Engineering Synthetic Cells through Module Integration and Evolution

Restrepo Sierra, A.M.

DOI
10.4233/uuid:2f2dfc76-5e29-4a04-84c1-3d337e3bf645

Publication date
2024

Document Version
Final published version

Citation (APA)
Restrepo Sierra, A. M. (2024). *Engineering Synthetic Cells through Module Integration and Evolution*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:2f2dfc76-5e29-4a04-84c1-3d337e3bf645

Engineering Synthetic Cells through Module Integration and Evolution

Dissertation
for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen
Chair of the Board for Doctorates
to be defended publicly on
Thursday 11th April 2024 at 10:00 o'clock

by

Ana María RESTREPO SIERRA

Master of Science in Bioscience,
King Abdullah University of Science and Technology (KAUST), Saudi Arabia
born in Medellín, Colombia.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof. dr. C.J.A. Danelon	Toulouse Biotechnology Institute, promotor
Prof. dr. G.H. Koenderink	Delft University of Technology, promotor

Independent members:

Dr. Y. Rondelez	ESPCI, France
Dr. N.J. Claassens	Wageningen University & Research, NL
Dr. K. Adamala	University of Minnesota, USA
Prof. dr. N. Ichihashi	University of Tokyo, Japan
Prof. dr. ir. P.A.S. Daran-Lapujade	Delft University of Technology, NL
Prof. dr. N. Dekker	Reserve member. Delft University of Technology, NL

"There is a theory which states that if ever anyone discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable. There is another theory which states that this has already happened"

– Douglas Adams, The Restaurant at the End of the Universe

Keywords: synthetic biology, synthetic cell, liposomes, cell-free gene expression, module integration, DNA replication, phospholipid biosynthesis, in vitro evolution.

Printed by: Gildeprint
Cover by: Ana María Restrepo Sierra

Casimir PhD-series
ISBN/EAN: 978-94-6384-563-2
An electronic copy of this dissertation is available at https://repository.tudelft.nl/

Table of contents

1

General Introduction

Abstract

Life, the most complex and admirable machine that one could think of has evolved over billions of years to display a beautiful variety of mechanisms that keep cells adapting, self-maintaining, reproducing, and evolving. If we think about it, what is this magic? What are the mechanisms behind life's origins and wonderful coordination? Attracted by these intricates, different scientific disciplines have for long studied all life's scales to grasp the fundamental principles of life [1]. In particular, the synthetic biology field has set an ultimate goal of discerning life until the point that a minimal synthetic cell can be fully recreated in a controlled laboratory set-up [2–5]. Synthetic cells, modular enough to be crafted by scientists, could not only reveal fundamental insights of how life works, but can also help unlock great biotechnological applications that lie beyond the reach of our current technologies and understanding of life.

Synthetic cells: re-imagining life from a top-down or bottom-up approach

The starting point to devise a machine that can be called "alive" is to define what are the basic functions of life? After Darwin's 19th-century theories regarding the origins of life on a *"warm little pond with all sorts of ammonia & phosphoric salts,—light, heat, electricity etcetera present, that a protein compound was chemically formed, ready to undergo still more complex changes"* [6], the 20th century witnessed a surge in conceptual exploration into the nature and essential traits of life. Erwin Schrödinger in his renowned 1944 book, *"What is Life?"* suggested that life, like any other natural phenomena, could be entirely comprehended as a physical process, governed by the principles of chemistry and physics [7]. Later on, Tibot Gánti, already in 1971, proposed a theoretical protocell model **(Fig. 1 a)** indicating that a functional minimal cell must at least encompass three essential components: (i) a boundary to separate the system from the environment, (ii) a metabolic network to self-sustain the system's chemical reactions, and (iii) an information storage molecule (i.e, the DNA), responsible to encode and pass-on information through generations [8]. Moreover, by 1980, the search for a minimal genome started with Harold J. Morowitz proposing *Mycoplasma genitalium*, containing only 525 genes, as a useful starting organism to understand essential principles of life [9]. On the quest to translate these proposed frameworks into a concrete reality, the synthetic biology field, starting to shape up as a research field in the late 1990s [5], embraced the construction of a minimal living entity, or minimal cell from two research lines: a bottom-up and a top-down approach **(Fig. 1 b).**

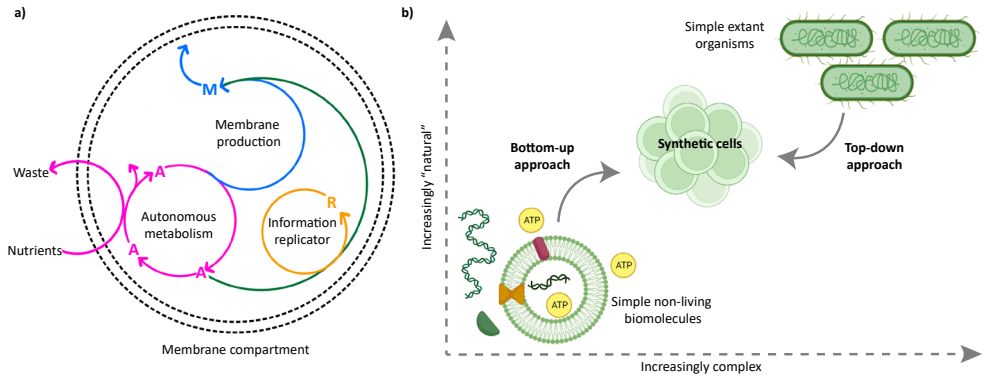


Fig. 1 a) Outlined illustration of the Chemoton. The conceptual Chemoton model, introduced by Tibor Gánti in 1971 [8], illustrates the foundational elements of early life: a membrane boundary, a metabolic network, and a replicator for information flow. **b)** Illustration of current strategies to build a synthetic cell: bottom-up and top-down approaches. Top-down approach starting from simple extant organisms. The organism complexity is reduced to reach a simplified or redesigned cell that can still self-sustain and be called alive. Bottom-up approach focuses on assembling basic non-living components to eventually realize a complex enough autonomous synthetic cell. The schematic illustration was inspired from [10] and created with BioRender.com.

Top-down approach

The top-down approach, mostly led by the J. Craig Venter institute, takes an extant organism and "simplifies" it to maintain the minimal functionalities needed to sustain life. Nearly two decades ago, researchers managed to pare down the genome of *M. genitalium*, the known organism with the smallest genome, to only 382 genes. Curiously though, even with this reduction, the functions of 28% of the resulting proteins remained unknown [11]. Subsequent to this work, Gibson et al. gave a strong step forward to build a synthetic cell by showing that a cell could function with a fully synthetic genome. A *M. mycoides* 1079 kb genome was chemically synthesized and transferred into an *M. Capricomum* cell to re-create a new *M. mycoides* organism fully ruled by a synthetic genome. The nearly synthetic cell was called JCVI-syn1.0 [12]. More recently, in 2016, JCVI-syn1.0 was utilized to engineer an even more minimal synthetic cell with a design, build, and test (DBT) iterative process. Researchers managed to obtain JCVI-syn3.0, a new working approximation to a minimal cell with a further reduced genome (531 kb), capable of self-reproduction [13]. Five main categories of cellular functions could be identified from the gene-expression profile of JCVI-syn3.0: cytosolic metabolism, cell membrane, gene-expression from genome information, preservation of genome information, and unassigned. Interestingly, JCVI-syn3.0 seems to be polymorphous with an altered growth rate when compared to its predecessor (JCVI-syn1.0), and still has 149 genes with unknown biological functions. Overall, even if some processes remain elusive, JCVI-syn3.0 has been a great framework to study essential biological processes on cell aliveness. Some studied mechanisms and useful applications include: understanding minimal cell mechanisms through modelling [14,15], overviews of the mechanics of minimal cell division [16], synthetic production of vaccines [17], genetic engineering and genome assembly methods [13,18,19], and most recently, great insights about minimal cell evolution [20].

Bottom-up approach

The bottom-up strategy seeks to emulate life's essential functions, starting from elemental non-living units and gradually piecing them together to eventually create a functioning cell-like entity [21]. Here, stripped down biomolecules from an origins of life scenario can be a useful starting point to build up the main building blocks of life. From this origins of life perspective, the 20th century brought great discoveries on the central dogma of molecular biology [22], and insightful theories on the RNA world to explain the origins of life [23]. Great progress has been accomplished on understanding prebiotic reactions for the formation, reproduction, and Darwinian evolution of essential biomolecules [24–27]. However, some primitive reaction conditions are still difficult to imitate, and a full understanding on the origins of essential life's biomolecules (i.e DNA, RNA, amino acids and proteins) is still an ongoing unsolved task. Thus, a bottom-up biochemical perspective to build a synthetic cell is complemented by the synthetic biology field where the main building blocks are already 'pre-made' biomolecules (RNA, DNA, proteins) resulting from

evolution. Bottom-up synthetic biology explores cell-like functions and evaluates their function and robustness within cell-like compartments. Strategies have not been limited by functions from only one type of organism, nor by already existing functions [28]. Whether from a virus, bacterial or eukaryotic cell, any useful life-like essential machinery can be implemented. For instance, Baldauf et al., explains in [29] how an actomyosin cytoskeleton from eukaryotic cells could be a useful tool for a synthetic cell division mechanism, while Kretschmer et al., focuses in [30] on promising bacterial cell division strategies for synthetic cells. Furthermore, the integration of transcription and translation mechanisms from *E. coli*, coupled with bacteriophage components (i.e., T7 or SP6 RNAPs), has sparked synthetic cell research for exploring DNA-encoded protein machineries [31–34]. With the latter, bacterial and viral-based functions have been explored inside gene-expressing compartments. Some examples include phospholipid biosynthesis [35], cytoskeleton networks [36], and FtsZ division mechanisms [37–39] from *E.coli*, and viral-based DNA replication machineries as a robust mechanism for a synthetic cell DNA self-replication module [40].

Overall, while top-down approaches follow a reductionist path, dissecting and modifying existing biological entities to reduce complexity, bottom-up approaches take a holistic perspective, assembling components to create systems from the ground up. Despite having distinct inspirations, these methods can nicely complement each other to foster iterative refinement, cross-validation, and for eventually reaching a better understanding of life. Recently, Sakai et al. demonstrated how both top-down and bottom-up strategies could start to complement each other [41]. There, a modified version of JCVI-syn3.0, JCVI-syn3A, containing an extra 12 kb for a nearly normal growth rate and morphology, was utilized for generating a nearly minimal cell-free system that supported in vitro protein expression. Clearly, a collaborative fusion of reductionism and holistic assembly techniques could pave the way for the eventual realization of a fully functional and adaptable synthetic cell.

A DNA-based approach for building a synthetic cell

Synthetic vesicles with DNA-encoded functions

In the Danelon lab, we focus our research on reconstructing life's functionalities from DNA-encoded machineries inside cell-like compartments **(Fig. 2)**. We visualize a synthetic cell as an eventually autonomous entity with DNA encoded information, and capable of self-sustenance, reproduction, evolution, and heredity. To perform the flow of information from DNA into proteins, we utilize the PURE system [31] (Protein synthesis Using Recombinant Elements) as our main gene-expression platform containing purified protein components from T7 bacteriophage for transcription (T7 RNAP), and from *E.coli* for translation. Even though cell-free extracts have been explored to create "man-made synthetic cells" inside femtoliter oil-based compartments [42], or inside lipid vesicles [43,44], the PURE system offers higher modularity, production yields, and control, when compared to standard cell-free reaction systems [31]. Interestingly, already a few years after PURE system development in 2001, its encapsulation inside cell-like compartments became almost a research line by its

own. Research groups have focused on characterizing in-vesiculo PURE system dynamics [45], compared different commercially available PURE system kits [46], studied PURE-based membrane protein synthesis [47] and, like us, started to utilize the PURE system as synthetic cell framework for the reconstitution of DNA encoded synthetic cell machineries [48,49].

For our cell-like compartment, we utilize semi-permeable lipid vesicles of a phospholipid composition that resembles the *E. coli* membrane one [46,50]. Similar to standard liposome production techniques, such as the lipid swelling methods [51], we perform a gentle swelling technique with ~200-300 µm lipid-coated beads instead of a flat or round glass surface [46]. With this strategy, we increase the lipid film to swelling solution ratios, are able to utilize smaller volumes for the swelling solution, and avoid the implementation of oil or organic solvents that may interfere with the efficiency of our encapsulated reactions. Our bead-based vesicle swelling technique allows for the production of a cell-sized polydisperse liposome population with a median of ~4 µm diameter.

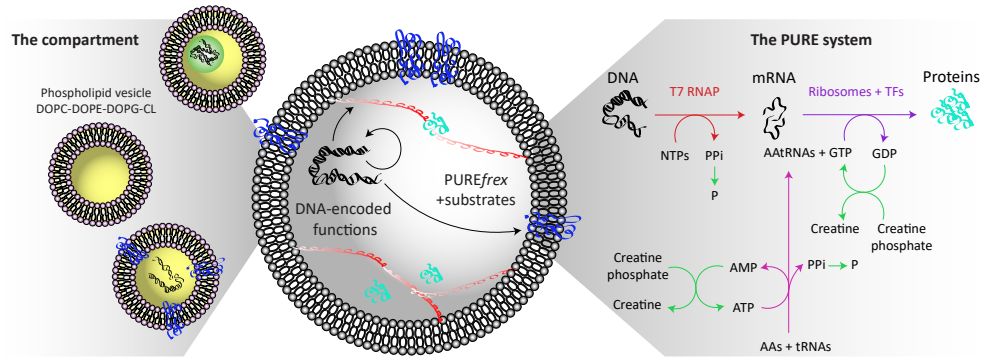


Fig. 2 Synthetic cells with DNA-encoded functions. Illustration of phospholipid synthetic vesicles with encapsulated DNA and PURE system for the production of synthetic cell relevant protein machineries. The PURE system acts as a main metabolic machinery for transcription and translation of the DNA-encoded proteins with an effective energy regeneration system that includes creatine kinase, myokinase, nucleoside-diphosphate kinase, and pyrophosphatase. Whenever needed, the PURE system can be supplemented or adapted with additional substrates, co-factors, or additional purified proteins (i.e chaperons [36]) for the proper folding and/or functioning of the synthetic-cell module to reconstitute. The compartment that encapsulates the DNA template, PURE, and needed additional substrates is a lipid bilayer vesicle formed with an *E. coli*-based composition of amphiphilic phospholipid molecules with a hydrophilic polar head group and a hydrophobic non-polar tail.

Engineering synthetic cells via module integration and evolution

Despite the recent advances in the reconstitution of biological functions, either from purified proteins or DNA-encoded ones, the level of molecular and organizational complexity reached so far is still insufficient to fully emulate functional life-like properties [35,36,38]. The rational fine-tuning of components can enhance biomodule performance, although this alone still falls short of achieving complete functionality. Moreover, realizing a full library of individually optimized modules is not enough to build a synthetic cell.

Eventually, biomodules need to integrate within the same synthetic cell without compromising individual module functionality [52,53]. However, modules are generally not inherently compatible, as they come from different organisms, and were evolved within their unique environmental context. Thus, the challenge lies not only in ensuring individual functionalities but also in composing a synergized network of originally unrelated biomodules, surpassing the constraints of their evolutionary divergence.

In-vesiculo protein expression can be considered as a good starting point for attempts at module integration. However, besides testing DNA-encoded modules, and the development of energy supply strategies for potentiating protein expression, not much research has been devoted to understanding module integration and coordination in a synthetic cell context. Only recently, Ichihashi's group published an interesting work on the compatibility between transcription, translation, and DNA replication in solution [54]. Their work shows some of the inhibitory effects behind the coordination of these three central dogma reactions. They highlight how adjusting magnesium concentrations can help improve transcription and translation when all three reactions are coupled. Clearly, evaluating module compatibility and rational engineering are useful starting strategies to reach biomodule synergy. However, as the number of integrated modules grows, more complex cross-talks can arise as a result of nonlinear relationships, evolutionary history, and context-specific factors. In such a scenario, a rational engineering approach alone might not be enough to tackle module incompatibility or improve module synergy. Thus, in the Danelon lab, we propose to complement the rational approach for module integration with the implementation of evolution as a fundamental biomodule for synthetic cell engineering. To our advantage, successful in-vitro evolutionary campaigns do not necessitate previous structure-function relationships, which fits our starting synthetic-cell set-up where most biomodules do not have a pre-defined function-relationship. We envision that, as also implemented in nature, evolution can be a powerful tool to accelerate the emergence of synthetic cells with advanced functionalities [55] **(Fig. 3)**.

The start of plausible in vitro evolution within gene-expressing microcompartments can be appointed to Griffiths work in 1998, were the first cell-like genotype to phenotype link was established within water-in-oil droplets [42]. Their work helped demonstrate the need of a strong genotype to phenotype link for a successful selection and enrichment of genetic variants in compartmentalized in vitro evolutionary campaigns. Besides industrial directed evolution applications [56], synthetic cell and origins of life research has also touched upon evolution (i.e., Darwinian evolution) for understanding fundamentals about life's origins, and adaptation capabilities of synthetic cells. The J. Craig Venter institute recently demonstrated how a nearly synthetic cell can rapidly adapt and evolve, undergoing genome minimization with distinct target selection, including some still mysterious proteins with unknown functionalities [20]. Moreover, Ichihashi et al., studied in [57] the evolution of RNA self-replicators over a 600-generation experiment within a cell-like system, and in ref [58] further developed into the study of RNA replicator networks resulting from the long-

term RNA evolution experiments. Interestingly, this work demonstrates how emergent RNA lineages, including host and parasites, cope together to help each other replicate. The group of Adamala and Szostak also reported an interesting resource-based competition model for Darwinian evolution inside protocells [59]. With an RNA-based genome, they illustrate that ribozymes capable of synthesizing short hydrophobic peptides can accelerate protocell growth, conferring a selective advantage over other protocells that could not grow along. Even if not showing evolution per se, their work confers a nice insight into protocell competitive growth and adaptation, already touched upon by Szostak in 2004 [60]. In general, notwithstanding the existing research, there is still a long learning road on understanding life's foundations. What is evident now is that evolution has adeptly guided life to adjust and endure over various selection pressures. **It is incredibly motivating to think how evolution can once more be harnessed to eventually fabricate a synthetic cell in a laboratory setting.**

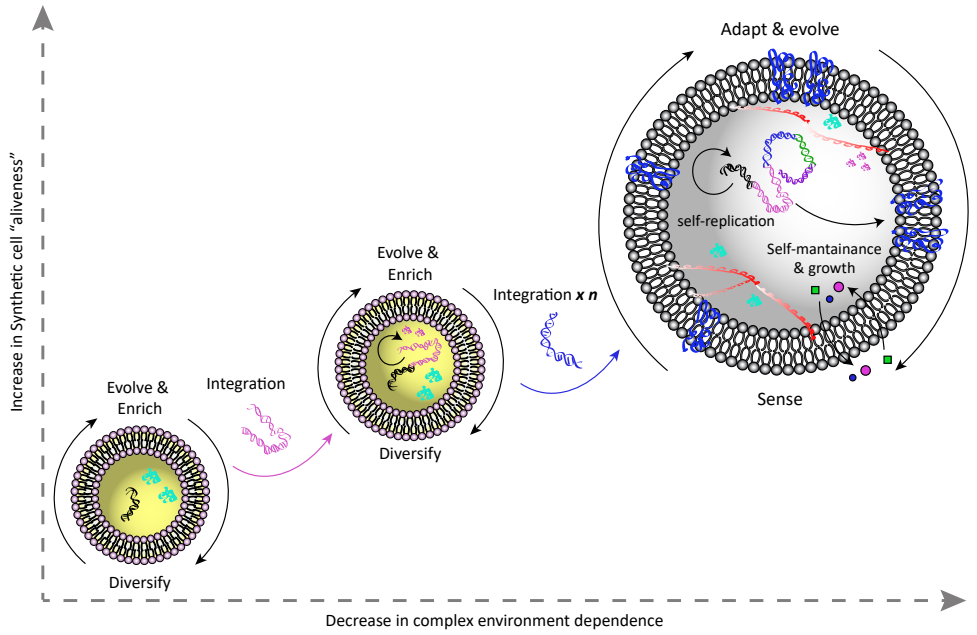


Fig. 3 In vitro evolution for building a synthetic cell. Through a systematic evolutionary approach with iterative rounds of module integration, diversification, and evolution inside synthetic vesicles, we envision the production of an autonomous synthetic cell with coordinated advanced functionalities.

With this thesis, we aim to contribute to the aim of engineering a synthetic cell through evolution and biomodule integration. Throughout four cool years of research, we first explored the capabilities of the Phi29 DNA replication module previously reconstituted inside gene-expressing vesicles [40], as a basal mechanism for (i) improving in-vitro evolutionary campaigns inside synthetic vesicles, (ii) evolving a DNA self-replicator within a synthetic cell framework, and (iii) integrating three essential synthetic cell modules: DNA self-replication, gene-expression, and phospholipid biosynthesis within synthetic

vesicles. Secondly, we explored high-throughput phenotype characterization as a must-perform step for characterizing and identifying optimization goals of various synthetic cell populations. Finally, we motivate the implementation of high-throughput image-based techniques for selection and sorting strategies on future evolutionary campaigns. We further outline and explain possible subsequent research steps to follow up on the hereby presented research towards building a synthetic cell. Briefly, the next chapters of this thesis discuss the following:

- **Chapter 2** introduces **CADGE: Clonal Amplification Enhanced Gene Expression** inside gene-expressing synthetic vesicles. With CADGE, we aim to tackle the low phenotypic output that results from low input DNA concentrations on in-vesiculo protein evolution efforts. For the development of CADGE, we utilize the Phi29 DNA replication machinery for the orthogonal amplification of a linear protein-coding dsDNA template under simultaneous transcription and translation. We demonstrate that CADGE can boost the protein expression and phenotypic output of a few soluble and membrane-associated proteins inside gene-expressing vesicles. We further illustrate how CADGE can improve in-vesiculo protein evolution campaigns by enabling the enrichment and recovery of a DNA variant selected with a positive feed-back loop strategy, or a high-throughput fluorescence-based sorting strategy.
- **Chapter 3** illustrates the **adaptive evolution of a self-replicating DNA inside synthetic vesicles**. Considering that evolution is a life's must-process for optimization, adaptation and heredity, we studied the evolution of a minimal DNA self-replicator system within gene-expressing synthetic vesicles. Under two different in-liposome evolutionary configurations: intermittent and continuous, we managed to (i) achieve DNA self-amplification from low DNA starting concentrations, (ii) introduce in-situ DNA diversity, and (iii) enrich for better performing DNA self-replicators. Within only few evolutionary rounds, we show the emergence and persistence of synonymous and non-synonymous mutations that allow the enrichment of advantageous self-replicators. By NGS sequencing and reverse engineering we further characterize and conclude on some of the most prominently enriched non-synonymous mutations on the DNA polymerase.
- **Chapter 4** presents **synthetic cells with a combined central dogma and synthesis of phospholipids**. We demonstrate the integration of three essential hallmarks of life inside cell-like vesicle compartments: DNA replication, membrane biosynthesis, and a transcription-translation machinery for gene-expression. We show how simultaneous gene-expression from a single synthetic genome enables DNA self-replication and phospholipid biosynthesis inside liposome compartments. We demonstrate that both DNA replication and phospholipid biosynthesis modules are surprisingly compatible and minimally affected by the presence of each other substrates, co-factors, or intermediary reaction compounds. We found, however, that co-expression of both

modules from our single genome negatively influences the overall occurrence of liposomes with active phospholipid biosynthesis, and leads to a decrease in DNA amplification yields.

- **Chapter 5** introduces **Imaging Flow Cytometry (IFC) for high-throughput phenotyping of synthetic cells**. We introduce IFC to the synthetic cell community as an information-rich imaging method for the rapid phenotyping of synthetic cells. With currently available IFC and data analysis software, we provide comprehensive pipelines for assessing synthetic cell populations with different phenotypic traits. We show how robust population statistics can be obtained from running only a few microliters of a liposome suspension sample (~60 thousands of liposomes per microliter), and discuss the importance of high-throughput analysis for effectively engineering synthetic cells.
- Finally, in **Chapter 6** we explore our **current progress and future plans for developing artificial cells with a semi-rational evolutionary approach**. On the scope of a design-build-test-learn (DBTL) methodology for creating synthetic cells, we discuss possible strategies for: (i) building or improving DNA templates for better (co)biomodule performance and their upcoming evolution, and (ii) screening and selecting the fittest synthetic cell variants from an evolutionary campaign. On this last point, we illustrate current and emergent technologies for assisting synthetic cell evolution, including ones that could cope well with complex phenotypic traits. Lastly, we delve into near-future work on the evolution of synthetic vesicles with integrated DNA self-replication and phospholipid biosynthesis, an immediate upcoming step from what is presented in chapter 4 of this thesis.

References

1. Scharf, C. et al. A Strategy for Origins of Life Research. Astrobiology 15, 1031–1042 (2015).
2. Dzieciol, A. J. & Mann, S. Designs for life: protocell models in the laboratory. Chem Soc Rev 41, 79–85 (2012).
3. Frischmon, C., Sorenson, C., Winikoff, M. & Adamala, K. P. Build-a-Cell: Engineering a Synthetic Cell Community. Life 11, 1176 (2021).
4. Sj, M. On biochemical constructors and synthetic cells. Interface focus 13, (2023).
5. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. Nat Rev Microbiol 12, 381–390 (2014).
6. https://www.darwinproject.ac.uk/letter/DCP-LETT-7471.xml. Darwin Correspondence Project
7. Schrödinger, E. What is Life? The Physical Aspect of the Living Cell. (Cambridge University Press, 1944).
8. Gánti, T. Organization of chemical reactions into dividing and metabolizing units: The chemotons. Biosystems 7, 15–21 (1975).
9. Morowitz, H. J. The completeness of molecular biology. Isr J Med Sci 20, 750–753 (1984).
10. Wikmark, O.-G. et al. Synthetic biology: biosafety and contribution to addressing societal challenges. (2016). doi:10.13140/RG.2.2.29987.25121.
11. Glass, J. I. et al. Essential genes of a minimal bacterium. Proc Natl Acad Sci U S A 103, 425–430 (2006).
12. Gibson, D. G. et al. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. Science 329, 52–56 (2010).
13. Hutchison, C. A. et al. Design and synthesis of a minimal bacterial genome. Science 351, aad6253 (2016).
14. Thornburg, Z. R. et al. Fundamental behaviors emerge from simulations of a living minimal cell. Cell 185, 345-360. e28 (2022).
15. Thornburg, Z. R. et al. Kinetic Modeling of the Genetic Information Processes in a Minimal Cell. Frontiers in Molecular Biosciences 6, (2019).
16. Pelletier, J. F., Glass, J. I. & Strychalski, E. A. Cellular mechanics during division of a genomically minimal cell. Trends in Cell Biology 32, 900–907 (2022).
17. Dormitzer, P. R. et al. Synthetic generation of influenza vaccine viruses for rapid response to pandemics. Sci Transl Med 5, 185ra68 (2013).
18. Karas, B. J. et al. Designer diatom episomes delivered by bacterial conjugation. Nat Commun 6, 6925 (2015).
19. Hernandez Hernandez, D. et al. Improved Combinatorial Assembly and Barcode Sequencing for Gene-Sized DNA Constructs. ACS Synth. Biol. 12, 2778–2782 (2023).
20. Moger-Reischer, R. Z. et al. Evolution of a minimal cell. Nature 620, 122–127 (2023).
21. Powell, K. How biologists are creating life-like cells from scratch. Nature 563, 172–175 (2018).
22. Crick, F. Central Dogma of Molecular Biology. Nature 227, 561–563 (1970).
23. Gilbert, W. Origin of life: The RNA world. Nature 319, 618–618 (1986).
24. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature 459, 239–242 (2009).
25. Ameta, S. et al. Darwinian properties and their trade-offs in autocatalytic RNA reaction networks. Nat Commun 12, 842 (2021).
26. Adamski, P. et al. From self-replication to replicator systems en route to de novo life. Nat Rev Chem 4, 386–403 (2020).
27. Biscans, A. Exploring the Emergence of RNA Nucleosides and Nucleotides on the Early Earth. Life 8, 57 (2018).
28. Göpfrich, K., Platzman, I. & Spatz, J. P. Mastering Complexity: Towards Bottom-up Construction of Multifunctional Eukaryotic Synthetic Cells. Trends Biotechnol 36, 938–951 (2018).
29. Baldauf, L., van Buren, L., Fanalista, F. & Koenderink, G. H. Actomyosin-Driven Division of a Synthetic Cell. ACS Synth Biol 11, 3120–3133 (2022).
30. Kretschmer, S., Ganzinger, K. A., Franquelim, H. G. & Schwille, P. Synthetic cell division via membrane-transforming molecular assemblies. BMC Biology 17, 43 (2019)
31. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. Nat Biotechnol 19, 751–755 (2001).

32. Gonzales, D. T., Suraritdechachai, S. & Tang, T.-Y. D. Compartmentalized Cell-Free Expression Systems for Building Synthetic Cells. Adv Biochem Eng Biotechnol 186, 77–101 (2023).
33. Buddingh', B. C. & van Hest, J. C. M. Artificial Cells: Synthetic Compartments with Life-like Functionality and Adaptivity. Acc Chem Res 50, 769–777 (2017).
34. Nourian, Z., Scott, A. & Danelon, C. Toward the assembly of a minimal divisome. Syst Synth Biol 8, 237–247 (2014).
35. Blanken, D., Foschepoth, D., Serrão, A. C. & Danelon, C. Genetically controlled membrane synthesis in liposomes. Nat Commun 11, 4317 (2020).
36. Kattan, J., Doerr, A., Dogterom, M. & Danelon, C. Shaping Liposomes by Cell-Free Expressed Bacterial Microtubules. ACS Synth. Biol. 10, 2447–2455 (2021).
37. Godino, E. et al. Cell-free biogenesis of bacterial division proto-rings that can constrict liposomes. Commun Biol 3, 1–11 (2020).
38. Godino, E. et al. De novo synthesized Min proteins drive oscillatory liposome deformation and regulate FtsA-FtsZ cytoskeletal patterns. Nat Commun 10, 4969 (2019).
39. Kohyama, S., Merino-Salomón, A. & Schwille, P. In vitro assembly, positioning and contraction of a division ring in minimal cells. Nat Commun 13, 6098 (2022).
40. van Nies, P. et al. Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. Nat Commun 9, 1–12 (2018).
41. Sakai, A. et al. Cell-Free Expression System Derived from a Near-Minimal Synthetic Bacterium. ACS Synth. Biol. 12, 1616–1623 (2023).
42. Tawfik, D. S. & Griffiths, A. D. Man-made cell-like compartments for molecular evolution. Nat Biotechnol 16, 652–656 (1998).
43. Noireaux, V. & Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. Proc Natl Acad Sci U S A 101, 17669–17674 (2004).
44. Nomura, S. M. et al. Gene Expression within Cell-Sized Lipid Vesicles. ChemBioChem 4, 1172–1175 (2003).
45. Saito, H. et al. Time-Resolved Tracking of a Minimum Gene Expression System Reconstituted in Giant Liposomes. ChemBioChem 10, 1640–1643 (2009).
46. Blanken, D., van Nies, P. & Danelon, C. Quantitative imaging of gene-expressing liposomes reveals rare favorable phenotypes. Phys. Biol. 16, 045002 (2019).
47. Kuruma, Y. & Ueda, T. The PURE system for the cell-free synthesis of membrane proteins. Nat Protoc 10, 1328–1344 (2015).
48. Yoshida, A., Kohyama, S., Fujiwara, K., Nishikawa, S. & Doi, N. Regulation of spatiotemporal patterning in artificial cells by a defined protein expression system. Chemical Science 10, 11064–11072 (2019).
49. Okauchi, H. & Ichihashi, N. Continuous Cell-Free Replication and Evolution of Artificial Genomic DNA in a Compartmentalized Gene Expression System. ACS Synth. Biol. 10, 3507–3517 (2021).
50. de Souza, T. P., Fahr, A., Luisi, P. L. & Stano, P. Spontaneous Encapsulation and Concentration of Biological Macromolecules in Liposomes: An Intriguing Phenomenon and Its Relevance in Origins of Life. J Mol Evol 79, 179–192 (2014).
51. Shohda, K., Takahashi, K. & Suyama, A. A method of gentle hydration to prepare oil-free giant unilamellar vesicles that can confine enzymatic reactions. Biochem Biophys Rep 3, 76–82 (2015).
52. Caschera, F. & Noireaux, V. Integration of biological parts toward the synthesis of a minimal cell. Current Opinion in Chemical Biology 22, 85–91 (2014).
53. Guindani, C., da Silva, L. C., Cao, S., Ivanov, T. & Landfester, K. Synthetic Cells: From Simple Bio-Inspired Modules to Sophisticated Integrated Systems. Angewandte Chemie International Edition 61, e202110855 (2022).
54. Seo, K. & Ichihashi, N. Investigation of Compatibility between DNA Replication, Transcription, and Translation for in Vitro Central Dogma. ACS Synth. Biol. (2023) doi:10.1021/acssynbio.3c00130.
55. Abil, Z. & Danelon, C. Roadmap to Building a Cell: An Evolutionary Approach. Frontiers in Bioengineering and Biotechnology 8, (2020).

56. Holstein, J. M., Gylstorff, C. & Hollfelder, F. Cell-free Directed Evolution of a Protease in Microdroplets at Ultrahigh Throughput. ACS Synth. Biol. 10, 252–257 (2021).
57. Ichihashi, N. et al. Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment. Nat Commun 4, 2494 (2013).
58. Mizuuchi, R., Furubayashi, T. & Ichihashi, N. Evolutionary transition from a single RNA replicator to a multiple replicator network. Nat Commun 13, 1460 (2022).
59. Adamala, K. & Szostak, J. W. Competition between model protocells driven by an encapsulated catalyst. Nature Chem 5, 495–501 (2013).
60. Chen, I. A., Roberts, R. W. & Szostak, J. W. The emergence of competition between model protocells. Science 305, 1474–1476 (2004).

2

CADGE: Clonal Amplification-Enhanced Gene Expression in Synthetic Vesicles

Abstract

In cell-free gene expression, low input DNA concentration severely limits the phenotypic output, which may impair in vitro protein evolution efforts. We address this challenge by developing CADGE, a strategy that is based on clonal isothermal amplification of a linear gene-encoding dsDNA template by the minimal Phi29 replication machinery and in situ transcription-translation. We demonstrate the utility of CADGE in bulk and in clonal liposome microcompartments to boost the phenotypic output of soluble and membrane-associated proteins, as well as to facilitate the recovery of encapsulated DNA. Moreover, we report that CADGE enables the enrichment of a DNA variant from a mock gene library either via a positive feedback loop-based selection or via high-throughput screening. This new biological tool can be implemented for cell-free protein engineering and the construction of a synthetic cell.

This chapter is taken from an already published manuscript with co-first authorship with Zhanar Abil: Abil, Z.*, Restrepo Sierra, A. M.* & Danelon, C. Clonal Amplification-Enhanced Gene Expression in Synthetic Vesicles. ACS Synth. Biol. 12, 1187–1203 (2023).

*Denotes equal contribution.

Introduction

Inspired by natural selection, directed evolution has become a powerful tool in synthetic biology. This engineering approach encompasses cycles of genetic diversification and enrichment of rare desired variants, allowing for accelerated protein evolution even with limited a priori knowledge about the structure-function relationships [1–3]. Directed evolution enabled engineering of a plethora of proteins, genetic pathways, and even genomes to generate variants with improved or tailor-made properties [4–11]. Incorporation of directed evolution principles to the construction framework of a synthetic cell has recently been proposed [12]. Compartmentalized gene expression in liposomes [13–15] has gained considerable momentum in the last few years, with methodological advances that have improved the yield of functional vesicles [16,17], enabling the reconstitution of complex biological functions, such as DNA replication [18], phospholipid synthesis [19], membrane deformation processes [20–22] and light-triggered ATP synthesis [23]. Moving forward to optimizing and integrating cellular modules may require a system's level evolutionary approach [12].

Over the past decades, numerous in vivo and cell-free methodologies for gene expression of the targeted phenotypes have been developed. In vivo methodologies have been the most common, since a suitable host organism could provide low-cost gene expression with a reliable yield [24]. However, cell-free systems have emerged as an alternative and attractive platform due to the higher degree of controllability and freedom from the constrains related to cell survival [25,26]. Cell-free protein synthesis enabled engineering of a number of proteins, including membrane or cytotoxic proteins [27,28] as well as peptides and proteins that incorporate unnatural amino acids [29–31]. Cell-free protein expression can be accomplished using either cell lysates [32] or in a reconstituted transcription-translation system such as the PURE system [33].

A pre-requisite for directed evolution is a genotype-phenotype link. In cell-free systems, this link is often implemented through ribosomal, mRNA, or other cell-free macromolecular display technologies [34], although these techniques are often limited to evolution of peptide and protein binding affinities. For evolution of an enzyme's catalytic turnover, however, compartmentalization in emulsion droplets [35] or liposomes [36] is more appropriate. Such biomimetic compartments are also often used as the chassis for engineering towards construction of an artificial cell [37]. Finally, liposomes are exceptionally suited for evolution of membrane proteins, requiring a lipid bilayer for solubility and activity [27].

However, coupling the gene expression and enrichment steps in a cell-free system within a microcompartment is often limited by the low yield of synthetized proteins from a single DNA template. Although detectable activity of cell-free expressed proteins arising from a single gene copy has been demonstrated in some experimental conditions [27,38–40], it is hardly surprising that below a certain threshold, template DNA concentration is a limiting factor for in vitro protein expression [15,41–44]. In fact, production of full-length proteins in reconstituted systems ceases before NTPs and amino acids get depleted, and efforts to increase the

amount of protein from low DNA concentrations remain frustrated [45]. Therefore, clonal amplification of expression templates is a generic solution to enhance protein yield and activity readout, as well as the recovery of selected DNA variants.

A major challenge in cell-free directed evolution is the coupling of DNA amplification from single template copies with gene expression and quantitation of the activity of the protein of interest for fitness assignment in one environment. For example, rolling circle amplification (RCA) based on the Phi29 DNA polymerase and replication cycle reaction (RCR) based on a reconstituted *E. coli* replisome are compatible with droplet microcompartments [28,46]. However, RCR has not been combined with in vitro transcription-translation (IVTT) in a one-pot reaction yet, and PCR requires heating steps that are incompatible with IVTT in one-pot reactions. On the other hand, combination of RCA with gene expression is only possible after optimization of some components for transcription and translation to minimize cross-inhibition effects [47–50], proscribing the use of standard commercial kits for IVTT. Thus, so far, DNA templates cannot be amplified efficiently in the same solution where the cell-free system is performed. Therefore, multiple-step workflows have been implemented, which require droplet-based microfluidic handling [28,51–53] or bead-display [54–58].

In this study, we simplify the in vitro evolution methodology by a single isothermal, clonal amplification-enhanced gene expression, or CADGE. The strategy relies on the protein-primed replication machinery of the *Bacillus subtilis* bacteriophage Phi29 [59] consisting of DNA polymerase (DNAP, encoded by gene *p2*), terminal protein (TP, encoded by gene *p3*), double-stranded DNA-binding protein (DSB, encoded by gene *p6*), and single-stranded DNA-binding protein (SSB, encoded by gene *p5*), and requires prior flanking of the gene of interest (*GOI)* with Phi29 origins of replication (ori) using a standard recombinant DNA technique of choice. The Phi29 DNAP is chosen largely due to its strand-displacement activity, a relatively rare property for a family B DNA polymerase [60,61] This activity enables it to displace the non-template DNA strand at ambient temperatures, thus ensuring compatibility with cell-free transcription-translation. In addition, Phi29 DNAP has an excellent processivity [60,62], which could be useful for efficiently replicating long and multigene DNA templates. To initiate the replication, DNAP forms a complex with TP [63], and the heterodimer is recruited to the replication origins, a process that is facilitated by DSB [64]. DSB activates the replication initiation by forming a multimeric nucleoprotein complex at the origins of replication [65], whereas TP primes the DNA synthesis at each end, remaining covalently attached to the 5'-end of the daughter strand [66]. After successful priming, DNAP dissociates from the complex and continues the polymerization activity [67]. SSB is another auxiliary protein, which assists in the replication by stabilizing the displaced strand [68]. Using this system, we previously realized transcription-translation-coupled self-replication of a two-gene construct [18]. Herein, we demonstrate that transcription-translation-coupled amplification of orthogonal genes can be achieved in bulk and in liposome compartments, improving the expression level of a gene of interest (*GOI)*. As a proof-of-concept, we show

the enrichment of an ori-*GOI* from a mock library encapsulated in liposomes, a key step toward cell-free protein evolution. Moreover, we apply CADGE to enable the screening of protein functions that are relevant in the field of synthetic cell construction.

Results

Design of CADGE

The CADGE strategy involves the following minimal requirements **(Fig. 1a,b)**:

1. A *GOI* is inserted between the 191-bp-oriL and 194-bp-oriR origins of replication of the Phi29 genome, although the 68-bp minimal origins could potentially be used as well [69]. The DNA template must be linearized with the origins at each end of the molecule, which can be achieved by PCR amplification from ori-containing plasmid DNA. Moreover, the linear DNA has to be phosphorylated at each 5'-end, which can be done by using 5'-phosphorylated primers. One, two [18] or, in principle, more genes can be encoded on a single ori-flanked DNA template. Hereafter, we refer to such linear constructs as ori-*GOI*.

2. The PURE*frex*2.0 system is chosen for IVTT because of its higher purity and reduced nuclease activity compared to other commercial PURE systems [45]. Thus, the linear DNA construct contains regulatory elements compatible with gene expression in PURE*frex*2.0. These comprise a T7 promoter, g10 leader sequence, *E. coli* ribosome binding site and a transcription terminator (e.g. T7 and vesicular stomatitis virus terminators)

3. The system requires four minimal protein components of the Phi29 replication machinery: DNAP, TP, SSB, and DSB, plus dNTPs and ammonium sulphate for the efficient dimerization of the replication initiation complex [70] **(Fig. 1a,b)**. DNAP and TP can either be introduced in a purified form (purCADGE) or in situ expressed from a separate DNA construct (expCADGE). In the latter configuration, the two genes *p2* and *p3* are introduced on a single plasmid, self-replication being prohibited by the circular nature of the DNA. Although SSB and DSB can be functionally expressed in the PURE system [18], we recommend supplying them as purified proteins since they are required at micromolar concentrations and their cell-free expression would create a burden on the transcription-translation apparatus. The linear replication product in CADGE is essentially identical to the parental DNA molecule – except for the fact that TP is covalently bound at the 5'-end of each daughter strand. In the current protocol, the 5'-TP is lost with subsequent PCR amplification during recovery of the total DNA from liposomes. Thus, the resulting recovered DNA is identical in its structure to the original template DNA and does not require any additional processing between rounds of evolution.

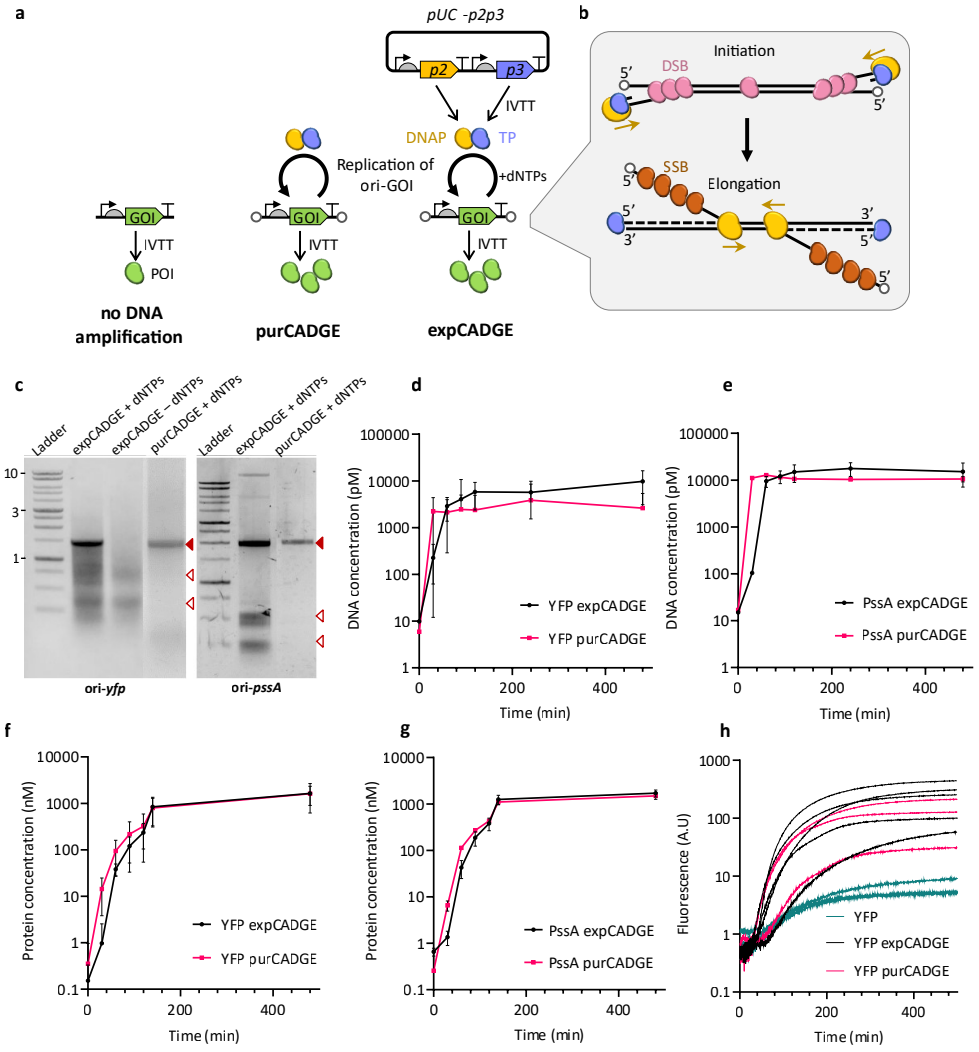


Fig. 1 Principles and validation of the CADGE strategy in bulk reactions. a) Schematics of the different gene expression configurations used in this study. **b)** Schematic of linear DNA replication by the Phi29 minimal DNA replication machinery. **c)** Visualization of amplified DNA in CADGE samples via agarose gel electrophoresis. Filled red arrowheads indicate expected product size, empty red arrowheads indicate unfinished smaller products. **d, e)** Time-course analysis of ori-*yfp* **(d)** and ori-*pssA* **(e)** DNA amplification in bulk CADGE reactions via absolute qPCR quantitation. **f, g)** Time-course analysis of YFP and PssA expression in bulk CADGE reactions via absolute LC-MS/MS quantitation. **h)** Kinetic measurements of YFP fluorescence from gene expression with and without CADGE. Different curves of same color are biological replicates.

Validation of CADGE in bulk reactions

We first evaluated the performance of CADGE on amplifying a *GOI* in bulk reactions. Two ori-GOI fragments encoding either the enhanced yellow fluorescent protein (eYFP) or *Escherichia coli* phosphatidylserine synthase (PssA) (ori-*yfp* and ori-*pssA*, respectively) were

constructed. The DNA template was added to PURE*frex*2.0 in the presence of the DNA replication machinery, dNTPs, and ammonium sulphate, and the solution was incubated at 30 °C. With both purCADGE and expCADGE, we found that the full-length DNA **(Fig. 1c)** can be amplified to saturation by two to three orders of magnitude from an input template concentration of 10 pM within two hours, as confirmed by absolute quantification by qPCR **(Fig. 1d,e)**. Although full-length replication products are the most abundant DNA species, shorter fragments corresponding to incomplete polymerization products are also visible in the gel, especially with expCADGE **(Fig. 1c)**.

To test if template amplification is accompanied with an increase in protein expression levels, we quantified the concentrations of eYFP and PssA by liquid chromatography-coupled mass spectrometry (LC-MS), and confirmed the production of both proteins to up to 1 mM, with no noticeable differences between purCADGE and expCADGE **(Fig. 1f,g)**. These amounts of protein expression were comparable to the generally reported cell-free protein synthesis levels [20,71,72], but with considerably (two to three orders of magnitude) less input of template DNA. The concentrations of YFP and PssA increased by at least 6-fold with expCADGE compared to unamplified gene conditions **(Fig. S1)(Fig. S2)**. Importantly, fluorescence kinetics measurements show that in the absence of DNA amplification, only a very low level of YFP was expressed even after several hours of incubation **(Fig. 1h)**. This finding indicates that the enhanced protein expression is the direct result of gene amplification. Kinetic analysis of protein synthesis gives apparent maximum translation rates (defined as the highest slope) comprised between 3.8 and 6.4 nM min^{-1} and a time before saturation of about 300 min **(Fig. 1f,g)(Table S1)**. These values are consistent with previous data obtained with nanomolar concentrations of DNA template [17,45], suggesting that CADGE does not significantly delay or slow down protein production. In fact, the apparent YFP production rate increases 25-fold with purCADGE and 40-fold with expCADGE compared to the condition with unamplified ori-*yfp* **(Fig. 1h)**.

CADGE improves phenotypic output in liposomes

We next sought to demonstrate that CADGE is able to increase the number of liposomes with detectable amounts of synthesized protein starting from clonal quantities of DNA molecules **(Fig. 2a)**. To this end, the construct ori-*yfp* and the CADGE components were encapsulated in a polydisperse population of liposomes, the bilayer of which is composed of biologically relevant lipids found in the composition of *E. coli* plasma membrane [18]. Ori-*yfp* was introduced at 10 pM bulk concentration, corresponding to an expected average number of DNA molecules per liposome $\lambda = 0.2$ (Methods section) if one assumes an average liposome radius of 2 µm [19]. To confine the IVTT and replication reactions to the interior of the liposomes, we introduced DNase I to the outer phase of the liposome population, which yielded a concentration of left-over DNA inside vesicles of around 100 fM **(Fig. 2b)**. The extent of clonal amplification was assessed by comparing endpoint data (typically overnight incubation at 30 °C) with (+) and without (–) dNTPs. To recover the DNA for analysis, we

heat-inactivated DNase I and released the DNA from the liposomes by dilution in water. Quantification by qPCR revealed that in both purCADGE and expCADGE, over 100-times more DNA was obtained in the presence of dNTPs than in the absence, thereof **(Fig. 2b)**, suggesting that DNA was amplified inside the liposomes.

To assess the effect of DNA amplification on gene expression, we analysed individual liposomes for YFP signal using flow cytometry **(Fig. 2c-e)**. We confirmed that under both CADGE configurations, and in the presence of dNTPs, a higher percentage of liposomes exhibited a YFP fluorescence above the background level **(Fig. 2c,d)**. This was expected from the strongly reduced protein expression level at low DNA concentrations **(Fig. S3)**. Interestingly, the mean intensity of YFP-expressing liposomes was about 5-fold higher in the presence of dNTPs compared to those in the absence but also compared to samples that contained none of the components of the minimal replication machinery, and the range of intensity values expanded across almost two orders of magnitude **(Fig. 2c,e)**. Similar observation was made from fluorescence imaging of individual liposomes **(Fig. 2f)**. These results suggest that clonal amplification does not only boost gene expression to exceed the threshold for measurable activity, but also increases the dynamic range of the phenotypic output. Although the percentage of YFP-expressing liposomes was slightly higher with purCADGE compared to expCADGE **(Fig. 2d)**, the intensity profiles were similar **(Fig. 2e)**, suggesting that co-expression of *p2* and *p3* genes does not significantly affect the production of protein of interest (POI) in liposomes. Similar conclusion could be reached from bulk reactions **(Fig. 1f,g)**.

We noticed that the percentage of YFP-expressing liposomes was lower in the –dNTPs sample when compared to the condition where replication reagents were omitted (YFP only)**(Fig. 2d)**. This suggests that some replication components may inhibit transcription-translation. We tested this hypothesis by varying DSB and SSB concentrations in ori-*yfp* bulk reactions and found that reduced amounts of DSB led to higher expression of ori-*yfp*, while changing SSB concentrations had little effect **(Fig. S4)**. Considering that DSB is a Phi29 transcription regulator of early and late genes [73], it is possible that binding of DSB to the DNA template inhibits gene expression in vitro. Therefore, we decided to lower DSB concentration down to either 52.5 or 105 µg/mL, in order to mitigate inhibitory effects without compromising DNA replication.

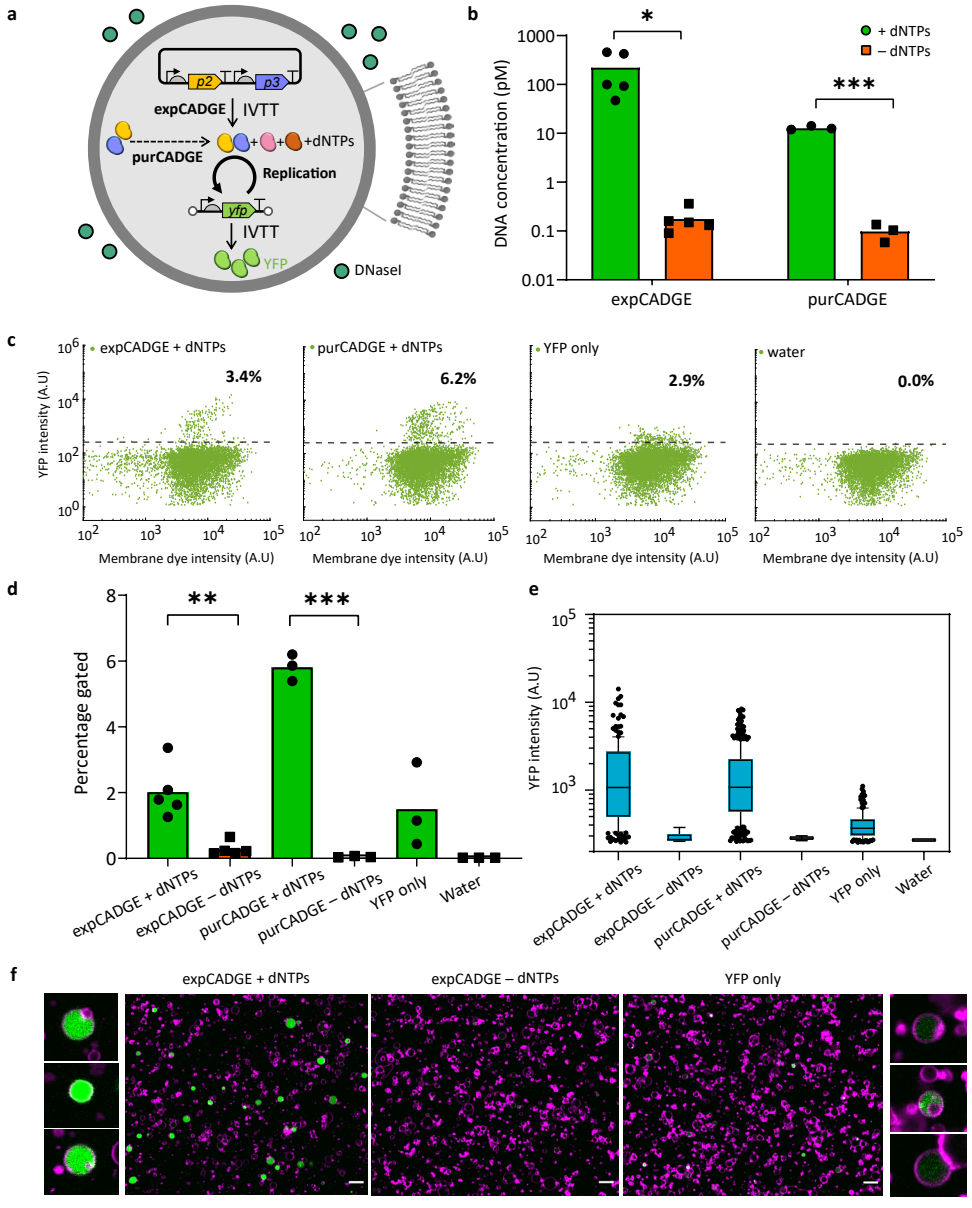


Fig. 2 Effect of in-liposome CADGE on the phenotypic output of a reporter gene. a) Schematic of in-vesiculo reporter gene amplification and expression via CADGE. **b)** Absolute quantitation of ori-*yfp* DNA by qPCR in lysed liposomes. Total 10 pM input template DNA concentration was used, which was reduced due to externally supplied DNase I. **c)** Populational variation of in-vesiculo YFP fluorescence in CADGE samples measured by flow cytometry. **d)** Quantitation of the fraction of liposomes showing above-background YFP fluorescence estimated by the horizontal gate in (c). **e)** Histogram representation of flow cytometric data of YFP expression in CADGE liposomes. **f)** Confocal microscopy imaging of CADGE and un-amplified samples. Magenta: Texas Red (membrane dye). Green: YFP. Scale bars: 5 µm. *P ≤ 0.05; ** P ≤ 0.01; *** P ≤ 0.001.

