Semi-permeable artificial cells and their applications

Damian Van Raad
2023

A thesis submitted for the fulfilment of requirements of the Doctor of Philosophy

Research School of Chemistry

The Australian National University

Declaration of original work

This thesis, and the results contained, are my own original work except where due reference has been made. The contents of this thesis are a product of my research degree candidature and have not been submitted for any other degree or diploma

Acknowledgements

I would like to begin by expressing my heartfelt gratitude to my supervisor, Professor Thomas Huber. I feel incredibly fortunate that he gave me the chance to join his group six years ago when I expressed my interest. He provided me with the incredible opportunity to embark on this exciting journey as a researcher in this field. Throughout my PhD, Thomas has been an excellent guide, offering me invaluable scientific, and moral support, and I couldn't have asked for a more meticulous and caring supervisor. Our countless discussions have been profoundly significant, and I will always cherish them as some of the most memorable moments of my career.

I am deeply grateful to Professor Gottfried Otting for his unwavering support and guidance throughout my research, particularly during the intricate process of conducting NMR experiments. His selfless dedication and valuable insights have been instrumental in shaping my work. I cannot express enough my appreciation for the significant amount of time and effort he has devoted to providing me with critical advice that has helped me in advancing my research.

I would like to express my sincere gratitude to all the members, past and present, of the Huber and Otting groups who have helped, supported, and befriended me during the past four years. Notable mentions Sven Ullrich, Henry Orton, and Elwy Abdelkader for always being there for a chinwag.

I would like to take a moment to express my heartfelt gratitude to God and my parents, who have been my constant support and source of inspiration throughout my journey. Without their unwavering love, encouragement, and guidance, I would not have been able to achieve my goals and aspirations.

To God, I am deeply grateful for blessing me with the strength, wisdom, and resilience to overcome challenges and pursue my dreams. Your grace and mercy have been my guiding light and my hope in difficult times.

I dedicate this thesis to Phoebe Jean Hodge, who has been a constant source of love, support, and inspiration in my life. Phoebe, your kindness, generosity, and selflessness have been a shining example to me, and I am grateful for all the ways in which you have enriched my life. I am honoured to dedicate this thesis to you as a small token of my appreciation.

Abbreviations

AA	Amino acid
aaRS/RS	Amino acyl tRNA Synthetase
AcP	Acetyl Phosphate
CAT	Chloramphenicol acetyl transferase
CFPS	Cell Free Protein synthesis
CK	Creatine kinase
CP	Creatine Phosphate
E. coli	*Escherichia coli*
EF-Tu	Elongation factor - thermo unstable
EF- G	Elongation factor - G
FACS	Fluorescent activated cell-sorting
Fig	Figure
FSC	Forward scatter
GFP	Green fluorescent protein
Glu^-	Glutamate
GMML	Glycerol minimal media
GSH	Glutathione reduced
GSSG	Glutathione oxidised
HMP	Hexametaphosphate
IPTG	Isopropyl β-D-1- thiogalactopyranoside
LbL	Layer by layer assembly
MD	Maltodextrin
M.j	*Methanocaldococcus jannaschii*
M.b	*Methanosarcina barkeri*
M.m	*Methanosarcina mezei*

mRNA	Messenger ribonucleic acid
ncAA	Non-canonical amino acid
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at 600nm
S30	Supernatant 30,000g
S12	Supernatant 12,000g
SAIL	Stereo Array Isotope Labeling
SSC	Side scatter
PaF	p-acetyl-L-phenylalanine
PDI	Protein disulfide isomerase
PEP	Phosphoenol pyruvate
PI	Protease inhibitor
pSer	O-phospho-L-serine
PPK	Polyphosphate kinase
PpiB	peptidyl-prolyl cis-trans isomerase B
pThr	O-phospho-L-threonine
PTM	Post-translational modification
PylRS	Pyrrolysl amino acyl tRNA synthetase
RF1	Release factor 1
RNAP	Ribonucleic acid polymerase
$tRNA_{sup}$/$tRNA_{CUA}$	Suppressor tRNA
tRNA	Transfer ribonucleic acid
TyrRS	Tyrosyl amino acyl tRNA synthetase
WT	Wild type

Abstract

Biotechnology depends on our ability to produce proteins rapidly, in high purity and at large scale. *In vivo* protein production in bacterial cells is generally the preferred choice to economically produce large quantities of protein, and cell-free protein synthesis is applied for specialized applications or when proteins are required quickly.

The scope of this thesis is the development of the encapsulated cell-like structures (eCells), which can be considered a hybrid approach at the intersection of *in vivo* and *in-vitro* protein expression. Chapter 2 of this thesis introduces a method for efficiently producing encapsulated bacterial lysate which can be applied to cell-free protein synthesis for recombinant protein expression. This facile method produces eCells using layer-by-layer polymer assembly to coat bacterial cells prior to lysis of their cell wall. As a result, eCells maintain many of the advantages of living cells while at the same time allow full control over the chemical environment. Individual eCells are separated by fluorescent activated cell-sorting (FACS) which highlights the femto-liter size and demarcation of unique protein features housed in a semipermeable polymer capsule.

Chapter 2 focusses on site specific non-canonical amino acid incorporation (ncAA) utilised in eCells, with amino acyl-tRNA synthetases such as the pyrrolysyl-tRNA synthetase which is typically inactive when purified and added to standard cell-free protein synthesis. Using PylRS synthetase, photocaged-cysteine and N^{ε}-(tert-butoxycarbonyl)-L-lysine was incorporated into peptidylprolyl cis−trans isomerase B (PpiB) using eCells.

eCells are not living entities, although amino acid anabolism has been shown to be functional. Chapter 3 and 4 demonstrate the application of eCells to isotope label proteins atom selectively and a reduction of the cost of isotope labelling proteins for NMR spectroscopy. The use of eCells, the strategically chosen precursor and the omission of the amino acid from the reaction mixture allow the production of proteins with residue and atom specific isotope labelling patterns, increasing the sensitivity of ^{13}C HSQC spectra with the elimination of one-bond ^{13}C-^{13}C scalar couplings. Chapter 3 introduces new approaches to selectively label methyl groups of valine, leucine, isoleucine and alanine residues in proteins. In addition, stereospecific and residue specific labelling of pro-*S* signals of leucine and valine residues has been achieved in eCells using the precursor 2-^{13}C-methylacetolactate in a perdeuterated protein environment. Chapter 4 explores labelling approaches for aromatic residues for NMR

spectroscopy. Starting from 3-^{13}C-pyruvate, the proteins PpiB and ubiquitin were selectively labeled in the ε and δ positions of Phe/Tyr and analysed by NMR.

eCells provide advantages over cell-free protein synthesis and *in vivo* expression. In Chapter 5 (i) it is demonstrated that the ability to influence protein folding and change the redox environment is an attractive feature of eCell CFPS. Human protein disulfide isomerase (PDI) which is expressed prior to encapsulation, helps proteins express in a soluble disulfide bonded form. Using eCells and PDI, two proteins that previously could not be expressed actively in bacterial cells, neurotrophic growth factor neuritin and active human light-chain enterokinase, were expressed.

Chapter 5 (ii) highlights that eCells have shown to be effective regarding bioremediation and (iii) biosensing. Using this system in conjunction with specific biocatalytic enzymes for an organophosphate pesticide, Paraoxon-ethyl, the contaminant can be inactivated and made inert, eliminating it from the natural environment. eCells are also able to detect heavy metals such as Hg^{2+} with a lower limit of detection and higher sensitivity than *in vivo* biosensing. Chapter 5 (iv) introduces that eCells can be made with other derivatives of *E. coli* and using these eCells, site specific introduction of phosphorylated groups can be achieved.

Table of Contents

List of Publications*

This thesis is founded on five manuscripts that serve as the basis for comprehensive and innovative research on semi-permeable artificial cells, also known as eCells. The included manuscripts delve into the design of eCells, relevant theories, and potential applications of eCell technology. The use of these manuscripts has enabled the advancement of the knowledge surrounding this topic and has contributed substantially to the existing body of knowledge within this field. Consequently, my thesis presents a significant and valuable contribution to this field while reflecting my effort in building on and extending the existing knowledge.

Cell-free protein synthesis – the powerhouse of protein expression*

Damian Van Raad, Gottfried Otting and Thomas Huber

Submission to Journal of Microbial Biotechnology- WILEY

In this work as the main investigator, I was responsible for carrying out and drafting the initial version of the manuscript.

Damian Van Raad and Thomas Huber. (2021**).** ***In vitro*** **protein synthesis in semipermeable artificial cells**. ACS Synth. Biol., 10, 1237-1244. https://doi.org/10.1021/acssynbio.1c00044

In this work as the main investigator, I was responsible for carrying out all aspects of protein expression, conducting kinetic experiments, thoroughly analyzing all the data obtained, and drafting the initial version of the manuscript.

Van Raad, D., Otting, G. & Huber, T. (2023) **Cell-free Synthesis of Proteins with Selectively ^{13}C-Labeled Methyl Groups from Inexpensive Precursors**. Mag. Res., 4, 187-197. https://doi.org/10.5194/mr-4-187-2023

In this work as the main investigator, I was responsible for carrying out all aspects of protein expression, conducting NMR experiments, thoroughly analyzing all the data obtained, and drafting the initial version of the manuscript.

Van Raad, D., Huber, T. & Otting, G. (2023) **Improved spectral resolution of [^{13}C,^{1}H]-HSQC spectra of aromatic amino acid residues in proteins produced by cell-free synthesis from inexpensive ^{13}C-labelled precursors.** J. Biomol. NMR., 78, 1 - 8. https://doi.org/10.1007/s10858-023-00420-9

In this work as the main investigator, I was responsible for carrying out all aspects of protein expression, conducting NMR experiments, thoroughly analyzing all the data obtained, and drafting the initial version of the manuscript.

Van Raad, D. & Huber, T**. eCell Technology for Cell-Free Protein Synthesis, Biosensing, and Remediation**. Adv. Biochem. Eng. Biotechnol. Springer, Berlin, Heidelberg. 1-18. https://doi.org/10.1007/10_2023_225

In this work as the main investigator, I was responsible for carrying out all aspects of protein expression, conducting NMR experiments, thoroughly analyzing all the data obtained, and drafting the initial version of the manuscript.

Damian Van Raad and Thomas Huber. (2022). *U.S. Patent No. US20220090046A1*. Washington, DC: U.S. Patent and Trademark Office.

In this patent as a co-inventor, I was responsible for carrying out and liaising with the Australian patent office (IP Australia), the technology transfer office of The ANU and patent lawyers.

CHAPTER ONE

Cell-free protein synthesis – the powerhouse of protein expression

1.1 Abstract

Cell-free protein synthesis (CFPS) is a versatile technology that is regularly used in biotechnology as a platform to rapidly produce biologics, proteins with post-translational modifications (PTMs), proteins and peptides containing noncanonical amino acids, and other high value protein products. CFPS has evolved significantly over time, increasing recombinant protein yield, and simplifying the process of cell lysate preparation. Major advancements to increase protein production yields include extending the productive half-life time of CFPS and have relied on the reduction or recycling of inorganic phosphate (P_i) using glycolytic intermediates as P_i acceptors. Cell lysates from new organisms have been prepared and tested for their ability to increase protein yield, with *Vibrio natriegens* being an efficient alternative to *Escherichia coli,* providing a larger volume of translationally active lysate per liter of growth media and the ability to prepare lysate from the stationary growth phase. CFPS has been applied to many applications, largely successfully. Post-translational modifications (PTMs) have been predominantly carried out in eukaryotic extracts. However, recent advancements in prokaryotic lysate have shown significant progress in achieving PTMs such as N-linked glycosylation, prenylation, and phosphorylation. Homogenous proteins with specialized PTMs can be readily produced in *E. coli* extract through genetic code expansion and targeted glycosylation/prenylation. NMR labelling techniques are efficient and cost effective when applied in CFPS, where reduction in homonuclear coupling can be achieved through the addition of isotopologue amino acids (AAs).

1.2 Introduction

Cell-free protein synthesis (CFPS) platforms have come to age since Nirenberg and Matthaei used cell extracts in the elucidation of the genetic code (Nirenberg and Matthaei, 1961). Still largely based on the crude cell extracts of organisms, CFPS has been readily accepted as a

powerful tool in synthetic biology to produce protein. It has been applied to the production of glycosylated protein in bacterial lysates, discovery of malaria vaccine candidates and paper-based biosensors for SARS-CoV-2 detection (Tsuboi et al., 2008; Hershewe et al., 2020; Cao et al., 2021). CFPS is an open system that is amenable to chemical change, allowing researchers to be unconstrained by the limitations of traditional *in vivo* cell cultures. As the homeostasis of a living cell is removed, all energy can be directed to the production of the recombinant protein, with unencumbered control over the chemical environment (Carlson et al., 2012). This allows production of proteins in a wide variety of chemical environments. CFPS has been applied in aqueous-organic biphasic systems, with oxidizing chemical environments, and from lyophilized lysates (Oh et al., 2006; Liu et al., 2020; Yang et al., 2021), highlighting its versatility as a protein production platform.

Although CFPS is a matured technology, further improvements are necessary to reach protein yields comparable with *in vivo* expression (Perez et al., 2016). A particular shortcoming of CFPS is the limited lifetime of the protein translation machinery. CFPS systems are generally most productive within the first 1.5 - 6 hours, and protein production rates are significantly reduced after (Pedersen et al., 2011). Transcription machinery, and consequently mRNA concentration, appear not to be limiting factors in CFPS reactions when T7 bacteriophage RNA polymerase is used, which generates a large amount of mRNA that can be reused (Kim and Swartz, 2001). The main issue surrounding the cessation of protein translation is attributed to the accumulation of inorganic phosphate (P_i) during the CFPS reaction. It has been hypothesized that high concentrations of P_i chelates magnesium ions, and the lack of free magnesium stalls protein synthesis (Kim et al., 2011).

CFPS reaction yields drastically change when the concentration of magnesium changes or magnesium is sequestered by P_i. Kim and co-workers showed that there is a minimum required magnesium concentration in ribosomal protein synthesis, and when free magnesium concentration drops below 12 mM, protein synthesis is inhibited (Kim et al., 2006 (a)). P_i is inhibitory to protein synthesis at concentrations greater than 30 mM (Kim and Swartz, 2001). In CFPS, >20 mM P_i is generated after 3 hrs incubation at 37^{o}C when creatine phosphate (CP)/creatine kinase (CK) is used as an energy source (Li et al., 2017). Ribosomal translation is

the main energy consuming process in CFPS using 4.5 to 5.9 ATP equivalents for each amino acid addition to the polypeptide chain. 2 ATP molecules are required for activation of the amino acid, 1 GTP for elongation factor thermo unstable transferring the $_{aa}$tRNA to the A site of the ribosome and 1 GTP for elongation factor G to catalyze translocation of the ribosome (Amthor, 2000, W; Liu et al., 2015; Chen et al., 2016). As a result, approximately 40 mM of ATP equivalence is consumed in the synthesis of 1 mg of a typical 25 kDa protein (Calhoun and Swartz, 2007). The high energy consumption and the inhibitory effect of free phosphate makes choosing the ATP generation method critical.

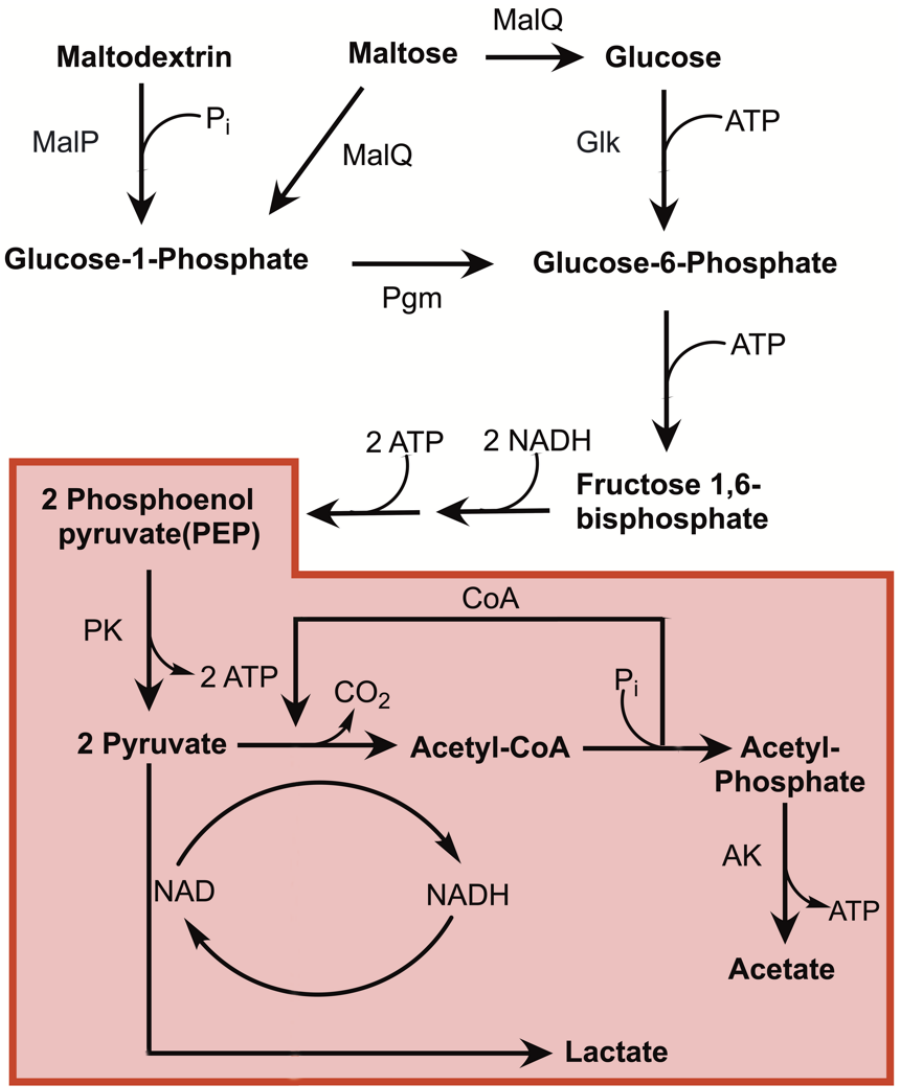
1.3 Energy generation

Energy generation methods in CFPS can be classed into two categories: (i) Indirect ATP regeneration systems using glycolytic metabolites of pyruvate, glucose and glucose derivatives, and (ii) direct ATP regeneration using a single enzyme (Creatine phosphate; CK, PEP; PK and AcP;AcK) (Lian et al., 2014). Substrates in direct ATP regeneration are expensive and are reliant on compounds with a high-energy phosphate bond, such as phospho-enolpyruvate (PEP), creatine phosphate (CP), or acetyl phosphate (AcP) (Ryabova et al., 1995; Wang and Percival, 2009). These high energy molecules are added to the reaction to regenerate ATP in a simple substrate level phosphorylation reaction, producing high concentrations of P_i in a short period of time (Kim and Swartz, 2000).

Indirect ATP regeneration systems have been developed over the years to reduce cost, extend the longevity of CFPS reactions, and increase productivity of the system (Whittaker, 2013; Banks et al., 2022). Activation of the metabolic pathways in glycolysis leads to efficient ATP regeneration and protein synthesis (Foshag et al., 2018). This process is facilitated by the utilization of catabolic metabolites and intermediates found within glycolysis (Calhoun and Swartz, 2005). Phosphorylation, a crucial step in glycolysis, recycles P_i and enables cyclic regeneration of ATP (Dopp et al., 2019). Currently, indirect ATP regeneration methods utilize the PANOx system, where higher glycolytic metabolites energize indirect regeneration methods (Jewett and Swartz, 2004a). The metabolic intermediate produced during glycolysis is transformed into phospho-enolpyruvate (PEP), which undergoes dephosphorylation to form ATP and is subsequently channeled into the PANOx system (Caschera and Noireaux, 2014). In order to improve energy

generation, combined energy systems utilizing both creatine phosphate (CP) and glucose have been used as energy sources.

The rate of ATP generation with indirect methods is relatively low during the initial stage of incubation when glucose is used as an energy source, as ATP is consumed for priming the glycolytic pathway (Kim et al., 2007). Using a combined energy system, phosphate is released quickly from CP and is used to drive the glycolytic pathway for the priming of glucose catabolism, which resulted in sustained ATP supply without accumulation of P_i resulting in higher protein yield.



PANOx System

Figure 1.1) The indirect regeneration of ATP in CFPS involves multistep reactions catalyzed by enzymes and substrates. Maltodextrin phosphorylase (MalP) mediates the indirect phosphorylation of ATP by bypassing the direct ATP phosphorylation by glucokinase (Glk). This strategy conserves ATP usage and reduces the overall P_i levels in the CFPS reaction. The PANOx system (Red) incorporates glycolytic metabolites, eliminating the requirement for the addition of expensive PEP and enabling the regeneration of ATP. The abbreviations for MalQ and Pgm are 4-alpha-glucanotransferase and phosphoglucomutase respectively.

An indirect ATP regeneration system is established using a combination of maltodextrin (MD) and hexametaphosphate (HMP) in conjunction with the native polyphosphate kinase. This system functions through the phosphorylation of the long glucose chains present in maltodextrin, utilizing inorganic phosphate (P_i) as the source of energy (as depicted in Figure 1.1). This method eliminates the need for the ATP-dependent activation of glucose and instead converts maltodextrin into glucose-1-phosphate, thereby facilitating the recycling of P_i from the reaction and enabling efficient ATP regeneration (Caschera and Noireaux, 2015). Although indirect glycolytic energy regeneration methods have managed to significantly extend the CFPS reaction time, efficient protein production by CFPS is still limited to a short period.

Apparatus-based developments have further improved the P_i removal from the reaction mixture (Schoborg et al., 2014), with some setups increasing the protein synthesis rates up to 40 hours (Spirin et al., 1988). The most common CFPS formats are batch, continuous exchange, and continuous flow, with each system providing different advantages (Katzen et al., 2005, Zemella et al., 2015) (Figure 1.2). Conventional batch systems provide the user with a high throughput method that can be automated, easily scaled and is relatively convenient to set up (Dondapati et al., 2020). However, batch formats tend to suffer from low productivity due to short sustained reaction times in CFPS, which can be in part mitigated with the use of an indirect P_i regeneration system using glucose derivatives, such as maltodextrin/hexametaphosphate rather than direct regeneration methods (Jewett and Swartz, 2004(b)). Batch systems are particularly amenable to these forms of ATP regeneration as only small amounts of co-factors need to be supplied, reducing the cost of CFPS (Sitaraman et al., 2004). More productive systems for CFPS are based

on continuous buffer exchange, which is still amenable to automation (Aoki et al., 2009). In a continuous exchange system, a regenerated cellulose membrane separates the ‘inner’ and ‘outer’ buffers. The CFPS reaction occurs within a permeable 8 - 10 kDa dialysis tubing which retains all macromolecules necessary for translation, while smaller molecules can freely diffuse between the inner and outer buffers. This was first demonstrated to generate 1.4 mg/ml of chloramphenicol acetyltransferase (CAT) over 14 hours (Kim and Choi, 1996). The continuous exchange system was developed further by using a condensed cell extract to increase the reaction yield to 6 mg/ml of CAT (Kigawa et al., 1999).

Continuous flow cell free (CFCF) format is similar to continuous buffer exchange and supplies reactants and removes inhibitory byproducts, which increases reaction duration and provides higher protein yields (Nishimura et al., 1995). The energy rich feed solution is continuously pumped, either through peristaltic or HPLC pumps into the reaction chamber, while the expressed protein and other byproducts are extracted through an ultrafiltration membrane to be collected by the user (Shirokov et al., 2002). This format has been applied with a variety of cell extracts, for example, wheat-germ extract which was able to produce a preparative amount of biologically active interleukin-6 (IL-6) from a CFCF reaction. Interestingly, the CFCF system exhibited sustained activity over a period of 52 hours, with a constant production of IL-6 (Volyanik et al., 1993). The CFCF system was able to generate 80 μg of interleukin-6 (IL-6) with a protein purity of 80%.

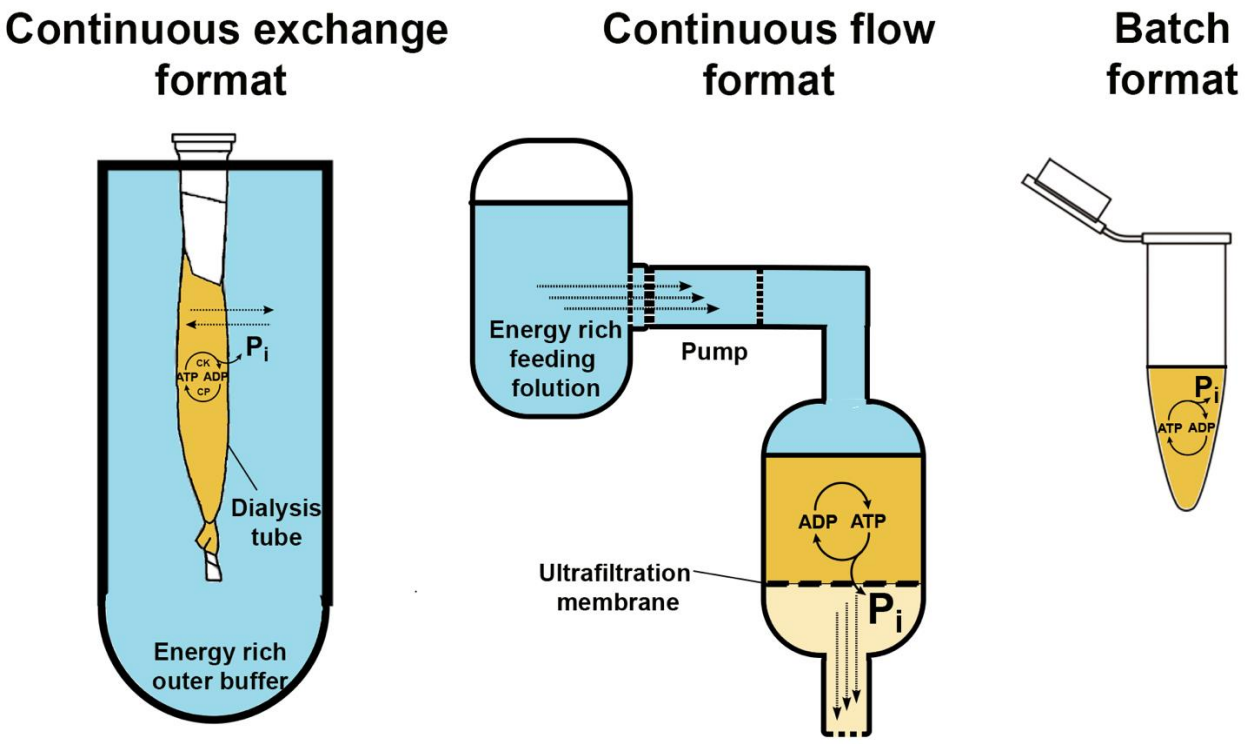


Figure 1.2) A graphical representation of three common systems utilized in CFPS reactions: batch, continuous exchange, and continuous flow. In the continuous exchange system, a dialysis tube containing the reaction mixture is immersed in a reaction buffer-like solution, facilitating removal of inhibitory small molecules from the reaction compartment. In the continuous flow system, a feeding solution is continuously pumped into the reaction chamber where products and inhibitory small molecules are separated via an ultrafiltration membrane. In the batch system, the lysate is mixed with reaction buffer without separating inhibitory byproducts.

1.4 Lysate preparation

CFPS lysates have been prepared using various organisms, with *E. coli* lysates being the most commonly employed, owing to their ability to grow to high OD_{600} in standard baffled flasks or bioreactors, utilizing cost-effective media (Ganjave et al., 2022). The integration of phage promoters with the initial S30 system has enabled the effective synthesis of desired proteins, thereby propelling the platform's ongoing development and broad utilization. (Nevin and Pratt, 1991). Subsequently, the platform has undergone continual evolution, contributing to its broad applicability across various scientific domains. Among the CFPS platforms, *E. coli* lysate has

been identified as the most productive and has been prepared using various techniques (Gregorio et al., 2019).

Despite being readily accessible due to its commercial availability, *E. coli* extracts are costly, prone to degradation during freeze-thaw cycles, and require special shipping conditions (Dopp et al., 2019). In the laboratory setting cell lysis can be readily completed using self-lysing (i.e. XJb (DE3)*) cell lines or by using mechanical processes such as a bead beating, homogenization or sonication (Shrestha et al., 2012). These procedures for making cell lysate can be laborious and often requires expensive reagents. New developments have therefore emphasized reducing the cost of lysate preparation by optimizing high-speed centrifugation, dialysis, and pre- incubation steps (Kwon and Jewett, 2015; Katsura et al., 2017).

Attempts have been made to optimize the ribosome content of lysates to enhance their productivity through the utilization of runoff reactions. A recent study concluded that performing a runoff reaction leads to an improvement in the productivity of cell extracts due to the increase in the availability of rRNA for protein translation (Kim et al., 2020).

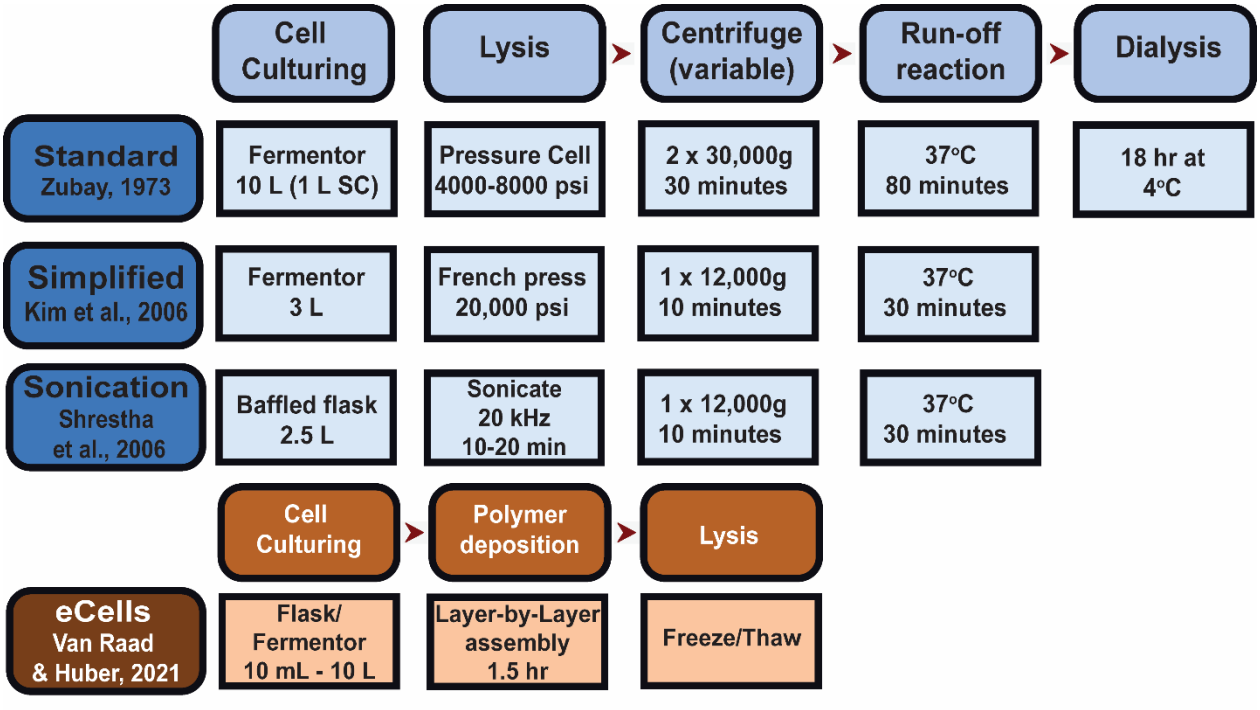


Figure 1.3) Evolution of lysate preparation methodologies over time is depicted in the figure. Progressive streamlining of the process and improvements in lysis techniques and culture scale have resulted in more efficient and time-saving lysate production, with the preparation time now reduced to less than 24 hours from the initiation of culturing.

The use of *V. natriegens*, a halophilic gram-negative marine bacterium, has been proposed as an alternative for the development of CFPS systems. The preparation of lysate from *V. natriegens* results in a significant increase in the ribosome content in the cell extract, as compared to *E. coli*. This is because *V. natriegens* possesses a ribosome content that is 30 - 60 % higher than *E. coli*, which enhances the potential for efficient protein translation in CFPS systems. (Xu et al., 2022). The ribosome content in the *V. natriegens* extract produced by Failmezger and co-workers was quantified to be four times higher compared to their similarly prepared *E. coli* extract (Failmezger et al., 2018). This is due to the fast growth rate of *V. natriegens*, with a cell division time of less than 10 minutes, enabling the acquisition of high cell densities required for extract preparation at a faster rate compared to *E. coli* (Hoffart et al., 2017).

V. natriegens extracts demonstrate limited compatibility with indirect methods of CFPS, as evidenced by a marked decrease in protein synthesis rates compared to coupled transcription-translation (CP/CK) based CFPS. As a result, the utilization of *V. natriegens* extracts is restricted to direct methods of CFPS, such as CP/CK systems (Failmezger et al., 2018). *V. natriegens* exhibits remarkable versatility in its lysate preparation, as harvesting at the stationary phase (OD_{600} 7.5) results in an increase in protein synthesis efficiency compared to exponential growth. This can be attributed to the fact that the ribosome content in *V. natriegens* remains elevated during stationary phase, while in standard *E. coli* extracts, ribosome content decreases in this phase. The yield of extract obtained from stationary phase *V. natriegens* lysate is significantly higher, with values ranging from 8-12 mL/L, compared to 1-3 mL/L obtained from *E. coli* lysate (des Soye et al., 2018).

Leishmania tarentolae, a unicellular eukaryotic flagellate, has emerged as a promising organism for CFPS. It possesses several favorable attributes for CFPS, including the capacity for large-scale fermentation, genetic modifiability, and the provision of a translational active lysate that

can be utilized to express a broad array of eukaryotic proteins. This system employs a coupled lysate with a species-independent translational sequence (SITS) that has the potential to circumvent the early initiation steps and drive the formation of the initiation complex for translation (Mureev et al., 2008). The *L. tarentolae* CFPS system demonstrated the functional expression of mammalian, *L. tarentolae*, and *Plasmodium falciparum* Rab GTPases. The system has consistently shown higher protein quality compared to *E. coli*-based CFPS, with an average yield improvement of five-fold. Furthermore, a majority of the expressed GTPases were found to be in their natively folded state, which is indicative of proper protein folding and function (Kovtun et al., 2010).

Extract preparation has evolved and moved towards lower centrifugation speeds, which has been shown to not compromise protein synthesis yields while making CFPS lysates accessible to labs without specialized equipment. Supernatant 12,000g, or S12, extract preparation is the use of lower centrifugal speed. The introduction of S12 extract preparation reduced the cost of protein production to 1 $ (USD) per milligram of protein by removing the costly high-speed centrifugation and dialysis steps of the S30 extract, with 1.2 mg/ml of chloramphenicol acetyltransferase (CAT) being generated (Kim et al., 2006). Further work by Kim and co-workers enhanced this by generating 1.8 mg/ml of chloramphenicol acetyltransferase using a S12 batch system (Kim et al., 2008). The implementation of sonication in the S12 preparation demonstrated a more efficient method to optimize the preparation of the extract for CFPS. (Shrestha et al., 2012). Sonication for S12 lysates was found to result in productivity comparable to lysates produced through high-pressure homogenization. Research has shown that productive *E. coli* lysates can be generated through sonication of the cell mass for a duration of 10-20 minutes at a frequency of 20 kHz. The yield of the resulting lysate is comparable to that of the traditional S30 preparation method.

Ezore and co-workers presented a method for producing a more active insect cell extract (ICE) that by avoiding harsh cell lysis conditions and utilizing a freeze-thaw cycle which allowed the system to achieve higher productivity of protein synthesis (Ezore et al., 2006). The study demonstrated a 400% increase in the yield of ICE CFPS compared to the productivity of the ICE obtained from mechanical homogenization. Despite the improvements in cell extract production,

the process remains labor-intensive, making it a significant cost to laboratories. The activity of cell lysate can also vary significantly between batches and change with storage, leading to decreased performance or even inactivity (Miguez et al., 2019).

1.5 CFPS platforms and PTMs

E. coli platforms are characterized by high productivity and relatively low cost compared to eukaryotic extracts, but they have limitations in terms of producing proteins with post-translational modifications (PTMs) (Smolskaya et al., 2020). Recent progress has been made in N-linked (asparagine-linked) glycosylation in *E. coli* through the transfer of eukaryotic and bacterial glycosylation systems to laboratory strains (Feldman et al., 2005). Specifically, the discovery of a bacterial glycosylation pathway from *Campylobacter jejuni*, has enabled the successful introduction of N-linked glycosylation to commonly used *E. coli* strains. (Szymanski et al., 1999). The *C. jejuni* N-linked glycosylation machinery is compatible with *E. coli* extracts due to the absence of endogenous glycosylation machinery, thus allowing for the functional transfer of *C. jejuni* glycosylation machinery into *E. coli* systems. The first CFPS system that incorporated the *C. jejuni* glycosylation locus into *E. coli* using purified components was able to successfully express glycosylated proteins *in vitro*. (Guarino and Delisa, 2012). By utilizing the *C. jejuni* glycosylation machinery in *E. coli* CFPS systems, researchers have opened new possibilities for the engineering and production of diverse recombinant glycan structures in proteins (Wacker et al., 2002).

Prenylation is the addition of an aliphatic moiety to a protein at a cysteine in a CAAX site (A is an aliphatic amino acid and X can be any amino acid) and typically occurs in eukaryotes (Jiang et al., 2018). This PTM acts in aiding fast and dynamic protein shuttling between the cytosol and membrane (Saini et al., 2009). Previously prenylation of proteins was only possible using eukaryotic extracts, such as insect cell extract (ICE), through its preparation and preservation of farnesyltransferase and geranylgeranyltransferase-I, II and II (FTase, GGTase I, GGTase II and GGTase III) (Suzuki et al., 2007). Enriching S30 lysates with prenylation machinery such as FTase and GGTase I from *Rattus norvegicus* has allowed efficient prenylation of protein targets. It has been found that prenyltransferase heterodimers of greater than > 70 kDa express well in *E. coli* and provide soluble, active protein. Using an enriched CFPS system, protein targets with a

CAAX motif attached to the C-terminus are synthesized *in situ*, prenylated and characterized using a fluorescent derivative of farnesylpyrophosphate or geranylpyrophosphate (Kai et al., 2022). The addition of these hydrophobic moieties to a protein target rendered them insoluble. Using detergents during *in situ* production did not overcome the solubility issue, with only one detergent, Brij-58, providing increasing yield of roughly 25% of soluble, geranylgeranylated protein approximately 25% over the control. The addition of nanodiscs proved more successful, with a linear increase of soluble protein with the addition of higher concentrations of nanodisc (Kai et al., 2022). The formation of a reversible membrane-binding switch in prenylated proteins was then characterized using confocal microscopy-based assays.

The use of eukaryotic cell lysates in CFPS provides a means for incorporating complex post-translational modifications (PTMs) into proteins, although yields of expressed proteins tend to be low. The insect cell extract (ICE) method has proven to be versatile in its ability to generate functional proteins for therapeutic applications, including membrane proteins. Single-chain variable fragments (scFv), mimic molecules for antibodies, can also be produced in CFPS systems through the utilization of the open chemical environment. By manipulating the redox potential in the ICE system, functional scFv fragments can be produced with a yield of 20 µg/mL. This was achieved through the addition of reduced/oxidized glutathione to the reaction mixture, which was conducted under reducing conditions (Stech et al., 2012). In a related study, ICE combined with exogenously added Sf21 microsomes has been demonstrated to enable the insertion of functional membrane proteins with PTMs, specifically N-linked glycosylation (Ezure et al., 2014). The successful characterization of three membrane proteins, namely the K+ channel KvAP, the pro-tumor necrosis factor pro-TNF, and the TA-protein synaptobrevin II, was achieved by their insertion into Sf21 microsomes. While yields were not reported, all of these proteins were glycosylated (Ezure et al., 2014).

Protein phosphorylation plays critical roles in the homeostasis of cellular processes, i.e cell cycle, growth, apoptosis, and signal transduction pathways. Many enzymes and receptors are activated/deactivated by phosphorylation and dephosphorylation events (Ardito et al., 2017). The utilization of wheat germ extract (WGE) and insect cell extract (ICE) has been shown to be an effective approach to produce eukaryotic proteins. It was previously assumed that the utilization

of WGE would result in a limited number of phosphorylated PTMs. However, recent findings have indicated the endogenous phosphorylation of NS5A protein from Hepatitis C Virus (HCV) when using WGE (Badillo et al., 2017). A more recent study by Böckmann and co-workers has shown that using WGE to produce the large envelope protein from duck hepatitis B has confirmed heterogenous phosphorylation primarily at S118, with minor phosphorylation products being at T79, T89, S117 generating 0.11 mg in a 0.25 ml batch CFPS reaction (David et al., 2019).

1.6 Incorporation of non-canonical amino acids

CFPS allows for the controlled synthesis of proteins outside of living cells, making it possible to manipulate and expand the genetic code beyond the 20 canonical amino acids. Residue-specific incorporation refers to the complete replacement of a natural amino acid with its non-canonical amino acid (ncAA) analog in a protein (Johnson et al., 2010). This method exploits the ability of endogenous amino acyl tRNA synthetases (aaRS) to recognize and promiscuously charge amino acids analogues, which relies on limiting the availability of the corresponding canonical amino acid in the reaction buffer to promote incorporation of the desired ncAA. However, natural aaRS enzymes typically exhibit a preference for their cognate canonical amino acids over ncAA analogs, which presents a challenge for this approach (Worst et al., 2016).

In contrast, site-specific incorporation involves introducing ncAA at a specific location(s) in a protein sequence. To incorporate an ncAA site specifically, an orthogonal aaRS and suppressor tRNA ($tRNA_{sup}$ or $tRNA_{CUA}$) is used to suppress an amber (TAG) codon and incorporate an ncAA (Wang et al., 2000). Site specific incorporation of an ncAA requires the supplementation of ncAAs in the CFPS reaction buffer. It also requires purification and supplementation of $tRNA_{sup}$ and aaRS that can selectively recognize and charge the desired ncAA onto the $tRNA_{sup}$ (Chin, 2017). By selecting and incorporating different ncAAs researchers can create proteins with new functions, such as increased stability, novel catalytic activity, or improved fluorescence.

There have been numerous successful examples of ncAA incorporation using CFPS. For instance, there has been development of a portable toolkit that integrates CFPS and protein immobilization in one pot, simplifying the process of protein-based biomaterial synthesis. They

created a set of plasmids that fuses the N-terminus of superfolder green fluorescent protein (sfGFP) with different peptide tags, enabling the immobilization of the protein on tailored materials using different immobilization chemistries. The incorporation of azide ncAA derivatives into the nascent protein further enables selective immobilization using copper-free click reactions (Benítez-Mateos et al., 2018).

