University of Glasgow

Cell-free protein systems and *in vitro* display methods as compelling tools for high-throughput screening

A thesis submitted to the University of Glasgow for the degree of **Doctor of Philosophy**

School of Engineering

College of Science and Engineering

Sabrina Lynn Galiñanes Reyes

October 2020

“Progress is made by trial and failure; the failures are generally a hundred times more numerous than the successes, yet they are usually left unchronicled. The reason is that the investigator feels that even though he has failed in achieving an expected result, some other more fortunate experimenter may succeed, and it is unwise to discourage his attempts.”

William Ramsay

From 'Radium and its Products', Harper’s Magazine (Dec 1904), 110, No. 655, 52.

Table of Contents

Acknowledgements

The research carried to write this thesis started in October 2016 and ended in April 2020. My research allowed me to work in two countries, Scotland and Japan. Inside Japan, I had the pleasure to work in three different organizations: Earth-Life Science Institute (ELSI) (Tokyo Institute of Technology), University of Osaka, and Japan Agency for Marine-Earth Science and Technology (JAMSTEC). This amazing opportunity allowed me to meet incredible people with so many different backgrounds, beliefs, and points of views and coming from different parts of the world. This part of my PhD made me grow as a professional but also as a person. All of the people I met left a mark in me, some more than others, but every single interaction helped to shape me into the person I am today. To start with, I want to thank everyone that I have had the pleasure to meet during my life, because whether it was good or bad, they all taught me something, and it is that knowledge that I use every day to advance through life. As for the people that I want to thank more personally:

Dr Soichiro Tsuda, as I would not have had this opportunity without him. When I started my MSc with him back in 2014 I admired his hard work and novel ideas and projects in mind, which led me to start my PhD with him. It was thanks to him I had the amazing opportunity to do part of my research in Japan and during that, fulfilling one of my time long dreams of visiting such an amazing country. He was always a caring person that took care of me and taught me so many useful and

practical skills. Even after his departure, he did his best to support me. I wish you and your family all the best in the future.

Prof. Leroy Cronin, for giving Dr Tsuda and I some much-needed laboratory space to perform our experiments. Also for providing all his valuable feedback and guidance to help solve the problems we encountered. His impressive mind and ideas will always be a model for me to look to.

Dr Yutetsu Kuruma, for being my other guiding light during my PhD, especially when big changes came along and I felt lost. You were the one who taught me the main block of my laboratory skills and you were always there when I had questions and was stuck. I always enjoyed our science and non-science talks, especially when they happened over a glass of sake at an izakaya. You are a great person with a heart of gold, never change!

Sumie Eto and **Mai Fujimi**, for being the best laboratories buddies I could have ever had. Laboratory work is always hard, especially after so many hours, but you two made it much smoother. I loved our talks filled with laughter, going for sashimi lunch together and always having a snack on my table from both of you. Also thank you for being there with all my laboratory questions, even when I made mistakes. You are both wonderful people and I am so glad you are in my life. You will both do marvellous in your new path in life.

Dr Kosuke Fujishima, for allowing me to work with you and take part in another amazing project. I truthfully enjoyed collaborating with you and your team, you were always so nice and tactful that it was a pleasure discussing science and just talking in general. You always listened to what I had to say, even when I felt like an inexperienced student. You have great potential as a supervisor and a big heart. You are a marvellous scientist, father and husband, you have a wonderful family and I am completely sure **Sana-chan** will grow up to be as amazing as her father.

Dr Julien Reboud, for taking care of me and all the paperwork and problems I drag along. I was dropped onto you midway through the PhD, but you took me under your wings and helped me a great deal, especially with the writing thesis part. This shows your dedication to your students and that is one of the best skills a supervisor can have.

I want to also personally thank the Lord Kelvin Adam Smith (LKAS) PhD Scholarship for funding me through all the research. Also the ELSI Origins Network (EON) for giving me the opportunity at the beginning of performing part of my research in Japan.

In this section, I want to put an asterisk for a very big thanks. I want to thank the millions of people who helped during the COVID-19 pandemic crisis. To all the health and sanitary people who worked endlessly to treat patients and save lives,

even when it meant risking themselves. To all the scientists and researchers who pulled long working shifts to bring treatment and understanding of this unpredicted virus. To all the personnel that had to continue working and getting exposed to maintain necessities running and prevent societies from collapsing. You are all our saviours, and you deserve all our respect and be recognized as the amazing people you all are.

All of the work behind research and endless investigation could not be possible without the help and dedication of some people that make all the paperwork, answering endless questions and way too often mistakes get resolved in the blink of an eye. I want to thank all the personnel from the organizations I have been in. The University of Glasgow, especially **Rebekah Derrett**, for your help with the scholarships process and inquiries; to **Julia Deans** for all her patience with my non-ending questions; to the Finance Office for always making sure I get my money back as soon as possible; and to the Insurance Office, for covering me when I was abroad. To the staff from ELSI that helped so much with all the visa paperwork along with my accommodation, stipend and laboratory equipment, **Kyoko Akiyama**, **Reiko Nagano**, **Harumi Tanaka** and **Kanako Yagame**. Also my special gratitude to the supporting staff from the University of Osaka and JAMSTEC.

This journey would have not been possible without the constant support of all my close friends. Even if are all now scattered around the world, we continue to keep in contact and remember the great memories we made together.

To my Glasgow gang, **Brynley Pearlstone**, **Joe Hennell**, **Steph Woodings**, **Christopher Delaney**, **Daniel Rosner and Shijoy Mathew**; ever since we all met at the university student residence we have had a blast. All those meetings at our flats, the restaurant runs for delicious food, the super fun board games evenings and of course the annual Eurovision gathering. It all made for superb memories and really fun moments. I am very proud of what all of you have accomplished so far, so keep it up. Even if I haven't been able to see her for a very long time, I always carry **Maria Rosa Castro Gonzalez** in my thoughts and heart as she is always there, one call away, when I need to talk and have a great laugh. I had the time of my life during our Thailand trip and I am happy you got to tie the knot to a great man that adores you. To my best Italian buddy, **Lucia Gastoldi**. We have always found in each other great confidants to let go of things and say everything that was on our mind. A friendship that started with coffee breaks in Tokyo and that will last forever. I know I can count on you and we will always be there for each other. Life has amazing experiences reserved for such an amazing person as you. Never give up, I believe in you.

A very special thanks to my brother from another mother, **Vasilis Ntouros**, for being the best friend/brother/confident that anyone could ever ask for. You have one of the biggest and strongest hearts and souls I have ever seen in a person. You

are a magnificent guy and you deserve all the best life can give you. I am forever in debt to you from your kindness.

To my Japan princesses gang, **Irene Bonati**, **Tony Jia and Alexandre Baccouche**. You were always there to make sure we would enjoy the finest afternoon tea Tokyo had to offer and to propose super fun trips when we would play kiwi and eat mikans like there was no tomorrow. And while the pandemic lockdown was on place, to our bi-weekly zoom meeting full of laughter and online board games. Along with them, a big thanks to the other members of the Tokyo gang, **Aishwarya Paknikar**, **Chaitanya Giri and Richard Gilliams.** I will always cherish our late nights of izakayas and karaoke, and of course the strong friendship we developed over that year that continues and overcomes any distance. Continuing on the same country, I want to express my gratitude to **Hugo Villanúa Vega** for making me part of his life and all the help he provided me. Also for all the video gaming sessions we had and the not so successful Japanese studying sessions that turned to hours long chatting about everything filled with Simpson's references. Of course also to his partner in life, **Yang Zhang**, for making me laugh and feel relaxed with her adorable unexpected comments and actions and her sweet and adorable personality. I wish you both all the best in this new chapter in your life!

To my longest lasting friends from my motherland, **Enrique León Calvo** y **Dunia Escribano**. From the very beginning to the very end, through the good and the bad, through the long-lasting distance, you were always there. To listen to me, to give me good advice, and to make me laugh. You were my anchor when I was

homesick and constantly made sure I would wake up to hundreds of text messages of very funny memes to start my day with a smile. I love you both with all my heart.

For the big finale, it is time for the most important people of my existence. To my everything, **Maia Kokawa**. I never imagined I would find everything I ever dream of on the opposite side of the world, but life gives you such pleasant surprises once in a while. You are the definition of perfection in my eyes. You are caring, patient, loving and hardworking. You are adorable and beautiful in equal parts. We share so many things that every day by your side is wonderful. To many more years of restaurant hunting for the best food and travelling as far as we can go. Of course, none of this would have ever been possible without the unconditional love and support from my parents, **Evelyn Reyes Gonzalez** and **Jaime Galiñanes Mondelo.** You are the best parents any person in the entire universe could have ever asked for. We have had a difficult life, you more than me, but you never stopped supporting me and making me grow and advance as a person and professional. No matter how hard the situation was, you always had time for me and made sure to fill me with love and endless opportunities to me make my path in life. Everything I am today is because of you. I love you both more than words can express.

Publications

At the date of thesis submission, this paper has been submitted and reviewed in ACS Synthetic Biology and Biotechnology and Bioengineering and was resubmitted to Biotechnology and Bioengineering and accepted in January 2021. The results of this paper form the basis of chapter 4.

- **Reyes, SG**, Kuruma, Y, Fujimi, M, *et al*. PURE mRNA display and cDNA display provide rapid detection of core epitope motif via high-throughput sequencing. *Biotechnol Bioeng*. **118**: 1702-1715 (2021). https://doi.org/10.1002/bit.27696

At the date of thesis submission, this paper has been submitted and reviewed in ACS Synthetic Biology and is being submitted to PLOS One in the upcoming months after discussion with the corresponding author. The results of this paper form the basis of chapter 3.

- Uncovering cell-free protein expression dynamics by a promoter library with diverse strengths. **Sabrina Galiñanes Reyes**, Yutetsu Kuruma, Soichiro Tsuda. bioRxiv 214593; doi: https://doi.org/10.1101/214593

During the research carried out in this thesis, collaborative work was also performed and two publications resulted from this:

- Yoshida, M., Hinkley, T., Tsuda, S., Abul-Haija, Y. M., McBurney, R. T., Kulikov V., Mathieson, J. S., **Reyes, S. G**., Castro, M. D., Cronin, L. Using Evolutionary Algorithms and Machine Learning to Explore Sequence Space for the Discovery of Antimicrobial Peptides. *Chem*. **4**, 533-543 (2018). https://doi.org/10.1016/j.chempr.2018.01.005

- Yoshida, M., **Reyes, S.**, Tsuda, S. *et al.* Time-programmable drug dosing allows the manipulation, suppression and reversal of antibiotic drug resistance in vitro. *Nat Commun.* **8**, 15589 (2017). https://doi.org/10.1038/ncomms15589

Abstract

Synthetic biology has become a promising field that aims at developing and using tools to work with recombinant DNA. In this field, cell-free protein expression systems have become a valuable asset to enable the *in vitro* transcription-translation of recombinant proteins, as functional elements of synthetic biology. These systems are not dependent on a living organism and consequently offer full control of the reactions' composition and environment, thus enabling protein expression in situations where *in vivo* systems would not perform efficiently. In this work, we aimed to explore their applications for *in vitro* display techniques, for protein and peptide evolution in drug discovery. Cell-free systems have the potential to allow for a higher number of library candidates to be selected and to enable the use of recombinant or unnatural candidates. These unnatural candidates are elements not used by organisms naturally, whether because they are toxic, they don't have the metabolisms to process them or just because they are fully synthetic. The most researched targets for drugs are membrane proteins, but they are also some of the most challenging, as they require a proper lipid membrane to fold and settle correctly.

The work presented in this thesis is focused on linking cell-free systems, *in vitro* display methods and membrane proteins, by characterising the effects of specific components on performance in a systematic step-by-step manner. The thesis first describes the uncovering of the underlying dynamics of protein expression in two

different types of cell-free systems, namely cell-extracts and purified components. A set of T7 promoter variants was constructed and tested in both systems, and the protein expression levels recorded and analysed. Both systems are driven by different expression dynamics for protein and mRNA. These expression dynamics represent the behaviour of certain parameters involved in synthesis, regulation, degradation, bottlenecks, etc … The limiting factors of both systems were identified for optimization of protein expression. Following conclusions from this analysis, purified components for protein translation were adopted and applied to both mRNA and cDNA display techniques. The results demonstrated the ability of the cell-free systems to provide a screening/selection method producing highly stable peptide conjugates and high sample purification. This proof of concept was tested and verified with the FLAG epitope, as a thoroughly characterised system. Several motifs with high affinity were obtained after 4 rounds of selection and further sequenced. Building further on these developments, cell-free systems were used to produce CX3CR1, a membrane protein from the G-protein-coupled receptors (GPCRs) family, within two types of synthetic lipid membranes, liposomes and nanodiscs. The thesis finishes by providing potential directions for the possible use of the cell-free expression systems, mRNA display and GPCR proteins for the creation of a peptide screening and selection method that could be used in the future for drug screening of membrane proteins.

Author's Declaration

The research performed in this thesis has been a collaborative effort and enhances the interdisciplinary contribution to other fields. I would like to acknowledge the following people for their contribution:

Dr Soichiro Tsuda, for assisting in the research carried out in chapter 3 by aiding with the data visualization (through R and Python) and the paper writing. Also for helping and teaching about the modelling of both cell-free systems.

Dr Kosuke Fushijima, for running hand by hand the research carried out in Chapter 4 by aiding with sequence data analysis (through Python) and the paper writing. To Dr Masako Yamazaki for being involved in the experimental part of Chapter 3 for the screening and selection cycles and the sequencing sample preparation.

Dr Yutetsu Kuruma for providing help and guidance during chapter 5 by teaching the procedures of preparation and purification of liposomes and nanodiscs. Also for allowing the use of the radiolabelling methodology to be used in his laboratory.

Dr Julien Reboud for providing guidance during the whole PhD process, but more specifically, for the valuable feedback provided on the final written version of the thesis.

Abbreviations

BLM: black lipid membrane

cDNA: complementary Deoxyribonucleic acid

DFHBI : (5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2,3-dimethyl- 4H-Imidazol-4-one

DMPC: dimyristoylphosphatidylcholine

DNA: Deoxyribonucleic acid

dNTP : Deoxynucleoside triphosphate

E. coli: Escherichia coli

EDTA: Ethylenediaminetetraacetic acid

FACS: fluorescence-activated cell sorting

FSC: Forward Scattering

GFP: Green Fluorescent Protein

GPCR: G protein-coupled receptors

GSH: reduced glutathione

GUV: giant unilamellar vesicle

ICF: intracellular fluid

LB: Lysogeny Broth, Luria broth, Lennox broth, or Luria–Bertani medium

MACS: magnetic-activated cell sorting

MLV: multilamellar vesicles

MP: membrane proteins

mRNA: messenger Ribonucleic acid

MSP: Membrane Scaffold Protein

MSP: membrane scaffold proteins

NMR: Nuclear Magnetic Resonance

PCR: Polymerase Chain Reaction

PDVF: polyvinylidene fluoride

PNK: Polynucleotide Kinase

POPC: 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine

POPS: 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine

PURE: Protein synthesis Using Recombinant Elements

RBS: Ribosome Binding Site

RNA: Ribonucleic acid

RNAP: RNA Polymerase

RT: reverse transcription

sfGFP: Superfolder Green Fluorescent Protein

SLB: supported lipid bilayers

SSC: Side Scattering

ssDNA: single-stranded Deoxyribonucleic acid

SUVs: small unilamellar vesicles

TAE: Tris-Acetate-EDTA

TBE : Tris-borate-EDTA

t-BLM: Tethered bilayer lipid membranes

WT: Wild Type

Thesis Summary

This thesis explores the applications of cell-free protein expression systems in synthetic biology together with *in vitro* display methods and synthetic membranes for membrane protein production.

As a quick summary of the thesis chapters:

Chapter 1: Introduction. This section is a literature review of the background information necessary to understand and follow this thesis.

Chapter 2. Materials and Methodology. This chapter explains in detail the procedures and protocols carried out during the experimental part of this thesis.

Chapter 3. Cell-free protein expression systems and promoter library mutants: uncovering their dynamics and tackling optimization processes. This topic covers research done on two different cell-free systems in order to discover how they behave and how can they be optimized using data analysis and mathematical modelling.

Chapter 4. In vitro display techniques and cell-free systems: a refined and clean method for in vitro selection of drug targeted peptides. Explanation of a methodology that could be used for screening of peptide candidates for drug discovery using cell-free systems and mutant libraries.

Chapter 5: Production of Lipid Vesicles and Nanodiscs Synthetic Membranes and G-protein-coupled receptors (GPCRs) using the PURE Cell-free System. Expression of membrane proteins using cell-free system and two different types of synthetic membranes, liposomes and nanodiscs.

Chapter 6: Conclusions and Future work. Conclusions of the thesis and explanation of posterior work that could be carried out to continue the research performed.

Further explanation of the thesis aims and how all chapters come together will be done in section 1.6 Thesis Aims and Summary so as to provide reader first with a background knowledge of the research field of the thesis.

Chapter 1. General Introduction

1.1 Introduction Outline

Major research and understanding of several topics were necessary prior in order to carry out the work in this thesis. The introduction of this thesis will focus on all the theoretical work and investigation performed before the experimental part. It has been divided into several sections for better explanation.

The first section will focus on synthetic biology (Section 1.2); then, cell-free systems (Section 1.3); *in vitro* display techniques (Section 1.4); artificial lipid membranes and membrane proteins (Section 1.5); and to finalise, and outline of the thesis (Section 1.6).

Section 1.2 will start with a summary of synthetic biology origins and their characteristics (Section 1.2.1). Then, it will go deeper into the major uses for synthetic biology, like protein engineering (Section 1.2.2), biosensors (Section 1.2.3), and artificial cells (Section 1.2.4).

In Section 1.3, cell-free protein systems will be discussed. To start with, a quick explanation of their characteristics and origins (Section 1.3.1); then, it will continue to explore the two main types of cell-free systems (Section 1.3.2); and to finalise, a segment to explore their applications (Section 1.3.3).

Section 1.4 will explain in detail different types of display methods. The first part (Section 1.4.1) will talk about display methods, their properties and uses in drug

screening. Then several types will be explained: mRNA and cDNA display (Section 1.4.2), ribosome display (Section 1.4.3), phage display (Section 1.4.4), cell display (Section 1.4.5) and liposome display (Section 1.4.6).

In Section 1.5, artificial lipids membranes (Section 1.5.1) and GPCR membrane proteins will be the major focus (Section 1.5.2).

To conclude, Section 1.6, will be an outline of the thesis.

1.2 Synthetic Biology

1.2.1 Origins and Characteristics

In the last few years, there has been an increasing demand for research tools that offer a broad area to work with to solve very specific problems in short periods. In this context, synthetic biology has gained popularity for meeting those criteria, since it offers a great range of available options to work with recombinant DNA [1].

Synthetic biology involves redesigning organisms for different goals. It focuses on the basics of genetic engineering to create new biological parts and/or systems, or redesign aspects already present in nature to give organisms new abilities. For example, it has been possible to engineer organisms that produce biofuels [2]. As these features come very handy for multiple purposes, researchers and companies are harvesting this field to tackle many problems. Some areas where it is being used, for example, are agriculture [3], food industry [4], and medicine[5]. Furthermore, thanks to the advancements in genetic engineering and the lower price for DNA synthesis and sequencing, this field is rapidly growing. Also, standardisation, regulation, and registration of biological parts and genes have been a great asset [6]. This aspect allows anyone accesses to a collection of genetic parts that are the genetic building bricks (BioBricks) [7] (Figure 1.A) for the assembly of biological systems. The catalogue includes many different types of elements, such as backbones, promoters, terminators, reporters, and protein domains. Also, the Synthetic gene database (SGDB) gives access to artificially engineered genes [8].

Within synthetic biology, there are two traditional ways of building new biological systems [9,10]. The top-down approach has been the more common way to perform synthetic biology for years. This path seeks to insert new biological functions into unmodified organisms, using external elements (Figure 1.B). The newly added features range from inserting genes that proportionate new properties, to producing pathways that trigger a response from an external signal. Although it uses mechanisms similar to those in single gene manipulation, the overall process tends to be more complex. Integrating new functions into host cells without any side effects is challenging. Some of the mechanisms used to perform this are BioBricks [7], the Cas9-CRISPR system [11], or artificial DNA [12]. An excellent example of this approach can be seen in yeast cells producing the precursor to antimalarial drug [13]. This achievement alone serves a high purpose, as it brings the possibility for affordable mass production of the drug.

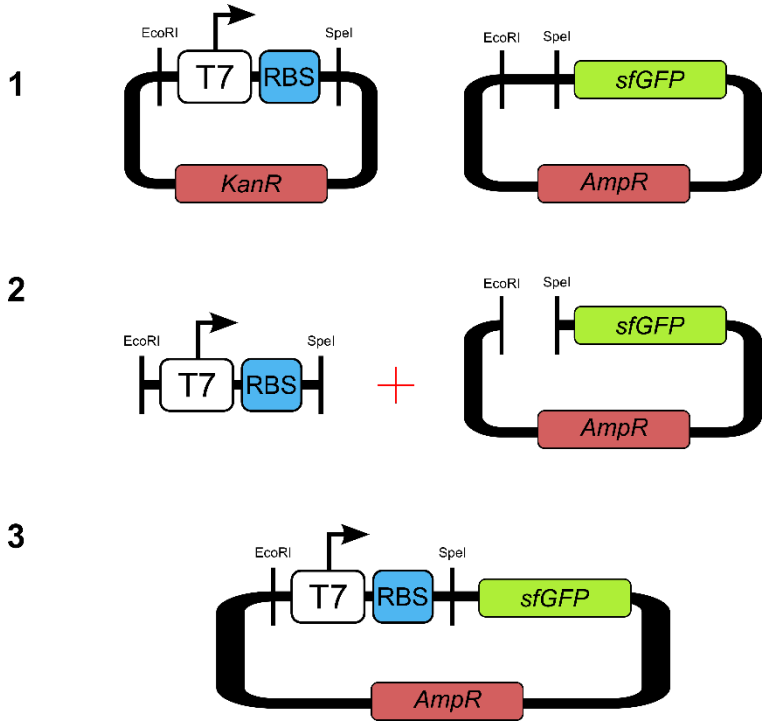


Figure 1.A: **Step by step on how to build a biological genetic circuit from BioBricks.** (1) The first step is to select the appropriate BioBricks from the catalogue and make sure their restriction sites match; (2) then both samples are cut with the same restriction enzymes; (3) and to finish, both cut samples are ligated together to get the final genetic circuit.

On the opposite side of this approach, is the bottom-up way. This method focuses on using individual parts ('non-living' components) to create full functional organisms or new biological systems, with specific desired properties [14] (Figure 1.B). The building blocks can be natural, modified, and/or synthetic in nature. One of the main ways this approach is used is to construct proto-cells or artificial cells [15] (Section 1.2.4). These cells try to mimic the structure and functions of real cells by encapsulating certain components in a synthetic membrane. Emphasis is being put

to achieve segregation within the cell, like organelles [16]. Although it is a good way to avoid the complexity of full organisms, like ensuring cell survival, this is also more challenging than the top-down approach, as the most appealing general objective involves creating life from scratch. A way to ease the difficulties this approach brings is to use extracts from lysed cells since they already contain almost all the necessary components for protein production. These are called a cell-free system (Section 1.3). The bottom-up approach is quite challenging and still relatively new, so most of the advancements done are still basic and in the early stages, such as DNA replication in an enclosed space [17]. Achievement of a fully synthetic cell includes the inclusion to sense and adapt to external and internal changes, and/or having the capability of full autonomous replication.

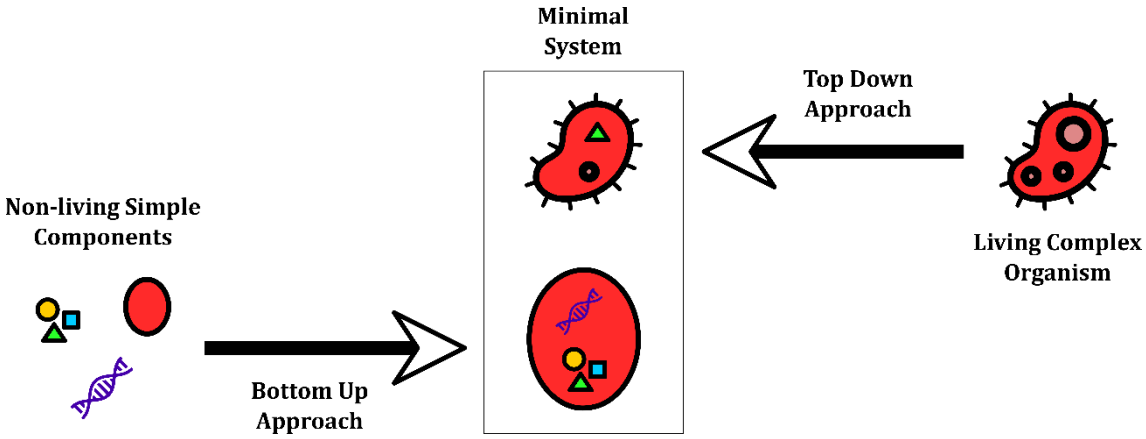


Figure 1.B. Synthetic biology approaches: top-down and bottom-up. Both pathways show possible courses to reach a minimum system. The top-down approach can eliminate non-essential genes from already existing organisms and/or add additional elements. The bottom-down pathway uses simple non-living components together to make them function in sync.

For both approaches, a clear challenge is knowing which minimal components are necessary for basic cell functions (cell division, energy regeneration, sensing and adapting, etc ...). Making all these pathways work together in a non-natural *in vitro* environment can be challenging. As already mentioned, the majority of work that is being carried out uses the top-down approach *in vivo*. This brings some limitations, being the most problematic one, interferences between the host system and the introduced one. Other disadvantages include long production periods, weak environment monitoring, or laborious sample uptake. These constraints are to be expected when working with living organisms since noxious metabolites in the cell, unbalanced growth or low product yields are common factors.

1.2.2 Synthetic Biology Applications: Protein Engineering

Synthetic biology can build novel biomolecular elements, systems, and pathways, to modify organisms. *De novo* engineering of synthetic circuits is a promising way to tackle difficulties in industry and academic research [18].

The more widespread application of synthetic biology is synthesis and production of products; both *in vivo* and *in vitro*. For *in vivo* production, the traditional approach involves the engineering of an organism by introducing foreign genes and posteriorly optimizing the production by directed evolution. For this case, the most used organisms are yeast (*Saccharomyces* being the dominant genus) and bacteria (mostly, *Escherichia coli* (*E. coli*)) are the preferred targets to work with for commercial production. Although the main industry objective is product

production, this is not always the case, as sometimes plants can be engineered directly, for example to improve their nutritional value, add new components for a better diet [19]; or just to withstand harsh environments [20].

Protein engineering, or modification, is another application of synthetic biology. Natural proteins can be modified; for example, by performing changes to their DNA sequence and recording protein efficiency and function. These modifications can be specific or random, like for example directed evolution, which tries to perform natural selection on the proteins to make them evolve into a specific path for a specific goal. The changes performed will create a mutant protein library in which candidates can be selected through screening methods (Figure 1.C). But also just *de novo* proteins can be created for new, mimicked, or better functions [21,22]. This is done mostly in the industry field, as companies look to produce industrial enzymes with high activity, to optimise the yield and make them more effective. These new versions of enzymes can help in making specific products, like lactose-free milk [23], or just lower the prices of the final product [24].

But the industry is not the only one taking advantage of recombinant proteins. Microreactors can be engineered to produce components in specific places at any given time. This can be done without the need for complex equipment and using simple elements. NASA recently started to research the prospect of astronauts producing resources where they are located to eliminate the need for shuttle delivery from Earth [25]. On Mars, in particular, synthetic biology could lead to production processes based on local resources, making it a powerful tool in the

development of manned outposts with less dependence on Earth. Work has gone into developing plant strains that can cope with the harsh Martian environment, using similar techniques to those employed to increase resilience to certain environmental factors in crops [26].

When it comes to creating a new protein function or modifying existing ones, the process can be laborious and time-consuming. For this, computational techniques have started to be used, such as bioinformatics, mathematical models, databases and algorithms (Figure 1.C). When performed properly, the results can yield enzymes having an exponential increase in non-natural product specificity [27].

Modifications can also be done to expand the natural elements used by organisms. For example, generally, only 20 amino acids are coded genetically to be used for protein production, but by modifying the genetic code, alternative unnatural amino acids can be present and used (like L-2-Aminobutyric acid) [28,29]. To do this, tRNAs and tRNA synthetases, from other organisms, are engineered into the host cell. Just as adding more amino acids to the genetic code is possible, the opposite also happens. By limiting the number or types of amino acids used by a certain organism, information can be gathered from certain proteins and enzymes and their functionality [30].

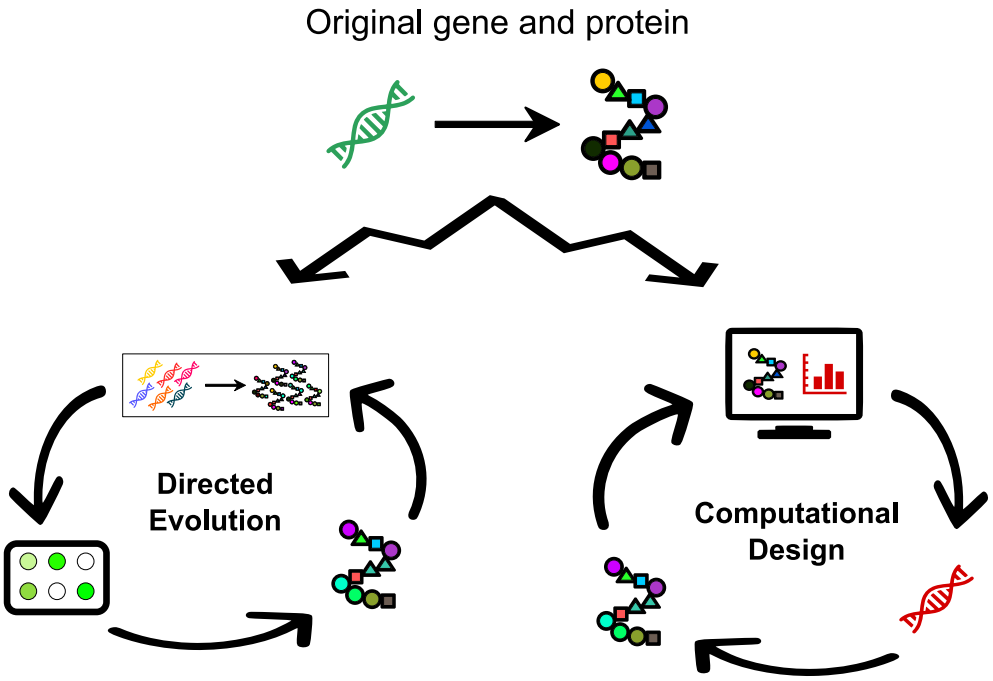


Figure 1.C: **Two different approaches to protein engineering**. Parting from an original gene and its protein, two different paths can be taken for the engineering of new features: (Left) Directed Evolution, in which a DNA library is made and screening of each protein is done individually to select a potential candidate; (Right) Computational Design, uses several computational tools to search in extensive DNA and protein databases and predicts a potential candidate.

Just as proteins can be designed, so can nucleic acids. It is possible to insert information into the sequence of a synthetic DNA molecule [31]. Usually, the use of certain algorithms enables the design of these types of genetic systems. As it happens with unnatural amino acids, the same is possible with unnatural nucleic acids [32]. This is usually done by genetically modifying an organism to allow the absorption and integration of said nucleic acids.

1.2.3 Synthetic Biology Applications: Biosensors

A sensor is a device that detects events or changes in its environment and sends the information to other devices to process it. Sensors are used frequently everywhere; especially in biological organisms as they have a multitude of regulatory circuits for sensing and responding to environmental changes. By taking advantage of these characteristics, biosensors can be engineered. A biosensor is an engineered organism that can sense a certain phenomenon in their surroundings and respond accordingly. A basic example of this is the capability of the bacteria *Aliivibrio fischeri* to detect and regulate the production of the luciferin enzyme, which produces light (bioluminescence). A more advanced example is an engineered *E.coli* bacteria that can detect TNT (trinitrotoluene) and DNT (dinitro), and posteriorly produce GFP (green fluorescence protein) [33]. When it comes to engineering organisms as biosensors, it can be done at the transcriptional, translational or post-translational level (Figure 1.D; A). The first stage in gene expression is the transcription, and all the elements involved in it can serve as biosensors. The main focus is on the promoters and their associated factors. This is mostly because there is already a great number of known promoters that respond to environmental changes, like for example, the *lac* operon. When done at the translational level, RNA molecules are the main target, especially non-coding RNA (Figure 1.D; B). These have a wide variety of important functions in the cell; like editing or cutting other RNA molecules or even regulating gene expression directly [34]. Since their ability to regulate depends on their particular sequence, they are

highly flexible and can have a broad range of targets, therefore allowing for fine-tuning of the biosensor system. Post-translational biosensing (Figure 1.D; C), is characterised by its high diversity and complexity. It depends on a high number of factors and interacting components, such as the connection between molecules or the ramifications of protein cascades. In this case, the main element to be used is protein receptors. While promoters and RNA molecules are mainly selected from naturally found ones, protein receptors can be designed *de novo* [18,33].

When these genetic circuits are present in more intricate patterns, they can be used to build biological computers [35]. These encircle a variety of genetic pathways with logical gates (e.g. AND, NOT, OR) that perform operations similar to those that computers do. Each pathway usually produces a molecule that induces or represses another pathway (activator, repressor sites), therefore creating an intricate web of systems interacting and working together. Studies have used this approach to make a proof-of-concept therapy that can detect and kill cancer cells [36].

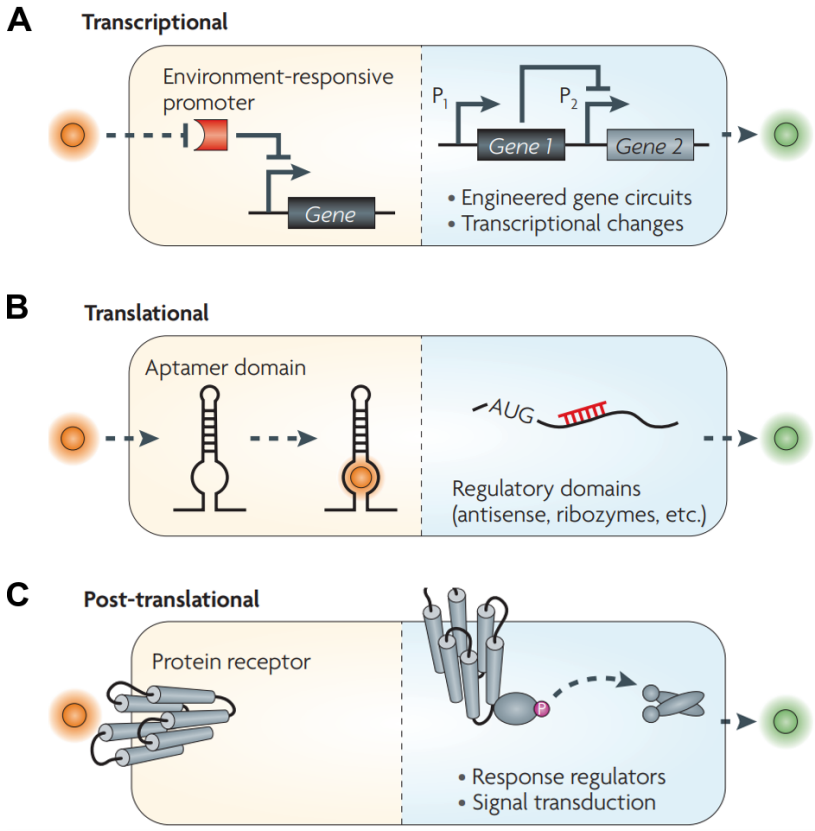


Figure 1.D: **Types of synthetic biosensors**. (A) Transcriptional biosensors mainly focus on environment responsive promoters linked to genetic circuits. (B) Translational biosensors are typically built by linking RNA aptamers to regulatory domains. (C) Post-translational biosensors are done by using protein receptors (usually in the membrane) that trigger signal cascades. Image used is from Khalil *et al*. (Nature Reviews, 2010) [18].

1.2.4 Synthetic Biology Applications: Artificial Life and Artificial Cells

Another use of synthetic biology is the possibility to create artificial life. This is based on the *in vitro* creation of organisms from non-living biomolecules. The two pathways of synthetic life are: understanding the origins of life and what makes something "alive" [37] and to use them as engineered cells, for drug delivery [38], as bioreactors [39], pollution treatment organisms, etc [40]. For the definition of what makes something "alive", some characteristics have to be all present in an

organism: 1. Presence of metabolic reactions that allow for all the other characteristics listed here to perform, 2. Organization and coordination of all the parts that constitute the organism, 3. Ability to perform regulated growth and therefore reproduction to create new organism, 4. The capability to sense and respond to signals and changes in the environment, and 5. Being able to undergo evolution.

When it comes to creating synthetic cells that are "alive", being able to sustain their metabolism on their own is essential. This can be done by up-taking elements found in their environment; creating ion gradients and recycling energy molecules. Additionally, the ability to store genetic information for all these processes is necessary along with the capability to mutate and reproduce. But so far this has not been possible [41]. The closest anyone came to this, was by creating a full synthetic bacterial genome and introducing it to a bacterial host with no genome [42]. This experiment was aimed to reveal the minimal genes necessary for life. The resulting bacteria cells came were created by inserting and deleting genes until a viable genome was obtained. The bacteria was called *Mycoplasma mycoides* JCVI-syn3.0 (Syn3.0), it had a total of 473 genes (Figure 1.E) [43] and was able to grow and reproduce itself (for comparison, *E. coli* has 4,500 genes). One peculiar aspect of this experiment is that only 80% of the original genes (from the donor) were essential. Apart from helping understand the minimal requirements for "life", Syn3.0 became very practical for laboratory use since its small genome made it possible to grow and reproduce much faster than many laboratory strains. Due to

its reduced genome, the time and nutrients the bacteria use to duplicate, and therefore reproduce, is lesser than the average bacterial strains, making it save time and resources when working with it. But most important, the creation of Syn3.0 helped in propelling the research of artificial cells, along with its development and customisation. It showed that even the simplest of cells can still be quite complex in nature and that the current reach of synthetic biology is not enough yet and needs to be improved.

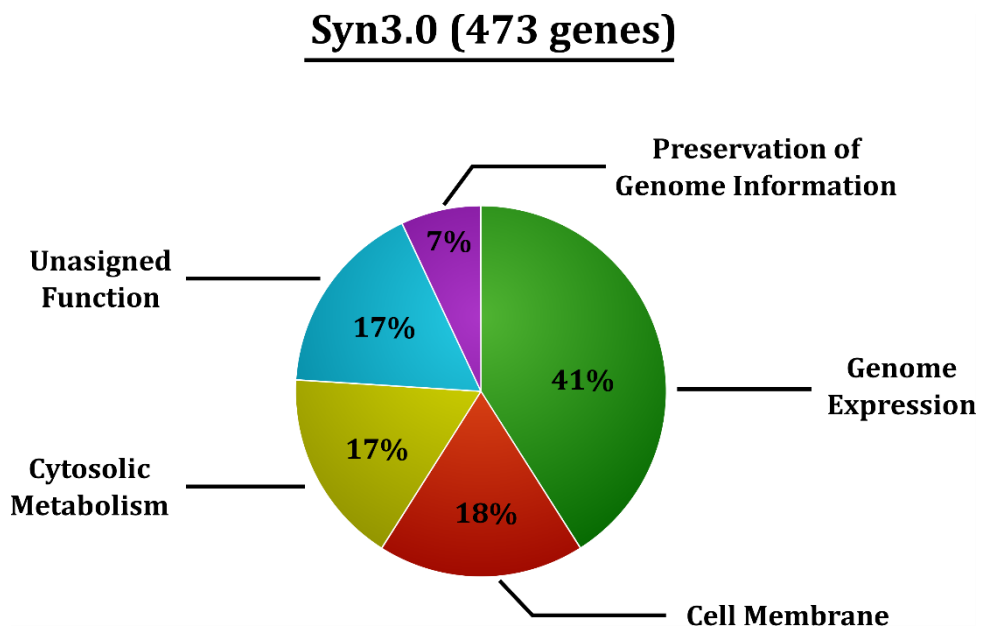


Figure 1.E: **Pie chart of the different types of genes that encompass the first artificial species, *Mycoplasma laboratorium*.** The organism, also called Syn3.0, had a total of 473 genes. The type of gene distribution is shown in the chart. It is important to note that a significant amount of genes are of unknown function.

Another close call in creating artificial life was done when scientists managed to create a completely artificial genome for *E. coli*, in which they reduced the natural number of codons from 64 to 59 [44]. In most cases, when it comes to synthesising an artificial organism, *Mycoplasma genitalium* (the original organism used for Syn3.0) or *Escherichia coli* with reduced genes, are used as starting points. These bacteria already have reduced genomes and can be easily cultivated in laboratory conditions.

So far all the examples explained were performed *in vivo*, using organisms as backbones for the synthetic genomes, so in other words, a top-down approach. When it comes to the bottom-up path, building a *de novo* artificial cell involves using non-living elements. The most basic cell model would involve: a phospholipid bilayer vesicle (lipid membrane), encapsulated synthetic DNA, and many biological systems to self-maintain and reproduce (Figure 1.F). When it comes to self-maintenance, active transport of energy molecules, cofactors and building blocks (nucleotides and amino acids) through the membrane from outside to the inside is necessary [45–47]. The main challenge encountered has been the synthesis of a minimal genome that includes all necessary genes for the production of the components involved in the protocell's development. Although it has not been possible yet, slowly these encountered issues are being mitigated [48,49]. One way to discover the minimum requirements for a protocell is to try and recreate the conditions believed to be present when the first cells arose. By re-enacting prebiotic conditions, synthetic cells can be formed "naturally" on their

own, for example, by allowing RNA polymers to be encapsulated by lipid vesicles through chemical reactions [50].

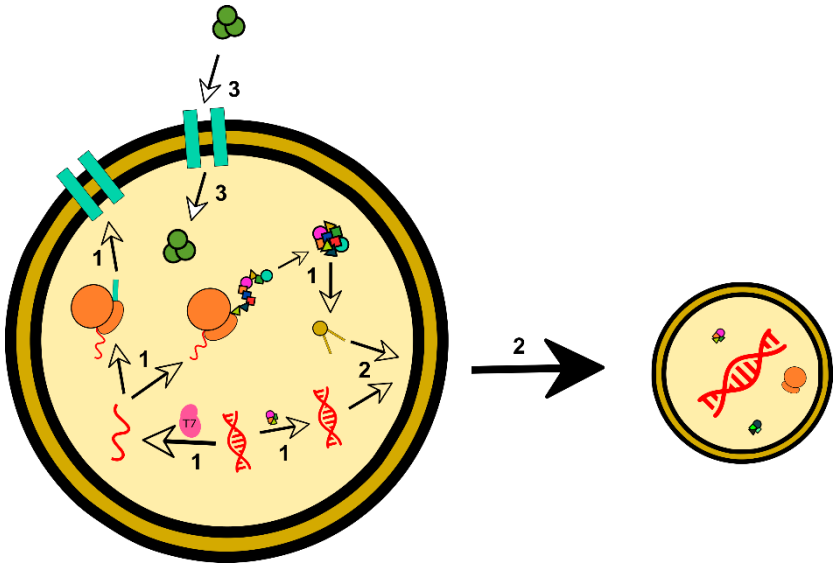


Figure 1.F: **Scheme of the minimum requirements of an artificial cell.** The protocell, needs of a lipid membrane (dark yellow) encapsulating a cytosol (light yellow) and all the necessary elements. The cell needs to able to sustain a metabolism (1) of DNA replication; RNA, enzyme and metabolites production and their transportation. Also, a system of DNA encapsulation through lipid membrane growth (2) for cell division. Finally, an active transportation system (3) permits the movement of all crucial components inside the cell.

Since artificial cells are minimal versions of organisms, it is quite possible to engineer them to eliminate all the drawbacks of using living organisms. For example, they can be used as bioreactors, biosensors and drug transporters at the same time. Since they are of biological origin, rejection risks if used on humans are reduced, as well as the production costs and time. Some studies have managed to engineer bacteria that, upon the detection of cancer cells (with protein to protein interaction) or the detection of a cancerous environment (for example, hypoxia), produce therapeutic molecules [51–53]. But since the treatment of diseases with living

organisms is not recommended for fear of infection and off-target side effects; researchers are trying to work more with more bottom-up approaches for artificial cells. Another way to avoid using foreign organisms is to engineer directly cells from the immune system (for example, T-cells) to make them more receptive to tumours [54].

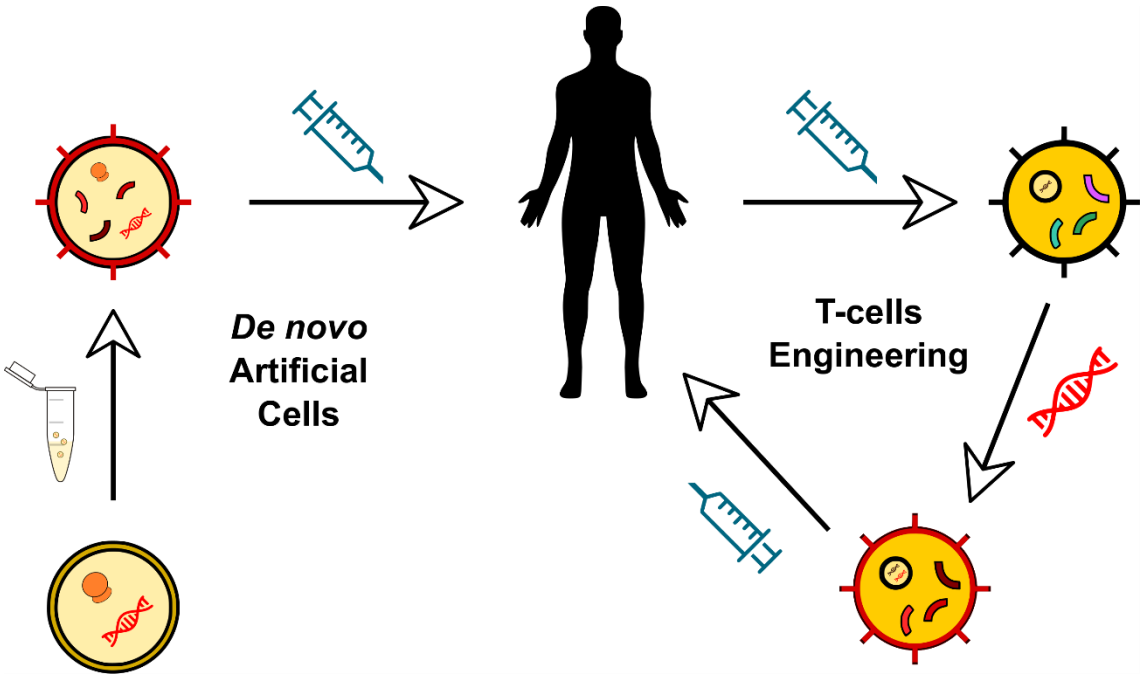


Figure 1.G: **The two different ways artificial cells can be used for therapy.** The first way (left) is to create *de novo* protocells with a piece of transcription-translation machinery and genetic code to create the appropriate proteins and enzymes for the detection and elimination of the target cell. The second method (right) is the removal, from the patient, of cells from the immune system (like T-cells) and engineer them to enhance their functionality. Whether they are *de novo* or engineered, both types of cells are applied to the patient for illness treatment.

1.3 Cell-Free Protein Systems

1.3.1 Properties and Characteristics

Cell-free (CF) systems, also called transcription-translation (TX-TL) systems, are a type of *in vitro* mechanism. They are used to research and recreate a biological system that happens inside cells, but without using living organisms (Figure 1.H). They have been increasingly popular for synthetic biology and metabolic engineering applications in recent years [55–57]. They bring certain advantages to reduce or eliminate issues that happen when working with complex living organisms. They prevail especially in the most influential fields of protein evolution and drug delivery therapy [58–64]. Some of the benefits of CF synthetic biology are the decreased amount of total experimental time [65] since it does not require to grow and maintain live organisms. It also reduces the efforts on analysis and reaction optimisation, mainly as the reaction can be easily modified and monitored. Other positive aspects of CF systems are the ability to avoid the production of by-products or the presence of unwanted cell mass [66]. One of the most appealing advantages is the ability to incorporate reactions that are not possible with living organisms, like, production of toxic compounds [67] or reactions in extreme conditions [68], which are conditions (temperature, pH, etc ...) which fall outside the range in which an organism can survive naturally.

These systems often produce a substantial yield of protein products; however, some issues remain. One of them is the lack of an energy regeneration system. This leads to the use of expensive phosphoenolpyruvate (PEP) as a supplement for the

reaction. Although PEP improves the ATP regeneration issue, the final costs of the reactions increase significantly. Due to this, a good amount of research is focusing on finding more affordable energy sources [69,70], such as glucose and glucose-6-phosphate, while avoiding the decrease in protein yield. Even though several studies managed to achieve this [71], and decrease even further the cell-free reaction costs [72], PEP remains the most extensively used energy source. Apart from that, the quality of several reaction components is an issue. This happens due to their origin, mostly from prokaryote cell extracts. But this is also been researched and improved greatly with the use of recombinant purified components [59]. While ongoing research continues to mitigate these issues and some others [73,74], cell-free systems using recombinant protein elements [59,75,76], offer a contaminant-free alternative with a final significant protein yield [77].

One of the main and most important components of CF systems are the proteins and enzymes that carry out most reactions. They can be obtained through different types of methods, such as cell lysis to obtain cell extracts; or from purified components protein expression. All these ways to obtain a CF system provide a wide assortment of experimental diversity and high reaction customisation. Also, CF reactions can be performed with basic affordable laboratory equipment. Whether it comes from cell extracts or purified components, obtaining the main proportion of enzymes and proteins takes up a high amount of time and cost, so this is another aspect being improved in CF systems [78].

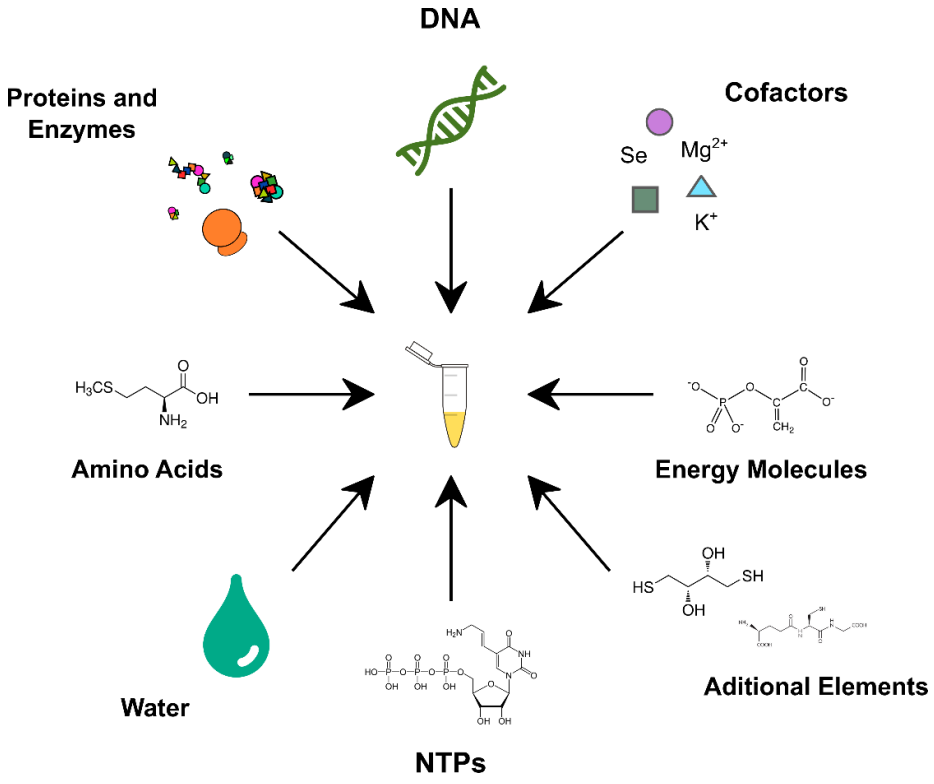


Figure 1.H: **Scheme showing the necessary components for a cell-free expression system.** There are three main factors for a CF reaction. A cell extract (or purified components) that contains the machinery necessary for TX-TL. A genetic component for the production of the desired product. And a reaction buffer containing cofactors, amino acids, salts, etc so the system can work properly and the enzymes have an energy source.

1.3.2 Types of Cell-Free Systems: Cell-extract vs Purified Components

The first CF systems were obtained through cell extract lysates. These lysates come from a living organism, such as bacteria, yeast, fall armyworm, wheat germ, tobacco, rabbit reticulocytes or HeLa cell line [79–85]. These lysates are taken from their intracellular fluid (ICF), which contains water, ions and macromolecules. It is mainly the TX-TL molecules that are necessary, like RNA polymerase, ribosomes,

tRNA synthetases, etc. All the other necessary components are usually added to the lysate, such as amino acids, nucleotides and buffers (Figure 1.H). These lysates, or subcellular fractions, can be obtained through several cell disruption methods (Figure 1.I) [86], like mechanical grinding [82], ultrasonic waves [87], high pressure[88], thermal lysis [89], biochemical rupture [90], enzymatic disruption [91] etc.

Some methods are dependent on what type of elements are being isolated, if they are temperature sensitive, thermal lysis is not adequate. Others might be used for very specific organisms, such as the enzymatic disruption using the chitinase enzyme for yeast cells. And some even use a mixture of several methods to overcome certain problems, such as the lower cell rupture efficiency by chemical methods, in which case, it is combined with a mechanical way [92].

The CF systems that use cell lysates are more affected by certain problems, like degradation of DNA, RNA and proteins due to nucleases and proteases found in the extracts. Degradation is a controlled way used by organisms to maintain a balance, but outside the host, there is no control of this process and these enzymes degenerate their targets without any regulation. Because of these issues, some of the most modern CF systems do not use cell lysates, instead, they use purified components [75].

Method	Sub-method	Technique	Advantages	Disadvantages
Mechanical	–	**High pressure** **Beads / Milling**	- Highly Efficient - Non-specific	- Generated heat causes product damage - Expensive - Difficult purification
Non-mechanical	**Physical**	**Thermal lysis**	- Easy - Non-specific	- Expensive - Generated heat causes product damage
		Osmotic shock	- Good for sensitive products	- Specific
		Cavitation	- Low energy and temperature required - Non-specific - Large scale possible	- Posible damage to sensitive product - Expensive - Difficult purification
	Chemical	**Alkaline lysis**	- Good for sensitive products - Non-specific	- Slow
		Detergent lysis	- Good release of proteins	- Damage to sensitive products - Expensive - Difficult purification - Low efficiency
	Biological	**Enzymatic lysis**	- Cell type specificity - Protein stability	- Low efficiency - Expensive - Need of detergents for bacteria lysis

Figure 1.I: Detailed overview of some types of cell lysis methods. The "Methods" column makes a distinction between techniques that rely or not on mechanical disruption. The "Sub-method" column provides additional details whether the technique is physical, chemical or biological. The "Technique" column gives some examples of each category, and the "Advantages" (green) and "Disadvantages" (red) column explains some of the techniques' benefits and restrains respectively.

E. coli - based systems are the most well-established systems due to the high yield for protein expression. But they are still being actively improved to simplify the

preparation methods [93–97] and obtain even better yield [69,98]. Protein synthesis Using Recombinant Elements (PURE) system is another type of cell-free system [59,99]. In contrast to the extract-based cell-free systems, purified components were reconstituted for protein synthesis (Figure 1.J).

