and protein led to a huge revolution in biology in the fifties/sixties. Advanced researches unravelled the remarkable complexity of the RNA field and the existence of multiple RNA derivatives, such as the well-known ribosomal RNA (rRNA), transfer RNA (tRNA), recently circular RNA and a big class of non-coding RNA like the small interfering RNA (siRNA). The list is long.

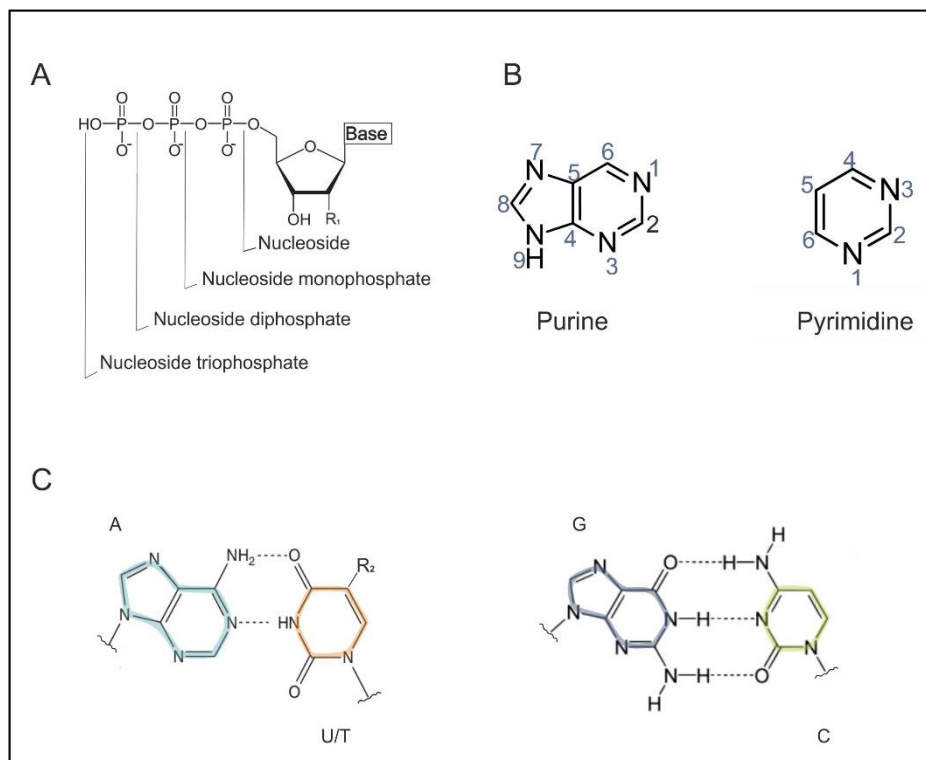


Figure 1. 1: Structure of Nucleic Acids.

(A) Schematic representation of a nucleoside and nucleotides (nucleoside monophosphate, diphosphate, and triphosphate). For DNA $R_1=H$ and for RNA $R_1=OH$. (B) Nomenclature of purine and pyrimidine. (C) The Watson-Crick pairs between bases: Adenosine A (blue) and uridine U ($R_2=H$) for RNA, and thymidine T ($R_2=CH_3$) for DNA (orange). On the right side, guanosine G (purple) and cytosine C (green).

1.1.1 mRNA biogenesis

The eukaryotic mRNA's life cycle begins in the nucleus. There, RNA polymerase transcribes the nucleotide sequence of the gene DNA into mRNA. In eukaryotic, there are three different polymerases (pol) namely pol I, pol II, and pol III. Only polymerase II showed the ability to support the maturation of the nascent mRNA³. Additional mediators are required for the transcription, such as the general transcription factors (GTF), interacting from one side with the polymerase II, and from the other side with a DNA sequence-site defined as core promoter. This latest consists of a core transcription element that contains a TATAA sequence (known as "TATA-box"), the initiator sequence (Inr) and possible other "*cis-acting*" sequences as regulatory elements. Briefly, transcription initiates when a transcription factor TFIID binds to the TATA box⁴⁻⁷, while another factor TFIIA stabilizes the complex. Other factors such as TFIIB (required for an accurate site selection) and TFIIF are recruited progressively to constitute the final binding platform for the RNA polymerase, and then of TFIIIE and TFIIH. One of the two TFIIH subunits denatures the double helix of the DNA sequence around the initiation site and transcription takes place a few nucleotides after the TATA box sequence^{8,9} (Figure 1. 2). Regulation mechanism is facilitated through *cis*-acting elements to up- or down-regulate transcription. They are located upstream or downstream of the promoter core: called the "Enhancers" and "Silencers", which are recognized by activator and repressor proteins, respectively. They can operate independently from their long-distance location, through a loop formation to distant approximate the activator/repressor proteins to transcription complex, which is bound to the promoter¹⁰. The synthesized precursor mRNA undergoes rapidly extensive processing resulting into a competent mRNA able to fulfil main function in protein expression.

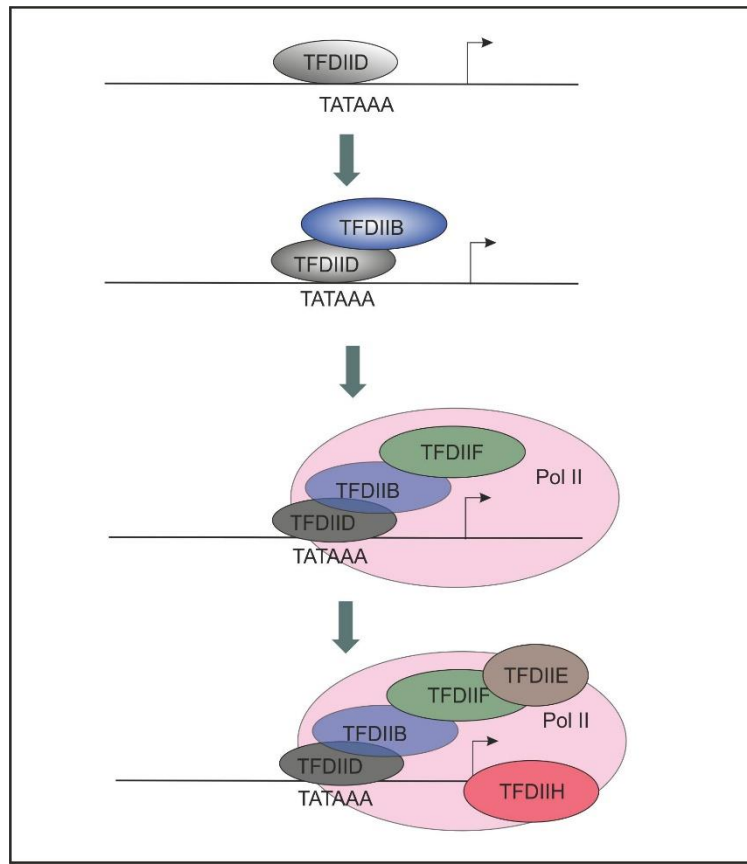


Figure 1. 2: The step-wise assembly of pre-initiation complex assembly in eukaryotes.

The straight black line refers to the promoter DNA. The factor TFIID binds to the TATA sequence upstream of the transcription start site. Then, TFIIB, TFIIF and RNA polymerase II (pol II) are recruited. The complex formation is terminated TFIIE and TFIIH binding. This figure was adapted from ⁹.

1.1.2 The pre-mRNA processing and functional activity

The mRNA synthesis and maturation are often described as two separated pathways, happening in a sequential manner. Efforts over the last decades have revolutionized this understanding and revealed that both events were in fact spatio-temporal intimately coordinated. This phenomenon was described by ‘Miller spread’ electron micrographs, demonstrating the Pol II was still bound to the DNA template, in the same time, the introns were being excised from nascent transcripts ¹¹. A marking feature of a large subunit composing the Pol II is the presence of a unique carboxy-terminal domain (CTD) ¹². Depending on the phosphorylation state of the amino acids in the CTD, it could mediate an electrostatic interaction regulating the implication of the pol II in the transcription machinery, the

deployment of splicing factors, and the binding of the 5' capping cleavage and polyadenylation machinery¹³⁻¹⁵. Maturation of the pre-mRNA on the 5' end consisted of the addition of a cap, by attaching a 7N-methylguanosine *via* a 5'-5' linkage, referred as m⁷G cap or m⁷GpppNpN-, called cap0 (Figure 1.3). Once in the cytoplasm, the cap structure is crucial for the canonical cap-dependent translation in eukaryotic mRNA¹⁶. Using SAM (S-adenosyl methionine) as a methyl donor¹⁷, additional methylation by the 2'O-methyltransferase to the second position of the initiating nucleotide and/or also on the next one, generates the cap 1 (m⁷Gppp^{m2}NpN-) or cap 2 (m⁷Gppp^{m2}Np^{m2}N-), respectively.

Usually, pre-mRNA molecule contains coding sequences (exons) that are interrupted by numerous non-coding sequences (introns), and are shortened by splicing the introns and joining the exons together. Globally, the splicing is achieved *via* two transesterification steps. At a first place, the cleavage at the 5' end of the intron (donor site), accompanied by binding of the 5' end to the 2' hydroxyl group of an "Adenosine" residue, known as the branch point, located between 20-40 nucleotides in upstream of the 3' end of the intron. This results in lariat structure formation. In a second step, cleavage took place at the 3' end of the intron and the two exons were ligated together. The formed lariat was released and rapidly degraded¹⁸. Moreover, the same single gene can also result in a variety of proteins, by alternative inclusion or exclusion of particular exons, described as the "alternative splicing"¹⁹.

Given the vital involvement of the cap in the mRNA process, its removal is highly regulated and organized in a targeted manner^{20,21}. Decay machinery including decapping enzymes Nudix hydrolase Dcp1, Dcp2, DcpS, and Nudt16 were identified to undergo the mRNA degradation from the 5' to 3' direction^{21,22}.

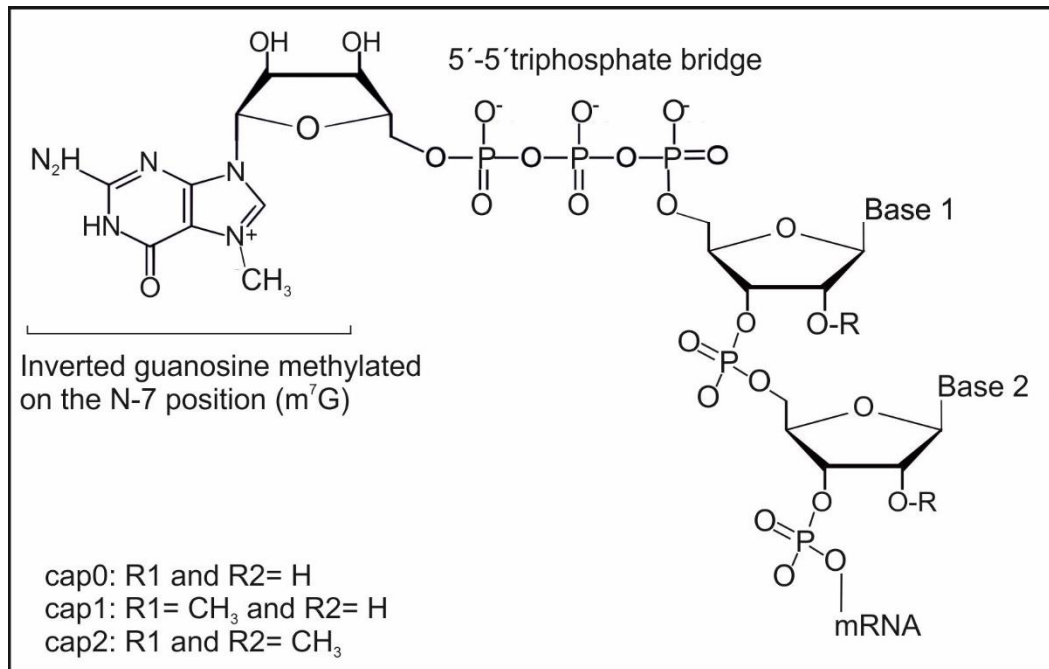


Figure 1. 3: The eukaryotic mRNA 5'cap structure.

Figure adapted from ⁴¹⁵.

Maturation at the 3' end of the pre-mRNA transcripts is processed by the addition of a poly(A) tail. This reaction of polyadenylation consists of a two-steps process, of cleavage and polyadenylation reaction. Briefly, the cleavage and polyadenylation specificity factor (CPSF), recognizes the polyadenylation signal "AAUAAA" (or a homolog) and catalyzes pre-mRNA cleavage ²³. Then, the cleavage stimulation factor (CSTF) and the cleavage factors (CFI and CFII) are recruited. These subunits specifically recognize the U/GU-rich sequence of the downstream element (DSE) ^{24,25}. Together with other factors, they bind directly to the mRNA to form a core complex and activate the 3' end processing. Finally, the poly(A) polymerase (PAP) joined the complex and added the poly-adenosine tail in a co-transcriptional manner. Indeed, an interaction between the CTD of the pol II with the polyadenylation factors CPSF and CSTF could be identified supporting further the notion that pre-mRNA processing is co-transcriptional ¹⁴.

Furthermore, an interaction between polyadenylation factors CFI and CPSF, and the U1 snRNP and U2 snRNP (components from the splicing machinery) was described. Consequently, the

splicing of the last intron could be facilitated by the recognition of the poly(A) tail, and maintained a balance between polyadenylation and splicing. A role of poly(A) tail to terminate transcription was as well reported ²⁶. Similar to the cap structure, poly(A) tails contributes also to the preparation of the mRNA nuclear export to the cytoplasm ²⁷, where it contributes in the translation step ^{28,29}, and can be also detrimental for their fate in the cells as the mRNA turnover can be dependent on the time required to “deadenylate” them.

1.1.2.1 Cap-dependent eukaryotic mRNA translation

In eukaryotes, the recruitment of ribosomes by the mRNA is achieved by several steps. The eIF4F, an initiation factor complex, includes DEAD-box RNA helicase eIF4A, the eIF4G as a scaffold protein, and the cap binding factor eIF4E. The latter allows the binding to the cap by interacting with the methyl moiety of the guanosine residue. The assembly of the 40S ribosomal subunit to initiation factors (such eIF3) and the eIF2–GTP–tRNA^{Met} ternary complex form together the 43S pre-initiation complex (43S PIC). The association of the 43S PIC to the mRNA is facilitated by the interaction between the eIF4G and eIF3. Together, they form the 48S initiation complex which then “scans” in the 5’ to the 3’ direction to encounter the AUG start codon, located within the Kozak sequence. Pioneering studies by Marilyn Kozak reported the sequence RCCAUGG (R = purine, A or G) to be the optimal context for AUG recognition in eukaryotes ³⁰. In the other hand, the scaffold translation factor eIF4G, a component from the eIF4F, interacts with the poly(A) binding protein (PABP), which is associated to the 3’ poly(A) tail and lead to the circularization of the mRNA, known as “closed loop” ³¹. This form stimulates the mRNA translation and supports as well the ribosome recycling step ³², (Figure 1. 4 A). Lastly, while some translation factors are released and 60S subunit is recruited to form the elongation-competent 80S ribosomes (reviewed in ³³).

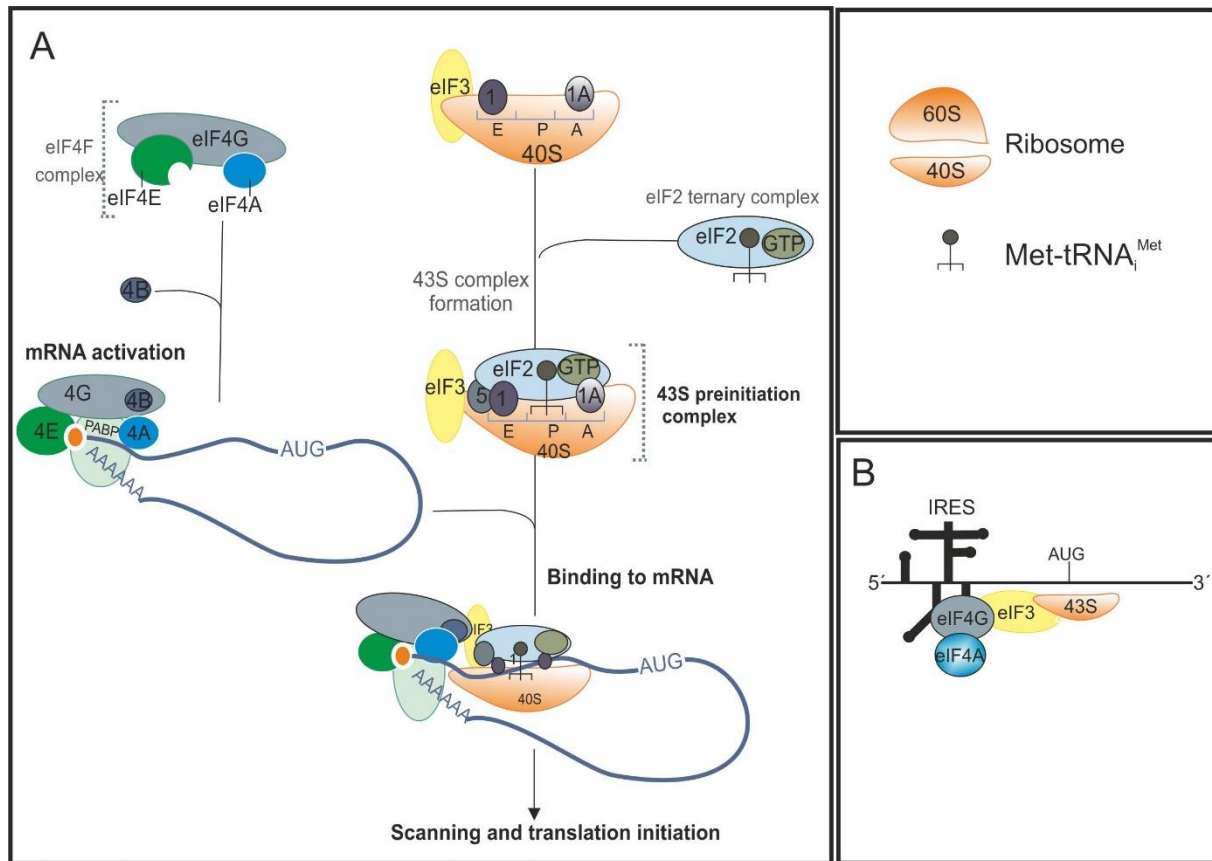


Figure 1. 4: Models of translation initiation.

(A) The translation pathway includes a series of sub-steps and several binding factors. During translation initiation, the assembly of the Met-tRNA^{Met} with the initiation factor 2 (eIF2)–GTP–Met-tRNA^{Met} to form a ternary complex. The mRNA is selected by the complex eIF4F binding to the cap structure (orange circle), including an ATP-dependent helicase eIF4A to unwind proximal secondary structures. The eIF4F complex includes as well the eIF4G, a scaffold protein that harbours binding domains for PABP, eIF4E, eIF4A, and eIF3. Other factors bind to the small ribosomal subunit 40S, such as eIF1 (1), eIF1A (1A), eIF5 to form together the pre-initiation complex 43S PIC. The latter is recruited to the 5' end, where it starts scanning of the 5'UTR in 5' to 3' direction. Once recognizing the initiation codon, 48S initiation complex is formed by the attachment of the 60S subunit, accompanied with the displacement of some initiation factor eIF1 and eIF3, while eIF5 mediated the hydrolysis of eIF2-bound GTP and Pi release. The resulting complex (not shown here), is called 80S initiation complex which is competent to start elongation. Figure adapted from references ^{33,416}. (B) Schematic diagram of cap-independent translation. In this example, translation initiation is mediated by the encephalomyocarditis virus (EMCV) IRES.

1.1.2.2 Cap-independent translation pathway

An alternative mechanism to cap-dependent translation can be initiated by recruiting ribosomes internally through the internal ribosome entry site (IRES). Early discoveries identified this sequence in the viral genome, and subsequent studies showed IRES presence in many mRNAs, i.e. DAP5 ^{34–36}. This mechanism allows to directly recruit the translational machinery to an internal initiation codon, involving *trans-acting* factors (not always) ³⁷.

In the case at hand, the encephalomyocarditis virus (EMCV) IRES from the picornavirus family was used. It includes structural domains H, I, J-K, and L upstream of the initiation codon AUG (Figure 1. 5). This initiation codon is also known as the AUG-11 which is the authentic initiation codon, while translation may rarely occur as well, on a near AUG (AUG-12) situated a few codons further downstream³⁸. The sequence is divided into 5 domains, containing a conserved motif GNRA (N stands for any nucleotide, and R for purine) in domain 3 which enhances the IRES structural body and function³⁹⁻⁴¹. The EMCV contains as well a pyrimidine-rich motif in domain 2 and domain 5 that provides binding of site for the polypyrimidine tract binding protein (PBP), and eIF4B, PTB and other RNA-binding proteins⁴²⁻⁴⁵.

In general, IRES structure is not static and enables the interaction with components from the translation machinery, recruiting a non-canonical eIF4G protein (eIF4G2) that directly recognizes an IRES to initiate translation complex assembly (Figure 1. 4 B). For translation initiation, the current translation model suggests that EMCV IRES requires the core of the 43S complex, eIF4A which strongly enhances the binding of the central domain of eIF4G to J-K domains of EMCV-IRES in adenosine triphosphate (ATP) dependent manner^{38,46,47}. Indeed, it requires almost the same factors as the cap-dependent initiation mechanism except for eIF4E⁴⁸⁻⁵⁰. Among other different IRESs, the Hepatitis C virus IRES which was reported to bind directly to eIF3 and the 40S ribosomal subunit and the cricket paralysis virus (CrPV) IRES which was described to be eIF-independent, thus, it could bypass the need for the entire factors⁵¹,

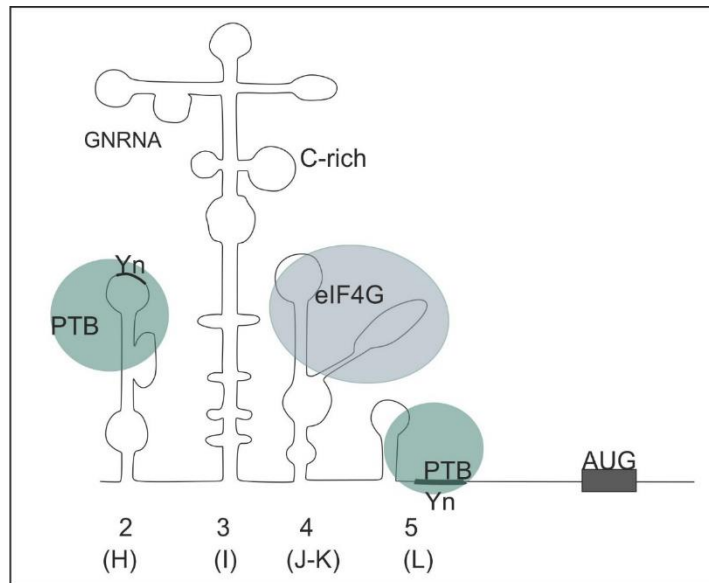


Figure 1. 5: Schematic diagram of the EMCV IRES secondary structure main features.

The conserved motifs (GNRA) in domain 3, C-rich site, and the pyrimidine tract Yn are indicated. The eIF4G translation factor's binding site is depicted by a purple circle in domains J-K in the domain 4 and the polypyrimidine binding protein (PTB) is recruited by the pyrimidine-rich Yn motif (green circles). Figure adapted from ⁴⁵.

1.2 Natural modifications in RNA

Besides the four regular nucleosides (adenosine, guanosine, cytidine, and uridine) composing the RNA sequence, other chemical structures were identified: small modifications installed on the ribose or the base of the nucleosides, such as methylations, deaminations, thiolations, glycosylations, isomerizations (some examples are illustrated in Figure 1. 6). Together, these marks are part of the epitranscriptomics RNA field ⁵². Progress and development of sequencing and detection technologies contributed widely in the mapping of modifications transcriptome wide ⁵³, with a growing repertoire accounting for over 170 distinct modifications ⁵⁴. They were not found only widespread across the three phylogenetic domains (Archaea, Bacteria, and Eukarya), but also mapped abundantly in both coding RNA (i.e. mRNA and coding circRNA ZNF609 ⁵⁵) and non-coding RNA ⁵⁶, reviewed in ⁵⁷. The discovery of mediating enzymes, responsible for the edition of certain modifications, had a positive impact for the understanding of the RNA modification's regulation.

Modifications in ribosomal and transfer RNA (rRNA and tRNA), two necessary components of translation machinery, are not only beneficial but also necessary for their proper function, and can directly influence mRNA decoding during protein synthesis ⁵⁸. Eukaryotic

rRNA are extensively modified with for example (Ψ) and by the addition of a methyl group at the 2'-O position of the ribose sugar residues (2'-O-Me). The latter could be also mapped due to its higher resistance to alkaline-mediated hydrolysis (⁵⁹, and review ^{60,61}). In this case, RNA sequences as the C/D and H/ACA boxes of small nuclear RNA (snoRNAs) are guiding the installation of those modifications ⁶². In general, modifications are not randomly distributed but tend to cluster in functionally important regions of the ribosome, near by the peptidyl transferase center in the large ribosomal subunit, and around the decoding site in the small ribosomal subunit, suggesting a role in translation fidelity and a correct positioning of tRNA and mRNA ^{63–66}.

The tRNA is the second most abundant specie of RNA in cells (following rRNA), carrying amino acids during protein synthesis. An average of 13 modifications per molecule was estimated in tRNA ⁶⁷. Their prevalence may be affected by metabolic states, the general stress in cells, and tRNA aminoacylation ^{68,69}. The most elaborate tRNA modifications were located in the anticodon loop, in particular positions 34 and 37, as they are highly conserved in eukaryotes and can directly contribute for the decoding regulation. Position 34, also called the wobble position, base pairs with the third base in the mRNA codon and has a role in deciphering the degenerate genetic code ^{70, 71}. Modifications at this position were described to fine-tune wobble pairing by extending or limiting the number of the pairing codons and the lack of certain modification could be lethal ^{72,73}. As an example, when position 34 is a uridine, it can be modified into 5-taurinomethyl-2-thiouridine, having consequences on the stabilization of the codon-anticodon interaction. This modification mutation can trigger serious neurological disorder such as mitochondrial encephalomyopathy, lactic acidosis and stroke-like episodes (MELAS), and (myoclonic epilepsy with ragged-red fibers (MERRF) ^{74–76}. Another example is the tRNA^{Phe} derived from patients with nonsyndromic X-linked intellectual disability disease. It was found to contain a 2'-O-methylation hypomodification at the guanosine position 34 (Gm34) ⁷⁷. More related examples are reviewed in ⁷⁸. Position 37, adjacent to the 3'-side of the anticodon, supports the interaction between the anticodon to the codon by improving intrastand stacking within the anticodon loop and may impact on translation, affecting consequently the cellular growth ^{79,80}. It was reported for instance that the deficiency of 1-methylguanosine in *Salmonella typhimurium* tRNA can affect the frameshifting ⁸¹. Modifications distributed in the main tRNA body are known to influence the structure and the proper geometry of the tRNA. The finest example by Helm and colleagues reported the requirement for the methylation at adenosine 9 (A9) to enable the canonical cloverleaf folding ⁸². Beyond their central role in

epitranscriptomic, modifications on tRNA may be hijacked by viral infection, to make a benefit for their proper replication^{83,84}.

1.2.1 Modifications in eukaryotic mRNA

Similarly to rRNA and tRNA, eukaryotic mRNA harbour several types of modifications, the most abundant ones being: *N*⁶-methyladenosine (*m*⁶A)^{85,86}, 5-methylcytosine (*m*⁵C)⁸⁷, Ψ^{88,89}, *N*¹-methyladenosine (*m*¹A)⁹⁰ and 2'-O-Me, (Figure 1. 6)⁹¹. In general, modifications have the potential to be implicated in various steps of the mRNA life cycle, from maturation and processing to translation⁹².

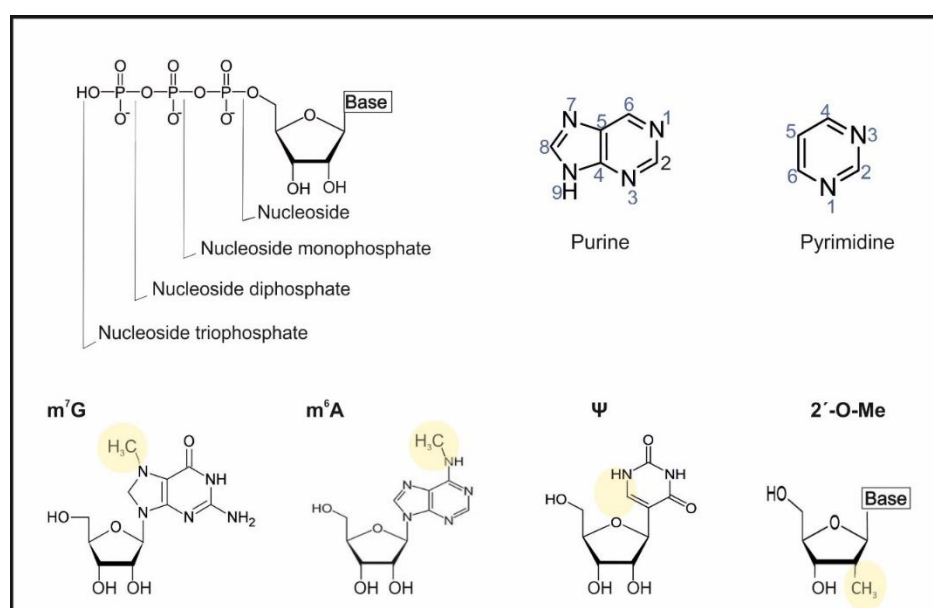


Figure 1. 6: Examples of RNA modifications.

Schematic representations of nucleoside and nucleotide (top left), and nomenclature of purine and pyrimidine (top right). Examples of RNA modifications (down).

Starting on the 5' end of the transcript, one of the earliest identified modification is the *m*⁷G, as a hallmark of the cap structure at the 5' end⁹³. The adjacent cap nucleotide (+1 position) could contain a 2'-O-methylation defining the mRNA as a “self” and avoiding the induction of the immune response mediated by the cellular sensors retinoic acid-inducible

gene I protein (RIG-I), the melanoma differentiation-associated protein 5 (MDA5) and the type I interferon effectors^{94–99}. An additional methyl group was reported on the +1 position on the adenosine position N6, the so-called *N*6, 2'-O-dimethyladenosine (*m*⁶Am) deriving

from an (Am)^{100,101}. In general, the double modification provided further resistance against enzymatic decapping reactions¹⁰².

Internal modifications in the mRNA body could be also detected, such as *N*¹-methyladenosine (m¹A). It was mapped in the 5' untranslated sequences⁹⁰, and tends to occur in structured regions with higher GC content^{103,104}. Since this modification is able to block Watson-Crick base pairing¹⁰⁵, it could also alter the predicted secondary/ternary structure of mRNA. In their work on mapping m¹A in eukaryotic cells, Dominissini and colleagues reported that this modification was found enriched around canonical and alternative translation initiation sites, and can be dynamically regulated in response to physiological stimuli by promoting protein levels⁹⁰. However, the main role of m¹A in mRNA requires further investigations. Indeed, *in vitro* translation study of m¹A-modified mRNA suggested that such modification can modulate translation in a position-dependent manner within a codon, without compromising translation fidelity¹⁰⁶. The identification of 5-methylcytosine (5mC) modification in DNA was previously performed by bisulfite treatment and sequencing based methods¹⁰⁷. This method was adapted for mapping of m⁵C in mRNA because of the harsh conditions of bisulfite-treatment. Many groups resorted to combining bisulfite-sequencing with other methods, i.e. mass-spectrometry and immunoprecipitation, and reported that m⁵C sites are likely to be distributed along the mRNA transcript in a cell/tissue type specific manner. Two groups reported that this modification was preferentially localized in the 5' UTR, around the start codons^{107,108}, while another study reported an enrichment in the CDS¹⁰⁹. Furthermore, recent data revealed also an overlap of the m⁵C sites with RNA binding-protein sites, such as binding sites of the Argonaute protein or the export adaptor protein ALYREF^{108,110}. Consequently, it suggested a multifaceted involvement in various processes.

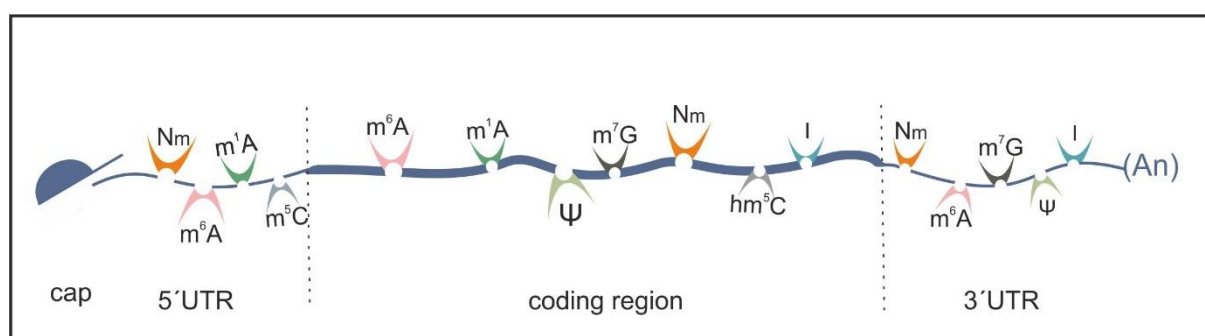


Figure 1. 7: Overview of the distribution of chemical modifications in eukaryotic mRNA.

Examples of most detected modifications: inosine (I), N_m (2'-O-Me), m⁵C, m⁶A, m⁷G, ac⁴C (N⁴ – acetylcytidine), hm⁵C (5-hydroxymethylcytidine). Figure adapted from¹¹⁹.

1.2.1.1 *m⁶A in mRNA*

The *m⁶A* modification is one of the most abundant internal mRNA modification in eukaryotes and was originally described in the seventies^{85,111,112} (Figure 1. 6). In mammalian mRNA, an average of 1 to 3 sites per a given mRNA transcript can be methylated^{111,113}. Transcriptome-wide mapping identified that methylations occur in a preferred sequence motif “DRACH “ (D=G/A/U, R=G/A, H=A/U/C)^{114,115}. There are more occurrences of the methylation consensus motif in transcripts than there are actually detected *m⁶A*. Hence only some and not all consensus motifs are methylated¹¹⁶. The methylation is generated by a “writer” complex consisting of methyltransferase like 3 (METTL3) supported by the allosteric activator methyltransferase like 14 (METTL14)^{85,117,118}, and is co-transcriptionally installed^{119,120}. Other associated proteins are expected to contribute to maintaining the complex activity and the specificity of the methylation, by augmenting the translocation into nuclear speckles, such as the interactor VIRMA (also known as KIAA1429) and Wilms tumor 1 associated protein (WTAP)^{121–123}, respectively. In the speckles, the methylation can be removed by demethylases, the fat mass and obesity-associated protein (FTO), and alkylation repair homolog protein 5 (ALKBH5)^{124–126}. Together, writers and erasers enable the reversibility of epigenetic events to maintain the balance to the overall biological system process. Deregulation of *m⁶A* by ALKBH5 knockout in mice had consequences on sperm development and mice fertility¹²⁴. Compiling work from several studies in the transcriptome-wide mapping of modifications revealed, that *m⁶A* is enriched in the 3’ UTR near stop codons and the main coding region, while the 5’UTR is relatively low methylated^{114,117}. Accordingly, *m⁶A* can affect the dynamic regulation of protein expression and mRNA stability, through its ability to influence local secondary structures and the recruitment of binding proteins. Reader family of YTH domain family (YTHDF1, YTHDF2, YTHDF3, and YTHDC1), was characterized to bind selectively *m⁶A* modified mRNA^{114,127,128,129}. In the nucleus, YTHDC1 was reported to have a role in alternative splicing by favouring interaction with the splicing factor SRSF3 while blocking SRSF10, facilitating exon inclusion^{128,130}. During the mRNA’s journey between the nucleus to the cytoplasm, *m⁶A* modification was found to promote mRNA nuclear export mediated by YTHDC1 binding to targeted *m⁶A*-transcripts¹³¹. Understanding of how *m⁶A* modifications exert their role remains a central question. It was reported that *m⁶A* located in the 5’ UTR can directly recruit initiation translation factors (eIF3), or be mediated by another factor (ABCF1) to enable cap-independent translation^{132,133}, (Figure 1. 8 A).

Wang and colleagues reported that the binding sites of the reader YTHDF1 were found to overlap with the m⁶A peaks, mostly in the 3' UTR around the stop codon. They suggested a translation model, where YTHDF1 actively promotes protein synthesis of m⁶A-modified transcripts, by interacting with several factors from the translation machinery, in particular with eIF3. Their association forms a “closed-loop”, similar to the canonical cap-dependent translation loop, promotes translation initiation ¹³⁴, (Figure 1. 8 B). Such a mechanism could guide translation in response to stress stimuli of selective transcripts when the canonical cap-dependent translation is compromised ^{135–137}. Notably, m⁶A methylation in 5' UTR increased upon heat shock stress, a process supported by YTHDF2 which probably translocates to the nucleus in order to preserve 5'UTR methylation transcripts by limiting the m⁶A ‘eraser’ FTO from demethylation ¹²⁹.

Previous thermodynamic studies reported the ability of m⁶A to influence mRNA folding and altering local structures to regulate binding protein regulation, i.e. the binding to heterogeneous nuclear ribonucleoprotein C (HNRNPC), a protein involved in alternative splicing ^{138,139}. In a recent study in a study employing *in vitro* translation assays, m⁶A localized in the mRNA coding region induced rather a translation elongation attenuation, and the effect was strongly position-dependent, with a major effect obtained when methylation was in the first codon position (m⁶AAA) ¹⁴⁰. Focussing on the function mechanism, Choi and colleagues suggested a model where m⁶A, presumably, affected the decoding process of ribosomes and tRNA accommodation ¹⁴¹. In addition to their implication in mRNA expression, another role in regulating mRNA decay was assigned to m⁶A readers. This could be shown, for instance, by the ability of reader YTHDF2 to directly recruit the deadenylase complex to regulate mRNA decay, ensuring a dynamic control of the lifetime of target transcripts ^{127,142}. Besides, the enriching of m⁶A in the 3'UTR upstream of the stop codons in mammalian mRNA could be associated to a potential role in regulating miRNA interference ¹¹⁷. While much of the work focused on uncovering the implication of m⁶A in biological mRNA pathways, concerns about the effective mapping of m⁶A were raised, in particular with antibody-based strategies. In this regard, antibody

specificity could be doubted in particular to cross-reactivity reactions with closely related chemical species ^{101,143,144}.

Given the overall methylation participation in essential biological processes of mRNA transcripts, it would not be surprising to elaborate connections with diseases including cancer. Consistent with this, the writer protein METTL3 was found to be involved in the promotion of the translation of important oncogenes such as EGFR and TAZ and can support cancer cell growth¹⁴⁵. Concerning the eraser protein FTO, it was detected in high levels in human AML samples and enables the acceleration of leukemogenesis ¹⁴⁶. Very recently, the differential regulation of the level of m⁶A methylation of certain genes was linked to Alzheimer's Disease ¹⁴⁷, and other potential involvements in other diseases are reviewed in ¹⁴⁸.

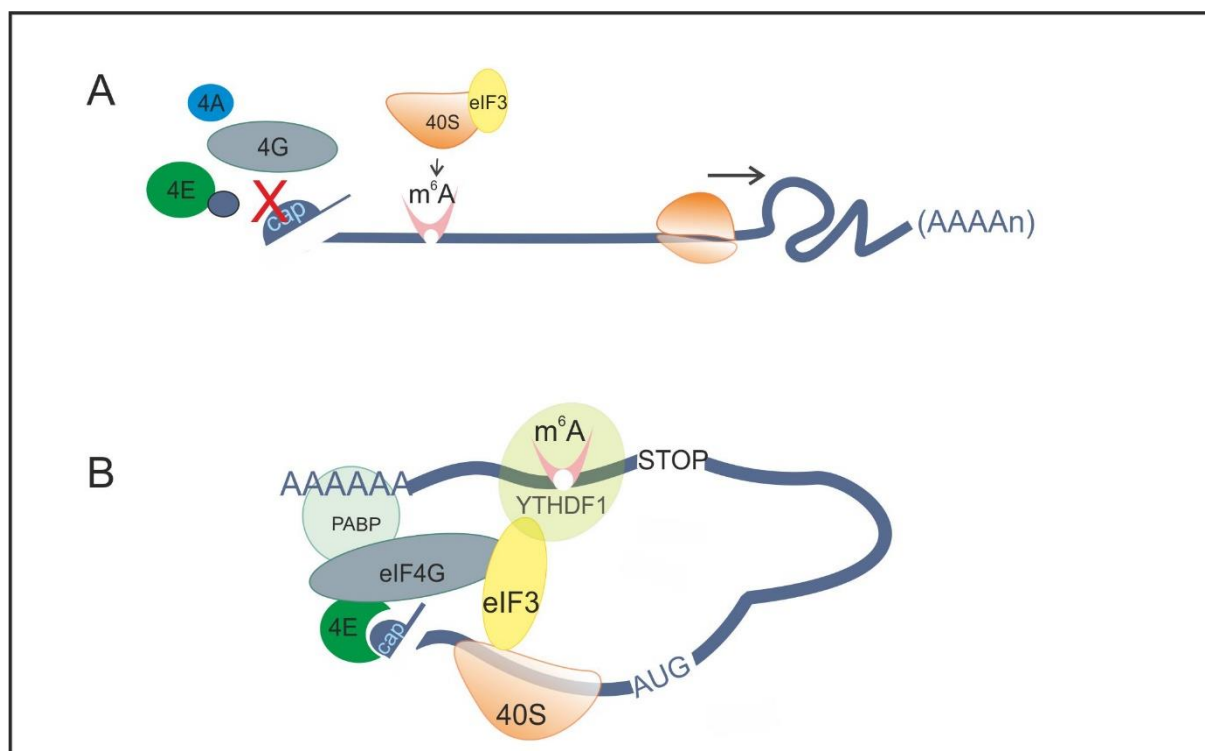


Figure 1. 8: Translation initiation driven by m⁶A.

(A) Translation can be initiated by m⁶A residues in the 5'UTR in a cap-independent manner. The modification is directly recognized by the eukaryotic initiation factor 3 (eIF3), which recruits the translation machinery to initiate translation. (B) When localized near the 3' UTR of mRNA transcripts, m⁶A can promote translation. YTHDF1 can bind to m⁶A and interact with the eIF3 through an mRNA looping mechanism. Figure adapted from ^{132, 134}.

1.2.1.2 Pseudouridine in mRNA

Pseudouridine (Ψ) is the first modified nucleoside discovered in the fifties and initially described as the “fifth RNA nucleotide”^{149–151}. It differs from uridine by the formation of C-C bond instead of the canonical C–N glycosidic bond^{152,153} (Figure 1. 6). The detection of Ψ in mRNA was recent, with the establishment of PseudoU-seq for mapping Ψ at a single-base resolution^{88,89}. Compiling work from several groups revealed a widespread prevalence of Ψ in eukaryotic mRNA estimated to 0.2%–0.6% of uridines converted to pseudouridines, which is comparable to the content of m⁶A in mRNA transcripts^{154–156}. Notably, pseudouridines are distributed along the 5' UTR, coding sequences and 3' UTR^{88,155,157–159}. Generally, the pseudouridylation can be catalyzed by RNA pseudouridine synthase (Pus) alone, or *via* an RNA-dependent mechanism. The latter is mediated by an RNA component called box H/ACA on a small nuclear RNA (snoRNA), associated with a set of evolutionary conserved core proteins. The snoRNA guides pseudouridine modifications at specific sites, through base-pairing with substrate RNA, and involves enzymatic proteins as dyskerin or Cbf5, in human and yeast respectively, to catalyze the U-to- ψ isomerization reaction^{160,161}. While the general mechanism for several RNA species was described, the basis in mRNA target recognition for pseudouridylation is still not well clear. Emerging prediction of pseudouridine synthase genes without an attributed function, may expand target substrate to mRNA molecules. In line with this, Schwartz and colleagues suggested the presence of a consensus sequence that guides site-specific PUSs (pseudouridine synthase PUS7) to catalyse pseudouridylation in mRNA. Importantly, PUS7 proteins could induce the pseudouridylation level under heat shock conditions, while their deletion decreased it⁸⁸. Such regulation control suggests a role in enhancing transcript stability and a distinct role in response to stress stimuli (reviewed in¹⁶²). In the same context, another group revealed mRNA pseudouridylation by PUS1 directed by structure-depend recognition of a shared sequence motif between these substrates¹⁶³.

Given the distribution along the mRNA transcript, it is likely plausible that Ψ interferes with distinct steps of mRNA's life cycle. A potential role in the formation of the spliceosome complex was suggested by Chen and colleagues since this process requires a polypyrimidine (C and U) tract near the 3' splice site. The conversion of U to Ψ in this site led to a splicing defect by affecting the binding of a splicing auxiliary factor (U2AF) to the polypyrimidine tract in *Xenopus* oocytes¹⁶⁴.

Owing to its chemical property, the isomerization does not affect classical base pairing at the Watson–Crick edge with adenosine but rather enhance it, and enables a more stable pairing with any of the principal bases ^{165,166}. A contribution to increasing phosphodiester back rigidity and base stacking was also described ^{165,167}, and may contribute to influencing RNA secondary structure. Because Ψ was considered to expand the genetic code by partially rewiring it ^{168,169}, it can reasonably anticipate a biological role. Artificial pseudouridylation studies performed by two groups, Karijovich and colleagues and Fernandez and colleagues, have shown that Ψ in stop codons directs the nonsense suppression by converting them to sense codons resulting in altered protein synthesis ^{169,170}. While such a feature could be conserved between the reported species, research is still required to resolve this question. In a followed study using a fully reconstituted bacterial translation system and human cells, Eyler and colleagues have demonstrated that single Ψ exerted an overall low-level peptide expression. They suggested that Ψ increased the rate of near-cognate tRNA by reducing the ability of the *E. coli* ribosome to discriminate between cognate and near-/noncognate aa-tRNAs. This depended strongly on the codon- and nucleotide position within a codon. Mechanically, Ψ perturbed the acceptor stem of tRNA^{Phe} and prevented a proper positioning in the A site of the peptidyl transferase center on the large ribosomal subunit ¹⁷¹. Hoernes and colleagues supported as well the low yield of full-length protein caused by Ψ in the coding region, without recoding miscoding events ¹⁴⁰. In contrast, Kariko and colleagues revealed that base pairing between Ψ and A could be interpreted by tRNA during mRNA decoding and resulted in the translation of functional proteins and increased protein yields in living cells ¹⁷². Interestingly, they also reported a translation alteration by pseudouridine-modified mRNA using other translation systems, namely wheat Germ extract and a bacterial translation system. So far, the mechanisms of function are not well understood. However, it is clear that Ψ modulates translation and affects the decoding fidelity, while more efforts will be needed to clarify the discrepancy between the different studies.

1.2.1.3 2'-O-methylation in mRNA

Continuous efforts in modifications detection revealed that the ribose can be methylated at the second position to form 2'-O-methylated (Nm or 2'-O-Me) nucleosides (Figure 1. 6). This modification was recently detected within the coding sequence of mRNA transcripts ^{91,173}, with an enrichment in codons for certain amino acids ⁹¹. Because of its chemical structure, 2'-O-Me is one of the RNA modifications that do not directly alter the Watson-Crick base pairing. For non-coding RNA such as rRNA, tRNA and miRNA, the addition of 2'-O-Me can be mediated

by two mechanisms: One is mediated by standalone methyltransferases that recognize their targets based on particular structures and sequences, and a second one, through a ribonucleoprotein complex containing C/D-box small nucleolar RNA (snoRNA) which guide to the targets ^{174,175}. A recent study proposed that the second mechanism was followed for the 2'-O-Me installation in mRNA ¹⁷⁶. However, this needs to be further confirmed.

While the biological role in mRNA is still ambiguous, it is now known that the second base adjacent to the 5' cap in mRNA, can be 2'-O-methylated, conferring resistance to nucleases, and is implicated in self/non-self-recognition by the immune system ^{95,177}. Indeed, ribose methylations in RNA were described to act as immunosuppressive, therefore avoiding Toll-like receptor (TLR) activation ¹⁷⁸. Certain sequences including a Gm residue are known to have an antagonistic effect on the activation of Toll like receptor 7 ^{179–181}, hence, are important suppressors of RNA-induced pattern recognition receptor (PRRs) activation ¹⁸². The implication of 2'-O-Me in translation was reported using a synthetic modified-mRNA, and revealed the potential of this modification to repress protein expression in a sequence context-dependent manner ^{140,183–185}, without potential in inducing miscoding events.

1.2.2 Modification detection *via* chromatography

The thin-layer chromatography (TLC) is one of the earliest methods that was used for numerous base modifications investigations. One way was for the localization of modification within an RNA sequence was based on the *in vivo* radioactive labelling of RNA, by supplementing, for instance, bacterial culture by radioactive labeled phosphate (³²P). The transcripts (polyadenylated RNA) were isolated, digested to mononucleotides, and analyzed by TLC. Likewise, the separation between the ³²P-labeled modified and the unmodified nucleotides could be separated for quantification and comparison to synthetic modified nucleotides or nucleosides for assignment ¹⁸⁶. The physiological features such as the net charge, polarity, and hydrophobicity that differ between modified and their unmodified equivalent nucleotides permit the identification of both species by chromatographic mobility, which resulted in a reference map ^{186–188}. Due to nucleic acids ultraviolet absorption capacity, the TLC permits also the identification of modifications from unlabeled samples. However, this requires bigger quantities (a few micrograms) as compared to the detection with radioactivity. In early works, Davis and Allen reported the discovery of an additional nucleotide in RNA isolated from yeast. They described their finding as the fifth nucleoside, which was later called pseudouridine ¹⁵⁰.

In another study on rRNA from *Xenopus laevis* and *man*, authors reported the detection of 2'-O-Me by TLC ¹⁸⁹.

Another revolution came with RNA post-labelling using T4-polynucleotide kinase (PNK), which allowed the 5'-end [³²P] labeling. This was a big step in the field, as the labeling allowed the identification of the end nucleotide in any RNA fragment ^{187,190}. Another alternative could be *via* direct incorporation of $\alpha^{32}\text{P}$ -NTPs into RNA during *in vitro* transcription ¹⁹¹(used in this work).

1.3 Synthetic mRNA design and applications

The use of synthetic mRNA contributed to the fundamental understanding of the cellular mRNA, which brought their application to broader research for therapy. Initial manipulations on mRNA backs to the early eighties. Synthetic mRNA was created by *in vitro* transcription (IVT) using phages RNA polymerase SP6 and the biological activity was validated by mRNA injection into the cytoplasm of frog oocytes and by *in vitro* translation using wheat Germs extract. However, the fragility of the mRNA molecule was one of the reasons that obstructed rapid research progress and more efforts were invested for this purpose. Since IVT mRNA molecules would be lacking the genomic integration which was a major concern in particular for DNA-based therapy, and due its transient presence in the cell permitting temporal functioning and one-step protein expression in the cytoplasm, there was a focus to improve the synthetic mRNA molecule as a novel therapeutic tool. Cumulated knowledge allowed to define basic features of mRNA required for the molecule stability and functionality in cells, essentially the 5' cap, 5'- and 3'-UTRs of interest, and the poly(A) tail.

1.3.1 Basic structural elements

The presence of a 5'cap structure would enhance translation, by bind the eukaryotic translation initiation factor eIF4E ¹⁹². It can be inserted by adding cap analogues during *in vitro* transcription or by using dedicated capping enzymes in a second approach (posttranscriptional). In the first possibility, while the addition of a cap analogue was known for increasing translation efficiency, a fair amount of the mRNA remains uncapped due to the competition from the starting nucleotide GTP when using a promoter sequence of T7 RNA polymerase ^{193–196}. Besides, cap analogue incorporation can lead to two possible orientations: either a correct or a reverse form ^{193,197}, while up to 50% IVT mRNA carried the correct form recognized by the translation apparatus One way to overcome the undesired orientation was by blocking the

elongation at the m⁷G using a 3'-O-methyl instead the 3'-OH of the m⁷G moiety^{195,197} (reviewed in¹⁹⁸, Figure 1. 9). Improving the stability of RNA-based drugs against RNases could be engineered by chemical modifications in the cap structure (Figure 1. 9). The substituting for instance of the O atoms of the triphosphate bridge, achieving O-to-X substitutions (where X could be a single atom or a group of atoms), was studied to increase the resistance to decapping enzymes and enhance the interaction with eIF4E^{195,199,200}.

