



**UNIVERSIDADE DE LISBOA**  
**INSTITUTO SUPERIOR TÉCNICO**

**Towards Cost-Effective mRNA-based Therapeutics Manufacturing  
Processes**

**Sara Alexandra Peça de Sousa Rosa**

**Supervisor:** Doctor Ana Margarida Nunes da Mata Pires de Azevedo

**Co-Supervisors:** Doctor Duarte Miguel de França Teixeira dos Prazeres  
Doctor Marco Paulo Cardoso Marques

Thesis approved in public session to obtain the PhD Degree in  
**Biotechnology and Biosciences**

Jury final classification: **Pass with Distinction**

**2025**



**UNIVERSIDADE DE LISBOA**  
**INSTITUTO SUPERIOR TÉCNICO**

**Towards Cost-Effective mRNA-based Therapeutics  
Manufacturing Processes**

**Sara Alexandra Peça de Sousa Rosa**

**Supervisor:** Doctor Ana Margarida Nunes da Mata Pires de Azevedo

**Co-Supervisors:** Doctor Duarte Miguel de França Teixeira dos Prazeres  
Doctor Marco Paulo Cardoso Marques

Thesis approved in public session to obtain the PhD Degree in  
**Biotechnology and Biosciences**

Jury final classification: **Pass with Distinction**

**Jury**

**Chairperson:** Doctor Arsénio do Carmo Sales Mendes Fialho, Instituto Superior Técnico, Universidade de Lisboa

**Members of the Committee:**

Doctor Nicolas Szita, Biochemical Engineering Department, University College London, Reino Unido

Doctor Ana Cecília Afonso Roque, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa

Doctor Maria Raquel Múrias dos Santos Aires Barros, Instituto Superior Técnico, Universidade de Lisboa

Doctor Gabriel António Amaro Monteiro, Instituto Superior Técnico, Universidade de Lisboa

Doctor Marco Paulo Cardoso Marques, Biochemical Engineering Department, University College London, Reino Unido

Doctor Marília Clemente Velez Mateus, Instituto Superior Técnico, Universidade de Lisboa

**Funding Institution**

FCT: Fundação para a Ciência e a Tecnologia

**2025**



# Abstract

mRNA is an emerging technology in the vaccine field. Owing to its precision and safety, as well as flexible manufacture, this technology is attractive for multiple applications, from prophylactic and cancer treatments to metabolic and genetic diseases. One of the most unique characteristics of these vaccines is the simple and flexible manufacturing process: mRNA is produced in cell-free reaction, where DNA template is transcribed into mRNA catalysed by an RNA polymerase (IVT). The highly defined nature of this cell-free reaction makes the process ideal for optimisation. Using a Bayesian optimization method to automate the experiment design, we were able to optimise mRNA production of the IVT reaction to achieve  $12 \text{ g}_{\text{mRNA}} \cdot \text{L}^{-1}$  of reaction in just 2 hours.

During IVT reaction, by-products, namely truncated mRNA and dsRNA are co-produced. A well-established manufacturing platform requires a tight control on the presence dsRNA throughout the process. We tackle this by exploring the impact the DNA sequence in dsRNA formation during IVT. Optimising the non-coding regions has led to a decrease of 30% in the production of dsRNA. Additionally, new monitoring strategy was developed where RNase T1 was combined with RP-HPLC. This has allowed to quantify dsRNA throughout the manufacturing process without interference of other impurities. The manufacturing process was further optimised by exploring affinity chromatography to directly capture mRNA from IVT reactions. Using an AI approach, a 7.5-fold increase in the resin binding capacity was obtained. Additionally, multimodal chromatography was explored to separate mRNA from dsRNA in a one-step process, which potentially can reduce the overall manufacturing costs. Using this method, a overall recovery yield of  $81 \pm 5\%$  was achieved, with a purity of  $88 \pm 2\%$  and no detectable concentration of DNA.

In the end, the results obtained contribute to state-of-art of mRNA vaccines manufacturing and will contribute to the development of sustainable, flexible and cost-effective manufacturing process.

**Key-words:** mRNA; Vaccines; Manufacturing Process; Production; Purification

# Resumo

mRNA é uma tecnologia emergente no campo das vacinas. Devido à sua precisão e segurança, bem como ao seu método de fabrico flexível, esta tecnologia é atrativa para múltiplas aplicações, desde tratamentos profiláticos e de cancro, até doenças genéticas e metabólicas. Uma característica única destas vacinas é o seu método de fabrico ser simples e flexível: mRNA é produzido numa reação livre-de-células, onde o molde de DNA é transcrito em mRNA (IVT), numa reação catalisada por a RNA polimerase. Um método baseado em inteligência artificial foi utilizado para a optimização da reação de IVT, obtendo-se uma produção de RNA de  $12 \text{ g}_{\text{mRNA}} \cdot \text{L}^{-1}$  em apenas duas horas reacionais.

Uma plataforma de manufactura bem estabelecida requer um controlo apertado à presença de dsRNA durante o processo. Nós abordamos este desafio explorando o impacto da sequência de DNA na formação de dsRNA durante a IVT. A optimização das regiões não codificantes levou a uma diminuição de 30% na produção de dsRNA. Uma nova estratégia de monitorização, onde a RNase T1 foi combinada com RP-HPLC, foi desenvolvida. Ista permite quantificar a dsRNA durante o processo sem interferência das impurezas. O processo de manufactura foi também optimizado ao explorar cromatografia de afinidade para capturar diretamente mRNA das reações de IVT. O modelo de IA foi usado para aumentar 7,5 vezes a capacidade da resina. Explorou-se cromatografia multimodal para separar mRNA do dsRNA num processo de um passo, o que potencialmente pode reduzir o custo geral de manufactura. Usando este método, antigiou-se uma recuperação de  $81 \pm 5\%$  de mRNA, com uma pureza de  $88 \pm 2\%$  e sem detectar a presença de DNA.

Os resultados obtidos contribuem para o estado-da-arte da manufactura de vacinas de mRNA e vai contribuir para o desenvolvimento de um processo sustentável, flexível e económico.

**Palavras-chave:** mRNA; Vacinas; Processo de manufactura; Produção; Purificação

# Agradecimentos /Acknowledgments

Durante estes últimos 4 anos (e durante toda a minha vida profissional) tive a sorte de conhecer pessoas que me marcaram e que sem elas nada disto teria sido possível. Espero não me esquecer de ninguém!

Primeiro que tudo, quero agradecer à professora Ana Azevedo. Se nunca tivesse respondido ao meu e-mail e dado a oportunidade de voltar para ciência. Obrigada por sempre ter respeitado as minhas ideias e por me ter deixado voar. Se não tivesse aceitado ser minha orientadora, hoje não estaria a escrever estes agradecimentos.

Também quero agradecer ao meu co-orientador professor Miguel Prazeres por todas as discussões e conselhos durante todo este tempo.

Marco, não há palavras suficientes para te agradecer. Foste (e és) muito mais que um co-orientador. Nunca duvidaste das minhas capacidades e sempre fizeste o possível e o impossível para me ajudar neste caminho, mesmo quando estava em Portugal. Tens o maior coração que eu alguma vez vi. Tenho muita sorte de te ter conhecido. És uma inspiração como cientista e (sobretudo) como pessoa.

Também quero agradecer a todos os alunos que mentorei durante este tempo. Começando por ti, Ricardo, que para além da sorte de te ter mentorado, também te tornaste um amigo para a vida. Ana e Mariana, espero que vos tenha ajudado na recta final das vossas vidas académicas.

Nos 4 anos que tive no iBB, tive a sorte de conhecer pessoas incríveis. Aos meus colegas de doutoramento, Leonor, Cris, Diogo, Jorgito e Miguel, muito obrigada por todos os momentos que passamos. Quero também agradecer à minha vizinha de lab, Renata, com quem passei todo o tempo de pandemia. Ricardo e dona Rosa, muito obrigada por todas as conversas e companhia. Sem vocês todo o tempo que passei no iBB teria sido muito aborrecido.

Felizmente, durante a minha vida fui colecionando pessoas que espero que me vão acompanhar para a vida. Sara Salazar, muito obrigada por todas as conversas. À Ana Quendera e Sofia, que continuemos muitos anos a partilhar as nossas aventuras. Carina, Catarina e Luisa, obrigado por todo o apoio.

When I arrived to UCL, I had a difficult time to adjust. Nevertheless, I was very welcomed by everyone, and I feel grateful for all the support that everyone gave me.

I was lucky have to mentored 6 amazing students, Shuran, Brandon, Nivan, Yves, Agust and Yustika.

Yustika, you are a ray of light to all those you meet.

To all my colleagues, Julian, Paras, Steven, Georgia, April and Celia. Thank you so much for all the moments we shared. To Baz and Kyle. To my little "Italians" Nusrat and Teresa! Thank you so much for everything!!!!

Aos meus amigos, Clara, Ricardo, Claudia, Rita. Espero que continuemos a construir memórias juntos.

Quando cheguei a Londres, ganhei uma família. E tenho a sorte de essa família trabalhar comigo todos os dias. Muito obrigada por tudo Salomé. And, of course Gu!

Também quero agradecer à minha família. Tenho a sorte de ter uma irmã-mãe, que sempre me tratou como filha, e que me deu duas sobrinhas que adoro como se fossem minhas. Aos meus pais, que me sempre apoiaram incondicionalmente, mesmo não percebendo muito bem esta ideia de desistir do emprego para fazer ciência.

Ao Davide, que mudou de país por mim. Não existe eu sem nós.

E à minha avó que neste momento está a lutar pela vida. Espero um dia ter 1/10 da força que tu tens.

# Abbreviation List

AEX	Anion exchange
AES	Amino-terminal enhancer of split
ARES	Adenylate-uridylate-rich elements
AML	Acute myeloid leukemia
ATPS	Aqueous two-phase system
AI	Artificial Intelligence
ARCA	Anti-reverse cap analog
Bp	base pair
CAR	Chimeric antigen receptor
CC $\alpha$	Decision limit
CCb	Detection capability
cDNA	Complementary DNA
cGMP	Current Good Manufacturing Process
circRNA	Circular RNA
CPP	Cell penetrating peptides
CV	Column volume
DC	Dendritic cells
DBC	Dynamic binding capacity
dsRNA	Double stranded RNA
<i>E.coli</i>	<i>Escherichia coli</i>
EGFP	Enhanced GFP
eIF	Eukaryotic-translation initiation factor
EDTA	Ethylenediaminetetraacetic acid

GMP	Good Manufacturing Practice
GTP	Guanosine triphosphate
HIC	Hydrophobic interaction
HPLC	High performance liquid chromatography
IFN	Interferon
IP	Ion pair
ISPR	<i>In situ</i> product recovery
IRES	Internal ribosome entry site
IVT	<i>in vitro</i> transcription
LiCl	Lithium chloride
LNP	Lipid nanoparticle
LOD	Limit of detection
LOQ	Limit of quantitation
Mes	2-( <i>N</i> -morpholino)ethanesulfonic acid
miRNA	microRNA
mRNA	Messenger RNA
NK	Natural killing cells
NSP	Non-structural proteins
NTP	Nucleoside triphosphate
ORF	Open reading frame
PABP	Poly(A)-binding protein
PAT	Process Analytical Technology
PCR	Polymerase chain reaction
pDNA	Plasmid DNA
QbD	Quality by design

RNAP	RNA polymerase
RP	Reverse-phase
RDRP	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
SAM	S-Adenosyl methionine
saRNA	Self-amplifying RNA
SEC	Size exclusion chromatography
ssRNA	Single stranded RNA
SFPR	Substrate feed and product recovery
SPTFF	Single pass tangential flow filtration
taRNA	Trans-amplifying RNA
TFF	Tangential Flow Filtration
TLR	Toll-like receptor
TNF	Tumour necrosis factor
Tris	Tris(hydroxymethyl)aminomethane
VCC	Vaccinia capping system
RT-PCR	Real time PCR
UTR	Untranslated region
WHO	World Health Organisation
Wt	Wild type

# Table of Contents

Abstract.....	I
Resumo.....	II
Agradecimentos /Acknowledgments.....	III
Abbreviation List.....	V
Chapter I – General Introduction .....	1
Introduction.....	2
The rise of mRNA technology .....	4
mRNA Vaccine Structure .....	4
mRNA Delivery.....	6
Applications.....	7
mRNA Manufacturing: from upstream to downstream.....	9
New perspectives .....	14
mRNA Safety and Quality .....	15
Concluding Remarks .....	15
References .....	16
Chapter II – The ABC of mRNA Manufacturing: a holistic perspective into process design and optimisation.....	25
Introduction.....	26
mRNA vaccines .....	27
mRNA designs.....	29
Untranslated regions .....	30
Poly-A tail.....	32
5' cap .....	32
Sequence optimisation .....	32
Modified nucleosides.....	33
Manufacturing steps, challenges and Quality Control .....	34
Plasmid design.....	34
pDNA manufacturing .....	34
In vitro transcription .....	36
Capping.....	38
Impurities challenge .....	39
Quality.....	42
Formulation .....	44

Current and future manufacturing demands.....	46
DNA template production .....	46
mRNA manufacturing.....	48
Downstream processing.....	49
Formulation.....	49
What is next?.....	50
References.....	51
Chapter III – Aims and Objectives .....	67
Thesis Aim.....	68
Chapters Outlook.....	69
References.....	71
Chapter IV – (RP)UHPLC/UV analytical method to quantify dsRNA during mRNA vaccine manufacturing process .....	72
Introduction.....	73
Materials and methods.....	74
Results and Discussion.....	78
Conclusions.....	83
References.....	84
Supplementary Material.....	87
Chapter V - Maximizing mRNA vaccine production with Bayesian optimization .....	91
Introduction.....	92
Materials and Methods.....	93
Results .....	98
Discussion .....	102
Conclusion.....	108
References.....	108
Supplementary Material.....	113
Chapter VI - Comprehensive evaluation of T7 promoter for enhanced yield and quality in mRNA production .....	118
Introduction.....	119
Material and methods .....	120
Results .....	125
Discussion .....	135
Conclusions .....	137
References.....	138

Supporting Information .....	141
Chapter VII – Simplifying the mRNA manufacturing process: Exploring oligo-dT ligand as a capture step after <i>in vitro</i> transcription reactions .....	148
Introduction.....	149
Materials and methods.....	151
Results and Discussion.....	154
Conclusions .....	161
References .....	162
Chapter VIII – Multimodal chromatography as a one-step purification process for mRNA vaccines .....	166
Introduction.....	167
Materials and Methods.....	169
Results and discussion .....	173
Conclusions .....	181
References .....	182
Chapter IX – Conclusions and Future work .....	186
References .....	190

# Chapter I – General Introduction

---

**Key words:** Vaccination; mRNA vaccines; mRNA manufacturing; mRNA applications

This chapter has been published as: Rosa, Sara Sousa, Duarte MF Prazeres, Ana M. Azevedo, and Marco PC Marques. "mRNA vaccines manufacturing: Challenges and bottlenecks." *Vaccine* 39, no. 16 (2021): 2190-2200.

# Introduction

Vaccines are one of the greatest advances in medicine and an important public health tool, as they not only prevent infection, morbidity and mortality individually, but also reduce and eliminate disease prevalence locally, ultimately leading to eradication of disease globally [1]. Since the development of the smallpox vaccine in 1798 [2] and rabies vaccine in 1885 [3], vaccine technology progressed from the use of inactivated and attenuated pathogens, to the use of subunits that only contain those pathogen components that can trigger an immunologic response (Figure 1). Key milestones include the development of virus-like particle vaccines, recombinant viral-vectored vaccines, and toxoids, polysaccharides or protein-based vaccines, which can be conjugated with different protein carriers to improve immune response.

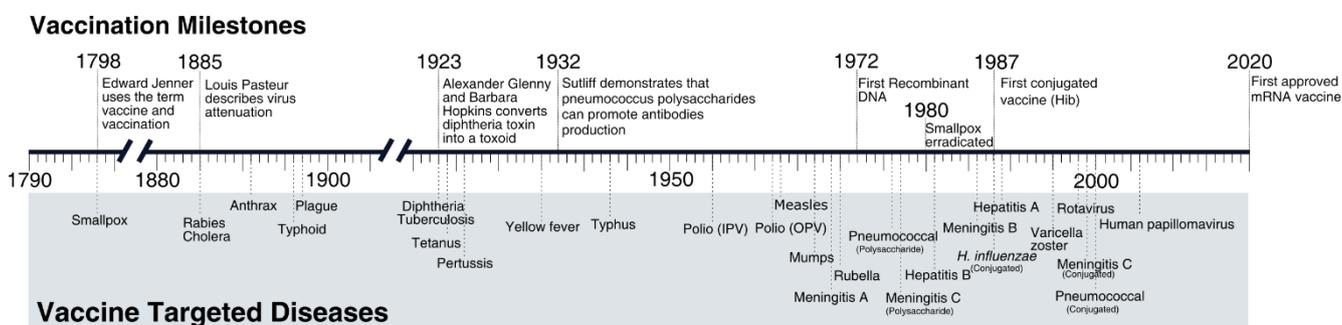


Figure 1. Vaccination targets and milestones adapted from [4,5].

Vaccines save 6 million lives every year and are one of the major responsible for the increase of human longevity [6]. Their impact on the economic viability of the healthcare system is also very large, since vaccines lower the treatment costs of diseases [7], and reduce the impact and risk of outbreaks [8]. Additionally, by preventing bacterial infection and, subsequently, reducing the need for antibiotic treatment, vaccines can have an impact on antimicrobial resistance [9]. The use of vaccines goes beyond prevention of infectious diseases. Technology advances coupled with progress in target selection and understanding of the immunosuppressive mechanisms have led to the development of therapeutic cancer vaccines [10].

Despite the proven effectiveness of current vaccines, there is still room for improvement in the vaccine technology field. Traditional attenuated and inactivated vaccines are still widely used today (e.g., Bacillus Calmette–Guérin vaccine, BCG and Inactivated Polio vaccine, IPV) owing to their robustness and stability. However, they present safety concerns due to the use of whole pathogens and in many cases, they do not have a defined composition. In the case of toxoid and subunit vaccines, and despite their safety and stability profile, the use of adjuvants is required for a strong immune response and the protection lifetime is limited (Table 1).

The manufacturing of new vaccines is typically a lengthy (6 to 36 months), challenging and expensive process, as no standard process is available [11,12]. To deliver effective, precise, and consistent vaccines it is imperative to use good manufacturing practice (GMP) compliant equipment, facilities, and procedures. However, this is costly and difficult to implement at a large scale. Vaccines developed on the basis of traditional technology have failed to respond effectively to several diseases, such as malaria, tuberculosis, AIDS or flu. Furthermore, SARS and Ebola epidemic outbreaks and, more recently, the COVID-19 pandemic, show that many of the current platforms are not well suited for a very fast, efficient, and cost-effective response.

New vaccine technology approaches are thus necessary to improve our response to outbreaks and enable vaccination worldwide. Ideally, a new vaccine should be safe, effective, stable, available to all populations and not susceptible to antigenic variance [13]. The manufacturing must be reliable, efficient, low-cost, and flexible to allow on-demand production. Viral vectors and DNA technology are two cutting-edge platforms that have the flexibility and characteristics to support faster vaccine development and manufacturing [14]. However, the costly and complex manufacturing of viral vectored vaccines and the poor immunogenicity presented by DNA vaccines (Table 1) can make them unattractive for some clinical applications.

Table 1. Advantages (+) and disadvantages (x) of the currently available types of vaccines.

Properties	Subunit					
	Inactivated	Live attenuated	Toxoid	(conjugate; protein-based; polysaccharide)	Viral Vectors	DNA
Humoral and cellular immune response	✗	+			+	
Lasting protection		+	✗	✗		
Stability	+		+	+	+	+
Safety	+	✗	+	+	✗	+
Manufacturing	+		+	✗	✗	+
Presence of adjuvants		+	✗	✗	+	✗
Cold chain		✗			✗	+

## The rise of mRNA technology

mRNA vaccines have reached the spotlight during the Covid-19 pandemic, as the forefront technology used for the development of vaccines by many companies. In fact, a mRNA vaccine candidate was the first to reach phase I clinical trials [15]. The potential of mRNA vaccines was first hinted at in 1990, when the *in vivo* expression of a protein was observed after injecting the coding mRNA into mouse skeletal muscle [16]. These early experiments proved that *in vitro* transcribed mRNA (IVT) can induce the production of proteins in live tissues. During the following 10 years, several studies demonstrated that mRNA could induce an immunologic response to the expressed protein in many mammalian cell types both *in vitro* and *in vivo* [17–19].

mRNA technology presents several advantages that makes it an attractive alternative over traditional vaccines or even DNA vaccines. Unlike attenuated or inactivated vaccines, mRNA is precise as it will only express a specific antigen and induce a directed immune response. Additionally, it promotes both humoral and cellular immune response and induces the innate immune system [20]. Compared with DNA-based vaccines, mRNA is more effective, since expression does not require nuclear entry, and safer, since the probability of random genome integration is virtually zero [21,22]. Additionally, expression of the coded antigens is transient since mRNA is quickly degraded by cellular processes, with no traces found after 2-3 days [23]. The flexible nature of the mRNA vaccine platform is also advantageous for manufacturing since a change in the encoded antigen does not affect the mRNA backbone physical-chemical characteristics [24], and hence allow production to be standardised. Additionally, since production is based on an *in vitro* cell-free transcription reaction, safety concerns regarding the presence of cell-derived impurities and viral contaminants commonly found in other platforms are minimised.

### mRNA Vaccine Structure

Construction of mRNA vaccines requires the insertion of the encoded antigen in a DNA template from where the mRNA is transcribed *in vitro*. Unlike DNA, mRNA only needs to reach the cytosol, where it will be transcribed into the antigen *in vivo*, using the cell machinery. This way, any desired sequence can be designed, produced *in vitro*, and delivered to any type of cell [21]. Inside the cells, RNA is recognised by endosomal or cytosolic receptors, which can lead to the activation of the type I interferon (IFN-I) pathway, and to the promotion of the production of chemokines and proinflammatory cytokines. These signal molecules lead to antigen-presenting cell (APC) activation and, subsequently, to a strong adaptive response [25].

The structure of mRNA vaccines is similar to eukaryotic mRNA - a single-stranded molecule with a cap at the 5' end, a poly(A) tail at the 3' end and an open reading frame (ORF) flanked

by untranslated regions (UTR) [20]. The 5' cap is an important component as it enables the translation initiation by binding to a eukaryotic translation initiation factor (eIF4E) [26]. Different structures are possible for the 5' cap. The Cap 0 structure, which features a methyl-7 guanine nucleotide linked to the 5' position through a 5' triphosphate, is the simplest. The Cap 1 structure is achieved by the methylation of the mRNA first nucleotide at the ribose 2'-O position. Both caps can be added during *in vitro* mRNA transcription using a synthetic cap analogue [27] or the proprietary Cap dinucleotide CleanCap® [28]. Another capping approach uses a post-transcription enzymatic reaction based on the vaccinia capping system [29]. This modification brings with it a number of advantages as it improves the translation initiation by recruiting translation initiation factors, protects the synthetic mRNA against exonuclease degradation [30], and avoids an innate immunity overactivation response [25]. The addition of a 3' poly(A) tail also improves mRNA stability and translational activities, as it protects mRNA from nuclease degradation by the poly(A)-binding protein (PABP) [31]. This tail can be added to the transcript by inserting a poly(A) sequence in the DNA template or by an enzymatic reaction [27]. Tail size optimization is an important factor for the stabilization and expression of mRNA. Longer poly-A tails can improve mRNA stability and translation. However, this effect is not linear, and the best tail size is dependent on cell type [31]. The untranslated regions (UTRs) are responsible for the transcription regulation and mRNA stability. These regions strongly affect translation efficiency as the sequences used are involved in the translation machinery recognition, recruitment, and mRNA trafficking. Strategies to modulate the innate immune response, such as the introduction of unnatural nucleosides (NTPs), and to improve translation efficiency, by using codon optimisation, are also commonly used in mRNA production [27, 28].

Two forms of mRNA structure are being extensively studied for vaccine applications: conventional or non-replicating mRNA and self-amplifying mRNA. In the conventional mRNA form, the antigen of choice is only flanked by UTR regions, a 3' poly(A) tail and a 5' cap. This form presents several advantages - molecules are simple and small, and the possibility of unwanted immune response is lowered since no other proteins are encoded [32]. However, this mRNA expression is limited to its transient nature, and higher mRNA doses may be necessary to achieve high expression [33]. Efforts have been made to overcome this bottleneck by using sequence optimization and formulation [34]. Self-amplifying mRNA (saRNA) is based on the addition of a viral replicase gene to enable the mRNA to self-replicate. Usually, sequences of single-stranded RNA viruses, such as alphaviruses, flaviviruses, and picornaviruses, are used [35]. Upon cytoplasm delivery, this type of mRNA produces high levels of the antigen of interest. Despite the use of viral genes, no viral infectious particles or virus-like-particles are observed during expression, reducing the safety concerns [21]. Evaluation of an saRNA vaccine for protection of mouse models against H1N1/PR8 infection showed that a 64-fold lower dose was

required to induce an immunologic response when compared with the conventional mRNA vaccine counterpart [36].

Trans-amplifying mRNA (taRNA) is a new structural modality of mRNA vaccines. The taRNA results from the splitting of the self-amplifying mRNA in a system with two templates, one containing the gene of interest and a second containing the replicase system. The amplification is performed *in trans* by the replicase in the cytoplasm. This system presents some advantages over saRNA since it is safer, more versatile and cost-effective to manufacture, as the production of shorter RNAs with high yield and high quality is less challenging. taRNA has already been used to protect mice against influenza with results showing induction of antibodies and protection[37].

## mRNA Delivery

mRNA must cross the cell membrane to reach the cytosol. This is challenging due to the negative charge of the molecule, its relatively large size (300-5000 kDa) and degradability, which can hamper its passive pass through the cell membrane [38]. To overcome this, mRNA can be delivered using different strategies including: i) direct injection of naked mRNA; ii) conjugation with lipid-based carriers, polymers, or peptides; iii) via transfection of dendritic cells (DC) [39].

The induction of an immune response by injection of naked mRNA in conventional and self-amplifying forms has been widely reported [40–44]. However, mRNA delivery can be limited by the presence of extracellular exonucleases in the target tissues, inefficient cell uptake or unsuccessful endosomal release [27]. Liposomes or lipid nanoparticles (LNPs) are one of the most promising mRNA delivery tools [45]. For example, LNP-mediated delivery of mRNA vaccines against Zika and influenza has shown encouraging results [46–49]. Although less explored, polymer-based delivery systems can also be used. Polyethylenimine (PEI) systems were successfully implemented as a strategy to deliver mRNA to cells [50], and intranasally [51]. Additionally, PEI-based systems improved the response to sa-mRNA vaccines in skin explants [52] and in mice [36]. Peptide-based delivery is a less explored system, as only protamine has been evaluated in clinical trials [53]. New delivery approaches include the use of cationic cell-penetrating peptides (CPPs) and anionic peptides. CPPs systems have proved to improve T-Cell immunity response *in vivo* [54], modulate innate immune response and enhance protein expression in both DC and human cancer cells *in vitro* [55,56]. mRNA polyplexes conjugated with an anion peptide, exhibited an increase in cellular uptake without inducing cytotoxicity in DC cells [57].

Despite the efforts to improve mRNA delivery, there are still challenges that must be considered, such as the delivery efficiency, cell targeting, materials safety, route of administration and vaccine thermostability. This topic is extensively revised elsewhere [39].

## Applications

Since Wolf *et al.* [16] showed that proteins can be produced from *in vitro* transcribed mRNA in live tissues, mRNA vaccines have been demonstrating efficacy in a number of applications [58]. The first record of a clinical trial using mRNA technology based on RNA-pulsed DC cancer vaccine dates back to 2003 [59]. Today, more than 140 clinical trials can be found that use mRNA to address different conditions such as cancer or infectious disease (Figure 2).

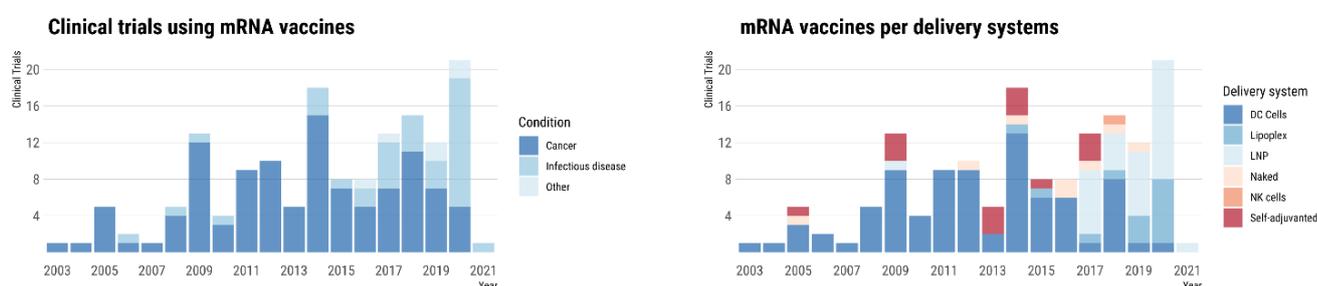


Figure 2. Breakdown of mRNA vaccines clinical trials filed per year according to disease type (left) and delivery system (right).

From the first applications, mRNA has emerged as a potential therapy for cancer. *Boczkowski et al* [60] produced one of the first breakthroughs by using mRNA to generate vaccines based on RNA-pulsed dendritic cells (DC) against tumour cells. Using this system, the antigen-presenting immune response was induced, and tumour regression was observed. Since then, mRNA-based DC vaccines have shown their potential in cancer applications in over 70 completed clinical trials. Recently, a phase I study where RNA transduced DCs were evaluated as a post-remission therapy in acute myeloid leukaemia (AML) was published [61]. This treatment induced an immune response with a positive relation between higher survival rate of patients with  $\leq 65$  years. The use of mRNA has also been explored to engineer T- or Natural Killer (NK) cells to express chimeric antigen receptor (CAR) that are used as a cancer cell therapy [62,63]. In fact, this system was successfully implemented in a phase I clinical trial designed to evaluate its potential in the treatment of colorectal cancers [64].

The direct injection of mRNA is a more cost-effective delivery alternative to DC vaccines. *In vivo* delivery of the naked, complexed, or encapsulated mRNA can be successfully performed by a number of administration routes such as intradermal, intramuscular, intranasal, intratumoral, intranodal or even intravenous [45]. Using this method, a dose consisting of only a few tenths or hundreds of micrograms of mRNA (10-250  $\mu\text{g}$ ) is administered to each patient to trigger an immune response [65]. The first clinical trial evaluating direct injection used naked mRNA in

patients with melanoma [66]. This approach was feasible and safe but no clinical effectiveness was observed. Self-adjuvanted RNAActive® vaccines is a technology developed by CureVac that uses a mixture of protamine-complexed and naked mRNA to improve the immunostimulatory effect of the vaccine [67]. This technology was successfully applied in phase I and I/II clinical trials targeting liver [68], prostate [69], lungs [70] and melanoma [71] cancers. New delivery approaches using lipoplexes and LNPs have been extensively used in clinical trials studies in the last couple of years. Recent results show that both technologies can be successfully applied to treat melanoma [72], lymphoma [73,74], and solid tumours [75,76].

Cancer is currently the target of choice for mRNA technology. Over 50% of the clinical trials focus on the treatment of melanomas, prostate and brain cancer (Figure 3), with most of the trials still in the early phases (I and II). The lack of benchmarks for cancer treatment hampers the evaluation of the vaccine’s effectiveness beyond the safety profile and the immunological response [21]. However, this is not the case for infectious diseases since many conventional vaccines are available to serve as benchmarks to validate the new mRNA vaccines.

Owing to its versatility and flexible manufacture, mRNA is an excellent platform for the development of prophylactic or therapeutic vaccines against infectious diseases (Figure 3). The first studies using mRNA technology for infectious diseases therapeutics targeted HIV. Using DC-based and naked delivery systems, phase I and II clinical trials presented mixed results despite the vaccine’s safe profile [77], as a lack of an efficient immunologic response against HIV was observed [78,79].

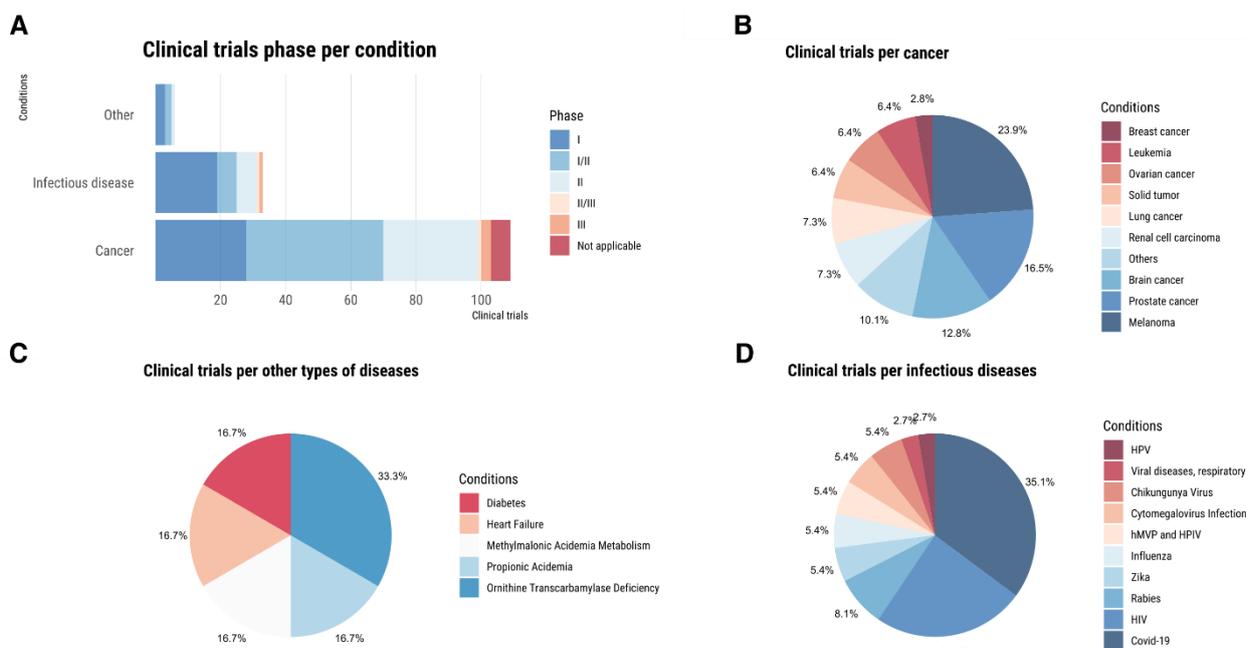


Figure 3. Distribution of clinical trials from <https://clinicaltrials.gov/> and <http://www.isrctn.com/> using mRNA vaccines per condition and phase (A), types of cancer (B), other disease types, and (D) infectious diseases.

Prophylactic vaccines using mRNA technology were also directed to rabies, with the first clinical trial using a self-adjuvanted delivery system [80]. Interestingly, this trial showed that the vaccine effectiveness depended on the route of administration, as only those patients that received the vaccine via needle-free devices produced antibodies above the WHO predefined titre ( $\geq 0.5$  IU mL<sup>-1</sup>). A new formulated mRNA vaccine based on LNPs delivery system is currently being evaluated in a phase I clinical trial [25].

mRNA technology is a perfect fit to overcome the bottlenecks faced by the conventional influenza vaccine. Indeed, studies on influenza immunisation provided the first demonstration of the efficacy of mRNA vaccines against infectious diseases in animals models (mice, ferrets and pigs) [81]. An LNP-based vaccine encoding H10N8 and H7N9 is currently being evaluated in phase I clinical trials. The first published results demonstrated that the H10N8 encoding vaccine was safe and triggered a robust prophylactic immunity [48].

mRNA vaccines have also shown promising results against other infectious diseases. For example, experiments with an LNP-based system against Zika have been performed in cells, mice and primates [46,47]. Currently, phase I clinical trials against Zika virus, Chikungunya virus, and a phase II trial against Human Cytomegalovirus using LNPs-bases systems are ongoing.

During the current Covid-2019 pandemic, mRNA vaccines took the spotlight as the first vaccines to be approved for the prophylactic treatment. Furthermore, at least nine clinical trials can be found using mRNA technology, two of which are in phase III. Three recently published studies describe encouraging results obtained in phase I clinical trials using LNP-based systems [82–86]. All studies reported a safe profile with mild to moderate reactions, despite the greater reactogenicity observed following the administration of the second dose. Furthermore, an immunologic response was also observed in all studies, thus supporting the advance of this technology to late-stage clinical evaluation. A recent phase III study reported an efficacy of 95% [86].

## mRNA Manufacturing: from upstream to downstream

One of the most important advantages of mRNA over conventional vaccines is its relatively simple manufacturing. To produce the mRNA product with specific quality attributes, a series of manufacturing steps must be carried out. Currently, a well-established manufacturing platform is still lacking and a number of combinations of steps is possible. These can be grouped into the upstream processing, which comprises the enzymatic generation of mRNA, and the downstream processing, which includes the unit operations required to purify the mRNA product

(Figure 4). These are complemented with LNP formulation and Fill-to-Finish steps [92]. Nonetheless, the choice of the unit operations is still dependent on the purpose. For example, a lab scale production usually consists of a one-step synthesis reaction followed by a nuclease digestion and a precipitation [58]. The exact unit operations used can have an impact on the manufacturing price [92] and on the cost per dose. Ultimately, the cost will be greatly influenced by the quantity of RNA per dose, production titres and production scale used. The purchase price of 5' cap analogue and modified UTP seem to have some impact on the cost [92].

mRNA is produced in a cell-free system and uses no animal derived raw materials. Cell-derived impurities or adventitious contaminations are thus absent, which makes the manufacturing of these molecules safer [58, 65]. The *in vitro* transcription (IVT) enzymatic reaction used to generate mRNA relies on T7, SP6 or T3 RNA polymerases to catalyse the synthesis of the target mRNA from the corresponding DNA template (Figure 4). This template must be produced in advance, usually by linearisation of a purified plasmid or by amplification of the region of interest using PCR. Apart from the linear DNA template, the IVT components must then include an RNA polymerase, nucleotide triphosphates (NTPs) substrates, the polymerase cofactor MgCl<sub>2</sub>, a pH buffer containing polyamine and antioxidants [33,89]. The reaction only takes a few hours in contrast with the time-consuming processes used to manufacture conventional vaccines. Furthermore, this reduced time lowers the probability for contamination to occur [65]. In general, milligrams of mRNA per millilitre of reaction can be obtained [90]. Additionally, the production process can be standardized as it is not dependent on the antigen encoded in the template.

As for mRNA capping, it can be performed during the IVT reaction by substituting a part of the guanosine triphosphate (GTP) substrate for a cap analog [91]. Alternatively, mRNA can be capped in a second enzymatic reaction using the vaccinia capping enzyme (VCC) and a methyl donor as a substrate (Figure 4). Although the capping efficiency of this method is higher (100% compared to 60-80% obtained with the use of a cap analog), the process with cap analogs is faster as it does not require the set-up of a second enzymatic reaction [25]. However, due to their price, cap analogues can have an impact on production costs [92], especially if large scale manufacturing is considered. Nevertheless, a cost analysis should be performed to compare the costs of the one-step and two-step production options [93]. Alternatively, co-transcriptional capping can be performed using CleanCap® Reagent AG [28]. Although this method does not compete with GTP and delivers a Cap 1 construct, it requires the use of templates with a modified T7 promoter.

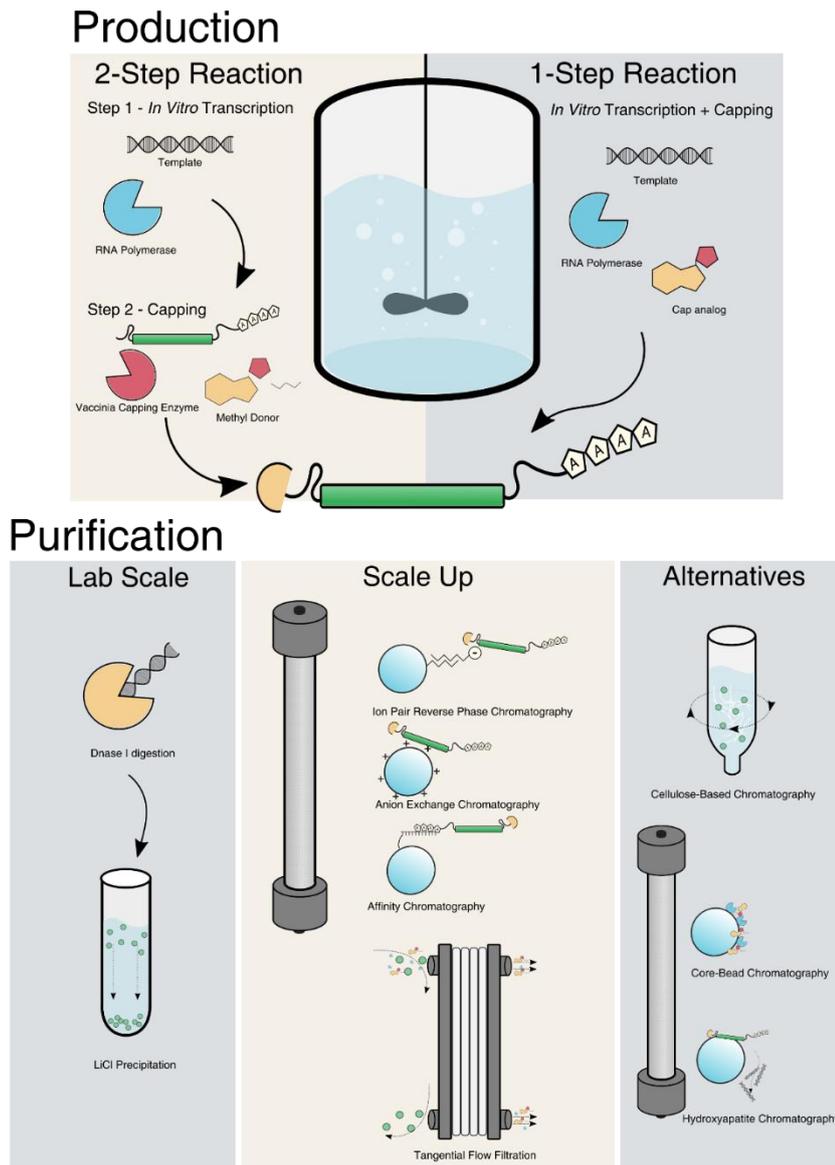


Figure 4. Schematic representation of the production and purification steps of a mRNA vaccines manufacturing process. mRNA production can be performed in a one-step enzymatic reaction, where a capping analog is used, or in a two-step reaction, where the capping is performed using vaccinia capping enzyme. mRNA purification process at lab scale consists of Dnase I digestion followed by LiCl precipitation. Purification at a larger scale is obtained using well-established chromatographic strategies coupled with tangential flow filtration. Alternatively, new types of chromatography can be used to complement the standard purification.

Although several commercial kits are available to produce mRNA for preclinical studies at laboratory scale, their costs are high [94]. The generation of mRNA by IVT at large scale and under current good manufacturing practice (cGMP) conditions is also challenging. For example, the specialised components of the IVT reaction must be acquired from certified suppliers that guarantee that all the material is animal component-free and GMP-grade. Furthermore, the availability of large amounts of these materials is limited and purchasing costs are high [58]. This is true, for example, in the case of the enzymes used for translation and capping.

Nevertheless, the expedite and simple nature of the production process is expected to lower production and operational costs when compared with the cell-based manufacturing of other biologicals such as proteins, antibodies, plasmid DNA and virus-like particles [94].

Once the mRNA is generated by IVT, it must be isolated and purified from the reaction mixture using multiple purification steps to achieve clinical purity standards (Figure 4). The reaction mixture contains not only the desired product, but also a number of impurities, which includes enzymes, residual NTPs and DNA template, and aberrant mRNAs formed during the IVT. Traditional lab scale purification methods are based on DNA removal by DNase digestion followed by lithium chloride (LiCl) precipitation [31,58]. However, these methods do not allow the removal of aberrant mRNA species such as dsRNA and truncated RNA fragments. The removal of these product-related impurities is crucial for mRNA performance, as they lower translation efficiency and modify the immunostimulatory profile. For example, a 10-1000-fold increase in protein production was observed when nucleoside-modified mRNA was purified by reverse phase HPLC prior to delivery to primary DC [95].

Chromatography is a mainstream purification process widely accepted in the pharmaceutical industry. Its high popularity is derived from several attributes such as selectivity, versatility, scalability and cost-effectiveness [96]. The first published protocol for large scale purification of synthetically produced RNA oligonucleotides used size exclusion chromatography (SEC) in a gravity-flow mode to separate molecules according to size. [97]. Further studies applying SEC with fast performance liquid chromatography were performed [98,99]. These techniques allowed a preparative scale purification process, achieving high purity and high yields. However, SEC presents limitations, as it is not able to remove similar size impurities, such as dsDNA.

The use of ion pair reverse-phase chromatography (IPC) proved to be an excellent method for mRNA purification [44,95,100,101]. In IPC, the negatively charged sugar-phosphate backbone of the oligonucleotides will pair with quaternary ammonium compounds present in the mobile phase (in this case triethylammonium acetate) to become lipophilic and then interact with the stationary phase of a reverse-phase chromatography column [90]. Elution is then performed with a gradient of an adequate solvent, e.g., acetonitrile. Using this approach, dsRNA impurities are effectively removed while maintaining the process's high yield. However, IPC is challenging and costly to scale, and the use of toxic reagents such as acetonitrile, is not desirable. A new cellulose-based chromatography process for the removal of dsRNA has been described that leverages the ability of dsRNA to bind to cellulose in presence of ethanol [102]. This method reported a mRNA yield of >65% with a dsRNA removal of over 90%. Still, the removal of other impurities was not addressed, and thus the introduction of pre-purification steps is likely to be required.

Ion exchange chromatography (IEC) can also be used to purify mRNA at large scale. This technique explores the charge difference between the target mRNA species and the different impurities. For example, weak anion exchange chromatography has been successfully implemented to separate mRNA from IVT impurities [103]. IEC presents several advantages: it is scalable and cost-effective; it allows the separation of longer RNA transcripts; and it presents higher binding capacities (when compared with IPC) [104]. Nevertheless, this chromatography must be performed under denaturing conditions. This makes the process more complex as it requires a mobile phase heater and a tight control of the temperature during chromatography.

Affinity based separation is another mRNA purification approach. A single-stranded sequence of deoxythymidine (dT) - Oligo dT - is routinely used for the capture of mRNA in laboratory applications. This sequence binds to the poly-A tails present in the mRNA. Chromatographic beads with immobilized oligo dT could thus be used for the process scale purification using affinity chromatography: the poly-A tails of the single stranded mRNA produced during IVT would bind to the stationary phase while impurities are washed out. This way, IVT unconsumed reagents, the DNA template and dsRNA could be efficiently removed [105]. While high purity products can be obtained using affinity chromatography, several drawbacks are present such as low binding capacities and a less cost-effective process.

The removal of small size impurities can also be achieved while concentrating or diafiltrating solutions by tangential flow filtration (TFF) [106,107]. Core bead chromatography can also be used for this purpose [108]. In this case, small impurities are trapped inside the beads, and the product will be in the flowthrough. However, both techniques rely on DNase digestion or denaturing agents to remove high size molecules such as the DNA template or the polymerase. DNA removal can also be achieved using hydroxyapatite chromatography without the use of a DNase [108]. As a polishing step, hydrophobic interaction chromatography (HIC) can be applied using connective interaction media monolith (CIM) containing OH or SO<sub>3</sub> ligands [109].

Large scale adaptations of the traditional laboratory scale mRNA purification methods are also being explored. For example, mRNA precipitation can be combined with TFF technique [106]. During TFF, the membrane captures the precipitated mRNA product while other impurities are removed by diafiltration. The product is then eluted by re-solubilizing the mRNA. Furthermore, DNA template removal can be achieved by performing the digestion with immobilised DNase [110]. Another approach is to use tagged DNA template that can then be removed after IVT using affinity chromatography [110]. Despite being scalable, these methods present a limited effectiveness since they only focus on the removal of some specific impurities and hence must be coupled with other purification steps.

## New perspectives

The current IVT mRNA production methods must be improved to move mRNA technology to commercialisation and to support market demand. As process yields and production scale have an impact on the manufacturing costs and consequentially on the cost per dose [92], we speculate that continuous processing would have a particular advantage to lower costs. Continuous processing is already used in the chemical and pharmaceutical industry to run flexible and cost-effective processes and will ultimately offer on demand production. Additionally, the process integration made possible by continuous manufacturing may also reduce operation time and facilitate automation and process analytical technologies (PAT), which can result in a higher productivity and higher product quality [111,112]. The relative simplicity of mRNA manufacturing makes the process well suited for continuous processing, and in particular at a microfluidic scale (Figure 5). At this scale, reaction rates can be accelerated under specific conditions, the use of expensive reagents can be minimised, and cascade reactions can be compartmentalised easily [113]. Further, *in situ* product removal (ISPR) and substrate feed and product recovery (SFPR) strategies can be implemented in flow to facilitate process control, recirculation, and re-use of compounds [113]. These strategies will allow the separation of molecules, such as enzymes (if free enzymes are used), co-factors or NTPs, that can be recirculated in the process. Different unit operations, such as TFF, aqueous two-phase systems (ATPS) or precipitation, could be used for this purpose. These potentially will lower the burden on the downstream processing as well as the overall processing costs. Furthermore, the proposed system could be coupled with a microfluidic formulation step, in which the mRNA is encapsulated into lipid nanoparticles (LNPs) [115]. This would allow the establishment of continuous mRNA processing until the fill-to-finish steps.

Downstream processing, together with fill-to-finish, is still the major bottleneck in the mRNA vaccine production due to the lack of well-established and cost-effective processes. Despite the effort to develop methods that achieve high purity products, most of them are coupled with the traditional precipitation or nuclease digestion techniques [102,108]. Moreover, most methods are not cost-effective which can make the process infeasible for the market needs. Alternative cost-effective techniques, such as a single-pass tangential flow filtration (SPTFF) or aqueous two-phase systems (ATPS), that can be applied in a continuous mode, could potentially improve the process time and manufacturing flexibility while reducing cost and maintaining the quality [113]. Additionally, new chromatographic operation modes can overcome the need for having multiple mRNA purification steps (Figure 5). For example, the use of multimodal chromatography is highly promising as the combination of interactions between the molecule and the matrix could

result in an integrated and intensified purification process without the need for multiple chromatographic steps [114].

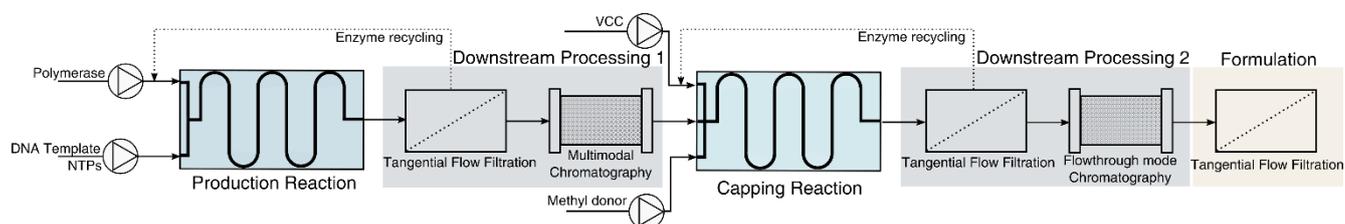


Figure 5. Conceptual design of a continuous manufacturing process for the production of mRNA vaccines. The process is composed of a 2-step enzymatic reaction in continuous form, followed by enzyme recycling using tangential flow filtration strategies and two multimodal chromatography steps, one in bind-elute mode for the intermediate purification, and a second in flowthrough mode for polishing. Formulation is achieved using a third tangential flow filtration module.

## mRNA Safety and Quality

mRNA manufacturing is advantageous when compared to the production of most biologicals since it does not require the use of cell cultures. Owing to its fast reaction time, the risk of contamination is lower than what is observed with other complex vaccine manufacturing processes. Additionally, the non-integrative nature and the transient expression inside the cells favours the mRNA safety profile [58,116].

Regulation guidelines for the evaluation of quality, safety and efficacy of RNA-based prophylactic vaccines for infection diseases are now being considered [117]. The emphasis is now on the establishment of manufacturing processes that can deliver a high quality and consistent product. Specifications for a number of critical process steps and acceptance criteria, intermediates, drug substances (DS) and drug product (DP) must therefore be defined, e.g., in terms of product yields, and analytical technologies that allows for rigorous product quantification and characterisation (product identity, purity and quality). mRNA quality can be assessed using several analytical techniques, such as gel electrophoresis and high-performance liquid chromatography (HPLC) [116], while the identity can be assured using sequencing techniques, such as reverse transcription polymerase chain reaction (RT-PCR) or next-generation sequencing [117]. The presence of residual amounts of DNA, enzymes and solvents [118], as well as dsRNA and truncated RNA fragments, must be determined. Additionally, as a general quality control, aspects like the presence of endotoxins, overall sterility and mRNA stability, must also be evaluated [117].

## Concluding Remarks

mRNA is a rising star in the field of biopharmaceuticals. The interest in this new type of vaccine derives from the flexibility, safety, and precision that these vaccines present when compared to

conventional approaches. The growing number of clinical trials for cancer therapies and infectious diseases demonstrates an increased interest from the industry to release these types of vaccines to the market. mRNA vaccines are precise, safe and flexible, which can be easily manufactured on a large scale for clinical grade applications. These vaccines can be an answer to quickly respond to epidemic outbreaks in terms of manufacturing.

However, to achieve this status, the development of sustainable and cost-effective manufacturing processes must be addressed. Although the IVT reaction of mRNA is safer and quicker than most of the established vaccines production, it relies on the use of expensive and limited materials. Downstream processing of the vaccine is still poorly established, and it is dependent on methods that lack scalability and cost-effectiveness. Moving the process to continuous manufacturing can overcome these bottlenecks. We propose a microfluidics approach with the compartmentalisation of enzymatic reactions coupled with *in situ* product removal (ISPR) and substrate feed and product recovery (SDPR) modules and the use of multimodal chromatography to replace the use of multiple chromatographic steps (Figure 5). The use of new production methods that allow the reuse and recirculation of compounds integrated with high-throughput purification and well-defined analytical methods in a continuous manufacturing process can be the answer for a sustainable, flexible and cost-effective vaccine manufacture that can allow an on-demand response.

## References

- [1] Andre F, Booy R, Bock H, Clemens J, Datta S, John T, et al. Vaccination greatly reduces disease, disability, death and inequity worldwide. *Bull World Health Organ* 2008;86:140–6. <https://doi.org/10.2471/BLT.07.040089>.
- [2] Jenner E. An inquiry into the causes and effects of the variolae vaccinae, a disease discovered in some of the western counties of England, particularly Gloucestershire, and known by the name of the cow pox. Sampson Low; 1798.
- [3] Pasteur L. Méthode pour prévenir la rage après morsure. *Comptes rendus Hebd Des séances De l'Académie Des Sci* 1885; 101:765–772.
- [4] Plotkin SA, Plotkin SL. The development of vaccines: how the past led to the future. *Nat Rev Microbiol* 2011;9:889–93. <https://doi.org/10.1038/nrmicro2668>.
- [5] Rappuoli R. Timeline: Vaccines. *Cell* 2020;183:552. <https://doi.org/10.1016/j.cell.2020.09.039>.
- [6] Plotkin S. History of vaccination. *Proc Natl Acad Sci* 2014;111:12283–12287. <https://doi.org/10.1073/pnas.1400472111>.
- [7] Kennedy J. Vaccine Hesitancy: A Growing Concern. *Pediatr Drugs* 2020;22:105–11. <https://doi.org/10.1007/s40272-020-00385-4>.
- [8] Bloom DE, Fan VY, Sevilla J. The broad socioeconomic benefits of vaccination. *Sci Transl Med* 2018;10. <https://doi.org/10.1126/scitranslmed.aaj2345>.

- [9] Bloom DE, Black S, Salisbury D, Rappuoli R. Antimicrobial resistance and the role of vaccines. *Proc Natl Acad Sci U S A* 2018;115:12868–71. <https://doi.org/10.1073/pnas.1717157115>.
- [10] Hollingsworth RE, Jansen K. Turning the corner on therapeutic cancer vaccines. *Npj Vaccines* 2019;4:1–10. <https://doi.org/10.1038/s41541-019-0103-y>.
- [11] Drury G, Jolliffe S, Mukhopadhyay TK. Process mapping of vaccines: Understanding the limitations in current response to emerging epidemic threats. *Vaccine* 2019;37:2415–21. <https://doi.org/10.1016/j.vaccine.2019.01.050>.
- [12] Mao HH, Chao S. Advances in Vaccines. *Curr Appl Pharm Biotechnol* 2020:155–188.
- [13] Oyston P, Robinson K. The current challenges for vaccine development. *J Med Microbiol* 2012;61:889–894. <https://doi.org/10.1099/jmm.0.039180-0>.
- [14] Rauch S, Jasny E, Schmidt KE, Petsch B. New Vaccine Technologies to Combat Outbreak Situations. *Front Immunol* 2018;9. <https://doi.org/10.3389/fimmu.2018.01963>.
- [15] Le TT, Andreadakis Z, Kumar A, Román RG, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. *Nat Rev Drug Discov* 2020;19:305–6. <https://doi.org/10.1038/d41573-020-00073-5>.
- [16] Wolff J, Malone R, Williams P, Chong W, Acsadi G, Jani A, et al. Direct gene transfer into mouse muscle in vivo. *Science* 1990;247:1465. <https://doi.org/10.1126/science.1690918>.
- [17] Conry RM, LoBuglio AF, Wright M, Sumerel L, Pike MJ, Johanning F, et al. Characterization of a Messenger RNA Polynucleotide Vaccine Vector. *Cancer Res* 1995; 55:1397–400.
- [18] Hoerr I, Obst R, Rammensee H-G, Jung G. In vivo application of RNA leads to induction of specific cytotoxic T lymphocytes and antibodies. *Eur J Immunol* 2000;30:1–7. <https://doi.org/10.1002/1521-4141>.
- [19] Qiu P, Ziegelhoffer P, Sun J, Yang N. Gene gun delivery of mRNA in situ results in efficient transgene expression and genetic immunization. *Gene Ther* 1996; 3:262–268.
- [20] Pollard C, De Koker S, Saelens X, Vanham G, Grooten J. Challenges and advances towards the rational design of mRNA vaccines. *Trends Mol Med* 2013;19:705–13. <https://doi.org/10.1016/j.molmed.2013.09.002>.
- [21] Maruggi G, Zhang C, Li J, Ulmer JB, Yu D. mRNA as a Transformative Technology for Vaccine Development to Control Infectious Diseases. *Mol Ther* 2019;27:757–72. <https://doi.org/10.1016/j.ymthe.2019.01.020>.
- [22] Sullenger BA, Nair S. From the RNA world to the clinic. *Science* 2016;352:1417–20. <https://doi.org/10.1126/science.aad8709>.
- [23] Steinle H, Behring A, Schlensak C, Wendel HP, Avci-Adali M. Concise review: application of in vitro transcribed messenger RNA for cellular engineering and reprogramming: progress and challenges. *Stem Cells* 2017;35:68–79. <https://doi.org/10.1002/stem.2402>.
- [24] Armbruster N, Jasny E, Petsch B. Advances in RNA vaccines for preventive indications: A case study of a vaccine against rabies. *Vaccines* 2019;7:132. <https://doi.org/10.3390/vaccines7040132>.
- [25] Linares-Fernández S, Lacroix C, Exposito J-Y, Verrier B. Tailoring mRNA Vaccine to Balance Innate/Adaptive Immune Response. *Trends Mol Med* 2020;26:311–23. <https://doi.org/10.1016/j.molmed.2019.10.002>.
- [26] Sonenberg N, Gingras A-C. The mRNA 5' cap-binding protein eIF4E and control of cell growth. *Curr*

- Opin Cell Biol 1998;10:268–275. [https://doi.org/10.1016/s0955-0674\(98\)80150-6](https://doi.org/10.1016/s0955-0674(98)80150-6).
- [27] Sahin U, Karikó K, Türeci Ö. mRNA-based therapeutics—developing a new class of drugs. *Nat Rev Drug Discov* 2014;13:759. <https://doi.org/10.1038/nrd4278>.
- [28] Tusup M, French LE, De Matos M, Gatfield D, Kundig T, Pascolo S. Design of in vitro Transcribed mRNA Vectors for Research and Therapy. *Chim Int J Chem* 2019;73:391–394. <https://doi.org/10.2533/chimia.2019.391>.
- [29] Yisraeli JK, Melton DA. Synthesis of long, capped transcripts in Vitro by SP6 and T7 RNA polymerases. *RNA Process. Part Gen. Methods*, vol. 180, Academic Press; 1989, p. 42–50. [https://doi.org/10.1016/0076-6879\(89\)80090-4](https://doi.org/10.1016/0076-6879(89)80090-4).
- [30] Ramanathan A, Robb GB, Chan S-H. mRNA capping: biological functions and applications. *Nucleic Acids Res* 2016;44:7511–7526. <https://doi.org/10.1093/nar/gkw551>.
- [31] Kwon H, Kim M, Seo Y, Moon YS, Lee HJ, Lee K, et al. Emergence of synthetic mRNA: In vitro synthesis of mRNA and its applications in regenerative medicine. *Biomaterials* 2018;156:172–93. <https://doi.org/10.1016/j.biomaterials.2017.11.034>.
- [32] Schlake T, Thess A, Fotin-Mleczek M, Kallen K-J. Developing mRNA-vaccine technologies. *RNA Biol* 2012;9:1319–30. <https://doi.org/10.4161/rna.22269>.
- [33] Geall AJ, Mandl CW, Ulmer JB. RNA: The new revolution in nucleic acid vaccines. *Semin Immunol* 2013;25:152–9. <https://doi.org/10.1016/j.smim.2013.05.001>.
- [34] Ross J. mRNA stability in mammalian cells. *Microbiol Mol Biol Rev* 1995; 59:423–450.
- [35] Lundstrom K. Replicon RNA viral vectors as vaccines. *Vaccines* 2016;4:39. <https://doi.org/10.3390/vaccines4040039>.
- [36] Vogel AB, Lambert L, Kinnear E, Busse D, Erbar S, Reuter KC, et al. Self-Amplifying RNA Vaccines Give Equivalent Protection against Influenza to mRNA Vaccines but at Much Lower Doses. *Mol Ther* 2018;26:446–55. <https://doi.org/10.1016/j.ymthe.2017.11.017>.
- [37] Beissert T, Perkovic M, Vogel A, Erbar S, Walzer KC, Hempel T, et al. A Trans-amplifying RNA vaccine strategy for induction of potent protective immunity. *Mol Ther* 2020;28:119–28. <https://doi.org/10.1016/j.ymthe.2019.09.009>.
- [38] Weng Y, Li C, Yang T, Hu B, Zhang M, Guo S, et al. The challenge and prospect of mRNA therapeutics landscape. *Biotechnol Adv* 2020;40:107534. <https://doi.org/10.1016/j.biotechadv.2020.107534>.
- [39] Zeng C, Zhang C, Walker PG, Dong Y. Formulation and delivery technologies for mRNA Vaccines 2020:1–40. [https://doi.org/10.1007/82\\_2020\\_217](https://doi.org/10.1007/82_2020_217).
- [40] Johanning F, Conry R, LoBuglio A, Wright M, Sumerel L, Pike M, et al. A Sindbis virus mRNA polynucleotide vector achieves prolonged and high level heterologous gene expression in vivo. *Nucleic Acids Res* 1995;23:1495–1501. <https://doi.org/10.1093/nar/23.9.1495>.
- [41] Diken M, Kreiter S, Selmi A, Britten C, Huber C, Türeci Ö, et al. Selective uptake of naked vaccine RNA by dendritic cells is driven by macropinocytosis and abrogated upon DC maturation. *Gene Ther* 2011;18:702–708. <https://doi.org/10.1038/gt.2011.17>.
- [42] Kreiter S, Selmi A, Diken M, Koslowski M, Britten CM, Huber C, et al. Intranodal vaccination with naked antigen-encoding RNA elicits potent prophylactic and therapeutic antitumoral immunity. *Cancer Res* 2010;70:9031–9040. <https://doi.org/10.1158/0008-5472.CAN-10-0699>.
- [43] Kyte JA, Aamdal S, Dueland S, Sæbøe-Larsen S, Inderberg EM, Madsbu UE, et al. Immune response and long-term clinical outcome in advanced melanoma patients vaccinated with tumor-mRNA-

- transfected dendritic cells. *OncolImmunology* 2016;5:e1232237. <https://doi.org/10.1080/2162402X.2016.1232237>.
- [44] Probst J, Weide B, Scheel B, Pichler BJ, Hoerr I, Rammensee H-G, et al. Spontaneous cellular uptake of exogenous messenger RNA in vivo is nucleic acid-specific, saturable and ion dependent. *Gene Ther* 2007;14:1175–80. <https://doi.org/10.1038/sj.gt.3302964>.
- [45] Kowalski PS, Rudra A, Miao L, Anderson DG. Delivering the Messenger: Advances in Technologies for Therapeutic mRNA Delivery. *Mol Ther* 2019;27:710–28. <https://doi.org/10.1016/j.ymthe.2019.02.012>.
- [46] Pardi N, Hogan MJ, Pelc RS, Muramatsu H, Andersen H, DeMaso CR, et al. Zika virus protection by a single low-dose nucleoside-modified mRNA vaccination. *Nature* 2017;543:248–51. <https://doi.org/10.1038/nature21428>.
- [47] Richner JM, Himansu S, Dowd KA, Butler SL, Salazar V, Fox JM, et al. Modified mRNA Vaccines Protect against Zika Virus Infection. *Cell* 2017;168:1114–1125.e10. <https://doi.org/10.1016/j.cell.2017.02.017>.
- [48] Bahl K, Senn JJ, Yuzhakov O, Bulychev A, Brito LA, Hassett KJ, et al. Preclinical and Clinical Demonstration of Immunogenicity by mRNA Vaccines against H10N8 and H7N9 Influenza Viruses. *Mol Ther* 2017;25:1316–27. <https://doi.org/10.1016/j.ymthe.2017.03.035>.
- [49] Feldman RA, Fuhr R, Smolenov I, (Mick) Ribeiro A, Panther L, Watson M, et al. mRNA vaccines against H10N8 and H7N9 influenza viruses of pandemic potential are immunogenic and well tolerated in healthy adults in phase 1 randomized clinical trials. *Vaccine* 2019;37:3326–34. <https://doi.org/10.1016/j.vaccine.2019.04.074>.
- [50] Zhao M, Li M, Zhang Z, Gong T, Sun X. Induction of HIV-1 gag specific immune responses by cationic micelles mediated delivery of gag mRNA. *Drug Deliv* 2016;23:2596–607. <https://doi.org/10.3109/10717544.2015.1038856>.
- [51] Li M, Zhao M, Fu Y, Li Y, Gong T, Zhang Z, et al. Enhanced intranasal delivery of mRNA vaccine by overcoming the nasal epithelial barrier via intra-and paracellular pathways. *J Controlled Release* 2016;228:9–19. <https://doi.org/10.1016/j.jconrel.2016.02.043>.
- [52] Blakney AK, Abdouni Y, Yilmaz G, Liu R, McKay PF, Bouton CR, et al. Mannosylated Poly(ethylene imine) Copolymers Enhance saRNA Uptake and Expression in Human Skin Explants. *Biomacromolecules* 2020;21:2482–92. <https://doi.org/10.1021/acs.biomac.0c00445>.
- [53] Weide B, Pascolo S, Scheel B, Derhovanessian E, Pflugfelder A, Eigentler TK, et al. Direct Injection of Protamine-protected mRNA: Results of a Phase 1/2 Vaccination Trial in Metastatic Melanoma Patients. *J Immunother* 2009;32:498–507. <https://doi.org/10.1097/CJI.0b013e3181a00068>.
- [54] Udhayakumar VK, Beuckelaer AD, McCaffrey J, McCrudden CM, Kirschman JL, Vanover D, et al. Arginine-Rich Peptide-Based mRNA Nanocomplexes Efficiently Instigate Cytotoxic T Cell Immunity Dependent on the Amphipathic Organization of the Peptide. *Adv Healthc Mater* 2017;6:1601412. <https://doi.org/10.1002/adhm.201601412>.
- [55] Coolen A-L, Lacroix C, Mercier-Gouy P, Delaune E, Monge C, Exposito J-Y, et al. Poly(lactic acid) nanoparticles and cell-penetrating peptide potentiate mRNA-based vaccine expression in dendritic cells triggering their activation. *Biomaterials* 2019;195:23–37. <https://doi.org/10.1016/j.biomaterials.2018.12.019>.
- [56] Bell GD, Yang Y, Leung E, Krissansen GW. mRNA transfection by a Xentry-protamine cell-penetrating peptide is enhanced by TLR antagonist E6446. *PLOS ONE* 2018;13:e0201464.

<https://doi.org/10.1371/journal.pone.0201464>.