The PURE system [59,100] is made of several components that are used for transcription (e.g., MetRS), translation (e.g., IF3 factor), aminoacylation (e.g., amino acyl-tRNA synthetase), and energy regeneration (e.g., creatine phosphate). All factors are purified individually to a high degree and assembled with substrates in a buffer ready to use. Since all the components are produced and purified individually, only the ones involved in the transcription/translation system are included. This removes from the system unwanted components like proteases and nucleases. Furthermore, the system allows for adjustments to the composition of each reaction individually. PURE*frex* 1.0 is the commercial version of the PURE system. Lately, more research has focused on simplification, robustness and low-cost for reconstruction [101] as well as alternative energy sources [102] in the PURE system. As the concentrations of all the components in the PURE system are known, it is suitable for systematic studies of cell-free protein expression, such as optimisation of component concentrations for better yield [103].

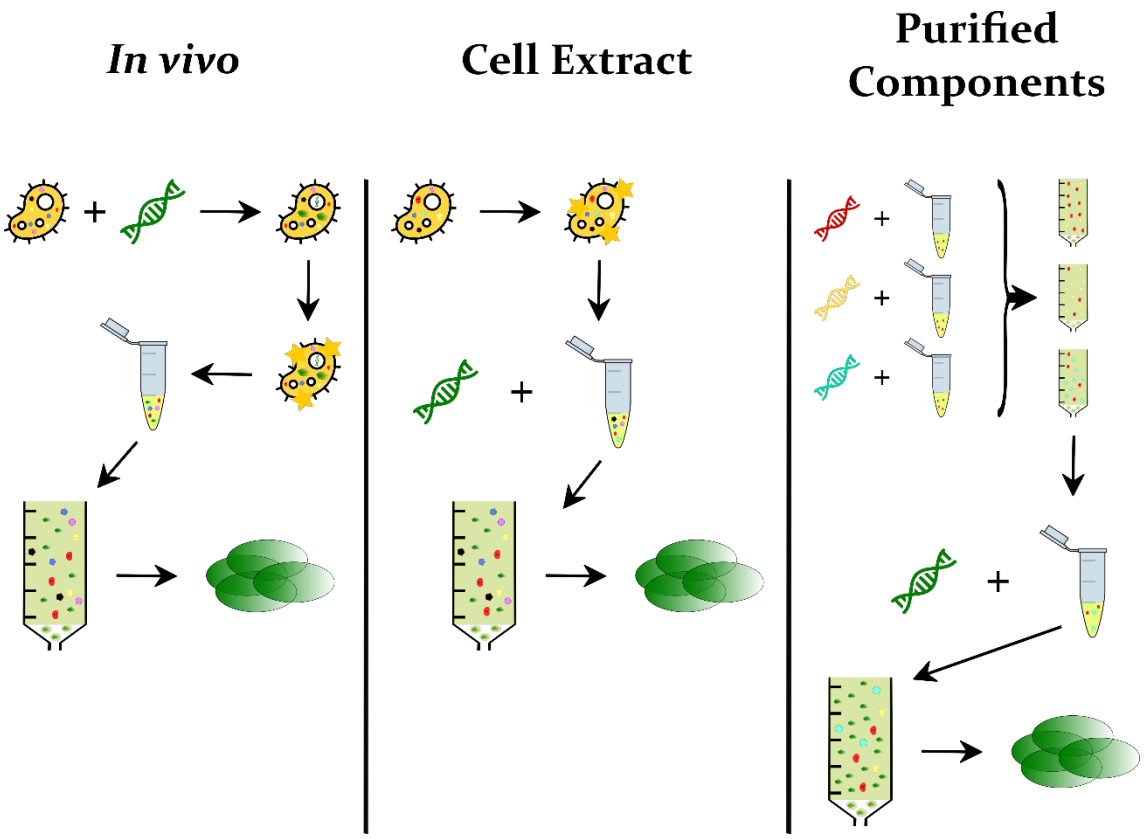


Figure 1.J: **Comparison of three different ways for protein expression.** The left column shows the *in vivo* way, in which organisms are transformed with a DNA molecule so it produces the desired protein, after the product can be obtained by cell lysis (optional if external protein) and purified. The middle column shows a cell extract-based cell-free system that takes advantage of a cell cytosol machinery to produce the desired product in vitro. The right column is also a cell-free system but based on purified components, that will also be used in vitro to obtain the desired product.

A recent computational study pointed out that there are more than 240 components and nearly 1000 reactions involved in the translation process of the PURE system [104]. The complex cell-free expression dynamics can be an issue especially when multiple proteins are being expressed, such as the reconstitution of the Sec translocon [105] or ATP synthase [99]. The correct balance of each synthesised proteins needs to be adjusted for the whole complex to be functional. Although

this is usually done by titrating the amount of each DNA input, protein expression levels do not linearly correlate with the amount of DNA [106,107], especially when a strong promoter is used [108]. Because of this nonlinear nature of cell-free protein expression, finding an optimal balance for multiple proteins can be a daunting challenge.

1.3.3 Cell-Free Protein Systems Applications

Using cell-free systems quickly gained popularity as it tackled constraints encountered when using living organisms, like cellular toxicity or unnatural environment conditions [67,109,110]. They are also starting to be used as a more economical way to synthesise biological elements instead of the traditional microbial fermentation [111]. They have a wide range of applications (Figure 1.K) from the rapid characterisation of genetic constructs (typically linear DNA constructs) [112] to protein expression in microcompartments for *in vitro* evolution [113,114].

The most straightforward use for CF systems is protein synthesis. *In vitro* reactions can very easily be controlled and monitored since they are not enclosed by membranes. From the production of proteins with tagged amino acids [115] to establishing a stable protein production system with a continuous flow of resources and removal of products [116]. Apart from these examples, fusion proteins made by CF systems are being an asset in the medicinal field to use as possible vaccines [117].

Another use for CF systems is real-time monitoring and alteration of metabolic pathways and their components. By using certain enzymes in an *in vitro* environment, it is possible to track the synthesis of products, the depletion of resources and each elements concentration at any given moment. Furthermore, the reaction can be altered anytime by adding, removing or altering the components. This can be very useful to optimise metabolic pathways and product synthesis [118,119].

As it was previously mentioned (Section 1.2.2) the incorporation of certain elements, like non-natural amino acids, is one of the uses for synthetic biology, but when working inside a cell it can be challenging. CF systems can be used when the amino acid metabolism is lacking or has been altered inside an organism. This way amino acids can be labelled for protein characterisation using NRM spectroscopy [120].

Directed evolution is another way CF systems come in handy. This method has been used for high throughput engineering, screening and selection of proteins. Using CF systems in this stance can be useful when dealing with difficult proteins (multimeric proteins, membrane proteins, and proteins with expanded amino acids) [121]. One of the main advantages of using both cell-free systems and directed evolution are that it allows for the use of large libraries (up to 10^{13} variants to test) since there are limitations of transformation efficiency when using living organisms (around 10^{9} variants). It also allows to control the conditions of the reaction. When it comes to membrane proteins, this combination, along with lipid

vesicles, can be really useful as membrane proteins are quite difficult to characterise and screen since they are dependent on a membrane [122].

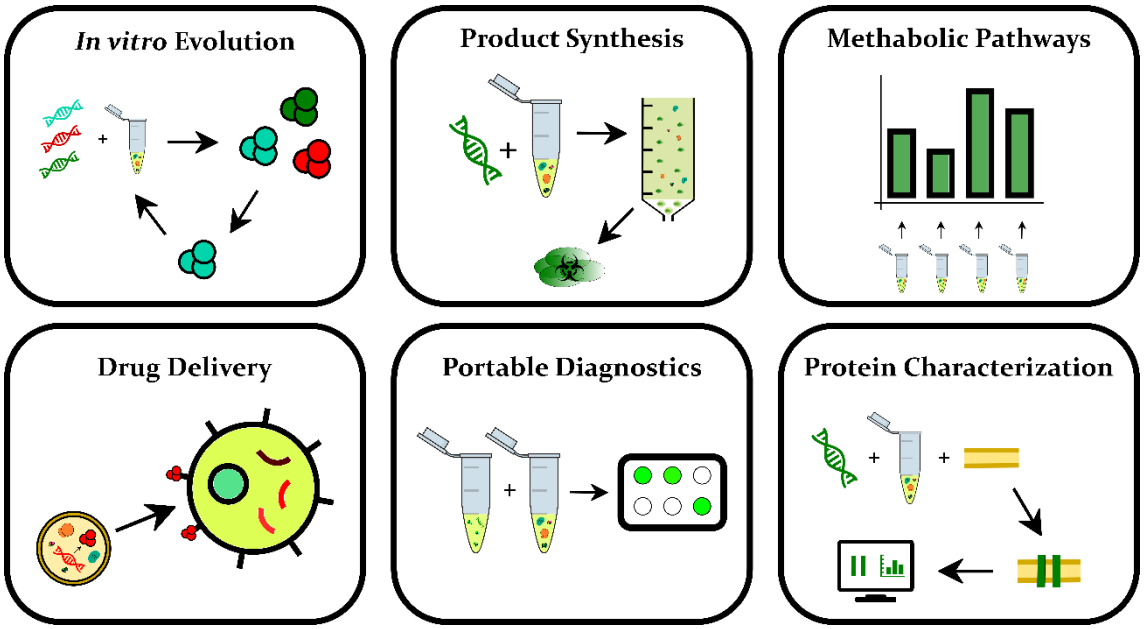


Figure 1.K: **Applications of cell-free protein system.** Cell-free systems have a wide variety of uses. (Top-left) *in vitro* protein evolution, (top-middle) protein synthesis for industry and research, (top-right) real-time monitoring and alteration for optimisation of metabolic pathways, (bottom-left) *in situ* drug delivery when encapsulated in lipid vesicles, (bottom-middle) as portable methods for illness diagnosis and, (bottom-right) for protein characterisation of membrane proteins.

A major characteristic of cell-free systems is the lack of engineered organisms and all the difficulties working with them brings. As a result, these systems have started to be quite promising when it comes to performing experimental research and diagnosis in remote locations where delivery, manufacturing or handling of laboratory products is impossible. Still, some issues remain with reagents' stability, durability and safety, requiring them to be in perpetual cold storage. Some studies

have been focusing on making a cell-free system more durable and stable at elevated temperatures [123,124] so they could be used as portable diagnosis tools.

Encapsulating CF systems inside lipid vesicles allows the creation of microreactors [125,126]. This method is being used frequently to study the origin of life since these artificial protocells could show how the first cells on Earth appeared [37]. Also, these microreactors are useful as vehicles for specific targeted drug delivery [127–130]. These liposome delivery systems are quite "trendy" as they overcome several problems found in delivery systems such as the possibility to deliver both hydrophobic and hydrophilic drugs and the insulation it provides to the encapsulated drug. They can prolong the drug's lifetime and decreases the probability of degradation [131]. Although liposomes seem like very suitable tools when delivering drugs, they still have their downsides like short half-life or premature rupture [131]. To overcome these difficulties, studies have focused on making hybrid vesicles with lipids and polymers to improve their stability [132].

1.4 *In vitro* and *in vivo* Display Techniques

1.4.1 Properties and Drug Screening

There has always been a high constant demand for the discovery and production of antibodies. One of the most established methods for antibody selection is hybridomas. This method produces a large number of antibodies by injecting mammals (usual mice) with a certain antigen that starts an immune response. Antibodies, produced by B lymphocytes, are then harvested from the animal and fused with B cell cancer cells. This new cell line, called hybridoma, can produce the monoclonal antibodies and be grown in the laboratory. But as it can be expected, this method is time-consuming, needs high resources, and is a disadvantage when it comes to animal ethics.

A different approach to antibody screening and production are *in vitro* display techniques. This technology encompasses a multitude of different methods to screen, select and optimise many antibodies from large libraries. The core of the technique is based on the fact that for each different antibody (phenotype) there is a specific genetic sequence (genotype). So by testing the capabilities of one specific antibody, its genetic sequence can be known and even evolved to enhance its function. When compared to the conventional techniques (hybridomas), display techniques stand out in their ability to test new (*de novo*) sequences, which immediately brings a much wider diversity of candidates to select from [133]. Furthermore, *in vitro* techniques are not restrained by the in *vivo* experimental

constraints, and are faster and more economical than working with animals (Figure 1.L).

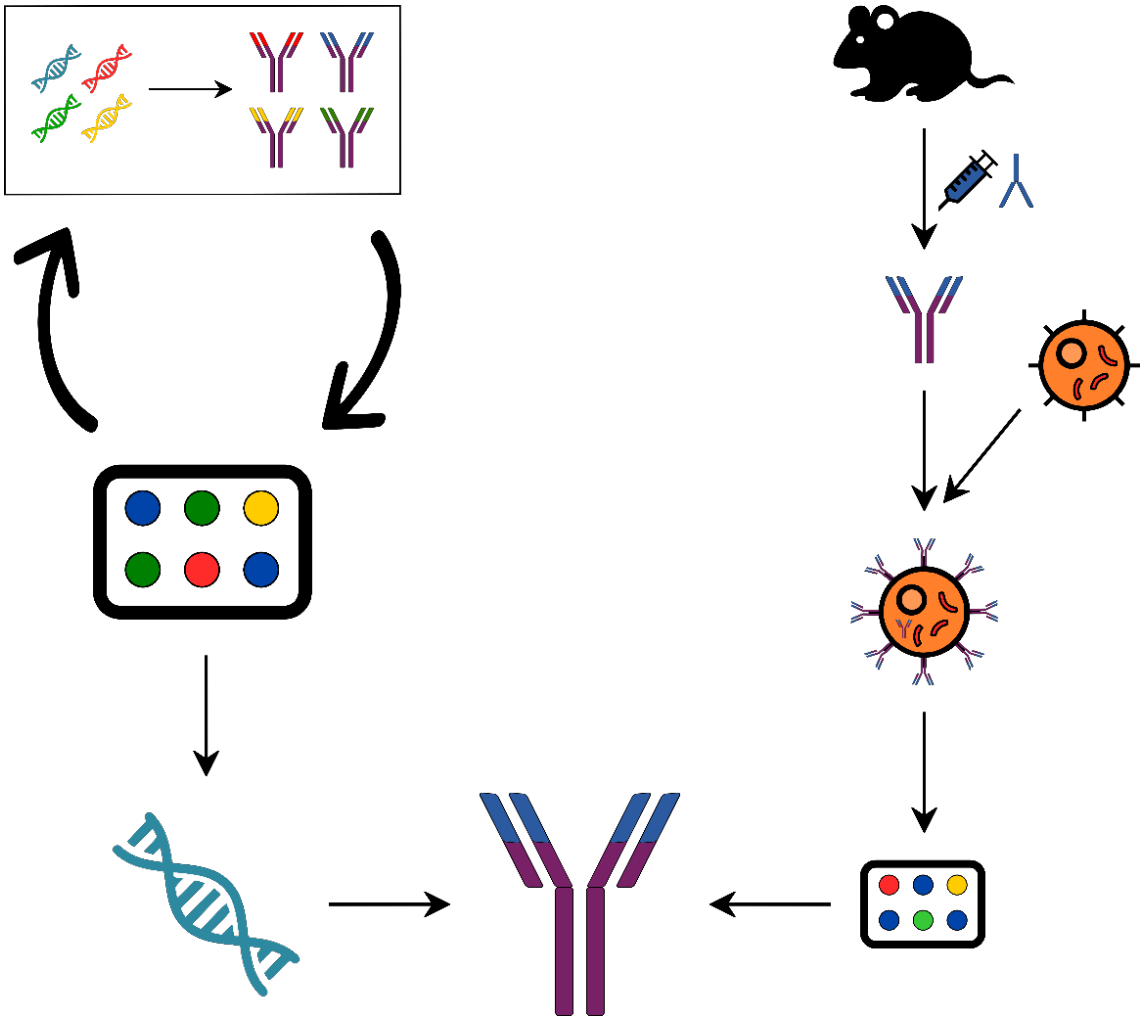


Figure 1.L: ***In vitro* and *in vivo* antibody screening and selection techniques.** The left side of the image shows a general way to perform *in vitro* selection of antibodies. This starts with a library and through rounds of testing and selection, the desired antibody is found and its DNA sequence obtained to produce more of it. The right side of the image shows the traditional *in vivo* (hybridoma) method, in which mice are injected with a specific antigen and the antibodies they produce are harvested. These antibodies are fused with B cancer cells to make a new cell line that can be used to harvest antibodies.

Different display techniques are available to choose from: phage display [134], yeast display [135,136], ribosome display [137], liposome display [138], DNA display [139–141], cDNA display [142,143], mRNA display [144–146]. Although the majority of them are performed *in vitro*, some still are performed by using microorganisms (yeast display, bacteria display, etc …). But even so, they all use different strategies for the same purpose: to couple genotype to phenotype; and therefore all have become indispensable tools for directed evolution. Among the display methods, *in vitro* approaches can screen the highest number of molecules, up to 10^{15} for mRNA, cDNA, and ribosome display. For both phage and bacterial display, the library size is around 10^{11} sequences (yeast display is even smaller), since they are dependent on the transformation efficiency. Furthermore, cell-based display methods are restricted to the use of natural elements. Using bigger libraries helps with the probability to select rare and unusual sequences and improves the total diversity. Moreover, *in vitro* reactions can be easily modified to suit a specific environment for functional screening [147]. Whether one or another display technique is used, if it very important to link one genotype to its one phenotype, otherwise, false positives could occur.

1.4.2 mRNA Display and cDNA Display

This method (Figure 1.M), is used primarily for *in vitro* protein, and/or peptide evolution. The main procedure starts with a DNA library for proteins or the peptide of interest. Depending on both the *in vitro* transcription and the translation systems to be used, the specific elements (promoter, RBS, etc …) need to be included in the DNA constructs. This library is first transcribed into

an mRNA library and further purified with a gel to be later ligated to a puromycin molecule using a T4 ligase system.

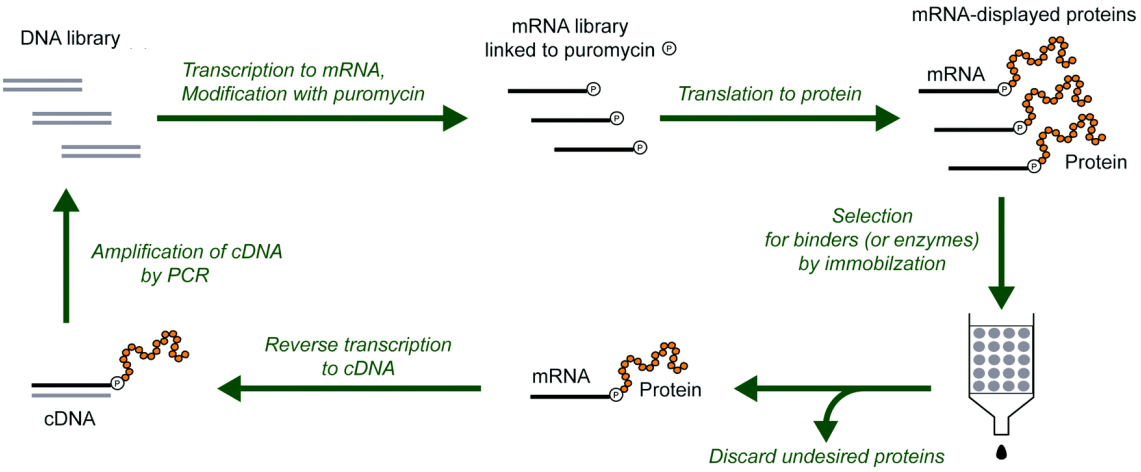


Figure 1.M: mRNA display overview. Parting from a DNA library, the molecules are transcribed into mRNA and then puromycin molecules attached to them. Later, the mRNA library is translated and the resulting product is a library of mRNA-protein conjugates. The proteins with the highest affinity, are selected and separated from the rest. The selected variants are reverse transcribed to cDNA and their numbers increased by PCR. Several further rounds of selection are performed to minimise the number of undesired proteins. Image used is from Blanco et al. (Physical Chemistry Chemical Physics, 2020) [148].

These ligated mRNA molecules are then translated, and their resulting peptides (or proteins) become attached to their coding mRNA molecules, by a covalent link to the puromycin molecule. The mRNA-peptide complexes can be further immobilised through a selection step, such as affinity chromatography, in which the sample attaches itself to a column, a well plate or magnetic beads coated with the immobilized ligand. The ones that do bind are then chosen to go through reverse transcription which gives a cDNA molecule sequence for each specific

bound peptide. Error-prone PCR can be also performed to increase the diversity for posterior cycles. If several cycles of selection are performed, the number of molecules with a weak affinity can be minimised. Afterwards, by sequencing the obtained cDNA molecules, the sequences of the peptides with high affinity to a specific molecule can be known.

Puromycin is the molecule used to bind the mRNA to the peptide. It is an antibiotic protein synthesis inhibitor that causes early translation termination. Puromycin is an analogue of the 3' end of a tyrosyl-tRNA with a part of its structure similar to adenosine, and the other to tyrosine (Figure 1.N) As a result, it causes a premature release of translation products. Puromycin is connected to the 3' end of each mRNA molecule after the transcription from DNA; so, when it comes to the translation step, once the ribosome reaches the 3' end of the mRNA molecule, the fused puromycin will enter the ribosome's A site and be incorporated into the peptide that is being formed. The mRNA-polypeptide conjugate is then released from the ribosome. The puromycin attachment in itself is usually not enough, and some other modifications are needed. Extra oligonucleotides and spacers are needed, in the DNA sequence, to provide flexibility and enough space for the puromycin to enter the A site of the ribosome [149].

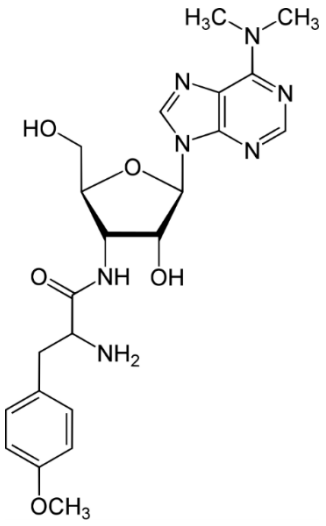


Figure 1.N: **Molecular structure of the puromycin molecule.** Puromycin is an analogue of the 3' end of a tyrosyl-tRNA with a part of its structure similar to adenosine, and the other to tyrosine.

As mentioned in the previous section, there are other different display methods, but mRNA brings certain advantages in comparison to them [150,151]. To start with, since it is an *in vitro* method, it already avoids the problems when working with yeast or bacteria cells, like reduced library size or increased method time and cost. When it comes to comparing mRNA display to other *in vitro* methods a unique advantage observed is the use of a covalent bond of the mRNA molecule and its peptide. For example, ribosome display has a non-covalent bond between the ribosome, the mRNA and the peptide, which hinders the complex stability and in return reduces the efficiency of the binding process. Apart from that, the mRNA-peptide complex is much smaller than that of its ribosome display counterpart, which helps reduce greatly the probability of unwanted peptide interactions (like

peptide to ribosome). So in this sense, mRNA display has the potential for fast evolution and selection of high-affinity diagnostic and therapeutic antibodies [152].

Cell-free systems offer a great advantage when it comes to tight control over the reaction environment, so when coupled with display techniques, they are useful for high-throughput screening and directed evolution of peptide/proteins [122,153,154]. But some problems arise due to the instability of the mRNA-protein conjugates, which makes cell lysate-based translation systems have limited utility for mRNA display [155–157]. While this issue is usually addressed by using of RNAse inhibitors and nuclease-free chemicals, to minimise RNA degradation [144,158], advancements of reconstituted contaminant-free "PURE" translation system have made *in vitro* display methods more popular for screening antibodies [159,160] and functional peptidomimetics [161]. Since the PURE system operates primarily with reconstituted components, it offers increased stability of mRNA-protein conjugates and also allows for greater flexibility in selection condition when combined with the mRNA display.

In addition to the already mentioned ways to improve the stability of the mRNA molecule, a variation of the mRNA display can be used. The cDNA display method (Figure 1.O), converts unstable mRNA-peptide conjugates into mRNA/cDNA-peptide conjugates. This new conjugate is advantageous under conditions where RNA instability is an issue during the selection step, such as targeting cell surface antigens under the presence of cellular ribonucleases [142,162].

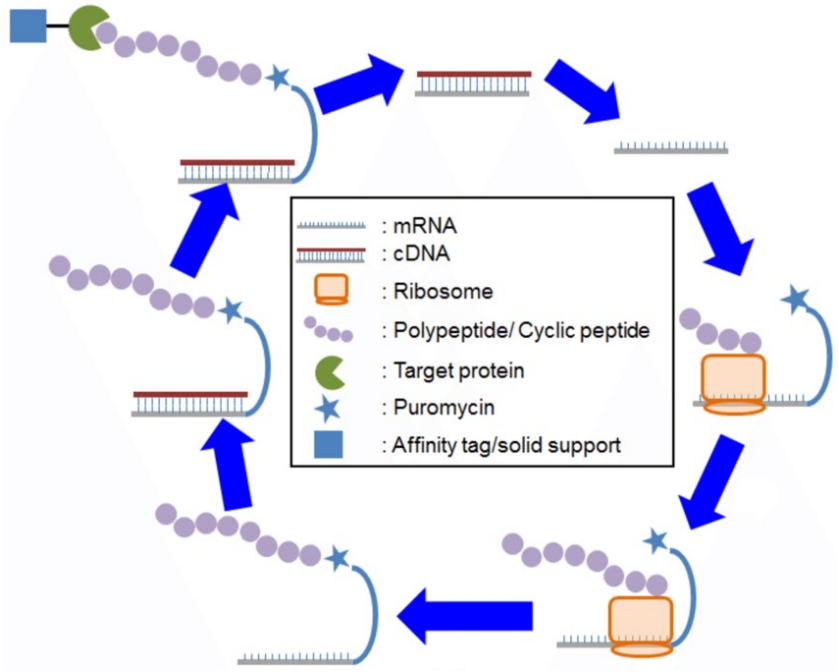


Figure 1.O: **cDNA display overview.** Almost identical to the mRNA display, this method helps with the stabilisation of the mRNA molecule by going through a reverse transcription step after the translation process. This mRNA/cDNA duplex is stronger to the formation of secondary structures and unwanted interactions with other elements. Image used is from Ong *et al.* (Current Topics in Medicinal Chemistry, 2020) [163].

1.4.3 Ribosome Display

Ribosome display (Figure 1.P) is an *in vitro* selection process for proteins and peptides. Much like mRNA display, it involves the translated peptide being bound to its mRNA molecule, but it does so through the ribosome. The mRNA-protein-ribosome complex uses the ribosome as the connection molecule between the other two components. These complexes also go through a binding process in which those with high affinity are selected so that later the mRNA can be

transformed into cDNA by reverse-transcription. The cDNA molecule is later amplified and sequenced to obtain the desired peptide sequence. Ribosome display is suitable for displaying big libraries and working with special requirements, such as amino acid modification, and unstable proteins.

This method also begins with a library of DNA sequences, but the constructs have a spacer sequence lacking a stop codon at their 3' end. The library then goes through transcription and translation, but the absence of a stop codon prevents the binding of release factors and therefore the mRNA-ribosome-peptide conjugate cannot be dismantled. The stability of the mRNA-ribosome-peptide is usually aided by lowering the temperature of the reaction and the addition of cations (like Mg2+). The spacer sequence in the mRNA molecule gets attached to the peptidyl tRNA allowing the peptide to go out of the ribosome and fold. Posteriorly, the complexes go through a binding process and once the peptides are attached to their ligands and restrained from movement, elution of the mRNA molecules is performed using high salt concentrations, chelating agents, or ligands. The mRNA is then reverse transcribed back into cDNA which goes through further selection cycles.

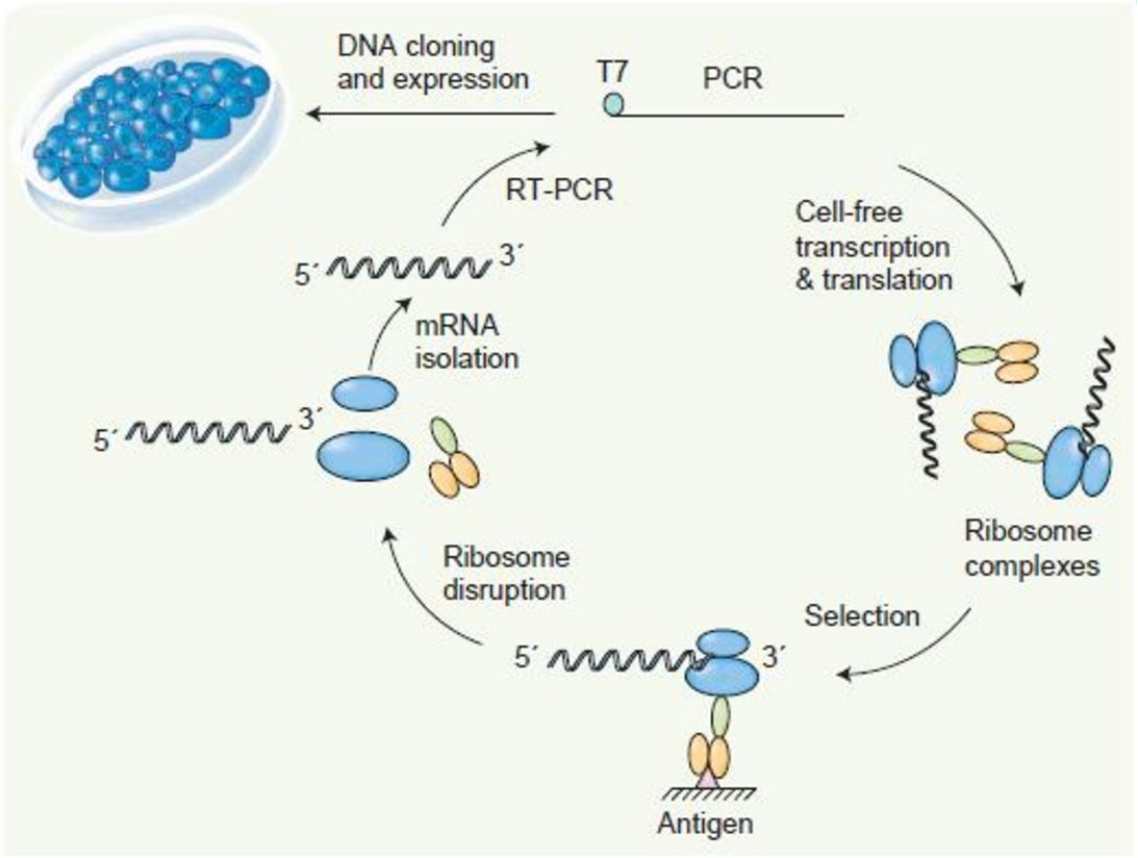


Figure 1.P: **Ribosome display overview.** Starting from a DNA library, and after being transcribed and translated, an mRNA-ribosome-peptide complex is formed. This conjugate goes through a selection process in which the mRNA molecules of the bound peptides are eluted and reverse transcribed to obtain cDNA sequences. Image used is from www.creative-biolabs.com.

Ribosome display provides a fast way to identify and produce the proteins that are selected, whether their sequence is known or not. Furthermore, as with all the *in vitro* methods, large pools of sequences can be made and analysed rapidly and the mRNA-ribosome-peptide complex keeps them from degrading or interacting with each other. Although it has a few drawbacks, like the lack of a stable enough link within the ribosomal complex or the higher probability of the peptides interacting

with the ribosome, it is still widely used for antibody and protein engineering regarding therapeutic approaches [164].

1.4.4 Phage Display

Phage display (Figure 1.Q) is the most widely used *in vivo* display technique for protein-protein, protein-peptide, and protein-DNA interactions. It uses bacteriophage viruses as a way to create a link between the genetic information encoded inside them, and the proteins that they display. A specific target gene, that codes for a certain protein, is inserted inside the reading frame for the virus coat proteins. This results in the phage displaying on its surface the protein of interest while retaining the inserted gene inside. Since the displayed protein is on the outside of the phage, it can be easily used for binding and selection processes. Depending on the protein, peptide or antibody of interest, different phage viruses can be used, such as M13, T4, T7, and λ [165].

Phage display starts with a DNA gene library which is cloned into the virus coat proteins. Different cloning sites are used at the same time to ensure the desired genes are inserted in all available reading frames and no problems arise during the translation process. This hybrid library (gene of interest and coat protein gene) is inserted inside bacteria cells, usually, *E. coli*, though transduction so the assembly of the hybrid DNA into the virus particles can be done. Sometimes, depending on the type of genetic construct made, a helper phage (or a bacterial packaging cell line) might be needed to aid with the assembly process. Once the viruses are assembled, they are released from the bacteria and can then be used for binding

processes. Since the displayed protein is on the virus surface, they will bind to the protein target and stay immobilised, while the ones that do not bind, are washed away. Posteriorly, the selected viruses can be eluted using a low-pH elution buffer along with sonication and used to transfect bacteria cells again to produce more phage. In this case, performing several selection cycles is called "panning". After several cycles of panning, the final phage can be used to recover the DNA sequence for sequencing and identification of the target protein.

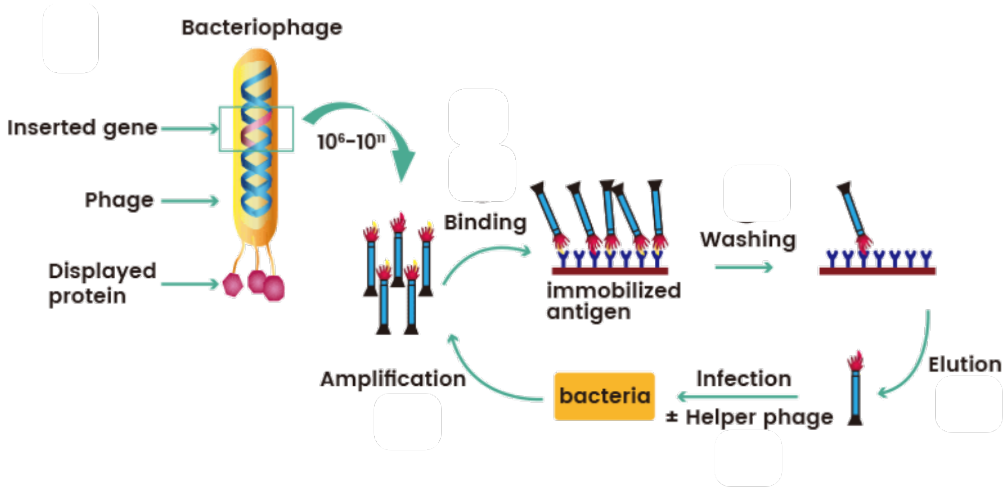


Figure 1.Q. Phage display overview. The process starts with a virus library in which the inserted gene encodes for a surface protein. Then through amplification of the virus numbers by bacteria transfection, posterior selection processes, and repetition through a round of selection, a protein candidate is selected and its encoding gene sequenced. Image used is from www.sinobiological.com.

Applications of phage display are bound to the same principle of display methods, to create a link between the displayed protein (phenotype) and its encoding gene (genotype). This method is widely used for *in vitro* protein evolution and therefore a popular tool for finding new ligands to certain target proteins (drug discovery).

It has been used to find, for example, antigens to tumours cells so that they can aid in diagnosis [166] but also antibodies for cancer treatment [167]. The main reasons phage display became so popular for antibody screening, when compared to hybridomas, is the ability to do large scale production, the high speed and easiness of the method, but also the lack of animal use. Furthermore, it is also possible to screen for toxic antigens as the viability of the bacteria cells is only necessary for a short period.

1.4.5 Cell Display: Bacterial and Yeast Display

Organisms can be used to perform *in vivo* display. As all display techniques, it is used for the discovery and selection of peptides, proteins and antibodies. This process can mainly be separated into two types of cells used: yeast and bacteria.

Yeast display (Figure 1.R), as its name suggests, uses yeast cells to display a library of recombinant proteins on the cell wall surface [168]. The gene, of the protein of interest, is encoded with the Aga2p protein gene, so the result is a fusion protein. The Aga2p protein is used by yeast to interact with other cells during reproduction cycles, and as such, it is displayed on the cell surface and away from other proteins. The selection process is done mainly through magnetic-activated cell sorting (MACS) or fluorescence-activated cell sorting (FACS) techniques, which both are performed in real-time. When it comes to the advantages of yeast display, it is worth mentioning that since eukaryotic cells are being used, it is a more suitable environment for the folding, post-translational modification, and translocation of human proteins. But even if beneficial for human protein screening, this method

has one of the lowest library sizes to work with (~10^8) and glycosylation of proteins is different in yeast and mammalian cells.

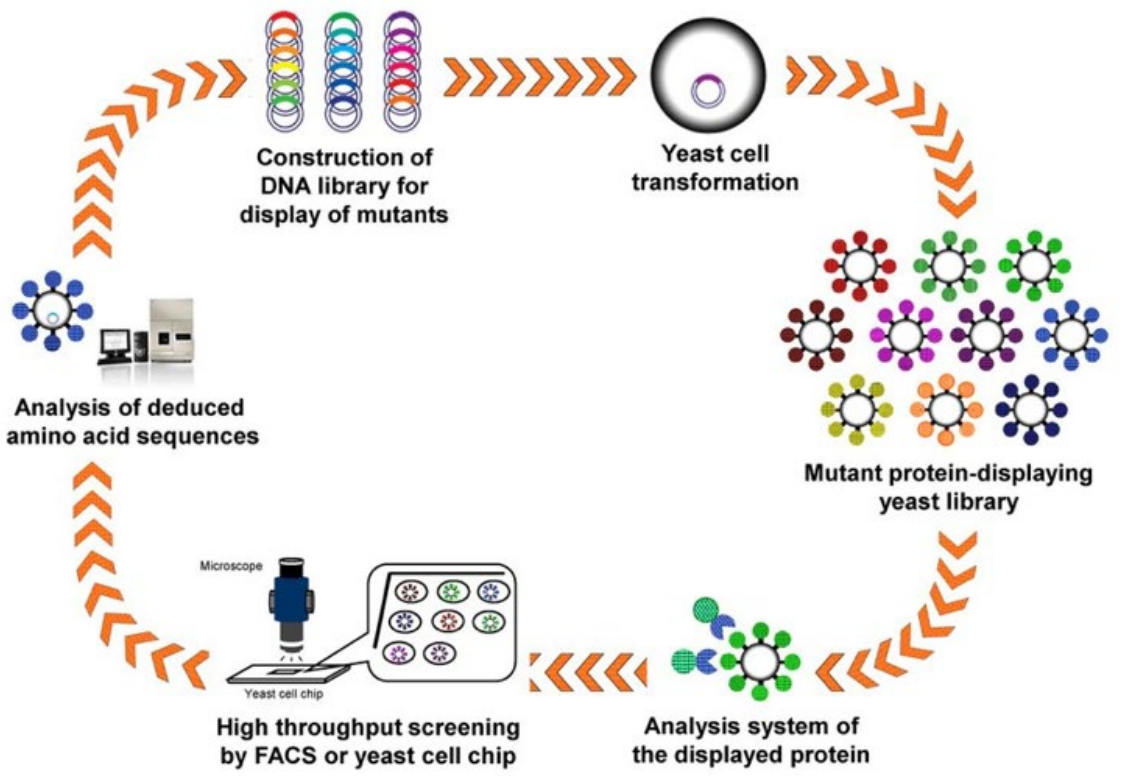


Figure 1.R: Yeast display overview. This display method starts with the coding of the gene of interest with the gene of the Aga2p surface protein. As a fusion protein, it is directed to the cell surface where it can be screened and selected using cell sorting methods. Image used is from www.creative-biolabs.com.

Bacterial display, on the other hand, uses bacteria cells instead of displaying engineered proteins. Much like yeast display, it starts with a DNA library, but in this case, the genes are coded with any surface protein (called a scaffold). These surface proteins are used for a wide range of functions, like cell recognition,

interaction, and signalling; and as with all surface proteins, they are translocated from the cytoplasm to the membrane. Since gram-negative bacteria have a periplasmic space (not found in gram-positive bacteria), they have a more difficult time translocating proteins to the surface. Once the fusion protein is displayed on the bacterial surface, cell sorting methods can be used for the selection process (FACS or MACS). The bacterial display can be used for all the typical applications of display methods: antibody mapping, peptide, and protein screening, etc. But there is a specific application used with this method, called vaccine delivery, in which they use the bacteria cells (with the displayed surface protein) as a delivery system directly for the vaccine. To do this, the bacteria cell has to be weakened so it is no longer pathogenic, or safe bacteria are used instead (food-grade bacteria). This way of delivery brings some advantages, such as the use of other proteins expressed on the bacteria surface as adjuvants (an agent that improves the immune response of a vaccine).

1.4.6 Liposome Display

Liposome display (Figure 1.S) is a recently discovered *in vitro* technique that involves the use of liposome and cell-free systems for the screening and selection of membrane proteins. This approach uses the liposome and cell-free system combination to create a micro-reactor or an artificial cell. Inside the liposome, the gene of interest, such for example a DNA molecule, is transcribed and translated by a cell-free system. The produced membrane protein is translocated to the liposome membrane and displayed there. These liposomes, as with bacteria and

yeast, can be screened and selected using cell sorting techniques (like FACS). The DNA encoding the membrane protein can be easily recovered from the liposomes to forgo more rounds of selection [122].

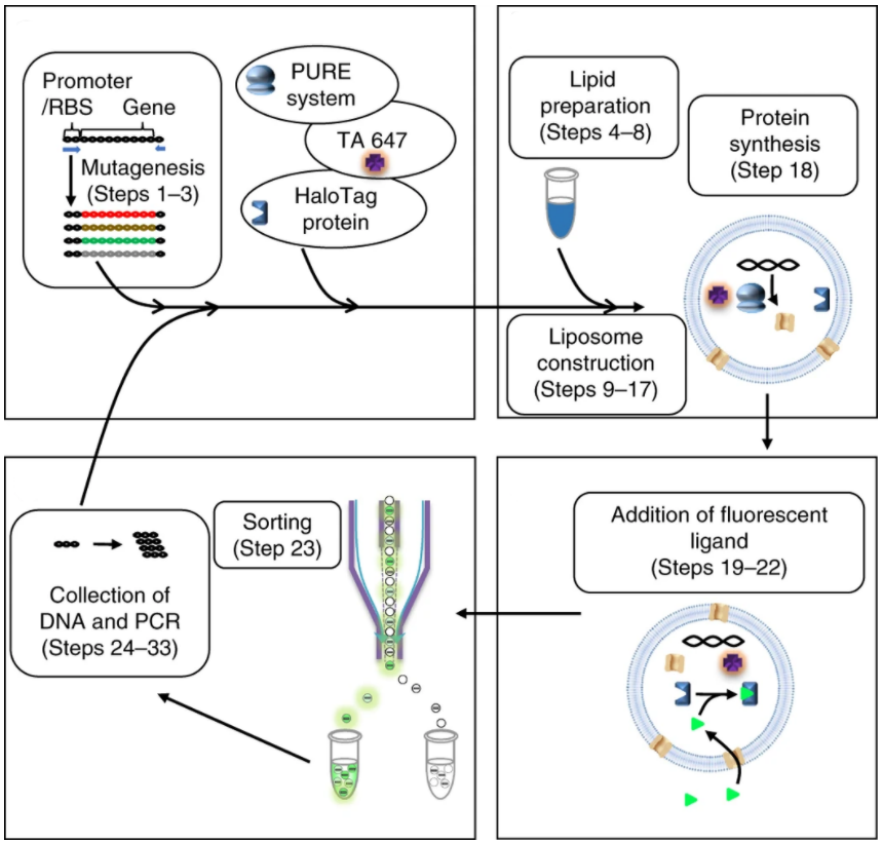


Figure 1.S: Liposome display overview. A novel method based on the use of both liposomes and cell-free systems to create artificial cells so membrane proteins can be expressed and displayed on the surface of the liposome. TA647 is a fluorescent molecule used in the inner solution of the liposome as a volume marker to know the size of the liposomes. The sorting and collection processes are performed by FACS analysis. Image used is from Fujii et al.[122].

Although the number of the DNA library is not as big as other display methods (~108), this technique offers speed in the selection process and the benefits of both in vitro and in vivo processes. Since it is an in vitro process, it allows for a high

degree of protein engineering and reaction environment control. But since the liposome mimics an artificial cell, it makes it possible to work with membrane proteins (and not only soluble globular proteins) and use cell-sorting methods for the screening and selection process.

1.5 Artificial Lipid Membranes and Membrane Proteins

1.5.1 Artificial Lipid Membranes Composition, Characteristics and Uses.

Artificial lipid membranes, or model lipid bilayers, are composed of a double layer of lipid molecules, that were assembled *in vitro*, and that try to mimic the natural cell membranes found in cells [169]. These synthetic membranes can be used to study the properties of biological membranes in a simple controlled *in vitro* environment [170] or just to construct artificial cells [171]. Furthermore, since they can be constructed with natural or synthetic elements, with one or several lipid mixtures, these membranes models are useful tools for the study and discovery of novel lipids. These advantages provide easier ways to improve the stability and fluidity of membranes. Furthermore, a wide range of model membranes can be constructed for different experimental demands. When these synthetic membranes are produced, they can stay in a soluble state or be anchored to a solid surface or substrate. If they are static, their stability increases making them more ideal to work with.

The simplest model bilayer is called black lipid membrane (BLM). To make them, first, a small fissure is made in a layer of a hydrophobic material, then the area around this fissure is coated with a solution of lipids (in a hydrophobic solvent). After drying, a salt solution is added on both sides and another lipid solution is coated for a second time. Between the organic and the aqueous phase, two lipid monolayer form which ends up fusing creating a bilayer (Figure 1.T, A). This bilayer separates two opposite chambers, which makes it ideal to study membrane proteins like ion channels and transportation proteins [172]. Membrane proteins can be spontaneously inserted into the bilayer by solubilising the protein into the aqueous solution. The main disadvantage of this membrane model is its short lifespan (measured in hours) and the contamination of the bilayer with unwanted solvents that can be trapped in-between the two monolayers.

A variation of the BLM model is the tethered bilayer lipid membrane (t-BLM). This method uses a BLM and anchors it to a solid surface to increase the membrane stability. Since unwanted interactions between the membrane (or the membrane protein) and the surface are a possibility, a material with inert chemistry is usually used, for example, gold. For the binding between the surface material and the membrane, thiolipids (lipid derivatives) are used (Figure 1.T, B). The bilayer is formed after the lipid monolayer is tethered to the gold surface by exposing it to a lipid solution [173]. The space between the hydrophobic head is usually bigger than in the BLM model, which makes it beneficial to work with bigger membrane proteins.

When the synthetic membrane is not attached to a solid surface by thiolipids or any other components, and just by the hydrophobic heads of one of the layers, it is called supported lipid bilayers (SLB) [174]. Due to the configuration of this model, only one of the monolayers is exposed. This characteristic brings enhanced stability to the membrane lasting weeks, which helps when using certain characterisation methods that use vibrations or movements that may break the bilayer.

If the lipid bilayer closes around itself, creating a sphere, and leaving a solution inside, it is a vesicle. Since they mimic very well the physiology of a cell, vesicles have been used widely to study the dynamics of cell membranes. They are produced in a very easy straightforward way, by exposing dehydrated lipids to a polar solution, which in turn makes the lipids close around themselves by creating an orb (Figure 1.T, C). Since the creation of the vesicles happens spontaneously, these tend to be of various sizes and contain several monolayers, reason why they are called multilamellar vesicles (MLV). Different methods, such as sonication, can be used to break these initial vesicles (or liposomes) into smaller unilamellar ones (SUVs). Since they can be prepared economically and fast, they are preferred for bulk studies and long term storage. SUVs are sometimes too small for certain studies, such as detailed fluorescence microscope imaging, and as a result, bigger versions can also be created, called giant unilamellar vesicle (GUV). But GUVs do take more manufacturing time, are more fragile to work with, and produce fewer numbers of total vesicles than SUVs. Another possible way to obtain vesicles is to

just isolate them from organisms or cells [175], but since they are of natural origin, they have pre-determined compositions and so this is not usually the preferred method. Since liposomes can be used to enclose markers or drugs inside themselves or to just express membrane proteins on their surfaces, they are widely used as tools for diagnosis and treatment [176]. Vesicles can help in the correct temporal delivery of drugs (chemicals, oligonucleotides, genes, and recombinant proteins) without compromising their quality, and providing reduced toxicity to the host. Another practical use for these liposome microreactors is to take advantage of their lipid bilayer as a location for membrane proteins to get attached to. This method is mainly used as a way to study membrane proteins and their properties [177]. Studying this type of proteins has proven to be challenging, mainly due to the difficulty in trying to find optimal conditions *in vitro* that would allow membrane proteins to keep their original correct conformation and functionality [178]. Liposomes provide an enclosed environment for the membrane protein to be produced and, at the same time, allow for its insertion and correct conformation into the liposome membrane. The liposomes' membrane offers a natural environment for the membrane proteins to be studied in terms of their structure and functionality [179]. The correct composition of the liposome's lipid bilayer is fundamental for the membrane protein and therefore the optimisation of this process is necessary for each different membrane protein produced [180]. Although liposome-based vesicles seem like a favourable drug delivery method and an easy and fast way to study membrane proteins, they have some fundamental issues due

to the heterogenous multilamellar condition of the liposomes. This issue can lead to different heterogeneous conformations of the same proteins and when it comes to oligomer proteins, differences in their monomers [181].

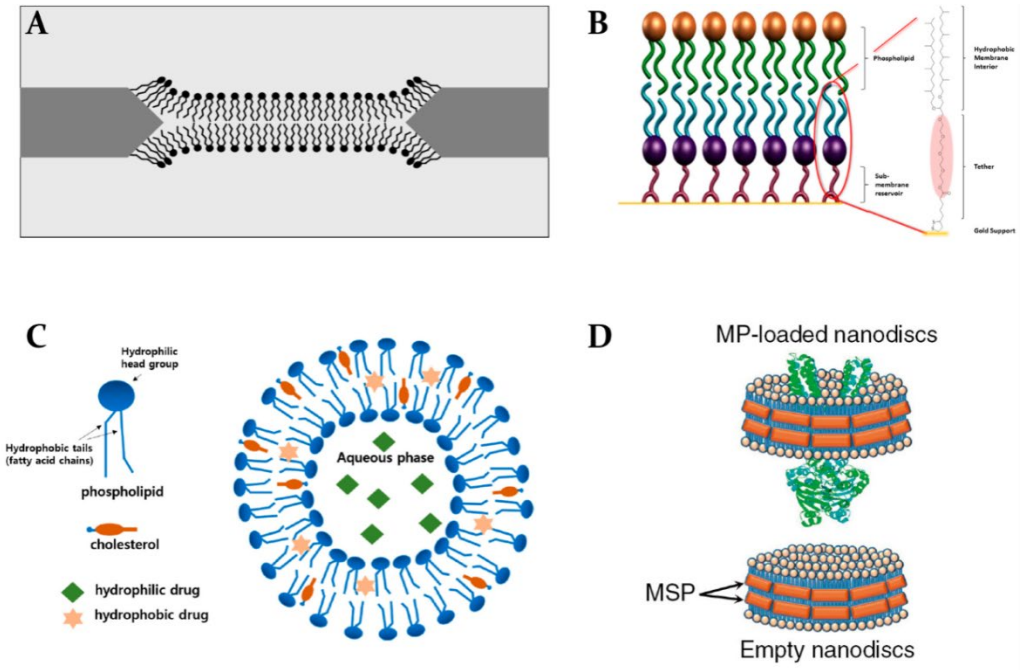


Figure 1.T: Types of synthetic lipid bilayers. (A) The simplest model bilayer, black lipid membrane (BLM); (B) an anchored version of a synthetic lipid membrane, tethered bilayer lipid membrane (t-BLM); (C) a liposome or vesicle; and (D) a figure of an empty nanodisc and one with an embedded membrane protein. Images used are from: (A) Hąc-Wydro *et al.*[182], (B) Andersson *et al.*[173], (C) Lee [183], and (D) Zoghbi *et al.*[184].

As an alternative, nanometric lipid bilayers, or nanodiscs, have started to gain popularity to perform *in vitro* experiments of membrane proteins [185]. These nanodiscs are also synthetic model membranes composed of a phospholipid bilayer and two membrane scaffold proteins (MSP) surrounding the hydrophobic region of the phospholipids (Figure 1.T, D). The popularity of these nanodiscs

comes from the fact that they offer a favourable and flexible way to study membrane proteins (MP) as the size and lipid composition can be easily adjusted. But also due to the increased stability which allows for elongated storage and analysis of the MP-nanodisc complex [181]. Another convenience is that, because of the aforementioned advantages, it is possible to regulate the oligomerisation of oligomeric proteins [185]. The membrane proteins produced in nanodiscs have shown to be quite stable for further analysis such as Nuclear Magnetic Resonance (NMR) [185–187]. Small drawbacks when working with nanodiscs can be some issues with loss of lipids when compared to liposomes, [188] and the time-consuming process of producing the MSP and finding the optimal lipid composition. But once these issues are dealt with, the nanodiscs can be produced in big quantities and stored for very long periods which makes them eventually very convenient in saving time.

1.5.2 Membrane Proteins: G-Protein Coupled Receptors (GPCRs) and their Applications for Drug Screening.

When a protein interacts or is part of a biological membrane, it falls into the membrane protein category. Almost 1/3 of all the human proteins are membrane proteins, so they play an important role in drug targeting. These proteins are quite difficult to characterise since they always depend on a lipid bilayer to properly localise to, and therefore determining the optimal conditions to ensure the correct conformation of the membrane proteins is a challenge. These membrane proteins play important roles for the cell [189]. One of their main functions is to serve as receptors, they help cells communicate between themselves and identify each

other, but also respond to external signals. Another function is that they help transport molecules and ions through the membrane in either direction. Apart from those two main roles of being receptors and transporters, membrane proteins can carry out a variety of other activities (enzymes, connectors, anchors, etc). Independently of their functions, they also fall into two categories according to their localisation on the cell membrane: Integral and peripheral. Both types usually go through post-translation modifications so that they can bind to the lipid membrane.

Integral membrane proteins are always located on the membrane and they tend to cross the whole membrane from one side to the other (Figure 1.V). Since they are fully integrated in the membrane, it is quite difficult to isolate them, as it would require to free them from the membrane using detergents or nonpolar solvents, which tends to damage the protein. If they cross more than once the membrane, they are called polytopic, and if they only cross it once they are called bitopic. Some do not fully cross the membrane all the way and are only attached to one of the monolayers, these are called monotopic. Another way to classify integral membrane proteins is depending on the motifs found in their secondary structure and how they assemble in their tertiary structure: alpha-helical (single or as a bundle) and beta-barrels (Figure 1.U). They can also be classified depending on which side of the membrane their C- and N- terminus ends up.