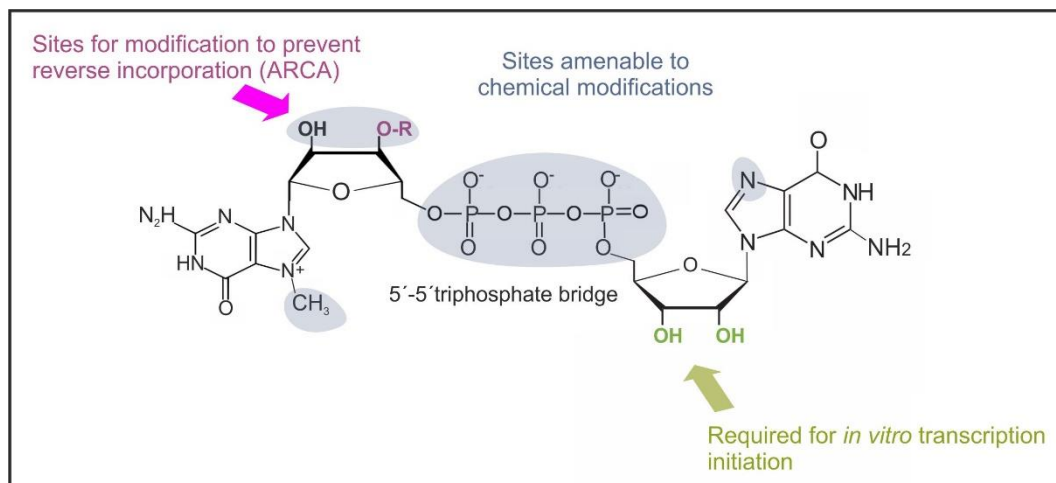


Figure 1. 9: Structures of cap analog.

The requirements for a cap analog to be incorporated into mRNA and to produce anti-reverse cap analogs (ARCAs). The R=H corresponds to the conventional cap analog, and R=CH₃ corresponds to the anti-reverse cap analog (ARCA, 3'-O-Me-m⁷G(5')pppG). Figure adapted from¹⁹⁸.

An alternative capping could be performed by enzymatic reaction, after IVT, which was described to be more efficient using for instance the *vaccinia* capping enzyme¹⁹⁵. In both cases, the remaining free 5' triphosphate generated by *in vitro* transcription could be avoided for certain applications, i.e. cellular transfection or mRNA vaccination²⁰¹, as the free 5'triphosphate may trigger pathogen-associated molecular patterns (PAMPs) *via* the activation of RNA sensors RIG-I. In this regard, eliminating the free 5'-triphosphate could be possible through an extensive treatment with phosphatase followed by a purification step, which is time and material consuming. The second alternative capping can take place post-transcriptionally, and usually performed by ready-to-use capping enzyme kits.

At the 3'-end, eukaryotic mRNA usually carry an adenine tract poly(A) tail structure that plays a double role to increase resistance against nucleases and ensure the interaction with PABP. Investigations of many studies on artificial mRNA, suggested the application of an average poly(A) length between 64-250 residues^{172,202–205}.

The poly(A) tail can be added either by the insertion of a poly(dT) sequence in the coding DNA template, or after IVT by enzymatic reaction using available polyadenylation kits from yeast and *E.Coli* polymerase.

In eukaryotic translation, ribosomes start scanning through along the 5' end seeking for the start codon (AUG). Therefore, the utilization of a strong and optimized Kozak sequence context is essential for an accurate translation initiation³⁰. The codon composition of the coding sequence may be also optimized, for instance, by replacing rare codons by equivalent frequently used and adapted to the cellular background²⁰⁶, and to avoid internal translation initiated on non-AUG start codon such as the CUG start codon^{207–210}.

1.3.2 Modified nucleosides insertion into synthetic mRNA

Exogenous IVT mRNA can inherently stimulate natural viral RNA alerting defence system, by activating pattern recognition receptors (PRR). They activate interferon inducers, in particular, TLRs in the endosomes, such as TLR3 which was found to detect dsRNA while TLR7 recognizes the ssRNA^{211–214}. Other RNA sensors like retinoic acid-inducible gene I protein (RIG-I) and melanoma differentiation-associated protein 5 (MDA5), were identified as two major cytosolic sensors to contribute in the detection of uncapped mRNA (5' three phosphate) and cap0^{201,215,216}. This leads to protein expression alteration and suppression through several mechanisms, such as ribosomal stalling and RNA degradation. One effect could be mediated by the protein kinase RNA-activated (PKR), which subsequently, phosphorylate the α -subunit of translation initiation factor 2 (eIF2 α) and repress translation initiation^{217–220}. The incorporation of modified nucleotides (such as Ψ , m⁶A, m⁵C, 2-thiouridine (s²U), pseudouridine, or 2'-O-Me in mRNA transcript was reported to significantly reduced the immunogenicity¹⁷⁸. Modified *in vitro* transcribed mRNA has been pivotal in the RNA-based research and paved the way for investigations about modification effect on translation decoding mechanism and potential and potential use as vaccines. The application of pseudouridine-modified mRNA was initially reported by Kariko and colleagues using fully Ψ -modified mRNA²²¹ and also by partial substituted Ψ –modified mRNA by other groups^{222–224}. The

insertion of modified nucleoside triphosphates could be processed *via* multiple strategies and depending on the desired focus. Kariko and colleagues addressed the synthesis of modified mRNA by direct random incorporation of modified nucleotide triphosphates during *in vitro* transcription. In a more targeted insertion, Meyer and colleagues harnessed the use of a T7 RNA polymerase to insert an m⁶A at the 5' end of the sequence using N⁶-methyladenosine 5'-monophosphate (m⁶AMP) ^{132,225}.

In a part of the study about the effect of inosine on the translational machinery, Licht and colleagues used a DNA template designed in a way that enables the *in vitro* transcription of an RNA containing a single inosine codon in the sequence. The sophisticated design consists of the addition of ITP instead of GTP during IVT. The encoded short peptide allowed a functional study of the effect of inosine to induce context-dependent recoding and translational stalling ²²⁶.

Modified RNA can be generated by chemical synthesis to create site-specifically modified RNA, however, this method is limited to short RNA (<40 nt) ²²⁷. For longer RNA molecules, RNA ligation was the most described technique to join together synthetic or *in vitro* transcribed RNA oligoribonucleotides with the aid of DNA splint to bring closer the RNA ends ^{228,229}. Originally this method was used to show the importance of 2'-hydroxyl at the 3'splice site for the splicing process. Thereby, a nuclear pre-messenger RNA substrate was created wherein the 2'-hydroxyl group at either splice site was substituted with a single hydrogen or 2'-O-methyl group ²²⁹.

1.3.3 Applications of modified mRNA

Advances in the delivery formulation to protect mRNA cargo in parallel to the improvement of the transcript's structural body by introducing modified nucleosides extended the IVT mRNA scope toward clinical trials. Conceptually, the antigen of choice from the pathogen target is identified, the sequence is cloned into the DNA template, finally, mRNA is transcribed *in vitro*, and the vaccine is delivered into the subject. As a result, for the continued efforts, synthetic mRNA is being explored for different fields of application. The IVT mRNA itself can be also harnessed as an added intrinsic adjuvant ^{230,231}. Indeed, unmodified IVT mRNA was reported to stimulate human TLR7, and TLR8, while the insertion of certain modification (for instance pseudouridine, m⁵C and s²U) diminished the RNA ability to induce them ¹⁷⁸. This finding was pivotal for further applications of mRNA. It was reported that a more significant improvement in terms of increasing protein expression could be obtained by

combining together two modifications, the m⁵C and Ψ in mRNA. An example of application was described in cellular reprogramming in fibroblast to induce pluripotent stem cells ²³². Another application was done by the administration of m⁵C/s²U-modified mRNA encoding surfactant protein B (SP-B) to treat lethal con-genital lung disease in mice lacking the latter protein ²²⁴. The twice weekly administration of modified-mRNA protected the mice from respiratory failure and prolonged the average life span of the animal with a necessity for a continuous treatment.

Extended studies showed that the introduction of *N1*-methyl pseudouridine outperformed pseudouridine in cell-based and mice experiments ^{219,233,234}. Deploying modified-mRNA vaccination started to take a part in epidemic and pandemic diseases caused by viral infections. A few years ago, the Zika virus has emerged as a pandemic manifested by neuropathology in new-borns and adults. A study reported the development of modified mRNA encoding the pre-membrane and envelope (prM-E) of glycoproteins of a 2013 Zika virus, to support immunity against the Zika virus in mice model ²³⁶. The administration maintained immunization from 3 months to 5 months in mice and macaques ^{235,236}. Pardi and colleagues described the administration of m¹Ψ-containing mRNAs encoding the light and heavy chains of chains VRC01 (an anti-HIV), protected mice against intravenous HIV challenge ²³⁷. Upon the pandemic of *SARS-CoV-2* virus, causing the coronavirus disease 2019 (COVID-19), in the world wide starting late 2019, an mRNA-based vaccine candidate (BNT162b2) was suggested by BioNTech and Pfizer. It encoded the trimerized receptor-binding domain spike glycoprotein of SARS-CoV-2 and the vaccination consists of 2 doses administrated 3 weeks apart ²³⁸.

1.4 Circular RNA

1.4.1 Biogenesis of circular RNA

Circular RNA, or simply circRNA having a size range varying between 100 bp to 4 kb, in eukaryotes (reviewed in ²³⁹). Their 5' and 3' ends are engaged in a covalently closed structure, which confers them protection against exonucleases ^{240–242}. The half-life of the endogenous circRNA was found to be at least 4 fold higher than that of linear mRNA. In eukaryotes, their abundance was estimated as 0.1–10% of their linear counterparts ²⁴³. The first hints about circRNA emerged on the early nineties in human ^{244,245}, and mouse *Sry* genes ²⁴⁶. Nevertheless, the lack of adapted analysis methods hampered further investigations. For years, circRNA were more perceived as rare events or artifacts of RNA splicing ^{241,245–247},

since introns lariats can evade debranching and degradation steps during the splicing process^{248–250}. Thus, they were presumed unlikely to play a significant biological role. Furthermore, circRNAs were found to encode for viral genome^{251,252}, and can be generated as an intermediate reaction during rRNA and tRNA processing^{247,253}.

The RNA circularization was reported to arise from back-splicing^{254–259}, an alternative for linear RNA splicing. Broadly speaking, canonical eukaryotic pre-mRNA splicing is catalyzed by the spliceosomal machinery to remove introns and link exons, wherein an upstream 5' splice donor site is joined to a downstream 3' acceptor site. Meanwhile, in back-splicing, a downstream 5' splice donor site is linked to an upstream 3' splice acceptor^{245,246,260,261}, leading to an RNA circle ligation by a 3'-5' phosphodiester bond at the splice junction site (BSJ). Recent research revealed that the back-splicing process benefited from the canonical factors of the spliceosomal machinery^{256,262} and that both processes were coupled and coordinated²⁶³. Consequently, circRNA could originate from exons (exonic circRNA), introns (intronic circRNA) and exon-intron circRNAs²⁶⁴, with a predominance detection of exonic circRNA^{265,266} (Figure 1. 10). Different models were suggested to generate circRNA, i.e., intronic circularization intermediate lariats or exon skipping (reviewed in²⁶⁷) originate during canonical splicing of linear RNA, and exon circularization occurring by permuted intron-exon (PIE) sequence, based on the autocatalytic ribozyme activity of the intron I group to form exonic circRNA²⁶⁸. The abundance of circRNA showed the existence of positive and also negative correlation relative to their respective linear transcripts^{243,258,261,269,270}. This relationship may dependent on the cell-type and the cellular metabolic conditions^{243,259}.

On a more fundamental level, the mechanism regulation by which the spliceosome machinery to proceed to back-splicing over linear splicing, or *vice-versa*, is largely unclear. Recent advancements revealed the presence of *trans*-factors and *cis*-elements that could promote back-splicing, such as the intron sequences flanking the exon(s)-forming circRNA (Figure 1. 10)²⁷¹²⁷². Besides their unusual length²⁶¹²⁶⁵, these sequences can base pairing across introns or within the directly flanking involve complementary sequences (i.e. ALU repeats), that can be in inverted orientations upstream and downstream, facilitation a back-splice junction and the exon circularization^{254,261,263} (Figure 1. 10).

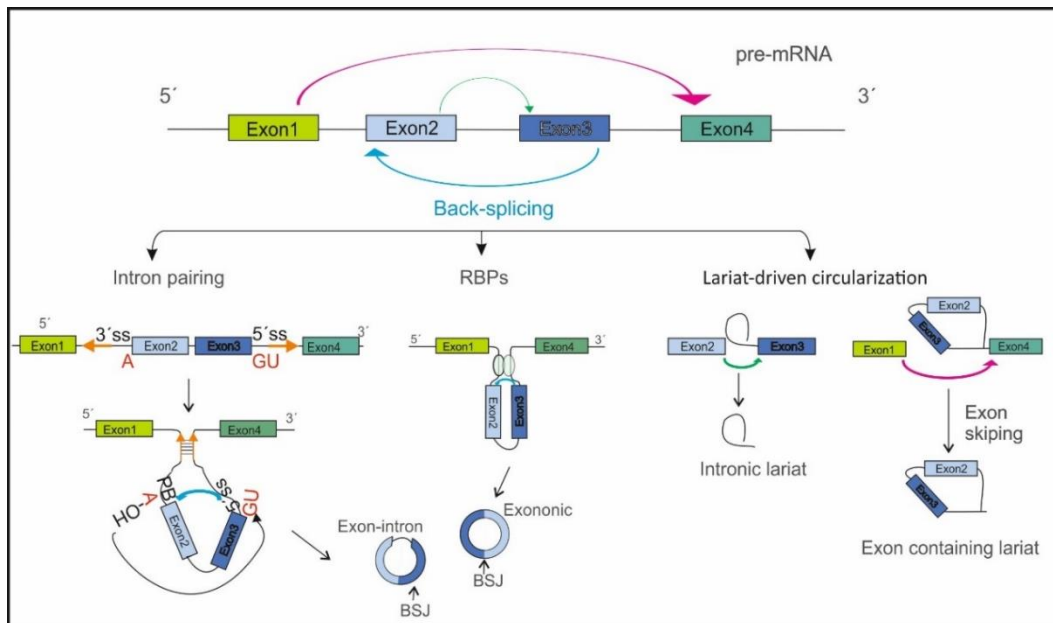


Figure 1. 10: Overview of circular RNA biogenesis.

The generation of circular RNA can be mediated by a non-canonical splicing pathway (back-splicing, arrow in pink). The mechanism can be guided by intron pairing flanking the circularized exons or by RNA-binding proteins (RBP) that interact with the specific motifs in the flanking introns to bring closer exons and create a back-splice junction (BSJ). The resulting circular RNA can be purely exonic or can retain an intron. During pre-mRNA canonical linear splicing (arrow in green), removed intron can be circularized in a lariat form generating circular RNA, and exon-skipping events (arrow) may generate exon-containing lariats. Figure adapted from reference ^{417,418}.

Another pathway was shown to be mediated by an RNA-binding protein. In their study, Ashwal-Fluss and colleagues reported that the circularization of an exon can compete with linear pre-mRNA splicing ²⁵⁴. They identified the *muscleblind* protein (MBL) that can bind to MBL binding sites localized on the circMbl flanking introns, which allowed to bring closer the splice signals and promoted circRNA formation. Very recently, the biogenesis and the accumulation of a circRNA subset was reported to be modulated by the presence of an m⁶A modification ²⁷³.

1.4.2 Detection and characterization of circular RNA

In the early nineties, initial studies reported that circRNA could be visualized by electron microscopy ²⁷⁴. Several biochemical approaches for a comprehensive evaluation of circular

RNA species were developed (reviewed here ²⁷⁵). Initial studies have been focusing on the separation of circular RNA viroids using polyacrylamide gel electrophoresis (PAGE), based on the distinct electrophoretic mobility of the circular form from the linear one ^{276,277}. Different protocols involving one and two dimensions electrophoresis were reported, i.e. a native first dimension (1D) followed by a denaturing 2D or running under denaturing conditions for both system while optimizing running parameters, i.e. the ratio of the N,N'-Methylene-bis-acrylamide (gel cross-link) or the concentration of the gels ^{277, 278}. In their study using circular RNAs synthesized *in vitro*, Feldstein and colleagues revealed the separation of the circular RNA based on their poor migration through highly cross-linked gels relative to less cross-linked gels ²⁷⁸. An alternative assay relied on a weak hydrolysis, using enzymatic RNase H degradation or chemical hydrolysis ^{259,279,280}. Indeed, such degradation was more evident for circRNA, as the introduction of a single nick could break the circle and generate a single linear molecules with the expected size ²⁸¹.

Traditional RNAseq pipelines made circRNA out of the profiling process ²⁸² due to the absence of a linear conformation and a polyadenylated tail at the 3' end. Improving the selection criteria by depleting both polyadenylated RNAs and rRNAs, and the enrichment for the circular RNA with RNase R allowed the identification of intronic circRNA as well as exonic circRNA, with a discrimination of lariats and trans-spliced species (based on the BSJ sequence) ^{258,261}. Indeed, the approach focused on back-spliced (BSJ) sequence reads to screen and predict circRNA. Direct detection of circRNA by quantitative PCR could be allowed by primers in divergent orientation to amplify the BSJ site, followed by gel analysis and could be verified by Sanger sequencing.

Systematic prediction using a computational method to validate the read-mapped regions that contained splicing junctions delineated that (i) they are expressed across species of all kingdoms of life, i.e. in mouse, *Caenorhabditis elegans*, *Drosophila*, humans ^{254,258,259,283,284} and in archaea ²⁵³, and (ii) despite their a low level abundance in different species ^{258,259,261,265,285}, they may constituted up to 15% of the actively transcribed genes ²⁶¹.

1.4.3 Biological implications of circular RNA

Recent studies have found that circRNAs exhibited a cell type- or tissue-specific expression ^{243,259,269} and could be expressed in a developmental stage-specific manner ^{259,286}. In addition, since some circRNAs could be much higher expressed as compared to their respective linear mRNAs ^{258,287}, a potential role was associated to pathological initiation and progression

in many diseases, as in cancers^{243,288–291}, for instance in hepatocellular carcinoma²⁹², human atherosclerosis^{265,287}, neurological diseases^{269,293}, Alzheimer²⁹⁴, and cardiac hypertrophy (reviewed²⁹⁵). CircRNAs were also found to be remarkably stable in the body fluids which suggested them as a prognostic biomarker for standard clinical blood samples^{296,297}. Recent studies have indicated that certain circRNAs contain multiple binding sites of miRNA and acting as “sponges” to store, modulate their level and translocate them to subcellular locations^{265,298, 299}. The CDR1as exonic circRNA RNA was one of the first circRNAs to be characterized and was shown to contain more than 70 binding sites for a micro RNA (miR7)²⁵⁹. This interaction did not trigger the circRNA cleavage but was suggested to rather a sponge to stabilize and transport miRNAs to neurons³⁰⁰. Besides to sponging miRNAs, some circRNAs can interact with proteins. They can regulate their activity, by promoting for instance the enzymatic activity as described with the CircACC1 which was identified to stabilize and promote the enzymatic activity of the AMP-activated protein kinase (AMPK) having a role in metabolic homeostasis³⁰¹. In their study about myotonic dystrophy, Ashwal-Fluss and colleagues reported that *muscleblind* protein (MBL) promoted circMbl generation. This latter contained several MBL binding sites suggesting possible regulation of the excess of MBL proteins by “sponging” them and thus reducing their availability²⁵⁴. Another role was assigned to circRNAs to modulate the transcription and promoting transcription of their parental genes. Li and colleagues reported a network interaction on one side between circRNA with U1 snRNP through RNA–RNA interactions, while the formed complex may further interact with the Pol II transcription complex at the promoters²⁶⁴.

Furthermore, naturally occurring circRNAs could also be identified to derive from actively coding genes and accumulated in the cytoplasm^{261,265}. Emerging evidence reporting the presence of an AUG translation initiation codon in some endogenous circRNA and their association with polysomes supported their protein expression capacity^{302,303}. The circSHPRH was reported to encode a short peptide 146 amino acids, which in turn, is involved in tumor repressor activity in human glioblastoma disease³⁰⁴. Several studies published as well predicted circRNA with coding potential related especially to cancer (reviewed in^{287,290}), such as in hepatocellular carcinoma³⁰⁵ and breast cancer³⁰⁶.

1.4.4 Synthetic circular RNA

To generate of synthetic circRNA, several groups relied on natural biogenesis pathways *in vitro* and *in vivo*. One strategy was described based on self-splicing-induced introns circularization. An initial report by Puttaraju & Been described the generation of circular RNA by permuting the order the exon and the group I intron from both the *Anabena* tRNA intron and the *Tetrahymena* intron generated a circular RNA exon *in vitro* ²⁶⁸. The design of the permuted intron-exon (PIE) consists of an end-to-end fused exon which are placed in half intron sequences, which self-splice to release circRNA. Thus, several reports applied this strategy by replacing the exon by another sequence, to produce circRNA *in vivo*, mostly in *E.coli* and less examples in yeast ^{307–309}. While PIE was previously reported for short RNA circularization, it was recently shown that it could be successfully applied to generate circRNA up to 5 kb in length *in vitro* ²⁸¹. This required the introduction of some structural optimizations, such as the addition of homology arms at the 5' and 3' ends of the precursor RNA to enhance further the spatial proximity of 5' and 3' splice sites ^{310,311}. Based on induced intron-pairing, by forming a double-stranded RNA structure to bring in proximity the splice sites, studies reported the generation of coding and non-coding circular RNA ^{312,313}. Circularization could be also performed *in vitro* using an enzymatic reaction. In this case, the RNA precursor could be transcribed *in vitro* then subjected to circularization by complementary annealing to an RNA or DNA “splint”, bringing together both the 3'phosphatase and 5'OH ends. Circular products could be then obtained by incubating with a ligase, T4 RNA ligase I or DNA/RNA ligase II ³¹⁴. Other strategies could be also adapted for circularization (review in ³¹⁵).

1.4.5 Translation of circular RNA

An early study by M. Kozak revealed that ribosomes were unable to bind to small circular RNA in eukaryotes ³¹⁶, supporting the classification of circRNA as non-coding RNA. Moreover, if circRNA could be translated, it would not rely on the canonical translation pathway as discussed in 1.1.2. Nevertheless, the discovery of another translation initiation mechanism through IRES sequence, which was encountered mostly in virus and also in some endogenous mRNA ^{37,317}, opened new perspectives to the existence of alternative translation models. Further identification of circRNA associated with polysomes has sparked renewed interest in the study of the coding capacity of circRNA ³¹⁸. A scheme recapitulating the suggested models/pathways of translation from circRNA is illustrated by (Figure 1. 11).

1.4.5.1 IRES-driven translation

In one of the best characterized examples, Legnini and colleagues reported that circ-ZNF609 was associated with heavy polysomes and expressed protein. The translation was shown to be led by a specific features of the UTR sequence. Indeed, when this sequence was inserted in a bicistronic linear RNA construct, translation was allowed mimicking an IRES-driven translation. However, the synthesis of a circ-ZNF609 by *in vitro* ligation could not express proteins, suggesting a translation mechanism that requires the formation of circRNA in a splicing-dependent manner *in vivo* ³¹⁹. Moreover, circ-ZNF609-encoded a protein that lacked two zinc-finger domains, usually contained by the zinc-finger protein 609, proposing a possible antagonistic (as dominant-negative) or cooperative contribution activity.

Another example have been proposed by Pamudurti and colleagues, who reported a cap-independent translation from circRNA generated from the *muscleblind* locus, expressed in fly head extracts. In this case, the translation was also promoted by the presence of particular UTR elements (cUTRs). Interestingly, using a bicistronic reporter constructs and an engineered circRNA containing the cUTRs, they showed a resistance to inhibition of cap-dependent translation by the circRNA, while this was not observed when cUTR was inserted in an RNA linear version ²⁸⁵.

During the last two decades, the interest to understand the translation mechanism mediated by cellular circRNA was facilitated by engineered circRNAs. Chen and Sarnow reported one of the first studies on the ability of a synthetic circRNA to drive translation *in vitro* ³²⁰. They used a splint ligation strategy to circularize an RNA containing an IRES upstream of a green fluorescent protein (GFP) open reading frame. Later on, independent studies provided further evidence on the translation capacity of engineered circRNA *in vitro* and *in vivo*, in the absence of translation elements (i.e. IRES) ^{303,312}.

1.4.5.2 *m*⁶A-driven translation

Genome-wide methylation mapping pattern revealed a subset of endogenous circular RNA enriched with *m*⁶A methylation sites ³¹⁸. Moreover, it has been found that the *m*⁶A reader (YTHDF1/YTHDF2) and writers (METTL3/ METTL14), employed by linear mRNA, were shared with those of *m*⁶A-circRNA ³²¹. Studying the ability of *m*⁶A-dependent translation activity from circular RNA form, Yang and colleagues reported the presence of a single *m*⁶A motif upstream of the coding region in a synthetic circRNA that was sufficient to promote translation *in vivo* which could be modulated by demethylase FTO and the *m*⁶A writers ³¹⁸. The

best described example of naturally occurring circRNA was the circ-ZNF609 which was found also methylated and served as a study model by Di Timeotio and colleagues^{55,273}. In a continuous work to their previous publication⁵⁵, they reported that translation was m⁶A cap-independent and was not sensitive to the eIF3, in contrast to earlier publication with m⁶A in linear RNA^{132,133}. In addition, the translation could be modulated by the physical interaction between YTHDF3 and eIF4G2, consistent with the findings obtained in the earlier mentioned report on synthetic m⁶A-circRNA. They also revealed that the knock down of METTL3 decreased the ZNF609 circRNA accumulation and that YTHDC1 was also required to promote circRNA production, which suggested the role of m⁶A modification as a key regulator in circular RNA expression and production promotion.

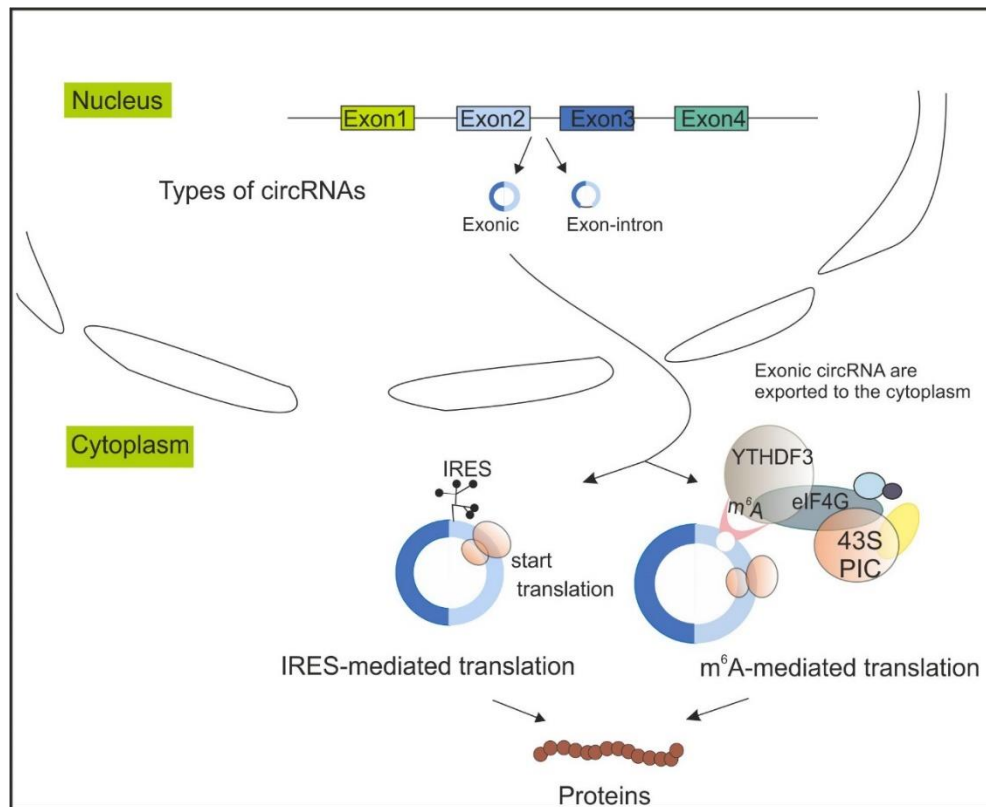


Figure 1. 11: Translation of circular RNA.

Potential models of translation of circular RNA in eukaryotic cell. The exon-intron circRNAs cumulate in the nucleus while Exonic circRNAs are exported to the nucleus. Models of translation initiation in circRNA proposed a translation pathway to be mediated by Internal Ribosome Entry Site (IRES) - and N⁶-methyladenosines (m⁶A)-mediated cap-independent translation initiation. The latter mechanism is likely promoted by YTHDF3 and eIF4G2. Figure adapted and modified from reference²⁹⁰.

2 Motivation and objectives

Eukaryotic mRNAs have been reported to harbor numerous modifications, with a potential involvement in important mRNA biological pathways. In nature, the installation of certain modifications was reported to be regulated and site-specific. The presented study here aimed to generate mRNA carrying naturally occurring modifications to study their implication on protein expression. As study model, investigating the impact of modifications in the context of IRES-driven translation was of interest, since there is a lack of data to this regard. One employed strategy here relied on the capacity of T7 RNA polymerase to introduce modified nucleoside triphosphate during *in vitro* transcription. Although being straightforward, it lacks the control on the precise incorporation site and is limited to modifications whose triphosphates can be catalyzed by T7 polymerase. Therefore, it was of interest to engineer point-modified mRNAs harbouring modifications to target critical positions in the mRNA sequence, such as the Kozak sequence motif, the start AUG codon, and as well as in other codons from the coding sequence. This was achieved by joining together three fragments to reconstitute one large point-modified mRNA, using a one-step enzymatic ligation that was initially validated to join multiple short RNA fragments. Besides the point-modified mRNA synthesis, optimizing a purification method was also considered to ensure the ligated mRNA quality (i.e. reducing gel impurities), which would be required for certain downstream applications, i.e. injection of RNA on a HPLC column.

Synthetic randomly modified mRNA was also harnessed in another topic, to address the efficacy of the immunoprecipitation sequencing (MeRIP) experiment, since it was challenged by antibodies selectivity leading to off-target bindings.

Assessing the influence of the RNA form on translation was also investigated by generating circular modified and unmodified RNA, using enzymatic ligation. For the validation and purification of synthetic circRNA, available methods and techniques are traditional and laborious. Most of the published data are based on gel electrophoresis or also enzymatic digestion. Establishing an alternative method was addressed here, for a direct separation and elution of circRNA *via* RP-HPLC. In a future experiment, circularization of point-modified mRNA was intended to provide a closer insight to the effect of modification according to its position within the RNA sequence, and would elucidate the mechanism behind it.

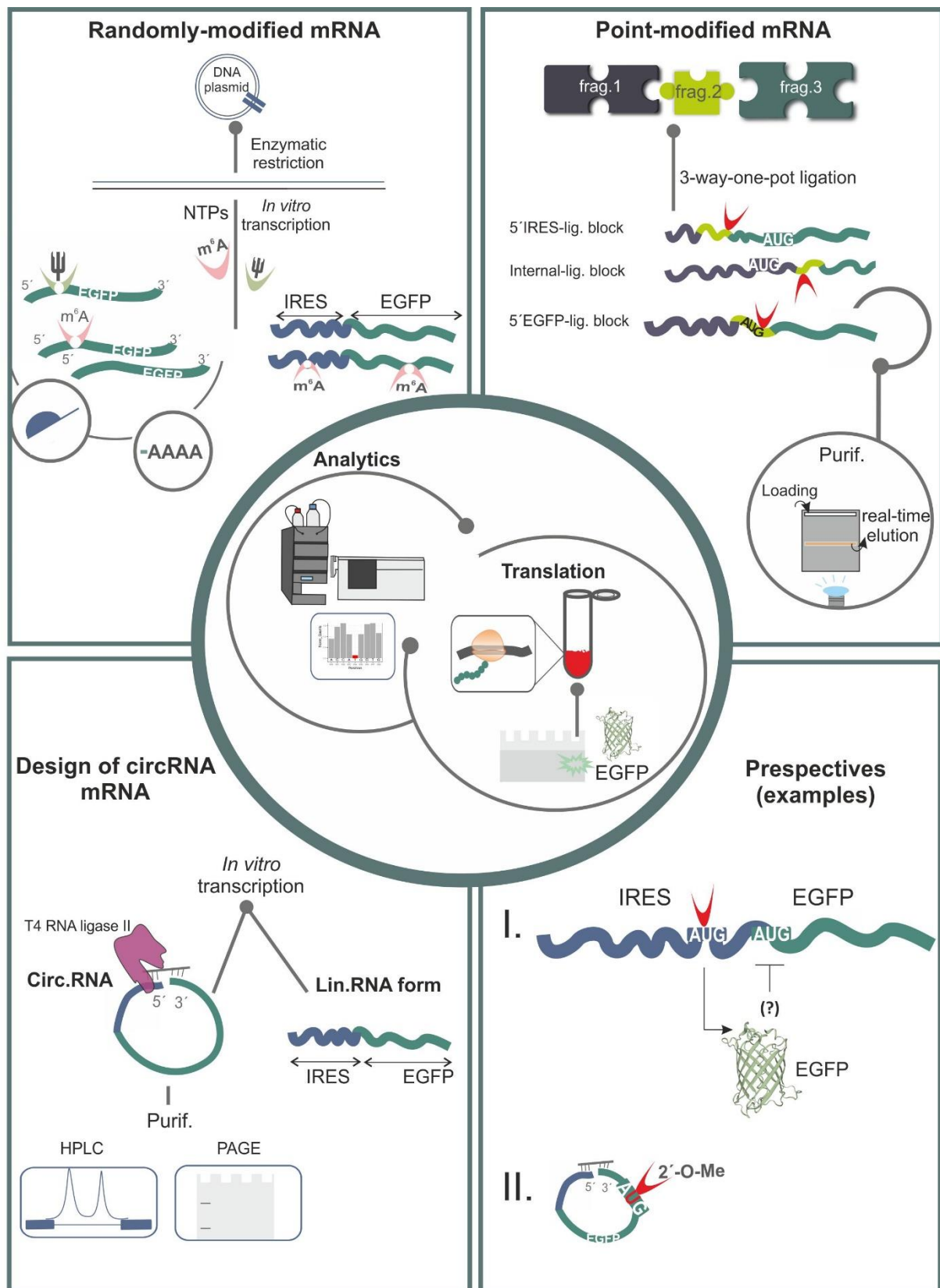


Figure 2.1: Global overview of the present work and possible perspectives.

3 Results and discussion

3.1 Engineering of nucleoside-modified mRNA

Modifications such as *N*⁶-methyladenosine (m⁶A)^{85,112}, 2'-O-methylation (N_m, 2'-O-Me)⁹¹ and pseudouridine (Ψ) were mapped, among others, in the coding sequence of mRNA (Please refer to Figure 1. 6 in the introduction). This suggests a putative implication and tuning of the protein expression. However, their physiological role has not been completely defined.

To elucidate their roles in protein synthesis, various strategies were developed to introduce modifications into artificial mRNA. Indeed, the insertion of various modifications, in particular Ψ and m⁶A, was shown accessible by the utilization of T7 RNA polymerase¹⁷⁸, while certain modifications, such as ribose methylations, are not good T7-substrate and can terminate run-off transcription^{178,322}. This approach allows a certain control regarding the choice of the nucleoside triphosphate to be substituted, and full or partial substitution by the equivalent modified one.

Despite this benefit, the incorporation remains random without a control over the exact position within the mRNA sequence. Alternatively, splint ligation can be employed as an approach to insert the desired modified nucleoside triphosphate in RNA, in a target-specific way^{323,324} called here “point-modified mRNA”.

Harnessing both approaches, the following work aimed to generate, in a first part, a series of IRES (internal ribosome entry site) and non-IRES EGFP mRNA, bearing randomly incorporated Ψ and m⁶A. In a second part, the preparation of single-point modified IRES-EGFP mRNA is described.

3.1.1 Generation of randomly-modified mRNA

In this work, a set of cap-dependent as well as cap-independent mRNA were synthesized.

For cap-independent translation, the EMCV IRES was implemented upstream of the open reading frame (ORF) of an EGFP. It is one of the common widely used IRES element for its ability to mediate accurate and stable translation in different cell lines^{325,326}. Because there are various mutated versions of EMCV IRES available in literature³²⁵, it should be mentioned here that the IRES sequence was conserved with exception of mutating the start codon 11 (among

other two start codons 10, 11 and 12³²⁷) to allow translation using the next start codon (start codon 12), which was located in a strong Kozak sequence upstream of the EGFP ORF.

Another especial consideration applied to the template design, was the insertion of a T7 RNA polymerase promoter. This enzyme was shown able to synthesize modified RNA including functional groups, i.e. alkyne or azido groups for click reactions^{328–330} and base modifications, i.e. m⁵C, m⁶A, m⁵U, and s²U¹⁷⁸. Interestingly, such capacity can be accompanied with a reduced T7-catalytic efficiency with for instance m⁶A modification¹⁷⁸. The size of the modified substituent groups is of matter and bulky groups, i.e. dibenzofuryl derivatives of ATP and GTP or 2'-O-Me could not be catalyzed by the T7 RNA polymerase^{322,331}. This would presumably lead to an unequal incorporation efficiency of modified *versus* unmodified nucleotides.

Here, either adenosines or uridines were replaced with m⁶A and Ψ respectively, by incorporating the modified nucleotides during *in vitro* transcription. The ratio of one particular modified NTP relative to the corresponding unmodified NTP were adapted to estimate an incorporation percentage of: 2.5%, 30%, 60% and 100% relative to unmodified respective NTP). After transcription, the full-length transcripts were analysed by capillary electrophoresis (Figure 3.1A) showing that modified mRNAs migrated similarly to their unmodified counterparts, based on their size.

The assumed incorporation ratios were confirmed by the application of modified as well as unmodified mRNAs for analysis on the nucleoside level. After mRNA digestion, the resulting nucleosides were applied for quantitative HPLC and MS analysis. The estimated Ψ nucleoside (Ψ:U) content was nearly proportional to the expected ratios added to the *in vitro* transcription (IVT) experiment.

When introducing m⁶A during IVT, the incorporation was weak. In fact, with high incorporation percentages, the measured ratios by HPLC and MS were comparable to the ratio of the m⁶A:A present in the reaction pool. However, at low m⁶A percentages, quantification could be mostly achieved by MS only. Moreover, comparing data from m⁶A incorporation in naked EGFP mRNA (>800 nt) and the IRES-EGFP mRNA (>1300 nt), suggests that the incorporation was independently from the RNA sequence context or its length (compare m⁶A Figure 3.1B, Figure 2B). Such a weak incorporation was also observed in previous work by Karikó and colleagues¹⁷⁸.

The quantification of the modification by MS compared to the expected values are recapitulated in Figure 3.1C.

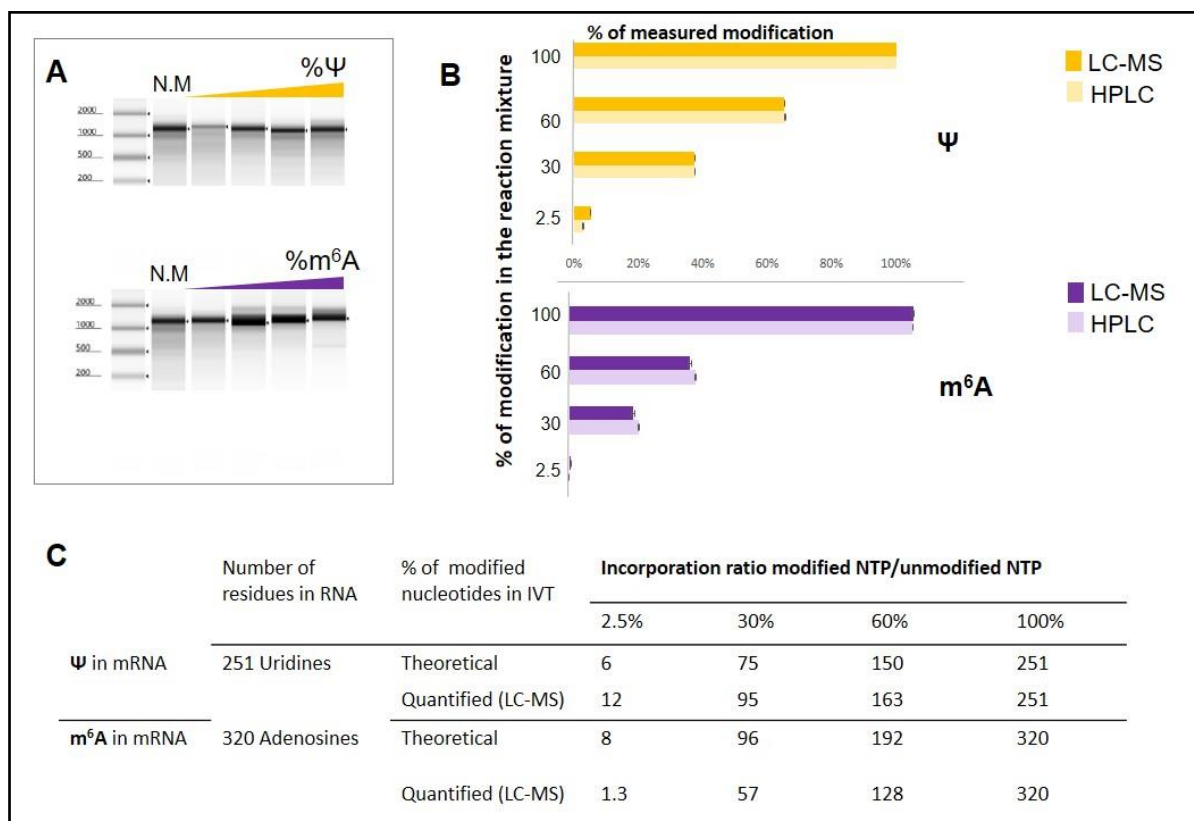


Figure 3. 1: Synthesis and analysis of modified IRES-EGFP mRNA.

(A) *In vitro* transcribed IRES-EGFP mRNA containing increasing ratios (2.5%, 30%, 60% and 100% of (Ψ:U) or (m⁶A:A) were separated on capillary electrophoresis (Agilent, TapeStation). Unmodified mRNA serves as a reference (N.M.). (B) Transcripts were digested and subjected to LC-MS quantification of the modified nucleosides. No corresponding modification was detected in the unmodified transcripts (not shown). For absolute quantification stable isotope-labelled nucleosides (¹³C) from *S. cerevisiae* were used as an internal standard and mixed in an equal amount with the analyzed sample. Error bars are shown, indicating \pm SD of technical triplicates. (C) Determination of the number of modified nucleoside content per mRNA molecule. The theoretical percentages correspond to the calculated number of m⁶A or Ψ based on their expected amount incorporated in the transcription reaction relative to the number of their unmodified NTP (320 A and 251 U) per mRNA molecule with a length of 1369 nt. The detected values correspond to the measured amounts determined by LC-MS, relative to the unmodified NTP derivatives.

Another set of EGFP mRNA was prepared without using an IRES. In this case, additional structural elements were added to generate either capped-mRNA, capped/poly(A)-mRNA, poly(A)-mRNA or naked mRNA (without neither a cap and nor a poly(A)). In addition, modified m⁶A and Ψ were as well incorporated during IVT at different ratio (2.5% 30%, 60% and 100% relative to the unmodified nucleotide). Interestingly, the quantification by LC-MS shows again that Ψ was behaving as a uridine substrate and was almost efficiently incorporated by T7 RNA polymerase, while this was not the case for the m⁶A incorporation (Figure 3. 2A).

Of the two possibilities to introduce a cap structure we opted for the enzymatic posttranscriptional *in vitro* capping using *vaccinia* capping enzyme, and more details are developed in the introduction part (section 1.2.3). This allowed a one-step reaction combining the latter enzyme (to achieve a cap0 structure) with a cap 2'-O-methyltransferase resulted in a final cap1 structure ³³², under experiment conditions provided by the manufacturer's recommendation and previous data from the Helm lab. Capped mRNAs were applied to LC-MS investigation for the detection of m⁷G and 2'-O-methylguanosine. As displayed by Figure 6.1, both modifications were detected. However, the quantification analysis showed sub-stoichiometric amounts of the concerning modifications, hinting at only partial formation of cap1. Since the reaction conditions were already optimized by a former colleague (Dr. Isabell Hellmuth) ³³⁰, it could be possible that results obtained by LC-MS reflected rather the difficulty of the nuclease P1 enzyme to hydrolyze the 5'-5' triphosphate bond and would result in incomplete digestion ³³³.

Similarly, the 3' end of the purified mRNAs was also processed by the addition of a poly(A) tail employing an *E. coli* Poly(A) Polymerase ³³⁴. Applying the lowest enzyme concentration suggested by the manufacturer with varied incubation times did not result in an extension on the 3' end (data not shown). Therefore, the enzyme concentration was optimized and the resulting reaction was analysed by microscale thermophoresis (MST). Such technique can be used to study nucleic acid hybridization by assessing their movement in a temperature gradient ^{335,336}. The presence of the newly synthesized poly(A) tail could be then evaluated by MST, by increasing concentrations of the polyadenylated RNA, while the amount of a fluorescently labeled poly(dT) oligo (complementary to the poly(A)) tail was constant (Figure 3. 2 B). Hybridization of the probe to polyadenylated RNA resulted in drastically changed thermophoretic behaviour, which was used to obtain dose-response curves for the RNA titration series (Figure 3. 2 B and C). Compared to the negative control (without a poly(A)), the addition of 15 units (U) or 20 U enzyme allowed similar binding to the probe (Figure 3. 2 C). According to the manufacturer, applying 15 U enzyme for the tailing reaction would generate an approximately 170 poly(A) tail length, and this concentration was used for further experiments. The addition of the poly(A) tail to the other mRNA constructs (modified or not with m⁶A and Ψ) generated similar MST profiles (Figure 3. 2 C). Overall, the poly(A) tail distribution was very homogenous for m⁶A-modified RNA and slightly more diverse for Ψ-modified constructs.

After tailing reaction, the produced RNA was also evaluated by application in capillary electrophoresis (TapeStation), which clearly showed a migration shift after the addition of the

poly(A), ensuring the homogeneity of the polyadenylated mRNA species (Figure 3. 2 D). It is to note also that when samples from the polyadenylation reaction were loaded without prior purification, the running on TapeStation was not based on the RNA size as compared to traditional agarose electrophoresis (please compare the migration on both gels in Figure 6. 2). Hence, the purification of the sample should be considered.

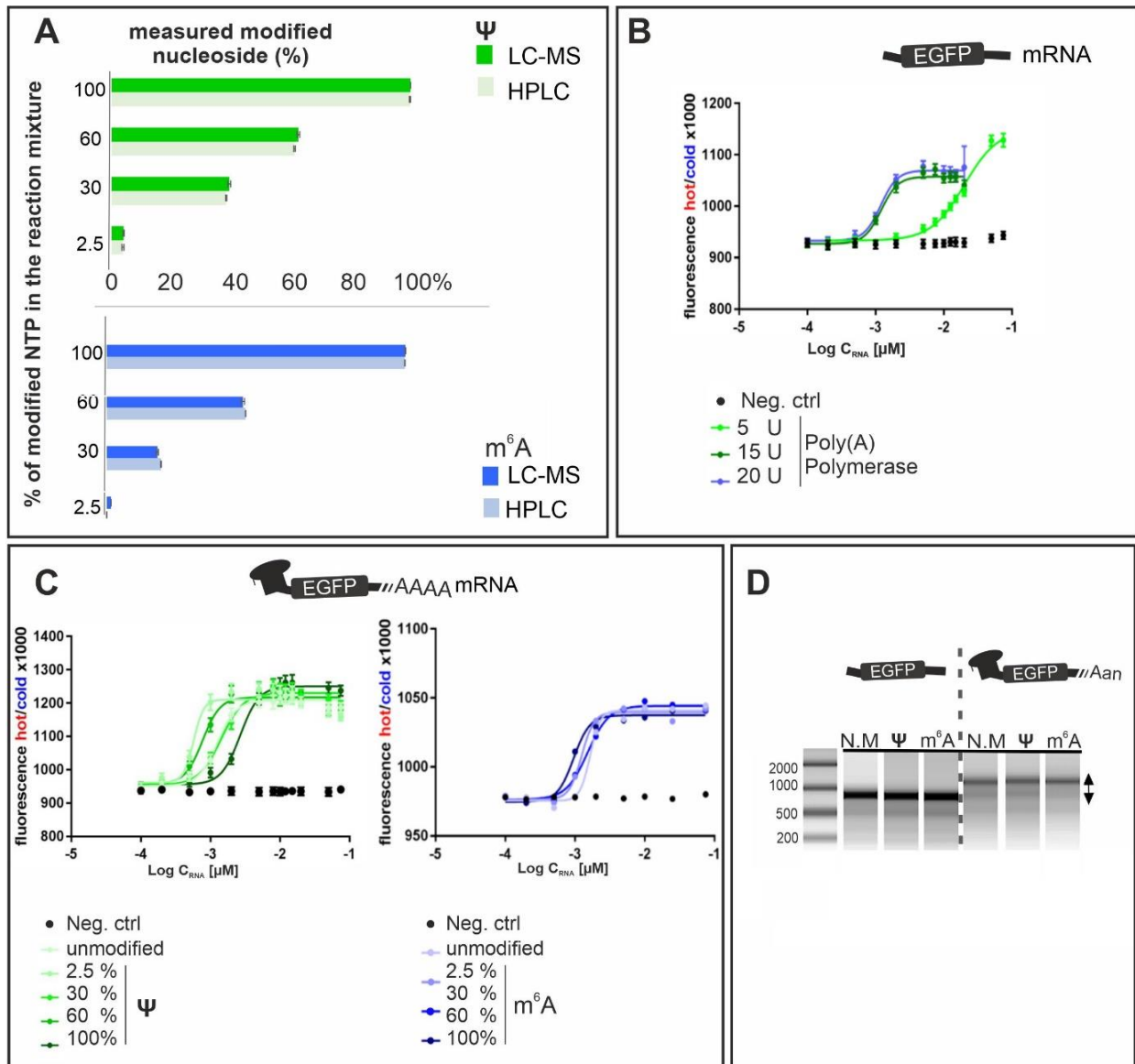


Figure 3. 2: Engineering and synthesis of modified, capped and polyadenylated mRNA.

(A) Base modification analysis by LC-MS. Modified mRNA transcripts were applied to LC-MS for the quantification of the incorporated modified nucleotides relative to the total quantified amount of respective unmodified and modified nucleoside. No corresponding modification was detected in the non-modified transcripts (not shown). For absolute quantification, stable isotope-labelled nucleosides (¹³C) from *S. cerevisiae* were used as an internal standard and mixed in an equal amount with the analysed sample. Error bars indicate ± SD of technical triplicates. (B) Binding curve generated from the fluorescence intensity of the fluorescent oligo(dT) bound to the newly synthesized poly(A) tail. Polyadenylation was realized using increasing amounts of *E. coli* poly(A) polymerase (5, 15 and 20 units). In black is the negative control without a poly(A). Optimal enzyme concentration was used to generate poly(A) tail in capped m⁶A and Ψ-modified mRNA (C). As controls, tailing on capped unmodified mRNA as well as on naked mRNA (negative control) were used. (D) Modified EGFP mRNA (without an EMCV IRES) was synthesized by implementing modified nucleotides in the *in vitro* transcription mixture to the corresponding ratios 2.5%, 30%, 60% and 100% relative to their non-modified NTP equivalent. To these RNAs, an additional step to add a cap and poly(A) structure allowed to create: naked (without cap, no poly(A)), capped mRNA, capped/poly(A) mRNA and poly(A) mRNA. Aliquots of naked and cap/poly(A) mRNA either: non-modified (N.M) or 100% modified mRNA were separated on capillary electrophoresis (Agilent, TapeStation).

3.1.2 Design and engineering of large point-modified mRNA

3.1.2.1 3-way-one pot ligation concept

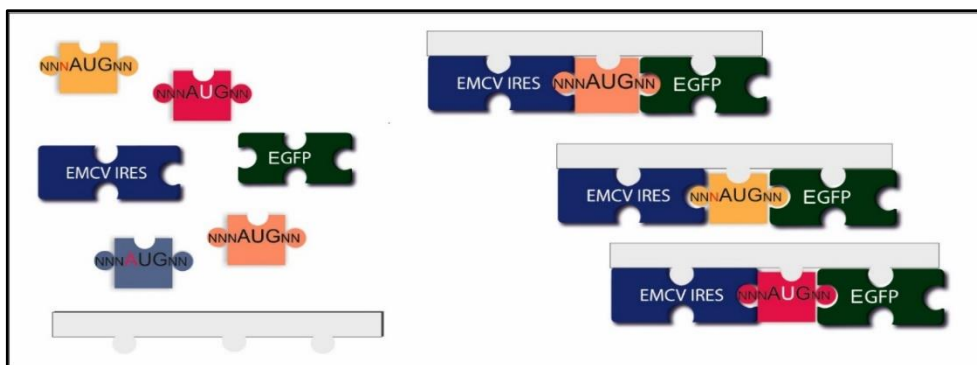
To have an overview about the impact of one single modification on mRNA translation, a large synthetic mRNA (>1300 nucleotides), was synthesized to harbour a single modification in site-specific way (called also point-modified RNA) using RNA ligation method^{132,169,205,226,330}.