- [57] Lou B, De Koker S, Lau CYJ, Hennink WE, Mastrobattista E. mRNA Polyplexes with Post-Conjugated GALA Peptides Efficiently Target, Transfect, and Activate Antigen Presenting Cells. *Bioconjug Chem* 2019;30:461–75. <https://doi.org/10.1021/acs.bioconjchem.8b00524>.
- [58] Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines—a new era in vaccinology. *Nat Rev Drug Discov* 2018;17:261. <https://doi.org/10.1038/nrd.2017.243>.
- [59] ClinicalTrials.gov. Immunotherapy in Treating Patients With Metastatic Breast Cancer. Identifier: NCT00003432 2003. <https://clinicaltrials.gov/ct2/show/NCT00003432> (accessed December 7, 2020).
- [60] Boczkowski D, Nair SK, Snyder D, Gilboa E. Dendritic cells pulsed with RNA are potent antigen-presenting cells in vitro and in vivo. *J Exp Med* 1996;184:465–72. <https://doi.org/10.1084/jem.184.2.465>.
- [61] Lichtenegger FS, Schnorfeil FM, Rothe M, Deiser K, Altmann T, Bücklein VL, et al. Toll-like receptor 7/8-matured RNA-transduced dendritic cells as post-remission therapy in acute myeloid leukaemia: results of a phase I trial. *Clin Transl Immunol* 2020;9:e1117. <https://doi.org/10.1002/cti2.1117>.
- [62] Beatty GL, Haas AR, Maus MV, Torigian DA, Soulen MC, Plesa G, et al. Mesothelin-Specific Chimeric Antigen Receptor mRNA-Engineered T Cells Induce Antitumor Activity in Solid Malignancies. *Cancer Immunol Res* 2014;2:112–20. <https://doi.org/10.1158/2326-6066.CIR-13-0170>.
- [63] Hung C-F, Xu X, Li L, Ma Y, Jin Q, Viley A, et al. Development of Anti-Human Mesothelin-Targeted Chimeric Antigen Receptor Messenger RNA-Transfected Peripheral Blood Lymphocytes for Ovarian Cancer Therapy. *Hum Gene Ther* 2018;29:614–25. <https://doi.org/10.1089/hum.2017.080>.
- [64] Xiao L, Cen D, Gan H, Sun Y, Huang N, Xiong H, et al. Adoptive Transfer of NKG2D CAR mRNA-Engineered Natural Killer Cells in Colorectal Cancer Patients. *Mol Ther* 2019;27:1114–25. <https://doi.org/10.1016/j.ymthe.2019.03.011>.
- [65] Pascolo S. The messenger’s great message for vaccination. *Expert Rev Vaccines* 2015;14:153–6. <https://doi.org/10.1586/14760584.2015.1000871>.
- [66] Weide B, Carralot J-P, Reese A, Scheel B, Eigentler TK, Hoerr I, et al. Results of the First Phase I/II Clinical Vaccination Trial With Direct Injection of mRNA. *J Immunother* 2008;31:180–8. <https://doi.org/10.1097/CJI.0b013e31815ce501>.
- [67] Kallen K-J, Heidenreich R, Schnee M, Petsch B, Schlake T, Thess A, et al. A novel, disruptive vaccination technology: Self-adjuvanted RNActive® vaccines. *Hum Vaccines Immunother* 2013;9:2263–76. <https://doi.org/10.4161/hv.25181>.
- [68] Buonaguro L, Mayer A, Loeffler M, Missel S, Accolla R, Ma YT, et al. Abstract LB-094: Hepavac-101 first-in-man clinical trial of a multi-peptide-based vaccine for hepatocellular carcinoma. *Cancer Res* 2020;80:LB-094-LB-094. <https://doi.org/10.1158/1538-7445.AM2020-LB-094>.
- [69] Kübler H, Scheel B, Gnad-Vogt U, Miller K, Schultze-Seemann W, vom Dorp F, et al. Self-adjuvanted mRNA vaccination in advanced prostate cancer patients: a first-in-man phase I/IIa study. *J Immunother Cancer* 2015;3:26. <https://doi.org/10.1186/s40425-015-0068-y>.
- [70] Sebastian M, Schröder A, Scheel B, Hong HS, Muth A, von Boehmer L, et al. 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* 2019;68:799–812. <https://doi.org/10.1007/s00262-019-02315-x>.
- [71] Matsui A, Uchida S, Ishii T, Itaka K, Kataoka K. Messenger RNA-based therapeutics for the treatment of apoptosis-associated diseases. *Sci Rep* 2015;5:1–10. <https://doi.org/10.1038/srep15810>.
- [72] Sahin U, Oehm P, Derhovanessian E, Jabulowsky RA, Vormehr M, Gold M, et al. An RNA vaccine drives immunity in checkpoint-inhibitor-treated melanoma. *Nature* 2020;585:107–12. <https://doi.org/10.1038/s41586-020-2537-9>.
- [73] Patel MR, Bauer TM, Jimeno A, Wang D, LoRusso P, Do KT, et al. A phase I study of mRNA-2752, a lipid nanoparticle encapsulating mRNAs encoding human OX40L, IL-23, and IL-36γ, for intratumoral (iTu) injection alone and in combination with durvalumab. *J Clin Oncol* 2020;38:3092–3092. [https://doi.org/10.1200/JCO.2020.38.15\\_suppl.3092](https://doi.org/10.1200/JCO.2020.38.15_suppl.3092).
- [74] Kranz LM, Diken M, Haas H, Kreiter S, Loquai C, Reuter KC, et al. Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature* 2016;534:396–401. <https://doi.org/10.1038/nature18300>.
- [75] Burris HA, Patel MR, Cho DC, Clarke JM, Gutierrez M, Zaks TZ, et al. A phase I multicenter study to assess the safety, tolerability, and immunogenicity of mRNA-4157 alone in patients with resected solid tumors and in combination with pembrolizumab in patients with unresectable solid tumors. *J Clin Oncol* 2019;37:2523–2523. [https://doi.org/10.1200/JCO.2019.37.15\\_suppl.2523](https://doi.org/10.1200/JCO.2019.37.15_suppl.2523).
- [76] Cafri G, Gartner JJ, Hopson K, Meehan RS, Zaks TZ, Robbins P, et al. Immunogenicity and tolerability of personalized mRNA vaccine mRNA-4650 encoding defined neoantigens expressed by the autologous cancer. *J Clin Oncol* 2019;37:2643–2643. [https://doi.org/10.1200/JCO.2019.37.15\\_suppl.2643](https://doi.org/10.1200/JCO.2019.37.15_suppl.2643).
- [77] Leal L, Guardo AC, Morón-López S, Salgado M, Mothe B, Heirman C, et al. Phase I clinical trial of an intranodally administered mRNA-based therapeutic vaccine against HIV-1 infection. *AIDS Lond Engl* 2018;32:2533–45. <https://doi.org/10.1097/QAD.0000000000002026>.
- [78] Gandhi RT, Kwon DS, Macklin EA, Shopis JR, McLean AP, McBrine N, et al. Immunization of HIV-1-Infected Persons With Autologous Dendritic Cells Transfected With mRNA Encoding HIV-1 Gag and Nef: Results of a Randomized, Placebo-Controlled Clinical Trial. *J Acquir Immune Defic Syndr* 1999 2016;71:246–53. <https://doi.org/10.1097/QAI.0000000000000852>.
- [79] Allard SD, De Keersmaecker B, de Goede AL, Verschuren EJ, Koetsveld J, Reedijk ML, et al. A phase I/IIa immunotherapy trial of HIV-1-infected patients with Tat, Rev and Nef expressing dendritic cells followed by treatment interruption. *Clin Immunol* 2012;142:252–68. <https://doi.org/10.1016/j.clim.2011.10.010>.
- [80] Alberer M, Gnad-Vogt U, Hong HS, Mehr KT, Backert L, Finak G, et al. Safety and immunogenicity of a mRNA rabies vaccine in healthy adults: an open-label, non-randomised, prospective, first-in-human phase 1 clinical trial. *The Lancet* 2017;390:1511–20. [https://doi.org/10.1016/S0140-6736\(17\)31665-3](https://doi.org/10.1016/S0140-6736(17)31665-3).
- [81] Petsch B, Schnee M, Vogel AB, Lange E, Hoffmann B, Voss D, et al. Protective efficacy of in vitro synthesized, specific mRNA vaccines against influenza A virus infection. *Nat Biotechnol* 2012;30:1210–6. <https://doi.org/10.1038/nbt.2436>.
- [82] Mulligan MJ, Lyke KE, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Phase 1/2 study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature* 2020:1–5. <https://doi.org/10.1038/s41586-020->

2639-4.

- [83] Jackson LA, Anderson EJ, Roupael NG, Roberts PC, Makhene M, Coler RN, et al. An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. *N Engl J Med* 2020 <https://doi.org/10.1056/NEJMoa2022483>.
- [84] Anderson EJ, Roupael NG, Widge AT, Jackson LA, Roberts PC, Makhene M, et al. Safety and Immunogenicity of SARS-CoV-2 mRNA-1273 Vaccine in Older Adults. *N Engl J Med* 2020; <https://doi.org/10.1056/NEJMoa2028436>.
- [85] Widge AT, Roupael NG, Jackson LA, Anderson EJ, Roberts PC, Makhene M, et al. Durability of Responses after SARS-CoV-2 mRNA-1273 Vaccination. *N Engl J Med*. <https://doi.org/10.1056/NEJMc2032195>.
- [86] Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* 2020;383:2603–15. <https://doi.org/10.1056/NEJMoa2034577>.
- [87] Sun N, Ning B, Hansson KM, Bruce AC, Seaman SA, Zhang C, et al. Modified VEGF-A mRNA induces sustained multifaceted microvascular response and accelerates diabetic wound healing. *Sci Rep* 2018;8:17509. <https://doi.org/10.1038/s41598-018-35570-6>.
- [88] Gan L-M, Lagerström-Fermér M, Carlsson LG, Arfvidsson C, Egnell A-C, Rudvik A, et al. Intradermal delivery of modified mRNA encoding VEGF-A in patients with type 2 diabetes. *Nat Commun* 2019;10:871. <https://doi.org/10.1038/s41467-019-08852-4>.
- [89] Fuchs A-L, Neu A, Sprangers R. A general method for rapid and cost-efficient large-scale production of 5' capped RNA. *RNA* 2016;22:1454–66. <https://doi.org/10.1261/rna.056614.116>.
- [90] Baronti L, Karlsson H, Marušič M, Petzold K. A guide to large-scale RNA sample preparation. *Anal Bioanal Chem* 2018;410:3239–52. <https://doi.org/10.1007/s00216-018-0943-8>.
- [91] Wochner A, Roos T, Ketterer T. Methods and means for enhancing rna production. US20170114378A1, 2017.
- [92] Kis Z, Kontoravdi C, Shattock R, Shah N. Resources, Production Scales and Time Required for Producing RNA Vaccines for the Global Pandemic Demand. *Vaccines* 2021;9:3. <https://doi.org/10.3390/vaccines9010003>.
- [93] Kis Z, Shattock R, Shah N, Kontoravdi C. Emerging Technologies for Low-Cost, Rapid Vaccine Manufacture. *Biotechnol J* 2019;14:1800376. <https://doi.org/10.1002/biot.201800376>.
- [94] Pascolo S. Messenger RNA: The Inexpensive Biopharmaceutical. *J Multidiscip Eng Sci Technol JMEST* 2017; 4:6937–41.
- [95] Kariko K, Muramatsu H, Ludwig J, Weissman D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Res* 2011;39:e142–e142. <https://doi.org/10.1093/nar/gkr695>.
- [96] Akash MSH, Rehman K. Column Chromatography. In: Akash MSH, Rehman K, editors. *Essent. Pharm. Anal.*, Singapore: Springer Singapore; 2020, p. 167–74. [https://doi.org/10.1007/978-981-15-1547-7\\_13](https://doi.org/10.1007/978-981-15-1547-7_13).
- [97] Lukavsky PJ, Puglisi JD. Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides. *RNA* 2004;10:889–93. <https://doi.org/10.1261/rna.5264804>.
- [98] Kim I, McKenna SA, Puglisi EV, Puglisi JD. Rapid purification of RNAs using fast performance liquid

- chromatography (FPLC). *RNA* 2007;13:289–94. <https://doi.org/10.1261/rna.342607>.
- [99] McKenna SA, Kim I, Puglisi EV, Lindhout DA, Aitken CE, Marshall RA, et al. Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nat Protoc* 2007;2:3270–7. <https://doi.org/10.1038/nprot.2007.480>.
- [100] Pascolo S. Messenger RNA-based vaccines. *Expert Opin Biol Ther* 2004;4:1285–94. <https://doi.org/10.1517/14712598.4.8.1285>.
- [101] Weissman D, Pardi N, Muramatsu H, Karikó K. HPLC Purification of In Vitro Transcribed Long RNA. In: Rabinovich PM, editor. *Synth. Messenger RNA Cell Metab. Modul. Methods Protoc.*, Totowa, NJ: Humana Press; 2013, p. 43–54. [https://doi.org/10.1007/978-1-62703-260-5\\_3](https://doi.org/10.1007/978-1-62703-260-5_3).
- [102] Baiersdörfer M, Boros G, Muramatsu H, Mahiny A, Vlatkovic I, Sahin U, et al. A Facile Method for the Removal of dsRNA Contaminant from In Vitro-Transcribed mRNA. *Mol Ther - Nucleic Acids* 2019;15:26–35. <https://doi.org/10.1016/j.omtn.2019.02.018>.
- [103] Easton LE, Shibata Y, Lukavsky PJ. Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* 2010;16:647–53. <https://doi.org/10.1261/rna.1862210>.
- [104] Issa WJ, Barberio JL, Aunins JG, Afeyan NB. Ion exchange purification of mrna. US/2016/0024141 A1, 2016.
- [105] Bancel S, Issa WJ, Aunins JG, Chakraborty T. Manufacturing methods for production of RNA transcripts. WO/2014/152027; PCT/US2014/026835; US20160024547A1, 2016.
- [106] Heartlein M, DeRosa F, Dias A, Karve S. Methods for purification of messenger rna. WO/2014/152966A1, 2014.
- [107] Funkner A, Dorner S, Sewing S, Johannes K, Broghammer N, Ketterer T, et al. Method for producing and purifying RNA, comprising at least one step of tangential flow filtration. PCT/EP2016/062152; WO/2016/193206, 2018.
- [108] Von Der Mülbe F, Reidel L, Ketterer T, Gontcharova L, Bauer S, Pascolo S, et al. Method for producing rna. PCT/EP2015/000959; US10017826B2; WO/2016/180430A1, 2015.
- [109] Funkner A, Sewing S, Strobel I, Mutzke T. Method for purifying rna. WO/2018/096179A1, 2018.
- [110] Issa WJ, Wang Y, Bancel S. Removal of DNA fragments in mRNA production process. WO/2014/152030A1; US10077439B2, 2016.
- [111] Walther J, Godawat R, Hwang C, Abe Y, Sinclair A, Konstantinov K. The business impact of an integrated continuous biomanufacturing platform for recombinant protein production. *J Biotechnol* 2015;213:3–12. <https://doi.org/10.1016/j.jbiotec.2015.05.010>.
- [112] Kapoor Y, Meyer RF, Meyer BK, DiNunzio JC, Bhambhani A, Stanbro J, et al. Flexible Manufacturing: The Future State of Drug Product Development and Commercialization in the Pharmaceutical Industry. *J Pharm Innov* 2020. <https://doi.org/10.1007/s12247-019-09426-z>.
- [113] Gruber P, Marques MPC, O’Sullivan B, Baganz F, Wohlgemuth R, Szita N. Conscious coupling: The challenges and opportunities of cascading enzymatic microreactors. *Biotechnol J* 2017; 12:1700030. <https://doi.org/10.1002/biot.201700030>.
- [117] Fisher AC, Kamga M-H, Agarabi C, Brorson K, Lee SL, Yoon S. The Current Scientific and Regulatory Landscape in Advancing Integrated Continuous Biopharmaceutical Manufacturing. *Trends Biotechnol* 2019; 37:253–67. <https://doi.org/10.1016/j.tibtech.2018.08.008>.
- [114] Halan V, Maity S, Bhambure R, Rathore AS. Multimodal Chromatography for Purification of

Biotherapeutics - A Review. *Curr Protein Pept Sci* 2019. <https://doi.org/10.2174/1389203718666171020103559>.

- [115] Dimov N, Kastner E, Hussain M, Perrie Y, Szita N. Formation and purification of tailored liposomes for drug delivery using a module-based micro continuous-flow system. *Sci Rep* 2017; 7:12045. <https://doi.org/10.1038/s41598-017-11533-1>.
- [116] Poveda C, Biter AB, Bottazzi ME, Strych U. Establishing Preferred Product Characterization for the Evaluation of RNA Vaccine Antigens. *Vaccines* 2019; 7:131. <https://doi.org/10.3390/vaccines7040131>.
- [117] World Health Organization. Call for Public Consultation – Evaluation of the quality, safety and efficacy of RNA-based prophylactic vaccines for infectious diseases: regulatory considerations n.d. <https://www.who.int/news-room/articles-detail/call-for-public-consultation-evaluation-of-the-quality-safety-and-efficacy-of-rna-based-prophylactic-vaccines-for-infectious-diseases-regulatory-considerations> (accessed January 31, 2021).
- [118] Schmid A. Considerations for Producing mRNA Vaccines for Clinical Trials. *RNA Vaccines* 2017:237–51. [https://doi.org/10.1007/978-1-4939-6481-9\\_15](https://doi.org/10.1007/978-1-4939-6481-9_15).

# Chapter II – The ABC of mRNA Manufacturing: a holistic perspective into process design and optimisation

---

**Key words:** mRNA manufacturing; Process design: Patient-centric approach

# Introduction

Vaccination is one of the most effective methods to control and prevent diseases. Recently, it has been estimated that during the COVID-19 pandemic event vaccines prevented a total of 151 million new infections and up to 620 thousand deaths to occur[1]. This was only possible because of the reduced development time of novel vaccine technology platforms that presented high efficacy and flexibility[2], allowing the manufacturing and delivery of these vaccines in a record time. This is in stark contrast with traditional vaccine development pipeline, where takes an average of 15 years, between the initial discovery and the vaccine licensure and policy recommendation [3]. This presents a major hurdle when a quick response is required in a case of an epidemic event. Evidently, other challenges arise that are related to the highly divergent nature of viruses due to high mutation rates, which impact the immunisation coverage[4].

mRNA vaccines are a new class platform technology that potentially overcome these bottlenecks. Their versatile was shown during the COVID-19 pandemic and these vaccines are ideal candidates for the ambition to develop a safe, effective and accessible response to epidemic events in 100 days[5]. The safety aspect relates transient expression of the gene of interest in the cytosol and precision to the stimulation of both innate and adaptive immune response[6]. The manufacturing is fairly straightforward: mRNA is produced in a cell-free system, where template DNA containing the gene of interest sequence, is *in vitro* transcribed. Since the physico-chemical characteristics of the mRNA are independent of the gene of interest, the process is flexible and can be standardised. mRNA vaccines are not only restricted to the prophylactic treatment field but are now being explored to be used for cancer treatment, protein replacement or genome engineering therapies[7].

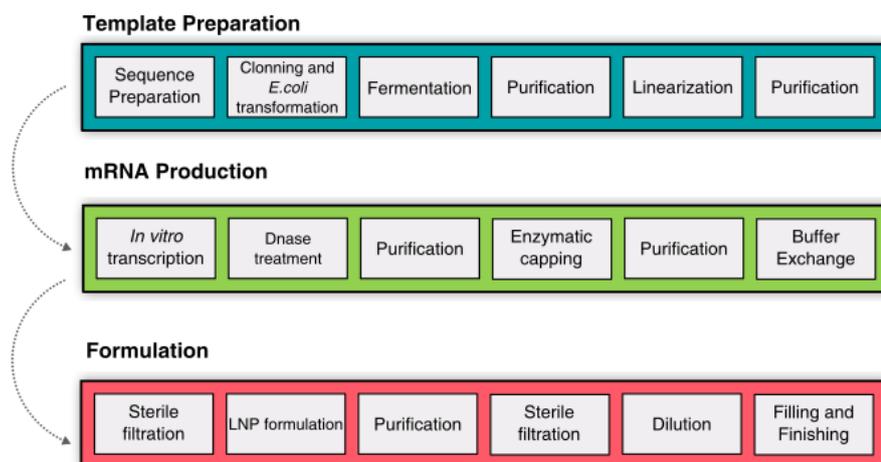


Figure 1. Schematic representation of all the steps of mRNA manufacturing process, since the development of the template sequence and DNA production, to the mRNA production, purification and formulation.

In this review, we look into all the steps of the mRNA vaccines manufacturing process, since the design of the template to the formulation steps (Figure 1) in patient-centric point of view. The flexibility advantage of the mRNA technology platform is also applied to its manufacturing process. DNA Template sequence can be optimised to maximise protein expression (e.g. codon optimisation), as well as mRNA stability and immunogenicity can be modulated by the non-coding regions used. DNA template production delivers an optimised template sequence in a linear form. Multiple production and purification strategies, cell-free and *E.coli* based, can be used according to the quantity range ( $\mu\text{g}$  or  $\text{mg}$ ) required. mRNA is produced by an enzymatic reaction that uses the pDNA produced and NTPs as substrates. This production can have different modalities (e.g. fed-batch, batch), and multiple purification strategies can be followed during downstream processing, depending on the quantity and quality required. mRNA delivery can only be successfully achieved using carrier-based system. These systems not only allow to stabilise, but its optimisation allows for an efficient and targeted delivery. Fully characterisation of the mRNA and its impurities throughout the process enable a tight manufacturing process control and precise product characterisation that are required to deliver high-quality product with little batch-to-batch variability.

## mRNA vaccines

The core of the messenger RNA (mRNA) is that a single-stranded RNA is responsible to carry genetic information from the nucleus to the cell cytosol, where the cell machinery will translate this information into a protein. This mechanism, present in both prokaryotes and eukaryotes, is the principle of the mRNA vaccines - any synthetic transcripts can induce the expression of protein *in situ* once it reaches the cytosol. This approach allows remarkable flexibility in the protein of interest, and the post-translational modification of these proteins is completed by the host cells, bypassing costly and inflexible cell-based manufacturing processes[8].

In eukaryotes, DNA is transcribed into RNA by the RNA polymerase II (Pol II) inside the cell nucleus. The mRNA is then processed, firstly by the addition of a 5' cap by converting the pppN 5' terminus into a m<sup>7</sup>GpppN using three enzymes (triphosphatase, guanyl transferase, and methyltransferase)[9]. The pre-mRNA is spliced to remove non-coding regions, and the transcription is finalised by the addition of a poly-A tail by a process named polyadenylation[10]. An eukaryotic mature mRNA presents a coding region flanked by 5' and 3' untranslated regions (UTRs)[11], a 5' cap and poly-A tail. mRNA vaccines mimic this structure (Figure 2.a). The gene of interest is flanked by the UTRs. These non-protein-coding sequences are crucial in controlling the levels of *in-situ* expression of the protein of interest. The 5' terminal nucleotide, which consists of 7-methylguanosine, has an important role in reducing the level of activation of innate immune mechanisms, as well as protection against exonuclease degradation[6, 12]. The 3' end of eukaryotic mRNA typically contains a terminal region known as the poly(A) tail. This region,

which typically consists of around 200 adenine nucleotides, have an important role in translation efficiency and mRNA stability. Owin to being highly conserved, the tail is an ideal target for downstream processing.

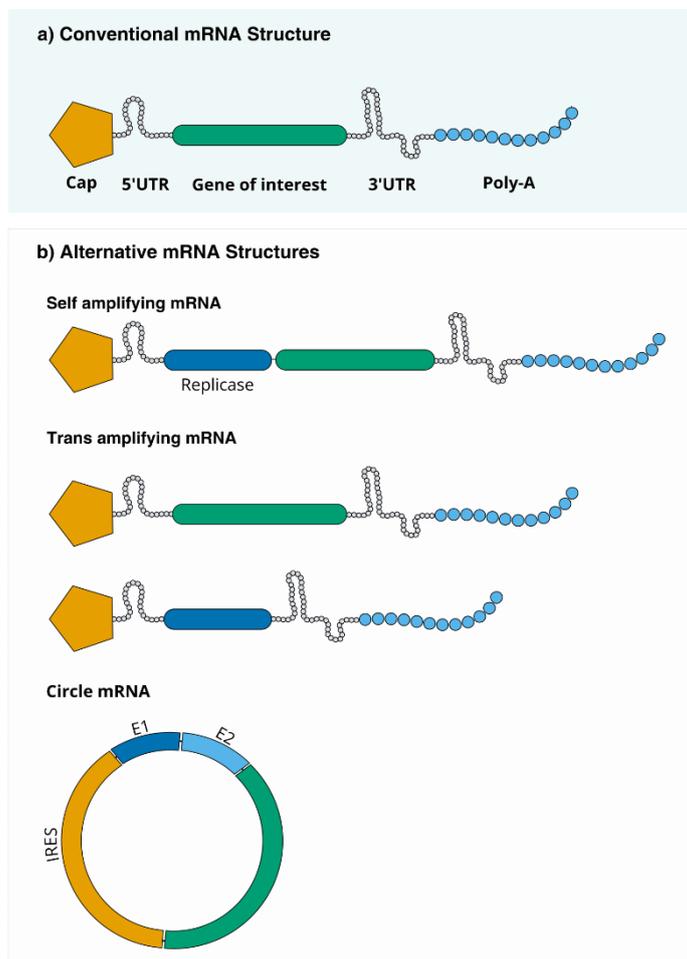


Figure 2. Conventional mRNA structure (a) and structure alternative (b)

Alternatives to conventional mRNA vaccine structures are being explored to improve the vaccine efficiency and stability as well as decrease doses required (Figure 2.b). In addition to the coding section for the protein of interest, self-amplifying mRNA contains sections encoding non-structural proteins (NSPs) which form replicative machinery *in situ*. The NSP encoding region is translated and forms RNA-dependent RNA polymerase (RDRP). RDRP synthesises a negative-sense strand from the saRNA, which contains both NSP and protein of interest regions. From this negative sense strand, two different positive sense strands are then transcribed. One encodes the NSPs and protein of interest; and one only encodes the protein of interest. The enhanced transcription of the protein coding region leads to enhanced levels of translation and therefore expression. This replicative machinery is typically based on an alphavirus genome[13]. Due to the additional regions encoding NSPs, the overall strand length of saRNA constructs can often exceed 10 kb. An advantage of *in situ* amplification of this RNA transcript is that a smaller dose is required than non-amplifying mRNA to elicit the same level of expression of the protein

of interest. Vogel *et al.*[14] showed that a 1.25 µg dose of an saRNA influenza vaccine was required to produce protective immunity in mice, whilst an equivalent non-amplifying mRNA vaccine required 80 µg, which means that an equivalent immune response is achieved with a 64 x smaller dosage.

An alternative mRNA that also explores the use of RDRPs is the trans-amplifying RNA (taRNA). In this case, these polymerases are encoded on a separate RNA strand to the coding sequence for the protein of interest. This creates a split-vector system with the transreplicase encoding strand, and strand containing the gene of interest (transreplicon) administered in tandem. The strand encoding the RDRPs may itself be self-amplifying, or non-amplifying. taRNA using a non-replicating replicase template was shown to exhibit expression as efficient as cis-acting saRNA. Owing to the shorter length of the taRNA, it overcomes the synthesis bottleneck of the saRNA[15].

Circular RNA (circRNA) is single stranded mRNA, that commonly occurs in eukaryotic cells, that is covalently bound end to end, as a continuous loop[16]. Synthetic circRNA is now emerging as a drug substance with a wide variety of potential clinical applications. Owing to the closed-loop nature of circRNA, it lacks 5' and 3' ends, as well as 5' cap. Translation initiation relies on the use of internal ribosome entry sites[17]. This type of RNA is not susceptible to exonuclease degradation, presents a higher range of temperature stability and longer expression times[18].

## mRNA designs

Non-coding regions in mRNA design have a strong impact on mRNA stability, delivery efficiency, overall protein expression and on the life-time within cell itself. The expression can be regulated by the specific cis-regulatory elements, which will interact with the trans-acting factors, present in the non-coding fraction of the DNA. These include promoters, enhancers and other elements that will produce the mature mRNA[19]. Furthermore, the post-transcription regulation is usually also achieved by these cis-regulatory elements, and ultimately modulate the subcellular localisation, translation efficiency and stability of mRNA.

The sequence itself can also have an impact on the efficiency and stability of mRNA, as well as secondary structures or the use of modified nucleosides. In their COVID vaccines, BioNTech and Moderna performed codon optimisation and used modified uracil to improve mRNA vaccine efficiency[20]. However, while BioNTech used an amino-terminal enhancer of split (AES) mRNA and the mitochondrial encoded 12S ribosomal RNA as 3' UTR, Moderna used a Homo sapiens haemoglobin subunit alpha 1 gene. These are notable examples of different strategies used to optimise non-coding regions (Table 1).

## Untranslated regions

The 5'UTR sequence and secondary structure impacts translation efficiency, and therefore the optimisation of this sequence can significantly increase translation. To initiate translation, the eukaryotic initiation factor 4F (eIF4F) complex binds to the 5' 7-methyl guanosine cap[21]. The eukaryotic translation initiation factor 4 G (eIF4G) interacts with poly-A binding proteins (PABP), circularising the mRNA[22], and the 43S pre-initiation complex will scan the 5' UTR until the initiation codon is recognised [23]. Afterwards, the large ribosomal subunit (60s) joins the previous complex to form the elongation-competent 80s ribosome and the translation elongation is initiated.

To achieve maximal mRNA translation a strong Kozak consensus sequence is usually required[24]. The use of the 5' UTR beta-globin in combination with a strong Kozak sequence can increase the translation efficiency and duration[25]. Nevertheless, a 5'UTR complement factor 3 (C3) and cytochrome p450E1 are potentially more efficient[26]. Translation yields can also be improved with high ribosome loading synthetic sequences but this approach is highly dependent on the targeted cell lines[27]. The use of minimal 5' synthetic UTRs (< 15 nucleotides) can yield higher expressions levels compared to human alpha-globin 5'-UTR[28]. 5'UTRs from TOP genes[29], namely HSD17B4[30] showed improved immunogenicity when compared with 5'UTR used in CVnCoV. Recruitment of eIF4G plays an important role in translation efficiency. The use of a 40 nt aptamer corresponding to the eIF4G binding site in combination with a strong Kozak sequence can also increase the translation efficiency by several folds due to direct recruiting[31].

UTRs can also be made up from highly expressed genes from the targeted organs (e.g. albumin or serum amyloid A in liver-expressed mRNA; Myoglobin or Myosin for the muscle; CD45 or CD18 for leukocytes; and CD36, GLUT4 or adiponectin for adipose tissue) [32]. Optimisation of 5'UTR sequences can be facilitated by computational approaches since it allows explore high quantities of sequences in a quick manner[33]. Predictive model have been applied to tune sequences for optimal translation[34]. Using a *de novo* design approach, a new 5'UTR sequence containing Kozak sequence and minimal secondary structures, proved to be more efficient than previously reported sequences by two-fold [35]. Deep-learning approaches can predict optimal UTR sequences from given amino acid sequences of a target protein[36].

The 3' UTR sequence plays an important role in mRNA localisation, translation and stability. This region is rich in cis-regulatory elements, namely adenylate uridylylate (AU-rich) elements (AREs) that are responsible for the transient response of the mRNA[37]. Different trans-acting factors include regulatory proteins and microRNAs (miRNA), and their action that can be modelled in response to external stimuli. 3'UTR  $\beta$  and  $\alpha$  globin have been successfully explored

to increase the stability and translation efficiency of mRNA[38,39] with mRNA half-life increased when two tandemly reiterated  $\beta$ -globin 3' UTRs are used [40]. However, the combination of naturally occurring 3'UTRs with AES-mtRNR1 and mtRNR1-AES proved to improve mRNA translation efficiency by increasing stability when compared with tandem  $\beta$ -globin[41,42]. The miRNA binding sites can also be used to modulate translation response[43]. As in the case of the 5'UTR, the effect of 3'UTR sequence modification is cell-type dependent. Tandem repeats of  $\beta$ -globin 3'-UTRs produce less protein in human pluripotent stem cells[44] or have a less pronounced effect on immature DC cells[45]. The replacement of both 3' and 5' UTR by UTRs from a highly stable cellular mRNA like the human cytochrome b-245 alpha chain (CYBA), improved translation in NIH3T3 and A549 cells[46]. Furthermore, it has been shown that secondary structures that include the internal ribosome entry site (IRES) can be used in both 5' and 3' UTR and a high level of translation can be obtained even in the absence of the 5'cap and the poly-A tail[47].

Table 1. Resue of 5'UTR and 3'UTR used in mRNA vaccines templates, as well as optimisation strategies that can be followed to maximise cellular response. \*Sequence used by Moderna. \*\* Sequence used by BioNTech

### 5'UTR

Sequences used:	
$\beta$ -globin	
$\alpha$ -globin	
Complement Factor 3	In combination with
Cytochrome p4502E1	Kosak sequence
eIF4G Binding Site	
Alternative optimisation strategies	
Target cells	
Minimal 5' synthetic UTRs	
Expressed genes from the targeted organs	
Computational approaches:	
Minimise GC	
Predict optimal UTR sequences	

### 3'UTR

Sequences used:	
$\beta$ -globin or 2 x $\beta$ -globin	
$\alpha$ -globin	
Haemoglobin subunit alpha 1 gene (HBA1)*	
Amino-terminal enhancer of split (AES)**	
AES-mtRNR1 or mtRNR1-AES	
internal ribosome entry site (IRES)	
Alternative optimisation strategies	
Target cells	
miRNA binding sites	

## Poly-A tail

A Poly-A tail is a sequential repetition of adenosine nucleosides at the 3' end of the mRNA. It plays an important role in the modulation of translation as it binds to PABP proteins that will interact with the eukaryotic translation initiation factors (eIFs) to form a closed-loop structure[48]. Initial reports showed that mRNA with longer poly-A tails (100-120 bp) had a better performance than mRNA with shorter tails[45,49]. Nevertheless, genes that are highly expressed present short poly-A tails[50], while long tails are associated with poorly translated genes[51]. Sequence optimisation has been explored with the use of non-adenosine nucleotides. Cytosine-containing sequences in the 3' region can protect the mRNA against CNOT complex-mediated mRNA degradation[52]. A histone-stem-loop sequence can be added in combination with a poly-A tail to stabilise the mRNA and increase translation [53,54]. Additionally, a 30 nucleotide poly-C tail can be used at the 3' terminal[30,55].

The insertion of the poly-A tail can be performed either enzymatically or incorporating the sequence into the DNA template. The latter is preferred since the enzymatic addition does not deliver a precise mRNA sequence[56]. Recombination can occur during the amplification of the plasmid DNA when plasmid encodes for homopolymeric stretches, however this effect can be minimised by the addition of a sequence spacer [57,58].

## 5' cap

The 5' cap corresponds to a guanosine with a methylation in the N7 amine that is linked to the first nucleotide of the mRNA by a reverse 5' to 5' triphosphate[59]. Translation efficiency is highly dependent on the cap since recruitment of the eukaryotic initiation factor 4E (eIF4E) which will subsequently bind to the eIF4F, circularising the mRNA by binding to PABP protein[21]. The cap can present different conformations: Cap-0 (cap with one methylation), Cap-1 (2'-OH methylation in the first nucleoside of the mRNA sequence), and Cap-2 (2'-OH methylation of the first two nucleosides). Cap-0 is recognised by the pattern recognition receptor (PRR) inside the cells as an exogenous RNA, triggering a strong innate immunity response by the cells[12]. The Cap-1 structure has only been described in eukaryotic mRNAs, and its recognition by the PRR is lower, which is the immunogenicity response [60]. This is the most used structure to cap mRNA vaccines[20].

## Sequence optimisation

mRNA structure can influence mRNA translation, stability and immunogenicity. A possible strategy is to optimise the coding sequence (CDS) by maximising GC content. GC-rich CDS presents a more efficient translation as its control is associated with mRNA decay pathways while and AU-rich CDS is controlled by translation regulators and miRNAs[61]. In fact, mRNA

with increased GC content showed to express more protein by 100 times[62]. Additionally, a lower AU content decreases the innate immune response. The presence of uridine in mRNA leads to the expression of pro-inflammatory cytokines that can block antigen expression[63,64]. Uridine-rich RNA, such as viral genomic RNA, are recognised by Toll-Like receptors (TLR) 7 and 8 which lead to the activation of the pro-inflammatory pathway[65]. Maximising GC content will lead to an increased expression *in vitro*, and therapeutically relevant protein levels are achieved in large animals[66].

An alternative perspective on mRNA CDS optimisation is to use the secondary structures as a central point in the design approach. mRNA is highly susceptible to be chemically degraded by hydrolysis[67]. However, the presence of secondary structures in the molecule can significantly reduce this effect[68]. This effect is position dependent as the presence of secondary structures in the first 30 nucleotides of the 5'UTR region do not seem to favour expression, nevertheless, highly structured regions downstream of these 30 nucleotides positively impacted the protein production[69]. A combinatorial approach that assessed translation efficiency and stability in both cells and solution showed that highly structured mRNA can improve stability and expression, and that in cell stability plays a more important role for protein expression[70].

Computational models based on thermodynamic (minimising the free- energy[71]) are helpful to visualise and optimise the mRNA secondary structures. Artificial intelligence and machine learning models are now gaining popularity due to their ability to predict based on multiple parameters trained models and combined with wet lab data[72]. Model predicted structures hydrolysis resistance obtained a 2-fold increase in mRNA half-life[68]. These types of models also allow to combine different optimisation strategies. An algorithm that combines the optimisation of structural stability and codon usage improved antibody titres up to 128 times *in vivo* [73]. Computational optimisation design for mRNA vaccines is extensively reviewed elsewhere[74].

## Modified nucleosides

A groundbreaking discovery that enabled mRNA technology to be used in clinical applications was the incorporation of modified nucleotides. Karikó et al[75] showed that the presence of modified nucleosides suppressed the immune-stimulatory effect of the mRNA. Natural occurring nucleosides are recognised by the PRRs that induce activation of the type I interferon (IFN-I) pathway and lead to a strong immune response[12]. The presence of modified nucleosides, such as pseudouridine ( $\Psi$ ), 5-methylcytidine (m5C), N6-methyladenosine (m6A), 5-methyluridine (m5U), or 2-thiouridine (s2U) suppressed the activation of the TLR3, TLR7, and TLR8[75], and consequently, the pro-inflammatory pathway. The use of pseudouridine proved also to improve protein expression, due to an increase in ribosome recruiting[76], and mRNA

stability[64]. Additionally, the production of double strand RNA (ds-RNA), an impurity formed during IVT, is prevented[77]. The modified nucleosides were used in the COVID vaccines BNT162b1 and BNT162b2[78], and mRNA-1273[79,80].

## Manufacturing steps, challenges and Quality Control

mRNA vaccines are produced in a cell free enzymatic reaction catalysed by RNA polymerases, and in the presence of other enzymes. During this process, a DNA template is *in vitro* transcribed (IVT) into mRNA using nucleotides as a co-substrate. The DNA template will contain a promoter region, the gene of interest, and the optimised untranslated regions (see subsection *Plasmid design*). The poly-A tail can be encoded directly in the DNA template sequence or by using a poly-A polymerase (see subsection *Poly-A tail*) [81], with the former being preferential due to better size control[40]. The mRNA cap can be added in a one-pot reaction by using cap analogs, or alternatively by the use of enzymes (see subsection *capping*). The IVT will produce several immunogenic by-products and process related impurities and an efficient removal is mandatory to achieve high-quality vaccines (see subsection *Impurities challenge*). Analytical methodologies that assess product quality and process control must be applied throughout the manufacturing process to ensure a tight control (see subsection *Quality*). In the end, mRNA must be formulated to ensure mRNA is efficiently delivered to the target cell cytosol (see subsection *Formulation*).

### Plasmid design

The DNA template used for the production of mRNA is comprised of a promoter sequence that is recognised specifically by the RNA polymerase, sequentially followed by the 5'UTR region, the gene of interest to be expressed, the 3'UTR region and the poly-A tail. Unlike eukaryotic mRNA produced in the cell nucleus, the gene of interest used in mRNA vaccines should only be composed by coding DNA (cDNA)[82]. A major hurdle in the plasmid production is the occurrence of recombination of encoded homopolymeric stretches (e.g. poly-A) during amplification[83]. This is overcome by the use of vectors that allow to expand the poly-A size using of type II restriction enzymes and ligation[84]; by the using linear plasmids based on coliphage N15[85]; or by segmenting the poly-A tail by the addition of spacers[57], [86].

### pDNA manufacturing

The DNA templates can be obtained either chemically or produced in *E. coli*. Amplification using PCR is commonly used for the production of small scale batches in the µg range[6,87]. For large scale production, milligrams of plasmid DNA (pDNA) are produced in *E. coli* and subsequently linearised using restriction enzymes[88,89] and purified with standardised workflows[90]. This method presents a number of advantages over the PCR since it reduces

the introduction of aberrant species, it is easily scalable, and can deliver high purity templates[91]. Typically, high number of copies of plasmid are obtained ( $1000 \text{ cell}^{-1}$  [92]) and final titers of  $2 \text{ g pDNA L}^{-1}$  [93]. Cell-free technologies based on rolling circle amplification are being tested for *in vitro* plasmid production at large scale[94,95].

## Template Manufacturing

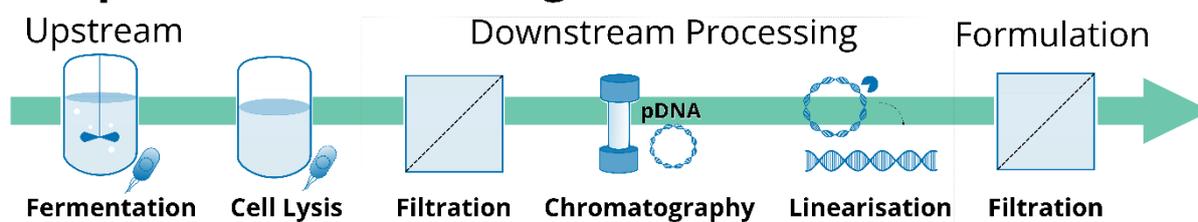


Figure 3. Representation of a *E.Coli*-based DNA template manufacturing process

The pDNA is produced either in fermentation batch mode (low cell density cultivation) or in fed-batch mode (high cell-density cultivation, dry cell mass of  $\sim 30\text{--}60 \text{ g L}^{-1}$ )[90]. Afterwards, pDNA is recovered by a cycle of alkaline lysis and neutralisation [96] with a significant removal of protein complexes and cell debris. Pre-purification steps can be performed by the use of precipitation steps (e.g. using alcohol, polyethylene glycol, calcium chloride and/or ammonium sulphate[97]) or membrane separation (e.g. using tangential flow filtration [98]). To increase levels of purity pre-chromatography, RNase can be used during these steps to lower RNA concentration. Chromatographic modalities are chosen according to the pDNA characteristics, and can range from anion exchange[99], hydrophobic interaction[100,101], affinity[102], size exclusion[103] to newer technology as multimodal chromatography[104–106]. Nonetheless, due to the size of pDNA, in the nanometre scale, and its impact on binding capacity and in separation using traditional modalities, superporous matrices are favoured (e.g. monolith or membranes)[107,108].

A critical step in the template production is the linearisation. This step, although it is not mandatory for IVT itself, it allows to produce mRNA molecules with the exact same length as the template due to the nicking of the sequence downstream of the poly-A tail. A sequence recognised by a restriction enzyme must be placed adjacent to the poly-A tail to increase uniformisation of the mRNA template, as termination sequences are not recognised by the RNA polymerase *in vitro*[82]. Importantly, restriction enzymes are chosen based on the type (e.g. IIS enzymes are preferable because they recognize asymmetric sequences and cleave at a defined distance), capability to produce 5' overhangs or blunt ends, and GMP grade availability (if required). This process usually performed in batch mode at  $37^\circ\text{C}$  from 1h up to an overnight.

Reaction conditions and parameters are enzyme dependent with the reaction can be stopped by heat or addition of EDTA. To purify the linearised DNA tangential flow filtration can be performed (e.g. 100 kDa PES membranes with a load of 0.6mg cm<sup>2</sup> [109]) or use chromatography or silica-based filters [91]. Nonetheless, to obtain linearized DNA at scale is cost and time intensive so alternative methods are being sought. Immobilisation of the restriction enzyme can be used to overcome these bottlenecks by allowing easy separation of enzymes from reaction media. Enzyme optimisation by mutation have been performed to increase activity and reusability of the enzyme[110] but this can present an hurdle for GMP implementation. Alternatives have been tested to avoid linearisation itself, such as the use of supercoiled pDNA. In this particular case, a transcription terminator is added to the sequence downstream of the poly-A tail. This form of pDNA forms stable complexes during transcription initiation[111], however their function in an *in vitro* system is limited since longer transcripts are produced[82] and will produce lower yields than linear pDNA. Nevertheless, a tangential flow filtration step is usually performed to formulate pDNA to the desired buffer.

Since the linear DNA used for the process is considered a starting material, it is required to follow guidelines that allow the materials to be suitable for GMP production. When preparing linear DNA using plasmid DNA it is required to establish cell banks[112]. Additionally, information regarding the origin of the DNA sequence and the plasmid map, as well as generation of the host cell line and transformation and purification of the host cell line, should be provided[113]. pDNA identity can be assessed using sequencing or restriction digestion[113], being the last used as a qualitative guide[114]. The linearisation percentage should also be assessed. pDNA concentration can be measured by spectrophotometric assay. Product and process-related impurities, must be characterised[115]. Residual host cell DNA and proteins, as well as the presence of endotoxins, must be assessed. Analytical methodology include quantitative PCR or HPLC; bichinchonic acid assays or ELISAs; and LAL-based assays, respectively. Other tests include bioburden, presence of mycoplasma and antibiotics, pH, and osmolality[114].

## In vitro transcription

mRNA is produced in an *in vitro* transcription reaction catalysed by RNA polymerases, and using nucleotides and the template DNA as substrates. This process presents an advantage over traditional biological manufacturing as it is cell-free based, free from cell derived impurities and adventitious contaminations[81]. The IVT is a highly controlled reaction, flexible as it can be easily scalable and adaptable to different mRNA products without loss of yield. In a couple of hours, this process can yield g<sub>mRNA</sub> L<sup>-1</sup>, reaching 12 g<sub>mRNA</sub> L<sup>-1</sup> [116,117].

A number of RNA polymerases have been reported to be used in the IVT reaction, namely T3, SP6 or T7 RNA Polymerase. T7 RNA polymerase (RNAP) is the most popular among them, due to single-subunit structure, highly specificity towards the T7 promoter, efficient production of long transcripts, and independence of protein co-factors[118,119]. The T7 RNAP kinetics can be divided into three different stages, and are preponderant in the formation of the mRNA and by-products. Mechanistically, the RNAP firstly binds to the promoter sequence, melts the two strands of DNA and starts the transcription. During this phase, RNAP is still bonded to the promoter which results in a RNA-DNA hybrid that is extended until the 8-9 bp[120,121]. Abortive RNA fragments are released during this phase. Elongation itself starts when the RNAP releases the promoter and forms a stable elongation complex[122]. When the enzyme reaches the end of the linearised template, the termination occurs This step can lead to the addition of sequences in the 3' end of the mRNA (see subsection *Impurities challenge*)[123].

Regarding other reaction components present in the IVT, and an important salt is magnesium. This salt is a cofactor for the T7 RNAP and its presence is required for the enzyme to bind to the DNA template[124]. Furthermore, magnesium forms a bond with the NTPs allowing the formation of a phosphodiester bond with the RNA, and release a pyrophosphate[125]. Magnesium can be added in the reaction as  $MgCl_2$  or as  $Mg(CH_3COO)_2$ , nevertheless chlorine can have an inhibitor effect on the reaction[126]. Reported optimal concentrations of  $Mg(CH_3COO)_2$  are between 40 and 60 mM[116,125], while  $MgCl_2$  concentrations described are between 20 and 30 mM[117,127].

The pyrophosphate released during IVT can have multiple inhibitory effects and impact final yields. It can cross-link with the free  $Mg^{2+}$  and precipitate, decreasing ion concentration in the solution[126,128] and potentially capturing the DNA template[129]. Additionally, pyrophosphate can compete for the binding site of the T7 RNAP in both free enzyme and the stable elongation form[130]. To avoid this, inorganic pyrophosphatase (Ppase) can be added to the IVT reaction, catalysing the hydrolysis of pyrophosphate to form orthophosphate. To this date, no consensus on the use of Ppase in the IVT has been reached. Nevertheless, high yield reported processes use pyrophosphatase in concentration ranging 1 to 8 U mL<sup>-1</sup> [91,116,117,127].

The concentration of NTPs is important to obtain a minimum reaction rate and typically ranged between 4 and 8 mM[91,116,117,127], with concentrations above 7 mM having a positive impact in the reaction[116,131]. Whilst modified nucleosides (see subsection *Modified nucleosides*) can be added in the same concentration range without compromising the reaction efficiency, the occurrence of errors and the sequence fidelity is highly dependent on the specific modified nucleoside used[76].

Tris or Hepes buffer are added to the reaction in order to buffer hydrogen ions produced during an IVT reaction [132]. Typically, reactions are performed at a pH close to 8 [91,133] but expected to decrease as reaction progresses. Continuous monitoring of the pH during IVT showed that the pH decreases to 6.4 in the first hour of the reaction[117]. Lower pH values (6.5-7.5) are reported to improve transcription[116,126]. Reaction temperature can have an impact on the promoter binding and production of by-products [134,135]. Temperatures ranging from 37°C to 43°C are described[116]. Using thermostable T7RNAP, the temperature can be increased up to 50°C[135]. Template concentration can range between 30 to 90 nM, and T7 RNAP between 6000 to 8000 U x mL<sup>-1</sup>, and [116,117,136]. Rnase inhibitors can be added when it is not possible to guarantee a free RNase environment.

Enhancers can be added to the reaction to improve mRNA production. Since RNAP oxidises easily, dithiothreitol (DTT) can be added to the reaction to avoid protein oxidation[137]. Spermidine presence in the IVT reaction can increase production up to 10 times[138]. This can be explained due its ability to condensate DNA by neutralising its negative charges[139]. Concentrations between 1 and 3 mM are reported to deliver optimal results[116,140,141]. Less explored enhancers include DMSO and chaotropic agents. DMSO in low concentrations can condensate T7 RNAP structure and increase its activity[119]. Urea and formamide were evaluated due their ability to act as a chaotropic agent at mild concentrations to lower the production of by-products during IVT[142].

## Capping

To fully manufacture a mRNA that is translation efficient, it is required to modify the guanosine triphosphate (GTP), that is usually the first nucleotide, into a Cap-1. Cap analogs have been synthesised to be used in the IVT reaction. This is a one-step IVT reaction, where the incorporation of the first GTP is substituted by the cap analog. This can be achieved by introducing a cap analog in IVT reaction in a concentration several folds higher than the GTP (usually in a 4:1 ratio)[143]. The traditional cap analog, m<sup>7</sup>GpppG, produces two isomeric forms of the mRNA, can be incorporated in both forward and reverse position, however, only the forward position is efficiently translated[144]. To overcome this, anti-reverse cap analogs (ARCA) that are exclusively incorporated in the forward orientation, can be used[145]. The major drawbacks of this one-step IVT reaction is the yield of capped mRNA obtained, which is usually 80%[143,146], and it only delivers a Cap-0 structure. CleanCap AG, a co-transcriptional reagent, can be used to deliver Cap-1 structure in one-pot IVT reaction, with a yield of 94%[133]. This was achieved by extend the cap analog from a dimer to a trimer, which allows for the methylation of the guanosine in the structure[147]. Nevertheless, to use this system it is required to have a 5'-AG at the downstream of the T7 promoter sequence[133]. A propargyl group added to the trinucleotide analog can enable further modifications of the mRNA[148].

Cap-0 can also be achieved by a second enzymatic reaction. RNA capping enzymes, such as Vaccinia capping enzyme (VCE) or Faustovirus capping enzyme (FCE), are able to produce Cap-0 by having three different activities: 1) RNA triphosphatase (TPase), which removes the phosphate to form a 5'diphosphate in the nascent of the mRNA; 2) an RNA Guanylyltransferase (GTase), that transfers a GMP from a GTP to the 5' end of the mRNA; and 3) an RNA guanine-N7-methyltransferase (MTase), which methylates the N7 position of the guanine base using S-adenosyl methionine (SAM)[149]. The enzymes require the addition of MgCl<sub>2</sub> as a co-factor, for the activities of the TPase and GTase; GTP as a GMP donor; and SAM. Typically, the reactions are performed at 37 °C for 2 hours. The main difference between FCE and VCE relies on the higher specific activity, higher activity towards mRNA containing secondary structure, and a broader temperature range of the first[150]. This process achieves 100% capping yield[6]. Cap-1 can be produced by the cap 2-O-methyltransferase enzyme and using SAM as the methyl donor. Both of these reactions can be performed in one-pot format without compromising the process yield [136,150].

## Impurities challenge

After the production of the mRNA by IVT, the process delivers not only include the product itself, but also a number of impurities that can be classified as process and product-related impurities. Process-related impurities include all the reagents, enzymes, NTPs and the template added in the IVT and/or capping reaction. Product-related impurities correspond to products that are misformed during the IVT, such as DNA-RNA fragments, abortive transcripts and double-stranded RNA (ds-RNA). These impurities are usually removed using a combination of purification steps that can include digestion, precipitation, filtration or chromatography, or a combination of the above[6].

To design a cost-effective downstream platform, it is mandatory to well characterise the physico-chemical characteristics of the product as well as of main impurities. This can include size, charge, hydrophobicity, composition, structure, and stability[151]. mRNA is a large molecule, with a size ranging 1-15 kbp[48]. mRNA sugar-phosphate backbone provides negatively charge to the molecule, and purine and pyrimidine bases provide the hydrophobicity. The mRNA secondary structure also influences the molecule availability of certain regions to be used in the purification process. mRNA is also very stable *in vitro*, which allows it to be precipitated, frozen, resuspended and heated up to 90°C without damage[152].

It is relatively easy to separate the mRNA from the smaller impurities, such as NTPs and enzymes. Purification methods at lab scale are usually based on lithium chloride precipitation[81,153]. The separation is achieved due to the interaction of the lithium cations with the negative backbone of the mRNA, which leads to precipitation. This process does not

precipitate proteins, it is less efficient in precipitating DNA and small size mRNA (<100 bp). Tangential flow filtration (TFF) can also be used to separate mRNA from smaller size impurities[109,154]. Size exclusion chromatography (SEC) techniques can also be used to separate mRNA from smaller size impurities[155,156]. One major drawback of these techniques the separation from DNA template or truncated and double stranded fragments cannot be achieved.

The DNA template is a challenging impurity owing to its physico-chemical similarities with the mRNA. Within the cells, DNA, even at residual quantities, leads to the IFN pathway to be activated, and, consequently to a strong innate immune response. Additionally, it can be integrated into the host genome. According to the information available, a limit of 330 ng of DNA per mg of mRNA is acceptable[157]. The main structural difference between DNA and the mRNA is that the DNA double helix. Due to its double stranded nature, DNA is less hydrophobic than RNA[158]. Removal of the DNA can be achieved enzymatically, by the use of Dnase I in free of immobilised form[159,160]. This enzyme hydrolyses DNA nonspecifically into small fragments. To efficiently remove the fragment and the enzyme added, it is required to couple at least a second purification step.

Chromatography approaches that use the difference in characteristics of between mRNA and DNA can be also applied. Hydroxyapatite chromatography can be used to separate DNA from RNA[95], as the DNA binds with more strength to the resin, and with a RNA recovery yield of 80%[161]. Core bead chromatography, a multimodal technology that combines the separation according to size with anion exchange interactions, can be used in combination with Dnase I digestion[95,127]. Anion exchange was also explored for the separation of mRNA from DNA templates. Due to the differences in charge related with the backbone, DNA will present a higher affinity towards the anionic ligands, while RNA will be eluted first[155]. This method allows purifying mRNA with sizes up to 10 kbp[162]. To achieve separation it is necessary to block intramolecular hydrogen bonds using denaturing conditions, whether with temperature, or chaotropic agents[151]. Small RNA abortive mRNA fragments can also be separated using this technique.

IVT produces by-products that include truncated mRNA and ds-RNA. During the binding stage of the RNAP, short transcripts are released in an event named abortive cycling[163]. Additionally, IVT reactions often produce mRNA that are shorter (n-i) or longer (n+i) than the template used[164]. T7 RNAP also presents RNA-dependent and template independent transcription capabilities[165]. T7 RNAP is able to use RNA using a cis mechanism - by folding back on itself forming intramolecular duplexes - or a trans mechanism, where it binds to the equivalent region to form an intermolecular duplex[166,167]. The short abortive mRNA fragments and the full-length can also produce ds-RNA by RNA-templated transcription[123].

The presence of ds-RNA in cells will induce the IFN-I pathway, as well as TNF- $\alpha$  and IFN- $\gamma$  which are linked to a strong immune response[12]. This process can cause translation inhibition, and it can be responsible for the induction of uncontrolled immune-inflammatory reactions, such as myocarditis [168].

Removal of the ds-RNA can increase the protein production in cells by 10-1000 folds[169]. To achieve this, a number of strategies can be applied, from engineering T7 RNAP enzyme to lower the ds production[123,135,170], the addition of DNA oligonucleotides complementary to the 3' end during IVT[171], to the use of different purification strategies. By exploring the structural differences between ss-RNA and ds-RNA, ion pair reverse-phase chromatography (IPC) was successfully implemented in separation of single stranded RNA (ss-RNA)[169,172]. An ion-pair, usually a quaternary ammonium compound, is added to the mobile phase, and in contact with the RNA, it will bind to the negatively charged backbone. Since the separation is dependent on backbone of each molecule, it is possible to separate dsRNA from ssRNA. However, this technique presents major challenges, such as resin capacity, scalability, and the use of toxic reagents (e.g. Acetonitrile). Cellulose based separation can be used to remove dsRNA from IVT samples. This process is based on hydrophilic interaction between the cellulose and the 2'-hydroxyl residues that are present in higher concentration in ds-mRNA. 90% of ds-RNA can be removed with this method[173]. Nevertheless, it requires to be coupled with a pre-purification step to remove other impurities. Non chromatographic methods can include the use of nucleases, namely RNase III, which recognises and cleaves ds-RNA[174].

A popular approach for purification of mRNA is the use of affinity chromatography. The separation is achieved due to the functional group, a sequence of deoxythymidine (dT), as affinity will capture through base-pair hydrogen bonding the poly-A tail present in the mRNA, while the impurities, such as the NTPs and enzymes, are washed out[175]. This type of purification can be found in different matrices, from traditional resins[159], monolith[176], to magnetic macroparticles[177]. This was the method of choice of Moderna to purify Covid-19 mRNA vaccine[151]. However, the low binding capacity and costs are major drawbacks when using this separation technology. Additionally, separation from the ds-RNA may not be achieved[178]. Due to the difference in hydrophobicity between ssRNA, ds-RNA and DNA, hydrophobic interaction chromatography (HIC) can be used for mRNA purification. Nevertheless, owing to the ability of the other impurities, such as proteins and aggregates, to bind as to the matrix[151], this method may only be used as a polishing step[179].

New chromatographic approaches can be explored for the purification of mRNA. Hydrogen bonding chromatography was shown to be able to separate ss-RNA using pH or phosphate/pyrophosphate gradient[179]. Weak anion exchange and hydrogen bonding multimodal [179,180] and a weak anion exchange multimodal chromatography[181] were also

successfully applied to separate mRNA from pDNA. Multimodal technologies may allow to intensify purification without the need of multiple chromatographic steps[6].

## Quality

In order to meet the consistency and safety required throughout the commercial batches, it is required to strictly follow manufacturing practices and product specifications. A quality by design (QbD) approach can accelerate the development as it is centred in the integration of quality considerations within the process[182]. The knowledge required should be acquired by the establishment of a Quality Target Product profile (QTPP), based on the patient's needs, and by a list of critical quality attributes (CQAs) associated[157]. Critical process parameters (CPPs), which are variables that can have an impact on CQAs, need to be set for each step of the manufacturing process. To create the knowledge required, it is imperative to fully characterise both products and impurities at the different stages of the process.

For the final product, a rigorous characterisation should comprise chemical, physical and biological methodologies[112]. mRNA identity is a critical point in the process, and it can be obtained by sequencing (RT-PCR, NGS or sanger sequencing)[183]. mRNA quality and integrity can be assessed from multiple angles and considering all the parts that constitute a final mRNA vaccine. Firstly, consistent expression of a complete protein should be evaluated[112]. Capping efficiency, length of poly-A tail and mRNA structure can be evaluated using analytical LC-UV/MS and LC-MS; RP-HPLC; and CD or DSC, respectively[157]. The presence of process impurities, namely enzymes, can be assessed by bicinchoninic acid assays or HPLC, while DNA can be quantified by qPCR. The presence of truncated and ds-RNA can be evaluated by electrophoresis, RP-HPLC or ELISAs. mRNA aggregates can be assessed by size exclusion chromatography[184]. LNP quality evaluations include particle size distribution and polydispersity, encapsulation efficiency and lipid composition. Other quality aspects include endotoxins quantification, product sterility and stability[183].

During manufacturing, a tight control is a key to have a successful process. Acquiring quality process data and at real time, not only provides important information regarding the relationship between process parameters (PPs) and QCAs, but can allow for a process control using process analytical technology (PAT) framework, and ultimately, a real-time release testing (RTRT)[185]. New analytical methods have been emerging to better characterise mRNA quantity and quality, as well its impurities, during the manufacturing process. At-line control of mRNA production during the IVT allowed to increase the process yield[186]. The method uses a multimodal monolith to control mRNA production and NTPs consumption[176,186]. Nevertheless, this method lacks the ability to assess the NTP individual consumption. Reverse-phase chromatography[187] and anion exchange chromatography can be used to quantify

mRNA[188–190]. On-line monitoring can be achieved using molecular beacon assays[135,191], or by using fluorescence resonance energy transfer (FRET) assay[192].

Capping efficiency can be determined using chromatography-based methods. Phenyl boronate chromatography can be used in conjunction with Rnase T1 and T2 to separate the cap structures from the 3'-nucleoside monophosphates[193]. RNA 5' polyphosphatase can be used to remove the phosphate from the uncapped 5' end, and is subsequently digested by terminator 5' phosphate dependent exonuclease, which completely digests the uncapped mRNA[194]. IMac PrimaS™ can be coupled with this digestion strategy to quantify the undigested mRNA[195]. This method does not allow separation between cap 0 or cap 1. LC-MS can be used to yield cap efficiency. Streptavidin-coated magnetic beads can purify 5' end fragments obtained using probes in combination with RNase H, to be then analysed by LC-MS[196]. This method was further refined, by using fluorescently RNase H cleaved fragments and optimised purification step[197]. Ribozymes, catalytically active RNA molecules, can be designed to cleave near the 5'end of mRNA and replace the use of probes and RNase I[198]. Anion exchange chromatography can be used instead of LC-MS[199]. Nevertheless, these methods require intermediate steps that can influence the analytic outcome. anti-m7G antibody can be used in an ELISA to detect capping [199].

The characterisation of dsRNA presence during the manufacturing is also critical. dsRNA is commonly detected by immunoassay techniques, namely dot blot and ELISA, using a anti-double-stranded RNA antibody. These methods are labour intensive and do not allow for an at-line detection. Lateral flow strip assays can be used to expedite the process[200]. RNase T1 can be used to digest ssRNA and it can be coupled with mRNA quantification methods[116,123].

As aforementioned, T7 RNAP can produce shorter or longer fragments. Since poly-A tail is a QCA it is mandatory to characterise poly-A size and heterogeneity during the manufacturing process. Poly-A analysis can be performed using similar strategies applied to the capping efficiency, using probes and Rnase I, or by using RNase T1[201], followed by chromatography or LC-MS[189]. Capillary gel electrophoresis is an alternative to the LC-MS with the same resolution[202].

Capillary gel electrophoresis can be used to assess quality and quantity of the mRNA during the manufacturing pt, as it allows to separate truncated mRNA fragments, quantify the full mRNA, and analyse the poly-A size, when coupled with enzymatic digestion[203]. Microfluidics chip electrophoresis techniques can be used to assess mRNA capping efficacy[204], and for dsRNA quantification[205]. These techniques have several advantages over traditional analytical methods as they offer precise control, are usually faster, and allow for high throughput.

## Formulation

mRNA vaccines are large and negatively charged biomolecules. Additionally, despite being stable *in vitro*[152], mRNA is susceptible to degradation by RNases, nucleases, helicases, polymerases and chaperones[67]. mRNA needs to reach the cytosol, and although delivery of naked mRNA is successful[206–208], the effectiveness is reduced[209]. Formulation is essential to deliver the mRNA to the cell's cytosol while maintaining its integrity, and to lower undesirable immunologic responses[210]. A number of delivery strategies can be applied to mRNA vaccines, from naked and self-adjuvant form[211,212], *ex vivo* using dendritic cells (DC)[213], or by using carrier-based systems.

Carrier-based systems are the most popular delivery methods for mRNA vaccines[6]. This category includes polymers, peptides and lipid-based systems. Polymers can form spheres, and, when directly mixed with the mRNA, and since they are positively charged, they interact with the negatively charged mRNA and form polyplexes[214]. Polymers used include Polyethylenimine (PEI), polylysine (PLL), and polyamidoamine (PAMAM)[20]. PEI has been successfully used to deliver mRNA[215–217]. Although this system is highly stable and allows long-term storage[218], they present high polydispersity and cytotoxicity, and challenging biodegradation[219].

Cationic peptides, such as protamine, can be used to deliver mRNA, and proved not only to deliver protection against RNases, but also to modulate immunologic profile[209,220,221]. The less explored cell-penetrating peptides, that owing to its low charges and ability to target specific cells, can be used by themselves[222], or in combination with lipid or polymers[223], to modulate the delivery response[20].

Lipid-based systems, namely LNPs, are the most common delivery system for mRNA vaccines, and the one used for the both approved vaccines[224]. LNP are composed of four different lipids, each one having a different function (Figure 4). The ionisable cationic phospholipid interacts with the negatively charged mRNA, forming the core. The outer layer is composed of neutral auxiliary phospholipids that mimic the cell membrane, glycol modified phospholipids that improve hydrophobic and increase the endocytosis[225], and cholesterol to promote mRNA intake[210]. Ionisable cationic lipids are composed of a tertiary amine that is positively charged at acidic pH, and neutral at physiological pH. The positive head will interact with the negative backbone of the mRNA in an acidic environment, then, the pH is increased to neutral (pH 7.4), and the hydrophobic interactions between the different lipids surpass the electrostatic interactions, and the outer layer is organised[226]. There is no consensus for the structural organisation of the LNP-mRNA, but recent studies suggest that the mRNA is densely packed,

and the ionisable lipids have a structural role, interacting by hydrogen bonds[227,228]. A single monolayer of lipids is formed on the outside[224].

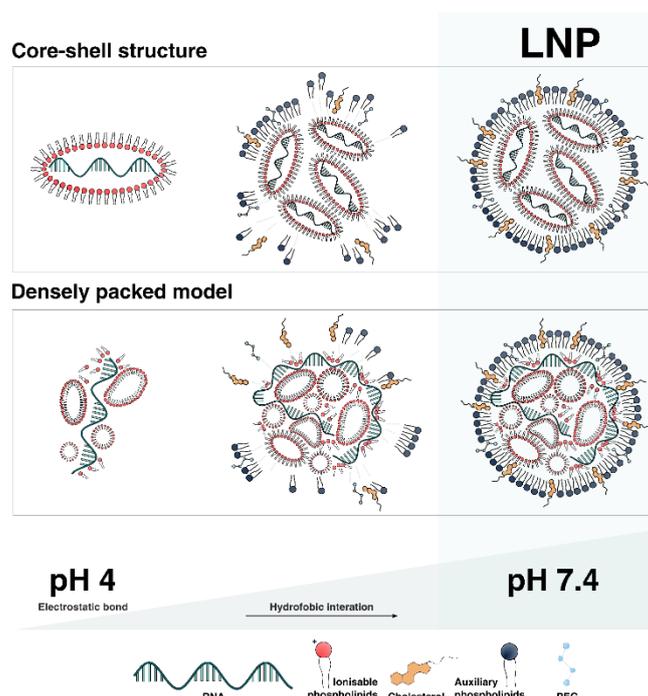


Figure 4. Structural elements for LNP-mRNA. Core-shell lipid structure production, and densely packed model structure production

LNPs manufacturing is divided in two steps: 1) mixing of the lipids formulation with the mRNA in a low pH environment, followed by 2) an increase of the pH environment to a neutral pH (7.4). A tight control is required during the formulation manufacturing steps, as it highly impacts the vesicles and encapsulation efficiency[229]. The most common method of mixing is a continuous injection of ethanol dissolved lipid solution into a water-based solution, where the mRNA is dissolved, to condensate and form micelles[230]. The mixing rate and percentage have a strong impact on the size owing to its influence on lipids polarity[231]. Most common mixing methods are microfluidic based[232]. Buffer adjustment can be achieved by adding a stock solution, or by dialysis or tangential flow filtration. These methods can then be coupled with evaporation or lyophilisation techniques[233]. This process usually yields 70-90%.[234] Removal of empty LNP particles, or separation of LNPs with different loads can be achieved using size exclusion chromatography[235].

CQAs during formulation include the particle's size distribution and polydispersity, encapsulation efficiency and lipid composition. Particle size can be assessed using dynamic light scattering (DLS), tuneable resistive pulse sensing methods (TRPS), and nanosight tracking analysis (NTA). The first two methods also can also be used to measure the particle's zeta potential. HPLC and LC-MS can be used to evaluate lipids composition[157]. Capping efficiency can be calculated by measuring the mRNA concentration in the supernatant. Another important QCA is

the mRNA quality during the formulation steps. Not only the mRNA identity must be assessed, but also the secondary structure. Circular dichroism (CD)[236] and differential scanning calorimetry (DSC)[157] can be employed for this task. LNP-mRNA stability studies evaluate the physico-chemical characteristics, such as mRNA identity and structure and quality, as well as biological performance during the shelf time of the final product[237].

## Current and future manufacturing demands

Pharmaceutical manufacturing is based in a mass manufacturing paradigm, where the product is considered to be one-size-fits-all[238]. In terms of manufacturing, this centralised and fixed model delivers high-quality and highly reproducible products in a mass quantity and lower production costs[239]. mRNA vaccines present a unique combination of features, such as precision and flexibility, and the recent advances improving their stability, delivery, and immune response, makes this technology an ideal candidate to be used in multiple conditions, from prophylactic treatments for both existing and new variants; cancer treatments; metabolic and genetic diseases; or protein replacement treatments[7,48,240]. Owing to the lack of flexibility, efficiency, and speed of the traditional mass manufacturing approaches, decentralised and patient-centric frameworks, where the manufacturing process is customisable to support the local and patient needs[239], are gaining interest[157], in particular in the personalised medicine field. mRNA technology is a perfect fit to this pivotal approach, as its manufacturing platform is flexible enough that can be easily adapted, and different strategies can be followed according to the patient's needs.

When planning the manufacturing process for mRNA vaccines, one should consider a patient-centric approach. Manufacturing requirements to deliver a prophylactic vaccine in hundreds of millions of doses in a short amount of time, as it was required during Covid-19 pandemic event, is not the same as delivering a personalised cancer vaccine. The manufacturing adaptability of mRNA vaccines starts with the design of the DNA template, where the non-coding regions, such 5'UTR, can be modified to improve the vaccine efficiency, and can be applied across the DNA template and mRNA production, capping, purification steps, and formulation.

## DNA template production

The first step of the design of the mRNA. Multiple strategies can be applied to enhance mRNA stability and efficiency, and modulate immune response, from the non-coding regions chosen, to the use of sequence optimisation approaches based on secondary structures. Nevertheless, sequence chosen can also impact mRNA production, and in particular, the production of by-products. T7 promoter sequences[241,242] and AT rich areas[243] can have a strong impact on the abortive cycling, thus reducing short dsRNA by-products[244]. Inhibition of the production of dsRNA by *cis* self-priming extension can be achieved by using a DNA that is complementary to

12-29 nucleotides at the 3' end of the mRNA produced immobilised in combination with T7 Rnase for a solid phase approach[245,246]. Uridine depletion can also decrease the production of by-products[247].

The designed DNA template can be produced either by *E.coli* fermentation followed by purification and linearisation, or chemically by PCR reaction. PCR can be ideal for small batch productions, as it requires less purification steps and it can be used to amplify directly from cDNA libraries[87]. This method is extremely flexible as, when coupled with molecular biology techniques, allows to modify and produce templates in a time efficient manner. However, the lack of scalability, cost-efficiency, and accuracy during replication makes this method not suitable for larger scale applications.

pDNA production by *E.coli* fermentation is a well-established method to achieve high quantities of pDNA. However, to achieve highly pure linearised pDNA requires multiple steps that decrease the process yield and increase process costs. After fermentation, multiple purification steps are used to achieve pure pDNA. pDNA quality can be assessed by the percentage of isoforms in the product, namely supercoiled DNA. This CQA is a measure of the DNA homogeneity, and has an impact on the quality of the DNA obtained after linearisation. However, this is not a quality attribute considered for the linearised DNA used as a starting material for IVT[114], and its purification can imply the addition of purification steps, making the process more lengthy and costly. Linearisation is a critical step in the preparation of the DNA template, and the percentage of linear in the final product is a CQA of the raw material[183]. Linearisation using restriction enzymes always requires optimisation, and this optimisation is dependent on the enzyme used. Compatibility of the buffer components present in the DNA solution should be always analysed as the enzyme activity can decrease in the presence of salts, detergents, organic solvents or chelating agents. Separation from enzyme, as well as the reaction components, is mandatory to be performed after linearisation. However, undigested pDNA and fragments with incomplete cleavage are also by-products obtained during this reaction. The removal of these impurities is more complex, and can require additional steps.

When assembling the linear pDNA manufacturing process, one should only consider the CQA in the different steps of the process, as well as, the step yield, cost, time of each step. To achieve high quality pDNA, a tray of multiple purification steps can be required, and a balance between the quality of the final product and the process tray must be achieved. Additionally, the quality attributes can differ from product to product and the manufacturing platform may be required to be adapted. Optimisation of the production of linearised pDNA should focus on the construction of a cost-effective and flexible platform and considering that the quality and homogeneity of the linearised DNA used as a raw material for IVT directly impacts not only the mRNA production,

but also the quality of the mRNA obtained. Recent studies highlight the importance of a high quality DNA template in IVT by-products[248].

## mRNA manufacturing

Owing to its simplicity and flexibility, IVT can be adapted into different modalities - batch, fed-batch, depending on the market needs. Traditionally, IVT is performed in batch mode. This well-established mode allows to produce up to 12 gmRNA.L<sup>-1</sup>[116]. However, cost-of-goods (GoGs) have a strong impact during the mRNA manufacture process[249], particularly the IVT reaction, owing to the use of enzymes, which can correspond up to 50% of production costs[117].