CADGE with a positive feedback loop

We then implemented expCADGE with a positive feedback loop coupling POI synthesis back to DNA replication. The autocatalytic framework of this selection strategy may offer a more effective and efficient alternative to fluorescence-based screening methods. Ori-*p3* template coding for TP was introduced at 10 pM concentration (λ = 0.2), supplemented with an excess amount of plasmid encoding solely the DNAP **(Fig. 3a)**, and encapsulated in liposomes. We hypothesized that an initial seed expression of TP could kick off the replication of ori-*p3* with the expressed DNAP and yield increasing amounts of ori-*p3*. Quantitative PCR showed that the *p3* gene was amplified inside liposomes by three orders of magnitude in the presence of dNTPs **(Fig. 3b)** compared to the –dNTPs control. The DNA intercalating dye dsGreen was used as a fluorescent marker to assess DNA amplification in single vesicles by flow cytometry. A fraction of liposomes with increased dsGreen fluorescence **(Fig. 3c,d)** compared to the background was detected in the presence of dNTPs, which corroborates that self-amplification of DNA took place.

The high amplification of ori-*p3* prompted us to experimentally determine the bulk concentration of DNA template below which the amplification is 'clonal'. Experimental validation of the λ = 1 regime is important considering the polydispersity of the liposome population, which differs from our assumption of constant volume (Methods section). Therefore, we performed a mock enrichment experiment by co-encapsulating ori-*p3* and an equimolar amount of unrelated DNA, also flanked with replication origins (here ori-*p6*) **(Fig. 3e)**. In this scenario DNA replication is conditional to the presence of both DNAP and TP. Therefore, ori-*p6* can only be replicated when co-encapsulated with ori-*p3*, i.e. under non-clonal conditions where more than a single molecule of ori-GOI is present in a liposome. Conversely, an enrichment of ori-*p3* over ori-*p6* would indicate that amplification is mostly clonal. After a single round of selection, ori-*p3* was enriched 114-fold and 37-fold over ori-*p6* when starting with 10 pM and 50 pM DNA concentrations, respectively **(Fig. 3f,g)**. This result confirms that in-liposome amplification of ori-GOI is mostly clonal and that CADGE is suitable for in vitro directed evolution purposes.

Amplification of ori-*p6* was however not totally prohibited, even at 10 pM input mixture concentration **(Fig. 3f)**. The latter is not unexpected considering that the estimated probability of co-occupancy of the ori-*p3* and ori-*p6* templates is not zero but is $(1\text{-}e^{-\lambda})^2$ = 0.15) with 50 pM input mixture concentration (λ = 0.5 for each ori-GOI). Together, the significant enrichment of ori-*p3* over ori-*p6* experimentally validates that 10 pM (and to some extent 50 pM) concentration is enough to keep a strong genotype to phenotype link in our polydisperse liposome population. This experiment also implies that, as long as DNA replication can be coupled to a POI activity, TP or any other POI can be potentially evolved using this selection scheme.

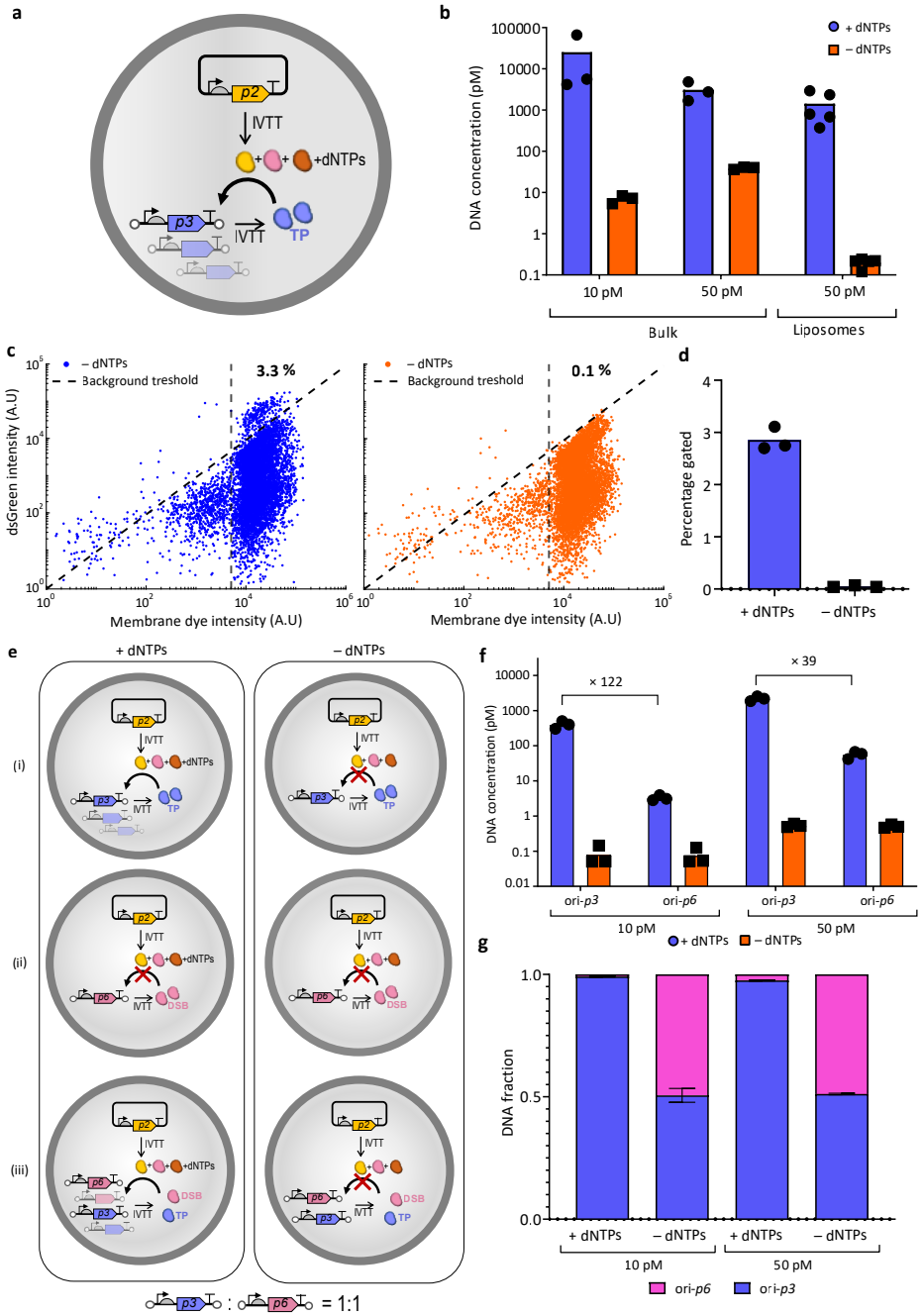


Fig. 3 CADGE with a positive feedback loop. a) Schematic of in-vesiculo ori-*p3* DNA amplification and expression via CADGE. **b)** Absolute quantitation of ori-*p3* DNA by qPCR in lysed liposomes. Total 10 pM input template DNA concentration was used, which was reduced due to externally supplied DNase I. **c)** Populational variation of dsGreen fluorescence in CADGE liposomes measured by flow cytometry. **d)** Quantitation of the fraction of liposomes showing above-background dsGreen fluorescence estimated by the diagonal gate in (c). **e)** Schematics of the different gene expression configurations used in (f) and (g). **f)** Absolute DNA quantitation of two genes by qPCR of liposome suspensions after a single round of selection. Total 10 pM or 50 pM input template DNA concentration was used, which was reduced due to externally supplied DNase I. **g)** Calculated fractions of the two genes in the mixture after a single round of selection. Error bars indicate standard deviation from three biological replicates.

CADGE improves enrichment efficiency of a GOI based on high-throughput screening

Next, we asked whether CADGE may be beneficial for in vitro protein evolution via fluorescence-based screening. To this end, we performed a mock enrichment experiment at a clonal expression condition with 10 pM ori-GOI, i.e. $\lambda = 0.2$. We aimed to enrich the DNA template ori-*yfp* from a mock library containing an excess of the unrelated template ori-*minD* based on the fluorescence of expressed YFP by fluorescence activated cell sorting (FACS). For this, the ori-*yfp* DNA template was mixed with 10-fold excess of ori-*minD* template and the DNA/PURE mixture was encapsulated in liposomes at a total 10 pM DNA concentration **(Fig. 4a)**. At such a low template DNA concentration (1 pM of ori-*yfp*), expression of YFP is significantly reduced compared to higher DNA concentrations typically used in cell-free reactions, leading to low signal-to-noise ratio **(Fig 4b)**. As expected from previous results, CADGE liposomes exhibited higher dynamic range of YFP fluorescence compared to liposomes that contained the same input DNA mixture concentration but no replication factors **(Fig. 4b)**. Up to 3-fold increase of the mean intensity of YFP-positive liposomes was measured upon gene amplification **(Fig. 4d)**. For sorting, two stringency conditions were tested: the "all-gate", which encompassed the top 1% of all the liposomes (applied to both non-amplified and CADGE samples), and the "high-gate", which included only the top (0.2%) of the high-intensity liposomes (applied to CADGE samples only). It was reproducibly difficult to recover the full-length DNA by PCR from the non-amplified liposome samples, while full-length DNA from liposomes with implemented CADGE was easily recovered **(Fig. 4e)**. This finding can be explained by higher DNA titers in the sorted liposomes from CADGE samples. Indeed, as assayed by qPCR, ori-*yfp* and ori-*minD* mixtures in liposomes were considerably (both more than a 100-fold) and uniformly (i.e. two genes amplified equally in a single sample) amplified with both purCADGE and expCADGE (pre-sorted samples)**(Fig 4f-h)**. Furthermore, qPCR analysis of sorted liposome samples gave the enrichment efficiency of ori-*yfp* over ori-*minD* **(Table S2)**. Using the more stringent condition of "high-gate" in CADGE samples results in improved purity of YFP sorting compared to "all-gate" in both CADGE and non-amplified samples **(Fig. 4h)(Fig. S5).**

Fluorescent proteins expressed from single copies of templates in biomimetic compartments can only be enriched several-fold per sorting round [36], likely due to low signal-to-noise ratios. Our findings show that CADGE improves enrichment efficiency by enabling selection of liposomes with more stringent fluorescence criteria (enrichment efficiency can reach 89 compared to 31 without amplification)**(Table S2)** and DNA recovery in a single round of mock enrichment experiment, and thus suggest that CADGE may facilitate in vitro protein evolution.

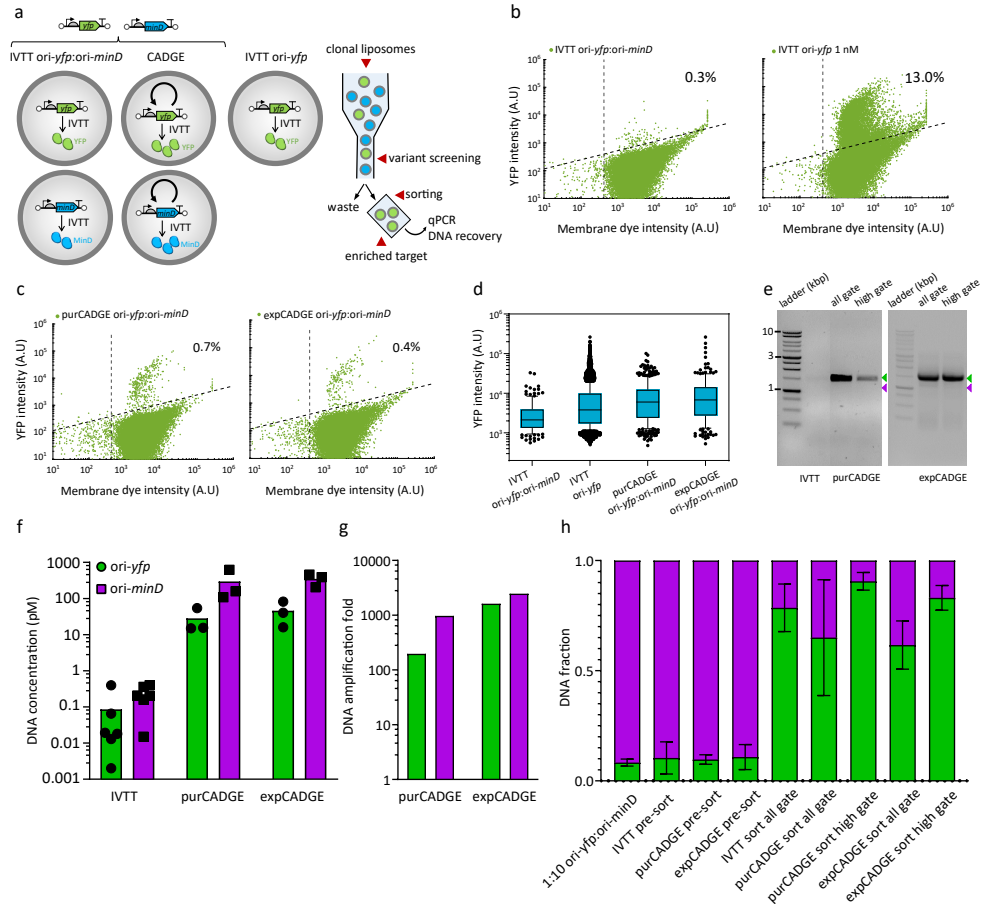


Fig. 4 Enrichment of clonally amplified GOI via high-throughput FACS screening. a) Schematics of clonal gene expression and enrichment experimental design for FACS. **b)** Flow cytometry analysis of liposomes prepared from indicated DNA template mixtures in the PURE system: 10 pM total with 1:1 ori-*yfp*:ori-*minD* DNA mixture (left) and 1 nM ori-*yfp* DNA (right). **c)** Flow cytometry analysis of CADGE liposomes prepared from 10 pM total input ori-*yfp*:ori-*minD* template mixture. **e)** Agarose gel electrophoresis images of recovered DNA from sorted CADGE liposomes. Green arrowhead indicates the expected DNA size for ori-*yfp* (2 kb), purple arrowhead indicates the expected size for ori-*minD* (~ 1.4 kb). **f)** Absolute DNA quantitation by qPCR of pre-sort liposome suspensions. **g)** DNA amplification in pre-sort purCADGE and expCADGE liposome samples. Amplification was calculated as DNA concentration of a specific gene in end-point samples with dNTPs compared to end-point samples without dNTPs. Colour coding is the same as in (f). **h)** Fractions of ori-*minD* and ori-*yfp* DNA mixtures recovered from pre-sort or post-sort liposomes as calculated from absolute DNA quantification by qPCR. Error bars indicate standard deviation from three to seven biological replicates.

CADGE improves phenotypic output of synthetic cell modules

We previously proposed in vitro evolution as a route to build a synthetic minimal cell [12]. Here, we seek to exploit CADGE for improving the expression of genes that are relevant for the construction of functional cellular modules. One candidate gene is *pssA* from the Kennedy phospholipid biosynthesis pathway [74,75]. The *E. coli pssA* gene encodes an

enzyme that conjugates cytidine diphosphate-diacylglycerol (CDP-DAG) with L-serine to produce cytidine monophosphate and phosphatidylserine (PS) **(Fig. 5a)**, a precursor of phosphatidylethanolamine. To assay the activity of in vesiculo synthesized PssA enzyme, we encapsulated PURE and the DNA encoding the *pssA* gene in phospholipid liposomes containing 5 mol% CDP-DAG and digested the external liposomal DNA with DNase I **(Fig. 5b)**. Since PssA is active as a membrane-associated enzyme, PS would be incorporated on the inner leaflet of the membrane [76,77]. However, as previously suggested [19], we expected some flipping of phospholipids to the outer membrane, such that the enzymatic activity of entrapped PssA could be detected by externally staining the liposomes with a PS-specific probe. To this end, we implemented C2-domain of lactadherin protein (LactC2) fused to a fluorescent protein like mCherry or eGFP **(Fig. 5b)** [19]. By flow cytometric analysis of LactC2-mCherry- and Acridine Orange- (membrane marker) stained liposomes, we observed that PS production (and, we assumed, gene expression) reduces considerably at limiting template DNA concentrations (10 and 50 pM DNA) compared to 1 nM **(Fig. 5c,d)(Fig. S6)**. Alternatively, we stained the liposomes with LactC2-eGFP and Texas Red (membrane dye), and imaged them by confocal microscopy **(Fig. S7)**. We found that limiting the template DNA concentration visibly reduces LactC2 binding, suggesting that pssA gene expression is diminished at low input DNA concentrations.

To test if clonal DNA amplification can improve PssA expression, we co-encapsulated the linear ori-*pssA* DNA fragment (10 or 50 pM) with the required additives for either purCADGE or expCADGE **(Fig. 5e)** and incubated at 30 °C for 4 hours. Using qPCR, we confirmed 10- to 100-fold amplification of the *pssA* gene compared to −dNTP controls with input ori-*pssA* concentrations of 10 pM, in both CADGE configurations **(Fig. 5f)**. Even though PS synthesis was detectable in −dNTP samples, the number of liposomes exhibiting a PS-positive phenotype and mean intensity of recruited LactC2-eGFP increased with functional CADGE (+dNTPs) **(Fig. 5g-i)(Fig. S8, S9)**. Overall, these findings demonstrate that, within a synthetic cell context, clonal amplification of template DNA can improve phospholipid headgroup conversion from in vitro expressed PssA protein.

Besides gene-directed phospholipid production, we decided to explore the benefit of CADGE for implementation of the Min system in clonal liposomes. The Min system is involved in the spatial organization of cytokinesis events in *E. coli* [78] and is therefore a relevant protein system for synthetic cell division. MinD is an ATP-dependent membrane binding protein that recruits MinC, an FtsZ-polymerization inhibitor [79]. We assembled expCADGE reactions with 10 pM ori-*minD* DNA and purified eGFP-MinC as a reporter of the activity of synthesized MinD **(Fig. 6a)**, and encapsulated the mixture in liposomes. Quantitative PCR data showed that ori-*minD* was clonally amplified almost a thousand-fold compared to −dNTPs control samples **(Fig. 6b)**. Confocal imaging and analysis of eGFP-MinC fluorescence distribution in the lumen and at the membrane revealed that in the vast majority of the liposomes, the basal amount of expressed MinD is not sufficient to recruit eGFP-MinC to the membrane (−dNTPs)**(Fig. 6c-e)**. Using expCADGE, a larger fraction of

liposomes exhibited an excess fluorescence signal at the membrane (+dNTPs)**(Fig. 6c-e)**, indicating that clonal amplification led to a re-localization of MinC through improved expression of functional MinD.

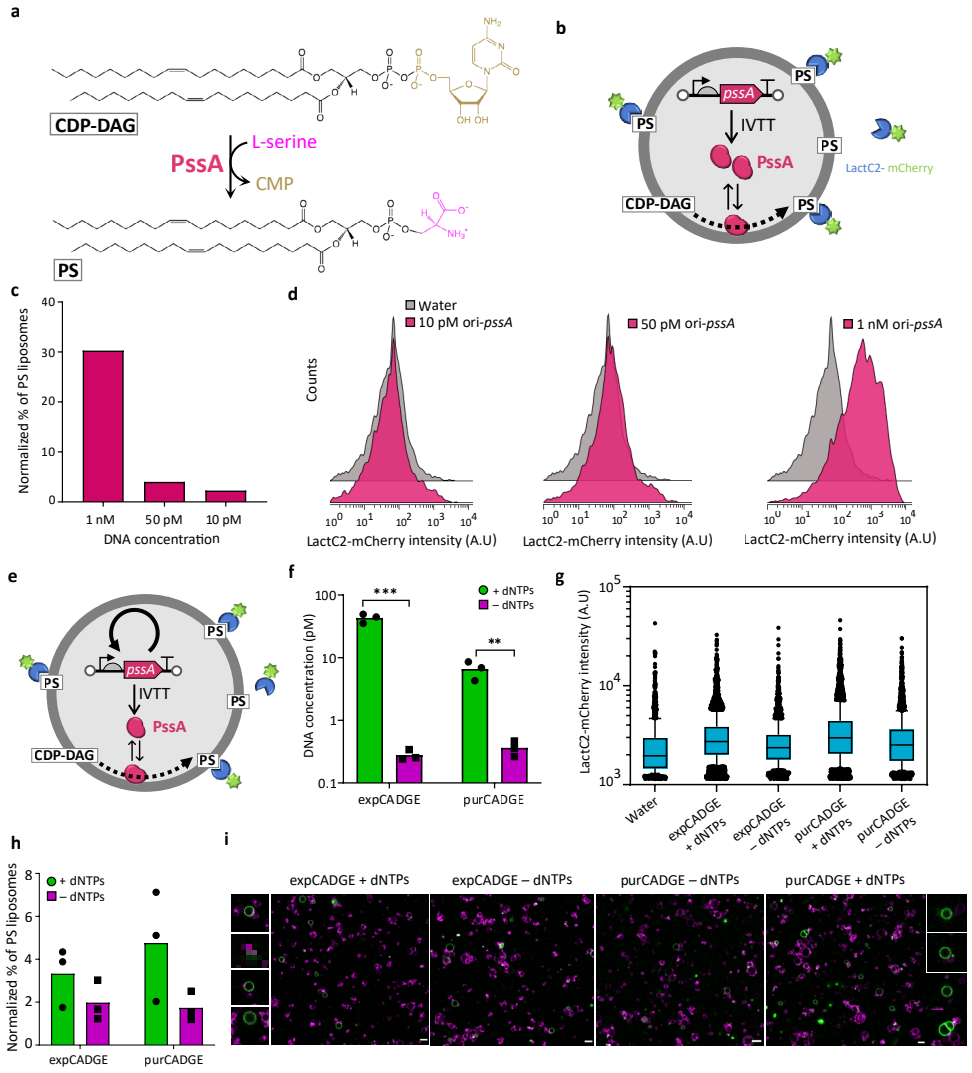


Fig. 5 Application of CADGE for improving the enzymatic catalysis of phospholipid headgroup conversion. a) Schematic of CDP-DAG conversion to PS by PssA. **b)** Schematic of in-vesiculo-expressed PssA enzyme activity and detection of PS-positive liposomes by LactC2-mCherry binding. Percentage quantitation **c)** and histograms **d)** of PS-positive liposomes expressing PssA in the PURE system, as assayed by flow cytometry (raw data in Fig S6). **e)** Schematic of CADGE liposomes expressing PssA enzyme and detection of PS-positive liposomes by LactC2-mCherry binding. **f)** Absolute DNA quantitation by qPCR of lysed CADGE liposome samples. ** $P \leq 0.01$; *** $P \leq 0.001$. **g)** Histograms and **h)** quantitation of PS-positive CADGE liposomes expressing PssA, as assayed by flow cytometry (raw data in Fig S8, S9). **i)** Confocal microscopy of CADGE liposomes expressing PssA. Green, LactC2-eGFP; Magenta, Texas Red-conjugated lipids. Scale bars: 5 µm.

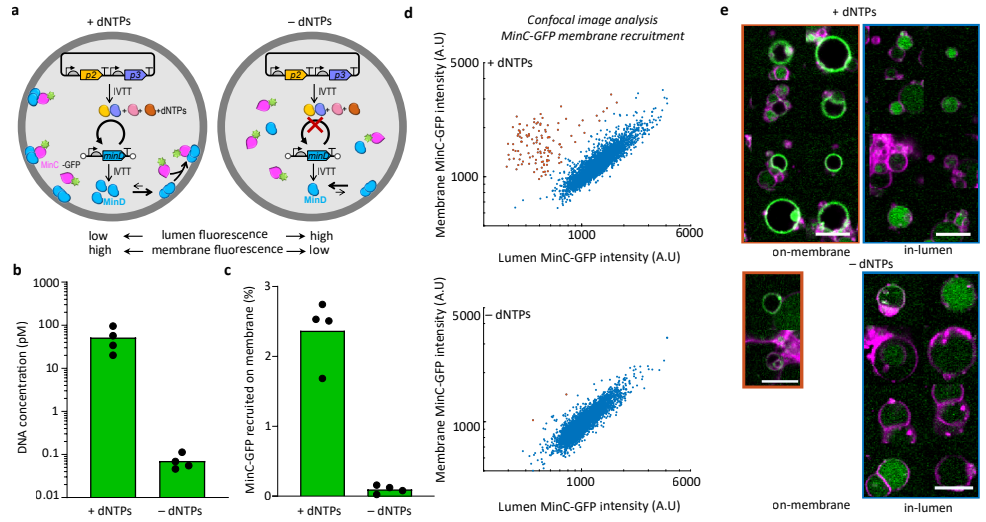


Fig. 6. Application of CADGE for facilitating the membrane recruitment of Min proteins. a) Schematics of MinC-MinD membrane binding assay in expCADGE liposomes. **b)** Absolute quantitation of ori-*minD* DNA by qPCR of lysed expCADGE liposome samples. Individual symbols are independent biological repeats. **c)** Percentage of MinC-GFP recruited to the membrane, as obtained from fluorescence confocal images. **d)** Scatter plots of the membrane and lumen fluorescence intensities in individual liposomes. The liposomes marked in orange dots display a marked recruitment of MinC-GFP to the membrane. **e)** Montage of randomly selected liposome images taken from the data shown in **d)**. Scale bars: 5 µm.

Discussion

Herein, we established CADGE, a single-step DNA amplification and in situ transcription-translation strategy that can be used for improving clonal gene expression. Our findings suggest that CADGE can be instrumental in facilitating in vitro evolution of a variety of genes, including those that are important for synthetic cell construction. The general applicability of this strategy is enabled by only a few requirements: the possibility of in vitro expression of the *GOI*, and in vitro activity of the POI.

Advantages of CADGE for clonal amplification, compared to previous strategies [28,51–56,80] include minimum effort (i.e. a single-step amplification and expression), time (around two hours to set up the reaction), and instrumentation (although microfluidic devices for micro compartmentalization or screening can be implemented, if necessary). The benefit of using CADGE post-enrichment lies in the simplicity of the protocol since only PCR is required to proceed to another round of encapsulation/enrichment, and in the improved DNA recovery yield.

Despite a number of advantages, CADGE is not without some limitations. Thus, DSB was found to exhibit some inhibitory effect on gene expression, which could be mitigated to some extent by optimized DSB concentrations **(Fig. S2)**. Moreover, in our fluorescence measurements of CADGE samples, the measured percentage of liposomes with a

fluorescence signal above the activity threshold (3 to 5% with the studied ori-*GOI*) is lower than the predicted value of $(1\text{-}e^{-\lambda}) \times 100 = 0.18\%$, with $\lambda = 0.2$ at 10 pM DNA, if one assumes that all liposomes with at least one DNA copy would give a signal and that all liposomes have a fixed diameter of 4 µm. More accurate percentage values calculated from histograms of liposome sizes in different samples are reported in **Table S3**, but the conclusions are unchanged. The discrepancy between the measured and predicted fraction of 'active' liposomes suggests that (i) some DNA molecules are transcriptionally inactive or depleted into the lipid film [44], thus lowering the apparent λ, (ii) only liposomes with particularly high concentration of amplified DNA or reporter protein cross the fluorescence detection threshold in our flow cytometer-based activity assay, or (iii) the encapsulation of input DNA does not follow a Poisson distribution. We suspect that under some conditions **(Fig. 2d)** and **(Fig. 5h)** (with ori-*yfp* and ori-*pssA*, respectively), competition for resources during *p2* and *p3* expression may limit the yield of synthesized POI in expCADGE. Therefore, optimization of the concentration of the *p2-p3* expression plasmid might be necessary for effective channelling of resources toward expression of POI. This drawback may be alleviated to some extent by using purified DNAP and TP (purCADGE), see **(Fig. 2)** and **(Fig. 5)**. Another limitation is the current lack of commercial availability of some of the required components, such as purified TP, SSB, or DSB. However, TP (together with DNAP) can be expressed from a plasmid in situ (expCADGE), while Phi29 SSB might in principle be replaced with a commercially available alternative (such as *E. coli* SSB), provided that it is shown to be compatible. Overall, we recommend that the optimality of expCADGE vs purCADGE, as well as the optimal DSB and *p2-p3* expression plasmid concentrations should be determined on a case-by-case basis for each *GOI*. We also believe that the yield of POI production per DNA template could be further improved through buffer optimization, in particular the concentrations of magnesium and potassium glutamate [81], tRNAs and NTPs [49].

PURE*frex*2.0 in vitro transcription-translation system was used here in its standard composition. However, other promoters than T7, such as the bacteriophage SP6 [19,82], T3 [83] or native *E. coli* [84] promoters, could also be used in combination with their cognate RNA polymerase supplied in the reaction mixture. One challenge may reside in the management of collision events between the Phi29 DNA polymerase and the RNA polymerases originating from different organisms [85,86]. Moreover, cell lysates, especially from *E. coli* [87], could be utilized as a cheaper cell-free expression system, in particular for biomanufacturing purposes [32]. The extract could be modified to avoid the degradation of linear PCR fragments by exonucleases, for instance by supplementing inhibitors of RecBCD (ExoV), such as GamS protein [88] or c-DNA oligonucleotides [89] or using Δ*RecBCD E. coli* strains 90. Alternatively, purified TP-bound DNA [18] could be used as a template in cell lysates, assuming that the parental TP hinders exonuclease digestion. Application of CADGE in eukaryotic cell extracts – e.g. from insect cells, wheat germ, rabbit reticulocytes, and human cells – might be useful for the production and engineering of proteins with post-translational modifications, such as glycosylation and phosphorylation. While protein yields may remain low even with clonal gene amplification compared to *E. coli*-based cell-free systems, the increased amount of DNA may be sufficient for the recovery of interesting gene variants. In general, codon usage of the *GOI* may be optimized for the chosen cell-free translation system, which should not influence much the DNA replication efficacy given the high template tolerance of the Phi29 DNAP.

In the shown examples, the genotype-to-phenotype coupling was established using phospholipid vesicles. Liposomes are uniquely suited for cell-free evolution of peripheral and transmembrane proteins [27,39], and for their tuneable membrane permeability, which could be relevant to assay the activity of POI through external addition of substrates or cofactors. The lipid-coated bead approach for liposome production [91] was chosen for its simplicity as it does not require specialized equipment, for the easy storage and distribution across laboratories of pre-assembled lipid films deposited on glass microbeads, and for its high biocompatibility due to the absence of organic solvent. One of the major drawbacks of using liposomes prepared by the swelling method is the heterogeneity of liposome sizes **(Fig. S10)** and encapsulation of DNA, PURE or CADGE components. Other methods for the preparation of more homogeneous liposomes in size and encapsulation efficiency, such as enhanced continuous droplet interface crossing encapsulation [92] and double-emulsion microfluidics [93], could in principle be utilized as well.

Other types of microcompartments can also potentially be combined with CADGE: water-in-oil emulsion droplets [52], microfabricated chambers [40], and peptide-based compartments [94], Emulsion droplets are particularly appealing for their highly monodisperse sizes and because they have already been integrated in microfluidic-based screening platforms for directed evolution of water-soluble enzymes [28,95].

Application of gene expressing liposomes empowered with clonal amplification is also relevant to build a synthetic cell from the ground-up. When applied to essential genes, CADGE-assisted directed evolution might accelerate the optimization of individual cellular modules and their integration to achieve higher level functions [12]. Considering the excellent processivity of Phi29 DNAP [60,62], the application of CADGE to long synthetic genomes can reasonably be envisaged. Through the example of TP **(Fig. 3)**, we showed the implementation of a positive feedback loop, where the *GOI* can itself assist in its own amplification, thereby circumventing the need for screening. This reaction scheme may in principle be expanded to self-amplification of polymerases and gene circuits based on DNA polymerization, such as in compartmentalized self-replication [96] and compartmentalized partnered replication [97]. Moreover, self-organization and catalytic activity of the peripheral membrane proteins MinD and PssA were detectable by isothermal DNA amplification from clonal amounts. This strategy might be particularly useful for the in vitro evolution of cellular functions starting from a single copy of ori-*GOI* library variants encapsulated in liposomes. The replicating template may contain single or multiple genes encoding entire pathways and

multiprotein complexes. For instance, application of CADGE to phospholipid-synthesizing enzymes of the Kennedy pathway located upstream (PlsB, PlsC and CdsA) and downstream (Psd) of PssA, could aid in optimizing synthetic cell growth through directed evolution.

Finally, we anticipate that performing CADGE under mutagenic conditions could extend its utility for in situ library production. For example, a mutator DNA polymerase [98] or mutagenic factors, such as Mn^{2+} and dNTP analogues, could be employed for genetic diversification directly within liposomes, bypassing the step of external gene library preparation. Such an error-prone CADGE strategy might be particularly interesting for introducing random mutations across long (> 10 kbp) DNA templates, for instance large synthetic genomes for the evolutionary construction of a minimal cell [12].

Materials and Methods

Buffers and solutions.

All buffers and solutions were made using Milli-Q grade water with 18.2 MΩ resistivity (Millipore, USA). Chemicals were purchased from Sigma-Aldrich unless otherwise indicated.

Construct design.

G365 (pUC-ori-*yfp*) was constructed by subcloning of the YFP gene (amplified by primers 1106 ChD/1107 ChD from plasmid G79) into Phi29 origins-containing vector G9618 (amplified by primers 1104 ChD /1105 ChD) via the Gibson Assembly method [99]. G368 (pUC-ori-*pssA*) was cloned by subcloning of the *pssA* gene (amplified by primers 1115 ChD /1116 ChD from plasmid G149) into Phi29 origins-containing vector G96 (amplified by primers 1104 ChD /1105 ChD) via the Gibson Assembly method. Plasmid G338 (pUC-ori-*p3*) was obtained as a result of subcloning the fragment ori-*p2p3*, which was PCR-amplified from plasmid G95 (plasmid encoding for ori-*p2p3*) [18] using the primers 961 ChD /962 ChD (with overhangs containing KpnI and HindIII restriction sites) into the KpnI-HindIII-linearized pUC19 vector, during which a spontaneous recombination event flipped out the entire (T7) promoter-*p2*-(vsv)terminator fragment, only leaving the shorter oriL-(T7)promoter-*p3*-(t7) terminator-oriR insert. G437 (pUC-ori-*minD*) was obtained by subcloning the *minD* gene (amplified by primers 91 ChD /397 ChD from plasmid pUC57-*minD*) [20] into the Phi29 origins-containing vector G365 (amplified by primers 535 ChD/562 ChD) via the Gibson Assembly method. The cloning of G85 (pUC57-*p2*) was previously reported [18]. All the plasmids were cloned by heat-shock transformation of *E. coli* Top10 strain, and plasmids were extracted from individual cultures outgrown in LB/ampicillin (50 µg/mL) using the PURE Yield Plasmid Miniprep kit (Promega). Individual clones were screened and confirmed by Sanger sequencing at Macrogen-Europe B.V. Primer sequences and plasmid descriptions can be found in the **Table S4** and **Table S5.**

Linear DNA templates were prepared by PCR using 5'-phosphorylated primers (491 ChD /492 ChD). Reactions were set up in 100 µL volume, 500 nM each primer, 200 µM dNTPs,

10 pg/µL DNA template, and 2 units of Phusion High-Fidelity DNA Polymerase (NEB) in HF Phusion buffer, and thermal cycling was performed as follows: 98 °C for 30 sec for initial denaturation, and thermal cycling at (98 °C for 5 sec, 72 °C for 90 sec) × 20, and final extension at 72 °C for 5 min. Extra care was taken to not over-amplify the DNA by too many thermal cycles, as it was found to adversely affect the quality of purified DNA. The amplified PCR fragments were purified using QIAquick PCR purification buffers (Qiagen) and RNeasy MinElute Cleanup columns (Qiagen) using the manufacturer's guidelines for QIAquick PCR purification, except for longer pre-elution column drying step (4 min at 10,000 g with open columns), and elution with 14 µL ultrapure water (Merck Milli-Q) in the final step. The purified DNA was quantified by Nanodrop 2000c spectrophotometer (Isogen Life Science) and further analysed for size and purity by gel electrophoresis.

Purification of DNAP, TP, SSB, DSB, LactC2-eGFP and LactC2-mCherry.

Purified Phi29 DNA replication proteins were produced as described in [18]. Stock concentrations and storage buffers are: DNAP (320 ng/µL in 50 mM Tris, pH 7.5, 0.5 M NaCl, 1 mM EDTA, 7 mM β-mercaptoethanol (BME), 50% glycerol), TP (400 ng/µL in 25 mM Tris, pH 7.5, 0.5M NaCl, 1 mM EDTA, 7 mM BME, 0.025% Tween 20, 50% glycerol), SSB (10 mg/mL in 50mM Tris, pH 7.5, 60 mM ammonium sulphate, 1 mM EDTA, 7 mM BME, 50% glycerol), DSB (10 mg/mL in 50 mM Tris, pH 7.5, 0.1 M ammonium sulphate, 1 mM EDTA, 7 mM BME, 50% glycerol). The proteins were aliquoted and stored at −80 °C. The DNAP and TP proteins were diluted before immediate use into PURE*frex*2.0 solution I (GeneFrontier). Both genes encoding for LactC2-eGFP and LactC2-mCherry were cloned into pET11 vector, under control of the T7-LacO promotor and in frame with an N-terminal His-tag. LactC2-mCherry was expressed in *E. coli* BL21(DE3) (NEB) and LactC2-eGFP protein was expressed in *E. coli* strain ER2566 (NEB). Overnight pre-cultures were prepared in Luria Broth (LB) medium containing 50 µg/mL ampicillin. The overnight cultures were diluted 1:100 in fresh LB medium with 50 µg/mL ampicillin and incubated at 37 °C while shaking, until an OD600 of 0.4-0.6 was reached. Protein expression was induced by adding 1 mM isopropyl β-D-1-thiogalactopyranoside. The cells were incubated at 26 °C for 4 h or overnight at 16 °C while shaking, and harvested at 4,000 × g for 15 min. Pellet of 1 L cells was resuspended in 10 mL lysis buffer (50 mM HEPES-KOH, pH 7.5, 500 mM NaCl, 10% glycerol). The cells were disrupted by sonication on ice, using 7 pulses of 30 seconds and 1 min intervals, with an amplitude of 40%. The cell suspension was centrifuged for 30 min at 30,000 × g at 4 °C to remove the cell debris. To the cell-free extract, 10 mM imidazole and SetIII protease inhibitor-EDTA-free (1:1000 dilution, Calbiochem) were added. The proteins were purified with HisPure Ni-NTA resin (ThermoScientific). The Ni-NTA (~3 mL) was equilibrated with buffer (50 mM HEPES-KOH, 500 mM NaCl, 10% glycerol, 10 mM imidazole, pH 7.5). The cell-free extract was mixed with the equilibrated resin and incubated for 1 to 16 h while tumbling in the cold room. After incubation the resin with bound protein was transferred into a gravity column, the unbound fraction was removed by gravity and subsequently the resin was washed with 20 equivalent volume wash buffer (50 mM HEPES-KOH, 500 mM

NaCl, 10% glycerol, 40 mM imidazole, pH 7.5). The protein was eluted with 5 mL elution buffer (50 mM HEPES-KOH, 500 mM NaCl, 500 mM imidazole, 10% glycerol, pH 7.5) and fractions of ~1 mL were collected. The fluorescent fractions were pooled together and buffer exchanged with storage buffer (50 mM HEPES-KOH, pH 7.5, 150 mM NaCl, 10% glycerol) using a 10-MWCO Amicon Ultra-15 centrifugal filter unit (Merck). The concentration of the protein was determined with a Bradford assay.

CADGE in bulk reactions.

Bulk reactions were set up in PURE*frex*2.0 (GeneFrontier). A 20-µL reaction consisted of 10 µL solution I, 1 µL solution II, 2 µL solution III, 20 mM ammonium sulphate, 300 µM dNTPs, 375 µg/mL purified Phi29 SSB protein, 105 µg/mL purified Phi29 DSB protein, 0.6 units/µL of Superase·In RNase inhibitor (Ambion), 10 pM target DNA and either plasmid DNA (2 nM plasmid G85 encoding for the *p2* gene in ori-*p3* clonal amplification experiments or 1 nM G340 encoding for *p2* and *p3* genes in ori-*yfp*, ori-minD, and ori-*pssA* experiments) or 3 ng/µL each purified Phi29 DNAP and TP. Reactions were incubated in a nuclease-free PCR tube (VWR) in a Thermal Cycler (C1000 Touch, Biorad) at a default temperature of 30 °C. Incubation time was indicated when appropriate.

To analyse the reactions by gel electrophoresis, 10 µL reaction was treated with 0.2 mg/mL RNase A (Promega), 0.25 units RNase One (Promega) at 30 °C for 2 h, followed by 1 mg/mL Proteinase K (Thermo Scientific) at 37 °C for 1 h, and column-purified using the QIAquick PCR purification buffers (Qiagen) and RNeasy MinElute Cleanup columns (Qiagen) using the manufacturer's guidelines for QIAquick PCR purification, except for longer pre-elution column drying step (4 min at 10,000 g with open columns), and elution with 14 µL ultrapure water (Merck Milli-Q) in the final step. A fraction (6 µL) of the eluate was mixed with an equal volume of 6× purple gel loading dye (NEB) and loaded in 1% agarose gel with ethidium bromide, following which DNA was separated using an electrophoresis system (Bio-Rad). The BenchTop 1-kb DNA Ladder (Promega) was used to estimate the size of DNA.

Mass spectrometry

LC-MS/MS analysis with QconCATs was employed for the absolute quantification of de novo synthesized proteins in bulk PURE reactions. Pre-ran PURE reaction solutions were mixed with one third volume of heavily labelled QconCAT(^{15}N) [72] in a 50 mM Tris (pH 8.0) buffer containing 1 mM $CaCl_2$. The samples were then incubated at 90 °C for 10 min and cooled down to 4 °C. Trypsin was then added at a 250 mg/mL final concentration and digestion incubation was carried out overnight at 37 °C. The trypsin digested samples were treated with TFA 10% and centrifuged for 10 min. The supernatant was then transferred to a glass vial with a small insert for LC-MS/MS analysis. Measurements were performed on a 6460 Triple Quad LCMS system (Agilent Technologies, USA) using Skyline software. [100] Samples of 5.5 µL were injected per run into an ACQUITY UPLC Peptide CSH C18 Column (Waters Corporation, USA). The peptides were separated in a gradient of buffer A (25 mM formic acid in Milli-Q water) and buffer B (50 mM formic acid in acetonitrile) at a flow rate of 500 µL per minute and at a column temperature of 40 °C. The column was initially equilibrated with 98% buffer A. After sample injection, buffer A gradient was changed to 70% (over the first 20 min), 60% (over the next 4 min), and 20% (over the next 30 sec). This final ratio was conserved for another 30 sec and the column was finally flushed with 98% buffer A to equilibrate it for the next run. The selected peptides and their transitions for both synthesized proteins and heavily labelled QconCATs were measured by multiple reaction monitoring (MRM). The recorded LC-MS/MS data was analysed with Skyline for fraction calculation between unlabelled and labelled peptides (^{14}N/^{15}N ratio) on both cell-free core/produced proteins and the initially added QconCATs. With these fraction values, and considering the regular concentration of core ribosomal peptides within PURE system (2 µM), we could estimate the concentration of the cell-free expressed proteins using the following equation

$$\frac{\left(\frac{^{14}\mathrm{N}}{^{15}\mathrm{N}}\right)_{\mathrm{PP}} \times 4}{\left(\frac{^{14}\mathrm{N}}{^{15}\mathrm{N}}\right)_{\mathrm{CPP1}} + \left(\frac{^{14}\mathrm{N}}{^{15}\mathrm{N}}\right)_{\mathrm{CPP2}}} \quad (1)$$

where PP refers to the detected peptide of produced protein, CPP1 refers to the detected peptide 1 of the core ribosomal protein (GVVVAIDK), and CPP2 refers to the detected peptide 2 of the core ribosomal protein (VVGQLGQVLGPR). MS/MS measurement details for each of the analysed proteins can be found in **Table S6**.

In-vesiculo protein expression.

The procedure was adapted from [18] with minor modifications. To prepare lipid-coated beads as precursors of liposomes with the standard lipid composition, in a 5-mL round-bottom glass flask, a primary lipid mixture was prepared consisting of DOPC (50.8 mol%), DOPE (35.6 mol%), DOPG (11.5 mol%), and 18:1 cardiolipin (2.1 mol%). The resulting mixture was additionally spiked with DSPE-PEG(2000)-biotin (1 mass%) and DHPE-Texas Red (0.5 mass%) for a total mass of 2 mg. Finally, the lipid mixture was complemented with 25.4 µmol of rhamnose (Sigma-Aldrich) dissolved in methanol. To prepare liposomes containing CDP-DAG, the primary lipid mixture composition was modified as following: DOPC (47.5 mol%), DOPE (34.2 mol%), DOPG (11.4 mol%), 18:1 cardiolipin (1.9 mol%), and CDP-DAG (5 mol%), with additional DSPE-PEG(2000)-biotin (1 mass%) and, if indicated, DHPE-TexasRed (0.5 mass%) for a total mass of 2 mg. All lipids were purchased at Avanti Polar Lipids and dissolved in chloroform, except the DHPE-Texas Red (Invitrogen). Finally, 600 mg of 212–300-µm glass beads (Sigma-Aldrich) were added to the lipid/rhamnose solution, and the organic solvent was removed by rotary evaporation at 200 mbar for 2 h at room temperature (rotary evaporator, Heidolph), followed by overnight desiccation. The dried lipid-coated beads were stored under argon at −20 °C until use. A 20-µL PURE*frex*2.0 reaction solution was assembled from 10 µL buffer solution, 1 µL enzyme solution, 2 µL

ribosome solution, and indicated amount of input DNA template in RNase-free Milli-Q water. To the well-mixed reaction, 10 mg lipid-coated beads, already pre-desiccated for at least 30 min before use, were added. The 1.5-mL Eppendorf tube containing the bead-PURE*frex*2.0 mixture was gently rotated on an automatic tube rotator (VWR) at 4 °C along its axis for 30 min for uniform liposome swelling. The mixtures were then subjected to four freeze/thaw cycles (5 sec in liquid nitrogen followed by 10 min on ice). From this step onwards, the liposome suspension was handled gently and only with cut pipette tips to prevent liposome breakage. Finally, 10 µL of the supernatant liposome suspension (the beads sediment to the bottom of the tube) was transferred to a PCR tube, where it was mixed with 0.5 units of DNase I (NEB). The reactions were incubated at 30 °C in a thermocycler for the indicated time periods.

In-vesiculo clonal amplification and expression of genes.

The liposome suspensions were assembled as above, except that the necessary CADGE components were pre-mixed with the PURE*frex*2.0 solution prior to the addition of the lipid-coated beads and swelling. The following compounds were supplemented (all final concentrations): 20 mM of ammonium sulphate, 0.75 U/µL SUPERase (Ambion), 10-50 pM template DNA (as indicated), 375 µg/µL purified SSB, 21-52.5 µg/µL of purified DSB, 1-3 ng/µL each of purified Phi29 DNAP and TP proteins or 250 pM of the *p2-p3* encoding G340 plasmid, and 300 µM of PCR Nucleotide mix (Promega). The liposome suspensions were incubated at 30 °C in a thermocycler for the indicated time periods.

Quantitative PCR.

Upon completion of the bulk or in-liposome CADGE reactions at 30 °C, 2-µL samples were harvested, and heated at 75 °C in the thermocycler for 15 min to inactivate the DNase I, and diluted 100-fold in Milli-Q water prior to addition to the qPCR mixtures. Ten microliter reactions consisted of Power-UP SYBR Green Master Mix (Applied Biosystems), 400 nM each primer (1121 ChD/1122 ChD for *yfp*, 980 ChD/981 ChD for *p3*, 1125 ChD/1126 ChD for *pssA*, 1208 ChD/1209 ChD for minD), and 1 µL of diluted sample. The thermal cycling and data collection were performed on Quantstudio 5 Real-Time PCR instrument (Thermo Fisher), using the thermal cycling protocol 2 min at 50 °C, 5 min at 94 °C, (15 sec at 94 °C, 15 sec at 56 °C, 30 sec at 68 °C) × 45, 5 min at 68 °C, followed by melting curve from 65 °C to 95 °C. The concentration of nucleic acids was calibrated using 10-fold serial dilutions of corresponding standard DNA templates ranging from 1 fM to 1 nM. Data were analysed using the Quantstudio Design and Analysis software v1.4.3 Software (Thermo Fisher).

Flow cytometry.

The liposome suspension (1-3 µL) was diluted in 300 µL (final volume) PB buffer consisting of 20 mM HEPES-KOH, pH 7.6, 180 mM potassium glutamate, and 14 mM magnesium acetate. To remove any remaining beads or large debris, the diluted liposome suspension was gently filtered through the 35-µm nylon mesh of the cell-strainer cap from the 5-mL

round-bottom polystyrene test tubes (Falcon). When indicated, dsGreen (Lumiprobe) dye was added at a 1:100,000 stock concentration to stain dsDNA, or Acridine Orange (6 µM) and LactC2-mCherry protein (300 nM) were added to stain the liposome membrane and phosphatidylserine, respectively. The mixture was incubated for 1 h at room temperature to equilibrate binding. Liposomes were screened with the FACSCelesta flow cytometer (BD Biosciences) using the 488-nm laser and 530/30 filter for detection of dsGreen, GFP, YFP, or Acridine Orange, and the 561-nm laser and 610/20 filter for detection of PE-Texas Red or mCherry. The following acquisition parameters were used: photon multiplier tube voltages set at 375 V for forward scatter, 260 V for side scatter (SSC), dsGreen detection at 500 V, PE-Texas Red detection at 300-370 V, YFP detection at 550 V, GFP detection at 700 V, mCherry detection at 370 V, Acridine Orange detection at 400 V, threshold for SSC at 200 V, sample flow 1 (~1000 events/s), injection volume 50-200 µL, recording of 10-100,000 total events. Data were analysed using Cytobank (https://community.cytobank.org/). Raw data was pre-processed as described in **Fig. S11** to filter out possible aggregates and debris.

Confocal microscopy and image analysis.

A custom-made glass imaging chamber was functionalized with BSA-biotin:BSA and Neutravidin as previously described [17]. The liposome suspension (3-7 µL) was supplemented with PB buffer to a maximum volume of 7 µL and transferred into the functionalized chamber. The LactC2-GFP probe was used at a final concentration of ~260 nM. After 30 to 60 min incubation at room temperature to let the liposomes sediment, the sample was imaged with a Nikon A1R Laser scanning confocal microscope equipped with a ×100 objective and operated via the NIS Elements software (Nikon). The laser lines 488 nm (for MinC-eGFP), 514 nm (for YFP) and 561 nm (for DHPE-Texas Red and LactC2-mCherry) were used in combination with appropriate emission filters. The position of the focal plane was manually adjusted to image as many liposomes as possible across their equatorial plane. Image analysis was performed using ImageJ (https://imagej.nih.gov/ij/) and an in-house developed code, called SMELDit, which enables the identification of individual liposomes, as well as the quantification of fluorescence signals at the membrane and in the lumen.

Mock enrichment of p3 gene.

The linear DNA constructs ori-*p3* and ori-*p6* were mixed at a 1:1 molar ratio for a total DNA concentration of either 10 pM or 50 pM in PURE*frex*2.0 solutions containing 20 mM ammonium sulphate, 300 µM dNTPs, 375 µg/mL purified SSB, 105 µg/mL purified DSB, and 0.6 units/µL of Superase·In RNase inhibitor. The reactions were also supplemented with 2 nM of plasmid DNA encoding for Phi29 DNAP (G85 plasmid). The well-mixed solution was encapsulated in liposomes as described above. Then, 5 µL of bead-free liposome suspension was transferred to a PCR tube, where it was mixed with 0.25 units of DNase 1 (Thermo Scientific), and incubated at 30 °C for 16 h. Upon completion, 2-µL samples were harvested from both + and – dNTPs reactions for quantitative PCR as described above. The enrichment efficiency of ori-*p3* over ori-*p6* was calculated as:

$$\frac{\text{fraction } p3 \text{ (+dNTPs)}/\text{fraction } p6 \text{ (+dNTPs)}}{\text{fraction } p3 \text{ (−dNTPs)}/\text{fraction } p6 \text{ (−dNTPs)}} \qquad (2)$$

Mock enrichment of yfp gene.

The linear DNA constructs ori-*yfp* and ori-*minD* were mixed at 1:10 molar ratio (1 pM ori-*yfp* and 9 pM ori-*minD* final concentrations) in either gene expression solution (PURE*frex*2.0: 50% v/v solution I, 5% v/v solution II, and 10% v/v solution III supplemented with 0.6 units/µL of Superase·In RNase inhibitor) or gene expression-coupled replication solution (PURE*frex*2.0 with an addition of 20 mM ammonium sulphate, 300 µM dNTPs, 375 µg/mL purified SSB, 52.5 µg/mL purified DSB, 3 ng/µL purified Phi29 DNAP, 3 ng/µL purified TP, and 0.6 units/µL of Superase·In RNase inhibitor). The well-mixed solution was encapsulated in liposomes as described above. Then, 10 µL of bead-free liposome suspension was transferred to a PCR tube, where it was mixed with 0.5 units of Proteinase K (Thermo Scientific), and incubated at 30 °C for 16 h. Three microliter of liposome suspension was mixed with 497 µL PB buffer and filtered through the 35-µm nylon mesh of the cell-strainer cap from the 5-mL round-bottom polystyrene test tubes (Falcon).

