



OPEN ACCESS

EDITED BY

Carla Silva,
University of Minho, Portugal

REVIEWED BY

Joao C. Guimaraes,
University of Lisbon, Portugal
Muriel Bardor,
Université de Rouen, France

*CORRESPONDENCE

Alexandra Moreira,
✉ alexandra.moreira@i3s.up.pt

RECEIVED 14 December 2023

ACCEPTED 25 January 2024

PUBLISHED 06 February 2024

CITATION

Reis-Claro I, Silva MI, Moutinho A, Garcia BC,
Pereira-Castro I and Moreira A (2024),
Application of the iPLUS non-coding sequence
in improving biopharmaceuticals production.
Front. Bioeng. Biotechnol. 12:1355957.
doi: 10.3389/fbioe.2024.1355957

COPYRIGHT

© 2024 Reis-Claro, Silva, Moutinho, Garcia,
Pereira-Castro and Moreira. This is an open-
access article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Application of the iPLUS non-coding sequence in improving biopharmaceuticals production

Inês Reis-Claro¹, Maria Inês Silva¹, Ana Moutinho¹,
Beatriz C. Garcia¹, Isabel Pereira-Castro^{1,2} and
Alexandra Moreira^{1,2,3*}

¹Gene Regulation, i3S—Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal, ²BMC—Instituto de Biologia Molecular e Celular, Universidade do Porto, Porto, Portugal, ³ICBAS—Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, Portugal

The biotechnological landscape has witnessed significant growth in biological therapeutics particularly in the field of recombinant protein production. Here we investigate the function of 3'UTR *cis*-regulatory elements in increasing mRNA and protein levels in different biological therapeutics and model systems, spanning from monoclonal antibodies to mRNA vaccines. We explore the regulatory function of iPLUS - a universal sequence capable of consistently augmenting recombinant protein levels. By incorporating iPLUS in a vector to express a monoclonal antibody used in immunotherapy, in a mammalian cell line used by the industry (ExpiCHO), trastuzumab production increases by 2-fold. As yeast *Pichia pastoris* is widely used in the manufacture of industrial enzymes and pharmaceuticals, we then used iPLUS in tandem (3x) and iPLUSv2 (a variant of iPLUS) to provide proof-of-concept data that it increases the production of a reporter protein more than 100-fold. As iPLUS functions by also increasing mRNA levels, we hypothesize that these sequences could be used as an asset in the mRNA vaccine industry. In fact, by including iPLUSv2 downstream of Spike we were able to double its production. Moreover, the same effect was observed when we introduced iPLUSv2 downstream of MAGEC2, a tumor-specific antigen tested for cancer mRNA vaccines. Taken together, our study provides data (TLR4) showing that iPLUS may be used as a valuable asset in a variety of systems used by the biotech and biopharmaceutical industry. Our results underscore the critical role of non-coding sequences in controlling gene expression, offering a promising avenue to accelerate, enhance, and cost-effectively optimize biopharmaceutical production processes.

KEYWORDS

cis-regulatory sequences, mRNA therapeutics, recombinant protein expression, USE, 3'UTR, iPLUS

1 Introduction

The recombinant protein production market has grown immensely over the past years, with a strong impact in biological therapeutics (O'Flaherty et al., 2020; Tihanyi and Nyitray, 2020). Since the first time the Food and Drug Administration (FDA) approved insulin to be used as a therapeutics, 41 years have passed and several hundred more have been approved

(Dingermann, 2008; Kesik-Brodacka, 2018). Recombinant proteins of therapeutic interest include monoclonal antibodies, hormones, enzymes, among other classes of proteins and/or peptides (O'Flaherty et al., 2020). Other biopharmaceutical whose importance and usage rose with the COVID-19 pandemic is the mRNA vaccines. mRNA as a therapeutics has the advantage of delivering the specific sequence of a known antigen, instructing the cells to manufacture the protein, which then culminates in an immunologic response specific for that immunogen (Pardi et al., 2018). Additionally, as they are simple to use and manufacture, mRNA vaccines have a variety of applications, from therapeutic vaccination for multiple cancers, to preventive immunization towards different pathogens, including new pandemics and outbreaks (Rauch et al., 2018).

The mRNAs untranslated regions (UTRs) have been known for a long time to have a crucial role in gene expression regulation (Gruber and Zavolan, 2019; Pereira-Castro and Moreira, 2021). There is a large variety of *cis*-regulatory elements present in the UTRs, either in the 5' or the 3' (Matoulkova et al., 2012). The 5'UTR has an important role in translation initiation, becoming an important region for regulation of translational efficiency (Jackson et al., 2010). For example, the presence of the internal ribosomal entry sites (IRES) sequence in the 5'UTR, found originally in certain virus, promotes the recruitment of the ribosomes, improving translation initiation. As so, this sequence is commonly employed in different expression systems (Johnson et al., 2017). The Kozak motif is also a commonly employed 9-nucleotide conserved sequence in expression vectors, as it is critical for the recognition of the translation initiation site by the ribosomes in eukaryotic mRNAs (Hernández et al., 2019). As for the 3'UTRs, their ability to control the mRNA fate comes from numerous *cis*-regulatory elements, by allowing the binding of *trans*-acting factors, such as RNA-binding proteins (RBPs) and microRNAs (miRNAs) (Mayr, 2019). The binding of these factors will impact gene expression regulation, not only by affecting 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 (Mayr, 2017; Pereira-Castro and Moreira, 2021). Usually, each 3'UTR regulatory sequence is tested in a gene and cell-type specific context (Vallazza et al., 2015). Sequences from the α -globin and β -globin, for example, are highly employed in expression systems since it is known that, in a physiological context, they promote a high stability to their mRNAs (Jiang et al., 2006). Another example of a broadly used sequence, is the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE), a 597 nucleotide-long sequence, that has been showed to increase luciferase and GFP production levels by five to eight-fold, when subcloned in the 3' of the cDNAs subcloned in expression vectors (Zufferey et al., 1999).

Recently, genome-wide studies have been conducted in the search for active regulatory sequences naturally present in the genomes, revealing new sequences that may be employed in the production of recombinant proteins of high commercial value by the biopharma and biotech industries (Sample et al., 2019; Cao et al., 2021; Leppke et al., 2022; Kirshina et al., 2023). For the 5'UTR, Cao et al. (2021) employed high-throughput screening in combination with sequence engineering, by initially analyzing 5'UTRs that

naturally reveal good translation efficiencies, followed by synthetic modification of these sequences to improve their effectiveness even further. The results obtained were validated in several cell models and the increase of protein expression was corroborated also in proteins with therapeutic value, such as vascular endothelial growth factor (VEGF). This approach has also been applied to the search for regulatory regions in the 3'UTR, also in the context of mRNA-based therapeutics, such as cancer vaccines, by increasing the mRNA stability and protein expression (Leppke et al., 2022; Kirshina et al., 2023). For example, the application of a combination of mtRNR1 (mitochondrial rRNA) and AES (Amino-Terminal Enhancer Of Split) 3'UTR sequences, promoted a T cell immune response against a vaccine antigen, as well as an increase in the delivery of reprogramming factors to induce pluripotency of differentiated cells (Orlandini von Niessen et al., 2019). These efforts to find new non-coding regulatory sequences reveal a yet unexplored field to optimize the expression vectors used in the production of biological therapeutics.

Certainly, *cis*-regulatory sequences present in the UTRs can be used to increase mRNA expression, as well as protein production, either in large scale as for recombinant protein production, or in the cells in the context of an immunization with an mRNA vaccine (Gómez-Aguado et al., 2020). For example, the human albumin 3'UTR increases the production of secreted luciferase by 2-3 fold in CHO cells (Pearson et al., 2012). The structure of the mRNAs in the vaccines mimics endogenous mRNAs (Pardi et al., 2018) and their design has been optimized to elicit a strong and specific immunological response. Delivery of the mRNA vaccines must promote the antigen's expression at levels sufficient for detection of immune responses (Karikó et al., 1999). Thus, 5'UTR and 3'UTR secondary structures, 3'UTR length, and different *cis*-regulatory elements present in both the 3'UTR and 5'UTR contribute to achieve high expression levels (Gómez-Aguado et al., 2020; Jackson et al., 2020; Castillo-Hair and Seelig, 2022).

The upstream sequence elements (USEs), located upstream of polyadenylation signals (PAS), are a good example of *cis*-regulatory sequences present in the 3'UTR. USEs are involved in cleavage and polyadenylation (CPA) efficiency, by allowing the binding of the CPA machinery (Zhao et al., 1999; Proudfoot, 2011; Darmon and Lutz, 2012; Oliveira et al., 2019). USEs possess a characteristic pyrimidine and/or U-rich stretch (Chan et al., 2011) and function as binding sites for various RBPs, modulating gene expression (Carswell and Alwine, 1989; Moreira et al., 1998; Danckwardt et al., 2007). Therefore, USEs have the potential to be employed in different applications besides their physiological role in CPA. For example, the subcloning of different USEs in viral vectors was tested to improve their 3'end processing efficiency by avoiding readthrough, consequently increasing both their efficiency and biosafety (Schambach et al., 2007). Also, it was demonstrated that the use of the USE SV40-derived sequence, as well as its repetition in tandem 2 times, led to an increase of 45%–100% in gene expression, depending on the promoters used (Schambach et al., 2007).