1.7 Applications of CFPS for NMR spectroscopy

The utilization of fluorinated derivatives for site-specific replacement of amino acid residues in proteins has been reported multiple times using CFPS. A recent study showed the global replacement of leucine, valine, and alanine residues with their fluorinated derivatives in the B1 domain of Streptococcal protein G (GB1) using CFPS (Maleckis et al., 2022). The resulting fluorine chemical shifts were dispersed, providing a useful means for the detection of protein structure and dynamics using NMR spectroscopy.

CFPS is widely utilized in NMR spectroscopy due to its cost-effectiveness and ability to label proteins on a residue-specific basis. An optimized system based on S30 extract was used, focusing on maximizing protein yield while minimizing the usage of ^{15}N-labeled amino acids. The optimized system facilitated the selective incorporation of ^{15}N-amino acids into 19 specific residues of human cyclophilin A, yielding clear ^{15}N-HSQC spectra without the use of auxotrophic strains (Ozawa et al., 2004).

CFPS is also practical in terms of the expression of proteins with differential isotope labelling schemes. Previous work has focused on chemical and enzymatic synthesis of AAs with isotope labelling patterns that have isolated spin systems devoid of homonuclear coupling for the assignment of large protein complexes (Sprangers and Kay., 2007). An example is the synthesis of Stereo Array Isotope Labeling (SAIL) amino acids which provide spectral simplification and sensitivity enhancement of proteins produced, as the non-exchangeable side chains have reduced spectral overlap (Kainosho and Güntert., 2009). SAIL amino acids have been applied to aromatic residues with alternate ^{13}C/^{2}H labelling schemes, labelling the δ, ε, and ζ of the aromatic ring with deuteration (Takeda et al., 2010). It has also been applied to methyl groups of AAs, with stereospecific isotope labelling with both ^{13}C and ^{2}H (Gans et al., 2010). All 20 canonical amino

acids have been chemically synthesized into SAIL amino acids, although expense and difficulty in synthesis is a limiting fact in the widespread use of these amino acids in CFPS.

Selective isotope labeling is important for studying protein behaviour and applications, but metabolic conversions can cause isotope scrambling and complicate interpretation of results. A strategy proposed for achieving cleaner selective isotope labeling in CFPS reactions involves reducing *E. coli* S30 extracts with $NaBH_4$, a strong reducing agent, to inactivate all pyridoxal-phosphate (PLP) dependent enzymes. In this way, conversion between different amino acids is supressed, resulting in cleaner isotope labeling, which stabilizes hydrogens bound to α-carbons against exchange with H_2O, minimizing the loss of α-deuterons during protein production from perdeuterated amino acids in a H_2O solution. It was shown that this method allows for production highly perdeuterated proteins without the need to back-exchange labile deuterons for protons, making it a more economical alternative to other methods for achieving isotope labeling. The methodology opens the possibility of producing proteins with exceptionally clean selective ^{15}N-labelling and uniformly $^{2}H/^{15}N/^{13}C$-labelled proteins in H_2O with high levels of perdeuteration while all amides are fully protonated (Su et al., 2011).

One limitation of traditional NMR techniques is that they can be difficult to interpret when applied to large protein complexes, as there are often many overlapping signals that make it difficult to distinguish individual chemical shifts (Huang and Kalodimos, 2013). To overcome this challenge, an elegant method was developed that involves segmental isotope labeling and selective amino acid-type labeling of individual domains using CFPS. The approach involves the production of individual domains of the protein, with selective labeling of specific amino acids. The labeled domains are then ligated together to form the complete protein, which is subsequently analysed with NMR spectroscopy. This was illustrated using two different proteins complexes: two C-terminal RNA-recognition motifs of the human polypyrimidine tract-binding protein, and two highly homologous helix-turn-helix domains of the human glutamyl-prolyl-tRNA synthetase. By selectively labeling specific amino acids within each domain, the authors were able to obtain clearer NMR spectra and more easily assign specific residues to different domains. This approach is particularly useful for multidomain proteins, where segmental labeling can reduce spectral overlap and make chemical shifts easily interpretable. This method presents

an innovative and effective approach for studying the structure of large, complex proteins using NMR spectroscopy (Michel et al., 2013).

The use of differentially labeled isotopologue precursors have been utilized for *in vivo* expression, with subsequent generation of protein with isolated spin systems and reduced spectral overlap. The culture is grown with a precursor that is funnelled into amino acid biosynthesis that produces protein with a specific isotope labelling pattern with alternate ^{13}C, ^{15}N or ^{2}H labels being incorporated (Schörghuber et al., 2018). The cost of supplementing these precursors is a major drawback, as large amounts of precursor are used for *in vivo* expression. Moreover, the use of these precursors is not compatible with current CFPS systems as the full complement of metabolic enzymes are not present within the S30 extract (Linser et al., 2014).

1.8 Limitations of CFPS

Extract based CFPS systems are a favored approach for conducting protein expression *in vitro*, mainly due to its high yield and simplicity. The lysate preparation for CFPS, however, is a laborious and time-consuming procedure that entails several steps, such as cell disruption, washing, and centrifugation (Carlson et al., 2012). These steps require specialized equipment, expensive reagents, and a significant amount of time and effort (Apponyi et al., 2008). Moreover, the application of CFPS in the high-throughput selection of mutants can be limited, as researchers need to employ microfluidics, which is hindered by time and complex apparatus. This impedes the practicality of screening millions of mutants, a necessity in many cases (Xiao et al., 2015).

Pyrrolysyl-tRNA synthetases (PylRS) derived from *Methanosarcina mazei* and *Methanosarcina barkeri* exhibit the capability to amino acylate a diverse range of ncAAs. Subsequently, significant engineering efforts have been focused on enhancing the productivity of PylRSs charging ncAA onto $tRNA_{sup}$ with bulky and non-bulky ncAA structures. While PylRS is commonly utilized in vivo, the preparation of PylRS for CFPS presents challenges. Specifically, PylRS derived from *M. mazei* and *M. barkeri* are susceptible to being inactive and are difficult to purify in their active form, which is primarily attributed to the low solubility of the N-terminal domain (Seki et al., 2020). Consequently, alternative strategies are needed to stabilize and purify PylRS to ensure their bioactivity during CFPS (Chemla et al., 2015).

The process of preparing lysates disrupts numerous metabolic pathways by rendering crucial enzymes inactive or eliminating them from the lysate. Consequently, applications reliant on these enzymes become infeasible. One such limitation is the difficulty in achieving isotope labeling through precursor incorporation, including ^{13}C glucose or other specifically labeled metabolites, due to the absence of a complete bacterial metabolism in CFPS lysate which is necessary for precursor biosynthesis into amino acids. (Linser et al., 2014). Although Stereo-Array Isotope Labeling (SAIL) is available, it is associated with high costs, which limits its practical application (Kainosho and Güntert, 2009). Moreover, no selective methyl-labeling methods using precursors for CFPS have been developed, which restricts their application in the structural and dynamic studies of large proteins. Specifically, achieving $^{13}C^{1}H$ methyl-labeled Ile, Leu, and Val (ILV) in a deuterated background, which is a widely used technique for investigating protein structure and dynamics, has not been feasible through CFPS, apart from using SAIL amino acids. Therefore, alternative approaches are needed to address these limitations and expand the utility of CFPS in isotope labeling studies.

1.8.1 Thesis Chapters

Chapter 2 introduces a novel method for efficiently producing encapsulated bacterial lysate or eCells that can be used for CFPS for recombinant protein expression. This chapter demonstrates the use of genetic code expansion through site-specific non-canonical amino acid incorporation (ncAA) in eCells, using amino acyl-tRNA synthetases like PylRS in CFPS to incorporate photocaged-cysteine (PCC) and N^{ε}-(tert-butoxycarbonyl)-L-lysine (boc-lysine) into peptidylprolyl cis−trans isomerase B (PpiB). This highlights the potential of eCells for use in a wide range of applications, including the efficient and cost-effective production of non-living encapsulated cells for the synthesis of proteins and fluorescent activated cell sorting (FACS). In addition, the chapter explores the cost-effectiveness and the simplification of lysate preparation using eCells and demonstrates that they are competitive in relation to traditional CFPS.

Chapter 3 introduces new approaches to selectively label the methyl groups of valine, leucine, isoleucine, and alanine residues in proteins using eCells. This chapter also highlights the strategic selection of precursor compounds and the omission of certain amino acids from the reaction

mixture to produce proteins with residue and atom-specific isotope labeling patterns, which can significantly reduce the cost of isotope labeling proteins for NMR spectroscopy. The chapter also explores the stereospecific labeling of pro-*S* signals of leucine and valine residues using the precursor 2-^{13}C-methylacetolactate. Through this chapter, it is shown that eCells have potential for reducing the amount of precursor compounds needed to produce proteins and can be further amenable to the 'unlabeling' of amino acid residues by their addition to the CFPS reaction mixture.

Chapter 4 of the thesis focuses on the development of new approaches for isotope labeling of aromatic amino acid residues for NMR spectroscopy. The chapter explains how the proteins peptidylprolyl cis−trans isomerase B and ubiquitin were selectively labeled in the ε and δ positions of Phe/Tyr using 3-^{13}C-pyruvate and isotopologues of glucose. 4-^{13}C erythrose and 3-^{13}C-pyruvate were also used to elevate the labelling efficiency in the δ positions for a combinatorial labelling strategy in-vitro.

Chapter 5 highlights the ability of eCells to modify the expression environment for protein folding and change the redox environment. It also shows that eCells are effective for bioremediation and biosensing of contaminants. The chapter introduces the functionalization of BL21(DE3)* derivatives for homogenous site-specific phosphorylation using a , dual incorporation of L-phospho-O-serine and toxic 3-fluoro-alanine, and the refolding of proteins using human protein disulfide isomerase (PDI) in the eCell system. Finally, it describes the successful expression of two previously difficult-to-express proteins, neuritin and active human light-chain enterokinase, in eCells.

1.9 References

1. Amthor, J. S. (2000). The McCree-de Wit-Penning de Vries-Thornley respiration paradigms: 30 years later. Ann. Bot., 86, 1-20.
2. Aoki, M., Matsuda, T., Tomo, Y., Miyata, Y., Inoue, M., Kigawa, T., & Yokoyama, S. (2009). Automated system for high-throughput protein production using the dialysis cell-free method. Protein Expr. Purif., 68, 128-136.

3. Ardito, F., Giuliani, M., Perrone, D., Troiano, G., & Lo Muzio, L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). Int. J. Mol. Med., 40, 271-280.
4. Badillo, A., Receveur-Brechot, V., Sarrazin, S., Cantrelle, F. X., Delolme, F., Fogeron, M. L., Molle, J., Montserret, R., Bockmann, A., Bartenschlager, R., Lohmann, V., Lippens, G., Ricard-Blum, S., Hanoulle, X., & Penin, F. (2017). Overall structural model of NS5A protein from hepatitis C virus and modulation by mutations conferring resistance of virus replication to cyclosporin A. Biochem., 56, 3029-3048.
5. Banks, A. M., Whitfield, C. J., Brown, S. R., Fulton, D. A., Goodchild, S. A., Grant, C., Love, J., Lendrem, D. W., Fieldsend, J. E., & Howard, T. P. (2022). Key reaction components affect the kinetics and performance robustness of cell-free protein synthesis reactions. Comput. Struct. Biotech. J., 20, 218-229.
6. Benítez-Mateos, A. I., Llarena, I., Sánchez-Iglesias, A. & López-Gallego, F. (2018. Expanding One-Pot Cell-Free Protein Synthesis and Immobilization for On-Demand Manufacturing of Biomaterials. ACS. Synth. Biol., 16, 875-884.
7. Calhoun, K. A., & Swartz, J. R. (2005). Energizing cell-free protein synthesis with glucose metabolism. Biotechnol. Bioeng., 90, 606-613.
8. Calhoun, K. A., & Swartz, J. R. (2007). Energy systems for ATP Regeneration in cell-free protein synthesis reactions. Methods Mol. Biol., 375, 3-17.
9. Cao, M., Sun, Q., Zhang, X., Ma, Y., & Wang, J. (2021). Detection and differentiation of respiratory syncytial virus subgroups A and B with colorimetric toehold switch sensors in a paper-based cell-free system. Biosens. Bioelectron., 182, 113-119.
10. Carlson, E. D., Gan, R., Hodgman, C. E., & Jewett, M. C. (2012). Cell-free protein synthesis: applications come of age. Biotechnol. Adv., 30, 1185-1194.
11. Caschera, F., & Noireaux, V. (2014). Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. Biochimie, 99, 162-168.
12. Caschera, F., & Noireaux, V. (2015). A cost-effective polyphosphate-based metabolism fuels an all *E. coli* cell-free expression system. Metab. Eng., 27, 29-37.
13. Chen, C., Cui, X., Beausang, J. F., Zhang, H., Farrell, I., Cooperman, B. S., & Goldman, Y. E. (2016). Elongation factor G initiates translocation through a power stroke. Proc. Natl. Acad. Sci. U.S.A., 113, 7515-7520.

14. Chin, J. W. (2017). Expanding and reprogramming the genetic code. Nature., 83, 53-60.
15. David, G., Fogeron, M. L., Montserret, R., Lecoq, L., Page, A., Delolme, F., Nassal, M., & Böckmann, A. (2019). Phosphorylation and alternative translation on wheat germ cell-free protein synthesis of the DHBV large envelope protein. Front. Mol. Biosci., 6, 1-7.
16. Des Soye, B. J., Davidson, S. R., Weinstock, M. T., Gibson, D. G., & Jewett, M. C. (2018). Establishing a high-yielding cell-free protein synthesis platform derived from Vibrio natriegens. ACS Synth. Biol., 7, 2245-2255.
17. Dondapati, S. K., Stech, M., Zemella, A., & Kubick, S. (2020). Cell-free protein synthesis: a promising option for future drug development. BioDrugs., 34, 327-348.
18. Dopp, B. J. L., Tamiev, D. D., & Reuel, N. F. (2019). Cell-free supplement mixtures: Elucidating the history and biochemical utility of additives used to support *in vitro* protein synthesis in *E. coli* extract. Biotechnol. Adv., 37, 246-258.
19. Dopp, J. L., Jo, Y. R., & Reuel, N. F. (2019). Methods to reduce variability in *E. coli*-based cell-free protein expression experiments. Synth. Syst. Biotechnol., 4, 204-211.
20. Ezore, T., Suzuki, T., Higashide, S., Shintani, E., Endo, K., Kobayashi, S. I., Shikata, M., Ito, M., Tanimizu, K., & Nishimura, O. (2006). Cell-free protein synthesis system prepared from insect cells by freeze-thawing. Biotechnol. Prog., 22, 1570-1577.
21. Ezure, T., Nanatani, K., Sato, Y., Suzuki, S., Aizawa, K., Souma, S., Ito, M., Hohsaka, T., von Heijine, G., Utsumi, T., Abe, K., Ando, E., & Uozumi, N. (2014). A cell-free translocation system using extracts of cultured insect cells to yield functional membrane proteins. PLOS ONE., 9 1-17.
22. Failmezger, J., Scholz, S., Blombach, B., & Siemann-Herzberg, M. (2018). Cell-free protein synthesis from fast-growing Vibrio natriegens. Front. Microbiol., 9, 1-10.
23. Feldman, M. F., Wacker, M., Hernandez, M., Hitchen, P. G., Marolda, C. L., Kowarik, M., Morris, H. R., Dell, A., Valvano, M. A., & Aebi, M. (2005). Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. Proc. Natl. Acad. Sci. U.S.A., 102, 3016-3021.
24. Foshag, D., Henrich, E., Hiller, E., Schäfer, M., Kerger, C., Burger-Kentischer, A., Diaz-Moreno, I., García-Mauriño, S. M., Dötsch, V., Rupp, S., & Bernhard, F. (2018). The *E. coli* S30 lysate proteome: A prototype for cell-free protein production. New Biotechnol., 40, 245-260.

25. Ganjave, S. D., Dodia, H., Sunder, A. V., Madhu, S., & Wangikar, P. P. (2022). High cell density cultivation of *E. coli* in shake flasks for the production of recombinant proteins. Biotechnol. Rep., 33, 2215-2232.
26. Gans, P., Hamelin, O., Sounier, R., Ayala, I., Durá, M. A., Amero, C. D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M. J., & Boisbouvier, J. (2010). Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. Angew. Chem. Int. Ed., 49, 1958-1962.
27. Gregorio, N. E., Levine, M. Z., & Oza, J. P. (2019). A user's guide to cell-free protein synthesis. Methods Protoc., 2, 1-34.
28. Guarino, C., & Delisa, M. P. (2012). A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. Glycobiology, 22, 596-601.
29. Hershewe, J., Kightlinger, W., & Jewett, M. C. (2020). Cell-free systems for accelerating glycoprotein expression and biomanufacturing. J. Ind. Microbiol. Biotechnol., 47, 977-991.
30. Hoffart, E., Grenz, S., Lange, J., Nitschel, R., Müller, F., Schwentner, A., Feith, A., Lenfers-Lücker, M., Takors, R., & Blombach, B. (2017). High substrate uptake rates empower Vibrio natriegens as production host for industrial biotechnology. Appl. Environ. Microbiol., 83, 1-10.
31. Huang, C. & Kalodimos, C. G. (2017). Structures of Large Protein Complexes Determined by Nuclear Magnetic Resonance Spectroscopy. Annu. Rev. Biophys., 46, 317-336.
32. Jewett, M. C., & Swartz, J. R. (2004a). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol. Bioeng., 86, 19-26.
33. Jewett, M. C., & Swartz, J. R. (2004b). Substrate replenishment extends protein synthesis with an *in vitro* translation system designed to mimic the cytoplasm. Biotechnol. Bioeng., 87, 465-471.
34. Jiang, H., Zhang, X., Chen, X., Aramsangtienchai, P., Tong, Z., & Lin, H. (2018). Protein lipidation: occurrence, mechanisms, biological functions, and enabling technologies. Chem. Rev., 118, 919-988.

35. Johnson, J. A., Lu, Y. Y., Van Deventer, J. A. & Tirrell, D. A. (2010). Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. Curr. Opin. Chem. Biol., 14, 774-80.
36. Kainosho, M., & Güntert, P. (2009). SAIL – stereo-array isotope labeling. Quart. Rev. Biophys., 42, 247-300.
37. Katsura, K., Matsuda, T., Tomabechi, Y., Yonemochi, M., Hanada, K., Ohsawa, N., Sakamoto, K., Takemoto, C., & Shirouzu, M. (2017). A reproducible and scalable procedure for preparing bacterial extracts for cell-free protein synthesis. J. Biochem., 162, 357-369.
38. Katzen, F., Chang, G., & Kudlicki, W. (2005). The past, present and future of cell-free protein synthesis. Trends Biotechnol., 23, 150-156.
39. Kigawa, T., Yabuki, T., Yoshida, Y., Tsutsui, M., Ito, Y., Shibata, T., & Yokoyama, S. (1999). Cell-free production and stable-isotope labeling of milligram quantities of proteins. FEBS Lett., 442, 15-19.
40. Kim, D. M., & Choi, C. Y. (1996). A semicontinuous prokaryotic coupled transcription/translation system using a dialysis membrane. Biotechnol. Prog., 12, 645-649.
41. Kim, D. M. & Swartz, J. R. (2000) Prolonging cell-free protein synthesis by selective reagent additions. Biotechnol. Prog., 16, 385-90.
42. Kim, H. C., Kim, T. W., & Kim, D. M. (2011). Prolonged production of proteins in a cell-free protein synthesis system using polymeric carbohydrates as an energy source. Process Biochem., 46, 1366-1369.
43. Kim, T. W., Keum, J. W., Oh, I. S., Choi, C. Y., Park, C. G., & Kim, D. M. (2006 (a)). Simple procedures for the construction of a robust and cost-effective cell-free protein synthesis system. J. Biotechnol., 126, 554-561.
44. Kim, T. W., Kim, D. M., & Choi, C. Y. (2006). Rapid production of milligram quantities of proteins in a batch cell-free protein synthesis system. J. Biotechnol., 123, 373-380.
45. Kim, T. W., Oh, I. S., Keum, J. W., Kwon, Y. C., Byun, J. Y., Lee, K. H., Choi, C. Y., & Kim, D. M. (2007). Prolonged cell-free protein synthesis using dual energy sources: Combined use of creatine phosphate and glucose for the efficient supply of ATP and retarded accumulation of phosphate. Biotechnol. Bioeng., 97, 1510-1515.

46. Kim, T. W., Kim, H. C., Oh, I. S., & Kim, D. M. (2008). A highly efficient and economical cell-free protein synthesis system using the S12 extract of Escherichia coli. Biotechnol. Bioprocess Eng., 13, 464-469.
47. Kim, J., Copeland, C. E., Seki, K., Vögeli, B., & Kwon, Y. C. (2020). Tuning the cell-free protein synthesis system for biomanufacturing of monomeric human filaggrin. Front. Bioeng. Biotechnol., 8, 1-14.
48. Ozawa, K., Headlam, M. J., Schaeffer, P. M., Henderson, B. R., Dixon, N. E., & Otting, G. (2004). Optimization of an Escherichia coli system for cell-free synthesis of selectively ^{15}N-labeled proteins for rapid analysis by NMR spectroscopy. Eur. J. Biochem., 271, 4084-4093.
49. Kovtun, O., Mureev, S., Johnston, W., & Alexandrov, K. (2010). Towards the Construction of Expressed Proteomes Using a Leishmania tarentolae Based Cell-Free Expression System. PLOS ONE, 5, e14388.
50. Kwon, Y. C., & Jewett, M. C. (2015). High-throughput preparation methods of crude extract for robust cell-free protein synthesis. Sci. Rep., 5, 1-8.
51. Kai, L., Sonal, Heermann, T., & Schwille, P. (2022). One-pot synthesis of prenylated proteins utilizing *E. coli* cell-free expression. bioRxiv, 1-25.
52. Li, J., Zhang, C., Huang, P., Kuru, E., Forster-Benson, E. T. C., Li, T. & Church G. M. (2017). Dissecting limiting factors of the Protein synthesis Using Recombinant Elements (PURE) system. Transl. (Austin)., 5, 1-14.
53. Lian, Q., Cao, H., & Wang, F. (2014). The cost-efficiency realization in the Escherichia coli-based cell-free protein synthesis systems. Biotechnol. Appl. Biochem., 61, 2351-2367.
54. Liu, W., Chen, C., Kavaliauskas, D., Knudsen, C. R., Goldman, Y. E., & Cooperman, B. S. (2015). EF-Tu dynamics during pre-translocation complex formation: EF-Tu GDP exits the ribosome via two different pathways. Nucleic Acids Res., 43, 9519-9528.
55. Liu, W. Q., Wu, C., Jewett, M. C., & Li, J. (2020). Cell-free protein synthesis enables one-pot cascade biotransformation in an aqueous-organic biphasic system. Biotechnol. Bioeng., 117, 4001-4008.

56. Linser, R., Gelev, V., Hagn, F., Arthanari, H., Hyberts, S. G. & Wagner, G. (2014). Selective methyl labeling of eukaryotic membrane proteins using cell-free expression. J Am. Chem. So.. 136, 11308-11310.
57. Maleckis, A., Abdelkader, E. H., Herath, I. D., & Otting, G. (2022). Synthesis of fluorinated leucines, valines and alanines for use in protein NMR. Org. Biomol. Chem., 20(7), 2424-2432.
58. Miguez, A. M., McNerney, M. P., & Styczynski, M. P. (2019). Metabolic profiling of Escherichia coli-based cell-free expression systems for process optimization. Ind. Eng. Chem. Res., 58, 22472-22482.
59. Michel, E., Skrisovska, L., Wüthrich, K. & Allain, F. H. (2013) Amino acid-selective segmental isotope labeling of multidomain proteins for structural biology. Chembiochem., 14, 457-466.
60. Mureev, S., Kovtun, O., Nguyen, T. T. U., & Alexandrov, K. (2009). Species-independent translational leaders facilitate cell-free expression. Nature Biotechnology, 27, 747-752.
61. Nevin, D. E., & Pratt, J. M. (1991). A coupled *in vitro* transcription-translation system for the exclusive synthesis of polypeptides expressed from the T7 promoter. FEBS Lett., 285, 259-263.
62. Nirenberg, M. W., & Matthaei, J. H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. Proc. Natl. Acad. Sci. USA, 47, 1588-1602.
63. Nishimura, N., Kitaoka, Y., & Niwano, M. (1995). Enhancement of protein synthesis in continuous-flow, cell-free system by improvement of membrane permeation. J. Ferment. Bioeng., 80, 403-405.
64. Oh, I. S., Kim, D. M., Kim, T. W., Park, C. G., & Choi, C. Y. (2006). Providing an oxidizing environment for the cell-free expression of disulfide-containing proteins by overwhelming the reducing activity of Escherichia coli S30 extract. Biotechnol. Prog., 22, 1225-1228.
65. Pedersen, A., Hellberg, K., Enberg, J., & Karlsson, B. G. (2011). Rational improvement of cell-free protein synthesis. New. Biotechnol., 28, 218-224.
66. Perez, J. G., Stark, J. C., & Jewett, M. C. (2016). Cell-free synthetic biology: engineering beyond the cell. Cold Spring Harb. Perspect. Biol., 8, 1-25.

67. Rogerson, D. T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S. M., Huguenin-Dezot, N., Muqit, M. M., Fry, A. M., Bayliss, R., & Chin J. W. (2015). Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nat. Chem. Biol., 11, 496-503.
68. Ryabova, L. A., Vinokurov, L. M., Shekhovtsova, E. A., Alakhov, Y. B., & Spirin, A. S. (1995). Acetyl phosphate as an energy source for bacterial cell-free translation systems. Anal. Biochem., 226, 184-186.
69. Saini, D. K., Chisari, M., & Gautam, N. (2009). Shuttling and translocation of heterotrimeric G proteins and Ras. Trends Pharmacol. Sci., 30, 278-286.
70. Schoborg, J. A., Hodgman, C. E., Anderson, M. J., & Jewett, M. C. (2014). Substrate replenishment and byproduct removal improve yeast cell-free protein synthesis. Biotechnol. J., 9, 630-640.
71. Schörghuber, J., Geist, Leonhard, Gerald Platzer, Feichtinger, Michael, Marilena Bisaccia, Scheibelberger, Lukas, Weber, F., Konrat, R., & Lichtenecker, R. J. (2018). Late metabolic precursors for selective aromatic residue labeling. J. Biomol. NMR., 71, 129-140.
72. Seki, E., Yanagisawa, T., Kuratani, M., Sakamoto, K. & Yokoyama, S. (2020). Fully productive cell-free genetic code expansion by structure-based engineering of Methanomethylophilus alvus pyrrolysyl-tRNA synthetase. ACS. Synth. Biol., 9, 718–732
73. Shirokov, V. A., Simonenko, P. N., Biryukov, S. V., & Spirin, A. S. (2002). Continuous-flow and continuous-exchange cell-free translation systems and reactors. Cell-Free Transl. Syst., 91-107.
74. Shrestha, P., Holland, T. M., & Bundy, B. C. (2012). Streamlined extract preparation for Escherichia coli-based cell-free protein synthesis by sonication or bead vortex mixing. BioTechniques., 53, 163-174.
75. Sitaraman, K., Esposito, D., Klarmann, G., le Grice, S. F., Hartley, J. L., & Chatterjee, D. K. (2004). A novel cell-free protein synthesis system. J. Biotechnology., 110, 257-263.
76. Smolskaya, S., Logashina, Y. A., & Andreev, Y. A. (2020). Escherichia coli extract-based cell-free expression system as an alternative for difficult-to-obtain protein biosynthesis. Int. J. Mol. Sci., 21, 1-21

77. Sprangers, R., & Kay, L. E. (2007). Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature., 445, 618-622.
78. Spirin, A. S., Baranov, V. I., Ryabova, L. A., Ovodov, S. Y., & Alakhov, Y. B. (1988). A continuous cell-free translation system capable of producing polypeptides in high yield. Science., 242, 1162-1164.
79. Stech, M., Merk, H., Schenk, J. A., Stöcklein, W. F., Wüstenhagen, D. A., Micheel, B., Duschl, C., Bier, F. F., & Kubick, S. (2012). Production of functional antibody fragments in a vesicle-based eukaryotic cell-free translation system. J. Biotechnology., 164, 220-231.
80. Su, X. C., Loh, C. T., Qi, R. & Otting, G. (2011). Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enymes for selective ^{15}N-labelling and production of perdeuterated proteins in H_2O. J. Biomol. NMR., 50, 35-42.
81. Suzuki, T., Ito, M., Ezure, T., Shikata, M., Ando, E., Utsumi, T., Tsunasawa, S., & Nishimura, O. (2007). Protein prenylation in an insect cell-free protein synthesis system and identification of products by mass spectrometry. Proteomics., 7, 1942-1950.
82. Szymanski, C. M., Ruijin, Y., Ewing, C. P., Trust, T. J., & Guerry, P. (1999). Evidence for a system of general protein glycosylation in Campylobacter jejuni. Mol. Microbiology., 32, 1022-1030.
83. Takeda, M., Ono, A. M., Terauchi, T., & Kainosho, M. (2010). Application of SAIL phenylalanine and tyrosine with alternative isotope-labeling patterns for protein structure determination. J. Biomol. NMR., 46, 45-49.
84. Tsuboi, T., Takeo, S., Iriko, H., Jin, L., Tsuchimochi, M., Matsuda, S., Han, E. T., Otsuki, H., Kaneko, O., Sattabongkot, J., Udomsangpetch, R., Sawasaki, T., Torii, M., & Endo, Y. (2008). Wheat germ cell-free system-based production of malaria proteins for discovery of novel vaccine candidates. Infect. Immun., 76, 1702-1708.
85. Van Raad, D., & Huber, T. (2021). *In vitro* protein synthesis in semipermeable artificial cells. ACS Synth. Biol., 10, 1237-1244.
86. Venkat, S., Chen, H., Gan, Q., & Fan, C. (2019). The application of cell-free protein synthesis in genetic code expansion for post-translational modifications. Front. Pharmacol., 10, 1-6.

87. Volyanik, E. V., Dalley, A., McKay, I. A., Leigh, I., Williams, N. S., & Bustin, S. A. (1993). Synthesis of preparative amounts of biologically active Interleukin-6 using a continuous-flow cell-free translation system. Anal. Biochem., 214, 289-294.
88. Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W., & Aebi, M. (2002). N-linked glycosylation in Campylobacter jejuni and its functional transfer into *E. coli*. Science., 298, 1790-1793.
89. Wang, L., Magliery, T. J., Liu, D. R. & Schultz, P. G. (2000). A new functional suppressor tRNA/aminoacyl− tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. J. Am. Chem. Soc., 122, 5010–5011.
90. Wang, Y. & Percival, Y. H. P. (2009). Cell-free protein synthesis energized by slowly-metabolized maltodextrin. BMC Biotechnol., 9, 1-8.
91. Worst, E. G., Exner, M. P., De Simone, A., Schenkelberger, M., Noireaux, V., Budisa, N. & Ott, A. (2016). Residue-specific Incorporation of Noncanonical Amino Acids into Model Proteins Using an Escherichia coli Cell-free Transcription-translation System. J. Vis. Exp., 114, 1-14.
92. Whittaker, J. W. (2013) Cell-free protein synthesis: The state of the art. Biotechnol. Lett., 35, 143-152.
93. Xiao, H., Bao, Z. & Zhao, H. (2015) High Throughput Screening and Selection Methods for Directed Enzyme Evolution. Ind. Eng. Chem. Res. 54, 4011-4020
94. Xu, J., Yang, S. & Yang, L. (2022) Vibrio natriegens as a host for rapid biotechnology. Trends Biotechnol., 40, 381-384.
95. Yang, J., Cui, Y., Cao, Z., Ma, S. & Lu, Y. (2021) Strategy exploration for developing robust lyophilized cell-free systems. Biotechnol. Notes., 16, 44-50.
96. Zemella, A., Thoring, L., Hoffmeister, C. & Kubick, S. (2015) Cell-Free protein Synthesis: pros and cons of prokaryotic and eukaryotic systems. ChemBioChem., 16, 2420-2431.
97. Zhang, M. S., Brunner, S. F., Huguenin-Dezot, N., Liang, A. D., Schmied, W. H., Rogerson, D. T. & Chin, J. W. (2017) Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. Nat. Methods., 14, 729-736.

CHAPTER TWO

In vitro protein synthesis in semi-permeable artificial cells

2.1 Abstract

A novel cell free protein synthesis (CFPS) system utilizing Layer-by-Layer (LbL) polymer assembly was developed to reduce the operational cost of conventional CFPS. This yielded an encapsulated cell system, dubbed 'eCells', that successfully performs *in vitro* CFPS and allows cost effective incorporation of non-canonical amino acids into proteins. The use of eCells in CFPS circumvents the need for traditional cell lysate preparation and purification of amino acyl-tRNA synthetases (aaRS) while still retaining the small scale of an *in vitro* reaction. eCells were found to be 55% as productive as standard dialysis CFPS at 13% of the cost. The reaction was shown to be scalable over a large range of reaction volume and the crowding environment in eCells confers a stabilizing effect on marginally stable proteins, such as the pyrrolysl tRNA synthetase (PylRS), providing a means for their application in *in vitro* protein expression. Photocaged-cysteine (PCC) and N^{ε}-(tert-butoxycarbonyl)-L-lysine (boc-lysine) was incorporated into peptidyl-prolyl cis-trans isomerase B (PpiB) using small amounts of ncAA with adequate yield of protein. Fluorescent activated cell sorting (FACS) was used to demonstrate the partition of the lysate within the eCells in contrast to standard one pot cell lysate-based methods.

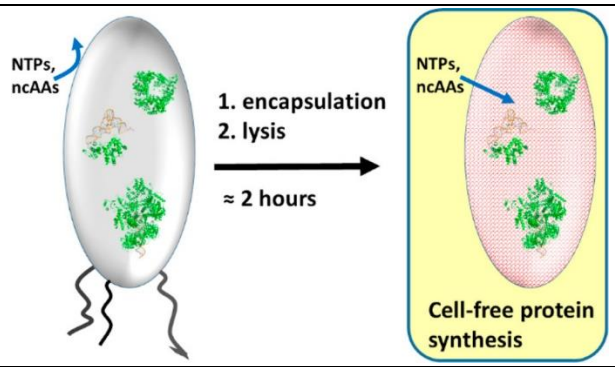


Figure 2.1) The following diagram illustrates the process of in vitro protein synthesis using capsules (eCells) created through layer-by-layer (LbL) assembly. The diagram outlines the system's permeability and the production timeframe from the point of inoculation.

2.2 Introduction

Cell-Free protein synthesis (CFPS) is a versatile tool for functional and structural biology, allowing rapid and inexpensive production of proteins. The method was first introduced more than 50 years ago by Nirenberg and Matthaei, enabling them to decipher the genetic code (Nirenberg and Matthaei., 1961). Since then, remarkable improvements in reaction stability and protein yields have been achieved by carefully optimizing conditions under which the reactions are performed (Ayoubi-Joshaghani et al., 2020). Combined with today's ease to commercially obtain kits for *in vitro* protein synthesis, CFPS is increasingly used in many applications of basic research, biotechnology, and health sciences.

Traditional CFPS is characterised by the use of a cell lysate that holds all the necessary components for energy production, transcription, translation and protein folding (Jewett & Swartz., 2004; Rosenblum and Cooperman., 2014). The lysate is a processed extract of the cytosol from a plurality of mono-clonal organisms and is used in conjunction with a buffer containing small molecule components essential to sustaining protein production. These components encompass amino acids, nucleotides, an energy production system, cofactors, and various salts. Lastly, a suitable DNA template is provided, which encodes the target protein for production.

To achieve even higher reaction controllability of CFPS, Ueda et al fully reconstituted an *in vitro* protein synthesis system from only ribosomes and essential elements of the *E. coli* translation system (Shimizu et al., 2014). In this PURE - "protein synthesis using recombinant elements" – system, all macromolecules required for protein synthesis (except for ribosomes) are individually expressed, purified and then titered into a reaction mixture. While producing the PURE system is labor intensive and thus costly, this disadvantage is greatly offset by the abstinence of proteolytic and nucleolytic enzymes in the reaction mixture which can reduce protein product yield.

Two commonly used *in vitro* protein synthesis systems are batch and dialysis CFPS. Batch CFPS is characterized by using small molecules coupled directly with transcription/translation components in a lysate as a one-pot mixture to generate protein. Dialysis mode CFPS is the same reaction mixture held within a regenerated cellulose membrane in an external equivalent of CFPS buffer. This regenerated cellulose membrane

separates the ‘inner’ and ‘outer’ buffer. The CFPS reaction occurs within the permeable 10-15 kDA dialysis tubing which retains all macromolecules necessary for translation, while smaller molecules can diffuse between the inner and outer buffer. The speed and ease with which CFPS can be performed and the full control over its reaction conditions make it possible to use CFPS where *in vivo* protein expression cannot be applied. Toxins, biopharmaceutical proteins and membrane proteins have been shown to be successfully expressed in CFPS and proteins were produced *in-situ* for high-throughput, arrayed analyses (Hovijitra et al., 2009; Cai et al., 2015; Shelby et al., 2019; Zhuang et al., 2020).

Although CFPS has become a viable alternative to traditional *in vivo* expression systems, it requires development. Major limiting factors are intense energy costs, expensive reagents, and lower reaction yields (Lian et al., 2014). Furthermore, standard *E. coli* lysates are 20 times more dilute than the cytosolic compartment of the cell (Rosenblum and Cooperman., 2014). This decreases rates of elongation, translation, and protein expression, and further affects stability of enzymes in the lysate. Compartmentalized systems where macromolecular translation/transcription components are confined have been developed to emulate *in vivo* systems.

Biomimetic confinements of lysate have been produced through various compartmentalization techniques. These methods include oil in water droplets, coacervates, lipid vesicles, liposomes and polymersomes (Shinde and Nagarsenker., 2009; Nishimura et al., 2012; Torre et al., 2014; Rampioni et al., 2019; Ayoubi-Joshaghani et al., 2020). These compartmentalized systems can be utilized to house genetic information and biomolecules related to transcription and translation. Liposomes are a common compartmentalization technique which can emulate biochemical reactions in a controlled *in vitro* environment. They are small vesicles comprising of a lipid bilayer and are often used as a cell model. For example, in 2009 Swartz et al synthesized functional aquaporin Z in synthetic liposomes with CFPS. They reported a 40 times increase in yield of functional aquaporin and into the membrane correctly inserted aquaporins over standard *in vivo* methodology. The aquaporin loaded liposomes were able to act as a simple cell model and allowed permeability measurements for a variety of biologically interesting molecules (Hovijitra et al., 2009)
One of the unique characteristics of living cells is their shielding boundary that separates their strongly homeostatic inside from the varying extracellular environment. This membrane-based compartmentalization of molecules is required to perform basic and complex functions

(Lai et al., 2015). Gram-negative bacteria have a highly negatively charged outer membrane due to the anionic moieties of lipopolysaccharide (Malanovic., 2016). We hypothesized that the physical properties of *E. coli*'s outer membrane allow cell surface modification to turn bacterial cells into semi-permeable containers for *in vitro* protein synthesis.

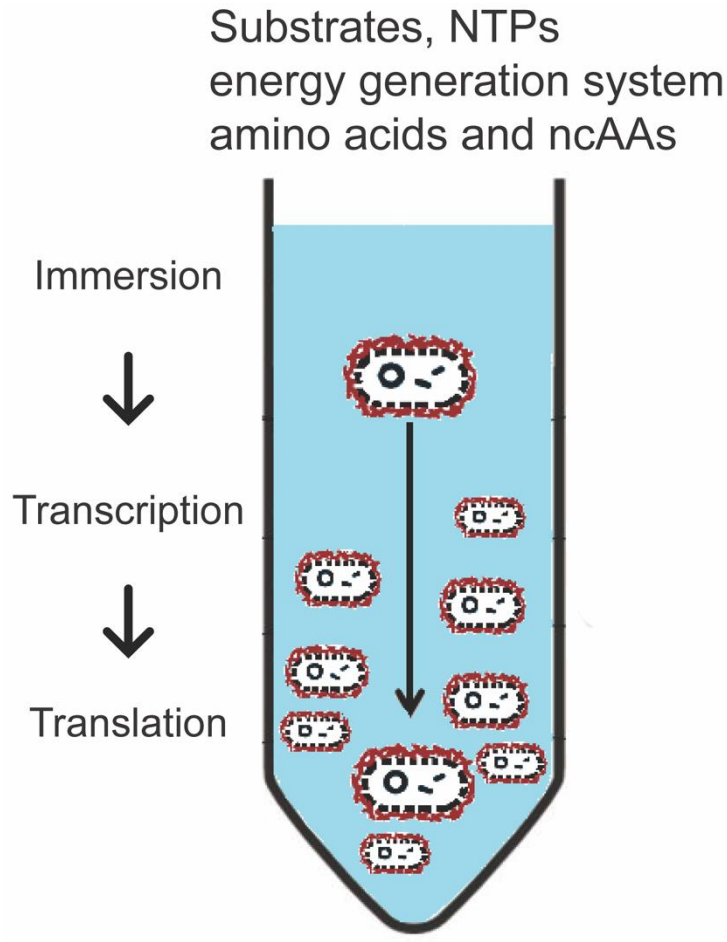


Figure 2.2) A schematic of *in vitro* protein synthesis using capsules produced through LbL assembly. All substrates required for protein synthesis are present in the CFPS buffer and can freely diffuse into the capsule**.**

Bio-polyelectrolytes have been used extensively to modify the surface of cells, especially *E. coli*. Layer-by-layer polymer assembly (LbL) relies on the adsorption of long chains of polyelectrolytes on a surface, to various levels of thickness, and operates through electrostatic interactions (Hillberg and Tabrizian., 2006). When a negatively charged surface is exposed to a cationic based polymer, the net charge of the surface changes due to adhesion and overcompensation by the polyion (Ariga et al., 2007). Iterations with polyelectrolytes of alternating net-charge can be repeated until a desired thickness of polymer is achieved (Tong

et al., 2012). Chitosan and alginate are commonly used bio-polymers for LbL deposition. Chitosan is a form of deacetylated chitin, which has a positively charged amino group in its monomer unit (Lee et al., 2013). Conversely, alginic acid, often derived from brown algae, consists of anionic monomer units (Lee et al., 2012). LbL assembly for cell surface modification has been applied as early as 2006 for bio-recognition (Hillberg and Tabrizian., 2006). Chitosan and alginate were found to be biocompatible and able to be adsorbed onto the cell surface without affecting cell viability. To confirm deposition of each layer of polyelectrolyte, zeta potential measurement was used, showing the charge alternation with successive deposition of cationic polymer and anionic polymer. It has been shown that visual confirmation of deposition of polyelectrolytes is possible. Imaging of these coated bacteria using TEM shows nanometer range deposition of polyelectrolytes, encapsulating them.

Cell encapsulation with chitosan and alginate was also employed for library screening of detergent stable mutants of G coupled protein receptors (Scott and Plückthun., 2013). Plückthun et al applied two layers of the polyelectrolytes onto the outer membrane of *E. coli* cells which expressed a library of G coupled protein receptor mutants linked to sfGFP to test retention (Yong and Scott., 2015). The native *E. coli* membranes and peptidoglycan layer were then dissolved using detergents, and microcapsules were selected using fluorescence activated cell sorting (FACS). The collected 1 % capsules with the highest fluorescence also contained highly detergent stable protein mutants.

