U.PORTO
ICBAS | INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
SCHOOL OF MEDICINE AND BIOMEDICAL SCIENCES

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

2.º CICLO

FCUP
ICBAS
2022

U.PORTO

Regulation of gene expression by noncoding sequences in health and disease

Inês Reis Claro

FC

U.PORTO
ICBAS | INSTITUTO DE CIÊNCIAS BIOMÉDICAS ABEL SALAZAR
SCHOOL OF MEDICINE AND BIOMEDICAL SCIENCES

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Regulation of gene expression by noncoding sequences in health and disease

Inês Reis Claro

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto e ao Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto

2022

MSc

Regulation of gene expression by noncoding sequences in health and disease

Inês Reis Claro
Masters in Biochemistry
Chemistry and Biochemistry Department
2022

Supervisor
Isabel Pereira-Castro, PhD, Junior Investigator, i3S, UP

Co-supervisor
Alexandra Moreira, PhD, Principal Investigator, ICBAS, i3S, UP
Lúcio Lara Santos, MD, PhD, Professor of Surgical Oncology, ICBAS, UP


INSTITUTO DE INVESTIGAÇÃO
E INOVAÇÃO EM SAÚDE
UNIVERSIDADE DO PORTO


U.PORTO
ICBAS
INSTITUTO DE CIÊNCIAS
BIOMÉDICAS ABEL SALAZAR
SCHOOL OF MEDICINE AND
BIOMEDICAL SCIENCES


U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Somewhere,
something incredible is waiting to be known.

Carl Sagan

Sworn Statement

I, Inês Reis Claro, enrolled in the Master Degree of Biochemistry at the Faculty of Sciences of the University of Porto hereby declare, in accordance with the provisions of paragraph a) of Article 14 of the Code of Ethical Conduct of the University of Porto, that the content of this dissertation reflects perspectives, research work and my own interpretations at the time of its submission.

By submitting this dissertation, I also declare that it contains the results of my own research work and contributions that have not been previously submitted to this or any other institution.

I further declare that all references to other authors fully comply with the rules of attribution and are referenced in the text by citation and identified in the bibliographic references section. This dissertation does not include any content whose reproduction is protected by copyright laws.

I am aware that the practice of plagiarism and self-plagiarism constitute a form of academic offense.

Inês Reis Claro

30/07/2022

Acknowledgements

First, I must give my biggest “thank you” to my supervisor, Doctor Isabel Pereira-Castro, and to my co-supervisor, Doctor Alexandra Moreira, for giving me the incredible opportunity of conducting my master thesis at the Gene Regulation Group. Isabel, this year would not have been the same without all your incentive, support, and constant help. Alexandra, your genuine passion for science made everything look easier, especially when the results were not the expected ones. Thank you so much for everything, I know I have grown enormously as a scientist due to you!

Secondly, I must thank all the remaining and past members of the Gene Regulation Group, who also accompanied and supported me during this year, giving me the opportunity to learn and to help. To Ana, Alexandre and Filipa, but mostly to Joana! All our lab bench talks, and lunch breaks made all the difference.

I also have to show my appreciation to my fellow biochemistry friends, André and Telmo, and also to Mafalda, who helped me throughout my journey, since I first started.

Last, but never the least, I am deeply thankful to my parents, who always supported my choices, helped when they were not easy, and most importantly, have unconditionally been by my side. Thank you so much!

Resumo

Num mRNA maduro, diversas sequências reguladoras estão presentes nas regiões não traduzidas (UTRs), nomeadamente no 3'UTR. Esta regulação pode ser conseguida através de sequências *trans*-reguladoras, como os microRNA (miRNAs) ou proteínas que se ligam ao RNA (RBPs), que atuam em coordenação com sequências *cis*-reguladoras, como por exemplo os *upstream sequence elements* (USEs), importantes no processamento do 3'. O nosso grupo descobriu um USE com funções fisiológicas importantes em *Drosophila melanogaster*, com um papel na modulação da poliadenilação alternativa e afetando a produção dos níveis necessários da proteína Polo, essencial para a viabilidade das moscas da fruta. Adicionalmente, foi demonstrado que esta sequência de 28 nucleótidos, rica em uracilos, que foi depois denominada iPLUS (*increase Protein Levels Universal Sequence*), promove um aumento na produção de proteínas recombinantes, quando subclonada a jusante de um gene-reporter. Isto foi comprovado em vários modelos para além de *Drosophila*, nomeadamente em peixe-zebra, levedura e linhas celulares de mamíferos.

Dadas as características reguladoras do iPLUS, foi proposto que esta sequência podia ser aplicada à tecnologia das vacinas de mRNA, de forma a aumentar a produção proteica do antigénio e a respetiva estabilidade do mRNA, ambos aspetos importantes em vacinas de mRNA. O nosso objetivo era avaliar se o iPLUS e uma das suas variantes (a região mais conservada do iPLUS repetida três vezes – iPLUS seq3 3X) podiam melhorar as vacinas de mRNA, pondo a hipótese de que a subclonagem destas sequências a jusante do cDNA do respetivo antigénio, poderia levar a um aumento da produção proteica e na estabilidade do mRNA. Com esse propósito, subclonamos o iPLUS e o iPLUS seq3 3X a jusante do cDNA da proteína *Spike* do vírus SARS-CoV-2. Planeamos avaliar o efeito destas sequências em duas variantes do SARS-CoV-2, a variante original de Wuhan, e a variante já dominante, Ómicron.

Ensaios de transfeção de linhas celulares, especificamente HeLa e ExpiCHO-S, revelaram que o iPLUS seq3 3X, quando subclonado a jusante do cDNA da *Spike* de Wuhan, promove um aumento na sua produção proteica e nos níveis de mRNA, quando comparado com o controlo que não possui esta sequência. Este aumento chega mesmo às 2.5-vezes, para a produção proteica da Spike em ExpiCHO-S. A sequência iPLUS não altera ou leva mesmo a uma redução nos níveis proteicos e de mRNA. Para a variante Ómicron, também não foi observado nenhum aumento com ambas as sequências iPLUS testadas, contrastando com os resultados obtidos para a

variante de Wuhan. Posteriormente, realizamos também ensaios *in vitro*, que consistiram na produção de mRNA contendo o cDNA da *Spike* de Wuhan, ou na presença do iPLUS seq3 3X, ou sem esta sequência. A transfeção de células com estas moléculas de mRNA revelou que a presença do IPLUS seq3 3X leva a uma diminuição da estabilidade do mRNA, o que foi reforçado por dados obtidos por ensaios com Actinomicina D. Estes resultados são algo contraditórios, uma vez que o iPLUS seq3 3X leva, por um lado a um aumento na produção da proteína *Spike* e nos seus níveis de mRNA, e por outro lado a uma diminuição da estabilidade desse mesmo mRNA, o que sugere que esta sequência pode estar a aumentar a transcrição ou a eficiência da tradução, e não a estabilidade do mRNA.

Em suma, ao longo desta tese demonstramos que estas sequências curtas, não codificantes, presentes no 3'UTR, são capazes de afetar consideravelmente diferentes aspetos relacionados com a regulação dos mRNAs, desde a transcrição à tradução. Adicionalmente, apresentamos dados que apontam para o facto de a região codificante poder afetar a respetiva transcrição e tradução, e constatamos que o iPLUS e o iPLUS seq3 3X têm funções diferentes, nalguns casos mesmo contrários. Assim, muitas questões levantadas permanecem por responder, abrindo novos caminhos para investigação, ainda que tenhamos provado claramente que o iPLUS seq3 3X pode ser um ótimo recurso para a produção proteica e de mRNA da *Spike* de Wuhan.

Palavras-chave: RNA mensageiro, upstream sequence elements (USE), 3'UTR, *severe acute respiratory syndrome coronavirus-2* (SARS-CoV-2), iPLUS, produção proteica, estabilidade de mRNA

Abstract

In a mature mRNA several expression regulatory sequences are present in the untranslated regions (UTRs), namely in the 3'UTR. Regulation can be achieved by *trans*-acting sequences, such as microRNAs (miRNAs) and RNA-binding proteins (RBPs), which act in coordination with *cis*-regulatory sequences, such as the upstream sequence elements (USEs), crucial for the pre-mRNA 3'end processing. Our group has discovered an USE with essential physiological roles in *Drosophila melanogaster,* modulating alternative polyadenylation and the production of the required levels of the Polo protein, crucial for the fly viability. It was further demonstrated that this non-coding 28-nucleotide long, U-rich sequence that was later named iPLUS (increase Protein Levels Universal Sequence), promotes an increase in recombinant protein production, when subcloned downstream of a reporter gene. This effect was observed in several models besides *Drosophila*, namely zebrafish, yeast and mammalian cell lines.

Given the regulatory characteristics of iPLUS, it was hypothesized that this sequence could be applied to mRNA vaccine's technology, in order to improve the antigen's protein production and the respective mRNA stability, both important aspects of an mRNA molecule. We aimed to evaluate if iPLUS and one of its variants (iPLUS most conserved region in tandem three times – iPLUS seq3 3X) could improve mRNA vaccines, hypothesizing that the subcloning of these sequences downstream of the antigen's cDNA could lead to an increase in protein production and mRNA stability. With that purpose, we subcloned iPLUS and iPLUS seq3 3X downstream of the Spike's protein cDNA from the SARS-CoV-2 virus. We intended to evaluate the effect of these sequences on two SARS-CoV-2 variants, the original Wuhan variant, and the now dominant Omicron.

Cell transfection assays, consisting in the transfection of both HeLa and ExpiCHO-S cell lines with plasmids with or without the iPLUS sequences, revealed that iPLUS seq3 3X, when subcloned downstream of Spike's cDNA from the Wuhan variant, leads to an increase in both Spike protein and mRNA levels, when compared to the control that lacks this sequence. This increase reaches even 2.5-fold for Spike's protein production in ExpiCHO-S cells, which are commonly used in the Pharma industry. The iPLUS sequence either does not impact or leads to a decrease in both protein and mRNA. For the Omicron variant, no increase was observed with both iPLUS sequences tested, contrasting with the results obtained with the Wuhan variant. Afterwards, i*n vitro* assays were carried out to evaluate the role of iPLUS seq3 3X in mRNA stability. Cell transfections with *in vitro* transcribed mRNAs containing Wuhan's Spike cDNA either in

the presence of iPLUS seq3 3X or lacking this sequence demonstrated that iPLUS seq3 3X led to a decrease in mRNA stability, which was also reinforced by data obtained with Actinomycin D assays. These results seem contradictory, with iPLUS seq3 3X leading both to an increase in Spike's protein production and mRNA levels, but to a decrease in the respective mRNA stability, which may suggest that this sequence may be increasing transcription or translation efficiency, rather than mRNA stability.

In short, throughout this thesis we showed that these short, non-coding sequences present in the 3'UTR can strongly impact several aspects of an mRNA fate and regulation, from transcription to translation. Additionally, we observed that iPLUS and iPLUS seq3 3X have different functions, sometimes even contrary. As so, many questions were raised and remain unanswered opening new avenues for research, even though we clearly prove that iPLUS seq3 3X can be an asset for Wuhan Spike's protein and mRNA production.

Keywords: messenger RNA, upstream sequence elements (USE), 3'UTR, severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2), iPLUS, protein production, mRNA stability.

Table of contents

Sworn Statement 2
Acknowledgements 3
Abstract 6
Table of contents 8
List of Figures 10
List of Abbreviations 12
Introduction 15
From transcription to translation: the importance of the mRNA 16
3'end Processing: Polyadenylation 19
The impact of the 3'UTR in gene expression regulation 20
Role of *cis*-regulatory sequences on gene expression regulation 23
Increased Protein Levels Universal Sequence (iPLUS) – an USE with practical applications 27
mRNA and Health 29
Virus and pandemics: the COVID-19 outbreak 29
iPLUS as an asset in COVID-19 mRNA vaccines - iCoV2PLUS 31
Aims 35
Materials and Methods 38
1. Plasmid construction 39
1.1. DNA preparation 39
1.1.1. Polymerase Chain Reaction (PCR) 39
1.1.2. Agarose Gel Electrophoresis 39
1.1.3. Restriction enzyme digestion and DNA modification 39
1.1.4. Sephadex DNA purification 40
1.2. Subcloning 41
1.2.1. Ligation reactions 41
1.2.2. Transformation of competent bacteria 41
1.2.3. Colony PCR 41
1.2.4. Sanger Sequencing 42
1.3. SpikeHu/Om_iPLUSes plasmid construction 42
1.4. Wuhan/Omicron_iPLUSes and Wuhan/Omicron plasmid construction 43
1.5. SpikeHu_polyA and SpikeHu_seq3 3X_polyA plasmid construction 44
2. Cell culture and transfection 44

2.1. HeLa cells culture 44
2.2. ExpiCHO-S cells culture 44
2.3. HeLa cells plasmid transfection 45
2.4. Hela cells mRNA transfections 45
2.5. ExpiCHO-S cells transfection 46
3. Protein extraction and Western Blot 47
3.1. Whole-cell protein extraction 47
3.2. Protein quantification 47
3.3. Western Blot 48
4. RNA extraction, cDNA synthesis and RT-qPCR 48
4.1. RNA extraction 48
4.2. DNase I treatment and cDNA synthesis 49
4.3. Quantitative reverse transcription PCR (RT-qPCR) 50
5. Actinomycin D treatment 51
6. *In vitro* transcription 51
6.1. Phenol/chloroform purification 51
6.2. T7 *in vitro* transcription 52
7. Statistical analyses 53
Results and Discussion 55
1. Cell transfections studies 56
1.1. iPLUS seq3 3X leads to an increase in Spike's production in HeLa cells 56
1.2. In ExpiCHO-S cells, iPLUS seq3 3X also increases Spike production 63
1.3. Efficiency of transfection remains constant, in both HeLa and ExpiCHO-S cells, for the Wuhan variant 67
2. mRNA stability assays 71
2.1. iPLUS seq3 3X decreases the mRNA stability of an *in vitro* transcribed *Spike* mRNA 71
2.2. Actinomycin D assays reinforce the destabilizing effect that iPLUS seq3 3X exerts in *Spike*'s mRNA 75
Conclusions and Future Perspectives 79
References 83

List of Figures

Figure 1 – Molecular steps from DNA to proteins. 18
Figure 2 – *Cis*-regulatory sequences relevant for CPA. 19
Figure 3 – Alternative polyadenylation (APA) can affect the proteome in a qualitative and/or quantitative manner. 22
Figure 4 – Upstream sequence elements (USEs) can impact 3'end formation. 24
Figure 5 – *Drosophila*'s *polo* protein levels are regulated by the *cis*-acting element USE present in the 3'UTR. 27
Figure 6 – iPLUS promotes an increase in protein levels, in several model organisms.. 28
Figure 7 – Mutations present in Omicron's Spike protein. 30
Figure 8 – Representative scheme illustrating the basic components of an mRNA molecule used for *in vitro* transcription, with the possible mechanisms employed to increase the effectiveness of transcription and mRNA stability. 33
Figure 9 – Study workflow.. 36
Figure 10 – iPLUS and iPLUS seq3 3X do not show an improvement in Spike's protein production or mRNA levels, after transfection of HeLa cells with the pcDNA™3.1 plasmids, in comparison to constructs lacking these sequences. 58
Figure 11 – iPLUS seq3 3X increases Wuhan Spike's protein production and mRNA levels in HeLa cells, in comparison with the plasmid that lacks this sequence. 60
Figure 12 – iPLUS and iPLUS seq3 3X, when subcloned downstream of Omicron's Spike cDNA, do not impact the respective protein production or mRNA levels..... 61
Figure 13 – iPLUS seq3 3X increases Wuhan Spike's protein production and mRNA levels, in comparison with the Wuhan plasmid that lack this sequence, in ExpiCHO-S cells. 65
Figure 14 – In Spike's Omicron variant, both iPLUS and iPLUS seq3 3X lead to a decrease in Spike's protein production and the respective mRNA levels. 66
Figure 15 – eGFP protein expression is kept constant upon Wuhan Spike's plasmids co-transfection with pCS2_eGFP, whilst maintaining the iPLUS seq3 3X increase in protein production, in both HeLa and ExpiCHO-S cell lines..... 68
Figure 16 – *In vitro* transcription workflow. 73
Figure 17 – iPLUS seq3 3X decreases the mRNA stability of an *in vitro* transcribed Spike mRNA in HeLa cells. 74

Figure 18 – mRNA stability assays using Actinomycin D show that the presence of iPLUS seq3 3X downstream of Spike's cDNA leads to a decrease in mRNA stability, when compared to the control that lacks this sequence. 76
Figure 19 – Schematic representation of the workflow for testing our templates *in vivo.* 81

List of Abbreviations

ACE2 angiotensin converting enzyme-2

ActD Actinomycin D

APA Alternative Polyadenylation

ARE AU-rich elements

ATP adenosine triphosphate

bGH bovine growth hormone

bp base pairs

cDNA complementary DNA

CDS coding sequence

CFIIm cleavage factor II

CFIm cleavage factor I

CHO Chinese Hamster Ovary

CMV cytomegalovirus

COVID-19 coronavirus disease 2019

CPA cleavage and polyadenylation

CPSF cleavage and polyadenylation specificity factor

CR-APA coding-region APA

CTP cytidine triphosphate

CtsF cleavage stimulation factor

DMEM Dulbecco's Modified Eagle Medium

DNA deoxyribonucleic acid

dNTP deoxynucleotide

DSE downstream sequence element

DTT dithiothreitol

GFP green fluorescent protein

GTP guanosine triphosphate

Heph Hepaestus

HSV Herpes Simplex Virus

HIV-1 Human Immunodeficiency Virus 1

IgM immunoglobin M

iPLUS increased Protein Levels Universal Sequence

LB Luria Broth

LNP lipid nanoparticles

miRNA micro RNA

mRNA messenger RNA

ORF open-reading frame

PAP polyadenylation polymerase

PAS polyadenylation signal

PBS phosphate-buffered saline

PCR polymerase chain reaction

PEG Polyethylene glycol

Pol II RNA polymerase II

Poly(a) polyadenosine

pre-mRNA precursor RNA

PTB polypyrimidine binding protein

QuickCIP Quick calf intestinal alkaline phosphatase

RBD receptor-binding domain

RBP RNA-binding protein

RNA ribonucleic acid

rRNA ribosomal RNA

RT reverse transcriptase

RT-qPCR Quantitative reverse transcription PCR

SARS-CoV-2 Severe Acute Respiratory Syndrome Coronavirus-2

SDS-PAGE sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SSIV SuperScript IV

SV40 simian virus 40

TAE Tris Acetate-EDTA

TBS-T 1x Tris-Buffered saline with 0.1% of Tween 20 detergent

TK thymidine kinase

tRNA transfer RNA

U1A U1 small nuclear ribonucleoprotein-A protein

USE upstream sequence element

UTP uridine triphosphate

UTR untranslated region

VOC variants of concern

WPRE Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element

Introduction

From transcription to translation: the importance of the mRNA

Our genetic information is stored as DNA, a molecule that consists in nitrogenous bases (adenine, thymine, cytosine, and guanidine), attached to a backbone of sugars (deoxyribose) and phosphate groups. Two chains of this polymer are linked through hydrogen bonds between complementary bases, providing the well-known double-helix pattern of DNA discovered by Watson and Crick in 1970 (1, 2). DNA can then be transcribed into RNA molecules, also constituted by sugars (in this case, ribose), phosphate groups and nitrogenous bases (adenine, uracil, cytosine, and guanidine). RNA has the particularity of being, for the most part, a single-stranded molecule and several types exist, including: messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA) (3). rRNA, as the name implies, is the primary component of ribosomes; tRNAs are responsible for carrying the corresponding amino acids to the ribosome, depending on the expressed codon, (3); and mRNAs provide the information for ribosomes to produce proteins in the cytoplasm (4).

DNA will serve as a template for the RNA to be formed and the process of gene transcription comprises a series of events, as illustrated in **Figure 1**. First, RNA polymerase must be assembled onto the promoter of the gene to be transcribed (5). In eukaryotes, RNA polymerase II (Pol II) is a crucial enzyme for mRNA transcription, being responsible for the transcription of all protein-coding genes as well as some non-coding RNAs. Pol II activity leads to the formation of a precursor messenger RNA (pre-mRNA), through a process that involves three main steps: initiation, elongation and termination (6). Pre-mRNA molecules must go through a process of maturation, to become able of migrating to the cytosol, where ribosomes are present and proteins are produced (5). In fact, if alterations or dysregulations in the pre-mRNA processing occur, the cell can be compromised, and different pathological conditions may appear (7).

RNA processing to form a mature mRNA includes the introduction of a 7-methylguanosine cap in the 5’ end, the steps related to splicing – intron excision and consequent exon ligation, and the addition of a long tail of adenines in the 3’ end of the pre-mRNA, the polyadenylation (polyA) tail. All these processing events are very important for the fate of the newly transcribed mRNA. Moreover, transcription of the pre-mRNA molecule takes place at the same time as the RNA processing events, so, maturation of the mRNA is mostly co-transcriptional, as these two processes also seem to be mechanistically interdependent (5).

Concerning the pre-mRNA processing steps, the capping at the 5’end occurs after the first 25 to 30 nucleotides are added to the nascent transcript. This cap will be important

for the recruitment of proteins associated with the other steps of mRNA maturation and export from the nucleus, but will also protect the transcripts from degradation (8).

Splicing is the process through which non-coding regions, present in the genomic sequence as introns, are removed from the pre-mRNA (intron excision), leaving only the exons, that are then joined together (exon ligation) (9). In some cases, different splicing events can occur for one given gene, with alternative exon inclusions and exclusions, resulting in a high diversity of possible transcripts through alternative splicing. In fact, a single gene can provide several thousands of mRNA isoforms, due to these alternative splicing events, translating into transcript and protein diversity (10).

PolyA tails, which are added by the 3'end cleavage and polyadenylation machinery (CPA), are present, with the exception of histone mRNAs, in all mature eukaryotic mRNAs (11). PolyA tails contain 150-200 adenosines in complex organisms, although this number can vary (12). As a matter of fact, variation in the length of the polyA tail can impact the fate of the mRNA transcript in the cytoplasm, with short polyA tails being associated with degradation pathways, while long polyA tails are more prone to promote mRNA stabilization (13). Besides mRNA stabilization, the pre-mRNA 3'end processing will also impact the export of the mature transcript from the nucleus to the cytoplasm and the translation efficiency of the mRNA (14). Interestingly, when the polyA site, which is the place where the polyA tail is added, of a mRNA is substituted with the polyA site of a rRNA (not polyadenylated), the rate of transport to the cytoplasm suffers a large decrease, impacting also the protein levels (15). When the cleavage of the nascent transcript and subsequent polyadenylation occurs in different locations for the same gene, due to the selection of alternative polyA sites, this also increases transcriptome diversity by originating isoforms with different 3'end lengths by alternative polyadenylation (APA) (16), as it will be discussed with more detail ahead.

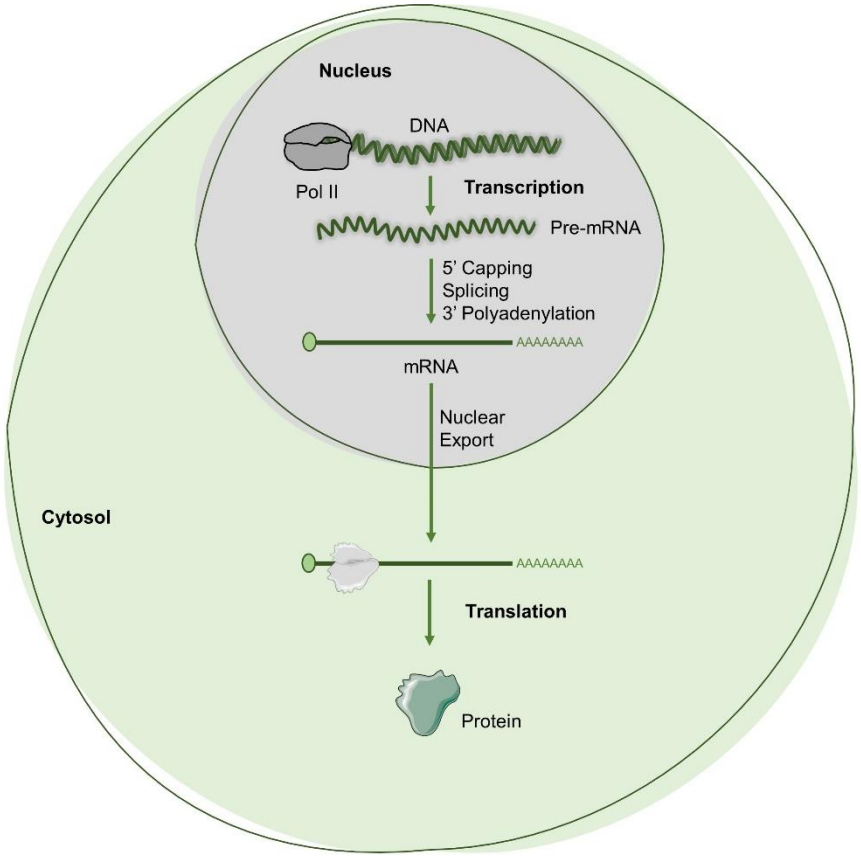


Figure 1 – Molecular steps from DNA to proteins. RNA polymerase II (Pol II) will transcribe a determined gene, producing a pre-mRNA molecule. Pre-mRNA processing steps (5'capping, splicing and 3'polyadenylation) must occur to produce a mature mRNA, capable of exiting the nucleus to the cytosol, where translation into the corresponding protein will take place. Adapted from Alberts *et al.*, 2015.

Alternative splicing and alternative polyadenylation impact gene expression regulation and can be seen as modulators of this process, since they can originate numerous transcripts from a single gene (17). Gene expression regulation is crucial, since it will guarantee that at a given time, the cell possesses the correct proteins, at a proper abundance, to maintain its normal function. Transcription of defined genes allows cells to develop a determined phenotype, a phenomenon essential for multicellular organisms. mRNA is a key component for the fine balance required in a cell, and as so, it is also highly regulated regarding its levels of production and its presence in different parts of the organism and the cell, in critical time points (18). Therefore, gene expression regulation is of extreme importance and in cases where dysregulation occur, some diseases may appear, such as cancer (18, 19).

3’end Processing: Polyadenylation

To terminate transcription of a determined pre-mRNA, an endonucleolytic cleavage at the 3’end must occur, being this the first step of polyadenylation. A signal in the pre-mRNA molecule is needed for the CPA machinery to recognize where the cleavage should occur, and this signal is known as the polyadenylation signal (PAS) (20). The most conserved signal and the stronger in polyadenylation efficiency is the AAUAAA hexanucleotide sequence, known as the canonical PAS (21), but several variations to this sequence have also been described, although less efficient in polyadenylation (22). The PAS is located around 15 to 30 nucleotides upstream of the cleavage and polyA site, usually a CA dinucleotide, being then followed by the polymerization of a tail of polyadenosines by the polyA polymerase (PAP). This addition corresponds to the second step of polyadenylation (21).