We have previously found that *Drosophila's polo* gene possesses a USE upstream of a weak PAS, highly conserved among *Drosophila* genomes (Oliveira et al., 2019). This 28-nucleotides long, pyrimidine-rich USE presents a physiological role in increasing

Polo's protein levels (Oliveira et al., 2019). This result led us to hypothesize that this mechanism is conserved, therefore, its function was investigated in proof-of-concept experiments in other genes and organisms. In *Danio rerio*, the microinjection of a plasmid containing this USE downstream of the green fluorescent protein (GFP) was used as a reporter and showed that the USE increases GFP expression levels by ~2-fold, indicating that the function of the USE is conserved in vertebrates (Eufrásio et al., 2023). We found that three RBPs (HuR, hnRNP C and PTBP1) bind to the USE in human cells, and in *Drosophila*, Hephastus (the ortholog of the human PTBP1), enhances the use of the PAS through interaction with *polo*'s USE leading to higher mRNA and protein levels (Oliveira et al., 2019; Eufrásio et al., 2023). In mammalian cell lines, specifically HeLa cells, it has previously been shown that this sequence led to an increase in a reporter's activity (Eufrásio et al., 2023).

The consistent increase in protein levels observed when *polo*'s USE was inserted in the 3'UTR led us to secure its potential in biopharmaceutical's formulations usage, patenting it as iPLUS, which stands for increase Protein Levels Universal Sequence (Oliveira et al., 2020). iPLUS fits in the need for non-coding regulatory regions to improve effectiveness of gene expression, as the biopharmaceutical and biotechnology industries requires their improvement in the manufacture processes with the goal of lowering production costs.

In this work we explored the potentialities of iPLUS in increasing the efficiency of mRNA and protein production for different biopharmaceuticals. We demonstrated that iPLUS, and a variant of this sequence that we named iPLUSv2, which consists of 3 repeats of the most conserved 8 nucleotide sequence stretch of iPLUS/USE (Oliveira et al., 2019) increases mRNA and protein levels in different models and for various proteins of biopharmaceutical interest. These results reinforce the importance of 3' non-coding sequences in the regulation of gene expression, highlighting their potential use in the process of manufacture to accelerate, improve and reduce the costs of the biopharmaceutical's production.

2 Materials and methods

2.1 Plasmids construction

2.1.1 Trastuzumab_LC and HC

The plasmid pcDNA3.4-TOPO (Thermo Fisher Scientific) was used as a backbone to subclone the trastuzumab LC (light chain) and HC (heavy chain) upstream of the WPRE sequence. The pcDNA3.4-TOPO was digested with AgeI-HF (NEB), the restriction site was blunted using Klenow (NEB), followed by digestion with XbaI (NEB). LC and HC were amplified by PCR using the phosphorylated primer pair LC+XbaI F; LC-R and HC+XbaI F; HC-R (Supplementary Table S1), pVITRO1-Trastuzumab-IgG1/κ [a gift from Andrew Beavil, Addgene plasmid #61883 (Dodev et al., 2014)] as the template DNA, and Phusion High Fidelity DNA polymerase (ThermoFisher Scientific) following manufacturer recommendations, for cloning in pcDNA3.4-TOPO in the XbaI and blunted AgeI restriction sites.

2.1.2 Trastuzumab_LC_iPLUS

Firstly, pGEM-4 (Promega) digested with SacI (NEB) and KpnI (NEB) was used to subclone the annealed oligos containing iPLUS sequence (AATTTATTTGTTTTTGGCCCTTCCCTT—Supplementary Table S1) flanked by digested SacI and KpnI restriction sites (pGEM-4-iPLUS). pGEM-4-iPLUS was digested with BamHI (NEB) and EcoRI (NEB) in order to obtain the iPLUS sequence. Trastuzumab_LC was linearized by an inverted PCR using WPRE+BamHI-F and WPRE+EcoRI-R primers (Supplementary Table S1) and Phusion High Fidelity DNA polymerase (Thermo Fisher Scientific). Linearized plasmid was digested with BamHI and EcoRI to clone the iPLUS sequence downstream of the WPRE sequence and upstream the HSV TK PAS.

2.1.3 Trastuzumab_HC_iPLUS

An inverted PCR using Trastuzumab_HC was performed as for Trastuzumab_LC_iPLUS. To obtain Trastuzumab_HC_iPLUS, the linearized plasmid and pGEM-4-iPLUS were digested with KpnI and the restriction site was blunted using Klenow, followed by EcoRI digestion.

2.1.4 pTZ-p04_GFP_iPLUS 3X and pTZ-p04_GFP_iPLUSv2

The pTZ-p04_GFP was already prepared and is a proprietary vector of Trenzyme GmbH. The iPLUS 3X (iPLUS in tandem three times) and iPLUSv2 (TTGTTTTTTTTGTTTTTTTGT) inserts were made by Trenzyme by primer design with flanking restriction sites followed by PCR amplification for cloning downstream of the GFP coding sequence in the *Pichia* expression vector pTZ-p04_GFP.

2.1.5 Spike

The CoV2-Spike-D614G plasmid [a gift from Jennifer Doudna, Addgene plasmid #177960; (Syed et al., 2021)], containing Spike's coding sequence from the Wuhan variant of SARS-Cov-2, was used for the preparation of the Spike's cDNA by digestion with SacI and XhoI (NEB) and creating blunt ends using Klenow. The Trastuzumab_LC_iPLUS was digested with EcoRV (NEB) and XbaI (NEB), followed by blunting with Klenow to remove trastuzumab LC and insert Spike cDNA (Spike_iPLUS). The iPLUS sequence was removed by EcoRV and XbaI digestion, followed by blunting with Klenow and re-ligation of the plasmid.

2.1.6 Spike_iPLUSv2

The Spike_iPLUS plasmid was used as backbone for the preparation of the vectors containing iPLUSv2 by removal of iPLUS by digestion with EcoRI and KpnI. For the iPLUSv2 preparation, pGEM-4 (Promega) digested with SacI and KpnI was used to subclone the annealed oligos containing iPLUSv2 sequence (Supplementary Table S1) flanked by digested SacI and KpnI restriction sites (pGEM-4-iPLUSv2). A PCR amplification of the pGEM4_iPLUSv2 was performed to obtain iPLUSv2 using pGEM4-F and pGEM4-R primers (Supplementary Table S1) and the PCR product was digested with EcoRI and KpnI and cloned in the digested Spike_iPLUS plasmid backbone to obtain Spike_iPLUSv2.

2.1.7 MAGEC2

The pDONR223_MAGEC2_WT [a gift from Jesse Boehm & William Hahn & David Root, Addgene plasmid #81862; (Kim et al., 2016)] was used as template to amplify *MAGEC2* by PCR using MAGEC2_F_XbaI and MAGEC2_R_EcoRV primers (Supplementary Table S1). The purified PCR product was digested with *XbaI* and *EcoRV* and cloned in the Trastuzumab_LC plasmid digested with the same restriction enzymes to remove *LC* and insert *MAGEC2*.

2.1.8 MAGEC2_iPLUSv2

An inverse PCR was performed using primers containing the iPLUSv2 sequence (MAGEC2_inv_F and MAGEC2_inv_R; Supplementary Table S1) using MAGEC2 as DNA template. PCR products were treated with *DpnI* (NEB) to eliminate parental DNA and ligated using T4 DNA ligase (NEB).

Sanger Sequencing was performed at the i3S Genomic Core Facility using the 3500 Genetic Analyzer sequencer (Applied Biosystems) to confirm that all vectors used in this study had the correct sequence.

2.2 Cell 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 37°C, with 5% CO₂, and were subcultured every 3–4 days.

ExpiCHO-S 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 37°C, with 8% CO₂ and 80% of relative humidity, with agitation at 125 rpm. When cells reached a concentration between 4×10^6 and 6×10^6 viable cells/mL, routine splitting to the desired dilution was performed, following the manufacturer's procedure.

2.3 Cell transfection

HeLa cells were transfected with the different vectors using Lipofectamine 2000 reagent (Thermo Fisher Scientific). For that, HeLa cells were seeded in 24-well plates, in a concentration that allowed them to reach 70%–90% confluency in the following day. For transfections, 1 µL of Lipofectamine 2000 and 500 ng of each plasmid was diluted in Opti-MEM (Thermo Fisher Scientific), following the manufacturer's protocol. Cells were allowed to grow at 37°C for 48 h, after which they were collected and processed, either for total RNA or protein extraction. ExpiCHO-S cells were transfected using ExpiFectamine CHO reagent (Thermo Fisher Scientific), following the manufacturer's protocol. In the day prior to transfection (day -1), cells were diluted to a final density of 4×10^6 viable cells/mL, in order to reach a density of 7×10^6 – 10×10^6 viable cells/mL in the following day (day 0). On the day of transfection, cell viability was assessed, and cells were split to a concentration of 6×10^6 viable cells/mL in a 25 mL final volume.

Cells were transfected with 20 µg of plasmid DNA in a final volume of 20 µL and, for Trastuzumab constructs, a 2:1 LC:HC ratio was followed. Following the Standard Protocol for transfection, ExpiFectamine CHO Enhancer and ExpiCHO Feed (Thermo Fisher Scientific) were added the day after transfection (day 1) and cells were kept at 37°C with 8% CO₂ for the remaining 9 days. At the time of cell harvest for downstream analysis, cell viability was at least 75%.

2.4 RNA extraction

Total RNA extraction was performed using TRIzol Reagent (Ambion by Life Technologies) following the manufacturer's standard protocol, with the exception that all RNA precipitations were carried out at -80°C, during at least 2 h, and using 1 µL of glycogen (15 mg/mL; Roche). Quantification and quality assessment of RNA was performed using the Nanodrop 1000 Spectrophotometer (Thermo Fisher Scientific). RNA samples were stored at -80°C until further use.

2.5 DNase I treatment

To eliminate any potential contamination from genomic and/or plasmid DNA in the RNA samples, these were incubated with DNase I (Roche) for 2 h at 37°C following the manufacturer protocol.