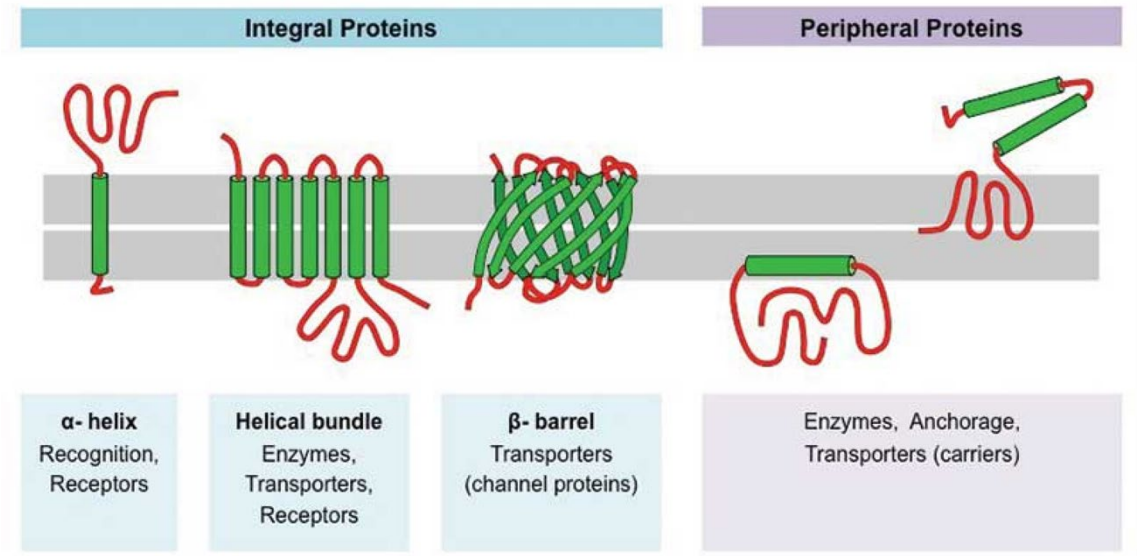


Figure 1.U. Types of membrane proteins. The main division of membrane protein is whether they are integral or peripheral. Inside the integral proteins, depending on the conformation they take, they can be part of the alpha-helix (single or bundle) or the beta-barrel. All membrane proteins have a wide verity of different functions. Image used is from www.ib.bioninja.com

Peripheral membrane proteins are only temporarily attached to one end of the lipid membrane or they are attached to integral proteins (Figure 1.U). Since they are not embedded in the membrane, they are much easier to isolate, only requiring treatment with a polar solution. They attach themselves to the lipid membrane through a mixture of hydrophobic, electrostatic, and other non-covalent links. Just as with the integral membrane proteins, they have a wide variety of functions, from transporters to acting as enzymes.

One of the main reasons membrane proteins are so important to study is that they are optimal targets for clinical and drug treatments of various diseases, especially for cancer, since they tend to overexpress themselves [190] and/or alter their post-

translational modifications [191] in cancerous cells. Also, they are easy targets as they are found in the surface of every cell which makes them accessible from the outside, but also because they have numerous antibodies as targets [192]. The family of membrane proteins most important as pharmaceutical targets are the G protein-coupled receptors (GPCR). The GPCR family of receptors is found in the surface of the lipid membrane of eukaryotic cells. They have specific roles to sense external stimuli and send an internal cascade of signals that will alter the gene expression of the cell (Figure 1.V). These proteins, the largest membrane receptors family, can recognise a huge variety of different ligands: neurotransmitters, mediators, hormones, peptides, lipids, amino acids and nucleic acids, as well as being able to respond to light. All these ligands can either way bind to the extracellular loops and N-terminus end, or the intramembrane helices. Since they cross the lipid membrane seven times, they are also known as seven-transmembrane receptors. As the name suggests, they are coupled to G proteins (signal transmitters). Once a ligand binds to the GPCR, it causes a conformational change which in return activates the G protein. Since they are receptors to so many different types of ligands and they are involved in many of the cell's signaling pathways, they tend to be highly involved in many diseases (cancerous, endocrinal, immunological, etc) when they suffer mutations or polymorphisms [193]. Around half of the commercial drugs in the market act as targets of GPCRs.

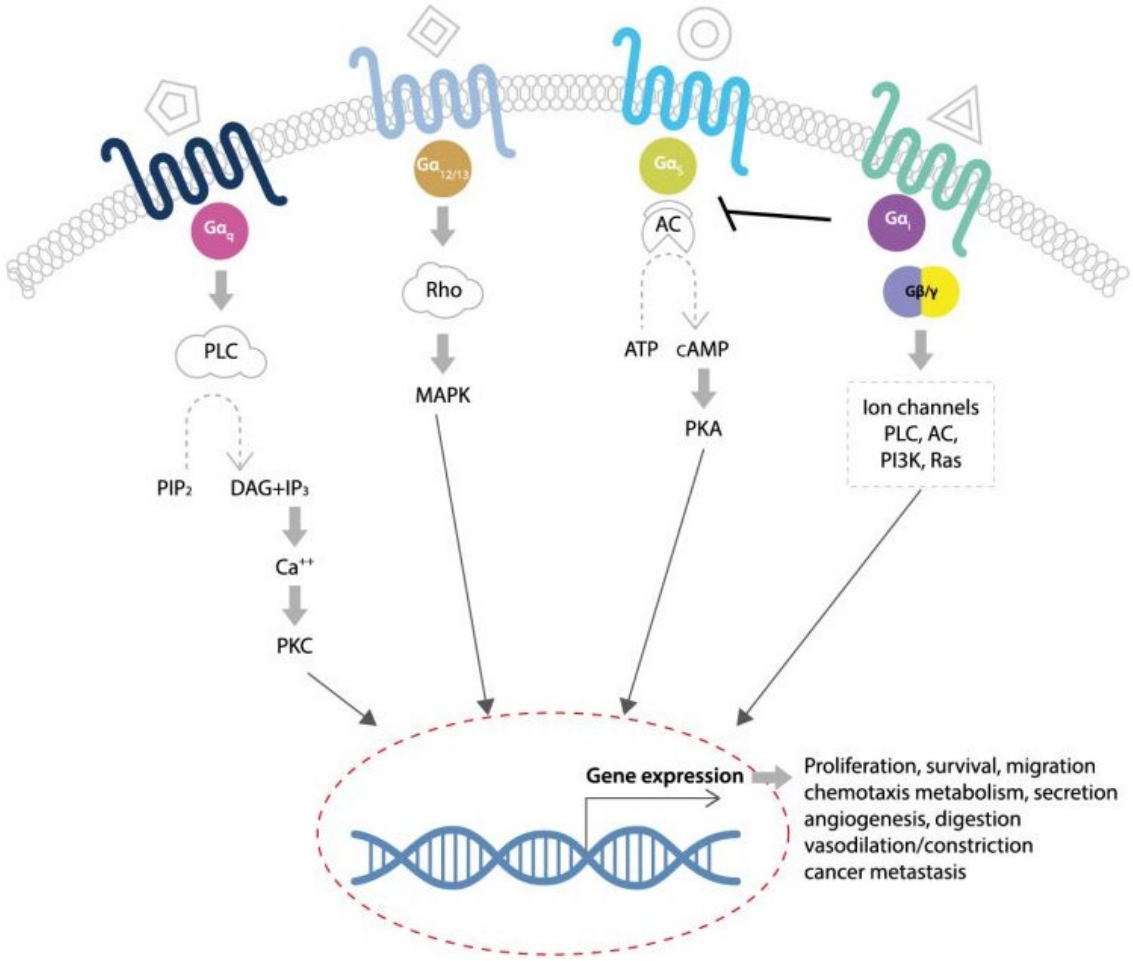


Figure 1.V. Different signalling pathways of GPCRs. The image shows details of the GPCRs reacting to external stimuli and all the possible ways they induce a cascade of signals inside the cell which has the ultimate goal of altering gene expression to adapt to the new signal. Image used is from www. gpcr.co.kr.

GPCRs are of great importance due to their involvement in human pathophysiology and because they are easily handled in the pharmaceutical industry [194]. Their genetic variation, along with the fact that they are attractive targets for pharmaceutical drugs, leads to a constant necessity to find new possible antibodies and proteins to target GPCRs.

1.6 Thesis Aims

The main aim of this thesis is to understand *in vitro* cell-free systems and display techniques that could help create a novel interaction of both methods to screen and select from a wide library of peptides so that possible drug targets against GPCRs protein (expressed in synthetic membranes) could be found. In order to carry out this, several objectives were followed.

Chapter 3 starts by performing a study of two different types of cell-free systems: based on cell extracts and based on purified components. This study was done to understand the dynamics behind each system and how they behave with a library of different samples, as well as to find ways to improve their weak spots. From this chapter, an optimal cell-free system (PURE system) was selected to work with for the rest of the chapters. For this chapter, two objectives were in mind:

1. Data analysis and modelling of the performance of two cell-free protein expression systems by using a promoter mutant library, for the optimization of product yield and minimization of degradation.

Cell-free protein expression systems are widely used for many applications regarding synthetic biology and metabolic engineering. More information was gathered from the dynamics of protein expression within these systems, including

on the impact of feedback loops and possible bottlenecks. A systems approach was used to uncover the underlying dynamics of protein expression in cell-free systems.

2. Design and analysis of mutant libraries in cell-free systems, to better understand relevant interacting nucleotides or sections within the promoter and identify beneficial mutants.

Coupled with the previous objective, a mutant library of a promoter displayed a range of protein level expression. This information could be valuable for performing cell-free system reactions in which individual tuning of protein expression is needed due to the presence of several genetic circuits. Also, beneficial mutants could enhance the final protein yield. The results obtained could also feedback into the first objective by feeding the model information about changes during transcription. Further exploration of the interactions of each mutant nucleotide (single nucleotide changes within the promoter) with the polymerase protein provided a starting point for the discovery of enhanced promoter sequences.

Continuing, chapter 4 shows a possible method for the screening and selection of peptides using the cell-free system chosen in chapter 3, the PURE system, and two different *in vitro* display methods, mRNA and cDNA display. The main objective was:

3. Determination of a methodology that could be used for clean and efficient sequence screening of candidates for drug screening using *in vitro* display techniques with cell-free systems (objective 1) and mutant libraries (objective 2).

The use of an appropriate cell-free system improved the efficiency of *in vitro* display techniques to prolong the stability of peptide conjugates. Also, it diminished the presence of components that could create undesired interactions with the ligand candidates. These advantages could provide a faster enrichment rate and a higher number of selected peptide candidates than traditional methodologies. These two aspects may prove useful when working with time-constraints of difficult proteins (like membrane proteins).

Chapter 5 is aimed at exploring synthetic lipid membranes and cell-free systems, for the production of G-protein-coupled receptors (GPCRs) membrane protein, a receptor from the G protein-coupled receptors (GPCRs) family, which are highly researched targets for drug screening in medical therapy. By using the PURE system, the CX3CR1 membrane protein receptor was produced in their natural conformation *in vitro*, in two types of synthetic membranes: liposomes and nanodiscs. This chapter also proposes a type of DNA library that could be used to

screen for possible candidates for CX3CR1 when using an *in vitro* display method. The main objective of this chapter was:

4. Expression of membrane proteins with the PURE cell-free system using two different types of synthetic membranes: liposomes and nanodiscs.

Synthesis and purification of membrane proteins using different synthetic membranes provided insight into the difficulties encountered when working with membrane proteins and how to overcome them. The use of an appropriate cell-free system could enhance the protein yield, stability of the membrane-protein complex and efficiency of the purification steps.

In summary, this thesis presents research made in-depth to analyse the dynamics of cell-free systems and screening selection methods. The combination of a transcription-translation system from purified components with *in vitro* selection allowed the production of highly stable peptide conjugates from a DNA library. All the proposed objectives could further be combined into an efficient and rapid method to screen and select for possible drug targets against membrane proteins. This could overcome the difficulties found in high throughput screening of these types of proteins.

1.7 Chapter 1 References

1. Keasling, J. D. Synthetic biology and the development of tools for metabolic engineering. *Metab Eng* **14**, 189–195 (2012).

2. Majidian, P., Tabatabaei, M., Zeinolabedini, M., Naghshbandi, M. P. & Chisti, Y. Metabolic engineering of microorganisms for biofuel production. *Renewable and Sustainable Energy Reviews* (2018). doi:10.1016/j.rser.2017.10.085

3. Wurtzel, E. T. *et al.* Revolutionizing agriculture with synthetic biology. *Nat. Plants* (2019). doi:10.1038/s41477-019-0539-0

4. Tyagi, A. *et al.* Synthetic Biology: Applications in the Food Sector. *Crit. Rev. Food Sci. Nutr.* (2016). doi:10.1080/10408398.2013.782534

5. Jain, K. K. Synthetic biology and personalized medicine. *Medical Principles and Practice* (2013). doi:10.1159/000341794

6. Endy, D. Foundations for engineering biology. *Nature* (2005). doi:10.1038/nature04342

7. Røkke, G., Korvald, E., Pahr, J., Øyås, O. & Lale, R. BioBrick assembly standards and techniques and associated software tools. *Methods Mol. Biol.* **1116**, 1–24 (2014).

8. Wu, G. *et al.* SGDB: A database of synthetic genes re-designed for optimizing protein over-expression. *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkl648

9. Wikmark, O.-G. et al. Synthetic biology: biosafety and contribution to addressing societal challenges. *GenØk Cent. Biosaf. Norw.* 1–69 (2016). doi:10.13140/RG.2.2.29987.25121

10. Schwille, P. Jump-starting life? Fundamental aspects of synthetic biology. *J. Cell Biol.* **210**, 687–690 (2015).

11. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32**, 347–55 (2014).

12. Cheng, A. A. & Lu, T. K. Synthetic Biology : An Emerging Engineering Discipline. *Annu. Rev. Biomed. Eng.* **14**, 155–178 (2012).

13. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* (2006). doi:10.1038/nature04640

14. Schwille, P. Bottom-up synthetic biology: Engineering in a Tinkerer's World. *Science* (2011). doi:10.1126/science.1211701

15. Xu, C., Hu, S. & Chen, X. Artificial cells: from basic science to applications. *Materials Today* (2016). doi:10.1016/j.mattod.2016.02.020

16. Elani, Y. *et al.* Constructing vesicle-based artificial cells with embedded living cells as organelle-like modules. *Sci. Rep.* (2018). doi:10.1038/s41598-018-22263-3

17. Kurihara, K. *et al.* Self-reproduction of supramolecular giant vesicles combined with the amplification of encapsulated DNA. *Nat. Chem.* (2011). doi:10.1038/nchem.1127

18. Khalil, A. S. & Collins, J. J. Synthetic biology: Applications come of age. *Nature Reviews Genetics* (2010). doi:10.1038/nrg2775

19. Tien Lea, D., Duc Chua, H. & Quynh Lea, N. Improving Nutritional Quality of Plant Proteins Through Genetic Engineering. *Curr. Genomics* (2016). doi:10.2174/1389202917666160202215934

20. Azadi, H. *et al.* Genetically Modified Crops: Towards Agricultural Growth, Agricultural Development, or Agricultural Sustainability? *Food Rev. Int.* (2015). doi:10.1080/87559129.2014.994816

21. Armbruster, B. N., Li, X., Pausch, M. H., Herlitze, S. & Roth, B. L. Evolving the lock to fit the key to create a family of G protein-coupled receptors

potently activated by an inert ligand. *Proc. Natl. Acad. Sci. U. S. A.* (2007). doi:10.1073/pnas.0700293104

22. Koder, R. L. *et al.* Design and engineering of an O2 transport protein. *Nature* (2009). doi:10.1038/nature07841

23. Corredig, M. Advanced Dairy Chemistry. Volume 1. Dairy Proteins. *Trends Food Sci. Technol.* (2004). doi:10.1016/j.tifs.2003.10.003

24. Wiltschi, B. *et al.* Enzymes revolutionize the bioproduction of value-added compounds: From enzyme discovery to special applications. *Biotechnology Advances* (2020). doi:10.1016/j.biotechadv.2020.107520

25. Menezes, A. A., Cumbers, J., Hogan, J. A. & Arkin, A. P. Towards synthetic biological approaches to resource utilization on space missions. *J. R. Soc. Interface* (2015). doi:10.1098/rsif.2014.0715

26. Verseux, C. N., Paulino-Lima, I. G., Baqué, M., Billi, D. & Rothschild, L. J. Synthetic biology for space exploration: Promises and societal implications. in *Ambivalences of Creating Life: Societal and Philosophical Dimensions of Synthetic Biology* (2015). doi:10.1007/978-3-319-21088-9_4

27. Mak, W. S. *et al.* Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat. Commun.* (2015). doi:10.1038/ncomms10005

28. Zhang, W. H., Otting, G. & Jackson, C. J. Protein engineering with unnatural amino acids. *Current Opinion in Structural Biology* (2013). doi:10.1016/j.sbi.2013.06.009

29. Link, A. J., Mock, M. L. & Tirrell, D. A. Non-canonical amino acids in protein engineering. *Current Opinion in Biotechnology* (2003). doi:10.1016/j.copbio.2003.10.011

30. Walter, K. U., Vamvaca, K. & Hilvert, D. An active enzyme constructed from

a 9-amino acid alphabet. *J. Biol. Chem.* (2005). doi:10.1074/jbc.M507210200

31. Church, G. M., Gao, Y. & Kosuri, S. Next-generation digital information storage in DNA. *Science* (2012). doi:10.1126/science.1226355

32. Malyshev, D. A. *et al.* A semi-synthetic organism with an expanded genetic alphabet. *Nature* (2014). doi:10.1038/nature13314

33. Belkin, S. *et al.* Remote detection of buried landmines using a bacterial sensor. *Nature Biotechnology* (2017). doi:10.1038/nbt.3791

34. Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics* (2001). doi:10.1038/35103511

35. Purcell, O. & Lu, T. K. Synthetic analog and digital circuits for cellular computation and memory. *Current Opinion in Biotechnology* (2014). doi:10.1016/j.copbio.2014.04.009

36. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science (80-.).* (2011). doi:10.1126/science.1205527

37. Stano, P. & Mavelli, F. Protocells Models in Origin of Life and Synthetic Biology. *Life* **5**, 1700–1702 (2015).

38. Chang, T. M. S. ARTIFICIAL CELL evolves into nanomedicine, biotherapeutics, blood substitutes, drug delivery, enzyme/gene therapy, cancer therapy, cell/stem cell therapy, nanoparticles, liposomes, bioencapsulation, replicating synthetic cells, cell encapsulation/scaffold. *Artificial Cells, Nanomedicine and Biotechnology* (2019). doi:10.1080/21691401.2019.1577885

39. Elani, Y., Law, R. V. & Ces, O. Protein synthesis in artificial cells: Using compartmentalisation for spatial organisation in vesicle bioreactors. *Phys. Chem. Chem. Phys.* (2015). doi:10.1039/c4cp05933f

40. Pohorille, A. & Deamer, D. Artificial cells: Prospects for biotechnology. *Trends in Biotechnology* (2002). doi:10.1016/S0167-7799(02)01909-1

41. Deamer, D. A giant step towards artificial life? *Trends in Biotechnology* (2005). doi:10.1016/j.tibtech.2005.05.008

42. Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science (80-.).* (2010). doi:10.1126/science.1190719

43. Sleator, R. D. JCVI-syn3.0 – A synthetic genome stripped bare! *Bioengineered* (2016). doi:10.1080/21655979.2016.1175847

44. Fredens, J. *et al.* Total synthesis of Escherichia coli with a recoded genome. *Nature* (2019). doi:10.1038/s41586-019-1192-5

45. Szostak, J. W., Bartel, D. P. & Luisi, P. L. Synthesizing life. *Nature* (2001). doi:10.1038/35053176

46. Noireaux, V., Maeda, Y. T. & Libchaber, A. Development of an artificial cell, from self-organization to computation and self-reproduction. *Proceedings of the National Academy of Sciences of the United States of America* (2011). doi:10.1073/pnas.1017075108

47. Bhattacharya, A., Brea, R. J. & Devaraj, N. K. De novo vesicle formation and growth: An integrative approach to artificial cells. *Chemical Science* (2017). doi:10.1039/c7sc02339a

48. Xu, B. Y., Xu, J. & Yomo, T. A protocell with fusion and division. *Biochemical Society Transactions* (2019). doi:10.1042/BST20190576

49. Berhanu, S., Ueda, T. & Kuruma, Y. Artificial photosynthetic cell producing energy for protein synthesis. *Nat. Commun.* (2019). doi:10.1038/s41467-019-09147-4

50. Gilbert, W. Origin of life: The RNA world. *Nature* (1986). doi:10.1038/319618a0

51. Zu, C. & Wang, J. Tumor-colonizing bacteria: A potential tumor targeting therapy. *Critical Reviews in Microbiology* (2014). doi:10.3109/1040841X.2013.776511

52. Gujrati, V. *et al.* Bioengineered bacterial outer membrane vesicles as cell-specific drug-delivery vehicles for cancer therapy. *ACS Nano* (2014). doi:10.1021/nn405724x

53. Krinsky, N. *et al.* Synthetic Cells Synthesize Therapeutic Proteins inside Tumors. *Adv. Healthc. Mater.* (2018). doi:10.1002/adhm.201701163

54. Park, T. S., Rosenberg, S. A. & Morgan, R. A. Treating cancer with genetically engineered T cells. *Trends in Biotechnology* (2011). doi:10.1016/j.tibtech.2011.04.009

55. Villarreal, F. & Tan, C. Cell-free systems in the new age of synthetic biology. *Frontiers of Chemical Science and Engineering* **11**, 58–65 (2017).

56. Perez, J. G., Stark, J. C. & Jewett, M. C. Cell-free synthetic biology: Engineering beyond the cell. *Cold Spring Harb. Perspect. Biol.* **8**, a023853 (2016).

57. Smith, M. T., Bennett, A. M., Hunt, J. M. & Bundy, B. C. Creating a completely 'cell-free' system for protein synthesis. *Biotechnol. Prog.* **31**, 1716–1719 (2015).

58. Smith, M. T., Wilding, K. M., Hunt, J. M., Bennett, A. M. & Bundy, B. C. The emerging age of cell-free synthetic biology. *FEBS Letters* **588**, 2755–2761 (2014).

59. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–5 (2001).

60. Kanter, G. *et al.* Cell-free production of scFv fusion proteins: An efficient approach for personalized lymphoma vaccines. *Blood* **109**, 3393–3399 (2007).

61. Boyer, M. E. Cell-Free Synthesis and Maturation of [FeFe] Hydrogenases.

Biotechnol. Bioeng. **99**, 59–67 (2008).

62. Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V. & Murray, R. M. Linear DNA for rapid prototyping of synthetic biological circuits in an. *ACS Synth. Biol* **3**, 387–397 (2014).

63. Albayrak, C. & Swartz, J. R. Cell-free co-production of an orthogonal transfer RNA activates efficient site-specific non-natural amino acid incorporation. *Nucleic Acids Res.* **41**, 5949–5963 (2013).

64. Bundy, B. C., Franciszkowicz, M. J. & Swartz, J. R. Escherichia coli-based cell-free synthesis of virus-like particles. *Biotechnol. Bioeng.* **100**, 28–37 (2008).

65. Zhu, Z., Kin Tam, T., Sun, F., You, C. & Percival Zhang, Y. H. A high-energy-density sugar biobattery based on a synthetic enzymatic pathway. *Nat. Commun.* (2014). doi:10.1038/ncomms4026

66. Zhang, Y. H. P., Evans, B. R., Mielenz, J. R., Hopkins, R. C. & Adams, M. W. W. High-yield hydrogen production from starch and water by a synthetic enzymatic pathway. *PLoS One* (2007). doi:10.1371/journal.pone.0000456

67. Kay, J. E. & Jewett, M. C. A cell-free system for production of 2,3-butanediol is robust to growth-toxic compounds. *Metab. Eng. Commun.* (2020). doi:10.1016/j.mec.2019.e00114

68. Ohno-iwashita, Y., Oshima, T. & Imahori, K. Protein synthesis in a cell-free system from an extreme thermophile: Effects of preincubation in the cold on polyuridylic acid-dependent polyphenylalanine synthesis at high temperature. *J. Biochem.* (1976). doi:10.1093/oxfordjournals.jbchem.a131178

69. Caschera, F. & Noireaux, V. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie* **99**, 162–168 (2014).

70. Caschera, F. & Noireaux, V. Integration of biological parts toward the

synthesis of a minimal cell. *Curr. Opin. Chem. Biol.* **22**, 85–91 (2014).

71. Kim, D. M. & Swartz, J. R. Prolonging cell-free protein synthesis with a novel ATP regeneration system. *Biotechnol. Bioeng.* (1999). doi:10.1002/(SICI)1097-0290(1999)66:3<180::AID-BIT6>3.0.CO;2-S

72. Kim, T. W. *et al.* An economical and highly productive cell-free protein synthesis system utilizing fructose-1,6-bisphosphate as an energy source. *J. Biotechnol.* (2007). doi:10.1016/j.jbiotec.2007.05.002

73. Didovyk, A., Tonooka, T., Tsimring, L. & Hasty, J. Rapid and Scalable Preparation of Bacterial Lysates for Cell-Free Gene Expression. *ACS Synth. Biol.* (2017). doi:10.1021/acssynbio.7b00253

74. Fujiwara, K. & Doi, N. Biochemical preparation of cell extract for cell-free protein synthesis without physical disruption. *PLoS One* (2016). doi:10.1371/journal.pone.0154614

75. Shimizu, Y., Kuruma, Y., Kanamori, T. & Ueda, T. The PURE system for protein production. *Methods Mol. Biol.* **1118**, 275–284 (2014).

76. Kuruma, Y. & Ueda, T. The PURE system for the cell-free synthesis of membrane proteins. *Nat. Protoc.* (2015). doi:10.1038/nprot.2015.082

77. Kazuta, Y., Matsuura, T., Ichihashi, N. & Yomo, T. Synthesis of milligram quantities of proteins using a reconstituted in vitro protein synthesis system. *J. Biosci. Bioeng.* (2014). doi:10.1016/j.jbiosc.2014.04.019

78. Tae-Wan Kim, Jung-Won Keum, In-Seok Oh, Cha-Yong Choi, Chang-Gil Park, D.-M. K. Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. *J. Biotechnol.* **126**, 554–561 (2006).

79. Jackson, R. J. & Hunt, T. Preparation and use of nuclease-treated rabbit reticulocyte lysates for the translation of eukaryotic messenger RNA. *Methods Enzymol.* (1983). doi:10.1016/S0076-6879(83)96008-1

80. Anderson, C. W., Straus, J. W. & Dudock, B. S. Preparation of a cell-free protein-synthesizing system from wheat germ. *Methods Enzymol.* (1983). doi:10.1016/0076-6879(83)01044-7

81. Yadavalli, R. & Sam-Yellowe, T. Hela based cell free expression systems for expression of Plasmodium rhoptry proteins. *J. Vis. Exp.* (2015). doi:10.3791/52772

82. Sun, Z. Z. *et al.* Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *J. Vis. Exp.* e50762 (2013). doi:10.3791/50762

83. Wang, X., Zhao, L. & Zhao, K. N. An optimized yeast cell-free lysate system for in vitro translation of human virus mRNA. *Methods Mol. Biol.* (2014). doi:10.1007/978-1-62703-782-2_14

84. Ezure, T., Suzuki, T. & Ando, E. A cell-free protein synthesis system from insect cells. *Methods Mol. Biol.* (2014). doi:10.1007/978-1-62703-782-2_20

85. Buntru, M., Vogel, S., Spiegel, H. & Schillberg, S. Tobacco BY-2 cell-free lysate: An alternative and highly-productive plant-based in vitro translation system. *BMC Biotechnol.* (2014). doi:10.1186/1472-6750-14-37

86. Islam, M. S., Aryasomayajula, A. & Selvaganapathy, P. R. A review on macroscale and microscale cell lysis methods. *Micromachines* (2017). doi:10.3390/mi8030083

87. Zhang, P., Zhang, G. & Wang, W. Ultrasonic treatment of biological sludge: Floc disintegration, cell lysis and inactivation. *Bioresour. Technol.* **98**, 207–210 (2007).

88. Kwon, Y.-C. & Jewett, M. C. High-throughput preparation methods of crude extract for robust cell-free protein synthesis. *Sci Rep.* **5**, (2015).

89. Johnson, B. H. & Hecht, M. H. Recombinant proteins can be isolated from E.

coli cells by repeated cycles of freezing and thawing. *Bio/Technology* (1994). doi:10.1038/nbt1294-1357

90. Fujiwara, K. & Doi, N. Biochemical preparation of cell extract for cell-free protein synthesis without physical disruption. *PLoS One* **11**, 1–15 (2016).

91. Salazar, O. & Asenjo, J. A. Enzymatic lysis of microbial cells. *Biotechnology Letters* (2007). doi:10.1007/s10529-007-9345-2

92. Balasundaram, B., Harrison, S. & Bracewell, D. G. Advances in product release strategies and impact on bioprocess design. *Trends in Biotechnology* (2009). doi:10.1016/j.tibtech.2009.04.004

93. Didovyk, A., Tonooka, T., Tsimring, L. & Hasty, J. Rapid and scalable preparation of bacterial lysates for cell-free gene expression. *ACS Synth. Biol.* acssynbio.7b00253 (2017). doi:10.1021/acssynbio.7b00253

94. Sun, Z. Z. *et al.* Protocols for Implementing an *Escherichia coli* Based TX-TL Cell-Free Expression System for Synthetic Biology. *J. Vis. Exp.* (2013). doi:10.3791/50762

95. Fujiwara, K. & Doi, N. Biochemical preparation of cell extract for cell-free protein synthesis without physical disruption. *PLoS One* **11**, (2016).

96. Li, J., Wang, H., Kwon, Y. C. & Jewett, M. C. Establishing a high yielding streptomyces-based cell-free protein synthesis system. *Biotechnol. Bioeng.* **114**, 1343–1353 (2017).

97. Shrestha, P., Holland, T. M. & Bundy, B. C. Streamlined extract preparation for Escherichia coli-based cell-free protein synthesis by sonication or bead vortex mixing. *Biotechniques* **53**, 163–174 (2012).

98. Caschera, F. & Noireaux, V. A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system. *Metab. Eng.* **27**, 29–37 (2015).

99. Kuruma, Y. & Ueda, T. The PURE system for the cell-free synthesis of membrane proteins. *Nat. Protoc.* **10**, 1328–44 (2015).

100. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* (2005). doi:10.1016/j.ymeth.2005.04.006

101. Lavickova, B. & Maerkl, S. J. A Simple, Robust, and Low-Cost Method to Produce the PURE Cell-Free System. *ACS Synthetic Biology* (2019). doi:10.1021/acssynbio.8b00427

102. Wang, P. H. *et al.* A Bifunctional Polyphosphate Kinase Driving the Regeneration of Nucleoside Triphosphate and Reconstituted Cell-Free Protein Synthesis. *ACS Synth. Biol.* (2019). doi:10.1021/acssynbio.9b00456

103. Kazuta, Y., Matsuura, T., Ichihashi, N. & Yomo, T. Synthesis of milligram quantities of proteins using a reconstituted in vitro protein synthesis system. *J. Biosci. Bioeng.* **118**, 554–557 (2014).

104. Matsuura, T., Tanimura, N., Hosoda, K., Yomo, T. & Shimizu, Y. Reaction dynamics analysis of a reconstituted Escherichia coli protein translation system by computational modeling. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1336–E1344 (2017).

105. Matsubayashi, H., Kuruma, Y. & Ueda, T. In vitro synthesis of the E. coli sec translocon from DNA. *Angew. Chemie - Int. Ed.* **53**, 7535–7538 (2014).

106. Noireaux, V., Bar-Ziv, R. & Libchaber, A. Principles of cell-free genetic circuit assembly. *Proc. Natl. Acad. Sci.* **100**, 12672–12677 (2003).

107. Noireaux, V. & Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl. Acad. Sci.* **101**, 17669–17674 (2004).

108. Ahn, J. H., Kang, T. J. & Kim, D. M. Tuning the expression level of recombinant proteins by modulating mRNA stability in a cell-free protein synthesis system. *Biotechnol. Bioeng.* **101**, 422–427 (2008).

109. Tinafar, A., Jaenes, K. & Pardee, K. Synthetic Biology Goes Cell-Free. *BMC Biology* (2019). doi:10.1186/s12915-019-0685-x

110. Bowie, J. U. *et al.* Synthetic Biochemistry: The Bio-inspired Cell-Free Approach to Commodity Chemical Production. *Trends in Biotechnology* (2020). doi:10.1016/j.tibtech.2019.12.024

111. Percival Zhang, Y. H. Production of biocommodities and bioelectricity by cell-free synthetic enzymatic pathway biotransformations: Challenges and opportunities. *Biotechnology and Bioengineering* (2010). doi:10.1002/bit.22630

112. Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V. & Murray, R. M. Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synth. Biol.* **3**, 387–397 (2014).

113. Agresti, J. J. *et al.* Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl. Acad. Sci.* **107**, 4004–4009 (2010).

114. Fujii, S., Matsuura, T., Sunami, T., Kazuta, Y. & Yomo, T. In vitro evolution of \alphadsdkdlmfcl,-hemolysin using a liposome display. *Proc. Natl. Acad. Sci.* **110**, 16796–16801 (2013).

115. NIRENBERG, M. W. & MATTHAEI, J. H. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U. S. A.* (1961). doi:10.1073/pnas.47.10.1588

116. Spirin, A. S., Baranov, V. I., Ryabova, L. A., Ovodov, S. Y. & Alakhov, Y. B. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science (80-.).* (1988). doi:10.1126/science.3055301

117. Kanter, G. *et al.* Cell-free production of scFv fusion proteins: An efficient approach for personalized lymphoma vaccines. *Blood* (2007).

doi:10.1182/blood-2006-07-030593

118. Bujara, M., Schümperli, M., Pellaux, R., Heinemann, M. & Panke, S. Optimization of a blueprint for in vitro glycolysis by metabolic real-time analysis. *Nat. Chem. Biol.* (2011). doi:10.1038/nchembio.541

119. Jin, X., Kightlinger, W. & Hong, S. H. Optimizing Cell-Free Protein Synthesis for Increased Yield and Activity of Colicins. *Methods Protoc.* (2019). doi:10.3390/mps2020028

120. Kigawa, T., Muto, Y. & Yokoyama, S. Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis. *J. Biomol. NMR* (1995). doi:10.1007/BF00211776

121. Dodevski, I., Markou, G. C. & Sarkar, C. A. Conceptual and methodological advances in cell-free directed evolution. *Current Opinion in Structural Biology* (2015). doi:10.1016/j.sbi.2015.04.008

122. Fujii, S. *et al.* Liposome display for in vitro selection and evolution of membrane proteins. *Nat. Protoc.* **9**, 1578–91 (2014).

123. Karig, D. K., Bessling, S., Thielen, P., Zhang, S. & Wolfe, J. Preservation of protein expression systems at elevated temperatures for portable therapeutic production. *J. R. Soc. Interface* (2017). doi:10.1098/rsif.2016.1039

124. Wilding, K. M., Zhao, E. L., Earl, C. C. & Bundy, B. C. Thermostable lyoprotectant-enhanced cell-free protein synthesis for on-demand endotoxin-free therapeutic production. *N. Biotechnol.* (2019). doi:10.1016/j.nbt.2019.07.004

125. Nourian, Z., Roelofsen, W. & Danelon, C. Triggered gene expression in fed-vesicle microreactors with a multifunctional membrane. *Angew. Chemie - Int. Ed.* **51**, 3114–3118 (2012).

126. Zhang, Y., Ruder, W. C. & LeDuc, P. R. Artificial cells: building bioinspired

systems using small-scale biology. *Trends Biotechnol.* **26**, 14–20 (2008).

127. Amidi, M., de Raad, M., Crommelin, D. J. A., Hennink, W. E. & Mastrobattista, E. Antigen-expressing immunostimulatory liposomes as a genetically programmable synthetic vaccine. *Syst. Synth. Biol.* **5**, 21–31 (2011).

128. Weber, W. & Fussenegger, M. Emerging biomedical applications of synthetic biology. *Nat. Rev. Genet.* **13**, 21–35 (2012).

129. Anderson, J. C., Voigt, C. A. & Arkin, A. P. Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* **3**, 133 (2007).

130. Leduc, P. R. *et al.* Towards an in vivo biologically inspired nanofactory. *Nat. Nanotechnol.* **2**, 3–7 (2007).

131. Jain, S., Jain, V. & Mahajan, S. C. Lipid Based Vesicular Drug Delivery Systems. *Adv. Pharm.* (2014). doi:10.1155/2014/574673

132. Magnani, C. *et al.* Hybrid vesicles from lipids and block copolymers: Phase behavior from the micro- to the nano-scale. *Colloids Surfaces B Biointerfaces* 1–11 (2018). doi:10.1016/j.colsurfb.2018.01.042

133. Brekke, O. H. & Sandlie, I. Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nature Reviews Drug Discovery* (2003). doi:10.1038/nrd984

134. Ledsgaard, L., Kilstrup, M., Karatt-Vellatt, A., McCafferty, J. & Laustsen, A. H. Basics of antibody phage display technology. *Toxins* (2018). doi:10.3390/toxins10060236

135. Boder, E. T. & Wittrup, K. D. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.* (1997). doi:10.1038/nbt0697-553

136. Cherf, G. M. & Cochran, J. R. Applications of yeast surface display for protein engineering. *Methods Mol. Biol.* (2015). doi:10.1007/978-1-4939-2748-7_8

137. Zahnd, C., Amstutz, P. & Plückthun, A. Ribosome display: Selecting and evolving proteins in vitro that specifically bind to a target. *Nat. Methods* (2007). doi:10.1038/nmeth1003

138. Fujii, S., Matsuura, T., Sunami, T., Kazuta, Y. & Yomo, T. In vitro evolution of α-hemolysin using a liposome display. *Proc. Natl. Acad. Sci. U. S. A.* (2013). doi:10.1073/pnas.1314585110

139. Doi, N. & Yanagawa, H. STABLE: Protein-DNA fusion system for screening of combinatorial protein libraries in vitro. *FEBS Lett.* (1999). doi:10.1016/S0014-5793(99)01041-8

140. Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T. & Yanagawa, H. DNA display for in vitro selection of diverse peptide libraries. *Nucleic Acids Res.* (2003). doi:10.1093/nar/gng119

141. Bertschinger, J. & Neri, D. Covalent DNA display as a novel tool for directed evolution of proteins in vitro. *Protein Eng. Des. Sel.* (2004). doi:10.1093/protein/gzh082

142. Yamaguchi, J. *et al.* cDNA display: A novel screening method for functional disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions. *Nucleic Acids Res.* (2009). doi:10.1093/nar/gkp514

143. Naimuddin, M. & Kubo, T. A High Performance Platform Based on cDNA Display for Efficient Synthesis of Protein Fusions and Accelerated Directed Evolution. *ACS Comb. Sci.* (2016). doi:10.1021/acscombsci.5b00139

144. Seelig, B. MRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. *Nat. Protoc.* (2011). doi:10.1038/nprot.2011.312

145. Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the in vitro selection of peptides and proteins. *Proc. Natl. Acad. Sci. U. S. A.* (1997). doi:10.1073/pnas.94.23.12297

146. Nemoto, N., Miyamoto-Sato, E., Husimi, Y. & Yanagawa, H. In vitro virus: Bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett.* (1997). doi:10.1016/S0014-5793(97)01026-0

147. Josephson, K., Ricardo, A. & Szostak, J. W. MRNA display: From basic principles to macrocycle drug discovery. *Drug Discovery Today* (2014). doi:10.1016/j.drudis.2013.10.011

148. Blanco, C., Verbanic, S., Seelig, B. & Chen, I. A. High throughput sequencing of: In vitro selections of mRNA-displayed peptides: Data analysis and applications. *Physical Chemistry Chemical Physics* (2020). doi:10.1039/c9cp05912a

149. Liu, R., Barrick, J. E., Szostak, J. W. & Roberts, R. W. [19] Optimized synthesis of RNA-protein fusions for in vitro protein selection. *Methods in Enzymology* (2000). doi:10.1016/s0076-6879(00)18058-9

150. Wang, H. & Liu, R. Advantages of mRNA display selections over other selection techniques for investigation of protein-protein interactions. *Expert Review of Proteomics* (2011). doi:10.1586/epr.11.15

151. Newton, M. S., Cabezas-Perusse, Y., Tong, C. L. & Seelig, B. In Vitro Selection of Peptides and Proteins-Advantages of mRNA Display. *ACS Synthetic Biology* (2020). doi:10.1021/acssynbio.9b00419

152. Fukuda, I. *et al.* In vitro evolution of single-chain antibodies using mRNA display. *Nucleic Acids Res.* (2006). doi:10.1093/nar/gkl618

153. Dodevski, I., Markou, G. C. & Sarkar, C. A. Conceptual and methodological advances in cell-free directed evolution. *Current Opinion in Structural Biology* (2015). doi:10.1016/j.sbi.2015.04.008

154. Contreras-Llano, L. E. & Tan, C. High-throughput screening of biomolecules using cell-free gene expression systems. *Synth. Biol.* (2018).

doi:10.1093/synbio/ysy012

155. Hino, M. *et al.* Efficiency of cell-free protein synthesis based on a crude cell extract from Escherichia coli, wheat germ, and rabbit reticulocytes. *J. Biotechnol.* (2008). doi:10.1016/j.jbiotec.2007.08.008

156. Shin, J. & Noireaux, V. Study of messenger RNA inactivation and protein degradation in an Escherichia coli cell-free expression system. *J. Biol. Eng.* (2010). doi:10.1186/1754-1611-4-9

157. Opyrchal, M., Anderson, J. R., Sokoloski, K. J., Wilusz, C. J. & Wilusz, J. A cell-free mRNA stability assay reveals conservation of the enzymes and mechanisms of mRNA decay between mosquito and mammalian cell lines. *Insect Biochem. Mol. Biol.* (2005). doi:10.1016/j.ibmb.2005.08.004

158. Newton, M. S., Cabezas-Perusse, Y., Tong, C. L. & Seelig, B. In Vitro Selection of Peptides and Proteins - Advantages of mRNA Display. *ACS Synthetic Biology* (2020). doi:10.1021/acssynbio.9b00419

159. Kanamori, T., Fujino, Y. & Ueda, T. PURE ribosome display and its application in antibody technology. *Biochimica et Biophysica Acta - Proteins and Proteomics* (2014). doi:10.1016/j.bbapap.2014.04.007

160. Nagumo, Y., Fujiwara, K., Horisawa, K., Yanagawa, H. & Doi, N. PURE mRNA display for in vitro selection of single-chain antibodies. *J. Biochem.* (2016). doi:10.1093/jb/mvv131

161. Bashiruddin, N. K. & Suga, H. Construction and screening of vast libraries of natural product-like macrocyclic peptides using in vitro display technologies. *Current Opinion in Chemical Biology* (2015). doi:10.1016/j.cbpa.2014.11.011

162. Ueno, S. & Nemoto, N. CDNA display: Rapid stabilization of mrna display. *Methods Mol. Biol.* (2012). doi:10.1007/978-1-61779-379-0_8

163. Ong, Y. S. *et al.* Recent Advances in Synthesis and Identification of Cyclic

Peptides for Bioapplications. *Curr. Top. Med. Chem.* (2017). doi:10.2174/1568026617666170224121658

164. Kunamneni, A., Ye, C., Bradfute, S. B. & Durvasula, R. Ribosome display for the rapid generation of high-affinity Zika-neutralizing single-chain antibodies. *PLoS One* (2018). doi:10.1371/journal.pone.0205743

165. Ledsgaard, L., Kilstrup, M., Karatt-Vellatt, A., McCafferty, J. & Laustsen, A. H. Basics of antibody phage display technology. *Toxins* (2018). doi:10.3390/toxins10060236

166. Deutscher, S. L. Phage display in molecular imaging and diagnosis of cancer. *Chem. Rev.* (2010). doi:10.1021/cr900317f

167. Popkov, M., Rader, C. & Barbas, C. F. Isolation of human prostate cancer cell reactive antibodies using phage display technology. *J. Immunol. Methods* (2004). doi:10.1016/j.jim.2004.05.004

168. Rosowski, S. *et al.* A novel one-step approach for the construction of yeast surface display Fab antibody libraries. *Microb. Cell Fact.* (2018). doi:10.1186/s12934-017-0853-z

169. Siontorou, C. G., Nikoleli, G. P., Nikolelis, D. P. & Karapetis, S. K. Artificial lipid membranes: Past, present, and future. *Membranes* (2017). doi:10.3390/membranes7030038

170. Bochicchio, D., Panizon, E., Monticelli, L. & Rossi, G. Interaction of hydrophobic polymers with model lipid bilayers. *Sci. Rep.* (2017). doi:10.1038/s41598-017-06668-0

171. Buddingh', B. C. & Van Hest, J. C. M. Artificial Cells: Synthetic Compartments with Life-like Functionality and Adaptivity. *Acc. Chem. Res.* (2017). doi:10.1021/acs.accounts.6b00512

172. Frey, C. M., Barth, H., Kranz, C. & Mizaikoff, B. Horizontal black lipid bilayer

membranes for studying pore-forming toxins. *Anal. Methods* (2018). doi:10.1039/c8ay01122b

173. Andersson, J. & Köper, I. Tethered and polymer supported bilayer lipid membranes: Structure and function. *Membranes* (2016). doi:10.3390/membranes6020030

174. Richter, R. P., Bérat, R. & Brisson, A. R. Formation of solid-supported lipid bilayers: An integrated view. *Langmuir* (2006). doi:10.1021/la052687c

175. Akbarzadeh, A. *et al.* Liposome: Classification, preparation, and applications. *Nanoscale Res. Lett.* (2013). doi:10.1186/1556-276X-8-102

176. Deshpande, P. P., Biswas, S. & Torchilin, V. P. Current trends in the use of liposomes for tumor targeting. *Nanomedicine* (2013). doi:10.2217/nnm.13.118

177. Verchère, A., Broutin, I. & Picard, M. Reconstitution of membrane proteins in liposomes. in *Methods in Molecular Biology* (2017). doi:10.1007/978-1-4939-7151-0_14

178. Engel, A. & Gaub, H. E. Structure and Mechanics of Membrane Proteins. *Annu. Rev. Biochem.* (2008). doi:10.1146/annurev.biochem.77.062706.154450

179. Park, S. H. *et al.* Structure of the chemokine receptor CXCR1 in phospholipid bilayers. *Nature* (2012). doi:10.1038/nature11580

180. Hickey, K. D. & Buhr, M. M. Lipid Bilayer Composition Affects Transmembrane Protein Orientation and Function. *J. Lipids* (2011). doi:10.1155/2011/208457

181. Catoire, L. J., Warnet, X. L. & Warschawski, D. E. Micelles, bicelles, amphipols, nanodiscs, iposomes, or intact cells: The hitchhiker's guide to the study of membrane proteins by NMR. in *Membrane Proteins Production for Structural Analysis* (2014). doi:10.1007/978-1-4939-0662-8_12

182. Hąc-Wydro, K. & Dynarowicz-Łątka, P. Biomedical applications of the

Langmuir monolayer technique. *Ann. UMCS, Chem.* (2010). doi:10.2478/v10063-008-0027-2

183. Lee, M. K. Liposomes for enhanced bioavailability of water-insoluble drugs: In vivo evidence and recent approaches. *Pharmaceutics* (2020). doi:10.3390/pharmaceutics12030264

184. Zoghbi, M. E. & Altenberg, G. A. Membrane protein reconstitution in nanodiscs for luminescence spectroscopy studies. *Nanotechnology Reviews* (2017). doi:10.1515/ntrev-2016-0078

185. Ritchie, T. K. *et al.* Chapter 11 Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods in Enzymology* (2009). doi:10.1016/S0076-6879(09)64011-8

186. Nath, A., Atkins, W. M. & Sligar, S. G. Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins. *Biochemistry* (2007). doi:10.1021/bi602371n

187. Bayburt, T. H. & Sligar, S. G. Membrane protein assembly into Nanodiscs. *FEBS Lett.* (2010). doi:10.1016/j.febslet.2009.10.024

188. Nakano, M. *et al.* Static and dynamic properties of phospholipid bilayer nanodiscs. *J. Am. Chem. Soc.* (2009). doi:10.1021/ja9017013

189. Cho, W. & Stahelin, R. V. Membrane-protein interactions in cell signaling and membrane trafficking. *Annual Review of Biophysics and Biomolecular Structure* (2005). doi:10.1146/annurev.biophys.33.110502.133337

190. Kampen, K. R. Membrane proteins: The key players of a cancer cell. *Journal of Membrane Biology* (2011). doi:10.1007/s00232-011-9381-7

191. Leth-Larsen, R., Lund, R. R. & Ditzel, H. J. Plasma Membrane Proteomics and Its Application in Clinical Cancer Biomarker Discovery. *Mol. Cell. Proteomics* (2010). doi:10.1074/mcp.R900006-MCP200

192. Grimm, D. *et al.* Diagnostic and Therapeutic Use of Membrane Proteins in Cancer Cells. *Curr. Med. Chem.* (2011). doi:10.2174/092986711794088344

193. Hauser, A. S. *et al.* Pharmacogenomics of GPCR Drug Targets. *Cell* (2018). doi:10.1016/j.cell.2017.11.033

194. Hauser, A. S., Attwood, M. M., Rask-Andersen, M., Schiöth, H. B. & Gloriam, D. E. Trends in GPCR drug discovery: New agents, targets and indications. *Nat. Rev. Drug Discov.* (2017). doi:10.1038/nrd.2017.178

Chapter 2. Materials and Methodology

2.1 Experiments for Chapter 3

2.1.1 General Materials and Verification Procedures

Unless otherwise specified, all the reagents, bacterial growth mediums, nucleotides sequences, primers and constructs, were acquired from Thermo Fisher Scientific. The PURE*frex* 1.0 cell-free system was purchased from GeneFrontier and the Expressway system (Expressway™ Maxi Cell-Free *E. coli* Expression System) from Thermo Fisher Scientific. The Zyppy Plasmid Miniprep and ZymoPURE Plasmid Midiprep kits were acquired from Zymo Research; the Wizard SV Gel and PCR Clean-Up System were bought from Promega, and the (5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2,3-dimethyl-4H-Imidazol-4-one (DFHBI) fluorophore from Sigma Aldrich. The competent bacteria cells used were *E. coli* BL21-CodonPlus (DE3)-RIPL, from Agilent Technologies. All the PCR reactions were performed using KOD Hot Start Master Mix, by EMD Millipore.

A thermocycler (T100 Thermocycler, BioRad) was used for the incubation of all PCR reactions, which were subsequently verified for size in a 1.5% agarose Tris-Acetate-EDTA (TAE) gel (UltraPure Agarose), with a constant voltage of 100V. For staining DNA products in the agarose gels, SYBR Gold Nucleic Acid Stain was used along with 1 kb DNA Ladder (NEB) for size comparison.

DNA quantification was performed by UV-Vis spectrophotometry (Nanodrop 2000c, Thermo Fisher Scientific). All the cell-free reactions were incubated in a

plate reader (Infinite 200 PRO, Tecan) and the incubation steps for the competent bacteria cells were performed in an incubator (BioShaker BR-43FL, Taitec). Centrifugation steps were performed in a microcentrifuge (model 5415 R, Eppendorf), with different tubes adaptors depending on the size of the tubes used.

2.1.2 Plasmid Stock and Production

The plasmid sfGFP-pET32b (supplied by Yoshihiro Shimizu's laboratory group, RIKEN) containing sfGFP (Superfolder Green Fluorescent Protein) as the reporter gene, was used as the template to create all the genetic constructs. *E. coli* BL21-CodonPlus(DE3)-RIPL competent cells were transformed with the plasmid sfGFP-pET32b following the manufacturer protocol, and after, grown overnight at 37°C in LB (Lysogeny Broth) agar plates with 50 µg/ml of ampicillin, for colony selection. The positive colonies were incubated in 10 ml of 2xYT medium, with 50 µg/ml of ampicillin, overnight at 37°C and 200 rpm. The plasmid, sfGFP-pET32b, was extracted from the bacterial colonies using the Zyppy Plasmid Miniprep and ZymoPURE Plasmid Midiprep Kits, and the remaining bacterial culture used as a plasmid stock by storing the cells with 70% glycerol at -80C.

2.1.3 Linear DNA Constructs

The plasmid obtained was used to make a linear construct using the sfGFP gene along with an RBS (Ribosome Binding Site) and a T7 promoter. This was done by two-step PCR. The first one performed used the obtained sfGFP-pET32b plasmid as the DNA template, to add the RBS region using the primers RBS-sfGFP-F and sf-GFP-R (Table 2.A). The PCR conditions were as follows: Initial Denaturation (94°C,

120 sec), 35 cycles of denaturation (98°C, 10 sec) and annealing/extension (68°C, 30 sec), and Final Extension (68°C, 120 sec). The result constructs, called RBS-sfGFP, was then used as the DNA template for the second round of PCR using the primers T7-RBS-sfGFP-F and sfGFP-R (Table 2.A) with the following conditions: Initial denaturation (94°C, 120 sec), 35 cycles of denaturation (98°C, 10 sec), annealing (30°C, 30 sec) and extension (70°C, 30 sec), and final extension (70°C, 120 sec). The construct from this final PCR, T7-RBS-sfGFP, which contains the consensus T7 promoter sequence, was used throughout the experiments as the reference sequence (also named as "WT" for Wild Type).

Other linear DNA constructs containing variants of only the T7 promoter, along with consensus RBS and sfGFP gene sequences, were constructed using also a two-step PCR by following the protocol for the cell-free system, PURE*frex* 1.0: The first PCR, used the already obtained WT construct as the template DNA. A different forward primer was used for each variant (Table 2.A) but the reverse primer, sfGFP-R, was common for them all. For making all the variants, the PCR conditions were: Initial denaturation (95°C, 120 sec), 35 cycles of denaturation (95°C, 20 sec), annealing (67°C, 10 sec) and extension (70°C, 15 sec), and final extension (70°C, 120 sec).

Table 2.A: Sequences and names of the oligo primers used for creating the WT consensus and the promoter variants constructs.

Name	Sequence
T7-smut-17A	GAAATAAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-17C	GAAATCAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-17G	GAAATGAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-16C	GAAATTCATACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-16G	GAAATTGATACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-16T	GAAATTTATACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-15C	GAAATTACTACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-15G	GAAATTAGTACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-15T	GAAATTATTACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-14A	GAAATTAAAACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-14C	GAAATTAACACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-14G	GAAATTAAGACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-13C	GAAATTAATCCGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-13G	GAAATTAATGCGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-13T	GAAATTAATTCGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-12A	GAAATTAATAAGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-12G	GAAATTAATAGGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-12T	GAAATTAATATGACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-11A	GAAATTAATACAACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-11C	GAAATTAATACCACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-11T	GAAATTAATACTACTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-10C	GAAATTAATACGCCTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-10G	GAAATTAATACGGCTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-10T	GAAATTAATACGTCTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-9A	GAAATTAATACGAATCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-9G	GAAATTAATACGAGTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-9T	GAAATTAATACGATTCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-8A	GAAATTAATACGACACACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-8C	GAAATTAATACGACCCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-8G	GAAATTAATACGACGCACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-7A	GAAATTAATACGACTAACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-7G	GAAATTAATACGACTGACTATAGGGAGACCACAACGGTTTCCCTC

T7-smut-7T	GAAATTAATACGACTTACTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-6C	GAAATTAATACGACTCCCTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-6G	GAAATTAATACGACTCGCTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-6T	GAAATTAATACGACTCTCTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-5A	GAAATTAATACGACTCAATATAGGGAGACCACAACGGTTTCCCTC
T7-smut-5G	GAAATTAATACGACTCAGTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-5T	GAAATTAATACGACTCATTATAGGGAGACCACAACGGTTTCCCTC
T7-smut-4A	GAAATTAATACGACTCACAATAGGGAGACCACAACGGTTTCCCTC
T7-smut-4C	GAAATTAATACGACTCACCATAGGGAGACCACAACGGTTTCCCTC
T7-smut-4G	GAAATTAATACGACTCACGATAGGGAGACCACAACGGTTTCCCTC
T7-smut-3C	GAAATTAATACGACTCACTCTAGGGAGACCACAACGGTTTCCCTC
T7-smut-3G	GAAATTAATACGACTCACTGTAGGGAGACCACAACGGTTTCCCTC
T7-smut-3T	GAAATTAATACGACTCACTTTAGGGAGACCACAACGGTTTCCCTC
T7-smut-2A	GAAATTAATACGACTCACTAAAGGGAGACCACAACGGTTTCCCTC
T7-smut-2C	GAAATTAATACGACTCACTACAGGGAGACCACAACGGTTTCCCTC
T7-smut-2G	GAAATTAATACGACTCACTAGAGGGAGACCACAACGGTTTCCCTC
T7-smut-1C	GAAATTAATACGACTCACTATCGGGAGACCACAACGGTTTCCCTC
T7-smut-1G	GAAATTAATACGACTCACTATGGGGAGACCACAACGGTTTCCCTC
T7-smut-1T	GAAATTAATACGACTCACTATTGGGAGACCACAACGGTTTCCCTC
RBS-sfGFP-F	AAGGAGATATACCAATGAGTAAAGGAGAAGAACTTTTCACTGGAG TTGTCC
T7-RBS-sfGFP-F	GAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCT AGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCA
sf-GFP-R	AAAGAGGAGAAATACTAGATGAGTAAAGGAGAAGAACTTTTCACT GGAGTTGTC

Using the original WT DNA construct, along with all its variations, extra bases were added at both the 5' end and the 3' end. The sequence attached at the 5' end was a non-coding random sequence, and that at the 3' end contained a T7 terminator sequence. Specific forward primers for each variant were used along with a common reverse primer, 5p3p_Common_R (Table 2.B). The PCR conditions were: Initial denaturation (95°C, 120 sec), 40 cycles of denaturation (95°C, 20 sec),

annealing (56°C, 10 sec) and extension (70°C, 30 sec), and final extension (70°C, 10 min). For the WT sample with extra bases only at the 5' end, its specific forward primer was used as described above with the difference of the reverse primer, which was 5p_Common2_R (Table 2.B). The PCR conditions were the same as those for the 5' and 3' extra bases sequences.