Fluorescence-activated cell sorting was conducted on FACSMelody (BD Biosciences). Lasers PE-CF594(YG) and FITC-BB515, 100-µm nozzle, 23.14 PSI pressure and 34.2 kHz drop frequency were used. Photon multiplier tube voltages applied were 320 V for forward scatter, 455 for side scatter, 337 V for Texas Red, and 673 V for GFP, and a threshold of 359 V at the side scatter was applied. Liposomes with 1% highest YFP signal were sorted out from liposomes prepared in gene expression solution ("all-gate"), and the same gate was applied to the liposomes prepared in gene expression-coupled replication solution or an adjusted gate including only 0.2% highest YFP signal ("high-gate"). Around 50,000 (low-gate) or 10,000 (high-gate) liposomes were sorted into a 1.5 mL Eppendorf tube. Liposomes from the "all-gate" were further concentrated by centrifugation at 12,000 g for 3 min, and removing three fourth of the supernatant volume. The proteinase K was heat inactivated at 95 °C for 5 min.

The DNA contained in sorted liposomes was used as a template for PCR amplification using phosphorylated primers (ChD 491/ChD 492). Reactions were set up in 100 µL volume, 300 nM each primer, 400 µM dNTPs, 10 µL sorted, heat-inactivated liposome solution, and 2 units of KOD Xtreme Hotstart DNA polymerase in Xtreme buffer, and thermal cycling was performed as follows: 2 min at 94 °C for polymerase activation, and thermal cycling at (98 °C for 10 sec, 65 °C for 20 sec, 68 °C for 1.5 min) × 30. The amplified PCR fragments were purified using QIAquick PCR purification buffers (Qiagen) and RNeasy MinElute Cleanup columns (Qiagen) using the manufacturer's guidelines for QIAquick PCR purification, except for longer pre-elution column drying step (4 min at 10,000 g with open columns), and elution with 14 µL ultrapure water (Merck Milli-Q) in the final step. The purified DNA

was quantified by the Nanodrop 2000c spectrophotometer (Isogen Life Science). The enrichment efficiency of ori-*yfp* over ori-*minD* was calculated as:

$$\frac{\text{fraction } yfp \text{ (post sort)}/\text{fraction } minD \text{ (post sort)}}{\text{fraction } yfp \text{ (pre sort)}/\text{fraction } minD \text{ (pre sort)}} \qquad (3)$$

Statistical analysis of DNA occupancy.

The probability that a liposome contains k molecules of DNA (k = 0, 1, 2, 3, …) according to a Poisson distribution is:

$$P(k) = \frac{\lambda^k}{k!}e^{-\lambda} \qquad (4)$$

where λ is the expected average number of input DNA molecules per liposome. It can be calculated as a function of the diameter d of the liposomes and the bulk concentration C of input DNA templates, as:

$$\lambda = \frac{\pi N_A C d^3}{6} \qquad (5)$$

where N_A is the Avogadro constant. A CADGE reaction may occur in a liposome if one or more copies of linear DNA template is encapsulated, whose corresponding probability is given by:

$$P(k \geq 1) = 1 - e^{-\lambda} \qquad (6)$$

With expCADGE, the concentration of *p2-p3*-plasmid largely exceeds that of ori-*GOI*, such that only the concentration of ori-*GOI* template limits the percentage of liposomes exhibiting CADGE: $P(k_{p2\text{-}p3\text{-plasmid}} \geq 1) \times P(k_{\text{ori-}GOI} \geq 1) \approx P(k_{\text{ori-}GOI} \geq 1)$.

Statistics

Box and whiskers plots in **(Fig. 2)**, **(Fig. 4)** and **(Fig. 5)** have the following characteristics: middle line is the median, the whiskers of the plot are drawn from the 10th percentile up to the 90th, any data point outside the whiskers is drawn as an individual point.

Data and code availability statement

Flow cytometry data were analysed using Cytobank (https://community.cytobank.org/). MATLAB scripts and a user manual for SMELDit are made available upon request.

Acknowledgements

The authors are grateful to Margarita Salas and Mario Mencía Caballero (Universidad Autónoma de Madrid), and Alicia del Prado and Miguel de Vega (Centro de Biología Molecular Severo Ochoa, Madrid) for kindly providing them with the purified DNAP, TP, SSB, and DSB proteins. They thank Yannick Rondelez and Thibault Di Meo (ESPCI, Paris) for fruitful discussions. They also thank Ilja Westerlaken for the purification of LactC2-mCherry and LactC2-eGFP proteins, and Elisa Godino for helping with the preparation and

imaging of MinD-containing liposome samples. The authors are grateful to Duco Blanken, Flora Yang, and Anne Doerr for their assistance with the mass spectrometry experimental setup. Finally, they are thankful to Tess Bevers for assistance with cloning of the ori-*p3* plasmid and Mats van Tongeren for developing SMELDit. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 707404, and from the Netherlands Organization for Scientific Research (NWO/OCW) via the "BaSyC–Building a Synthetic Cell" Gravitation grant (024.003.019).

Supplementary Information

CADGE experimental condition	Estimated production rate
PssA expCADGE	6.4 nM/min
PssA purCADGE	5.3 nM/min
YFP expCADGE	3.8 nM/min
YFP purCADGE	4.1 nM/min

Table S1. Estimated protein production rate from LC/MS kinetics data. The highest slope of the curve is calculated.

Experimental condition	Enrichment efficiency
IVTT (all gate sorting)	31.51
expCADGE (all gate sorting)	13.24
expCADGE (high gate sorting)	40.43
purCADGE (all gate sorting)	17.32
purCADGE (high gate sorting)	89.69

Table S2. Enrichment efficiency of ori-*yfp*.

	P (k ≥ 1)	
Sample type	**d is fixed (mean)**	**d is distributed**
YFP (Fig. 2)	0.214	0.248
YFP (Fig. 2)	0.169	0.211
PssA (Fig. 5)	0.094	0.125
MinD (Fig. 6)	0.246	0.274
MinD (Fig. 6)	0.261	0.288

Table S3. Probabilities of having one or more DNA molecules per liposome, P(k ≥1).

Name	Purpose	Sequence
1106 ChD	YFP gene subcloning	CCGTTTAGAGGCCCCAAGGG
1107 ChD	YFP gene subcloning	CTTCGTCTGTGTCGCATGTGAAATTAATACGACTCACTA TAGGGAGACCACAACG
1104 ChD	amplification of vector	TAGCATAACCCCTTGGGGC
1105 ChD	amplification of vector	CCTATAGTGAGTCGTATTAATTTCACATGCGAC
1115 ChD	pssA gene subcloning	CTTCGTCTGTGTCGCATGTGAAATTAATACGACTCACTA TAGGGGAATTGTGAGC
1116 ChD	pssA gene subcloning	AACCCCTCAAGACCCGTTTAGAG
961 ChD	TP gene cloning	ACGTGGTACCAAAGTAAGCCCCCACCCTCACATG
962 ChD	TP gene cloning	AGCTAAGCTTAAAGTAGGGTACAGCGACAACATACAC
491 ChD	PCR for IVTTR	5-PHOS/AAAGTAAGCCCCCACCCTCACATG
492 ChD	PCR for IVTTR	5-PHOS/AAAGTAGGGTACAGCGACAACATACAC
1121 ChD	YFP detection	TGCAACTGGCTGACCACTAC
1122 ChD	YFP detection	AATGATTGTCCGGCAGCAGA
980 ChD	p3 detection	ACGGCTGAAATTGACATCCCG
981 ChD	p3 detection	CCAGGCGTTGAACTTCTTTGG
1125 ChD	pssA-qPCR-F	AACAGGATGACGGTGGCAAA
1126 ChD	pssA-qPCR-R	GGAACATCTACGCCCGGATT
1208 ChD	MinD detection	CGCGACTCTGACCGTATTT
1209 ChD	MinD detection	AGCATGTCACCTCTGCTTAC

Table S4. DNA primer sequence and purpose

plasmid name	plasmid description
G365	Contains the DNA unit for the expression of YFP fluorescence protein. Transcription is regulated by a T7 promoter and T7 terminator sequences. The entire CDS unit is placed in between right and left origins of replication from the Φ29 DNA replication machinery.
G368	Contains the DNA unit for the expression of the phospholipid biosynthesis protein PssA. Transcription is regulated by a T7 promoter and T7 terminator sequences. The entire CDS unit is placed in between right and left origins of replication from the Φ29 DNA replication machinery.

G437	Contains the DNA unit for the expression of MinD protein. Transcription is regulated by a T7 promoter and T7 terminator sequences. The entire CDS unit is placed in between right and left origins of replication from the Φ29 DNA replication machinery.
G338	Contains the DNA unit for the expression of the Φ29 terminal protein TP. Transcription is regulated by a T7 promoter and T7 terminator sequences. The entire CDS unit is placed in between right and left origins of replication from the Φ29 DNA replication machinery.
G85	Contains the DNA unit for the expression of Φ29 DNA polymerase. Transcription is regulated by a T7 promoter and vsv terminator sequences.
G95	Contains the DNA sequence for the expression of DNAP and TP. Each protein expression is independently regulated by a T7 promoter and a terminator sequence. DNAP unit uses a vsv terminator. TP unit utilizes a T7 terminator sequence. The entire CDS encoding for DNAP and TP is placed in between right and left origins of replication from the Φ29 DNA replication machinery.

Table S5. Plasmid DNA description.

Protein	**Compound name.light/heavy**	**Precursor ion *(m/z)***	**Product ion *(m/z)***	**Collision energy (eV)**	**Accelerator voltage (eV)**	**Ion name**
PSSA	DLQSIADYPVK.light	624.8272	805.4454	20.4	4	y7
PSSA	DLQSIADYPVK.light	624.8272	692.3614	20.4	4	y6
PSSA	DLQSIADYPVK.light	624.8272	506.2973	20.4	4	y4
PSSA	DLQSIADYPVK.light	624.8272	343.2340	20.4	4	y3
PSSA	DLQSIADYPVK.light	624.8272	357.1769	20.4	4	b3
PSSA. QconCAT	DLQSIADYPVK.heavy	631.3079	813.4217	20.4	4	y7
PSSA. QconCAT	DLQSIADYPVK.heavy	631.3079	699.3406	20.4	4	y6
PSSA. QconCAT	DLQSIADYPVK.heavy	631.3079	511.2825	20.4	4	y4
PSSA. QconCAT	DLQSIADYPVK.heavy	631.3079	347.2221	20.4	4	y3
PSSA. QconCAT	DLQSIADYPVK.heavy	631.3079	361.1650	20.4	4	b3
YFP	FEGDTLVNR.light	525.7644	903.4530	17.3	4	y8
YFP	FEGDTLVNR.light	525.7644	774.4104	17.3	4	y7
YFP	FEGDTLVNR.light	525.7644	717.3890	17.3	4	y6

YFP	FEGDTLVNR.light	525.7644	602.3620	17.3	4	y5
YFP	FEGDTLVNR.light	525.7644	501.3144	17.3	4	y4
YFP	FEGDTLVNR.light	525.7644	449.1667	17.3	4	b4
YFP. QconCAT	FEGDTLVNR.heavy	532.2451	915.4175	17.3	4	y8
YFP. QconCAT	FEGDTLVNR.heavy	532.2451	785.3778	17.3	4	y7
YFP. QconCAT	FEGDTLVNR.heavy	532.2451	727.3593	17.3	4	y6
YFP. QconCAT	FEGDTLVNR.heavy	532.2451	611.3354	17.3	4	y5
YFP. QconCAT	FEGDTLVNR.heavy	532.2451	509.2906	17.3	4	y4
YFP. QconCAT	FEGDTLVNR.heavy	532.2451	453.1548	17.3	4	b4
Ribosomal protein S4 (YFP quantification)	LSDYGVQLR.light	525.7826	850.4417	17.3	4	y7
Ribosomal protein S4 (YFP quantification)	LSDYGVQLR.light	525.7826	735.4148	17.3	4	y6
Ribosomal protein S4 (YFP quantification)	LSDYGVQLR.light	525.7826	572.3515	17.3	4	y5
Ribosomal protein S4 (YFP quantification)	LSDYGVQLR.light	525.7826	635.3035	17.3	4	b6
Ribosomal protein S4. QconCAT (YFP quantification)	LSDYGVQLR.heavy	525.7826	861.4091	17.3	4	y7
Ribosomal protein S4. QconCAT (YFP quantification)	LSDYGVQLR.heavy	525.7826	745.3851	17.3	4	y6
Ribosomal protein S4. QconCAT (YFP quantification)	LSDYGVQLR.heavy	525.7826	581.3248	17.3	4	y5
Ribosomal protein S4. QconCAT (YFP quantification)	LSDYGVQLR.heavy	525.7826	641.2857	17.3	4	b6

Ribosomal protein L6 (YFP quantification)	APVVVPAGVDVK.light	575.8452	883.5247	18.9	4	y9
Ribosomal protein L6 (YFP quantification)	APVVVPAGVDVK.light	575.8452	784.4563	18.9	4	y8
Ribosomal protein L6 (YFP quantification)	APVVVPAGVDVK.light	575.8452	685.3879	18.9	4	y7
Ribosomal protein L6 (YFP quantification)	APVVVPAGVDVK.light	575.8452	268.1656	18.9	4	b3
Ribosomal protein L6 (YFP quantification)	APVVVPAGVDVK.light	575.8452	367.2340	18.9	4	b4
Ribosomal protein L6 (YFP quantification)	APVVVPAGVDVK.light	575.8452	466.3024	18.9	4	b5
Ribosomal protein L6. QconCAT (YFP quantification)	APVVVPAGVDVK.heavy	582.3259	893.4951	18.9	4	y9
Ribosomal protein L6. QconCAT (YFP quantification)	APVVVPAGVDVK.heavy	582.3259	793.4296	18.9	4	y8
Ribosomal protein L6. QconCAT (YFP quantification)	APVVVPAGVDVK.heavy	582.3259	693.3642	18.9	4	y7
Ribosomal protein L6. QconCAT (YFP quantification)	APVVVPAGVDVK.heavy	582.3259	271.1567	18.9	4	b3
Ribosomal protein L6. QconCAT (YFP quantification)	APVVVPAGVDVK.heavy	582.3259	371.2221	18.9	4	b4
Ribosomal protein L6. QconCAT (YFP quantification)	APVVVPAGVDVK.heavy	582.3259	471.2876	18.9	4	b5
Ribosomal protein S1 (PSSA quantification)	GVVVAIDK.light	400.7475	644.3978	13.4	4	y6
Ribosomal protein S1 (PSSA quantification)	GVVVAIDK.light	400.7475	545.3293	13.4	4	y5
Ribosomal protein S1 (PSSA quantification)	GVVVAIDK.light	400.7475	446.2609	13.4	4	y4

Ribosomal protein S1 (PSSA quantification)	GVVVAIDK.light	400.7475	426.2711	13.4	4	b5
Ribosomal protein S1. QconCAT (PSSA quantification)	GVVVAIDK.heavy	405.234	651.3770	13.4	4	y6
Ribosomal protein S1. QconCAT (PSSA quantification)	GVVVAIDK.heavy	405.234	551.3115	13.4	4	y5
Ribosomal protein S1. QconCAT (PSSA quantification)	GVVVAIDK.heavy	405.234	451.2461	13.4	4	y4
Ribosomal protein S1. QconCAT (PSSA quantification)	GVVVAIDK.heavy	405.234	431.2563	13.4	4	b5
Ribosomal protein L1 (PSSA quantification)	VVGQLGQVLGPR.light	611.8670	1024.5898	20	4	y10
Ribosomal protein L1 (PSSA quantification)	VVGQLGQVLGPR.light	611.8670	839.5098	20	4	y8
Ribosomal protein L1 (PSSA quantification)	VVGQLGQVLGPR.light	611.8670	726.4257	20	4	y7
Ribosomal protein L1 (PSSA quantification)	VVGQLGQVLGPR.light	611.8670	442.2772	20	4	y4
Ribosomal protein L1 (PSSA quantification)	VVGQLGQVLGPR.light	611.8670	329.1932	20	4	y3
Ribosomal protein L1 (PSSA quantification)	VVGQLGQVLGPR.heavy	620.3418	1039.5453	20	4	y10
Ribosomal protein L1 (PSSA quantification)	VVGQLGQVLGPR.heavy	620.3418	851.4742	20	4	y8
Ribosomal protein L1 (PSSA quantification)	VVGQLGQVLGPR.heavy	620.3418	737.3931	20	4	y7
Ribosomal protein L1 (PSSA quantification)	VVGQLGQVLGPR.heavy	620.3418	449.2565	20	4	y4
Ribosomal protein L1 (PSSA quantification)	VVGQLGQVLGPR.heavy	620.3418	335.1754	20	4	y3
Ribosomal protein S4 (PSSA quantification)	AALELAEQR.light	500.7747	745.3839	16.5	4	y6

Ribosomal protein S4 (PSSA quantification)	AALELAEQR.light	500.7747	616.3413	16.5	4	y5
Ribosomal protein S4 (PSSA quantification)	AALELAEQR.light	500.7747	503.2572	16.5	4	y4
Ribosomal protein S4 (PSSA quantification)	AALELAEQR.light	500.7747	432.2201	16.5	4	y3
Ribosomal protein S4 (PSSA quantification)	AALELAEQR.light	500.7747	385.2082	16.5	4	b4
Ribosomal protein S4 (PSSA quantification)	AALELAEQR.heavy	507.2555	755.3542	16.5	4	y6
Ribosomal protein S4 (PSSA quantification)	AALELAEQR.heavy	507.2555	625.3146	16.5	4	y5
Ribosomal protein S4 (PSSA quantification)	AALELAEQR.heavy	507.2555	511.2335	16.5	4	y4
Ribosomal protein S4 (PSSA quantification)	AALELAEQR.heavy	507.2555	439.1994	16.5	4	y3
Ribosomal protein S4 (PSSA quantification)	AALELAEQR.heavy	507.2555	389.1963	16.5	4	b4

Table S6. Transitions of the MS/MS measurements for the proteolytic peptides of the indicated proteins.

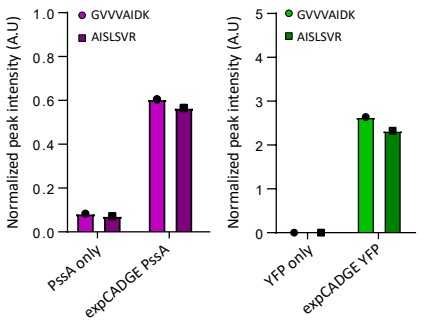


Fig S1. Relative protein quantification from bulk reactions with and without expCADGE. The maximum peak intensity of each POI proteolytic peptide was normalized to the maximum peak intensity of one of the two S1 ribosomal protein signature peptides (GVVVAIDK or AISLSVR) present in PURE system.

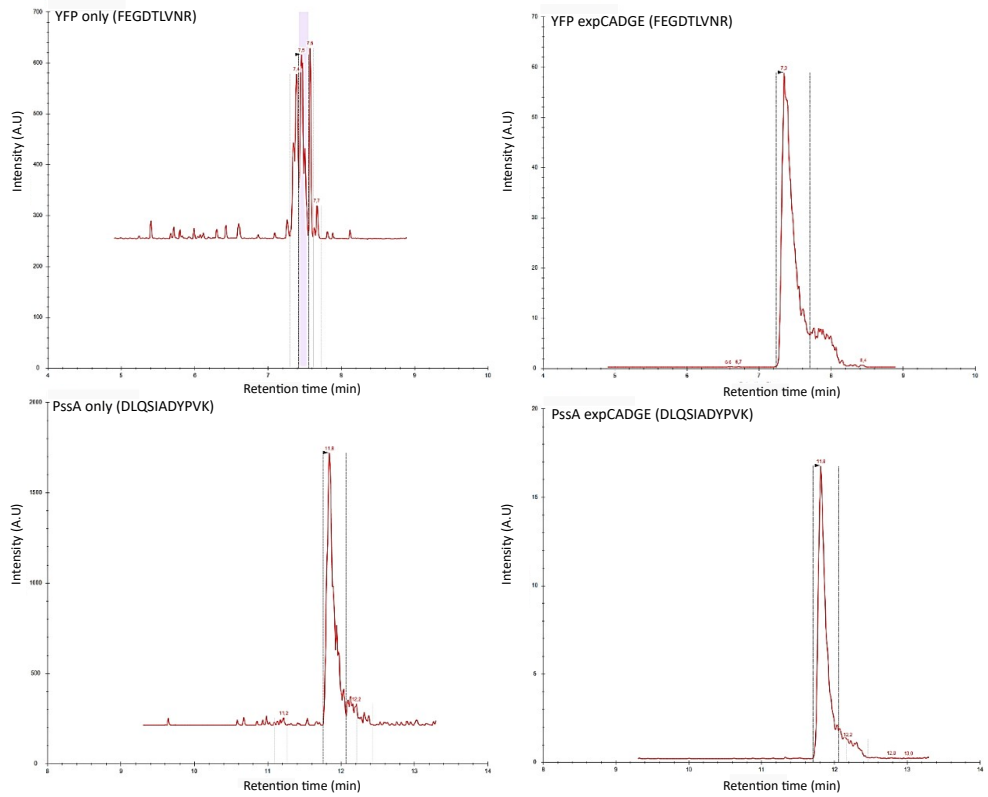


Fig. S2 LC-MS raw data peaks of the YFP and PssA proteolytic peptides used for relative quantification.

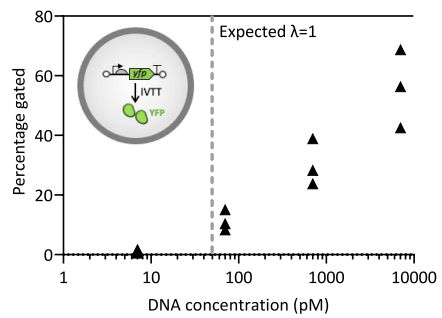


Fig. S3 Effect of DNA concentration on YFP protein expression without gene amplification. Fluorescence from individual liposomes was analyzed by flow cytometry. The vertical dashed line indicates the DNA concentration that theoretically corresponds to λ = 1 if one assumes a monodisperse population of liposomes with a diameter of 4 µm and a random (Poisson) partitioning of DNA.

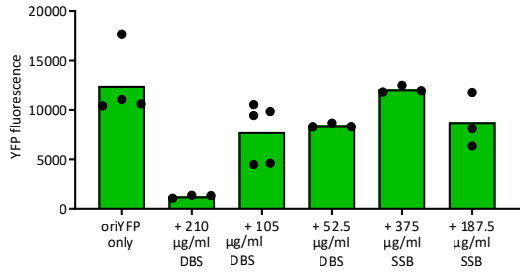


Fig. S4 End-point YFP fluorescence measurements from ori-*yfp* bulk IVTT reactions. Protein expression can be inhibited under high DSB concentrations (210 µg/ml).

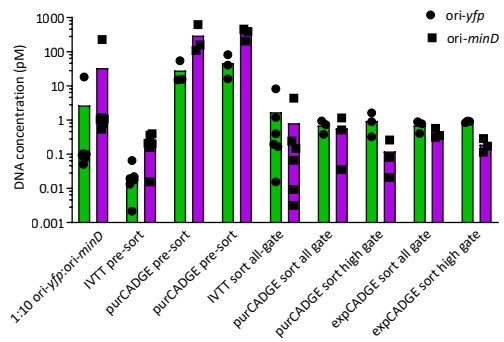


Fig. S5 Individual qPCR data from Fig. 4h. Enrichment of ori-*yfp* over ori-*minD*. Each symbol represents a biological repeat. 'IVTT' indicates a reaction without DNA replication.

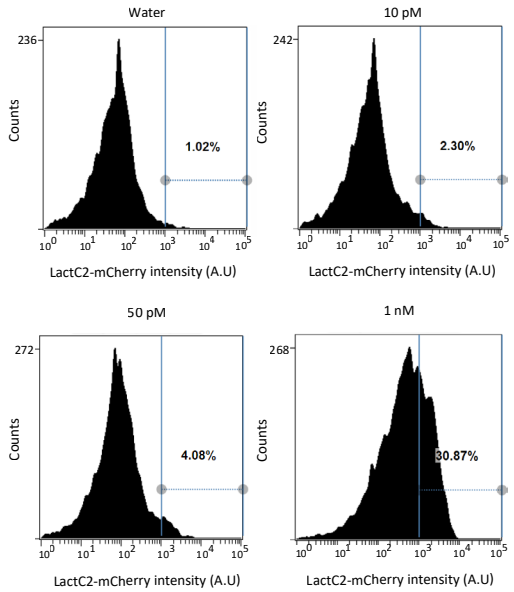


Fig. S6 Raw FACS data of liposome samples with appended gating line as used in Fig. 5c,d.

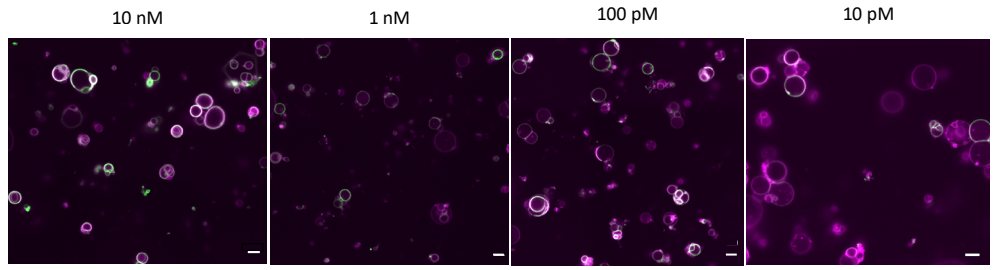


Fig. S7 In-liposome expression of ori-*pssA* under different DNA concentrations (10 nM to 10 pM) without DNA replication. Liposome membrane dye (Texas-red) is colored in magenta and PS binding protein LactC2-eGFP is colored in green. Overlay of the two colors is displayed in white. Lowering the concentration of *pssA* gene reduces the number of liposomes with membrane-recruited LactC2-eGFP. Scale bars are 5 µm.

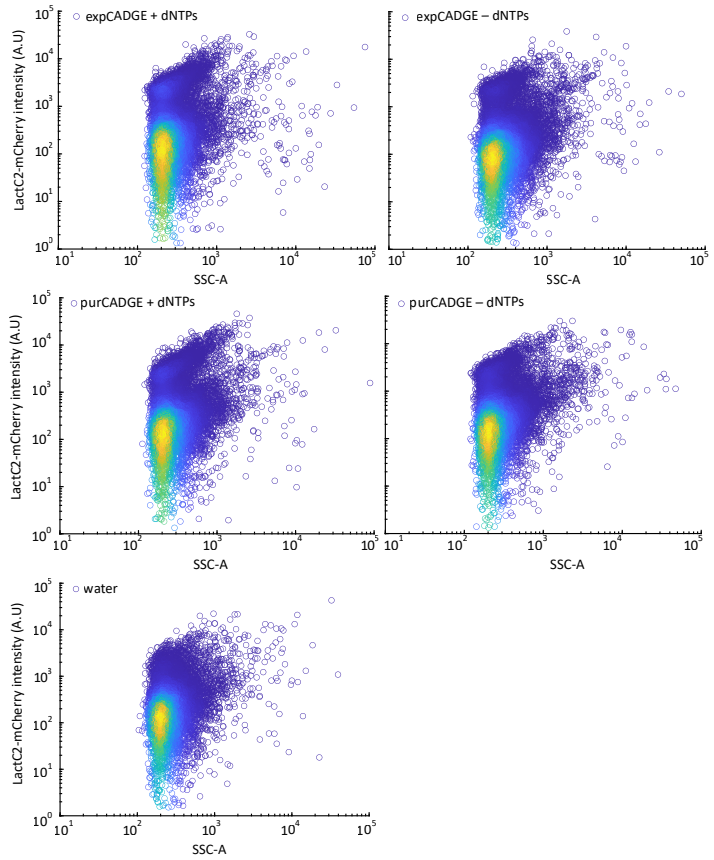


Fig S8. FACS data of liposome samples analyzed in Fig. 5g.

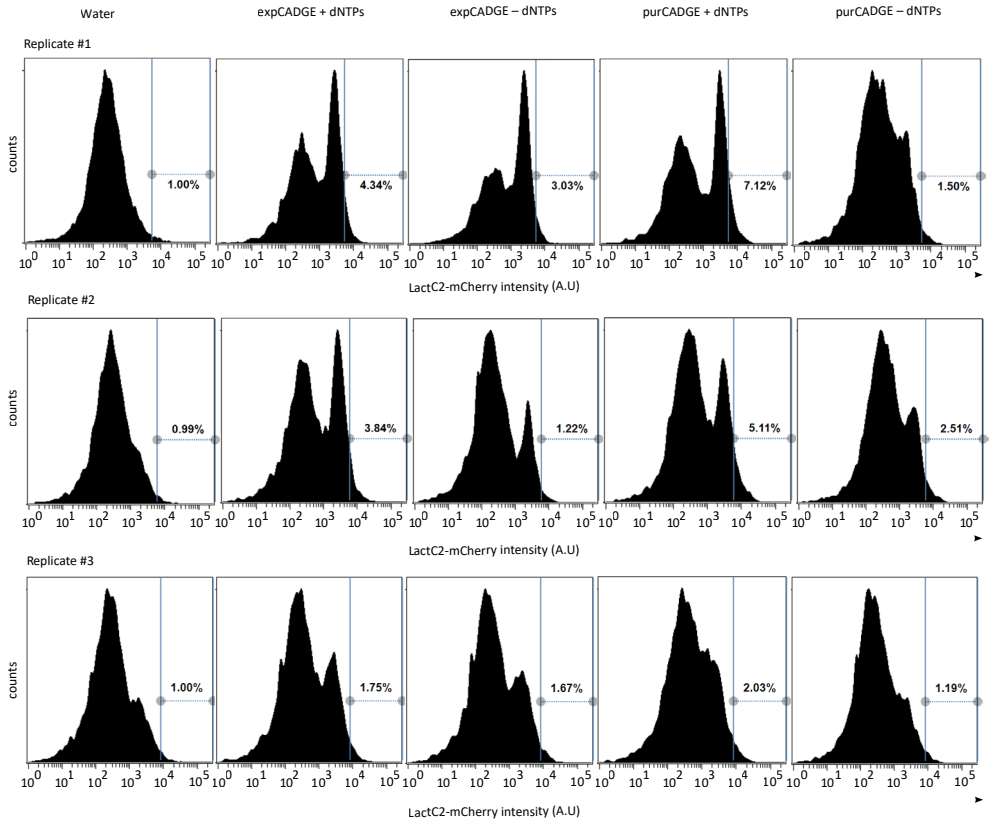


Fig. S9. Liposomes FACS data and gating strategy for Fig. 5h.

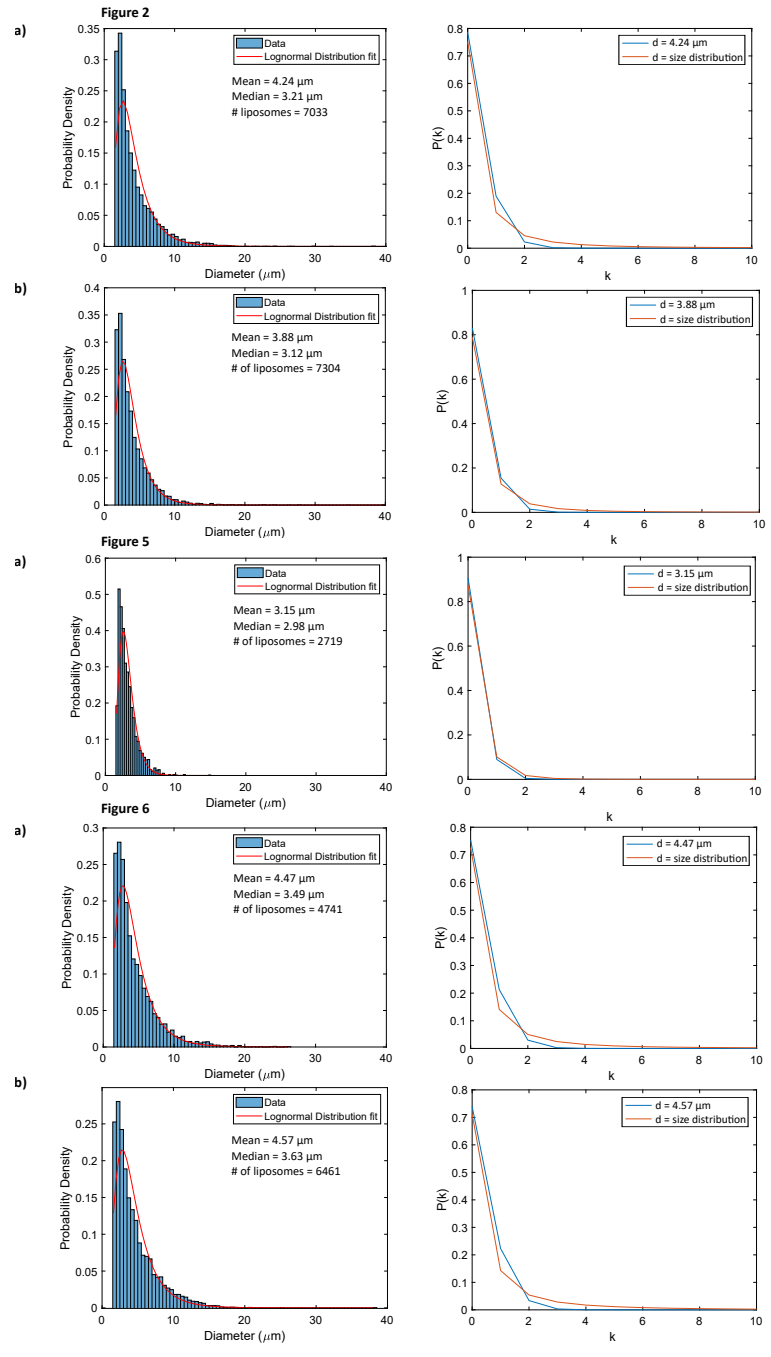


Fig. S10 Quantitative analysis of liposome size distribution and DNA occupancy. Histograms of liposome sizes for different samples (left). The mean and median diameter values, as well as the number of liposomes analyzed per sample are appended on the graphs. The distributions were fitted with a lognormal function (red curves). The probability of having k DNA molecules on average per liposome was calculated by assuming a fixed diameter (distribution mean, in blue) or from the actual size distribution in the corresponding sample (in red).

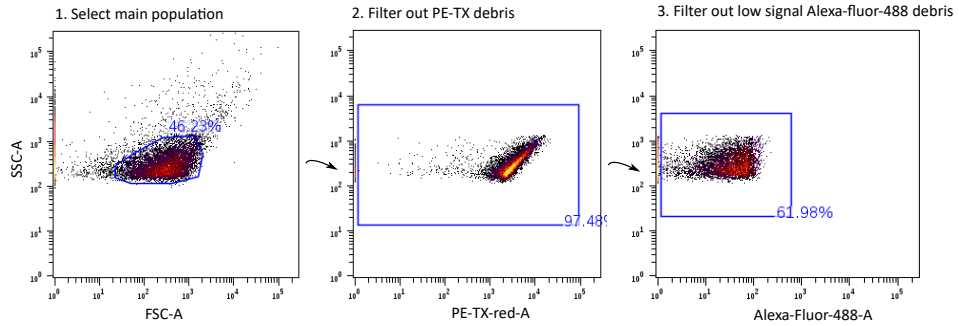


Fig. S11 Data processing for FACS data. Filtering gate to remove liposomal debris in PssA expressing-liposomes.

References

1. Romero, P. A., and Arnold, F. H. (2009) Exploring Protein Fitness Landscapes by Directed Evolution. Nat. Rev. Mol. Cell Biol. 10. 866–876.
2. Packer, M. S., and Liu, D. R. (2015) Methods for the Directed Evolution of Proteins. Nat. Rev. Genet. 16. 379–394.
3. Wang, Y., Xue, P., Cao, M., Yu, T., Lane, S. T., and Zhao, H. (2021) Directed Evolution: Methodologies and Applications. Chem. Rev. 121. 12384–12444.
4. Bornscheuer, U. T., Hauer, B., Jaeger, K. E., and Schwaneberg, U. (2019) Directed Evolution Empowered Redesign of Natural Proteins for the Sustainable Production of Chemicals and Pharmaceuticals. Angew. Chemie - Int. Ed. 58. 36–40.
5. Turner, N. J. (2009) Directed Evolution Drives the next Generation of Biocatalysts. Nat. Chem. Biol. 5. 567–573.
6. Engqvist, M. K. M., and Rabe, K. S. (2019) Applications of Protein Engineering and Directed Evolution in Plant Research. Plant Physiol. 179. 907–917.
7. Hida, K., Hanes, J., and Ostermeier, M. (2007) Directed Evolution for Drug and Nucleic Acid Delivery. Adv. Drug Deliv. Rev. 59. 1562–1578.
8. Xiong, A. S., Peng, R. H., Zhuang, J., Davies, J., Zhang, J., and Yao, Q. H. (2012) Advances in Directed Molecular Evolution of Reporter Genes. Crit. Rev. Biotechnol. 32. 133–142.
9. Johannes, T. W., and Zhao, H. (2006) Directed Evolution of Enzymes and Biosynthetic Pathways. Curr. Opin. Microbiol. 9. 261–267.
10. Chatterjee, R., and Yuan, L. (2006) Directed Evolution of Metabolic Pathways. Trends Biotechnol. 24. 28–38.
11. Zhou, S., Wu, Y., Xie, Z. X., Jia, B., and Yuan, Y. J. (2021) Directed Genome Evolution Driven by Structural Rearrangement Techniques. Chem. Soc. Rev. 50. 12788–12807.
12. Abil, Z., and Danelon, C. (2020) Roadmap to Building a Cell : An Evolutionary Approach. Front. Bioeng. Biotechnol. 8. DOI: 10.3389/fbioe.2020.00927.
13. Yu, W., Sato, K., Wakabayashi, M., Nakaishi, T., Ko-Mitamura, E. P., Shima, Y., Itaru Urabe, T., and Yomo, E. (2001) Synthesis of Functional Protein in Liposome. J. Biosci. Bioeng. 92. 590–593.
14. Nomura, S. I. M., Tsumoto, K., Hamada, T., Akiyoshi, K., Nakatani, Y., and Yoshikawa, K. (2003) Gene Expression within Cell-Sized Lipid Vesicles. Chembiochem. 4. 1172–1175.
15. Noireaux, V., and Libchaber, A. (2004) A Vesicle Bioreactor as a Step toward an Artificial Cell Assembly. Proc. Natl. Acad. Sci. U. S. A. 101. 17669–17674.
16. Fujii, S., Matsuura, T., Sunami, T., Nishikawa, T., Kazuta, Y., and Yomo, T. (2014) Liposome Display for in Vitro Selection and Evolution of Membrane Proteins. Nat. Protoc. 9. 1578–1591.
17. Blanken, D., Van Nies, P., and Danelon, C. (2019) Quantitative Imaging of Gene-Expressing Liposomes Reveals Rare Favorable Phenotypes. Phys. Biol. 16. DOI: 10.1088/1478-3975/ab0c62.
18. van Nies, P., Westerlaken, I., Blanken, D., Salas, M., Mencía, M., and Danelon, C. (2018) Self-Replication of DNA by Its Encoded Proteins in Liposome-Based Synthetic Cells. Nat. Commun. 9. DOI: 10.1038/s41467-018-03926-1.

19. Blanken, D., Foschepoth, D., Serrão, A. C., and Danelon, C. (2020) Genetically Controlled Membrane Synthesis in Liposomes. Nat. Commun. 11. DOI: 10.1038/s41467-020-17863-5.
20. Godino, E., López, J. N., Foschepoth, D., Cleij, C., Doerr, A., Castellà, C. F., and Danelon, C. (2019) De Novo Synthesized Min Proteins Drive Oscillatory Liposome Deformation and Regulate FtsA-FtsZ Cytoskeletal Patterns. Nat. Commun. 10. DOI: 10.1038/s41467-019-12932-w.
21. Godino, E., López, J. N., Zarguit, I., Doerr, A., Jimenez, M., Rivas, G., and Danelon, C. (2020) Cell-Free Biogenesis of Bacterial Division Proto-Rings That Can Constrict Liposomes. Commun. Biol. 3. DOI: 10.1038/S42003-020-01258-9.
22. Kattan, J., Doerr, A., Dogterom, M., and Danelon, C. (2021) Shaping Liposomes by Cell-Free Expressed Bacterial Microtubules. ACS Synth. Biol. 10. 2447–2455.
23. Berhanu, S., Ueda, T., and Kuruma, Y. (2019) Artificial Photosynthetic Cell Producing Energy for Protein Synthesis. Nat. Commun. 10. DOI: 10.1038/s41467-019-09147-4.
24. Pourmir, A., and Johannes, T. W. (2012) Directed Evolution: Selection of the Host Organism. Comput. Struct. Biotechnol. J. 2. DOI: 10.5936/CSBJ.201209012.
25. Shimizu, Y., Kuruma, Y., Ying, B. W., Umekage, S., and Ueda, T. (2006) Cell-Free Translation Systems for Protein Engineering. FEBS J. 273. 4133–4140.
26. Perez, J. G., Stark, J. C., and Jewett, M. C. (2016) Cell-Free Synthetic Biology: Engineering Beyond the Cell. Cold Spring Harb. Perspect. Biol. 8. DOI: 10.1101/CSHPERSPECT.A023853.
27. Fujii, S., Matsuura, T., Sunami, T., Kazuta, Y., and Yomo, T. (2013) In Vitro Evolution of α-Hemolysin Using a Liposome Display. Proc Natl Acad Sci U.S.A. 110. 16796–16801.
28. Holstein, J. M., Gylstorff, C., and Hollfelder, F. (2021) Cell-Free Directed Evolution of a Protease in Microdroplets at Ultrahigh Throughput. ACS Synth. Biol. 10. 252–257.
29. Newton, M. S., Cabezas-Perusse, Y., Tong, C. L., and Seelig, B. (2020) In Vitro Selection of Peptides and Proteins - Advantages of MRNA Display. ACS Synth. Biol. 9. 181–190.
30. Watts, R. E., and Forster, A. C. (2012) Update on Pure Translation Display with Unnatural Amino Acid Incorporation. Methods Mol Biol. 805. 349–365.
31. Uyeda, A., Watanabe, T., Kato, Y., Watanabe, H., Yomo, T., Hohsaka, T., and Matsuura, T. (2015) Liposome-Based in Vitro Evolution of Aminoacyl-TRNA Synthetase for Enhanced Pyrrolysine Derivative Incorporation. ChemBioChem. 16. 1797–1802.
32. Silverman, A. D., Karim, A. S., and Jewett, M. C. (2020) Cell-Free Gene Expression: An Expanded Repertoire of Applications. Nat. Rev. Genet. 21. 151–170.
33. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-Free Translation Reconstituted with Purified Components. Nat. Biotechnol. 19. 751–755.
34. Dodevski, I., Markou, G. C., and Sarkar, C. A. (2015) Conceptual and Methodological Advances in Cell-Free Directed Evolution. Curr. Opin. Struct. Biol. 33. DOI: 10.1016/j.sbi.2015.04.008.
35. Tawfik, D. S., and Griffiths, A. D. (1998) Man-Made Cell-like Compartments for Molecular Evolution. Nat. Biotechnol. 16. 652–656.
36. Sunami, T., Sato, K., Matsuura, T., Tsukada, K., Urabe, I., and Yomo, T. (2006) Femtoliter Compartment in Liposomes for in Vitro Selection of Proteins. Anal. Biochem. 357. 128–136.
37. Cho, E., and Lu, Y. (2020) Compartmentalizing Cell-Free Systems: Toward Creating Life-Like Artificial Cells and Beyond. ACS Synth. Biol. 9. 2881–2901.
38. Nishikawa, T., Sunami, T., Matsuura, T., Ichihashi, N., and Yomo, T. (2012) Construction of a Gene Screening System Using Giant Unilamellar Liposomes and a Fluorescence-Activated Cell Sorter. Anal. Chem. 84. 5017–5024.
39. Uyeda, A., Nakayama, S., Kato, Y., Watanabe, H., and Matsuura, T. (2016) Construction of an in Vitro Gene Screening System of the E. Coli EmrE Transporter Using Liposome Display. Anal. Chem. 88. 12028–12035.
40. Zhang, Y., Minagawa, Y., Kizoe, H., Miyazaki, K., Iino, R., Ueno, H., Tabata, K. V., Shimane, Y., and Noji, H. (2019) Accurate High-Throughput Screening Based on Digital Protein Synthesis in a Massively Parallel Femtoliter Droplet

Array. Sci. Adv. 5. DOI: 10.1126/SCIADV.AAV8185.

41. Stögbauer, T., Windhager, L., Zimmer, R., and Rädler, J. O. (2012) Experiment and Mathematical Modeling of Gene Expression Dynamics in a Cell-Free System. Integr. Biol. 4. 494–501.
42. Chory, J., and Kaplan, S. (1982) The in Vitro Transcription-Translation of DNA and RNA Templates by Extracts of Rhodopseudomonas Sphaeroides. Optimization and Comparison of Template Specificity with Escherichia Coli Extracts and in Vivo Synthesis. J. Biol. Chem. 257. 15110–15121.
43. Niederholtmeyer, H., Xu, L., and Maerkl, S. J. (2013) Real-Time MRNA Measurement during an in Vitro Transcription and Translation Reaction Using Binary Probes. ACS Synth. Biol. 2. 411–417.
44. Nourian, Z., and Danelon, C. (2013) Linking Genotype and Phenotype in Protein Synthesizing Liposomes with External Supply of Resources. ACS Synth. Biol. 2. DOI: 10.1021/sb300125z.
45. Doerr, A., De Reus, E., Van Nies, P., Van Der Haar, M., Wei, K., Kattan, J., Wahl, A., and Danelon, C. (2019) Modelling Cell-Free RNA and Protein Synthesis with Minimal Systems. Phys. Biol. 16. DOI: 10.1088/1478-3975/aaf33d.
46. Ueno, H., Sawada, H., Soga, N., Sano, M., Nara, S., Tabata, K. V., Su'etsugu, M., and Noji, H. (2021) Amplification of over 100 Kbp DNA from Single Template Molecules in Femtoliter Droplets. ACS Synth. Biol. 10. 2179–2186.
47. Libicher, K., Hornberger, R., Heymann, M., and Mutschler, H. (2020) In Vitro Self-Replication and Multicistronic Expression of Large Synthetic Genomes. Nat. Commun. 11. DOI: 10.1038/s41467-020-14694-2.
48. Okauchi, H., and Ichihashi, N. (2021) Continuous Cell-Free Replication and Evolution of Artificial Genomic DNA in a Compartmentalized Gene Expression System. ACS Synth. Biol. 10. 3507–3517.
49. Sakatani, Y., Ichihashi, N., Kazuta, Y., and Yomo, T. (2015) A Transcription and Translation-Coupled DNA Replication System Using Rolling-Circle Replication. Sci. Rep. 5. DOI: 10.1038/SREP10404.
50. Sakatani, Y., Yomo, T., and Ichihashi, N. (2018) Self-Replication of Circular DNA by a Self-Encoded DNA Polymerase through Rolling-Circle Replication and Recombination OPEN. Sci. Rep. 8. DOI: 10.1038/s41598-018-31585-1.
51. Mazutis, L., Araghi, A. F., Miller, O. J., Baret, J.-C., Frenz, L., Janoshazi, A., Valé Rie Taly, ´, Miller, B. J., Hutchison, J. B., Link, D., Griffiths, A. D., and Ryckelynck, M. (2009) Droplet-Based Microfluidic Systems for High-Throughput Single DNA Molecule Isothermal Amplification and Analysis. Anal. Chem. 81. 4813–4821.
52. Fallah-Araghi, A., Baret, J. C., Ryckelynck, M., and Griffiths, A. D. (2012) A Completely in Vitro Ultrahigh-Throughput Droplet-Based Microfluidic Screening System for Protein Engineering and Directed Evolution. Lab Chip. 12. 882–891.
53. Galinis, R., Stonyte, G., aidotas Kiseliovas, V., apolas Zilionis, R., Studer, S., Hilvert, D., Janulaitis, A., and Mazutis, L. (2016) DNA Nanoparticles for Improved Protein Synthesis In Vitro. Angew. Chemie Int. Ed. 55. 3120–3123.
54. Paul, S., Stang, A., Lennartz, K., Tenbusch, M., and Überla, K. (2013) Selection of a T7 Promoter Mutant with Enhanced in Vitro Activity by a Novel Multi-Copy Bead Display Approach for in Vitro Evolution. Nucleic Acids Res. 41. DOI: 10.1093/nar/gks940.
55. Diamante, L., Gatti-Lafranconi, P., Schaerli, Y., and Hollfelder, F. (2013) In Vitro Affinity Screening of Protein and Peptide Binders by Megavalent Bead Surface Display. Protein Eng. Des. Sel. 26. 713–724.
56. Lindenburg, L., Huovinen, T., Van De Wiel, K., Herger, M., Snaith, M. R., and Hollfelder, F. (2020) Split & Mix Assembly of DNA Libraries for Ultrahigh Throughput On-Bead Screening of Functional Proteins. Nucleic Acids Res. 48. DOI: 10.1093/nar/gkaa270.
57. Plesa, C., Sidore, A. M., Lubock, N. B., Zhang, D., and Kosuri, S. (2018) Multiplexed Gene Synthesis in Emulsions for Exploring Protein Functional Landscapes. Science. 359. 343–347.
58. Restrepo Sierra, A. M., Arold, S. T., and Grünberg, R. (2022) Efficient Multi-Gene Expression in Cell-Free Droplet Microreactors. PLoS One. 17. DOI: 10.1371/journal.pone.0260420.
59. Mencía, M., Gella, P., Camacho, A., De Vega, M., and Salas, M. (2011) Terminal Protein-Primed Amplification of Heterologous DNA with a Minimal Replication System Based on Phage ϕ29. Proc. Natl. Acad. Sci. 108. 18655–18660.
60. Rodríguez, I., Lázaro, J. M., Blanco, L., Kamtekar, S., Berman, A. J., Wang, J., Steitz, T. A., Salas, M., and De Vega, M. (2005) A Specific Subdomain in ϕ29 DNA Polymerase Confers Both Processivity and Strand-Displacement Capacity. Proc. Natl. Acad. Sci. U. S. A. 102. 6407–6412.