Here we demonstrate that the concept of LbL encapsulation of bacterial cells can be extended to generate semi-permeable artificial cells with femto-liter volume which are apt for *in vitro* protein synthesis. By LbL deposition of alternate charged polyions onto the cell surface, rigid shells are formed that trap all components of the bacteria. High internal pressure, between 3-5 atmospheres, exists within *E. coli* cells and its structure is stabilized by a highly crosslinked peptidoglycan layer (Fischetti., 2008). Only a small number of these crosslinks need to be broken for the plasma membrane to be ruptured and cause hypotonic lysis. We use a hydrolytic enzyme to lyse the native cell wall of the bacteria, leaving a layered and semi-permeable capsule surrounding all constituents of the cell. We refer to these polyelectrolyte encapsulated cells with the cell wall lysed as eCells.

An important feature of these semi-permeable artificial cells is that the entire macromolecular repertoire of a living cell is present in approximately the same composition, concentration and volume, which presents an advantage over lysate based *in vitro* expression. Enzymes that are

associated with the membrane are generally lost through lysate preparation and as a result, metabolic pathways can be more fragmented in traditional CFPS.

The ancillary pathways which assist protein translation are hard to quantify in the S30 extract. In particular, the PURE system consists of reconstituted proteins directly related to protein transcription/translation, individually purified and used for protein expression (Shimizu et al., 2014). Within the PURE system there is a low level of protein expression indicative of a complex network of processes required for protein production. Encapsulated protein synthesis holds the entire repertoire of protein elements of an *in vivo* system and after lysis they are contained in a polyelectrolyte capsule. This expanded repertoire allows for the function and use of multi-step enzymatic pathways that require proteins associated with the membrane or other proteins lost during the preparation of the cell lysate. These capsules could provide added efficiency to *in vitro* systems that utilise membrane proteins in multi-step enzymatic cascades, either for energy-generation or for a specific by-product.

2.3 Materials and methods

2.31 Polyelectrolytes

Polyelectrolytes were purchased from Sigma; low molecular weight chitosan (448869, 50,000-190,000 Da) and sodium alginate.

2.32 CFPS protocol

The protocol for CFPS was adapted from Noireux and Caschera's hexametaphosphate(HMP)/ maltodextrin phosphate recycling system (Caschera and Noireux., 2015). The HMP/maltodextrin CFPS buffer contains 0.9 mM of UTP and CTP, 50 mM HEPES, 1.5 mM GTP, 1.5 mM ATP, 0.68 μM folinic acid, 0.64 mM cAMP, 1.7 mM DTT, 3.5 mM of amino acid mix, KGlu 60 mM, MgGlu 6 mM, 2% v/v PEG-8000, 5 mM CoA, 35 mM maltodextrin, 30 mM 3-PGA, 1.2 mM HMP and 10 mM NAD. Cyclic HMP was linearised at 100 ^{o}C for 5 minutes and then used after cooling.

CFPS using a creatine phosphate as energy source was used to express peptidyl-prolyl cis-trans isomerase B (PpiB) containing photocaged cysteine (PCC), as higher protein yields are obtained with this more efficient ATP regeneration system. The creatine phosphate CFPS buffer contains 0.9 mM of UTP and CTP, 50 mM HEPES, 1.5 mM GTP, 1.5 mM ATP, 0.68

μM folinic acid, 0.64 mM cAMP, 1.7 mM DTT, 3.5 mM of amino acid mix, KGlu 60 mM, MgGlu 6 mM, 2% v/v PEG-8000, 250 μg/ml of creatine kinase, 80 mM of creatine phosphate. Roche complete mini^{TM} PI inhibitor was added to CFPS buffer after being dissolved in 10 mL at 10% v/v. IPTG, and, where required, non-canonical amino acids were at 1 mM concentration when added to the final buffer. Photocaged cysteine (PCC) was an exception, and a higher amino acid concentration of 3.5 mM (or 33 mg in 30 mL) in creatine phosphate CFPS buffer was used.

2.33 Competent cell procedure and transformation

The transformation process involved the use of electrocompetent cells. These cells were initially grown in super optimal broth (SOB) media until they reached the desired growth phase of 0.7 OD_{600}. Subsequently, the cells were harvested by centrifugation and washed twice with ice-cold 10% glycerol solution. The washed cells were then resuspended in 3 mL of ice-cold glycerol solution and then snap frozen in 50 μl aliquots.

For the transformation step, a Bio-Rad MicroPulser™ was employed. The plasmid DNA, prepared according to standard protocols, was mixed with the resuspended cells. The mixture was transferred to a pre-chilled electroporation cuvette with 2 mM width, ensuring efficient electrical contact. The electroporation cuvette containing the mixture was then subjected to an electrical pulse using the BioRad MicroPulser™ device, allowing for the introduction of the plasmid into the cells through transient permeabilization of the cell membrane.

2.34 Encapsulation procedure

E. coli XJb(DE3)* cells were transformed by electroporation with the plasmid of interest prior to encapsulation and grown in LB medium at 37 ^{o}C in baffled flasks with shaking at 180 rpm. Endolysin production was induced at the time of inoculation with a final concentration of 3 mM arabinose. Cells were grown to OD_{600} = 0.6, harvested by centrifugation at 2,773 g and washed three times with 20 mL PBS-E buffer (137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4 and 1 mM EDTA, pH 7.4). For coating with chitosan, 1 g of the cells were resuspended in 20 mL of 0.25 mg/mL chitosan in PBS-E with vigorous shaking for 20 minutes. 1 g of the cell pellet was washed with 20 mL PBS-E pH 6.0 three times to remove excess chitosan and then resuspended in 20 mL of 0.25 mg/mL alginate PBS-E solution and

subjected to vigorous shaking for 20 minutes. XJb cells were then washed 3 times with PBS-E pH 6.0, resuspended in PBS-E pH 7.4 and directly stored at -80^{o} C in 1 ml aliquots in PBS-E buffer. For larger expression, the entire capsule pellet was frozen and then used. To thaw both, a water bath at RT was used with the tube of cell pellet or aliquot being immersed until thawed.

2.34 eCells with CFPS protocol

Frozen aliquots of encapsulated cells were thawed as described in 2.33 and resuspended in CFPS buffer. The buffer volumes ranged between 0.5 mL and 30 mL depending on the weight of cell pellet used (20 mg – 1 g). Lysis of the cell wall occurs spontaneously during thawing. CFPS for each experiment was conducted at 37^{o}C in a shaker at 180 rpm. For fluorescence, readings were taken every 1.5 hours for 12 hours and a final endpoint reading was taken at 24 hours later using a SpectraMax fluorescence reader, with SoftMax pro as the measurement software. Protein expression was conducted with 500 µl of CFPS buffer using the HMP and maltodextrin method of energy generation with the addition of protease inhibitor as described in 2.32. Each of the fluorescence readings were of 100 µl of CFPS immersed encapsulated cells.

2.35 Protein purification and Plasmid constructs

The gene of monomeric neon-green Green fluorescent protein (nGFP) in a plasmid with pCloDF13 origin of replication and amber stop codon interrupted at position 13 was used to test the incorporation of non-canonical amino acids. 'Wild type' (WT) neon-green monomeric GFP was expressed from a pET vector with pBR322 origin of replication. A derivative *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase (TyrRS) and $tRNA_{sup}$ pair was used for incorporation of para-acetyl-phenylalanine (PaF) (Wang et al., 2003). This aaRS/$tRNA_{sup}$ pair is under a GlnS promoter in a plasmid with pUC origin of replication. The same construct with Methanosarcina mazei pyrrolysl-tRNA synthetase/$tRNA_{sup}$ was used to incorporate N^{ε}-(*tert*-butoxycarbonyl)-L-lysine (boc-lysine) into nGFP, mCherry RFP and PpiB, at positions A13TAG, A13TAG and A147TAG respectively. For WT PpiB expression in eCells a plasmid with the pCloDF13 origin of replication, a T7 promoter, lac operator and lacI was used to tightly control expression and minimize unwanted background expression. For WT PpiB expression in dialysis mode a plasmid with the pCloDF13 origin of replication a T7 promoter

was used. Plasmids used with a pCloDF13 origin had spectinomycin resistance, synthetase plasmids with a pUC origin and a glnS promoter had kanamycin resistance. pET vectors used for WT GFP, had ampicillin resistance.

The gene for PpiB with an amber site at position 147, was induced for expression from an encapsulated pellet of 1 g of XJb(DE3)* from 500 mL of LB media. The encapsulation procedure was conducted as described in 2.34. 3 mL of CFPS was used to induce expression of the thawed eCells with 1 mM boc-lysine. For photocaged cysteine incorporation, cells were cultured in 1 L and then encapsulated as described in 2.34. The volume of CFPS buffer was 30 mL with 3.5 mM of photocaged cysteine.

For protein extraction, the eCell reaction was topped up to 30 mL with binding buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% v/v glycerol). The eCells were lightly sonicated for 2 minutes at 20 kHZ to disrupt the polymer bag. After a VX 22g VWR high speed refrigerated centrifuge with a R15A fixed angle rotor was used to clarify the homogenised eCells. The eCell mixture was in a 50 mL falcon tube and spun at 20,000 g for 15 minutes at 4^{o}C. For protein purification, 1 ml Ni-NTA His GraviTrap TALON columns (GE Healthcare) were used. The column was equilibrated with 10 mL of binding buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5% v/v glycerol). Then the clarified eCell sample was passed through the column. After 10 mL of wash buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 10 mM imidazole, 5% v/v glycerol) was run through the column. The protein was eluted with 5 mL of elution buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 300 mM imidazole 5% v/v glycerol). After purification, an approximate yield of 0.5 mg of boc-lysine incorporated at position 147 PPiB was obtained. For photocaged cysteine incorporated PPiB, the protein yield was 0.65 mg. Protein concentration was determined using the Thermo Scientific Nanodrop™ One/Onec Microvolume UV spectrophotometer, employing A280 and the extinction coefficient of 9960 $M^{-1}cm^{-1}$. The proteins were then analysed by mass spectrometry.

2.36 Mass Spectrometry analysis

High resolution electro spray ionization mass spectrometry was performed in positive mode on an Orbitrap Elite mass spectrometer and on Orbitrap Fusion™ Tribrid™ mass spectrometer coupled with an UltiMate 3000 UHPLC (Thermo Scientific, USA). Samples were injected into the mass analyzer via an Agilent ZORBAX SB-C3 Rapid Resolution HT

Threaded Column, using an acetonitrile gradient (10–85 %) and 0.1 % formic acid. Mass spectra were deconvoluted and processed using the Xcalibur software package.

2.37 Fluorescent activated cell sorting (FACS)

XJb(DE3)* cells were transformed with a plasmid containing a pCloDF13 origin of replication, *Methanosarcina mazei* PylRS/$tRNA_{sup}$ under a glnS promoter and a mRFP gene which was amber interrupted at position 13. The cells were grown at 37^{o}C to OD_{600} of 0.6 and then encapsulated. The *in vitro* expression was conducted with 3 mL of creatine phosphate based CFPS with and without 1 mM boc-lysine for 12 hours. 10^6 eCells were analysed using FACS on an Aria II high speed cell sorter. For the negative controls, the same procedure for encapsulation and expression was used and ncAA and/or the plasmid coding for the tRNA synthetase were omitted.

2.38 Yield comparison Dialysis CFPS and eCell CFPS procedure

S30 cell extract was prepared from a 20 L culture of *E. coli* strain 3071(RF1 removed) as detailed in Apponyi et al and yielded 100 mL of S30 extract (Apponyi et al., 2008). This extract was used for the subsequent dialysis CFPS. PCR was performed on the overexpression T7 PpiB plasmid to amplify the gene. The gene was then cyclised and used as 20% of the inner volume of the inner buffer of dialysis CFPS (Wu et al., 2007).
lac PpiB was transformed into XJb(DE3)* cells, grown in 1 L media and then encapsulated. This yielded 2 g of eCells. 6 x 300 mg of eCells were placed as aliquots for further expression. The eCells aliquots were immersed in 11 mL creatine phosphate CFPS buffer with 287 mg of Creatine phosphate and expressed at 37^{o}C overnight.

A w/v comparison of eCell CFPS and dialysis CFPS was conducted in 11 mL of creatine phosphate CFPS buffer. 300 mg of eCells was weighed out and then directly compared with 300 µl of S30 extract. Exactly 287 mg of creatine phosphate was used for both expressions amounting to 80 mM as a final concentration. The CFPS protocol for dialysis CFPS was conducted as detailed in Apponyi et al (Apponyi et al., 2008). The experiments were done in triplicate.

Expression yields of PpiB with boc-lysine incorporated at position A147 and WT PpiB were compared as follows. Two plasmids, one with A147 PpiB under a T7 promoter and one with

PylRS/$tRNA_{sup}$, were transformed into XJb(DE3)* cells and then grown in LB media. The cell mass was encapsulated at OD 0.6 and this yielded 2 g of eCells for expression. 300 mg of eCells was weighed out and then directly used for eCell expression in 11 mL reaction buffer. 287 mg of creatine phosphate was used for expression amounting to 80 mM as a final concentration.

2.4 Results

2.41 Diffusion of WT nGFP

An expression plasmid with WT mNeonGreenGFP as reporter gene was expressed in the XJb cell line and eCells prepared as described in the method section. Based on this 26.9 kDa monomeric form of GFP, the passive diffusion of macromolecules from the encapsulation can be monitored by fluorescence measurements. The two layers of polyelectrolyte were hypothesized to be able to retain macromolecules while being permeable to small molecules. Figure 2.3 shows the amount of nGFP remaining in encapsulated cells over the time course of 24 hours. The main diffusion of nGFP into the surrounding solution occurs over the course of six hours, and of the 100% of nGFP present at the beginning of the diffusion experiment, approximately 10% are present 24 hours later. The fact that substantial nGFP diffusion from the eCells is observed substantiates that the bacterial cell wall is lysed in the eCell preparation, as nGFP is not able to escape from living cells with intact membranes. The half-retention time of 26.9 kDa nGFP in the encapsulations is 3.46 hours modelled by a fluorescence change due to passive diffusion.

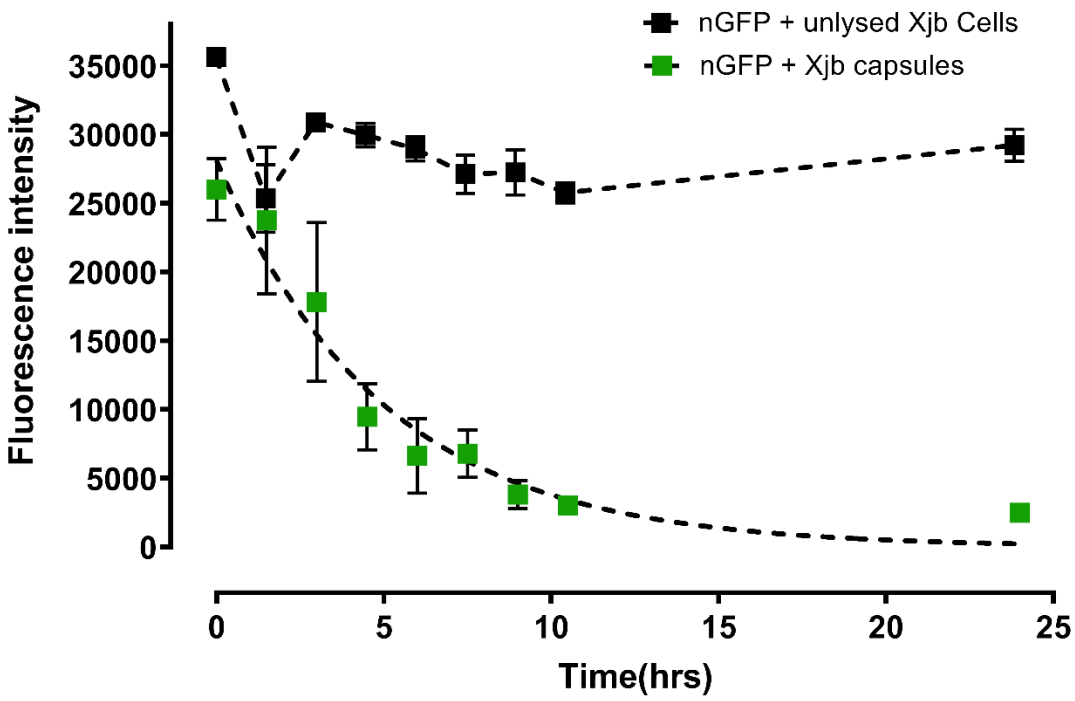


Figure 2.3) Induced, encapsulated and lysed XJb(DE3)* with WT nGFP in PBS-E buffer in comparison to live XJb Cells with nGFP expressed both done in triplicate. The relative diffusion of GFP from an encapsulated system over the course of 24 hrs, with the half-point of GFP diffusion at approximately 4 hours. The nGFP fluorescence of the live culture remains static throughout the experiment. Experiments were done in triplicate; error bars represent standard deviation.

Based on this observation we hypothesized that other larger biomolecules with varying physio-chemical properties are also likely held within the polyelectrolyte shell, and protein components for transcription and translation remain functionally active to sustain *in vitro* protein synthesis. To test this hypothesis, we performed cell-free protein synthesis using eCells from *E. coli*.

2.42 Encapsulated Cell-Free protein synthesis

Figure 2.4 shows the cell-free protein synthesis of nGFP in eCells with and without external supplement of nucleotides, amino acids and small molecules necessary for ATP production. As expected, eCells which did not have these small molecules externally supplied but contained 1 mM IPTG to induce transcription, showed no significant *in vitro* production of

nGFP. The lack of nGFP production indicates complete depolarization of the cells and thus no residual cell activity remains. In difference, eCells in cell-free protein synthesis buffer produce nGFP over 10 hours, validating that transcription and translation components are effectively retained in eCells and are active over the critical period where most protein is produced in conventional CFPS. Larger macro-molecules (nearly all proteins important for transcription and translation) will diffuse out of eCells much slower. In addition, highly charged macromolecules, such as tRNAs, will encounter a high kinetic barrier reducing the rate of diffusion.

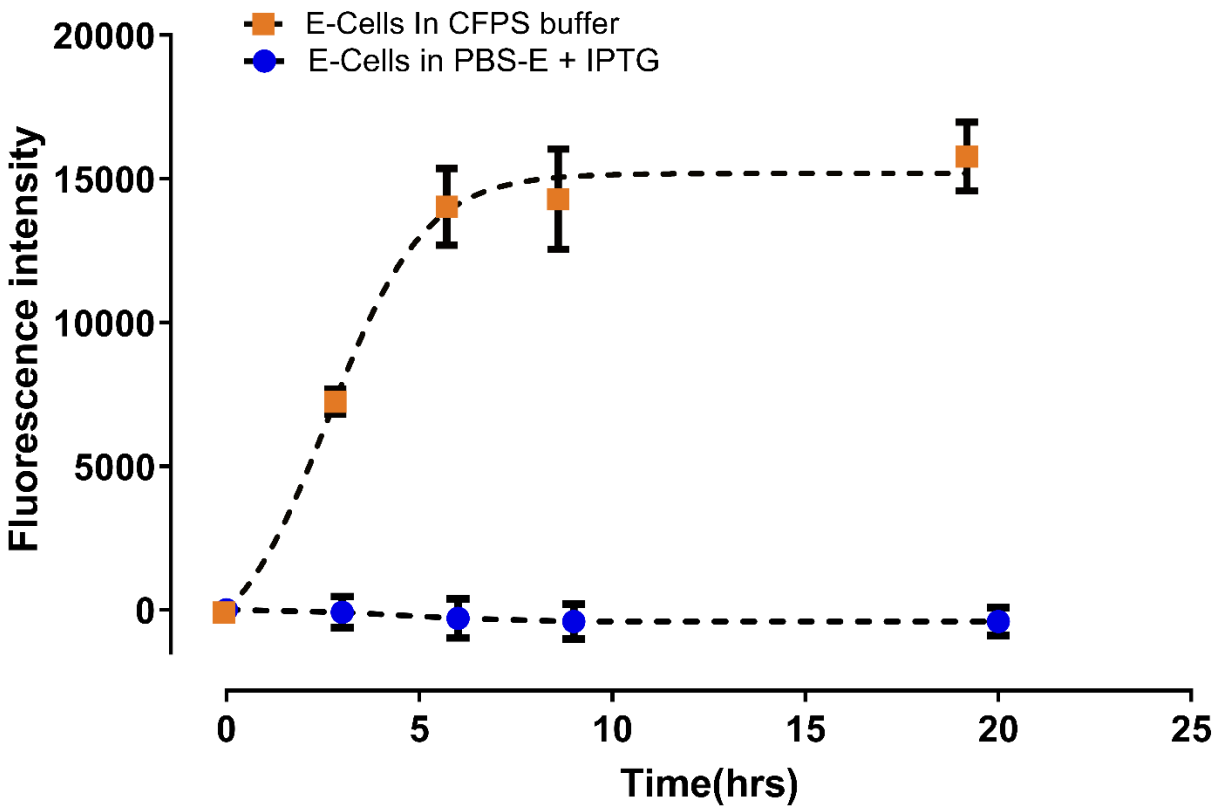


Figure 2.4) *In vitro* protein production of WT nGFP in encapsulated and lysed eCells using HMP/maltodextrin cell free protein synthesis mode. A substantial production of GFP in this method, exceeding that of the background control of PBS-E IPTG, indicating that *in vitro* protein synthesis occurred. All fluorescence expression and measurment was done in triplicate.

2.43 ncAA incorporation

Next, we illustrated that eCells also are apt to produce proteins containing non-canonical amino acids (ncAA) when the ncAA is externally supplied. Using a *M. jannaschii* suppressor tRNA and tyrosyl tRNA synthetase mutant pair to selectively encode the amber stop codon for p-acetyl-L-phenylalanine (PaF), the nGFP gene in a pCDF plasmid with amber interruption at amino acid Leu13 was expressed to produce a fluorescent reporter. Results shown in Figure 2.5 confirm the ability for ncAAs to be incorporated into an encapsulated CFPS system. As there was no tRNA synthetase or tRNA supplemented within the CFPS reaction, it highlights the utility of the capsules as the orthogonal tRNA/tRNA synthetase pair is expressed during cell preparation.

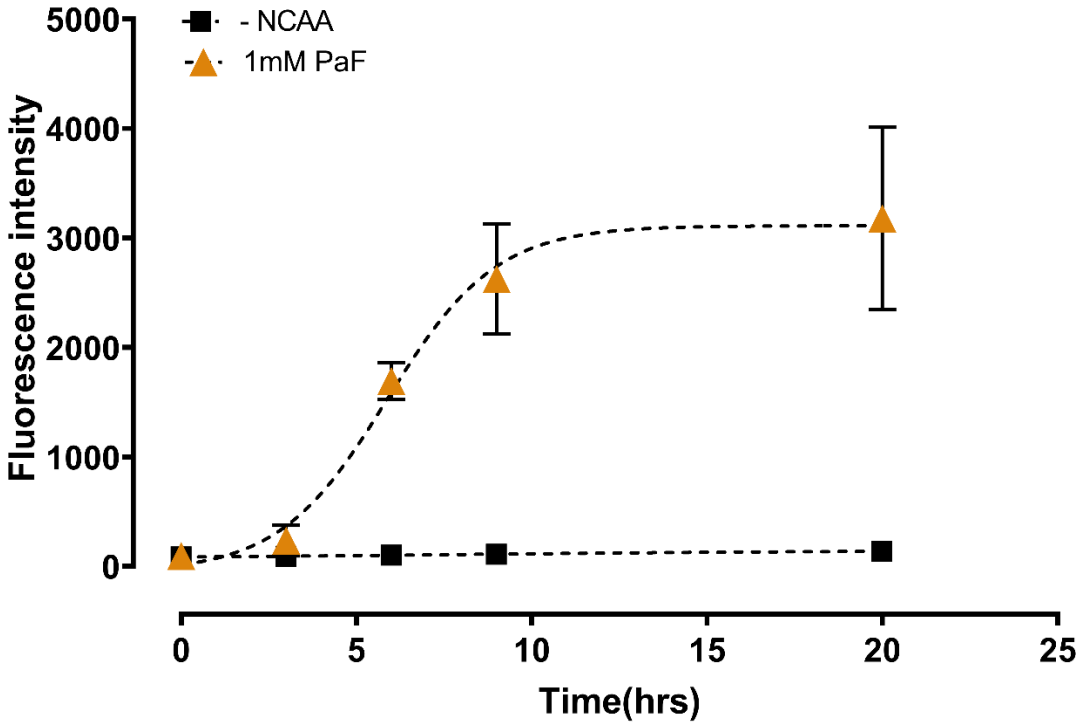


Figure 2.5) Fluorescence intensity using nGFPamb and a previously engineered *Methanocaldococcus jannaschii* TyrRS/$tRNA_{sup}$ system. The reported fidelity of incorporation using this $tRNA_{sup}$/tRNA synthetase pair is 99.8% from (Wang et al., 2003)

We then explored if the cell-like environment of macromolecules in the eCells is beneficial for stability and activity of enzymes. Using an established example of pyrrolysl-tRNA synthetase (PylRS) to incorporate a non-canonical amino acid (ncAA) we demonstrate the ease in which a traditionally difficult ncAA incorporation can be achieved using encapsulated CFPS. The experiment is based on the work by Yokoyama and co-workers who previously

engineered a *Methanosarcina mazei* PylRS/$tRNA_{sup}$ pair for N^{ε}-(tert-butoxycarbonyl)-L-lysine (boc-lysine) incorporation into proteins (Yanagisawa et al., 2008).

Results in figure 2.6 show a clear difference of fluorescence with 1 mM boc-lysine and without the ncAA, indicating that there was significant incorporation of boc-lysine in eCell CFPS. A small amount of amber suppression with canonical amino acids is also observed in the control experiment omitting boc-lysine in the reaction buffer, which is also reported previously (Odoi et al., 2013). Mis-acylation of (WT) suppressor tRNA by *M. barkeri* pyrrolysyl-RS is common and this causes low expression of nGFP in the control experiment. In difference, this is not seen when PaF was incorporated into nGFP, as the Mj. TyrRS/$tRNA_{sup}$ pair has a reported high fidelity for ncAA incorporation of 99.8% and very low tendency for mis-incorporation (Wang et al., 2003).

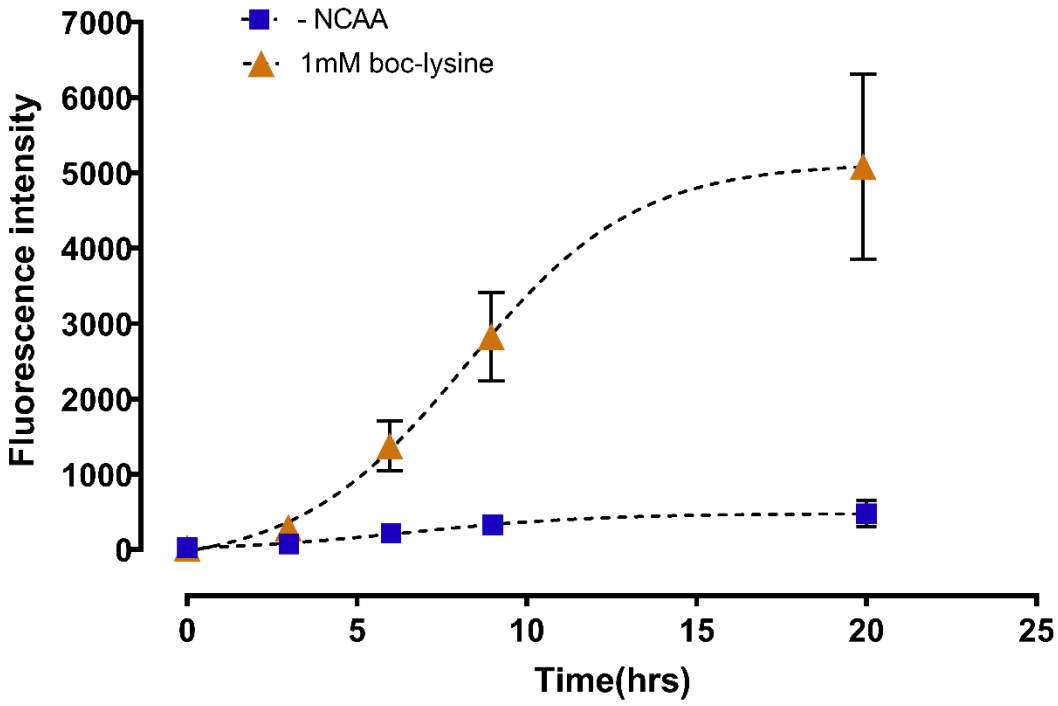


Figure 2.6) Fluorescence intensity using A13 nGFP using boc-lysine lysine and PylRS/$tRNA_{sup}$ system. The negative control produced small amounts of nGFP due to misacylation of the tRNA by promiscuity of the aaRS towards canonical amino acids.

To confirm that ncAA incorporation occurred successfully, a large scale encapsulation and eCell CFPS was conducted in 3 mL of MD/HMP CFPS buffer, with A147amber PPiB as reporter protein. Full length PPiB expressed in high yield (0.5 mg; Figure 2.7A) with modest

requirement of the ncAA; only 0.74 mg of boc-lysine was supplemented to the eCell CFPS reaction.

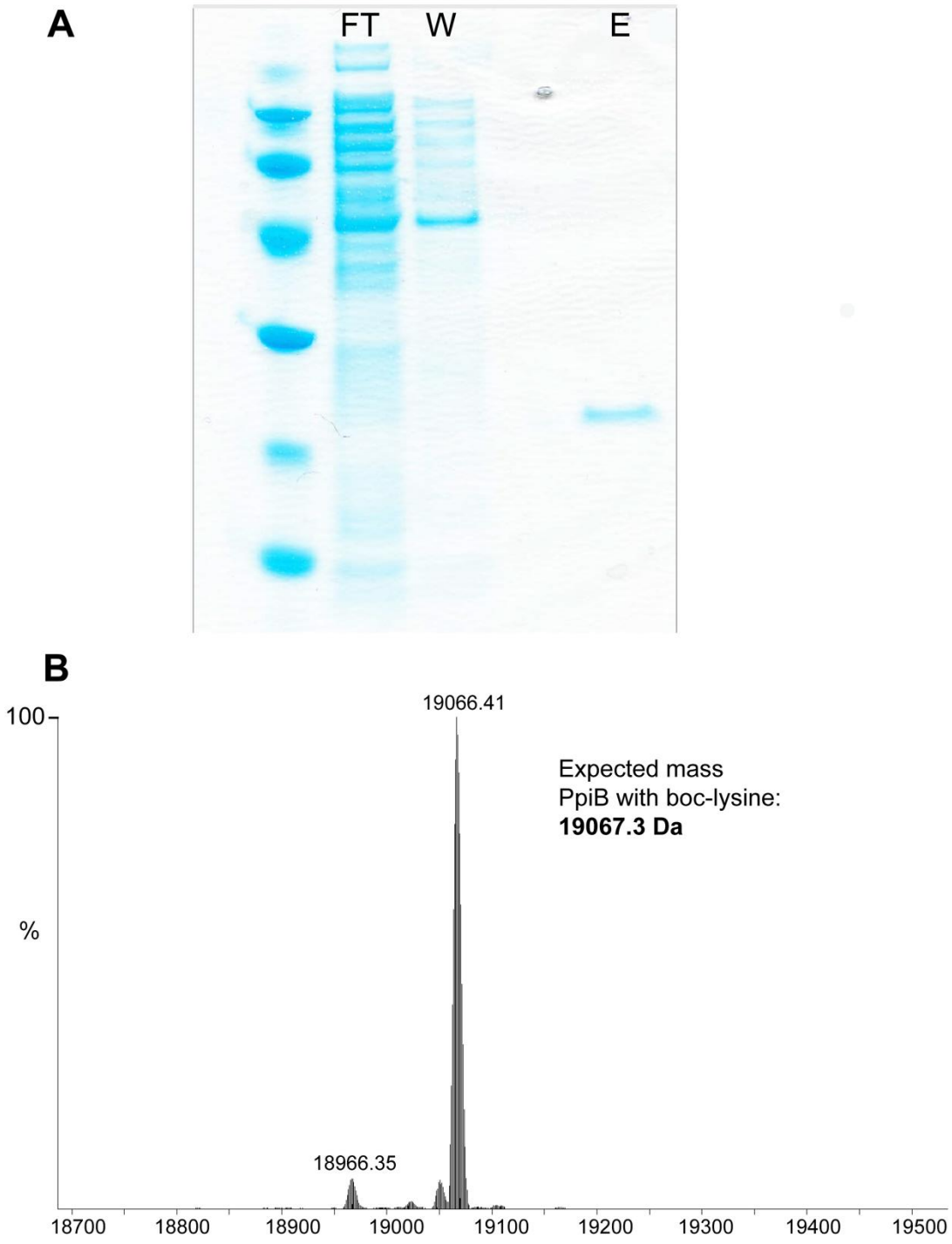


Figure 2.7) A.) SDS-PAGE gel of different fractions during the purification of PPiB, showing flow through (FT) wash(W) and elution (E) **B.)** Whole protein mass spectrum of PPiB with boc-lysine at position 147. Expected mass: 19067.3 Da. Other peaks are related to natural amino acid incorporation.

Mass spectrometry confirmed the successful incorporation of boc-lysine into PPiB. The major mass peak at 19066.4 Da corresponds to boc-lysine containing PPiB. The second, smaller peak at 18966.3 Da corresponds to amber codon suppression with glutamine. Using eCell CFPS, we demonstrate that chemically interesting ncAAs which are expensive when used *in vivo* can be used at a much lower cost using eCells. This is highlighted with the cost regarding eCells incorporating boc-lysine, where 1 g of boc-lysine is required to supplement 1 L of media costing $117, while eCells reactions utilise orders of magnitude less costing only ¢0.5 per 3 mL reaction.

To compare expression yields between dialysis CFPS and eCell CFPS, 300 mg of eCells and 300 µl of S30 extract were used to express WT PpiB in a creatine phosphate based CFPS. The same amount of creatine phosphate was used in both expressions. eCells generated 2.21 mg (±0.2 mg), while dialysis CFPS generated 3.99 mg (±1.2 mg). this means that eCells are approximately half as productive as dialysis CFPS (Appendix figure 1). Yields for eCell expression with ncAA incorporation was also assessed. Amber 147 PpiB was used to report incorporation of boc-lysine with PylRS/$tRNA_{sup}$. 300 mg of eCells with the reporter plasmid amber 147 PpiB and the plasmid with PylRS/$tRNA_{sup}$ were used for expression and yielded 0.46 mg (±0.1 mg) of PpiB with boc-lysine incorporated. This is 21% of the WT expression yield, which is unsurprising as incorporation of ncAA reduces expression yield significantly. The yield between conventional dialysis CFPS and eCells expression is competitive especially when cost is evaluated.

Photocaged-cysteine (PCC) is cysteine with a photocage group covalently linked to the thiol group (Welegedara et al., 2018). Site-specific incorporation of this amino acid was achieved using a suppressor tRNA from *M. mazei* and a *M. barkeri* pyrrolysyl-tRNA synthetase, with mutations to specifically recognize the ncAA. Due to difficulties in purification of the active tRNA synthetase, site-specific incorporation of PCC into proteins is not feasible *in vitro* and highlights the importance of eCells as a viable method for producing proteins with ncAA.

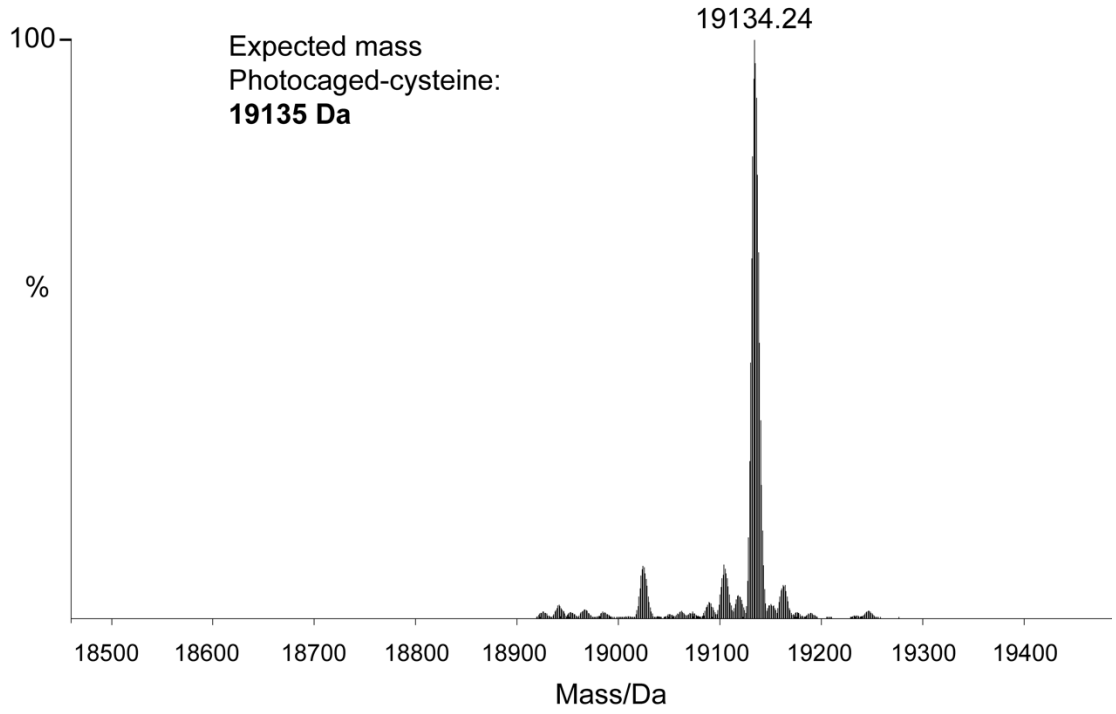


Figure 2.8) Mass spectrum of PpiB with photocaged-cysteine incorporated at position 147. The calculated mass is 19135 Da. Only a small amount of the protein is found to be decaged with a mass of 18941 Da.

As shown in figure 2.8, PCC was confirmed to be incorporated in PPiB at position A147 with a yield of 0.65 mg protein produced from a 30 mL eCell CFPS reaction. The expected mass is 19,135 Da with the expected mass of the decaged cysteine at 18941 Da. The successful incorporation of this PCC highlights the economical use of ncAA and the robustness of the system.

2.44 FACS of eCells

A further advantage of cell-free protein synthesis in nano-encapsulation over cell-free protein synthesis using cell lysate is that multiple synthesis reactions can be performed simultaneously and later decoded. To demonstrate, we used an amber interrupted fluorescent protein (amber 13 RFP) to report on successful amber suppression with a *M. Mazei* PylRS/tRNA pair. FACS was used to identify the population of capsules with RFP expressed in three conditions: with PylRS/$tRNA_{sup}$ and ncAA, with PylRS/$tRNA_{sup}$ and no ncAA and reporter without the synthetase. Figure 2.9 summarizes the FACS results from sorting eCells.

The histograms show a significant shift in capsule populations under the three conditions. In presence of the substrate ncAA, boc-lysine and the aaRS during expression, approximately one third of capsules contain high amounts of full length, functional RFP. In contrast, encapsulations without PylRS/$tRNA_{sup}$ and in absence of ncAA exhibit no red fluorescence, and only a small number of capsules display red fluorescence when only the ncAA was omitted. The latter result likely stems from amber suppression with natural amino acids.

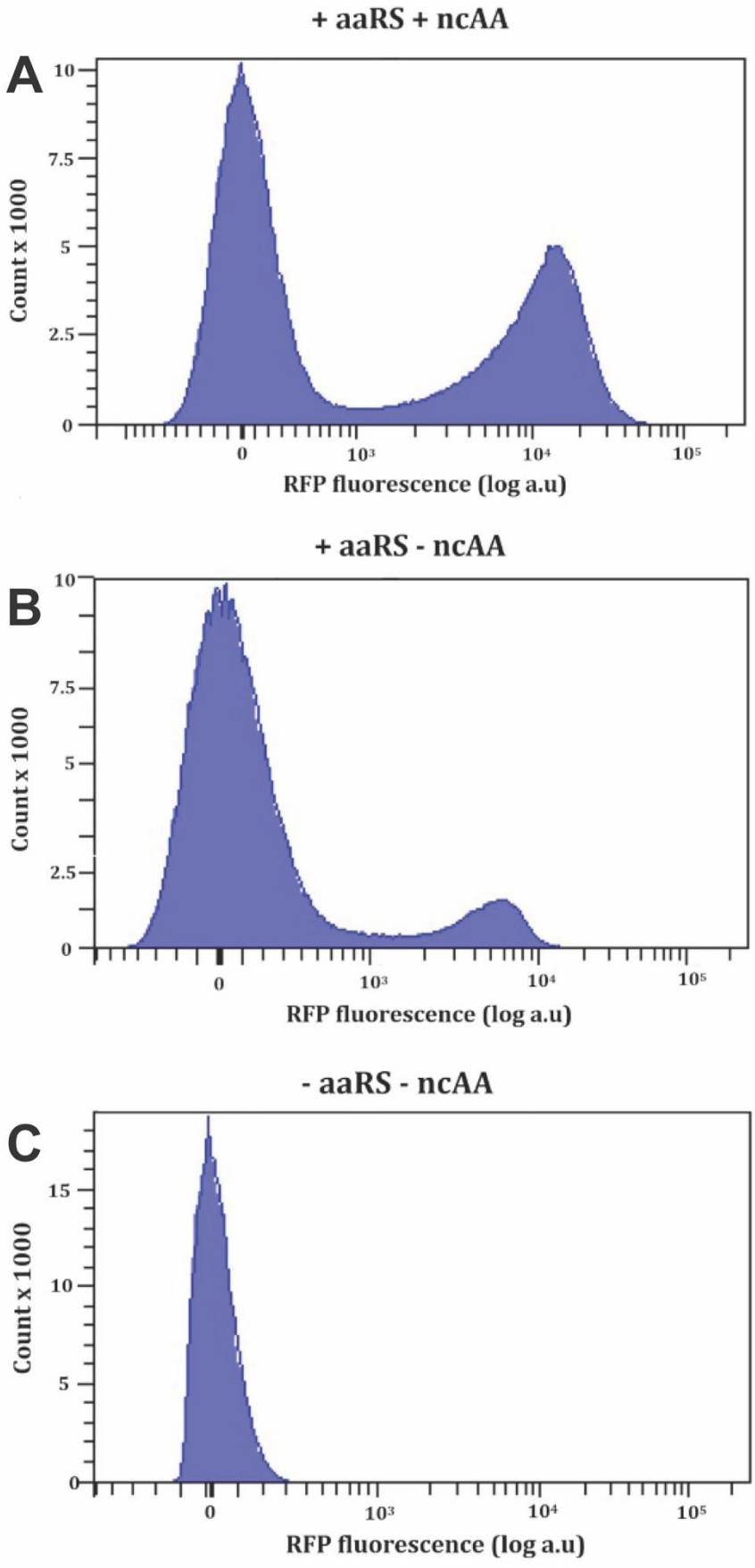

+ aaRS + ncAA
A
Count x 1000
10
7.5
5
2.5
0
0
10³
10⁴
10⁵
RFP fluorescence (log a.u)
+ aaRS - ncAA
B
Count x 1000
10
7.5
5
2.5
0
0
10³
10⁴
10⁵
RFP fluorescence (log a.u)
- aaRS - ncAA
C
Count x 1000
15
10
5
0
0
10³
10⁴
10⁵
RFP fluorescence (log a.u)

Figure 2.9) Histograms of eCells transformed that have been induced with CFPS for expression of amber RFP with boc-lysine incorporated at position 13. **A.)** FACS histogram with the PylRS and ncAA present in the eCells during expression **B.)** FACS histogram of CFPS induced eCells with PylRS present, but expression was conducted in the absence of boc-lysine. **C.)** FACS histogram of eCells expressed without PylRS and omitting ncAA
After a dual transformation of the plasmids with aaRS/tRNA pair and the reporter gene, CFPS can be started the same day as inoculation, without the need for the time-consuming purification step of the tRNA synthetase. Encapsulated CFPS provides a platform that can easily perform ncAA incorporation and streamlines *in vitro* expression of ncAA modified proteins.