Besides the PAS, other *cis*-regulatory elements are important for the efficiency and usage of the polyA site, especially in assisting weaker PAS (23, 24). These regulatory elements can be found both upstream or downstream of the PAS and are, respectively, named upstream sequence elements (USE) or downstream sequence elements (DSE). DSEs appear approximately 20 nucleotides downstream of the polyA site and are commonly U/GU-rich, although they present some sequence variability (25). Cleavage and polyadenylation will occur somewhere between the PAS and the DSE, as represented in **Figure 2** (14, 26).

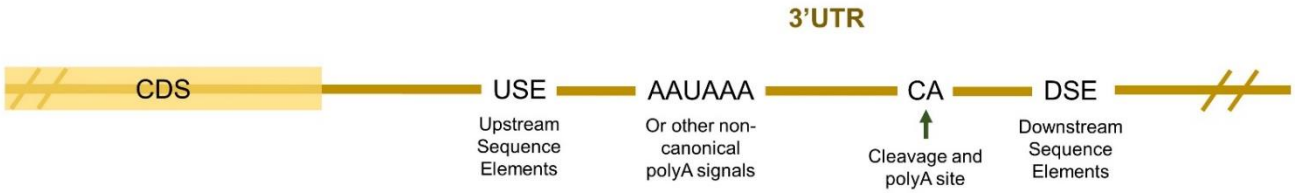


Figure 2 – *Cis*-regulatory sequences relevant for CPA. Representation of the consensus sequences for cleavage and polyadenylation. Adapted from Proudfoot, 2011.

Notably, over 80 proteins are associated with the multi-subunit complex responsible for pre-mRNAs 3’end formation (27). Nevertheless, there are 4 main protein complexes that compose the cleavage and polyadenylation core factors, accounting for approximately 20 proteins (23). These protein complexes are the cleavage and polyadenylation specificity factor (CPSF) complex, the cleavage factor I (CFIm), cleavage factor II (CFIIm) and the cleavage stimulation factor (CstF) complex (27).

The CPSF complex is composed by six subunits (23), namely CPSF-73, the endonuclease responsible for pre-mRNA cleavage, CPSF-30 and WD repeat domain

33 (WDR33), that recognize and bind the PAS, and factor interacting with PAPOLA and CPSF1 (FIPL1), that bind U-rich sequences, such as the USEs (23, 28). This complex is recruited as soon as Pol II starts transcribing the nascent transcript (29). The cleavage stimulation factor (CstF) is comprised by three subunits, with CstF64 being crucial for the recognition of DSEs (30). The CFIm complex is also important for the selection of the PAS and includes three subunits: CFIm 25, CFIm 59 and CFIm 68 (31, 32), while the CFIIm complex comprises two subunits: CLP1 and PCF11 (33). When all these complexes are in place, PAP is recruited and the addition of adenosines by this enzyme will, in turn, stimulate the activity of the poly(A) binding protein nuclear 1 (PABPN1), important for the polyA tail length definition (13, 14).

Given the complexity and the numerous components involved in the pre-mRNA 3'end processing, small alterations in this equilibrium, namely in the concentration of individual proteins required for this process or by mutations in important *cis*-regulatory sequences, can translate into differential gene expression and polyadenylation and this can significantly impact the normal functioning of the cell (23, 29, 34).

The impact of the 3'UTR in gene expression regulation

Human genes present several PASs in their sequence, so, for a given gene, different PAS can be chosen, resulting in different mRNA isoforms. This phenomenon is known as APA and occurs in approximately 70% of our genes (35). If more than one PAS is present in the 3'UTR, the resulting transcripts will differ in their 3'UTR length, but their coding region may remain the same, therefore, a similar protein is produced. This type of APA is commonly denominated as 3'UTR-APA and is present in more than 50% of the human genes (35, 36). If, on the other hand, an alternative PAS appears in the coding region, either in an intron or in an exon, then this event will result in different proteins once transcription occurs, as represented in **Figure 3.A**. These types of APA are known as coding-region APA (CR-APA), when the PAS is located on an exon, or intronic APA (or IPA), when the PAS is located on an intron (37).

APA has been related to different states of the cell, meaning that a certain 3'UTR profile can be characteristic of a cellular state or disease, due to its ability to modulate gene expression regulation and the outcome of the mRNA (17). It has been demonstrated that cells in a proliferative state will tend to use more frequently a proximal PAS, producing shorter 3'UTRs and mRNA isoforms (38, 39), which has also been described for cancer cells (40, 41), activated T cells (38) or during cellular stress (42). On the contrary, in some cases, cells tend to favor the use of distal PAS, resulting

in 3'UTR lengthening. This is true for cells during differentiation (43) or senescence (44), as well as for some specific cell types, namely central nervous system cells (45).

A great example of the importance of APA for the normal physiological functioning of our immune system is the shift between the membrane bound and the soluble form of immunoglobulin M (IgM). The IgM heavy chain gene has two possible PAS, one located in an intron and another one in the 3'UTR, consequently resulting in the production of two different proteins. For plasma cells, the use of the PAS located in an intron leads to the production of a shorter protein with the loss of the transmembrane domain of the IgM, leading to its secretion (46). This is an example that shows that APA in an intronic region can alter the proteome in a qualitative manner, affecting proteome diversity due to the generation of different protein isoforms, even though, for this to happen, the resulting mRNA needs to meet the necessary requirements to reach to translation (7, 12). This is also an excellent example of the tight regulation of the 3'end machinery, since for this APA event to occur in plasma cells, the levels of CstF64 must increase upon B cell activation (47).

On the other hand, 3'UTR-APA can impact gene expression in a more quantitative manner, as represented in **Figure 3.B** (7, 12). It is known that the 3'UTR is a vital location for gene expression regulation, since it harbors various *cis*-regulating elements, such as binding sites for microRNAs (miRNA) and RNA-binding proteins (RBPs), in addition to the presence of several regulatory sequences, namely AU-rich elements (AREs), USEs and DSEs (23, 37). miRNAs are crucial for gene regulation, due to their ability to impact mRNA stability and subsequent translation (23). RBPs can bind to *cis*-acting elements in the 3'UTR, functioning as *trans*-regulatory factors. Binding of RBPs can result in the recruitment of other effector proteins, promoting a wide range of events, also impacting gene expression (48). With these facts in mind, longer 3'UTRs resulting from the selection of distal PASs may possess more of these regulatory elements, becoming more susceptible to regulation. On the contrary, mRNA isoforms with shorter 3'UTRs will be able to avoid much of that regulation, culminating in higher protein expression (37, 40).

CD5, a transmembrane glycoprotein that can be found on the surface of T-cells, is an example of this type of regulation. *CD5* can be produced by selection of three different PAS, resulting in three different mRNA isoforms that only vary in their 3'UTR length (49). It has been demonstrated through genome wide studies that when T-cells became activated, a shift occurs, and these cells start to produce mRNA isoforms with shorter 3'UTRs, due to the selection of proximal PAS (50). This is also the case of *CD5*, the shorter mRNA isoform is the most transcribed when T-cells became

activated, which associated with the binding of an RBP (polypyrimidine tract binding protein (PTBP1)) and the loss of the binding site for miR-204, will lead to an increase in the expression of the corresponding protein (49).

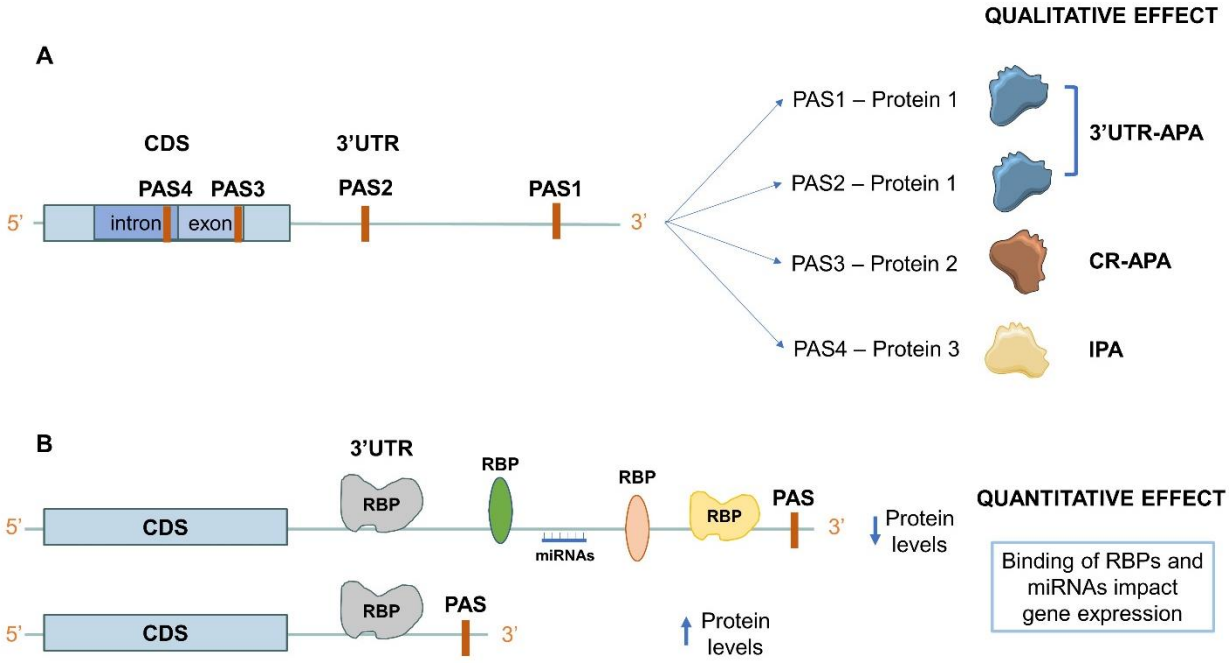


Figure 3 – Alternative polyadenylation (APA) can affect the proteome in a qualitative and/or quantitative manner. A) When different polyadenylation signal (PAS) are present alongside a transcript sequence, different phenomena can occur. If the two PAS are located on the 3' untranslated region (3'UTR), then the different mRNA isoforms will only vary in their 3'UTR length, leading to the production of a similar protein (3'UTR-APA). On the other hand, if a PAS is present on the coding or the intronic sequence, either on an exon or an intron, the resulting protein will be different (CR-APA and IPA). This impacts the proteome on a qualitative manner, affecting the proteins present. **B)** APA can also impact the proteome in a quantitative manner, since different *cis*-acting elements, such as binding sites for microRNAs (miRNAs) or RNA-binding proteins (RBPs) can be present in two different mRNA isoforms, resulting from 3'UTR-APA. The decrease in 3'UTR length leads to a decrease in the number of *cis*-acting elements present, aiding the shorter mRNA isoform in being less susceptible to regulation. These, in some cases, leads to a higher levels of protein production resulting from the shorter mRNA isoforms. Adapted from Hollerer *et al*., 2014.

However, it is worth mentioning that this relation may not be a universal truth since some studies did not find this relation between shortening of the 3'UTR and protein production increase (50). For example, in *Drosophila melanogaster*'s *polo* gene, that will be discussed in more detail ahead, the distal mRNA isoform is more efficiently translated into protein and is responsible for most of the protein production (51).

The 3'UTR length will not only affect the protein expression levels, but also other important factors for the fate of the transcript, namely its stability, subcellular localization, translation efficiency and the function of the resulting protein (48, 52).

Taken together, all these mechanisms show the potential of gene expression regulation by the 3'UTR and APA, as well as the critical effects that their dysregulation may have on our physiological functioning.

Role of *cis*-regulatory sequences on gene expression regulation

As previously described, *cis*-regulatory sequences play a key role in pre-mRNA 3'end processing (23). There is a huge variety of *cis*-regulatory elements that are present in the untranslated regions, either in the 5' or the 3' (53). The DSEs and USEs were already mentioned, but many others exist. On the 3'UTR, for example, we can find AU-rich elements (AREs) that promote a decrease in mRNA stability and translation progress (54), or GU-rich elements (GRE), with related functions to AREs, also capable of impacting mRNA decay, among other post-transcriptional events (55). In the case of iron responsive elements (IREs), for example, these can be found either in the 3'UTR or in the 5'UTR and play important roles in the regulation of iron metabolism. Countless others *cis*-regulatory elements could be named (53), but throughout this thesis we will focus on one specific type of *cis*-regulatory element, the USE.

USEs are the auxiliary upstream elements that were found not only in cellular mRNAs but also in viral ones. These sequences lack a clear consensus motif, with their only similarity being the characteristic U-rich sequence (25, 56). These *cis*-regulatory sequences can harbor several binding sites to different RBPs, namely proteins of the CPA machinery, and as so, these are important for the efficiency of the mRNA 3'end formation and PAS selection, as illustrated in **Figure 4** (23, 34, 57-59). In fact, RBPs play an important role in gene expression regulation, through their binding not only to specific sequences, but also in a structural dependent manner, impacting the fate of the transcripts, as it will be described in the following examples (34, 48, 59-61).

USEs were initially found in viral genomes, where they are widely present (62-64). The simian virus 40 (SV40) is an example of this. Two different PAS exist on its genome, termed early and late PAS, with the late PAS being considerably more efficient than the early one. Carswell *et al*. found that an USE is present upstream of the late PAS, being essential for its efficiency, since deletion of this USE leads to a drastic reduction of the late mRNA levels (62). In fact, this USE allows the binding of U1 small nuclear ribonucleoprotein-A protein (U1 snRNP-A; U1A) (58), an important protein for PAS selection, that acts by inhibiting the selection of premature PAS in nascent transcripts (65). In turn, U1A interacts with CPSF160, which increases the polyadenylation efficiency and 3'end formation (58).

The Human Immunodeficiency Virus 1 (HIV-1) viral genome presents another example containing an USE. This virus has an inherent termination redundancy given that it possesses different termination sites. In order to produce the correct pre-mRNA, one of

these canonical PAS in the 5'end has to be disregarded, and a similar one present in the 3'end is recognized and used by the CPA machinery. An USE with important functions for the efficient processing of the later mRNA transcript was found in the 3'end, having a role in the stabilization of CPA protein complexes in the PAS. In fact, CPSF160 also interacts directly with this USE (66). Interestingly, the mechanism behind HIV-1 USE involves the formation of an RNA stem-lop structure that brings together the USE and the PAS. Therefore, another layer of complexity to the USE's function is related to spatial requirements (61).

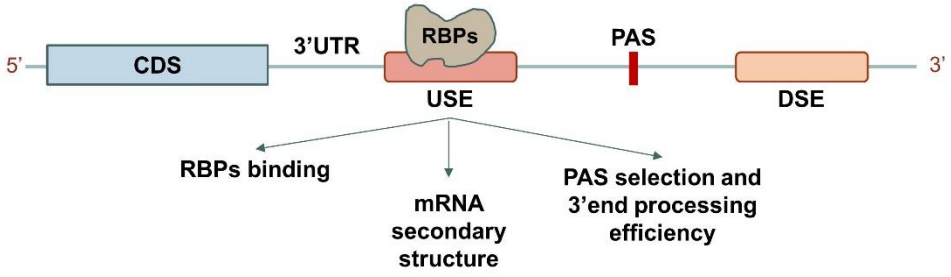


Figure 4 – Upstream sequence elements (USEs) can impact 3'end formation. USE's can function as a binding site for RNA-binding proteins (RBPs), namely the cleavage and polyadenylation machinery (CPA), promote the formation of specific mRNA secondary structures, in addition to its effect in polyadenylation signal (PAS) selection. All of these factors impact the 3'end processing efficiency.

In the case of adenovirus, five different mRNA transcripts can arise from five different PAS present at the major late transcription unit (MLTU), being these named L1 to L5 (67). USEs can be found in L1 (63), L3 (68), and L4 (69). For L3, it was demonstrated that the USE present is important also to increase the processing efficiency, through the formation of a stable CPA complex. Interestingly, this can explain the differences in the PAS choice between L1, a closer PAS, and L3. L1 USE's does not impact PAS strength, but when the upstream sequence from the L1 is substituted to the upstream sequence from L3, maintaining the remaining elements of L1, the processing efficiency is increased (70). For adenovirus-2 L4, the USE is important due to the formation of a hairpin in a region that involves both the USE and the PAS. In fact, the deletion of the USE's sequence or the disruption of this secondary structure, can lead to a 50% reduction the in cleavage and polyadenylation efficiency (60).

The Epstein–Barr virus (EBV) DNA polymerase (pol) mRNA possesses two different PAS in its sequence. The longer isoform contains an unusual and low in efficiency PAS, UAUAAA, that is still used for CPA. Analysis of the sequences flanking the PAS led to the discovery of an USE, whose mutation culminated in a decrease of the 3'end processing efficiency by three-fold (71). Remarkably, the sequence of this USE can also be found in tandem repeats in the Figwort mosaic virus, a plant virus, that

possesses the same unusual PAS. In this case, the presence of the USE upstream of the UAUAAA PAS, strikingly, leads to the disregard of a more proximal canonical PAS, with cleavage and polyadenylation occurring at that distal site (64).

An interesting application for these sequences explored by Schambach *et al*., consists in the subcloning of the USEs in viral vectors, enhancing its 3'end processing, in order to improve the vector's efficiency, avoiding readthrough (72). Notably, an increase in viral biosafety may also be achieved. Upon a possible integration of the viral DNA into the hosts genome, readthrough of cellular genes can lead to the accumulation of proto-oncogenes, triggering malignant transformation. Curiously, this study evaluated various USEs from both viral and cellular origin, such as the adenovirus L3 USE or the C2 complement discussed ahead, but the authors found that the SV40 USE in tandem two times was the one that gave better results regarding the titer levels and gene expression (72). In fact, USEs are interchangeable as they present functional similarities. To give an example, it was demonstrated that the HIV USE is able to replace the USE from the SV40 late PAS (73), but others examples exist (72, 74).

As referred before, some examples of USEs with important functions in eukaryotic genes were also identified. Regarding the example of the human ciclooxygenase-2 (COX-2), whose expression is induced in response to specific external cues, its sequence encodes two different, but commonly used PAS, one more proximal and other more distal. The proximal PAS is not the canonical one, but a weaker variation, AUUAAA, however, is still highly used in comparison with the distal stronger canonical PAS. Of note, upstream of this proximal PAS, several USEs are present and if these are mutated, the expression levels of this shorter mRNA isoform decreases (75). This reinforces the importance of the USEs in the selection of weaker PAS.

Several other examples where USEs have been identified can be named, such as the Complement C2 (76), prothrombin (77) or Lamin B2 (78, 79). Complement C2 has an USE of little over 50 nucleotides, that when mutated, affects the full efficiency of PAS selection. In fact, this sequence is crucial for both cleavage and polyadenylation reactions, with different mechanisms and RBPs involved (59). PTBP1 binds to *C2* USE's increasing the efficiency of cleavage. Additionally, this USE increases the efficiency of binding of CstF64 to the PAS, enhancing polyA tail addition (59). Interestingly, *C2* USE is highly conserved across five different mammals (76).

The human prothrombin also possesses an USE in the 3'UTR. This gene is associated with gain-of-function mutations of its 3' end cleavage site, increasing the chances of suffering a venous thrombosis in carriers. The A containing 3'cleavage site CA (the most common) is more efficient for cleavage and polyadenylation than the usually

found in prothrombin, CG. Besides this, the DSE found in this gene is weak. To counter this, prothrombin possesses a USE that compensates the low activity of the other *cis*-regulatory elements present (57). The physiological mechanism behind this USE involves inhibitory proteins (FBP2 and FBP3), that while bound repress transcription of prothrombin's mRNA, but upon stimuli detach and allow the USE to increase 3'end formation, by binding of stimulatory factors (80).

For Lamin B2, an USE present in the 3'UTR was also found. In this case, the USE is important due to its ability to stabilize the binding to CPSF, therefore promoting mRNA 3'end formation, even at the level of PAP's polyA tail addition (78).

Other example where USEs have a function in polyadenylation and gene expression regulation is *polo*, a *Drosophila*'s gene coding for a serine-threonine cell cycle kinase (81). This protein produces two different mRNA isoforms, varying on their 3'UTR length. The usage of the distal PAS seems to be essential for *Drosophila'*s viability, as transgenic flies that lack the distal PAS present an inefficient proliferation of histoblasts, the precursor cells of the abdomen, leading to high fly lethality (51). This happens because the *polo* longer mRNA isoform, originated by APA, is more efficiently translated than the shorter mRNA isoform. Therefore, when the longer mRNA isoform is missing, Polo's protein levels are insufficient and flies present a strong abdominal phenotype, associated with increase lethality (51).

The Gene Regulation group discovered that *polo* possesses a USE in the 3'UTR, upstream of the proximal PAS being this 28-nucleotide, pyrimidine-rich sequence, highly conserved among *Drosophila* genomes (34). Additionally, a recent bioinformatic analysis done by our group showed that the most conserved region of the USE (TTGTTTT), present in a maximum distance of 450 nucleotides upstream the *polo*'s non-canonical proximal PAS (ATTAAA), appears in 31 genes shared by mouse, zebrafish and human (unpublished data). This points to a possible role of the USE in the regulation of a subset of genes, conserved in the 3'UTR of different species. Generation of transgenic flies lacking the USE leads to the appearance of phenotypes related to low polo protein levels, pointing to an involvement of this sequence in the modulation of PAS selection, as represented in **Figure 5**. Besides that, the lack of this sequence also affects the cell cycle kinases Aurora B and Msp1, Polo's targets, culminating in mitotic abnormalities, such as aneuploidies. This clearly shows that this USE has an important physiological role in the regulation of *polo* protein levels (34). Furthermore, upon an RNA-protein pulldown assay, the Hephaestus (Heph) protein, an orthologue of the human PTBP1, was identified as binding the USE, being essential for distal PAS selection. These findings were reinforced by the evidence that *heph* mutants

present lower levels of the distal *polo* mRNA isoform as well as lower polo protein levels. In fact, *heph* mutants also present mitotic abnormalities, due to low activity levels of polo in the kinetochores of dividing neuroblasts (34).

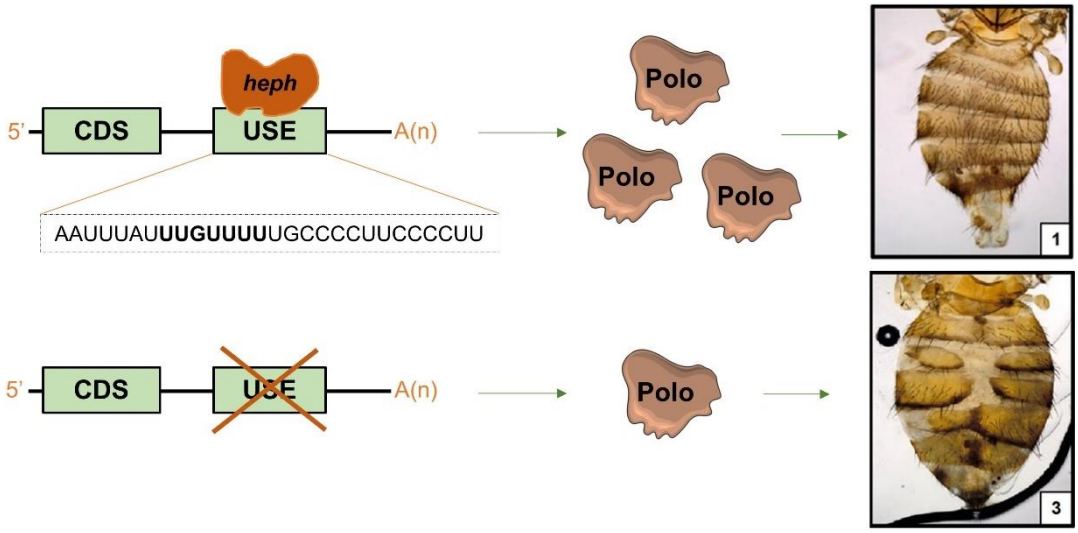


Figure 5 – *Drosophila*'s *polo* protein levels are regulated by the *cis*-acting element USE present in the 3'UTR. The USE has a crucial role in *polo*'s distal PAS selection, which is more efficiently translated, allowing the production of the correct amount of Polo protein. When the USE sequence is not present, *polo*'s protein levels decrease, leading to the appearance of abdominal defects. Furthermore, *heph*, the RBP that binds to the USE is represented, as well as the sequence that corresponds to *polo*'s USE, with the most conserved regions in bold. Adapted from Oliveira *et al.*, 2019.

Increased Protein Levels Universal Sequence (iPLUS) – an USE with practical applications

After the discovery that this 28-nucleotide pyrimidine-rich sequence located upstream of the proximal PAS in *Drosophila*'s *polo* was able to increase the corresponding protein levels, the group focused on investigating *polo*'s USE mechanism of action, and if it could impact additional genes and organisms. With that purpose, proof-of-concept experiments were carried out in other genes and organisms. Concretely, in *Danio rerio*, the microinjection of a plasmid containing this USE downstream of the coding sequence of the green fluorescent protein (GFP) was used as a reporter. The experiment demonstrated that, when in comparison to a mutated USE, this 28-nucleotide sequence was able to increase GFP expression levels (82). In HeLa cells, it was also demonstrated that the USE can increase luciferase activity (82). Moreover, experiments in yeast (*Pichia pastoris*), and in other mammalian models, such as the adapted Chinese hamster ovary (ExpiCHO-S) cells, also support the role of USE in the increase of protein levels in different biological models (unpublished data), as represented in **Figure 6**.

All the results involving different biological systems had a consistent increase in protein levels when this USE was present and led to the patent of this sequence, that was named iPLUS, which stands for increase Protein Levels Universal Sequence (82).

More recently, the iPLUS sequence was subcloned downstream of a monoclonal antibody used in cancer immunotherapy. iPLUS doubled the protein levels of this antibody, when compared to a plasmid lacking iPLUS (unpublished results and **Figure 6**). In these assays, the iPLUS sequence was also associated with an increase in mRNA levels of the monoclonal antibody, when compared to plasmids without the iPLUS (unpublished results) indicating a possible role in mRNA stabilization.