61. Kamtekar, S., Berman, A. J., Wang, J., Lázaro, J. M., De Vega, M., Blanco, L., Salas, M., and Steitz, T. A. (2004) Insights into Strand Displacement and Processivity from the Crystal Structure of the Protein-Primed DNA Polymerase of Bacteriophage ϕ29. Mol. Cell. 16. 609–618.
62. Blanco, L., Bernads, A., Lharo, J. M., Martins, G., Garmendia, C., and Salasn, M. (1989) Highly Efficient DNA Synthesis by the Phage Phi29 DNA Polymerase. J. Biol. Chem. 264. 8935–8940.
63. Kamtekar, S., Berman, A. J., Wang, J., Lá Zaro, J. M., De Vega, M., Blanco, L., Salas, M., and Steitz, T. A. (2006) The Phi29 DNA Polymerase:Protein-Primer Structure Suggests a Model for the Initiation to Elongation Transition. EMBO J. 25. 1335–1343.
64. Blanco, L., Gutierrez, J., Lazaro, J. M., Bernad, A., and Salas, M. (1986) Replication of Phage Phi29 DNA in Vitro: Role of the Viral Protein P6 in Initiation and Elongation. Nucleic Acids Res. 14. 4923–4937.
65. Freire, R., Serrano, M., Salas, M., and Hermoso, J. M. (1996) Activation of Replication Origins in Phi29-Related Phages Requires the Recognition of Initiation Proteins to Specific Nucleoprotein Complexes. J. Biol. Chem. 271. 31000–31007.
66. Penalva, M. A., and Salas, M. (1982) Initiation of Phage Phi29 DNA Replication in Vitro: Formation of a Covalent Complex between the Terminal Protein, P3, and 5'-DAMP. Proc. Nati Acad. Sci. USA. 79. 5522–5526.
67. Méndez, J., Blanco, L., and Salas, M. (1997) Protein-Primed DNA Replication: A Transition between Two Modes of Priming by a Unique DNA Polymerase. EMBO J. 16. 2519–2527.
68. Martìin, G., Làzaro, J., Mèndez, E., and Salas, M. (1989) Characterization of the Phage Phi 29 Protein P5 as a Single-Stranded DNA Binding Protein. Function in Phi 29 DNA-Protein P3 Replication. Nucleic Acids Res. 17. 3663–3672.
69. Serrano, M., Gutierrez, J., Prieto, I., Hermoso, J. M., and Salas, M. (1989) Signals at the Bacteriophage Phi 29 DNA Replication Origins Required for Protein P6 Binding and Activity. EMBO J. 8. DOI: 10.1002/j.1460-2075.1989.tb03584.x.
70. Blanco, L., Prieto, I., Gutierrez, J., Bernad, A., Lazaro, J. M., Hermoso, J. M., and Salas, M. (1987) Effect of NH4+ Ions on Phi29 DNA-Protein P3 Replication: Formation of a Complex between the Terminal Protein and the DNA Polymerase. J. Virol. 61. 3983–3991.
71. Carlson, E. D., Gan, R., Hodgman, C. E., and Jewett, M. C. (2012) Cell-Free Protein Synthesis: Applications Come of Age. Biotechnol. Adv. 30. 1185–1194.
72. Doerr, A., Foschepoth, D., Forster, A. C., and Danelon, C. (2021) In Vitro Synthesis of 32 Translation-Factor Proteins from a Single Template Reveals Impaired Ribosomal Processivity. Sci. Rep. 11. DOI: 10.1038/S41598-020-80827-8.
73. Meijer, W. J. J., Horcajadas, J. A., and Salas, M. (2001) ϕ29 Family of Phages. Microbiol. Mol. Biol. Rev. 65. 261–287.
74. Raetz, C. R. H. (1976) Phosphatidylserine Synthetase Mutants of Escherichia Coli. Genetic Mapping and Membrane Phospholipid Composition. J. Biol. Chem. 251. 3242–3249.
75. DeChavigny, A., Heacock, P. N., and Dowhan, W. (1991) Sequence and Inactivation of the Pss Gene of Escherichia Coli. Phosphatidylethanolamine May Not Be Essential for Cell Viability. J. Biol. Chem. 266. 5323–5332.
76. Larson, T. J., and Dowhan, W. (1976) Ribosomal-Associated Phosphatidylserine Synthetase from Escherichia Coli: Purification by Substrate-Specific Elution from Phosphocellulose Using Cytidine 5'-Diphospho-1,2-Diacyl-Sn-Glycerol. Biochemistry. 15. 5212–5218.
77. Hirabayashi, T., Larson, T. J., and Dowhan, W. (1976) Membrane-Associated Phosphatidylglycerophosphate Synthetase from Escherichia Coli: Purification by Substrate Affinity Chromatography on Cytidine 5' Diphospho-1,2-Diacyl-Sn-Glycerol Sepharose. Biochemistry. 15. 5205–5211.
78. Rowlett, V. W., and Margolin, W. (2015) The Min System and Other Nucleoid-Independent Regulators of Z Ring Positioning. Front. Microbiol. 6. DOI: 10.3389/FMICB.2015.00478.
79. Rowlett, V. W., and Margolin, W. (2013) The Bacterial Min System. Curr. Biol. 23. DOI: 10.1016/J.CUB.2013.05.024.
80. Lu, W. C., Levy, M., Kincaid, R., and Ellington, A. D. (2014) Directed Evolution of the Substrate Specificity of Biotin Ligase. Biotechnol. Bioeng. 111. 1071–1081.
81. Borkowski, O., Koch, M., Zettor, A., Pandi, A., Batista, A. C., Soudier, P., and Faulon, J. L. (2020) Large Scale Active-Learning-Guided Exploration for in Vitro Protein Production Optimization. Nat. Commun. 11. DOI: 10.1038/

S41467-020-15798-5.
82. Ishikawa, K., Sato, K., Shima, Y., Urabe, I., and Yomo, T. (2004) Expression of a Cascading Genetic Network within Liposomes. FEBS Lett. 576. 387–390.
83. Kobori, S., Ichihashi, N., Kazuta, Y., and Yomo, T. (2013) A Controllable Gene Expression System in Liposomes That Includes a Positive Feedback Loop. Mol. Biosyst. 9. 1282–1285.
84. Maddalena, L. L. D., Niederholtmeyer, H., Turtola, M., Swank, Z. N., Belogurov, G. A., and Maerkl, S. J. (2016) GreA and GreB Enhance Expression of Escherichia Coli RNA Polymerase Promoters in a Reconstituted Transcription-Translation System. ACS Synth. Biol. 5. 929–935.
85. Elías-Arnanz, M., and Salas, M. (1997) Bacteriophage Phi29 DNA Replication Arrest Caused by Codirectional Collisions with the Transcription Machinery. EMBO J. 16. 5775–5783.
86. Elías-Arnanz, M., and Salas, M. (1999) Resolution of Head-on Collisions between the Transcription Machinery and Bacteriophage Phi29 DNA Polymerase Is Dependent on RNA Polymerase Translocation. EMBO J. 18. 5675–5682.
87. Marshall, R., and Noireaux, V. (2018) Synthetic Biology with an All E. Coli TXTL System: Quantitative Characterization of Regulatory Elements and Gene Circuits. Methods Mol. Biol. 1772. 61–93.
88. Sitaraman, K., Esposito, D., Klarmann, G., Le Grice, S. F., Hartley, J. L., and Chatterjee, D. K. (2004) A Novel Cell-Free Protein Synthesis System. J. Biotechnol. 110. 257–263.
89. Marshall, R., Maxwell, C. S., Collins, S. P., Beisel, C. L., and Noireaux, V. (2017) Short DNA Containing χ Sites Enhances DNA Stability and Gene Expression in E. Coli Cell-Free Transcription-Translation Systems. Biotechnol. Bioeng. 114. 2137–2141.
90. Batista, A. C., Levrier, A., Soudier, P., Voyvodic, P. L., Achmedov, T., Reif-Trauttmansdorff, T., DeVisch, A., Cohen-Gonsaud, M., Faulon, J. L., Beisel, C. L., Bonnet, J., and Kushwaha, M. (2022) Differentially Optimized Cell-Free Buffer Enables Robust Expression from Unprotected Linear DNA in Exonuclease-Deficient Extracts. ACS Synth. Biol. 11. 732–746.
91. Nourian, Z., Roelofsen, W., and Danelon, C. (2012) Triggered Gene Expression in Fed-Vesicle Microreactors with a Multifunctional Membrane. Angew. Chemie Int. Ed. 51. 3114–3118.
92. Van de Cauter, L., Fanalista, F., Van Buren, L., De Franceschi, N., Godino, E., Bouw, S., Danelon, C., Dekker, C., Koenderink, G. H., and Ganzinger, K. A. (2021) Optimized CDICE for Efficient Reconstitution of Biological Systems in Giant Unilamellar Vesicles. ACS Synth. Biol. 10. 1690–1702.
93. Gonzales, D. T., Yandrapalli, N., Robinson, T., Zechner, C., and Tang, T. Y. D. (2022) Cell-Free Gene Expression Dynamics in Synthetic Cell Populations. ACS Synth. Biol. 11. 205–215.
94. Vogele, K., Frank, T., Gasser, L., Goetzfried, M. A., Hackl, M. W., Sieber, S. A., Simmel, F. C., and Pirzer, T. (2018) Towards Synthetic Cells Using Peptide-Based Reaction Compartments. Nat. Commun. 9. DOI: 10.1038/S41467-018-06379-8.
95. Obexer, R., Godina, A., Garrabou, X., Mittl, P. R. E., Baker, D., Griffiths, A. D., and Hilvert, D. (2017) Emergence of a Catalytic Tetrad during Evolution of a Highly Active Artificial Aldolase. Nat. Chem. 9. 50–56.
96. Ghadessy, F. J., Ong, J. L., and Holliger, P. (2001) Directed Evolution of Polymerase Function by Compartmentalized Self-Replication. Proc. Natl. Acad. Sci. U. S. A. 98. 4552–4557.
97. Ellefson, J. W., Meyer, A. J., Hughes, R. A., Cannon, J. R., Brodbelt, J. S., and Ellington, A. D. (2014) Directed Evolution of Genetic Parts and Circuits by Compartmentalized Partnered Replication. Nat. Biotechnol. 32. 97–101.
98. de Vega, M., Lazaro, J. M., Salas, M., and Blanco, L. (1996) Primer-Terminus Stabilization at the 3'-5' Exonuclease Active Site of Phi29 DNA Polymerase. Involvement of Two Amino Acid Residues Highly Conserved in Proofreading DNA Polymerases. EMBO J. 15. 1182–1192.
99. Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. Nat. Methods. 6. 343–345.
100. MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: An Open Source Document Editor for Creating and Analyzing Targeted Proteomics Experiments. Bioinformatics. 26. 966–968.

3

Adaptive Evolution of Self-replicating DNA in a Synthetic Protocell

Abstract

The replication of the information-carrying molecule, along with the introduction of mutations and selection of the fittest variants, are fundamental principles that drive evolution in biology. We and others have postulated that the re-construction of a synthetic cell in the laboratory will be contingent on the development of a genetic self-replicator capable of undergoing Darwinian evolution. Although DNA-based life dominates, the in vitro reconstruction of an evolving DNA self-replicator has remained elusive. We hereby emulate in liposome compartments the principles according to which life propagates information and evolves. Using two different experimental configurations supporting intermittent or continuous evolution (i.e., without DNA recovery by PCR and re-encapsulation), we demonstrate sustainable replication of a linear DNA template encoding the DNA polymerase and terminal protein from the Phi29 bacteriophage, expressed in the 'protein synthesis using recombinant elements' (PURE) system. The self-replicator is able to survive across multiple rounds of replication-coupled transcription-translation reactions in liposomes, and, within only a few evolution rounds, it accumulates mutations conferring a selection advantage. Combined data from next generation sequencing with reverse engineering of some of the enriched mutations reveal nontrivial and context-dependent effects of the introduced mutations. The present results are foundational to build up genetic complexity in an evolving synthetic cell, as well as to study evolutionary processes in a minimal cell-free system.

The content of this chapter is based on a co-first authorship manuscript in preparation with Zhanar Abil* and Christophe Danelon. Other co-authors include Andreea Stan, and Amélie Châne, Miguel de Vega, Alicia del Prado, and Yannick Rondelez.

*Denotes equal contribution.

Introduction

The holy grail of modern synthetic biology is the construction of synthetic life. The many discoveries that lie in the path to creating synthetic life will spearhead advances in biomedicine, biotechnology, and fundamental biology [1–6]. In the quest for constructing artificial life, a number of life's features have been reconstituted in a cell-free environment [7–13], although a functionally integrated, autonomous synthetic cell seems still out of reach. One of the remarkable abilities of extant living forms is evolution, i.e., the ability to genetically diversify and adapt to changing environments. This ability is responsible for terrestrial Life's extraordinary robustness that allowed it to continuously exist and survive multiple geological calamities in the past 3.5-3.8 billion years [14,15]. In fact, some theories suggest that evolvability was the cause of life's emergence in the first place [16]. We therefore asked: Can we recreate such evolvability in a non-living synthetic biochemical system? Furthermore, can this ability assist us in our efforts to build synthetic life and better understand living processes?.

Evolution, whether it is chemical or biological, is enabled by the ability of individuals in a population to replicate (reproduce), diversify (mutate), and amplify differentially (undergo natural selection). Diversification and differential amplification are both outcomes of variability in replication. Thus, replication is a key prerequisite for an ability to evolve. Replication refers to the process by which two or more copies of the genomic material are produced from one parental molecule. This process is crucial for information continuity during cell proliferation, dynamic stability of a population, and generation of diversity in a population, all of which are prerequisites for the long-term stability and adaptability of a population of replicators. Clearly, as the appearance and evolution of molecular replicators were critical steps in the origin of life on Earth, their in vitro reconstitution represents a major step in crafting synthetic cells from the ground up.

Multiple examples of Darwinian evolution of oligomeric or polymeric replicators in cell-free environments have been demonstrated previously [17,6,18]. When considering the realization of a synthetic system capable of self-replication and evolution, the chemistry on which these processes are based may differ from that found in contemporary organisms; yet the functions that embody life must be retained. A variety of self-replicating systems have been created in the laboratory. For example, non-enzymatic self-replication based on autocatalytic template production [19,20], cross-catalysing RNA replicators [21], self-replicating peptides [22–25], vesicles [26], micelles [27–29], supramolecular polymerization [30], and cooperative replicating RNA networks [31]. These studies showed that populations of molecular replicators can respond to selection pressure, exhibit exponential growth, feature emergent traits of heritability and selection, and in a few cases are capable of undergoing Darwinian evolution. However, in such systems with unified genotype and phenotype (i.e., the information carrier and information processing machine are not physically separated), heredity and evolution are restricted due to limited structural space [32].

In contrast, the separation of genotype and phenotype would increase the randomness of the sequence-structure map, and thus, significantly increase a system's ability to evolve [32]. Such a system was studied in vitro by Ichihashi and colleagues, who built a translation-coupled RNA replication system [33]. They performed evolutionary experiments of self-replicating RNA molecules by self-encoded Qβ replicase in droplet compartments [33–38]. However, it would be exceedingly challenging to develop a form of synthetic life that is fully encoded on an RNA-genome. RNA is highly unstable when compared to DNA, most extant life is DNA-based, and the majority of currently available tools for regulation and processing of nucleic acids are based on systems with a DNA genome. Moreover, RNA-based Qβ-replicase system suffers from poor template generality, which can limit genome expansion to encode more functionalities.

Bacteriophage Phi29-based minimal DNA replication is a promising strategy for building the synthetic genome of a protocell [7,39]. Rolling circle replication (RCR) has been proposed [39] but its application for replicating genetic material requires a recombination step, which until this day has not been very efficient [40]. The recombinase considerably inhibits the replication step, limiting DNA amplification [40]. Moreover, RCR with or without recombination results in a complex mixture of monomeric (circular) and polymeric (repetitive) clones [41–43]. In contrast, the protein-primed linear DNA replication that we previously implemented inside gene-expressing vesicles [7] enables complete self-regeneration after each incubation round. With our system, the replication products are linear monomeric clones that are identical to their parent, thus better mimicking extant natural life's information flow. Moreover, since reproduction of the parental DNA does not require further processing steps [7,44], it is most suitable for building synthetic biological systems via an evolutionary approach [5]. However, it remains unknown whether such a protein-primed minimal DNA self-replicator is capable to persist and adapt in a compartmentalized cell-free environment.

Herein, we report on the synthetic evolution of transcription-translation-coupled self-replicating DNA molecules in protocellular compartments. Sustained replication of a sequence-optimized and an engineered mutator DNA is established inside PURE system-containing phospholipid vesicles. Moreover, to circumvent the finite pool of substrates and limited lifetime of the biochemical processes, while enabling redistribution of the DNA content across vesicles, a freeze-thaw cycle-based method is applied. Improved self-amplification, and enrichment of synonymous and nonsynonymous mutations are demonstrated, providing an evolutionary path for the emergence of a self-replicating synthetic cell.

Results and Discussion

Overall design strategy

For the design of our DNA self-replicator, we drew inspiration from the replication mechanism of the *Bacillus subtilis* bacteriophage Phi29 genome. In vitro replication of hete-

rologous DNA with four purified proteins has already been reported [45]. Moreover, in an earlier study, we showed that a synthetic DNA encoding the Phi29 DNA polymerase (DNAP, from gene *p2*) and terminal protein (TP, from gene *p3*), named *ori-p2p3*, can be self-amplified when expressed in the PURE system in the presence of purified auxiliary proteins (double-stranded and single-stranded binding proteins: DSB and SSB), and dNTPs. The linear DNA template encompasses two transcriptional units and two origins of replication, one at each end **(Fig. 1a)**. Each gene was codon optimized for improved expression with an *E. coli*-based translation machinery and was cloned between a T7 promoter and either vsv-r1 and vsv-r2 or a T7 transcription terminator, thus constituting a chimeric, synthetic DNA construct.

To maintain a genotype to phenotype link as well as to limit the propagation of molecular parasites, the *ori-p2p3* template, along with PURE*frex*2.0 and the replication substrates and cofactors for in vitro transcription-translation-coupled replication (IVTTR), were compartmentalized inside liposomes with a lipid composition that resembles that of the *E. coli* inner membrane [7]. Due to the importance of membrane vesicles in extant biology, we found it pertinent to investigate the compatibility of our DNA self-replicator with liposomes (vs. emulsion droplets or other artificial scaffolds for synthetic cells), and the possibility of performing evolution in such a combined system.

Replicator engineering

First, we aimed to optimize the sequence of our original template *ori-p2p3* for long-term evolutionary experiments. During IVTTR, the main self-replication product of the size 3.2 kb was generated, as well as an unexpected additional band of the size 1.4 kb **(Fig. 1)**. With the concern that this fragment could be a self-produced "molecular parasite" that could significantly hinder an evolutionary experiment, we decided to explore the nature of this shorter fragment and find possible ways to prevent its re-appearance. Sanger sequencing of the gel-purified 1.4 kb DNA fragment revealed that this band corresponded to a shortened *ori-p2p3* that missed the DNAP gene-expression cassette (T7 promoter-*p2*-T7 terminator), but still had the origins and the entire TP transcriptional unit intact. Thus, it could potentially act as a potent parasite. We hypothesized that the deletion of the DNAP cassette was a result of a recombination event between two identical 83-bp DNA sequences on the *ori-p2p3* template that spanned the T7 promoter, the g10 leader sequence, and a ribosome binding site upstream the two genes. To diminish the recombination frequency during IVTTR, we modified the repeated T7 g10 leader sequence upstream of the TP gene on the template **(Fig. 1b)**. We hypothesized that the strong stem-loop structure that the original leader sequence forms on the transcribed mRNA, as predicted by RNAfold web server [46], could be substituted with an artificial sequence with a similar mRNA secondary structure for avoiding recombination without affecting gene expression of TP **(Fig. 1b)**. We therefore constructed a modified *ori-p2p3* template wherein a string of 33 bp of the leader sequence upstream of the TP gene was replaced by an alternative sequence that was predicted to form

a similar mRNA stem loop despite no primary sequence similarity to the original leader sequence. This resulted in a modified *ori-p2p3* template (we called *mod-ori-p2p3*) with a similar TP expression **(Fig. 1c)** and self-replication ability as the original *ori-p2p3* **(Fig. 1 d,e)**, corroborating the idea that the secondary RNA structure of the g10 leader sequence, and not its primary sequence, is somehow crucial for gene expression. More importantly, *mod-ori-p2p3* produced significantly less of the smaller 1.4 kb-product **(Fig. 1 d,e)**, supporting the recombination hypothesis and suggesting that *mod-ori-p2p3* is a better starting template for our in vitro evolutionary experiments.

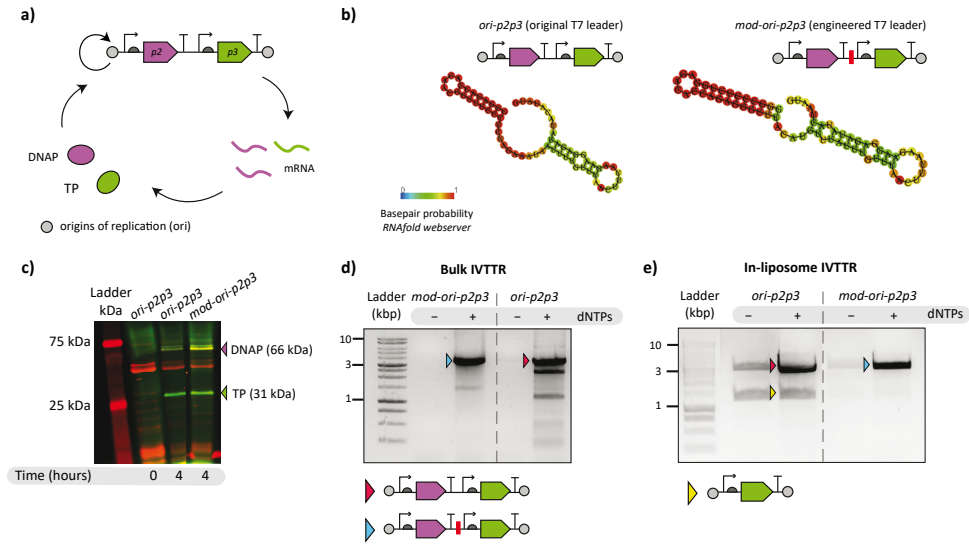


Fig. 1 Engineering of a DNA self-replicator. a) Schematic illustration of a DNA self-replicator encoding DNAP and TP from the Phi29 bacteriophage DNA replication apparatus. **b)** Predicted leader sequence RNA structures [46] from *ori-p2p3* and *mod-ori-p2p3* DNA templates. **c)** SDS-PAGE gel showing protein production from bulk IVTT reactions with *ori-p2p3* and *mod-ori-p2p3* DNA template. Each reaction solution was supplemented with Green-Lys reagent for co-translational protein labelling [47]. **d)** Agarose gel electrophoresis of recovered DNA from bulk IVTTR with *ori-p2p3* or *mod-ori-p2p3* DNA template. **e)** Agarose gel electrophoresis of PCR-recovered DNA from in-vesiculo IVTTR with *ori-p2p3* or *mod-ori-p2p3* DNA template.

Reducing the set of proteins for self-replication in liposomes

Next, we sought to optimize the compartmentalized IVTTR reaction. We have previously shown that both SSB and DSB are required in high amounts for efficient self-amplification of *ori-p2p3* in bulk IVTTR reactions [7]. However, we also observed that in an orthogonal DNA amplification setting (i.e., where the amplified DNA encodes a protein other than DNAP and TP), a reduced amount of DSB improved the yield of expressed protein from the replicating template [44]. We therefore investigated if reducing DSB concentration could improve self-amplification in liposomes. Surprisingly, we found that *mod-ori-p2p3* self-amplifies efficiently in liposomes even in complete absence of DSB **(Fig. S1 a,b)**. However, consistent with bulk IVTTR, that strongly relies on the presence of DSB, template DNA

outside of liposomes did not amplify without DSB **(Fig. S1 c-f)**. Therefore, we chose to omit DSB in our follow up evolution protocol.

Error-prone IVTTR with an exonuclease-deficient DNAP variant

We wondered if it is possible to modulate the rate of evolution by implementing a more error-prone variant of Phi29 DNAP. We explored the position F62 **(Fig. 2a)** (reported as F65 by some authors, due to discrepancies in the delimitation of the translation initiation site. Our delimitation corresponds to the NCBI Reference Sequence: NC_011048.1). The F62 residue is located in the N-terminal exonuclease domain of DNAP **(Fig. 2b)**, and is involved in binding to the ssDNA that is predicted to form when the template/primer DNA melts [48]. The F62Y mutation was reported to reduce exonucleolytic activity of DNAP and increase the frequency of nucleotide misincorporation in vitro, while only mildly affecting the DNAP's TP-deoxynucleotidylation, TP-DNA initiation, and TP-DNA amplification [48]. To test TP-primed DNA amplification activity by the F62Y variant of DNAP, we first constructed a plasmid containing the circularized *mod-ori-p2(F62Y)p3* (hence it cannot self-replicate) and performed orthogonal DNA amplification of an origin-flanked unrelated gene, *ori-pssA* (*pssA* coding for *E. coli* phosphatidylserine (PS) synthase), in a bulk IVTTR reaction. Absolute quantification of the *pssA* gene by quantitative polymerase chain reaction (qPCR) before and after 16 hours of reaction revealed a yield of amplified DNA comparable to that performed by wildtype (WT) DNAP **(Fig. 2c)**. Secondly, we compared the in-liposome self-replication activity of *mod-ori-p2p3* and *mod-ori-p2(F62Y)p3*, and found that both versions reached similar amplification folds in our system **(Fig. 2d)**. We therefore decided to also use this mutator polymerase, along with the wildtype DNAP, in some of our evolution experiments.

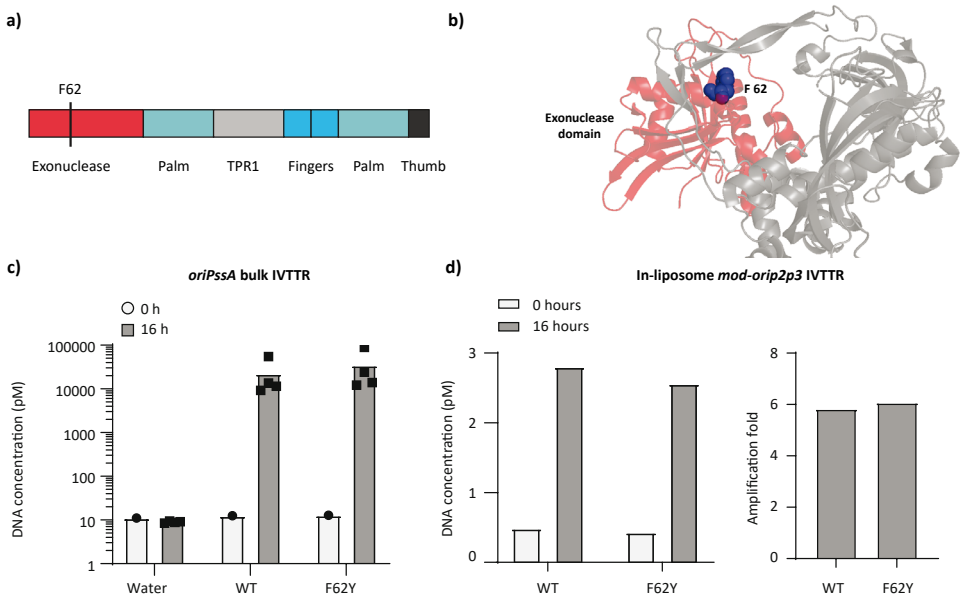


Fig. 2 A DNA self-replicator with deficient proof-reading activity. a) Schematic of Phi29 DNAP protein domains along its protein sequence with indicated F62 as the chosen amino acid to alter for engineering a more error-prone DNAP variant. **b)** Structural model of Phi29 DNAP (PDB code 2PY5) with the highlighted exonuclease domain in red, and F62 position in blue. The adapted protein structure was generated with PyMOL Molecular Graphics System. **c)** Absolute DNA quantification of *ori-pssA* from orthogonal bulk IVTTR using a nonreplicating plasmid of *mod-ori-p2p3* (DNAP WT) and *mod-ori-p2(F62Y)p3* (DNAP (F62Y)). **d)** Absolute DNA quantification and amplification folds from in-liposome self-replication reactions with *mod-ori-p2p3* or *mod-ori-p2(F62Y)p3* as DNA template.

Enrichment of an active self-replicator from a larger pool of DNA

We next asked if our in-liposome IVTTR method is viable for in vitro self-selection of DNA replicators. The selection principle of active (or more active) self-replicators would be based on their ability to clonally and differentially amplify within individual liposomes, thus producing more copies of themselves in the total population of DNA for the next round of evolution **(Fig. S2 a)**. The absolute amount of inactive (or less active) DNA molecules would remain the same or increase to a smaller extent, decreasing their fraction in the total population of DNA in the next generation of replicators. This phenomenon would be similar in concept to natural selection in populations of living organisms and to compartmentalized self-replication (CSR, methodology for directed evolution of polymerases [49]). The main methodological difference when compared to CSR is the absence of an in vivo gene expression step, which also obviates molecular cloning, and thus significantly simplifies the protocol. Another difference is that in our compartmentalized IVTTR, all the steps, including transcription, translation, and replication happen in a single pot, in a continuously self-enhancing positive feedback loop, which in principle could allow an increase in enrichment efficiency.

To test for the viability of our selection method, we performed an enrichment experiment starting from a mock library consisting of *mod-ori-p2p3* and a 50-fold molar excess of an unrelated gene (*plsB*, coding for *E. coli* glycerol-3-phosphate acyltransferase) of similar length and flanked with Phi29 replication origins, *ori-plsB*. To ensure stringent genotype-to-phenotype link, we encapsulated 10 pM DNA mixture in liposomes, a concentration permitting on average 0.2 DNA molecule per liposome considering an average vesicle diameter of 4 µm and a Poissonian partitioning [44]. To allow IVTTR exclusively within liposomes, the DNA outside of liposomes was digested with externally added DNase I. After 16 hours of in-liposome IVTTR at 30 °C, we measured the concentration of each of the genes (*p2* and *plsB*) before and after incubation by absolute qPCR quantitation, and discovered that the fraction of *mod-ori-p2p3* in the mixture had shifted from 1.7% to 16% **(Fig. S2b)**, which is a roughly 10-fold increase in *mod-ori-p2p3* fraction. These results imply that our developed strategy of in-liposome IVTTR of a self-replicator can support an in vitro evolution campaign.

Survival of self-replicators over multiple rounds of intermittent evolution

We asked whether it is possible to use an in-liposome intermittent protocol to establish

a persistent population of an evolving self-replicator system **(Fig. 3a)**. In other words, throughout the repeated cycles of experimental evolution, would the DNA gradually (i) stop replicating (go extinct), (ii) retain its initial self-replication activity (neutral drift), or (iii) improve its self-replication activity (adaptive evolution). We started our evolutionary campaign with *mod-ori-p2p3* linear PCR fragment. At each round of in vitro evolution, we encapsulated in liposomes the IVTTR reaction mix (no DSB added) along with 10 pM DNA (λ = 0.2). Amplification reaction outside of liposomes was prohibited by adding DNase I after vesicle formation. Clonal amplification of self-replicators was performed at 30 °C for 16 hours, after which the DNase I was thermally inactivated. The DNA was then released from the vesicles by an osmotic shock, diluted, and further amplified using conventional bulk PCR. The expected *mod-ori-p2p3* size (~3.2 kb) was verified by agarose gel electrophoresis, and, whenever needed (i.e., if extra bands of lower DNA sizes were observed), the full-length *mod-ori-p2p3* DNA band was gel-purified to limit the possible propagation of molecular parasites to the next IVTTR round. The resulting DNA was carried on to the next round of evolution and encapsulated again at λ = 0.2. The sequence diversity was allowed to accumulate passively by Phi29 DNA polymerase in situ during IVTTR (10^{-5} to 10^{-6} subs/base/doubling [50–52], and by PCR DNA recovery with KOD DNA polymerase ($0.7\text{-}1.2\times10^{-5}$) [53,54].

We performed 12 rounds of intermittent in-liposome evolution, and called this evolutionary campaign Int-WT(1). We quantified the initial and final amounts of DNA at each round and discovered that within 12 rounds of in vitro evolution, the amplification of self-replicating DNA improved at least 5-fold **(Fig. 3 b,c)**. Likely due to the size selection by the DNA PCR recovery and band excision from gel electrophoresis, the length of the amplified DNA did not change over the course of evolution, and no additional DNA products were observed **(Fig. 3d)**. We repeated the Int-WT(1) in a separate evolutionary experiment (Int-WT(2)), and the DNA amplification profile of this experiment once again confirmed persistent self-amplification of the DNA replicator throughout the evolution campaign **(Fig. 3e)**. This time, however, there was neither tangible improvement, nor deterioration of DNA replication levels by the end of the experiment. Perhaps a different evolutionary trajectory was undertaken on Int-WT(2), leading to no detectable improvement in amplification fold by qPCR.

To investigate the impact of a higher rate of mutation accumulation during IVTTR on the evolutionary dynamics, we applied the same protocol for in vitro evolution of *mod-ori-p2(F62Y)p3* variant, and called this campaign Int-MUT. Interestingly, the amplification of self-replicating DNA improved two rounds earlier than in the int-WT(1) campaign **(Fig. 3f)**. By the 9-th round of evolution, we increased the selection stringency of Int-MUT evolution by reducing the IVTTR incubation time from 16 hours to 4 hours at 30 °C. As a result, the replication yield dropped in the last 3 rounds of evolution, after which we stopped the evolution campaign. Overall, we conclude that compartmentalized, transcription-translation-coupled self-replication of DNA using an intermittent evolution protocol is compatible with the survival of functional DNA replicators. In two instances over three

independent evolution campaigns, the DNA replicator could self-improve within only 10 rounds of evolution **(Fig. 3g)**.

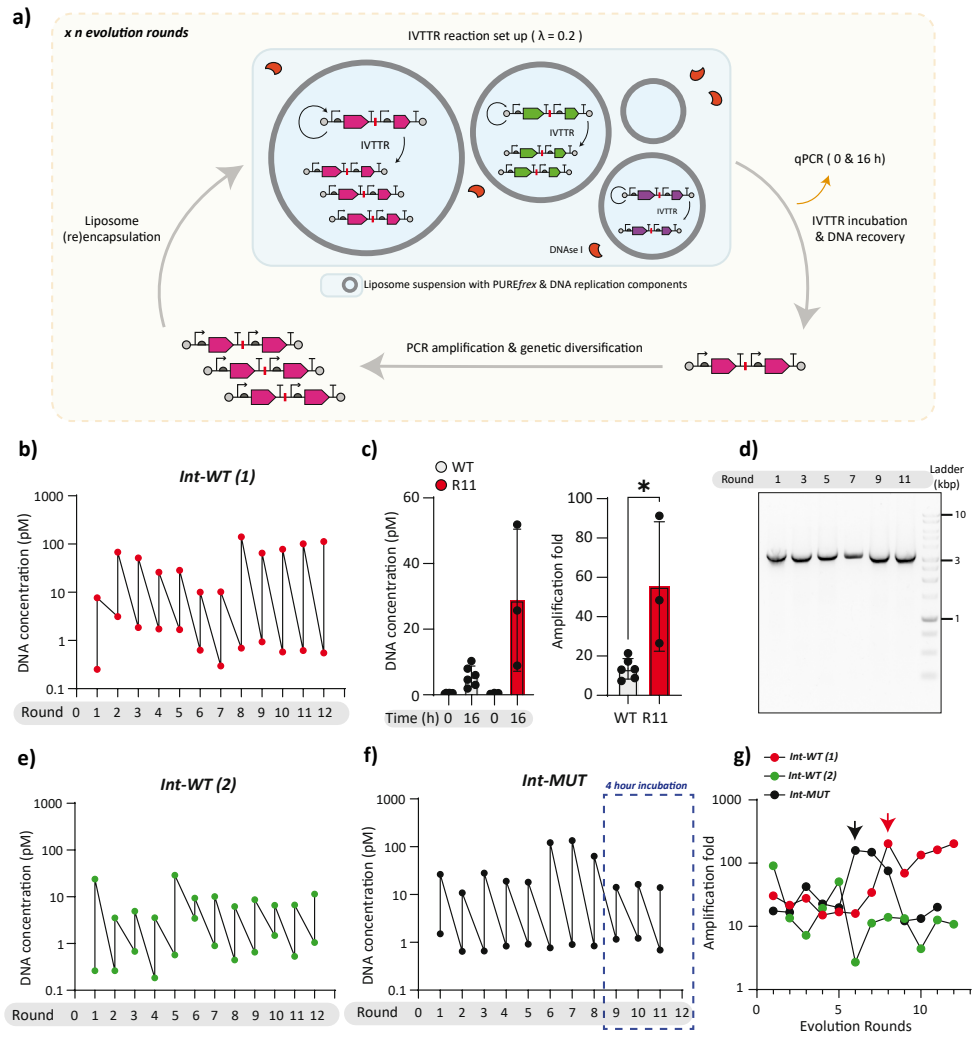


Fig. 3 Intermittent evolution of *mod-ori-p2p3* and *mod-ori-p2(F62Y)p3*. **a)** Schematic illustration of the experimental set-up for the intermittent evolution campaign of DNA self-replicators. **b)** Absolute DNA quantification along the evolution rounds of an intermittent evolution campaign, Int-WT(1). The target region for qPCR quantification (~200 bp) belongs to the *p2* gene of *mod-ori-p2p3*. **c)** Absolute DNA quantification and amplification folds from in-liposome IVTTR reactions with parental *mod-ori-p2p3* PCR (WT) and PCR recovered DNA from round 11 (R11) on Int-WT(1) campaign. **d)** Agarose gel electrophoresis of PCR-recovered DNA from six different rounds of Int-WT(1). **e)** Absolute DNA quantification along the evolution rounds of Int-WT(2), a biological replicate of Int-WT(1). **f)** Absolute DNA quantification along Int-MUT, the intermittent evolutionary campaign of *mod-ori-p2(F62Y)p3*, an error-prone replicator variant. The target region for qPCR was the same as in Int-WT campaigns. Highlighted in a blue square, the last three evolution rounds were incubated for 4 hours instead of 16 hours. **g)** Calculated DNA amplification folds from all evolution rounds on the three campaigns.

Black and red arrow heads point at the highest amplification folds obtained in R6 for Int-WT(l), and in R8 for Int-MUT. *P ≤ 0.05.

Freeze-thaw cycle-assisted continuous evolution of mod-ori-p2p3

Next, we investigated whether it is possible to minimize the researcher intervention in the in vitro evolution process via a more streamlined, "semi-continuous" evolution protocol. In such a system, amplified DNA would be passed on to daughter liposome compartments via liposome fusion and fission, obviating out-of-liposome PCR amplification and controlled re-encapsulation of DNA **(Fig. 4a)**. We devised a strategy that enables mimics of growth/division to be implemented with liposomes for repeated IVTTR without isolation of replicator DNA. Through fusion of 'feeding vesicles' onto the exhausted in-liposome IVTTR reactions, we supplied an excess of lipids and replenished the reaction with fresh PURE system constituents and, DNA replication factors. Both fusion and fission events are promoted by freeze-thaw (FT) cycles, during which releasing and stochastically re-entrapping the DNA content are expected. A similar protocol was used by Tsuji et al. for the replication of RNA over multiple rounds of liposome cultivation [55]. The assay required supplementing the replicating enzyme after each cycle and evolution was not demonstrated [55]. Serial transfer methods under compartmentalized conditions have also been applied to water-in-oil droplets for repeated RNA replication [56].

Fig. 4a summarizes the main steps of the continuous evolution cycle: (i) in-liposome IVTTR, (ii) 100-fold or 10-fold dilution with a solution containing feeding vesicles (same composition as the 'self-replicator vesicles' except that DNA was omitted), (iii) application of a FT cycle to promote liposome fusion-fission, hence proliferation, while pooling and stochastically re-entrapping the DNA content, (iv) continue with step (i), etc. We reasoned that the transfer of genetic information coupled to a new round of IVTTR would enable propagation of the self-replicator if DNA amplification overcompensates for the dilution effect caused by the addition of the feeding vesicles. The process is called 'semi-continuous' as it obviates extraction, out-of-liposome amplification, and re-encapsulation of DNA. However, it is not yet autonomous as compared with in vivo continuous evolution methods [57].

Membrane fusion and content mixing were confirmed by performing separate assays using flow cytometry **(Fig. S3)**. DNA leakage into the outer solution during FT was estimated to be 50% **(Fig. S4a)**, thus around half of amplified DNA would still remain inside liposomes (old and fresh), which we assumed was enough to preserve a sufficient amount of compartmentalized self-replicator for a next round of IVTTR. No external DNase was added in step (i) of the campaign to avoid digestion of encapsulated DNA caused by membrane permeabilization during FT. The template DNA *mod-ori-p2p3* was used and this evolution campaign was called Con-WT. While DSB is not essential for in-liposome IVTTR **(Fig. 3, Fig. S1)**, we observed a gradual decrease in DNA concentration during repeated FT cycles when DSB was omitted **(Fig. S4b)** (100-fold liposome dilution regime between rounds was

used). In the presence of DSB, DNA replicators persisted over at least five cycles **(Fig. 4b)** (at a 100-fold dilution regime between rounds). Accumulation of the full-length replicator in the course of evolution was verified by running PCR-amplified DNA (using ori-binding primers) from each round on an agarose gel **(Fig. 4c)**, suggesting that continuous DNA evolution is possible in the presence of liposomes. However, since unlike in the intermittent evolutionary approach, no DNase I was added in the external aqueous phase, it is not clear if the observed DNA population was due to replication solely inside of liposomes or also to external replication.

To limit replication of external DNA, we performed an independent evolution experiment, this time by reducing the IVTTR duration from 16 to 4 hours. We reasoned that external DNA amplification kinetics may be slower than that of internal **(Fig. S1f)**, due to the beneficial effects of molecular crowding and confinement on gene expression inside of liposomes. Total (inside and outside of liposomes) DNA concentration did not noticeably change in round 1 within 4 hours of IVTTR **(Fig. 4d)**, which is likely due to external DNA amplification kinetics being not high enough to reach a log phase within 4 hours **(Fig. S1f)**, and internal DNA amplification being unnoticeable as a result of internal DNA being only a small fraction of the total DNA concentration. Therefore, in this evolutionary experiment, the dilution factor was reduced from 100 to 10-fold to maintain a sufficient amount of DNA for the next round. As expected, no changes of DNA concentration were observed after one IVTTR round. During the three following rounds, total DNA concentration remained relatively constant due to roughly equal DNA amplification and dilution rates. Finally, DNA concentration gradually increased ~1700-fold from round 4 to round 8, with a ~225-fold amplification at round 8 alone **(Fig. 4d).** Analysis of DNA species flanked with origins of replication revealed an accumulation of the full-length replicator but also of lower-sized products that appeared at round 6 **(Fig. 4e)**. These results show persistent survival of the DNA self-replicator, indicating that DNA amplification rate improved since round 1 of continuous evolution in the presence of liposomes. However, it is still not clear whether the observed DNA replication, and possibly evolution, happens inside or outside of liposomes.

On the importance of compartmentalization

It has long been recognized that compartmentalization is important for the functional selection of self-replicating systems [58]. In particular, spatial organization can prevent the spread of nonfunctional replicons named parasites, resulting in the survival of compartments enriched with self-replicating molecules that would otherwise become extinct as parasites take over. The role of compartmentalization has been experimentally verified using PCR and RNA replication systems [59]. Moreover, transient compartmentalization was shown to be sufficient for selecting functional RNA replicators and purging the parasites [33]. To clarify the effect of liposomal compartmentalization on the evolutionary dynamics of our DNA self-replicator, we carried out a serial dilution experiment in the absense of liposomes. The parental DNA template *mod-ori-p2p3* was subjected to alternating steps of bulk IVTTR and

10-fold dilution with fresh PURE system and DNA replication components **(Fig. 4f)**. We here set the reaction time to 16 hours because the yield of amplified DNA is poor after 4 hours of non-compartmentalized IVTTR reaction. As anticipated, the DNA concentration gradually decreased and self-replication was totally suppressed at round 6 **(Fig. 4g)**. Agarose gel analysis of DNA samples after full-length recovery PCR revealed the presence of short replication products already at round 2, while the amount of the full-length replicator gradually diminishes **(Fig. 4h)**. These results suggest that parasite takeover was responsible for the extinction of self-replicators, demonstrating the importance of compartmentalizing liposomes for sustainable IVTTR.

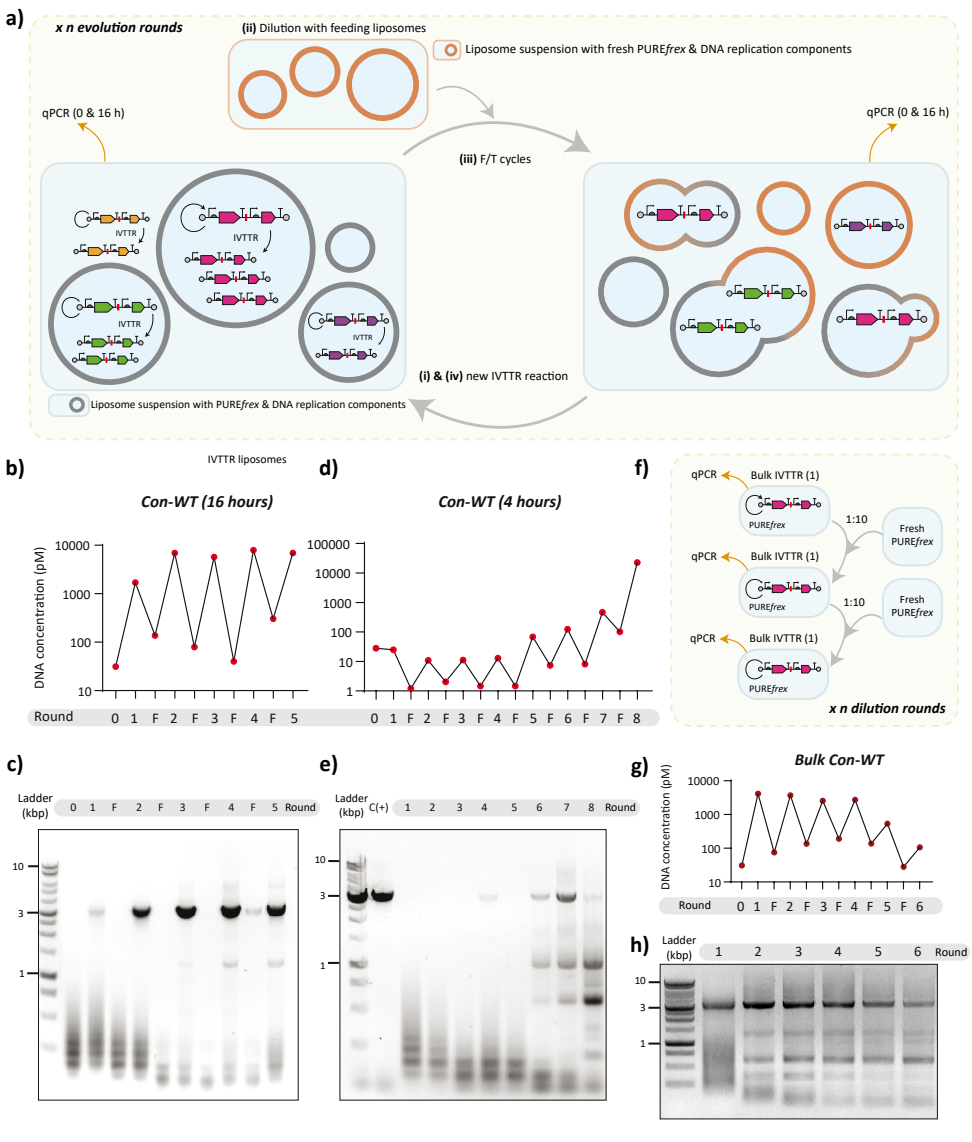


Fig. 4 Continuous evolution of *mod-ori-p2p3*. **a)** Schematic illustration of the experimental set-up for a continuous evolution campaign inside gene-expressing liposomes. Step (i) in-liposome IVTTR with the *mod-ori-p2p3* replicator, (ii) 10 or 100-fold dilution with a feeding vesicle solution, (iii) application of a FT cycle to promote liposome fusion-fission, and (iv) IVTTR starting the next round of continuous evolution. **b)** Absolute DNA quantification along the evolution rounds of the Con-WT campaign. Each IVTTR was incubated 16 hours and liposomes were diluted 100 times with feeding vesicles. The target region for qPCR quantification (~200 bp) belongs to the *p2* gene. 'F' letters indicate 'feeding' rounds with fresh vesicle solution (step (ii) and (iii)) **c)** Agarose gel electrophoresis of PCR-recovered DNA from all evolution rounds in Con-WT (16 hours). **d)** Absolute DNA quantification along the evolution rounds of an independent Con-WT campaign, where the IVTTR incubation time was set to 4 hours and dilution factor in step (ii) to ten. **e)** Agarose gel electrophoresis of PCR-recovered DNA from all evolution rounds of Con-WT (4 hours). **f)** Schematic illustration of the experimental set-up for the bulk serial transfer campaign (Bulk-WT). Bulk IVTTR reactions were incubated 16 hours Then, fresh PURE system and DNA replication factors were supplied in a 10-fold dilution step for starting the next round of IVTTR. **g)** Absolute DNA quantification along evolution rounds of the Bulk-WT experiment. **h)** Agarose gel electrophoresis of PCR-amplified DNA from all rounds of Bulk-WT.

Quantitative PCR targeting multiple regions scanning the entire length of *mod-ori-p2p3* was performed directly from liposome-diluted samples before and after each round of the three continuous evolution experiments. We wanted to examine whether different regions of the construct would amplify better or worse **(Fig. S5a)**, which may result from template-switching and recombination events that generate additional molecular species. In the two Con-WT campaigns (16 and 4 hours), the *p2* gene driving replication follows the dynamic pattern of the other targeted regions **(Fig. S5 b,c)**. In contrast, the abundance of *p2* in the bulk evolution experiment decreased faster than the other regions suggesting that shorter parasites outcompeted the self-replicator **(Fig. S5d)**. The takeover of molecular parasites and extinction of the full-length replicator in the absence of liposomes suggests that continuous evolution of a replicator in the presence of liposomes most likely happens inside of liposomes.

Overall, the rapid emergence, within a few evolution rounds, and persistence of a DNA self-replicator with improved activity are experimentally demonstrated in both intermittent and continuous evolutionary settings. We next sought to determine which genetic variations were acquired that conferred a selection advantage.

Emergence of DNA variants and fixation dynamics

To investigate the evolutionary processes that took place during the campaigns Int-WT(1), Int-WT(2), Int-MUT, and Con-WT, we deep-sequenced the PCR-amplified products of IVTTR at different evolution rounds using the Illumina next generation sequencing (NGS) technology after random fragmentation. We mapped and extracted the frequency of occurrence of all the point mutations that were detected at a frequency of at least 1% in at least one round. We found that in the evolving Int-WT(1) population, some of the mutations increased in frequency earlier in the rounds and decreased in the later rounds, while some of the others increased in frequency and became dominant in the later rounds **(Table 1) (Fig. S6)**. In particular, the nonsynonymous mutations S79G and A80T in the DNAP gene reached 67% and 26% frequencies by round 11 in the evolutionary campaign Int-WT(1).

For Int-WT(2), we observed an accumulation of different nonsynonymous mutations **(Table 1) (Fig. S6)** that, as shown above **(Fig. 3)**, did not seem to improve self-replication within the 12 rounds of intermittent evolution. Interestingly though, NGS results show that S79G and A80T mutations started to appear at low frequency on the later evolution rounds in Int-WT(2) (0.1-1%). A synonymous mutation on S79 position appeared on R6 (~0.1%), but on R11, S79G emerged with a low frequency (~0.6%). Similarly, A80T appeared at R6 with a low frequency (~0.1%), but accumulated through the consecutive rounds, reaching >1% frequency on R11. Both mutations would probably have become more dominant – similarly as in Int-WT(1) – if we had continued Int-WT(2) over 12 rounds. In the Int-MUT evolution experiment, a similar pattern of genetic diversification was observed throughout the evolution. However, we also observed a rapid and simultaneous takeover of a set of nine mutations, which became dominant by round 7 (87-90%). These included mutations in the DNAP gene Y62F (double mutation and reversal to WT from F62Y), A80T, I67 silent, E158G, F234L, in the TP gene S189G and L263P, and in the intergenic region 2997 A>G (T7 terminus) **(Table 1)**. This pattern suggests a contamination event from one of the later rounds of the Int-WT(1) evolution into the Int-MUT evolution, since all of these mutations can be found in the Int-WT(1) dataset, although they never reached allele frequencies this high in Int-WT(1). Unfortunately, the early-in-the-evolution replacement of the F62Y mutation meant that we could not reliably analyze its contribution to the evolution's course.

Two silent mutations in the *p2* gene K475 and K121, both found only in Int-MUT, are worthy of note. In both cases, we have two consecutive lysines, which together are encoded by 6A's in a row. Both of these silent mutations are AAA to AAG mutations in the first of the two lysine codons that disrupted the homopolymer runs of 6 A's. Repeated nucleotides are known to be a source of DNA polymerase-mediated frame-shift mutations in coding sequences [60], thus making them potential hubs for deleterious mutational hotspots. We hypothesize that these homopolymeric runs could act as local sources of genetic instability that would result in outcompetition by a more stable synonymous replicator. However, as it was also previously reported that in consecutive lysine sequences, homopolymeric A's can result in ribosome sliding and poorer translation [61], we cannot completely exclude protein expression level effect. Interestingly, the AAG codon at positions 475 and 121 of the *p2* gene is the original codon found in the Phi29 genome. They both have been substituted by the more frequent lysine codon AAA by the codon-optimization step to generate the parental *ori-p2p3* construct **(Table S1)**, but in vitro evolution changed them back to their native sequence. Other examples of such a codon reversal include I67 synonymous mutation, found on both Int-WT(1) and Int-MUT. Here, the ATT codon was reverted to ATC, which is originally present in the Phi29 genome and has a lower codon frequency **(Table S1)**. Mutations K475K, K121K, and I67I suggest that the genetic diversity in Int-MUT accumulated partly due to contaminating DNA from Int-WT(1), but also partly independently of Int-WT(1), generating a unique evolutionary pathway.

Some other enriched mutations in both Int-WT(1), Int-WT(2), and Int-MUT included

additional T's in the vsv-r1, vsv-r2, and T7 terminators. This could be another example of a mutational hotspot due to homopolymeric runs [60], in this case of T's, which may have been selected due to improved transcription termination [62]. In Con-WT, a synonymous mutation in DNAP gene restored the original Phi29 codon sequence (Y344Y with TAC >TAT) **(Table S1)**. However, in this case, TAT is a more frequent codon than TAC in *E. coli*. Synonymous mutations in the coding sequences may regulate protein expression profiles or even protein folding by controlling local translation rate [63].

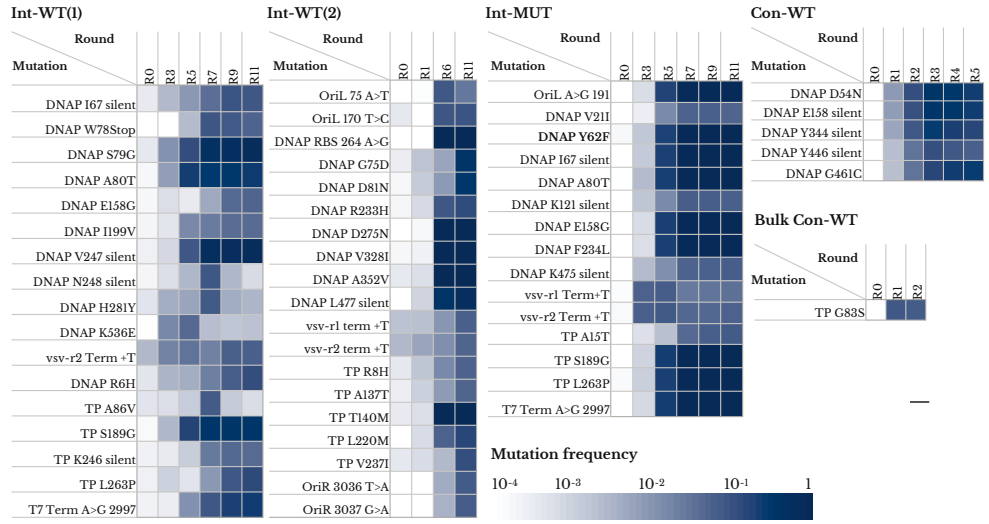


Table 1 Analysis of evolutionary patterns. Heatmap of mutation frequencies enriched to at least 5% at any round of in vitro evolution. Int-WT(1), trial 1 of intermittent in-vesiculo evolution on a starting WT (codon-optimized) sequence of *mod-ori-p2p3*. Int-WT(2), trial 2 of intermittent in-vesiculo evolution on a starting WT (codon-optimized) sequence of *mod-ori-p2p3*. Int-MUT, intermittent in-vesiculo evolution on a starting F62Y variant of codon-optimized *mod-ori-p2p3*. Con-WT(16 hours), continuous in-vesiculo evolution on a starting WT (codon-optimized) sequence of *mod-ori-p2p3*. And, Bulk-WT, serial transfer of bulk IVTTR reaction starting with a WT (codon-optimized) sequence of *mod-ori-p2p3*.

To estimate the accumulation of genetic diversity throughout the evolution, we integrated all the frequencies of mutations that were found above a certain threshold to approximate an average number of mutations found per single DNA molecule in the population. The frequency threshold was set to 0.1% since any mutation found at lower frequencies was likely a result of sequencing error (Phred scores were generally above 35). We found that the average number of mutations per DNA molecule increased from 0 before the start of evolution, gradually to 4 and 7 by round 11 of Int-WT(1) and Int-WT(2)/Int-MUT, respectively **(Fig. S7)**. In Con-WT, only one mutation on average per DNA molecule was reached after 5 rounds **(Fig. S7)**. These results suggest that the mutational load in our experiment was enough to generate quite a significant genetic diversity for selection to take place. We also estimated the total number of positions mutated (which were found to be at frequencies of at least 0.1%) in the entire DNA population, and observed that the number increased from 0

to over 600 by R11 of Int-WT(1) and Int-WT(2), and over 800 of Int-MUT **(Fig. S7)**.

Notably, the mutation accumulation rate was much lower for the in vesiculo Con-WT experiment, plateauing at around 1 mutation/molecule after round 3, whereas the intermittent evolution method kept accumulating mutations throughout 10 rounds of evolution **(Fig. S7)**. Total mutated positions per evolution round was also much lower for the continuous evolution method **(Fig. S7)**. These data suggest that the PCR recovery used between rounds of mutation in Int-WT introduces a significant amount of the DNA diversity. Therefore, using a mutator DNAP variant might be more beneficial for faster accumulation of mutations in continuous evolution.

Overall, the improved amplification (fitness increase) in Int-WT(1) and Int-MUT, and enrichment of mutations in all intermittent evolutionary campaigns suggest that our system can support evolutionary adaptation.

Characterization of enriched variants

We next characterized some of the most enriched missense mutations in Int-WT(1) and Int-MUT. To assess whether mutations S79G and A80T in the DNAP gene were sufficient to improve DNA amplification, we created single and double mutants starting from the parental template, and subjected them to in-liposome and bulk IVTTR. In-liposome IVTTR of the two single-mutant constructs led to a 4-fold improvement in the final DNA yield compared to the parental template (*mod-ori-p2p3*), while the enhancement of the combined mutations (S79G-A80T) brought only moderate improvement, if any **(Fig. 5 a,b)**. Since similar amplification increase was observed in R11 of Int-WT(1) compared to the parental sequence **(Fig. 3b)**, we conclude that S79G or A80T mutations independently accounted for the majority of the self-amplification improvement observed in the population. Additionally, in-vesiculo IVTTR kinetics showed a slightly higher apparent self-replication rate for S79G variant DNA when compared to the parental (WT) template **(Fig 5c)**.