2.5 Discussion

CFPS is a technology which allows synthesis of proteins that are difficult to produce or have chemically interesting groups incorporated. Here we have developed the eCell technique where one can reliably control environmental factors and cost-effectively produce these specialized proteins. The eCell system circumvents issues with enzymes that are difficult to purify and use *in vitro*, as it is the case with PylRS. The aaRS/$tRNA_{sup}$ purification process for PylRS is complex, which complicates ncAA incorporation. Many ncAAs that are incorporated with PylRS mutants are expensive to produce and so far could only be used in *in vivo* protein expression. A previously described method to circumvent this complex purification was by creating a S30 extract transformed with the cognate aaRS/$tRNA_{sup}$ pair. A culture expressing the aaRS/$tRNA_{sup}$ was grown for ~13 hours and then the S30 extract was prepared. This was successful in a standard CFPS setting and allowed for ncAA incorporation in a PylRS system (Chemla et al., 2015).

eCells allow the use of PylRSs more easily as they combine advantages of *in vivo* and *in vitro* protein expression. The stabilizing effect in the eCells allow the use and storage of less stable proteins, as is the case with PylRS. This has been shown in eCells with the incorporation of two PylRS based ncAA into PPiB at adequate yields. The crowded environment in eCells have attributes similar to what is encountered in living cells and eCells are compartmentalized like single cells, rather than a whole lysate. Further, the entire protein repertoire is contained within these capsules which could allow the use of other cellular machinery beyond that of transcription and translation. eCells are also comparable to other *in vitro* CFPS systems with

the numerous advantages that are associated with CFPS. The compartmentalization of proteins in eCells, which are permeable to small molecules, mimic features from dialysis CFPS. During CFPS eCells are immersed and resuspended in CFPS buffer. This allows higher translation yield, because phosphate, which in high concentration inhibit translation, can freely diffuse to the larger, outer reaction buffer. The chemical environment of the eCells can be altered to fit precisely the desired experimental conditions and reaction can be repeatedly scaled. CFPS in eCells is highly versatile and can be used with conventional and non-conventional energy generation systems such as CP and MD/HMP, which furthers its accessibility as a platform.

Difficult and expensive ncAA, such as PCC, can be incorporated at a much lower cost than traditionally possible with *in vivo* methods. This is due to the large reduction of reaction volume, the inexpensive production of the eCells and the very modest requirement of ncAA. While the protein expression yield with eCell CFPS is markedly lower than what can be achieved with *in vivo* expression, overall costs are significantly reduced when expensive supplements are necessary. For example, *in vivo* expression yield for PPiB containing a single PCC was 4.5 mg per liter cell culture, while the yield of the same protein when using eCell CFPS was 0.65 mg (Welegedara et al., 2018). Considering the reaction volume in the eCell CFPS was only 30 mL, utilising the ncAA in eCells could be considered an alternative, as only 32 mg of the amino acid consumed, rather than 314 mg used in a 1 L cell culture. Although eCells could be an alternative, the culturing conditions are different, with eCells being immersed as an aggregate into CFPS reaction buffer, while cells are resuspended uniformly typically at 37ºC in LB media, which is more practical. When compared to traditional CFPS, it still retains its effectiveness as the yield is reduced by 45% (Appendix figure 1). This is furthered when the cost and speed of eCell preparation is taken into consideration, accentuating eCells as competitive and productive *in vitro* system.

Moreover, the screening capabilities of eCells is highlighted with the use of small amounts of both eCells and ncAA. Experiments with a fluorescent reporter used the equivalent of cells from 17 mL of LB in 500 µL of CFPS buffer. This coupled with approximately 100 µg of ncAA being used for both, makes the system convenient for small scale CFPS. Because of this economic use of expensive supplements, we believe that eCells are a viable method to screen PylRS variants for their incorporation of hard to incorporate ncAAs. It is important to

note that there is misincorporation of glutamine in the FACS screen amounting to approximately 10 % of the protein as seen in figure 2.8 B.

Cost is reduced further through the exclusion of dialysis tubing, continuous flow or other apparatus that are used to increase the reaction of CFPS. These apparatuses complicate CFPS and increases costs, restricting the technology to be applied in specialized laboratories only. In contrast, the eCell method is economical and relatively quick to perform when compared to the preparation of S30 or cell lysate for conventional CFPS (Apponyi et al., 2008). Methods of lysate preparation are known to take an extensive amount of time, cost and effort to prepare. The process requires cell breakage, washing with various buffers, centrifugation of large volumes at high speed, and require expensive reagents. The standard cost of 20 L S30 extract preparation as detailed in Apponyi et al was tabulated to be $645 AUD from reagents readily available on Sigma-Aldrich, while eCells were found to be $86 for the same volume. Furthermore, eCell preparation can be completed within a timeframe of 2 hours and then used immediately after a single freeze thaw cycle. Inexpensive polyelectrolytes, such as chitosan and sodium alginate, are used for the LbL assembly and the encapsulated cells can be conveniently stored at -80^{o}C with no or little loss of function over time. eCells are not simply a new *in vitro* method but are individual capsules housing genetic material. This demarcation of capsule is important, as it confers unique properties of a single cell, but still allows explicit control over the chemical environment. In this way, minute quantities of eCells and ncAAs are employed to explore the feasibility of integrating non-canonical amino acids into CFPS and FACS. By utilizing eCells, the genetic information encapsulated within these entities can be deliberately and independently separated from the expressed protein, thereby enabling a controlled investigation of the relationship between genotype and phenotype.

The eCell system distinguishes itself from other capsule preparation systems such as microfluidics in several key aspects. Unlike microfluidic capsules, which often require complex engineering and precise control over fluid flow, the eCell system simplifies the process by encapsulating bacteria, eliminating the need for elaborate preparations (Ayoubi-Joshaghani et al., 2020). This encapsulation retains the small cell-like structure within in vitro reactions in a high throughput manner, providing a cost-effective and accessible method for protein synthesis. In contrast, liposomes, although capable of encapsulating lysate, lack the

native cell-like structure and some components that are naturally found in eCells which allow for the retention of some characteristics of living cells.

It is important to highlight that the eCell technology exhibits certain limitations in contrast to conventional CFPS. This arises from the requirement of consistent production between batches of eCells with different vectors. This necessity contrasts with the standard S30 lysate preparation, which offers greater convenience. This is due to the adaptability of S30 to accommodate plasmid addition in the reaction mixture ad hoc, as well as its stable storage capability at -80°C. When compared to *in-vivo* systems, the utilization of eCell technology for protein expression is more expensive. This is primarily due to the requirement of substantial labour and increased reagent costs as opposed to expression in media. The costs associated with reagents have been tabulated by Warfel et al. in 2023. Specifically, the cost per mL of CFPS buffer amounts to USD $4.93. Consequently, for a 30 mL expression, the total cost of CFPS reagents would reach approximately $147.9 not taking into consideration labour cost. In contrast, conducting the same culturing process in LB media would cost a significantly lower amount of $2.4, accentuating the practicality of in-vivo expression.

Although this is the case for in-vivo expression, the eCell technology provides a cost-effective alternative to conventional CFPS implementations. This is due to the easy variation in size from laboratory scale screening experiments (as small as a single cell encapsulation) to pilot scale industrial processes of 100 mL or more eCell CFPS reaction volume. While eCells present a cost-effective alternative to traditional cell-free protein synthesis (CFPS) systems, it is important to acknowledge that the eCell system requires substantial optimization and characterization to adequately evaluate its performance relative to standard CFPS and its diverse advantages.

2.6 References

1. Apponyi, M. A., Ozawa, K., Dixon, N. E. & Otting, G. (2008). Cell-free protein synthesis for analysis by NMR spectroscopy. Methods Mol. Biol., 426, 257-268.
2. Ariga, K., Hill, J. P. & Ji, Q. (2007). Layer-by-layer assembly as a versatile bottom-up nanofabrication technique for exploratory research and realistic application. Phys. Chem. Chem. Phys., 9, 2319-2340.

3. Tong, W., Song, X. & Gao, C. (2012). Layer-by-layer assembly of microcapsules and their biomedical applications. Chem. Soc. Rev. 41, 6103-6124.
4. Ayoubi-Joshaghani, M.H., Dianat-Moghadam, H., Seidi, K., Jahanban-Esfahlan, A., Zare, P. & Jahanban-Esfahlan, R. (2020). Cell-free protein synthesis: The transition from batch reactions to minimal cells and microfluidic devices. Biotechnol. Bioeng., 117, 1204–1229.
5. Cai, Q., Hanson, J.A., Steiner, A.R., Tran, Q., Masikat, M.R., Chen, R., Zawada J.F., Sato, A.K., Hallam, T.J. & Yin, G. (2015). A simplified and robust protocol for immunoglobulin expression in Escherichia coli cell-free protein synthesis systems. Biotechnol. Prog., 31, 823–831.
6. Caschera, F. & Noireaux, V. (2015). A cost-effective polyphosphate-based metabolism fuels an all *E. coli* cell-free expression system. Metab. Eng., 27, 29–37.
7. Chemla, Y., Ozer, E., Schlesinger, O., Noireaux, V. & Alfonta, L. (2015). Genetically Expanded Cell-Free Protein Synthesis Using Endogenous Pyrrolysyl Orthogonal Translation System. Biotechnol. Bioeng., 112, 1663–1672.
8. Fischetti, V. A. (2008). Bacteriophage lysins as effective antibacterials. Curr. Opin. Microbiol., 11, 393–400.
9. Hillberg, A. L. & Tabrizian, M. (2006). Biorecognition through layer-by-layer polyelectrolyte assembly: In-situ hybridization on living cells. Biomacromolecules., 7, 2742-2750.
10. Hovijitra, N. T., Wuu, J. J., Peaker, B. & Swartz, J. R. (2009). Cell-free synthesis of functional aquaporin Z in synthetic liposomes. Biotechnol. Bioeng., 104, 40–49.
11. Jewett, M.C. & Swartz, J.R. (2004). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol. Bioeng., 86, 19–26.
12. Lai, S. N., Zhou, X., Ouyang, X., Zhou, H., Liang, Y., Xia, J. & Zheng, B. (2020). Artificial cells capable of long-lived protein synthesis by using aptamer grafted polymer hydrogel. ACS Synth. Biol., 9, 76–83.
13. Lee, D. W., Lim, C., Israelachvili, J. N. & Hwang, D. S. (2013). Strong adhesion and cohesion of chitosan in aqueous solutions. Langmuir, 29, 14222–14229.
14. Lee, K. Y. & Mooney, D. J. (2012). Alginate: Properties and biomedical applications. Prog. Polym. Sci., 37, 106–126.

15. Lian, Q., Cao, H. & Wang, F. (2014). The cost-efficiency realization in the Escherichia coli-based cell-free protein synthesis systems. Appl. Biochem. Biotechnol., 174, 2351–2367.
16. Nirenberg, M.W. & Matthaei, J.H. (1961). The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. Proc. Natl. Acad. Sci. U.S.A., 47, 1588–1602.
17. Malanovic, N. (2016). Gram-positive bacterial cell envelopes: The impact on the activity of antimicrobial peptides. Biochim. Biophys. Acta-Biomembr., 1858, 936–946.
18. Nishimura, K., Matsura, T., Nishimura, K., Sunami, T., Suzuki, H. & Yomo, T. (2012). Cell-free protein synthesis inside giant unilamellar vesicles analyzed by flow cytometry. Langmuir, 28, 8426–8432.
19. Odoi, K. A., Huang, Y., Rezenom, Y. H. & Liu, W. R. (2013). Nonsense and Sense Suppression Abilities of Original and Derivative Methanosarcina mazei Pyrrolysyl-tRNA Synthetase-tRNA Pyl Pairs in the Escherichia coli BL21(DE3) Cell Strain. PLoS One, 8, e57035.
20. Rampioni, G., D'Angelo, F., Leoni, L. & Stano, P. (2019). Gene-expressing liposomes as synthetic cells for molecular communication studies. Front. Bioeng. Biotechnol., 7, 1-10.
21. Rosenblum, G. & Cooperman, B.S. (2014). Engine out of the chassis: Cell-free protein synthesis and its uses. FEBS Lett., 588, 261–268.
22. Scott, D. J. & Plückthun, A. (2013). Direct molecular evolution of detergent-stable g protein-coupled receptors using polymer encapsulated cells. J. Mol. Biol., 425, 662–677.
23. Shelby, M.L., He, W., Dang, A.T., Kuhl, T.L. & Coleman, M.A. (2019). Cell-Free Co-Translational Approaches for Producing Mammalian Receptors: Expanding the Cell-Free Expression Toolbox Using Nanolipoproteins. Front. Pharmacol., 10, 744.
24. Shimizu, Y., Kuruma, Y., Kanamori, T. & Ueda, T. (2014). The PURE System for Protein Production. In Methods in molecular biology (Clifton, N.J.) vol. 1118, 275–284.
25. Shinde, U. A. & Nagarsenker, M. S. (2009). Characterization of gelatin-sodium alginate complex coacervation system. Indian J. Pharm. Sci., 71, 313–317.
26. Tong, W., Song, X. & Gao, C. (2012). Layer-by-layer assembly of microcapsules and their biomedical applications. Chem. Soc. Rev., 41, 6103.

27. Torre, P., Keating, C. D. & Mansy, S. S. (2014). Multiphase water-in-oil emulsion droplets for cell-free transcription-translation. Langmuir, 30, 5695–5699.
28. Wang, L., Zhang, Z., Brock, A. & Schultz, P. G. (2003). Addition of the keto functional group to the genetic code of Escherichia coli. Proc. Natl. Acad. Sci., 100, 56–61.
29. Warfel, K. F., Williams. A., Wong. D. A., Sobol, S.E., Desai, P., Li, J., Chang, Y.F., DeLisa, M.P., Karim, A.S. & Jewett, M. C. (2023) A Low-Cost, Thermostable, Cell-Free Protein Synthesis Platform for On-Demand Production of Conjugate Vaccines. ACS. Synth. Biol. 12, 95-107.
30. Welegedara, A. P., Adams, L. A., Huber, T., Graham, B. & Otting, G. (2018). Site-Specific Incorporation of Selenocysteine by Genetic Encoding as a Photocaged Unnatural Amino Acid. Bioconjug. Chem., 29, 2257–2264.
31. Wu, P. S. C., Ozawa, K., Lim, S. P., Vasudevan, S. G., Dixon, N. E. & Otting, G. (2007). Cell-Free Transcription/Translation from PCR-Amplified DNA for High-Throughput NMR Studies. Angew. Chemie Int. Ed., 46, 3356–3358.
32. Yanagisawa, T., Ishii, R., Fukunaga, R., Koyabashi, T., Sakamoto, K. & Yokoyama, S. (2008). Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode Nε-(o-Azidobenzyloxycarbonyl) Lysine for Site-Specific Protein Modification. Chem. Biol., 15, 1187–1197.
33. Yong, K. J. & Scott, D. J. (2015). Rapid directed evolution of stabilized proteins with cellular high-throughput encapsulation solubilization and screening (CHESS). Biotechnol. Bioeng., 112, 438–446.
34. Zhuang, L., Huang, S., Wan-Qiu, L., Karim, A.S., Jewett, M.C. & Li, J. (2020). Total *in vitro* biosynthesis of the nonribosomal macrolactone peptide valinomycin. Metab. Eng., 60, 37–44.

CHAPTER THREE

Cell-free Synthesis of Proteins with Selectively ^{13}C-Labeled Methyl Groups from Inexpensive Precursors

3.1 Abstract

The novel eCell system maintains the activity of the entire repertoire of metabolic *E. coli* enzymes in cell-free protein synthesis. We show that this can be harnessed to produce proteins with selectively ^{13}C-labeled amino acids from inexpensive ^{13}C-labeled precursors. The system is demonstrated with selective ^{13}C-labelling of methyl groups in the proteins ubiquitin and peptidyl-prolyl *cis–trans* isomerase B. Starting from 3-^{13}C-pyruvate, ^{13}C-HSQC cross-peaks are obtained devoid of one-bond ^{13}C-^{13}C scalar couplings. Starting from 2-^{13}C-methyl-acetolactate, single methyl groups of valine and leucine are labeled. Labelling efficiencies are 70 % or higher, and the method allows to produce perdeuterated proteins with protonated methyl groups in residue-selective manner. The system uses the isotope-labeled precursors sparingly and is more readily scalable than conventional cell-free systems.

3.2 Introduction

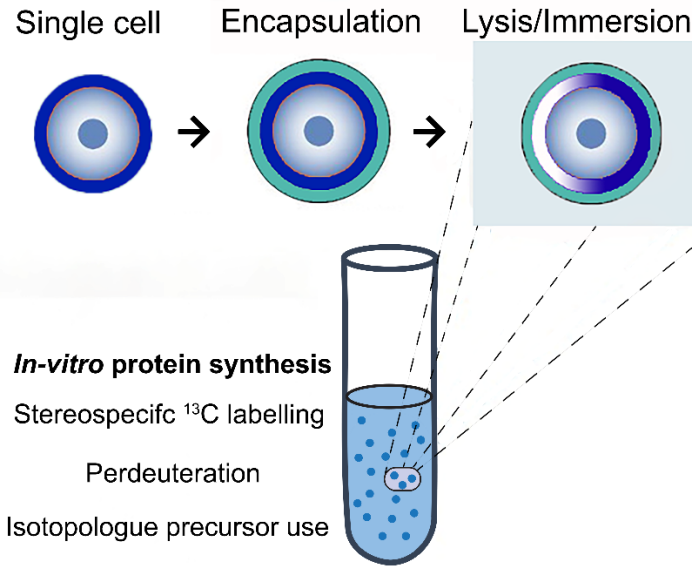


Figure 3.1) The schematic illustrates a simple step-by-step process involved in eCell production, along with its diverse applications specifically within the realm of in-vitro protein synthesis and NMR.

The NMR resonance assignments of high-molecular weight proteins critically depends on the availability of samples enriched with stable isotopes (Tugarinov et al., 2006). Conventional strategies based on uniformly ^{13}C-enriched proteins usually employ [U-^{13}C]-glucose as the (de facto) only carbon source in minimal media (Filipp et al., 2009). The ^{13}C-enrichment of proteins enables the sensitive recording of heteronuclear correlation spectra such as 2D ^{13}C-HSQC spectra, which are particularly sensitive for methyl groups. Methyl groups play a privileged role in the NMR analysis of large protein systems in solution, as their signals can be observed for macromolecular complexes as large as 1 MDa (Boswell and Latham, 2018). Methyl-bearing amino acids are abundant not only in the hydrophobic core of globular proteins but also in hydrophobic ligand binding pockets (Otten et al., 2010). Methyl groups thus serve as useful probes for the analysis of protein structure, dynamics and function (Schütz and Sprangers, 2020).

Among the amino acids with methyl groups, the spectral regions of the methyl groups of isoleucine, leucine and valine (ILV) overlap in a ^{13}C-HSQC spectrum (Rasia et al., 2012). This poses a problem for large proteins, which not only contain many methyl groups but also feature broad NMR signals (Lange et al., 2012). Furthermore, uniformly ^{13}C-labeled proteins feature ^{13}C-^{13}C couplings, in particular one-bond ^{13}C-^{13}C couplings $^{1}J_{CC}$, which lead to broad multiplets in the ^{13}C-dimension of ^{13}C-HSQC spectra. Several strategies have been devised to resolve the methyl cross-peaks of ILV residues. (i) Protein samples can be produced from amino acid mixtures containing only a single amino acid with isotope enrichment (Schubert et al., 2006; Schörghuber et al., 2018). Notably, however, suitably labeled amino acids may be available commercially but can be expensive (Takeda et al., 2010). Some of the most affordable versions are uniformly enriched with ^{13}C, which retains the problem of ^{13}C-^{13}C couplings. Alternatively, isotope-labeled precursors can be supplied, but this usually does not solve the problem of ^{13}C-^{13}C couplings (Hajduk et al., 2000), and the approach calls for the use of auxotrophic strains to avoid isotope scrambling (Whittaker, 2007; Lazarova et al., 2018). (ii) ^{13}C-^{1}H correlation spectra can be recorded with homonuclear ^{13}C-decoupling in the ^{13}C-dimension, either by recording as a constant-time experiment (Vuister and Bax, 1992), which sacrifices sensitivity, or by band-selective decoupling (Behera et al., 2020), which

cannot decouple the ^{13}C multiplet of leucine methyls as the ^{13}C chemical shifts of their coupling partners are too close. (iii) Stereospecific-selective labelling of single methyl groups of each valine and leucine side chain can be achieved by providing 2-^{13}C-methylacetolactate in the growth medium (Gans et al., 2010). This method avoids $^{1}J_{CC}$ couplings and has been successfully applied to protein complexes up to 1 MDa (Sprangers and Kay, 2007), but relies on uncompromised biosynthesis pathways for leucine and valine and thus requires *in vivo* protein production and relatively large quantities of the expensive precursor. The cost issue is magnified for large protein complexes, which are not amenable to multi-dimensional NMR experiments, and where the assignment of the methyl cross-peaks therefore is achieved by site-directed mutagenesis. A case in point is the 468-kDa multimeric aminopeptidase PhTET2, where the assignment of the alanine $C^{\beta}H_3$ and isoleucine $C^{\delta}H_3$ groups alone consumed 3.2 L of media with expensive ^{13}C-labeled precursors (Amero et al., 2011).

The optimal labelling scheme would be selective by amino acid type, deliver a labelling pattern where the methyl groups are enriched with ^{13}C, and their directly bonded carbon is ^{12}C (Kasinath et al., 2013). Furthermore, it should be amenable to cell-free protein synthesis (CFPS), which uses isotope labeled compounds sparingly (Torizawa et al., 2004). Unfortunately, the biosynthesis of ^{13}C-labeled amino acids is compromised in *in vitro* protein expression systems (Linser et al., 2014), although a limited degree of metabolism can be restored by re-introducing certain cofactors (Jewett et al., 2008). For example, metabolites from glycolysis can be used for energy generation in CFPS, if cofactors such as NAD^+ and CoA are provided (Kim and Swartz, 2001). Energy generation systems have also been based on phosphoenol pyruvate (PEP), as well as pyruvate, glucose and maltodextrin (Caschera and Noireaux, 2015). In our hands, these systems proved to be more difficult to establish presumably because of their dependence on the activity of multiple enzymes from the glycolytic pathway. The problem of maintaining the activity of enzymes required for energy regeneration in CFPS has been solved, however, by the recently established eCell system (Van Raad and Huber, 2021).

eCells are bacterial cells coated with polymers, where the cell wall has been lysed (Van Raad and Huber, 2021). The resulting cells can no longer replicate, but they still contain all biomacromolecules required for protein synthesis, while their porous polymer coat gives low-molecular weight compounds free access to the cytosol. eCells thus are ideal vehicles for CFPS. We hypothesized that eCells also preserve the activity of the enzymes involved in the

biosynthesis of amino acids and therefore allow the production of methyl-labeled amino acids from inexpensive precursors such as 3-^{13}C-pyruvate or ^{13}C-glucose. The present work demonstrates that this is indeed the case.

Different precursors can be used for selective ^{13}C-enrichment of ILV methyl groups. Well-established isotopologue precursors used in *in vivo* protein expression include α-ketobutyrate and α-ketoisovalerate (Goto et al., 1999), which is a key precursor for both leucine and valine synthesis (Lundström et al., 2007). In general, precursors close to the final stages of biosynthesis of methyl bearing amino acids allow the labelling of proteins with high selectivity (Schörghuber et al., 2018). The last precursor for selective stereospecific methyl labelling of leucine and valine is 2-methyl-4-acetolactate (Figure 3.2B), whereas early precursors in the biosynthetic pathways of amino acids, such as pyruvate, cannot label methyl groups in a stereospecific manner. In the following we demonstrate the excellent utility of eCells to produce proteins with selectively ^{13}C-labeled methyl groups from different inexpensive precursors, including pyruvate, 2-methyl-4-acetolactate and glucose.

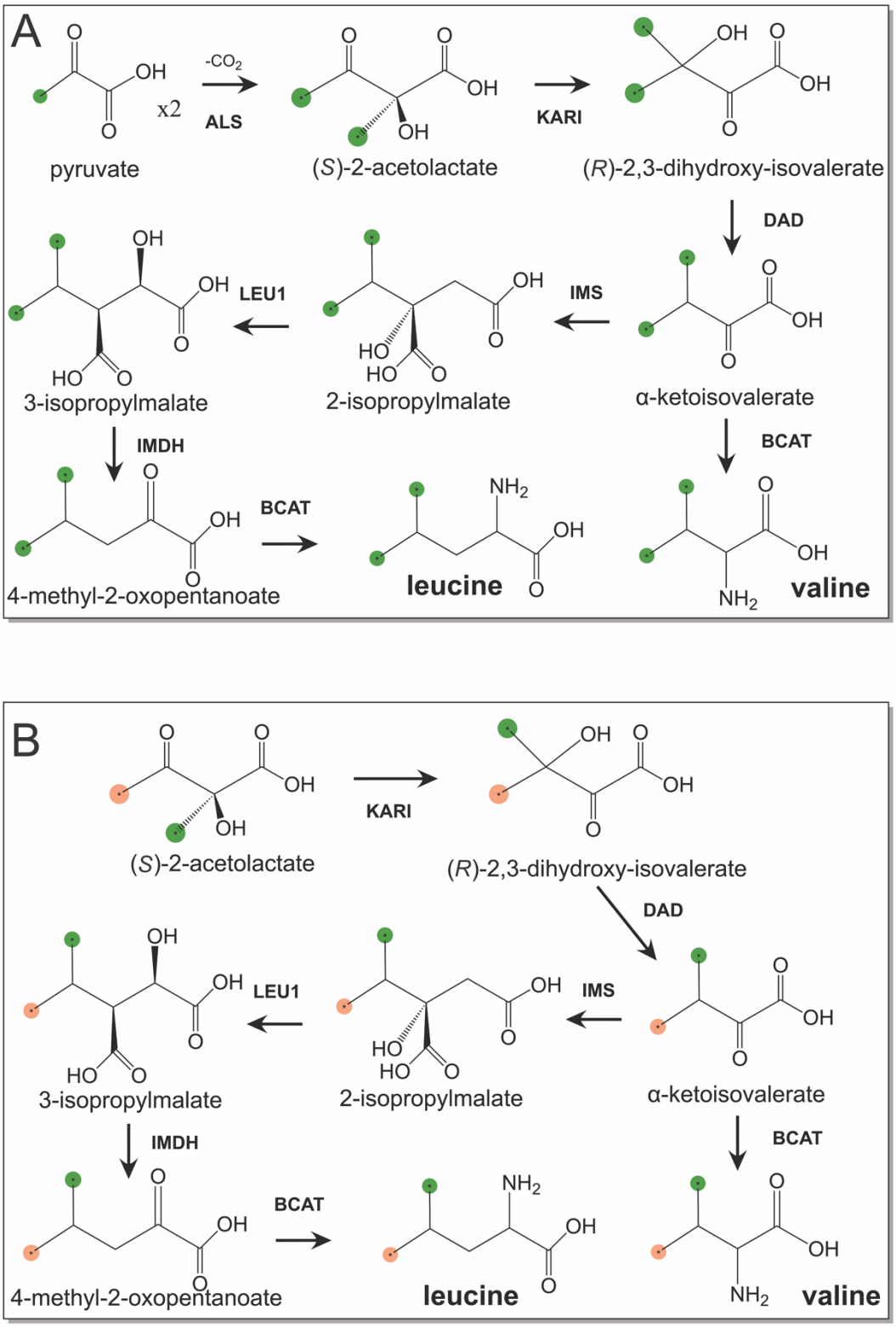


Figure 3.2) Biosynthetic pathways of leucine and valine from isotope-labeled precursors. ^{13}C-labeled methyl groups are identified by green balls, and methyl groups at natural isotopic abundance are highlighted by orange balls. **A.)** Biosynthetic pathway starting from 3-^{13}C-labeled pyruvate. **B.)** Stereoselective biosynthetic pathway starting from (*S*)-2-acetolactate.

Abbreviations used: KARI, ketol-acid reductoisomerase; DAD, dihydroxy-acid dehydratase; IMS, 2-isopropylmalate synthase; LEU1, 3-isopropylmalate dehydratase; IMDH, 3-isopropylmalate dehydrogenase; BCAT, Branched-chain aminotransferase.

3.3 Materials and methods

3.31 Amino acids, precursors and materials

The polyelectrolytes low-molecular weight chitosan (50,000 – 190,000 Da) and sodium alginate were purchased from Merck. The ethyl ester of 2-^{13}C-methyl-4-$^{2}H_3$-acetolactate (ethyl-2-hydroxy-2-^{13}C-methyl-3-oxobutanoate) was purchased from Cambridge Isotope Laboratories (CIL; USA). Perdeuterated amino acids were from CIL and Martek Isotopes (USA). 3-^{13}C-pyruvate was from Sigma-Aldrich.

3.32 Plasmids

A plasmid was constructed with the pCloDF13 origin of replication, the gene of the *E. coli* peptidyl–prolyl *cis–trans* isomerase PpiB with C-terminal His_6-tag under control of the T7 promoter and a spectinomycin resistance gene, generating the plasmid pCDF PpiB CTH. For ubiquitin expression a plasmid was constructed with pCloDF13 origin of replication, the spectinomycin resistance gene and the gene of ubiquitin under control of the T7 promoter (plasmid pCDF Ubi CTH). A *lac* operator was inserted in front of the T7 promoter to reduce background protein expression prior to induction which reduces the ^{13}C labelling efficiency.

3.33 Production of eCells

E. coli XJb cells were transformed with either pCDF Ubi CTH or pCDF PpiB CTH and grown in LB medium at 37 ^{o}C in baffled flasks with shaking at 180 rpm. Endolysin production was induced at the time of inoculation with a final concentration of 3 mM arabinose. Cells were grown to OD_{600} = 0.6, harvested by centrifugation and washed three times with PBS-E buffer (phosphate-buffered saline with 1 mM EDTA, pH 7.4). For coating with chitosan, the cells were resuspended in 0.25 mg/mL chitosan in PBS-E with vigorous shaking for 20 minutes. The cell pellet was washed with PBS-E pH 6.0 three times to remove excess chitosan and then resuspended in 0.25 mg/mL alginate PBS-E solution and subjected

to vigorous shaking for 20 minutes. The cells were then washed 3 times with PBS-E pH 6.0, resuspended in PBS-E pH 7.4 and stored at -80 ^{o}C in PBS-E buffer.

3.34 Production of deuterated eCells

5 g sodium pyruvate was dissolved in 50 mL D_2O and the pH adjusted with 0.1 mM KOD to pH 11. The solution was stirred overnight at 95 ^{o}C to exchange the protons of pyruvate for deuterium. 500 mL M9 minimal media was prepared in D_2O with 22 mM KH_2PO_4, 42 mM Na_2HPO_4, 8.6 mM NaCl, 18.6 mM NH_4Cl, 500 µL 1 mg/mL thiamine (vitamin B6), 0.1 mM $CaCl_2$, 250 µL 1000x metal mixture (50 mM $FeCl_3$, 10 mM $MnCl_2$, 10 mM $ZnSO_4$, 2 mM $CoCl_2$, 2 mM $CuCl_2$ and 2 mM $NiCl_2$), 5 mM $MgSO_4$, 3 mM arabinose and 25 mg/mL spectinomycin. The H-D exchange in pyruvate was confirmed by NMR. The deuterated pyruvate was added to the dry mixture of buffer salts and the final pyruvate-M9 medium made up to 500 mL, adjusted to pH 7.2 and filter sterilised prior to inoculation.

XJb(DE3)* cells that had been transformed with pCDF PpiB CTH were trained for production of perdeuterated proteins in a protocol adapted from that reported by Li and Byrd (2022). 15 mL of an overnight starter culture of pCDF PpiB CTH was diluted with 15 mL of deuterated pyruvate-M9 medium and incubated at 37 ^{O}C with shaking at 180 rpm. When the OD_{600} reached 1.0, the cells were again diluted with 30 mL of deuterated pyruvate-M9 medium and incubated a second time. Upon reaching OD_{600} = 1.0, the 60 mL culture was spun down, the cells transferred to a 50 mL culture and growth continued overnight at 37 ^{O}C with shaking at 180 rpm. The 50 mL culture was added to 400 mL of deuterated pyruvate-M9 medium and left to grow until OD_{600} = 0.75 was reached, after which the cells were encapsulated as described in Section 3.33

.

3.35 CFPS systems

The protocol for pyruvate-based CFPS was adapted from the phosphate recycling system by Jewett and Swartz (2004). The CFPS buffer contained 0.9 mM UTP and CTP, 50 mM HEPES, 1.5 mM GTP, 1.5 mM ATP, 0.68 µM folinic acid, 0.64 mM cAMP, 1.7 mM DTT, 3.5 mM of each amino acid (apart from the amino acid(s) to be synthesized by the eCells for isotope enrichment), 60 mM K-Glu, 8 mM Mg-Glu, 2% v/v PEG-8000, 4 mM sodium oxalate, 0.25 mM CoA and 0.33 mM NAD^+. Roche cOmplete^{TM} Mini protease inhibitor

cocktail was added to the CFPS buffer. Of the volume following dissolution of one tablet in 10 mL water, 10% were added to the CFPS reaction. The reaction was conducted with 33 mM pyruvate.

The protocol for glucose-based CFPS was likewise adapted from the previously published phosphate recycling system (Jewett and Swartz, 2004). The glucose CFPS buffer contained the same components as the pyruvate-based CFPS protocol, but with 10 mM sodium phosphate dibasic pH 7.5 and without sodium oxalate and pyruvate. The reaction was conducted with 30 mM glucose.

The CFPS system using creatine phosphate and creatine kinase as energy source contained the same components as the pyruvate-based CFPS protocol, but without sodium oxalate, pyruvate, CoA or NAD^+ and adding 250 μg/mL creatine kinase, 80 mM creatine phosphate and 6 mM Mg-Glu instead of 8 mM Mg-Glu.

Following addition of the isotopically enriched precursor, the CFPS buffers for each of these reactions were adjusted to pH 7.5. Frozen aliquots of encapsulated cells were thawed, and the pellet immersed in CFPS buffer. CFPS for each experiment was conducted at 37 ^oC with shaking at 180 rpm.

3.36 Acetolactate labelling

2-^{13}C-methyl-4-$^{2}H_3$-acetolactate as the source for prochiral methyl groups was set free from the ethyl ester by incubating in H_2O with 0.1 M NaOH (NaOD for deuteration experiment) (pH 13) at 37 ^oC for 30 minutes. The compound was tested both in pyruvate-based CFPS and in CFPS with the creatine-phosphate/creatine kinase system. The CFPS reaction was conducted in 15 mL buffer with 0.1 mM $NADP^+$, 3.5 mM 2-^{13}C-methyl-4-$^{2}H_3$-acetolactate and 0.2 mM penoxsulam to inhibit the acetolactate synthase (ALS) enzyme. Ubiquitin was produced from 300 mg eCells and purified using His-Gravitrap columns (GE Healthcare, USA).

For perdeuterated CFPS, all buffer stocks were dissolved in D_2O, and the pH was adjusted with KOD to pH 7.2. The creatine phosphate based CFPS reaction was conducted in 20 mL

D_2O buffer with 5 mM 2-^{13}C-methyl-4-$^{2}H_3$-acetolactate, 0.1 mM $NADP^+$, 1 mM of all amino acids in perdeuterated form excluding valine and 0.2 mM penoxsulam to inhibit the acetolactate synthase (ALS) enzyme. PpiB was produced from 800 mg eCells and purified using His-Gravitrap columns (GE Healthcare, USA).

3.37 Labelling with 3-^{13}C-pyruvate or 1-^{13}C glucose

Dry 3-^{13}C-pyruvate was added to 15 mL CFPS buffer at 33 mM final concentration. Leucine, valine and isoleucine were omitted from the amino acid mixture to allow for ^{13}C-labelling of their methyl groups. Ubiquitin and PpiB were expressed using 300 mg eCells and purified using His-Gravitrap columns. To illustrate the scalability of the reaction, ubiquitin samples were also produced with specific labelling of alanine and valine in 5 mL CFPS buffer using 300 mg eCells with the amino acid of interest omitted from the amino acid mixture. PpiB with ^{13}C-labeled valine was produced in 20 mL pyruvate-based CFPS buffer with 1 g eCells with valine omitted from the amino acid mixture.

To test the performance of 1-^{13}C-glucose as ^{13}C source, dry 1-^{13}C glucose was added to 5 mL glucose-based CFPS buffer at 30 mM final concentration (Appendix figure 3). Leucine and valine were omitted from the amino acid mixture to allow for labelling of their methyl groups. To assess the potential of glutamate in the buffer in diluting the ^{13}C-label, reactions were conducted with a buffer containing 60 mM K-Glu/8 mM Mg-Glu or 100 mM adipic acid/8 mM $MgCl_2$.

3.38 NMR spectroscopy and isotope labelling yields

All NMR spectra were recorded at 25 ^{o}C using 5 mm NMR tubes and a Bruker 800 or 600 MHz NMR spectrometer equipped with TCI cryoprobes.

The isotope labelling efficiency of leucine residues in ubiquitin was assessed by integrating the ^{1}H-NMR signals of the γ_2-methyl group of Leu50 and its ^{13}C satellites, which are resolved in the 1D NMR spectrum. For samples without isotope-labeled leucine, the ^{13}C-HSQC cross-peak intensities of the labeled residues were compared with those of an internal standard of 0.1 mM 3-^{13}C-pyruvate.

3.4 Results

3.41 Ubiquitin with ^{13}C-labeled methyl groups in alanine, leucine and valine made from 3-^{13}C-pyruvate

The biosynthetic methyl labelling strategies were validated using ubiquitin as a model protein. The ^{13}C-label was provided by 3-^{13}C-pyruvate, which served both as carbon source for amino acid synthesis and energy source for protein production. Omission of leucine and valine from the reaction mixture allows for detection of ^{13}C-labeled valine and leucine produced from pyruvate during the cell-free reaction. As singly ^{13}C-labeled pyruvate contains no neighbouring ^{13}C atoms, the methyl groups of leucines and valines are expected to not show any $^{1}J_{CC}$ coupling. This expectation was borne out in the experiment, where the cross-peaks revealed no splittings in the ^{13}C-dimension (Figure 3.3). Therefore, this labelling scheme delivers better spectral resolution than uniform ^{13}C-labelling schemes, where multiplet splittings due to $^{1}J_{CC}$ couplings can be avoided only by specific pulse sequences that compromise sensitivity (Vuister and Bax, 1992; Behera et al., 2022). As the biosynthetic pathways remained intact, the ^{13}C-label was subject to incorporation into a range of amino acids and thus prone to some isotope scrambling. For example, isotopic enrichment was also detected for alanine (due to direct conversion of pyruvate to alanine by alanine-transaminase) and the δ_2-methyl group of isoleucine. Importantly, however, the labelling efficiency of the isopropyl groups of leucine and valine was high (about 70%).

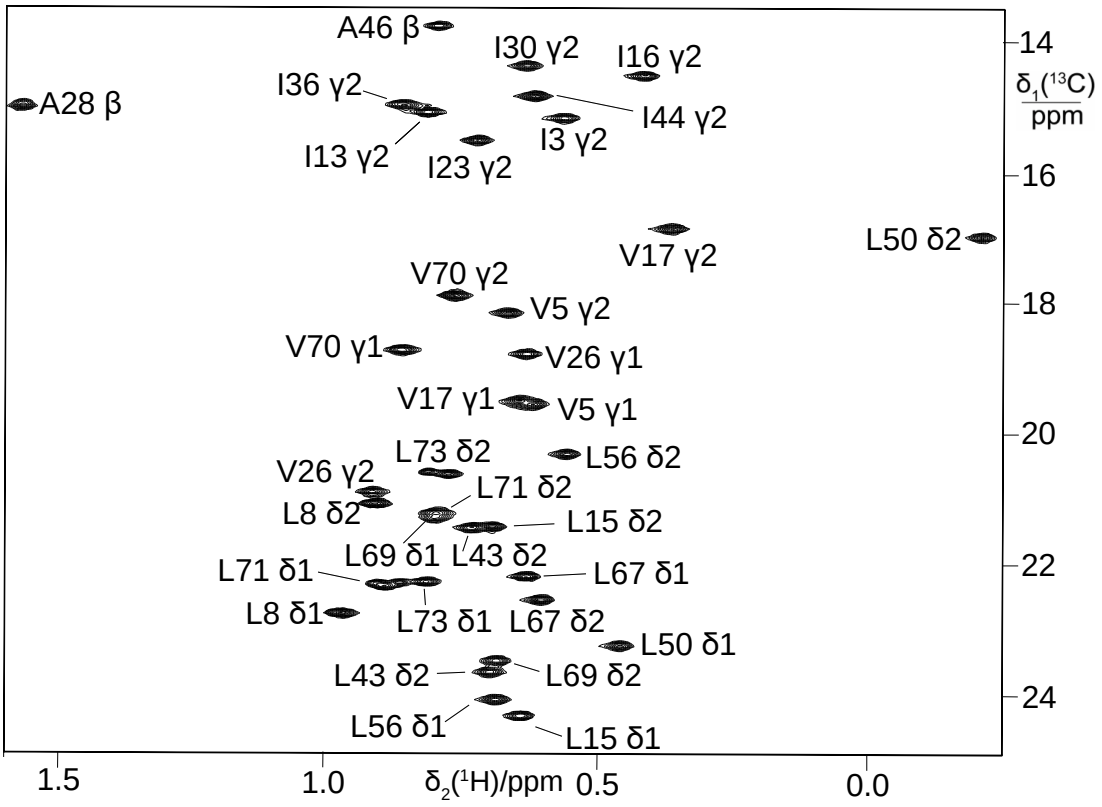
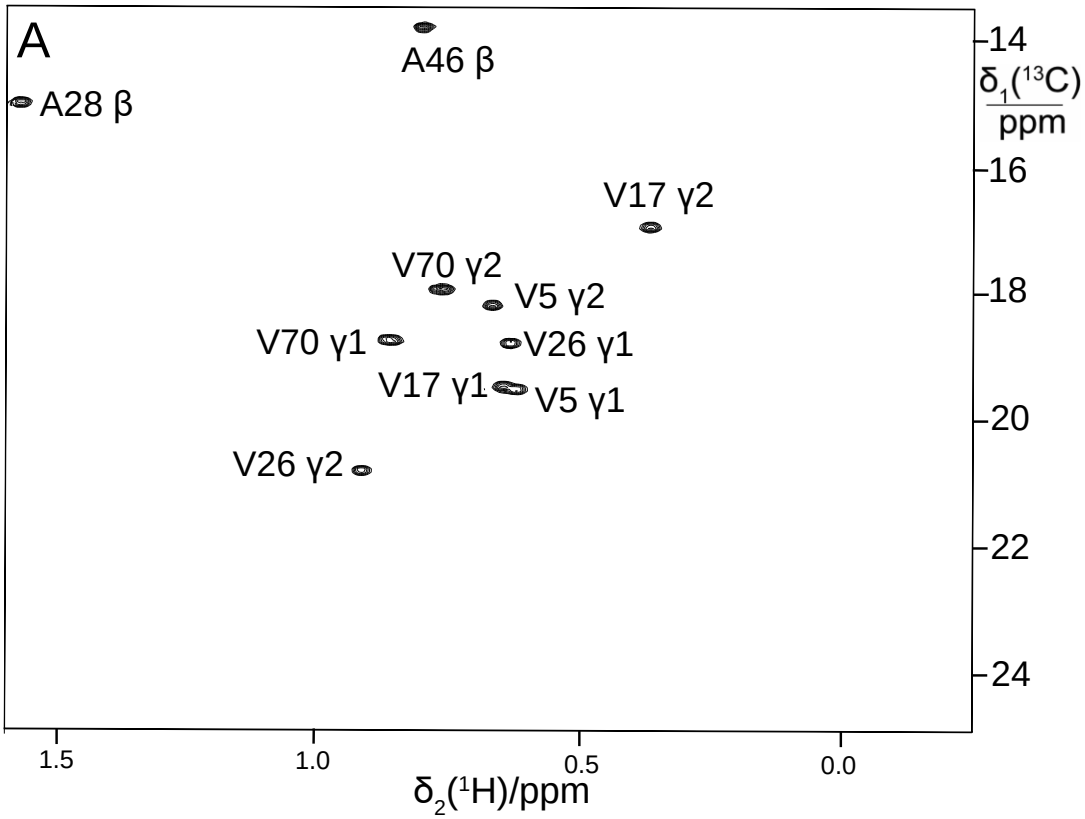


Figure 3.3) ^{13}C-HSQC spectrum of ubiquitin produced from 3-^{13}C-pyruvate by eCell CFPS, resulting in uniform ^{13}C-labelling of both isopropyl methyl groups of leucine and valine. The protein yield was 0.7 mg from 10 mL eCell CFPS reaction, and the level of isotope labelling was 70%.

