Building synthetic multicellular systems from the bottom-up

DISSERTATION

Zur Erlangung des akademischen Grades
Doctor of Philosophy (Ph.D.)

vorgelegt
der Fakultät Mathematik und Naturwissenschaften
der Technischen Universität Dresden

von
David T. Gonzales
geboren am 1. August 1987 in Toronto, Kanada

Vorgelegt am 7. Dezember 2021
Verteidigt am 2. Juni 2022

Die Dissertation wurde in der Zeit von Oktober 2017 bis Dezember 2021 im Max Planck Institut für Molekulare Zellbiologie un Genetik angefertigt.

Gutachter:

- Prof. A. Francis Stewart, Ph.D.
- Prof. Christophe Danelon, Ph.D.

Promotionskomission:

- Prof. Dr. Christian Dahmann (Vorsitzender)
- Prof. A. Francis Stewart, Ph.D.
- Prof. Dr. Thorsten Mascher
- Prof. Dr. Marion Ansorge-Schuhmacher
- T.-Y. Dora Tang, Ph.D.

Tag der Verteidigung: 2. Juni 2022

For my family and friends.

Acknowledgements

First and foremost, I would like to express my deepest thanks and appreciation to my PhD supervisors Dora Tang and Christoph Zechner. Together, they provided me with a great interdisciplinary research environment under the ELBE program. They gave me the freedom to learn at my own pace and provided constant support, even while starting their own families during the later stages of my PhD. I am grateful for the in-depth scientific discussions, and their patience and optimism towards me and my gradual progress.

I would also like to thank Carl Modes for the all the valuable feedback he provided throughout my PhD in our joint Modes/Zechner group meetings, and his honest and encouraging comments in my paper and thesis drafts. I would like to thank Ivo Sbalzarini and Francis Stewart for the guidance they provided during my TAC meetings. Their suggestions helped shape my project towards a clearer goal. I am especially grateful for the opportunity Ivo and Christoph gave me to tutor in their Stochastic Modeling and Simulation courses, which has been one of the most fun and fulfilling times during my PhD.

I am very thankful for the fruitful collaborations with Tom Robinson and Naresh Yandrapalli in MPI-KG, Potsdam. Naresh taught me how to use their microfluidic devices and provided all the microfluidic chips during my PhD. They helped me with my endless troubleshooting questions and were always prompt and reliable for my next request of chips.

I would also like to thank all my labmates in the Tang lab and Zechner lab. In particular, Giorgio Fracasso for teaching me how to make cell extracts, Yvonne Körner for helping me throughout my PhD with extract making and purifying proteins, Celina Love for showing me how to make my first liposomes, Basusree Ghosh for teaching me how to make RNA gels, Mengfei Gao and Archishman Ghosh for reviewing my thesis chapters and providing valuable feedback, and Lorenzo Duso and Tommaso Bianucci for helping me with questions on statistics and Julia. Even more than their scientific support, I am grateful for the friendships made in my PhD. I especially enjoyed all the lively discussions in the lab, our in-person and online lab meetings, lunches, and coffee breaks. The after-work swimming with Elisa and Tommaso and skating with Anne kept me healthy during paper and thesis writing!

All this work would have not been possible without the amazing facilities of MPI-CBG. I would like to thank the Light Microscopy Facility, Protein Expression, Purification, Characterization Facility, Scientific Computing Facility, Technology Development Studio, International Office, and PhD Office for their indispensable help inside and outside the lab.

Lastly, my biggest thanks to my family for their unconditional love and support. And to Ysa for being my constant companion throughout my PhD.

David T. Gonzales
Dresden, December 2021

Abstract

Biological cell populations, such as in tissues or microbial communities, are constantly subject to different sources of noise and variability. Despite this, multicellular systems are still able to function properly because cells coordinate with each other by communication. Using biological model systems to study this multiscalar process can be challenging because of their innate complexity. In this thesis, we address this challenge by building a synthetic multicellular system using bottom-up *in vitro* assembly approaches. Using this platform, we aim to study the effect of cell-to-cell communication to population variability in a minimal and simplified context. To achieve this, we require a synthetic cell population with (i) quantifiable gene expression dynamics, (ii) customizable population variability, and (iii) intercellular communication. Having these characteristics will allow us to test different initial configurations of population variability and monitor population gene expression dynamics with and without cell-to-cell communication. To generate these synthetic cell populations, reconstituted cell-free expression systems (CFES) are encapsulated into monodisperse-sized liposomes using double-emulsion microfluidics. Both transcription and translation levels are simultaneously monitored and quantified to develop models of cell-free gene expression dynamics and differentiate between bulk and encapsulated formats. Population variability was then incorporated by combining different batches of cells to create distinct subpopulations or by using a two-inlet double-emulsion microfluidic device to generate single populations with a large dispersion of encapsulated DNA template. Lastly, genetic circuits based on the quorum sensing system of *Vibrio fischeri* are used to implement diffusion-mediated intercellular signalling. Quorum sensing gene circuits in *Escherichia coli* extract-based CFES were tested in bulk and phase transfer-generated synthetic cells. Together with these experimental systems, corresponding models of synthetic cell populations that can account for population variability and secrete-and-sensing communication are developed using mixed-effects models and moment dynamics. Overall, this work leverages CFES and microfluidic technologies to reproducibly generate a simplified *in vitro* model of multicellular systems that can be easily monitored spatiotemporally to study multi-scalar processes.

Zusammenfassung

Biologische Zellpopulationen, z.B. in Geweben oder mikrobiellen Gemeinschaften, sind ständig verschiedenen Quellen von Rauschen und Variabilität ausgesetzt. Trotzdem sind multizelluläre Systeme in der Lage, ordnungsgemäß zu funktionieren, weil sich die Zellen durch Kommunikation miteinander abstimmen. Die Verwendung biologischer Modellsysteme zur Untersuchung dieses multiskalaren Prozesses kann aufgrund ihrer angeborenen Komplexität eine Herausforderung darstellen. In dieser Arbeit gehen wir diese Herausforderung an, indem wir ein synthetisches multizelluläres System mit Hilfe von Bottom-up-*in vitro*-Assembly-Ansätzen aufbauen. Mit Hilfe dieser Plattform wollen wir die Auswirkungen der Kommunikation von Zelle zu Zelle auf die Populationsvariabilität in einem minimalen und vereinfachten Kontext untersuchen. Um dies zu erreichen, benötigen wir eine synthetische Zellpopulation mit (i) quantifizierbarer Genexpressionsdynamik, (ii) anpassbarer Populationsvariabilität und (iii) interzellulärer Kommunikation. Mit diesen Eigenschaften können wir verschiedene Ausgangskonfigurationen der Populationsvariabilität testen und die Genexpressionsdynamik der Population mit und ohne Zell-zu-Zell-Kommunikation beobachten. Um diese synthetischen Zellpopulationen zu erzeugen, werden rekonstituierte zellfreie Expressionssysteme (CFES) mit Hilfe der Doppelemulsions-Mikrofluidik in monodisperse Liposomen eingekapselt. Sowohl die Transkriptions- als auch die Translationsraten werden gleichzeitig überwacht und quantifiziert, um Modelle für die Dynamik der zellfreien Genexpression zu entwickeln und zwischen Bulk- und verkapselten Formaten zu unterscheiden. Die Variabilität der Populationen wurde dann durch die Kombination verschiedener Zellchargen zur Bildung unterschiedlicher Subpopulationen oder durch die Verwendung einer mikrofluidischen Doppelemulsionsvorrichtung mit zwei Einlässen zur Erzeugung einzelner Populationen mit einer großen Streuung der eingekapselten DNA-Vorlage einbezogen. Schließlich werden genetische Schaltkreise auf der Grundlage des Quorum-Sensing-Systems von *Vibrio fischeri* verwendet, um diffusionsvermittelte interzelluläre Signalübertragung zu implementieren. Quorum-Sensing-Genkreisläufe in CFES auf der Basis von *Escherichia coli*-Extrakten wurden in synthetischen Zellen getestet, die durch Bulk- und Phasentransfer erzeugt wurden. Zusammen mit diesen experimentellen Systemen wurden entsprechende Modelle synthetischer Zellpopulationen entwickelt, die die Populationsvariabilität und die Sekretions- und Sensing-Kommunikation mit Hilfe von Mixed-Effects-Modellen und Momentendynamik berücksichtigen können. Insgesamt nutzt diese Arbeit CFES- und Mikrofluidik-Technologien, um reproduzierbar ein vereinfachtes *in vitro*-Modell multizellulärer Systeme zu erzeugen, das leicht raum-zeitlich überwacht werden kann, um multiskalare Prozesse zu untersuchen.

Contents

Preface

Cells in biological multicellular systems are never exactly identical. Even in an isogenic cell population, cells are constantly subject to different sources of noise and variability [71]. Despite this, biological multicellar systems, are still able to function properly because cells can coordinate by intercellular communication. In this work, we are interested in studying the effect of cell-to-cell communication to population variability using a bottom-up approach. A bottom-up approach involves reconstituting biological systems using simple building blocks, such as synthetic or purified components of the cell and non-living molecular species. This approach has been used for decades in the field of biochemistry to help us understand the functions of individual cellular components in biology. By isolating the components of interest, biological systems can be more easily investigated in a simplified context. In more recent years, reconstitution techniques have been applied in the field of synthetic biology to building minimal synthetic cells from the bottom-up. Building towards a synthetic cell offers new insights and questions on how the mechanisms of life can work as a whole within a more complex system. It has pushed us to think about how to compartmentalize reactions [192, 261, 255, 49, 159], shape cells [263, 139, 53, 74], maintain out-of-equilibrium systems [147, 21, 24], and drive cell division [155, 92, 245]. It has also brought about close collaborations with other fields in physics, engineering, and biotechnology to utilize or develop new tools such as microfluidics [52, 20, 289], cell-free expression systems (CFES) [227, 145], and 3D printing [273]. Progress in building synthetic single cells further motivated work in creating multicellular systems. In biology, multicellularity enables larger and more complex lifeforms, but also requires additional means of intercellular communication, 3D organization, and differentiation to coordinate cells and control growth and population variability [101]. Recent efforts have already demonstrated some of these aspects in bottom-up assembly of synthetic multicellular systems. For instance, 2 and 3D-printed structures composed of lipid-stabilized water-in-oil droplets have been made to express membrane pore proteins to facilitate cell-cell signalling [29, 66]. Coupled with positive or negative feedback genetic circuits, these can be made to simulate simple artificial morphogen signalling and differentiation in a defined multicellular geometry [66, 65]. Artificial cell mimics made of a porous polymer shell containing a nucleus-like DNA-hydrogel have also been made to communicate via diffusible protein signals to activate gene expression [184]. Quorum sensing artificial cells that can produce and sense diffusible signalling molecules were interfaced with actual quorum sensing bacteria [221, 148, 206]. These examples demonstrate our growing capability to assemble complex synthetic multicellular systems with biological features. In the future, bottom-up synthetic systems could form the basis of unique multicellular biomaterials or an *in vitro* platform where intercellular signalling systems can be reconstituted and tested in a minimal multicellular context.

Aim and approach

We aim to build a platform to study the interplay between cell-to-cell communication and population variability using bottom-up assembly approaches. To achieve this, we require a synthetic cell population with (i) quantifiable gene expression dynamics, (ii) customizable population variability, and (iii) intercellular communication. Having these characteristics will allow us to test different initial configurations of population variability and monitor population gene expression dynamics with and without cell-to-cell communication. Using recent advances in double-emulsion microfluidics, we encapsulated CFES into large populations of monodisperse-sized liposomes to generate synthetic cell populations. Gene expression dynamics in these synthetic cell populations were quantified in both transcription and translation levels using a Broccoli RNA aptamer and mCherry protein reporter. Different coarse-grained models of resource-limited cell-free gene expression were then tested and fit against the quantified gene expression dynamics. We used the Akaike information criterion (AIC) for model selection and analyzed identifiability and confidence intervals of the fit parameters by profile likelihood analysis. The selected model was used to quantify differences between bulk and encapsulated gene expression rate parameters. We then produced synthetic cell populations with different distributions of encapsulated DNA analogous to biological cell populations that exhibit heterogeneity and variability by mixing different batches of synthetic cells together or by using a two-inlet microfluidic device. Intercellular communication in the synthetic cell population was incorporated by expressing a quorum sensing genetic circuit with the CFES. Lastly, we developed models that account for population variability and secrete-and-sensing communication using mixed-effects models and moment-based analysis that can be applied with our experimental platform.

Outline

The contents of this thesis is organized as follows. In Chapter 1, we introduce the field of bottom-up synthetic biology, discuss the key components needed to build synthetic multicellular systems, and present state-of-the-art examples. Chapter 2 provides a background on cell-free expression systems (CFES) and their encapsulation into lipid vesicles or liposomes, which are the main building blocks used in this work. Chapter 3 presents our workflow to generate synthetic cell populations using double-emulsion microfluidics to encapsulate CFES. We show how gene expression dynamics can be monitored and quantified in synthetic cell populations. This is then used to develop a coarse-grained model of resource-limited cell-free gene expression and infer transcription and translation rate parameters in both bulk and encapsulated CFES formats. In Chapter 4, we integrate a quorum sensing gene circuit in CFES and encapsulate these into synthetic cells with the goal of building intercellularly communicating populations. We then show how synthetic cell populations with different distributions of encapsulated DNA can be prepared by mixing different populations together to get distinct subpopulations, or by using a two-inlet microfluidic device to generate increased variability in a single population. Lastly, in Chapter 5, we introduce a mixed-effects model to account for cell-to-cell variability within a synthetic cell population and a moment-based model of communicating cell populations, which can account for population variability and stochastic gene expression.

Related publications

In collaboration with other groups, the work presented in this thesis has resulted in three publications. Chapter 1 is based on the review article:

Gonzales DT, Zechner C, Tang TYD. **Building synthetic multicellular systems using bottom-up approaches.** *Current Opinion in Systems Biology*, 24: 56-63, 2020 [96].

Chapter 3 shows results from:

Gonzales DT, Yandrapalli N, Robinson T, Zechner C, Tang TYD. **Cell-free gene expression dynamics in synthetic cell populations.** *ACS Synthetic Biology*, 11(1):205-215, 2022 [95].

Parts of Chapter 5 are from:

Gonzales DT, Tang TYD, Zechner C. **Moment-based analysis of biochemical networks in a heterogeneous population of communicating cells.** *IEEE 58th Conference on Decision and Control (CDC)*, 939-944, 2019 [94].

Chapter 1

Bottom-up multicellular systems*

Biological systems represent a key example of how simple building blocks make complex multiscale systems. Single cells coordinate with each other on the molecular level through intercellular communication and feedback loops to form multicellular systems that can exhibit population-level behaviour and functions. For instance, bacterial communities use quorum sensing to coordinate gene expression in response to population-level properties such as the total concentration of secreted autoinducer molecules [281]. Oscillations in biofilm growth of a microbial community arise from long-range metabolic co-dependence between interior and peripheral cells in the colony [157]. Emergent properties that arise from the dynamics of single cells and their intercellular interactions can also be observed in more complex biological systems. The formation of digits during limb development is proposed to be caused by periodic patterns from a Turing reaction-diffusion network of morphogens within the limb bud tissue [207]. Boundaries between cell compartments can further provide a physical means of discretizing space in multicellular systems such as observed in the development of skin scale colour pattern in lizards [164]. These selected examples provide a mechanistic understanding of how emergent behaviour in biological systems can arise from intercellular communication, reaction-diffusion, and multi-compartmentalization. If these key features can be abstracted and built into bottom-up synthetic multicellular systems, then emergent properties could potentially be assembled from scratch (Figure 1.1A). Although lacking in complexity compared to living systems, bottom-up design and construction of synthetic multicellular systems can help describe the essential elements that recapitulate emergent properties [219]. A bottom-up approach provides a well-defined and simple experimental system that can help us understand the emergence of complex behavior from first principles [98]. In the past decade, significant progress has been made in building synthetic cell units by encapsulating chemical or enzymatic reactions within micron-sized compartments [214, 84]. There is now an extensive toolkit of membrane-bound compartments, as well as reactions which can be encapsulated and supported within the compartments, that utilise a wide range of different molecules. For example, membrane-bound compartments can be generated by the self-assembly of lipids in water (liposomes) or oil (water-oil emulsions and droplet interface bilayers), amphiphilic

*This chapter is based on the review article: Gonzales DT, Zechner C, Tang TYD. **Building synthetic multicellular systems using bottom-up approaches.** *Current Opinion in Systems Biology*, 24: 56-63, 2020 [96].

di-block copolymers (polymersomes), and protein-polymer conjugates (proteinosomes) (Figure 1.1B) [240]. These compartments can encapsulate chemical reactions such as DNA strand displacement (DSD) reactions [223], Polymerase/Exonuclease/Nickase Dynamic Network Assembly (PEN DNA) reactions [178, 16], cell-free transcription and translation or cell-free expression systems (CFES) [143], and enzyme cascades [70, 31] where reaction products can act as intercellular signalling molecules (Figure 1.1C). Improvements in microfluidic [57, 265] and droplet printing technologies [28] have further enabled the generation of homogeneous or user-defined heterogeneous synthetic cell populations [20, 10, 286] in custom 2D or 3D arrangements. This offers a robust and versatile platform for manipulating single cell reaction parameters and population heterogeneity with a high level of control that is not easily accessible or attainable in standard bulk methodologies. Taken together, these technologies have enabled bottom-up synthetic biology to produce and characterize compartmentalised reactions as individual and isolated cells and provide the basis for developing synthetic multicellular systems with networks of interacting synthetic cellular units. In this chapter, we focus on recent advances in the emerging area of bottom-up approaches to create biologically-inspired systems that bridge scales from single cells to multicellular systems. Specifically, we consider how intercellular communication and feedback loops can be integrated into populations of synthetic cells to invoke coordinated behaviours that are synonymous with multicellular systems.

1.1 Intercellular communication

Using signalling molecules, individual cells are able to send and receive information from other cells, or the ensemble, and coordinate functions as a population [46]. There are a number of routes that biological cells use to communicate efficiently. Communication can occur by direct diffusion of signalling molecules through membranes and membrane channels. For non-membrane diffusible signalling molecules, membrane-bound protein receptors can relay extracellular signals into the cell by signal transduction or via membrane vesicle transport processes such as endo/exocytosis of outer membrane vesicles (OMVs) that are found in bacteria [262]. These biological mechanisms often rely on complex processes and molecular structures which can be experimentally challenging to reconstitute within *in vitro* systems. Here we discuss strategies to integrate intercellular communication into minimal synthetic systems.

1.1.1 Membrane-diffusive signalling

Most examples of synthetic cell-cell communication utilise diffusion-mediated communication as a route to coordinate behaviours between individual cells [15]. Compartmentalised reactions in lipid vesicles or DIBs can be coordinated or triggered via passive diffusion of membrane-diffusible small molecules, or diffusion facilitated through incorporated membrane protein pores and ion channels (*e.g* α-hemolysin (αHL) pore [1, 254], DNA origami nanopore [260], MscL channel [109]) for larger or charged molecules (Figure 1.2A-B). The substrate or product of any enzyme reaction can act as a signalling molecule that is sequentially modified by a series of encapsulated enzymatic reactions as it travels between different compartments, thereby organizing the cells into a primitive network [70, 31]. Signalling molecules generated *in situ* by CFES, as demonstrated

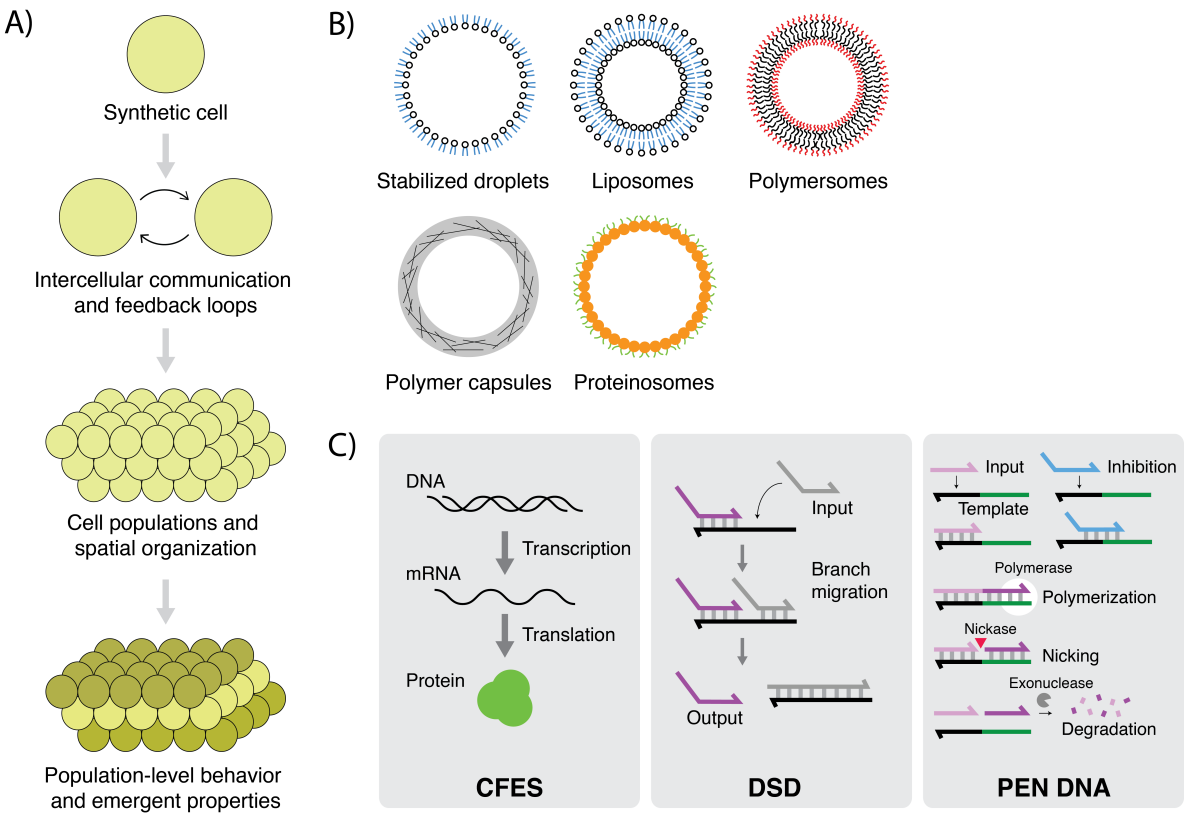

Figure 1.1: Building bottom-up synthetic multicellular systems. (A) Synthetic cells must communicate intercellularly and have feedback systems within cell populations to exhibit population-level emergent behaviour. (B) Compartments can be formed from self-assembling bilayers/monolayers of lipids to make stabilized droplets or liposomes, diblock copolymers to make polymersomes, and cross-linked polymers to make polymerized capsules or protein-polymer conjugate proteinosomes. (C) Examples of reaction types, such as cell-free expression systems (CFES), DNA strand displacement (DSD) reactions, and polymerase/exonuclease/nickase dynamic network assembly (PEN DNA) reactions, that can be encapsulated in synthetic cells.

by the production of N-acyl homoserine lactones (AHL) by AHL synthase (LuxI) expressed from a DNA plasmid, can recapitulate bacterial quorum sensing gene circuits within synthetic cells [148]. These examples demonstrate the open and modular platforms in chemistry and synthetic biology that can be used to create communicating cell populations with life-like features.In general, the mobility of the signalling molecule is dependent on its diffusive properties with respect to the membrane. The different molecular building blocks used to construct the membrane can impart specific permeability properties to the compartment. Liposomes allow only small and uncharged molecules to diffuse through the lipid bilayer membrane. In contrast, compartments formed from porous acrylate polymer membranes permit transport of larger macromolecules, such as TetR-sfGFP (50.1 kDa) and T3 RNA polymerase proteins (98.8 kDa), to neighbouring cells [184] (Figure 1.2C). A strength of synthetic cellular systems is that their membrane properties are inherently tuneable, allowing the diffusivity of signalling molecules to be modulated at will. For example, the membrane permeability of proteinosomes can be controlled by incorporating environment-responsive molecules that affect the membrane protein-polymer structure [298, 205]. Reaction molecules inside the synthetic

cell can also be modified to provide specific containment or diffusive properties. DNA-based reactions, such as DSD and PEN DNA reactions, have been compartmentalised in water-in-oil droplets [108], proteinosomes [121], and terpolymer-stabilised coacervate droplets [161]. The short single-stranded DNA (ssDNA) molecules (<100 bp) can typically diffuse through proteinosome membranes and out of coacervate droplets. However, the membrane can retain DNA template molecules that have been immobilised or localised inside these compartments by binding them to larger streptavidin beads [121, 291] or self-assembling supramolecular nanoscaffolds [161] (Figure 1.2D).

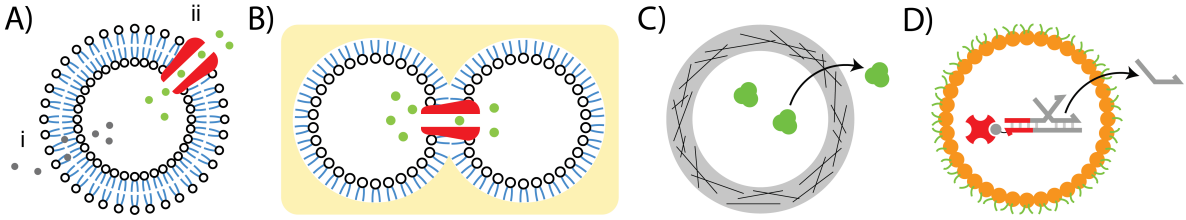

Figure 1.2: Diffusion-mediated intercellular communication between synthetic cells. (A) Membrane-diffusible molecules (i) and larger or charged molecules that are transported through membrane channels (ii) can act as intercellular signalling molecules. (B) Membrane channel proteins in DIBs can facilitate diffusion of signalling molecules for cell-cell communication. (C) Larger molecules such as expressed enzymes can diffuse through porous acrylate polymer membranes. (D) Free ssDNA (<100 bp) from DSD reactions can pass proteinosome membranes as signal molecules, while streptavidin-bound DNA templates are localized inside the proteinosome.

1.1.2 Signal transduction across the membrane

While it has been shown that passive diffusion across a membrane can be effective in driving cell-cell communication, this also limits the ability of the compartment to maintain concentration gradients between the intercellular and extracellular space. This is a key feature of biological systems and important for maintaining a potential for generating energy via proton-motive forces across the membrane. Biological cells use signal transduction systems to internalise signals while maintaining the separation between intra and extracellular environments. This is facilitated by membrane proteins (*e.g.* histidine kinases, GPCRs, tyrosine kinase receptors) that oligomerize, transduce, or change conformation in response to ligand binding. While many natural membrane transducer proteins have been reconstituted in lipid membranes [191] or in polymersomes [171] for structural and functional studies, integrating these into a synthetic cell is challenging because it requires further coupling of the transducer to reaction networks within the cell. Instead, synthetic membrane transducers, which are simpler and easier to reconstitute, have been used in artificial cell applications [271]. Bernitzki and Schrader (2009) used two transmembrane units that dimerize with a di/tri-cationic primary messenger which results in a FRET signal between the inner leaflet headgroups [22]. More recently, Lister *et al.* (2017) designed a synthetic membrane transducer composed of a hydrophobic scaffold with an exposed ligand-binding pocket at one end, and a fluorescent pyrene pair reporter group at the opposite end. Binding of a carboxylate ligand induces a conformational change that results in a fluorescent readout due to an increase of proximity between the pyrene pair [154]. Langton *et al.* (2017) designed

membrane translocators that, depending on the pH of the environment, move between the outer and the inner leaflet of the bilayer membrane to expose active headgroup sites. Increasing the pH deprotonates the outer leaflet headgroup making it apolar [141]. This allows the translocation of the transducer exposing the active site to the inner leaflet for the response reaction inside the synthetic cell [142]. A similar artificial translocator has also been designed to respond to Cu(II) ion availability [140]. These examples demonstrate the possibilities of designing novel transducers which can be integrated into microcompartments for specific and specialised modes of communication between synthetic cells. For example, contact-dependent signalling could be engineered in synthetic cells by having both ligands and receptors bound to the membrane similar to the synthetic Delta-Notch signalling system that has been implemented in mammalian cells [182].

1.2 Spatial organization of cell populations

As intercellular communication is predominantly driven by the diffusion of signalling molecules between cells - the distance and arrangement within a 1, 2, or 3-dimensional space can significantly affect the strength and degree of cell-cell interactions within the population. Therefore, rational control of the spatial organization of synthetic cells can regulate the diffusive signalling properties and the overall population behaviour. One approach to achieving this is by tuning and regulating adhesion between compartments. Lipid vesicles will spontaneously adhere to one another at small distances by weak intermolecular forces and osmotic deflation [195]. The functionalization of a lipid membrane with light-activated adhesion protein pairs (iLID/Nano and nMagHigh/pMagHigh) [39, 125] or complementary DNA linkers [229] can promote specific adhesion between different liposome species (Figure 1.3A). Adhesion can also facilitate contact-dependent communication by direct exchange of cellular contents through fusion events [1] offering an alternative route to cellular communication compared to diffusion. Moreover, cell-cell adhesion can facilitate the construction of higher-order multicellular structures as demonstrated in prototissue spheroids formed by chemically cross-linking proteinosomes together [91] or multicompartmentalized lipid vesicles generated from multicompartment double emulsion templates using microfluidics [48] (Figure 1.3B).

Recent progress in the area of microfluidics and 3D droplet printing has enabled control over spatial localization of compartments. Using microfluidic traps, hundreds of cells can be packed together into chambers [290] or individually arranged into 2D grids with specified distances [126, 121, 291] (Figure 1.4A). For example, Joesaar *et al.* (2019) and Yang *et al.* (2020) used 2D arrays of proteinosomes encapsulating DSD reactions to create populations that communicate by diffusive signalling. By selectively triggering different signalling sources (sender cells), the integrated spatiotemporal activation of the receiver cells in the population can be observed [121, 291]. Aside from localizing the single cell units, these microfluidic traps also allow a quick exchange of external solution by flow through. Droplet printing has further opened exciting possibilities to generate specific patterns and structures of 3D synthetic tissues. Villar *et al.* (2013) was one of the first to produce custom 3D structures composed of droplet interface bilayers (DIBs) by printing water droplets into an oil phase containing lipids. The lipids self-assemble on the water-oil interface to generate a lipid monolayer that stabilise the

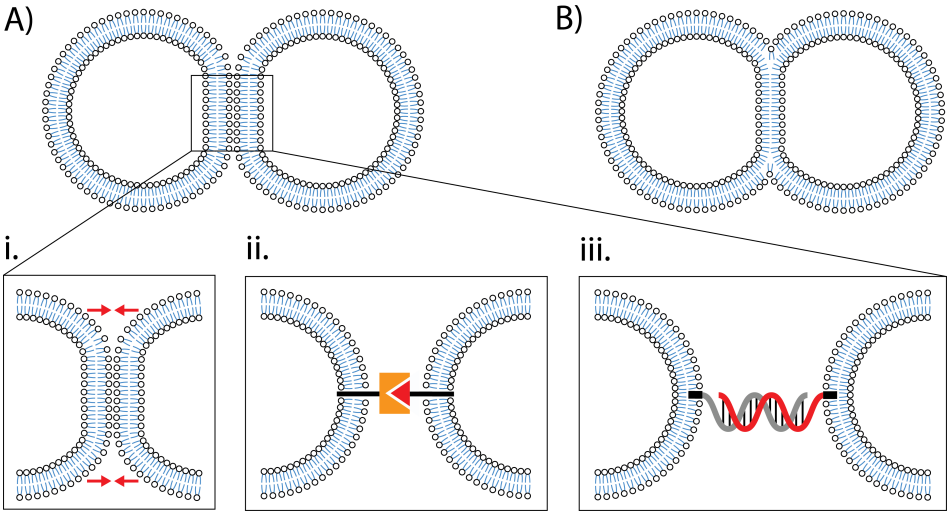

Figure 1.3: Cell-cell adhesion in liposome synthetic cells. (A) Liposomes can be directly adhered together using (i) attractive electrostatic interactions, (ii) cell adhesion molecules, and (iii) hybridizing DNA linkers. (B) Multicompartment lipid vesicles can also form multicellular systems where each compartment shares a lipid bilayer with its neighbour.

droplets. Droplet interface bilayers are formed when the droplets come into contact with one another [28]. Different packing arrangements, such as hexagonal, square, or amorphous packing, can be controlled by the balance between surface tensions of the lipid monolayer and bilayer which determines the equilibrium contact angle between DIBs [9] (Figure 1.4B). These 3D structures can also be modularly constructed and then assembled into larger centimeter sized synthetic tissues [7]. Additional properties and features of the 3D-printed DIBs can be incorporated by encapsulating different components into the cellular population. For example, 3D structures of DIBs with different osmolarities [273] and hydrogel droplets [62] were shown to fold and curl due to the shrinking and growing of individual droplets reacting to osmotic pressure and light or temperature, respectively. Booth *et. al* (2016) encapsulated CFES with a light-activated DNA promoter for an αHL gene and selectively connected individual DIBs to each other by localised light activation [29]. Using a two-step dewetting process, these 3D-printed DIBs can be further transferred into an aqueous phase to create a multicellular lipid vesicle structure [8]. This provides many new opportunities to create synthetic tissues that, unlike DIB structures immersed in an oil phase, can interact with the external environment. This was demonstrated in the same study by integrating αHL pore protiens in the multicellular structures to release a fluorescent cargo or sense Ca^{2+} ions in the outer aqueous environment. These examples demonstrate the high degree of spatial control that is possible within synthetic tissues and will be useful to study how inhomogeneous properties in interacting cellular units can affect population behaviour.

1.3 Feedback in multicellular systems

Many examples of intercellular communication in artificial cell populations are implemented as a unidirectional or sender-receiver relationship between cells. Integrating feedback loops couple all cells together into a closed loop control system that can regulate the population and help in the self-maintainence of out-of-equilibrium behaviour

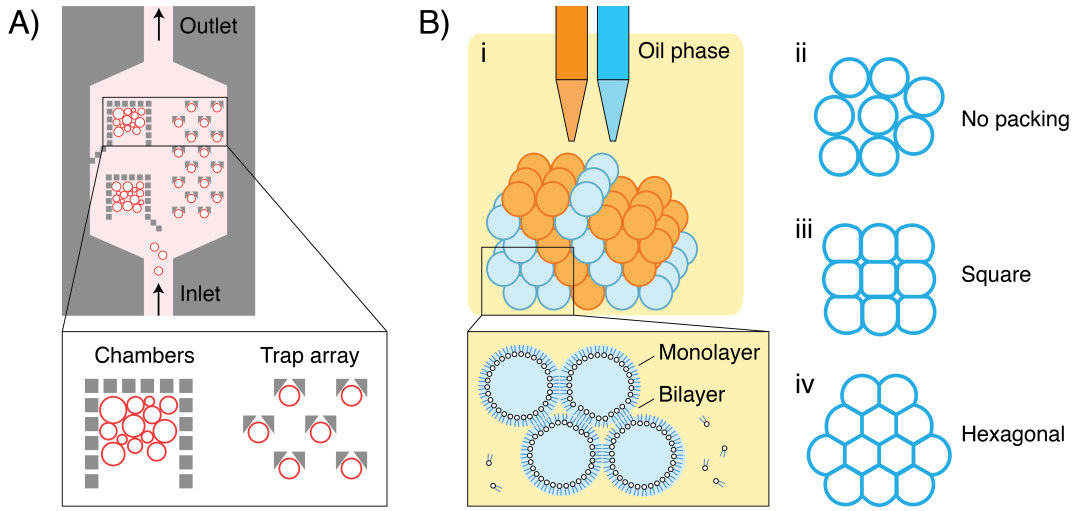

Figure 1.4: Spatial organisation of synthetic cell populations. (A) Synthetic cell populations can be packed together in microfluidic chambers [290] or trap arrays in a regular grid [126]. (B) (i) 3D-printed droplets can be used to make custom tissue-like structures of DIB synthetic cells [273]. The arrangement of the DIBs can be controlled by the balance between surface tensions of the lipid monolayer and bilayer into (ii) no packing, (iii) square packing, and (iv) hexagonal packing [9]. All illustrations are based on figures in their respective references.

within the system. In recent years, communicating artificial cell populations have been engineered with bidirectional communication or feedback systems to self-regulate population behaviour. For example, by coupling two populations of proteinosomes that activate and inhibit each other using DSD reactions, Joesaar *et al.* (2019) observed a self-regulated transient response in the activator cell population [121] (Figure 1.5A). Dupin *et al.* (2019) built sender and receiver artificial cells which mimicked cell differentiation with a positive feedback switch [66]. Sender cells contained arabinose that diffuses into the receiver cells through αHL pore proteins, further triggering αHL expression by arabinose induction in a positive feedback loop (Figure 1.5B). Qiao *et al.* (2019) used two different predation strategies to couple different synthethic cellular compartments to create an intercellular feedback system [204]. The cell population consists of proteinosomes containing glucose oxidase, pH-insensitive coacervate droplets that are attached to the proteinosome, and pH-sensitive coacervate droplets containing proteinase K. The reaction is initiated by adding glucose, which is oxidised by glucose oxidase to produce H_2O_2 in the proteinosome. This lowers the local pH by the decomposition of H_2O_2 to H^+ ions in water and disassembles the pH-sensitive coacervates, releasing proteinase K into solution. Proteinase K is then sequestered into the pH-insensitive coacervate and degrades the proteinosome membrane. This results in the dilution of glucose oxidase enzymes in a negative feedback (Figure 1.5C). These examples demonstrate our growing capability to integrate feedback control into artificial multicellular systems. Further application of feedback loops could provide more complex behaviour in bottom-up synthetic cell populations. In synthetic microbial communities, bistability [266], oscillations [222], gene expression noise reduction [27], and robust control [13] were achieved by engineering intercellular feedback loops.

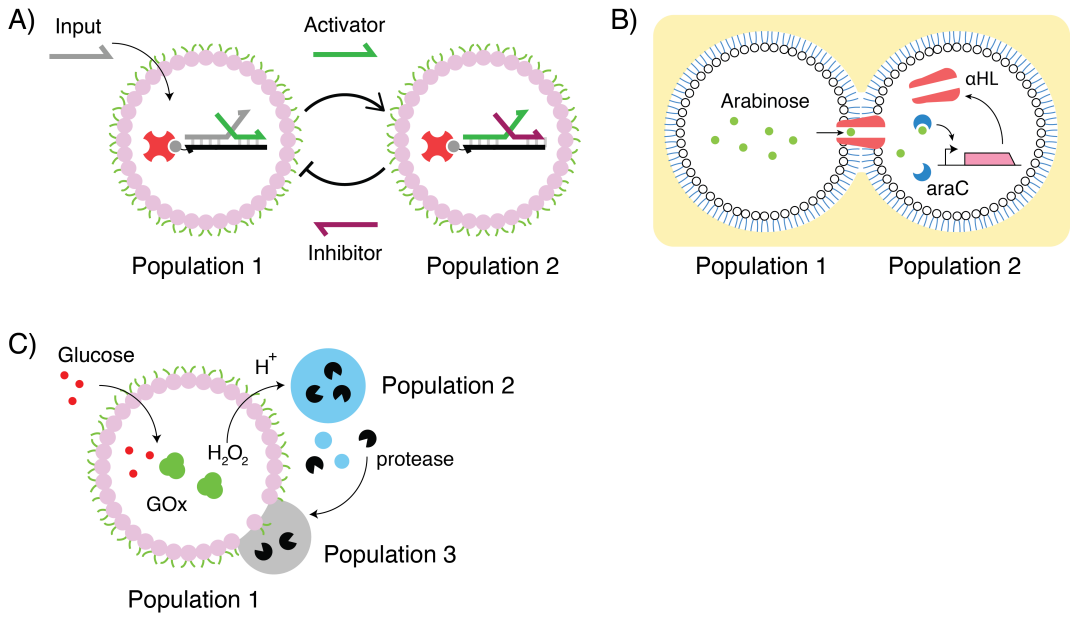

Figure 1.5: Intercellular feedback loops in synthetic cells. (A) Negative feedback in proteinosome populations with DNA strand displacement (DSD) reactions. Upon addition of the input strand, population 1 releases the activator strand for population 2, which responds by releasing the inhibitor strand to deactivate population 1 [121]. (B) Positive feedback in DIBs with CFES. Population 1 are sender cells containing arabinose inducer molecules that diffuse through αHL pores to population 2. This induces αHL expression in a pBAD-αHL gene circuit to promote further diffusion of arabinose [66]. (C) Negative feedack or response-retaliation in a synthetic protocell community. Upon addition of glucose, glucose oxidase (GOx) encapsulated in proteinosomes (Population 1) produces H_2O_2. Decomposition of H_2O_2 releases H^+ ions that disassembles Population 2 coacervates due to the pH change and releases protease enzymes into solution. The proteases are then sequestered by another coacervate population that is attached to the proteinosomes (Population 3). Because of the proximity of the proteases to the proteinosomes, the proteinosome membrane is digested and the GOx is released and diluted in solution and deactivated [204].

1.4 Conclusions

So far, we have focused on recent developments in spatial organization, intercellular communication and feedback systems as key features for the design and construction of minimal synthetic multicellular systems. A key challenge in the field is the maintenance of out-of-equilibrium reactions. The integration of degradation and turnover of reaction products [226] and improved energy and resource regeneration [234], into communicating populations of synthetic cells will provide a route to produce sustained out of equilibrium behaviours. Moreover, population heterogeneity, stochastic reactions [190], dynamic compartments [245], and the integration of intracellular feedback loops from gene circuits engineered in natural cells [153, 232] will increase the robustness and diversity of behaviours within populations of minimal synthetic cells. In the later chapters, we address two aspects by integrating population heterogeneity and intercellular communication in a synthetic cell population. With the development of more complex and integrated experimental systems, collaborations with theoreticians will provide a route to design, build and test emergent properties in multicellular systems [98]. For

example, dynamic models of synthetic multicellular systems can help predict which ranges of parameters will result in emergent properties. In turn, bottom-up synthetic multicellular systems can be more easily modified to achieve these parameter values as compared to biological model systems. The exciting development of minimal synthetic multicellular systems opens new avenues in synthetic biology, theory, biology and materials in the upcoming years. It can help develop new tools for complex multiscalar assembly, make bio-inspired materials and devices, and provide a minimal multicellular context to study intercellular signalling systems. In our specific case, we aim to build a simplified platform to study the effect of cell-to-cell communication to population variability.

Chapter 2

Building blocks: cell-free expression and liposomes

The experimental work in this thesis is based on two technologies: (1) cell-free expression systems (CFES) and (2) liposome encapsulation. In this chapter, we present a background on both topics that make up the building blocks of our synthetic cell populations. CFES are *in vitro* reconstitutions of the transcription and translation machinery found in cells. These *in vitro* systems are open and easily manipulated, which have allowed researchers to investigate the inner workings of the cell without the added complexity of cell growth and division. We use *E. coli*-based CFES that are either derived from cell extracts or reconstituted from purified components. We next describe the different methods used to encapsulate CFES into liposomes. Encapsulation of CFES into lipid vesicles or liposomes compartmentalizes these reactions into single cell units. Apart from closely resembling biological cells, the lipid bilayer membrane of liposomes provides a semi-permeable boundary for small singalling molecules to diffuse through and the potential for integration of functional membrane proteins or pore proteins.

2.1 *In vitro* reconstitution of gene expression

The systematic reconstitution of biological processes involve purifying and recombining biological components or molecular species to recreate a function. One of the earliest studies of reconstitution was conducted by Eduard Buchner (Tübingen, 1897), who showed that fermentation could still occur in cell-free yeast extracts [30]. Although conducted using crude cell extracts, Buchner strengthened the growing idea that processes occurring in living organisms can be summarized into a collection of biochemical reactions. Cell extracts provide an open system of the native cellular machinery, making it possible to conveniently investigate the underlying biochemistry in the cell. Hoagland *et al.* (1957) used extracts of rat liver and mouse tumor cells with microsomes to study peptide bond formation in protein synthesis [111]. Kirsch *et al.* (1960) further demonstrated that the ribosomes dettached from the microsomes would still function to incorporate amino acids into polypeptides [134]. Notably, the genetic coding mechanism was discovered with the help of cell extract studies. Nirenberg and Matthaei (1961) demonstrated how polyuridylic acid RNA templates are translated into poly-L-phenylalanine peptides in *Escherichia coli* extracts - hinting at the translation of the UUU codon to phenylalanine in the genetic code [189]. In later years,

DNA replication was reconstituted using *X. laevis* egg (1981) [81] and *E. coli* (1986) [26] cell extracts. One of the main difficulties of using cell extracts was that it always had to be prepared fresh for each experiment. To address this, Matthaei and Nirenberg (1961) established a protocol to prepare *E. coli* cell extracts that are capable of amino acid incorporation into protein even after storage for several months at -15°C [170]. Extracts prepared were named S-30 and S-100, according to the supernatant fractions obtained from crude *E. coli* cell extracts after centrifugation speeds of 30,000×g and 100,000×g, respectively. With this method, Nathans *et al.* (1962) showed the first fully synthesized protein (coat protein of coliphage f2) from RNA templates using *E. coli* cell-free extracts [183]. DeVries and Zubay (1967) next showed coupled transcription and translation from a DNA template (encoding β-galactosidase) in cell-free extracts [54]. The method for the S-30 extract was further modified and optimized by the laboratory of Zubay to improve the expressed protein yield [299].

Today, the process and yield of cell-free *E. coli* extracts have improved dramatically. Methods to produce cell-free *E. coli* extracts are simpler as compared to earlier protocols, making it more accessible and cost-effective for researchers. Cell-free reactions from *E. coli* extracts can cost as little as 0.0105 USD/μL (for materials and reagents only) and produce up to 0.75 mg/mL of protein [249]. Other more recent studies report 0.9 mg/ml [149], 2.3 mg/ml [37], and 4 mg/ml [87] of expressed protein in a cell-free batch reaction. In comparison, the protocol reported by Zubay (1973) produced cell-free extracts that expressed approximately 1 μg/mL of protein [299]. Extract protocols have also been developed with other prokaryotic organisms such as *Vibrio natriegens* [50, 287], *Bacillus subtilis* [128], *Pseudomonas putida* [279], *Streptomyces venezuelae* [151], and *Bacillus megaterium* [180]. These are typically used as a fast prototyping platform for specific organisms of interest or to provide new functionalities not native in *E. coli* extracts. For example, the marine bacterium *V. natriegens* has a reported doubling time of 10 minutes [68] - the fastest growing bacterium known to date. Due to this extremely fast growth rate and potentially enhanced protein translation rates, it has gained interest for applications in recombinant protein expression [285]. *Streptomyces*-based cell-free expression systems are sought because of the capability of *Streptomyces* to express natural product gene clusters and high GC-content genes [151]. Different extracts can even be pooled together to provide diverse regulatory mechanisms within one reaction [292]. Cell-free expression systems are also produced from eukaryotic organisms such as yeast [112], wheat germ [252], insect [72], HeLa [174], rabbit reticulocyte [243], and plant cells [33, 34]. Compared to prokaryotic-based extracts, eukaryotic-based extracts have a greater capacity to carry out posttranslational modifications necessary for many functional and complex proteins. For instance, cell-free expression systems from *Spodoptera frugiperda 21* (Sf21) insect and Tobacco Bright Yellow 2 (BY-2) plant cell extracts are able to actively glycosylate expressed proteins [257, 34]. Recent advances however have also been made using extracts from engineered *E. coli* strains pre-enriched with glycosylation components such as oligosaccharyltransferases and lipid-linked oligosaccharides. Cell-free expression systems made from these strains only require the addition of the DNA template encoding the protein-target for glycosylation [119, 129]. Eukaryotic systems are also suspected to provide a more favorable environment for the proper folding of multi-domain proteins as a result of slower polypeptide elongation rates and cotranslational mechanism of chaperones. This

was further supported by a systematic comparison between *E. coli* and wheat germ cell-free expression of random multi-domain proteins [110]. The main disadvantage of eukaryotic cell-free extracts are their protein expression yields and difficult preparation methods relative to *E. coli* extracts. For instance, a commercial Sf21 insect cell extract kit (TnT T7 Insect Cell Extract Protein Expression System, Promega) can produce up to 75 μg/mL, while a Tobacco BY-2 cell-free expression system can produce up to 100 μg/mL of protein from a DNA template [34], which are significantly lower than *E. coli* extracts. More recently however, a Tobacco BY-2 cell-free expression system was commercialized and further optimized to produce up to 3 mg/mL of protein (ALiCE Cell Free Protein Expression Kit, LenioBio).

Cell-free expression systems have undoubtedly contributed significantly to our fundamental understanding of gene expression. But in recent years, this technology has also shown use in the applied fields of synthetic biology as well as in the biotechnology industry. Owing to its fast and open platform, cell-free expression systems are used for rapid prototyping and screening of gene regulatory parts, genetic circuits, and protein libraries. Unlike with whole cells, experimental setup of an established cell-free system platform avoids the time-consuming step of DNA transformation and cell culture. Furthermore, these cell-free systems have been shown to accurately match *in vivo* characteristics in *E. coli* [41]. Niederholtmeyer *et al.* (2015) used cell-free systems to rapidly characterize genetic circuit oscillators and subsequently applied these in *E. coli* cells [185]. Moore *et al.* (2018) also showed matching promoter and RBS characteristics between *in vivo* and *in vitro* cell-free expression with *B. megaterium* [180]. Further expanding its utility, larger biosynthetic pathways can also be tested and optimized in cell-free systems. For instance, crude extracts enriched with enzymes can be mixed at different ratios to help optimize biosynthetic pathways for n-butanol [122] or mevalonate [63] production. Industrial use of cell-free reactions to produce molecules of interest directly is mainly limited by its ability to scale to larger volumes and its prohibitive cost. Low value biomolecules such as subtilisin, will not be economically feasible, but high-value pharmaceuticals, such as drug-conjugated antibodies, could be viable as cell-free products [235]. Voloshin and Swartz (2008) determined that oxygen availability was crucial for scale-up and demonstrated 1L cell-free reactions in a stirred tank reactor that produced up to 435 μg/ml of IGF-1 [274]. Zawada *et al.* (2011) produced a cytokine human granulocyteâmacrophage colony-stimulating factor (GM-CSF) at 700 μg/ml in a 100L batch [295] - the largest cell-free batch reaction done. Cell-free systems have also been developed into biosensors for medical, environmental, and forensic diagnostics by coupling nucleic acid sensors or metabolic pathways with transcription-factors that trigger a visual output. This has been shown to specifically detect target molecules like RNA sequences from the Novovirus, Zika, and Ebola viruses using RNA toehold switches [197, 198, 160] and other substances such as benzoic acid, hippuric acid, and cocaine [275], or Hg(II) and gamma-hydroxybutyrate [99] using novel metabolic enzymes or transcription factors. Lastly, one of the most attractive aspects of cell-free systems is their potential to reduce distribution costs and decentralize production of biomolecules. Cell-free expression systems are amenable to freeze-drying, which can remove the cold chain transport and distribution requirements of many biomolecular products. Diagnostic kits based on cell-free systems can be freeze-dried onto paper for convenient use in the field [197]. Conjugate vaccines, produced from freeze-dried *E. coli*-based

cell-free systems called iVAX (*in vitro* conjugate vaccine expression), have been shown effective against bacterial pathogens in mice [242]. Freeze-dried cell-free diagnostic kits with a glucose output compatible with off-the-shelf glucose meters provide a practical approach to low-cost diagnostics [11]. The field of personalized medicine could also potentially use cell-free systems at point-of-care. Extracts from human blood-derived leukocytes have been demonstrated to express nano luciferase (Nluc), Granulocyte-colony stimulating factor (G-CSF), and Erythropoietin (EPO) [35]. Although many of the applications of cell-free systems mentioned serve as proofs-of-principle to date, new and innovative solutions are continuously being developed to improve yields, specificity, and reproducibility.

2.1.1 *E. coli* extract-based cell-free expression

An *E. coli* extract-based CFES (also called TXTL for transcription-translation) consists of three main components: the extract, reaction mix, and DNA template. These components are mixed together and incubated at 16-37 °C for 1-24 hours for gene expression and protein production. The *E. coli* cell extract or lysate contains all the necessary enzymes for transcription and translation. The reaction mix supplies the nucleotide and amino acid substrates, tRNAs, energy resources, cofactors, polyamines, molecular crowders, and salts. The DNA template provides the gene of interest with the regulatory sequence elements (*e.g.* promoter, ribosomal binding site, and terminator) for gene expression. There are many variations of extract-based CFES preparation protocols. Modern *E. coli* extract-based CFES can produce up to 4 mg/mL of protein [87] as compared to early efforts that produced 1 μg/mL of protein [299]. The steady improvements in protein yield of *E. coli* extract-based CFES are a result of many rounds of rational and systematic optimization of methods and formulations. In this section, we discuss the development and rational behind current protocols and describe the purpose of specific ingredients in an *E. coli* extract-based CFES.

E. coli strains for cell extracts

The *E. coli* strain used for extract production is chosen depending on the requirements or application of the cell-free expression system. Many of the strains used for CFES are also used for protein expression and purification as both typically aim for high protein yields. For example, *E. coli* BL21 strains have been heavily used as host strains for extracts due to their lack of lon [294] and ompT [100] proteases which prevent protein degradation. Its derivative, *E. coli* BL21 (DE3) [248], is used to make extracts that contain the T7 RNA polymerase for T7 RNAP-directed transcription [149]. Having the T7 RNA polymerase (or any other enzyme) expressed in the extract itself avoids the need for extra purification steps and external addition into the cell-free reaction mixture. *E. coli* strains have also been developed to provide stable environments for linear DNA. For instance, *E. coli* NMR5 is modified by removing the *endA* endonuclease I and *recD* exonuclease V genes [172]. Purified GamS protein, a RecBCD inhibitor from lambda phage, can also be externally added into extract-based CFES to protect linear DNA templates [250, 55]. These modifications allow the use of linear DNA templates instead of plasmid DNA in CFES reactions to further reduce experimental times in design-build-test cycles [250]. The 59.T7.Opt strain is genomically recoded to replace all native amber stop codons to ochre codons (UAA), release

factor RF1-deficient, and genome integrated T7 RNA polymerase. This allows efficient incorporation of non-cannonical amino acids (ncAAs) using the amber codon (UAG) in extract-based CFES [166, 51]. *E. coli* strains can also be transformed with plasmids expressing enzymes needed in the extract. This provides a convenient method to enrich extracts with specific enzymes and does not require genomic integration. For example, *E. coli* BL21 Gold (DE3) harboring a pAD-LyseR plasmid was used to aid in the extract production method. The pAD-LyseR plasmid constitutively and weakly expresses the R endolysin that weakens the inner membrane of *E. coli* to facilitate cell lysis by a simple freeze-thaw cycle [55]. In the same study, the pACYC-FLAG-dN6-His plasmid was also transformed in the same strain to enrich extracts with an engineered covalently-linked ClpX protease hexamer for increased degradation of ssr-A tagged proteins [55]. BL21 Rosetta2 (DE3) strains harbor a plasmid expressing rare tRNA codons for efficient translation of recombinant genes *in vivo* as well as in extracts [249, 236]. Different strains and plasmids can also be used to produce enriched extracts that can perform post-translational modifications on expressed proteins such as disulfide bond formation [61] and glycosylation [119, 242]. The SHuffle T7 Express lysY *E. coli* strain (NEB, USA) constitutively expresses disulfide bond isomerase C (DsbC) to support disulfide bond formation. Extracts from this strain have been demonstrated to more efficiently express luciferase (from *Gaussia princeps*) that contains five disulfide bonds as compared to BL21 Star (DE3) extracts [61]. The CLM24 strain has a *waaL* knock-out, which leads to an accumulation of O-polysaccharide antigens or lipid-linked oligosaccharides (LLO) in the inner membrane [75]. CLM24 strains harboring plasmids that express specific glycans and oligosaccharyltransferase (OST) can be used to produce enriched extracts that can support N-linked glycosylation of cell-free expressed acceptor proteins [196, 119, 242]. This strategy can be further used to build full biosynthetic pathways by mixing combinations of enriched extracts as demonstrated for mevalonate synthesis in cell-free metabolic engineering [63].

Lysate preparation

The cell extract is produced by growing the selected *E. coli* strain, pelleting and lysing the cells to collect the cytosol, and clarifying the lysate (Figure 2.1). A rich media, typically 2xYTP (yeast extract, tryptone, phosphate buffer), is used to culture cells at 37 °C for extract preparation. The phosphate buffer composed of potassium phosphate mono and di-basic solutions are included to help maintain a stable pH and lower phophatase activity in the lysate [131]. Cells are harvested during the mid-log phase of growth for optimal cell-free expression. Harvesting either too early or too late results in lower downstream cell-free protein expression. Despite ribosome levels being maximal at the early-log phase [73], early-log phase extracts result in lower CFES activity as compared to mid-log phase extracts [133]. Modifications of the traditional 2xYTP media have been used to prolong the window of cell harvest for convenience, as well as to increase cell density and consequently the amount of lysate produced. For example, cell-free autoinduction (CFAI) media, prepared by increasing the buffering capacity and providing 82 mM glycerol, 11.7 mM D-lactose, and 2.8 mM D-glucose as additional carbon sources in 2xYTP media, results in equally active extracts harvested from cultures at OD_{600} values of 2.5 and 10. This capacity to maintain cell-free expression activity from different growth stages was attributed to the continued activity of the pentose phosphate pathway [150]. In a separate study, removal of glucose in the

culture media also resulted in higher bacterial cell-free expression activity as compared to extracts made from glucose-fed cultures [236].

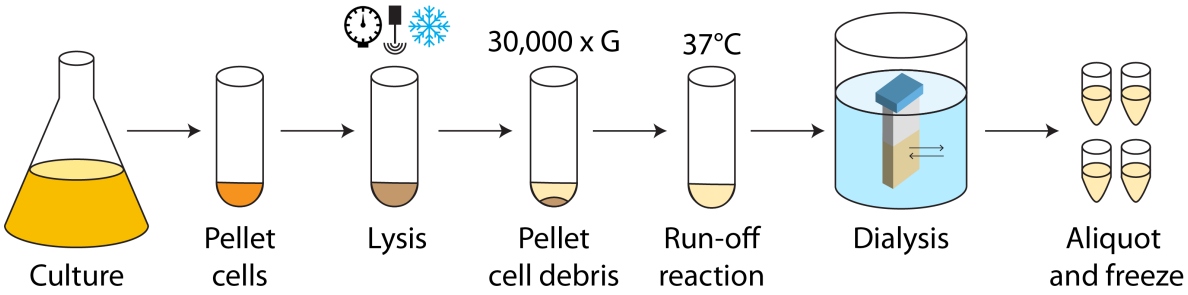

Figure 2.1: Lysate preparation for CFES. The selected strain is grown in rich media, pelleted, and lysed in S30A buffer. Cell lysis can be facilitated by bead-beating, sonication, high pressure (*e.g.* french press or emulsifier), or freeze-thaw cycles with enzymatic digestion. After lysis, the lysate supernatant is clarified by centrifugation at 30,000 xg, incubated at 37 °C for 1-2 hours in a run-off reaction, and then dialyzed in S30B buffer.

After culturing, cells are pelleted by centrifugation, washed, and resuspended in S30A buffer at pH 7.5. The S30A buffer is composed of either 10 mM tris-acetate, 14 mM magnesium acetate (MgOAc), 60 mM potassium acetate (KOAc), and 2 mM dithiotheitol (DTT) [156] or 10 mM tris-acetate (TrisOAc), 14 mM magnesium glutamate (MgGlu), 60 mM potassium glutamate (KGlu), and 2mM DTT [249]. The tris-acetate buffer helps maintain a stable pH. Magnesium and potassium are essential ions required for may nucleic acid-protein interactions and enzymatic activity [120]. Acetate or glutamate counter-ions have been used empirically [87]. Glutamate salts more closely match *in vivo* conditions as compared to acetate as glutamate is the principal anion produced in osmotically stressed *E. coli* cells [209]. However, using acetate or glutamate in CFES did not show any significant differences in protein expression [120]. DTT is provided to prevent the formation of disulfide bonds that may inactivate T7 RNA polymerase [220]. The cell pellet is resuspended in S30A buffer at a ratio of 1-2 mL lysis buffer for 1 g of pellet (wet weight) to be able to obtain a high concentration of protein in the extracts after lysis. Cell lysis can be facilitated by bead beating, sonication, pressure, or enzymatic methods. Bead beating and sonication methods are convenient for preparation of small batches of extract (up to 20 mL extract) [87]. However, both methods are prone to heating due to the prolonged exposure of the sample to high frequency energy [249, 149]. In contrast, pressure methods using a French press or continuous flow emulsifier avoids excessive sample heating because cells are subjected to high pressure changes through a nozzle only 1-3 times for lysis [60]. Pressure methods can be used for medium to large scale batches of extract (up to 200 mL extract) [87]. An alternative lysis method is by autolysis using an expressed R endolysin to facilitate cell lysis by a freeze-thaw cycle [55].

After lysis, crude lysates are clarified by centrifugation, incubated at 37 °C for 1-2 hours, and then dialyzed by a second S30B buffer. The centrifugation speed was originally set at 30,000 $\times g$ to ensure proper clarification and removal of cell debris, aggregates, and precipitates [170]. However, more accessible protocols using lower centrifugation speeds at 12,000 $\times g$ have also been successfully implemented [60]. After

clarification, the lysate is incubated at 37 °C for 1-2 hours. This step is called the *run-off reaction* where ribosomes are allowed to *run-off* their native transcripts to free them for cell-free translation. Additionally, short oligonucleotides sheared from the lysis step are also further digested by endogenous exonucleases during the run-off reaction [249]. After the run-off reaction, the lysate is dialyzed in S30B buffer at 4 °C for 3 hours. The S30B buffer is composed of 5 mM tris-acetate, 14 mM magnesium glutamate, 60 mM potassium glutamate, and 1 mM DTT at pH 8.2 [249]. Including the post-processing steps (run-off reaction and dialysis) improves transcriptional activity of native *E. coli* promoters [236]. Centrifugation at 12,000 $\times g$ between post-processing steps are also done to further clarify the lysate. The final lysate is flash frozen in 50-100 μL aliquots using liquid nitrogen and stored at -80 °C.

Reaction buffer

The reaction buffer supplies the CFE reaction with the necessary substrates and cofactors for RNA and protein synthesis, a favorable environment for enzymatic reactions, and an energy regeneration system. This is prepared as a separate solution and combined with the lysate and DNA template in the final CFES reaction. A standard reaction buffer is composed of nucleotides (ATP, UTP, CTP, GTP), the 20 amino acids, transfer RNA (tRNA), coenzyme A (CoA), β-nicotinamide adenine dinucleotide (NAD), folinic acid, spermidine, and an energy regeneration substrate such as 2-phosphoenolpyruvate (PEP) (Figure 2.2B). These components are dissolved in a solution of MgGlu, KGlu, and HEPES buffer at a pH of 7-7.5. The concentrations of the lysate, MgGlu, and KGlu are titrated for each batch of lysate produced to optimize protein expression yield. The final CFES reaction mix should have approximately 8.9-9.9 mg/mL of total protein from the lysate, 0-10 mM MgGlu, and 20-140 mM KGlu [249]. A standard final CFE reaction mix is described in Table 2.1. Additional cofactors can also be included to improve transcriptional regulation and gene expression yield. For example, addition of cyclic adenosine monophosphate (cAMP) improves gene expression and repression efficiency in *lac* regulated genes in CFES [40]. cAMP binds to an allosteric catabolite repressor protein (CRP) which interacts with regulatory DNA sequences in the lac promoter [201, 211]. Oxalate inhibits phosphoenolpyruvate synthetase that depletes the supply of ATP [131]. The *E. coli* cytosol is macromolecularly crowded at concentrations exceeding 200 mg/mL of protein and biopolymers (20-fold higher than a standard CFES solution) [209]. To mimic the macromolecular crowding environment in the cytosol, polyethylene glycol (PEG 8000, molecular weight (MW): 7000-9000 g/mol) or Ficoll 70 (MW: 70000 g/mol) can be included in CFE reactions at 2% v/v or 0-100 mg/mL, respectively. The addition of PEG was suggested to stabilize mRNA transcripts in the extract [120].