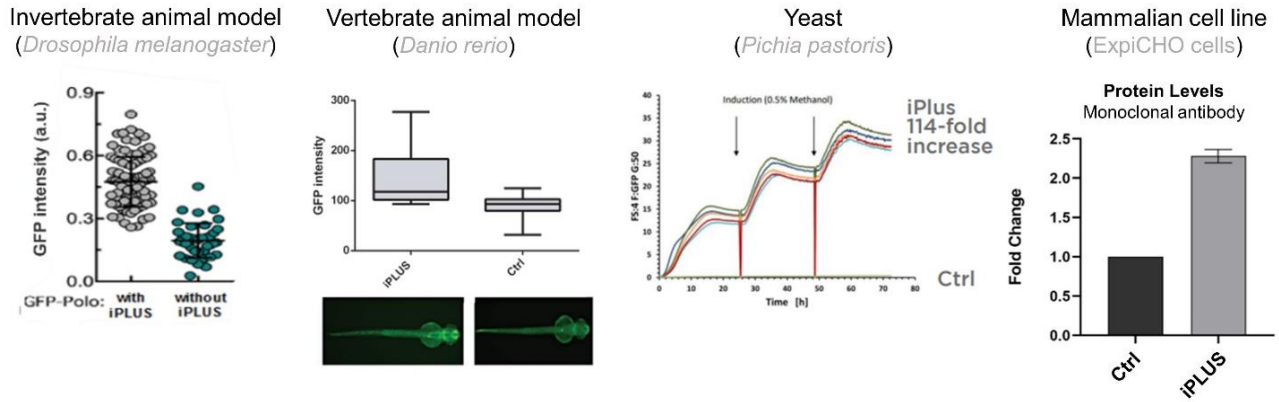


Figure 6 – iPLUS promotes an increase in protein levels, in several model organisms. iPLUS lead to an increase in GFP's expression in *Drosophila melanogaster*, *Danio rerio* and *Pichia pastoris*, and also to an increase in the production of a monoclonal antibody, in ExpiCHO-S cells. Adapted from (82) and unpublished results.

In addition, some variations of the iPLUS sequence have also been tested with good results, such as the 28-nucleotide long sequence repeated three times, or the most conserved region, UUGUUUU, in tandem three times, that was named iPLUS seq3 3X.

The iPLUS and sequence variants can be of extreme value, due to a wide group of advantages: a) it is a noncoding sequence present in the 3'UTR, so it will not affect the coding region of the resulting protein; b) it can be easily inserted in different vectors, without difficult techniques or procedures; c) it is compatible with different expression systems demonstrating its versatility; and d) it will allow an increase in the yield of protein production, resulting in a decrease of the costs for the production of a desired protein (82). These are all advantages that can be extremely valuable in an industrial environment, namely for pharmaceutic or biotechnological companies, since a bigger yield associated with reduced costs is always a desirable outcome.

iPLUS discovery and its biological relevancy gives even more emphasis to the critical role that mRNA non-coding *cis*-regulatory sequences can exert, in the regulation of gene expression.

mRNA and Health

Virus and pandemics: the COVID-19 outbreak

Several pandemic outbreaks have marked history over the years, with virus being responsible for some of those, namely the Spanish flu between 1918 and 1919, and the new coronavirus disease 2019 (COVID-19) outbreak, among others (83). COVID-19 is the disease caused by a recently discovered coronavirus, called severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) (84). SARS-CoV-2 is an RNA virus with a single-stranded positive RNA genome, being these the viruses with the largest genomes known (85). SARS-CoV-2 first appeared in the city of Wuhan, in China, but rapidly spread to all over the globe. Symptoms are very broad, with the more typical being fever, headaches, cough, and in some cases, the infection can culminate in acute respiratory distress syndrome (ARDS). Older people or with co-morbidities are more susceptible to develop worse complications (86). To this day, around 600 million cases of SARS-CoV-2 infection have been reported worldwide, with almost 6.5 million of deaths related (87). This brings a huge burden to the health care systems, with higher expenses, but mostly, culminates in lower quality of life and higher mortality (83, 88).

SARS-CoV-2 is an enveloped virus, containing four major structural proteins: membrane (M) , nucleocapsid (N), envelope (E) and the spike protein (85). The Spike protein is responsible for the binding of SARS-CoV-2 to the host cells and subsequent fusion, being present at the surface envelope of the virus. This homotrimer glycosylated protein has an N-terminal part, called S1, which forms an external globular domain where the receptor-binding domain (RBD) is located, and a C-terminal part, or S2, which functions as a transmembrane anchor and contains the membrane fusion peptide (89). A subdomain present in the RBD is responsible for the binding to the angiotensin converting enzyme-2 (ACE2) receptor, which then allows the viral entry into cells. On the viral envelope, Spike may exhibit two different conformations, either a closed one that is unable to bind to ACE2, or an open conformation, where RBD is exposed for binding (90). The Spike-ACE2 binding leads to a proteolytic cleavage on S2, still maintaining the Spike protein together by non-covalent bonds. Normally, the cleavage is performed by a cellular serine protease, the transmembrane serine protease-2 (TMPRSS2), which activates the Spike protein, by exposing the fusion peptide and promoting a more favorable conformation for viral entry (89). Being a single-strand positive virus, its genome can act as mRNA, bind to ribosomes and start

protein production immediately, in order to produce virions for further release (91). The Spike protein, more specifically its N-terminal domain, is the target of almost all the antibodies elicited by infection or vaccination (92, 93).

Since the beginning of the pandemic, several variants of concern (VOC) have emerged from the originally discovered Wuhan-Hu-1 variant. Mutations are the cause of this, and the higher the transmission rate, the higher is the probability of mutations to occurs, affecting the biology of infection. By order of appearance, five major variants have been identified: Alpha (B.1.1.7), Beta (B.1.351), Gamma (P1), Delta (B.1.617.2) and Omicron (B.1.1.529) variants (94). The Omicron variant firstly appeared in November 2021, and rapidly spread, leading to an increase in infections worldwide, with already described sublineages (95). Interestingly, the Omicron variant presents a total of 37 mutations in Spike when compared to the original Wuhan variant, as represented in **Figure 7**, a few coincident with the ones found in the other VOCs, but others were newly described. Remarkably, 15 of these mutations are on the RBD, impacting the binding to ACE2 (96). The appearance of these mutations confers to Omicron a higher strength of binding to the ACE2 receptors (97). Furthermore, the antigenicity is also altered with Omicron, which is able to escape most antibodies produced following vaccination and after infection with other SARS-Cov-2 variants. Different conformations of the Spike protein trimer result from these mutations, further explaining the immune evasion (98, 99). In short, Omicron has a higher capability of infecting host cells, with higher transmissibility, even though the severity of disease is lower (98).

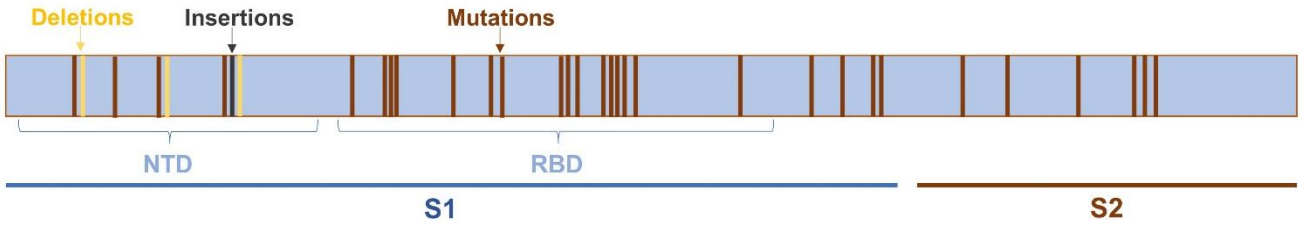


Figure 7 – Mutations present in Omicron's Spike protein. Schematic representation of the full-length sequence of the Omicron's Spike protein. Deletions are represented as yellow bars; insertions are in grey and mutations in brown. NTD – N-terminal domain. RBD – Receptor-binding domain. Adapted from Cui *et al*., 2022.

iPLUS as an asset in COVID-19 mRNA vaccines - iCoV2PLUS

With all the results regarding the value of iPLUS in mind, it was hypothesized that this sequence could be an important tool to be applied in the mRNA vaccines industry, which had an enormous expansion with the SARS-CoV-2 pandemic.

mRNA as a therapeutic agent has been on the table for a long time, but due to the instability of this molecule and the costs associated with it, the idea stayed mainly in paper until the emergence of this pandemic (100). In 1987, Malone *et al.* first realized that when mRNA molecules were mixed with fat droplets, they gained the ability of being delivered inside of the cell, actually leading to the synthesis of the respective protein (101). This was a huge milestone for the mRNA vaccine technology.

mRNA as a drug has the advantage of delivering the specific sequence of a known antigen, instructing our cells to manufacture the protein, which then culminates in an immunologic response specific for that immunogen (102). As so, it has several safety advantages, namely the non-infectious nature of the mRNA molecule, that will be naturally degraded by our cells, and its inability to integrate the genome (comparing to both attenuated virus vaccines and DNA vaccines) (103). Besides that, it offers the possibility of adapting the sequence in a quicker and simpler way, if a new strain of virus appears, also offering the possibility of immunization with more than one mRNA-encoding antigens (104). Indeed, mRNA vaccines are believed to be a useful resource in a vast group of applications, from therapeutic vaccination regarding multiple cancers, to a preventive immunization towards a variety of pathogens, including new pandemics and outbreaks that can once again affect the whole world (104, 105). In fact, Moderna and Pfizer/BioNtech developed their vaccines against SARS-CoV-2 in record time, within less than a year from the time that the virus genome was sequenced (106).

For the SARS-CoV-2 virus, most of the mRNA vaccines are currently using the full-length Spike protein sequence, providing further epitopes, than a minority that is using a shorter sequence corresponding to the RBD (107). Additionally, two mutations for proline residues are present in the most widely used vaccines, in order to maintain Spike's in its pre-fusion conformation. After immunization, a Spike protein, glycosylated, homotrimeric, membrane-anchored and exposed at the surface of cells, that resembles the native conformation of Spike from SARS-CoV-2, will elicit the production of protective antibodies (106).

The basic components of an mRNA construct are the coding region of the transcript of interest with an open-reading frame (ORF), a 5'UTR and a 3'UTR, a 5'cap and a 3'

polyA tail, as illustrated in **Figure 8**. The first step is to generate a plasmid for *in vitro* transcription, containing the coding region of the antigen and a promoter, such as the T7 RNA polymerase promoter, so that transcription can be carried out by an RNA polymerase (108). The addition of the 5'cap is of extreme importance for a mature mRNA, as discussed previously, and it can be achieved enzymatically, usually by the recombinant enzyme *Vaccinia Virus*, or during the *in vitro* transcription process, utilizing synthetic cap analogues. Cap analogues enable a simpler reaction setup but present a few limitations when in comparison to the enzymatic capping, namely these can be incorporated in the reverse orientation, leading to a decrease in translation. Moreover, cap analogues compete with guanosine triphosphate (GTP), the initiator nucleotide, resulting in some uncapped mRNA molecules (109). Regarding the addition of the 3'polyA tail, the same is true, it can be added by recombinant PAP, or it can be encoded in the plasmid DNA, being the latter preferable, since it allows for more reproductible lengths (110). The resulting mRNA molecule needs to be purified, being this a crucial step for the efficacy of mRNA as a therapeutic agent (102).

The use of modified nucleotides during transcription, such as pseudouridine, is also being employed, since this alteration can decrease the immune response in our body towards the mRNA molecule that is being injected. This strategy is used in Pfizer and Moderna's COVID-19 vaccines, for example (100, 111). Another approach that can be employed to increase the translation is the use of optimized codons. Synonymous codons can lead to the addition of the same amino acid, but the substitution of some of these less common codons for the more frequent ones, can increase the efficiency of protein production (102, 110).

The path from the injection site to the interior of our cells, where mRNA will be translated, can also be critical for its fate. Strategies must be employed for this intrinsically instable molecule to reach the cells in a viable state, and this is also a field with intense research (102). Nowadays, the most used system are the lipid nanoparticles (LNP) formulations. LNPs provide protection to the mRNA from nucleases highly present in our blood stream, improving its stability. LNPs are also important for efficient delivery to the cytoplasm of cells, when in comparison to naked mRNA molecules, and additionally aid in the escape from the endosomal pathway, that would also lead to mRNA degradation (107, 112).

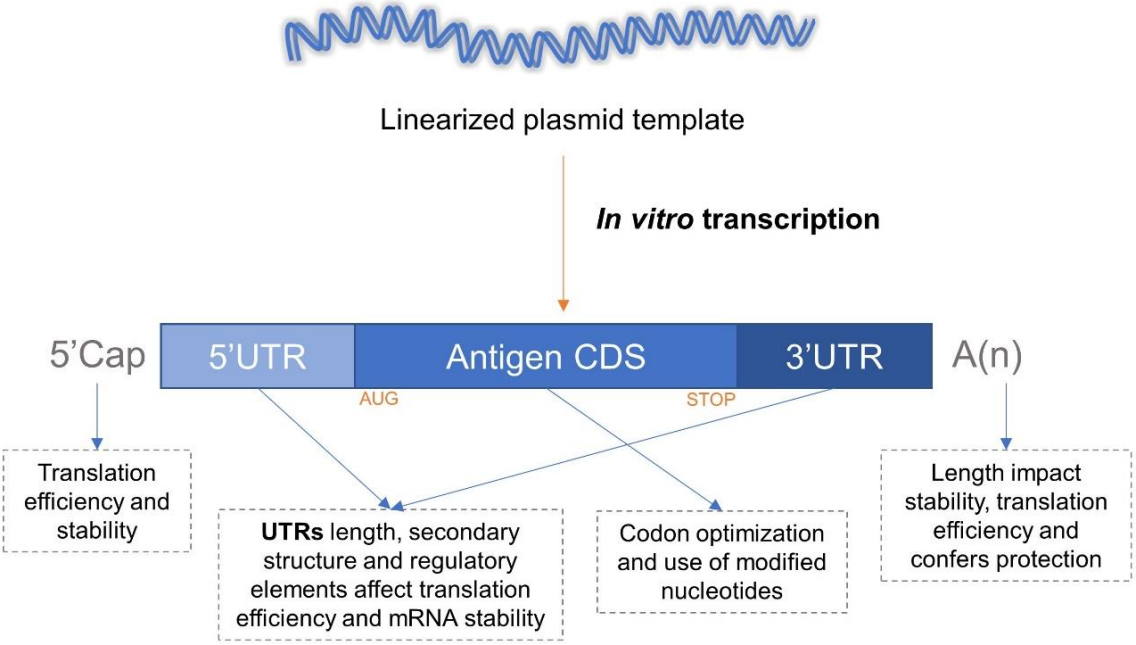


Figure 8 – Representative scheme illustrating the basic components of an mRNA molecule used for *in vitro* transcription, with the possible mechanisms employed to increase the effectiveness of transcription and mRNA stability. First, a plasmid template needs to be linearized, in order to process for the *in vitro* transcription. A 5'cap and a 3'polyA tail, a 5' and 3'UTRs and an ORF containing the coding sequence of the desired antigen are the basic components of an mRNA construct. Four factors regarding the mRNA construct can be improved to increase the performance of the construct: 5'Cap efficiency, UTRs length, secondary structure and regulatory elements, modifications of the coding sequence (CDS) and polyA tail length. Adapted from Gómez-Aguado *et al.*, 2020, and Jackson *et al.*, 2020.

As stated in the previous sections, UTR's can be of extreme importance for improving gene expression, and even though it is not a coding region, its sequence is designed with great care. 5'UTR secondary structures, 3'UTR length, and different regulatory elements present in both the 3'UTR and 5'UTR, can impact translation effectiveness (110, 113). Regarding the 5'UTR, the presence of the internal ribosomal entry sites (IRES) sequence, that promotes the binding of the ribosome, can improve translation initiation (114). The presence of specific sequences in the 3'UTR can impact, as referred before, mRNA stability and the consequent protein expression, but these sequences have been poorly investigated (110). Sequences from the α-globin and β-globin are commonly added in these contexts, since it is known that in a physiological context, their mRNAs are highly stable, due to sequences present in its 3'UTR (115). Likewise, the regulatory element of woodchuck hepatitis virus (WPRE) is a commonly used sequence in plasmids that works in a posttranscriptional manner. The presence of this element in the sense orientation of the 3'UTR promotes an increase in the expression of different proteins, such as GFP and luciferase (116).

iPLUS fits perfectly in this necessity for non-coding regulatory regions that can improve effectiveness of gene expression (82). Biotech and Pharma industries are constantly

searching for ways to improve their formulations, and iPLUS can be a way of ensuring that. The experimental results regarding this sequence point to its ability to improve not only the levels of protein expression, but also mRNA stability. Both aspects are important for the vaccine companies, since it would allow a more effective vaccine, aiding in the problems related to mRNA stability. It would also promote a more effective immunization, due to improved levels of protein transcription, triggering a more robust immune response. As so, it was hypothesized that iPLUS can provide an improvement in the field of mRNA vaccines.

Aims

The 3'UTR is a vital location for gene expression regulation, due to the presence of *cis*-acting elements, and the respective binding of *trans*-acting factors. In *Drosophila*'s *polo* gene, a USE *cis*-acting element located upstream of the proximal polyA site was found by our group with an essential role for the fruit fly development, impacting PAS selection and *polo*'s protein production (34, 51). Several proof-of-concept experiments carried out revealed that the USE, later patented as iPLUS, was able to increase protein production levels of a reporter gene, in several models besides *Drosophila*, such as mammalian cell lines, yeast and zebrafish (82). Furthermore, the USE has also been related to an increase in mRNA levels (unpublished data).

With the known properties of iPLUS in mind, we hypothesized that the iPLUS and variant sequences could be applied to the mRNA vaccines industry by providing an improvement in Spike mRNA stability and yield of protein production, both of extreme importance for the Pharma industry. For that, Spike from two different SARS-CoV-2 variants were studied, the original Wuhan variant still being used in mRNA vaccines, and the more recent and now dominant Omicron variant. The iPLUS and one of the iPLUS variants, iPLUS seq3 3X were tested by subcloning these sequences downstream Spike's cDNA.

Our main aim was to evaluate if iPLUS can in fact be an asset for the mRNA vaccines technology, by increasing Spike mRNA stability and/or enhancing protein production, and also to understand its mode of action. For that, two different types of assays were designed (**Figure 9**) with the purpose of studying two specific objectives:

1) To unravel if iPLUS and/or one of its variants promotes an increase in Spike's protein production and the corresponding mRNA levels, by performing transfection studies in HeLa and ExpiCHO-S cell lines.
2) To evaluate if the iPLUS and/or one of its variants are able to increase Spike's mRNA stability, by using *in vitro* transcription and Actinomycin D assays.

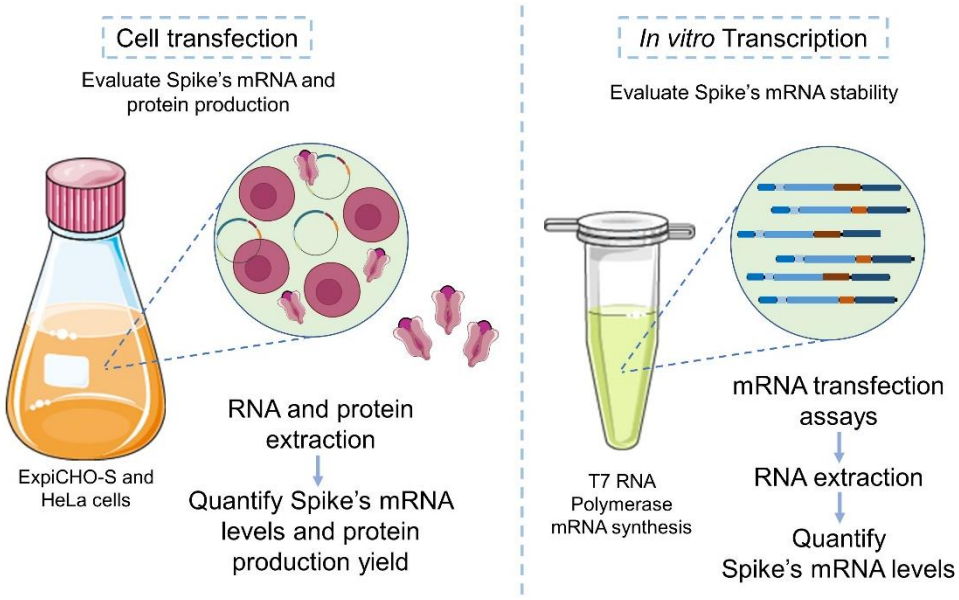


Figure 9 – Study workflow. Two different types of assays were designed. First, with the transfection assays, we wanted to evaluate the effect of iPLUS and one of its variants on Spike's protein and mRNA levels. For this, the iPLUS sequences were subcloned downstream of Spike's cDNA, and ExpiCHO-S and HeLa cell lines were transfected with the respective plasmids. RNA and protein were extracted for the required analyses. Secondly, we wanted to evaluate the respective effect of these sequences on mRNA stability. For this, mRNA was *in vitro* transcribed, and then also transfected in HeLa cells. This allowed us to evaluate Spike's mRNA levels after interaction with the normal cellular components, and by doing so, calculate its stability. Additionally, cells were transfected, followed by the addition of Actinomycin D, in order to evaluate mRNA half-lives, by which we can infer its stability.

Materials and Methods

1. Plasmid construction

1.1. DNA preparation

1.1.1. Polymerase Chain Reaction (PCR)

PCR reactions were set up by combining 15 ng of the correspondent DNA template (**Table 1**), 1x HF Phusion Buffer (Thermo Fisher Scientifics), 0.5 µM of both forward and reverse primers, 200 µM of dNTPs (Thermo Fisher Scientifics), 0.02 U/µL of the polymerase Phusion (Thermo Fisher Scientifics), and nuclease-free water (HyClone™ HyPure Molecular Biology Grade Water; GE Healthcare Life Sciences), to a final volume of 20 µL. Manufactures standard protocol was followed for the PCR programs choice, and primers and plasmid templates used are listed in **Table 1**.

Table 1. List of primers used to construct the plasmids.

Primer name	Sequence (5'-3')	Application
pGem-F	AGGTGACACTATAGAATACACG	Amplification of iPLUS and iPLUS seq3 3X sequences cloned in pGEM4 plasmids
pGem-R	CTGCAGGTCGACTCTAGA	
Spike_NotI	AATCGCGGCCGCTAACTAGAGAACCCACTGCTTA	Amplification of theT7 promoter, Spike coding region, and the iPLUS sequences (when present), from the pcDNA3.1 plasmids
BGH-R	TAGAAGGCACAGTCGAGG	

1.1.2. Agarose Gel Electrophoresis

Agarose gels were made using TAE 1x (Tris Acetate-EDTA) running buffer to dilute the proper amount of agarose, according to the percentage required, and 4 µL of SYBR Safe DNA Gel Stain (Invitrogen) for each 100 mL of TAE 1x was added. As a molecular weight ladder, 10 µL of GeneRuler™ DNA Ladder Mix (Thermo Fisher Scientific) was used. When needed, loading buffer (LB) 6x was included in the samples, to a concentration of 1x, to allow the observation of the nucleic acids. Electrophoresis was performed according to the specificities of each experiment and the results were analyzed in a Gel Doc™ XR+ System (Bio-rad).

1.1.3. Restriction enzyme digestion and DNA modification

Manufacturer's instructions were followed regarding the restriction enzyme digestions and the DNA modifications, with minor alterations. Briefly, all restriction enzymes (NEB) reactions were carried using 5-10 U of enzyme for each µg of DNA, in a final volume of 50 µL, with 1x restriction enzyme Buffer (NEB). Reaction mixtures were incubated for at least 1h30min, at 37º C. When sequential digestions in different buffers were

required, a purification step with the Macherey-Nagel™ NucleoSpin™ Gel and PCR clean-up kit (Thermo Fisher Scientific) was performed.

DNA Polymerase I, Large (Klenow) Fragment (NEB) was used for blunting DNA ends, either by filling in the 5' overhangs or by degradation of the 3' overhangs. For this enzyme, the reaction mixture was supplemented with dNTPs to a final concentration of 33 µM each, in 1x enzyme restriction Buffer and at least 1 U of Klenow per µg of DNA. Samples were incubated for a minimum of 30 min at 25º C.

The Quick calf intestinal alkaline phosphatase (QuickCIP) was used for dephosphorylation of most of the digested backbones, and it acts by removing the 5' and the 3'phosphate groups, as a mean to increase cloning efficiency. A µL of enzyme (5000 U/µL) was used for each pmol of DNA ends, with the respective Buffer at 1x. Samples were incubated for at least 30 min, at 37º C.

1.1.4. Sephadex DNA purification

To extract DNA from an agarose gel, bands were cut in the Gel DocTM XR+ System using UV light, and then were placed at -80º C for at least 20 min. To prepare the column for purification, 800 µL of Sephadex G-50 resin solution (6,67% w/v) were packed into each column, followed by a centrifugation of 4 min at 2220 g. The packed Sephadex columns were then placed in a clean Eppendorf. At the same time, bands were exposed to a heat shock, to help separate the DNA from the agarose gel, by going from the freezer to a thermoblock at 42º C, for at least 4 min, followed by vortex and spin down. The liquid and remaining solid part of the agarose bands were placed on top of the Sephadex-packed columns. Next, Eppendorfs were centrifuge for 10 min, at 4400 g, and the volume extracted was used to carry out the rest of the protocol. To the liquid fraction, 1/3 volumes of ammonium acetate (10M), 10/3 volumes of absolute ethanol and 1 µL of glycogen (20 mg/mL, Life Technologies) were added. This mixture was vortexed for about 10 s and then centrifuged for 5 min, at 14000 g. Supernatant was discarded, and the pellet, where the DNA is present, was washed twice with 1 mL of 70% ethanol, by centrifugation for 5 min at 14000 g. Once again, the supernatant was discarded, and the pellet was left to air dry. DNA pellet was eluted with a proper amount of nuclease-free water, depending on the strongness of the initial bands. Finally, concentrations were measured at a Nanodrop 1000 Spectrometer (Thermo Fisher Scientifics).

1.2. Subcloning

1.2.1. Ligation reactions

T4 DNA Ligase (Thermo Fisher Scientific) was used to perform ligations between the prepared backbones and inserts. For sticky-ends ligation and re-circularization reactions, 4 U of T4 DNA ligase (5 U/µL) was added to a mixture containing 2 µL of T4 DNA Ligase Buffer (Thermo Fisher Scientific), plasmid backbone (normally 100 ng for reaction) and insert (backbone/insert ratio was 1:5 for the majority of the reactions), plus nuclease-free water to a final volume of 20 µL. For blunt-end ligation reactions, the reaction setup is similar, except 5 U of T4 DNA Ligase (5 U/µL) were used, plus 2 µL of 50% PEG 4000 Solution were added (Thermo Fisher Scientific), increasing the viscosity of the solution, to promote the probability of ligation.
In parallel, two controls were prepared: one containing the backbone and the T4 DNA ligase, and one containing the backbone without the T4 DNA Ligase. These controls allowed us to understand if the backbone had more tendency to re-circularize than to allow the entry of the insert (backbone plus ligase) and also to understand if there was uncut plasmid present (backbone minus ligase). The ligation reaction mixtures were incubated overnight at room temperature.