The absence of ^{13}C-enrichment of the δ-methyl group of isoleucine is a signature of the biosynthetic pathway, where one pyruvate molecule is linked with unlabeled acetyl-CoA to form α-ketobutyrate as the precursor of isoleucine, channelling the ^{13}C label into the γ_2-methyl rather than the δ-methyl group.

As the cell-free reaction was performed in a buffer containing high concentrations of glutamate, we speculated that the degree of isotope labelling could be increased by substituting glutamic acid for adipic acid, which is not easily converted into amino acids (Jia et al., 2009). Using 100 mM adipic acid/8 mM $MgCl_2$ instead of 60 mM K-Glu/8 mM Mg-Glu, however, did not increase the labelling efficiency and slightly decreased the protein yield (data not shown).


A
A46 β
A28 β
V17 γ2
V70 γ2
V5 γ2
V70 γ1
V26 γ1
V17 γ1
V5 γ1
V26 γ2
14
16
18
20
22
24
δ1(13C)
ppm
1.5
1.0
0.5
0.0
δ2(1H)/ppm

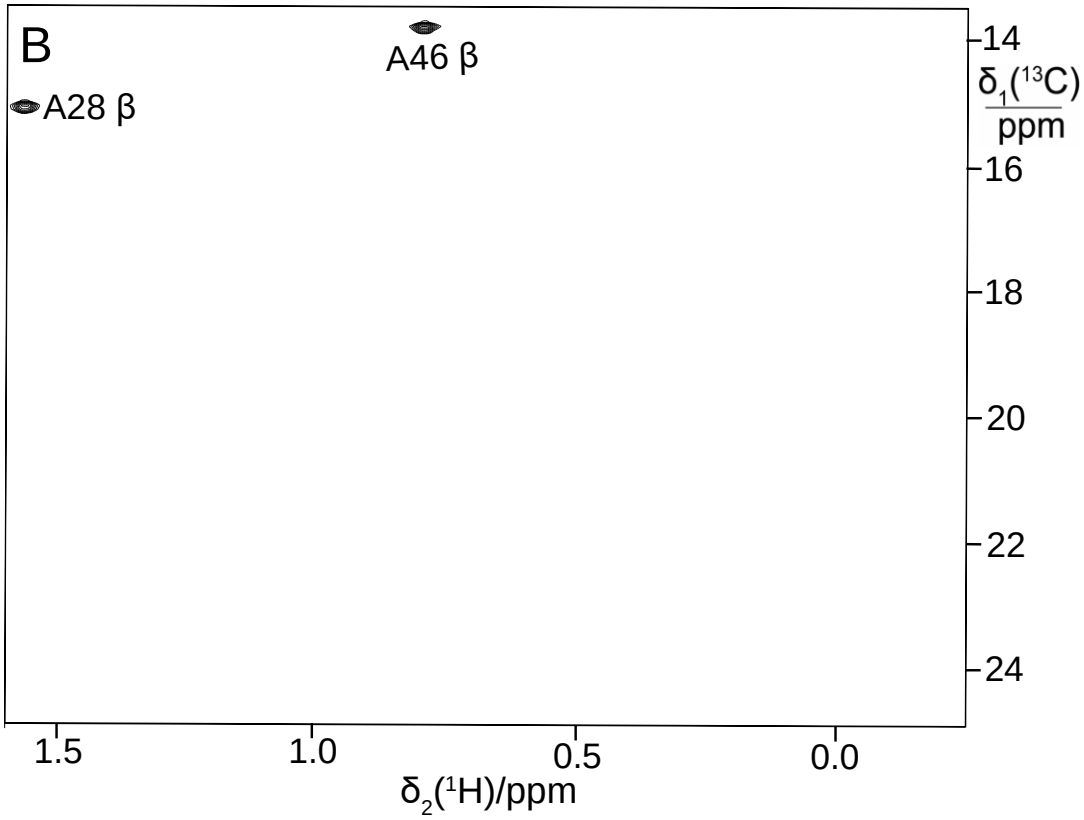

B
A46 β
A28 β
14
16
18
20
22
24
δ1(13C)
ppm
1.5
1.0
0.5
0.0
δ2(1H)/ppm

Figure 3.4) ^{13}C-HSQC spectra of ubiquitin expressed in a 5 mL CFPS reaction using 300 mg eCells with 3-^{13}C-pyruvate. **A.)** Valine was omitted from the amino acid mixture. Protein yield 1.6 mg, ^{13}C-enrichment of the valine methyl groups >70%. **B.)** Alanine was omitted from the amino acid mixture. Protein yield 1.35 mg, ^{13}C-enrichment of the alanine methyl groups >85%.

Starting from 3-^{13}C-pyruvate for biosynthesis, the selectivity of isotope labelling was enhanced by 'unlabelling' the amino acids not of interest for labelling, which is achieved simply by adding them to the CFPS reaction at natural isotopic abundance. For example, the ^{13}C label was apparent only in the valine methyl groups when only valine was omitted from the amino acid mixture (Figure 3.4A), and only alanine peaks were observed when only alanine was left out (Figure 3.4B).

3.42 Ubiquitin with ^{13}C-labeled methyl groups made from 2-^{13}C-methyl-acetolactate

2-^{13}C-methyl-acetolactate has been shown to allow the *in vivo* production of proteins with stereospecifically labeled isopropyl groups of valine and leucine (Gans et al., 2010). To test the performance of this approach with eCells, a sample of ubiquitin was prepared with the provision of 2-^{13}C-methyl-acetolactate and penoxsulam, which is a bactericidal acetolactate synthase (ALS) inhibitor that blocks the biosynthetic conversion of pyruvate to acetolactate, thus abolishing the synthesis of leucine and valine from pyruvate. Both the unlabeled pyruvate and creatine phosphate ATP regeneration systems were used. Both resulted in stereoselective labelling with similar labelling efficiency, highlighting the absence of any significant isotopic dilution by the addition of pyruvate at natural isotopic abundance. Figure 3.5 shows that the prochiral *S*-methyl groups of ubiquitin were stereoselectively labeled as expected.

Although the ALS inhibitor did not entirely prevent the incorporation of unlabeled valine and leucine, presumably due to the unlabeled amino acids already present in the eCells prior to protein production, the isotope labelling efficiency nevertheless reached 70%. Importantly, the eCell system enabled production of this selectively ^{13}C-labeled sample from less than 6 mg methyl-acetolactate precursor, and no ^{13}C labelling of pro-*R* methyl groups was detectable. The effectiveness of the ALS inhibitor in preventing the production of unlabeled

valine and leucine was confirmed by comparison with the isotope labelling efficiency when the CFPS was performed using the widely used ATP-regeneration system with creatine phosphate and creatine kinase (Kigawa et al., 1999; Apponyi et al., 2008). The same isotope labelling efficiency was obtained and the same protein yield (0.7 mg).

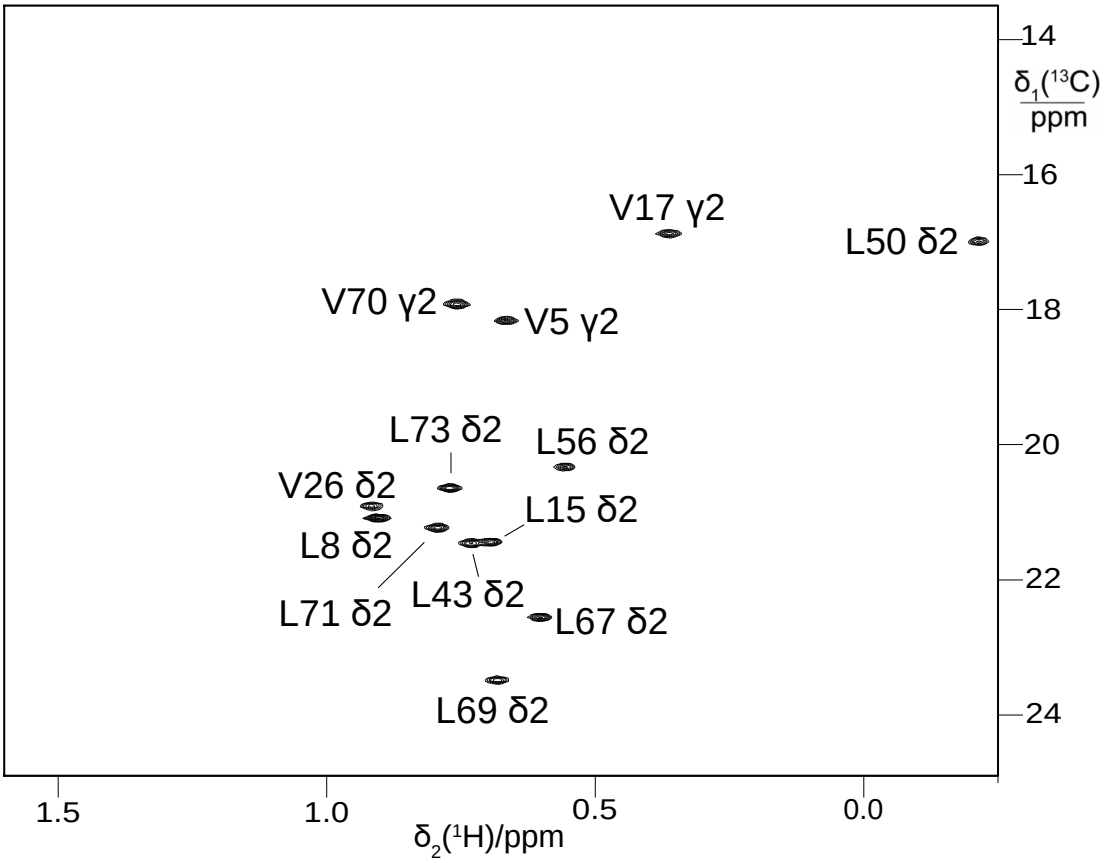
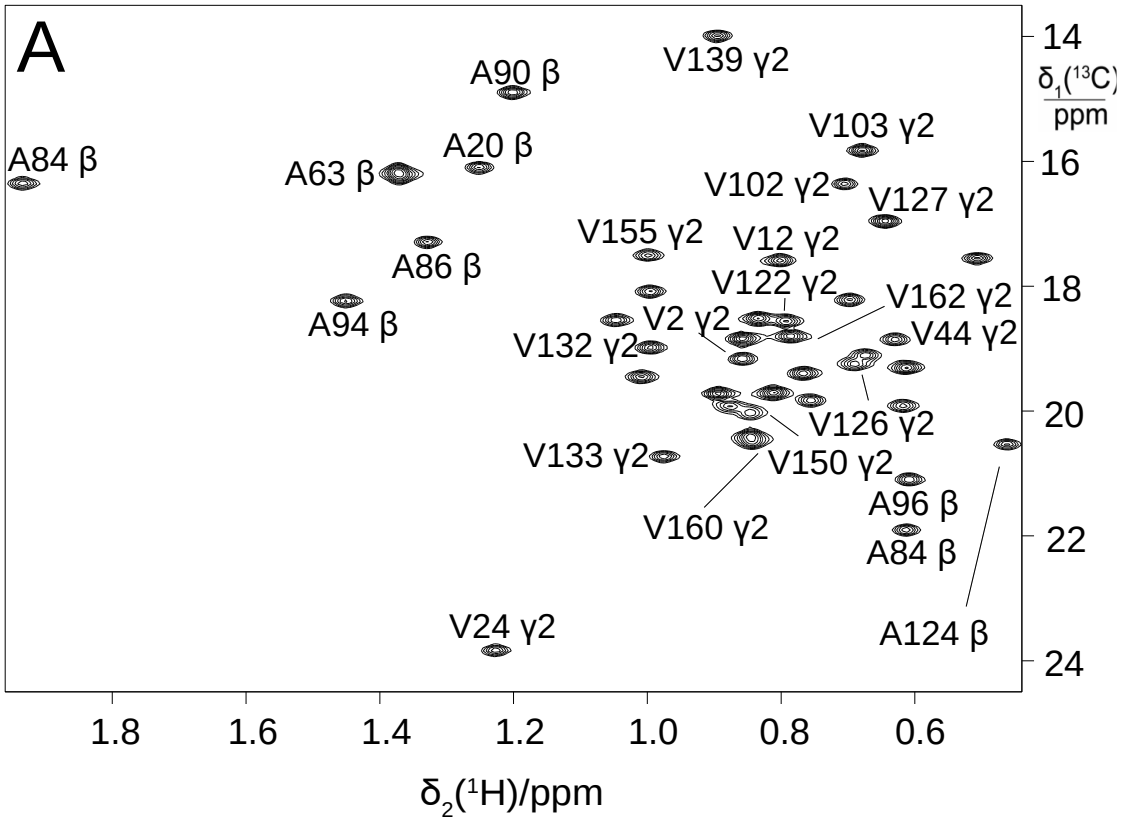


Figure 3.5) ^{13}C-HSQC spectrum of ubiquitin with labelling of the pro-*S* methyl groups in leucine and valine by using site-specifically ^{13}C-labeled acetolactate in eCell CFPS. Protein yield 0.7 mg, isotope labelling efficiency 70%.

3.43 PpiB with stereospecific ^{13}C-labeled methyl groups in valine

To illustrate the broad applicability of the eCell approach to produce perdeuterated proteins, it was also applied to the *E. coli* peptidyl-prolyl *cis–trans* isomerase B (PpiB), which is a 19 kDa protein. Figure 3.6A shows the ^{13}C-HSQC cross-peaks of PpiB prepared with 3-^{13}C-pyruvate while omitting valine. Although the methyl groups of alanine residues are also observed, no two cross-peaks overlap to the extent that they cannot be recognized as separate cross-peaks. Figure 3.6B shows the ^{13}C-HSQC cross-peaks of perdeuterated PpiB made by eCell CFPS using perdeuterated eCells, pyruvate deuterated by H-D exchange and 2-^{13}C-methyl-4-$^{2}H_3$-acetolactate. All amino acids were provided in perdeuterated form and valine was omitted. This resulted in stereoselective labelling of the pro-*S* groups of valine residues in PpiB with a high labelling efficiency (ca. 90%) and adequate yield (1.32 mg). The deuteration level of the protein was high, as shown by a 1D ^{1}H-NMR spectrum. (Appendix figure 5)


A
A90 β
V139 γ2
A84 β
A20 β
A63 β
V103 γ2
V102 γ2
V127 γ2
V155 γ2
V12 γ2
A86 β
V122 γ2
V162 γ2
A94 β
V2 γ2
V44 γ2
V132 γ2
V126 γ2
V133 γ2
V150 γ2
A96 β
V160 γ2
A84 β
V24 γ2
A124 β
14
16
18
20
22
24
δ1(13C) ppm
1.8
1.6
1.4
1.2
1.0
0.8
0.6
δ2(1H)/ppm

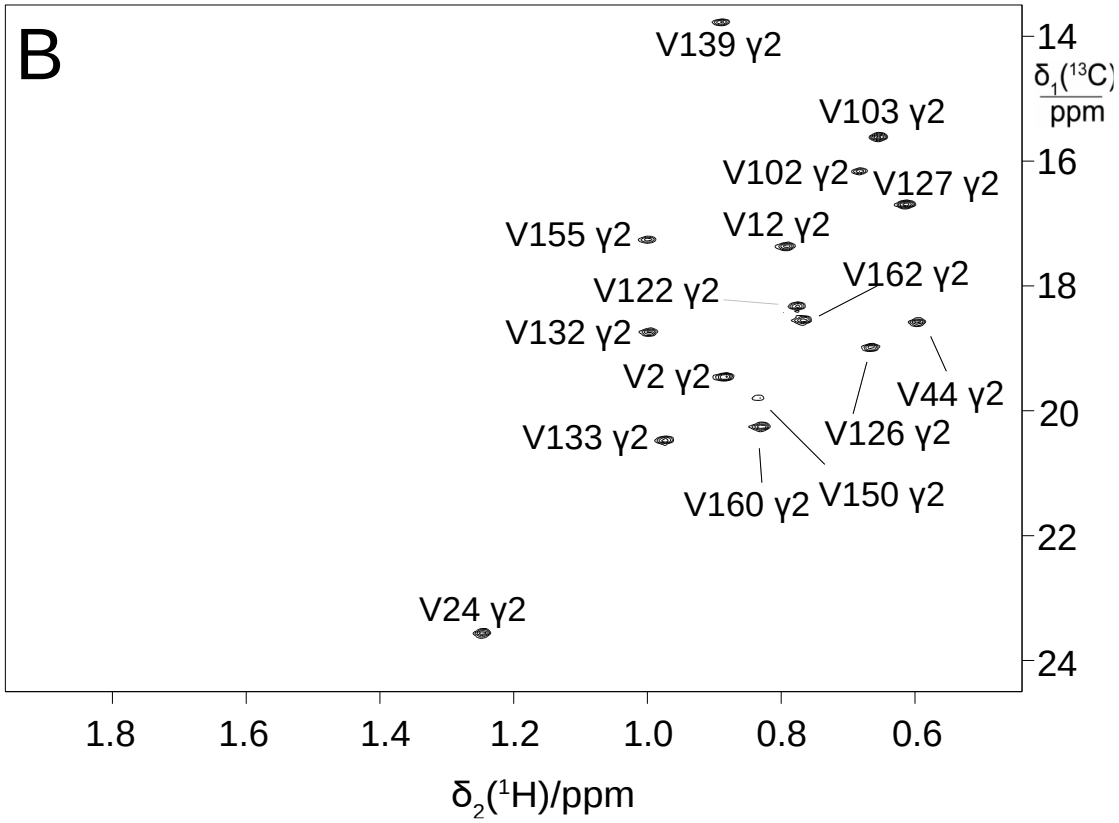

B
V139 γ2
V103 γ2
V102 γ2
V127 γ2
V155 γ2
V12 γ2
V122 γ2
V162 γ2
V132 γ2
V2 γ2
V44 γ2
V126 γ2
V133 γ2
V160 γ2
V150 γ2
V24 γ2
14
16
18
20
22
24
δ1(13C) ppm
1.8
1.6
1.4
1.2
1.0
0.8
0.6
δ2(1H)/ppm

Figure 3.6) Selective ^{13}C-labelling of the methyl groups of alanine and valine residues in PpiB produced by eCell CFPS. **A.)** ^{13}C-HSQC spectrum of PpiB produced from 3-^{13}C-pyruvate with valine omitted. Published assignments are shown (BMRB file 11451). The spectrum also displays the cross-peaks of the γ_1-methyl groups, but their assignments have not been reported. Protein yield 2.2 mg, isotope labelling efficiency >75%. **B.)** ^{13}C-HSQC spectrum of PpiB produced from 2-^{13}C-methyl-acetolactate by CFPS with valine omitted in an eCell CFPS reaction in D_2O using deuterated eCells. The ^{13}C-HSQC spectrum illustrates the selective labelling of the pro-*S*-methyl groups of valine in a perdeuterated protein. The protein yield was 1.3 mg, and the ^{13}C-labelling level was 90%.

3.44 eCell CFPS for stereospecific assignments by biosynthetically directed fractional ^{13}C-labelling

Biosynthetic fractional ^{13}C-labelling is a well-established approach to obtain stereospecific assignments of isopropyl methyl groups (Senn et al., 1989; Neri et al., 1989). Starting from a mixture of 10% uniformly ^{13}C-labeled glucose and 90% glucose at natural isotopic abundance, the ^{13}C-NMR spectrum of pro-*R* methyl groups displays splittings due to $^1J_{CC}$ couplings while the pro-*S* methyl groups do not. The approach is inexpensive as only little isotope-labeled glucose is needed. To explore whether eCells maintain the required biosynthetic pathway, a sample of ubiquitin was prepared from a mixture of ^{13}C-labeled and unlabeled glucose. The ^{13}C-HSQC spectrum showed the multiplet finestructures expected for the pro-*R* and pro-*S* methyl groups (Figure 3.7).

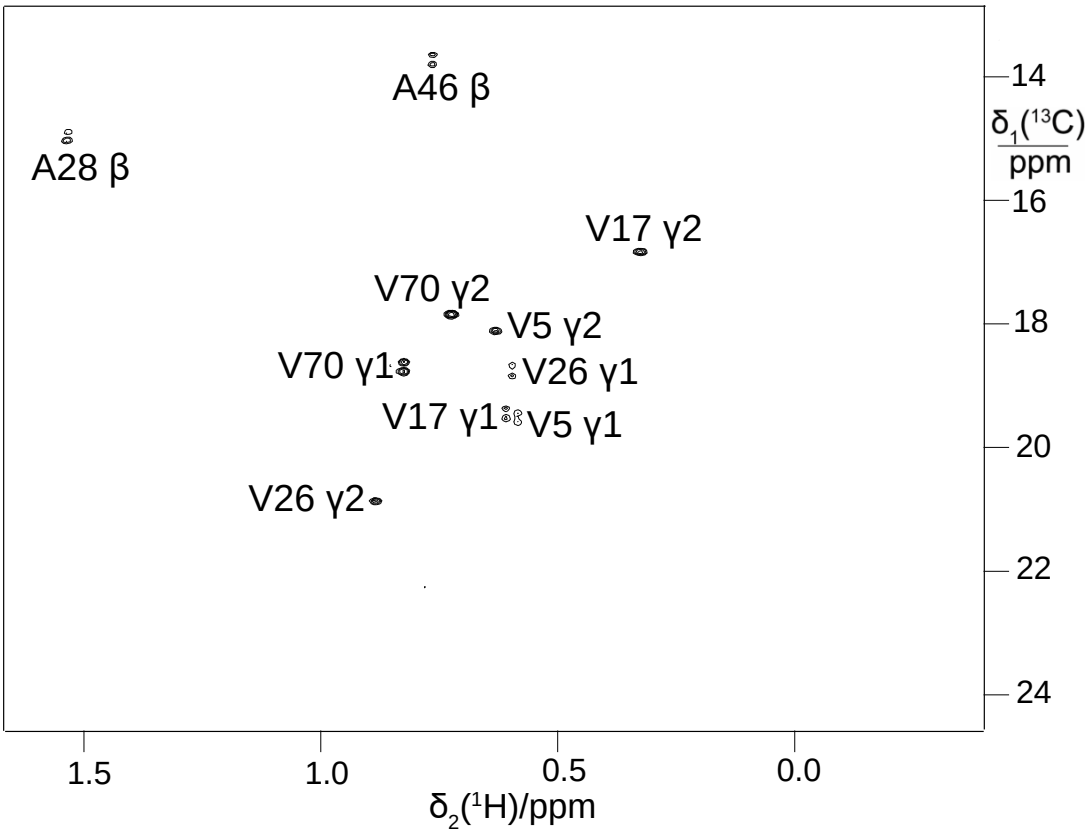


Figure 3.7) ^{13}C-HSQC spectrum of ubiquitin produced by eCell CFPS from a mixture of 10% uniformly ^{13}C-labeled glucose and 90% glucose at natural isotopic abundance. Protein yield 3.8 mg, labelling efficiency below 10%.

3.5 Discussion

Biosynthetic pathways naturally established in bacterial cells can be exploited to produce selectively ^{13}C-labeled proteins in a cell-free reaction that delivers protein in yields sufficient for NMR analysis. In contrast to conventional CFPS reactions, which are based on cell extracts prepared by high-speed centrifugation, the eCells used in the present work much better preserve the activities of the natural complement of biosynthetic enzymes of the parent live *E. coli* cells. In this way, eCells combine intact biosynthetic pathways with some of the fundamental advantages of conventional CFPS, namely the low requirement of amino acids (Torizawa et al., 2004), greater likelihood of compatibility with toxic proteins that do not disrupt transcription/translation and facile modification of the solution conditions with regard to compounds small enough to enter the eCells. Importantly, eCells can be produced rapidly and easily, but require preparation of new batches when different vectors are required. Once

prepared, they can be stored at -80 ºC for years without loss of activity. Table 3.1 highlights the experimental costing for isotope-labeled precursors used in single experiments of the present work.

Table 3.1) Comparison of precursors and their contribution to the cost of eCell CFPS reaction with 300 mg eCells.

^{13}C-labelled precursor	Precursor cost (USD)	Cost of precursor for one reaction (USD)	Total protein yield (mg/mL)	Labelling degree	Position labelled
2-^{13}C-methyl-4-$^{2}H_3$-acetolactate	\$1722/g	\$14	0.7	90%	V= γ_2 L= δ_2
3-^{13}C-pyruvate	\$866/g	\$34	0.8	70%	V= γ_2, γ_1 L= δ_2, δ_1 I=γ_2
10% [U-^{13}C]-glucose + 90% unlabelled glucose	\$258/g	\$2	3.8	10%	V= γ_2^3
1-^{13}C-glucose	\$282/g	\$14	1.4	44%	V= γ_2, γ_1 L= δ_2, δ_1

[1] Prices from Cambridge Isotope Laboratories (https://www.isotope.com), Omicron Biochemicals Inc. (https://www.omicronbio.com) and Apollo Scientific accessed 3rd April 2023.

The present work demonstrates the use of eCells for selective, and also stereospecific, ^{13}C-labelling of the methyl groups of leucine and valine. ^{13}C-labeled methyl groups are privileged for NMR investigations of large proteins, as their magnetization relaxes more slowly than the magnetization of any other amino acid moiety (Tugarinov and Kay, 2006). Resolving the spectral overlap between the ^{13}C-HSQC cross-peaks of different methyl groups, however, is not straightforward, as the ^{13}C-NMR spectra of proteins produced from uniformly ^{13}C-enriched precursors feature multiplet splittings governed by the large $^{1}J_{CC}$ coupling constant. In principle, the splitting could be avoided if the protein were made from suitably isotope-labeled amino acids, but leucine and valine with stereospecific single ^{13}C-enrichment of only one of the isopropyl methyl groups are prohibitively expensive or demanding to prepare (Linser et al., 2014).

As stereospecific ^{13}C-labelling achieved by *in vivo* protein expression with 2-^{13}C-methyl-4-acetolactate (Gans et al., 2010) has become very popular, this compound has become available commercially. (We found the deuterated isotopologue 2-^{13}C-methyl-4-$^{2}H_3$-acetolactate to be more readily available than the undeuterated analogue, but the selective ^{13}C-labelling strategy would be beneficial also without deuteration.) Our results show that this compound can be used in protein production by eCell CFPS, and the resulting ^{13}C-HSQC spectra clearly identify the pro-*S* methyl groups. Alternatively, eCell CFPS also allows the stereospecific distinction of the isopropyl methyl groups by the classical method of biosynthetic fractional ^{13}C-labelling that uses an inexpensive mixture of uniformly ^{13}C-labeled glucose with an excess of glucose at natural isotopic abundance (Neri et al., 1989). Although this scheme allows stereospecific assignments at extraordinarily low cost as far as ^{13}C-labeled glucose is concerned, the level of isotope labelling associated with this scheme is intrinsically low, and we therefore prefer 2-^{13}C-methyl-4-acetolactate for stereospecific assignments, which also minimizes cross-peak overlap by avoiding $^{1}J_{CC}$ multiplet splittings.

The present study also outlines a protocol for the cell-free synthesis of perdeuterated proteins utilizing eCells generated from deuterated media. This approach has some benefit as it is practical and economical. Additional versatility arises from the straightforward way, in which spectral simplification can be obtained by the provision of deuterated amino acids. The high levels of ^{13}C incorporation (90%) and deuteration (estimated to be >95%) are comparable with *in vivo* protein preparations and favourable for good sensitivity of NMR experiments of large protein complexes (Appendix figure 4) (O'Brien et al., 2018).

Pyruvate plays a central role in bacterial biosynthesis and, as shown in the present work, singly ^{13}C-labeled pyruvate is suitable as a relatively inexpensive precursor for labelling methyl groups of leucine and valine with high levels of ^{13}C-enrichment. If, at the same time, unlabeled leucine or valine is provided in the CFPS reaction to suppress their respective cross-peaks, the cross-peaks of the amino acid omitted can be observed selectively. The increased spectral resolution afforded by this scheme is particularly beneficial for larger proteins. Furthermore, inactivation of transaminases by reduction with $NaBH_4$ (Su et al., 2011) may allow extending this approach to the selective ^{15}N-labelling of amino acids from ^{15}N-ammonium salt. These experiments are currently in progress.

In principle, using 1-^{13}C-glucose as the carbon source delivers the same selectivity of isotope labelling as 3-^{13}C-pyruvate (Lundström et al., 2007) but, as glycolysis breaks the glucose down into 3-^{13}C-pyruvate and unlabeled pyruvate, glucose simultaneously labeled in the 1 and 6 position is required to avoid the dilution with unlabeled pyruvate (Loquet et al., 2011). We therefore prefer 3-^{13}C-pyruvate.

As pyruvate can be converted to alanine by a single enzyme, it is difficult to suppress the cross-peaks of the $C^{\beta}H_3$ groups of alanine when starting from ^{13}C-labeled pyruvate. The addition of an excess of unlabeled alanine to the reaction would dilute the labeled pyruvate with unlabeled pyruvate, and inhibition of the alanine aminotransferase by reduction with $NaBH_4$ would also inhibit the transaminase that installs the amino group on leucine and valine by transfer from glutamate. We therefore propose to identify the alanine cross-peaks with a sample, where the isotope labelling of leucine and valine is suppressed by provision of these amino acids in unlabeled form (Figure 3.4B).

Starting from pyruvate, we found it difficult to achieve labelling efficiencies much above 70%. We attribute this to an isotope dilution effect due to a pool of unlabeled amino acids present in the eCells. Attempts to dialyse eCells in a large volume of buffer for an extended period of time reduced the protein yield as the eCells lose activity by gradually leaking biomacromolecules (Van Raad and Huber, 2021). Notably, proteins produced *in vivo* from various ^{13}C-labeled glucose isotopomers are likewise subject to isotopic dilution, and examples with ~45% labelling efficiency have been reported (Lundström et al., 2009; Loquet et al., 2011; Weininger, 2017).

As proteins slowly leak through the porous polymer coating, the lifetime of eCells is limited to about 8 h at 37 °C, which limits protein yields. We note, however, that it is recommended not to conduct *in vivo* protein expression from selectively labelled precursors for too long either, to avoid isotope scrambling by precursor recycling as stated in (Kurauskas et al., 2017).

One key distinction lies in the capacity of eCells to maintain small reaction volumes, similar to CFPS reactions. This characteristic brings about a notable reduction in precursor cost, which is a crucial practical advantage. Additionally, eCells offer the capability to utilize inhibitors for selectively modulating amino acid biosynthesis pathways. This precise control

over carbon sources enables the specific placement of carbon within the amino acid structure. In contrast, standard CFPS lysates lack the comprehensive array of proteins involved in amino acid biosynthesis, which is a defining characteristic of eCells. As a result, eCells provide a unique platform at the intersection of in-vivo and in-vitro protein expression, facilitating the selective enrichment of ^{13}C labels within the anabolism of amino acids for protein expression.

While the eCell system offers advantages for selective labeling and NMR studies, there are still some discrepancies between the labeling patterns and efficiency compared to in vivo labeling strategies. Specifically, it remains to be determined whether the supplementation of methyl bearing amino acids within *in vivo* experiments effectively suppresses ^{13}C methyl cross peaks, especially when perdeuterating proteins (Rowlinson et al., 2022). Similarly, it is essential to explore whether supplementation of amino acid biosynthesis inhibitors can selectively modulate biosynthesis pathways in-vivo. As within eCells this allows for precise control over the carbon source, facilitating the label to be funnelled into the desired position of the amino acid structure.

3.6 Conclusions

In summary, the eCell platform opens new possibilities for the selective ^{13}C-enrichment of methyl groups in proteins. It combines high levels of isotope enrichment with low cost of isotope-enriched precursors. The protocol for eCell preparation is uncomplicated, and the open nature of eCells provides exquisite control over the chemical environment, so that different isotope labelling is achieved simply by use of different reaction buffers. In contrast to conventional CFPS based on dialysis systems, where protein yields depend on good contact between inside and outside buffers and, therefore, the geometry of the setup, the eCell system can readily be scaled in volume. We anticipate that it will find many more uses beyond those demonstrated in the present work.

3.7 References

1. Apponyi, M. A., Ozawa, K., Dixon, N. E. & Otting, G. (2008). Cell-free protein synthesis for analysis by NMR spectroscopy. Methods. Mol. Biol., 426, 257-268.
2. Amero, C., Asunción Durá, M., Noirclerc-Savoye, M., Perollier, A., Gallet, B., Plevin, M. J., Vernet, T., Franzetti, B. & Boisbouvier, J. (2011). A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. J. Biomol. NMR, 50, 229-236.
3. Behera, S. P., Dubey, A., Chen, W. N., De Paula, V. S., Zhang, M., Sgourakis, N. G., Bermel, W., Wagner, G., Coote, P. W. & Arthanari, H. (2020). Nearest-neighbor NMR spectroscopy: categorizing spectral peaks by their adjacent nuclei. Nat. Commun., 11, 5547.
4. Boswell, Z. K. & Latham, M. P. (2018). Methyl-Based NMR Spectroscopy Methods for uncovering structural dynamics in large proteins and protein complexes. Biochemistry, 58, 144-155.
5. Caschera, F. & Noireaux, V. (2015). A cost-effective polyphosphate-based metabolism fuels an all *E. coli* cell-free expression system. Metabol. Engin., 27, 29-37.
6. Filipp, F. V., Sinha, N., Jairam, L., Bradley, J. & Opella, S. J. (2009). Labeling strategies for ^{13}C-detected aligned-sample solid-state NMR of proteins. J. Magn. Reson., 201, 121-130.
7. Gans, P., Hamelin, O., Sounier, R., Ayala, I., Durá, M. A., Amero, C. D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M. J. & Boisbouvier, J. (2010). Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. Angew. Chem. Int. Ed., 49, 1958-1962.
8. Goto, N. K., Gardner, K. H. & Mueller, G. A. (1999). A robust and cost-effective method for the production of Val, Leu, Ile methyl-protonated ^{15}N, ^{13}C, ^{2}H-labeled proteins ligand-induced folding of a two-component signaling receiver domain. J. Biomol. NMR., 13, 369-374.
9. Hajduk, P. J., Augeri, D. J., Mack, J., Mendoza, R., Yang, J., Betz, S. F. & Fesik, S. W. (2000). NMR-based screening of proteins containing ^{13}C-labeled methyl groups. J. Am. Chem. Soc., 122, 7898-7904.
10. Jewett, M. C. & Swartz, J. R. (2004). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol. Bioeng., 86, 19-26.

11. Jewett, M. C., Calhoun, K. A., Voloshin, A., Wu, J. J. & Swartz, J. R. (2008). An integrated cell-free metabolic platform for protein production and synthetic biology. Mol. Sys. Biol., 4, 220.
12. Jia, X., Ozawa, K., Loscha, K. & Otting, G. (2009). Glutarate and N-acetyl-L-glutamate buffers for cell-free synthesis of selectively ^{15}N-labeled proteins. J. Biomol. NMR, 44, 59-67.
13. Kasinath, V., Valentine, K. G. & Wand, A. J. (2013). A ^{13}C labeling strategy reveals a range of aromatic side chain motion in calmodulin. J. Am. Chem. Soc., 135, 9560-9563.
14. Kim, D. M. & Swartz, J. R. (2001). Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. Biotechnol. Bioeng., 74, 309-316.
15. Kurauskas, V., Schanda, P. & Sounier, R. (2017). Methyl-specific isotope labelling strategies for NMR studies of membrane proteins. Meth. Mol. Biol., 1635, 109-123.
16. Lange, O. F., Rossi, P., Sgourakis, N. G., Song, Y., Lee, H. W., Aramini, J. M., Ertekin, A., Xiao, R., Acton, T. B., Montelione, G. T. & Baker, D. (2012). Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. Proc. Natl. Acad. Sci., 109, 10873–10878.
17. Lazarova, M., Löhr, F., Rues, R.-B., Kleebach, R., Dötsch, V. & Bernhard, F. (2018). Precursor-based selective methyl labeling of cell-free synthesized proteins. ACS Chem. Biol., 13, 2170-2178.
18. Linser, R., Gelev, V., Hagn, F., Arthanari, H., Hyberts, S. G. & Wagner, G. (2014). Selective methyl labeling of eukaryotic membrane proteins using cell-free expression. J. Am. Chem. Soc., 136, 11308-11310.
19. Loquet, A., Lv, G., Giller, K., Becker, S. & Lange, A. (2011). ^{13}C spin dilution for simplified and complete solid-state NMR resonance assignment of insoluble biological assemblies. J. Am. Chem. Soc., 133, 4722-4725.
20. Lundström, P., Kaare, T., Carstensen, T., Bezsonova, I., Wiesner, S., Hansen, D. F., Religa, L. R., Akke, M. & Kay, E. L. (2007). Fractional ^{13}C enrichment of isolated carbons using [1-^{13}C]-or [2-^{13}C]-glucose facilitates the accurate measurement of dynamics at backbone Cα and side-chain methyl positions in proteins, J. Biomol. NMR., 38, 199–212.
21. Lundström, P., Lin, H. & Kay, L. E. (2009) Measuring ^{13}C chemical shifts of invisible excited states in proteins by relaxation dispersion NMR spectroscopy, J. Biomol. NMR., 44, 139–155.

22. Neri, D., Szyperski, T., Otting, G., Senn, H. & Wüthrich, K. (1989). Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. Biochemistry., 28, 7510-7516.
23. Martemyanov, K. A., Shirokov, V. A., Kurnasov, O. V., Gudkov, A. T. & Spirin, A. S. (2001). Cell-free production of biologically active polypeptides: application to the synthesis of antibacterial peptide cecropin. Prot. Exp. Purif., 21, 456-461.
24. Otten, R., Chu, B., Krewulak, K. D., Vogel, H. J. & Mulder, F. A. A. (2010). Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins. J. Am. Chem. Soc., 132, 2952-2956.
25. Schörghuber, J., Geist, L., Platzer, G., Feichtinger, M., Bisaccia, M., Scheibelberger, L., Weber, F., Konrat, R., & Lichtenecker, R. J. (2018). Late metabolic precursors for selective aromatic residue labeling. J. Biomol. NMR., 71, 129-140.
26. Schubert, M., Manolikas, T., Rogowski, M. & Meier, B. H. (2006). Solid-state NMR spectroscopy of 10% ^{13}C labeled ubiquitin: spectral simplification and stereospecific assignment of isopropyl groups. J. Biomol. NMR., 35, 167-173.
27. Senn, H., Werner, B., Messerle, B. A., Weber, C., Traber, R. & Wüthrich, K. (1989). Stereospecific assignment of the methyl ^{1}H NMR lines of valine and leucine in polypeptides by nonrandom ^{13}C labelling. FEBS Lett., 249, 113-118.
28. Sprangers, R. & Kay, L. E. (2007). Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature., 445, 618-622.
29. Su, X. C., Loh, C. T., Qi, R. & Otting, G. (2011). Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enzymes for selective ^{15}N-labelling and production of perdeuterated proteins in H_2O. J. Biomol. NMR., 50, 35-42.
30. Rasia, R. M., Brutscher, B. & Plevin, M. J. (2012). Selective isotopic unlabeling of proteins using metabolic precursors: application to NMR assignment of intrinsically disordered proteins. ChemBioChem., 13, 732-739.
31. Takeda, M., Ono, A. M., Terauchi, T. & Kainosho, M. (2010). Application of SAIL phenylalanine and tyrosine with alternative isotope-labeling patterns for protein structure determination. J. Biomol. NMR., 46, 45-49.
32. Torizawa, T., Shimizu, M., Taoka, M., Miyano, H. & Kainosho, M. (2004). Efficient production of isotopically labeled proteins by cell-free synthesis: a practical protocol. J. Biomol. NMR., 30, 311-325.

33. Tugarinov, V., Kanelis, V. & Kay, L. E. (2006). Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. Nat. Protoc., 1, 749-754.
34. Van Raad, D. & Huber, T. (2021). *In vitro* protein synthesis in semipermeable artificial cells. ACS Synth. Biol., 10, 1237-1244.
35. Vuister, G. W. & Bax, A. (1992). Resolution enhancement and spectral editing of uniformly 13C-enriched proteins by homonuclear broadband ^{13}C decoupling. J. Magn. Reson., 98, 428-435.
36. Whittaker, J. W. (2007). Selective isotopic labeling of recombinant proteins using amino acid auxotroph strains. Meth. Mol. Biol., 389, 175-187.
37. Weininger, U. (2017) Site-selective -C labeling of proteins using erythrose.. J. Biomol. NMR., 67, 191-200.