Other high-energy molecules such as creatine phosphate (CP) or acetyl phosphate (AcP) have also been used for energy regeneration. Similar to PEP, these molecules can phosphorylate ADP to ATP through kinase-catalyzed reactions (pyruvate kinase (Pyk) for PEP, creatine kinase (Crk) for CP, acetyl kinase (Ack) fpr AcP). Alternative energy regeneration systems that produce ATP through metabolic pathways have also been demonstrated to sustain CFE reactions. Kim and Swartz (2001) showed that pyruvate can be metabolized in extract-based CFES by endogenous enzymes with the addition of NAD cofactor to produce AcP and subsequently ATP [132] (Figure 2.2A). Conveniently, PEP is converted to pyruvate in the traditional CFE energy regeneration system and

Table 2.1: Components of an *E. coli* extract-based CFES reaction mix.

Component	Concentration	Purpose
Lysate	8.9-9.9 mg/mL	Crude extract from *E. coli* that provides enzymes (*e.g.* RNAP, ribosomes) and co-factors needed for transcription and translation.
Nucleotides	0.85-1.5 mM	ATP, CTP, GTP, UTP. Transcription building blocks for RNA synthesis.
Amino acids	1.5-5 mM	20 amino acids. Translation building blocks for protein synthesis.
tRNA	0.17-0.2 mg/mL	Guides amino acids to the ribosome complex based on the mRNA template triplet codon. Although tRNA is already present in the lysate, it is supplemented in the final CFE reaction mix.
CoA	0.26-0.27 mM	Coenzyme A. Cofactor for energy regeneration. CoA and pyruvate are condensed to make acetyl-CoA in the presence of NAD and then acetylphosphate. Acetylphosphate regenerates ATP from ADP [132].
NAD	0.33-0.40 mM	β-Nicotinamide adenine dinucleotide. *See above.*
Folinic acid	0.068-0.1 mM	Formyl donor substrate for N-formylmethionine (fMet) required for translation initiation.
Spermidine	1-1.5 mM	Ubiquitous molecule in the cell involved in many cellular processes including stimulating assembly of the 30S ribosomal subunit [69, 117]. Binds with the negatively charged backbone of nucleic acids and ribosomes. Improves efficiency and fidelity of *in vitro* gene expression [102].
PEP	30-50 mM	Phosphoenolpyruvate. Secondary energy source. Provides a high energy phosphate donor group that is transferred to ADP by pyruvate kinase to generate ATP [132].
DNA plasmid	1-20 nM	DNA template for transcription.

addition of NAD facilitates the secondary utilization of pyruvate [132]. Glycolytic intermediates such as glucose-6-phosphate (G6P) can also be metabolized to pyruvate via glycolysis to generate ATP. The conversion of one G6P molecule to pyruvate through the glycolysis pathway produces three ATP molecules as compared to only one ATP per PEP converted to pyruvate. This results in a greater protein expression yield with

G6P compared to PEP as an energy source. However, it was observed that addition of oxalate in G6P-powered CFE reactions reduced ATP concentrations [132]. Metabolism of maltodextrin or maltose are also used as low-cost ATP regeneration strategies that have the added benefit of consuming accumulated inorganic phosphate. Maltodextrin and maltose are metabolized into glucose or glucose-1-phosphate (G1P), which then enter the glycolysis pathway to produce ATP [280, 37] (Figure 2.2C).

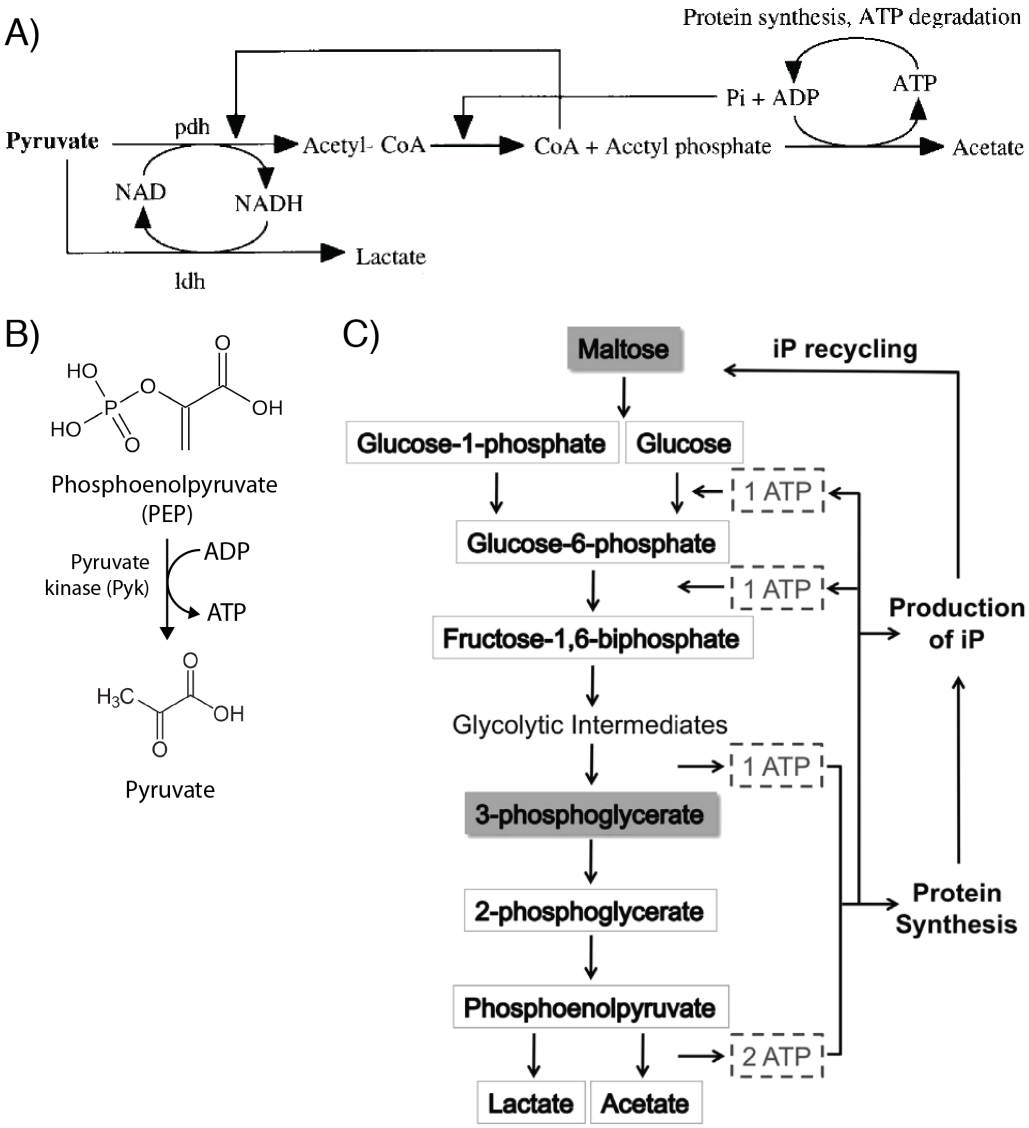

Figure 2.2: Energy regeneration reactions for CFES. (A) Pyruvate metabolism to produce acetyl phosphate to regenerate ATP. Image from [132]. Reproduced with permission. © 2001 John Wiley & Sons, Inc. (B) Phosphoenolpyruvate regenerates ATP from ADP using pyruvate kinase (Pyk). (C) Maltose-based metabolism for ATP regeneration also results in inorganic phosphate utilization. Image from [37]. Reproduced with permission. © 2014 Elsevier Masson SAS. All rights reserved.

2.1.2 PURE cell-free expression system

A second approach to cell-free expression is by using fully purified elements and reconstituting the entire gene expression system from the bottom-up. Unlike the extract-based CFES, this approach provides a completely defined reaction composition that is advantageous for direct testing of individual components. In addition, a purified and reconstituted CFES will lack nucleases and proteases that are normally present in an S30 extract, making it an ideal starting platform for bottom-up gene expression systems. The first attempt of a fully reconstituted CFES was reported by the group of Weissbach (1977). Although their purified system did not fully support protein synthesis, they identified three partially purified lysate fractions that restored protein synthesis when added to their purified system. In total, their reconstituted system was composed of of 33 purified *E. coli* factors, ribosomes, Ehrlich ascite extracts for aminoacyl-tRNA synthetases, and the three *E. coli* lysate fractions [138]. Several groups then used different methods to circumvent limitations in the purification of aminoacyl-tRNA synthetases by using pre-charged aminoacyl tRNAs [83] or partially purified aminoacyl-tRNA synthetases [200]. Improved techniques in protein purification led to the PURE (protein synthesis using recombinant elements) system developed by the group of Ueda in 2001. The PURE system is composed of 32 His-tagged purified proteins from *E. coli* in addition to tRNAs, NTPs, amino acids, and other components for energy regeneration (Table 2.2) [227]. Unlike the extract-based CFES, the PURE system is a completely defined mix with each ingredient individually added. This makes it possible to test all components of the CFES one at a time. For instance, Matsuura *et al.* (2009) exhaustively altered the concentrations of each of the 69 components of the PURE system to study interactions between components and their effects on protein yield. Using a Bahadur expansion analysis, they showed that only one to two-component epistatic interactions resulted in significant differences in protein translation, while more than two-component interactions were negligible [168]. Doerr *et al.* (2019) subjected the PURE CFES to a wide range of experimental conditions and targeted perturbations of ATP-dependent RNA helicase (hrpA), IF1, IF2, peptidyl-tRNA-hydrolase (PTH), and a mutated T7 RNAP to develop a mechanistic kinetic model that can globally fit the experimental results [59]. While providing a minimal CFES that is well-defined, the PURE system is costly and labor-intensive to produce. Fortunately, the preparation of the PURE system can be simplified by purifying several protein-producing *E. coli* strains in a single co-culture and purification step (OnePot PURE system) [145]. This method significantly reduces the labor and cost of the PURE CFES system.

Despite being a reduced or minimal system for gene expression, the PURE and extract-based CFES are still extremely complex in terms of its composition and biochemical reactions. To illustrate, a large-scale kinetic model describing the PURE CFES used a total of 241 species and 968 reactions [169]. The extract-based CFES is even more complex because of the many unidentified proteins and metabolites present in the lysate [86]. Further characterizing CFES reactions will be necessary to identify resource limitations and help develop accurate kinetic models. Spirin *et al.* (1988) first pointed out that the exhaustion of specific components can limit gene expression in CFES. By dialyzing a CFE reaction in a feeding buffer containing additional ATP, GTP, amino acids, and PEP, gene expression was prolonged to up to 40 hours [239]. Targeted methods, such as luciferase assays for ATP and colorimetric assays for inorganic phos-

Table 2.2: Components of the PURE system [227].

Translation factors	*Other enzymes*
Initiation factor 1 (IF1)	Methionyl-tRNA transformylase
Initiation factor 2 (IF2)	Ribosomes
Initiation factor 3 (IF3)	Creatine kinase
Elongation factor G (EF-G)	Myokinase
Elongation factor Tu (EF-Tu)	Nucleoside diphosphate kinase
Elongation factor Ts (EF-Ts)	Pyrophosphatase
Release factor 1 (RF1)	T7 RNA polymerase
Release factor 3 (RF3)	
Ribosome release factor (RRF)	*Energy sources*
	ATP
Aminoacyl-tRNA synthetases	GTP
AlaRS	CTP
ArgRS	UTP
AsnRS	Creatine phosphate
AspRS	
CysRS	*Buffers*
GlnRS	HEPES
GluRS	KOH
GlyRS	Potassium glutamate
HisRS	Magnesium glutamate
IleRS	Spermidine
LeuRS	DTT
LysRS	
MetRS	*Other components*
PheRS	20 amino acids
ProRS	10-formyl-5,6,7,8-tetrahydrofolic acid
SerRS	tRNAs
ThrRS	
TrpRS	*Template*
TyrRS	DNA/mRNA
ValRS	

phates and NAD^+/NADH, were later used to measure specific changes in composition over time in CFE reactions [130, 131, 132, 63]. High-performance liquid chromatography (HPLC) enabled the measurement of amino acids and several metabolites such as glucose, acetate, lactate, pyruvate, succinate, oxaloacetate, phosphate, and ethanol [131, 120, 63]. More recently, liquid chromatography–mass spectrometry (LC-MS) and gas chromatography–mass spectrometry (GC-MS) methods were used for proteomic and metabolomic characterization of extract-based CFES. These studies revealed proteomic and metabolomic changes in lysates prepared using different methods [79, 173], detected the completeness of translated proteins and post-translational modifications [116], and identified proteins required for specific pathways and heterologous functions [86]. Metabolic and proteomic profiling offer a wealth of information - identifying up to 800-1839 proteins [86, 79, 116] and 260 analytes [173] in CFES. These comprehensive

datasets can be used to further standardize CFES and provide insight into mechanisms and interactions of gene expression reactions coupled with metabolic pathways.

2.2 Compartmentalizing CFES with liposomes

Building functional synthetic cells from the bottom-up requires robust methods of compartment assembly. Thanks to the wide application and potential use of liposomes in membrane science, drug delivery, and artificial cells, many different methods of liposome formation have been developed. Each of these methods provide different advantages and disadvantages in terms of ease-of-use, compatibility with the encapsulated CFES reaction, and properties of liposomes produced. In this section, we focus on some bulk and microfluidic methods for unilamellar liposome production.

Phospholipids, in particular phosphatidylcholines (PC), make up the majority of biological cell membranes and can form lipid bilayers in aqueous media due to their amphiphilic nature [6]. The molecular structure of the phospholipids greatly affect the properties of the lipid bilayer. For instance, headgroup size and charge and chain length and saturation all influence the packing density, fluidity, bending stiffness, and spontaneous curvature of the lipid bilayer in liposomes [175]. In particular, the molecular shape of the phospholipid can favor different arrangements or packing behavior. Phosphatidylcholines provide a cylindrical shape that favorably organizes into lamellar or lipid bilayer structures (Figure 2.3). Generating liposomes in the laboratory using phosphatidylcholines such as L-α-phosphatidylcholine (Egg PC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), and 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) have been well-established [210, 12, 199]. However, there are also a variety of other phospholipid types and synthetic molecules composed of amphiphilic block copolymers that can be used to create liposomes or *polymersomes*, respectively [58].

In biological cells, membrane composition can affect the physical properties and functions of the membrane, such as for recruitment of membrane proteins [144] or signalling in regulatory pathways through membrane sensor proteins [17]. Similarly, additives can also be included in the lipid bilayer of liposomes to modulate physical properties and function. For example, cholesterol can decrease membrane fluidity at high temperatures by restricting phospholipid movement, but increase fluidity at low temperatures by interfering with the tight packing of the phospholipids [175]. In liposome drug delivery systems, up to 2% Polyethylene glycol (PEG)-lipid conjugates or PEGlylated lipids can be used in the phospholipid composition to increase stability of liposomes. The PEG group sterically prevents interactions of the lipid bilayer with the environment, prolonging its circulation half-life in the environment [135, 181]. To form and maintain stable vesicles, it is critical to consider environmental factors such as pH, temperature, osmolarity, and external mechanical forces. The osmolarity of the inner and outer solutions of liposomes should be balanced to prevent swelling and bursting. Osmolarity or osmotic concentration is the total solute concentration of a solution and provides a measure of the osmotic pressure that can be exerted through a semi-permeable membrane between two solutions. In practice, osmolarity, with units of Osmomoles/L, is determined indirectly by measuring the freezing point depression or vapor point depression against a standard set of known solutions. However, these

colligative properties differ for different solutes [251]. As a result, osmolarity measurements of complex mixtures such as cell extracts can have deviations from its actual value.

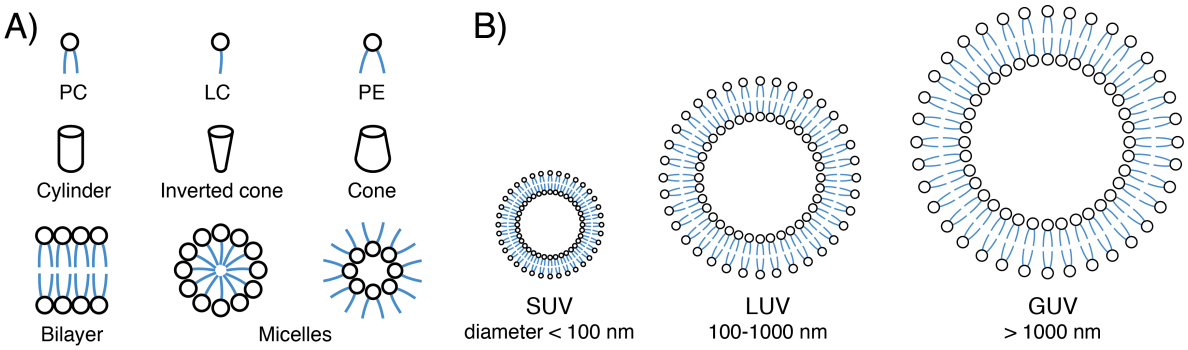

Figure 2.3: Phospholipid and liposome classification. (A) The organization of phospholipids depend on the chemical structure and molecular shape of the phospholipid. Phosphatidylcholines (PC) have a cylindrical molecular shape that results in a lamellar or bilayer organization. Lysophosphatidylcholine (LC) and phosphatidylethanolamine (PE) lipids have a cone-shape that results in micelle organization. (B) Unilamellar liposomes classified by size: small unilamellar vesicles (SUV) have diameters less than 100 nm, large unilamellar vesicles (LUV) with diameters between 100-1000 nm, and giant unilamellar vesicles (GUV) with diameters larger than 1000 nm. Illustrations not drawn to scale.

2.2.1 Bulk methods for GUV production

Here we describe several bulk methods of GUV production. These methods typically result in polydisperse liposomes and have the advantage of being convenient to produce without much specialized equipment.

Lipid film swelling

Liposomes were first synthesized in the laboratory in the late 1960s with a lipid film swelling or gentle hydration method [45, 210]. To produce liposomes, phospholipids dissolved in chloroform are deposited on a glass surface and dried. The dry lipid film is hydrated with water, which causes the lipid film to swell and form liposomes that detach from the surface with gentle shaking or flow (Figure 2.4A). Liposomes produced using this method are polydisperse and multilamellar, depending on the amount of lipids deposited on the surface. While the method is relatively easy, the yield is low because the lipids need to be correctly oriented to form the bilayer. In addition, high ionic strength solutions inhibit liposome formation due to electrostatic effects. By using negatively charged lipids or neutral lipids in a solution with divalent cations, repulsive forces can stabilize unilamellar liposomes and oppose adhesive forces between membranes [4, 5]. For instance, by using 10% negatively charged lipids, Akashi *et al.* (1996) prepared GUVs with up to 100 mM ionic salt solutions [4]. Depositing the lipid film on glass beads can increase liposome yield by increasing the surface area for liposome formation [56]. Additionally, doping the lipids with sugar helps separate the lipid films during swelling as water permeates the lipid layers due to osmotic pressure [264]. A gel-assisted hydration method was also developed that improved the yield

and robustness of liposome formation. This involved drying the lipid film on top of an agarose gel or polyvinyl alcohol (PVA) layer which partially dissolves and swells to promote liposome formation [113, 284].

Electroformation

In electroformation, lipid films are deposited on either platinum or indium titanium oxide (ITO) electrodes, immersed in water, and then subjected to a direct or alternating (DC or AC) electric current (Figure 2.4B). The electric current causes fluctuations that facilitate bilayer separation and bending for fast liposome formation [12]. Relatively uniform sized liposomes can be produced by depositing a micropatterned lipid film, which sets a growth limit within each patch [258]. Initially, electroformation was limited to low salt conditions and lipid films deposited from organic solvents. Potts *et al.* (2008) modified the protocol by creating lipid film deposits from aqueous proteoliposomes lipid solutions. The proteoliposomes are SUVs or LUVs generated using sonication, extrusion, or reverse phase evaporation techniques and provides smooth and properly oriented membranes on the electrodes that more easily form unilamellar vesicles [203]. Additionally, by optimizing both electroformation parameters and osmotic pressure for swelling, GUVs can be efficiently produced at physiologically relevant conditions that are crucial for native membrane reconstitution and membrane protein studies. This combined electroswelling methodology has been successfully used to integrate functional sarcoplasmic reticulum Ca^2-ATPase and the H^+ pump bacteriorhodopsin [90], as well as native membranes and lipid mixtures into liposomes [179].

Inverse emulsion phase transfer

The inverse emulsion phase transfer method, established by Pautot *et al.* (2003), uses water-in-oil droplet templates to form unilamellar liposomes [199]. First, a water-in-oil emulsion is made from the desired inner solution and lipid-oil mixture. Due to their amphiphatic nature, the lipids assemble to form a monolayer around each water droplet interface. The emulsion is then gently poured over a layered solution of the aqueous outer solution (bottom) and lipid-oil mix (top). The inner solution is made denser than the oil and outer solution so that the droplets sink by centrifugation and pass through a second lipid monolayer interface into the outer solution to form the liposome (Figure 2.4C). This method can be used to make either symmetric or asymmetric liposomes, depending on the lipid oil mix used to make the emulsion and layered solution. The insertion orientation of membrane proteins can also be controlled by adding the protein in the inner or outer solution [288]. A prerequisite for the inverse emulsion phase transfer method is the density difference of the inner and outer solutions. This is done using sugar-based density gradients (*e.g.* sucrose and glucose in the inner and outer solution, respectively) to obtain a density difference while maintaining osmotic balance. The relative speed in producing liposomes and the capability to have different inner and outer solutions without extra washing steps make this method particularly suitable for encapsulating CFES [192, 190, 38]. Moga *et al.* (2019) reported a thorough study of the inverse emulsion phase transfer method [177]. Here, the different parameters such as the sugar density gradient, centrifugation speed and time, volume ratios of the phases, and incubation time for the lipid monolayer to form were systematically optimized to provide guidelines on efficient liposome formation by the inverse

emulsion phase transfer method. The main drawback of this method is the presence of trace amounts of oil in the lipid bilayer, which is why most biophysical studies on lipid membranes opt to use oil-free liposome formation methods.

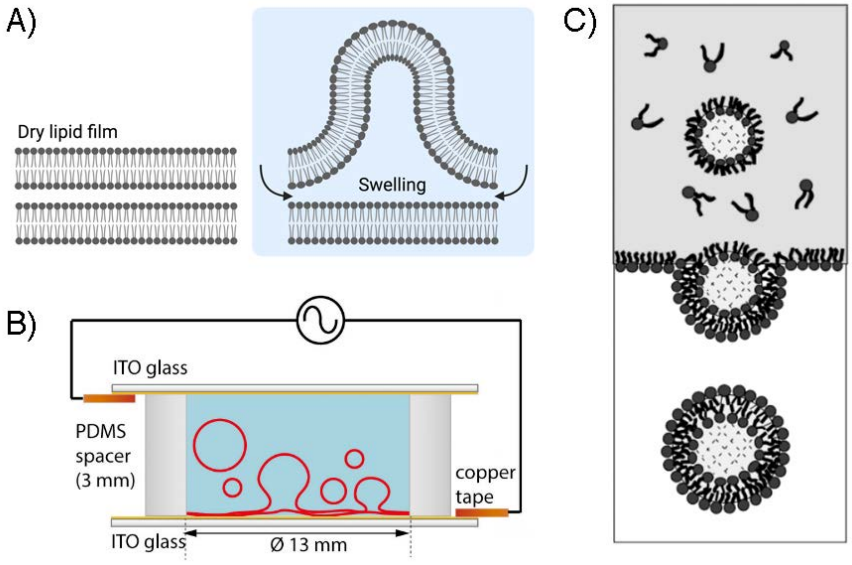

Figure 2.4: Bulk methods for GUV production.(A) When water penetrates the space between dried lipid bilayers, they can dettach and form unilamellar or multilamellar liposomes. Image created with BioRender.com under a paid subscription (Max Planck Gesellschaft). (B) Electroformation of liposomes from lipid films on ITO electrodes immersed in buffer. Two ITO electrodes (one coated with dried lipid films) facing each other in a chamber is filler with buffer. An AC electric field is applied and GUVs form at the ITO surface with dried lipids. Image from [244] licensed under CC BY 4.0. (C) Inverse emulsion phase transfer method. The top layer is an emulsion with water-in-oil droplet templates. The oil phase contains phospholipids that arrange in the water-oil interfaces as monolayers. The bilayer is formed as the droplets are driven down into the bottom aqueous layer. Image reproduced with permission from [199]. © 2003 American Chemical Society. All rights reserved.

One-pot GUV assembly

Recently, Göpfrich *et al.* (2019) developed a one-pot method to produce GUVs [97]. Similar to the inverse emulsion phase transfer method, it uses water-in-oil emulsion droplet templates. However, instead of having lipids dissolved in the oil phase, SUVs and proteoliposomes in the aqueous phase fuse to the droplet interface to form the GUV. The oil phase contains PEG-based surfactants to stabilize the GUVs. Addition of a droplet-destabilizing agent (perfluoro-1-octanol, PFO) displaces the PEG-based surfactants and causes the GUVs to fuse into the oil-water interface and be released into the aqueous phase. Unilamellarity of the resulting GUVs were confirmed by fluorescence recovery after photobleaching (FRAP) measurements of the membrane to determine the fluidity of the lipids, CryoTEM imaging of the membrane, and membrane integration of the α-hemolysin pore protein. This method was shown to be robust with a variety of charged or neutral lipids, as well different buffer solutions and complex media (Dulbecco's Modified Eagle Medium, DMEM).

2.2.2 Microfluidic methods for GUV production

Microfluidic methods of GUV production addresses the issues of reproducibility and encapsulation efficiency in bulk methods. Using controlled flow rates at low Reynolds numbers in microfluidic systems, uniform-sized water-in-oil emulsion droplets are produced and transferred into the aqueous phase with high efficiency. Microfluidic approaches require more equipment (pressure/syringe pumps, glass capillary or polydimethylsiloxane (PDMS) microfluidic chips, microscope with high-speed camera) and effort to prepare. Fortunately, there is a growing community in the field of microfluidics and efforts to make the technology more open and accessible [136], as well as commercially available. Notably, all microfluidic methods are oil-based methods. However, biologically compatible oils and dewetting techniques can be used to reduce leftover oil.

Droplet emulsion transfer

The droplet emulsion transfer method creates water-in-oil droplets in a flow-focusing microfluidic device, and then transfers these into an aqueous solution to form the liposomes. This was first demonstrated by Tan *et al.* (2006), where the oil phase is composed of oleic acid with phospholipids to stabilize the droplets and form a lipid monolayer at the oil-water interface. Afterwards, the droplets are collected and injected into a 10-30% ethanol and water solution one droplet at a time. The oleic acid dissolves in ethanol, leaving the lipids to assemble into a bilayer to form the liposome. This method was demonstrated to encapsulate solutions containing polystyrene beads, GFP protein, yeast cells, and HeLa and MCF7 breast cancer cells in DMEM media with cell viability of more than two hours [253]. However, it was also observed that the carboxylic group of oleic acid can change the pH of the inner solution. As a result, encapsulated GFP fluorescence was quenched in the water-in-oil droplets and re-adjusting the pH to 10 allowed fluorescence of the GFP again. Asymmetric lipid bilayers can be made in a similar approach by transferring the droplets into a layered solution of aqueous outer solution and a different lipid solution from the water-in-oil droplets [115]. Later, a fully integrated microfluidic device was developed to handle the phase transfer on-chip in a second junction where the water-in-oil droplets are deflected into the aqueous phase flow [167] (Figure 2.5). Unlike the droplet emulsion transfer method, this is a direct translation of the bulk inverse emulsion phase transfer because the lipid bilayer is formed as the droplet crosses the oil-water interface. The method was shown to produce liposomes with an encapsulation efficiency of 83% using dodecane or hexadecane with DOPC lipids for the lipid-oil phase. The liposomes retained dextran (10 kDa) and fluorescein dye (332 Da), while addition of α-hemolysin resulted in permeability of the fluoroscein dye. A unique droplet emulsion transfer device was also recently developed based on a continuous droplet interface crossing encapsulation (cDICE) [267]. The cDICE device is composed of a lipid-oil and outer aqueous solution interface maintained by centrifugal force. Water-in-oil droplets are generated from an aqueous capillary injection into the lipid oil phase. These droplets move through the oil-water interface by centrifugation to form liposomes. This method produced liposomes with an average diameter of 12 μm with a coefficient of variation of 47%, which results from polydisperse droplets made from the capillary jet.

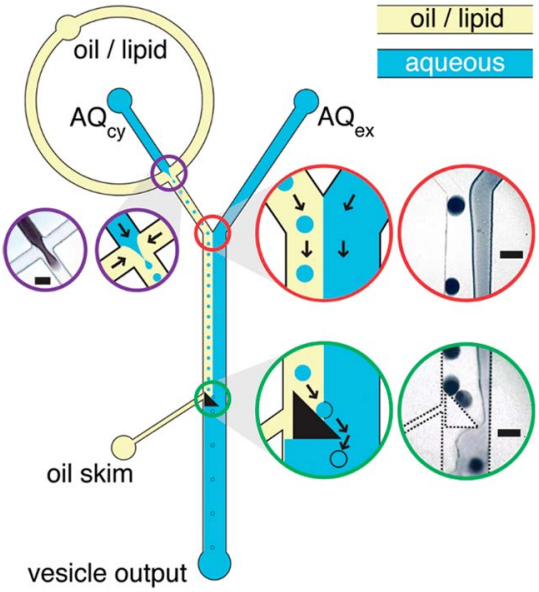

Figure 2.5: On-chip phase transfer method. Water-in-oil droplets are generated in the first junction and then deflected by a post (black triangle) into the aqueous layer to form the GUVs. Image reproduced with permission from [167]. © 2011 American Chemical Society.

Double-emulsion microfluidics

Double-emulsion microfluidic devices were first developed by Shum *et al.* (2008) [231]. This was composed of glass capillaries that focus inner aqueous and middle lipid-oil solutions that break up at the collection tube to form double-emulsions (Figure 2.6A). The phospholipids form monolayers at the interfaces to stabilize the double-emulsion droplets. The oil layer is then removed of the toluene and chloroform solvent by slow evaporation or dewetting. A PDMS-based microfluidic device was later designed to produce double-emulsion templates for liposome and polymersome production [202]. This design is composed of two flow-focusing junctions in series that encapsulates the inner aqueous solution into water-in-oil droplets and then again into the outer aqueous solution to make water-in-oil-in-water double emulsions. To avoid the double emulsions wetting onto the PDMS, the outlet channels are pretreated with a hydrophilic coacervate coating composed of poly(diallyldimethylammonium chloride) (PDADMAC) and poly(sodium 4-styrenesulfonate) (PSS). Additional components, such as Synperonic F18, glycerol, and PDADMAC, were included in the outer solution to stabilize the double emulsions during liposome production. Excess oleic acid used as the oil phase solvent is then removed by ethanol extraction. At the same time, a microfluidic design was also developed for octanol-assisted liposome assembly (OLA). This uses a six-way PDMS junction to form the double emulsions. The lipid-oil phase uses 1-octanol as the solvent, which spontaneously buds of the double emulsion to minimize interfacial membrane tension [52] (Figure 2.6B). A chacterization study of OLA-generated liposomes validated that lateral diffusion of lipids in the membrane measured by FRAP are not significantly different from electroformed liposomes [216]. Multicellular liposomes of up to thirty compartments have also been made by encapsulating multiple water droplets in the double emulsion and then removing the remaining octanol using a glass capillary device [48] (Figure 2.6C). Further improvements were made in the

PDMS double-emulsion design in [202] by including serpentine modules in the input and output channels to help stabilize flows, using octanol as the oil phase solvent, and constricting the output double-emulsions or liposomes at each turn of the serpentine module to pinch off excess oil in the lipid bilayers [289]. The stabilized flows allowed the production of liposomes without the need for additional surfactants that can affect membrane properties.

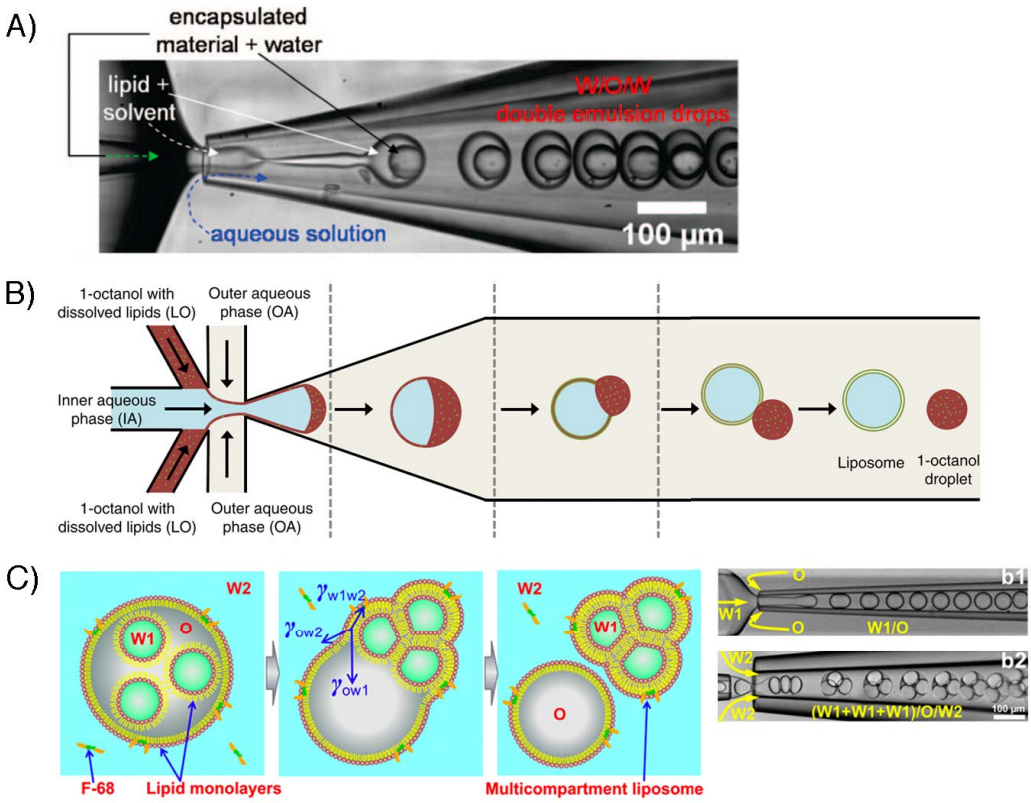

Figure 2.6: Double-emulsion microfluidics for liposome production. (A) Glass capillary device for generating double emulsion droplets. Image from [231]. © 2008 American Chemical Society. (B) Multicompartment liposomes are made by encapsulating more than one water-in-oil droplet in the double emulsion. Image reproduced with permission from [48]. © 2016 American Chemical Society. (C) Octanol-assisted liposome assembly from double emulsion microfluidics. Double emulsions are generated in a six-way junction PDMS chip. The interfacial tension of octanol and water is higher than the lipid membrane. This results in an octanol pocket forming on the liposome that spontaneously buds off. Image from [52] licensed under CC BY 4.0.

Phase-transfered 3D printed GUVs

Most microfluidic methods described have been focused on high-throughput production of monodisperse GUVs. However, building multicellular systems also require control in the 3D structure and arrangement of the single cell units. Recently, Alcineso *et al.* (2021) developed a method where 3D printed water droplets stabilized in an oil-lipid phase or droplet-bilayer interface (DIBs) [273, 29, 9] (previously described in Section 1.2) are dewetted into an aqueous phase to form defined multicellular liposome structures [8] (Figure 2.7). Unlike the multicompartment double emulsion method [48],

3D printing allows precise control of the geometry of the multicellular structure, as well as the organization of sub-populations. Here, dewetting of the printed DIBs were facilitated by a two-step phase transfer process. First, lipids were exchanged from 1,2-diphytanoyl-sn-glycero-3-phosphocholine (DPhPC) lipids used for droplet printing to POPC. This helps lower the interfacial tension in the lipid membrane and quickly remove excess undecane:silicone oil (36:65% v/v) phase used during the aqueous transfer in the second step. These synthetic tissue structures were then shown to successfully integrate α-hemolysin pore proteins in their lipid bilayers to sense and release molecules, encapsulate coacervates, and express protein from CFES.

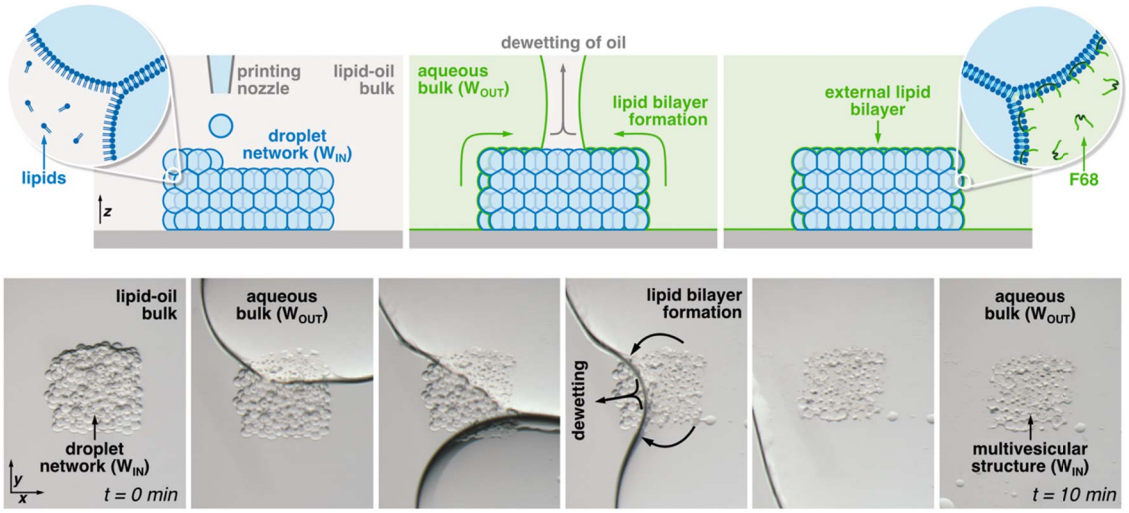

Figure 2.7: 3D droplet printed multivesicular structures are printed as droplets with lipid bilayer interfaces and dewetted into an aqueous phase to produce large multicompartment liposome structures. Image from [8]. © 2021 Wiley-VCH GmbH.

Chapter 3

Gene expression dynamics in synthetic cell populations*

3.1 Background

Due to its resemblance to biological cells, CFES encapsulated in liposomes have been established as one of the most popular and utilized type of synthetic cells. As single cell units, CFES in liposomes have been used to make synthetic cells that can implement transcription and translation [192], gene expression cascades [118], genetic circuits for intercellular communication [1], and CFES coupled with other metabolic processes like ATP production [21] and DNA replication [270]. Despite these successes, compartmentalized cell-free transcription and translation dynamics has not yet been simultaneously quantified and modelled. While CFES are dramatically reduced in complexity compared to biological cells, they still contain as many as thirty-seven enzymes and thirty-two small molecule compounds or substrates in PURE systems [228, 145] and even more for crude extract-based CFES. This makes it challenging to collect sufficient data to test existing CFES models or to develop tractable models that rely on the knowledge of precise chemical species as a function of time. Fortunately, there is an increasing effort towards proteomic and metabolic analysis that can provide more detailed and quantitative molecular information to profile CFES reactions [120, 86, 32, 173]. Alternatively, coarse-grained CFES models circumvent the need to measure all molecular species by focusing on the conversions of only a few species such as DNA, RNA, protein, RNA polymerase, and ribosomes. Several models have already been demonstrated to faithfully capture quantified cell-free gene expression dynamics in bulk solutions. For instance, coarse-grained models using first order and Michaelis-Menten kinetics have described cell-free transcription and translation dynamics [124, 233, 247] and extended to include concentrations of RNA polymerase and ribosomes [103, 165]. Dynamic models that include the initiation, elongation, and termination steps of translation [188, 59] or central carbon metabolism [114] have been used to identify bottlenecks in transcription and translation which can be experimentally relieved to improve protein productivity in

*This chapter is based on the article: Gonzales DT, Yandrapalli N, Robinson T, Zechner C, Tang TYD. **Cell-free gene expression dynamics in synthetic cell populations.** *ACS Synthetic Biology*, 11(1):205-215, 2022. [95]

CFES. These examples demonstrate how quantitative coarse-grained models can provide a better understanding of the CFES building blocks and will be crucial for further engineering more complex synthetic multicellular systems.

Many studies have shown that compartmentalization significantly affects the encapsulated CFES. For example, CFES encapsulated in small cell-sized liposomes can result in stochastic gene expression [190] and rare phenotypes, such as high gene expression not observed in bulk reactions [25]. The outer solution of the semi-permeable liposome compartments can also act as a source or sink of material. For instance, by providing a *feeding* buffer in the outer solution, the gene expression yield and activity of liposome encapsulated CFES can be increased and prolonged [192]. For this reason, it is also important to be able to quantify compartmentalized CFES reactions to test current CFES models. This will provide insights into the physical effects of encapsulation on cell-free gene expression, such as trans-membrane diffusion and surface effects, that will be crucial in building and understanding multicellular systems. Bulk methods of liposome production result in synthetic cells with large variations in cell size and gene expression profiles [193, 25, 213, 38]. Using bulk methods can be advantageous due to their accessibility and opportunity in exploring a large random space in terms of encapsulation and size [25] without specialised equipment. However, it can also be advantageous to generate uniform populations of synthetic cells for reproducibility and predictability. Microfluidic and droplet printing techniques can generate synthetic cell populations with greater throughput, control, and uniformity compared to standard bulk methods [265, 159]. This has already been shown to be effective in generating monodisperse synthetic cells to study the effect of macromolecular crowding on gene expression without the use of synthetic crowding agents [272] and to qualitatively monitor Spinach2 RNA aptamer transcription dynamics [49]. This is especially important for quantitative approaches as it enables the generation of statistically robust data that is amenable to accurate modelling.

Methodologies to monitor mRNA and protein dynamics in cell-free systems have been demonstrated by utilizing fluorescence resonance energy transfer (FRET) donor-acceptor pairs for mRNA [247, 186], fluorescent proteins such as GFP and YFP [25, 213], and fluorescent Spinach RNA aptamers simultaneously with YFP [187, 269] or mCherry [193]. However, expressed mRNA and protein levels in encapsulated CFES have only been measured as either relative fluorescence units within the synthetic cells or quantified concentrations for mRNA or protein levels separately. To the best of our knowledge, an absolute and simultaneous quantification of both transcription and translation dynamics within liposome synthetic cell populations has not yet been presented. In this chapter, we (i) applied methods of mRNA and protein monitoring in CFES reactions to quantify and test models of cell-free gene expression dynamics and (ii) encapsulated these reactions into liposomes using double-emulsion microfluidics [289] to generate a monodisperse population of synthetic cells. These address both challenges of monodispersity and quantification of gene expression dynamics in synthetic cell populations. mRNA and protein dynamics are simultaneously quantified using a fluorescent Broccoli RNA aptamer [78, 76] and mCherry protein reporters and converted into absolute concentration units by a standard calibration curve. Bulk reaction experiments were used to develop and select from several variations of a resource-limited cell-free

gene expression model in [247]. Overall, this work combines bottom-up assembly with mathematical modelling to provide a quantitative outlook of liposome compartmentalized gene expression dynamics in synthetic cell populations. This can help facilitate direct comparisons between experiments of bulk and compartmentalized CFES reactions and provide a basis for the design and construction of multicellular systems using controlled assembly methods and accurately modelling their modular single cell units.

3.2 Methodology

3.2.1 Plasmid design

The plasmids pEXP5-NT/6xHis eGFP [152] and pEXP5-NT/6xHis mCherry [255] were kindly provided by J. L. Ross Anderson, University of Bristol. These plasmids consist of a constitutive T7 RNA polymerase-mediated promoter with a strong ribosomal binding site to express 6xHis-tagged eGFP and mCherry fluorescent proteins, respectively. To monitor both transcription and translation dynamics in CFES, we constructed the pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid (Figure 3.1). This plasmid consists of a constitutive T7 RNA polymerase-mediated promoter to express a red fluorescent mCherry protein and two copies of a dimeric Broccoli RNA aptamer stabilized by the F30 stem-loop [78, 77] between the stop codon of mCherry and the terminator of the gene construct. The pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid was made by inserting a F30-2xdBroccoli fragment downstream the mCherry stop codon and upstream the terminator of the mCherry gene in the pEXP5-NT/6xHis mCherry plasmid. This results in transcribed mRNA that includes the F30-2xBroccoli sequence but a translated protein without the F30-2xBroccoli sequence. The F30 structure acts as a stable RNA scaffold for the two dimeric Broccoli units (2xdBroccoli) [76]. Broccoli binds and activates the fluorescence of the small molecule (Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one (DFHBI) (Sigma) [78]. All plasmids have a high-copy number origin of replication. Plasmid construction protocols and sequences are further described in Appendix B. The pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid was sequenced confirmed by Sanger sequencing and is available in Addgene (www.addgene.org, plasmid ID 169233).

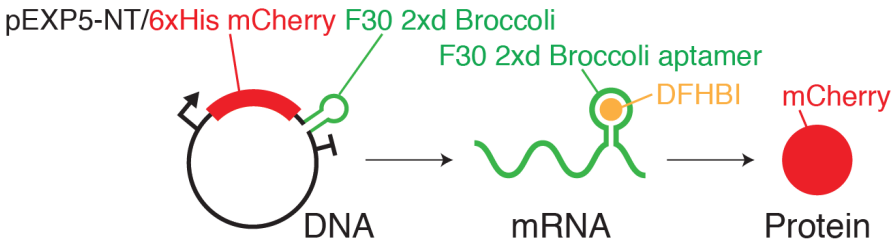

Figure 3.1: The pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid contains a constitutive T7 RNAP-mediated promoter expressing 6xHis mCherry with a F30-2xdBroccoli RNA aptamer tag. A small molecule dye DFHBI becomes green fluorescent upon binding with the Broccoli RNA aptamer (472 nm max excitation/507 nm max emission [78]). mCherry is translated from the mRNA transcript to provide a red fluorescent signal (587 nm max excitation/610 nm max emission (www.fpbase.org)).

3.2.2 Bulk CFES experiments

Bulk CFES expression experiments were run using a standard half-volume (12.5 μL) reaction mix of the PURExpress In Vitro Protein Synthesis Kit (NEB, USA). All CFES experiments were supplemented with sucrose at a final concentration of 80.4 mM. The additional sucrose was included to balance the osmolarity between inner and outer buffer solutions in the encapsulated experiments, but also included in the bulk experiments to maintain the same reaction conditions. To detect levels of the Broccoli RNA aptamer, 10 μM of DFHBI was added in CFES reactions using the pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid or its purified transcripts. A standard reaction is described in Table 3.1.

Table 3.1: Standard PURExpress master mix. DFHBI is added only when the DNA plasmid or mRNA template used contains the Broccoli RNA aptamer. This standard PURExpress CFES mastermix has an osmolarity of approx. 1200-1500 mOsmol/kg.

Component	**Volume (μL)**
Solution A	5
Solution B	3.75
RNAse Inhibitor (NEB)	0.25
Sucrose (1.5 M)	0.67
DFHBI (500 μM)	0.25
Plasmid DNA or mRNA	x
Water	Fill to 12.5
TOTAL	12.5

All plasmid DNA templates were prepared and purified by ethanol precipitation using the QIAGEN Plasmid Maxi Kit (QIAGEN, Germany) and then dissolved in nuclease-free water. 6xHis mCherry F30-2xdBroccoli mRNA transcripts were prepared by in vitro transcription of the pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid using the HiScribe T7 High Yield RNA Synthesis Kit (NEB, USA), treated with DNAse I (NEB, USA), purified using the QIAGEN RNeasy Mini kit (QIAGEN, Germany), and dissolved in nuclease-free water. Triplicate CFES reactions with the required DNA or mRNA template concentrations were prepared in 384-well plates (Greiner Bio-One, Austria), sealed with a clear film (Thermo, USA), and incubated in a Spark 20M plate reader (TECAN, Switzerland) at 30 °C. Fluorescence measurements were undertaken for each sample at 10-minute intervals for 8 hours. Excitation and emission wavelengths used were 485/535 nm, 570/620 nm, and 450/510 nm with a bandwidth of ±20 nm each, for eGFP, mCherry, and Broccoli RNA respectively. Fluorescence values were then converted into concentration units using a linear calibration curve from serial dilutions of purified eGFP protein, mCherry protein, and Broccoli RNA in the same CFES reaction mix and plate reader acquisition settings. In all our calibrations (including confocal calibrations for encapsulated CFES), we assume that our purified standards contain only complete and properly folded fluorescent protein or RNA aptamer. Further details for the calibration are available in Appendix E.

3.2.3 Encapsulated CFES experiments

Inverse emulsion phase transfer

Liposomes generated using the bulk inverse emulsion phase transfer method were prepared using a modified protocol as described in [177]. Briefly, 7.5 μL of PURExpress CFES with pEXP5-NT/eGFP plasmid DNA (Table 3.1) is added into a 1.5 mL microcentrifuge tube with 375 μL of lipid oil phase consisting of 0.39 mM 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC) (Avanti, USA) and 1.25 μM N-(Texas Red sulfonyl)-1,2-dihexadecanoyl-sn-glycero-3-phosphoethanolamine, triethylammonium salt (Texas Red DHPE) (Biotium, USA) in mineral oil (Sigma, USA). To prepare the lipid oil phase, POPC (36 μL of 32.5 μM in chloroform) and Texas Red DHPE (1μL of 1mM in chloroform) are mixed together in a glass test tube, dried under flowing nitrogen gas for 5 minutes and in vacuum for 30 minutes, resuspended in 2 mL mineral oil, and incubated at 37 °C in a Sonorex sonicator bath (Bandelin, Germany) for 1 hour. The mixture is then emulsified by running the tube across a microcentrifuge tube rack ten times. Then, 100 μL of the emulsion is gently layered on top of a prepared 100 μL outer feeding buffer solution (Table 3.2) + 40 μL lipid oil phase solution in BSA-coated wells of 96-well plate. The plate is next centrifuged at 3000xg for 10 minutes to produce liposomes that settle at the bottom of the plate. The plate is then incubated at 30 °C overnight (approx. 15 hours) and then imaged using confocal microscopy. This method is illustrated in Figure 3.2. The osmolarities of the inner and outer solutions were measured by freezing-point depression using a freezing point osmometer (Osmomat 3000, Gonotec) calibrated with water, 300 mOsmol/kg, and 2000 mOsmol/kg NaCl solution standards. Large differences in osmolarities (>20 mOsmol/kg) were balanced by adding water or 1.5 mM glucose to the outer solution when necessary. Confocal microscopy was done with an LSM 880 with Airyscan, 40X/1.2 C-Apochromat objective). Excitation/detection wavelengths are 488 nm/505-515 nm for eGFP protein and 594 nm/605-615 nm for Texas Red DHPE. Liposomes were segmented manually in Fiji 1.53c [217] and size and expressed eGFP fluorescence distributions collected using Python (v3.6) with Scikit-image [268]. Details for image analysis are found in Appendix G.

Double-emulsion microfluidics

CFES reactions were encapsulated into liposomes using a double-emulsion microfluidic device and methodology as presented in [289] and illustrated in Figure 3.2. Inner CFES solutions were prepared with a plasmid DNA template similarly to the bulk CFES experiments (Table 3.1). The lipid oil phase was composed of 1-Octanol (Sigma, USA) with 6.5mM of L-α-phosphatidylcholine (Egg PC) phospholipids (Avanti, USA), and 53.3 μM of 1,1'-Dioctadecyl-3,3,3',3'-Tetramethylindodicarbocyanine, 4-Chloro benzenesulfonate Salt (DiD) fluorescent dye (Invitrogen, USA). The outer aqueous solution was composed of a CFES feeding buffer solution modified from [38], which contains NTPs (6mM ATP (Sigma, USA), 4mM CTP (Sigma, USA), 4 mM UTP (Sigma, USA), 6 mM GTP (Roche, Switzerland)), amino acids (0.5 mM each) (Sigma, USA), 1.5 mM spermidine(Sigma, USA), 1.5 mM dithiothreitol (DTT) (Thermo, USA), 0.02 mM folinic acid (Sigma, USA), 280mM potassium glutamate (Sigma, USA), 20 mM magnesium glutamate (Sigma, USA), 100mM HEPES (Roth, Germany), 480 mM glucose (Sigma, USA), and 2% (w/v) Pluronic F-68 (Gibco, USA) at pH 7.6 (Table 3.2).

Table 3.2: Outer feeding solution buffer recipes for inverse emulsion or microfluidic-generated synthetic cells. Solutions were neutralized to the final pH by adding 15% (w/v) KOH.

Component	Inverse emulsion	Microfluidic
ATP (mM)	5	6
CTP (mM)	1.5	4
UTP (mM)	1.5	4
GTP (mM)	3	6
Amino acids (mM)	0.5	0.5
Spermidine (mM)	1.5	1.5
DTT (mM)	1.5	1.5
Folinic acid (mM)	0.02	0.02
K-Glutamate (mM)	280	280
Mg-Glutamate (mM)	20	20
HEPES (mM)	150	100
Glucose (mM)	200	480
Pluronic acid (% (w/v))	0	2
DFHBI (μM)		
pH	7.35	7.6
Osmolarity (mOsmol/kg)	1339	1441

10 μM DFHBI was added into the outer solution for experiments that used the Broccoli aptamer. An outer solution composed of PURExpress without plasmid template (Table 3.3) was also prepared to test outer solutions with the same or diluted compositions as the inner PURExpress CFES. The dilution mix was composed of the same concentration of 10 μM DFHBI for the Broccoli aptamer, 2% pluronic acid to stabilize the synthetic cells, and glucose to maintain the same osmolarity of the outer solution. Similar to the inverse emulsion phase transfer method, inner and outer solution osmolarities were measured by freezing-point depression and balanced by adding water or 1.5 mM glucose to the outer solution when necessary. To generate synthetic cells, the inner CFES, lipid oil phase, and outer buffer solutions were dispensed through the pretreated microfluidic device using three pressure-regulated pumps (Dolomite Mitos P-Pump, UK) at approximate pressures of 70:80:80 mbar, respectively. This resulted in a flow regime where the inner CFES solution was encapsulated into double-emulsions to form liposomes. Flow regimes and double-emulsion formation in the microfluidic device were monitored under brightfield using a Zeiss Andor Axiovert 200M with a 5x/0.15 Plan-Neofluar Ph1 M27 objective and PCO Dimax S4 Monochrome sCMOS high-speed camera. The cells were collected into a microcentrifuge tube, and then placed in channels made from parafilm channels sandwiched between a microscope slide (76x26x1 mm) and cover slip (24x60 mm). The ends of the channels were sealed with Twinsil Speed silicone (Picodent, Germany) to avoid evaporation. We can generate hundreds to thousands of synthetic cells per microfluidic session, but typically prepare up to a hundred synthetic cells in a glass slide for imaging. Prepared synthetic cell populations were imaged by confocal laser scanning microscopy using an inverted Zeiss LSM 880 with Airyscan and a 10X/0.45 Plan-Apochromat M27 objective. The samples were maintained at 30 °C. Laser excitation wavelengths were 488 nm, 488 nm, 561 nm, 633 nm for Broccoli RNA, eGFP protein, mCherry protein, and DiD dye, respectively.

Emission wavelengths were 499-561 nm, 499-561 nm, 579-641 nm, and 640-720 nm for Broccoli RNA, eGFP protein, mCherry protein, and DiD dye detection, respectively. Images were focused at the equator of the synthetic cells and then acquired every 5 minutes for a total of 12 hours. Z-stack images of the samples were taken at the 12-hour endpoint. Timelapse and z-stack images were processed using Fiji (v1.53c) [217] and Python (v3.6) with Scikit-image [268]. Synthetic cells were segmented and fluorescence values for each cell were taken and converted to concentration units using a linear calibration curve from serial dilutions of purified eGFP protein, mCherry protein, and Broccoli RNA in bulk CFES reaction solutions with the same confocal microscopy acquisition settings. Further details for calibration, microfluidic chip fabrication and pretreatment, and image analysis are available in Appendix E, F, and G, respectively.

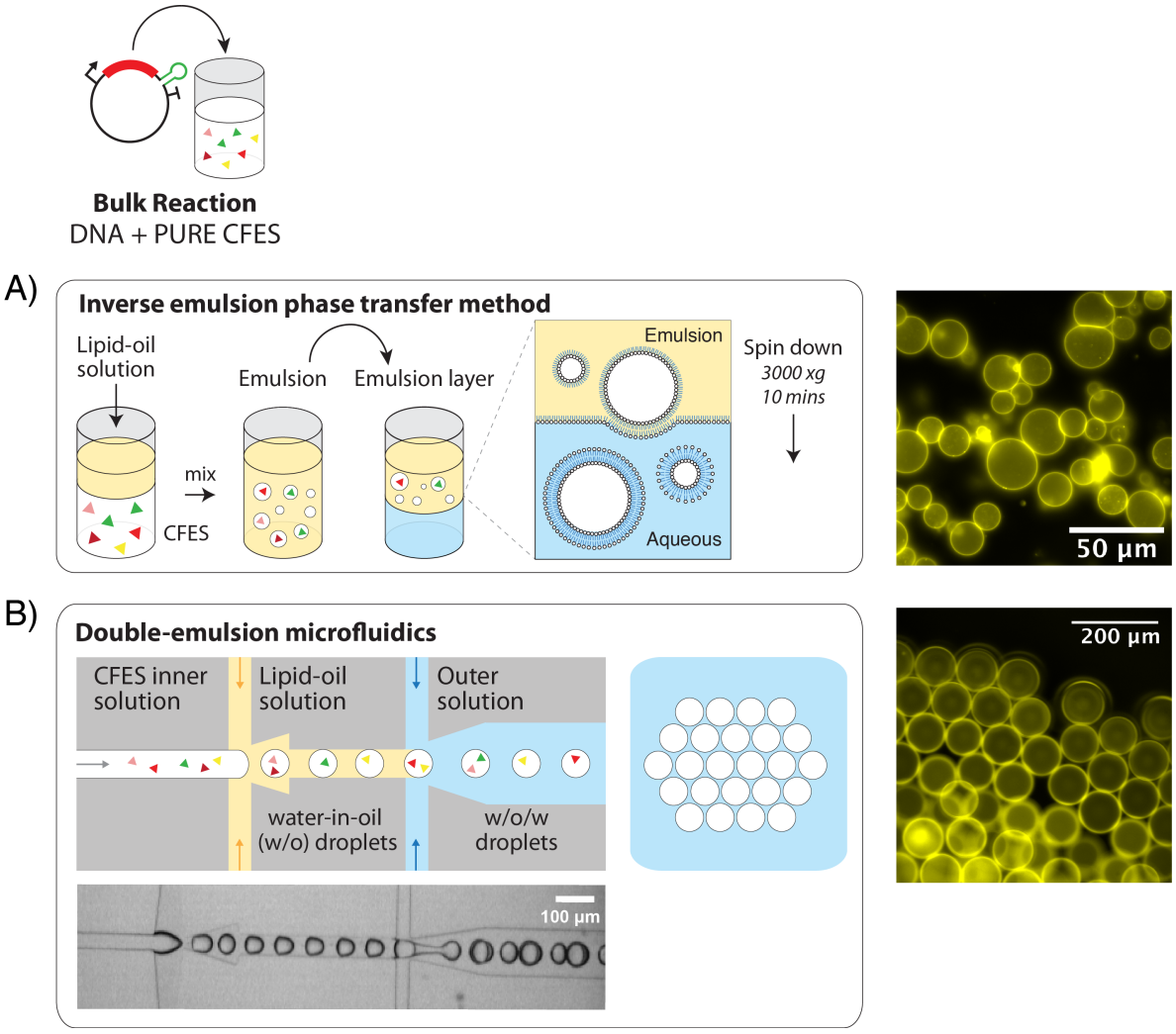

Figure 3.2: Encapsulation of CFES by inverse emulsion and microfluidics. (A) Inverse emulsion phase transfer results in smaller and polydisperse liposomes. (B) The double-emulsion microfluidic generated liposomes results in the larger and more monodisperse liposomes. Liposomes are composed of POPC phospholipids for the inverse emulsion phase transfer-made liposomes and Egg PC for the microfluidic-generated liposomes with Texas Red DHPE encapsulating a 200 mM sucrose solution. The outer solution is composed of 200 mM glucose. A brightfield image of the microfluidic device during liposome production is shown below the illustration.

Table 3.3: PURExpress outer feeding solution and dilution buffer recipes.

Component	PURE outer	Dilution buffer
Solution A	100	0
Solution B	75	0
RNAse inhibitor (NEB)	5	0
DFHBI (500 μM/ 4 mM)	5	3
Pluronic acid (10% (w/v))	50	240
Glucose (1.5 M)	13.3	815
Water	13.2	142
TOTAL	261.5	1200
Osmolarity (mOsmol/kg)	1363	1351

3.2.4 CFES model selection and parameter estimation

A cell-free gene expression model was developed based on a previously published the resource-limited gene expression model [247]. We tested seven variations of the model, including the original model by Stögbauer *et al.* (2012) [247], using mass-action or Michaelis-Menten kinetics for transcription and translation, as well as the degradation and consumption of transcription resources and translation resources. A detailed derivation of the seven models are described in Appendix H. These models were fit on the Broccoli RNA aptamer and mCherry protein timeseries data from our bulk experiments. The agreement between the experimental data and model was measured by the negative natural logarithm of the likelihood (L) of the model parameters given the experimental data

$$-ln(L) = -ln(p(\boldsymbol{Y}|\boldsymbol{\theta})) \tag{3.1}$$

where $\boldsymbol{\theta} = \boldsymbol{\theta_1}, ..., \boldsymbol{\theta_k}$ is the set of parameters for the model and $\boldsymbol{Y}$ is the experimental data. The term on RHS is the log-likelihood of observing data $\boldsymbol{Y}$ given model parameters $\boldsymbol{\theta}$. Rate parameters of a model were estimated by minimizing the log-likelihood

$$\hat{\boldsymbol{\theta}} = \text{argmin}_{\boldsymbol{\theta}}(-ln(p(\boldsymbol{Y}|\boldsymbol{\theta}))) \tag{3.2}$$

where $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator (MLE) of the model parameters. The different models were ranked according to the Akaike information criterion (AIC) [3]

$$\text{AIC} = 2k - 2ln(\hat{L}) \tag{3.3}$$

where k is the number of parameters and $\hat{L}$ is the likelihood evaluated at the MLE $\hat{\boldsymbol{\theta}}$. The AIC measures the quality of fit of the model to the data using the likelihood penalized by the total number of parameters. The model with the lowest AIC value was selected and used for both bulk and synthetic cell population experiments. Profile likelihoods and likelihood-based confidence intervals from the parameter estimates were calculated to assess parameter identifiability in the different models [208, 163]. The profile likelihoods of each parameter were calculated by

$$\text{PL}(\boldsymbol{\theta_i}) = \min_{\boldsymbol{\theta}_{j\neq i}}(-ln(p(\boldsymbol{Y}|\boldsymbol{\theta}))) \tag{3.4}$$

which is the minimum of the negative log-likelihood with respect to all parameters $\boldsymbol{\theta}_{j\neq i}$ while holding the parameter θ_i fixed. Likelihood-based confidence intervals of each

parameter were estimated by the regions in

$$\{\theta_i | \text{PL}(\theta_i) + ln(p(\boldsymbol{Y}|\hat{\boldsymbol{\theta}})) < \chi^2(\alpha, df)\} \tag{3.5}$$

where $\chi^2(\alpha, df)$ is the chi-squared distribution with $\alpha = 0.95$ confidence level and df degrees of freedom, which is the number of parameters of the model [208]. Graphically, this means that the confidence intervals are found at the intersections of the profile likelihood and the significance threshold horizontal line at $-ln(p(\boldsymbol{Y}|\hat{\boldsymbol{\theta}})) + \chi^2(\alpha, df)$. Profile likelihoods of each parameter between a range greater and less than $\hat{\boldsymbol{\theta}}$ by a factor of 10 are calculated to plot the profile likelihoods. 95% likelihood-based confidence intervals of the parameters from the profile likelihoods are calculated from the upper and lower intersection of the profile likelihood and the significance threshold line. Structural non-identifiable parameters are indicated by flat profile likelihoods below the significance threshold in both -∞ and ∞ directions. Practical non-identifiable parameters are indicated by profile likelihoods below the significance threshold in both or either $-\infty$ and/or ∞ directions, but still have a unique minimum [208]. Further details of the model selection, profile likelihood analysis, and parameter estimation are available in Appendix I.