1.2.2. Transformation of competent bacteria

Ligation reactions were added in a 1:10 (v/v) ratio to chemically competent bacteria (TOP10 and DH5α *E. coli* strains). Cells were incubated on ice for 15 min, followed by a heat-shock of 1 min at 42º C, after which the samples were placed again on ice for 5 min. Then, 4 volumes of Luria Broth (LB) were carefully added to the competent bacteria, followed by an incubation of 1 h at 37º C, with 220 rpm of agitation. The resultant bacteria were plated in LB agar plates, with Ampicillin (100 µg/L). For plasmid transformation, 100 µL of the bacterial suspension were plated directly in LB agar, while for ligation reaction, all volume was centrifuged for 2 min at 400 g, the supernatant was discarded, and bacteria were plated. LB agar petri dishes were incubated overnight at 37º C, and in the next day the results were analyzed, to decide if a colony PCR should be performed or follow for direct inoculation.

1.2.3. Colony PCR

To screen individual colonies, especially in the case of blunt end reactions, colony PCR was performed using GoTaq Flexi DNA Polymerase (Promega). A master mix was prepared containing 2 µL of 5x Green GoTaq Flexi Buffer (Promega), 0.75 µL of $MgCl_2$

(25 mM), 0.5 µL of dNTPs mix (10 mM each), 0.5 µL of the forward and 0.5 µL of the reverse primer (10 µM), plus 0.15 µL (0.75 U) of GoTaq Flexi DNA Polymerase, adding nuclease-free water until a final volume of 10 µL for each sample. A pipette tip was used to select the desired colony from the LB agar plate, and this tip was then dipped into the PCR mix. A 96-well plate was prepared as a mirror for the performed PCR, containing 100 µL of LB supplemented with ampicillin (100 µg/ mL). The pipette tip was retrieved from the PCR mix and dipped in the well containing LB plus antibiotic, and this plate was then incubated at 37º C. As a negative control, a clean pipette from the same box was dipped in a PCR tube containing the master mix, being then dipped in the 96-well mirror plate.

The PCR was carried in a thermocycler according with the following program: initial cell lysis at 95º C for 5 min; 35 cycles of 1 min at 95º C, 30 s at the corresponding annealing temperature, followed by extension at 72º C for a period depending on the expected amplicon size (1 min/1 kb); the final extension was performed at 72º C for 7 min. PCR products were then loaded in an agarose gel and run, normally, for 30 min at 120 V. After analyzing the electrophoretic results, the positive colonies were reinoculated using the bacteria grown in the mirror 96-well plate, in 5 mL of LB supplemented with ampicillin (100 µg/ mL). The colonies were incubated overnight at 37º C, with an agitation of 220 rpm. The following day, plasmid DNA was extracted by using the ZR Plasmid Miniprep kit (ZYMO RESEARCH), accordingly to the manufacturer's protocol.

1.2.4. Sanger Sequencing

The Sanger Sequencing procedure was performed at the i3S Genomic Core Facility to ensure that all plasmids used in this work had the correct sequence. Succinctly, the reaction mixture contained 5 µL of the plasmid (100 ng/µL), the desired primer (10 µM), BigDye Terminator v3.1 Cycle Sequencing Kit and 5x Sequencing Buffer (Applied Biosystems), plus nuclease-free water. The sequencing products are then purified using illustra™ Sephadex™ G50 Fine DNA Grade (GE Healthcare) and sequencing was performed by 3500 Genetic Analyzer sequencer (Applied Biosystems).

1.3. SpikeHu/Om_iPLUSes plasmid construction

A plasmid containing Spike's cDNA sequence from the Wuhan original variant of SARS-Cov-2 was acquired from Addgene, with the original name pcDNA3.1-SARS2-Spike (117), but hereafter, it will be renamed as SpikeHu. Likewise, a plasmid

containing Spike's cDNA sequence from the Omicron variant was acquired from GenScript Biotech (pcDNA3.1-SARS-CoV-2 Omicron Strain) and renamed SpikeOm.
These plasmids were used to subclone our sequences of interest, iPLUS and iPLUS seq3 3X, downstream of the Spike's coding regions. Given that the plasmid backbones were the same (pcDNA3.1), the cloning plan fitted both SpikeHu and SpikeOm plasmids, and consisted of a digestion with *Xho*I, followed by the blunting of the restriction sites with Klenow. After a purification step, digestion with *Xba*I was performed, leaving this cohesive end. Finally, plasmids were dephosphorylated using QuickCIP, leaving the backbone ready for cloning. For the inserts, the in-house plasmids, pGEM4_iPLUS and pGEM4_seq3 3x were used as templates, in order to amplify by PCR the iPLUS sequence, and its variant, iPLUS seq3 3X. PCR products were purified and subjected to a sequential digestion, initially with *EcoR*I followed by blunting with Klenow, and then with *Xba*I, matching to the cohesive end present in the backbones. The respective backbones and inserts were ligated using the T4 DNA ligase and competent bacteria were transformed as previously described.

1.4. Wuhan/Omicron_iPLUSes and Wuhan/Omicron plasmid construction

For the construction of the plasmids with the pcDNA3.4 backbone containing both Spike's cDNA from the Wuhan or Omicron variant and iPLUS or iPLUS seq3 3X, the in-house pcDNA3.4_LC_WPRE_iPLUS was digested with the restriction enzymes *EcoR*V and *Xba*I, followed by blunting with Klenow, to remove the LC sequence. Ultimately, the plasmid was dephosphorylated with QuickCIP. SpikeHu and SpikeOm plasmids were used for the preparation of the Spike's cDNA as an insert, by being doubly digested with *Sac*I and *Xho*I, followed by the blunting of the restriction sites using Klenow. T4 DNA ligase was used for the ligation of the backbone and inserts, and competent bacteria were transformed. This way, the pcDNA3.4 plasmids containing iPLUS and the two different cDNA variants were obtained, and these plasmids were named Wuhan_iPLUS and Omicron_iPLUS.
These plasmids were then used as backbones for the preparation of the plasmids containing the seq3 3X variant of the iPLUS. Once again, a sequential digestion was required, with the plasmids being digested with *EcoR*I, and after a purification step, with *Kpn*I. The DNA was then dephosphorylated using QuickCIP. Once again, a PCR amplification as described above, using pGEM4_seq3 3X as a template, was performed to obtain the iPLUS seq3 3X sequence, that was then digested with the same enzymes as these backbones (*EcoR*I and *Kpn*I), allowing for a cohesive end ligation on both sides. The plasmids obtained upon this cloning procedure were named

Wuhan_seq3 3X and Omicron_seq3 3X. T4 DNA ligase was used for ligation and competent bacteria were then transformed.

For the construction of the pcDNA3.4 plasmids lacking the iPLUS sequences, to be used as controls, Wuhan_iPLUS and Omicron_iPLUS were sequentially digested using *EcoR*I, followed by *Kpn*I and blunt-end formation using Klenow. The backbones were then allowed to re-circularize in the presence of T4 DNA Ligase, and competent bacteria were transformed as previously described.

1.5. SpikeHu_polyA and SpikeHu_seq3 3X_polyA plasmid construction

pUC57-Mini, the plasmid bought to GenScript containing a polyA tail with a short linker, was digested with *Not*I High Fidelity (HF) and *Hpa*I. Additionally, SpikeHu and SpikeHu_seq3 3X (pcDNA3.1 backbone) were used as PCR templates for the amplification of the T7 promoter, Wuhan Spike's cDNA and, in the case of SpikeHu_seq3 3X, iPLUS seq3 3X. For this, the primers SpikeHu_NotI and BGH-R were used. PCR products were digested with *Not*I-HF, to match the cohesive end in pUC57-Mini. The respective backbones and inserts were ligated using the T4 DNA ligase and competent bacteria were transformed.

2. Cell culture and transfection

2.1. HeLa cells culture

HeLa cells were grown and maintained with Dulbecco's Modified Eagle Medium (DMEM) with GlutaMAX (Gibco, Thermo Fisher Scientific), supplemented with 10% fetal bovine serum (FBS; Gibco, Thermo Fisher Scientific) and 1% penicillin-streptomycin solution (Gibco, Thermo Fisher Scientific). Cells were maintained in an incubator at 37º C, with 5% CO_2. Every 3 to 4 days, depending on the confluency of the cells, the medium was removed, cells were washed with 2 mL of 1x PBS (Gibco, Thermo Fisher Scientific), then were trypsinized at 37º C, with 500 µL TrypLE™ Express reagent (Gibco, Thermo Fisher Scientific). After that, cells were resuspended in fresh medium and moved to a new, sterile T25, with the appropriate dilution.

Cells were tested for Mycoplasma contamination, using a species-specific PCR, and no contamination was detected.

2.2. ExpiCHO-S cells culture

ExpiCHO-S suspension cell line (Thermo Fisher Scientific) was maintained in 30 mL of ExpiCHO™ Expression Medium (Gibco, Thermo Fisher Scientific), in 125 mL

Nalgene™ Single-Use PETG Erlenmeyer Flasks with Plain Bottom (Thermo Fisher Scientific). Cells were kept in an incubator at 37º C, with 8% CO_2 and 80% of relative humidity, with agitation at 125 rpm. When cells reached a concentration between $4x10^6$ to $6x10^6$ viable cells/mL, routine splitting to the desired dilution was performed, following the manufactures procedure.

An inverted light microscope (Zeiss Axiovert 25) was used to observe cell morphology. Cells were diluted in a 1:10 (v/v) ratio in Trypan Blue (Gibco, Thermo Fisher Scientific) and placed in a Neubauer chamber (Marienfeld Superior), to assess cell density and cell viability.

2.3. HeLa cells plasmid transfection

Hela cells transfection with the pcDNA3.1 and pcDNA3.4 plasmids was performed using Lipofectamine 2000 reagent (Thermo Fisher Scientific), accordingly to the manufactures standard protocol. First, HeLa cells were seeded the day before transfection in 24-well plates, in a concentration that allowed them to reach 70 to 90% confluency, at the time of transfection. Usually, 18 x 10^4 cells were plated in complete DMEM, for each experimental condition. To transfect cells, 500 ng of each plasmid was used. For the co-transfection with the pCS2_eGFP plasmid, 100 ng of this plasmid was used, plus 400 ng of each pcDNA3.4 plasmid (Wuhan/Omicron, Wuhan/Omicron_iPLUS and Wuhan/Omicron_seq3 3x). Briefly, the respective 500 ng of each plasmid were diluted in 25 µL of Opti-MEM (Thermo Fisher Scientific). In parallel, 1 µL of Lipofectamine 2000 was also diluted in 25 µL of Opti-MEM, followed by an incubation of 5 min at room temperature. Both mixes were combined, followed by another 5 min incubation at room temperature. After this, the whole volume was added to the respective well, drop wise, where the medium was previously changed to 500 µL of DMEM without antibiotics. Cells were allowed to grow at 37º C for 48 h, after which they were collected and processed, either for total RNA or protein extraction.

In the co-transfection experiments, to evaluate the efficiency of transfection, eGFP expression was also evaluated using a ZOE inverted fluorescence microscope (Bio-Rad).

2.4. Hela cells mRNA transfections

Three different protocols from three different companies were tested regarding mRNA transfection, always following the manufacturer's instructions. HeLa cells were seeded in complete DMEM the day prior to transfection and were allowed to grow until 70%

confluency was reached, a confluency in the recommended range for all three protocols.

iLipoP® (Adjuvartis) required the preparation of an mRNA formulation the day prior to transfection. mRNA was first diluted to a final concentration of 40 ng/µL in nuclease-free water. From the mRNA dilution, 5 µL were mixed with 5 µL of iLipoP® solution, at room temperature, and the mixture was vortexed for 2 s. Two volumes of cell-penetrating peptide (CPP; Adjuvartis) were added, at room temperature, and once again the mixture was vortexed for 2 s. The mRNA formulation was stored for approximately 24 h, at 4º C, to stabilize the mRNA complexes. On the day of transfection, this mRNA formulation was equilibrated at room temperature prior to transfection, and then diluted 10 times with Opti-MEM. The medium from each well was removed, and the mRNA formulation diluted in Opti-MEM was added. After 2h of incubation at 37º C with 5% CO_2, this formulation was removed, and media changed to complete DMEM.

Lipofectamine 2000 reagent was used following the same protocol as for plasmid transfection, this time using 500 ng of the *in vitro* transcribed mRNAs.

For JetMESSENGER™ reagent (Polyplus transfection), 500 ng of mRNA were diluted with 50 µL of mRNA buffer (Polyplus transfection), the mixture was vortexed and spun down. Following that, 1 µL of JetMESSENGER™ reagent was added, samples were vortexed and spun down. An incubation period of 15 min at room temperature was carried out, and the mixture was added drop wise to the respective well, where previously the medium had been changed to DMEM without antibiotics.

With the iLipoP® protocol, cells were incubated for 24 h at 37º C with 5% CO_2, while for Lipofectamine 2000 and JetMESSENGER™, transfected cells were incubated for 48 h under the same conditions, after which cells were collected for total RNA extraction.

2.5. ExpiCHO-S cells transfection

The ExpiFectamine™ CHO reagent (Gibco, Thermo Fisher Scientific) was used for ExpiCHO-S cells transfection, according to the manufactures protocol. Briefly, in the day prior to transfection, cells were split to a concentration of $4x10^6$ viable cells/mL and were allowed to grow until a cell density of $7x10^6$-$10x10^6$ viable cells/mL. Before transfection, cell viability was checked, and cells were split to a concentration of $6x10^6$ viable cells/mL. Cells were transfected with 18 µg of each pcDNA3.4 plasmid, and for the co-transfection experiments, an additional 5 µg of pCS2_eGFP was added. Briefly, plasmid DNA was diluted in 1 mL of OptiPRO™ SFM, while 80 µL ExpiFectamine™ CHO reagent was diluted in 920 µL of OptiPRO™ SFM. All reagents were cold, but the

setup was performed at room temperature. Both mixtures were combined, allowed to incubate for approximately 1 min, after which the whole volume was added to the flasks, drop by drop, while swirling the flask. On the day after transfection, 150 µL of ExpiFectamine™ CHO Enhancer (Thermo Fisher Scientific) and 6 mL of ExpiCHO™ Feed (Thermo Fisher Scientific) were added, and cells were kept in the incubator at 37º C, with CO_2 at 8% and agitation of 125 rpm, for the next 8 days.

Nine days post-transfection, at the time of cell harvest, cell viability should be at least 75% and that was assed using Tripan Blue and an inverted light microscope (Zeiss Axiovert 25), before proceeding to total RNA or protein extraction.

3. Protein extraction and Western Blot

3.1. Whole-cell protein extraction

For HeLa cells grown in 24-well plates, medium was first removed, cells were washed with 500 µL of 1x PBS, and 60 µL of NP-40 lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40) supplemented with 1x protease inhibitors (Sigma) was added to each well. The plate was incubated for 20 min on ice, cells were then lysed, and the liquid was retrieved to an Eppendorf. For ExpiCHO-S cells, 1 mL of each culture was centrifuge for 5 min at 300 g, the supernatant was discarded and 50 µL of NP-40 supplemented with 1x protease inhibitor was added. Cells were incubated for 20 min on ice. For both cell lines, samples were centrifuge at 4º C, for 5 min at 14000 g to remove any cell debris, the supernatant was retrieved to a new Eppendorf and then store at -80º C.

3.2. Protein quantification

Whole-cell protein extracts were quantified using the Quick Bradford reagent (Bio-rad). Samples were first diluted in lysis buffer, and 5 µL were loaded in duplicates into a 96-well flat-bottom plate. In addition, 5 µL of eight Bovine Serum Albumin (BSA) standards of known concentration (0.000, 0.125, 0.250, 0.500, 0.750, 1.000, 1.500 and 2.000 mg/mL) were added in duplicate, in order to plot a standard curve for protein quantification. Afterwards, 250 µL of the Quick Bradford reagent were added to each well, and the mixture was incubated for 5 min at room temperature. Finally, absorbance at 595 nm was measure in a microplate reader (Synergy 2).

3.3. Western Blot

After calculating the protein concentration of each sample, a constant amount of protein was prepared for the Western Blot analyses (normally 30 µg of the whole-cell protein extracts). Laemmli Buffer 2x (Bio-Rad) with 5% of β-mercaptoethanol was added to the samples in a 1:1 ratio, that were then boiled for 5 min at 95º C.

Samples were run in a 3% stacking and 10% resolving Tris-glycine SDS-polyacrylamide gel for 80 V for approximately 15 min, until samples entered the running gel and were aligned. Afterwards, the voltage was altered to 180 V for approximately 1h15min. The gel was then transferred into a nitrocellulose membrane, using a dry transfer method with the iBlot Gel Transfer Device (Thermo Fisher Scientific).

Afterwards, the membrane was blocked with 5% non-fat dried milk, prepared with a mixture of 1x Tris-Buffered saline with 0.1% of Tween 20 detergent (TBS-T), for 45 min. Secondly, the membrane was incubated for 1h15min with the respective primary antibody diluted in TBS-T containing 5% non-fat dried milk, followed by four washes with TBS-T. Next, the membranes were incubated for 1 h with the secondary antibody, also diluted in 5% non-fat dried milk with TBS-T. Finally, four washes with TBS-T were performed, after which a chemiluminescent agent (ECL Prime, GE Healthcare) was added on top of the membranes and incubated for 5 min, to allow protein detection. Immunoblots were then visualized on the ChemiDoc XRS+System (Bio-Rad), and protein quantification was performed on Image lab Software (Bio-rad).

The antibodies used in this study were: anti-SARS-CoV-2 Spike mouse monoclonal antibody (1A9; GeneTex; 1:2000 dilution), anti-α-tubulin mouse monoclonal antibody (B-5-1-2; Sigma, 1:150000 dilution), anti-GAPDH mouse monoclonal antibody (ab8245; Abcam; 1:1500000), anti-GFP mouse monoclonal antibody (B-2; Santa Cruz Biotechnologies; 1:5000 dilution) and goat anti-mouse IgG-HRP secondary antibody (Santa Cruz Biotechnologies).

4. RNA extraction, cDNA synthesis and RT-qPCR

4.1. RNA extraction

Total RNA extraction was performed using TRIzol (Ambion® by Life Technologies™), according to the manufacturer protocol. For HeLa cells grown in 24-well plates, medium was first removed, cells were washed with 500 µL of 1x PBS, and 500 µL of TRIzol was added. For ExpiCHO-S cells, 500 µL of culture were retrieved to a nuclease-free Eppendorf and centrifuge for 5 min, at 300 g. Supernatant was discard, and 500 µL of TRIzol was added.

Cells were incubated for 5 min and were then lysed. Following this, samples were either transfer to a nuclease-free Eppendorf and then frozen at -80º C, or were directly transfer to the PhaseMaker tubes (Thermo Fisher Scientifics), to proceed with RNA extraction. The PhaseMaker tubes were initially centrifuge for 30 s, at 16000 g. Samples were added to PhaseMaker tubes at room temperature, followed by the addition of 100 µL of chloroform (Merck), and tubes were vigorously shaken for about 15 s. Samples were incubated for 5 min at room temperature, and then centrifuge at 16000 g for 5 min, at 4º C. This step allows the physical separation of the organic and aqueous phases, by a resin present in the PhaseMaker tubes, with the aqueous phase remaining on top of this resin. The aqueous phase was collected to a new nuclease-free Eppendorf and kept on ice and 1 µL of nuclease-free glycogen (20 mg/mL) and one volume (v/v) of isopropanol was added to the sample. Samples were vortexed and incubated at -80º C for at least 2 h, to promote total RNA precipitation.

Samples were later thawed at room temperature and centrifuge at 4º C, for 20 min at 12000 g, supernatant was removed, and two washes with 500 µL of 75% ice cold ethanol were performed for 10 min, 12000 g, at 4º C. Lastly, pellet was air dried at room temperature, followed by resuspension on the appropriate amount of nuclease-free water, usually 15 µL for HeLa cells and 50 µL for ExpiCHO-S cells.

Quantification and quality check of the RNA was performed using the Nanodrop 1000 Spectrophotometer, being the quality analyses based on the ratios of absorbance at 280 nm and 260 nm. Samples were then stored at -80º C.

4.2. DNase I treatment and cDNA synthesis

Since genomic and/or plasmid DNA can be present in these samples, a treatment with DNase I was performed, to eliminate any possible contaminates for the next experiments. In the plasmid transfection experiments, 250 ng of total RNA was used for DNase I treatment followed by cDNA synthesis. For the mRNA transfection experiments, 1000 ng of total RNA was used.

Briefly, total RNA was first diluted in nuclease-free water, to a final volume of 9.8 µL. For DNase I treatment, 1.2 µL of 10x DNase I recombinant RNase-free incubation buffer (Roche) and 1 µL of Recombinant DNase I (Roche) were added, reaching a final volume of 12 µL. Samples were incubated for 2h30min (for plasmid transfection experiments) or 45 min (for mRNA transfection experiments) at 37º C, followed by 10 min at 75º C, to inactivate the enzyme. This procedure was followed by cDNA synthesis. For that, 1 µL of Random Hexamers (50 µM) and 1 µL of dNTPs mix (10nM each) were added to the previous DNase I treated samples. Mixtures were incubated 5

min at 65º C, followed by 5 min at 4º C. After, 4 µL of 5x SSIV Buffer (Invitrogen), 1 µL of dithiothreitol (DTT; 100 mM; Invitrogen), 0.5 µL of Ribolock Recombinant RNase Inhibitor (40 U/µL; Thermo Fisher Scientific) and 0.5 µL of Superscript IV (SSIV) Reverse Transcriptase (200 U/µL; Invitrogen) were added to each sample. As a control for possible genomic or plasmid DNA contaminations, samples were prepared in double and the same mixtures were added, only substituting SSIV RT for water (*minus* RT samples). Samples were incubated for 10 min at 23º C, 10 min at 55º C, and finally, 10 min at 80º C to inactivate the enzyme. Samples were stored at -20º C.

4.3. Quantitative reverse transcription PCR (RT-qPCR)

To quantify mRNA levels, several primers, represented in **Table 2**, were designed with the help of the online platforms Ensembl and NCBI Primer Blast.

We started by analyzing primer efficiency to ensure that each primer pair had good amplification efficiencies. This was done by performing RT-qPCR with several sequential dilutions of cDNA (1 or undiluted, 1:10, 1:100 and 1:1000), to obtain a standard curve. For a primer to be considered efficient, the slope of this standard curve should be around -3.3, which would reflect that the primer amplifies in the same manner independently of the different cDNA concentrations. We used the slope of the standard curve to calculate the primer pair amplification efficiency (E) as following:

$$E = -1 + 10^{\left(\frac{-1}{slope}\right)}$$

Next, we calculated the efficiency percentage (% E):

$$\% E = (E - 1) \times 100$$

We considered a primer efficient at the respective temperature if the % E was between 90 and 110%.

For each RT-qPCR, a master mix was prepared, each well containing 5 µL of SYBR Select Master Mix (Applied Biosystems, Life Technologies), 0.125 µL of the specific forward and 0.125 µL of the specific reverse primer (10 µM) and 3.75 µL of water, reaching a total of 9 µL. Then, 1 µL of cDNA was added per well, or 1 µL of water in the non-template controls. The 7500 Fast Real-Time PCR System thermocycler (Applied Biosystems) was used for all the reactions, and the program was the following: 2 min at 50º C, 2 min at 95º C; 40 cycles at 95º C for 15 s, 60º C for 1 min; the melt curve stage occurs at 95º C for 15 s, 60º C for 1 min, 95º C for 30 s and 60º C for 15 s.

Table 2. **Primers used for RT-qPCR.**

Primer name	Sequence (5'-3')
SpikeHu_F	ATGACTCGAGTCTAGAACAAAA
SpikeHu_R	AGCTCGAGTCTAGAACAA
SpikeOm_F	GGTGTTCGCCCAAGTGAAAC
SpikeOm_R	CGA ATT CCT CGA GTT T
GAPDH-F	TGGAGTCTACTGGCGTCTT
GAPDH-R	TGTCATATTTCTCGTGGTTCA
18S F	GCAGAATCCACGCCAGTACAAGA
18S R	CCCTCTATGGGCCCGAATCTT
eGFP_F	GAGCTGAAGGGCATCGACTT
eGFP_R	TTCTGCTTGTCGGCCATGAT
MCL1_F	ACCGGCGCCAAGGACACAA
MCL1_R	AAGGCCGTCTCGTGGTTGC

5. Actinomycin D treatment

In order to study Spike's mRNA stability, transcription inhibition with Actinomycin D (Sigma) was performed. Firstly, HeLa cells were cultured in complete DMEM, in 24-well plates, and transfected with the Wuhan and Wuhan_seq3 3X plasmids, following the same protocol described in section 2.3. with Lipofectamine 2000, being incubated at 37º C, with 5% CO_2 overnight. In the next morning, the medium was changed to complete DMEM with Actinomycin D, at a final concentration of 5 µg/mL, and cells were treated for 2, 4, 6, 8, 24, 48 and 72 h. During the time course of each experiment, cells were kept in the incubator. In the 0 h timepoint, no Actinomycin D was added. To stop the experiment at the defined timepoint, the plates were removed from the incubator and total RNA extraction was performed, as described in section 4.1. mRNA levels were measured by RT-qPCR, using primers to amplify Wuhan Spike's and GAPDH mRNAs (**Table 2**). *GAPDH* was used as a negative control, due to its constant levels upon Actinomycin D treatment. *MCL1* was also used as a control since its mRNA half-life is already known (118), and it was tested to ensure that the protocol was effective.