Table 2.B: Sequences and names of the oligo primers used for creating the WT consensus and the promoter variants construct with additional bases at 3' and 5' and the ones with only additional bases at 5'. Also included are the oligo primers used for creating the RNA spinach aptamer constructs.

Name	Sequences
5p3p_WT_F	TTCGTTTTATTTGATGCCGAAATTAATACGACTCACTATAGGG
5p3p_17A_F	TTCGTTTTATTTGATGCCGAAATAAATACGACTCACTATAGGG
5p3p_17C_F	TTCGTTTTATTTGATGCCGAAATCAATACGACTCACTATAGG
5p3p_17G_F	TTCGTTTTATTTGATGCCGAAATGAATACGACTCACTATAGGG
5p3p_16C_F	TTCGTTTTATTTGATGCCGAAATTCATACGACTCACTATAGG
5p3p_16G_F	TTCGTTTTATTTGATGCCGAAATTGATACGACTCACTATAGGG
5p3p_16T_F	TTCGTTTTATTTGATGCCGAAATTTATACGACTCACTATAGGG
5p3p_15C_F	TTCGTTTTATTTGATGCCGAAATTACTACGACTCACTATAGGG
5p3p_15G_F	TTCGTTTTATTTGATGCCGAAATTAGTACGACTCACTATAGGG
5p3p_15T_F	TTCGTTTTATTTGATGCCGAAATTATTACGACTCACTATAGGG
5p3p_14A_F	TTCGTTTTATTTGATGCCGAAATTAAAACGACTCACTATAGGG
5p3p_14C_F	TTCGTTTTATTTGATGCCGAAATTAACACGACTCACTATAGGG
5p3p_14G_F	TTCGTTTTATTTGATGCCGAAATTAAGACGACTCACTATAGGG
5p3p_13C_F	TTCGTTTTATTTGATGCCGAAATTAATCCGACTCACTATAGGG
5p3p_13G_F	TTCGTTTTATTTGATGCCGAAATTAATGCGACTCACTATAGGG
5p3p_13T_F	TTCGTTTTATTTGATGCCGAAATTAATTCGACTCACTATAGGG
5p3p_12A_F	TTCGTTTTATTTGATGCCGAAATTAATAAGACTCACTATAGGG
5p3p_12G_F	TTCGTTTTATTTGATGCCGAAATTAATAGGACTCACTATAGGG
5p3p_12T_F	TTCGTTTTATTTGATGCCGAAATTAATATGACTCACTATAGGG
5p3p_11A_F	TTCGTTTTATTTGATGCCGAAATTAATACAACTCACTATAGGG
5p3p_11C_F	TTCGTTTTATTTGATGCCGAAATTAATACCACTCACTATAGGG

5p3p_11T_F	TTCGTTTTATTTGATGCCGAAATTAATACTACTCACTATAGGG
5p3p_10C_F	TTCGTTTTATTTGATGCCGAAATTAATACGCCTCACTATAGGG
5p3p_10G_F	TTCGTTTTATTTGATGCCGAAATTAATACGGCTCACTATAGGG
5p3p_10T_F	TTCGTTTTATTTGATGCCGAAATTAATACGTCTCACTATAGGG
5p3p_Common_R	TCAACAGGAGTCCAAGCAAAAAACCCCTCAAGACCCGTTTAGAGGC CCCAAGGGGTTATGCTAGTATAAACGCAGAAAGGCC
5p_Common2_R	TATAAACGCAGAAAGGCCCACCC
Aptamer1-F	GGGAGACCACAACGGTTTCCCTCGCCCGGATAGCTCAGTCGGTAGA GC
Aptamer2-F	CGGATAGCTCAGTCGGTAGAGCAGCGGCCGGACGCAACTGAATGA AATGGTGAAGGACGGGTCCAGGTGTGGCTGCTTCG
Aptamer3-R	GGACGCGACTAGTTACGGAGCTCACACTCTACTCAACAAGCTGCAC TGCCGAAGCAGCCACACCTGGACCC
Aptamer4-R	TGGCGCCCGAACAGGGACTTGAACCCTGGACCCGCGGCCGGACGCG ACTAGTTACGGAGCT

2.1.4 Spinach RNA Aptamer Constructs

Several linear DNA templates, containing the spinach aptamer, were constructed. A first PCR was performed using four primers at the same time: Aptamer1-F, Aptamer2-F, Aptamer3-R and Aptamer4-R (Table 2.B), to create the RNA aptamer and tRNA scaffold sequence. The PCR conditions were: Initial denaturation (95°C, 120 sec), 40 cycles of denaturation (95°C, 20 sec), annealing phase (62°C, 10 sec) and extension (70°C, 5 sec), and final extension (70°C, 10 min).

The PCR product obtained was purified and immediately used as a template for a second PCR to attach to it the promoter region of some of the constructs. The forward primers that were used to make the initial set of constructs (Table 1) and a common reverse primer, Aptamer4-R (Table 2), were used for this second PCR. The following conditions were used: Initial denaturation (95°C, 120 sec), 40 cycles

of denaturation (95°C, 20 sec), annealing phase (57°C, 10 sec) and extension (70°C, 5 sec), and final extension (70°C, 10 min). The forward primers used were the ones for the WT construct and for the 17A, 17C, 16C, 16T and 11A variants.

2.1.5 Cell-Free Transcription-Translation Reaction

Unless otherwise indicated, all the reactions using the PURE system were performed with a final volume of 20 µl and using 24 ng (1 µl) of each linear DNA construct. For the reactions using the Expressway system, the total volume of each reaction was 26.6 µl with 235 ng (5 µl) of linear DNA. Stocks of DNA samples were made for accurate reproducibility and all experiments were performed in duplicate. The reaction components for both systems were assembled in a master mix for each one and then the specifically required volume added to the wells of a black flat-bottom 384 well-plate (Nunc 384 black well-plate) along with their corresponding amount of linear DNA template. The well-plates were covered with a transparent sealing membrane (Breath-Easy, Sigma-Aldrich) to avoid evaporation and afterwards incubated in a plate reader at 37°C for 12-20 h. During the incubation period, GFP fluorescence was measured and recorded every 15 min (excitation: 395 nm; emission: 509 nm).

2.1.6 RNA Spinach Aptamer Cell-Free Transcription Reaction

For the RNA spinach aptamer experiments, the reactions with the PURE and Expressway systems were performed the same as stipulated before and following the same conditions, but DFHBI fluorophore was added to each reaction to a final

concentration of 20 μM and its fluorescence measured instead (excitation: 460 nm; emission: 502 nm).

2.1.7 Mathematical Analysis and Modelling

A computer simulation was performed using custom programs written in Python and R[1]. For numerical simulation of differential equations, SciPy module[2] was used. Nonlinear fitting of experimental data for the estimation of system parameters in the differential equations was performed by least-square fitting using leastsq function in the SciPy module. The rate of reaction and the maximum expression level for each time-course fluorescence data were calculated in R using growthcurver package [3].

2.2 Experiments for Chapter 4

2.2.1 General Materials and Verification Procedures

Unless otherwise specified, all the reagents were acquired from Thermo Fisher Scientific. All the oligonucleotide sequences were ordered from FASMAC Inc., (Table 2.C). The DNA Polymerase I, Klenow Fragment (Large Fragment of the *E. coli* DNA Polymerase I) and PrimeScript One Step RT-PCR Kit Ver.2, were bought from Takara Bio; the Wizard SV Gel and PCR Clean-Up System from Promega; the HiScribe™ T7 Quick High Yield RNA Synthesis Kit and all the T4 DNA Ligase products from New England Biolabs (NEB); and both the NucleoSpin® RNA Clean-up and the NucleoSpin RNA Clean-up XS kits, from Macherey-Nagel. The PURE*frex* 1.0 cell-free system was purchased from GeneFrontier.

All DNA library samples transcribed mRNA libraries and their FITC-puromycin-FITC DNA tag ligated products (mRNA-tag) were confirmed for the correct size before proceeding to the next steps. Unless otherwise stated, the DNA samples were run on a 2% agarose Tris-Acetate-EDTA gel (UltraPure Agarose), with a constant voltage of 100V; the transcribed mRNA (and mRNA-tag) samples were resolved using a mini size (8 ml) 8 M Urea 6% polyacrylamide Tris-borate-EDTA (TBE) gel at constant 50 mA; and the translated mRNA-peptide conjugate samples in a polyacrylamide SDS-PAGE gel, with a 3.5% stacking gel and a 10% resolving gel containing 8 M urea, at a constant 50 mA. For detecting both DNA and mRNA products, SYBR Gold Nucleic Acid Stain was used and the gels were visualised in a transilluminator (FUJIFILM Wako Pure Chemical). The electroelution processes were performed in an Electro-Eluter by Bio-Rad (model 422) along with all its components.

The Western-blot reactions performed were done using polyvinylidene fluoride (PVDF) membranes (Bio-Rad), and a Trans-Blot SD Cell (Bio-Rad) apparatus in semi-dry conditions at 110 mA for 60 minutes. The blocking solution used was PVDF Blocking Reagent for Can Get Signal (Toyobo), and the incubation reagent was Can Get Signal, solution 2 (Toyobo). The membrane was revealed with chemiluminescence using ECL Select Western Blotting Detection Reagent (GE Healthcare).

The peptide conjugate selection process was performed using Anti-FLAG M2 Magnetic Beads and a DynaMag-2 magnetic stand, bought from Sigma-Aldrich and

Thermo Fisher Scientific respectively. All the incubation procedures were done in a circular rotator (Labinco L28, Salford Scientific Supplies).

For next-generation sequencing (NGS), Illumina MiSeq was used. Quantification for sequencing was carried out using Qubit 2.0 Fluorometer and Qubit 1X dsDNA HS Assay Kit (Thermo Fisher Scientific). The sequencing products used were NEBNext Ultra II DNA Library Prep with Sample Purification Beads and NEBNext Multiplex Oligos for Illumina (Index Primers Set 1) (NEB). Agilent 2100 Bioanalyzer (Agilent Technologies) was used for length verification and the final samples run with the Miseq Reagent Kit v3 600 cycles.

Table 2.C: Sequences and names of all the oligonucleotides for the DNA libraries' construction, DNA-tag ligation, reverse transcription and real-time PCR procedures. Nomenclature is as follows: N = A, C, G and T; B = C, G and T; R = A and G; Y = C and T; p = 5'-phosphorylation; Spacer 18 = Hexaethylene glycol; F-dT = Fluorescein-dT and Puro = Puromycin.

Name	Sequence (5' to 3')
Display-F	GTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAAT AATTTTGTTTAACTTTAAGAAGGAGATATACCAATGGAC
Display-10aa-random-R	TTTTTCACCTGATCCGCTGCCTTTCTGTTTACCNNBRNNNNBNNBRNN NNBNNBRNNNNBNNBGTCCATTGGTATATCTCC
Display-FLAG-random-R	TTTTTCACCTGATCCGCTGCCTTTCTGTTTACCNYYNYYNYYGTCGTCC TTGTAGTCCATTGGTATATCTCC
FLAG-control-R	TTTTTCACCTGATCCGCTGCCTTTCTGTTTACCCTTGTCATCGTCGTCCT TGTAGTCCATTGGTATATCTCC
RTPCR-F	GTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTC
RTPCR-R	TTTTTCACCTGATCCGCTGCCTTTCTG
Puromycin-DNA tag	pCCCTTCACCTGATCCGCTGAAAAAAAAAAAAAAAAA(Spacer 18) (Spacer 18) (F-dT) (Spacer 18) CC (Puro)

DNA quantification was performed by UV-Vis spectrophotometry (Nanodrop 2000c, Thermo Fisher Scientific). All the cell-free reactions were incubated in a thermomixer (Eppendorf Thermomixer R, Eppendorf). Centrifugation steps were performed in a microcentrifuge (model 5415 R, Eppendorf), with different tubes adaptors depending on the size of the tubes used.

2.2.2 Linear DNA Library Construction

Linear DNA libraries were constructed by annealing and extension using the DNA Polymerase I, Klenow Fragment Kit and 3 µMof each of two single-stranded DNA (ssDNA) oligonucleotides. In the making of the two original libraries, the Display-F and Display-FLAG-random-R oligos were used for the FLAG-random library while Display-F and Display-10aa-random-R oligos were used for the 10aa-random library. To make the fixed FLAG (DYKDDDDK) sequence, the oligos used were Display-F and FLAG-control-R. For all these reactions, each oligo pair was mixed with a final concentration of 200 µM dNTPs in Klenow buffer, and annealed by heating at 92°C for 30 seconds followed by cooling to room temperature. Then 10U of the Klenow Fragment was added to the reaction mix and an extension step was performed at 37°C for 1 hour, followed by inactivation of the enzyme at 50°C for 15 min. Both DNA libraries (FLAG-random and 10aa-random) were further purified by column using the Wizard SV Gel and PCR Clean-Up System and quantified by Nanodrop.

2.2.3 *In vitro* Transcription and DNA-tag Ligation

In vitro transcription was performed using HiScribe™ T7 Quick High Yield RNA Synthesis Kit with approximately 500 ng (11 pmol, 6.6 x 10^{12} molecules) of each linear DNA library at 37°C for 6 h according to the kit's protocol. The transcribed product was cleaned up afterwards using the NucleoSpin® RNA Clean-up kit and measured with Nanodrop. To anneal the transcribed mRNA library with a puromycin-FITC DNA tag, first, a 100 µL reaction was set up consisting of final concentrations of 4 µM transcribed mRNA, 6 µM puromycin-FITC DNA tag (Table 2.C) and 1 mM ATP in T4 ligation buffer. Second, the reaction was incubated at 90°C for 30 sec and cooled to room temperature with a 1°C/sec decline rate to correctly anneal the mRNA library with the puromycin-FITC DNA tag. And third, 3U of T4 PNK and 20U of T4 RNA ligase were added to the reaction mixture and further incubated at 25°C for 30 min.

2.2.4 Sample Product Purification: Electroelution and Ethanol Precipitation

The ligation product, mRNA-puromycin-FITC (mRNA-tag), was run on a 40 ml (8M Urea 6% polyacrylamide) TBE gel for purification before being used for *in vitro* translation. Using a clean scalpel and under a blue light LED transilluminator, the strips containing the mRNA-tag were cut from the Urea-TBE gel and followed by standard electroelution gel purification. First, the cut gel's strips were crushed into small pieces and placed inside the Electro-Eluter tubes (Bio-Rad) with a membrane cap (12kDa cut-off) attached to the tubes. The tubes were filled with 1x TBE buffer

and run at 10 mA/tube for 30 min. Then, the polarity was reversed for 1 min to release the mRNA-tag from the membrane. The eluted sample solution was recovered and subjected to downstream ethanol precipitation, in which 10% sample volume of 3 M sodium acetate was added to the eluted sample solution, and after 3 volumes of 99.5% ethanol (FUJIFILM Wako Pure Chemical). The final solution was incubated overnight at -20°C for increased yield and subsequently centrifuged at 4°C, 20,000 RCF for 60 min. The supernatant was carefully discarded, and 1 ml of 70% ethanol was added to the white pellet. A second centrifugation step was performed at 4°C, 20,000 x g for 15 min. The supernatant was again carefully discarded and the pellet was air-dried at room temperature for 30 min. Finally, the pellet was re-suspended in RNase free water and quantified by Nanodrop.

2.2.5 *In vitro* Translation Using Cell-free System for mRNA Display

From the recovered mRNA-tag samples, a total of 2 µg was used for *in vitro* translation using PURE*frex* 1.0 in a 100 µl final volume reaction mix. The optimised conditions for efficient mRNA-peptide conjugate formation are as follows: initial translation at 37°C for 30 min, followed by addition of a salt mix to a final concentration of 32.5 mM MgCl2 and 375 mMKCl, and then a second incubation at 37°C for 60 min. After this step, an equal volume of 2x Laemmli sample buffer (Bio-Rad) was added, mixed and centrifuged at 10,000 x g, 1 min and room temperature to remove the salt precipitant. The supernatant, containing the translated product (mRNA-peptide), was recovered and further resolved on a polyacrylamide SDS-

PAGE gel (3.5% stacking gel and 8 M urea 10% resolving gel). Afterwards, the gel bands corresponding to mRNA-peptide conjugate were excised and purified by electroelution and ethanol precipitation.

2.2.6 Optimisation of mRNA-peptide Conjugate Formation

The formation of the mRNA-peptide conjugate using PURE*frex*1.0 described in the paragraph 2.2.5, was firstly optimised through experimentation of three different reaction conditions: translation time, salt mix concentration, and incubation time after salt mix addition. All test reactions were performed at a 10 µl scale. For the translation time, 0, 5, 15, 30 and 60 minutes at 37°C were tested, before the addition of a specific salt mix (final concentration, 32.5 mM $MgCl_2$ and 375 mM KCl) followed by 60 minutes of incubation. For the optimisation of the salt mix concentration, firstly a 30 minutes translation was performed at 37°C, then either no salt or 1x, 1/2x, 1/4x, 1/8x concentrations of 65 mM $MgCl_2$ and 750 mM KCl salt mixed added. These salt concentration parameters (65 mM $MgCl_2$ and 750 mM KCl) were originally reported to promote a covalent bond between puromycin and polypeptide chain [4,5]. For the incubation time after salt addition, a 30 minutes translation was initially performed, and then salt added to final 32.5 mM $MgCl_2$, 375 mM KCl concentration, then 0, 10, 30, 60 and 90 minutes of posterior incubation tested.

2.2.7 Trypsin Digestion

Trypsin digestion was carried out by using 75 ng mRNA-tag sample (with and without translation under optimal condition), 1 mM $CaCl_2$, 50 mM ammonium

bicarbonate, and 0.5 µg of Trypsin Gold (Promega). Each reaction was incubated at 37°C for 30 min and resolved on a polyacrylamide SDS-PAGE gel (3.5% stacking gel and 8 M urea 10% resolving gel).

2.2.8 Reverse Transcription for cDNA Synthesis

Reverse transcription (RT) was performed against purified mRNA-peptide conjugates to obtain mRNA/cDNA duplexes for cDNA display using PrimeScript One Step RT-PCR Kit Ver.2. The initial peptide library (0th round) was prepared by mixing both FLAG-random and 10aa-random mRNA-peptide conjugate samples on a 1:10,000 molar ratio respectively. Approximately 200 ng (~2 x 10^{12} molecules) of this initial peptide library was reverse transcribed in a 50 µl reaction with a final concentration of 0.4 µM of RTPCR-R primer (Table 2.C) in 1 step buffer. Tubes were incubated at 80°C for 30 seconds, then cooled down to 4°C for primer annealing. After, the PrimeScript 1 step Enzyme Mix was added and incubated at 48°C for 20 min. An "RT–"tube was also prepared as a negative control, without adding the enzyme mix. After the incubation, both samples (RT-: mRNA-peptide and RT+: mRNA/cDNA-peptide) were purified using the NucleoSpin RNA Clean-up XS kit and quantified by Nanodrop.

2.2.9 Western Blot Analysis

Western blot analysis was performed to confirm the physical bond between the peptide and their mRNA sequences in the mRNA-peptide or mRNA/cDNA-peptide conjugates. A control mRNA sequence (Table 2.C) encoding the FLAG epitope (DYKDDDDK) was ligated with the puromycin-FITC DNA tag and translated

under optimised conditions. Samples were run on a 10-20% polyacrylamide SDS-PAGE gel (SuperSep Ace, FUJIFILM Wako) at constant 30 mA until the samples stack and reach the separating gel, and after at constant 50 mA for additional 40 minutes. The samples in the gel were semi-dry transferred to a PDVF membrane, blocking was carried out using the PVDF Blocking Reagent for Can Get Signal and, the membrane was incubated using Anti DYKDDDDK tag monoclonal antibody peroxidase-conjugated (FUJIFILM Wako Pure Chemical) and Can Get Signal, solution 2 with a 1: 12,000 ratio, respectively. The membrane was then revealed with chemiluminescence using ECL Select Western Blotting Detection Reagent.

2.2.10 Selection Using Anti-FLAG M2 Antibody

First, the Anti-FLAG M2 Magnetic Beads needed to be washed, so 40 µl of beads were mixed with 1 ml of TBS-T buffer (50 mM Tris-HCl, pH 7.4, with 150 mM NaCl, 1 mM EDTA, 0.2% Tween 20), incubated in a circular rotor for 5 min and then the supernatant extracted using the DynaMag-2 magnetic stand. This process was repeated two times. Afterwards, as two different samples, 100 ng of each initial library, the mRNA-peptide and the mRNA/cDNA-peptide, were diluted with 500 µl of TBS-T buffer. These diluted samples were mixed with the already washed beads and incubated with gentle mixing in the circular rotator for 1 hour to allow strong binding. After the incubation process, the supernatant was removed again with the magnetic stand, and the beads were further washed first three times with TBS-T and then another three times with TBS buffer. The incubation time for each washing step in the circular rotor was 5 min each. Through the 1st to 3rd rounds,

beads were re-suspended in 50 µl of TBS buffer and solutions were directly applied to the RT-PCR reaction (see 2.2.11 section) to generate a double-strand DNA library for the next selection round and sequencing. For the 4th round samples though, the elution phase was performed in several steps using a FLAG Peptide (Sigma-Aldrich) to compete and remove the conjugates, by increasing the FLAG Peptide concentrations from 4 µg/ml to 20 µg/ml and 100 µg/ml. Each of the eluted fractions, along with the remaining beads, were subjected to RT-PCR.

2.2.11 RT-PCR for Amplifying Selected Sequences

A total of 10 µl of re-suspended magnetic beads with bound mRNA-, mRNA/cDNA-peptide conjugates (from round 1 to 3)were used as a substrate for a 50 µl RT-PCR reaction using PrimeScript One Step RT-PCR Kit Ver.2 with 0.4 µM of each RTPCR-F and RTPCR-R primers (Table 2.C). The reaction mix was incubated at 80°C for 30 seconds, then cooled down to 4°C for primer annealing, and afterwards followed by the addition of the PrimeScript enzyme with incubation at 48°C for 20 min to complete the reverse transcription reaction. The reaction mixture was then split into a total 8 tubes to be able to recover the samples at different PCR cycles (0, 5, 10, 15, 20, 25, 30 and 35 to be more specific), and this way, check the amplification efficiency and optimal numbers of cycles needed so over-amplification of the DNA could be avoided. PCR conditions were: initial denaturation at 94°C for 60 seconds, followed by 30 cycles of 94°C for 15 seconds, 62°C for 30 seconds and 72°C for 25 seconds, then final extension at 68°C for 30 seconds. The samples were run on a 2% TAE agarose gel stained by SYBR Gold and the optimal number of cycles was

determined based on the band intensity and the absence of extra bands which are caused by non-specific amplified products. Finally, 50 µl of the RT-PCR reaction was once again conducted with optimal PCR cycles (20 cycles for rounds 1 and 2, and 15 cycles for round 3) and the product was purified using Wizard SV Gel and PCR Clean-Up System for the next round of selection and posterior sequencing.

2.2.12 Next-Generation Sequencing

Sequencing was performed for every round of DNA library (0 to 4 rounds). Firstly, quantification of the samples was carried out using Qubit 2.0 Fluorometer and Qubit 1X dsDNA HS Assay Kit. A total of 50 ng of each DNA library sample was used as an input for the NEBNext Ultra II DNA Library Prep with Sample Purification Beads and also the NEBNext Multiplex Oligos for Illumina (Index Primers Set 1) were used to multiplex the samples. The length of the ready DNA Library samples was determined with Agilent 2100 Bioanalyzer, then quantified by Qubit 2.0 Fluorometer and finally run on the MiSeqillumina platform using the Miseq Reagent Kit v3 600 cycles.

2.2.13 Sequence Analysis

The sequence information, from the MiSeqillumina platform, was retrieved as raw FASTQ files. The sequencing part corresponding to the coding regions was extracted using a pattern search program written by Perl script and only the sequences with an 8 nucleotide perfect match upstream of the start codon and downstream of the random region, were considered for downstream analysis. The selected sequences, still in FASTQ format, were further processed using

FASTAptamer-count[6] to rank sort the sequences based on their read counts. Consensus sequence logos were created using WebLogo 3[7] based on the top 50 reads with specific conditions (fixed letter and position).

2.3 Experiments for Chapter 5

2.3.1 General Materials and Verification Procedures

Unless otherwise specified, all the reagents, chemicals and vials were acquired from Thermo Fisher Scientific. The lipids were bought from Avanti Polar Lipids Inc; the PURE*frex* systems (version 2.0, 2.1 and DS supplements) from Gene Frontiers. The Wizard Plus SV Minipreps DNA Purification System from Promega, the BL21 (DE3) Competent *E. Coli* cells and the Bio-Beads SM-2 Resin from New England Biolabs.

Unless otherwise stated, size verification of DNA samples was performed on a 1.5% agarose Tris-Acetate-EDTA gel (UltraPure Agarose), with a constant voltage of 100V; while for peptide and protein samples it was in a pre-cast 10-20% polyacrylamide SDS-PAGE gel (SuperSep Ace, FUJIFILM Wako) at constant 30 mA until the samples were stacked and reached the separating gel, and after that, at constant 50 mA for an additional 40 minutes. For detecting DNA products, SYBR Gold Nucleic Acid Stain was used and for proteins, the staining used was Oriole Fluorescent Gel Staining (Bio-Rad). DNA markers used for the agarose gel were 1 kb DNA Ladder (NEB) and the protein ladder for the SDS-PAGE gel was Precision Plus Protein Dual ColorStandards (Bio-Rad).

DNA constructs were done using the KOD -Plus- Neo PCR Kit (Toyobo). All the primer sequences were ordered from FASMAC Inc., (Table 2.D).

Visualisation of vesicles was done by fluorescence microscopy using an Aarson 40x - 1000x Fluorescence Microscope (RSB-14, Aarson Scientific), and also by Fluorescence-Activated Cell Sorting (FACS) using a BD FACS Aria II (BD Biosciences).

The protein purification process was performed using either MagneHIS Ni-Particles (Promega) or Anti-HA-tag mAb-Magnetic Beads Monoclonal Antibody (MBL). Also, DynaMag-2 magnetic stand (Thermo Fisher Scientific) was used. All the incubation procedures were done in a circular rotator (Labinco L28, Salford Scientific Supplies).

DNA quantification was performed by UV-Vis spectrophotometry (Nanodrop 2000c, Thermo Fisher Scientific). All the cell-free reactions were incubated in a thermomixer (Eppendorf Thermomixer R, Eppendorf). Centrifugation steps were performed in a microcentrifuge (model 5415 R, Eppendorf), with different tubes adaptors depending on the size of the tubes used.

2.3.2 DNA Constructs, Primers and PCR

The plasmid for MSP1D1 protein was obtained from Addgene (#20061) and used as a circular construct for protein expression. The plasmid for the CX3CR1 receptor (pCX3CR1) was obtained from the Takuya Ueda Laboratory (University of Tokyo) [8]. The plasmid for sfGFP (psfGFP) was obtained from a colleague in the same laboratory (Earth-Life Science Institute). The CX3CR1 samples were used as linear

DNA templates for protein expression. The first fragment used a forward primer containing a T7 promoter (with a 17A mutation) and RBS sequence and a reverse primer with an HA-tag. For this, two-step PCR was performed using the plasmid, pCX3CR1, the forward primer T7_CX3CR1_HA_F1 and the reverse primers T7_CX3CR1_HA_R (Table 2.D) were used with the following the first PCR conditions: Initial Denaturation (94°C, 120 sec), 35 cycles of denaturation (98°C, 10 sec) and annealing/extension (68°C, 40 sec), and Final Extension (68°C, 120 sec). The result constructs, called RBS-CX3CR1_HA, was then used as the DNA template for the second round of PCR using the primers T7_CX3CR1_HA_F2 and T7_CX3CR1_HA_R (Table 2.D) with the following second PCR conditions: Initial denaturation (94°C, 120 sec), 35 cycles of denaturation (98°C, 10 sec), annealing (58°C, 30 sec) and extension (68°C, 40 sec), and final extension (68°C, 120 sec). The construct from this final PCR was referred to as T7-CX3CR1_HA.

The second fragment, CX3CR1-sfGFP was constructed by overlap PCR. It fused an sfGFP sequence to the C-terminus of a CX3CR1 fragment. The sfGFP fragment was done using the psfGFP plasmid, the sfGFP_F forward primer and the sfGFP_R reverse primer (Table 2.D) with the following PCR conditions: Initial denaturation (94°C, 120 sec), 35 cycles of denaturation (98°C, 10 sec), annealing (67°C, 30 sec) and extension (68°C, 27 sec), and final extension (68°C, 120 sec). The CX3CR1 fragment was done using a DNA template, the RBS-CX3CR1_HA fragment mentioned above. A T7 promoter was added and the HA tag replaced with the overlapping sequence. The primers used were T7_CX3CR1_HA_F1 and

T7_CX3CR1_R (Table 2.D) with the following PCR conditions: Initial denaturation (94°C, 120 sec), 35 cycles of denaturation (98°C, 10 sec), annealing (58°C, 30 sec) and extension (68°C, 36 sec), and final extension (68°C, 120 sec). For the final overlapping PCR reaction, both sfGFP and CX3CR1 fragments were annealed with the following CPR conditions: Initial denaturation (94°C, 120 sec), 35 cycles of denaturation (98°C, 10 sec), annealing (63°C, 30 sec) and extension (68°C, 60 sec), and final extension (68°C, 120 sec). The construct from this final PCR was referred to as T7-CX3CR1_sfGFP.

Table 2.D: Sequences and names of the oligo primers used for creating all the CX3CR1 constructs.

Name	Sequence (5' to 3')
T7_CX3CR1_HA_F1	5'- TTTAAGAAGGAGATATACCAATGGATCAGTTCCCG – 3'
T7_CX3CR1_HA_R	5'- TGCTAGTTAAGCGTAATCTGGAACATCGTATGGGTACAGCAGCAGCAGTGC – 3'
T7_CX3CR1_HA_F2	5' - GAAATAAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAATT TTGTTTAACTTTAAGAAGGAGATATACCA – 3'
sfGFP_F	5'- ACGGTCTCTGCGATCTCCCAATGAGTAAAGGAGAA – 3'
sfGFP_R	5'- GCGCCAGTGCGACGCTATTATTTGTAGAGCTCATC – 3'
T7_CX3CR1_R	5'- GCACTGCTGCTGCTGTGGGAGATCGCAGAGACCGT – 3'

2.3.3 Giant Unilamellar Vesicles (GUVs) Preparation

Liposomes were, produced by water-in-oil (w/o) emulsion transfer [9]. The lipid mixture used was POPC: Cholesterol (1: 1). First, a 6 cubic ml vial is cleaned thoroughly and all traces of dust inside it are removed with pressure air. Once completed, dry lipids and paraffin oil are put inside following the calculations for the desired molarity. On top of that, a small stirrer is introduced with a pincer to

avoid dust going inside. The solution is then mixed at room temperature and 1,000 rpm for 2-3 h (or overnight at 4°C) in a magnetic stirrer (BT Lab Systems). In the meantime, the inner and outer solution of the liposomes is made. When testing new molarity, ratio or new lipids themselves, 200 mM of Glucose and 200 mM of Sucrose are used for the outer and inner solution respectively. Once good conditions are found, the PURE*frex* system is used instead. For this, the inner solution is the normal PURE mixture as it is stipulated in the protocol, while the outer solution consists of the non-protein components of the PURE system, minus also the tRNAs, and complemented with 200 mM of Glucose.

In an Eppendorf tube, first 200 µl of the outer solution is set at the bottom of the tube, then directly on top of it, 100 µl of the mixed lipid solution is poured slowly. In another Eppendorf tube, 200 µl of the lipid solution along 5 µl of the Inner solution is mixed by flicking the tube hard with a finger until the colour goes from transparent to slightly white. This lipid-inner solution mixture is slowly mixed inside the top layer of lipid solution from the first Eppendorf with a pipette. Once finished, the Eppendorf tube is centrifuged 300 g for 10 minutes, and then at 2,500 g for another 10 minutes. After this step, a visible white pellet at the bottom of the tube appears. With the help of a pipette, the top layer of the lipid solution is slowly removed out of the tube and discarded until only the outer solution and the white pellet remain. From this point onwards, the white pellet can be slowly and carefully removed with a pipette and poured into another clean Eppendorf, or a small hole can be made at the bottom of the Eppendorf, where the while pellet is, with the

help of a needle, to let a couple of droplets containing the freshly made vesicles out of the tube with pressure by closing the lid. Whether one or the other procedure is performed, a clean tube will be containing the new vesicles with the as little amount of outer solution as possible, to dissolve the sample as little as possible.

To check the vesicles by fluorescence microscope, a tape separator with a hole in the middle was put between the two glass slides, and 2 µl of vesicles sample were placed in the middle of the hole.

2.3.4 Fluorescence-Activated Cell Sorting (FACS)

For vesicle visualisation using FACS, around 50-100 µl of vesicles at a 1:10 dilution rate was used each time. The experiment was carried out at room temperature when testing the vesicles and at 4°C when using the PURE system to avoid more product being produced. The flow rate was set to 1.0 and the FSC (Forward Scattering) and SSC (Side Scattering) thresholds to 200 both (all arbitrary units). The voltage was set to variable values that would allow for vesicles' visualisation in the plot graph.

2.3.5 Membrane Scaffold Protein (MSP) Production, PURE*frex*2.0

For the making of the membrane scaffold protein, MSP1D1, some already stipulated protocols were followed [8,10] with some slight modifications.

The DH5α *E. coli* bacteria containing the plasmid with the MSP1D1 gene (Addgene, #20061) was first grown in 10 ml of LB medium containing Kanamycin (50 µg/ ml) overnight at 37°C with moderate shaking. The plasmid was extracted using the Wizard Plus SV Minipreps DNA Purification System and quantified with

Nanodrop. For storing and further plasmid expression purposes, the plasmid was transformed into *E. coli* BL21 (DE3) cells following the transformation protocol and the following day, selected colonies stored with 70% Glycerol at -80°C, while other colonies were grown in another 10 ml of LB medium containing Kanamycin (50 µg/ml) for 6 h at 37°C with moderate shaking. The plasmid was again extruded using the aforementioned miniprep kit.

The MSP1D1 proteins were expressed in PURE*frex* 2.0 since it yields a higher protein output. The reaction was carried out in a final volume of 100 µl, using 10 ng of the MSP1D1 plasmid and following the company's protocol by incubating the reaction at 37°C for 6 h.

2.3.6 Membrane Scaffold Protein (MSP) Purification, MagneHis Ni-Particles

The MagneHis Ni-Particles were used to purify the MSP1D1 since these proteins have a 7-His tag at their N terminal. First, 50 µl of the beads were cleaned thoroughly two times with Lysis Buffer (50 mMTris HCl pH 8.07, 300 mM NaCl, 5 mM MgCl2) by washing them with the mentioned buffer in a circular rotator for 5 min each at room temperature.

Parting from 100 µl of the PURE*frex* 2.0 reaction, 5 µl of it were taken as the "PURE" fraction, for posterior analysis. The remaining 95 µl were mixed with 400 µl of Lysis Buffer and the 50 µl of the washed beads (with no liquid) in a circular rotator for 30 min at room temperature, or overnight at 4°C. The tube was then placed in the Magnetic Stand for 2 min, and after the liquid taken out and put inside an

Eppendorf tube as the “Flow-Through” fraction. To the remaining beads, 500 µl of Lysis Buffer with 20 mM Imidazole was added and the sample further incubated in a circular rotor for 10 min. After placing the tube in the magnetic stand, the liquid was transferred to an Eppendorf tube as “Wash 1”. The washing process was repeated and a second fraction, “Wash 2” was obtained. The process was finalised by adding to the beads another 500 µl of Lysis Buffer but this time with 500 mM Imidazole. The mixture was incubated for the last time in the circular rotor for another 10 minutes and then placed in the magnetic stand where the liquid was taken as the “Elution” fraction. All the fractions were flash-frozen and stored at -80°C.

The “Elution” fraction was later used for dialysis to remove the imidazole. The sample was put inside a dialysis cassette (Micro Float-A-Lyzer, Repligen) which was left inside a tank filled with 1 L of Lysis Buffer, overnight at 4°C with mild rotation.

2.3.7 Membrane Scaffold Protein (MSP) Concentration and Verification

After the dialysis of the MSP1D1 sample (section 2.3.5), 5 µl were stored on ice. Buffer exchange was performed with PBS using Amicon Ultra Centrifugal Filters (Merck Millipore) and the sample was also concentrated to a total volume of 50 µl. All the fractions, along with the dialysis and buffer exchange samples were analysed on two different 10–20% SDS-PAGE gels, for the correct size verification, at constant 30 mA until the samples stack and reach the separating gel, and after at constant 50 mA for additional 40 minutes. One gel was stained with Oriole

Fluorescente Gel Stain for immediate visualisation and the other was subjected to semi-dry conditions and transferred to a PDVF membrane. Blocking was carried out using the PVDF Blocking Reagent for Can Get Signal and, the membrane was incubated using Anti 6xHistidine monoclonal antibody (9C11), Peroxidase-conjugated (FUJIFILM Wako Chemicals) and Can Get Signal, solution 2 with a 1: 10,000 ratio, respectively. The membrane was then revealed with chemiluminescence using ECL Select Western Blotting Detection Reagent.

The sample was posteriorly flash-frozen for -80°C storage. The protein concentration for the MSP1D1 was calculated by Nanodrop.

2.3.8 Nanodiscs Production, Bio-Beads

The whole procedure was done following already stipulated protocols [11–14] with some modifications. When syringes were needed, metallic and glass ones were selected, instead of plastic, due to the use of chloroform. Syringes were pre-rinsed with chloroform and vials cleaned from dust with pressured air. The lipids used for the construction of the nanodiscs were POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine), POPS (1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine) and Cholesterol in a 72: 20: 8 proportion respectively. The ratio of membrane scaffold protein to lipid mixture was 1: 60 (MSP1D1: lipid mixture). All the lipids were dissolved in Chloroform.

First, the calculated volume of lipid mixture (POPC: POPS: Cholesterol) was poured into a 6 cc crystal vial and N2 gas was forced inside to dry the mixture, creating this way a dry film. The vial was then covered in aluminium foil and

further dried overnight by leaving it at room temperature in a vacuum so that all the humidity is removed. The next day, the dry film was hydrated with NanodiscBuffer (20 mMTris/HCl, pH 7.4 and 100 mM NaCl) containing 100 mM Na cholate. During the hydration process, the glass vial was put into a water bath at 60°C with sonication to ensure all the lipid dry films are fully dissolved. The process was aided by vortexing and/or flicking. Once dissolved, the whole volume was passed into a 2 ml tube and the MSP1D1 proteins were added. The sample was further incubated at 4°C for 1 h and kept in ice afterwards.

In the meantime, the Bio-Beads SM-2 Resin was prepared to be used with the sample. First, the top 5 cm of a 25 ml pipette was cut off with a saw and the tube cleaned well with pressured air. Cotton was then pushed down the cut pipette, to where the tip is. The cut pipette was supported vertically with a holder and 25 ml of MiliQ water was run through it. Then around 2.5 gr of beads (4-5 ml) was put inside the cut pipette with 25 ml of 100% methanol. Before all the liquid came out of the cut pipette, the beads were pipetted up and down the tube to wash them thoroughly. Posteriorly, the beads were further washed 3 times with MiliQ water and finally another 3 times with Nanodisc Buffer (without the detergent). During the last wash, again before all the liquid came out the bottom, the flow was stopped with paraffin wax paper and the beads in the Nanodisc Buffer (around 15 ml) were collected into a Falcon tube.

For every microliter of sample 1 mg of dry beads (no liquid) is necessary. To the lipid-detergent-MSP1D1 sample, the dry bio-beads were added and the mixture was

mixed overnight at 4°C in a circular rotator. The following day, the mixed sample was centrifuged at 4°C for 10 minutes and 12,000 rpm. The supernatant was transferred into another Eppendorf tube and kept at 4°C.

2.3.9 Nanodiscs Purification and Concentration, AKTA Chromatography

Further purification of the nanodiscs was performed using AKTA chromatography (ÄKTA pure 25L, Cytiva) to separate the correctly assembled nanodiscs samples from the aberrant ones. The column used was HiLoad 16/600 Superdex 200 pg.

The machine was cleaned and flushed with Nanodisc Buffer (20 mMTris/HCl, pH 7.4 and 100 mMNaCl) before injecting the sample. The 500 µl tube line was used for running the sample through it and a flow rate of 0.2 ml/min was selected. The entire run lasted for 2h and a total of 24 fractions were collected, each one with 1 ml of volume. A total of 10 fractions were selected to determine if they contain the purified nanodiscs from the AKTA plot. The gel was run with 5 µl from each selected fraction.

A total of 5 fractions containing most of the purified nanodiscs were selected for concentration using Amicon Ultra-15 Centrifugal Filter (Merck). A final volume of 1.3 ml was collected and stored at -80°C by flash freezing in 5 µl and 10 µl aliquots. The concentration of the Nanodiscs was determined by Nanodrop at 131 µM.

2.3.10 G Protein-Coupled Receptors (GPCR) Production, PURE*frex* 2.1 and DS Supplement

The PURE*frex* 2.1 version was selected first for its increase in product yield but more so, for the compatibility with the DS Supplement, which aids in the post-

transcriptional formation of sulphur bonds between proteins or subsections to secure the protein functionality. In the case CX3CR1 (C-C Chemokine receptor), this supplement was necessary. The reducing agent used was reduced glutathione (GSH). For this reaction, the nanodiscs previously produced (section 2.4.8) were supplemented at a final concentration of 10 µM and the template DNA, T7-CX3CR1-HA at a final concentration of 5 nM. The final volume of the reaction was 100 µl and it was incubated for 6 h at 37°C.

2.3.11 G Protein-Coupled Receptors (GPCR) Purification, Anti-HA-tag mAb-Magnetic Beads.

The Anti-HA-tag mAb-Magnetic Beads were used to purify the CX3CR1 receptors that assembled correctly, from the empty nanodiscs. First, 50 µl of the beads were cleaned thoroughly 2 times with PBS buffer by washing them in a circular rotator for 5 min each at room temperature. Parting from 100 µl of the PURE*frex* 2.1 DS Supplement reaction, 5 µl of it were stored as the "PURE" fraction. The remaining 95 µl were mixed with 400 µl of PBS Buffer and the 50 µl of the washed beads (with no liquid) in a circular rotator for 30 min at room temperature, or overnight at 4°C. The tube was then placed in a Magnetic Stand for 2 min. The liquid was taken out and put inside an Eppendorf tube as the "Flow-Through" fraction. To the beads, 500 µl of PBS Buffer was added and the solution mixed in a circular rotor for 10 min. After placing the tube in the magnetic stand, the liquid was then transferred to an Eppendorf tube as the "Wash 1" fraction. The washing process was repeated and a second fraction, "Wash 2" obtained. The process was finalised by adding to

the beads 200 µl of HA-peptide. The mixture was incubated in the rotor for another 30 minutes and then placed in the magnetic stand where the liquid was removed and taken as the "Elution" fraction. All the fractions were flash-frozen and stored at -80°C.

All the fractions were analysed on a 10–20% SDS-PAGE gel, for correct size verification, at constant 30 mA until the samples stack and reach the separating gel, and after at constant 50 mA for additional 40 minutes. The gel was stained with Oriole Fluorescent Gel Stain for immediate visualisation.

2.3.12 Radiolabelling

Radiolabelling was used to determine the concentration of CX3CR1 produced and the efficiency of the purification process. The radioactive amino acid used in this case was Methionine [S^{35}] (Methionine L-[S^{35}], PerkinElmer).

Parting from 49 µl of PURE*frex* 2.1 DS Supplement reaction, 1 µl of Methionine S^{35} was added. The reaction was incubated at 37°C for 3 h. Afterwards, the samples were purified as stipulated in section 2.3.11.

All the fractions obtained during the purification process were run in a pre-cast 10–20% SDS-PAGE gel and run at constant 120 V for 60 min and then at 200 V for another 45 min. The gel was then incubated in Fixing Solution (30% Methanol and 10% Acetic Acid) for 30 min with mild shaking. After the fixing step, the gel was dried to remove all radioactive water and humidity from it. Using a gel dryer (AE-3711RapiDry mini, Atto), the gel was dehydrated at 80°C during 1 h.

In the meantime, several dilutions of the Methionine [S^{35}] were made for a standard curve. The labelled amino acid was diluted with 50 mM HEPES pH 7.5 in the following dilutions: 1/500, 1/1,000, ½,000, 1/5,000, 1/10,000, 1/50,000, 1/100,000 and 1/200,000. A total of 10 µl of each dilution was spotted onto a rectangular piece of filter paper and each dilution spots labelled. The spotted samples were then left to dry.

Once the dilutions and the samples were dried in their respective filter papers, they were put inside an autoradiography cassette (Fisher Scientific) along with an intensifying screen (Fisher Scientific) and incubated in the dark for 2 days. After this period, the intensifying screen was revealed using a phosphorimager (Typhoon FLA 7000, GE Healthcare).

2.4 References for Chapter 2

1. R Development Core Team. R: A language and environment for statistical computing. *Vienna, Austria* (2017). R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
2. Jones, E., Oliphant, T., Peterson, P. & Others. SciPy: Open Source Scientific Tools for Python, 2001 (http://www.scipy.org/). *Http://Www.Scipy.Org/* (2015).
3. Sprouffske, K. & Wagner, A. Growthcurver: An R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics* (2016). doi:10.1186/s12859-016-1016-7
4. Naimuddin, M. *et al.* Directed evolution of a three-finger neurotoxin by using cDNA display yields antagonists as well as agonists of interleukin-6 receptor signaling. *Mol. Brain* (2011). doi:10.1186/1756-6606-4-2
5. Naimuddin, M. & Kubo, T. A High Performance Platform Based on cDNA Display for Efficient Synthesis of Protein Fusions and Accelerated Directed Evolution. *ACS Comb. Sci.* (2016). doi:10.1021/acscombsci.5b00139
6. Alam, K. K., Chang, J. L. & Burke, D. H. FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol. Ther. - Nucleic Acids* (2015). doi:10.1038/mtna.2015.4
7. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* (2004). doi:10.1101/gr.849004
8. Gessesse, B., Nagaike, T., Nagata, K., Shimizu, Y. & Ueda, T. G-Protein Coupled Receptor Protein Synthesis on a Lipid Bilayer Using a Reconstituted Cell-Free Protein Synthesis System. *Life* (2018). doi:10.3390/life8040054
9. Fujii, S. *et al.* Liposome display for in vitro selection and evolution of membrane proteins. *Nat. Protoc.* **9**, 1578–91 (2014).
10. Ritchie, T. K. *et al.* Chapter 11 Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs. *Methods in Enzymology* (2009). doi:10.1016/S0076-6879(09)64011-8
11. Gessesse, B., Nagaike, T., Nagata, K., Shimizu, Y. & Ueda, T. G-Protein Coupled Receptor Protein Synthesis on a Lipid Bilayer Using a Reconstituted Cell-Free Protein Synthesis System. *Life* **8**, 54 (2018).
12. Bayburt, T. H. & Sligar, S. G. Membrane protein assembly into Nanodiscs. *FEBS Lett.* (2010). doi:10.1016/j.febslet.2009.10.024
13. Denisov, I. G., Grinkova, Y. V., Lazarides, A. A. & Sligar, S. G. Directed Self-

Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *J. Am. Chem. Soc.* (2004). doi:10.1021/ja0393574

14. Lamichhane, R., Liu, J., Pauszek III, R. & Millar, D. Fluorophore Labeling, Nanodisc Reconstitution and Single-molecule Observation of a G Protein-coupled Receptor. *BIO-PROTOCOL* (2017). doi:10.21769/bioprotoc.2332

Chapter 3. Cell-free protein expression systems and promoter library mutants: uncovering their dynamics and tackling optimization processes

3.1 Chapter 3 Aims and Summary

Cell-free protein expression systems have been widely used for applications such as synthetic biology and metabolic engineering. But even if broadly used, little is known about how protein expression works in cell-free systems. Furthermore, there are gaps in understanding the molecular interactions taking place, their dynamics, and how these can contribute to bottlenecks in the process. This chapter describes a systems approach to uncovering the underlying dynamics of protein expression in cell-free systems.

A set of T7 promoter variants was constructed to express superfolder GFP (sfGFP) in two different cell-free systems: one with reconstituted purified elements and the other based on *E. coli* extracts. The maximum expression level and the rate of protein synthesis as responses to the transcription rate change were different in the two cell-free systems, suggesting they could be driven by different expression dynamics. This observation was confirmed by a mathematical model for each cell-free system. It revealed negative feedback in the mRNA-protein translation by the PURE system and also identified limiting factors that could be optimised to enhance protein expression in the two cell-free systems (Fig.3.A). For the PURE system, these dynamics can be extrapolated to other types of systems based on

purified components as long as they maintain the same level of contaminant-free environment. Differences could occur if different proteins and components were to be used. That is why the results obtained in this chapter would probably not match many other types of extract-based system. These systems differ much since they come from a wide range of different organisms, natural or modified.

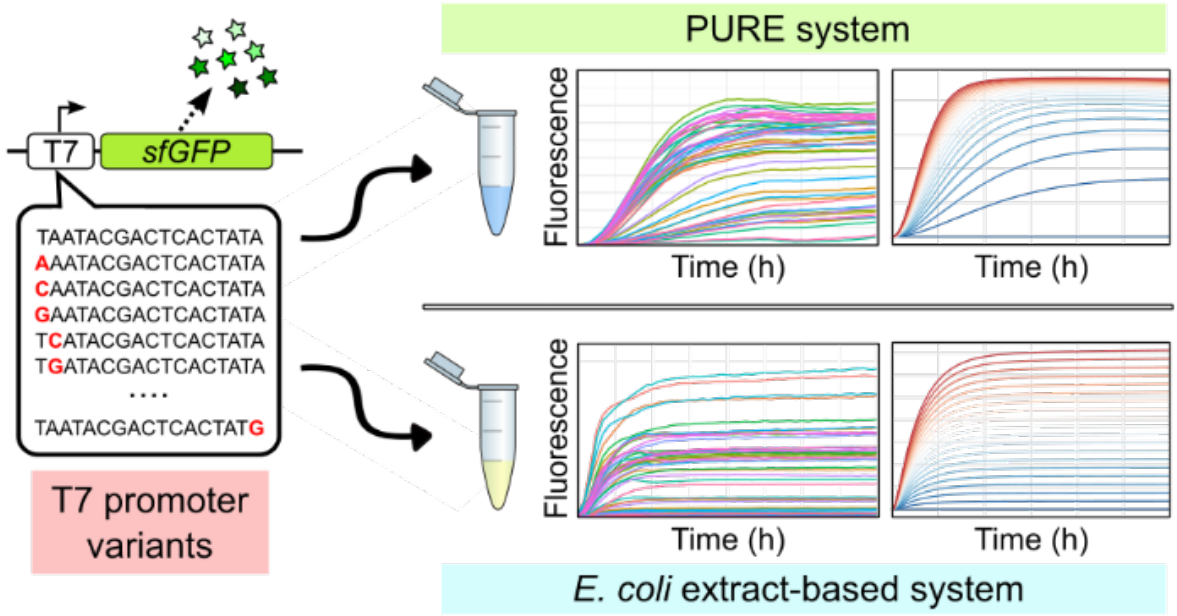


Figure 3.A: Summary of the experimental approach carried out for Chapter 3. From left to right, first, a series of T7 promoter variants are created and tested in two different cell-free systems. The protein expression of sfGFP is measured in both and mathematical models are run with the information gathered to reveal the dynamics of both systems.

3.2 Preparation of T7 Promoter Library Variants Constructs

The T7 promoter has been universally used for *in vitro* and *in vivo* experiments [1,2], mostly due to the high promoter-specific strength of the T7 RNA Polymerase (RNAP).

To gather information about its performance in cell-free systems, screens of the promoter sequence against transcription rates and translated product yield have been performed [3]. However such studies have focused on multiple mutations and high throughput screening, which prevented a systematic analysis that could generate rational understanding of the bottlenecks. Here we used single mutations and studied the dynamics of two very different cell-free systems.

Alterations were made to the 17 base-pair long recognition site of the T7 promoter sequence (TAATACGACTCACTATA) and a library constructed of T7 promoter variants, that encompassed all possible single base-pair alterations (Fig.3.B), yielding a total of 51 variants. Each of these variants was ligated downstream to a common ribosome binding site (RBS) and sfGFP. The original T7 promoter sequence will be referred in this chapter as the "core sequence" or “WT” from now on for easier referral.

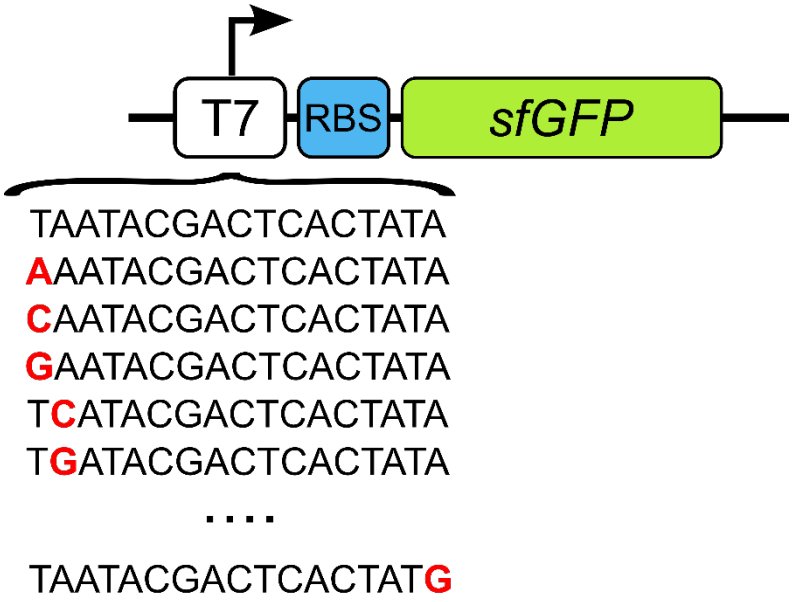


Figure 3.B: Schematic of the linear DNA templates used for protein expression in this chapter. Each DNA sequence contains the consensus core sequence (or one of its 51 variants) (white), a ribosome binding site (RBS) (blue) and an sfGFP gene (green). Single base-pair substitutions in the T7 promoter variants are highlighted in red.