2.6 cDNA synthesis

Following DNase I treatment, 1 µL of Random Hexamers (50 µM) and 1 µL of dNTPs mix (10 nM each) (Thermo Fisher Scientific) were added and samples were incubated at 65°C for 5 min followed by 5 min at 4°C. After that, 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 negative control to detect genomic DNA, the same mixtures were prepared for each sample, but without the addition of SSIV RT (-RT). Samples were incubated for 10 min at 23°C, 10 min at 55°C, and finally, 10 min at 80°C to inactivate the enzyme. Samples were stored at -20°C until further use.

2.7 Reverse transcription quantitative real-time PCR (RT-qPCR)

RT-qPCRs were prepared using 1 µL of cDNA and 9 µL of a mixture containing 5 µL of 2x SYBR Select Master Mix (Applied Biosystems), 0.125 µL (0.125 µM) of each primer (listed in Supplementary Table S1) and 3.75 µL of nuclease-free water. Samples were run in triplicate on a 7500 Fast Real-Time PCR System (Applied Biosystems), using the program: 50°C for 2 min, 95°C for 2 min; 40 cycles at 95°C for 15 s, 60°C for 1 min followed by a melt curve stage. Non-template controls and -RT samples were

also used in each experiment. Results were analyzed using the $\Delta\Delta Ct$ method (Normalized fold expression) using *ACTB* (for ExpiCHO-S cells) or *18S* (for HeLa cells) reference genes to normalize gene expression (Pfaffl, 2001).

2.8 Enzyme-linked immunosorbent assay (ELISA)

For recombinant trastuzumab protein quantification, the human IgG (total) uncoated ELISA kit with plates (Invitrogen) was used. For sample preparation, ExpiCHO-S cells were collected and centrifuged at 4,000 g for 30 min. The supernatant containing the recombinant protein was recovered and used for the ELISA protocol using 1:250 or 1:500 dilutions. For each condition, duplicates were made and non-transfected cells were used as negative controls. The ELISA manufacturer's protocol was followed and the 96-well plate was read at 450 nm and 630 nm in a Synergy2 plate reader (BioTek).

2.9 Western blot

Whole-cell protein extracts were obtained by cell lysis with NP-40 cell lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40), supplemented with 1x protease inhibitor cocktail (Sigma, P8340). Protein extracts were quantified using the Quick Bradford reagent (Bio-Rad) and 20–40 μ g were added to Laemmli Buffer 2x (Bio-Rad) with 5% of β -mercaptoethanol in a 1:1 ratio, and boiled for 5 min at 95°C. Samples were run in a 3% stacking and 10% resolving Tris-glycine SDS-polyacrylamide gel and transferred into a nitrocellulose membrane, using a dry transfer method with the iBlot Gel Transfer Device (Thermo Fisher Scientific). The membrane was incubated with the respective primary antibody diluted in TBS-0.1% Tween 20 containing 5% non-fat dried milk, followed by three washes with TBS-0.1% Tween 20 and incubation with the appropriate secondary antibody. The antibodies used in this study were: anti-SARS-CoV-2 Spike mouse monoclonal antibody (1A9; GeneTex; 1:2,000 dilution), anti- α -tubulin mouse monoclonal antibody (B-5-1-2; Sigma; 1:150,000 dilution), anti-GAPDH mouse monoclonal antibody (ab8245; Abcam; 1:1,500,000 dilution), recombinant anti-MAGEC2 antibody (ab209667; Abcam; 1:2,500 dilution), goat anti-mouse IgG-HRP secondary antibody (Santa Cruz Biotechnologies; 1:10,000 dilution) and goat anti-rabbit IgG-HRP secondary antibody (Santa Cruz Biotechnologies; 1:10,000 dilution). Detection was achieved using enhanced chemiluminescence (ECL Prime, GE Healthcare) and the immunoblots were visualized on the ChemiDoc XRS + System (Bio-Rad). Quantification of the protein intensity was performed using the Image lab Software (Bio-Rad).

2.10 Recombinant GFP protein expression in *Pichia pastoris*

These experiments were done as a subcontracted service by Trenzyme GmbH (Konstanz, Germany). After transformation of the expression vectors pTZ_P04_GFP (control plasmid without iPLUS

sequences), pTZ_P04_GFP_iPLUS 3X and pTZ_P04_GFP_iPLUSv2 into *Pichia pastoris* strain KM71H, positive integrants (stable clones) were selected by their ability to grow in presence of Zeocin. After confirmation of proper integration at the AOX locus by PCR based methods, five stable clones per construct were analyzed in duplicates for their ability to synthesize GFP after induction with methanol (0.5% methanol every 24 h) in a 72 h experiment. Screening process was done in a microbioreactor system (BioLector, equipped with 48-well “flower plates”) in 48 -well format, sample volume up to 1,000 μ L per well in all clones in parallel.

2.11 Statistical analyses

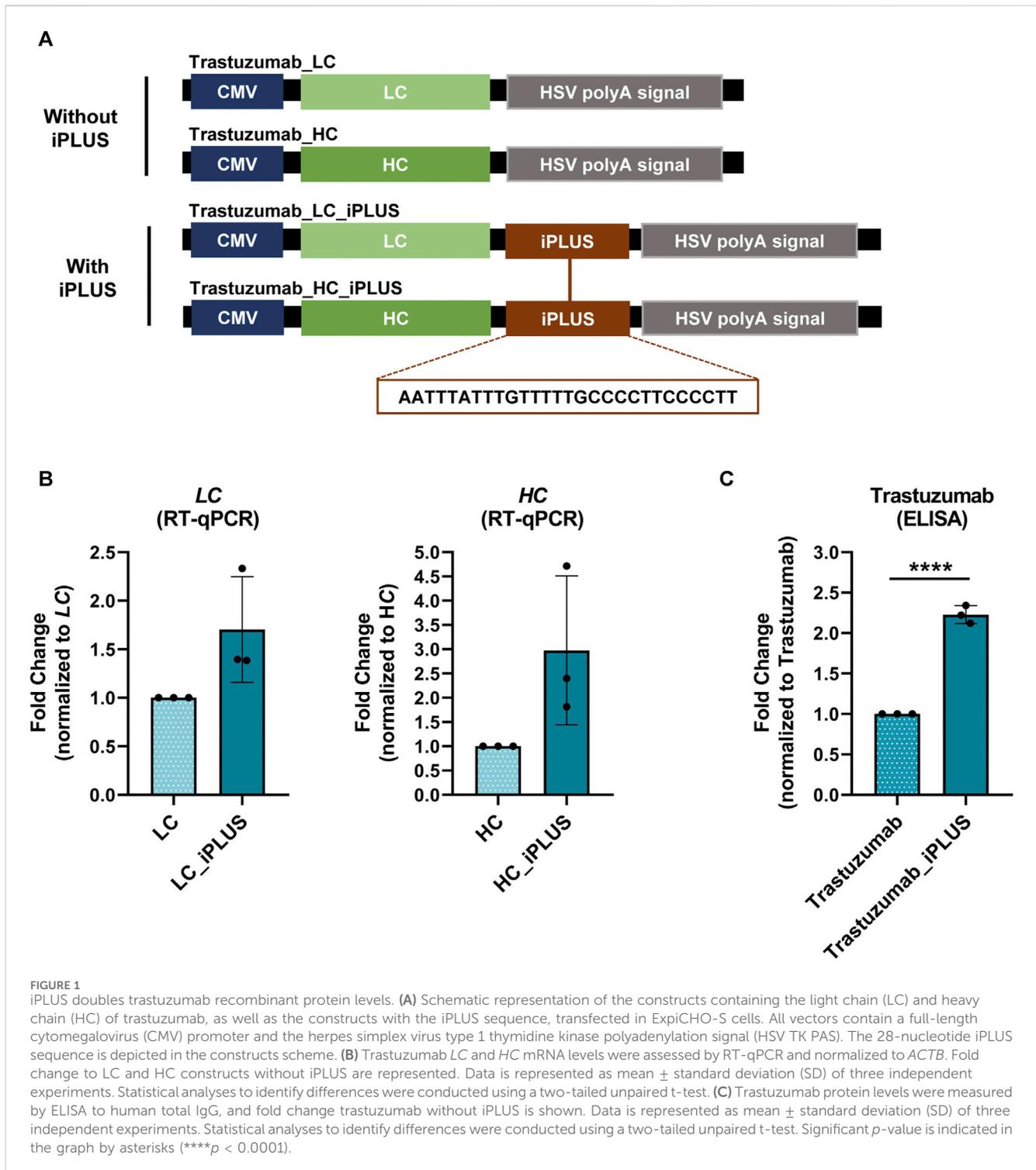
Data are presented in the graphs as mean \pm SD (standard deviation) and are from three independent experiments. Statistical differences were tested using two-tailed unpaired t-test. Only *p*-values <0.05 were considered statistically significant. All graphs were done using GraphPad Prism version 8.01 (GraphPad Software).

3 Results

3.1 Incorporation of iPLUS in a mammalian expression vector doubles the production of trastuzumab, a recombinant monoclonal antibody used in cancer immunotherapy

The value of the biopharmaceutical industry has been pushed by the recombinant monoclonal antibodies (mAbs) market (Walsh, 2018). Therefore, we applied iPLUS to the production of trastuzumab, a human recombinant monoclonal antibody used in cancer immunotherapy (Albanell and Baselga, 1999; Krasniqi et al., 2019). As other mAbs, trastuzumab is composed of four polypeptide chains: two identical light chains (LC) and two identical heavy chains (HC), bound by covalent bonds, mainly disulfide bonds, forming a heterodimeric Y-shaped structure (Castelli et al., 2019). We started by subcloning trastuzumab LC and HC chains separately in the pcDNA3.4 expression vector (Figure 1A). We chose pcDNA3.4 as it is widely used in industrial settings for monoclonal antibodies production (Kosuge et al., 2020). This vector contains a full-length CMV promoter that drives high levels of expression, and was modified by the vector's industry (Thermo Fisher Scientific) to include the WPRE downstream of the multiple cloning site, which increases the expression of several proteins, such as GFP and luciferase (Zufferey et al., 1999).