6. *In vitro* transcription

6.1. Phenol/chloroform purification

To purify nucleic acids throughout the several steps of *in vitro* transcription, samples were firstly diluted to a final volume of 200 µL, with nuclease-free water. Phenol/chloroform with isoamyl alcohol (25:24:1) was added to the samples in a 1:1 ratio, that were then vortexed and centrifuge for 5 min at 13000 g. The aqueous phase,

where nucleic acids are present, was then retrieved to a nuclease-free Eppendorf, and 200 µL of chloroform was added. Samples were vortexed, and centrifuge for 5 min at 13000 g, after which this new aqueous phase was again retrieved to a new Eppendorf. To this, sodium acetate (AcNa, 3 M, pH 5.2) was added in a 1:10 ratio, plus 2 volumes of cold absolute ethanol and 1 µL of glycogen. Samples were allowed to precipitate overnight, at -20º C in the case of the linearized plasmids for *in vitro* transcription, or at -80º C in the case of the obtained mRNA (for mRNA, all the centrifugations mentioned above were performed at 4º C). After this, samples were centrifuge at 4º C for 15 min, at 13000 g, the supernatant was discarded, and the pellet was washed with 500 µL of 70% cold ethanol, by centrifugation at 4º C, for 5 min, at 13000g. Supernatant was discarded, the pellet was air-dried for around 5 to 10 min and then eluted in the desired volume of nuclease-free water. Nucleic acid concentration was once again measured using the NanoDrop 1000 Spectrophotometer, also used to evaluate the quality of the obtained mRNA.

6.2. T7 *in vitro* transcription

The HighYield T7 mRNA Synthesis Kit ($me^1\Psi$-UTP; Jena Bioscience), was used for synthesis of N1-Methylpseudouridine-modified mRNA, accordingly to the manufacturer's procedure, with minor alterations. To produce a mature mRNA, a m^7G(5')ppp(5')G Cap Analog (Ambion®, Applied Biosystems) was added in the same concentration recommended for the nucleotides. For the mRNA to incorporate the Cap analog efficiently, GTP is added in a 1:4 ratio of the Cap Analog.

Initially, 3 µg of Wuhan_polyA and Wuhan_seq3 3X_polyA plasmids were linearized with the restriction enzyme *EcoR*I (NEB), in 1x *EcoR*I Buffer (NEB), for 2 h at 37º C. A small amount of the reaction was ran in an agarose gel to confirm that the plasmid had been completely linearized, followed by template purification with phenol/chloroform as in section 6.1.

Briefly, all components were combined at room temperature, in a final volume of 20 µL, with the following order: nuclease-free water, 2 µL of HighYield T7 Reaction Buffer, 2 µL of DTT (100 mM), CTP, ATP, N^1-MethylpseudoUTP and Cap Analog to a final concentration of 7.5 mM, GTP to a final concentration of 1.875 mM, 1 µg of template DNA, and finally, 2 µL of HighYield T7 RNA Polymerase Mix. The mixture was vortexed and spun down, and then incubated at 37º C for 2 h, in the dark. After this, a purification with Illustra™ ProbeQuant™ G-50 Micro Columns (GE Healthcare) was performed. Afterwards, 1 µL of Recombinant DNase I (Roche) was added to the solution, in 1x DNase I recombinant RNase-free incubation buffer (Roche), and

samples were incubated for 40 min at 37º C, followed by a phenol/chloroform purification.

mRNA was quantified using NanoDrop 1000 Spectrophotometer and sent for the i3S Genomic Core Facility for analyses with the 2100 BioAnalyzer (Agilent Technologies). Shortly, the Agilent RNA 6000 Pico assay was performed, that consists in capillary gel electrophoresis on a chip, separating samples by size. mRNA is heat denatured (2 min, at 70º C), as well as an RNA ladder, and samples are run simultaneously. Bioanalyzer allows the visualization of the molecules, to confirm the length of the mRNA, and evaluation of the quality of the RNA, by analyses of the degraded products present.

7. Statistical analyses

In graphs containing more than one experiment, data is presented as mean +/- standard deviation (SD). The analyses for statistical differences were performed either with one-way ANOVA corrected for multiple comparisons using the Dunnett multiple comparisons test, or with two-tailed unpaired *t* test. Only p-values < 0.05 were considered statistically significant, where * $p<0.05$, ** $p< 0.01$, and ns stands for not significant. All graphs were done using GraphPad Prism version 8.01.

Results and Discussion

1. Cell transfections studies

1.1. iPLUS seq3 3X leads to an increase in Spike's production in HeLa cells

The COVID-19 pandemic was a turning point for mRNA vaccines, opening the possibility for these to be rapidly tested and placed on the market. In fact, until the end of 2019, there were 15 mRNA vaccine candidates on clinical trials, even though none of them had passed to Phase III trials. The BNT162b2 mRNA vaccine from Pfizer/BioNTech against SARS-CoV-2 was the first mRNA drug to ever be approved (119). An mRNA vaccine has 5 main components, a 5'Cap, a 3'polyA tail, an UTR both in the 3' and the 5', in addition to the coding region of the desired antigen (113). Several properties of these vaccines can still be improved, and further research is needed to fully understand their effects on a long term. For example, to date, the most widely vaccines used for immunization require a two-dose shot. It would be beneficial to increase the vaccine's efficiency in order to ensure that a single dose can be administered.

With the problematic associated with the COVID-19 pandemic and the rising of the mRNA vaccines in mind, we decided to evaluate if the subcloning of iPLUS downstream of Spike's cDNA, could lead to an improvement in important characteristics in these constructs, namely their stability and subsequent protein production. We also decided to test a variant of the iPLUS, iPLUS seq3 3X, that corresponds to the most conserved region of iPLUS (TTGTTTT) in tandem three times. As so, iPLUS and iPLUS seq3 3X were subcloned downstream of Spike's cDNA, from two different variants, the original and still used in most vaccines, Wuhan variant, but also the more recent and already dominant Omicron variant. We intended to evaluate if iPLUS could aid the improvement of the existing mRNA vaccines but also in the development of new mRNA vaccines containing the Omicron variant, which are only now starting to appear on the market, in order to increase people's immunization specifically directed to the variant that is now in circulation.

We started by constructing the plasmids necessary for the cell transfection assays (**Figure 9**) that will allow us to evaluate the influence of both iPLUS sequences, in both Spike mRNA levels and the corresponding protein expression.

For that, iPLUS and iPLUS seq3 3X were subcloned downstream of Spike's cDNA from both the Wuhan and the Omicron variant. The plasmids used had the pcDNA3.1

backbone and the cDNAs for Spike variants (SpikeHu and SpikeOm) and the iPLUS and iPLUS seq3 3X sequences were inserted, obtaining 6 different plasmids: SpikeHu/Om, SpikeHu/Om_iPLUS and SpikeHu/Om_seq3 3X. The plasmids contain a T7 RNA polymerase promoter, a cytomegalovirus (CMV) enhancer-promoter, the cDNA, in this case, Spike's cDNA from the Wuhan or the Omicron variant, and a PAS from the bovine growth hormone (bGH), as illustrated in **Figure 10.A**.

These plasmids were then used for a preliminary experience in HeLa cells. Results show that, for the Wuhan variant, the presence of iPLUS and iPLUS seq3 3X downstream of Spike's cDNA led to a slight decrease in the protein levels, while they do not significantly impact the mRNA levels, as we can see in **Figure 10.B.** In fact, regarding the mRNA levels, we observed a slight increase for iPLUS and a decrease for iPLUS seq3 3X. For the Omicron variant, on the other hand, both our sequences led to a marked decrease in both protein and mRNA levels, as we can see in **Figure 10.C**. The mRNA levels decrease by 75% upon the presence of iPLUS, when in comparison to the control, and decrease 50% in the presence of iPLUS seq3 3X, correlating with the levels of protein. Notably, comparing only the Spike's from the two variants without iPLUS, SpikeOm seems to be more expressed than SpikeHu (**Figures 10.B** and **10.C**).

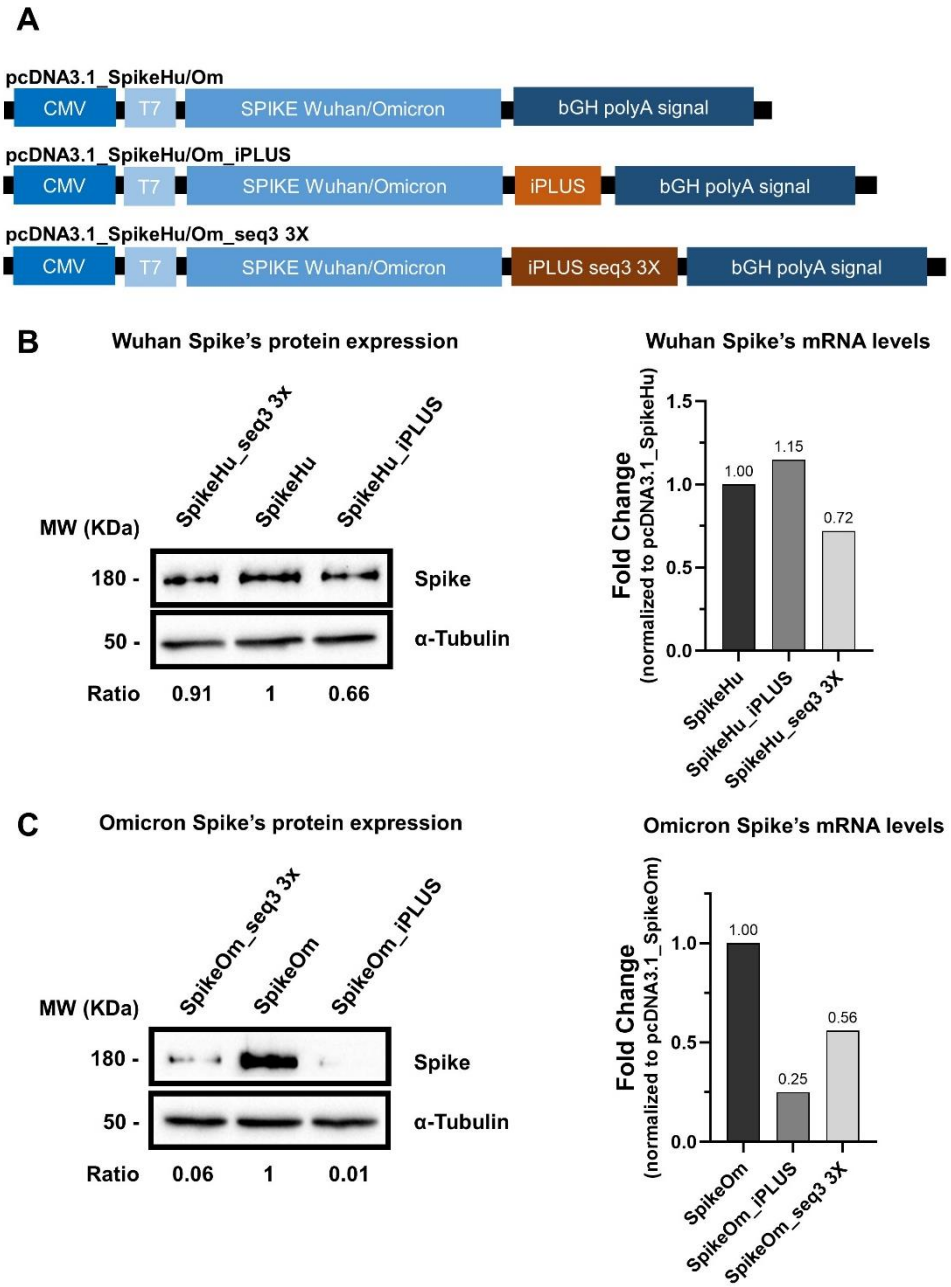


Figure 10 – iPLUS and iPLUS seq3 3X do not show an improvement in Spike's protein production or mRNA levels, after transfection of HeLa cells with the pcDNA™3.1 plasmids, in comparison to constructs lacking these sequences. A) Schematic representation of the pcDNA3.1 constructs, containing either Spike's cDNA from the Wuhan or the Omicron variant, and iPLUS or its variant, iPLUS seq3 3X. Additionally, the plasmids contain a T7 promoter, a CMV enhancer/promoter, and the bGH PAS. These plasmids were used for HeLa cells transfection in these preliminary experiments. **B)** Western Blot and Spike's mRNA levels show that the presence of iPLUS or iPLUS seq3 3X downstream of Spike from the Wuhan variant, does not improve Spike's protein production neither the corresponding mRNA levels. Wuhan Spike's protein production was evaluated by Western Blot, with α-Tubulin as a loading control. Quantification of the protein levels normalized to the loading control, were used to calculate the fold change, represented below each image as ratio, being normalized to SpikeHu. mRNA levels were measured by RT-qPCR, using the internal control *18S*, and fold change was then calculated by normalization to *SpikeHu*, being represented above each bar. n=1 **C)** For the Omicron variant, the presence of iPLUS and iPLUS seq3 3X leads to a strong decrease in both Spike's protein production and the corresponding mRNA levels. Omicron Spike's protein production was evaluated by Western Blot, also with α-Tubulin as a loading control. Quantification of the protein levels normalized to α-Tubulin, was used to calculate the fold change, represented below each image as the ratio, being normalized to SpikeOm. mRNA levels were measured by RT-qPCR, using the internal control *18S*, and fold change was then calculated using *SpikeOm* for normalization. The respective value is represented above each bar. n=1

The results obtained using the pcDNA3.1 plasmids and the iPLUS sequences did not seem promising in HeLa cells, but a second strategy already planned was chosen to continue the project. The pcDNA™3.4 TOPO™ had already been used in our lab, with the presence of iPLUS with good results. This was the plasmid used as a backbone in the experiments regarding the monoclonal antibody protein production, in ExpiCHO-S cells, where iPLUS insertion doubled the recombinant antibody expression (**Figure 6**). Besides that, pcDNA3.4 based vectors are normally used in the industry for recombinant protein production, being this another advantage in the use of these plasmid backbones.

As so, the next step was to subclone the cDNA from both Spike's variants into the pcDNA3.4 plasmids containing the iPLUS sequences by substituting the cDNA of the monoclonal antibody by either the SpikeHu or SpikeOm variants. These plasmids were named Wuhan/Omicron, Wuhan/Omicron_iPLUS and Wuhan/Omicron_iPLUS seq3 3X. The pcDNA3.4 constructs possess the full-length CMV immediate-early promoter/enhancer, the PAS from the herpes simplex virus (HSV) type 1 thymidine kinase (TK) and the already mentioned WPRE. The latter is present downstream of the gene of interest, in the 3'UTR, promoting the formation of a secondary structure, which increases gene expression. Of note, WPRE is not present in the pcDNA3.1 plasmid used in the previous experiments. iPLUS and iPLUS seq3 3X were thus subcloned downstream of the WPRE sequence, and upstream of the existing HSV-TK PAS in the pcDNA3.4 backbones. The sequence components of these plasmids are represented in **Figure 11.A** and **Figure 12.A**.

These plasmids were then used for the transfection of HeLa cells, in 4 independent experiments. The results presented in **Figure 11.B, C** and **D**, show that, for the Wuhan variant, the presence of iPLUS seq3 3X downstream of Spike's cDNA, leads to an increase in both Spike's protein production and mRNA levels, when in comparison to the plasmid without the iPLUS sequences. For the protein expression, this increase is approximately 1.9-fold (**Figure 11.C**), while for the mRNA levels the increase is approximately 1.4-fold (**Figure 11.D**). It is thus clear that iPLUS seq3 3X leads to a significant increase in both the mRNA and protein levels of the Wuhan variant.

On the contrary, the presence of the iPLUS sequence itself downstream of Spike's coding sequence, led to a decrease in both the protein and mRNA levels (**Figure 11.C** and **11.D**). This decrease was more accentuated on the mRNA levels, with a 40% reduction when compared to the control, but it is also observed for the protein levels, as shown in the Western Blot panel on **Figure 11.B**, although some variability in the

results leads to lack of statistical significance, which could be because these are four completely independent experiments (n=4). Nevertheless, the mean of the fold ratio values for the protein expression is approximately 0.45, therefore the decrease we are observing in both the mRNA and protein levels are correlated.

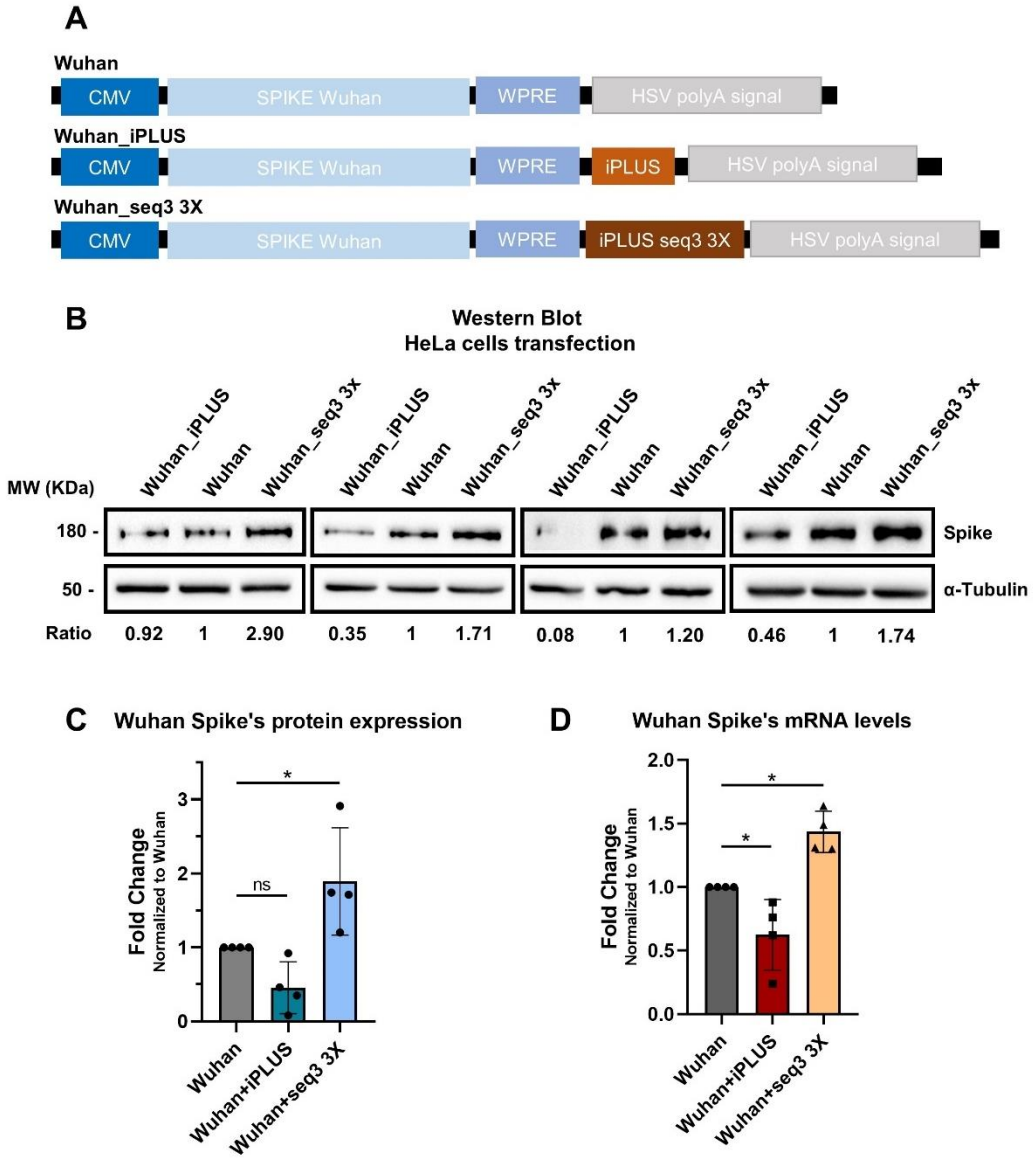


Figure 11 – iPLUS seq3 3X increases Wuhan Spike's protein production and mRNA levels in HeLa cells, in comparison with the plasmid that lacks this sequence. A) Schematic representation of the pcDNA3.4 constructs, named Wuhan for short, containing Wuhan Spike's cDNA alone or with the iPLUS or its variant, iPLUS seq3 3X. Additionally, these plasmids contain a full-length CMV promoter, WPRE and the HSV TK PAS. iPLUS sequences are located between the WPRE and the PAS. **B)** Spike's protein production was evaluated by Western Blot, represented in the panel. Four independent experiments (n=4) were performed and used for Spike's protein quantification. α-Tubulin was used as a loading control. Protein quantification was calculated using the loading control for normalization, followed by fold change analyses, normalized to the Wuhan plasmid. The fold change values are represented below each image and plotted in **C)**. **C)** Protein quantification levels are plotted for each condition, for four independent experiments. **D)** mRNA levels were measured by RT-qPCR, also for four independent experiments (n=4). The mRNA levels were normalized to the internal control *18S*, and fold change was then calculated in relation to the Wuhan sample. Data are shown as mean +/- standard deviation (SD) for four independent experiments, both for protein quantification and mRNA levels. Statistical significancy was determined by one-way ANOVA corrected for multiple comparisons using the Dunnett multiple comparisons test where: *=p<0.05 and ns means not significant.

For the Omicron variant, the results obtained are once again different from the ones obtained with the Wuhan variant. In fact, as it can be observed in **Figure 12.B, C** and **D**, iPLUS and iPLUS seq3 3X sequences, when subcloned downstream of Omicron Spike's cDNA, do not have an impact neither in the protein production, nor in the mRNA levels, evidenced by the lack of statistical significance.

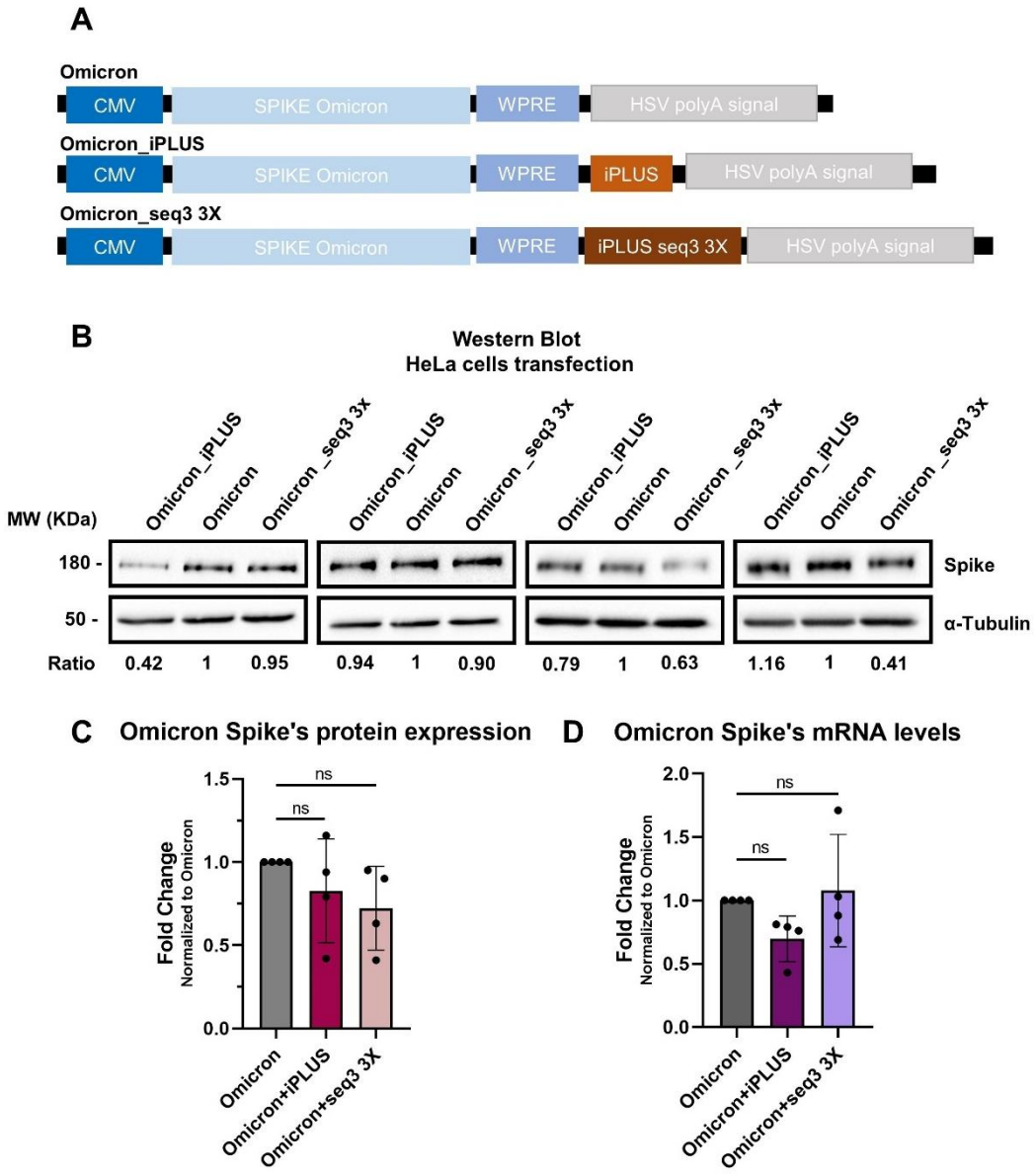


Figure 12 – iPLUS and iPLUS seq3 3X, when subcloned downstream of Omicron's Spike cDNA, do not impact the respective protein production or mRNA levels. A) Schematic representation of the pcDNA3.4 constructs, containing Omicron Spike's cDNA and iPLUS or its variant, iPLUS seq3 3X. Additionally, these plasmids contain a full-length CMV promoter, WPRE and the HSV TK PAS. **B)** Western Blots for Spike's protein production evaluation. Four independent experiments (n=4) were performed and used for Spike's protein quantification. α-Tubulin was used as a loading control. Protein quantification was calculated using the loading control for normalization, followed by fold change analyses, normalized to the Omicron plasmid without iPLUS sequences. The fold change values are represented below each image and plotted in **C). C)** Protein quantification levels plotted for each condition, for four independent experiments. **D)** mRNA levels were measured by RT-qPCR, also for four different experiments (n=4). mRNA levels were normalized to the internal control 18S, and fold change was then calculated in relation to the Omicron's condition. Data shown in C and D correspond to the mean +/- SD of n=4 for both protein quantification and mRNA levels. Statistical significancy was determined by one-way ANOVA corrected for multiple comparisons using the Dunnett multiple comparisons test where: ns means not significant.

Curiously, the only difference between the Wuhan and Omicron plasmids is the Spike cDNA (**Figure 11.A** and **12.A**). Our results clearly show that the iPLUS sequence does not produce the same effect for the Wuhan and Omicron Spike variants, and the same holds true for iPLUS seq3 3X, since no increase is observed with the Omicron_iPLUS seq3 3X plasmid. The same results were obtained with the initial plasmids pcDNA3.1 (**Figure 10**), where we see that iPLUS and iPLUS seq3 3X do not impact Wuhan's protein production or mRNA levels in comparison with the plasmid that lacks this sequence (**Figure 10.B**), but for Omicron's Spike there is a significant decrease (**Figure 10.C**). This points to the hypothesis that the coding region plays a role in controlling Spike mRNA and protein levels. We know that these sequences only differ by 37 mutations, corresponding to the same protein only with slight changes specific to its conformation, which makes these results even more surprising. Specifically, the two Spike sequences used in this project share 84% of similarity (analysis done using Needleman-Wunsch alignment of two sequences, in BLAST® from NCBI).