Under bulk IVTTR conditions starting with higher DNA concentrations (1-2 nM), no differences in DNA replication were observed between the parental DNA (round 0 PCR of *mod-ori-p2p3*), the reverse-engineered *mod-ori-p2p3* template harboring the S79G and A80T mutations, and the PCR-recovered DNA from round 11 on Int-WT(1) **(Fig 5d)**. This finding suggests a mutational fitness advantage in response to compartmentalization in liposomes, where macromolecular crowding, confinement, or membrane effects could play a role. Furthermore, we tested whether end-point evolved DNA could replicate under bulk IVTTR conditions without DSB, as this protein was absent in the intermittent evolution campaigns. Utilizing PCR-recovered DNA from round 11 of Int-WT1 as template, we showed that the evolved DNA was not apt to amplify without DSB in bulk **(Fig. S8)**. Together, these results suggest that our self-replicating system underwent evolutionary adaptation to the specific in-liposome IVTTR condition.

Further characterization of single variants did not reveal the exact mechanism by which

this improvement is achieved. End-point in vitro protein expression assays with GreenLys and liquid chromatography-mass spectrometry showed no improvement in the amounts of synthesized proteins **(Fig. S9)(Fig. S10)**. To disentangle protein property (e.g., activity, stability) from DNA template effects, we characterized the purified S79G, A80T, and the double mutant DNAP variants. None of them differ from the WT in their ability to replicate the original *mod-ori-p2p3* template DNA in bulk IVTT reactions **(Fig. S11 a,b)**. Hence, it is unlikely that the amino acid residue substitution in the translated DNAP protein improves its replication activity. Moreover, under different replication conditions, with or without coupled transcription, the DNA template harbouring the S79G mutation was not amplified better than the WT DNA template by purified WT or DNAP variants **(Fig. S11 a-c)**. This result indicates that S79G alone did not improve the template replicability either.

Curiously, we also observed that both purified S79G and A80T DNAP variants fail to replicate DNA in a Phi29 replication buffer **(Fig. S11 d,e)**, both in protein-primed and DNA-primed settings **(Fig. S12)**. Since the mutations are located in an exonuclease domain of DNAP, where residues often contribute to the stabilization of the template/primer complex [64,65], we suspected a defect in the interaction of the mutant proteins with DNA in certain conditions. To test this idea, we measured the affinity of the mutant proteins to the template/primer complex by electrophoretic mobility shift assay and observed reduced DNA binding **(Fig. S12)**. This result suggests that the replication deficiency the mutants display in certain conditions (e.g., the Phi29 replication buffer) and/or when replication is initiated by a DNA-oligomer primed mechanism, may be explained by the impaired stabilization of the 5' end of the template strand. Nevertheless, it is not clear if such destabilization of the DNAP-DNA complex could improve the protein-primed DNA replication in a higher ionic-strength and crowded environment, such as the conditions we employed in our self-replication system. In any case, the opposite effects of these mutations in different conditions suggests that this evolution experiment resulted in DNAP variants that are more fit only in the specific environment in which they were selected, in agreement with the directed evolution maxim 'you get what you select for' [66].

Reverse engineering of the silent mutations K121 and K475 (DNAP gene), and K188 (TP gene) was also performed. All single mutations led to a similar DNA amplification yield as the parental DNA **(Fig. 5a,b)**, suggesting that any beneficial effects may only become apparent in a richer genetic context, or that they act as hotspot stabilizing mutations without any direct effect on expression or replication.

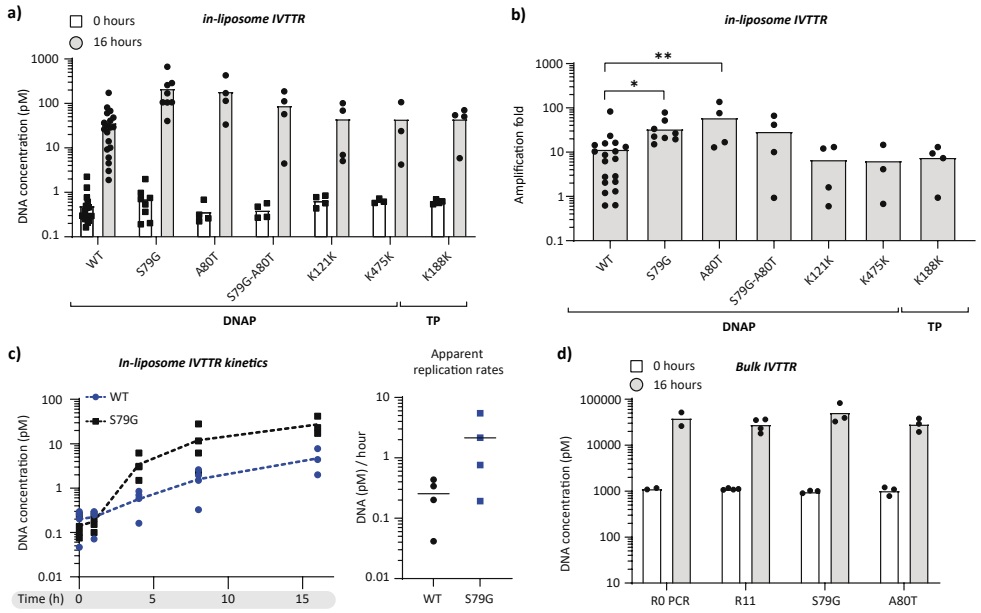


Fig. 5 Reverse engineering and characterization of fixed end-point mutations. a) Absolute DNA quantification from in-liposome IVTTR reactions with reverse-engineered self-replicators. qPCR target region (~200 bp) belongs to the *p2* gene in *mod-ori-p2p3*. **b)** Estimated DNA amplification folds (end-point DNA concentration / DNA concentration at time zero) from in-liposome IVTTR reactions displayed in panel a. An unpaired t-test was performed to compare amplification folds between WT and S79G or A80T variants **c)** Absolute DNA quantification at different time points of in-liposome IVTTR reactions (left panel). qPCR target region (~200 bp) belongs to the *p2* gene. DNA template was *mod-ori-p2p3* WT or S79G single mutant. Three biological repeats were performed per condition. Dashed lines connect the calculated mean values per time point across the replicates. Apparent maximum DNA replication rates, defined as the highest slopes in the kinetic curves, were calculated for each replicate (right panel). Horizontal dashed lines correspond to the mean value across WT and S79G replicates **d)** Absolute DNA quantification from bulk IVTTR reactions with parental *mod-ori-p2p3* (R0 PCR) employed in Int-WT(1), PCR-recovered DNA from round 11 of Int-WT(1) (R11), and the two reverse-engineered *mod-ori-p2p3* mutants encoding for DNAP(S79G) and DNAP(A80T). *P ≤ 0.05; ** P ≤ 0.01.

Conclusion and Outlook

This work shows that a de novo designed DNA self-replicator is capable of undergoing sustainable amplification and adaptability in a synthetic protocellular environment. Our primary goal was not to determine the most plausible evolutionary path for the origin of life on Earth, but to understand broader evolutionary principles and processes that can lead to the emergence of self-replicating, functionally integrated entities, and ultimately a synthetic cell. The discovery that adaptive evolution arose faster (< 10 rounds) than with repetitive-DNA replication [40], where mutation effects average out due to multiple gene copies per molecule, and faster than with the RNA/Qβ-replicase system [33,37] (> 100 rounds), makes our DNA self-replicating mechanism a good candidate for implementation in an evolving synthetic cell. For instance, we envision that the integration of more genes in this minimal DNA self-replicator constitutes the next step for co-evolving multiple cellular

functions in vitro through self-improvement.

Our self-replicators, both *mod-ori-p2p3* and to larger extent the mutator *mod-ori-p2(F62Y)p3*, enable the emergence and maintenance of genetic diversity in liposome populations such that selection can operate. A number of fixed mutations have been identified. Yet, the exact causes for mutational fitness advantage will have to be further investigated. Although the fixed mutations S79G and A80T located in the exonuclease domain of DNAP are sufficient to increase the replication yields in liposomes, no improvement of the activity of the purified DNAP variants could be detected. The results indicate that selection for increased self-replication efficiency is specific to the in-liposome IVTTR environment in which the mutations emerged, and that (genetic) effects other than protein property may also contribute to selection. Moreover, the specific factors defining the selection pressure for adaptive evolution remain to be explored. It is clear however that the hybrid nature of the self-replicating system (replication proteins are from the Phi29 phage, RNAP is from the T7 bacteriophage, the translation machinery is from *E. coli*, the DNA template (codon-optimization, regulatory elements, gene positioning, etc.) is synthetic, and liposomes are artificial lipid-based containers) applies on its own a strong evolutionary pressure. It will be interesting to adjust the selection conditions to alter replicative fitness.

Considering that the sequence space is strongly reduced compared to living organisms, including nearly minimal bacterial cells [67], it may be easier to understand the first principles of self-replicating systems due to the fewer targets on which positive selection can act. This provides an experimental testbed to evaluate hypotheses on the fundamental concepts of evolution in living systems and to predict how minimal cells respond to changing situations. Finally, our platform can be used to model viral replication of genomic DNA through transcription-translation in bacterial host organisms, as well as the underlying evolutionary mechanisms, with implications in the development of new therapeutic methods.

Materials and Methods

Buffers and solutions

All buffers and solutions were made using Milli-Q grade water with 18.2 MΩ resistivity (Millipore, USA). Chemicals were purchased from Sigma-Aldrich unless otherwise indicated.

Construction of DNA fragments for IVTTR reactions

Plasmid G340, which contains *mod-ori-p2p3* construct with mutated T7 leader upstream of *p3* gene, was prepared from G95 plasmid (original *ori-p2p3* construct from [7]). The fragment encoding the T7 mutated leader sequence was prepared by primer extension of the overlapping primer pair 1058 and 1060 ChD. The genes *p2* and *p3* were amplified from G95 plasmid using the primer pairs 1049/1056 ChD for *p2*, and 1057/1052 ChD for *p3*. The three fragments were assembled into a KpnI and HindIII-linearized pUC19 vector with Gibson Assembly [68]. Plasmid G371, containing *mod-orip2(F62Y)p3* and encoding for

Phi29 DNAP(F62Y), was cloned from G340 plasmid by focused PCR mutagenesis using 948/1132 ChD, and 1131/1137 ChD as primer pairs. The two overlapping DNA fragments were assembled into KpnI/PmeI-linearized G340 plasmid using the Gibson Assembly method [68]. Reverse engineered plasmids containing point mutations enriched over the evolutionary campaigns (G559 for DNAP(S79G), G570 for DNAP(A80T), G569 for DNAP(S79G&A80T), G560 for DNAP(K121K), G561 for DNAP(K475K), and G562 for TP(K188K)), were constructed by mutagenesis PCR utilizing G340 as a DNA template. After PCR, the reactions were treated with Dpn1 for digesting the parental G340 DNA template. The primer pairs used for each DNAP and TP mutagenesis PCR can be found in **Table S2**. All the plasmids were cloned by heat-shock transformation of *E. coli* Top10 strain, and plasmids were extracted from individual cultures outgrown in ampicillin containing LB using Promega PURE yield Plasmid Miniprep kit. Individual clones were screened and confirmed by Sanger sequencing.

To prepare linear *mod-ori-p2p3* DNA fragments for IVTTR experiments, a PCR was performed with phosphorylated primers 491 and 492 ChD. Reactions were set up in 100 µL volume, 500 nM each primer, 200 µM dNTP, ~10 pg/mL DNA template, and 2 units of Phusion High-Fidelity DNA Polymerase (NEB) in HF Phusion buffer. Thermal cycling was performed as follows: 98 °C 30 second initial denaturation, 20 cycles of (98 °C for 5 sec., 72 °C for 3 min.), and final extension at 72 °C for 5 min. Extra care was taken to not over-amplify the DNA by too many thermal cycles, as it was found to adversely affect the quality of purified DNA. The amplified PCR fragments were purified using Qiagen QIAquick PCR purification buffers and Qiagen RNeasy MinElute Cleanup columns using the manufacturer's guidelines for QIAquick PCR purification, except for longer pre-elution column drying step (4 min. at 10,000 g with open columns), and elution with 14 µL MilliQ water in the final step. The purified DNA was quantified by Nanodrop 2000c spectrophotometer (Isogen Life Science) and further analysed for size and purity by agarose gel electrophoresis.

Bulk IVTTR

Bulk replication reactions were set up in PURE*frex* 2.0 (GeneFrontier). A 20-µL reaction consisted of 10 µL solution I, 1 µL solution II, 2 µL solution III, 20 mM ammonium sulphate, 300 mM dNTPs, 375 µg/mL purified Phi29 SSB protein, 105 µg/mL purified Phi29 DSB protein, and 0.6 units/µL of SUPERase·In RNase inhibitor (Themo Fisher), and template DNA at the indicated amount (~2.3 nM for **Fig. 1**, ~50 pM for **Fig. 4**, and ~1 nM for **Fig. 5**). Reactions were incubated in a nuclease-free PCR tube (VWR) in a Thermal Cycler (C1000 Touch, Biorad) at a default temperature of 30 °C. Incubation time was indicated when appropriate, variating from 4 to 16 hours. To analyse the reactions by gel electrophoresis, 10 µL reaction was treated with 0.2 mg/mL RNase A (Promega), 0.25 units RNase One (Promega) at 30 °C for 1-2 hours, followed by 1 mg/mL Proteinase K (Thermo Scientific) at 37 °C for 1-2 hours, and column-purified using the QIAquick PCR purification buffers (Qiagen) and RNeasy MinElute Cleanup columns (Qiagen) using the manufacturer's guidelines for QIAquick PCR purification, except for longer pre-elution column drying step (4 minutes at 10,000 g with open columns), and elution with 14 µL MilliQ water in the final step. A fraction (generally 6 µL) of the eluate was mixed with an equal volume of 6x purple gel loading dye (NEB) and loaded in 1% agarose gel with ethidium bromide, following which DNA was separated using an electrophoresis system (Bio-Rad). The BenchTop 1-kb DNA Ladder (Promega) was used to estimate the size of DNA.

Lipid-coated bead preparation

The procedure was adapted from [7] with minor modifications. To prepare lipid-coated beads, a lipid mixture consisting of DOPC (50.8 mol%), DOPE (35.6 mol%), DOPG (11.5 mol%), cardiolipin (2.1 mol%), DSPE-PEG(2000)-biotin (1 mass%) and DHPE-TexasRed (0.5 mass%) for a total mass of 2 mg and 25.4 µmol of rhamnose (Sigma-Aldrich) dissolved in methanol was assembled in a 5-mL round-bottom glass flask. All lipids were purchased at Avanti Polar Lipids and dissolved in chloroform, except the DHPE-TexasRed membrane dye (Invitrogen). Finally, 600 mg of 212–300-µm glass beads (Sigma-Aldrich) were added to the lipid solution, and the organic solvent was removed by of rotary evaporation at 200 mbar for ~2 h, followed by lipid beads collection, aliquoting, and overnight desiccation in individual 2 mL Eppendorf tubes. The dried lipid-coated beads were stored under argon at −20 °C.

Intermittent evolution of self-replicating DNA

Replication reactions were set up in PUREfrex 2.0 (GeneFrontier). A 10-µL reaction consisted of 5 µL solution I, 0.5 µL solution II, 1 µL solution III, 20 mM ammonium sulphate, 300 mM dNTPs, 375 µg/mL purified Phi29 SSB protein, 0.6 units/µL of Superase·In RNase inhibitor (Thermo Fisher), and 10 pM template DNA was prepared in a 1.5 mL Eppendorf tube. To the well-mixed reaction, 5 mg lipid-coated beads, already pre-desiccated for at least 20-30 min before use, were added. The 1.5 mL-Eppendorf tube containing the bead-PUREfrex mixture was next gently rotated on an automatic tube rotator (VWR) at 4 °C along its axis for 30 minutes for uniform liposome swelling. The mixtures were then subjected to four freeze/thaw cycles (5 seconds in liquid nitrogen followed by 10 minutes on ice). Using a cut pipette tip, 5 µL of bead-free liposome suspension (the beads sediment to the bottom of the tube) was transferred to a PCR tube, where it was mixed with 0.5 units of DNase I (NEB). Reactions were incubated in a nuclease-free PCR tube (VWR) in a Thermal Cycler (C1000 Touch, Biorad) at a default temperature of 30 °C for 20 minutes (for 0 hour sample), or 4-16 hours (whenever indicated), after which the DNase I was heat-inactivated at 75 °C for 15 minutes. Liposome suspension was then diluted 100-fold in milli-Q water. Diluted IVTTR reactions were used as templates for PCR amplification using phosphorylated primers 491 and 492 ChD. For this, reactions were set up in 100 µL volume, 300 nM each primer, 400 µM dNTPs, 10 µL diluted liposome suspension, and 2 units of KOD Xtreme Hotstart DNA polymerase in Xtreme buffer. Thermal cycling was performed as follows: 2 min at 94 °C for polymerase activation, and 25-30 cycles of (98 °C for 10 sec., 65 °C for 20 sec, 68

°C for 1.5 min). Extra care was taken to not over-amplify the DNA by too many cycles, as it was found to negatively affect DNA recovery during the next round of evolution. The amplified PCR fragments were size-separated on a 1% agarose gel containing SYBR Safe by gel electrophoresis. Whenever having additional bands on the gel electrophoresis (even if slight) that did not correspond to the expected *mod-ori-p2p3* band size (~3.2 kb), the DNA band with the expected size was excised and purified using QIAquick Gel Extraction Kit buffers (Qiagen) and RNeasy MinElute Cleanup columns (Qiagen) using the manufacturer's guidelines for gel extraction, except for longer pre-elution column drying step (4 minutes at 10,000 g with open columns). Final DNA elution was done with 14 µL of MilliQ water. The purified DNA was quantified by Nanodrop 2000c spectrophotometer (Isogen Life Science) and utilized as DNA template for the upcoming evolutionary round.

Quantitative PCR (qPCR)

10 µL reactions consisted of PowerUP SYBR Green Master Mix (Applied Biosystems), 400 nM each primer targeting the *p2* gene (976/977 ChD), and 1 µL of 100-fold diluted sample. The thermal cycling and data collection were performed on Quantstudio *5* Real-Time PCR instrument (Thermo Fisher), using the thermal cycling protocol 2 min at *50* °C, *5* min at 94 °C, 4*5* cycles of (1*5* sec at 94 °C, 1*5* sec at *56* °C, 30 sec at 68 °C), *5* min at 68 °C, and a melting curve from 6*5* °C to 9*5* °C. The concentration of nucleic acids was calibrated using 10-fold serial dilutions of corresponding standard DNA templates ranging from 1 fM to 1 nM. The data was analysed using the Quantstudio *5* Software (Thermo Fisher). DNA amplification folds were calculated with DNA concentrations at the IVTTR reaction end-point (generally 16 hours) / DNA concentrations at the starting point of incubation (0 hours).

NGS of evolutionary intermediates: Library preparation, sequencing, and data analysis.

DNA was PCR-amplified from 100-fold diluted liposome suspensions of evolutionary rounds as follows. Reactions were set up in 200 µL volume, 300 nM each primer, 400 µM dNTP, 20 µL of the diluted liposome suspension, and 4 units of KOD Xtreme Hotstart DNA polymerase in Xtreme buffer. Thermal cycling was performed as follows: 2 min at 94 °C for polymerase activation, and 25-30 cycles of (98 °C for 10 sec., 65 C for 20 sec, 68 °C for 1.5 min). The DNA was then purified using QIAquick PCR purification buffers (Qiagen) and RNeasy MinElute Cleanup columns (Qiagen) using the manufacturer's guidelines for QIAquick PCR purification, except for longer pre-elution column drying step (4 minutes at 10,000 g with open columns), and elution with 30 µL MilliQ water in the final step. The purified DNA was then prepared for deep sequencing using the Illumina Truseq DNA PCR free library preparation kit and deep sequenced using the Novaseq 6000 platform 150 bp paired end sequencing at Macrogen-Europe B.V.

To analyse NGS data, we utilized Galaxy, a web-based open-source platform for big data analysis at Usegalaxy.org. Workflows and data are available upon request. Using Galaxy available packages, we performed the following analysis steps. We mapped the paired reads

to the *mod-ori-p2p3* DNA sequence using the BWA software package [69,70] in BAM format using default options. Next, we used the MergeSamFiles tool to merge BAM datasets from different rounds of evolution into one set and marked duplicates to examine the aligned records for duplicate molecules. We then used the BamLeftAlign tool to realign indels in homopolymers and microsatellite repeats. We next applied the Filter tool to filter data on read mapping quality (≥20) and proper read pairing. We then utilized the FreeBayes tool, a bayesian genetic variant detector [71,72] to map and quantify the misalignments. The expected mutation rate was set to 0.0001. Requirement of minimal fraction of observations was set to 0.01 (for retrieving a list of all variants above 1%) or 0.001 (for quantifying all mutations above 0.1%). Requirement for the minimal count of observations supporting an alternate allele was set to 10. The data was then converted from VCF to tab-delimited format using the VCFtoTab-delimited tool. Further analysis, such as quantification of mutations above specified thresholds, and corrections of semantical errors on frequency calculations were performed in excel.

Purification of DNA polymerases, TP, SSB, and DSB proteins

Wildtype Phi29 DNA polymerase was expressed and purified as described in [73]. Terminal Protein, was expressed and purified as described in [45]. Single-stranded DNA binding protein and DSB were expressed and purified as described in [74] and [45], respectively. The DNA constructs for protein expression and purification of DNAP variants were obtained by site-directed mutagenesis using pJLPM as DNA template (a derivative of pT7-4w2) [75] containing the *p2* gene encoding for WT DNAP. Upon cloning, the entire *p2* gene was sequenced to verify the presence of the desired mutations and the absence of additional ones. All DNAP variants were expressed in *E.coli* BL21(DE3) cells and further purified essentially as described for the wildtype DNAP [73].

Bulk DNA replication with purified protein variants in PURE background

Bulk replication-transcription reactions were carried out with a modified PURE system that did not contain solution III (ribosomes). Replication-only reactions were performed with a customized PURE solution II minus T7 RNA polymerase (GeneFrontier Corp.). A 20 µL reaction solution was assembled with 10 µL solution I, 1 µL solution II, 10-20 mM ammonium sulphate, 300 µM dNTPs, 3 ng/mL of purified DNA polymerase variant, 3 ng/mL purified terminal protein, 375 µg/mL SSB protein, 105 µg/mL purified DSB protein, and 2 nM of indicated template DNA. Reactions were incubated for 16 hours in a thermal cycler (C1000 Touch, Biorad) at a temperature of 30 °C. Before and after incubation, 1 µL of sample was taken for DNA quantification with qPCR. For DNA sample analysis by gel electrophoresis, reaction solutions were incubated with 1 µL RNase A (4 mg/mL RNAse A solution, Promega), and 1 µL RNase One (10 units/µL RNase ONE Ribonuclease solution, Promega) for 1 or 2 hours at 30 °C. Solutions were then supplemented with 1.5 µL EDTA (100 mM), 1.5 µL SDS (1%), and 10 to 20 mg Proteinase K (Thermo Scientific). Samples were incubated at 50°C for 4 hours, and column-purified using the QIAquick PCR purification

buffers (Qiagen) and RNeasy MinElute Cleanup columns (Qiagen) using the manufacturer's guidelines for QIAquick PCR purification, except for an additional pre-elution column drying step (7 minutes at 10,000 g with open columns), and 10 to 20 minutes column incubation with 14 µL of ultrapure water (Merck Milli-Q) as the eluant for the final step. A 7 µL fraction of the eluate was mixed with 3 µL of 6x purple gel loading dye (NEB) and loaded in 0.7-1% agarose gel with ethidium bromide, following which DNA was separated using an electrophoresis system (Bio-Rad). A Bench Top 1-kb DNA Ladder (Promega) was used to estimate the size of DNA.

Bulk DNA replication with purified protein variants in replication buffer

A 20 µL replication-only reaction solution was assembled, consisting of 1x Phi29 replication buffer (50 mM Tris-HCl (pH 7.5), 10 MgCl2, 5% glycerol, 1 mM DTT, 10 or 20 mM of amonium sulphate) (NEB), 0.625 ng/µL of purified DNAP, 1.25 ng/µL of purified TP, 375 µg/mL purified SSB, 105 µg/mL purified DSB, 400 µM of dNTPs, and 2 nM of linear DNA template. When required, purified DNAP and TP protein stocks were diluted with a buffer containing 100 mM NaCl, 0.05% Tween-20, and 25 mM Tris-HCl, prior to their addition into the reaction solution. Samples were incubated for 16 hours in a thermal cycler (C1000 Touch, Biorad) at a temperature of 30 °C. Before and after incubation, 1 µL of sample was taken for DNA quantification with qPCR. For gel electrophoresis analysis, sample treatment and preparation of agarose gels were as described above.

Co-translational labelling and gel fluorescence imaging of expressed proteins

Standard 20 µL PUREfrex2.0 (GeneFrontier Corp.) reaction solutions were assembled on ice (10 µL solution I, 1 µL solution II, 0.5 µL solution III, 0.6 units/µL of Superase·In RNase inhibitor (Ambion), and 1 nM of linear template DNA) and supplemented with 1 µL of BODIPY-Lys-tRNALys (FluoroTect GreenLys, Promega) to incorporate fluorescently labelled lysine residues in the synthesized proteins. Samples were incubated at 37 °C for 16 hours for protein expression, after which they were treated with 1 µL RNase A (4 mg/mL RNAse A solution, Promega), and 1 µL RNase One (10 units/µL RNase ONE Ribonuclease solution, Promega) for 1-4 hours at 37 °C. Ten microliters of treated samples were mixed with 4x Laemmli Sample buffer and DTT to reach a 15 µL volume and a final concentration of 1x Laemmli Sample buffer and 10 mM DTT, and were denatured for 5 min at 95 °C. Samples were analysed on a freshly prepared 12% SDS polyacrylamide gel electrophoresis (PAGE) gel. The loaded SDS gels were run for 15 min at 110 V, followed by 45 min at 180 V. Fluorescence imaging of the translation products was performed with a fluorescence gel imager (Typhoon, Amersham Biosciences). After fluorescence detection, the gels were stained overnight with Coomassie Brilliant Blue, de-stained overnight, and imaged on a Bio-Rad ChemiDoc Imager.

Continuous replication and evolution in liposomes.

Swelling solutions for in-liposome IVTTR were prepared using 50 pM *mod-ori-p2p3* DNA template, PUREfrex2.0, and DNA replication substrates, as explained above for the intermittent evolution protocol, except that DSB was included at a concentration of 8 µM. Feeding vesicles were produced with the same protocol, except that DNA was not added, aliquoted and stored at −80 °C directly after the last freezing step. For IVTTR, samples were incubated for 4 or 16 hours at 30 °C. Before and after incubation, 2 µL were collected with a cut pipette tip for DNA quantitation by qPCR, heat-inactivated at 75 °C for 15 minutes, and the samples were kept at −20 °C until further use. When indicated, 0.5 µL of DNase I (Promega) was added to the liposomes in order to digest the outer DNA, so that only the DNA present inside liposomes was quantified. To allow the DNase to act, the IVTTR solution was then incubated for 20 min at 30 °C before harvesting the time zero sample. To start a next evolution round, IVTTR-liposome samples were diluted either 100x (Con-WT 16 h) or 10x (Con-WT 4 h) with the feeding vesicle solution. The 100x dilution was realized in a two-step 10x dilution starting from 3 µL of IVTTR-liposome solution mixed with 27 µL of feeding vesicles. We gently pipetted up and down with a cut tip and kept a 2 µL sample to quantify the DNA concentration after feeding by qPCR. The remaining liposome solution was centrifuged for 5 minutes at 16000 r.c.f. at 4 °C. The tube was then dipped into liquid nitrogen for 5 seconds and left to thaw on ice for 10 minutes. Finally, liposomes were gently resuspended with a cut pipette tip, and incubated at 30 °C for a new IVTTR cycle. The procedure was repeated for a total of 4-8 cycles.

Serial transfer of bulk IVTTR reactions.

A 20 µL IVTTR reaction solution was assembled according to the protocol for preparation of the swelling solution for continuous replication and evolution in liposomes. The sample was incubated for 16 hours at 30 °C. A 2 µL sample was taken before and after the incubation step. After incubation, 2 µL of the IVTTR solution was diluted 10x with a feeding solution of the same composition except that DNA was omitted. After gentle pipetting up and down, the next IVTTR round was started by incubating at 30 °C for 16 hours. The procedure was repeated for a total of 6 rounds.

Supplementary Materials and Methods

LC-MS protein quantification **(Fig. S10)**

LC-MS/MS analysis was employed for the relative quantification of de novo synthesized DNAP and TP in bulk PURE reactions. 9-10 µL of the PURE reaction (no older than one week, stored at −20 °C) was mixed with 3 µL of heavily labelled QconCAT(^{15}N) [76], stored in a 50 mM Tris (pH 8.0) buffer containing 1 mM $CaCl_2$ to attempt absolute quantification as in [44], and with 12-13 µL of freshly prepared digestion buffer (12.5 mM tris-base, 12.5 mM tris-HCl, 1 mM $CaCl_2$, 5 mM TCEP). Next, the mixture was vortexed heavily, supplemented with 3.6 µL of 50 mM iodocamide, and incubated in the dark for 15 minutes. Then, 10 µL of trypsin (0.2 µg/µL) were added to each sample and the mixture was incubated overnight at 37 °C for protein digestion. The trypsin digested samples were centrifuged at

maximum speed (~14000-16000 g) for 30 minutes. 15 µL of the supernatant were collected and supplemented with 5-6 µL of 0.2% formid acid. The pH was checked with a pH strip to confirm the acidification of the solution (pH ~2-4). The mixture was then transferred to a glass vial with a small insert for LC-MS/MS analysis. Measurements were performed on a 6460 Triple Quad LCMS system (Agilent Technologies, USA). ~5.5 µL of sample were injected per run into an ACQUITY UPLC Peptide CSH C18 Column (Waters Corporation, USA). The peptides were separated in a gradient of buffer A (25 mM formic acid in Milli-Q water) and buffer B (50 mM formic acid in acetonitrile) at a flow rate of 500 µL per minute and at a column temperature of 40 °C. The column was initially equilibrated with 98% buffer A. After sample injection, buffer A gradient was changed to 70% (over the first 20 min), 60% (over the next 4 min), and 20% (over the next 30 sec). This final ratio was conserved for another 30 sec and the column was finally flushed with 98% buffer A to equilibrate it for the next run. The selected peptides and their transitions for both synthesized proteins and heavily labelled QconCATs were measured by multiple reaction monitoring (MRM). The recorded LC-MS/MS data was analyzed with Skyline for fraction calculation between unlabelled peptides from DNAP and TP proteins. MS/MS measurement details for each of the analysed proteins can be found in **Table S3**.

Flow cytometry for liposome fusion assays **(Fig. S4)**

Liposomes were produced from lipid-coated beads prepared as explained above using 0.5 mol% of either Texas Red or Oregon Green membrane dyes. Swelling solutions consisted either of PURE*frex*2.0 or an mCherry-encoding DNA (2 nM) in homemade PURE buffer (PB) consisting of 20 mM HEPES-KOH, pH 7.6, 180 mM potassium glutamate, and 14 mM magnesium acetate.. Vesicle fusion by F/T was achieved by mixing equivalent amounts (either 5 µL or 10 µL) of two different liposome populations, centrifuging for 1.5 minutes at 16000 r.c.f, flash-freezing the sample tube in liquid nitrogen, and thawing on ice. For assaying liposome content mixing, samples were incubated for 3-6 hours at 37 °C to allow for the expression of mCherry. One microliter of liposome samples was taken before and after F/T, diluted in 149 µL of swelling buffer, and filtered in 5 mL Falcon tubes (BD Falcon) with cell-strainer caps. Filtered diluted samples were pipetted into 96 U-shaped wells for flow cytometry analysis on a FACS Celesta flow cytometer (BD Biosciences). Liposomes were screened using the 488-nm laser line with 530/30 filter for detection of Oregon green, and the 561-nm laser line with 610/20 filter for detection of mCherry and Texas Red. Photon multiplier tube voltages were manually adjusted between 370-500 V for both laser lines, 375 V for the forward scatter light, and 260 V for the side scatter light. Loader settings were set to 50 µL injection volume with no mixing and 800 µL wash between sample runs. For each sample ~20000 events were recorded. The raw flow cytometry data was analysed and pre-processed to filter out possible aggregates and liposome debris using Cytobank (https://community.cytobank.org/), as previously described in [44].

DNA templates, substrates and nucleotides **(Fig. S12)**

Unlabelled nucleotides were purchased from GE Healthcare. The [γ-^{32}P]ATP (3,000 Ci/mmol) and [α-^{32}P]dATP (3,000 Ci/mmol) were supplied by PerkinElmer. Oligonucleotides sp1 (5'-GATCACAGTGAGTAC), sp1c+6 (5'-TCTATTGTACTCACTGTGATC), and M13 Universal Primer (5'-GTAAAACGACGGCCAGT) were purchased from Sigma-Aldrich. T4 polynucleotide kinase (T4PNK) was purchased from New England Biolabs. Oligonucleotide sp1 was 5'-labelled with ^{32}P using [γ-^{32}P]ATP (10 µCi) and T4PNK and further hybridized to oligonucleotide sp1c+6 (1:2 ratio) to get the primer/template substrate sp1/sp1c+6 for the Exonuclease/Polymerization balance assays (see below). Oligonucleotides were annealed in the presence of 50 mM Tris- HCl (pH 7.5) and 0.2 M NaCl, heating to 90 °C for 10 min before slowly cooling to room temperature overnight. M13mp18 (+) strand ssDNA (Sigma-Aldrich) was hybridized to the universal primer as described above, and the resulting molecule was used as a primer/template complex to analyse processive DNA polymerization coupled to strand displacement by the wildtype and variants of Phi29 DNAP. Terminal protein-Phi29 DNAP complex (TP-DNA) was obtained as described in [77].

Primed M13 DNA replication assay **(Fig. S13)**

The incubation mixture contained, in 25 µL, 50 mM Tris–HCl (pH 7.5) 10 mM $MgCl_2$, 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/mL of BSA, 40 mM dNTPs and [α-^{32}P]dATP (1 µCi), 4.2 nM of primed M13mp18 ssDNA, and 60 nM of either the wildtype or the indicated mutant Phi29 DNA polymerase. After incubation at 30 °C for the indicated times, the reactions were stopped by adding 10 mM EDTA-0.1% SDS and the samples were filtered through Sephadex G-50 spin columns. For size analyses of the synthesized DNA, the labelled DNA was denatured by treatment with 0.7 M NaOH and subjected to electrophoresis in alkaline 0.7% agarose gels, as described in [78]. After electrophoresis the gels were dried and autoradiographed.

TP-DNA amplification assay **(Fig S12)**

The assay was performed essentially as described in [79]. The reaction mixture contained in a final volume of 25 µL, 50 mM Tris–HCl (pH 7.5) 10 mM $MgCl_2$, 20 mM ammonium sulphate, 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/mL BSA, 80 mM of each dNTP and [alpha-^{32}P]dATP (1 µCi), 15 pM of TP-DNA, 3 nM of either wildtype or the indicated DNA polymerase variant, 6 nM of TP, 30 µM of SSB and 30 µM of DSB. After incubation for the indicated times at 30 °C, samples were processed as described for the TP-DNA replication assay and subjected to electrophoresis in alkaline 0.7% agarose gels, as described [78]. After electrophoresis, the gels were dried and autoradiographed.

Exonuclease/polymerase balance assay **(Fig. S12)**

In a final volume of 12.5 µL, the incubation mixture contained 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, 4% (v/v) glycerol, 0.1 mg/mL BSA, 1 nM 5'-labelled sp1/sp1c+6 substrate (a primer/template structure that contains a 6-nt 5'-protruding end, and therefore

can be used as substrate for DNA-dependent DNA polymerization and also for the exonuclease activity), 30 nM wildtype or mutant Phi29 DNA polymerase, and the indicated increasing concentrations of the four dNTPs (0-150 nM). After incubation for 5 min at 25 °C, the reaction was stopped by adding EDTA up to a final concentration of 10 mM. Reaction products were resolved by electrophoresis in 7 M urea-20% polyacrylamide gels and autoradiography. Polymerization or 3'-5' exonucleolysis was detected as an increase or decrease, respectively, in the size (15-mer) of the 5'-labelled primer.

Electrophoretic mobility shift assay (EMSA) **(Fig. S12)**

The incubation mixture contained, in a final volume of 20 µL, 50 mM Tris-HCl (pH 7.5), 20 mM ammonium sulphate, 0.1 mg/mL BSA, 0.7 nM 5'-labelled sp1/sp1c+6 primer/template hybrid, and the indicated amount of wildtype or mutant DNA polymerase. After incubation for 5 min at 4 °C, the samples were subjected to electrophoresis in precooled 4% (w/v) polyacrylamide gels [80:1 acrylamide/bis-acrylamide (w/w)] containing 12 mM Tris acetate (pH 7.5) and 1 mM EDTA, and run at 4 °C in the same buffer at 8 V/cm [80]. After autoradiography, a stable interaction between the enzyme and the DNA was detected as a shift (retardation) in the migrating position of the labelled DNA.

Acknoledgements

We would like to thank Ilja Westerlaken for single clone preparation and Sanger sequencing analysis and Elisa Godino for assisting with characterization of some DNA variants and flow cytometry experiments . The research was funded by the NWO Gravitation programs "Building a synthetic cell" (BaSyC) and NanoFront. CD acknowledges funding from ANR pour CPJ (ANR-2023-004). In memoriam of Andreea R. Stan, who sadly passed away and did not see the final version of the paper.

Supplementary Information

Evolution campaign (at least 5% frequency)	AA position	Codon in parental *mod-orip-2p3*	Codon in original Phi29 genome	DNA mutation	Comments
Int-WT1	I67	ATT (0.49)	ATC (0.39)	ATC (0.39)	Recovered Phi29 original DNA sequence
	V247	GTG (0.35)	GTT (0.28)	GTA (0.17)	
	N248	AAT (0.49)	AAT (0.49)	AAC (0.51)	Changed from original Phi29 sequence
Int-WT2	L477	CTG (0.35)	TTG (0.13)	CTT (0.12)	
Int-Mut	I67	ATT (0.49)	ATC (0.39)	ATC (0.39)	Recovered Phi29 original DNA sequence
	K121	AAA (0.74)	AAG (0.26)	AAG (0.26)	Followed by string of A's
	K475	AAA (0.74)	AAG (0.26)	AAG (0.26)	Followed by string of A's
Con-WT (4 h)	E158	GAA (0.68)	GAA (0.68)	GAG (0.32)	Changed from original Phi29 sequence

Table. S1 Codon bias analysis on at least *5%* synonymous mutations from all evolutionary campaigns. Numeric fractions for codon usage estimation on each codon were taken from GenScript Codon Usage Frequency Table (https://www.genscript.com/tools/codon-frequency-table)

Primer pair	Sequence (5' → 3')	Purpose
1058 ChD	GTCGACTGCTAATACGACTCACTATAGGGCCCTCTGGAGACACCAGAGGG	Cloning G340 plasmid containing *mod-ori-p2p3*. PCR fragment containing modified T7 leader sequence.
1060 ChD	CGGGCTGCGTGCCATTAGTATATCTCCTTCTTAAAGTTAAACAAATAAACATGT	
1049 ChD	GCAAGGCGATTAAGTTGGGTAACG	Cloning G340 plasmid containing *mod-ori-p2p3*. PCR fragment containing *p2* gene
1056 ChD	TGTCTCCAGAGGGCCCTATAGTGAGTCGTATTAGCAGTCGAC	
1057 ChD	ACATGTTTATTTGTTTAACTTTAAGAAGGAGATATACTAATGGCACGCAGCCCG	Cloning G340 plasmid containing *mod-ori-p2p3*. PCR fragment containing *p3* gene
1052 ChD	TGTGTGGAATTGTGAGCGGATAAC	
948 ChD	TGTAAAACGACGGCCAGT	Cloning G371 plasmid containing *mod-ori-p2p3* encoding for Φ29 DNAP(F62Y). Fragment 1 for Gibson assembly
1132 ChD	GTTGATAATGAACGCACCATCGTATTTCAGATTGTGGAAGTAC	
1131 ChD	GTACTTCCACAATCTGAAATACGATGGTGCGTTCATTATCAAC	Cloning G371 plasmid containing *mod-ori-p2p3* encoding for Φ29 DNAP(F62Y). Fragment 2 for Gibson assembly
1137 ChD	GGCGGTCATGCGATCCAG	
1324 ChD	ATGGGGCGCCGATGGTCTGCCGAACACC	Cloning G559 plasmid containing *mod-ori-p2(S79G)p3*.
1325 ChD	TCGGCGCCCCATTTAAAGCCATTACGTTCCAG	
1326 ChD	CTGAAGAAACTGCCGTTTCCGGTGAAG	Cloning G560 plasmid containing *mod-ori-p2(K121K)p3*.
1327 ChD	GTTTCTTCAGGCTGTCATAGATCACGGTATG	
1328 ChD	GAAGAAACTGGGTTATTGGGCACACGAATC	Cloning G561 plasmid containing *mod-ori-p2(K475K)p3*.
1329 ChD	GTTTCTTCGGGTCAACGATATCTTTAATCACATCC	
1330 ChD	ATCAAGAGCGTCGAAGGCTCATTTAACTCGTTC	Cloning G562 plasmid containing *mod-ori-p2p3(K188K)*.
1331 ChD	ACGCTCTTGATAAAATTCAGTTGCAGCTGAATC	
1344 ChD	ATGGGGCACCGATGGTCTGCCGAACACC	Cloning G569 plasmid containing *mod-ori-p2(S79G&A80T)p3*
1345 ChD	ATCGGTGCCCCATTTAAAGCCATTACGTTCC	
1346 ChD	TGGAGCACCGATGGTCTGCCGAACACC	Cloning G570 plasmid containing *mod-ori-p2(A80T)p3*
1347 ChD	CATCGGTGCTCCATTTAAAGCCATTACGTTCC	
491 ChD	P-AAAGTAAGCCCCCACCCTCACATG	PCR to produce *ori-p2p3*, *mod-ori-p2p3*, and all *mod-ori-p2p3* reversed engineered versions.
492 ChD	P-AAAGTAGGGTACAGCGACAACATACAC	
976 ChD	GGATGAAGACTACCCGCTGC	qPCR amplicon targeting *p2* gene.
977 ChD	ACAGGTCTGCGATTTCACCG	

Table. S2 Primer pairs used for PCR and qPCR

Protein	Compound name	Precursor ion *(m/z)*	Product ion *(m/z)*	Ion name
DNAP	ENGALGFR	432,2221	492,2929	y4
DNAP	ENGALGFR	432,2221	563,3300	y5
DNAP	ENGALGFR	432,2221	620,3515	y6
DNAP	ENGALGFR	432,2221	734,3944	y7
TP	IAEIER	365,7083	314,1710	b3
TP	IAEIER	365,7083	304,1615	y2
TP	IAEIER	365,7083	417,2456	y3
TP	IAEIER	365,7083	546,2882	y4
TP	IAEIER	365,7083	617,3253	y5

Table. S3 Transitions of the MS/MS measurements for the proteolytic peptides of DNAP

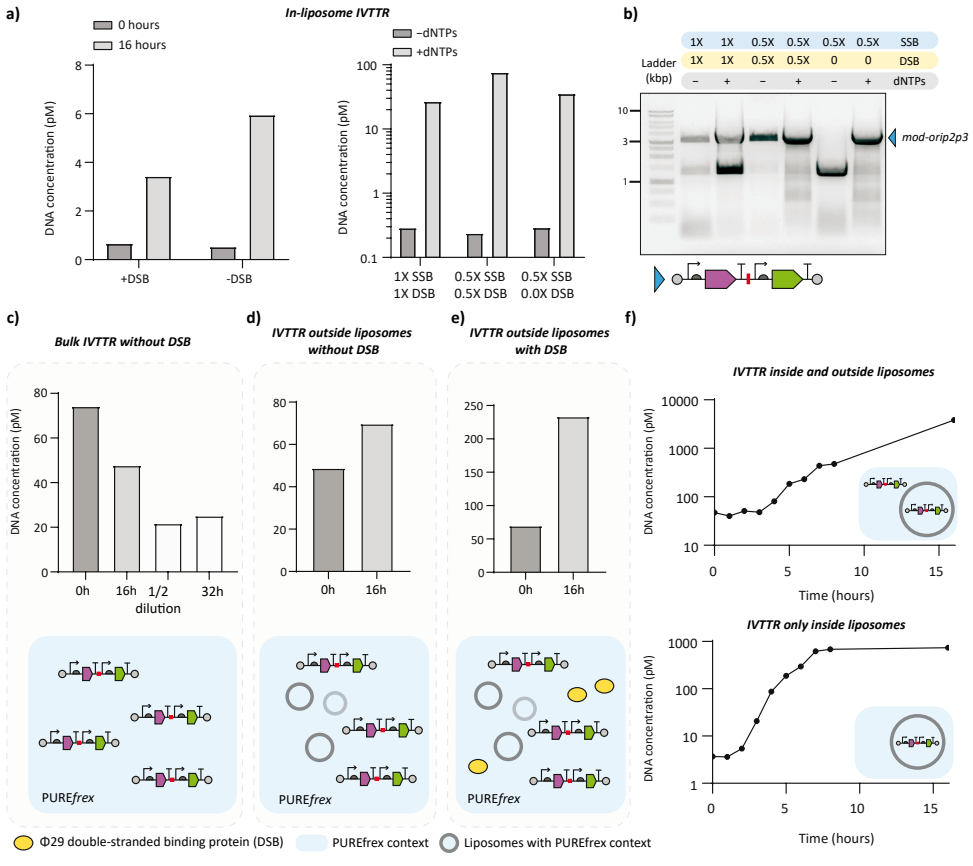


Fig. S1 In-liposome and bulk IVTTR dependence on DNA replication auxiliary proteins DSB and SSB. a) Absolute DNA quantification from in-liposome IVTTR reactions with different amounts of DSB and SSB. Amplification by qPCR targeted the *p2* gene of the *mod-ori-p2p3* DNA construct (same in all panels). **b)** Agarose gel electrophoresis with PCR-recovered DNA from panel a. **c)** Absolute DNA quantification from bulk IVTTR reactions without DSB protein. Samples were taken at different time points, with ½ dilution with fresh PURE after 16 hours incubation, and another 16 hours of incubation **e)** Absolute DNA quantification from IVTTR reactions without DSB and with *mod-ori-p2p3* added outside

preformed liposomes. **f)** Absolute DNA quantification from IVTTR reactions containing DSB. *mod-ori-p2p3* was present both inside and outside of liposomes (top panel), or only inside (bottom panel).

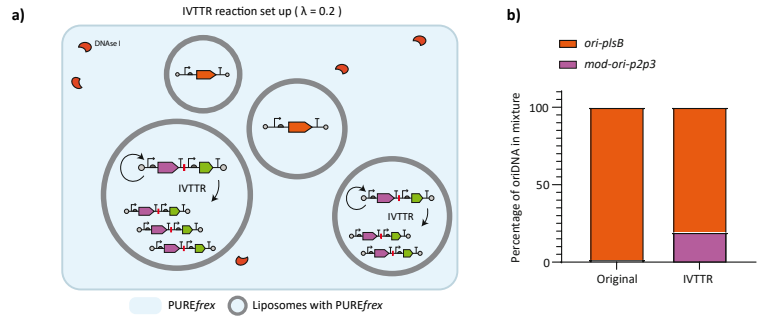


Fig. S2 a) Scheme of *mod-ori-p2p3* enrichment from a mixture with non-self-amplifying DNA (*ori-plsB*) inside gene-expressing liposomes. **b)** Calculated percentages of each DNA species in the original DNA mixture and after IVTTR.

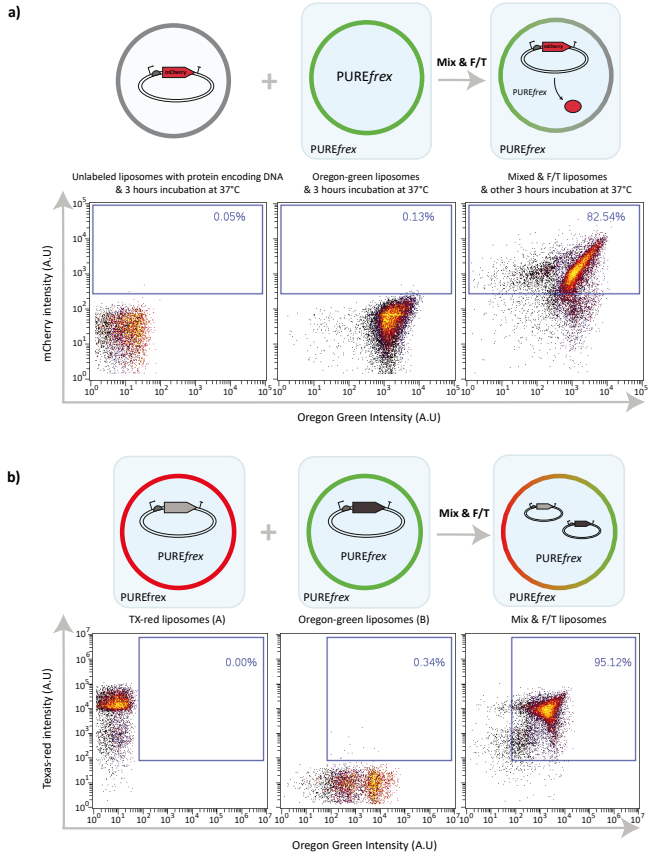


Fig. S3 Liposome content mixing and membrane fusion are promoted by FT. a) Flow cytometry scatter plots from unlabelled liposomes encapsulating the mCherry gene (left panel), Oregon-green labelled liposomes with encapsulated PURE system (middle panel), and mixed samples exposed to FT and incubation for gene expression (right panel). **b)** Flow cytometry scatter plots from PURE containing liposomes stained either with Texas Red (left panel) or Oregon-green

(middle panel) membrane dye. After liposome mixture and FT, a new population of liposomes exhibiting both membrane fluorophores appeared as a result of lipid mixing (right panel).

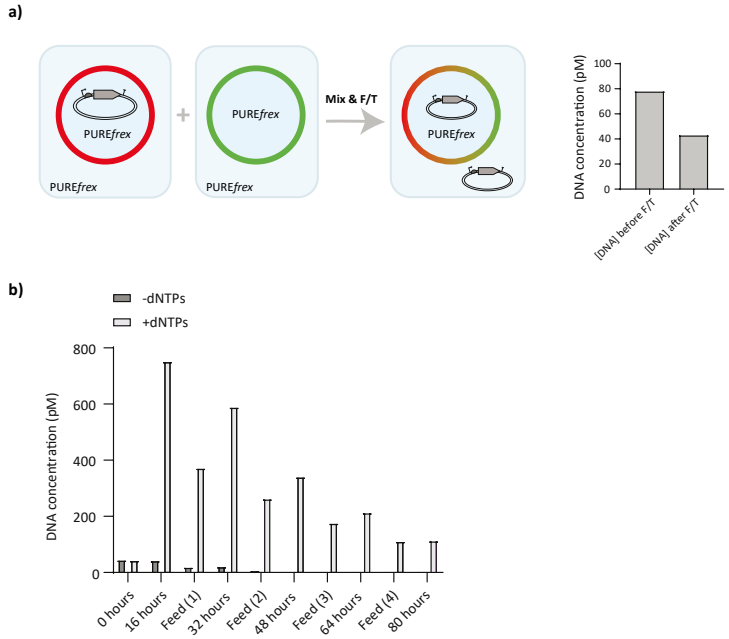


Fig. S4 Estimation of DNA leakage caused by F/T. a) DNA loss after FT was estimated to be ~50%. DNA concentrations measured by qPCR after mixing and FT of two PURE containing liposome populations. One population co-encapsulated DNA, not the other. DNA outside of liposomes was digested by DNase I and the amount of DNA was assessed by qPCR. **b)** Omitting DSB causes a gradual decrease in DNA concentration after a few serial dilution rounds. Absolute DNA quantification from repeated in-liposome IVTTR reactions with *mod-ori-p2p3* as DNA template (continuous evolution method) and no DSB protein (also no DNAse I).

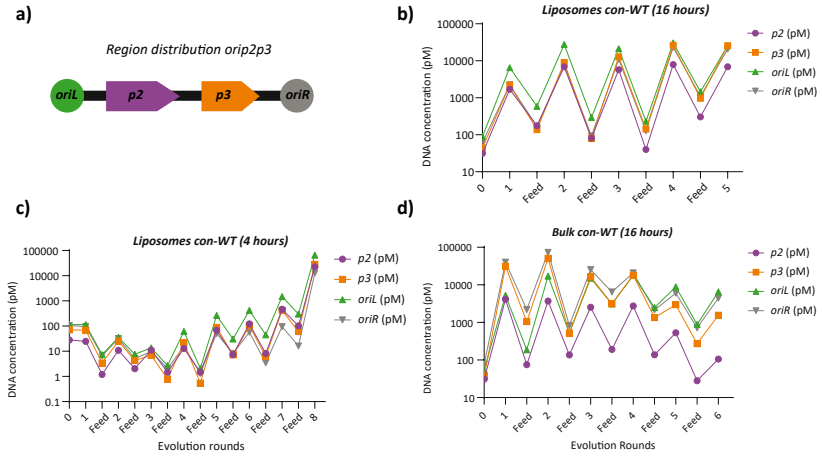


Fig. S5 a) Schematic illustration of *mod-ori-p2p3* self-replicator regions that were targeted by qPCR. **b-d)** Absolute DNA quantification of the different targeted regions from samples in Con-WT(16 h) **(b)**, Con-WT(4 h) **(c)**, and Bulk-WT **(d)**. Color coding is the same as in panel a.

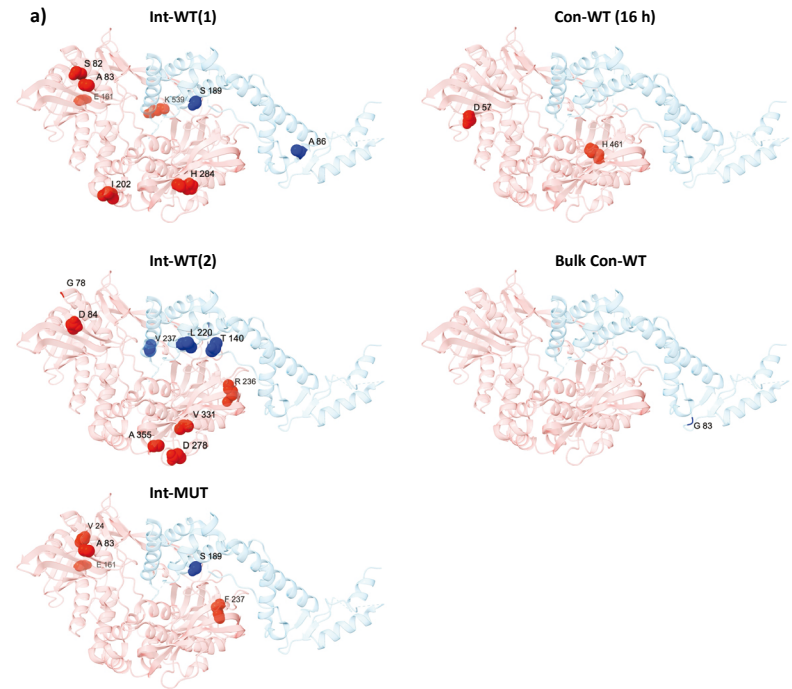


Fig. S6 DNAP and TP protein residues on which mutations were detected at frequencies of at least 5% on all evolutionary campaigns (Int-WT(1), Int-WT(2), Con-WT (16h), and Bulk-WT (16 h)). Panel **a)** contains the residue locations on DNAP-TP protein complex (PDB 2EX3). Panel **b)** contains the residue locations on DNAP primer-template complex (PDB 2PZS). DNAP is colored in pink. TP in sky blue. DNA strand in green. Primer in yellow. DNAP amino acid residues in red, and TP amino acid residues in blue. The figures were generated using The Open-Source Molecular Graphics System, v. 2.5.0, Schrödinger, LLC (Open-Source PyMOL is Copyright (C) Schrodinger, LLC.)