IVT reaction was successfully implemented in fed-batch mode[117,250,251]. In all cases, the reaction is fed to maintain the substrate concentration, and it achieved a mRNA production of 9 to 12 gmRNA.L<sup>-1</sup>. The optimisation of these processes also required the implementation of process analytical technologies (PAT) that allow a tight IVT control both on the product production and substrates consumptions. This paved the way for the first applications of continuous mRNA production. A microfluidic device equipped with micromixers was successfully applied for continuous mRNA production[252]. A continuous manufacturing approach brings a number of advantages, such as the reduction of variance in product quality[253], and a constant process operation in optimal conditions[254]. This has an impact on manufacturing costs[6,249], and coupled with reduction in quality assurance costs obtained by real-time-release-testing (RTRT)[253,255], can allow for a significant reduction in the vaccine cost per dose. Nevertheless, there are still gaps in the application of this technology in a continuous manufacturing form.

Multiple perspectives should be considered when choosing the most adequate capping strategy. The post-transcriptional capping requires the use of a second enzymatic reaction to achieve cap-1. Usually, it is required to first purify the mRNA to achieve the best capping yields, which makes the process longer and more complex, and ultimately increases the manufacturing costs. When using a cap analog or the Cleancap, a one-pot IVT and capping reaction can be performed, which makes the process faster. However, enzymatic capping efficiency can reach 100% while cap analogs yield between 60-80%[12,256,256]. One advantage of Cleancap over cap analogs is that it delivers Cap-1 and with a higher capping yield (94%)[133]. Nevertheless the use of Cleancap presents two major drawbacks: 1) it is the most impactful material in mRNA manufacturing costs[249], and 2) the process dependent on a single supplier (TriLink Biotechnologies). Availability of raw materials can be a constraint for larger scale demand[257]. The two mRNA approved vaccines followed different strategies, Moderna used post-transcriptional enzymatic reaction[79], while BioNTech used the co-transcriptional approach using Cleancap[78].

## Downstream processing

mRNA purification can follow a variety of strategies with multiple combinations of different purification steps. When setting up the downstream platform, one should consider not only the CQA, but also the scale required and the product application, as they may have a strong impact on the manufacturing process. Identify the impurities that can critically impact the downstream and select the adequate purification tray have impactful influence on process. For example, residual DNA specification is present for the final active substance[115,258]. The most straightforward approach to remove DNA is the use of DNase. This simple and scalable approach can be coupled with a precipitation, TFF, or a single step of chromatography to remove the remaining impurities. However, enzymatic digestion will not present 100% of yield, and it is one more impurity that is added that requires it to be removed from the final product. Also, unless it can be reused, this raw material can impact the manufacturing costs, specially at higher scales. The use of a chromatography strategy to remove DNA can be more cost-effective for larger scales. Chromatography and TFF devices can be found validated and ready-to-use in a number of sizes for different process scales, which facilitates their use for GMP-grade production. However, the use of multiple steps of chromatography can make the process lengthy and less cost-effective.

The determination and control of the dsRNA concentration in the final product is recommended[183]. However, there is still no specification for the active substance. Both approved vaccines assess dsRNA concentration and compared with the available data from previous batches[115,258]. Ideally, dsRNA production is controlled during IVT by using engineered T7 RNAP, which can make the process dependent on single suppliers, and ultimately increase the costs. Purification processes include the use of an additional enzyme, or the implementation of purification strategies, such as RP-HPLC or cellulose-based separation. These processes require to be coupled with additional purification steps to remove the remaining impurities, which can make the process less cost-effective, especially at larger scale. Developing single-step purification[6], where the dsRNA is also removed, can significantly impact the downstream process cost, enabling a more sustainable manufacturing of mRNA vaccines at larger scale.

## Formulation

One particularly characteristic of the formulation step is that it can be successfully performed in a continuous form from the mixing and diafiltration step[259], to the final lyophilisation step[260]. This process can be also implemented in a microfluidics scale, which presents a number of advantages, such as it allows for a tighter process control through PAT and minimise the raw material costs[6,232]. However, the lack of thermostability of lipid nanoparticles can be also

attributed to the manufacturing process. The diafiltration method does not remove the organic solvent in an efficient manner[261], and the presence of these molecules, even at low concentration (1-2%), can lead to membrane disruption and mRNA leakage[262]. Additionally, lyophilisation is an expensive and time consuming step[261] that is difficult to implement in a flexible manufacturing process. Improving mRNA-LNP stability can also include the optimisation and implementation of alternative manufacturing steps that makes the final product less dependent on storage temperature. This can include improving the separation process or implement new strategies that do not require the use of organic solvents.

## What is next?

mRNA vaccine technology only reached the spotlight during Covid-19 pandemic event, and although there is a great amount on the research on the use mRNA technology to multiple applications, and over 60 clinical trials are ongoing[7], there is still room for improvement to make this technology reach its full potential. Focusing the manufacturing point-of-view, most of the substantial advances, namely the optimisation of the mRNA production [116,117], were made in the last couple of years. Additionally, an effort has been made to improve the gap on analytical methodologies that fully characterise mRNA product and its impurities and that can be used to tightly manufacturing process. New separation methodologies are being developed to further improve the delivery high quality mRNA [180]. Improving mRNA-LNPs stabilisation to decrease the mRNA dependence on cold-chain to enable is also a hot-topic in mRNA manufacturing research.

Nevertheless, enabling mRNA technology to be widely available is highly dependent on the control of the manufacturing costs. Continuous manufacturing can be an answer to control process cost, as well as reduce operation time and facilitate automation and process analytical technology[6]. Nevertheless, the use of this modality may not be suitable to all the requirements of all mRNA application in a patient-centric perspective. New cell-based mRNA production modes are being explored[263]. New purification strategies based on widely established precipitation methods[264] are being implemented to increase process speed, and new materials[265] are being explored to improve existing purification method capacity and selectivity. Continuous pDNA linearisation methods to streamline template preparation, and continuous capping strategies to simplify the capping process, are being developed [266]. In the end, state-of-art of mRNA vaccines manufacturing still needs to be expanded in order to the development sustainable, flexible and cost-effective manufacturing process, making the technology affordable to all.

## References

- [1] Yang J, Vaghela S, Yarnoff B, De Boisvilliers S, Di Fusco M, Wiemken TL, et al. Estimated global public health and economic impact of COVID-19 vaccines in the pre-omicron era using real-world empirical data. *Expert Rev Vaccines* 2023;22:54–65. <https://doi.org/10.1080/14760584.2023.2157817>.
- [2] Le TT, Andreadakis Z, Kumar A, Román RG, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. *Nat Rev Drug Discov* 2020;19:305–6. <https://doi.org/10.1038/d41573-020-00073-5>.
- [3] Black S, Bloom DE, Kaslow DC, Pecetta S, Rappuoli R. Transforming vaccine development. *Semin Immunol* 2020;50:101413. <https://doi.org/10.1016/j.smim.2020.101413>.
- [4] Brisse M, Vrba SM, Kirk N, Liang Y, Ly H. Emerging Concepts and Technologies in Vaccine Development. *Front Immunol* 2020;11.
- [5] Kumar A, Blum J, Thanh Le T, Havelange N, Magini D, Yoon I-K. The mRNA vaccine development landscape for infectious diseases. *Nat Rev Drug Discov* 2022;21:333–4. <https://doi.org/10.1038/d41573-022-00035-z>.
- [6] Rosa SS, Prazeres DMF, Azevedo AM, Marques MPC. mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine* 2021;39:2190–200. <https://doi.org/10.1016/j.vaccine.2021.03.038>.
- [7] Wang Y-S, Kumari M, Chen G-H, Hong M-H, Yuan JP-Y, Tsai J-L, et al. mRNA-based vaccines and therapeutics: an in-depth survey of current and upcoming clinical applications. *J Biomed Sci* 2023;30:1–35. <https://doi.org/10.1186/s12929-023-00977-5>.
- [8] Jackson NAC, Kester KE, Casimiro D, Gurunathan S, DeRosa F. The promise of mRNA vaccines: a biotech and industrial perspective. *Npj Vaccines* 2020;5:1–6. <https://doi.org/10.1038/s41541-020-0159-8>.
- [9] Moore MJ, Proudfoot NJ. Pre-mRNA Processing Reaches Back to Transcription and Ahead to Translation. *Cell* 2009;136:688–700. <https://doi.org/10.1016/j.cell.2009.02.001>.
- [10] Desterro J, Bak-Gordon P, Carmo-Fonseca M. Targeting mRNA processing as an anticancer strategy. *Nat Rev Drug Discov* 2020;19:112–29. <https://doi.org/10.1038/s41573-019-0042-3>.
- [11] Moulleron H, Delcourt V, Roucou X. Death of a dogma: eukaryotic mRNAs can code for more than one protein. *Nucleic Acids Res* 2016;44:14–23. <https://doi.org/10.1093/nar/gkv1218>.
- [12] Linares-Fernández S, Lacroix C, Exposito J-Y, Verrier B. Tailoring mRNA Vaccine to Balance Innate/Adaptive Immune Response. *Trends Mol Med* 2020;26:311–23. <https://doi.org/10.1016/j.molmed.2019.10.002>.
- [13] Brito LA, Kommareddy S, Maione D, Uematsu Y, Giovani C, Berlanda Scorza F, et al. Chapter Seven - Self-Amplifying mRNA Vaccines. In: Huang L, Liu D, Wagner E, editors. *Adv. Genet.*, vol. 89, Academic Press; 2015, p. 179–233. <https://doi.org/10.1016/bs.adgen.2014.10.005>.
- [14] Vogel AB, Lambert L, Kinnear E, Busse D, Erbar S, Reuter KC, et al. Self-amplifying RNA vaccines give equivalent protection against influenza to mRNA vaccines but at much lower doses. *Mol Ther* 2018;26:446–55.
- [15] Beissert T, Perkovic M, Vogel A, Erbar S, Walzer KC, Hempel T, et al. A Trans-amplifying RNA vaccine strategy for induction of potent protective immunity. *Mol Ther* 2020;28:119–28. <https://doi.org/10.1016/j.ymthe.2019.09.009>.
- [16] Bai Y, Liu D, He Q, Liu J, Mao Q, Liang Z. Research progress on circular RNA vaccines. *Front Immunol* 2023;13:1091797. <https://doi.org/10.3389/fimmu.2022.1091797>.

- [17] Marques R, Lacerda R, Romão L. Internal Ribosome Entry Site (IRES)-Mediated Translation and Its Potential for Novel mRNA-Based Therapy Development. *Biomedicines* 2022;10:1865. <https://doi.org/10.3390/biomedicines10081865>.
- [18] Chen C, Wei H, Zhang K, Li Z, Wei T, Tang C, et al. A flexible, efficient, and scalable platform to produce circular RNAs as new therapeutics 2022:2022.05.31.494115. <https://doi.org/10.1101/2022.05.31.494115>.
- [19] Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene* 2001;276:73–81. [https://doi.org/10.1016/S0378-1119\(01\)00674-6](https://doi.org/10.1016/S0378-1119(01)00674-6).
- [20] Fang E, Liu X, Li M, Zhang Z, Song L, Zhu B, et al. Advances in COVID-19 mRNA vaccine development. *Signal Transduct Target Ther* 2022;7:1–31. <https://doi.org/10.1038/s41392-022-00950-y>.
- [21] Izidoro MS, Sokabe M, Villa N, Merrick WC, Fraser CS. Human eukaryotic initiation factor 4E (eIF4E) and the nucleotide-bound state of eIF4A regulate eIF4F binding to RNA. *J Biol Chem* 2022;298. <https://doi.org/10.1016/j.jbc.2022.102368>.
- [22] Leppek K, Das R, Barna M. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol* 2018;19:158–74. <https://doi.org/10.1038/nrm.2017.103>.
- [23] Adomavicius T, Guaita M, Zhou Y, Jennings MD, Latif Z, Roseman AM, et al. The structural basis of translational control by eIF2 phosphorylation. *Nat Commun* 2019;10:2136. <https://doi.org/10.1038/s41467-019-10167-3>.
- [24] Acevedo JM, Hoermann B, Schlimbach T, Teleman AA. Changes in global translation elongation or initiation rates shape the proteome via the Kozak sequence. *Sci Rep* 2018;8:4018. <https://doi.org/10.1038/s41598-018-22330-9>.
- [25] Zarghampoor F, Azarpira N, Khatami SR, Behzad-Behbahani A, Foroughmand AM. Improved translation efficiency of therapeutic mRNA. *Gene* 2019;707:231–8. <https://doi.org/10.1016/j.gene.2019.05.008>.
- [26] Asrani KH, Farelli JD, Stahley MR, Miller RL, Cheng CJ, Subramanian RR, et al. Optimization of mRNA untranslated regions for improved expression of therapeutic mRNA. *RNA Biol* 2018;15:756–62. <https://doi.org/10.1080/15476286.2018.1450054>.
- [27] Linares-Fernández S, Moreno J, Lambert E, Mercier-Gouy P, Vachez L, Verrier B, et al. Combining an optimized mRNA template with a double purification process allows strong expression of in vitro transcribed mRNA. *Mol Ther - Nucleic Acids* 2021;26:945–56. <https://doi.org/10.1016/j.omtn.2021.10.007>.
- [28] Trepotec Z, Aneja MK, Geiger J, Hasenpusch G, Plank C, Rudolph C. Maximizing the Translational Yield of mRNA Therapeutics by Minimizing 5'-UTRs. *Tissue Eng Part A* 2019;25:69–79. <https://doi.org/10.1089/ten.tea.2017.0485>.
- [29] Thess A. Artificial nucleic acid molecules comprising a 5'top utr. WO2013143700A2, 2013.
- [30] 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;601:410–4. <https://doi.org/10.1038/s41586-021-04231-6>.
- [31] Tusup M, Kundig T, Pascolo S. An eIF4G-recruiting aptamer increases the functionality of in vitro transcribed mRNA. *EPH-Int J Med Health Sci* ISSN 2456-6063 2018;4:29–37.
- [32] Fougerolles A de, Guild J. Modified polynucleotides for the production of secreted proteins. US10385106B2, 2019.
- [33] Angermueller C, Pärnamaa T, Parts L, Stegle O. Deep learning for computational biology. *Mol Syst Biol* 2016;12:878. <https://doi.org/10.15252/msb.20156651>.
- [34] Sample PJ, Wang B, Reid DW, Presnyak V, McFadyen IJ, Morris DR, et al. Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat Biotechnol* 2019;37:803–9. <https://doi.org/10.1038/s41587-019-0164-5>.

- [35] Zeng C, Hou X, Yan J, Zhang C, Li W, Zhao W, et al. Leveraging mRNA Sequences and Nanoparticles to Deliver SARS-CoV-2 Antigens In Vivo. *Adv Mater* 2020;32:2004452. <https://doi.org/10.1002/adma.202004452>.
- [36] Gong H, Wen J, Luo R, Feng Y, Guo J, Fu H, et al. Integrated mRNA sequence optimization using deep learning. *Brief Bioinform* 2023;24:bbad001. <https://doi.org/10.1093/bib/bbad001>.
- [37] 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;9:563–76. <https://doi.org/10.4161/rna.20231>.
- [38] Karikó K, Kuo A, Barnathan ES. Overexpression of urokinase receptor in mammalian cells following administration of the in vitro transcribed encoding mRNA. *Gene Ther* 1999;6:1092–100. <https://doi.org/10.1038/sj.gt.3300930>.
- [39] Carralot J-P, Probst J, Hoerr I, Scheel B, Teufel R, Jung G, et al. Polarization of immunity induced by direct injection of naked sequence-stabilized mRNA vaccines. *Cell Mol Life Sci CMLS* 2004;61:2418–24. <https://doi.org/10.1007/s00018-004-4255-0>.
- [40] Holtkamp S, Kreiter S, Selmi A, Simon P, Koslowski M, Huber C, et al. Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood* 2006;108:4009–17. <https://doi.org/10.1182/blood-2006-04-015024>.
- [41] von Niessen AGO, Poleganov MA, Rechner C, Plaschke A, Kranz LM, Fesser S, et al. Improving mRNA-based therapeutic gene delivery by expression-augmenting 3' UTRs identified by cellular library screening. *Mol Ther* 2019;27:824–36.
- [42] NIESSEN AOV, FESSER S, Vallazza B, Beissert T, Kuhn A, Sahin U. 3' utr sequences for stabilization of rna. WO2017059902A1, 2017.
- [43] Chakraborty T, Bancel S, Hoge SG, Roy A, Fougerolles A de, Afeyan NB. Terminally modified RNA. US10155029B2, 2018.
- [44] Sultana N, Hadas Y, Sharkar MTK, Kaur K, Magadum A, Kurian AA, et al. Optimization of 5' Untranslated Region of Modified mRNA for Use in Cardiac or Hepatic Ischemic Injury. *Mol Ther Methods Clin Dev* 2020;17:622–33. <https://doi.org/10.1016/j.omtm.2020.03.019>.
- [45] Holtkamp S, Kreiter S, Selmi A, Simon P, Koslowski M, Huber C, et al. Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood* 2006;108:4009–17. <https://doi.org/10.1182/blood-2006-04-015024>.
- [46] Ferizi M, Aneja MK, Balmayor ER, Badiyan ZS, Mykhaylyk O, Rudolph C, et al. Human cellular CYBA UTR sequences increase mRNA translation without affecting the half-life of recombinant RNA transcripts. *Sci Rep* 2016;6:39149. <https://doi.org/10.1038/srep39149>.
- [47] Solodushko V, Fouty B. Terminal hairpins improve protein expression in IRES-initiated mRNA in the absence of a cap and polyadenylated tail. *Gene Ther* 2023;1–8. <https://doi.org/10.1038/s41434-023-00391-4>.
- [48] Qin S, Tang X, Chen Y, Chen K, Fan N, Xiao W, et al. mRNA-based therapeutics: powerful and versatile tools to combat diseases. *Signal Transduct Target Ther* 2022;7:1–35. <https://doi.org/10.1038/s41392-022-01007-w>.
- [49] mRNA transfection of dendritic cells: Synergistic effect of ARCA mRNA capping with Poly(A) chains in cis and in trans for a high protein expression level. *Biochem Biophys Res Commun* 2006;340:1062–8. <https://doi.org/10.1016/j.bbrc.2005.12.105>.
- [50] Lima SA, Chipman LB, Nicholson AL, Chen Y-H, Yee BA, Yeo GW, et al. Short poly(A) tails are a conserved feature of highly expressed genes. *Nat Struct Mol Biol* 2017;24:1057–63. <https://doi.org/10.1038/nsmb.3499>.
- [51] Nicholson AL, Pasquinelli AE. Tales of Detailed Poly(A) Tails. *Trends Cell Biol* 2019;29:191–200. <https://doi.org/10.1016/j.tcb.2018.11.002>.

- [52] Li CY, Liang Z, Hu Y, Zhang H, Setiasabda KD, Li J, et al. Cytidine-containing tails robustly enhance and prolong protein production of synthetic mRNA in cell and in vivo. *Mol Ther - Nucleic Acids* 2022;30:300–10. <https://doi.org/10.1016/j.omtn.2022.10.003>.
- [53] Roth N, Schön J, Hoffmann D, Thran M, Thess A, Mueller SO, et al. Optimised Non-Coding Regions of mRNA SARS-CoV-2 Vaccine CV2CoV Improves Homologous and Heterologous Neutralising Antibody Responses. *Vaccines* 2022;10:1251. <https://doi.org/10.3390/vaccines10081251>.
- [54] THEβ A, Schlake T, Probst J. Nucleic acid comprising or coding for a histone stem-loop and a poly(a) sequence or a polyadenylation signal for increasing the expression of an encoded tumour antigen. WO2013120500A1, 2013.
- [55] Baumhof P, RAUCH S, KOWALCZYK A, Lutz J, JASNY E, Petsch B, et al. Optimized nucleic acid molecules. US20200399322A1, 2020.
- [56] Sahin U, Karikó K, Türeci Ö. mRNA-based therapeutics—developing a new class of drugs. *Nat Rev Drug Discov* 2014;13:759.
- [57] Trepotec Z, Geiger J, Plank C, Aneja MK, Rudolph C. Segmented poly(A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life. *RNA* 2019;25:507–18. <https://doi.org/10.1261/rna.069286.118>.
- [58] EBERLE F, Sahin U, Kuhn A, Vallazza B, DIKEN M. Stabilization of poly(a) sequence encoding dna sequences. WO2016005004A1, 2016.
- [59] Ramanathan A, Robb GB, Chan S-H. mRNA capping: biological functions and applications. *Nucleic Acids Res* 2016;44:7511–26.
- [60] Hyde JL, Diamond MS. Innate immune restriction and antagonism of viral RNA lacking 2'-O methylation. *Virology* 2015;479–480:66–74. <https://doi.org/10.1016/j.virol.2015.01.019>.
- [61] Courel M, Clément Y, Bossevain C, Foretek D, Vidal Cruchez O, Yi Z, et al. GC content shapes mRNA storage and decay in human cells. *eLife* 2019;8:e49708. <https://doi.org/10.7554/eLife.49708>.
- [62] Kudla G, Lipinski L, Caffin F, Helwak A, Zylicz M. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS biology*. 2006 Jun;4(6):e180.
- [63] Verbeke R, Hogan MJ, Loré K, Pardi N. Innate immune mechanisms of mRNA vaccines. *Immunity* 2022;55:1993–2005. <https://doi.org/10.1016/j.immuni.2022.10.014>.
- [64] Karikó K, Muramatsu H, Welsh FA, Ludwig J, Kato H, Akira S, et al. Incorporation of Pseudouridine Into mRNA Yields Superior Nonimmunogenic Vector With Increased Translational Capacity and Biological Stability. *Mol Ther* 2008;16:1833–40. <https://doi.org/10.1038/mt.2008.200>.
- [65] Morais P, Adachi H, Yu Y-T. The Critical Contribution of Pseudouridine to mRNA COVID-19 Vaccines. *Front Cell Dev Biol* 2021;9.
- [66] Thess A, Grund S, Mui BL, Hope MJ, Baumhof P, Fotin-Mleczek M, et al. Sequence-engineered mRNA Without Chemical Nucleoside Modifications Enables an Effective Protein Therapy in Large Animals. *Mol Ther* 2015;23:1456–64. <https://doi.org/10.1038/mt.2015.103>.
- [67] Houseley J, Tollervey D. The many pathways of RNA degradation. *Cell* 2009;136:763–76.
- [68] Wayment-Steele HK, Kim DS, Choe CA, Nicol JJ, Wellington-Oguri R, Watkins AM, et al. Theoretical basis for stabilizing messenger RNA through secondary structure design. *Nucleic Acids Res* 2021;49:10604–17. <https://doi.org/10.1093/nar/gkab764>.
- [69] Mauger DM, Cabral BJ, Presnyak V, Su SV, Reid DW, Goodman B, et al. mRNA structure regulates protein expression through changes in functional half-life. *Proc Natl Acad Sci* 2019;116:24075–83. <https://doi.org/10.1073/pnas.1908052116>.

- [70] Leppek K, Byeon GW, Kladwang W, Wayment-Steele HK, Kerr CH, Xu AF, et al. Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics. *Nat Commun* 2022;13:1536. <https://doi.org/10.1038/s41467-022-28776-w>.
- [71] Fallmann J, Will S, Engelhardt J, Grüning B, Backofen R, Stadler PF. Recent advances in RNA folding. *J Biotechnol* 2017;261:97–104. <https://doi.org/10.1016/j.jbiotec.2017.07.007>.
- [72] Sato K, Hamada M. Recent trends in RNA informatics: a review of machine learning and deep learning for RNA secondary structure prediction and RNA drug discovery. *Brief Bioinform* 2023;24:bbad186. <https://doi.org/10.1093/bib/bbad186>.
- [73] Zhang H, Zhang L, Lin A, Xu C, Li Z, Liu K, et al. Algorithm for Optimized mRNA Design Improves Stability and Immunogenicity. *Nature* 2023:1–3. <https://doi.org/10.1038/s41586-023-06127-z>.
- [74] Kim Y-A, Mousavi K, Yazdi A, Zwierzyna M, Cardinali M, Fox D, et al. Computational design of mRNA vaccines. *Vaccine* 2023. <https://doi.org/10.1016/j.vaccine.2023.07.024>.
- [75] Karikó K, Buckstein M, Ni H, Weissman D. Suppression of RNA Recognition by Toll-like Receptors: The Impact of Nucleoside Modification and the Evolutionary Origin of RNA. *Immunity* 2005;23:165–75. <https://doi.org/10.1016/j.immuni.2005.06.008>.
- [76] Svitkin YV, Cheng YM, Chakraborty T, Presnyak V, John M, Sonenberg N. N1-methylpseudouridine in mRNA enhances translation through eIF2 $\alpha$ -dependent and independent mechanisms by increasing ribosome density. *Nucleic Acids Res* 2017;45:6023–36. <https://doi.org/10.1093/nar/gkx135>.
- [77] Litvinova VR, Rudometov AP, Karpenko LI, Ilyichev AA. mRNA Vaccine Platform: mRNA Production and Delivery. *Russ J Bioorganic Chem* 2023;49:220–35. <https://doi.org/10.1134/S1068162023020152>.
- [78] Vogel AB, Kanevsky I, Che Y, Swanson KA, Muik A, Vormehr M, et al. BNT162b vaccines protect rhesus macaques from SARS-CoV-2. *Nature* 2021;592:283–9. <https://doi.org/10.1038/s41586-021-03275-y>.
- [79] Corbett KS, Edwards DK, Leist SR, Abiona OM, Boyoglu-Barnum S, Gillespie RA, et al. SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature* 2020;586:567–71. <https://doi.org/10.1038/s41586-020-2622-0>.
- [80] Nelson J, Sorensen EW, Mintri S, Rabideau AE, Zheng W, Besin G, et al. Impact of mRNA chemistry and manufacturing process on innate immune activation. *Sci Adv* 2020;6:eaaz6893. <https://doi.org/10.1126/sciadv.aaz6893>.
- [81] Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines—a new era in vaccinology. *Nat Rev Drug Discov* 2018;17:261. <https://doi.org/10.1038/nrd.2017.243>.
- [82] Pascolo S. Vaccination With Messenger RNA. In: Saltzman WM, Shen H, Brandsma JL, editors. *DNA Vaccines Methods Protoc.*, Totowa, NJ: Humana Press; 2006, p. 23–40. <https://doi.org/10.1385/1-59745-168-1:23>.
- [83] Godiska R, Mead D, Dhodda V, Wu C, Hochstein R, Karsi A, et al. Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic Acids Res* 2010;38:e88. <https://doi.org/10.1093/nar/gkp1181>.
- [84] A convenient method to generate and maintain poly(A)-encoding DNA sequences required for in vitro transcription of mRNA n.d. <https://doi.org/10.2144/btn-2018-0120>.
- [85] Grier AE, Burleigh S, Sahni J, Clough CA, Cardot V, Choe DC, et al. pEVL: A Linear Plasmid for Generating mRNA IVT Templates With Extended Encoded Poly(A) Sequences. *Mol Ther - Nucleic Acids* 2016;5. <https://doi.org/10.1038/mtna.2016.21>.
- [86] Nance KD, Meier JL. Modifications in an Emergency: The Role of N1-Methylpseudouridine in COVID-19 Vaccines. *ACS Cent Sci* 2021;7:748–56. <https://doi.org/10.1021/acscentsci.1c00197>.

- [87] Wei H, Rong Z, Liu L, Sang Y, Yang J, Wang S. Streamlined and on-demand preparation of mRNA products on a universal integrated platform. *Microsyst Nanoeng* 2023;9:1–10. <https://doi.org/10.1038/s41378-023-00538-8>.
- [88] Niazi SK. RNA Therapeutics: A Healthcare Paradigm Shift. *Biomedicines* 2023;11:1275. <https://doi.org/10.3390/biomedicines11051275>.
- [89] Pascolo S. Messenger RNA-based vaccines. *Expert Opin Biol Ther* 2004;4:1285–94. <https://doi.org/10.1517/14712598.4.8.1285>.
- [90] Urthaler J, Schuchnigg H, Garidel P, Huber H. Industrial Manufacturing of Plasmid-DNA Products for Gene Vaccination and Therapy. In: Thalhamer J, Weiss R, Scheiblhofer S, editors. *Gene Vaccines*, Vienna: Springer; 2012, p. 311–30. [https://doi.org/10.1007/978-3-7091-0439-2\\_16](https://doi.org/10.1007/978-3-7091-0439-2_16).
- [91] Bancel S, Issa WJ, Aunins JG, Chakraborty T. Manufacturing methods for production of RNA transcripts. WO/2014/152027; PCT/US2014/026835; US20160024547A1, 2016.
- [92] Schmeer M, Schleef M. Pharmaceutical Grade Large-Scale Plasmid DNA Manufacturing Process. In: Rinaldi M, Fioretti D, Iurescia S, editors. *DNA Vaccines Methods Protoc.*, New York, NY: Springer; 2014, p. 219–40. [https://doi.org/10.1007/978-1-4939-0410-5\\_14](https://doi.org/10.1007/978-1-4939-0410-5_14).
- [93] Gonçalves GAL, Prather KLJ, Monteiro GA, Carnes AE, Prazeres DMF. Plasmid DNA production with *Escherichia coli* GALG20, a *pgi*-gene knockout strain: Fermentation strategies and impact on downstream processing. *J Biotechnol* 2014;186:119–27. <https://doi.org/10.1016/j.jbiotec.2014.06.008>.
- [94] Ohlson J. Plasmid manufacture is the bottleneck of the genetic medicine revolution. *Drug Discov Today* 2020;25:1891–3. <https://doi.org/10.1016/j.drudis.2020.09.040>.
- [95] EBER FJ, SEWING S, WAGNER W. Method of producing rna from circular dna and corresponding template dna. US20200040370A1, 2020.
- [96] Voß C. Production of plasmid DNA for pharmaceutical use. In: El-Gewely MR, editor. *Biotechnol. Annu. Rev.*, vol. 13, Elsevier; 2007, p. 201–22. [https://doi.org/10.1016/S1387-2656\(07\)13008-8](https://doi.org/10.1016/S1387-2656(07)13008-8).
- [97] Diogo MM, Queiroz JA, Prazeres DMF. Chromatography of plasmid DNA. *J Chromatogr A* 2005;1069:3–22. <https://doi.org/10.1016/j.chroma.2004.09.050>.
- [98] Kahn DW, Butler MD, Cohen DL, Gordon M, Kahn JW, Winkler ME. Purification of plasmid DNA by tangential flow filtration. *Biotechnol Bioeng* 2000;69:101–6. [https://doi.org/10.1002/\(SICI\)1097-0290\(20000705\)69:1<101::AID-BIT12>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-0290(20000705)69:1<101::AID-BIT12>3.0.CO;2-1).
- [99] Beck J, Biechele M, Repik C, Gruber P, Furtmüller PG, Hahn R. Desorption of plasmid DNA from anion exchangers: Salt concentration at elution is independent of plasmid size and load. *J Sep Sci* 2023;46:2200943. <https://doi.org/10.1002/jssc.202200943>.
- [100] Bo H, Wang J, Chen Q, Shen H, Wu F, Shao H, et al. Using a single hydrophobic-interaction chromatography to purify pharmaceutical-grade supercoiled plasmid DNA from other isoforms. *Pharm Biol* 2013;51:42–8. <https://doi.org/10.3109/13880209.2012.703678>.
- [101] Božič K, Sedlar A, Kralj Š, Černigoj U, Štrancar A, Sekirnik R. Selective hydrophobic interaction chromatography for high purity of supercoiled DNA plasmids. *Biotechnol Bioeng* 2024;121:1739–49.
- [102] Sousa F, Prazeres DMF, Queiroz JA. Affinity chromatography approaches to overcome the challenges of purifying plasmid DNA. *Trends Biotechnol* 2008;26:518–25. <https://doi.org/10.1016/j.tibtech.2008.05.005>.
- [103] Li L-Z, Liu Y, Sun M-S, Shao Y-M. Effect of salt on purification of plasmid DNA using size-exclusion chromatography. *J Chromatogr A* 2007;1139:228–35. <https://doi.org/10.1016/j.chroma.2006.11.027>.
- [104] Silva-Santos AR, Rosa SS, Prazeres DMF, Azevedo AM. Purification of Plasmid DNA by Multimodal Chromatography. In: Sousa Â, editor. *DNA Vaccines Methods Protoc.*, New

- York, NY: Springer US; 2021, p. 193–205. [https://doi.org/10.1007/978-1-0716-0872-2\\_10](https://doi.org/10.1007/978-1-0716-0872-2_10).
- [105] Matos T, Queiroz JA, Bülow L. Plasmid DNA purification using a multimodal chromatography resin. *J Mol Recognit* 2014;27:184–9. <https://doi.org/10.1002/jmr.2349>.
- [106] Silva-Santos AR, Alves CPA, Prazeres DMF, Azevedo AM. Separation of plasmid DNA topoisomers by multimodal chromatography. *Anal Biochem* 2016;503:68–70. <https://doi.org/10.1016/j.ab.2016.03.012>.
- [107] Valente JFA, Queiroz JA, Sousa F. Dilemma on plasmid DNA purification: binding capacity vs selectivity. *J Chromatogr A* 2021;1637:461848. <https://doi.org/10.1016/j.chroma.2020.461848>.
- [108] Kodermac MŠ, Rotar S, Dolenc D, Božič K, Štrancar A, Černigoj U. Grafting chromatographic monoliths with charged linear polymers for highly productive and selective protein or plasmid DNA purification. *Sep Purif Technol* 2025;353:128253. <https://doi.org/10.1016/j.seppur.2024.128253>.
- [109] Funkner A, Dorner S, Sewing S, Johannes K, Broghammer N, Ketterer T, et al. Method for producing and purifying RNA, comprising at least one step of tangential flow filtration. PCT/EP2016/062152; WO/2016/193206, 2018.
- [110] Roos T, PANAHA BY, Kunze M. Method for in vitro transcription using an immobilized restriction enzyme. EP3289077B1, 2020.
- [111] Diaz GA, Rong M, McAllister WT, Durbin RK. The Stability of Abortively Cycling T7 RNA Polymerase Complexes Depends upon Template Conformation. *Biochemistry* 1996;35:10837–43. <https://doi.org/10.1021/bi960488+>.
- [112] Evaluation of the quality, safety and efficacy of messenger RNA vaccines for the prevention of infectious diseases: regulatory considerations n.d. [https://cdn.who.int/media/docs/default-source/biologicals/vaccine-standardization/annex-3---mrna-vaccines\\_who\\_trs\\_1039\\_web-2.pdf?sfvrsn=6e31a112\\_1&download=true](https://cdn.who.int/media/docs/default-source/biologicals/vaccine-standardization/annex-3---mrna-vaccines_who_trs_1039_web-2.pdf?sfvrsn=6e31a112_1&download=true) (accessed August 27, 2023).
- [113] Schmeer M, Buchholz T, Schleef M. Plasmid DNA Manufacturing for Indirect and Direct Clinical Applications. *Hum Gene Ther* 2017;28:856–61. <https://doi.org/10.1089/hum.2017.159>.
- [114] Plasmid release specifications – response to CGT industry feedback. n.d.
- [115] Pfizer E. Assessment Report Comirnaty. Procedure No. EMEA/H/C/005735/0000. Netherlands: European Medicines Agency; 2021.
- [116] Rosa SS, Nunes D, Antunes L, Prazeres DMF, Marques MPC, Azevedo AM. Maximizing mRNA vaccine production with Bayesian optimization. *Biotechnol Bioeng* 2022;119:3127–39. <https://doi.org/10.1002/bit.28216>.
- [117] Skok J, Megušar P, Vodopivec T, Pregeljč D, Mencin N, Korenč M, et al. Gram-Scale mRNA Production Using a 250-mL Single-Use Bioreactor. *Chem Ing Tech* 2022;94:1928–35. <https://doi.org/10.1002/cite.202200133>.
- [118] Borkotoky S, Kumar Meena C, Bhalerao GM, Murali A. An in-silico glimpse into the pH dependent structural changes of T7 RNA polymerase: a protein with simplicity. *Sci Rep* 2017;7:6290. <https://doi.org/10.1038/s41598-017-06586-1>.
- [119] Borkotoky S, Murali A. The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *Int J Biol Macromol* 2018;118:49–56. <https://doi.org/10.1016/j.ijbiomac.2018.05.198>.
- [120] Guo Q, Nayak D, Brieba LG, Sousa R. Major Conformational Changes During T7RNAP Transcription Initiation Coincide with, and are Required for, Promoter Release. *J Mol Biol* 2005;353:256–70. <https://doi.org/10.1016/j.jmb.2005.08.016>.
- [121] Tahirov TH, Temiakov D, Anikin M, Patlan V, McAllister WT, Vassylyev DG, et al. Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* 2002;420:43–50. <https://doi.org/10.1038/nature01129>.

- [122] Temiakov D, Montesana PE, Ma K, Mustaev A, Borukhov S, McAllister WT. The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc Natl Acad Sci* 2000;97:14109–14. <https://doi.org/10.1073/pnas.250473197>.
- [123] Dousis A, Ravichandran K, Hobert EM, Moore MJ, Rabideau AE. An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts. *Nat Biotechnol* 2023;41:560–8. <https://doi.org/10.1038/s41587-022-01525-6>.
- [124] Gunderson SI, Chapman KA, Burgess RR. Interactions of T7 RNA polymerase with T7 late promoters measured by footprinting with methidiumpropyl-EDTA-iron(II). *ACS Publ* 2002. <https://doi.org/10.1021/bi00380a007>.
- [125] Breckenridge NC, Davis RH. Optimization of repeated-batch transcription for RNA production. *Biotechnol Bioeng* 2000;69:679–87. [https://doi.org/10.1002/1097-0290\(20000920\)69:6<679::AID-BIT12>3.0.CO;2-6](https://doi.org/10.1002/1097-0290(20000920)69:6<679::AID-BIT12>3.0.CO;2-6).
- [126] Kern JA, Davis RH. Application of solution equilibrium analysis to in vitro RNA transcription. *Biotechnol Prog* 1997;13:747–56. <https://doi.org/10.1021/bp970094p>.
- [127] Von Der Mülbe F, Reidel L, Ketterer T, Gontcharova L, Bauer S, Pascolo S, et al. Method for producing rna. PCT/EP2015/000959; US10017826B2; WO/2016/180430A1, 2015.
- [128] Feyrer H, Munteanu R, Baronti L, Petzold K. One-Pot Production of RNA in High Yield and Purity Through Cleaving Tandem Transcripts. *Molecules* 2020;25:1142. <https://doi.org/10.3390/molecules25051142>.
- [129] Braatz R, Stover NM, Ganko K. Mechanistic Modeling of In Vitro Transcription. Preprints; 2023. <https://doi.org/10.22541/au.169299542.25533105/v1>.
- [130] Arnold S, Siemann M, Scharnweber K, Werner M, Baumann S, Reuss M. Kinetic modeling and simulation of in vitro transcription by phage T7 RNA polymerase. *Biotechnol Bioeng* 2001;72:548–61. [https://doi.org/10.1002/1097-0290\(20010305\)72:5<548::AID-BIT1019>3.0.CO;2-2](https://doi.org/10.1002/1097-0290(20010305)72:5<548::AID-BIT1019>3.0.CO;2-2).
- [131] Pokrovskaya ID, Gurevich VV. In Vitro Transcription: Preparative RNA Yields in Analytical Scale Reactions. *Anal Biochem* 1994;220:420–3. <https://doi.org/10.1006/abio.1994.1360>.
- [132] Young JS, Ramirez WF, Davis RH. Modeling and optimization of a batch process for in vitro RNA production. *Biotechnol Bioeng* 1997;56:210–20. [https://doi.org/10.1002/\(SICI\)1097-0290\(19971020\)56:2<210::AID-BIT10>3.0.CO;2-K](https://doi.org/10.1002/(SICI)1097-0290(19971020)56:2<210::AID-BIT10>3.0.CO;2-K).
- [133] Henderson JM, Ujita A, Hill E, Yousif-Rosales S, Smith C, Ko N, et al. Cap 1 Messenger RNA Synthesis with Co-transcriptional CleanCap® Analog by In Vitro Transcription. *Curr Protoc* 2021;1:e39. <https://doi.org/10.1002/cpz1.39>.
- [134] Oakley JL, Pascale JA, Coleman JE. T7 RNA polymerase. Conformation, functional groups, and promoter binding. *ACS Publ* 2002. <https://doi.org/10.1021/bi00692a019>.
- [135] Wu MZ, Asahara H, Tzertzinis G, Roy B. Synthesis of low immunogenicity RNA with high-temperature in vitro transcription. *RNA* 2020;26:345–60. <https://doi.org/10.1261/rna.073858.119>.
- [136] Bancel S, Issa WJ, Aunins JG, Chakraborty T. Manufacturing methods for production of rna transcripts. EP3578663A1, 2019.
- [137] Kartje ZJ, Janis HI, Mukhopadhyay S, Gagnon KT. Revisiting T7 RNA polymerase transcription in vitro with the Broccoli RNA aptamer as a simplified real-time fluorescent reporter. *J Biol Chem* 2021;296. <https://doi.org/10.1074/jbc.RA120.014553>.
- [138] Watanabe Y, Igarashi K, Mitsui K, Hirose S. Differential stimulation by polyamines of phage DNA-directed in vitro synthesis of proteins. *Biochim Biophys Acta* 1983;740:362–8. [https://doi.org/10.1016/0167-4781\(83\)90083-0](https://doi.org/10.1016/0167-4781(83)90083-0).
- [139] Gosule LC, Schellman JA. DNA condensation with polyamines: I. Spectroscopic studies. *J Mol Biol* 1978;121:311–26. [https://doi.org/10.1016/0022-2836\(78\)90366-2](https://doi.org/10.1016/0022-2836(78)90366-2).

- [140] Fuchs E. The Interdependence of Magnesium with Spermidine and Phosphoenolpyruvate in an Enzyme-Synthesizing System in vitro. *Eur J Biochem* 1976;63:15–22. <https://doi.org/10.1111/j.1432-1033.1976.tb10201.x>.
- [141] Moussatché N. Polyamines stimulate DNA-dependent RNA synthesis catalyzed by vaccinia virus. *Biochim Biophys Acta BBA - Gene Struct Expr* 1985;826:113–20. [https://doi.org/10.1016/0167-4781\(85\)90116-2](https://doi.org/10.1016/0167-4781(85)90116-2).
- [142] Piao X, Yadav V, Wang E, Chang W, Tau L, Lindenmuth BE, et al. Double-stranded RNA reduction by chaotropic agents during in vitro transcription of messenger RNA. *Mol Ther - Nucleic Acids* 2022;29:618–24. <https://doi.org/10.1016/j.omtn.2022.08.001>.
- [143] Labourier E, Pasloske BL. Methods and compositions for preparing capped rna. WO2006004648A1, 2006.
- [144] Kore AR, Charles I. Synthesis and evaluation of 2'-O-allyl substituted dinucleotide cap analog for mRNA translation. *Bioorg Med Chem* 2010;18:8061–5. <https://doi.org/10.1016/j.bmc.2010.09.013>.
- [145] Stepinski J, Waddell C, Stolarski R, Darzynkiewicz E, Rhoads RE. Synthesis and properties of mRNAs containing the novel “anti-reverse” cap analogs 7-methyl(3'-O-methyl)GpppG and 7-methyl(3'-deoxy)GpppG. *RNA* 2001;7:1486–95.
- [146] Wochner A, Roos T, Ketterer T. Methods and means for enhancing rna production. US20170114378A1, 2017.
- [147] Hogrefe RI, Lebedev A, McCaffrey AP, Shin D. Compositions and methods for synthesizing 5'-capped rnas. US20180273576A1, 2018.
- [148] Senthilvelan A, Vonderfecht T, Shanmugasundaram M, Potter J, Kore AR. Click-iT trinucleotide cap analog: Synthesis, mRNA translation, and detection. *Bioorg Med Chem* 2023;77:117128. <https://doi.org/10.1016/j.bmc.2022.117128>.
- [149] Fuchs A-L, Neu A, Sprangers R. A general method for rapid and cost-efficient large-scale production of 5' capped RNA. *RNA* 2016;22:1454–66. <https://doi.org/10.1261/rna.056614.116>.
- [150] Chan SH, Molé CN, Nye D, Mitchell L, Dai N, Buss J, et al. Biochemical characterization of mRNA capping enzyme from Faustovirus. *RNA* 2023:rna.079738.123. <https://doi.org/10.1261/rna.079738.123>.
- [151] Feng X, Su Z, Cheng Y, Ma G, Zhang S. Messenger RNA chromatographic purification: advances and challenges. *J Chromatogr A* 2023;1707:464321. <https://doi.org/10.1016/j.chroma.2023.464321>.
- [152] Pascolo S. Synthetic Messenger RNA-Based Vaccines: From Scorn to Hype. *Viruses* 2021;13:270. <https://doi.org/10.3390/v13020270>.
- [153] Walker SE, Lorsch J. Chapter Nineteen - RNA Purification – Precipitation Methods. In: Lorsch J, editor. *Methods Enzymol.*, vol. 530, Academic Press; 2013, p. 337–43. <https://doi.org/10.1016/B978-0-12-420037-1.00019-1>.
- [154] DeRosa F, Dias A, Karve S, Heartlein M. Methods for purification of messenger RNA. US9850269B2, 2017.
- [155] Easton LE, Shibata Y, Lukavsky PJ. Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* 2010;16:647–53. <https://doi.org/10.1261/rna.1862210>.
- [156] McKenna SA, Kim I, Puglisi EV, Lindhout DA, Aitken CE, Marshall RA, et al. Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nat Protoc* 2007;2:3270–7. <https://doi.org/10.1038/nprot.2007.480>.
- [157] Daniel S, Kis Z, Kontoravdi C, Shah N. Quality by Design for enabling RNA platform production processes. *Trends Biotechnol* 2022;40:1213–28. <https://doi.org/10.1016/j.tibtech.2022.03.012>.
- [158] Diogo MM, Queiroz JA, Prazeres DMF. Assessment of purity and quantification of plasmid DNA in process solutions using high-performance hydrophobic interaction

- chromatography. *J Chromatogr A* 2003;998:109–17. [https://doi.org/10.1016/S0021-9673\(03\)00618-6](https://doi.org/10.1016/S0021-9673(03)00618-6).
- [159] Cui T, Fakhfakh K, Turney H, Güler-Gane G, Toloczko A, Hulley M, et al. Comprehensive studies on building a scalable downstream process for mRNAs to enable mRNA therapeutics. *Biotechnol Prog* 2023;39:e3301. <https://doi.org/10.1002/btpr.3301>.
- [160] Issa WJ, Wang Y, Bancel S. Removal of DNA fragments in mRNA production process. WO/2014/152030A1; US10077439B2, 2016.
- [161] Scorza FB, Wen Y, Geall A, Porter F. Rna purification methods. US20210214388A1, 2021.
- [162] Issa WJ, Barberio JL, Aunins JG, Afeyan NB. Ion exchange purification of mrna. US/2016/0024141 A1, 2016.
- [163] Martin CT, Muller DK, Coleman JE. Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* 1988;27:3966–74. <https://doi.org/10.1021/bi00411a012>.
- [164] Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* 1987;15:8783–98. <https://doi.org/10.1093/nar/15.21.8783>.
- [165] Cazenave C, Uhlenbeck OC. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc Natl Acad Sci U S A* 1994;91:6972–6.
- [166] Gholamalipour Y, Karunanayake Mudiyanseilage A, Martin CT. 3' end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character-RNA-Seq analyses. *Nucleic Acids Res* 2018;46:9253–63. <https://doi.org/10.1093/nar/gky796>.
- [167] Mu X, Greenwald E, Ahmad S, Hur S. An origin of the immunogenicity of in vitro transcribed RNA. *Nucleic Acids Res* 2018;46:5239–49. <https://doi.org/10.1093/nar/gky177>.
- [168] Milano G, Gal J, Creisson A, Chamorey E. Myocarditis and COVID-19 mRNA vaccines: a mechanistic hypothesis involving dsRNA. *Future Virol* 2021.
- [169] Kariko K, Muramatsu H, Ludwig J, Weissman D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Res* 2011;39:e142–e142. <https://doi.org/10.1093/nar/gkr695>.
- [170] Miller M, Alvizo O, Chng C, Jenne S, Mayo M, Mukherjee A, et al. An Engineered T7 RNA Polymerase for efficient co-transcriptional capping with reduced dsRNA byproducts in mRNA synthesis 2022:2022.09.01.506264. <https://doi.org/10.1101/2022.09.01.506264>.
- [171] Gholamalipour Y, Johnson WC, Martin CT. Efficient inhibition of RNA self-primed extension by addition of competing 3'-capture DNA-improved RNA synthesis by T7 RNA polymerase. *Nucleic Acids Res* 2019;47:e118. <https://doi.org/10.1093/nar/gkz700>.
- [172] Weissman D, Pardi N, Muramatsu H, Karikó K. HPLC Purification of In Vitro Transcribed Long RNA. In: Rabinovich PM, editor. *Synth. Messenger RNA Cell Metab. Modul. Methods Protoc.*, Totowa, NJ: Humana Press; 2013, p. 43–54. [https://doi.org/10.1007/978-1-62703-260-5\\_3](https://doi.org/10.1007/978-1-62703-260-5_3).
- [173] Baiersdörfer M, Boros G, Muramatsu H, Mahiny A, Vlatkovic I, Sahin U, et al. A Facile Method for the Removal of dsRNA Contaminant from In Vitro-Transcribed mRNA. *Mol Ther - Nucleic Acids* 2019;15:26–35. <https://doi.org/10.1016/j.omtn.2019.02.018>.
- [174] Miracco EJ. Rna affinity purification. US20220290125A1, 2022.
- [175] Mencin N, Štepec D, Margon A, Vidič J, Dolenc D, Simčič T, et al. Development and scale-up of oligo-dT monolithic chromatographic column for mRNA capture through understanding of base-pairing interactions. *Sep Purif Technol* 2023;304:122320. <https://doi.org/10.1016/j.seppur.2022.122320>.
- [176] Korenč M, Mencin N, Puc J, Skok J, Nemeč KŠ, Celjar AM, Gagnon P, Štrancar A. Chromatographic purification with CIMmultus™ Oligo dT increases mRNA stability. *Cell and Gene Therapy Insights*. 2021 Oct 13. [177] Wommer L, Meiers P, Kockler I,

- Ulber R, Kampeis P. Development of a 3D-printed single-use separation chamber for use in mRNA-based vaccine production with magnetic microparticles. *Eng Life Sci* 2021;21:573–88. <https://doi.org/10.1002/elsc.202000120>.
- [178] Grinsted J, Liddell J, Bouleghlimat E, Kwok KY, Taylor G, Marques MPC, et al. Purification of therapeutic & prophylactic mRNA by affinity chromatography. *Cell Gene Ther Insights* 2022;8:335–49. <https://doi.org/10.18609/cgti.2022.049>.
- [179] Two new capture options for improved purification of large mRNA. *BioInsights* n.d. <https://www.insights.bio/cell-and-gene-therapy-insights/journal/article/1712/Two-new-capture-options-for-improved-purification-of-large-mRNA> (accessed March 26, 2023).
- [180] Megušar P, Miklavčič R, Korenč M, Ličen J, Vodopivec T, Černigoj U, et al. Scalable multimodal weak anion exchange chromatographic purification for stable mRNA drug substance. *ELECTROPHORESIS* n.d.;n/a. <https://doi.org/10.1002/elps.202300106>.
- [181] Miklavčič R, Megušar P, Kodermac ŠM, Bakalar B, Dolenc D, Sekirnik R, et al. High Recovery Chromatographic Purification of mRNA at Room Temperature and Neutral pH. *Int J Mol Sci* 2023;24:14267. <https://doi.org/10.3390/ijms241814267>.
- [182] Castellanos MM, Gressard H, Li X, Magagnoli C, Moriconi A, Stranges D, et al. CMC Strategies and Advanced Technologies for Vaccine Development to Boost Acceleration and Pandemic Preparedness. *Vaccines* 2023;11:1153. <https://doi.org/10.3390/vaccines11071153>.
- [183] World Health Organization. Call for Public Consultation – Evaluation of the quality, safety and efficacy of RNA-based prophylactic vaccines for infectious diseases: regulatory considerations n.d. <https://www.who.int/news-room/articles-detail/call-for-public-consultation-evaluation-of-the-quality-safety-and-efficacy-of-rna-based-prophylactic-vaccines-for-infectious-diseases-regulatory-considerations> (accessed January 31, 2024).
- [184] De Vos J, Morreel K, Alvarez P, Vanluchene H, Vankeirsbilck R, Sandra P, et al. Evaluation of size-exclusion chromatography, multi-angle light scattering detection and mass photometry for the characterization of mRNA. *J Chromatogr A* 2024;1719:464756. <https://doi.org/10.1016/j.chroma.2024.464756>.
- [185] Sacher S, Poms J, Rehrl J, Khinast JG. PAT implementation for advanced process control in solid dosage manufacturing – A practical guide. *Int J Pharm* 2022;613:121408. <https://doi.org/10.1016/j.ijpharm.2021.121408>.
- [186] Pregeljč D, Skok J, Vodopivec T, Mencin N, Krušič A, Ličen J, et al. Increasing yield of in vitro transcription reaction with at-line high pressure liquid chromatography monitoring. *Biotechnol Bioeng* 2023;120:737–47. <https://doi.org/10.1002/bit.28299>.
- [187] Issa W, Packer M. Methods for hplc analysis. US20210163919A1, 2021.
- [188] Kanavarioti A. HPLC methods for purity evaluation of man-made single-stranded RNAs. *Sci Rep* 2019;9:1019. <https://doi.org/10.1038/s41598-018-37642-z>.
- [189] Fekete S, Doneanu C, Addepalli B, Gaye M, Nguyen J, Alden B, et al. Challenges and emerging trends in liquid chromatography-based analyses of mRNA pharmaceuticals. *J Pharm Biomed Anal* 2023;224:115174. <https://doi.org/10.1016/j.jpba.2022.115174>.
- [190] Welbourne EN, Loveday KA, Nair A, Nourafkan E, Qu J, Cook K, et al. Anion exchange HPLC monitoring of mRNA in vitro transcription reactions to support mRNA manufacturing process development. *Front Mol Biosci* 2024;11:1250833.
- [191] Marras SAE, Gold B, Kramer FR, Smith I, Tyagi S. Real-time measurement of in vitro transcription. *Nucleic Acids Res* 2004;32:e72. <https://doi.org/10.1093/nar/gnh068>.
- [192] Zhang F, Wang Y, Wang X, Dong H, Chen M, Du N, et al. RT-IVT method allows multiplex real-time quantification of in vitro transcriptional mRNA production. *Commun Biol* 2023;6:1–11. <https://doi.org/10.1038/s42003-023-04830-1>.
- [193] Theus SA, Liarakos CD. A simple assay for determining the capping efficiencies of RNA polymerases used for in vitro transcription. *BioTechniques* 1990;9:610–2, 614–5.

- [194] Chiron S, Jais P. Non-radioactive monitoring assay for capping of messenger RNA. *J. Genet. Genomics*, vol. 1, 2019, p. 46–9.
- [195] Pregeljč D, Skok J, Vodopivec T, Mencin N, Krušič A, Ličen J, et al. Increasing yield of in vitro transcription reaction with at-line high pressure liquid chromatography monitoring. *Biotechnol Bioeng* 2023;120:737–47. <https://doi.org/10.1002/bit.28299>.
- [196] Beverly M, Dell A, Parmar P, Houghton L. Label-free analysis of mRNA capping efficiency using RNase H probes and LC-MS. *Anal Bioanal Chem* 2016;408:5021–30. <https://doi.org/10.1007/s00216-016-9605-x>.
- [197] Chan SH, Whipple JM, Dai N, Kelley TM, Withers K, Tzertzinis G, et al. RNase H-based analysis of synthetic mRNA 5' cap incorporation. *RNA* 2022;28:1144–55. <https://doi.org/10.1261/rna.079173.122>.
- [198] Vlatkovic I, Ludwig J, Boros G, Szabó GT, Reichert J, Buff M, et al. Ribozyme Assays to Quantify the Capping Efficiency of In Vitro-Transcribed mRNA. *Pharmaceutics* 2022;14:328. <https://doi.org/10.3390/pharmaceutics14020328>.
- [199] Heartlein M, DeRosa F, Dias A. Quantitative assessment for cap efficiency of messenger RNA. US11365437B2, 2022.
- [200] Luo D, Wu Z, Wang D, Zhang J, Shao F, Wang S, et al. Lateral flow immunoassay for rapid and sensitive detection of dsRNA contaminants in in vitro-transcribed mRNA products. *Mol Ther - Nucleic Acids* 2023;32:445–53. <https://doi.org/10.1016/j.omtn.2023.04.005>.
- [201] Beverly M, Hagen C, Slack O. Poly A tail length analysis of in vitro transcribed mRNA by LC-MS. *Anal Bioanal Chem* 2018;410:1667–77. <https://doi.org/10.1007/s00216-017-0840-6>.
- [202] Di Grandi D, Dayeh DM, Kaur K, Chen Y, Henderson S, Moon Y, et al. A Single-Nucleotide Resolution Capillary Gel Electrophoresis Workflow for Poly(A) Tail Characterization in the Development of mRNA Therapeutics and Vaccines. *J Pharm Biomed Anal* 2023;115692. <https://doi.org/10.1016/j.jpba.2023.115692>.
- [203] Warzak DA, Pike WA, Luttgaharm KD. Capillary electrophoresis methods for determining the IVT mRNA critical quality attributes of size and purity. *SLAS Technol* 2023. <https://doi.org/10.1016/j.slast.2023.06.005>.
- [204] Sun Y, Lu Z, Miller M, Perroud T, Tong Y. Application of microfluidic chip electrophoresis for high-throughput nucleic acid fluorescence fragment analysis assays. *NAR Genomics Bioinforma* 2023;5:lqad011. <https://doi.org/10.1093/nargab/lqad011>.
- [205] Peña ACD, Li N, Vaduva M, Bwanali L, Tripathi A. A microfluidic electrophoretic dual dynamic staining method for the identification and relative quantitation of dsRNA contaminants in mRNA vaccines. *Analyst* 2023;148:3758–67. <https://doi.org/10.1039/D3AN00281K>.
- [206] Kreiter S, Selmi A, Diken M, Koslowski M, Britten CM, Huber C, et al. Intranodal vaccination with naked antigen-encoding RNA elicits potent prophylactic and therapeutic antitumoral immunity. *Cancer Res* 2010;70:9031–40. <https://doi.org/10.1158/0008-5472.CAN-10-0699>.
- [207] Kreiter S, Selmi A, Diken M, Koslowski M, Britten CM, Huber C, et al. Intranodal vaccination with naked antigen-encoding RNA elicits potent prophylactic and therapeutic antitumoral immunity. *Cancer Res* 2010;70:9031–40. <https://doi.org/10.1158/0008-5472.CAN-10-0699>.
- [208] Khoury HJ, Collins RH, Blum W, Stiff PS, Elias L, Lebkowski JS, et al. Immune responses and long-term disease recurrence status after telomerase-based dendritic cell immunotherapy in patients with acute myeloid leukemia. *Cancer* 2017;123:3061–72. <https://doi.org/10.1002/cncr.30696>.
- [209] Weide B, Carralot J-P, Reese A, Scheel B, Eigentler TK, Hoerr I, et al. Results of the First Phase I/II Clinical Vaccination Trial With Direct Injection of mRNA. *J Immunother* 2008;31:180–8. <https://doi.org/10.1097/CJI.0b013e31815ce501>.

- [210] Li M, Li Y, Li S, Jia L, Wang H, Li M, et al. The nano delivery systems and applications of mRNA. *Eur J Med Chem* 2022;227:113910. <https://doi.org/10.1016/j.ejmech.2021.113910>.
- [211] Buonaguro L, Mayer-Mokler A, Accolla R, Ma YT, Heidenreich R, Avallone A, et al. HepaVac-101 first-in-man therapeutic cancer vaccine phase I/II clinical trial for hepatocellular carcinoma patients. *J Clin Oncol* 2018;36:TPS3135–TPS3135. [https://doi.org/10.1200/JCO.2018.36.15\\_suppl.TPS3135](https://doi.org/10.1200/JCO.2018.36.15_suppl.TPS3135).
- [212] Matsui A, Uchida S, Ishii T, Itaka K, Kataoka K. Messenger RNA-based therapeutics for the treatment of apoptosis-associated diseases. *Sci Rep* 2015;5:1–10. <https://doi.org/10.1038/srep15810>.
- [213] Lichtenegger FS, Schnorfeil FM, Rothe M, Deiser K, Altmann T, Bücklein VL, et al. Toll-like receptor 7/8-matured RNA-transduced dendritic cells as post-remission therapy in acute myeloid leukaemia: results of a phase I trial. *Clin Transl Immunol* 2020;9:e1117. <https://doi.org/10.1002/cti2.1117>.
- [214] Haas H, Herrero JM, SCHLEGEL AMG, Erbar S. Rna formulations suitable for therapy. *US20220362388A1*, 2022.
- [215] Blakney AK, Zhu Y, McKay PF, Bouton CR, Yeow J, Tang J, et al. Big Is Beautiful: Enhanced saRNA Delivery and Immunogenicity by a Higher Molecular Weight, Bioreducible, Cationic Polymer. *ACS Nano* 2020;14:5711–27. <https://doi.org/10.1021/acsnano.0c00326>.
- [216] Li M, Li Y, Peng K, Wang Y, Gong T, Zhang Z, et al. Engineering intranasal mRNA vaccines to enhance lymph node trafficking and immune responses. *Acta Biomater* 2017;64:237–48. <https://doi.org/10.1016/j.actbio.2017.10.019>.
- [217] Rui Y, Wilson DR, Tzeng SY, Yamagata HM, Sudhakar D, Conge M, et al. High-throughput and high-content bioassay enables tuning of polyester nanoparticles for cellular uptake, endosomal escape, and systemic in vivo delivery of mRNA. *Sci Adv* 2022;8:eabk2855. <https://doi.org/10.1126/sciadv.abk2855>.
- [218] Yang W, Mixich L, Boonstra E, Cabral H. Polymer-Based mRNA Delivery Strategies for Advanced Therapies. *Adv Healthc Mater* n.d.;n/a:2202688. <https://doi.org/10.1002/adhm.202202688>.
- [219] Ramachandran S, Satapathy SR, Dutta T. Delivery Strategies for mRNA Vaccines. *Pharm Med* 2022;36:11–20. <https://doi.org/10.1007/s40290-021-00417-5>.
- [220] Armbruster N, Jasny E, Petsch B. Advances in RNA vaccines for preventive indications: A case study of a vaccine against rabies. *Vaccines* 2019;7:132. <https://doi.org/10.3390/vaccines7040132>.
- [221] Hoerr I, Obst R, Rammensee H-G, Jung G. In vivo application of RNA leads to induction of specific cytotoxic T lymphocytes and antibodies. *Eur J Immunol* 2000;30:1–7. <https://doi.org/10.1002/1521-4141>.
- [222] Udhayakumar VK, Beuckelaer AD, McCaffrey J, McCrudden CM, Kirschman JL, Vanover D, et al. Arginine-Rich Peptide-Based mRNA Nanocomplexes Efficiently Instigate Cytotoxic T Cell Immunity Dependent on the Amphipathic Organization of the Peptide. *Adv Healthc Mater* 2017;6:1601412. <https://doi.org/10.1002/adhm.201601412>.
- [223] Lou B, De Koker S, Lau CYJ, Hennink WE, Mastrobattista E. mRNA Polyplexes with Post-Conjugated GALA Peptides Efficiently Target, Transfect, and Activate Antigen Presenting Cells. *Bioconjug Chem* 2019;30:461–75. <https://doi.org/10.1021/acs.bioconjchem.8b00524>.
- [224] Kim J, Eygeris Y, Gupta M, Sahay G. Self-assembled mRNA vaccines. *Adv Drug Deliv Rev* 2021;170:83–112. <https://doi.org/10.1016/j.addr.2020.12.014>.
- [225] Mui BL, Tam YK, Jayaraman M, Ansell SM, Du X, Tam YYC, et al. Influence of Polyethylene Glycol Lipid Desorption Rates on Pharmacokinetics and Pharmacodynamics of siRNA Lipid Nanoparticles. *Mol Ther - Nucleic Acids* 2013;2:e139. <https://doi.org/10.1038/mtna.2013.66>.

- [226] Hajj KA, Whitehead KA. Tools for translation: non-viral materials for therapeutic mRNA delivery. *Nat Rev Mater* 2017;2:1–17.
- [227] Szebeni J, Kiss B, Bozó T, Turjeman K, Levi-Kalisman Y, Barenholz Y, et al. New insights into the structure of Comirnaty Covid-19 vaccine: A theory on soft nanoparticles with mRNA-lipid supercoils stabilized by hydrogen bonds 2022:2022.12.02.518611. <https://doi.org/10.1101/2022.12.02.518611>.
- [228] Rissanou AN, Ouranidis A, Karatasos K. Complexation of single stranded RNA with an ionizable lipid: an all-atom molecular dynamics simulation study. *Soft Matter* 2020;16:6993–7005. <https://doi.org/10.1039/D0SM00736F>.
- [229] Guevara ML, Persano F, Persano S. Advances in Lipid Nanoparticles for mRNA-Based Cancer Immunotherapy. *Front Chem* 2020;8.
- [230] Derosa F, KARVE S, Sarode A. Improved processes of preparing mRNA-loaded lipid nanoparticles. AU2021226578A1, 2022.
- [231] Reichmuth AM, Oberli MA, Jaklenec A, Langer R, Blankschtein D. mRNA vaccine delivery using lipid nanoparticles. *Ther Deliv* 2016;7:319–34. <https://doi.org/10.4155/tde-2016-0006>.
- [232] Maeki M, Uno S, Niwa A, Okada Y, Tokeshi M. Microfluidic technologies and devices for lipid nanoparticle-based RNA delivery. *J Controlled Release* 2022;344:80–96. <https://doi.org/10.1016/j.jconrel.2022.02.017>.
- [233] Brito L, Chen D, GAMBER GG, Geall A, Love K, ZABAWA T, et al. Lipids and lipid compositions for the delivery of active agents. US10426737B2, 2019.
- [234] Leung AKK, Tam YYC, Cullis PR. Chapter Four - Lipid Nanoparticles for Short Interfering RNA Delivery. In: Huang L, Liu D, Wagner E, editors. *Adv. Genet.*, vol. 88, Academic Press; 2014, p. 71–110. <https://doi.org/10.1016/B978-0-12-800148-6.00004-3>.
- [235] Jia X, Liu Y, Wagner AM, Chen M, Zhao Y, Smith KJ, et al. Enabling online determination of the size-dependent RNA content of lipid nanoparticle-based RNA formulations. *J Chromatogr B* 2021;1186:123015. <https://doi.org/10.1016/j.jchromb.2021.123015>.
- [236] Haas H, BACIC T, Schumacher J. Method for determining at least one parameter of a sample composition comprising nucleic acid, such as rna, and optionally particles. US20220381748A1, 2022.
- [237] Cheng F, Wang Y, Bai Y, Liang Z, Mao Q, Liu D, et al. Research Advances on the Stability of mRNA Vaccines. *Viruses* 2023;15:668. <https://doi.org/10.3390/v15030668>.
- [238] Siiskonen M, Malmqvist J, Folestad S. Integrated product and manufacturing system platforms supporting the design of personalized medicines. *J Manuf Syst* 2020;56:281–95. <https://doi.org/10.1016/j.jmsy.2020.06.016>.
- [239] Algorri M, Abernathy MJ, Cauchon NS, Christian TR, Lamm CF, Moore CMV. Re-Envisioning Pharmaceutical Manufacturing: Increasing Agility for Global Patient Access. *J Pharm Sci* 2022;111:593–607. <https://doi.org/10.1016/j.xphs.2021.08.032>.
- [240] Kairuz D, Samudh N, Ely A, Arbuthnot P, Bloom K. Advancing mRNA technologies for therapies and vaccines: An African context. *Front Immunol* 2022;13.
- [241] Paul S, Stang A, Lennartz K, Tenbusch M, Überla K. Selection of a T7 promoter mutant with enhanced in vitro activity by a novel multi-copy bead display approach for in vitro evolution. *Nucleic Acids Res* 2013;41:e29. <https://doi.org/10.1093/nar/gks940>.
- [242] Padilla R, Sousa R. A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Res* 2002;30:e138.
- [243] Extended Upstream A-T Sequence Increases T7 Promoter Strength\* - *Journal of Biological Chemistry* n.d. [https://www.jbc.org/article/S0021-9258\(20\)59037-6/fulltext](https://www.jbc.org/article/S0021-9258(20)59037-6/fulltext) (accessed September 19, 2023).

- [244] Sari Y, Sousa Rosa S, Jeffries J, Marques MPC. Comprehensive evaluation of T7 promoter for enhanced yield and quality in mRNA production. *Sci Rep* 2024;14:9655. <https://doi.org/10.1038/s41598-024-59978-5>.
- [245] Martin C, Cavac E, Malagodapathiranage KH, Sui S, Perry SL, Gholamalipour Y. Novel enzymatic methods to generate high yields of sequence specific rna oligonucleotides with extreme precision. US20230265477A1, 2023.
- [246] Martin C, Cavac E, Malagodapathiranage KH, Sui S, Perry SL, Gholamalipour Y. Enzymatic methods to generate high yields of sequence specific rna with extreme precision. US11578348B2, 2023.
- [247] Vaidyanathan S, Azizian KT, Haque AKMA, Henderson JM, Hendel A, Shore S, et al. Uridine Depletion and Chemical Modification Increase Cas9 mRNA Activity and Reduce Immunogenicity without HPLC Purification. *Mol Ther - Nucleic Acids* 2018;12:530–42. <https://doi.org/10.1016/j.omtn.2018.06.010>.
- [248] Martínez J, Lampaya V, Larraga A, Magallón H, Casabona D. Purification of linearized template plasmid DNA decreases double-stranded RNA formation during IVT reaction. *Front Mol Biosci* 2023;10.
- [249] Kis Z, Kontoravdi C, Shattock R, Shah N. Resources, Production Scales and Time Required for Producing RNA Vaccines for the Global Pandemic Demand. *Vaccines* 2021;9:3. <https://doi.org/10.3390/vaccines9010003>.
- [250] Elich J, Rabideau AE, Shamashkin M, Philpot R, Fritz B, Wojciechowski P. Fed-batch in vitro transcription process. US20220145381A1, 2022.
- [251] Kern JA, Davis RH. Application of a fed-batch system to produce RNA by in vitro transcription. *Biotechnol Prog* 1999;15:174–84. <https://doi.org/10.1021/bp990008g>.
- [252] Choi I, Ahn G-Y, Kim ES, Hwang SH, Park H-J, Yoon S, et al. Microfluidic Bioreactor with Fibrous Micromixers for In Vitro mRNA Transcription. *Nano Lett* 2023;23:7897–905. <https://doi.org/10.1021/acs.nanolett.3c01699>.
- [253] Schmidt A, Helgers H, Vetter FL, Juckers A, Strube J. Fast and Flexible mRNA Vaccine Manufacturing as a Solution to Pandemic Situations by Adopting Chemical Engineering Good Practice—Continuous Autonomous Operation in Stainless Steel Equipment Concepts. *Processes* 2021;9:1874. <https://doi.org/10.3390/pr9111874>.
- [254] Understand Advanced Process Control 2016. <https://www.aisc.org/resources/publications/cep/2016/june/understand-advanced-process-control> (accessed November 13, 2023).
- [255] Schmidt A, Helgers H, Vetter FL, Zobel-Roos S, Hengelbrock A, Strube J. Process Automation and Control Strategy by Quality-by-Design in Total Continuous mRNA Manufacturing Platforms. *Processes* 2022;10:1783. <https://doi.org/10.3390/pr10091783>.
- [256] DeRosa F, Dias A, Heartlein M, Karve S. Methods for Purification of Messenger RNA. US20220389403A1, 2022.
- [257] Pascolo S. Synthetic Messenger RNA-Based Vaccines: From Scorn to Hype. *Viruses* 2021;13:270. <https://doi.org/10.3390/v13020270>.
- [258] COVID E. Vaccine Moderna EPAR public assessment report 2021 [updated 11 March 2021].
- [259] Hengelbrock A, Schmidt A, Strube J. Formulation of Nucleic Acids by Encapsulation in Lipid Nanoparticles for Continuous Production of mRNA. *Processes* 2023;11:1718. <https://doi.org/10.3390/pr11061718>.
- [260] Lamoot A, Lammens J, Lombaerde ED, Zhong Z, Gontsarik M, Chen Y, et al. Successful batch and continuous lyophilization of mRNA LNP formulations depend on cryoprotectants and ionizable lipids. *Biomater Sci* 2023. <https://doi.org/10.1039/D2BM02031A>.