3.3 Results

3.3.1 Quantitative gene expression dynamics in bulk CFES

To monitor both transcription and translation dynamics in CFES, we constructed the pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid. This plasmid consists of a constitutive T7 RNA Polymerase-mediated promoter to express a red fluorescent mCherry protein and two copies of a dimeric Broccoli RNA aptamer stabilized by the F30 stem-loop [78, 76] between the stop codon of mCherry and the terminator of the gene construct (Figure 3.1). Binding of a small molecule dye 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI) to the Broccoli RNA aptamer results in a green fluorescence signal. This allows simultaneous fluorescence monitoring of transcribed mRNA and reporter protein levels. PURExpress CFES reactions were titrated with pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA or purified mRNA transcripts from the same plasmid. Reaction mixtures were incubated at 30 °C and monitored for mRNA and protein levels over time in a fluorescence well plate reader (Figure 3.3A). Relative fluorescence units were converted into nM concentration units using calibration curves from serial dilutions of Broccoli RNA aptamer and mCherry protein in the same reaction mix composition and acquisition settings (*see Appendix*). We observed typical profiles of gene expression in CFES, where signal is first detected from transcription of mRNA followed by translation of mCherry protein (Figure 3.3B-C). The gene expression profiles show a plateau at ~3 hours for mRNA and ~5 hours for protein. Our results also show that the endpoint protein concentrations increase with increasing plasmid DNA or mRNA transcript concentrations until a saturation concentration of approximately 5 nM for plasmid DNA and 800 nM for mRNA transcript. Rates of transcription reach a maximum at the initial point, while translation rates peak at 1.5-2 hours (Figure 3.4) and then gradually decrease. These results indicate that gene expression rates and yield are dependent on both the consumption

and degradation of resources. If gene expression was dependent solely on the consumption of resources the final protein production would be constant regardless of DNA or mRNA input. This hypothesis is supported by previous work where the addition of fresh ribosomes to PURExpress after exhaustion restores gene expression activity [247] and that a delayed addition of DNA template into PURExpress after incubation results in reduced rates and yield of gene expression [247, 59].

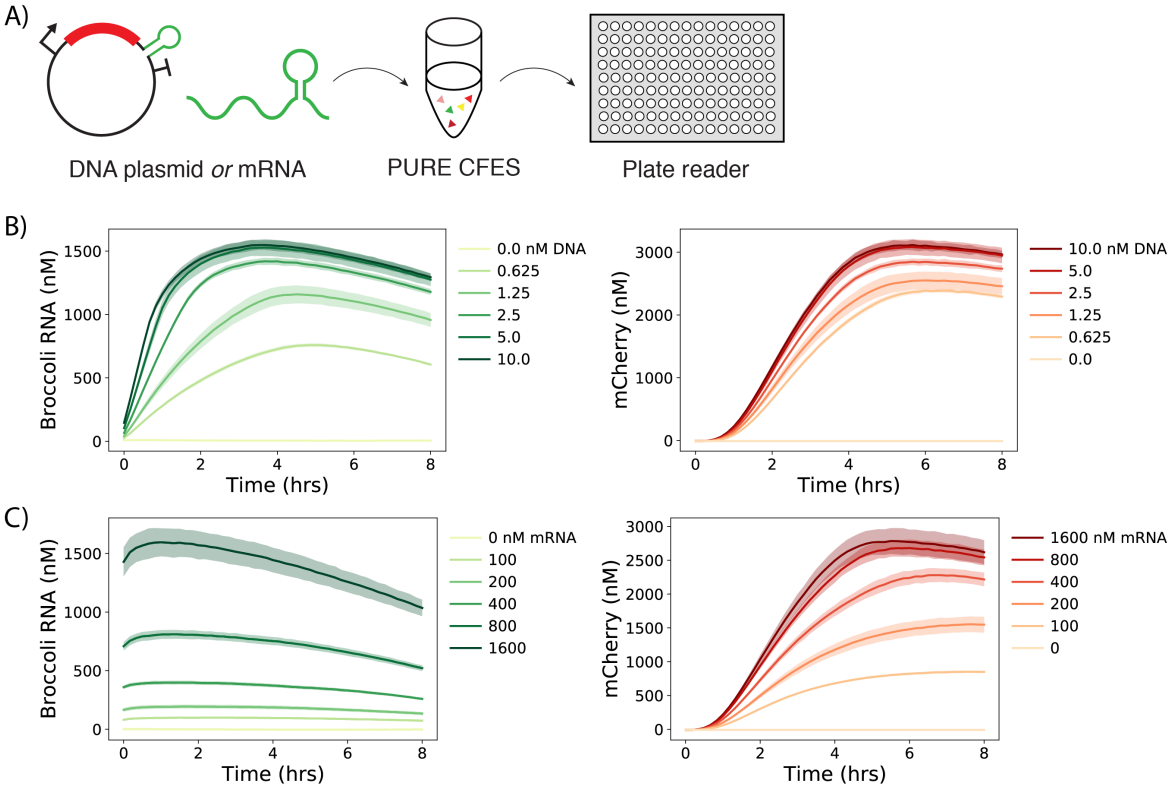

Figure 3.3: Monitoring transcription and translation in bulk CFES. (A) DNA plasmid or mRNA titration bulk CFES experiment workflow. DNA or mRNA templates are added into a master mix of PURExpress CFES (NEB, USA) with DFHBI at different DNA concentrations. Fluoresence signal of both expressed Broccoli RNA aptamer and mCherry protein are monitored over time in a plate reader. RFU values are converted to concentrations (μM) using a standard calibration curve. (B) mRNA Broccoli and mCherry protein expression levels over time from bulk PURExpress CFES titrated with varying concentrations of pEXP5-NT/6xHis mCherry F30-2xdBroccoli DNA plasmid. (C) mRNA and mCherry protein expression levels over time from bulk PURExpress CFES titrated with varying concentrations of purified 6xHis mCherry F30-2xdBroccoli RNA transcripts. Solid lines and shaded areas correspond to mean and standard deviation values from triplicate experiments.

In addition to DNA and mRNA titrations, the maturation rate of the mCherry protein in PURExpress was measured using a modified RNAseA assay previously described by Garamella *et al.* (2016) [85]. Briefly, a standard bulk PURExpress reaction with DFHBI was prepared with 300 nM 6xHis mCherry F30-2xdBroccoli mRNA template. The reaction was incubated at 30 °C for 30 mins to allow the CFES to produce mCherry protein. After 30 mins, 300 nM of RNAseA was added into the CFES reaction to degrade all mRNA and halt translation. mRNA Broccoli aptamer and mCherry protein

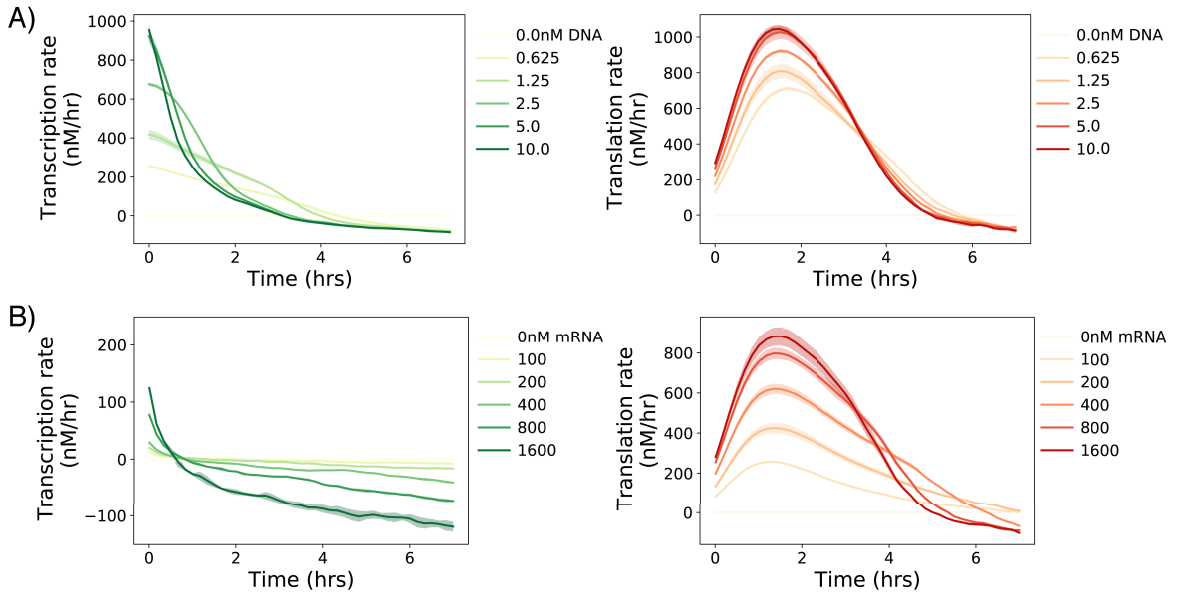

Figure 3.4: Transcription and translation rates in bulk CFES. (A) Transcription and translation rates over time from bulk PURExpress CFES titrated with varying concentrations of pEXP5-NT/6xHis mCherry F30-2xdBroccoli DNA plasmid. (B) Transcription and translation rates over time from bulk PURExpress CFES titrated with varying concentrations of purified 6xHis mCherry F30-2xdBroccoli RNA transcripts. Solid lines and shaded areas correspond to mean and standard deviation values from triplicate experiments.

fluorescence values were monitored (Figure 3.5A). The increase of mCherry signal after all mRNA is degraded is attributed to the maturation of the protein and fit to a first order kinetic model

$$\frac{d[\text{Protein}^*]}{dt} = k_{\text{mat}}[\text{Protein}] \tag{3.6}$$

$$[\text{Protein}^*] = [\text{Protein}^*]_0 + [\text{Protein}]_0(1 - e^{-k_{\text{mat}}t}) \tag{3.7}$$

where Protein is the unfolded protein, Protein* is the mature and fluorescent protein, and k_{mat} is the maturation rate parameter. The RNA signal was depleted after 2 hours after the addition of RNAseA (Figure 3.5B). Fitting the increase of mCherry fluorescence in the 2-4 hour window using Equation 3.7 results in a maturation rate of $k_{\text{mat}} = 2.15 \pm 0.12$ hr^{-1} (Figure 3.5C). This corresponds to a maturation half-time ($t_{0.5}$) of 19.31 ± 2.24 mins, which is also comparable to previous reports of mCherry maturation at 15 mins in E. coli [224].

3.3.2 CFES model selection and parameter fitting

To describe the dynamics of cell-free gene expression, a coarse-grained model based on Stögbauer et al. (2012) [247] was developed to quantitatively compare results across experiments and literature values. This model accounts for both transcription and translation dynamics driven by a limited pool of resources for gene expression. Transcription (TsR) and translation resources (TlR), are assigned unitless quantities initialized at 1 and then gradually decreases to 0 as it is consumed by transcription or translation, and degradation. These species serve as a phenomenological proxy to account for the cumulative effect of different limiting factors that fuel transcription and

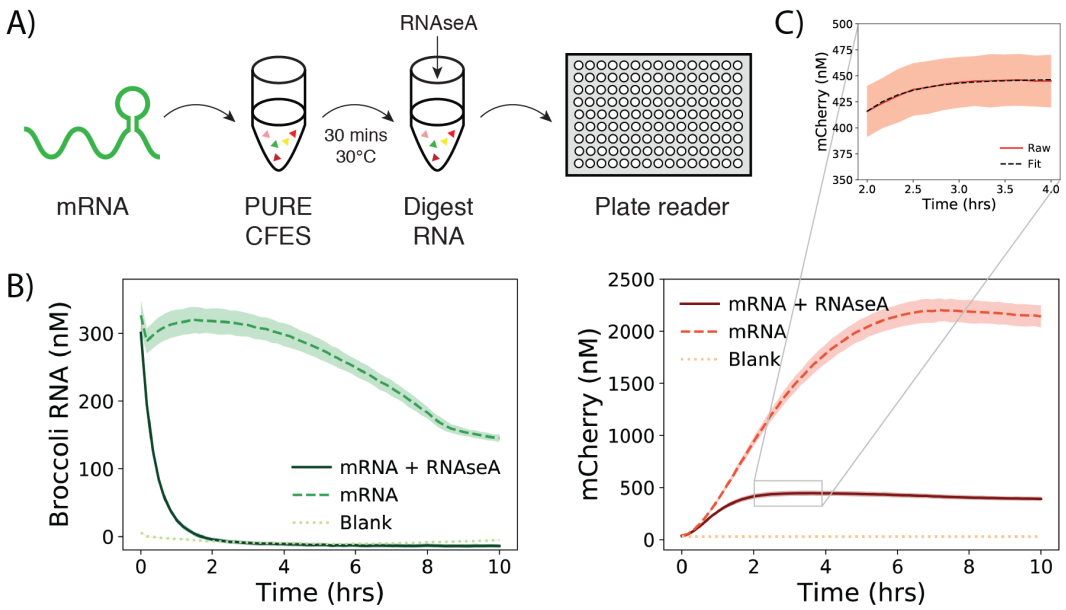

Figure 3.5: RNAseA assay for mCherry maturation rate. (A) 300 nM 6xHis mCherry F30-2xdBroccoli mRNA is added into the PURExpress CFES (NEB, USA) and incubated for 30 mins at °C. Afterwards, 300 nM RNAseA is added to deplete the mRNA and then protein expression is monitored in the plate reader. (B) mRNA Broccoli and mCherry protein levels over time after addition of RNAseA. Solid line is the sample with both mRNA and RNAseA added, dashed line is for mRNA without RNAseA, and dotted line is the blank. (C) Inset from mCherry signal in (B) from 2-4 hours showing the fit of Equation 3.7 where $k_{\text{mat}} = 2.15 \pm 0.12$ hr^{-1}. Solid red line is the mean of the data and black dashed line is the model fit. Shaded areas correspond to mean and standard deviation values from triplicate experiments.

translation processes, such as RNA polymerase, ribosome concentrations, NTP, amino acids, and other energy resources. Based on this model, we generated seven candidate CFES models composed of a system of delay and ordinary differential equations (Appendix H). We used our bulk experimental results from both transcription and translation dynamics, as well as the RNAseA assay for mCherry maturation, to guide model selection. Candidate models were ranked among each other using the Akaike information criterion (AIC) [3]. Profile likelihoods were then used to determine the parameter identifiability and confidence intervals for each of the candidate models [208, 163]. The best-scoring model resulting from this analysis is shown in Equation 3.8-3.13 (model 2 in Appendix H).

$$\frac{d[\text{DNA}]}{dt} = 0 \tag{3.8}$$

$$\frac{d[\text{RNA}]}{dt} = \frac{k_r \text{TsR}[\text{DNA}]}{K_r + [\text{DNA}]} - \delta_r[\text{RNA}] \tag{3.9}$$

$$\frac{d[\text{Protein}]}{dt} = \frac{k_p \text{TlR}[\text{RNA}](t-\tau_d)}{K_p + [\text{RNA}](t-\tau_d)} - k_{\text{mat}}[\text{Protein}] \tag{3.10}$$

$$\frac{d[\text{Protein}^*]}{dt} = k_{\text{mat}}[\text{Protein}] \tag{3.11}$$

$$\frac{\text{TsR}}{dt} = -\frac{a k_r \text{TsR}[\text{DNA}]}{K_r + [\text{DNA}]} - \delta_{\text{TsR}}\text{TsR} \tag{3.12}$$

$$\frac{d\text{TlR}}{dt} = -\frac{b k_p \text{TlR}[\text{RNA}](t-\tau_d)}{K_p + [\text{RNA}](t-\tau_d)} - \frac{\delta_{\text{TlR}}\text{TlR}}{K_l + \text{TlR}} \tag{3.13}$$

This model uses Michaelis-Menten type kinetics for transcription and translation. Translation is additionally modeled by a delay differential equation with a time delay (τ_d) to account for the time delay of protein expression observed in our mRNA titration experiments, which was necessary despite having a protein maturation step. Transcription and translation resources (TsR and TlR) are consumed by transcription and translation processes and also degraded independently. These are consumed during transcription and translation with a scaling factor, a and b respectively. Both resources spontaneously degrade with first order and Michaelis-Menten kinetics for TsR and TlR, respectively. RNA degradation and mCherry protein maturation are assigned first-order reactions. In contrast to the previously published model [247], TsR degradation was included to account for the independent exhaustion of transcription resources. Lastly, we included a time lag (τ_l) in the fitting procedure of the model to account for the time between starting the CFES reaction and acquiring the first data point. This was negligible for our bulk experiments that took less than 10 minutes from adding the DNA or RNA template into the CFE bulk reactions to acquiring the first data points in the plate reader. However, it was important for the encapsulated experiments which had longer sample preparation times of ~30 minutes. The final model is illustrated in Figure 3.6A with a total of 16 free parameters. These parameters include all the rate parameters k_r, K_r, δ_r, k_p, K_p, δ_{TsR}, δ_{TlR}, K_l, a, b, τ_d, τ_l, σ_r, and σ_p. The parameters σ_r and σ_p are standard deviation parameters in a gaussian noise model for the Broccoli RNA and mCherry measurements, respectively. k_{mat} is fixed at 2.15 ± 0.12 hr^{-1} from the RNAseA assay for mCherry maturation (Figure 3.5). Initial values of [DNA] and [RNA] are defined by the experiment. Initial [Protein] and [Protein*] values are set at zero. Initial transcription and translation resources, TsR and TlR, are set at 1.0. The model fit by MLE is shown in Figure 3.6B-C. The optimized rate parameters and likelihood-based confidence intervals are shown in Table 3.4. Profile likelihoods of each parameter are shown in Figure 3.7.

All parameters except for K_l and τ_l are well-identifiable as indicated by profile likelihoods that have a local minimum at $\hat{\theta}_i$ (Figure 3.7). The estimates of K_l and τ_l are in the order of 10^{-6} and 10^{-9}, respectively. Varying these parameters within one order of magnitude does not significantly affect the model fit, which results in their flat profile likelihoods. Overall, the model captures the general behavior of gene

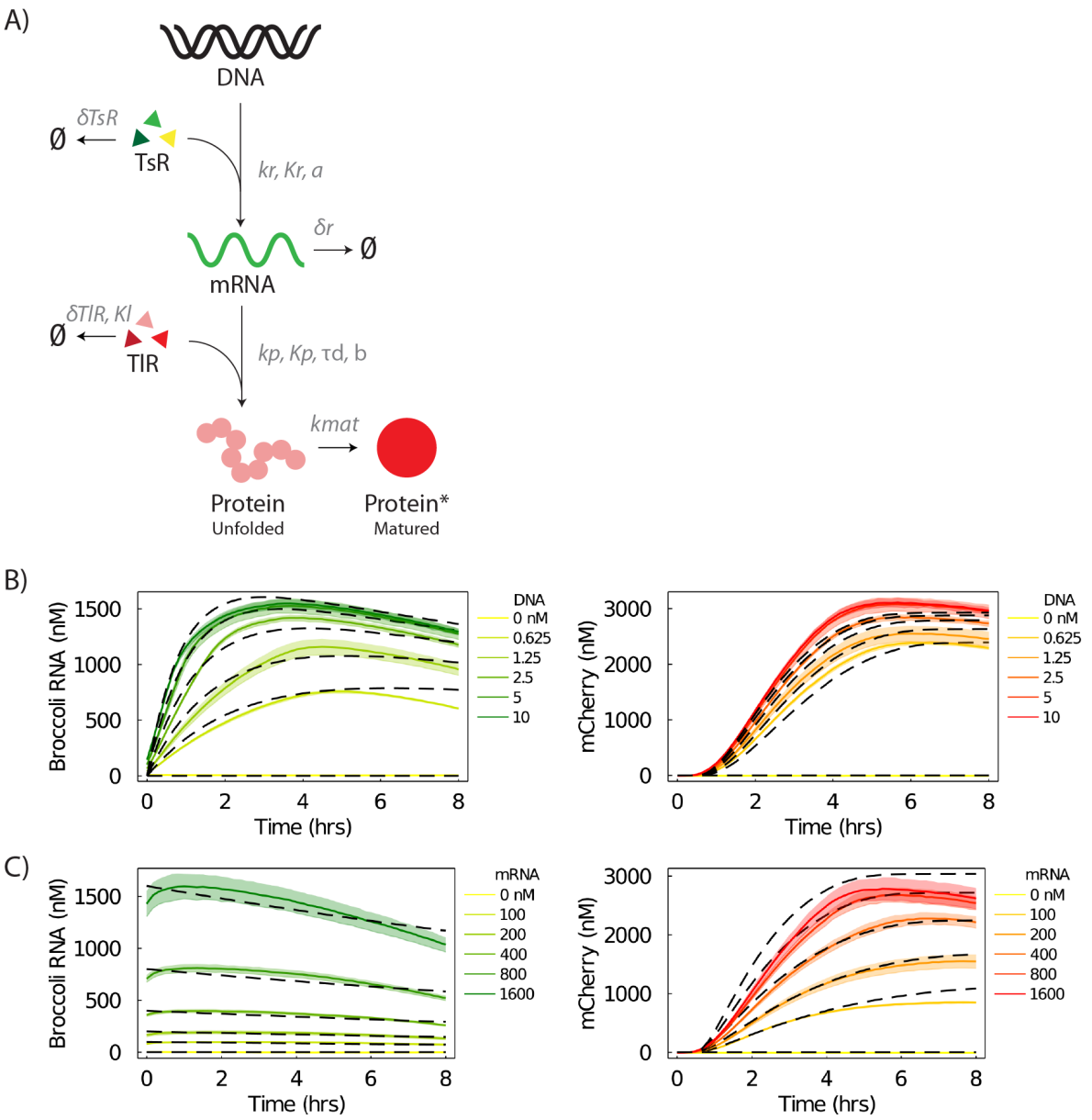

Figure 3.6: Resource-limited CFES model and bulk fit. (A) Illustration of the resource-limited gene expression model for CFES. Parameters are k_r: RNA transcription rate, K_r: Dissociation constant between RNAP and DNA, δ_r: RNA degradation rate, k_p: Protein translation rate, K_p: Dissociation constant between ribosome and RNA, k_{mat}: mCherry maturation rate, δ_{TsR}: TsR degradation rate, δ_{TlR}: TlR degradation rate, K_l: Michaelis-Menten constant for TlR degradation, a: Scaling factor for consumption of TsR with transcription, b: Scaling factor for consumption of TlR with translation, and τ_d: Time-delay for protein translation. (B) Experiment (colored lines) and model fit (black dashed lines) of Broccoli mRNA and mCherry protein expression levels over time from bulk PURExpress CFES titrated with varying concentrations of pEXP5-NT/6xHis mCherry F30-2xdBroccoli DNA plasmid. (C) Experiment (colored lines) and model fit (black dashed lines) of Broccoli mRNA and mCherry protein expression levels over time from bulk PURExpress CFES titrated with varying concentrations of purified 6xHis mCherry F30-2xdBroccoli RNA transcripts. Solid lines and shaded areas are the mean and standard deviations of the experimental data from three technical replicates.

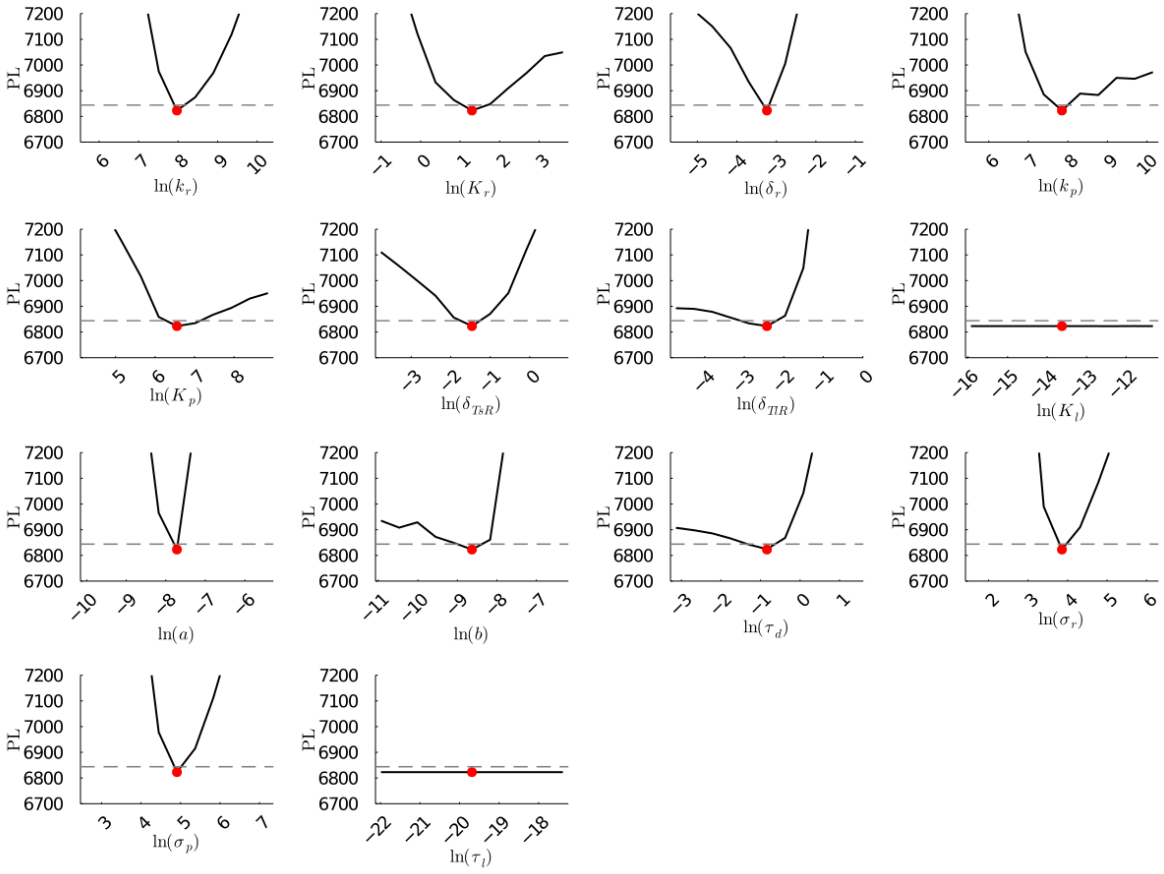

Figure 3.7: Profile likelihoods of bulk CFES model parameters. Each plot corresponds to a parameter in the model and additional fitting parameters (σ_r and σ_p). The y-axis of each plot is the negative log likelihood of the model given the corresponding parameter value in the x-axis with all other parameters reoptimized (Equation 3.4). The red dot shows the optimized parameter set with the minimum negative log likelihood. The dashed grey line is the 95% significance threshold line from (Equation 3.5). The intersections of the significance threshold line and profile likelihood are the likelihood-based confidence intervals of the optimized parameter.

expression dynamics across different initial DNA and RNA conditions. The remaining quantitative mismatch is likely due to additional chemical complexity not caputured by our coarse-grained model. Assuming that a standard PURExpress reaction contains 100 nM of T7 RNA polymerase [228] and 2.4 μM of ribosomes (NEB), T7 RNA polymerase transcription and ribosome translation rates are approximately 8.2-11.1 NTP/s and 0.20-0.28 amino acid/s, respectively (calculated from $k_r = 2728 - 3674$ nM/hr for a 1087 bp transcript and $k_p = 2211 - 3108$ nM/hr for a 777 aa protein in Table 1). These values are lower than the reported in vivo rates in E. coli bacterial cells (230 ± 20 NTP/s [93] and 8-18 amino acid/s [44]). However, the polymerase transcription rates and ribosome translation rates are similar to previous work in PURExpress expressing GFP at 37 °C (2.2 NTP/s and 0.03 amino acid/s, respectively) [247]. Using a FRET sensor to measure RNA transcription in PURExpress, initial transcription rates from 10 nM of DNA plasmid template was previously measured at 7 nM/min [186]. This is also comparable to our initial transcription rate measurements at 15.9 nM/min for 10 nM of DNA plasmid (Figure 3.4)A. The differences could be attributed to different

Table 3.4: Parameter estimates ($\hat{\boldsymbol{\theta}}_{\boldsymbol{bulk}}$) and likelihood-based 95% confidence intervals (CI) from bulk DNA and RNA titration experiments. Parameters with CIs at -∞ and/or ∞ are non/weakly-identifiable within one order of magnitude from $\hat{\boldsymbol{\theta}}_{\boldsymbol{bulk}}$.

Parameter	Description	$\hat{\boldsymbol{\theta}}_{\boldsymbol{bulk}}$	95% CI	Units
k_r	RNA transcription rate	2894	2728–3674	nM/hr
K_r	Dissociation constant between RNAP and DNA	3.67	2.89–5.68	nM
δ_r	RNA degradation rate	0.0392	0.0361–0.0422	1/hr
k_p	Protein translation rate	2689	2211–3108	nM/hr
K_p	Dissociation constant between ribosome and RNA	703	530–1347	nM
k_{mat}	mCherry maturation rate	2.15	($\pm$0.12)	1/hr
δ_{TsR}	TsR degradation rate	0.231	0.171–0.298	1/hr
δ_{TlR}	TlR degradation rate	0.0884	0.0441–0.1187	1/hr
K_l	Michaelis-Menten constant for TlR degradation	1.21E-6	-∞–∞	-
a	Scaling factor for consumption of TsR with transcription	4.45E-4	4.18E-4–4.57E-4	-
b	Scaling factor for consumption of TlR with translation	1.78E-4	1.18E-4–2.42E-4	-
τ_d	Time-delay for protein translation	0.433	0.254–0.560	hr
σ_r	Stdev. of Broccoli RNA measurements	47.3	44.8–54.8	nM
σ_p	Stdev. of mCherry protein measurements	135.6	128.0–156.0	nM
τ_l	Time lag between reaction start and data collection	2.81E-9	-∞–∞	hr

reaction conditions, T7 RNA polymerase concentrations, the encoding gene, and/or batch-to-batch variability of the expression system.

3.3.3 Generating monodisperse synthetic cell populations

Having established a quantitative model for cell-free gene expression in bulk reactions, we next test its applicability on populations of compartmentalized reactions. To this end, PURExpress CFES was encapsulated in lipid-based synthetic cell populations using either a bulk inverse emulsion phase transfer method [177] or a double-emulsion microfluidic methodology [202, 289] as described in Section 3.2.3. The inner solution was composed of the PURExpress CFES and a plasmid DNA for constitutive T7 RNAP-mediated expression of a fluorescent protein gene (eGFP or mCherry). Confocal microscopy images for the inverted emulsion and microfluidic-generated synthetic cells were segmented to obtain relative fluorescence units (RFU) of expressed protein in single cells in each population. These were then used to calculate the coefficient of variation (CV) of the distribution of expressed protein in each cell population. The CV allows comparison of the variability of distributions with different scales of measure-

ment. The inverted emulsion method generated liposomes with a mean radius of 8.0 μm and a coefficient of variation (CV) of 0.32 (Figure 3.8A and D). In comparison, the microfluidic-generated synthetic cells were larger with a mean radius of 29.0 μm and exhibited lower size variation with a CV of 0.09 (Figure 3.8B and E) as expected. Protein expression in the inverted emulsion-generated synthetic cells also showed a greater variation (mean RFU 126.72 and CV 0.49) compared to synthetic cells produced in microfluidics (mean RFU 37.0 and CV 0.05). These results are in agreement with previous studies of phase transfer generated cells with expressed protein concentration CVs ranging from 0.20-0.80 [190]. We further show that simultaneous encapsulation of two plasmids in a microfluidic-generated synthetic cell population results in expression of both eGFP and mCherry protein in each cell at a consistent ratio (3.11 ± 0.133 eGFP:mCherry RFU) (Figure 3.8C and F). This demonstrates the robustness of our synthetic cell production where the inner CFES solution is well-mixed and microfluidic method maintains the homogeneity throughout encapsulation. The increased variance in phase transfer-generated cells is likely a combined result of fluctuations in cell size and encapsulation. In contrast, the synthetic cell populations generated using double-emulsion microfluidics resulted in larger and more uniform cell populations making them highly suitable for our quantitative analysis. In addition, it was also observed that the fluorescence from expression of the pEXP5-NT/6xHis eGFP plasmid is decreased in the two plasmids synthetic cells (Mean RFU 14.1, Figure 3.8F) as compared to the single plasmid synthetic cells (Mean RFU 37.0, Figure 3.8E). This is a result of gene expression resources being split between the expression of both eGFP and mCherry proteins in the two plasmid synthetic cells.

3.3.4 Gene expression dynamics in synthetic cell populations

Using our microfluidic platform, we generated synthetic cell populations comprised of large populations of monodisperse liposome-encapsulated CFES and quantified RNA and protein levels over time to study transcription and translation dynamics using fluorescence microscopy methods. To alleviate non-identifiabilities during model fitting, we prepared three populations of synthetic cells with different DNA concentrations (1.75, 3.5, and 7.0 nM of pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA) during one microfluidic session from one batch of CFES master mix and outer feeding buffer solution. RNA and protein levels in the synthetic cell populations were then monitored with confocal microscopy at 30 °C for 12 hours (Figure 3.9). Similar to the bulk CFES experiments, relative fluorescence units were converted into absolute concentrations using a standard calibration curve to obtain quantified RNA and protein dynamics in the synthetic cell populations (Figure 3.10 and 3.11). Cell sizes from the three populations containing different plasmid DNA concentrations were monodisperse at $\sim$30 μm radius with a coefficient of variation (CV) ranging from 0.04-0.065. The variability of gene expression from mRNA to protein remained constant with CV values ranging from 0.02-0.03 (Table 3.5). This indicates a low degree of variability in translation across the synthetic cells as CV values were not altered between mRNA and protein levels. Based on cell size and concentration measurements, copy numbers of DNA, mRNA, and protein molecules in a single synthetic cell are estimated to be in the order of 10^5, 10^7, and 10^8, respectively. Other components required for gene expression in the PUR-

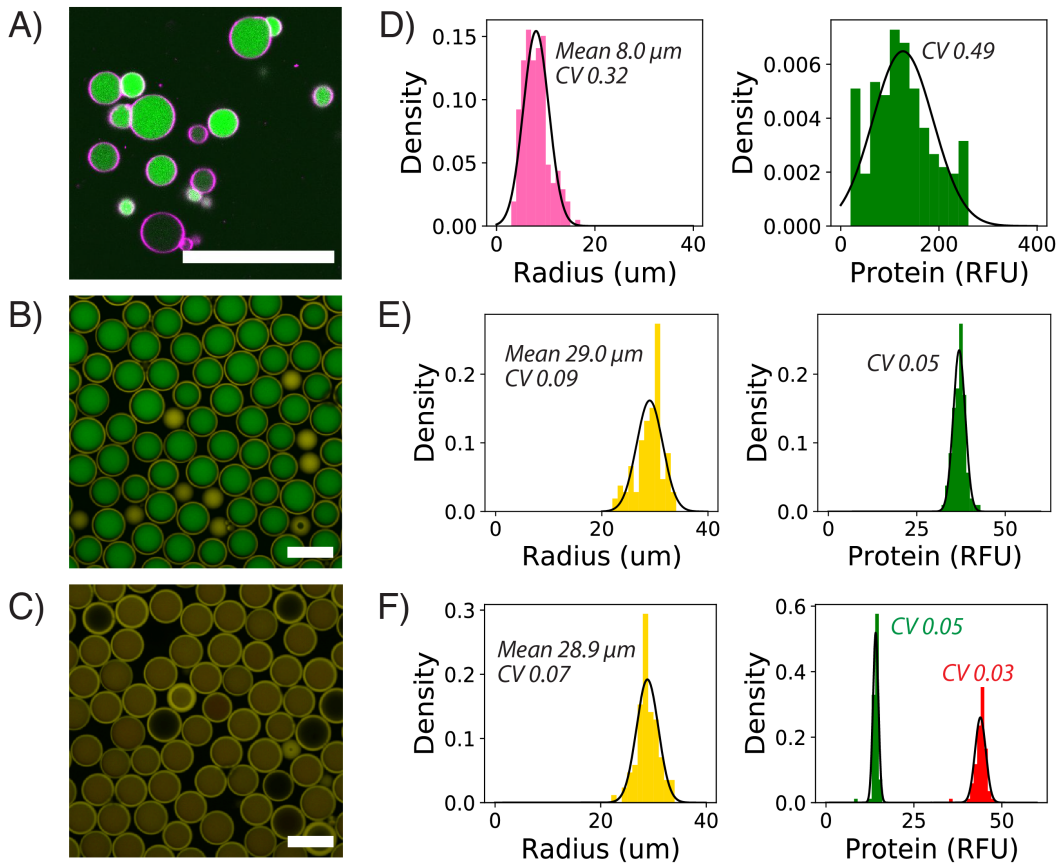

Figure 3.8: Variability in synthetic cell populations. (A) Synthetic cell population expressing eGFP protein from 1.17 nM pEXP5-NT/6xHis eGFP plasmid DNA generated using the bulk inverse emulsion phase transfer method. (B) Microfluidic-generated synthetic cells expressing eGFP protein from 4.5 nM pEXP5-NT/6xHis eGFP F30-2xdBroccoli plasmid DNA. (C) Merged image of synthetic cell population expressing both eGFP and mCherry protein from two plasmids (4.5 nM pEXP5-NT/6xHis eGFP and 4.5 nM pEXP5-NT/6xHis mCherry plasmid DNA. Endpoint distributions of radius and protein RFU of the synthetic cell populations in (A-C): (D) phase transfer-generated synthetic cell population (206 cells) expressing eGFP protein, (E) single microfluidic-generated population (106 cells) expressing eGFP protein, and (F) two plasmid population (85 cells) expressing eGFP (green) and mCherry (red) protein. Black lines are Gaussian distribution fits based on the mean and variance of the histograms. Only CV values are listed in the protein histograms since RFU values are not relatively comparable among the different populations. All images are taken at endpoint after 12 hours incubation at 30 °C using confocal microscopy (40X objective for A and 10X for B and C). Scale bars are all 100 μm.

Express CFES are also present in similar or higher concentrations [190, 227], such that stochastic effects associated with low copy numbers should be virtually absent. Time scales of active gene expression were comparable between bulk (Figure 3.3B) and encapsulated reactions (Figure 3.10A) (approx. 8-13 hrs). Maximum gene expression rates and endpoint mRNA and protein concentrations differ between the bulk expression and compartmentalized expression (Figure 3.3-3.4 and Figure 3.10-3.11). In particular, protein expression in the liposomes are consistently lower than in bulk reactions. We hypothesized that deviations of gene expression dynamics between bulk and encapsulated formats are due to the different chemical conditions. Specifically, the composition

of the outer feeding buffer solution can affect the inner CFES reaction by diffusion of material across the semi-permeable membrane [192]. In preparing inner and outer solutions, we ensured that inner CFES and outer solutions were osmotically balanced by matching freezing-point osmometer measurements. However, the outer feeding buffer and inner PURExpress CFES were still prepared with slightly different compositions.

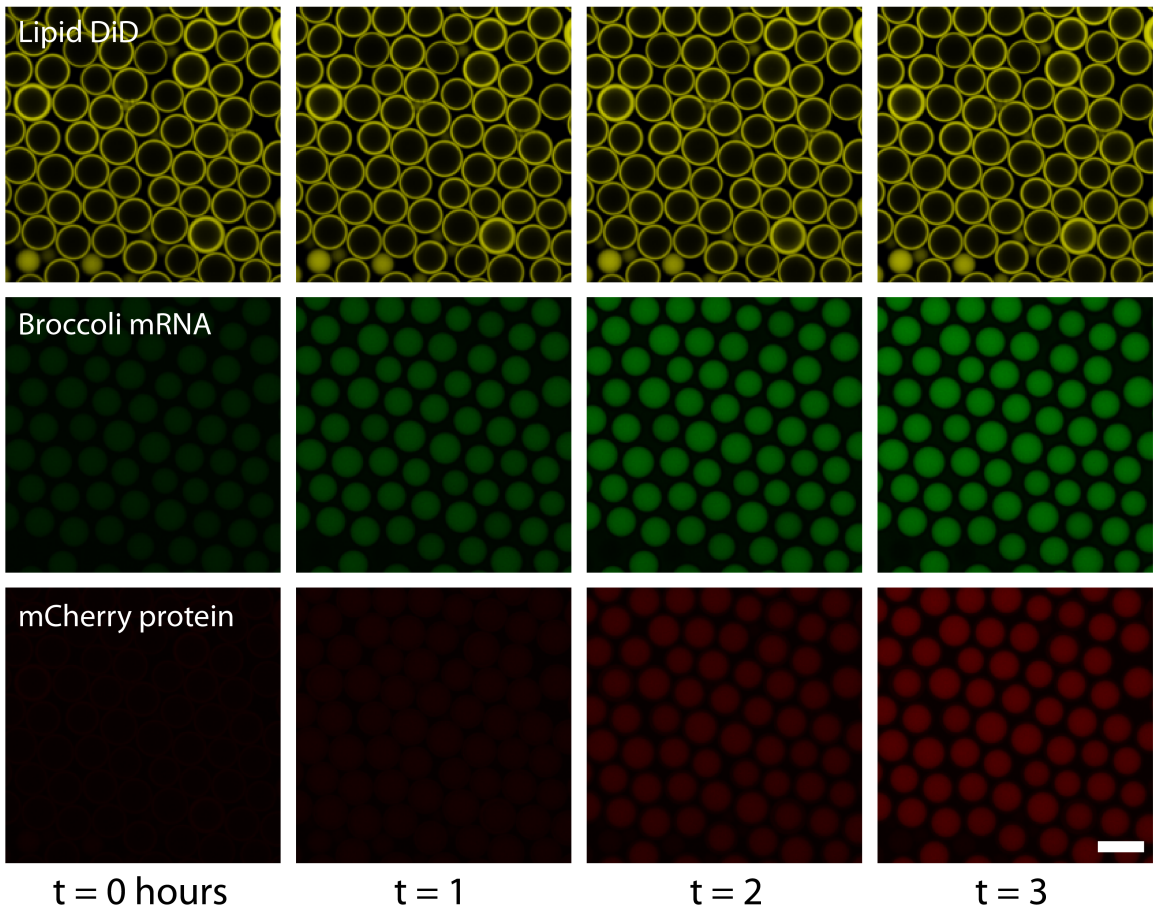

Figure 3.9: Timelapse confocal images of a synthetic cell population containing PURE CFES and 3.5 nM pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA. Images are divided into three channels: DiD dye-tagged lipid membrane (top row), Broccoli mRNA (middle row), and mCherry protein (bottom row). Timelapse images were taken every 5 mins for a total of 12 hours with incubation at 30 °C using confocal microscopy. Scale bars are all 100 μm.

Table 3.5: Endpoint gene expression of synthetic cell populations. Mean and standard deviation size and endpoint mRNA and protein expression of microfluidic-generated synthetic cell populations. Values in parentheses are the coefficients of variation (CV). Total number of cells analyzed are 82, 85, and 78 for populations with 1.75, 3.5, and 7.0 nM of pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA, respectively.

DNA (nM)	**Radius (μm)**	**Broccoli RNA (nM)**	**mCherry (nM)**
1.75	29.8 ± 1.4 (0.048)	759.0 ± 17.3 (0.022)	1240.5 ± 38.4 (0.031)
3.5	30.4 ± 1.9 (0.064)	973.8 ± 21.6 (0.022)	1892.7 ± 51.5 (0.028)
7.0	32.0 ± 1.4 (0.043)	1093.7 ± 33.5 (0.03)	1973.9 ± 59.1 (0.030)

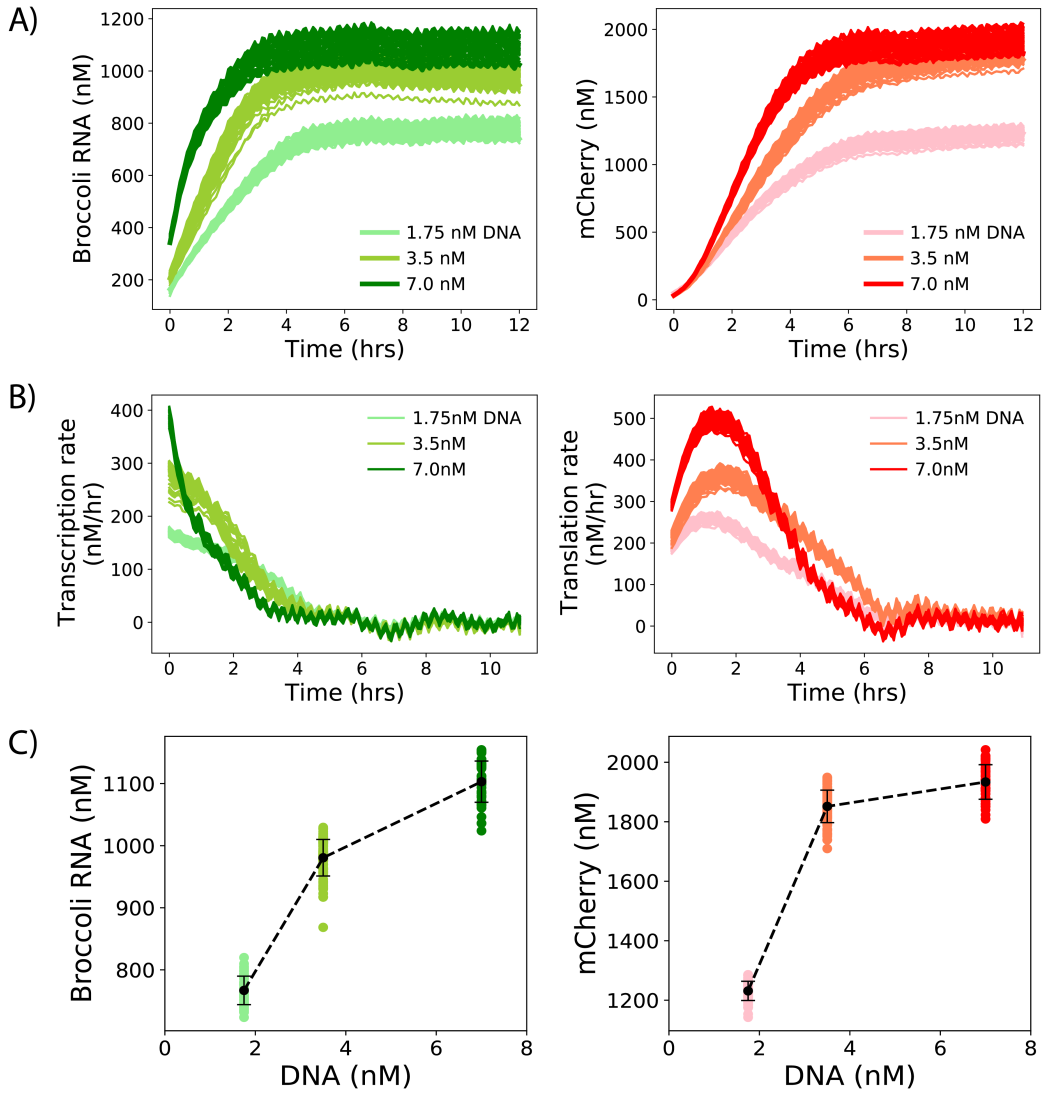

Figure 3.10: Quantified transcription and translation dynamics in synthetic cell populations. (A) Single-cell traces of mRNA and protein expression in three synthetic cell populations with 1.75 nM, 3.5 nM, and 7.0 nM of pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA. RFU values are converted into nM concentrations units. (B) Rates of transcription and translation calculated from (A) using a rolling window average of 1 hour. (C) Endpoint RNA and protein concentrations of synthetic cell populations at the different DNA concentrations. Each dot is a single synthetic cell at endpoint. Black dots with error bars are mean and standard deviation of the populations. Total number of cells analyzed are 82, 85, and 78 for populations with 1.75 nM, 3.5 nM, and 7.0 nM of pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA, respectively.

To quantify gene expression dynamics, mean RNA and protein dynamics from all three synthetic cell populations were globally fit to the resource-limited CFES model in Equation 3.8-3.13 (Figure 3.12). Sample preparation of the synthetic cells typically took 0.5-1 hour due to the encapsulation of different plasmid concentrations. As a result, the initial points of gene expression were not fully captured in the timeseries data. The fitted rate parameters and 95% confidence limits obtained by fitting the experimental data to the model are shown in Table 3.6. The identifiable parameters are comparable to

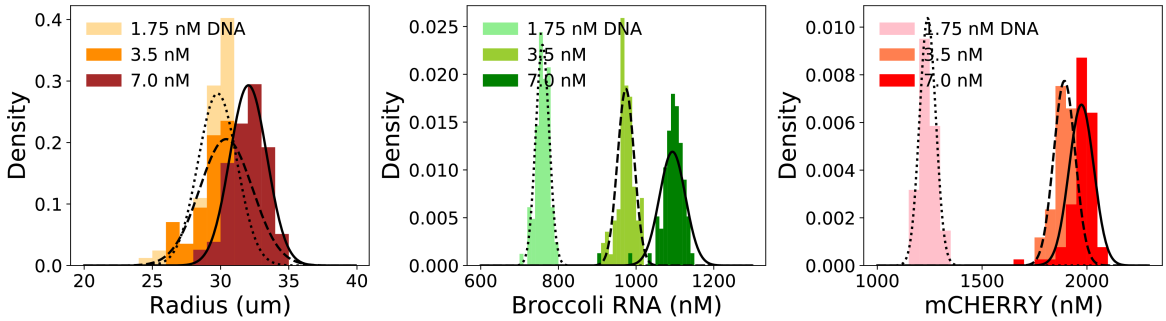

Figure 3.11: Endpoint distributions of radius, mRNA, and protein concentrations of the synthetic cell populations. Data is obtained from maximum values of z-stack images at endpoint after 30 °C incubation for 12 hours. Histograms are fit to univariate normal distributions and correlations are fit to bivariate normal distributions. Total number of cells analyzed are 82, 85, and 78 for populations with 1.75 nM, 3.5 nM, and 7.0 nM of pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA, respectively.

the bulk reaction parameters within one order of magnitude. However, more parameters were weakly identifiable as only three DNA concentrations were considered for model fitting in the synthetic cell populations. Profile likelihood analysis in Figure 3.13 showed that the parameters K_l and τ_d were weakly identifiable, and b is non-identifiable. It is important to note that the parameter estimation was performed on only one batch of experiments. This was done to avoid the known batch-to-batch variability in CFES. However, we also observed that different batches of synthetic cell populations prepared on different days can result in different endpoint protein concentrations and maximum translation rates with a batch-wise CV of 0.10 and 0.16, respectively ((Figure 3.14)). This is comparable to previously reported batch CV values of expressed eGFP or RFP in bulk PURE systems at 0.05-0.2 [42, 145].

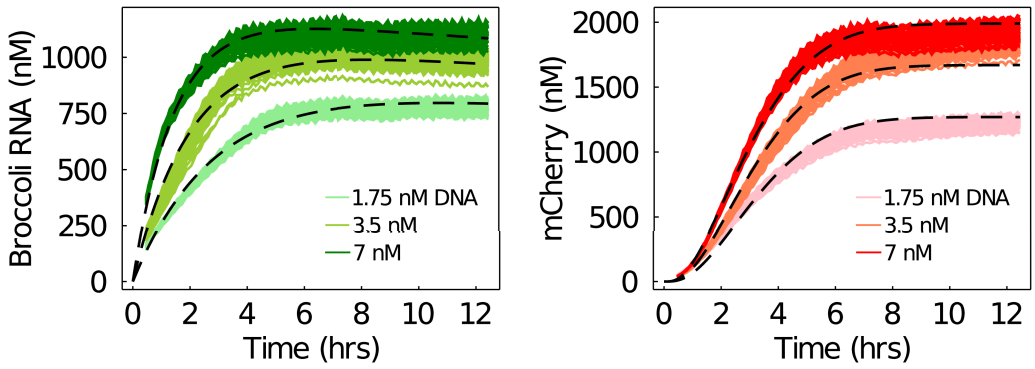

Figure 3.12: CFE model fit on gene expression dynamics in synthetic cell populations. Single-cell traces of mRNA and protein expression in three synthetic cell populations with 1.75 nM, 3.5 nM, and 7.0 nM of pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA as in Figure 3.10A are shown with the global model fit to the resource-limited CFES model in Equation 3.8-3.13 in black dashed lines.

To determine whether the composition of the outer solution would significantly affect gene expression in the microfluidic-generated synthetic cells, synthetic cell populations were prepared with different dilutions of the outer solution composed of PURExpress

Table 3.6: Parameter estimates ($\hat{\boldsymbol{\theta}}_{\boldsymbol{pop}}$) and likelihood-based 95% confidence intervals (CI) from synthetic cell populations. Parameters with CIs at -∞ and/or ∞ are non/weakly-identifiable within one order of magnitude from $\hat{\boldsymbol{\theta}}_{\boldsymbol{pop}}$.

Parameter	Description	$\hat{\boldsymbol{\theta}}_{\boldsymbol{pop}}$	95% CI	Units
k_r	RNA transcription rate	1899	1631–3537	nM/hr
K_r	Dissociation constant between RNAP and DNA	8.86	6.97–18.66	nM
δ_r	RNA degradation rate	0.0081	0.00239–0.0143	1/hr
k_p	Protein translation rate	1954	1617–2696	nM/hr
K_p	Dissociation constant between ribosome and RNA	1319	819–2038	nM
k_{mat}	mCherry maturation rate	2.15	($\pm$0.12)	1/hr
δ_{TsR}	TsR degradation rate	0.154	0.136–0.175	1/hr
δ_{TlR}	TlR degradation rate	0.244	0.184–0.684	1/hr
K_l	Michaelis-Menten constant for TlR degradation	0.232	-∞–0.713	-
a	Scaling factor for consumption of TsR with transcription	6.60E-4	6.21E-4–6.74E-4	-
b	Scaling factor for consumption of TlR with translation	4.46E-13	-∞–∞	-
τ_d	Time-delay for protein translation	0.0576	-∞–0.279	hr
σ_r	Stdev. of Broccoli RNA measurements	23.48	21.83–28.25	nM
σ_p	Stdev. of mCherry protein measurements	71.58	66.4–86.1	nM
τ_l	Time lag between reaction start and data collection	0.457	0.342–0.535	hr

CFES without DNA. An inner solution was prepared according to Table 3.1 with 3 nM of pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA. The outer solution was also composed of PURExpress CFES, with glucose, pluronic acid, and no DNA (Table 3.3). Aliquots of the liposomes were diluted with the dilution mix to result in 100, 66, 50, 33, 10% PURExpress CFES outer solutions. Despite the inner, outer, and dilution solutions being osmotically balanced, differences in composition between diluted outer solutions and the inner solutions resulted in material flux into the liposomes. This is because osmolarities are indirectly measured and balanced using freezing point depression measurements. As a result, comparing complex solutions of different compositions will have deviations from the actual osmolarities [251]. This results in liposomes swelling or bursting in the diluted outer solution. Diluted outer solutions resulted in lower expression of RNA and protein in the synthetic cell populations, which shows that the composition of the outer solution influences the dynamics of encapsulated CFES reaction (Figure 3.15). The higher expression in the undiluted outer solution agrees with previous experiments showing that a feeding outer solution can improve gene expression in liposome encapsulated CFES [192, 36].

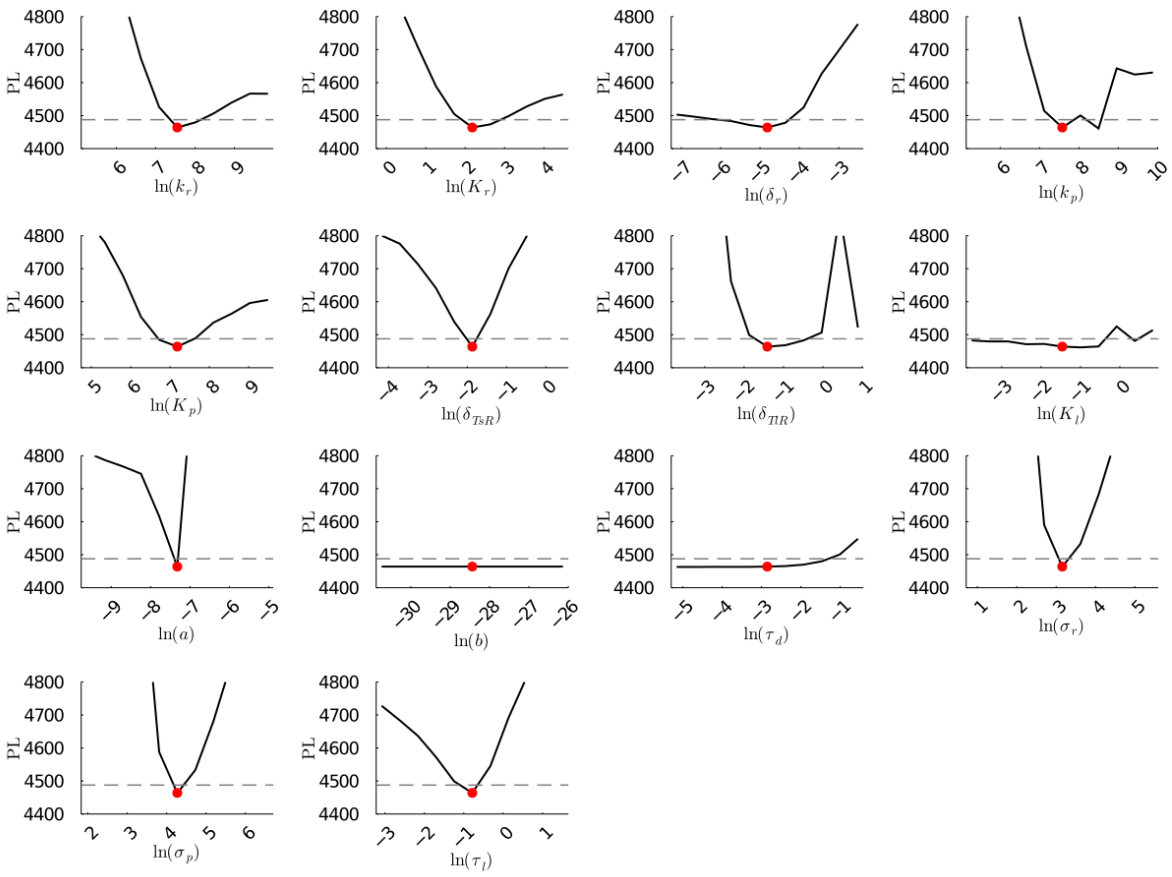

Figure 3.13: Profile likelihoods of the synthetic cell populations. Each plot corresponds to a parameter in the model and additional fitting parameters (σ_r and σ_p). The y-axis of each plot is the negative log likelihood of the model given the corresponding parameter value in the x-axis with all other parameters reoptimized (Equation 3.4). The red dot shows the optimized parameter set with the minimum negative log likelihood. The dashed grey line is the 95% significance threshold line from (Equation 3.5). The intersections of the significance threshold line and profile likelihood are the likelihood-based confidence intervals of the optimized parameter.

3.4 Conclusions

In summary, our study tested different variations of a coarse-grained model of CFES reactions using simultaneously quantified RNA and protein dynamics with likelihood-based methods for model selection and parameter identification. By using a coarse-grained model, gene expression parameters were estimated without knowledge of the full composition of the CFES. This is particularly useful for crude extract systems or proprietary CFES such as the NEB PURExpress CFES that we used in this study. Several models have been developed to include more details of CFE reactions such as initiation and elongation factors [59] or multiple translating ribosomes on an mRNA template [165]. These models provide a more detailed interpretation of the data, but also require either additional information on the time-varying states of these gene expression factors or additional unknown parameters that can result in over-parameterized models and non-identifiability. In the present study, a coarse-grained model of transcription and translation was able recapitulate the full gene expression dynamics across

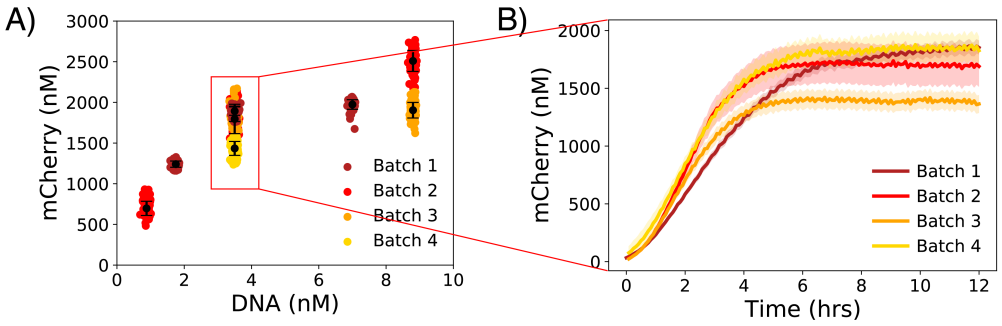

Figure 3.14: Protein expression of batches of synthetic cell populations. (A) Endpoint concentrations of synthetic cell populations at different pEXP5-NT/6xHis mCherry F30 2xd-Broccoli plasmid DNA concentrations from different batches. Batches are performed on a different day with a new set of inner CFES, feeding outer buffer, and lipid phase solutions. Black dots with error bars are mean and standard deviation of the populations. (B) Dynamics of protein expression in batches of synthetic cell populations with 3.5 nM of pEXP5-NT/6xHis mCherry F30 2xd-Broccoli plasmid DNA. Lines and shaded areas show the mean and standard deviation of each synthetic cell population.

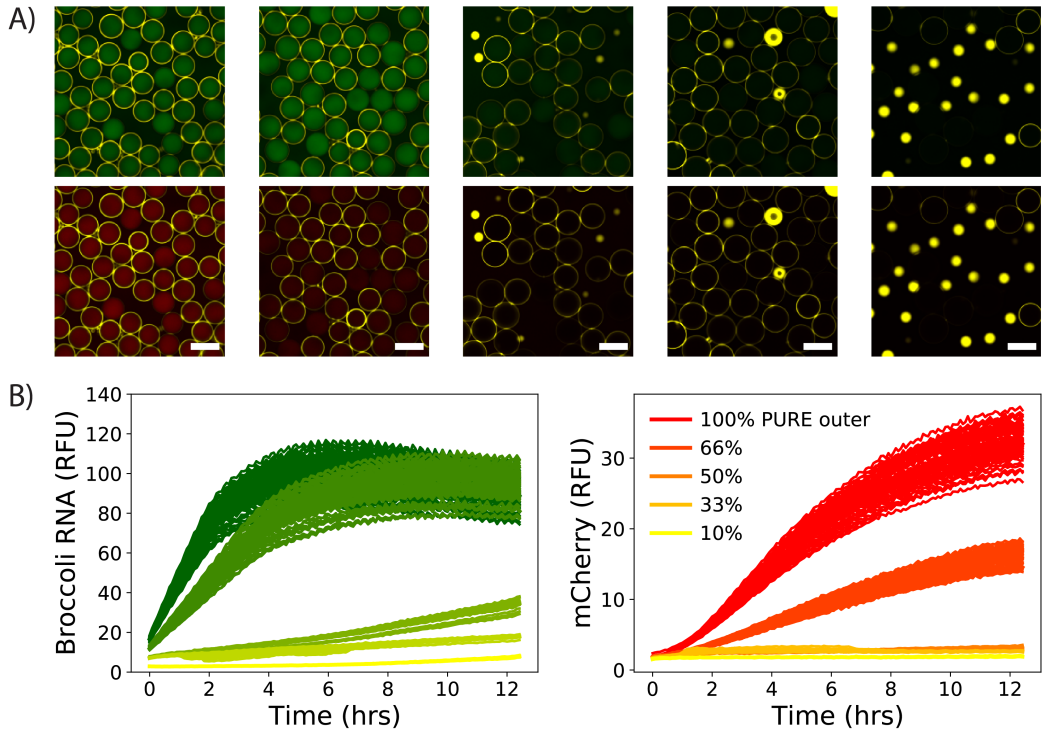

Figure 3.15: Effects of outer buffer dilution in synthetic cell populations. (A) Confocal images of synthetic cell populations with 3 nM pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA in decreasing percentage of PURExpress CFES outer solution (100, 66, 50, 33, and 10 % from left to right) at endpoint after 30 °C incubation for 12.5 hours. Top row is for mRNA Broccoli channel (green) and bottom row is for mCherry protein (red) both overlaid with the lipid DiD dye channel (yellow). (B) Single-cell gene expression dynamics of expressed RNA and protein of the same synthetic cell populations. Total number of cells analyzed are 66, 69, 17, 16, and 17 for 100, 66, 50, 33, and 10% PURExpress outer solution populations, respectively. Scale bars are all 100 μm.