Therefore, the IRES-EGFP mRNA (IRES EMCV upstream of the EGFP sequence in mRNA) was used as a study model, for two reasons: (i) it allows to circumvent the need for a capping step required for the translation, and (ii) the effect of point-modifications was not yet explored in the case of translation driven by IRES-activity.

Similar to the construction of a piece of the puzzle, RNA ligation allows here to reconstitute a whole RNA molecule by reassembling together different pieces (fragments) together (Figure 3.3). The concept idea here is straightforward and consisted first in splitting the full-length mRNA sequence into 3 main fragments considering the following: the fragment containing the modification (called fragment 2) is the one in the middle of the overall mRNA sequence and was synthesized chemically by solid-phase. Therefore, its length was limited to 20-40 nucleotides³³⁷. The other two puzzle pieces flanking fragment 2 (located upstream (fragment 1) or downstream (fragment 3) of fragment 2) were produced by a T7 *in vitro* transcription (IVT).

Figure 3.3: The puzzle concept.

The pieces in green (EGFP) and in blue (EMCV IRES) play the role of a scaffold. The middle piece may hold the modification (in yellow, dark grey, red) or not (orange) on/around the start codon. Finally, the key piece (in grey), is the splint DNA that joins the three pieces (fragments) together, resulting in the creation of several RNA combinations.



Given the T7 promoter sequence (5'TAATACGACTCACTATAG 3'), the design of these two fragments should consider the presence of at least two G nucleotides at the RNA 5' end for an efficient transcription³³⁸. Finally, the full-length RNA can be reconstituted again by joining an equimolar amount of the 3 fragments in one single-step (the so-called 3-way-one-pot ligation) and allow ligation in the presence of a T4 RNA ligase II. The order and the correct orientation were assured by the presence of a splint DNA.

3.1.2.2 Generation of point-modified RNA via 3-way-one-pot ligation

In theory, the 3-way-one-pot ligation provides an open choice to virtually place the modification at any given site in an mRNA that is over 1300 nt long. The generation of such modified mRNA requires the incorporation of chemically synthesized modified RNA (fragment 2) by ligation between several other fragments. Given that the fragment 2 was designed to harbouring each time one single distinct modification implemented at different spot of the sequence (Figure 4 A), several combinations of modified-RNA constructs can be generated. The question of feasibility implies not only the ability to ligate those together, but also, to separate the final product from the unreacted fragments in the reaction mixture.

Therefore, two selection criteria were established: one regarding the efficiency of the ligation, and another one for a proper recovery method that can be evaluated in theory or tested in practice (Please refer to section 3.1.2.3).

Based on this, one construct allowing the separation of the ligated product based on the size of that is clearly larger than the other fragments was selected (Figure 3. 4 A). This construct allows

to study point-modifications around the start codon, since fragment 2 is located in the beginning of the 5' end of the EGFP ORF.

Initially, extensive optimizations of the 3-way-one-pot ligation were performed regarding, the length of the splint DNA (39, 59, 82 and 89 nts), the hybridization conditions and the final enzyme concentration. The evaluation of ligation efficiency was the best with longer splints as 82 and 89 nts DNA oligos. In addition, hybridization was performed by incubating at 75°C, followed by a cooling down, to allow the splint to hybridize together the three fragments. Ligation efficacy was comparable with the different cooling conditions: smooth cooling (during 30 and 60 minutes), by direct incubation on ice, or by cooling at room temperature for 15 minutes (data not shown). Finally, the incubation with the T4 RNA ligase II was performed at 16°C over night. Likewise, the RNA integrity was more preserved during a long incubation, despite a reduced ligation efficacy.

Here, as seen in Figure 3. 4 B and C, the ligation was efficient and reconstituted the full-length IRES-EGFP mRNA. This result validated the possibility to create long single modified mRNA *via* a splint ligation approach. Initially, ligation in one-pot reaction was reported to create smaller RNA i.e. tRNA and synthetic RNA^{324,339}. Moreover, in the seldom published examples of large point-modified mRNA, no more than two fragments were ligated together¹⁸⁵. Here, the advantage of the 3-way-one pot reaction was bypassing the stepwise ligation, through which a ligation product was first generated using two out of three fragments, to be purified, through which a ligation product is first generated from two out of three fragments, purified and used again for the final ligation step with the remaining fragment. Assuming the heavy losses after each purification step (by gel elution for instance) and the time consumption, one can imagine the high amount of starting material to be provided for the stepwise reaction.

Further characterization of the obtained ligation product was conducted using hybridization of a fluorescent DNA-probe which is complementary to fragment 2. As expected, a fluorescent signal was detected for the ligation product, which migrated on the same level to a positive control (full-length IVT mRNA) (Figure 3. 4 C, lane 1 and 5). The fluorescence was absent in the controls, consisting of single IVT fragments (Figure 3. 4 C, lanes 3 and 4). Nevertheless, fluorescence could be detected from the hybridization with reaction mixtures of either one of the IVT-fragments with the middle fragment 2 (Figure 3. 4 C, lanes 8 and 9) demonstrated as expected correct ligation orientation in splinted two-way ligations. The full-length ligated product was also applied for sequencing which confirmed the mostly the expected sequence. To note, insignificant contamination with additional nucleotides could be detected at the 3'

extremity of the transcripts. This longer extension was previously described to be added by T7 RNA polymerase³⁴⁰.

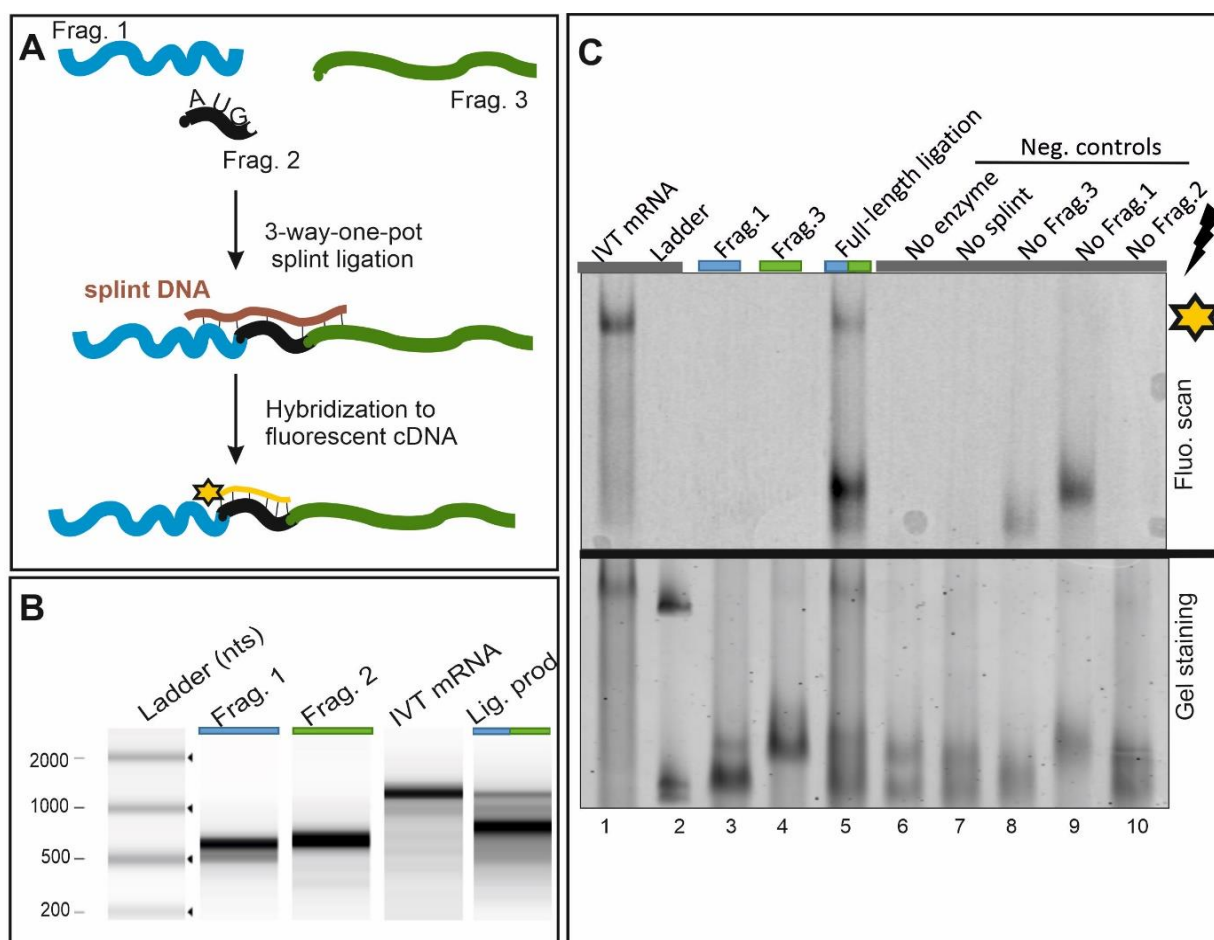


Figure 3. 4: 3-way-one-pot synthesis of large mRNA harbouring single-point modification.

(A) Concept overview of the design of splint ligated mRNA to introduce targeted modifications within the middle fragment harbouring the start codon (AUG) (Frag. 2). In blue is the IRES part and green defines the EGFP sequence. The full-length mRNA was reconstituted out of three fragments ligated together in a 3-way-one pot splint ligation with the help of T4 RNA ligase 2. After DNase I digestion, the ligation product was validated by capillary electrophoresis (TapeStation) (B), and by a 2% agarose gel. The hybridization with a fluorescein-labelled DNA which is complementary to the middle fragment was also performed (C, upper gel). After the fluorescence scan, the gel was stained with GelRed (C, lower gel). Full-length mRNA produced by *in vitro* transcription (IVT), single fragments (Fragment 1, 2 and 3) as well as ligations (lane 6-10, C) that were performed in the absence of either the splint, enzyme or one of the main fragments were used as controls (C).

Next, different modifications could be inserted into the mRNA by the 3-way-one-pot splint ligation. Using this approach render it possible to insert certain modifications (in particular 2'-ribose methylation), which cannot be catalysed by T7 polymerase *via* their triphosphates³²². Overall, ligation was as efficient as using a non-modified fragment 2 and successfully generated a long IRES-EGFP mRNA containing: Ψ and 2'-O-Me at different positions around the start codon. Correct incorporation of the modified nucleotides was verified

by LC-MS as well as the ribose-methylation specific detection by RiboMethSeq method (vide infra).

As example, the construct bearing single Um residue in the middle of the second codon GUG (pos. 641) was used also for LC-MS analysis (Figure 3. 5 A). Quantification of the amount of U_m on nucleoside level showed no significant amount of the modification in the unmodified ligation product but almost exactly one U_m per mRNA for the construct 2'-O-methylated. Besides, RiboMethSeq analysis allowed to detect one 2' O-ribose methylation at the expected position 641 (Figure 3. 5 B). Indeed, the presence of 2'-O-methylation at the ribose moiety increases the protection of the phosphodiester bond in RNA from random nucleolytic cleavage at alkaline conditions in a RiboMethSeq protocol ³⁴¹. This can be evaluated as normalized number after mapping of the sequencing reads (at 5'-end and 3'-end) to the reference sequence, resulting in a coverage profile. The presence of a 2'-O-Me is characterized by the presence of a drop in the coverage profile, resulting from cleavage protection. Other selected 2'-O-Me samples were also applied for RiboMethSeq analysis (Figure 6. 3).

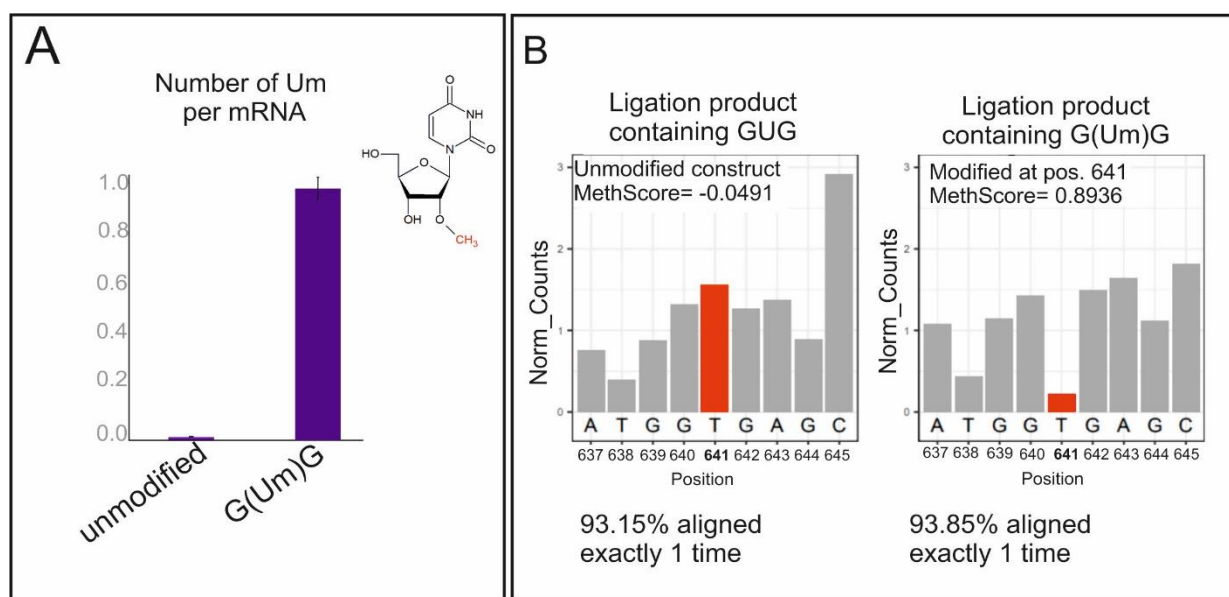


Figure 3. 5: Modification detection in ligated RNA.

(A) Absolute LC-MS/MS quantification of the number of 2'-O-methylated Uridines per IRES-EGFP ligated mRNA molecule. After ligation using fragment 2 harboring 2'-O-ribose methylation uridine (Um), the ligated product was purified and digested for LC-MS analysis. For absolute quantification ¹³C, stable isotope-labelled nucleosides from *S. cerevisiae* were used as an internal standard and mixed in an equal amount with the analyzed sample. An equivalent of one Um could be detected with the modified RNA, while it was absent in ligation reference. Errors bars indicating \pm SD of technical duplicates. (B) RiboMethSeq protection profiles for unmodified and 2'-O-methylated constructs. Normalized RiboMethSeq cleavage profiles around the position 641 are shown for the unmodified and modified construct. Position of modification is shown by a red bar. The MethScores (representing modification level) for position 641 are given for each construct.

It is to note that other constructs were generated and were supposed to contain m⁶A at defined positions. However, the characterization by LC-MS analysis clearly showed a contamination that occurred during the solid-phase synthesis of the modified RNA oligo with other modified nucleotides desoxy-m⁶A (Figure 6. 4). Therefore, this set of point-modified RNA constructs was excluded.

In another application of the 3-way-one-pot ligation, other constructs were created to enable the introduction of modifications this time within the EGFP open reading frame region, which was called “internal-lig block” or in 5' end of the IRES sequence (5' IRES block) (Figure 6. 5 right and left, respectively). In these examples, final ligated product could be also obtained with the “internal-lig block” with a clear separation from non-ligated fragment (1, 2 and 3). Meanwhile, despite the presence of a band corresponding to the expected final size with the “5' IRES block” ligation, further validation steps were required. Indeed different factors, and in particular the length of the fragments should be carefully taken into account. As depicted by the example of the “5' IRES block” in Figure 6. 5 A left, fragment 1 was too short (~80 nts), while fragment 3 was very close in length to the final product (5' IRES block). As shown by the agarose gel, the separation between the ligated product and the free fragment 3 would be difficult if not possible. The validation of the incorporation of both, fragment 1 (~80 nts) and fragment 2 (20 nts), was assayed by hybridization to two complementary fluorescently labelled DNA oligos: one to bind fragment 1 and another one to bind fragment 2) (Figure 6. 5 B left). In both cases, the fluorescence scan showed a band (in the ligation product line) migrating at the same level as the expected size of the full-length mRNA (IVT full-length), which validated the final ligation product.

At this stage, the focus was oriented toward finding an adequate purification method, as presumably, traditional agarose gel elution would result in co-elution of free fragment 3. An alternative was the injection of the ligation product into an HPLC (Figure 6. 6 A). Results demonstrated a large peak of the ligated RNA was eluted roughly at the same time as the IVT full-length mRNA. Fractions were then collected and analyzed on agarose gel, showing a difficulty to distinguish the ligation 5' IRES block from the unreacted fragment 3 (Figure 6. 6 B).

3.1.2.3 Development of a purification method for ligated mRNA

Crucial step during the manufacturing process of synthetic mRNA is the recovery step. It is necessary to prevent side reactions that would be triggered by residual contaminants (salts, proteins, etc. ...) and also to remove the starting material and side reaction products. Multiple standard purification procedures based on chemicals or columns and gel electrophoresis are available. Very recently, the purity of RNA preparation was highlighted as an important factor for applications on living cells, to avoid RNA-sensing by cellular immune responses³⁴². However, a compromise between a scalable purification method and the purity factor is underestimated in literature.

To this point, options were quite limited in terms of choice of purification approaches, in particular applied for RNA ligation products. So far, gel electrophoresis of various types and concentrations was commonly used^{140,303,343–345} and other techniques like magnetic bead-based poly(A) purification remained scarce¹⁸⁵. Aside of featuring multiple problems in classic gel electrophoresis, especially for upscaling the production of large RNA, the impact of the purification step on the RNA quality was underestimated.

Moreover, a challenge could be encountered during the recovery of RNA ligation product. In some cases (Figure 6. 5 A) the reliable size separation *via* gel becomes more important for the isolation of ligated products with similar size range length to their initial substrate fragments. Of interest, developing an optimal purification method must consider, yield and purity of the sample, besides a comfortable handling of the method itself. To address this topic, comparative purification methods using a variety of HPLC (based on a Reversed phase RP18 column) and gel elution were assayed (Please refer to the overview in Figure 6.7).

The purification of the ligation product on HPLC was followed by the analysis of the collected fractions on an agarose gel. As illustrated by Figure 3. 6, gel electrophoresis shows that the separation between the final length ligated product from the other unreacted fragment was not efficient.

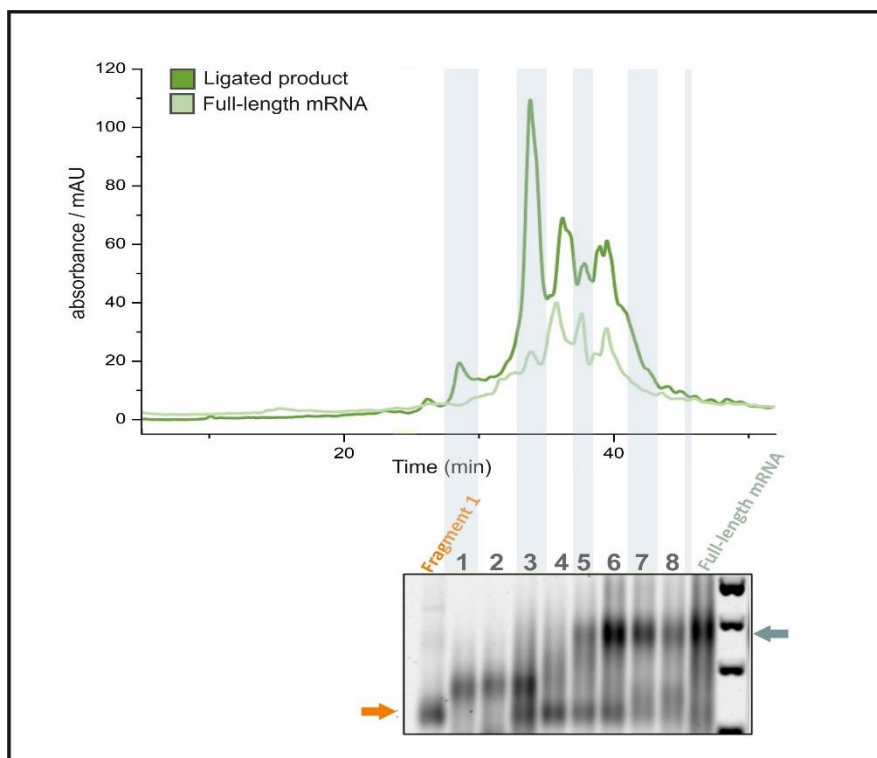


Figure 3. 6: Purification of linear ligated mRNA on RP-HPLC.

Ligation reaction mixture was loaded on HPLC column (dark green). The full-length mRNA, obtained by *in vitro* transcription, was also loaded in another run as a control (light green). The fractions were collected (fraction 1-8), EtOH precipitated and separated on a 2 % gel agarose. One of the fragments (fragment 1 in orange arrow) and the IVT full-length mRNA (green arrow) were also loaded as controls for the gel electrophoresis. The gel was stained with GelRed.

Considering the overall length of the RNA, the extraction from denaturing polyacrylamide gel by excision of the corresponding band followed by a passive elution was precluded. Since the total length of the mRNA (1369 nts) differs significantly from the length of the initial fragments (628 and 722 nts) and the analysis by agarose gel showed sufficient separation (Figure 3. 4 B,C), ligated RNA was tested for agarose gel elution. Therefore, purification was tested by adapting the protocol from Moore *et al.* with some adjustments³⁴⁶. Initial recovery showed a repeated occurrence of solid material in re-dissolved RNA preparations after workup. This material, which occurred in amounts sufficiently abundant to disrupt subsequent HPLC analysis/purification, turned out to be residual agarose particles from the gel purification.

It is important to outline that (i) the ratio of RNA to the gel slice volume is decisive, (ii) that minimum contamination by agarose is unavoidable by this technique, and (iii) that for small amounts of isolated mRNAs, such as those issuing from ligation reactions, said ratio would likely be prohibitively unfavorable for practical application. In an attempt to reduce the number

of agarose particles, filtration of the excised RNA containing agarose slices *via* glass wool was assayed. While the yields for both purification methods were comparable, a significant reduction in agarose content could not be achieved, according to visual inspection of the pellets. The quality control of the pellets was also assessed after RNA purification by means of: low melting/gelling agarose, followed by phenol/chloroform extraction and gel electro-elution. In line with this, a protocol for the purification, based on electro-elution was developed and named, ‘real-time gel elution’ (Figure 3. 7). Briefly, the presence of a fluorescent staining dye (here SYBR®Gold), excited by blue light, allowed to monitor the visualization of the RNA fragment migration within the agarose gel and therefore the stepwise recovery out of a precast trough without the need for gel excision (Figure 6. 8). Afterward, the eluate was spin-filtered using a pore size of 0.2 μm .

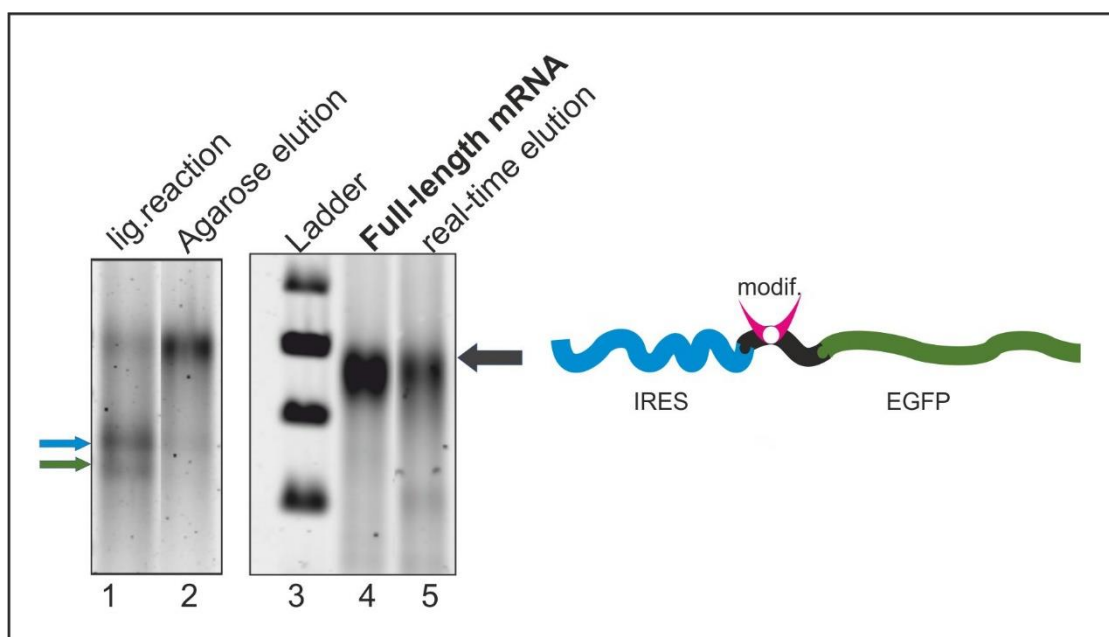


Figure 3. 7 : Purification of RNA ligation by 3-way-one-pot ligation.

After synthesis of the 5'EGFP-lig. Block (scheme), the ligation reaction was loaded on low melting agarose (lane 2) for conventional gel-elution or by real-time gel elution (lane 5). As references, aliquots from the ligation reaction (lane 1) and full-length mRNA transcribed by *in vitro* transcription (lane 4) were also loaded. Blue and green arrow corresponds to the fragment 1 and fragment 3, respectively. Ladder used: GeneRuler 100bp Plus DNA Ladder. The gel was stained with Sybr®Gold.

After precipitation, the remaining agarose particle contamination was characterized by scattering intensities (Figure 3. 8) indicated that the phenol/chloroform extraction produced the

most and largest particles, whereas the real-time gel elution generated the fewest and smallest (smaller than $1 \mu\text{m}^2$).

particle size	Real-time gel elution	Glaswool filtration	Phe/ CHCl_3 -extraction
Small	7	2	6
Medium	3	6	8
Large	1	3	4

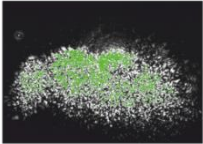
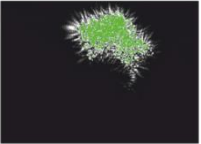

		
Large particle Typical Image from Phe/ CHCl_3 -extraction	Medium particle Typical Image from Glaswool filtration	Small particle Typical Image from Real-time gel elution

Figure 3. 8: Gel particle characterization after RNA real-time gel elution.

Analysing agarose purified ligations samples by Nanosight LM10 showed different size distribution of agarose particles for Real-time gel elution, glass wool filtration and phenol-chloroform extraction. A manually inspection of 8 random chosen positions with particles inside the analysing window were made. The amount of counted agarose particles is given in the table after the different RNA purifications. Typical sizes for large, medium and small particles are shown in the pictures.

Consequently, the purified mRNA was subjected to translation in order to estimate the effect of the remaining agarose particles on preserving the biological RNA functionality (further discussion in the next part).

3.2 Establishment of mRNA functionality analysis via an in-gel fluorescence scan

Several studies showed the possibility to study protein expression *via* a variety of cell-free translation systems, trying to mimic the physiological cellular environment^{140,347–350}. For a first approach, *in vitro* systems is simpler and less complicated than in cell-based assay. Indeed, they are exempt of several factors including cellular transfection, uptake and delivery can define the efficiency of protein expression. Therefore, translation here was investigated in commercial cell-free system.

All mRNAs in this work encoded for the enhanced green fluorescent protein (EGFP), a codon-optimized version of the original GFP protein^{351–355}. Its bioluminescence feature³⁵⁶ and its distinguished stability permit a straightforward read-out of the fluorescence emitted by functional protein. Hence, this provides an overview of potential alteration of the translation machinery by modifications.

Initially, translation was performed in cell-free HeLa cell extract which failed to translate the IRES-EGFP mRNA. An alternative was to use nuclease-treated rabbit reticulocyte lysate (RRL)^{357–361}.

Conceptually, the translation of all mRNA in this work was carried out in RRL system, then protein synthesis was monitored by in-gel fluorescence detection³⁶². Remarkably, the folded fluorescent protein was stable under the rather harsh conditions of a conventional SDS-PAGE, when samples were prepared in non-reductive conditions and without pre-heating. Such an outstanding advantage to assess the integrity of the EGFP protein by a direct and quick detection to verify the fluorescence activity, may bypass the use of further analysis procedures.

Following the electrophoresis, the gel was directly scanned and excited by the corresponding wavelength (488 nm) revealing fluorescent band which corresponds to EGFP (example in Figure 3. 9, lane 2). Interestingly, the in-gel fluorescence scan could also allow the detection of the protein ladder on the same gel by changing the settings of the excitation and emission wavelengths (Only excitation by a red laser, 633 nm). Most of the ladder bands were visualized, in particular around the EGFP bands. Posterior Coomassie staining accurately confirmed the ladder, indicating the possibility to rely on fluorescence excitation without additional staining. The molecular weight of the EGFP is ~27 kDa, even though the observed band exhibited a migration not fitting to the calculated weight. This migration behavior is well described under

these conditions^{362–364} which clearly preserved the conformational structure of the EGFP. In this case, this reflects that the expressed protein is intact and correctly folded.

Considering these different parameters, nuclease-treated RRL was programmed with increasing transcripts concentration (9, 18, 36, 45, 182 and 273 nM) and the produced proteins were separated on SDS-PAGE. Translation was also tested on nuclease-untreated RRL using ³⁵S-methionine validated the same amount. High amounts of mRNA were saturating for protein synthesis, while 20 nM of mRNA, which was in the linear part of the curve above the limit of detection, was used for optimal translation (Figure 6. 9).

In a next step, the impact of remaining gel impurities on mRNA, translation was assessed in commercial nuclease-treated RRL (as previously discussed in section 3.1.2.3). Therefore, a comparative experiment was conducted by the translation of *in vitro* transcribed mRNA, purified by either a commercially available silica-based column or real-time gel elution. The produced proteins were separated on SDS-PAGE followed by in-gel detection. As illustrated in Figure 3. 9 (lanes 2 and 3), fluorescent bands could be detected, and as expected, no band was observed for the negative control, where the mRNA was omitted (Figure 3. 9, lane1).

The comparable fluorescence intensity of protein expressed by mRNA, either purified by real-time gel elution or by silica-column, validated that residual gel particles did not interfere with translation. Hence, the compatibility of the developed purification method with the translation system here allowed its use for the follow up experiments in this work.

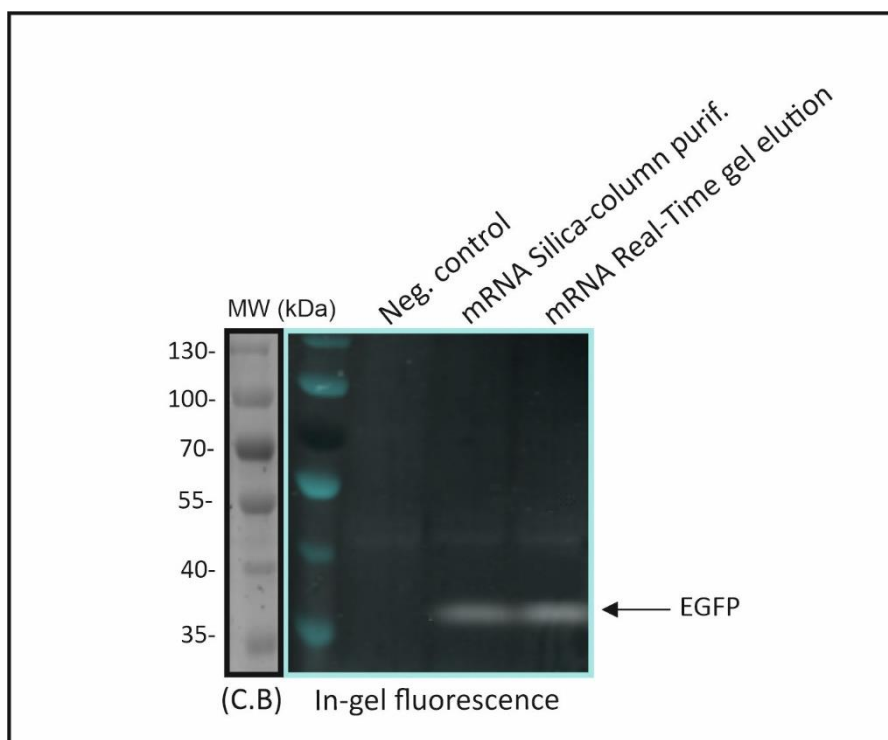


Figure 3. 9: Monitoring protein expression by in-gel fluorescence detection.

In vitro translation of IRES-EGFP mRNA purified either by silica-column or by real-time elution. They were incubated in nuclease-treated RRL then aliquots from the reaction mixture were loaded on a 10% SDS-PAGE. As negative control, mRNA was omitted in the reaction. The EGFP fluorescence was measured in-gel by fluorescence scanning, using a blue laser (wavelength settings: 488 nm excitation/ 522 nm emission filter). On the left side, the gel was stained with Coomassie blue (C.B), then scanned using red laser 633 nm to reveal the weight size ladder bands.

3.3 Translation-tuning by modified nucleosides

In the light of advanced detection methods in mapping and identification of modifications in RNA^{365,366}, a wave of interest was generated to understand particularly their contribution to protein expression. In fact, mRNA translation may hinge on the insertion of a single methyl group within the RNA sequence. One of the most studied modification (m^6A) was shown to promote translation when localized in the 5' UTR of the mRNA sequence, while it may slow down the translation elongation rate when localized in the coding region^{120,132,141}. Furthermore, certain modifications can procure new features for the synthetic mRNA, i.e. preventing the activation of TLRs and RIG-I^{221,224,367–369}.

So far, the inclusion of an IRES sequence within the context of point modified mRNA has not been yet investigated.

Therefore, to have a more detailed picture of the impact of mRNA modifications on protein expression, base and ribose-modified nucleoside triphosphate derivatives (m^6A , Ψ , and 2'-O-Me) were introduced, either at a single-defined position in a large reconstituted mRNA IRES-EGFP sequence (as described in section 3.1 and 3.2), or by random incorporation (m^6A , Ψ) in IRES and non-IRES EGFP mRNA.

3.3.1 Influence of single point-modified nucleosides on mRNA translation

3.3.1.1 Impact of single point-modifications located in the coding sequence

Point modifications on the ribose (2'-O-Me) were introduced, separately, upstream of the start codon through the two successive codons in the IRES-EGFP mRNA. Figure 3. 10 displays the efficacy of the protein produced by *in vitro* translation in the nuclease-treated RRL system.

Taking a global view by studying the effect on codon scale showed that, protein expression did not exhibit a regular pattern when 2'-O-Me was introduced at the first and at the wobble position (third position of the codon). Indeed, protein synthesis was marginally affected, enhanced or inhibited depending on the codon context and position.

The introduction of a ribose methylation at the second nucleotide showed either a pronounced inhibitory effect, reducing the translation by about 60% (as compared to the unmodified RNA), or a complete repression. This effect was present on the codon 1 (GUG) and codon 2 (AGC), which would suggest it would apply to other codons. This faithfully recapitulated findings from three studies, on a short model mRNA¹⁸³, on a longer mRNA in a bacterial *in vitro* translation system¹⁴⁰, and upon transfection into eukaryotic cells¹⁸⁵.

It is noteworthy that the observed translational characteristic at position 2 was studied in the context of cap-dependent mRNA. Thus, this work here provides an additional confirmation of this effect extrapolated on cap-independent translation by using an IRES EMCV.

One study associated the decoding impairment to an excessive rejection of cognate tRNA during proofreading. This lead to stalling translation with the greatest impact on position 2 as described previously by J. Choi and colleagues¹⁸³. They also reported that codons providing G-C base-pairing reduced the stalling duration, with “CC_mC” codon having the lowest stalling duration compared to others in the study (stalling duration increase: AA_mA>UU_mU>CU_mC>CC_mC). Thereby, this stalling was also mirrored by the low translation efficacy.

Interestingly, this could not be observed here, since for instance with the “GU_mG” (codon 1), translation was still negatively affected despite the presence of the G-C base pairing. Therefore, the observed effect of the ribose (2'-O-Me) on translation efficiency here, was rather modification-position-dependent regardless of the presence of G-C base pairing in this codon, which is also confirmed by the codon 2 (AGC) (Figure 3. 10).

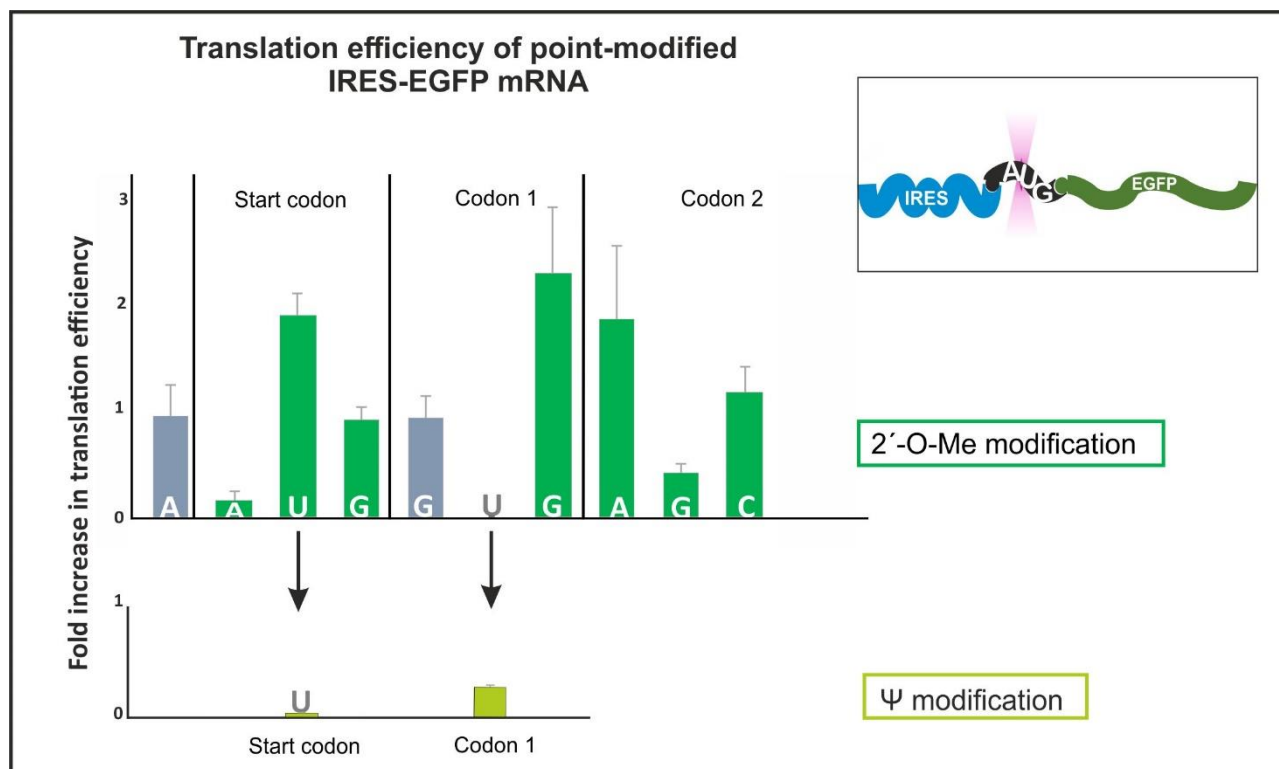


Figure 3. 10: Translation of ligated point-modified IRES-EGFP mRNA.

Point-modified ligated IRES-EGFP mRNA and unmodified one were subjected to *in vitro* translation in RRL system. Aliquots from the translation mixture were separated on a 10% SDS-PAGE followed by EGFP fluorescence in-gel detection. The fluorescence intensity was measured, relative to the non-modified mRNA. Each bar corresponds to individual single-point 2'-O-Me modification, introduced at the indicated nucleotide (letters) from the mRNA sequence (dark green). Dark grey color corresponds to modified nucleotides involved into the Kozak sequence (A_mCCAUGG_m). Light green bars correspond to single-point introduction of pseudouridine in (position 2/start codon) and (position 2/codon 1) only. Error bars indicating \pm SD of biological triplicates.

To understand how boosting or inhibiting can affect the synthesis of a full-length and functional EGFP protein, the kinetic of protein production (of individually 2'-O-Me-modified and non-modified ligated constructs) was recorded (Figure 3.11). Focusing on early time points of the translation reaction, initial fluorescence signal appeared already starting from 5 minutes incubation. The intensity of the signal correlated with protein production: the higher the protein level, the higher is the signal intensity (the curve of: position 3 in the start codon, position 2 in

the codon 1, and position 1 in codon 2). In addition, with low protein level production (example position 2 in codon 2), it is clear that the appearance of a first signal required additional 5 minutes to be detected (relative to non-modified RNA, in grey). Hence, translation was delayed and was time dependent.

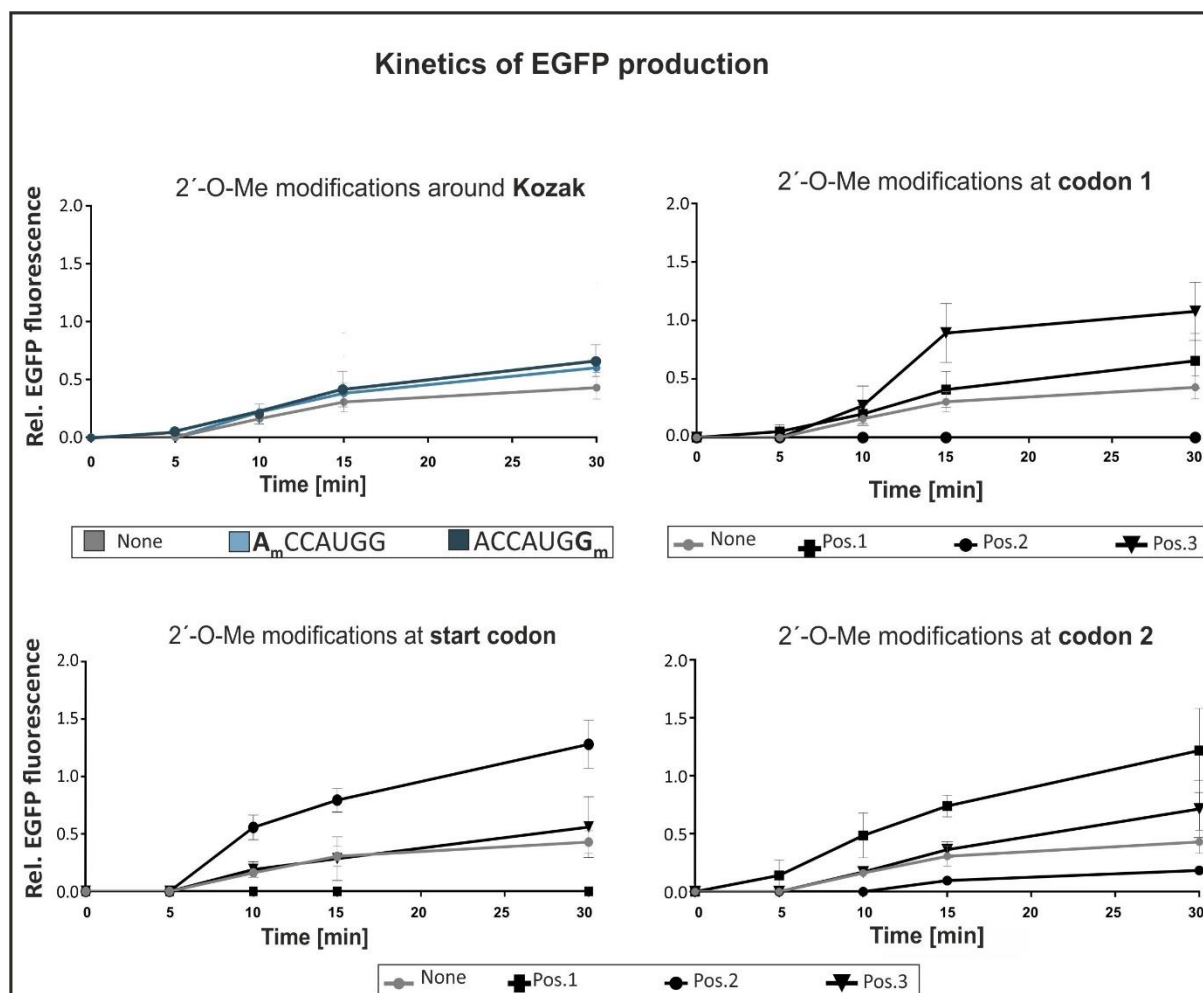


Figure 3. 11: Kinetics of the EGFP protein production from point-modified mRNA.

In the upper left curves, time points from translation reaction of 2'-O-Me inserted in the Kozak sequence (light and dark blue) are shown here until 30 min and protein production was measured relative to non-modified construct at 90 min (complete incubation time). The line in grey corresponds to the unmodified mRNA. Black lines correspond to 2'-O-Me inserted in the first position (square), second (circle) or the third (triangle) from one codon. Each time point is the mean of three independent measurements (\pm SD).

Translation rescue:

In an attempt to rescue the negative translation impact, ligated mRNA harbouring modifications in the second position, encoding for amino acid valine and serine, were incubated with a purified fraction from total tRNA enriched with eukaryotic cognate tRNA³⁵⁸. Data illustrated in Figure 3.12 showed that translation was still inhibited. The same result was obtained with tRNA Serine (data not shown). It could be possible that the concentration of the cognate tRNA in the mixture was not enough to rescue the translation and this could be further studied. Another hypothesis could be that ribose methylation provoked ribosomal alteration by steric clash, as reported previously¹⁸³, however, further investigations are required.

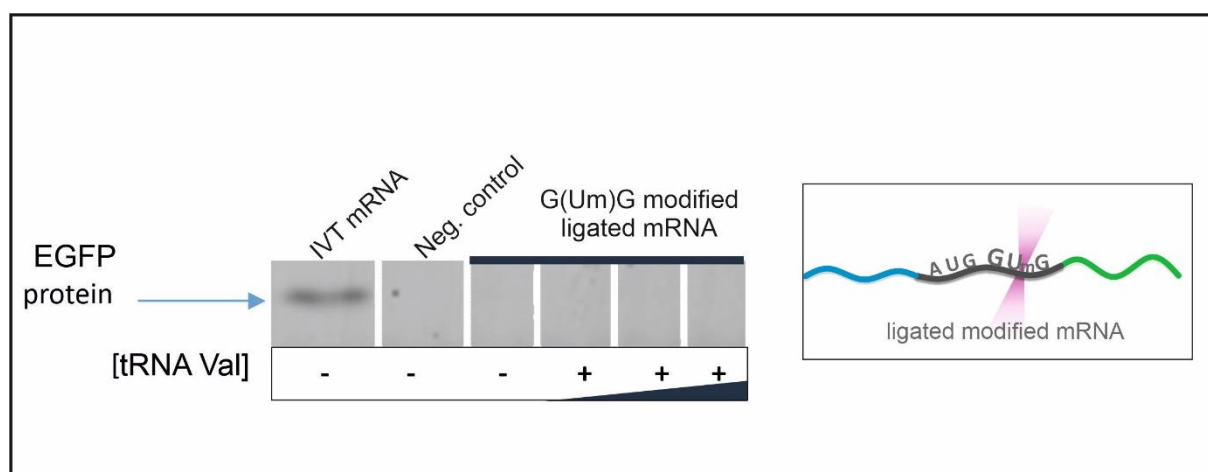


Figure 3. 12: Translation inhibition effect's rescue.

Ligated mRNA harbouring 2'-O-Me modification at the second position of the second codon was applied for *in vitro* translation and supplemented (+) or not (-) with an excess cognate tRNA Valine (0.6 µg, 1.2 µg and 2 µg of tRNA total fraction enriched with tRNA Valine) with respect to the highest inhibiting total tRNA per reaction 200µg/ mL³⁵⁸. An aliquot of translation with *in vitro* (IVT) transcribed full-length IRES-EGFP mRNA, negative control (no added mRNA) and with the ligated modified mRNA without treatment were run additionally on the same 10% SDS-PAGE. After gel electrophoresis, the produced EGFP protein was detected by in-gel fluorescence.

3.3.1.2 Modification impact: Focus on the start codon and the flanking sequence

Deciphering the role of ribose and base modifications on the protein output was mostly studied by introducing modifications in UTRs, within the coding sequence and near/at the stop codon

for translation elongation studies^{106,117,169,185}. However; It is unknown the effect of modifications on the start codon.

In this work, one of the targeted sites within the mRNA sequence was the start codon and the surrounding Kozak site. This latter is a motif (RCCAUGG) that is found in eukaryotes around the AUG start codon and recognized for protein translation initiation^{370,371}. The A nucleotide of the "AUG" is called +1 and the nucleotide upstream is called -1. Positions +4 and -3 have the dominant effect on translation initiation and provide a 'strong' consensus motif. In this context, purine nucleotides are favoured at the positions +4 (i.e., G) and -3 (either an A or G). Here, positions -3 and +4 in the Kozak sequence were modified by the introduction of ribose methylated nucleotides 2'-O-Me. The analysis of the *in vitro* translation product (in the nuclease-treated RRL) showed that translation efficiency was only slightly affected at both positions: A_mCCAUGG_m. This suggests that 2'-O-Me was tolerated by the scanning mechanism for the translation initiation (Figure 3. 10, dark grey colour).

In addition, 2'-O-Me modifications were also introduced, separately, to the nucleotides of the AUG start codon triplet (Figure 3. 10). The translation efficiency relative to the non-modified ligated mRNA, showed that 2'-O-Me at position 1 and 3 of the start codon reduced translation initiation in a higher magnitude when inserted in position 1. Meanwhile, a striking effect was observed for the first time on the second position of the start codon (AU_mG) with an almost 2-fold increased stimulation of the protein synthesis.

Taking all together, the inhibitory feature attributed to the second position as described previously in literature, and validated also by this work in codon 1 and codon 2, is indeed independent from: the codon context, translation system (*in vitro* or in cells), and the mRNA coding sequences (i.e. EGFP reporter mRNA was used here, while others used ErmCL reporter mRNA^{140,185}). But more important, this effect clearly excluded the start codon.

Similar to 2'-O-Me modification, pseudouridine (Ψ) was also inserted at the second position of the start codon. In contrast to 2'-O-Me, Ψ modification led to translation inhibition (Figure 3. 10, bright green). Based on these findings, the influence of Ψ inserted in the second position was also investigated in another codon (codon 1).

As illustrated by Figure 3. 10 (light green), translation efficiency was reduced to (~24%) when present in codon 1. Indeed, one single Ψ was enough to impede or reduce the overall protein production, which was also reported in prokaryotic *in vitro* translation systems¹⁴⁰.

Furthermore, previous study reported that kinetic observations associated this effect as a consequence of a reduced peptide bond formation and guanosine-5'-triphosphate hydrolysis by EF-Tu (a translation elongation factor), hence a reduced elongation rate ¹⁷¹. In the same study, they also reported the resolution of the crystal structure of the *Thermus thermophilus* 70S ribosome with tRNA^{Phe} bound to ΨUU in the A site, which demonstrated that the tRNA was prevented from a proper positioning in the 70S ribosomal peptidyl transferase center (PTC).

3.3.2 Influence of random-modification incorporation on mRNA translation

3.3.2.1 Random incorporation of Ψ

The influence of single Ψ inserted at targeted positions was described in the previous part. However, since several papers reported the advantage of mRNA, harbouring multiple Ψ, on translation, it was sought here to realize a comparative study by preparing randomly incorporated Ψ in IRES-EGFP mRNA (Figure 3. 12).

As described in section 3.1.1, the addition of 2.5% Ψ to the *in vitro* transcription reaction resulted into the substitution of 5% of the uridines in the mRNA with Ψ, which is the equivalent to ~12 occupied positions in a ~1300 nt full-length mRNA. Results illustrated by Figure 3. 12 A, show that translation was already reduced with the lowest content of Ψ. This negative effect was also obtained with Ψ point-modified mRNA (inserted at position 2 of the start codon and the first codon). Since the random incorporation of Ψ by T7 RNA polymerase can be statistically distributed on 251 possible positions, the probability that one or even both of those single positions to be occupied by Ψ is low. This suggests that other yet unrevealed positions along the sequence influenced the translation negatively. Moreover, the translation inhibition was stronger with increasing Ψ modification in mRNA. Interestingly, the complete substitution with Ψ (100%) did not completely suppress the translation, while single Ψ inserted at the start codon could inhibit it (Please refer to Figure 3. 10 for comparison). This suggests possible presence of other critical positions along the RNA sequence with disruptive or stabilizing potential.