Although most research regarding *cis*-acting elements is focused on the UTRs, some examples have also been described within the coding region. The coding sequence does not simply code for the amino acids and several other vital features are present (126). Binding sites for miRNAs and RBPs have been described within the coding sequence of some mammalian genes, affecting the respective mRNA stability and subsequent translation, in an interconnected way (120-122). As a matter of fact, several *in silico* analysis have reported that within SARS-CoV-2 genome several binding sites for some human miRNAs are present, providing an antiviral mechanism (123, 124). These binding sites are not only predicted to be present in the 3'UTR, but also, specifically along the coding sequence of several viral proteins, including Spike (125). We can speculate that maybe the sequence differences between Wuhan and Omicron Spike's cDNA can generate different binding sites for regulatory *trans*-acting factors, which may explain the differences observed between Wuhan and Omicron with iPLUS seq3 3X. In fact, a preliminary *in silico* analyses showed that there are different miRNAs target sites in the cDNA of Wuhan and Omicron Spike's used in this study. Additionally, it is also known that RNA virus present RNA secondary and even tertiary structures, important for their life cycle regulation, and SARS-Cov-2 is one of these cases (88). Perchance, different secondary structures are present in these mRNAs, leading to differences in mRNA stability and in the translation efficacy.

All these factors can be possible explanations for the differences we are seeing from Wuhan and Omicron Spike's cDNA, but in order to clarify these results several additional experiments would be required, such as an *in silico* analysis of possible

mRNA secondary structures, binding sites for RBPs, miRNAs or mRNA regulatory sequences, and subsequent *in vitro* characterization.

It is interesting that we are observing these contrasting results between Spike from the Wuhan and Omicron variants, as they reveal that new mechanisms for Spike expression are in place, which need to be further investigated to understand the molecular mechanisms of disease and to produce improved mRNA vaccines. Additionally, we clearly demonstrated that in HeLa cells, iPLUS seq3 3X increases Wuhan Spike's protein expression, by 1.9-fold, and the mRNA levels, by 1.4-fold. The effect observed for iPLUS seq3 3X in HeLa cells is extremely interesting from the point of view of a possible application to the mRNA vaccines. This may allow for an increase in the antigen production upon immunization, improving the efficacy of this process.

1.2. In ExpiCHO-S cells, iPLUS seq3 3X also increases Spike production

It is important to guarantee that the results obtained in HeLa cells for iPLUS are also maintained in other cellular models. As we are studying *cis*-regulatory sequences, and even though we still do not fully understand its mechanism of action, we know that RBPs, including Heph, bind to these sequences in *Drosophila*, leading to an increase in mRNA levels (34). We cannot rule out the possibility that the concrete cellular environment encountered in a specific cell line, regarding its proteome and transcriptome, can impact the way that these sequences affect protein production and mRNA stability.

As so, it was also our objective to evaluate the effect of these iPLUS sequences, on Spike's protein production and mRNA levels, in another cell line from a different species. The ExpiCHO-S cell line was chosen as a model for this project due to its common use in the industry for recombinant protein production (127). The ExpiCHO-S™ cell line is derived from the Chinese Hamster Ovarian CHO-S cell line, allowing for a higher protein production, and is grown in a serum-, protein-, chemical- animal-origin-free medium and can be maintained at high densities, in suspension. Possible partners in the future may be interested in learning how our plasmids function in this mammalian model, since CHO cells have become a widely used platform for the initial phases of screening in protein production (128). In fact, besides our focus on mRNA vaccines, we cannot exclude the possible application of our developed plasmids in Spike's protein production with higher yields. Recombinant Spike protein can be applied in the study of the biology of this virus, for subsequent production of specific antibodies, that can be

used in therapeutics, or to assemble clinical diagnostic tools, such as serological tests (129-131).

With that purpose, we transfected the ExpiCHO-S™ cell line with the Wuhan and Omicron plasmids used previously in the HeLa cells experiments (**Figure 11.A** and **12.A**). As it can be observed in **Figure 13.A, B** and **C**, we obtained with the ExpiCHO-S cell line the same results observed in HeLa cells, where iPLUS seq3 3X is able to promote an increase in both Spike's mRNA levels and respective protein production.

It can also be observed that the two-band profile displayed in the Western Blots for the Wuhan Spike protein in ExpiCHO-S cells, is very different from the single band encountered in HeLa cells. This is probably due to the time course of the experiment. For HeLa cells, the period between transfection and collection of the cells is 48 h, but for ExpiCHO-S cells the transfection protocol requires 9 days in culture until collection. It is known that Spike can be present on Western Blots as a full-length Spike band, of 180 kDa, or appears as a 90 kDa band, corresponding to its cleaved subunits (131, 132). As so, the 9 days period possibly allows for this cleavage to occur, resulting in the appearance of two different Spike bands. Of note, we are evaluating two different mammalian systems, the ExpiCHO-S cells derived from Chinese hamster and the HeLa cells derived from human, which could also explain the differences in the Spike protein profile.

For iPLUS seq3 3X, we observe a significant increase in Spike's protein levels for the full-length 180 kDa protein (**Figure 13.A**), where the increase is of ~2.5-fold (**Figure 13.A**; left), whereas this increase is not statistically significant for the cleaved band of 90 kDa (**Figure 13.B**; right). This is due to the variability between the three independent experiments (n=3) and that this process of protein cleavage is probably much more variable than the production of the full-length protein. Even though there is no statistical significance, we can observe that all ratio values (ratio down) presented in **Figure 13.A** point to an increase in the 90 kDa band in the iPLUS seq3 3X plasmid. For the mRNA levels, we observe a 1.6-fold increase in Spike's levels, when iPLUS seq3 3X is present (**Figure 13.C**). As so, it is very clear that indeed iPLUS seq3 3X induces an increase in both Spike's protein and mRNA levels, when compared to the Wuhan control. It is not uncommon that the protein levels present the stronger increase in ExpiCHO-S cells, since these cells are specifically designed to produce recombinant proteins.

On the other hand, the original iPLUS sequence has the opposite effect, promoting a strong decrease in protein production (approximately 70% for both the full-length and cleaved Spike; **Figure 13.B**), while not impacting the mRNA levels (**Figure 13.C**),

when compared with the control, Wuhan. We can conclude that iPLUS is clearly not promoting the same effects as iPLUS seq3 3X, neither in HeLa or ExpiCHO-S cells, and unexpectedly it does not increase Spike protein production or mRNA levels.

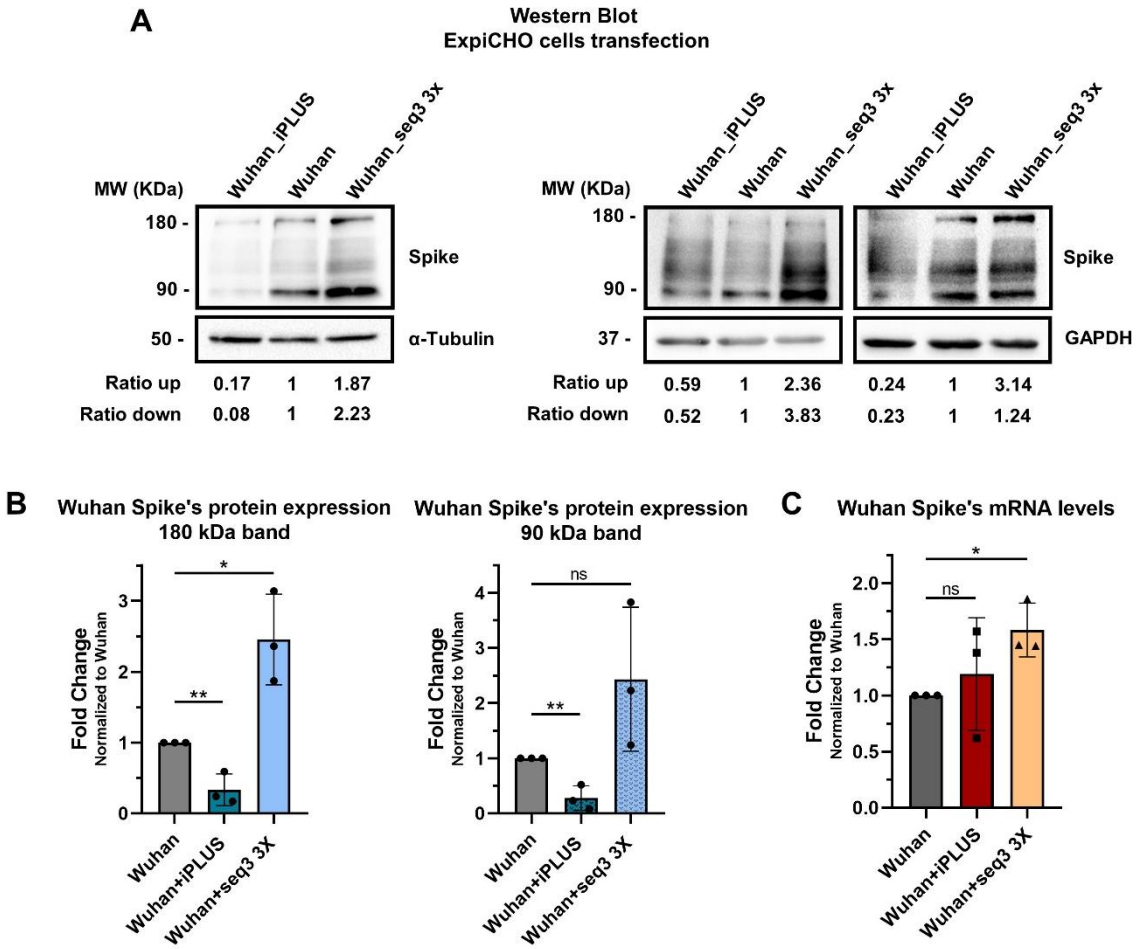


Figure 13 – iPLUS seq3 3X increases Wuhan Spike's protein production and mRNA levels, in comparison with the Wuhan plasmid that lack this sequence, in ExpiCHO-S cells. A) Spike's protein production was evaluated by Western Blot, in three independent experiments (n=3). Two different molecular weight bands appear, one related to the 180 kDa full-length Spike protein, and another corresponding to a 90 kDa Spike's cleaved subunit. α-Tubulin or GAPDH were used as loading controls. Protein quantification was calculated using the respective loading control for normalization, followed by fold change analyses, normalized to the Wuhan plasmid band of each molecular weight. The fold change values are represented below each image and plotted in **B)**, with ratio up referring to the 180 kDa band, and ratio down referring to the 90 kDa band. **B)** Graphs represent protein quantification of Spike, where the left graph has the ratio values corresponding to n=3 for the upper band, of 180 kDa, and the right graph has the ratio values for the lower band, of 90 kDa. **C)** mRNA levels were measured by RT-qPCR (n=3). The internal control *ACTB* was used for normalizing mRNA levels, and fold change was calculated in relation to the Wuhan sample. In all graphs, results are plotted as mean +/- SD of the three experiments. Statistical analysis was performed using a two-tailed unpaired t test, where: *=p<0.05, **=p<0.01 and ns corresponds to not significant. The analysis was done specifically between the Wuhan and Wuhan_iPLUS samples, or Wuhan and Wuhan_seq3 3X, even though both results are plotted on the same graph. The plasmids used in this experiments are represented in **Figure 11.A**.

For the transfection of our set of plasmids containing the Omicron variant on ExpiCHO-S cells, the results also showed a similar tendency to the ones found with HeLa cells, thus only one experiment was performed in these cells. In fact, as it can be observed in **Figure 14.A** and **B**, the presence of both iPLUS and iPLUS seq3 3X leads to a decrease in both Omicron Spike's protein and mRNA levels. For the mRNA levels, a decrease of 50% for iPLUS and 40% for iPLUS seq3 3X was observed (**Figure 14.B**). Regarding Omicron Spike's protein production, a decrease between 70 and 80% is observed in all bands (**Figure 14.A**). For the full-length Omicron Spike (180 kDa band), bands cannot be quantified since they appear very fade, however, we can visually recognize that a decrease is present regarding both iPLUS and iPLUS seq3 3X, when compared to the Omicron control sample. Here, another band appears with more intensity than in the Western Blots regarding the Wuhan plasmids, with a molecular weight of approximately 110 kDa. As a matter of fact, it is known that an important difference between Spike from the Wuhan and the Omicron variant is related to their conformation, with Omicron presenting a much more compact structure (97). Moreover, Omicron presents mutations in the vicinity of the cleavage site, with an impact in its cleavage by cellular proteases (133, 134). It can be the case here that these differences are reflected in the distinct profiles visualized. Nevertheless, it is clear that both these sequences are impacting negatively the production of the respective mRNA and protein.

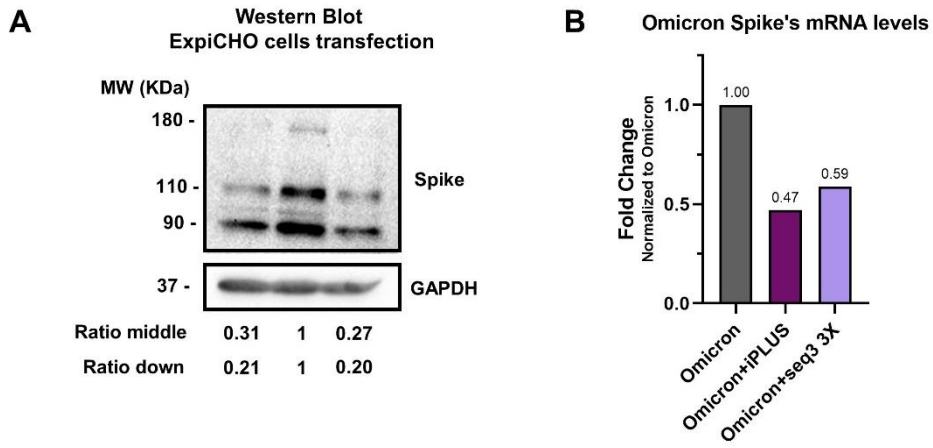


Figure 14 – In Spike's Omicron variant, both iPLUS and iPLUS seq3 3X lead to a decrease in Spike's protein production and the respective mRNA levels. A) Spike's protein production was evaluated by Western Blot, in ExpiCHO-S cells. GAPDH was used as a loading control. Protein quantification was calculated using the loading control for normalization, for the different molecular weight bands. This was followed by fold change analyses, normalized to the respective Omicron control sample band of each molecular weight. Fold change values are represented below each image, with ratio middle corresponding to the 110 kDa band and ratio down referring to the 90 kDa band, both resulting from Spike's protein cleavage. **B)** *Spike* mRNA levels were measured by RT-qPCR, for one experiment. mRNA levels were normalized to the internal control *ACTB*, and fold change was calculated in relation to the Omicron sample.

1.3. Efficiency of transfection remains constant, in both HeLa and ExpiCHO-S cells, for the Wuhan variant

Although the transfection protocol used in each cell line, for all conditions, is kept the same, a concern raised during the course of the experiments was the possibility that different transfection efficiencies could be associated with the different plasmids. If this was happening, the interesting results we were obtaining could be due to differences in transfection efficiency and not to the primary effect of each sequence itself. Different strategies could have been employed to solve this problem, such as a new cloning to insert a reporter gene in our plasmids, providing us a bicistronic plasmid, with a readout specific for each cell that was transfected. This could then be use for normalization of the results. Another strategy, which was not done due to time constrains, is to select transfected cells with Geneticin™. As a matter of fact, pcDNA3.4 expression vectors contain the resistance gene (*neo*) for this antibiotic. By performing the selection of cells, we would only be evaluating the ones that were in fact transfected with our plasmids, giving more accurate results. The strategy followed in order to try to bypass that issue, was the co-transfection of the Wuhan and Omicron sets of plasmids, with a GFP reporter plasmid already used in the lab, named pCS2_eGFP, that contains the coding sequence of eGFP under the control of a CMV promoter. By doing this, we expected to see a constant level of eGFP between all conditions, demonstrating that the transfection protocol is equally effective, having eGFP as a readout.

Due to time constrains, only a preliminary experiment was done for each cell line, but as observed in **Figure 15**, eGFP protein levels remain relatively constant in most of our experimental conditions. For HeLa cells, this experience was performed for both the Wuhan and the Omicron set of plasmids, while for ExpiCHO-S, only the Wuhan plasmids where used. For the Wuhan variant, in both HeLa and ExpiCHO-S, we can visually see in the Western Blots, presented in **Figure 15.A and C**, that the levels of eGFP protein are relatively constant between the three conditions, and this is also reinforced by the protein quantification of eGFP (Ratio eGFP below the Western Blots).

Moreover, the tendency of iPLUS seq3 3X leading to an increase in Spike's protein production is maintained, as shown in the Western Blots (ratio Spike) in **Figure 15.A** and **C**. The increase with iPLUS seq3 3X in Wuhan Spike's protein production reaches even two-fold for the 90 kDa band in ExpiCHO-S cells, as well as in HeLa, where the increase is approximately 1.9-fold, as observed previously in the experiments without co-transfection.

However, when we look at the results of the Omicron set of plasmids co-transfected with pCS2_eGFP, the values regarding protein expression of eGFP, presented in **Figure 15.B**, are more variable, especially in the Omicron_seq3 3X condition. This can indicate that we have lower transfection efficiencies in the Omicron_seq3 3X condition, even though we do not think that if the efficiencies were similar, it would be sufficient to observe an increase in Spike levels in comparison to the control. Nevertheless, further experiments are required, especially for the Omicron plasmids.

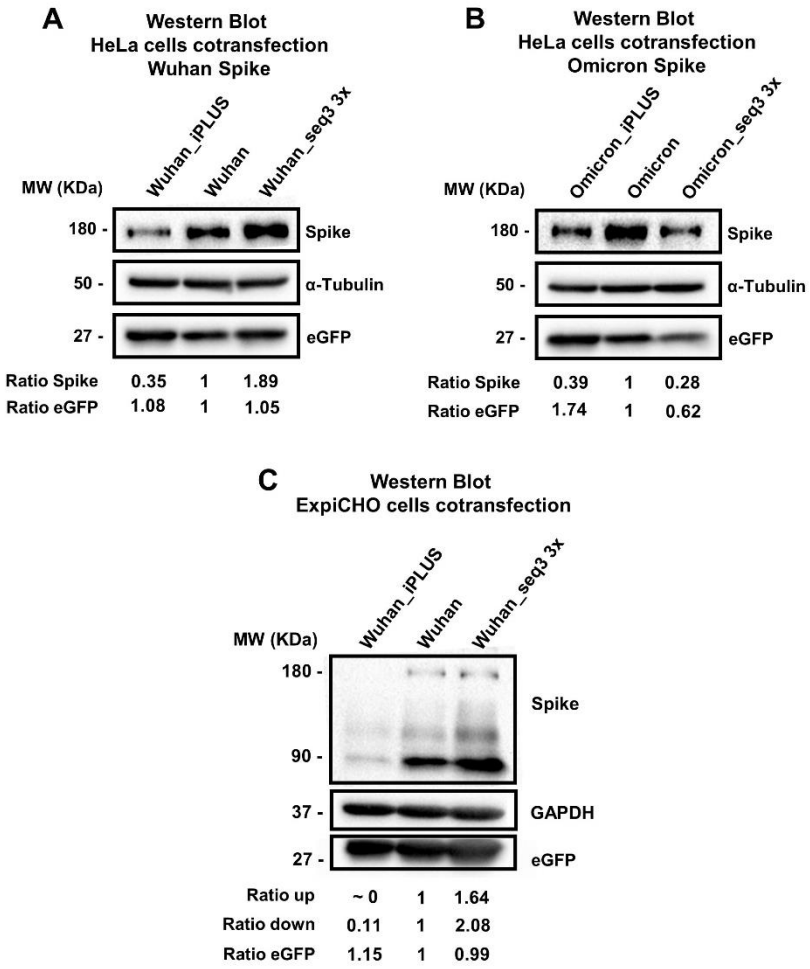


Figure 15 – eGFP protein expression is kept constant upon Wuhan Spike's plasmids co-transfection with pCS2_eGFP, whilst maintaining the iPLUS seq3 3X increase in protein production, in both HeLa and ExpiCHO-S cell lines. A) HeLa cells co-transfection with the Wuhan set of plasmids (control, iPLUS and iPLUS seq3 3X) and pCS2_eGFP, showed that eGFP protein expression is maintained relatively constant. For the Western Blot analysis, α-Tubulin was used as a loading control. Protein quantification, for both Spike and eGFP, was normalized to the loading control, and fold change was then calculated in relation to the Wuhan sample. **B)** HeLa cells co-transfection with the Omicron plasmids (control, iPLUS and iPLUS seq3 3X) and pCS2_eGFP showed a higher variation regarding eGFP, protein levels. For Western Blot analysis, α-Tubulin was used as a loading control and also for normalization of the values upon the quantification of Spike's and eGFP's protein levels. Fold change was then calculated using the Omicron control condition for normalization. **C)** ExpiCHO-S cells were co-transfected with the Wuhan plasmids and pCS2_eGFP, showing a similar trend as HeLa, with eGFP protein expression remaining constant between all three conditions. In here, GAPDH was used as a loading control for the Western Blot analyses and fold change was calculated as above. In all Western blots, the fold change values are represented as the ratios below each image, and in the case of ExpiCHO-S experiment, ratio up refers to the full-length Spike protein and ratio down refers to the 90 kDa cleaved subunit of Spike. Results are shown for one single experiment (n=1).

With all the results presented until this point, we have demonstrated that: 1) our sequences, iPLUS and iPLUS seq3 3X, when subcloned downstream of Spike cDNA from the Omicron variant, do not improve neither its protein production nor its mRNA levels; 2) iPLUS, when subcloned downstream of the Wuhan Spike cDNA, also does not impact Spike's protein or mRNA levels, actually leading to a strong decrease in protein production in ExpiCHO-S cells, reaching 80%; 3) on the contrary, iPLUS seq3 3X, that corresponds to the most conserved region of iPLUS in tandem three times, leads to an increase, that can go up to 2-fold, in both mRNA levels and protein production, when subcloned downstream of Spike's cDNA from the Wuhan variant. Obtaining this result in both HeLa and ExpiCHO-S cell lines reinforced the fact that this iPLUS variant sequence can be a valuable asset for Spike's recombinant protein production or mRNA vaccines.

It is also interesting to compare the first results regarding iPLUS seq3 3X effect obtained with the pcDNA3.1 plasmids (**Figure 10**), to the results obtained with the pcDNA3.4 plasmids (**Figures 11** and **13**). For the Wuhan variant, iPLUS seq3 3X did not cause any significant effect in our preliminary results with the pcDNA3.1 plasmids, neither in the protein nor in the mRNA levels (**Figure 10.B**). Contrastingly, for the pcDNA3.4 plasmids and in two different models, iPLUS seq3 3X leads to an increase in both Spike's protein expression and mRNA levels (**Figure 11.C** and **D**, and **Figure 13.B** and **C)**. This can be explained in at least two ways, concerning the differences between these plasmids:

1) the specific polyA signal (bGH for the pcDNA3.1's and HSV TK for the pcDNA3.4's) is impacting mRNA processing and consequent mRNA and protein levels. In fact, it is known that polyadenylation of an mRNA is crucial for its nuclear export, stability, and consequent translation, as was discussed in the introduction. This will obviously impact gene expression (135), so it can be reasonable to believe that this differences can be impacting our results. Some studies have addressed these effects (136, 137), but recently, Wang *et al*., compared the effect of five different PAS, including bGH and HSV TK PASs, in transgene expression upon transfection of CHO cells. Interestingly, here they found that transient expression of eGFP was significantly decrease when the HSV TK PAS was presented, when compared to the bGH PAS (138). Nevertheless, this does not exclude that in the context of our plasmids, iPLUS sequences together with the PAS, have different effects.

2) the presence of the WPRE sequence is playing a role in the effect of iPLUS seq3 3X. The group had already data showing that iPLUS subcloned in the pcDNA3.4 vector presented good results in what concerns an increase in monoclonal antibody protein

production. Besides that, it is important to note that it has been shown that the WPRE can act synergistically with USEs (72), which may be the case here.
Further experiments would be needed to unravel the molecular mechanism involved in iPLUS-dependent regulation of Spike expression, starting with the removal of the WPRE sequence from the pcDNA3.4 plasmids, to evaluate if the results obtained would still be the same. The same goes for the PAS, which could be swapped in an attempt to understand if it played a role in our findings.

Regarding the different results from iPLUS and iPLUS seq3 3X with the pcDNA3.4 plasmid, some explanations can be hypothesized. The two sequences present many similarities, only differing in length by 7 nucleotides, with iPLUS being a 28-nucletide long sequence, and iPLUS seq3 3X having a length of 21 nucleotides. Nevertheless, iPLUS seq3 3X present different binding sites for *trans*-acting factors, in particular RBPs that typically need 6 nucleotides to bind to, created by the junction of the iPLUS most conserved region three times. Additionally, this sequence as a higher U-richness when compared to iPLUS. This may promote the binding of different RBPs, possibly explaining the different effects we are obtaining. Additionally, we cannot rule out the possibility of different mRNA secondary structures promoted by these sequences to originate contrasting outcomes regarding mRNA 3'end formation, mRNA stability, or translation efficiency. These are all known common regulatory effects from sequences present in the 3'UTRs, as it was our aim to investigate. It is still striking how these small, non-coding sequences, with many similarities, can impact so strongly and so differently in the production of both the mRNA and the protein for Spike.

With all of this in mind, the next step in this project was to evaluate the effect that iPLUS seq3 3X exerts in mRNA stability, giving its good results in the cell transfection assays with the Wuhan variant.

2. mRNA stability assays

2.1. iPLUS seq3 3X decreases the mRNA stability of an *in vitro* transcribed *Spike* mRNA

As iPLUS seq3 3X increases Spike's protein production and mRNA levels, we wanted to further understand the mechanism behind this observation. An increase in mRNA stability could explain the increase in both Spike's mRNA and protein levels. Higher stability would allow the mRNA to remain intact in the cell for longer periods, when compared to the mRNA lacking this sequence, promoting higher levels of this molecule, correlating with more protein translation. If this is the case, an increase in mRNA stability associated with iPLUS seq3 3X, would be extremely beneficial for mRNA vaccines. It is known that a crucial bottleneck of the mRNA vaccine industry is the inherent instability of this molecule, which in truth was one the motives that contributed to the delay in a more generalized use of this technology. Every component that can improve this point will be an advantage. This will not only contribute to the manufacturing point of view, but it will also be beneficial for people's immunization. A more stable mRNA will reach our cells more intact, also increasing the resulting production of the antigen, culminating in a stronger immune response.