- [261] De A, Ko YT. Why mRNA-ionizable LNPs formulations are so short-lived: causes and way-out. *Expert Opin Drug Deliv* 2023;20:175–87. <https://doi.org/10.1080/17425247.2023.2162876>.
- [262] Kimura N, Maeki M, Sato Y, Ishida A, Tani H, Harashima H, et al. Development of a Microfluidic-Based Post-Treatment Process for Size-Controlled Lipid Nanoparticles and Application to siRNA Delivery. *ACS Appl Mater Interfaces* 2020;12:34011–20. <https://doi.org/10.1021/acsami.0c05489>.
- [263] Curry E, Muir G, Qu J, Kis Z, Hulley M, Brown A. Engineering an Escherichia coli based in vivo mRNA manufacturing platform. *Biotechnol Bioeng* 2024;121:1912–26. <https://doi.org/10.1002/bit.28684>.
- [264] Feng X, Li Z, Su Z, Che S, Dai B, Cheng Y, et al. Rapid and high recovery isolation of mRNA from *in vitro* transcription system by ammonium sulphate precipitation at room temperature. *Sep Purif Technol* 2024;336:126331. <https://doi.org/10.1016/j.seppur.2024.126331>.
- [265] Dewar EA, Guterstam P, Holland D, Lindman S, Lundbäck P, Brito dos Santos S, et al. Improved mRNA affinity chromatography binding capacity and throughput using an oligo-dT immobilized electrospun polymer nanofiber adsorbent. *J Chromatogr A* 2024;1717:464670. <https://doi.org/10.1016/j.chroma.2024.464670>.
- [266] García-Marquina G, Zhang A, Sproviero M, Fang Y, Gardner AF, Robb GB, et al. Streamlined DNA template preparation and co-transcriptional 5' capped RNA synthesis enabled by solid-phase catalysis 2023:2023.10.28.564520. <https://doi.org/10.1101/2023.10.28.564520>.

# Chapter III – Aims and Objectives

---

## Thesis Aim

mRNA is an emerging alternative technology in the vaccine field. Owing to its precision and safety, as well as flexible manufacture, this technology is attractive for multiple applications, from prophylactic and cancer treatments to metabolic and genetic diseases[1]. One of the main advantages of this technology is the reduced time that is required to develop and deliver a new vaccine. In fact, during Covid-19 pandemic event, a mRNA-based vaccines was the first to reach clinical trials[2], and was approved by FDA in less than one year since the first Covid-19 case was detected[3].

A flexible and cost-effective manufacturing platform is required to allow on-demand deliver of mRNA vaccines. mRNA manufacturing process is one of the main advantages of this technology: mRNA is produced in a cell-free enzymatic cascade reactions [4] that allows to achieve grams of product in a matter of hours[5]. The cell-free nature of the process, and independence of the target sequence, allows this platform to be performed in a standardised fashion. Downstream processing consists of a multi-step purification platform that can include a range of operation units, namely, precipitation, enzymatic digestion, chromatography or tangential flow filtration, or a combination of the above[4]. Nevertheless, the manufacturing relies on the use in the use of expensive raw materials, and it is dependent on methodologies that lack scalability and cost-effectiveness to achieve the high purity required for clinical applications.

To enabling the mRNA technology to reach its full potential and meet the market demands is necessary to improve the manufacturing process state-of-art. Continuous manufacturing has the potential to circumvent the current bottlenecks[4]. Nevertheless, with the current process knowledge, scaling can be limits owing to the materials and equipment availability, and lack of knowledge of the manufacturing process itself[6]. Deliver an efficient manufacturing and cost-effective manufacturing process using that uses the existing resources requires an optimisation of all the process steps.

The aim of this thesis (Figure 1) is to design and develop flexible and cost-effective operation units that can improve the mRNA manufacturing platform in order to maximise product production yield, while improve mRNA quality, and that can be easily adapted to different process modalities, namely continuous manufacturing.

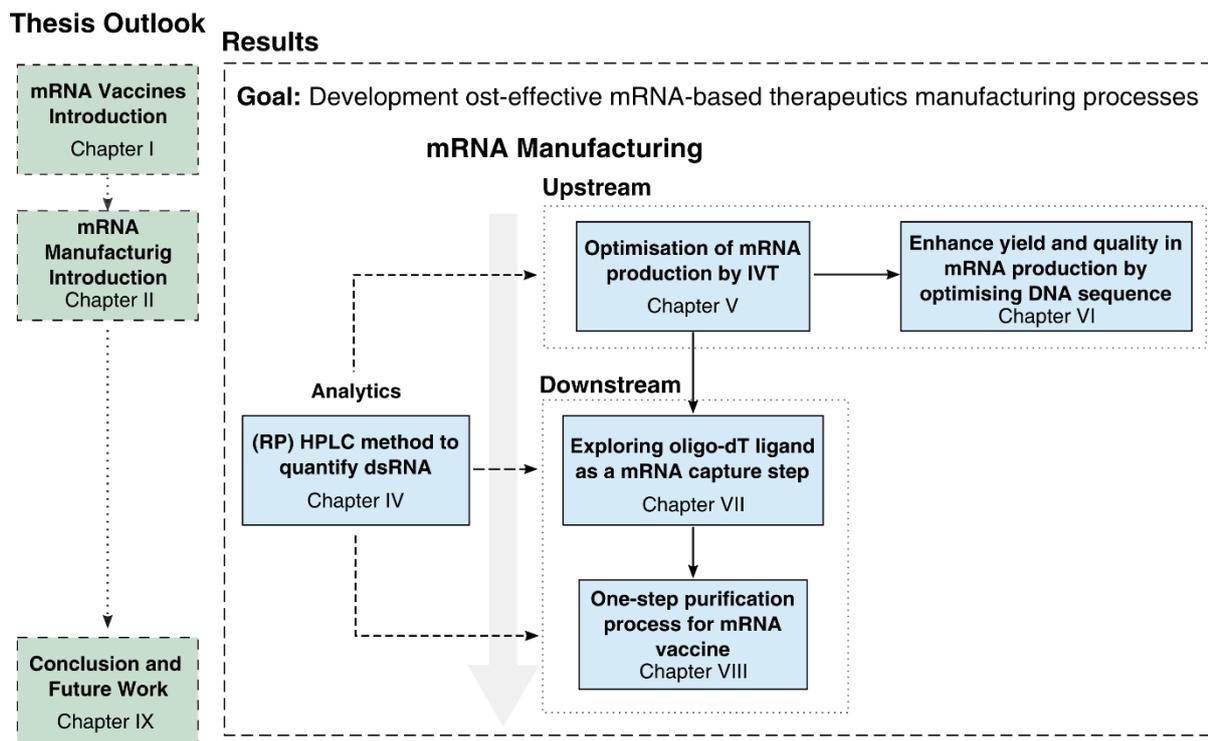


Figure 2. Schematic representation of thesis outlook, goals and chapters

## Chapters Outlook

One of the first challenges when implementing a manufacturing process is correctly characterise the product and its impurities. Analytical techniques should be precise and reliable to rigorously characterise the product and/or impurities. Additionally, minimal sample processing and handling, the ability adapt the method to automation, are pivotal characteristics in an analytical technique. To quantify dsRNA in the different stages of the manufacturing process, a reverse phase (RP)HPLC method was implemented. Chapter IV describes a reverse-phase HPLC method can be used to quantify total mRNA coupled with an enzymatic digestion, to allow direct quantification dsRNA. This method performance was compared with well-established dsRNA detection immunoassay techniques. HPLC showed to precise, and less prone to operational errors, with minimal variance in quantification.

mRNA is *in vitro* transcribed (IVT) from a DNA template catalysed by an RNA polymerase. This well-defined can easily be controlled, which makes this process ideal for optimisation. An objective of this work was to maximise mRNA production yield. This was achieved by optimising the IVT reaction conditions and analysing mRNA production using the analytical methodology established in chapter IV. In chapter V, a data-driven optimisation method - Bayesian optimization – was used to achieve optimal mRNA production of 12 g L<sup>-1</sup> total approximately 2 h. The results obtained correspond to a IVT yield increase of a factor of two in half of the time when comparing with industry standards and data reported in literature.

Further optimisation was performed by exploring the impact of the DNA template on the reaction. Increasing production rates can lead to the production of by-products. A cost-effective process allows to deliver yield products with reduced impurities in order to avoid an intensive purification tray. In chapter VI, the effect on mRNA production of the sequence at the T7 promoter region is evaluated. In particular, AT-rich sequences were inserted in the downstream region of the T7 promoter. This allowed for a notable increase in mRNA titres compared to the wildtype T7 promoter, and to reduce in 30% the production by-products, namely dsRNA.

A flexible and cost-effective manufacturing process requires a well-established downstream processing tray. This a major bottleneck in the mRNA manufacturing process. One objective of this work was to simplify the downstream tray by optimising the existing purification methods, as well as use new separation technologies to deliver high quality mRNA at a high yield. One of the most popular mRNA purification approaches is the use of affinity chromatography. In this case, a deoxythymidine (dT) sequence ligand has affinity towards the poly-A tail present in the mRNA. Optimisation of this step is critical for the successful implementation of this method without resorting to multiple purification steps. In Chapter VII, the Machine Learning algorithm was employed to optimise the dynamic binding capacity (DBC). Using this model, an increase of 7.5-fold from the initial conditions was obtained, achieving  $1.8 \text{ mg mL}^{-1}$  resin, in only 20 runs. Additionally, the behaviour of the major impurities, namely DNA template and dsRNA, were evaluated throughout the chromatographic separation. This study showed that affinity chromatography can be used as a capture step after IVT.

Achieve complete separation of the mRNA from its impurities in a one-step purification process is the main goal of the optimisation of the downstream process. In Chapter VIII Nuvia aPrime, a multimodal resin that combines strong anion exchanger with a phenyl group, to purify IVT samples, was explored to achieve a high yield and yield quality product. Optimisation of the binding conditions allowed to flow through the main process and product-related impurities, namely NTPs, DNA template and dsRNA, while maximising the mRNA binding. This one-step process achieved a mRNA recovery yield of  $81 \pm 5\%$ , with a purity of  $88 \pm 2\%$  with no detectable concentration of DNA and a reduction on dsRNA of  $65\%$ .

The knowledge developed throughout this work contribute to state-of-art of mRNA vaccines manufacturing and can be applied to the development sustainable, flexible and cost-effective manufacturing process.

This thesis was developed in the framework of the Biotechnology and Biosciences PhD program of Instituto Superior Técnico (IST) (Lisbon, Portugal), and funded by Fundação para a Ciência e a Tecnologia (FCT) [SFRH/BD/148437/2019]. The research work involved a in a collaboration between the Institute for Biotechnology and Bioengineering in Instituto Superior

Técnico (Lisbon, Portugal), and the University College London, Department of Biochemical Engineering. During the development of the project, a number of collaborations with different experts were placed, namely with the LASIGE Computer Science and Engineering Research Centre at Faculdade de Ciências de Universidade de Lisboa.

## References

- [1] Wang Y-S, Kumari M, Chen G-H, Hong M-H, Yuan JP-Y, Tsai J-L, et al. mRNA-based vaccines and therapeutics: an in-depth survey of current and upcoming clinical applications. *J Biomed Sci* 2023;30:1–35. <https://doi.org/10.1186/s12929-023-00977-5>.
- [2] Le TT, Andreadakis Z, Kumar A, Román RG, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. *Nat Rev Drug Discov* 2020;19:305–6. <https://doi.org/10.1038/d41573-020-00073-5>.
- [3] Verbeke R, Lentacker I, De Smedt SC, Dewitte H. The dawn of mRNA vaccines: The COVID-19 case. *J Controlled Release* 2021;333:511–20. <https://doi.org/10.1016/j.jconrel.2021.03.043>.
- [4] Rosa SS, Prazeres DMF, Azevedo AM, Marques MPC. mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine* 2021;39:2190–200. <https://doi.org/10.1016/j.vaccine.2021.03.038>.
- [5] Baronti L, Karlsson H, Marušič M, Petzold K. A guide to large-scale RNA sample preparation. *Anal Bioanal Chem* 2018;410:3239–52. <https://doi.org/10.1007/s00216-018-0943-8>.
- [6] Mishra RS, Kumar R, Dhingra S, Sengupta S, Sharma T, Gautam GD. Adaptive grey model (AGM) approach for judgemental forecasting in short-term manufacturing demand. *Mater Today Proc* 2022. <https://doi.org/10.1016/j.matpr.2021.12.531>.

# Chapter IV – (RP)UHPLC/UV analytical method to quantify dsRNA during mRNA vaccine manufacturing process

---

**Key words:** Analytical techniques; HPLC; dsRNA; quantification

## Introduction

There is widespread interest in mRNA vaccines within both academia and industry since the COVID-19 pandemic. Today, there are over 60 ongoing clinical trials that use mRNA vaccines for a variety of treatments that include prophylactic and cancer treatments, protein replacement or gene editing[1]. When compared with traditional vaccine platforms, mRNA vaccines present a number of advantages, namely the fast production, the safety profile, or the vaccine effectiveness. From a clinical and manufacturing point-of-view, consistency of high-quality products is required with little batch-to-batch variability. This can only be achieved with a precise product characterisation and a tight manufacturing process control. Therefore, analytical procedures that rigorously characterise the mRNA and impurities (process- and product-related impurities)[2] throughout the manufacturing process are required[3].

mRNA vaccines are usually produced in a cell-free system where a linearised DNA strand is *in vitro* transcribed (IVT) into a mRNA strand. RNA polymerases (e.g. T7 RNA polymerase) transcribe the DNA with nucleotides as co-substrates to produce grams per litre of mRNA[4–7]. Other reaction components include co-factors, enhancers and additional enzymes, with temperature and pH as critical reaction parameters. The exploration of different reaction conditions coupled with a tight production characterisation[8] led already an improvement in reaction yields, with reported production in the range of 12 g/L[7,9]. However, despite the tight control, there are impurities produced throughout the IVT process. These can be classified into process related (enzymes, residual NTPs, or DNA template) and product related impurities (malformed mRNAs)[2,7]. During the IVT reaction, the T7 RNA polymerase can release truncated mRNA molecules, or produce complementary RNA strands that can hybridise and produce double-stranded mRNA (dsRNA). The presence of this particular impurity in the final product must be avoided since it can impact translation and trigger a strong immune response[10], which ultimately can lead to an uncontrolled immune-inflammatory reaction[11]. In the absence of dsRNA, protein expression within cells can be increased by 10-1000 folds[12]. Therefore, optimised reaction conditions or new purification methods are necessary to eliminate dsRNA as well as precise analytical procedures to quantify this impurity. Nevertheless, there is still a lack of a defined concentration limits that are established by the regulatory agencies for dsRNA[3].

From a reaction optimisation perspective, multiple approaches have been followed to reduce dsRNA such as engineering T7 RNA polymerase[13,14], or blocking the 3' end with complementary oligonucleotides to avoid overextension[15]. New chromatographic modalities can be applied, exploring physico-chemical differences between mRNA and dsRNA [16]. Nonetheless, dsRNA needs to be carefully monitored during the manufacturing process itself and in the final product[3]. Analytically, dsRNA can be detected and characterised by several

methods, including electrophoresis[17,18], immunoassays that use anti-double-stranded RNA antibodies such as dot blot, ELISA or lateral flow strip assay (LFSA)[19]; asymmetric flow field flow fractionation (A4F)[20], or even chromatographic methods[8,21–23]. Nevertheless, there is a lack of well-established methods that can be applied in the manufacturing process which will quantify dsRNA that meet the regulatory requirements [24], are quick, specific and minimise impurities interference, can be easily adapted to the different process stages, from the production to the different purification steps, and to the different modalities of the mRNA process (batch, fed-batch, or continuous).

In this communication, we explore a method to quantify dsRNA in the different stages of the manufacturing process. A previously established reverse-phase HPLC method that quantifies total mRNA was coupled to an enzymatic digestion step that digests ssRNA, allowing to directly measure dsRNA. The developed RP-HPLC method achieved lower limit of detection (LOD) and limit of quantification (LOQ),  $1.41 \times 10^{-2} \pm 1.78 \times 10^{-3} \text{ g L}^{-1}$ , and  $4.27 \times 10^{-2} \pm 5.4 \times 10^{-3} \text{ g L}^{-1}$ , respectively, compared to the current golden standard, the dot blot. Furthermore, the decision limit ( $CC\alpha$ ,  $9.95 \times 10^{-2} \pm 1.26 \times 10^{-3} \text{ g L}^{-1}$  and detection capability ( $CC\beta$ ,  $1.70 \times 10^{-2} \pm 2.14 \times 10^{-3} \text{ g L}^{-1}$ ) are low, and obtained for single sample under 30 min. When compared with standard methods (e.g. ELISA and dot blot assay[3]. HPLC outperformed these and proved to be precise, and less prone to operational errors, even in spiking studies with process-related impurities (NTPs and DNA) where minimal variance in quantification is observed. The implemented RP-HPLC is robust to use during mRNA manufacturing process and that can be adapted for Process Analytical Technology (PAT) purposes in the future

## Materials and methods

Unless otherwise stated, all chemicals and reagents were purchased from ThermoFisher Scientific (UK).

### Template plasmid DNA production

Template design and plasmid production was performed as previously described(7). Briefly, GFP gene (GenBank Accession #AAB02572.1) is flanked by 5'UTR containing the T7 RNA polymerase promoter, eukaryotic translation factor binding site and a Kozak consensus sequence[25], and by a 3'UTR composed by two  $\beta$ -globin tandem repeats and followed by a 120 bp poly-A tail segmented with a 6 bp [26]. A pUC7 containing a kanamycin resistance is used as a plasmid vector with plasmid propagation performed in *E. coli* NEB 10-beta (New England Biolabs, UK). The pDNA is obtained by performing an overnight culture in LB media at 37 °C and purified using GeneJET Plasmid Miniprep Kit.

### **Template production by touchdown polymerase chain reaction**

The forward and reverse primers were previously described[7]. A T7 RNA polymerase promoter sequence was added to the 5' end of the complementary strand using the reverse primer. DNA template for mRNA IVT is produced by touchdown PCR (Applied Biosystems™ Veriti™ 96-Well Thermal Cycler, ThermoFisher Scientific, UK). Briefly, the reaction mixture comprises 250 ng mL<sup>-1</sup> of template plasmid, 0.4 μM of forward and reverse primer, 1x VeriFi™ Buffer, 1x VeriMax Enhancer, and 0.02 U μL<sup>-1</sup> high-fidelity VeriFi™ DNA Polymerase (PCR Biosystems, UK). The PCR conditions are for the denaturation step 98°C for 30 s, 20 cycles with a denaturation step at 94°C for 15 s, a annealing step at 65-55°C for 30s and extension step at 72°C for 45 s. This is followed by 20 cycles of an annealing step at 55°C for 30 s and an extension step at 72°C. The final extension is performed at 72°C for 2 min. The obtained PCR product is purified and concentrated using a GeneJET PCR Purification Kit. Final concentrations are quantified on a NanoDrop™ One Microvolume UV-Vis Spectrophotometer (ThermoFisher Scientific, UK).

### **dsRNA *in vitro* transcription reactions**

dsRNA is produced in an IVT reaction using previously described reaction conditions[7]. Briefly, 89 nM of DNA template is mixed with 7.7 mM NTPs, 5.3 mM dithiothreitol, 50 mM magnesium acetate, 40 mM Tris-HCl pH 6.8, 2.3 mM spermidine, 8 U mL<sup>-1</sup> inorganic pyrophosphatase and 7750 U mL<sup>-1</sup> of T7 RNA polymerase. To this, 1500 U mL<sup>-1</sup> Rnase inhibitor is added to the mix to avoid degradation. The mixture is incubated at 43 °C for two hours on an Applied Biosystems™ Veriti™ 96-Well Thermal Cycler (ThermoFisher Scientific, UK). After incubation, the produced strands of RNA are annealed to form dsRNA by diluting the reaction mixture twice with WFI and incubating in a decreasing temperature gradient temperature (from 85°C to 30°C) with a 2 min step at each temperature.

### **dsRNA purification**

The DNA template and the remaining mRNA are digested using TURBO™ DNase, and RNase T1, respectively. The enzymes are added to a concentration of 0.04 U μL<sup>-1</sup> (TURBO™ DNase), and 20 U μL<sup>-1</sup> (RNase T1). The samples are incubated at 37°C for 30 min and purified afterwards using a MEGAclear™ Transcription Clean-Up kit. After purification, the dsRNA is precipitated overnight at -20°C by adding 500 mM pH 5 ammonium acetate and 2.5 volumes of ethanol. The samples are centrifuged at 15000 x g for 15 min and supernatant discarded. The pellet is air dried and WFI water is added to resuspend the pellet to the desired final concentration. Concentration is determined on a NanoDrop™ One Microvolume UV-Vis Spectrophotometer (ThermoFisher Scientific, UK) using a conversion factor of 47 μg. mL<sup>-1</sup> . Abs<sub>260</sub><sup>-1</sup> [27].

## **Analytical methodologies**

### **Gel electrophoresis**

Samples obtained from the IVT were digested with T1 RNAse as previously described and analysed by gel electrophoresis [7]. Briefly, a 2% (w/v) agarose gel was prepared with 0.5× TBE buffer with 5.5 mM of magnesium chloride and prestained with SYBR® Safe DNA Gel Stain and run at 100 V for one hour. 1 kb Plus DNA Ladder (New England Biolabs, UK) was used for analysis.

The samples were blotted onto 0.45 µm nitrocellulose membrane using a pipette tip. The membrane was dried and incubated with 3% (w/v) of BSA (Sigma-Aldrich, USA) in 1× PBS-Tween buffer, containing 1× PBS (10 mM sodium phosphates, 2.68 mM potassium chloride, and 140 mM calcium chloride) and 0.05% (v/v) of Tween-20 (Sigma-Aldrich, USA), for an hour at 22 °C. The membrane was subsequently incubated for an hour with 1:1500 dilution of SCICONS™ J2 mouse anti-dsRNA IgG2a monoclonal antibody (Nordic-MUbio, Netherlands) as primary antibody using 3% (w/v) BSA in 1x PBS-Tween buffer. The membrane was washed by incubating for 5 min in the PBS-Tween buffer and repeated three times. Afterwards, the membrane was incubated with 1:2000 dilution of secondary antibody, mouse IgG horseradish peroxidase (HRP)-conjugated antibody (R&D Systems, UK) for an hour, followed by three washing steps. Prior to the chemiluminescence exposure, the membrane was treated with Pierce™ ECL Western Blotting Substrate and incubated for 3 min. All the incubation steps were performed at 22 °C. The visualisation was performed through chemiluminescence exposure using an Amersham™ Imager 600 (GE Healthcare, UK).

### **Reverse-phase high performance liquid chromatography**

HPLC analysis(7,28) was performed on an UltiMate 3000 UHPLC System (Thermo Fisher Scientific, UK) equipped with a VWD-3400 RS Rapid Detector. All species were analysed on a RP-DNApac column (2.1 × 100 nm, Thermo Fisher Scientific, UK) at 80 °C with detection at 260 nm. The column was pre-equilibrated with TAE buffer (100 mM Tris acetate, pH 7, 2.5 mM EDTA), with initial flow rate set to 0.2 mL min<sup>-1</sup>. After a 1 min washing step, the flow rate was increased to 0.35 mL.min<sup>-1</sup>, at a gradient of 0.25 m.min<sup>-1</sup> gradient over 30 sec. A first elution gradient is performed to 6% of the elution buffer (TAE buffer, 25% acetonitrile) for 30 s at 0.35 mL. min<sup>-1</sup>, followed by a gradient of 0.4 mL min<sup>-1</sup> over 4 min until 76.5% elution buffer is reached, finalising with gradient to 100% elution buffer for 1 min. The column is washed with the elution buffer for 3 min and re-equilibrated with TAE buffer for 6 min at 0.4 mL min<sup>-1</sup>.

### **Enzyme-linked immunosorbent assay**

The SCICONS mouse dsRNA ELISA kit (J2 based) ELISA (Nordic MUBio, NL) was used to detect dsRNA. Briefly, 96 well plates were treated with anti-dsRNA coating antibody diluted in

PBS and incubated overnight at 4°C. The plates were then incubated with 1%BSA in PBS for 1h at 37°C and were washed with PBS-T (1× PBS with 0.5% Tween 20) three times. 100 uL of samples were added to each well and incubated for 1 hour at 37 °C. Afterwards, the plates were washed 4 times with PBS-T and incubated with the anti-dsRNA antibody at 37°C for 1 hour. After another washing step, HRP-conjugated goat-anti mouse secondary antibody diluted in PBS with 1% BSA is added and incubated at 37°C for 1 hour. The plates are washed a final time, and TMB substrate is added. The plates are incubated between 15 to 30 min at 22 °C. The reaction is stopped by adding H<sub>2</sub>SO<sub>4</sub> to a final concentration of 1M. The absorbance is read at 450 nm.

### Spiking Studies

Spiking studies were performed by mixing dsRNA with DNA or NTPs (Table 1). RNA without any spiking components, isolated NTPs and DNA and the mixture of both impurities were used as controls. The dsRNA samples were quantified using HPLC and dot blot methods.

Table 2 Concentration of dsRNA DNA and NTPs concentrations used for the spiking studies.

Species	Concentration	
	min.	Max.
RNA (g L <sup>-1</sup> )	0.1	0.4
DNA (nM)	10	90
NTPs (mM)	2	7.5

### Data and statistical analysis

All the statistical analysis was performed in R. At least three independent experiments and three dependent measures were performed for each experiment, and all the data is results are presented as mean ± standard deviation. One-way ANOVA followed by Tukey’s test were used for analysis. A p value < 0.05 was considered to indicate statistical significance. Normality plot of the residuals can be found in Supplemental Information Figures S1 and S2, for the calibration curves and dsRNA spiking studies, respectively. The Shapiro-Wilk test of the ANOVA residuals was performed, and no violation of normality was detected.

### Range validation

The limits of detection (LOD) and quantification (LOQ) were calculated based on the standard error of the intercept ( $\sigma$ ) and the slope (S) of the calibration curves at a signal-to-noise ratio of 3.3 (LOD) and 10 (LOQ) (29), according to:

$$\text{LOD}=3.3\sigma/S \quad (1)$$

$$\text{LOQ}=10\sigma/S \quad (2)$$

where  $S$  and  $\sigma$  are the slope and standard deviation of the response, respectively. The decision limit ( $CC\alpha$ ) and decision capability ( $CC\beta$ ) were calculated considering a 2.33 factor, which corresponds to 1% of false positive risk, and a 1.64 factor, which corresponds to a 5% false negative risk with regards to  $CC\alpha$ [24,30], according to

$$CC\alpha = 2.33\sigma/S \quad (3)$$

$$CC\beta = CC\alpha + 1.64\sigma/S \quad (4)$$

## Results and Discussion

We have explored reverse phase chromatography (RP-HPLC) for the detection of dsRNA based on the physico-chemical properties of mRNA, namely the hydrophobicity[31]. Preparative chromatography, namely ion-pair (IP-RP-HPLC), has already been used to separate ssRNA from dsRNA[12,32]. The addition of ion-pair to the mobile phase will interact with the negatively charged backbone of oligonucleotides and will allow separation according to the number of backbones available(33). In this work, the mobile phase is composed of Tris buffer and EDTA, instead of an ion-pair and thus the separation is achieved by the hydrophobic portions present in the mRNA (e.g poly(A) tail) [28]. To facilitate detection, it is necessary to ensure that solely dsRNA is present in samples. To achieve this, an enzymatic digestion prior to injection is performed using RNase T1[7,13]. The RNase T1 is an endoribonuclease that cleaves single-stranded RNA at the guanosine residues. When added to an IVT reaction, Rnase T1 will digest all the ssRNA, and the dsRNA can be directly quantified by RNA quantification methods. The semi-quantitative assays such as the dot blot or ELISA can be performed without the use of this additional enzymatic digestion since they use specific antibodies (eg. dsRNA antibody j2)[34]. However, with these assays, no direct read out of the samples is possible and often require multiple assay preparation steps, which can potentially increase the associated operator error.

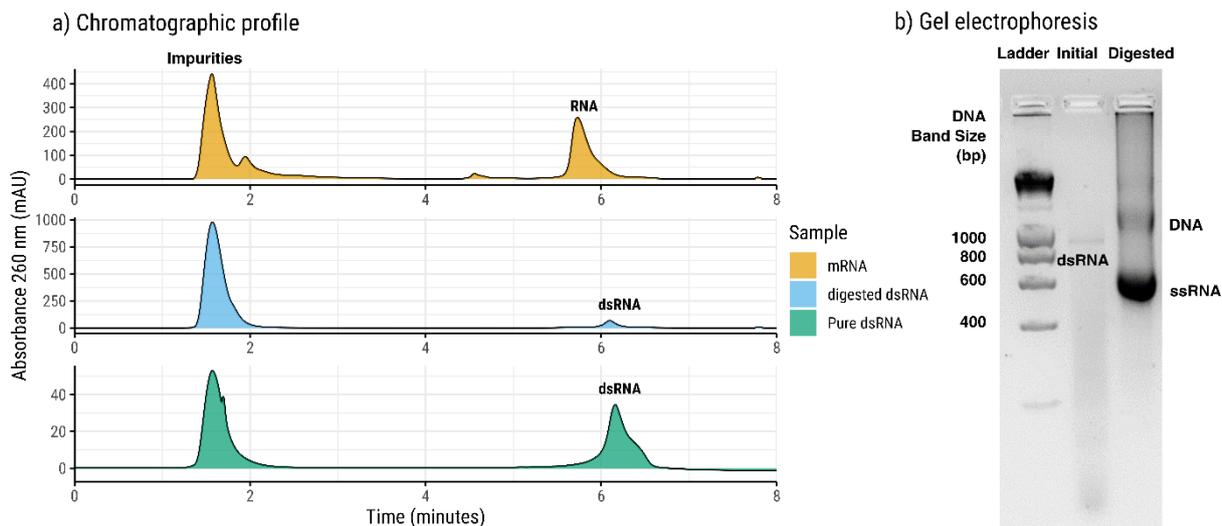


Figure 3. Analysis of dsRNA by RP-HPLC and corresponding gel electrophoresis of the samples. a) Chromatographic profile of total mRNA (yellow) and dsRNA before (blue) and after purification (green) obtained by the developed RP-HPLC method. b) Agarose gel electrophoresis of the total mRNA (initial) and after digestion with RNase T1 (digested).

The RP-HPLC separates molecules according to its hydrophobicity (Figure 1a). Small size impurities (e.g. reaction components) do not usually bind to the stationary phase and are eluted during the wash phase (first peak in the chromatograms), while mRNA is eluted during the gradient phase. This is achieved by increasing the organic solvent concentration in the mobile phase. After digestion with the RNase T1 enzyme, the digested ssRNA is eluted during the wash phase while the dsRNA is bound to the stationary phase and requires organic solvent to be eluted. After digestion, there is a shift in retention time of the RNA peak. This can be attributed to the difference in the molecular weight between dsRNA and ssRNA (Figure 1b). The dsRNA produced during IVT will have single stranded regions, possible interacting with the stationary phase, while after digestion, only double strand regions are present.

We have validated the HPLC method by analysing the different regulatory requirements (e.g. specificity, range, accuracy and precision, and/or robustness [24]) and compare it to dot blot and commercially available dsRNA detection ELISA kits (Figure 2). To achieve this, 6-point calibration curves were prepared with dsRNA concentrations ranging from 0.025 to 1 g L<sup>-1</sup>, 0.05 to 1 g L<sup>-1</sup> and 0.005 to 0.075 mg L<sup>-1</sup>, for the HPLC, dot blot and ELISA, respectively. The standard curves were evaluated using three independent samples. In the HPLC method, linearity is observed throughout the range of sample concentration used with a high correlation coefficient,  $R^2 > 0.99$  (Figure 2a). In contrast, the dot blot only presents linear relationship at concentrations below 0.8 g L<sup>-1</sup> with  $R^2 > 0.98$  (Figure 2b). However, the residuals evaluation data show that there is homoscedasticity (Supplementary Information, Figure S1.b), which means that although the  $R^2$  is lower, the variance between samples is similar. The ELISA method is most specific of the methods evaluated, as it can detect the smallest concentration of dsRNA. However, this assay has a very narrow range of detection. Linearity is only achieved at concentration below 0.07 mg

L<sup>-1</sup> (Figure 2c) with a low correlation coefficient,  $R^2 > 0.95$ . The lower correlation can be attributed the high sensitivity of the method which makes the process more prone to operator errors. The residuals analysis (Supplementary Information, Figure S1) and the respective model coefficient (p-value) evaluation show that there is a strong correlation between the variables for the three methods, confirming the goodness of fit of the resulting models.

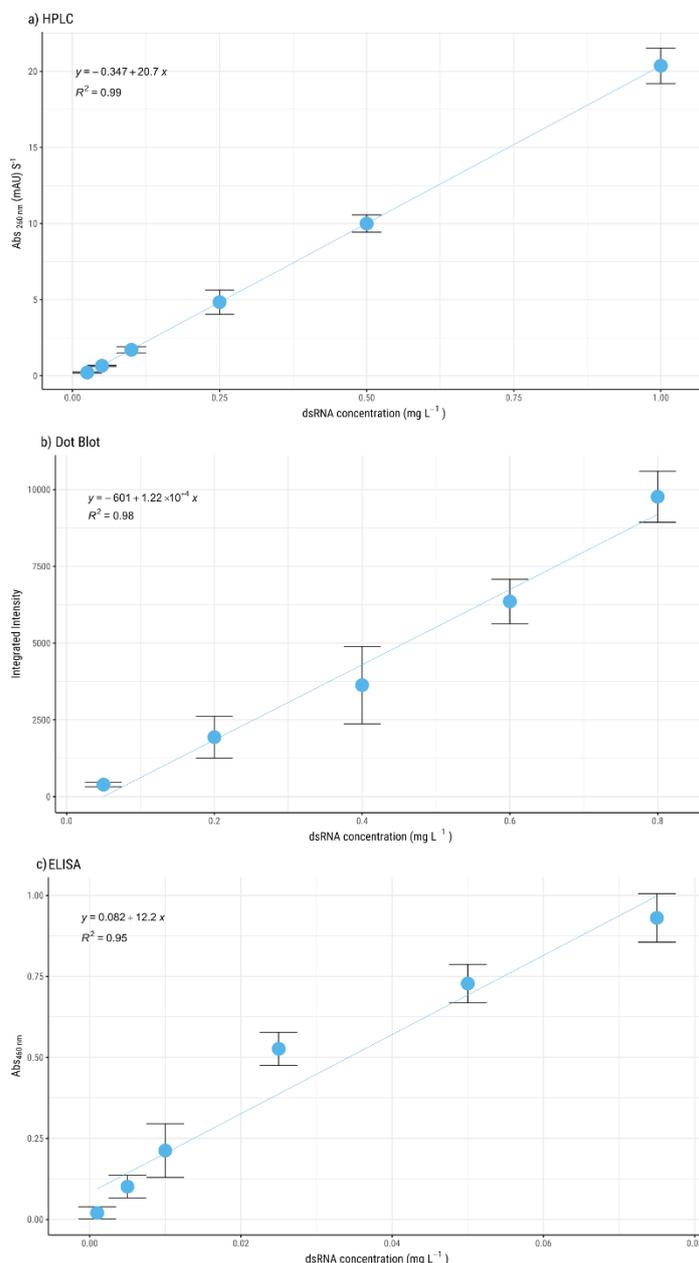


Figure 2. dsRNA calibration curves and corresponding linear regression obtained for the analytical method used: HPLC (a), Dot Blot (b) and ELISA (c). Three independent linear regression analyses were performed, and 3 independent samples were analysed.

The limit of detection (LOD) and limit of quantification (LOQ) were calculated for the three methods (Table 2). ELISA presents the lowest LOD and LOQ,  $1.51 \times 10^{-5} \pm 3.21 \times 10^{-6}$  and  $4.56 \times 10^{-5} \pm 9.73 \times 10^{-6}$  g L<sup>-1</sup>, respectively. The dot blot presents the highest LOD ( $0.201 \pm 0.06$  g

L<sup>-1</sup>) and LOQ (0.609 ± 0.181 g L<sup>-1</sup>) whilst the HPLC presents 14 times lower LOD and LOQ (0.014 ± 0.002, 0.043 ± 0.005 g L<sup>-1</sup>), when compared with dot blot assay. The decision limit (CC $\alpha$ ) and decision capability (CC $\beta$ ) were evaluated for all methods. CC $\alpha$  represents the lowest concentration obtained which is reliable (1% of false positive risk), while CC $\beta$  is the lowest concentration possible to measure with an error probability of 5% (false negative risk).

ELISA assay presents the lowest CC $\alpha$  and CC $\beta$  values (1.06x10<sup>-5</sup> ± 2.27x10<sup>-6</sup> and 1.81x 10<sup>-5</sup> ± 3.86x10<sup>-6</sup> g L<sup>-1</sup>) while in contrast, dot blot presents the highest values (0.142 ± 0.042 and 0.242 ±0.07 g L<sup>-1</sup>). Since ELISA presents the lowest LOD and LOQ evaluated, and it can be an ideal to use with highly pure samples with residual amounts of impurities, namely dsRNA. Currently approved vaccines control the dsRNA levels to be as low as possible throughout the manufacturing process[35,36] with no defined value. Currently, the dot blot assay is the method recommended to characterise dsRNA throughout the mRNA manufacturing process [3] and evaluation values obtained support this. Nevertheless, the developed HPLC method shows a larger detection range and limits, presenting itself as a more versatile quantification method. Additionally, these ranges can be extended as the maximum concentration of dsRNA that can be quantified by HPLC is contingent on the saturation capacity of the column itself, which can be adapted by selecting different column's diameters and lengths.

Table 2. Range characterisation of the three dsRNA quantification methods evaluated (HPLC, dot blot and ELISA) in terms of limit of detection (LOD), limit of quantification (LOQ), decision limit (CC $\alpha$ ) and the detection capability (CC $\beta$ ).

Parameter	Analytical Method					
	HPLC		Dot Blot		ELISA	
	$\bar{x}$ (g L <sup>-1</sup> )	$\sigma$ (g L <sup>-1</sup> )	$\bar{x}$ (g L <sup>-1</sup> )	$\sigma$ (g L <sup>-1</sup> )	$\bar{x}$ (g L <sup>-1</sup> )	$\sigma$ (g L <sup>-1</sup> )
LOD	1.41 x 10 <sup>-2</sup>	1.78 x 10 <sup>-3</sup>	2.01 x 10 <sup>-1</sup>	5.96 x 10 <sup>-2</sup>	1.51 x 10 <sup>-5</sup>	3.21 x 10 <sup>-6</sup>
LOQ	4.27 x 10 <sup>-2</sup>	5.40 x 10 <sup>-3</sup>	6.09 x 10 <sup>-1</sup>	1.81 x 10 <sup>-1</sup>	4.56 x 10 <sup>-5</sup>	9.73 x 10 <sup>-6</sup>
CC $\alpha$	9.95 x 10 <sup>-3</sup>	1.26 x 10 <sup>-3</sup>	1.42 x 10 <sup>-1</sup>	4.21 x 10 <sup>-2</sup>	1.06 x 10 <sup>-5</sup>	2.27 x 10 <sup>-6</sup>
CC $\beta$	1.70 x 10 <sup>-2</sup>	2.14 x 10 <sup>-3</sup>	2.42 x 10 <sup>-1</sup>	7.17 x 10 <sup>-2</sup>	1.81 x 10 <sup>-5</sup>	3.86 x 10 <sup>-6</sup>

To evaluate further the accuracy and precision of the HPLC method implemented, a spiking study was performed. Pure dsRNA samples were spiked with two process related impurities that typically can be present in process samples, namely NTPs and DNA template (Table 1). The concentration of the impurities tested are in the range commonly encountered in the manufacturing process of mRNA vaccines. Additional controls with the spiking reagents (DNA and NTPs) and their mixtures with the different concentration evaluated were also performed. The HPLC precision performance was compared with dot blot assay as similar range can be used in both methods. HPLC analysis shows that the values measured have a more uniform distribution when compared with the dot blot assay (Figure. 3). Statistical analysis performed show that a significant difference is observed in the dot blot performed with a high concentration

of dsRNA (Figure 3d,  $0.4 \text{ g L}^{-1}$ ), and in the presence of a high concentration of NTPs (Supplementary Information, Table S1-S4). This may be due to the occurrence of unspecific binding of the antibody used to perform the dot blot assay.

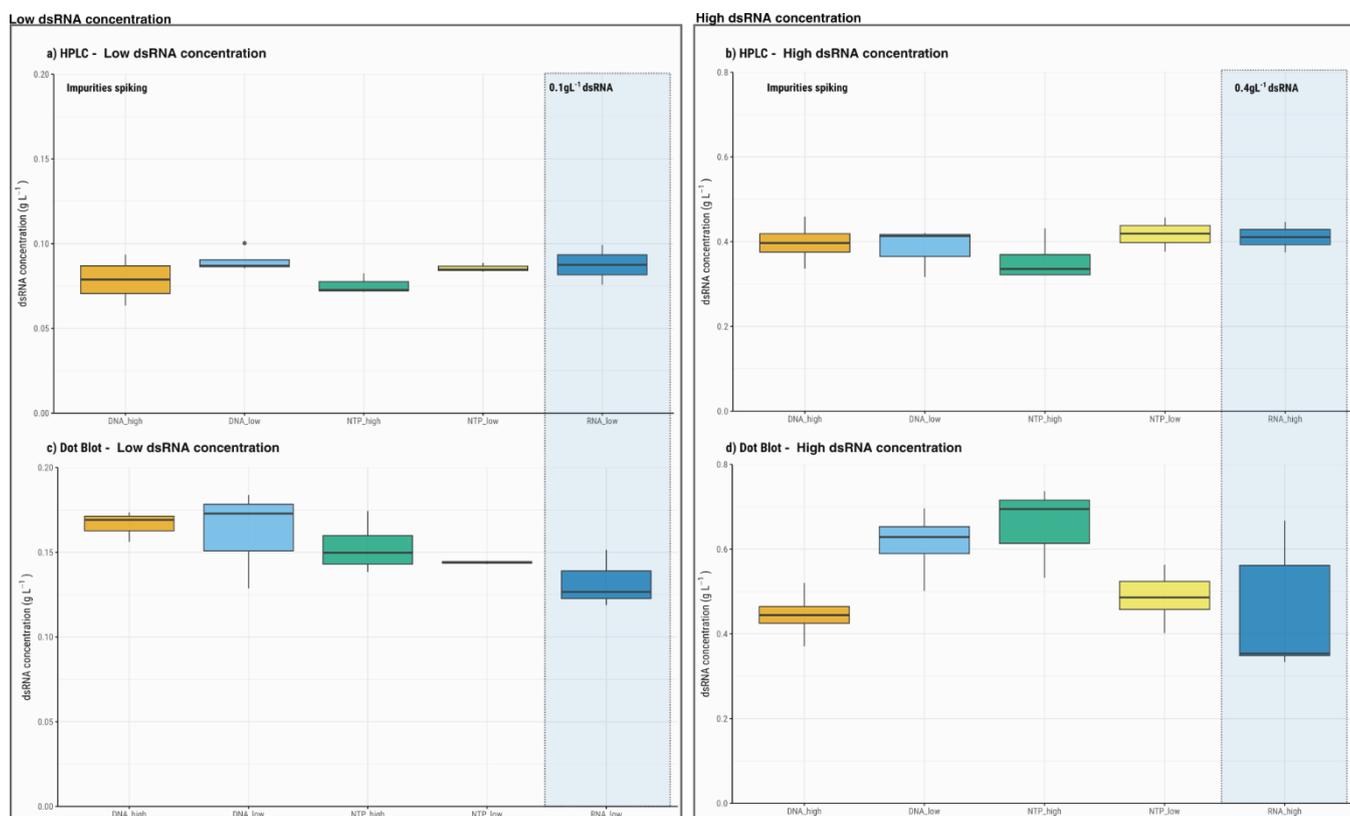


Figure 4 Spiking studies and respective distribution of dsRNA concentration obtained using HPLC and dot blot. HPLC results using  $0.1 \text{ g L}^{-1}$  (a) and  $0.4 \text{ g L}^{-1}$  of dsRNA (b). Dot Blot results using  $0.4 \text{ g L}^{-1}$  (c) and  $0.2 \text{ g L}^{-1}$  of dsRNA (d). DNA\_high and DNA\_low corresponds to a spiking with 90 and 10 nM of DNA template, respectively. NTP\_low and NTP\_high corresponds to 2 and 7.5 mM, respectively. Blue shades areas correspond to the samples without spiking. At least 3 independent samples were analysed for each condition

No significant differences in the dsRNA concentrations were found when using HPLC, indicating that this the method is more accurate (Supplementary Information, Table S1-S4). Comparing the concentrations obtained for high and low range of dsRNA concentrations, the results show that there are no significant differences between both methods. For high concentration of dsRNA ( $0.4 \text{ g L}^{-1}$ ), HPLC measured  $0.41 \pm 0.05 \text{ g L}^{-1}$ , while with dot blot assay, a concentration of  $0.45 \pm 0.151 \text{ g L}^{-1}$  was obtained. For low dsRNA concentrations ( $0.1 \text{ g L}^{-1}$ ),  $0.09 \pm 0.016 \text{ g L}^{-1}$  and  $0.13 \pm 0.017 \text{ g L}^{-1}$  were obtained for the HPLC and dot blot assay, respectively. Overall, the HPLC method presents a better precision and range, as it presented a better correlation between the concentration range ( $0.025$  to  $1 \text{ g.L}^{-1}$ ) evaluated when compared with Dot Blot assay, but ELISA presents the highest smallest range ( $0.005$  to  $0.075 \text{ mg L}^{-1}$ ).

Nevertheless, when comparing analytical methods, it is also required to look into the overall analytical performance (Table 3). Although ELISA is the method that can accurately measure the smallest concentration of dsRNA present in samples, the results are solely obtained after 12 hours due to the multiple incubation steps involved. Additionally, the multiple steps involved in this assay also increases operator error occurrence. In terms of time-to-result, the dot blot is the faster method among the immunoassays (three hours). Nonetheless, as the ELISA, it requires multiple incubations steps and reagents additions. The HPLC is the fastest for single sample, requiring minimum sample preparation and operator input to run the method. However, process time can increase if an additional digestion step with RNase T1 is required. Nevertheless, throughput is limited to HPLC sampler capacity and method running time. Higher throughput are obtained with the ELISA and dot blot. On the other hand, the HPLC method is a fully quantitative method while dot blot relies on a semi quantitative analysis based on densitometry measurements.

Table 3. Summary of the method validation and performance overview of the three analytical methods evaluated for the measurement of dsRNA concentration in a single sample.

	Method validation			Performance overview			
	Sensibility	Range	Precision	Assay Time	Complexity	Operator error	Detection
ELISA	+++	+	n/a	> 12 h	++	+++	Quantitative
Dot blot	++	++	++	> 3 h	+	++	Semi-quantitative
HPLC	++	+++	+++	<1 h	+++	+	Quantitative

Cost can also be a predominant factor when choosing the analytical method to be used. By the nature of the assay itself, the immuno-assays are costly. In particular, the dot blot costs are associated with the primary antibodies used. Comparatively, these are up to 10x higher than the RNase T1 enzyme required for HPLC assay, but cost will be diluted with increased throughput. Evidently, HPLC costs are strongly anchored with the stationary phase and equipment used. Finally, at manufacturing level, HPLC can be used as at-line method integrated in the process workflow[37] allowing data to be obtained faster and consequentially enabling a tight process control.

## Conclusions

dsRNA is a product related impurity produced during the mRNA vaccine manufacturing. This impurity has a strong impact on mRNA performance within the cells as it decreases the translation rate and increases the inflammatory response, and presence of this particular impurity in the final product must be quantified and avoided.

In this work, the use of HPLC for dsRNA quantification is explored, and compared with established immunoassays, namely the dot blot and ELISA. With the developed RP-HPLC

method it was possible to quantify dsRNA in sample in under 30min. The sensitivity and precision of this methods is high, with a broader detection range 0.025 to 1 g L<sup>-1</sup> and minimum impurity detection interference. From a regulatory perspective, it achieved the lowest limit of detection (LOD) and limit of quantification (LOQ),  $1.41 \times 10^{-2} \pm 1.78 \times 10^{-3}$  g L<sup>-1</sup>, and  $4.27 \times 10^{-2} \pm 5.4 \times 10^{-3}$  g L<sup>-1</sup>, respectively, compared to the current golden standard, the dot blot. The occurrence of false positive or negatives with this method is also low given the decision limit and detection capability obtained.

Additionally, the implementation of the HPLC method requires minimum operator input and sample handling with throughput achieved with method running time. This, combined with the ability to use this method at-line and prone to automation makes the HPLC an ideal method to be used to quantify dsRNA throughout the manufacturing process. Precise and reliable analytical assays are of paramount importance to have well-established manufacturing process that delivers high quality products.

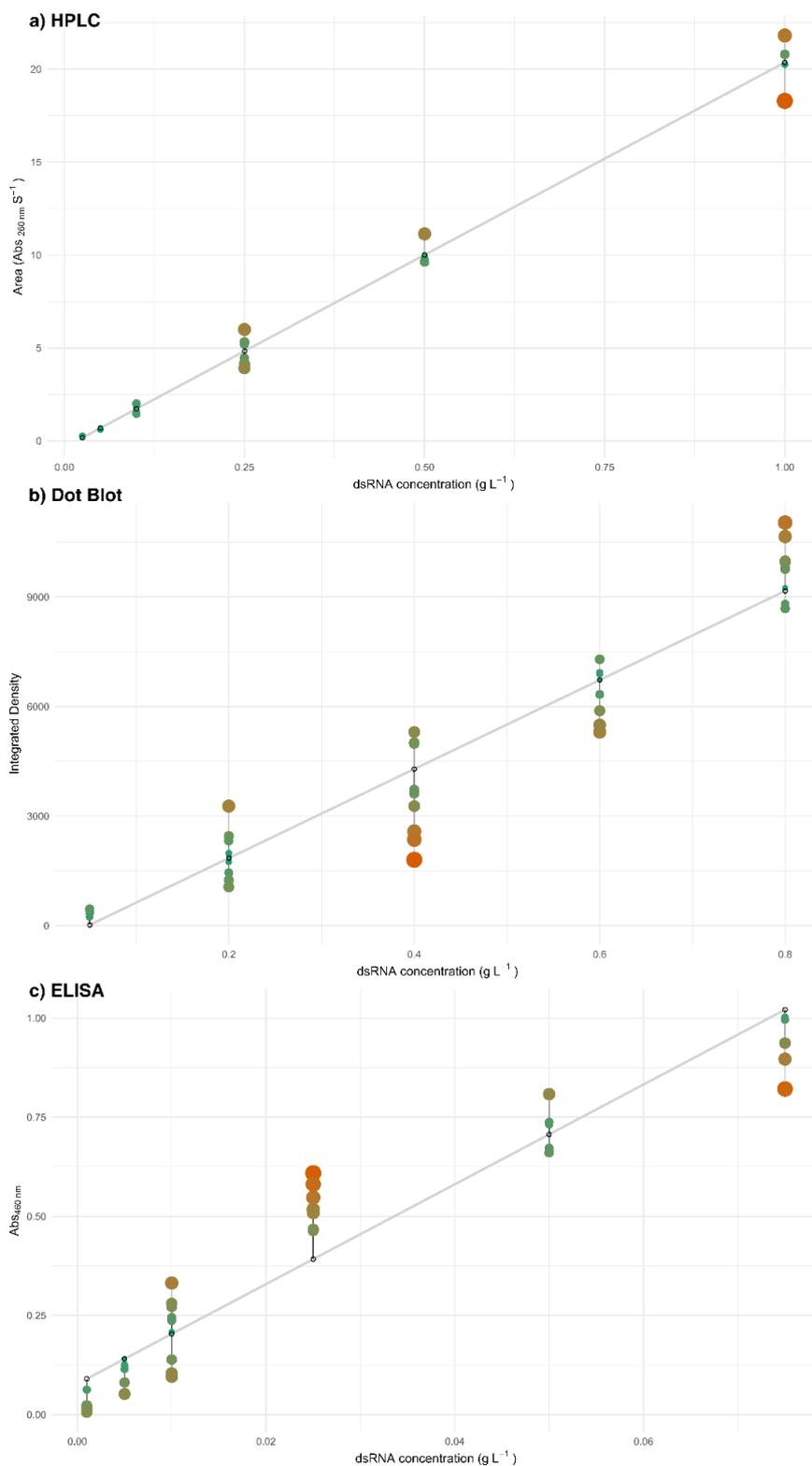
## References

- [1] Wang YS, Kumari M, Chen GH, Hong MH, Yuan JPY, Tsai JL, et al. mRNA-based vaccines and therapeutics: an in-depth survey of current and upcoming clinical applications. *J Biomed Sci.* 2023 Dec;30(1):1–35.
- [2] Grinsted J, Liddell J, Bouleghlimat E, Kwok KY, Taylor G, Marques MPC, et al. Purification of therapeutic & prophylactic mRNA by affinity chromatography. *Cell Gene Therapy Insights.* 2022 Mar 31;8(2):335–49.
- [3] Organization WH, others. Evaluation of the Quality, Safety and Efficacy of RNA-Based Prophylactic Vaccines for Infectious Diseases: Regulatory Considerations. Retrieved January. 2020;1:2022.
- [4] Bancel S, Issa WJ, Aunins JG, Chakraborty T. Manufacturing methods for production of RNA transcripts. Google Patents; 2018.
- [5] Henderson JM, Ujita A, Hill E, Yousif-Rosales S, Smith C, Ko N, et al. Cap 1 Messenger RNA Synthesis with Co-transcriptional CleanCap® Analog by In Vitro Transcription. *Current Protocols.* 2021;1(2):e39.
- [6] Wochner A, Roos T, Ketterer T. Methods and means for enhancing rna production. US20170114378A1, 2017.
- [7] Rosa SS, Nunes D, Antunes L, Prazeres DMF, Marques MPC, Azevedo AM. Maximizing mRNA vaccine production with Bayesian optimization. *Biotechnology and Bioengineering.* 2022;119(11):3127–39.
- [8] Pregeljč D, Skok J, Vodopivec T, Mencin N, Krušič A, Ličen J, et al. Increasing yield of in vitro transcription reaction with at-line high pressure liquid chromatography monitoring. *Biotechnology and Bioengineering.* 2023;120(3):737–47.
- [9] Skok J, Megušar P, Vodopivec T, Pregeljč D, Mencin N, Korenč M, et al. Gram-Scale mRNA Production Using a 250-mL Single-Use Bioreactor. *Chemie Ingenieur Technik.* 2022;94(12):1928–35.
- [10] Linares-Fernández S, Lacroix C, Exposito JY, Verrier B. Tailoring mRNA Vaccine to Balance Innate/Adaptive Immune Response. *Trends in Molecular Medicine.* 2020 Mar 1;26(3):311–23.

- [11] Milano G, Gal J, Creisson A, Chamorey E. Myocarditis and COVID-19 mRNA vaccines: a mechanistic hypothesis involving dsRNA. *Future Virology*. 2021;(0).
- [12] Kariko K, Muramatsu H, Ludwig J, Weissman D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic acids research*. 2011;39(21):e142–e142.
- [13] Dousis A, Ravichandran K, Hobert EM, Moore MJ, Rabideau AE. An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts. *Nat Biotechnol*. 2023 Apr;41(4):560–8.
- [14] Miller M, Alvizo O, Chng C, Jenne S, Mayo M, Mukherjee A, et al. An Engineered T7 RNA Polymerase for efficient co-transcriptional capping with reduced dsRNA byproducts in mRNA synthesis. *bioRxiv*. 2022;2022–09.
- [15] Gholamalipour Y, Karunanayake Mudiyansele A, Martin CT. 3' end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character-RNA-Seq analyses. *Nucleic Acids Res*. 2018 Oct 12;46(18):9253–63.
- [16] Rosa SS, Prazeres DMF, Azevedo AM, Marques MPC. mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine*. 2021 Apr;39(16):2190–200.
- [17] Peña ACD, Vaduva M, S. Li N, Shah S, Frej MB, Tripathi A. Enzymatic isolation and microfluidic electrophoresis analysis of residual dsRNA impurities in mRNA vaccines and therapeutics. *Analyst*. 2024;149(5):1509–17.
- [18] Peña ACD, Li N, Vaduva M, Bwanali L, Tripathi A. A microfluidic electrophoretic dual dynamic staining method for the identification and relative quantitation of dsRNA contaminants in mRNA vaccines. *Analyst*. 2023;148(16):3758–67.
- [19] Luo D, Wu Z, Wang D, Zhang J, Shao F, Wang S, et al. Lateral flow immunoassay for rapid and sensitive detection of dsRNA contaminants in in vitro-transcribed mRNA products. *Molecular Therapy - Nucleic Acids*. 2023 Jun 13;32:445–53.
- [20] Eskelin K, Lampi M, Coustau C, Imani J, Kogel KH, Poranen MM. Analysis and purification of ssRNA and dsRNA molecules using asymmetrical flow field flow fractionation. *Journal of Chromatography A*. 2022 Nov 8;1683:463525.
- [21] Nwokeoji AO, Kung AW, Kilby PM, Portwood DE, Dickman MJ. Purification and characterisation of dsRNA using ion pair reverse phase chromatography and mass spectrometry. *Journal of Chromatography A*. 2017 Feb 10;1484:14–25.
- [22] De Vos J, Morreel K, Alvarez P, Vanluchene H, Vankeirsbilck R, Sandra P, et al. Evaluation of size-exclusion chromatography, multi-angle light scattering detection and mass photometry for the characterization of mRNA. *Journal of Chromatography A*. 2024 Mar 29;1719:464756.
- [23] Camperi J, Lippold S, Ayalew L, Roper B, Shao S, Freund E, et al. Comprehensive Impurity Profiling of mRNA: Evaluating Current Technologies and Advanced Analytical Techniques. *Analytical Chemistry*. 2024;
- [24] Guideline, I. C. H. H. Validation of Analytical Procedures Q2 (R2); ICH: Geneva, Switzerland, 2022.
- [25] Tusup M, French LE, De Matos M, Gatfield D, Kundig T, Pascolo S. Design of in vitro Transcribed mRNA Vectors for Research and Therapy. *CHIMIA International Journal for Chemistry*. 2019;73(5):391–4.
- [26] Trepotec Z, Geiger J, Plank C, Aneja MK, Rudolph C. Segmented poly(A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life. *RNA*. 2019 Apr 1;25(4):507–18.
- [27] Nwokeoji AO, Kilby PM, Portwood DE, Dickman MJ. Accurate Quantification of Nucleic Acids Using Hypochromicity Measurements in Conjunction with UV Spectrophotometry. *Anal Chem*. 2017 Dec 19;89(24):13567–74.
- [28] Williamlssa M. Methods for hplc analysis. US20210163919A1; 2021.

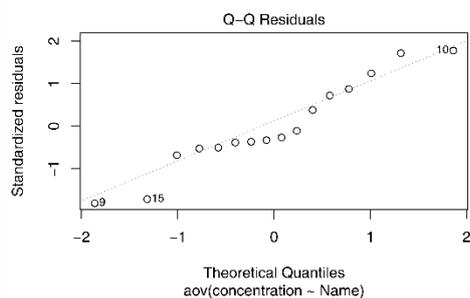
- [29] Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU (Text with EEA relevance. ) [Internet]. OJ L, 32017R0746 May 5, 2017. Available from: <http://data.europa.eu/eli/reg/2017/746/oj/eng>
- [30] Van Loco J, Jànosi A, Impens S, Fraselle S, Cornet V, Degroodt JM. Calculation of the decision limit ( $CC\alpha$ ) and the detection capability ( $CC\beta$ ) for banned substances: The imperfect marriage between the quantitative and the qualitative criteria. *Analytica Chimica Acta*. 2007 Mar 14;586(1):8–12.
- [31] Aguilar MI. Reversed-phase high-performance liquid chromatography. HPLC of peptides and proteins: Methods and protocols. 2004;9–22.
- [32] Weissman D, Pardi N, Muramatsu H, Karikó K. HPLC purification of in vitro transcribed long RNA. *Synthetic messenger RNA and cell metabolism modulation: Methods and protocols*. 2013;43–54.
- [33] Baronti L, Karlsson H, Marušič M, Petzold K. A guide to large-scale RNA sample preparation. *Anal Bioanal Chem*. 2018 May 1;410(14):3239–52.
- [34] Schonborn J, Oberstraß J, Breyel E, Tittgen J, Schumacher J, Lukacs N. Monoclonal antibodies to double-stranded RNA as probes of RNA structure in crude nucleic acid extracts. *Nucleic Acids Research*. 1991 Jun 11;19(11):2993–3000.
- [34] Moderna E. Assessment Report COVID-19 Vaccine Moderna. Procedure No. EMEA/H/C/005791/0000. Netherlands: European Medicines Agency; 2021.
- [35] Pfizer E. Assessment Report Comirnaty. Procedure No. EMEA/H/C/005735/0000. Netherlands: European Medicines Agency; 2021.
- [36] Pregeljč D, Skok J, Vodopivec T, Mencin N, Krušič A, Ličen J, et al. Increasing yield of in vitro transcription reaction with at-line high pressure liquid chromatography monitoring. *Biotechnology and Bioengineering*. 2023;120(3):737–47.

# Supplementary Material

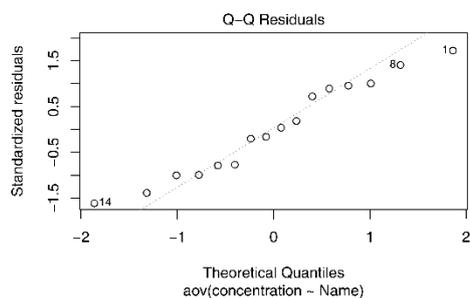


**Supplementary Figure S1.** Residual plots for the calibration curves used to calculate dsRNA concentration using the HPLC (a), the Dot Blot (b) and ELISA (c).

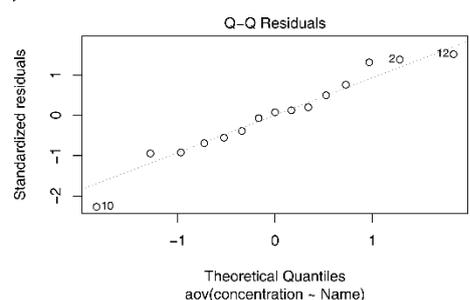
a) HPLC low\_RNA



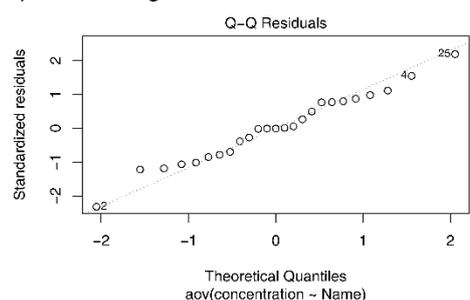
b) HPLC high\_RNA



c) Dot Blot low\_RNA



d) Dot Blot high\_RNA



**Supplementary Table S1.** Dot Blot Tukey test results for high dsRNA concentration (0.4 g.L<sup>-1</sup>). Concentration used for DNA correspond to 10 and 90 nM and to NTPs to 2 and 7.6 mM, for low and high, respectively.

		Confidence Interval at 95%			
		Mean Difference	Lower bound	Upper bound	adjusted P value
DNA_high	DNA_low	0.169	-0.063	0.4	0.228
	NTP_high	0.319	0.108	0.531	0.002
	NTP_low	0.042	-0.17	0.253	0.975
	RNA_high	0.008	-0.212	0.228	1
DNA_low	NTP_low	0.151	-0.061	0.362	0.246
	DNA_low	-0.127	-0.338	0.085	0.404
	RNA_high	-0.161	-0.381	0.059	0.223
NTP_low	NTP_high	-0.277	-0.467	-0.088	0.002
RNA_high	NTP_high	-0.312	-0.51	-0.113	0.001
	NTP_low	-0.034	-0.233	0.164	0.985

**Supplementary Table S2.** Dot Blot Tukey test results for low dsRNA concentration. (0.1 g.L<sup>-1</sup>). Concentration used for DNA correspond to 10 and 90 nM and to NTPs to 2 and 7.6 mM, for low and high, respectively.

		Confidence Interval at 95%			
		Mean Diference	Lower bound	Uper bound	adjusted P value
DNA_high	DNA_low	-0.004	-0.052	0.044	0.998
	NTP_high	-0.013	-0.058	0.032	0.864
	NTP_low	-0.022	-0.076	0.031	0.663
	RNA_low	-0.034	-0.082	0.014	0.213
DNA_low	NTP_low	-0.009	-0.054	0.036	0.965
	DNA_low	-0.018	-0.071	0.036	0.808
	RNA_low	-0.03	-0.078	0.018	0.322
NTP_low	NTP_high	-0.009	-0.06	0.042	0.974
RNA_low	NTP_high	-0.021	-0.066	0.024	0.571
	NTP_low	-0.012	-0.065	0.042	0.947

**Supplementary Table S3.** HPLC Tukey test results for high dsRNA concentration (0.4 g.L<sup>-1</sup>). Concentration used for DNA correspond to 10 and 90 nM and to NTPs to 2 and 7.6 mM, for low and high, respectively.

		Confidence Interval at 95%			
		Mean Diference	Lower bound	Uper bound	adjusted P value
DNA_high	DNA_low	-0.014	-0.139	0.111	0.996
	NTP_high	-0.041	-0.157	0.075	0.775
	NTP_low	0.02	-0.105	0.145	0.984
	RNA_high	0.014	-0.129	0.156	0.998
DNA_low	NTP_low	-0.028	-0.153	0.098	0.95
	DNA_low	0.034	-0.1	0.168	0.919
	RNA_high	0.027	-0.122	0.177	0.974
NTP_low	NTP_high	0.061	-0.064	0.187	0.534
RNA_high	NTP_high	0.055	-0.087	0.197	0.724
	NTP_low	-0.007	-0.156	0.143	1

**Supplementary Table S4.** HPLC Tukey test results for low dsRNA concentration (0.1 g.L<sup>-1</sup>). Concentration used for DNA correspond to 10 and 90 nM and to NTPs to 2 and 7.6 mM, for low and high, respectively.

		Confidence Interval at 95%			
		Mean Diference	Lower bound	Uper bound	adjusted P value
DNA_high	DNA_low	0.011	-0.011	0.033	0.503
	NTP_high	-0.003	-0.027	0.021	0.992
	NTP_low	0.007	-0.017	0.031	0.874
	RNA_low	0.009	-0.018	0.036	0.824
DNA_low	NTP_low	-0.014	-0.038	0.01	0.352
	DNA_low	-0.004	-0.028	0.02	0.976
	RNA_low	-0.002	-0.029	0.025	0.998
NTP_low	NTP_high	0.01	-0.015	0.036	0.71
RNA_low	NTP_high	0.012	-0.017	0.04	0.666
	NTP_low	0.002	-0.027	0.03	0.999

# Chapter V - Maximizing mRNA vaccine production with Bayesian optimization

---

**Key words:** mRNA; IVT; Optimization; Manufacturing; Artificial Intelligence

This chapter has been published as: Rosa, Sara Sousa, Davide Nunes, Luis Antunes, Duarte MF Prazeres, Marco PC Marques, and Ana M. Azevedo. "Maximizing mRNA vaccine production with Bayesian optimization." *Biotechnology and Bioengineering* 119, no. 11 (2022): 3127-3139.

## Introduction

Delivering vaccines in a short time is key to control disease outbreaks, as recently shown during the Covid-19 pandemic. However, this is a demanding task, since traditional vaccine manufacturing is complex, time-consuming and lacks the flexibility required for a timely and global response [1, 2]. During the Covid-19 pandemic event, mRNA technology allowed for the delivery of an approved vaccine in a record-breaking time of less than one year [3]. This success is attributed to the inherent flexibility and precision of mRNA vaccines, as the same vaccine backbone can be used for multiple targets, and only the gene of interest is expressed. Additionally, mRNA vaccines can be manufactured in a standardized fashion, allowing for the production of different vaccine targets with the same platform [4]. Nonetheless, global demand has peaked for COVID-19 vaccines and manufacturers are struggling with the supply chain [5]. Scaling up the manufacturing process can be limited, as a result of constraints in materials and equipment availability, and lack of knowledge on the manufacturing process itself [6]. Enabling on-demand vaccine production requires an efficient and cost-effective manufacturing process that makes optimal use of existing resources.

mRNA vaccines are produced in a cell-free system that *in vitro* transcribes the desired DNA template into a mRNA molecule using a RNA polymerase and nucleoside triphosphates (NTPs) as substrate. Other components (e.g. magnesium, RNase inhibitors, and inorganic pyrophosphatase) are added to the reaction that is performed under controlled pH and temperature conditions [7, 8], delivering grams of product per litre of reaction in a matter of hours. The fine balance between reaction components and conditions will dictate the reaction outcome, i.e., the quantity and quality of mRNA produced. The large number of variables in an *in vitro* transcription (IVT) reaction can constitute an optimization challenge and a critical problem for Design of Experiment (DoE) methods traditionally employed for reaction optimization [9]. These conventional DoE methodologies can introduce human bias in parameter factor selection and often imply oversimplified assumptions about the parameter relationships. Furthermore, typical DoE approaches can be time and cost prohibitive (e.g., for reactions that require 12 parameters to be evaluated with each parameter having 3 levels to be considered, a full factorial design requires  $5 \times 10^5$  ( $3^{12}$ ) experimental runs to be performed). Routinely, to reduce the number of optimization runs, some variables are kept constant. These fixed values frequently come from the operator's intuition or from insights from existing literature, often depicting similarly constrained settings[10]. Thus, new methods are necessary to overcome the traditional limitations of DoE approaches. We therefore propose Bayesian optimization as a methodology that allows for a scientist-in-the-middle approach[11], where domain knowledge, complemented by model analysis/explanations, must be sought after, to fine tune the search for optimal reaction conditions.

Bayesian optimization is an iterative global optimization method, adequate for problems in which we want to maximize a black-box objective function over a bounded set of variables[12, 13]. The method is appropriate for settings in which the unknown function is expensive to evaluate (e.g. in terms of time and/or resources), meaning that it is desirable to spend computational time making better choices about where to seek the best parameters. Fundamentally, Bayesian optimization has two main components: a surrogate model, and an acquisition function. The surrogate model mimics the behaviour of the unknown expensive function being optimized, while being computationally cheaper to evaluate. It provides a prior probability distribution over all possible objective functions, representing our belief about the function's properties such as amplitude and smoothness. The prior distribution is updated with each new measurement to produce a more accurate posterior distribution. The next point to be evaluated is determined by the acquisition function, which is based on the mean ( $\mu$ ) and standard deviation ( $\sigma$ ) of the surrogate model. For example, one can compute the maximum expected improvement (EI)[12] over the current best result. The confidence intervals allow us to quantify model uncertainty and decide when to stop the search.

The experiment design challenge is not unique to IVT reactions. In fact, design problems are pervasive in both scientific and industrial settings. Bayesian optimization emerged as a powerful solution for varied design problems[14] in domains like synthetic chemistry[15] or machine learning[16], among many others. In this article, we present for the first time a set of Machine Learning techniques to design, guide and analyse experimental processes. In particular, our approach based on Bayesian optimization and explanation models was applied to the production of mRNA vaccines by IVT. The data-driven method we used allowed to achieve optimal mRNA production while dealing with a large parameter space, whilst significantly reducing the number of required experimental runs. Optimal IVT conditions were found in 60 runs with a maximum of  $12g.L^{-1}$  total mRNA being produced in approximately 2 hours, which corresponds to a IVT yield increase of a factor of two in half of the time, when compared to the best available baseline reaction. These results reinforce the potential of Bayesian optimization to be applied on the optimization of biochemical reactions for industrial applications.