Our hypothesis was that the inclusion of the iPLUS in the pcDNA3.4 expression vector could further improve the yield of production. We thus incorporated in the LC and HC containing vectors the 28-nucleotide long iPLUS non-coding sequence downstream of WPRE and upstream of the herpes simplex virus (HSV) type 1 thymidine kinase (TK) PAS (Figure 1A). Co-transfection of LC:HC antibody chains containing vectors with and without iPLUS was performed in ExpiCHO-S cells and antibody protein levels were quantified by an IgG ELISA assay and mRNA levels evaluated by RT-qPCR. ExpiCHO-S cells were selected as they are commonly used in the biopharmaceutical



industries for therapeutic recombinant protein production due to their ability to survive in animal-component-free medium and grow in suspension at high densities. As can be observed in Figure 1B, vectors containing iPLUS result in an increase in LC and HC mRNA levels, as well as in a 2.2-fold increase in trastuzumab protein levels (Figure 1C). These results show that iPLUS doubles the production of this important therapeutic monoclonal antibody at the laboratory scale.

3.2 iPLUS 3X and the iPLUS variant sequence iPLUSv2 highly improve recombinant protein expression levels in *Pichia pastoris*

The next step was to evaluate the use of iPLUS in yeast. *Pichia pastoris* may use methanol as a carbon source, allowing the induction of protein expression in a highly controlled manner through the AOX1 promoter, making it especially suitable for

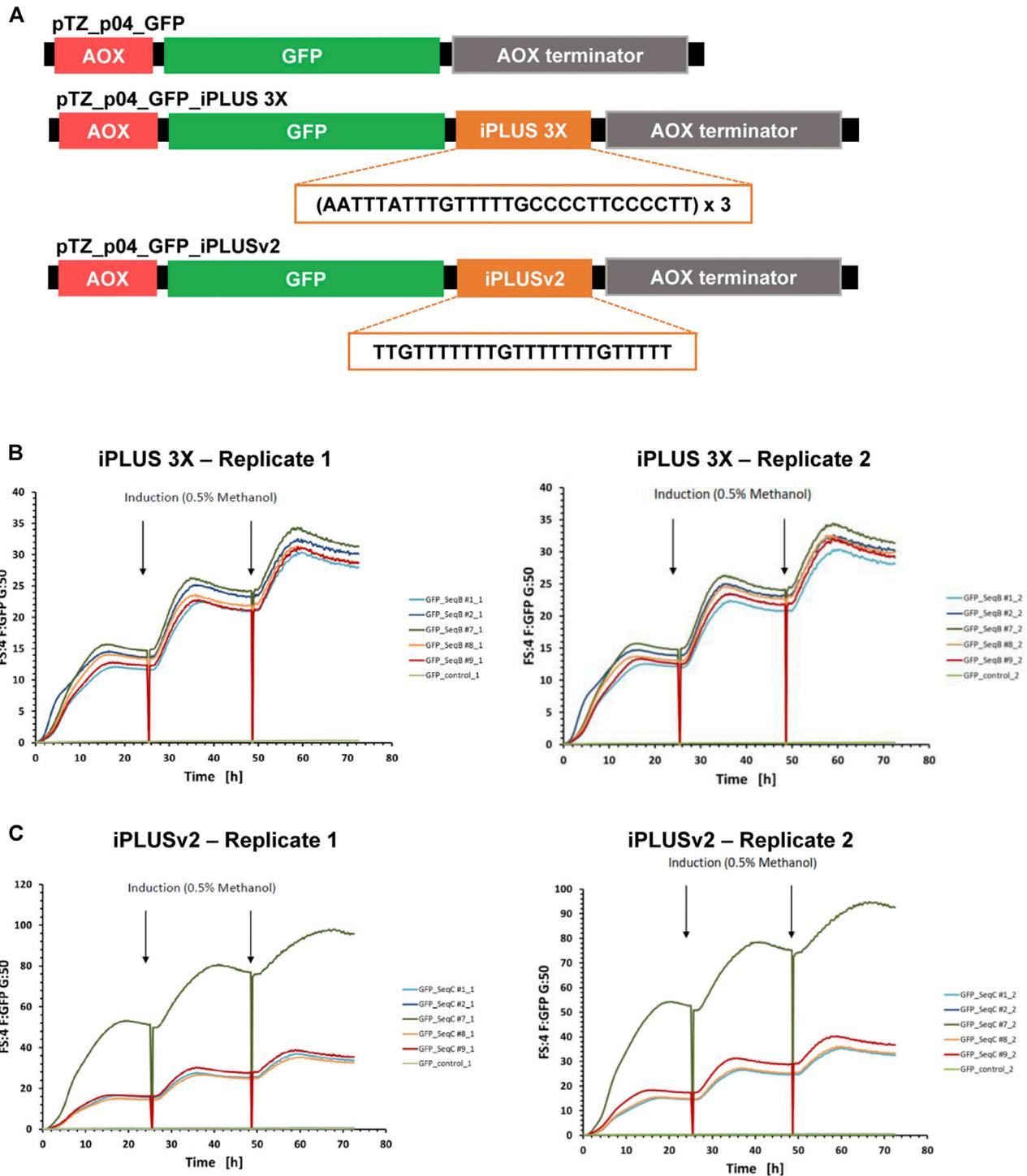


FIGURE 2
iPLUS 3X and iPLUSv2 boost GFP expression levels in *Pichia pastoris*. **(A)** Scheme of the *Pichia pastoris* pTZ_p04 expression constructs containing the GFP coding sequence, as well as the constructs with the iPLUS sequence in tandem 3X and the iPLUS variant iPLUSv2. All pTZ_p04 vectors contain a methanol inducible alcohol oxidase gene I (AOX) promoter and an AOX terminator. The 84-nucleotide iPLUS 3X sequence and the 24-nucleotide iPLUSv2 sequence introduced downstream the GFP coding sequence are depicted in the constructs. **(B)** Both iPLUS 3X and **(C)** iPLUSv2 increase GFP expression in *Pichia*. Five different *Pichia* stable clones for iPLUS 3X **(B)** and iPLUSv2 **(C)** were analysed in duplicates (replicates 1 and 2) in a microbioreactor system for their ability to synthesize GFP after induction with 0.5% methanol at 24 and 48 h time-points. iPLUS containing constructs show a strong increase in GFP expression, 114-fold for iPLUS 3X and 303-fold for iPLUSv2, in the best performing clones (clones #7), in comparison with the control clone expressing GFP (GFP-control; light green).

large-scale and industrial applications. In fact, *Pichia* is widely used by the biopharmaceutical industry to produce recombinant proteins (Cregg et al., 2000).

We used the GFP reporter protein as a read-out for the use of iPLUS in *Pichia pastoris*, and started by making GFP constructs with and without iPLUS (GFP-control) sequences by gene synthesis. The iPLUS constructs contained the iPLUS sequence in tandem 3X (iPLUS 3X), to test if the inclusion of several iPLUS repeats would increase protein levels, and a variant sequence of iPLUS (iPLUSv2), which corresponds to the most conserved region of the iPLUS contained in *Drosophila's polo* gene (Oliveira et al., 2019), also repeated 3 times (Figure 2A). These sequences were then cloned into a *Pichia* expression vector having GFP under the control of methanol inducible AOX promoter (Figure 2A). After *Pichia* transformation with the different constructs, five stable clones (positive integrants) of each construct were selected and analyzed for their ability to synthesize GFP after induction with methanol. The parallel screening of all clones in a microbioreactor system demonstrates that both iPLUS 3X (Figure 2B) and iPLUSv2 (Figure 2C) can significantly increase GFP expression in *Pichia*. In comparison to the control GFP without any iPLUS sequence (Supplementary Figure S1), the best performing clones showed an increase in GFP expression of 114-fold for iPLUS 3X and 303-fold for iPLUSv2.

This proof-of-concept experiment underscore the versatility and efficacy of iPLUS and reveals that different iPLUS variants, such as the new iPLUSv2, enhance protein production in *Pichia pastoris* expression systems. Furthermore, it clearly shows that iPLUS may be a valuable asset for the improvement of *Pichia* expression vectors with biopharmaceutical applications.

3.3 iPLUSv2 enhances Spike's mRNA and protein levels in different cell lines

In the post COVID-19 pandemic landscape, mRNA vaccines are positioned as a dynamic and groundbreaking technology platform for addressing global health threats (Lorentzen et al., 2022). However, moving from a novel technology to a cornerstone of vaccination strategies in a short time opens space for improvements, such as in the capacity of increasing antigen production levels.

Given our results in improving trastuzumab expression in mammalian cells, as well as with the new variant iPLUSv2 in *Pichia*, we hypothesized that the inclusion of the new activator sequence, iPLUSv2, in the 3' end of Spike's coding sequence could be a valuable asset for the mRNA vaccine technology. In fact, the inclusion of *cis*-regulatory sequences in the 3' end of Spike was also applied by Moderna and Pfizer in their COVID-19 mRNA vaccines (Xia, 2021).

As in the trastuzumab experiments, the pcDNA3.4 constructs were used (Figure 3A), with iPLUSv2 downstream of Spike. We tested two different cell lines, HeLa and ExpiCHO-S. Human HeLa cells are excellent models for proof-of-concept experiments, due to their easiness in manipulation and analyses. ExpiCHO-S cells were selected not only because they are commonly used by the industry but also as iPLUS increases trastuzumab expression in these cells (Figure 1).