DNA and RNA titration experiments. While we have focused on a simple constitutively expressed gene in this model, it can be readily extended to more complex gene circuitry, CFES characteristics, and protein maturation properties. We then showed that large populations of highly monodisperse synthetic cells can be reproducibly generated using double-emulsion microfluidics. Gene expression in these synthetic cells are also uniform and deterministic. Using our methodologies, we demonstrated that bulk and encapsulated CFES reactions result in different gene expression dynamics. These differences are attributed to the semi-permeable lipid membrane, which allows the exchange of ions and water that alters the internal composition of the synthetic cells. This emphasizes the importance of the physical environment to compartmentalized biochemical reactions. Our demonstrated high degree of control over synthetic cell production, as well as relative ease of analysis compared to synthetic cells with high variability generated by bulk encapsulation methods, will be critical for bottom-up synthetic biology to build synthetic multicellular systems.

Chapter 4

Variability and communication in synthetic cell populations

4.1 Background

Biological multicellular systems are subject to many sources of noise and variation. Even isogenic populations of cells will unavoidably exhibit variability and heterogeneity [162]. For instance, protein expression in bacterial *E. coli* cells can fluctuate from cell to cell in culture. This can be a result of stochastic biochemical reactions due to low copy numbers of reactants (intrinsic noise) or fluctuations of states, concentrations, and locations of molecules that provide differences in cell activities (extrinsic noise) [71]. In populations of single cell organisms, variability can be highly advantageous through bet-hedging. Unlike adaptive strategies where cells sense and respond to the environment, a standing diverse pool of cell types can ensure survival of some cells against sudden stress factors [64]. Variability in cell populations can also provide fractional or graded responses to stimuli. For instance, adipocytes maintain tissue size with a low rate of renewal without de-differentiation due to population heterogeneity stemming from protein expression noise in combination with multiple positive feedback regulation [2]. Intercellular variability can also increase the amount of information transfer by summing the cell-to-cell variable dose responses as demonstrated in multiple cultured myotubes [276]. In other cases, variability can be undesired and biological systems have strategies to reduce population variability. For example, *Myxococcus xanthus* bacteria can rescue non-motile mutant cells within the population by outer membrane exchange (OEM). This process passes on outer membrane motility proteins from healthy strains to the mutant strains to restore motility [194]. Intercellular communication can also help coordinate a population-wide response and reduce population variability. Quorum sensing cells secrete and sense a small signalling autoinducer molecule which diffuses around the external environment for cells in the population to detect. This allows the cells to sense population density and then respond collectively. Quorum sensing communication, first described in the bioluminescent marine bacterium *Vibrio fischeri* as the Lux quorum sensing system, has been suggested to reduce population variability by diffusional dissipation of the external signalling molecule [256]. Coupled with a positive feedback gene circuit, it can provide a switch-like behavior that triggers the entire population when the AHL autoinducer is past a critical concentration. This positive feedback configuration is found in the bioluminescent marine bacterium *V. fischeri*

[281] (Figure 4.1A). The quorum sensing system is also known to be coupled to negative feedback regulation to repress biofilm formation in *Vibrio cholerae*, the causative agent in the diarrheal disease cholera. At low cell densities and low autoinducer concentrations, the autoinducer receptors function as kinases and trigger downstream expression of genes for biofilm formation. At high autoinducer concentrations, the autoinducer receptors function as phosphatases to stop biofilm formation and potentially promote dispersal of *V. cholerae* [282]. Engineering the quorum sensing system with negative feedback regulation has also been demonstrated in *E. coli* bacteria to reduce population variability in cell cultures [27].

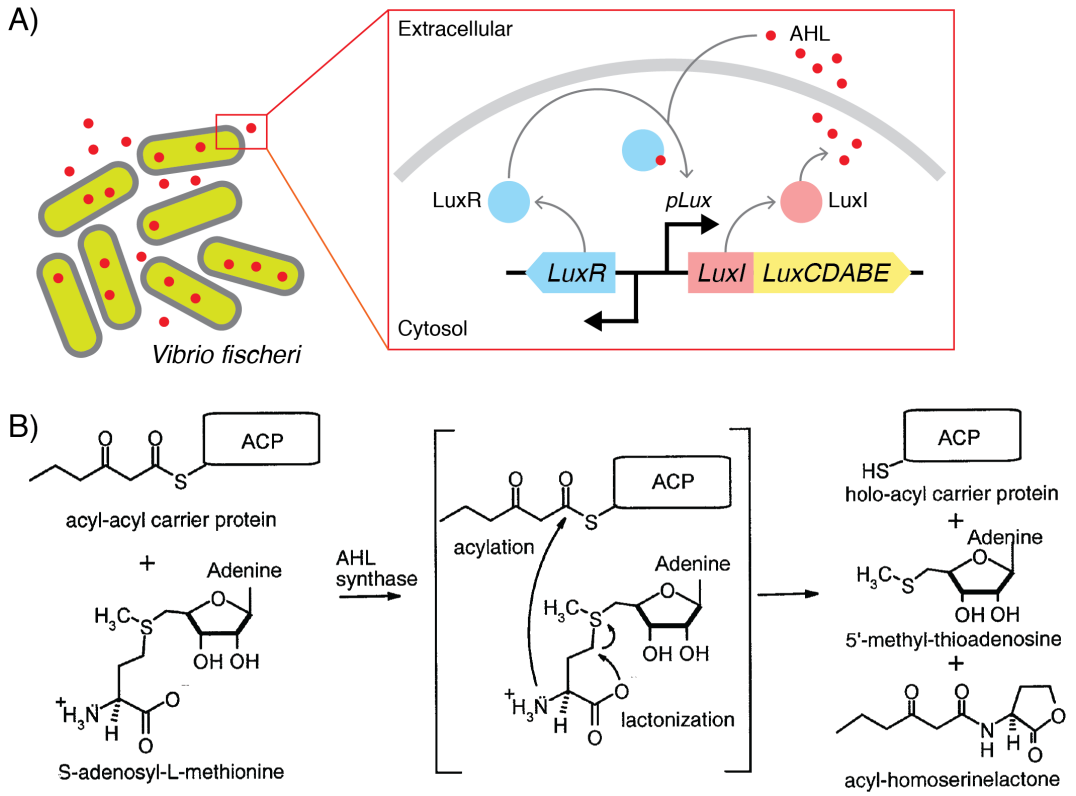

Figure 4.1: Quorum sensing system in *Vibrio fischeri*. (A) *V. fischeri* bacteria secrete a small and membrane-diffusible AHL molecule, N-(3-Oxohexanoyl)-L-homoserine lactone (3OC6-HSL), into the environment. AHL concentrations increase as the population density of the bacteria increases. At a critical concentration of AHL, the *lux* operon is induced to express genes for bioluminescence. Inset: Positive feedback of the Lux quorum sensing system. LuxR and LuxI control downstream expression of the *lux* operon. LuxI is the acyl-homoserine lactone (AHL) synthase that produces AHL. LuxR is a cytoplasmic transcription factor that complexes with AHL, binds with the *lux* promoter (*pLux*), and activates gene expression in the *lux* operon. The *lux* operon also includes the *luxI* gene to further express LuxI and produce AHL. (B) Synthesis of 3OC6-HSL by LuxI AHL synthase from its precursor substrates 3-Oxohexanoyl-acyl carrier protein (ACP) and S-adenosyl-L-methionine (SAM). Image from [283]. Reproduced with permission. © 2002 Elsevier.

In this chapter, we aim to incorporate variability and quorum sensing communication in synthetic cell populations. By providing these two characteristics in synthetic cell populations, the effect of intercellular communication on population variability can

be investigated in a minimal *in vitro* context. Different quorum sensing gene circuits have already been shown to function in bulk CFES reactions. For example, the Lux-type quorum transcription factors (LuxR, LasR, RpaR) can induce transcription of their cognate promoters (pLux, pLas, pRpa) with the addition of their cognate AHLs (N-(3-oxohexanoyl)-L-homoserine lactone (3OC6-HSL), N-3-oxododecanoyl homoserine lactone (3OC12-HSL), p-coumaroyl-HSL) in an *E. coli* extract-based CFES [104]. Cell-free expressed AHL synthases LuxI and RhlI have also been shown to be capable of synthesizing 3OC6-HSL and C4-HSL, respectively, in CFES. However, addition of the AHL precursors, 3-oxohexanoyl-acyl carrier protein (ACP) or butyryl-ACP and S-adenosyl-L-methionine (SAM), are required for the AHL synthase LuxI to produce AHL in PURE CFES systems [148, 206] (Figure 4.1B). In contrast, LuxI in *E. coli* extract-based CFES has been demonstrated to produce the 3OC6 AHL *in situ* to induce transcription of the LuxR and pLux quorum sensing system [55]. This indicates that the AHL precursors are readily present in the lysates. Quorum sensing systems in CFES have also been encapsulated into liposomes by the inverse emulsion phase transfer method and made to interact with quorum sensing bacterial cells [148, 206]. However, a population of only synthetic cells encapsulating the cell-free quorum sensing systems in liposomes and interacting among each other has not yet been demonstrated. Here, we present our current progress in building a Lux quorum sensing synthetic cell population by encapsulating *E. coli* extract-based CFES into liposomes. First, we produced *E. coli* extract-based CFES and tested the functionality of our quorum sensing gene circuits in bulk CFES reactions. These were then encapsulated into liposomes using the inverse emulsion phase transfer method. To incorporate variability in the synthetic cell populations, we utilize a two-inlet double-emulsion microfluidic design from [225] to generate monodisperse liposomes with high encapsulated DNA concentration variability. This was demonstrated by encapsulating a PURExpress CFES with varying concentrations of pEXP5-NT/6xHis mCherry F30-2xdBroccoli DNA plasmid. In the near future, we plan to combine the two results of quorum sensing in CFES and double-emulsion microfluidics to generate heterogeneous communicating synthetic cell populations.

4.2 Methodology

4.2.1 Preparation of *E. coli* extract-based CFES

An *E. coli* extract-based CFES was prepared using a modified protocol based on the work of Levin *et al.* (2019) [149]. The CFES is composed of the extract, solution A, solution B, and the DNA template for gene expression. To prepare the extract, an *E. coli* BL21 (DE3) starter culture was inoculated in 200 mL LB media and incubated overnight at 37 °C with 180 rpm shaking. This was then used to inoculate the production culture of 500 mL 2xYTP media at a starting OD600 of 0.01. The production culture was incubated at 37 °C with 180 rpm shaking until an OD600 of 1.6 (approx. 2-3 hours). The cells were pelleted by centrifugation for 10 mins at 5000 xg using an Avanti Centrifuge J26-XP with a JLA-8.1000 rotor (Beckman Coulter, USA), transferred to a pre-weighed 50 mL tube, and washed and pelleted three times with 30 mL S30A buffer (14 mM MgGlu, 60 mM KGlu, 50 mM Tris, and 2 mM DTT, titrated with glacial acetic acid to pH 7.7). The washed pellet is weighed (∼2.5 g of wet pellet/500 mL culture),

then flash frozen with liquid nitrogen, and stored at -80 °C overnight. The next day, 1 mL S30A buffer per 1 g wet pellet weight was added to the frozen pellet to thaw and resuspend by vortexing. The resuspension is transferred to 1.5 mL aliquots in 2 mL microcentrifuge tubes and then sonicated on an ice-water bath to prevent overheating of the sample. Sonication was done by 10 cycles of 10 s pulse and 30 s rest at 25% amplitude (Branson Digital Sonifier Model 450-D with a 418-A probe). The resulting lysate should turn a darker brown compared with the pre-sonicated resuspension. Immediately after sonication, 2 μL of 1 M DTT is added into the 2 mL microcentrifuge tube. The sonicated lysates are then centrifuged at 18000 xg for 10 mins. The clear supernatant was pooled and collected into a 10 mL tube and incubated at 37 °C with 250 rpm with the cap open for the 1 hour run-off reaction. After the run-off reaction, the lysate was clarified by centrifugation at 10000 xg for 10 mins to remove the precipitates. The clarified lysate was then dialyzed in S30B buffer (14 mM MgGlu, 60 mM KGlu, 5 mM Tris, and 2 mM DTT, titrated with 2 M Tris to pH 8.2) for 3 hours at 4 °C using a Spectra/Por 2 Dialysis Membrane with a 12-14 kDa molecular weight cut-off (Repligen, USA). After dialysis, the lysate is clarified by centrifugation at 10000 xg for 10 mins, split into 50 μL aliquots, flash frozen with liquid nitrogen, and then stored at -80 °C until use. Solution A is composed of 12.4 mM ATP (Sigma, USA), 8.7 mM GTP (Roche, Switzerland), 8.7 mM CTP (Sigma, USA), 8.7 mM UTP (Sigma, USA), 0.68 mM folinic acid (Sigma, USA), 0.176 mg/mL tRNA (Roche, Switzerland), 2.7 mM β-nicotinamide adenine dinucleotide (Sigma, USA), 1.8 mM Coenzyme A (CoA), 27.2 mM oxalic acid (Roth, Germany), 6.8 mM putrescine (Sigma, USA), 10.1 mM spermidine (Sigma, USA), and 774.5 mM HEPES (Roth, Germany). Solution B is composed of 14.3 mM of each of the twenty amino acids (Sigma, USA), 236.3 mM phosphoenolpyruvate (PEP), 71.6 mM magnesium glutamate (MgGlu) (Sigma, USA), and 970.8 mM potassium glutamate (KGlu) (Sigma, USA). Further details of all the chemicals used are listed in Appendix A. The final *E. coli* extract-based CFES master mix is prepared according to Table 4.1. This CFES mix is used as bulk reactions or encapsulated by double-emulsion microfluidics and incubated at 30 °C for 2-5 hours.

Table 4.1: Standard *E. coli* extract-based CFES master mix.

Component	**Volume (μL)**
Solution A	1.8
Solution B	1.75
MgGlu (200 mM)	0.125
Extract	5
Plasmid DNA	x
Water	Fill to 12.5
TOTAL	12.5

Expression activity of the homemade *E. coli* extract-based was compared to standard reactions of commercially available PURExpress CFES (NEB, USA) and myTXTL Sigma 70 Master Mix Kit (Arbor Biosciences, USA) (Table 4.2 and 4.3). Positive control plasmids, pEXP5-NT/6xHis eGFP and p70a(2) deGFP, were used to test constitutive T7 RNAP and σ_{70}-mediated transcriptional activity of the extracts, respectively (Figure 4.2A). The plasmid pEXP5-NT/6xHis eGFP was kindly provided by J.L. Ross

Anderson, University of Bristol and p70a(2) deGFP is the positive control plasmid from the myTXTL Sigma 70 Master Mix Kit. The plasmid p70a(2) deGFP expresses a truncated form of eGFP that is optimized for CFES and has similar fluorescence properties of eGFP [230]. All plasmid DNA templates were prepared and purified by ethanol precipitation using the QIAGEN Plasmid Maxi Kit (QIAGEN, Germany) and then dissolved in nuclease-free water. The bulk CFES reactions were prepared in 384-well plates (Greiner Bio-One, Austria), sealed with a clear plate film (Thermo, USA), and incubated in a Genios Pro plate reader (TECAN, Switzerland) at 30 °C. Fluorescence measurements were taken for each sample at 10-minute intervals for 2-6 hours. Excitation/emission filter wavelengths for eGFP or deGFP detection were 485/535 nm through a 320-500 nm reflection and 520-800 nm transmission dichroic mirror with a gain of 25, averaged over 10 reads, and an integration time of 40 μs.

Table 4.2: NEB PURExpress CFES master mix.

Component	**Volume (μL)**
Solution A	5
Solution B	3.75
RNAseA inhibitor	0.25
Sucrose (1.5M)	0.67
Plasmid DNA	x
Water	Fill to 12.5
TOTAL	12.5

Table 4.3: Arbor Biosciences myTXTL Sigma70 CFES master mix. T7 RNAP (NEB, USA) is added to the master mix only for T7 RNAP-mediated transcription of the pEXP5-NT/6xHis eGFP plasmid.

Component	**Volume (μL)**
Sigma70 MM	5
T7 RNAP (50 units/μL)	0.25
Plasmid DNA	x
Water	Fill to 12.0
TOTAL	12.0

4.2.2 Quorum sensing plasmid constructs

The quorum sensing system from *Vibrio fischeri* was used to create a genetic circuits that allow AHL diffusion-mediated communication between synthetic cells. Separate plasmids expressing LuxI, LuxR, and eGFP proteins under transcriptional control of either T7 RNA polymerase (T7 RNAP) or the *lux* promoter (pLux) were constructed to test the functionality of the different parts of the quorum sensing gene circuit. Specifically, the plasmids pEXP5-NT/6xHis LuxRopt, pEXP5-NT/6xHis LuxIopt, pEXP5-NT/pLux 6xHis eGFP, and pEXP5-NT/pLux 6xHis LuxIopt were constructed under the same high-copy ampicillin resistance pEXP5-NT vector backbone (Invitrogen, USA) (Figure 4.2B). Plasmids pEXP5-NT/6xHis LuxRopt and pEXP5-NT/6xHis LuxIopt have a constitutive T7 RNA polymerase-mediated promoter and

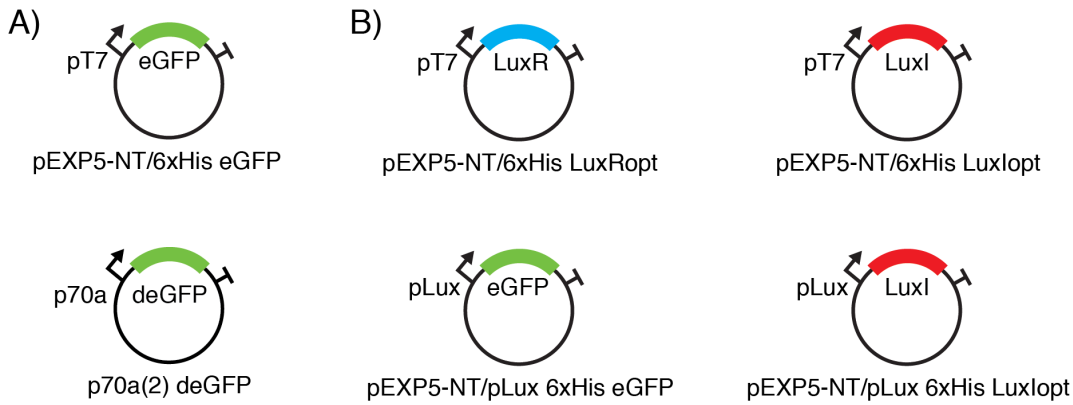

Figure 4.2: Quorum sensing gene cicruits for intercellular communication. (A) Positive control plasmids pEXP5-NT/6xHis eGFP and p70a(2) deGFP for constitutive T7 RNAP and σ_{70}-mediated transcription, respectively. (B) Quorum sensing plasmids pEXP5-NT/6xHis LuxRopt, pEXP5-NT/6xHis LuxIopt, pEXP5-NT/pLux 6xHis eGFP, and pEXP5-NT/pLux 6xHis LuxIopt. Promoters are annotated with either pT7, pLux, or p70a for T7 RNAP constitutive, LuxR-AHL inducible, or σ_{70} constitutive promoters, respectively.

strong ribosomal binding site (RBS) for expression of LuxR and LuxI, respectively. Plasmids pEXP5-NT/pLux 6xHis eGFP and pEXP5-NT/pLux 6xHis LuxIopt are controlled by the *lux* promoter that is induced by the LuxR-AHL transcription factor complex and strong RBS. The genes *luxR* and *luxI* were codon optimized for *E. coli* K12 expression with a relative threshold for rare codons of 0.3 in Geneious (v11.0.2, www.geneious.com) to avoid rare codon usage. The quorum sensing plasmids were first tested in bulk *E. coli* extract-based CFES reactions by preparing a master mix according to Table 4.1, adding different combinations and concentrations of the plasmids pEXP5-NT/6xHis LuxRopt, pEXP5-NT/6xHis LuxIopt, pEXP5-NT/pLux 6xHis eGFP, and pEXP5-NT/pLux 6xHis LuxIopt, as well as N-(3-oxohexanoyl)-L-homoserine lactone (3OC6-HSL) or also called N-(β-ketocaproyl)-L-homoserine lactone (Sigma, USA), and monitoring for eGFP expression in a plate reader over time at 30 °C as described in Section 4.2.1. Plasmid construction protocols and sequences are further described in Appendix B. All plasmids were sequence confirmed by Sanger sequencing.

4.2.3 Encapsulated quorum sensing CFES reactions

Combinations of the quorum sensing plasmids were used to make up the three different quorum sensing gene circuits: a sender circuit, receiver circuit, and positive feedback circuit (Figure 4.3). The *sender gene circuit* is composed of the pEXP5-NT/6xHis LuxIopt, which constitutively expresses the AHL synthase LuxI. The LuxI enzyme is able to produce AHL using precursors available in the *E. coli* extract to induce the receiver circuit. The *receiver gene circuit* is composed of both pEXP5-NT/6xHis LuxRopt and pEXP5-NT/pLux 6xHis eGFP plasmids. Addition of AHL that is externally added or produced by the sender circuit to the receiver gene circuit will result in the formation of the LuxR-AHL complex and induction of eGFP expression. The *positive feedback gene circuit* is composed of the pEXP5-NT/6xHis LuxRopt, pEXP5-NT/pLux 6xHis LuxIopt, and pEXP5-NT/pLux 6xHis eGFP plasmids. In the positive feedback circuit, addition of AHL will result in the formation of the LuxR-AHL complex and induction

of LuxI and eGFP expression. The LuxI enzyme then produces more AHL to further induce itself in a positive feedback. *E. coli* extract-based CFES are prepared according to Table 4.1 with the different quorum sensing gene circuits.

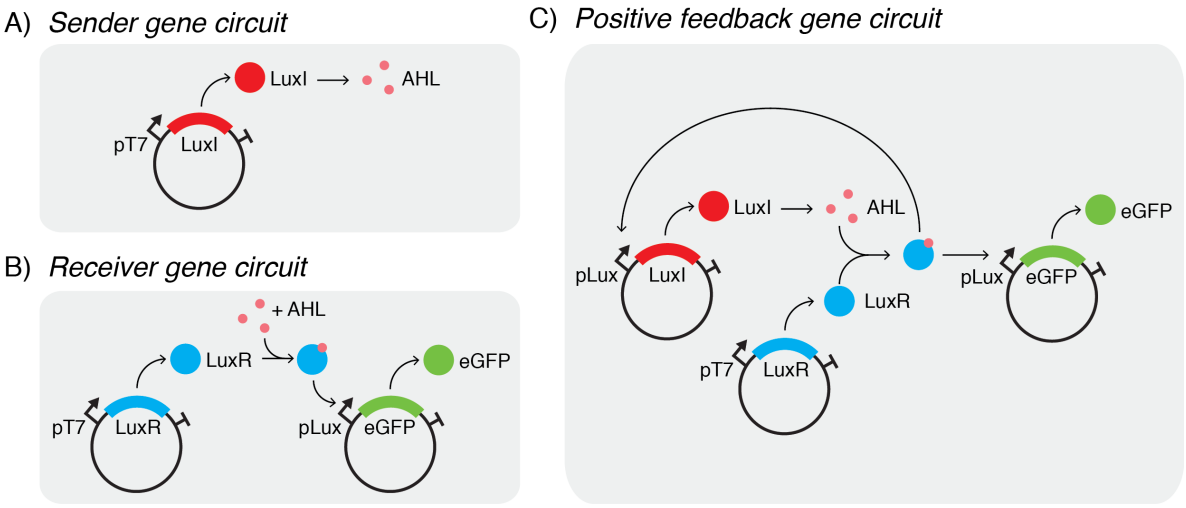

Figure 4.3: Quorum sensing gene cicruits for intercellular communication. (A) The sender gene circuit constitutively expresses LuxI and produces the AHL signal. (B) The receiver gene circuit constitutively expresses LuxR and can induce the plasmid pEXP5-NT/pLux 6xHis eGFP with the addition of AHL. (C) The positive feedback gene circuit also constitutively expresses LuxR, and both *luxI* and *egfp* genes are regulated by the *lux* promoter. The addition of AHL will result in a positive feedback where LuxI and eGFP are expressed and produces more AHL.

Inverse emulsion phase transfer

Encapsulation was first done by inverse emulsion phase transfer as described in Section 3.2.3, but using the *E. coli* extract-based CFES inner and outer feeding buffer solution prepared according to Table 4.4 with their final osmolarities balanced. Synthetic cells encapsulating the receiver gene circuit (0.5 nM pEXP5-NT/6xHis LuxRopt + 10 nM for pEXP5-NT/pLux 6xHis eGFP) was tested by externally adding 200 nM AHL in the outer solution after inverse emulsion phase transfer. Synthetic cells encapsulating the sender gene circuit (0.5 nM pEXP5-NT/6xHis LuxIopt) were tested by combining them with receiver synthetic cells to determine if they are able to cause induction of gene expression in the receiver cells. Sender cells and receiver cells were made by successive inverse emulsion phase transfer of receiver and sender gene circuits in the *E. coli* extract-based CFES as shown in Figure 4.4, where the receiver synthetic cells in solution is used as the outer buffer for phase transfer inverse emulsion of sender synthetic cells. Texas Red DHPE or DiD lipid dyes are used for the receiver and sender cells, respectively, to differentiate the two cell types. After the quorum sensing synthetic cells were generated and induced, samples were incubated at 30 °C for 4 hours and then imaged under confocal microscopy. Confocal microscopy was done with an LSM 880 with Airyscan and C-Apochromat 40x/1.2 W autocorr M27 objective. Excitation/detection wavelengths are 488 nm/505-515 nm for eGFP protein, 594 nm/605-615

nm for Texas Red DHPE, and 633/640-720 nm for DiD dye.

Table 4.4: Inner and outer solutions for *E. coli* extract-based CFES synthetic cells. Volumes are given for one bulk reaction volume (1X). For generating a synthetic cell population, 9X and 45X volumes are prepared for the inner and outer solutions, respectively. Pluronic acid is added only for microfluidic-generated synthetic cells and not for inverse emulsion phase transfer-generated synthetic cells.

Component (μL)	**Inner**	**Outer**
Solution A	1.8	1.8
Solution B	1.75	1.75
200 mM MgGlu	0.125	0.125
Extract	5	0
S30B	0	5
Plasmid DNA	x	0
10% Pluronic acid	0	(0.3125)
Water	Fill to 12.5	Fill to 12.5
TOTAL	12.5	12.5
Osmolarity (mOsmol/kg)	667	667

Microfluidic-generated synthetic cells

Synthetic cells were also generated by microfluidics as described in Section 3.2.3, but using the *E. coli* extract-based CFES inner and outer feeding buffer solution prepared according to Table 4.4 and a lipid-oil solution composed of 6.5 mM of L-α-phosphatidylcholine (Egg PC) phospholipids (Avanti, USA), and 53.3 μM of DiD (Thermo, USA) or 8 μM Texas Red DHPE (Biotium, USA) fluorescent dye. However, unlike the PURExpress CFES, we found that *E. coli* extract-based CFES was not compatible with the octanol oil phase solution. This was observed in microfluidic-generated synthetic cells encapsulating *E. coli* extract-based CFES with a constitutive pEXP5-NT/6xHis eGFP DNA plasmid (10 nM) that did not express any fluorescent eGFP protein. We further confirmed that the oil was responsible for the decrease in gene expression by overlaying bulk 12.5 μL reactions of *E. coli* extract-based CFES with 10 nM pEXP5-NT/6xHis eGFP DNA plasmid with 2 μL of different oil phases (mineral oil, octanol, and oleic acid) on top. Only the CFES with mineral oil overlay expressed eGFP protein, while the other oil phases inhibited eGFP expression. All reactions were incubated for 4 hours at 30 °C before imaging or measuring eGFP fluorescence in the plate reader. Bulk CFES reactions were prepared in 384-well plates (Greiner Bio-One, Austria), sealed with a clear plate film (Thermo, USA), and measured using a Genios Pro plate reader (TECAN, Switzerland). Excitation/emission filter wavelengths for eGFP or deGFP detection were 485/535 nm through a 320-500 nm reflection and 520-800 nm transmission dichroic mirror with a gain of 25, averaged over 10 reads, and an integration time of 40 μs. Synthetic cell populations were imaged using fluorescence widefield microscopy (Andor Axiovert 200M, Zeiss) with a 20x/0.4 LD A-Plan Ph2 objective and Andor Zyla PLUS sCMOS camera. Fluorescence excitation/emission was set at 550nm through a ROX filter set (excitation bandpass 575$\pm$15 nm, beam splitter HC BS 596 nm, emission BP 641$\pm$75 nm) for Texas Red DHPE and

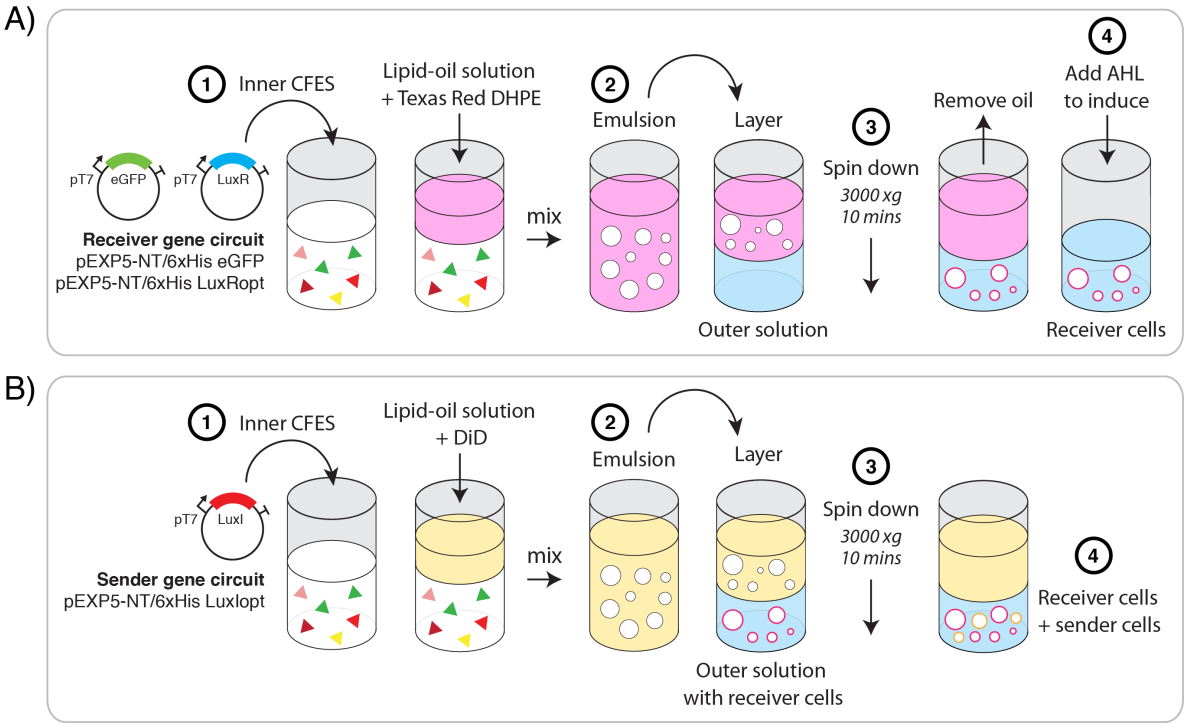

Figure 4.4: Quorum sensing synthetic cells by inverse emulsion phase transfer. (A) Receiver synthetic cells are made by encapsulating 0.5 nM pEXP5-NT/6xHis LuxRopt + 10 nM for pEXP5-NT/pLux 6xHis eGFP in *E. coli* extract-based CFES into liposomes tagged with Texas Red DHPE. The top oil layer is removed before either inducing the synthetic cells with AHL, or using it as the outer solution for generating sender synthetic cells. (B) Sender synthetic cells are made by encapsulating 0.5 nM pEXP5-NT/6xHis LuxIopt in *E. coli* extract-based CFES into liposomes tagged with DiD dye. The outer solution used already contains receiver cells from (A) so that the resulting cell population after phase transfer will contain both receiver and sender cells.

a GFP/Alexa 488/FITC filter set (excitation bandpass 449-489 nm, dichroic longpass 497 nm, emission bandpass 502-549 nm) for eGFP.

4.2.4 Generating heterogeneous synthetic cell populations

Using a two-inlet double-emulsion microfluidic device [225], PURExpress CFES reactions were encapsulated into synthetic cell populations with varying plasmid DNA concentration distributions. This was achieved by preparing the two inner PURExpress CFES solutions according to Table 3.1 with different pEXP5-NT/6xHis mCherry F30-2xdBroccoli DNA plasmid concentrations (1 nM and 10 nM) and adjusting the pressure ratio of the inlets during synthetic cell production (Figure 4.5). The two inner CFES solutions, lipid oil phase, and outer buffer solution (Table 3.2) were dispensed at approximate pressures of 50-70:50-70:80:80 mbar, respectively, to encapsulate the inner solutions into double-emulsions to form liposomes. The lipid oil phase was composed of 1-Octanol (Sigma, USA) with 6.5mM of L-α-phosphatidylcholine (Egg PC) phospholipids (Avanti, USA), and 53.3 μM of 1,1'-Dioctadecyl-3,3,3',3'-Tetramethylindodicarbocyanine, 4-Chloro benzenesulfonate Salt (DiD) fluorescent dye (Invitrogen, USA). Flow regimes and double-emulsion formation in the microfluidic de-

vice were monitored under brightfield microscopy using a Zeiss Andor Axiovert 200M with a 5x/0.15 Plan-Neofluar Ph1 M27 objective and PCO Dimax S4 Monochrome sCMOS high-speed camera. The synthetic cells were collected into a microcentrifuge tube and then placed in channels made from parafilm sandwiched between a microscope slide (76x26x1 mm) and cover slip (24x60 mm). The ends of the channels were sealed with Twinsil Speed silicone (Picodent, Germany) to avoid evaporation. Prepared synthetic cell populations were imaged by confocal laser scanning microscopy using an inverted Zeiss LSM 880 with Airyscan and a 10X/0.45 Plan-Apochromat M27 objective. The samples were maintained at 30 °C. Laser excitation wavelengths were 488 nm, 488 nm, 561 nm, 633 nm for Broccoli RNA, eGFP protein, mCherry protein, and DiD dye, respectively. Emission wavelengths were 499-561 nm, 499-561 nm, 579-641 nm, and 640-720 nm for Broccoli RNA, eGFP protein, mCherry protein, and DiD dye detection, respectively. Images were focused at the equator of the synthetic cells and then acquired every 5 mins for a total of 12 hours. Timelapse images were processed using Fiji (v1.53c) [217] and Python (v3.6) with Scikit-image [268]. Further details for calibration, microfluidic chip fabrication and pretreatment, and image analysis are available in Appendix E, F, and G, respectively.

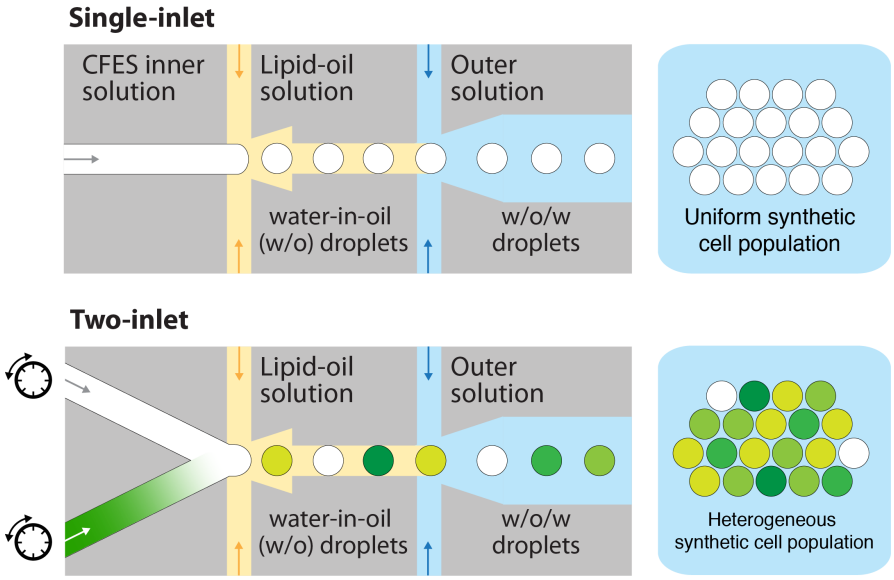

Figure 4.5: Single and two-inlet double emulsion microfluidics. The single-inlet microfluidic device (top) results in the production of monodisperse synthetic cells with uniform concentrations of encapsulated contents. The two-inlet microfluidic device (bottom) can generate monodisperse synthetic cells with varying encapsulated contents by providing inlet solutions with different compositions and adjusting pressure ratios between the two inlets during synthetic cell production.

4.3 Results

4.3.1 Testing the quorum sensing in bulk CFES

First, the expression activity of the homemade *E. coli* extract-based CFES was compared with standard reactions of commercially available CFES. Both T7 RNAP and

native σ_{70}-mediated transcription were tested using the pEXP5-NT/6xHis eGFP and p70(2) deGFP plasmids, respectively. T7 RNAP-mediated transcription is greated in our homemade *E. coli* extract-based CFES as compared to the PURExpress CFES and myTXTL Sigma70 CFES supplemented with T7 RNAP. However, σ_{70}-mediated transcription in the homemade extract is lower as compared to the activity of the myTXTL Sigma70 CFES using the p70(2) deGFP plasmid (Figure 4.6). Nevertheless, this demonstrates that both T7 RNAP and σ_{70}-mediated transcription are active and sufficient for our quorum sensing plasmids with either T7 RNAP (pT7) or σ_{70}-mediated *lux* (pLux) promoters. Next, bulk *E. coli* extract-based CFES reactions were used to test if the quorum sensing gene circuits function as expected. The receiver gene circuit (pEXP5-NT/pLux 6xHis eGFP and pEXP5-NT/ 6xHis LuxRopt) in the *E. coli* extract-based CFES shows that addition of increasing amounts of AHL (0-200 nM) results in greater eGFP expression. Alternatively, addition of pEXP5-NT/6xHis LuxIopt plasmid is sufficient to induce the *lux* promoter and express eGFP (Figure 4.7). We also show that as little as 0.5 nM of pEXP5-NT/6xHis LuxIopt or pEXP5-NT/pLux 6xHis LuxIopt plasmid results in maximum eGFP expression of the quorum sensing receiver gene circuit, while higher concentrations result in a decreased eGFP output (Figure 4.8). This can be a result of resource-sharing in the CFES where the different plasmids share in the same pool of expression resources which we have also observed in Figure 3.8E and F of the previous section between eGFP and mCherry expression. Notably, both constitutive pEXP5-NT/6xHis LuxIopt and AHL inducible pEXP5-NT/pLux 6xHis LuxIopt plasmids resulted in roughly the same eGFP expression despite having no external AHL added. This could be a result of the leaky expression of the *lux* promoter. Overall, these results demonstrate that both LuxR protein and AHL are required to induce eGFP expression under the *lux* promoter. In addition, expressed LuxI in our homemade *E. coli* extract-based CFES can produce the AHL inducer for the quorum sensing system.

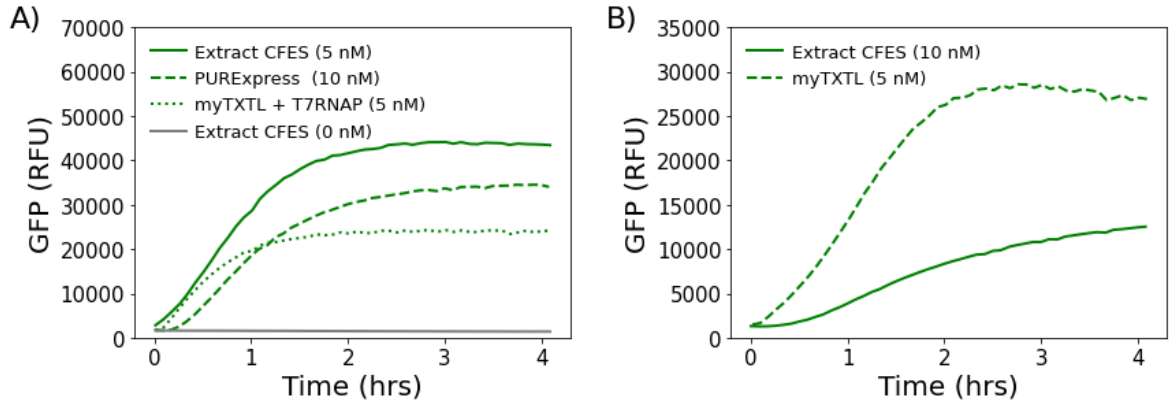

Figure 4.6: Gene expression activity of the *E. coli* extract-based CFES. (A) Comparison of constitutive T7 RNAP-mediated expression in the homemade extract with PURExpress CFES and myTXTL Sigma70 extract-based CFES using the pEXP5-NT/6xHis eGFP plasmid. (B) Comparison of constitutive σ_{70}-mediated expression in the homemade extract-based CFES with myTXTL Sigma70 CFES using the p70(2) deGFP plasmid. To allow T7 RNAP-mediated expression, 0.25 μL of T7 RNAP was added into in the myTXTL Sigma70 CFES master mix. Plasmid concentrations used are indicated in the figure legends.

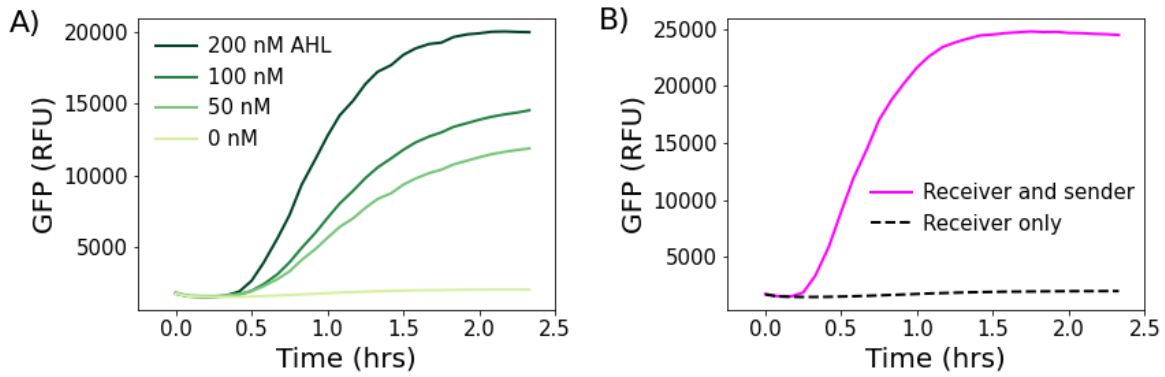

Figure 4.7: Testing quorum sensing plasmids in bulk CFES. (A) *E. coli* extract-based CFES with 20 nM pEXP5-NT/pLux 6xHis eGFP and 2.0 nM pEXP5-NT/6xHis LuxRopt titrated with 50, 100, and 200 nM AHL. (B) *E. coli* extract-based CFES with the receiver gene circuit only and both receiver and sender gene circuits. Plasmid concentrations for the gene circuits are 20 nM pEXP5-NT/pLux 6xHis eGFP and 2.0 nM pEXP5-NT/6xHis LuxRopt for the receiver gene circuit, and 2.0 nM pEXP5-NT/6xHis LuxIopt for the sender gene circuit.

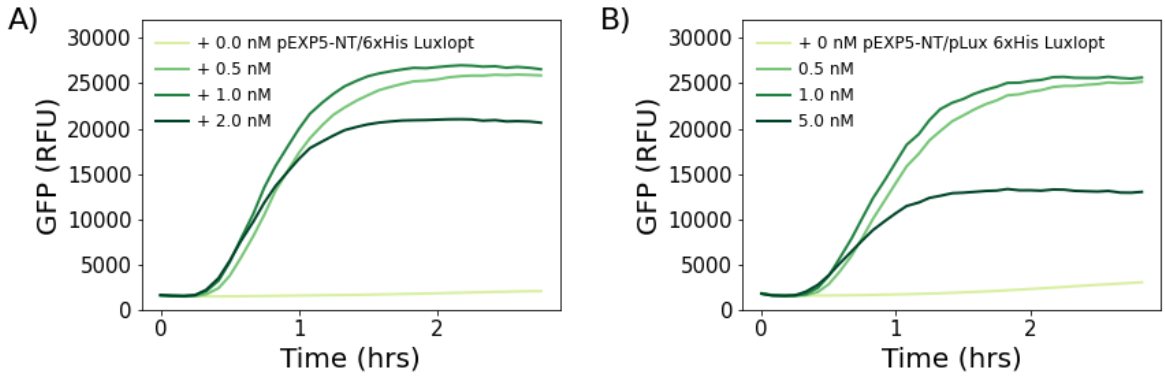

Figure 4.8: Titrating LuxI-producing plasmids in quorum sensing bulk CFES. (A) *E. coli* extract-based CFES with 10 nM pEXP5-NT/pLux 6xHis eGFP and 0.5 nM pEXP5-NT/6xHis LuxRopt supplemented with increasing amounts of pEXP5-NT/6xHis LuxIopt. (B) *E. coli* extract-based CFES with 10 nM pEXP5-NT/pLux 6xHis eGFP and 0.5 nM pEXP5-NT/6xHis LuxRopt supplemented with increasing amounts of pEXP5-NT/pLux 6xHis LuxIopt.

4.3.2 Quorum sensing in synthetic cell populations

We next tested if the quorum sensing gene circuits would function in *E. coli* extract-based CFES encapsulated into liposomes. Using the inverse emulsion phase transfer method, synthetic cells containing either the quorum sensing receiver gene circuit or sender gene circuit were generated. Receiver synthetic cells were inducible by the addition of 100 nM 3OC6-homoserine lactone in the external solution (Figure 4.9A and B). However, sender synthetic cells were unable to induce expression of receiver cells in close proximity (Figure 4.9C). We suspect that this is due to dilution of the AHL signal molecule as the total number and volume of the sender cells is much less than the volume of the external solution. Assuming a population of 1000 sender cells with an average diameter of 20 μm in a 100 μL outer solution, the dilution factor would be in the range of $1/10^5$. Generating quorum sensing synthetic cell populations using our double-emulsion microfluidic method would avoid this dilution problem as synthetic

cells are larger and packed more densely in a low volume channel (up to 7 μL). This was initially attempted, but both externally induced receiver cells and a mixed population of receiver and sender cells did not show any eGFP expression (Figure 4.10). We found that the 1-octanol used for the oil phase was not compatible with the *E. coli* extract-based CFES. This was confirmed by overlaying different oil phases on bulk CFES reactions and monitoring constitutive gene expression of eGFP from the pEXP5-NT/6xHis eGFP plasmid DNA. Mineral oil did not affect cell-free gene expression but 1-octanol and oleic acid inhibited cell-free gene expression. We hypothesize that this is due to the functional groups of the oil phase components affecting the CFES environment, such as oleic acid lowering the pH [253]. We are currently investigating alternative oil phase solutions that are compatible with both the *E. coli* extract-based CFES and double-emulsion microfluidics (*e.g.* squalene, decane, or mixtures of octanol and mineral oil).

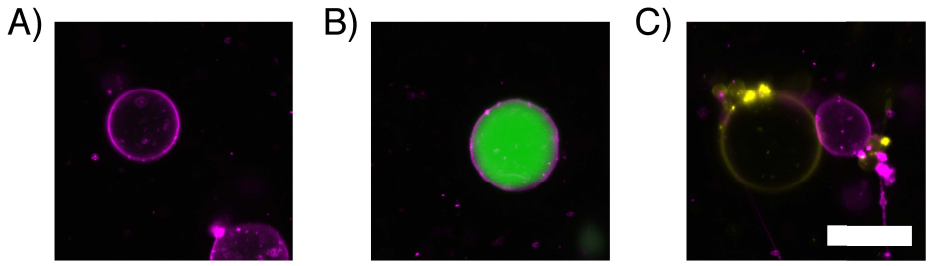

Figure 4.9: Encapsulated quorum sensing gene circuits in *E. coli* extract-based CFES. Confocal images are taken at endpoint after a 4 hour incubation at 30 °C. (A) A receiver synthetic cell without AHL. (B) A receiver synthetic cell with 100 nM AHL (3OC6-HSL) induction to resulting in eGFP expression (green). (C) Receiver (magenta) and sender (yellow) synthetic cells in close proximity did not result in induction of the receiver cell by the sender cell to express eGFP. Liposomes were generated by inverse emulsion method. Scale bar is at 25 μm.

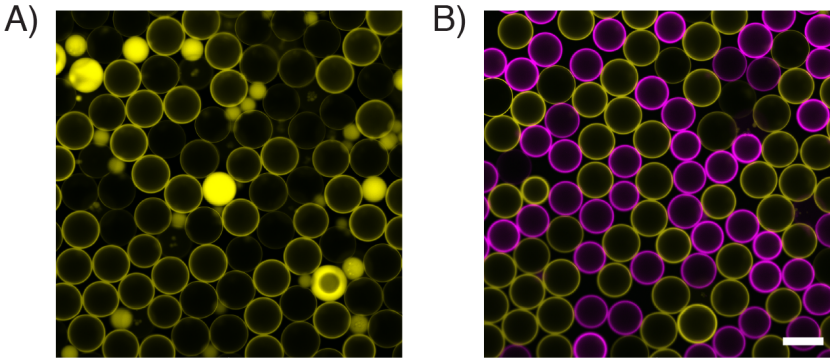

Figure 4.10: Quorum sensing synthetic cells generated by double-emulsion microfluidics. Confocal images are taken at endpoint after a 4 hour incubation at 30 °C. (A) Receiver synthetic cells induced with 100 nM AHL (3OC6-HSL) to induce eGFP expression. (B) Receiver (yellow) and sender (magenta) quorum sensing cells in a mixed synthetic cell population for sender cells to induce the receiver cells. Induction of eGFP expression in both cases was not observed due to oil incompatibility with the CFES. Scale bar is at 100 μm.

4.3.3 Variability & heterogeneity in synthetic cell populations

Heterogeneous synthetic cell populations were generated using a two-inlet microfluidic chip. By varying the pressures of the two inner solutions during production between 50-70 mbar, different ratios of the two inner solutions are mixed into the synthetic cells. Figure 4.11 shows the resulting synthetic cell population from a two-inlet microfluidic device and compares this with a population generated from a single-inlet microfluidic device. The synthetic cell population generated by a single-inlet microfluidic chip had expressed Broccoli RNA and protein coefficient of variation (CV) values of 0.02-0.03 (Table 3.5). The two-inlet microfluidic chip-generated synthetic cell population has a CV of 0.29 and 0.17 for expressed Broccoli RNA and mCherry protein, respectively Figure 4.12. This is despite similar low size variations between them (0.04-0.07 and 0.075 for single and two-inlet populations, respectively), indicating that the increased variability is a result of different concentrations of plasmid DNA encapsulated with the inner solution. A heterogeneous synthetic cell population can also be made by mixing together synthetic cells from different batches. To demonstrate this, synthetic cells from Section 3.3.4 encapsulating different pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid DNA concentrations (1.75, 3.5, and 7.0 nM) were mixed together to form a heterogeneous population (Figure 4.13). This results in a heterogeneous population with distinct subpopulations instead of a single population with a large variance that is produced using the two-inlet microfluidic device.

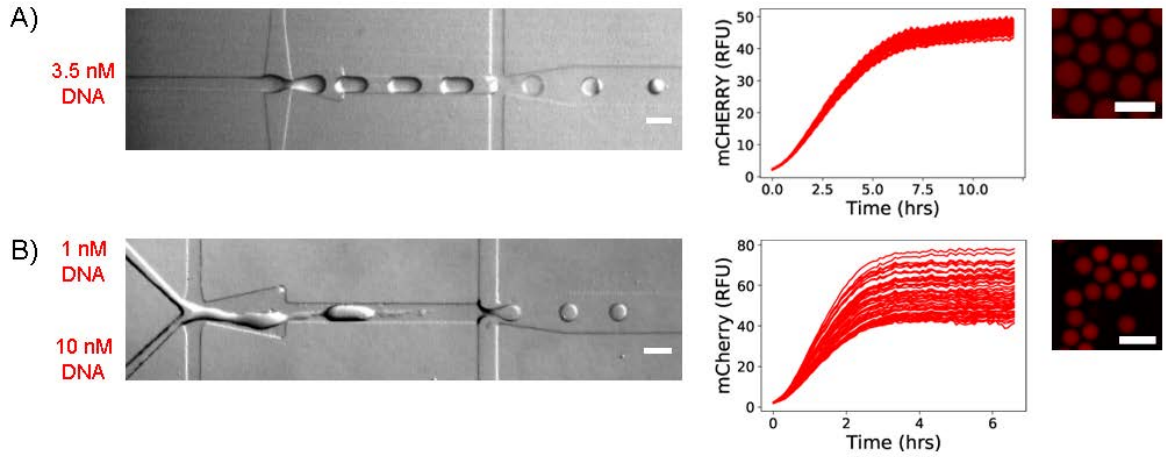

Figure 4.11: Comparison of single inlet vs. two-inlet microfluidics. (A) A single inlet microfluidic device encapsulating inner PURE CFES solution with 3.5 nM pEXP5-NT/6xHis mCherry F30 2xdBroccoli plasmid DNA results in liposomes with low variance of gene expression where endpoint protein RFU CV is 0.03. (B) In contrast, the two-inlet device results in a greater variance of endpoint protein RFU CV at 0.17. The variance of encapsulated DNA was increased by randomly changing the ratio of the two inlet pressures during production. Each inlet had PURE CFES with either 1 nM or 10 nM pEXP5-NT/6xHis mCherry F30 2xdBroccoli plasmid DNA. Representative endpoint images of expressed mCherry for each microfluidic device is shown on the right. Scale bars are all at 100 μm.

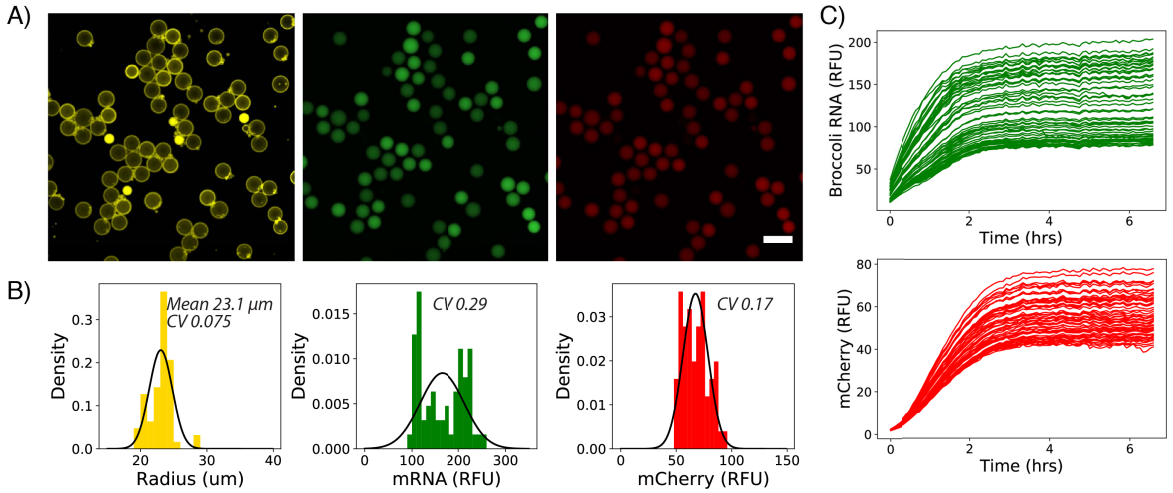

Figure 4.12: Heterogeneous synthetic cell population made by two-inlet microfluidics. Synthetic cells encapsulating PURExpress CFES with 1-10 nM pEXP5-NT/6xHis mCherry F30 2xdBroccoli plasmid DNA. (A) Confocal images of DiD dye-tagged lipid membrane (left), expressed Broccoli mRNA (middle), and mCherry protein (right) at endpoint after 8 hours of incubation at 30 °C. Scale bar is at 100 μm. (B) Histograms of radius (μm), Broccoli mRNA RFU, and mCherry protein RFU of the synthetic cell population at the same endpoint. Black lines are gaussian distribution fits on the histograms. Coefficient of variation (CV) values of each plot are noted. (C) Single cell timeseries of the synthetic cell population for Broccoli mRNA and mCherry protein.

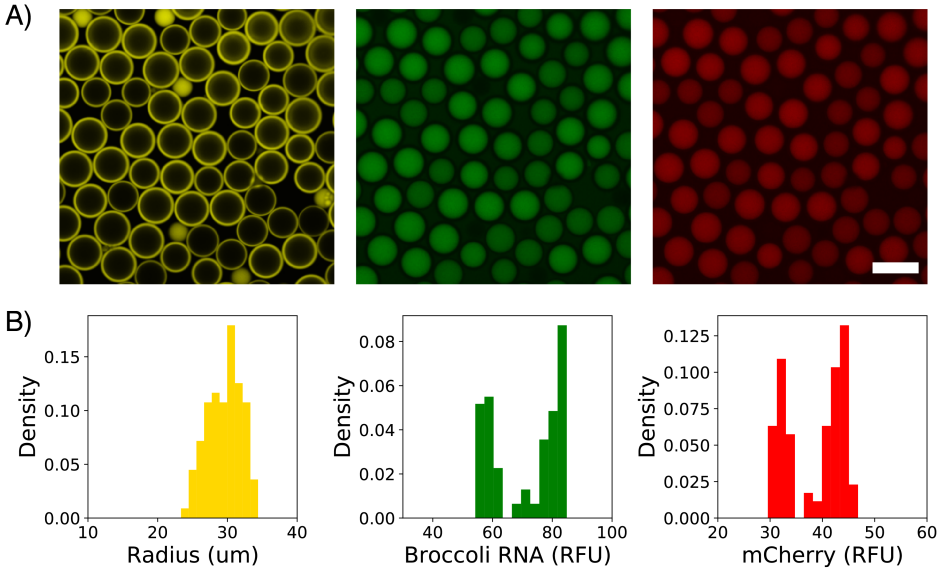

Figure 4.13: Heterogeneous population from mixed batches of synthetic cells. After production of the three synthetic cell populations presented in Section 3.3.4 (Figure 3.9-3.10), the three synthetic cell populations were combined to form a heterogeneous population with distinct subpopulations. (A) Endpoint confocal images of the heterogeneous synthetic cell population for the lipid, Broccoli RNA, and mCherry protein channels (left to right) after 12 hours of incubation at 30 °C. Scale bar is at 100 μm. (B) Histograms of radius (μm), Broccoli RNA (RF), and mCherry protein (RFU) of the synthetic cell population at the same endpoint.

4.4 Conclusions

We presented our current progress in building a synthetic cell population that is capable of quorum sensing communication. By utilizing an *E. coli* extract-based CFES, we avoid the need to externally provide of the precursors of the AHL signal molecule (SAM and acylACP). The quorum sensing gene circuits were shown to function in bulk and liposomes generated by inverse emulsion. However, microfluidic-generated synthetic cells were incompatible with the *E. coli* extract-based CFES. Specifically, the octanol oil phase inhibited gene expression in *E. coli* extract-based CFES. We are currently searching for an alternative for octanol that is compatible with both microfluidics and *E. coli* extract-based CFES. We also presented a method to incorporate variability in a monodisperse synthetic cell population. Heterogeneous synthetic cell populations can be made by using a two-inlet microfluidic device or by simply mixing two batches of synthetic cell populations together to provide a single population with a large variance or several distinct subpopulations, respectively. Combined together, these can be used as an *in vitro* platform to study the interplay between cell-to-cell communication and population variability in a minimal context. Although very simple and minimal, this synthetic multicellular system can potentially be modified with different encapsulated gene circuits with feedback loops or different modes of intercellular signalling to create more complex systems in the future.

Chapter 5

Modeling variability & communication in synthetic cell populations*

In the recent decades, biological research has greatly benefited from new technologies that allow investigation of biological systems in multiple scales and dimensions. For example, fluorescence-activated cell sorting (FACS) by flow cytometry can be used to provide snapshot data of the distribution of cells growing in a bacterial culture over time [14]. Unlike bulk methods that measure only population averaged values, single cell resolution data provides more information about the distribution of the population. This added information can be useful in determining the underlying reaction and signalling networks that can shape population distributions within biological multicellular systems. For instance, secrete-and-sensing yeast cells can respond to stimuli in a uni or bi-modal fashion depending on the strength of positive feedback against the signalling molecule [293]. Paracrine signalling is proposed to reduce the response variability of cells around a wound site by the action of a fast diffusing epidermal growth factor (EGF). The fast diffusion of EGF averages-out in the external environment and triggers a uniform and coordinated response in the nearby cells [105]. Spatiotemporal data of cells can be further tracked to get single cell trajectories over time in fluorescence microscopy videos [127]. Unlike snapshot data, single cell trajectories additionally provide information on the genealogy or history of individual cells, which can clarify correlations in cell variables [277]. To take advantage of single cell resolution data, many types of mathematical models have been developed to simulate biological multicellular systems. These include models that treat each cell as an individual in cell ensemble models or agent-based models, or population density models that provide a density description of the population dynamics. Intercellular variability can be considered by assigning cell parameters or initial conditions as random variables, or *random effects*, over the population. Cell ensemble and agent-based models are relatively easy and flexible to setup. Intracellular dynamics can be modeled using differential equations derived from a reaction or rule-based network. However, these become computationally expensive

*This chapter is partially based on the article: Gonzales DT, Tang TYD, Zechner C. **Moment-based analysis of biochemical networks in a heterogeneous population of communicating cells.** *IEEE 58th Conference on Decision and Control (CDC)*, 939-944, 2019. [94]

to simulate large cell populations, making iterative methods for the inverse problem of parameter estimation challenging. In addition, large cell ensembles are required to accurately obtain population distributions [277, 241]. Population density models avoid the need to simulate individual cells by considering only the dynamics of the population distribution or its moments. For example, the dynamics of a cell population can be described by a chemical master equation (CME) or cell population master equation (CPME), which can account for intracellular stochastic reactions (intrinsic noise) and cell-to-cell variability (extrinsic noise) in the population [241, 296]. However, solutions of complex population density models can be analytically intractable and require approximations and numerical approaches to solve [296].

In this chapter, we present two methods to model a synthetic cell population with a fixed population size. The first uses a cell ensemble approach to model a population of independent (*i.e.* non-communicating) and identically-distributed cells. To account for cell-to-cell variability, parameters and initial conditions of individual cells are drawn from a distribution and simulated using a set of deterministic reaction rate equations with either random or fixed parameters in a mixed-effects model. The second method is a population density model for secrete-and-sensing communicating cells [293, 19], such as the quorum sensing system described in Chapter 4. Here, we consider cells that are subject to both stochastic chemical reactions (intrinsic noise) and cell-to-cell parameter variability (extrinsic noise) and capture the lower-order moment dynamics of the cell population from the CME [296]. However, because cells are no longer independent in a population of communicating cells, the number of equations required will be dependent on the number of cells in the population. By applying a symmetry-based model reduction based on the work of Batmanov *et al.* (2012) [19], the reduced model becomes independent of the population size and computationally efficient.