3.3 The PURE System vs. the Expressway System

To investigate the expression levels of sfGFP in each DNA construct, they needed to be tested in a cell-free system. The cell-free systems to be used with the constructs had to include the T7 RNA Polymerase protein as all of the DNA constructs use the T7 promoter (or variations of it). The cell-free systems, also called *in vitro* transcription-translation systems, were selected to represent the two competing approaches currently used in the field, namely either purified components (PURE system (PURE*frex* 1.0, GeneFrontier)[4]) or cell extracts *(E.coli* extract-based cell-free system (Expressway™ Cell-Free E. coli Expression System, Thermo Fisher Scientific)) (Fig.3.C).

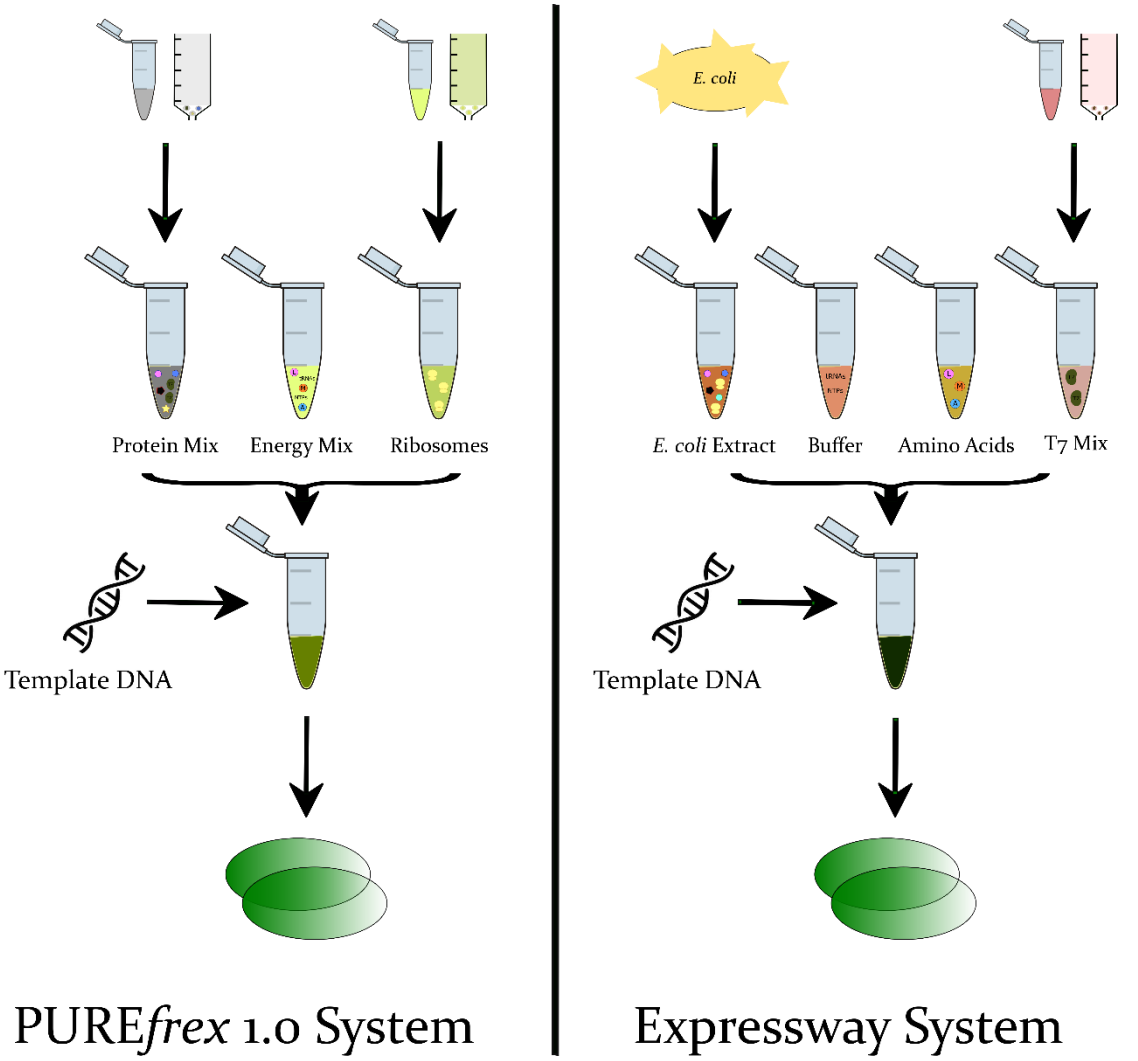


Figure 3.C: Schematics of both the PURExpress 1.0 Cell-free System and the Expressway Cell-free system. The PURE system (left) contains all the minimum necessary components to perform a transcription-translation reaction. All its protein and enzymatic components are recombinant and were produced *in vitro* to be later purified by chromatography. It is composed of three solutions: the protein mix contain all the proteins and enzymes necessary (minus the ribosomes); the energy mix contains all the chemical components, factors and co-factors and buffers for the proteins and enzymes to work; and the ribosome mix contains the purified ribosomes. The Expressway system (right) is based on an *E. coli* cell extract which contains a mixture of protein and enzymes, which is supplemented by buffers, amino acids and the T7 RNA Polymerase.

3.4 mRNA and Protein Degradation

To measure the protein yield for all constructs and both cell-free systems, each sample was incubated with either PURE*frex* 1.0 and Expressway and the sfGFP fluorescence measured every 15 min for at least for 12 h. All the experiments were performed at least in duplicate.

A few constructs were selected to represent the wide variation of sfGFP fluorescence obtained from the DNA library. The results can be seen in Figure 3.D (A and B), where a time-course of the sfGFP fluorescence is plotted (the core sequence is denominated as "WT"). Results for all variants are provided in Appendix (Section A.1, Figure A.1.1.).

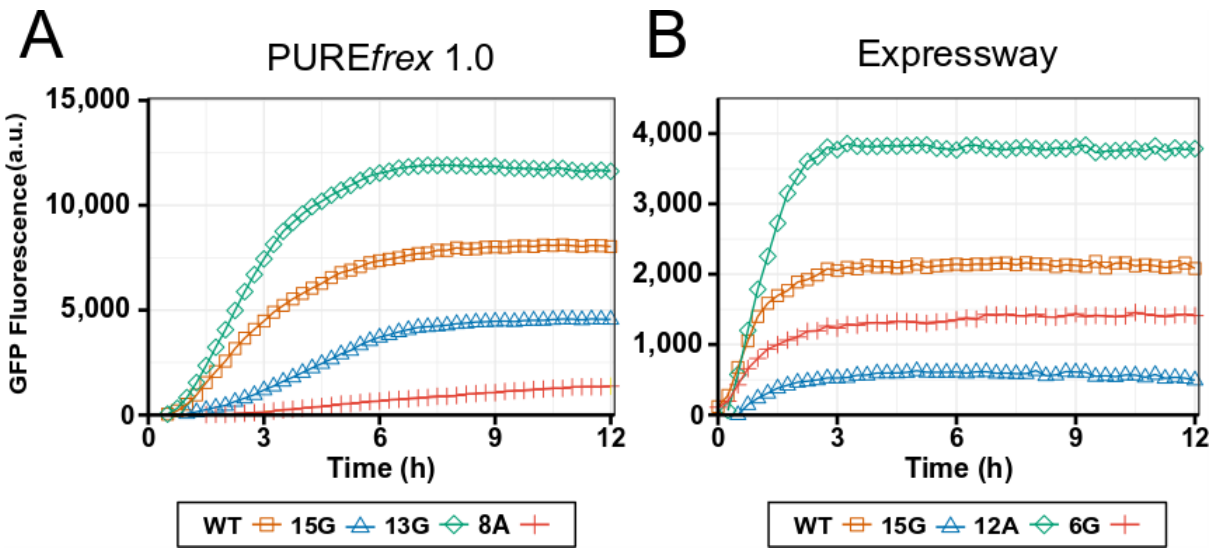


Figure 3.D: Protein expression in the PURE and extract-based cell-free system using T7 promoter variants. (A) PURE*frex* System and (B) Expressway System. Fluorescence measurements of sfGFP expression were made every 15 min with the consensus promoter sequence (denoted as WT) and three variants.

In the PURE system, the reaction took an average of 7.5 h (± 1.5 h) for saturation, while for the extract-based system it took an average of 3.5 h (± 0.5 h) to saturate. The first thing to mention about the plots is the difference in sfGFP fluorescence achieved from each variant. This shows that each mutation performed in the T7 promoter affected the final levels of protein expression. When looking at each plot, they all have the same shape. This includes a first exponential phase in which protein production is happening. But as resources are used and they eventually ran out, the plot comes to a saturation point in which the protein production halts.

A possible reason for the difference between both systems can be the rate of both DNA and RNA degradation. Since the PURE system is made from recombinant elements, the reaction is expected to last longer as it is free from DNase and RNase[5] elements; without having into consideration possible contaminants that could fall into the reaction while performing it. This feature would allow the reaction to last longer as the substrates for both the T7 RNA polymerase and ribosome (DNA and mRNA respectively) would degrade at a much slower pace. On the other hand, since the majority of the extract-based system proteins and enzymes come from the *E. coli* cytoplasm (except for the T7 RNAP), it could also include components that are part of the bacteria metabolism. Some of those elements could be nucleases that degrade mRNA and DNA, which in return would cause early termination of the protein synthesis due to lack of the two main substrates. Also, as the sequences have only five additional base-pairs upstream of the T7 promoter

sequence, these linear DNA templates were probably digested by nucleases and lost the promoter sequence faster in the Expressway system.

Furthermore, some promoter variants had higher final expression levels than those of the core sequence (e.g. 13G in PURE and 12A in Expressway). These improvements were not as common in previous studies with T7 promoter variants, whether in linear [6] or plasmid form [7].

To better compare different variants and their expression levels, the final fluorescence numbers were normalized to that of the core sequence. This was done by averaging the two values obtained from each variant (duplicate experiment) and then dividing that value to that of the wild type. A value similar to 1 indicates a similar level of protein expression than that of the core sequence. A value above 1 indicates a higher level of protein expression when compared to the wild type version and so on. When visualizing these normalized number in descending order (Figure 3.E, A and B), a spectrum of protein expression levels can be seen in both cases. This trait is ideal when trying to perform high throughput screening in a library of samples and for identifying an optimal candidate. Out of all the 51 variants, some showed an improvement in expression levels, compared to the core sequence. In the PURE system, 12 of them did while in the extract-based system, 9 did. The highest yield change averaged at a 1.5-fold increase for the PURE system (1.4 and 1.6 for each duplicate), while in the extract-based system it was a 3.6-fold increase (3.4 and 3.8 for each duplicate).

Although the protein expression increment in some variants in the extract-based system was high, the absolute expression level in the extract-based system was lower when compared to that in the PURE system (Figure 3.J). For example, using the core “WT” sequence in the PURE system led to fluorescence levels 3 times higher than in the extract (Figure 3.D). This could be due to proteases, mainly absent in the PURE system, but not specifically removed in the extract. These proteases degrade proteins, including the newly synthesized sfGFP, and therefore could limit the output.

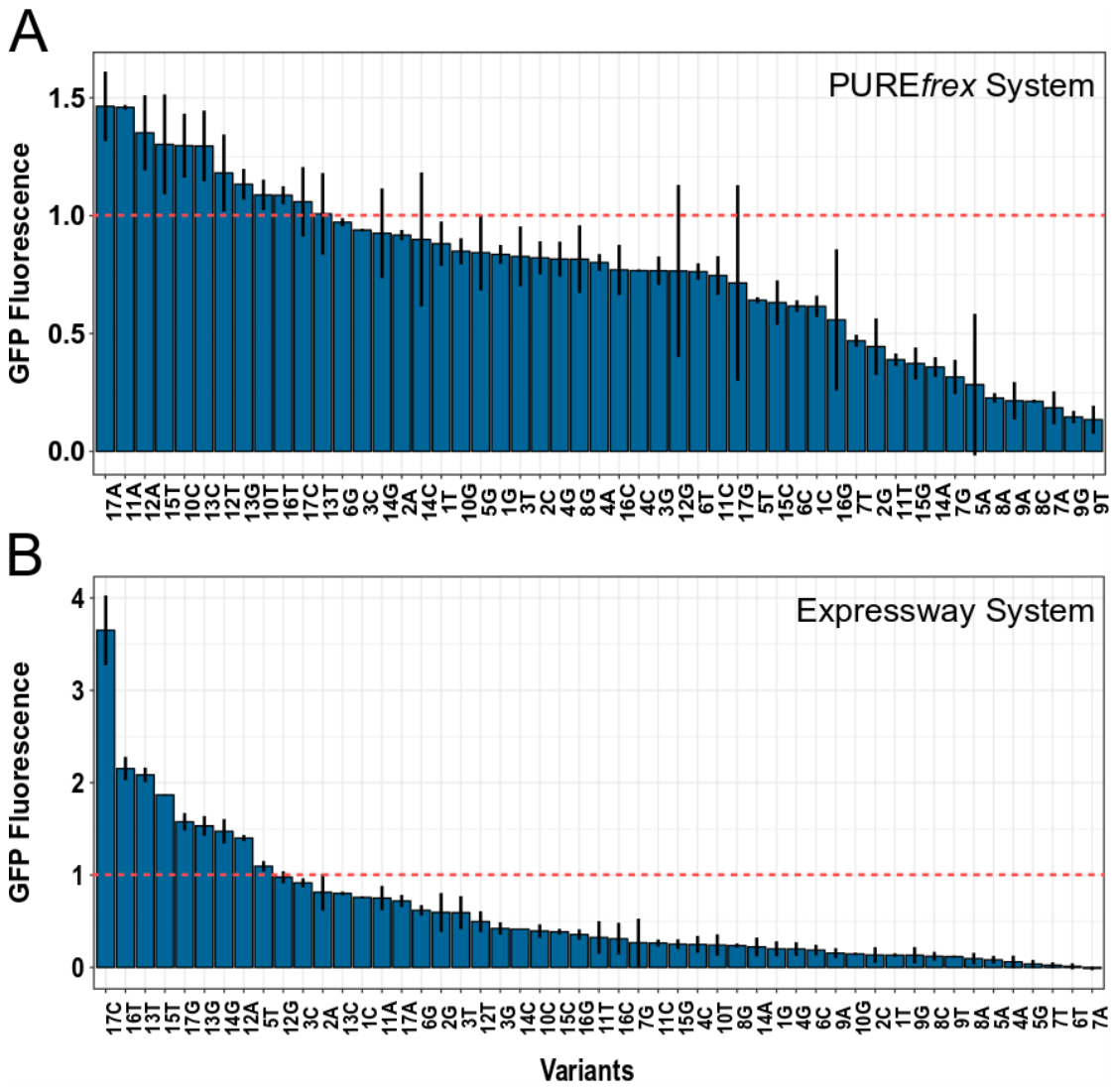


Figure 3.E: Relative sfGFP fluorescence with 51 T7 promoter variants, normalized to that with the consensus sequence (indicated by the red line) for the PURE*frex* System (A) and the Expressway System (B). Each variant was identified by the position and substituted base. For example, "17A" indicates a base at position -17 was substituted to adenine. The black bars represent the upper and lower values of the averaged value for each par of duplicates.

3.5 Reproducibility

To test the reproducibility level of both system, two sets of values were obtained from experiments performed on two different days. This reproducibility level was calculated by plotting both normalized values against each other. The R-squared value (R^2), represents how close the data are to the fitted regression line. This line represents the relationship between the x-scale value and the y-scale value. The closer the R^2 value is to 1, the more similar both values are and the more reproducible the data is. The regression line does not necessarily have to pass through the 0-0 origin value. This is due to the fact that this line tries to accommodate as much as possible passing through all the data points, and since the blank was already subtracted from each point, there is not a 0-0 value.

The extract-based system had higher reproducibility levels (R^2=0.97) between the two experiments than the PURE system (R^2=0.74). (Figure 3.F, A and B).

In the PURE system, proteins are often partially translated and consequently functionally inactive due to ribosome stalling on the mRNA [8]. On the other hand, the Expressway system is made from bacterial extracts and therefore it contains additional elements that can rescue stalled ribosomes; such as the alternative ribosome-rescue factor A (ArfA), release factor 2 (RF-2) [9] and the elongation factor P (EF-P) [10,11]. These factors, among others, could help increase the quality of the produced proteins and therefore increases the level of reproducibility between experiments.

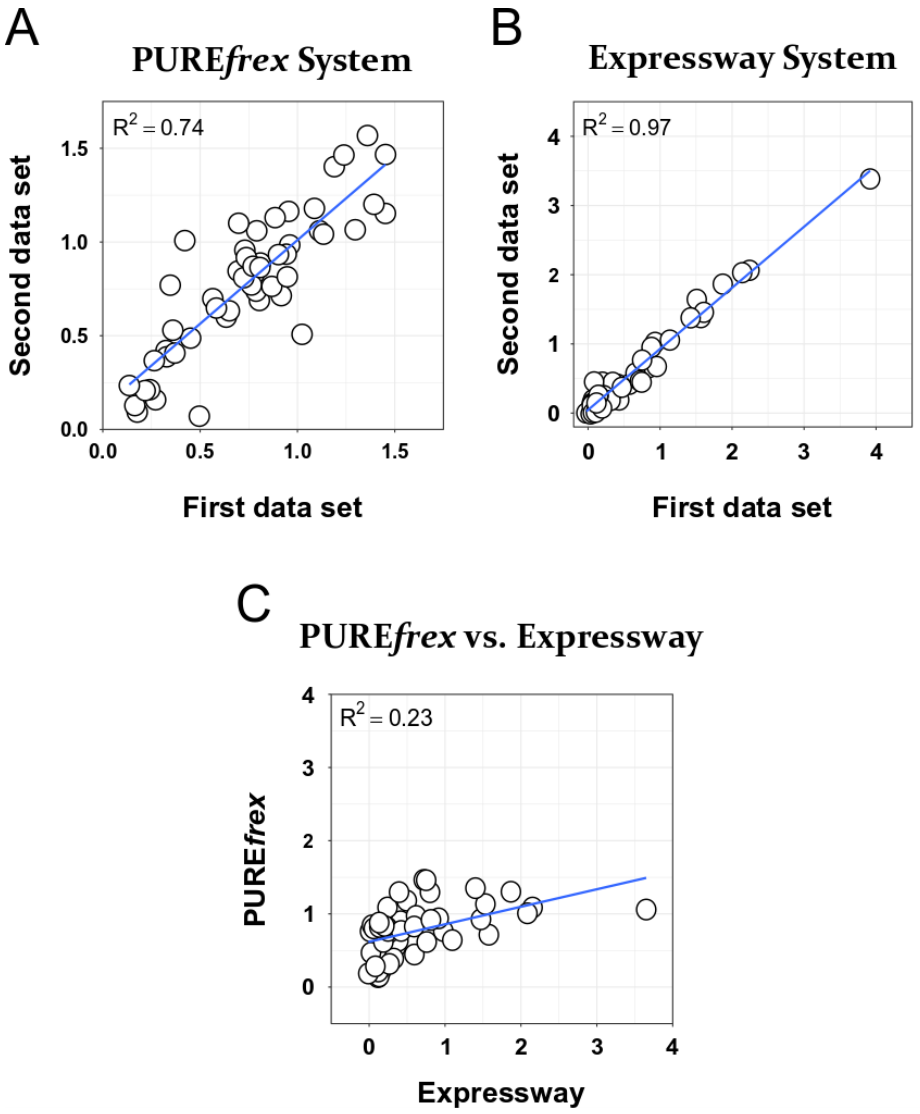


Figure 3.F: Scatter plots of the relative sfGFP fluorescence for each of the variants. (A and B) Scatter plots of duplicated experimental data for the PURE and extract-based systems, respectively. (C) Scatter plot of the averaged values (of both days) obtained with the PURE system against those of the extract-based system. The coefficient of determination R^2 was shown in the upper left part of the plots. The blue line indicates the regression line.

A previous study [5] pointed out that the PURE system could produce ~4-fold more protein than an *E. coli* extract-based system. Since some proteins in the PURE system were very likely incompletely translated or even non-functional and the

reproducibility levels were not very high, our data shows consistency with the previous study by displaying approximately a 3-fold increase in protein expression levels compared to that with the extract-based system (Figure 3.J).

These results also show that there is not much correlation between each variant's promoter strength of both systems (Figure 3.F, C; R^2=0.23), confirming our previous observation linked to the different compositions of each system. Although some outliers' samples can be seen in the plot (more easily seen in Figure 3.I), omitting these values will most likely result in different values for R^2. In order to identify better if the outliers should be omitted or not, further experimental data should be obtained first, to confirm the reproducibility values obtained and to determine why they are outliers in the first place.

3.6 Substitution Location within the Promoter Region

To study the effect of the position of the substitutions in the variants, the data is also displayed as heat maps of relative protein expressions (Figure 3.G, A and B).

This representation shows that most of the variants that show improvement in the levels of protein expression (higher levels than the core sequence) are the ones located in the upper region of the promoter. This region corresponds to the polymerase binding domain, where the T7 RNAP makes its first contact with the promoter sequence (position -17 to -13) [12]. In contrast, substitutions that follow this mentioned region (-9 and -5 for PURE and -11 and -4 for the extract-based), significantly reduce the protein expression level. This is consistent with previous

studies that show that bases around the -12 to -5 region, are crucial for the T7 RNAP to recognize the promoter sequence [7,13], and thus, substitutions introduced around this area are prone to interfere negatively with the protein-promoter recognition. Even further downstream of the regions already mentioned, substitutions affected more negatively to protein expression in both cell-free systems, but substantially more in the Expressway System.

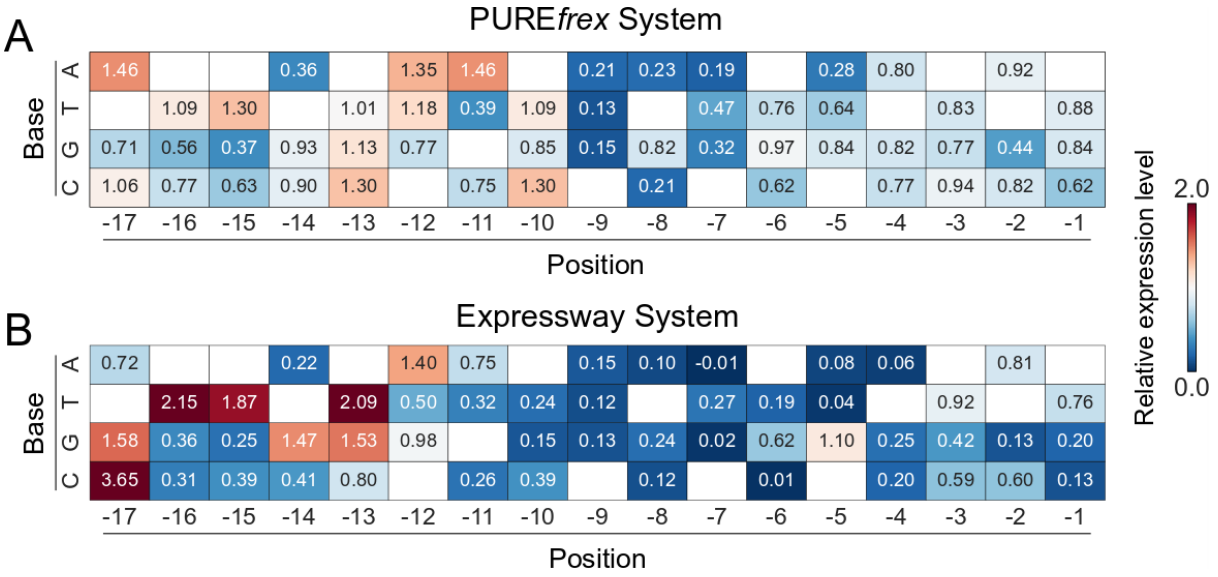


Figure 3.G: Heat maps of the relative sfGFP fluorescence of the 51 T7 promoter variants. In the PURE (A) and extract-based systems (B). The normalized data is also presented in Figure 3.E (A and B). Colours represent fold change in the final expression level relative to that of the consensus promoter.

3.7 Effects of Extended Sequence Length

As mentioned in Section 3.3, one of the reasons believed to affect the lower general yield of the extract-based system, along with the shorter time it takes to saturate;

might be the presence of nucleases that could degrade the DNA (and mRNA) quickly by attacking the 5' end where the promoter is located.

To test this hypothesis further and investigate the behaviours of the variants, a new different type of DNA construct was produced. This new version included an additional 20 base-pair random sequence located upstream of the T7 promoter (at the 5' end); and additional bases downstream the sfGFP gene (at the 3' end). The sequence added at the 3' end is that of the T7 promoter, to test if its presence can affect the overall protein yield. From this point, these new sequences will be referred to as "extended sequence" (Figure 3. H).

Since the variants located in the regions -17 to -10 seemed to be more significant to the final protein expression levels (whether beneficial or deleterious), it was decided to only construct the extended versions of those variants.

As done before with the non-extended sequences, a heat map of the normalized values of the new extended constructs was made (Figure 3.H). Overall, the fluorescence values of the variants decreased in most of the cases when being compared to the values of the core sequence.

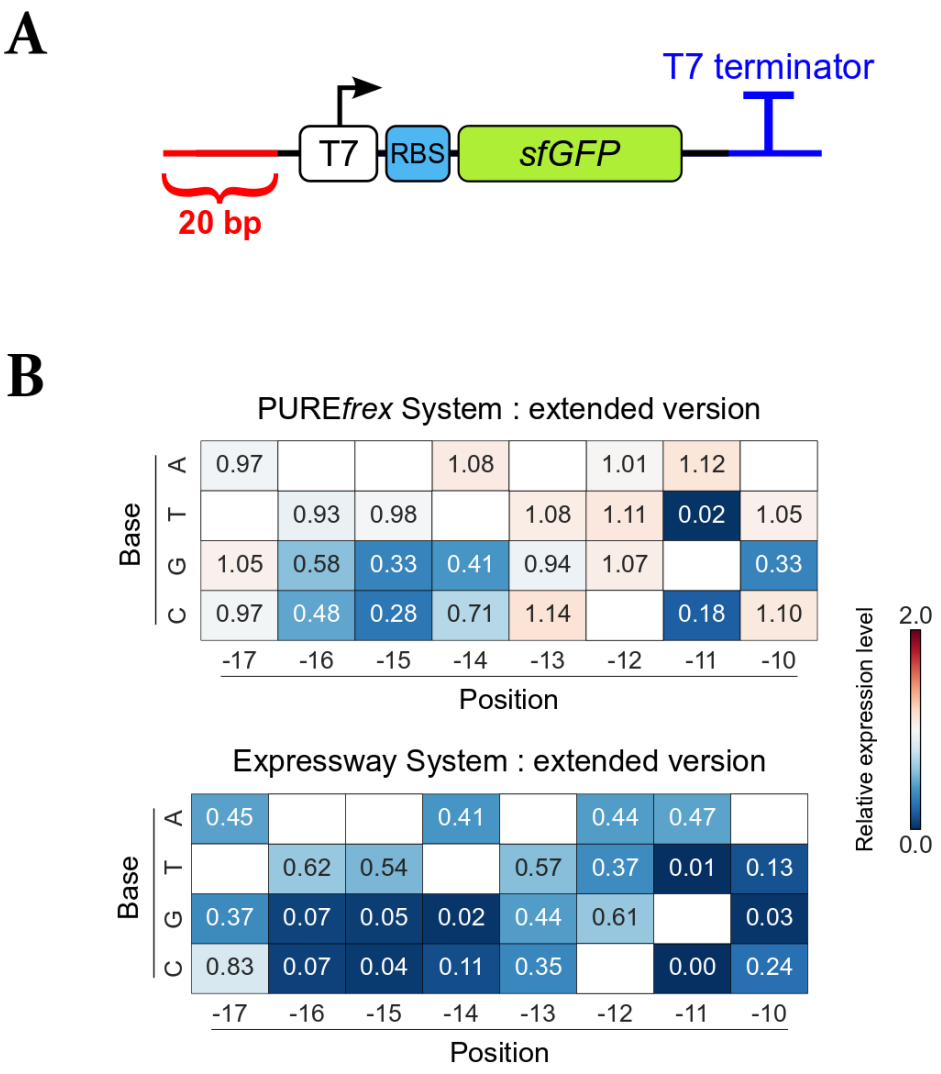


Figure 3.H: Schematic of the extended linear DNA templates and heat maps of the relative GFP fluorescence of 21 T7 promoter variants with extended base pairs at 5' and 3' ends. (A)The extended version features the core sequence (Figure 3.B) with extra only bases at the 5'-end (red); or with those plus extra bases at the 3'-end (dark blue). The sequence attached at the 3'-end contains a T7 terminator sequence. Each DNA sequence contains the consensus core sequence (or one of its 51 variants) (white), a ribosome binding site (RBS) (light blue) and an sfGFP gene (green). (B) For the PURE system (upper) and extract-based system (lower). Colours represent fold change in the final expression level relative to that of the consensus promoter.

This data shows that the extra bases had a more positive outcome for the core sequence (10-fold increase) that it had for the rest of the variants (average 4-fold increase) (Appendix, Section A.1, Table A.1.1.), therefore making the normalized values lower than those of the sequences without extra bases. A possibility for the decrease in these values would be a more stable binding of the T7 RNAP to the linear DNA due to the extra bases. As mentioned in Section 3.3, most substitutions located in the -17 to -10 region in the T7 promoter could have helped the T7 TNAP attach itself more securely to the DNA; so, it could be that the extra bases added at the 5'-end also had a similar, but lower, effect on the T7 RNAP, therefore cancelling the greater beneficial effects of the base substitution. However, the correlation between extended and non-extended sequences showed good values when it comes to the protein expression levels for both cell-free systems (Appendix, Section A.1., Figure A.1.3). This suggests that the gene expression levels by the T7 variants are consistent overall. In general, the regression line value shows a positive correlation between protected and unprotected samples, therefore proving that the extra bases protection was universal for all variants.

3.8 Absolute Fluorescence Levels and Rate of Reaction

Even though each extended sequence variant's normalized fluorescence levels decreased, in general, when comparing the absolute expression levels between protected and unprotected samples, the fluorescence values increased more than 4-fold on average for the extract-based system when having extra bases (Figure 3.I,

A) (Appendix, Section A.1, Table A.1.1.). But for the PURE system, the sfGFP levels for the extended sequences were similar to those with no extra bases (Appendix, Section A.1, Table A.1.2.).

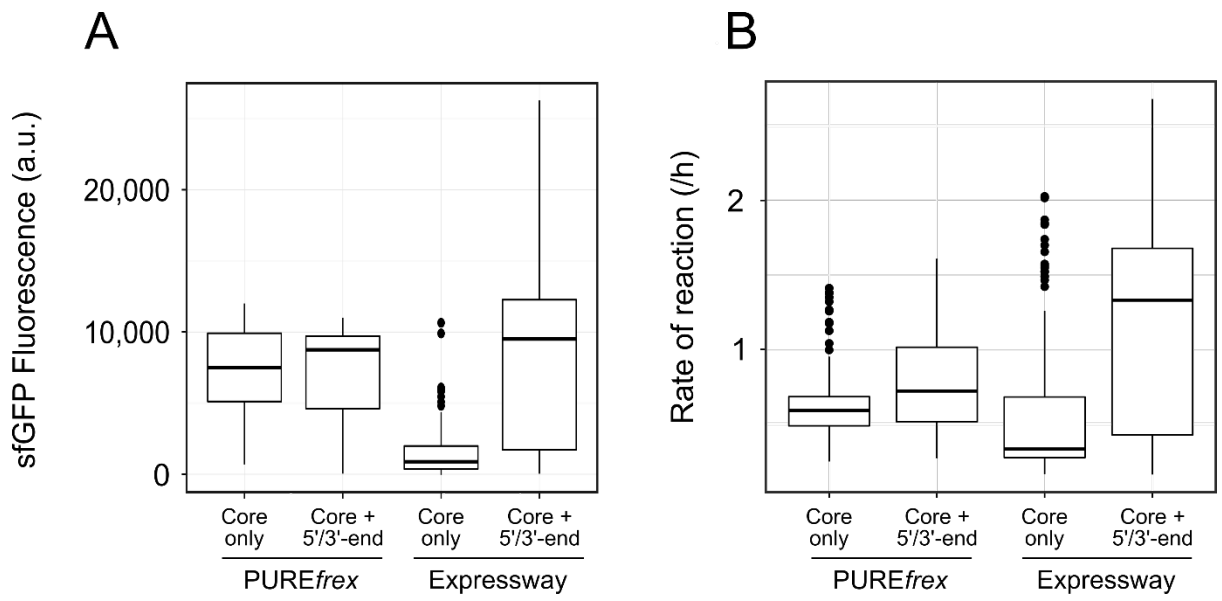


Figure 3.I: (A) Boxplot of absolute sfGFP fluorescence for the normal and extended DNA constructs in the PURE and extract-based systems. (B) Boxplot of the rate of reaction from all DNA constructs. The rate was calculated as the slope of a logistic curve fitted to individual fluorescence time-course data. The black thick line represents the median, and the box shows the first and third quartile. The upper and lower whiskers indicate 50% of the values higher or lower than the median, respectively. Black dots are outliers. Notes: (A) fluorescence expression levels (a.u.) are the fluorescence levels (numbers) detected from the sfGFP protein. This correlates to the amount of mRNA being transcribed and therefore, to the strength of the promoter which participates in said transcription. (B) rate of reaction is the speed at which a reaction takes place, in this case, the speed at which the sfGFP protein is being produced. This is calculated by dividing the change in fluorescence values in a specific amount of time. The faster a protein is being produced, the higher this number will be.

Additional bases on each end of the extended sequenced could help to postpone the degradation by nucleases of the T7 promoter sequence as well as some coding regions of the sfGFP gene in the Expressway system. Since the PURE system does not contain DNases as it is made out of purified components, the added bases do not provide significant extra protection and therefore the expression levels remain the same. One way to test this theory could be to intentionally add nucleases to the PURE system and text whether the extended sequences do help in this case.

Second, the T7 terminator introduced on the 3'-end of the extended sequence could have also helped in the increase of protein yield. It has already been reported that the T7 terminator improves the stability of transcribed mRNA in the *E. coli*-based cell-free system by improving the expression yield more than 3-fold [14]. To determine if the T7 terminator did influence the increase in protein expression, another construct was made with only extra added bases at the 5'-end. For the extract-based system, the results showed a small increase in protein expression (Figure 3.J, A), suggesting that even though the T7 promoter might help with the mRNA stabilization in this system, it is less significant than the extended random sequence at the 5'-end. In the case of the PURE system, the effect of T7 terminator appeared to be limited (Figure 3.J, B).

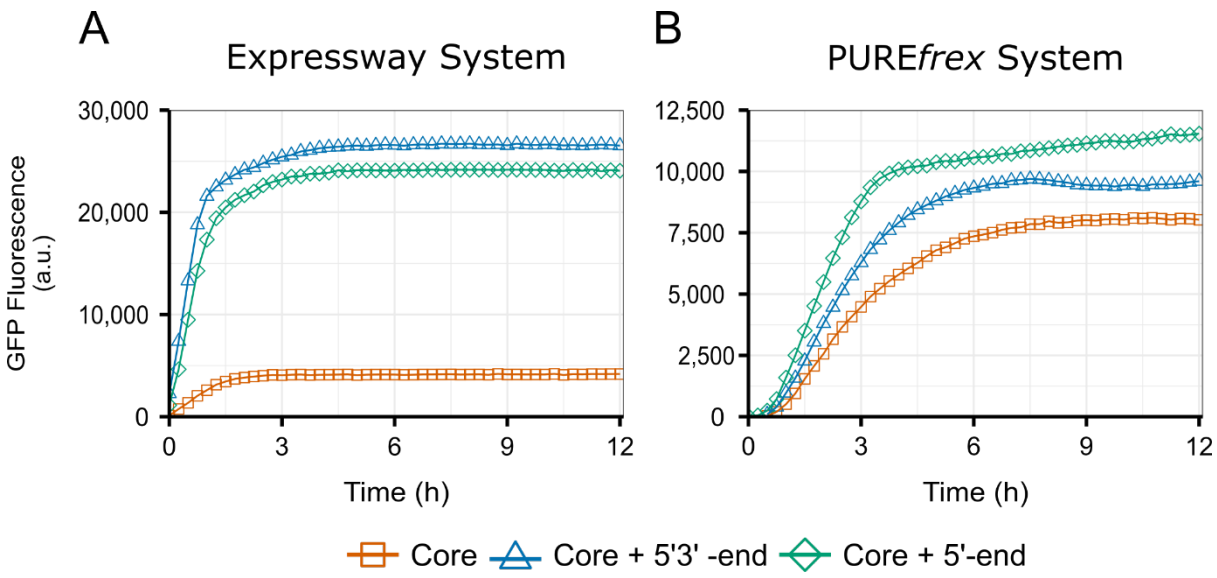


Figure 3.J: Time-course of GFP fluorescence over a period of 12 h using the T7 consensus sequence. Three different linear constructions, the core sequence (red squares), core sequence with extra bases at 5'-end (green diamonds) and core with extra bases at 5' and 3'-ends (blue triangles). (A) GFP expression in the extract-based system, and (B) the PURE system.

When looking at the rate of protein expression (Figure 3.I, B) of the normal and extended constructs in both cell-free systems, the extended sequence in the extract-based system showed a major increase compared to the non-extended sequence (1.3/h and 0.3/h respectively). In the PURE system, only a small increase was observed (0.8/h for extended and 0.6/h for non-extended).

When the rate and absolute expression levels (Figure 3.I) are plotted against each other, they show different modes of protein expression dynamics (Figure 3.K, A and B). In the case of the PURE system, the plot shows an exponential-like profile; in the low expression region (weak promoter variants) whereas, in the high expression region (strong promoter variants), the variants showed a wide range of

expression rates, although it appeared to have a limit of maximum expression level around at 12,500. On the other hand, when it comes to the extract-based system, the plot shows a diagonally-rotated profile. That is, if we take the plot from the PURE system and rotate in a diagonal (from lower left to upper right), the plot from the Expressway System appears. The rate of reaction seems to have a limit around 2/h while several variants were showing no fluorescence and thus very low rates of expression.

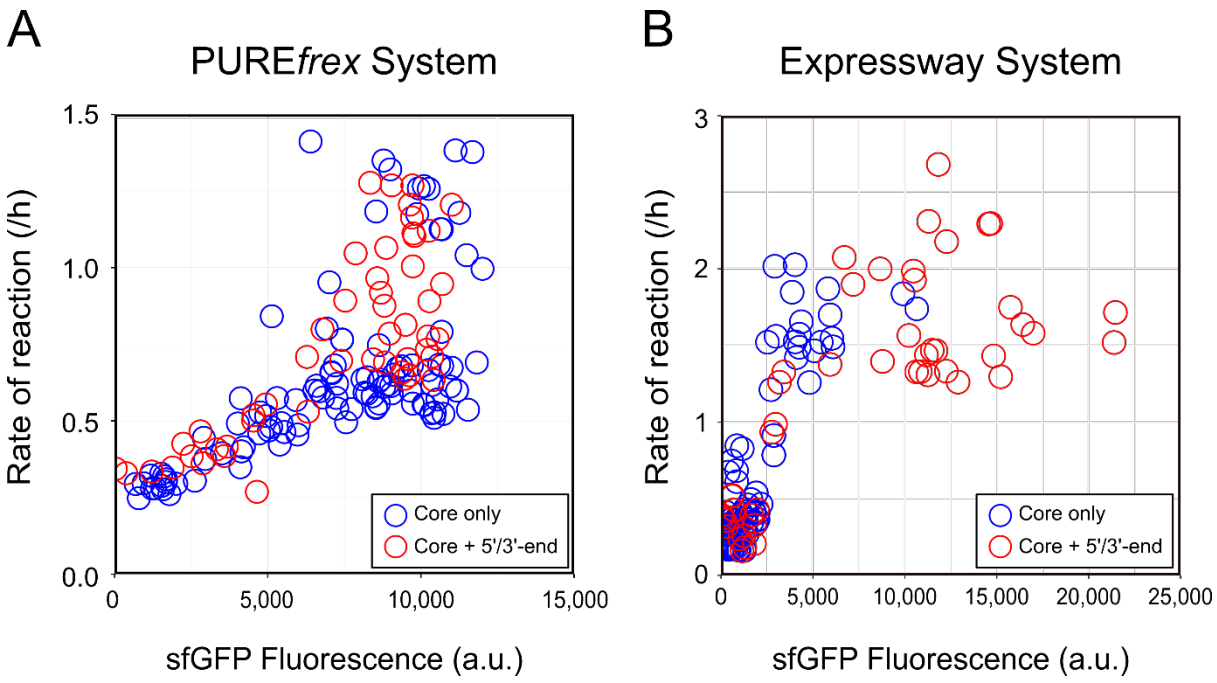


Figure 3.K: Scatter plots of the rate of protein expression against the absolute GFP fluorescence for the PURE (A) and extract-based systems (B). The same data as Figure 3.J. was used for the plots. Blue and red circles indicate the data for T7 promoter variants with the core and extended sequences, respectively.

To summarize the results obtained during the experimental part of the chapter:

- T7 promoter variants were tested in two different cell-free systems: extract-based and from purified components. The results showed a wide range of protein expression levels. Some of them showing higher levels than the core sequence, 1.5-fold and 3.6-fold increase in the PURE and Expressway system respectively.
- The reproducibility level of both systems was quite different. The extract-based system showed an R^2 value of 0.97 while the PURE system had a value of 0.74. This is probably due to the lack of ribosome rescue systems in the PURE system, as opposed to the extract-based one. This causes early termination of proteins and decreases the reproducibility rate.
- The mutations located around the -17 to -13 promoter region were most beneficial. This region is the polymerase binding site to the promoter. While mutations around the -11 to -4 region were mostly detrimental. This region is the polymerase recognition site of the promoter.
- Extension of the DNA constructs was performed and their protein levels tested. In the PURE system, the fluorescence levels stayed mostly the same and so did the rate of reaction (from 0.6/h to 0.8/h). But in the Expressway system, there was a 4-fold increment in the fluorescence level and the reaction rate raised from 0.3/h to 1.3/h on average. The extended sequence most likely helps delay the degradation of the DNA and mRNA by nucleases found in the cell-extracts, which are not present in the PURE system.

3.9 Cell-Free Systems Mathematical Models

Each sfGFP measurement time-course curve was fitted with model equations and the parameter values, k_{TX}, k_{TL}, d_D, d_R and d_P estimated. This systems approach was performed to better understand the dynamics of protein expression in cell-free systems. Cell-free systems are considered a grey-box model, there is previous knowledge to feed the model. A model of protein expression in a cell-free system can be described with the following equations (Equation 3.A, 1-3) [15–17]. A modified equation (Equation 3.A, 3* instead of 3), was used instead, to fit the PURE system data. The use of equation 3* instead of 3 for PURE will be explained later on.

$$(1) \quad \frac{\mathbf{d[DNA]}}{\mathbf{dt}} = -\boldsymbol{d_D}[\boldsymbol{DNA}]$$

$$(2) \quad \frac{\mathbf{d[mRNA]}}{\mathbf{dt}} = \boldsymbol{k_{TX}}[\boldsymbol{DNA}] - \boldsymbol{d_R}[\boldsymbol{mRNA}]$$

$$(3) \quad \frac{\mathbf{d[Protein]}}{\mathbf{dt}} = \boldsymbol{k_{TL}}[\boldsymbol{mRNA}] - \boldsymbol{d_P}[\boldsymbol{Protein}]$$

$$(3^*) \quad \frac{\mathbf{d[Protein]}}{\mathbf{dt}} = \boldsymbol{k_{TL}}[\boldsymbol{mRNA}](\mathbf{1} - \frac{[\boldsymbol{Protein}]}{\boldsymbol{K}}) - \boldsymbol{d_P}[\boldsymbol{Protein}]$$

Equation 3.A: Equations used for the mathematical models of protein expression for both cell-free systems. For the extract-based system, the equations used were 1 through 3; but for the PURE system the equations used were: 1, 2 and 3*.

In these equations, [DNA], [mRNA] and [Protein] are the concentrations of DNA, mRNA, and protein in the cell-free reaction. k_{TX} and k_{TL} are reaction constants for transcription and translation, respectively. d_D, d_R, and d_P are degradation constants

for DNA, mRNA, and protein, respectively. To estimate the system parameters, the inputs (concentration of template DNA) were varied and changes in protein expression levels were measured [16]. However, it is also possible to modify other parameters by adding/removing system components. For example, the addition of GamS (RecBCD nuclease inhibitor) to a crude extract-based cell-free system prevents DNA degradation [18], which corresponds to decreasing d_D in the equation (1) above (Equation 3.A). In this model, the transcription rate k_{TX} was modified and changes in protein expression levels measured. It is known that alterations to the consensus promoter sequence affect the binding affinity of T7 RNA polymerase (RNAP) to the promoter and thus changes the protein expression level [7,19,20]. Protein expression dynamics of these two types of cell-free systems are considered to be different owing to the lack of supplementary components in the PURE system [4,8,21]. So it is possible that the two cell-free systems would give different protein expression patterns in response to the varied transcription rates k_{TX}.

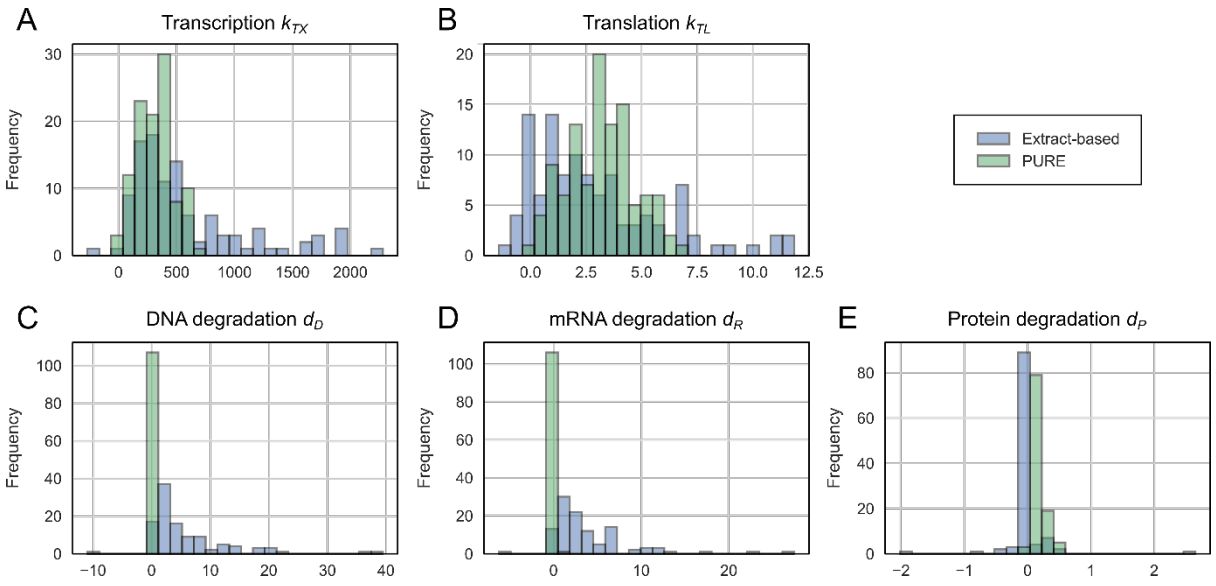


Figure 3.L: Histograms of estimated parameter values. (A) Transcription constant k_{TX}, (B) translation constant k_{TL}, (C-E) degradation constants for DNA d_D, mRNA d_R, and protein d_P, respectively. Note that different models were used to fit the parameters for the PURE and extract-based cell-free systems.

The fitted parameter values were consistent with the possible mechanisms behind the observed protein expression discussed in previous sections. The transcription parameter k_{TX} showed a wide distribution (over three orders of magnitude) (Figure 3.L, A) since each substituted base in the T7 promoter sequence changes the binding affinity of the T7 RNAP. On the opposite side, most of the translation parameter k_{TL} (Figure 3.M, B) fit within the same order of magnitude. The DNA/mRNA degradation parameters, d_D and d_R, between the two cell-free systems are different (Figure 3.L, C and D). The extract-based system shows a wider distribution than the PURE system, which had only one single peak. This could indicate that both the DNA and mRNA degrade very slowly in the PURE system,

as it lacks nucleases; but since they are present in the extract-based system, both DNA and mRNA are quickly digested. In both systems, the protein degradation value d_P (Figure 3.L, E) was close to zero. This is consistent with the experimental data as the measured GFP fluorescent signal did not decrease over a long period (Figure 3.D and Appendix, Section A.1, Figure A.1.1.).

When looking at the histograms of figure 3.L, few parameter values, especially those of degradations are in the negative range. Since the mathematical model tends to change the parameters linearly, it is possible that if some of experimental observed values are quite close to 0, the error range of the model can go into negative values, even though it is not biologically possible. The presence of outliers also makes the error range bigger. Due to the simplicity of the, modifications are needed to include the characteristics of complex non-linear system and also setting up rational upper and lower limits.

Altering the promoter sequence should correspond to modifying the transcription parameter k_{TX}, since the substitution in bases of the promoter affects directly the binding affinity of the T7 polymerase to the promoter, and therefore the transcription process gets influenced. Consequently, to simulate the different protein expressions levels by all the promoter variants, the parameter values were fitted for the consensus promoter sequence and only k_{TX} was changed from zero to the maximum fitted value.

The simulated protein expression in both systems is shown in Figure 3.M (left column for the extract-based system and the right column for the PURE system).

The simulation reproduces the behaviour of the protein expression in both systems; such as the halting of protein expression after 3-4 h in the extract-based system (Figure 3.M, A). For figure 3.M, the parameter K_{tx} was varied while the other remained fixed, since the K_{tx} represents the transcription constant and therefore is the value that most directly represent the effect the mutations on the T7 promoters had. The other parameters, can only be estimated and would have to be experimentally obtained to be able to vary then in the model more accurately.

When calculations were made for the rate of reaction and the maximum protein expression level (Figure 3.N), by fitting the simulated data to a logistic curve, it displayed a similar profile as shown in Figure 3.K.

The simulation which was performed for the PURE*frex* system data was done with a modified version of the equation (Equation 3.A; 1, 2 and 3*) instead. When performed using the same equations used for the extract-based system (Appendix, Section A.1, Figure A.1.4), the temporal expression patterns obtained matched the observed temporal expression profiles, but it did not reproduce the unique scatter pattern in Figure 3.K, A. This result implied that some factors were not being taken into consideration in the already mentioned theoretical model.

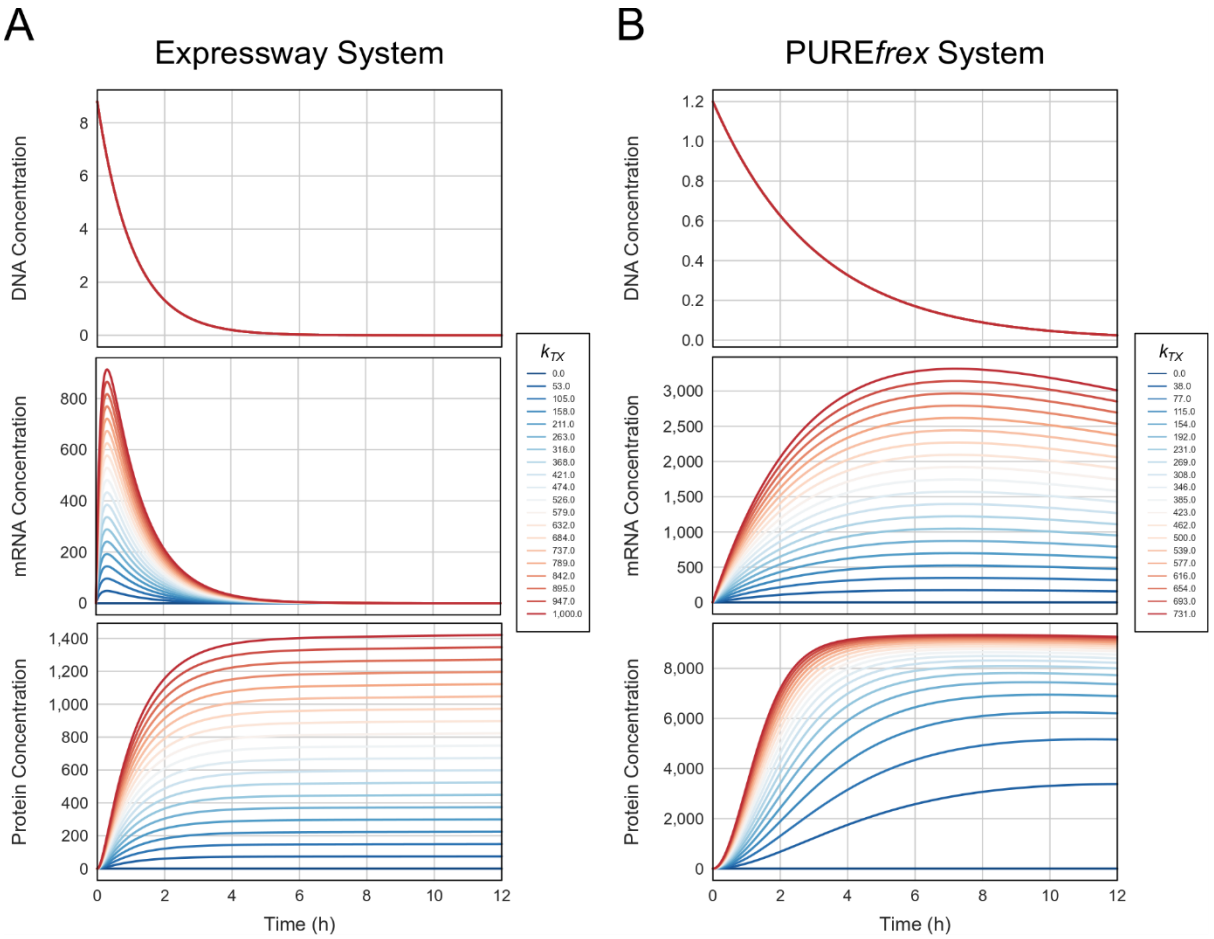


Figure 3.M: Simulated cell-free protein expression. Simulated cell-free protein expression of the extract-based system (A) and the PURE system (B). DNA concentration (top), mRNA concentration (middle), and protein expression (bottom) were shown. The parameter k_{TX} was varied (colours) while the other parameters were fixed. A modified equation (see text for details) was used to simulate protein expression for the PURE system.

As mentioned in Chapter 1, the main limitation concerning cell-free systems is the limited amount of energy sources and the accumulation of inhibitory by-products [5]. As further proof of this issue, when a cell-free reaction is being fed energy molecules and by-products are eliminated through dialysis, the final protein yield can increase up to 72-fold [22]. Also, recycling of inorganic phosphate can improve the total protein yield [23]. All these issues should be considered as negative feedback

to the protein production, so to incorporate this factor, the last equation was modified so that the term $(1 - [Protein]/K)$ represents the negative feedback. The proteins stop being produced as the protein expression level reaches the maximum defined by the carrying capacity constant K. Using the equations (1), (2), and (3*), the model was refitted for the experimental data to estimate the system parameters $k_{TX}, k_{TL}, d_D, d_R, d_P$, and the carrying capacity K. Figure 3.N (B) shows the simulation results performed with equation 3* for PURE and 3 for Expressway, for the rate of reaction against the maximum expression level.