As observed previously for *Pichia* (Figure 2), in HeLa cells the inclusion of iPLUSv2 leads to an increase in both mRNA and protein levels of Spike (Figures 3B–D), in comparison to the control vector without iPLUSv2. For the protein levels, this increase is approximately 2.1-fold, while for the mRNA levels the increase is approximately 1.4-fold. In ExpiCHO-S cells, iPLUSv2 promoted, once again, an increase in both Spike's mRNA and protein levels, as illustrated in Figures 3E–G. Here, the increase in protein production is approximately 2.5-fold, while for the mRNA levels the increase is approximately 1.6-fold.

These results clearly show that iPLUSv2 can increase Spike's mRNA levels and double protein production, in two different cell lines from different species (human and hamster).

3.4 MAGEC2 mRNA and protein levels increase with iPLUSv2

The mRNA vaccines are a potential immunotherapy solution to combat various types of cancer, prompting the immune system to destroy cancer cells by triggering a specific and enduring immune response against tumor antigens (TAs), inhibiting tumor growth (Miao et al., 2021). Nevertheless, in clinical trials it has been observed that although there is a certain level of positive response from patients (Papachristofilou et al., 2019), it does not seem to be a sufficiently robust response to cause tumor recession or an increase in overall survival for all the patients (Chehelgerdi and Chehelgerdi, 2023), leading to termination of their production. For that reason, pharmaceutical companies ought to incorporate novel molecular tools to improve cancer mRNA vaccines, to promote good immunological responses.

Therefore, our next step was to assess whether subcloning iPLUSv2 downstream of the melanoma antigen family C2 (MAGEC2) would result in increased mRNA levels and consequent protein production. For these proof-of-concept experiments, MAGEC2, a tumor-specific antigen for non-small cell lung cancer (NSCLC), was selected since it is included in CureVac's CV9202 NSCLC mRNA vaccine and whose information was made public (Papachristofilou et al., 2019). CV9202 is a vaccine encoding 6 tumor associated antigens, being MAGEC2 one of them. Antigen-specific immune responses were detected in a phase Ib study enrolling 26 patients with stage IV NSCL (Papachristofilou et al., 2019; Sebastian et al., 2019).

As before, the initial step was the construction of vectors containing the MAGEC2 with and without iPLUSv2 downstream of this antigen (Figure 4A) and these vectors were used to transfect HeLa cells. As it can be observed in Figures 4B, C, the inclusion of iPLUSv2 sequence in MAGEC2 vectors resulted in an increase in approximately 1.8-fold in MAGEC2 mRNA levels (Figure 4B) and a 2.8-fold increase in MAGEC2 protein levels, in comparison to vectors without this sequence (Figures 4C, D).

As observed with iPLUSv2 in Spike or iPLUS in trastuzumab, these results show that the presence of iPLUS non-coding sequences in vectors used for recombinant protein expression can at least double protein production. Moreover, the impact of iPLUSv2 in MAGEC2 bears substantial implications for its application in cancer mRNA vaccines, as the increase in antigen production certainly influences the efficiency of the process.

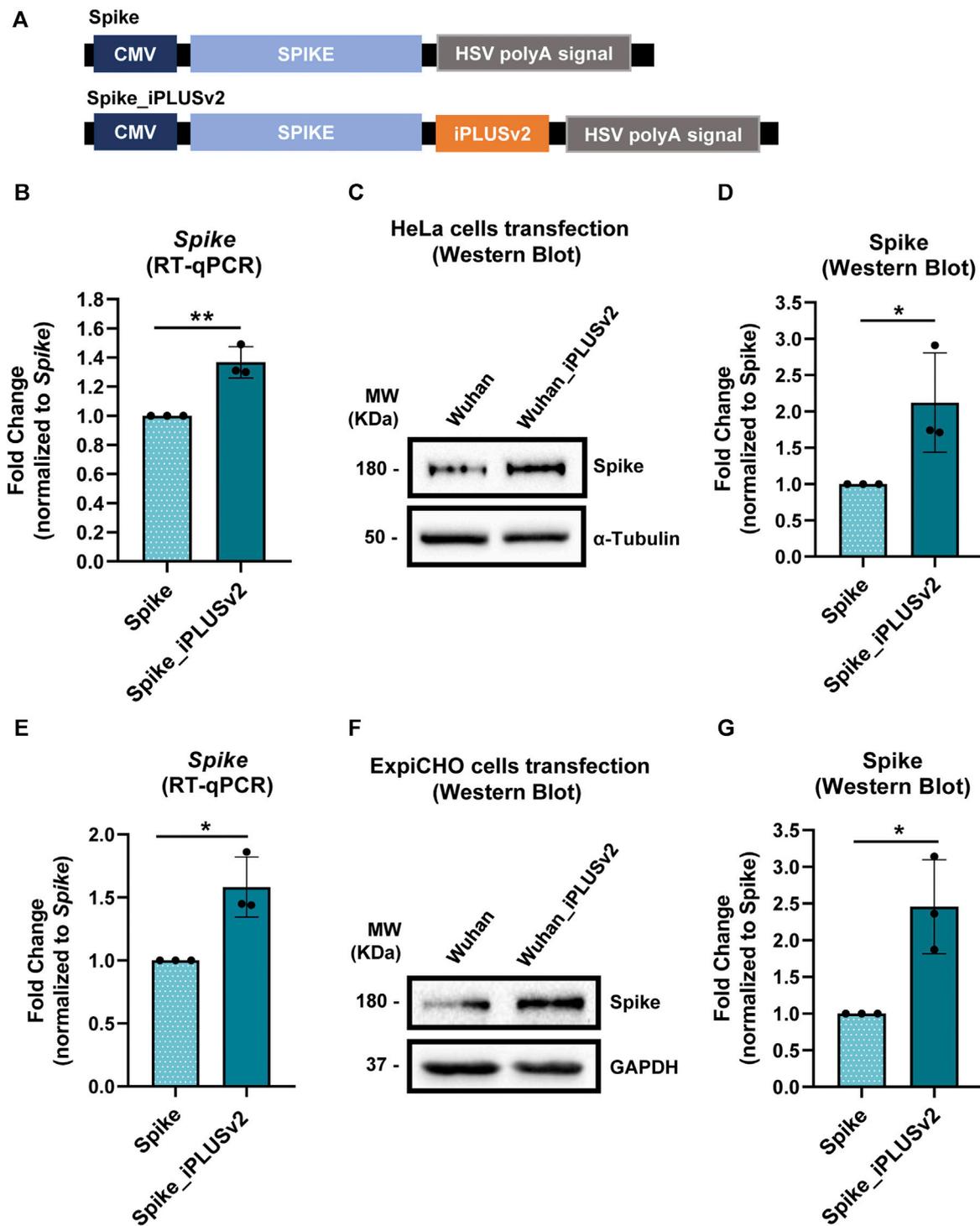


FIGURE 3

iPLUSv2 leads to an increase in Spike's mRNA levels and protein production, in both HeLa and ExpiCHO-S cell lines. (A) Schematic representation of the constructs containing Spike's cDNA and the iPLUS variant, iPLUSv2. Additionally, these vectors contain a full-length CMV promoter and the HSV TK PAS. These vectors were used for HeLa and ExpiCHO-S cells transfection, after which protein was quantified and mRNA levels were measured. (B) Spike's mRNA levels were measured by RT-qPCR after HeLa cells transfection and normalized to the housekeeping *18S*. Fold change was calculated in relation to Spike. Data are shown as mean + standard deviation (SD) for three different experiments ($n = 3$). Statistical significance was determined by two-tailed unpaired t-test. (C) Representative image of Spike's protein levels evaluated by Western blot after HeLa cells transfection, being the protein quantification levels plotted for each condition, for three independent experiments ($n = 3$) in (D). (D) Spike's protein quantification was calculated using α -Tubulin as the loading control for normalization, followed by fold change analyses, normalized to Spike. Data are shown as mean + standard deviation (SD). Statistical significance was determined by two-tailed unpaired t-test. (E) Spike's mRNA levels were measured by RT-qPCR after ExpiCHO-S cells transfection and normalized to the housekeeping *hACTB*. Fold change was calculated in relation to Spike. Data are shown as mean + standard deviation (SD) for three different experiments ($n = 3$). Statistical significance was determined by two-tailed unpaired t-test. (F) Representative image of Spike's

(Continued)

FIGURE 3 (Continued)

protein levels evaluated by Western blot after ExpiCHO-S cells transfection, being the protein quantification levels plotted for each condition, for three independent experiments ($n = 3$) in (G). (G) Spike's protein quantification was calculated using GAPDH as the loading control for normalization, followed by fold change analyses, normalized to Spike. Data are shown as mean + standard deviation (SD). Statistical significance was determined by two-tailed unpaired t-test. Significant p -value is indicated in the graph by asterisks ($*p < 0.05$ and $**p < 0.01$).