CHAPTER FOUR

Improved spectral resolution of [^{13}C,^{1}H]-HSQC spectra of aromatic amino acid residues in proteins produced by cell-free synthesis from inexpensive ^{13}C-labeled precursors

4.1 Abstract

A novel isotope labelling method of aromatic residues is described that relies on the active enzyme metabolism in eCells for efficient conversion of inexpensive ^{13}C-labeled precursors into amino acid isotopologues free of one-bond ^{13}C–^{13}C couplings. Selective ^{13}C-labelling of tyrosine and phenylalanine residues is achieved simply by using different compositions of the reaction buffers.

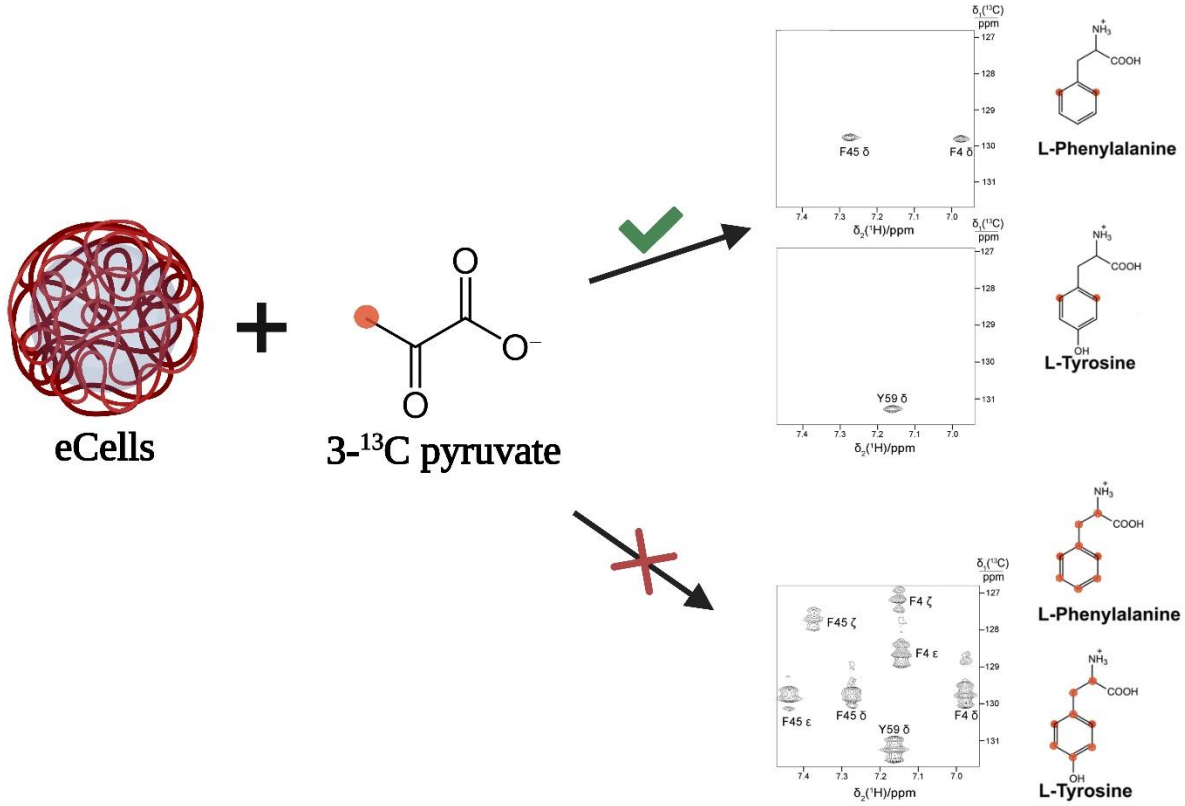


Figure 4.1) Overview of the biosynthesis of aromatic amino acids, specifically tyrosine and phenylalanine, starting from 3-^{13}C pyruvate and utilizing eCells. The eCells can produce proteins containing aromatic amino acids that are specifically labeled with ^{13}C in the δ position. This labeling strategy results in the appearance of single sharp peaks in the ^{13}C

HSQC spectra, reducing spectral complexity and minimizing potential overlap between ^{13}C-^{13}C signals. The ^{13}C-enriched positions are represented by red balls in the diagram.

4.2 Introduction

The NMR spectroscopic analysis of proteins of increasing molecular weight is hindered by cross-peak overlap. Selective isotope-labelling presents an elegant way for improving the spectral resolution (Verardi et al., 2012) with site-selective ^{13}C and ^{2}H labelling delivering greatly simplified protein NMR spectra, which benefit the analysis of large proteins (Takeuchi et al., 2007; Kainosho and Güntert, 2009; Takeda et al., 2010). As a drawback, the cost of sample preparation with isotope-labeled amino acids can be high (Gell et al., 2017). The present work presents a new strategy for selective ^{13}C-labelling of aromatic amino acids that uses inexpensive ^{13}C-labeled precursors to direct ^{13}C labels into specific positions of the aromatic side chains of phenylalanine (Phe), tryptophan (Trp) and tyrosine (Tyr) in a cell-free protein synthesis (CFPS) system.

Due to large one-bond ^{13}C–^{13}C couplings ($^{1}J_{CC}$) and limited chemical shift dispersion between the δ and ε CH groups of Tyr and δ, ε and ζ CH groups of Phe (Williams et al., 2015), the [^{13}C,^{1}H]-HSQC spectra of the aromatic rings of Phe, Trp and Tyr residues are prone to spectral overlap, when the proteins are produced with uniform ^{13}C-labelling (Torizawa et al., 2005; Milbradt et al., 2015). $^{1}J_{CC}$ couplings also hinder relaxation measurements of aromatic side chains to probe for conformational changes (Ban et al., 2017). This situation has triggered the development of various atom-specific labelling methods to reduce the complexity of ^{13}C-NMR spectra by eliminating $^{1}J_{CC}$ couplings (Törner et al., 2020; Schörghuber et al., 2018). As the chemical synthesis of suitably ^{13}C-labeled amino acids is expensive, alternative approaches have been explored that harness the anabolic pathways of amino acid synthesis in bacteria to produce specifically labeled protein from relatively inexpensive labeled precursors (Whittaker, 2007). To direct the anabolic pathways, the bacterial cells are grown in a minimal medium, where the carbon source is enriched with ^{13}C in specific positions, and the protein is produced *in vivo* (Gans et al., 2010; Kasinath et al., 2013; Lichtenecker et al., 2013). This labelling strategy is prone to isotope scrambling, which has led to protocols that limit the scrambling by the co-addition of uniformly 2H-labeled precursors and short periods of bacterial cell growth after induction to prevent precursor recycling (Kurauskas et al., 2017). In an extension of this approach, selected amino acids can

be “unlabeled” by their provision in the growth medium in their standard form, i.e., at natural isotopic abundance (Rasia et al., 2012).

To obtain the protein quantities required for NMR spectroscopy, *in vivo* protein expression requires significant amounts of the isotope-labeled amino acids or precursors, and it has been shown that modern cell-free protein synthesis (CFPS) systems incorporate amino acids in much higher yield (Torizawa et al., 2004). Unfortunately, many of the bacterial enzymes are rendered nonfunctional during the preparation of cell extracts, so that the less expensive isotope labelling approaches that depend on bacterial anabolic pathways are inaccessible to CFPS (Linser et al., 2014). As replenishing the cell-free extract with the requisite enzymes would be costly and unpractical, we explored the use of the recently developed eCell system for protein production under CFPS conditions. eCells are bacterial cells coated with layers of differently charged polyionic polymers and cell walls made porous by the expression of endolysin (Van Raad and Huber, 2021). The pores of eCells leak proteins only very slowly over a period of hours, whereas low-molecular weight compounds readily exchange between intra- and extracellular space. eCells can therefore be compared with traditional CFPS systems that use dialysis membranes separating inner and outer buffer solutions (Apponyi et al., 2008). Importantly, the mild preparation conditions of eCells maintains the activity of bacterial anabolic enzymes much better than cell extracts. The present work shows that the fully functional complement of biosynthetic *E. coli* enzymes required for the synthesis of Phe, Trp and Tyr, as well as the entire repertoire of transcription/translation components, is active in eCells and allows the production of proteins with site-specifically ^{13}C-labeled aromatic amino acids from inexpensive precursors.

The bacterial biosynthetic pathways of aromatic amino acid synthesis (Figure 4.2) show that pyruvate and erythrose are precursors that are relatively close to the final amino acid product. Phosphoenol pyruvate is combined with erythrose phosphate to produce Phe, Trp and Tyr, indicating that selective ^{13}C-labelling of the δ and ε positions of Tyr/Phe and the δ position of Trp can be achieved in eCells by the use of ^{13}C-labeled pyruvate and erythrose, in analogy to the situation *in vivo* (Lundström et al., 2007). This chapter shows that this is indeed the case.

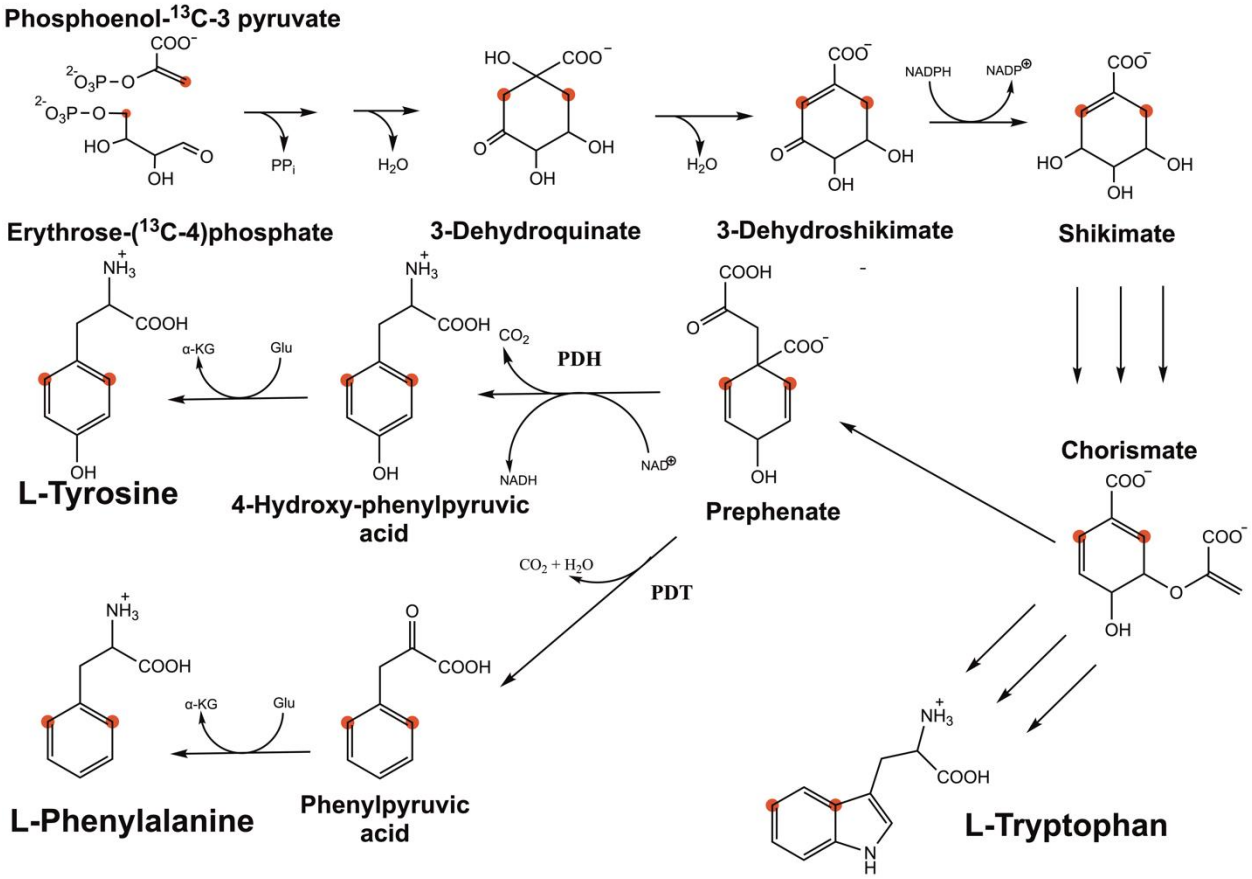


Figure 4.2) Biosynthetic pathway of the aromatic amino acids tyrosine, phenylalanine and tryptophan starting from phosphoenol-pyruvate and erythrose-4-phosphate. The present work used 4-^{13}C erythrose and 3-^{13}C pyruvate in eCells to produce proteins with specifically ^{13}C-labeled aromatic amino acids. Red balls mark the ^{13}C-enriched positions and trace their fate through the pathways.

4.3 Materials and Methods

4.31 Plasmids

The expression vectors used for PpiB and ubiquitin were described previously (Van Raad and Huber, 2021). They afford expression of the target proteins under control of a T7 promoter and *lac* operator and contain a spectinomycin resistance gene.

4.32 eCell production

The production of eCells followed a previously published protocol, using *E. coli* autolysis XJb(DE3)* cells, which have an endolysin gene in the genome under a P_{BAD} promoter (Van Raad et al., 2021). The culture was induced for endolysin production at the time of inoculation with a final concentration of 3 mM arabinose. The cells were grown to OD_{600} 0.6 and washed three times with PBS-E (phosphate-buffered saline containing 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 and 1 mM EDTA) pH 7.4 and resuspended in 0.25 mg/mL of chitosan in PBS-E solution. After vigorous shaking for 20 minutes, the cell pellet was washed with PBS-E pH 6.0 three times to remove excess chitosan and then resuspended in 0.25 mg/mL of alginate PBS-E solution and subjected to vigorous shaking for 20 minutes. The cells were then washed 3 times with PBS-E pH 6.0, resuspended in PBS-E pH 7.4 and directly stored at -80 ^{o}C in PBS-E buffer.

4.33 CFPS protocol

The protocol for eCell CFPS based on glucose and pyruvate was described previously (Van Raad et al., 2023). The protocol was adapted from the phosphate recycling system published by Jewett and co-workers (Jewett & Swartz, 2004). The CFPS buffer contained 0.9 mM UTP and CTP, 50 mM HEPES, 1.5 mM GTP, 1.5 mM ATP, 0.68 μM folinic acid, 0.64 mM cAMP, 1.7 mM DTT, 60 mM KGlu, 8 mM MgGlu, 2% v/v PEG-8000, 5 mM CoA, 33 mM pyruvate, 4 mM sodium oxalate, 0.25 mM CoA and 0.33 mM NAD. 3.5 mM of each amino acid was added except those targeted for isotope enrichment by the eCells. Na_2HPO_4 was omitted to promote the use of glucose for energy generation, and 4 mM sodium oxalate was used to promote the generation of ATP by the PANOx system of Jewett and Swartz (2004). Prior to each CFPS reaction, the buffer was adjusted to pH 7.5. Frozen aliquots of encapsulated cells were thawed and the pellet immersed in CFPS buffer. Except where stated otherwise, 1.5 g eCells produced from 1 L cell culture at OD_{600} 0.6 were suspended in 25 mL of CFPS buffer. CFPS was conducted for 12 h at 37 ^{o}C in a shaker at 180 rpm. The proteins were purified using His-Gravitrap columns (GE Healthcare).

4.34 Labelling with 3-^{13}C-pyruvate, 2-^{13}C glucose or with 1-^{13}C glucose and 4-^{13}C erythrose combined

To produce ubiquitin samples from 3-^{13}C-pyruvate for ^{13}C-labelling of the δ positions of tyrosine and phenylalanine, dry 3-^{13}C-pyruvate was added to 300 mg eCells in 10 mL CFPS buffer at 33 mM final concentration, and Tyr or Phe were omitted from the CFPS buffer.

To test the performance of 2-^{13}C-glucose as carbon source in the aromatic amino acid synthesis, dry 2-^{13}C glucose was added at 30 mM final concentration to 1.5 g eCells suspended in 25 mL CFPS buffer. Tyrosine, phenylalanine and tryptophan were omitted from the amino acid mixture to allow for labelling of the ε positions of tyrosine and phenylalanine and the tryptophan δ positions.

For increased levels of ^{13}C labelling of PpiB, 3-^{13}C-pyruvate was added to CFPS buffers at 33 mM final concentration combined with 4-^{13}C erythrose at 30 mM concentration. Tyrosine and phenylalanine were omitted from the CFPS reaction mixture.

For increased ^{13}C labelling of PpiB using pyruvate and an isotopologue of glucose, 3-^{13}C-pyruvate was added to CFPS buffer at 16.5 mM final concentration combined with 1-^{13}C glucose at 16 mM concentration. Tyrosine and phenylalanine were omitted from the reaction mixture. For all samples, 1 mM IPTG was added to the CFPS reaction mixture.

4.35 NMR spectroscopy and isotope labelling yields

[^{13}C,^{1}H]-HSQC spectra were recorded at 25 ^{o}C using Bruker 600 or 800 MHz NMR spectrometers equipped with TCI cryoprobes.

To determine the degree of isotope enrichment, the [^{13}C,^{1}H]-HSQC cross-peak integrals of the labeled residues were compared with those of an internal standard of 0.1 mM 3-^{13}C-pyruvate in spectra recorded with a recovery delay of 30 s between scans.

4.4 Results

4.41 Selective ^{13}C-labelling of the δ positions of Phe or Tyr of ubiquitin

The biosynthesis of aromatic amino acids under CFPS conditions was explored using eCells made from *E. coli* XJb(DE3)* cells. The eCells contained the expression plasmids for

ubiquitin or the *E. coli* prolyl *cis–trans* peptidyl isomerase B (PpiB). eCell CFPS supplied with 3-^{13}C pyruvate is expected to label a single δ carbon of phenylalanine and tyrosine (Figure 4.1), while for tryptophan, the ^{13}C label is expected to appear in the position of a quaternary carbon which does not produce a [^{13}C,^{1}H]-HSQC cross-peak.

The NMR spectra of ubiquitin confirmed these expectations. Ubiquitin contains two Phe, one Tyr and no Trp residue. As expected, two [^{13}C,^{1}H]-cross-peaks are observed for the sample prepared with Phe omitted and unlabeled Tyr added in the CFPS buffer (Figure 4.3A), whereas a single cross-peak is observed for the sample prepared with Tyr omitted and Phe added (Figure 4.3B). As anticipated, these cross-peaks show no evidence of $^{1}J_{CC}$ splittings, which are prominent in uniformly ^{13}C-labeled samples (Figure 4.3C). 300 mg eCells, 10 mL and 33 mM 3-^{13}C pyruvate were sufficient to produce 1.2 mg Phe $^{13}C^{\delta}$-labeled or 0.8 mg Tyr $^{13}C^{\delta}$-labeled purified ubiquitin with isotope labelling levels of 37% and 40%, respectively. As eCells are made from *E. coli* cells that contain some unlabeled amino acids, a somewhat reduced degree of isotope enrichment is expected as using 3-^{13}C pyruvate allows for a maximum of 50% labelling efficiency. We attempted to increase the level of isotope labelling by extensive dialysis of the eCells in phosphate buffered saline prior to use in CFPS, but this did not significantly increase the level of labelling. Importantly, the presence of unlabeled Tyr or Phe did not affect the selectivity of labelling.

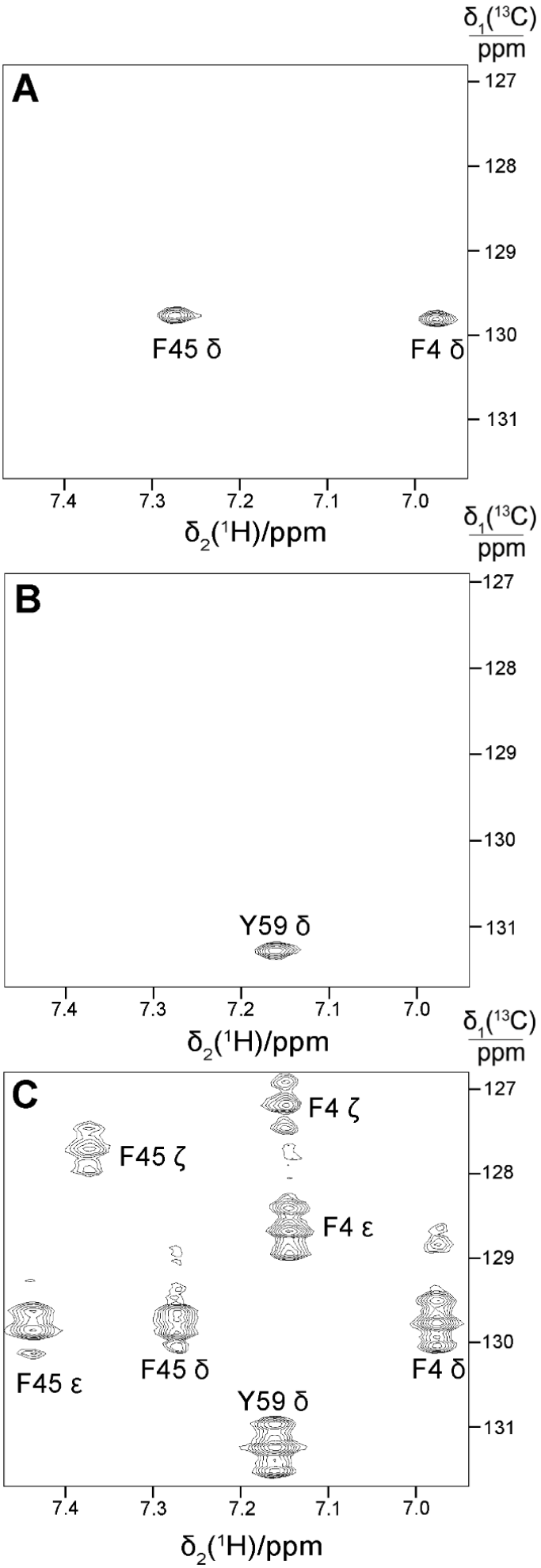

A
$\delta_1(^{13}C)$ ppm
127
128
129
130
131
F45 δ
F4 δ
7.4
7.3
7.2
7.1
7.0
$\delta_2(^1H)$/ppm
B
$\delta_1(^{13}C)$ ppm
127
128
129
130
131
Y59 δ
7.4
7.3
7.2
7.1
7.0
$\delta_2(^1H)$/ppm
C
$\delta_1(^{13}C)$ ppm
127
128
129
130
131
F4 ζ
F45 ζ
F4 ε
F45 ε
F45 δ
F4 δ
Y59 δ
7.4
7.3
7.2
7.1
7.0
$\delta_2(^1H)$/ppm

Figure 4.3) Selective ^{13}C-labelling of the δ positions of phenylalanine or tyrosine of ubiquitin produced by eCell CFPS with 3-^{13}C pyruvate as the carbon source. The plots show selected spectral regions from [^{13}C,^{1}H]-HSQC spectra using different labelling strategies. **A.)** C$^{\delta}$H cross-peaks of the two phenylalanine residues of ubiquitin. The protein was produced in the presence of 3.5 mM unlabeled amino acids, including Tyr but not Phe. Protein yield was 1.2 mg from 300 mg of eCells in 10 mL CFPS buffer. **B.)** C$^{\delta}$H cross-peak of the single tyrosine residue of ubiquitin. The protein was produced in the presence of 3.5 mM unlabeled amino acids, including Phe but not Tyr. Protein yield 0.8 mg. **C.)** Uniformly ^{13}C-labeled sample prepared using uniformly ^{13}C-labeled glucose in a 5 mL eCell CFPS reaction containing 300 mg eCells, omitting Tyr and Phe. The yield was 1.7 mg of purified protein.

4.42 Selective ^{13}C-labelling of the ε positions of Phe, Tyr and δ position of Trp in PpiB

To demonstrate the general applicability of the labelling protocol, a sample of PpiB was made using 33 mM 3-^{13}C pyruvate and 1.5 g eCells. 4.8 mg PpiB was obtained with a labelling level of about 36% (Appendix figure 2). To increase the labelling degree of the C$^{\delta}$ positions, we also prepared a sample using from 16 mM 1-^{13}C glucose and 16.5 mM 3-^{13}C pyruvate. The experiment yielded 2.7 mg protein per gram of eCells with about 50% ^{13}C labelling of the C$^{\delta}$ positions. The [^{13}C,^{1}H]-spectra of the C$^{\delta}$ and C$^{\varepsilon}$ preparation showed resolved cross-peaks for 10 Phe/3 Tyr and 9 Phe/3 Tyr free of multiplet splittings due to $^{1}J_{CC}$ couplings. PpiB contain 12 Phe and 3 Tyr residues, but some of them undergo ring flips in the intermediate time regime, where signals are broadened beyond detection (Takeda et al., 2010).

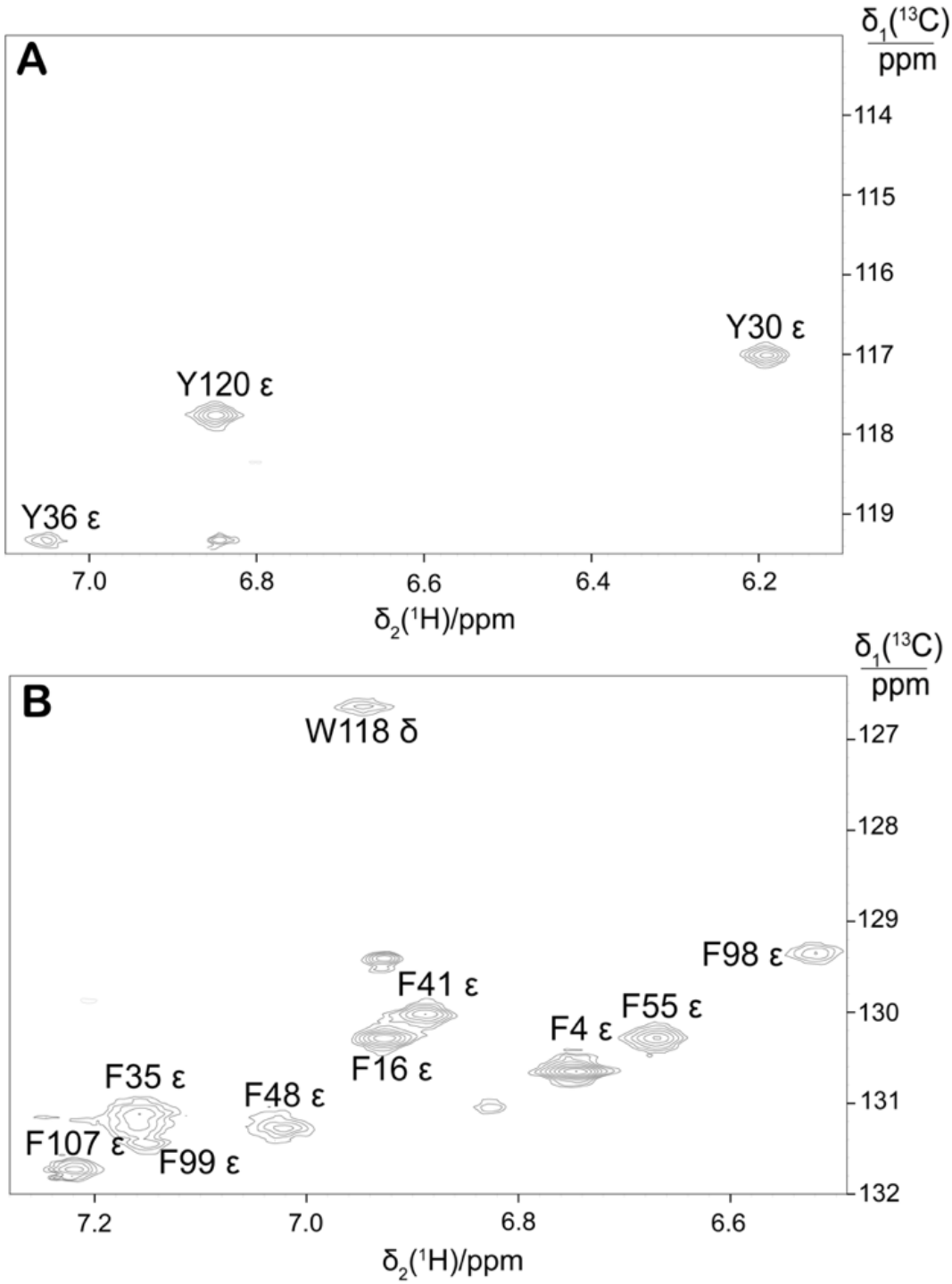


Figure 4.4) Selected spectral regions of the [^{1}H,^{13}C]-HSQC spectrum of PpiB produced by eCell CFPS with 2-^{13}C glucose as the carbon source and with tyrosine, phenylalanine and tryptophan omitted and all other amino acids at 3.5 mM. The protein yield was 4.8 mg and the isotope labelling level about 45%. **A.)** C$^{\varepsilon}$H cross-peaks of tyrosine and C$^{\delta}$H cross-peak of tryptophan. The weak unmarked cross-peak may arise from the C$^{\delta}$H group of His8, as the production of histidine from 2-^{13}C glucose directs ^{13}C to the δ position of the imidazole ring (Weininger, 2017). **B.)** C$^{\varepsilon}$H cross-peaks of phenylalanine. The two weak unmarked cross-

peaks may arise from the $C^{\varepsilon}H$ groups of Phe27 and Phe123, for which no assignments have been reported.

The biosynthetic pathways indicate that the ε carbons of Phe and Tyr and the δ carbon of Trp can be labeled also by using a different isotopologue precursor, 2-^{13}C glucose. Using 1.5 g eCells,g 25 mL CFPS buffer and 30 mM 2-^{13}C glucose yielded 4.8 mg PpiB with >45% labelling degree of the ε carbons of Phe and Tyr. Figure 4.4 shows that this approach likewise avoided splittings by $^{1}J_{CC}$ couplings.

4.43 Selective ^{13}C-labelling of the δ positions of Phe, Tyr in PpiB

While 3-^{13}C pyruvate enriches only one of two C^{δ} positions in the aromatic rings of Phe and Tyr, the biosynthetic pathways (Figure 4.2) suggest that the additional provision of 4-^{13}C erythrose will label both C^{δ} positions and thus increase the sensitivity of the NMR measurements. Using 1.5 g eCells, 33 mM of 3-^{13}C pyruvate and 30 mM of 4-^{13}C erythrose, 1.6 mg PpiB was obtained with a labelling level of 70%. Although the theoretical maximum of 100% labelling was not reached, the improved labelling level enabled the detection of the weak $C^{\delta}H$ cross-peak of Phe110, which has previously been reported to display broad signals due to ring flips in the intermediate time regime (Takeda et al., 2010), and the spectrum showed singlets in the ^{13}C dimension as expected (Figure 4.5).

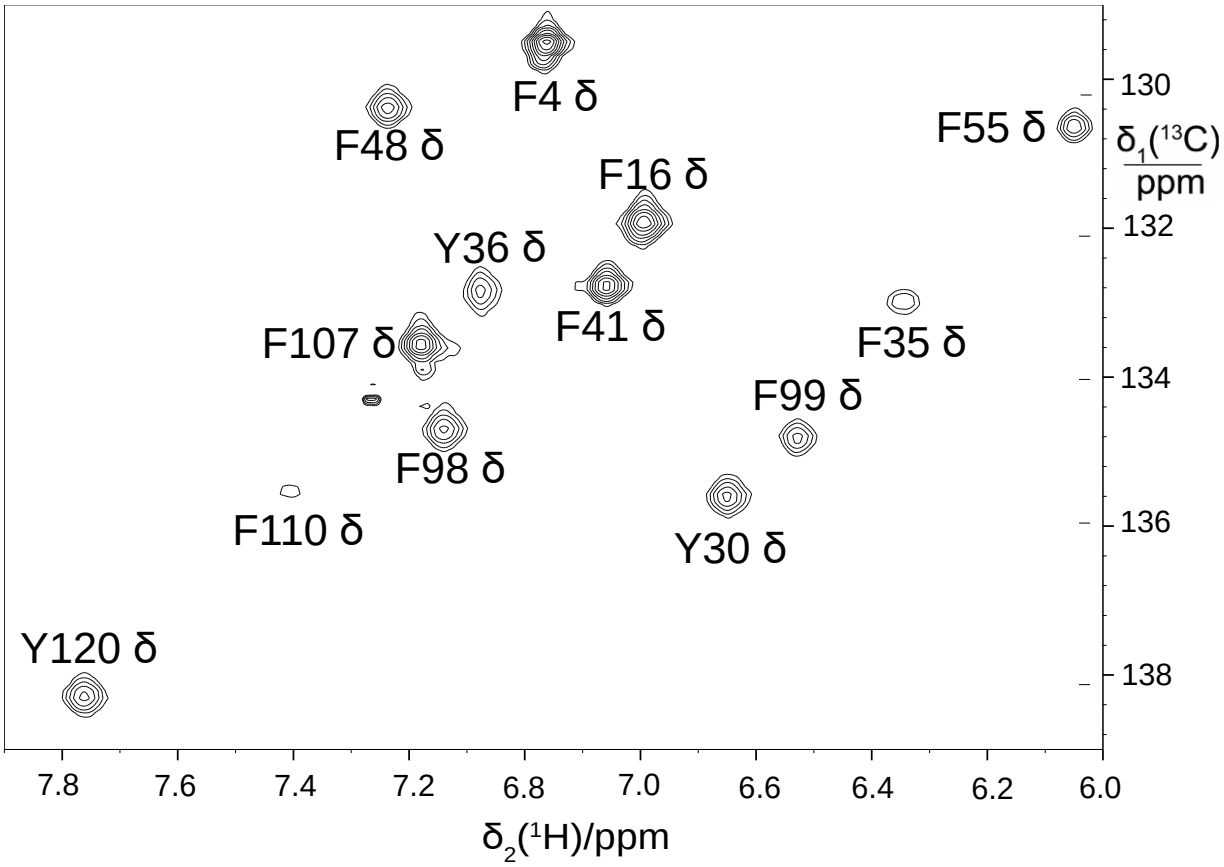


Figure 4.5) Selected spectral region of the [^{1}H,^{13}C]-HSQC spectrum of PpiB produced by eCell CFPS with 3-^{13}C pyruvate and 4-^{13}C erythrose as the carbon sources for Phe and Tyr. The yield was 1.6 mg of purified protein, and the level of isotope labelling was > 70%.

Table 4.1) Comparison of precursors and their contribution to the cost of eCell expression for PpiB expression using 1.5 g eCells in 25 mL CFPS buffer.[1]

Precursor/Energy source	Precursor cost (USD)	Labelling level (%) actual/max.	Protein yield (mg/mL)	Cost per reaction	Position labeled in Phe and Tyr
3-^{13}C pyruvate	\$866/g	36/50	4.8	\$36	δ
3-^{13}C pyruvate + 1-^{13}C glucose	\$866/g + \$282/g	50/100	2.7	\$26	δ
3-^{13}C pyruvate + 4-^{13}C erythrose	\$866/g + \$1340/g	70/100	1.6	\$104	δ
2-^{13}C glucose	\$305/g	46/50	4.8	\$15	ε

[1] Prices from Cambridge Isotope Laboratories (https://www.isotope.com) and Omicron Biochemicals, Inc. (https://www.omicronbio.com), accessed 4 Jan 2023.

4.5 Discussion

Due to their relatively rigid and hydrophobic character, aromatic residues are often overrepresented in ligand binding sites and hydrophobic cores of proteins (Makwana and Mahalakshmi, 2015). In enzymes, aromatic amino acids frequently form important interactions with substrates (Lanzarotti et al., 2011). This makes them excellent probes to assess 3D structure and ligand binding in pharmaceutical development and drug design (Brylinski, 2018). To make aromatic side chains more amenable to monitoring by NMR spectroscopy, the spectral simplification afforded by atom-specific labelling confers an important benefit.

In contrast to protein production *in vivo*, where isotopic scrambling is best suppressed by using late metabolic precursors in isotope labeled form together with auxotrophic bacterial strains, eCells provide a straightforward alternative route to atom-specific isotope labelling of aromatic amino acids starting from inexpensive and more readily available precursors. For example, the cost of isotope-labeled starting material for 4.8 mg PpiB produced from eCells was about 32 USD for 50 mg 2-^{13}C glucose (Table 4.1). The equivalent sample produced in vivo would require much more isotope-labelled compound to satisfy the larger culture volumes involved. Typical amounts of ^{13}C-labelled precursors used for expression in 1 litre of M9 minimal medium are 2.5 g ^{13}C glucose, 2.0 g pyruvate or 1.3 g erythrose (Cai et al., 1998; 2019; Kasinath et al., 2013). The economics of eCell CFPS thus appear competitive to in vivo protein expression in minimal media also when measured in terms of cost of ^{13}C-labelled precursor required per mg of isotope-labelled protein. We anticipate that actual outcomes will depend mostly on the intrinsic yields of the specific protein in eCell CFPS versus in vivo cell cultures. Like in unlabelling experiments conducted *in vivo* (Rowlinson et al., 2022), provision of unlabeled amino acids in eCell CFPS allows suppressing undesired cross-peaks of the unlabeled amino-acid residues. This approach works well for discriminating between Phe and Tyr, as the *E. coli* metabolism does not easily swap carbon atoms between these two residue types (Takeuchi et al., 2007). The presence of tyrosine has been shown to differentially inhibit prephenate dehydrogenase (PDH) activity, thereby preventing 4-

hydroxyphenyl pyruvic acid formation prior to the production of tyrosine. This inhibition results in a decrease in PDH activity of up to 95% (Hudson et al., 1983), which increases the availability of the carbon source for the biosynthesis of other aromatic amino acids. Because PDH is an irreversible catalytic step, the pool of unlabeled tyrosine accumulates and subsequently gets incorporated into the protein (Turnbull et al., 1990). The same mechanism has been shown to occur for phenylalanine, where the inhibition of prephenate dehydratase (PDT) by phenylalanine can reach up to 90% at a concentration of 50 μM (Gething et al., 1976). Most importantly, we found no evidence for eCell CFPS misdirecting ^{13}C into non-targeted positions on the aromatic ring of the labeled amino acid. This establishes eCell CFPS as an attractive *in vitro* tool for selective labelling starting from inexpensive and readily accessible precursors.

The present work demonstrates the labelling of the δ or ε carbons of aromatic amino acids by starting from different isotopologues. Compared to the sparse ^{13}C enrichment of aromatic amino acids *in vivo* reported previously (Loquet et al., 2011), eCell CFPS uses glucose isotopologues much more efficiently, resulting in greatly increased ^{13}C enrichment in the δ positions of Phe and Tyr residues, using either the precursors 3-^{13}C pyruvate (>36%) or 2-^{13}C glucose (>45%). Even higher levels of ^{13}C labelling (>70%) can be achieved for the δ position by the combined use of 3-^{13}C pyruvate and 4-^{13}C erythrose. Previous efforts to increase the labelling efficiency *in vivo* by the combined use of 1-^{13}C glucose and ^{13}C isotopologues of erythrose resulted in no significant increase of label incorporation (Weininger, 2017).

In experiments conducted to label the δ positions of aromatic residues of PpiB by expression from 3-^{13}C pyruvate and 1-^{13}C glucose in eCells, we obtained a labelling degree of only about 50%. Isotopic dilution due to a pool of unlabeled amino acids present in the eCells is difficult to avoid, and attempts to dialyse eCells in a large volume of buffer for an extended period to remove any unlabelled amino acids prior to eCell CFPS led to reduced protein yields, which may be expected as eCells gradually leak biomacromolecules (Van Raad and Huber, 2021). Notably, our estimate of the labelling degree is conservative, as we compared the cross-peak volumes observed for the aromatic C-H groups with the cross-peak of ^{13}C-pyruvate added to the solution as an internal standard. In a protein of the molecular weight of PpiB (19 kDa), transverse relaxation during the INEPT periods of the ^{13}C-HSQC spectra would have affected the protein signals more than those of the internal standard.

In vivo experiments conducted with deuterated pyruvate and 4-^{13}C erythrose have been shown to endow the d positions of Tyr and Phe with single ^{1}H-^{13}C pairs with a labelling degree of 67% while simultaneously installing deuterium in neighbouring positions (Kasinath et al., 2013), which is particularly beneficial for ^{13}C-relaxation measurements (Dreydoppel et al., 2021). While we did not perform experiments with deuterated compounds, we expect that the same labelling pattern would be obtained by eCell CFPS using the same precursors. This expectation is based on the result that eCells maintain the activities of the full complement of enzymes required for the biosynthesis of aromatic amino acids and thus perform in ways closely similar to live *E. coli*. For the same reason we anticipate that synthetic, commercially available late-stage precursors will perform as well in eCell CFPS as in *in vivo* protein expression.

4.6 Conclusions

The eCell system allows for the selective ^{13}C-enrichment of aromatic groups, from early precursors, including atom-selective labelling of ring positions without $^{1}J(^{13}C,^{13}C)$ couplings. High levels of isotope incorporation can be achieved with little expense for isotope-labeled precursors. Facile control over the chemical environment in eCell CFPS enables the ‘unlabelling' of specific residues by simple addition of the respective unlabeled amino acids. The preparation of eCells is straightforward, readily scalable and provides an alternative to costly dialysis membranes. Therefore, the eCell CFPS platform is provides a useful platform for the selective isotope labelling of proteins for NMR spectroscopy.

4.7 References

1. Akke, M. & Weininger, U. (2023) NMR studies of aromatic ring flips to probe conformational fluctuations in proteins. J. Phys. Chem. B 127:591–599.
2. Apponyi, M. A., Ozawa, K., Dixon, N. E. & Otting, G. (2008). Cell-free protein synthesis for analysis by NMR spectroscopy. Methods Mol. Biol., 426, 257-268.
3. Ban, D., Smith, C. A., de Groot, B. L., Griesinger, C. & Lee, D. (2017). Recent advances in measuring the kinetics of biomolecules by NMR relaxation dispersion spectroscopy. Arch. Biochem. Biophys., 628, 81–91.
4. Brylinski, M. (2018). Aromatic interactions at the ligand-protein interface: Implications for the development of docking scoring functions. Chem. Biol. Drug. Des., 91, 380–390.