In the extract-based system, mRNA was produced very rapidly and peaked around at 30 min after the start of the reaction and then decreased exponentially since the template DNA was degraded by nucleases present in the reaction mixture, and eventually reached to the base level at around 4 h. Whereas in the PURE system, the template DNA gets degraded at a much slower rate, resulting in the rate of mRNA production to be slower. Most of transcribed mRNA was still present even after 12 h.

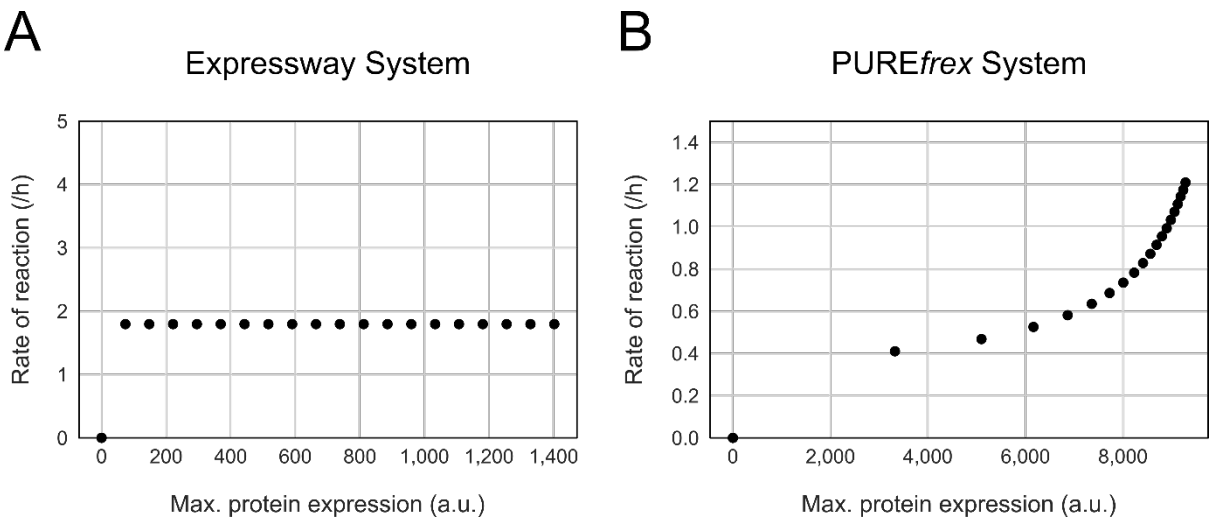


Figure 3.N: Scatter plots of the rate of protein expression against the maximum protein expression. The values were obtained by fitting the simulated data in Figure 3.N. (A and B) to a logistic curve.

When the parameter k_{TX} was linearly varied, the protein expression yield in the extract-based system also changed linearly because there were no nonlinear terms in the model equation. In the PURE system though, the response to this linear change was nonlinear because of the negative feedback already explained. This makes the system very sensitive to small changes in parameter values, which implies that the maximum expression levels could be biased to the higher region. Combined time-course plots for both cell-free systems (Appendix, Section A.1, Figure A.1.2) show similar patterns as the simulated cases in Figure 3.M, especially the cases with the extended sequences. Although it was not possible to experimentally measure k_{TX} (the binding affinity of T7 RNAP to each promoter variant sequence), these results may also validate the models.

3.10 RNA Spinach Aptamer

To look more deeply into the temporal dynamics of the mRNA, a final DNA construct was made. This consisted of two of the typical components, T7 promoter and RBS, but with an added Spinach aptamer (Figure 3.O), providing green fluorescence in the presence of a specific fluorophore, such as 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) [24,25]. Due to the green fluorescence of the aptamer, it was not possible to include the sfGFP gene in the DNA template as both peaks of emission wavelength overlap. Only a few constructs were selected for this experiment: the consensus T7 promoter sequence, two strong variants (17A and 11A for PURE; 17A and 16T for extract-based) and a weak variant (16C for both).

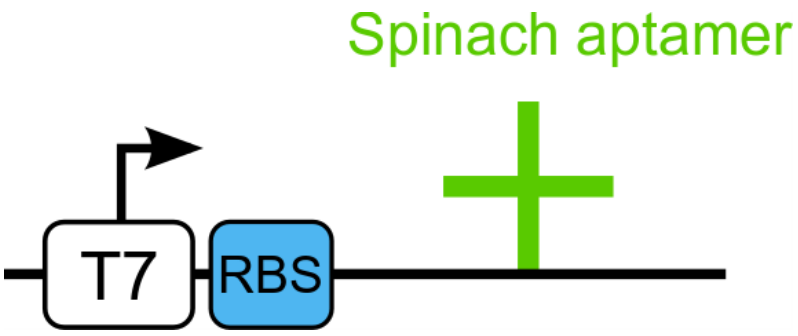


Figure 3.O: Linear DNA template with Spinach aptamer. The sequence contains the T7 promoter or its variant, RBS, and Spinach aptamer. It does not contain any genes to be expressed.

The measured mRNA concentration profile (Figure 3.P) was similar to the simulated mRNA concentration in both cell-free systems. It should be noted that no experimental information about mRNA concentration was included when

fitting the differential equations to the protein expression data. Nevertheless, the time-courses of mRNA concentration in the simulation reproduced those in the experiments. This result validates the theoretical models as a representation of the dynamics behind the cell-free protein expression.

Although it is known that the fluorescence of DFHBI molecules are sensitive to pH changed in the reaction, the pH was not monitored in this case. Therefore, the fluorescence levels obtained might not be fully representative of the mRNA levels. But since cell-free system try to mimic reactions that take place inside a cell, it is probable that the pH of the reaction did not vary too much outside a considerable range.

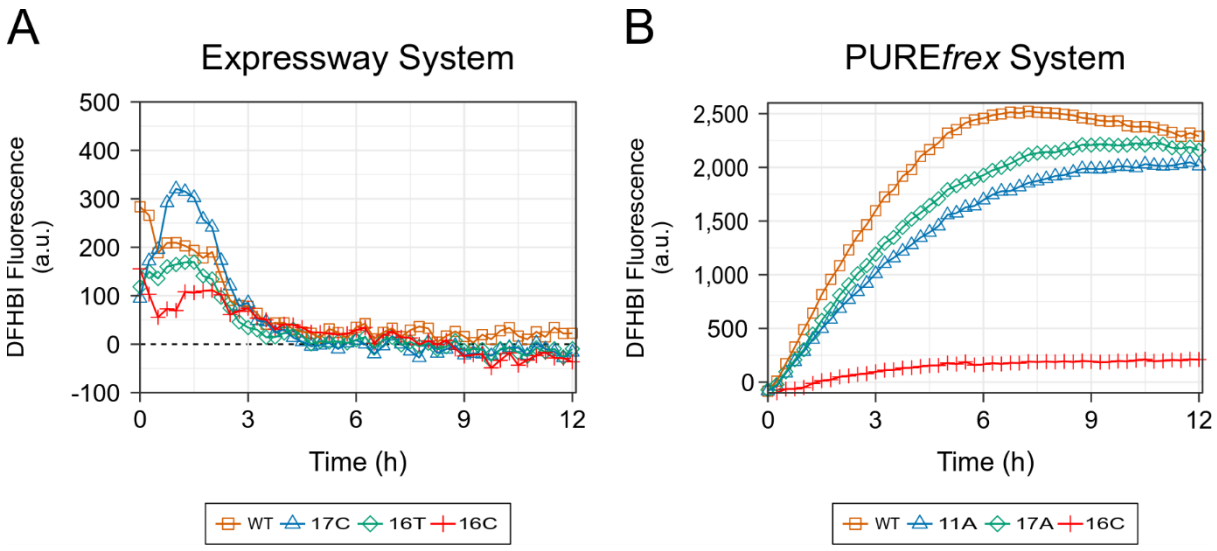


Figure 3.P: Fluorescence measurements of Spinach aptamer with the consensus promoter sequence (denoted as WT) and other three variants in the PURE and extract-based systems, respectively.

3.11 Conclusions

In this chapter, the behaviour of single base-pair substituted variants of the T7 promoter in two different cell-free systems was characterized. Although further understanding of cell-free protein expression mechanism at the molecular level is required to fully explain the complex expression patterns observed here, the experiments revealed different modes of protein expression dynamics in the PURE and extract-based systems, which were confirmed by analytical models. It was shown that even simple time-course measurement data of cell-free systems contains rich information and the combination with systems approach helps uncover the dynamics behind it.

These observations illustrate that the two cell-free systems have different limiting factors for improved protein expression. In the extract-based system, the protein expression was primarily limited by the availability of template DNA and transcribed mRNA. As they were quickly digested by nucleases in the reaction mixture, any method to prevent or slow down the degradation would improve the yield. For example, the addition of salmon sperm DNA could “occupy” the nucleases and slow down the degradation of the DNA as an easier and maybe more affordable way. But whenever possible, it is always better to keep the experimental reactions a simple as possible.

This is consistent with the experimental results in previous sections that the extended sequence showed significantly improved yield. In the PURE system, the limiting factors lie within the mRNA-protein translation. In the current form, the

negative feedback term in the model is still ambiguous and needs further biological elaboration. However, it has been shown that supplementing certain components like amino acids, tRNAs, magnesium, and proteins involved in ribosome recycling to the PURE system improves the final yield [5]. This is consistent with our model prediction.

As it has been shown before by previous studies, the systems approach has proven effective for engineering synthetic genetic circuits in the cell-free system [16,26]. This approach can also be effective to disentangle the complex dynamically-interacting factors in the cell-free systems and obtain deeper insights that are otherwise difficult to capture.

3.12 References for Chapter 3

1. Du, L., Gao, R. & Forster, A. C. Engineering multigene expression in vitro and in vivo with small terminators for T7 RNA polymerase. *Biotechnol. Bioeng.* (2009). doi:10.1002/bit.22491
2. Vethanayagan, J. G. G. & Flower, A. M. Decreased gene expression from T7 promoters may due to impaired production of active T7 RNA polymerase. *Microb. Cell Fact.* (2005). doi:10.1186/1475-2859-4-3
3. 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). doi:10.1093/nar/gks940
4. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–5 (2001).
5. Li, J. *et al.* Dissecting limiting factors of the Protein synthesis Using Recombinant Elements (PURE) system. *Translation* **5**, e1327006 (2017).
6. Chizzolini, F., Forlin, M., Cecchi, D. & Mansy, S. S. Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength. *ACS Synth. Biol.* **3**, 363–371 (2014).
7. Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* **39**, 10419–30 (2000).
8. Li, J., Gu, L., Aach, J. & Church, G. M. Improved cell-free RNA and protein synthesis system. *PLoS One* **9**, (2014).
9. Shimizu, Y. ArfA recruits RF2 into stalled ribosomes. *J. Mol. Biol.* **423**, 624–631 (2012).
10. Ude, S. *et al.* Translation Elongation Factor EF-P Alleviates Ribosome Stalling at Polyproline Stretches. *Science (80-.).* **339**, 82–85 (2013).
11. Doerfel, L. K. *et al.* EF-P Is Essential for Rapid Synthesis of Proteins Containing Consecutive Proline Residues. *Science (80-.).* **339**, 85–88 (2013).
12. Kochetkov, S. ., Rusakova, E. . & Tunitskaya, V. . Recent studies of T7 RNA polymerase mechanism. *FEBS Lett.* **440**, 264–267 (1998).
13. Jorgensen, E. D., Durbin, R. K., Risman, S. S. & McAllister, W. T. Specific contacts between the bacteriophage T3, T7, and SP6 RNA polymerases and their promoters. *J. Biol. Chem.* **266**, 645–651 (1991).
14. Ahn, J. H. *et al.* Cell-free synthesis of recombinant proteins from PCR-

amplified genes at a comparable productivity to that of plasmid-based reactions. *Biochem. Biophys. Res. Commun.* **338**, 1346–1352 (2005).

15. Karzbrun, E., Shin, J., Bar-Ziv, R. H. & Noireaux, V. Coarse-grained dynamics of protein synthesis in a cell-free system. *Phys. Rev. Lett.* **106**, (2011).

16. Siegal-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V. & Murray, R. M. Gene circuit performance characterization and resource usage in a cell-free 'breadboard'. *ACS Synth. Biol.* **3**, 416–425 (2014).

17. Ingalls, B. *Mathematical Modeling in Systems Biology: An Introduction*. (The MIT Press, 2012).

18. Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V. & Murray, R. M. Linear DNA for rapid prototyping of synthetic biological circuits in an *Escherichia coli* based TX-TL cell-free system. *ACS Synth. Biol.* **3**, 387–397 (2014).

19. Rong, M., He, B., McAllister, W. T. & Durbin, R. K. Promoter specificity determinants of T7 RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 515–519 (1998).

20. Chizzolini, F. *et al.* Cell-Free Translation Is More Variable than Transcription. *ACS Synth. Biol.* **6**, 638–647 (2017).

21. Hillebrecht, J. R. & Chong, S. A comparative study of protein synthesis in in vitro systems: from the prokaryotic reconstituted to the eukaryotic extract-based. *BMC Biotechnol.* **8**, 58 (2008).

22. Jackson, K., Kanamori, T., Ueda, T. & Hugh Fan, Z. Protein synthesis yield increased 72 times in the cell-free PURE system. *Integr. Biol.* **6**, 781 (2014).

23. Caschera, F. & Noireaux, V. A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system. *Metab. Eng.* **27**, 29–37 (2015).

24. van Nies, P. *et al.* Unbiased Tracking of the Progression of mRNA and Protein Synthesis in Bulk and in Liposome-Confined Reactions. *ChemBioChem* **14**, 1963–1966 (2013).

25. Chizzolini, F., Forlin, M., Cecchi, D. & Mansy, S. S. Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength. *ACS Synth. Biol.* **3**, 363–371 (2014).

26. Franco, E., Giordano, G., Forsberg, P. O. & Murray, R. M. Negative autoregulation matches production and demand in synthetic transcriptional networks. *ACS Synth. Biol.* **3**, 589–599 (2014).

Chapter 4. *In vitro* display techniques and cell-free systems: a refined and clean method for *in vitro* selection of drug targeted peptides

4.1 Chapter 4 Aims and Summary

Cell-free transcription-translation systems are excellent tools for protein expression. The PURE system, based on recombinant elements, has been shown to present high purity and high yield, and therefore it has been used for expression of both native and *de novo* proteins, as well as in experiments involving various display methods to select for functional polypeptides [1–3]. This chapter builds upon the knowledge gained in Chapter 3 to extend the use of cell-free systems with high-throughput screening methods.

A refined method was developed that involved the use of the PURE cell-free system and a display methodology for the preparation of stable mRNA and cDNA-peptide conjugates. The method was further validated for its use for *in vitro* selection processes. The conjugates of peptide-mRNA/cDNA obtained reached a formation efficiency of 40% (conjugation efficiency range is between 5% to 40% [4]). This was followed by precise gel purification to minimize the existence of carry-over of components, from the translation system, in the conjugate samples. Strict purification steps allowed for posterior clean and efficient sequence screening of the random peptide library. To test the efficiency of the method, the anti-FLAG M2 antibody was selected as a target for process validation. With a starting library of approximately 1.7×10^{12} random sequences and after 4 rounds of selection with two

defined display methods: mRNA and cDNA display; high-throughput sequencing showed enrichment of the FLAG epitope DYKDDD as well as a consensus FLAG epitope motif DYK(D/L/N)(L/Y/D/N/F)D. Furthermore, both mRNA and cDNA display methods resulted in a similar overall performance.

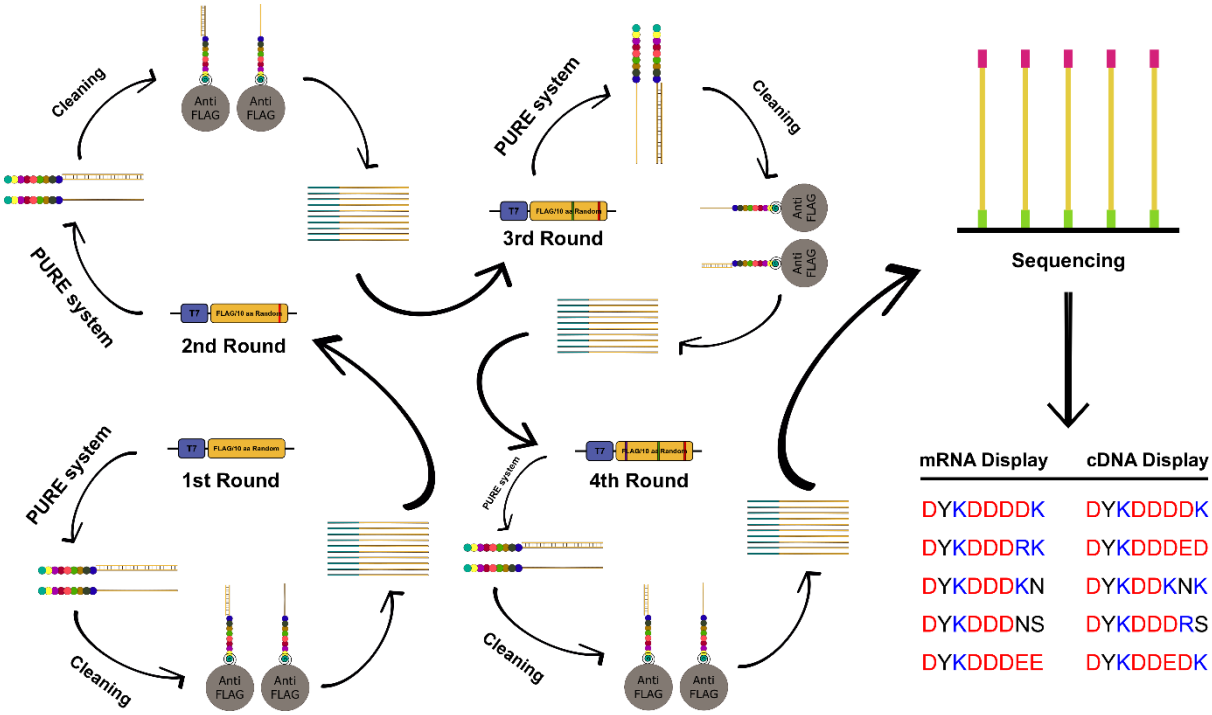


Figure 4.A: Summary of the experimental approach carried out for Chapter 4. From the bottom left corner, the 1st round of selection starts the experiment and it ends with the 4th round. Each round uses the PURE system for the translation step, meticulous cleaning of the constructs and two different display methods: mRNA and cDNA display. The selection process is performed using Anti-FLAG antibody beads. The final samples are sent for MiSeq sequencing and the results analysed.

4.2 Method Overview

The experiments carried out encompassed the development and optimization of two display methods: mRNA and cDNA display; utilizing a commercial version of the PURE protein expression system (PURE*frex* 1.0) [5,6]. During each round of selection, approximately 1,000 peptides were sequenced.

One of the important features of this method is the stability and high purity of both the mRNA-tag and mRNA/cDNA-peptide constructs, mainly due both the use of the PURE system during the translation step (as it is a nuclease- and protease-free environment) and the two-step gel purification performed (as it eliminates the majority of by-products).

The elimination of specific components and by-products (RNA, enzymes, polyamines, cofactors, and ions) from the samples during the selection rounds minimized non-specific interactions; with either the mRNA and cDNA-peptide conjugates or with other components; especially during the binding assay steps.

To test the validity of the method, a short FLAG epitope (DYKDDDDK) was selected due to its already known crystal structure [7]. Several display methods (phage, DNA and ribosome display) have already been successful in enriching the FLAG epitope motif [8–11]. One round of selection takes a minimum of 1 day (24 h) (Figure 4.B), which means the whole procedure can be completed in 1 week. Usually, for mRNA display, one round of selection takes around 2-7 days depending on the type of target. The high-throughput sequencing was performed after

collecting the DNA library samples from each round, to characterize the sequence variations after each round of selection.

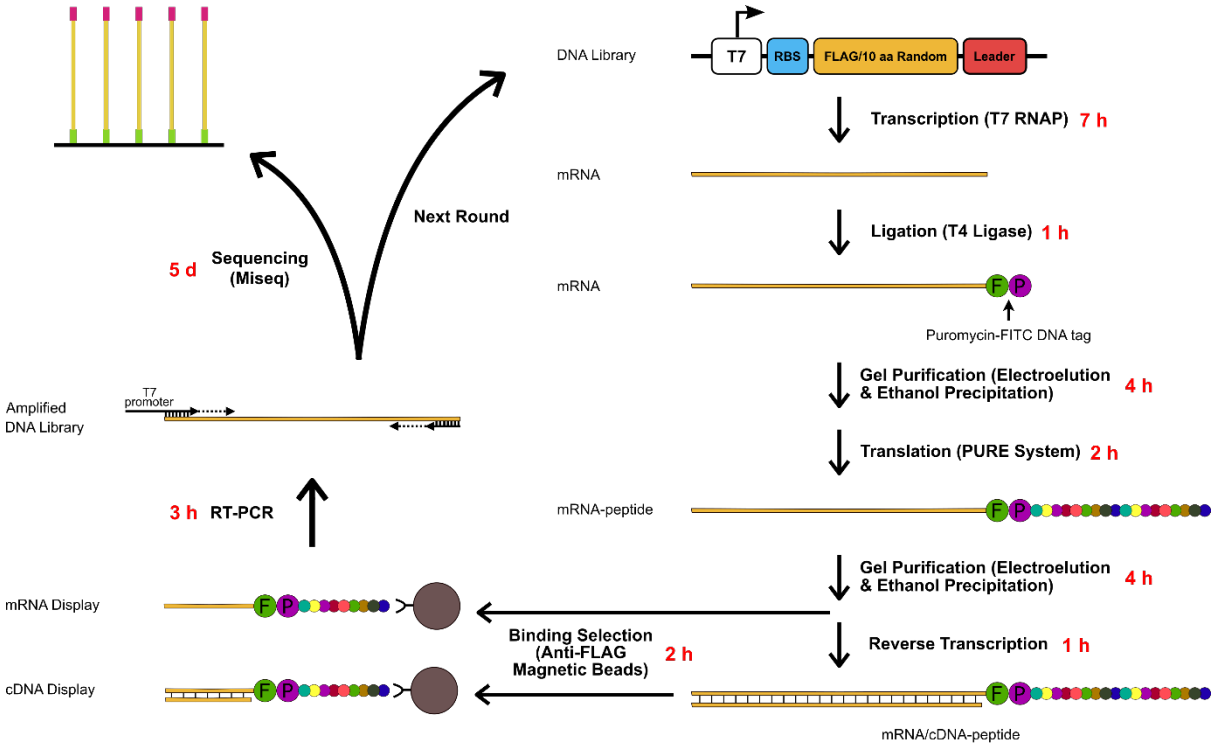


Figure 4.B: Schematic overview of one round of selection and the estimated time of each step. The initial round starts with a DNA library (designed with randomized sequences) that includes a T7 promoter upstream and a leader sequence (complementary to the puromycin-FITC-DNA tag) downstream. The DNA library is *in vitro* transcribed to an mRNA library and later ligated to the puromycin-FITC DNA tag. The resulting product purified with a gel and translated using the PURE*frex* 1.0 System. The resulting mRNA-peptide conjugate is purified a second time again by gel electrophoresis. The purified product could be used as it is (mRNA Display) or reverse transcribed (cDNA Display). Both products forego a binding assay against anti-FLAG M2 magnetic beads and the selected conjugates are further reverse transcribed to go into another round and later on, sequenced using the Illumina MiSeq system. Preparation time (hours) for each step is shown in red with exception of MiSeq sequencing (5 days) carried out only after all samples from all round were obtained.

4.3 DNA Library Design with Random Sequences

To oversee the trajectory of various sequences in each round, two different DNA libraries were designed. The 'FLAG-random library' contained a fixed sequence for the first five amino acids of the FLAG epitope (**DYKDD**DDK); followed by three consecutive degenerate "RRN" codons corresponding to seven amino acids (Asn, Lys, Asp, Glu, Ser, Arg and Gly) (Appendix, Section A.2, Figure A.2.1). The '10aa-random library' contained ten degenerate codons ("VNN" and "NNY") which corresponded to 16 and 17 different amino acids, respectively (Appendix, Section A.2, Figure A.2.1). The number of different sequences in both FLAG-random and 10aa-random libraries was 343 and 1.7×10^{12}, respectively. The FLAG-random library was used as a positive control to confirm the enrichment of the full FLAG epitope sequence in the early rounds. On the other hand, the 10aa-random library was used to validate the performance of the display method by exploring a wide range of random sequences to find several suitable ones that could bind to the anti-FLAG M2 antibody.

Since this method was used with the PURE system, both DNA libraries contained upstream a T7 promoter and ribosome binding site (RBS). The leader sequence at the 3' region was complementary to the puromycin-FITC DNA tag for an efficient ligation. The only difference between the two libraries was the randomized sequence regions.

4.4. T7 RNAP *In vitro* Transcription and T4 RNA Ligation

Following the construction of both DNA libraries, *in vitro* transcription with T7 RNA Polymerase was performed. After a purification step, the mRNA construct was ligated to the puromycin-FITC DNA tag with T4 Ligase enzyme. The ligation was carried out by the Y-ligation method [12] to connect the 3' end from the single-stranded mRNA construct to the phosphorylated 5' end of the DNA tag (Appendix, Section A.2, Figure A.2.2). The ligation procedure was confirmed by running the ligated products on 6 % polyacrylamide TBE gel with 8 M Urea and visualizing the band shift using SYBR Gold staining for the mRNA molecule and FITC fluorescence detection of the puromycin-FITC DNA tag (Figure 4.C).

After the ligation procedure, around 50% of the mRNA molecules were attached to the Puromycin-FITC DNA. The final product mRNA-tag product was further purified by gel using electroelution and ethanol precipitation to eliminate unligated products. The elimination of free non-ligated Puromycin-FITC DNA tag was important for the posterior translation of the mRNA-tag product since puromycin can interfere with the ribosomes.

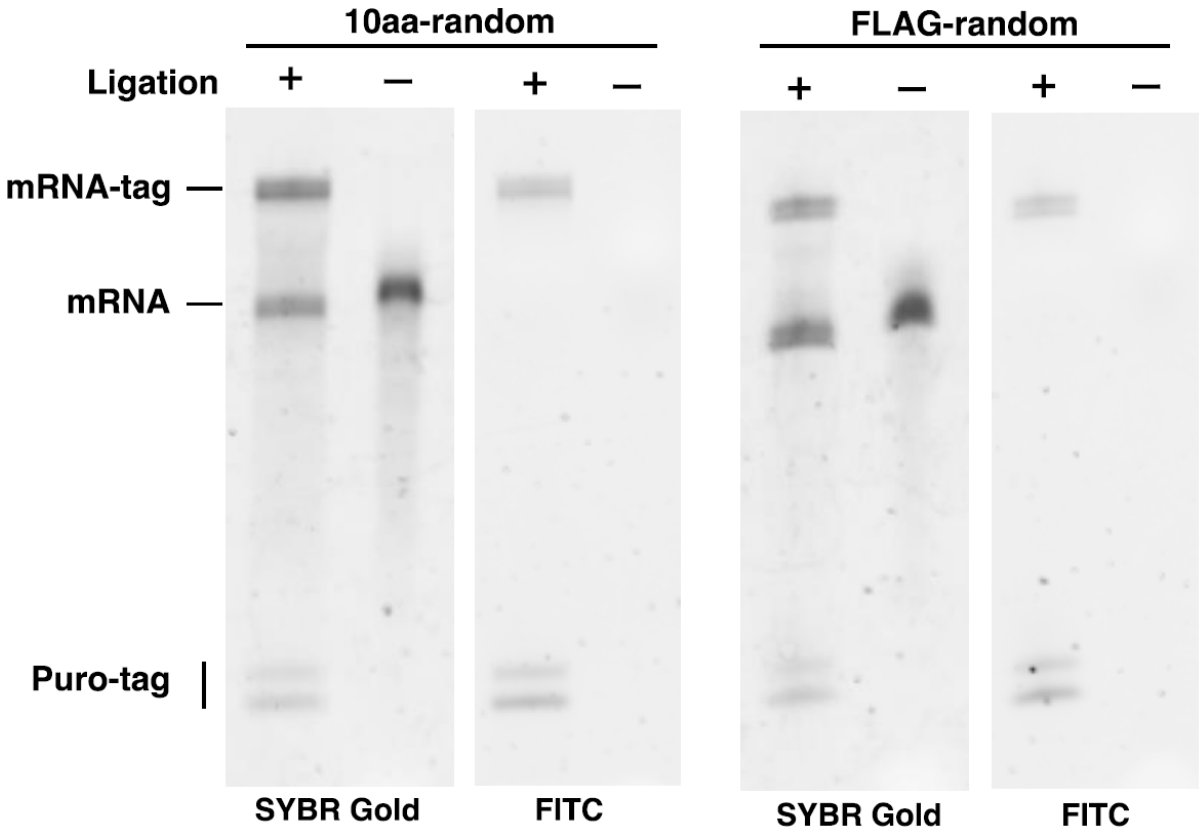


Figure 4.C: Puromycin-FITC DNA tag ligation gel images. Image of ligated samples resolved in 6 % polyacrylamide TBE with 8 M Urea gels with both SYBR GOLD staining and isothiocyanate fluorescence (FITC). The gels show both the FLAG-random and 10aa-random libraries samples with and without (negative control) the ligation process.

4.5 PURE system *in vitro* Translation Optimization

The recovered mRNA-tag from section 4.4 was posteriorly subjected to *in vitro* translation, to form mRNA-peptide conjugate constructs, using the commercial version of the cell-free PURE system, PURE*frex* 1.0. The PURE*frex* kit is known, as discussed in chapter 3, for its high purity as it does not contain nucleases and proteases and therefore is adapted for any display methods. The lack of nucleases already gives an important advantage for mRNA and cDNA display methods as it

allows for strong stability of the conjugates since their degradation in the reaction mixture is minimal. Since display methods and selection processes occur during the span of several days, the elimination of degradation, whether it is for nucleotides or peptides, is of high importance. Another important aspect of the PURE system is the high protein yield even if the rate of product formation depends greatly on other aspects, like incubation time and co-factors concentration. These conditions, if optimized, could help promote the efficient incorporation of the puromycin into the stalled ribosome at the mRNA-DNA border. Previous studies have tackled this by introducing rare codons at the 3' end [2] or by using long linkers in the puromycin-DNA tag [13]. So in this study, a rare GGA codon was placed at the 3' end and a polyA (18 nt) included in the puromycin-FITC DNA tag (Appendix, Section A.2, Figure A.2.2).

But even with these optimizations, as already mentioned, more can be done in terms of the incubation conditions. So before continuing with the selection round, a series of experiments were performed to optimize this process by testing 3 sets of different conditions in the translation reaction with the PURE system. The sample used for the optimization was the purified 10aa-random library mRNA-tag, and the parameters tested were: translation time, salt mix concentration (KCl and $MgCl_2$), and incubation time after salt addition.

The total incubation (translation) time reflects the time for the ribosomes to prepare and perform the translation of the coding region, which since it lacks a STOP codon, makes the ribosome stall at the 3' end of the mRNA. Since the optimal

incubation temperature for the PURE*frex* 1.0 System is 37 °C, that parameter was not changed and stayed the same. What was first tested was the initial incubation time before the addition of the salt mixture; from 0 to 60 minutes. It was found that after just 5 min, already 40% of the samples were translated; and further incubation did not increase significantly this number (Figure 4.D; A).

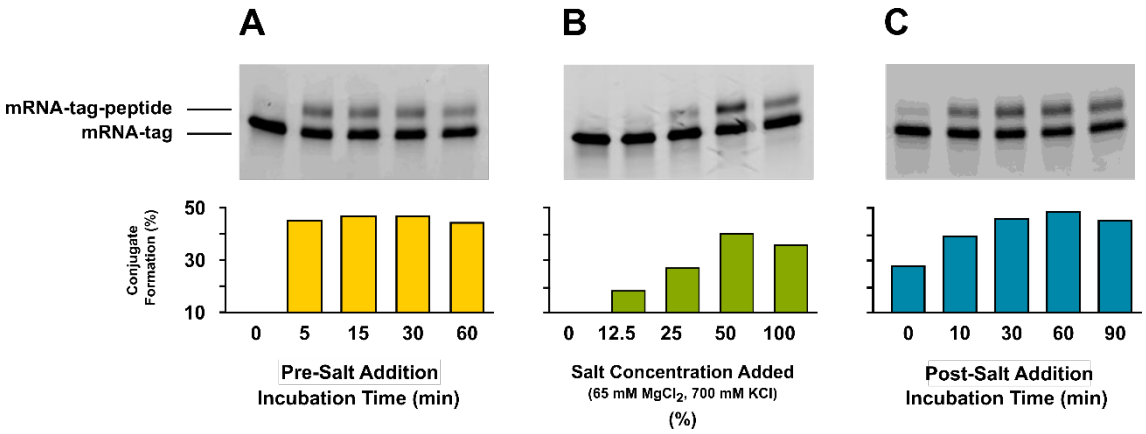


Figure 4.D: Gel images of the mRNA-peptide conjugate formation under various translation conditions. (A) The difference in length of translation before salt addition. The numbers indicate different incubation time (in minutes) carried out at 37 degrees Celsius. (B) The difference in salt concentrations with original concentration (100%) as 750 mM KCl, 65 mM $MgCl_2$ used in previous studies [13,14] (C) Difference in lengths of incubation time (in minutes) after salt addition.

The second thing that was tested was the concentration of salt mix to be added to the reaction after the initial incubation. Previous studies showed that high concentration of KCl improved the accessibility of the peptidyl-tRNA (bound to the ribosome) to the puromycin [15], and so this salt addition step has been adopted in recent mRNA and cDNA display methods [13,16,17]. If the salt mixture is not added,

conjugate formation is less than 10%, however, adding 32.5 mM $MgCl_2$ and 375 mM KCl to the mixture increases the formation rate to above 40% (Figure 4.D; B). Finally, the last parameter tested was the incubation time post-salt addition, as it helps the puromycin to react with the peptidyl-tRNA to form a covalent link between mRNA and peptide. A total of 60 minutes of incubation at 37°C after adding the already mentioned salt mixture, yielded a conjugate formation rate of 48.5% (Figure 4.D; C).

After the optimized conditions were tested and confirmed, the mRNA-peptide conjugates were synthesized from both libraries mRNA-tag samples. Gel purification followed with an SDS-Urea polyacrylamide gel.

4.6 Reverse Transcription for cDNA Display and Verification of mRNA-peptide and mRNA/cDNA-peptide Conjugates

With the purified mRNA-peptide conjugates, reverse transcription (RT) was performed to obtain mRNA/cDNA-peptide conjugates, for cDNA Display. Purification was also performed on the resulted samples.

Verification for both mRNA-peptide and mRNA/cDNA-peptide samples was done in two different ways: trypsin digestion and western blot. Trypsin digestion was performed since it targets lysine (Lys) and arginine (Arg) residues, and all peptides have 3 Lys resides close to the C-terminus region, where the DNA tag is (Appendix, Section A.2, Figure A.2.1). Therefore, after the digestion, the peptide is cut from the nucleotide part of the conjugate and a band shift in a Urea-SDS PAGE gel should

be visible. The digestion was applied for the 10aa-random mRNA-tag and mRNA-tag-peptide products.

After the digestion was performed for 30 min on the translated product, the upper band on the gel disappeared, indicating that the band corresponded to the mRNA-tag-peptide conjugate (Figure 4.E; A).

As for the western-blot, it was performed on the products (with and without RT) derived from the fixed FLAG-control library (Figure 4.E; B). FITC fluorescence signal was observed from the membrane transferred samples and the chemiluminescence signal was detected using an anti-FLAG-HRP antibody. Both signals overlap at the same position indicating the formation of peptides and their attachments to both mRNA-tag and mRNA/cDNA-tag samples. Under SDS-PAGE conditions, translated samples migrate more slowly (upper bands) than mRNA-tag products, as the peptide attachments make them heavier and bigger; but double-stranded mRNA/cDNA-peptide conjugates migrate faster than the single-stranded mRNA-peptide conjugate, due to the double charge from the phosphate backbone (two nucleotide strands).

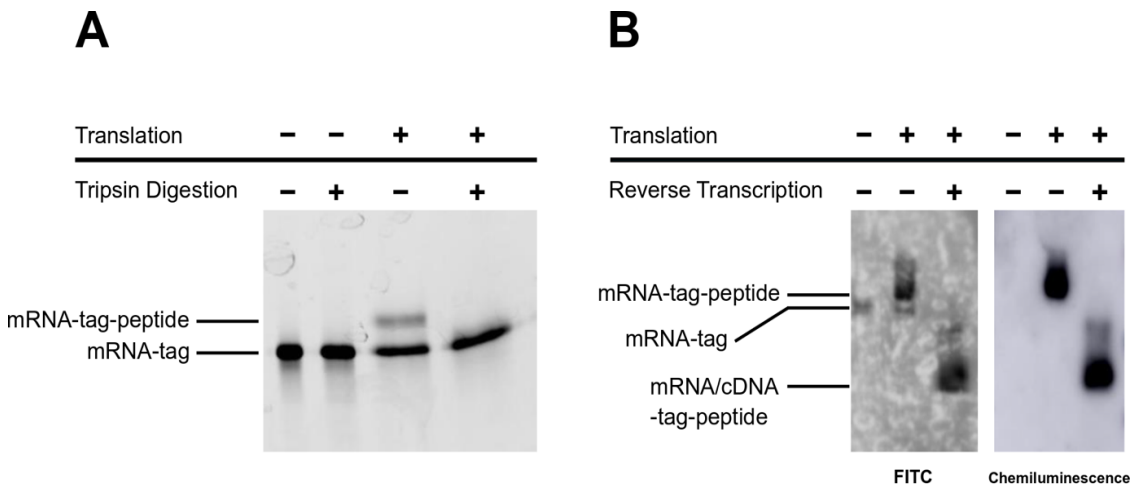


Figure 4.E: Confirmation of mRNA/cDNA-peptide conjugates. (A) Trypsin digestion of 10aa-random library mRNA-tag and mRNA-peptide conjugate products. (B) Western blotting of FLAG-random library mRNA-tag, mRNA-peptide and mRNA/cDNA-peptide conjugate products. For the western blot, the detection was done using FITC fluorescence for the DNA-tag located in all conjugate products (left); and chemiluminescent signal for the FLAG epitope.

4.7 *In vitro* Selection Assay and Reverse Transcription PCR

To validate the performance of both mRNA and cDNA display methods, a selection assay was conducted using anti-FLAG M2 antibody magnetic beads.

To further test the selection ability of both methods, the two libraries (FLAG-random and 10aa-random) conjugate peptide samples were mixed with a 1:10,000 molar ratio respectively. After the binding selection process was finished and several washing steps were carried out, the recovered beads were directly subjected to reverse transcription PCR (RT-PCR) to produce cDNA sequences of the selected samples in the binding assay. Before this though, a test was done to determine the optimal number of cycles of the RT-PCR to avoid over-amplification of the

samples. For each round of selection, the amplified product was checked every 5 cycles (from 0 to 35 cycles) (Appendix, Section A.2, Figure A.2.3).

Amplified DNA products, from both display methods, started being seen at the 15th cycle and were saturated at the 20th cycle. When looking at the band intensity of the 15th RT-PCR cycle, the final rounds showed a higher intensity, proving that the quantity of sample attached to the beads was higher as the rounds progressed. A non-specific band, probably from non-specific binding of primers, appeared in every lane at around 200 bp, but since it remained weak in intensity throughout multiple rounds, the rounds of selection were continued. The optimal number of cycles (for both display methods) for rounds 1 and 2 were settled at 20 cycles and for the 3 rd round, 16 cycles.

4.8 High-throughput MiSeq Sequencing Analysis

All the sequencing data from the MiSeq system (performed by the sequencing department of the Tokyo Institute of Technology) were collected and analysed with a Perl script to extract the coding regions. The extracted sequences were further evaluated by the FastAptamer software [18], to count and rank the unique sequences based on their total number of reads. The top 50 most abundant sequences from each round, of mRNA display, were further explored to confirm the enrichment of FLAG epitope during the selection. Sequences from the FLAG-random library resulted in a fixed aspartic acid (D) amino acid residue at the 6th position just after the first round of selection; followed by two positively charged

arginine (R) and lysine (K) residues at the variables 7th and 8th positions respectively (Figure 4.F; A). For the 10aa-random library, after three rounds of selection, the sequences settled on a conserved core FLAG motif (DYKxxD) in two different positions: one starting at position 2 and the other at position 5 (Figure 4.D; B). This binding motif has been reported previously to have a strong affinity to anti-FLAG M2 antibodies [8–11]. Further analysis was put into those two conserved motifs and a few features were found; namely a weakly conserved glutamic acid (E) immediately before the FLAG motif, (D/L/N) and (D/F/L/N/Y) located between the conserved lysine (K) and aspartic acid (D) with over 80% probability followed by a downstream proline (P) residue (Figure 4.F; C and D).

The next analysis performed was the enrichment rate of various FLAG motifs detected through both displays methods and their comparison. This step is important to determine the efficiency of both display methods.

Although from the 1st round to the 3rd, the samples sequenced were obtained directly from the magnetic beads, for the final round, the 4th, a competitive elution with a FLAG peptide (DYKDDDDK) was done by increasing the concentration during the elution process, and the samples obtained sequenced as normal.

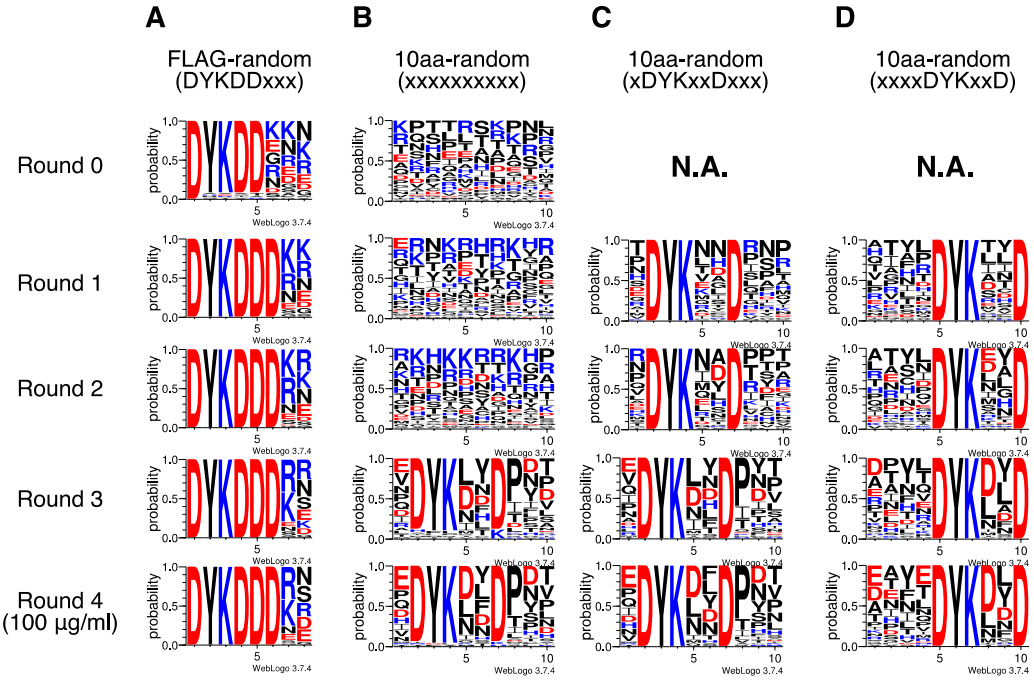


Figure 4.F: Consensus logo of top 50 abundant sequences from each selection round of mRNA display. A round-by-round consensus logo was created using Weblogo3 [19] using the top 50 most abundant sequences based on their RPM. Sequence logos derived from FLAG-random library (A), 10aa-random library (B), core FLAG motif bearing sequences within the 10aa-random library (C and D) are colour-coded based on their charge (negatively charged: red, positively charged: blue and others: black). For the 4th round, consensus logo was created using the sequences eluted with highest FLAG epitope tag concentration (100 µg/ml).

The expected frequency for the full FLAG epitope sequence (DYKDDDDK) in the initial library (zero round) was 0.29 reads per million and it was indeed not detected (1,768,526 sequences). But as the rounds advanced, the observed frequency went up, to 1,479 and 198 reads per million (3rd round) for mRNA and cDNA display, respectively (Figure 4.G; A). When focused on the consensus FLAG motif (DYKDDDxx) with a fixed aspartic acid (D) at the sixth position, the

enrichment numbers were 261,982 and 31,689 reads per million (for mRNA and cDNA display, respectively) (Figure 4.G; B).

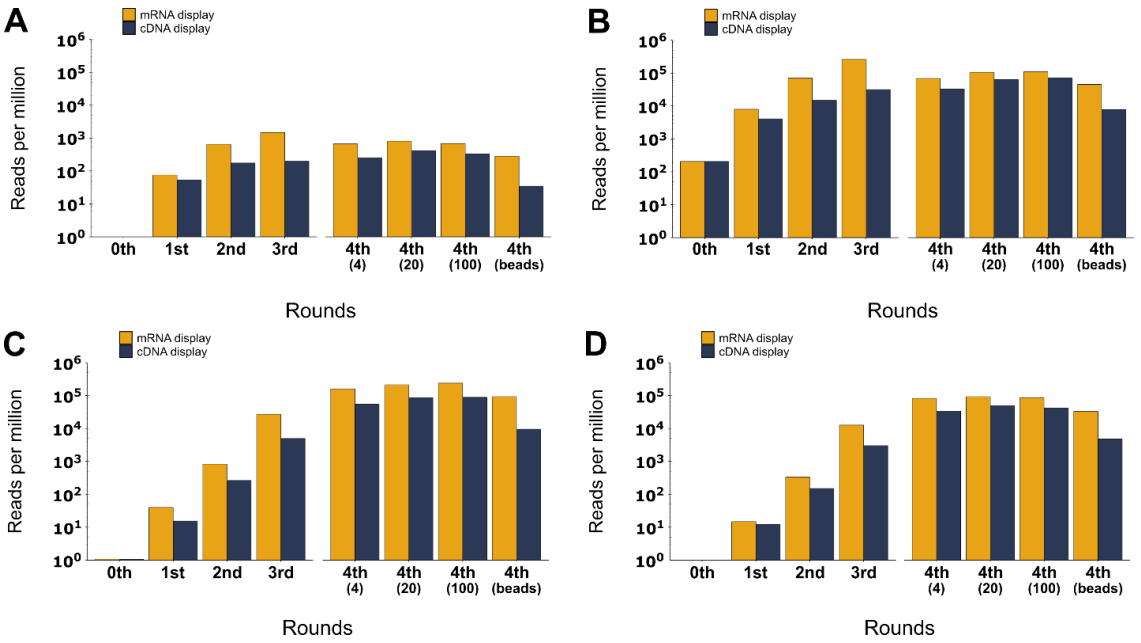


Figure 4.G: Round-by-round read counts for the enriched FLAG motifs. Sequence reads from initial library (0th) to 3rd round selection, 4th round with competitive elution with FLAG epitope tag at 4, 20, 100 μg/ml concentration and remaining beads are counted for (A) DYKDDDDK, (B) DYKDDDxx, (C) xDYK(D/L/N)(D/F/L/N/Y)Dxxx and (D) xxxxDYK(D/L/N)(D/F/L/N/Y)D sequences. For each round, read counts for mRNA display method (orange) and cDNA display (dark blue) are presented side-by-side.

This suggests that the 6th position is important as it enhances the binding of the sequence to the anti-FLAG M2 antibody. A majority of the sequences that were collected derived from the 10aa-random library, especially one type of consensus FLAG motifs: DYK(D/L/N)(D/F/L/N/Y)D, located at two different positions (Figure 4.G; C and D).

The highest enrichment of FLAG motifs was not seen from the sequences on the beads but from the elution done with the competitor FLAG peptide (20 or 100 μg/ml), reaching total 325,917 and 135,913 RPM for mRNA and cDNA display, respectively (Table 4.A; B). With the selection conditions chosen, mRNA display shows a higher enrichment rate of the consensus FLAG motif over cDNA display (Table 4.A; C). Additionally, with mRNA display, the number of FLAG motifs in the sequences that remain bound to the beads, after the highest elution was done, was greater than with cDNA display (Figure 4.G and Table 4.A; A and B). These results suggest that the mRNA/cDNA conjugate is more easily eluted from the beads.

Table 4.A: Read per million (RPM) counts of sequence motifs for each selection round.

A

		FLAG-random library			
		DYKDDDDK		DYKDDDxx	
		mRNA	cDNA	mRNA	cDNA
Rounds	0th	0.0	0.0	207.3	207.3
	1st	75.5	54.3	8,020.8	4,087.2
	2nd	654.8	176.1	70,554.6	14,886.9
	3rd	1,479.3	198.8	261,982.4	31,689.0
	4th (4 µg/ml)	667.4	254.5	69,109.0	33,061.4
	4th (20 µg/ml)	811.2	419.1	106,310.4	64,328.6
	4th (100 µg/ml)	679.7	330.1	109,619.2	71,565.1
	4th (beads)	281.1	34.9	45,249.6	7,824.0

B

		10aa-random library					
		xDYK(DLN)(DFLNY)Dxxx		xxxxDYK(DLN)(LYDNF)D		DYK(DLN)(LYDNF)D	
		mRNA	cDNA	mRNA	cDNA	mRNA	cDNA
Rounds	0th	1.1	1.1	0.0	0.0	1.1	1.1
	1st	39.9	15.5	14.8	12.2	54.7	27.7
	2nd	829.5	264.3	331.5	150.4	1,161.0	414.7
	3rd	26,969.3	4,972.2	12,722.6	3,008.6	39,691.9	7,980.8
	4th (4 µg/ml)	159,326.1	54,677.6	82,702.1	34,327.9	242,028.2	89,005.5
	4th (20 µg/ml)	213,819.3	87,011.1	92,021.7	48,902.1	305,841.0	135,913.2
	4th (100 µg/ml)	239,957.4	89,083.7	85,959.6	42,275.9	325,917.0	131,359.5
	4th (beads)	92,287.3	9,561.9	32,896.6	4,791.3	125,183.9	14,353.2

C

		FLAG-random library			
		DYKDDDDK		DYKDDDxx	
		mRNA	cDNA	mRNA	cDNA
Rounds	0th to 1st	N.A.	N.A.	38.7	19.7
	1st to 2nd	8.7	3.2	8.8	3.6
	2nd to 3rd	2.3	1.1	3.7	2.1
	3rd to 4th (100 µl/ml)	0.5	1.7	0.4	2.3
	4th (100 µl/ml to beads)	0.4	0.1	0.4	0.1

D

		10aa-random library					
		xDYK(DLN)(DFLNY)Dxxx		xxxxDYK(DLN)(LYDNF)D		DYK(DLN)(LYDNF)D	
		mRNA	cDNA	mRNA	cDNA	mRNA	cDNA
Rounds	0th to 1st	37.3	14.5	N.A.	N.A.	51.1	25.9
	1st to 2nd	20.8	17.1	22.4	12.4	21.2	15.0
	2nd to 3rd	32.5	18.8	38.4	20.0	34.2	19.2
	3rd to 4th (100 µl/ml)	8.9	17.9	6.8	14.1	8.2	16.5
	4th (100 µl/ml to beads)	0.4	0.1	0.4	0.1	0.4	0.1

A final analysis of the sequences was done to see how much importance some amino acids residues, from the FLAG motif, have over the binding strength to the anti-FLAG M2 antibody. Comparison of sequences lacking one of the four residues of the core FLAG motif (DYKxxD) was done (Figure 4.H). The lacking tyrosine (Y) or lysine (K) residues showed negative enrichment in the first round, indicating the importance of these residues for primary recognition of the anti-FLAG M2 antibody and posterior binding. The observed result is consistent with the previous selections [8–10] and the obtained M2 antibody structural data [7].

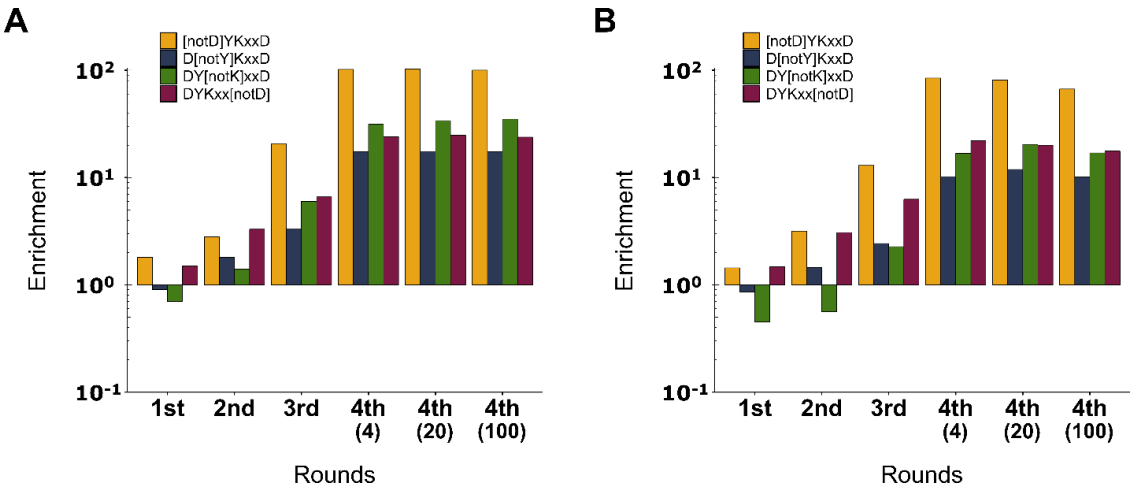


Figure 4.H: Round-by-round enrichment of the motifs lacking a single key residue from the core FLAG motif. Sequence reads for partial FLAG motif lacking each of the four key residues first D (yellow), Y (blue), K (green) and second D (dark red) are shown for each round during the (A) mRNA display and (B) cDNA display selection.

4.9 Conclusions

In this chapter, two display methods (mRNA and cDNA) were tested using the commercially available PURE*frex* 1.0 cell-free protein system. This purified cell-free system was chosen over cell-free systems based on cell lysates. This method proved efficient and stable in the synthesis and selection of mRNA and mRNA/cDNA-peptide conjugate. The PURE systems lack of contaminants allowed for minimal degradation of the sample as it was shown by the high conjugation efficiency (40%). In existing methodologies, degradation causes the conjugation efficiency to range between 5-40% [4]. Unwanted components during the cleaning steps were limited using electroelution to ensure the purity of the peptide conjugates and ethanol precipitation to secure as much sample recovery as possible. Although these latter two steps are common purification techniques, they could be important when selecting for a component of the PURE system such as components involved in the transcription and translations processes.

mRNA and cDNA display methods were also compared. The overall performance was validated by searching for epitopes, from approximately 1,012 sequences, using the anti-FLAG M2 antibody [7]. Posterior high-throughput sequencing, after three rounds of selection, led to the identification of the FLAG epitope (DYKDDDxx) and consensus FLAG motif sequences [DYK(DLN)(DFLNY)D]. This specific motif appeared in the only two sites where it could appear (if the restrictions imposed in the original library are taken into consideration) since only three sites in the mentioned library had degenerate codon sequences that can code for tyrosine, and

one of those three sites is too close to the C-terminal end of the random region and therefore cannot be part of the FLAG motif.

When comparing the performance of both methods, similar results were obtained, with the mRNA display having slightly higher enrichment than the cDNA display (Figure 4.H). Although this has only been performed once, and repetition should be done to confirm this statement, this could prove the consistency of the method which will be useful for future studies to explore the sequence and functional space of diverse polypeptides.