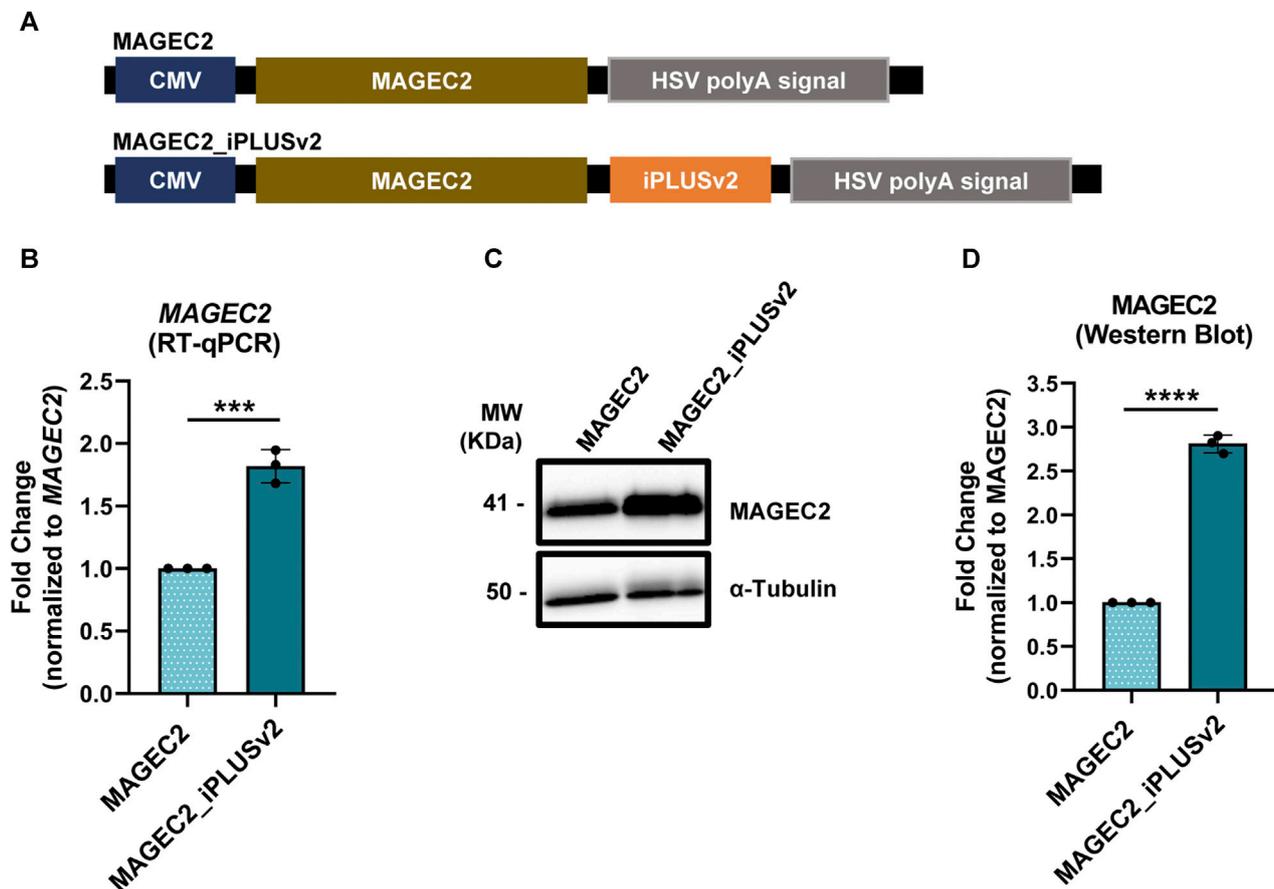


FIGURE 4

iPLUSv2 increases MAGEC2 mRNA and protein levels. (A) Schematic representation of the constructs, containing the MAGEC2 coding sequence and the iPLUS variant, iPLUSv2, transfected in HeLa cells. Vectors contain a full-length CMV promoter and the HSV TK PAS. (B) MAGEC2 mRNA levels increase in the presence of iPLUSv2. mRNA levels were measured by RT-qPCR and normalized to 18S. Fold change to MAGEC2 without iPLUSv2 is represented as mean \pm standard deviation (SD) of three independent experiments. Statistical analyses to identify differences were done using a two-tailed unpaired t-test. Significant p -value is indicated in the graph by asterisks ($***p < 0.001$). (C,D) MAGEC2 protein levels increase >2-fold in the presence of iPLUSv2. (C) Representative image of MAGEC2 protein levels evaluated by Western blot, with α -Tubulin as a loading control. (D) Quantification of MAGEC2 protein levels was calculated using the respective loading control for normalization, followed by fold change analyses to MAGEC2 without iPLUSv2. Data is represented as mean \pm standard deviation (SD) of three independent experiments. Statistical analyses to identify differences were conducted using a two-tailed unpaired t-test. Significant p -value is indicated in the graph by asterisks ($****p < 0.0001$).

4 Discussion

Cis-regulatory sequences present in the 3'UTR that impact both mRNA and protein production, should be further investigated to be applied to enhance the production of biopharmaceuticals (O'Flaherty et al., 2020; Kirshina et al., 2023).

The existence of U-rich USEs as auxiliary regulatory elements has been widely recognized since they were originally described in viral mRNAs (DeZazzo and Imperiale, 1989; Russnak and Ganem, 1990; Darmon and Lutz, 2012). In the Human Immunodeficiency Virus 1 (HIV-1), for example, different termination sites exist and it

is a USE in the vicinity of the 3' long terminal repeats (LTR) PAS that aids in the production of the correct length transcript, disregarding the PAS present in the 5'LTR (Valsamakis et al., 1991). It was initially thought that these sequences were specific to viral genomes, but this notion was questioned when different examples in cellular genes were discovered, such as in complement C2, Lamin B2, cyclooxygenase-2 or prothrombin, in which USEs mode of action has been described (Moreira et al., 1995; Brackenridge et al., 1997; Hall-Pogar et al., 2005; Danckwardt et al., 2011). For instance, we have showed that polo's USE in *Drosophila* is essential for polo distal PAS selection and has a

function in increasing *polo* mRNA and protein levels (Pinto et al., 2011; Oliveira et al., 2019). Moreover, microinjection of GFP-USE in one-cell stage zebrafish embryos increases GFP levels demonstrating a conservation of the role of the USE in increasing protein levels *in vivo* (Eufrásio et al., 2023). We then tested if the USE could function downstream of a reporter gene in a mammalian cell model, thus showing that the USE is also able to increase luciferase activity in HeLa cells (Oliveira et al., 2020). The results obtained with *polo*'s USE *in vivo* in two animal models and *in vitro* in HeLa cells prompt us to patent the use of this sequence in several applications including in the production of recombinant proteins with therapeutic use and mRNA vaccines and re-named it as iPLUS (Oliveira et al., 2020). We have previously described the mechanism of action of USEiPLUS and showed that iPLUS regulates several steps of gene expression. By liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis we demonstrated that PTBP1 binds to iPLUS RNA and using PTBP1 mutants we showed that it increases mRNA 3' end formation, and consequently, mRNA levels (Oliveira et al., 2019). Additionally, we showed by UV crosslinking and immunoprecipitation assays that HuR (Brennan and Steitz, 2001; Hinman and Lou, 2008; Ho et al., 2020), a RBP with a well-defined role in mRNA stability and translation efficiency, also binds to *in vitro* transcribed iPLUS RNA, thus increasing protein levels (Eufrásio et al., 2023). Despite the fact that USE's/iPLUS physiological roles and mechanistic mode of function have been thoroughly investigated, this is the first time that a USE/iPLUS has been employed with the purpose of boosting protein production in different expression systems.

Here, we have expanded upon our earlier findings by exploring the potential of the non-coding iPLUS sequence to enhance the production of relevant biopharmaceuticals, including monoclonal antibodies used in immunotherapy and antigens employed in mRNA vaccines. Our results show the versatility of iPLUS by demonstrating its efficacy in improving the production of relevant recombinant proteins across various models. We have showed that in a recombinant monoclonal antibody commonly used in cancer immunotherapy, trastuzumab, iPLUS increases *LC* and *HC* mRNA levels and doubles trastuzumab protein levels in ExpiCHO-S cells. Next, we showed that in *Pichia*, a strain of yeast extensively used in recombinant protein production, the inclusion of iPLUS and a new iPLUS variant sequence in *Pichia* expression vectors containing GFP strongly impacts GFP expression levels, in at least 100-fold. Upon these consistent results, we turned our attention to a different kind of biopharmaceutical, in high demand at the moment, mRNA vaccines. For Spike, the antigen used in the COVID-19 mRNA vaccines, we observed again an increase in both mRNA and protein levels, in two different models, HeLa and ExpiCHO-S cells. The same effect was observed when we evaluated a different antigen tested in a lung cancer mRNA vaccine, MAGEC2. These results clearly show that in the three tested models (HeLa, ExpiCHO-S and *Pichia*) and with various proteins, iPLUS or its variant sequence iPLUSv2, consistently exhibit a beneficial impact in increasing both mRNA and protein levels.

These results produce proof-of-concept data of the applicability of the use of iPLUS in the biopharmaceutical industry and pave the way for further scale up experiments, in an industrial context. In yeast, there is evidence that the regulatory role of sequences present

in the 3'UTR is conserved (Shalgi et al., 2005; Ito et al., 2020; Savinov et al., 2021), but this is still a rather unexplored field of research. When referring specifically to *Pichia pastoris*, the knowledge is even more limited, so there is still room for improvements in this area, as we have demonstrated here by applying iPLUS to this expression system. This also opens the possibility for applying this technology to other recombinant DNA technologies, to increase the production of antibodies for immunotherapy, different mRNA vaccines antigens and possibly other therapeutic proteins, such as hormones or cytokines. Furthermore, it is possible that iPLUS presents a synergistic effect with existing non-coding regulatory sequences located at the 3' end of expression systems, such as the WPRE, suggesting new applications for iPLUS and offering an opportunity to further enhance commercially available vectors.

Indeed, both iPLUS and iPLUSv2 offer numerous valuable advantages for their inclusion in vectors used in an industrial environment. As non-coding sequences, they avoid interference with the coding region of the protein or antigen of interest, avoiding additional regulatory steps. Their short size (28 or 25 nucleotides) facilitates easy subcloning into various expression systems or vectors without requiring complex laboratory procedures. Furthermore, iPLUS proves to be a versatile asset, as it functions across different expression systems, proteins, and mRNAs thus enabling the selection of a system tailored to the unique characteristics of each mRNA and protein (Oliveira et al., 2020).

In summary, iPLUS underscores the importance of the *cis*-acting elements present in the 3'UTR in the regulation of gene expression, highlighting its potential application for the development of different vectors and biologics. Our findings clearly show that iPLUS incorporation into vectors used in different models has the potential to be used by the biopharmaceutical industry to increase the yield of protein production, thereby reducing production costs.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used. Ethical approval was not required for the studies on animals in accordance with the local legislation and institutional requirements because only commercially available established cell lines were used.