These obtained data are also supported by prior studies using *in vitro* prokaryotic *E.coli* ²²¹ or *in vitro* eukaryotic *translation* systems (HeLa cells and wheat germ extract) ^{132,221} as well as HEK293T cells ¹⁸⁵.

In addition, the weak protein expression was also reflected by the kinetic experiment. As shown by (Figure 3. 13 B (purple)), the protein fluorescence was already visible after 5 minutes for the unmodified mRNA. However, modification incorporation resulted in an additional 5 to 10 minutes delay with high Ψ content. This suggests the requirement for a longer time for translation initiation or protein synthesis (elongation). Actually, different factors may define the production of a functional protein, i.e. translation initiation, elongation and termination steps, and also the correct protein folding. In this context, it is possible that the presence of multiple pseudouridines within the mRNA could result in truncated peptides or possibly miscoding events by amino acid substitutions ¹⁷¹, which lead to a loss of the fluorescence activity. For the latter hypothesis, it was previously reported that rewiring the genetic code was also observed when incorporating the pseudouridine in the stop codon, resulting in a nonsense repression ¹⁶⁹. However, no miscoding events were signalled in protein analysis of mRNA containing Ψ located in the coding region ¹⁴⁰.

On the other hand, contrast findings were also published in literature revealing a positive effect of Ψ on translation. In fact, it was reported that Ψ in mRNA stimulated rather the protein yields in rabbit reticulocyte extracts and upon human cell transfection cell lines and mice ^{221,224,232,372}, by modulating the activation of the immune system, thus leading to a higher protein production. The addition of a methyl group to pseudouridine ($m^1\Psi$) was also described for its stimulatory effect of protein expression to a greater degree than with Ψ ³⁷³. Interestingly, in the present work, the complete incorporation of $m^1\Psi$ in IRES-EGFP mRNA inhibited protein synthesis in the RRL lysate (data are not shown).

To recapitulate the findings in this work, inserting Ψ resulted in a negative impact on protein translation efficiency. One hypothesis could be that the observed effect is the consequence of Ψ incorporation in the IRES sequence which defected the secondary/functional structure.

3.3.2.2 *Random incorporation of m^6A*

In a comparative experiment and similar to pseudouridine incorporation, multiple m^6A randomly inserted into the sequence were tolerated and had in general a modest stimulation of

protein production while full substitution completely impeded translation as illustrated by (Figure 3.13 A). Moreover, translation kinetics of protein synthesis was comparable to that of the unmodified mRNA. Previous studies confirmed this finding, to some extent. In fact, a previous study demonstrated that the translation efficiency of mRNA harbouring up to 5% m⁶A was comparable to that of the non-modified mRNA ²²¹. However, complete substitution with m⁶A had rather inhibitory effect in the rabbit reticulocyte lysate and also *in vivo* (on primary cells and Hek cells) ^{120,221}. In a more targeted experiment, the consecutive m⁶A occupation within the same codon was demonstrated to be sufficient to terminate protein synthesis in an *E.Coli* free-cell system ¹⁴⁰. In contrast to these findings, random substitution of 50% of adenosines with m⁶A was reported to enhance the translation of naked mRNA (without a cap and a poly(A) tail) using HeLa cell extract. Thereby, it was demonstrated that m⁶A promoted cap-independent translation in cell lysates ¹³².

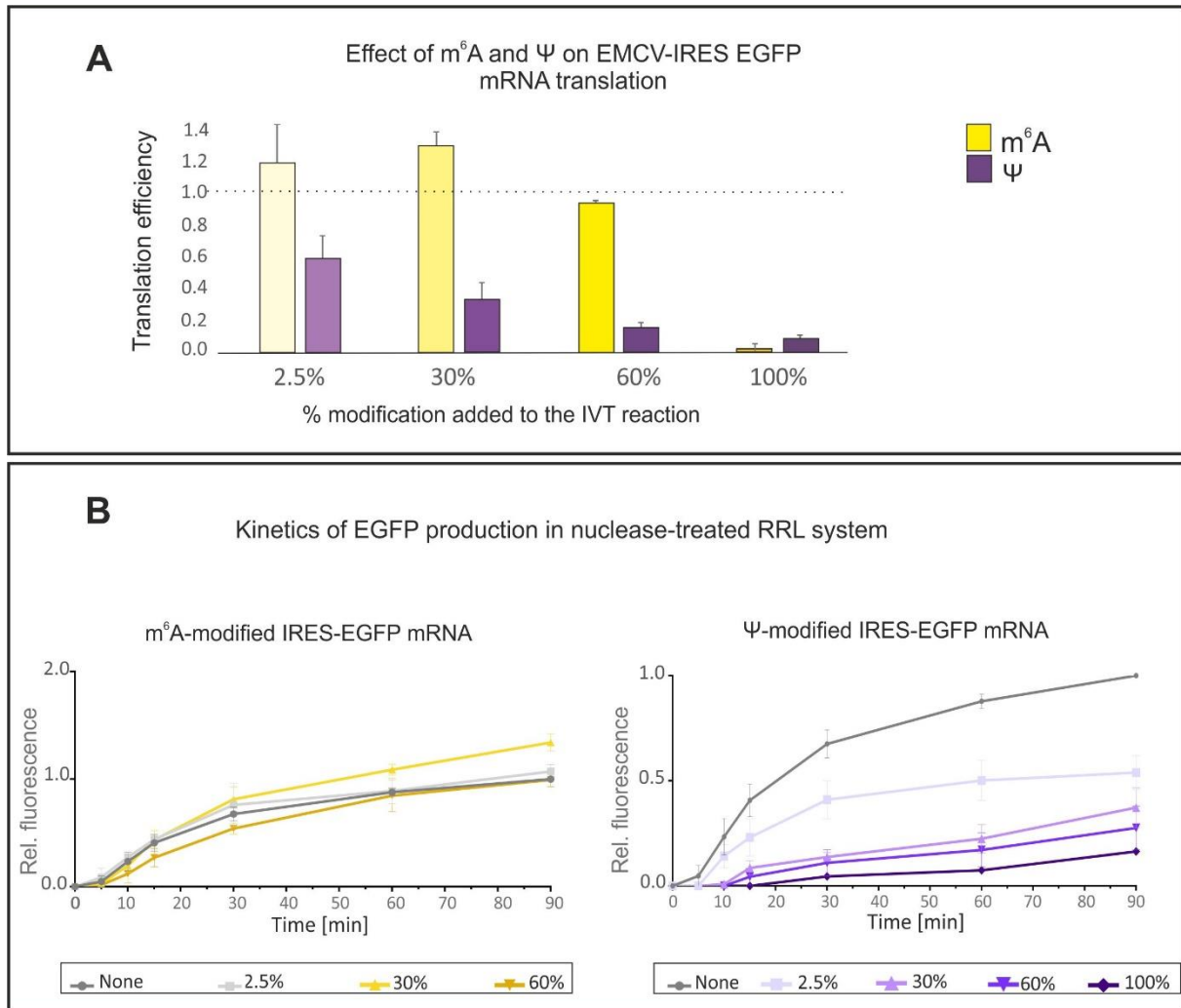


Figure 3.13: Translation efficiency of randomly modified IRES-EGFP mRNA.

(A) Modifications, either m^6A or Ψ were added to the *in vitro* transcription in the corresponding ratios (2.5%, 30%, 60%, and 100%) relative to the corresponding non-modified NTP, to create randomly modified IRES-EGFP mRNA. The effect of modifications was analysed after modified-mRNA translation in the *in vitro* translation system (nuclease-treated RRL). Aliquots were analyzed on 10% SDS-PAGE and scanned for EGFP fluorescence by in-gel detection. Translation efficiency was determined relative to the fluorescence intensity value obtained with non-modified mRNA (set to 1), and normalized to background. Error bars of triplicates increase relative to non-modified mRNA translation. Error bars indicating \pm SD of triplicates. The kinetics of the EGFP protein production was also studied (B). Protein fluorescence was detected by in-gel-fluorescence at different time points (0, 5, 10, 15, 30, 60 and 90 min), and the evolution was determined relative to 90 min incubation (set to one) of the non-modified mRNA. Error bars indicating \pm SD of biological triplicates.

At this stage of understanding, it appeared that different parameters (i.e. the used translation system) can determine as well the efficiency of the translation. However, the common point between those studies cited above, is that investigations on single modified mRNA were established using mRNA constructs without an IRES sequence^{106,117,132,140,169,185,374}. Thus, the results obtained in this work provided an evidence of the modification effect in the context of an IRES containing mRNA. Indeed, random modification incorporation in mRNA could have consequences on the correct folding of the IRES sequence, presumably, leading to a loss of its function. Previous studies revealed the importance of the integrity of stem-loops (stem-loop I, J and K) contained in the EMCV IRES, to ensure the mRNA translational efficiency^{46,375,376}. During the initial steps of translation, the initiation factor eIF4G interacts extensively with J-K domains and recruits eIF4A to the IRES. Together, this complex induces limited local conformational changes in the IRES, preparing a binding site for the ribosome.

Moreover, compared to unmodified adenosine, the introduction of a methyl group (m^6A) was described to induce destabilizing effects and less stable base pairing with uridines^{377,378}. Considering these features, it is probable that the presence of m^6A could be structural disruptive of the IRES required structure for translation.

Therefore, to analyse the impact of modifications independently from this factor, a set of EGFP mRNA harbouring either: a cap, cap/poly(A) tail, poly(A) tail or neither a cap nor a tail, was prepared using random incorporation of either m^6A or Ψ . Subsequently, the modified mRNA was applied for *in vitro* translation. As depicted in (Figure 3.14 A), the translation of a naked mRNA was enabled by the nuclease-treated RRL consistent with literature^{379,380}.

Similar to IRES-EGFP mRNA translation, increasing Ψ content in the different mRNA constructs reduced as well translation efficiency, except for the capped mRNA. Interestingly, the negative effect was not as strong as observed with the IRES-EGFP mRNA. In addition, complete substitution with Ψ retained at least ~40% of the translation capacity with the different mRNA constructs (in the presence of cap and poly(A) or naked mRNA) and almost 80% with capped mRNA (Figure 3.14 A). In contrast, translation was impeded when replacing all uridines with Ψ in IRES-EMCV (Figure 3.13 A). Moreover, the combination of the cap structure together with the insertion of Ψ retained most of the protein fluorescence signal with the different amounts of pseudouridines in mRNA (Figure 3.14 A, green).

As those experiments were performed side by side, results here suggested that the presence of an IRES element had an additional negative influence on Ψ -mRNA.

A kinetics study revealed that the low translation efficacy caused by increasing content of Ψ was also reflected by a delay of the detection of the first EGFP fluorescence detection. The delay was proportional to increasing modification content as compared to the non-modified mRNA (Compare the darkest colour (100% Ψ) to the grey line (unmodified mRNA) in each curve from the Figure 3. 14 B).

Results and discussion

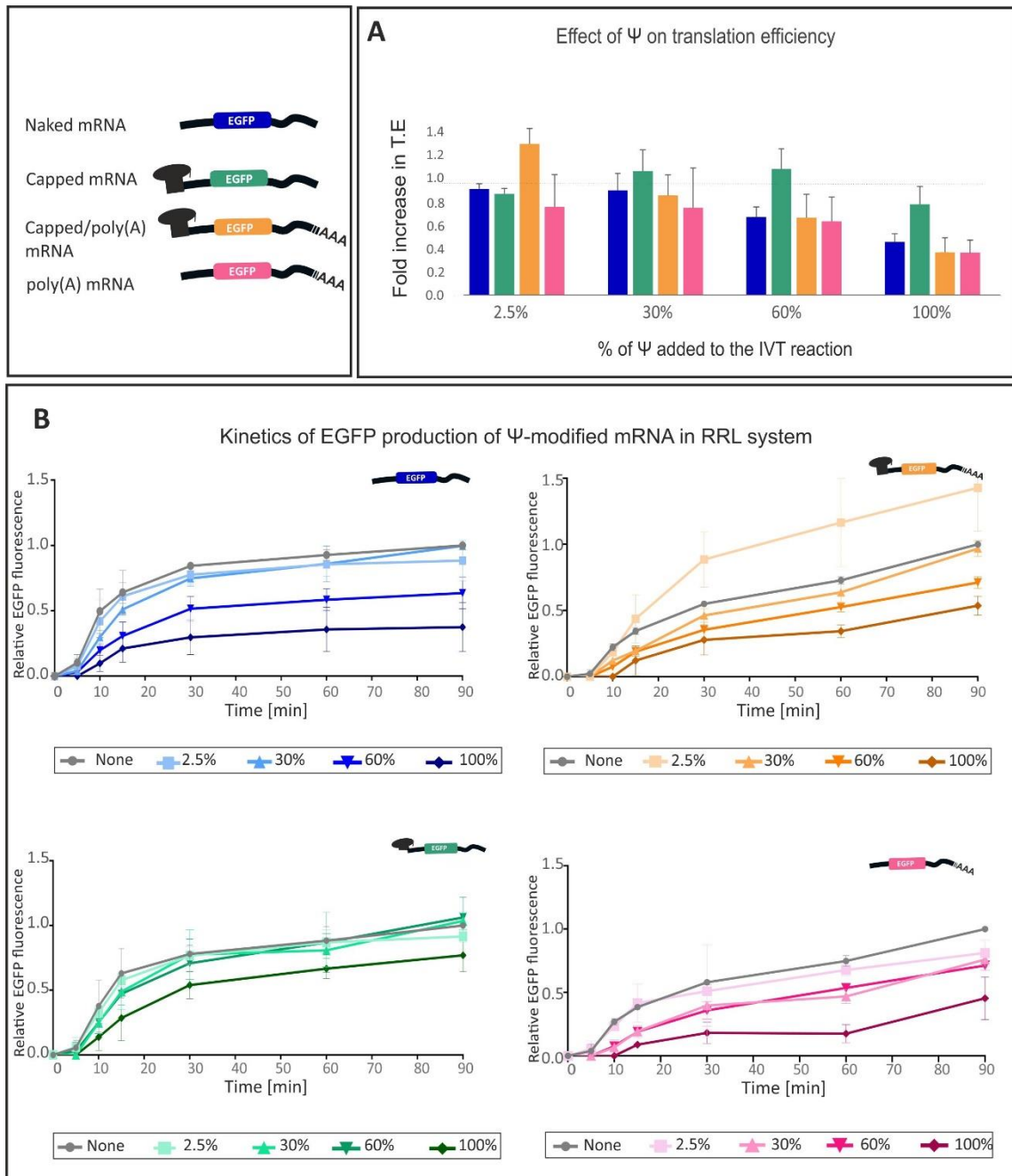


Figure 3. 14: Translation of pseudouridine modified EGFP mRNA.

(A) The Ψ modification was added to the *in vitro* transcription in the corresponding ratios (2.5%, 30%, 60%, and 100%) as described in 3.1.1. To the modified-mRNA, cap or poly(A) tail were added to generate: naked, capped, capped/poly(A) tail and poly(A) EGFP mRNA. For each mRNA construct, the translation efficiency was calculated relative to the corresponding non-modified form (set to one). Error bars indicating \pm SD of triplicates. (B) Time course experiment of EGFP production from modified and non-modified RNA was measured at 0, 5, 10, 15, 30, 60 and 90 minutes. The EGFP fluorescence was detected by in-gel scanning. The translation of the corresponding non-modified mRNA, incubated for 90 minutes, was set to one. Error bars indicate \pm SD of biological triplicates.

In a next set of experiments, m⁶A was incorporated in the mRNA sequence, similar to Ψ. Translation efficiency of naked and capped RNA (Figure 3. 15 A, purple and turquoise colours) correlated with the results obtained earlier with the IRES-EGFP mRNA (Figure 3. 13 A). Moreover, protein expression was sensitive to the additional structures: cap and poly(A) elements (Figure 3. 15 A, brown and green colors), which exhibited a superior protein production of at least 1.5 fold higher for the different m⁶A incorporation ratios. Same to results obtained with IRES-EGFP mRNA, the translation was dramatically altered by 100% replacement with m⁶A, which impeded protein expression.

Time course recorded at different time points showed, in contrast to pseudouridine incorporation, the detection of the first appearance of the EGFP fluorescence signal (between 0-10 minutes) here was not delayed and was recorded at the same time for the different m⁶A-constructs (Figure 3. 15 B), independently from additional structures. The protein production evolved in a proportional manner to the translation efficiency recorded for each modified-mRNA construct.

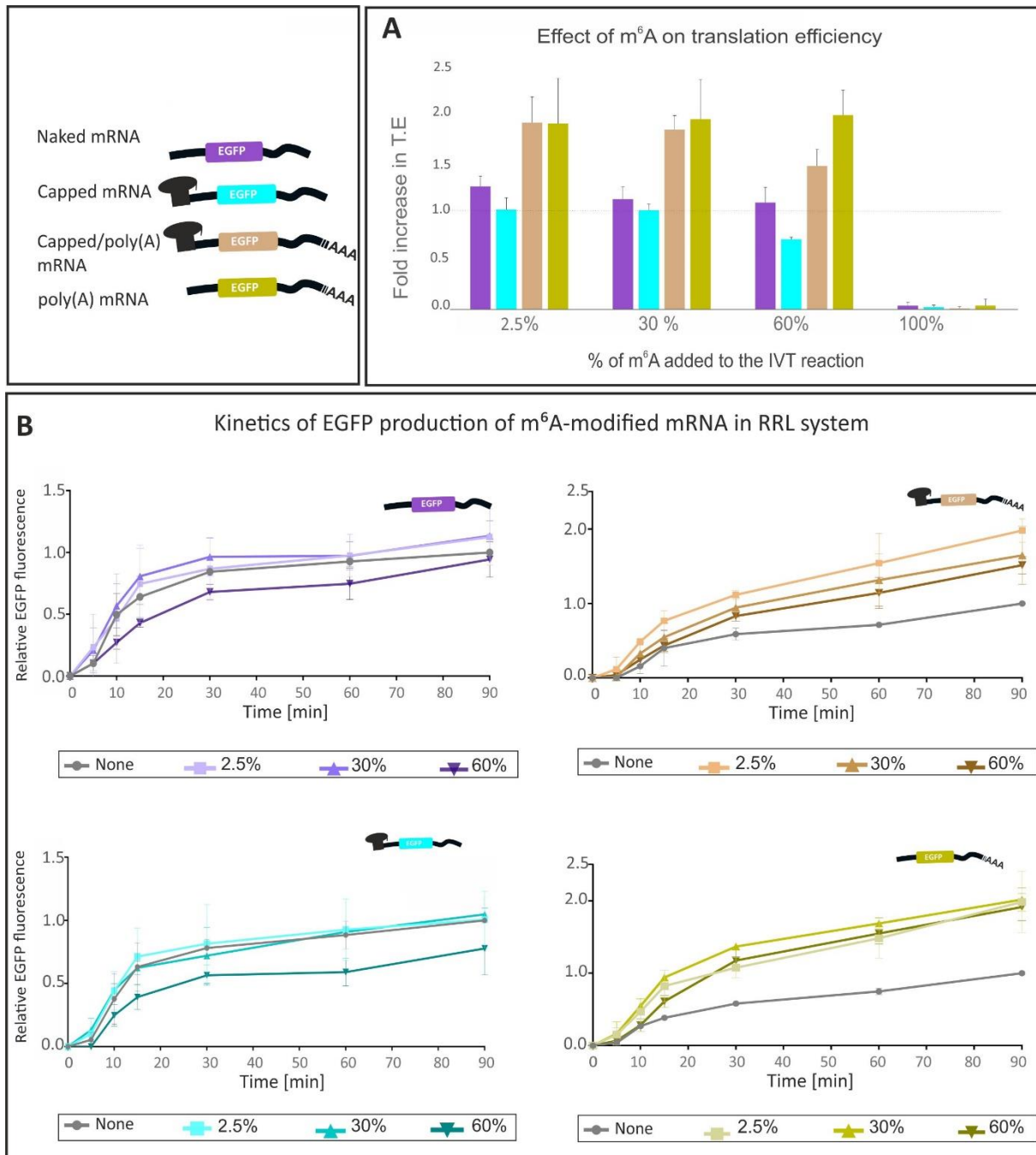


Figure 3. 15: Translation of m⁶A modified EGFP mRNA.

(A) The m⁶A modification was added to the *in vitro* transcription in the corresponding ratios (2.5%, 30%, 60%, and 100%). To the modified-mRNA, cap or poly(A) tail was added to generate: naked, capped, capped/poly(A) tail and poly(A) EGFP mRNA. All mRNAs were applied for *in vitro* translation in a cell-free nuclease-treated RRL system. For each mRNA construct, the translation efficiency was calculated relative to the corresponding non-modified form (set to one). Error bars indicating \pm SD of independent triplicates. (B) Translation aliquots at different time points (0, 5, 10, 15, 30, 60 and 90 min) were measured by in-gel for EGFP fluorescence to determine the kinetics of the EGFP protein synthesis relative to translation product of 90 minutes incubation of non-modified mRNA form (set to one). Error bars indicating \pm SD of biological triplicates.

Extensive studies endeavored to explain the mechanism of function related to m⁶A. It was previously shown the presence of a direct interaction between m⁶A containing mRNA to an initiation factor eIF3A, mediating translation in a cap-independent manner¹³². Considering this finding, randomly incorporated m⁶A in the capped-EGFP mRNA, harbouring an equivalent of 3 modifications (corresponding to 2.5% m⁶A added to IVT reaction) was used as a study model and applied for translation under cap-independent conditions. Thereby, an excess of a cap analog was added to the translation reaction which sequesters the eIF4E, an initiation factor necessary for cap-dependent translation. As illustrated by Figure 3. 16, the translation efficiency was reduced with the unmodified capped mRNA, as expected. Similarly, protein expression using m⁶A-containing mRNA was also negatively affected. Hence, it is clear that the presence of the modification alone could not bypass the canonical cap-dependent pathway and still required the availability of the free cap-dependent initiation factor for an efficient translation.

Collectively, on the light of the presented obtained results here and the conflicting findings from the literature, it turns out to be univocally to draw a clear conclusion. Diverse factors stand behind translation efficiency makes it difficult to generalize the effect of one modification on mRNA translation, such as (i) the used reporter gene sequence (ii) the structural elements constituting the mRNA architecture, i.e., cap-dependent and independent translation and (iii) the used translation system. For example, translating the same modified mRNA can show opposite effect in eukaryal and procaryal *in vitro* translation systems and showed also a possible dependency on cell types³⁷⁴. From a structural point of view, the modification would be involved into the creation of local structures that are recognized by certain interacting or translation proteins and ribosome. These sites would then favorite translation or reduce it³⁸¹. This would also suggest then the presence of some particular accessory proteins and pathways which are available in some translation systems or cells and not in others, and thus, would explain the contradictory results obtained *in vivo* and *in vitro*.

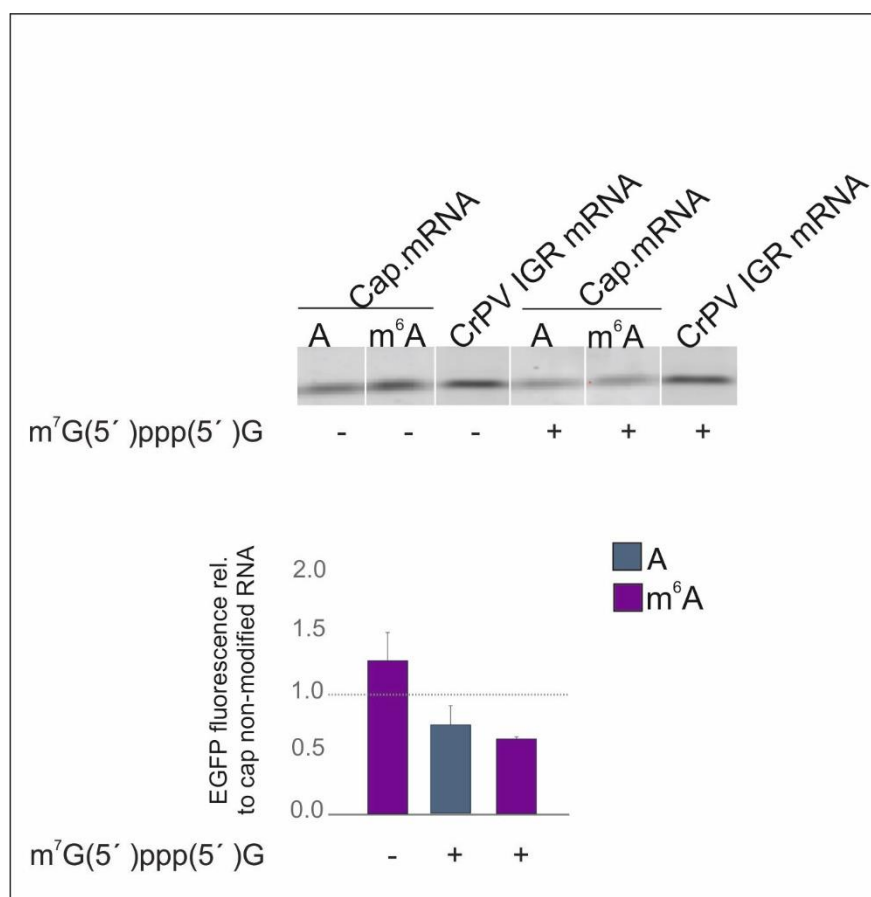


Figure 3. 16: m⁶A modified-RNA requires translation initiation factor eIF4E.

EGFP capped mRNA containing either m⁶A or non-modified adenosines was applied for translation in treated-RRL system. Free cap analog m⁷G(5')ppp(5')G was added to sequester cap-binding translation factor eIF4E. An aliquot from translation reactions was analysed on 10% SDS-PAGE followed by in-gel fluorescence scan. Translation of the capped mRNA was reduced by the addition of the cap analog. The CrPV IGR-EGFP mRNA, which is eIF4E-independent, was used as control. The presence of a cap analog was unable to inhibit its translation. Down, EGFP fluorescence was quantified relative to capped non-modified mRNA. Values here are shown as the average of two independent experiments.

3.4 Circular RNA

3.4.1 Design and engineering of circular mRNA

Over the past years, several efforts to improve artificial mRNA stability and efficiency suggested the use of cap structure, modified nucleosides, or codon optimization. The artificial circularization of mRNA molecules is an additional strategy in the search of sustainability and it permits (on another hand) the understanding of the transferability potential of circular RNA (circRNA) ^{55,257,275,303}.

3.4.2 Design and synthesis of circular RNA

Various strategies were developed for the circularization exogenous RNA based on enzymatic reactions (using DNA and RNA ligase), chemical ligations and by self-splicing introns (reviewed in ³¹⁵). The generation of circular RNA including an internal ribosome entry site (IRES) EMCV has been described using either introns permuted intron-exon splicing (PIE) (self-splicing introns) ²⁸¹ or splint ligation using a (DNA ligase I) ³¹⁴.

Here, linear mRNA molecule containing a wild type EMCV inserted upstream of an EGFP sequence ³⁷⁶ was circularized by an enzymatic approach using a T4 RNA ligase II

For an efficient nicking ³⁸², a splint DNA has been designed to be complementary to 30 nucleotides of each mRNA extremity, joining the 5' to 3' end extremities (Figure 3.17 A). Another aspect for the circularization reaction has been the synthesis of RNA in the presence of an excess of guanosine monophosphate (GMP) as a nucleotide initiator leading to a single phosphate. This permitted to create a single phosphate at the 5' end, which is desired to prime the ligation ^{314,383–385}. As depicted in Figure 3.17 B left, the different products of the circularization reaction were separated on a denaturing PAGE for analysis and revealed two bands: the first located at the same level as its linear precursor, while the second band migrated significantly slower. The incubation time was optimized and showed that incubation for 30 minutes at 37°C was sufficient to produce an additional band after gel analysis (presumably the ligated product), while long incubation reduced the RNA quality and the yield (Figure 6. 10).

Circularization reaction was separated on denaturing PAGE for analysis (Figure 3.17 B left) and revealed two bands: one band at the same level as for the linear precursor RNA and another one that migrated significantly slower than this latter.

This additional and retarded band is a typical characteristic of circRNA, and was observed in several studies ^{386,387}. Assuming that the upper band corresponds to the circular form, a yield of at least 45% of circularization efficiency could be achieved (estimated by ImageJ®). As expected, the absence of the enzyme did not produce an additional band.

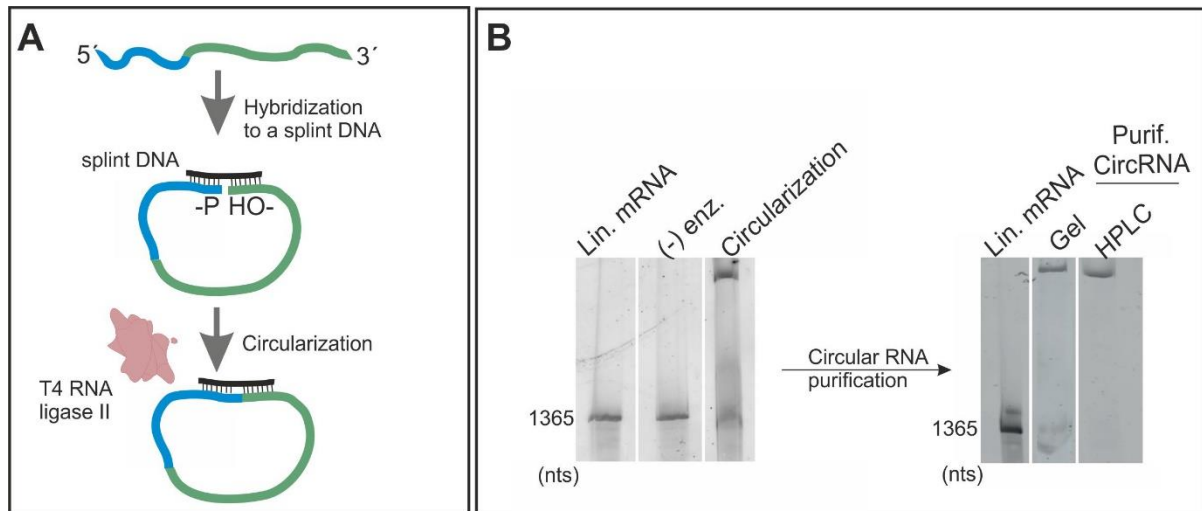


Figure 3. 17: *In vitro* synthesis of circular RNAs and purification.

(A) Schematic workflow of circular mRNA synthesis using a ligase. RNA was *in vitro* transcribed (IRES and EGFP sequences are in blue and green, respectively) and annealed to a complementary DNA oligonucleotide (splint DNA). The latter plays the role of a bridge to bring together the 5' end to the 3' end in proximity. Subsequently, the reaction was incubated in the presence of T4 RNA ligase II. (B) An aliquot from the circularization reaction, after DNase I treatment, was separated on 4% denaturing polyacrylamide gel electrophoresis (PAGE). As a control, the enzyme was omitted from the reaction (gel on the left). The upper band corresponding to the circularized form was either gel eluted or purified on HPLC (PAGE on the right). The bands were revealed using GelRed staining.

3.4.3 Circularization of modified mRNA precursor

Linear mRNA precursors were *in vitro* transcribed in the presence of different ratios (2.5%, 30%, 60%, and 100%) of modified nucleoside triphosphates (Ψ or m^6A) to their equivalent unmodified. CircRNAs were prepared under the same ligation conditions. The analysis of the ligation product on PAGE showed a comparable profile to the circularization of non-modified mRNA (data are not shown). It appears that complete substitution of the canonical nucleotides by m^6A or Ψ did not impede circularization. This finding supports other study that reported the failure of circularization of modified RNA using a commonly used method for circularizing large RNA molecules, based on on a spontaneous group I intron self-splicing system, while circularization based-enzyme rendered it possible³⁸⁸.

3.4.4 Circularity validation

The validation of the synthesis of a circular RNA was addressed by multiple strategies and tools in the literature (reviewed in ²⁷⁵). Some of them were experimented within this work, based on their accessibility in the lab.

3.4.4.1 Validation of RNA circularity by traditional validation strategies

3.4.4.1.1 Electrophoretic mobility

As mentioned earlier, the distinguishable migration of circular RNA and the linear counterpart on gel electrophoresis could be utilized as a method to identify circular structures. In this direction, urea-polyacrylamide gels containing different crosslinking were prepared by varying the ratio of acrylamide to *N,N'*-methylene-bis-acrylamide. As expected, the migration of the circRNA was always slower than that of the linear form in the three gels (Figure 3.18 A and Figure 6. 11). This could be also summarized by the decreasing ratio of circRNA migration of relative to its linear precursor (green circles), with increasing gel cross-linking (Figure 3.18 A). Meanwhile, a reference RNA kept an almost stable migration ratio relative to the precursor linear RNA. Thus, the slowly migration of circRNA, relative to linear precursor, was more delayed by increasing gel crosslinking which was consistent with previous reports ^{278,386}.

Another electrophoretic mobility assay was reported on a native agarose gel. In such case, linear and circular RNA migrated at the same level, in contrast to their migration on a denaturing PAGE ³⁸⁹. This was recapitulated in the obtained results here (Figure 3.18 B).

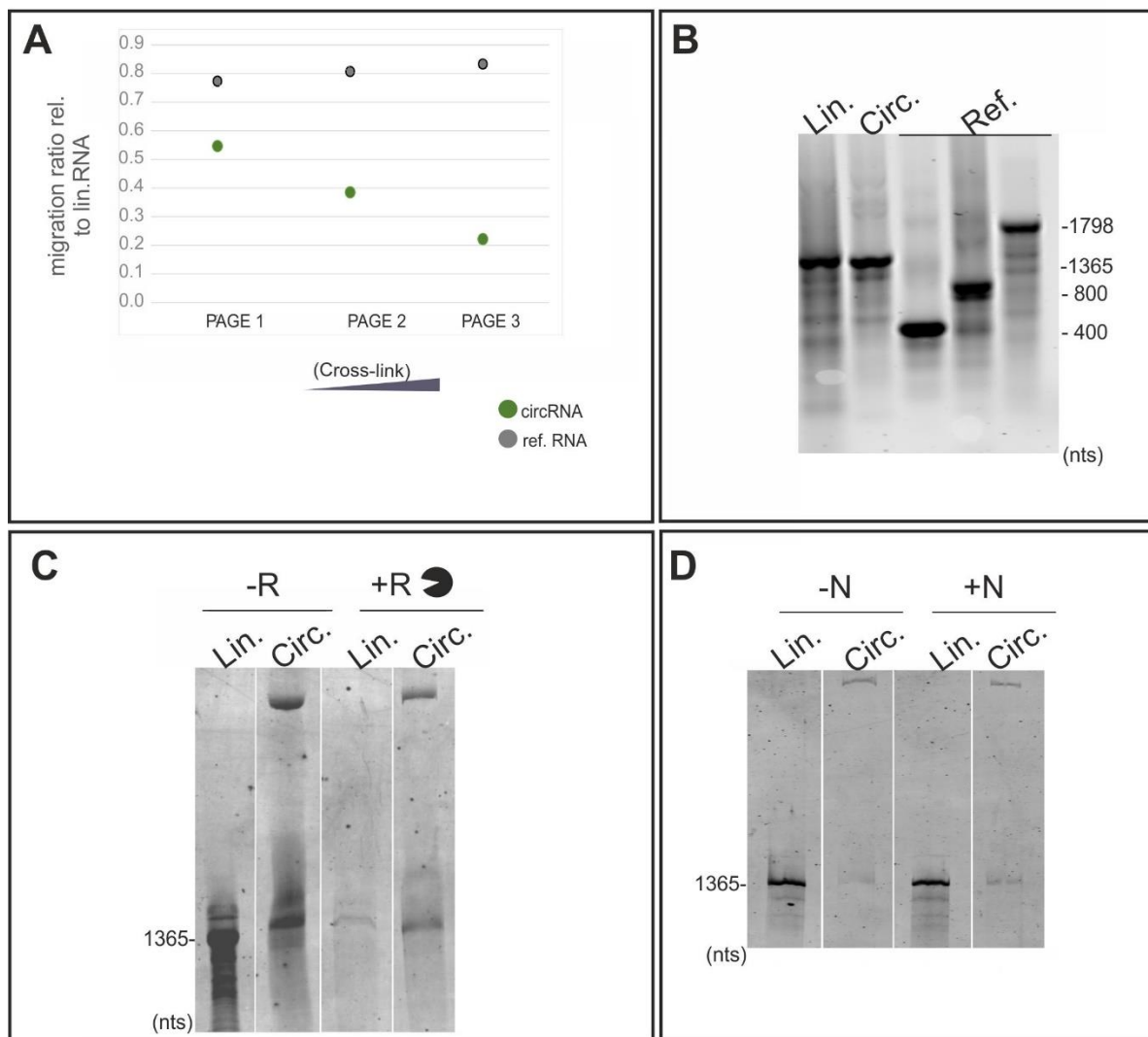


Figure 3. 18: Circularity verification by different biochemical assays.

(A) An aliquot from circular RNA reaction (circRNA) was separated in urea-denatured PAGE, with increasing degree of the crosslinking, the ratio of the bisacrylamide to acrylamide from PAGE 1 to PAGE 3. For each gel, the ratios of the migration distances, of circular RNA (circRNA) and a reference RNA (ref.) with a known size, were determined relative to linear RNA precursor (green corresponds to the ratio of circRNA/linear precursor and grey to the ratio of a ref. RNA to linear precursor). (B) Investigation of the circular RNA migration behaviour on a 1% native agarose gel. RNA fragments from a known size were used as references to assess the migration (Ref.). (C) Digestion of linear and circular RNA with RNase R (+R) on a 4% denaturing PAGE (the gel was stained with GelRed). (D) Gel-purified putative circular and linear RNA were subjected for partial hydrolysis (+N). Consequently, the linear form results in a smearing band while the circular form would be nicked, producing a linear RNA form and migrating similarly. Untreated samples were assigned as (-N). The nicking reaction was analyzed using a denaturing PAGE and stained by SybrGold®.

3.4.4.1.2 *Enzymatic-based treatment*

An enzymatic approach was adapted by treating the ligated product with RNase R. This enzyme belongs to the RNase II superfamily and acts *in vivo* as a quality control for abnormal and damaged mRNA^{390–393}. It is able to processively hydrolyze RNA from the 3' - 5' direction. Thus, it can be used to discriminate between lariat (loop structure) RNA and branch RNA.

Here, RNase R assay was initially performed on linear mRNA following the manufacturer's recommendations. Unexpectedly, the subsequent analysis on PAGE showed that the mRNA was left intact. Therefore, extensive optimization was assayed by varying the final concentration of the enzyme. This resulted in the degradation of a small portion of the linear mRNA, with increasing RNase R concentration (Figure 6. 12). One hypothesis was that the secondary structure of the mRNA may hinder the cleavage activity, although RNase R was described to process highly structured RNA^{394,395}. Therefore, mRNA was heat-denatured prior to digestion. Interestingly, the pre-heating step dramatically facilitated mRNA degradation (Figure 6. 12), which occurred already with the lowest enzyme concentration. Nevertheless, the presence of undigested mRNA was still present as a “shadow” band.

Based on these initial experiments, an aliquot from the circularization product was also treated under the same conditions with RNase R. As displayed by Figure 3.18 C, the intensity of the band corresponding to the linear mRNA was barely detectable. However, the intensity of the two bands from the ligation product decreased, as compared to their input sample (-R, without RNase R). This suggests that the RNase R digestion was not selective and did not exempt the circular RNA form. Importantly, previous researchers remarked that RNase R enzyme batches could differ from one to the other and could be contaminated with endonucleolytic enzyme activity²⁵⁸, which could be a plausible explanation for the observed random digestion and raise questions about the quality of the enzyme preparation.

Therefore, the results obtained here suggested that RNase R was not a reliable method to characterize circRNA in this case.

3.4.4.1.3 *Chemical-based treatment*

Another assay to validate circularity was described based on alkaline partial hydrolysis^{259,389}. Presumably, such treatment would result in single nicking which opens the circle and produce predominantly the linear form. Therefore, the putative band of circular RNA was gel-eluted and treated for partial alkaline hydrolysis. Different reaction conditions were tested by varying the

pH (8, 9, and 9.5) and the hydrolysis time incubation (30 sec., 2.5 min., and 5 min.). Analysis of samples on PAGE showed that long incubation (above 30 sec.) in high pH and at 90°C lead mostly to RNA degradation (data are not shown). Under optimal conditions (please refer to the materials and methods section), alkaline hydrolysis resulted in a smearing band of the linear RNA (Figure 3.18 D). Meanwhile, the intensity of the putative circular band was decreased in the benefit of the lower band whose intensity was slightly increased.

Despite different attempts to optimize the partial hydrolysis reaction, it is clear that the obtained results could be more valuable, with a more significant effect (a higher intensity of the lower band). Therefore, another strategy was used to investigate the circularity, which is described in the following part.

3.4.4.2 Circularization validation on RP-HPLC

The injection of circularization reaction injected in a reverse phase HPLC (RP-HPLC) resulted in the elution of two distinct major peaks: one was eluted at around 17 minutes and another one at 40-42 minutes (Figure 3.19 A, green). The eluted fractions were subsequently gel analyzed, and showed that they corresponded to the circular and linear forms, respectively (Figure 3.19 A and B). Moreover, gel analysis of the fractions recovered between 17 and 40 minutes revealed a mixture of circular and linear forms.

Thus, the different steps of the circularization reaction could be fractionated on HPLC and eluted in the following order: First the circular RNA, then the input reaction, and finally the linear precursor. Given the principle of function, the elution is based on hydrophobic interaction. Therefore, polar molecules would be eluted earlier in a proportional manner to the RNA size. Since both circular and linear mRNA have the same length, it seems that the circularization added new features allowing the circular RNA to be first eluted, probably by decreasing the interaction area with the stationary phase. Hence, aside from one study reporting the separation of circRNA on size exclusion chromatography ²⁸¹, findings here show for the first time the potential use of an alternative strategy.

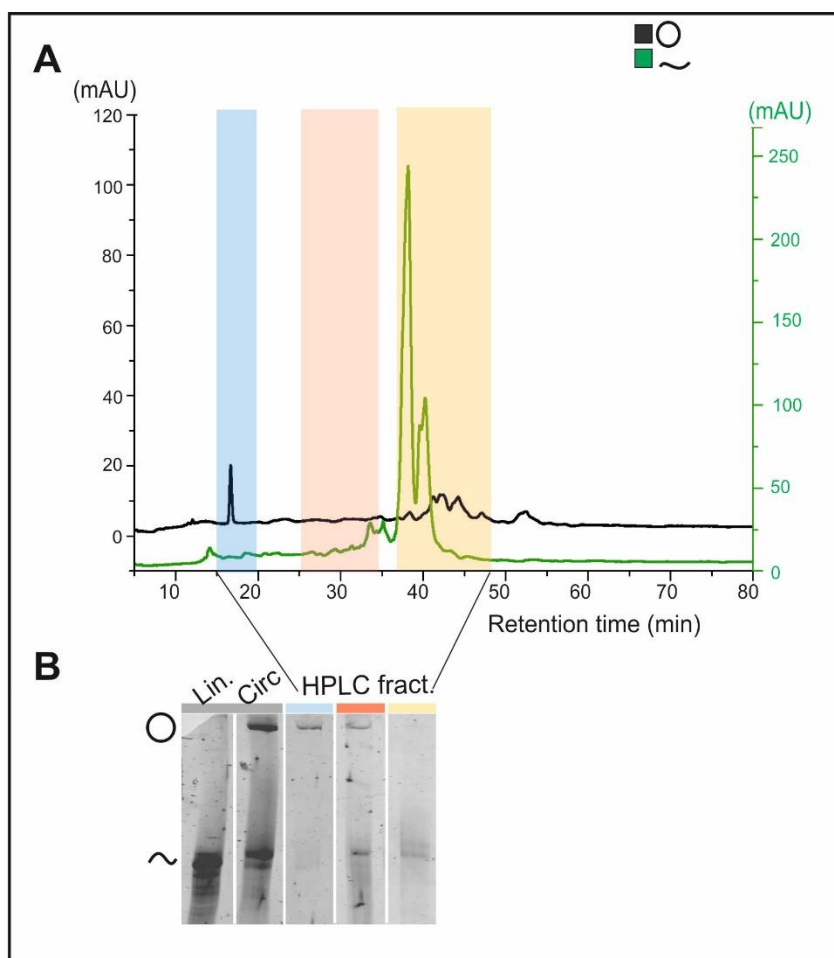


Figure 3. 19: Identification of RNA circularity *via* HPLC.

Circularization reaction was purified by phenol/chloroform then injected to RP-HPLC. An overlay of HPLC chromatogram of linear RNA (black) and circRNA (green) is shown in (A). The circle and the lane correspond to circular and linear RNA, respectively. Fractions were collected at the defined times (indicated by colors) and analyzed by separation on a 4% denaturing PAGE (B). Bands were visualized by SybrGold® staining.

Altogether, data here showed the generation of large circular RNA out of a ~1.3 kb precursor RNA using splint ligation (Figure 3.17 and 3.18), while first studies resorted to permuted intron-exon (PIE) strategy to produce circRNA in this range of size ^{315,388}. Interestingly, neither the structured IRES sequence, nor the presence of modification constituted a challenge for an efficient ligation.

3.4.5 Protein synthesis from circular RNA

Most of the circRNA found in nature were described to fulfill non-coding roles. They were postulated to contribute as gene expression regulators, i.e. as miRNA sponges^{259,265} or “mRNA trap” to sequester the translation initiation site³⁹⁶. Considering the fact that circRNA could originate from exons, besides their localization in the cytoplasm, an interest has been drawn to study their capacity to produce proteins^{258,259,261,263,397}. Consequently, few recent studies reported the translational capacity of a set of natural circRNAs^{55,285,398,399}. One alternative for better understanding and studying how circularized mRNA could direct the synthesis of a defined protein, could be performed *via* synthetic circRNA.

In this work, the circularized RNA, encoding for EGFP protein, was obtained by splint ligation (Figure 3. 17). Afterward, the corresponding band was gel-purified and was subjected to a comparative translation using nuclease-treated and untreated rabbit reticulocyte lysate (RRL). The translation product was separated on an SDS-PAGE and the ribosomal biosynthesis of a functional EGFP was monitored *via* an in-gel protein fluorescence scan (please refer to section 3.2 for more details). As depicted in (Figure 3. 20 A, upper gel), a band was detected with the linear and the circRNA translation, but not with the control. In addition, circRNA produced proteins that had homogeneous size distribution and migrated as “a monomeric band” to the predicted size of proteins translated by the equivalent linear RNA (~30 kDa). This result suggested that circRNA was able to initiate EGFP protein translation that was correctly folded with an active chromophore, consisting with previous report using engineered IRES-EGFP circRNA^{281,314}.

Next, circRNA was also programmed in untreated RRL system. This cell lysate was not nuclease treated and would recapitulate “near physiological cell conditions”⁴⁰⁰. After translation, the quantification of the expressed EGFP protein, revealed a less efficient translation than that obtained in the nuclease-treated one (Figure 3. 20 A). Indeed, translation efficiency was almost two times lower than that from the respective linear mRNA (Figure 3. 20 A and Figure 3. 21). While the translation yield from circRNA was not always reported in literature, such poor capacity of circRNA translation was in agreement with results from previous studies obtained in eukaryotic and prokaryotic translation systems^{55,312–314,401}.

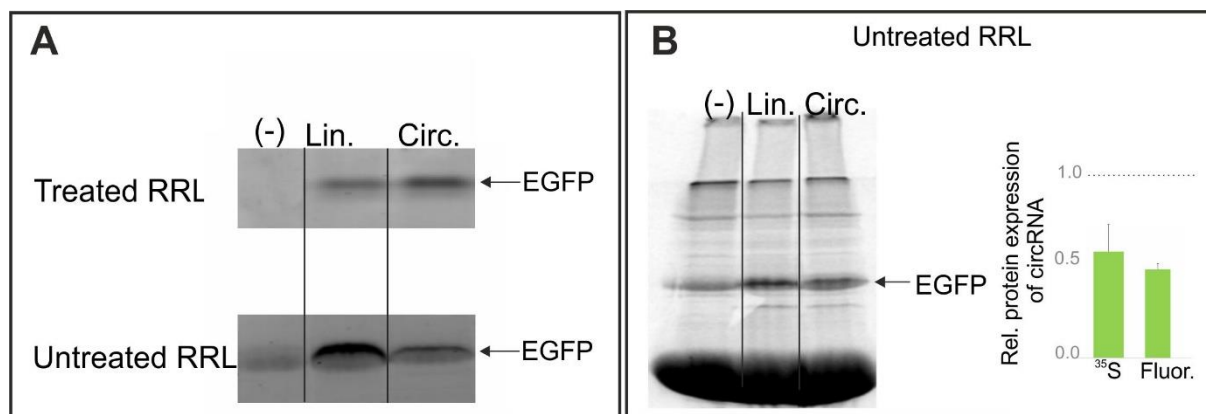


Figure 3. 20 : Translation of circular RNA in rabbit reticulocyte lysate (RRL).

(A) The translation efficiency of circRNA (Circ) compared to its linear (Lin) precursor was tested in treated and untreated RRL. After incubation at 30°C for 90 minutes, aliquots were analysed by separation on a 10% SDS-PAGE. Subsequently, produced EGFP protein was detected by in-gel fluorescence scanning using blue laser settings (488/520 nm as excitation and emission wavelengths, respectively). As controls, reactions containing linear RNA (Lin) or without RNA (-) were used. (B) Circular (Circ) and linear (Lin) RNA were also translated in untreated RRL in the presence of ^{35}S -methionine and analyzed by separation on a 10% SDS-PAGE. As a control, RNA was omitted from the translation reaction (-). On right side, protein expression from circular RNA was quantified after phosphor imaging or by in-Gel fluorescence, relative to the respective translation from linear mRNA. Data are presented here as an average (\pm) of two independent experiments.

It was previously reported that during the translation of circRNA, additional high molecular weight products may also arise from a continuous translation, whereby, ribosomes traverse the circular RNA template for multiple rounds like in the rolling circle amplification (RCA) mechanism^{303,312,402,403}. Such an event could result in an elongation of the protein at its C-terminus extremity, producing large, long-repeating polypeptides and, presumably, the loss of fluorescence of the EGFP protein

Therefore, to verify if such an event took place here, the circRNA was subjected to translation in the presence of radioactive methionine (^{35}S) in the untreated-RRL system. The produced protein was separated on an SDS-PAGE and analyzed by the exposure to a phosphorimager. As depicted in (Figure 3. 20 B), there are no larger proteins above the expected size, as compared to the negative control (first lane).

Altogether, with the previous results we hypothesized that translation occurred in a single round until encountered the stop codon, restricting the final size of the synthesized protein. Thus, the RCA event was excluded.

Importantly, the low translation efficiency from the produced proteins evaluated by ³⁵S-methionine incorporation, correlated also with the weak fluorescence intensity, confirming a total low EGFP protein production in untreated RRL (Figure 3. 20 B). One hypothesis is that the endogenous mRNA present in the untreated-RRL (while digested in treated-RRL) may have a role in the different translation capacity between both RRL systems. In fact, untreated-RRL contains endogenous mRNAs (mainly globin mRNA) which are actively translated and would be competing for translation with exogenous RNA templates³⁴⁹

If this was the case, the stronger reduction of translation capacity of circRNA (~2 times lower) revealed that circRNA could be concerned to the competition with the endogenous RNA. Therefore, further investigations are required to explore this hypothesis. One possibility could be through (i) a micrococcal nuclease treatment to inhibit translation of endogenous mRNA.

The mechanism of translation from circular RNA is not well elucidated. Based on the conventional eukaryotic cap-dependent translation, a free 5' end is required for translation initiation. However, both ends in the circular RNA are engaged to form the closed circle. Such an unusual structure would, presumably, mimic the semi-circle formed during canonical mRNA translation through the interaction between proteins that bind the cap and the poly(A) regions⁴⁰⁵, or it would hamper this complex formation.

Obviously, in this work, the small subunit ribosome 40S was still able to initiate translation on the circular RNA without the requirement for free 5' end scanning. Fundamentally, circRNA translation would be enabled through another mechanism, to allow internal initiation. A previous study by Chen and Sarnow reported the synthesis of a circular RNA containing IRES EMCV by splint ligation. They demonstrated that circRNA, without an EMCV element failed to translate proteins³¹⁴. Interestingly, two decades later, Abe and colleagues described the possibility to translate proteins from a synthesized circRNA in the absence of any particular element for internal ribosome entry, a poly-A tail, or a cap structure³⁰³. This finding raised again the question about the followed mechanism of translation of circRNA. Indeed, very recently, a handful of natural circRNAs have been found to contain specific sequences that mediated internal translation in a splice-dependent/cap-independent manner^{55,285,306}. Besides valuable insight about the natural circRNA formation and expression, they also reported a low protein expression level.