One difference found was the elution speed of both FLAG motif sequences conjugates, mRNA-peptide and mRNA/cDNA-peptide, with the latter being eluted slower from the magnetic beads. One possible reason could be that the mRNA-peptide conjugate is smaller in size (around 135 bp of cDNA lighter depending on the DNA library) and thus allows it to be more accessible to binding sites on the beads than the larger mRNA/cDNA-peptide conjugates. Also, the sequences from the cDNA display method appear to have more non-specific binding on the remaining beads after the elution process.

By showing that through a round-by-round sequencing process and a stepwise competitive elution with FLAG peptide (at the final 4th round), the consensus FLAG motif prevailed over other sequence variants; the method gained a strong point in favour since high-throughput sequencing is being used more and more as a powerful approach for various *in vitro* selection and evolution experiments. Additional enrichment of several amino acids adjacent to the FLAG motif in a

library dependent manner (FLAG-random vs 10aa-random) were also observed. This suggests a mutual interaction between residues so an optimal binding conformation can be formed. Furthermore, by comparing the enrichment patterns of sequences with a single residue difference, it was possible to highlight the residues with the most influence with regards to the anti-FLAG M2 antibody recognition process. Enrichment of core FLAG motifs lacking one of the four key residues (DYKxxD) indicates that tyrosine (Y) and lysine (K) appear as the two key residues essential for binding. Although this has been studied before, it has been done on phage display [9]. This pattern can be useful to further explore high-affinity epitope candidates and even discover "super epitopes variants", based on recent advancements of array-based high-throughput and quantitative protein assay [20].

4.10 References for Chapter 4

1. Godino, E. *et al.* De novo synthesized Min proteins drive oscillatory liposome deformation and regulate FtsA-FtsZ cytoskeletal patterns. *Nat. Commun.* (2019). doi:10.1038/s41467-019-12932-w

2. Nagumo, Y., Fujiwara, K., Horisawa, K., Yanagawa, H. & Doi, N. PURE mRNA display for in vitro selection of single-chain antibodies. *J. Biochem.* (2016). doi:10.1093/jb/mvv131

3. Ueda, T., Kanamori, T. & Ohashi, H. Ribosome display with the PURE technology. *Methods Mol. Biol.* (2010). doi:10.1007/978-1-60327-331-2_18

4. Wang, R., Cotten, S. W. & Liu, R. MRNA display using covalent coupling of MRNA to translated proteins. *Methods Mol. Biol.* (2012). doi:10.1007/978-1-61779-379-0_6

5. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–5 (2001).

6. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* (2005). doi:10.1016/j.ymeth.2005.04.006

7. Roosild, T. P., Castronovo, S. & Choe, S. Structure of anti-FLAG M2 Fab domain and its use in the stabilization of engineered membrane proteins. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* (2006). doi:10.1107/S1744309106029125

8. Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T. & Yanagawa, H. DNA display for in vitro selection of diverse peptide libraries. *Nucleic Acids Res.* (2003). doi:10.1093/nar/gng119

9. Srila, W. & Yamabhai, M. Identification of amino acid residues responsible for the binding to anti-FLAG™ M2 antibody using a phage display combinatorial peptide library. *Appl. Biochem. Biotechnol.* (2013). doi:10.1007/s12010-013-0326-8

10. Miceli, R. M., DeGraaf, M. E. & Fischer, H. D. Two-stage selection of sequences from a random phage display library delineates both core residues and permitted structural range within an epitope. *J. Immunol. Methods* (1994). doi:10.1016/0022-1759(94)90097-3

11. Osada, E., Shimizu, Y., Akbar, B. K., Kanamori, T. & Ueda, T. Epitope mapping using ribosome display in a reconstituted cell-free protein synthesis system. *J. Biochem.* (2009). doi:10.1093/jb/mvp027

12. Nishigaki, K., Taguchi, K., Kinoshita, Y., Aita, T. & Husimi, Y. Y-ligation: An efficient method for ligating single-stranded DNAs and RNAs with T4 RNA ligase. *Mol. Divers.* (1998). doi:10.1023/A:1009644028931

13. Naimuddin, M. & Kubo, T. A High Performance Platform Based on cDNA Display for Efficient Synthesis of Protein Fusions and Accelerated Directed Evolution. *ACS Comb. Sci.* (2016). doi:10.1021/acscombsci.5b00139

14. Naimuddin, M. *et al.* Directed evolution of a three-finger neurotoxin by using cDNA display yields antagonists as well as agonists of interleukin-6 receptor signaling. *Mol. Brain* (2011). doi:10.1186/1756-6606-4-2

15. Van Der Mast, C. A. & Bloemers, H. P. J. The puromycin reaction mediated by yeast ribosomes in high salt. *Mol. Biol. Rep.* (1973). doi:10.1007/BF00357646

16. Seelig, B. MRNA display for the selection and evolution of enzymes from in vitro-translated protein libraries. *Nat. Protoc.* (2011). doi:10.1038/nprot.2011.312

17. Yamaguchi, J. *et al.* cDNA display: A novel screening method for functional disulfide-rich peptides by solid-phase synthesis and stabilization of mRNA-protein fusions. *Nucleic Acids Res.* (2009). doi:10.1093/nar/gkp514

18. Alam, K. K., Chang, J. L. & Burke, D. H. FASTAptamer: A bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Mol. Ther. - Nucleic Acids* (2015). doi:10.1038/mtna.2015.4

19. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: A sequence logo generator. *Genome Res.* (2004). doi:10.1101/gr.849004

20. Layton, C. J., McMahon, P. L. & Greenleaf, W. J. Large-Scale, Quantitative Protein Assays on a High-Throughput DNA Sequencing Chip. *Mol. Cell* (2019). doi:10.1016/j.molcel.2019.02.019

Chapter 5: Production of Lipid Vesicles and Nanodiscs Synthetic Membranes and G-protein-coupled receptors (GPCRs) using the PURE Cell-free System

5.1 Chapter 5 Aims and Summary

Chapter 3 presented studies focused on cell-free system's dynamics whilst Chapter 4 used this information to select a cell-free system from purified components to test two different display methods: mRNA and cDNA display.

This Chapter focusses on using the PURE system within two different types of synthetic lipid membranes for the production of G-protein-coupled receptors (GPCRs). These types of membrane protein are attractive targets for drug screening since they are found in most cells. They perform a wide variety of functions and their localization, on the cell's surface, provides easy access. But they are more challenging to study than soluble proteins since they are dependent on lipid membranes to fold and settle properly. The main elements of this chapter include the use of a cell-free system, the PURE system; the selection of a type GPCR membrane protein, CX3CR1; and the use of two different synthetic lipid membranes, vesicles and nanodiscs. The CX3CR1 protein (from the human GPCR family) was selected as a target because it is involved in a wide range of inflammatory diseases [1], which makes it a potential target for cancer therapy [2]. It also has a known specific ligand, CX3CL1 [3]. This information could help check the veracity of the proposed method or design a peptide target library more readily.

CX3CR1 also has shown polymorphism, which offers targeted therapeutic options [4].

The chapter details the synthesis of the CX3CR1 protein in both lipid vesicles and nanodiscs (Figure 5.A). The design and production of a random peptide library will then be described to find suitable targets for the GPCR. This last section is the first step in performing screening and selection of peptide targets for GPCRs samples using display methods. The details of this will be outlined and expanded in Chapter 6 (Future Work Section).

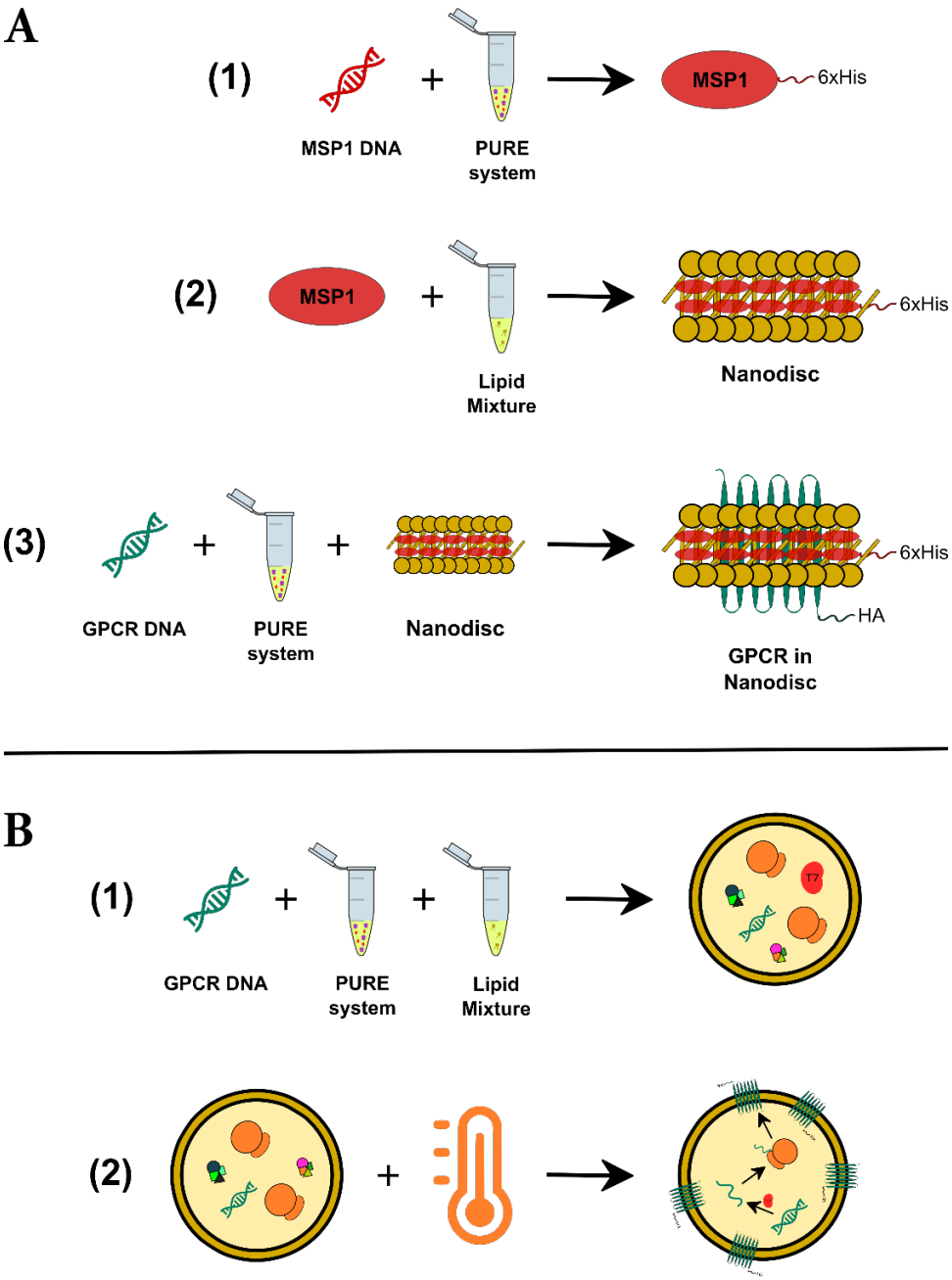


Figure 5.A: Summary of the experimental approach for Chapter 5. Experimental procedures carried out during this chapter. (A) Expression of GPCR in nanodiscs. The first step (1) used the PURE system to produce the MSP1D1 proteins, which are necessary to create nanodiscs, with a His-tag for future selection and purification. The second step (2) involves the full creation of the nanodiscs using the MSP1D1 protein and a lipid mixture. The third step (3) uses the nanodiscs and again the PURE system to produce GPCRs. (B) Expression of GPCR in liposomes. The first step (1) encapsulates the cell-free system and GPCR DNA inside the liposome. The second step (2) incubates the liposomes so the GPCR is produced and automatically inserted in the membrane.

5.2 Liposomes as Synthetic Membranes for GPCRs Production and Assembly: Microscope Visualisation and FACS Analysis.

The first lipid membrane model used for the production of GPCRs was liposomes and Giant unilamellar vesicles, (GUVs). Vesicles are usually more economical to produce and large quantities can be obtained in a single preparation process, and even stored for future use. The water-in-oil (w/o) emulsion transfer method [5–7] (Figure 5.B) was chosen to produce the lipid vesicles since the liposomes produced with this method have already been studied and their stability optimized for protein production [8]. The DNA of the product being synthesised, in this case, the GPCR membrane protein CX3CR1, is located in the inner solution, along with the PURE system. So when the liposomes are assembled, the membrane protein production occurs inside the liposome. It has been found that even though there are no translocon proteins located inside the liposome to carry the GPCRs to the membrane, membrane proteins can spontaneously integrate themselves on the liposome's membrane [9,10]. To visualize the GPCRs, these were cloned downstream with an sfGFP gene, so as a fusion protein when the GPCRs travels and assembles in the membrane, the sfGFP fluorescence should be visible on the liposome's surface, instead of inside where some of the proteins that fail to interact with the membrane remain. Furthermore, when creating the DNA construct, the T7 promoter had a mutation at the position 17, adenine (A) for tyrosine (T), since it was proven in chapter 3 that a few mutations in the T7 promoter enhance the final protein production.

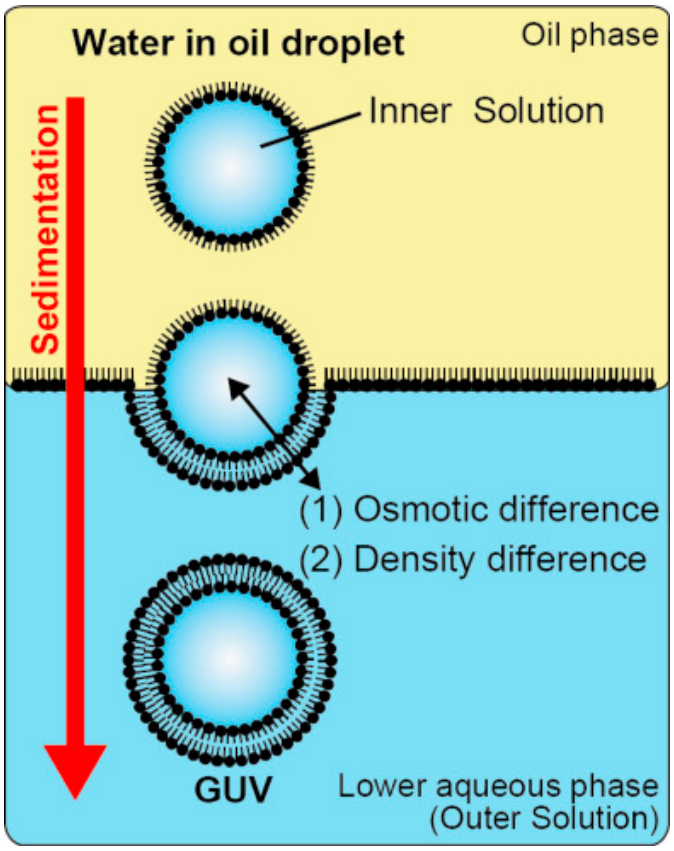


Figure 5.B: Schematic of the water in oil (w/o) emulsion-transfer method for GUV production. The method starts with a water droplet suspended in the upper oil phase while being protected by a lipid monolayer, which later will become the inner monolayer of the vesicle. By performing centrifugation, the w/o droplet sinks further and settles in the lower aqueous phase, which is the outer solution in which the GUVs will be suspended. When the liposome passes from the oil phase to the aqueous phase, it crosses a layer of lipids which quickly link themselves to the droplet forming a lipid bilayer. Image used is from Tsuji *et al.* (Journal of Bioscience and Bioengineering, 2018) [7].

To determine the optimal composition for the lipid, inner and outer solutions for the vesicles, a first test was performed. This consisted of producing vesicles according to Fujii et al. [6]. The lipid composition used was POPC: Cholesterol (1: 1). The inner solution consisted of the PURE system and a DNA sequence for sfGFP

(the WT construction from Chapter 3), to confirm the proper functioning of the transcription-translation system (Appendix, Section A.3, Figure A.3.1). The outer solution is composed of the non-protein elements of the PURE system but also without the tRNAs. This outer solution is also complemented with 200 mM glucose to maintain an osmosis balance. sfGFP was produced and visualized by microscope (Appendix, Figure A.3.1).

The next step involved the direct production of the hybrid GPCR protein, CX3CR1-sfGFP inside the vesicles and posterior analysis. The hybrid protein was seen on the surface of the vesicles under the microscope (Figure 5.C; A). The vesicles were also run through fluorescence-activated cell sorting (FACS) to analyse better the total population of the vesicles, along with their conformation and fluorescence level. FACS analysis was chosen to explore the characteristics of the whole population of liposomes. When plotting the liposomes, calculations can be performed to analyse the percentage of liposomes with certain characteristics and collect them to separate them from the rest of the population. This is useful when performing further experiments with the liposomes in order to have a more homogenous population with high number of the protein of interest.

When plotting the forward (FSC-A) and side (SSC-A) scattering values of each vesicle against each other, there is a certain region of the plot that corresponds to the unilamellar liposomes (Figure 5.C; B). These liposomes are usually of various sizes (wide range of FSC-A values) but with low SSC-A values since several monolayers divert the laser more and therefore increment the side scattering of

the vesicle, as it was shown in the Fujii et. al paper [6]. Furthermore, when comparing the sfGFP fluorescence levels (FITC-A) of the vesicles with ones with no DNA, the plot shifted (Figure 5.C; B). It can be seen that the lower plots (negative control) show no fluorescence further than the 10^2, being this the background noise. While the upper graphs (those of the sample) show population of vesicles with FITC signal (to the right of the 10^2 mark). However, less than 25% of the total vesicle population showed sfGFP fluorescence in the FACS analysis. This could be due to the limited concentration of the inner phase when manufacturing the vesicles, resulting in depleted liposomes. Studies have also shown that the PURE system is prone to generate partially translated proteins or functionally inactive, due to ribosome stalling [11], which is more likely for hybrid proteins, such as CX3CR1-sfGFP.

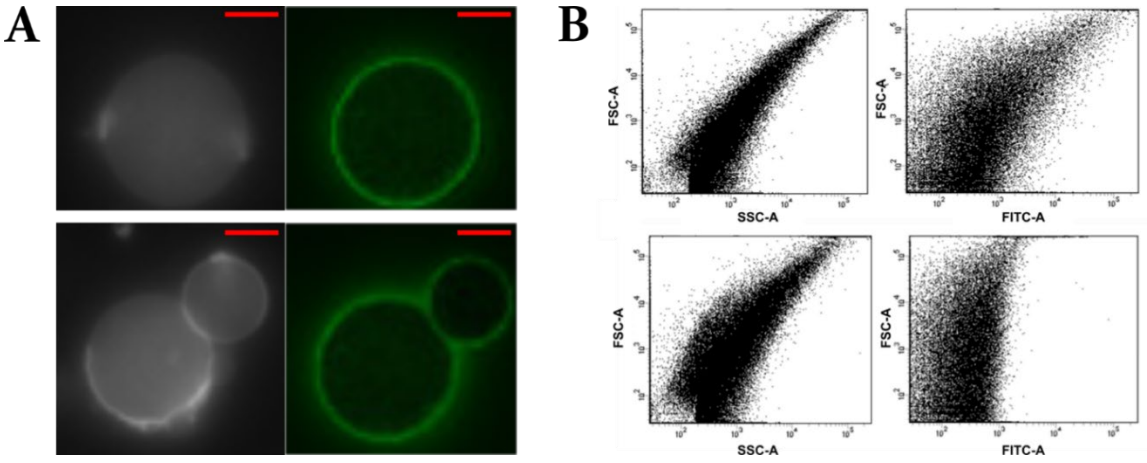


Figure 5.C: Lipid vesicles analysis using the PURE system and CX3CR1-sfGFP fusion protein. (A) Microscope images of vesicles with a lipid composition of POPC: Cholesterol (1: 1). The inner solution contains the PURE*frex* 1.0 system and CX3CR1-sfGFP DNA. The images on the left column correspond to phase contrast and the ones on the right column to the green fluorescence of CX3CR1-sfGFP migrating to the liposome membrane. The bar size, on the right upper corners of the images, is 25 µm. (B) Plots in 2D of the lipid vesicles population when run through FACS. The upper graphs represent the CX3CR1-sfGFP sample while the bottom graphs represent a negative control (PURE system with no DNA). The forward scattering value is represented by FSC-A and the side scattering value by SSC-A. The sfGFP fluorescence is represented by FITC-A. The total number of vesicles plotted in each graph is 100,000.

5.3 Nanodiscs as Synthetic Membranes for GPCRs Production and Assembly: MagneHis Purification and Western Blot Analysis.

Another lipid model chosen for the production of GPCRs was nanodiscs [12,13]. These are smaller than vesicles, and GPCRs inserted in them are accessible from both ends. This provides a significant advantage when performing selection methods as there is limited concern over the correct insertion of the protein in the membrane. Furthermore, detection can be easily be carried out by SDS-PAGE gel and western blot. Nanodiscs, like vesicles, remain stable when kept at low temperatures (4 to -20 degree Celsius), so they are useful for long experiments.

The first step in making nanodiscs is the expression and purification of the membrane scaffold proteins (MSP). In this case, the human MSP1D1 amphipathic protein was used since it has been shown to enable nanodiscs stabilization [14]. The plasmid pMSP1D1, contains the MSP1D1 gene, a His-tag on the N-terminus and a T7 promoter. Using the PURE*frex* 1.0 system, the proteins were produced. Purification was achieved using magnetic nickel particles attached to His-tag antibodies. Recovered MSP1D1-HisTag samples were then concentrated and stored. Quantification was determined by absorbance at 280 nm using a calculated coefficient of extinction of 21,430 $(mg/ml)^{-1}\ cm^{-1}$, obtaining 3.8 mg/ml of purified product.

Nanodiscs were prepared with a lipid mixture consisting of POPC, POPS and Cholesterol in a 72: 20: 8 ratio respectively. The lipid mixture was incubated with purified MSP1D1 protein in a 1: 60 ratio. The chosen ratios for the lipids and

proteins have been reported to form stable nanodiscs for GPCR insertion [12]. Elimination of remaining detergent in the mixture (which is the Na cholate from the Nanodisc buffer), to aid correct nanodisc assembly, was achieved by SM2 bio beads. A two-step purification process was performed on the recovered nanodiscs samples to eliminate aggregation and aberrant nanodiscs. The first one was done by ultra-centrifugation for big lipid and protein aggregations, since the precipitate faster and at lower speed. The second one by AKTA Chromatography for better sample recovery and concentration. Bigger particles and aggregations elute faster from the column than single nanodiscs. From the chromatography peak obtained (Figure 5.D; A), 10 fractions were run on an SDS-PAGE gel for sample size verification (Figure 5.D; B). 5 fractions were selected for sample concentration, and when finished, another SDS-PAGE gel run was performed (Figure 5.D; C).

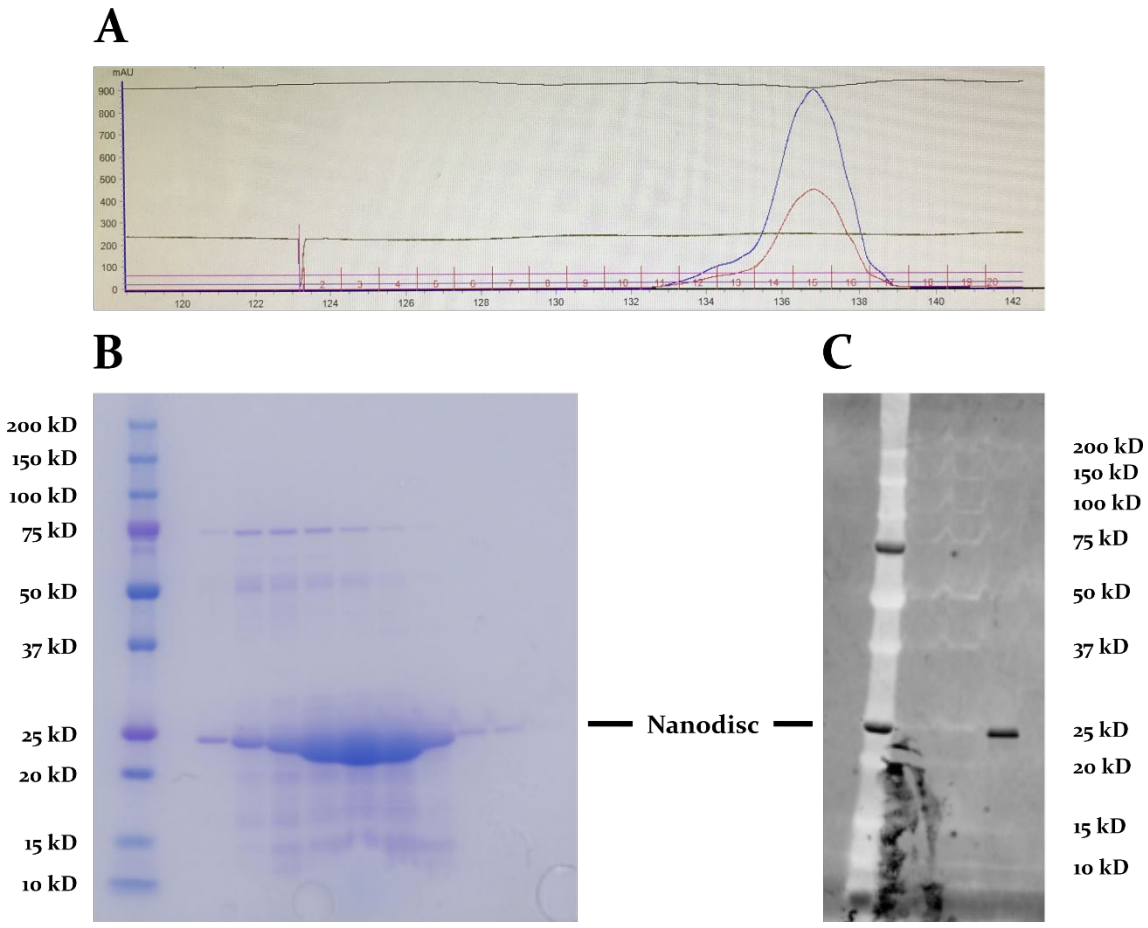


Figure 5.D: Nanodisc purification and verification. (A) Plot graph of the nanodisc sample as it passes through the AKTA chromatography channels. The red and blue lines represent the absorbance at 260 and 280 nm respectively. The x-axis represents the millilitres of sample/buffer that have passed through the column and just on top of it, the number of each fraction collected. The y-axis represents the mili-absorbance units. The blue and red peaks both show the nanodiscs. (B) SDS-PAGE gel of the selected fractions from the AKTA chromatography purification step. A total of 10 fractions were run on the gel. The nanodiscs band can be appreciated at 26 kDa. From these fractions, only 5 were selected for sample concentration. (C) A second SDS-PAGE gel shows the final purified and concentrated nanodisc sample.

For the final step, nanodiscs will be used along with the PURE system to produce CX3CR1. The DNA construct that encodes this protein gene, also contains a HaloTag sequence downstream the CX3CR1 gene. This tag will allow purification of the protein-nanodiscs complexes from those nanodiscs that are empty. The cell-free system selected was PURE*frex* 2.1 DS supplemented. This version differs from the basic PURE system in that it helps in the formation of disulphide bonds in the protein by replacing DTT for glutathione reduced (GSH). After the transcription-translation reaction with nanodiscs, the sample was purified using magnetic nickel particles attached to HA-tag antibodies. This ensured only the CX3CR1 samples were selected and the empty nanodiscs left behind. The collected fractions were run on an SDS-PAGE gel for visualization (Figure 5.E; B). The samples visualized in the SDS-PAGE gel also included the PURE system components. It was challenging to estimate the purification efficiency since some bands were superposed over the CX3CR1 band in the PURE fraction lane (Figure 5.E; A). The process was repeated but with Methionine [S^{35}] in the PURE system, (Figure 5E; B). The results showed that 0.021 mg of final purified CX3CR1 was obtained per ml of the PURE system. The purification efficiency was 63%. One important aspect to have in mind during the experiment is that the obtained bands in the gels corresponding to the CX3CR1 sample were lower than expected. The CX3CR1 protein has a molecular weight of 41 kDa. But the obtained bands in both gels were in the 34 kDa mark. A possible explanation comes from the studies performed by Rath *et al*. [15]. They showed that membrane proteins do not migrate in SDS-PAGE

gels accordingly to their molecular weight. Numerical factors have to be taken into consideration when calculating the molecular weight band on the gel. For proteins migrating faster, like CX3CR1, a correction factor of 0.82 has to be applied. In this case, the molecular weight of the observed band was divided by this factor, resulting in the expected original 41 kDa value.

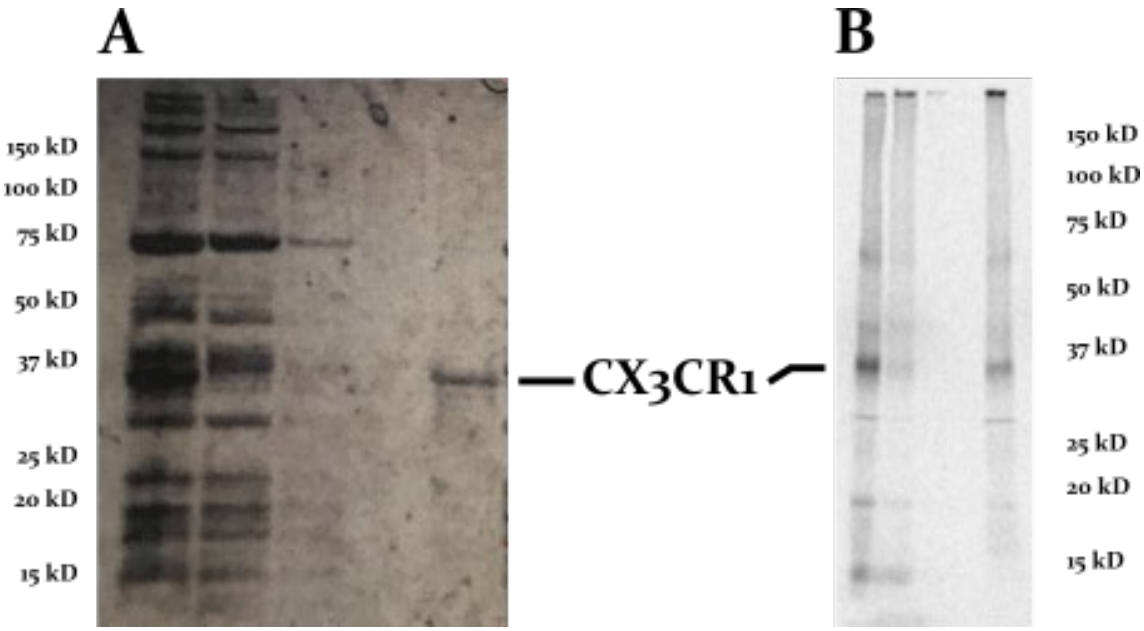


Figure 5.E: CX3CR1 purification and verification. Each lane represents a fraction from the purification process. From left to right: PURE fraction, flow-through fraction, 1st wash fraction, 2nd wash fraction, and elution fraction. (A) SDS-PAGE gel of the fractions from the HA-tag magnetic particles purification step. (B) SDS-PAGE gel of the fractions from the HA-tag magnetic particles purification step. The gel went through fixing and drying processes for posterior visualization of Methionine [S^{35}] signal in the bands.

When analysing the efficiency of the HA-tag magnetic purification process, the majority of the sample that was not purified was located in the flow-through fraction, around 35%. This indicated that this purification process needs to be optimized, as a big portion of the sample does not bind to the HA-tagged particles during the incubation period. This can be solved by finding out the optimal number of magnetic particles needed and incubation time and temperature. Another possible cause could be the HA-tag located in the CX3CR1 proteins. If the tag is not available, due to a short linker sequence, it will not be able to find and bind to the anti-HA tag from the magnetic beads. Furthermore, if the HA-tag has a considerable amount of hydrophobic amino acids, it is more likely to interact and bind to the nanodiscs, making it unavailable for binding to the beads.

The amount of CX3CR1 sample obtained, 0.021 mg, per ml of the PURE system is low when comparing to results obtained from previous studies, but still within an expected range (0.05 – 0.01 mg) [12]. Nanodiscs play an important role when it comes to protein yield. Their final concentration and size can affect the right incorporation and stability of membrane proteins. Thus, this factor is to be taken into consideration when making improvements to this aspect of the reaction.

5.4 DNA Library Design for Screening and Selection of Peptides as Drug Targets for GPCRs.

Following the successful production of CX3CR1 in both lipid vesicles and nanodiscs, these samples could be used for screening of peptide targets for GPCR interactions (important as drug targets).

Using the designs introduced in Chapter 4 as a backbone, a new method could be made for peptide testing on CX3CR1 (expressed in both vesicles and nanodiscs). The first step in this process is the creation of a random peptide library that will provide targets for the protein of interest. Since the PURE system is the preferred cell-free system, the use of the T7 promoter is mandatory. Even more, from chapter 2 several mutations on this promoter enhanced the product yield. A leader region on the 3' end, which complements the puromycin-FITC-DNA tag, is also one of the components of the library, for posterior puromycin ligation to the mRNA molecule (for both mRNA and cDNA display). For the design of the random peptide part, it is important to understand the natural ligands of CX3CR1 and their properties and conformations. The chemokine protein CX3CL1 (FKN) is a ligand of the CX3CR1 receptor. This ligand is around 400 amino acids long in humans, making it a long protein and not suitable for designing any random library. But the chemokine domain (FKN-CD) of CX3CL1 that binds to CX3CR1 [16], with a length of 76 amino acids [17,18], can be used for the library design. This domain binds to the extracellular loops of the receptor [17]. Even if the FKN-CD section of CX3CL1 is much shorter than the whole protein, it is still long if a full random library is to be made. The limit of

the number of sequences *in vitro* display methods can screen and select is around 10^{15}. This number corresponds to around a peptide 12 amino acids long, if all possible natural amino acids can be located in each spot. In the case of FKN-CD, this could be done by parts. Changing certain regions of the FKN-CD domain one at a time and analysing the dominant motifs in each region of the sequence (Figure 5.F; middle). A final optimized ligand can be designed by using the top motifs. Another way could be to just create a random peptide library that is around 10 amino acids long (Figure 5.F; bottom). Although the whole 76 amino acids FKN-CD attaches to CX3CR1, shorter peptides could reveal important binding regions and sequences. This option also allows to produce the random library more economically using long primers and PCR. Long sequences, otherwise, have to be ordered.

One last possibility could be to decrease the possible number of amino acids that can be inserted in each spot. By allowing, for example, 10 possible amino acids per spot, the maximum length of the random peptides increases to 15 amino acids. This could help save resources if some option is already known not to work for certain positions in the peptide. Whether one or another type of library is chosen, it is important to have a positive control (Figure 5.F; top). The untouched sequence of FKN-CD binds fully to CX3CR1. This can be used to test the veracity of the method like it was done in chapter 4.

Figure 5.F: Possible DNA libraries for the screening and selection of peptides targets for the CX3CR1 receptor. All libraries contain a T7 promoter, a ribosome binding site (RBS) and a leader region at the 3' end that complements the puromycin-FITC-DNA tag. (Top) The control DNA library encodes the gene for the chemokine domain (FKN-CD) of the CX3CL1 ligand. This is the naturally found ligand of the CX3CR1 membrane protein. (Middle) The long version of the random peptide library. Only the first 12 amino acids are randomized to limit the total number of possible sequences. (Bottom) The short version of the random peptide library. This only contains 12 randomized amino acids without the rest of the FKN-CD.

5.5 Outline and Conclusions

In this chapter, a first experiment involved the production of the CX3CR1 receptor, a GPCR membrane protein, in lipid vesicles. Liposomes were chosen as the first model of synthetic lipid membranes. The receptor production was carried out inside the liposome using the PURE system. This system is based on purified components and includes only the minimum necessary components for transcription-translation. This helps increase the stability of liposomes as it keeps the concentration of the inner solution elements low, therefore minimizing the change of bursting due to pressure. The visualization of the protein on the liposome surface was done by fusing CX3CR1 to sfGFP. The use of liposome for expression of membrane proteins is widely used. Liposomes are produced quickly and economically. They can also be stored for future uses. The main issue encountered is the lack of knowledge on the membrane protein inclusion direction. In the case of GPCRs, most ligands interact with the extracellular loops. If the loops are found inside the vesicle, the ligands cannot reach them.

The second synthetic membrane model used for CX3CR1 was nanodiscs. These have longer production time and require more laboratory equipment, so they may not be always affordable. But they are small which makes it possible to analyse them using methods as simple as an SDS-PAGE gel. They also permit the membrane protein to be reached by both sides. Storage for longer experimental periods is also possible. Although the efficiency of CX3CR1 production and purification was satisfactory, it can still be improved. For example by allowing

more free access to the HA-tag in the protein. Or by selecting a less hydrophobic tag to minimize interaction with the nanodisc. The amount of total protein produced, 0.021 mg/ml of PURE system, was on the low range. This issue is probably caused by several factors: the condition of the PURE system, the amount and size of the nanodiscs, and the design of the DNA construct used. All these parameters have room for optimization.

CX3CR1 was chosen as the GPCR candidate for this chapter. This protein is related to several types of diseases, and mutations found in it can cause cell mortality [19]. Since CX3CR1 shows polymorphisms, future treatments involving this protein could focus on individual targeted therapy. This membrane protein has a known natural ligand, CX3CL1. This knowledge comes in handy to design more specific sequences, or libraries, of possible targets for CX3CR1.

The design of a random peptide library for CX3CR1 was hindered due to the size of its known ligand, CX3CL1. The chemokine domain of the ligand that binds to the target protein is quite long. This reduces the number of amino acids that can be randomized in the sequence as *in vitro* display methods has a limit of sequences. A bigger number than the limit will just interfere with the process and many sequences could be lost. Smaller random peptides could still be tested for finding binding sequences. Also, shorter regions of the chemokine domain can still be randomized and new motifs found that have higher specificity for CX3CR1.

5.6 Chapter 5 References

1. D'Haese, J. G., Friess, H. & Ceyhan, G. O. Therapeutic potential of the chemokinereceptor duo fractalkine/CX3CR1: An update. *Expert Opinion on Therapeutic Targets* (2012). doi:10.1517/14728222.2012.682574

2. D'Haese, J. G., Demir, I. E., Friess, H. & Ceyhan, G. O. Fractalkine/CX3CR1: Why a single chemokine-receptor duo bears a major and unique therapeutic potential. *Expert Opin. Ther. Targets* **14**, 207–219 (2010).

3. Bazan, J. F. *et al.* A new class of membrane-bound chemokine with a CX3C motif. *Nature* (1997). doi:10.1038/385640a0

4. Niessner, A. *et al.* Opposite effects of CX3CR1 receptor polymorphisms V2491 and T280M on the development of acute coronary syndrome. A possible implication of fractalkine in inflammatory activation. *Thromb. Haemost.* (2005). doi:10.1160/TH04-11-0735

5. Pautot, S., Frisken, B. J. & Weitz, D. A. Production of unilamellar vesicles using an inverted emulsion. *Langmuir* (2003). doi:10.1021/la026100v

6. Fujii, S. *et al.* Liposome display for in vitro selection and evolution of membrane proteins. *Nat. Protoc.* **9**, 1578–91 (2014).

7. Tsuji, G., Sunami, T. & Ichihashi, N. Production of giant unilamellar vesicles by the water-in-oil emulsion-transfer method without high internal concentrations of sugars. *J. Biosci. Bioeng.* (2018). doi:10.1016/j.jbiosc.2018.04.019

8. Nishimura, K. *et al.* Cell-free protein synthesis inside giant unilamellar vesicles analyzed by flow cytometry. *Langmuir* (2012). doi:10.1021/la3001703

9. Kuruma, Y., Suzuki, T., Ono, S., Yoshida, M. & Ueda, T. Functional analysis of membranous F 0-a subunit of F 1F 0-ATP synthase by in vitro protein synthesis. *Biochem. J.* (2012). doi:10.1042/BJ20111284

10. Fujii, S., Matsuura, T., Sunami, T., Kazuta, Y. & Yomo, T. In vitro evolution of α-hemolysin using a liposome display. *Proc. Natl. Acad. Sci. U. S. A.* (2013). doi:10.1073/pnas.1314585110

11. Li, J., Gu, L., Aach, J. & Church, G. M. Improved cell-free RNA and protein synthesis system. *PLoS One* **9**, (2014).

12. Gessesse, B., Nagaike, T., Nagata, K., Shimizu, Y. & Ueda, T. G-Protein Coupled Receptor Protein Synthesis on a Lipid Bilayer Using a Reconstituted Cell-Free Protein Synthesis System. *Life* (2018). doi:10.3390/life8040054

13. Cai, Y. *et al.* Purification of family B G protein-coupled receptors using nanodiscs: Application to human glucagon-like peptide-1 receptor. *PLoS One*

(2017). doi:10.1371/journal.pone.0179568

14. Stoilova-McPhie, S., Grushin, K., Dalm, D. & Miller, J. Lipid nanotechnologies for structural studies of membrane-associated proteins. *Proteins Struct. Funct. Bioinforma.* (2014). doi:10.1002/prot.24631

15. Rath, A. & Deber, C. M. Correction factors for membrane protein molecular weight readouts on sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Anal. Biochem.* (2013). doi:10.1016/j.ab.2012.11.007

16. Fujita, M., Takada, Y. K. & Takada, Y. Integrins α v β 3 and α 4 β 1 Act as Coreceptors for Fractalkine, and the Integrin-Binding Defective Mutant of Fractalkine Is an Antagonist of CX3CR1 . *J. Immunol.* (2012). doi:10.4049/jimmunol.1200889

17. Mizoue, L. S., Bazan, J. F., Johnson, E. C. & Handel, T. M. Solution structure and dynamics of the CX3C chemokine domain of fractalkine and its interaction with an N-terminal fragment of CX3CR1. *Biochemistry* (1999). doi:10.1021/bi9820614

18. Wojdasiewicz, P. *et al.* The Chemokine CX3CL1 (Fractalkine) and its Receptor CX3CR1: Occurrence and Potential Role in Osteoarthritis. *Archivum Immunologiae et Therapiae Experimentalis* (2014). doi:10.1007/s00005-014-0275-0

19. Collar, A. L. *et al.* The homozygous CX3CR1-M280 mutation impairs human monocyte survival. *JCI insight* (2018). doi:10.1172/jci.insight.95417

Chapter 6: Conclusions and Future work

6.1 Discussion and Conclusions

Each of the chapters includes a summary of the work carried out described. This final chapter discusses conclusions jointly to form an overview of the whole thesis project. The main aim of this research was to explore the properties and boundaries of cell-free systems and *in vitro* display methods so that they could be used more extensively in conjunction.

The explored features of two different cell-free systems gave an insight into their dynamics and disclosed two different patterns for protein expression. Model analysis of the experimental results confirmed distinct pathways taken for cell-free systems from cell extracts and purified components. Results were consistent with the hypothesis that the extract-based system is mainly hindered by the presence of naturally found contaminants that decrease the availability of substrates. Methods to prevent or minimize this feature could be implemented in this type of cell-free system. For example, engineering of a minimal cell (top-down or bottom-up approach) could help eliminate these contaminants and enhance productivity. Once established, the economical and time benefits of the extract-based cell-free system could benefit many. The purified components based system (PURE system) had its limitation during protein translation, including depletion of resources and ribosome stalling. Since one of the main features of the PURE system is its ability for unbounded reaction modification and monitoring, these issues could be

resolved further by supplementing resources (amino acids, energy molecules, etc) and mechanisms to assist the reaction processes (ribosome recycling and rescuing factors) into the reaction to improve the final yield.

Apart from the cell-free discovered dynamics, single-base substitutions in a mutant library of T7 promoters allowed the discovery of variants with enhanced yield. With two simple, yet different, approaches, useful information could be obtained. Experiment data were recorded using straight forward time-course measurement of protein yield. Although a simple experimental approach, when combined with system modelling to create an interdisciplinary approach, useful information was revealed. Cell-free systems have potential and they are already used widely in different fields of synthetic biology. The dynamics of factor interaction could be easily disentangled with more effort.

The information obtained from this experimental approach, which included the testing of mutant libraries and the building of a mathematical model, provided a knowledge of the strengths and limitations of each system. When exploring display methods, such as mRNA or ribosome display, several aspects become important. One is the security that during the screening and selection process, that the peptides samples have high purity levels. Any presence of contaminants or undesired molecules could hinder the interaction of the peptides with the binding target. Peptides and mRNA molecules could also be degraded. This lowers the possibility of finding a good candidate and reduces efficiency. So when working with display method, it is important to ensure the cleanliness of the experimental

reaction. Also, to provide an efficient method of purification and sample recovery. During this thesis, these aspects were put to test. The inclusion of the PURE cell-free system into both mRNA and cDNA display was explored. The election of this cell-free system over others, especially cell extract-based ones, came down to the low levels of mRNA degradation. For both mRNA and cDNA display, an mRNA molecule is attached to each peptide. As a short unprotected single-strand molecule, its stability is already low and the probability for degradation high. Using the PURE system minimized the possibility of degradation by nucleases. It also reduced the interaction with other unnecessary elements that could in exchange destabilize the mRNA-peptide complex or disturb the selection process. The lack of contaminants and unwanted elements in the PURE system also allows efficient cleaning and purification procedures: the fewer components that need to be cleaned from the sample, the more efficient the procedure will be. The PURE system was used in both mRNA and cDNA display methods. Both of these were selected for their *in vitro* properties. They can process libraries up to 10^{15} in size. The complexes they form with the peptides are small and less prone to disassembly. And since they are performed *in vitro*, rounds of selection can be done much faster. When comparing both methods, their efficiencies almost matched. It was thought the cDNA display will show superior efficiency since the mRNA molecule becomes more stable by being attached to cDNA. And although in theory, this makes sense, in practice it was not seen. The most plausible explanation comes to the use of the PURE system and highly efficient cleaning

procedures. As it was already explained, these features improved the mRNA stability. With these conditions, the selection of mRNA display over cDNA display should be made. The mRNA-peptide complexes are smaller in size, which allows easier access to binding sites. But also by eliminating one single step, RT-PCR to form mRNA-cDNA, the number of extra components and procedures exposed to the mRNA-peptide complexes are reduced. And the whole selection process becomes faster.

The combination of mRNA display and the PURE system resulted in satisfactory screening and selection of binding motifs from a random peptide library. Further exploration of this combined method could enhance the discovery of high-affinity peptide candidates for drug discovery.

When it comes to the discovery of drug candidates for therapy treatment, one of the major candidates are membrane proteins. The location of these proteins, on the cell's surface, makes them easily accessed by drug molecules. They are also involved in a wide variety of cellular processes and pathways, which in return results in various diseases appearing when they malfunction. G protein-coupled receptors (GPCRs) are the most promising candidates for membrane proteins drug discovery. They are found mostly every cell, in significant quantities and are involved in almost every signaling pathway.

To research further into drug discovery for GPCRs, first more had to be understood about them. Since they are membrane proteins, they need the presence of synthetic lipid models to be produced and folded properly. The GPCR candidate

chosen to investigate this field of drug discovery in membrane proteins was the receptor CX3CR1. It is highly involved in several signalling pathways, making it important in disorders such as cancer, inflammatory syndromes and cell apoptosis. CX3CR1 also shows polymorphisms. This feature can be used in the future for individual therapy. Since different versions of the protein are found in a different individual, each one could answer differently to a specific antibody or peptide targets. The selection of synthetic membrane for CX3CR1 production came down to two choices: lipid vesicles and nanodiscs. Two options were used to investigate further how they both differ when being involved is screening and selection processes. Lipid vesicles act like bioreactors. They are basic synthetic lipid models that are affordable and quite fast to produce. But when it comes to the disadvantages of this model, the main aspect is the probability of wrong insertion direction of the membrane protein. The binding motif of the protein could end up being on the inside of the liposome. Nanodiscs do not have this problem, as they allow the membrane protein to be accessible from both ends. On the other hand, they are more costly to make and the process takes longer.

Production of CX3CR1 in both lipid models was also carried out by the PURE system, for the reasons already mentioned during the thesis. But apart from that, it was chosen for its ability to supplement the reaction with external components. In this case, elements were added to help the formation of disulphide bonds in the protein.

The production of CX3CR1 membrane protein was satisfactory, and the final aim of the thesis involved the design of a method to screen and select possible peptide targets for this protein. The use of the PURE system and mRNA display would be part of this methodology as they have been proven during this thesis to be incredibly valuable. The whole concept of this method as one will be explained more in detail in the next section (Section 6.2 Future Work). CX3CR1 has a natural ligand, called CX3CL1. The knowledge of this molecule and its binding sequence helped in the design of a possible DNA library for CX3CR1 drug screening. The chemokine domain of CX3CL1 binds to CX3CR1. But this domain is still quite long at 76 amino acids. Complete randomization of the whole domain is not possible as the total number of possible variants would be too high for any display system. But a variation of some of the smaller domain motifs is possible. In conjunction, all the motifs could give an idea of a variant chemokine domain protein with enhanced binding affinity for CX3CR1. Testing of smaller peptides instead of the whole domain is also possible and could yield important targets for the membrane protein.

To finalize, this thesis has brought insight into the use of cell-free systems, display methods, membrane protein and synthetic lipid membranes. In a partnership, the possibility to create a screening and selection method of drug targets for membrane proteins is possible. This thesis has demonstrated the advantages and efficiency of all the elements discussed. The next step to be taken is to use this in

conjunction in an interdisciplinary way to discover how far their possibilities can reach together.

6.2 Future Work

One important aspect that could be explored further is the molecular structure of the T7 promoter variants that both enhanced and lowered the final protein yield. The molecular aspect of the transcription process by the T7 RNA Polymerase (T7 RNAP) has been studied before [1]. And although explanations were hypothesised about the effect of the mutation's position within the promoter, more can be explored with the information already available about the T7 RNAP. By looking individually at each substitution and the molecular interaction the new nucleotide has with the T7 RNA polymerase, more can be learned about why each specific mutation has certain effects on the transcription process like what has already been done with the core T7 promoter sequence [2]. Combinations of more than one substitution in the promoter could enhance even more the activity of the T7 RNAP. Or just a random T7 promoter library could be further explored for intricate variations [3]. The existence of a T7 promoter library with a spectrum of promoter strength is valuable when designing genetic circuits. As every pathway, whether natural or synthetic, has different yield outcomes, multiple combinations could be designed in the same reaction environment.

When it comes to the membrane protein production in lipid vesicles, improvements are suggested here. The addition to the PURE system of certain elements can aid in the correct integration into the liposome surface. By

implementing the complex protein system known as translocation channel, the integration of GPCRs into the liposome membrane can be enhanced [4]. This could be used in further studies with different types of membrane proteins and GPCRs for liposome display methods.

The production and purification efficiency of CX3CR1 in nanodiscs is another aspect worth investing more time and resources. The purification carried out was performed using the HA-tag located in the CX3CR1 receptor. The first thing to improve would be to elongate the linker sequence between the receptor and the tag. This will allow better access of the HAtag to the antibody surface, and thus allowing more samples to be bound and not washed away. The second thing would be to test different tag. The HA tag has 44% hydrophobic amino acids. These are more prone to interact with the nanodiscs phospholipids and be unavailable for binding and selection. A possible option to use could be the FLAG-tag, as it has no hydrophobic amino acids. Apart from the attached tag, modification of both the PURE system components and size and amount of nanodiscs in the reaction, could improve the final product yield too. So these parameters also need optimization for each different membrane protein.

Different types of GPCRs are also worth testing. Although CX3CR1 was chosen, many more are also suitable targets, like CCR5. Furthermore, modifications of this protein could aid in their production or stability, making them more suitable for research. For example, insertion of a rubredoxin into the third intracellular loop of CCR5 enhanced thermostability [5–7].

The lipid composition used for both liposomes and nanodiscs was already pre-determined and for the expression of CX3CR1 it worked well. But this aspect has room to growth for the level of protein expression. Furthermore, if these samples were planned to be used in cells or for clinical trial, the composition is very important as the type of lipids used play an important role in membrane fluidity and absorption.

The final and most important aspect of future work for this thesis is the integration of all the acquired knowledge. The creation of a method that could screen and select possible targets for membrane proteins is the final goal. During this thesis, it was learnt the beneficial aspect of the PURE system, of the use of mRNA display for screening and selection processes. The production of a G-coupled receptor protein was carried out in two different types of synthetic lipid membranes. Even a possible random DNA library of possible drug targets was designed. The proposed system (Figure 6.A) is based on the mRNA display method used in chapter 4. It starts by using the PURE system to produce all the random mRNA-peptides conjugates from the designed library. The selection process will use the produced membrane receptors, in both vesicles and nanodiscs, to scan for high-affinity targets. By going through several rounds of selection, possible candidates and high binding motifs will be discovered.

GPCRs, like most membrane proteins, are extremely difficult to produce and purify efficiently. Also, due to the difficulty in working with membrane proteins, 3D structures of most of them are not available [8]. Screening of possible ligands for

membrane proteins not only can be performed for drug targeting purposes but also to understand better the binding and structures of the receptors they bind to. By analysing the sequences and motifs of the candidates with high affinity, patterns can be observed and hypothesis can be proposed about the membrane protein structure or binding regions. This approach could help dissect the structures of proteins from the same family or that have the same function. Predicting new candidates using computer modelling is limited by the lack of available data [9]. So the more data collected the better the aid in predicting new candidates for the target protein or other similar proteins. Furthermore, this lack of available data puts more focus on using High-Throughput Screening (HTS) methods that can cope with a high number of candidates in libraries and perform rapid testing [10]. Minimization of unwanted interactions of the candidates with other molecules is also necessary to increase as much as possible the efficiency of screening. The demanding conditions for drug screening of membrane proteins could be summary in categories: (1) Rapid and easy production and purification of membrane proteins, (2) high-throughput screening of big libraries and rapid testing, and (3) high efficiency and minimal side interactions of the candidates and/or membrane protein.

The proposed method (Figure 6.A) could meet these conditions. The production and purification of the GPCR receptor in both vesicles and nanodiscs were performed successfully. It was not a challenging task and could be performed within 1 day, as long as stocks of nanodiscs are present. The efficiency of

purification was acceptable (63% efficiency showed in chapter 5) and the amount of protein obtained enough to perform further experiments (0.021 mg/ml CX3CR1 quantification showed in chapter 5). But as already explained during this section, these parameters have the potential to be optimized. The use of mRNA display to screen and select for candidates already provides the possibility to test big libraries (10^{15}). And the fact that it is performed *in vitro* already reduces the time needed to perform one round of selection (24 h) and minimizes the presence of unwanted components. Finally, the PURE system, as it is composed of purified components, minimizes contaminants in the reaction that could cause unwanted interactions. It also aids in the efficiency rate for mRNA-peptide conjugates (40% rate showed in chapter 4). Combining all these elements could create an optimal system for the screening and selection of membrane proteins.