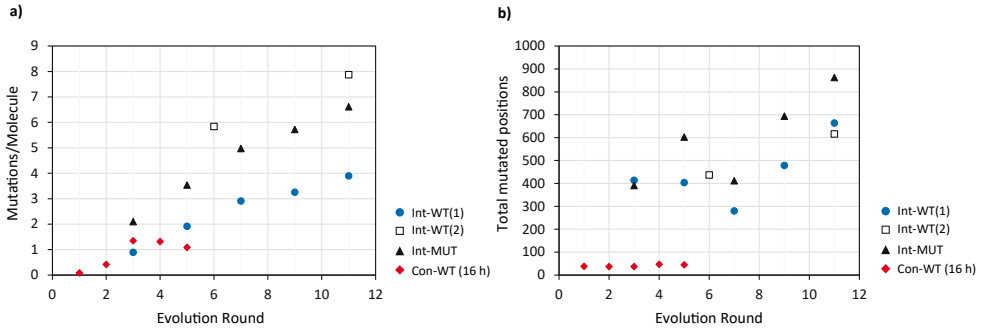


Fig. S7 a) Accumulation of mutations per round of evolution as mutations/molecule, and **b)** total mutated positions for all in-liposome evolutionary campaigns: intermittent evolution (Int-WT(1), Int-WT(2), Int-MUT), and continuous evolution (Con-WT).

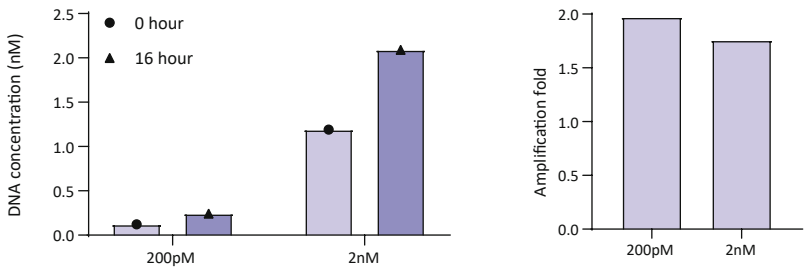
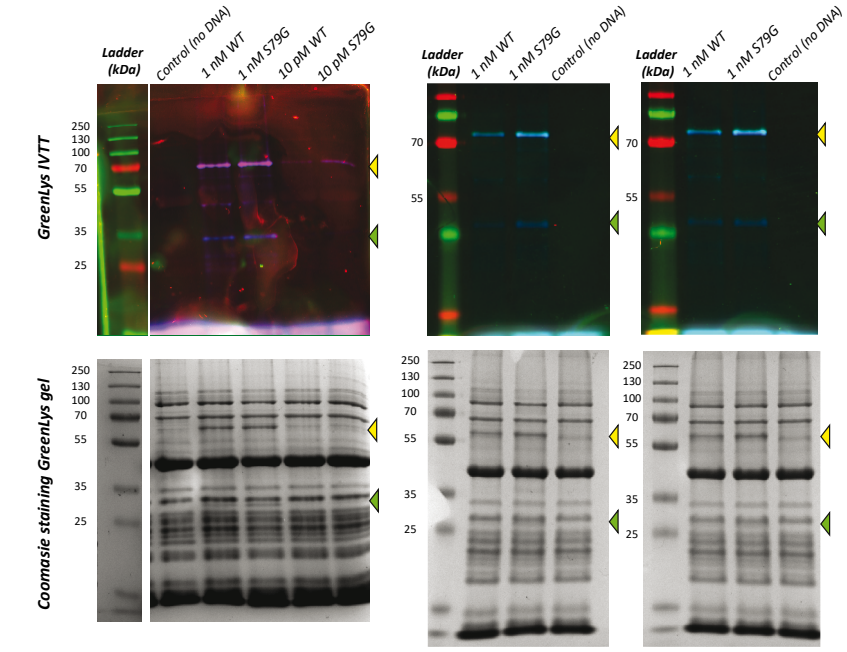


Fig. S8 Absolute DNA quantification on bulk IVTTR experiments using PCR-recovered DNA from round 11 of Int-WT(1) as DNA template. Two different concentrations were tested: 200 pM and 2 nM. A region in the *p2* gene from *mod-ori-p2p3* DNA was targeted for qPCR



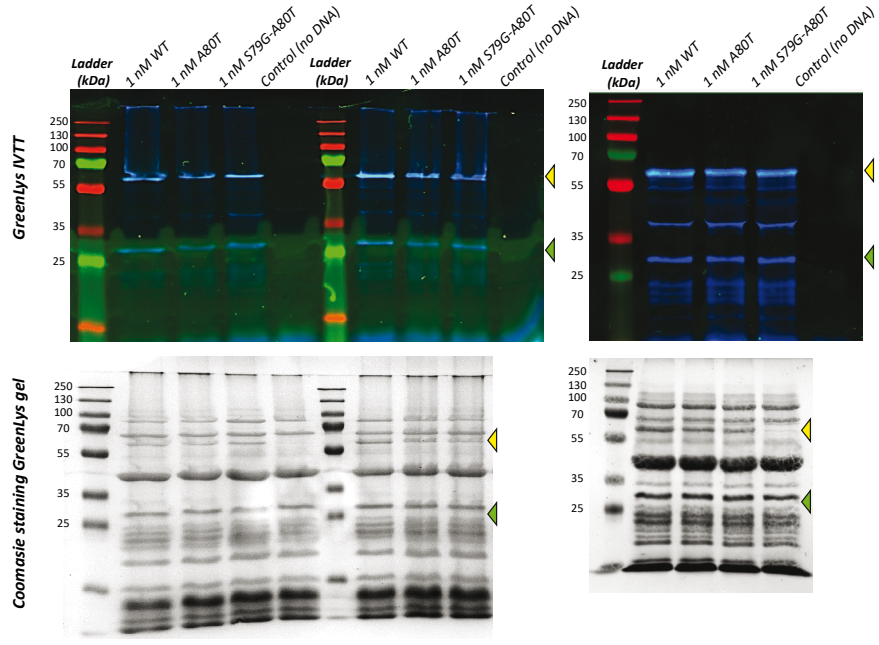


Fig. S9 Biological repeats of bulk IVTT protein production with parental *mod-ori-p2p3* and reversed engineered *mod-ori-p2p3* variants (DNAP(S79G) and DNAP(A80T)) as DNA templates. SDS-PAGE gels with Green-Lys protein labelling (upper gel), Coomassie rotein staining (bottom gel).

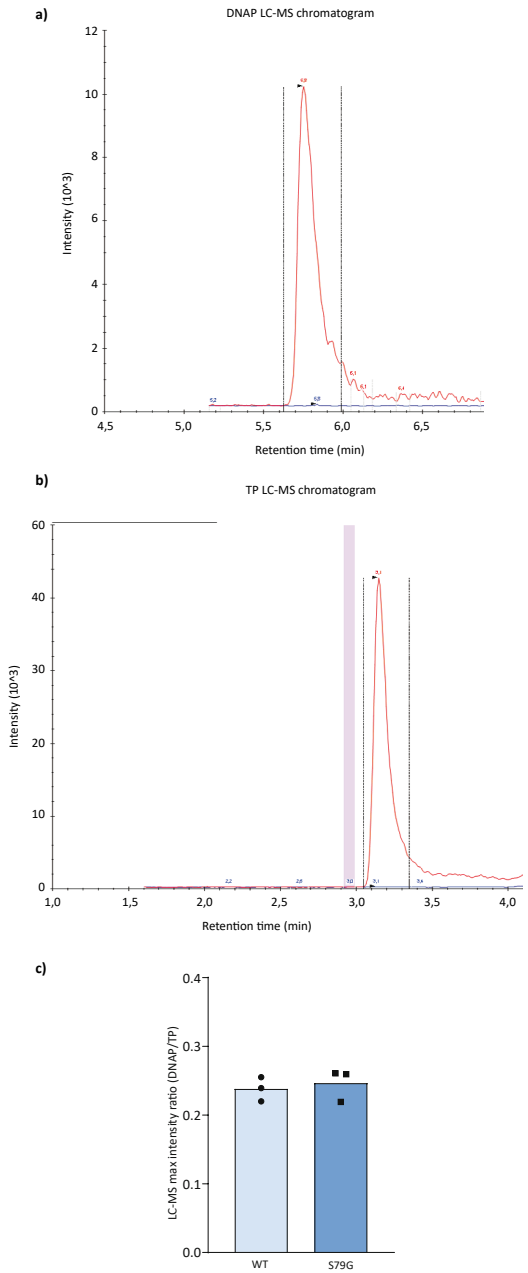


Fig. S10 LC-MS relative quantification of WT DNAP and S79G DNAP variant. Per sample, the maximum intensities recorded for DNAP, either WT or S79G, were normalized with the maximum recorded intensity of produced TP.

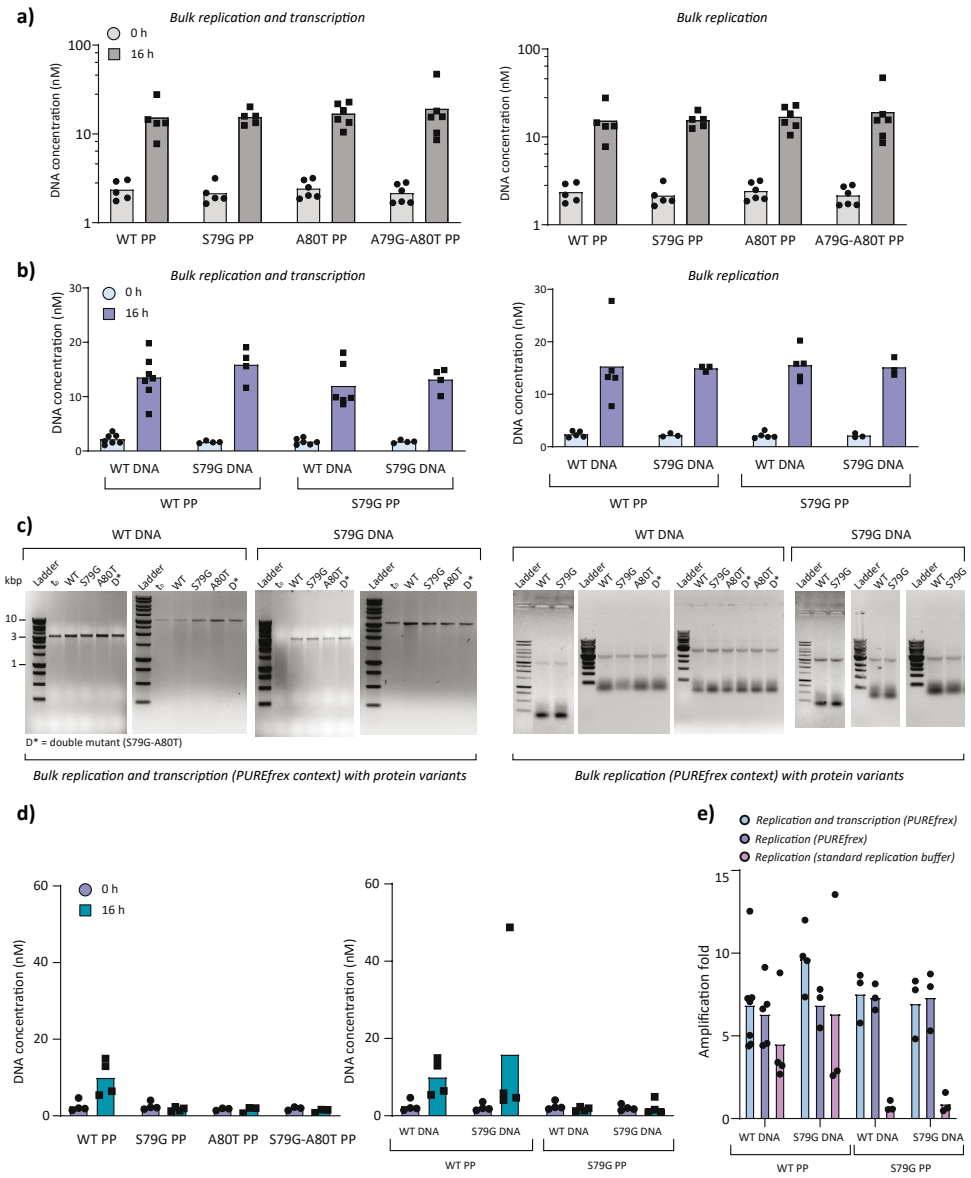


Fig. S11 Bulk replication assays of *mod-ori-p2p3* templates with purified DNAP protein variants: S79G, A80T, and double DNAP mutant. **a)** Absolute DNA quantification of bulk replication-transcription (left panel) or replication (right panel) samples with different DNAP variants, as indicated, in a PURE*frex* background. **b)** Absolute DNA quantification on bulk replication-transcription (left panel) or replication (right panel) samples with different DNA templates: *mod-ori-p2p3* or *mod-ori-p2(S79G)p3*, and different purified DNAP variants in a PURE*frex* background. **c)** Agarose gel electrophoresis of *mod-ori-p2p3* and *mod-ori-p2(S79G)p3* recovered DNA from the reaction samples shown in panel (a) and (b). **d)** Absolute DNA quantification on bulk replication reactions with different DNA templates: *mod-ori-p2p3* or *mod-ori-p2(S79G)p3*, and different purified DNAP variants. Reactions were performed in a standard Phi29 DNA replication buffer. **e)** Amplification fold (DNA concentration at 16 h / DNA concentration at 0 h) values for the different conditions tested. Individual data points per condition correspond to biological repeats.

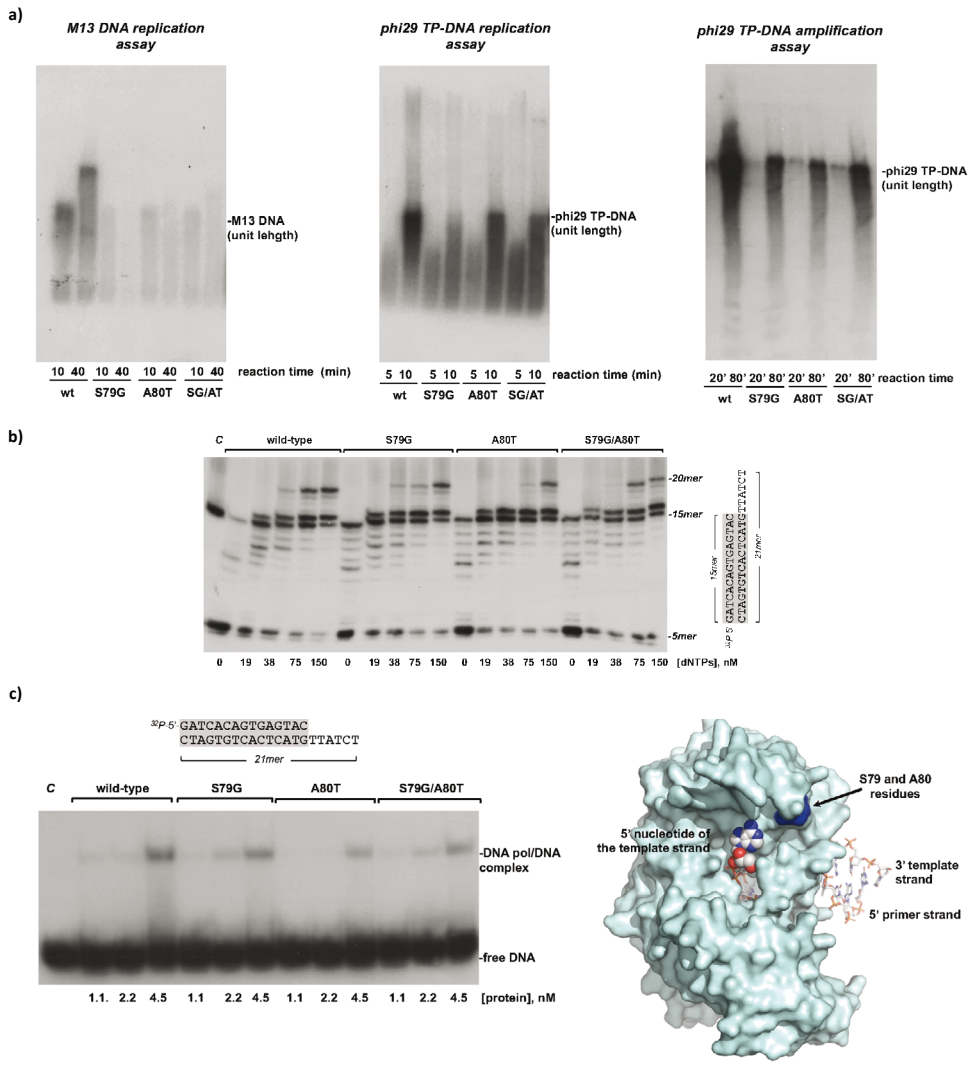


Fig. S12 a) M13 DNA replication assay (left panel). Replication of primed M13 DNA was carried out as described in Materials and Methods in the presence of 500 ng of M13 template, 10 mM $MgCl_2$, and 60 nM of either wild-type or mutant (S79G) DNA polymerase. The expected position of full-length M13 DNA is shown at the right side of the gel. TP-DNA replication assay (middle panel). The assay was performed essentially as described in Materials and Methods using 500 ng of TP-DNA, 10 mM $MgCl_2$,12 nM of TP and 12 nM of either wild-type or mutant (S79G) DNA polymerase. The migration position of unit-length of TP-DNA is indicated on the right side of the gel. DNA amplification assay (right panel). The assay was carried out as described in Material and Methods, in the presence of 5 ng of TP-DNA, 10 mM $MgCl_2$, 3 nM of DNA polymerase wild-type or the indicated mutant, 6 nM of TP, 30 µM of SSB and 30 µM of DBP. The migration position of unit-length of TP-DNA is indicated on the right side of the gel. **b)** Exonuclease/polymerization balance assay. The reaction was performed as described in Materials and Methods using 1 nM of dsDNA substrate (5' labelled molecule sp1/sp1c+6, depicted at the right of the figure), 10 mM $MgCl_2$, 30 nM of either wild-type or mutant DNA polymerase, the indicated concentration of dNTPs, and 5 min incubation at 25 °C. Polymerization or 3'-5' exonuclease activity is detected as an increase or decrease in the size of the labeled primer (15 mer). C: control lane without enzyme. **c)** On the left panel, interaction of wild-type and mutant DNA polymerases with a primer/template substrate (0.7 nM). The EMSA assay was performed as described in Materials and Methods with 0.7 nM of sp1/sp1c+6 and 5 min at 4 °C. The 5' labelled molecule sp1/sp1c+6, 15mer/21mer (depicted at the top of the figure) was incubated either with the wild-type or mutant enzymes. The bands corresponding to free DNA or to the DNA/DNAP complex were detected by autoradiography. C: control lane without enzyme. On the right panel surface representation of Phi29 DNA polymerase complexed with primer/template DNA. Crystallographic data are from Protein Data Bank ID code 2PZS [81]. DNA polymerase residues S79 and A80 are shown as spheres and the primer and template strands as sticks. Figure was generated using The Open-Source Molecular Graphics System, v. 2.5.0, Schrödinger, LLC (Open-Source PyMOL is Copyright (C) Schrodinger, LLC.).

References

1. Buddingh', B. C. & van Hest, J. C. M. Artificial Cells: Synthetic Compartments with Life-like Functionality and Adaptivity. Acc Chem Res 50, 769–777 (2017).
2. Xu, C., Hu, S. & Chen, X. Artificial cells: from basic science to applications. Mater Today (Kidlington) 19, 516–532 (2016).
3. Glass, J. I., Merryman, C., Wise, K. S., Hutchison, C. A. & Smith, H. O. Minimal Cells—Real and Imagined. Cold Spring Harb Perspect Biol 9, a023861 (2017).
4. Yewdall, N. A., Mason, A. F. & van Hest, J. C. M. The hallmarks of living systems: towards creating artificial cells. Interface Focus 8, 20180023 (2018).
5. Abil, Z. & Danelon, C. Roadmap to Building a Cell: An Evolutionary Approach. Frontiers in Bioengineering and Biotechnology 8, (2020).
6. Adamski, P. et al. From self-replication to replicator systems en route to de novo life. Nat Rev Chem 4, 386–403 (2020).
7. van Nies, P. et al. Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. Nat Commun 9, 1–12 (2018).
8. Godino, E. et al. De novo synthesized Min proteins drive oscillatory liposome deformation and regulate FtsA-FtsZ cytoskeletal patterns. Nat Commun 10, 4969 (2019).
9. Blanken, D., Foschepoth, D., Serrão, A. C. & Danelon, C. Genetically controlled membrane synthesis in liposomes. Nat Commun 11, 4317 (2020).
10. Godino, E. et al. Cell-free biogenesis of bacterial division proto-rings that can constrict liposomes. Commun Biol 3, 1–11 (2020).
11. Doerr, A., Foschepoth, D., Forster, A. C. & Danelon, C. In vitro synthesis of 32 translation-factor proteins from a single template reveals impaired ribosomal processivity. Sci Rep 11, 1898 (2021).
12. Kattan, J., Doerr, A., Dogterom, M. & Danelon, C. Shaping Liposomes by Cell-Free Expressed Bacterial Microtubules. ACS Synth. Biol. 10, 2447–2455 (2021).
13. Godino, E., Restrepo Sierra, A. M. & Danelon, C. Imaging Flow Cytometry for High-Throughput Phenotyping of Synthetic Cells. ACS Synth. Biol. 12, 2015–2028 (2023).
14. Daniel, I., Oger, P. & Winter, R. Origins of life and biochemistry under high-pressure conditions. Chem. Soc. Rev. 35, 858–875 (2006).
15. Lenton, T. M. Testing Gaia: The Effect of Life on Earth's Habitability and Regulation. Climatic Change 52, 409–422 (2002).
16. Annila, A. & Annila, E. Why did life emerge? International Journal of Astrobiology 7, 293–300 (2008).
17. Duim, H. & Otto, S. Towards open-ended evolution in self-replicating molecular systems. Beilstein J. Org. Chem. 13, 1189–1203 (2017).
18. Le Vay, K., Weise, L. I., Libicher, K., Mascarenhas, J. & Mutschler, H. Templated Self-Replication in Biomimetic Systems. Advanced Biosystems 3, 1800313 (2019).
19. von Kiedrowski, G., Wlotzka, B., Helbing, J., Matzen, M. & Jordan, S. Parabolic Growth of a Self-Replicating Hexadeoxynucleotide Bearing a 3'-5'-Phosphoamidate Linkage. Angewandte Chemie International Edition in English 30, 423–426 (1991).

20. von Kiedrowski, G. Minimal Replicator Theory I: Parabolic Versus Exponential Growth. in Bioorganic Chemistry Frontiers (eds. Dugas, H. & Schmidtchen, F. P.) 113–146 (Springer, 1993). doi:10.1007/978-3-642-78110-0_4.
21. Lincoln, T. A. & Joyce, G. F. Self-Sustained Replication of an RNA Enzyme. Science 323, 1229–1232 (2009).
22. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Ghadiri, M. R. A self-replicating peptide. Nature 382, 525–528 (1996).
23. Issac, R., Ham, Y. W. & Chmielewski, J. The design of self-replicating helical peptides. Curr Opin Struct Biol 11, 458–463 (2001).
24. Issac, R. & Chmielewski, J. Approaching Exponential Growth with a Self-Replicating Peptide. J. Am. Chem. Soc. 124, 6808–6809 (2002).
25. Mukherjee, R., Cohen-Luria, R., Wagner, N. & Ashkenasy, G. A bistable switch in dynamic thiodepsipeptide folding and template-directed ligation. Angew Chem Int Ed Engl 54, 12452–12456 (2015).
26. Zepik, H. H., Blöchliger, E. & Luisi, P. L. A Chemical Model of Homeostasis. Angew Chem Int Ed Engl 40, 199–202 (2001).
27. Bachmann, P. A., Walde, P., Luisi, P. L. & Lang, J. Self-replicating reverse micelles and chemical autopoiesis. ACS Publications https://pubs.acs.org/doi/pdf/10.1021/ja00178a073 (2002) doi:10.1021/ja00178a073.
28. Colomer, I., Morrow, S. M. & Fletcher, S. P. A transient self-assembling self-replicator. Nat Commun 9, 2239 (2018).
29. Morrow, S. M., Colomer, I. & Fletcher, S. P. A chemically fuelled self-replicator. Nat Commun 10, 1011 (2019).
30. Carnall, J. M. A. et al. Mechanosensitive Self-Replication Driven by Self-Organization. Science 327, 1502–1506 (2010).
31. Vaidya, N. et al. Spontaneous network formation among cooperative RNA replicators. Nature 491, 72–77 (2012).
32. Crutchfield, J. P. & Schuster, P. Evolutionary Dynamics: Exploring the Interplay of Selection, Accident, Neutrality, and Function. (Oxford University Press, 2003).
33. Ichihashi, N. et al. Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment. Nat Commun 4, 2494 (2013).
34. Bansho, Y. et al. Importance of parasite RNA species repression for prolonged translation-coupled RNA self-replication. Chem Biol 19, 478–487 (2012).
35. Mizuuchi, R., Ichihashi, N., Usui, K., Kazuta, Y. & Yomo, T. Adaptive Evolution of an Artificial RNA Genome to a Reduced Ribosome Environment. ACS Synth. Biol. 4, 292–298 (2015).
36. Furubayashi, T. et al. Emergence and diversification of a host-parasite RNA ecosystem through Darwinian evolution. eLife 9, e56038 (2020).
37. Mizuuchi, R. & Ichihashi, N. Sustainable replication and coevolution of cooperative RNAs in an artificial cell-like system. Nat Ecol Evol 2, 1654–1660 (2018).
38. Yoshiyama, T., Ichii, T., Yomo, T. & Ichihashi, N. Automated in vitro evolution of a translation-coupled RNA replication system in a droplet flow reactor. Sci Rep 8, 11867 (2018).
39. Forster, A. C. & Church, G. M. Towards synthesis of a minimal cell. Molecular Systems Biology 2, 45 (2006).
40. Okauchi, H. & Ichihashi, N. Continuous Cell-Free Replication and Evolution of Artificial Genomic DNA in a Compartmentalized Gene Expression System. ACS Synth. Biol. 10, 3507–3517 (2021).
41. Sakatani, Y., Yomo, T. & Ichihashi, N. Self-replication of circular DNA by a self-encoded DNA polymerase through rolling-circle replication and recombination. Sci Rep 8, 13089 (2018).
42. Sakatani, Y., Mizuuchi, R. & Ichihashi, N. In vitro evolution of phi29 DNA polymerases through compartmentalized gene expression and rolling-circle replication. Protein Eng Des Sel 32, 481–487 (2019).
43. Okauchi, H., Sakatani, Y., Otsuka, K. & Ichihashi, N. Minimization of Elements for Isothermal DNA Replication by an Evolutionary Approach. ACS Synth. Biol. 9, 1771–1780 (2020).
44. Abil, Z., Restrepo Sierra, A. M. & Danelon, C. Clonal Amplification-Enhanced Gene Expression in Synthetic Vesicles. ACS Synth. Biol. 12, 1187–1203 (2023).
45. Mencía, M., Gella, P., Camacho, A., de Vega, M. & Salas, M. Terminal protein-primed amplification of heterologous DNA with a minimal replication system based on phage ϕ29. Proceedings of the National Academy of Sciences 108, 18655–18660 (2011).

46. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA Websuite. Nucleic Acids Res 36, W70–W74 (2008).
47. Kobs, G. FluoroTectTM GreenLys in vitro Translation Labeling System.
48. de Vega, M., Lázaro, J. M., Salas, M. & Blanco, L. Mutational analysis of phi29 DNA polymerase residues acting as ssDNA ligands for 3'-5' exonucleolysis. J Mol Biol 279, 807–822 (1998).
49. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. Proceedings of the National Academy of Sciences 98, 4552–4557 (2001).
50. Esteban, J. A., Salas, M. & Blanco, L. Fidelity of phi 29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. J Biol Chem 268, 2719–2726 (1993).
51. Paez, J. G. et al. Genome coverage and sequence fidelity of phi29 polymerase-based multiple strand displacement whole genome amplification. Nucleic Acids Res 32, e71 (2004).
52. Soengas, M. S. et al. Site-directed mutagenesis at the Exo III motif of phi 29 DNA polymerase; overlapping structural domains for the 3'-5' exonuclease and strand-displacement activities. EMBO J 11, 4227–4237 (1992).
53. Potapov, V. & Ong, J. L. Examining Sources of Error in PCR by Single-Molecule Sequencing. PLOS ONE 12, e0169774 (2017).
54. McInerney, P., Adams, P. & Hadi, M. Z. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. Mol Biol Int 2014, 287430 (2014).
55. Tsuji, G., Fujii, S., Sunami, T. & Yomo, T. Sustainable proliferation of liposomes compatible with inner RNA replication. PNAS 113, 590–595 (2016).
56. Bansho, Y., Furubayashi, T., Ichihashi, N. & Yomo, T. Host–parasite oscillation dynamics and evolution in a compartmentalized RNA replication system. Proceedings of the National Academy of Sciences 113, 4045–4050 (2016).
57. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. Nature 472, 499–503 (2011).
58. Eigen, M. & Schuster, P. A principle of natural self-organization. Naturwissenschaften 64, 541–565 (1977).
59. Matsuura, T. et al. Importance of compartment formation for a self-encoding system. Proceedings of the National Academy of Sciences 99, 7514–7517 (2002).
60. Streisinger, G. et al. Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. Cold Spring Harb Symp Quant Biol 31, 77–84 (1966).
61. Koutmou, K. S. et al. Ribosomes slide on lysine-encoding homopolymeric A stretches. eLife 4, e05534 (2015).
62. Jeng, S. T., Gardner, J. F. & Gumport, R. I. Transcription termination in vitro by bacteriophage T7 RNA polymerase. The role of sequence elements within and surrounding a rho-independent transcription terminator. J Biol Chem 267, 19306–19312 (1992).
63. Nieuwkoop, T., Finger-Bou, M., Van Der Oost, J. & Claassens, N. J. The Ongoing Quest to Crack the Genetic Code for Protein Production. Molecular Cell 80, 193–209 (2020).
64. de Vega, M., Lazaro, J. M., Salas, M. & Blanco, L. Primer-terminus stabilization at the 3'-5' exonuclease active site of phi29 DNA polymerase. Involvement of two amino acid residues highly conserved in proofreading DNA polymerases. The EMBO Journal 15, 1182–1192 (1996).
65. de Vega, M., Blanco, L. & Salas, M. phi29 DNA polymerase residue Ser122, a single-stranded DNA ligand for 3'-5' exonucleolysis, is required to interact with the terminal protein. J Biol Chem 273, 28966–28977 (1998).
66. Zhao, H. & Arnold, F. H. Directed evolution converts subtilisin E into a functional equivalent of thermitase. Protein Engineering, Design and Selection 12, 47–53 (1999).
67. Moger-Reischer, R. Z. et al. Evolution of a minimal cell. Nature 1–6 (2023) doi:10.1038/s41586-023-06288-x.
68. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6, 343–345 (2009).
69. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).

70. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589–595 (2010).
71. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at http://arxiv.org/abs/1207.3907 (2012).
72. Tange, O. GNU Parallel: The Command-Line Power Tool.
73. Lázaro, J. M., Blanco, L. & Salas, M. Purification of bacteriophage phi 29 DNA polymerase. Methods Enzymol 262, 42–49 (1995).
74. Soengas, M. S., Gutiérrez, C. & Salas, M. Helix-destabilizing activity of phi 29 single-stranded DNA binding protein: effect on the elongation rate during strand displacement DNA replication. J Mol Biol 253, 517–529 (1995).
75. Del Prado, A. et al. Insights into the Determination of the Templating Nucleotide at the Initiation of ϕ29 DNA Replication. J Biol Chem 290, 27138–27145 (2015).
76. Doerr, A., Foschepoth, D., Forster, A. C. & Danelon, C. In vitro synthesis of 32 translation-factor proteins from a single template reveals impaired ribosomal processivity. Scientific Reports 11, DOI: 10.1038/S41598-020-80827-8 (2021).
77. Peñalva, M. A. & Salas, M. Initiation of phage phi 29 DNA replication in vitro: formation of a covalent complex between the terminal protein, p3, and 5'-dAMP. Proc Natl Acad Sci U S A 79, 5522–5526 (1982).
78. McDonell, M. W., Simon, M. N. & Studier, F. W. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J Mol Biol 110, 119–146 (1977).
79. Blanco, L., Lázaro, J. M., de Vega, M., Bonnin, A. & Salas, M. Terminal protein-primed DNA amplification. Proceedings of the National Academy of Sciences 91, 12198–12202 (1994).
80. Carthew, R. W., Chodosh, L. A. & Sharp, P. A. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell 43, 439–448 (1985).
81. Berman, A. J. et al. Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. EMBO J 26, 3494–3505 (2007).

4

Synthetic Cells with Integrated DNA Self-replication and Membrane Biosynthesis

Abstract

The emergence, organization and persistence of cellular life can be ascribed to the functional integration of biological processes. A systemic approach to reconstruct a synthetic cell therefore relies on the integration of a minimal – yet sufficient – number of cellular functions in vitro. In this study, we demonstrate the integration of three life's essential 'modules' inside liposome compartments: DNA self-replication (DNArep), phospholipid synthesis (PLsyn), and gene expression to support the synthesis of the earlier two functions. Both modules, expressed from a six-gene synthetic genome using the 'protein synthesis using recombinant elements' (PURE) system, are compatible and minimally affected by substrate or cofactor crosstalk. However, incorporating genes related to a secondary module reduces the occurrence of liposomes exhibiting an active PLsyn module, while decreasing the yield of amplified DNA by DNArep. Quantitative analysis of the different phenotypes under various conditions was performed by flow cytometry and high-content fluorescence imaging of large (thousands of liposomes per sample) populations of gene-expressing liposomes. We further discuss possible optimization routes to accelerate module integration towards engineering an autonomous synthetic cell. Moreover, the genetically programmed vesicles equipped with membrane synthesis and genome self-replication abilities provide a minimal cell-free platform to better understand the way cellular processes are coupled at the genetic and metabolic levels.

The content of this chapter is based on a manuscript in preparation with Federico Ramírez Gómez, Mats van Tongeren, and Christophe Danelon.

Introduction

The construction of a synthetic cell is a still standing endeavour at the intersection of bioscience and engineering. This ambitious goal involves the assembly of a simplified, yet fully functional cellular entity, designed to converge life's essential processes. Bottom-up synthetic cell research has focused on understanding life's fundamental mechanisms reconstituted inside cell-like compartments [1–3]. The repository of genetic and protein parts functionally assembled from the ground up keeps growing with studies using purified or in-situ synthesized proteins [4–6]. Some of the recently explored areas include compartments and growth [7–10,10–14] division [15–19], DNA replication [20–23], energy regeneration [24,25], and cell-cell communication [26].

While these studies have yielded valuable insights into the specifics of each biological mechanisms, they did not yet address the higher-ordered life's complexity that lies in the synergy and coordination of the diverse cellular processes [27,28]. This prompts a new question: How does a minimal synthetic cell orchestrate its machineries to be considered alive? Conceptual illustrations of intracellular coordination trace back to the 20th century, where Ganthi's chemoton model underscored what a functional protocell needs, namely coupled genetic information storage, metabolism, and membrane compartmentalization [29]. Modern cell biology has revealed the highly integrated nature of the different functions supporting life, from the accurate allocation of resource and energy [30] to the intimate coupling between DNA replication and growth for maintaining a correct genomic integrity and cell division events [31,32].

Although the necessity of integrating functional modules to build a synthetic cell has been acknowledged [27,33], experimental attempts remain scarce. Hence, a huge gap persists between in vivo research and theoretical inquiry [3,34,24,35], and the bottlenecks that will inevitably hinder functional synergy of the reconstituted modules, remain unexplored [36]. Inspired by Ganthi's protocell concept [29], we herein explored the integration of three life's essential modules inside liposome compartments: DNA self-replication, growth through internal phospholipid synthesis, and transcription-translation for the in-situ production of the proteins involved in the above processes [37] **(Fig. 1a)**. Gene expression and DNA replication ensure information processing, storage and heredity, and gene-encoded phospholipid biosynthesis enables self-sustained growth and reproduction of synthetic cells.

For the DNA replication module (DNArep), we drew inspiration from the replication mode of the Phi29 genome [38], which we previously implemented in PURE system [21]. With only four replication proteins **(Fig. 1b)**, two of which are encoded in the replicating template (i.e., DNAP and TP), exponential DNA amplification was achieved within gene-expressing vesicles. We hereby modified this synthetic *DNArep* replicator and integrated genes for phospholipid synthesis (*PLsyn*) module, leading to the *DNArep-PLsyn* genome. For the *PLsyn* module, we chose a reduced *E. coli* Kennedy pathway encompassing four enzymes for the sequential conversion of acyl-CoA and glycerol-3-phosphate precursors into

phosphatidylserine (PS), the last intermediate product for phosphatidylethanolamine (PE) synthesis **(Fig. 2c)**. Membrane synthesis through PS production within gene-expressing vesicles has already been demonstrated using a PS-specific fluorescent probe [11].

In the present work, we constructed for the first time synthetic cell models with integrated molecular machineries for DNA self-replication, membrane synthesis, and transcription-translation metabolism. Importantly, we postulate that experimental attempts to integrate rudimentary functions, followed by evolutionary engineering strategies, is a more valuable approach than optimizing the individual modules separately prior to combination. We demonstrate that both DNArep and PLsyn modules can be encoded and expressed from a single synthetic DNA genome. The unprecedented liposome sample quality enabled high-content imaging of large populations of vesicles providing quantitative insights on the effects of genetic and metabolic coupling of cellular functions in a minimal in vitro system.

Results and discussion

Design, assembly and cell-free expression of the DNArep-PLsyn synthetic genome

We pinned the work for combining synthetic cell modules by constructing a synthetic DNA genome, named *DNArep-PLsyn*, encoding both DNA replication (DNArep) and phospholipid synthesis (PLsyn) bio modules. Previously designed T7-based monocistronic DNA parts encoding either DNA self-replication proteins [21], or lipid-synthesizing enzymes [11] were integrated into a single linear template containing six genes (two for DNArep and four for PLsyn), and flanked with Phi29 origins of replication **(Fig. 1a)**. To construct the *DNArep-PLsyn* synthetic genome, we iterated throughout different cloning strategies and found that template complexity (i.e., repetitive elements) often led to recombination events in *E. coli* [39–41]. We then opted for an in vitro DNA assembly approach using overlapping PCR to stitch the *DNArep* and *PLsyn* genetic parts **(Fig. 1d)**. We successfully obtained a linear template with the expected size (~9600 bp), and the sequence was validated by nanopore sequencing **(Fig. 1e)(Fig. S1).**

Next, we confirmed gene expression from *DNArep-PLsyn* in PURE system. The reaction was supplemented with GreenLys reagent for co-translational protein labelling [42]. Expression levels of the synthesized proteins were qualitatively compared with that from the separate *DNArep* and *PLsyn* fragments that were utilized for generating *DNArep-PLsyn*. All six encoded proteins were produced at detectable levels **(Fig. 1f)(Fig. S2)**. Only a slight reduction of protein synthesis yield was observed for *DNArep-PLsyn* template compared to the separate expression of each individual genetic module. This could be caused by resource sharing when the number of genes increases, but the effect was less pronounced than expected. We conclude that *DNArep-PLsyn* acts as an effective template for expressing all necessary proteins involved in both DNArep and PLsyn modules.

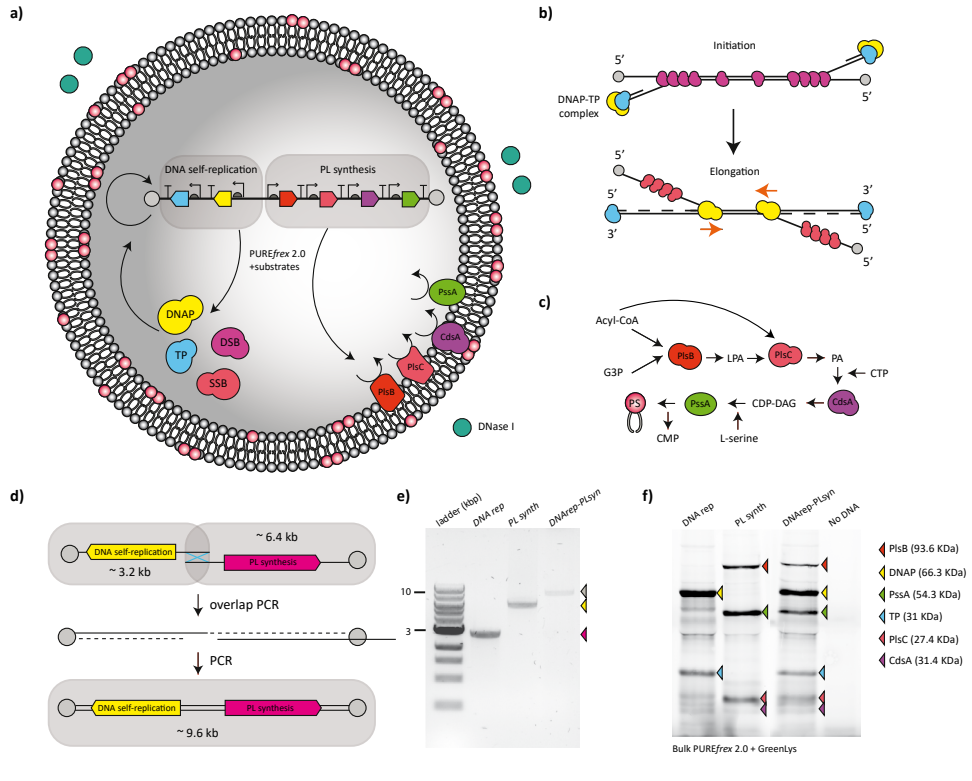


Fig. 1 A synthetic genome encoding both DNArep and PLsyn modules. a) Illustration of synthetic vesicles with coupled transcription-translation, DNA self-replication and phospholipid biosynthesis. DNAse outside of liposomes prevents external gene expression. **b)** Schematic of the initiation and elongation steps of the protein-primed DNA replication mechanism of the bacteriophage Phi29 [43]. **c)** Illustration of the four enzymatic steps of the *E. coli* Kennedy pathway to transform oleoyl-CoA (O-CoA) and glycerol 3-phosphate (G3P) to phosphatidylserine (PS). **d)** In vitro PCR assembly strategy to construct *DNArep-PLsyn* genome (~9.6 kb) containing a minimal DNA self-replication genetic module (*DNArep*) (~3.2 kb), and the partial Kennedy pathway module (*PLsyn*) (~6.4 kb). **e)** Agarose gel electrophoresis of the assembled *DNArep-PLsyn*, along with the *DNArep* and *PLsyn* DNA fragments used in its assembly. **f)** SDS-PAGE analysis of bulk IVTT reactions from assembled *DNArep-PLsyn* template, or the individual *DNArep* and *PLsyn* fragments. PURE system solutions were supplemented with GreenLys reagent for fluorescent labelling of synthesized proteins.

Integration of DNArep and PLsyn modules inside gene-expressing vesicles

- *Validation of DNArep and PLsyn module integration*

Our next aim was to evaluate and potentially optimize the simultaneous activity of the DNArep and PLsyn modules inside liposomes. Since each of the encoded modules may have a preferred reaction temperature (DNA replication works well at ~30 °C [21,38], while cell-free gene expression [44] and phospholipid biosynthesis [11] are effective at 37 °C), we decided to test different incubation temperatures. We encapsulated *DNArep-PLsyn* in liposomes together with PURE system, and the required substrates and cofactors for both DNA replication and phospholipid synthesis. We then ran the reactions at 30 °C, 34 °C, or 37 °C. After overnight incubation, DNA was stained with dsGreen intercalating dye [21,45],

and membrane-incorporated PS was labelled with the specific fluorescent probe LactC2-mCherry [11]. Samples were run on a flow cytometer and the intensity levels of the scattered light (as a general object reporter), dsGreen (DNArep activity reporter), and LactC2-mCherry (PLsyn activity reporter) were quantified **(Fig. 2a-c)**. Liposomes exhibiting functional DNArep or/and PLsyn modules were detected with all three temperatures when analysed separately **(Fig 2a)** or in combination **(Fig. 2b-c)(Fig. S3)** with a slightly higher occurrence for DNArep-active liposomes at 34°C **(Fig. 2a-c)**. Notably, a range of ~0.4 to 12% of the liposomes (corresponding to ~50 to 1200 liposomes across biological replicates at 34 °C) displayed both DNArep and PLsyn functions (ROI 2), while a larger fraction of liposomes was positive to either of the two modules (ROI 1 and ROI 4), or was inactive (ROI 3) **(Fig. 2b)**. Such a functional heterogeneity within the same clonal (here referring to the fact that one DNA species was used) population of liposomes is also observed in single-gene expression experiments [46]. This variability can be attributed to uneven loading or supply of substrates or cofactors, or to varying expression levels of the genetic modules between liposomes [11,46]. In addition, a significant variability across biological replicates (sample-to-sample heterogeneity) was observed **(Fig. 2b)**. For example, the percentage of DNArep-PLsyn-positive liposomes (ROI 2) was in average 2.9 ± 1.3 (mean ± s.e.m) across 8 biological replicates at 34 °C **(Fig. S3)**. Well acknowledged sources of variability in cell-free gene expression are DNA concentration and quality, especially at sub nanomolar amounts and in vitro assembled templates. In addition, variabilities on DNA self-amplification could perhaps influence PLsyn phenotype, hence ROI 2 occurrence. Nonetheless, our flow cytometry results demonstrate that functional integration of DNArep and PLsyn modules from a linear synthetic genome is possible, and can be obtained from different incubation temperatures.

To provide a more direct evidence of module activities, we assessed genome self-replication by absolute DNA quantification using qPCR and lipid synthesis by liquid chromatograhy-mass spectrometry (LC-MS). Two different regions were targeted for qPCR, one in the *p2* gene and one in the *pssA* gene. The results quantitatively confirmed that all tested temperatures were suitable for genome self-replication, with a slight preference for 34 °C **(Fig. 2d)**. We further investigated whether the full-length genome was amplified (vs. shorter fragments) by targeting all six genes by qPCR. These experiments were performed at 34 °C. Despite some variations in the concentration of replicated genes, the data clearly showed that the entire DNA between *p3* gene and *pssA* gene (~5000 bp apart) was amplified about 10-fold **(Extended Fig. 1)**. Small differences could arise from DNA replication arrest events, leading to incomplete fragment amplification [47,48]. Since qPCR amplifies only ~200-bp regions and the terminal origins of replication were not targeted, we also performed PCR DNA recovery from liposome samples followed by agarose gel analysis of the amplification products. The entire *DNArep-PLsyn* genome (within the resolution of agarose gel electrophoresis) could be recovered from diluted liposome reactions **(Extended Fig. 1c)(Fig S4)**. Shorter DNA species were also observed **(Fig. S4)**, suggesting that the DNArep-PLsyn genome may have

experienced incomplete self-replication or that smaller DNA fragments were generated during PCR recovery.

Production of PS and intermediate enzymatic products of the reconstituted phospholipid biosynthesis pathway was directly demonstrated by LC-MS. Accumulation of PA suggests that PlsC, but not PlsB, CdsA or PssA, may be limiting the yield of PS production **(Fig. 2e) (Fig. S5)**. Considering that phosphatidylglycerol (PG) accounts for 12% of the total lipids, we estimated that synthesized PS could represent 0.7% of the total lipid content after 16 hours incubation at 34 °C **(Fig. 2e)**. It is relevant to note that LC-MS measurements cover the entire liposome population, averaging out PS amount across all liposomes. Increasing initial PG quantities could be an interesting step to explore to ameliorate PssA activity and possibly PS production [11].

Overall, these results demonstrate for the first time the successful integration of two genetically programmed cellular modules, DNA self-replication and phospholipid membrane synthesis. Amplification of full-length synthetic genomes from encoded proteins and production of PS from a 4-step enzymatic reaction were shown at temperatures ranging from 30 °C to 37 °C, with a slightly higher DNA replication efficiency at 34 °C. Although the percentages of liposomes exhibiting the two active modules were rather low, the large population of gene-expressing vesicles that could be generated from a few microliter samples led to several hundreds of liposomes with the combined phenotype.

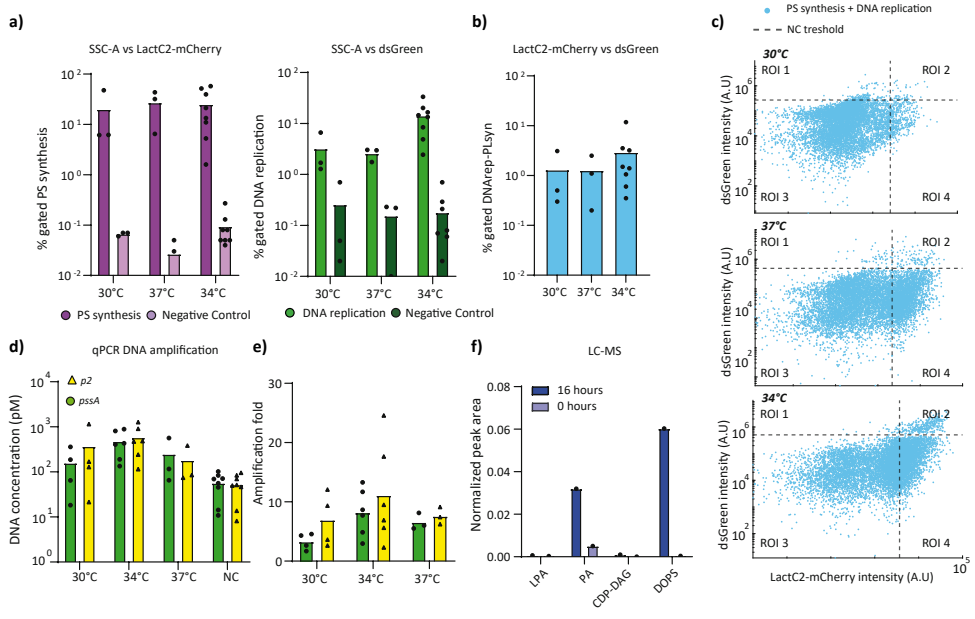


Fig. 2 Validation of DNArep and PLsyn protein activity inside gene-expressing liposomes at different incubation temperatures. a) Percentage of liposomes with active PS synthesis (left) and DNA replication (right) under 30, 37 and 34 °C incubation temperatures. Flow cytometry data are SSC-A vs. dsGreen for DNA replication and SSC-A vs. LactC2-mCherry for PS synthesis. Data points represent biological repeats and bar height the mean value. Raw data from

individual replicates can be found in Fig. S3. **b)** Percentage of liposomes exhibiting dual dsGreen and LactC2-mCherry signals at 30, 37 and 34 °C incubation temperatures. Joint phenotype populations were selected from LactC2-mCherry vs dsGreen scatter plots. Raw data from individual replicates can be found in Fig. S3. **c)** Flow cytometry scatter plots from liposome samples displaying four regions of interest (ROI 1-4) at all tested temperatures: DNArep-active liposomes are in ROI 1, PLsyn-active liposomes in ROI 4, and liposomes with both active DNArep and PLsyn modules are in ROI 2. Vertical and horizontal dashed lines indicate intensity threshold values that have been defined using control samples (see Fig. S3). Data from additional biological repeats can be found in Fig. S3. **d)** Absolute DNA quantification by qPCR of samples incubated at 30, 37, and 34 °C. qPCR target regions (~200 bp) are from *pssA* and *p2* genes. The negative control (NC) represents calculated DNA values at initial incubation points (0 hour). **e)** Amplification fold of *DNArep-PLsyn* DNA calculated from qPCR data in panel d: end-point (16 hours) DNA concentration / DNA concentration at time zero. Data points represent biological repeats and bar height the mean value. **f)** LC-MS detection of DOPS and PLsyn intermediate enzymatic products before and after expression of *DNArep-PLsyn*. Peak area for each compound was normalized to that of DOPG. Additional biological repeats and negative controls can be found in Fig. S5.

- *High-content imaging of DNArep and PLsyn phenotypes*

We next characterized liposomes expressing the *DNArep-PLsyn* at 34 °C by fluorescence confocal microscopy, allowing us to extract multidimensional features that are not possible to obtain with flow cytometry. In particular, we asked whether liposome size, lamellarity, or morphology could affect or be affected by module activity. Direct visualization of single vesicles also provides more precise characterization of the different phenotypes, hence a more accurate classification based on activity levels. Quantitative analysis was performed with an in-house developed software called SMELDit (see methods section) that enables automated liposome identification, feature analysis, and image recovery from scattered data plots. dsGreen and LactC2-mCherry signals were utilized as fluorescent markers for DNArep and PLsyn activity, respectively **(Fig. 3a)**. We observed that the addition of the substrates and cofactors, and expression of *DNArep-PLsyn* did not affect liposome sample quality **(Fig. 3a)(Fig. S6)**. Notably, images unravelled phenotypic traits that could not be inferred from flow cytometry data, such as the presence of bright dsGreen spots in the vesicle lumen, which result from active DNA replication. A similar observation has been reported during amplification of a shorter DNA self-replicator and was attributed to an induced condensation of highly concentrated DNA [21,45]. It is interesting to note that such a phenomenon is also possible with a 3-fold longer DNA template (~9.6 kbp vs. ~3.2 kbp) containing more expressed genes (6 vs. 2)

When aggregating data from all biological replicates, over 34000 liposomes were recognized. A phenotype map corresponding to the two-dimensional plot of LactC2-mCherry vs. dsGreen signals from single vesicles was generated **(Fig. 3b)**. Liposomes were classified according to four different phenotypes based on intensity thresholding, akin to flow cytometry data analysis (ROI 1-4) **(Fig. 2c)**. We found that ~8% of liposomes, corresponding to over 2900 liposomes, had coexistent DNA replication and PS synthesis (ROI 2). Vesicles with either active PLsyn (~10%, ROI 4) or active DNArep (~31%, ROI 1) module were more abundant **(Fig. 3b)**. We then questioned whether liposome sizes varied across the four regions, for example as a result of membrane synthesis. Vesicle size distribution

was computed for each phenotypic region **(Fig. S7)**. No marked differences in the median values of the apparent diameter between active (ROI 2 and 4) and inactive (ROI 1 and 3) PLsyn module were observed (3.7 ± 2.4 µm median across all ROIs), indicating that the yield of newly synthesized lipids is not sufficient for physical growth of liposomes.