In our case, the circular RNA contains already an IRES element. Such mechanism is well-described as an alternative mechanism for capped-mRNA translation^{317,406}, where translation

initiation was shown to bypass the requirement for the cap and eIF4E (cap-binding protein)^{48,50}. The question that could be raised was if the translation initiation on circRNA depended on the IRES element or on the circular RNA form itself. Therefore, it would be of interest to investigate its translation without IRES.

Given that the same RNA sequence containing natural translation regulatory sequences (found in circRNA) can drive a different translation pathway depending on the RNA form (circular or linear)²⁸⁵, this would explain also the difference of translation efficiency between linear and circRNA obtained here and further investigations in this concern are required.

3.4.6 Translation of HPLC-purified circRNA

In another experiment, circRNA was purified using RP-HPLC and programmed in RRL lysate for translation. Surprisingly, fluorescence quantification revealed a low translation capacity of almost 20% relative to linear precursor in nuclease-treated RRL, while it was repressed in nuclease-untreated RRL (Figure 3. 21). This finding did not support a previous study, describing that protein expression of HPLC purified circRNA (based on size exclusion) was superior to that of gel-eluted circRNA, translated eukaryotic cells²⁸¹. It could be probable that these observed differences in translation efficiency were associated to the difference of RNA purification methods and expression systems.

For further understanding, the HPLC-purified circRNA was added at an increasing concentration to the treated-RRL. While saturating concentration was quickly reached (Figure 6. 13), the lowest concentration for the detection limit was twofold higher with circRNA (1 pmol) than that of the linear RNA (0.5 pmol) (Figure 6. 9 and 6. 13). Deeper investigations are required to better understand this observation since many factors such as translation factors or the salts concentration could interfere with the translation efficacy.

One hypothesis could be that circRNA here was mimicking the “sponge” role, as it was described to bind to several miRNA molecules simultaneously²⁶⁵. By analogy, it could be speculated that circRNA behaved like a sponge for the translation apparatus factors which made them less available in the extract for translation. To have a better understanding of the ability of circRNA to engage and bind ribosomes, a sucrose gradient experiment could be performed.

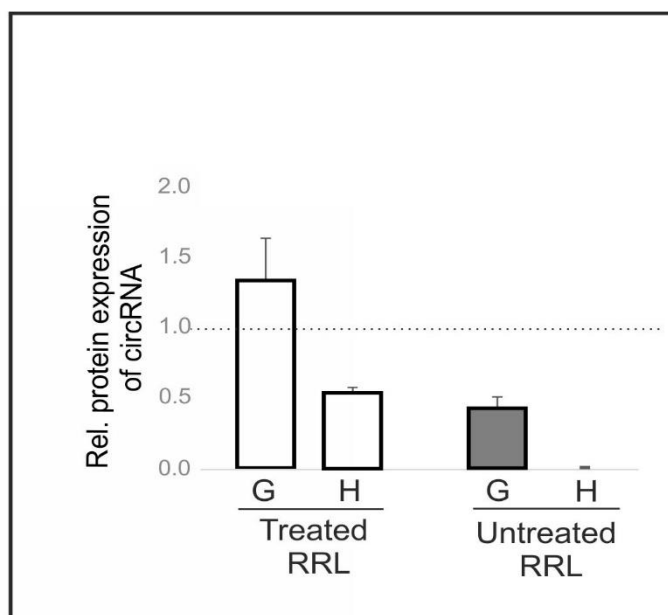


Figure 3. 21: Translation of circular RNA purified *via* HPLC or by gel-elution.

The translation efficiency of circRNA was tested in treated and untreated RRL. CircRNA purified either by gel-elution (PAGE) (G) or HPLC (H) was programmed for translation and the produced EGFP protein was quantified relative to that translated unmodified linear RNA. (All data are the average of two independent experiments \pm).

3.4.7 Translation of modified-circRNA

The influence of m^6A inserted internally in the circRNA sequence was investigated. Therefore, linear mRNA was synthesized by randomly replacing adenosines with m^6A , with increasing ratios of $m^6A:A$ (2.5%, 30%, 60%, and 100%), during *in vitro* transcription and was circularized as described previously. Gel-purified modified-circRNA was subjected to translation in an untreated RRL system. As illustrated in (Figure 3. 22), translation efficiency was comparable to the unmodified circRNA with low modified circRNA (with m^6A lower than 60% in IVT),, while increasing m^6A content was translation inhibitory. Furthermore, the full substitution of adenosines with m^6A repressed protein synthesis, similar to results obtained earlier with modified-linear RNA (containing or not an IRES structure in section 3.3.1 and 3.3.2). It appeared that circRNA was unlikely to benefit from the m^6A incorporation as observed in the previous experiment with the linear RNA (Please refer to Figure 3. 13 and 3. 15). However, it remains important to assay the translation of modified circRNA using nuclease-treated RRL.

In contrast, recent study revealed that a single m⁶A (situated upstream of the ORF) in circRNA was found to strongly induce protein translation, in a cap-independent manner³¹⁸.

This stimulatory effect was not observed in this work, and supports another study reporting translation reduction from a 10% m⁶A-circRNA relative to non-modified circRNA, in living cells³⁸⁸.

Further investigations are required to have a better insight into the mechanism of translation, such as the binding factors involved, the interacting proteins, and the necessity for the EMCV IRES element for translation of circRNA in the presence of m⁶A modification.

One probable hypothesis about the negative effect could be linked to the geometrical constraints which may affect RNA-protein interactions.

The effect of the full substitution of uridines with pseudouridines in circRNA was also assessed and showed translation repression (data not shown). Interestingly, in addition to effect of pseudouridine on mRNA translation depending on the present elements (i.e IRES or cap), the shape of the RNA may also have an impact on translation.

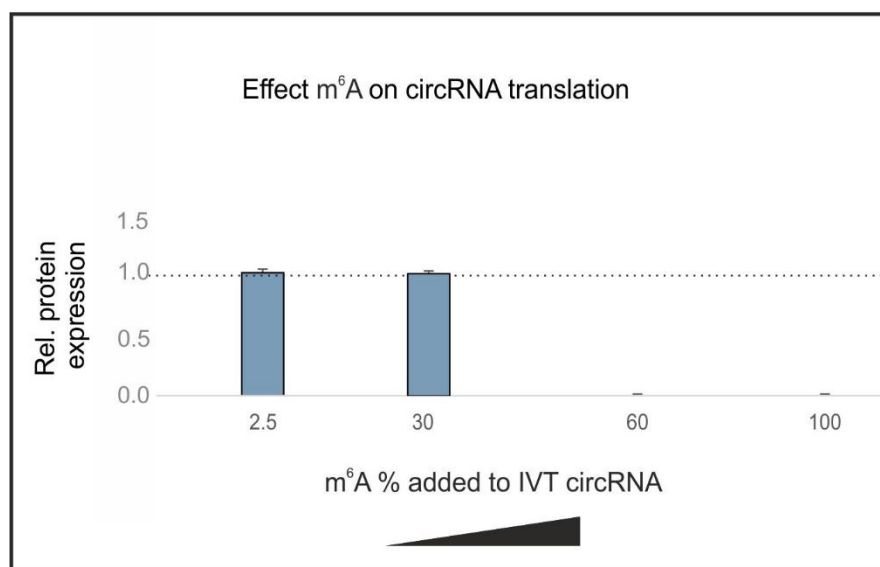


Figure 3. 22: Translation of m⁶A-modified circRNA.

Modified circular mRNA with randomly incorporated m⁶A during *in vitro* transcription (IVT) was subjected to translation in untreated RRL in the presence of S³⁵-Methionine. Quantification of EGFP protein was determined relative to that of the unmodified mRNA. (All data presented as the average of two independent experiments ±).

3.5 Application of randomly-modified mRNA: Determination of enrichment factors for modified RNA in MeRIP experiments

In this chapter, the emphasis lays on the development an approach to determine enrichment factors in MeRIP experiments.

RNA modifications were mapped in different RNA categories, such as tRNA and rRNA but also in messenger RNA (mRNA^{407,408}, or also miRNA⁹². The m⁶A is one of the most common internal modification in the mRNA and was reported for its important role in mRNA's life cycle, i.e. which plays a role in almost every stage of mRNA's life cycle, i.e. the mRNA translation³⁸¹ and degradation⁵⁸. Therefore, it was of interest to map the m⁶A modifications and uncover their biological relevance. The emphasis to study and detect m⁶A, generated several methods and the most commonly used is the MeRIP, based on a combination of immunoprecipitation with m⁶A-binding antibodies and next-generation sequencing (NGS)^{101,117,409}. The enrichment of the modified mRNA relies on the binding of a specific antibody to the methylated RNA. This provided a boost for m⁶A detection in particular in two seminal^{114,117}.

Although the accessibility and the availability of commercial modification-specific antibodies, there is a lack of data reporting the efficiency of the enrichment. Additionally, suspicion for off-targets was reported for various binding of antibodies⁴¹⁰. In addition to m⁶A detection, antibodies could detect also m⁶A harbouring an additional base modification (m⁶A_m) at lower abundance¹⁰¹.

This raises a pertinent question regarding the selectivity of IgG molecule to recognize chemical modification within the nucleic acid over unmodified one, based only on a single unique methyl group. From a dimensional point, one could evaluate the additional mass of one methyl group (14 Dalton) in an RNA fragment from 25 to 100 KDa (75-300 nucleotides) interacting with an IgG molecule of 150 KDa (Figure 3. 23). Considering the spatial distance between Fab copies, it is more probable to bind several modification sites, which could be a bias in the case of an RNA containing two or more modifications located in proximity.

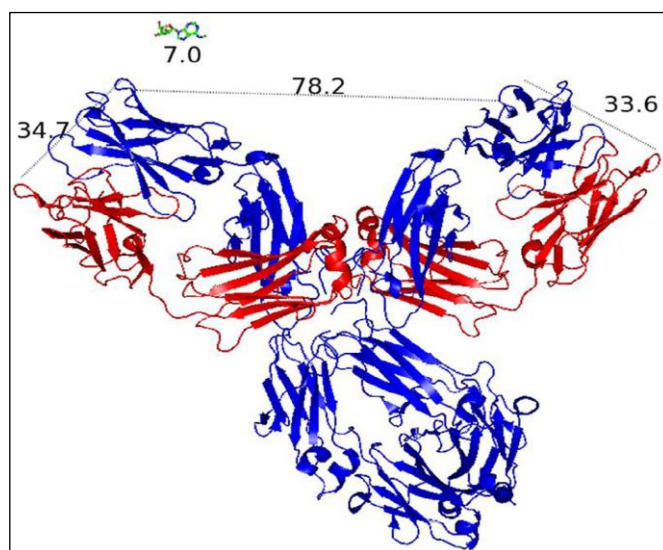


Figure 3. 23: Size comparison (Å) between potential binding surface of an IgG molecule and m⁶A residue.

For the color code, heavy chains are colored in blue and light chains of an IgG molecule are colored in red (PDB ref. 5dK3). Figure adapted from ¹⁴⁴.

These questions were addressed by the development of an assay to assess the effectiveness of the modification-antibody binding *via* enrichment factors after MeRIP. This was simplified by using a commercial antibody against m⁶A contained in the mRNA sequence. The influence of the m⁶A-content on MeRIP was also considered.

3.5.1 Method design and analysis

The RNA used in this work was prepared by *in vitro* transcription (IVT) using a defined sequence to minimize the sequence influence. Modified and unmodified mRNA were prepared in two separated transcription reactions (in the presence of not of m⁶ATP) using T7 RNA polymerase as described in section 3.1.1. A very important step is the radioactive labeling to distinguish modified from unmodified mRNA for relative quantification. In such, $\alpha^{32}\text{P}$ -UTP was introduced to label modified RNA, while unmodified was labeled with $\alpha^{32}\text{P}$ -ATP, during IVT (Figure 6. 14 (Appendix)). The determination of the enrichment factor was assayed on two sets of modified mRNA were prepared by the addition of either 2.5% or 50% m⁶A to A (adenosine) in the *in vitro* transcription reaction. This permits to adjust the average distance

between the m⁶A sites, which can be short or long with a Gaussian distribution. When 50% of adenosines are replaced with m⁶A, one out of two adenosine may be an m⁶A. Both preparations were mixed in equal amounts in µg and served as input to immobilized antibodies. Since the most commonly used antibodies in MeRIP of m⁶A research are commercial, the monoclonal mouse antibody from the synaptic system was used for the immunoprecipitation.

The recovered RNA was then treated with the nuclease NP1 to generate 5'-monophosphate mixtures allowing to conserve the radioactive labels were on the 5'-end (i.e. α-³²P-AMP and α-³²P-UMP). Subsequent analysis on a 1D TLC facilitated the recording of the positions of the radiolabelled nucleotides upon exposure of the TLC plates to phosphorImager screens (Figure 3. 24 A). A typical TLC revealed two large dark spots corresponding to the two ³²P-5'-monophosphate nucleotides which were mapped to a reference map (Figure 3. 24 B). This separation permitted to quantify the ratio between both radioactive signals which was used to determine the enrichment factor of the modified relative to the unmodified RNA.

The equal intensity spots of modified and unmodified mRNA validated the input mixture homogeneity (input, Figure 3. 24 A) and the enrichment factor could be determined as resumed by the equation.

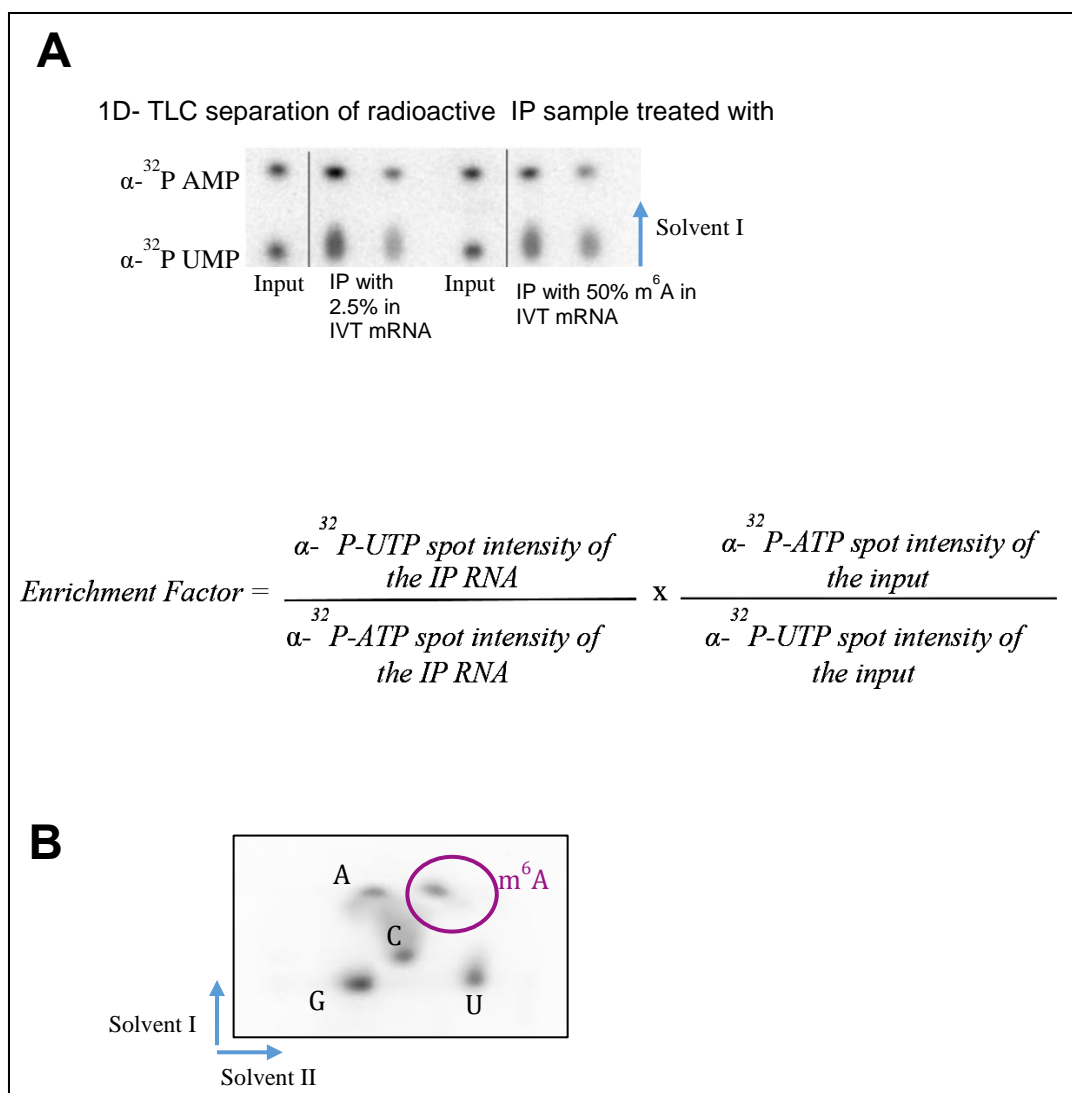


Figure 3. 24: TLC analysis and enrichment factor determination.

(A) After the immunoprecipitation, the recovered antibody-bound RNA was subjected to NP1 digestion and radioactive 5'-monophosphate nucleosides were separated on 1D TLC. The intensity of the spots permitted the quantification of the enrichment factor of an antibody binding m^6A modified mRNA (as resumed by the equation). A 2D chromatographic mobility of radioactive 5'-monophosphate nucleosides. (B) After the digestion of mRNA containing 50% m^6A by RNase T2 in order to label all the modified and unmodified bases in their 3' side allowing to identify m^6A spot Figure adapted from ¹⁴⁴.

3.5.2 Determination of the enrichment factor of modified RNA

After immunoprecipitation, the bound mRNA fraction was extracted and analyzed as explained in the previous section, revealing that the m^6A antibody selectively bound to modified mRNA as compared to the input (Figure 3. 25). The binding was stronger with mRNA containing 50% m^6A with an enrichment factor of 2 fold higher than the input.

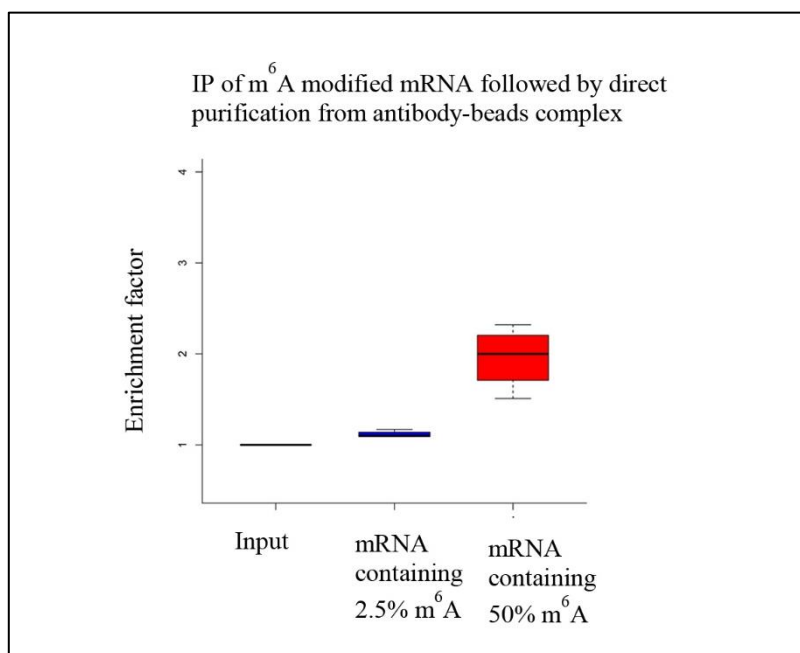


Figure 3. 25: Quantification of enrichment factor after MeRIP.

Immunoprecipitation was performed using a hot mixture of modified (low and highly modified-m⁶A mRNA) and unmodified mRNA. Modified mRNA was prepared by the addition of either 2.5% or 50% m⁶ATP to the *in vitro* transcription. Subsequently, bound mRNA was purified by TRI reagent extraction from antibody-beads fraction and prepared for 1D TLC analysis. The intensity of the revealed spots was normalized to the input and allowed the determination of the enrichment factor of antibody binding m⁶A modified mRNA. Error bars are expressed as means \pm SD of triplicates). Figure adapted from ¹⁴⁴.

Further optimizations of the method were explored by using different binding competitors. When the binding was competed by an excess of free m⁶A, the enrichment factor of low modified samples (2.5% m⁶A) was increased as compared to the direct extraction of bound mRNA from the beads-antibody, from 1.3 fold to almost 2 fold (Figure 3. 26, A green). A higher enrichment reaching 4 fold was obtained with 50% m⁶A modified mRNA as compared to the input (Figure 3. 26, B green).

The elution with free non-modified adenosine A was also assayed. As depicted in (Figure 3. 26, C and D green), most of the modified RNA was enriched in the antibody-beads complex fraction with an enrichment factor of 4 relative to the input, in the contrast to RNA elution with m⁶A. Moreover, the enriching was independent of the level of modification incorporation in mRNA.

Collectively, these data confirmed the selectivity of the antibody to detect m⁶A within an RNA mixture pool and provided an evaluation of MeRIP protocols. Since the m⁶A profiling has been shown to be 0.1%–0.5% of total adenosine residues in cellular RNA ^{86,104,410} and a typical

mRNA may contain 100–300 adenosines (including the polyA tail), the enrichment of m⁶A containing RNA fragments in the order of 4 times, as suggested by the findings here, would explain the functioning of m⁶A-MeRIP protocols applied for targeting particular modification in a relatively “rich” modified RNA (starting material). Nevertheless, findings showed also that antibodies with much higher affinity and specificity would be required to target modifications that are low abundance in the epitranscriptome.

This chapter were published in ¹⁴⁴.

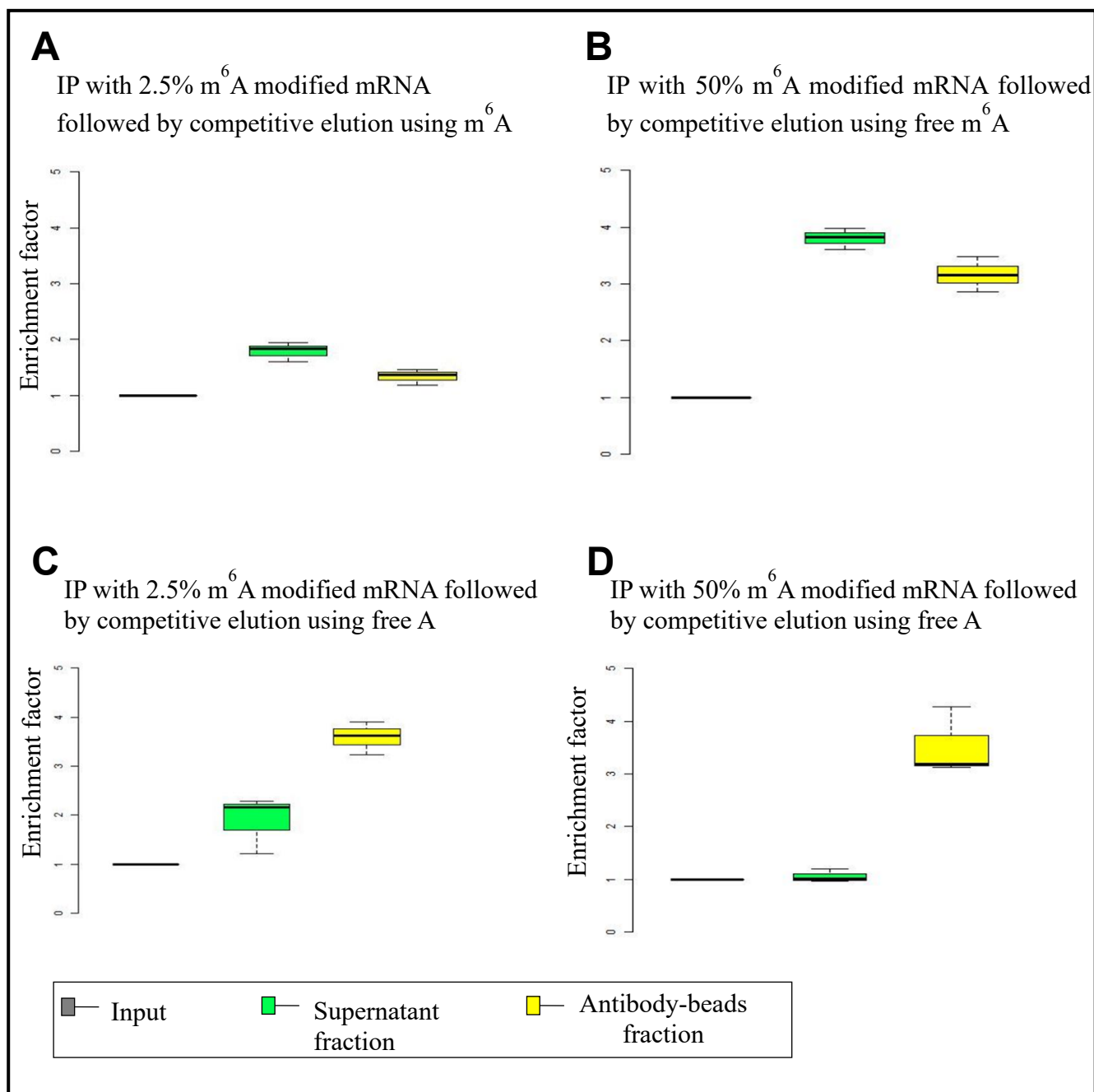


Figure 3. 26: Determination of enrichment factor for optimized IP.

Immunoprecipitation (IP) was performed using low and highly m⁶A-modified mRNA (2.5% and 50% m⁶ATP added to the IVT). Afterwards, the bound mRNA fraction was eluted by the addition of an excess of free m⁶A (**A**, **B**) or free Adenosine (**C**, **D**). The eluted fraction recovered from the supernatant was applied for ethanol precipitation (green), meanwhile, the RNA bound to antibody-beads fraction was extracted by a solvent (yellow). The radioactive intensity of modified IP RNA to unmodified RNA in each fraction permitted the quantification of the enrichment factor, as compared to the input (RNA mixture pool prior to IP). Error bars are expressed as means \pm SD of triplicates. Figure adapted from ¹⁴⁴.

Results and discussion

4 Conclusion and perspectives

4.1 Translation-tuning by modified mRNA

The influence of modifications on mRNA translation was investigated by the development of a systematic approach. The design and synthesis using an optimized single-step ligation (called here 3-way-one pot), initially described to ligate smaller RNA fragments. This allowed to synthesise a large single point-modified IRES-EGFP mRNA molecule (>1.3 kb), by offering the advantage to insert modifications at eventually any desired site. The sequence around the start codon was targeted to introduce modifications, namely the 2'-O-Me and pseudouridine. The synthesis of point-modified and unmodified RNA (called here "5'EGFP-lig. block") showed a comparable ligation efficiency. A critical step was the recovery of the ligation product. This was carefully addressed and optimized, as big gel particles were observed after an elution from conventional agarose gel. Such residuals could disturb the sample purity and further applications, i.e. RNA loading on an HPLC column. An analytical study was conducted by LC-MS and RiboMethSeq which successfully validated the presence of modifications in the final point-modified mRNA products. The analysis revealed also possible detection of other unexpected modifications with certain samples (supposed to contain m⁶A). These contaminations would presumably be originated from the phosphoramidite building block synthesis (those constructs were discarded from this study). Therefore, it is highly recommended for other scientists to make accurate verification of the purchased oligonucleotides. Altogether, the presented parameters in this work can be considered as prerequisites for the successful preparation to reconstitute a large point-modified mRNA out of three fragments, which was reproducible and up scalable. As a proof of concept, another construct was generated using the same mRNA reporter, in order to introduce modification in the middle of the coding sequence (Internal-lig. Block).

Since investigations from several other studies on the impact of mRNA modifications concerned internal sites in the coding sequence (CDS), near the stop codon or the UTR, the present study, here, strived to get a better understanding on modifications inserted at essential sites for translation initiation (start codon), besides to an internal insertion in the coding sequence. The overall effect of single modifications on protein production revealed that modifications finely tune translation in a position- and codon-dependent manner.

It was anticipated that replacing the 2'-OH in the second position of a codon with a bulky methyl group may potentially reduce or abolish translation^{140,176,185}. In this work, the insertion of a 2'-O-Me at the first or the third position from internal codons in CDS, did not generate a pattern. The translation was either slightly affected or stimulated. Meanwhile, the insertion of the 2'-O-Me in the first position from start codon was translation inhibitory. Strikingly, when 2'-O-Me was in the second position from the start codon, translation was dramatically stimulated to almost 2 fold (as compared to the unmodified RNA). To the author's knowledge, this may be the first report of such a relevant effect and suggested a new potential of this position in modulating translation efficiency. Moreover, the obtained effect could not be recapitulated by the insertion of another modification (pseudouridine). The insertion of a single Ψ at the second position from the first codon, reduced protein level production (to at least 75%). This inhibitory effect was consistent with another study (only 50% of translation reduction), where Ψ was positioned at the second position of a different codon in the CDS¹⁴⁰. Thereby, pseudouridine may act in position-dependent manner. Altogether, results here suggest a difference of the influence of 2'-O-Me and also pseudouridine when positioned in the coding sequence or in the start codon, in a position-dependent manner.

This negative impact of pseudouridine was also additive, with full substitution of U by Ψ in IRES-EGFP mRNA, which reduced translation to a minimum (~10%). Previous data reported on full substitution of U: Ψ opened a debate in literature. Protein level from fully Ψ -modified transcripts was negatively affected when translated in an *in vitro* prokaryotic translation system and the wheat Germ extract^{140,221}. Meanwhile, a stimulatory effect could be obtained using eukaryotic translation systems as the RRL or upon transfection in living cells²²¹.

Another point that was revealed by this work, was that fully Ψ -modified mRNA could fine-tune protein expression depending on the RNA structural features. While the addition of a cap structure in Ψ -modified mRNA promoted translation, the presence of an IRES sequence was inhibitory. Such a finding could be anticipated, as modifications may disrupt the secondary structure required for a functional IRES.

Based on these results, the role of another modification (m^6A) to modulate translation was also studied. As compared to Ψ -modified mRNA, the full substitution of adenosines with m^6A hindered translation of all mRNA constructs, independently from the present features/elements. Thus, modifications and the RNA structures/elements, i.e. IRES, could both modulate protein expression in the present system. The obtained data from these comparative findings would add another also another layer to the existing discoherency of findings from previously reported

studies. Indeed, several groups showed that mRNA modification's effect on protein expression may vary depending on (i) the used translational system, and (ii) the sequence context^{106,169,171,221,374}. Thus, it becomes difficult to generalize conclusions and more future investigations are required.

Altogether, findings here expanded point-modification studies, earlier studied in canonical translation context, to IRES-dependent translation mRNA. Future experiments may be devoted to further investigate, in particular, the modifications inserted in the start codon using a reporter without an IRES.

Since it was previously reported that modified-mRNA translation could also vary between *in vitro* translation systems from eukaryotic²²¹, it could be an alternative to use living cells which may offset this ambiguity. Modification insertion may have consequences on mRNA translation at different levels, as interruption of translation, ribosomal stalling and rewiring of the genetic code^{169,170}. Such events might be considered to induce amino acid miss-incorporation and thus the loss of protein fluorescence activity. Therefore, it can be of interest to combine functional study by in-gel fluorescence, side to side with protein production assessed by S³⁵-Methionine incorporation or western-blot to have a global view on the production of full-length and functional proteins. For a refined and more detailed research, purification of the produced protein, by immunoprecipitation for instance, and a subsequent mass spectrometry can provide a closer look on the identity of the protein sequence. More work is required to understand the mechanism behind it, addressing the influence of modifications on ribosomal association to form polysomes and possible existence of ribosomal stalling could be performed by toeprinting experiments.

Whereas 2'-O-Me at second position in codon in the CDS had a conservative inhibitory effect, an irregular pattern was obtained with 2'-O-Me at position 1 and 2. One might question the existence of an exclusivity position or a position-codon dependent modification effect. One strategy might be also by introducing modification (i.e. 2'-O-Me) at the same position/codon but at different sites within the coding sequence, or by modifying the same codons in a different RNA reporter sequences.

The established work here provided a versatile workflow and a valuable tool that can be exploited to explore the influence of modifications in other sites. As a proof-of-principle, it can be envisaged in future to study influence of modifications on translation elongation, using the successfully reconstituted internal ligation block mRNA.

Preliminary results obtained on single-modified IRES-EGFP mRNA, harbouring two close start codons (separated by 3 codons), resulted in the differential expression of two fluorescent translation products that could be separated on SDS-PAGE (as illustrated by the overview scheme). It is possible to hypothesize that a single modification can direct translation on either start codons. In a first attempt, a single pseudouridine was inserted at the second position from the downstream AUG start codon and led to a barely detectable translation product, which was translated from the upstream start codon. More detailed investigation and control experiments are required, and future experiments may include the insertion of other modifications as the m⁶A, as well in critical positions in the Kozak motif.

Lastly, given the distribution of various modifications along the mRNA sequence, it can be interesting for future research to make a combination of two point modifications, different or same modifications, and further study their capacity to collectively modulate protein expression, as well as their immunogenicity aspect upon cellular transfection.

Altogether, this work could expand the knowledge about modification involvement on protein expression that can be utilized as regulators of ribosomal translation. Considering the illustrated suggestions, this can be a feat evolve utility of the developed point-modified concept mRNA to another level.

4.2 Circular RNA

Findings here showed the possibility to generate a large and functional circular EGFP RNA (circRNA) with >1.3Kb. Notably, incorporating random modifications (m⁶A), from few nucleoside triphosphates to a full substitution of adenosines, did not perturb circularization efficiency. The validation of RNA circularity was initially addressed by harnessing the use of a RP-HPLC. Indeed, it allowed a clear separation between circular and linear mRNA forms. Circularization efficiency is underreported in literature. Otherwise, it depended on the followed circularization strategy and the RNA length^{310,388}. While the ligation was highly efficient (exceeding 50%), the recovery by yield either HPLC or gel-elution of the produced circRNA could be estimated to 10-20% of the initial material.

Since the discovery of circRNA, only few studies started to elucidate their coding capacity. Here, applying the engineered IRES-EGFP circRNA for translation in untreated RRL system, revealed that the EMCV IRES could drive translation. This suggested a potential alternative to recruit circRNA for translation instead of linear mRNA. Investigations on the effect of the

purification method revealed that gel-eluted circRNA elicited greater levels of protein production than circRNA purified by HPLC. One possibility could consider the stability of circRNA purified by either methods. Moreover, a time point study may be performed in future to explore if HPLC purified circRNA does not delay the translation initiation.

Moreover, translation efficiency from circRNA was strongly dependent on the used translation system (nuclease-treated commercial and nuclease-untreated RRL) and requires further investigations. Therefore, it can be of interest to prepare more controls and explore the effect in the context in other eukaryotic cell-free systems and by cellular transfection.

While partial incorporation of m⁶A stimulated the translation of linear mRNA, circRNA did not benefit from this effect. One may hypothesize the ability of modified-circRNA to assemble ribosomes for translation initiation. More work is required to have a better understanding on the modification effect, in particular by the circularization of the point-modified mRNA which would open new perspectives for the understanding of modification regulation in circular mRNA and as well in the linear counterpart.

5 Materials and methods

5.1 Materials

5.1.1 Instruments

Balances

Mettler Toledo Excellence Plus	Mettler Toledo (Gießen, Germany)
Sartorius Cubis Analytical Balance	Sartorius (Goettingen, Germany)
Mettler Toledo PM460	Mettler Toledo (Gießen, Germany)

Mixing & incubators

Thermoshaker Plus	Eppendorf (Hamburg, Germany)
VWR Digital Heatblock	VWR International (Radnor, USA)
GFL Schüttelinkubator 3032	Gesellschaft für Labortechnik (Burgwedel, Germany)
BIOER ThermoCell	BIOER (Hangzhou, China)
IKA RH basic 2	IKA-Werke GmbH & CO. KG (Staufen, Germany)
Heareus BB15	Thermo Fisher Scientific (Waltham, USA)
Ecotron	InForst HT (Swiss)
Vortex Mixer (7-2020)	neoLab (Heidelberg, Germany)

Centrifuges

Eppendorf Centrifuge 5810	Eppendorf (Hamburg, Germany)
Sprout mini-centrifuge	Biozym (Hessisch Oldendorf, Germany)
1 - 15 PK Sigma	Sigma (Osterode am Harz, Germany)
Avanti J25	Beckmann Coulter (Krefeld, Germany)
Eppendorf Centrifuge 5424 R	Eppendorf (Hamburg, Germany)

HPLC & columns

Agilent 1100 HPLC system	Agilent (Böblingen, Germany)
Agilent 1260 Infinity LC	Agilent (Santa Clara, USA)
YMC-Triart C8 HPLC column (3 µm particle size, 120 Å pore size, 150 mm length, 3mm inner diameter)	YMC Europe GmbH (Dinslaken, Germany)

Mass spectrometry

Agilent 6460 triple quadrupole	Böblingen (Germany)
--------------------------------	---------------------

Appendix

ESI; Micromass LCT	Waters (Eschborn, Germany)
Peak scientific GENIUS XE	Peak Scientific Instruments GmbH (Düren, Germany)
NITROGEN	

Miscellaneous

Agarose gel chamber	PeqLab (Erlangen, Germany)
Dark Reader Blue Light Transilluminator	Clare Chemical Research (USA)
LSG-400-20 NA vertical chamber	C.B.S. Scientific (San Diego, USA)
Consort EV232 power supply	Consort (Turnhout, Belgium)
Model 250/2.5 power supply	BioRad (München, Germany)
Typhoon TRIO+	GE Healthcare (Chicago, Illinois, USA)
Shaker (DOS-10L)	neoLab (Heidelberg, Germany)
Mini-PROTEAN® Tetra Vertical electrophoresis Cell	Bio-Rad (Feldkirchen, Germany)
Malvern NanoSight LM10	Malvern Panalytical (Germany).
pH measurements	
FiveEasy™ FE20 pH meter	Mettler Toledo (Gießen, Germany)
Pipetting	
Micropipettes (Pipetting Discovery comfort 2, 10, 20, 100, 200 and 1000 µL)	Abimed (Langen, Germany)
Pipette boy (Integra)	VWR (Darmstadt, Germany)
Ultrapure water purification system	MilliQ Millipore (Schwalbach, Germany)
Spectrophotometer	PeqLab (Erlangen, Germany)
NanoDrop ND 2000	
phosphorimager	GE Healthcare
Ultrapure water purification system	Millipore (Schwalbach, Germany)
MilliQ	
thin-layer plates fluorescent (20 × 20 cm, 0.1-mm layer) from Merck	Merck (Darmstadt, Germany)

5.2 Chemicals & Consumables

Gel electrophoresis and analysis

Bromphenol blue	Merck (Darmstadt, Germany)
Formamide	Carl Roth (Karlsruhe, Germany)
Ammonium persulfate	Carl Roth (Karlsruhe, Germany)
TEMED	Carl Roth (Karlsruhe, Germany)
Rotiphorese sequencing gel diluent)	Carl Roth (Karlsruhe, Germany)
Rotiphorese (10x) TBE buffer	Carl Roth (Karlsruhe, Germany)
Rotiphorese Gel 40% acrylamide mix (19:1)	Carl Roth (Karlsruhe, Germany)
Rotiphorese sequencing gel concentrate	Carl Roth (Karlsruhe, Germany)

Appendix

Rotiphorese® Gel B	Carl Roth (Karlsruhe, Germany)
Rotiphorese® Gel B (30 % acrylamide solution)	Carl Roth (Karlsruhe, Germany)
Rotiphorese 10x SDS-PAGE	Carl Roth (Karlsruhe, Germany)
Agarose	Biozym (Germany)

Staining solutions

SYBR®Gold nucleic acid gel stain 10000x	Thermo Fisher Scientific (Frankfurt, Germany)
GelRed 3x	Biotium (Hayward, USA)

Protein analytics

Tris-HCl	Carl Roth (Karlsruhe, Germany)
30% Acrylamide/Bis Solution, 29:1	Bio-Rad (Feldkirchen, Germany)
Rotiphorese sequencing gel buffer concentrate	Carl Roth (Karlsruhe, Germany)
TWEEN 20	Sigma Aldrich (Steinheim, Germany)
SDS	Roth (Karlsruhe, Germany)
Tris-HCl	Carl Roth (Karlsruhe, Germany)
tRNA fraction	
Rabbit Reticulocyte Lysate Nuclease-untreated	Gift from Dr. Franck Martin, Strasbourg
Rabbit Reticulocyte Lysate System, Nuclease Treated	Promega (Walldorf, Germany)

Bacterial culture and transformation

LB broth (Lennox)	Sigma-Aldrich (Steinheim, Germany)
LB broth with agar (Lennox)	Sigma Aldrich (Steinheim, Germany)
Petri dishes	Sarstedt (Nümbrecht, Germany)

Antibiotics:

Kanamycin	Fluka (Buchs, Germany)
Ampicillin	Roth (Karlsruhe, Germany)

Subcloning Efficiency™ DH5α competent Cells	Invitrogen™ (Germany)
---------------------------------------------	-----------------------

RNA/DNA isolation & EtOH precipitation and chemicals

Ammonium acetate	Merck (Darmstadt, Germany)
Glycogen	Thermo Fisher Scientific (Frankfurt, Germany)
Ethanol >99.5%	Carl Roth (Karlsruhe, Germany)
Roti-aqua-phenol (RNA)	Carl Roth (Karlsruhe, Germany)
Roti-phenol (DNA)	Carl Roth (Karlsruhe, Germany)

Appendix

Isobutyric acid	Sigma-Aldrich (Steinheim, Germany)
TRI Reagent®	Sigma-Aldrich (Steinheim, Germany)
Chloroform HPLC grade	Sigma-Aldrich (Steinheim, Germany)
Isobutyric acid	Sigma-Aldrich (Steinheim, Germany)
Ammonia solution (25%)	Merck (Darmstadt, Germany)

RNA *in vitro* synthesis

Dithiotreitol (DTT)	Thermo Fisher Scientific (Waltham, USA)
ATP ($\geq 90\%$, lyophilized)	Carl Roth (Karlsruhe, Germany)
GTP (1g, $\geq 90\%$, lyophilized)	Carl Roth (Karlsruhe, Germany)
UTP (100 mg, $\geq 90\%$, lyophilized)	Carl Roth (Karlsruhe, Germany)
CTP (1g, $\geq 98\%$, lyophilized)	Carl Roth (Karlsruhe, Germany)
Guanosine 5'-monophosphate	Sigma-Aldrich (Steinheim, Germany)
Bovine Serum Albumin (BSA)	Thermo Fisher Scientific (Waltham, USA)
N ⁶ -Methyl-ATP (100 mM)	Jena Bioscience (Jena, Germany)
Pseudo-UTP (100 mM)	Jena Bioscience (Jena, Germany)
$\alpha^{32}\text{P}$ -ATP (800 Ci/ mmol)	Hartmann analytic (Braunschweig, Germany)
$\alpha^{32}\text{P}$ -UTP (800 Ci/ mmol)	Hartmann analytic (Braunschweig, Germany)
N ¹ -Methylpseudo-UTP (100 mM)	Jena Bioscience (Jena, Germany)
Triton X-100	Sigma Aldrich (Steinheim, Germany)

HPLC

Triethylammonium acetate HPLC grade	Sigma Aldrich (Steinheim, Germany)
Acetonitrile HPLC grade	Honeywell (Morris Plains, USA)
Acetic acid LCMS grade	Sigma Aldrich (Steinheim, Germany)

5.3 Disposables

Nanosep MF Centrifugal Devices (0.45 μm)	Pall (New York, USA)
Eppendorf tubes (1.5, 2.0 mL)	Roth (Karlsruhe, Germany)
Falcon® tubes (15, 50 mL)	CellStar (Frickenhäusen, Germany)
0.2 μm solid phase filters	(Nanosep centrifugal device, Pall, USA)
Disposable serological pipettes	Sarstedt (Nümbrecht, Germany)
Petri dishes	Sarstedt (Nümbrecht, Germany)
Rotilabo-syringe filters (0.22 μm)	Carl Roth (Karlsruhe, Germany)
Pipette tips (10, 20, 100, 200, 1000 μL)	Carl Roth (Karlsruhe, Germany)

5.4 Enzymes, reagents and kits

T4 RNA ligase II (10 U/μL)	New England Biolabs (Germany)
DNase I (50 U/μL)	Thermo Fisher Scientific (Frankfurt, Germany)
T7 RNA Polymerase	self-prepared in our lab (AK Helm)
T7 RNA Polymerase (20 U/μL)	Thermo Fisher Scientific (Frankfurt, Germany)
T4 Polynucleotide kinase enzyme (10 U/μL)	Thermo Fisher Scientific (Waltham, USA)
RNasin® Ribonuclease Inhibitor	Promega (Walldorf, Germany)
RNase R (20 U/μL)	Epicenter (Madison, WI, USA)
Cfr42I (SacII) (10 U/μL)	Thermo Fisher Scientific (Frankfurt, Germany)
BamH I (10 U/μL)	Thermo Fisher Scientific (Frankfurt, Germany)
BcuI (SpeI) (10 U/μL)	Thermo Fisher Scientific (Frankfurt, Germany)
Vaccinia Capping System	New England Biolabs (Frankfurt am Main, Germany)
SmaI (20 U/μL)	New England Biolabs (Frankfurt am Main, Germany)
Nuclease P1 from <i>Penicillium citrinum</i> (lyophilized)	Sigma Aldrich (Steinheim, Germany)
Recombinant RNase T2	MoBiTec (Göttingen, Germany)
QIAGEN Plasmid Maxi Kit	Qiagen (Hilden, Germany)
MegaClear™ Kit	Thermo Fisher Scientific (Frankfurt, Germany)
Cap 2'-O-Methyltransferase (50 U/μL)	New England Biolabs (Frankfurt am Main, Germany)
m ⁷ G(5')ppp(5')G RNA Cap Structure Analog (40 mM)	New England Biolabs (Frankfurt am Main, Germany)
Protein G Sepharose™ 4 Fast Flow	GE Healthcare (Dornstadt, Germany)
m ⁶ A mouse monoclonal antibody	Synaptic Systems (Göttingen, Germany)
Vaccinia Capping System (10 U/μL)	New England Biolabs (Frankfurt am Main, Germany)

5.5 Buffers and solutions

All buffers and media were prepared either by using MilliQ-water or, if required for bacterial culture, with RNase/DNase- and endotoxin-free water (Zymo Research).

Buffer/solution

Composition

5x Straßbourg buffer (transcription buffer)	40 mM Tris-HCl (pH 8.1), 1 mM spermidine, 5 mM DTT, 0.01 % Triton X-100
---------------------------------------------	-------------------------------------------------------------------------

Appendix

PAGE loading buffer, denaturing	1x TBE (from 10x) and 90 % v/v formamide in water
PAGE loading buffer, denaturing, blue	1x TBE (from 10x), 90 % v/v formamide, 0.1 % xylene cyanol, and 0.1 % bromphenol blue in water
10 x PBS	1.37 M NaCl, 27 mM KCl, 17 mM KH ₂ PO ₄ and 100 mM Na ₂ HPO ₄ (pH 6.8)
RNA-IP buffer	150 mM LiCl, 0.5% NP-40, 10 mM Tris-HCl at pH 7.5
Wash buffer I	RNA-IP buffer plus 150 mM LiCl
Wash buffer II	RNA-IP buffer plus 300 mM LiCl
TEAA buffer (1 M stock)	1 mol trimethylamine, 1 mol acetic acid in 1 L Milli-Q water, pH 7.0
HPLC solvent A	1:10 dilution of stock buffer 1 M TEAA in Milli-Q water (pH 7.0)
Bacterial culture LB-medium	20 g LB-Broth mix were dissolved in 1 L distilled water, then autoclaved
Bacterial culture LB-agar	17,5 g LB-agar was added to 500 mL distilled water and autoclaved
Denaturing PAGE mixtures for the separation of nucleic acids	<p>For 100 mL 4% denat. PAGE: 16 mL of gel concentrate (25% solution (19:1)), 74 mL gel diluent and 10 mL sequencing gel buffer. For polymerization 400 µL 10% APS ammonium persulfate and 50 µL Tetramethylethyldiamin (TEMED). The gel was casted using 1 mm spacers and one 1 mm comb. After polymerization, lower spacer and the comb were removed, pockets were rigorously rinsed with 1x TBE, before sample loading.</p> <p>For the validation of circular RNA, the method was adapted from ²⁷⁸ with slight modifications and run was achieved in a one dimension run. A 100 mL of 4% PAGE included acrylamide:N,N'-methylene</p>

Appendix

bis-acrylamide ratios: 19:1, 39:1 and 59:1, in 1× TBE and 8 M urea.

10% SDS-PAGE for proteins

10% resolving gel: 3.3 mL Acrylamide/bis (30%, Bio-Rad), 100 µL 10% SDS, 3.75 mL, 1.5 M Tris HCl (Adjust to pH 8.8 with 6 N HCl), 2.85 mL distilled water, 10 µL TEMED and 60 µL of 10% APS.

4.5% Stacking gel: 750 µL Acrylamide/bis (30%), 650 µL 0.5 Tris HCl (Adjust to pH 6.8 with 6 N HCl), 3.56 mL distilled water, 50 µL of 10% SDS, 50 µL TEMED and 50 µL of 10% APS.

Solvent I

Solvent I was prepared in the following proportions: preparation: isobutyric acid_concentrated ammonia_- water (66_1_33 [v_v_v]). For example: in a final 500 mL volume, 330 mL of isobutyric acid (Sigma-Aldrich) was mixed with 5 mL of ammonia solution (25%, Merck) and 165 mL of ultrapure water (MilliQ).

Solvent II

isopropanol:conc.HCl:water (68:18:14) [v:v:v].

5.6 Softwares

CorelDRAW X7 Corel Corporation (Ottawa, Canada)

CorelDRAW V5 2020 Corel Corporation (Ottawa, Canada)

IrfanView created by Irfan Skiljan

Microsoft Office (Excel, Powerpoint, Access), Microsoft

ImagJ V5 (Madison, USA)

Typhoon scanner software GE Healthcare (Buckinghamshire, UK)

Ape V5

ChemBioDraw Ultra 12.0 CambridgeSoft/PerkinElmer (USA)

Mendeley Desktop 1.17.6 Mendeley Ltd. (London, UK)

GraphPad Prism 7.0 - statistical analysis GraphPad Software (San Diego, USA)

5.7 Oligonucleotides (all nucleotides were supplied from IBA (Göttingen))

Table 1 : Unmodified RNA oligonucleotides (middle fragment, fragment 2).

Name	Details	Sequence (5' to 3')
MH905	5'EGFP-lig. block	CCACAACCAUGGUGAGCAA
MH903	5'IRES-lig. block	UAACGUUACUGGCCGAAGCCGCUU

MH906	internal-lig block	UGGUGAACCGCAUCGAGCUGAA
-------	--------------------	------------------------

Table 2 : Modified RNA oligonucleotides (middle fragment, fragment 2).

Name	Details	Sequence (from 5' to 3')
MH 960	5'EGFP-lig. block	CCACAACCAUGG(X)GAGCAA, X= Ψ
MH 963	5'EGFP-lig. block	CCACAACCA(X)GGUGAGCAA, X= Ψ
MH 1052	5'EGFP-lig. block	CCACAACC(X)UGGUGAGCAA, X= A _m
MH 1053	5'EGFP-lig. block	CCACAACCA(X)GGUGAGCAA, X=U _m
MH 1054	5'EGFP-lig. block	CCACAACCAU(X)GUGAGCAA, X=G _m
MH 1055	5'EGFP-lig. block	CCACAACCAUG(X)UGAGCAA, X=G _m
MH 1056	5'EGFP-lig. block	CCACAACCAUGG(X)GAGCAA, X= U _m
MH 1057	5'EGFP-lig. block	CCACAACCAUGGU(X)AGCAA, X= G _m
MH 1058	5'EGFP-lig. block	CCACAACCAUGGUG(X)GCAA, X=A _m
MH 1059	5'EGFP-lig. block	CCACAACCAUGGUGA(X)CAA, X= G _m
MH 1060	5'EGFP-lig. block	CCACAACCAUGGUGAG(X)AA, X= C _m
MH 1099	5'EGFP-lig. block	CCACA(X)CCAUGGUGAGCAA, X=A _m

Table 3: Modified RNA oligonucleotides discarded from the study (please refer to 3.1.2).