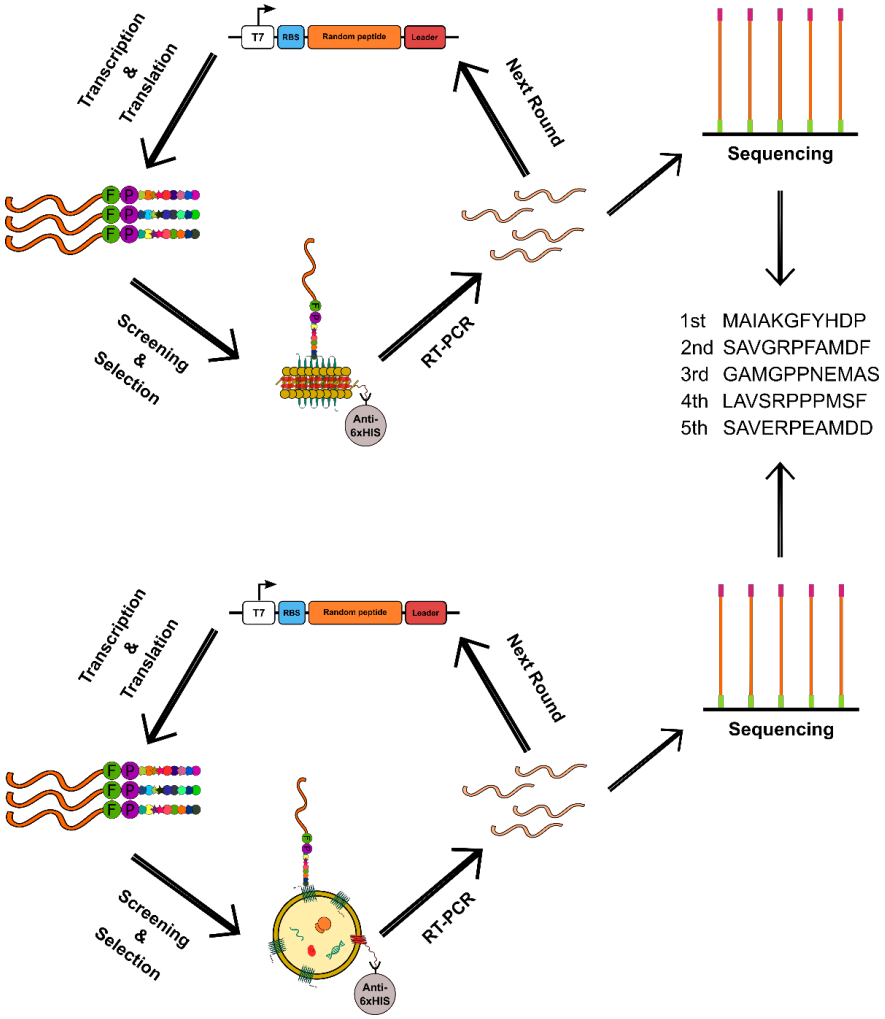


Figure 6.A: Scheme of possible mRNA display method for membrane proteins. The diagram shows how the display could be performed using membrane proteins in nanodiscs (upper) and lipid vesicles (lower). It starts by a random DNA library, which after translation by a cell-free system, it creates mRNA-peptides complexes. These conjugates are incubated with the membrane proteins to screen for possible targets. After binding and washing steps, the selected candidates will be sequenced. Several rounds can be performed to enhance the number of candidates with high binding affinity.

6.3 Chapter 6 References

1. Skinner, G. M., Baumann, C. G., Quinn, D. M., Molloy, J. E. & Hoggett, J. G. Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase: A single-molecule view of the transcription cycle. *J. Biol. Chem.* (2004). doi:10.1074/jbc.M310471200

2. Temiakov, D. *et al.* Structural basis for substrate selection by T7 RNA polymerase. *Cell* (2004). doi:10.1016/S0092-8674(04)00059-5

3. Komura, R., Aoki, W., Motone, K., Satomura, A. & Ueda, M. High-throughput evaluation of T7 promoter variants using biased randomization and DNA barcoding. *PLoS One* (2018). doi:10.1371/journal.pone.0196905

4. Kuruma, Y., Stano, P., Ueda, T. & Luisi, P. L. A synthetic biology approach to the construction of membrane proteins in semi-synthetic minimal cells. *Biochim. Biophys. Acta - Biomembr.* (2009). doi:10.1016/j.bbamem.2008.10.017

5. Tan, Q. *et al.* Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science (80-.).* (2013). doi:10.1126/science.1241475

6. Gessesse, B., Nagaike, T., Nagata, K., Shimizu, Y. & Ueda, T. G-Protein Coupled Receptor Protein Synthesis on a Lipid Bilayer Using a Reconstituted Cell-Free Protein Synthesis System. *Life* (2018). doi:10.3390/life8040054

7. Chun, E. *et al.* Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* (2012). doi:10.1016/j.str.2012.04.010

8. Martins, S. M. A., Trabuco, J. R. G., Monteiro, G. A. & Prazeres, D. M. GPCR screening and drug discovery: Challenges and latest trends. *Eur. Pharm. Rev.* (2012).

9. Jacob, L., Hoffmann, B., Stoven, V. & Vert, J. P. Virtual screening of GPCRs: An in silico chemogenomics approach. *BMC Bioinformatics* (2008). doi:10.1186/1471-2105-9-363

10. Sewing, A. & Cawkill, D. High-throughput lead finding and optimisation for GPCR targets. *Ernst Schering Foundation symposium proceedings* (2006). doi:10.1007/2789_2006_012

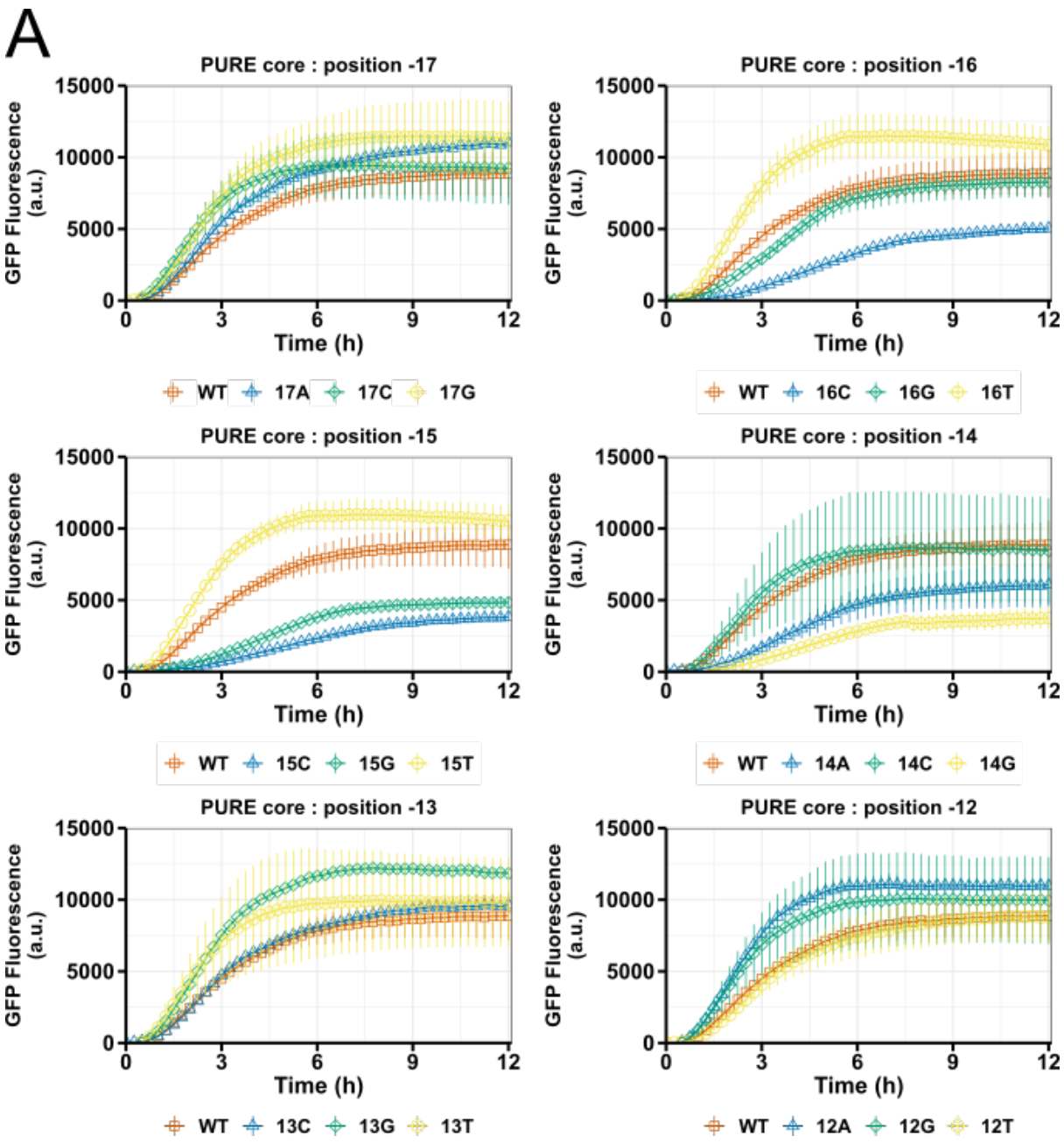
Appendix

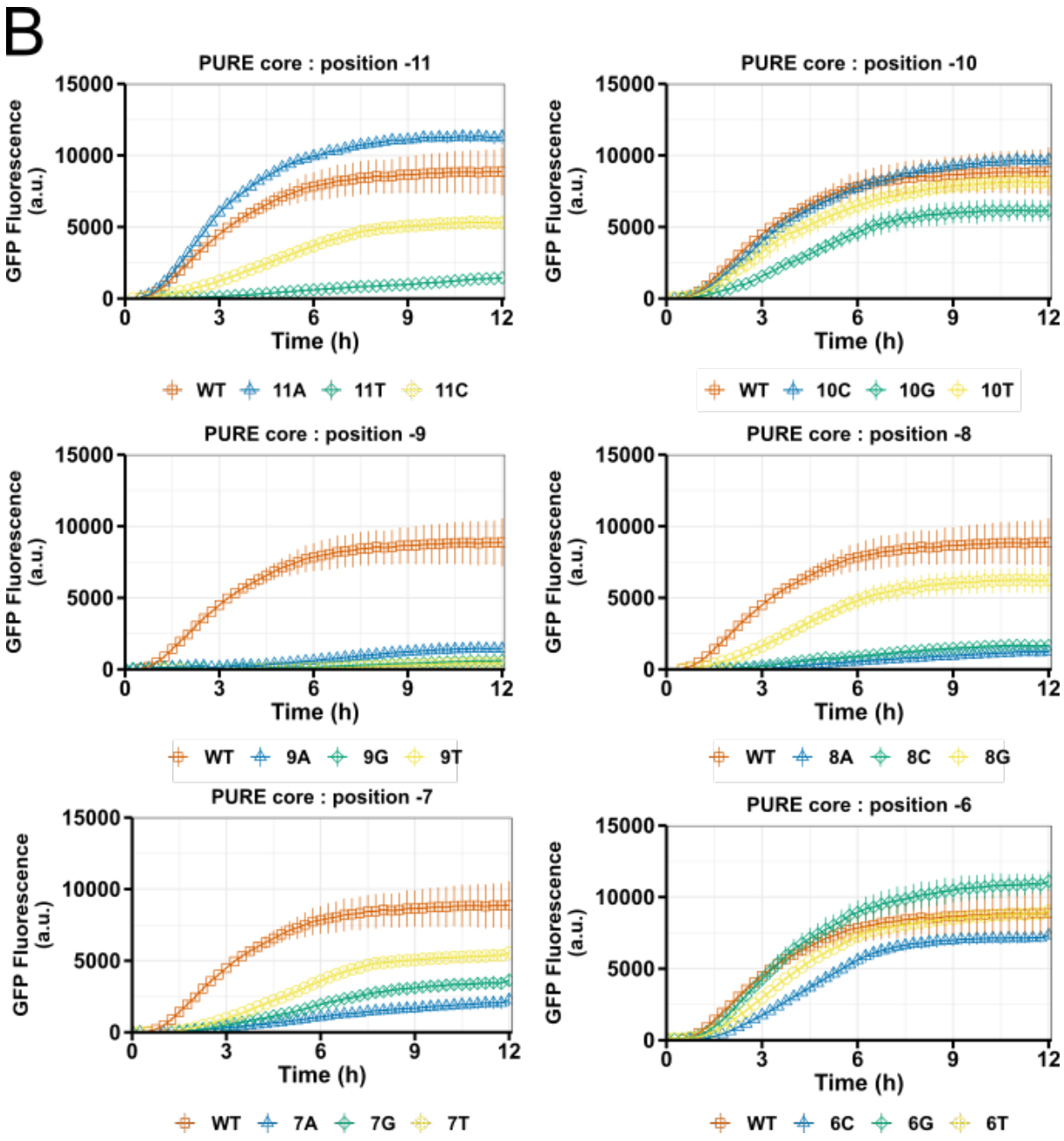
A.1 Appendix for Chapter 3

A.1.1 Single mutation plots

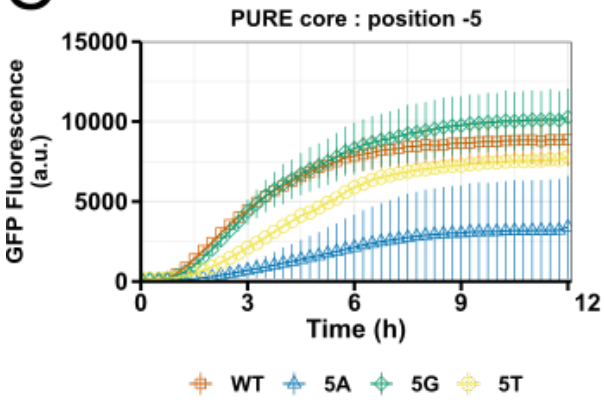
The individual plots of each variant (Figure A.1.1) were used to determine the expression rate of each one. It also shows differently the data found in the heatmaps (Chapter 3, Figure 3.G) as it can be seen that as the mutation advance more into the promoter sequence, they become more detrimental. Combined time-course plots for both cell-free systems (Figure A.1.2) shows better the wide range of protein expression variation that a mutant library provides. It also shows similar patterns as the simulated cases (Chapter 3, Figure 3.M) especially the cases with the extended sequences.

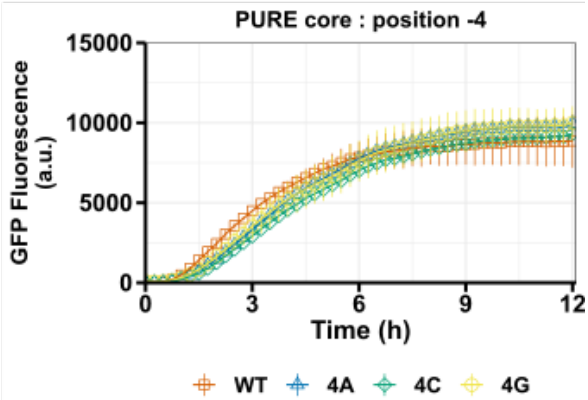
Figure A.1.1 (next page): Time-course of GFP fluorescence over a period of 12 h for all the 51 mutants. Each plot shows the time course of sfGFP expression by three variants for each position of the T7 promoter, along with that by the consensus sequence (denoted as WT) for comparison. (A, B, C) GFP expression in the PURE system, and (D, E, F) in the extract-based system.

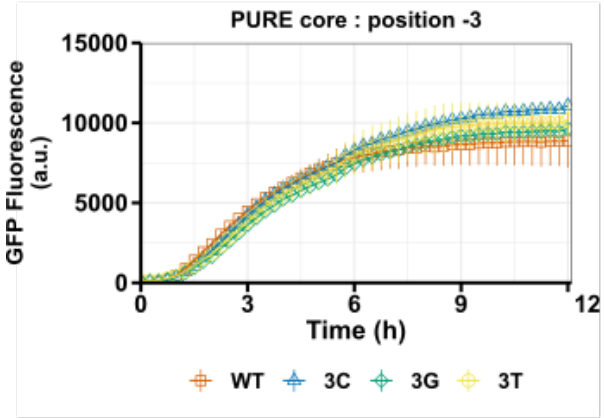


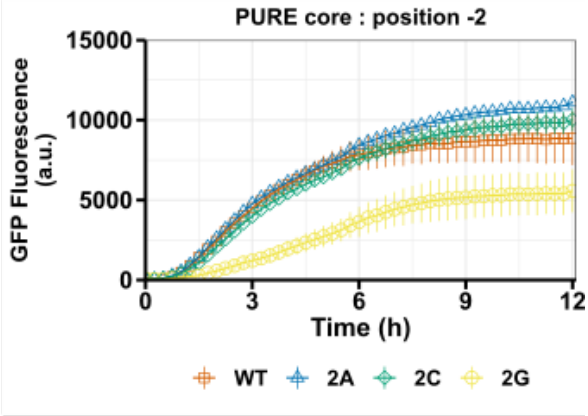

B
PURE core : position -11
PURE core : position -10
PURE core : position -9
PURE core : position -8
PURE core : position -7
PURE core : position -6
GFP Fluorescence (a.u.)
Time (h)
WT 11A 11T 11C
WT 10C 10G 10T
WT 9A 9G 9T
WT 8A 8C 8G
WT 7A 7G 7T
WT 6C 6G 6T

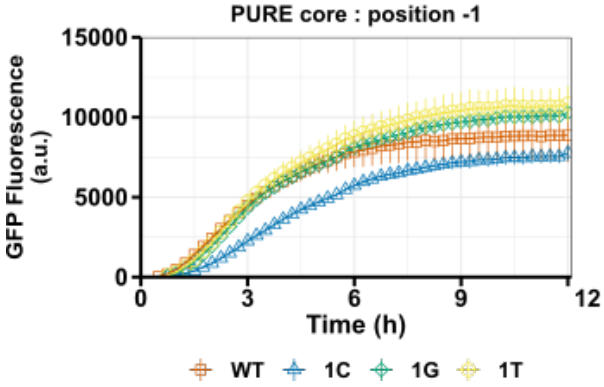
C


PURE core : position -5
GFP Fluorescence (a.u.)
15000
10000
5000
0
0
3
6
9
12
Time (h)
WT
5A
5G
5T

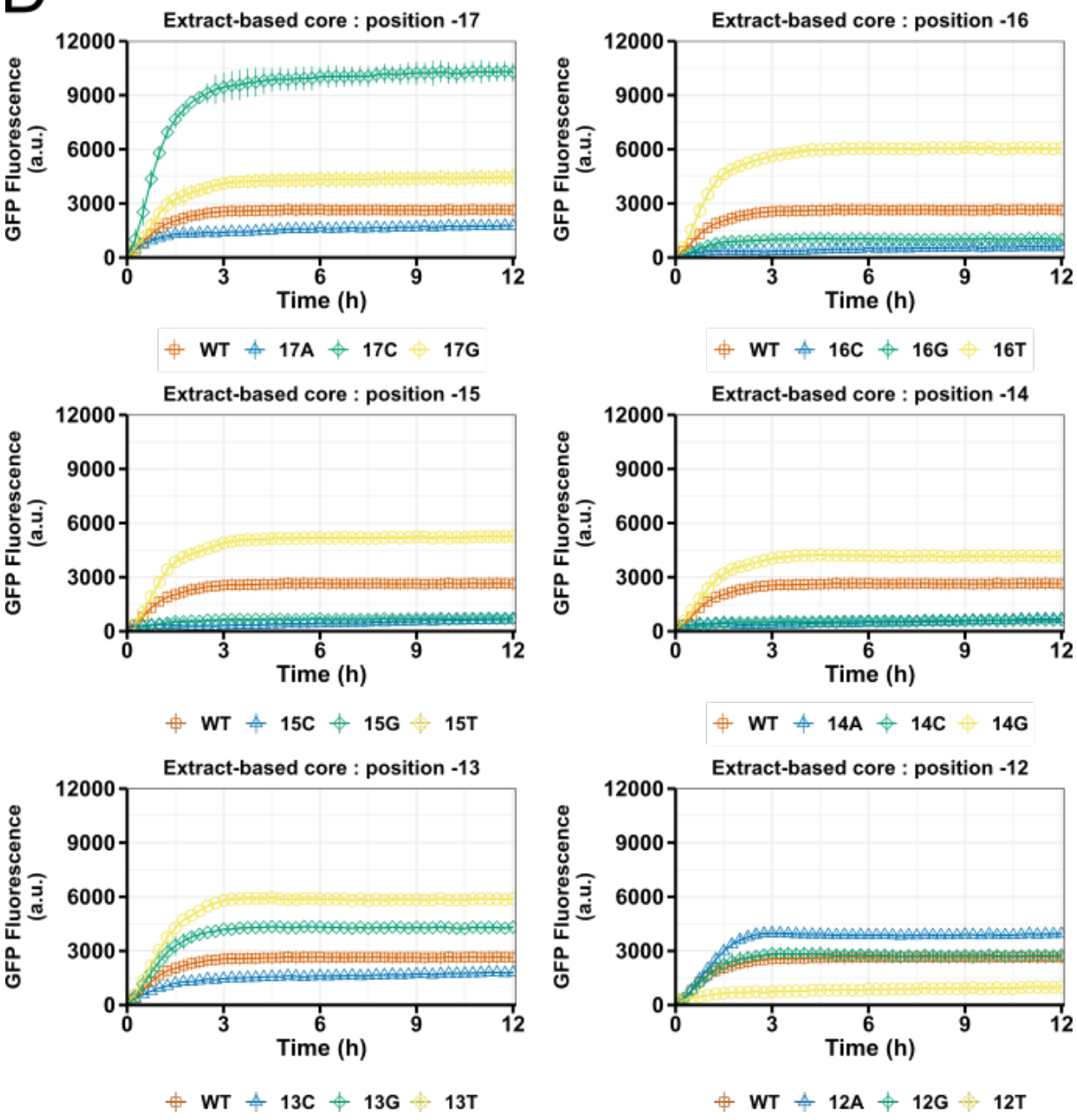

PURE core : position -4
GFP Fluorescence (a.u.)
15000
10000
5000
0
0
3
6
9
12
Time (h)
WT
4A
4C
4G


PURE core : position -3
GFP Fluorescence (a.u.)
15000
10000
5000
0
0
3
6
9
12
Time (h)
WT
3C
3G
3T

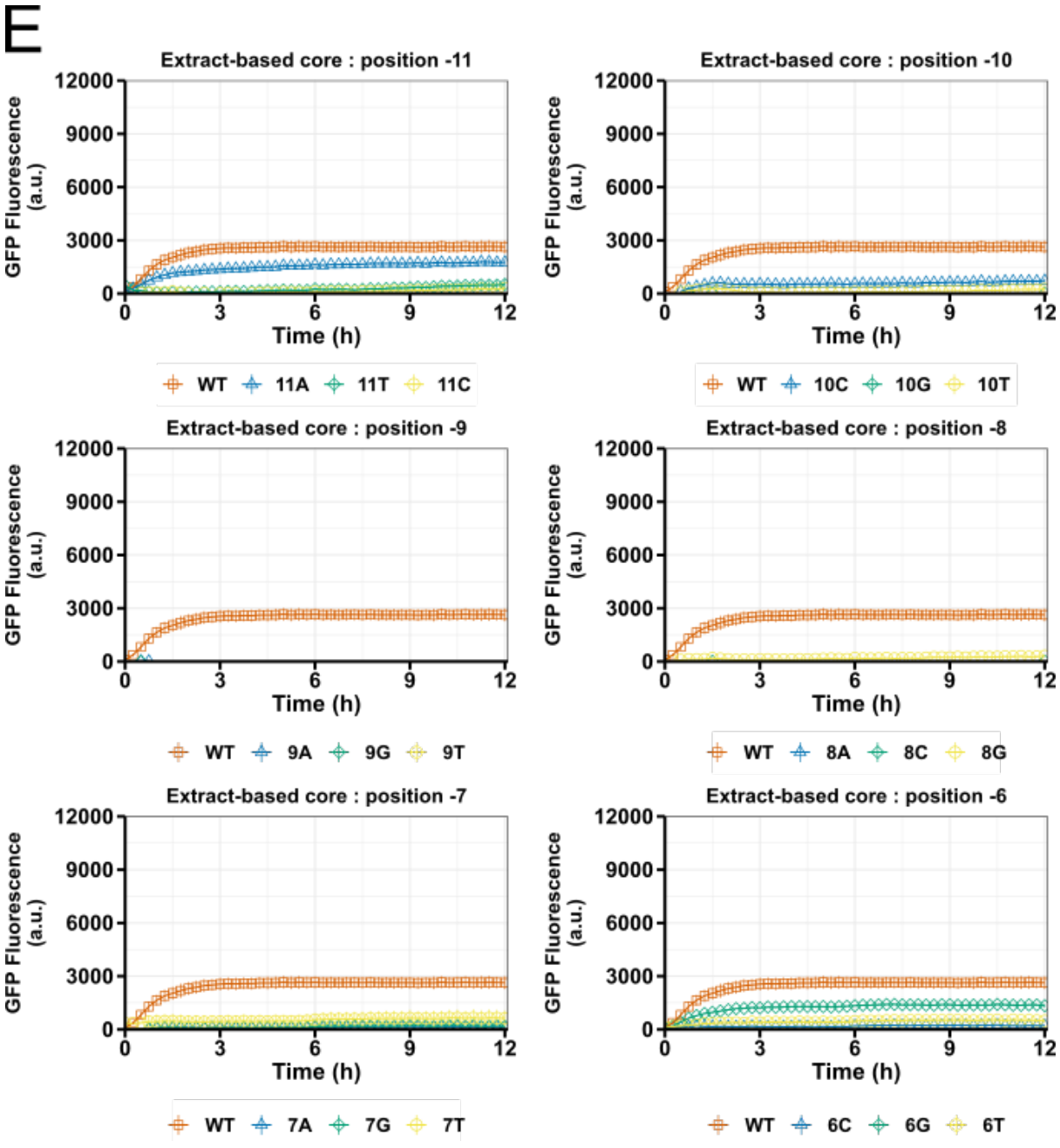

PURE core : position -2
GFP Fluorescence (a.u.)
15000
10000
5000
0
0
3
6
9
12
Time (h)
WT
2A
2C
2G


PURE core : position -1
GFP Fluorescence (a.u.)
15000
10000
5000
0
0
3
6
9
12
Time (h)
WT
1C
1G
1T

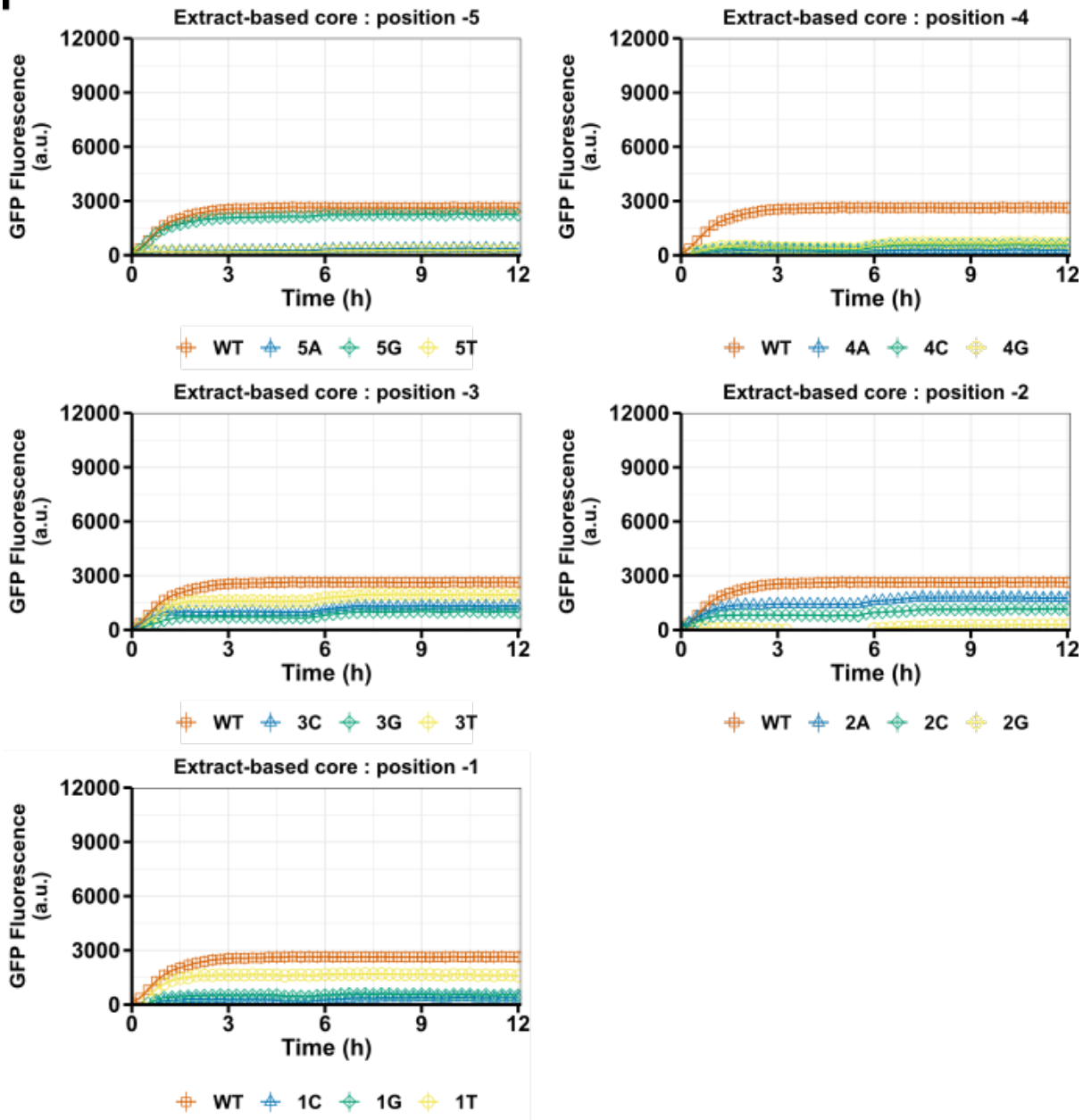
D


Extract-based core : position -17
GFP Fluorescence (a.u.)
Time (h)
WT 17A 17C 17G
Extract-based core : position -16
WT 16C 16G 16T
Extract-based core : position -15
WT 15C 15G 15T
Extract-based core : position -14
WT 14A 14C 14G
Extract-based core : position -13
WT 13C 13G 13T
Extract-based core : position -12
WT 12A 12G 12T

E


Extract-based core : position -11
GFP Fluorescence (a.u.)
Time (h)
WT 11A 11T 11C
Extract-based core : position -10
WT 10C 10G 10T
Extract-based core : position -9
WT 9A 9G 9T
Extract-based core : position -8
WT 8A 8C 8G
Extract-based core : position -7
WT 7A 7G 7T
Extract-based core : position -6
WT 6C 6G 6T

F


Extract-based core : position -5
GFP Fluorescence (a.u.)
12000
9000
6000
3000
0
0
3
6
9
12
Time (h)
WT 5A 5G 5T
Extract-based core : position -4
GFP Fluorescence (a.u.)
Time (h)
WT 4A 4C 4G
Extract-based core : position -3
GFP Fluorescence (a.u.)
Time (h)
WT 3C 3G 3T
Extract-based core : position -2
GFP Fluorescence (a.u.)
Time (h)
WT 2A 2C 2G
Extract-based core : position -1
GFP Fluorescence (a.u.)
Time (h)
WT 1C 1G 1T

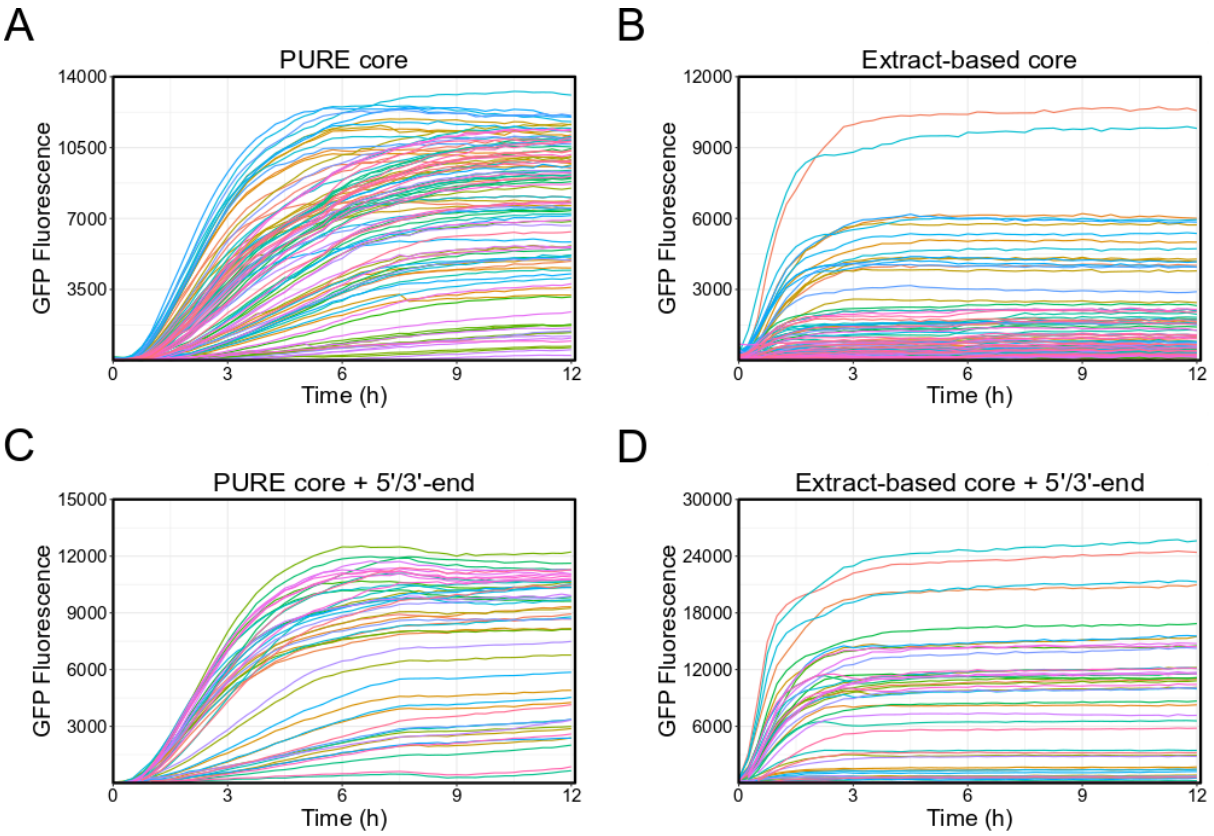


Figure A.1.2: Time-course of GFP fluorescence using the consensus T7 promoter sequence and its variants. The data used in Figure A.1.1 were plotted in one graph. (A) Time-course of protein expression for the core sequence variants in the PURE system, (B) the core sequence variants in the extract-based system, (C) the extended sequence variants in the PURE system, and (D) the extended variants in the extract-based system.

A.1.2 Reproducibility plots

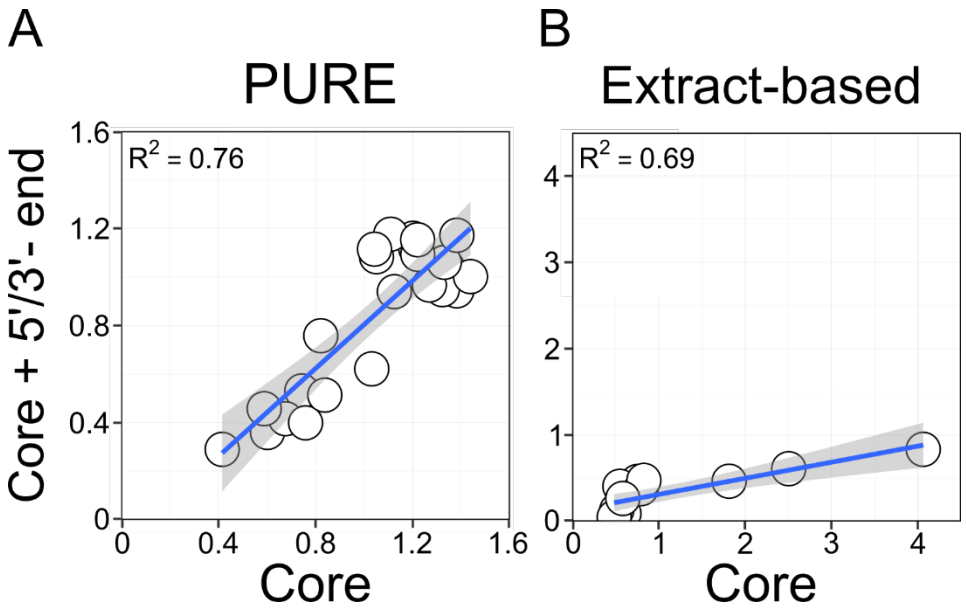


Figure A.1.3: **Scatter plots for comparing two linear DNA templates.** Normalized GFP fluorescence by the extended sequence (the core with extra bases at 5' and 3'-ends, y-axis) was plotted against that by the core sequence (x-axis). (**A**) The PURE system. (**B**) The extract-based system.

The reproducibility values (Figure A.1.2) comparing the core sequences and the extended versions show the difference in expression level that adding extra bases caused. The value is lower in the extract-based system proving that this base addition has a higher impact on this type of system. However, the correlation between extended and non-extended sequences showed good values suggesting that the gene expression levels are consistent overall.

A.1.3 Absolute expression levels

Table A.1.1: Fold change of the maximum GFP fluorescence (right column) for each variant (left column) expressed in the extract-based system. Values in the middle column show the absolute maximum values of a variant in the core and extended sequence, respectively.

Variant name	Core	Extended	Fold change
Consensus	2779	28456	10.24
-17A	1739.5	13117	7.54
-17C	10274.5	23652	2.30
-17G	4457	10596	2.38
-16C	579	2211	3.82
-16G	1011	2373	2.35
-16T	6066.5	17831	2.94
-15C	799.5	1487	1.86
-15G	714.5	1690	2.37
-15T	5270	15626	2.96
-14A	633	11837	18.70
-14C	883.5	3428	3.88
-14G	4148	1004	0.24
-13C	1973.5	10131	5.13
-13G	4316	12770	2.96
-13T	5878.5	16383	2.79
-12A	3950	12724	3.22
-12G	2760.5	17495	6.34
-12T	1105	10797	9.77
-11A	1823.5	13436	7.37
-11C	759.5	430	0.57
-11T	640.5	505	0.79
-10C	1132.5	7191	6.35
-10G	418	1036	2.48
-10T	696	3900	5.60
Average	2592.34	9604.21	4.60

The absolute expression levels increased more than four-fold on average for the extract-based system (Table A.1.1). For the PURE system, the sfGFP levels of the extended sequences were similar to those with no extra bases (Table A.1.2).

Table A.1.2: Fold change of the maximum GFP fluorescence (right column) for each variant (left column) expressed in the PURE system. Values in the middle column show the absolute maximum values of a variant in the core and extended sequence, respectively.

Variant name	Core	Extended	Fold change
Consensus	7424	9160.5	1.23
-17A	11243.5	9151	0.81
-17C	8517	9336	1.10
-17G	10274	10016.5	0.97
-16C	5056	4528	0.90
-16G	7906	5624	0.71
-16T	10061	9036	0.90
-15C	3821.5	2693.5	0.70
-15G	4532	3233.5	0.71
-15T	9436.5	9273.5	0.98
-14A	7693	10279	1.34
-14C	6153.5	6836.5	1.11
-14G	3563.5	3991.5	1.12
-13C	9738.5	9850	1.01
-13G	11084.5	8245.5	0.74
-13T	9304.5	9340.5	1.00
-12A	9901.5	8811	0.89
-12G	9132.5	9297	1.02
-12T	8813	9612.5	1.09
-11A	11248	9752.5	0.87
-11C	5540.5	1554	0.28
-11T	1675.5	208	0.12
-10C	9629.5	9571	0.99
-10G	6302.5	2913	0.46
-10T	8077	9100.5	1.13
Average	7845.16	7256.6	0.92

A.1.4 Modelling plots

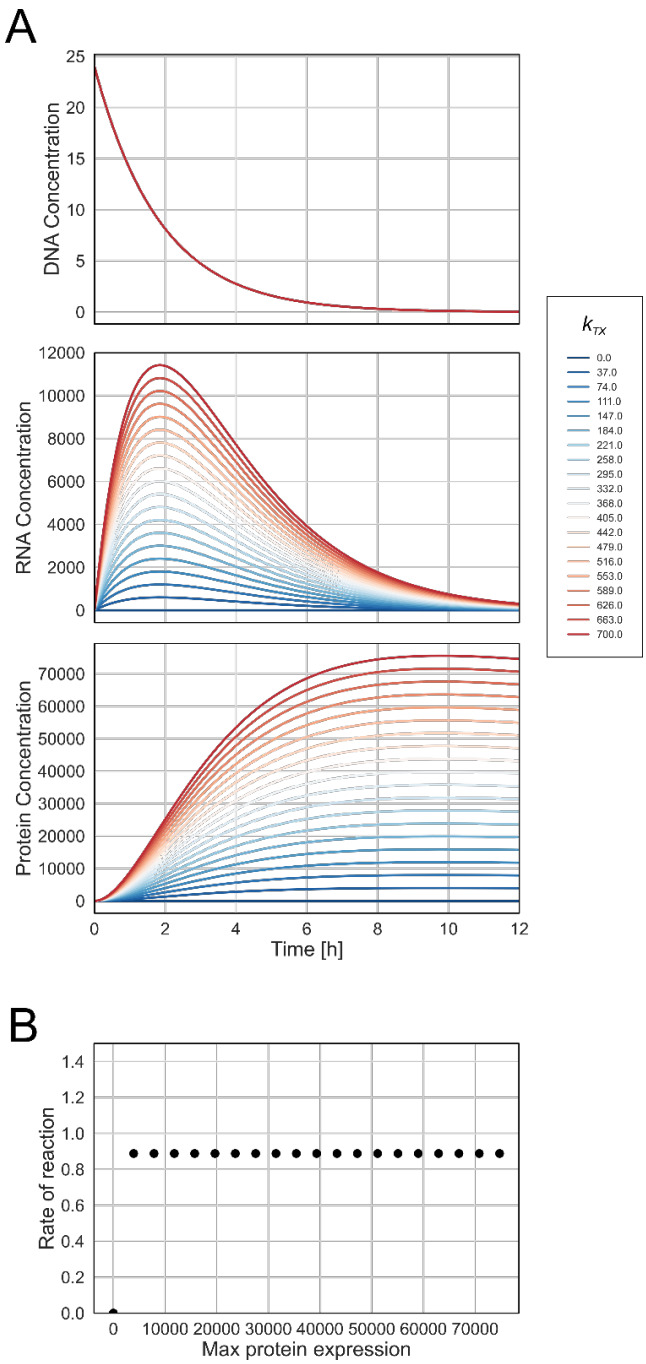


Figure A.1.4: (A) Simulated cell-free protein expression in the PURE system using the standard model. DNA concentration (top), mRNA concentration (middle), and protein expression (bottom) were shown. The parameter k_{TX} was varied (colours) while the other parameters were fixed. (B) Scatter plot of the rate of protein expression against the maximum protein expression. The values were obtained by fitting the simulated data in (A) to a logistic curve.

This plot (Figure A.1.3) shows the simulation performed on the PURE cell-free system with the same equations that were used for the extract-based system. Upon performing this simulation it could be seen that the expression patterns matched the observed ones, but the unique scatter pattern was not seen (Chapter 3, Figure 3.K, A). This implied that some factors were not being taken into consideration in the theoretical model and thus it was adjusted for the PURE system (Chapter 3, Equation 3.A).

A.2 Appendix for Chapter 4

A.2.1 DNA Sequences

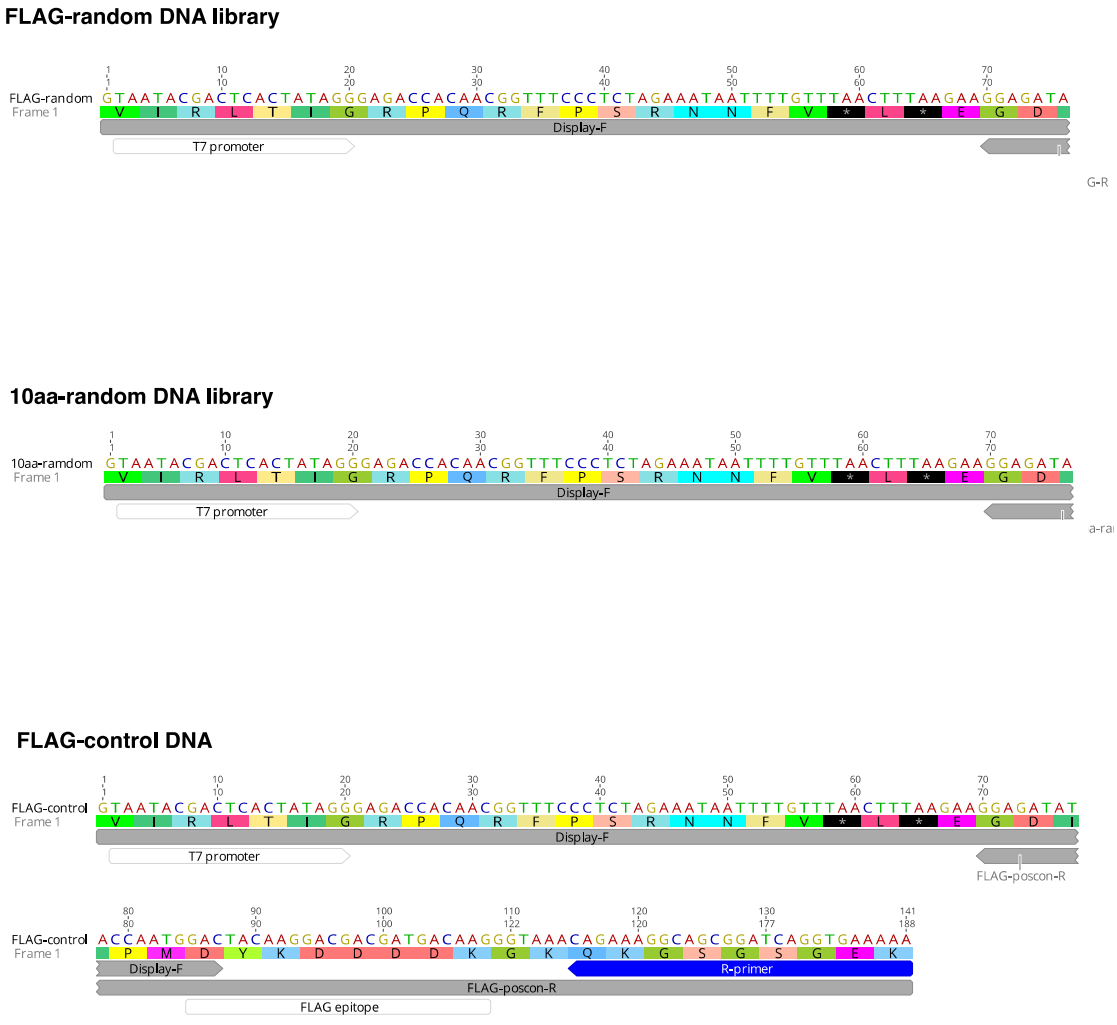


Figure A.2.1: DNA library design. Two DNA libraries (FLAG-random and 10aa-random) were designed to validate the performance of our newly refined mRNA, cDNA display method. FLAG-control DNA is used specifically for western blot analysis of mRNA-peptide and mRNA/cDNA-peptide conjugate confirmation.

The 'FLAG-random library' contained a fixed sequence for the first five amino acids of the FLAG epitope (DYKDDDDK); followed by three consecutive degenerate "RRN" codons corresponding to seven amino acids (Asn, Lys, Asp, Glu, Ser, Arg and Gly). The '10aa-random library' contained ten degenerate codons ("VNN" and "NNY") which corresponded to 16 and 17 different amino acids, respectively. The number of different sequences in both FLAG-random and 10aa-random libraries was 343 and 1.7×10^{12}, respectively.

Both DNA libraries contain a T7 promoter and ribosome binding site (RBS), for the use with a T7 RNA Polymerase. The leader sequence at the 3' region was complementary to the puromycin-FITC DNA tag (Figure A.2.2) for posterior ligation.

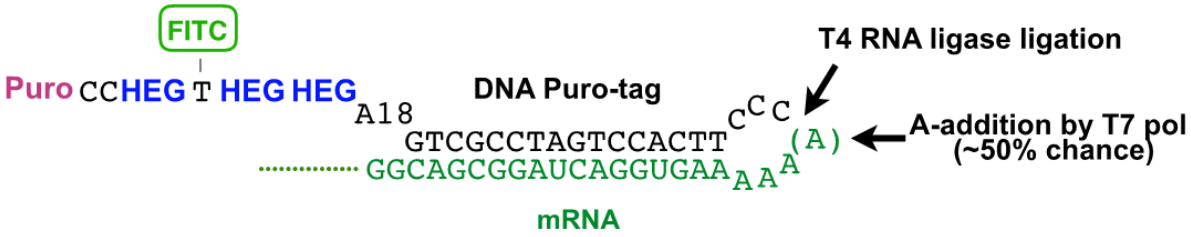


Figure A.2.2: Detailed image of the puromycin-FITC DNA tag ligation process. Parting from the DNA library, *in vitro* transcription was performed obtaining mRNA constructs. Afterwards, puromycin tags (DNA puro-tag) were attached by T4 RNA Ligase ligation.

The DNA Puro-tag (Figure A.2.2) is ligated to the mRNA sequence by the T4 RNA ligase by connecting the 3' end from the single-stranded mRNA construct to the phosphorylated 5' end of the DNA tag. This tag will later react with the peptidyl-tRNA to form a covalent link between mRNA and peptide. to connect. A rare GGA codon was placed at the 3' end and a polyA (18 nt) in the puromycin-FITC DNA tag to help the molecule incorporate into the stalled ribosome.

A.2.1 Gel images

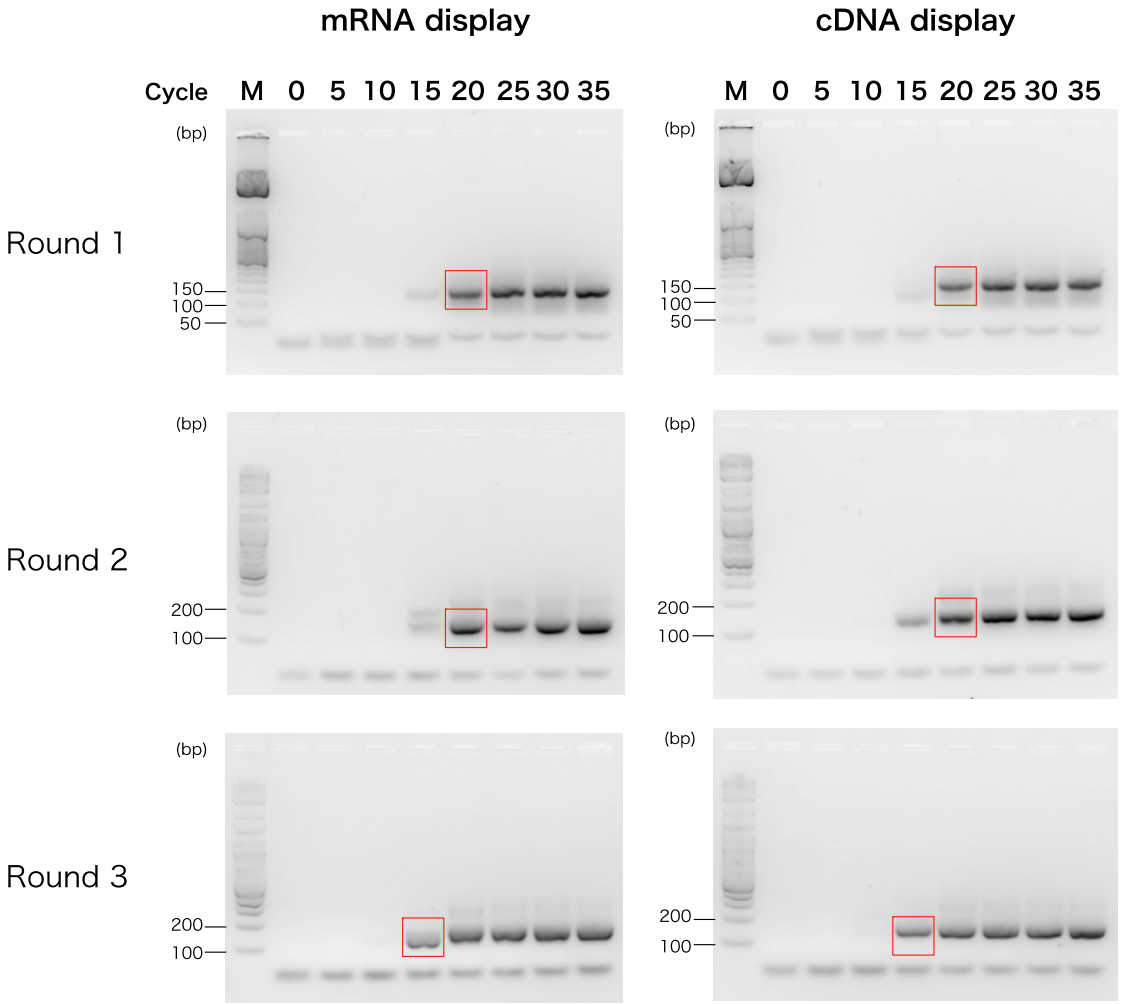


Figure A.2.3: Reverse transcription (RT)-PCR after selection. Affinity selected mRNA-peptide conjugates (mRNA display) and mRNA/cDNA-peptide conjugates (cDNA display) bound anti-FLAG M2 antibody magnetic beads were subjected to RT-PCR. For each round, total 6 µl x 8 reaction tubes were prepared for each display method to monitor the yield of amplified product for every 5 cycles (0 to 35 cycles). Expected size is a mixture of 141 bp (FLAG-random) and 150 bp (10aa-random).

A test was performed (Figure A.2.3) to determine the optimal number of cycles for the RT-PCR. This is to avoid over-amplification of the samples. For each round of selection, the amplified product was checked every 5 cycles (from 0 to 35 cycles). Amplified DNA products, started being seen at the 15th cycle and were saturated at the 20th cycle. Band intensity of the 15th RT-PCR cycle was higher than previous rounds. This proves that the quantity of the sample was higher as the rounds progressed. A non-specific band, probably from non-specific binding of primers, appeared in every lane at around 200 bp. The optimal number of cycles for the first two wounds was 20 cycles and for the 3 rd round, 16.

A.3 Appendix for Chapter 5

A.3.1 Microscope visualization

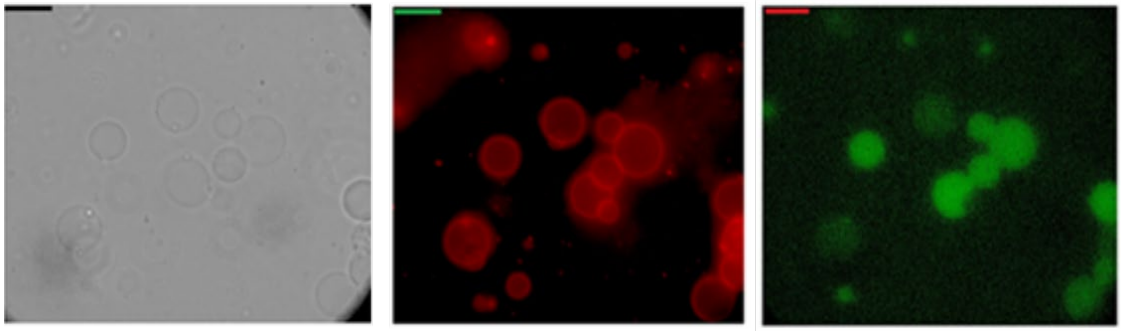


Figure A.3.1: Microscope images of lipid vesicles using the PURE system and sfGFP as the reporter gene. Vesicles have a lipid composition of POPC: Cholesterol (1: 1). The inner solution contains the PURE*frex* 1.0 system and sfGFP DNA. The image on the left is phase contrast. The one on the middle shows fluorescence of NileRed attached to the lipid membrane. The image of the right represents the green fluorescence of sfGFP being produced inside the vesicles by the PURE system. The bar size, on the left upper corner of each image, is 50 μm.