5. Cai, M., Huang, Y., Sakaguchi, K., Clore, G. M., Gronenborn, A. M. & Craigie, R. (1998). An efficient and cost-effective isotope labeling protocol for proteins expressed in shape Escherichia coli. J. Biomol. NMR., 11, 97–102.
6. Cai, M., Huang, Y., Craigie, R. & Clore. G. M. (2019). A simple protocol for expression of isotope-labeled proteins in Escherichia coli grown in shaker flasks at high cell density. J. Biomol. NMR., 73, 743–748.
7. Dreydoppel, M., Lichtenecker, R. J., Akke, M. & Weininger, U. (2021) ^{1}H $R_{1\rho}$ relaxation dispersion experiments in aromatic side chains. J. Biomol. NMR 75:383–392.
8. Fujiwara, T. & Ramamoorthy, A. (2006). How far can the sensitivity of NMR be increased? Ann. Rep. NMR Spectrosc., 58, 155–175.
9. Gans, P., Hamelin, O., Sounier, R., Ayala, I., Durá, M. A., Amero, C. D., Noirclerc-Savoye, M., Franzetti, B., Plevin, M. J. & Boisbouvier, J. (2010). Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. Angew. Chem. Int. Ed., 49, 1958–1962.
10. Gell, D. A., Kwan, A. H. & Mackay, J. P. (2017). NMR spectroscopy in the analysis of protein-protein interactions. In G. A. Webb (Ed.), Mod. Mag. Res., (pp. 1–34). Springer International Publishing.
11. Gething, M. J. H., Davidson, B. E. & Dopheide, T. A. A. (1976) Chorismate Mutase/Prephenate Dehydratase from Escherichia coli K12. Eur. J. Biochem., 71, 317–325.
12. Jewett, M. C. & Swartz, J. R. (2004). Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. Biotechnol. Bioeng., 86, 19–26.
13. Kainosho, M. & Güntert, P. (2009). SAIL—stereo-array isotope labeling. Q. Rev. Biophys., 42, 247–300.
14. Kasinath, V., Valentine, K. G. & Wand, A. J. (2013). A 13C labeling strategy reveals a range of aromatic side chain motion in calmodulin. J. Am. Chem. Soc., 135, 9560–9563.
15. Kurauskas, V., Schanda, P. & Sounier, R. (2017). Methyl-specific isotope labeling strategies for NMR studies of membrane proteins. Methods Mol. Biol., 1635, 109–123.

16. Lacabanne, D., Meier, B. H. & Böckmann A. (2018) Selective labeling and unlabeling strategies in protein solid-state NMR spectroscopy. J. Biomol. NMR. 71141-71150.
17. Lanzarotti, E., Biekofsky, R. R., Estrin, D. A., Marti, M. A. & Turjanski, A. G. (2011). Aromatic-aromatic interactions in proteins: Beyond the dimer. J. Chem. Inf. Model., 51, 1623–1633.
18. Lichtenecker, R. J., Weinhäupl, K., Schmid, W. & Konrat, R. (2013). α-Ketoacids as precursors for phenylalanine and tyrosine labeling in cell-based protein overexpression. J. Biomol. NMR., 57, 327-331.
19. Linser, R., Gelev, V., Hagn, F., Arthanari, H., Hyberts, S. G. & Wagner, G. (2014). Selective methyl labeling of eukaryotic membrane proteins using cell-free expression. J. Am. Chem. Soc., 136, 11308-11310.
20. Loquet, A., Lv, G., Giller, K., Becker, S. & Lange, A. (2011). ^{13}C spin dilution for simplified and complete solid-state NMR resonance assignment of insoluble biological assemblies. J. Am. Chem. Soc., 133, 4722-4725.
21. Lundström, P., Teilum, K., Carstensen, T., Bezsonova, I., Wiesner, S., Hansen, D. F., Religa, T. L., Akke, M. & Kay, L. E. (2007). Fractional ^{13}C enrichment of isolated carbons using [1-^{13}C]- or [2-^{13}C]-glucose facilitates the accurate measurement of dynamics at backbone Cα and side-chain methyl positions in proteins. J. Biomol. NMR., 38, 199-212.
22. Makwana, K. M. & Mahalakshmi, R. (2015). Implications of aromatic–aromatic interactions: From protein structures to peptide models. Prot. Sci., 24, 1920-1933.
23. Milbradt, A. G., Arthanari, H., Takeuchi, K., Boeszoermenyi, A., Hagn, F. & Wagner, G. (2015). Increased resolution of aromatic cross peaks using alternate ^{13}C labeling and TROSY. J. Biomol. NMR., 62, 291-301.
24. Pervushin K, Riek R, Wider G & Wüthrich K (1998) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in ^{13}C-labeled proteins. J. Am. Chem. Soc. 120:6394–6400.
25. Rasia, R. M., Brutscher, B. & Plevin, M. J. (2012). Selective isotopic unlabeling of proteins using metabolic precursors: Application to NMR assignment of intrinsically disordered proteins. ChemBioChem., 13, 732-739.
26. Rowlinson, B., Crublet, E., Kerfah, R. & Plevin, M. J. (2022) Specific isotopic labelling and reverse labelling for protein NMR spectroscopy: using metabolic precursors in sample preparation. Biochem. Soc. Trans. 50, 1555-1567.

27. Schörghuber, J., Geist, L., Platzer, G., Feichtinger, M., Bisaccia, M., Scheibelberger, L., Weber, F., Konrat, R. & Lichtenecker, R. J. (2018). Late metabolic precursors for selective aromatic residue labeling J. Biomol. NMR., 71, 129-140.
28. Takeda, M., Ono, A. M., Terauchi, T., Miyano, M. & Kainosho, M. (2010). Application of SAIL phenylalanine and tyrosine with alternative isotope-labeling patterns for protein structure determination. J. Biomol. NMR., 46, 45-49.
29. Takeuchi, K., Ng, E., Malia, T. J. & Wagner, G. (2007). 1-^{13}C amino acid selective labeling in a ^{2}H/^{15}N background for NMR studies of large proteins. J. Biomol. NMR., 38, 89-98.
30. Teilum, K., Brath, U., Lundström, P. & Akke, M. (2006) Biosynthetic ^{13}C labeling of aromatic side chains in proteins for NMR relaxation measurements. J. Am. Chem. Soc. 128, 2506–2507.
31. Torizawa, T., Shimizu, M., Taoka, M., Miyano, H. & Kainosho, M. (2004). Efficient production of isotopically labeled proteins by cell-free synthesis: A practical protocol. J. Biomol. NMR., 30, 311-325.
32. Torizawa, T., Ono, A. M., Terauchi, T. & Kainosho, M. (2005). NMR assignment methods for the aromatic ring resonances of phenylalanine and tyrosine residues in proteins. J. Am. Chem. Soc., 127, 12620-12626.
33. Törner, R., Awad, R., Gans, P., Brutscher, B. & Boisbouvier, J. (2020). Spectral editing of intra- and inter-chain methyl-methyl NOEs in protein complexes. J. Biomol. NMR., 74, 83-94.
34. Turnbull, J., Cleland, W. W. & Morrison, J. F. (1990) Chorismate mutase-prephenate dehydrogenase from Escherichia coli. 1. Kinetic characterization of the dehydrogenase reaction by use of alternative substrates. Biochemistry. 2910245-10254.
35. Van Raad, D. & Huber, T. (2021). *In vitro* protein synthesis in semipermeable artificial cells. ACS Synth. Bio., 10, 1237-1244.
36. Verardi, R., Traaseth, N. J., Masterson, L. R., Vostrikov, V. V. & Veglia, G. (2012). Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. Adv. Exp. Med. Biol., 992, 35-62.
37. Weininger, U. (2017). Site-selective ^{13}C labeling of proteins using erythrose. J. Biomol. NMR., 67, 191-200.

38. Whittaker, J. W. (2007). Selective isotopic labelling of recombinant proteins using amino acid auxotroph strains. Methods Mol. Biol., 389, 175-187.
39. Williams, J. K., Schmidt-Rohr, K., & Hong, M. (2015). Aromatic spectral editing techniques for magic-angle-spinning solid-state NMR spectroscopy of uniformly ^{13}C-labeled proteins. Solid. State. Nucl. Magn. Reson., 72, 118-126.

CHAPTER FIVE

eCell technology for cell-free protein synthesis, bio-sensing and remediation

5.1 Abstract

The eCell technology is a recently introduced, specialized protein production platform with uses in a multitude of biotechnological applications. This chapter summarizes the use of eCell technology in four selected application areas. Firstly, for detecting heavy metal ions, specifically mercury, in an *in vitro* protein expression system. Results show improved sensitivity and lower limit of detection compared to comparable *in vivo* systems. Secondly, eCells are semi-permeable, stable, and can be stored for extended periods of time, making them a portable and accessible technology for bioremediation of toxicants in extreme environments. Thirdly and fourthly, applications of eCell technology are shown to facilitate expression of correctly folded disulfide-rich proteins and incorporate chemically interesting derivatives of amino acids into proteins which are toxic to *in vivo* protein expression. Overall, eCell technology presents a cost-effective and efficient method for biosensing, bioremediation and protein production.

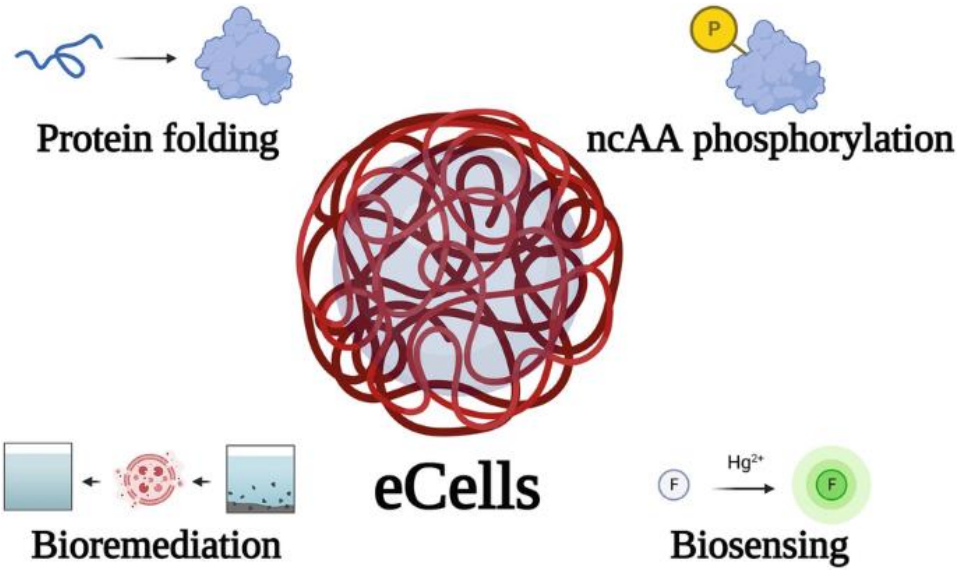


Figure 5.1) The eCell system is a versatile technology with applications in biosensing, bioremediation, protein folding, and the production of proteins with non-canonical amino acids (ncAA). It utilizes a semipermeable capsule to provide stability and protection for proteins, allowing for targeted detection of analytes, efficient removal of pollutants, controlled production of disulfide bonded proteins, and successful incorporation of phosphorylated ncAAs**.** Image produced using biorender.com.

5.2 Introduction:

The eCell system is a novel approach to *in vitro* cell-free protein synthesis that employs the principles of Layer-by-Layer (LbL) polymer assembly to encapsulate bacteria (Hillberg and Tabrizian, 2006). The encapsulation of bacteria circumvents the need for lysate preparation, while still retaining the small cell-like structure in *in vitro* reactions. This results in a cost-effective alternative andaccessible method for cell-free protein synthesis. The semipermeable capsule surrounding the bacteria is formed by the sequential deposition of charged polyelectrolytes onto the cell surface, which creates a rigid shell that retains all components of the bacteria (Scott & Plückthun, 2013). The native cell wall of the bacteria is lysed using a hydrolytic enzyme, leaving a layered and semipermeable capsule surrounding the cell's constituents (Van Raad and Huber, 2021). The eCell system offers several advantages, including versatility and accessibility, as it can be used with conventional and non-conventional energy generation systems. The plasmid is present for translation and requires

only the CFPS buffer to induce expression, making it an attractive alternate platform for cell-free protein synthesis. eCells are non-living entities that maintain many of the advantages of living cells while allowing full chemical control of the environment. By encapsulating cells, the risk of secondary contamination has the potential to be removed, and the efficiency of decontamination can be increased by using enzymes specific to the contaminants. Additionally, enzymes encapsulated within eCells demonstrate a potential for increased stability and a longer shelf life compared to isolated or purified enzymes used for bioremediation, but this has yet to be experimentally validated. By overexpressing bioremediation enzymes in eCells, an *in vitro* decontamination system can be applied to address toxicant and contamination issues such as pesticides and heavy metal detection (Liu et al., 2020; Mali et al., 2022).

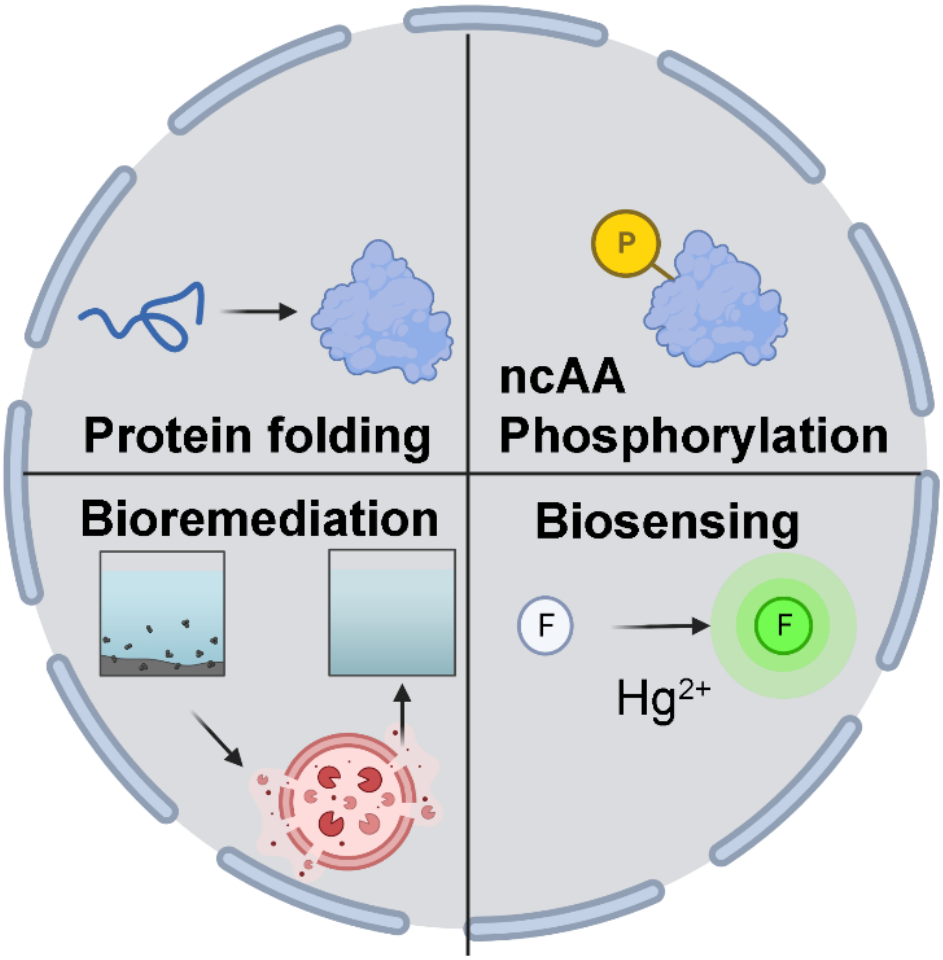


Figure 5.2) The eCell system has applications in the fields of biosensing, bioremediation, production of proteins with non-canonical amino acids (ncAA), and protein folding. In biosensing, the semipermeable capsule of the eCell system can be used as a stable platform

for the detection of analytes. In bioremediation, the eCell system can be used to encapsulate and protect proteins related to the biodegradation of pollutants, allowing for targeted and efficient removal of contaminants. The eCell system can be used to enhance correct protein folding with protein disulfide isomerase (PDI), a crucial enzyme in protein folding. The semipermeable capsule of the eCell system allows for the retention of both the protein and PDI, creating a controlled environment to produce disulfide bonded proteins. In addition eCells have been shown to be successful in incorporation of phosphorylated ncAAs.

In this chapter we utilise the eCell platform in a diverse array of applications some of which enable the detection of heavy metal ions at sub-ppb levels, using a published biosensor for Hg^{2+} ions from (Wan et al., 2019). The detection is straightforward, regulated and highly stringent. For bioremediation, an engineered phosphotriesterase (PTE), previously reported, is being employed for organophosphate decontamination in a range of extreme environments (Campbell et al., 2018). The utilization of eCells in saline environments serves as a demonstration of the robustness of the eCell technology. Additionally, acidic and basic environments pose significant challenges to bioremediation strategies as they can negatively impact the viability of the organism. The protective capsule of the eCell allows for the localization of enzymes, resulting in a stabilization effect and increased efficiency of decontamination in such environments.

The eCell platform has the potential to be a valuable addition to the toolkit of synthetic biology, particularly in the production of conformationally correct disulfide bonded proteins. Using eCell CFPS in combination with protein disulfide isomerase (PDI) and an altered redox potential of the endoplasmic reticulum (-208mV), it is potentially possible to achieve efficient and accurate folding of these difficult to produce disulfide-rich proteins. As demonstrated, this method allows for the soluble expression of neuritin, a human neurotrophic factor, in a soluble form. Additionally, active human light chain enterokinase was produced and was found to cleave its protein target after its DDDDK recognition site with some off target cleavage. These results highlight some of the potential applications of eCells in the field of synthetic biology. The utilization of eCell technology also allows for more advanced modifications of proteins, such as the incorporation of site-specific phosphorylation and/or fluorination of residues. The implementation of a previously published system for the incorporation of pThr and pSer into eCells does not only reduce reagent costs but also

demonstrates the functionalization of bacterial cell lines into eCells using an auxiliary plasmid. Additionally, utilizing eCell technology may address issues of reproducibility in the incorporation of pThr using *in vivo* and standard CFPS methods.

5.3 Biosensing

CFPS systems have been applied in biosensing of a wide range of analytes (Salehi et al., 2017; Hunt et al., 2022; Lu, 2022). CFPS has made biosensing more portable, through utilising on the go fluorescence/absorbance output and complex CFPS paper detection using freeze dried reagents (Soltani et al., 2018; Free et al., 2023). In-cell based biosensors have been used to demonstrate the detection of contaminants in natural environments and the sensitive detection of heavy metal ions (Gupta et al., 2019; Wang et al., 2019; Liu et al 2022). Biosensors for mercury ions have been developed through engineering of the mercury responsive MerR reporter system (Gupta et al., 2019). MerR responds to Hg^{2+} ions in a repressor-activator for its cognate promoter P_{merT}, which controls expression of a green fluorescent protein and allows output characterisation with Hg^{2+} concentrations to be detected at sub ppb levels (Wan et al., 2019). Despite the advantages of *in vivo* bioremediation and biosensing, *in situ* use of GMO presents significant biosafety concerns that prevent large scale use of living organisms in the natural environment (Davidson, 2005). Cell-free synthetic biology in conjunction with biosensing offers a promising mechanism for circumventing secondary contamination through GMO release (Karig, 2017). In this way, biosensors can be safely used to detect toxic and pathogenic contaminants in pristine natural environments.

To demonstrate the eCell technology for biosensing, a plasmid pSB3K3 which constitutively expresses the repressor MerR, which binds to Hg^{2+} ions and induces the transcription of P_{merT} for green-fluorescent protein (GFP) expression (addgene plasmid #123148) was transformed in XJb(DE3)* cell line for eCell CFPS and *in vivo* detection of mercury (Wan et al., 2019).

Expression of GFP was detected on a plate reader to compare Hg^{2+} detection and sensitivity between *in vivo* and *in vitro* systems. Ranges of Hg^{2+} concentrations from 0.001 μM to 1 μM were measured in the comparison of eCell with *in vivo* culture biosensing after overnight culturing and expression. A negative control of 0 μM was used to measure background fluorescence.

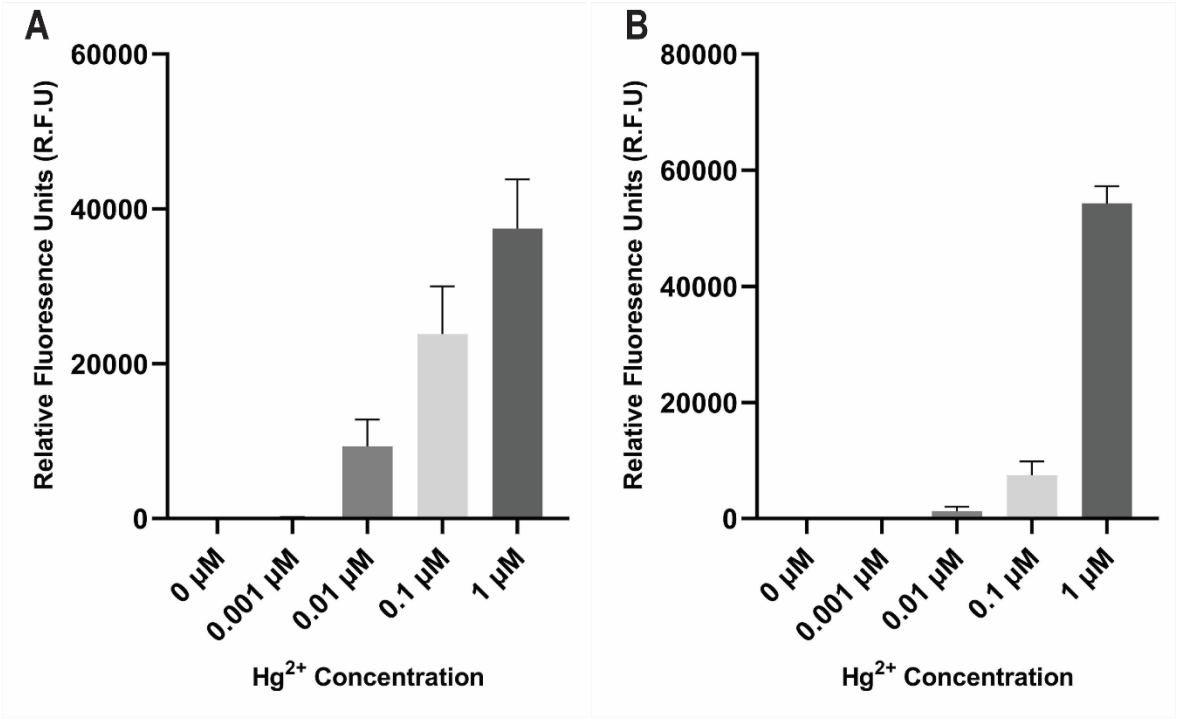


Figure 5.3) Sensor output from 0 µM, 0.001 µM, 0.01 µM, 0.1 µM and 1 µM of Hg^{2+} ions using GFP as a fluorescent reporter. (A) Fluorescent readout of GFP expressed in 3 mL of CFPS buffer using encapsulated and lysed eCells. (B) Fluorescent readout of GFP expressed in 25 mL *in vivo* expression of GFP. Samples were normalised by weighing 50 mg of wet cell-mass in 3 mL to 50 mg of eCells in 3 mL and measuring fluorescence from 100 µL.. Error bars represent standard deviation from triplicate expressions.

Sensitive detection of Hg^{2+} ions is accomplished in the eCell *in vitro* expression system. 700 mg of wet cell mass was generated from overnight expression of the *in-vivo* culture.
The sensitivity of the wet cell mass in our study corresponds to the sensitivity reported in the publication that forms the foundation for the biosensor. This is achieved by employing the same culturing conditions, which involve expression with the same temperature, media and subsequent fluorescence measurements, as described by the authors (Wan et al., 2019). Fluorescence-based analysis revealed a significant increase in GFP expression in eCells exposed to increasing concentrations of Hg^{2+}, indicating a high level of sensitivity in detecting the presence of the heavy metal ions. Comparison of the *in vitro* system to *in vivo* expression revealed a significant increase in sensitivity in the eCells, which is likely due to the higher permeability of the artificial cells to Hg^{2+} ions compared to living cells. The semi-

permeable nature of the eCells allows for passive diffusion of the metal ions, as opposed to relying on controlled cellular uptake, resulting in a more stringent detection limit. The limit of detection for Hg^{2+} using eCells was found to be 0.01 μM, which is approximately 10 times better than the limit of detection observed in *in vivo* expression (0.1 μM). The results obtained were consistent with previously published data using similar constructs for Hg^{2+} detection *in vivo*, further validating the utility of eCells as a sensitive and reliable biosensing system. The eCells have been demonstrated to be an efficient and cost-effective biosensing tool, as previous data has shown that they can be stored for extended periods of time at -80^{o}C without loss of protein expression capability and in this way have the potential to be used as a portable biosensing system (Van Raad and Huber, 2021). Furthermore, the utilization of eCells in natural resource analysis has the potential to pose minimal risk to the environment as they are non-living entities, but further studies are required to demonstrate this attribute.

5.4 Bioremediation

Bioremediation is the functional use of either microbes or reconstituted enzymes to remove toxic molecules from the natural environment (Sales da Silva et al., 2020). This has been used efficiently to transform toxicants into unreactive molecules, mitigating damage to natural resources (Padhan et al., 2021). The effectiveness of decontamination is linked to the efficiency of the enzyme. Genetic engineering has increased the efficiency of bioremediation enzymes, which is typically done by increasing the enzyme's catalytic efficiency. An example of this was through a directed evolution of an engineered phosphotriesterase (PTE) metalloenzyme originating from *Pseudomonas diminuta* (Campbell et al., 2018). This enzyme, which has been found in organisms living in organophosphate (parathion) contaminated soil, has been subjected to directed evolution to increase its catalytic specificity for phosphotriesters.

We hypothesized that eCells would be particularly suitable for bioremediation of toxicants in extreme environments. It was tested through the examination of their ability to inactivate the organophosphate pesticide derivative, paraoxon-ethyl. The eCells were engineered to express an active phosphotriesterase (PTE) enzyme, which facilitates the hydrolysis of paraoxon-ethyl into diethylphosphoric acid and p-nitrophenol. The latter compound absorbs light at a wavelength of 405nm, which was monitored by photometric measurements. The experiment

was conducted by encapsulating PTE-overexpressed XJb(DE3)* cells, lysing their cell wall, and then quantifying their ability to hydrolyse paraoxon-ethyl by measuring absorbance at 30 second intervals for 20 minutes. A total of 0.1 mg of eCells with overexpressed PTE and 200 µl of reaction buffer (20 mM Tris-HCL, 100 mM NaCl, 100 µM $ZnCl_2$, and 100 µM ethyl paraoxon) were used in the assay. The results of this experiment demonstrate the potential of eCells as a viable bioremediation tool for toxicants in extreme environments, specifically for the inactivation of organophosphate compounds. The % remediation values were calculated from the calibration curve of p-nitrophenol as described in Ambreen et al., 2020.

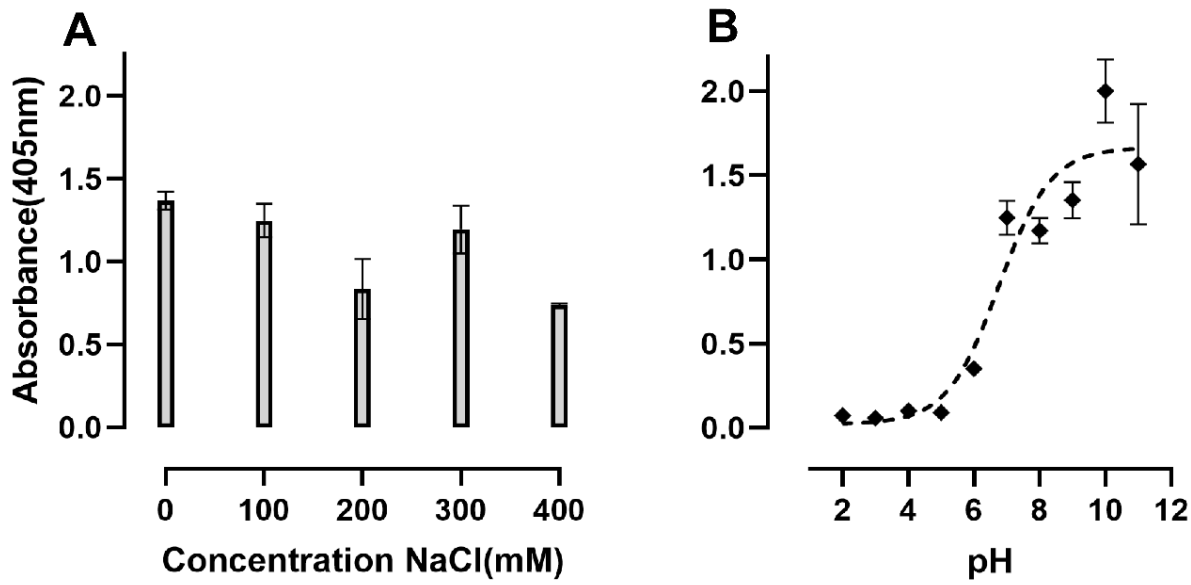


Figure 5.4) A.) Effect of salinity on PTE activity with each of the 20 min endpoints of different salinity being shown, ranging from 100mM to 400mM NaCl. Salinity affects PTE activity only to a small degree up to salt concentration of 300 mM but decreases at 400mM. **B.)** Effect of pH on PTE activity with the final time point of each reaction at pH 2-11. There is no activity at acidic pH but at higher pH the activity of the enzyme remains consistently active.

Figure 5.4 indicates that eCells containing PTE can effectively bioremediate paraoxon-ethyl, a toxic organophosphate pesticide derivative, in extreme environments, such as high salinity and pH. We observed robust PTE activity in eCells even at 400 mM (2.3% w/w) sodium chloride concentration, which approximates the salinity of sea water. The positive control,

without the presence of NaCl (0 mM), exhibited approximately 40% remediation of paraoxon-ethyl. However, as the salt concentration increased to 400 mM, the remediation percentage decreased, reaching approximately 25% in 20 minutes. The activity of the PTE enzyme was only slightly reduced in these conditions, retaining approximately 60% activity of the control. Additionally, we observed that PTE activity in eCells displayed a robustness to environmental pH changes, with increased activity at higher pH. The remediation of paraoxon-ethyl varied with different pH levels. At pH 7, 8, and 9, approximately 40% of paraoxon-ethyl was remediated. However, at pH 10 or 11, the remediation percentage increased to around 50%.However, acidic pH had a significant impact on the activity of encapsulated PTE, resulting in a significant decrease when exposed to pH less than 7. This loss of function at low pH is likely due to the compromised stability of the enzyme rather than changes in the properties of the eCells. On the other hand, PTE in eCells were able to withstand the effects of very basic pH with no effect on activity. The results suggest that eCells remove the need for purification of enzymes and allow their release in a wide variety of environments with the potential for reduced secondary contamination. This is important as current bioremediation strategies utilise living microbes to decontaminate excavated material. These findings demonstrate the potential utility of eCells as a bioremediation tool for real world environments but is important to validate this attribute through further studies, as the absence of secondary contamination has not been confirmed.

An important feature of eCells is that they can be lyophilized and retain some function after re-hydration. This was demonstrated by encapsulating and lysing overexpressed PTE XJb(DE3)* eCells and lyophilizing after snap freezing in two different buffers: phosphate buffered saline (PBS) and sucrose (0.1 g/L). Subsequently, 0.1 mg of the lyophilized eCells were incubated with a reaction buffer and absorbance measurements were taken at regular intervals over a 20-minute period.

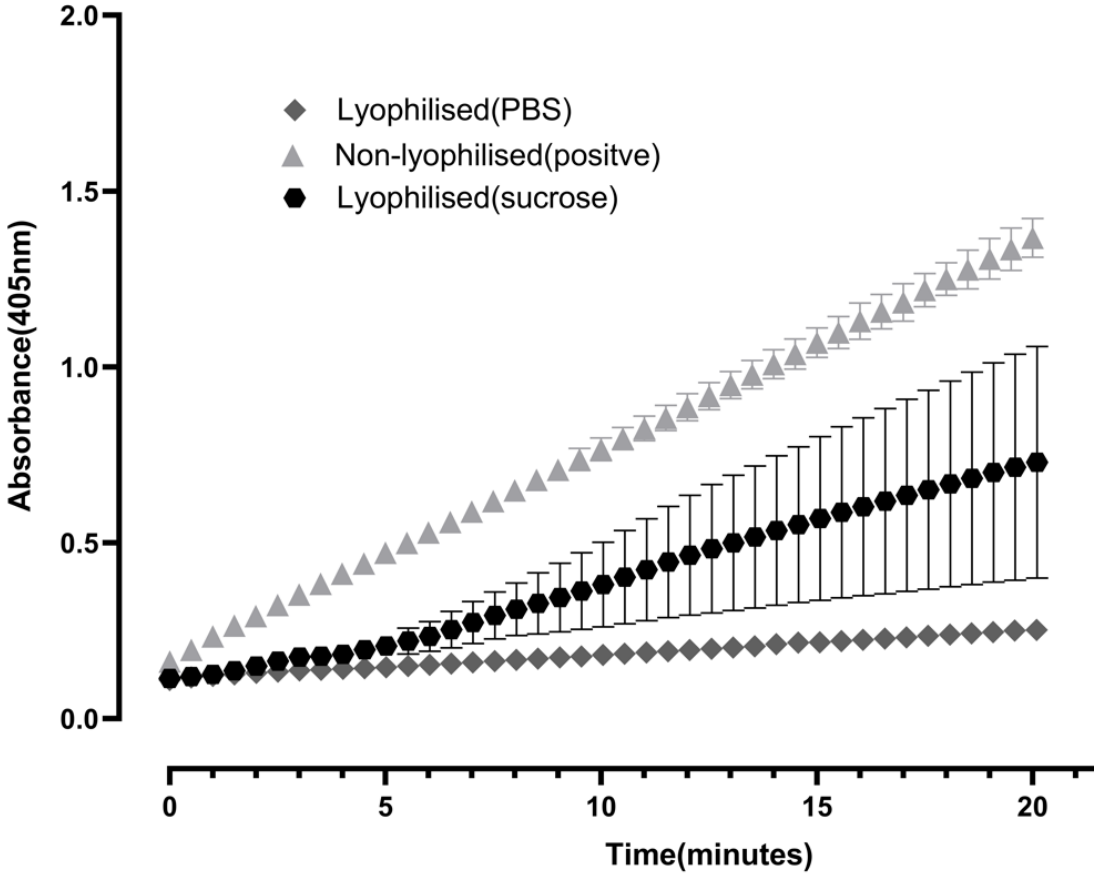


Figure 5.5) This figure depicts a time-dependent analysis of the hydrolysis of paraoxon using encapsulated cells that have been genetically modified to overexpress phosphotriesterase (PTE). The enzymatic activity of PTE was evaluated under lyophilization conditions. The experimental groups included a positive control of encapsulated cells with overexpressed PTE, lyophilized eCells with overexpressed PTE and lyoprotecting groups (sucrose), and eCells with overexpressed PTE that were lyophilized from phosphate-buffered saline (PBS) buffer. The results show that the introduction of lyoprotecting groups, such as sucrose, led to enhanced stabilization of overexpressed PTE, as demonstrated by the observed difference between lyophilization with and without sucrose.

The findings of the lyophilisation experiments demonstrate that eCells containing PTE exhibit some enzymatic activity after lyophilisation and storage in both PBS and sucrose buffer (Figure 5.5). The observed decline in PTE activity in eCells stored in PBS buffer as compared to those in sucrose buffer suggests that sucrose serves as an effective lyoprotectant for eCells. These outcomes suggest that eCell-based preservation methods can be an alternative to traditional lyophilisation of purified protein retaining activity. The positive control using a

standard curve of p-nitrophenol demonstrated approximately 40% remediation of paraoxon-ethyl within a 20-minute timeframe (Ambreen et al., 2020). In the lyophilization experiments involving sucrose and PBS, the remediation percentages were observed to be 15% and 5% respectively. Additionally, the confinement of PTE within eCells is advantageous, preventing dispersion of the enzyme in larger aqueous environments. Nonetheless, the lyophilisation process has a negative impact on the enzyme activity and optimizing the lyophilisation conditions such as temperature, time and buffer composition in future experiments would be critical to maintain enzyme stability/activity.

Like in biosensing, the use of eCells for bioremediation is simple, inexpensive, and does not require the purification of enzymes to inactivate toxicants. eCells have the potential to provide stability on marginally stable proteins and this may allow for the use and storage of proteins with greater protection than reconstituted enzymes. This may be due to the crowded environment of eCells, which act as compartmentalized single cells, rather than the dispersion of reconstituted enzymes. As a result, enzymes within capsules are potentially more stable and have a longer shelf life compared to isolated or purified enzymes used for bioremediation purposes. The activity of PTE overexpressed eCells is reliable in variable environments, with activity retained even when exposed to alkaline pH or high salinity (Figure 5.4A and 5.4B). Bioremediation using living microbes would often not be possible in these extreme environments. eCells can be applied to contamination issues such as pesticides, eliminating the need for microbial decontamination of excavated material. This can be achieved by overexpressing a specific enzyme to be applied to large amounts of contaminated excavated material or aqueous natural resources. The portability of eCells is enhanced by its ability to be lyophilized and retain some proportion of its bioremediation capacity. A loss of 50% activity is observed during the lyophilization process when using sucrose as a lyoprotectant (Figure 5.5). However, the retention of activity in a dry capsule form is advantageous for the portability of eCells.

5.5 Functional expression of disulfide-rich proteins

Preparation of disulfide-rich proteins using cell-based expression systems still suffers from several issues. Typically, these proteins are hard to express in bacterial cells and/or form inclusion bodies which require refolding using denaturants (Singh et al., 2015). This is laborious as it requires gradual reduction of high concentrations of a denaturant (i.e. urea,

guanidium, arginine) and is a slow process (Moghadam et al., 2015). In addition, the process of folding a protein from inclusion bodies and restoring its activity can be inconsistent when denaturants are used. An elegant way to improve the yield of correctly folded disulfide-rich proteins is through modification of the reductive potential of the CFPS lysate and the additional use of a disulfide isomerase. A homodimeric therapeutic biomolecule, porcine Interleukin-12 (IL-12) composed of two subunits of p40 was successfully refolded using a ratio of reduced/oxidised glutathione (GSH/GSSG), protein disulfide isomerase (PDI) from *Homo sapiens* and chaperones (Martens et al., 2000). After p40 overexpression in *E. coli* and solubilisation in urea, PDI using a ratio of GSH/GSSG was found to catalyse the formation and breakage of disulfide bonds into the conformationally correct structure. This PDI mediated redox renaturation was found to be the superior method for correct folding over spontaneous refolding with urea (Martens et al., 2000).

The use of eCells in biotechnology offers several advantages, including potential stability and control over the expression environment. By overexpressing protein disulfide isomerase (PDI) in eCells and altering the expression conditions to mimic the redox potential of the endoplasmic reticulum, it is possible to produce disulfide-rich proteins in a soluble, active form. This was demonstrated using a maltose binding protein fused to a DDDDK linker and a human light chain enterokinase (MBP-DDDDK-EK) with a 6x polyhistidine tag. The pET vector with ampicillin resistance gene, MBP-DDDDK-EK under T7 promoter/terminator and a C112S mutation for stability, was transformed alongside a p15a vector with kanamycin resistance with PDI under a P_{BAD} promoter (P_{BAD} PDI/MBP-DDDDK-EK). 1 L of LB media was inoculated with a 10 mL starter culture of XJb(DE3)* P_{BAD} PDI/EK-MBP. 3 mM of arabinose was added at the beginning of inoculation for the expression of PDI and endolysin and 1 g of eCells were made as described in (Van Raad and Huber, 2021). 1 g of eCells was used with 20 mL creatine phosphate based CFPS expression at pH 7.5 with 1.7 mM GSH and 0.3 mM GSSG and incubated at 37^{o}C overnight.

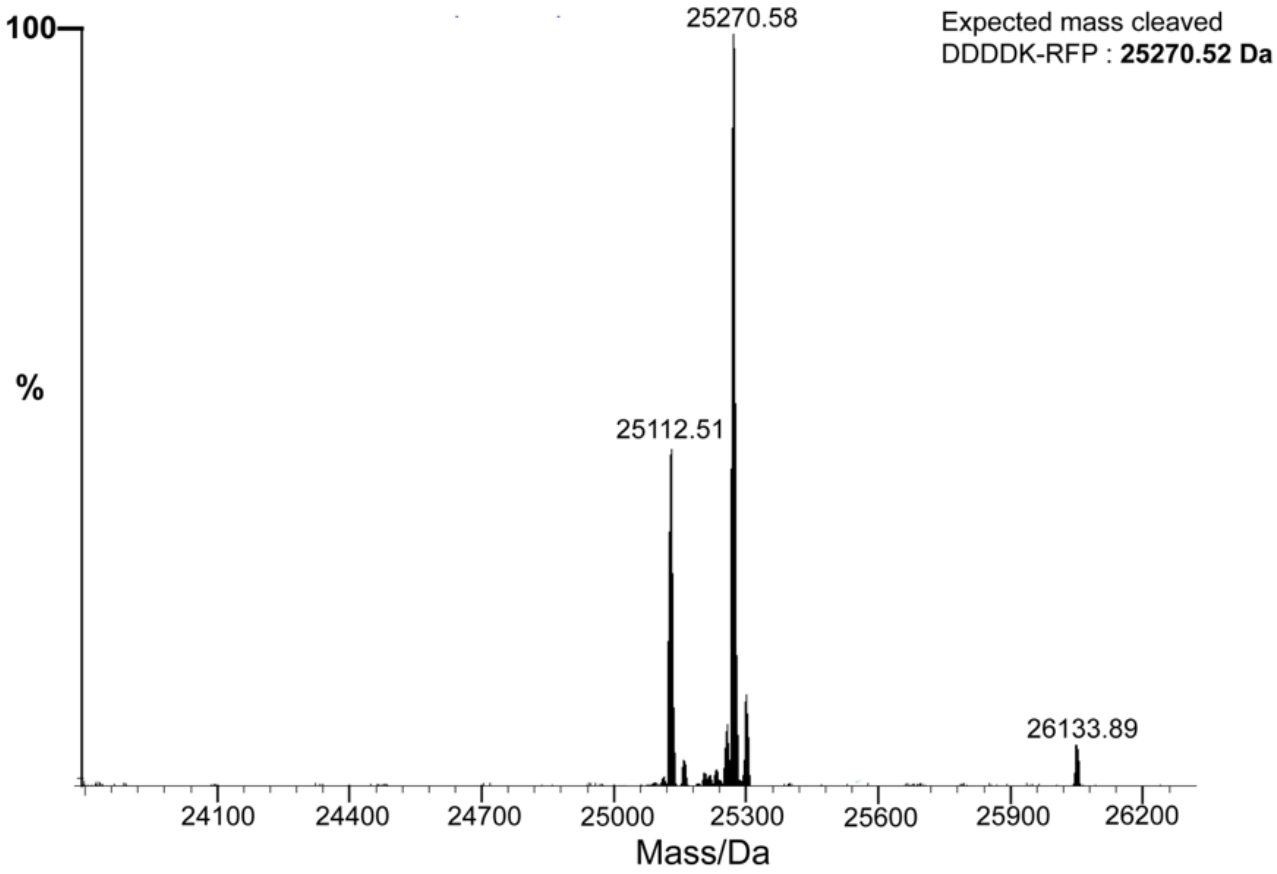


Figure 5.6) High-resolution MS of RFP modified with N-terminal 6xHis tag and DDDDK recognition site for human light chain enterokinase. The sample was incubated overnight with overexpressed human EK light chain purified from eCells. The signal at 25270.58 Da is the cleaved protein after the DDDDK site after 12 hrs of incubation at 37^{o}C.