Name	Details	Sequence (from 5' to 3')
MH 963	5'EGFP-lig. block	CCACAACCA(X)GGUGAGCAA, X= Ψ
MH 961	5'EGFP-lig. block	CCACAACCAUGGUG(X)GCAA, X= m ⁶ A
MH 1061	5'EGFP-lig. block	CCACA(X)CCAUGGUGAGCAA, X= m ⁶ A
MH 1062	5'EGFP-lig. block	CCACAACCAUGGUGAGC(X)A, X= m ⁶ A
MH 1082	5'EGFP-lig. block	CCAC(X)ACCAUGGUGAGCAA, X= m ⁶ A
MH 1083	5'EGFP-lig. block	CC(X)CAACCAUGGUGAGCAA, X= m ⁶ A

Table 4: DNA oligo.

Name	Details	Sequence (from 5' to 3')
MH 910	Splint DNA for 5' IRES-lig. block	TGGGCACCACCCCGGTGAACAGCTCCTC GCCCTTGCTCACCATGG
MH 911	Splint DNA for Internal-lig block	TTGCCGTCCTCCTTGAAGTCGATGCCCTT CAGCTCGATGCGGT
MH 1168	Splint DNA for 5'EGFP-lig. block	TGGGCACCACCCCGGTGAACAGCTCCTC GCCCTTGCTCACCATGGTTGTGGGATCC AAGCTTATCATCGTGTTTTTCAAAGG
MH 838	Splint DNA for circularization	TCGAGGGGGGGCCCGGTACCCAATTTCG CCGCGGTGGCGGCCGCTTTACTTGTACA GCTC
Name	Details	Sequence (from 5' to 3')
MH 964	Conjugated to cy5 dye, used for 5' IRES-lig. block	AGGGGGGGCCCGGTACCCAATTTCG
MH 955	Conjugated to fluorescein, used for 5' IRES-lig. block	CTTATTCCAAGCGGCTTCGGCCAGTAAC GTTAGGGGGGG
MH 957	Conjugated to fluorescein, used for 5'EGFP-lig. block	GCTCCTCGCCCTTGCTCACCATGGTTGTG GCCATATTATC

Table 5: pDNA vectors.

Name	Resistance gene	Supplier	Details
IRESeGFP5'-1-S1_pUC57	Ampicillin	GeneScript (Piscataway, USA)	Encoding fragment 1 used for 5'EGFP-lig. Block RNA synthesis
IRE5'eGFP-3_pUC57	Ampicillin	GeneScript (Piscataway, USA)	Encoding fragment 3, used for 5'EGFP block RNA synthesis
IRE-EMCV-S1_pUC57	Ampicillin	GeneScript (Piscataway, USA)	Encoding fragment 3, used for full-length EMCV-EGFP mRNA

Appendix

IRES5'eGFP-1 _pUC57	Ampicillin	GeneScript (Piscataway, USA)	Encoding fragment 1, used for 5' IRES-lig. Block RNA synthesis
IRES5'eGFP-3 _pUC57	Ampicillin	GeneScript (Piscataway, USA)	Encoding fragment 3, used for 5' IRES-lig. Block RNA synthesis
IRES-eGFPint2-1 _pUC57	Ampicillin	GeneScript (Piscataway, USA)	Encoding fragment 1, used for Internal-lig. Block RNA synthesis
IRES-eGFPint2-3 _pUC57	Ampicillin	GeneScript (Piscataway, USA)	Encoding fragment 3, used for Internal-lig. Block RNA synthesis
IRES-eGFPint2-3	Ampicillin	GeneScript (Piscataway, USA)	Encoding fragment 3, for Internal-lig. block
eGFP-1-UTR- Stop_PUC57	Ampicillin	GeneScript (Piscataway, USA)	Encoding full length EGFP mRNA
pPBSII IRES-eGFP plasmid	Ampicillin	Gift from Prof. Dr. Stefanie Dimmeler	Encoding for full- length wild IRES EMCV-EGFP mRNA

Table 6: RNA sequences

Name	Sequence
IRES-EGFP mRNA full length	GGGCGAATTGGGTACCGGGCCCCCTCGAGGTCATCGAATTCCGCCCTCTCCCTCCCCCCCCCTAACGTTACTG GCCGAAGCCGCTTGAATAAGGCCGCTGTGCGTTTGTCTATATGTTATTTTCACCATATGCGCTCTTTTGGCAAT GTGAGGGCCCGAAACCTGGCCCTGTCTCTTGTACGAGCATTCCTAGGGGTCTTTCCCTCTCGCCAAAGGAATGC AAGGTCTGTGAATGTCGTGAAGGAAGCAGTTCTCTGGAAGCTTCTTGAAGACAAACACGCTGTGTAGCGACCT TTGCAGGCAGCGGAACCCCCACCTGGCGACAGGTGCCTCTGCGGCCAAAAGCCACGTGTATAAGATACACCTGC AAAGGCGGCACAACCCAGTGCCACGTTGTGAGTTGGATAGTTGTGGAAGAGTCAAATGGCTCTCCTCAAGCGT ATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCATTGTATGGGATCTGATCTGGGGCCTCGGTGCACATGC TTTACATGTGTTTAGTCGAGGTAAAAAACGCTAGGCCCCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAA CACGATGATAAGCTTGGATCCCAACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGGCCATCCT GGTCGAGCTGGACGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCGTGCCCTGGCCACCTCGTGACCACCTGACC TACGGCGTGCAGTGCTTCAGCCGCTACCCGACCACATGAAGCAGCAGCTTCTTCAAGTCCGCCATGCCCGAAG GCTACGTCCAGGAGCGCACCATCTTCTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGG GCGACACCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCTGGGGACAAGC TGGAGTACAACATAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCA AGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACG GCCCCGTGCTGCTGCCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAAGCCCCAACGAGAAGCGCG ATCATATGGTCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGG CCGCCACC
EGFP mRNA full-length	GGGCGAAUUGGGUACCGGGCCCCCCCCUCGAAGACAAGCUUCCUGCAGGUCGACUCUAGAGGAUC CCGGGUACCGAGCUCGAAUUCGGCUUCCACCAUGGUGAGCAAGGGCGAGGAGCUGUUCACCGGG GUGGUGCCCCAUCCUGGUCGAGCUGGACGGCGACGUAACGGCCACAAGUUCAGCGUGUCCGGCG AGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCACCACCGGCAAGCU GCCCCUGCCCUGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACC CCGACCACAUGAAGCAGCAGCAGCUUCUUAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCG CACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGAC

Appendix

	<p>ACCCUGGUGAACCAGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCAACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUUAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUAUCCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCACCCAGUCCCGGUGAGCAAAGACCCCAACGAGAAGCGGAUCACAUAUGGUCCUGCUGGAGUUCGUGACCGCCGCGGAUACUCUCGGCAUGGACGAGCUGUACAAGUAAAAGCGCCGCCACC</p>
Wild type EMCV IRES-EGFP mRNA full length	<p>GGGCGAAUUGGGUACCGGGCCCCCCCCUCGAGGUCAUCGAAUUCGCCCCUCUCCUCCCCCCCCCUAACGUUACUGGCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUUGUCUAUAUGUUAUUUUCACCAUAUUGCCGUCUUUUGGCAUUGAGGGCCCGGAAACCUGGCCUGUCUUCUUGACGAGCAUUCUAGGGGUCUUUCCCCUCUCGCCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUUCUUGGAAGCUUCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAAACCCCCACCUGGCGACAGGUGCCUCUCGGGCCAAAAGCCACGUGUAUAAGAUACACCUGCAAAGCGGCACAACCCAGUGCCACGUUGUGAGUUGGAUAGUUGGAAAGAGUCAAAUGGCUCUCCUCAAGCGUAUUAACAAGGGGCGUAAGGAUGCCAGAAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGGCCUCGGUGCAUAUGCUUUAUGUGUUUAGUCGAGGUUAAAAAACGUCUAGGCCCCCGAACCACGGGGACGUGGUUUUCCUUUGAAAAACACGAUGAUAAUAUGGCCACAACCAUGGUGAGCAAGGGCGAGGAGCUGUACACCGGGGUGGCCAUCCUGGUCGAGCUGGACGGCGACGUAAGCGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGGUUCAUCUGCACCACCGGCAAGCUGCCCGUGCCCGAGGCCACCCUCGUGACCACCCUGACCUACGGGUGCAGUGCUUCAGCCGCUACCCCGACCAUGAAGCAGCAGCAGCUUCUUAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCUGCAGCUCGCCGACCACUACCAGCAGAACACCCCAUCGGCGACGGCCCCGUGCUGCGCCGACAACCACUACCUGAGCACCCAGUCCGCCUGAGCAAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUACUCUCGGCAUGGACGAGCUGUACAAGUAAAGCGGCCGCCACC</p>
Fragment 1 used for 5'EGFP-lig. Block RNA synthesis	<p>GGGCGAAUUGGGUACCGGGCCCCCCCCUCGAGGUCAUCGAAUUCGCCCCUCUCCUCCCCCCCCCUAACGUUACUGGCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUUGUCUAUAUGUUAUUUUCACCAUAUUGCCGUCUUUUGGCAUUGAGGGCCCGGAAACCUGGCCUGUCUUCUUGACGAGCAUUCUAGGGGUCUUUCCCCUCUCGCCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUUCUUGGAAGCUUCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAAACCCCCACCUGGCGACAGGUGCCUCUCGGGCCAAAAGCCACGUGUAUAAGAUACACCUGCAAAGCGGCACAACCCAGUGCCACGUUGUGAGUUGGAUAGUUGGAAAGAGUCAAAUGGCUCUCCUCAAGCGUAUUAACAAGGGGCGUAAGGAUGCCAGAAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGGCCUCGGUGCAUAUGCUUUAUGUGUUUAGUCGAGGUUAAAAAACGUCUAGGCCCCCGAACCACGGGGACGUGGUUUUCCUUUGAAAAACACGAUGAUAAAGCUUGGAUC</p>
Fragment 3 used for 5'EGFP-lig. Block RNA synthesis	<p>GGGCGAGGAGCUGUACACCGGGGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAAACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCACCACCGGCAAGCUGCCCGUGCCCGAGGCCACCCUCGUGACCACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCAUGAAGCAGCAGCAGCUUCUUAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGCGCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGACACCCUGGUGAACCGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCACUACCAGCAGAACACCCCAUCGGCGACGGCCCCGUGCUGGCCGACAACCACUACCUGAGCACCCAGUCCGCCUGAGCAAAAGACCCCAACGAGAAGCGCGAUCACAUUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUACUCUCGGCAUGGACGAGCUGUACAAGUAAGCGGCCGCCACC</p>
Fragment 1 used for 5' IRES-lig block RNA synthesis	<p>GGGCGAAUUGGGUACCGGGCCCCCCCCUCGAGGUCAUCGAAUUCGCCCCUCUCCUCCCCCCCCC</p>

Fragment 3 used for 5' IRES-lig block RNA synthesis	GGAAUAAGGCCGGUGUGCGUUUGUCUAUAUGUUAUUUUCCACCAUAUUGCCGUCUUUUGGCAAU GUGAGGGGCCGAAACCUGGCCUGUCUUCUUGACGAGCAUUCUAGGGGUCUUUCCCCUCUCG CCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAGCAGUCCUCUGGAAGCUUCUUGAA GACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAACCCCCACCUGGCGACAGGUGCCU CUGCGGCCAAAAGCCACGUGUAUAGAUAACACUGCAAAGGCGGCACAACCCAGUGCCACGUU GUGAGUUGGAUAGUUGUGGAAAAGAGUCAAAUGGCUCUCCUCAAGCGUAUUAACAAGGGGCG AAGGAUGCCCAAGGUACCCCAUUGUAUGGGAUCUGAUCUGGGGCCUCGGUGCACAUGCUUUA CAUGUGUUUAGUCGAGGUUAAAAAACGUCUAGGCCCCCCGAACCACGGGGACGUGGUUUUCU UUGAAAAACACGAUGAUAAUUGGCCACAACCAUGGUGAGCAAGGGCGAGGAGCUGUACCCGG GGUGGUGCCCAUCCUGGUCGAGCUGGACGGCGACGUAACCGGCCACAAGUUCAGCGUGUCCGGC GAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCCUGAAGUUAUCUGCACCACCGGCAAGC UGCCCGUGCCCUGGCCACCCUCGUGACACCCUGACCUACGGCGUGCAGUGCUUCAGCCGCUAC CCCGACCACAUGAAGCAGCAGACUUCUUAAGUCCGCCAUGCCCGAAGGCUACGUCCAGGAGC GCACCAUCUUCUUAAGGACGACGGCAACUACAAGACCCGCGCCGAGGUGAAGUUCGAGGGCGA CACCUGUGAACCAGCAUCGAGCUGAAGGGCAUCGACUUAAGGAGGACGGCAACAUCUUGGGG CACAAGCUGGAGUACAACUACAACAGCCACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAAGC GCAUCAAGGUGAACUUAAGAUCGCCACAACAUCGAGGACGGCAGCGUGCAGCUCGCCGACCA CUACCAGCAGAACACCCCCAUCGGCGACGGCCCCGUGCUGCUGCCCGACAACCACUACCUGAGCA CCCAGUCCGCCUGAGCAAAGACCCCAACGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGU GACCGCCGCCGGGAUCACUCUGGCAUGGACGAGCUGUACAAGUAAAGCGGCCGCCACC
Fragment 1 used for Internal-lig. Block RNA synthesis	GGGCGAAUUGGGUACCGGGCCCCCCCCUCGAGGUAUCGAAUUCGCCCCUCUCCUCCCCCCCCC CUAACGUUACUGGCCGAAGCCGCUUGGAAUAAGGCCGGUGUGCGUUUGUCUAUAUGUUAUUUUC CACCUAUUGCCGUCUUUUGGCAUUGUGAGGGCCCGGAAACCUGGCCUCUGCUUCUUGACGAGC AUUCCUAGGGGUCUUUCCCCUCUGCCAAAGGAAUGCAAGGUCUGUUGAAUGUCGUGAAGGAAG CAGUUCUCUGGAAGCUUCUUGAAGACAAACAACGUCUGUAGCGACCCUUUGCAGGCAGCGGAA CCCCCACCUGGCGACAGGUGCCUCUGCGGCCAAAAGCCACGUGUAUAGAUAACACUGCAAAG GCGGCACAACCCAGUGCCACGUUGUGAGUUGGAUAGUUGUGGAAAGAGUCAAAUGGCUCUCCU CAAGCGUAUUAACAAGGGGCGUAAGGAUGCCAGAGGUACCCCAUUGUAUGGGAUCUGAUCU GGGGCCUCGGUGCAUGCUUUAACUUGUUAUAGUCGAGGUUAAAAAACGUCUAGGCCCCCCG AACCACGGGGACGUGGUUUUCCUUUGAAAAACACGAUGAUAAAGCUUGGAUCCACAAACCAUGGU GAGCAAGGGCGAGGAGCUGUUCACCGGGGUGGUGCCAUCCUGGCGAGCUGGACGGCGACGUA AACGGCCACAAGUUCAGCGUGUCCGGCGAGGGCGAGGGCGAUGCCACCUACGGCAAGCUGACCC UGAAGUUAUCUGCACCACCGGCAAGCUGCCCGUGCCCGGGCCACCCUCGUGACCACCCUGACC UACGGCGUGCAGUGCUUCAGCCGCUACCCCGACCACAUGAAGCAGCAGACUUCUUAAGUCCG CCAUGCCCGAAGGCUACGUCCAGGAGCGACCAUCUUCUUAAGGACGACGGCAACUACAAGAC CCGCGCCGAGGUGAAGUUCGAGGGCGACACCC
Fragment 3 used for Internal-lig. Block RNA synthesis	GGGCAUCGACUUAAGGAGGACGGCAACAUCUGGGGCACAAGCUGGAGUACAACUACAACAGC CACAACGUCUAUAUCAUGGCCGACAAGCAGAAGAACGGCAUCAAGGUGAACUUAAGAUAUCCGCC ACAACUAGGAGCGGACGCGUGCAGCUCGCGGACCACUACCAGCAGAACACCCCAUCGGCGAC GGCCCCGUGCUGCGCCGACAACCACUACCUAGCAGCACCAGUCCGCCUAGCAAAGACCCCAA CGAGAAGCGCGAUCACAUGGUCCUGCUGGAGUUCGUGACCGCCGCCGGGAUCACUCUGGCAUG GACGAGCUGUACAAGUAAAGCGGCCGCCACC

5.8 Molecular biology methods and techniques

5.8.1 From DNA to RNA

Plasmid preparation and amplification

Plasmids DNA (pDNA) used in this work were either amplified in the lab or commercially prepared. In case of bacterial transformation, competent DH5 α *E.coli* strain (Invitrogen) were

transformed by pDNA (pPBSII IRES-EGFP or pGEM4Z64A-eGFP) following the manufacturer's instructions. Following heat shock, transformed bacteria were cultured in antibiotic-free liquid medium for 1 hour then plated on agar plates using Pasteur pipette. As growth control, plate with an additional antibiotic (Kanamycin, 50 µg/mL) and another without any. A single colony was selected from the agar plate considering the selective resistance to Ampicillin at 100 µg/mL, and was used to prepare pre-culture in 10 mL Luria broth (LB) incubated 8 hours at 37°C/ 350 rpm followed by an overnight culture in 150 mL LB (using 150 µL volume from pre-culture). At OD600 ~ 1.5, plasmid was extracted following the QIAGEN Plasmid Maxi Kit (QIAGEN).

DNA linearization

Circular plasmid vector with integrated EGFP sequence of the corresponding mRNA (GenScript, USA) was linearized by restriction enzymes: SacII (pPBSII IRES-eGFP, IRES5'eGFP-3_PUC57, IRES-EMCV-S1 _pUC57 and IRESeGFPint2-3_PUC57, and eGFP-1-UTR-Stop_PUC57), Sma I IRES5'eGFP-1_PUC57, BamHII (IRESeGFP5'-1-S1_PUC57 and IRESeGFPint2-1_PUC57), and BcuI (SpeI) (pGEM4Z64A-eGFP). The linearization was performed as described by the manufacturer protocol using the minimal enzyme units and an overnight incubation at 37°C. The linearized plasmid was purified using phenol/chloroform extraction and precipitated by ethanol-precipitation.

RNA synthesis by in vitro transcription and purification

The *in vitro* transcription was performed using a T7 RNA polymerase previously in-house expressed from an Addgene plasmid pQE 80L Kan or commercially available from Thermo Fisher Scientific. The DNA template was therefore incubated with a NTP mix at a final concentration of 5mM each in 40 mM Tris-HCl (pH 8.1), 1 mM Spermidine, 10 mM DTT and 0.01 % Triton X-100 as well as final concentration of 30 mM MgCl₂ and 2.5 µg/mL Bovine serum albumin (Thermo Fisher Scientific). In order to prepare the ligation step (if needed), the final concentration of GTP was reduced to 2 mM and GMP was added in an excess of 16 mM final concentration⁴¹¹. This step will facilitate the ligation of two RNA extremities: the 5' end mono-phosphate with the 3'-end hydroxyl group. Finally, the DNA Template was digested by adding 2 U DNase I per 1 µg DNA at 37 °C for 30 minutes. To purify the transcription products, the MegaClear™ Kit was used according to the manufacturer's protocol.

In vitro transcription of modified mRNA

Random modified mRNA synthesis was carried out in the same conditions as the IVT described above in the exception of: the amount of the modified NTPs were added to substitute their unmodified equivalent in the following ratios: 2.5%, 30%, 60% and 100% of the unmodified respective NTP. Subsequent purification was realized using the MegaClear™ Kit according to the according to the manufacturers protocol.

Capping

The addition of the cap was carried in one-step reaction uusing a combination of *Vaccinia Capping System* and mRNA Cap 2'O-methyltransferase (NEB) and incubated over 2 hours according to the manufacturer recommendations. Subsequently, samples were purified using the MEGAclear™ Kit (Invitrogen) silica spin column system and eluted in MilliQ water.

Polyadenylation

Polyadenylation reaction was carried out using purified transcripts (capped or non-capped) using an optimized final concentration of 75U/μL. The reaction was incubated at 37°C for 45 minutes.

5.8.2 Purification & concentration methods for oligonucleotides

Phenol-chloroform extraction

The recovery of the linearized DNA plasmid was carried out by two times extraction with chloroform and phenol (TE-buffered, Carl Roth) at a ratio of 2:1:1. The mixture was vortexed and centrifuged for 2 minutes at 14 000 rpm for phase separation. The upper phase was transferred into a new tube for another two times with volume/volume chloroform. One last extraction with diethyl ether was followed to remove residual phenol traces. The lower phase was saved and controlled for the presence of phenol by measuring the absorbance at 270 nm using Nanodrop®. In the case of phenol detection, the last extraction step can be repeated.

Ethanol precipitation

The precipitation of an oligonucleotide solution was carried out using 0.1x volumes of 5 M Ammonium acetate pH 5.2, 1 μL glycogen (5 mg/mL, Thermo Fisher Scientific, Germany) and 2-2.5 volumes of pure cold ethanol (kept at -80°C or -20°C). The mixture was vortexed and incubated over night at -20°C. In a next step, RNA sample was centrifuged (15.000 x g at -4°C) for 30-458 minutes, then, the supernatant was removed and the pellet wash by resuspension in

75% ethanol (pre-cooled at -20°C). The solution was again centrifuged for 15 minutes, followed by a second washing step. Finally, the pellet was re-suspended in MilliQ water.

5.8.3 Design and synthesis of ligated mRNA and analysis

Point-modified mRNA preparation

Using three fragments, namely fragments 1, 2, and 3, a long point-modified mRNA was constructed by 3-way-one-pot ligation. Therefore, fragments 1 and 3 were *in vitro* transcribed (IVT), while the middle oligonucleotide fragment was chemically synthesized and commercially ordered from IBA (Germany). Prior to ligation, 1 pmol of commercial oligonucleotide (unmodified oligo MH905) was phosphorylated at its 5' end using 0.9 U of T4 Polynucleotide kinase enzyme in a final volume adjusted to a final enzyme concentration of 1 U/μL. Additionally, to the buffer A, 5 mM DTT and 5 mM ATP were added and the reaction mixture was incubated 1h at 37°C. In the next step, equimolar amounts from IVT fragment 1 and fragment 2 were added to the phosphorylation reaction mixture, while the amount of the splint DNA (MH1168) was 2% less than the RNA fragments amount. For the preparation of “5' IRES-lig. block” and “internal-lig block”, splint DNA MH910 and MH911 were respectively used. The Ligation reaction was then incubated in 1x T4 RNA Ligase 2 for 4 minutes at 75 °C, and allowed to cool down to room temperature over 15 minutes. In the following step, the ligation was achieved by the addition of 1 U of T4 RNA Ligase 2 per 0.8 μg of RNA. The total volume was adjusted to an enzyme concentration of 0.57 U/μL and incubated overnight at 16 °C. Finally, DNA splint was degraded using 2 U of DNase I per 1 μg DNA and incubated at 37 °C for 30 minutes.

Hybridization to fluorescent DNA oligo

The successful incorporation (by ligation) of this middle fragment was verified by hybridization of a fluorescent cDNA-probe analysed on a native 2 % agarose gel, then separated on a 1% native. The gel was first scanned for fluorescence (Table 4), then stained with 1x GelRed.

Circularization of mRNA

For circularization, the *in vitro* transcribed mRNA was circularized using T4 RNA ligase 2. Therefore, mRNA was annealed to an equimolar 60-nt length DNA oligomer (splint DNA MH838) by incubating in 1x reaction buffer for 4 min at 75 °C, followed by cooling at room temperature for 15 min. Then 50 U T4 RNA ligase 2 was added to the mixture and the reaction was incubated at 37 °C for 30 min. The splint DNA was digested using 2 U of DNase I (Thermo Fisher Scientific, Germany) per 1 μg DNA and incubated at 37 °C/30 minutes.

Rnase R

For the RNA circularity validation, RNA samples from the circularization reaction was denatured by incubation at 85°C for 60 seconds and cooled on ice for 2 minutes. Then, 1U RNase R enzyme was applied per 0.3 µg RNA and the reaction was incubated 3 h/37°C.

RNA nicking assay

RNA was partially hydrolysed by incubating 100 ng gel purified circular RNA product in 50 mM NaHCO₃, pH 9 in a final volume of 20 µL. The reaction was incubated for 45 seconds at 90°C in water bath. Longer incubation led to RNA degradation. Control reaction was carried out by incubating on ice. Digestion was stopped by immediately re-suspending in denaturing RNA loading buffer and storing on ice, then analysed on 4% denaturing PAGE electrophoresis.

Agarose particle tracking

The respective purified ligation samples were brought to a total volume of 1 mL and injected into the sample chamber from a Malvern NanoSight LM10 (Malvern Panalytical, Germany). A 532 nm laser was scattered by the particles in suspension. This was visualized by a microscope with 20x magnification on which a sCMOS camera was mounted. Because of the unequal size and shape, the number of agarose particles was not evaluated automatically by the Nanoparticle tracking analysis software, but rather manually. Therefore, eight random positions of the camera were chosen and inspected visually. The categories of the particle size were big, medium, and small. All particles between estimated 1 µm² and about 3.5 µm² were classified as medium.

5.8.4 PAGE electrophoresis and circular RNA purification***Polyacrylamide gel electrophoresis (PAGE) and gel elution***

In order to assess the quality RNA 4% denaturing polyacrylamide gel solution was used and the gel was casted in glass frames of different sizes. RNA sample (at least 100 ng) was dissolved in 2x denaturation loading buffer. In addition, a DNA ladder and a blue marker (bromophenol blue and xylene cyanol) were also prepared with the loading buffer to assess the size and migration time. The electrophoresis was running in 1x TBE buffer at 12W for 3 hours in small glass frame (20 x 13 cm) and 7 hours and 10 minutes in bigger glass frame (20 x 20 cm), while cooling with a fan. After migration, the gels were carefully transferred into a staining solution containing one of the following intercalating agents: SYBR® Gold or GelRed™. After 15-20 minutes of shaking incubation, the gels were scanned using a Typhoon TRIO+ (GE Healthcare)

which permitted the visualisation of the different nucleic acids' bands. For gel extraction, GelRed solution was used for staining and band visualization. The scanned gel picture was printed on A4 paper and placed under the gel to guide the excision the corresponding band. After cutting, the gel piece was crushed into smaller pieces, and transferred into a 1.5 mL Eppendorf® tube to incubate in 250 µL of 0.5 M NH₄OAc (pH 5.0) for elution. After overnight incubation at 20°C/750 rpm, the mixture was purified *via* a Nanosep spin filter with 0.45 µm membrane. The separation between the elution buffer from the gel residues by centrifugation over 15 minutes at 12 000 rpm. Finally, RNA was precipitated by Ethanol precipitation.

Agarose gel electrophoresis

For the separation of DNA or RNA samples, 1% agarose gel was prepared and pre-stained with SYBR®Gold nucleic acid gel stain. The staining solution was mixed in 1x TBE (roughly 1/10 of the prepared gel solution) and added to the final mixture after heating step. In the case of fluorescent RNA sample, the gel was post-stained. A quantity of 75 ng-1 µg of RNA or DNA samples were mixed with the loading dye and gel electrophoresis was run in 1X TBE buffer at 70 Volts for 30 minutes then voltage was increase to 130 Volts.

For nucleic acid visualization, the gel was scanned with Typhoon 9400 (GE Healthcare).

After loading the samples, electrophoresis was carried out in 1x TBE buffer at 120 Volts for 90 minutes. If not pre-stained, gels were stained afterwards with either SYBR® Gold or GelRed™. Nucleic acid bands were visualized on a Typhoon TRIO+ (GE Healthcare) according to the settings illustrated in Table 1.

Scanning settings for Typhoon imager

In the case of detecting fluorescently RNA samples, gels were first scanned with the corresponding wavelength, then stained with Sybr®Gold or GelRed and allowed to scanning and analysis by the software ImageJ.

Table 7 : Settings for gel scans based on spectroscopic properties of staining dyes/labels.

		Laser excitation [nm]	Emission filter [nm]
Staining solution	SYBR [®] Gold	488	520 BP 40
	GelRed	532	610 BP 30
Fluorescent dye conjugated to DNA oligo	Fluorescein	488	520 BP 40
	Cy5	633	670 BP 30

BP: band pass filter; SP: short pass filter

RNA Analysis by TapeStation

RNA samples (75-150 ng) were analysed by TapeStation (4200, Agilent) according to the operating procedure of the manufacturer.

Point-modified-ligated RNA recovery by real-time gel elution

Purification of the ligated mRNA was conducted by real-time elution, adapted from the concept from E-Gel CloneWell (Thermo Fisher Scientific) ⁴¹², using Dark Reader Blue Light Transilluminator and a conventional agarose gel chamber. The gel was prepared to contain two large wells, one in the top and another one in the middle of the gel, by combining 10 wells together with a Sellotape (out of 12-well comb). In order to follow migration, 1% agarose gel was pre-stained with 1x SYBR[®]Gold. When the gel was formed, the ligation sample was mixed 2:1 with denaturing loading dye was applied to the upper well. The electrophoresis was run at 80 V over 30 minutes, then, the voltage was increased to 150 V. Once the sample reaches the second well, interest bands were collected using a 1000 µL pipet. The collection chamber was rinsed between the fractions with 1x TBE buffer. The eluted RNA was then precipitated by Ethanol-precipitation and dissolved in ultrapure water. A last step was carried by filtering the purified RNA through a 0.2 µm solid phase filters, to reduce further residual contamination with agarose particles.

5.8.5 Liquid chromatography (LC)

RNA analysis and purification via HPLC

For circular RNA analytics and elution, a YMC-Triart C18 (150 x 3 mm and 3 µm particle size) column from YMC Europe GmbH, Germany, combined with an Agilent 1100 HPLC series

coupled with a DAD (diode array detector) was used. In a prior step, RNA circularization reaction was first purified by phenol/chloroform, and an amount of 30-40 µg purified RNA was injected. The chromatography was realized by using 0.1 M triethylammonium acetate (pH 7.0, solvent A) and HPLC grade acetonitrile (solvent B) under denaturing conditions at 65 °C, with a flow rate of 0.4 mL/min as following:

Table 8 : HPLC gradient applied for for circular RNA

Time (min)	Solvent A (%)	Solvent B (%)
2	90	10
8	87.5	12,5
38	86,7	13,3
45	86,7	13,3
55	86,6	13,4
85	86,6	13,4
95	86,2	13,8
105	86,2	13,8
110	50	50
113	100	100
116	100	100
123	90	10

Table 9 : HPLC gradient applied for ligated linear RNA purification

Time (min)	Solvent A (%)	Solvent B (%)
5	90	10
6	87.5	12,5
36	86.7	13,5
41	86.7	13,5
51	86.6	13,8
57	86.6	13,8
65	86.2	100

5.8.6 Liquid chromatography-Tandem Mass spectrometry (LC-MS/MS)

Sample preparation

All samples for LC-MS (QQQ) analysis were digested to their nucleoside level as follows: 10 µg of isolated RNAs were mixed with 0.3 U nuclease P1, 0.1 U snake venom phosphodiesterase, 200 ng Pentostatin (adenosine deaminase inhibitor) and 500 ng Tetrahydrouridine (cytidine deaminase inhibitor) in 1/10 volume of 10× nuclease NP1 buffer (0.25 M NH₄OAc pH 5.0, 0.2 mM ZnCl₂). The mixture was incubated for 3–4 hours at 37 °C. After the first incubation 1/10

volume 10× fast alkaline phosphatase buffer (0.5 M NH₄OAc, pH 9.0) and 1 U fast alkaline phosphatase were added and the mixture incubated again for 5 hours at 37 °C. In the case of digestion of smaller RNA amounts the enzyme concentrations were adjusted.

LC-MS (QQQ) measurements and settings

For LC-MS measurements, an amount of 50 ng from the digested sample were analysed together with 50 ng internal standard *via* Agilent 1260 series coupled to a diode array detector (DAD) and Agilent 6460 Triple Quadrupole mass spectrometer equipped with an electrospray ion source. The solvents consisted of 5 mM ammonium acetate buffer (pH 5.3; solvent A) and LC-MS grade acetonitrile (solvent B; Honeywell). The elution started with 100% solvent A, followed by a linear gradient to 8% solvent B at 10 min and 40% solvent B after 20 min. The flow rate was 0.35 mL/min on a Synergi Fusion column (4 µM particle size, 80 Å pore size, 250 x 2.0 mm from Phenomenex). ESI parameters were as follows: gas temperature 350°C, gas flow 8 L/min, nebulizer pressure 50 psi, sheath gas temperature 350°C, sheath gas flow 12 L/min, capillary voltage 3000 V. The MS was operated in the positive ion mode using Agilent MassHunter software in the dynamic MRM (multiple reaction monitoring) mode. For quantification, a combination of external and internal calibration was applied as described previously ⁴¹³.

5.8.7 *In vitro* translation and protein analytics

***In vitro* translation in Nuclease-treated Rabbit Reticulocyte Lysates**

Rabbit Reticulocyte Lysates, nuclease treated (Promega) was programmed with 1 pmol mRNA and incubated for 90 minutes at 30 °C following the manufacturer instructions. Translation was stopped and aliquots of 2 µL were analysed for EGFP protein expression. Aliquots were mixed in non-reducing loading SDS-PAGE buffer (without β-mercaptoethanol) and loaded without pre-heating treatment on 10% SDS-polyacrylamide gel electrophoresis. Protein activity was directly assessed by in-gel fluorescence. The gel was placed in Typhoon scanner TRIO+ (GE Healthcare, United Kingdom) and fluorescence was detected by blue laser settings (488 / 532 nm excitation wave length and 526 nm short pass emission filter). Quantification of EGFP signal intensity was analysed by ImageJ normalized to the background. The cap dependency experiment was performed by a preincubation with 0.7 mM m⁷G(5')ppp(5')G RNA Cap structure (concentration adapted from ^{132,414} for 5 minutes at room temperature to sequester cap-binding proteins. Subsequently, the translation experiment and analysis were realized as described.

For the translation rescue experiment, 1 pmol mRNA was by incubated with increasing amounts (0.6 μ g, 1.2 μ g, and 2 μ g) of purified fractions from total tRNA enriched with eukaryotic cognate tRNA valine (fraction C5, provided by Gérard Keith). Translation reaction and analysis were performed as described.

***In vitro* translation in untreated Rabbit Reticulocyte Lysates**

To follow the protein synthesis, 1 pmol mRNA was programmed in untreated Rabbit Reticulocyte Lysate. Translation mixture contained 75 mM KCl, 0.5 mM MgCl₂, RNasin and a mixture of amino acids (Methionine is omitted), 1mM each, and ³⁵S-methionine in a final volume of 20 μ L. The reaction was incubated at 30°C for 1 hour. Protein production was analysed by separating 5 μ L aliquot (mixed with loading dye) on 15% SDS PAGE and visualized by phosphorimaging.

5.8.8 Immunoprecipitation and samples preparation (for section 3.5)

This part was published in Slama and colleagues ¹⁴⁴.

***In vitro* transcription of modified and radioactive labelled mRNA**

Cold RNA was *in vitro* transcribed (as described in section 5.8.1) using linearized plasmid pGEM4Z64A-eGFP as template. The modified mRNA was prepared either by adding 2.5% or 50% of m⁶ATP relative to ATP to the reaction mixture.

Radioactive labeled RNA (hot) was prepared similar to cold mRNA, in the exception of the addition of 10 μ Ci α ³²P-ATP (800 Ci/ mmol) or 10 μ Ci α ³²P-UTP (800 Ci/mmol) to unmodified RNA and modified transcription reaction, respectively. The cold transcribed mRNAs was quantified by measuring the absorbance at 260 nm using NandoDrop, while the radioactivity intensity of hot transcripts was measured using a scintillation counter Wallac 1409 (PerkinElmer).

RNA sample preparation for IP

Two RNA mixtures (mixture A and mixture B) were prepared to contain hot and cold RNA as following: After hot IVT, the total CPM was measured, a volume (V) could be determined to pipette 8000 CPM. The amount of RNA in this volume was determined using the concentration of the corresponding cold IVT (for example hot unmodified IVT and cold unmodified IVT), relying on the fact that both IVT reactions (cold/hot) were prepared in the same conditions. To this, cold mRNA was added to achieve a total amount of 8 μ g. Such that, mixture A was prepared to contain the modified RNA and a mixture B containing unmodified one. Both

mixtures A and B were then mixed to achieve the same amounts in μg of modified and unmodified RNA and were applied for immunoprecipitation.

Nuclease digestion for 1D TLC

RNA samples were digested in the presence of 1/ 10 V of nuclease P1 buffer (10:9 vol 250 mM NH_4OAc at pH 5.0; 1V 2 mM ZnCl_2) and 0.3 U of NP1 in a final volume of 5 μL . The reaction was incubated for 2 h and a half at 37 °C.

Nuclease digestion for 2D TLC

For the 2D TLC, the modified RNA sample (2 μg) was incubated for digestion in the presence of 2.5 U RNase T2 and 10 mM sodium acetate buffer (adjusted with acetic acid to a pH at 4.5) in a final volume of 5 μL .

Immunoprecipitation

Protein G Sepharose beads (60 μL) were washed twice in 500 μL by centrifuging at 1000 \times g/4 °C for 1 min and discarding the supernatant. Afterward, 2. 5 μg of the purified antibody was added to the beads and incubated in 1 mL 1xPBS overnight in a rotating wheel. The beads were then washed three times using an RNA-IP buffer (150 mM LiCl , 0.5% NP-40, 10 mM Tris-HCl at pH 7.5). The RNA mixture was applied to the antibody-beads in 500 μL RNA-IP buffer for 2 h/750 rpm at 4 °C. Afterward, beads were washed once with RNA-IP buffer, once with wash buffer I (RNA-IP buffer plus 150 mM LiCl) and wash buffer II (RNA-IP buffer plus 300 mM LiCl), and a final one with RNA-IP buffer. The supernatant was discarded, then RNA was isolated by extraction using TRI Reagent® followed by EtOH precipitation.

For elution by competition, 0.5 mg/mL m^6A (Sigma-Aldrich) or adenosine (Sigma-Aldrich) was added to 500 μL IP buffer and applied to the IP, and incubated for 2 h/ 750 rpm at 4 °C. RNA was recovered by EtOH precipitation of the supernatant fraction, while the RNA bound to the antibodies-beads fraction was extracted TRI Reagent® as described above.

First dimension TLC

The radioactive sample was digested with NP1 then spotted by altering spotting (0.5 μL) and drying steps on a cellulose-coated thin-layer plate fluorescent. After drying, the TLC plate was placed in a glass chamber for thin layer chromatography containing solvent I and covered. When the solvent gets to the top, the separation was stopped by drying the TLC plate under the

fume hood for around 15 minutes. The TLC was then exposed to a phosphorimager screen (GE Healthcare) for 1 hour and scanned by Typhoon TRIO+ (GE Healthcare) scanner.

Second dimension TLC

The radioactive labeled RNA was digested using RNase T2 and applied to a 1D-TLC with solvent I as described above. After the first running, the plate was dried and placed in a perpendicular position to the 1D, in another glass chamber containing solvent II. Once the solvent reached the top, the plate was dried, and the radioactive spots were revealed as described above.

6 Appendix

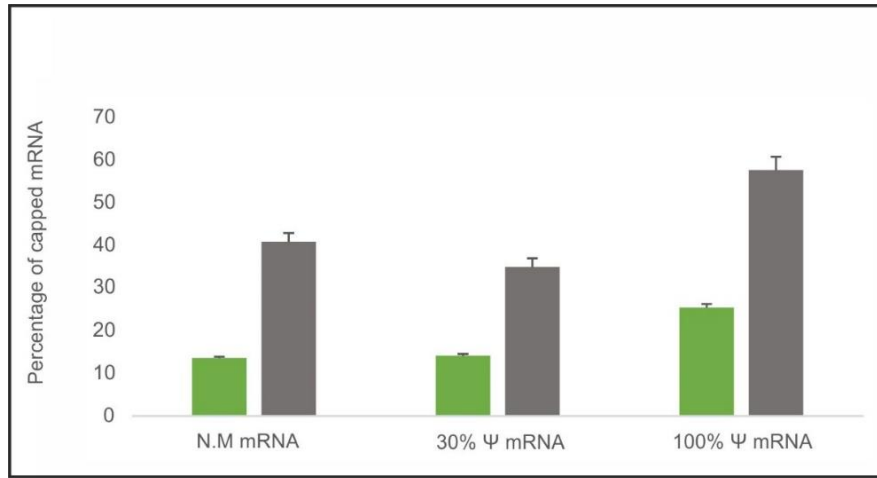


Figure 6. 1: Cap mRNA analysis.

The non-modified EGFP mRNA (N.M) was *in vitro* transcribed. Pseudouridine-modified mRNA were synthesized by the addition of pseudouridine at different ratio (30 and 100%) relative to uridines to the transcription reaction. After capping, mRNA samples were analyzed by LC-MS/MS for m⁷G (in green) and the 2'-O-Me detection (grey) detection. The Samples were incubated for digestion to the nucleoside level and analyzed via LC-MS/MS. For absolute quantification ¹³C, stable isotope-labelled nucleosides from *S. Cerevisiae* were used as an internal standard and mixed in an equal amount with the analyzed sample. Quantification of the percentage of capped mRNA was evaluated relative to cytosine quantification. The application of uncapped mRNA as negative control showed no detectable cap modifications. Errors bars indicating \pm SD of technical triplicates

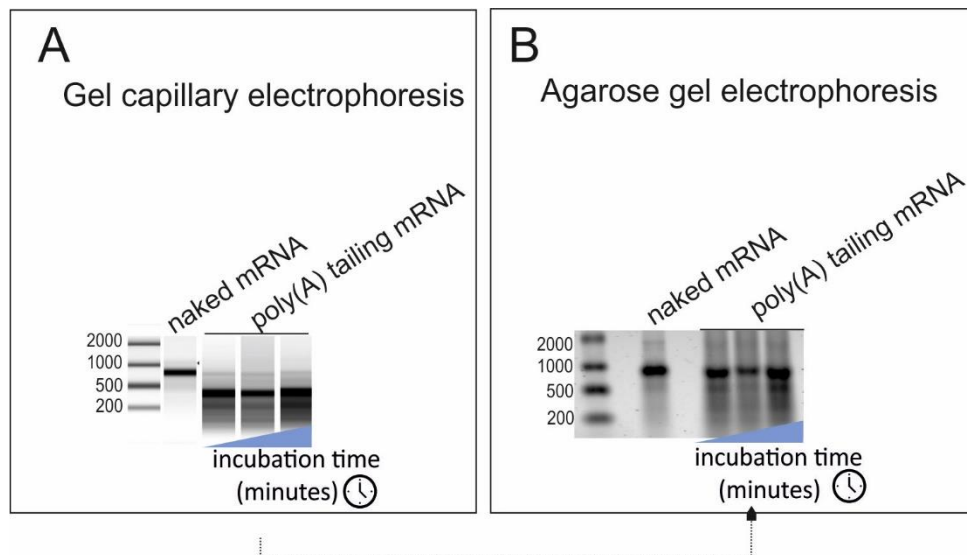


Figure 6. 2: Analysis of polyadenylated EGFP mRNA.

Poly(A) tailing using 5 units *E. Coli* Poly(A) and purified naked EGFP mRNA. The reaction was incubated at 37°C for 45, 90, and 180 minutes. Aliquots were analyzed by capillary electrophoresis (Agilent, TapeStation) (A) and the leftover mixture (unloaded by the machine) was loaded on a 1.5% agarose gel (B), showing a different migration behavior between both electrophoresis.

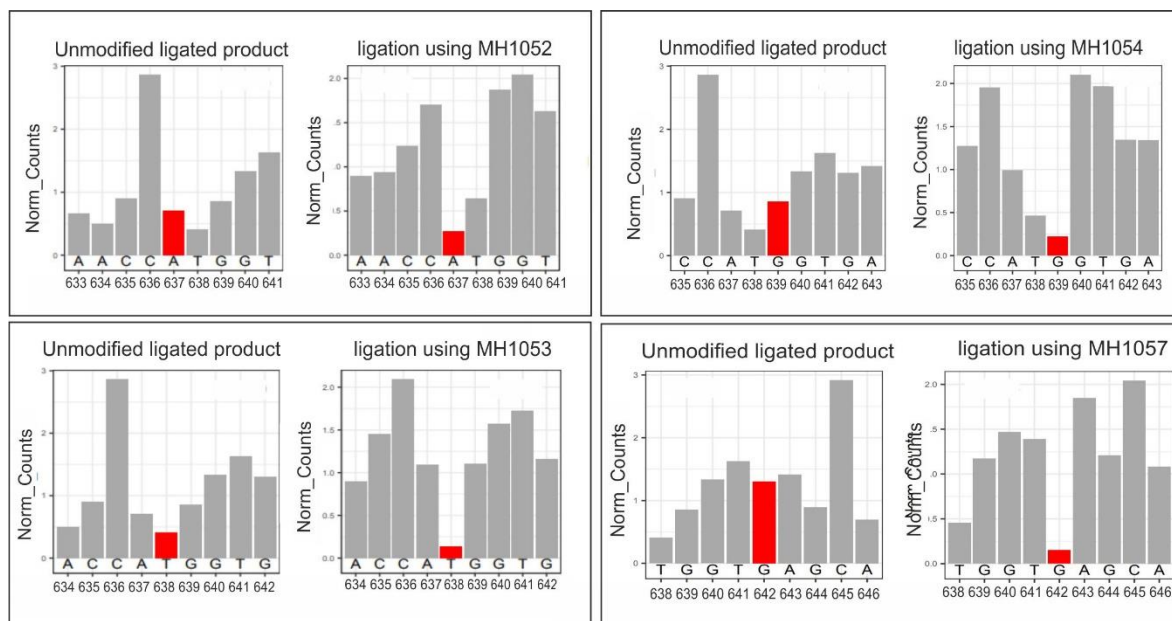


Figure 6. 3: Analysis of 2'-O-Me point-modified mRNA.

After ligation using either unmodified (MH905) or modified RNA oligo (MH1052, MH1053, MH1054 and MH1075), samples were applied for MHRiboMethSeq analysis. The protection profiles showing the normalized RiboMethSeq cleavage profiles around the corresponding positions are shown for the unmodified (unmodified ligated product) and modified construct. Position of modification is shown by a red bar.

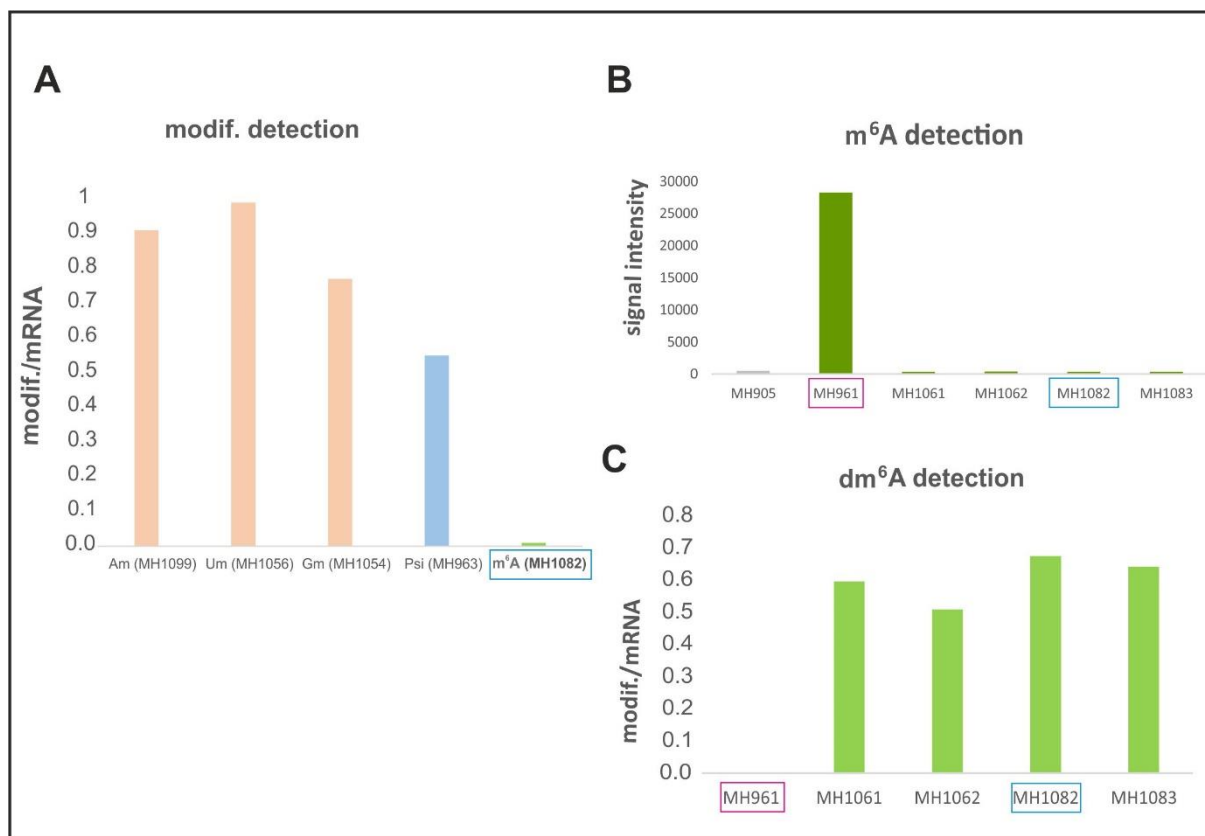


Figure 6. 4: Detection of unexpected modifications in m⁶A-point modified mRNA.

Absolute LC-MS/MS detection of m⁶A in IRES-EGFP ligated mRNA molecule. After ligation using fragment 2 harboring 2'-O-Me (MH1099, MH1058 and MH1054), pseudouridine (MH963) and m⁶A (MH1082) were applied for modification detection and quantification by LC-MS analysis relative to uridine. For absolute quantification ¹³C, stable isotope-labelled nucleosides from *S. cerevisiae* were used as an internal standard and mixed in an equal amount with the analyzed sample. An equivalent of one A_m, U_m, G_m and pseudouridine could be detected for one modified mRNA (**A**). Given the failure to detect m⁶A modified mRNA (MH1082, in blue square), other ligation products containing m⁶A at different sites (MH961, MH1061, MH1083, MH1062) were also applied to LC-MS (**B**). As a control, ligation using MH905 (non-modified oligo) was used and no modification was detected (**B**). All the RNA oligo showed the presence of dm⁶A, except MH961 (pink square) (**C**).

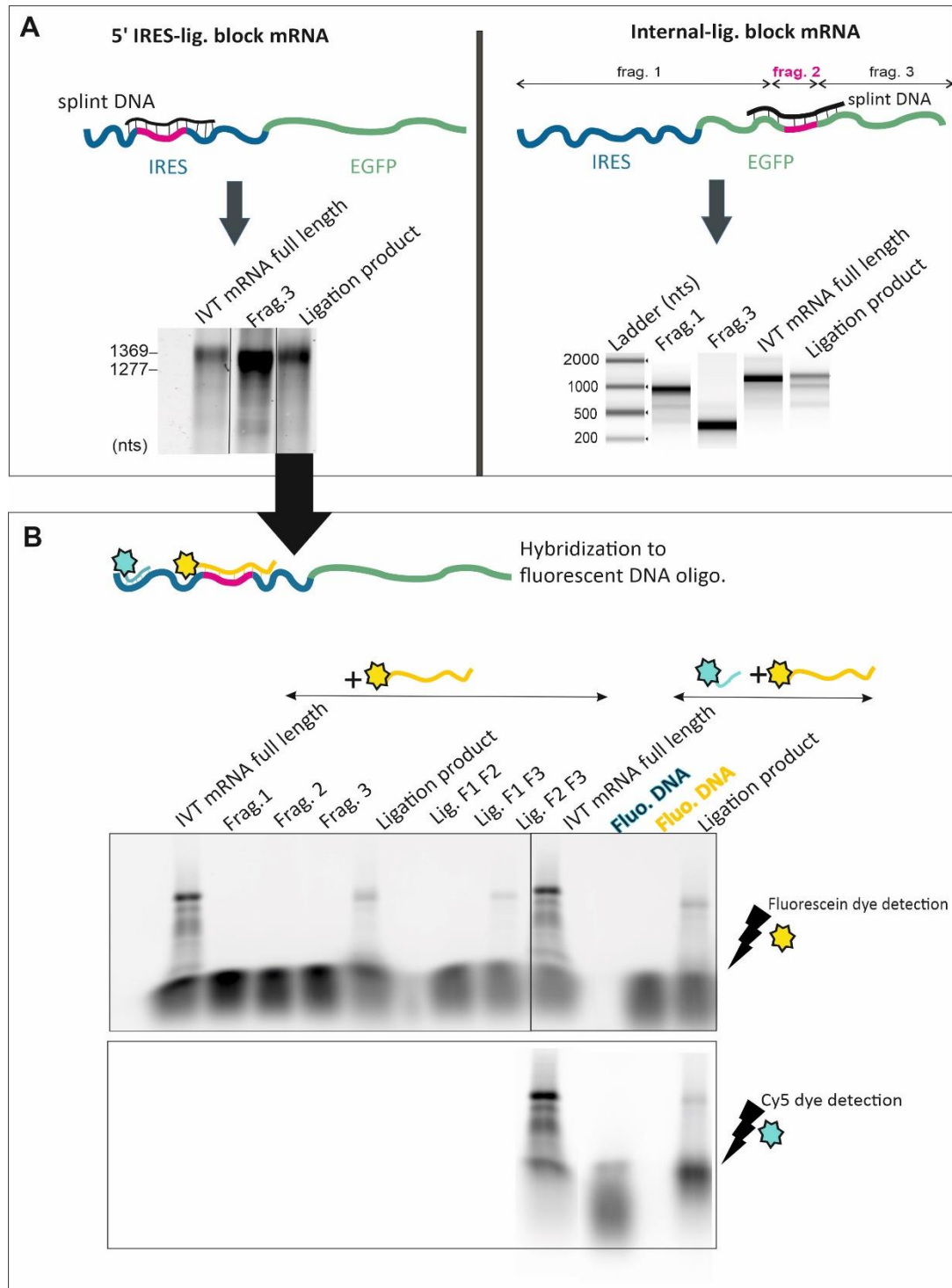


Figure 6. 5: Generation of 5' IRES-lig block and internal-lig block.