Therefore, to assess the iPLUS seq3 3X effect on mRNA stability, we performed *in vitro* transcription of our two constructs of interest: the control containing only the Wuhan's Spike cDNA and a construct containing iPLUS seq3 3X sequence subcloned downstream of Wuhan's Spike cDNA. For the *in vitro* transcription assays, we needed a T7 RNA polymerase promoter, for the RNA polymerase to initiate transcription. For this reason, we used the pcDNA3.1 plasmids, SpikeHu and SpikeHu_seq3 3X for subsequent cloning steps, given they already contain the T7 promoter, Wuhan's Spike cDNA and iPLUS seq3 3X (**Figure 10.A**).

As it was discussed before, the polyA tail is crucial to produce a mature mRNA, suitable for transfection into cells or for use as a vaccine. Normally, for these applications, the desired length should be higher than 100 nucleotides. This addition could be done enzymatically, but the direct incorporation of the polyA sequence in the template is more suitable since it allows for more reproducible additions. A problem that arises from this is that long, repetitive sequences as this can destabilize the plasmid where they are subcloned (119). We decided to follow the strategy used by Pfizer/BioNTech in its COVID-19 mRNA vaccine, by adding a short linker in the polyA

tail, producing a sequence A_{30}(10 bp UGC linker)A_{70} (139), as represented in **Figure 16.A**. Besides this, we also apply modified nucleotides, specifically N^1-Methylpseudouridine, considered to be crucial in mRNA vaccines, since its usage instead of uridine decreases the immune response to the mRNA molecule being inject and is being employed in most of the COVID-19 vaccines already on the market (100, 111).

To include of the polyA tail downstream of Spike's cDNA or downstream of iPLUS seq3 3X, we ordered a plasmid containing the polyA tail with the linker (pUC57-Mini) and included, by PCR amplification from the pcDNA3.1 plasmids, the T7 promoter, Wuhan Spike's cDNA and the iPLUS seq3 3X (in one of the constructs). These plasmids were named SpikeHu_polyA and SpikeHu_seq3 3X_polyA and are represented in **Figure 16.A**.

After the templates linearization and preparation was concluded, the *in vitro* transcription protocol, using the HighYield T7 mRNA Synthesis Kit (me1Ψ-UTP), was performed, as represented in **Figure16.B**. Following this, the capped and polyadenylated mRNA obtained were evaluated recurring to the Bioanalyzer, to ensure that the two *in vitro* mRNAs produced were not degraded. In fact, as we can observe in **Figure 16.C** both *Wuhan* and *Wuhan_seq3 3X* mRNAs present the expected profile, with low degradation, and correct molecular weight (around 4000 bp for the Wuhan, and an additional slight increase for Wuhan_seq3 3X, due to the presence of our sequence). Since mRNA is so fragile and easily degraded, this quality check was very important to guarantee that the next experiment could be performed, in the right conditions.

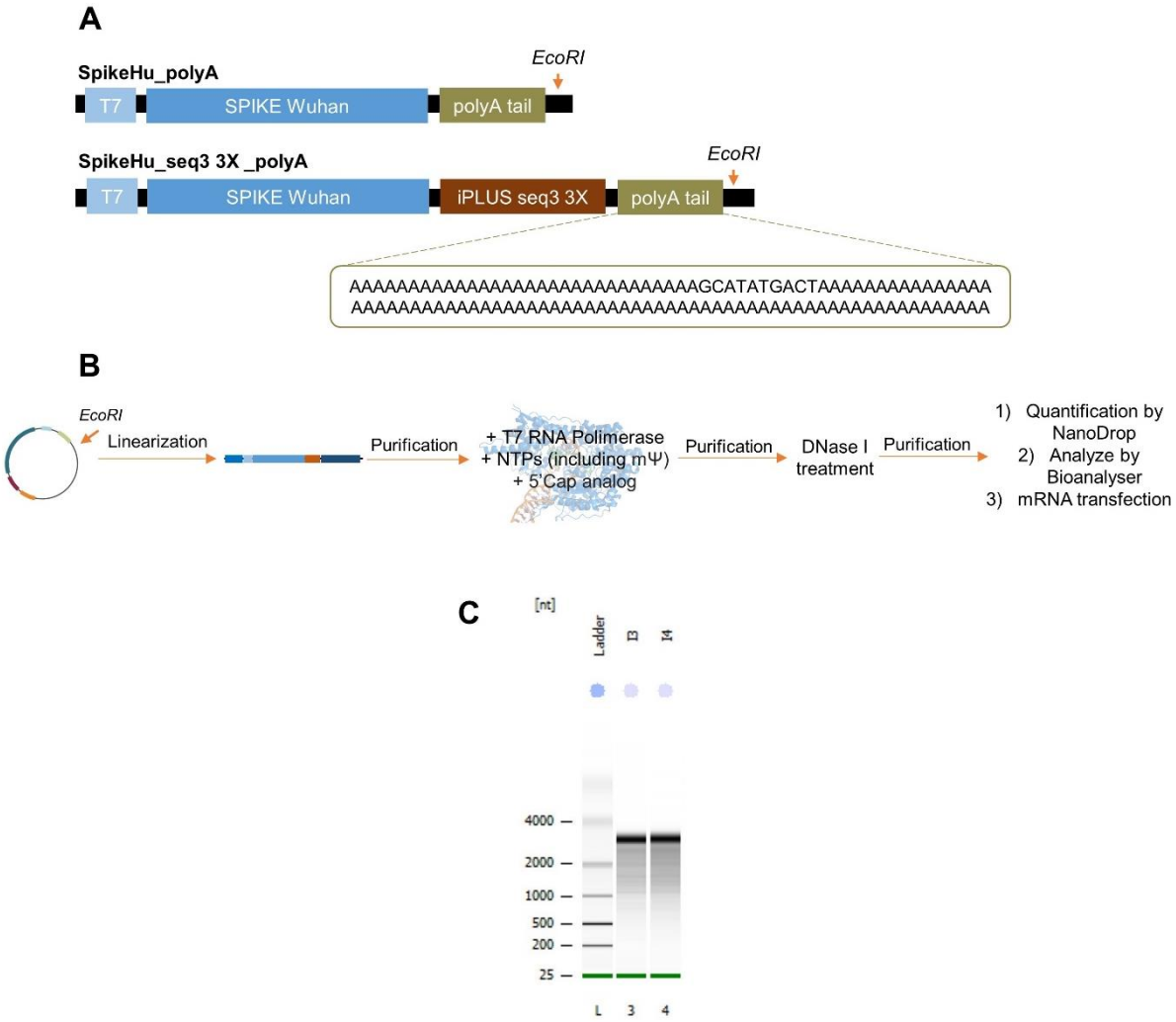


Figure 16 – *In vitro* transcription workflow. A) Schematic representation of the sequences present in the plasmids used as templates for *in vitro* transcription. A T7 promoter is necessary for *in vitro* transcription using the T7 RNA polymerase, upstream of Spike's cDNA. These plasmids have a polyA tail of defined length, subcloned downstream of Spike, in the control, or downstream of iPLUS seq3 3X. The polyA tail sequence is also represented, with the linker between adenosines indicated. The *Eco*RI restriction site indicated was used for plasmid linearization. **B)** *In vitro* transcription protocol. First, the plasmids need to be linearized with the restriction enzyme *Eco*RI. To remove any possible contamination, a phenol/chloroform purification was performed. In RNase free conditions, *in vitro* transcription was performed, using a T7 RNA polymerase and N^1-MethylpseudoUTP. A 5'cap was also added co-transcriptionally in order to produce a mature mRNA, using a 5'cap analog. A purification step, with G-50 columns was performed, followed by a DNase I treatment, to remove the DNA template. A final phenol/chloroform purification was carried out, and the concentrations of the obtained mRNA were measured in NanoDrop 1000. The mRNA *in vitro* transcribed was evaluated by Bioanalyser, and after those results, mRNA transfection experiments were carried out. **C)** Bioanalyser analysis of the *in vitro* transcribed mRNAs. I3 corresponds to the mRNA containing only Spike's coding region, while I4 corresponds to the mRNA containing Spike and iPLUS seq3 3X. The mRNA has the expected molecular size (around 4000 base pairs), that we can confirm by comparison to the RNA ladder ran alongside the samples, with I4 being slightly bigger, also as expected.

In order to evaluate the behavior of our *in vitro* transcribed mRNA, that resembles an mRNA vaccine inside a cellular environment we performed mRNA transfection assays in HeLa cells. As we wanted to obtain good transfection efficiencies, we tested three different transfection reagents from three different companies: iLipoP, Lipofectamine

2000 and JetMESSENGER. All have in common the formation of cationic carriers (either lipo- or polyplexes), containing the mRNA, for the following endocytosis by cells (140). One experiment in HeLa cells was performed using the companies' standard protocols. Cells were transfected with the mRNA and were incubated for 24 h in the case of iLipoP, and for 48 h in the case of Lipofectamine 2000 and JetMESSENGER. After collection of the cells and RNA extraction, the mRNA levels of *Spike* were evaluated by RT-qPCR. Since *Spike* mRNA does not exist basally on HeLa cells, we confirmed that we are indeed transfecting with success the mRNAs with these three reagents.

Unexpectedly, the results from all three transfections point to the same conclusion: iPLUS seq3 3X leads to a decrease in the levels of the *in vitro* transcribed *Spike* mRNA in comparison to the *Wuhan* control, suggesting that the iPLUS confers instability to the mRNA (**Figure 17)**. iPLUS seq3 3X seems to induce higher mRNA degradation, in comparison to the control, even though it was also able to promote a significant increase in Spike's protein production and mRNA levels, comparing to the same control, as observed in the previous plasmid transfections (**Figure 11.C** and **11.D** and **Figure 13.B** and **C**). iLipoP and Lipofectamine 2000 present more striking values than JetMESSENGER, but the global tendency is the same.

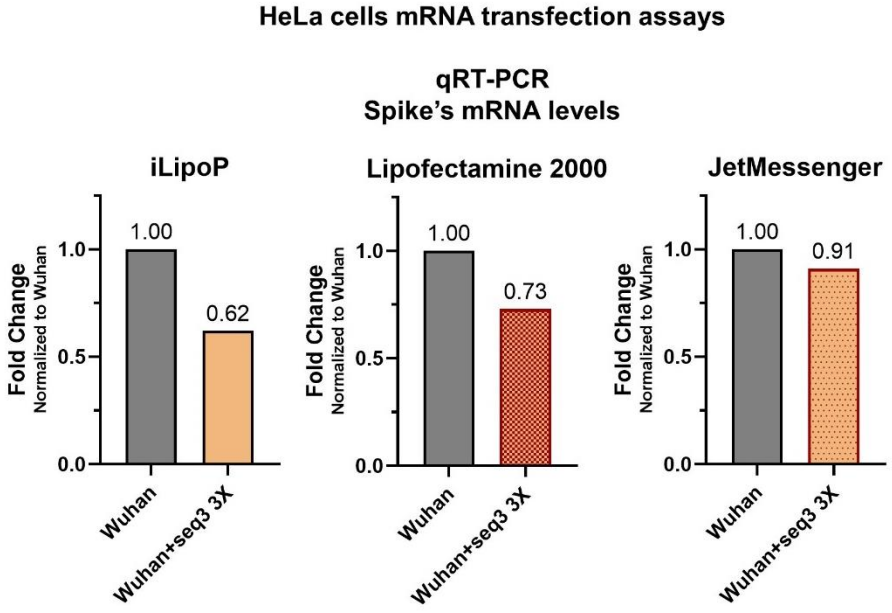


Figure 17 – iPLUS seq3 3X decreases the mRNA stability of an *in vitro* transcribed Spike mRNA in HeLa cells. The same amount of mRNA was transfected in the Wuhan and Wuhan+seq3 3X conditions, in each protocol. After 24h, for iLipoP, and 48h for Lipofectamine 2000 and JetMESSENGER, mRNA was extracted and then analyzed by RT-qPCR. All three different reagents tested point the same tendency, mRNA corresponding to the Wuhan+seq3 3X is present in lower quantities, when compared to the mRNA lacking this sequence, showing that iPLUS seq3 3X sequence downstream of Wuhan Spike's cDNA leads to a decrease in its mRNA stability. *18S* was used as the internal control, to normalize gene expression. Fold change was then calculated using the Wuhan condition for normalization.

As it was described above, the two mRNAs are similar, only differing in the presence of iPLUS seq3 3X, which apparently has an unexpected effect on mRNA stability. Next, and to confirm the results obtained with the transfection of the *in vitro* transcribed mRNA, Actinomycin D assays were performed, to further evaluate mRNA stability.

2.2. Actinomycin D assays reinforce the destabilizing effect that iPLUS seq3 3X exerts in *Spike*'s mRNA

The use of Actinomycin D (ActD) to evaluate mRNA stability has been widely used, since it allows the evaluation of mRNA decay rates. This compound is able to inhibit new transcription, due to its ability to bind to DNA, stopping the normal activity of the RNA polymerases. With this, new mRNA cannot be transcribed, and we can evaluate the outcome of the mRNA that remains in the cell or the mRNA half-life (141).

Since our gene of interest, *Spike*, is not normally present in cells, a first step of plasmid transfection was required for this experiment. For this, we used the Wuhan and Wuhan_seq3 3X plasmids (pcDNA3.4 backbone represented in **Figure 11.A**), since these were the ones that provided good results in the cell transfection assays. Cells were transfected with the plasmids and incubated overnight, and in the following day, ActD was added. RNA was extracted and expression was evaluated by RT-qPCR in the following timepoints of ActD incubation: 2, 4, 6, 8, 24, 48 and 72 hours, in addition to a timepoint of 0 h, where no ActD was added, to assess the basal levels of *Spike*. Schematic representation of this protocol is represented in **Figure 18.A**.

The results of the ActD treatment follow the same tendency as the previous mRNA transfection assays. iPLUS seq3 3X leads to a decrease in mRNA stability, when compared to the *Spike* mRNA that lacks this sequence (**Figure 18.B**). This assay allows us to conclude this by the comparison of the half-lives of the mRNA, which is assessed by the timepoint at which the mRNA levels decrease to half of the quantity present in the beginning. This data is shown in **Figure 18.B**, where we can observe that the Wuhan_seq3 3X mRNA has a half-life around 7 hours (Bordeaux line), presenting a strong decay already 2h after treatment with ActD, while the Wuhan mRNA has a half-life of little less than 24 hours (blue line), which shows a substantial difference.

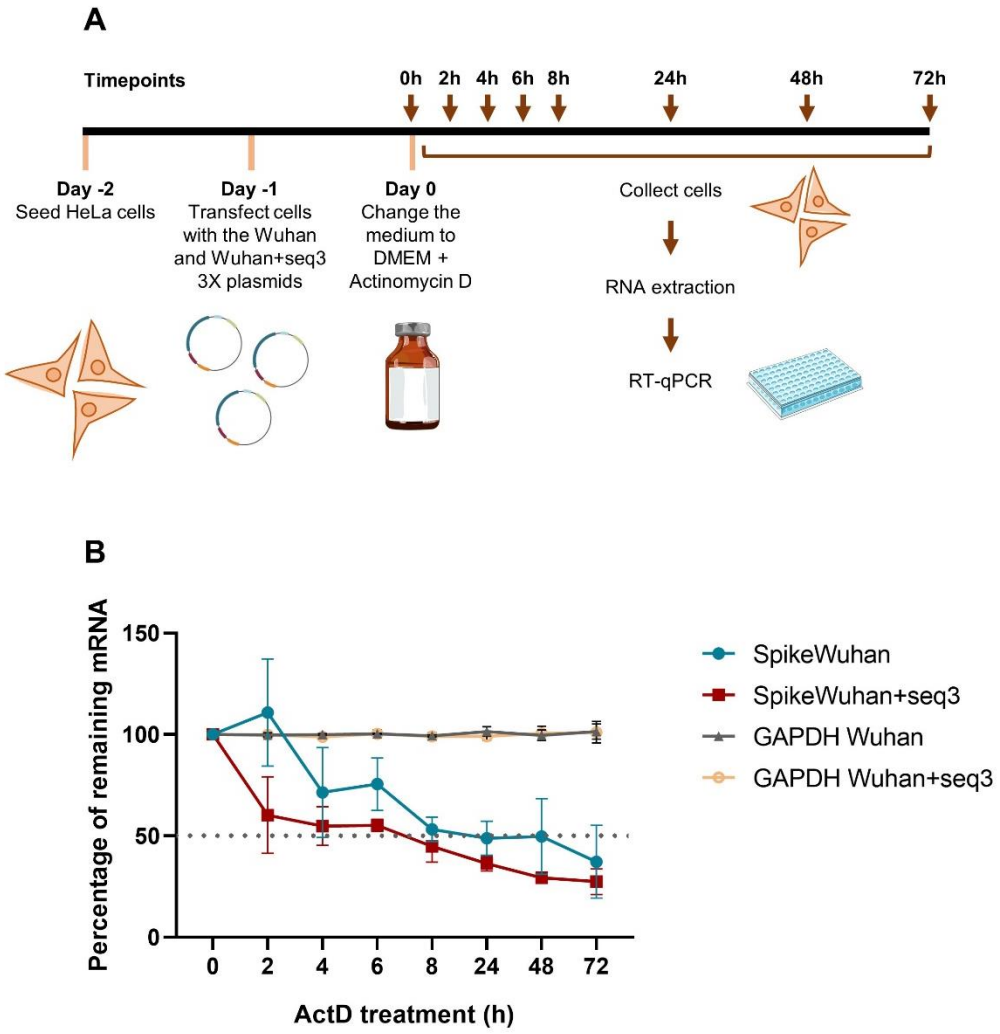


Figure 18 – mRNA stability assays using Actinomycin D show that the presence of iPLUS seq3 3X downstream of Spike's cDNA leads to a decrease in mRNA stability, when compared to the control that lacks this sequence. A) Schematic representation of the protocol followed for this experiment. On day -2, HeLa cells were seeded in 24-well plates, to guarantee approximately 70% confluency the next day. Cells were then transfected the day after with the Wuhan and Wuhan iPLUS seq3 3X plasmids. After an overnight incubation, the medium was changed to DMEM supplemented with Actinomycin D (5 µg/mL) in all conditions, except in the 0h timepoint, that serves as a control for the amount of mRNA basally present at the beginning of the experiment. Cells were treated during the represented timepoints, RNA was extracted using Trizol, and the samples were then analysed by RT-qPCR. **B)** mRNA containing iPLUS seq3 3X downstream of Spike's cDNA (bordeaux) has lower stability, when compared to the mRNA that lacks this sequence (blue), which we can conclude by the difference in the mRNA profiles. The plot shows the mean +/- standard deviation (SD) of the mRNA relative expression, normalized to *GAPDH*, that was used as a housekeeping gene. Three different experiments are represented (n=3). The expression levels of the stable *GAPDH* are also represented in grey (Wuhan) and pale pink (Wuhan+seq3 3X), for each timepoint. *MCL1* levels were also measured as a positive control with known mRNA half-life (not shown).

It is indeed surprising that iPLUS seq3 3X is able, on one hand, to increase Wuhan Spike's protein expression and the corresponding mRNA levels, and on the other hand, decrease the stability of this mRNA. These observations show that a different mechanism either than an increase in mRNAs stability is associated with the effect we observe in the cell transfection assays. In fact, it is important to reinforce that these

plasmids used for the Actinomycin D assays are the same used in the cell transfections, where not only the levels of protein, but even the levels of Spike's mRNA containing iPLUS seq3 3X were shown to be higher, in comparison to the control Wuhan.

One possible explanation is that iPLUS seq3 3X is acting on or affecting transcription. Maybe the Wuhan_seq3 3X plasmid is transcribed at higher levels than its control plasmid, Wuhan, producing high quantities of this mRNA, sufficient to overcome the higher rate of mRNA degradation, and still promoting the increase in protein levels. In fact, we always observe a stronger increase in Spike's protein levels in the presence of iPLUS seq3 3X in comparison to the control Wuhan, than the increase found in the mRNA levels. As so, it is possible that there is an increase in the efficiency of translation, perhaps with ribosomes being more capable of binding to the mRNAs that contain our sequence thus promoting their translation.

To our knowledge, these contrasting results regarding mRNA stability and protein and mRNA levels, have not been documented in other studies and may reveal other layers of gene expression regulation. Understandably, a decrease in mRNA stability is normally associated with a decrease in the respective mRNA levels and protein production. As so, to further understand this point, more experiments are required. It would be interesting to perform, for example, a nuclear run-on assay, where we could evaluate the levels of nascent Spike transcripts, in both Wuhan and Wuhan_seq3 3X. This is independent from the mRNA stability and would allow us to investigate if in fact Wuhan_seq3 3X is being transcribed more efficiently than the Wuhan control. It would also be interesting to test the efficiency of translation of both our mRNAs by, for instance, polysome profiling.

In fact, when we compare the two mRNAs, the only difference is the presence of this 21-nucleotide, non-coding, U-rich sequence. It is truly interesting how this small, non-coding sequence can impact so significantly several aspects of gene expression, regarding both the mRNA and the resulting protein, with mechanisms that remain to be elucidated, opening new avenues of research in the future.

Conclusions and Future Perspectives

First, we discovered that, for the SARS-CoV-2 Wuhan Spike variant subcloned in the pcDNA3.4 plasmid, the presence of iPLUS seq3 3X downstream of the cDNA, was able to promote an increase in Spike's protein production and mRNA levels. These results are further reinforced by the fact that they were validated in two different mammalian models, HeLa and ExpiCHO-S cells. The increase was always higher for Spike protein expression than for its mRNA levels, reaching 1.9-fold for HeLa cells and 2.5-fold for ExpiCHO-S cells (full-length Spike). For the iPLUS sequence, the results are not so interesting in the applicability point of view, since this sequence either does not impact the protein and mRNA levels or leads to a decrease. For the SARS-CoV-2 Omicron variant, both iPLUS and iPLUS seq3 3X either had no impact on the protein and mRNA levels of Spike (HeLa) or led to a decrease in both mRNA and protein levels (ExpiCHO-S cells). From a fundamental knowledge perspective, the functions of iPLUS and iPLUS variants in Spike's expression here described are new.

Secondly, and with the promising results found for iPLUS seq3 3X in the presence of the Wuhan variant, we evaluated the effect of this sequence in mRNA stability, when compared to the control lacking iPLUS seq3 3X. Strikingly, iPLUS seq3 3X leads to a decrease in *Spike*'s mRNA stability, proven both by HeLa cells transfection with *in vitro* transcribed mRNA and by ActD assays. Regarding mRNA transfection, with three different protocols we observed that the levels of mRNA are lower when iPLUS seq3 3X is present in the transfected mRNA. This was further reinforced by the results from the ActD assays, where the half-life of the Wuhan_seq3 3X mRNA was determined as approximately 7 h, while for the control Wuhan the half-life corresponds to approximately 24 h.

The mechanism behind iPLUS seq3 3X increase in Spike's protein expression and mRNA levels remains to be elucidated, since our hypothesis that it could be related to an increase in mRNA stability does not hold true. In fact, we are observing the opposite, with a less stable mRNA giving rise to an increase in protein and mRNA levels. This further proves the importance of keep studying the mechanisms of action behind iPLUS and the factors that may be binding to these iPLUS sequences and regulating Spike's expression. It is truly remarkable that a 21-nucleotide long, non-coding sequence can affect so significantly protein expression and mRNA levels and, even in a negative way, mRNA stability.

Our aims in this thesis were more focused on the specific biological function of these sequences, and we have proved that iPLUS seq3 3X could be a valuable asset to increase therapeutics for Pharma industries. Nevertheless, throughout this project several related questions were raised. The different results obtained with iPLUS and iPLUS seq3 3X, sequences that share many similarities, remain to be explained. Additionally, we observed that the coding sequence is able to influence the results obtained with both our sequences, since different results were obtained with the Wuhan and Omicron variant. This also raised important questions regarding the mode of action of iPLUS, which remains unknown.

Moreover, from the work developed in this thesis, some other new lines of investigation can be followed:

- The production of a stable cell line expressing Spike from the Wuhan variant, with iPLUS seq3 3X downstream of its sequence, can further solve our concerns regarding the efficiency of transfection. Moreover, this could be an important asset in the production of Spike's protein with high yields, for several biological applications.
- Evaluation of different iPLUS and iPLUS seq3 3X variants in this context. Could the addition of more tandem repeats of the most conserved region of iPLUS increase even more the effects observed with iPLUS seq3 3X? iPLUS seq3 6X or 12X could be tested, also to try to understand the role of this conserved sequence.
- The DNA templates developed for the *in vitro* transcription can be further used to test the effect of our mRNAs in mouse's immunization, both with the Wuhan and Wuhan_seq3 3X. For the moment, we are working in collaboration with other groups at i3S for the development of a LNP to encapsulate our mRNA, in order to prepare a compound with all the characteristics of an mRNA vaccine. We will then proceed with the immunization of a hACE2 mouse model already in use by our collaborators at the i3S, followed by evaluation of the mouse's capacity to fight a SARS-CoV-2 infection, as schematized in **Figure 19**. Additionally, we also intend to evaluate the immune response triggered after immunization with our mRNAs.

In short, throughout this thesis, we demonstrate how small, non-coding sequences present in the 3'UTR, can greatly affect the fate of an mRNA, regarding its levels, stability, and subsequent translation. These sequences can be assets for several applications, such as recombinant protein production or mRNA vaccines.

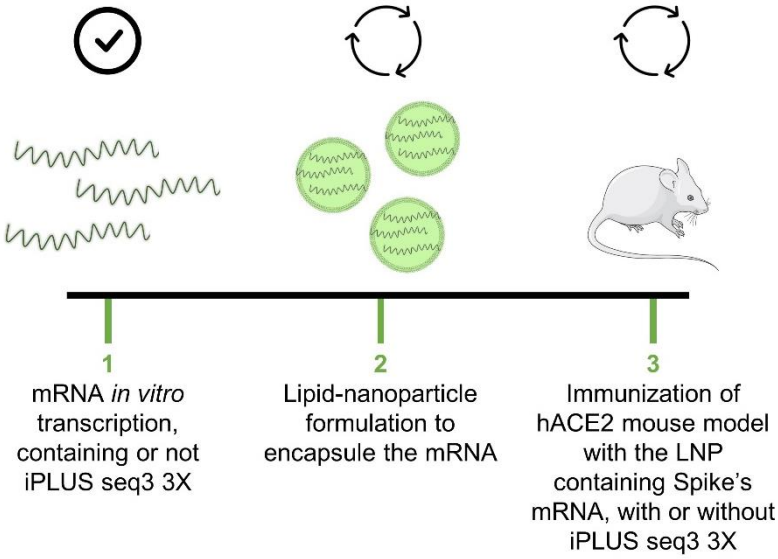


Figure 19 – Schematic representation of the workflow for testing our templates *in vivo*. The first task, that is already completed, was the preparation of the plasmid templates containing Spike's cDNA, either in the presence or not of iPLUS seq3 3X and their mRNA *in vitro* transcription. The second step, currently being developed, is the production of a lipid-nanoparticle (LNP) with the correct formulation, in order to allow the delivery of our mRNA molecules (with or without iPLUS seq3 3X) inside of cells. Finally, these molecules can be used for mouse immunization, giving us the possibility of testing if our sequence can in fact improve the capacity of the mouse to fight an infection with SARS-CoV-2.