5.1 Mixed-effects modeling for cell populations

Mixed-effects models describe the response of a system using model parameters that are either *fixed* or *random* to account for variability and correlations across a population and/or subpopulations. The fixed effects are parameters that are identical over the entire population, while random effects are parameters that vary within the population. Using a mixed-effects model is particularly useful for biological datasets that are highly structured and subject to noise and variability. This approach has been used extensively from early population pharmacokinetic studies [18] to more recent single cell resolution studies [80, 158]. Explicitly describing the mixed-effect structure of a model can help parameter inference if the specific sources of variability are appropriately identified. In addition, inferring the magnitude of variation of parameters and their correlations within the population can also be the central goal of the study [106]. Fixed and random effect parameters in mixed-effects models can be estimated by finding the parameter set that gives the maximum a posteriori probability (MAP) estimate from the observed data. However, directly calculating the MAP for mixed-effects models with large populations is computationally expensive if all population and individual cell parameters are estimated simultaneously. Instead, we apply an alternative approach similar to an expectation-maximization (EM) algorithm [278] where random and fixed parameters are estimated separately and iteratively until convergence, effectively reducing

the space over which optimization is done. In this section, we develop a mixed-effects model structure and infer its parameters from a simulated dataset of gene expression in a heterogeneous synthetic cell population.

5.1.1 Notation for mixed-effects models

In a mixed-effects model, we consider parameters that are either fixed or random following a certain distribution over the population. For instance, the output or measured data $\boldsymbol{Y_{i,j}}$ can be described by

$$\boldsymbol{Y_{i,j}} = f(t_{i,j}, \boldsymbol{\psi_i}) + \boldsymbol{\varepsilon_{i,j}} \tag{5.1}$$

where $f()$ is the model with input time $t_{i,j}$ and cell parameters $\boldsymbol{\psi_i}$. The indices i and j are for the cell and time index, respectively. The random number $\boldsymbol{\varepsilon_{i,j}}$ is the measurement error that we assume follows a normal distribution

$$\boldsymbol{\varepsilon_{i,j}} \sim N(\mathbf{0}, \boldsymbol{\sigma^2}). \tag{5.2}$$

The parameter $\boldsymbol{\psi_i}$ is a cell-specific vector that collects the fixed ($\boldsymbol{\theta}$) and random ($\boldsymbol{\phi_i}$) parameters

$$\boldsymbol{\psi_i} = \begin{pmatrix} \boldsymbol{\theta} \\ \boldsymbol{\phi_i} \end{pmatrix}. \tag{5.3}$$

We consider the random effect parameters follow a multivariate lognormal distribution

$$\boldsymbol{\phi_i} \sim LN(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{5.4}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the hyperparameters of the random effects. The lognormal distribution is used to avoid negative parameter values.

5.1.2 MAP estimation for mixed-effects models

Model parameters can be estimated by calculating the maximum a posteriori probability (MAP) of the model given the data. Specifically, we aim to maximize the posterior probability of the parameter set given the observed data

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma^2} | \boldsymbol{Y}) = \frac{p(\boldsymbol{Y}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma^2})}{p(\boldsymbol{Y})} \tag{5.5}$$

$$= \frac{p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\sigma^2}) p(\boldsymbol{\phi}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) p(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{p(\boldsymbol{Y})} \tag{5.6}$$

$$\propto p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\sigma^2}) p(\boldsymbol{\phi}|\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{5.7}$$

where Equation 5.7 follows from considering flat priors over the hyperparameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Minimizing the negative logarithm of Equation 5.7 with respect to all parameters to obtain the MAP estimate is usually analytically intractable and computationally expensive to solve. To address this problem, we use a heuristic approach that iterates between optimizing the fixed and random effects until convergence of MAP estimate. This avoids the difficult problem of minimizing all cell and population parameters together, which will lead to a very high-dimensional optimization problem with large cell populations. The method for optimization is illustrated in Algorithm 1.

Algorithm 1 Iterative method for MAP estimation of a mixed-effects model.

Set r_{iters} and f_{iters} ▷ Number of random and fixed effect iterations.
Initialize all parameters $\{\boldsymbol{\theta}, \boldsymbol{\sigma^2}, \boldsymbol{\phi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$
for k in 1:f_{iters} **do**
    $\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\sigma^2}} = \text{argmax}_{\boldsymbol{\theta},\boldsymbol{\sigma^2}}(p(\boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{\sigma^2}; \boldsymbol{\phi}))$ ▷ Optimize fixed effects.
    Update $\boldsymbol{\theta}$, $\boldsymbol{\sigma^2}$ with $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{\sigma^2}}$
    for j in 1:r_{iters} **do**
        for i in 1:cells **do** ▷ Optimize random parameters.
            $\hat{\boldsymbol{\phi}}_{\boldsymbol{i}} = \text{argmax}_{\boldsymbol{\phi_i}}(p(\boldsymbol{Y}|\boldsymbol{\phi_i}; \boldsymbol{\theta}, \boldsymbol{\sigma^2})p(\boldsymbol{\phi_i}|\boldsymbol{\mu}, \boldsymbol{\Sigma}))$
            Update $\boldsymbol{\phi_i}$ with $\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}$
        end for
        $\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}} = \text{argmax}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(p(\boldsymbol{\phi}|\boldsymbol{\mu}, \boldsymbol{\Sigma}))$ ▷ Optimize hyperparameters.
        Update $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\mu}}$,$\hat{\boldsymbol{\Sigma}}$
    end for
end for

5.1.3 Parameter estimation of heterogeneous cell populations

Using the method described in Section 5.1.2, we estimate the parameters of a simulated dataset of gene expression in heterogeneous synthetic cell populations modeled with mixed-effects. Gene expression dynamics in a single synthetic cell follows the set of differential equations

$$\frac{d[\text{DNA}]}{dt} = 0 \tag{5.8}$$

$$\frac{d[\text{RNA}]}{dt} = \frac{k_r[\text{DNA}]}{K_r + [\text{DNA}]} - \delta_r[\text{RNA}] \tag{5.9}$$

$$\frac{d[\text{Protein}]}{dt} = \frac{k_p[\text{RNA}]}{K_p + [\text{RNA}]} - \delta_p[\text{Protein}] \tag{5.10}$$

where k_r, K_r, k_p, K_p, δ_r, and δ_p are transcription rate, translation rate, RNA degradation rate, and protein degradation rate parameters, respectively. K_r, and K_p are Michaelis-Menten equilibrium constants for transcription and translation processes. For the generated dataset, we simulated two populations of synthetic cells with all rate parameters fixed at $k_r = 30$ nM/hr, $K_r = 20$ nM, $k_p = 80$ nM/hr, $K_p = 70$ nM, $\delta_r = 0.5$ 1/hr, and $\delta_p = 0.2$ 1/hr. However, initial DNA concentrations are log-normally distributed over the populations with parameters $\mu = 1$ and $\sigma = 0.5$ for population 1 and $\mu = 3$ and $\sigma = 0.3$ for population 2. This simulated dataset is similar to the experiment demonstrated in Section 4.3.3. Each synthetic cell population has a total of 15 cells. The simulated dataset has a Gaussian measurement error of $N(0, 0.25)$ for RNA and $N(0, 0.64)$ for protein is shown in Figure 5.1. Using this simulated dataset, we estimated the parameters of the model with the same random effect structure where all rate parameters are fixed and initial DNA concentrations are log-normally distributed for each population. We also assumed that the parameters K_r, K_p, δ_r, and δ_p were known so that the unknown parameters were only k_r, k_p, and initial DNA concentrations. Figure 5.2 and Table 5.1 show that gene expression dynamics, as well as fixed

and random effect parameters were accurately recovered by MAP estimation using Algorithm 1 using 50 fixed effect iterations (f_{iters}) with 5 inner random effect iterations (r_{iters}) each.

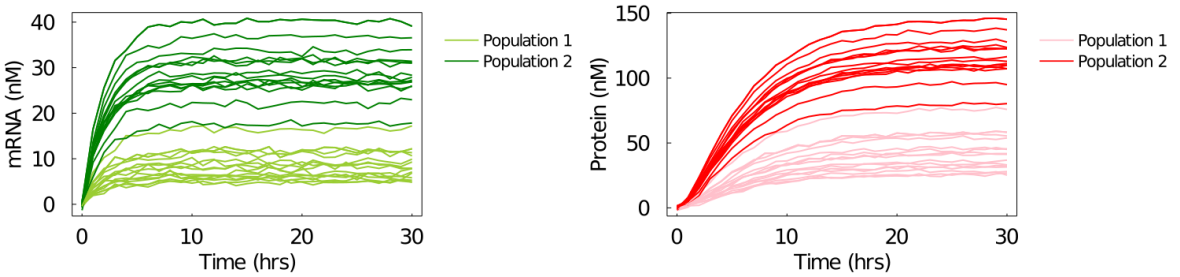

Figure 5.1: Simulated gene expression dynamics in two heterogeneous cell populations. RNA (left) and protein (right) dynamics are described by the model in Equation 5.8-5.10 with fixed rate parameters $k_r = 30$ nM/hr, $K_r = 20$ nM, $k_p = 80$ nM/hr, $K_p = 70$ nM, $\delta_r = 0.5$ 1/hr, and $\delta_p = 0.2$ 1/hr. Initial DNA concentrations are log-normally distributed with $LN(1, 0.5)$ for population 1 and $LN(3, 0.3)$ for population 2 with 15 cells each. All traces have a Gaussian measurement error of $N(0, 0.25)$ for RNA and $N(0, 0.64)$ for protein.

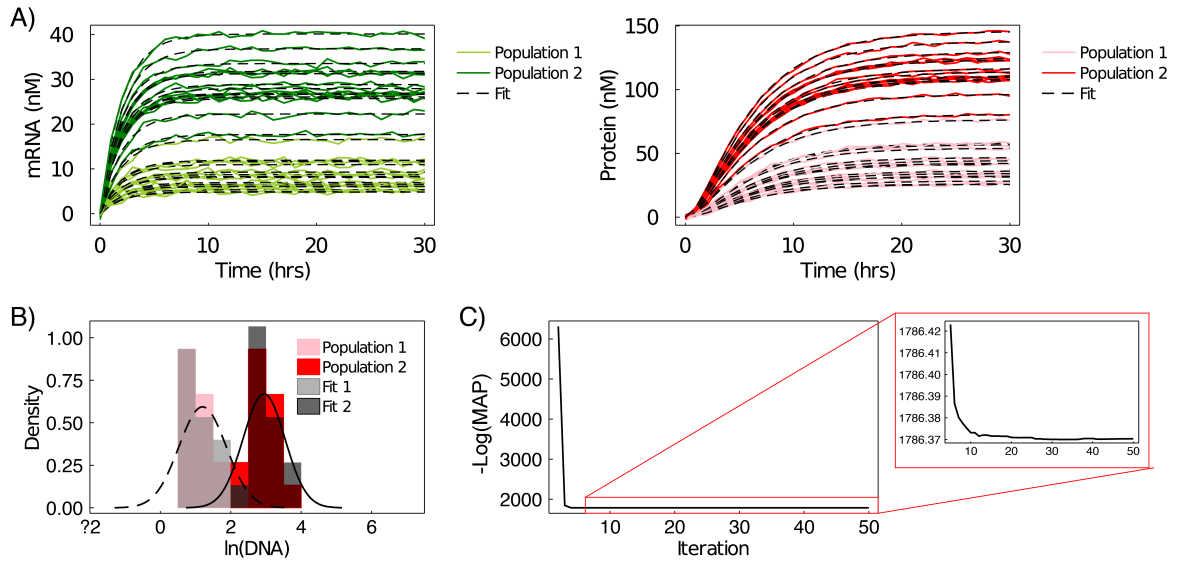

Figure 5.2: Parameter estimation of heterogeneous synthetic cell populations. (A) RNA and protein dynamics for the two synthetic cell populations and their corresponding fits (dashed lines). (B) Histograms of initial DNA concentrations of the heterogeneous synthetic cell populations (pink and red) with the fit initial DNA concentrations (grey and black). Dashed and solid lines show the log-normal distributions of the populations from the fit hyperparameters. Note that the x-axis scale is in log scale. (C) Negative logarithm of the MAP estimate showing convergence after 50 fixed effect iterations. Inset shows the values from iterations 5-50.

Table 5.1: True, initial, and fit parameter values of the simulated cell population. LN denotes a log-normal distribution with parameters μ and σ. All initial [DNA] values were 3 nM and 20 nM for individual cells from population 1 and 2, respectively.

Parameter	True	Initial	Fit
k_r	30	10	29.12
K_r	20	-	-
k_p	80	100	79.98
K_p	70	-	-
δ_r	0.5	-	-
δ_p	0.2	-	-
σ_r	0.5	0.1	0.48
σ_p	0.8	0.1	0.82
$[\text{DNA}]_1 \sim \text{LN}(\mu_1, \sigma_1)$	(1,0.5)	(1,1)	(1.20,0.45)
$[\text{DNA}]_2 \sim \text{LN}(\mu_2, \sigma_2)$	(3,0.3)	(1,1)	(2.94,0.35)

5.2 Moment dynamics of communicating cells

In this section, we demonstrate an approach to model an intercellularly communicating cell population with *both* intrinsic and extrinsic sources of variability. Here, cells are no longer independent of each other and biochemical reactions are considered stochastic. Although our cell-free expression systems typically have deterministic gene expression dynamics, using lower concentrations of DNA plasmids and encapsulating these into smaller compartments could potentially result in stochastic gene expression due to lower copy numbers. Computational models and algorithms to study stochastic biochemical networks inside living cells are most commonly based on the chemical master equation (CME), whose solution provides a time-dependent probability distribution over molecular concentrations [89]. CME-based models can account for the discrete and random nature of biochemical reactions (intrinsic noise) as well as population heterogeneity stemming from differences in each cell's microenvironment (extrinsic variability) [71]. The computational analysis of the CME is challenging but several efficient numerical techniques have been proposed in the past, including stochastic simulation algorithms [88], moment-based methods [237, 296] and combinations thereof [215, 67, 82]. However, the majority of existing approaches to study noise in cell populations rely on the assumption that individual cells act independently of each other. More concretely, each cell's dynamics is considered to be an independent and identically distributed realization of the same stochastic process. This assumption is violated in systems where cells communicate with one another to coordinate their behavior [238, 293]. Typical examples include quorum-sensing systems in bacterial colonies [246], or paracrine communication in higher organisms [105]. Understanding the interplay between cell-cell communication and the stochastic behavior of individual cells is an important challenge and demands for suitable mathematical approaches. However, extending standard techniques to account for cell-cell communication leads to computational difficulties, because the dimensionality of the resulting models increases with the number of cells in a population.

Recently, first attempts have been made to develop more tractable stochastic models of systems of communicating cells. In [238], for instance, the authors use a moment-

based method to study how neighbour-neighbour-coupling affects concentration fluctuations in a tissue. A related approach has been proposed in [19] to study community effects in cells that interact with each other by secreting and sensing certain signalling molecules. In particular, the authors show how the dimensionality of the considered model can be dramatically reduced by exploiting certain symmetries in the governing equations. However, both approaches account exclusively for intrinsic noise, whereas extrinsic sources have not been considered. In [27], the authors study how chemical communication via a quorum sensing molecule affects intrinsic and extrinsic noise in cell communities. To obtain tractable simulations, they used a tailored approach that combines stochastic simulations with a quasi-steady state approximation to eliminate fast variables from the model. In this section, we develop a moment-based approach to study noise from both intrinsic and extrinsic sources in secrete-and-sensing communicating cells [293, 19]. We first derive a system of differential equations that captures lower order moments of the population. Since the number of obtained equations grows combinatorially with number of considered cells, we perform a symmetry-based model reduction based on the work of Batmanov et al. (2012) [19]. The dimensionality of the reduced model is independent of the population size, which makes it computationally very efficient. We employ this approach to study how noise is affected by cell-to-cell communication in several biochemical networks.

5.2.1 Moment-based analysis of secrete-and-sensing cells

We consider a population of N genetically identical cells that communicate with each other through a diffusing signalling molecule such as a quorum sensing signalling molecule. Each individual cell i is associated with an identical set of S chemical species $\mathrm{X}_{i,1}, \ldots, \mathrm{X}_{i,S}$ that interact with one another via M biochemical reaction channels. Without loss of generality, we consider the first $S-1$ species to be confined to the intracellular environment of cell i. The Sth species corresponds to the signalling molecule, which can shuttle between the intra- and extracellular environment through transport reactions. For simplicity, we consider the external environment to be well-mixed, such that the import of signalling molecules into any cell i does not depend on the spatial configuration of the system. In total, the system can be described by a reaction network

$$\sum_{k=1}^{S} \alpha_{j,k} \mathrm{X}_{i,k} \rightarrow \sum_{k=1}^{S} \beta_{j,k} \mathrm{X}_{i,k} \tag{5.11}$$

$$\mathrm{X}_{i,S} \leftrightharpoons \mathrm{X}_{E}, \tag{5.12}$$

for $i = 1, \ldots, N$ and $j = 1, \ldots, M-2$. In Equation 5.11, $\alpha_{j,k}$ and $\beta_{j,k}$ correspond to the reactant and product coefficients of the respective reaction. In Equation 5.12, the species X_E denotes the signalling molecule in the external environment. In total, the network comprises $NS+1$ chemical species and NM chemical reactions. We define by $X_i(t) = (X_{i,1}(t), \ldots, X_{i,S}(t))$ the state of cell i, which collects the copy numbers of all species associated with this cell at time t. The state of the overall system is then given by $X(t) = (X_1(t), \ldots, X_N(t), X_E(t))$, with $X_E(t)$ as the number of signalling molecules in the external environment. Moreover, we denote by $\nu_{i,k} \in \mathbb{Z}^{NS+1}$ the stoichiometric change vector associated with the kth reaction channel of cell i consistent with reaction network (Equation 5.11-5.12).

5.2.2 Moment dynamics of heterogeneous cell communities

If both the intra- and extracellular environments are well-mixed, we can describe the state $X(t)$ as a continuous-time Markov chain, whose state probability distribution $P(x,t) = P(X(t) = x)$ admits a master equation of the form

$$\frac{\mathrm{d}P(x,t)}{\mathrm{d}t} = \sum_{i=1}^{N}\sum_{j=1}^{M}[a_{i,j}(x-\nu_{i,j},c_{i,j})P(x-\nu_{i,j},t) - a_{i,j}(x,c_{i,j})P(x,t)]. \tag{5.13}$$

In Equation 5.13, the function $a_{i,j}$ is the reaction propensity associated with reaction j in cell i. Throughout this article, we consider the propensities to obey the law of mass action such that $a_{i,j}(x,c_{i,j}) = c_{i,j}g_j(x_i,x_E)$ with x_i as the part of the state vector associated with cell i, x_E as the abundance of the signalling molecule in the environment, $c_{i,j} \in \mathbb{R}^+$ as a stochastic rate constant, and g_j as a polynomial. Note that while g_j is identical for all cells, we allow the individual rate constants $c_{i,j}$ to vary across the population. This provides a means to account for extrinsic sources of cell-to-cell variability [296]. We consider the reaction rate constants for each cell i to be independent random vectors $C_i = (C_{i,1},\dots,C_{i,M})$ drawn from a common probability distribution $C_i \sim p_c(\cdot)$ for $i = 1,\dots,N$. Note that deterministic (non-varying) reaction rates can be accounted for by letting p_c be a Dirac measure with respect to this parameter. With $C = (C_1,\dots,C_N)$, we can formulate a master equation for the conditional distribution $P(x,t \mid c) = P(X(t) = x \mid C = c)$, i.e.,

$$\frac{\mathrm{d}P(x,t \mid c)}{\mathrm{d}t} = \sum_{i=1}^{N}\sum_{j=1}^{M}[a_{i,j}(x-\nu_{i,j},c_{i,j})P(x-\nu_{i,j},t \mid c) - a_{i,j}(x,c_{i,j})P(x,t \mid c)]. \tag{5.14}$$

To analyze the cell community model, we resort to a moment-based approach, which provides a lower-dimensional description of the population and its heterogeneity. More precisely, we seek for the population moments

$$\begin{aligned}\langle\phi(X,C)\rangle &= \big\langle\langle\phi(X,C) \mid C\rangle\big\rangle \\ &= \left\langle \sum_{x\in\mathcal{X}} \phi(x,C)P(x,t \mid C) \right\rangle,\end{aligned} \tag{5.15}$$

with $\phi : (x,c) \to \mathbb{R}$ as a monomial in (x,c) and $\mathcal{X}$ as the domain of $X(t)$. Note that we omit the dependency of the moments on time for the sake of a compact notation. In order to derive a differential equation for the time evolution of $\langle\phi(X,C)\rangle$, we calculate the derivative of Equation 5.15 and insert the r.h.s. of Equation 5.14

$$\begin{aligned}\frac{\mathrm{d}\langle\phi(X,C)\rangle}{\mathrm{d}t} = \Bigg\langle \sum_{x\in\mathcal{X}} \phi(x,C) \sum_{i=1}^{N}\sum_{j=1}^{M} &[a_{i,j}(x-\nu_{i,j},C_{i,j})P(x-\nu_{i,j},t \mid C) \\ &- a_{i,j}(x,C_{i,j})P(x,t \mid C)]\Bigg\rangle.\end{aligned} \tag{5.16}$$

Using a change of variable, Equation 5.16 simplifies to

$$\frac{\mathrm{d}\langle\phi(X,C)\rangle}{\mathrm{d}t} = \left\langle \sum_{i=1}^{N}\sum_{j=1}^{M} \Big\langle \phi(X+\nu_{i,j},C)a_{i,j}(X,C_{i,j}) \mid C \Big\rangle - \Big\langle \phi(X,C)a_{i,j}(X,C_{i,j}) \mid C \Big\rangle \right\rangle, \tag{5.17}$$

where the inner brackets denote expectations conditionally on the random parameters C. Now, using double expectations, we obtain

$$\frac{\mathrm{d}\langle\phi(X,C)\rangle}{\mathrm{d}t} = \sum_{i=1}^{N}\sum_{j=1}^{M} \Big\langle \phi(X+\nu_{i,j},C)a_{i,j}(X,C_{i,j}) \Big\rangle - \Big\langle \phi(X,C)a_{i,j}(X,C_{i,j}) \Big\rangle. \tag{5.18}$$

The Equation 5.18 describes the time evolution of moments and cross moments for a heterogeneous population of secrete-and-sense cells. It can thus be seen as an extension of the moment equations derived in [296] to account for cell-to-cell communication. Indeed, if we set the transport rates to zero, all cells in the population become independent of each other such that the two approaches yield equivalent solutions. Note that depending on the details of the system, the moment equations from Equation 5.18 may not be closed. This is the case in the presence of second-order reactions, or even first-order reactions if their rate constants are randomly distributed across the population. Moment-closure approximation (MA) techniques provide a popular means to address this problem by imposing certain assumptions on the underlying state probability distribution [237, 296, 218]. More precisely, this allows us to replace the higher order moments that appear on the r.h.s. of Equation 5.18 by functions of the lower order moments. These functions are referred to as *closure functions* and their particular form depends on the distributional assumption we make. Popular choices include the normal [218] and lognormal [237] closure functions and we will adopt those in the present study. To check the accuracy of our MAs in this study, we compare them with moments calculated using the Stochastic Simulation Algorithm (SSA) [88]. Throughout this article, we consider moments of up to second order and replace all third-order moments that appear on the r.h.s. of (5.18) using MA functions provided in Table 5.2.

Table 5.2: Moment-closure approximation (MA) functions of the order three.

MA	$\langle \boldsymbol{X_1X_2X_3} \rangle$
Normal	$\langle X_1\rangle\langle X_2X_3\rangle + \langle X_2\rangle\langle X_1X_3\rangle + \langle X_3\rangle\langle X_1X_2\rangle - 2\langle X_1\rangle\langle X_2\rangle\langle X_3\rangle$
Lognormal	$\dfrac{\langle X_1X_2\rangle\langle X_2X_3\rangle\langle X_1X_3\rangle}{\langle X_1\rangle\langle X_2\rangle\langle X_3\rangle}$

5.2.3 Symmetry-based model reduction

While Equation 5.18 provides a more tractable description of the cell community than Equation 5.13, its dimensionality still scales combinatorially with the number of

considered cells N. In case we consider all first and second order moments, the number of equations is given by

$$K_{eq} = 2(NS+1) + \binom{NS+NM'+1}{2} - \binom{NM'}{2}, \tag{5.19}$$

where M' is the number of rate constants among the M reactions that vary from cell to cell. The first term in Equation 5.19 accounts for 1st and 2nd order moments of the chemical species, and the remaining terms account for all cross-moments of species and rate constants that change over time. The number of first and second order moment equations scales quadratically with N, which limits the above approach to relatively small population sizes. However, the moment system can be reduced substantially by taking into account the symmetries of the considered model as has been proposed in [19]. More precisely, if we consider all initial cell states $X_i(0)$ to be identically distributed, the moment dynamics of each cell will be equivalent and indistinguishable for all times $t > 0$ such that $\langle X_{i,k}\rangle = \langle X_{j,k}\rangle$, $\langle X_{i,k}^2\rangle = \langle X_{j,k}^2\rangle$, $\langle X_{i,k}X_{j,l}\rangle = \langle X_{m,k}X_{n,l}\rangle$, and $\langle X_{i,k}X_E\rangle = \langle X_{j,k}X_E\rangle$ for any $i \neq j$, $m \neq n$, k and l. Consequently, we can obtain a reduced model by considering only the moments up to the second order and cross-moments of the states of any two reference cells $X_i(t)$ and $X_j(t)$ as well as the amount of signalling molecules in the external environment $X_E(t)$. The resulting set of equations can be reduced further by eliminating all moments associated with one of the two cells j (e.g., $\langle X_{j,k}\rangle$ or $\langle X_{j,k}X_E\rangle$) since those are identical to the corresponding moments of cell i. In total, the required number of equations is then given by

$$\hat{K}_{eq} = 2(S+1) + \binom{S+1+M'}{2} + \binom{S+M'}{2} - 2\binom{M'}{2} + S, \tag{5.20}$$

which is thus independent of the population size.

5.2.4 Illustrative example

To demonstrate how the original system can be reduced based on symmetries, consider the toy model

$$\mathrm{X}_{i,1} \xrightarrow{C_{i,1}} \mathrm{X}_{i,2} \tag{5.21}$$

$$\mathrm{X}_{i,2} \underset{c_3}{\overset{c_2}{\rightleftharpoons}} \mathrm{X}_E, \tag{5.22}$$

with $i = 1, \ldots, N$. We consider the rate $C_{i,1}$ to be randomly distributed across the population, while c_2 and c_3 are identical in all cells. We distinguish the moment equations of the system between two cases. The first case concerns equations that are the same in the original and reduced model. These are equations of the moments that do not involve the signalling molecule in the external environment such as $\langle X_{i,1}\rangle$, $\langle X_{i,2}\rangle$, or $\langle X_{i,1}X_{i,2}\rangle$. For example, in both the original and reduced models, the expectation of species $X_{i,2}$ satisfies

$$\frac{\mathrm{d}\langle X_{i,2}\rangle}{\mathrm{d}t} = \langle C_{i,1}X_{i,1}\rangle + c_2\langle X_E\rangle - c_3\langle X_{i,2}\rangle. \tag{5.23}$$

The dynamics of moments and cross-moments involving the external signalling molecule depend on *all* cells in the population due to the transport reactions. In the case of

$\langle X_{i,2}X_E\rangle$, the original equation is:

$$\begin{aligned}\frac{\mathrm{d}\langle X_{i,2}X_E\rangle}{\mathrm{d}t} =&\langle C_{i,1}X_{i,1}X_E\rangle + c_2\langle X_E^2\rangle - c_2\langle X_E\rangle + c_3\langle X_{i,2}^2\rangle \\ &- c_3\langle X_{i,2}X_E\rangle - c_3\langle X_{i,2}\rangle + \sum_{k\neq i} c_3\langle X_{k,2}X_{i,2}\rangle - c_2N\langle X_{i,2}X_E\rangle.\end{aligned} \tag{5.24}$$

The terms in the sum capture the dependencies between two cells k and i. Since all terms in the sum are identical due to the symmetry of the population, they can be replaced by the contribution of any cell $j \neq i$ multiplied by $(N-1)$,

$$\begin{aligned}\frac{\mathrm{d}\langle X_{i,2}X_E\rangle}{\mathrm{d}t} =&\langle C_{i,1}X_{i,1}X_E\rangle + c_2\langle X_E^2\rangle - c_2\langle X_E\rangle + c_3\langle X_{i,2}^2\rangle \\ &- c_3\langle X_{i,2}X_E\rangle - c_3\langle X_{i,2}\rangle + c_3(N-1)\langle X_{j,2}X_{i,2}\rangle - c_2N\langle X_{i,2}X_E\rangle.\end{aligned} \tag{5.25}$$

Therefore, if we perform analogous manipulations for all other moments and cross-moments involving the signalling molecule (e.g., $\langle X_E\rangle$, $\langle X_{i,1}X_E\rangle$) and eliminate all remaining redundancies between cell i and j, we arrive at a system of 17 coupled differential equations, independent of the population size N. For further information on the symmetry-based model reduction approach, the reader may refer to [19].

5.2.5 Case studies

In this section, we use the described moment-based approach to study how cell-cell communication affects noise in different biochemical networks. Python scripts used for this study are available in www.github.com/zechnerlab/CommunityMoments. Moment equations were solved using a SciPy numerical solver (solve_ivp). SSA sample paths from Gillespie's Direct Method were simulated using Tellurium [43] and used to test the accuracy of the moment-based approach. For the sake of a compact notation, molecular species are assigned different letters and reaction rate parameters are assigned letters with superscripts.

Birth-death process

As a first example, we study a birth-death process in a heterogeneous population of interacting cells, *i.e.*,

$$\begin{aligned}\emptyset \xrightarrow{C_i^b} \mathrm{P}_i \xrightarrow{C_i^d} \emptyset \\ \mathrm{P}_i \underset{c^t}{\overset{c^t}{\rightleftharpoons}} \mathrm{Q},\end{aligned} \tag{5.26}$$

for $i = 1, \ldots, N$. Here, the birth and death reaction rate constants are independent random variables as indicated by capital letters C_i^b and C_i^d, while the transport rate c^t is fixed. For this reaction network, a symmetry-reduced model of up to the second order can be described using $\hat{K}_{eq} = 12$ differential equations regardless of the population size.

The goal of this case study is to study how cell-to-cell communication affects the variability in the abundance of P_i. To quantify variability, we use two metrics: the first one is the coefficient of variation (CV) defined as

$$\mathrm{CV}[P_i] = \sqrt{\frac{\langle P_i^2\rangle - \langle P_i\rangle^2}{\langle P_i\rangle^2}}, \tag{5.27}$$

for any $i = 1, \ldots, N$. The CV captures the expected variation in protein abundance inside single cells across *different* populations. Due to the symmetry, we have $\mathrm{CV}[P_i] = \mathrm{CV}[P_k]$ for any i and k. Furthermore, we define the pair variation (PV)

$$\mathrm{PV}[P_i, P_k] = \sqrt{\frac{\langle (P_i - P_k)^2 \rangle}{\langle P_i \rangle \langle P_k \rangle}}, \tag{5.28}$$

which captures the expected variation between two different cells i and k within the *same* population. Note that in the absence of cell-cell communication, (Equation 5.27) and (Equation 5.28) are identical up to a scaling factor of $\sqrt{2}$. For a communicating population, however, this is not the case due to correlations between cells in the population. In Figure 5.3A and B, we compare the first and second order moments of species P_i for the symmetry-reduced model of size $N = 10$ with stochastic simulations and found a good agreement between the two approaches. In Figure 5.3C-F, we show the dependency of $\mathrm{CV}[P_i]$ and $\mathrm{PV}[P_i,P_j]$ as a function of the transport rate as well as the population size. The coefficient of variation $\mathrm{CV}[P_i]$ decreases with increasing transport rates and to some extent also with the population size. While the pair variation $\mathrm{PV}[P_i,P_j]$ shows a similar decrease with increasing transport rates, it seems to be independent of the population size. We also show how extrinsic variability changes the steady-state variability of P_i by increasing the CV of the birth rate parameter C_i^b, while keeping all other parameters constant. Both $\mathrm{CV}[P_i]$ and $\mathrm{PV}[P_i,P_j]$ increase with increasing extrinsic variability, whereas large transport rates can attenuate this effect due to spatial averaging of P_i levels (Figure 5.4).

Autocatalytic circuit

Next, we focus on an autocatalytic system defined by

$$\begin{aligned} \emptyset &\xrightarrow{C_i^b} \mathrm{A}_i \xrightarrow{C_i^d} \emptyset \\ \mathrm{A}_i &\xrightarrow{c^a} \mathrm{A}_i + \mathrm{A}_i \\ \mathrm{A}_i &\underset{c^t}{\overset{c^t}{\rightleftharpoons}} \mathrm{B}. \end{aligned} \tag{5.29}$$

Here, we consider C_i^b and C_i^d to be randomly distributed across the population, while the autocatalytic rate c^a and transport rate c^t are fixed. In this example, the total number of reduced equations is $\hat{K}_{eq} = 12$. To obtain a closed set of moments, we applied the lognormal closure and checked its accuracy by comparing it to Monte Carlo estimates of the moments calculated over 1000 SSA realizations (Figure 5.5). We generally found a good agreement between the MA and the SSA. Similar to the birth-death systems, we observe that both CV and PV of species A_i decrease with increasing transport rates. Increasing the population size does not affect the PV, but decreases its CV.

To study the relationship between the CV and PV, we plotted $\mathrm{CV}[A_i]$ against $\mathrm{PV}[A_i,A_j]$ in Figure 5.6. As expected, CV and PV are related by a factor of $\sqrt{2}$ when $c^t = 0$. Slight deviations from this scaling are possible due to the approximations involved in the derivation of the moment equations. In the presence of communication, $\mathrm{PV}[A_i,A_j]$ drops below the $\sqrt{2}$ scaling law, indicating that the variability between cells in the same population is smaller than the variability of cells across different populations.

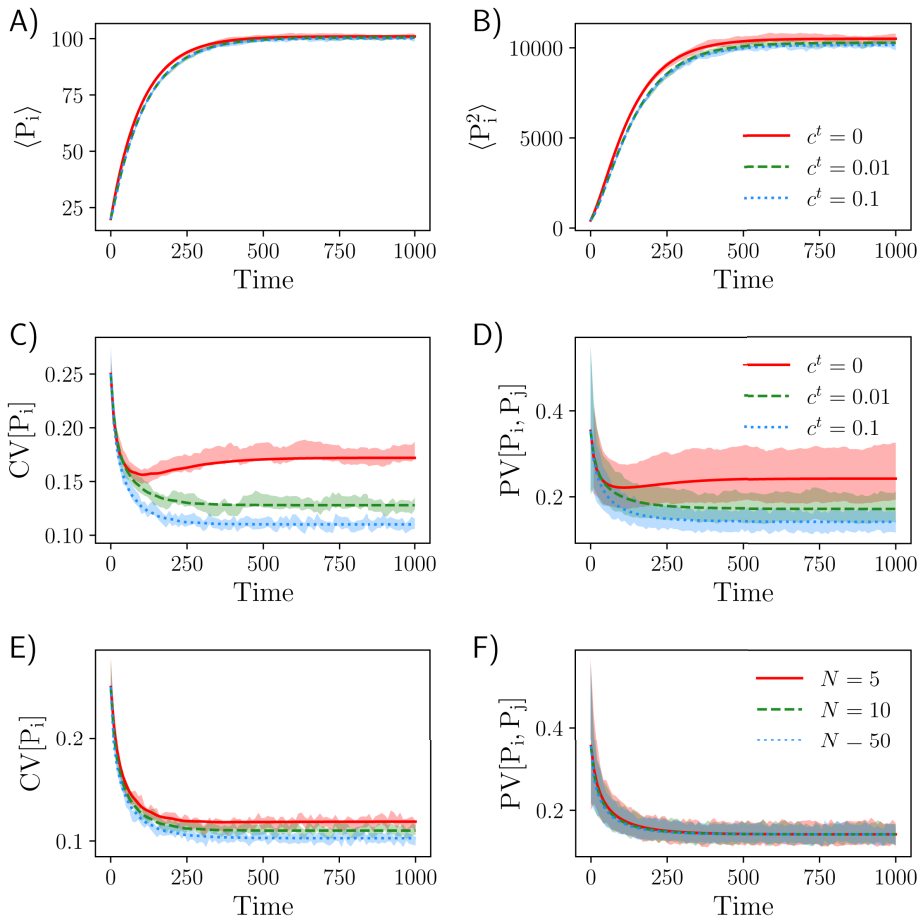

Figure 5.3: Moment dynamics of communicating cells with a birth-death reaction using the symmetry-reduced model (Normal MA). (A) First and (B) second order moment dynamics of species P_i with a population size of $N = 10$ and $c^t = [0, 0.1, 0.1]$. (C) CV$[P_i]$ and (D) PV$[P_i, P_j]$ decrease as transport rate increases for a fixed population size $N = 10$. Increasing the population size $N = [5, 10, 50]$ for a fixed $c^t = 0.1$ decreases the (E) CV$[P_i]$, but not (F) PV$[P_i, P_j]$. Other parameters and initial conditions were set to $\langle C_i^b \rangle = 1$, Var$[C_i^b] = 0.01$, $\langle C_i^d \rangle = 0.01$, Var$[C_i^d] = 1e-6$, $\langle P_i(0) \rangle = 20$, Var$[P_i(0)] = 25$, $\langle Q(0) \rangle = 0$, and Var$[Q(0)] = 0$. Shaded areas are bootstrapped 95% confidence intervals (CI) from 1000 SSA realizations.

Genetic feedback circuit

Lastly, we tested the moment-based method using a larger system. In particular, we focus on a genetic feedback circuit given by

$$\begin{aligned} \mathrm{D}_i &\xrightarrow{C_i^m} \mathrm{D}_i + \mathrm{M}_i & \mathrm{M}_i &\xrightarrow{d^m} \emptyset \\ \mathrm{M}_i &\xrightarrow{C_i^p} \mathrm{M}_i + \mathrm{P}_i & \mathrm{P}_i &\xrightarrow{d^p} \emptyset \\ \mathrm{D}_i + \mathrm{P}_i &\underset{c^d}{\overset{c^a}{\rightleftharpoons}} \mathrm{DP}_i & \mathrm{P}_i &\underset{c^t}{\overset{c^t}{\rightleftharpoons}} \mathrm{Q}, \\ \mathrm{DP}_i &\xrightarrow{c^f} \mathrm{DP}_i + \mathrm{M}_i \end{aligned} \tag{5.30}$$

for $i = 1, \ldots, N$. Here, we consider the reaction rate constants associated with transcription C_i^m and translation C_i^p to be randomly distributed across the population,

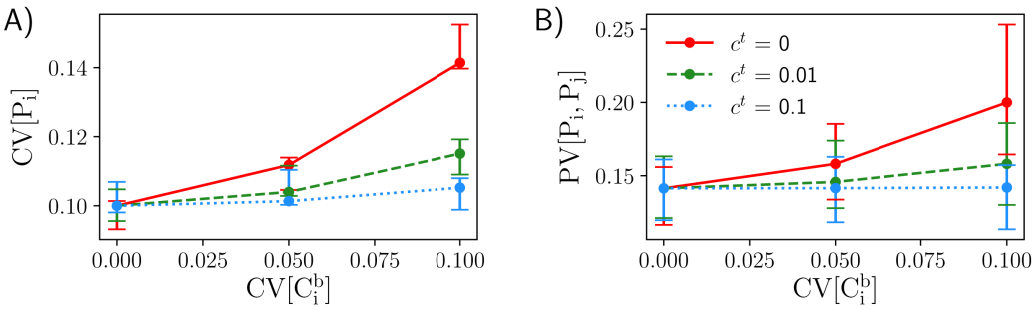

Figure 5.4: Variability reduction due to higher transport rates. Steady-state (A) CV$[P_i]$ and (B) PV$[P_i, P_j]$ for different degrees of population heterogeneity (CV$[C_i^b]$) in communicating cells are reduced for higher transport rates (c^t). Other parameters and initial conditions were set to $\langle C_i^b \rangle = 1$, $c^d = 0.01$, $\langle P_i(0) \rangle = 20$, Var$[P_i(0)] = 25$, $\langle Q(0) \rangle = 0$, Var$[Q(0)] = 0$, and $N = 10$. Steady-state moments were determined from the reduced model with normal closure at $t = 1000$. Error bars are bootstrapped 95% CIs from 1000 SSA realizations.

while all other rate parameters are fixed. Transcriptional feedback of the gene circuit is mediated by the protein product (P), which binds DNA (D) to form a complex DP. In the bound state, the gene can be transcribed with rate constant c^f. Depending on the ratio of the bound and unbound transcription rate, the feedback mechanism can either enhance ($c^f > \langle C_i^m \rangle$) or inhibit ($c^f < \langle C_i^m \rangle$) gene expression. We consider a population of $N = 10$ cells and applied a lognormal closure to solve for the reduced moment dynamics. The total number of reduced equations for this system is $\hat{K}_{eq} = 48$. Figure 5.7 shows that the time evolution of the mRNA moments obtained from the symmetry-reduced model and SSA are very similar to each other. To analyze the computational efficiency of the moment-based approach, we recorded run times for the positive feedback gene circuit for different population sizes N and compared it against the SSA (Table 5.3). As expected, the run time for the moment-based model does not increase with N while the run time associated with the SSA increases by orders of magnitude.

Table 5.3: Run times (s) of the reduced moment-based approach vs. SSA for the positive feedback gene circuit*.

N	**Reduced moments**	**SSA** (1 sample path)
10	0.068 ± 0.005	0.142 ± 0.004
100	0.065 ± 0.004	12.300 ± 0.123
1000	0.064 ± 0.005	2569.20 ± 52.93

*Mean and standard deviation from 10 replicates.

5.3 Conclusions

We presented two approaches for modeling cell populations with noise and variability. The first uses a mixed-effects model to account for cell-to-cell variability in a non-communicating cell population. The second uses a moment-based approach to model secrete-and-sensing communicating heterogeneous cell populations with stochastic gene expression. In principle, these methods apply to arbitrary reaction networks

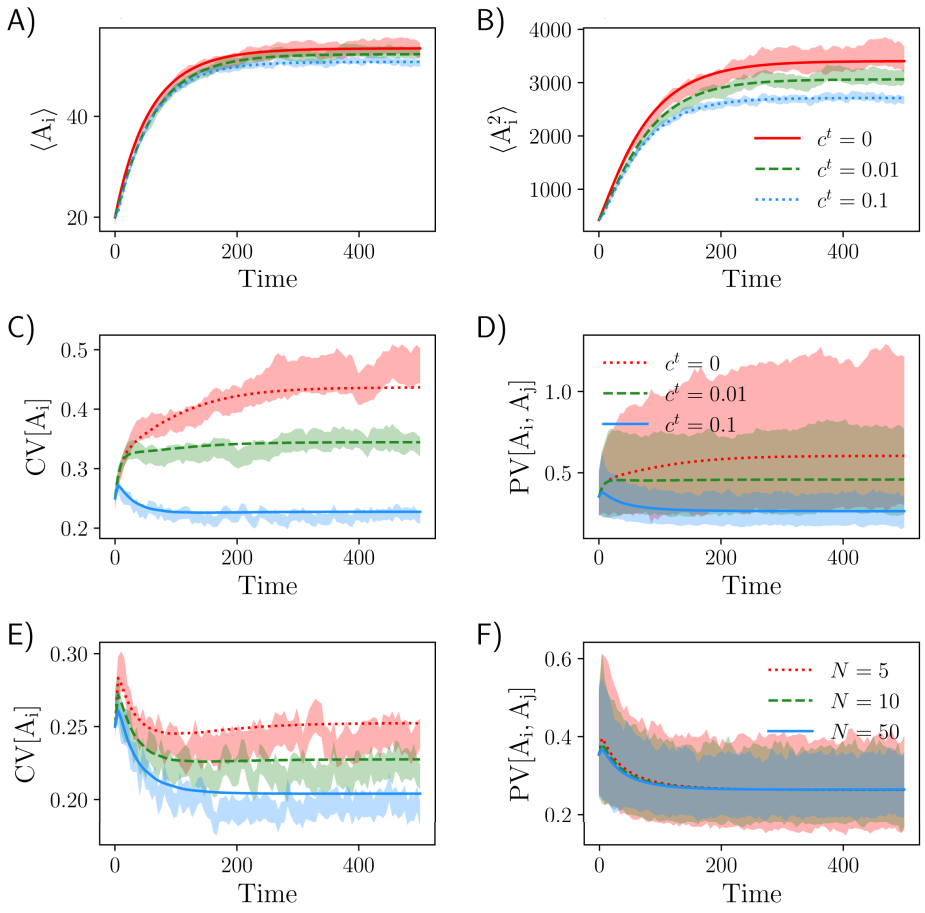

Figure 5.5: Moment dynamics of communicating cells with an autocatalytic reaction using the symmetry-reduced model (Lognormal MA). (A) First and (B) second order moment dynamics of species A_i for a population size of $N = 10$ and $c^t = [0, 0.01, 0.1]$. (C) CV$[A_i]$ and (D) PV$[A_i, A_j]$ decrease as transport rate increases for a fixed population size $N = 10$. Increasing the population size from $N = [5, 10, 50]$ for a fixed $c^t = 0.1$ decreases (E) CV$[A_i]$, while (F) PV$[A_i, A_j]$ remains the same. Other parameters and initial conditions were set to $\langle C_i^b \rangle = 1$, Var$[C_i^b] = 0.01$, $c^a = 0.08$, $\langle C_i^d \rangle = 0.1$, Var$[C_i^d] = 1e-4$, $\langle A_i(0) \rangle = 20$, Var$[A_i(0)] = 25$, $\langle B(0) \rangle = 0$, and Var$[B(0)] = 0$. Shaded areas are bootstrapped 95% CIs from 1000 SSA realizations.

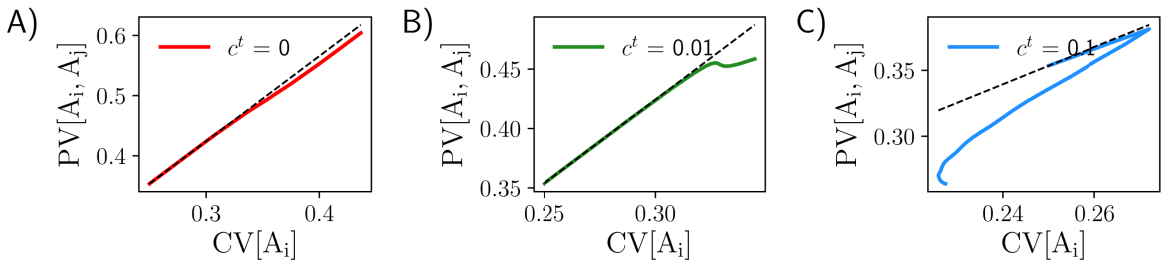

Figure 5.6: CV$[A_i]$ vs. PV$[A_i,A_j]$ of the autocatalytic network at different transport rates $c^t = [0, 0.01, 0.1]$ (A-C, respectively). CV$[A_i]$ and PV$[A_i,A_j]$ are related by a factor of $\sqrt{2}$ (diagonal) at $c^t = 0$, while deviations from this scaling occur when $c^t > 0$. Values of CV$[A_i]$ and PV$[A_i,A_j]$ were taken from the time-courses corresponding to the case $N = 10$ in Figure 5.5.

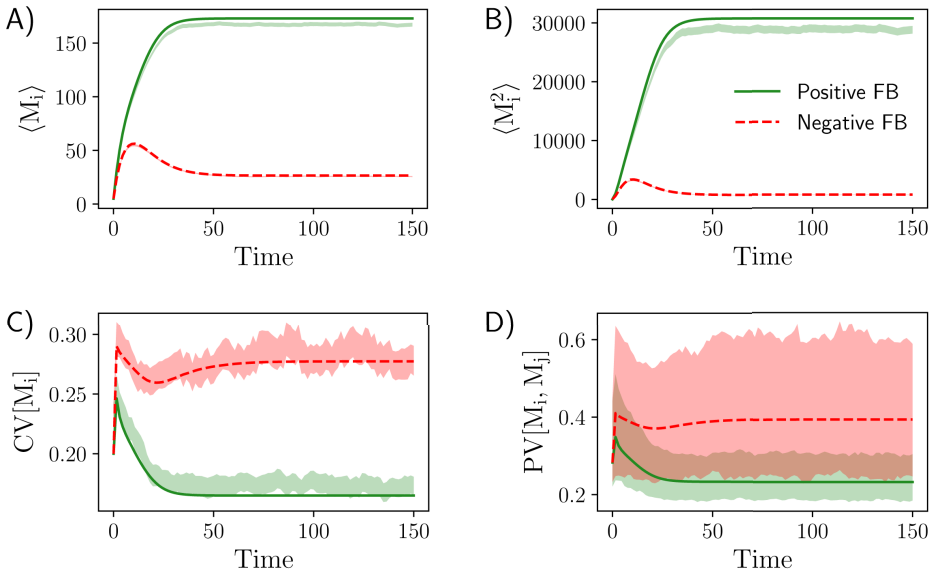

Figure 5.7: Moment dynamics of communicating cells with a feedback circuit using the symmetry-reduced model (Lognormal MA). (A) First and (B) second order moment dynamics of species M_i for populations of $N = 10$ cells with negative feedback ($c_i^f = 0.1$) and positive feedback ($c_i^f = 1$) and their corresponding time evolution of (C) CV$[M_i]$ and (D) PV$[M_i, M_j]$. Reaction rate parameters and initial conditions were set to $\langle D_i(0)\rangle = 30$, Var$[D_i(0)] = 25$, $\langle M_i(0)\rangle = \langle P_i(0)\rangle = \langle DP_i(0)\rangle = 5$, Var$[M_i(0)] =$ Var$[P_i(0)] =$ Var$[DP_i(0)] = 1$, $\langle Q(0)\rangle = 5$, Var$[Q(0)] = 0$, $\langle C_i^m\rangle = 0.5$, Var$[C_i^m] = 0.01$, $\langle C_i^p\rangle = 0.05$, Var$[C_i^p] = 2.5e-5$, $c^a = 0.01$, $c^d = 0.01$, $d^m = d_p = 0.2$, and $c^t = 0.8$. Shaded areas are bootstrapped 95% CIs from 1000 SSA realizations.

with a well-mixed external signalling molecule and suitable moment closure functions for the moment-based approach. Although both modeling approaches do not capture many additional complexities of biological cell populations such as cell growth, death, division, and other metabolic processes, these can be ideal for modeling our synthetic cell populations or . Importantly, methods to solve the inverse problem of parameter estimation can be used determine the working parameter space of quantified experiments. We demonstrated this in Section 5.1.3 by MAP estimation of simulated data from a non-linear mixed-effects model. Several other methods that use a restricted maximum likelihood (REML) approach [107], EM algorithm [278], or a stochastic approximation of the EM algorithm (SAEM) [137, 123] are also available, along with tools for model selection and identifiability analysis for non-linear mixed-effects models [47, 146, 158]. Moment-based inference has also been successfully demonstrated to capture intrinsic and extrinsic sources of variability in cell populations [296, 297, 212]. As our capability to build more complex bottom-up synthetic multicellular systems grow, these modeling approaches will be important for designing emergent or population-level behavior from single synthetic cells. For example, we aim to use these models to study the effect of cell-to-cell communication on a heterogeneous population of quorum sensing synthetic cells that we have demonstrated in Chapter 4.

Summary and outlook

In this work, we aimed to build a synthetic multicellular system by combining technologies of cell-free expression systems and microfluidics to study the effect of cell-to-cell communication to population variability in a minimal context. In Chapter 3, we established our platform to reproducibly encapsulate CFES into monodisperse synthetic cell populations by using a double-emulsion microfluidic device [289] and simultaneously quantify both transcription and translation dynamics using a Broccoli RNA aptamer [78] and mCherry protein. Importantly, this allowed us to test different coarse-grained resource-limited gene expression models [247] using model selection and obtain gene expression rate parameters and their respective confidence intervals by likelihood-based parameter estimation and profile likelihood analysis [208], respectively. By obtaining quantified rate parameters, we can begin to compare biologically relevant parameters between *in vivo*, and *in vitro* bulk and encapsulated formats of gene expression such as RNA polymerase transcription and ribosome translation rates. For instance, we found that T7 RNA polymerase transcription and ribosome translation rates in PURExpress were much lower than *in vivo* rates in *E. coli* by one to two orders of magnitude, but comparable to previous studies in PURExpress within the same or one order of magnitude.

In Chapter 4, we incorporated variability in these synthetic cell populations by either mixing different batches of synthetic cells to make a heterogeneous cell population with distinct sub-populations, or by using a two-inlet microfluidic device [225] to generate a population with a large variance of encapsulated plasmid DNA. In both cases, the synthetic cells are still monodisperse in size - providing single cell reproducibility that is important for assembling multi-scale structures and building predictable multicellular systems. This enables us to customize initial DNA distributions of the populations, which will result in different distributions of expressed protein over time depending on the gene circuit and strength of cell-to-cell coupling through intercellular communication. We then used an *E. coli* extract-based CFES to implement a quorum sensing gene circuit for diffusion-mediated cell-to-cell communication. The quorum sensing system functions in bulk reactions and synthetic cells generated by an inverse emulsion methodology. However, we found that the octanol oil phase in our double-emulsion microfluidics methodology was not compatible with the *E. coli* extract-based CFES. Different oil phases that are compatible to both *E. coli* extract-based CFES and double-emulsion microfluidics as well as other encapsulation methods, such as droplet interface bilayers, are currently being tested to encapsulate the quorum sensing gene circuits in extract-based CFES.

In Chapter 5, gene expression models of cell populations were developed to include population variability and both intrinsic and extrinsic noise with intercellular communication. In the first method, we used a mixed-effects model with deterministic rate equations to describe a synthetic cell population composed of independent cells. Population variability was included by letting certain parameters vary over the population. We then applied a heuristic optimization approach to solve the inverse problem of parameter inference. Using an iterative method that alternates between maximizing fixed and random effect parameters, rate parameters of the population were correctly inferred from simulated data. This approach is especially useful for our synthetic cell populations and some biological settings where the measured dynamics are deterministic but subject to extrinsic variability. In addition, different random effect structures of the model can be tested and compared to help determine the extrinsic sources of variability. In the second method, we developed a model of a cell population with both intrinsic and extrinsic noise [296] and intercellular communication through a diffusible signalling molecule. By assuming that the external environment is well-mixed or that the diffusion of the signalling molecule in the environment is fast, the number of moment equations to describe the population can be significantly decreased by a symmetry-based model reduction [19]. Although this approach can be applied to a quorum sensing system with both stochastic chemical reactions and population variability, the assumption of a well-mixed environment might not apply to our current setup of synthetic cell populations. If time scales of diffusion and the quorum sensing gene circuit are comparable to each other, spatiotemporal reaction-diffusion models will be necessary.

In the immediate future, we aim to build on our work and generate populations of quorum sensing communicating synthetic cells. Using our platform, single-cell gene expression dynamics can be quantitatively and spatiotemporally monitored to investigate the effect of cell-to-cell communication on population variability. Overall, we envision that engineered systems made of synthetic micron-sized compartments encapsulating molecular reactions, coupled with mathematical modeling, can help facilitate the design and construction of complex and multiscale chemical systems from the bottom-up. These systems can lead to exciting applications in tools for complex molecular assembly, bio-inspired materials, and *in vitro* platforms for intercellular signalling pathways.

Appendix A

Chemical and equipment list

Table A.1: Chemical list.

Name	**MW** (g/mole)	**Supplier**	**Catalog No.**
Amino acids			
L-Alanine (A)	89.09	Sigma, USA	A7627
L-Arginine (R)	174.20	Sigma, USA	A5006
L-Asparagine (N)	132.12	Sigma, USA	A0884
L-Aspartic acid (D)	133.10	Sigma, USA	A9256
L-Cysteine (C)	121.16	Sigma, USA	W326305
L-Glutamic acid (E)	147.13	Sigma, USA	G1251
L-Glutamine (Q)	146.14	Sigma, USA	G3126
Glycine (G)	75.07	Sigma, USA	G7126
L-Histidine (H)	155.15	Sigma, USA	H8000
L-Isoleucine (I)	131.17	Sigma, USA	I2752
L-Leucine (L)	131.17	Sigma, USA	L8000
L-Lysine (K)	146.19	Sigma, USA	L5501
L-Methionine (M)	149.21	Sigma, USA	M9625
L-Phenylalanine (F)	165.19	Sigma, USA	P2126
L-Proline (P)	115.13	Sigma, USA	P0380
L-Serine (S)	105.09	Sigma, USA	S4500
L-Threonine (T)	119.12	Sigma, USA	T8625
L-Tryptophan (W)	204.23	Sigma, USA	T0254
L-Tyrosine (Y)	181.19	Sigma, USA	T3754
L-Valine (V)	117.15	Sigma, USA	V0500
NTPs			
ATP	551.14	Sigma, USA	A26209
CTP	527.12	Sigma, USA	30320
GTP	523.18	Roche, CH	10106399001
UTP	586.12	Sigma, USA	94370

Name	**MW** (g/mole)	**Supplier**	**Catalog No.**
Enzymes			
Benzonase	-	MPI-CBG	-
DNAse I	-	NEB, USA	M0303
DpnI	-	NEB, USA	R0176
Phusion HF	-	NEB, USA	M0531
RNAseA	-	Carl Roth, DE	7156.1
Ladders			
1kb DNA ladder	-	NEB, USA	N3232
PageRuler Plus protein ladder	-	Thermo, USA	26616
Riboruler HR RNA ladder	-	Thermo, USA	SM1821
Dyes			
DFHBI	252.22	Sigma, USA	SML1627
DiD	1052.08	Invitrogen, USA	D7757
Gel Loading Dye	-	NEB, USA	B7024
GelRed	-	Biotium, USA	41003
Laemmli buffer	-	Bio-rad, USA	1610737
SYBRgold	-	Thermo, USA	S11494
Texas Red DHPE	1380.77	Biotium, USA	60027
Others			
FTCS	681.57	Sigma, USA	729965
1-Octanol	130.23	Sigma, USA	297887
40% Polyacrylamide	-	Bio-rad, USA	1610144
Agarose	-	Invitrogen, USA	16500
Ammonium peroxydisulphate	228.20	Carl Roth, DE	9592.3
Ampicillin	371.39	Sigma, USA	A9518
β-NAD	663.43	Sigma, USA	N1511
β-mercaptoethanol	78.13	Sigma, USA	M6250
Coenzyme A	767.53	Sigma, USA	C4282
Dithiothreitol	154.253	Thermo, USA	R0862
DMSO	78.13	Sigma, USA	D8418
Egg PC	770.123	Avanti, USA	840051
Folinic acid	511.50	Sigma, USA	F7878
D-(+)-Glucose	180.16	Sigma, USA	G8270
Glycerol	92.09	VWR, USA	24388.295
300 mOsmol/kg std.	-	Gonotec, DE	30.9.0020
2000 mOsmol/kg std.	-	Gonotec, DE	30.9.2000
HEPES	238.31	Carl Roth, DE	9105
Hydrochloric acid	36.46	Merck, USA	100317
Hydrogen peroxide	34.01	Sigma, USA	16911
Imidazole	68.08	AppliChem, DE	A1073
Magnesium chloride	203.30	Sigma, USA	105833
Magnesium glutamate	388.61	Sigma, USA	49605
Mineral oil (0.84 g/mL)	-	Sigma, USA	M5904

Name	**MW** (g/mole)	**Supplier**	**Catalog No.**
Nuclease-free water	-	Thermo, USA	AM9937
N-(3-oxohexanoyl)-L-homoserine lactone	213.23	Sigma, USA	K3007
Oxalic acid	126.07	Roth, DE	8879.1
PDADMAC (1.04 g/mL)	-	Sigma, USA	409014
Phosphoenolpyruvate	206.1	Roche, CH	10108294001
Pluronic acid F-68	-	Gibco, USA	24040032
Poly(4-styrenesulfonic acid)	75000	Sigma, USA	561223
POPC	760.076	Avanti, USA	850457C
Potassium glutamate	203.23	Sigma, USA	G1149
Potassium hydroxide	56.11	Sigma, USA	221473
Potassium phosphate dibasic	174.18	Sigma, USA	P8584
Potassium phosphate monobasic	136.086	Sigma, USA	P8709
Protease Inhibitor Cocktail	-	Roche, CH	COEDTAF
Putrescine	88.15	Sigma, USA	51799
Sodium chloride	58.44	Merck, USA	1.06404
Spermidine	145.25	Sigma, USA	S2626
Sucrose	342.30	Sigma, USA	S9378
SU-8 2025	-	Microchem, USA	-
TEMED	116.21	Carl Roth, DE	2367.3
Trizma base	121.14	Sigma, USA	T1503
Tris-Glycine-SDS buffer (10X)	-	Bio-rad, USA	1610772
tRNA from *E. coli* MRE600	-	Roche, CH	10109541001
Twinsil Speed silicone	-	Picodent, DE	1300 1002
Urea	60.06	Merck, USA	108487

Table A.2: Kit list.

Name	**Supplier**	**Catalog No.**
HiScribe T7 High Yield RNA Synthesis Kit	NEB, USA	E2040
NEBuilder HiFi DNA Assembly Kit	NEB, USA	E2621
PURExpress In Vitro Protein Synthesis Kit	NEB, USA	E6800
Plasmid Maxi Kit	QIAGEN, DE	12162
RNeasy Mini kit	QIAGEN, DE	74104
myTXTL Sigma 70 Master Mix Kit	Arbor Biosciences, USA	507024
QIAquick Gel Extraction Kit	QIAGEN, DE	28704
QIAquick PCR Purification Kit	QIAGEN, DE	28104
Quick Start Bradford Assay	Bio-Rad, USA	5000201

Table A.3: Consumables list.

Name	Supplier	Catalog No.
24x60 mm coverslip	Menzel-Gläser, DE	-
384-well plates, Lobase, Black	Greiner Bio-One, AT	788096
76x26x1 mm microscope slides	Marienfeld, DE	1000000
HisTrap FF column (5 mL)	Cytiva, USA	GE17-5255-01
Millex syringe filter (PVDF)	Millipore, USA	SLGV033RS
Adhesive PCR Plate Seals	Thermo, USA	AB0558
Spectra/Por2 Dialysis Membrane (12-14 kD)	Repligen, USA	132680T
TGX Stain-Free PAGE Gel, 15-well, 4-15%	Bio-Rad, USA	4568086

Table A.4: Equipment list.