The 3-way-splint ligation concept was applied to introduce single modifications internally within the EGFP CDS region (internal-lig block) and around the 5' end of the wild type IRES EMCV (5' IRES-lig block) (A). The gel-purified 5' IRES-lig block product was analyzed on agarose gel while the internal-lig block was analyzed by gel capillary electrophoresis (TapeStation). The 5' IRES-lig block was then applied to hybridization with fluorescent DNA complementary either to fragment 1 (lower gel) or fragment 2 (upper gel). In the left part, the samples were simultaneously hybridized to both fluorescent oligos.

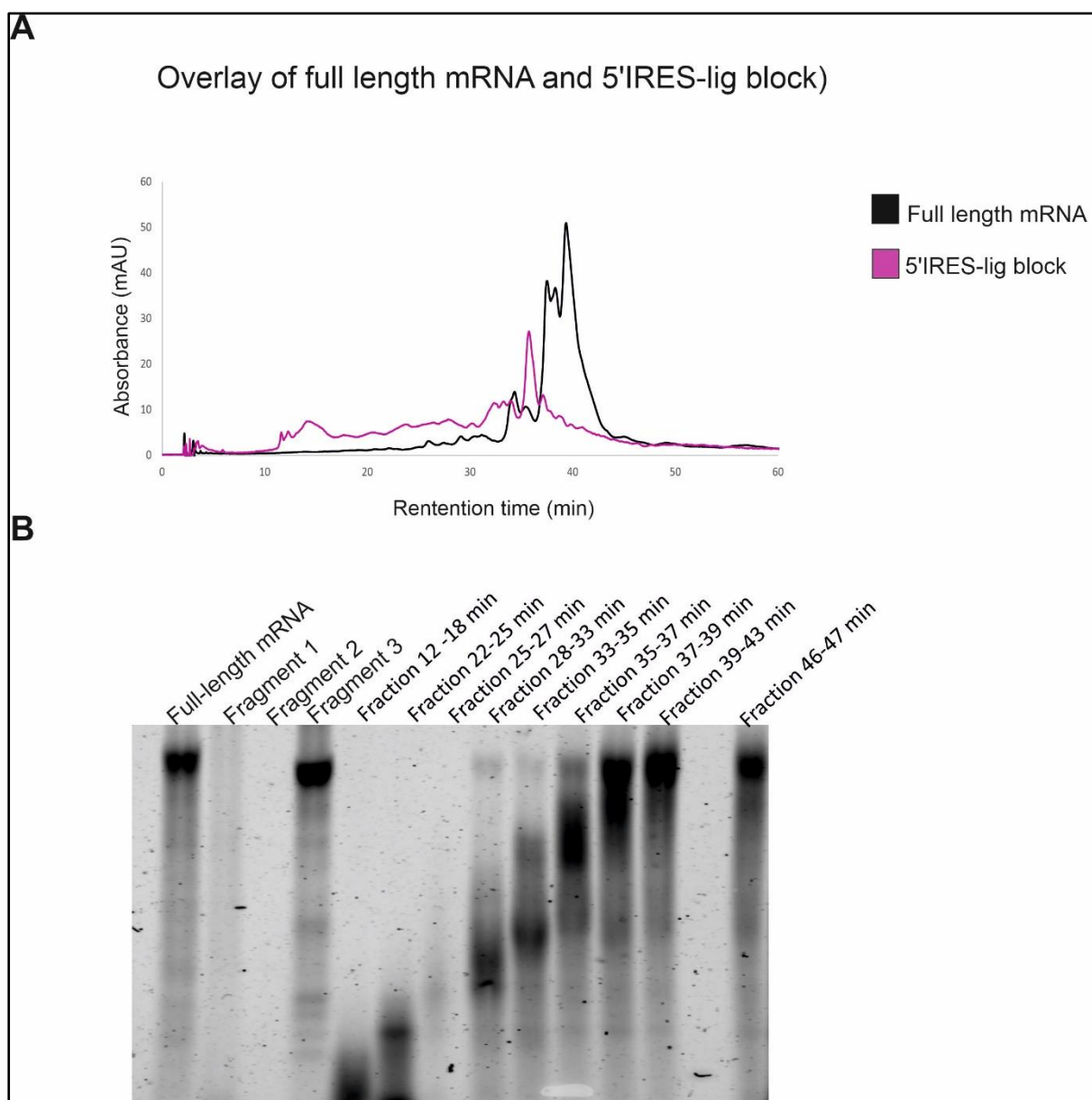


Figure 6. 6: HPLC purification of 5'IRES-lig block.

Synthesized 5'IRES-lig mRNA was phenol/chloroform purified and injected into HPLC. (B) During elution, fractions were collected and analyzed on a 1.5% agarose gel. For the visualization, GelRed was used as a staining solution.

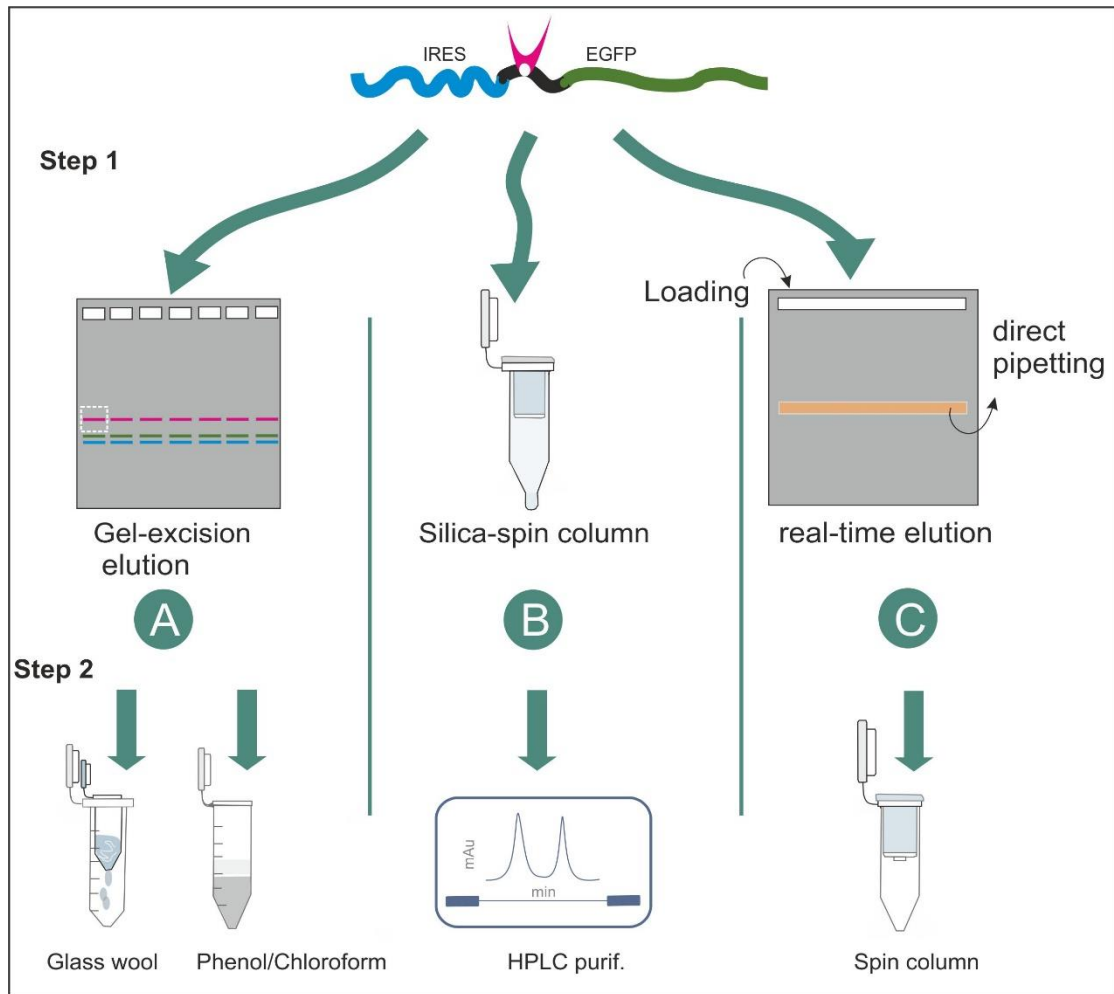


Figure 6. 7: Purification of ligated RNA.

(A) The ligated product was eluted from agarose gel followed by the excision of the corresponding band. Subsequently, the RNA was purified by phenol/chloroform extraction (using low melting gel agarose) or recovered by glass wool filtration (normal agarose gel). (B) Direct purification of the ligated RNA was performed by the application on the Silica spin column. (C) Another method was based on gel elution. The ligation mixture was loaded on a 1% agarose gel. The gel was prepared to contain a large pocket in the middle. This allows the recovery of the ligated product while migrating through the gel. Therefore, the electrophoresis apparatus was equipped with a transilluminator to allow the visualization and the follow up of the RNA migration in real-time. The recovered RNA was finally filtered through a solid phase filter with a pore size of 0.2 μm .

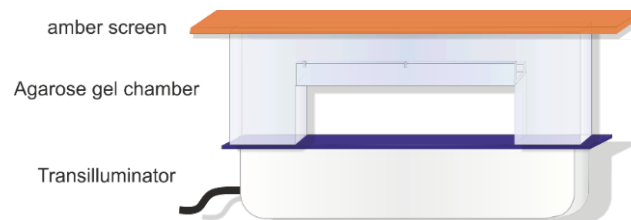


Figure 6. 8: Set-up tool for real-time gel elution.

The agarose gel lays on the top of a transilluminator, which emits visible blue light. This allows the visualization of migrating RNA and its elution. The orange screen allows the protection from the brightness of the blue light during RNA elution (pipetting).

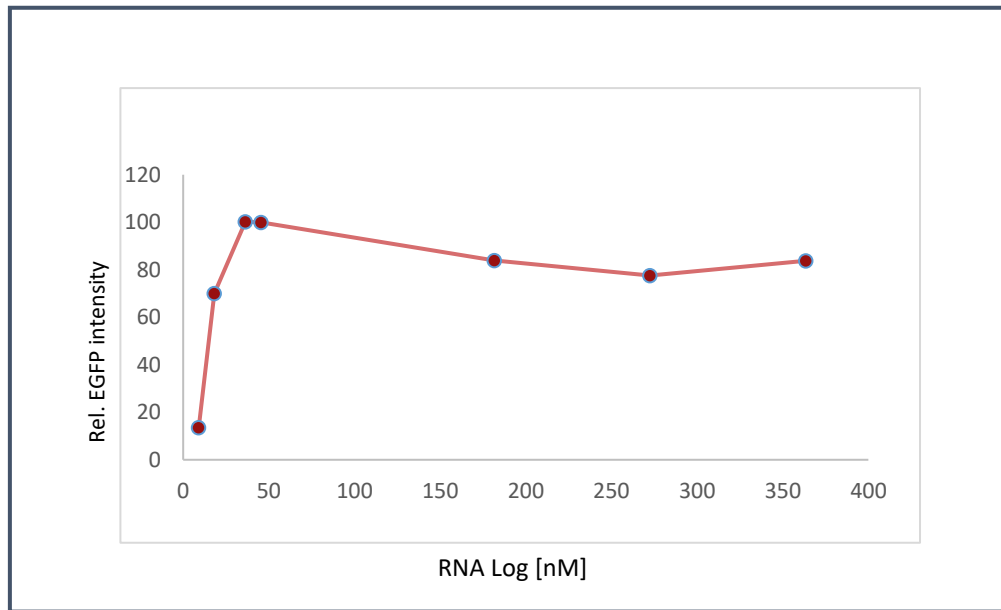


Figure 6. 9: Optimization of RNA concentration for translation in RRL system.

Nuclease treated RRL was programmed with increasing amounts of RNA (0.5, 1, 1.8, 2.2, 9, 13.5 and 18 pmol) in a final volume of 50 μ L. After translation, aliquots were separated on 10% SDS-PAGE, and EGFP fluorescence was in-gel detected by blue laser settings (488 nm/ 520 nm excitation/emission wavelengths) to 1.8 pmol (amount recommended by the manufacturer) .

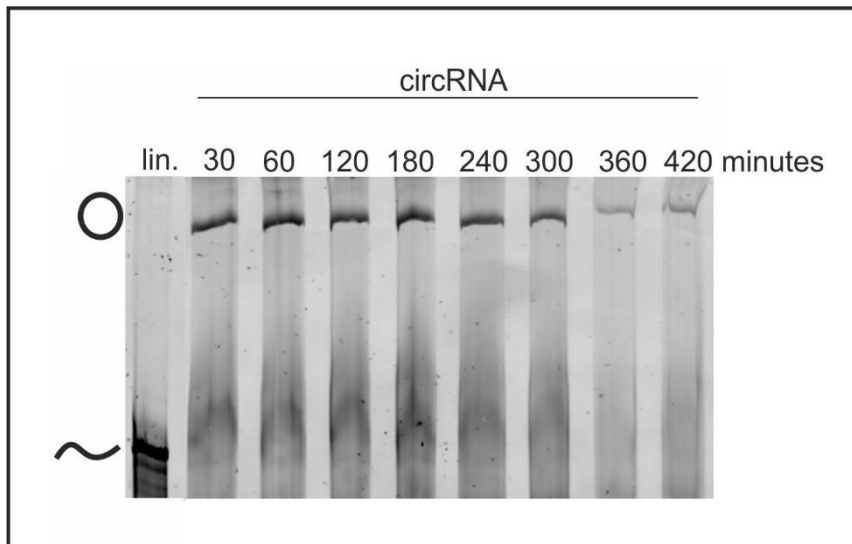


Figure 6. 10: Circularization reaction optimization.

For circularization, IRES EMCV mRNA was transcribed *in vitro*, hybridized to a splint DNA to join both ends together. After that, T4 RNA ligase II was added and the ligation reaction was incubated at 37°C for 30 , 60, 120, 180, 240, 300 ,360, and 420 minutes. Aliquots from the ligation reaction were separated on a 4% denaturing PAGE followed by staining with GelRed.

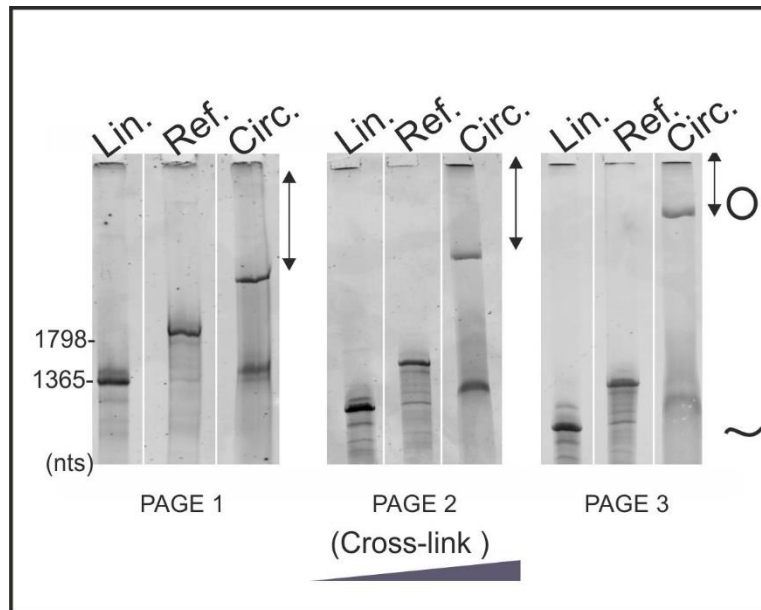


Figure 6. 11: Analysis of circular RNA migration on increasing cross-linking PAGE.

An aliquot from circular RNA reaction (Circ.) was separated in urea-denatured PAGE with different degrees of crosslinking, the ratio of the bisacrylamide to acrylamide (59:1, 39:1 and 19:1) to increase the crosslink from PAGE 1 to PAGE 3. Positions of circular and linear (Lin.) forms are indicated. A reference mRNA from a known size (1789 nts) was used (Ref.). The migration was carried out 7 hours and 10 minutes under cooling. The distance migration of the circular form was indicated by an arrow. The migration distance was recorded in Figure 3. 17. The gels were visualized by GelRed staining.

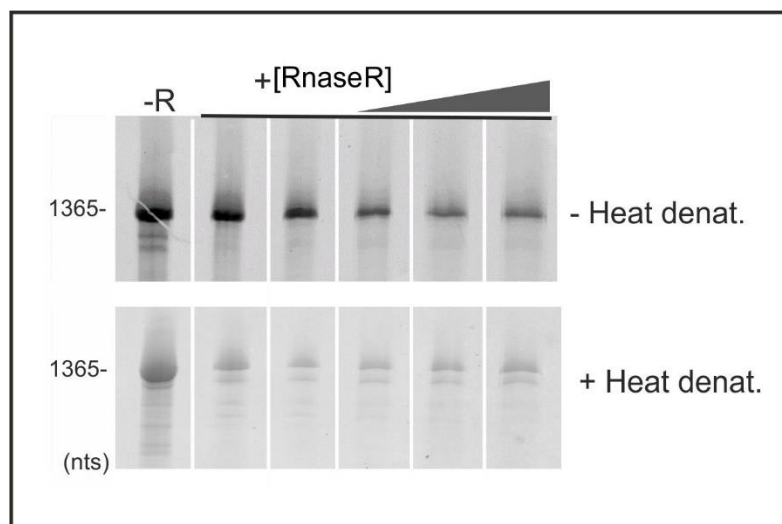


Figure 6. 12: RNA treatment with RNase R.

The *in vitro* transcribed mRNA (2.5 µg) was digested with increasing amount of RNase R (0.05, 0.1, 0.2, 0.4 and 0.6 U/µL). RNA was either directly digested (- preheating denaturing) or heat-denatured prior to digestion (+preheat denat.). After incubation for 3 hours at 37°C, aliquots from reactions were separated on a 6% denaturing PAGE. The gel was revealed by staining with GelRed.

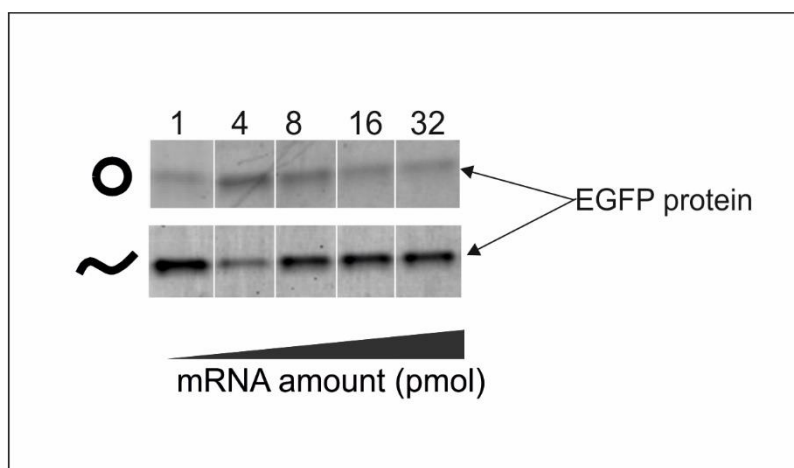


Figure 6. 13: Translation of HPLC-purified circRNA.

After circularization, RNA was purified by injection into RP-HPLC and the corresponding collected fraction was ethanol precipitated. Increasing amounts of purified circRNA were programmed for translation in nuclease treated RRL. Synthesized protein was separated on 10 SDS PAGE and analyzed for EGFP fluorescence by in-gel detection.

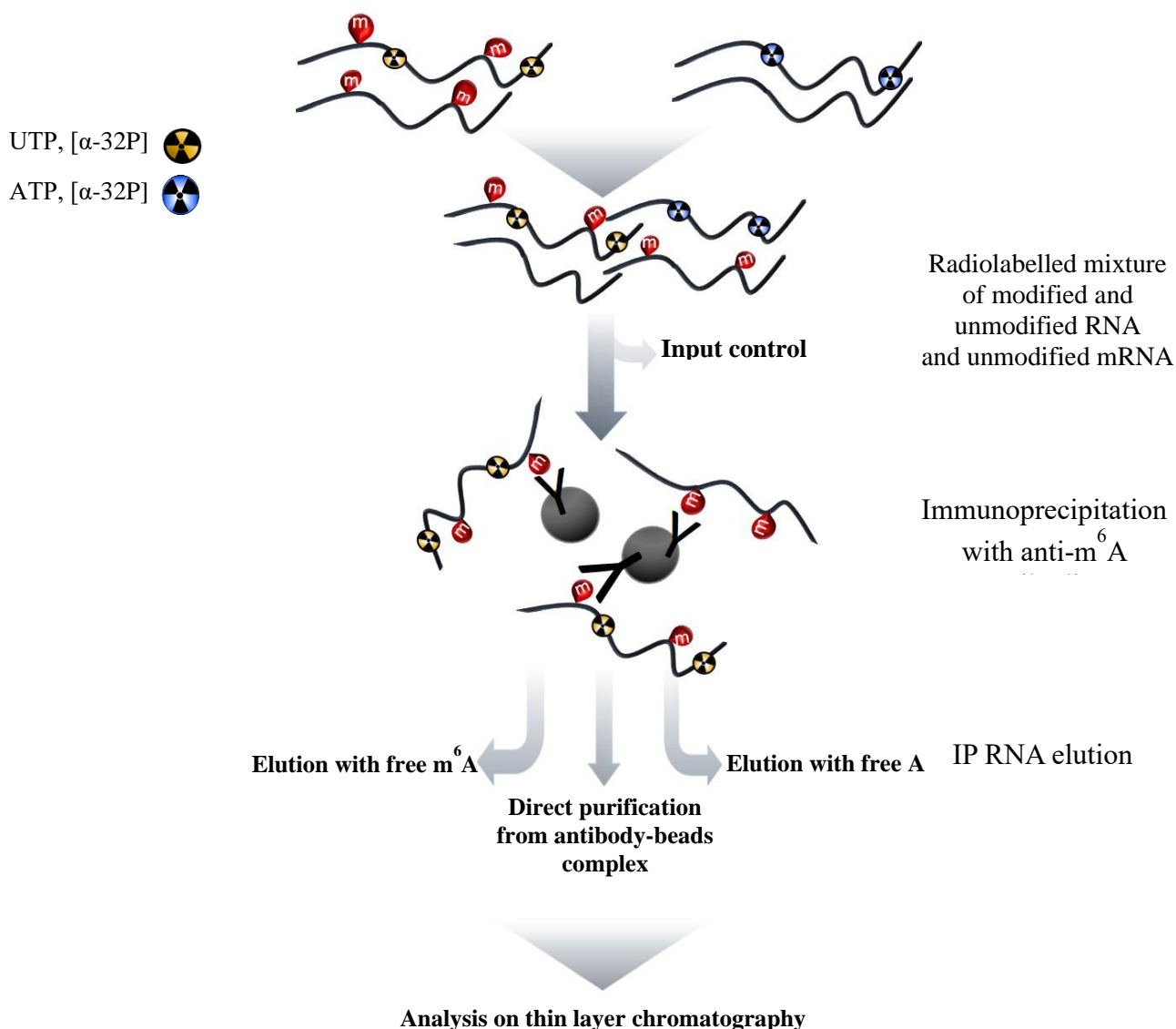


Figure 6. 14: Schematic overview of the m⁶A immunoprecipitation protocol for enrichment factor determination.

Modified and unmodified RNA were synthesized and labelled with ³²P *in vitro* and subjected to immunoprecipitation using m⁶A antibody. The bound RNA (IP) was extracted directly from antibody-beads complex, or eluted by competitive assay using free m⁶A or using free A residues. Eluted RNA in the supernatant fraction was ethanol precipitated while RNA in the antibody-beads fraction was solvent extracted. Immunoprecipitated RNA and input control were digested by nuclease P1. The obtained nucleosides were analyzed on TLC and enrichment factor was determined as described.

References

1. CRICK, F. Central Dogma of Molecular Biology. *Nature* **227**, 309–309 (1970).
2. Bogdanov, A. & Chemistry, A. O. Shabarova, A. Bogdanov Advanced Organic Chemistry of Nucleic Acids. *VCH* (1994).
3. Sisodia, S. S., Sollner-Webb, B. & Cleveland, D. W. Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing or polyadenylation. *Mol. Cell. Biol.* **7**, 3602–3612 (1987).
4. McCracken, S. *et al.* Yeast Transcriptional Regulatory Mechanisms. *Annu. Rev. Genet.* **495**, 481–486 (1997).
5. Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**, 327–335 (1996).
6. Leigh Zawel, D. R. Initiation of Transcription by RNA Polymerase II: A Multi-step Process. *Prog. Nucleic Acid Res. Mol. Biol.* **Volume 44**, Pages 67-108 (1993).
7. Struhl, K. Yeast Transcriptional Regulatory Mechanisms. *AMIL Rev. Genet.* **29**, 651–74 (1995).
8. Wang, W., Carey, M. & Gralla, J. a Y. D. Polymerase II Promoter Activation : Closed Complex. *Science (80-.).* **255**, 450–453 (1992).
9. Deng, W. & Roberts, S. G. E. TFIIB and the regulation of transcription by RNA polymerase II. *Chromosoma* **116**, 417–429 (2007).
10. Kulaeva, O. I., Nizovtseva, E. V., Polikanov, Y. S., Ulianov, S. V. & Studitsky, V. M. Distant Activation of Transcription: Mechanisms of Enhancer Action. *Mol. Cell. Biol.* **32**, 4892–4897 (2012).
11. Bieberstein, N. I., Oesterreich, F. C., Straube, K. & Neugebauer, K. M. First exon length controls active chromatin signatures and transcription. *Cell Rep.* **2**, 62–68 (2012).
12. Allison, L. A., Moyle, M., Shales, M. & James Ingles, C. Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42**, 599–610 (1985).
13. Mortillaro, M. J. *et al.* A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 8253–8257 (1996).
14. McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**, 357–361 (1997).
15. Du, L. & Warren, S. L. A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. **136**, 5–18 (1997).
16. Ramanathan, A., Robb, G. B. & Chan, S. H. mRNA capping: Biological functions and applications. *Nucleic Acids Res.* **44**, 7511–7526 (2016).
17. Lockless, S. W., Cheng, H. T., Hodel, A. E., Quijcho, F. A. & Gershon, P. D. Recognition of

References

- capped RNA substrates by VP39, the vaccinia virus- encoded mRNA cap-specific 2'-O-methyltransferase. *Biochemistry* **37**, 8564–8574 (1998).
18. Kessler, O., Jiang, Y. & Chasin, L. A. Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Mol. Cell. Biol.* **13**, 6211–6222 (1993).
 19. Black, D. L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
 20. Ewa Grudzien-Nogalska and Megerditch Kiledjian. New Insights into Decapping Enzymes and Selective mRNA Decay. *Physiol. Behav.* **176**, 139–148 (2017).
 21. Grudzien-Nogalska, E., Jiao, X., Song, M. G., Hart, R. P. & Kiledjian, M. Nudt3 is an mRNA decapping enzyme that modulates cell migration. *RNA* **22**, 773–781 (2016).
 22. Dunkley, T. & Parker, R. The DCP2 protein is required for mRNA decapping in *Saccharomyces cerevisiae* and contains a functional MutT motif. *EMBO J.* **18**, 5411–5422 (1999).
 23. Sadofsky, M. & Alwine, J. C. Sequences on the 3' side of hexanucleotide AAUAAA affect efficiency of cleavage at the polyadenylation site. *Mol. Cell. Biol.* **4**, 1460–1468 (1984).
 24. Takagaki, Y. & Manley, J. L. RNA recognition by the human polyadenylation factor CstF. *Mol. Cell. Biol.* **17**, 3907–3914 (1997).
 25. Yang, Q., Gilmartin, G. M. & Doublié, S. Structural basis of UGUA recognition by the Nudix protein CFIm25 and implications for a regulatory role in mRNA 3' processing. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10062–10067 (2010).
 26. Logan, J., Falck-Pedersen, E., Darnell, J. E. & Shenk, T. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc. Nat. Acad. Sci.* **84**, 8306–8310 (1987).
 27. Douglas R.Drummond, J. A. and A. C. The effect of capping and polyadenylation on the stability, movement and translation of synthetic messenger RNAs in *Xenopus* oocytes. *Nucleic Acids Res.* **11**, 7375–7394 (1985).
 28. Munroe, D. & Jacobson, A. mRNA poly(A) tail, a 3' enhancer of translational initiation. *Mol. Cell. Biol.* **10**, 3441–3455 (1990).
 29. Gray, N. K. Multiple portions of poly(A)-binding protein stimulate translation in vivo. *EMBO J.* **19**, 4723–4733 (2000).
 30. Kozak, M. Possible role of flanking nucleotides in recognition of the aug initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.* **9**, 5233–5252 (1981).
 31. Tomek, W. & Wollenhaupt, K. The 'closed loop model' in controlling mRNA translation during development. *Anim. Reprod. Sci.* **134**, 2–8 (2012).
 32. Bonderoff, J. M. & Lloyd, R. E. Time-dependent increase in ribosome processivity. *Nucleic Acids Res.* **38**, 7054–7067 (2010).
 33. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The mechanism of eukaryotic translation

References

- initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* **11**, 113–127 (2010).
34. Jang, S. K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* **62**, 2636–2643 (1988).
35. Holčík, M., Gordon, B. W. & Korneluk, R. G. The Internal Ribosome Entry Site-Mediated Translation of Antiapoptotic Protein XIAP Is Modulated by the Heterogeneous Nuclear Ribonucleoproteins C1 and C2. *Mol. Cell. Biol.* **23**, 280–288 (2003).
36. Martínez-Salas, E., Ramos, R., Lafuente, E. & De Quinto, S. L. Functional interactions in internal translation initiation directed by viral and cellular IRES elements. *J. Gen. Virol.* **82**, 973–984 (2001).
37. Jackson, R. J. The current status of vertebrate cellular mRNA IRESs. *Cold Spring Harb. Perspect. Biol.* **5**, (2013).
38. Kaminski, A., Howell, M. T. & Jackson, R. J. Initiation of encephalomyocarditis virus RNA translation: The authentic initiation site is not selected by a scanning mechanism. *EMBO J.* **9**, 3753–3759 (1990).
39. Thapar, R., Denmon, A. P. & Nikonowicz, E. P. Recognition modes of RNA tetraloops and tetraloop-like motifs by RNA-binding proteins. *Wiley Interdiscip. Rev. RNA* **5**, 49–67 (2014).
40. Fernández-Miragall, O., Ramos, R., Ramajo, J. & Martínez-Salas, E. Evidence of reciprocal tertiary interactions between conserved motifs involved in organizing RNA structure essential for internal initiation of translation. *RNA* **12**, 223–234 (2006).
41. López de Quinto, S. & Martínez-Salas, E. Conserved structural motifs located in distal loops of aphthovirus internal ribosome entry site domain 3 are required for internal initiation of translation. *J. Virol.* **71**, 4171–4175 (1997).
42. López De Quinto, S., Lafuente, E. & Martínez-Salas, E. IRES interaction with translation initiation factors: Functional characterization of novel RNA contacts with eIF3, eIF4B, and eIF4GII. *RNA* **7**, 1213–1226 (2001).
43. Luz, N. & Beck, E. Interaction of a cellular 57-kilodalton protein with the internal translation initiation site of foot-and-mouth disease virus. *J. Virol.* **65**, 6486–6494 (1991).
44. Ochs, K., Rust, R. & Niepmann, M. Translation Initiation Factor eIF4B Interacts with a Picornavirus Internal Ribosome Entry Site in both 48S and 80S Initiation Complexes Independently of Initiator AUG Location. *J. Virol.* **73**, 7505–7514 (1999).
45. Martinez-Salas, E., Francisco-Velilla, R., Fernandez-Chamorro, J. & Embarek, A. M. Insights into structural and mechanistic features of viral IRES elements. *Front. Microbiol.* **8**, 1–15 (2018).
46. Kolupaeva, V. G., Lomakin, I. B., Pestova, T. V. & Hellen, C. U. T. Eukaryotic Initiation Factors 4G and 4A Mediate Conformational Changes Downstream of the Initiation Codon of the Encephalomyocarditis Virus Internal Ribosomal Entry Site. *Mol. Cell. Biol.* **23**, 687–698 (2003).
47. Lomakin, I. B. & Hellen, C. U. T. Physical association of eukaryotic initiation factor 4G (eIF4G) with eIF4A strongly enhances binding of eIF4G to the internal ribosomal entry site of

References

- encephalomyocarditis.pdf. **20**, 6019–6029 (2000).
48. Pestova, T. V, Hellen, C. U. & Shatsky, I. N. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol. Cell. Biol.* **16**, 6859–6869 (1996).
49. Kolupaeva, V. G., Pestova, T. V., Hellen, C. U. T. & Shatsky, I. N. Translation eukaryotic initiation factor 4G recognizes a specific structural element within the internal ribosome entry site of encephalomyocarditis virus RNA. *J. Biol. Chem.* **273**, 18599–18604 (1998).
50. Pestova, T. V, Shatsky, I. N. & Hellen, C. U. Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Mol. Cell. Biol.* **16**, 6870–6878 (1996).
51. Jan, E. & Sarnow, P. Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus. *J. Mol. Biol.* **324**, 889–902 (2002).
52. He, C. Grand Challenge Commentary: RNA epigenetics? *Nat. Chem. Biol.* **6**, 863–865 (2010).
53. Motorin, Y. & Helm, M. Methods for RNA modification mapping using deep sequencing: Established and new emerging technologies. *Genes (Basel)*. **10**, (2019).
54. Boccaletto, P. *et al.* MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res.* **46**, D303–D307 (2018).
55. Legnini, I. *et al.* Circ-ZNF609 Is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell* **66**, 22-37.e9 (2017).
56. Pendleton, K. E. *et al.* The U6 snRNA m6A Methyltransferase METTL16 Regulates SAM Synthetase Intron Retention. *Cell* **169**, 824-835.e14 (2017).
57. Roundtree, I. A., Evans, M. E., Pan, T. & He, C. Dynamic RNA Modifications in Gene Expression Regulation. *Cell* **169**, 1187–1200 (2017).
58. Xiang, Y. *et al.* m6A RNA methylation regulates the UV-induced DNA damage response. *Nature* **543**, 573–576 (2017).
59. Marchand, V., Blanloeil-Oillo, F., Helm, M. & Motorin, Y. Illumina-based RiboMethSeq approach for mapping of 2'-O-Me residues in RNA. *Nucleic Acids Res.* **44**, (2016).
60. Sharma, S. & Lafontaine, D. L. J. ‘View From A Bridge’: A New Perspective on Eukaryotic rRNA Base Modification. *Trends Biochem. Sci.* **40**, 560–575 (2015).
61. Sergeeva, O. V., Bogdanov, A. A. & Sergiev, P. V. What do we know about ribosomal RNA methylation in *Escherichia coli*? *Biochimie* **117**, 110–118 (2015).
62. Bohnsack, N. J. W. and M. T. The box C/D and H/ACA snoRNPs: Key players in the modification, processing and the dynamic folding of ribosomal RNA. *Wiley Interdiscip Rev RNA* **3**, 397–414 (2012).
63. Bakin, A., Lane, B. G. & Ofengand, J. Clustering of Pseudouridine Residues around the Peptidyltransferase Center of Yeast Cytoplasmic and Mitochondrial Ribosomes. *Biochemistry* **33**, 13475–13483 (1994).
64. Zebarjadian, Y., King, T., Fournier, M. J., Clarke, L. & Carbon, J. Point Mutations in Yeast

References

- CBF5 Can Abolish In Vivo Pseudouridylation of rRNA. *Mol. Cell. Biol.* **19**, 7461–7472 (1999).
65. Tollervey, D., Lehtonen, H., Jansen, F., Kern, H. & Hurt, E. C. Temperature-sensitive mutations demonstrate roles for yeast fibrillarin in pre-rRNA processing, pre-rRNA methylation, and ribosome assembly. **72**, 443–457 (1993).
66. Raychaudhuri, S., Conrad, J., Hall, B. G. & Ofengand, J. A pseudouridine synthase required for the formation of two universally conserved pseudouridines in ribosomal RNA is essential for normal growth of *Escherichia coli*. *RNA* **4**, 1407–1417 (1998).
67. Pan, T. Modifications and functional genomics of human transfer RNA. *Cell Res.* **28**, 395–404 (2018).
68. Koh, C. S. & Sarin, L. P. Transfer RNA modification and infection – Implications for pathogenicity and host responses. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1861**, 419–432 (2018).
69. Madore, E. *et al.* Effect of modified nucleotides on *Escherichia coli* tRNA(Glu) structure and on its aminoacylation by glutamyl-tRNA synthetase. Predominant and distinct roles of the mnm5 and s2 modifications of U34. *Eur. J. Biochem.* **266**, 1128–1135 (1999).
70. Agris, P. F., Vendeix, F. A. P. & Graham, W. D. tRNA's Wobble Decoding of the Genome: 40 Years of Modification. *J. Mol. Biol.* **366**, 1–13 (2007).
71. Crick, F. H. C. Codon—anticodon pairing: The wobble hypothesis. *J. Mol. Biol.* **19**, 548–555 (1966).
72. Jackman, J. E. & Alfonzo, J. D. Transfer RNA modifications: Nature's combinatorial chemistry playground. *Wiley Interdiscip. Rev. RNA* **4**, 35–48 (2013).
73. Näsval, S. J., Chen, P. & Björk, G. R. The wobble hypothesis revisited: Uridine-5-oxyacetic acid is critical for reading of G-ending codons. *RNA* **13**, 2151–2164 (2007).
74. Schaffer, S. W., Jong, C. J., Ito, T. & Azuma, J. Role of taurine in the pathologies of MELAS and MERRF. *Amino Acids* **46**, 47–56 (2014).
75. Yasukawa, T. *et al.* Modification defect at anticodon wobble nucleotide of mitochondrial tRNAs(Leu)(UUR) with pathogenic mutations of mitochondrial myopathy, encephalopathy, lactic acidosis, and stroke-like episodes. *J. Biol. Chem.* **275**, 4251–4257 (2000).
76. Goto, Y. I., Nonaka, I. & Horai, S. A mutation in the tRNA^{Leu}(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature* **348**, 651–653 (1990).
77. Guy, M. P. *et al.* Defects in tRNA anticodon loop 2'-O-methylation are implicated in non-syndromic X-linked intellectual disability due to mutations in FTSJ1. *Hum Mutat.* **36**, 1176–1187 (2015).
78. Franz, M., Hagenau, L., Jensen, L. R. & Kuss, A. W. Role of transfer RNA modification and aminoacylation in the etiology of congenital intellectual disability. *J. Transl. Genet. Genomics* 50–70 (2020) doi:10.20517/jtgg.2020.13.
79. Tuorto, F. & Lyko, F. Genome recoding by tRNA modifications. *Open Biol.* **6**, 0–2 (2016).

References

80. Dihanich, M. E. *et al.* Isolation and characterization of MOD5, a gene required for isopentenylolation of cytoplasmic and mitochondrial tRNAs of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **7**, 177–184 (1987).
81. Björk, G. R. & Nilsson, K. 1-Methylguanosine-deficient tRNA of *Salmonella enterica* serovar Typhimurium affects thiamine metabolism. *J. Bacteriol.* **185**, 750–759 (2003).
82. Helm, M., Giegé, R. & Florentz, C. A Watson-Crick base-pair-disrupting methyl group (m1A9) is sufficient for cloverleaf folding of human mitochondrial tRNA(Lys). *Biochemistry* **38**, 13338–13346 (1999).
83. Mathew, S. F. *et al.* The highly conserved codon following the slippery sequence supports - 1 frameshift efficiency at the HIV-1 frameshift site. *PLoS One* **10**, 1–24 (2015).
84. Renda, M. J. *et al.* Mutation of the Methylated tRNA³Lys Residue A58 Disrupts Reverse Transcription and Inhibits Replication of Human Immunodeficiency Virus Type 1. *J. Virol.* **75**, 9671–9678 (2001).
85. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 3971–3975 (1974).
86. Dominissini, D., Moshitch-Moshkovitz, S., Salmon-Divon, M., Amariglio, N. & Rechavi, G. Transcriptome-wide mapping of N6-methyladenosine by m6A-seq based on immunocapturing and massively parallel sequencing. *Nat. Protoc.* **8**, 176–189 (2013).
87. Edelheit, S., Schwartz, S., Mumbach, M. R., Wurtzel, O. & Sorek, R. Transcriptome-Wide Mapping of 5-methylcytidine RNA Modifications in Bacteria, Archaea, and Yeast Reveals m5C within Archaeal mRNAs. *PLoS Genet.* **9**, (2013).
88. Schwartz, S. *et al.* Transcriptome-wide mapping reveals widespread dynamic-regulated pseudouridylation of ncRNA and mRNA. *Cell* **159**, 148–162 (2014).
89. Thomas M. Carlile, Maria F. Rojas-Duran, Boris Zinshteyn, Hakyung Shin, Kristen M. Bartoli, and W. V. G. Pseudouridine profiling reveals regulated mRNA pseudouridylation in yeast and human cells. *Nature* **515**, 143–146 (2014).
90. Dominissini, D. *et al.* The dynamic N1-methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).
91. Dai, Q. *et al.* Nm-seq maps 2'-O-methylation sites in human mRNA with base precision. *Nat Methods* **14**, 695–698 (2017).
92. Li, S. & Mason, C. E. The pivotal regulatory landscape of RNA modifications. *Annu. Rev. Genomics Hum. Genet.* **15**, 127–150 (2014).
93. Wei, C. M., Gershowitz, A. & Moss, B. Methylated nucleotides block 5' terminus of HeLa cell messenger RNA. *Cell* **4**, 379–386 (1975).
94. Ramanathan, A., Robb, G. B. & Chan, S. H. mRNA capping: Biological functions and applications. *Nucleic Acids Res.* **44**, 7511–7526 (2016).
95. Züst, R. *et al.* Ribose 2'-O-methylation provides a molecular signature for the distinction of self

References

- and non-self mRNA dependent on the RNA sensor Mda5. *Nat. Immunol.* **12**, 137–143 (2011).
96. Goodfellow, I. G. & Roberts, L. O. Molecules in Focus Eukaryotic initiation factor 4E. *Eur. PMC Funders Gr.* **40**, 2675–2680 (2011).
97. Stephen Buratowski. Progression through the RNA polymerase II CTD cycle Stephen. *Bone* **36**, 541–546 (2009).
98. Fuguo Jiang, Anand Ramanathan, Matthew T. Miller, Guo-Qing Tang, Michael Gale Jr., Smita S. Patel, and Joseph Marcotrigiano1Fuguo Jiang1,*, Anand Ramanathan Matthew T. Miller, Guo-Qing Tang, Michael Gale Jr., Smita S. Patel2, and J. M. Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Nature* **479**, 423–427 (2011).
99. Langberg, S. R. & Moss, B. Post-transcriptional modifications of mRNA. Purification and characterization of cap I and cap II RNA (nucleoside-2'-)-methyltransferases from HeLa cells. *J. Biol. Chem.* **256**, 10054–10060 (1981).
100. Sikorski, P. J. *et al.* The identity and methylation status of the first transcribed nucleotide in eukaryotic mRNA 5' cap modulates protein expression in living cells. *Nucleic Acids Res.* **48**, 1607–1626 (2020).
101. Linder, B. *et al.* Single-nucleotide resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12**, 767–772 (2015).
102. Mauer, J. *et al.* Reversible methylation of m6 Am in the 5' cap controls mRNA stability. *Nature* **541**, 371–375 (2017).
103. Li, X. *et al.* Transcriptome-wide mapping reveals reversible and dynamic N1-methyladenosine methylome. *Nat. Chem. Biol.* **12**, 311–316 (2016).
104. Dominissini, D. *et al.* The dynamic N1 -methyladenosine methylome in eukaryotic messenger RNA. *Nature* **530**, 441–446 (2016).
105. Hauenschild, R. *et al.* The reverse transcription signature of N-1-methyladenosine in RNA-Seq is sequence. **43**, 9950–9964 (2015).
106. You, C., Dai, X. & Wang, Y. Position-dependent effects of regioisomeric methylated adenine and guanine ribonucleosides on translation. *Nucleic Acids Res.* **45**, 9059–9067 (2017).
107. Amort, T. *et al.* Distinct 5-methylcytosine profiles in poly(A) RNA from mouse embryonic stem cells and brain. *Genome Biol.* **18**, 1–16 (2017).
108. Yang, X. *et al.* 5-methylcytosine promotes mRNA export-NSUN2 as the methyltransferase and ALYREF as an m 5 C reader. *Cell Res.* **27**, 606–625 (2017).
109. Cui, X. *et al.* 5-Methylcytosine RNA Methylation in Arabidopsis Thaliana. *Mol. Plant* **10**, 1387–1399 (2017).
110. Squires, J. E. *et al.* Widespread occurrence of 5-methylcytosine in human coding and non-coding RNA. *Nucleic Acids Res.* **40**, 5023–5033 (2012).
111. Wei, C. M., Gershowitz, A. & Moss, B. 5'-Terminal and Internal Methylated Nucleotide

References

- Sequences in Hela Cell mRNA. *Biochemistry* **15**, 397–401 (1976).
112. Perry, R. P. & Kelley, D. E. Existence of methylated messenger RNA in mouse L cells. *Cell* **1**, 37–42 (1974).
113. Perry, R. P., Kelley, D. E., Friderici, K. & Rottman, F. The methylated constituents of L cell messenger RNA: Evidence for an unusual cluster at the 5' terminus. *Cell* **4**, 387–394 (1975).
114. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).
115. Schwartz, S. *et al.* Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. *Cell Rep.* **8**, 284–296 (2014).
116. Wei, C. M. & Moss, B. Nucleotide Sequences at the N6-Methyladenosine Sites of HeLa Cell Messenger Ribonucleic Acid. *Biochemistry* **16**, 1672–1676 (1977).
117. Manuscript, A. & Codons, N. S. Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and Near Stop Codons. **149**, 1635–1646 (2013).
118. Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G. & Rottman, F. M. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA (N6-adenosine)-methyltransferase. *Rna* **3**, 1233–1247 (1997).
119. Zaccara, S., Ries, R. J. & Jaffrey, S. R. Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.* **20**, 608–624 (2019).
120. Slobodin, B. *et al.* Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. *Cell* **169**, 326–337.e12 (2017).
121. Zhong, S. *et al.* MTA is an Arabidopsis messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20**, 1278–1288 (2008).
122. Ping, X. L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* **24**, 177–189 (2014).
123. Schwartz, S. *et al.* Perturbation of m6A writers reveals two distinct classes of mRNA methylation at internal and 5' sites. **8**, 284–296 (2015).
124. Zheng, G. *et al.* ALKBH5 Is a Mammalian RNA Demethylase that Impacts RNA Metabolism and Mouse Fertility. *Mol Cell.* **49**, 18–29 (2013).
125. Zhao, X. *et al.* FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.* **24**, 1403–1419 (2014).
126. Jia, G. *et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol.* **7**, 885–887 (2012).
127. Wang, X. *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).
128. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol. Cell* **61**, 507–519 (2016).
129. Jun Zhou, Ji Wan, Xiangwei Gao, Xingqian Zhang, and S.-B. Q. Dynamic m6A mRNA

References

- methylation directs translational control of heat shock response. *Nature*. **526**, 591–594 (2015).
130. Carroll, S. M., Narayan, P. & Rottman, F. M. N6-methyladenosine residues in an intron-specific region of prolactin pre-mRNA. *Mol. Cell. Biol.* **10**, 4456–4465 (1990).
131. Roundtree, I. A. *et al.* YTHDC1 mediates nuclear export of N6-methyladenosine methylated mRNAs. *Elife* **6**, 1–28 (2017).
132. Meyer, K. D. *et al.* 5' UTR m6A Promotes Cap-Independent Translation. *Cell* **163**, 999–1010 (2015).
133. Ryan A. Coots, Xiao-Min Liu, Yuanhui Mao, Leiming Dong, Jun Zhou, Ji Wan, Xingqian Zhang, and S.-B. Q. m6A Facilitates eIF4F-Independent mRNA Translation. *Mol Cell*. **68(3)**, 504–514 (2017).
134. Xiao Wang, Boxuan Simen Zhao, Ian A. Roundtree, Zhike Lu, D. H. & Honghui Ma, Xiaocheng Weng, Kai Chen, Hailing Shi, and C. H. N6-methyladenosine Modulates Messenger RNA Translation Efficiency. *Physiol. Behav.* **176**, 100–106 (2016).
135. Zhou, J. *et al.* Dynamic m6 A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594 (2015).
136. Zhou, J. *et al.* N6-Methyladenosine Guides mRNA Alternative Translation during Integrated Stress Response. *Mol. Cell* **69**, 636–647.e7 (2018).
137. Jun Zhou, Ji Wan, Xin Erica Shu, Yuanhui Mao, Xiao-Min Liu, Xin Yuan, Xingqian Zhang, Martin E. Hess, Jens C. Brüning, and S.-B. Q. N6-Methyladenosine Guides mRNA Alternative Translation during Integrated Stress Response. *Mol Cell*. **69**, 636–647.e7 (2018).
138. Liu, N. *et al.* N6 -methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* **518**, 560–564 (2015).
139. Kierzek, E. & Kierzek, R. The synthesis of oligoribonucleotides containing N6-alkyladenosines and 2-methylthio-N6- alkyladenosines via post-synthetic modification of precursor oligomers. *Nucleic Acids Res.* **31**, 4461–4471 (2003).
140. Hoernes, T. P. *et al.* Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res.* **44**, 852–862 (2016).
141. Junhong Choi, Ka-Weng Jeong, Hasan Demirci, Jin Chen, Alexey Petrov, Arjun Prabhakar^{1,6}, Seán E. O’Leary¹, Dan Dominissini⁷, Gideon Rechavi, S. Michael Soltis⁵, Måns Ehrenberg, and J. D. P. N6 -methyladenosine in mRNA disrupts tRNA selection and translation elongation dynamics. *Nat Struct Mol Biol.* **23**, 110–115 (2016).
142. Du, H. *et al.* YTHDF2 destabilizes m 6 A-containing RNA through direct recruitment of the CCR4-NOT deadenylase complex. *Nat. Commun.* **7**, 1–11 (2016).
143. Helm, M., Lyko, F. & Motorin, Y. Limited antibody specificity compromises epitranscriptomic analyses. *Nat. Commun.* **10**, 5669 (2019).
144. Slama, K. *et al.* Determination of enrichment factors for modified RNA in MeRIP experiments. *Methods* **156**, 102–109 (2019).