The soluble MBP-DDDDK-EK was expressed in its active form using eCell technology. 200 µg of protein was produced from the 20 mL eCell CFPS utilising PDI and altered redox potential. To characterise whether the produced EK was active, 60 µl of 40 µM of red fluorescent protein (RFP) with a DDDDK motif preceding an N-terminal 6xHis tag (DDDDK-RFP) was incubated in cleavage buffer (20 mM Tris-HCL pH 8, 50 mM NaCl, 100 µM $CaCl_2$) with 6 µg of EK at 37^{o}C O/N. The cleaved product of DDDDK-RFP was the major species in MS spectrum, although there was also non-specific cleavage of RFP at 25112 Da, which corresponds to cleavage at Ser10, with the 158 difference to corresponding to a product with Ser10 and Ala9 as seen in figure 5.6. The non-specific cleavage could be through EK, or another co-purified protease present within the EK sample. The original purified, uncleaved DDDDK-RFP (26813.08 Da) was not present in the MS, indicating a complete cleavage reaction.

A further demonstration where the eCell technology produced a difficult protein is human neuritin, a disulfide bonded protein which has not been reported being expressed successfully. Neuritin is a human neurotrophic factor which stimulates neuronal cell growth and is linked to inflammation and allergy. Neuritin is a disulfide-linked homodimer, consisting of two 9.72 kDa polypeptide monomer subunits. A pET vector with an ampicillin resistance gene, neuritin with a N-terminal 6xHis tag and a TEV protease cleavage site under a T7 promoter/terminator was co-transformed into XJb(DE3)* with a p15a vector with kanamycin resistance with PDI under a P_{BAD} promoter. A 5 mL CFPS was conducted with 300 mg of PDI overexpressed eCells as described in (Van Raad and Huber, 2021). PDI is on an auxillary P_{BAD} plasmid which expresses when induced with arabinose.

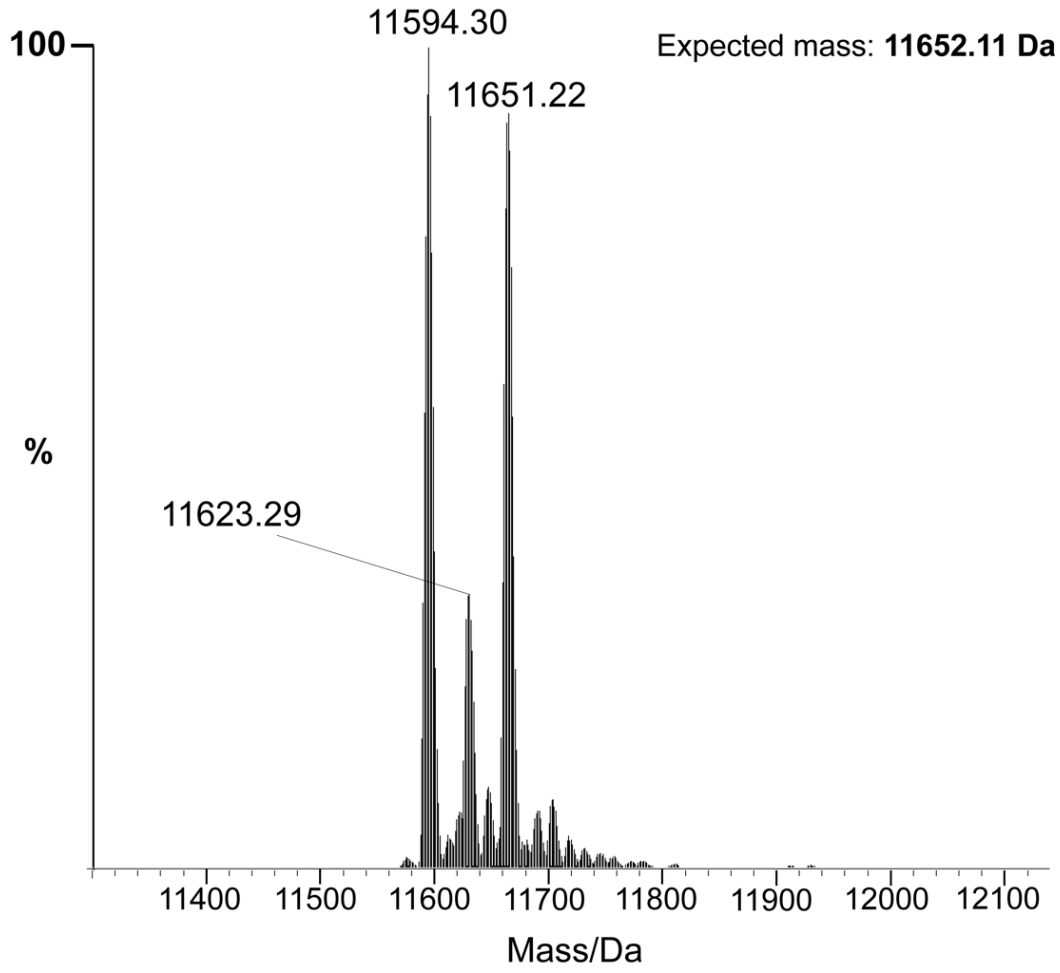


Figure 5.7) High-resolution mass spectrum of Neuritin with a N-terminal 6xHIS tag and a TEV cleavage site. The signals show the neuritin with all the 6 cysteines forming disulfides. The expected mass is 11652 Da for the disulfide bonded form and 11658 for reduced form. There are unidentified products at 11594.30 Da and 11623.29 Da.

The soluble protein obtained from eCell expression was 72 μg from a 5 mL reaction and 300 mg of eCells. Intact mass spectrometry confirms that an isoform of neuritin was produced, however, other unidentified soluble products of neuritin at 11594 Da and 11623 Da were also detected in the MS (Figure 5.7). The refolding of neuritin via protein disulfide isomerase (PDI) within the encapsulation during CFPS reaction was undetermined. The peptide has been oxidised and further work is required to determine if it is folded in its native fold. In contrast to eCell system, bacterial expression of neuritin produced protein in inclusion bodies, requiring refolding, and not guaranteeing the formation of the correct conformation of the protein. To further understand the structure of neuritin, the eCell expression conditions were scaled up to generate a sufficient amount of neuritin for NMR spectroscopy. A 100 mL creatine phosphate/CK CFPS reaction was conducted with about 3.5 grams of eCells, which produced 0.6 mg of soluble neuritin. The 1D ^{1}H NMR spectrum of this neuritin sample, however, showed it was heterogeneous, with broad chemical shifts for amides, methyl groups, and sidechain protons, possibly pointing to fast internal disulfide exchange in neuritin. Typically, *E. coli* expression systems have issues producing disulfide-rich proteins in a conformationally correct structure or in a soluble form, as the cytosol of *E. coli* is reducing, inhibiting the formation of disulfide bonds. There have been in-vivo systems used with success, such as the Shuffle strain of *E. coli* (Lénon et al., 2020). The unique genetic makeup of the Shuffle strain, with specific deletions and a peroxiredoxin enzyme mutation, enables the oxidative folding of disulfide bonds within the cytoplasm, allowing for the expression of soluble disulfide bonded proteins. An alternative is the use of altered redox potential at -208mV for the reshuffling of disulfides via PDI. This has been demonstrated previously *in vitro* but not in complement to a CFPS reaction. The openness of the CFPS reaction allows for precise fine tuning of the redox potential in the CFPS reaction, enhancing PDI ability to potentially form a correct target protein structure. The use of eCells has the additional advantage that PDI does not have to be purified or specialized cell extracts need to be used. eCells is an inexpensive and readily scalable *in vitro* system that has the potential to correctly fold difficult to produce proteins with multiple disulfide bonds, although further work is needed to determine optimal conditions for its use.

5.6 Production of proteins with non-standard amino acids

Using CFPS systems, site-specific protein phosphorylation has allowed the study of conformational shifts of proteins during phosphorylation events (Nishi et al., 2014). Threonine, serine, and tyrosine make up most phosphorylated residues, but it is difficult to specifically target individual residues for homogenous phosphorylation (Silvestroni et al., 2009). A reliable way to recreate phosphorylation at a specific residue is through site specific non-canonical amino acid (ncAA) incorporation (Qin and Liu, 2022). To incorporate a phosphorylated residue, an orthogonal amino-acyl tRNA synthetase (aaRS) and suppressor tRNA ($tRNA_{sup}$) pair is used to recode the amber (TAG) codon to introduce a phosphorylated ncAA into the polypeptide chain (Chen and Tsai, 2022). *E. coli* cell strains and aaRS/$tRNA_{sup}$ have been established for the incorporation of two phosphorylated ncAAs; phosphoserine (pSer) and phosphothreonine (pThr) (Rogerson et al., 2015; Zhang et al., 2017). In naturally occurring phosphoproteins, these two residues are post-translationally installed by kinases, and make up over 98% of all phosphorylation occurrences (Honkanen and Golden, 2002). To homogenously incorporate pSer and pThr translationally in *E. coli* cells, derivatives of BL21(DE3)* cell lines with knockouts for serB and serC were introduced. The knockout ΔserC prevents the production of phosphoserine (pSer) as the pThr aaRS/$tRNA_{sup}$ preferentially incorporates pSer rather than pThr, which is present in *E. coli* (Zhang et al., 2017). ΔserB prevents possible enzymatic dephosphorylation of pSer as the gene encoding phosphoserine phosphatase is inactivated (Rogerson et al., 2015). To utilise these cell strains in site specific protein phosphorylation in CFPS, producing a lysate based on these knockouts is essential.

Chemical modifications of proteins with toxic moieties are not possible to produce by *in vivo* expression. eCell technology is advantageous for complex ncAA incorporation, such as introducing site specific phosphorylation and/or fluorination of alanine, which is toxic to cells. Using eCells, we have incorporated several phosphorylated ncAAs that are difficult to express or have never been expressed *in vitro*. This expression has been done with different cell lines transformed with pACYC-based plasmid carrying a T7 lysozyme gene derivative (pLysS) for eCell generation. This allows the functionalisation of any gram-negative bacteria into encapsulated lysates (eCells), which then allow for their use in CFPS. In this case we use

BL21(DE3)* derivatives for the homogenous site specific phosphorylation of L-phospho-O-threonine (pThr) in E18TAG ubiquitin. As a proof of principle we produced the protein streptococcal protein G 1 (GB1) with site-specific phosphorylation of a serine residue and simultaneously replaced all alanine residue with 3-fluoro-alanine, which is highly toxic to cells but does not affect gene transcription and ribosomal translation of proteins.

The first step was to establish whether pThr incorporation was possible in an *in vitro* system. XJb(DE3)* cells were transformed with a pCloDF13 origin plasmid with a Ubiquitin gene under a T7 promoter which is amber interrupted at position 18 and with a 6xHis tag on the C-terminus (pCDF A18 Ubi) and the engineered pUC ThpRS tRNA(v2.0) pdux efSep previously reported. 1 g of eCells was produced from 1 L of LB media and then induced for CFPS expression using 20 mL CFPS buffer with 5 mM pThr supplemented. At the same time, the gene of a lytic enzyme on a biocompatible vector pLysS was transformed into BL21ΔserC with pCDF A18 Ubi alongside pUC ThpRS tRNA(v2.0) pdux efSep, to functionalise a non-autolysis strain of bacterium that has serC inactivated, preventing intracellular production of SeP. 1 g of BL21ΔserC eCells was produced from 1 L of LB media and then induced for CFPS expression using 20 mL CFPS buffer with 5 mM pThr supplemented.

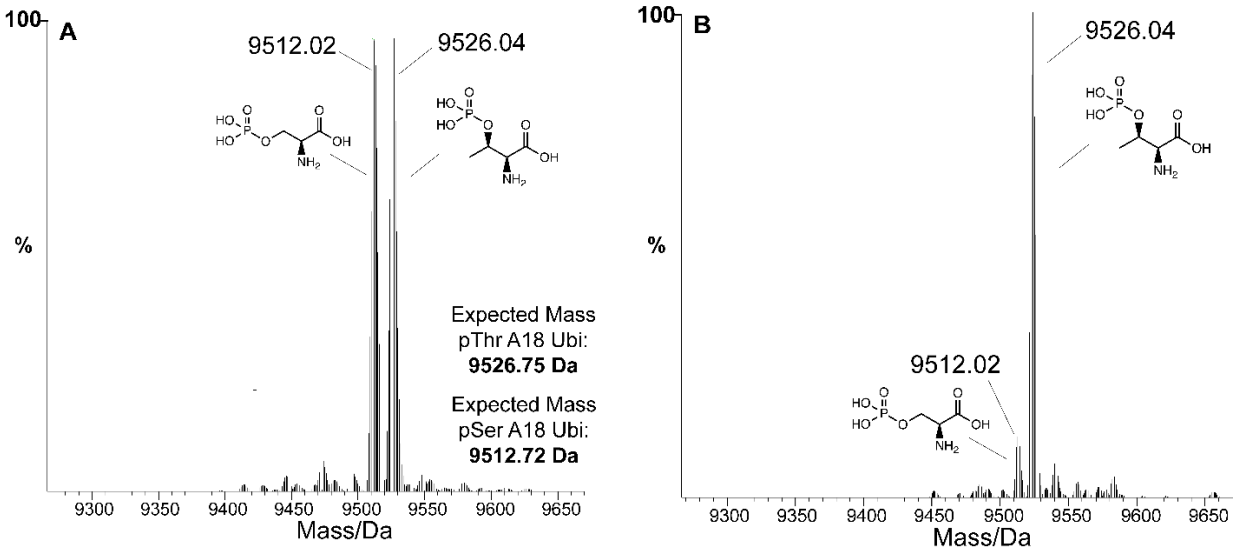


Figure 5.8) A.) Mass spectrum of ubiquitin showing both O-phospho-L-threonine and O-phospho-L-serine incorporated at position 18, due to the high level of O-phospho-L-serine using XJb(DE3)*eCells for expression. **B.)** Mass spectrum of ubiquitin with only O-phospho-L-threonine incorporated at position 18 using functionalised ΔSerC eCells w/pLysS. High-

resolution mass measurements were performed on an Orbitrap Elite mass spectrometer in positive mode at a resolution of 240,000.

The yield for XJb(DE3)* eCell expression was 0.95 mg, while BL21ΔserC was 0.72 mg. in Figure 5.8A shows the mass spectrometry analysis of the XJb(DE3)* eCell produced samples. The two observed mass peaks are for the incorporation of both pSer and pThr in equal amount at position 18 of ubiquitin. XJb(DE3)* has an active SerC gene increasing intracellular concentration of pSer, competing with pThr in the charging of tRNA and subsequent incorporation. Supplementing pThr at 5 mM concentration in the CFPS produced protein with 50% of the desired product. Functionalising BL21ΔserC into eCells avoid the intracellular production of pSer and allow for selective incorporation of pThr. As seen in Figure 5.8B the incorporation of pThr increased to approximately 90% using functionalised BL21ΔserC eCells, highlighting that reduction of intracellular pSer in CFPS systems increases incorporation and that bacterial cell lines can be functionalised using an auxillary plasmid for eCell CFPS. To produce proteins with multiple modifications, such as fluorination and phosphorylation, it is necessary to utilise eCells. Proteins with residue specific replacement of fluorinated groups, such as 3-fluoro-alanine, and phosphor-serine cannot be produced *in vivo*, due to cell toxicity and the difficulty of producing phosphorylated residues containing proteins in standard CFPS (Kollonitsch & Barash, 1973).

A vector with spectinomycin resistance, pCloDF13 origin with MASMTG-ENLFYQ-streptococcal protein G B1 gene (GB1) under a T7 promoter which is amber interrupted at position 24 and 28 was transformed into BL21ΔserB with the engineered phosphoserine orthogonal translation system (SepOTSλ) and PlysS. 0.6 g of BL21ΔserB eCells were used from this 3-plasmid system in a 15 mL CFPS with 4 mM pSer and 7 mM of DL- 3-fluoro-alanine replacing L-alanine in the reaction mixture.

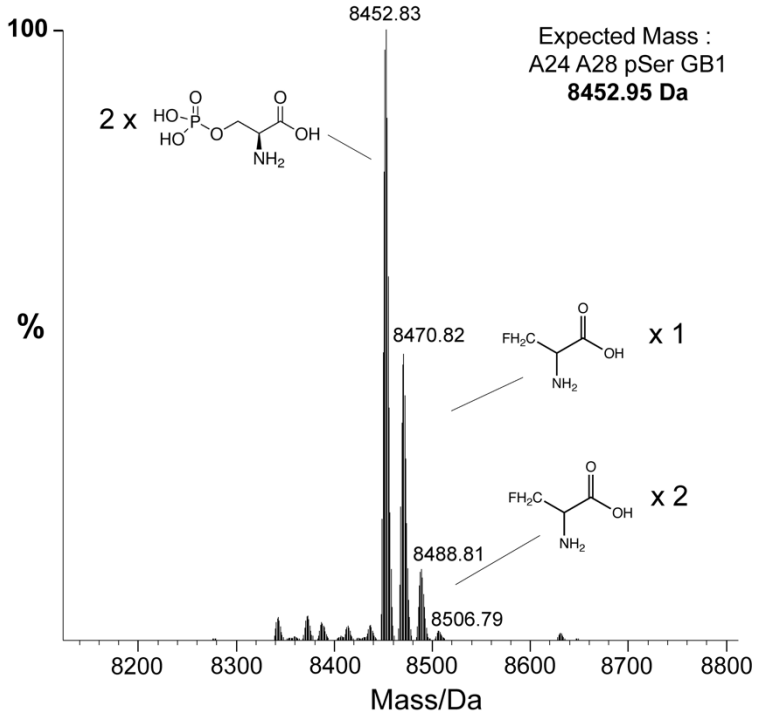


Figure 5.9) High-resolution mass spectrum of GB1 containing two phosphor-serine residues and alanine being replaced by 3-fluoro-alanine. The signals show the measure masses of the molecules and the expected (calculated) masses for GB1 with two phosphor-serine (pSer) and varying number of 3-fluoro-alanine replacing native alanine. This expression was done using functionalised ΔSerB eCells w/pLysS, 4mM pSer and 7mM 3-fluoro-alanine. The expression yield was or 0.4 mg from a 15 mL eCFPS reaction.

0.3 mg of protein was produced from 0.6 g of eCells of GB1 double incorporated pSer at A24 and A28. Figure 5.9 shows the mass spectrum of the produced protein. Both phosphoserine residues have been successfully introduced into the protein, but 3-fluoro-alanine replacement is only partially, with only one or two alanine residues out of the possible three replaced. These results suggest that alanine is either synthesized within eCells during the CFPS reaction or was partially retained in the eCells despite washing steps post-lysis. Using XJb(DE3)* cells for incorporation of pSer and 3-fluoro-alanine produced a more complicated to interpret mass spectrum, because both pSer and pThr were incorporated at the two amber sites and alanine residues were not fully replaced by 3-fluoro-alanine.

The functionalisation of *E. coli* derivatives into eCells using a lytic auxiliary plasmid has allowed for homogenous phosphorylation of different protein targets. In combination, the

residue specific fluorination of alanine has been applied to protein targets. This allows chemically interesting derivatives of amino acids to be incorporated into proteins for NMR spectroscopy, as ^{19}F is an attractive NMR active spin which can be detected readily with high sensitivity. The incorporation of native alanine alongside 3-fluoro-alanine highlights the issue of natural amino acids being present in the eCells prior to expression. Extensive dialysis steps have been applied, but important macromolecules related to transcription and translation can leak over time from the permeable encapsulation, ultimately reducing protein yield. Using an auxiliary plasmid, eCells are expanded in used to not only the XJb(DE3)* cell line, but to any gram negative bacteria. We demonstrate this by functionalising two *E. coli* cell lines, BL21ΔserC and BL21ΔserB for incorporation of pSer and pThr.

5.7 Materials and methods

5.71 Biosensing

E. coli XJb(DE3)* cells with P_{merT} GFP was grown in LB medium at 37 ^{o}C in baffled flasks with shaking at 180 rpm. Endolysin production was induced at the time of inoculation with a final concentration of 3 mM arabinose. Cells were grown to OD600 = 0.6, harvested by centrifugation at 2,773 g and washed three times with 20 mL PBS-E buffer (phosphate-buffered saline with 1 mM EDTA, pH 7.4). For coating with chitosan, 1 g of the cells were resuspended in 20 mL of 0.25 mg/mL chitosan in PBS-E with vigorous shaking for 20 minutes. 1 g of the cell pellet was washed with 20 mL PBS-E pH 6.0 three times to remove excess chitosan and then resuspended in 20 mL of 0.25 mg/mL alginate PBS-E solution and subjected to vigorous shaking for 20 minutes. The eCells were frozen in 50 mg aliquots. The frozen aliquots of encapsulated cells were thawed, spun down and the pellet immersed in 3 mL of CFPS buffer with varying concentrations of $Hg(OAc)_2$ from 0.01 μM to 1 μM, with a negative control of 0 μM being used. CFPS for each experiment was conducted at 37^{o}C in a shaker at 200 rpm. Fluorescence readings were taken at 12 hours using a SpectraMax fluorescence reader, with SoftMax pro as the measurement software.

5.72 Bioremediation

E. coli XJb(DE3)* cells housing a vector with a puC origin of replication, ampicillin resistance and a phosphotriesterase gene under a T7 promoter and terminator named

(PETMCSII PTE). This was grown in LB medium at 37 ^{o}C in baffled flasks with shaking at 180 rpm. Endolysin and PTE production was induced at the time of inoculation with a final concentration of 3 mM arabinose and 1 mM IPTG. Cells were grown overnight to express endolysin and PTE and then encapsulated as described in 5.71.

In each experiment, a total of 0.1 mg of eCells with overexpressed phosphotriesterase (PTE) and 200 µl of reaction buffer (20 mM Tris-HCl, 100 mM NaCl, 100 µM $ZnCl_2$, and 100 µM ethyl paraoxon) were used. Encapsulated cells were tested for their robustness in varying salinity levels. The cells were immersed in five different salt-containing buffers: 0 mM (0% w/w) 100 mM (0.5% w/w), 200 mM (1.1% w/w), 300 mM (1.7% w/w), and 400 mM (2.3% w/w) NaCl. Organophosphate (ethyl-paraoxon) was added to the solution at 100 µM, and the absorbance at 405 nm was monitored over a 20-minute period. For the investigation of the stability of encapsulated cells in highly acidic and basic environments, the cells were encapsulated and immersed in nine different pH solutions ranging from highly basic (pH 11) to highly acidic (pH 2) buffered by Tris-HCl. Organophosphate was titrated into the solutions, and the absorbance at 405 nm was measured. The encapsulation of pre-expressed PTE transformed cells, followed by lyophilization was done using two different buffers: phosphate buffered saline (PBS) and sucrose in (0.1g/L) in 1 mL. The cells were snap frozen in liquid nitrogen and then lyophilised with Labconco^{Tm} FreeZone 4.5 Liter Benchtop Freeze Dry System at 0.1 mbar of pressure and at -47^{o}C overnight. 0.1 mg of lyophilized cells were then added to 200 µl of reaction buffer (20 mM Tris-HCl, 100 mM NaCl, 100 µM $ZnCl_2$, and 100 µM ethyl paraoxon) the retained activity of the lyophilized cells was evaluated by measuring the hydrolysis activity, which produced p-nitrophenol that absorbs at 405 nm. The % remediation values were calculated from the calibration curve of p-nitrophenol as described in Ambreen et al., 2020.

5.73 Protein folding

E. coli XJb(DE3)* cells housing a vector with a p15a origin of replication, kanamycin resistance and protein disulfide isomerase (PDI) under an araBAD promoter and terminator was transformed with two proteins of interest. One protein vector had a puC origin of replication, ampicillin resistance, a human light chain enterokinase fused to maltose binding protein with a C112S mutation and a 6x polyhistidine tag under a T7 promoter and terminator named (MBP-DDDDK-EK). Another had a puC origin of replication, ampicillin resistance,

the neuritin gene with a 6x polyhistidine tag and a TEV protease recognition site ENLYFQ under a T7 promoter and terminator. Endolysin and PDI production was induced at the time of inoculation with a final concentration of 3 mM arabinose. Cells were grown to OD_{600} 0.7 to express PDI and endolysin and then encapsulated as described in 5.71. 1 g of eCells were produced and frozen. These eCells were thawed and used with 20 mL creatine phosphate based CFPS buffer at pH 7.5 with 1.7 mM GSH and 0.3 mM GSSG and incubated at 37^{o}C overnight. This ratio of GSH and GSSG amounted to a redox potential of -208 mv. 200 μg of protein was produced for MBP-EK-DDDDK. To assess the activity of the produced MBP-EK-DDDDK, a reaction mixture was prepared by incubating 60 μl of 40 μM red fluorescent protein (RFP) containing a DDDDK motif preceding an N-terminal 6 x polyhistadine tag (DDDDK-RFP) in cleavage buffer (20 mM Tris-HCL pH 8, 50 mM NaCl, 100 μM $CaCl_2$) with 6 μg of EK. The reaction mixture was then allowed to incubate overnight at 37°C.

72 μg of neuritin was produced from a 5 mL reaction and 300 mg of eCells. For protein purification, 1 ml His GraviTrap TALON columns (GE Healthcare) were used for both proteins. Protein concentration was determined using the Thermo Scientific Nanodrop™ One/Onec Microvolume UV spectrophotometer using A280 and the extinction coefficient of both proteins. The proteins were analysed using mass spectrometry.

5.74 Production of proteins with non-standard amino acids

E. coli BL21ΔserC and BL21ΔserB cells housing a vector containing a p15a origin of replication, chloramphenicol resistance and T7 lysozyme gene under an araBAD promoter and terminator (PlysS). BL21ΔserC and BL21ΔserB cells also contain a a plasmid with a CloDF13 origin, the gene for streptococcal protein G B1 domain with two amber sites at A24, A28 under a T7 promoter and terminator, with spectinomycin resistance (PCDF GB1 A24, A28). For BL21ΔserC the plasmid from Zhang et al., 2017 (addgene number #173899) called pUC pThrRS Sep_{tRNA}v2.0CUA OXB20-PduX EFSep (PduX EFSep) was transformed for phosphothreonine incorporation. For XJb(DE3)* expression only PduX EFSep and PCDF GB1 A24, A28 was transformed. For BL21ΔserB the plasmid from Rogerson et al., 2015 called SepOTSλ (addgene #173897) was transformed into cells containing PlysS and PCDF GB1 A24, A28. Endolysin and endolysin production was induced at the time of inoculation with a final concentration of 3 mM arabinose. Cells were grown to OD_{600} 0.7 to express and then encapsulated as described in 5.71. For BL21 ΔserC and XJb(DE3)* the same culturing

conditions and encapsulation was applied resulting in 1 g of eCells produced from 1 L of LB media and were frozen. The eCells were thawed and induced for CFPS expression using 20 mL CFPS buffer with 5 mM pThr supplemented. For BL21ΔserB 0.6 g of BL21ΔserB eCells were used from this 3-plasmid system in a 15 mL CFPS with 4 mM pSer and 7 mM of DL- 3-fluoro-alanine replacing L-alanine in the reaction mixture.

For protein purification, 1 ml His GraviTrap TALON columns (GE Healthcare) were used for both proteins. Protein concentration was determined using the Thermo Scientific Nanodrop™ One/Onec Microvolume UV spectrophotometer using A280 and the extinction coefficient of both proteins. The proteins were analysed using mass spectrometry.

5.75 Mass spectrometry

High resolution electro spray ionization mass spectrometry was performed in positive mode on an Orbitrap Elite mass spectrometer and on Orbitrap Fusion™ Tribrid™ mass spectrometer coupled with an UltiMate 3000 UHPLC (Thermo Scientific, USA). Samples were injected into the mass analyzer via an Agilent ZORBAX SB-C3 Rapid Resolution HT Threaded Column, using an acetonitrile gradient (10–85 %) and 0.1 % formic acid. Mass spectra were deconvoluted and processed using the Xcalibur software package.

5.8 Summary

In this study, an *in vitro* expression system was employed using eCells to detect heavy metal ions, specifically mercury. The eCell system was shown to successfully detect mercury ions at higher sensitivity compared to *in vivo* expression. The semi-permeable nature of the eCells capsules allows for passive diffusion of the metal ions, improving stringency. The eCells system is efficient, inexpensive, and can be stored for extended periods of time, making it a portable and accessible biosensing technology without risk to the environment. The eCell technology is a promising approach for bioremediation of toxicants in extreme environments due to its ability to encapsulate and protect enzymes, providing greater stability and shelf life compared to isolated or purified enzymes. Furthermore, the eCells exhibit robustness to environmental changes such as high salinity and pH, and their lyophilisation feature enhances their portability, making them a viable option for on-site bioremediation.

Additionally, eCell technology can be used for expressing disulfide-bonded proteins in a soluble, active form. This is accomplished by controlling the expression environment and overexpressing protein disulfide isomerase (PDI) to mimic the redox potential of the endoplasmic reticulum. The semipermeable capsule of the eCell system allows for the retention of both the protein and PDI, providing a controlled environment for conformationally correct protein folding. The efficacy of this technology was demonstrated using human light chain enterokinase, and a human neurotrophic factor called neuritin, which produced significant amounts of protein for NMR spectroscopy.

Moreover, this study presents a method for simultaneous incorporation of SeP, pThr, and 3-fluoro-alanine into proteins using eCell CFPS with *E. coli* derivatives and a lytic auxiliary plasmid. The method enables site-specific phosphorylation of different protein targets and the incorporation of chemically interesting derivatives of amino acids. The approach is quicker and less costly than other methods for producing capsule lysate.

In conclusion, eCell technology is a versatile and promising approach for various biotechnological applications. It has shown to be efficient, cost-effective, and environmentally friendly, with potential use in biosensing and bioremediation, as well as in the production of challenging-to-express proteins.

5.8 References:

1. Ambreen, S., Yasmin, A. & Aziz S. (2020). Isolation and characterization of organophosphorus phosphatases from Bacillus thuringiensis MB497 capable of degrading Chlorpyrifos, Triazophos and Dimethoate. Heliyon., 6, 1-9.
2. Campbell, E. C., Grant, J., Wang, Y., Sandhu, M., Williams, R. J., Nisbet, D. R. & Scott, C. J. (2018). Hydrogel-Immobilized Supercharged Proteins. Adv. Biosyst., 2, 1-11.
3. Chen, J. & Tsai, Y. H. (2022). Applications of genetic code expansion in studying protein post-translational modification. J. Mol. Biol., 434, 1-12.
4. Davison, J. (2005). Risk mitigation of genetically modified bacteria and plants designed for bioremediation. J. Ind. Microbiol. Biotechnol., 32, 639-650.

5. Free, T. J., Tucker, R. W., Simonson, K. M., Smith, S. A., Lindgren, C. M., Pitt, W. G. & Bundy, B. C. (2023). Engineering At-Home Dilution and Filtration Methods to Enable Paper-Based Colorimetric Biosensing in Human Blood with Cell-Free Protein Synthesis. Biosensors (Basel)., 13, 1-13.
6. Gupta, N., Renugopalakrishnan, V., Liepmann, D., Paulmurugan, R. & Malhotra, B. D. (2019). Cell-based biosensors: Recent trends, challenges and future perspectives. Biosens. Bioelectron., 141, 1-23.
7. Hillberg AL, Tabrizian M. (2006). Biorecognition through Layer-by-Layer Polyelectrolyte Assembly: In-Situ Hybridization on Living Cells. Biomacromolecules. 7, 2742-2750.
8. Mali, H., Shah, C., Raghunandan, B. H., Prajapati, A., Patel, D., Trivedi, U. & Subramanian, B. (2022). Organophosphate pesticides an emerging environmental contaminant: Pollution, toxicity, bioremediation progress, and remaining challenges. J. Env. Sci. 127. 234-250.
9. Honkanen, E. R. & Golden, T. (2002). Regulators of serine/threonine protein phosphatases at the dawn of a clinical era? Curr. Med. Chem., 9, 2055-2075.
10. Hunt, J. P., Galiardi, J., Free, T. J., Yang, S.O., Poole, D., Zhao, E.L., Anderson, J. L., Wood, D.W. & Bundy B. C. (2022) Mechanistic discoveries and simulation-guided assay optimization of portable hormone biosensors with cell-free protein synthesis. Biotechnol. J., 17, 1-8.
11. Karig, D. K. (2017). Cell-free synthetic biology for environmental sensing and remediation. Curr. Opin. Biotechnol., 45, 69-75.
12. Kollonitsch, J. & Barash, L. (1976) Organofluorine synthesis via photofluorination: 3-fluoro-D-alanine and 2-deuterio analogue, antibacterials related to the bacterial cell wall. J Am Chem Soc 98:5591–5593
13. Lénon. M., Ke. N., Szady. C., Sakhtah. H., Ren. G., Manta. B., Causey. B. & Berkmen M. (2020) Improved production of Humira antibody in the genetically engineered Escherichia coli SHuffle, by co-expression of human PDI-GPx7 fusions. Appl Microbiol Biotechnol. 104, 9693-9706.
14. Liu, W. Q., Wu, C., Jewett, M. C. & Li, J. (2020). Cell-free protein synthesis enables one-pot cascade biotransformation in an aqueous-organic biphasic system. Biotechnol. Bioeng., 117, 4001-4008.

15. Liu, C., Yu, H., Zhang, B., Liu, S., Liu, C. G., Li, F. & Song. H. (2022). Engineering whole-cell microbial biosensors: Design principles and applications in monitoring and treatment of heavy metals and organic pollutants. Biotechnol. Adv., 60, 1-20.
16. Lu, Y. (2022) Textile-embedded cell-free biosensors. Nat. Biomed. Eng., 6, 225-236.
17. Martens, E., Alloza, I., Scott, C. J., Billiau, A. & Vandenbroeck, K. (2000). Protein disulfide isomerase-mediated cell-free assembly of recombinant interleukin-12 p40 homodimers. Eur. J. Biochem., 267, 6679-6683.
18. Moghadam, M., Ganji, A., Varasteh, A., Falak, R. & Sankian, M. (2015). Refolding process of cysteine-rich proteins: Chitinase as a model. Rep. Biochem. Mol. Biol., 4, 19-24.
19. Nishi, H., Shaytan, A. & Panchenko, A. R. (2014). Physicochemical mechanisms of protein regulation by phosphorylation. Front. Genet., 5, 1-10.
20. Padhan, D., Rout, P. P., Kundu, R., Adhikary, S. & Padhi, P. P. (2021). Bioremediation of Heavy Metals and Other Toxic Substances by Microorganisms. Soil Bioremediation., 285-329.
21. Qin, X. & Liu, T. (2022). Recent advances in genetic code expansion techniques for protein phosphorylation studies. J. Mol. Biol., 434 1-12.
22. Rogerson, D. T., Sachdeva, A., Wang, K., Haq, T., Kazlauskaite, A., Hancock, S. M., Dezot, H. D., Muqit, M. M. K., Fry, A. M. F., Bayliss, R. & Chin, J. W. (2015). Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nat. Chem. Biol., 11, 496-503.
23. Sales da Silva, I. G., Gomes de Almeida, F. C., Padilha da Rocha e Silva, N. M., Casazza, A. A., Converti, A. & Asfora Sarubbo, L. (2020). Soil bioremediation: Overview of technologies and trends., Energies., 13, 1-25.
24. Salehi, A. S., Shakalli Tang, M. J., Smith, M.T., Hunt. J. M., Law, R.A., Wood, D.W. & Bundy, B. C. (2017) Cell-Free Protein Synthesis Approach to Biosensing hTRβ-Specific Endocrine Disruptors. Anal Chem., 89, 3395-3401.
25. Scott, D. J. & Plückthun, A. (2013). Direct molecular evolution of detergent-stable g protein-coupled receptors using polymer encapsulated cells. J. Mol. Biol., 425, 662–677.
26. Silvestroni, A., Jewell, K. A., Lin, W. J., Connelly, J. E., Ivancic, M. M., Tao, W. A. & Hu, F. (2009). Identification of serine/threonine kinase substrates in the human pathogen group B streptococcus. J. Proteome. Res., 8, 2563-2574.

27. Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M. & Panda, A. K. (2015). Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb. Cell. Fact., 14, 1-10.
28. Soltani, M., Davis, B. R., Ford, H., Nelson, J. A. D. & Bundy, B. C. (2018). Reengineering cell-free protein synthesis as a biosensor: Biosensing with transcription, translation, and protein-folding. Biochem. Eng. J., 138, 165-171.
29. Van Raad, D. & Huber, T. (2021). *In vitro* Protein Synthesis in Semipermeable Artificial Cells. ACS. Synth. Biol., 10, 1237-1244.
30. Wan, X., Volpetti, F., Petrova, E., French, C., Maerkl, S. J. & Wang, B. (2019). Cascaded amplifying circuits enable ultrasensitive cellular sensors for toxic metals. Nat. Chem. Biol., 15, 540-548.
31. Wang, X., Cheng, M., Yang, Q., Wei, H., Xia, A., Wang, L., Ben, Y., Zhou, Q., Yang. Z. & Huang, X. (2019). A living plant cell-based biosensor for real-time monitoring invisible damage of plant cells under heavy metal stress. Sci. Total Environ., 697, 1-10.
32. Zhang, M. S., Brunner, S. F., Huguenin-Dezot, N., Liang, A. D., Schmied, W. H., Rogerson, D. T. & Chin, J. W. (2017). Biosynthesis and genetic encoding of phosphothreonine through parallel selection and deep sequencing. Nat. Methods., 14, 729-736.

CHAPTER SIX

Conclusion and Future direction

This thesis presents the development and applications of the eCell system. The eCell system is a novel cell-free protein synthesis (CFPS) platform that utilizes Layer-by-Layer (LbL) polymer assembly to encapsulate the entire repertoire of transcription/translation enzymes in *E. coli*. This allows for cost-effective generation of capsule lysates which allow for the incorporation of non-canonical amino acids (ncAAs) into proteins. This system eliminates the need for traditional cell lysate preparation and purification of amino acyl-tRNA synthetases (aaRS), while still retaining the small scale of an *in vitro* reaction. eCells are semi-permeable, stable, and can be stored for extended periods of time, making them a portable and accessible technology for biosensing, bioremediation and protein production. They have been shown to facilitate expression of correctly folded disulfide-rich proteins and can be used to produce proteins with selectively ^{13}C-labeled amino acids from inexpensive ^{13}C-labeled precursors. Additionally, eCells have been demonstrated the partitioning of lysate using fluorescent activated cell sorting (FACS), linking phenotype to genotype. Overall, eCell technology presents a cost-effective and efficient method for a wide range of biotechnological applications. The eCell technology has also been applied in four areas, including detecting heavy metal ions, bioremediation of toxicants, facilitating expression of correctly folded disulfide-rich proteins, and incorporating chemically interesting derivatives of amino acids into proteins which are toxic to *in vivo* protein expression. Overall, eCell technology presents a cost-effective and efficient method for biosensing, bioremediation, and protein production.

6.1 Final Remarks

Potential future research areas involve the segregation of extensive libraries in eCells for the purpose of screening protein mutants. This is because eCell technology has a distinct advantage over microfluidics in terms of its ability to encapsulate many cells without being constrained by time or complicated apparatus. However, microfluidics offers an advantage in terms of precise control of encapsulation to produce only single-cell encapsulation, whereas

eCell technology may produce capsules containing multiple cells that can be gated away with FACS. Nevertheless, the ease and durability of the eCell system, combined with the ability to sort millions of mutants with FACS, offer an advantage over directed evolution based on microfluidics. This allows for the high-throughput selection of rare mutations that can improve protein stability under conditions that are typically incompatible with biological systems.

The application of cell-free lysates derived from various yeast strains has emerged as a powerful tool for studying eukaryotic protein expression *in vitro*. Yeast-based cell-free protein synthesis (CFPS) systems have the potential to overcome some of the limitations of traditional *in vivo* protein expression systems, such as toxicity and limited yields, while also providing high yields of functional proteins with post-translational modifications.

Among the diverse yeast strains, *S. cerevisiae*, a well-established model organism and a highly productive bio-manufacturing platform *in vivo*, has attracted considerable attention for its application in yeast-based CFPS. Recent studies have demonstrated the potential of *S. cerevisiae* lysates in expressing functional proteins in a cell-free environment, including difficult-to-express proteins and large multi-domain proteins. Given its advantages in large-scale production, ease of genetic manipulation, and the extensive knowledge of its metabolism and genetics, yeast-based CFPS is poised to play a significant role alongside other CFPS technologies.

The development of eCells derived from a eukaryotic system would significantly enhance the synthetic biologist's toolkit by offering an easily obtainable eukaryotic lysate. This CFPS system could be a significant expansion of the eCell capability with the ability to produce soluble proteins that are not accessible through an *E. coli*-based system. Moreover, enzymatic lysis of the yeast cell wall has been shown to preserve organelles and the nuclear envelope, which may enable more efficient CFPS. This aspect has significant implications for the expression of certain eukaryotic proteins that require post-translational modifications, such as glycosylation, disulfide bond formation, and proteolysis.

In addition to enhancing protein expression, implementing an autolysis system for yeast production could further increase the convenience of purifying and extracting protein from yeast systems. The notoriously difficult-to-rupture nature of yeast cell walls has been an issue

in the isolation of protein from yeast, particularly for non-conventional strains. An autolysis system would allow for the controlled degradation of the yeast cell wall, thereby facilitating protein isolation.

The proposed eukaryotic CFPS system would be a significant expansion of the eCell capability, with great potential for advancing fundamental research and industrial applications. While further work is required to optimize the system, the proposed yeast-based CFPS system holds great promise for producing complex eukaryotic proteins for use in diverse applications, including biopharmaceuticals and industrial biotechnology. Biosynthesis is a fundamental process by which living organisms, including plants, animals, and microorganisms, synthesize complex molecules from simpler building blocks. It involves a series of multistep enzymatic reactions that are tightly regulated and occur within specific cellular compartments. The discovery of natural pathways and advancements in gene technology have provided deeper insights into biosynthesis, allowing the manipulation of these pathways for the production of molecules that would otherwise be difficult to access using traditional synthetic chemical approaches.

In addition to metabolic engineering strategies, there has been a growing interest in developing biocatalytic cascades *in vivo* and *in vitro*. One such approach is the use of encapsulated cells (eCells), which are designed to hold enzymes within a semi-permeable polymer bag. This allows for the easy and cost-effective biosynthesis of high-cost small molecules by reconstituting enzymatic pathways. eCells are particularly useful in circumventing the need for purification of individual enzymes, as they can be added as a simple reagent and lysed by a freeze-thaw lytic step.

Overall, the development of eCells and other biocatalytic cascades has the potential to revolutionize the field of biosynthesis, making it more efficient, cost-effective, and accessible for the production of a wide range of valuable compounds.

Looking to the future, the integration of *in vitro* technologies has the potential to transform cell-free synthetic biology. The creation of eCells offers exciting opportunities for a variety of applications. By combining the principles of cellular metabolic engineering with cell-free reaction optimization, researchers can develop new approaches for metabolic engineering and other innovative developments in cell-free synthetic biology.

eCells are highly versatile systems that can be designed to mimic the characteristics of natural cells while providing greater flexibility and control. This system can be tailored to specific functions, such as enzyme production, drug delivery, and biosensing. They can also be used to study the fundamental principles of biological systems and to develop new biotechnologies.

The integration of eCells into cell-free synthetic biology offers several advantages over traditional approaches. By encapsulating enzymes and other biological components within a synthetic membrane, researchers can control the reactions and optimize the conditions for maximum efficiency. This approach also allows for the production of complex biological molecules that would be difficult to access using traditional synthetic chemical methods.