Name	Supplier	Catalog No.
Objectives		
100x/1.3 Oil Plan-Neofluar Ph3 M27	Zeiss, DE	420491-9910-000
10X/0.45 Plan-Apochromat M27	Zeiss, DE	420640-9900-000
40X/1.2 C-Apochromat W autocorr M27	Zeiss, DE	421767-9971-790
20X/0.4 LD A-Plan Ph2	Zeiss, DE	421051-9910-000
5x/0.15 Plan-Neofluar Ph1 M27	Zeiss, DE	420331-9911-000
Equipment		
Andor Axiovert 200M	Zeiss, DE	-
Avanti Centrifuge J26-XP	Beckman Coulter, USA	-
Biovision Gel Doc system	Biovision, USA	-
Branson Digital Sonifier	Emerson Electric, USA	450-D
Dimax S4 Monochrome sCMOS high-speed camera	PCO, DE	-
JLA-8.1000 rotor	Beckman Coulter, USA	363688
LSM 880 with Airyscan	Zeiss, DE	-
Mitos P-Pump	Dolomite, UK	3200016
NanoDrop 2000	Thermo, USA	ND-2000
Osmomat 3000	Gonotec, DE	-
pE-4000 illumination system	CoolLED, USA	-
Sonorex sonicator bath	Bandelin, DE	-
Spark 20M plate reader	TECAN, CH	-
Genios Pro plate reader	TECAN, CH	-
Spin Coater	Laurell Tech., USA	WS-650MZ-23B
Typhoon 9500 Fluo and Phospho Imager	GE, USA	-

Appendix B

Plasmid sequences and cloning

The strains and plasmids used in this study are listed and described in Table B.1 and Table B.2. Plasmids pEXP5-NT/6xHis eGFP [152] and pEXP5-NT/6xHis mCherry [255] were provided by JLR Anderson, University of Bristol. The plasmid p70a(2) deGFP was obtained from the myTXTL Sigma 70 Cell-Free Master Mix kit (Arbor Biosciences, USA). The pEXP5-NT/6xHis mCherry F30-2xdBroccoli and pEXP5-NT/6xHis mCherry F30-Broccoli plasmids were made by inserting the F30-2xdBroccoli and F30-Broccoli RNA aptamer sequences [78, 77] downstream the mCherry stop codon and upstream the terminator of the mCherry gene in the pEXP5-NT/6xHis mCherry. The F30-2xdBroccoli indicates two dimeric Broccoli RNA aptamers supported by an F30 stem loop, while F30-Broccoli has one Broccoli RNA aptamer supported by an F30 stem loop. To create the insert for pEXP5-NT/6xHis mCherry F30-2xdBroccoli and pEXP5-NT/6xHis mCherry F30-Broccoli plasmids, Broccoli RNA aptamer inserts were synthesized as ultramers with overlapping regions to the pEXP5-NT/6xHis mCherry vector plasmid. F30-2xdBroccoli ultramers 1 and 2 were annealed, PCR amplified against each other, and column purified (QIAquick PCR Purification Kit) to obtain the dsDNA F30-2xdBroccoli insert. The pEXP5-NT/6xHis mCherry plasmid vector was divided into two parts by PCR amplifying the template plasmid pEXP5-NT/6xHis mCherry using the primer pair NFB ptetO gBlock Fwd and mCherry Rev, and primer pair Tphi Fwd and DUR Vector 2. The PCR products were then digested with DpnI (NEB, USA) to remove the plasmid template and then gel purified (QIAquick Gel Extraction Kit) to obtain the vector fragments only. Plasmid assembly of the purified DNA parts for pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid was assembled using the NEBuilder HiFi DNA Assembly Kit (NEB) and transformed into *E. coli* DH5α. The pEXP5-NT/6xHis mCherry F30-Broccoli plasmid was similarly constructed with the same two vector parts and a synthesized F30-Broccoli ultramer. The quorum sensing plasmids were constructed from the base pEXP5-NT/6xHis eGFP plasmid. For the pEXP5-NT/pLux 6xHis eGFP plasmid, the T7 promoter was replaced with the LuxR-AHL inducible promotor (pLux promoter) from the *lux* operon. To prepare the vector, the pEXP5-NT/6xHis eGFP was digested with BglII and XbaI restriction enzymes (NEB, USA) to remove the T7 promoter, and then gel purified (QIAquick Gel Extraction Kit) to obtain the vector fragments only. The insert pLux promoter sequence was synthesized as an ultramer (pLux std ultramer), PCR amplified using the primer pair pLux Fwd and pLux Rev, digested with BglII and XbaI restriction enzymes, and then column purified (QIAquick PCR Purification Kit). The purified vector from the digested pEXP5-NT/6xHis eGFP and pLux promoter insert were lig-

ated together by T4 DNA Ligase (NEB, USA), and then transformed into *E. coli* DH5α. To create the pEXP5-NT/6xHis LuxRopt and pEXP5-NT/6xHis LuxIopt plasmids, *E. coli* codon optimized sequences for LuxR (LuxRopt) and LuxI (LuxIopt) were synthesized as double-stranded DNA gene blocks (IDT, USA) and then replaced into the pEXP5-NT/6xHis eGFP plasmid. The vector parts were prepared by PCR amplifying the pEXP5-NT/6xHis eGFP plasmid with primer pair TEV Rev and PIVEX2 rev, and primer pair T7 Fwd and PIVEX1 fwd. The PCR products were then digested with DpnI (NEB, USA) to remove the plasmid template and then gel purified (QIAquick Gel Extraction Kit) to obtain the vector fragments only. Plasmid assembly of the purified DNA parts for pEXP5-NT/6xHis LuxRopt and pEXP5-NT/6xHis LuxIopt plasmids were assembled using the NEBuilder HiFi DNA Assembly Kit (NEB) and transformed into *E. coli* DH5α. To construct the pEXP5-NT/pLux 6xHis LuxIopt plasmid, vector parts were PCR amplified from the pEXP5-NT/pLux 6xHis eGFP plasmid using the primer pair Amp Fwd and His Rev EX, and primer pair TAAT7 Fwd EX Amp Rev. The PCR products were then digested with DpnI (NEB, USA) to remove the plasmid template and then gel purified (QIAquick Gel Extraction Kit) to obtain the vector fragments only. The LuxIopt insert was prepared by PCR amplification using the primer pair LuxIopt Fwd and LuxIopt Rev and then column purified (QIAquick PCR Purification Kit). Plasmid assembly of the purified DNA parts for pEXP5-NT/pLux 6xHis LuxIopt plasmid were assembled using the NEBuilder HiFi DNA Assembly Kit (NEB) and transformed into *E. coli* DH5α. All plasmids have a high-copy number origin of replication. All plasmids were prepared and purified by ethanol precipitation from *E. coli* DH5α cultures using the Plasmid Maxi Kit (QIAGEN, USA) and measured by NanoDrop 2000 (Thermo, USA). Primer, ultramer, and gene block sequences are provided in Table B.3, Table B.4, and Table B.6. PCR cycling protocols are shown in Table B.5. All PCR, purification, and assembly methods were performed using the manufacturer's standard protocols. All primers and ultramers, and gene blocks were synthesized by Integrated DNA Technologies (IDT, www.idtdna.com). All assembled plasmids were confirmed by Sanger sequencing (GENEWIZ, www.genewiz.com) using the primers pEXP5NT insert Fwd and/or pEXP5NT insert Rev. The pEXP5-NT/6xHis mCherry F30-2xdBroccoli and pEXP5-NT/6xHis mCherry F30-Broccoli plasmids are deposited in Addgene (www.addgene.org, plasmid ID 169233-169234).

Table B.1: *E. coli* strains.

Strain	**Description**	**Source**
E. coli DH5α	High-efficiency competent cells for transformation.	NEB (C2987)
E. coli BL21 (DE3)	Competent cells for T7 RNAP-mediated expression.	NEB (C2527)

Table B.2: *E. coli* bacterial plasmids.

Plasmid	Description	Source
pEXP5-NT/6xHis eGFP	Constitutive T7 RNAP-mediated expression of 6xHis-eGFP.	[152]
pEXP5-NT/6xHis mCherry	Constitutive T7 RNAP-mediated expression of 6xHis-mCherry.	[255]
pEXP5-NT/6xHis mCherry F30-2xdBroccoli	Constitutive T7 RNAP-mediated expression of 6xHis-mCherry with F30-2xdBroccoli RNA aptamer.	This study.
pEXP5-NT/6xHis mCherry F30-Broccoli	Constitutive T7 RNAP-mediated expression of 6xHis-mCherry with F30-Broccoli RNA aptamer.	This study.
p70a(2) deGFP	Constitutive σ_{70}-mediated expression of truncated eGFP (deGFP).	Arbor Biosciences
pEXP5-NT/6xHis LuxRopt	Constitutive T7 RNAP-mediated expression of 6xHis-Lux.	This study.
pEXP5-NT/6xHis LuxIopt	Constitutive T7 RNAP-mediated expression of 6xHis-LuxI.	This study.
pEXP5-NT/pLux 6xHis eGFP	LuxR-AHL inducible expression of 6xHis eGFP.	This study.
pEXP5-NT/pLux 6xHis LuxIopt	LuxR-AHL inducible expression of 6xHis-LuxI.	This study.

Table B.3: Primers used in this study.

Primer	Sequence (5'-3')
NFB ptetO gBlock Fwd	TTTCTCCTTACGCATCTGT
mCherry Rev	ATCACCCTTTAGCTGCC
Tphi Fwd	CGAAAGGAAGCTGAGTTG
DUR vector 2	CTGGCTTAACTATGCGGC
pEXP5NT insert Fwd	TCGGTGATTCATTCTGCT
pEXP5NT insert Rev	GGTTATTGTCTCATGAGCG
TEV fwd	GGTAGCAGCGGCGAAAAC
T7 Rev	CTTCCTTTCGGGCTTTGTTAG
TEV Rev	AAAATACAGGTTTTCGCCG
PIVEX2 rev	CTCTGCTAATCCTGTTACCA
T7 Fwd	TCCGGCTGCTAACAAAG
PIVEX1 Fwd	ATCGTCTTGAGTCCAACC
pLux Fwd	AGTTATAGATCTGAGCTGGC
pLux Rev	ATAACTTCTAGACGACTATAACA
LuxIopt Fwd	ATGACCATTATGATTAAAAAATCGG
LuxIopt Rev	CGCTGCCAGGGCGTAATTTT
Amp Fwd	TATCATTGCAGCACTGGG
Amp Rev	TAGATAACTACGATACGGGAGG

Table B.4: Ultramers and extension primers used in this study.

Ultramer	Sequence (5'-3')
F30-2xdBroccoli 1	CCGGCGGCATGGACGAGCTGTACAAGGGCAGC
	TAAAGGGTGATCTTGCCATGTGTATGTGGGAG
	ACGGTCGGGTCCATCTGAGACGGTCGGGTCCA
	GATATTCGTATCTGTCGAGTAGAGTGTGGGCT
	CAGATGTCGAGTAGAGTGTGGGCTCCCACATA
	CTCTGATGATCCAGACGGTCGGGTCCATCTGA
F30-2xdBroccoli 2	TTATTGCTCAGCGGTGGCAGCAGCCAACTCAG
	CTTCCTTTCGGGCTTTGTTAGCAGCCGTTGCC
	ATGAATGATCCAGCCCACACTCTACTCGACAT
	CTGAGCCCACACTCTACTCGACAGATACGAAT
	ATCTGGACCCGACCGTCTCAGATGGACCCGAC
	CGTCTGGATCATCA
F30-Broccoli EX	ACTCAGCTTCCTTTCGGGCTTTGTTAGCAGTT
	GCCATGAATGATCCCGAAGGATCATCAGAGTA
	TGTGGGAGCCCACACTCTACTCGACAGATACG
	AATATCTGGACCCGACCGTCTCCCACATACAC
	ATGGCAACCGGATCACCCTTTAGCTGCC
pLux std	AGTTATAGATCTGAGCTGGCGTACGCGTTGAA
	CACTTCACAGATGATAGGGATTCGGGTAAAGA
	GCGTGTCATTGGGGGCTTATACAGGCGTTACA
	ATTACTTAACATAAGCACCTGTAGGATCGTAC
	AGGTTTACGCAAGAAAATGGTTTGTTATAGTC
	GTCTAGAAGTTAT
TAAT7 Fwd EX	AAAATTACGCCCTGGCAGCGTAAAGGGTGATC
	CGGCTGC
His Rev EX	ATTTTTTAATCATAATGGTCATGGACTGAAAA
	TACAGGTTTTCG

Table B.5: PCR cycling protocol for plasmid parts. Extension times were held between 0.5 or 1.5 minutes for PCR products < 1 kb or > 1 kb, respectively.

PCR cycle	Temperature	Time
Initial denaturation	98 °C	40 s
Denaturation	98 °C	10 s
Annealing	53 °C	15 s
Extension	72 °C	1.5/0.5 mins
Final extension	72 °C	5 mins
Hold	4 °C	∞

Table B.6: Gene blocks or dsDNA used in this study. Geneious v11.0.2 was used for the codon optimization of LuxR and LuxI sequences with the target organism *E. coli* K12 and a rare threshold of 0.3.

Ultramer	**Sequence (5'-3')**
LuxRopt	GGTAGCAGCGGCGAAAACCTGTATTTTCAGTCCATGAA
	AAACATCAATGCCGACGACACCTACCGCATTATCAATAA
	AATCAAAGCGTGTCGCAGCAATAATGACATCAATCAATG
	CCTGTCTGATATGACCAAAATGGTGCATTGTGAATATTA
	TCTGCTGGCGATTATTTATCCGCATTCTATGGTGAAATC
	TGATATTAGCATTCTGGATAATTACCCGAAAAAATGGCG
	TCAATATTATGATGACGCGAACCTTATCAAATATGACCC
	GATTGTGGATTATTCTAACTCCAATCATAGCCCAATCAA
	TTGGAATATTTTTGAAAACAATGCGGTGAATAAAAAAT
	CTCCGAATGTTATCAAAGAAGCGAAAACCAGCGGTCTG
	ATTACGGGCTTTAGTTTCCCAATTCATACGGCGAACAAT
	GGCTTCGGTATGCTGAGTTTTGCACATAGCGAAAAAGA
	CAACTATATTGATAGTCTGTTTCTGCACGCGTGTATGAA
	CATTCCGCTGATTGTTCCGTCTCTGGTTGATAATTATCG
	CAAAATCAATATTGCAAATAATAAAAGCAACAACGACCT
	GACCAAACGCGAAAAAGAATGTCTGGCGTGGGCGTGCG
	AAGGCAAATCGTCTTGGGATATTAGCAAAATTCTGGGT
	TGCAGTGAACGCACCGTCACCTTCCATCTGACCAATGCC
	CAAATGAAACTGAATACCACCAACCGCTGCCAAAGTATT
	TCTAAAGCAATTCTGACGGGTGCAATTGATTGCCCGTAT
	TTCAAAAACTAAAGGGTGATCCGGCTGCTAACAAAGCCC
	GAAAGG
LuxIopt	ATGACCATTATGATTAAAAAATCGGATTTTCTGGCAATT
	CCGTCGGAAGAATATAAAGGTATTCTGAGTCTGCGTTA
	TCAAGTGTTTAAACAACGCCTGGAATGGGACCTGGTTG
	TGGAAAATAACCTGGAAAGCGATGAATATGATAACAGC
	AATGCAGAATATATTTATGCGTGTGATGATACCGAAAA
	TGTGAGTGGCTGCTGGCGTCTGCTGCCGACCACCGGTG
	ATTATATGCTGAAAAGTGTTTTTCCGGAACTGCTGGGTC
	AACAGAGTGCGCCGAAAGATCCGAATATTGTCGAACTG
	AGTCGTTTTGCGGTGGGTAAAAATAGCAGCAAAATTAA
	TAACTCTGCGAGTGAAATTACCATGAAACTGTTTGAAGC
	GATTTATAAACACGCGGTTAGTCAAGGTATTACCGAATA
	TGTGACCGTGACCAGCACCGCAATTGAACGCTTTCTGAA
	ACGTATTAAAGTTCCGTGTCATCGTATTGGCGACAAAGA
	AATTCATGTGCTGGGTGATACCAAATCGGTTGTGCTGT
	CTATGCCGATTAATGAACAGTTTAAAAAAGCAGTCCTGA
	ATGCAGCGAACGACGAAAATTACGCCCTGGCAGCGTAA

Appendix C

mRNA expression and purification

The mRNA transcript of the pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid containing the mCherry gene and Broccoli RNA aptamer was transcribed *in vitro* and purified for RNA calibration and RNA titration experiments. To prepare the DNA template for *in vitro* transcription (IVT), the region containing the T7 promoter up to the terminator of the pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid was PCR amplified (Phusion HF) using primers pEXP5-NT Insert Fwd and Rev (Table B.3) and according to Table B.5 with an extension time of 45 s, digested with DpnI (NEB, 20 units/50 μL PCR reaction volume) to remove the plasmid template, and purified (QIAquick PCR Purification Kit). The purified PCR product was used as the template for IVT using the HiScribe T7 High Yield RNA Synthesis Kit which was used according to manufacturer's instructions with 5 hours incubation at 37 °C and DNAse I treatment. The IVT reaction product is then purified using the QIAGEN RNeasy kit as described by the manufacturer. The purified mRNA was confirmed by Nanodrop 260/280 ratios of ∼2.0, native agarose gel, and denaturing urea-PAGE. We observed two separate bands in the denaturing urea polyacrylamide gel for the 6xHis mCherry F30-2xdBroccoli RNA transcript. To check if these two bands still contain the Broccoli aptamer, an in-gel DFHBI staining was performed according to [77], which showed that both bands fluoresce with DFHBI. Therefore, both fractions had the complete RNA aptamer sequences. The two bands in the denaturing RNA gel could be a result of transcription read-through over the terminator of the PCR amplicon DNA template. However, we assumed that both species contained the F30-2xdBroccoli RNA aptamer because the DNA template is from a pure single band PCR product, both RNA bands are only slightly longer than the expected size (∼1224 bp), and both bands are fluorescent with the DFHBI from the in-gel staining. Therefore, using this purified mRNA for our calibration curves is consistent on a molar concentration basis. Purified mRNA was stored at -20 °C until further use. The agarose DNA gel, native and denaturing urea polyacrylamide gels, and DFHBI in-gel stains are shown in Figure C.1.

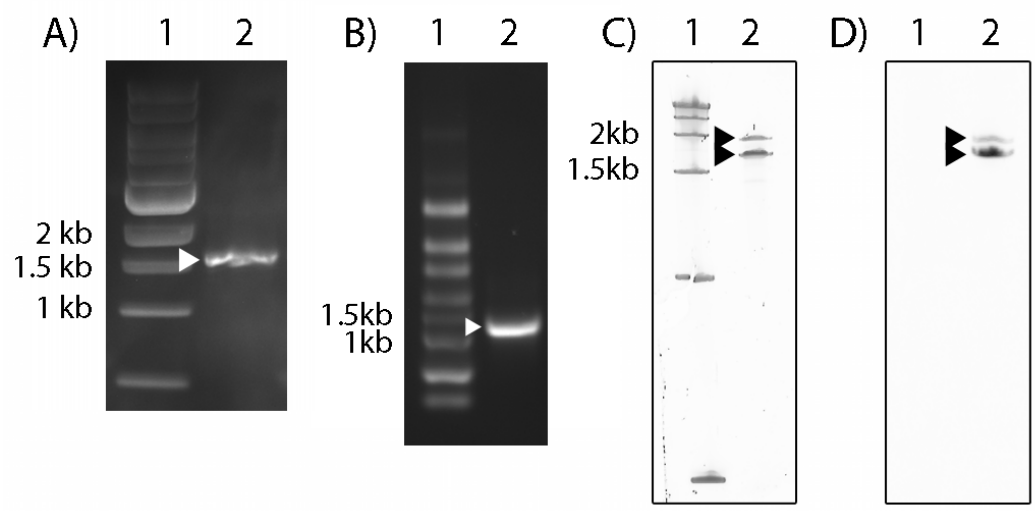

Figure C.1: *In vitro* transcription and purification of mRNA. (A) AGE of DNA template for IVT reaction amplified from pEXP5-NT/6xHis mCherry F30-2xdBroccoli plasmid using primers pEXP5NT Insert Fwd and Rev (2 % agarose, 150 V, 30 mins, stained with GelRed). Lane 1: 1 kb ladder, lane 2: PCR product (1534 bp). (B) Purified RNA from IVT reaction in a native agarose gel (1 % agarose, 120 V, 40 mins, stained with GelRed). Lane 1: Riboruler High Range RNA ladder, lanes 2-4: dilutions of purified RNA transcript (expected size 1224bp). (C) Purified RNA from IVT reaction in a denaturing urea polyacrylamide gel (3 % acrylamide, 8 M urea, 200 V, 45 mins, stained with SYBRgold). Lane 1: Riboruler High Range RNA ladder, lane 2: purified RNA transcript showing two separate bands (black arrows). (D) In-gel staining of purified RNA from IVT reaction in a denaturing urea polyacrylamide gel. After running the gel (3 % acrylamide, 8 M urea, 200 V, 45 mins), the gel was washed 3x with water for 5 minutes and then stained with 10 M DFHBI, 40 mM HEPES (pH 7.4), 100 mM KCl, and 1 mM MgCl2 for 30 mins before imaging. Lane 1: Riboruler High Range RNA ladder, lane 2: purified RNA transcript showing two DFHBI stained bands (black arrows). Agarose gels were imaged using a Biovision Gel Doc system with 365 nm UV excitation. Polyacrylamide gels were imaged using a Typhoon 9500 Fluo and Phospho Imager (BPB1-530DF20 filter).

Appendix D

Protein expression and purification

pEXP5-NT/6xHis eGFP and pEXP5-NT/6xHis mCherry plasmids were transformed into E. coli BL21 (DE3) strains and used for protein expression and purification of eGFP and mCherry 6xHis-tagged proteins for quantitative calibration. All culture media was composed of LB with 100 μg/mL Ampicillin. For both mCherry and eGFP protein expression, overnight cultures were used to inoculate 500 mL of production cultures with 40 μM IPTG to a starting OD600 of 0.015. After incubation at 37 °C and stirring at 180 rpm for approximately 4 hours (OD600$\sim$5.0), the cultures were centrifuged at 5000 rpm (Avanti Centrifuge J26-XP with JLA-8.1000 rotor, Beckman Coulter) for 5 minutes at 4 °C to pellet the cells. The media was removed and the cells resuspended in lysis buffer (50mL lysis buffer/L culture). The resuspended cells were lysed by high-pressure homogenization by running the cell dispersion through the Emulsiflex C5-Avestin at 1500 bar two times. The lysed extracts were then clarified by centrifugation at 16000 rpm for 30 minutes and filtered through a 0.45 μm membrane (Millex-HV Low Protein Binding Durapore Membrane (PVDF), 33 mm diameter filter unit, 0.45 μm pore size) to collect only the soluble protein fraction. Ni-Sepharose resin HisTrap columns (HisTrap FF 5 mL, Cytiva) were washed with water and pre-equilibriated with 25 mL of the same lysis buffer. The clarified extracts containing the His-tagged proteins were loaded on the column, washed with 25 mL lysis buffer, 25 mL high salt buffer, and 25 mL lysis buffer at a rate of 5 mL/min. The proteins were eluted from the column with elution buffer at 2 mL/min. The protein was collected at 18 fractions at 2 mL each. A Bradford assay (Bio-Rad) was run on each of the 18 fractions to determine which fraction from the column contained the His-tagged protein to collect. The selected fractions were placed in a 12-14 kDa cut-off dialysis membrane (Spectra/Por 2 Dialysis Membrane Standard RC Tubing MWCO: 12-14 kDa) and dialyzed at 4 °C overnight in storage buffer. After dialysis, aliquots of the purified proteins were flash frozen in liquid nitrogen and stored at -80 °C until use. All media and buffer recipes are listed in Table D.1. Purified proteins were confirmed by size using SDS or Native PAGE (Bio-Rad TGX Stain-Free PAGE Gel, 15-well, 4-15 %) and absorbance spectra using UV-vis scanning (NanoDrop 2000) (Figure D.1). To load the gels, 5 μL of sample was mixed with 5 muL of 2X Laemmli buffer, incubated at 95 °C for 5 mins, and then 7 μL was loaded onto the gel (Bio-Rad TGX Stain-Free PAGE Gel, 15-well, 4-15 %). Samples were run at 200 V for 30 mins in 1X Tris-Glycine-sodium dodecyl sulfate (SDS) or 1X Tris-Glycine buffer.

Table D.1: Media and buffer recipes for protein expression and purification.

Buffer	**Components**	**Media**
LB media	NaCl	10 g/L
	Tryptone	10 g/L
	Yeast extract	5 g/L
	Ampicillin	100 μ/mL
Lysis buffer	PBS	2X
	Imidazole	20 mM
	DTT	1 mM
	Benzonase	0.025 mg/L
	Protease inhibitor	1 tablet/L
	$MgCl_2$	1 mM
	pH (+ HCl/NaOH)	7.5
HiSalt buffer	PBS	6X
	pH (+ HCl/NaOH)	7.5
Elution buffer	PBS	2X
	DTT	0.3 mM
	Imidazole	250 mM
Storage buffer	Tris HCl	25 mM
	NaCl	150 mM
	Glycerol	5 %
	pH (+ HCl/NaOH)	8
1X PBS buffer	NaCl	137 mM
	KCl	2.7 mM
	Na_2HPO_4	10 mM
	KH_2PO_4	2 mM
	pH (+ HCl/NaOH)	7.5

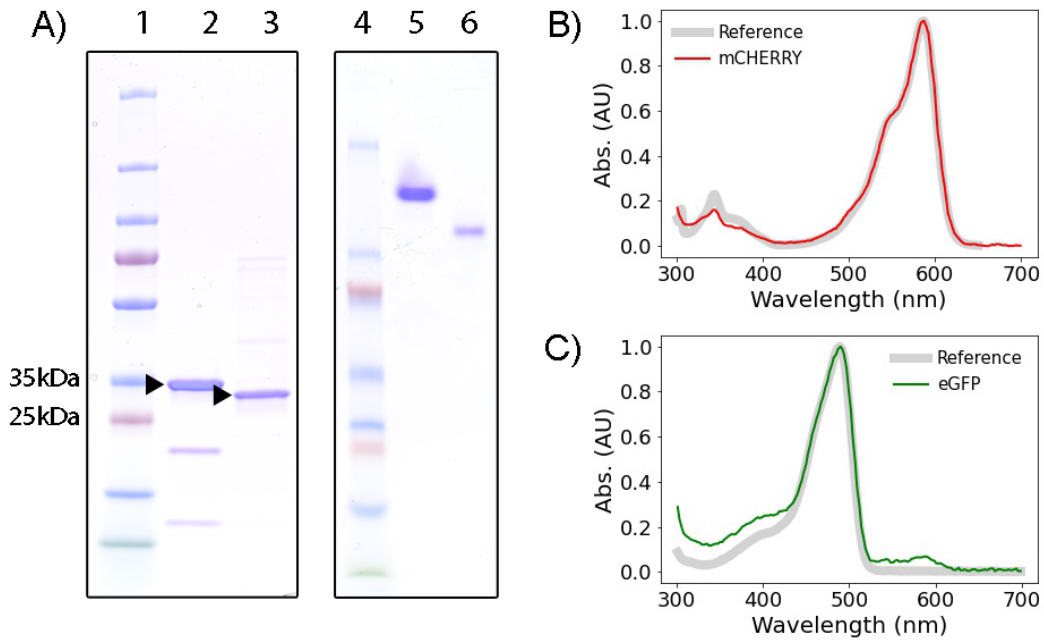

Figure D.1: Purified mCherry and eGFP proteins. (A) SDS PAGE of purified proteins. Lane 1 - PageRuler Plus protein ladder (Thermo, USA), lane 2 - purified mCherry (29.222 kDa), and lane 3 - purified eGFP (32.153 kDa). Low molecular weight bands (∼20 kDa and ∼12 kDa) visible in the mCherry lanes are self-cleaved products of mCherry as a result of sample boiling prior to SDS PAGE [5]. Target protein products in lanes 2 and 3 are marked with a black arrow. Lanes 4-6 are native non-denaturing gels of the same samples that show single bands of the proteins. Lane 1 - PageRuler Plus protein ladder (Thermo, USA), lane 2 - purified mCherry (29.222 kDa), and lane 3 - purified eGFP (32.153 kDa). Purified 6xHis mCherry (B) and 6xHis eGFP (C) absorbance spectra compared to its reference spectra (fpbase.org/protein/egfp/ and fpbase.org/protein/mCherry/).

Appendix E

mRNA and protein calibration

To obtain calibration curved for eGFP and mCherry, purified stock solutions of eGFP and mCherry proteins were quantified by Bradford assay (Quick Start Bradford Assay, Bio-Rad) with a BSA protein standard and NanoDrop A280 readings. The purified 6xHis mCherry F30-2xdBroccoli mRNA stock solution was quantified using NanoDrop RNA-40 readings. Mean values obtained from the Bradford assay were used for downstream calculations for protein concentrations and calibration (Table E.1). Serial dilutions of protein and mRNA stock solutions were prepared with a PURExpress master mix solution (Table 18) without plasmid DNA or mRNA, with and without DFHBI to ensure equivalent conditions between the gene expression reaction mixture and the calibration samples required for correct calibrations. These calibration standards were then loaded in a 384-well plate (12.5 μL/well) or loaded on a glass slide (2 μL/spot). Glass slides comprised of a parafilm border sandwiched by a glass slide (76x26x1 mm microscope slides, Marienfeld) and coverslip (24x60 mm, Menzel-Gläser) (Figure E.1). Linear calibration curves to convert relative fluorescence units (RFU) to concentration units (nM) were obtained for eGFP, mCherry, and mRNA F30-2xdBroccoli aptamer in both the plate reader Figure E.2-E.3) and confocal microscopy (Figure E.4-E.5) at the same acquisition settings for bulk and liposome-encapsulated experiments. Histograms of RFU per pixel in the ROIs of the calibration standards show increase of fluorescence distributions as a function of increasing standard concentration. Higher laser intensities were avoided to prevent bleaching effects of the CFES. Our methodologies provide an accurate method for quantification of mRNA and protein concentrations within the synthetic cells from confocal imaging. Excitation/emission wavelengths in the TECAN Spark 20M plate reader were 485/535 nm (Gain 50), 570/620 nm (Gain 70), and 450/510 nm (Gain 60) with a bandwidth of ±20 nm each, for eGFP, mCherry, and Broccoli RNA, respectively. For confocal microscopy (LSM 880 with Airyscan, 10X/0.45 Plan-Apochromat M27 objective), excitation/detection wavelengths are 488 nm/499-561 nm for RNA Broccoli and eGFP protein, and 561 nm/579-641 nm for mCherry protein. As the 488 nm laser power gradually decreased between experiments, the mRNA F30-2xdBroccoli aptamer calibration curve used for DNA titration in synthetic cell populations experiment was interpolated between two calibration curves based on the measured laser power (Table E.2).

Table E.1: eGFP protein, mCherry protein, and 6xHis mCherry F30-2xdBroccoli mRNA stock concentrations. Protein concentrations were measured by Bradford assay and A260/280 measurements on the Nanodrop (in parentheses). mRNA concentration was measured by NanoDrop RNA-40 readings.

Stock	**μM**	**A260/280**	**A260/230**
eGFP protein	79.34 ± 9.47 (79.93)	-	-
mCherry protein	226.75 ± 10.63 (287.72)	-	-
6xHis mCherry F30-2xdBroccoli mRNA	26.42	2	2.34

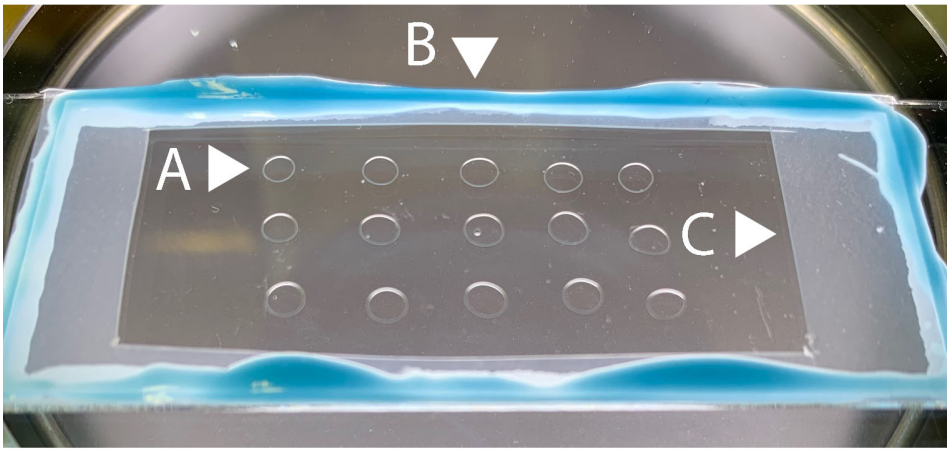

Figure E.1: Glass slide setup for confocal microscopy calibration. (A) 2 μL spots of standard solutions for calibration. (B) Twinsil Speed silicone sealing the glass slide and cover slip setup. (C) Parafilm layer between glass slide and coverslip.

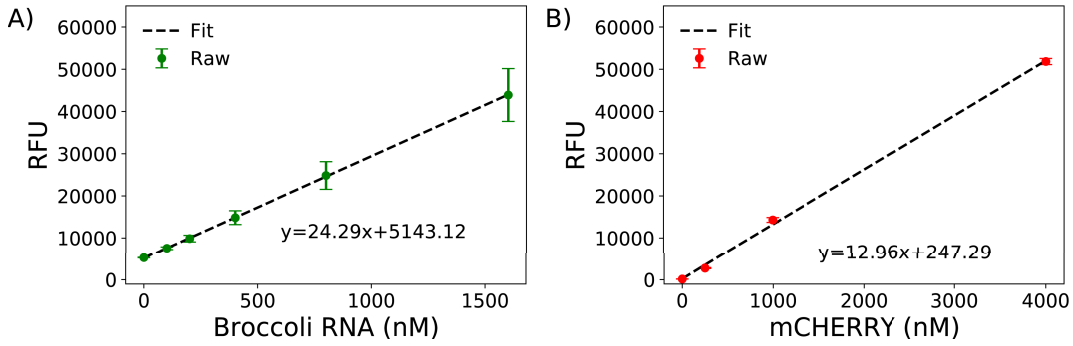

Figure E.2: Calibration of Broccoli RNA aptamer and mCherry protein in bulk plate reader format. Linear calibration curve of (A) F30-2xdBroccoli RNA aptamer and (B) mCherry protein in PURExpress master mix with 10 μM DFHBI in a bulk plate reader format. Mean and standard deviation bars are calculated from triplicate samples.

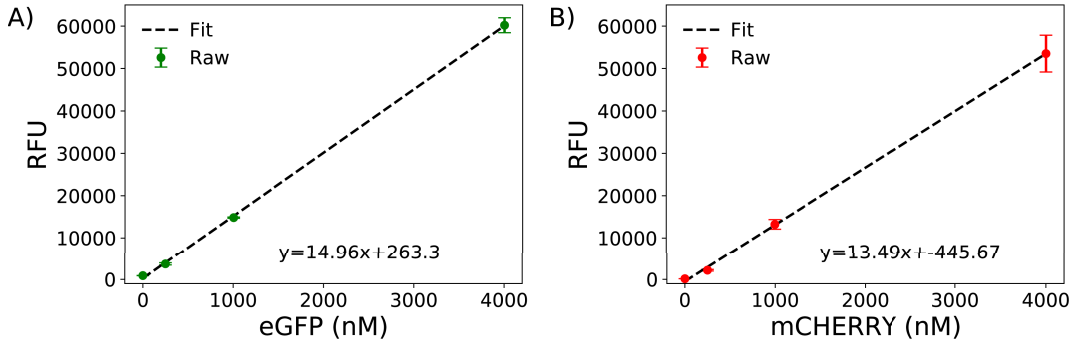

Figure E.3: Calibration of eGFP and mCherry protein in plate reader format. Linear calibration curve of (A) eGFP protein and (B) mCherry protein in PURExpress mastermix without DFHBI in a bulk plate reader format. Mean and standard deviation bars are calculated from triplicate samples.

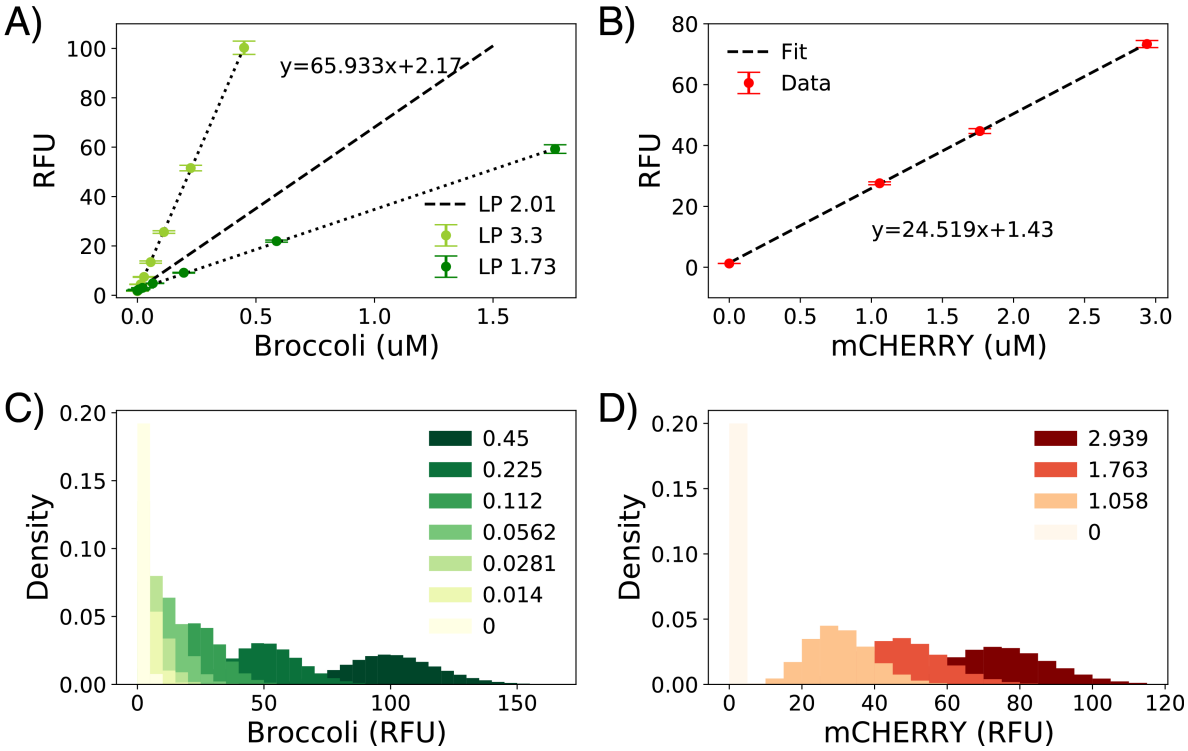

Figure E.4: Calibration curves of (A) F30-2xdBroccoli RNA aptamer and (B) mCherry protein in PURExpress master mix with 10 μM DFHBI obtained from confocal images. Dotted lines in (A) are F30-2xdBroccoli RNA aptamer calibration curves at laser powers 3.3 mW (light green) and 1.73 mW (green). The dashed line is the linear interpolated calibration curve at laser power 2.01 mW used for experiments described in Section 14. The calibration curve at 1.73 mW was used for experiments described in Section 15. Mean and standard deviation bars are calculated from the mean values of ROIs (Fig. 37). RFU distributions per pixel in the ROIs for (C) F30-2xdBroccoli RNA aptamer (LP 3.3) and (D) mCherry protein calibration standards.

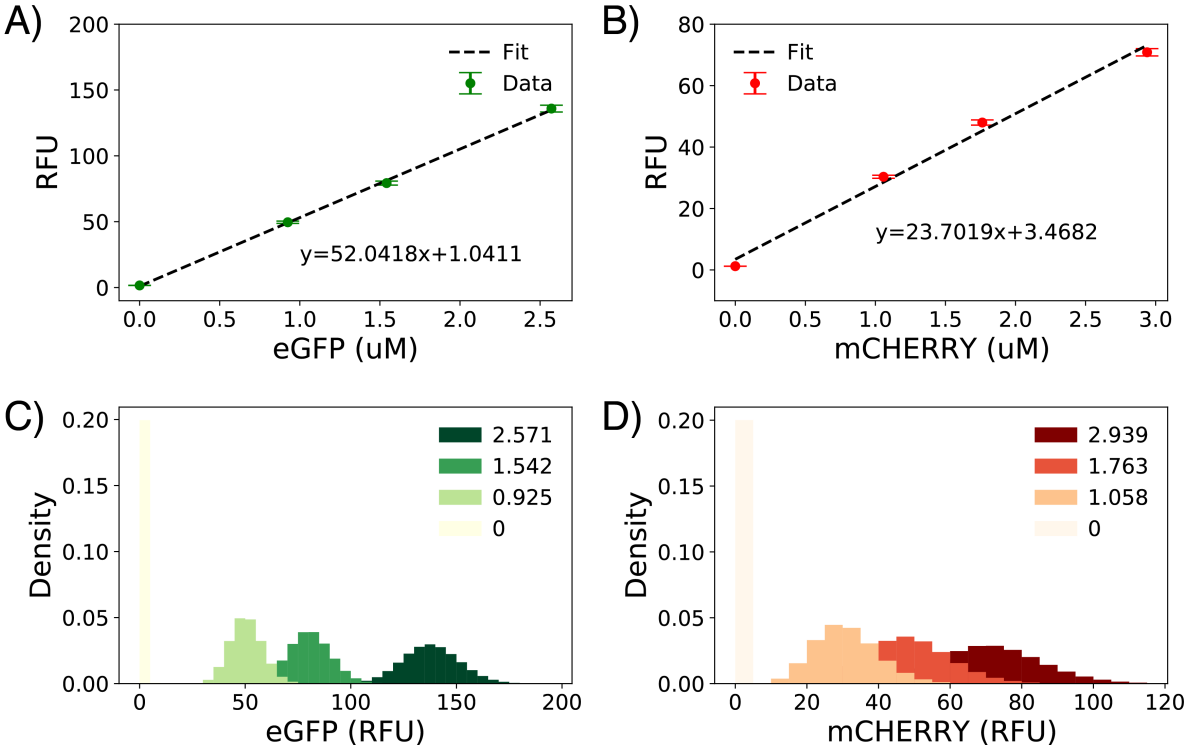

Figure E.5: Calibration curve of (A) eGFP protein and (B) mCherry protein in PURExpress mastermix without DFHBI obtained from optical confocal images. Mean and standard deviations are calculated from the mean values of ROIs (Fig. 37). RFU distributions per pixel obtained from the ROIs for (C) eGFP protein and (D) mCherry protein calibration standards show differentiation between the fluorescence intensities of the standard proteins.

Table E.2: Laser powers for quantified experiments.

Experiment	**Laser (nM)**	**Power (mW)**
RNA Broccoli	488	3.3
calibration 1	561	3.96
	633	1.61
eGFP and mCherry	488	1.73
protein calibration	561	3.22
	633	1.26
RNA Broccoli	488	1.73
calibration 2	561	3.22
	633	1.26
DNA titrations in	488	2.01
synthetic cell	561	3.16
populations	633	1.12
Two plasmids in a	488	1.73
synthetic cell	561	3.22
population	633	1.26

Appendix F

Microfluidic chip preparation and workflow

Design and fabrication

PDMS microfluidic chips were designed and fabricated as presented in [289] and [225] using soft-photolithography methods. All microfluidic chips used in this work were provided by the group of Tom Robinson (MPI-KG, Potsdam). Briefly, 4" Si-wafer master molds were prepared by spin coating (model no. WS-650MZ- 23NPPB, Laurell Tech. Corp.) SU8 2025 (Microchem Inc.) to a height of 50 μm. The SU8-coated Si-wafers were pre-baked, exposed with UV-light through a film mask with the requisite design (see supplementary CAD file in [289] for single-inlet design) for 7 s (MicroLithography Services), and then post-baked. SU8 development was then done by gently washing the wafer in developer solution (Microchem Inc.) for 3 mins before hard-baking the Si-wafers for 30 min at 200 °C. Lastly, the prepared master molds were silanized overnight (50 μL of 1H,1H,2H,2H-perfluorodecyltrichlorosilane) in a desiccator. PDMS with curing agent (10:1 mix) was degassed for 30 mins and poured over the prepared master molds. The PDMS was cured for 3 hours at 90 °C. After curing, the PDMS was peeled from the mold and inlet holes were punched using a 1mm biopsy puncher (Kai Europe GmbH) and then bonded onto glass cover slips by air plasma treatment (Plasma Cleaner PDC-002-CE, Harrick Plasma) at 600 mbar for 1 minute. After plasma treatment, the microfluidic chips were kept on a hotplate at 60 °C for 2 hours.

Pretreatment

Prior to encapsulation, the microfluidic device was pretreated with a polyelectrolyte solution to provide a hydrophilic coating in the channel between the second junction and the outlet (Figure F.1). First, a 1:3 (v/v) mixture of 37 % (w/v) HCl and 50 % (w/v) H2O2 was flushed through from the outlet towards the outer solution inlet by vacuum for 5 minutes to oxidize the channel walls. After washing with water, a 5 % (w/v) solution of positive polyelectrolyte poly(diallyldimethylammonium chloride) (PDADMAC, 1.04 g/mL, 1.375 n20/D) was flushed through for 10 minutes. After another water wash, a 2 % (w/v) solution of negative polyelectrolyte poly(4-styrenesulfonic acid) (PSS, MW ~75000) was flushed for 5 minutes to form a hydrophilic coating, and followed by a final water wash. Pretreated microfluidic chips were used on the same day of treatment.

Encapsulation of CFES in liposomes

Before encapsulation, the osmolarity of the inner and outer solutions were measured using a freezing point osmometer (Osmomat 3000, Gonotec) calibrated with water, 300 mOsmol/kg, and 2000 mOsmol/kg NaCl solution standards. The outer aqueous solution is balanced to the same osmolarity of the inner CFES solution using water or a solution of 1.5 M glucose. The lipid oil phase is composed of 1-Octanol with 6.5 mM of L-α-phosphatidylcholine (Egg PC) phospholipids, and 53.3 μM of 1,1'-Dioctadecyl-3,3,3',3'-Tetramethylindodicarbocyanine, 4-Chlorobenzenesulfonate Salt (DiD) or 8 μM N-(Texas Red sulfonyl)-1,2-dihexadecanoyl-sn-glycero-3-phosphoethanolamine, triethylammonium salt (Texas Red DHPE) (Biotium, USA) fluorescent dye. To prepare the lipid oil phase, Egg PC (300 μL of 32.5 μM in chloroform), and DiD or Texas Red DHPE dye dissolved n chloroform were mixed together in a glass test tube, dried under flowing nitrogen gas for 5 minutes and dried under vacuum for 30 minutes. The dry lipid film was resuspended in 1.5 mL 1-Octanol, and incubated at 37 °C in a sonication bath (Sonorex sonicator bath) for 1 hour. To generate synthetic cells, the inner CFES, lipid oil phase, and outer buffer solutions were dispensed through the pretreated microfluidic device using three pressure-regulated pumps (Dolomite Mitos P-Pump) at approximate pressures of 70:80:80 mbar or 0-70:50-70:80:80 mbar for the single and two-inlet microfluidic chips, respectively. This results in a flow regime where the inner CFES solution is encapsulated into double-emulsions to form liposomes. Flow regimes and double-emulsion formation in the microfluidic device were monitored under brightfield using a Zeiss Andor Axiovert 200M with a 5x/0.15 Plan-Neofluar Ph1 M27 objective and PCO Dimax S4 Monochrome sCMOS high-speed camera. The cells are collected into a microcentrifuge tube, and then placed in channels made from parafilm channels sandwiched between a microscope slide (76x26x1 mm) and cover slip (24x60 mm) (Figure F.2). The ends of the channels were sealed with Picodent Twinsil Speed silicone to avoid evaporation. Imaging was done on either a widefield or confocal microscope. A sample image of a synthetic cell population is shown in Figure F.3.

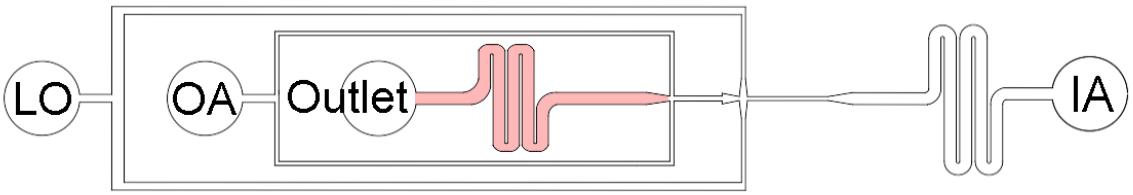

Figure F.1: Pretreatment of the microfluidic chip. Cleaning and coating solutions were sequentially flushed with water washed in-between from the outlet towards the outer aqueous solution inlet (OA) (marked by red arrows) by vacuum. This provides a hydrophilic polyelectrolyte coating in the outlet channel (pink shaded area) to prevent the double-emulsions wetting on the channel walls. IA: inner aqueous solution inlet. LO: lipid-oil phase inlet.

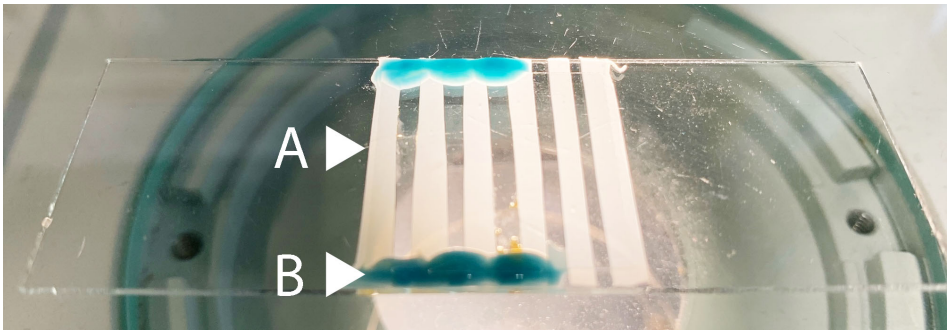

Figure F.2: Glass slide setup for synthetic cell populations. (A) Channels between parafilm strips where synthetic cells are loaded. (B) Twinsil Speed silicone sealing the ends of each channel.

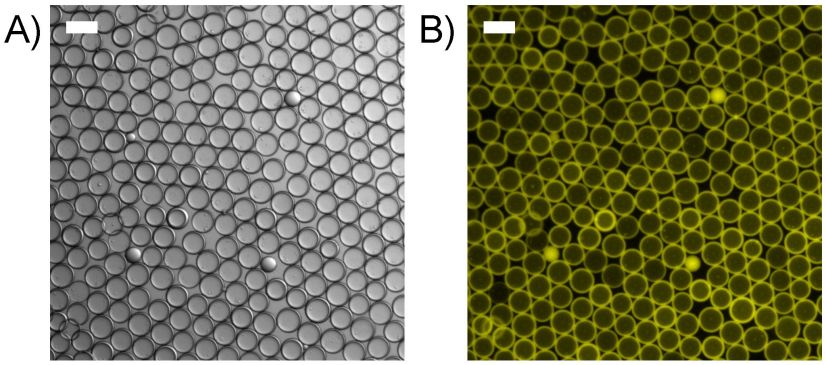

Figure F.3: Example of a microfluidic-generated synthetic population. Lipids are composed of Egg PC with DiD dye. (A) Brightfield image and (B) fluorescence image of the same population using a Zeiss Andor Axiovert 200M widefield microscope with a 5x/0.15 Plan-Neofluar Ph1 M27 objective. Fluorescence excitation was at 550 nm through a ROX filter set (excitation bandpass 575±15 nm, beam splitter HC BS 596 nm, emission BP 641±75 nm). White scale bars are 100 μm.

Appendix G

Image analysis of synthetic cell populations

Here, we briefly describe the steps for image acquisition and analysis of calibration, timeseries, and z-stack confocal microscopy data of microfluidic-generated synthetic cell populations. Image analysis was done in Python (v3.6) using the following packages: Numpy (v1.19.1), Pandas (v1.1.3), Scipy (v1.5.2), Matplotlib (v3.3.1), Pystackreg (v0.2.2) [259], Scikit-image (v0.16.2) [268], and H5py (v3.1.0).

Image acquisition

Confocal images of cell free expression encapsulated within lipid vesicles were acquired using an LSM 880 Airyscan inverted laser scanning confocal microscope with a 10X/0.45 Plan-Apochromat M27 objective. For timelapse imaging, the laser was focused at the equator of the synthetic cells and then imaged at the following excitation/emission wavelengths are 488nm/499-561 nm for RNA Broccoli and eGFP protein, 561 nm/579-641 nm for mCherry protein, and 633 nm/640-720 nm for DiD lipid dye every 5 minutes for a total of 8-12 hours. Z-stack images of the samples were taken at the endpoint focused at the top and bottom of the synthetic cells and then divided into 20 intervals.

Analysis of Timelapse images

Image stacks from timeseries confocal microscopy experiments were first registered to align the stack of each channel to its first image (Figure G.1). The last registered image on the lipid channel stack was then used as a reference image to segment cells of a synthetic cell population throughout the timeseries. A gaussian filter was used to smooth the reference image before thresholding and filling in small holes. The binary image was then labelled for each individual object and filtered by size and circularity to get only the synthetic cells. Cells with defects (*e.g.* oil drops or shadows from debris) were manually checked and removed. Cells near the edge of the image frame were also avoided to ensure only complete vesicles were analyzed (Figure G.2). Fluorescence measurements of each labeled synthetic cell were collected from a 10 μm diameter circle in the center of each cell (Figure G.3).

Z-stack images

A maximum projection of the lipid channel along the z-axis from z-stack endpoint images of the timeseries confocal microscopy experiments were used as the reference image for cell segmentation. The same workflow of gaussian filter > thresholding > filling > labelling > manual cleaning > filtering by circularity and size was then done on the reference image to segment the synthetic cells (Figure G.4). Cell sizes were obtained from the equivalent diameters of the segmented cells. Fluorescence measurements of each labeled synthetic cell was collected from the maximum value along the z-axis of a 10 μm diameter circle in the center of each cell. Single cell fluorescence measurements and sizes for each cell population were saved in csv files.

Image intensity correction (planar shading correction)

We observed an uneven and tilted illumination of the samples across the images. Low magnification objectives, such as the 20x objective which we used for our experiments, are sensitive to tilting errors because of their large field of views. Tilting was accounted for by using a planar shading correction [176]. This was done by fitting the single cell fluorescence data points or regions of interest (ROIs) in their respective coordinates to an equation of a plane and then tilting the plane back to a flat horizontal fixed at the origin. This method of correction was applied individually to all confocal images used in this study including the confocal images for calibration. To illustrate, Figure G.5 shows the uneven illumination in an image and the method of correction. Using a grid mask for bulk solution images or cell masks for synthetic cell population images (Figure G.3), RFU values of the ROIs were plot along the x and y-axes and then corrected to have an average RFU values to be constant across the plane (Figure G.6).

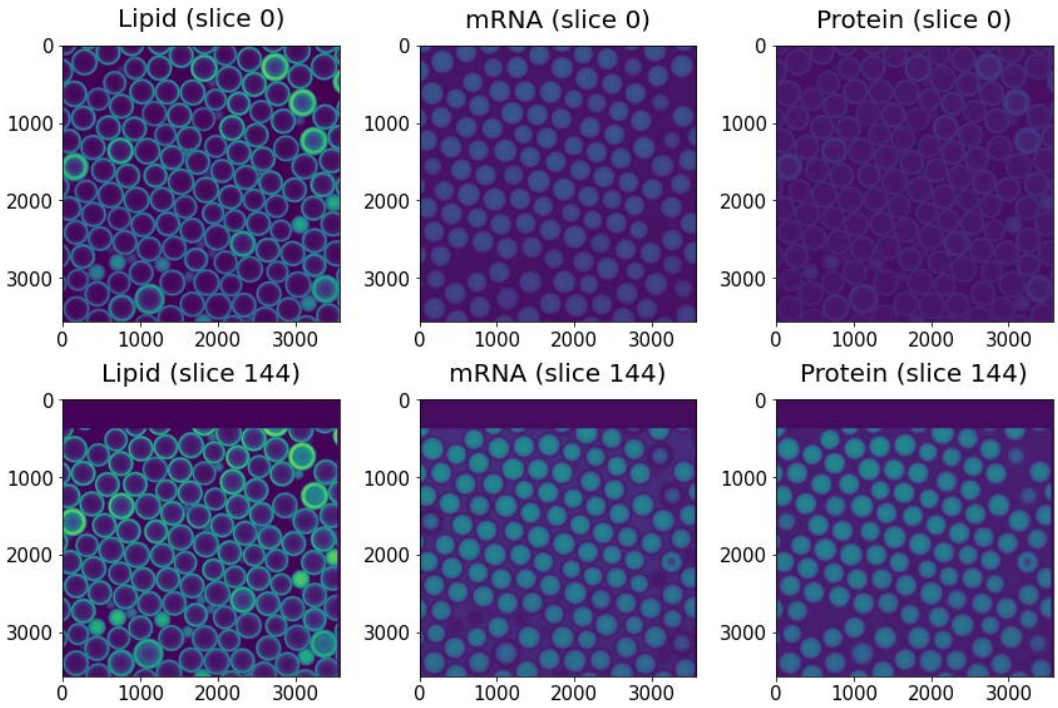

Figure G.1: Image registration of timeseries stacks. The three channels (lipid, mRNA, and protein) for the first (slice 0) and last (slice 144) image in the timeseries after registration using Pystackreg. X and Y axes are given in pixel units.

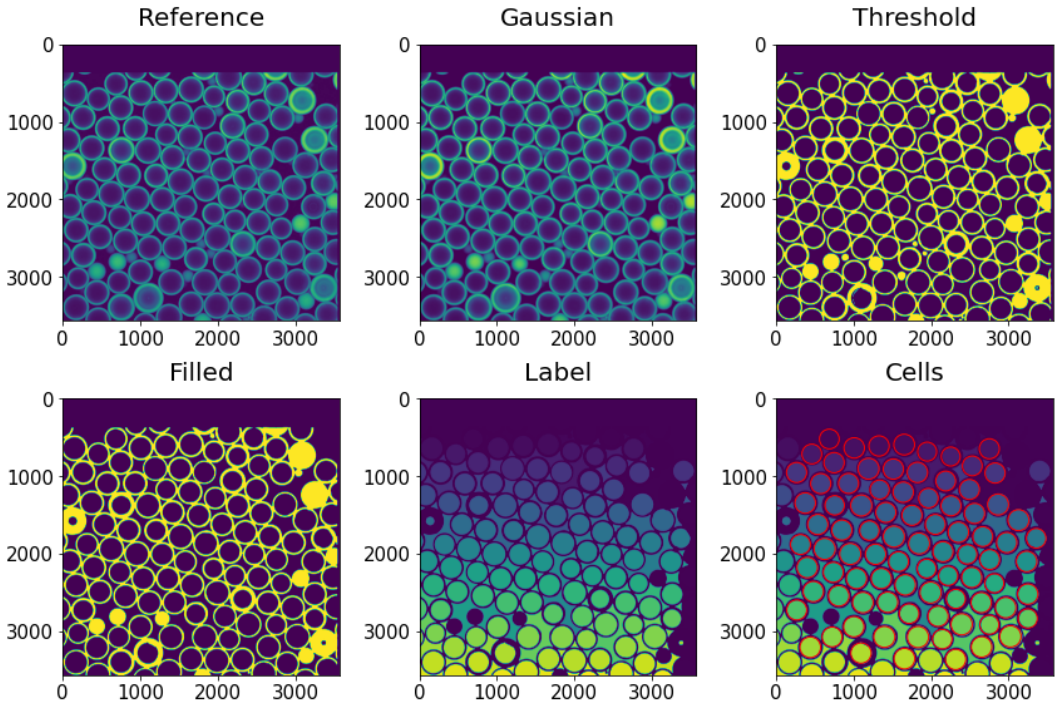

Figure G.2: Image segmentation in the lipid channel. From left to right: reference lipid image, gaussian filter, convert to binary by thresholding, filling in small holes, labelling objects, and filtering cells by shape and size (labeled in red circles).

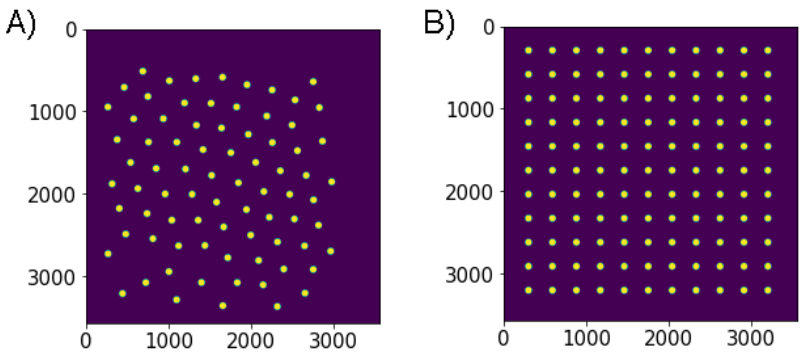

Figure G.3: Masks of a synthetic cell population and grid. ROIs were obtained by taking 10 μm diameter circles from (A) the center of the cells segmented in Figure G.2 or (B) in a uniform grid for calibration standards.

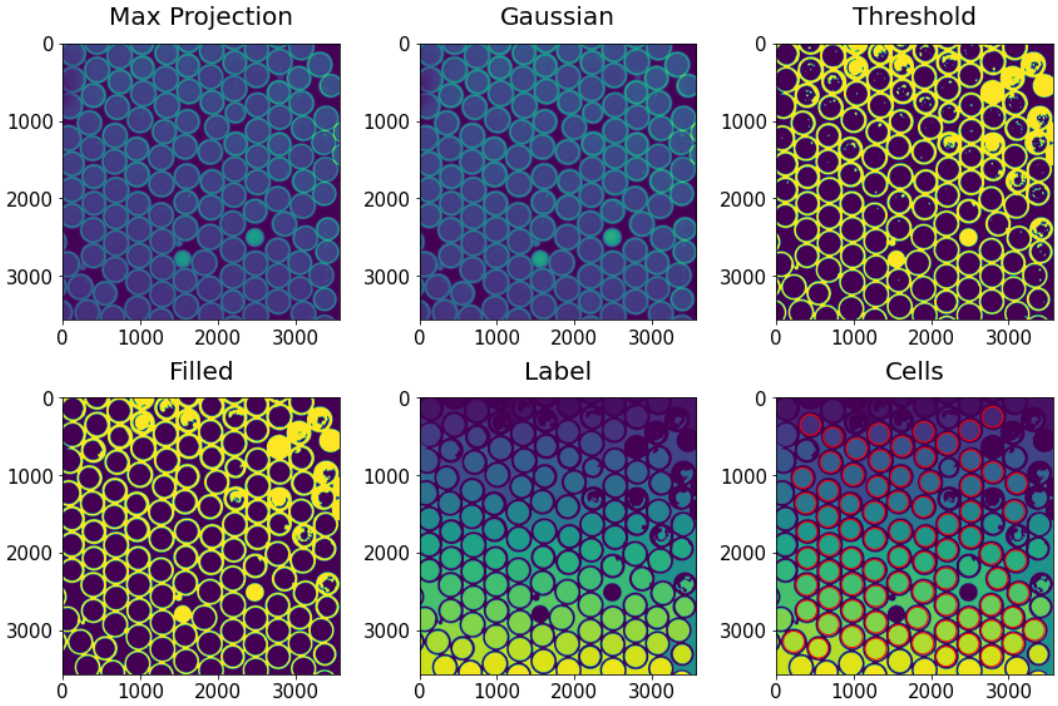

Figure G.4: Image segmentation of a z-stack. From left to right: maximum projection along z-axis, gaussian filter, convert to binary by thresholding, filling in small holes, labelling objects, and filtering cells by shape and size (labeled in red circles).

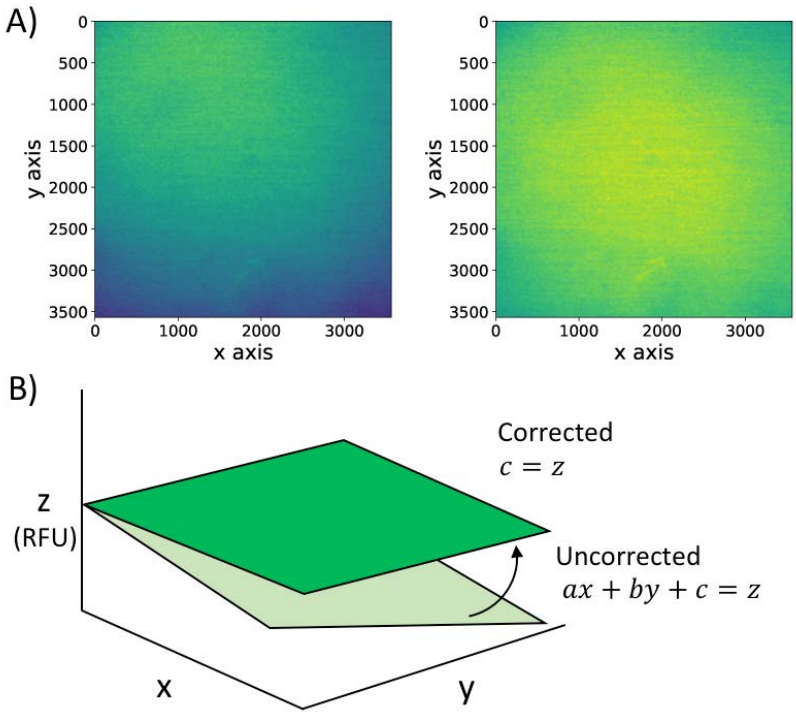

Figure G.5: Illustration of planar shading correction. (A) Example of confocal image of a bulk sample of eGFP protein with uneven and tilted illumination (left) and its planar shading corrected image (right). (B) Fluorescence values of single cells across the plane follow a gradient due to uneven and tilted illumination. This is corrected by enforcing the average fluorescence signal to be constant across the plane. To obtain the corrected data, fluorescence values were subtracted by (ax+by) from the plane fit of uncorrected data.