References

1. Ghannam JY, Wang J, Jan A. Biochemistry, DNA Structure. StatPearls. Treasure Island (FL): StatPearls Publishing. Copyright © 2022, StatPearls Publishing LLC.; 2022.
2. Crick F. Central dogma of molecular biology. Nature. 1970;227(5258):561-3.
3. Wang D, Farhana A. Biochemistry, RNA Structure. StatPearls. Treasure Island (FL): StatPearls Publishing. Copyright © 2022, StatPearls Publishing LLC.; 2022.
4. Mercadante AA, Dimri M, Mohiuddin SS. Biochemistry, Replication and Transcription. StatPearls. Treasure Island (FL): StatPearls Publishing. Copyright © 2022, StatPearls Publishing LLC.; 2022.
5. Bentley DL. Coupling mRNA processing with transcription in time and space. Nat Rev Genet. 2014;15(3):163-75.
6. Osman S, Cramer P. Structural Biology of RNA Polymerase II Transcription: 20 Years On. Annu Rev Cell Dev Biol. 2020;36:1-34.
7. Hollerer I, Grund K, Hentze MW, Kulozik AE. mRNA 3'end processing: A tale of the tail reaches the clinic. EMBO Mol Med. 2014;6(1):16-26.
8. Ramanathan A, Robb GB, Chan SH. mRNA capping: biological functions and applications. Nucleic Acids Res. 2016;44(16):7511-26.
9. Will CL, Lührmann R. Spliceosome structure and function. Cold Spring Harb Perspect Biol. 2011;3(7).
10. Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative splicing. Nature. 2010;463(7280):457-63.
11. Marzluff WF, Koreski KP. Birth and Death of Histone mRNAs. Trends Genet. 2017;33(10):745-59.
12. Lutz CS, Moreira A. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. Wiley Interdiscip Rev RNA. 2011;2(1):22-31.
13. Eckmann CR, Rammelt C, Wahle E. Control of poly(A) tail length. Wiley Interdiscip Rev RNA. 2011;2(3):348-61.
14. Mandel CR, Bai Y, Tong L. Protein factors in pre-mRNA 3'-end processing. Cell Mol Life Sci. 2008;65(7-8):1099-122.
15. Huang Y, Carmichael GG. Role of polyadenylation in nucleocytoplasmic transport of mRNA. Mol Cell Biol. 1996;16(4):1534-42.
16. Erson-Bensan AE, Can T. Alternative Polyadenylation: Another Foe in Cancer. Mol Cancer Res. 2016;14(6):507-17.

17. Curinha A, Oliveira Braz S, Pereira-Castro I, Cruz A, Moreira A. Implications of polyadenylation in health and disease. Nucleus. 2014;5(6):508-19.
18. Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biological networks. Nat Rev Genet. 2010;11(1):75-87.
19. Wong F, Gunawardena J. Gene Regulation in and out of Equilibrium. Annu Rev Biophys. 2020;49:199-226.
20. Gruber AJ, Zavolan M. Alternative cleavage and polyadenylation in health and disease. Nat Rev Genet. 2019;20(10):599-614.
21. Proudfoot NJ, Brownlee GG. 3' non-coding region sequences in eukaryotic messenger RNA. Nature. 1976;263(5574):211-4.
22. Gruber AR, Martin G, Keller W, Zavolan M. Means to an end: mechanisms of alternative polyadenylation of messenger RNA precursors. Wiley Interdiscip Rev RNA. 2014;5(2):183-96.
23. Pereira-Castro I, Moreira A. On the function and relevance of alternative 3'-UTRs in gene expression regulation. Wiley Interdiscip Rev RNA. 2021;12(5):e1653.
24. Zarudnaya MI, Kolomiets IM, Potyahaylo AL, Hovorun DM. Downstream elements of mammalian pre-mRNA polyadenylation signals: primary, secondary and higher-order structures. Nucleic Acids Res. 2003;31(5):1375-86.
25. Chan S, Choi EA, Shi Y. Pre-mRNA 3'-end processing complex assembly and function. Wiley Interdiscip Rev RNA. 2011;2(3):321-35.
26. Proudfoot NJ. Ending the message: poly(A) signals then and now. Genes Dev. 2011;25(17):1770-82.
27. Shi Y, Di Giammartino DC, Taylor D, Sarkeshik A, Rice WJ, Yates JR, 3rd, et al. Molecular architecture of the human pre-mRNA 3' processing complex. Mol Cell. 2009;33(3):365-76.
28. Kaufmann I, Martin G, Friedlein A, Langen H, Keller W. Human Fip1 is a subunit of CPSF that binds to U-rich RNA elements and stimulates poly(A) polymerase. Embo j. 2004;23(3):616-26.
29. Nourse J, Spada S, Danckwardt S. Emerging Roles of RNA 3'-end Cleavage and Polyadenylation in Pathogenesis, Diagnosis and Therapy of Human Disorders. Biomolecules. 2020;10(6).
30. MacDonald CC, Wilusz J, Shenk T. The 64-kilodalton subunit of the CstF polyadenylation factor binds to pre-mRNAs downstream of the cleavage site and influences cleavage site location. Mol Cell Biol. 1994;14(10):6647-54.

31. Yang Q, Gilmartin GM, Doublié S. Structural basis of UGUA recognition by the Nudix protein CFI(m)25 and implications for a regulatory role in mRNA 3' processing. Proc Natl Acad Sci U S A. 2010;107(22):10062-7.

32. Venkataraman K, Brown KM, Gilmartin GM. Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. Genes Dev. 2005;19(11):1315-27.

33. Schäfer P, Tüting C, Schönemann L, Kühn U, Treiber T, Treiber N, et al. Reconstitution of mammalian cleavage factor II involved in 3' processing of mRNA precursors. Rna. 2018;24(12):1721-37.

34. Oliveira MS, Freitas J, Pinto PAB, de Jesus A, Tavares J, Pinho M, et al. Cell Cycle Kinase Polo Is Controlled by a Widespread 3' Untranslated Region Regulatory Sequence in Drosophila melanogaster. Mol Cell Biol. 2019;39(15).

35. Derti A, Garrett-Engele P, Macisaac KD, Stevens RC, Sriram S, Chen R, et al. A quantitative atlas of polyadenylation in five mammals. Genome Res. 2012;22(6):1173-83.

36. Hoque M, Ji Z, Zheng D, Luo W, Li W, You B, et al. Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. Nat Methods. 2013;10(2):133-9.

37. Di Giammartino DC, Nishida K, Manley JL. Mechanisms and consequences of alternative polyadenylation. Mol Cell. 2011;43(6):853-66.

38. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. Science. 2008;320(5883):1643-7.

39. Elkon R, Drost J, van Haaften G, Jenal M, Schrier M, Oude Vrielink JA, et al. E2F mediates enhanced alternative polyadenylation in proliferation. Genome Biol. 2012;13(7):R59.

40. Mayr C, Bartel DP. Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. Cell. 2009;138(4):673-84.

41. Lai DP, Tan S, Kang YN, Wu J, Ooi HS, Chen J, et al. Genome-wide profiling of polyadenylation sites reveals a link between selective polyadenylation and cancer metastasis. Hum Mol Genet. 2015;24(12):3410-7.

42. Zheng D, Wang R, Ding Q, Wang T, Xie B, Wei L, et al. Cellular stress alters 3'UTR landscape through alternative polyadenylation and isoform-specific degradation. Nat Commun. 2018;9(1):2268.

43. Ji Z, Lee JY, Pan Z, Jiang B, Tian B. Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. Proc Natl Acad Sci U S A. 2009;106(17):7028-33.

44. Chen M, Lyu G, Han M, Nie H, Shen T, Chen W, et al. 3' UTR lengthening as a novel mechanism in regulating cellular senescence. Genome Res. 2018;28(3):285-94.

45. Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC. Widespread and extensive lengthening of 3' UTRs in the mammalian brain. Genome Res. 2013;23(5):812-25.

46. Early P, Rogers J, Davis M, Calame K, Bond M, Wall R, et al. Two mRNAs can be produced from a single immunoglobulin mu gene by alternative RNA processing pathways. Cell. 1980;20(2):313-9.

47. Takagaki Y, Seipelt RL, Peterson ML, Manley JL. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell. 1996;87(5):941-52.

48. Mayr C. Regulation by 3'-Untranslated Regions. Annu Rev Genet. 2017;51:171-94.

49. Domingues RG, Lago-Baldaia I, Pereira-Castro I, Fachini JM, Oliveira L, Drpic D, et al. CD5 expression is regulated during human T-cell activation by alternative polyadenylation, PTBP1, and miR-204. Eur J Immunol. 2016;46(6):1490-503.

50. Gruber AR, Martin G, Müller P, Schmidt A, Gruber AJ, Gumienny R, et al. Global 3' UTR shortening has a limited effect on protein abundance in proliferating T cells. Nat Commun. 2014;5:5465.

51. Pinto PA, Henriques T, Freitas MO, Martins T, Domingues RG, Wyrzykowska PS, et al. RNA polymerase II kinetics in polo polyadenylation signal selection. Embo j. 2011;30(12):2431-44.

52. Mayr C. What Are 3' UTRs Doing? Cold Spring Harb Perspect Biol. 2019;11(10).

53. Matoulkova E, Michalova E, Vojtesek B, Hrstka R. The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. RNA Biol. 2012;9(5):563-76.

54. Barreau C, Paillard L, Osborne HB. AU-rich elements and associated factors: are there unifying principles? Nucleic Acids Res. 2005;33(22):7138-50.

55. Vlasova IA, Bohjanen PR. Posttranscriptional regulation of gene networks by GU-rich elements and CELF proteins. RNA Biol. 2008;5(4):201-7.

56. Darmon SK, Lutz CS. Novel upstream and downstream sequence elements contribute to polyadenylation efficiency. RNA Biol. 2012;9(10):1255-65.

57. Danckwardt S, Kaufmann I, Gentzel M, Foerstner KU, Gantzert AS, Gehring NH, et al. Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. Embo j. 2007;26(11):2658-69.

58. Lutz CS, Murthy KG, Schek N, O'Connor JP, Manley JL, Alwine JC. Interaction between the U1 snRNP-A protein and the 160-kD subunit of cleavage-polyadenylation specificity factor increases polyadenylation efficiency in vitro. Genes Dev. 1996;10(3):325-37.

59. Moreira A, Takagaki Y, Brackenridge S, Wollerton M, Manley JL, Proudfoot NJ. The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. Genes Dev. 1998;12(16):2522-34.

60. Sittler A, Gallinaro H, Jacob M. The secondary structure of the adenovirus-2 L4 polyadenylation domain: evidence for a hairpin structure exposing the AAUAAA signal in its loop. J Mol Biol. 1995;248(3):525-40.

61. Gilmartin GM, Fleming ES, Oetjen J. Activation of HIV-1 pre-mRNA 3' processing in vitro requires both an upstream element and TAR. Embo j. 1992;11(12):4419-28.

62. Carswell S, Alwine JC. Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences. Mol Cell Biol. 1989;9(10):4248-58.

63. DeZazzo JD, Imperiale MJ. Sequences upstream of AAUAAA influence poly(A) site selection in a complex transcription unit. Mol Cell Biol. 1989;9(11):4951-61.

64. Sanfaçon H. Analysis of figwort mosaic virus (plant pararetrovirus) polyadenylation signal. Virology. 1994;198(1):39-49.

65. Kaida D, Berg MG, Younis I, Kasim M, Singh LN, Wan L, et al. U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. Nature. 2010;468(7324):664-8.

66. Gilmartin GM, Fleming ES, Oetjen J, Graveley BR. CPSF recognition of an HIV-1 mRNA 3'-processing enhancer: multiple sequence contacts involved in poly(A) site definition. Genes Dev. 1995;9(1):72-83.

67. Saha B, Parks RJ. Recent Advances in Novel Antiviral Therapies against Human Adenovirus. Microorganisms. 2020;8(9).

68. Prescott JC, Falck-Pedersen E. Varied poly(A) site efficiency in the adenovirus major late transcription unit. J Biol Chem. 1992;267(12):8175-81.

69. Sittler A, Gallinaro H, Jacob M. Upstream and downstream cis-acting elements for cleavage at the L4 polyadenylation site of adenovirus-2. Nucleic Acids Res. 1994;22(2):222-31.

70. Prescott J, Falck-Pedersen E. Sequence elements upstream of the 3' cleavage site confer substrate strength to the adenovirus L1 and L3 polyadenylation sites. Mol Cell Biol. 1994;14(7):4682-93.

71. Silver Key SC, Pagano JS. A noncanonical poly(A) signal, UAUAAA, and flanking elements in Epstein-Barr virus DNA polymerase mRNA function in cleavage and polyadenylation assays. Virology. 1997;234(1):147-59.

72. Schambach A, Galla M, Maetzig T, Loew R, Baum C. Improving transcriptional termination of self-inactivating gamma-retroviral and lentiviral vectors. Mol Ther. 2007;15(6):1167-73.

73. Valsamakis A, Zeichner S, Carswell S, Alwine JC. The human immunodeficiency virus type 1 polyadenylylation signal: a 3' long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylylation. Proc Natl Acad Sci U S A. 1991;88(6):2108-12.

74. Russnak R, Ganem D. Sequences 5' to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. Genes Dev. 1990;4(5):764-76.

75. Hall-Pogar T, Zhang H, Tian B, Lutz CS. Alternative polyadenylation of cyclooxygenase-2. Nucleic Acids Res. 2005;33(8):2565-79.

76. Moreira A, Wollerton M, Monks J, Proudfoot NJ. Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. Embo j. 1995;14(15):3809-19.

77. Sachchithananthan M, Stasinopoulos SJ, Wilusz J, Medcalf RL. The relationship between the prothrombin upstream sequence element and the G20210A polymorphism: the influence of a competitive environment for mRNA 3'-end formation. Nucleic Acids Res. 2005;33(3):1010-20.

78. Brackenridge S, Proudfoot NJ. Recruitment of a basal polyadenylation factor by the upstream sequence element of the human lamin B2 polyadenylation signal. Mol Cell Biol. 2000;20(8):2660-9.

79. Brackenridge S, Ashe HL, Giacca M, Proudfoot NJ. Transcription and polyadenylation in a short human intergenic region. Nucleic Acids Res. 1997;25(12):2326-36.

80. Danckwardt S, Gantzert AS, Macher-Goeppinger S, Probst HC, Gentzel M, Wilm M, et al. p38 MAPK controls prothrombin expression by regulated RNA 3' end processing. Mol Cell. 2011;41(3):298-310.

81. Llamazares S, Moreira A, Tavares A, Girdham C, Spruce BA, Gonzalez C, et al. polo encodes a protein kinase homolog required for mitosis in Drosophila. Genes Dev. 1991;5(12a):2153-65.

82. Oliveira MJ, A; Freitas, J; Moreira, A; Ribeiro Bessa, JC. Nucleic Acid to Activate Gene Expression and Protein Production, WO/2020/076174, 2020.
83. Piret J, Boivin G. Pandemics Throughout History. Front Microbiol. 2020;11:631736.
84. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 2020;382(8):727-33.
85. Schoeman D, Fielding BC. Coronavirus envelope protein: current knowledge. Virol J. 2019;16(1):69.
86. Chathappady House NN, Palissery S, Sebastian H. Corona Viruses: A Review on SARS, MERS and COVID-19. Microbiol Insights. 2021;14:11786361211002481.
87. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020;20(5):533-4.
88. Yang SL, Ponti RD, Wan Y, Huber RG. Computational and Experimental Approaches to Study the RNA Secondary Structures of RNA Viruses. Viruses. 2022;14(8).
89. Duan L, Zheng Q, Zhang H, Niu Y, Lou Y, Wang H. The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. Front Immunol. 2020;11:576622.
90. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell. 2020;181(2):281-92.e6.
91. Candido KL, Eich CR, de Fariña LO, Kadowaki MK, da Conceição Silva JL, Maller A, et al. Spike protein of SARS-CoV-2 variants: a brief review and practical implications. Braz J Microbiol. 2022:1-25.
92. McCallum M, De Marco A, Lempp FA, Tortorici MA, Pinto D, Walls AC, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell. 2021;184(9):2332-47.e16.
93. Cerutti G, Guo Y, Zhou T, Gorman J, Lee M, Rapp M, et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. Cell Host Microbe. 2021;29(5):819-33.e7.
94. Scovino AM, Dahab EC, Vieira GF, Freire-de-Lima L, Freire-de-Lima CG, Morrot A. SARS-CoV-2's Variants of Concern: A Brief Characterization. Front Immunol. 2022;13:834098.
95. Dhawan M, Saied AA, Mitra S, Alhumaydhi FA, Emran TB, Wilairatana P. Omicron variant (B.1.1.529) and its sublineages: What do we know so far amid the

emergence of recombinant variants of SARS-CoV-2? Biomed Pharmacother. 2022;154:113522.

96. Hoffmann M, Krüger N, Schulz S, Cossmann A, Rocha C, Kempf A, et al. The Omicron variant is highly resistant against antibody-mediated neutralization: Implications for control of the COVID-19 pandemic. Cell. 2022;185(3):447-56.e11.

97. Cui Z, Liu P, Wang N, Wang L, Fan K, Zhu Q, et al. Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 Omicron. Cell. 2022;185(5):860-71.e13.

98. Fan Y, Li X, Zhang L, Wan S, Zhang L, Zhou F. SARS-CoV-2 Omicron variant: recent progress and future perspectives. Signal Transduct Target Ther. 2022;7(1):141.

99. Cao Y, Wang J, Jian F, Xiao T, Song W, Yisimayi A, et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. Nature. 2022;602(7898):657-63.

100. Dolgin E. The tangled history of mRNA vaccines. Nature. 2021;597(7876):318-24.

101. Malone RW, Felgner PL, Verma IM. Cationic liposome-mediated RNA transfection. Proc Natl Acad Sci U S A. 1989;86(16):6077-81.

102. Jackson NAC, Kester KE, Casimiro D, Gurunathan S, DeRosa F. The promise of mRNA vaccines: a biotech and industrial perspective. NPJ Vaccines. 2020;5:11.

103. Zhang C, Maruggi G, Shan H, Li J. Advances in mRNA Vaccines for Infectious Diseases. Front Immunol. 2019;10:594.

104. Pilkington EH, Suys EJA, Trevaskis NL, Wheatley AK, Zukancic D, Algarni A, et al. From influenza to COVID-19: Lipid nanoparticle mRNA vaccines at the frontiers of infectious diseases. Acta Biomater. 2021;131:16-40.

105. Rauch S, Jasny E, Schmidt KE, Petsch B. New Vaccine Technologies to Combat Outbreak Situations. Front Immunol. 2018;9:1963.

106. Rijkers GT, Weterings N, Obregon-Henao A, Lepolder M, Dutt TS, van Overveld FJ, et al. Antigen Presentation of mRNA-Based and Virus-Vectored SARS-CoV-2 Vaccines. Vaccines (Basel). 2021;9(8).

107. Park JW, Lagniton PNP, Liu Y, Xu RH. mRNA vaccines for COVID-19: what, why and how. Int J Biol Sci. 2021;17(6):1446-60.

108. Weissman D. mRNA transcript therapy. Expert Rev Vaccines. 2015;14(2):265-81.

109. Muttach F, Muthmann N, Rentmeister A. Synthetic mRNA capping. Beilstein J Org Chem. 2017;13:2819-32.

110. Gómez-Aguado I, Rodríguez-Castejón J, Vicente-Pascual M, Rodríguez-Gascón A, Solinís M, Del Pozo-Rodríguez A. Nanomedicines to Deliver mRNA: State of the Art and Future Perspectives. Nanomaterials (Basel). 2020;10(2).

111. Anderson BR, Muramatsu H, Nallagatla SR, Bevilacqua PC, Sansing LH, Weissman D, et al. Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. Nucleic Acids Res. 2010;38(17):5884-92.

112. Buschmann MD, Carrasco MJ, Alishetty S, Paige M, Alameh MG, Weissman D. Nanomaterial Delivery Systems for mRNA Vaccines. Vaccines (Basel). 2021;9(1).

113. Gebre MS, Rauch S, Roth N, Yu J, Chandrashekar A, Mercado NB, et al. Optimization of non-coding regions for a non-modified mRNA COVID-19 vaccine. Nature. 2022;601(7893):410-4.

114. Johnson AG, Grosely R, Petrov AN, Puglisi JD. Dynamics of IRES-mediated translation. Philos Trans R Soc Lond B Biol Sci. 2017;372(1716).

115. Jiang Y, Xu XS, Russell JE. A nucleolin-binding 3' untranslated region element stabilizes beta-globin mRNA in vivo. Mol Cell Biol. 2006;26(6):2419-29.

116. Zufferey R, Donello JE, Trono D, Hope TJ. Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. J Virol. 1999;73(4):2886-92.

117. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020;581(7807):221-4.

118. Pereira-Castro I, Garcia BC, Curinha A, Neves-Costa A, Conde-Sousa E, Moita LF, et al. MCL1 alternative polyadenylation is essential for cell survival and mitochondria morphology. Cell Mol Life Sci. 2022;79(3):164.

119. Chaudhary N, Weissman D, Whitehead KA. mRNA vaccines for infectious diseases: principles, delivery and clinical translation. Nat Rev Drug Discov. 2021;20(11):817-38.

120. Lee EK, Gorospe M. Coding region: the neglected post-transcriptional code. RNA Biol. 2011;8(1):44-8.

121. Grzybowska EA, Wakula M. Protein Binding to Cis-Motifs in mRNAs Coding Sequence Is Common and Regulates Transcript Stability and the Rate of Translation. Cells. 2021;10(11).

122. Brümmer A, Hausser J. MicroRNA binding sites in the coding region of mRNAs: extending the repertoire of post-transcriptional gene regulation. Bioessays. 2014;36(6):617-26.

123. Abedi F, Rezaee R, Hayes AW, Nasiripour S, Karimi G. MicroRNAs and SARS-CoV-2 life cycle, pathogenesis, and mutations: biomarkers or therapeutic agents? Cell Cycle. 2021;20(2):143-53.

124. Ahmed JQ, Maulud SQ, Dhawan M, Priyanka, Choudhary OP, Jalal PJ, et al. MicroRNAs in the development of potential therapeutic targets against COVID-19: A narrative review. J Infect Public Health. 2022;15(7):788-99.

125. Marchi R, Sugita B, Centa A, Fonseca AS, Bortoletto S, Fiorentin K, et al. The role of microRNAs in modulating SARS-CoV-2 infection in human cells: a systematic review. Infect Genet Evol. 2021;91:104832.

126. Hunter M, Yuan P, Vavilala D, Fox M. Optimization of Protein Expression in Mammalian Cells. Curr Protoc Protein Sci. 2019;95(1):e77.

127. O'Flaherty R, Bergin A, Flampouri E, Mota LM, Obaidi I, Quigley A, et al. Mammalian cell culture for production of recombinant proteins: A review of the critical steps in their biomanufacturing. Biotechnol Adv. 2020;43:107552.

128. Gutiérrez-Granados S, Cervera L, Kamen AA, Gòdia F. Advancements in mammalian cell transient gene expression (TGE) technology for accelerated production of biologics. Crit Rev Biotechnol. 2018;38(6):918-40.

129. Schaub JM, Chou CW, Kuo HC, Javanmardi K, Hsieh CL, Goldsmith J, et al. Expression and characterization of SARS-CoV-2 spike proteins. Nat Protoc. 2021;16(11):5339-56.

130. Sinegubova MV, Orlova NA, Kovnir SV, Dayanova LK, Vorobiev, II. High-level expression of the monomeric SARS-CoV-2 S protein RBD 320-537 in stably transfected CHO cells by the EEF1A1-based plasmid vector. PLoS One. 2021;16(2):e0242890.

131. Stuible M, Gervais C, Lord-Dufour S, Perret S, L'Abbé D, Schrag J, et al. Rapid, high-yield production of full-length SARS-CoV-2 spike ectodomain by transient gene expression in CHO cells. J Biotechnol. 2021;326:21-7.

132. Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun. 2020;11(1):1620.

133. Meng B, Abdullahi A, Ferreira I, Goonawardane N, Saito A, Kimura I, et al. Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. Nature. 2022;603(7902):706-14.

134. Escalera A, Gonzalez-Reiche AS, Aslam S, Mena I, Laporte M, Pearl RL, et al. Mutations in SARS-CoV-2 variants of concern link to increased spike cleavage and virus transmission. Cell Host Microbe. 2022;30(3):373-87.e7.

135. Powell SK, Rivera-Soto R, Gray SJ. Viral expression cassette elements to enhance transgene target specificity and expression in gene therapy. Discov Med. 2015;19(102):49-57.

136. Azzoni AR, Ribeiro SC, Monteiro GA, Prazeres DM. The impact of polyadenylation signals on plasmid nuclease-resistance and transgene expression. J Gene Med. 2007;9(5):392-402.

137. Levitt N, Briggs D, Gil A, Proudfoot NJ. Definition of an efficient synthetic poly(A) site. Genes Dev. 1989;3(7):1019-25.

138. Wang XY, Du QJ, Zhang WL, Xu DH, Zhang X, Jia YL, et al. Enhanced Transgene Expression by Optimization of Poly A in Transfected CHO Cells. Front Bioeng Biotechnol. 2022;10:722722.

139. World Health Organization Messenger RNA Encoding the Full-Length SARS-CoV-2 Spike Glycoprotein.
https://web.archive.org/web/20210105162941/https://mednetcommunities.net/inn/db/media/docs/11889.doc2020

140. Tavernier G, Andries O, Demeester J, Sanders NN, De Smedt SC, Rejman J. mRNA as gene therapeutic: how to control protein expression. J Control Release. 2011;150(3):238-47.

141. Lai WS, Arvola RM, Goldstrohm AC, Blackshear PJ. Inhibiting transcription in cultured metazoan cells with actinomycin D to monitor mRNA turnover. Methods. 2019;155:77-87.