## Materials and Methods

### mRNA synthesis

Template design. The mRNA template comprises the EGFP gene (GenBank Accession #AAB02572.1) flanked by 5' and 3' UTR. The 5' UTR contains the T7 RNA polymerase promoter, an eukaryotic translation initiation factor 4 G eIF4G binding site and a Kozak consensus sequence[44]. Two beta-globin tandem repeats are used as a 3' UTR, followed by a 120 bp poly-A, segmented with a 6 bp spacer[45]. Two additional templates were constructed by fusing

the Covid-19 Receptor Bind Domain (RDB) gene (GenBank Accession #YP\_009724390.1) and the Scocas9[46] gene with the EGFP gene. All the templates are inserted in a puc57 vector with kanamycin resistance. Sequences used are found in Supplementary Table S2.

**Template production.** DNA template is obtained by polymerase chain reaction (PCR). The reaction mixture contained 20 ng. mL<sup>-1</sup> of plasmid, 0.4 μM of forward and reverse primers, 0.2 mM dNTP mix; 1 x Reaction buffer, 1 x Stabilizer Solution and 0.025 U.μL<sup>-1</sup> NZYProof DNA Polymerase (NZYtech, Portugal). Primer sequences are found in Supplementary Table S3. The reaction is prepared to a final volume of 1000 μL and further split into single 20 μL reactions. The PCR reaction is initiated by a denaturation step at 95 °C for 3 minutes, followed by 30 cycles of: 1) 30 seconds at 95 °C; 2) annealing at 57.5 °C for 30 s; 3) extension at 72 °C for 60 seconds per kbp. The final extension is performed at 72 °C for 60 s. The PCR product is purified and concentrated 20 times using Sera-Mag Select (Cytiva, Sweden) following the manufacturer instructions. Briefly, one volume of Sera-Mag select is added to the pooled PCR reactions. After incubation, the supernatant is removed and the beads are washed two times with 85% v/v ethanol. The purified template is then eluted in water for injection (WFI), and quantified by UV spectroscopy using NanoDrop (ThermoFisher, USA).

***in vitro* transcription reactions.** *in vitro* transcription (IVT) reactions are performed using T7 P&L RNA polymerase HC (Jena Biosciences, Germany). The purified PCR product is used as template, and natural NTPs (New England Biolabs, USA) are used as substrate. *E. coli* inorganic pyrophosphatase (New England Biolabs, USA) and RNase inhibitors (NZYtech, Portugal), are added to the reaction volume. The reaction medium also contains magnesium acetate (heptahydrate) and magnesium chloride, spermidine (Alfa-Aesar, USA), dithiothreitol (Sigma, Germany), 40 mM Tris-HCl (Fisher, USA). The final volume of 20 μL was made up with WFI. Reactions were carried out in a thermocycler (Biometra, Germany). Reaction parameters were varied between set boundaries during optimization experiments (Table 1). The obtained samples were quantified using reverse-phase high performance liquid chromatography (RP-HPLC).

**mRNA purification.** mRNA was purified using MEGAclean™ Transcription Clean-Up Kit (ThermoFisher, USA) following manufacturing instructions with minor adjustments. Briefly, 3 20 μL IVT reactions were pooled, 9 μL of TURBO™ DNase (ThermoFisher, USA), and 1 μL of 10x TURBO™ DNase Buffer (ThermoFisher, USA) were added, and the sample was incubated for 15 min at 37 °C. 350 μL of binding buffer and 250 μL of absolute ethanol were added to the sample. The sample was loaded into the spin filter and centrifuged at 15000xg for 1 min. The filter was washed and centrifuged in the previous conditions twice. For elution, 50 μL of elution buffer were added to the filter, incubated 5 min at 65 °C, and centrifuged at 15000xg for 1 min.

Elution was performed twice. The 100  $\mu\text{L}$  mRNA sample was further concentrated by precipitating with 10  $\mu\text{L}$  of 5 M ammonium acetate and 270  $\mu\text{L}$  absolute ethanol at  $-20^\circ\text{C}$  for 30 minutes. The sample was centrifuged at 15000xg for 10 min at  $4^\circ\text{C}$ . The obtained pellet was washed with 100  $\mu\text{L}$  75% ethanol and centrifuged using the previous conditions. The obtained pellet was let to dry and re-suspended in 20  $\mu\text{L}$  of elution buffer. The mRNA was quantified using Nanodrop 1 (ThermoFisher, USA) and the quality was evaluated by RP-HPLC.

IVT kinetics. IVT kinetics was studied with the reaction conditions described in Table 2 at volume of 65  $\mu\text{L}$ . Reactions were carried out in a thermocycler (Biometra, Germany) and 5  $\mu\text{L}$  samples were taken in a course of 5 hours. The IVT reaction was stopped by dilute the sample 8 times in 1x pH 7 TAE buffer (100 mM Tris acetate, pH 7, 2.5 mM EDTA). The obtained samples were evaluated by RP-HPLC and gel electrophoresis.

### Analytical methods

**mRNA quantification** . mRNA was quantified using reverse-phase high performance liquid chromatography (RP-HPLC) using the method adapted from [47]. A 2.1 x 100 nm RP-DNApac column and a guard column (3 x 10 nm) (ThermoFisher, USA) were used in a HPLC equipped with a column heater. Samples of 5  $\mu\text{L}$  diluted 8 times in 1x pH 7.4 TAE buffer (100 mM Tris acetate, pH 7, 2.5 mM EDTA), were injected in a pre-equilibrated column with TAE buffer. The samples were eluted using 1x TAE with 25% (v/v) acetonitrile. The run was performed at  $80^\circ\text{C}$ , and the absorbance was monitored at 260 and 280 nm. The run conditions are present in supplementary Table S4. The peak area corresponding to the elution of mRNA was considered for the evaluation. Calibration curves were constructed using purified mRNA samples with known concentrations in the range 0.5-16  $\text{g.L}^{-1}$ .

**Gel electrophoresis.** Samples obtained from the IVT kinetic study were analysed by gel electrophoresis. A 2% (w/v) agarose (Fisher Scientific, USA) was prepared with 0.5 x TBE buffer (ThermoScientific, USA) containing 5.5 mM of magnesium chloride (Fisher Scientific, USA) and pre-stained with ethidium bromide (ThermoFisher, USA). The gel was loaded with a 1  $\mu\text{L}$  mRNA sample diluted in 15  $\mu\text{L}$  of WFI and 4  $\mu\text{L}$  of 6 x purple Loading Dye (New England Biolabs, USA), and 4  $\mu\text{L}$  of NZYDNA ladder III (NZYTech, Portugal). The electrophoresis was performed at 100 V for 120 min using 0.5 x TBE buffer, 5.5 mM  $\text{MgCl}_2$ . The gels were scanned using an Axygen Gel Documentation System (Axygen, USA).

### Bayesian optimization

**Latin Hypercube Sampling (LHS).** Latin hypercube sampling[17] is used to generate a series of initial experiments before the Bayesian optimization process is guided by the surrogate model. An implementation is available in the scikit-optimize library[48]. LHS was used to sample an initial set of 16 configurations to guarantee that the initial batch of reaction parameters don't

overlap, and are sufficiently scattered over the candidate domains. The rest of the experiments are guided by the Bayesian optimization approach.

**optimization Cycle.** Two sequential experiments were performed. Overall, 16 configurations were sampled and evaluated from an initial LHS design, and used to initialize the Bayesian optimization cycle (Figure 1a). After the surrogate model is initialized, 3 to 5 reaction conditions are suggested by the Bayesian model, the outcome for the IVT reactions for these conditions is evaluated and used to update the model each time. The cycle was continued until no significant gains were observed for a total of 150 reactions.

**Gaussian Process.** A Gaussian process (GP) was chosen to be the surrogate model for the Bayesian optimization process. A Gaussian process is a stochastic process such that any finite sub-collection of random variables has a multivariate Gaussian distribution. A Gaussian process defines a prior distribution on functions  $f: X \rightarrow \mathbb{R}$  and can be thought of as the generalization of a Gaussian distribution over a finite vector space to a function space of finite dimension. Just as a Gaussian distribution is fully specified by its mean and covariance matrix, a GP is specified by a mean function  $m(x) = E[f(x)]$  and a positive definite covariance (kernel) function  $k(x,x) = E[(f(x) - \mu(x))(f(x) - \mu(x))]$ . The chosen kernel greatly impacts the resulting distribution on functions and can correspond to strong assumptions about them (e.g. smoothness and differentiability). The squared exponential kernel is often the default choice for Gaussian process regression, but sample functions with this covariance function are unrealistically smooth for most practical optimization problems, as such, a Matérn 5/2 kernel was chosen to be the kernel function for the GP:

$$k(x, x) = \theta_0 \left( 1 + \sqrt{5r^2(x, x)} + \frac{5}{3}r^2(x, x) \right) \exp \left\{ -\sqrt{5r^2(x, x)} \right\} \quad (1)$$

Matérn 5/2 yields twice differentiable sample functions, e.g. like quasi-Newton methods, without requiring the smoothness of the squared exponential. For the kernel hyperparameters (length scales, covariance amplitude, observation noise and constant mean) a point estimate of these parameters was used by optimizing the marginal likelihood under the Gaussian process. For a broader introduction to Gaussian Processes see [28, 49, 50].

**Acquisition Function.** The acquisition function gives us the candidates for the next reaction configurations to be evaluated on each optimization cycle. The expected improvement (EI) acquisition function[12] is chosen for the Bayesian optimization process. The EI criterion is computed as follows: Let  $f_{\max} = \max(y^{(1)}, \dots, y^{(n)})$  be the current best function value. Let us model of uncertainty at  $y(x)$  as a normally distributed random variable  $Y$  with mean and standard deviation given by the surrogate model. Weighing all the improvements, the portion of the uncertainty density that extends beyond the current  $f_{\max}$  by the corresponding density values,

will give the expected improvement. Formally, the improvement at a point  $x$  is given by the random variable

$I = \min(0, f_{\max} + Y)$  and it models the uncertainty about the objective function value at  $x$ . The EI is the expected value:

$$E[I(x)] \equiv E[\min(0, f_{\max} + Y)] = (f_{\max} + \hat{y}) \Phi\left(\frac{f_{\max} + \hat{y}}{s}\right) + \phi\left(\frac{f_{\max} + \hat{y}}{s}\right) \quad (2)$$

where  $\hat{y}$  and  $s$  are the surrogate model prediction and its standard error at  $x$ ,  $Y$  is  $\text{Normal}(\hat{y}, s^2)$ , and  $\phi(\cdot)$  and  $\Phi(\cdot)$  are the standard normal density and distribution functions. Selecting values where  $x$  maximizes the EI acquisition function or from at random within a certain distance from the maximum improvement gives a balance between exploration and exploitation.

**SHAP (SHapley Additive exPlanations) of Gaussian Process Estimator.** Understanding all predictive elements of our Bayesian optimization framework provides insight into how the model can be improved, and supports understanding of the process being modelled. SHAP[18] is a method based on the game theory concept of optimal Shapley values that views any explanations of a model's prediction as a model itself, an explanation model. We use this method to understand why our Gaussian Process estimator makes certain predictions. Independent variable values are interpreted as players in a coalition game from which Shapley values are computed. This approach unifies other methods as additive feature attribution methods, where an explanation model is a linear function of binary variables:

$M$

$$g(z') = \phi_0 + \sum_{j=1}^M \phi_j z_j' \quad (3)$$

$j=1$

where  $z' \in \{0,1\}^M$ ,  $M$  is the number of simplified input features, and  $\phi_i \in R$ . The explanation model attributes an effect  $\phi_i$  to each feature, and summing the effects of all feature attributions approximates the output  $f(x)$  of the original model.

The exact computation of SHAP values is challenging[18], so, to generate explanations for the Gaussian Process predictions, approximations are computed using a model-agnostic permutationbased explanation model that uses the Shapley sampling values method[51].

**Surrogate model comparison.** The generalization error of Random Forests (RF)[26] and Gradient Boosting Machines (GBM)[27] was compared to the generalization error for the Gaussian Process used throughout our experiments. Leave-one-out cross validation was performed on the reaction data (150 configurations including replicated runs). The surrogate model is trained on each data fold of size  $n-1$  and the absolute prediction error is measured on

the remaining data point. Figure S1 shows the distribution of errors for GP, RF, and GBM. Our initial choice of GP and kernel is comparable to other possible choices.

## Results

### IVT Optimization Workflow

In an IVT reaction, a RNA polymerase uses a target DNA template to synthesize a RNA molecule using NTPs as substrate. To optimize this reaction, twelve parameters were identified that could potentially influence the reaction outcome: the concentration of the enzymes (T7 RNA polymerase and inorganic phosphatase), RNase inhibitor, DNA template, NTPs, spermidine and dithiothreitol (DTT); the type of cofactor (magnesium acetate vs magnesium chloride) and its respective concentration; and reaction pH, temperature and time (Table 1). The Bayesian optimizer is fed experimental data, which corresponds to the concentration of mRNA produced, obtained by the configurations suggested by the model it maintains (the initial conditions being randomized). In each optimization round, the model is updated and new experimental conditions are drawn based on its current knowledge of the process with the goal of maximizing mRNA production. This process is cyclic (Figure 1a), proceeding as follows:

1. A Gaussian Process surrogate model is initialized with a Matérn 5/2 kernel;
2. The surrogate model is fed with a batch of random initial configurations taken from a latin hypercube sampling (LHS) design [17];
3. The reaction configurations are experimentally performed and evaluated by quantifying the synthesized mRNA by Reverse-phase High Performance Liquid Chromatography;
4. The surrogate model is updated with the evaluations of each reaction condition to construct a posterior distribution;
5. The expected improvement acquisition function[12] is computed based on the surrogate model and its maximum value is used to suggest the next reaction configurations;
6. Steps 3-4 are repeated until the convergence criteria are met, either the maximum budget for the total number of experiments is reached, and/or a number of experimental runs are performed without significant improvement.

Table 1: in vitro transcription (IVT) reaction parameters and evaluation metric (see also Fig 1c) (abbreviations: DTT - Dithiothreitol, NTP - Nucleoside triphosphate).

Name	Units	Type	Domain/Range
Cofactor	Cofactor choice	Categorical	MgAcetate, MgCl2
Cofactor Concentration	mM	Real number	[0,100]
DTT	mM	Real number	[0,10]
RNase Inhibitor	U.mL <sup>-1</sup>	Integer	[0,2000]
NTPs	mM	Real number	[1.0,10.0]

DNA Template	nm	Integer	[10,100]
Inorganic Pyrophosphatase	U.mL <sup>-1</sup>	Integer	[0,10]
Spermidine	mM	Real number	[0,10]
T7 RNA Polymerase	U.mL <sup>-1</sup>	Integer	[1000,50000]
Temperature	C°	Integer	[20,50]
Reaction Time	minutes	Integer	[10,300]
pH	—	Real Number	[6.5,8]
Evaluation	Concentration in g.L <sup>-1</sup>	Integer	...

### Optimization Analysis

The progress of the optimization procedure is tracked in multiple ways. We overview the exploration of the parameter space using a parallel coordinate plot (Figure 1c). This plot provides an overview over the possible optimal ranges for each parameter across the entire optimization process. We also keep track of the best configuration evaluations over time, which we use to determine if significant progress is being made in the optimization cycle. By analyzing the parallel coordinate plot for the entire experiment, it is possible to observe that the optimal mRNA IVT production is obtained when: i) magnesium acetate concentration is set between 40 and 70 mM; ii) NTP concentration is above 7 mM; iii) spermidine concentration is between 1 and 3 mM; iv) T7 RNA polymerase (T7 RNAP) concentration is between 6000 and 8000 U.mL<sup>-1</sup>; v) reaction temperature is between 37 and 45 °C; and vi) pH is lower than 7.5. The final optimal reaction conditions are found in the first 60 runs, after which no increase in mRNA concentration is observed (Figure 1c). In the end, we were to find multiple parameter configurations that yielded more than 10 grams per litre of reaction.

To better understand the IVT reaction being modelled we use two different mechanisms. Firstly, we compare the model predictions in terms of expected mRNA IVT production and uncertainty boundaries with our reaction empirical evaluations. Secondly, we build explanation models for the Gaussian process surrogate model (Figure 1e,f). The SHapley Additive exPlanations (SHAP)[18] give the overall parameter importance in terms of parameter impact related to the model predictions. As this surrogate model is fed with more data, and it becomes closer to the unknown function, in this case, the IVT reaction being modelled, the explanations become more informative about the actual impact of different reaction parameters on the amount of mRNA being produced. The evaluation of the coordinate plot combined with these mechanisms allowed us to understand which parameters had the most impact on surrogate model predictions. In this case, we can observe that: i) the pH seems to have the most importance and lower pH values, between 6.5 and 7.5, have a positive impact on the reaction; ii) cofactor concentration also has

a large impact on the reaction output with optimal range set between 40 and 60 mM; iii) high spermidine concentration seems to negatively impact the model, and values used should range between 1 and 3 mM; iv) high inorganic pyrophosphatase and DTT concentrations impact positively on the reaction; v) T7 RNAP presence impacts the model, but lower concentrations, between 6000 and 8000  $U.mL^{-1}$  produce more mRNA; vi) higher concentrations of DNA and NTPs, above 40 nm and 7 mM, respectively, influence the reaction positively (Figure 1c,e,f).

### IVT kinetics

In our experiments, reaction time was not explicitly optimized, since we did not evaluate the results taking the enzyme specific activity into account. Nevertheless, we further investigated the best reaction conditions in terms of the reaction performance over time. To achieve this, we chose the 6 best parameters combinations found by the optimization process, evaluated mRNA production over time, and compared these dynamics with a benchmark reaction (Table 2). This baseline corresponds to the reaction parameters listed in the Moderna patent [19], which result in an expected IVT yield of 5  $g.L^{-1}$ .

Table 2: mRNA production reaction parameters and mRNA concentration for highest production conditions after Bayesian optimization (Reaction 1-6) compared to the baseline reaction [19] condition (Reaction 7).

Reaction	1	2	3	4	5	6	7
Cofactor C (mm)	60.00	48.46	41.79	59.87	49.28	40.00	40.00
DTT (mm)	7.09	3.85	5.57	9.85	5.27	3.99	5.00
RNase I ( $U.mL^{-1}$ )	829	1072	1217	986	1474	1045	1000
NTPs (mm)	8.57	8.81	9.89	8.50	7.75	9.29	7.50
DNA Template (nm)	61	100	89	100	89	72	40
Ppase ( $U.mL^{-1}$ )	10	9	5	2	8	7	1
Spermidine (mm)	2.65	1.35	2.24	1.31	2.25	2.03	1.00
T7 RNAP ( $U.mL^{-1}$ )	7346	7320	6607	6166	7743	7748	7000
Temperature ( $^{\circ}C$ )	43	39	44	40	44	44	37
Time (min)	263	98	120	148	121	279	240
pH	6.89	6.80	6.65	6.78	6.67	6.60	8.00
mRNA C ( $g.L^{-1}$ )	12.61 $\pm 0.82$	10.76 $\pm 0.47$	11.76 $\pm 0.66$	12.27 $\pm 0.77$	12.18 $\pm 0.98$	11.52 $\pm 0.23$	7.64 $\pm 0.87$

To perform the kinetics analysis, we sampled the reactions over the course of 5h (Figure 2a). All the reactions found by the optimization process performed above the baseline reaction, producing at least 10  $g.L^{-1}$  of mRNA. We determined the best reaction out of the 6 configurations by taking into consideration the reaction yield and the production rate. Although during the optimization, reaction 1 and 4 produced the highest amount of mRNA, reaction 5 achieved the highest yield in less time as after 2 h (Table S1), corresponding to a final concentration of  $10.65 \pm 0.01 g.L^{-1}$ , 80% of the maximum reaction yield (Figure 2b). We also investigated the quality of the mRNA formed during the course of reaction 5 by agarose gel electrophoresis

(Figure 2c). After approximately 115 minutes, no significant changes in mRNA amount are detected, confirming that reaction performance peaks at the 2-hour reaction. A second band is observed to increase after 115 minutes that may correspond to reaction by-products such as double-strand mRNA (ds-mRNA) or aberrant mRNA.

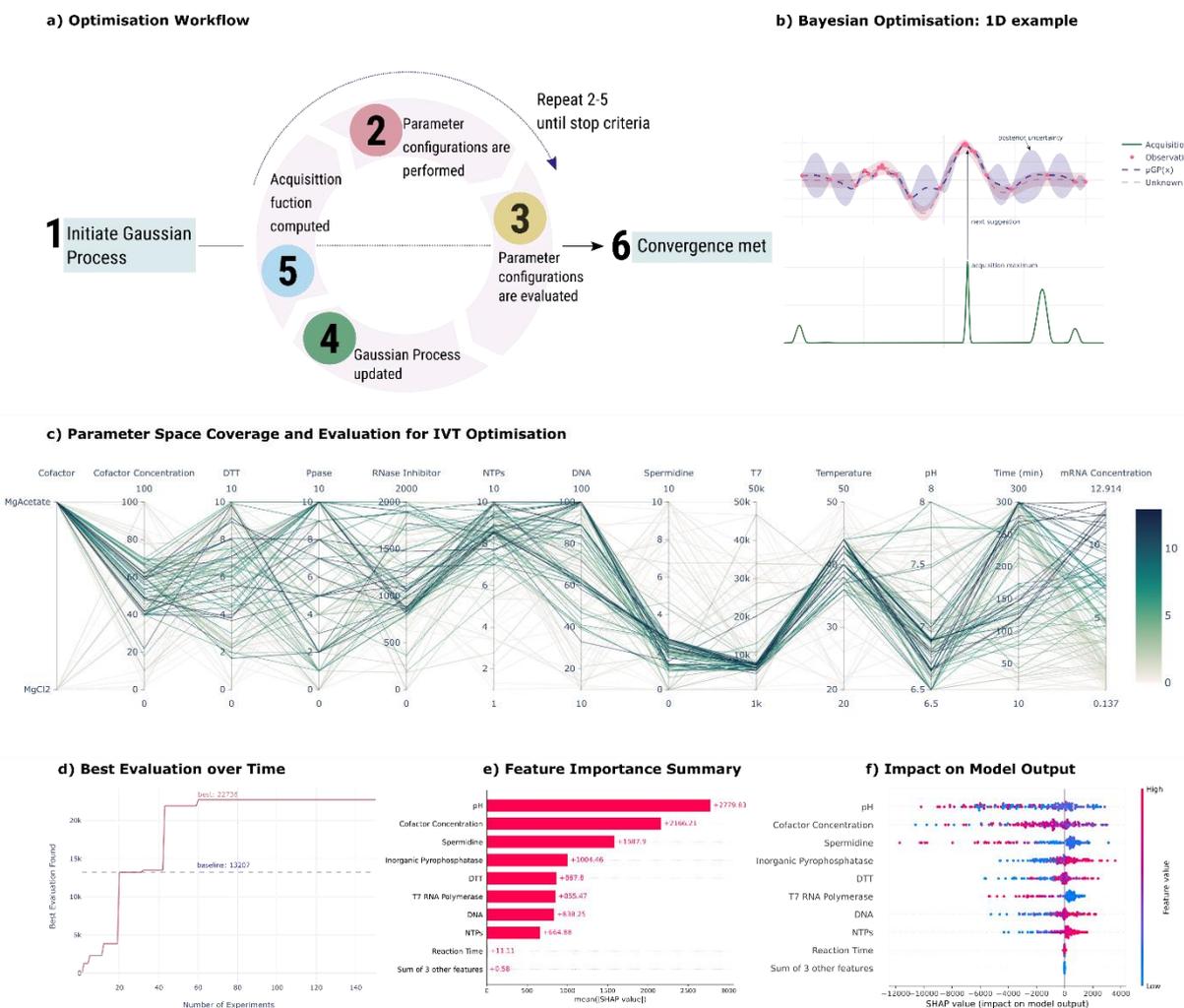


Figure 1: Bayesian optimization of mRNA IVT reaction. a, Bayesian optimization workflow. b, one-dimensional example of Bayesian optimization process using a Gaussian process surrogate model and corresponding acquisition function, maximized to select the next set of parameters to be tested. The surrogate model is plotted as the posterior mean, with the shaded region representing a posterior distribution uncertainty of  $2\sigma$  units. c, All parameter configurations for all the IVT experimental runs along with their respective evaluation in mRNA concentration ( $g.L^{-1}$ ). d, Convergence plot depicting the best evaluation throughout all the IVT experimental runs, and convergence to the optimum. e, Feature importance summary computed from the average SHapley Additive exPlanation (SHAP) values [18] computed for the Gaussian Process regressor predictions across all IVT experimental data. f, Impact of feature value in model prediction value for the Gaussian Process regressor used as surrogate model in the Bayesian optimization process.

### IVT performance validation

To validate the optimized reaction performance, we compared the mRNA production using different size templates. We produced templates containing EGFP, the Covid-19 Receptor Bind Domain fused to EGFP (RDB\_EGFP), and the Cas9 gene also fused to EGFP (Cas9\_EGFP), with 1195, 1864 and 5299 bp, respectively. We investigated the differences between reaction rates (Figure 2d), and concentration of mRNA produced after 2 hours (Figure 2e). Changing the size of the template does not have an impact on the reaction using the optimized conditions as all the evaluated templates produced more than  $10 \text{ g.L}^{-1}$  of mRNA. However, with larger templates the mRNA concentration tends to reduce from a concentration of  $9.7 \pm 0.29 \text{ g.L}^{-1}$  after 145 minutes of reaction time to a concentration of  $7.1 \pm 0.19 \text{ g.L}^{-1}$  at the end of reaction. In spite of this reduction,  $10 \text{ g.L}^{-1}$  of mRNA are still produced within two hours of reaction time.

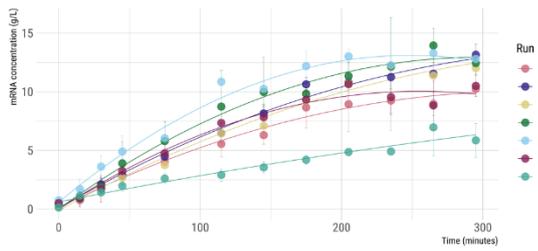
### Optimal IVT production and comparison

Finally, we compared our best reaction conditions with different reaction conditions found in the literature [32] and patents [19, 20] (Figure 2f). Using our optimal reaction parameters and with only 2 hours of reaction time, we were able to produce approximately  $12 \text{ g.L}^{-1}$  of mRNA, which corresponds to a two-fold increase compared with the best baseline (Table 2; reaction 7).

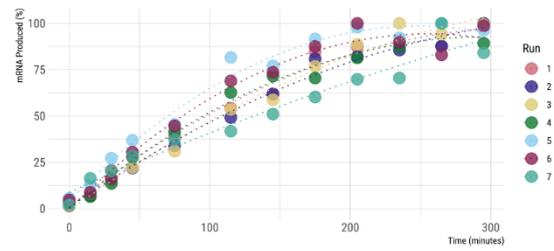
## Discussion

IVT reaction optimization has become increasingly important not only due to the growing interest in using RNA molecules in a number of diagnostic and therapeutic applications, but especially due to the rise of mRNA vaccine technology. However, existing literature is focused on reaction modelling [21, 22, 23, 24] or on DoE [25] methodologies. Optimization methods only consider the exploration of small parameter spaces and often assume that the relationships between reaction parameters are given by pre-established enzymatic dynamic models. In this work, we demonstrate the effectiveness of the Bayesian optimization methodology when applied to the production of mRNA by IVT, a multi-component reaction that depends on 12 different reaction parameters. Using this method, we were able to optimize the reaction in only 60 runs. We also make use of interpretability techniques, in particular, explanation models based on Shapley Values [18] for the Gaussian Process surrogate model. The combination of the obtained results with the explanation models allowed to bridge these explanations/interpretations with pre-existing knowledge about the IVT reaction being modelled.

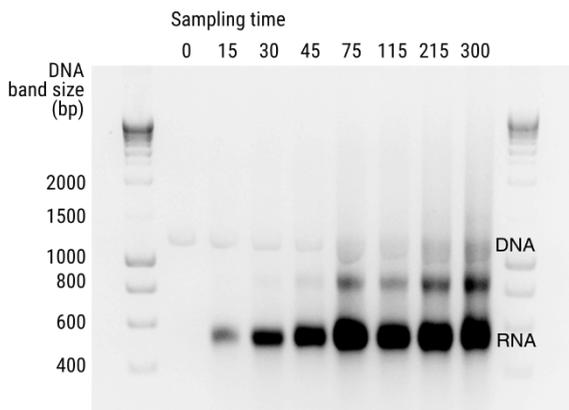
**a) mRNA Production Kinetics**



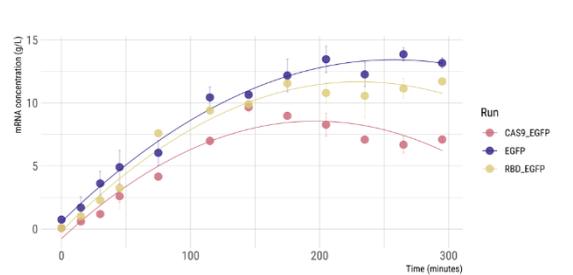
**b) Percentage of mRNA Produced**



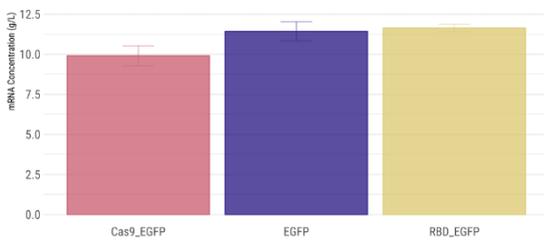
**c) Agarose Gel Electrophoresis**



**d) mRNA Production Rate with Different Sizes Templates**



**e) mRNA Production with Different Sizes Templates**



**f) Benchmark mRNA Production**

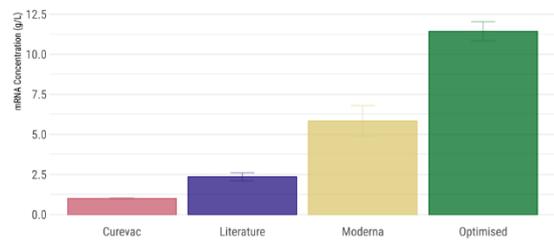


Figure 2: mRNA Production Analysis. a, mRNA production kinetics of the 6 best runs obtained by Bayesian optimization and the benchmark reaction19 for a time course of 5 hours. Error bars represent standard deviation obtained for each point. Trendline function used: 2nd order polynomial. b, Percentage of mRNA produced as a function of minutes of reaction time considering 100% the highest mRNA concentration produced for each set of runs. Error bars represent standard deviation obtained for each point. Trendline function used: 2nd order polynomial. c, Agarose gel electrophoresis analysis of the reaction mixture at the defined setpoints for the best production run (condition 5). d, mRNA production kinetics for a time course of 5 hours using the best run parameters (condition 5), and templates with the different sizes (EGFP - 1195 bp; RBD\_EGFP - 1864 bp ) Cas9\_EGFP - 5299 bp). Error bars represent standard deviation obtained for each point. Trendline function used: 2nd order polynomial. e, mRNA production using the parameters of the Run 5 and 2 hours of reaction time with the 3 different size templates (EGFP, RBD\_EGFP, Cas9\_EGFP). f, mRNA production concentration in grams per litre using EGFP template for the following runs: optimized - Run 5 and 2 hours of reaction time; Moderna [19]; Literature [32], Curevac [20].

For any optimization problem, the most important metric to be defined is the objective evaluation. In the case of IVT reactions, the goal is to maximize the quantity of mRNA produced for a specific set of reaction parameters. To achieve this, we used the mRNA concentration as the surrogate metric. The initial set of 16 reactions to be evaluated were sampled at random using Latin Hypercube Sampling (LHS)[17]. This is a type of stratified Monte Carlo (MC) sampling where the range of each variable is partitioned into  $N$  non-overlapping intervals on the basis of equal probability size  $1/N$ . The number of partitions is equal to the number of required samples. LHS guarantees that the initial parameter configurations do not overlap and are sufficiently scattered over the target parameter space. This is often important to guarantee that the model is seeded with different regions of the target response surface. Nevertheless, since sampling overlapping configurations is extremely unlikely with 12 parameters and we have a small number of initial experiments, this sampling strategy can be as good as random uniform.

Public datasets on IVT reactions for mRNA production are currently non-existent, so there is no basis to estimate the impact of the different components that compose our Bayesian optimization approach. As such, a sensible starting point is chosen for the surrogate model, acquisition function, and hyperparameters, namely: a Gaussian Process (GP) surrogate model with a Matérn 5/2 kernel and the Expected Improvement (EI) acquisition function. This setting does not make assumptions about the underlying structure of the target problem while still being capable of tackling a wide range of domains. Nevertheless, there are multiple sensible choices for surrogate models besides a Gaussian Process. We performed a post-hoc comparison of generalization error between Random Forests (RF)[26], Gradient Boosting Machines (GBM)[27], and Gaussian Processes (GP)[28] and found no significant differences between them in terms of estimated generalization power (Figure S1). Techniques such as deep kernel learning[29] can be used to combine the structural properties of deep learning architectures with the non-parametric flexibility of Gaussian Processes and present a promising research opportunity for future work since they would and would allow for the creation of specialized kernels for this (or similar) domains.

Finally, in terms of the optimization process, no better reaction conditions were found after 60 runs with an incremental increase in mRNA concentration (Figure 1d). However, this optimization only considered the maximization of mRNA production as a goal. This opens an avenue for future work where multiple-objective optimization can be used to seek the Pareto frontier of IVT reactions using multiple metrics. In other words, the set of optimal trade-offs where improving one metric means deteriorating another, which can include minimizing expensive reagent concentrations, reaction time, or estimated reaction cost, while maximizing reaction yield.

By combining the interactive nature of the optimization cycle and the insights gained from the explanation models, we were able to both validate the model predictions and gain insight about the reaction itself. This allowed the construction of a more coherent view of the dynamics of IVT reactions in general. IVT reactions employ an RNA polymerase and a DNA template to generate RNA using the NTPs as substrate and magnesium as a cofactor. T7 RNA polymerase (T7 RNAP) is a 98 kDa and single subunit enzyme that catalyzes RNA synthesis of very long transcripts without auxiliary transcription factors[30, 31]. Due to its characteristics, T7 RNAP is widely used for the production of mRNA by IVT. Understanding the influence of each parameter on the IVT reaction yield, i.e. amount of mRNA produced, is fundamental not only to validate our model predictions, but also to gain a better understanding of the reaction dynamics. To achieve this, we compared the parameters importance and their impact in the surrogate model predictions combined with the information given by the coordinate plot, with existing literature focusing on the impact of those parameters (Figure 1c,e,f). The model predicted that pH, co-factor concentration, spermidine and inorganic pyrophosphatase have the most impact on the reaction outcome, followed by DTT, T7 RNAP, DNA, NTPs and reaction time (Figure 1e). The least impactful parameters predicted include reaction temperature, type of co-factor and RNase inhibitor. In more detail, the model predicts that a low pH value significantly and positively impacts the reaction (Figure 1e,f). Typically, IVT reactions are performed using higher pH values, e.g. 7.9 or 8[19, 32]. However, by exploring a wider range of pH, we found that values between 6.5 and 7.2 improve the transcription rate. These results are in line with previous work where an optimal transcription rate is reported for values of pH between 7.0 and 7.5 [21].

Magnesium also plays an important role in the IVT reaction. Its presence is required to bind T7 RNAP to the template[33]. Additionally, magnesium-NTP complexes are used to form phosphodiester bonds with the RNA chain[24]. During this reaction, a pyrophosphate is released as a by-product. Depending on the concentration of  $Mg^{2+}$  present in the reaction, free pyrophosphate can cross-link with free  $Mg^{2+}$  and precipitate due to the formation of long aggregates. Our results showed that the mRNA production rate is increased with concentrations of magnesium acetate between 40 and 60 mM. It is important to maintain a high concentration of free  $Mg^{2+}$  to ensure that the cofactor does not limit the reaction. Young et al predicted an optimal range between 50 and 60 mM[24]. Additionally, the presence of counter ions can also inhibit mRNA production. Through the optimization process we also found that magnesium acetate is preferred since it can be used in higher concentrations than magnesium chloride. These results were consistent with previous work [21, 25].

The pyrophosphate by-product may inhibit IVT as it reduces the free  $Mg^{2+}$ . To avoid this, an inorganic pyrophosphatase (PPase) can be used to catalyze the hydrolysis of pyrophosphate. This leads to the formation of orthophosphate, releases  $Mg^{2+}$ , and, ultimately, increases mRNA

production [21]. The use of PPase in IVT is not consensual, with some reports claiming that it does not influence mRNA production if the  $Mg^{2+}$  is present in sufficient concentration[21, 22]. However, our findings show that higher concentrations of PPase positively influence the mRNA production (Figure 1f). Previous works also observed this improvement when using PPase concentrations between 5 to 10  $U.mL^{-1}$ [34, 35]. This positive influence can be particularly observed when using high concentration of NTPs[22]. We hypothesize that PPase is important to maintain a threshold concentration of  $Mg^{2+}$  and hence sustain a high mRNA production rate.

Spermidine is an aliphatic polyamine with high affinity towards nucleic acids that neutralizes negative charges, and consequently, promotes condensation and aggregation of DNA[34]. In IVT, spermidine plays an important role in the transcription initiation as it stabilizes the DNA-enzyme complex[35]. Its presence in IVT reaction is fundamental as it can lead to an increase of up to 10 times when using T7 RNAP as enzyme[36]. Here we found that spermidine can also have an inhibitory effect when present in high concentrations. However, when used in concentrations between 1 and 3 mM, it influences the mRNA production positively. These results are in line with optimal conditions found in previous reports[37, 38].

As previously described, there is a close relation between the substrate concentration and the cofactor, as NTPs and  $Mg^{2+}$  form complexes that are added to the nascent mRNA chains. NTPs concentration must be high enough to promote the reaction. However, its increase beyond a certain point does not have a significant influence on the reaction output[24]. Our findings suggest that NTPs concentration above 7 mM have a positive impact on mRNA production, and in line with the literature[39].

We observed that the optimal T7 RNAP concentration lies between 6000 and 8000  $U.mL^{-1}$ , and that values above this point have a negative impact on the reaction yield. This shows that increasing enzyme concentration, i.e. volumetric activity, does not translate into higher mRNA production due to diffusion limitation and solubility challenges. Another important parameter is the template concentration. It should be high enough to guarantee that it is not the limiting factor in the reaction. We observed that DNA concentrations above 40 nm should be used to guarantee optimal production. DTT is a reducing agent that plays an important role to maintain the enzymes active during the transcription. Reaction temperature, although it does not seem to have an impact on the reaction, affects the binding of the enzyme to the template promoter. To achieve total binding [40], temperatures of at least  $37^{\circ}C$  should be used. Additionally, high temperature could also reduce the formation of ds-mRNA [41], a reaction by-product that leads to a decrease in vaccine translation efficiency and modify the immunostimulatory profile when present in the final product [4]. The model predicts that the optimal temperature in the system should be between  $37$  and  $44^{\circ}C$ .

A larger and more coherent picture of the impact of reaction parameters on IVT has emerged from the optimization results herein presented. Furthermore, multiple reaction conditions were found that lead to mRNA concentrations higher than  $10 \text{ g.L}^{-1}$ . In terms of production rate, we observed that the reaction time can be reduced to 2 hours. Ultimately, we found a set of IVT reaction parameters able to produce  $12 \text{ g.L}^{-1}$  in two hours of reaction time (Figure 2f). This means a 2 fold increase in half of time when compared with the benchmark IVT, which is only able to produce roughly  $6 \text{ g.L}^{-1}$  in 4 hours. Additionally, increasing template size did not have an impact on production using the optimized conditions as all the templates evaluated produced more than  $10 \text{ g.L}^{-1}$  of mRNA (Figure 2e). Nevertheless, increasing the reaction time above two hours for longer DNA templates is not beneficial as a decrease in mRNA concentration is observed (Figure 2d). This can be explained by the formation of a precipitate, which can lead to the precipitation of the mRNA already formed, and consequently, to the generation of aberrant mRNA species.

Another important point is the quality of the mRNA produced. This metric was not considered in the optimization evaluation because it introduces a new criterion, transforming the problem into a multi-objective optimization scenario. While this is the natural progression of the present work, a single objective setting allowed us to validate all the components involved before considering tradeoff decisions when searching for the optimal reaction conditions. The presence of by-products in the final product has a strong impact both in mRNA translation efficiency within the patient cells and in the immunostimulatory profile[4]. mRNA quality attributes include the homogeneity of the sample and the presence of aberrant mRNA species such as ds-RNA, which can be analysed by a number of analytical techniques such as electrophoresis, HPLC or ELISA[42]. Additionally, mRNA structure and identity could be also analysed using circular dichroism (CD), RT-PCR or next-generation sequencing[43].

Overall, by using Bayesian optimization, we were able to increase the mRNA IVT production 2-fold in half of when compared with benchmark reactions. This optimization approach proved to be cost-effective, as it only required 60 reactions to achieve optimal parameter combinations. Using Machine Learning techniques in conjunction with the experimenter's observations and intuitions, were the key to detect human error in the reaction preparation and assess model improvement over time. This reveals the importance of interactivity and explainability in a scientist-in-the-middle approach to problems being solved with techniques from Machine Learning. This allowed us to better understand the parameter's impact on the model, make a connection with existing literature, and ultimately increase the existing know-how on IVT reactions to create more efficient and flexible processes, and, ultimately to increase the global manufacturing capacity to respond to disease outbreaks, epidemics, and to pandemic scenarios like the ongoing Covid-19 global pandemic. Additionally, these results reinforce the potential of

Bayesian optimization to be applied on the optimization of biochemical reactions for industrial applications.

## Conclusion

Overall, by using Bayesian optimization, we were able to increase the mRNA IVT production twofold, up to  $12 \text{ g}_{\text{mRNA}} \cdot \text{L}^{-1}$ , in under two hours when compared to published industry standards and data reported in literature. This optimization approach proved to be cost-effective, as it only required 60 reactions to achieve optimal parameter combinations. Using Machine Learning techniques in combination with the experimental observations and intuitions, were the key to ultimately detect human error in the reaction preparation and assess model improvement over time. This reveals the importance of interactivity and explainability in a scientist-in-the-middle approach to problems being solved with techniques from Machine Learning. This allowed to better understand the IVT reaction parameter's impact on the model increasing the existing know-how on IVT reactions to create in the future more efficient and flexible processes. The results obtained can potentially increase the global manufacturing capacity of mRNA vaccines. Additionally, these results reinforce the potential of Bayesian optimization to be applied on the optimization of (bio)chemical reactions for industrial applications.

## References

- [1] Divya Hosangadi et al. "Enabling Emergency Mass Vaccination: Innovations in Manufacturing and Administration during a Pandemic". In: *Vaccine* 38.26 (2020), pp. 4167–4169. issn: 0264-410X. doi: 10.1016/j.vaccine.2020.04.037.
- [2] Benoit Hayman, Rajinder Suri, and Sai D. Prasad. "COVID-19 Vaccine Capacity: Challenges and Mitigation – The DCVMN Perspective". In: *Vaccine* 39.35 (Aug. 2021), pp. 4932–4937. issn: 0264-410X. doi: 10.1016/j.vaccine.2021.07.007.
- [3] Philip Ball. "The Lightning-Fast Quest for COVID Vaccines — and What It Means for Other Diseases". In: *Nature* 589.7840 (7840 2020), pp. 16–18. doi: 10.1038/d41586-020-03626-1. [4] Sara Sousa Rosa et al. "mRNA Vaccines Manufacturing: Challenges and Bottlenecks". In: *Vaccine* 39.16 (Apr. 2021), pp. 2190–2200. issn: 0264410X. doi: 10.1016/j.vaccine.2021.03.038.
- [5] Olivier J. Wouters et al. "Challenges in Ensuring Global Access to COVID-19 Vaccines: Production, Affordability, Allocation, and Deployment". In: *The Lancet* 397.10278 (2021), pp. 1023–1034. issn: 0140-6736, 1474-547X. doi: 10.1016/S0140-6736(21)00306-8.
- [6] R.S. Mishra et al. "Adaptive grey model (AGM) approach for judgemental forecasting in short-term manufacturing demand". In: *Materials Today: Proceedings* (2022). issn: 22147853. doi: <https://doi.org/10.1016/j.matpr.2021.12.531>.
- [7] Michael J. Chamberlin. "3 Bacterial DNA-Dependent RNA Polymerases". In: *The Enzymes*. Ed. by Paul D. Boyer. Vol. 15. Nucleic Acids Part B. Academic Press, Jan. 1982, pp. 61–86. doi: 10.1016/S1874-6047(08)60275-9.

- [8] Andrew J Geall, Christian W Mandl, and Jeffrey B Ulmer. "RNA: the new revolution in nucleic acid vaccines". In: *Seminars in immunology*. Vol. 25. 2. Elsevier. 2013, pp. 152–159. doi: 10.1016/j.smim.2013.05.001.
- [9] Damien van de Berg et al. "Quality by Design Modelling to Support Rapid RNA Vaccine Production against Emerging Infectious Diseases". In: *npj Vaccines* 6.1 (1 2021), pp. 1–10. issn: 2059-0105. doi: 10.1038/s41541-021-00322-7.
- [10] Rolf Carlson and Johan E Carlson. *Design and optimization in organic synthesis*. Elsevier, 2005. isbn: 9780080455273.
- [11] Luis Antunes et al. "E\*plora v.0: Principia for Strategic Exploration of Social Simulation Experiments Design Space". In: *Advancing Social Simulation: The First World Congress*. Ed. by Shingo Takahashi, David Sallach, and Juliette Rouchier. Springer Japan, 2007, pp. 295–306. isbn: 978-4-431-73167-2. doi: 10.1007/978-4-431-73167-2\_27.
- [12] J. Močkus. "On Bayesian Methods for Seeking the Extremum". In: *Optimization Techniques IFIP Technical Conference Novosibirsk, July 1–7, 1974*. Ed. by G. I. Marchuk. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 1975, pp. 400–404. isbn: 978-3-540-37497-8. doi: 10.1007/3-540-07165-2\_55.
- [13] Donald R Jones, Matthias Schonlau, and William J Welch. "Efficient global optimization of expensive black-box functions". In: *Journal of Global optimization* 13.4 (1998), pp. 455–492. doi: <https://doi.org/10.1023/A:1008306431147>.
- [14] Bobak Shahriari et al. "Taking the Human Out of the Loop: A Review of Bayesian Optimization". In: *Proceedings of the IEEE* 104.1 (2016), pp. 148–175. issn: 0018-9219, 1558-2256. doi: 10.1109/JPROC.2015.2494218.
- [15] Benjamin J. Shields et al. "Bayesian Reaction Optimization as a Tool for Chemical Synthesis". In: *Nature* 590.7844 (7844 Feb. 2021), pp. 89–96. issn: 1476-4687. doi: 10.1038/s41586-021-03213-y.
- [16] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. "Practical bayesian optimization of machine learning algorithms". In: *Advances in neural information processing systems* 25 (2012).
- [17] Michael D McKay, Richard J Beckman, and William J Conover. "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code". In: *Technometrics* 42.1 (2000), pp. 55–61. doi: 10.1080/00401706.2000.10485979.
- [18] Scott M Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, pp. 4768–4777.
- [19] Stephane Bancel et al. "Manufacturing methods for production of rna transcripts". US Patent App. 14/777,190. Jan. 2016.
- [20] Aniela Wochner, Tilmann Roos, and Thomas Ketterer. "Methods and means for enhancing rna production". 0040526A1. US Patent 10,837,039. Nov. 2021.
- [21] J. A. Kern and R. H. Davis. "Application of Solution Equilibrium Analysis to in Vitro RNA Transcription". In: *Biotechnology Progress* 13.6 (1997), pp. 747–756. issn: 8756-7938. doi:

10.1021/bp970094p.

- [22] Jeffrey A Kern and Robert H Davis. "Application of a fed-batch system to produce RNA by in vitro transcription". In: *Biotechnology progress* 15.2 (1999), pp. 174–184. doi: 10.1021/bp990008g.
- [23] Naomi C Breckenridge and Robert H Davis. "Optimization of repeated-batch transcription for RNA production". In: *Biotechnology and bioengineering* 69.6 (2000), pp. 679–687. doi: 10.1002/1097-0290(20000920)69:6<679::aid-bit12>3.0.co;2-6.
- [24] Jennifer S. Young, W. Fred Ramirez, and Robert H. Davis. "Modeling and Optimization of a Batch Process for in Vitro RNA Production". In: *Biotechnology and Bioengineering* 56.2 (1997), pp. 210–220. issn: 1097-0290. doi: 10.1002/(SICI)1097-0290(19971020)56:2<210::AID-BIT10>3.0.CO;2-K.
- [25] Karnyart Samnuan et al. "Design-of-Experiments In Vitro Transcription Yield Optimization of Self-Amplifying RNA". In: *bioRxiv* (2021). doi: 10.1101/2021.01.08.425833.
- [26] Leo Breiman. "Random Forests". In: *Machine Learning* 45 (2004), pp. 5–32. doi: 10.1023/A:1010933404324.
- [27] Jerome H Friedman. "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics* (2001), pp. 1189–1232. doi: 10.1214/aos/1013203451.
- [28] D. J. C. Mackay. "Gaussian processes-a replacement for supervised neural networks?" In: *Lecture notes for a tutorial at NIPS* (1997).
- [29] Andrew Gordon Wilson et al. "Deep Kernel Learning". In: *Artificial Intelligence and Statistics*. PMLR. 2016, pp. 370–378.
- [30] Subhomoi Borkotoky et al. "An in-silico glimpse into the pH dependent structural changes of T7 RNA polymerase: a protein with simplicity". In: *Scientific reports* 7.1 (2017), pp. 1–12. doi: 10.1038/s41598-017-06586-1.
- [31] Subhomoi Borkotoky and Ayaluru Murali. "The Highly Efficient T7 RNA Polymerase: A Wonder Macromolecule in Biological Realm". In: *International Journal of Biological Macromolecules* 118 (Oct. 2018), pp. 49–56. issn: 0141-8130. doi: 10.1016/j.ijbiomac.2018.05.198.
- [32] Jordana M Henderson et al. "Cap 1 Messenger RNA Synthesis with Co-transcriptional CleanCap® Analog by In Vitro Transcription". In: *Current protocols* 1.2 (2021), e39. doi: 10.1002/cpz1.39.
- [33] Samuel I Gunderson, Kenneth A Chapman, and Richard R Burgess. "Interactions of T7 RNA polymerase with T7 late promoters measured by footprinting with methidiumpropyl-EDTAiron (II)". In: *Biochemistry* 26.6 (1987), pp. 1539–1546. doi: 10.1021/bi00380a007.
- [34] Leonard C. Gosule and John A. Schellman. "DNA Condensation with Polyamines: I. Spectroscopic Studies". In: *Journal of Molecular Biology* 121.3 (May 1978), pp. 311–326. issn: 0022-2836. doi: 10.1016/0022-2836(78)90366-2.

- [35] Magali Frugier et al. "Synthetic Polyamines Stimulate in Vitro Transcription by T7 RNA Polymerase". In: *Nucleic Acids Research* 22.14 (July 1994), pp. 2784–2790. issn: 0305-1048. doi: 10.1093/nar/22.14.2784.
- [36] Y. Watanabe et al. "Differential Stimulation by Polyamines of Phage DNA-directed in Vitro Synthesis of Proteins". In: *Biochimica Et Biophysica Acta* 740.4 (1983), pp. 362–368. issn: 0006-3002. doi: 10.1016/0167-4781(83)90083-0.
- [37] Nissin Moussatché. "Polyamines Stimulate DNA-dependent RNA Synthesis Catalyzed by Vaccinia Virus". In: *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 826.2 (Nov. 1985), pp. 113–120. issn: 0167-4781. doi: 10.1016/0167-4781(85)90116-2.
- [38] Eckart Fuchs. "The Interdependence of Magnesium with Spermidine and Phosphoenolpyruvate in an Enzyme-Synthesizing System in Vitro". In: *European Journal of Biochemistry* 63.1 (1976), pp. 15–22. issn: 1432-1033. doi: 10.1111/j.1432-1033.1976.tb10201.x.
- [39] I. D. Pokrovskaya and V. V. Gurevich. "In Vitro Transcription: Preparative RNA Yields in Analytical Scale Reactions". In: *Analytical Biochemistry* 220.2 (Aug. 1994), pp. 420–423. issn: 0003-2697. doi: 10.1006/abio.1994.1360. pmid: 7526740.
- [40] John L Oakley, Judith A Pascale, and Joseph E Coleman. "T7 RNA polymerase. Conformation, functional groups, and promoter binding". In: *Biochemistry* 14.21 (1975), pp. 4684– 4691.
- [41] Monica Z. Wu et al. "Synthesis of Low Immunogenicity RNA with High-Temperature in Vitro Transcription". In: *RNA* 26.3 (Mar. 2020), pp. 345–360. issn: 1355-8382, 1469-9001. doi: 10.1261/rna.073858.119. pmid: 31900329. url: <http://rnajournal.cshlp.org/content/26/3/345> (visited on 12/05/2021).
- [42] Cristina Poveda et al. "Establishing preferred product characterization for the evaluation of RNA vaccine antigens". In: *Vaccines* 7.4 (2019), p. 131. doi: 10.3390/vaccines7040131.
- [43] World Health Organization. "Evaluation of the quality, safety and efficacy of RNA-based prophylactic vaccines for infectious diseases: regulatory considerations". In: (2020). url: <https://www.who.int/news-room/articles-detail/call-for-public-consultationevaluation-of-the-quality-safety-and-efficacy-of-rna-based-prophylacticvaccines-for-infectious-diseases-regulatory-considerations> (visited on 01/01/2022).
- [44] Marina Tusup et al. "Design of *in Vitro* Transcribed mRNA Vectors for Research and Therapy". In: *CHIMIA International Journal for Chemistry* 73.5 (May 2019), pp. 391–394. issn: 0009-4293. doi: 10.2533/chimia.2019.391.
- [45] Zeljka Trepotec et al. "Segmented poly (A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life". In: *RNA* 25.4 (2019), pp. 507–518. doi: 10.1261/rna.069286.118.
- [46] He Huang et al. "One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*". In: *Acta biochimica et biophysica Sinica* 47.4 (2015), pp. 231–243. doi: 10.1093/ abbs/gmv007.
- [47] Meredith Packer William Issa. "Methods for hplc analysis". US20210163919A1. 2021.
- [48] Tim Head et al. *Scikit-Optimize/Scikit-Optimize*. Version v0.8.1. Zenodo, Sept. 2020. doi: 10.5281/zenodo.4014775.

- [49] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Nov. 2005. isbn: 9780262256834. doi: 10.7551/mitpress/3206. 001.0001.
- [50] Christopher K. I. Williams. “Regression with Gaussian Processes”. In: *Mathematics of Neural Networks*. Ed. by Stephen W. Ellacott, John C. Mason, and Iain J. Anderson. Red. by Ramesh Sharda. Vol. 8. Operations Research/Computer Science Interfaces Series. Boston, MA: Springer US, 1997, pp. 378–382. isbn: 978-1-4613-7794-8 978-1-4615-6099-9. doi: 10.1007/978-1-4615-6099-9\_66.
- [51] Erik Štrumbelj and Igor Kononenko. “Explaining Prediction Models and Individual Predictions with Feature Contributions”. In: *Knowledge and Information Systems* 41 (2013), pp. 647–665. doi: 10.1007/s10115-013-0679-x.

## Supplementary Material

Supplementary Table S1: Optimization of mRNA production in concentration ( $g_{mRNA}.L^{-1}$ ) for each run (1-7) presented in Table 2 and the values obtained in the kinetics studies (Figure 2a) at 80% of maximum production obtained during the kinetics reactions for each run and the corresponding times (min).

Run	Optimization Studies		Kinetics Studies	
	Concentration $g.L^{-1}$	Time (min)	Concentration $g.L^{-1}$	Time (min)
1	$12.61 \pm 0.82$	263	$8.19 \pm 1.76$	175
2	$10.76 \pm 0.47$	98	$10.63 \pm 0.6$	175
3	$11.76 \pm 0.66$	120	$9.27 \pm 0.51$	175
4	$12.27 \pm 0.77$	148	$11.34 \pm 1.19$	205
5	$12.18 \pm 0.98$	121	$10.85 \pm 0.26$	115
6	$11.52 \pm 0.23$	279	$8.52 \pm 0.93$	145
7	$7.64 \pm 0.87$	240	$5.57 \pm 0.3$	235

Supplementary Table S2. Sequences in the 5'-UTR, 3'-UTR, poly-A, and the gene of interest for DNA templates EGFP, TA\_EGFP, and T7 RNAP\_EGFP.

Name	Size (bp)	Sequence
5'-UTR	45	ACTCACTATTTGTTTTTCGCGCCAGTTGCAAAAAGTGTGCCACC
3'-UTR	284	GAGAGCTCGCTTTCTTGCTGTCCAATTTCTATTTAAAGGTTCCCTTTGTTCCCTAA GTCCAACACTAAACTGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTG CCTAATAAAAAACATTTATTTTCATTGCTGCGTCGAGAGCTCGCTTTCTTGCTG TCCAATTTCTATTTAAAGGTTCCCTTTGTTCCCTAAGTCCAACACTAAACTGGGG GATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATT TTCATTGCTGCGTC
Poly-A	126	AA AAAAAAAAATGCATAA AAAAAAAAAAAAAAAAAAAAA
<b>Genes</b>		
EGFP	720	ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGTGGTGCCCATCCTGGTCGAG CTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTG CCCCTGCCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTC AGCCGCTACCCCGACCACATGAAGCAGCAGCACTTCTTCAAGTCCGCCATGCC GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAG ACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATC AAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCC GACCACTACCAGCAGAACACCCCATCGGCGACGGCCCGTGTCTGCTGCCCGAC

---

AACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGC  
GATCACATGGTCTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATG  
GACGAGCTGTACAAGTAA

TA\_EGFP 2002 ATGAACAGCAACAAAGCGATGATGGCGCGCCGCAGCGATGCGGTGCCGCGCGGC  
GTGGGCCAGATTTCATCCGATTTTCGCGGAACGCGCGGAAAACCTGCCGCGTGTGG  
GATGTGGAAGGCCGCGAATATCTGGATTTTTCGCGGCGGCATTGCGGTGCTGAAC  
ACCGGCCATCTGCATCCGCAGGTGGTGGCGGCGGTGGAAGATCAGCTGAAGAAA  
CTGAGCCATACCTGCTTTCAGGTGCTGGCGTATGAACCGTATCTGGCGCTGTGC  
GAGAAAATGAACCAGAAAAGTGCCGGGCGATTTTTCGAAGAAAACCCCTGCTGGTG  
ACCACCGGCAGCGAAGCGGTGAAAACGCGGTGAAAATTGCGCGCGCGGCGACC  
GGCCGCGAGCGGCGGATTCGCTTTACCGGCGCGGCGCATGGCCGCACCCATTAT  
ACCCTGAGCCTGACCGGCAAAGTGAACCCGTATAGCGCGGCATGGGCTGATG  
CCGGGCCATGTGTATCGCGCGCTGTATCCGTGCGCGCTGCATGGCGTGAGCGAT  
GATGAAGCGATTGCGAGCATTTCATCGCATTTTCAAGAACGATGCGGCGCCGGAA  
GATATTGCGGCGATTATTATTGAACCGGTGCAGGGCGAAGGCGGCTTTTATGCG  
GCGAGCCCGGCGTTTATGCAGCGCCTGCGCGCGCTGTGCGATGAACATGGCATT  
ATGCTGATTGCGGATGAAGTGCAGAGCGGCGCGGGCCGCACCGGCACCCGTGTT  
GCGATGGAACAGATGGGCGTGGCGGCGGATATTACCACCTTTGCGAAAAGCATT  
GCGGGCGGCTTTCCGCTGGCGGGCGTGACCGGCCGCGCGGAAGTGAATGGATGCG  
ATTGCGCCGGGCGGCTGGGCGGCACCTATGCGGGCAACCCGATTGCGTGCGCG  
GCGGCGCTGGCGGTGCTGCAGATTTTCGAACAGGAAAACCTGCTGGAGAAAAGCG  
AACAGCTGGGCGATACCTGCGCCAGGGCCTGCTGGCGATTGCGGAAGATCAT  
CCGAAAATTGGCGATGTGCGCGGCTGGGCGCGATGATTGCGATTGAACTGTTT  
GAAGAAGGCGATCATAGCCGCCCGAACGCGCGCCTGACCGCGGATATTGTGGCG  
CGCGCGCGGATAAAGGCTGATTCTGCTGAGCTGCGGCCCGTATTATAACGTG  
CTGCGCATTCTGGTGCCGCTGACCATTGAAGAAGCGCAGATTGAACAGGGCCTG  
AAAATTATTGCGGATTGCTTTAGCGAAGCGAAAACAGGCGCATGGTGAGCAAGGG  
CGAGGAGCTGTTTACCGGGGTGGTGGCCATCCTGGTTCGAGCTGGACGGCGACGT  
AAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG  
CAAGCTGACCCTGAAGTTCATCTGCACCACCGCAAGCTGCCCGTGCCCTGGCC  
CACCTCGTGACCACCTGACCTACGGCGTGCACTGCTTCAGCCGCTACCCCGA  
CCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCA  
GGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGT  
GAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT  
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACAGCCA  
CAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAA  
GATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCA  
GAACACCCCATCGGCGACGGCCCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT  
CACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCTCT  
GCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAA  
GTAA

T7 RNAP\_ 3370 ATGAACACGATTAACATCGCTAAGAACGACTTCTCTGACATCGAACTGGCTGCT  
EGFP ATCCCGTTTCAACACTCTGGCTGACCATTACGGTGAGCGTTTAGCTCGCGAACAG  
TTGGCCCTTGAGCATGAGTCTTACGAGATGGGTGAAGCAGCCTTCCGCAAGATG  
TTTGAGCGTCAACTTAAAGCTGGTGGAGTTGCGGATAACGCTGCCGCCAAGCCT  
CTCATCACTACCCTACTCCCTAAGATGATTGCACGCATCAACGACTGGTTTTGAG  
GAAGTGAAGCTAAGCGCGGCAAGCGCCGACAGCCTTCCAGTTTCTGCAAGAA  
ATCAAGCCGGAAGCCGTAGCGTACATCACCATTAAGACCACTCTGGCTTGCCTA  
ACCAGTGCTGACAATAACAACCGTTCAGGCTGTAGCAAGCGCAATCGGTGCGGCC  
ATTGAGGACGAGGCTCGCTTCGGTTCGTATCCGTGACCTTGAAGCTAAGCACTTC  
AAGAAAACGTTGAGGAACAACCAACAAGCGGTAGGGCACGTCTACAAGAAA  
GCATTTATGCAAGTTGTCGAGGCTGACATGCTCTTAAGGGTCTACTCGGTGGC  
GAGGCGTGGTCTTCGTGGCATAAGGAAGACTCTATTTCATGTAGGAGTACGCTGC  
ATCGAGATGCTCATTGAGTCAACCGGAATGGTTAGCTTACACCGCCAAAATGCT  
GGCGTAGTAGGTCAAGACTCTGAGACTATCGAACTCGCACCTGAATACGCTGAG  
GCTATCGCAACCCGTGCAGGTGCGCTGGCTGGCATCTCTCCGATGTTCCAACCT  
TGCCTAGTTCTCCTAAGCCGTGGACTGGCATTACTGGTGGTGGCTATTGGGCT

---

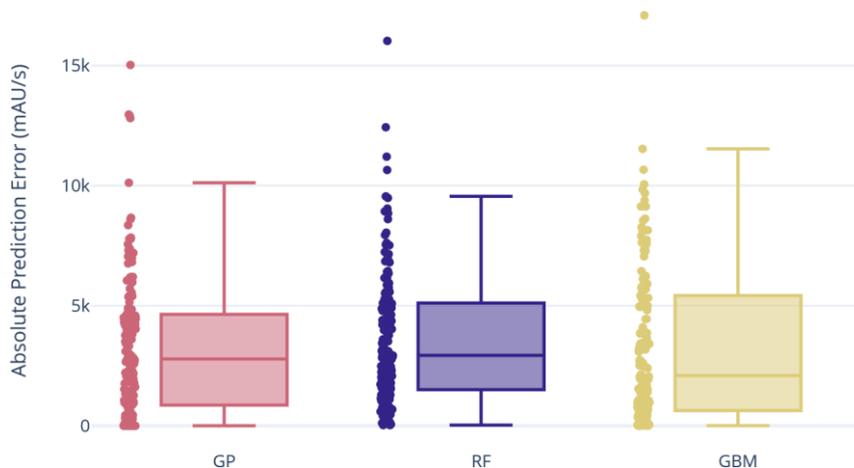
AACGGTCGTCGTCCTCTGGCGCTGGTGGTACTCACAGTAAGAAAGCACTGATG  
CGCTACGAAGACGTTTACATGCCTGAGGTGTACAAAGCGATTAACATTGCGCAA  
AACACCGCATGGAAAATCAACAAGAAAGTCCTAGCGGTGCGCAACGTAATCACC  
AAGTGGAAAGCATTGTCCGGTCGAGGACATCCCTGCGATTGAGCGTGAAGAACTC  
CCGATGAAACCGGAAGACATCGACATGAATCCTGAGGCTCTCACCGCGTGGAAA  
CGTGCTGCCGCTGCTGTGTACCGCAAGGACAGGGCTCGCAAGTCTCGCCGTATC  
AGCCTTGAGTTCATGCTTGAGCAAGCCAATAAGTTTGCTAACCATAAAGGCCATC  
TGGTTCCTTACAACATGGACTGGCGCGGTGCTGTTTACGCCGTGTCAATGTTT  
AACCCGCAAGGTAACGATATGACCAAAGGACTGCTTACGCTGGCGAAAAGGTAAA  
CCAATCGGTAAGGAAGGTTACTACTGGCTGAAAATCCACGGTGCAAACGTGTCG  
GGTGTGATAAGGTTCCGTTCCCTGAGCGCATCAAGTTCATTGAGGAAAACAC  
GAGAACATCATGGCTTGCCTAAGTCTCCACTGGAGAACACTTGGTGGGCTGAG  
CAAGATTCTCCGTTCTGCTTCCCTGCGTTCCTGCTTTGAGTACGCTGGGGTACAG  
CACCACGGCCTGAGCTATAACTGCTCCCTTCCGCTGGCGTTTGACGGGTCTTGC  
TCTGGCATCCAGCACTTCTCCGCGATGCTCCGAGATGAGGTAGGTGGTCGCGCG  
GTTAACTTGCTTCCCTAGTGAGACCGTTTACAGGACATCTACGGGATTGTTGCTAAG  
AAAGTCAACGAGATTCTACAAGCAGACGCAATCAATGGGACCGATAACGAAGTA  
GTTACCGTGACCGATGAGAACACTGGTGAATCTCTGAGAAAAGTCAAGCTGGGC  
ACTAAGGCACTGGCTGGTCAATGGCTGGCTCACGGTGTACTCGCAGTGTGACT  
AAGCGTTCAGTCATGACGCTGGCTTACGGGTCCAAAGAGTTCGGCTTCCGTC  
CAAGTGCTGGAAGATACCATTACAGCCAGCTATTGATTCGGGCAAGGGTCCGATG  
TTCCTACTAGCCGAATCAGGCTGCTGGATAACATGGCTAAGCTGATTTGGGAATCT  
GTGAGCGTGACGGTGGTAGCTGCGGTTGAAGCAATGAACACTGGCTTAAGTCTGCT  
GCTAAGCTGCTGGCTGCTGAGGTCAAAGATAAGAAGACTGGAGAGATTCTTCGC  
AAGCGTTGCGCTGTGCATTGGGTAACCTCCTGATGGTTTTCCCTGTGTGGCAGGAA  
TACAAGAAGCCTATTACAGACGCGCTTGAACCTGATGTTCCCTCGGTGAGTCCGC  
TTACAGCCTACCATTAAACACCAACAAGATAGCGAGATTGATGCACACAAACAG  
GAGTCTGGTATCGCTCCTAATTTGTACACAGCCAAGACGGTAGCCACCTTCGT  
AAGACTGTAGTGTGGGCACACGAGAAGTACGGAATCGAATCTTTTGCCTGATT  
CACGACTCCTTCGGTACCATTCCGGCTGACGCTGCGAACCTGTTCAAAGCAGTG  
CGCGAAACTATGGTTGACACATATGAGTCTTGTGATGTACTGGCTGATTTCTAC  
GACCAGTTCGCTGACCAGTTGCACGAGTCTCAATTGGACAAAATGCCAGCACTT  
CCGGCTAAAGGTAACCTGAACCTCCGTGACATCTTAGAGTCCGACTTCGCGTTC  
GCGCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGTGGTGCCCATCCTGGT  
CGAGCTGGACGGCGACGTAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGA  
GGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAA  
GCTGCCCGTGCCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTG  
CTTCAGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCAT  
GCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA  
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGA  
GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGA  
GTACAACACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGG  
CATCAAGGTGAACCTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCT  
CGCCGACCACTACCAGCAGAACACCCCATCGGGCAGGGCCCCGTGCTGCTGCC  
CGACAACCACTACCTGAGCACCAGTCCGCCCTGAGCAAAGACCCCAACGAGAA  
GCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGG  
CATGGACGAGCTGTACAAGTAA

Supplementary Table S3: Primer sequence used in DNA template production for all the three templates.

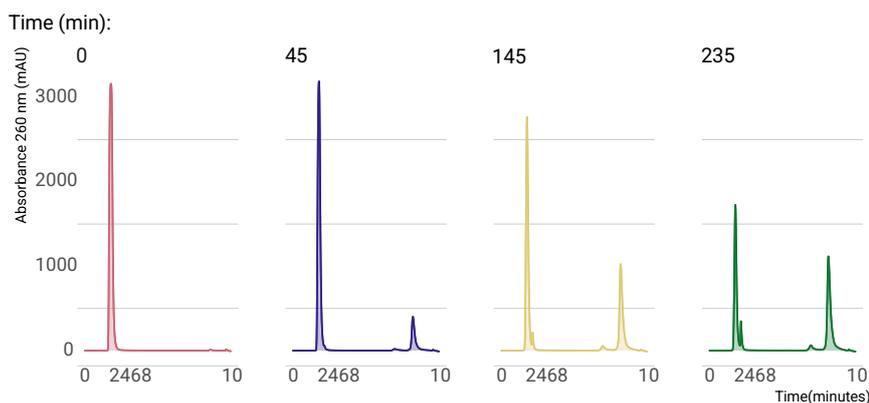
Primers	Sequence
Forward	TAA TAC GAC TCA CTA TAG GGA CT CAC TAT TTG TTT T
Reverse	TTT TTT TTT TTT TTT TTT TTT TTT TTT ATG CA

Supplementary Table S4: Established RP-HPLC gradient method. The binding corresponds to the mobile phase A composed of 1 X TAE, and the elution corresponds to the mobile phase B composed of 1 X TAE, 25% acetonitrile (v/V). Adapted from [1].