Author contributions

IR-C: Investigation, Methodology, Writing—original draft, Writing—review and editing. MS: Investigation, Methodology, Writing—original draft, Writing—review and editing. AnM: Investigation, Methodology, Writing—original draft, Writing—review and editing. BG: Investigation, Methodology,

Writing—original draft, Writing—review and editing. IP-C: Investigation, Methodology, Supervision, Writing—original draft, Writing—review and editing. AIM: Conceptualization, Funding acquisition, Project administration, Supervision, Writing—original draft, Writing—review and editing.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by Programa Gilead Génese, ref 13854; by CaixaImpulse, la Caixa Foundation (LCF/TR/CI19/52460018); by National Funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., under the project UIDB/04293/2020; by FCT under the project EXPL/SAU-PUB/1073/2021 (<http://doi.org/10.54499/EXPL/SAU-PUB/1073/2021>) and by Programa Operacional Regional do Norte and co-funded by European Regional Development Fund under the project “The Porto Comprehensive Cancer Center” with the reference NORTE-01-0145-FEDER-072678—Consórcio PORTO. CCC—Porto Comprehensive Cancer Center. IP-C is funded by a Researcher contract DOI: 10.54499/DL57/2016/CP1355/CT0016.

Acknowledgments

We are very thankful to all the current and past members of the Gene Regulation group, as well as to our academic and industrial

colleagues for their insightful discussions. We are grateful to the Research and Innovation Unit at i3S for their support, to Mariana Coelho as an industry liaison and to Trenzyme GmbH as a service provider in the *Pichia* experiments.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2024.1355957/full#supplementary-material>

References

- Albanell, J., and Baselga, J. (1999). Trastuzumab, a humanized anti-HER2 monoclonal antibody, for the treatment of breast cancer. *Drugs Today (Barc)* 35 (12), 931–946. doi:10.1358/dot.1999.35.12.564040
- Brackenridge, S., Ashe, H. L., Giacca, M., and Proudfoot, N. J. (1997). Transcription and polyadenylation in a short human intergenic region. *Nucleic Acids Res.* 25 (12), 2326–2335. doi:10.1093/nar/25.12.2326
- Brennan, C. M., and Steitz, J. A. (2001). HuR and mRNA stability. *Cell Mol. Life Sci.* 58 (2), 266–277. doi:10.1007/pl000000854
- Cao, J., Novoa, E. M., Zhang, Z., Chen, W. C. W., Liu, D., Choi, G. C. G., et al. (2021). High-throughput 5' UTR engineering for enhanced protein production in non-viral gene therapies. *Nat. Commun.* 12 (1), 4138. doi:10.1038/s41467-021-24436-7
- Carswell, S., and Alwine, J. C. (1989). Efficiency of utilization of the simian virus 40 late polyadenylation site: effects of upstream sequences. *Mol. Cell Biol.* 9 (10), 4248–4258. doi:10.1128/mcb.9.10.4248
- Castelli, M. S., McGonigle, P., and Hornby, P. J. (2019). The pharmacology and therapeutic applications of monoclonal antibodies. *Pharmacol. Res. Perspect.* 7 (6), e00535. doi:10.1002/prp2.535
- Castillo-Hair, S. M., and Seelig, G. (2022). Machine learning for designing next-generation mRNA therapeutics. *Acc. Chem. Res.* 55 (1), 24–34. doi:10.1021/acs.accounts.1c00621
- Chan, S., Choi, E. A., and Shi, Y. (2011). Pre-mRNA 3'-end processing complex assembly and function. *Wiley Interdiscip. Rev. RNA* 2 (3), 321–335. doi:10.1002/wrna.54
- Chehelgerdi, M., and Chehelgerdi, M. (2023). The use of RNA-based treatments in the field of cancer immunotherapy. *Mol. Cancer* 22 (1), 106. doi:10.1186/s12943-023-01807-w
- Cregg, J. M., Cereghino, J. L., Shi, J., and Higgins, D. R. (2000). Recombinant protein expression in *Pichia pastoris*. *Mol. Biotechnol.* 16 (1), 23–52. doi:10.1385/mb:16:1:23
- Danckwardt, S., Gantzer, A. S., Macher-Goeppinger, S., Probst, H. C., Gentzel, M., Wilm, M., et al. (2011). p38 MAPK controls prothrombin expression by regulated RNA 3' end processing. *Mol. Cell* 41 (3), 298–310. doi:10.1016/j.molcel.2010.12.032
- Danckwardt, S., Kaufmann, I., Gentzel, M., Foerster, K. U., Gantzer, A. S., Gehring, N. H., et al. (2007). Splicing factors stimulate polyadenylation via USEs at non-canonical 3' end formation signals. *Embo J.* 26 (11), 2658–2669. doi:10.1038/sj.emboj.7601699
- Darmon, S. K., and Lutz, C. S. (2012). Novel upstream and downstream sequence elements contribute to polyadenylation efficiency. *RNA Biol.* 9 (10), 1255–1265. doi:10.4161/rna.21957
- DeZazzo, J. D., and Imperiale, M. J. (1989). Sequences upstream of AAUAAA influence poly(A) site selection in a complex transcription unit. *Mol. Cell Biol.* 9 (11), 4951–4961. doi:10.1128/mcb.9.11.4951
- Dingermann, T. (2008). Recombinant therapeutic proteins: production platforms and challenges. *Biotechnol. J.* 3 (1), 90–97. doi:10.1002/biot.200700214
- Dodev, T. S., Karagiannis, P., Gilbert, A. E., Josephs, D. H., Bowen, H., James, L. K., et al. (2014). A tool kit for rapid cloning and expression of recombinant antibodies. *Sci. Rep.* 4, 5885. doi:10.1038/srep05885
- Eufrásio, A., Azevedo, J., Machado, J., Ferreira, A., Moutinho, A., Henriques, F., et al. (2023). Unravelling the regulatory function of a short 3'UTR sequence in zebrafish and in human cells. *bioRxiv*. doi:10.1101/2023.11.15.567165
- Gómez-Aguado, I., Rodríguez-Castejón, J., Vicente-Pascual, M., Rodríguez-Gascón, A., Solinis, M., and Del Pozo-Rodríguez, A. (2020). Nanomedicines to deliver mRNA: state of the art and future perspectives. *Nanomater. (Basel)* 10 (2), 364. doi:10.3390/nano10020364
- Gruber, A. J., and Zavolan, M. (2019). Alternative cleavage and polyadenylation in health and disease. *Nat. Rev. Genet.* 20 (10), 599–614. doi:10.1038/s41576-019-0145-z
- Hall-Pogar, T., Zhang, H., Tian, B., and Lutz, C. S. (2005). Alternative polyadenylation of cyclooxygenase-2. *Nucleic Acids Res.* 33 (8), 2565–2579. doi:10.1093/nar/gki544
- Hernández, G., Osnaya, V. G., and Pérez-Martínez, X. (2019). Conservation and variability of the AUG initiation codon context in eukaryotes. *Trends Biochem. Sci.* 44 (12), 1009–1021. doi:10.1016/j.tibs.2019.07.001
- Hinman, M. N., and Lou, H. (2008). Diverse molecular functions of Hu proteins. *Cell Mol. Life Sci.* 65 (20), 3168–3181. doi:10.1007/s00018-008-8252-6
- Ho, J. J. D., Balukoff, N. C., Theodoridis, P. R., Wang, M., Krieger, J. R., Schatz, J. H., et al. (2020). A network of RNA-binding proteins controls translation efficiency to activate anaerobic metabolism. *Nat. Commun.* 11 (1), 2677. doi:10.1038/s41467-020-16504-1
- Ito, Y., Terai, G., Ishigami, M., Hashiba, N., Nakamura, Y., Bamba, T., et al. (2020). Exchange of endogenous and heterogeneous yeast terminators in *Pichia pastoris* to tune mRNA stability and gene expression. *Nucleic Acids Res.* 48 (22), 13000–13012. doi:10.1093/nar/gkaa1066