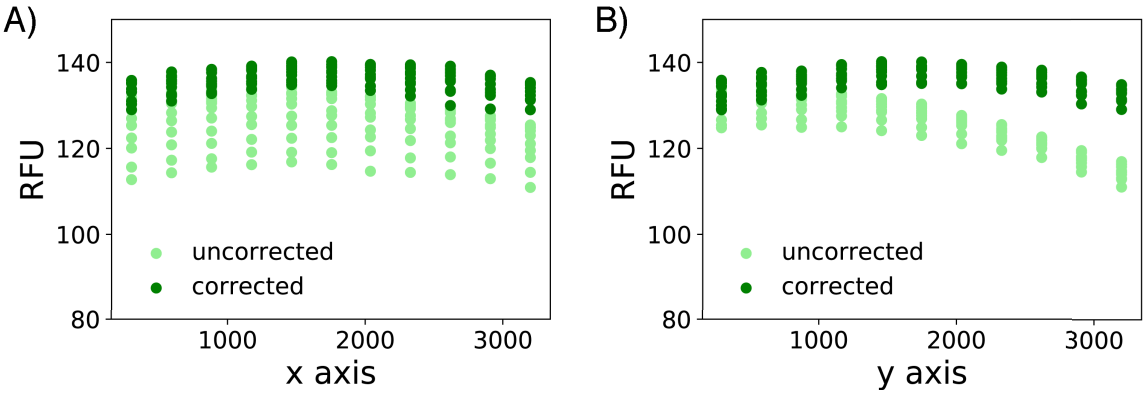

Figure G.6: Example of corrected fluorescence values. RFU values of ROIs with respect to the (A) x and (B) y-axes before and after correction from uneven illumination.

Appendix H

Resource-limited cell-free gene expression model

Seven different candidate models for cell-free gene expression (summarized in Table H.1), based on and including the resource-limited gene expression model in [247], were used to fit our bulk and encapsulated RNA Broccoli and mCherry protein expression timeseries data. The general structure of the models consists of either mass-action, Michaelis-Menten, or Hill kinetics to describe the different steps of transcription and translation to provide a coarse-grained model of gene expression. In addition, gene expression was modeled as a resource-limited process. Transcription and translation processes consume TsR and TlR species, which are transcription and translation resources, respectively. To illustrate, the resource-limited gene expression model can be described by the reaction network with their respective reaction rate constants in Equation H.1-H.8.

$$\text{DNA} + \text{RNAP} \xleftrightarrow{k_{f1}/k_{b1}} \text{DNA:RNAP} \tag{H.1}$$

$$\text{DNA:RNAP} + \text{TsR} \xrightarrow{k_{cat1}} \text{DNA} + \text{RNAP} + \text{RNA} \tag{H.2}$$

$$\text{RNA} + \text{Rb} \xleftrightarrow{k_{f2}/k_{b2}} \text{RNA:Rb} \tag{H.3}$$

$$\text{RNA:Rb} + \text{TlR} \xrightarrow{k_{cat2}} \text{RNA} + \text{Rb} + \text{Protein} \tag{H.4}$$

$$\text{Protein} \xrightarrow{k_{mat}} \text{Protein}^{*} \tag{H.5}$$

$$\text{RNA} \xrightarrow{\delta_{\text{RNA}}} \tag{H.6}$$

$$\text{TsR} \xrightarrow{\delta_{\text{TsR}}} \tag{H.7}$$

$$\text{TlR} \xrightarrow{\delta_{\text{TlR}}} \tag{H.8}$$

For transcription, DNA is first reversibly bound to RNA polymerase (RNAP) to form a DNA:RNAP complex (Equation H.1). This complex can consume transcription resources (TsR) to produce RNA (Equation H.2). In the process, DNA and RNAP dissociate from each other. Following mass-action kinetics, the rate equation for RNA production can be written as

$$\frac{d[\text{RNA}]}{dt} = k_{cat1}[\text{DNA:RNAP}][\text{TsR}]. \tag{H.9}$$

Assuming that the binding/unbinding of DNA and RNAP is fast and in quasi-equilibrium

$$k_{f1}[\text{DNA}][\text{RNAP}] = k_{b1}[\text{DNA:RNAP}]. \tag{H.10}$$

Additionally, assuming that total RNAP concentration is conserved,

$$k_{f1}[\text{DNA}]([\text{RNAP}]_0 - [\text{DNA:RNAP}]) = k_{b1}[\text{DNA:RNAP}] \tag{H.11}$$

$$[\text{DNA:RNAP}] = \frac{k_{f1}[\text{RNAP}]_0[\text{DNA}]}{k_{b1} + k_{f1}[\text{DNA}]} \tag{H.12}$$

where $[\text{RNAP}]_0$ is the initial amount of RNA polymerase in the reaction. Replacing [DNA:RNAP] in the rate equation of RNA production and grouping constants together results in

$$\frac{d[\text{RNA}]}{dt} = \frac{k_{cat1}k_{f1}[\text{RNAP}]_0[\text{TsR}][\text{DNA}]}{k_{b1} + k_{f1}[\text{DNA}]} \tag{H.13}$$

$$\frac{d[\text{RNA}]}{dt} = \frac{k_r[\text{TsR}][\text{DNA}]}{K_r + [\text{DNA}]} \tag{H.14}$$

where $k_r = k_{cat1}[\text{RNAP}]_0$ is an effective rate constant for transcription and $K_r = k_{b1}/k_{f1}$ is the dissociation constant between DNA and RNAP. A similar derivation can be done for translation of protein (Equation H.3-H.4)

$$\frac{d[\text{Protein}]}{dt} = \frac{k_p[\text{TlR}][\text{Protein}]}{K_p + [\text{Protein}]} \tag{H.15}$$

where k_p is an effective rate constant for translation and K_p is the dissociation constant between RNA and ribosome (Rb). The remaining steps for protein maturation, RNA degradation (Equation H.5), TsR degradation (Equation H.7), and TlR degradation (Equation H.8) in the reaction network are modeled as first order reactions. Altogether, the reaction network of Equation H.1-H.8 can be written as a system of ordinary differential equations

$$\frac{d[\text{DNA}]}{dt} = 0 \tag{H.16}$$

$$\frac{d\text{RNA}}{dt} = \frac{k_r[\text{TsR}][\text{DNA}]}{K_r + [\text{DNA}]} - \delta_r[\text{RNA}] \tag{H.17}$$

$$\frac{d[\text{Protein}]}{dt} = \frac{k_p[\text{TlR}][\text{RNA}]}{K_p + [\text{RNA}]} - k_{mat}[\text{Protein}] \tag{H.18}$$

$$\frac{d[\text{Protein}^*]}{dt} = k_{mat}[\text{Protein}] \tag{H.19}$$

$$\frac{d[\text{TsR}]}{dt} = -\frac{ak_r[\text{TsR}][\text{DNA}]}{K_r + [\text{DNA}]} - \delta_{\text{TsR}}[\text{TsR}] \tag{H.20}$$

$$\frac{d[\text{TlR}]}{dt} = -\frac{bk_p[\text{TlR}][\text{RNA}]}{K_p + [\text{RNA}]} - \delta_{\text{TlR}}[\text{TlR}] \tag{H.21}$$

TsR and TlR are assigned unitless quantities starting from 1 and are consumed consumed during transcription and translation, scaled by the factors a and b, respectively. Delay differential equations can be used to describe translation to account for the ~0.5

hour delay of protein expression observed in mRNA titration bulk experiments. For example, Equation H.18 can be written to include a time-delay (τ_d), as

$$\frac{d[\text{Protein}]}{dt} = \frac{k_p[\text{TlR}][\text{RNA}](t-\tau_d)}{K_p + [\text{RNA}](t-\tau_d)} - k_{mat}[\text{Protein}] \tag{H.22}$$

where $[\text{RNA}](t-\tau_d)$ is a function of RNA at time $(t-\tau_d)$.

Different combinations of mass-action or Michaelis-Menten kinetics are used for the degradation of TsR and TlR in the candidate models. However, unlike the Michaelis-Menten kinetics of transcription and translation, we do not derive TsR and TlR degradation kinetics from a particular reaction network because TsR and TlR are abstracted species of transcription and translation resources. We chose to represent their degradation dynamics as either first order or Michaelis-Menten kinetics for simplicity. For example, Model 1 is shown in Equation H.23-H.28, which includes Michaelis-Menten type kinetics for resource degradation (2nd term in the RHS of Equation H.27 and H.28). Model 7 is the model provided by Stögbauer et al. (2012) [247] without a delay differential equation for translation, while Model 5 is Model 7 with a delay differential equation for the translation step. Each parameter is described in Table H.2.

$$\frac{d[\text{DNA}]}{dt} = 0 \tag{H.23}$$

$$\frac{d\text{RNA}}{dt} = \frac{k_r[\text{TsR}][\text{DNA}]}{K_r + [\text{DNA}]} - \delta_r[\text{RNA}] \tag{H.24}$$

$$\frac{d[\text{Protein}]}{dt} = \frac{k_p[\text{TlR}][\text{RNA}]}{K_p + [\text{RNA}]} - k_{mat}[\text{Protein}] \tag{H.25}$$

$$\frac{d[\text{Protein}^*]}{dt} = k_{mat}[\text{Protein}] \tag{H.26}$$

$$\frac{d[\text{TsR}]}{dt} = -\frac{ak_r[\text{TsR}][\text{DNA}]}{K_r + [\text{DNA}]} - \frac{\delta_{\text{TsR}}[\text{TsR}]}{K_s + [\text{TsR}]} \tag{H.27}$$

$$\frac{d[\text{TlR}]}{dt} = -\frac{bk_p[\text{TlR}][\text{RNA}]}{K_p + [\text{RNA}]} - \frac{\delta_{\text{TlR}}[\text{TlR}]}{K_l + [\text{TlR}]} \tag{H.28}$$

Table H.1: Summary of seven candidate CFES models. Each model is composed of the different reaction steps modeled by mass-action kinetics (MA), Michaelis-Menten kinetics (MM), or Michaelis-Menten kinetics with delay (DM). Model 7 is the model provided in Stögbauer et al. (2012) [247]

Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Transcription	MM	MM	MM	MM	MM	MM	MM
Translation	DM	DM	DM	DM	DM	DM	MM
RNA degradation	MA	MA	MA	MA	MA	MA	MA
Protein maturation	MA	MA	MA	MA	MA	MA	MA
TsR consumption	MM	MM	MM	MM	MM	MM	MM
TlR consumption	DM	DM	DM	-	-	DM	-
TsR degradation	MM	MA	MM	MM	-	-	-
TlR degradation	MM	MM	MA	-	MM	-	MM

Table H.2: Rate parameters of resource-limited gene expression models.

Parameter	**Units**	**Description**
k_r	nM/hr	RNA transcription rate.
K_r	nM	Dissociation constant between RNAP and DNA.
δ_r	1/hr	Degradation rate of RNA.
k_p	nM/hr	Protein translation rate.
K_p	nM	Dissociation constant between ribosome and RNA.
k_{mat}	1/hr	Protein maturation rate.
δ_{TsR}	1/hr	TsR degradation rate.
K_s	-	Michaelis-Menten constant of TsR degradation.
δ_{TlR}	1/hr	TlR degradation rate.
K_l	-	Michaelis-Menten constant of TlR degradation.
a	-	Scaling factor for consumption of TsR with transcription.
b	-	Scaling factor for consumption of TlR with translation.
τ_d	hr	Time-delay of translation.

Appendix I

Parameter estimation and model selection

All methods for parameter estimation, model selection, and profile likelihood analysis were written in Julia (v1.5.3) [23] using the packages DifferentialEquations, Distributions, Optim, Plots, DataFrames, DelimitedFiles, and CSV.

Model selection and profile likelihoods

To rank the seven candidate models, we calculated the Akaike information criterion (AIC) from fit parameter estimates. The AIC values of each model are shown in Table I.1. Fits of the models with bulk experiments are shown in Figure I.1-I.7. Model 7 is the same model provided by Stögbauer et al. (2012) [247], but resulted in the highest AIC (and lowest log-likelihood). This was due to the initial delay of mCherry signal in both DNA and RNA titration experiments and endpoints of mRNA in the DNA titration experiments not fitting well. We then modified this model to include a delay in translation (Models 1-6) and degradation of transcription resources (Models 1-4) to account for observed deviations from experimental data. These modifications improved the AIC and qualitative fit. Overall, Model 2 was selected to be presented in the main text because it had the lowest AIC score, which accounts for both goodness-of-fit through the log-likelihood score, as well as model complexity by a penalty term for number of parameters in the AIC. Profile likelihoods of each parameter between a range greater and less than $\hat{\theta}$ by a factor of 10 at 11 equally spaced increments in of Models 1-6 are shown in Figure I.8-I.13.

Table I.1: AIC of the CFES models based on bulk experiments.

Model	Log-likelihood	Parameters	AIC
1	-6822.78	15	13675.56
2	-6822.87	14	13673.74
3	-6887.60	14	13803.21
4	-6922.83	14	13873.65
5	-7255.11	12	14534.21
6	-7208.64	11	14439.28
7	-7354.29	11	14730.58

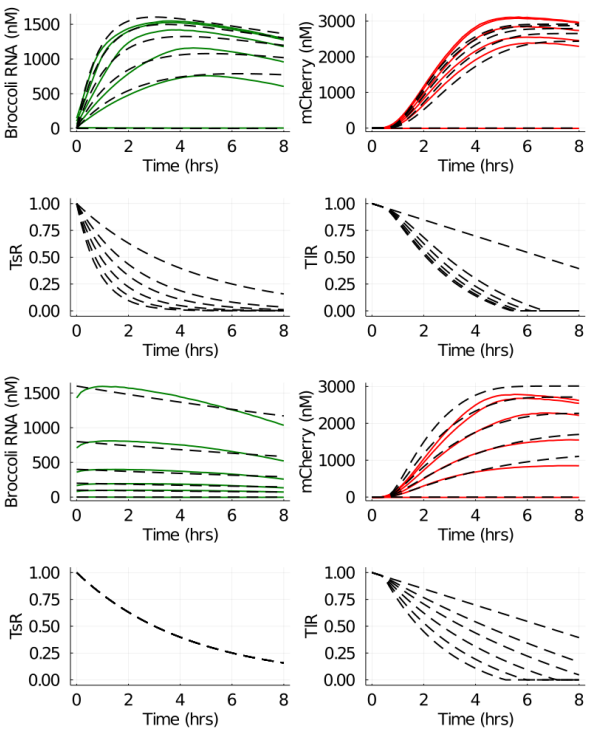

Figure I.1: Model 1 fit with DNA and RNA titration bulk experiments. Colored lines are mean timeseries experimental data for the different species RNA, Protein, TsR, and TlR. Dashed lines are the model simulation from the parameter estimates.

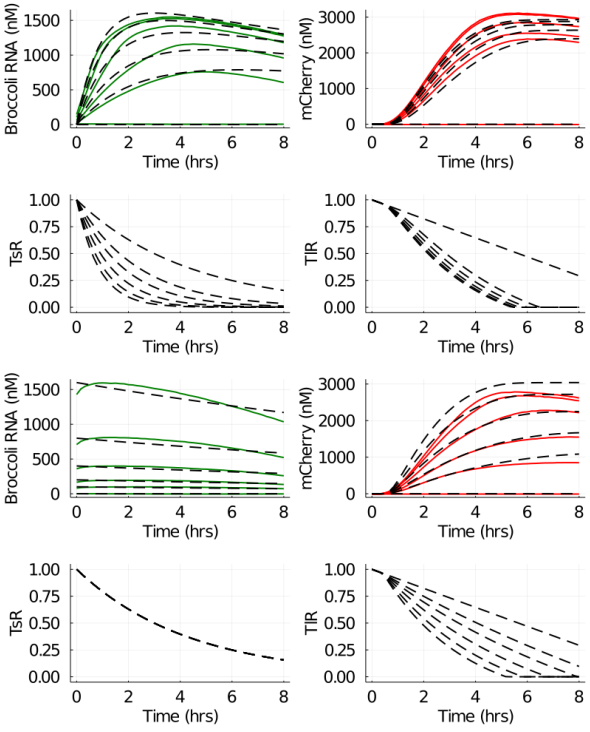

Figure I.2: Model 2 fit with DNA and RNA titration bulk experiments. Colored lines are mean timeseries experimental data for the different species RNA, Protein, TsR, and TlR. Dashed lines are the model simulation from the parameter estimates.

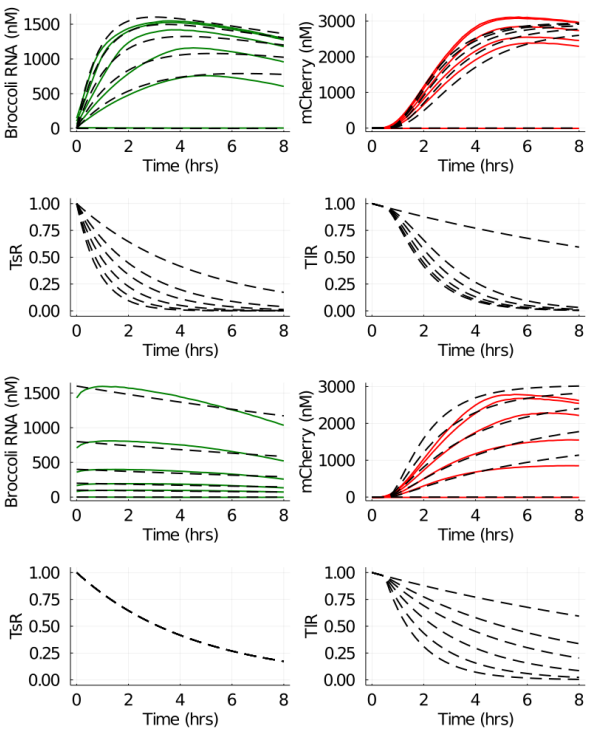

Figure I.3: Model 3 fit with DNA and RNA titration bulk experiments. Colored lines are mean timeseries experimental data for the different species RNA, Protein, TsR, and TlR. Dashed lines are the model simulation from the parameter estimates.

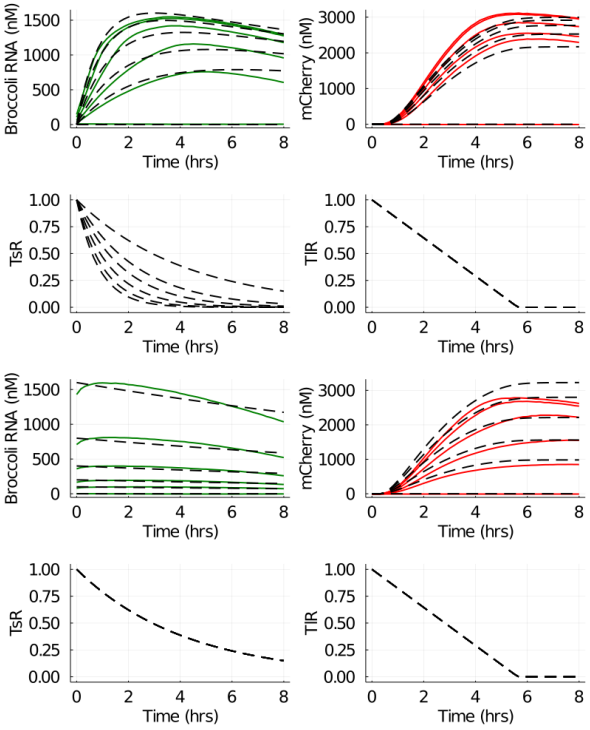

Figure I.4: Model 4 fit with DNA and RNA titration bulk experiments. Colored lines are mean timeseries experimental data for the different species RNA, Protein, TsR, and TlR. Dashed lines are the model simulation from the parameter estimates.

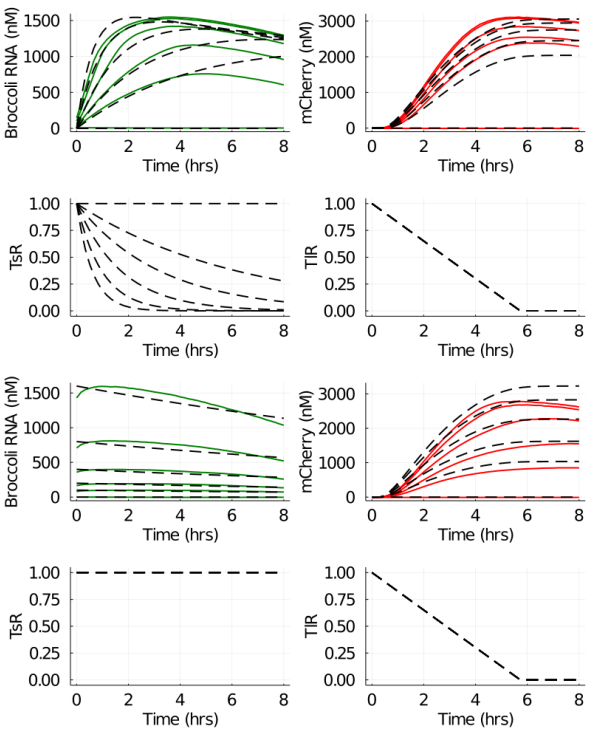

Figure I.5: Model 5 fit with DNA and RNA titration bulk experiments. Colored lines are mean timeseries experimental data for the different species RNA, Protein, TsR, and TlR. Dashed lines are the model simulation from the parameter estimates.

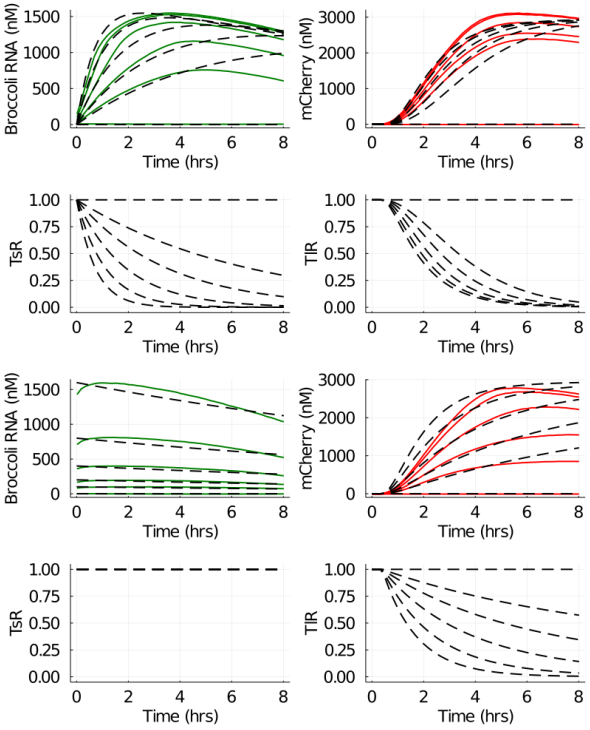

Figure I.6: Model 6 fit with DNA and RNA titration bulk experiments. Colored lines are mean timeseries experimental data for the different species RNA, Protein, TsR, and TlR. Dashed lines are the model simulation from the parameter estimates.

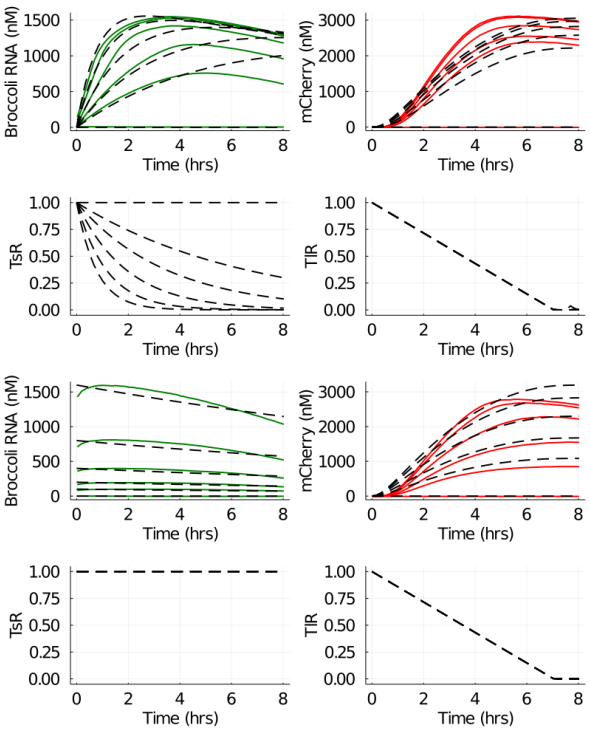

Figure I.7: Model 7 fit with DNA and RNA titration bulk experiments. Colored lines are mean timeseries experimental data for the different species RNA, Protein, TsR, and TlR. Dashed lines are the model simulation from the parameter estimates.

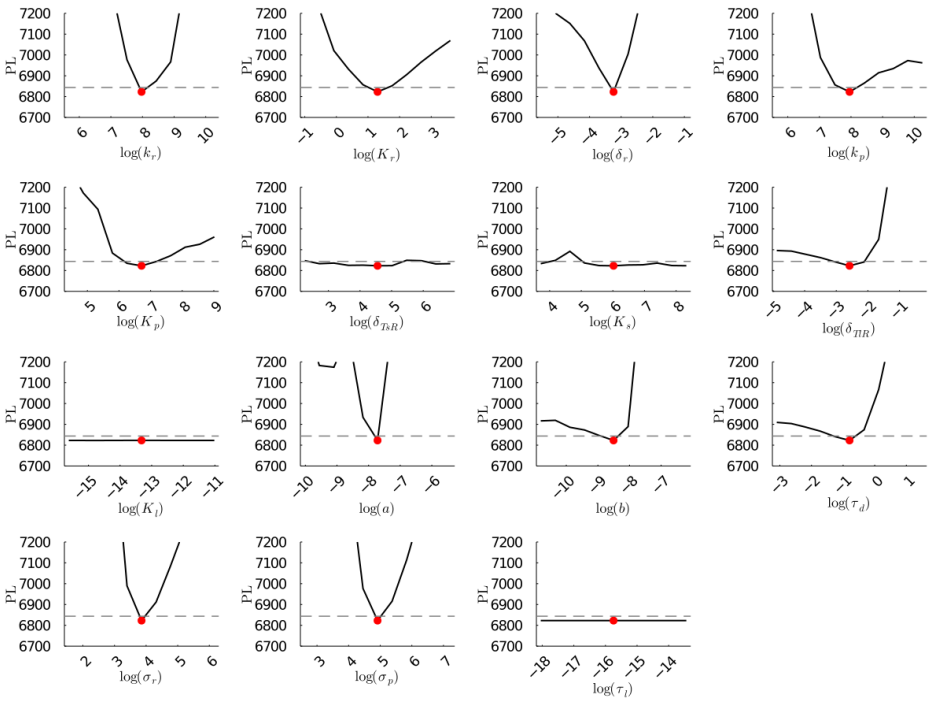

Figure I.8: Profile likelihoods of parameters from Model 1. The y-axis of each plot is the negative log likelihood of the model given the corresponding parameter value in the x-axis with all other parameters re-optimized. The red dot shows the optimized parameter set with the negative MLE. The dashed grey line is the 95% significance threshold line. All logarithms are in natural log.

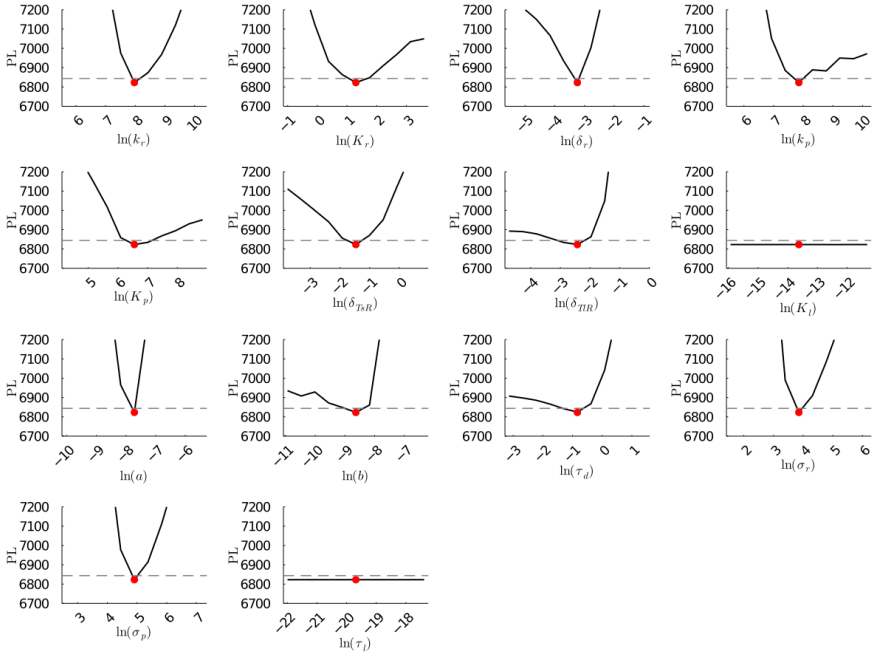

Figure I.9: Profile likelihoods of parameters from Model 2. The y-axis of each plot is the negative log likelihood of the model given the corresponding parameter value in the x-axis with all other parameters re-optimized. The red dot shows the optimized parameter set with the negative MLE. The dashed grey line is the 95% significance threshold line. All logarithms are in natural log.

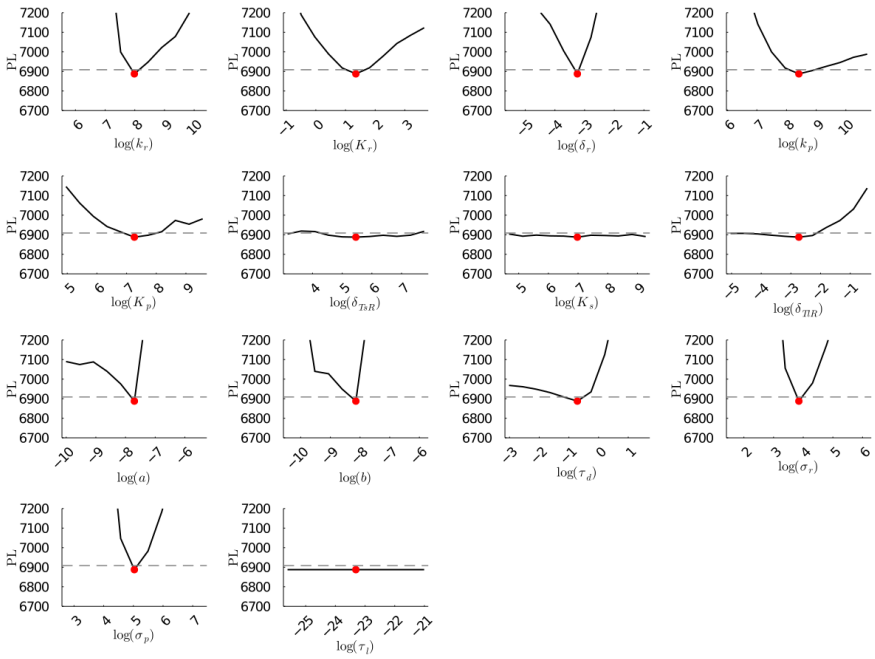

Figure I.10: Profile likelihoods of parameters from Model 3. The y-axis of each plot is the negative log likelihood of the model given the corresponding parameter value in the x-axis with all other parameters re-optimized. The red dot shows the optimized parameter set with the negative MLE. The dashed grey line is the 95% significance threshold line. All logarithms are in natural log.

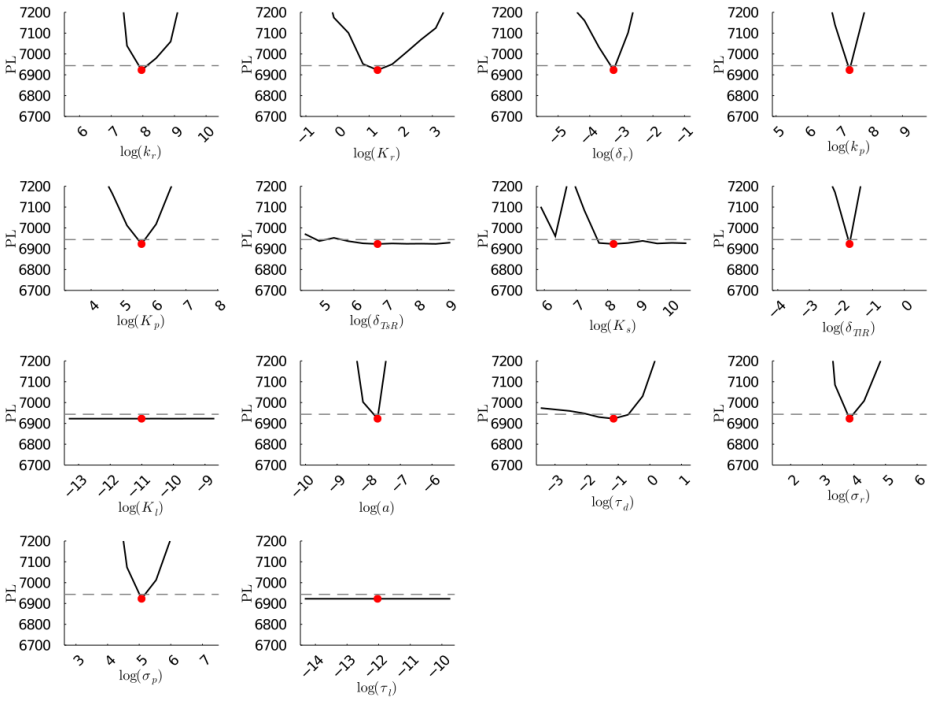

Figure I.11: Profile likelihoods of parameters from Model 4. The y-axis of each plot is the negative log likelihood of the model given the corresponding parameter value in the x-axis with all other parameters re-optimized. The red dot shows the optimized parameter set with the negative MLE. The dashed grey line is the 95% significance threshold line. All logarithms are in natural log.

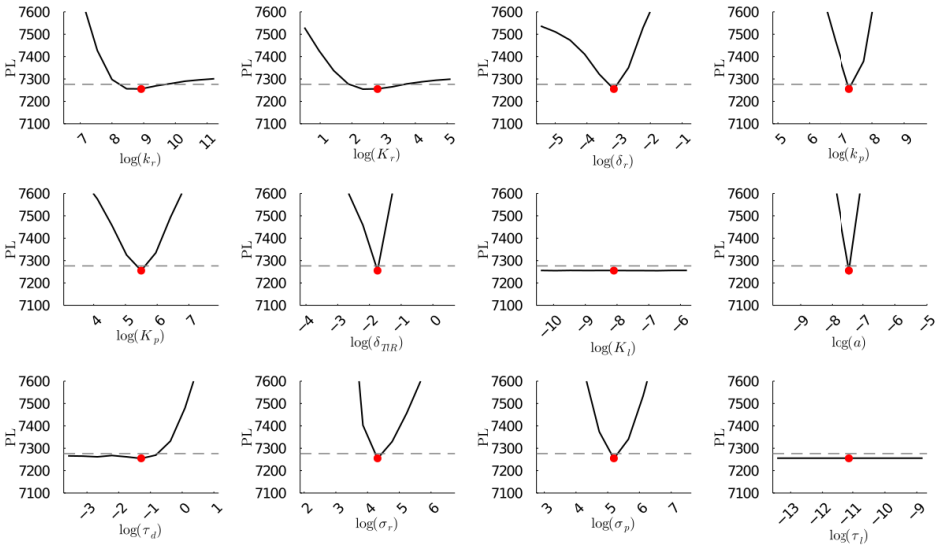

Figure I.12: Profile likelihoods of parameters from Model 5. The y-axis of each plot is the negative log likelihood of the model given the corresponding parameter value in the x-axis with all other parameters re-optimized. The red dot shows the optimized parameter set with the negative MLE. The dashed grey line is the 95% significance threshold line. All logarithms are in natural log.

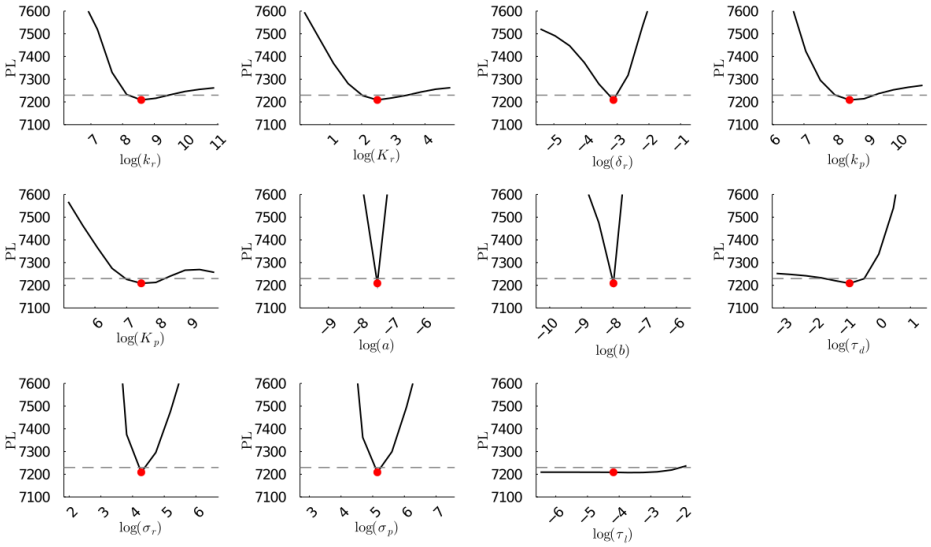

Figure I.13: Profile likelihoods of parameters from Model 6. The y-axis of each plot is the negative log likelihood of the model given the corresponding parameter value in the x-axis with all other parameters re-optimized. The red dot shows the optimized parameter set with the negative MLE. The dashed grey line is the 95% significance threshold line. All logarithms are in natural log.

List of Figures

List of Tables

List of Abbreviations

3D	three dimensional
ACP	acyl carrier protein
AHL	alpha-homoserine lactone
AIC	Akaike information criterion
bp	basepair
BP	bandpass
cDICE	continuous droplet interface crossing encapsulation
CFAI	cell-free autoinduction
CFE	cell-free expression
CFES	cell-free expression system
CI	confidence interval
CME	chemical master equation
CPME	cell population master equation
CV	coefficient of variation
DFHBI	(Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-on
DIB	droplet interface bilayer
DM	Michaelis-Menten kinetics with delay
DMEM	Dulbecco's modified eagle medium
DNA	deoxyribonucleic acid
EGF	epidermal growth factor
EM	expectation-maximization
FACS	fluorescence-activated cell sorting
FRAP	fluorescence recovery after photobleaching
FRET	fluorescence resonance energy transfer
GC-MS	gas chromatography–mass spectrometry
GFP	green fluorescent protein
GUV	giant unilamellar vesicle
HPLC	high-performance liquid chromatography
HSL	homoserine lactone
IA	inner aqueous phase
IPTG	isopropyl β-D-1-thiogalactopyranoside
IVT	*in vitro* transcription
IVTT	*in vitro* transcription and translation
LC-MS	liquid chromatography–mass spectrometry
LO	lipid-oil phase
LUV	large unilamellar vesicle

MA	moment-closure approximation
MA	mass-action
MAP	maximum a posteriori probability
MLE	maximum likelihood estimation
MM	Michaelis-Menten
MW	molecular weight
MWCO	molecular weight cut-off
NEB	New England Biolabs
NTP	nucleoside triphosphate
OA	outer aqueous phase
OEM	outer envelope membrane
OLA	octanol-assisted liposome assembly
PCR	polymerase chain reaction
PDMS	polydimethylsiloxane
PL	profile likelihood
PURE	protein synthesis using recombinant elements
PV	pairwise variation
Rb	Ribosome
REML	restricted maximum likelihood
RFU	relative fluorescence units
RNA	ribonucleic acid
RNAP	RNA polymerase
ROI	region of interest
SAEM	stochastic approximation of expectation-maximization
SAM	S-adenosyl-L-methionine
Stdev	standard deviation
SUV	small unilamellar vesicle
TlR	translation resources
TsR	transcription resources
TXTL	transcription and translation
WO	water-in-oil
WOW	water-in-oil-in-water

Bibliography

[1] K. P. Adamala, D. A. Martin-Alarcon, K. R. Guthrie-Honea, and E. S. Boyden. Engineering genetic circuit interactions within and between synthetic minimal cells. *Nature Chemistry* 9:5 (2017), pp. 431–439. DOI: 10.1038/nchem.2644.

[2] R. Ahrends, A. Ota, K. M. Kovary, T. Kudo, B. O. Park, and M. N. Teruel. Controlling low rates of cell differentiation through noise and ultrahigh feedback. *Science* 344:6190 (2014), pp. 1384–1389. DOI: 10.1126/science.1252079.

[3] H. Akaike. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* 19:6 (1974), pp. 716–723. DOI: 10.1109/TAC.1974.1100705.

[4] K. Akashi, H. Miyata, H. Itoh, and K. Kinosita. Preparation of giant liposomes in physiological conditions and their characterization under an optical microscope. *Biophysical Journal* 71:6 (1996), pp. 3242–3250. DOI: 10.1016/S0006-3495(96)79517-6.

[5] K. I. Akashi, H. Miyata, H. Itoh, and K. Kinosita. Formation of giant liposomes promotedby divalent cations: Critical role of electrostatic repulsion. *Biophysical Journal* 74:6 (1998), pp. 2973–2982. DOI: 10.1016/S0006-3495(98)78004-X.

[6] B. Alberts, A. Johnson, J. Lewis, P. Walter, M. Raff, and K. Roberts. Molecular Biology of the Cell: 6th Edition. Garland Science, 2017. ISBN: 9781317563754.

[7] A. Alcinesio, I. Cazimoglu, G. R. Kimmerly, V. Restrepo Schild, R. Krishna Kumar, and H. Bayley. Modular Synthetic Tissues from 3D-Printed Building Blocks. *Advanced Functional Materials* (2021), p. 2107773. DOI: 10.1002/adfm.202107773.

[8] A. Alcinesio, R. Krishna Kumar, and H. Bayley. Functional multivesicular structures with controlled architecture from 3D-printed droplet networks. *ChemSystemsChem* (2021), syst.202100036. DOI: 10.1002/syst.202100036.

[9] A. Alcinesio, O. J. Meacock, R. G. Allan, C. Monico, V. Restrepo Schild, I. Cazimoglu, M. T. Cornall, R. Krishna Kumar, and H. Bayley. Controlled packing and single-droplet resolution of 3D-printed functional synthetic tissues. *Nature Communications* 11:1 (2020), p. 2105. DOI: 10.1038/s41467-020-15953-y.

[10] S. Allazetta, A. Negro, and M. P. Lutolf. Microfluidic Programming of Compositional Hydrogel Landscapes. *Macromolecular Rapid Communications* 38:15 (2017), p. 1700255. DOI: 10.1002/marc.201700255.

[11] E. Amalfitano, M. Karlikow, M. Norouzi, K. Jaenes, S. Cicek, F. Masum, P. Sadat Mousavi, Y. Guo, L. Tang, A. Sydor, D. Ma, J. D. Pearson, D. Trcka, M. Pinette, A. Ambagala, S. Babiuk, B. Pickering, J. Wrana, R. Bremner, T. Mazzulli, D. Sinton, J. H. Brumell, A. A. Green, and K. Pardee. A glucose meter interface for point-of-care gene circuit-based diagnostics. *Nature Communications* 12:1 (2021), pp. 1–10. DOI: 10.1038/s41467-020-20639-6.

[12] M. I. Angelova and D. S. Dimitrov. Liposome electroformation. *Faraday Discussions of the Chemical Society* 81 (1986), p. 303. DOI: 10.1039/dc9868100303.

[13] S. K. Aoki, G. Lillacci, A. Gupta, A. Baumschlager, D. Schweingruber, and M. Khammash. A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature* 570:7762 (2019), pp. 533–537. DOI: 10.1038/s41586-019-1321-1.

[14] N. B. Arnfinnsdottir, A. V. Bjørkøy, R. Lale, and M. Sletmoen. Heterogeneity in GFP expression in isogenic populations of P. putida KT2440 investigated using flow cytometry and bacterial microarrays. *RSC Advances* 6:42 (2016), pp. 36198–36206. DOI: 10.1039/C5RA23757B.

[15] L. Aufinger and F. C. Simmel. Establishing Communication Between Artificial Cells. *Chemistry – A European Journal* 25:55 (2019), pp. 12659–12670. DOI: 10.1002/chem.201901726.

[16] A. Baccouche, K. Montagne, A. Padirac, T. Fujii, and Y. Rondelez. Dynamic DNA-toolbox reaction circuits: A walkthrough. *Methods* 67:2 (2014), pp. 234–249. DOI: 10.1016/j.ymeth.2014.01.015.

[17] S. Ballweg, E. Sezgin, M. Doktorova, R. Covino, J. Reinhard, D. Wunnicke, I. Hänelt, I. Levental, G. Hummer, and R. Ernst. Regulation of lipid saturation without sensing membrane fluidity. *Nature Communications* 11:1 (2020), p. 756. DOI: 10.1038/s41467-020-14528-1.

[18] M. K. Al-Banna, A. W. Kelman, and B. Whiting. Experimental design and efficient parameter estimation in population pharmacokinetics. *Journal of Pharmacokinetics and Biopharmaceutics* 18:4 (1990), pp. 347–360. DOI: 10.1007/BF01062273.

[19] K. Batmanov, C. Kuttler, F. Lemaire, C. Lhoussaine, and C. Versari. "Symmetry-Based Model Reduction for Approximate Stochastic Analysis". *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 7605 LNBI. 2012, pp. 49–68. DOI: 10.1007/978-3-642-33636-2_5.

[20] T. Beneyton, C. Love, M. Girault, T.-Y. D. Tang, and J.-C. Baret. High-Throughput Synthesis and Screening of Functional Coacervates Using Microfluidics. *ChemSystemsChem* 2:6 (2020), syst.202000022. DOI: 10.1002/syst.202000022.

[21] S. Berhanu, T. Ueda, and Y. Kuruma. Artificial photosynthetic cell producing energy for protein synthesis. *Nature Communications* 10:1 (2019). DOI: 10.1038/s41467-019-09147-4.

[22] K. Bernitzki and T. Schrader. Entirely Artificial Signal Transduction with a Primary Messenger. *Angewandte Chemie International Edition* 48:43 (2009), pp. 8001–8005. DOI: 10.1002/anie.200902973.

[23] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A fresh approach to numerical computing. *SIAM Review* 59:1 (2017), pp. 65–98. DOI: 10.1137/141000671.

[24] O. Biner, J. G. Fedor, Z. Yin, and J. Hirst. Bottom-Up Construction of a Minimal System for Cellular Respiration and Energy Regeneration. *ACS Synthetic Biology* (2020). DOI: 10.1021/acssynbio.0c00110.

[25] D. Blanken, P. van Nies, and C. Danelon. Quantitative imaging of gene-expressing liposomes reveals rare favorable phenotypes. *Physical Biology* 16:4 (2019), p. 045002. DOI: 10.1088/1478-3975/ab0c62.

[26] J. J. Blow and R. A. Laskey. Initiation of DNA replication in nuclei and purified DNA by a cell-free extract of Xenopus eggs. *Cell* 47:4 (1986), pp. 577–587. DOI: 10.1016/0092-8674(86)90622-7.

[27] Y. Boada, A. Vignoni, and J. Picó. Engineered Control of Genetic Variability Reveals Interplay among Quorum Sensing, Feedback Regulation, and Biochemical Noise. *ACS Synthetic Biology* 6:10 (2017), pp. 1903–1912. DOI: 10.1021/acssynbio.7b00087.

[28] M. J. Booth, V. Restrepo Schild, F. G. Downs, and H. Bayley. Functional aqueous droplet networks. *Molecular BioSystems* 13:9 (2017), pp. 1658–1691. DOI: 10.1039/C7MB00192D.

[29] M. J. Booth, V. R. Schild, A. D. Graham, S. N. Olof, and H. Bayley. Light-activated communication in synthetic tissues. *Science Advances* 2:4 (2016), pp. 1–12. DOI: 10.1126/sciadv.1600056.

[30] E. Buchner. Alkoholische Gährung ohne Hefezellen. *Berichte der deutschen chemischen Gesellschaft* 30:1 (1897), pp. 117–124. DOI: 10.1002/cber.18970300121.

[31] B. C. Buddingh, J. Elzinga, and J. C. M. van Hest. Intercellular communication between artificial cells by allosteric amplification of a molecular signal. *Nature Communications* 11:1 (2020), p. 1652. DOI: 10.1038/s41467-020-15482-8.

[32] M. Bujara, M. Schümperli, R. Pellaux, M. Heinemann, and S. Panke. Optimization of a blueprint for in vitro glycolysis by metabolic real-time analysis. *Nature Chemical Biology* 7:5 (2011), pp. 271–277. DOI: 10.1038/nchembio.541.

[33] M. Buntru, S. Vogel, H. Spiegel, and S. Schillberg. Tobacco BY-2 cell-free lysate: An alternative and highly-productive plant-based in vitro translation system. *BMC Biotechnology* 14 (2014), pp. 1–11. DOI: 10.1186/1472-6750-14-37.

[34] M. Buntru, S. Vogel, K. Stoff, H. Spiegel, and S. Schillberg. A versatile coupled cell-free transcription-translation system based on tobacco BY-2 cell lysates. *Biotechnology and Bioengineering* 112:5 (2015), pp. 867–878. DOI: 10.1002/bit.25502.

[35] D. Burgenson, C. Gurramkonda, M. Pilli, X. Ge, A. Andar, Y. Kostov, L. Tolosa, and G. Rao. Rapid recombinant protein expression in cell-free extracts from human blood. *Scientific Reports* 8:1 (2018), pp. 1–7. DOI: 10.1038/s41598-018-27846-8.

[36] F. Caschera, J. W. Lee, K. K. Y. Ho, A. P. Liu, and M. C. Jewett. Cell-free compartmentalized protein synthesis inside double emulsion templated liposomes with in vitro synthesized and assembled ribosomes. *Chem. Commun.* 52:31 (2016), pp. 5467–5469. DOI: 10.1039/C6CC00223D.

[37] F. Caschera and V. Noireaux. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription-translation system. *Biochimie* 99:1 (2014), pp. 162–168. DOI: 10.1016/j.biochi.2013.11.025.

[38] P. M. Caveney, S. E. Norred, C. W. Chin, J. B. Boreyko, B. S. Razooky, S. T. Retterer, C. P. Collier, and M. L. Simpson. Resource Sharing Controls Gene Expression Bursting. *ACS Synthetic Biology* 6:2 (2017), pp. 334–343. DOI: 10.1021/acssynbio.6b00189.

[39] T. Chakraborty, S. M. Bartelt, J. Steinkühler, R. Dimova, and S. V. Wegner. Light controlled cell-to-cell adhesion and chemical communication in minimal synthetic cells. *Chemical Communications* 55:64 (2019), pp. 9448–9451. DOI: 10.1039/C9CC04768A.

[40] D. A. Chambers and G. Zubay. The stimulatory effect of cyclic adenosine 3'5'-monophosphate on DNA-directed synthesis of beta-galactosidase in a cell-free system. *Proceedings of the National Academy of Sciences of the United States of America* 63:1 (1969), pp. 118–122. DOI: 10.1073/pnas.63.1.118.

[41] J. Chappell, K. Jensen, and P. S. Freemont. Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Research* 41:5 (2013), pp. 3471–3481. DOI: 10.1093/nar/gkt052.

[42] F. Chizzolini, M. Forlin, N. Yeh Martín, G. Berloffa, D. Cecchi, and S. S. Mansy. Cell-Free Translation Is More Variable than Transcription. *ACS Synthetic Biology* 6:4 (2017), pp. 638–647. DOI: 10.1021/acssynbio.6b00250.

[43] K. Choi, J. K. Medley, M. König, K. Stocking, L. Smith, S. Gu, and H. M. Sauro. Tellurium: An extensible python-based modeling environment for systems and synthetic biology. *Biosystems* 171:March (2018), pp. 74–79. DOI: 10.1016/j.biosystems.2018.07.006.

[44] X. Dai, M. Zhu, M. Warren, R. Balakrishnan, V. Patsalo, H. Okano, J. R. Williamson, K. Fredrick, Y. P. Wang, and T. Hwa. Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. *Nature Microbiology* 2:December 2016 (2016). DOI: 10.1038/nmicrobiol.2016.231.

[45] J. De Gier, J. Mandersloot, and L. Van Deenen. Lipid composition and permeability of liposomes. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 150:4 (1968), pp. 666–675. DOI: 10.1016/0005-2736(68)90056-4.

[46] D. Del Vecchio, A. J. Dy, and Y. Qian. Control theory meets synthetic biology. *Journal of The Royal Society Interface* 13:120 (2016), p. 20160380. DOI: 10.1098/rsif.2016.0380.

[47] M. Delattre, M. Lavielle, and M.-A. Poursat. A note on BIC in mixed-effects models. *Electronic Journal of Statistics* 8:1 (2014), pp. 456–475. DOI: 10.1214/14-EJS890.

[48] N.-N. Deng, M. Yelleswarapu, and W. T. S. Huck. Monodisperse Uni- and Multicompartment Liposomes. *Journal of the American Chemical Society* 138:24 (2016), pp. 7584–7591. DOI: 10.1021/jacs.6b02107.

[49] N.-N. Deng and W. T. S. Huck. Microfluidic Formation of Monodisperse Coacervate Organelles in Liposomes. *Angewandte Chemie* 129:33 (2017), pp. 9868–9872. DOI: 10.1002/ange.201703145.

[50] B. J. Des Soye, S. R. Davidson, M. T. Weinstock, D. G. Gibson, and M. C. Jewett. Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from Vibrio natriegens. *ACS Synthetic Biology* 7:9 (2018), pp. 2245–2255. DOI: 10.1021/acssynbio.8b00252.

[51] B. J. Des Soye, V. R. Gerbasi, P. M. Thomas, N. L. Kelleher, and M. C. Jewett. A Highly Productive, One-Pot Cell-Free Protein Synthesis Platform Based on Genomically Recoded Escherichia coli. *Cell Chemical Biology* 26:12 (2019), 1743–1754.e9. DOI: 10.1016/j.chembiol.2019.10.008.

[52] S. Deshpande, Y. Caspi, A. E. Meijering, and C. Dekker. Octanol-assisted liposome assembly on chip. *Nature Communications* 7 (2016), pp. 1–9. DOI: 10.1038/ncomms10447.

[53] S. Deshpande, W. K. Spoelstra, M. van Doorn, J. Kerssemakers, and C. Dekker. Mechanical Division of Cell-Sized Liposomes. *ACS Nano* 12:3 (2018), pp. 2560–2568. DOI: 10.1021/acsnano.7b08411.

[54] J. K. DeVries and G. Zubay. DNA-directed peptide synthesis. II. The synthesis of the alpha-fragment of the enzyme beta-galactosidase. *Proceedings of the National Academy of Sciences* 57:4 (1967), pp. 1010–1012. DOI: 10.1073/pnas.57.4.1010.

[55] A. Didovyk, T. Tonooka, L. Tsimring, and J. Hasty. Rapid and Scalable Preparation of Bacterial Lysates for Cell-Free Gene Expression. *ACS Synthetic Biology* 6:12 (2017), pp. 2198–2208. DOI: 10.1021/acssynbio.7b00253.

[56] D. Dimitrov, J. Li, M. Angelova, and R. Jain. Surface effects in preparation of cell-size liposomes. *FEBS Letters* 176:2 (1984), pp. 398–400. DOI: 10.1016/0014-5793(84)81205-3.

[57] Y. Ding, P. D. Howes, and A. J. DeMello. Recent Advances in Droplet Microfluidics. *Analytical Chemistry* 92:1 (2020), pp. 132–149. DOI: 10.1021/acs.analchem.9b05047.

[58] B. M. Discher, Y. Y. Won, D. S. Ege, J. C. Lee, F. S. Bates, D. E. Discher, and D. A. Hammer. Polymersomes: Tough vesicles made from diblock copolymers. *Science* 284:5417 (1999), pp. 1143–1146. DOI: 10.1126/science.284.5417.1143.

[59] A. Doerr, E. De Reus, P. Van Nies, M. Van Der Haar, K. Wei, J. Kattan, A. Wahl, and C. Danelon. Modelling cell-free RNA and protein synthesis with minimal systems. *Physical Biology* 16:2 (2019). DOI: 10.1088/1478-3975/aaf33d.

[60] J. L. Dopp and N. F. Reuel. Process optimization for scalable E. coli extract preparation for cell-free protein synthesis. *Biochemical Engineering Journal* 138 (2018), pp. 21–28. DOI: 10.1016/j.bej.2018.06.021.

[61] J. L. Dopp and N. F. Reuel. Simple, functional, inexpensive cell extract for in vitro prototyping of proteins with disulfide bonds. *Biochemical Engineering Journal* 164:September (2020), p. 107790. DOI: 10.1016/j.bej.2020.107790.

[62] F. G. Downs, D. J. Lunn, M. J. Booth, J. B. Sauer, W. J. Ramsay, R. G. Klemperer, C. J. Hawker, and H. Bayley. Multi-responsive hydrogel structures from patterned droplet networks. *Nature Chemistry* 12:4 (2020), pp. 363–371. DOI: 10.1038/s41557-020-0444-1.

[63] Q. M. Dudley, K. C. Anderson, and M. C. Jewett. Cell-Free Mixing of Escherichia coli Crude Extracts to Prototype and Rationally Engineer High-Titer Mevalonate Synthesis. *ACS Synthetic Biology* 5:12 (2016), pp. 1578–1588. DOI: 10.1021/acssynbio.6b00154.

[64] H. Dueck, J. Eberwine, and J. Kim. Variation is function: Are single cell differences functionally important?: Testing the hypothesis that single cell variation is required for aggregate function. *BioEssays* 38:2 (2016), pp. 172–180. DOI: 10.1002/bies.201500124.

[65] A. Dupin, L. Aufinger, I. Styazhkin, F. Rothfischer, B. Kaufmann, S. Schwarz, N. Galensowske, H. C. Schaumann, and F. C. Simmel. Synthetic cell-based materials extract positional information from morphogen gradients. *bioRxiv* (2021), pp. 1–19. DOI: https://doi.org/10.1101/2021.04.25.441320.

[66] A. Dupin and F. C. Simmel. Signalling and differentiation in emulsion-based multicompartmentalized in vitro gene circuits. *Nature Chemistry* 11:1 (2019), pp. 32–39. DOI: 10.1038/s41557-018-0174-9.

[67] L. Duso and C. Zechner. Selected-node stochastic simulation algorithm. *Journal of Chemical Physics* 148:16 (2018). DOI: 10.1063/1.5021242.

[68] R. Eagon. Pseudomonas natriegens, a marine bacterium with a generation time of less than 10 minutes. *Journal of Bacteriology* 83:4 (1962), pp. 736–737.

[69] G. Echandi and I. D. Algranati. Defective 30S ribosomal particles in a polyamine auxotroph of Escherichia coli. *Biochemical and Biophysical Research Communications* 67:3 (1975), pp. 1185–1191. DOI: 10.1016/0006-291X(75)90798-6.

[70] Y. Elani, R. V. Law, and O. Ces. Vesicle-based artificial cells as chemical microreactors with spatially segregated reaction pathways. *Nature Communications* 5:1 (2014), p. 5305. DOI: 10.1038/ncomms6305.

[71] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science* 297:5584 (2002), pp. 1183–1186. DOI: 10.1126/science.1070919.

[72] T. Ezore, T. Suzuki, S. Higashide, E. Shintani, K. Endo, S. I. Kobayashi, M. Shikata, M. Ito, K. Tanimizu, and O. Nishimura. Cell-free protein synthesis system prepared from insect cells by freeze-thawing. *Biotechnology Progress* 22:6 (2006), pp. 1570–1577. DOI: 10.1021/bp060110v.

[73] J. Failmezger, M. Rauter, R. Nitschel, M. Kraml, and M. Siemann-Herzberg. Cell-free protein synthesis from non-growing, stressed Escherichia coli. *Scientific Reports* 7:1 (2017), pp. 1–10. DOI: 10.1038/s41598-017-16767-7.

[74] F. Fanalista, A. Birnie, R. Maan, F. Burla, K. Charles, G. Pawlik, S. Deshpande, G. H. Koenderink, M. Dogterom, and C. Dekker. Shape and Size Control of Artificial Cells for Bottom-Up Biology. *ACS Nano* 13:5 (2019), pp. 5439–5450. DOI: 10.1021/acsnano.9b00220.

[75] M. F. Feldman, M. Wacker, M. Hernandez, P. G. Hitchen, C. L. Marolda, M. Kowarik, H. R. Morris, A. Dell, M. A. Valvano, and M. Aebi. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 102:8 (2005), pp. 3016–3021. DOI: 10.1073/pnas.0500044102.

[76] G. S. Filonov and S. R. Jaffrey. RNA Imaging with Dimeric Broccoli in Live Bacterial and Mammalian Cells. *Current protocols in chemical biology* 8:1 (2016), pp. 1–28. DOI: 10.1002/9780470559277.ch150174.

[77] G. S. Filonov, C. W. Kam, W. Song, and S. R. Jaffrey. In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. *Chemistry and Biology* 22:5 (2015), pp. 649–660. DOI: 10.1016/j.chembiol.2015.04.018.

[78] G. S. Filonov, J. D. Moon, N. Svensen, and S. R. Jaffrey. Broccoli: Rapid Selection of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and Directed Evolution. *Journal of the American Chemical Society* 136:46 (2014), pp. 16299–16308. DOI: 10.1021/ja508478x.

[79] D. Foshag, E. Henrich, E. Hiller, M. Schäfer, C. Kerger, A. Burger-Kentischer, I. Diaz-Moreno, S. M. García-Mauriño, V. Dötsch, S. Rupp, and F. Bernhard. The E. coli S30 lysate proteome: A prototype for cell-free protein production. *New Biotechnology* 40:May 2017 (2018), pp. 245–260. DOI: 10.1016/j.nbt.2017.09.005.

[80] F. Fröhlich, A. Reiser, L. Fink, D. Woschée, T. Ligon, F. J. Theis, J. O. Rädler, and J. Hasenauer. Multi-experiment nonlinear mixed effect modeling of single-cell translation kinetics after transfection. *npj Systems Biology and Applications* 4:1 (2018), p. 42. DOI: 10.1038/s41540-018-0079-7.

[81] R. S. Fuller, J. M. Kaguni, and A. Kornberg. Enzymatic replication of the origin of the Escherichia coli chromosome. *Proceedings of the National Academy of Sciences* 78:12 (1981), pp. 7370–7374. DOI: 10.1073/pnas.78.12.7370.

[82] A. Ganguly, D. Altintan, and H. Koeppl. Jump-Diffusion Approximation of Stochastic Reaction Dynamics: Error Bounds and Algorithms. *Multiscale Modeling Simulation* 13:4 (2015), pp. 1390–1419. DOI: 10.1137/140983471.

[83] M. C. Ganoza, C. Cunningham, and R. M. Green. Isolation and point of action of a factor from Escherichia coli required to reconstruct translation. *Proceedings of the National Academy of Sciences of the United States of America* 82:6 (1985), pp. 1648–1652. DOI: 10.1073/pnas.82.6.1648.

[84] K. A. Ganzinger and P. Schwille. More from less – bottom-up reconstitution of cell biology. *Journal of Cell Science* 132:4 (2019), jcs227488. DOI: 10.1242/jcs.227488.

[85] J. Garamella, R. Marshall, M. Rustad, and V. Noireaux. The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Synthetic Biology* 5:4 (2016), pp. 344–355. DOI: 10.1021/acssynbio.5b00296.