To account for the variability across biological replicates, we constructed the phenotype map for each replicate sample **(Fig. 3c)(Fig. S8)**. Despite clear variations in the percentages of gated liposomes in each region, all replicates were populated with vesicles exhibiting simultaneous DNArep and PLsyn activity **(Fig. 3c)**. Finally, we examined the LactC2-mCherry and dsGreen intensity values for every liposome as this may reveal differences in the efficacy of module activity. From both pooled data and individual replicates, we observed no strong differences between regions **(Fig. 3d)(Fig. S8)**. This result suggests that DNArep activity is not lessened when coupled with PLsyn activity, and vice versa, in liposomes where at least one module is active. This finding point to a robust compatibility between the two functions, with minor crosstalk effects in the presence of both sets of module-specific substrates and cofactors.

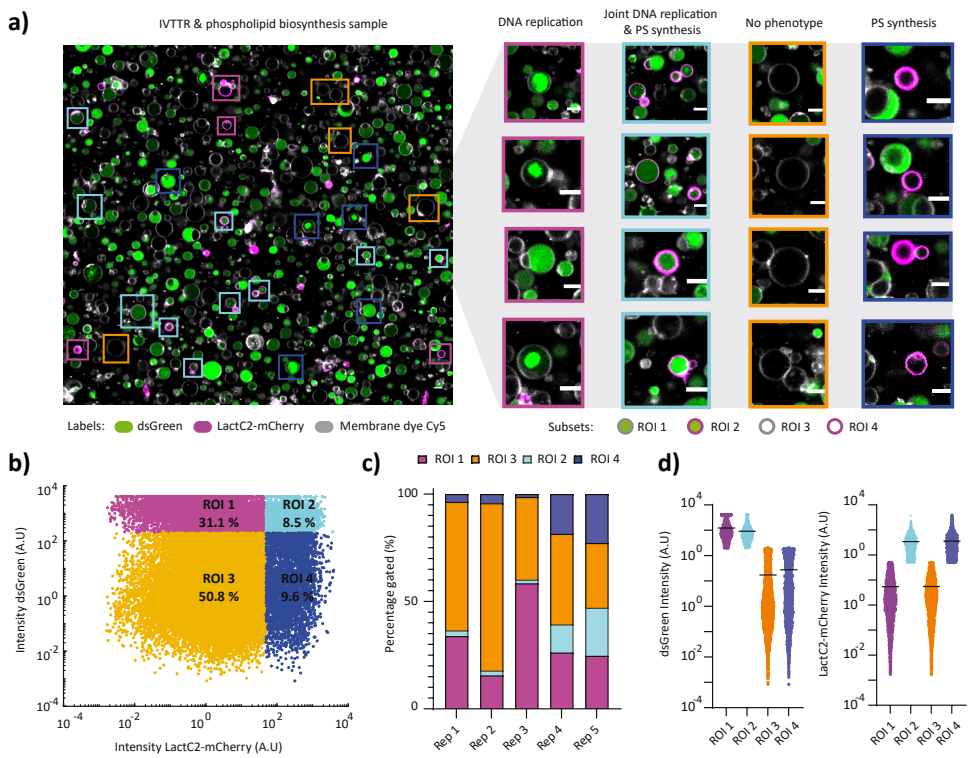


Fig. 3 High-content imaging of DNArep and PLsyn active phenotypes. a) Confocal microscopy images from gene-expressing liposomes with complete DNArep and PLsyn reaction conditions. Membrane dye (Cy5) is coloured in white, LactC2-mCherry in magenta, and dsGreen in green. Scale bar is 5 µm. Four distinct liposome phenotypes used for classification are highlighted: DNArep (ROI 1), dual DNArep and PLsyn (ROI 2), no module activity detected (ROI 3), and PLsyn (ROI 4). **b)** LactC2-mCherry vs. dsGreen phenotype map from SMELD-it image analysis on all biological repeats (~34000 liposomes). Identified population subsets are gated into ROI 1-4 and coloured as in panel a. Percentages of liposomes per ROI are appended and were calculated from the pooled dataset. Phenotype maps from individual biological repeats, as well as minus DNA negative control samples can be found in Fig S8. **c)** Phenotype map (gated ROIs) from individual biological repeats. Dual phenotype region (ROI 2) is present in all replicates with at least ~100 identified liposomes. Specifically, 273 liposomes on Rep 1, 159 liposomes for Rep 2, 94 liposomes for Rep 3, 234 liposomes for Rep 4, and 2204 liposomes on Rep 5. **d)** dsGreen and LactC2-mCherry intensity profiles across all ROIs from panel b. Each dot represents an SMELD-it identified liposome. Horizontal line indicates the mean of each data cluster.

Crosstalk effects between DNArep and PLsyn modules

- *Influence of turning a module ON and OFF on the activity of the other module*

To better understand the influence that active DNArep or PLsyn modules may have on each other, we assayed liposomes containing the full synthetic genome, this time by adding either of the two sets of substrates/cofactors (DNArep or PLsyn) **(Fig. 4a)**. An additional condition was tested, where all DNArep substrates/cofactors were supplied, except for dNTPs. This turns DNArep module OFF but keeps all the other substrates present (i.e., SSB, DSB, ammonium sulphate). We reasoned that possible inhibitory effects may arise by the substrates themselves or by intermediate reaction products. Moreover, we speculated that DNA processing by the Phi29 DNA polymerase may either have a beneficial effect on PS synthesis by increasing the yield of synthesized enzymes through genome amplification [49] or have an adverse effect by hindering gene expression through collision events between DNA-interacting proteins (DSB or DNA polymerase vs. RNA polymerase) [48–50].

Following the same protocol as described above, we verified that DNArep and PLsyn were only active when their corresponding substrates/cofactors were present **(Fig. 4b,c)(Fig. S9)**, confirming that nonspecific staining with dsGreen and LactC2-mCherry was negligible. Using data pooled from all biological replicates, we found that the occurrence of DNArep-active liposomes (ROI 1+2) reduced only from ~38% to ~31% when PLsyn substrates were supplied, while the occurrence of PLsyn-active liposomes (ROI 4+2) reduced only from ~18% to ~15%/~10% (with/out dNTPs) when DNArep substrates/cofactors were supplemented **(Fig. 4d)(Fig. S10a)**. Moreover, the intensity distributions of DNArep and PLsyn activity reporters were similar regardless of the presence or absence of the substrates from the other module **(Fig. 4e)**. These results suggest that the activity of a given module is not significantly influenced when switching the other module ON.

When examining individual replicates, we found a higher variability on the occurrence of PLsyn-active liposomes (ROI 2+4) when reactions contained all DNA replication substrates compared to in their absence **(Fig. S10b)**. This suggests that the sample-to-sample variations in the fraction of PLsyn-active liposomes when both modules were ON could be caused by some DNArep activity or by its substrates/cofactors. We hereby conclude that the appearance of liposomes displaying an active phenotype is not affected by inhibitory effects caused by substrate/cofactor crosstalk nor by module co-activity.

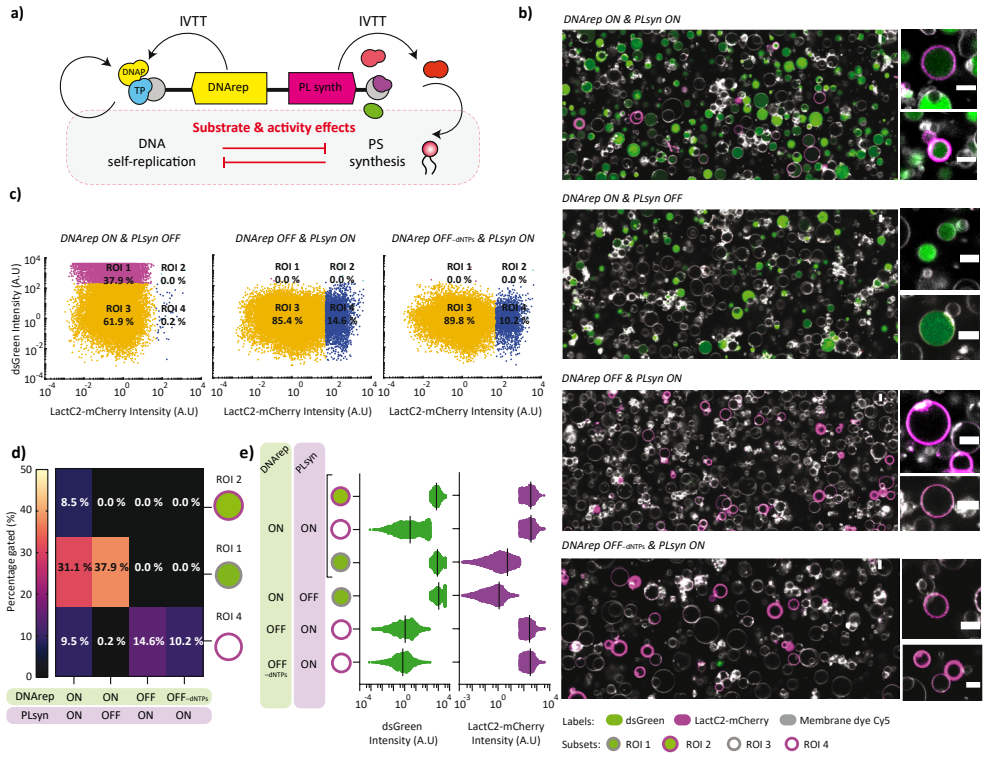


Fig. 4 Effects of turning a module ON and OFF on the activity of the other module. a) Schematic of co-active DNArep and PLsyn with an emphasis on substrate and co-activity crosstalk effects. **b)** Confocal microscopy images of liposome samples with different substrate additions to trigger module activity. ON and OFF labelling indicates presence (ON) or absence (OFF) of substrates/cofactors for either DNArep or PLsyn. Liposome membrane dye (Cy5) is coloured in white, LactC2-mCherry in magenta, and dsGreen in green. Scale bar is 5 µm. **c)** Phenotype scatter plots from SMELDit image analysis (LactC2-mCherry vs. dsGreen) on all biological repeats with only one active module. Classified liposome subpopulations are labelled as ROI 1-4 and gated in different colours as in Fig. 3b. Appended percentages are calculated from the pooled dataset including all biological repeats. Scatter plots from the individual repeats can be found in Fig. S9. **d)** Phenotype heatmap with gated percentage values for ROIs 1,2,4 calculated across all replicates with active and/or inactive modules. Percentages for individual repeats can be found in Figs. S9 and S10. **e)** dsGreen and LactC2-mCherry intensity profiles across gated ROIs from panel d. Each dot represents a SMELDit identified liposome. Vertical lines indicate the mean value of each data cluster.

- *Interplay between DNA template and module co-expression on DNArep and PLsyn activity*

Having established that functional integration of the DNArep and PLsyn modules was minimally affected by substrate/cofactor crosstalk effects, we studied what the influence of introducing the genes of a module could have on the activity of the other. In other words, we asked whether co-expression of the two genetic modules *DNArep-PLsyn*, and the nature of the DNA template, could influence the propensity to form active liposomes. For this, we compared liposome populations with *DNArep-PLsyn* genome against liposomes with DNA templates encoding only the genes of a single module, i.e., either *DNArep* or

PLsyn. Here, plausible effects that may compromise module activity from *DNArep-PLsyn* genome encompass gene-expression resource sharing [25,44,51] and premature transcription termination caused by collision events between DNA interacting proteins, such as the DNA polymerase and RNA polymerase [48]. Both effects become more prominent as the number of genes increases. As observed above from *DNArep-PLsyn* PCR recovery **(Extended Fig. 1)**, we cannot exclude the possibility of impaired replication of the full-length template, which would occur more frequently with *DNArep-PLsyn* than with the shorter *DNArep* template.

We prepared liposomes containing either *DNArep-PLsyn*, *DNArep*, or *PLsyn* DNA, along with the full set of substrates and cofactors. As expected, microscopy images showed that the appearance of liposome phenotypes was directed by the encapsulated DNA program **(Fig. 5b,c)(Fig. S11)**. Interestingly, the percentages of DNArep-positive liposomes were similar with and without co-expression of the *PLsyn* genes, decreasing only from ~45% (ROI 1) to ~40% (ROI 1+2) when PLsyn was co-expressed **(Fig. 5d)**. Conversely, the percentages of PLsyn-positive liposomes dropped from ~38% (ROI4) to ~18% (ROI4+2) when *DNArep* was co-expressed **(Fig. 5d)(Fig. S12)**, suggesting that PLsyn activity is more sensitive to genetic background and expression burden than DNArep activity. This effect may also limit dual-module activity in liposomes containing *DNArep-PLsyn*, explaining the higher prevalence of a single phenotype in *PLsyn* and *DNArep*-containing liposomes (ROI 4, ~38% on PLsyn and ~44% on DNArep), compared to those with *DNArep-PLsyn* displaying both phenotypes (ROI 2, ~8%) **(Fig 5d)**.

While the occurrence of liposomes exhibiting an active PLsyn module was influenced by the co-expression of *DNArep*, we noticed that the intensity distributions reporting the levels of DNArep and PLsyn activity were similar under single- and double-genetic module expression conditions **(Fig. 5e)**. For PLsyn, this suggests that, above a detectable activity threshold, PS production yield was not affected by co-expression of DNArep genes. For DNA replication, however, dsGreen signal intensity is proportional to DNA quantity, which accounts for DNA length and amplification fold. Therefore, similar dsGreen intensities from replicated *DNArep-PLsyn* and *DNArep* templates suggest that the amplification fold of *DNArep-PLsyn* is lower than that of *DNArep* given its larger size (~9.6 kbp vs. ~2.3 kbp). To test this hypothesis, we performed absolute DNA quantitation by qPCR, confirming that *DNArep-PLsyn* replicates at a lower yield than *DNArep* (~10-fold vs. ~100-fold) **(Fig. 5f,g)**. Considering that the yield of synthesized DNAP and TP does not differ much from the *DNArep-PLsyn* or *DNArep* templates (in the absence of module-specific substrates and cofactors), we speculate that the processivity of DNAP – and not replication initiation – might be the amplification bottleneck. This can be surprising, given that the Phi29 genome has a size of 20 kbp. However, the fully synthetic nature of the liposomal environment highly differs from the conditions for natural Phi29 genome replication in the bacterial host.

While the above experiments focused on the end-point characterization of module

functionalities, we then examined how phenotype appearance developed in the course of gene expression. Liposome samples containing either *DNArep-PLsyn*, *DNArep* or *PLsyn* DNA in the presence of all substrates and cofactors for both modules were incubated for different times, and were analysed by flow cytometry and qPCR. When expression was directed by *DNArep-PLsyn*, liposomes with joint phenotypes (ROI 2) appeared only after 4 hours of incubation and their occurrence plateaued after 8 hours, while events corresponding to either DNArep (ROI 1) or PLsyn (ROI 4) activity appeared earlier, at 1-4 hours, and their occurrence plateaued after 4 hours **(Extended Fig. 2a,b)(Fig. S13)**. Expression of single-module DNA templates confirmed that liposomes exhibiting DNArep or PLsyn could already be observed after 1-4 hour incubation **(Extended Fig. 2a,b)(Fig. S13)**. These results indicate that simultaneous build-up of PS and DNA above detectable threshold (ROI 2) from *DNArep-PLsyn* expression was delayed compared to a situation where only one module was expressed. Additionally, 8 hours of incubation was sufficient to reach the maximum percentage of liposomes displaying coupled DNArep and PLsyn (ROI 2). Direct DNA quantitation by qPCR revealed that both *DNArep-PLsyn* and *DNArep* template concentrations reached a steady state after 4 hours of incubation **(Extended Fig. 2c)**, although the amplification yields differed. These results show that DNA length and/or the number of transcriptional units reduce the yield of replicated template when coupled to gene expression but the maximum amount of DNA is reached at similar times.

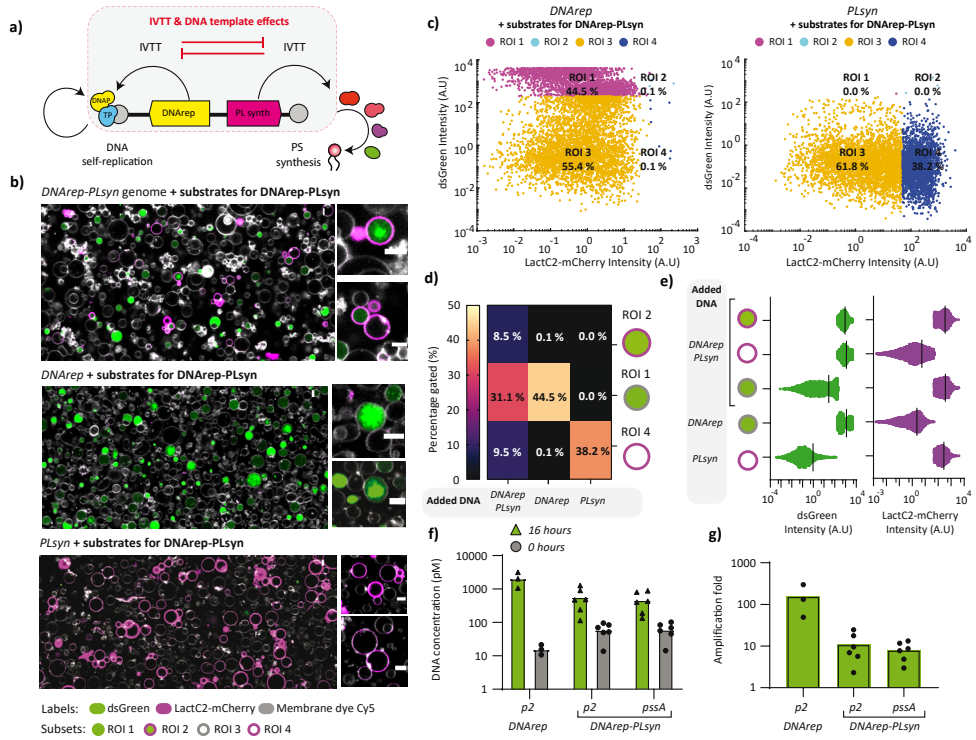


Fig. 5 Effects of DNA template and co-expression of genetic modules on DNArep and PLsyn activity. a) Schematic of the expression of the *DNArep* and *PLsyn* genetic modules and possible crosstalk effects. **b)** Confocal microscopy images of liposome samples with all substrates and cofactors but with different DNA templates as indicated. Liposome membrane dye (Cy5) is coloured in white, LactC2-mCherry in magenta, and dsGreen in green. Scale bar is 5 µm. **c)** Phenotype scatter plots from SMELDit image analysis (LactC2-mCherry vs. dsGreen) on all biological repeats with either *DNArep* or *PLsyn* as DNA template. Classified liposomes in ROI 1-4 are gated in different colours. Appended percentages are calculated from the pooled data of all biological repeats. Scatter plots from individual biological replicates can be found in Fig. S11. **d)** Phenotype heatmap constructed from all repeats with different template conditions: *DNArep-PLsyn*, *DNArep* and *PLsyn* DNAs. **e)** dsGreen and LactC2-mCherry intensity profiles across ROIs from panel d. Each dot represents a SMELD-it identified liposome. Vertical lines indicate the mean of each data cluster. **f)** Absolute DNA quantification from liposome samples containing either *DNArep-PLsyn* genome or the minimal self-replicator *DNArep*. The targeted regions on the *pssA* and *p2* genes are indicated. **g)** Amplification fold of *DNArep-PLsyn* and *DNArep* DNA templates calculated from qPCR data in panel f: end-point (16 hours) DNA concentration / DNA concentration at time zero. Data points represent biological repeats and bar height the mean value.

Discussion and outlook

We hereby accomplished the integration of three essential synthetic cell modules: DNA self-replication, PL synthesis for compartment growth, and a minimal metabolism for transcription-translation. Self-replication of DNA and PL biosynthesis were genetically programmed on a six-gene DNA construct co-encapsulated with the PURE system machinery inside liposomes. Expression of both genetic modules resulted in the appearance of vesicles with coupled DNA amplification and membrane biosynthesis. Although joint-phenotypes did not dominate the liposome population, we routinely obtained more than hundred vesicles with coupled activities per few microliters of sample.

We found that the activity of a given module does not affect the activity of the other from the expressed *DNArep-PLsyn* DNA. However, the presence of additional genes reduced the occurrence of PS producing liposomes and the yield of DNA amplification. To remedy this negative effect, different designs of the *DNArep-PLsyn* genome could be tested to optimize the metabolic balance and resource allocation for gene expression. For instance, gene organization in the form of operons [52], or a different combination of regulatory elements (i.e., variable RBSs, SP6 or T7 promoters, T7 or vsv terminators) [53–56] could be attempted to modulate the amounts of the gene expression products. Alternatively, stringent temporal control over gene expression may be realized by implementing genetic circuits, for example ON/OFF switches regulated by specific signals [57,58]. With this, biological processes, such as genome replication, transcription, and PL synthesis can be properly synchronized in time. Rational engineering could also aid in reducing intrinsic gene-expression variability [59], or to ameliorate protein specific properties like dsDNA affinity for improving DNAP processivity [60].

As a complementary strategy to rational design, we propose to utilize evolutionary tools to optimize DNArep and PLsyn synergy [33]. Considering that translation is a major gene-expression bottleneck [61–63], RBS libraries can be scanned for finding the best (not necessarily the highest) *DNArep-PLsyn* expression levels that lead to a higher occurrence of liposomes

with both functions. The RBS calculator [64–67] already aided in vivo pathway optimization [68], and could also be used to fine-tune synthetic cell pathways. Complementary, evolutionary goals could aim at enhancing protein activity by mutagenizing the protein encoding sequence [22]. Moreover, evolution of the *DNArep-PLsyn* genome may occur through multiple rounds of replication-coupled transcription-translation by accumulating advantageous mutations during in situ reactions [69–71]. DNA replication in clonal liposome populations has shown to enhance phenotypic output and DNA enrichment efficiency (chapter 2 of this thesis) [49]. Selection of the desired variants can be done by screening and sorting liposomes based on the intensity of the PLsyn- and DNArep-reporting probes (LactC2-mCherry and dsGreen), using fluorescence activated cell-sorting (FACS) [11,72] .

Our work lays the ground for future module integration efforts. More genes involved in other cellular processes could be introduced to our *DNArep-PLsyn* partial genome, such as division machinery [15,17,73], and PURE system self-regeneration [24,25,61,74]. As the level of genetic and functional complexity increases, directed evolution stands out as the most suitable strategy to accelerate synergy [33]. Here, technological development in image-based screening and selection of liposomes will become invaluable to select complex phenotypic traits that can no longer be identified by FACS [75,76] (e.g., membrane deformation, and protein or DNA localization inside vesicles [45]).

Materials and Methods

Buffers and solutions

All buffers were made with MilliQ grade water with 18.2 MΩ resistivity (Millipore, USA). All chemicals were purchased from Sigma-Aldrich unless indicated otherwise.

DNA construct design

Plasmid G363, utilized to construct *DNArep-PLsyn* genome, was assembled using a stepwise Golden Gate ligation between six independent transcriptional cassettes (with *plsB*, *plsC*, *cdsA*, *pssA*, *p2*, and *p3* genes) and a Phi29 origins flanked backbone [21]. First, two-fragment ligations were performed with *plsB* and *plsC*, *cdsA* and *pssA*, and *p2* and *p3*. Then, these two-gene cassettes were ligated into the Phi29 ori-flanked vector [21]. Plasmid G435, containing the ori-flanked *DNArep* genes and utilized for *DNArep-PLsyn* genome assembly, was derived from a previously cloned construct encoding for Phi29 DNA replication proteins (G95) [21]. Plasmid G555 harbouring the ori-flanked *PLsyn* gene pathway was constructed by subcloning the four-gene *PLsyn* fragment from G363 (digested out with Xho1 and Nco1 restriction enzymes), into a Phi29 origins flanked vector [21], also digested with Xho1 and Nco1. PCR fragments for *DNArep-PLsyn* genome assembly were prepared from G363 with 5'-phosphorylated 491 ChD and 1302 ChD primers (*PL synth* fragment), and from G435 with 5'-phosphorylated 492 ChD and 1289 ChD primers (*DNA rep* fragment) **(Fig. 1e)**. Linear DNA fragments containing either of the two genetic modules were prepared by PCR using 5'-phosphorylated primers 491 and 492 ChD, and KOD Xtreme Hot Start

DNA polymerase. The reaction solution contained (final concentrations) 1X Extreme Buffer (MERCK), 0.02 U/µL KOD DNA polymerase (MERCK), ~0.3 ng/µL DNA template, 300 nM forward and reverse primers, 6 µM dNTPs, and MilliQ water up to 50 µL final volume. The thermocycler protocol was set to 94 °C for 2 min for polymerase activation, followed by 25-30 cycles of 10 s 98 °C, 20 s 60 to 68 °C, 30-60 s/kb at 68 °C depending on the length of the desired amplicon. All PCR amplicons were verified for correct DNA length by 0.7-1% agarose gel electrophoresis before PCR clean-up with QIAquick PCR purification kit (Qiagen). Whenever needed (e.g., if additional unexpected bands were observed in the gel electrophoresis), DNA was purified directly from the agarose gel with a QIAquick Gel Extraction Kit (Qiagen). Both PCR and gel purification standard protocols were modified with a longer column drying step of ~5 min at 10000 g (Eppendorf Centrifuge 5415 R) before DNA elution with MilliQ water. Purified DNAs were quantified by Nanodrop 2000c spectrophotometer (Isogen Life Science). Plasmid and PCR descriptions can be found on **Table. S1** and **Table S2**, respectively. All DNA sequences can be provided upon request.

In vitro assembly of the DNArep-PLsyn genome

DNArep-PLsyn genome was constructed by overlap PCR from *DNA rep* fragment and *PL synth* fragment with two main PCR steps: (i) plasmid overlap and full product DNA extension, and (ii) addition of 491 and 492 ChD primers and PCR amplification of the full *DNArep-PLsyn* product. 47 µL reactions were prepared with final concentrations of 1X Extreme Buffer (MERCK), 0.02 U/µL KOD DNA polymerase (MERCK), ~0.3 ng/µL DNA template, 6 µM dNTPs, and MilliQ water. The thermocycler was programmed for 2 min at 94°C for polymerase activation, followed by 5 cycles of (10 s at 98°C, 20 s at 60°C and 3 min and 40 s at 68°C), and 20 cycles of (10 s at 98°C, 7 min at 68°C). For the last 20 cycles of the overlap PCR program, primers 491 and 492 ChD were added to the reaction to a final concentration of 300 nM each.

Purification of SSB, DSB, and LactC2-mCherry

Purified SSB and DSB Phi29 auxiliary proteins were produced and stored in −80°C as previously described in [21] . SSB stock concentration was 10 mg/mL, stored in a buffer with 50 mM Tris, pH 7.5, 60 mM ammonium sulphate, 1 mM EDTA, 7 mM β-mercaptoethanol (BME), and 50% glycerol. DSB stock concentration was 10 mg/mL, stored in a buffer with 50 mM Tris, pH 7.5, 0.1 M ammonium sulphate, 1 mM EDTA, 7 mM BME, and 50% glycerol. LactC2-mCherry protein stock was produced and stored in −80°C as described in [49]. LactC2-mCherry stock concentrations were determined by a Bradford assay, and stored in a buffer with 50 mM HEPES-KOH, pH 7.5, 150 mM NaCl, and 10% glycerol. Before liposome staining, the protein stock vial was centrifuged at maximum speed (Eppendorf Centrifuge 5415 R) for 10 min to spin down protein aggregates. If necessary, LactC2-mCherry protein stock was diluted in homemade PURE buffer (PB) consisting of 180 mM potassium glutamate monohydrate, 14 mM magnesium acetate tetrahydrate, and 20 mM HEPES at a pH of 7.6 (adjusted with potassium hydroxide) before usage.

In-liposome gene expression

Lipid-coated beads were prepared as explained in [11,21] with minor modifications. A lipid mixture was prepared with chloroform-dissolved lipids (Avanti Polar Lipids) in a 5-mL round-bottom flask. The solution contained 49.5% DOPC, 33.7% DOPE, 12% DOPG, 3.8% 18:1 CL, and 1% DSPE-PEG-biotin mass composition, for a total mass of lipids of 2.02 mg. For confocal microscopy experiments the primary lipid composition was adjusted to include 0.05% mass of DSPE-PEG(2000)-Cy5. The resulting mix was supplemented with 25.4 µmol of rhamnose (Sigma-Aldrich) dissolved in methanol. 600 mg of 212–300-µm glass beads (Sigma-Aldrich) were added to the 5-mL round-bottom flask containing the lipid/rhamnose solution, and the solvent was evaporated with a rotary evaporator (Heidolph) for 2 hours at room temperature and 200 mbar. The lipid-coated beads were recovered and further dried in a desiccator overnight. The dried lipid-coated beads were flushed with argon and stored at −20 °C until use. In 1.5 mL Eppendorf tubes, PURE*frex*2.0 (GeneFrontier) IVTT reactions were assembled as recommended by the manufacturer using 500 pM DNA template and 0.75 U/µL of Superase·In RNase inhibitor (Thermo Fisher). When indicated, the IVTT mixture was supplemented with the required substrates and cofactors for DNA replication (300 µM dNTPs, 20 mM ammonium sulphate, 0.75 mg/mL SSB, and 0.21 mg/mL DSB, final concentrations) or/and PL synthesis (1 mM CTP, 500 µM G3P, 500 µM L-Serine, and 5 mM β-mercaptoethanol, final concentrations. The oleoyl-CoA (O-CoA) precursor was added in a next step as described below). 10 mg of lipid-coated beads freshly desiccated for 20-30 min were added into a 1.5 mL Eppendorf tube containingto 20 µL of swelling solution. The sample in the Eppendorf tube was kept in a 4 °C room for gentle rotation with an automatic tube rotator (VWR) for 45-60 min, and was then subjected to four freeze-thaw cycles with alternating steps of dipping in liquid nitrogen and thawing on ice. Liposomes were recovered with a cut pipette tip and transferred to a PCR tube containing DNase One (NEB) for a final concentration of 0.1 U/µL. For PLsyn activity, the sample was further transferred to another PCR tube with a pre-deposited 1 µg O-CoA (Avanti Polar Lipids) film, prepared as described in [11]. With the added liposome suspension, the final O-CoA concentration was 176 µM, and when needed (i.e., liposome suspension volumes changed), the pre-deposited O-CoA quantity was adjusted to maintain the same final concentration. Samples were incubated at 30, 37, or 34 °C for 16 hours. A maximum of 15 µL liposome suspension was handled per PCR tube. If higher volumes were needed, the suspension was distributed across different PCR tubes with independently added O-CoA dried lipid film.

Flow cytometry

In a PCR tube, a liposome suspension (1.5 µL) was mixed in a 1:1 ratio with a 1000x diluted dsGreen (Lumiprobe) stock solution in PB and the sample was left to incubate in the dark at room temperature for 30 min. The 3 µL liposome-dsGreen mixture was further diluted by adding 147 µL of a 1:1000 dsGreen:PB solution. To remove possible remaining glass beads, the 150 µL solution was filtered through a 35-µm nylon mesh of a cell-strainer cap from

5-mL round-bottom polystyrene test tubes (Falcon). With a large volume pipette tip, 138.5 µL of the filtered solution was transferred to a 2 mL round-bottom tube to which 1.5 µL of 1:1000 dsGreen:PB solution and 10 µL of LactC2-mCherry probe were added to obtain a final LactC2-mCherry concentration of 300 nM and a final volume of 150 µL. The sample was incubated for 1 hour before injection in a FACSCelesta flow cytometer (BD Biosciences) set up with a 488-nm laser and 530/30 filter for detection of dsGreen, and with a 561-nm laser and 610/20 filter for detection of LactC2-mCherry. Photon multiplier tube voltages were 375 V for forward scatter, 260 V for side scatter, 370 V for LactC2-mCherry, and 550 V for dsGreen detection. Loader settings were set to 50 µL injection volume with no mixing and 800 µL wash between sample runs. For each sample ~20000 events were recorded. The raw data were analysed and pre-processed using Cytobank (https://community.cytobank.org/) to filter out possible aggregates and liposome debris as previously described in [49]. Text files with all SSC-A, dsGreen, and LactC2-mCherry intensity values were exported from Cytobank and plotted with MATLAB.

Confocal microscopy

Liposome samples (2 µL) were transferred in custom-made glass chambers functionalized with BSA-biotin:BSA and Neutravidin, as previously described in [11], and pre-filled with 13 µL of a staining solution (1X dsGreen and 3 µM LactC2-mCherry diluted in PB). Chambers were incubated in the dark at room temperature for 1 hour. Confocal microscopy imaging was carried out on a Nikon Eclipse Ti (NIS-Elements AR software) using a 100X oil immersion objective. Laser settings for image acquisition were set to: 488-nm laser with 20 HV, -10 offset, and 1.0 intensity for dsGreen, 561-nm laser with 50 HV, -10 offset, and 1.00 intensity for LactC2-mCherry, and 640 nm laser with 95 HV, -5 offset, and 5.00 intensity for Cy5 membrane dye. Each sample was imaged by automated acquisition of 10 by 10 fields of view, stitched together with a ~5% overlap. The sample height was adjusted manually to detect as many liposomes as possible, while also avoiding background debris.

Image analysis

Confocal images were analysed manually with Fiji [77] and automatically with SMELDit, an in-house-developed MATLAB script to automatically extract single liposome features while indexing each analysed liposome. In short, the Cy5 and mCherry channels of each image are combined and convolved with a Laplacian filter kernel to determine membrane boundaries. To set what pixels belong to the inside of each liposome a binarization step, with a consistent cutoff based on previous data, followed by a filling and erosion step were utilized. The resulting binary image displayed separate segments, each representing the lumen of individual liposomes. This step was followed by an additional selection for filtering out the segments that could correspond to lipid aggregates or other noise sources. Then, all segments were analysed individually for a circularity (C) check defined by $C=(P^2/4\pi A)$, where P is the perimeter length and A is the area of the segment. A 'true liposome' threshold was set for C values between 0.5 and 2.0. LacC2-mCherry aggregates were filtered out by

rejecting events whenever LactC2-mCherry intensities were higher than a pre-set cutoff inside the lumen. The segments that passed these extra filtering steps were considered liposomes and were saved individually as 60 by 60-pixel cropped images with a given ID number. For each individual liposome SMELDit measures the apparent radius, average Cy5 intensity and variance on the membrane, average LactC2-mCherry intensity and variance on the membrane, and average dsGreen intensity and variance on the lumen. Per sample, SMELDit displays all single-liposome extracted data in a scatter/histogram interactive GUI on which the user can draw regions of interest (ROI) for extracting information about liposome subpopulations. Here, ROI 1-4 were defined in SMELDit using negative controls for thresholding. Finally, once ROIs are drawn, example liposomes from the ROI are retrieved as a liposome-montage. ROI liposome data can be saved separately and ROI coordinates can also be saved and transferred to analyse another sample. The SMELDit MATLAB code is available upon request.

Quantitative PCR analysis

2 µL of liposome suspension was collected, incubated 15 min at 75 °C for DNAseI heat inactivation, and 100X diluted in MilliQ water. 10 µL qPCR reactions were prepared with 1X PowerUP SYBR Green Master Mix (Applied Biosystems), 400 nM of each forward and reverse primer (976/977 ChD for *p2*, 980/981 ChD for *p3*, 1125/1126 ChD for *pssA*, 1119/1120 for *plsB*, 1410/1411 ChD for *plsC*, 1408/1409 ChD for *cdsA*), and 1 µL of the diluted liposome sample. Solutions were transferred to a qPCR 96-well plate (Thermo Fisher) that were sealed with an adhesive transparent film (Thermo Fisher) and spined down for 15 seconds. Measurements were performed on a Quantstudio 5 Real-Time PCR instrument (Thermo Fisher) using the protocol: 2 min at 50 °C, 5 min at 94 °C, 45 cycles of 15 sec at 94 °C, 15 sec at 56 °C, 30 sec at 68 °C, 5 min at 68 °C, and a final melting curve stage from 65 °C to 95 °C. Sample DNA concentrations were calculated from standard curves generated using DNA templates of known concentrations ranging from 1 fM to 1 nM (9 µL of qPCR reaction + 1 µL of DNA). Data were further analysed with Quantstudio Design and Analysis software v1.4.3 Software (Thermo Fisher).

Recovery of DNArep-PLsyn DNA from liposomes

Liposome samples diluted 100 times in MilliQ water for qPCR measurement were also utilized also for DNA PCR recovery. Reactions were set up to either amplify three fragments or a single, near full-length, fragment from the *DNArep-PLsyn* genome. 20-50 µL reaction solutions were assembled in a 1X Xtreme Buffer with 300 nM of each primer (all primers and details about the corresponding PCR amplification targets can be found on **Table S2**). 2-5 µL of the diluted liposome solution, and 0.02 U/µL KOD DNA polymerase. The thermal cycler was programmed to follow 2 min at 94 °C for polymerase activation, and 30 cycles of (98 °C for 10 sec, 60 °C for 20 sec, 68 °C for 2 min for fragments A,B,C, and 5 min for one-fragment PCR-recovery). PCR products were analysed by agarose gel electrophoresis.

Bulk IVTT reactions

10 µL bulk IVTT reactions were performed with PURE*frex* 2.0 using 500 pM DNA template according to the manufacturer's guidelines. To visualize synthesized protein products, the reaction was supplemented with 1 µL of GreenLys:MilliQ water solution (FluoroTect GreenLys, Promega) and incubated for 16 hours at 37 °C. Samples were then supplemented with 1 µL of RNase A (4 mg/mL) and 1 µL of RNase One (10 U/µL), and incubated for 1-2 hours at 37 °C for complete RNA digestion. 10 µL of the RNA-digested sample were mixed with 1X Laemmli sample buffer and 10 mM of DTT, final concentrations. The reaction mixtures were incubated at 95 °C for 5 min and loaded on a 12% SDS-PAGE gel that was run for 1 hour at 100 V, followed by another 50-60 min at 130-160 V. GreenLys labelled proteins were visualized on a fluorescence gel imager (Typhoon, Amersham Biosciences) using Cy2 (488 nm), Cy3 (532 nm), and Cy5 (635 nm) lasers with band-pass filters of 515-535 nm for Cy2, 560-580 nm for Cy3, and 655-685 for Cy5. Laser PMT voltages were set to 500 for Cy2 and automatic adjustment for Cy3 and Cy5. The SDS-PAGE gel was further stained overnight with Instant Blue (expedeon), destained the next day with MilliQ water, and visualized with a ChemiDoc imaging system (Bio-Rad).

Analysis of lipid content by LC-MS

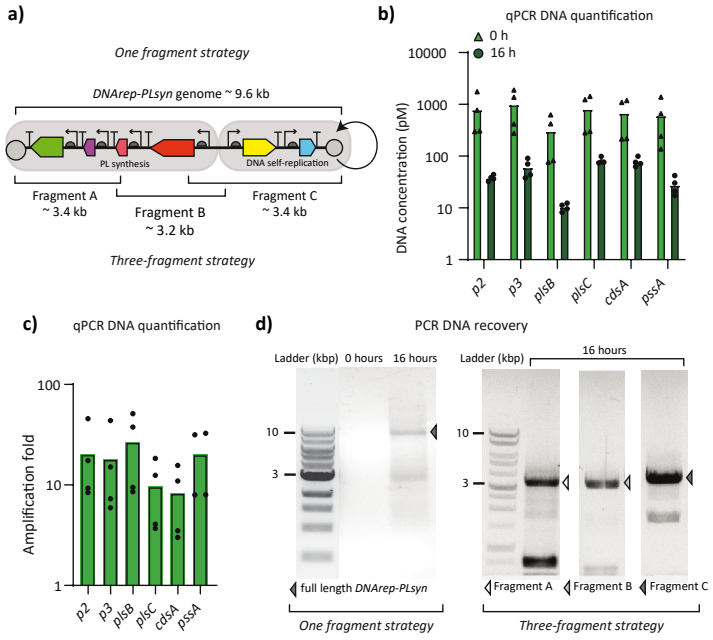
2-4 µL of liposome suspension were diluted 10X in a sample preparation solution consisting of a 1:1:98 ratio of 0.5 M EDTA : 200 mM acetylacetone : 100% methanol. The solution was sonicated for 10 min, and centrifuged at a maximum speed (Eppendorf Centrifuge 5415 R) and room temperature for 5-10 min. 15 µL of the supernatant with the soluble lipid phase were transferred to a 25 µL glass insert within a 2 mL LC-MS glass vial, further diluted 4X with the sample preparation solution, and stored under argon at −20 °C until use (less than a week). For LC-MS sample analysis, a ~10 µL sample, kept at 4 °C, were injected into a 6460 Triple Quad LC-MS stocked with a ACQUITY UPLC Peptide CSH C18 Column with a mobile phase A of 0.05% ammonium hydroxide and 2 mM acetylacetone in water, and a mobile phase B of % 2-propanol, 20% acetonitrile, 0.05% ammonium hydroxide and 2 mM acetylacetone, at a flow rate of 300 µL/min and column temperature of 60 °C. An A:B ratio of 70:30 was set to equilibrate the column. Upon sample injection, the A:B ratio was slowly changed to 100% mobile phase B and kept for two min. Then, the 70:30 ratio was gradually restored and kept until the end of each sample run. Transitions were established in our previous work [78]. Data were analysed with Skyline-daily. Peak areas were exported and normalized to DOPG peak areas. When two injections were done per sample, the averaged peak area was considered.

Acknowledgments

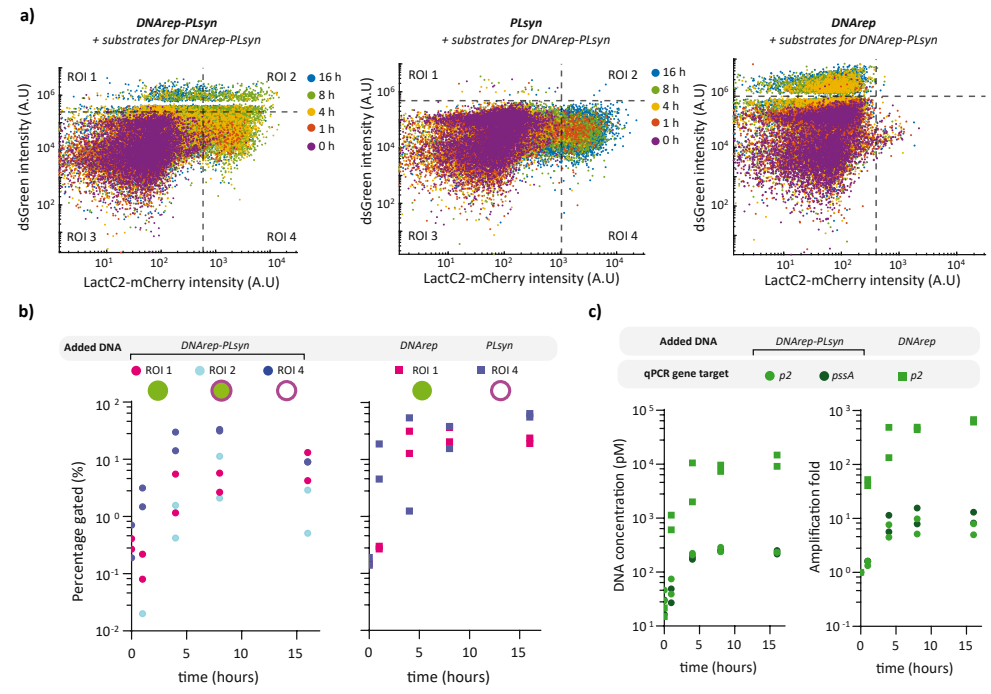
The authors are thankful to Miguel de Vega and Alicia del Prado from the Centro de Biología Molecular Severo Ochoa, Madrid, for generously providing the purified SSB and DSB proteins. They also thank Adja Zoumaro-Djayoon for her support on the LC-MS

lipidomic experiments, Ilja Westerlaken for her assistance in constructing *DNArep-PLsyn* genome batches, and to Laura Sierra for performing DNArep-PLsyn kinetic experiments. Funding for this project was provided by the Netherlands Organization for Scientific Research (NWO/OCW) through the "BaSyC, Building a Synthetic Cell" Gravitation grant (024.003.019).

Extended Figures



Extended Fig. 1 *DNArep-PLsyn* genome self-replicates in liposomes under joint-module reaction conditions at 34 °C. **a)** Illustration of the two DNA recovery strategies consisting of either three or one fragment that was PCR-amplified from *DNArep-PLsyn*. **b)** Absolute DNA quantification of *DNArep-PLsyn* containing liposomes at time zero (dark green) and after 16 hours (light green) of sample incubation. The gene names of the targeted regions (~200 bp) are indicated. **c)** DNA amplification fold calculated from the data shown in b. Each data point within a target gene represents a biological repeat. **d)** Agarose gel electrophoresis of PCR-recovered *DNArep-PLsyn* with both PCR recovery strategies as illustrated in panel (a).



Extended Fig. 2 In-vesiculo DNArep and PLsyn phenotype appearance along the course of gene expression. **a)** Superposed flow cytometry scatter plots from liposome samples at different incubation time points. The samples contained *DNArep-PLsyn*, *DNArep*, or *PLsyn* DNA templates and all DNArep-PLsyn substrates and cofactors. Additional repeats can be found in Fig. S13. **b)** Kinetics of ROIs gated liposomes for the three DNA template conditions. Data points from two biological repeats are displayed. **c)** DNA concentration kinetics from liposome samples containing either *DNArep-PLsyn* or *DNArep* DNA template. Targeted regions for qPCR were located on the *pssA* and *p2* genes of *DNArep-PLsyn* genome, and on *p2* gene on *DNArep* template. DNA amplification fold at different time points was calculated by dividing the concentration of DNA to that at time zero.

Supplementary information

Plasmid name	Plasmid description
G363	Plasmid encoding for DNArep and PLsyn proteins (DNAP, TP, PlsB, PlsC, CdsA, PssA). All genes are present as individual transcriptional cassettes under the control of a SP6 promoter (DNArep machinery) and T7 promoter (PLsyn machinery), ribosomal binding sites, and T7 transcription terminators. The combined transcriptional units are flanked by Φ29 origin of replication sequences oriL and oriR as oriL-transcriptional units-oriR [21]. DNA sequence is available upon request.
G435	Plasmid encoding for DNArep proteins (DNAP and TP). Both genes are present as individual transcriptional cassettes, under the control of a T7 promoter, ribosomal binding sites, and T7, vsv-r1 and vsv-r2 transcription terminators (for DNAP and TP, respectively). The combined transcriptional units are flanked by Φ29 origin of replication sequences oriL and oriR as oriL-transcriptional units-oriR [21]. DNA sequence is available upon request.
G*555*	Plasmid encoding for PLsyn proteins (PlsB, PlsC, CdsA, PssA). All four genes are present as individual transcriptional cassettes, under the control of a T7 promoter, ribosomal binding sites, and T7 transcription terminators. The combined transcriptional units are flanked by Φ29 origin of replication sequences oriL and oriR as oriL-transcriptional units-oriR [21]. DNA sequence is available upon request.

Table S1. List of plasmids.

Primer pair	DNA sequence (*5'* → *3'*)	Purpose
491 ChD **1302 ChD**	P-AAAGTAAGCCCCCACCCTCACATG GTATTAATTTCACATGCGACAGAATTCGCGG CCGCTTCTAG	PCR amplicon to assemble *DNArep-PLsyn* genome ($PLsyn_{frag}$). DNA template: G363
1289 ChD **492 ChD**	GCGAATTCTGTCGCATGTGAAATTAATACGA CTCACTATAGGGA P-AAAGTAGGGTACAGCGACAACATACAC	PCR amplicon to assemble *DNArep-PLsyn* ($DNArep_{frag}$). DNA template: G435
491 ChD **492 ChD**	P-AAAGTAAGCCCCCACCCTCACATG P-AAAGTAGGGTACAGCGACAACATACAC	PCR to produce *DNArep* and *PLsyn* DNA templates. DNA templates: G435 (for *DNArep*), and G*555* (for *PLsyn*).
1459 ChD **1460 ChD**	CTCCTAATATCGACATAATCCGTCGATCCTCG CCCCATTGACCGACTATCTTCGACAAG	PCR amplicon for one-fragment *DNArep-PLsyn* DNA recovery. 32 bp away from oriR and oriL.
1378 ChD **1410 ChD**	AAAGTAAGCCCCCACCCTCACATG CGACAGAAACAATCGCACTAAAG	Fragment A for three-fragment *DNArep-PLsyn* DNA recovery.
1411 ChD **1290 ChD**	CGGGAACATCCAGATGGAAATA ATGCGACAGAATTCGCGGCCGCTTC	Fragment B for three-fragment *DNArep-PLsyn* DNA recovery.
756 ChD **424 ChD**	ACTTCGCCTTTTTCACGCCC AAAGTAGGGTACAGCGACAACATACAC	Fragment C for three-fragment *DNArep-PLsyn* DNA recovery.
976 ChD **977 ChD**	GGATGAAGACTACCCGCTGC ACAGGTCTGCGATTTCACCG	qPCR amplicon of targeted region in *p2* gene.
980 ChD **981 ChD**	ACGGCTGAAATTGACATCCCG CCAGGCGTTGAACTTCTTTGG	qPCR amplicon of targeted region in *p3* gene.

1125 ChD **1126 ChD**	AACAGGATGACGGTGGCAAA GGAACATCTACGCCCGGATT	qPCR amplicon of targeted region in *pssA* gene.
1119 ChD **1120 ChD**	TCTCCCGCGACGTATTGATG AATAACGTCCGGCAACTCGT	qPCR amplicon of targeted region in *plsB* gene.
1410 ChD **1411 ChD**	GGATGAAGACTACCCGCTGC ACAGGTCTGCGATTTCACCG	qPCR amplicon of targeted region in *plsC* gene.
1408 ChD **1409 ChD**	CGACAGAAACAATCGCACTAAAG CGGGAACATCCAGATGGAAATA	qPCR amplicon of targeted region in *cdsA* gene.

Table S2. List of primers used for PCR and qPCR

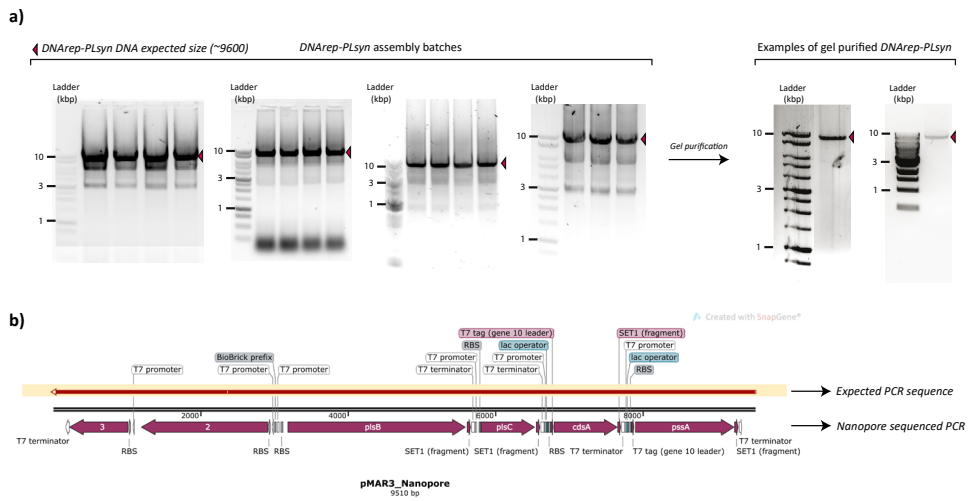


Fig. S1 ***DNArep-PLsyn*** **genome assembly by overlap PCR. a)** Agarose gel electrophoresis of different *DNArep-PLsyn* genome assembly batches with (right) and without (left) gel purification. **b)** DNA sequence alignment (from SnapGene) between expected *DNArep-PLsyn* DNA sequence and a nanopore sequenced *DNArep-PLsyn* assembly batch (www.plasmidsaurus.com).