- Jackson, N. A. C., Kester, K. E., Casimiro, D., Gurunathan, S., and DeRosa, F. (2020). The promise of mRNA vaccines: a biotech and industrial perspective. *NPJ Vaccines* 5, 11. doi:10.1038/s41541-020-0159-8
- Jackson, R. J., Hellen, C. U. T., and Pestova, T. V. (2010). The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11 (2), 113–127. doi:10.1038/nrm2838
- Jiang, Y., Xu, X. S., and Russell, J. E. (2006). A nucleolin-binding 3' untranslated region element stabilizes β -globin mRNA *in vivo*. *Mol. Cell Biol.* 26 (6), 2419–2429. doi:10.1128/mcb.26.6.2419-2429.2006
- Johnson, A. G., Grosely, R., Petrov, A. N., and Puglisi, J. D. (2017). Dynamics of IRES-mediated translation. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 372, 20160177. doi:10.1098/rstb.2016.0177 (1716)
- Karikó, K., Kuo, A., and Barnathan, E. (1999). Overexpression of urokinase receptor in mammalian cells following administration of the *in vitro* transcribed encoding mRNA. *Gene Ther.* 6 (6), 1092–1100. doi:10.1038/sj.gt.3300930
- Kesik-Brodacka, M. (2018). Progress in biopharmaceutical development. *Biotechnol. Appl. Biochem.* 65 (3), 306–322. doi:10.1002/bab.1617
- Kim, E., Ilic, N., Shrestha, Y., Zou, L., Kamburov, A., Zhu, C., et al. (2016). Systematic functional interrogation of rare cancer variants identifies oncogenic alleles. *Cancer Discov.* 6 (7), 714–726. doi:10.1158/2159-8290.cd-16-0160
- Kirshina, A., Vasileva, O., Kunyk, D., Seregina, K., Muslimov, A., Ivanov, R., et al. (2023). Effects of combinations of untranslated-region sequences on translation of mRNA. *Biomolecules* 13 (11), 1677. doi:10.3390/biom13111677
- Kosuge, H., Nagatoishi, S., Kiyoshi, M., Ishii-Watabe, A., Tanaka, T., Terao, Y., et al. (2020). Highly sensitive HPLC analysis and biophysical characterization of N-glycans of IgG-Fc domain in comparison between CHO and 293 cells using FcγRIIIa ligand. *Biotechnol. Prog.* 36 (6), e3016. doi:10.1002/btpr.3016
- Krasniqi, E., Barchiesi, G., Pizzuti, L., Mazzotta, M., Venuti, A., Maugeri-Saccà, M., et al. (2019). Immunotherapy in HER2-positive breast cancer: state of the art and future perspectives. *J. Hematol. Oncol.* 12 (1), 111. doi:10.1186/s13045-019-0798-2
- Leppeck, K., Byeon, G. W., Kladwang, W., Wayment-Steele, H. K., Kerr, C. H., Xu, A. F., et al. (2022). Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. *Nat. Commun.* 13 (1), 1536. doi:10.1038/s41467-022-28776-w
- Lorentzen, C. L., Haanen, J. B., Met, Ö., and Svane, I. M. (2022). Clinical advances and ongoing trials of mRNA vaccines for cancer treatment. *Lancet Oncol.* 23 (10), e450–e458. doi:10.1016/s1470-2045(22)00372-2
- Matoulova, E., Michalova, E., Vojtesek, B., and Hrstka, R. (2012). The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol.* 9 (5), 563–576. doi:10.4161/rna.20231
- Mayr, C. (2017). Regulation by 3'-untranslated regions. *Annu. Rev. Genet.* 51, 171–194. doi:10.1146/annurev-genet-120116-024704
- Mayr, C. (2019). What are 3' UTRs doing? *Cold Spring Harb. Perspect. Biol.* 11 (10), a034728. doi:10.1101/cshperspect.a034728
- Miao, L., Zhang, Y., and Huang, L. (2021). mRNA vaccine for cancer immunotherapy. *Mol. Cancer* 20 (1), 41. doi:10.1186/s12943-021-01335-5
- Moreira, A., Takagaki, Y., Brackenridge, S., Wollerton, M., Manley, J. L., and Proudfoot, N. J. (1998). The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev.* 12 (16), 2522–2534. doi:10.1101/gad.12.16.2522
- Moreira, A., Wollerton, M., Monks, J., and Proudfoot, N. J. (1995). Upstream sequence elements enhance poly(A) site efficiency of the C2 complement gene and are phylogenetically conserved. *Embo J.* 14 (15), 3809–3819. doi:10.1002/j.1460-2075.1995.tb00050.x
- O'Flaherty, R., Bergin, A., Flampouri, E., Mota, L. M., Obaidi, I., Quigley, A., et al. (2020). Mammalian cell culture for production of recombinant proteins: a review of the critical steps in their biomanufacturing. *Biotechnol. Adv.* 43, 107552. doi:10.1016/j.biotechadv.2020.107552
- Oliveira, M. J., Ana, J., Freitas, J., Moreira, A., Ribeiro, B., and Carlos, J. (2020). *Nucleic acid to activate gene expression and protein production. WO/2020/076174.*
- Oliveira, M. S., Freitas, J., Pinto, P. A. B., de Jesus, A., Tavares, J., Pinho, M., et al. (2019). Cell cycle kinase polo is controlled by a widespread 3' untranslated region regulatory sequence in *Drosophila melanogaster*. *Mol. Cell Biol.* 39 (15), e00581-. doi:10.1128/mcb.00581-18
- Orlandini von Niessen, A. G., Poleganov, M. A., Rechner, C., Plaschke, A., Kranz, L. M., Fesser, S., et al. (2019). Improving mRNA-based therapeutic gene delivery by expression-augmenting 3' UTRs identified by cellular library screening. *Mol. Ther.* 27 (4), 824–836. doi:10.1016/j.jymthe.2018.12.011
- Papachristofilou, A., Hipp, M. M., Klinkhardt, U., Früh, M., Sebastian, M., Weiss, C., et al. (2019). Phase Ib evaluation of a self-adjuvanted protamine formulated mRNA-based active cancer immunotherapy, BI1361849 (CV9202), combined with local radiation treatment in patients with stage IV non-small cell lung cancer. *J. Immunother. Cancer* 7 (1), 38. doi:10.1186/s40425-019-0520-5
- Pardi, N., Hogan, M. J., Porter, F. W., and Weissman, D. (2018). mRNA vaccines - a new era in vaccinology. *Nat. Rev. Drug Discov.* 17 (4), 261–279. doi:10.1038/nrd.2017.243
- Pearson, M. J., Khazaipou, S., Optun, A., Pryme, I. F., Stern, B., and Hesketh, J. E. (2012). Albumin 3'untranslated region facilitates increased recombinant protein production from Chinese hamster ovary cells. *Biotechnol. J.* 7 (11), 1405–1411. doi:10.1002/biot.201200044
- Pereira-Castro, I., and Moreira, A. (2021). On the function and relevance of alternative 3'-UTRs in gene expression regulation. *Wiley Interdiscip. Rev. RNA* 12 (5), e1653. doi:10.1002/wrna.1653
- Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29 (9), e45–e45. doi:10.1093/nar/29.9.e45
- Pinto, P. A., Henriques, T., Freitas, M. O., Martins, T., Domingues, R. G., Wyrzykowska, P. S., et al. (2011). RNA polymerase II kinetics in polo polyadenylation signal selection. *Embo J.* 30 (12), 2431–2444. doi:10.1038/emboj.2011.156
- Proudfoot, N. J. (2011). Ending the message: poly(A) signals then and now. *Genes Dev.* 25 (17), 1770–1782. doi:10.1101/gad.17268411
- Rauch, S., Jasny, E., Schmidt, K. E., and Petsch, B. (2018). New vaccine technologies to combat outbreak situations. *Front. Immunol.* 9, 1963. doi:10.3389/fimmu.2018.01963
- Russnak, R., and Ganem, D. (1990). Sequences 5' to the polyadenylation signal mediate differential poly(A) site use in hepatitis B viruses. *Genes Dev.* 4 (5), 764–776. doi:10.1101/gad.4.5.764
- Sample, P. J., Wang, B., Reid, D. W., Presnyak, V., McFadyen, I. J., Morris, D. R., et al. (2019). Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat. Biotechnol.* 37 (7), 803–809. doi:10.1038/s41587-019-0164-5
- Savinov, A., Brandsen, B. M., Angell, B. E., Cuperus, J. T., and Fields, S. (2021). Effects of sequence motifs in the yeast 3' untranslated region determined from massively parallel assays of random sequences. *Genome Biol.* 22 (1), 293. doi:10.1186/s13059-021-02509-6
- Schambach, A., Galla, M., Maetzig, T., Loew, R., and Baum, C. (2007). Improving transcriptional termination of self-inactivating gamma-retroviral and lentiviral vectors. *Mol. Ther.* 15 (6), 1167–1173. doi:10.1038/sj.mt.6300152
- Sebastian, M., Schröder, A., Scheel, B., Hong, H. S., Muth, A., von Boehmer, L., et al. (2019). A phase I/IIa study of the mRNA-based cancer immunotherapy CV9201 in patients with stage IIIB/IV non-small cell lung cancer. *Cancer Immunol. Immunother.* 68 (5), 799–812. doi:10.1007/s00262-019-02315-x
- Shalgi, R., Lapidot, M., Shamir, R., and Pilpel, Y. (2005). A catalog of stability-associated sequence elements in 3' UTRs of yeast mRNAs. *Genome Biol.* 6 (10), R86. doi:10.1186/gb-2005-6-10-r86
- Syed, A. M., Taha, T. Y., Tabata, T., Chen, I. P., Ciling, A., Khalid, M. M., et al. (2021). Rapid assessment of SARS-CoV-2-evolved variants using virus-like particles. *Science* 374 (6575), 1626–1632. doi:10.1126/science.abc16184
- Tihanyi, B., and Nyitray, L. (2020). Recent advances in CHO cell line development for recombinant protein production. *Drug Discov. Today Technol.* 38, 25–34. doi:10.1016/j.ddtec.2021.02.003
- Vallazza, B., Petri, S., Poleganov, M. A., Eberle, F., Kuhn, A. N., and Sahin, U. (2015). Recombinant messenger RNA technology and its application in cancer immunotherapy, transcript replacement therapies, pluripotent stem cell induction, and beyond. *Wiley Interdiscip. Rev. RNA* 6 (5), 471–499. doi:10.1002/wrna.1288
- Valsamakis, A., Zeichner, S., Carswell, S., and Alwine, J. C. (1991). The human immunodeficiency virus type 1 polyadenylation signal: a 3' long terminal repeat element upstream of the AAUAAA necessary for efficient polyadenylation. *Proc. Natl. Acad. Sci. U. S. A.* 88 (6), 2108–2112. doi:10.1073/pnas.88.6.2108
- Walsh, G. (2018). Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* 36 (12), 1136–1145. doi:10.1038/nbt.4305
- Xia, X. (2021). Detailed dissection and critical evaluation of the pfizer/BioNTech and Moderna mRNA vaccines. *Vaccines (Basel)* 9 (7), 734. doi:10.3390/vaccines9070734
- Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* 63 (2), 405–445. doi:10.1128/mmb.63.2.405-445.1999
- Zufferey, R., Donello, J. E., Trono, D., and Hope, T. J. (1999). Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J. Virol.* 73 (4), 2886–2892. doi:10.1128/jvi.73.4.2886-2892.1999