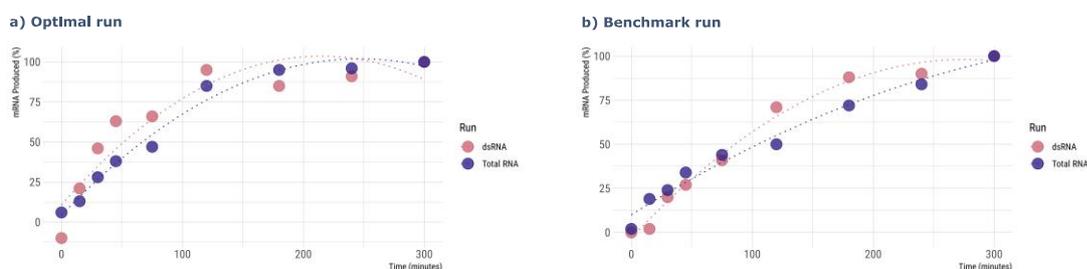
Time (min)	Binding %	Elution %	Flow mL.min <sup>-1</sup>
0	100	0	0.2
1	100	0	0.2
1.5	100	0	0.25
2	94	6	0.35
6	76.5	23.5	0.4
7	0	100	0.4
10	0	100	0.4
11	100	0	0.4
20	100	0	0.4



Supplementary Figure S1: Comparison of Gaussian Process (GP), Random Forests (RF) [2], and Gradient Boosting Machines (GBM)[3] surrogate models using leave-one-out cross validation. The median error for the GP of approximately 3000 mAU.s<sup>-1</sup> corresponds to 1.8 g<sub>mRNA</sub>.L<sup>-1</sup> of error.



Supplementary Figure S2: Chromatographic profiles of the RP-HPLC runs performed with samples of the parameter setup 5 of the kinetics runs at different time point. The peak at retention time of 2 minutes corresponds to the impurities (NTPs, enzymes), and the peak at 8 minutes corresponds to the mRNA synthesized.



Supplementary Figure S3: Percentage of total mRNA and dsRNA produced in as a function of reaction time considering 100% the highest mRNA concentration produced for each set of runs for both total. a) Total RNA and dsRNA produced using optimal conditions (Run 5). b) Total RNA and dsRNA produced using benchmark conditions [4]. A 2<sup>nd</sup> order polynomial function was used as a trendline for visualisation purposes. dsRNA was obtained by digesting each sample with Rnase T1 and analysed by HPLC as described in methods.

## References

- [1] Meredith Packer William Issa. "Methods for hplc analysis". US20210163919A1. 2021.
- [2] Leo Breiman. "Random Forests". In: *Machine Learning* 45 (2004), pp. 5–32. doi: 10.1023/A: 1010933404324.
- [3] Jerome H Friedman. "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics* (2001), pp. 1189–1232. doi: 10.1214/aos/1013203451.
- [4] Stephane Bancel et al. "Manufacturing methods for production of rna transcripts". US Patent App. 14/777,190. Jan. 2016.

# Chapter VI - Comprehensive evaluation of T7 promoter for enhanced yield and quality in mRNA production

---

**Key words:** mRNA; DNA template; Optimization; dsRNA; T7 promoter

This chapter has been published as: RSari, Yustika, Sara Sousa Rosa, Jack Jeffries, and Marco PC Marques. "Comprehensive evaluation of T7 promoter for enhanced yield and quality in mRNA production." *Scientific Reports* 14, no. 1 (2024): 9655.

## Introduction

Vaccines are pivotal in mitigating severe health complications and hold the capacity to control the spread of infectious diseases, potentially eradicating these in populations[1,2]. Over the long term, the savings in healthcare expenses and reduction in mortality rates can prove economically advantageous [1,3,4]. After years of dedicated research on vaccine technology, the mRNA vaccine received full approval in 2020 to tackle COVID-19 pandemic[5,6] with the BNT162b2 mRNA vaccine (Pfizer-BioNTech), achieved WHO emergency use authorization less than a year after the official pandemic declaration[7,9]. Between the date of approval and November 2023, approximately  $1.18 \times 10^{12}$  doses of Pfizer-BioNTech vaccines, and  $3.27 \times 10^{11}$  of the Moderna vaccines were administered worldwide[10]. With the remarkable success against SARS-CoV-2, research in mRNA vaccines is expanding and intensifying against other infectious diseases, such as Zika, HIV[11,12], and a spectrum of cancers[13]. Despite this, the cost of these vaccines is still prohibitive for LMICs, mainly driven by the cost of goods[14].

The prominent feature of mRNA vaccines is their rapid manufacturing which can be swiftly developed using a pathogen's gene sequences[15]. The mRNA vaccine is designed to mimic eukaryotic mRNA, and is composed of a 5' cap, an open reading frame (ORF) encoding a specific antigen, untranslated regions (5'UTR and 3'UTR), and a poly(A) tail[16,17]. Inherently, these vaccines are precise and simple, since after administration only the antigen encoded by the targeted gene[18] is translated. Furthermore, their safety profile is high since the mRNA administered has a transient expression[6,19,20]. However, specific modifications to the mRNA nucleotide improve the stability and translational efficiency, prolonging its in vivo half-life, essential for effectiveness as a vaccine[15,21]. mRNA is produced in a cell-free in vitro transcription reaction from a linear DNA template, using RNA polymerase as a catalyst and nucleoside triphosphates (NTPs) as co-substrates[6], apart from other reaction components[22,23]. Typical production titres are between 2 and 5 g L<sup>-1</sup>[24-27] but recent studies have shown that titres can be increased to 12 g L<sup>-1</sup> in batch or fed-batch mode[28,29].

T7 RNA polymerase (T7 RNAP) has been predominantly used for IVT reactions[30,31], due to the high fidelity displayed, and consists of a single subunit and is highly processive. The transcription process can however generate double-stranded RNA (dsRNA) contaminants through different mechanisms (e.g. random priming of abortive transcripts[32], antisense transcription[33], turn-around transcription and self-primed extension of product RNA[34-36]) which hinder mRNA translation efficiency[37] and can ultimately compromise vaccine safety[33,38,39]. Therefore, T7 RNAP efficiency and activity are continuously optimised, such as the development of mutant versions that reduce abortive products and immunostimulatory byproducts[40,41]. Additionally, modifying the T7 promoter region facilitates an optimal

interaction between T7 RNAP and the DNA template, thereby facilitating the initiation and elongation of transcription, resulting in increased mRNA production[42,43]. Extensive research has been conducted to assess the impact of various T7 promoter variants aimed at increasing total mRNA produced, such as performing modifications in different T7 promoter regions—from the core region[44] to upstream[45] and downstream[42] regions. Nevertheless, these approaches did not explore the T7 RNAP promoter optimisation focusing on the production of mRNA vaccines in a transcription system.

In this contribution, we explored the effect of modifying DNA templates for the synthesis of mRNA in terms of process yield and quality. Site-directed mutagenesis was used in the T7 promoter region with the transcription performance markedly enhanced. The AT-rich insertion in the downstream region of the T7 promoter allowed for a notable increase in mRNA titres compared to the wildtype T7 promoter, reaching a maximum of 14 g L<sup>-1</sup> in approximately 2 h. The mRNA titres up to 12 g L<sup>-1</sup> were also achieved in 45 min of IVT reaction, thereby reducing the required reaction time. mRNA quality was increased by minimising the dsRNA concentration as an undesirable byproduct. The results obtained outperformed the wildtype T7 promoter by decreasing the dsRNA production by up to 30%. A decrease was also observed with the increase in template size, but less significant (up to 20%). The modifications the promoter sequence did not alter significantly the initiation of the translation within the cells. The results highlighted the potential of an AT-rich sequence in the downstream region of the T7 promoter as a strategic modification to improve the quantity and quality of mRNA production via in vitro transcription, increasing the cost-effectiveness of mRNA manufacturing[14,23].

## Material and methods

Unless otherwise stated, all chemicals and reagents were purchased from ThermoFisher Scientific (UK).

### Template construction for mRNA synthesis

#### Template design

The mRNA template consists of the EGFP gene (GenBank Accession #AAB02572.1) flanked by two untranslated regions (5'-UTR and 3'-UTR) and followed by a poly-A sequence. The 5'-UTR comprises three elements: the wildtype or mutant promoter of T7 RNA polymerase, a binding site of eukaryotic translation initiation factor eIF4G, and a Kozak consensus sequence[16,23]. The 3'-UTR utilises two tandem repeats of 3'-UTR from the human  $\beta$ -globin gene. The poly-A sequence (120 bp) is segmented with a 6 bp spacer[46]. Additional templates were assembled by fusing the EGFP gene with the *Klebsiella pneumoniae* transaminase gene (GenBank Accession #AF074934.1), and the EGFP gene with the T7 RNA polymerase gene (GenBank Accession #NP\_041960.1). Sequences used in this study are presented in

Supporting Information Table S1. All the mRNA templates are inserted in a pUC57 plasmid vector with kanamycin resistance.

## **Promoter modification and plasmid construction**

### **Site-directed mutagenesis**

Sited-directed mutagenesis was performed to mutate or add downstream and upstream insertion in the promoter region using Q5<sup>®</sup> Site-Directed Mutagenesis Kit (New England Biolabs, UK). A plasmid control, comprising of mRNA template with wildtype T7 promoter adapted from Rosa et al.[23] was used as the template for mutagenesis. The amplification of reaction mix (1 × Q5<sup>®</sup> Hot Start High-Fidelity Master Mix, 0.5 μM forward primer, 0.5 μM reverse primer, and 25 ng plasmid template, and Gibco™ Water for Injection, WFI) was performed through touchdown polymerase chain reaction (TD-PCR) using Applied Biosystems™ Veriti™ 96-Well Thermal Cycler (ThermoFisher Scientific, UK) with the following detailed cycle conditions: an initial denaturation step at 98 °C for 30 s, 10 cycles of 10 s at 98 °C and annealing at 66–57 °C (for samples: T7#4, T7c62, T7c62\_T7#4, and T7DI\_1 to T7DI\_11) or 70–61 °C (for samples: T7Max and T7Max\_T7#4) for 30 s. The annealing temperature decreased 1 °C per cycle and an extension step was performed at 72 °C for 30 s per kb. This was followed by 20 cycles of 10 s at 98 °C; annealing at 57.5 °C (for samples: T7#4, T7c62, T7c62\_T7#4, and T7DI\_1 to T7DI\_11) or 61.5 °C (for samples: T7Max and T7Max\_T7#4) for 30 s; and extension at 72 °C for 30 s per kb. The final extension was executed at 72 °C for 2 min. Afterwards, 1 μL of TD-PCR products were treated with 1 × Kinase, Ligase, and DpnI (KLD) enzyme mix and buffer (New England Biolabs, UK), and adjusted with WFI to a final volume of 10 μL. The KLD mix was incubated for 5 min at 21 °C and 5 μL of the mix was used for transformation using a heat shock method. All plasmids and primers used in this study are presented in Supporting Information Table S2 and S3, respectively.

### **Gibson assembly**

Using the Gibson assembly method, two plasmids, pT7wt\_TA\_EGFP and pT7wt\_T7 RNAP\_EGFP, were obtained. This was achieved by separately integrating the genes for *K. pneumoniae* transaminase (TA) and T7 RNA polymerase (T7 RNAP) into the pT7wt\_EGFP template. TA and T7 RNAP genes were amplified and isolated from pET29A\_TA and pET29A\_T7 RNAP using PCR with specific primers. The linearisation of vector pT7wt\_EGFP also performed through PCR. All the PCR reactions were performed using high-fidelity VeriFi™ DNA Polymerase, VeriFi™ Buffer, and VeriMax Enhancer (PCR Biosystems, UK). The PCR products of linear vector pT7wt\_EGFP, and isolated TA and T7 RNAP genes were analysed using agarose-gel electrophoresis and the gel containing correct sizes of DNA bands were further isolated and purified. The purified linear vector pT7wt\_EGFP, and the insert (TA gene or

T7 RNAP gene) were assembled using Gibson Assembly Master Mix (New England Bioscience, UK). For each reaction, the same mass (0.1 pmol) of insert and vector were mixed with 10  $\mu\text{L}$  of 2  $\times$  Gibson Assembly Master Mix and adjusted with WFI to a total volume of 20  $\mu\text{L}$ . The reactions were performed at 50  $^{\circ}\text{C}$  for 2 h. Following the incubation, 2  $\mu\text{L}$  of the assembly products were subsequently transformed to *E. coli* NEB 10-beta (New England Biolabs, UK).

### **Molecular cloning**

Chemically competent *E. coli* NEB 10-beta cells (New England Biolabs, UK) were prepared by the calcium chloride method and used for routine transformation. Transformation of plasmids was performed using the heat-shock method. Transformed cells were plated on Luria–Bertani agar media (25 g/L Miller LB broth (Sigma-Aldrich) with 15 g/L culture media agar (MP Biomedicals, USA)) with kanamycin (50  $\mu\text{g}/\text{mL}$ ) and overnight incubated at 37  $^{\circ}\text{C}$ . Colony PCR was performed using high-fidelity VeriFi™ DNA Polymerase with VeriFi™ Buffer and VeriMax Enhancer (PCR Biosystems, UK). Plasmid DNA was purified using GeneJET Plasmid Miniprep Kit, following the protocol by the manufacturer.

### **Plasmid verification and sequencing**

NanoDrop™ One Microvolume UV–Vis Spectrophotometer (ThermoFisher Scientific, UK) was used to measure the concentration of purified plasmids. Purified plasmids were digested using EcoRI (New England Biolabs, UK) and LguI/SapI with CutSmart® buffer (New England Biolabs, UK) for one hour incubation at 37  $^{\circ}\text{C}$ , followed by inactivation at 65  $^{\circ}\text{C}$  for 20 min. Approximately 100 ng/ $\mu\text{L}$  of the purified plasmids were Sanger sequenced (Eurofins Genomics, UK).

### **Agarose gel electrophoresis**

To analyse the PCR and digestion products, 1% (w/v) of agarose (Sigma-Aldrich, USA) was prepared using 0.5  $\times$  TBE buffer (45 mM Tris–borate and 1 mM EDTA), Invitrogen SYBR® Safe DNA Gel Stain (1:10,000 dilution), and run at 100 V for one hour. Purple Gel Loading Dye (New England Biolabs, UK) was used to load the samples into the gel and 1 kb Plus DNA Ladder (New England Biolabs, UK) for analysis.

### **mRNA and dsRNA synthesis**

#### **Template production**

DNA templates for IVT reactions were produced through touchdown polymerase chain reaction (TD-PCR). The TD-PCR reaction mixture contained between 200 and 250 ng  $\text{mL}^{-1}$  of plasmid, 0.4  $\mu\text{M}$  of forward and reverse primers, 1  $\times$  VeriFi™ Buffer, 1  $\times$  VeriMax Enhancer, and 0.02 U  $\mu\text{L}^{-1}$  high-fidelity VeriFi™ DNA Polymerase (PCR Biosystems, UK). The reaction mixture was prepared to a total volume of 500  $\mu\text{L}$  and split into 50  $\mu\text{L}$  reaction per tube. Detail of plasmids (as templates) and primers are found in S2 and S3, respectively. Supportive Information Table

S. The TD-PCR was performed using a Applied Biosystems™ Veriti™ 96-Well Thermal Cycler (ThermoFisher Scientific, UK) with an initial denaturation step at 95 °C for 1 min, 10 cycles of 15 s at 95 °C; annealing at 60–51 °C (for samples: T7#4 and T7DI\_1 to T7DI\_11) or 66–57 °C (for samples: T7wt, T7Max, T7c62, T7Max\_T7#4, and T7c62\_T7#4) for 30 s with annealing temperature decreased 1 °C per cycle. Extension was performed at 72 °C for 30 s per kb, followed by 20 cycles of 15 s at 95 °C; annealing at 51.5 °C (for samples: T7#4 and T7DI\_1 to T7DI\_11) or 58 °C (for samples: T7wt, T7Max, T7c62, T7Max\_T7#4, and T7c62\_T7#4) for 30 s and extension at 72 °C for 30 s per kb. The final extension was executed at 72 °C for 2 min. The TD-PCR product was purified using GeneJET PCR Purification Kit (ThermoFisher Scientific, UK) following the manufacturer's instructions. A 10 × concentrated TD-PCR product was obtained from the purification step and further quantified using NanoDrop™ One Microvolume UV–Vis Spectrophotometer (ThermoFisher Scientific, UK).

### ***In vitro* transcription (IVT) reactions**

The IVT reaction conditions were adapted from Rosa et al. (2022)[23]. The IVT reaction mixture contained 89 nM of linear DNA template (purified TD-PCR product), 7.75 mM of each NTP (ATP, GTP, CTP, and UTP), 5.3 mM of DTT, 49 mM of Mg-acetate (Santa Cruz Biotechnology, USA), 40 mM pH 6.5 Tris buffer, 2.3 mM of spermidine, 0.008 U  $\mu\text{L}^{-1}$  of *Saccharomyces cerevisiae* inorganic pyrophosphatase, 1.48 U  $\mu\text{L}^{-1}$  of RiboShield™ RNase Inhibitor (PCR Biosystems, UK), 7.7 U  $\mu\text{L}^{-1}$  of bacteriophage T7 RNA polymerase, and was made up to a final volume of 20  $\mu\text{L}$  (for sample measurements) or 100  $\mu\text{L}$  (for the calibration curve) with water for injection (WFI). The IVT was performed at 43 °C for 2 h. The mRNA produced from IVT was quantified using reverse-phase high-performance liquid chromatography (RP-HPLC) described in Section *mRNA quantification*.

To produce dsRNA for the calibration curve, a subsequent incubation was performed after IVT to facilitate the dsRNA hybridisation. The 100  $\mu\text{L}$  of dsRNA IVT product was diluted to 200  $\mu\text{L}$  with water for Injection (WFI) and incubated at gradient temperature 85 °C to 23 °C for 2 min at each temperature.

### **RNA purification**

The mRNA purification for the calibration curve was performed using MEGAclean™ Transcription Clean-Up Kit as instructed by the manufacturer with slight modifications. The 100  $\mu\text{L}$  of IVT product was treated with 2  $\mu\text{L}$  of TURBO™ DNase and incubated for 15 min at 37 °C. For dsRNA purification, after the hybridisation incubation step, 2  $\mu\text{L}$  of TURBO™ DNase and 2  $\mu\text{L}$  of RNase T1 were added to 200  $\mu\text{L}$  of diluted dsRNA IVT product and incubated at 37 °C for 15 min. The 350  $\mu\text{L}$  of binding solution and 250  $\mu\text{L}$  of 100% v/v ethanol were added to the samples and then loaded into the filter cartridge for centrifugation (15,000 × g, 1 min, 21 °C).

The filter was washed with 500  $\mu\text{L}$  of wash solution and centrifuged under the same condition in the previous step twice. The mRNA was eluted with 50  $\mu\text{L}$  of elution buffer, followed by 5 min incubation at 65 °C and centrifugation at 15,000  $\times g$  for 1 min at 21 °C. The elution step was repeated in the same previous condition. The 100  $\mu\text{L}$  of purified mRNA was further precipitated with 10  $\mu\text{L}$  of 5 M ammonium acetate and 275  $\mu\text{L}$  of 100% v/v ethanol, and then overnight incubated at – 20 °C. Samples were centrifuged (top speed) at 4 °C for 15 min. The supernatant was discarded, and the obtained pellet was air-dried to remove the remaining ethanol. The pellet was resuspended in 40  $\mu\text{L}$  of elution buffer. The concentrated purified mRNA sample was then quantified using NanoDrop™ One Microvolume UV–Vis Spectrophotometer (ThermoFisher Scientific, UK) and RP-HPLC (*mRNA quantification*).

### **mRNA capping for expression studies**

One-pot cap-1 reactions were performed using Faustovirus Capping Enzyme (New England Biolabs, UK) and Cap 2'-O-methyltransferase (New England Biolabs, UK). Briefly, 50  $\mu\text{g}$  of purified mRNA was added to a reaction containing 1X FCE capping buffer, 0.5 mM GTP, 2 mM S-adenosylmethionine, 1  $\mu\text{L}$  of Rnase inhibitor, 1 U  $\mu\text{L}^{-1}$  of Faustovirus Capping Enzyme, 4 U  $\mu\text{L}^{-1}$  of Cap 2'-O-methyltransferase, and WFI water to a final volume of 50  $\mu\text{L}$ . The samples were incubated at 37°C without shaking for 2 h. Afterwards, the samples were purified as described in Section *RNA purification*. The pellets were resuspended in 10  $\mu\text{L}$  of WFI water, and quantified using NanoDrop™ One Microvolume UV–Vis Spectrophotometer (ThermoFisher Scientific, UK).

### **Analytical methods**

#### **mRNA quantification**

##### **RP-HPLC**

The total mRNA concentration was quantified using the established RP-HPLC gradient method adapted from Issa and Packer [47] and Rosa et al. [23]. An UltiMate™ 3000 UHPLC System with a 2.1  $\times$  100 mm DNAPac™ RP column and a 3  $\times$  10 mm guard column was used. 5  $\mu\text{L}$  of each sample, diluted 6 times, was run in the pre-equilibrated column with TAE buffer absorbance measured at 260 nm. Elution was achieved by a gradient elution using TAE buffer with 25% acetonitrile. The runs were performed at 80 °C with the following conditions: After injection, the column is washed for 1 min and 0.2 mL  $\times$  min<sup>-1</sup>. The flow is gradually increased to 0.25 mL  $\times$  min<sup>-1</sup> for 30 s. A gradient to 6% of elution buffer and 0.35 mL  $\times$  min<sup>-1</sup> is applied for 30 s, followed by a gradient to 76.5% of elution buffer at 0.4 mL  $\times$  min<sup>-1</sup> for 4 min, and a final gradient to 100% elution buffer for 1 min. The column is then washed with 100% elution buffer for 3 min, and re-equilibrated with the binding buffer for 6 min.

#### **Agarose gel electrophoresis**

A 2% (w/v) of agarose with 0.5 × TBE buffer (45 mM Tris–borate and 1 mM EDTA) and 5.5 mM of magnesium chloride was prepared and pre-stained with Invitrogen SYBR<sup>®</sup> Safe DNA Gel Stain (1:10,000 dilution). The gel was loaded with 1.5 µL of mRNA sample diluted in WFI into a final volume of 10 µL and 2 µL of 6 × purple Loading Dye (New England Biolabs, UK). A 5 µL of 1 kb Plus DNA Ladder (New England Biolabs, UK) was used as the molecular marker. The electrophoresis was run at 100 V for 75 min using 0.5 × TBE buffer containing 5.5 mM of MgCl<sub>2</sub>. The gel was visualised using Amersham™ Imager 600 (GE Healthcare, UK).

### **dsRNA quantification**

The dsRNA concentration was measured using the RP-HPLC method adapted from Issa et al. [47] described in Section *mRNA quantification*. Samples of 10 µL were treated with 0.5 µL of RNase T1 and incubated at 37 °C for 15 min to digest the ssRNA (Supplementary Information Figure S2).

### **Protein expression**

EGFP expression was performed using the 1-Step Human Coupled IVT kit. Capped mRNA (Section *mRNA capping for expression studies*) was diluted to a final concentration of 1 g L<sup>-1</sup> and 2 µL were added to the reaction mixture. Positive and negative controls were the kit GFP control and WFI water, respectively. The samples were incubated for 6 h at 30 °C without agitation. The samples were diluted 1:2 with WFI water to a final volume of 50 µL and the EGFP fluorescence was measured using Infinite Pro 200 (Tecan, USA).

### **Statistical analysis**

Statistical analyses were conducted using GraphPad Prism (version 10.0.2): one-way ANOVA with the Brown-Forsythe test (determine the standard deviation for duplicate IVT experiments for mRNA concentration and dsRNA level across promoter variants), Dunnett's and Tukey's multiple comparisons test (compare the total mRNA concentration and the dsRNA level across promoter variants) and two-way ANOVA followed by Tukey's multiple comparisons test (analysing and comparing the mRNA concentration and dsRNA level in T7 promoter variants over a range of template sizes).

## **Results**

### **T7 promoter modifications**

In this study, we evaluated a total of 16 different T7 promoter variants, detailed in Table 1. For comparison, three reported T7 promoter variants were used as a positive control, namely T7#4[42], T7Max[43], and T7c62[44], were selected based on their improved transcriptional performance relative to the wild-type T7 promoter both in IVT systems and cell-free

transcription/translation systems. Each control exhibits a modification in a specific region of the T7 promoter, namely in the upstream, downstream, or within the core promoter region. Particularly, the T7c62 variant carries nucleotide substitutions within the core promoter region, at positions -4 (A substituted for T), -1 (C for A), and +2 (A for G). This variant was reported to demonstrate approximately 2-fold higher protein expression level than wildtype T7 promoter[44] in cells. The specific sequences in both upstream (-22 to -18)[43,45] and downstream (+4 to +8)[42] regions of the T7 promoter also have been documented to improve the transcription levels. T7Max incorporates an upstream element (AATTC) at positions -22 to -18, which has been linked to increased gene expression in *in vitro* systems[43]. T7#4 contains an AT-rich downstream element, 'ATAAT', at positions +4 to +8[42]. This promoter was employed as a representative variant containing an AT-rich downstream element, previously demonstrated to improve T7 promoter activity, with amplicon abundances increasing by over a 5-fold range compared to GC-rich combinations[42]. T7Max\_T7#4 and T7c62\_T7#4 promoters were also constructed to evaluate the synergistic effects of incorporating modifications from disparate regions. Furthermore, a library of AT-rich downstream variants that are composed by 11 variants, T7DI\_1 to T7DI\_11, and that contains different AT-rich combinations at positions +4 to +8, was created. The variants T7DI\_1 to T7DI\_11 aimed to assess the effects of alternate AT-rich sequences and evaluate the sequence-specificity of these elements on modulating transcriptional activity. The hypothesis is that AT-rich sequences can facilitate DNA unwinding during the initiation of the transcription.

Table 1. T7 promoter variants used in this study. The sample ID for the T7 promoter variants T7#4[42], T7Max[43], and T7c62[44] references the original. T7Max\_T7#4 and T7c62\_T7#4 combine both T7#4 sequence with the T7Max and T7c62 modifications. 11 promoter variants in this study are designated with the prefix "T7DI," followed by a numerical identifier. T7DI corresponds to T7 promoter with a specific downstream sequence, while the accompanying number indicates the unique arrangement of AT-rich sequences located at positions +4 to +8 in the downstream region.

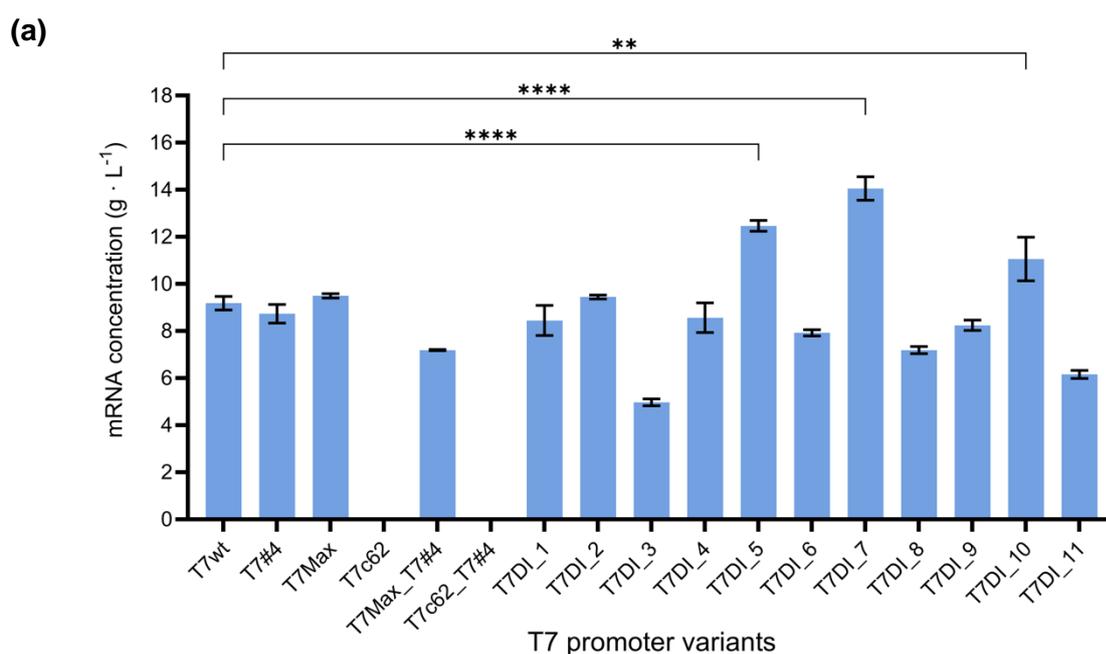
Name	Recognition Region															Initiation Region														
	-22	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-5	-1	+1	+2	+3	+4	+5	+6	+7	+8					
T7wt (wildtype)						T	A	A	T	A	C	G	A	C	T	C	A	T	A	G	G	G								
T7#4						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	A	T	A	A	T
T7Max	A	A	T	T	C	T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	A				
T7c62						T	A	A	T	A	C	G	A	C	T	C	A	C	A	T	C	C	G	G	A	G				
T7Max_T7#4	A	A	T	T	C	T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	A	T	A	A	T
T7c62_T7#4						T	A	A	T	A	C	G	A	C	T	C	A	C	A	T	C	C	G	G	A	T	A	A	A	T
T7DI_1						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	A	A	T	A	A
T7DI_2						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	T	A	A	A	A
T7DI_3						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	A	A	T	T	A
T7DI_4						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	A	T	T	A	A
T7DI_5						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	T	T	A	A	A
T7DI_6						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	A	T	T	T	A
T7DI_7						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	T	T	T	A	A
T7DI_8						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	A	T	T	A	T
T7DI_9						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	A	T	A	T	T
T7DI_10						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	T	A	T	A	T
T7DI_11						T	A	A	T	A	C	G	A	C	T	C	A	C	T	A	T	A	G	G	G	T	T	T	T	A

## Impact of T7 promotor modification on IVT performance

### Specific AT-rich downstream elements

The linear DNA templates encoding for EGFP with the T7 promoter variant were used for *in vitro* transcription (Figure 1). Several promoter variants from the library designed that contain specific AT-rich downstream sequences significantly outperformed the wild-type T7 promoter (T7wt), producing at least 10 g·L<sup>-1</sup> within 2 h (Figure 1a). T7DI\_7 achieved the highest mRNA concentration at 14.05 ± 0.5 g·L<sup>-1</sup>, marking approximately 1.5-fold increment relative to T7wt (9.18 ± 0.29 g·L<sup>-1</sup>). This was followed by T7DI\_5 and T7DI\_10 with 1.4-fold and 1.2-fold mRNA yield compared to T7wt. Several promoter variants, i.e. T7DI\_1, T7DI\_2, T7DI\_4, and T7DI\_9, produced similar amounts of mRNA as T7wt, ranging from 8.24 to 9.45 g·L<sup>-1</sup> (P>0.05). In contrast, the other variants exhibited lower mRNA yield compared to T7wt, with T7DI\_3 showing the lowest concentration at 4.97 ± 0.15 g·L<sup>-1</sup> (P<0.05) (Figure 1a).

The control T7 promoter variants T7#4[42] and T7Max[43] exhibited no significant differences in mRNA production compared to T7wt (P> 0.05). However, the T7Max\_T7#4 promoter, which combines modifications from both upstream (-22 to -18) and downstream (+4 to +8) regions produced lower mRNA levels (7.19 ± 0.02 g·L<sup>-1</sup>) indicating the possible counteractive effects of combined elements (Figure 1a). The T7c62[44] and T7c62\_T7#4 promoters produced no mRNA (Figure 1a, 1b). Although T7c62 was previously reported to demonstrate higher expression level[44], no mRNA was produced after 2 hours of IVT in this study. A similar result was also reported with no activity for the RNA broccoli aptamer transcribed using the T7c62 promoter variant[43].



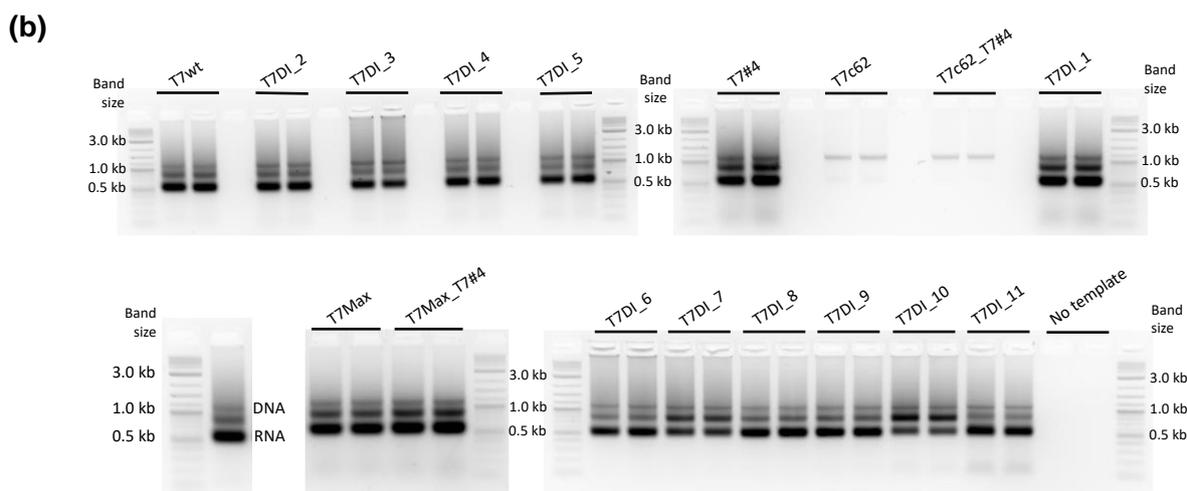


Figure 1. The mRNA production profiles of the different T7 promoter variants. IVT reactions were performed according to Rosa et al. (2022)<sup>23</sup> at 43°C for 2 h. (a) Total mRNA produced and quantified by RP-HPLC. Error bars represent the standard deviation for  $n = 2$ . One-way ANOVA with Brown-Forsythe test and followed by Dunnett's multiple comparisons test, \*\* and \*\*\*\* denote p-values of 0.0012 and <0.0001 respectively. (b) Agarose-gel electrophoresis analysis of IVT products. The mRNA produced is indicated by the 600 nt RNA band while the 1.2 kb band represents the linear DNA template used. Gels do not present normalised mRNA quantity, and the variances observed are a result of independent IVT reactions. The cropped gel images are displayed to improve clarity and conciseness. The original gels are available in Supporting Information Figure S1.

### Kinetic analysis of T7 promoter modification and impact on production yields

Three promoter variants were selected for subsequent kinetic analyses based on the production profile during the screening phase, namely final concentration of approximately  $10 \text{ g}\cdot\text{L}^{-1}$  after 2 h (Figure 1). The T7DI\_5, and T7DI\_7 promoters achieved higher production yields compared to T7wt, and T7DI\_2 presented the least variance in production yield. The reaction profile shows that T7DI\_2 and T7DI\_5 reached a sheiling of total mRNA produced in 45 minutes, achieving maximum concentrations of  $10.67 \pm 0.06$  and  $10.01 \pm 0.28 \text{ g}\cdot\text{L}^{-1}$ , respectively. After 120-minute reaction time, no significant differences in total mRNA concentrations were observed for T7DI\_2 and T7DI\_5 promoters (Figure 2). In contrast, T7DI\_7 exhibited a lower production rate than T7DI\_2 and T7DI\_5, but still outperforming T7wt. Initial kinetic measurements within the first 15 minutes indicated similar production rates among T7DI\_2, T7DI\_5, and T7DI\_7. After the 15-minute time point, T7DI\_7 showed a diminished rate relative to T7DI\_2 and T7DI\_5, followed by a gradual increase in mRNA production, peaking at the 120-minute mark. In contrast, no such increase in mRNA yield was observed for T7DI\_2 and T7DI\_5 post-45 minutes. The findings demonstrate that, following a 2 h incubation in IVT, the promoter variants produced comparable quantities of mRNA regardless of the different transcriptional rates. This time frame is in line observations obtained with optimised IVT conditions[23,48].

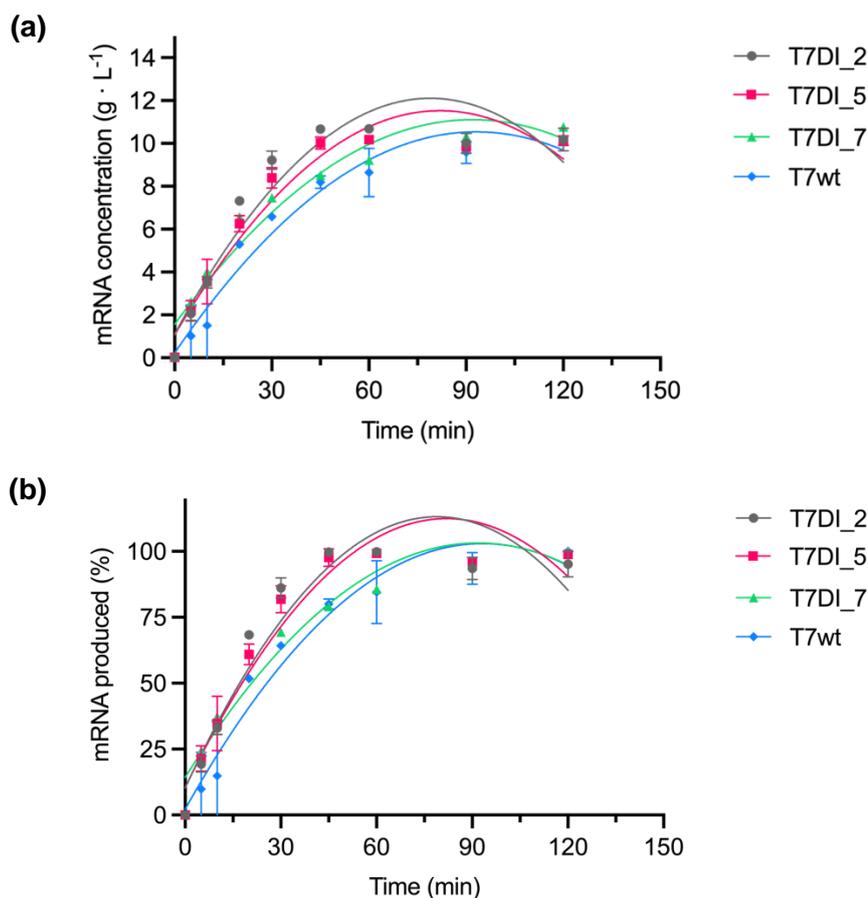


Figure 2. The kinetics analysis of T7 promoter variants. IVT reactions were performed according to Rosa et al. (2022)[23] 43°C for 2 h. (a) mRNA concentration (g·L<sup>-1</sup>) for a time course of 2 h. (b) mRNA produced as a function of reaction time, with the maximum mRNA concentration achieved corresponding to 100%. Error bars represent the standard deviation for n = 2 replicates. The lines correspond to a second order polynomial (quadratic) function.

### T7 promoter variants with specific AT-rich downstream elements produce less dsRNA byproduct

Table 2. The comparison of total mRNA and dsRNA concentrations for three T7 promoter candidates. The IVT reactions were prepared following the method described by Rosa et al. (2022)[23] and the reactions were conducted at 43°C for 2 hours.

Promoter Candidates	Total mRNA (g·L <sup>-1</sup> )	dsRNA/mRNA total (mg·g <sup>-1</sup> )
T7wt (wildtype)	9.18 ± 0.29	181.34 ± 1.32
T7DI_2	9.45 ± 0.08	96.51 ± 1.46
T7DI_5	12.47 ± 0.23	138.47 ± 1.08
T7DI_7	14.05 ± 0.50	92.22 ± 2.84

The evaluation of T7 promoter candidates focused on achieving higher mRNA yields while reducing dsRNA impurities (Table 2). Prior to quantification, the IVT products were treated with RNase T1 to degrade the single-strand RNA (ssRNA). The T7 promoter variants T7DI\_2, T7DI\_5, and T7DI\_7, which contain specific AT-rich downstream elements, produced

significantly lower amounts of dsRNA byproduct compared to T7wt. Among these, T7DI\_7 showed the minimal residual dsRNA at a concentration of  $0.63 \pm 0.07 \text{ g}\cdot\text{L}^{-1}$ , a 39.4% reduction in comparison to T7wt ( $1.04 \pm 0.08 \text{ g}\cdot\text{L}^{-1}$ ). T7DI\_2 and T7DI\_5 followed with dsRNA concentrations of  $0.76 \pm 0.06 \text{ g}\cdot\text{L}^{-1}$  and  $0.9 \pm 0.07 \text{ g}\cdot\text{L}^{-1}$ , marking a reduction of 27.1% and 13.2% respectively, relative to T7wt. T7DI\_2, T7DI\_5, and T7DI\_7 also exhibited lower ratios of dsRNA byproduct per gram of total mRNA produced compared to T7wt. T7wt produced dsRNA at a ratio of  $181.34 \pm 1.32 \text{ mg}\cdot\text{g}^{-1}$ , corresponding to approximately 18% of the total mRNA produced (Figure 3b). T7DI\_5 showed a decrease of 23.6% in the dsRNA/mRNA<sub>total</sub> ratio compared to T7wt, with a ratio of  $138.47 \pm 1.08 \text{ mg}\cdot\text{g}^{-1}$  (13.85% of the total mRNA produced). In addition, both T7DI\_2 and T7DI\_7 achieved a 46–49% reduction in dsRNA concentration compared to T7wt and were approximately 30–33% lower than T7DI\_5 (Figure 3b). To analyse the impact of the promoter candidates to the protein production, the 3 candidates (T7DI\_2, T7DI\_5, T7DI\_7) and the control (T7wt) were produced, capped and evaluated using a commercial-available *in vitro* protein expression kit based on HeLa cells. No significant differences were observed in the fluorescence expressed between the three candidates and between the candidates and the promoter wildtype (Figure 3C). This means that the modifications at the downstream of the promoter have no effect on the initiation of the translation within the cells.

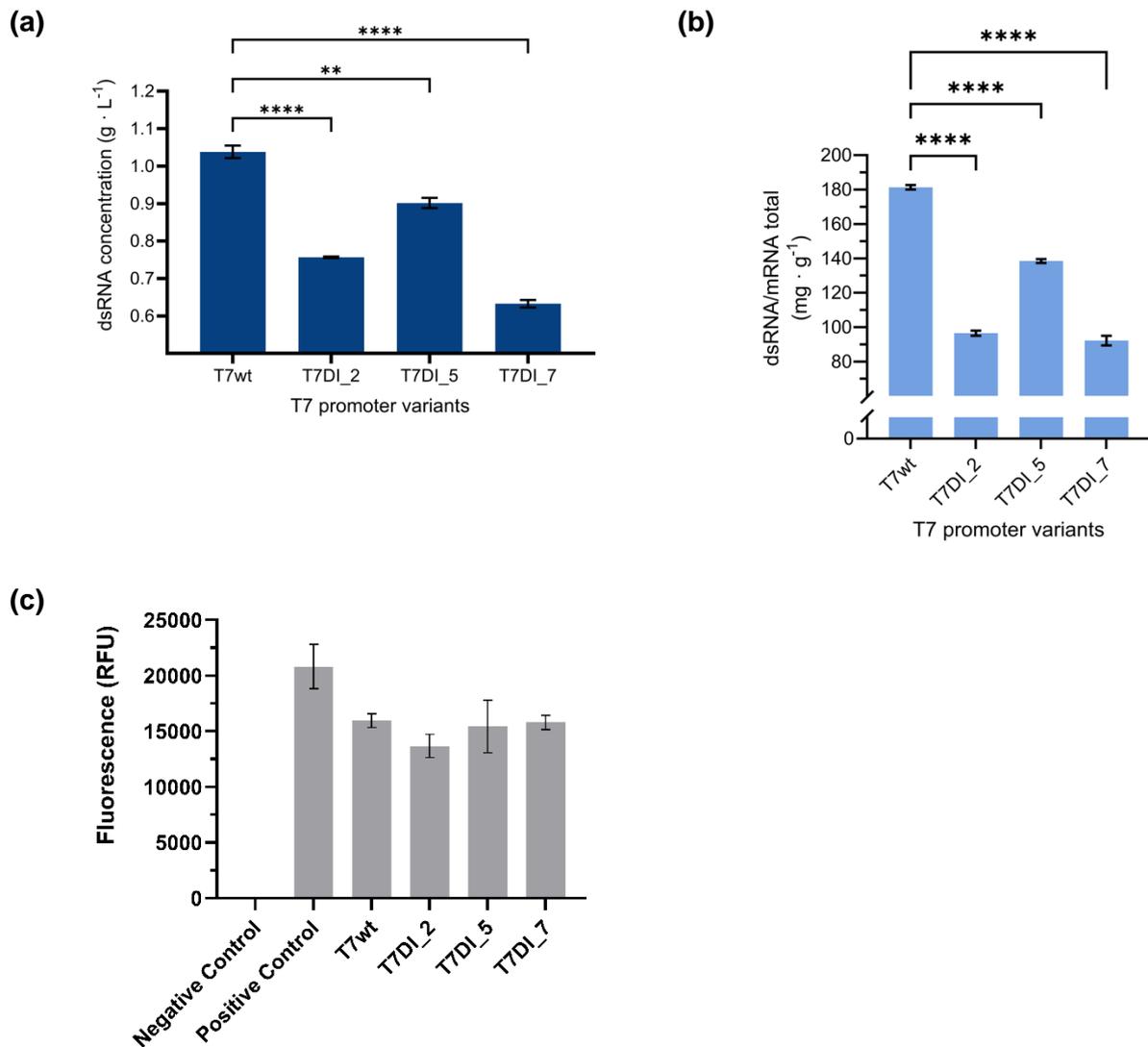


Figure 3. The dsRNA byproduct level of T7 promoter variants and the respective protein production analysis. The IVT reaction mix was prepared based on Rosa et al. (2022)[23] and the reactions were performed at 43°C for 2 h. Prior to quantification, RNase T1 was added to each IVT product to degrade the ssRNA. (a) dsRNA concentration ( $\text{g} \cdot \text{L}^{-1}$ ) quantified by RP-HPLC. (b) Concentration of dsRNA per gram mRNA ( $\text{mg} \cdot \text{g}^{-1}$ ) produced in IVT. Error bars represent the standard deviation ( $n = 2$ ). (c) EGFP production analysis (RFU) of the 3 promoter variants. Positive control corresponds to the use of the GFP control, and negative control WFI water. One-way ANOVA with Brown-Forsythe test followed by Tukey's multiple comparisons test, \*\* and \*\*\*\* denote p-values of 0.0012 and  $<0.0001$ , respectively. No significant differences ( $p > 0.05$ ) do not have p-value annotations.

## Evaluation of different sizes of templates and the impact on IVT

The mRNA production profiles of T7 promoter variants were evaluated using three different sizes of templates: 1195 pb (containing EGFP gene), 2483 bp (fused *Klebsiella pneumoniae* transaminase and EGFP genes) and 3851 bp (fused T7 RNA polymerase and EGFP genes). For this study, we selected the sequences that produced the least amount of dsRNA (T7DI\_2 and T7DI\_7) (Figure 4) to evaluate their consistency in maintaining low dsRNA levels when used to produce mRNA from different sizes of templates. Minimizing dsRNA is critical because it can influence the efficiency of mRNA translation[37] and compromise the safety of vaccines[38]. For the 1.2 kb EGFP template, the T7DI\_7 promoter produced  $14.05 \pm 0.5 \text{ g}\cdot\text{L}^{-1}$ , representing a 1.5-fold increase in comparison to T7wt, which produced  $9.18 \pm 0.29 \text{ g}\cdot\text{L}^{-1}$ . In contrast, the T7DI\_2 produced  $9.45 \pm 0.08 \text{ g}\cdot\text{L}^{-1}$ , showing no statistically significant difference from the T7wt ( $P > 0.05$ ). However, with the larger 2.5 kb TA\_EGFP template, T7DI\_2 produced  $7.50 \pm 0.10 \text{ g}\cdot\text{L}^{-1}$ , outperforming both T7DI\_7 and T7wt by 1.12-fold and 1.38-fold, respectively ( $P < 0.05$ ). For the 3.9 kb T7 RNAP\_EGFP template, no significant differences in mRNA production were observed between T7wt and T7DI\_2, or between T7DI\_2 and T7DI\_7 ( $P > 0.05$ ). Changing the template size from 1.2 kb to 2.5 kb influenced the mRNA yields across T7 promoter variants. For the 2.5 kb TA\_EGFP template, T7DI\_7 exhibited a 52.3% decrease, followed by 40.7% and 20.6% reductions in T7wt and T7DI\_2, respectively, when compared to the 1.2 kb EGFP template ( $P < 0.001$ ). A similar trend was also observed when comparing the 3.9 kb (T7 RNAP\_EGFP) template to the 1.2 kb (EGFP) template, marking decreases in mRNA production in T7 RNAP\_EGFP by 40.7%, 17.6%, and 15.8% for T7DI\_7, T7wt, and T7DI\_2, respectively ( $P < 0.001$ ). Rosa et al. (2022)[23] also reported that a larger 5.3 kb template, encoding fused Cas9 and EGFP genes, performed a 27% reduction in final mRNA concentration after a 2 h reaction compared to a smaller 1.2 kb template encoding EGFP gene<sup>23</sup>. However, changing the template size from 2.5 kb (TA\_EGFP) to 3.9 kb (T7 RNAP\_EGFP) did not significantly impact the mRNA production. T7DI\_2 consistently produced 7.5 to 7.9  $\text{g}\cdot\text{L}^{-1}$ , with modest increases observed in T7wt and T7DI\_7. Among the T7 promoter variants examined, T7DI\_2 exhibited minimal variation in mRNA yield across different templates.

The dsRNA profile was also evaluated for the three different sizes of mRNA. In the 1.2 kb EGFP template, T7wt exhibited the highest dsRNA byproduct level at a ratio of  $113.15 \pm 5.36 \text{ mg per gram of mRNA total (mg}\cdot\text{g}^{-1})$  (Figure 4b). These ratios were approximately 1.4-fold and 1.3-fold higher than those produced by T7DI\_2 and T7DI\_7, respectively. The same trend was observed in larger templates, although with lower degree. Interestingly, the effect decreases with the increase of template size. The dsRNA/mRNA<sub>total</sub> ratio was reduced by 18% in the 2.5 kb TA\_EGFP template compared to the 1.2 kb EGFP template (Figure 4b). Changing

the template size to 3.9 kb (T7 RNAP\_EGFP) in T7wt also reduced the dsRNA/mRNA<sub>total</sub> ratio by 12% relative to EGFP. T7DI\_2 also demonstrated to have a stronger impact on dsRNA byproduct levels compared with T7DI\_7.

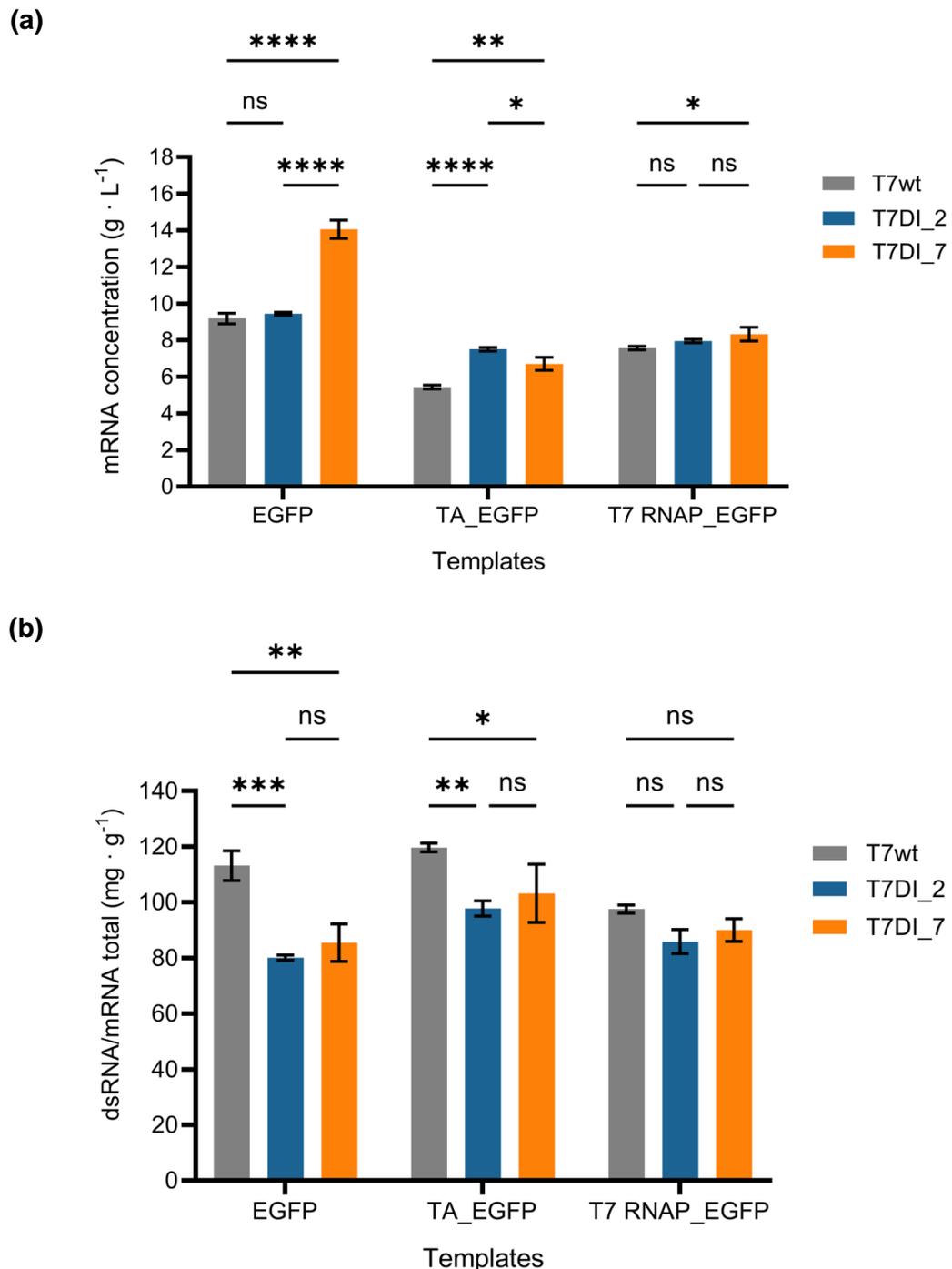


Figure 4. The profiles of mRNA production and dsRNA byproduct in T7 promoter variants across diverse template sizes. The IVT reaction mix was prepared based on Rosa et al. (2022)[23]. Two hours of IVT reactions were performed using T7 promoter variants (T7wt, T7DI\_2, and T7DI\_7) with three different sizes of templates (EGFP—1195 bp; TA\_EGFP—2483 bp; T7 RNAP\_EGFP—3851 bp). The mRNA and dsRNA concentrations were quantified

using RP-HPLC. (a) The mRNA production ( $\text{g}\cdot\text{L}^{-1}$ ) and (b) the ratio of dsRNA byproduct per gram mRNA total ( $\text{mg}\cdot\text{g}^{-1}$ ) across different sizes of templates. For a – b: error bars represent the standard deviation for  $n = 2$  replicates. Two-way ANOVA followed by Tukey's multiple comparisons, the significance levels are indicated as follows: 'ns' means no significant difference ( $p > 0.05$ ), \* for p-values between 0.01 and 0.045, \*\* for p-values between 0.001 and 0.005, \*\*\* for p-value of 0.0003, and \*\*\*\* for p-values of  $<0.0001$ . IVT, in vitro transcription; EGFP, enhanced green fluorescence protein; TA, Transaminase; T7 RNAP, T7 RNA polymerase; RP-HPLC, reverse phase-high performance liquid chromatography.

## Discussion

T7 RNA polymerase (RNAP) is the enzyme of choice to be used in the IVT system, owing to its simple structure, efficient production of long transcripts, and high specificity towards the T7 promoter[30,49]. Improving IVT system productivity may rely on improving transcriptional efficiency and mRNA quality. Different strategies can be used to lower the dsRNA presence, for example, engineer the T7 RNAP. A newly engineered T7 RNAP has demonstrated a significant reduction in dsRNA production during IVT compared to the wild type (WT)[40]. Despite this advance in reducing dsRNA impurities, the mutant variant produced slightly less total RNA compared to the WT[40], and it has not yet been available commercially. Through promoter optimisation, a 46–49% reduction in dsRNA concentration compared to the wild-type (WT) can be achieved, as demonstrated by the results with T7DI\_2 and T7DI\_7 (Figure 3). This optimization not only reduces dsRNA levels but also outperforms the WT in production rate (Figure 2) and total RNA yield during IVT (Figure 1a). Therefore, optimising T7 promoter sequences emerges as a viable and efficient strategy for enhancing transcription performance and minimizing dsRNA impurities in IVT.

In this work, we assessed the effect of modification in the promoter regions in the overall yield of the IVT reaction. We compared 11 variants that contained a AT-rich region at the promoter downstream, and modifications previously reported to enhance T7 RNAP transcriptional performance were used as controls. It is posited that the AT-rich element in the downstream region (+4 to +8) may facilitate DNA template unwinding and initiate the transcription bubble[42], thus enhancing the transcription performance and total mRNA produced. Extended AT-rich sequences in the upstream region have been shown to enhance the stability of the polymerase-promoter complex by inhibiting dissociation[45] and improve the *in vitro* protein synthesis[43], whereas downstream AT-rich motifs facilitate the unwinding of the DNA double helix during transcription initiation[42].

Several promoter variants outperformed the wild type T7 promoter, namely T7DI\_7 (1.5-fold increment), T7DI\_5 (1.4-fold) and T7DI\_10 (1.2-fold). Nonetheless, the downstream AT-rich sequences produced varied mRNA concentrations. The higher production rates observed underline the positive effect of the AT-rich downstream element at positions +4 to +8, which

may facilitate the unwinding of the double-stranded DNA template and the initiation of a transcription bubble which commences at the -4 position and extends downstream[42,50]. Lower production yields are also observed. The reduced mRNA yield observed in T7Max\_T7#4, underscores the importance of further investigation into the cumulative effects of these multiple promoter modifications. The T7c62 promoter contains mutations at positions -4 (A for T), -1 (C for A), and +2 (A for G)[44]. These positions encompass the TATA sequence from -4 to -1, which serves as the unwinding region and plays a pivotal role in the formation of transcription bubble[50]. Nucleotides at positions -1 (A) and -4 (T) are highly conserved across bacteriophage promoters and are notably AT-rich[51]. Furthermore, templates with a guanine (G) triplet at positions +1 to +3 of the T7 promoter were transcribed more robustly and may prevent premature dissociation of abortive transcripts[42]. These observations suggest that mutations at positions -4, -1, and +2 within T7c62 might have a profound effect on transcriptional activity, potentially explaining the lack of mRNA produced.

While the varied promoter modifications might produce similar mRNA concentrations, their underlying mechanism by which transcription performance is modulated could be different. The different transcriptional rates in T7DI\_2, T7DI\_5, and T7DI\_7 might arise from distinct AT combinations which highlights the sequence-specific manner of the AT-rich downstream element. Nonetheless, the mechanism underlying the distinct effects of AT combinations on transcription rates remains unknown and requires further investigation.

Considering the reduction in the dsRNA levels, it suggests that the AT-rich downstream elements may influence the stability of the transcription initiation complex and minimize the generation of abortive transcripts. During the initiation of transcription, T7 RNAP binds to the T7 promoter and synthesizes short RNAs or abortive transcripts via a mechanism known as abortive cycling[52,53,54]. These abortive transcripts can either anneal to each other or interact with T7 RNAP through RNA-templated transcriptional capabilities, giving rise to short dsRNA molecules[34,35,53]. Therefore, minimizing the formation of abortive transcripts during transcription initiation potentially influences the concentration of dsRNA byproduct produced in the IVT reaction. AT-rich sequences can have an impact on the stabilisation of the T7 RNAP-DNA complex[45]. AT-rich sequences in the upstream region of the T7 promoter have been reported to increase the stability of the polymerase-promoter complex by reducing the dissociation rate constant[45]. The AT-rich sequences in the downstream region might exert similar stabilizing effects on the initiation complex, which may influence the levels of dsRNA by-product. Additionally, varying dsRNA concentrations among the promoter variants with AT-rich downstream elements also highlight the sequence-specific characteristic of this element.

The variability in final mRNA concentrations across various template sizes suggests that the mRNA production is potentially influenced by the specific gene sequences encoded within the template rather than by the size, highlighting the sequence-dependent factors in transcriptional efficiency. Specific sequence characteristics have been implicated in the formation of abortive transcripts or truncated mRNA species, which can affect the efficiency of full-length mRNA synthesis[52,53]. In addition, the presence of specific sequences that are energetically favourable to RNA dimerization can trigger the formation of self-complementary mRNA[35,53]. In this condition, the mRNA duplex can interact with RNA polymerase through RNA-templated transcription capabilities, thereby influencing the production of canonical mRNA molecules[34,35,55].

The dsRNA level varies with different lengths of mRNA[40]. The percentage of dsRNA produced per total mRNA (% w/w) is reported to decrease from approximately 7.5% in ~850 nt mRNA to 2% in ~1500 nt mRNA[40]. However, a modest increase to roughly 4% for ~2900 nt mRNA was observed compared to ~1500 nt mRNA[40]. The variations in dsRNA levels across different templates suggest that formation of extended loopback dsRNA species[35,56]. Nevertheless, longer templates will produce dsRNA formation is a sequence-dependent mechanism[52,53]. mRNA can also act as templates for T7 RNA polymerase, leading to the lower amounts of mRNA strands. It is also noteworthy that the formation of abortive transcripts is sequence-specific[52,53]. The dsRNA formation is potentially more influenced by particular sequences encoded in the template that encourage the formation of RNA dimers[35,53] than by the size of the gene itself.

The relatively lower levels of dsRNA exhibited by T7DI\_2, along with its consistent performance across templates, suggest that a specific AT-rich downstream element may influence the dsRNA formation through mechanisms related to abortive cycling. The synthesis of short dsRNA species and abortive transcripts is associated with the initiation phase of the transcription[36,52]. Considering that the position of this element is at +4 to +8 downstream of the T7 promoter, where the initiation of transcription bubble occurs[42,50], it might stabilize the initiation-to-elongation transition during transcription, thereby reducing the number of abortive cycles. Nevertheless, this hypothesis needs further investigation.

## Conclusions

We demonstrated that modifications to the DNA template sequence have improved mRNA yields and quality. An AT-rich region at the downstream of the promoter resulted in an increase in the mRNA production of  $14 \text{ g}\cdot\text{L}^{-1}$  in approximately 2 hours. Analysing the mRNA production profile, it is observed that the promoter variants peak the mRNA production at 45 min, impacting overall production times. The increase in mRNA production is accompanied by a

reduction in dsRNA formation of at least 18% mostly due to a reduction in abortive cycling. Similar observations were made when with different pDNA templates (~1200bp to ~3900 bp). The optimisation of the non-coding regions sequences can lead to a positive impact on vaccine effectivity and stability but also to increase production yields and product quality during vaccine manufacturing. This is of paramount importance if a rapid response is required in events of future epidemics with quality on-demand productions. An increase in mRNA quality with a reduction in intensive purification operations will undoubtedly influence the manufacturing process's cost-effectiveness, ultimately making these vaccines affordable to all.

## References

- [1] Bloom, D. E., Fan, V. Y. & Sevilla, J. P. The broad socioeconomic benefits of vaccination. *Sci. Transl. Med.* **10**, eaaj2345 (2018).
- [2] Kayser, V. & Ramzan, I. Vaccines and vaccination: history and emerging issues. *Hum. Vaccines Immunother.* **17**, 5255–5268 (2021).
- [3] Andre, F. *et al.* Vaccination greatly reduces disease, disability, death and inequity worldwide. *Bull. World Health Organ.* **86**, 140–146 (2008).
- [4] LARGERON, N., Lévy, P., Wasem, J. & Bresse, X. Role of vaccination in the sustainability of healthcare systems. *J. Mark. Access Health Policy* **3**, 27043 (2015).
- [5] Lamb, Y. N. BNT162b2 mRNA COVID-19 Vaccine: First Approval. *Drugs* **81**, 495–501 (2021).
- [6] Rosa, S. S., Prazeres, D. M. F., Azevedo, A. M. & Marques, M. P. C. mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine* **39**, 2190–2200 (2021).
- [7] Li, M. *et al.* COVID-19 vaccine development: milestones, lessons and prospects. *Signal Transduct. Target. Ther.* **7**, 146 (2022).
- [8] Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020).
- [9] World Health Organization (WHO). WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020> (2020).
- [10] Mathieu, E. *et al.* Coronavirus (COVID-19) Vaccinations. *Coronavirus Pandemic (COVID-19)* <https://ourworldindata.org/covid-vaccinations>.
- [11] Pardi, N. *et al.* Administration of nucleoside-modified mRNA encoding broadly neutralizing antibody protects humanized mice from HIV-1 challenge. *Nat. Commun.* **8**, 14630 (2017).
- [12] Richner, J. M. *et al.* Modified mRNA Vaccines Protect against Zika Virus Infection. *Cell* **168**, 1114–1125.e10 (2017).
- [13] Lorentzen, C. L., Haanen, J. B., Met, Ö. & Svane, I. M. Clinical advances and ongoing trials of mRNA vaccines for cancer treatment. *Lancet Oncol.* **23**, e450–e458 (2022).
- [14] Kis, Z., Kontoravdi, C., Shattock, R. & Shah, N. Resources, Production Scales and Time Required for Producing RNA Vaccines for the Global Pandemic Demand. *Vaccines* **9**, 3 (2020).
- [15] Pardi, N., Hogan, M. J., Porter, F. W. & Weissman, D. mRNA vaccines — a new era in vaccinology. *Nat. Rev. Drug Discov.* **17**, 261–279 (2018).
- [16] Tusup, M. *et al.* Design of in vitro Transcribed mRNA Vectors for Research and Therapy. *CHIMIA* **73**, 391 (2019).
- [17] Kwon, S., Kwon, M., Im, S., Lee, K. & Lee, H. mRNA vaccines: the most recent clinical applications of synthetic mRNA. *Arch. Pharm. Res.* **45**, 245–262 (2022).

- [18] Qin, F. *et al.* A Guide to Nucleic Acid Vaccines in the Prevention and Treatment of Infectious Diseases and Cancers: From Basic Principles to Current Applications. *Front. Cell Dev. Biol.* **9**, 633776 (2021).
- [19] Fang, E. *et al.* Advances in COVID-19 mRNA vaccine development. *Signal Transduct. Target. Ther.* **7**, 94 (2022).
- [20] Schlake, T., Thess, A., Fotin-Mleczek, M. & Kallen, K.-J. Developing mRNA-vaccine technologies. *RNA Biol.* **9**, 1319–1330 (2012).
- [21] Karikó, K. *et al.* Incorporation of Pseudouridine Into mRNA Yields Superior Nonimmunogenic Vector With Increased Translational Capacity and Biological Stability. *Mol. Ther.* **16**, 1833–1840 (2008).
- [22] Geall, A. J., Mandl, C. W. & Ulmer, J. B. RNA: The new revolution in nucleic acid vaccines. *Semin. Immunol.* **25**, 152–159 (2013).
- [23] Rosa, S. S. *et al.* Maximizing mRNA vaccine production with Bayesian optimization. *Biotechnol. Bioeng.* **119**, 3127–3139 (2022).
- [24] Brandys, P. *et al.* A mRNA Vaccine Encoding for a RBD 60-mer Nanoparticle Elicits Neutralizing Antibodies and Protective Immunity Against the SARS-CoV-2 Delta Variant in Transgenic K18-hACE2 Mice. *Front. Immunol.* **13**, 912898 (2022).
- [25] Henderson, J. M. *et al.* Cap 1 Messenger RNA Synthesis with Co-transcriptional CleanCap<sup>®</sup> Analog by In Vitro Transcription. *Curr. Protoc.* **1**, (2021).
- [26] Whitley, J. *et al.* Development of mRNA manufacturing for vaccines and therapeutics: mRNA platform requirements and development of a scalable production process to support early phase clinical trials. *Transl. Res.* **242**, 38–55 (2022).
- [27] Wochner, A., Roos, T. & Ketterer, T. Methods and Means for Enhancing RNA Production. (2017).
- [28] Pregeljic, D. *et al.* Increasing yield of IVT reaction with at-line HPLC monitoring. <https://www.authorea.com/users/502841/articles/582756-increasing-yield-of-ivt-reaction-with-at-line-hplc-monitoring?commit=7b142ef22783da58ef53b64c64bed530240a1716> (2022) doi:10.22541/au.166119012.22572977/v2.
- [29] Skok, J. *et al.* Gram-Scale mRNA Production Using a 250-mL Single-Use Bioreactor. *Chem. Ing. Tech.* **94**, 1928–1935 (2022).
- [30] Borkotoky, S. & Murali, A. The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *Int. J. Biol. Macromol.* **118**, 49–56 (2018).
- [31] Duan, B., Wu, S., Da, L.-T. & Yu, J. A Critical Residue Selectively Recruits Nucleotides for T7 RNA Polymerase Transcription Fidelity Control. *Biophys. J.* **107**, 2130–2140 (2014).
- [32] Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798 (1987).
- [33] Mu, X., Greenwald, E., Ahmad, S. & Hur, S. An origin of the immunogenicity of in vitro transcribed RNA. *Nucleic Acids Res.* **46**, 5239–5249 (2018).
- [34] Cazenave, C. & Uhlenbeck, O. C. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc. Natl. Acad. Sci.* **91**, 6972–6976 (1994).
- [35] Gholamalipour, Y., Karunanayake Mudiyansele, A. & Martin, C. T. 3' end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character—RNA-Seq analyses. *Nucleic Acids Res.* **46**, 9253–9263 (2018).
- [36] Triana-Alonso, F. J., Dabrowski, M., Wadzack, J. & Nierhaus, K. H. Self-coded 3'-Extension of Run-off Transcripts Produces Aberrant Products during in Vitro Transcription with T7 RNA Polymerase. *J. Biol. Chem.* **270**, 6298–6307 (1995).
- [37] Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P. & Sharp, P. A. Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev.* **13**, 3191–3197 (1999).
- [38] Jin, B. *et al.* Immunomodulatory Effects of dsRNA and Its Potential as Vaccine Adjuvant. *J. Biomed. Biotechnol.* **2010**, 1–17 (2010).
- [39] Milano, G., Gal, J., Creisson, A. & Chamorey, E. Myocarditis and COVID-19 mRNA vaccines: a mechanistic hypothesis involving dsRNA. *Future Virol.* **17**, 191–196 (2022).

- [40] Dousis, A., Ravichandran, K., Hobert, E. M., Moore, M. J. & Rabideau, A. E. An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts. *Nat. Biotechnol.* (2022) doi:10.1038/s41587-022-01525-6.
- [41] Tang, G.-Q. *et al.* Relaxed Rotational and Scrunching Changes in P266L Mutant of T7 RNA Polymerase Reduce Short Abortive RNAs while Delaying Transition into Elongation. *PLoS ONE* **9**, e91859 (2014).
- [42] Conrad, T., Plumbom, I., Alcobendas, M., Vidal, R. & Sauer, S. Maximizing transcription of nucleic acids with efficient T7 promoters. *Commun. Biol.* **3**, 439 (2020).
- [43] Deich, C. *et al.* T7Max transcription system. *J. Biol. Eng.* **17**, 4 (2023).
- [44] Paul, S., Stang, A., Lennartz, K., Tenbusch, M. & Überla, K. Selection of a T7 promoter mutant with enhanced in vitro activity by a novel multi-copy bead display approach for in vitro evolution. *Nucleic Acids Res.* **41**, e29–e29 (2013).
- [45] Tang, G.-Q., Bandwar, R. P. & Patel, S. S. Extended Upstream A-T Sequence Increases T7 Promoter Strength. *J. Biol. Chem.* **280**, 40707–40713 (2005).
- [46] Trepotec, Z., Geiger, J., Plank, C., Aneja, M. K. & Rudolph, C. Segmented poly(A) tails significantly reduce recombination of plasmid DNA without affecting mRNA translation efficiency or half-life. *RNA* **25**, 507–518 (2019).
- [47] Issa, William & Packer, Meredith. Methods for HPLC analysis. 52 (2021).
- [48] Samnuan, K., Blakney, A. K., McKay, P. F. & Shattock, R. J. Design-of-experiments in vitro transcription yield optimization of self- amplifying RNA. (2022).
- [49] Borkotoky, S., Kumar Meena, C., Bhalerao, G. M. & Murali, A. An in-silico glimpse into the pH dependent structural changes of T7 RNA polymerase: a protein with simplicity. *Sci. Rep.* **7**, 6290 (2017).
- [50] Sousa, R. T7 RNA Polymerase. in *Encyclopedia of Biological Chemistry* 355–359 (Elsevier, 2013). doi:10.1016/B978-0-12-378630-2.00267-X.
- [51] Padmanabhan, R., Sarcar, S. N. & Miller, D. L. Promoter Length Affects the Initiation of T7 RNA Polymerase In Vitro: New Insights into Promoter/Polymerase Co-evolution. *J. Mol. Evol.* **88**, 179–193 (2020).
- [52] Martin, C. T., Muller, D. K. & Coleman, J. E. Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* **27**, 3966–3974 (1988).
- [53] Nacheva, G. A. & Berzal-Herranz, A. Preventing undesired RNA-primed RNA extension catalyzed by T7 RNA polymerase. *Eur. J. Biochem.* **270**, 1458–1465 (2003).
- [54] Gong, P. & Martin, C. T. Mechanism of Instability in Abortive Cycling by T7 RNA Polymerase. *J. Biol. Chem.* **281**, 23533–23544 (2006).
- [55] Arnaud-Barbe, N. Transcription of RNA templates by T7 RNA polymerase. *Nucleic Acids Res.* **26**, 3550–3554 (1998).
- [56] Cazenave, C. & Uhlenbeck, O. C. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc. Natl. Acad. Sci.* **91**, 6972–6976 (1994).