[86] D. Garenne, C. L. Beisel, and V. Noireaux. Characterization of the all-E. coli transcription-translation system myTXTL by mass spectrometry. *Rapid Communications in Mass Spectrometry* 33:11 (2019), pp. 1036–1048. DOI: 10.1002/rcm.8438.

[87] D. Garenne, S. Thompson, A. Brisson, A. Khakimzhan, and V. Noireaux. The all- E. Coli TXTL Toolbox 3.0: New Capabilities of a Cell-Free Synthetic Biology Platform. *Synthetic Biology* 5:4 (2021), pp. 344–355. DOI: 10.1093/synbio/ysab017.

[88] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* 81:25 (1977), pp. 2340–2361. DOI: 10.1021/j100540a008.

[89] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications* 188:1-3 (1992), pp. 404–425. DOI: 10.1016/0378-4371(92)90283-V.

[90] P. Girard, J. Pécréaux, G. Lenoir, P. Falson, J.-L. Rigaud, and P. Bassereau. A New Method for the Reconstitution of Membrane Proteins into Giant Unilamellar Vesicles. *Biophysical Journal* 87:1 (2004), pp. 419–429. DOI: 10.1529/biophysj.104.040360.

[91] P. Gobbo, A. J. Patil, M. Li, R. Harniman, W. H. Briscoe, and S. Mann. Programmed assembly of synthetic protocells into thermoresponsive prototissues. *Nature Materials* 17:12 (2018), pp. 1145–1153. DOI: 10.1038/s41563-018-0183-5.

[92] E. Godino, J. N. López, D. Foschepoth, C. Cleij, A. Doerr, C. F. Castellà, and C. Danelon. De novo synthesized Min proteins drive oscillatory liposome deformation and regulate FtsA-FtsZ cytoskeletal patterns. *Nature Communications* 10:1 (2019), p. 4969. DOI: 10.1038/s41467-019-12932-w.

[93] M. Golomb and M. Chamberlin. Characterization of T7 specific ribonucleic acid polymerase. IV. Resolution of the major in vitro transcripts by gel electrophoresis. *Journal of Biological Chemistry* 249:9 (1974), pp. 2858–2863. DOI: 10.1016/S0021-9258(19)42709-9.

[94] D. T. Gonzales, T.-Y. Dora Tang, and C. Zechner. Moment-based analysis of biochemical networks in a heterogeneous population of communicating cells. *2019 IEEE 58th Conference on Decision and Control (CDC)* Cdc (2020), pp. 939–944. DOI: 10.1109/cdc40024.2019.9029457.

[95] D. T. Gonzales, N. Yandrapalli, T. Robinson, C. Zechner, and T.-Y. D. Tang. Cell-Free Gene Expression Dynamics in Synthetic Cell Populations. *ACS Synthetic Biology* 11:1 (2022), pp. 205–215. DOI: 10.1021/acssynbio.1c00376.

[96] D. T. Gonzales, C. Zechner, and T.-Y. Dora Tang. Building synthetic multicellular systems using bottom–up approaches. *Current Opinion in Systems Biology* 24 (2020), pp. 56–63. DOI: 10.1016/j.coisb.2020.10.005.

[97] K. Göpfrich, B. Haller, O. Staufer, Y. Dreher, U. Mersdorf, I. Platzman, and J. P. Spatz. One-Pot Assembly of Complex Giant Unilamellar Vesicle-Based Synthetic Cells. *ACS Synthetic Biology* 8:5 (2019), pp. 937–947. DOI: 10.1021/acssynbio.9b00034.

[98] T. E. Gorochowski, S. Hauert, J.-u. Kreft, L. Marucci, N. R. Stillman, T.-y. D. Tang, L. Bandiera, V. Bartoli, D. O. R. Dixon, A. J. H. Fedorec, H. Fellermann, A. G. Fletcher, T. Foster, L. Giuggioli, A. Matyjaszkiewicz, S. McCormick, S. Montes Olivas, J. Naylor, A. Rubio Denniss, and D. Ward. Toward Engineering Biosystems With Emergent Collective Functions. *Frontiers in Bioengineering and Biotechnology* 8 (2020), p. 705. DOI: 10.3389/fbioe.2020.00705.

[99] A. Gräwe, A. Dreyer, T. Vornholt, U. Barteczko, L. Buchholz, G. Drews, U. L. Ho, M. E. Jackowski, M. Kracht, J. Lüders, T. Bleckwehl, L. Rositzka, M. Ruwe, M. Wittchen, P. Lutter, K. Müller, and J. Kalinowski. A paper-based, cell-free biosensor system for the detection of heavy metals and date rape drugs. *PLoS ONE* 14:3 (2019), pp. 1–22. DOI: 10.1371/journal.pone.0210940.

[100] J. Grodberg and J. J. Dunn. ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *Journal of Bacteriology* 170:3 (1988), pp. 1245–1253. DOI: 10.1128/jb.170.3.1245-1253.1988.

[101] R. K. Grosberg and R. R. Strathmann. The Evolution of Multicellularity: A Minor Major Transition? *Annual Review of Ecology, Evolution, and Systematics* 38:1 (2007), pp. 621–654. DOI: 10.1146/annurev.ecolsys.36.102403.114735.

[102] R. I. Gumport. Effects of Spermidine on the Rna Polymerase Reaction. *Annals of the New York Academy of Sciences* 171:3 (1970), pp. 915–938. DOI: 10.1111/j.1749-6632.1970.tb39399.x.

[103] A. Gyorgy and R. M. Murray. Quantifying resource competition and its effects in the TX-TL system. *2016 IEEE 55th Conference on Decision and Control, CDC 2016* 1:Cdc (2016), pp. 3363–3368. DOI: 10.1109/CDC.2016.7798775.

[104] A. D. Halleran and R. M. Murray. Cell-Free and in Vivo Characterization of Lux, Las, and Rpa Quorum Activation Systems in E. coli. *ACS Synthetic Biology* 7:2 (2018), pp. 752–755. DOI: 10.1021/acssynbio.7b00376.

[105] L. N. Handly, A. Pilko, and R. Wollman. Paracrine communication maximizes cellular response fidelity in wound signaling. *eLife* 4:OCTOBER2015 (2015), pp. 1–18. DOI: 10.7554/eLife.09652.

[106] X. A. Harrison, L. Donaldson, M. E. Correa-Cano, J. Evans, D. N. Fisher, C. E. Goodwin, B. S. Robinson, D. J. Hodgson, and R. Inger. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6:5 (2018), e4794. DOI: 10.7717/peerj.4794.

[107] D. A. Harville. Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72:358 (1977), pp. 320–338. DOI: 10.1080/01621459.1977.10480998.

[108] K. Hasatani, M. Leocmach, A. J. Genot, A. Estévez-Torres, T. Fujii, and Y. Rondelez. High-throughput and long-term observation of compartmentalized biochemical oscillators. *Chemical Communications* 49:73 (2013), p. 8090. DOI: 10.1039/c3cc44323j.

[109] J. W. Hindley, D. G. Zheleva, Y. Elani, K. Charalambous, L. M. Barter, P. J. Booth, C. L. Bevan, R. V. Law, and O. Ces. Building a synthetic mechanosensitive signaling pathway in compartmentalized artificial cells. *Proceedings of the National Academy of Sciences of the United States of America* 116:34 (2019), pp. 16711–16716. DOI: 10.1073/pnas.1903500116.

[110] N. Hirano, T. Sawasaki, Y. Tozawa, Y. Endo, and K. Takai. Tolerance for random recombination of domains in prokaryotic and eukaryotic translation systems: Limited interdomain misfolding in a eukaryotic translation system. *Proteins: Structure, Function, and Bioinformatics* 64:2 (2006), pp. 343–354. DOI: 10.1002/prot.21008.

[111] M. B. Hoagland, P. C. Zamecnik, and M. L. Stephenson. Intermediate reactions in protein biosynthesis. *Biochimica et Biophysica Acta* 24:889 (1957), pp. 215–216. DOI: 10.1016/0006-3002(57)90175-0.

[112] C. E. Hodgman and M. C. Jewett. Optimized extract preparation methods and reaction conditions for improved yeast cell-free protein synthesis. *Biotechnology and Bioengineering* 110:10 (2013), pp. 2643–2654. DOI: 10.1002/bit.24942.

[113] K. S. Horger, D. J. Estes, R. Capone, and M. Mayer. Films of Agarose Enable Rapid Formation of Giant Liposomes in Solutions of Physiologic Ionic Strength. *Journal of the American Chemical Society* 131:5 (2009), pp. 1810–1819. DOI: 10.1021/ja805625u.

[114] N. Horvath, M. Vilkhovoy, J. A. Wayman, K. Calhoun, J. Swartz, and J. D. Varner. Toward a genome scale sequence specific dynamic model of cell-free protein synthesis in Escherichia coli. *Metabolic Engineering Communications* 10:November 2018 (2020), e00113. DOI: 10.1016/j.mec.2019.e00113.

[115] P. C. Hu, S. Li, and N. Malmstadt. Microfluidic fabrication of asymmetric giant lipid vesicles. *ACS Applied Materials and Interfaces* 3:5 (2011), pp. 1434–1440. DOI: 10.1021/am101191d.

[116] G. B. Hurst, K. G. Asano, C. J. Doktycz, E. J. Consoli, W. L. Doktycz, C. M. Foster, J. L. Morrell-Falvey, R. F. Standaert, and M. J. Doktycz. Proteomics-Based Tools for Evaluation of Cell-Free Protein Synthesis. *Analytical Chemistry* 89:21 (2017), pp. 11443–11451. DOI: 10.1021/acs.analchem.7b02555.

[117] K. Igarashi and K. Kashiwagi. Effects of polyamines on protein synthesis and growth of Escherichia coli. *Journal of Biological Chemistry* 293:48 (2018), pp. 18702–18709. DOI: 10.1074/jbc.TM118.003465.

[118] K. Ishikawa, K. Sato, Y. Shima, I. Urabe, and T. Yomo. Expression of a cascading genetic network within liposomes. *FEBS Letters* 576:3 (2004), pp. 387–390. DOI: 10.1016/j.febslet.2004.09.046.

[119] T. Jaroentomeechai, J. C. Stark, A. Natarajan, C. J. Glasscock, L. E. Yates, K. J. Hsu, M. Mrksich, M. C. Jewett, and M. P. Delisa. Single-pot glycoprotein biosynthesis using a cell-free transcription-translation system enriched with glycosylation machinery. *Nature Communications* 9:1 (2018), pp. 1–11. DOI: 10.1038/s41467-018-05110-x.

[120] M. C. Jewett and J. R. Swartz. Substrate replenishment extends protein synthesis with an in vitro translation system designed to mimic the cytoplasm. *Biotechnology and Bioengineering* 87:4 (2004), pp. 465–471. DOI: 10.1002/bit.20139.

[121] A. Joesaar, S. Yang, B. Bögels, A. van der Linden, P. Pieters, B. V. V. S. P. Kumar, N. Dalchau, A. Phillips, S. Mann, and T. F. A. de Greef. DNA-based communication in populations of synthetic protocells. *Nature Nanotechnology* 14:4 (2019), pp. 369–378. DOI: 10.1038/s41565-019-0399-9.

[122] A. S. Karim and M. C. Jewett. A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metabolic Engineering* 36 (2016), pp. 116–126. DOI: 10.1016/j.ymben.2016.03.002.

[123] M. Karlsson, D. L. Janzén, L. Durrieu, A. Colman-Lerner, M. C. Kjellsson, and G. Cedersund. Nonlinear mixed-effects modelling for single cell estimation: when, why, and how to use it. *BMC Systems Biology* 9:1 (2015), p. 52. DOI: 10.1186/s12918-015-0203-x.

[124] E. Karzbrun, J. Shin, R. H. Bar-Ziv, and V. Noireaux. Coarse-grained dynamics of protein synthesis in a cell-free system. *Physical Review Letters* 106:4 (2011), pp. 1–4. DOI: 10.1103/PhysRevLett.106.048104.

[125] F. Kawano, H. Suzuki, A. Furuya, and M. Sato. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nature Communications* 6:1 (2015), p. 6256. DOI: 10.1038/ncomms7256.

[126] Y. Kazayama, T. Teshima, T. Osaki, S. Takeuchi, and T. Toyota. Integrated Microfluidic System for Size-Based Selection and Trapping of Giant Vesicles. *Analytical Chemistry* 88:2 (2016), pp. 1111–1116. DOI: 10.1021/acs.analchem.5b03772.

[127] P. J. Keller, A. D. Schmidt, J. Wittbrodt, and E. H. Stelzer. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science* 322 (2008), pp. 1065–1069. DOI: 10.1126/science.1162493.

[128] R. Kelwick, A. J. Webb, J. T. MacDonald, and P. S. Freemont. Development of a Bacillus subtilis cell-free transcription-translation system for prototyping regulatory elements. *Metabolic Engineering* 38:September (2016), pp. 370–381. DOI: 10.1016/j.ymben.2016.09.008.

[129] W. Kightlinger, K. E. Duncker, A. Ramesh, A. H. Thames, A. Natarajan, J. C. Stark, A. Yang, L. Lin, M. Mrksich, M. P. DeLisa, and M. C. Jewett. A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nature Communications* 10:1 (2019). DOI: 10.1038/s41467-019-12024-9.

[130] D. M. Kim and J. R. Swartz. Prolonging cell-free protein synthesis with a novel ATP regeneration system. *Biotechnology and Bioengineering* 66:3 (1999), pp. 180–188. DOI: 10.1002/(SICI)1097-0290(1999)66:3<180::AID-BIT6>3.0.CO;2-S.

[131] D. M. Kim and J. R. Swartz. Oxalate improves protein synthesis by enhancing ATP supply in a cell-free system derived from Escherichia coli. *Biotechnology Letters* 22:19 (2000), pp. 1537–1542. DOI: 10.1023/A:1005624811710.

[132] D. M. Kim and J. R. Swartz. Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnology and Bioengineering* 74:4 (2001), pp. 309–316. DOI: 10.1002/bit.1121.

[133] J. Kim, C. E. Copeland, S. R. Padumane, and Y. C. Kwon. A crude extract preparation and optimization from a genomically engineered escherichia coli for the cell-free protein synthesis system: Practical laboratory guideline. *Methods and Protocols* 2:3 (2019), pp. 1–15. DOI: 10.3390/mps2030068.

[134] J. Kirsch, P. Siekevitz, and G. Palade. Amino Acid Incorporation in Vitro by Ribonucleoprotein Particles Detached from Guinea Pig Liver Microsomes. *Journal of Biological Chemistry* 235:5 (1960), pp. 1419–1424. DOI: 10.1016/S0021-9258(18)69423-2.

[135] A. L. Klibanov, K. Maruyama, V. P. Torchilin, and L. Huang. Amphipathic polyethyleneglycols effectively prolong the circulation time of liposomes. *FEBS Letters* 268:1 (1990), pp. 235–237. DOI: 10.1016/0014-5793(90)81016-H.

[136] D. S. Kong, T. A. Thorsen, J. Babb, S. T. Wick, J. J. Gam, R. Weiss, and P. A. Carr. Open-source, community-driven microfluidics with Metafluidics. *Nature Biotechnology* 35:6 (2017), pp. 523–529. DOI: 10.1038/nbt.3873.

[137] E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics and Data Analysis* 49:4 (2005), pp. 1020–1038. DOI: 10.1016/j.csda.2004.07.002.

[138] H. Kung, B. Redfield, B. V. Treadwell, B. Eskin, C. Spears, and H. Weissbach. DNA-directed in vitro synthesis of β-galactosidase. Studies with purified factors. *Journal of Biological Chemistry* 252:19 (1977), pp. 6889–6894. DOI: 10.1016/s0021-9258(17)39933-7.

[139] C. Kurokawa, K. Fujiwara, M. Morita, I. Kawamata, Y. Kawagishi, A. Sakai, Y. Murayama, S.-i. M. Nomura, S. Murata, M. Takinoue, and M. Yanagisawa. DNA cytoskeleton for stabilizing artificial cells. *Proceedings of the National Academy of Sciences* 114:28 (2017), pp. 7228–7233. DOI: 10.1073/pnas.1702208114.

[140] M. J. Langton, F. Keymeulen, M. Ciaccia, N. H. Williams, and C. A. Hunter. Controlled membrane translocation provides a mechanism for signal transduction and amplification. *Nature Chemistry* 9:5 (2017), pp. 426–430. DOI: 10.1038/nchem.2678.

[141] M. J. Langton, L. M. Scriven, N. H. Williams, and C. A. Hunter. Triggered Release from Lipid Bilayer Vesicles by an Artificial Transmembrane Signal Transduction System. *Journal of the American Chemical Society* 139:44 (2017), pp. 15768–15773. DOI: 10.1021/jacs.7b07747.

[142] M. J. Langton, N. H. Williams, and C. A. Hunter. Recognition-Controlled Membrane Translocation for Signal Transduction across Lipid Bilayers. *Journal of the American Chemical Society* 139:18 (2017), pp. 6461–6466. DOI: 10.1021/jacs.7b02345.

[143] N. Laohakunakorn, L. Grasemann, B. Lavickova, G. Michielin, A. Shahein, Z. Swank, and S. J. Maerkl. Bottom-Up Construction of Complex Biomolecular Systems With Cell-Free Synthetic Biology. *Frontiers in Bioengineering and Biotechnology* 8:March (2020), p. 213. DOI: 10.3389/fbioe.2020.00213.

[144] J. B. Larsen, C. Kennard, S. L. Pedersen, K. J. Jensen, M. J. Uline, N. S. Hatzakis, and D. Stamou. Membrane Curvature and Lipid Composition Synergize To Regulate N-Ras Anchor Recruitment. *Biophysical Journal* 113:6 (2017), pp. 1269–1279. DOI: 10.1016/j.bpj.2017.06.051.

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

[146] M. Lavielle and L. Aarons. What do we mean by identifiability in mixed effects models? *Journal of Pharmacokinetics and Pharmacodynamics* 43:1 (2016), pp. 111–122. DOI: 10.1007/s10928-015-9459-4.

[147] K. Y. Lee, S. J. Park, K. A. Lee, S. H. Kim, H. Kim, Y. Meroz, L. Mahadevan, K. H. Jung, T. K. Ahn, K. K. Parker, and K. Shin. Photosynthetic artificial organelles sustain and control ATP-dependent reactions in a protocellular system. *Nature Biotechnology* 36:6 (2018), pp. 530–535. DOI: 10.1038/nbt.4140.

[148] R. Lentini, N. Y. Martín, M. Forlin, L. Belmonte, J. Fontana, M. Cornella, L. Martini, S. Tamburini, W. E. Bentley, O. Jousson, and S. S. Mansy. Two-Way Chemical Communication between Artificial and Natural Cells. *ACS Central Science* 3:2 (2017), pp. 117–123. DOI: 10.1021/acscentsci.6b00330.

[149] M. Z. Levine, N. E. Gregorio, M. C. Jewett, K. R. Watts, and J. P. Oza. Escherichia coli-Based Cell-Free Protein Synthesis: Protocols for a robust, flexible, and accessible platform technology. *Journal of visualized experiments* 144 (2019), pp. 1–11. DOI: 10.3791/58882.

[150] M. Z. Levine, B. So, A. C. Mullin, R. Fanter, K. Dillard, K. R. Watts, M. R. La Frano, and J. P. Oza. Activation of Energy Metabolism through Growth Media Reformulation Enables a 24-Hour Workflow for Cell-Free Expression. *ACS Synthetic Biology* 9:10 (2020), pp. 2765–2774. DOI: 10.1021/acssynbio.0c00283.

[151] J. Li, H. Wang, Y. C. Kwon, and M. C. Jewett. Establishing a high yielding streptomyces-based cell-free protein synthesis system. *Biotechnology and Bioengineering* 114:6 (2017), pp. 1343–1353. DOI: 10.1002/bit.26253.

[152] M. Li, D. C. Green, J. L. R. Anderson, B. P. Binks, and S. Mann. In vitro gene expression and enzyme catalysis in bio-inorganic protocells. *Chemical Science* 2:9 (2011), p. 1739. DOI: 10.1039/c1sc00183c.

[153] G. Lillacci, Y. Benenson, and M. Khammash. Synthetic control systems for high performance gene expression in mammalian cells. *Nucleic Acids Research* 46:18 (2018), pp. 9855–9863. DOI: 10.1093/nar/gky795.

[154] F. G. A. Lister, B. A. F. Le Bailly, S. J. Webb, and J. Clayden. Ligand-modulated conformational switching in a fully synthetic membrane-bound receptor. *Nature Chemistry* 9:5 (2017), pp. 420–425. DOI: 10.1038/nchem.2736.

[155] T. Litschel, B. Ramm, R. Maas, M. Heymann, and P. Schwille. Beating Vesicles: Encapsulated Protein Oscillations Cause Dynamic Membrane Deformations. *Angewandte Chemie - International Edition* 57:50 (2018), pp. 16286–16290. DOI: 10.1002/anie.201808750.

[156] D. V. Liu, J. F. Zawada, and J. R. Swartz. Streamlining Escherichia Coli S30 extract preparation for economical cell-free protein synthesis. *Biotechnology Progress* 21:2 (2005), pp. 460–465. DOI: 10.1021/bp049789y.

[157] J. Liu, A. Prindle, J. Humphries, M. Gabalda-Sagarra, M. Asally, D.-y. D. Lee, S. Ly, J. Garcia-Ojalvo, and G. M. Süel. Metabolic co-dependence gives rise to collective oscillations within biofilms. *Nature* 523:7562 (2015), pp. 550–554. DOI: 10.1038/nature14660.

[158] C. Loos, K. Moeller, F. Fröhlich, T. Hucho, and J. Hasenauer. A Hierarchical, Data-Driven Approach to Modeling Single-Cell Populations Predicts Latent Causes of Cell-To-Cell Variability. *Cell Systems* 6:5 (2018), 593–603.e13. DOI: 10.1016/j.cels.2018.04.008.

[159] C. Love, J. Steinkühler, D. T. Gonzales, N. Yandrapalli, T. Robinson, R. Dimova, and T. Y. Tang. Reversible pH-Responsive Coacervate Formation in Lipid Vesicles Activates Dormant Enzymatic Reactions. *Angewandte Chemie - International Edition* 59:15 (2020), pp. 5950–5957. DOI: 10.1002/anie.201914893.

[160] D. Ma, L. Shen, K. Wu, C. W. Diehnelt, and A. A. Green. Low-cost detection of norovirus using paper-based cell-free systems and synbody-based viral enrichment. *Synthetic Biology* 3:1 (2018), pp. 1–11. DOI: 10.1093/synbio/ysy018.

[161] E. Magdalena Estirado, A. F. Mason, M. Á. Alemán García, J. C. M. van Hest, and L. Brunsveld. Supramolecular Nanoscaffolds within Cytomimetic Protocells as Signal Localization Hubs. *Journal of the American Chemical Society* 142:20 (2020), pp. 9106–9111. DOI: 10.1021/jacs.0c01732.

[162] N. Maheshri and E. K. O'Shea. Living with Noisy Genes: How Cells Function Reliably with Inherent Variability in Gene Expression. *Annual Review of Biophysics and Biomolecular Structure* 36:1 (2007), pp. 413–434. DOI: 10.1146/annurev.biophys.36.040306.132705.

[163] T. Maiwald, H. Hass, B. Steiert, J. Vanlier, R. Engesser, A. Raue, F. Kipkeew, H. H. Bock, D. Kaschek, C. Kreutz, and J. Timmer. Driving the model to its limit: Profile likelihood based model reduction. *PLoS ONE* 11:9 (2016), pp. 1–18. DOI: 10.1371/journal.pone.0162366.

[164] L. Manukyan, S. A. Montandon, A. Fofonjka, S. Smirnov, and M. C. Milinkovitch. A living mesoscopic cellular automaton made of skin scales. *Nature* 544:7649 (2017), pp. 173–179. DOI: 10.1038/nature22031.

[165] R. Marshall and V. Noireaux. Quantitative modeling of transcription and translation of an all-E. coli cell-free system. *Scientific Reports* 9:1 (2019), pp. 1–12. DOI: 10.1038/s41598-019-48468-8.

[166] R. W. Martin, B. J. Des Soye, Y. C. Kwon, J. Kay, R. G. Davis, P. M. Thomas, N. I. Majewska, C. X. Chen, R. D. Marcum, M. G. Weiss, A. E. Stoddart, M. Amiram, A. K. Ranji Charna, J. R. Patel, F. J. Isaacs, N. L. Kelleher, S. H. Hong, and M. C. Jewett. Cell-free protein synthesis from genomically recoded bacteria enables multisite incorporation of noncanonical amino acids. *Nature Communications* 9:1 (2018), pp. 1–9. DOI: 10.1038/s41467-018-03469-5.

[167] S. Matosevic and B. M. Paegel. Stepwise Synthesis of Giant Unilamellar Vesicles on a Microfluidic Assembly Line. *Journal of the American Chemical Society* 133:9 (2011), pp. 2798–2800. DOI: 10.1021/ja109137s.

[168] T. Matsuura, Y. Kazuta, T. Aita, J. Adachi, and T. Yomo. Quantifying epistatic interactions among the components constituting the protein translation system. *Molecular Systems Biology* 5:297 (2009), pp. 1–10. DOI: 10.1038/msb.2009.50.

[169] T. Matsuura, N. Tanimura, K. Hosoda, T. Yomo, and Y. Shimizu. Reaction dynamics analysis of a reconstituted Escherichia coli protein translation system by computational modeling. *Proceedings of the National Academy of Sciences of the United States of America* 114:8 (2017), E1336–E1344. DOI: 10.1073/pnas.1615351114.

[170] J. H. Matthaei and M. W. Nirenberg. CHARACTERISTICS AND STABILIZATION OF DNAASE-SENSITIVE PROTEIN SYNTHESIS IN E. COLI EXTRACTS. *Proceedings of the National Academy of Sciences* 47:10 (1961), pp. 1580–1588. DOI: 10.1073/pnas.47.10.1580.

[171] S. May, M. Andreasson-Ochsner, Z. Fu, Y. X. Low, D. Tan, H.-P. M. de Hoog, S. Ritz, M. Nallani, and E.-K. Sinner. In Vitro Expressed GPCR Inserted in Polymersome Membranes for Ligand-Binding Studies. *Angewandte Chemie International Edition* 52:2 (2013), pp. 749–753. DOI: 10.1002/anie.201204645.

[172] N. Michel-Reydellet, K. Woodrow, and J. Swartz. Increasing PCR Fragment Stability and Protein Yields in a Cell-Free System with Genetically Modified *Escherichia coli* Extracts. *Journal of Molecular Microbiology and Biotechnology* 9:1 (2005), pp. 26–34. DOI: 10.1159/000088143.

[173] A. M. Miguez, M. P. McNerney, and M. P. Styczynski. Metabolic Profiling of Escherichia coli-Based Cell-Free Expression Systems for Process Optimization. *Industrial and Engineering Chemistry Research* 58:50 (2019), pp. 22472–22482. DOI: 10.1021/acs.iecr.9b03565.

[174] S. Mikami, T. Kobayashi, M. Masutani, S. Yokoyama, and H. Imataka. A human cell-derived in vitro coupled transcription/translation system optimized for production of recombinant proteins. *Protein Expression and Purification* 62:2 (2008), pp. 190–198. DOI: 10.1016/j.pep.2008.09.002.

[175] I. M. Mladenov and M. Hadzhilazova. The Many Faces of Elastica: Biological Membranes. Springer, 2017. DOI: 10.1007/978-3-319-61244-7_3.

[176] M. K. Moëll and L. A. Donaldson. Shading Correction Methods for Digital Image Analysis of Confocal Wood Images. *IAWA Journal* 28:3 (2007), pp. 349–364. DOI: 10.1163/22941932-90001646.

[177] A. Moga, N. Yandrapalli, R. Dimova, and T. Robinson. Optimization of the Inverted Emulsion Method for High-Yield Production of Biomimetic Giant Unilamellar Vesicles. *ChemBioChem* 20:20 (2019), pp. 2674–2682. DOI: 10.1002/cbic.201900529.

[178] K. Montagne, R. Plasson, Y. Sakai, T. Fujii, and Y. Rondelez. Programming an in vitro DNA oscillator using a molecular networking strategy. *Molecular Systems Biology* 7:1 (2011), p. 466. DOI: 10.1038/msb.2010.120.

[179] L. R. Montes, A. Alonso, F. M. Goñi, and L. A. Bagatolli. Giant unilamellar vesicles electroformed from native membranes and organic lipid mixtures under physiological conditions. *Biophysical Journal* 93:10 (2007). DOI: 10.1529/biophysj.107.116228.

[180] S. J. Moore, J. T. MacDonald, S. Wienecke, A. Ishwarbhai, A. Tsipa, R. Aw, N. Kylilis, D. J. Bell, D. W. McClymont, K. Jensen, K. M. Polizzi, R. Biedendieck, and P. S. Freemont. Rapid acquisition and model-based analysis of cell-free transcription–translation reactions from nonmodel bacteria. *Proceedings of the National Academy of Sciences* 115:19 (2018), E4340–E4349. DOI: 10.1073/pnas.1715806115.

[181] A. Mori, A. Chonn, L. S. Choi, A. Israels, M. A. Monck, and P. R. Cullis. Stabilization and regulated fusion of liposomes containing a cationic lipid using amphipathic polyethyleneglycol derivatives. *Journal of Liposome Research* 8:2 (1998), pp. 195–211. DOI: 10.3109/08982109809035526.

[182] L. Morsut, K. T. Roybal, X. Xiong, R. M. Gordley, S. M. Coyle, M. Thomson, and W. A. Lim. Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* 164:4 (2016), pp. 780–791. DOI: 10.1016/j.cell.2016.01.012.

[183] D. Nathans, G. Notani, J. H. Schwartz, and N. D. Zinder. Biosynthesis of the coat protein of Coliphage f2 by *E. coli* extracts. *Proceedings of the National Academy of Sciences* 48:8 (1962), pp. 1424–1431. DOI: 10.1073/pnas.48.8.1424.

[184] H. Niederholtmeyer, C. Chaggan, and N. K. Devaraj. Communication and quorum sensing in non-living mimics of eukaryotic cells. *Nature Communications* 9:1 (2018), p. 5027. DOI: 10.1038/s41467-018-07473-7.

[185] H. Niederholtmeyer, Z. Z. Sun, Y. Hori, E. Yeung, A. Verpoorte, R. M. Murray, and S. J. Maerkl. Rapid cell-free forward engineering of novel genetic ring oscillators. *eLife* 4 (2015), pp. 1–18. DOI: 10.7554/eLife.09771.

[186] H. Niederholtmeyer, L. Xu, and S. J. Maerkl. Real-time mRNA measurement during an in vitro transcription and translation reaction using binary probes. *ACS Synthetic Biology* 2:8 (2013), pp. 411–417. DOI: 10.1021/sb300104f.

[187] P. van Nies, Z. Nourian, M. Kok, R. van Wijk, J. Moeskops, I. Westerlaken, J. M. Poolman, R. Eelkema, J. H. van Esch, Y. Kuruma, T. Ueda, and C. Danelon. Unbiased tracking of the progression of mRNA and protein synthesis in bulk and in liposome-confined reactions. *ChemBioChem* 14:15 (2013), pp. 1963–1966. DOI: 10.1002/cbic.201300449.

[188] A. Nieß, J. Failmezger, M. Kuschel, M. Siemann-Herzberg, and R. Takors. Experimentally Validated Model Enables Debottlenecking of in Vitro Protein Synthesis and Identifies a Control Shift under in Vivo Conditions. *ACS Synthetic Biology* 6:10 (2017), pp. 1913–1921. DOI: 10.1021/acssynbio.7b00117.

[189] M. W. Nirenberg and J. H. Matthaei. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. *Proceedings of the National Academy of Sciences* 47:10 (1961), pp. 1588–1602. DOI: 10.1073/pnas.47.10.1588.

[190] K. Nishimura, S. Tsuru, H. Suzuki, and T. Yomo. Stochasticity in Gene Expression in a Cell-Sized Compartment. *ACS Synthetic Biology* 4:5 (2015), pp. 566–576. DOI: 10.1021/sb500249g.

[191] T. Niwa, Y. Sasaki, E. Uemura, S. Nakamura, M. Akiyama, M. Ando, S. Sawada, S.-a. Mukai, T. Ueda, H. Taguchi, and K. Akiyoshi. Comprehensive study of liposome-assisted synthesis of membrane proteins using a reconstituted cell-free translation system. *Scientific Reports* 5:1 (2016), p. 18025. DOI: 10.1038/srep18025.

[192] V. Noireaux and A. Libchaber. A vesicle bioreactor as a step toward an artificial cell assembly. *Proceedings of the National Academy of Sciences* 101:51 (2004), pp. 17669–17674. DOI: 10.1073/pnas.0408236101.

[193] S. E. Norred, P. M. Caveney, G. Chauhan, L. K. Collier, C. P. Collier, S. M. Abel, and M. L. Simpson. Macromolecular Crowding Induces Spatial Correlations That Control Gene Expression Bursting Patterns. *ACS Synthetic Biology* 7:5 (2018), pp. 1251–1258. DOI: 10.1021/acssynbio.8b00139.

[194] E. Nudleman, D. Wall, and D. Kaiser. Cell-to-Cell Transfer of Bacterial Outer Membrane Lipoproteins. *Science* 309:5731 (2005), pp. 125–127. DOI: 10.1126/science.1112440.

[195] A. Oda, C. Watanabe, N. Aoki, and M. Yanagisawa. Liposomal adhesion via electrostatic interactions and osmotic deflation increase membrane tension and lipid diffusion coefficient. *Soft Matter* 16:18 (2020), pp. 4549–4554. DOI: 10.1039/D0SM00416B.

[196] A. A. Ollis, S. Zhang, A. C. Fisher, and M. P. Delisa. Engineered oligosaccharyltransferases with greatly relaxed acceptor-site specificity. *Nature Chemical Biology* 10:10 (2014), pp. 816–822. DOI: 10.1038/nchembio.1609.

[197] K. Pardee, A. A. Green, T. Ferrante, D. E. Cameron, A. Daleykeyser, P. Yin, and J. J. Collins. Paper-based synthetic gene networks. *Cell* 159:4 (2014), pp. 940–954. DOI: 10.1016/j.cell.2014.10.004.

[198] K. Pardee, A. A. Green, M. K. Takahashi, D. Braff, G. Lambert, J. W. Lee, T. Ferrante, D. Ma, N. Donghia, M. Fan, N. M. Daringer, I. Bosch, D. M. Dudley, D. H. O'Connor, L. Gehrke, and J. J. Collins. Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell* 165:5 (2016), pp. 1255–1266. DOI: 10.1016/j.cell.2016.04.059.

[199] S. Pautot, B. J. Frisken, and D. A. Weitz. Production of unilamellar vesicles using an inverted emulsion. *Langmuir* 19:7 (2003), pp. 2870–2879. DOI: 10.1021/la026100v.

[200] M. Y. Pavlov and M. Ehrenberg. Rate of translation of natural mRNAs in an optimized in vitro system. *Archives of Biochemistry and Biophysics* 328:1 (1996), pp. 9–16. DOI: 10.1006/abbi.1996.0136.

[201] M. Perros and T. A. Steitz. DNA Looping and Lac Repressor - CAP Interaction (Technical Comments). *Science* 274:10 (1996), pp. 1929–1932.

[202] J. Petit, I. Polenz, J.-C. Baret, S. Herminghaus, and O. Bäumchen. Vesicles-on-a-chip: A universal microfluidic platform for the assembly of liposomes and polymersomes. *The European Physical Journal E* 39:6 (2016), p. 59. DOI: 10.1140/epje/i2016-16059-8.

[203] T. Pott, H. Bouvrais, and P. Méléard. Giant unilamellar vesicle formation under physiologically relevant conditions. *Chemistry and Physics of Lipids* 154:2 (2008), pp. 115–119. DOI: 10.1016/j.chemphyslip.2008.03.008.

[204] Y. Qiao, M. Li, D. Qiu, and S. Mann. Response-Retaliation Behavior in Synthetic Protocell Communities. *Angewandte Chemie* 131:49 (2019), pp. 17922–17927. DOI: 10.1002/ange.201909313.

[205] J. Qin, S. Wei, H. Xu, L. Wang, P. Hao, B. Peng, Y. Zhang, P. Shi, and X. Zan. The Bioinspired Facile Method to Efficiently Generate Diverse Proteinosomes with pH Switchable Permeability. *Advanced Materials Interfaces* 7:14 (2020), p. 2000329. DOI: 10.1002/admi.202000329.

[206] G. Rampioni, F. D'Angelo, M. Messina, A. Zennaro, Y. Kuruma, D. Tofani, L. Leoni, and P. Stano. Synthetic cells produce a quorum sensing chemical signal perceived by Pseudomonas aeruginosa. *Chemical Communications* 54:17 (2018), pp. 2090–2093. DOI: 10.1039/C7CC09678J.

[207] J. Raspopovic, L. Marcon, L. Russo, and J. Sharpe. Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. *Science* 345:6196 (2014), pp. 566–570. DOI: 10.1126/science.1252960.

[208] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* 25:15 (2009), pp. 1923–1929. DOI: 10.1093/bioinformatics/btp358.

[209] T. M. Record, E. S. Courtenay, S. Cayley, and H. J. Guttman. Biophysical compensation mechanisms buffering E. coli protein-nucleic acid interactions against changing environments. *Trends in Biochemical Sciences* 23:5 (1998), pp. 190–194. DOI: 10.1016/S0968-0004(98)01207-9.

[210] J. P. Reeves and R. M. Dowben. Formation and properties of thin-walled phospholipid vesicles. *Journal of Cellular Physiology* 73:1 (1969), pp. 49–60. DOI: 10.1002/jcp.1040730108.

[211] W. S. Reznikoff. The lactose operon-controlling elements: a complex paradigm. *Molecular Microbiology* 6:17 (1992), pp. 2419–2422. DOI: 10.1111/j.1365-2958.1992.tb01416.x.

[212] J. Ruess and J. Lygeros. Moment-Based Methods for Parameter Inference and Experiment Design for Stochastic Biochemical Reaction Networks. *ACM Transactions on Modeling and Computer Simulation* 25:2 (2015), pp. 1–25. DOI: 10.1145/2688906.

[213] H. Saito, Y. Kato, M. Le Berre, A. Yamada, T. Inoue, K. Yosikawa, and D. Baigl. Time-resolved tracking of a minimum gene expression system reconstituted in giant liposomes. *ChemBioChem* 10:10 (2009), pp. 1640–1643. DOI: 10.1002/cbic.200900205.

[214] A. Salehi-Reyhani, O. Ces, and Y. Elani. Artificial cell mimics as simplified models for the study of cell biology. *Experimental Biology and Medicine* 242:13 (2017), pp. 1309–1317. DOI: 10.1177/1535370217711441.

[215] H. Salis and Y. Kaznessis. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *The Journal of Chemical Physics* 122:5 (2005), p. 054103. DOI: 10.1063/1.1835951.

[216] M. Schaich, D. Sobota, H. Sleath, J. Cama, and U. F. Keyser. Characterization of lipid composition and diffusivity in OLA generated vesicles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1862:9 (2020), p. 183359. DOI: 10.1016/j.bbamem.2020.183359.

[217] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: An open-source platform for biological-image analysis. *Nature Methods* 9:7 (2012), pp. 676–682. DOI: 10.1038/nmeth.2019.

[218] D. Schnoerr, G. Sanguinetti, and R. Grima. Comparison of different moment-closure approximations for stochastic chemical kinetics. *Journal of Chemical Physics* 143:18 (2015). DOI: 10.1063/1.4934990.

[219] E. Schrödinger. What is life? The physical aspect of the living cell. 1944.

[220] D. Schwarz, F. Junge, F. Durst, N. Frölich, B. Schneider, S. Reckel, S. Sobhanifar, V. Dötsch, and F. Bernhard. Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nature protocols* 2:11 (2007), pp. 2945–2957. DOI: 10.1038/nprot.2007.426.

[221] M. Schwarz-Schilling, L. Aufinger, A. Mückl, and F. C. Simmel. Chemical communication between bacteria and cell-free gene expression systems within linear chains of emulsion droplets. *Integrative Biology* 8:4 (2016), pp. 564–570. DOI: 10.1039/C5IB00301F.

[222] S. R. Scott, M. O. Din, P. Bittihn, L. Xiong, L. S. Tsimring, and J. Hasty. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis. *Nature Microbiology* 2:8 (2017), p. 17083. DOI: 10.1038/nmicrobiol.2017.83.

[223] G. Seelig, D. Soloveichik, D. Y. Zhang, and E. Winfree. Enzyme-Free Nucleic Acid Logic Circuits. *Science* 314:5805 (2006), pp. 1585–1588. DOI: 10.1126/science.1132493.

[224] N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. Giepmans, A. E. Palmer, and R. Y. Tsien. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* 22:12 (2004), pp. 1567–1572. DOI: 10.1038/nbt1037.

[225] S. C. Shetty, N. Yandrapalli, K. Pinkwart, D. Krafft, T. Vidakovic-Koch, I. Ivanov, and T. Robinson. Directed Signaling Cascades in Monodisperse Artificial Eukaryotic Cells. *ACS Nano* (2021), acsnano.1c04219. DOI: 10.1021/acsnano.1c04219.

[226] X. Shi, T. Wu, C. M. Cole, N. K. Devaraj, and S. Joseph. Optimization of ClpXP activity and protein synthesis in an E. coli extract-based cell-free expression system. *Scientific Reports* 8:1 (2018), p. 3488. DOI: 10.1038/s41598-018-21739-6.

[227] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, and T. Ueda. Cell-free translation reconstituted with purified components. *Nature Biotechnology* 19:8 (2001), pp. 751–755. DOI: 10.1038/90802.

[228] Y. Shimizu, T. Kanamori, and T. Ueda. Protein synthesis by pure translation systems. *Methods* 36:3 (2005), pp. 299–304. DOI: 10.1016/j.ymeth.2005.04.006.

[229] S. F. Shimobayashi, B. M. Mognetti, L. Parolini, D. Orsi, P. Cicuta, and L. Di Michele. Direct measurement of DNA-mediated adhesion between lipid bilayers. *Physical Chemistry Chemical Physics* 17:24 (2015), pp. 15615–15628. DOI: 10.1039/C5CP01340B.

[230] J. Shin and V. Noireaux. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *Journal of Biological Engineering* 4 (2010), pp. 2–10. DOI: 10.1186/1754-1611-4-8.

[231] H. C. Shum, D. Lee, I. Yoon, T. Kodger, and D. A. Weitz. Double Emulsion Templated Monodisperse Phospholipid Vesicles. *Langmuir* 24:15 (2008), pp. 7651–7653. DOI: 10.1021/la801833a.

[232] V. Siciliano, F. Menolascina, L. Marucci, C. Fracassi, I. Garzilli, M. N. Moretti, and D. di Bernardo. Construction and Modelling of an Inducible Positive Feedback Loop Stably Integrated in a Mammalian Cell-Line. *PLoS Computational Biology* 7:6 (2011). Ed. by U. Ohler, e1002074. DOI: 10.1371/journal.pcbi.1002074.

[233] D. Siegal-Gaskins, Z. A. Tuza, J. Kim, V. Noireaux, and R. M. Murray. Gene circuit performance characterization and resource usage in a cell-free breadboard. *ACS Synthetic Biology* 3:6 (2014), pp. 416–425. DOI: 10.1021/sb400203p.

[234] H. R. Sikkema, B. F. Gaastra, T. Pols, and B. Poolman. Cell Fuelling and Metabolic Energy Conservation in Synthetic Cells. *ChemBioChem* 20:20 (2019), pp. 2581–2592. DOI: 10.1002/cbic.201900398.

[235] A. D. Silverman, A. S. Karim, and M. C. Jewett. Cell-free gene expression: an expanded repertoire of applications. *Nature Reviews Genetics* 21:3 (2020), pp. 151–170. DOI: 10.1038/s41576-019-0186-3.

[236] A. D. Silverman, N. Kelley-Loughnane, J. B. Lucks, and M. C. Jewett. Deconstructing Cell-Free Extract Preparation for in Vitro Activation of Transcriptional Genetic Circuitry. *ACS Synthetic Biology* 8:2 (2019), pp. 403–414. DOI: 10.1021/acssynbio.8b00430.

[237] A. Singh and J. P. Hespanha. Lognormal Moment Closures for Biochemical Reactions. *Proceedings of the 45th IEEE Conference on Decision and Control* 2 (2006), pp. 2063–2068. DOI: 10.1109/CDC.2006.376994.

[238] S. Smith and R. Grima. Single-cell variability in multicellular organisms. *Nature Communications* 9:1 (2018), p. 345. DOI: 10.1038/s41467-017-02710-x.

[239] A. Spirin, V. Baranov, L. Ryabova, S. Ovodov, and Y. Alakhov. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science* 242:4882 (1988), pp. 1162–1164. DOI: 10.1126/science.3055301.

[240] W. K. Spoelstra, S. Deshpande, and C. Dekker. Tailoring the appearance: what will synthetic cells look like? *Current Opinion in Biotechnology* 51 (2018), pp. 47–56. DOI: 10.1016/j.copbio.2017.11.005.

[241] M. Stamatakis and K. Zygourakis. A mathematical and computational approach for integrating the major sources of cell population heterogeneity. *Journal of Theoretical Biology* 266:1 (2010), pp. 41–61. DOI: 10.1016/j.jtbi.2010.06.002.

[242] J. C. Stark, T. Jaroentomeechai, T. D. Moeller, J. M. Hershewe, K. F. Warfel, B. S. Moricz, A. M. Martini, R. S. Dubner, K. J. Hsu, T. C. Stevenson, B. D. Jones, M. P. DeLisa, and M. C. Jewett. On-demand biomanufacturing of protective conjugate vaccines. *Science Advances* 7:6 (2021). DOI: 10.1126/sciadv.abe9444.

[243] J. Stavnezer and R. C. C. Huang. Synthesis of a Mouse Immunoglobulin Light Chain in a Rabbit Reticulocyte Cell-free System. *Nature New Biology* 230 (1971), pp. 172–176.

[244] H. Stein, S. Spindler, N. Bonakdar, C. Wang, and V. Sandoghdar. Production of isolated giant unilamellar vesicles under high salt concentrations. *Frontiers in Physiology* 8:FEB (2017), pp. 1–16. DOI: 10.3389/fphys.2017.00063.

[245] J. Steinkühler, R. L. Knorr, Z. Zhao, T. Bhatia, S. M. Bartelt, S. Wegner, R. Dimova, and R. Lipowsky. Controlled division of cell-sized vesicles by low densities of membrane-bound proteins. *Nature Communications* 11:1 (2020), p. 905. DOI: 10.1038/s41467-020-14696-0.

[246] A. M. Stevens, K. M. Dolan, and E. P. Greenberg. Synergistic binding of the Vibrio fischeri LuxR transcriptional activator domain and RNA polymerase to the lux promoter region. *Proceedings of the National Academy of Sciences of the United States of America* 91:26 (1994), pp. 12619–12623. DOI: 10.1073/pnas.91.26.12619.

[247] T. Stögbauer, L. Windhager, R. Zimmer, and J. O. Rädler. Experiment and mathematical modeling of gene expression dynamics in a cell-free system. *Integrative Biology* 4:5 (2012), p. 494. DOI: 10.1039/c2ib00102k.

[248] F. W. Studier and B. A. Moffatt. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* 189:1 (1986), pp. 113–130. DOI: 10.1016/0022-2836(86)90385-2.

[249] Z. Z. Sun, C. A. Hayes, J. Shin, F. Caschera, R. M. Murray, and V. Noireaux. Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. *Journal of Visualized Experiments* 79 (2013), pp. 1–14. DOI: 10.3791/50762.

[250] Z. Z. Sun, E. Yeung, C. A. Hayes, V. Noireaux, and R. M. Murray. Linear DNA for rapid prototyping of synthetic biological circuits in an escherichia coli based TX-TL cell-free system. *ACS Synthetic Biology* 3:6 (2014), pp. 387–397. DOI: 10.1021/sb400131a.

[251] T. E. Sweeney and C. A. Beuchat. Limitations of methods of osmometry: measuring the osmolality of biological fluids. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 264:3 (1993), R469–R480. DOI: 10.1152/ajpregu.1993.264.3.R469.

[252] K. Takai, T. Sawasaki, and Y. Endo. Practical cell-free protein synthesis system using purified wheat embryos. *Nature Protocols* 5:2 (2010), pp. 227–238. DOI: 10.1038/nprot.2009.207.

[253] Y.-C. Tan, K. Hettiarachchi, M. Siu, Y.-R. Pan, and A. P. Lee. Controlled Microfluidic Encapsulation of Cells, Proteins, and Microbeads in Lipid Vesicles. *Journal of the American Chemical Society* 128:17 (2006), pp. 5656–5658. DOI: 10.1021/ja056641h.

[254] T.-Y. D. Tang, D. Cecchi, G. Fracasso, D. Accardi, A. Coutable-Pennarun, S. S. Mansy, A. W. Perriman, J. L. R. Anderson, and S. Mann. Gene-Mediated Chemical Communication in Synthetic Protocell Communities. *ACS Synthetic Biology* 7:2 (2018), pp. 339–346. DOI: 10.1021/acssynbio.7b00306.

[255] T.-Y. D. Tang, D. van Swaay, A. DeMello, J. L. Ross Anderson, and S. Mann. In vitro gene expression within membrane-free coacervate protocells. *Chemical Communications* 51:57 (2015), pp. 11429–11432. DOI: 10.1039/C5CC04220H.

[256] Y. Tanouchi, D. Tu, J. Kim, and L. You. Noise reduction by diffusional dissipation in a minimal quorum sensing motif. *PLoS Computational Biology* 4:8 (2008), pp. 4–11. DOI: 10.1371/journal.pcbi.1000167.

[257] H. Tarui, S. Imanishi, and T. Hara. A novel cell-free translation/glycosylation system prepared from insect cells. *Journal of Bioscience and Bioengineering* 90:5 (2000), pp. 508–514. DOI: 10.1016/S1389-1723(01)80031-1.

[258] P. Taylor, C. Xu, P. D. I. Fletcher, and V. N. Paunov. A novel technique for preparation of monodisperse giant liposomes. *Chemical Communications* 3:14 (2003), p. 1732. DOI: 10.1039/b304059c.

[259] P. Thévenaz, U. E. Ruttimann, and M. Unser. A pyramid approach to subpixel registration based on intensity. *IEEE Transactions on Image Processing* 7:1 (1998), pp. 27–41. DOI: 10.1109/83.650848.

[260] R. P. Thomsen, M. G. Malle, A. H. Okholm, S. Krishnan, S. S. Bohr, R. S. Sørensen, O. Ries, S. Vogel, F. C. Simmel, N. S. Hatzakis, and J. Kjems. A large size-selective DNA nanopore with sensing applications. *Nature Communications* 10:1 (2019), p. 5655. DOI: 10.1038/s41467-019-13284-1.

[261] P. Torre, C. D. Keating, and S. S. Mansy. Multiphase water-in-oil emulsion droplets for cell-free transcription-translation. *Langmuir* 30:20 (2014), pp. 5695–5699. DOI: 10.1021/la404146g.

[262] M. Toyofuku, K. Morinaga, Y. Hashimoto, J. Uhl, H. Shimamura, H. Inaba, P. Schmitt-Kopplin, L. Eberl, and N. Nomura. Membrane vesicle-mediated bacterial communication. *The ISME Journal* 11:6 (2017), pp. 1504–1509. DOI: 10.1038/ismej.2017.13.

[263] F. C. Tsai and G. H. Koenderink. Shape control of lipid bilayer membranes by confined actin bundles. 11:45 (2015), pp. 8834–8847. DOI: 10.1039/c5sm01583a.

[264] K. Tsumoto, H. Matsuo, M. Tomita, and T. Yoshimura. Efficient formation of giant liposomes through the gentle hydration of phosphatidylcholine films doped with sugar. *Colloids and Surfaces B: Biointerfaces* 68:1 (2009), pp. 98–105. DOI: 10.1016/j.colsurfb.2008.09.023.

[265] M. Ugrinic, A. DeMello, and T.-Y. D. Tang. Microfluidic Tools for Bottom-Up Synthetic Cellularity. *Chem* 5:7 (2019), pp. 1727–1742. DOI: 10.1016/j.chempr.2019.03.012.

[266] A. Urrios, J. Macia, R. Manzoni, N. Conde, A. Bonforti, E. de Nadal, F. Posas, and R. Solé. A Synthetic Multicellular Memory Device. *ACS Synthetic Biology* 5:8 (2016), pp. 862–873. DOI: 10.1021/acssynbio.5b00252.

[267] L. Van de Cauter, F. Fanalista, L. van Buren, N. De Franceschi, E. Godino, S. Bouw, C. Danelon, C. Dekker, G. H. Koenderink, and K. A. Ganzinger. Optimized cDICE for Efficient Reconstitution of Biological Systems in Giant Unilamellar Vesicles. *ACS Synthetic Biology* 10:7 (2021), pp. 1690–1702. DOI: 10.1021/acssynbio.1c00068.

[268] S. Van Der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu. Scikit-image: Image processing in python. *PeerJ* 2014:1 (2014), pp. 1–18. DOI: 10.7717/peerj.453.

[269] P. Van Nies, A. S. Canton, Z. Nourian, and C. Danelon. Monitoring mRNA and protein levels in bulk and in model vesicle-based artificial cells. Vol. 550. 2015, pp. 187–214. DOI: 10.1016/bs.mie.2014.10.048.

[270] P. Van Nies, I. Westerlaken, D. Blanken, M. Salas, M. Mencía, and C. Danelon. Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. *Nature Communications* 9:1 (2018), pp. 1–12. DOI: 10.1038/s41467-018-03926-1.

[271] S. Vanuytsel, J. Carniello, and M. I. Wallace. Artificial Signal Transduction across Membranes. *ChemBioChem* 20:20 (2019), pp. 2569–2580. DOI: 10.1002/cbic.201900254.

[272] M. A. Vibhute, M. H. Schaap, R. J. Maas, F. H. Nelissen, E. Spruijt, H. A. Heus, M. M. Hansen, and W. T. Huck. Transcription and Translation in Cytomimetic Protocells Perform Most Efficiently at Distinct Macromolecular Crowding Conditions. *ACS Synthetic Biology* 9:10 (2020), pp. 2797–2807. DOI: 10.1021/acssynbio.0c00330.

[273] G. Villar, A. D. Graham, and H. Bayley. A Tissue-Like Printed Material. *Science* 340:6128 (2013), pp. 48–52. DOI: 10.1126/science.1229495.

[274] A. M. Voloshin and J. R. Swartz. "Large-scale batch reactions for cell-free protein synthesis". *Cell-free protein synthesis*. 2008.

[275] P. L. Voyvodic, A. Pandi, M. Koch, I. Conejero, E. Valjent, P. Courtet, E. Renard, J. L. Faulon, and J. Bonnet. Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. *Nature Communications* 10:1 (2019), pp. 1–8. DOI: 10.1038/s41467-019-09722-9.

[276] T. Wada, K. ichi Hironaka, M. Wataya, M. Fujii, M. Eto, S. Uda, D. Hoshino, K. Kunida, H. Inoue, H. Kubota, T. Takizawa, Y. Karasawa, H. Nakatomi, N. Saito, H. Hamaguchi, Y. Furuichi, Y. Manabe, N. L. Fujii, and S. Kuroda. Single-Cell Information Analysis Reveals That Skeletal Muscles Incorporate Cell-to-Cell Variability as Information Not Noise. *Cell Reports* 32:9 (2020), p. 108051. DOI: 10.1016/j.celrep.2020.108051.

[277] S. Waldherr. Estimation methods for heterogeneous cell population models in systems biology. *Journal of the Royal Society Interface* 15:147 (2018). DOI: 10.1098/rsif.2018.0530.

[278] S. Walker. An EM Algorithm for Nonlinear Random Effects Models. *Biometrics* 52:3 (1996), p. 934. DOI: 10.2307/2533054.

[279] H. Wang, J. Li, and M. C. Jewett. Development of a Pseudomonas putida cell-free protein synthesis platform for rapid screening of gene regulatory elements. *Synthetic Biology* 3:1 (2018), pp. 1–7. DOI: 10.1093/synbio/ysy003.

[280] Y. Wang and Y. H. Percival. Cell-free protein synthesis energized by slowly-metabolized maltodextrin. *BMC Biotechnology* 9 (2009), pp. 1–8. DOI: 10.1186/1472-6750-9-58.

[281] C. M. Waters and B. L. Bassler. QUORUM SENSING: Cell-to-Cell Communication in Bacteria. *Annual Review of Cell and Developmental Biology* 21:1 (2005), pp. 319–346. DOI: 10.1146/annurev.cellbio.21.012704.131001.

[282] C. M. Waters, W. Lu, J. D. Rabinowitz, and B. L. Bassler. Quorum sensing controls biofilm formation in Vibrio cholerae through modulation of cyclic Di-GMP levels and repression of vpsT. *Journal of Bacteriology* 190:7 (2008), pp. 2527–2536. DOI: 10.1128/JB.01756-07.

[283] W. T. Watson, T. D. Minogue, D. L. Val, S. B. von Bodman, and M. E. Churchill. Structural Basis and Specificity of Acyl-Homoserine Lactone Signal Production in Bacterial Quorum Sensing. *Molecular Cell* 9:3 (2002), pp. 685–694. DOI: 10.1016/S1097-2765(02)00480-X.

[284] A. Weinberger, F. C. Tsai, G. H. Koenderink, T. F. Schmidt, R. Itri, W. Meier, T. Schmatko, A. Schröder, and C. Marques. Gel-assisted formation of giant unilamellar vesicles. *Biophysical Journal* 105:1 (2013), pp. 154–164. DOI: 10.1016/j.bpj.2013.05.024.

[285] M. T. Weinstock, E. D. Hesek, C. M. Wilson, and D. G. Gibson. Vibrio natriegens as a fast-growing host for molecular biology. *Nature Methods* 13:10 (2016), pp. 849–851. DOI: 10.1038/nmeth.3970.

[286] M. Weiss, J. P. Frohnmayer, L. T. Benk, B. Haller, J.-W. Janiesch, T. Heitkamp, M. Börsch, R. B. Lira, R. Dimova, R. Lipowsky, E. Bodenschatz, J.-C. Baret, T. Vidakovic-Koch, K. Sundmacher, I. Platzman, and J. P. Spatz. Sequential bottom-up assembly of mechanically stabilized synthetic cells by microfluidics. *Nature Materials* 17:1 (2018), pp. 89–96. DOI: 10.1038/nmat5005.

[287] D. J. Wiegand, H. H. Lee, N. Ostrov, and G. M. Church. Cell-free Protein Expression Using the Rapidly Growing Bacterium ¡em¿Vibrio natriegens¡/em¿. *Journal of Visualized Experiments* 145 (2019), pp. 1–14. DOI: 10.3791/59495.

[288] M. Yanagisawa, M. Iwamoto, A. Kato, K. Yoshikawa, and S. Oiki. Oriented Reconstitution of a Membrane Protein in a Giant Unilamellar Vesicle: Experimental Verification with the Potassium Channel KcsA. *Journal of the American Chemical Society* 133:30 (2011), pp. 11774–11779. DOI: 10.1021/ja2040859.

[289] N. Yandrapalli, J. Petit, O. Bäumchen, and T. Robinson. Surfactant-free production of biomimetic giant unilamellar vesicles using PDMS-based microfluidics. *Communications Chemistry 2021 4:1* 4:1 (2021), pp. 1–10. DOI: 10.1038/s42004-021-00530-1.

[290] N. Yandrapalli and T. Robinson. Ultra-high capacity microfluidic trapping of giant vesicles for high-throughput membrane studies. *Lab on a Chip* 19:4 (2019), pp. 626–633. DOI: 10.1039/C8LC01275J.

[291] S. Yang, P. A. Pieters, A. Joesaar, B. W. A. Bögels, R. Brouwers, I. Myrgorodska, S. Mann, and T. F. A. de Greef. Light-Activated Signaling in DNA-Encoded Sender–Receiver Architectures. *ACS Nano* 14:11 (2020), pp. 15992–16002. DOI: 10.1021/acsnano.0c07537.

[292] S. S. Yim, N. I. Johns, J. Park, A. L. Gomes, R. M. McBee, M. Richardson, C. Ronda, S. P. Chen, D. Garenne, V. Noireaux, and H. H. Wang. Multiplex transcriptional characterizations across diverse bacterial species using cell-free systems. *Molecular Systems Biology* 15:8 (2019), pp. 1–15. DOI: 10.15252/msb.20198875.

[293] H. Youk and W. A. Lim. Secreting and sensing the same molecule allows cells to achieve versatile social behaviors. *Science* 343:6171 (2014), p. 1242782. DOI: 10.1126/science.1242782.

[294] R. A. Young and R. W. Davis. Efficient isolation of genes by using antibody probes. *Proceedings of the National Academy of Sciences of the United States of America* 80:5 I (1983), pp. 1194–1198. DOI: 10.1073/pnas.80.5.1194.

[295] J. F. Zawada, G. Yin, A. R. Steiner, J. Yang, A. Naresh, S. M. Roy, D. S. Gold, H. G. Heinsohn, and C. J. Murray. Microscale to manufacturing scale-up of cell-free cytokine production-a new approach for shortening protein production development timelines. *Biotechnology and Bioengineering* 108:7 (2011), pp. 1570–1578. DOI: 10.1002/bit.23103.

[296] C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl. Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences* 109:21 (2012), pp. 8340–8345. DOI: 10.1073/pnas.1200161109.

[297] C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koeppl. Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods* 11:2 (2014), pp. 197–202. DOI: 10.1038/nmeth.2794.

[298] P. Zhou, X. Liu, G. Wu, P. Wen, L. Wang, Y. Huang, and X. Huang. Programmable Modulation of Membrane Permeability of Proteinosome upon Multiple Stimuli Responses. *ACS Macro Letters* 5:8 (2016), pp. 961–966. DOI: 10.1021/acsmacrolett.6b00506.

[299] G. Zubay. In Vitro Synthesis of Protein in Microbial Systems. *Annual Review of Genetics* 7:1 (1973), pp. 267–287. DOI: 10.1146/annurev.ge.07.120173.001411.

Erklärung des Promovierenden zum Antrag auf Eröffnung des Promotionsverfahrens

1. Die folgende Promotionsordnung in ihrer gültigen Fassung erkenne ich an:
   - Bereich Mathematik und Naturwissenschaften - Promotionsordnung vom 23.02.2011
2. Die Promotion wurde an folgendem Institut/an folgender Professur durchgeführt:
   - Prof. A. Francis Stewart, Ph.D.
3. Folgende Personen haben die Promotion wissenschaftlich betreut und/oder mich bei der Auswahl und Auswertung des Materials sowie bei der Herstellung des Manuskripts unterstützt:
   - T.-Y. Dora Tang, Ph.D.
   - Christoph Zechner, Ph.D.
4. Ich bestätige, dass für meine Person bisher keine früheren, erfolglosen Promotionsverfahren stattgefunden haben. Andernfalls habe ich diesem Antrag eine Erklärung bzw. Nachweise beigelegt, in dem ersichtlich ist, wo, wann, mit welchem Thema und mit welchem Bescheid diese Promotionsversuche stattgefunden haben.
5. Ich versichere weiterhin, dass
   (a) ich die vorliegende Arbeit mit dem Titel "Building synthetic multicellular systems from the bottom-up" ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel selbst angefertigt habe. Hilfe Dritter wurde nur in wissenschaftlich vertretbarem und prüfungsrechtlich zulässigem Ausmaß in Anspruch genommen. Es sind keine unzulässigen geldwerten Leistungen, weder unmittelbar noch mittelbar, im Zusammenhang mit dem Inhalt der vorliegenden Dissertation an Dritte erfolgt.
   (b) die aus fremden Quellen direkt oder indirekt übernommenen Gedanken als solche kenntlich gemacht sind.
   (c) ich die vorliegende Arbeit bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zwecke einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt habe.
6. Mir ist bekannt, dass die Nichteinhaltung dieser Erklärung oder unrichtige Angaben zum Verfahrensabbruch oder zum nachträglichen Entzug des Doktortitels führen können.

David T. Gonzales
Dresden, 7. Dezember 2021