References

145. Lin, S., Choe, J., Du, P., Triboulet, R. & Gregory, R. I. The m6A Methyltransferase METTL3 Promotes Translation in Human Cancer Cells. *Mol. Cell* **62**, 335–345 (2016).
146. Su, R. *et al.* Fto Plays an Oncogenic Role in Acute Myeloid Leukemia As a N6-Methyladenosine RNA Demethylase. *Blood* **128**, 2706–2706 (2016).
147. Han, M. *et al.* Abnormality of m6A mRNA Methylation Is Involved in Alzheimer's Disease. *Front. Neurosci.* **14**, 1–9 (2020).
148. Zhang, C., Fu, J. & Zhou, Y. A review in research progress concerning m6A methylation and immunoregulation. *Front. Immunol.* **10**, (2019).
149. WALDO E. COHN, E. V. Nucleoside-5'-Phosphates from Ribonucleic Acid. **167**, 483–484 (1951).
150. Allen, F. F. D. and F. W. Ribonucleic acids from yeast which contain a fifth nucleotide. *J. Biol. Chem.* **227**, 907–915 (1957).
151. Li, X., Ma, S. & Yi, C. Pseudouridine: the fifth RNA nucleotide with renewed interests. *Curr. Opin. Chem. Biol.* **33**, 108–116 (2016).
152. CHUAN-TAO YU, F. W. A. Studies of an isomer of uridine isolated from ribonucleic acids. *Biochim. Biophys. Acta* **32**, 405–406 (1959).
153. Cohn, W. E. 5-Ribosyl uracil, a carbon-carbon ribofuranosyl nucleoside in ribonucleic acids. *BBA - Biochim. Biophys. Acta* **32**, 569–571 (1959).
154. Tardu, M., Jones, J. D., Kennedy, R. T., Lin, Q. & Koutmou, K. S. Identification and Quantification of Modified Nucleosides in *Saccharomyces cerevisiae* mRNAs. *ACS Chem. Biol.* **14**, 1403–1409 (2019).
155. Li, X. *et al.* Chemical pulldown reveals dynamic pseudouridylation of the mammalian transcriptome. *Nat. Chem. Biol.* **11**, 592–597 (2015).
156. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m6A RNA methylation. *Nat. Rev. Genet.* **15**, 293–306 (2014).
157. Lovejoy, A. F., Riordan, D. P. & Brown, P. O. Transcriptome-wide mapping of pseudouridines: Pseudouridine synthases modify specific mRNAs in *S. cerevisiae*. *PLoS One* **9**, (2014).
158. Charette, M. & Gray, M. W. Pseudouridine in RNA: What, where, how, and why. *IUBMB Life* **49**, 341–351 (2000).
159. Wendy V. Gilbert, T. A. B. and C. S. Messenger RNA modifications: form, distribution, and function. *Science*. **352**, 1408–1412 (2016).
160. Penzo, M., Guerrieri, A. N., Zacchini, F., Treré, D. & Montanaro, L. RNA pseudouridylation in physiology and medicine: For better and for worse. *Genes (Basel)*. **8**, 1–14 (2017).
161. Rintala-Dempsey, A. C. & Kothe, U. Eukaryotic stand-alone pseudouridine synthases—RNA modifying enzymes and emerging regulators of gene expression? *RNA Biol.* **14**, 1185–1196 (2017).
162. Ge, J. & Yu, Y. T. RNA pseudouridylation: New insights into an old modification. *Trends*

References

- Biochem. Sci.* **38**, 210–218 (2013).
163. Carlile, T. M. *et al.* mRNA structure determines modification by pseudouridine synthase 1. *Nat. Chem. Biol.* **15**, 966–974 (2019).
 164. Chen, C., Zhao, X., Kierzek, R. & Yu, Y.-T. A Flexible RNA Backbone within the Polypyrimidine Tract Is Required for U2AF65 Binding and Pre-mRNA Splicing InVivo. *Mol. Cell. Biol.* **30**, 4108–4119 (2010).
 165. Hudson, G. A., Bloomingdale, R. J. & Znosko, B. M. Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides. *RNA* **19**, 1474–1482 (2013).
 166. Yu, Y. T. & Meier, U. T. RNA-guided isomerization of uridine to pseudouridine - Pseudouridylation. *RNA Biol.* **11**, 1483–1494 (2014).
 167. Zhou, K. I. *et al.* Pseudouridines have context-dependent mutation and stop rates in high-throughput sequencing. *RNA Biol.* **15**, 892–900 (2018).
 168. Parisien, M., Yi, C. & Pan, T. Rationalization and prediction of selective decoding of pseudouridine-modified nonsense and sense codons. *RNA* **18**, 355–367 (2012).
 169. Karijovich, J. & Yu, Y.-T. Modifying the genetic code: Converting nonsense codons into sense codons by targeted pseudouridylation. *Nature* **474**, 395–398 (2012).
 170. Israel S. Fernández, Chyan Leong Ng, Ann C. Kelley, Guowei Wu, Yi-Tao Yu, and V. R. Unusual base pairing during the decoding of a stop codon by the ribosome. *Nature* **500**, 107–110 (2013).
 171. Eyler, D. E. *et al.* Pseudouridinylation of mRNA coding sequences alters translation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23068–23074 (2019).
 172. Anderson, B. R. *et al.* Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. *Nucleic Acids Res.* **38**, 5884–5892 (2010).
 173. Bartoli, K., Schaening, C., Carlile, T. & Gilbert, W. Conserved Methyltransferase Spb1 Targets mRNAs for Regulated Modification with 2'-O-Methyl Ribose. *bioRxiv* 271916 (2018) doi:10.1101/271916.
 174. Shubina, M. Y., Musinova, Y. R. & Sheval, E. V. Nucleolar methyltransferase fibrillarin: Evolution of structure and functions. *Biochem.* **81**, 941–950 (2016).
 175. Somme, J. *et al.* Characterization of two homologous 2'-O-methyltransferases showing different specificities for their tRNA substrates. *RNA* **20**, 1257–1271 (2014).
 176. Elliott, B. A. *et al.* Modification of messenger RNA by 2'-O-methylation regulates gene expression in vivo. *Nat. Commun.* **10**, 1–9 (2019).
 177. Adamiak, D. A., Rypniewski, W. R., Milecki, J. & Adamiak, R. W. Crystal structure of 2'-O-Me(CGCGCG)₂, an RNA duplex at 1.30 Å resolution. Hydration pattern of 2'-O-methylated RNA. *Nucleic Acids Res.* **22**, 4599–4607 (1997).
 178. Karikó, K., Buckstein, M., Ni, H. & Weissman, D. Suppression of RNA recognition by Toll-like

References

- receptors: The impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* **23**, 165–175 (2005).
179. Gehrig, S. *et al.* Identification of modifications in microbial, native tRNA that suppress immunostimulatory activity. *J. Exp. Med.* **209**, 225–233 (2012).
180. Jöckel, S. *et al.* The 2'-O-methylation status of a single guanosine controls transfer RNA-mediated toll-like receptor 7 activation or inhibition. *J. Exp. Med.* **209**, 235–241 (2012).
181. Eigenbrod, T. *et al.* *Recognition of Specified RNA Modifications by the Innate Immune System. Methods in Enzymology* vol. 560 (Elsevier Inc., 2015).
182. Jennifer L. Hyde, M. S. D. Innate immune restriction and antagonism of viral RNA lacking 2'-O methylation. *Virology*. **479**, 66–74 (2015).
183. Choi, J. *et al.* 2'-O-methylation in mRNA disrupts tRNA decoding during translation elongation. *Nat. Struct. Mol. Biol.* **25**, 208–216 (2018).
184. Hoernes, T. P. & Erlacher, M. D. Translating the epitranscriptome. *Wiley Interdiscip. Rev. RNA* **8**, 1–18 (2017).
185. Hoernes, T. P. *et al.* Eukaryotic translation elongation is modulated by single natural nucleotide derivatives in the coding sequences of mRNAs. *Genes (Basel)*. **10**, 1–12 (2019).
186. Grosjean, H., Keith, G. & Droogmans, L. Detection and Quantification of Modified Nucleotides in RNA Using Thin-Layer Chromatography. *Methods Mol. Biol.* **265**, 265:357–91 (2004).
187. M Silberklang, A M Gillum, U. L. R. Use of in Vitro 3~p Labeling in the Sequence Analysis of Nonradioactive tRNAs. *Methods Enzym.* **2**, 59:58-109 (1979).
188. Keith, G. Mobilities of modified ribonucleotides on two-dimensional cellulose thin-layer chromatography. *Biochimie* **77**, 142–144 (1995).
189. Madenf, B. E. H. Identification of the Locations of the Methyl Groups in 18 S Ribosomal RNA from *Xenopus laevis* and Man. *J. Mol. Biol.* 681–699 (1986).
190. Székely, M. & Sanger, F. Use of polynucleotide kinase in fingerprinting non-radioactive nucleic acids. *J. Mol. Biol.* **43**, 607–617 (1969).
191. Huang, C. & Yu, Y. T. Synthesis and labeling of RNA in vitro. *Curr. Protoc. Mol. Biol.* 1–19 (2013) doi:10.1002/0471142727.mb0415s102.
192. Modrak-wojcik, A. *et al.* Eukaryotic translation initiation is controlled by cooperativity effects within ternary complexes of 4E-BP1 , eIF4E , and the mRNA 5' cap. *FEBS Lett.* **587**, 3928–3934 (2013).
193. Jemielity, J. *et al.* Novel 'anti-reverse' cap analogs with superior translational properties. *Rna* **9**, 1108–1122 (2003).
194. Wang, P. & Na, J. Synthetic Messenger RNA and Cell Metabolism Modulation. *Methods Protoc.* **969**, 221–233 (2013).
195. Grudzien, E. *et al.* Novel cap analogs for in vitro synthesis of mRNAs with high translational efficiency. *Rna* **10**, 1479–1487 (2004).

References

196. Quabius, E. S. & Krupp, G. Synthetic mRNAs for manipulating cellular phenotypes: An overview. *N. Biotechnol.* **32**, 229–235 (2015).
197. Pasquinelli, A. E., Dahlberg, J. E. & Lund, E. Reverse 5' caps in RNAs made in vitro by phage RNA polymerases. *RNA (New York, N.Y.)* vol. 1 957–967 (1995).
198. Warminski, M., Sikorski, P. J., Kowalska, J. & Jemielity, J. Applications of Phosphate Modification and Labeling to Study (m)RNA Caps. *Top. Curr. Chem.* **375**, 211–239 (2017).
199. Kowalska, J. *et al.* Synthesis, properties, and biological activity of boranophosphate analogs of the mRNA cap: Versatile tools for manipulation of therapeutically relevant cap-dependent processes. *Nucleic Acids Res.* **42**, 10245–10264 (2014).
200. Su, W. *et al.* Translation, stability, and resistance to decapping of mRNAs containing caps substituted in the triphosphate chain with BH₃, Se, and NH. *Rna* **17**, 978–988 (2011).
201. Hornung, V. *et al.* Ligand for RIG-I. *Science (80-.)*. **314**, 994–997 (2006).
202. Jing Peng, Elizabeth L. Murray, and D. R. S. In vivo and in vitro analysis of poly(A) length effects on mRNA translation. *Methods Mol Biol.* **419**, 215–230 (2008).
203. Aimee L. Jalkanen, Stephen J. Coleman, and J. W. Determinants and Implications of mRNA Poly(A) Tail Size - Does this. *Semin Cell Dev Biol.* **0**, 24–32 (2014).
204. Mockey, M. *et al.* mRNA transfection of dendritic cells: Synergistic effect of ARCA mRNA capping with Poly(A) chains in cis and in trans for a high protein expression level. *Biochem. Biophys. Res. Commun.* **340**, 1062–1068 (2006).
205. Hoernes, T. P. *et al.* Atomic mutagenesis of stop codon nucleotides reveals the chemical prerequisites for release factor-mediated peptide release. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E382–E389 (2018).
206. Gustafsson, C., Govindarajan, S. & Minshull, J. Codon bias and heterologous protein expression. **22**, (2004).
207. Malarkannan, S., Horng, T., Shih, P. P., Schwab, S. & Shastri, N. Presentation of Out-of-Frame Peptide / MHC Class I Complexes by a Novel Translation Initiation Mechanism. **10**, 681–690 (1999).
208. Gardin, J. *et al.* Measurement of average decoding rates of the 61 sense codons in vivo. 1–20 (2014) doi:10.7554/eLife.03735.
209. Hussmann, J. A., Patchett, S., Johnson, A. & Sawyer, S. Understanding Biases in Ribosome Profiling Experiments Reveals Signatures of Translation Dynamics in Yeast. 1–25 (2015) doi:10.1371/journal.pgen.1005732.
210. Weinberg, D. E. *et al.* Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.* 2016 **14**, 1787–1799 (2016).
211. Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S. & Reis E Sousa, C. Innate Antiviral Responses by Means of TLR7-Mediated Recognition of Single-Stranded RNA. *Science (80-.)*. **303**, 1529–1531 (2004).

References

212. Alexopoulou, L. Recognition of double-stranded RNA and activation of NF- κ B by Toll-like receptor 3. *Nature* **413**, 732–738 (2001).
213. Heil, F. *et al.* Species-Specific Recognition of Single-Stranded RNA via Toll-like Receptor 7 and 8. **303**, 1526–1530 (2004).
214. Heil, F. *et al.* Species-Specific Recognition of Single-Stranded RNA via Toll-like Receptor 7 and 8. *Science* (80-.). **303**, 1526–1530 (2004).
215. Andreas Pichlmair, Oliver Schulz, Choon Ping Tan, Tanja Naslund, 2 Peter Liljestrom Friedemann Weber, C. R. e S. RIG-I-Mediated Antiviral Responses to Single-Stranded RNA Bearing 5'-Phosphates. *Science* (80-.). **314**, 997–1001 (2006).
216. Yoneyama, M. *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.* **5**, 730–737 (2004).
217. Biology, C., Einstein, A. & Road, E. Double-Stranded Poliovirus RNA Inhibits Initiation of Protein Synthesis by Reticulocyte Lysates. *Proc. Nat. Acad. Sci.* **68**, 1075–1078 (1971).
218. Balachandran, S. *et al.* Essential Role for the dsRNA-Dependent Protein Kinase PKR in Innate Immunity to Viral Infection. **13**, 129–141 (2000).
219. Svitkin, Y. V *et al.* N1-methyl-pseudouridine in mRNA enhances translation through eIF2 α -dependent and independent mechanisms by increasing ribosome density. *Nucleic Acids Res.* **45**, 6023–6036 (2017).
220. Nallagatla, S. R. *et al.* 5'-Triphosphate – Dependent Activation of PKR by RNAs with Short Stem-Loops. *Science* (80-.). **318**, 1455–1458 (2007).
221. Karikó, K. *et al.* Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol. Ther.* **16**, 1833–1840 (2008).
222. Karikó, K., Muramatsu, H., Keller, J. M. & Weissman, D. Increased erythropoiesis in mice injected with submicrogram quantities of pseudouridine-containing mRNA encoding erythropoietin. *Mol. Ther.* **20**, 948–953 (2012).
223. Yoshioka, N. *et al.* Efficient generation of human iPSCs by a synthetic self-replicative RNA. *Cell Stem Cell* **13**, 246–254 (2013).
224. Kormann, M. S. D. *et al.* Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. *Nat. Biotechnol.* **29**, 154–159 (2011).
225. Contreras, R., Cheroute, H., Degraeve, W. & Fiers, W. Simple, efficient in vitro synthesis of capped RNA useful for direct expression of cloned eukaryotic genes. *Nucleic Acids Res.* **10**, 6353–6381 (1982).
226. Licht, K. *et al.* Inosine induces context-dependent recoding and translational stalling. *Nucl. Acid Res.* **47**, 3–14 (2019).
227. Mcgregor, A., Rao, M. V., Duckworth, G., Stockley, P. G. & Connolly, B. A. Preparation of oligoribonucleotides containing 4-thiouridine using Fpmp chemistry . Photo-crosslinking to

References

- RNA binding proteins using 350 nm irradiation. *Nucleic Acids Res* **24**, 3173–3180 (1996).
228. Wyatt, J. R., Erik, J. S. & Steitz, J. A. Site-specific cross-linking of mammalian U5 snRNP to the 5' splice site before the first step of pre-mRNA splicing. *Genes Dev* **6**, 2542–53 (1992).
229. Moore, M. J. & Sharp, P. A. Site-Specific Modification of Pre-mRNA: The 2'-Hydroxyl Groups at the Splice Sites. *Science (80-.)*. **256**, 992–997 (1992).
230. P.C.Huijgens *et al.* Ifosfamide and VP- 162 13 Combination Chemotherapy Combined with Ablative Chemotherapy and Autologous Marrow Transplantation as Salvage Treatment for Malignant Lymphoma. *Eur. J. Cancer Clin. Oncol.* **24**, 483–486 (1988).
231. Bourquin, C. *et al.* Immunostimulatory RNA oligonucleotides trigger an antigen-specific cytotoxic T-cell and IgG2a response. *Immunobiology* **109**, 2953–2960 (2007).
232. Warren, L. *et al.* Highly Efficient Reprogramming to Pluripotency and Directed Differentiation of Human Cells with Synthetic Modified mRNA. 618–630 (2010) doi:10.1016/j.stem.2010.08.012.
233. Parr, C. J. C. *et al.* N1 -Methylpseudouridine substitution enhances the performance of synthetic mRNA switches in cells. *Nucleic Acids Res.* **48**, 1–7 (2020).
234. Andries, O. *et al.* N1 -methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice. *J. Control. Release* **217**, 337–344 (2015).
235. Pardi, N. *et al.* Zika virus protection by a single low-dose nucleoside-modified mRNA vaccination. *Nat. Publ. Gr.* (2017) doi:10.1038/nature21428.
236. Richner, J. M. *et al.* Modified mRNA Vaccines Protect against Zika Virus Infection. *Cell* **168**, 1114–1125.e10 (2017).
237. Pardi, N. *et al.* Administration of nucleoside-modified mRNA encoding broadly neutralizing antibody protects humanized mice from HIV-1 challenge. *Nat. Commun.* **8**, 6–13 (2017).
238. Walsh, E. E. *et al.* RNA-Based COVID-19 Vaccine BNT162b2 Selected for a Pivotal Efficacy Study. *medRxiv Prepr. Serv. Heal. Sci.* (2020) doi:https://doi.org/10.1101/2020.08.17.20176651.
239. Lasda, Erika, Parker, R. Circular RNAs: diversity of form and function ERIKA. *Ref. Libr.* 115–123 (1996) doi:10.1261/rna.047126.114.DIFFERENT.
240. Harland, R. & Misher, L. Stability of RNA in developing *Xenopus* embryos and identification of a destabilizing sequence in TFIIEA messenger RNA. *Development* **102**, 837–852 (1988).
241. Cocquerelle, C., Mascrez, B., Hétiuin, D. & Bailleul, B. Mis-splicing yields circular RNA molecules. *FASEB J.* **7**, 155–160 (1993).
242. Barrett, S. P. & Salzman, J. Circular RNAs: Analysis, expression and potential functions. *Dev.* **143**, 1838–1847 (2016).
243. Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L. & Brown, P. O. Cell-Type Specific Features of Circular RNA Expression. *PLoS Genet.* **9**, (2013).

References

244. Cocquerelle, C., Daubersies, P., Majerus, M. A., Kerckaert, J. P. & Bailleul, B. Splicing with inverted order of exons occurs proximal to large introns. *EMBO J.* **11**, 1095–1098 (1992).
245. Nigro, J. M. *et al.* Scrambled exons. *Cell* **64**, 607–613 (1991).
246. Capel, B. *et al.* Circular transcripts of the testis-determining gene Sry in adult mouse testis. *Cell* **73**, 1019–1030 (1993).
247. Grabowski, P. J., Zaug, A. J. & Cech, T. R. The intervening sequence of the ribosomal RNA precursor is converted to a circular RNA in isolated nuclei of tetrahymena. *Cell* **23**, 467–476 (1981).
248. Talhouarne, G. J. S. & Gall, J. G. Lariat intronic RNAs in the cytoplasm of *Xenopus tropicalis* oocytes. *Rna* **20**, 1476–1487 (2014).
249. Zhang, Y. *et al.* Circular Intronic Long Noncoding RNAs. *Mol. Cell* **51**, 792–806 (2013).
250. Gardner, E. J., Nizami, Z. F., Conover Talbot, J. & Gall, J. G. Stable intronic sequence RNA (sisRNA), a new class of noncoding RNA from the oocyte nucleus of *Xenopus tropicalis*. *Genes Dev.* **26**, 2550–2559 (2012).
251. Kos, A., Dijkemat, R., Meide, P. H. Van Der & Schellekens, H. The hepatitis delta (B) virus possesses a circular RNA A. **379**, 1985–1987 (1986).
252. Sanger, H. L., Klotzt, G., Riesnert, D., Gross, H. J. & Albrecht, K. Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. **73**, 3852–3856 (1976).
253. Danan, M., Schwartz, S., Edelheit, S. & Sorek, R. Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.* **40**, 3131–3142 (2012).
254. Ashwal-Fluss, R. *et al.* CircRNA Biogenesis competes with Pre-mRNA splicing. *Mol. Cell* **56**, 55–66 (2014).
255. Kelly, S., Greenman, C., Cook, P. R. & Papantonis, A. Exon Skipping Is Correlated with Exon Circularization. *J. Mol. Biol.* **427**, 2414–2417 (2015).
256. Starke, S. *et al.* Exon circularization requires canonical splice signals. *Cell Rep.* **10**, 103–111 (2015).
257. Chen, L. L. & Yang, L. Regulation of circRNA biogenesis. *RNA Biol.* **12**, 381–388 (2015).
258. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**, (2012).
259. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
260. Hansen, T. B. *et al.* miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.* **30**, 4414–4422 (2011).
261. Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157 (2013).

References

262. Wang, Y. & Wang, Z. Efficient backsplicing produces translatable circular mRNAs. *RNA* **21**, 172–179 (2015).
263. Zhang, X. O. *et al.* Complementary sequence-mediated exon circularization. *Cell* **159**, 134–147 (2014).
264. Li, Z. *et al.* Exon-intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.* **22**, 256–264 (2015).
265. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388 (2013).
266. Huang, C., Liang, D., Tatomer, D. C. & Wilusz, J. E. A length-dependent evolutionarily conserved pathway controls nuclear export of circular RNAs. *Genes Dev.* **32**, 639–644 (2018).
267. Ebbesen, K. K., Kjems, J. & Hansen, T. B. Circular RNAs: Identification, biogenesis and function. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859**, 163–168 (2016).
268. Puttaraju, M. & Been, M. Group I permuted intron-exon (PIE) sequences self-splice to produce circular exons. *Nucleic Acids Res.* **20**, 5357–5364 (1992).
269. Rybak-Wolf, A. *et al.* Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Mol. Cell* **58**, 870–885 (2014).
270. Xu, K. *et al.* Annotation and functional clustering of circRNA expression in rhesus macaque brain during aging. *Cell Discov.* **4**, 1–18 (2018).
271. Dubin, R. A., Kazmi, M. A. & Ostrer, H. Inverted repeats are necessary for circularization of the mouse testis Sry transcript. *Gene* **167**, 245–248 (1995).
272. Pasmán, Z., Been, M. D. & Garcia-Blanco, M. A. Exon circularization in mammalian nuclear extracts. *RNA* vol. 2 603–610 (1996).
273. Di Timoteo, G. *et al.* Modulation of circRNA Metabolism by m6A Modification. *Cell Rep.* **31**, (2020).
274. Coca-Prados, M. & Hsu, M.-T. Electron microscopic evidence for circular form of RNA in the cytoplasm of eukaryotic cells. *Nature* **280**, 0–1 (1979).
275. Sharpless, W. R. J. and N. E. Detecting and characterizing circular RNAs William. *Nat Biotechnol* **5**, 738–749 (2015).
276. Sarokin, L. & Carlson, M. Discrimination between RNA circles, interlocked RNA circles and lariats using two-dimensional polyacrylamide gel electrophoresis. *Methods* **14**, 6597–6605 (1988).
277. Schumacher, J., Randles, J. W. & Riesner, D. A two-dimensional electrophoretic technique for the detection of circular viroids and virusoids. *Anal. Biochem.* **135**, 288–295 (1983).
278. Feldstein, P. A., Levy, L., Randles, J. W. & Owens, R. A. Synthesis and two-dimensional electrophoretic analysis of mixed populations of circular and linear RNAs. *Nucleic Acids Res.* **25**, 4850–4854 (1997).
279. Preußner, C. *et al.* Selective release of circRNAs in platelet-derived extracellular vesicles. *J.*

References

- Extracell. Vesicles* **7**, (2018).
280. Diegelman, A. M. & Kool, E. T. Generation of circular RNAs and trans-cleaving catalytic RNAs by rolling transcription of circular DNA oligonucleotides encoding hairpin ribozymes. *Nucleic Acids Res.* **26**, 3235–3241 (1998).
 281. Wesselhoeft, R. A., Kowalski, P. S. & Anderson, D. G. Engineering circular RNA for potent and stable translation in eukaryotic cells. *Nat. Commun.* **9**, (2018).
 282. Zeng, X., Lin, W., Guo, M. & Zou, Q. A comprehensive overview and evaluation of circular RNA detection tools. *PLoS Comput. Biol.* **15**, 1–21 (2017).
 283. Vicens, Q. & Westhof, E. Biogenesis of circular RNAs. *Cell* **159**, 13–14 (2014).
 284. Guo, J. U., Agarwal, V., Guo, H. & Bartel, D. P. Expanded identification and characterization of mammalian circular RNAs. *Genome Biol.* **15**, 1–14 (2014).
 285. Pamudurti, N. R. *et al.* Translation of CircRNAs. *Mol. Cell* **66**, 9-21.e7 (2017).
 286. Salzman, J. Circular RNA Expression Its Potential Regulation and Function. *Trends Genet.* **32**, 309–316 (2016).
 287. Burd, C. E. *et al.* Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.* **6**, 1–15 (2010).
 288. Yang, Q. *et al.* Circular RNA expression profiles during the differentiation of mouse neural stem cells. *BMC Syst. Biol.* **12**, (2018).
 289. Wu, A. *et al.* Upregulated hsa_circ_0005785 Facilitates Cell Growth and Metastasis of Hepatocellular Carcinoma Through the miR-578/APRIL Axis. *Front. Oncol.* **10**, 1–13 (2020).
 290. Lei, M., Zheng, G., Ning, Q., Zheng, J. & Dong, D. Translation and functional roles of circular RNAs in human cancer. *Mol. Cancer* **19**, 1–9 (2020).
 291. Lei, B., Tian, Z., Fan, W. & Ni, B. Circular RNA: A novel biomarker and therapeutic target for human cancers. *Int. J. Med. Sci.* **16**, 292–301 (2019).
 292. Yao, R., Zou, H. & Liao, W. Prospect of circular RNA in Hepatocellular carcinoma: A novel potential biomarker and therapeutic target. *Front. Oncol.* **8**, 1–12 (2018).
 293. Kleaveland, B., Shi, C. Y., Stefano, J. & Bartel, D. P. A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. *Cell* **174**, 350-362.e17 (2018).
 294. Dube, U. *et al.* An atlas of cortical circular RNA expression in Alzheimer disease brains demonstrates clinical and pathological associations. *Nat. Neurosci.* **22**, 1903–1912 (2019).
 295. Kastner, N. *et al.* New Insights and Current Approaches in Cardiac Hypertrophy Cell Culture, Tissue Engineering Models, and Novel Pathways Involving Non-Coding RNA. *Front. Pharmacol.* **11**, 1–7 (2020).
 296. Memczak, S., Papavasileiou, P., Peters, O. & Rajewsky, N. Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *PLoS One* **10**, 1–13 (2015).
 297. Jae Hoon Bahn, Qing Zhang, Feng Li, Tak-Ming Chan , Xianzhi Lin, Y. & Kim, David T.W.

References

- Wong, and X. X. The Landscape of MicroRNA, Piwi-Interacting RNA, and Circular RNA in Human Saliva. **61**, 221–230 (2015).
298. Li, Y. *et al.* Circ HIPK 3 sponges miR-558 to suppress heparanase expression in bladder cancer cells. *EMBO Rep.* **18**, 1646–1659 (2017).
299. Han, D. *et al.* Circular RNA circMTO1 acts as the sponge of microRNA-9 to suppress hepatocellular carcinoma progression. *Hepatology* **66**, 1151–1164 (2017).
300. Piwecka, M. *et al.* Loss of a mammalian circular RNA locus causes miRNA deregulation and affects brain function. *Science (80-.).* **357**, (2017).
301. Li, Q. *et al.* CircACC1 Regulates Assembly and Activation of AMPK Complex under Metabolic Stress. *Cell Metab.* **30**, 157-173.e7 (2019).
302. Wilusz, J. E. Circular RNAs: Unexpected outputs of many protein-coding genes. *RNA Biol.* **14**, 1007–1017 (2017).
303. Abe, N. *et al.* Rolling Circle Translation of Circular RNA in Living Human Cells. *Sci. Rep.* **5**, 1–9 (2015).
304. Zhang, M. *et al.* A novel protein encoded by the circular form of the SHPRH gene suppresses glioma tumorigenesis. *Oncogene* **37**, 1805–1814 (2018).
305. Liang, W. C. *et al.* Translation of the circular RNA circ β -catenin promotes liver cancer cell growth through activation of the Wnt pathway. *Genome Biol.* **20**, 1–12 (2019).
306. Yang, Y. *et al.* Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *J. Natl. Cancer Inst.* **110**, 304–315 (2018).
307. Umekage, S. & Kikuchi, Y. In vitro and in vivo production and purification of circular RNA aptamer. *J. Biotechnol.* **139**, 265–272 (2009).
308. Ford, E. & Ares, M. ribozymes derived from a group I intron of phage T4 Synthesis of circular RNA in bacteria and yeast using RNA cyclase. *Proc. Nat. Acad. Sci.* **91**, 3117–3121 (1994).
309. Puttaraju, M. & Been, M. D. Circular ribozymes generated in Escherichia coli using group I self-splicing permuted intron-exon sequences. *J. Biol. Chem.* **271**, 26081–26087 (1996).
310. Ford, E. & Ares, M. Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozymes derived from a group I intron of phage T4. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 3117–3121 (1994).
311. Chen, Y. G. *et al.* Sensing Self and Foreign Circular RNAs by Intron Identity. *Mol. Cell* **67**, 228-238.e5 (2017).
312. Perriman, R. & Ares, M. Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. *RNA* **4**, 1047–1054 (1998).
313. Kramer, M. C. *et al.* Combinatorial control of Drosophila circular RNA expression by intronic repeats, hnRNPs, and SR proteins. *Genes Dev.* **29**, 2168–2182 (2015).
314. Chen, C. Y. & Sarnow, P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science (80-.).* **268**, 415–417 (1995).

References

315. Petkovic, S. & Müller, S. RNA circularization strategies in vivo and in vitro. *Nucleic Acids Res.* **43**, 2454–2465 (2015).
316. KOZAK, M. Inability of circular mRNA to attach to eukaryotic ribosomes IN. *Nature* **699**, 82–85 (1979).
317. Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320–325 (1988).
318. Yang, Y. *et al.* Extensive translation of circular RNAs driven by N⁶-methyladenosine. *Cell Res.* **27**, 626–641 (2017).
319. Legnini, I. *et al.* Circ-ZNF609 is a Circular RNA that Can Be Translated and Functions in Myogenesis. *Mol. Cell* **66**, 22–37 (2017).
320. Chen, C. Y. & Sarnow, P. Initiation of protein synthesis by the eukaryotic translational apparatus on circular RNAs. *Science (80-.)*. **268**, 415–417 (1995).
321. Li, A. *et al.* Cytoplasmic m⁶A reader YTHDF3 promotes mRNA translation. *Cell Res.* **27**, 444–447 (2017).
322. Padilla, R. & Sousa, R. A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Res.* **30**, 1–4 (2002).
323. Kurschat, W. C., Müller, J., Wombacher, R. & Helm, M. Optimizing splinted ligation of highly structured small RNAs. *RNA* **11**, 1909–1914 (2005).
324. Hengesbach, M. *et al.* RNA intramolecular dynamics by single-molecule FRET. *Curr. Protoc. Nucleic Acid Chem.* 1–22 (2008) doi:10.1002/0471142700.nc1112s34.
325. Bochkov, Y. A. & Palmenberg, A. C. Translational efficiency of EMCV IRES in bicistronic vectors is dependent upon IRES sequence and gene location. *Biotechniques* **41**, 283–292 (2006).
326. Borman, A. M., Le Mercier, P., Girard, M. & Kean, K. M. Comparison of Picornaviral IRES-Driven Internal Initiation of Translation in Cultured Cells of Different Origins. *Nucleic Acids Res.* **25**, 925–932 (1996).
327. Kaminski, A., Belsham, G. J. & Jackson, R. J. Translation of encephalomyocarditis virus RNA: Parameters influencing the selection of the internal initiation site. *EMBO J.* **13**, 1673–1681 (1994).
328. Sawant, A. A., Mukherjee, P. P., Jangid, R. K., Galande, S. & Srivatsan, S. G. A clickable UTP analog for the posttranscriptional chemical labeling and imaging of RNA. *Org. Biomol. Chem.* **14**, 5832–5842 (2016).
329. Sawant, A. A. *et al.* A versatile toolbox for posttranscriptional chemical labeling and imaging of RNA. *Nucleic Acids Res.* **44**, e16 (2016).
330. Hellmuth, I. *et al.* Bioconjugation of small molecules to RNA impedes its recognition by toll-like receptor 7. *Front. Immunol.* **8**, 1–13 (2017).
331. Milisavljevič, N., Perlíková, P., Pohl, R. & Hocek, M. Enzymatic synthesis of base-modified RNA by T7 RNA polymerase. A systematic study and comparison of 5-substituted pyrimidine

References

- and 7-substituted 7-deazapurine nucleoside triphosphates as substrates. *Org. Biomol. Chem.* **16**, 5800–5807 (2018).
332. Barbosa, E. & Moss, B. mRNA(nucleoside-2'-)-methyltransferase from Vaccinia Virus. *J. Biol. Chem.* **253**, 7692–7697 (1978).
 333. Chen, D., Luongo, C. L., Nibert, M. L. & Patton, J. T. Rotavirus open cores catalyze 5'-capping and methylation of exogenous RNA: Evidence that VP3 is a methyltransferase. *Virology* **265**, 120–130 (1999).
 334. Cao, G. J. & Sarkar, N. Identification of the gene for an Escherichia coli poly(A) polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10380–10384 (1992).
 335. Jerabek-Willemsen, M. *et al.* MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* **1077**, 101–113 (2014).
 336. Jacob, D. *et al.* Absolute Quantification of Noncoding RNA by Microscale Thermophoresis. *Angew. Chemie - Int. Ed.* **58**, 9565–9569 (2019).
 337. Pitsch, S. & Weiss, P. A. Chemical Synthesis of RNA Sequences with 2'- O - [(Triisopropylsilyl)oxy]methyl-protected Ribonucleoside Phosphoramidites. *Curr. Protoc. Nucleic Acid Chem.* **7**, 1–15 (2001).
 338. Sousa, R. and Mukherjee, S. T7 RNA polymerase. *Prog Nucleic Acid Res Mol Biol* **73**, (2013).
 339. Stark, M. R., Pleiss, J. A., Deras, M., Scaringe, S. A. & Rader, S. D. An RNA ligase-mediated method for the efficient creation of large, synthetic RNAs. *RNA* **12**, 2014–2019 (2006).
 340. Gholamalipour, Y., Karunanayake Mudiyanse, A. & Martin, C. T. NAR breakthrough article 3' end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character—RNA-Seq analyses. *Nucleic Acids Res.* **46**, 9253–9263 (2018).
 341. Motorin, Y. & Marchand, V. Detection and analysis of RNA ribose 2'-o-methylations: Challenges and solutions. *Genes (Basel)*. **9**, 1–14 (2018).
 342. Karikó, K., Muramatsu, H., Ludwig, J. & Weissman, D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Res.* **39**, 1–10 (2011).
 343. Baronti, L., Karlsson, H., Marušič, M. & Petzold, K. A guide to large-scale RNA sample preparation. *Anal. Bioanal. Chem.* **410**, 3239–3252 (2018).
 344. Summer, H., Grämer, R. & Dröge, P. Denaturing urea polyacrylamide gel electrophoresis (Urea PAGE). *J. Vis. Exp.* 3–5 (2009) doi:10.3791/1485.
 345. Frémont, P., Dionne, F. T. & Rogers, P. A. Recovery of biologically functional messenger RNA from agarose gels by passive elution. *Anal. Biochem.* **156**, 508–514 (1986).
 346. Moore, D., Dowhan, D., Chory, J. & Ribaudo, R. K. Isolation and Purification of Large DNA Restriction Fragments from Agarose Gels. *Curr. Protoc. Mol. Biol.* **59**, (2002).
 347. Castagnetti, S., Hentze, M. W., Ephrussi, A. & Gebauer, F. Control of oskar mRNA translation by Bruno in a novel cell-free system from Drosophila ovaries. *Development* **127**, 1063–1068

References

- (2000).
348. Michel, Y. M., Poncet, D., Piron, M., Kean, K. M. & Borman, A. M. 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**, 32268–32276 (2000).
 349. Svitkin, Y. V. *et al.* Eukaryotic Translation Initiation Factor 4E Availability Controls the Switch between Cap-Dependent and Internal Ribosomal Entry Site-Mediated Translation. *Mol. Cell. Biol.* **25**, 10556–10565 (2005).
 350. Hillebrecht, J. R. & Chong, S. A comparative study of protein synthesis in in vitro systems: From the prokaryotic reconstituted to the eukaryotic extract-based. *BMC Biotechnol.* **8**, 1–9 (2008).
 351. Yang, T. T., Cheng, L. & Kain, S. R. Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* **24**, 4592–4593 (1996).
 352. Bokman, S. H. & Ward, W. W. Renaturation of Aequorea green-fluorescent protein. *Biochem. Biophys. Res. Commun.* **101**, 1372–1380 (1981).
 353. Muratovska, A. & Eccles, M. R. Conjugate for efficient delivery of short interfering RNA (siRNA) into mammalian cells. *FEBS Lett.* **558**, 63–68 (2004).
 354. Piyaraj, P. The Green Fluorescent Protein (GFP). *Bangkok Med. J.* **05**, 101–102 (2013).
 355. Olliver, L., Grobler-Rabie, A. & Boyd, C. D. In vitro translation of messenger RNA in a wheat germ extract cell-free system. *Methods Mol. Biol.* **86**, 229–233 (1998).
 356. Martin Chalfie, Yuan Tu, Ghia Euskirchen, William W. Ward, D. C. P. Green fluorescent protein as a marker gene expression. *Science (80-.).* **52**, 1766–1767 (1994).
 357. Beckler, G. S., Thompson, D. & Van Oosbree, T. In vitro translation using rabbit reticulocyte lysate. *Methods Mol. Biol.* **37**, 215–232 (1995).
 358. Jackson, R. J. & Hunt, T. Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA. *Methods Enzymol.* **96**, 50–74 (1983).
 359. PELHAM, H. R. B. & JACKSON, R. J. An Efficient mRNA-Dependent Translation System from Reticulocyte Lysates. *Eur. J. Biochem.* **67**, 247–256 (1976).
 360. Borman, A. M., Bailly, J. luc, Girard, M. & Kean, K. M. Picornavirus internal ribosome entry segments: Comparison of translation efficiency and the requirements for optimal internal initiation of translation in vitro. *Nucleic Acids Res.* **23**, 3656–3663 (1995).
 361. Anastasina, M., Terenin, I., Butcher, S. J. & Kainov, D. E. A technique to increase protein yield in a rabbit reticulocyte lysate translation system. *Biotechniques* **56**, 36–39 (2014).
 362. Inouye, S. & Tsuji, F. I. Aequorea green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. *FEBS Lett.* **341**, 277–280 (1994).
 363. Geertsma, E. R., Groeneveld, M., Slotboom, D. J. & Poolman, B. Quality control of overexpressed membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5722–5727 (2008).
 364. Aokp, T. protein and its application to Western blotting. *FEBS Lett.* **384**, 193–197 (1996).

References

365. Nachtergaele, S. & He, C. Chemical Modifications in the Life of an mRNA Transcript. *Annu. Rev. Genet.* **52**, 349–372 (2018).
366. Helm, M. & Motorin, Y. Detecting RNA modifications in the epitranscriptome: Predict and validate. *Nat. Rev. Genet.* **18**, 275–291 (2017).
367. Picard-Jean, F. *et al.* 2'-O-methylation of the mRNA cap protects RNAs from decapping and degradation by DXO. *PLoS One* **13**, 9–10 (2018).
368. Lu, F. *et al.* Efficient Generation of Transgenic Buffalos (*Bubalus bubalis*) by Nuclear Transfer of Fetal Fibroblasts Expressing Enhanced Green Fluorescent Protein. *Sci. Rep.* **8**, 1–10 (2018).
369. Durbin, A. F., Wang, C., Marcotrigiano, J. & Gehrke, L. RNAs containing modified nucleotides fail to trigger RIG-I conformational changes for innate immune signaling. *MBio* **7**, 1–11 (2016).
370. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
371. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger rNAS. *Nucleic Acids Res.* **15**, 8125–8148 (1987).
372. Anderson, B. R. *et al.* Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. *Nucleic Acids Res.* **38**, 5884–5892 (2010).
373. Kauffman, K. J. *et al.* Efficacy and immunogenicity of unmodified and pseudouridine-modified mRNA delivered systemically with lipid nanoparticles in vivo. *Biomaterials* **109**, 78–87 (2016).
374. Li, B., Luo, X. & Dong, Y. Effects of Chemically Modified Messenger RNA on Protein Expression. *Bioconjug. Chem.* **27**, 849–853 (2016).
375. Lomakin, I. B., Hellen, C. U. T. & Pestova, T. V. Physical Association of Eukaryotic Initiation Factor 4G (eIF4G) with eIF4A Strongly Enhances Binding of eIF4G to the Internal Ribosomal Entry Site of Encephalomyocarditis Virus and Is Required for Internal Initiation of Translation. *Mol. Cell. Biol.* **20**, 6019–6029 (2000).
376. Duke, G. M., Hoffman, M. A. & Palmenberg, A. C. Sequence and structural elements that contribute to efficient encephalomyocarditis virus RNA translation. *J. Virol.* **66**, 1602–1609 (1992).
377. Engel, J. D. & Von Hippel, P. H. Effects of methylation on the stability of nucleic acid conformations. *J. Biol. Chem.* **253**, 927–934 (1978).
378. Roost, C. *et al.* Structure and thermodynamics of N6-methyladenosine in RNA: A spring-loaded base modification. *J. Am. Chem. Soc.* **137**, 2107–2115 (2015).
379. Iizuka, N., Najita, L., Franzusoff, A. & Sarnow, P. Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **14**, 7322–7330 (1994).
380. De Gregorio, E., Preiss, T. & Hentze, M. W. Translational activation of uncapped mRNAs by the central part of human eIF4G is 5' end-dependent. *RNA* **4**, 828–836 (1998).
381. Mao, Y. *et al.* m6A in mRNA coding regions promotes translation via the RNA helicase-

References

- containing YTHDC2. *Nat. Commun.* **10**, 1–11 (2019).
382. Nandakumar, J., Ho, C. K., Lima, C. D. & Shuman, S. RNA substrate specificity and structure-guided mutational analysis of bacteriophage T4 RNA ligase 2. *J. Biol. Chem.* **279**, 31337–31347 (2004).
383. Ho, C. K. & Shuman, S. Bacteriophage T4 RNA ligase 2 (gp24.1) exemplifies a family of RNA ligases found in all phylogenetic domains. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12709–12714 (2002).
384. Nandakumar, J., Shuman, S. & Lima, C. D. RNA Ligase Structures Reveal the Basis for RNA Specificity and Conformational Changes that Drive Ligation Forward. *Cell* **127**, 71–84 (2006).
385. Petkovic, S. & Müller, S. RNA self-processing: Formation of cyclic species and concatemers from a small engineered RNA. *FEBS Lett.* **587**, 2435–2440 (2013).
386. Harley, C. B. *et al.* Discrimination between RNA circles, interlocked RNA circles and lariats using two-dimefile:///C:/Users/kawt0/OneDrive/Documents/2020/2020 Phd/thesis/REVIEWS/circular RNA/Production and Purification of Artificial Circular RNA Sponges for Application in Mol. **16**, 7269–7285 (1988).
387. Breuer, J. & Rossbach, O. Production and Purification of Artificial Circular RNA Sponges for Application in Molecular Biology and Medicine. *Methods Protoc.* **3**, 42 (2020).
388. Wesselhoeft, R. A. *et al.* RNA Circularization Diminishes Immunogenicity and Can Extend Translation Duration In Vivo. *Mol. Cell* **74**, 508-520.e4 (2019).
389. Siaw, G. E. L., Liuc, I. F., Lin, P. Y., Been, M. D. & Hsieh, T. S. DNA and RNA topoisomerase activities of Top3 β are promoted by mediator protein Tudor domain-containing protein 3. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5544–E5551 (2016).
390. Domingues, S. *et al.* The role of RNase R in trans-translation and ribosomal quality control. *Biochimie* **114**, 113–118 (2015).
391. Venkataraman, K., Guja, K. E., Garcia-Diaz, M. & Karzai, A. W. Non-stop mRNA decay: A special attribute of trans-translation mediated ribosome rescue. *Front. Microbiol.* **5**, 1–12 (2014).
392. Cheng, Z. F. & Deutscher, M. P. An important role for RNase R in mRNA decay. *Mol. Cell* **17**, 313–318 (2005).
393. Suzuki, H. *et al.* Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Res.* **34**, (2006).
394. Li, Z., Reimers, S., Pandit, S. & Deutscher, M. P. RNA quality control: Degradation of defective transfer RNA. *EMBO J.* **21**, 1132–1138 (2002).
395. Matos, R. G., Barbas, A., Gómez-Puertas, P. & Arraiano, C. M. Swapping the domains of exoribonucleases RNase II and RNase R: Conferring upon RNase II the ability to degrade ds RNA. *Proteins Struct. Funct. Bioinforma.* **79**, 1853–1867 (2011).
396. Chao, C. W., Chan, D. C., Kuo, A. & Leder, P. The mouse formin (Fmn) gene: Abundant circular RNA transcripts and gene-targeted deletion analysis. *Mol. Med.* **4**, 614–628 (1998).

References

397. Zhou, C. *et al.* Genome-Wide Maps of m6A circRNAs Identify Widespread and Cell-Type-Specific Methylation Patterns that Are Distinct from mRNAs. *Cell Rep.* **20**, 2262–2276 (2017).
398. Mo, D. *et al.* A universal approach to investigate circRNA protein coding function. *Sci. Rep.* **9**, 1–13 (2019).
399. Greco, S., Cardinali, B., Falcone, G. & Martelli, F. Circular rnas in muscle function and disease. *Int. J. Mol. Sci.* **19**, 1–18 (2018).
400. Rifo, R. S., Ricci, E. P., Décimo, D., Moncorgé, O. & Ohlmann, T. Back to basics: The untreated rabbit reticulocyte lysate as a competitive system to recapitulate cap/poly(A) synergy and the selective advantage of IRES-driven translation. *Nucleic Acids Res.* **35**, (2007).
401. Zhao, J. *et al.* Transforming activity of an oncoprotein-encoding circular RNA from human papillomavirus. *Nat. Commun.* **10**, (2019).
402. Abe, N. *et al.* Rolling circle amplification in a prokaryotic translation system using small circular RNA. *Angew. Chemie - Int. Ed.* **52**, 7004–7008 (2013).
403. Zhao, W., Ali, M. M., Brook, M. A. & Li, Y. Rolling Circle Amplification: Applications in Nanotechnology and Biodetection with Functional Nucleic Acids. *Angew. Chemie* **120**, 6428–6436 (2008).
404. Pause, A. *et al.* Insulin-dependent stimulation of protein synthesis by phosphorylation of a regulator of 5' -cap function. *Nature* **371**, (1994).
405. Sonenberg, N. & Hinnebusch, A. G. Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell* **136**, 731–745 (2009).
406. Kozak, M. The scanning model for translation: An update. *J. Cell Biol.* **108**, 229–241 (1989).
407. (Ed.), H. G. Fine-Tuning of RNA Functions by Modification and Editing. *Springer-Verlag* **12**, (2005).
408. Chengqi Yi and Tao Pan. Cellular Dynamics of RNA Modification. *ACC Chem* **44**, 1380–1388 (2011).
409. Czerwonec, A. *et al.* MODOMICS: A database of RNA modification pathways. 2008 update. *Nucleic Acids Res.* **37**, 118–121 (2009).
410. Schwartz, S. *et al.* High-Resolution mapping reveals a conserved, widespread, dynamic mRNA methylation program in yeast meiosis. *Cell* **155**, 1409–1421 (2013).
411. Martin, C. T. & Coleman, J. E. T7 RNA Polymerase Does Not Interact with the 5'-Phosphate of the Initiating Nucleotide. *Biochemistry* **28**, 2760–2762 (1989).
412. Gibson, J. F., Kelso, S. & Skevington, J. H. Band-cutting no more: A method for the isolation and purification of target PCR bands from multiplex PCR products using new technology. *Mol. Phylogenet. Evol.* **56**, 1126–1128 (2010).
413. Kellner, S. *et al.* Absolute and relative quantification of RNA modifications via biosynthetic isotopomers. *Nucleic Acids Res.* **42**, 1–10 (2014).
414. Martin, F. *et al.* Cap-Assisted Internal Initiation of Translation of Histone H4. *Mol. Cell* **41**, 197–

References

- 209 (2011).
415. Galloway, A. & Cowling, V. H. mRNA cap regulation in mammalian cell function and fate. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1862**, 270–279 (2019).
416. Hinnebusch, A. G. & Lorsch, J. R. The mechanism of eukaryotic translation initiation: New insights and challenges. *Cold Spring Harb. Perspect. Biol.* **4**, 1–26 (2012).
417. Aufiero, S., Reckman, Y. J., Pinto, Y. M. & Creemers, E. E. Circular RNAs open a new chapter in cardiovascular biology. *Nat. Rev. Cardiol.* **16**, 503–514 (2019).
418. Costello, A., Lao, N. T., Barron, N. & Clynes, M. Reinventing the Wheel: Synthetic Circular RNAs for Mammalian Cell Engineering. *Trends Biotechnol.* **38**, 217–230 (2020).

Publications and Collaborations

- (1) **Slama, K.**, Galliot, A., Weichmann, F., Hertler, J., Feederle, R., Meister, G., & Helm, M. Determination of enrichment factors for modified RNA in MeRIP experiments. *Methods* 156, 102–109 (2019).
- (2) Pannwitt, S., **Slama, K.**, Depoix, F., Helm, M. & Schneider, D. Against Expectations: Unassisted RNA Adsorption onto Negatively Charged Lipid Bilayers. *Langmuir* 35, 14704–14711 (2019).
- (3) Weichmann, F., Hett, R., Schepers, A., Ito-Kureha, T., Flatley, A., **Slama, K.**, Hastert, F. D., Angstman, N. B., Cristina Cardoso, M., König, J., Hüttelmaier, S., Dieterich, C., Canzar, S., Helm, M., Heissmeyer, V., Feederle, R., & Meister, G. Validation strategies for antibodies targeting modified ribonucleotides. *RNA* 26, 1489–1506 (2020).
- (4) Hellmuth, I., Freund, I., Schlöder, J., Seidu-Larry, S., Thüring K., **Slama, K.**, Langhanki J., Kaloyanova, S., Eigenbrod, T., Krumb, M., Röhm, S., Peneva, K., Opatz, T., Jonuleit, H., H Dalpke A., Helm, M. Bioconjugation of small molecules to RNA impedes its recognition by toll-like receptor 7. *Front. Immunol.* 8, 1–13 (2017).

In process:

Slama, K., Hertler J., Marchand V., Motorin Y., & Helm M. Synthesis of point-modified mRNA

Oral Presentations

Slama K., "In vitro circular mRNA synthesis", PhD Student Meeting (SPP1784), TU Dortmund, Dortmund, Germany, 04 - 05 June 2018

Slama K., "In vitro RNA construction by splint ligation", Exploring new frontiers of tRNA biology, Institute of Ph, 26-2rmaceutical and Biomedical Sciences, Mainz, 26 - 27 April 2017

Poster Presentations

- (1) The 10th Meeting of the GBM Study Section RNA-Biochemistry & Synthetic RNA Biology Workshop 4 -7 October 2018 in Bonn (Germany)
- (2) Conference for Nucleic Acids, Immunity and Genome Defence, 4-6 September 2018 in Brno (Czech Republic)
- (3) The 1st Symposium on Nucleic Acid Modifications, , 04 - 06 September 2017 in Mainz, (Germany)

Curriculum Vitae

Kaouthar Slama

Nationality: Tunisian
 Birthday: 12.01.1991
 Birthplace: Paris (France)

Education:

Since 01 /2016 to date	Ph.D. student at Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg-University, Mainz (Germany)
02/2015-07/2015	Master 2 internship, Topic: “Conjugation of siRNA with various ligands and study of the impact of conjugation on RNAi pathway” Johannes Gutenberg-University, Mainz (Germany)
09/2014 – 09/2015	International Master Biosciences and health engineering-RNAES, Faculty of Sciences and Technologies, University of Lorraine, Nancy (France)
01-02/2014	Master internship, Topic:” Development of a protocol for immunoprecipitation of protein “chondrocalcin” in order to study post-translational modifications in the case of osteoarthritis”, Proteomics facility in the Biopole Nancy (France)
09/2013 – 06/2014	Master (first half) of BioSciences and Health Engineering- (Biochemistry, Molecular Biology and Cellular Regulation), Faculty of Sciences and Technologies, University of Lorraine, Nancy (France)
02-04/2013	Bachelor internship, Topic: “Screening for producing alpha_galactosidase bacteria strain”, Biotechnology Center Sfax, (Tunisia)
09/2011-06/2012	Bachelor in Analytical and Experimental biology, University of Gabes (Tunisia)
09/2008 –06/2009	High school degree in Natural sciences. High school Houmt-Souk-Djerba (Tunisia)

 Place, Date

 Kaouthar Slama