# Supporting Information

Table S1. Sequences of the the 5'-UTR, 3'-UTR, poly-A, and the gene of interest for DNA templates EGFP, TA\_EGFP, and T7 RNAP\_EGFP, used in this study.

Name	Size (bp)	Sequence
5'-UTR	45	ACTCACTATTTGTTTTTCGCGCCCAGTTGCAAAAAGTGTGCGCCACC
3'-UTR	284	GAGAGCTCGCTTTCCTTGCTGTCCAATTTCTATTAAAGGTTCCCTTTGTTCCCTAA GTCCAACTACTAAACTGGGGGATATTATGAAGGGCCTTGAGCATCTGGATTCTG CCTAATAAAAAACATTTATTTTCATTGCTGCGTCGAGAGCTCGCTTTCCTTGCTG TCCAATTTCTATTAAAGGTTCCCTTTGTTCCCTAAGTCCAACTACTAAACTGGGG GATATTATGAAGGGCCTTGAGCATCTGGATTCTGCCTAATAAAAAACATTTATT TTCATTGCTGCGTC
Poly-A	126	AA AAAAAAAAATGCATAA AAAAAAAAAAAAAAAAAAAAA
<b>Genes</b>		
EGFP	720	ATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAG CTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGC GATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACGGCAAGCTG CCCGTGCCCTGGCCCACCCTCGTGACCACCTGACCTACGGCGTGCAGTGCTTC AGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCC GAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAG ACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTG AAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTAC AACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATC AAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCC GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGAC AACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGC GATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATG GACGAGCTGTACAAGTAA
TA_EGFP	2002	ATGAACAGCAACAAAGCGATGATGGCGCGCCGAGCGATGCGGTGCCGCGCGGC GTGGGCCAGATTCATCCGATTTTCGCGGAACGCGCGGAAAACCTGCCGCTGTGG GATGTGGAAGGCCGGAATATCTGGATTTTGCGGGCGGCATTGCGGTGCTGAAC ACCGGCCATCTGCATCCGCAGGTGGTGGCGGCGGTGGAAGATCAGCTGAAGAAA CTGAGCCATACCTGCTTTCAGGTGCTGGCGTATGAACCGTATCTGGCGCTGTGC GAGAAAATGAACCAGAAAGTGCCGGGCGATTTTGCGAAGAAAACCTGCTGGTG ACCACCGGCAGCGAAGCGGTGAAAACGCGGTGAAAATTGCGCGCGCGGCGACC GGCCGAGCGGCGGATTCGTTTTACCGGCGCGGCGCATGGCCGCACCCATTAT ACCCTGAGCCTGACCGGCAAAGTGAACCCGTATAGCGCGGGCATGGCCCTGATG CCGGGCCATGTGTATCGCGCGCTGTATCCGTGCGCGCTGCATGGCGTGAGCGAT GATGAAGCGATTGCGAGCATTCATCGCATTTTCAAGAACGATGCGGCGCCGGAA GATATTGCGGCGATTATTATTGAACCGGTGCAGGGCGAAGGCGGCTTTTATGCG GCGAGCCCCGGGCTTTATGCAGCGCCTGCGCGCGCTGTGCGATGAACATGGCATT ATGCTGATTGCGGATGAAGTGCAGAGCGGCGCGGGCCGCACCGGCACCCCTGTTT

GCGATGGAACAGATGGGCGTGGCGGGCGGATATTACCACCTTTGCGAAAAGCATT  
GCGGGCGGCTTTCCGCTGGCGGGCGTGACCGGCCGCGCGGAAGTGATGGATGCG  
ATTGCGCCGGGCGGCCTGGGCGGCACCTATGCGGGCAACCCGATTGCGTGCGCG  
GCGGGCGCTGGCGGTGCTGCAGATTTTCGAACAGGAAAACCTGCTGGAGAAAGCG  
AACCAGCTGGGCGATACCCTGCGCCAGGGCCTGCTGGCGATTGCGGAAGATCAT  
CCGGAAATTGGCGATGTGCGCGGCCTGGGCGCGATGATTGCGATTGAAGTGT  
GAAGAAGGCGATCATAGCCGCCCCGAACGCGCGCCTGACCGCGGATATTGTGGCG  
CGCGCGCGGATAAAGGCCTGATTCTGCTGAGCTGCGGGCCGATTTATAACGTG  
CTGCGCATTCTGGTGCCGCTGACCATTGAAGAAGCGCAGATTGAACAGGGCCTG  
AAAATTATTGCGGATTGCTTTAGCGAAGCGAAACAGGCGCATGTTGAGCAAGGG  
CGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGT  
AAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG  
CAAGCTGACCCTGAAGTTCATCTGCACCACCGCAAGCTGCCCGTGCCCTGGCC  
CACCTCGTGACCACCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGA  
CCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCA  
GGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGT  
GAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTT  
CAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACACTACAACAGCCA  
CAACGTCTATATCATGGCCGACAAGCAGAAGACGGCATCAAGGTGAAGTTC  
GATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCA  
GAACACCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAG  
CACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC  
GCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGAGCTGTACAA  
GTAA

T7 RNAP\_ 3370  
EGFP

ATGAACACGATTAACATCGCTAAGAACGACTTCTCTGACATCGAACTGGCTGCT  
ATCCCGTTCAACACTCTGGCTGACCATTACGGTGAGCGTTTAGCTCGCGAACAG  
TTGGCCCTTGAGCATGAGTCTTACGAGATGGGTGAAGCACGCTTCCGCAAGATG  
TTTGAGCGTCAACTTAAAGCTGGTGAGGTTGCGGATAACGCTGCCGCCAAGCCT  
CTCATCACTACCCTACTCCCTAAGATGATTGCACGCATCAACGACTGGTTTGAG  
GAAGTGAAAGCTAAGCGCGGCAAGCGCCCCGACAGCCTTCCAGTTCCTGCAAGAA  
ATCAAGCCGGAAGCCGTAGCGTACATCACCATTAAGACCACTCTGGCTTGCCTA  
ACCAGTGCTGACAATACAACCGTTCAGGCTGTAGCAAGCGCAATCGGTGGGGCC  
ATTGAGGACGAGGCTCGCTTCGGTTCGTATCCGTGACCTTGAAGCTAAGCACTTC  
AAGAAAAACGTTGAGGAACAACCTCAACAAGCGCGTAGGGCACGTCTACAAGAAA  
GCATTTATGCAAGTGTGTCGAGGCTGACATGCTCTCTAAGGGTCTACTCGGTGGC  
GAGGCGTGGTCTTCGTGGCATAAGGAAGACTCTATTATGTAGGAGTACGCTGC  
ATCGAGATGCTCATTGAGTCAACCGGAATGGTTAGCTTACACCGCCAAAATGCT  
GGCGTAGTAGGTTCAAGACTCTGAGACTATCGAACTCGCACCTGAATACGCTGAG  
GCTATCGCAACCCGTGCAGGTGCGCTGGCTGGCATCTCTCCGATGTTCCAACCT  
TGCGTAGTTCCTCCTAAGCCGTGGACTGGCATTACTGGTGGTGGCTATTGGGCT  
AACGGTCGTCTCCTCTGGCGCTGGTGCCTACTCACAGTAAGAAAGCACTGATG  
CGCTACGAAGACGTTTACATGCCTGAGGTGTACAAAGCGATTAACATTTGCGCAA  
AACACCGCATGGAAAATCAACAAGAAAGTCCTAGCGGTGCGCAACGTAATCACC  
AAGTGGAAGCATTGTCCGGTCGAGGACATCCCTGCGATTGAGCGTGAAGAACTC  
CCGATGAAACCGGAAGACATCGACATGAATCCTGAGGCTCTCACCGGTGGAAA  
CGTGCTGCCGCTGCTGTGTACCGCAAGGACAGGGCTCGCAAGTCTCGCCGTATC  
AGCCTTGAGTTCATGCTTGAGCAAGCCAATAAGTTTGCTAACCATAAGGCCATC  
TGTTCCCTTACAACATGGACTGGCGCGGTGCTGTTTACGCCGTGCAATGTTT  
AACCCGCAAGGTAACGATATGACCAAAGGACTGCTTACGCTGGCGAAAGGTA  
CCAATCGGTAAGGAAGGTTACTACTGGCTGAAAATCCACGGTGCAAACCTGTGCG  
GGTGTGATAAGGTTCCGTTCCCTGAGCGCATCAAGTTCATTGAGGAAAACCC  
GAGAACATCATGGCTTGGCTAAGTCTCCACTGGGAAACACTGGTGGGCTGAG  
CAAGATTCTCCGTTCTGCTTCCCTGCGTTCGCTTTGAGTACGCTGGGGTACAG  
CACACGGCCTGAGCTATAACTGCTCCCTTCCGCTGGCGTTGACGGGTCTTGC  
TCTGGCATCCAGCACTTCTCCGCGATGCTCCGAGATGAGGTGAGGTGGTTCGCGCG  
GTTAACTTGCTTCCCTAGTGAGACCGTTCAGGACATCTACGGGATGTTGCTAAG  
AAAGTCAACGAGATTCTACAAGCAGACGCAATCAATGGGACCGATAACGAAGTA

GTTACCGTGACCGATGAGAACACTGGTCAAATCTCTGAGAAAGTCAAGCTGGGC  
 ACTAAGGCACTGGCTGGTCAATGGCTGGCTCACGGTGTACTCGCAGTGTGACT  
 AAGCGTTCAGTCATGACGCTGGCTTACGGGTCCAAAGAGTTCCGGCTTCCGTCAA  
 CAAGTGCTGGAAGATAACCATTACAGCCAGCTATTGATTCCGGCAAGGGTCCGATG  
 TTCACTCAGCCGAATCAGGCTGCTGGATACATGGCTAAGCTGATTTGGGAATCT  
 GTGAGCGTGACGGTGGTAGCTGCGGTTGAAGCAATGAACTGGCTTAAGTCTGCT  
 GCTAAGCTGCTGGCTGCTGAGGTCAAAGATAAGAAGACTGGAGAGATTCTTCGC  
 AAGCGTTGCGCTGTGCATTGGGTAACCTCCTGATGGTTTTCCCTGTGTGGCAGGAA  
 TACAAGAAGCCTATTCAGACGCGCTTGAACCTGATGTTCCCTCGGTGAGTTCCGC  
 TTACAGCCTACCATTAACACCAACAAAGATAGCGAGATTGATGCACACAAACAG  
 GAGTCTGGTATCGCTCCTAACTTTGTACACAGCCAAGACGGTAGCCACCTTCGT  
 AAGACTGTAGTGTGGGCACACGAGAAGTACGGAATCGAATCTTTTGCAGTATT  
 CACGACTCCTTCGGTACCATTCCGGCTGACGCTGCGAACCTGTTCAAAGCAGTG  
 CGCGAAACTATGGTTGACACATATGAGTCTTGTGATGTACTGGCTGATTTCTAC  
 GACCAGTTCGCTGACCAGTTGCACGAGTCTCAATTGGACAAAATGCCAGCACTT  
 CCGGCTAAAGGTAACCTTGAACCTCCGTGACATCTTAGAGTCGGACTTCGCGTTC  
 GCGCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGGCCATCCTGGT  
 CGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGA  
 GGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCACCGGCAA  
 GCTGCCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTG  
 CTTTACAGCCGCTACCCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCAT  
 GCCCCAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA  
 CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGA  
 GCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGA  
 GTACAACACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGG  
 CATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCT  
 CGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCC  
 CGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAA  
 GCGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCGGGGATCACTCTCGG  
 CATGGACGAGCTGTACAAGTAA

**Table S2.** Plasmid used in this study

Plasmids	Relevant characteristics	Source or references
pT7wt_EGFP	pUC57-Kan vector, wildtype T7 promoter, EGFP gene	23
pT7#4_EGFP	pUC57-Kan vector, T7#4 promoter, EGFP gene	42
pT7Max_EGFP	pUC57-Kan vector, T7Max promoter, EGFP gene	43
pT7c62_EGFP	pUC57-Kan vector, T7c62 promoter, EGFP gene	44
pT7Max_T7#4_EGFP	pUC57-Kan vector, T7Max_T7#4 promoter, EGFP gene	42,43
pT7c62_T7#4_EGFP	pUC57-Kan vector, T7c62_T7#4 promoter, EGFP gene	42,44
pT7DI_1_EGFP	pUC57-Kan vector, T7DI_1 promoter, EGFP gene	In this study
pT7DI_2_EGFP	pUC57-Kan vector, T7DI_2 promoter, EGFP gene	In this study
pT7DI_3_EGFP	pUC57-Kan vector, T7DI_3 promoter, EGFP gene	In this study
pT7DI_4_EGFP	pUC57-Kan vector, T7DI_4 promoter, EGFP gene	In this study
pT7DI_5_EGFP	pUC57-Kan vector, T7DI_5 promoter, EGFP gene	In this study

pT7DI_6_EGFP	pUC57-Kan vector, T7DI_6 promoter, EGFP gene	In this study
pT7DI_7_EGFP	pUC57-Kan vector, T7DI_7 promoter, EGFP gene	In this study
pT7DI_8_EGFP	pUC57-Kan vector, T7DI_8 promoter, EGFP gene	In this study
pT7DI_9_EGFP	pUC57-Kan vector, T7DI_9 promoter, EGFP gene	In this study
pT7DI_10_EGFP	pUC57-Kan vector, T7DI_10 promoter, EGFP gene	In this study
pT7DI_11_EGFP	pUC57-Kan vector, T7DI_11 promoter, EGFP gene	In this study
pET29A_TA	pET29A vector with <i>K. pneumoniae</i> transaminase gene	Lab collection
pET29A_T7 RNAP	pET29A vector with T7 RNA polymerase gene	Lab collection
pT7wt_TA_EGFP	Wildtype T7 promoter, fused <i>K. pneumoniae</i> transaminase and EGFP genes	In this study
pT7wt_T7 RNAP_EGFP	Wildtype T7 promoter, fused T7 RNA polymerase and EGFP genes	In this study
pT7DI_2_TA_EGFP	T7DI_2 promoter variant, fused <i>K. pneumoniae</i> transaminase and EGFP genes	In this study
pT7DI_2_T7 RNAP_EGFP	T7DI_2 promoter variant, fused T7 RNA polymerase and EGFP genes	In this study
pT7DI_7_TA_EGFP	T7DI_7 promoter variant, fused <i>K. pneumoniae</i> transaminase and EGFP genes	In this study
pT7DI_7_T7 RNAP_EGFP	T7DI_7 promoter variant, fused T7 RNA polymerase and EGFP genes	In this study

**Table S3.** List of primers used in this study.

Primer name	Sequence (5' → 3')	Template
<b>A. Site-directed mutagenesis</b>		
T7#4_FWD / T7c62T7#4_FWD	ATAATACTCACTATTTGTTTTTCGC	pT7wt_EGFP
T7Max_FWD / T7c62_FWD	AATTCTAATACGACTCACTATAGGG CGGAGACTCACTATTTGTTTTTCGC	pT7wt_EGFP pT7wt_EGFP
T7MaxT7#4_FWD	AATTCTAATACGACTCACTATAGGG	pT7#4_EGFP
T7DI1_FWD / T7DI2_FWD	AATAAACTCACTATTTGTTTTTCGC TAAAAACTCACTATTTGTTTTTCGC	pT7wt_EGFP pT7wt_EGFP
T7DI3_FWD / T7DI4_FWD	AATTA ACTCACTATTTGTTTTTCGC ATTAAACTCACTATTTGTTTTTCGC	pT7wt_TA_EGFP pT7wt_T7 RNAP_EGFP pT7wt_EGFP pT7wt_EGFP
T7DI5_FWD / T7DI6_FWD	TTAAAACTCACTATTTGTTTTTCGC ATTTAACTCACTATTTGTTTTTCGC	pT7wt_EGFP pT7wt_EGFP



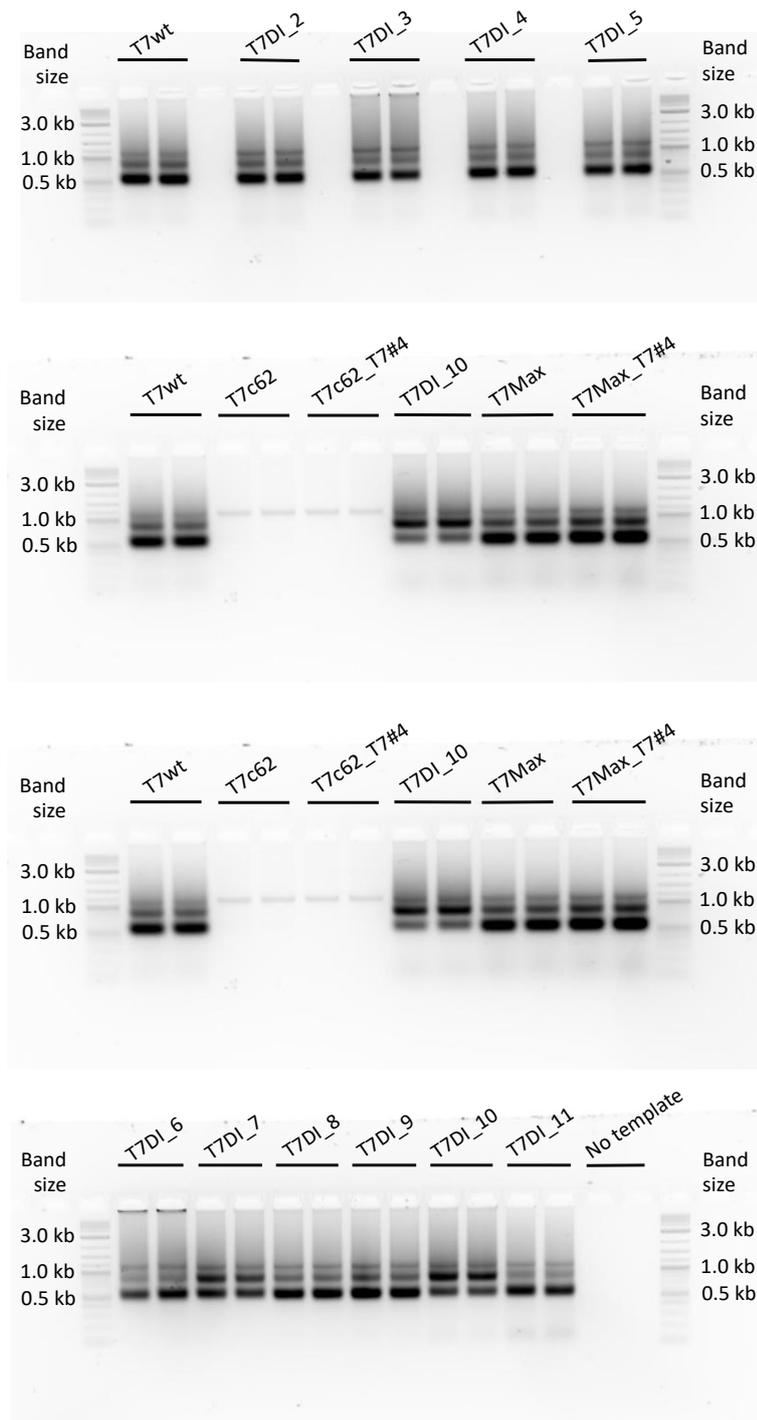


Figure S1. The agarose-gel electrophoresis analysis of IVT products (original gel images). IVT reactions were performed at 43°C for 2 h, based on the protocol in Rosa et al. (2022)[23]. The mRNA produced in the IVT is indicated by the 600 nt RNA band while the linear DNA template used in the reaction is represented by a 1.2 kb band. mRNA is produced in IVT reactions utilising T7 promoter variants, except for T7c62 and T7c62\_T7#4, where no mRNA band is observed. Gels do not provide a normalised quantity of mRNA and cannot be used to assess the mRNA yield.

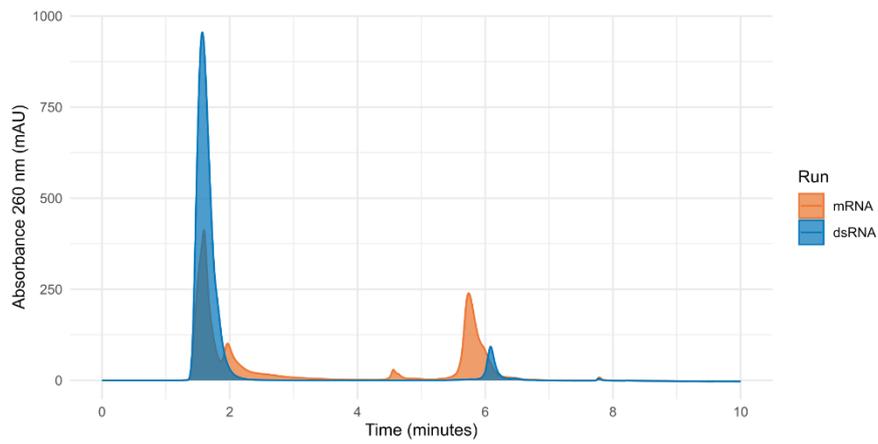


Figure S2. Chromatographic profiles of the samples containing total mRNA (blue) and dsRNA (red). The dsRNA chromatogram was obtained after digestion with Rnase T1 as described in section 2.3.2. Peaks at 5.5 and 6.2 min are total mRNA and dsRNA, respectively.

# Chapter VII – Simplifying the mRNA manufacturing process: Exploring oligo-dT ligand as a capture step after *in vitro* transcription reactions

---

**Keys words:** mRNA; Downstream processing; Affinity chromatography; purification; IVT

## Introduction

Messenger RNA (mRNA) is a single-stranded nucleic acid molecule that is responsible to carry the genetic information stored in the nucleus to the cytosol where it is translated into proteins. This forms the basis for mRNA vaccines, a synthetic transcript (DNA template containing the target gene) that use the cell machinery to express antigens *in situ* in order to induce an immune response [1]. The manufacturing of these vaccines is fairly straightforward, comprising the DNA template production based on the gene of interest chosen, synthesis of the drug substance (mRNA) in a cell free system, and the formulation into a drug product using carrier-bases systems[2,3]. This simplicity and the efforts of academia, industry and regulatory agencies allowed these vaccines (vaccines BNT162b2 and mRNA-1273) to be developed, manufactured and deployed in a record time to tackle the COVID-19 pandemic [4]. Since then, the application of these vaccines has broaden, spanning from vaccines to protein-encoding therapies[5], currently with over 60 clinical trials in the pipeline[6]. This interest in this vaccine technology is attributed also to the safety, non-integrative, and the transient expression nature of the mRNA. Furthermore, the induction of both humoral and cellular immune response the expression of the gene of interest is favoured. Nonetheless, challenges of these vaccines are the low thermal stability (requiring cold-chain), the effective duration of the immune response (requiring immunisation boosts) and the public acceptance.

From a molecular perspective, mRNA is composed of the gene of interest flanked by a backbone consisting of a 5' cap and a UTR in the 5' molecule end, and a second UTR and a poly-A tail in the 3' molecule end [7]. mRNA manufacturing process is relatively simple. The mRNA is usually produced enzymatically in an *in vitro* transcription reaction catalysed by RNA polymerase, using DNA template containing the targeted gene, and NTPs as a co-substrates [8–11]. Yields over  $10 \text{ g}_{\text{mRNA}} \text{ L}^{-1}_{\text{reaction}}$  are now obtained in batch[9] and fed-batch[12] mode in approximately 2 hours. The capping of the 5' end is necessary to provide protection against endonuclease and improving sensing and protein expression. This can be added during this step, using a cap-analog in a step entitled co-transcriptional capping[2], or by a second enzymatic reaction[13]. After IVT, it is common to have besides the mRNA, process- ( e.g. T7 RNA polymerase and other enzymes, unused NTPs and the DNA template) and product-related impurities (e.g. partial transcripts or double stranded mRNA, dsRNA[2,3]. The removal of product-related impurities represent a critical step on the mRNA performance as its presence in the final product is linked to a decrease in mRNA translation and to the induction of uncontrolled immune-inflammatory reactions[14]. Despite the fairly standardised manufacturing process[2]. It is important to operate the downstream processing under

optimise conditions, (to minimise product losses and simultaneously decrease manufacturing cost), operate within regulator guidelines[15], and allow the processing to be product agnostic.

The downstream processing is composed of several filtration and chromatographic steps, used in sequence, to achieve the separation of the mRNA from the different process and product-related impurities. In particular, the chromatography separation has received particular interest as it allows to explore physico-chemical properties to separate mRNA from its impurities. Different chromatography techniques have been explored such as, size exclusion[16,17], ion-pair reverse phase[18,19] and anion-exchange chromatography[20]. All the later have their associated challenges, such as inefficient separation of mRNA from the product-related impurities, or the use of organic solvent to achieve separation. New types of chromatography have been explored to overcome these challenges, such as anion exchange or multimodal chromatography. Core bead chromatography can separate mRNA from smaller and negatively charged impurities in a flowthrough mode[21]. However, it requires the use of enzymes to remove similar size impurities, such as DNase to remove the DNA template. Weak-anion exchange chromatography requires the use of denaturing conditions to achieve separation. Weak anion exchange multimodal monolith can achieve separation of mRNA from DNA template[22,23]. Hydroxyapatite[21] and hydrophobic interaction chromatography[24,25] can be combined with different purification steps to deliver a highly pure ssRNA.

One particular chromatography type is the use of oligo deoxythymidine (dT) affinity towards the poly-A tail [3,26]. Oligo deoxythymidine (oligo dT) consists of single stranded thymidine DNA nucleotide chains. This ligand is immobilised on solid chromatography matrices such as beaded resins, monoliths, or membranes. Linkers and polysaccharide extenders may also be introduced between the oligo dT and matrix to improve access of the feed material to the oligo dT[3,27]. This type of chromatography is widely popular due to its specificity and different supports are commercially available. The use of this ligand is extensively studied to purify RNA containing poly-adenylated molecules from different solutions[27–29]. In recent years, oligo-dT affinity has been explored for the purification of mRNA vaccines[30–32]. Nevertheless, these studies focus the interaction of mRNA with the solid-phase and knowledge on the behaviour of the separation of the mRNA with process- and product impurities is missing.

In this study, we explore oligo-dT as a solid-phase to purify mRNA directly after an IVT reaction. We hypothesise that this ligand could be use as primary capture step to remove all the process- and product-impurities, avoiding the use of additional chromatographic or even filtration steps. Machine Learning routines, previously used to optimise IVT conditions[9], were employed to optimise the dynamic binding capacity (DBC) by exploring the salt concentration in the mobile phase, the mRNA concentration in the feed, and the overall operation residence



with RNase free water at a ration of 1:2 and incubated at 80°C for 5 min. The temperature was decreased by 1°C every 90s min until it reached 25°C.

### **RNA purification**

dsRNA samples were incubated with RNase T1 and Turbo DNase (1 uL per 20 uL of IVT reaction) for 30 minutes at 37°C. mRNA samples are incubated with only Turbo DNase, using the same concentration. dsRNA and mRNA IVT samples were purified using a MEGAclean™ Transcription Clean-Up kit according to the manufacturer instructions. RNAs samples were further concentrated by ethanol precipitation by adding 500 mM pH 5 ammonium acetate and 2.5 volumes of ethanol. The samples were centrifuged at 15000 x g for 15 min and resuspended in WFI water to the desired final concentration.

### **Chromatography runs**

1mL POROS™ GoPure™ Oligo (dT)25 pre-packed column (ThermoFisher Scientific, US) was used in an AKTA Avant (Cytiva, Sweden) system equipped with a multiwavelength cell set at 260 nm. All chromatography parameters were controlled and monitored using the UNICORN 6.1 software (Cytiva, Sweden). 10 mM pH 7.4 Tris, 2mM EDTA, 2M NaCl (Buffer A), 10 mM pH 7.4 Tris, 2mM EDTA (Buffer B), and Mili-Q water were used as the mobile phase. The conductivity was adjusted by mixing Buffer A and Buffer B with equilibration step performed at the desired conductivity for 5 CVs at 300 cm.h<sup>-1</sup>. Samples were conditioned with a mixture of Buffer A and B to the desired conductivity. After injection, the column was washed with the equilibration buffer (A mixture of A and B at the desired conductivity) for 2 CVs. The injection flow was adjusted to the desired residence time. A wash with 95% of buffer B (5 mS.cm<sup>-1</sup>) was performed for 5 CVs at 300 cm.h<sup>-1</sup>. mRNA was eluted by a step elution of 5 CVs with Milli-Q water, or a gradient elution with water of 20 CV, at a flow rate of 300 cm.h<sup>-1</sup>. During the binding studies, 0.2 mL fractions were collected during injection with 2 mL collected for the remaining steps. Samples were evaluated and quantified by agarose gel electrophoresis and HPLC analysis (see section *Analytical methods*).

### **Bayesian optimisation**

The Bayesian optimisation methodology and configuration used was previously developed [9]. Briefly, DBC<sub>10</sub> was set as an estimator and mRNA feed concentration (g.L<sup>-1</sup>), resident time (min) and NaCl concentration (mM) were set as parameters. Latin hypercube sampling (LHS) was used to generate initial experiments before the optimisation cycle with Gaussian Process (GP) chosen to be the surrogate model for the Bayesian optimisation process. A Matérn 5/2 kernel was chosen to be the kernel function for the GP. Additionally, and Expected Improvement (EI) acquisition function was added to guide with exploration and/or exploitation

of the Bayesian optimisation cycle. The predictions of the GP estimators ( $DBC_{10}$ ) were evaluated by the use of Shapley values.

## Analytical methods

### HPLC analysis

mRNA was quantified by reverse-phase HPLC (RP-HPLC) as described previously [9,33]. Briefly, a DNAPac RP column (2.11x100 mm) and guard column guard column (3 × 10 nm) were used in a UHPLC equipped with a multiwavelength detector (ThermoFisher, USA). Samples of 2.5 uL were diluted 6 times before injection in TAE buffer (100 mM Tris-acetate, 2.5 mM EDTA, pH 7) in a pre-equilibrated column with the same buffer at a flow rate of 0.2 mL min<sup>-1</sup>. After a 1 min washing step, a gradient step with 1xTAE buffer, pH 7 and 25% (v/v) acetonitrile was performed at 0.35 mL min<sup>-1</sup> for 30s until 6% of 1xTAE buffer, pH 7, 25% (v/v) acetonitrile was reached. This was followed by a second gradient step to 76.5% of 1xTAE buffer, pH 7, 25% (v/v) acetonitrile, at 0.4 mL min<sup>-1</sup> for 4 min. A cleaning was performed at 100% of the same buffer for 3 min with column re-equilibration with a 1xTAE buffer for 4 mins at 0.4 mL min<sup>-1</sup>. The run was performed at 80°C, and the absorbance was monitored at 260 nm.

### Agarose gel electrophoresis

Samples were run on a 2% (w/v) agarose gel electrophoresis. Briefly, the gel was prepared with 0.5x TBE buffer (45 mM Tris–borate and 1 mM EDTA) and 5.5 mM magnesium chloride, and stained with Invitrogen SYBR® Safe DNA Gel Stain (1:10,000 dilution). The samples were prepared by adding 6× purple Loading Dye (New England Biolabs, UK) according to manufacturer instructions. The electrophoresis was run at 100 V for 75 min, and visualised using an Amersham™ Imager 600 (GE Healthcare, UK).

### Evaluation of binding capacity

The dynamic binding capacity (DBC) was defined as the amount of mRNA that binds to the solid-phase until a 10% of breakthrough was achieved ( $DBC_{10}$ ) and defined by

$$DBC_{10} = (V_{10} - V_0) \frac{mRNA_{feed}}{CV} \quad (1)$$

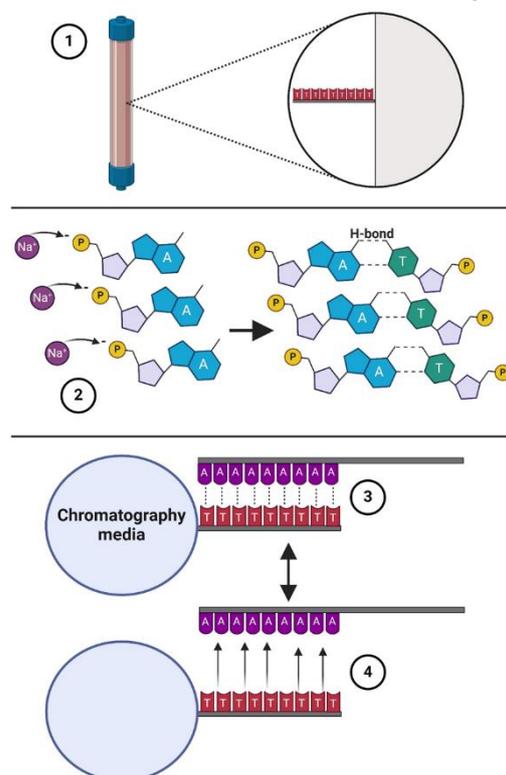
where  $V_{10}$ ,  $V_0$ ,  $CV$  are the volume where the concentration on the flow corresponds to 10% of the mRNA feed concentration, the delay volume, and the column volume, respectively, and

$mRNA_{\text{feed}}$  is the mRNA feed concentration. The column volume was determined under non-binding conditions by the injection of 1M of NaCl.

## Results and Discussion

### Physico-chemical properties of mRNA and interaction with oligo-dT ligands

mRNA is a single-stranded molecule composed of ribonucleotides linked by the 5' phosphate to a 3' hydroxyl group of the previous nucleotide, exhibiting an overall negative charge with an isoelectric point between 2 to 2.5 [34]. This results from the anionic phosphate groups that link the nucleotides together by phosphodiester bonds. Hydrogen bonds can be formed between the nitrogenous bases, impacting secondary structures and stabilisation of the RNA structure, and ultimately also the purification of mRNA during manufacturing. In particular, oligo-dT ligands used in affinity chromatography takes advantage of the hydrogen bond that can be formed between the poly-A tail structure present in the mRNA, and the single stranded thymidine DNA nucleotide (Figure 1). To achieve this, the dominant electrostatic interactions that usually causes repulsion between the RNA and the ligand, must be repressed in favour of the hydrogen bonding. Usually, sodium chloride is added to the mobile phase during binding to release free sodium cations that will associate with the anionic phosphate groups and shield the negative charges of each group from one another, and ultimately reducing the negative repulsion between the groups. Upon association of sodium cations with the anionic phosphates, the nucleotide chain of both RNA and ligand is stabilised and due to this, it



becomes favourable for the poly-A tail to undergo hydrogen bonding with the complementary oligo-dT chain. Chaotropic agents may be used to modulate poly-A tail and ligand interaction.

Figure 1 schematic representation of the mRNA binding to the Oligo-dT ligand. Step 1 - Oligo dT is an affinity ligand which can be used to capture mRNA by the poly A tail, when immobilised on a solid chromatography matrix. Step 2- Addition of sodium chloride stabilises mRNA and shields the electrostatic repulsion between phosphate groups on separate strands. This induces base pairing with oligo dT. Step 3 - Base pairing between poly adenine and thymidine immobilises the whole mRNA molecule on the chromatography matrix. This immobilisation can be maintained with a high salt concentration. Step 4 - Elution of mRNA is achieved by reducing salt concentration, causing poly A to unbind from oligo dT. Made with Biorender.

mRNA produced by *in vitro* transcription contains multiple process related impurities that can influence binding (e.g. reaction pH and salts) and therefore mRNA must be conditioned into the appropriate mobile phase before column loading. This can be achieved by a tangential flow filtration operating in diafiltration mode[24,35] or simply by dilution. Choosing the correct mobile phase in each step of the chromatographic process is critical to ensure the best yield and separation efficiency. Traditionally, Tris-EDTA buffer (TE) or phosphate-EDTA, pH 7.4 are usually used in mRNA purification processes [26,32,36]. EDTA minimising nonspecific interactions between mRNA and impurities [37]. NaCl concentrations typically vary between 0.5 to 2M in the binding mobile phase with other salts, such as lithium chloride (ranging between 0.5 to 2.5M), could also be used [2].

For the benchmark runs, pH 7.4 TE buffer containing 1M of NaCl was the binding mobile phase of choice and used to condition the IVT samples. Initial binding was performed at a flow rate of 300 cm.h<sup>-1</sup>, corresponding to 1 min residence time, with a wash step at 5 mS.cm<sup>-1</sup> conductivity. This would enable the removal of both process- and product-related impurities that could bind to the stationary phase with the mRNA being eluted solely with water (Figure 2). Under these conditions, the DNA template (process related impurity) is removed during the binding and wash step (Figure 2B). With the washing step. no DNases are required to be added to the process before the chromatography to remove the DNA. By analysing the HPLC chromatograms, a recovery yield of mRNA is 82% with an overall impurity removal of 89%. Regarding the DBC<sub>10</sub> obtained under the evaluated conditions, this showed low capacity (0.24 mg<sub>RNA</sub>.mg<sub>resin</sub><sup>-1</sup>) but are in line with previous studies for identical residence time and using pure mRNA samples[30]. It is therefore clear that DBC should be maximised by optimising the binding conditions.

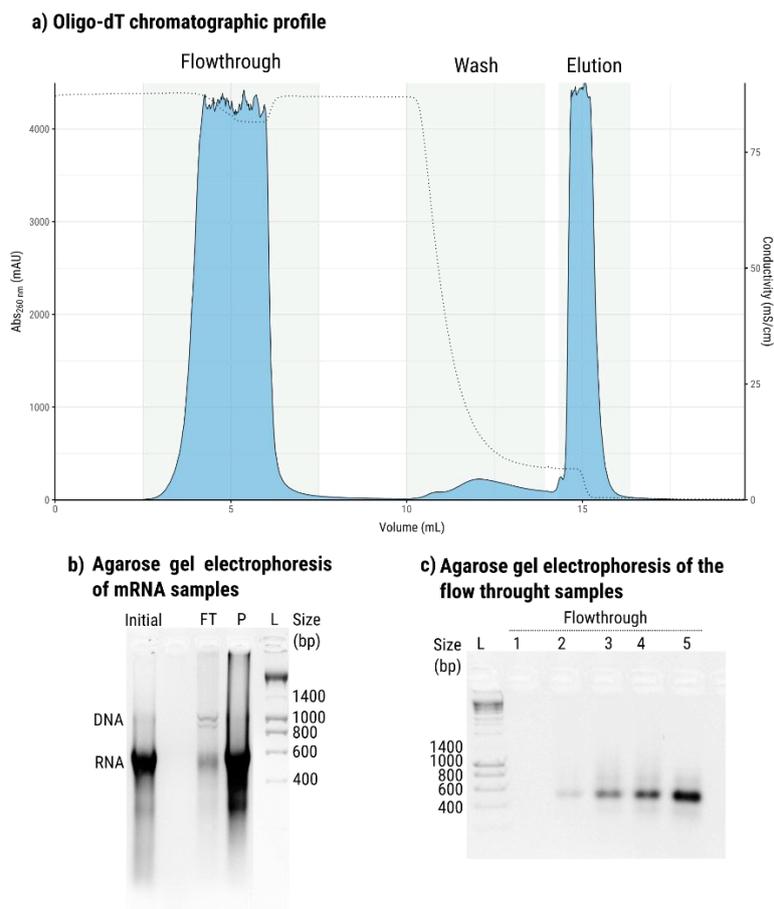


Figure 2. Chromatographic mRNA purification from IVT reactions using 1mL POROS™ GoPure™ Oligo (dT)25 pre-packed columns. a) Purification of mRNA from IVT sample conditioned with binding buffer (TE buffer, pH 7.4, 1M NaCl). Column is equilibrated with a binding buffer and washed after injection with 2CV to remove unbounded impurities. A second wash is performed with a mixture of binding buffer with 5% TE buffer, pH 7.4 at 5 mS.cm<sup>-1</sup> for 5 CVs. The mRNA is eluted with Mili-Q water for 5 CV at 300 cm.h<sup>-1</sup>. Dotted and solid lines correspond to the conductivity and absorbance at 260 nm, respectively. b) Gel agarose electrophoresis of the samples obtained from different steps of the chromatographic run. Lanes labelling: Initial correspond to IVT sample, FT to flowthrough during injection, P to the mRNA elution peak, and L corresponds to the ladder. c) Gel agarose electrophoresis of the samples obtained by the fractionation of the flowthrough during injection. Lanes 1 to 5 correspond to the volume between 2.5 to 3.5 mL (0.2 mL fractions) residence times.

### Increasing binding capacity by applying Bayesian optimisation

A Bayesian optimiser was used to optimise binding conditions. This methodology was previously used to maximise mRNA production, exploring the different IVT reaction parameters, resulting in a 2 fold improvement in yield and significant reduction in reaction time [9]. The benefits of this methodology are the reduced experiments necessary to achieve a optimal conditions compared to traditional DoE approaches whilst assessing the relative individual parameter effect by analysing the model sensitivity. Considering the physico-chemical properties of both mRNA and ligand, the level of impurities and ultimately process cost, we have chosen three parameters to optimise, namely, the NaCl concentration in the

mobile phase, the mRNA feed concentration and the residence time (Table 1). NaCl concentration plays an important role in the balance between the hydrogen bonding forces and the electrostatic interactions between the mRNA and the ligand used (section 3.1). The mRNA feed concentration will impact binding but also how the impurities will interact with the ligand. Additionally, low product concentrations require higher processing volumes having a strong impact on process costs. Finally, the mRNA itself is a large biomolecule (1 to 100 nm) that can have multiple conformations, depending on its sequence and on the chosen binding conditions, influencing transport phenomena (e.g. adsorption and diffusion).

Table 1. Chromatography parameters used in the Bayesian optimisation methodology to increase the dynamic binding capacity (DBC).

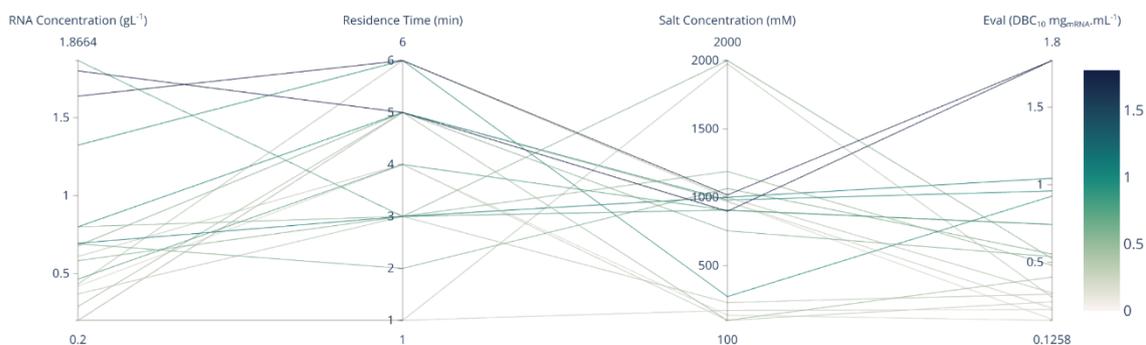
Parameters	Range	
	Min.	Max.
mRNA feed concentration (g.L <sup>-1</sup> )	0.2	2
Column residence time (min)	1	6
NaCl concentration (mM)	100	2000

The initial conditions of the Bayesian optimisation model used were randomised and the resulting calculated DBC<sub>10</sub> for each experimental design suggested was fed back into the model. This allowed the model to be updated after each optimisation cycle with new experimental conditions suggested until an optimised condition was proposed (maximum DBC<sub>10</sub>). After 20 iterations, the model converged to an optimal parameter configuration, achieving a maximum DBC<sub>10</sub> of 1.8 mg<sub>ma</sub>.mL<sub>resin</sub><sup>-1</sup> (Figure 2a). In general, higher DBC<sub>10</sub> are obtained when mRNA feed concentrations are above 1.5 g.L<sup>-1</sup> and the residence time is above 5 min, and medium NaCl concentrations are between 0.8 and 1.2 M. These results were supported by the explanation models and the overall importance of each parameter regarding its impact in the model prediction given by the SHapley Additive exPlanation (SHAP) values.

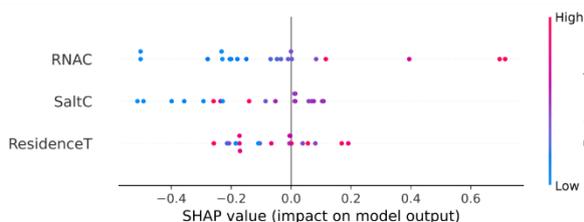
According to this evaluation, the mRNA feed concentration is the parameter with the highest estimated impact on the model predictions with the extreme mRNA concentrations (1.6 to 1.8 mg.mL<sup>-1</sup>) impacting the model. On the other hand, NaCl concentration at concentrations of approx. 1M seems to impact positively the model while lower concentrations lower (~0.1M) and higher (2M) impact negatively the results. NaCl concentrations have already been reported as having a significant impact on DBCs for other stationary phases[26,32,38], such as monoliths, with optimal results obtained with NaCl concentrations between 0.5 and 1.25 M. Higher concentrations of this salt can potentially create instability of the mRNA itself, resulting

in precipitation and corresponding loss of the molecule [26]. Regarding the residence times, higher values seem to impact the model positively. However, the relative impact is lower compared to the other parameters considered. As for the salt, the effect of the residence time on the DBCs is highly dependent on the stationary phase used[30], and attributed to the difference pore sizes that resins present compared with monoliths or fibres. The optimisation methodology allowed to increase the DBC<sub>10</sub> 7.5x from 0.24 to 1.8 mg<sub>mRNA</sub>.mL<sub>resin</sub><sup>-1</sup> with solely 20 experimental runs. The DBC<sub>10</sub> achieved are comparable with values obtained for pure mRNA samples [30,32] using directly IVT samples, showing that most of the impurities do not interact with the solid phase. It is worthy to mention that mRNA size and the mRNA sequence itself can impact the resin capacity, i.e. the DBC<sub>10</sub> can be process specific.

#### a) Parameter Space Coverage and Evaluation



#### b) Impact on Model Output



#### c) Feature Importance Summary

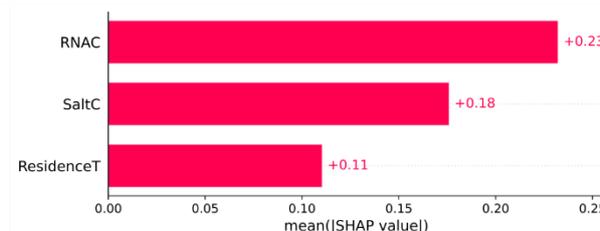


Figure 3. Optimisation of the binding capacity (DBC) of 1mL POROS™ GoPure™ Oligo (dT)25 pre-packed columns to be used for the separation of mRNA from IVT impurities using Bayesian optimisation. a) Parameter space coverage (mRNA feed concentration, feed concentration and salt concentration) for all the DBC runs performed with their respective DBC<sub>10</sub> evaluation (mg<sub>mRNA</sub>.mL<sub>resin</sub><sup>-1</sup>) b) Impact of the evaluated parameter values on the model prediction values for the Gaussian Process regressor used as the surrogate model for the Bayesian optimisation methodology. RNAC corresponds to RNA concentration (g.L<sup>-1</sup>), SaltC to NaCl concentration (mM) and ResidenceT to residence time (min). c) Parameter (feature) importance summary computed from the average SHapley Additive exPlanation (SHAP) values for the Gaussian Process regressor predictions across all DBC<sub>10</sub> experimental data. RNAC corresponds to RNA concentration (g.L<sup>-1</sup>), SaltC to NaCl concentration (mM) and ResidenceT to residence time (min).

## Evaluation of process-related impurities binding on oligo-dT ligands

DNA is a process-related impurity that is required to be removed after IVT reactions since the presence of DNA in the final product can lead to the activation of the IFN pathway within the cells causing a strong immune response[14]. To date, the acceptable limit of DNA in the drug product is  $330 \text{ ng}_{\text{DNA}} \cdot \text{mg}_{\text{RNA}}^{-1}$ [39]. The removal of DNA from the IVT mixture is challenging due to the physico-chemical similarity with mRNA. Traditionally, DNA is removed by an initial digestion using DNase I enzymes [40,41] and followed by a purification step such as chromatographic step or a tangential flow filtration. To avoid the use of an intermediate digestion in order to reduce processing time and costs, we explored the binding ability of the DNA to the stationary phase with increasing the salt concentration in the mobile phase (Figure 3). By analysing the chromatograms obtained for DNA binding, represented in Figure 3a, it is possible to observe that DNA binds with an increase of salt concentration in the binding buffer. In fact, DNA binds to the oligo-dT ligand with a maximum yield of  $71\% \pm 0.5$ , at 2 M NaCl (Figure 3b) with  $45\% \pm 0.3$  binding at the optimal binding conditions for mRNA (section 3.2). These results are corroborated with the agarose gel analysis (Figure 3c). DNA can be eluted from the column separating this impurity from the mRNA by performing a washing at a low salt concentration,  $5 \text{ mS} \cdot \text{cm}^{-1}$ . This challenge in separation can be attributed to the hydrogen bonding forces which bind to the linear small PCR derived DNA.

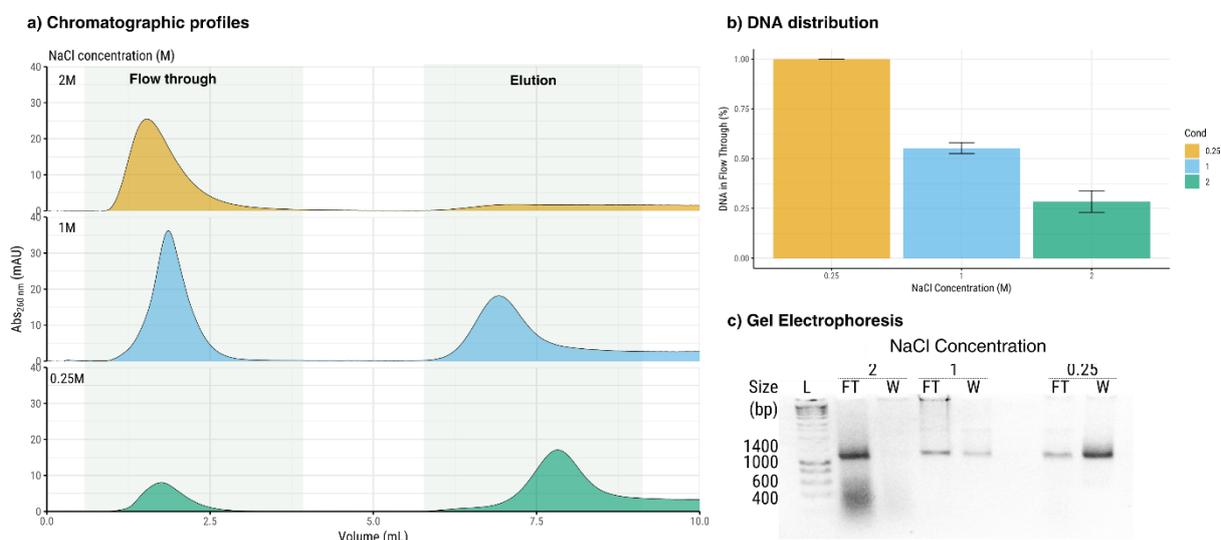
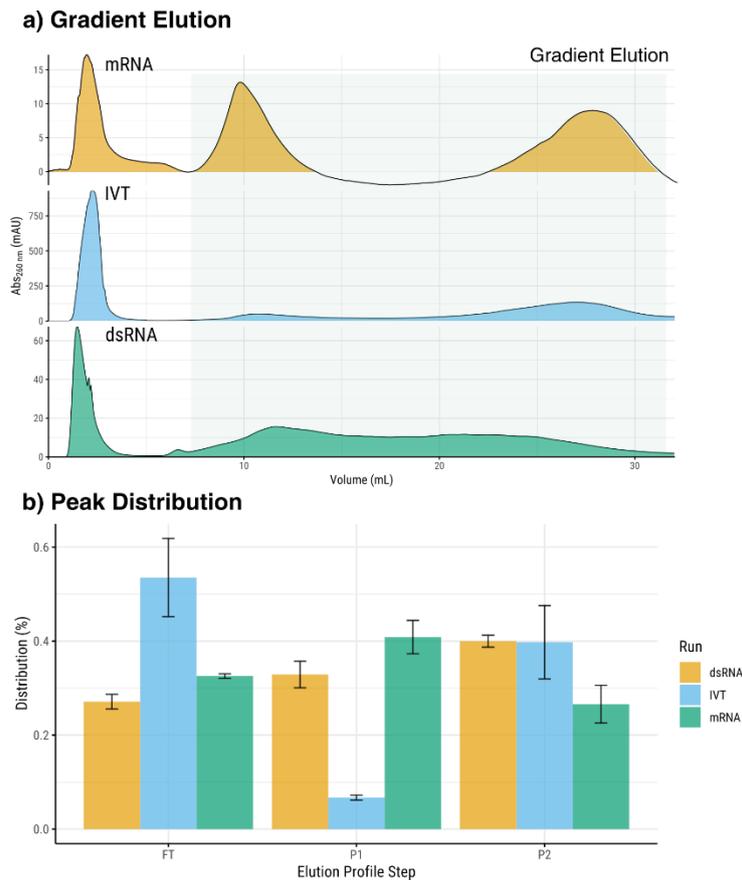


Figure 3. DNA binding studies on 1mL POROS™ GoPure™ Oligo (dT)25 pre-packed column in optimal conditions to bind mRNA directly from IVT reactions. a) Chromatographic profiles of DNA under different salt concentrations i.e. 0.25M NaCl (green), 1M NaCl (blue) and 2M NaCl (yellow) in TE buffer, pH 7.4. The column was equilibrated with the corresponding binding buffer for 5CV and washed with 2 CV. The DNA was eluted with TE buffer, pH 7.4. b) Percentage of DNA present in the flowthrough of the binding step under the different salt concentrations tested. c) Agarose gel electrophoresis of the DNA obtained during the flow through (FT) and elution (W) of the different NaCl concentrations evaluated. Lane L corresponds to the DNA ladder.

IVT reactions produce shorter abortive mRNA products and double stranded RNA (dsRNA). These have to be removed as that can be responsible for translation inhibition, and strong immune responses, ultimately leading to uncontrolled immune-inflammatory reactions such as myocarditis[42]. On the other, the removal of dsRNA can increase the protein production up to 10-1000 fold[18]. As the DNA, the dsRNA also binds to the oligo-dT ligand, and can be eluted during the elution step (data not shown). This can be attributed to the existing of regions that not double stranded such as the poly-A tail region or binding to DNA itself. This effect was further evaluated by exploring the elution differences of ssRNA and dsRNA using pure samples and IVT samples under a shallow gradient from 5 mS.cm<sup>-1</sup> washing buffer to water for 20 CVs (Figure 4).

Under these conditions, a different dsRNA population are observed in the elution peaks (Figure 4a). Since the dsRNA is produced during IVT by multiple stages, short transcripts can be released during the binding stage of the RNA polymerase enzyme to the DNA template (abortive cycling) [43], hybridising with the ssRNA. Furthermore, the RNA polymerase can also produce shorter (n-i) or longer (n+i) transcripts[44] forming intermolecular and intramolecular duplexes[45,46]. Additionally, T7 RNA polymerase can use the RNA as a template[47], producing full length dsRNA strands[48]. This results in a dsRNA population that can have different binding strengths within the population itself and to the ssRNA. The dsRNA population influence the elution of ssRNA when an IVT sample is directly injected. This is observed in the dsRNA peak distribution which is similar the IVT sample (P2), while mRNA is slightly shifted to the left (P1) (Figure 4.b).

The salt can have an impact on the secondary structures of the nucleic acids. Previous studies have shows that the elution of plasmid DNA from anion exchange chromatography is independent of the size, and that the conformation changes have a higher impact on the molecule's behaviour[49]. Achieving separation of all mRNA populations only resourcing with differences in hydrogen bonding seems difficult to achieve, as changes in the forces not only affect the interaction with mRNA and ligand but also with all the mRNA species present in the IVT mixture.



**Figure 4.** Elution profiles of mRNA and dsRNA using 1mL POROS™ GoPure™ Oligo (dT)25 pre-packed column under optimal conditions to bind mRNA directly from IVT reactions. a) Sample containing purified dsRNA (yellow), IVT (blue) and pure mRNA (green) are injected and washed as previously described (see Figure 1). A gradient elution to water is performed for 20 CV at a flow rate of 300 cm.h<sup>-1</sup>. b) Peak distribution analysis of purified dsRNA (yellow), IVT (blue) and pure mRNA (green), by % percentage of the total area obtained by peak integration. FT corresponds to flow through during binding, and P1 and P2 to corresponds the two peaks observed during gradient elution, respectively.

## Conclusions

Oligo-dT can be successfully used to capture and purify mRNA directly from IVT reaction. Exploring optimal binding conditions parameters, namely NaCl concentration, mRNA feed concentration, and residence time, using the Bayesian optimisation previously explored[9], was performed with the goal of maximising dynamic binding capacity at 10% of breakthrough (DBC10). This methodology allowed to improve 7.5 times capacity obtained initially, with only 20 iterations. The algorithm converged to optimal parameter configurations, achieving a DBC10 of 1.8 mg<sub>ma</sub>.mL<sub>resin</sub><sup>-1</sup>.

Impurities present in the final product present a major hurdle to the effect of the mRNA within the cells, as they can modulate its response. Owing to its physico-chemical similarities to

mRNA, DNA template and dsRNA present a challenge in its separation from the target. We explored the behaviour of these two impurities throughout the chromatography step. In the case of DNA, the concentration NaCl used during binding can strongly influence its interaction with the resin. At high NaCl concentrations, the hydrogen bonding becomes the prominent force, and binding to the ligand can occur. Nevertheless, elution can be achieved with a washing step, and the use of enzymes to digest can be avoided, which ultimately can decrease the manufacturing costs. dsRNA populations seem to interact with both mRNA and ligand, and no complete separation can be achieved. Exploration of the hydrogen binding forces impact on the ligand binding to different mRNA sequences and as well as in the potential product-related impurities can further enhance the potential of this affinity chromatography to simplify the mRNA manufacturing process.

## References

- [1] Sullenger BA, Nair S. From the RNA world to the clinic. *Science* 2016;352:1417–20. <https://doi.org/10.1126/science.aad8709>.
- [2] Rosa SS, Prazeres DMF, Azevedo AM, Marques MPC. mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine* 2021;39:2190–200. <https://doi.org/10.1016/j.vaccine.2021.03.038>.
- [3] Grinsted J, Liddell J, Bouleghlimat E, Kwok KY, Taylor G, Marques MPC, et al. Purification of therapeutic & prophylactic mRNA by affinity chromatography. *Cell Gene Ther Insights* 2022;8:335–49. <https://doi.org/10.18609/cgti.2022.049>.
- [4] Curreri A, Sankholkar D, Mitragotri S, Zhao Z. RNA therapeutics in the clinic. *Bioeng Transl Med* 2023;8:e10374. <https://doi.org/10.1002/btm2.10374>.
- [5] Barbier AJ, Jiang AY, Zhang P, Wooster R, Anderson DG. The clinical progress of mRNA vaccines and immunotherapies. *Nat Biotechnol* 2022;40:840–54. <https://doi.org/10.1038/s41587-022-01294-2>.
- [6] Wang Y-S, Kumari M, Chen G-H, Hong M-H, Yuan JP-Y, Tsai J-L, et al. mRNA-based vaccines and therapeutics: an in-depth survey of current and upcoming clinical applications. *J Biomed Sci* 2023;30:1–35. <https://doi.org/10.1186/s12929-023-00977-5>.
- [7] Pollard C, De Koker S, Saelens X, Vanham G, Grooten J. Challenges and advances towards the rational design of mRNA vaccines. *Trends Mol Med* 2013;19:705–13. <https://doi.org/10.1016/j.molmed.2013.09.002>.
- [8] Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines—a new era in vaccinology. *Nat Rev Drug Discov* 2018;17:261. <https://doi.org/10.1038/nrd.2017.243>.
- [9] Rosa SS, Nunes D, Antunes L, Prazeres DMF, Marques MPC, Azevedo AM. Maximizing mRNA vaccine production with Bayesian optimization. *Biotechnol Bioeng* 2022;119:3127–39. <https://doi.org/10.1002/bit.28216>.
- [10] Bancel S, Issa WJ, Aunins JG, Chakraborty T. Manufacturing methods for production of RNA transcripts 2018.
- [11] Wochner A, Roos T, Ketterer T. Methods and means for enhancing rna production. US20170114378A1, 2017.
- [12] Skok J, Megušar P, Vodopivec T, Pregeljč D, Mencin N, Korenč M, et al. Gram-Scale mRNA Production Using a 250-mL Single-Use Bioreactor. *Chem Ing Tech* 2022;94:1928–35. <https://doi.org/10.1002/cite.202200133>.

- [13] Henderson JM, Ujita A, Hill E, Yousif-Rosales S, Smith C, Ko N, et al. Cap 1 Messenger RNA Synthesis with Co-transcriptional CleanCap® Analog by In Vitro Transcription. *Curr Protoc* 2021;1:e39. <https://doi.org/10.1002/cpz1.39>.
- [14] Linares-Fernández S, Lacroix C, Exposito J-Y, Verrier B. Tailoring mRNA Vaccine to Balance Innate/Adaptive Immune Response. *Trends Mol Med* 2020;26:311–23. <https://doi.org/10.1016/j.molmed.2019.10.002>.
- [15] Organization WH, others. Evaluation of the Quality, Safety and Efficacy of RNA-Based Prophylactic Vaccines for Infectious Diseases: Regulatory Considerations. Retrieved January 2020;1:2022.
- [16] McKenna SA, Kim I, Puglisi EV, Lindhout DA, Aitken CE, Marshall RA, et al. Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nat Protoc* 2007;2:3270–7. <https://doi.org/10.1038/nprot.2007.480>.
- [17] Kim I, McKenna SA, Puglisi EV, Puglisi JD. Rapid purification of RNAs using fast performance liquid chromatography (FPLC). *RNA* 2007;13:289–94. <https://doi.org/10.1261/rna.342607>.
- [18] Kariko K, Muramatsu H, Ludwig J, Weissman D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Res* 2011;39:e142–e142. <https://doi.org/10.1093/nar/gkr695>.
- [19] Weissman D, Pardi N, Muramatsu H, Karikó K. HPLC Purification of In Vitro Transcribed Long RNA. In: Rabinovich PM, editor. *Synth. Messenger RNA Cell Metab. Modul. Methods Protoc.*, Totowa, NJ: Humana Press; 2013, p. 43–54. [https://doi.org/10.1007/978-1-62703-260-5\\_3](https://doi.org/10.1007/978-1-62703-260-5_3).
- [20] Easton LE, Shibata Y, Lukavsky PJ. Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* 2010;16:647–53. <https://doi.org/10.1261/rna.1862210>.
- [21] Von Der Mülbe F, Reidel L, Ketterer T, Gontcharova L, Bauer S, Pascolo S, et al. Method for producing rna. PCT/EP2015/000959; US10017826B2; WO/2016/180430A1, 2015.
- [22] Miklavčič R, Megušar P, Kodermac ŠM, Bakalar B, Dolenc D, Sekirnik R, et al. High Recovery Chromatographic Purification of mRNA at Room Temperature and Neutral pH. *Int J Mol Sci* 2023;24:14267. <https://doi.org/10.3390/ijms241814267>.
- [23] Megušar P, Miklavčič R, Korenč M, Ličen J, Vodopivec T, Černigoj U, et al. Scalable multimodal weak anion exchange chromatographic purification for stable mRNA drug substance. *ELECTROPHORESIS* n.d.;n/a. <https://doi.org/10.1002/elps.202300106>.
- [24] Funkner A, Dorner S, Sewing S, Johannes K, Broghammer N, Ketterer T, et al. Method for producing and purifying RNA, comprising at least one step of tangential flow filtration. PCT/EP2016/062152; WO/2016/193206, 2018.
- [25] Gagnon P, Goričar B, Peršič Š, Černigoj U, Štrancar AJ. Two new capture options for improved purification of large mRNA. *Cell Gene Ther. Insights.* 2020;6:1035-46.
- [26] Mencin N, Štepec D, Margon A, Vidič J, Dolenc D, Simčič T, et al. Development and scale-up of oligo-dT monolithic chromatographic column for mRNA capture through understanding of base-pairing interactions. *Sep Purif Technol* 2023;304:122320. <https://doi.org/10.1016/j.seppur.2022.122320>.
- [27] Sau SP, Larsen AC, Chaput JC. Automated solid-phase synthesis of high capacity oligo-dT cellulose for affinity purification of poly-A tagged biomolecules. *Bioorg Med Chem Lett* 2014;24:5692–4. <https://doi.org/10.1016/j.bmcl.2014.10.065>.
- [28] Aviv H, Leder P. Purification of Biologically Active Globin Messenger RNA by Chromatography on Oligothymidylic acid-Cellulose. *Proc Natl Acad Sci* 1972;69:1408–12. <https://doi.org/10.1073/pnas.69.6.1408>.

- [29] Engel BJ, Grindel BJ, Gray JP, Millward SW. Purification of poly-dA oligonucleotides and mRNA-protein fusions with dT25-OAS resin. *Bioorg Med Chem Lett* 2020;30:126934. <https://doi.org/10.1016/j.bmcl.2019.126934>.
- [30] Dewar EA, Guterstam P, Holland D, Lindman S, Lundbäck P, Brito dos Santos S, et al. Improved mRNA affinity chromatography binding capacity and throughput using an oligo-dT immobilized electrospun polymer nanofiber adsorbent. *J Chromatogr A* 2024;1717:464670. <https://doi.org/10.1016/j.chroma.2024.464670>.
- [31] Mencin N, Štepec D, Margon A, Vidič J, Dolenc D, Simčič T, et al. Development and scale-up of oligo-dT monolithic chromatographic column for mRNA capture through understanding of base-pairing interactions. *Sep Purif Technol* 2023;304:122320. <https://doi.org/10.1016/j.seppur.2022.122320>.
- [32] Tan Y-Z, Qiao L-Z, Wang S-S, Zhang J, Qian J, Zhu M, et al. Enhanced adsorption performance of varying-length mRNA on Oligo dT affinity resins through optimal pore size and grafting. *Biochem Eng J* 2024;203:109213. <https://doi.org/10.1016/j.bej.2023.109213>.
- [33] Sari Y, Sousa Rosa S, Jeffries J, Marques MPC. Comprehensive evaluation of T7 promoter for enhanced yield and quality in mRNA production. *Sci Rep* 2024;14:9655. <https://doi.org/10.1038/s41598-024-59978-5>.
- [34] Feng X, Su Z, Cheng Y, Ma G, Zhang S. Messenger RNA chromatographic purification: advances and challenges. *J Chromatogr A* 2023;1707:464321. <https://doi.org/10.1016/j.chroma.2023.464321>.
- [35] Heartlein M, DeRosa F, Dias A, Karve S. Methods for purification of messenger rna. WO/2014/152966A1, 2014.
- [36] Ribonucleic acid purification, 2014.
- [37] Gagnon P. Purification of Nucleic Acids: a handbook for Purification of DNA Plasmids and MRNA for Gene Therapy and Vaccines. BIA Separations; 2020.
- [38] Flook K. Supporting development of mRNA-based therapies by addressing large-scale purification challenges. *Cell and Gene Therapy Insights* 7.5 2021: 489-502
- [39] Daniel S, Kis Z, Kontoravdi C, Shah N. Quality by Design for enabling RNA platform production processes. *Trends Biotechnol* 2022;40:1213–28. <https://doi.org/10.1016/j.tibtech.2022.03.012>.
- [40] Cui T, Fakhfakh K, Turney H, Güler-Gane G, Toloczko A, Hulley M, et al. Comprehensive studies on building a scalable downstream process for mRNAs to enable mRNA therapeutics. *Biotechnol Prog* 2023;39:e3301. <https://doi.org/10.1002/btpr.3301>.
- [41] Issa WJ, Wang Y, Bancel S. Removal of DNA fragments in mRNA production process. WO/2014/152030A1; US10077439B2, 2016.
- [42] Milano G, Gal J, Creisson A, Chamorey E. Myocarditis and COVID-19 mRNA vaccines: a mechanistic hypothesis involving dsRNA. *Future Virol* 2021.
- [43] Martin CT, Muller DK, Coleman JE. Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* 1988;27:3966–74. <https://doi.org/10.1021/bi00411a012>.
- [44] Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* 1987;15:8783–98. <https://doi.org/10.1093/nar/15.21.8783>.
- [45] Gholamalipour Y, Karunanayake Mudiyanseilage A, Martin CT. 3' end additions by T7 RNA polymerase are RNA self-templated, distributive and diverse in character-RNA-Seq analyses. *Nucleic Acids Res* 2018;46:9253–63. <https://doi.org/10.1093/nar/gky796>.

- [46] Mu X, Greenwald E, Ahmad S, Hur S. An origin of the immunogenicity of in vitro transcribed RNA. *Nucleic Acids Res* 2018;46:5239–49. <https://doi.org/10.1093/nar/gky177>.
- [47] Cazenave C, Uhlenbeck OC. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc Natl Acad Sci U S A* 1994;91:6972–6.
- [48] Dousis A, Ravichandran K, Hobert EM, Moore MJ, Rabideau AE. An engineered T7 RNA polymerase that produces mRNA free of immunostimulatory byproducts. *Nat Biotechnol* 2023;41:560–8. <https://doi.org/10.1038/s41587-022-01525-6>.
- [49] Beck J, Biechele M, Repik C, Gruber P, Furtmüller PG, Hahn R. Desorption of plasmid DNA from anion exchangers: Salt concentration at elution is independent of plasmid size and load. *J Sep Sci* 2023;46:2200943. <https://doi.org/10.1002/jssc.202200943>.

# Chapter VIII – Multimodal chromatography as a one-step purification process for mRNA vaccines

---

**Key words:** Multimodal chromatography; mRNA; Downstream processing; Impurities removal; Purification

## Introduction

mRNA is rapidly ascending as one of most popular technologies beyond COVID-19 prophylactic vaccines. The technology is being explored in multiple applications, from prophylactic and cancer treatments, to protein replacement or gene editing [1], and multiple efforts have been made to improve its stability and effectiveness. Recent advances include the development of different mRNA modalities, such as self-amplifying [2] or circular RNA[3,4], as well as improving delivery systems[5]. The popularity can be attributed to the number of advantages that mRNA has over traditional technologies, such its precision and safety, as well as flexible manufacture[6]. mRNA is produced in an *in vitro* transcription reaction (IVT), where a DNA template is transcribed into mRNA catalysed by an RNA polymerase[7–10]. This reaction can yield up to  $12 \text{ g}_{\text{mRNA}} \text{ L}^{-1}_{\text{reaction}}$ , batch[8] and fed-batch mode[11]. The manufacturing process also includes a capping step that can be a second enzymatic reaction[6], or co-transcribed during IVT[12].

mRNA manufacture process, although it is an advantage over cell-based vaccines owing to its simplicity and flexibility, it is still a bottleneck due to the lack of a well-established platform[6]. In particular, the mRNA downstream can be composed of a combination of multiple unit operations that can include enzymatic digestion, precipitation, filtration or chromatography. Costs associated with downstream operation can reach 70% of the total manufacturing cost of biopharmaceuticals[13]. Chromatography is usually the unit operation of choice for the downstream manufacturing owing to its selectivity, versatility and scalability[14], which allows it to achieve a high purity product and high process yields. Thus, purification of mRNA using chromatography is also widely explored. More established methods include size exclusion (SEC) [15,16], anion-exchange (AEX)[17], or affinity chromatography [18–22]. However, these chromatography methods do not assure complete separation of product-related impurities, and often require to be coupled with enzymes or additional purification steps to ensure a high quality product. Ion-pair reverse phase (IP-RP) can be also used to purify mRNA, in particular to separate the RNA from dsRNA produced during IVT[23,24]. Hydroxyapatite[25] and hydrophobic interaction chromatography (HIC)[26,27] can be used as a polishing step to achieve high quality mRNA.

Multimodal resins can overcome the need of multiple purifications steps, as they combine multiple type of interactions, such as ion exchange, hydrogen-bonding or hydrophobic interactions, that can work in combination or individually, to increase the selectivity and achieve a higher purity[28]. Multimodal chromatography usually presents increased resolution, higher binding capacity and higher salt tolerance when compared with single-mode[29,30]. Different multimodal modes have been explored for the separation of mRNA. Core bead

chromatography combined with enzymatic digestion can separate mRNA from smaller and negatively charged impurities in a flowthrough mode[25]. Weak anion exchange combined with hydrogen bond was used to separate mRNA from pDNA[27], and a multimodal weak anion exchange matrix can separate mRNA from pDNA from pure[31] and IVT samples[32]. Nevertheless, the separation from dsRNA produced during IVT was not achieved.

Nuvia aPrime presents a number of characteristics that makes this resin an ideal candidate to use for the purification of high-quality mRNA. Nuvia matrix consists of a hydrophilic polymer, with a median particle diameter of 50  $\mu\text{m}$ , a density of ligand of 100  $\mu\text{eq.mL}^{-1}$  and a bimodal pore distribution containing large and small pores [33,34]. The ligand is composed of a quaternary group, as a strong anion exchanger, combined with a phenyl group, for hydrophobicity. These types of mixed-mode resins are widely applied in the purification of antibodies as a result of their ability to bind to nucleic acids, while antibodies flow through the solid phase[35]. The higher salt tolerance exhibited by these resins, attributed to the hydrophobicity of the ligand, results in higher capacity and selectivity[36,37]. Owing to this, these mixed-mode modalities have been successfully explored to purify nucleic acids in different stages of the downstream process. It was used to purify pDNA from crude *Escherichia coli* lysates [38], as polishing step to separate different pDNA isoforms[28,39] and minicircles[40], or to separate single stranded (ss)DNA scaffolds from double stranded (ds)DNA to produce high quality DNA-origami nanostructures[41].

Nuvia aPrime differs from other commercially available resins due to its lack of ethyl and methyl groups, which makes the ligand less hydrophobic[42], and a longer linker between the quaternary amine and the aromatic ring, which can potentially cause a weaker retention[43].

We explore the use of this multimodal chromatography to separate mRNA from product and process-related impurities derived from IVT, in a one-step purification process. The differences in the physico-chemical characteristics of the mRNA, namely its single strand nature relative to its impurities, allows to take advantage of the hydrophobicity forces to bind to the multimodal solid phase with a higher strength. By adjusting the conductivity value between 52 to 53  $\text{mS.cm}^{-1}$  in the binding mobile phase with sodium chloride (NaCl), it is possible to flow through the main impurities, namely NTPs, DNA template and dsRNA. The influence of the addition of different additives, namely EDTA, polysorbate and chaotropic salts, in the mobile phases to improve separation and yield, were also explored. The best elution conditions were achieved by exploring different pH during elution. Increase to a basic pH during elution allowed to improve mRNA recovery yield up to 84%. In the end, the use of 0.2 M glycine, pH 11, combined with 2M of NaCl and 2mM EDTA is sufficient to achieve a mRNA recovery mRNA recovery

yields  $81\pm 5\%$  with a purity of  $88\pm 2\%$ . (RP)HPLC and agarose gel analysis shows no presence of DNA and a reduction on dsRNA to  $0.07 \text{ g}_{\text{dsRNA}} \cdot \text{g}_{\text{RNA}}^{-1}$ .

Nuvia aPrime can be used to purify IVT samples, and achieve a high yield and high-quality product, without resorting to the use of enzymes or pre-purification steps. Additionally, separation is achieved in one-step, as the impurities flow through the solid phase during binding, and pure mRNA is eluted by increasing pH and NaCl. This simple methodology can potentially lower the process time, which ultimately can lead to a decrease in costs of mRNA manufacturing.

## Materials and Methods

All chemicals and reagents used to produce mRNA were purchased from ThermoFisher Scientific (USA), and all reagents used to optimise the chromatographic step were purchased from Sigma-Aldrich (USA), unless otherwise stated.

### **mRNA, DNA and dsRNA production**

#### **DNA template design and production**

Template design and plasmid production was performed as previously described [8,44]. The DNA template is produced by touchdown polymerase chain reaction. The forward and reverse primers used can be found elsewhere [8,44]. dsRNA template was produced by adding A T7 RNA polymerase promoter to the 5' end of the complementary strand using the reverse primer used. DNA template for IVT is produced by touchdown PCR (Applied Biosystems™ Veriti™ 96-Well Thermal Cycler, ThermoFisher Scientific, UK). Briefly, the reaction mixture comprises  $250 \text{ ng mL}^{-1}$  of template plasmid,  $0.4 \text{ }\mu\text{M}$  of forward and reverse primer, 1x VeriFi™ Buffer, 1x VeriMax Enhancer, and  $0.02 \text{ U }\mu\text{L}^{-1}$  high-fidelity VeriFi™ DNA Polymerase (PCR Biosystems, UK). The PCR conditions are the following: a denaturation step  $98^\circ\text{C}$  for 30 s; 20 cycles with a denaturation step at  $94^\circ\text{C}$  for 15 s, followed by an annealing step at  $65\text{-}55^\circ\text{C}$  for 30s, and extension step at  $72^\circ\text{C}$  for 1 min; 20 cycles of an annealing step at  $55^\circ\text{C}$  for 30 s and an extension step at  $72^\circ\text{C}$  for 1 min .; and a final extension is performed at  $72^\circ\text{C}$  for 2 min. The obtained PCR product is purified and concentrated using a GeneJET PCR Purification Kit. DNA is quantified using NanoDrop™ One Microvolume UV-Vis Spectrophotometer (ThermoFisher Scientific, UK).

#### **mRNA and dsRNA production**

mRNA and dsRNA were produced by IVT following the optimal conditions described in [8]. Briefly, 89 nM of template are mixed with a solution containing 7.7 mM NTPs, 5.3 mM dithiothreitol, 50 mM magnesium acetate, 40 mM Tris-HCl pH 6.8, 2.3 mM spermidine, 8 U

mL<sup>-1</sup> inorganic pyrophosphatase, 500 U mL<sup>-1</sup> Rnase inhibitor, and 7750 U mL<sup>-1</sup> of T7 RNA polymerase. This mixture is incubated at 43 °C for two hours on an Applied Biosystems™ Veriti™ 96-Well Thermal Cycler (ThermoFisher Scientific, UK). IVT samples were further purified or used directly in the chromatography optimisation. dsRNA IVT samples were diluted 1:2 with WFI water and further incubated in a decreasing temperature gradient, from 85°C to 30°C, using the same thermal cycler, to ensure annealing.

### **RNA purification**

mRNA and dsRNA samples were purified using MEGAclean™ Transcription Clean-Up kit. mRNA samples were incubated with 0.04 U µL<sup>-1</sup> of TURBO™ DNase for 30 min at 37 °C to remove the DNA template before purification. dsRNA samples were incubated with 20 U µL<sup>-1</sup> RNase T1 and 0.04 U µL<sup>-1</sup> of TURBO™ DNase for 30 min at 37 °C to remove DNA template and ssRNA species. After purification, the samples were precipitated -20°C by adding 500 mM pH 5 ammonium acetate and 2.5 volumes of ethanol. The samples were then centrifuged at 15000 x g for 15 min, air dried and resuspended by WFI water.

### **Chromatography optimisation**

Foresight Nuvia aPrime 4A Column, 1 mL pre-packed column (Bio-Rad, USA) was used in an AKTA Avant system (Cytiva, Sweden), equipped with a multiwavelength cell set at 260 nm. All chromatography parameters were controlled and monitored using the UNICORN 6.1 software (Cytiva, Sweden). All the runs were performed at a 1 ml.min<sup>-1</sup> (300 cm.h<sup>-1</sup>). Injection volume was set to 1.2 times the volume of the loop used. Column was cleaned after each run with 1M NaOH for 5 CV.

### **Gradient elution protocol**

Column is pre-equilibrated with 5 CV with the binding buffer, the sample is injected and then washed during 2 CV with the binding buffer. Gradient elution to 100% of the elution buffer (10 mM Tris, 2M NaCl, pH 7.4) is performed during 20 CV. Binding buffer conductivity was set to 40 mS.cm<sup>-1</sup> by mixing 10 mM Tris, pH 7.4 with 10 mM Tris, 2M NaCl, pH 7.4.

### **Step elution**

Column is pre-equilibrated with 5 CV with the binding buffer, the sample is injected and then washed during 2 CV with the binding buffer. A step elution to 100% of the elution buffer is performed during 5 CV.

### Binding optimisation

Mixtures of buffer A (10 mM Tris, pH 7.4) with buffer B (10 mM Tris, 2M NaCl, pH 7.4) were performing using AKTA Avant system (Cytiva, Sweden) to obtain the conductivity varying between 90 to 99 mS.cm<sup>-1</sup>, which corresponds to a pump B mixture of 50 to 55%, respectively. IVT samples were spiked with pure dsRNA. The step elution protocol was followed in this step.

### EDTA addition to the mobile phase

500 mM EDTA, pH 8, stock was prepared and added to both Buffer A (10 mM Tris, pH 7.4) and Buffer B (10 mM Tris, 2 M NaCl pH 7.4). Binding buffer conductivity was set to a conductivity of 52 to 53 mS/cm, corresponding to a pump B % of ~53%. The step elution protocol was followed in this step.

### Addition of additives to elution mobile phase

Buffer A corresponds to 10 mM Tris, 2 mM EDTA, pH 7.4, and Buffer B to 10 mM Tris, 2 mM 2 M NaCl pH 7.4. Binding buffer conductivity was set to a conductivity of 52 to 53 mS/cm, corresponding to a pump B % of ~53%. Elution buffer corresponds to 10 mM Tris, 2 mM 2 M NaCl pH 7.4 combined with the reagents and concentrations presented in Table 1. The step elution protocol was followed in this step.

Table 1. Additives names, composition and concentrations used in the elution buffer (10 mM Tris, 2 mM 2 M NaCl pH 7.4) to improve mRNA elution

Name	Composition	Concentration	Unit
Bellow CMC	Polysorbate 20	0.02	mg.mL <sup>-1</sup>
Above	Polysorbate 20	0.06	mg.mL <sup>-1</sup>
Mg <sup>2+</sup>	MgCl <sub>2</sub>	0.1	M
		0.25	M
		0.5	M
		1	M
		2	M
Gu-HCl	Gu-HCl	300	mM

### pH elution evaluation

As in the previous studies, binding buffer conductivity was set to a conductivity of 52 to 53 mS.cm<sup>-1</sup>, corresponding to a pump B % of ~53% of the mixture of Buffer A and B. Elution buffer corresponds to the buffer presented in Table 2 combined with 2M NaCl and 2M EDTA. The step elution protocol was followed in this step.

Table 2. Names, composition and concentrations used in buffer composition used to evaluate the effect of different pH. To all the buffers, 2M NaCl and 2 mM EDTA are also added.

Name	Composition	Concentration	Unit
pH 6	Mes	200	mM
pH 7.5	Tris	10	mM
pH 9	Tris	10	mM
pH 11	Glycine	400	mM
pH 13	NaOH	5	mM

### Chromatographic analysis

Peak areas were obtained using peak integration tools in Unicorn software. Peak was used to calculate the percentage of mRNA eluted to the total IVT injected, as well as mRNA yields from pure injected samples.

### Analytical methodologies

#### Gel electrophoresis

Samples obtained were analysed by gel electrophoresis[8,44]. Briefly, a 2% (w/v) agarose gel was prepared with 0.5× TBE buffer with 5.5 mM of magnesium chloride and pre-stained with SYBR® Safe DNA Gel Stain and run at 100 V for one hour. Densitometry analysis was performed using GelAnalyzer software (version 23.1.1).

#### Reverse-phase high performance liquid chromatography

HPLC analysis [8,44,45] was performed on an UltiMate 3000 UHPLC System (Thermo Fisher Scientific, UK) equipped with a VWD-3400 RS Rapid Detector, using RP-DNApac column (2.1 × 100 nm, Thermo Fisher Scientific, UK) at 80 °C with detection at 260 nm. Briefly, the column was pre-equilibrated with TAE buffer (100 mM Tris acetate, pH 7, 2.5 mM EDTA), with initial flow rate set to 0.2 mL min<sup>-1</sup>. After a 1 min washing step, the flow rate was increased to 0.35 mL.min<sup>-1</sup>, at a gradient of 0.25 m.min<sup>-1</sup> gradient over 30 sec. A first elution gradient is performed to 6% of the elution buffer (TAE buffer, 25% acetonitrile) for 30 s at 0.35 mL. min<sup>-1</sup>, followed by a gradient of 0.4 mL min<sup>-1</sup> over 4 min until 76.5% elution buffer is reached. Column is cleaned with a gradient to 100% elution buffer for 1 min, followed by 100% elution buffer for 3 min and re-equilibrated with TAE buffer for 6 min at 0.4 mL min<sup>-1</sup>. For analysis of mRNA concentration, samples were diluted 3 x with TAE buffer.

dsRNA and DNA analysis is performed by a digestion of 5  $\mu\text{L}$  of mRNA sample with 1  $\mu\text{L}$  of RNase T1 and RNase A, respectively, prior to injection, for 30 min at 37°C. Calibration curves were constructed using purified mRNA and dsRNA samples with known concentrations.

## Results and discussion

### Resin screening

Evaluating other commercially available multimodal resins implemented for the purification of pDNA[38,39] showed to be a challenge owing to the high hydrophobicity nature of the mRNA. This was already observed in a different anion-exchange multimodal resin for pDNA samples, where the RNA was only eluted during cleaning-in-place (CIP) [28]. The hydrogen bond forces between bases and the ligand can be associated with the ethyl (or methyl) groups may present challenge to elute mRNA. Nuvia aPrime was successfully used to bind and elute mRNA (Figure 1). The first studies, the binding was performed using a conductivity of  $\sim 75$  mS/cm, followed by a wash, and a gradient elution to 2M of NaCl. Binding with no salt resulted in no observation of peaks during the elution gradient step. We hypothesise that, without the presence of salt, electrostatic forces that result from the phosphate groups in the mRNA backbone, are predominant. When adding salt during elution, hydrophobic forces provided by the bases become dominant and the mRNA binds to the phenyl group and does not elute. Salt plays an important role in the conformation of the nucleic acids, which may influence its interactions with the ligands. pDNA elution from anionic exchange columns seems to be dependent on its conformation rather than its size [46]. The presence of salt may expose the bases and increase the strength of the hydrophobic forces during separation. This can be observed during the elution of dsRNA, as it elutes first than single stranded mRNA (Figure 1), at a conductivity between 90 to 100 mS/cm. This can be attributed to its double-stranded nature, where the bases are paired with each other, which lowers the hydrophobicity of the molecule. ssDNA elutes at a higher conductivity force ( $\sim 130$  mS/cm), owing to its exposed bases.

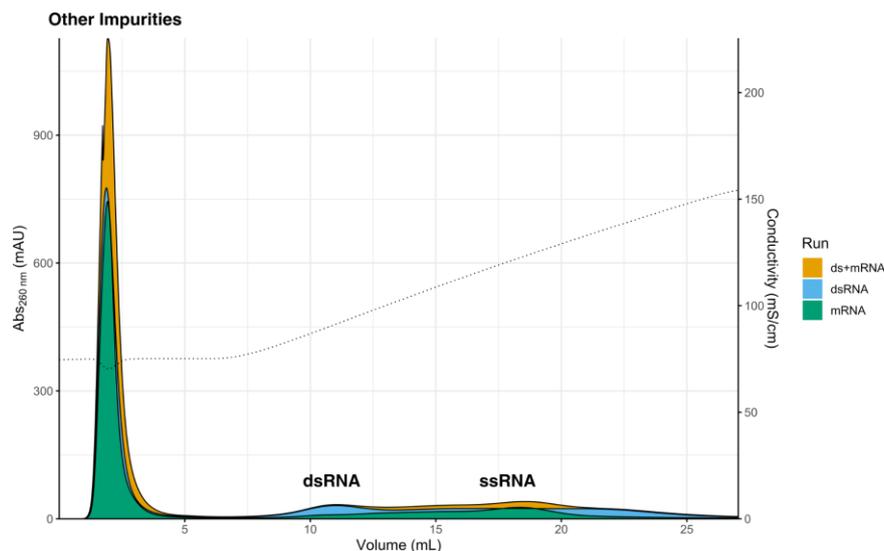


Figure 1. Chromatographic profile of dsRNA (blue), mRNA (green) and mRNA spiked with dsRNA (yellow) using Nuvia aPrime resin. Column was pre-equilibrated with a mixture of 10 mM Tris, pH 7.4 with 10 mM Tris, pH 7.4, 2M NaCl to a conductivity of ~74 mS/cm for 5 CVs. After injection, the column was washed with 2 CVs of the same buffer, followed by a gradient elution to 100% of 10 mM Tris, pH 7.4, 2M NaCl for 20 CV.

### Optimising binding conditions

A pivotal step in the optimisation of the separation of the ssRNA from the dsRNA is adjusting the salt concentration during binding. The salt concentration should be high enough to flow through the main impurities, namely DNA, NTPs and the dsRNA, but low enough to guarantee that the ssRNA still binds to the column. A first evaluation was performed by setting the binding conductivity at 52 and 53 mS/cm, followed by a step elution with 2M of NaCl (Figure 2). The impurities chosen to be analysed were DNA template, NTPs and pure dsRNA, owing to its similarity with the final product. NTPs are the substrate used during the IVT to produce the mRNA[6]. Owing to their physico-chemical similarities with mRNA, DNA and dsRNA are difficult-to-process impurities. However, their presence in the final product can modulate the immunologic response within the cells[47]. In fact, the removal of dsRNA improved 10-1000 fold the protein production in primary dendritic cells (DC)[24]. By analysing the chromatographic profile of the binding and elution of impurities (Figure 2), it is clear that none bind to the stationary phase with the conductivity ranging from 52 (Figure 2.a) to 53 mS/cm (Figure 2.b).

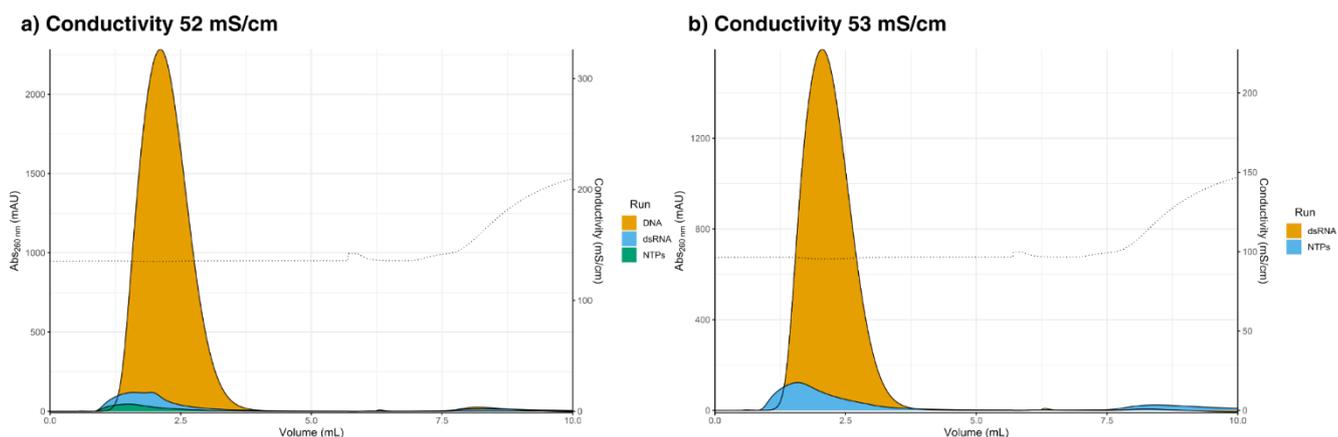


Figure 2. Chromatographic profile of DNA (yellow) and dsRNA (blue) and NTPs (green) using Nuvia aPrime resin. a) Binding conditions of  $\sim 52 \text{ mS}\cdot\text{cm}^{-1}$ , followed by a step elution to 100% of 10 mM Tris, pH 7.4, 2M NaCl. b) Binding conditions of  $\sim 53 \text{ mS}\cdot\text{cm}^{-1}$ , followed by a step elution to 100% of 10 mM Tris, pH 7.4, 2M NaCl.

To evaluate the mRNA binding and impurities removal during binding step, IVT samples spiked with purified dsRNA were injected with different concentrations of NaCl in the binding buffer, ranging from 92 to 99  $\text{mS}\cdot\text{cm}^{-1}$ , which corresponds to a B mixture of 50 to 55% (Figure 3). The chromatographic profiles obtained are observed in Figure 3.a). The different profiles show a higher mRNA retention during binding until a B% of 54, which corresponds to a conductivity between 92 to 93  $\text{mS}/\text{cm}$ . This may indicate that within these conductivity values, the hydrophobic forces play a more predominant role during binding. The samples obtained from the flow through, and elution were evaluated in terms of mRNA concentration by gel densitometry (Figure 3.b). With conductivity values above 97  $\text{mS}/\text{cm}$  (54%), the mRNA retention in the stationary phase decreases considerably, from 80% to below 60%, which may indicate that at this point, the hydrophobic forces are not strong enough to keep the mRNA retained, and it flows through the stationary phase. Owing to the double-stranded nature of the DNA, the hydrophobic forces are usually lower than the ssRNA and it should elute at lower NaCl concentrations. In fact, this difference has been widely explored to separate DNA from RNA by hydrophobic interaction chromatography[48]. By analysing the agarose gel electrophoresis (Figure 3.c), DNA is always found in the flow through sample in all the conditions evaluated. dsRNA showed the same profile as the DNA, and the no significant differences were found throughout the flow through and elution samples for all the conditions. NaCl concentration of  $\sim 92 \text{ mS}/\text{cm}$  was set for further studies.

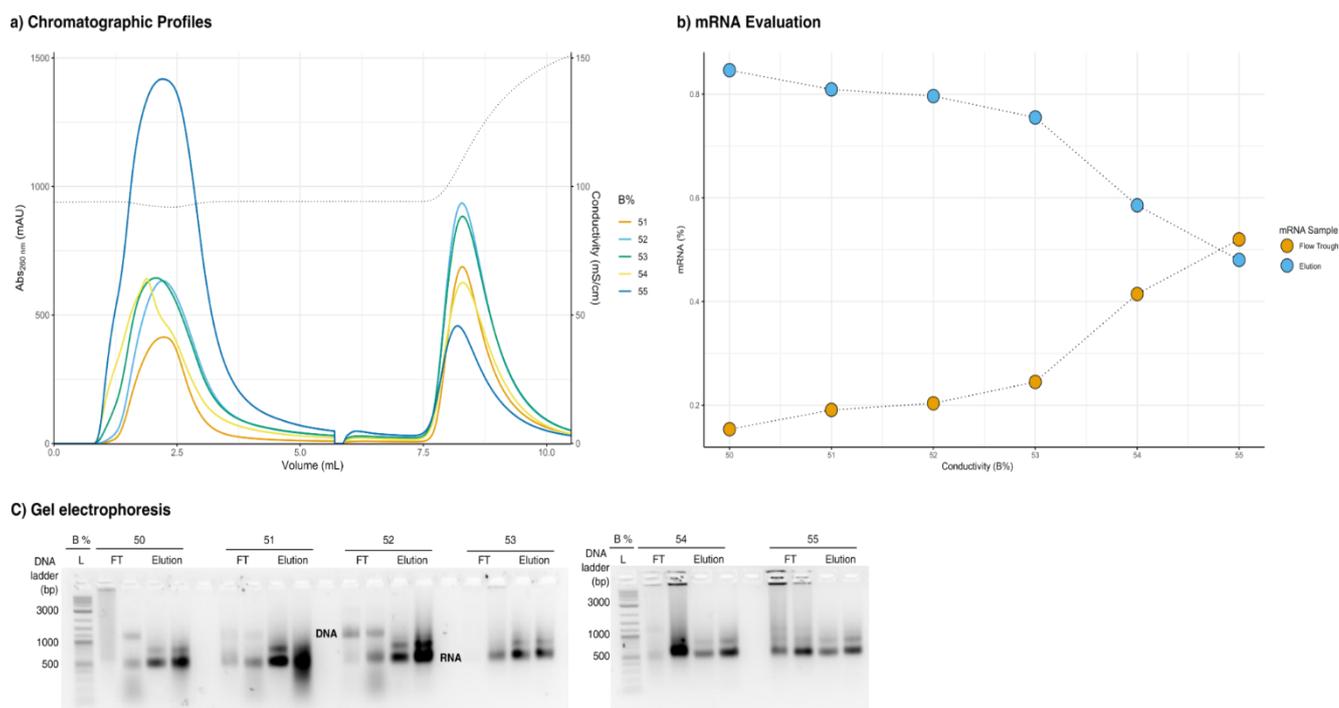


Figure 3. Binding conditions optimisation. a) Chromatographic profiles of IVT samples spiked with dsRNA with different binding buffers with varying conductivities (92 to 99  $\text{mS}\cdot\text{cm}^{-1}$ ) obtained by mixing a percentage of 10 mM Tris, pH 7.4, 2M NaCl (%B) with 10 mM Tris, pH 7.4. b) Evaluation of mRNA retention and elution from the solid phase during with the different binding conditions, c) Agarose gel electrophoresis of the fractions during binding (FT) and elution obtained for the different %B evaluated.

### Effect of EDTA in mobile phase

EDTA is a quenching agent widely used in nucleic acids processing owing to its ability to prevent DNase and RNase degradation. It can be added after IVT to stop the reaction, and protect the mRNA from degradation, but in combination with salts, it also can block nonspecific interactions that can influence the interactions between mRNA and the solid phase[49]. In fact, the use of EDTA combined with salt concentration during improved separation of ds species in a hydrogen bonding chromatography[27]. Additionally, EDTA combined with higher pH can improve separation between pDNA and mRNA in a weak anion exchange monolith[31]. The influence of the EDTA concentration during the chromatography step was evaluated by adding EDTA to the mobile phases used in concentrations ranging from 0 to 8 mM (Figure 4.a). The resulting separation was evaluated in terms of mRNA yield in elution, and percentage of mRNA eluted to the total IVT injected. The EDTA influence on the binding and elution observed is minimal, which means that the metal contamination in both IVT and mobile phases used is not directly affecting the separation, and separation is only achieved by the increased salt concentration (within the EDTA concentrations evaluated). A concentration of 2 mM of EDTA

was set as a component of the mobile phases used, as its presence can prevent RNA degradation by RNases during purification [21,50].

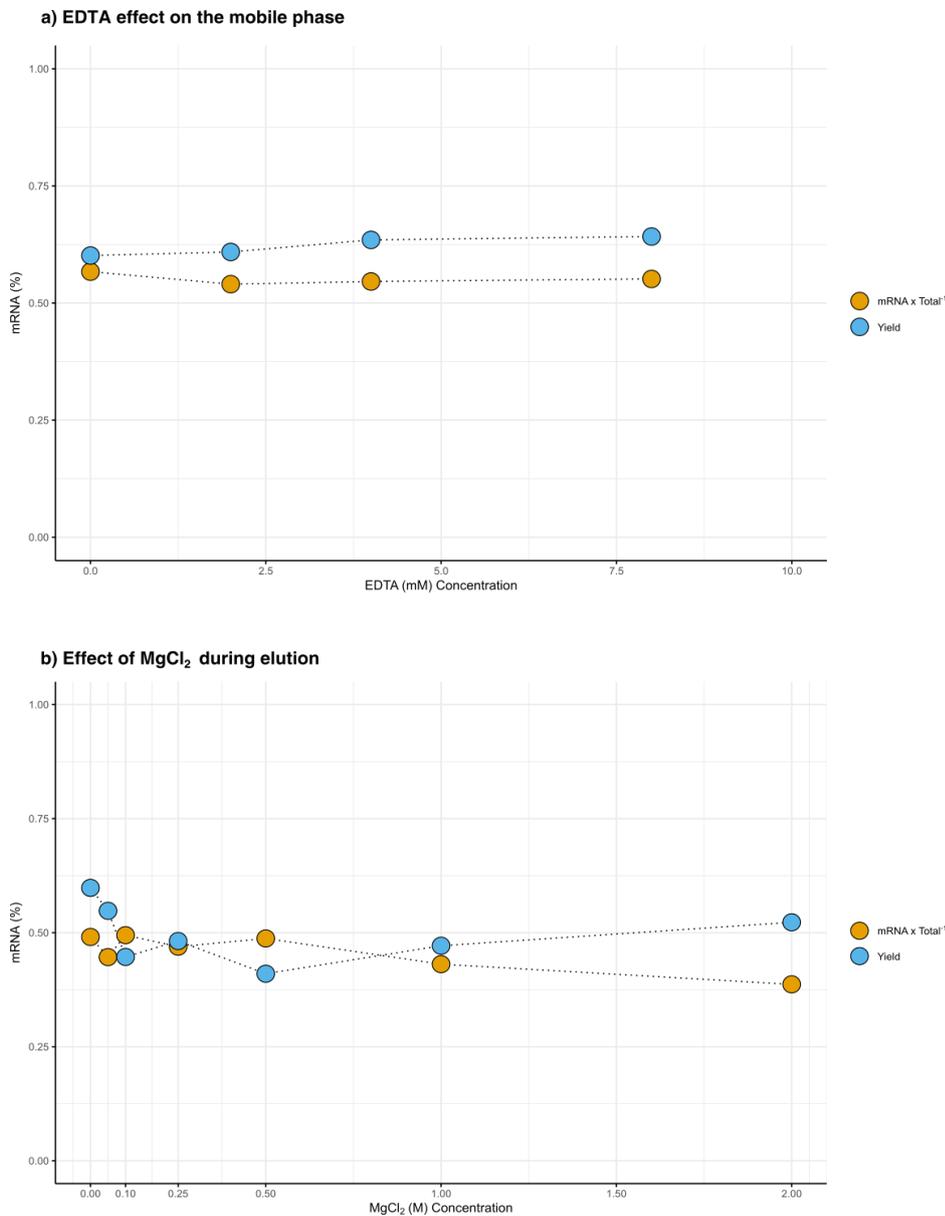


Figure 4. Effect of additives on mRNA elution. a) The effect of the addition of EDTA in varying concentration (0 to 8 mM) to both binding and elution buffer in terms of mRNA recovery to total IVT injected (orange dots) and mRNA recovery yield (blue dots). b) The effect of the addition of MgCl<sub>2</sub> in varying concentrations (0 to 2M) in the elution buffer in terms of mRNA recovery to total IVT injected (orange dots) and mRNA recovery yield (blue dots).

### The effect of additives on the elution profile

Mg<sup>2+</sup> plays an important role in the folding of RNA and stabilisation of tertiary structures[51]. Mg<sup>2+</sup> ions can interact with the RNA phosphate backbone, lowering the hydrostatic

interactions, and ultimately favouring interactions between the bases. This condensation effect might affect the binding and elution of the mRNA from the multimodal ligand. We explore this by adding  $\text{MgCl}_2$  to the elution mobile phase in different concentrations ranging from 5.5 mM to 2 M (Figure 4. b). Addition of  $\text{MgCl}_2$  was already explored to improve mRNA binding of the mRNA to oligo-dT ligand[22]. Additionally, addition of small concentrations of  $\text{MgCl}_2$  can improve separation of high molecular weight species [52] during size exclusion chromatography. We hypothesise that the condensation effect is responsible for the lower mRNA yield observed for the small  $\text{MgCl}_2$  concentrations. Nevertheless, this effect seems to be less visible with higher concentrations of  $\text{MgCl}_2$ . This may be derived from the chaotropic nature of  $\text{MgCl}_2$ , which can lead to a salting in effect, and ultimately decrease the retention derived from the hydrophobic aromatic ring[53].

Focusing on the improvement of mRNA recovery yields, the addition of polysorbate 20 to the elution buffer was evaluated. This non-ionic surfactant is widely used in antibody formulation, as it reduces protein aggregation and surface adsorption[54], owing to its hydrophobic nature. The addition of polysorbate 20 during the downstream processing steps can improve product recovery yields[55]. We explored the addition of two concentrations of polysorbate, below and above critical micelle concentration (CMC), and we evaluated its impact on mRNA recovery. Polysorbate 20 concentration below CMC showed no effect on mRNA recovery. However, polysorbate 20 concentrations above CMC, showed a slight improvement (6-8%) on mRNA yield. Increasing Polysorbate concentrations may have a positive effect on mRNA recovery, but they were not explored within the scope of this study.

The addition of other chaotropic agents, namely guanidine hydrochloride (Gu-HCl), was also explored. Gu-HCl in combination with NaCl seem to improve mRNA binding capacity to oligo-dT ligands[22], and improves isoform pDNA separation in anion exchange chromatography[56]. Gu-HCl has a similar effect as  $\text{MgCl}_2$ , as it can weaken hydrogen bonding and hydrophobic interactions[57]. In our experiments, the addition of 0.3 M Gu-HCl to elution mobile phase showed similar effects as  $\text{MgCl}_2$  at similar concentrations.

### **pH effect on elution**

pH gradients have been used to separate mRNA from pDNA using a multimodal weak anion exchange and hydrogen bond[27]. Lowering the pH during binding to pH 5 also allows to separate mRNA from NTPs and pDNA using a weak anion exchange monolith[31]. NaOH is used to achieve full elution of mRNA from an anion exchange high-performance liquid chromatography[58]. The effect of pH was evaluated by changing the elution mobile phase buffer, while maintaining NaCl and EDTA concentration (Figure 5). It is observed that by increasing the pH, mRNA recovery is also increased. Increasing the pH from 7.5 to 9 leads to

a 10% increase on mRNA recovery. Varying the pH from 9 to 11, 5% more mRNA is recovered, and from 11 to 13, a recovery yield of 84% is achieved. Higher pH values seem to lower the hydrophobic interaction between mRNA and the multimodal ligand. Since the use of NaOH can lead to irreversible denaturation of the mRNA, pH was set to 11, using glycine as a buffer. Glycine concentration was also varied between 0.2 to 0.4 M, and no difference in mRNA yield was observed.

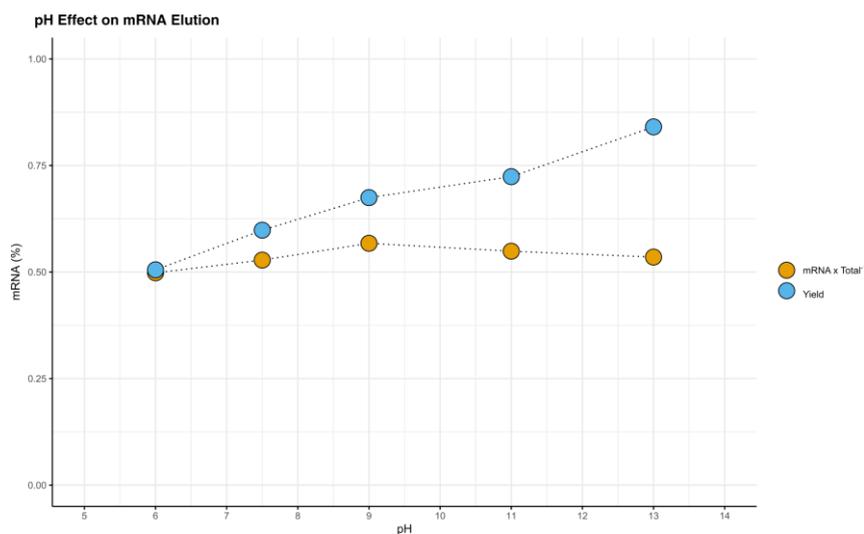


Figure 5. Effect of increasing pH during elution. mRNA elution was evaluated regarding mRNA recovery to total IVT injected (orange dots) and mRNA recovery yield (blue dots).

By optimising the elution, we were able to set up a one-step purification step based on the differences in charge and in hydrophobicity between the main impurities, namely NTPs, DNA template and dsRNA. These impurities flow through the stationary phase, while mRNA, owing to its higher hydrophobic nature, binds. Elution is achieved by increasing pH and salt in the mobile phase. The separation was evaluated by injecting samples containing pure dsRNA and pure mRNA, as well as IVT, in the optimised separation conditions (Figure 6). It is observed that most of the dsRNA species are eluted in the flow through, while mRNA binds to the stationary phase. A mRNA recovery yield of  $79\pm 5\%$  was obtained. No significant difference on mRNA recovery between IVT or pure mRNA samples is obtained.

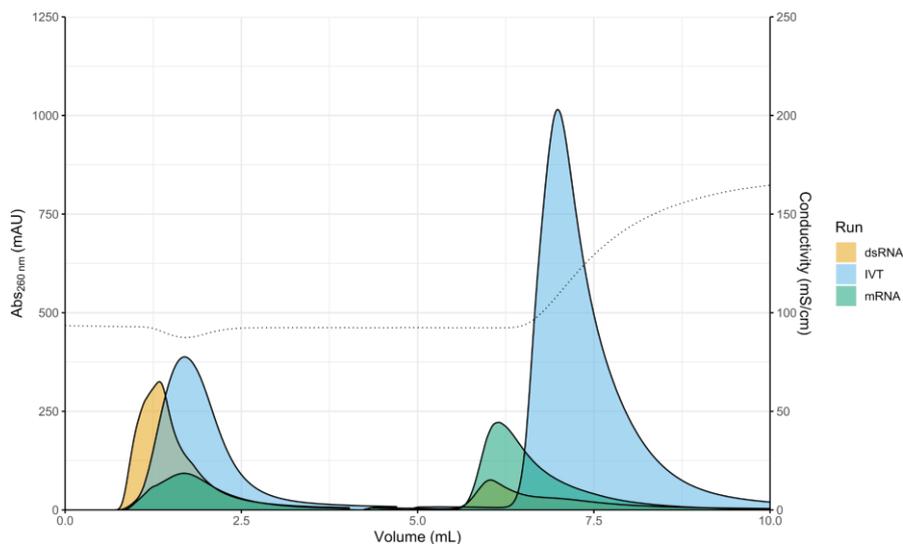


Figure 6. Chromatographic profile of dsRNA (yellow), mRNA (green) and IVT (blue) using optimised conditions. Column was equilibrated (5 CVs) and washed (2 CVs) with 10 mM Tris, 2 mM EDTA, pH 7.4 mixed with 10 mM Tris, 2 mM EDTA, 2 M NaCl, pH 7.4 to a conductivity of  $\sim 53$  mS/cm. Elution was achieved by doing a step elution with 100% of 200 mM glycine, 2 mM EDTA, 2 M NaCl for 5 CVs.

## Process validation

To validate the process, multiple IVT sample injections were performed and analysed in terms of mRNA recovery, DNA and dsRNA removal, and purity by (RP)HPLC and agarose gel electrophoresis (Figure 7). mRNA recovery yields  $81 \pm 5\%$  with a purity of  $88 \pm 2\%$ . The yield is in line with the results previously obtained with pure mRNA samples (Figure 6). The small fragments observed corresponding to the  $12 \pm 2\%$  in the (RP)HPLC analysis (Figure 7.b) may be derived by sample handling after purification, namely the precipitation process used to concentrate mRNA after purification. Tangential flow filtration (TFF) step may be implemented to concentrate and diafiltrate mRNA after purification[26], which may further improve mRNA purity. DNA removal was also analysed. RNase A was used in the purified sample to digest ds and ssRNA, in order to analyse the presence of other impurities, namely the DNA template. No DNA is detected in the final sample by agarose gel electrophoresis (Figure 7.b) or by (RP)HPLC analysis after purification and after digestion (Figure 7.c). Finally, using this method, the concentration of dsRNA was decreased to  $0.07 \text{ g}_{\text{dsRNA}} \cdot \text{g}_{\text{RNA}}^{-1}$ , which corresponds to a removal of 65%.

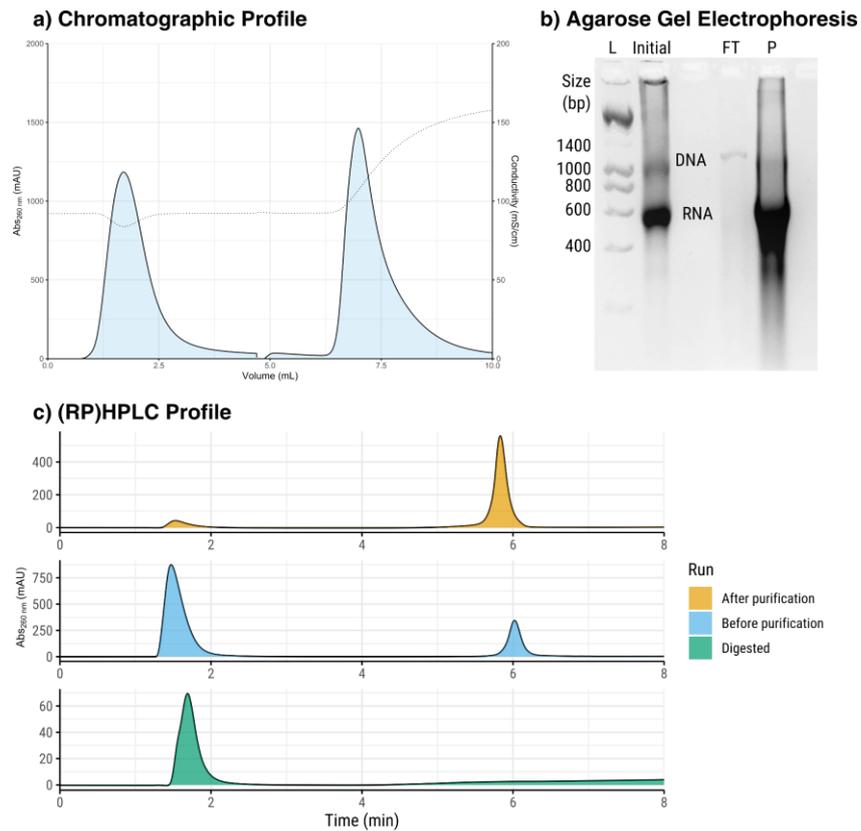


Figure 6. Evaluation of the mRNA yield and quality obtained by the one-step optimised purification process. a) Chromatographic profile of a IVT using optimised conditions. Column was equilibrated (5 CVs) and washed (2 CVs) with 10 mM Tris, 2 mM EDTA, pH 7.4 mixed with 10 mM Tris, 2 mM EDTA, 2 M NaCl, pH 7.4 to a conductivity of ~53 mS/cm. Elution was achieved by doing a step elution with 100% of 200 mM glycine, 2 mM EDTA, 2 M NaCl for 5 CVs. b) Agarose gel electrophoresis of the fractions of the flow through (FT) and peak (P) obtained by the optimised purification. c) HPLC analysis of a purified mRNA sample (yellow), IVT sample before purification (Blue) and after digestion with RNase A (Green).

## Conclusions

Nuvia aPrime, a multimodal resin that combines strong anion exchanger and hydrophobic interaction, can be used to separate highly pure mRNA from IVT samples. This scalable purification method explores the physico-chemical differences between the mRNA and its product and process-related impurities, namely its single-stranded nature.

Binding conditions were optimised by adjusting the conductivity required to flow through the main impurities, namely DNA template and dsRNA, while maximising the binding of mRNA to the solid phase. This is achieved with conductivity corresponding to ~52 to 53 mS.cm<sup>-1</sup>. The addition of EDTA was adjusted to 2 mM in both mobile phases to ensure protection against RNase during the chromatographic step. The effect of different additives in the elution step, such as Polysorbate 20 and magnesium chloride (MgCl<sub>2</sub>), were explored in order to maximise mRNA recovery. Nevertheless, optimal elution conditions were only obtained by increasing the

pH. This highlights the importance of hydrophobic interaction forces during the separation of the mRNA using this multimodal ligand. The combined effect of high pH and high salt concentration allowed to achieve a mRNA recovery yields  $81\pm 5\%$  with a purity of  $88\pm 2\%$ , with a concentration of dsRNA of  $0.07 \text{ g}_{\text{dsRNA}} \cdot \text{g}_{\text{RNA}}^{-1}$  and a complete separation from the DNA. This simple approach allows not only to decrease the number of steps required to achieve a high purity product, as well as decrease the process time of the chromatographic step, which ultimately can lead to a more cost-effective mRNA manufacturing process.

## References

- [1] Wang Y-S, Kumari M, Chen G-H, Hong M-H, Yuan JP-Y, Tsai J-L, et al. mRNA-based vaccines and therapeutics: an in-depth survey of current and upcoming clinical applications. *J Biomed Sci* 2023;30:1–35. <https://doi.org/10.1186/s12929-023-00977-5>.
- [2] Vogel AB, Lambert L, Kinnear E, Busse D, Erbar S, Reuter KC, et al. Self-amplifying RNA vaccines give equivalent protection against influenza to mRNA vaccines but at much lower doses. *Molecular Therapy* 2018;26:446–55.
- [3] Liu X, Zhang Y, Zhou S, Dain L, Mei L, Zhu G. Circular RNA: An emerging frontier in RNA therapeutic targets, RNA therapeutics, and mRNA vaccines. *Journal of Controlled Release* 2022;348:84–94. <https://doi.org/10.1016/j.jconrel.2022.05.043>.
- [4] Qu L, Yi Z, Shen Y, Lin L, Chen F, Xu Y, et al. Circular RNA vaccines against SARS-CoV-2 and emerging variants. *Cell* 2022;185:1728-1744.e16. <https://doi.org/10.1016/j.cell.2022.03.044>.
- [5] Witten J, Hu Y, Langer R, Anderson DG. Recent advances in nanoparticulate RNA delivery systems. *Proceedings of the National Academy of Sciences* 2024;121:e2307798120. <https://doi.org/10.1073/pnas.2307798120>.
- [6] Rosa SS, Prazeres DMF, Azevedo AM, Marques MPC. mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine* 2021;39:2190–200. <https://doi.org/10.1016/j.vaccine.2021.03.038>.
- [7] Pardi N, Hogan MJ, Porter FW, Weissman D. mRNA vaccines—a new era in vaccinology. *Nature Reviews Drug Discovery* 2018;17:261. <https://doi.org/10.1038/nrd.2017.243>.
- [8] Rosa SS, Nunes D, Antunes L, Prazeres DMF, Marques MPC, Azevedo AM. Maximizing mRNA vaccine production with Bayesian optimization. *Biotechnology and Bioengineering* 2022;119:3127–39. <https://doi.org/10.1002/bit.28216>.
- [9] Bancel S, Issa WJ, Aunins JG, Chakraborty T. Manufacturing methods for production of RNA transcripts 2018.
- [10] Wochner A, Roos T, Ketterer T. Methods and means for enhancing rna production. US20170114378A1, 2017.
- [11] Skok J, Megušar P, Vodopivec T, Pregeljc D, Mencin N, Korenč M, et al. Gram-Scale mRNA Production Using a 250-mL Single-Use Bioreactor. *Chemie Ingenieur Technik* 2022;94:1928–35. <https://doi.org/10.1002/cite.202200133>.
- [12] Henderson JM, Ujita A, Hill E, Yousif-Rosales S, Smith C, Ko N, et al. Cap 1 Messenger RNA Synthesis with Co-transcriptional CleanCap® Analog by In Vitro Transcription. *Current Protocols* 2021;1:e39. <https://doi.org/10.1002/cpz1.39>.
- [13] Mehta A. Downstream Processing for Biopharmaceuticals Recovery. In: Arora D, Sharma C, Jaglan S, Lichtfouse E, editors. *Pharmaceuticals from Microbes: The Bioengineering Perspective*, Cham: Springer International Publishing; 2019, p. 163–90. [https://doi.org/10.1007/978-3-030-01881-8\\_6](https://doi.org/10.1007/978-3-030-01881-8_6).

- [14] Akash MSH, Rehman K. Column Chromatography. Essentials of Pharmaceutical Analysis, Springer; 2020, p. 167–74.
- [15] McKenna SA, Kim I, Puglisi EV, Lindhout DA, Aitken CE, Marshall RA, et al. Purification and characterization of transcribed RNAs using gel filtration chromatography. *Nature Protocols* 2007;2:3270–7. <https://doi.org/10.1038/nprot.2007.480>.
- [16] Kim I, McKenna SA, Puglisi EV, Puglisi JD. Rapid purification of RNAs using fast performance liquid chromatography (FPLC). *RNA* 2007;13:289–94. <https://doi.org/10.1261/rna.342607>.
- [17] Easton LE, Shibata Y, Lukavsky PJ. Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography. *RNA* 2010;16:647–53. <https://doi.org/10.1261/rna.1862210>.
- [18] Grinsted J, Liddell J, Bouleghlimat E, Kwok KY, Taylor G, Marques MPC, et al. Purification of therapeutic & prophylactic mRNA by affinity chromatography. *Cell Gene Therapy Insights* 2022;8:335–49. <https://doi.org/10.18609/cgti.2022.049>.
- [19] Dewar EA, Guterstam P, Holland D, Lindman S, Lundbäck P, Brito dos Santos S, et al. Improved mRNA affinity chromatography binding capacity and throughput using an oligo-dT immobilized electrospun polymer nanofiber adsorbent. *Journal of Chromatography A* 2024;1717:464670. <https://doi.org/10.1016/j.chroma.2024.464670>.
- [20] Tan Y-Z, Qiao L-Z, Wang S-S, Zhang J, Qian J, Zhu M, et al. Enhanced adsorption performance of varying-length mRNA on Oligo dT affinity resins through optimal pore size and grafting. *Biochemical Engineering Journal* 2024;203:109213. <https://doi.org/10.1016/j.bej.2023.109213>.
- [21] Mencin N, Štepec D, Margon A, Vidič J, Dolenc D, Simčič T, et al. Development and scale-up of oligo-dT monolithic chromatographic column for mRNA capture through understanding of base-pairing interactions. *Separation and Purification Technology* 2023;304:122320. <https://doi.org/10.1016/j.seppur.2022.122320>.
- [22] Mencin N, Krušič A, Licen J, others. Increasing dynamic binding capacity of oligo (dT) for mRNA purification: experimental results using CIM 96-well plates. *Bioprocess Int* 2022.
- [23] Weissman D, Pardi N, Muramatsu H, Karikó K. HPLC purification of in vitro transcribed long RNA. *Synthetic Messenger RNA and Cell Metabolism Modulation: Methods and Protocols* 2013:43–54.
- [24] Kariko K, Muramatsu H, Ludwig J, Weissman D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Research* 2011;39:e142–e142. <https://doi.org/10.1093/nar/gkr695>.
- [25] Von Der Mülbe F, Reidel L, Ketterer T, Gontcharova L, Bauer S, Pascolo S, et al. Method for producing rna. PCT/EP2015/000959; US10017826B2; WO/2016/180430A1, 2015.
- [26] Funkner A, Dorner S, Sewing S, Johannes K, Broghammer N, Ketterer T, et al. Method for producing and purifying RNA, comprising at least one step of tangential flow filtration. PCT/EP2016/062152; WO/2016/193206, 2018.
- [27] Gagnon P, Goričar B, Peršič Š, Černigoj U, Štrancar A. Two new capture options for improved purification of large mRNA. *Cell Gene Therapy Insights* 2020;6:1035–46. <https://doi.org/10.18609/cgti.2020.114>.
- [28] Silva-Santos AR, Rosa SS, Prazeres DMF, Azevedo AM. Purification of Plasmid DNA by Multimodal Chromatography. In: Sousa Â, editor. *DNA Vaccines: Methods and Protocols*, New York, NY: Springer US; 2021, p. 193–205. [https://doi.org/10.1007/978-1-0716-0872-2\\_10](https://doi.org/10.1007/978-1-0716-0872-2_10).
- [29] Pinto IF, Aires-Barros MR, Azevedo AM. Multimodal chromatography: debottlenecking the downstream processing of monoclonal antibodies. *Pharmaceutical Bioprocessing* 2015;3:263–79. <https://doi.org/10.4155/pbp.15.7>.

- [30] Zhao G, Dong X-Y, Sun Y. Ligands for mixed-mode protein chromatography: Principles, characteristics and design. *J Biotechnol* 2009;144:3–11. <https://doi.org/10.1016/j.jbiotec.2009.04.009>.
- [31] Miklavčič R, Megušar P, Kodermac ŠM, Bakalar B, Dolenc D, Sekirnik R, et al. High Recovery Chromatographic Purification of mRNA at Room Temperature and Neutral pH. *International Journal of Molecular Sciences* 2023;24:14267. <https://doi.org/10.3390/ijms241814267>.
- [32] Megušar P, Miklavčič R, Korenč M, Ličen J, Vodopivec T, Černigoj U, et al. Scalable multimodal weak anion exchange chromatographic purification for stable mRNA drug substance. *ELECTROPHORESIS* 2023;44:1978–88. <https://doi.org/10.1002/elps.202300106>.
- [33] Roberts JA, Carta G. Protein adsorption and separation with monomodal and multimodal anion exchange chromatography resins. Part I. Multicomponent adsorption properties and frontal chromatography. *Journal of Chemical Technology & Biotechnology* 2022;97:3292–305. <https://doi.org/10.1002/jctb.7239>.
- [34] Watson MP, Winters MA, Sagar SL, Konz JO. Sterilizing Filtration of Plasmid DNA: Effects of Plasmid Concentration, Molecular Weight, and Conformation. *Biotechnology Progress* 2006;22:465–70. <https://doi.org/10.1021/bp050280s>.
- [35] Gagnon P. IgG Aggregate Removal by Charged-Hydrophobic Mixed Mode Chromatography. *Current Pharmaceutical Biotechnology* n.d.;10:434–9. <https://doi.org/10.2174/138920109788488888>.
- [36] Yoshimoto N, Itoh D, Isakari Y, Podgornik A, Yamamoto S. Salt tolerant chromatography provides salt tolerance and a better selectivity for protein monomer separations. *Biotechnology Journal* 2015;10:1929–34. <https://doi.org/10.1002/biot.201400550>.
- [37] Santarelli X, Cabanne C. Mixed Mode Chromatography: A Novel Way Toward New Selectivity. *Current Protein and Peptide Science* 2019;20:14–21. <https://doi.org/10.2174/1389203718666171024121137>.
- [38] Matos T, Queiroz JA, Bülow L. Plasmid DNA purification using a multimodal chromatography resin. *Journal of Molecular Recognition* 2014;27:184–9. <https://doi.org/10.1002/jmr.2349>.
- [39] Silva-Santos AR, Alves CPA, Prazeres DMF, Azevedo AM. A process for supercoiled plasmid DNA purification based on multimodal chromatography. *Separation and Purification Technology* 2017;182:94–100. <https://doi.org/10.1016/j.seppur.2017.03.042>.
- [40] Silva-Santos AR, Alves CPA, Monteiro G, Azevedo AM, Prazeres DMF. Multimodal chromatography of supercoiled minicircles: A closer look into DNA-ligand interactions. *Separation and Purification Technology* 2019;212:161–70. <https://doi.org/10.1016/j.seppur.2018.11.015>.
- [41] Silva-Santos AR, Paulo PMR, F. Prazeres DM. Scalable purification of single stranded DNA scaffolds for biomanufacturing DNA-origami nanostructures: Exploring anion-exchange and multimodal chromatography. *Separation and Purification Technology* 2022;298:121623. <https://doi.org/10.1016/j.seppur.2022.121623>.
- [42] Roberts JA, Kimerer L, Carta G. Effects of molecule size and resin structure on protein adsorption on multimodal anion exchange chromatography media. *Journal of Chromatography A* 2020;1628:461444. <https://doi.org/10.1016/j.chroma.2020.461444>.
- [43] Robinson J, Snyder MA, Belisle C, Liao J, Chen H, He X, et al. Investigating the impact of aromatic ring substitutions on selectivity for a multimodal anion exchange prototype library. *Journal of Chromatography A* 2018;1569:101–9. <https://doi.org/10.1016/j.chroma.2018.07.049>.

- [44] Sari Y, Sousa Rosa S, Jeffries J, Marques MPC. Comprehensive evaluation of T7 promoter for enhanced yield and quality in mRNA production. *Sci Rep* 2024;14:9655. <https://doi.org/10.1038/s41598-024-59978-5>.
- [45] Issa W, Packer M. Methods for hplc analysis. US20210163919A1, 2021.
- [46] Beck J, Biechele M, Repik C, Gruber P, Furtmüller PG, Hahn R. Desorption of plasmid DNA from anion exchangers: Salt concentration at elution is independent of plasmid size and load. *Journal of Separation Science* 2023;46:2200943. <https://doi.org/10.1002/jssc.202200943>.
- [47] Linares-Fernández S, Lacroix C, Exposito J-Y, Verrier B. Tailoring mRNA Vaccine to Balance Innate/Adaptive Immune Response. *Trends in Molecular Medicine* 2020;26:311–23. <https://doi.org/10.1016/j.molmed.2019.10.002>.
- [48] Diogo MM, Queiroz JA, Prazeres DMF. Chromatography of plasmid DNA. *Journal of Chromatography A* 2005;1069:3–22. <https://doi.org/10.1016/j.chroma.2004.09.050>.
- [49] Gagnon P. Purification of Nucleic Acids: a handbook for Purification of DNA Plasmids and MRNA for Gene Therapy and Vaccines. BIA Separations; 2020.
- [50] Dayeh DM, Cika J, Moon Y, Henderson S, Di Grandi D, Fu Y, et al. Comprehensive chromatographic assessment of forced degraded in vitro transcribed mRNA. *Journal of Chromatography A* 2024;1722:464885. <https://doi.org/10.1016/j.chroma.2024.464885>.
- [51] Misra VK, Draper DE. The linkage between magnesium binding and RNA folding1. *Journal of Molecular Biology* 2002;317:507–21. <https://doi.org/10.1006/jmbi.2002.5422>.
- [52] D'Atri V, Lardeux H, Goyon A, Imiolek M, Fekete S, Lauber M, et al. Optimizing Messenger RNA Analysis Using Ultra-Wide Pore Size Exclusion Chromatography Columns. *International Journal of Molecular Sciences* 2024;25:6254. <https://doi.org/10.3390/ijms25116254>.
- [53] Shukla AA, Peterson J, Sorge L, Lewis P, Thomas S, Waugh S. Preparative Purification of a Recombinant Protein by Hydrophobic Interaction Chromatography: Modulation of Selectivity by the Use of Chaotropic Additives. *Biotechnology Progress* 2002;18:556–64. <https://doi.org/10.1021/bp020038a>.
- [54] Cui Y, Cui P, Chen B, Li S, Guan H. Monoclonal antibodies: formulations of marketed products and recent advances in novel delivery system. *Drug Development and Industrial Pharmacy* 2017;43:519–30. <https://doi.org/10.1080/03639045.2017.1278768>.
- [55] Moleirinho MG, Rosa S, Carrondo MJT, Silva RJS, Hagner-McWhirter Å, Ahlén G, et al. Clinical-Grade Oncolytic Adenovirus Purification Using Polysorbate 20 as an Alternative for Cell Lysis. *Current Gene Therapy* 2018;18:366–74. <https://doi.org/10.2174/1566523218666181109141257>.
- [56] Černigoj U, Vidič J, Ferjančič A, Sinur U, Božič K, Mencin N, et al. Guanidine improves DEAE anion exchange-based analytical separation of plasmid DNA. *ELECTROPHORESIS* 2021;42:2619–25. <https://doi.org/10.1002/elps.202100210>.
- [57] Tsumoto K, Ejima D, Senczuk AM, Kita Y, Arakawa T. Effects of salts on protein–surface interactions: applications for column chromatography. *Journal of Pharmaceutical Sciences* 2007;96:1677–90. <https://doi.org/10.1002/jps.20821>.
- [58] Welbourne EN, Loveday KA, Nair A, Nourafkan E, Qu J, Cook K, et al. Anion exchange HPLC monitoring of mRNA in vitro transcription reactions to support mRNA manufacturing process development. *Frontiers in Molecular Biosciences* 2024;11:1250833.

# Chapter IX – Conclusions and Future work

---

Vaccines are the second most effective way of preventing diseases, only after clean water. During the recent Covid-19 pandemic event, it is estimated that, only in Europe, vaccination saved 1.4 millions lives of individual aged  $\geq 25$  years[1]. Several vaccine technologies have been developed throughout time, ultimately work on the same principle, stimulate the immune response to recognize a pathogen or part of a pathogen.

One of the great challenges on vaccine development is to design vaccine platform that is able to delivers a safe and effective vaccine that is able to respond epidemic events in only 100 days [2]. mRNA vaccines are a (almost) perfect technology to solve this challenge. mRNA vaccines are based on the principle that any target gene can be deliver as a mRNA molecule in any type of cell and have an immunologic response[3]. Owing to its precision, safeness and effectivity, this technology because widely popular in recent year. Today, over 60 ongoing clinical trials can be found, with a variety of applications that include prophylactic and cancer treatments, protein replacement or gene editing[4].

One of the strong points of mRNA technology is its simple and flexible manufacturing process. mRNA is produced in a cell-free enzymatic cascade reactions[3]. This is a well-defined reaction , that allows to reach a production of grams per litre of reaction in a matter of hours[5]. The mRNA manufacturing process can be divided in three main steps: 1) DNA template production; 2) mRNA production; 3) mRNA formulation. Firstly, the target sequence must be inserted into a DNA plasmid, and a linearised DNA template must be produced. This template is then used to produce mRNA in an *in vitro* transcription that uses an RNA polymerase as a catalyst, and nucleotides as a substrate. A capping structure must be added to the obtained mRNA, by either a second enzymatic reaction, or by a co-transcriptional reaction. To achieve a high-quality product, mRNA must be purified from the process and product-related impurities. Purification is a critical step as it can lead to a increase of 1000 fold in protein production inside cells[6]. Owing to its charge and size, mRNA needs to the formulated to be efficiently delivered to the target cells.

Although the advantages that the mRNA manufacturing presents over traditional vaccines, the process is still a bottleneck for this technology to achieve its full potential. The current state-of-art of the manufacturing process is still lacking, and process scalability is limited to the material and equipment available[7]. In order to deliver an efficient and cost-effective manufacturing process, it is required to extensive optimise each step.

This work is focused on the optimisation of the mRNA production process. An integrated perspective of process was used as the strategy for the design and development of each step, and with the goal of maximising production yield, while improving mRNA quality, and with the flexibility required to be easily adapted to different modes of operation.

The first objective of this work was to develop analytical methods that can be used throughout the manufacturing process, and that are accurate, robust and fast. Owing to their resolution, speed and availability, HPLC-based methods are ideal candidates to be used as a standard analytical technique to characterise mRNA and its impurities. A reverse-phase (RP) HPLC method was implemented to quantify mRNA and then adapted to quantify one of the main impurities produced during IVT, dsRNA. This was achieved by coupling a RNase treatment before analysis. The RNase chosen only to digest single-stranded RNA, leaving the dsRNA to be quantified by the (RP)HPLC. By using this technique, it is possible to quantify dsRNA in sample in under 30 min, with a similar sensitivity and precision with the current golden standard, dot blot. The technique showed to be robust enough to be during mRNA manufacturing process and that can be easily adapted for Process Analytical Technology (PAT) for future purposes.

IVT reaction is the upstream operation in the mRNA production tray. Owing to its defined nature of the IVT reaction makes this process highly prone to optimisation. Nevertheless, it is a complex reaction, as it contains a large number of variables, such multiple enzymes, template and nucleotides, co-factors and enhancers, that can affect the production outcome. A standard mRNA production can yield up to  $5 \text{ g}_{\text{mRNA}} \cdot \text{L}^{-1}$  in 4 hours. Nevertheless, there was no consensus in the reaction parameters. Using AI approaches to automate the experiment design, the IVT production yield was maximised to produce  $12 \text{ g}_{\text{mRNA}} \cdot \text{L}^{-1}$  of reaction in just 2 hours[8]. Notably, this approach only required a total of 60 reactions to achieve optimal reaction conditions from a total of 12 parameters evaluated. The results obtained outperform published industry standards and data reported in literature in terms of both achievable reaction yield and reduction of production time.

However, increasing IVT yields can also lead to an increase in the production of unwanted by-products, namely dsRNA. In an integrated perspective point-of-view reducing the dsRNA to residual levels can avoid intensive purification steps, which ultimately can make manufacturing process more cost-effective. Further optimisation of the IVT was performed by evaluation the effect of the DNA template on the reaction itself. Modifications downstream of the T7 RNA polymerase promoter were performed to evaluate its effect on mRNA yields and dsRNA production. In particular, transcription performance was optimised by modifying the sequence downstream of the T7 promoter with additional AT-rich sequences. These variants exhibited up to a 30% reduction in dsRNA byproduct levels compared to a wildtype T7 promoter, and have similar EGFP protein expression[9]. The results show non-coding regions can have an impact on mRNA production yields and quality, its effect should be further explored.

In order to achieve a quality product, mRNA must be purified after IVT. The mRNA downstream processing is dependent on multiple purification steps to achieve separation from the difficult-to-process impurities. Affinity chromatography is a popular method to purify mRNA. In particular, the commercially available oligo-dT ligand, that binds to the poly-A tail present in the mRNA molecules, is used in the study. Nevertheless, there is still a lack of knowledge in understanding how this ligand can be used to obtain pure mRNA from unpurified IVT samples. Optimal binding conditions were explored using a Machine Learning approach for sequential data-driven design-of-experiments. Using this model, an increase of 7.5-fold from the initial conditions was obtained, achieving  $1.8 \text{ mg}\cdot\text{mL}^{-1}$  resin, in only 20 runs. Looking into the impurities, namely DNA and dsRNA, we evaluated its behaviour throughout the purification step. The hydrogen bonding forces can lead to the binding of DNA, depending on the mobile phase force, but its elution can be achieved with a low salt wash before elution. The different dsRNA populations can interact with both ligand and the target mRNA through the hydrogen bonding forces, and its complete separation only using this affinity chromatography is challenging. Nevertheless, oligo-dT ligands can be used as a capture step after IVT, without resorting to multiple purification steps.

New chromatographic modalities were also explored in order to achieve complete separation of the mRNA from its impurities in a one-step purification process. Nuvia aPrime, a multimodal resin that combines strong anion exchanger with a phenyl group was used to purify directly IVT samples, without resorting to the use of enzymes or pre-purification step. Optimisation of the binding conditions allowed to flow through the main process and product-related impurities, namely NTPs, DNA template and dsRNA, while maximising the mRNA binding. In order to maximise mRNA recovery, pH was increased to 11. This allows to achieve a RNA recovery yield of  $81\pm 5\%$ , with a purity of  $88\pm 2\%$  with no detectable concentration of DNA and a reduction on dsRNA of 65%. Using this simple one-step method, a high yield process and high quality product can be obtained without the need of a tray of purification steps in the mRNA manufacturing platform. This can potentially lower the process time, which ultimately can lead to a decrease in costs of mRNA manufacturing.

Capping is still a bottleneck of the mRNA process, as it is the most impactful raw material in the mRNA manufacturing costs[10]. The use of Cleancap is a preferable method as it allows to achieve a complete mRNA molecule in one step. However, to this day there is only one supplier of this raw material. Enzymatic reaction based on viral capping enzymes can overcome this dependence, but usually requires a second enzymatic step, which increases the processing time of the manufacturing process. Design and development of capping strategies based on enzymatic reaction and that are scalable and cost effective, as well as that facilitate the translation to continuous processing, is one of the trends on mRNA manufacturing process

optimisation. This can be achieved by optimising of the enzymatic capping reaction and the analytical tools required to evaluate reaction progression. One-pot, where IVT and capping reactions are performed at the same time, or sequential enzymatic reactions, can be solution to simplify the enzymatic capping process and make more it attractive to be implemented in the mRNA manufacturing process.

The need for a well-established and reproducible manufacture process that renders high product quality at a lower cost is shifting the biomanufacturer industry from batch to continuous mode. Additionally, the reuse and recirculation of compounds integrated with high-throughput purification and well-defined analytical methods that can be applied in a continuous mode can make this process more sustainable and cost-effective. Nevertheless, this still requires the design of new unit operations that allow to recover unused raw materials, and that can be operated in continuous mode. Enzymes and the DNA template are costly raw materials, and its re-use should make the process more cost-effective. Nevertheless, separation of separation of these material from mRNA without losing activity can be challenging. Immobilisation of DNA and/or enzymes can be an alternative to achieve separation from mRNA without resorting to multiple purification steps. Additionally, it can have a positive impact on product-related impurities production.

In the end, the results obtained contribute to state-of-art of mRNA vaccines manufacturing and will contribute to the development sustainable, flexible and cost-effective manufacturing process, enabling this technology to be more affordable to all.

## References

- [1] Network TWERS. Estimated number of lives directly saved by COVID-19 vaccination programs in the WHO European Region, December 2020 to March 2023 2024:2024.01.12.24301206. <https://doi.org/10.1101/2024.01.12.24301206>.
- [2] Kumar A, Blum J, Thanh Le T, Havelange N, Magini D, Yoon I-K. The mRNA vaccine development landscape for infectious diseases. *Nature Reviews Drug Discovery* 2022;21:333–4. <https://doi.org/10.1038/d41573-022-00035-z>.
- [3] Rosa SS, Prazeres DMF, Azevedo AM, Marques MPC. mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine* 2021;39:2190–200. <https://doi.org/10.1016/j.vaccine.2021.03.038>.
- [4] Wang Y-S, Kumari M, Chen G-H, Hong M-H, Yuan JP-Y, Tsai J-L, et al. mRNA-based vaccines and therapeutics: an in-depth survey of current and upcoming clinical applications. *J Biomed Sci* 2023;30:1–35. <https://doi.org/10.1186/s12929-023-00977-5>.
- [5] Baronti L, Karlsson H, Marušič M, Petzold K. A guide to large-scale RNA sample preparation. *Anal Bioanal Chem* 2018;410:3239–52. <https://doi.org/10.1007/s00216-018-0943-8>.

- [6] Kariko K, Muramatsu H, Ludwig J, Weissman D. Generating the optimal mRNA for therapy: HPLC purification eliminates immune activation and improves translation of nucleoside-modified, protein-encoding mRNA. *Nucleic Acids Research* 2011;39:e142–e142. <https://doi.org/10.1093/nar/gkr695>.
- [7] Mishra RS, Kumar R, Dhingra S, Sengupta S, Sharma T, Gautam GD. Adaptive grey model (AGM) approach for judgemental forecasting in short-term manufacturing demand. *Materials Today: Proceedings* 2022. <https://doi.org/10.1016/j.matpr.2021.12.531>.
- [8] Rosa SS, Nunes D, Antunes L, Prazeres DMF, Marques MPC, Azevedo AM. Maximizing mRNA vaccine production with Bayesian optimization. *Biotechnology and Bioengineering* 2022;119:3127–39. <https://doi.org/10.1002/bit.28216>.
- [9] Sari Y, Sousa Rosa S, Jeffries J, Marques MPC. Comprehensive evaluation of T7 promoter for enhanced yield and quality in mRNA production. *Sci Rep* 2024;14:9655. <https://doi.org/10.1038/s41598-024-59978-5>.
- [10] Kis Z, Kontoravdi C, Shattock R, Shah N. Resources, Production Scales and Time Required for Producing RNA Vaccines for the Global Pandemic Demand. *Vaccines* 2021;9:3. <https://doi.org/10.3390/vaccines9010003>.