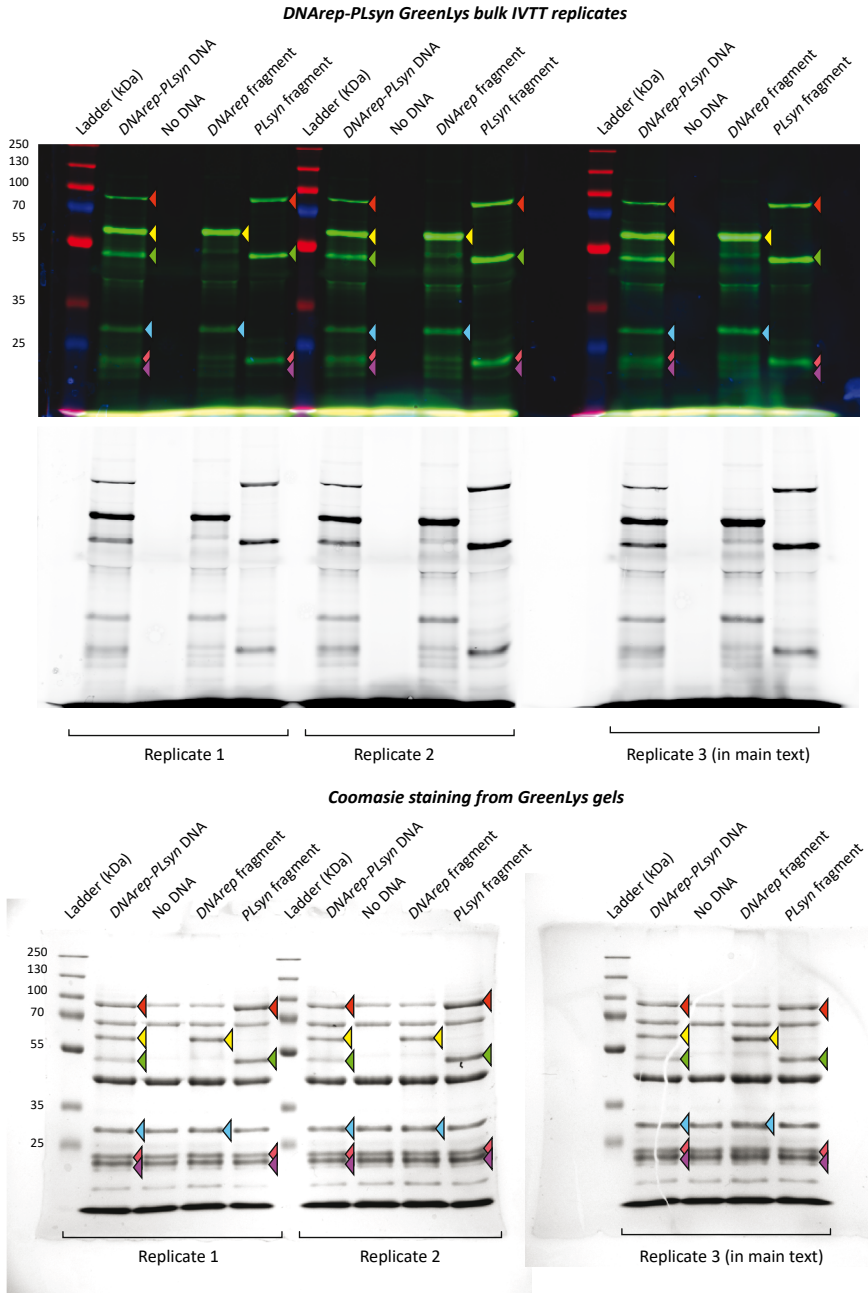
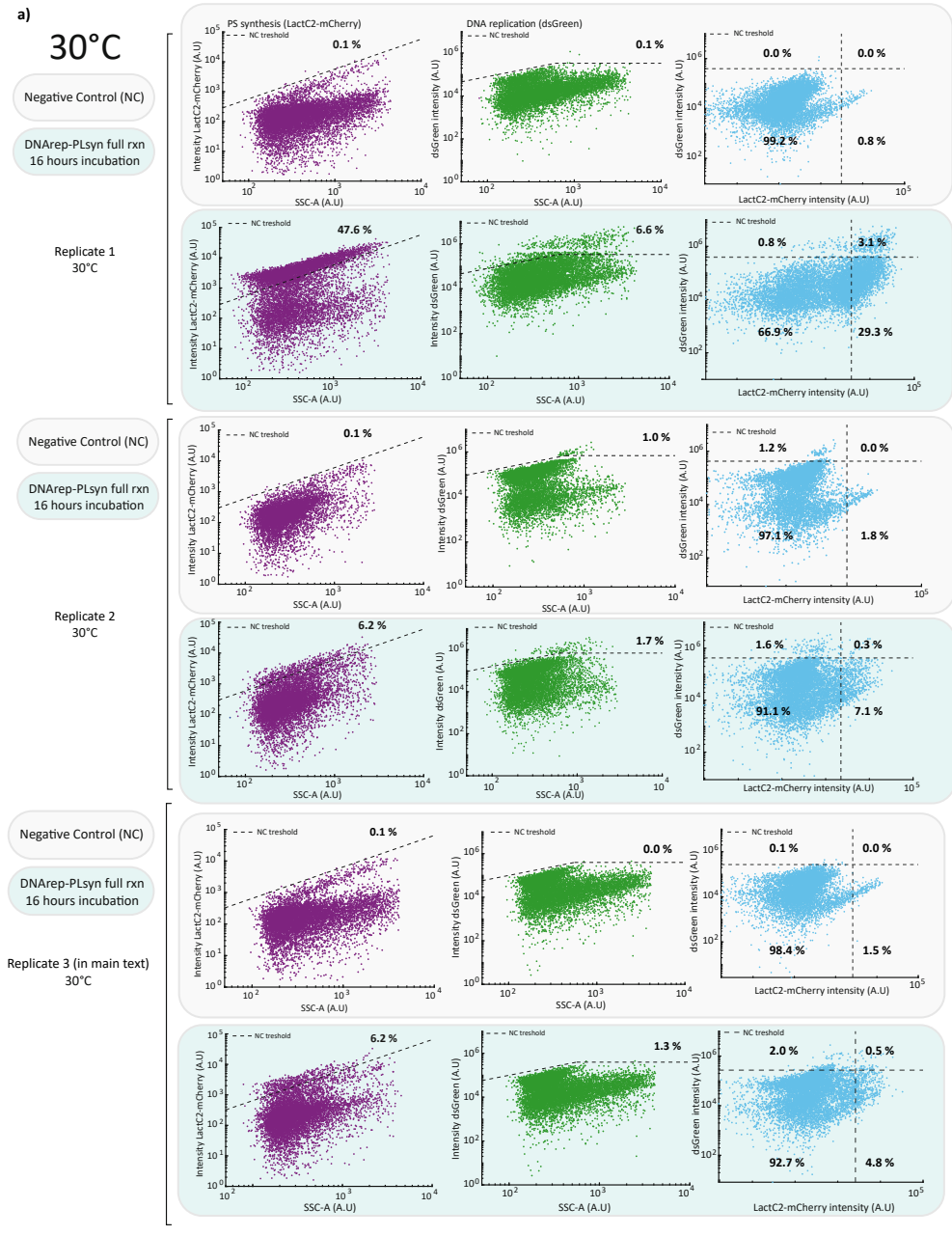


Fig. S2 Biological repeats of bulk IVTT protein production with *DNArep-PLsyn*, *DNA rep* fragment and *PL synth* fragment as DNA templates (Fig. 1f). SDS-PAGE gels with Green-Lys protein labelling (upper gel), and Coomassie protein staining (bottom gel).

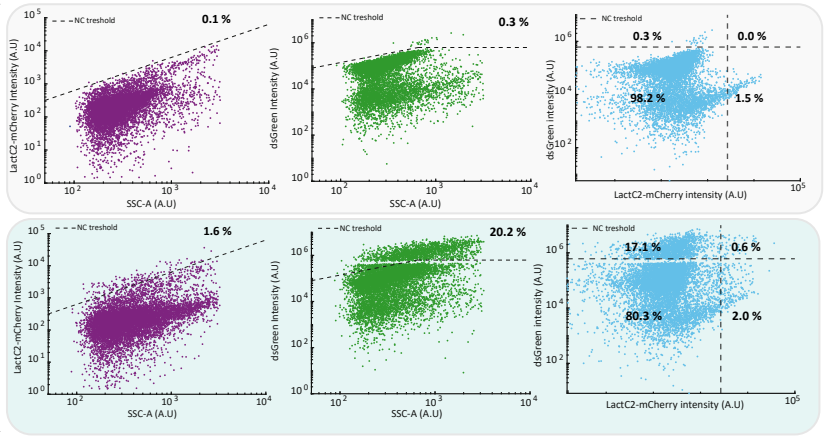


34°C

Negative Control (NC)

DNArep-PLsyn full rxn 16 hours incubation

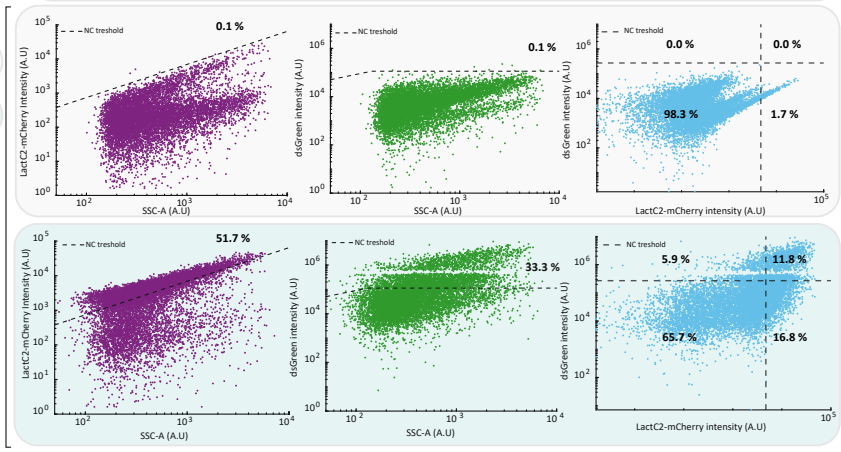
Replicate 1 34°C



Negative Control (NC)

DNArep-PLsyn full rxn 16 hours incubation

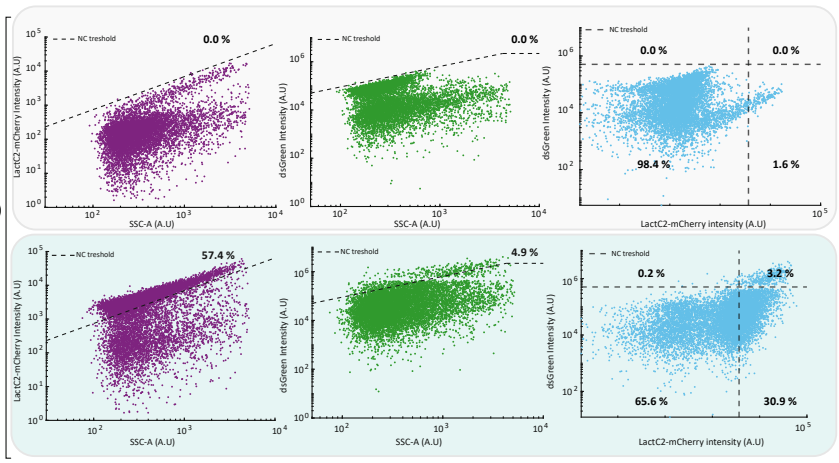
Replicate 2 34°C



Negative Control (NC)

DNArep-PLsyn full rxn 16 hours incubation

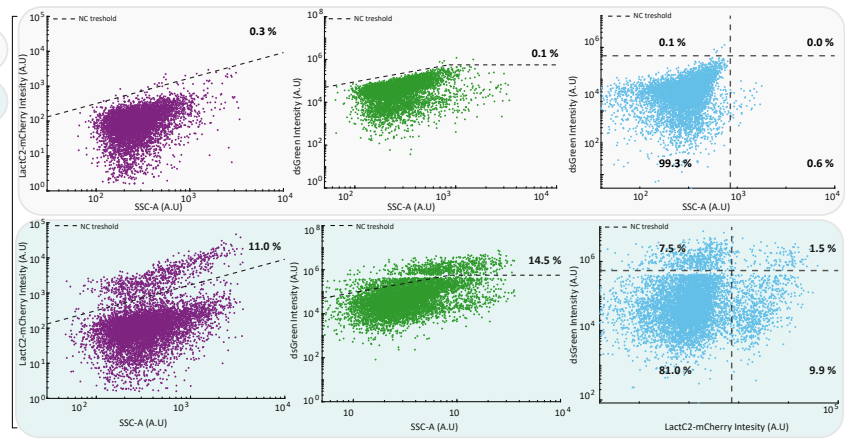
Replicate 3 (in main text) 34°C



Negative Control (NC)

DNArep-PLsyn full rxn 16 hours incubation

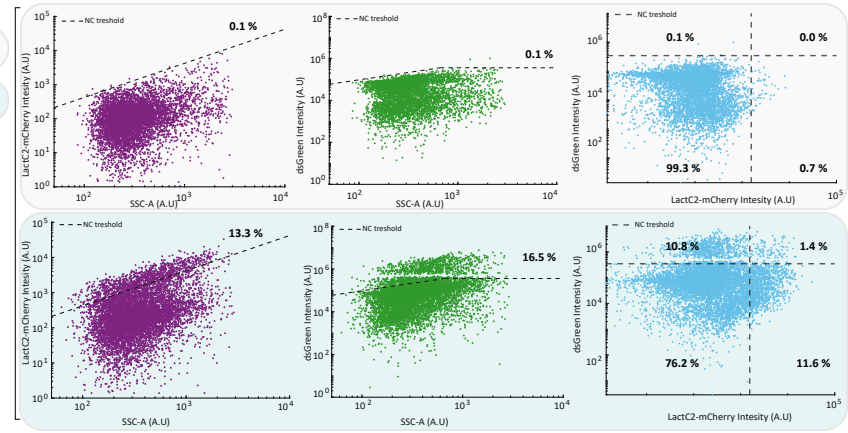
Replicate 5 34°C



Negative Control (NC)

DNArep-PLsyn full rxn 16 hours incubation

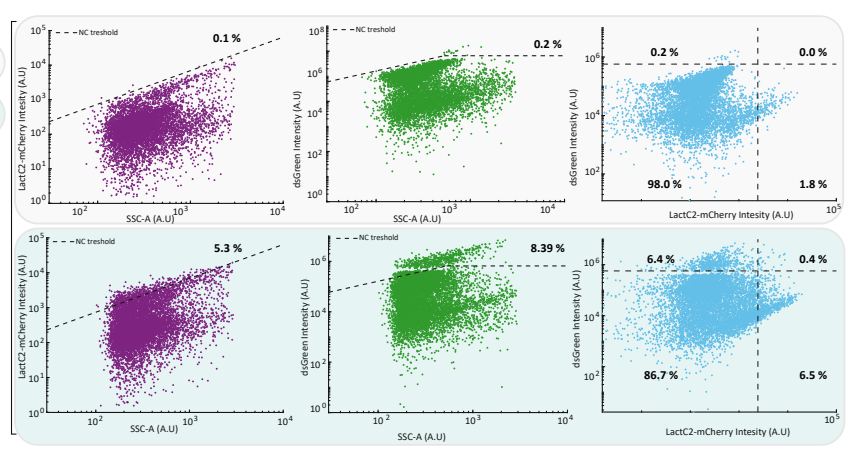
Replicate 4 34°C



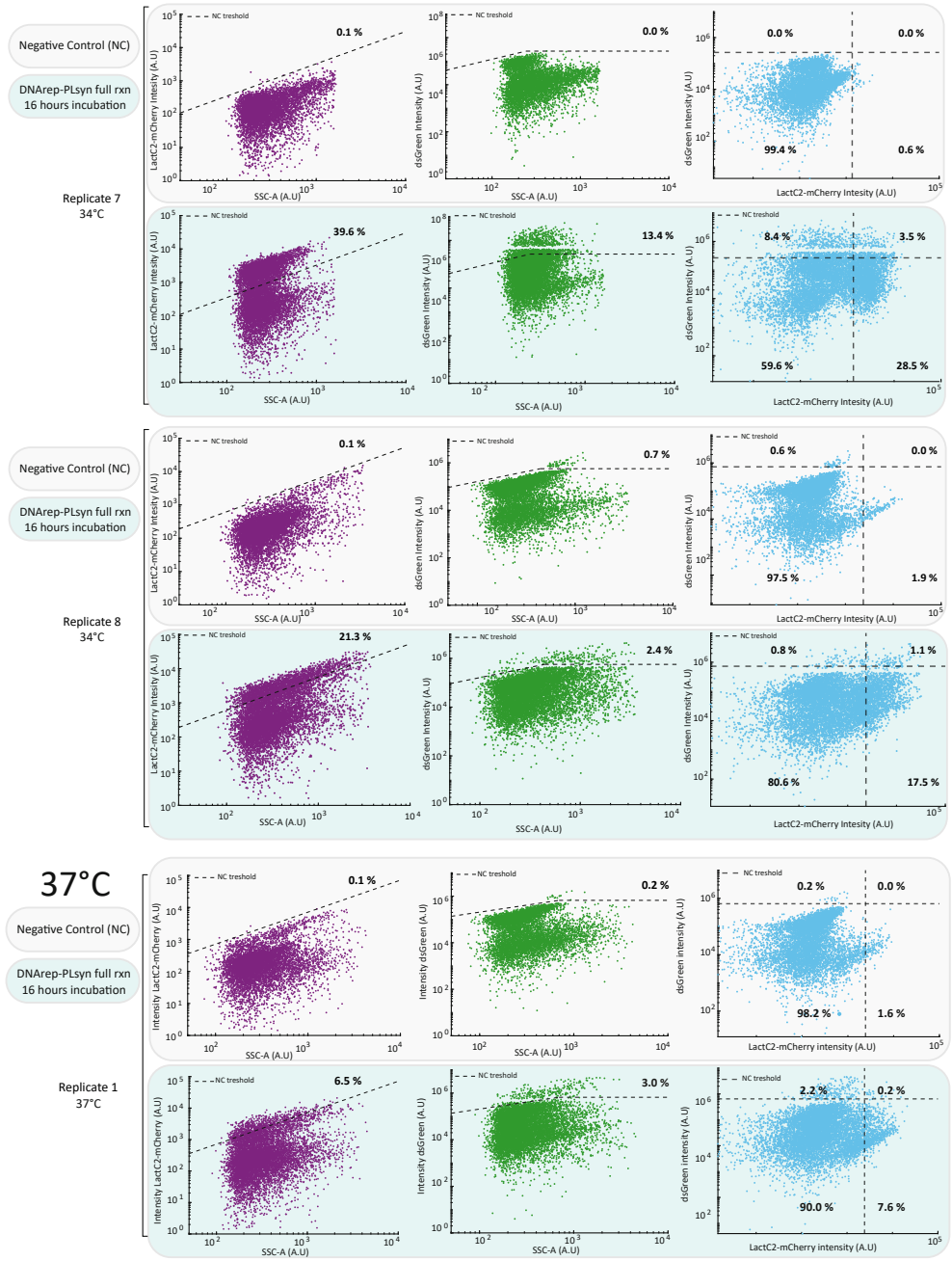
Negative Control (NC)

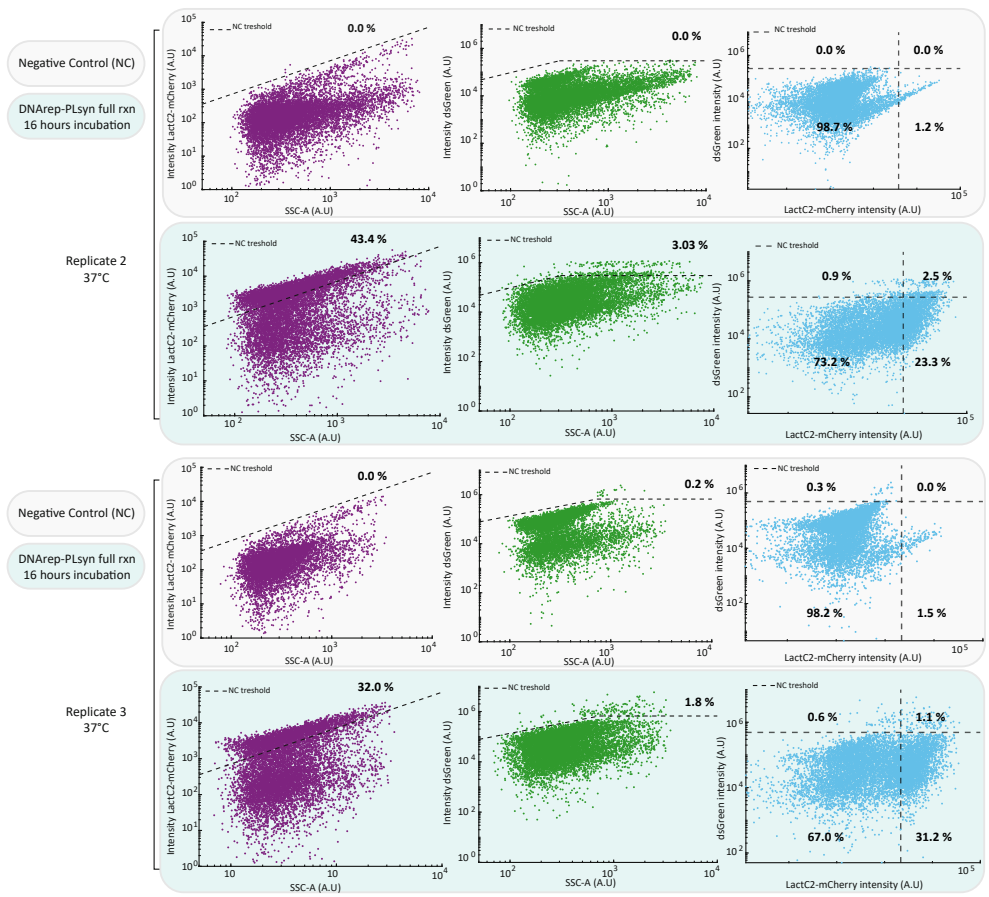
DNArep-PLsyn full rxn 16 hours incubation

Replicate 6 34°C



4





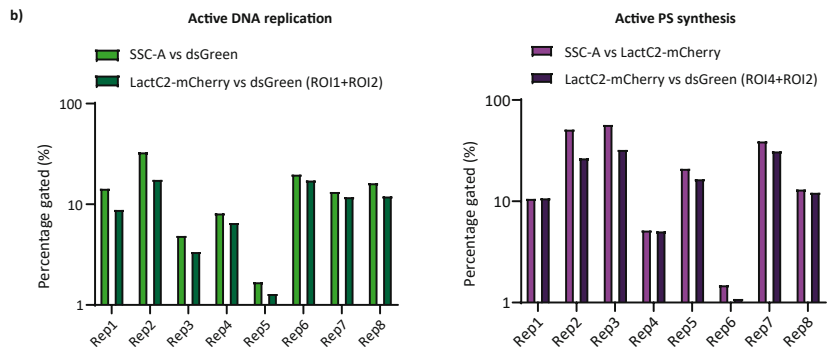


Fig. S3 a) Flow cytometry scatter plots from all the biological repeats of DNArep-PLsyn activity assays shown in Fig. 2. Samples were incubated at 30, 34, or 37 °C, as indicated. **b)** Bar plot representation of the percentage of gated liposomes calculated from the SSC-A and fluorescence signals, as specified in the legend, for 8 biological replicates. DNA-rep- and PLsyn-active liposomes were classified based on intensity thresholds as described in the Methods section.

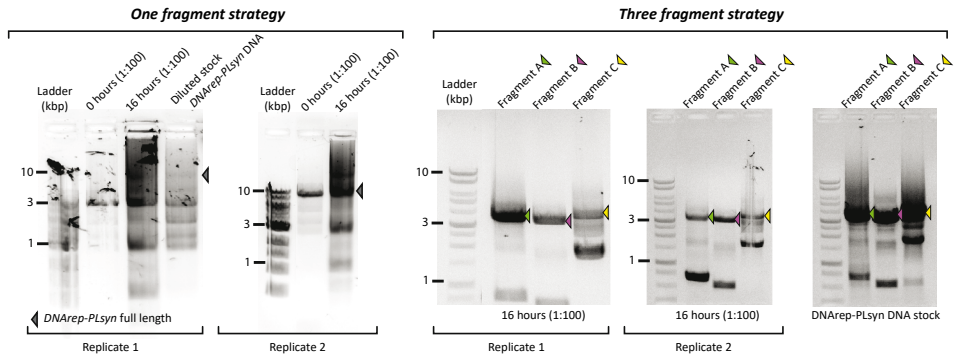
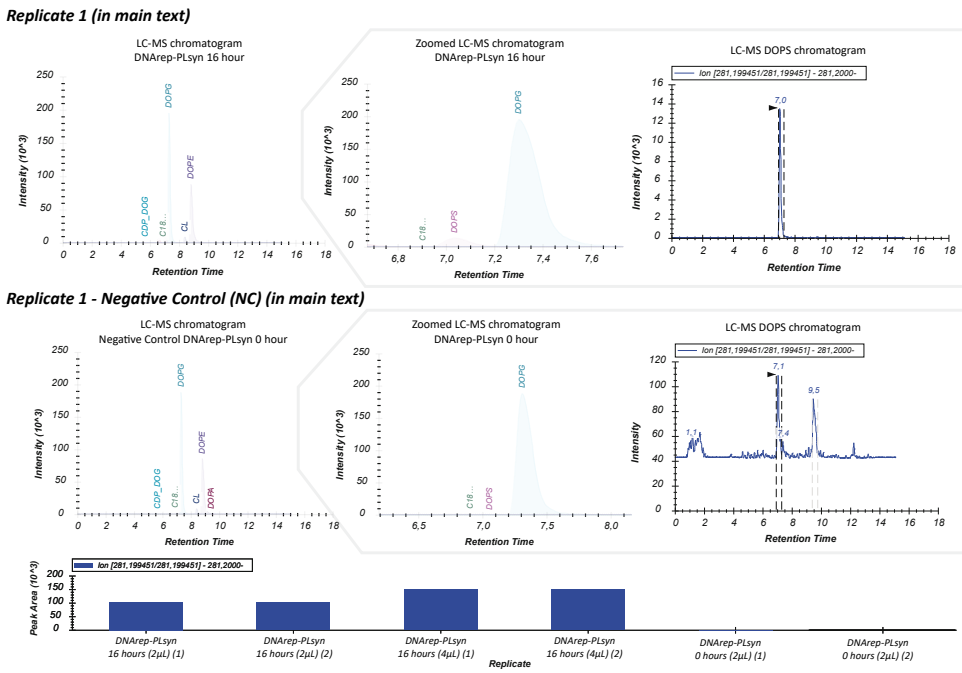


Fig. S4 Agarose gel electrophoresis of PCR-recovered DNA from the *DNArep-PLsyn* template isolated from liposome samples. A one- or three-fragment (A, B, C) recovery strategy was used, as indicated.



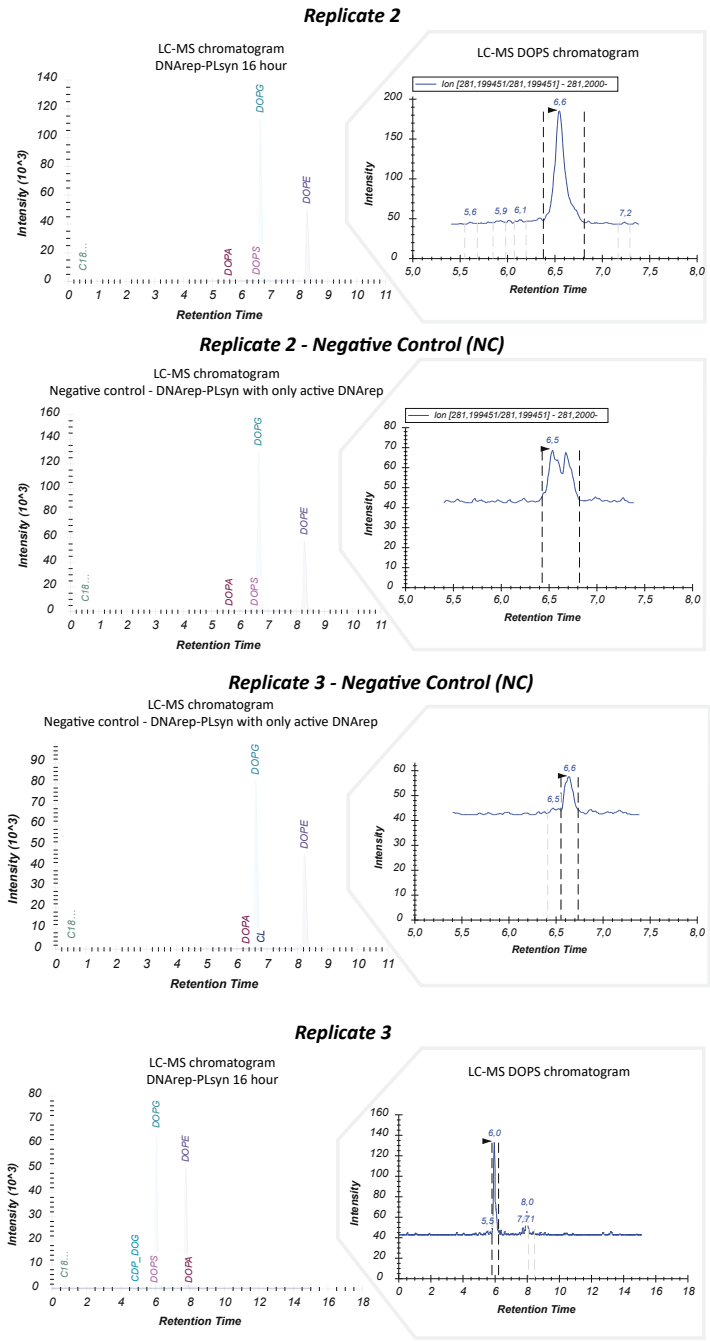


Fig. S5 LC-MS chromatograms of the lipid content from liposome samples with expressed DNArep-PLsyn protein machinery, and full set of substrates and cofactors. All biological repeats show clear PS intensity peaks (inserts) when compared to the negative controls (time zero). Intensity values are in arbitrary unit. For biological replicate 1, DOPS peak areas from technical replicates are represented as bar plots.

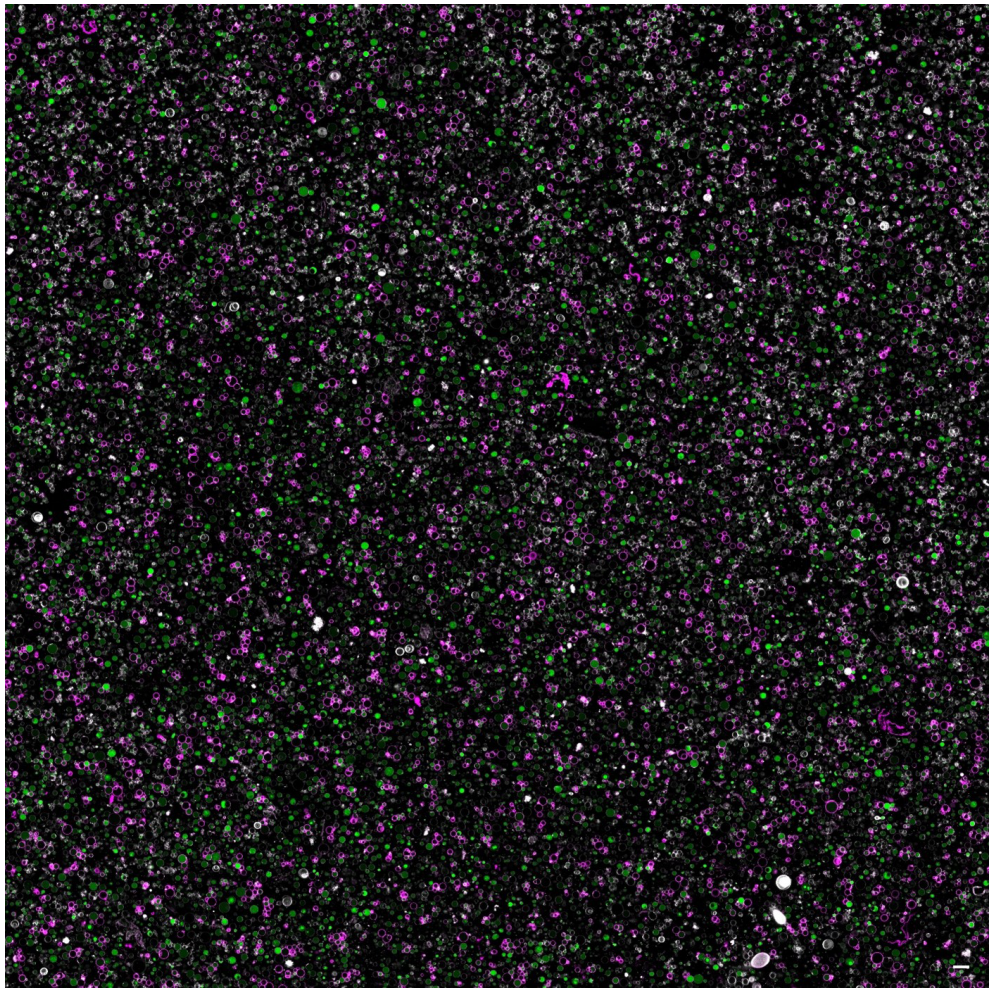


Fig. S6 Large field-of-view confocal image (~325x325 µm) from a liposome population expressing *DNArep-PLsyn* genome with the full set of substrates and cofactors for module activation. White, Cy5 membrane dye; magenta, LactC2-mCherry; green, dsGreen. The scale bar indicates 5 µm.

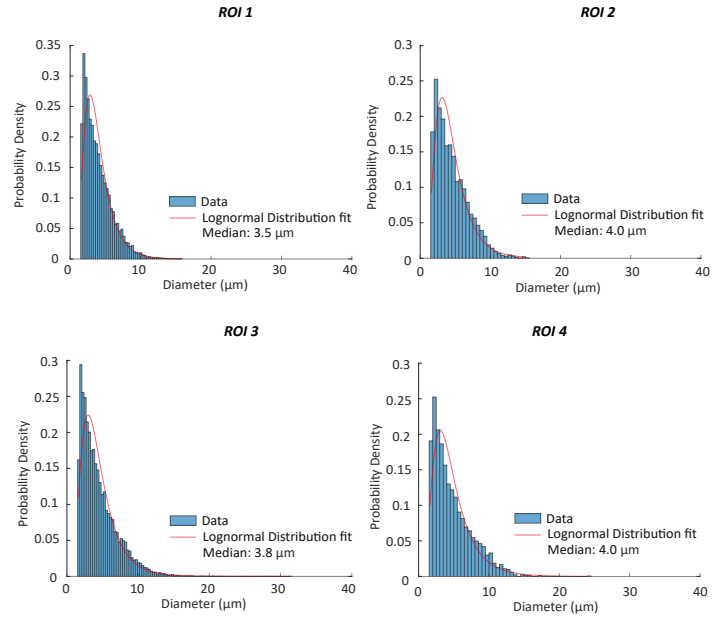
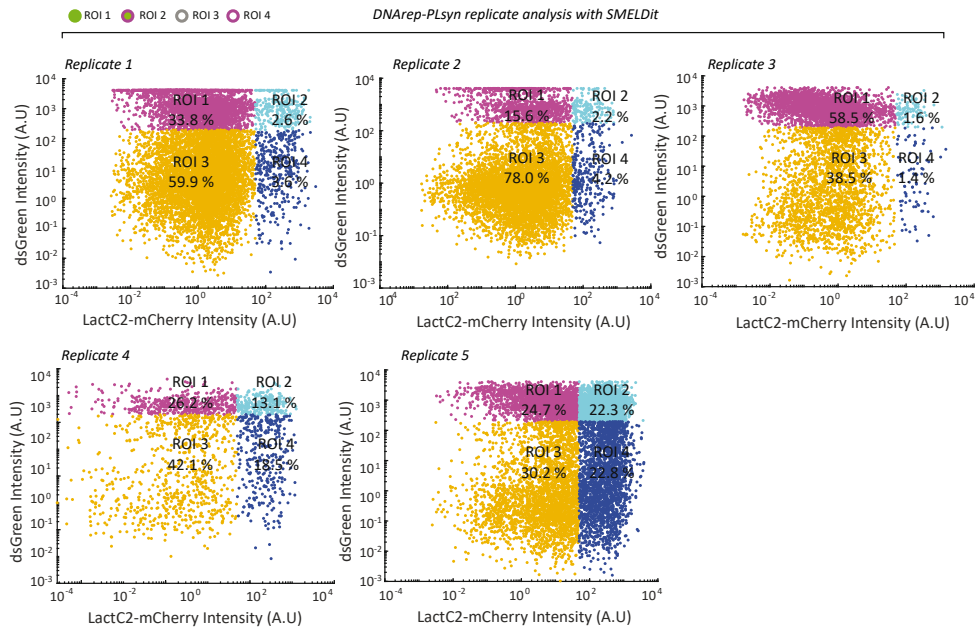


Fig. S7 Vesicle size distribution calculated for each ROI from *DNArep-PLsyn*-expressing liposomes. Data from multiple replicate samples were pooled for the analysis. The red curve is the lognormal distribution fit. The median liposome diameter value for each ROI is appended.



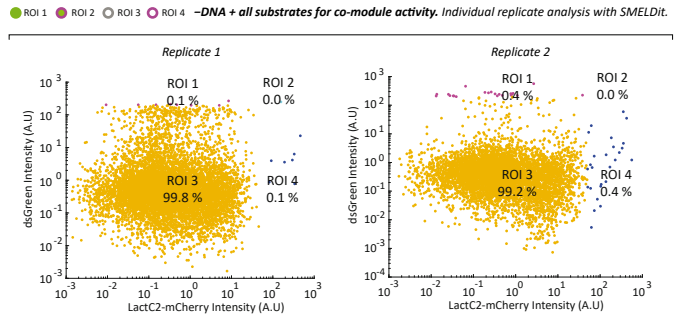


Fig. S8 Phenotype scatter plots from SMELDit image analysis (LactC2-mCherry vs. dsGreen) of liposome populations expressing the *DNArep-PLsyn* genome (negative controls with no DNA), in the presence of all substrates and cofactors for dual module activity. Individual biological repeats from pooled data shown in Fig. 3. Displayed ROI percentages were calculated for each replicate. Negative control samples together with experiments shown in Fig. 4 and 5 were used to define the intensity thresholds for classification into four ROIs.

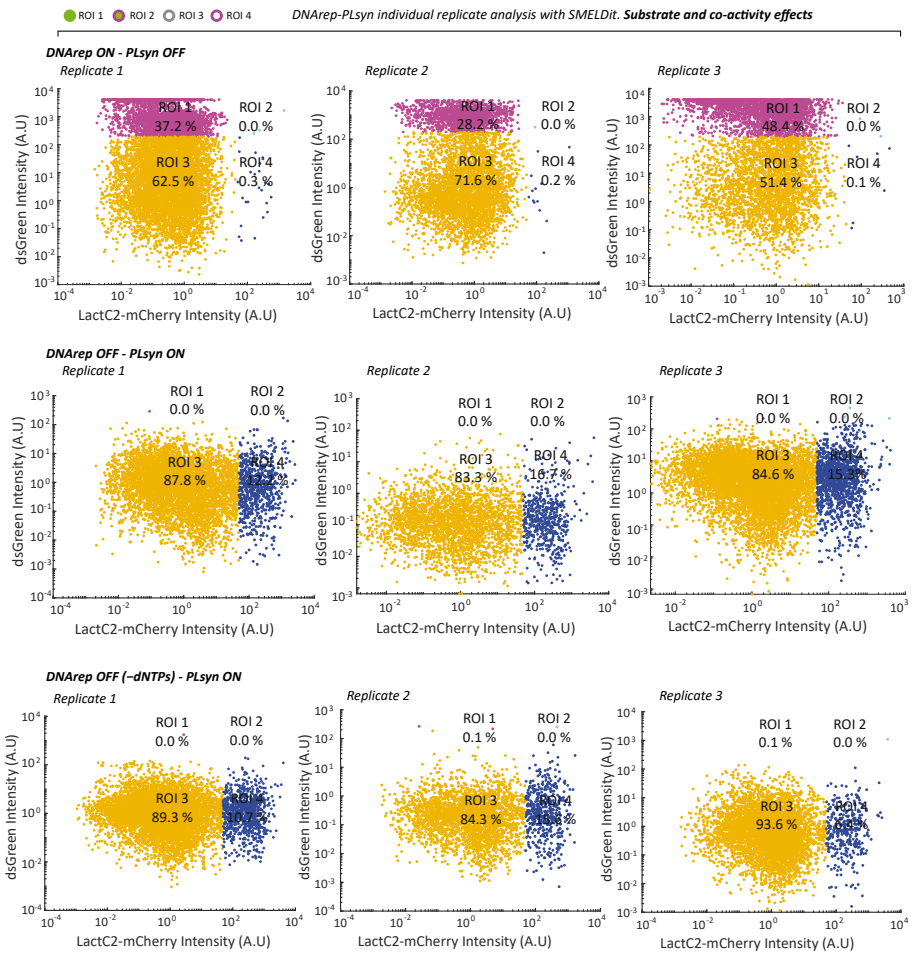


Fig. S9 Phenotype scatter plots from SMELDit image analysis (LactC2-mCherry vs. dsGreen) of liposome populations expressing the *DNArep-PLsyn* genome in the presence of substrates and cofactors to activate only DNArep (ON) or only PLsyn (ON). Individual biological repeats from pooled data shown in Fig. 4. Displayed ROI percentages were calculated for each replicate.

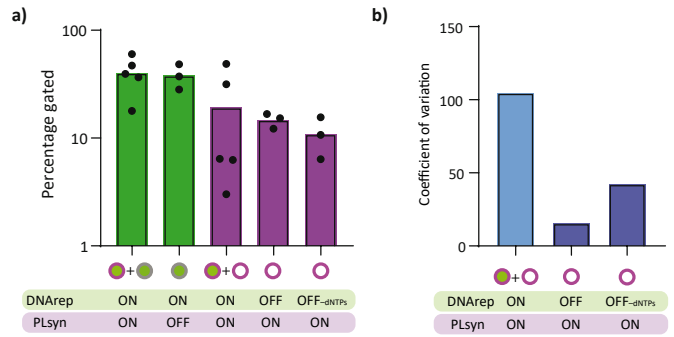


Fig. S10 a) Percentage of liposomes expressing the *DNArep-PLsyn* genome with an active DNArep or/and PLsyn module. Data points for each activation condition are from individual biological repeats. **b)** Coefficient of variation calculated from the data shown in a for PLsyn-active conditions.

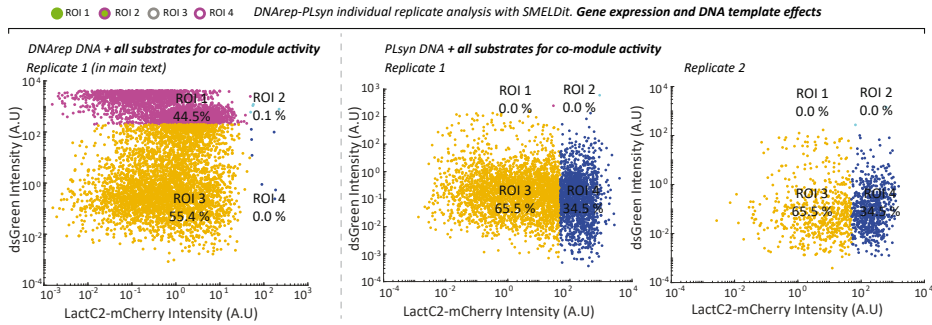


Fig. S11 Phenotype scatter plots from SMELDit image analysis (LactC2-mCherry vs. dsGreen) of liposome populations expressing either *DNArep-PLsyn*, *DNArep*, or *PLsyn* DNA in the presence of all substrates and cofactors. Individual biological repeats from pooled data shown in Fig. 5. Displayed ROI percentages were calculated for each replicate.

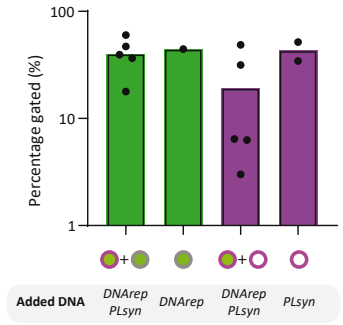


Fig. S12 Percentage of liposomes expressing either *DNArep-PLsyn*, *DNArep*, or *PLsyn* DNA in the presence of all substrates and cofactors. Data points for each condition are from individual biological repeats.

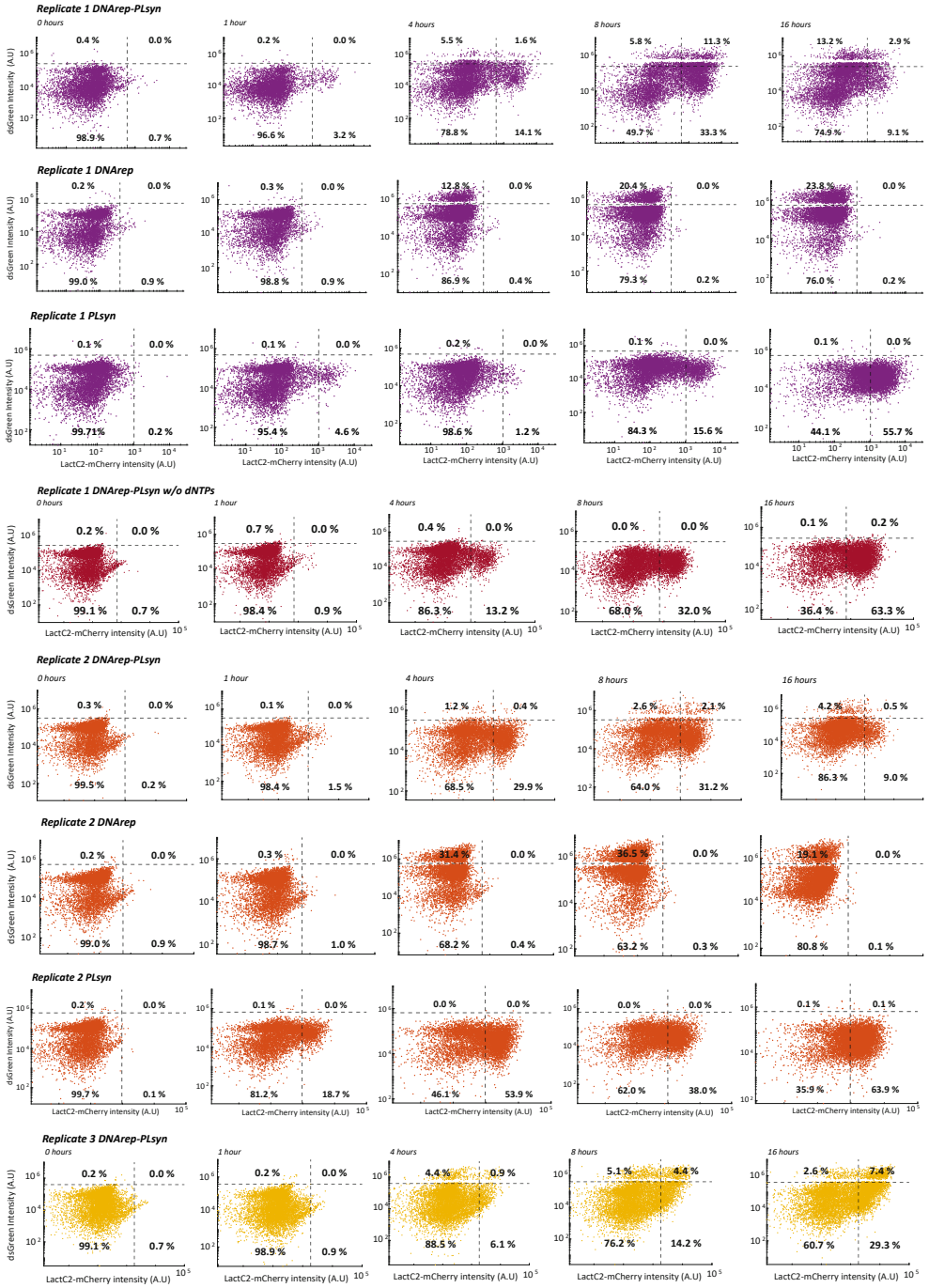


Fig. S13 Flow cytometry scatter plots (LactC2-mCherry vs. dsGreen) of liposome samples analysed at different incubation times for expression of *DNArep-PLsyn*, *DNArep* or *PLsyn* DNA under full substrates/cofactors condition. Data from individual replicate experiments are shown.

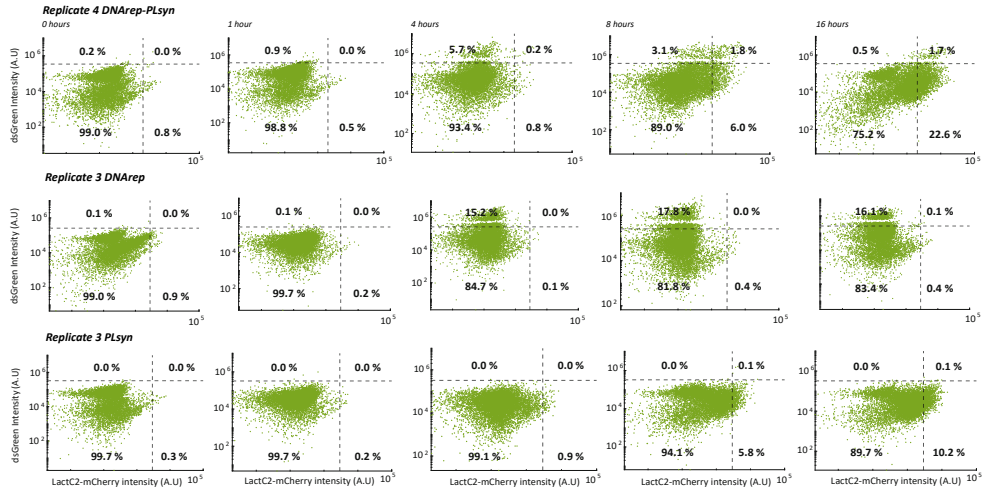


Fig. S14 Flow cytometry scatter plots (LactC2-mCherry vs. dsGreen) of liposome samples analysed at different incubation times for expression of *DNArep-PLsyn*, *DNArep* or *PLsyn* DNA under full substrates/cofactors condition. Contrary to Fig. S13, these experiments have been conducted following a slightly different protocol for sample incubation and collection. Here, reactions were incubated in the same tube over all the incubation times. The starting volume of the liposome solution was set to ~10 µL from which three samples of ~2 µL were collected after 1, 4, and 8 hours. The remaining solution was utilized for the 16-hour measurement.

References

1. Buddingh', B. C. & van Hest, J. C. M. Artificial Cells: Synthetic Compartments with Life-like Functionality and Adaptivity. Acc Chem Res *50*, 769–777 (2017).
2. Supramaniam, P., Ces, O. & Salehi-Reyhani, A. Microfluidics for Artificial Life: Techniques for Bottom-Up Synthetic Biology. Micromachines (Basel) 10, 299 (2019).
3. Guindani, C., da Silva, L. C., Cao, S., Ivanov, T. & Landfester, K. Synthetic Cells: From Simple Bio-Inspired Modules to Sophisticated Integrated Systems. Angewandte Chemie International Edition 61, e202110855 (2022).
4. Gaut, N. J. & Adamala, K. P. Reconstituting Natural Cell Elements in Synthetic Cells. Adv Biol (Weinh) *5*, e2000188 (2021).
5. Hirschi, S., Ward, T. R., Meier, W. P., Müller, D. J. & Fotiadis, D. Synthetic Biology: Bottom-Up Assembly of Molecular Systems. Chem Rev 122, 16294–16328 (2022).
6. Yue, K. et al. Bottom-Up Synthetic Biology Using Cell-Free Protein Synthesis. Adv Biochem Eng Biotechnol 185, 1–20 (2023).
7. Schmitt, C., Lippert, A. H., Bonakdar, N., Sandoghdar, V. & Voll, L. M. Compartmentalization and Transport in Synthetic Vesicles. Frontiers in Bioengineering and Biotechnology 4, (2016).
8. Vogele, K. et al. Towards synthetic cells using peptide-based reaction compartments. Nat Commun 9, 3862 (2018).
9. Exterkate, M., Caforio, A., Stuart, M. C. A. & Driessen, A. J. M. Growing Membranes In Vitro by Continuous Phospholipid Biosynthesis from Free Fatty Acids. ACS Synth. Biol. 7, 153–165 (2018).
10. Bhattacharya, A., Brea, R. J., Niederholtmeyer, H. & Devaraj, N. K. A minimal biochemical route towards de novo formation of synthetic phospholipid membranes. Nat Commun 10, 300 (2019).
11. Blanken, D., Foschepoth, D., Serrão, A. C. & Danelon, C. Genetically controlled membrane synthesis in liposomes. Nat Commun 11, 4317 (2020).
12. Eto, S. et al. Phospholipid synthesis inside phospholipid membrane vesicles. Commun Biol *5*, 1–11 (2022).
13. Gonzales, D. T., Suraritdechachai, S. & Tang, T.-Y. D. Compartmentalized Cell-Free Expression Systems for Building

Synthetic Cells. Adv Biochem Eng Biotechnol 186, 77–101 (2023).
14. van de Cauter, L., van Buren, L., Koenderink, G. H. & Ganzinger, K. A. Exploring Giant Unilamellar Vesicle Production for Artificial Cells - Current Challenges and Future Directions. Small Methods e2300416 (2023) doi:10.1002/smtd.202300416.
15. Godino, E. et al. De novo synthesized Min proteins drive oscillatory liposome deformation and regulate FtsA-FtsZ cytoskeletal patterns. Nat Commun 10, 4969 (2019).
16. Kretschmer, S., Ganzinger, K. A., Franquelim, H. G. & Schwille, P. Synthetic cell division via membrane-transforming molecular assemblies. BMC Biology 17, 43 (2019).
17. Godino, E. et al. Cell-free biogenesis of bacterial division proto-rings that can constrict liposomes. Commun Biol 3, 1–11 (2020).
18. Baldauf, L., van Buren, L., Fanalista, F. & Koenderink, G. H. Actomyosin-Driven Division of a Synthetic Cell. ACS Synth Biol 11, 3120–3133 (2022).
19. Rubio-Sánchez, R., Mognetti, B. M., Cicuta, P. & Di Michele, L. DNA-Origami Line-Actants Control Domain Organization and Fission in Synthetic Membranes. J. Am. Chem. Soc. 145, 11265–11275 (2023).
20. Sakatani, Y., Yomo, T. & Ichihashi, N. Self-replication of circular DNA by a self-encoded DNA polymerase through rolling-circle replication and recombination. Sci Rep 8, 13089 (2018).
21. van Nies, P. et al. Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. Nat Commun 9, 1–12 (2018).
22. Okauchi, H. & Ichihashi, N. Continuous Cell-Free Replication and Evolution of Artificial Genomic DNA in a Compartmentalized Gene Expression System. ACS Synth. Biol. 10, 3507–3517 (2021).
23. Libicher, K., Hornberger, R., Heymann, M. & Mutschler, H. In vitro self-replication and multicistronic expression of large synthetic genomes. Nat Commun 11, 904 (2020).
24. De Capitani, J. & Mutschler, H. The Long Road to a Synthetic Self-Replicating Central Dogma. Biochemistry 62, 1221–1232 (2023).
25. Lavickova, B., Laohakunakorn, N. & Maerkl, S. J. A partially self-regenerating synthetic cell. Nat Commun 11, 6340 (2020).
26. Smith, J. M., Chowdhry, R. & Booth, M. J. Controlling Synthetic Cell-Cell Communication. Frontiers in Molecular Biosciences 8, (2022).
27. Caschera, F. & Noireaux, V. Integration of biological parts toward the synthesis of a minimal cell. Current Opinion in Chemical Biology 22, 85–91 (2014).
28. Göpfrich, K., Platzman, I. & Spatz, J. P. Mastering Complexity: Towards Bottom-up Construction of Multifunctional Eukaryotic Synthetic Cells. Trends Biotechnol 36, 938–951 (2018).
29. Gánti, T. Organization of chemical reactions into dividing and metabolizing units: The chemotons. Biosystems 7, 15–21 (1975).
30. Balakrishnan, R., de Silva, R. T., Hwa, T. & Cremer, J. Suboptimal resource allocation in changing environments constrains response and growth in bacteria. Mol Syst Biol 17, e10597 (2021).
31. Burby, P. E. & Simmons, L. A. Regulation of Cell Division in Bacteria by Monitoring Genome Integrity and DNA Replication Status. J Bacteriol 202, e00408-19 (2020).
32. Mukherjee, A. et al. Plasticity of growth laws tunes resource allocation strategies in bacteria. bioRxiv 2023.08.22.554312 (2023) doi:10.1101/2023.08.22.554312.
33. Abil, Z. & Danelon, C. Roadmap to Building a Cell: An Evolutionary Approach. Frontiers in Bioengineering and Biotechnology 8, (2020).
34. Bailoni, E. & Poolman, B. ATP Recycling Fuels Sustainable Glycerol 3-Phosphate Formation in Synthetic Cells Fed by Dynamic Dialysis. ACS Synth. Biol. 11, 2348–2360 (2022).
35. Bailoni, E. et al. Minimal Out-of-Equilibrium Metabolism for Synthetic Cells: A Membrane Perspective. ACS Synth Biol 12, 922–946 (2023).
36. Seo, K. & Ichihashi, N. Investigation of Compatibility between DNA Replication, Transcription, and Translation for

in Vitro Central Dogma. ACS Synth. Biol. (2023) doi:10.1021/acssynbio.3c00130.
37. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. Nat Biotechnol 19, 751–755 (2001).
38. Blanco, L. & Salas, M. Replication of phage phi 29 DNA with purified terminal protein and DNA polymerase: synthesis of full-length phi 29 DNA. Proc Natl Acad Sci U S A 82, 6404–6408 (1985).
39. Godiska, R. et al. Linear plasmid vector for cloning of repetitive or unstable sequences in Escherichia coli. Nucleic Acids Research 38, e88 (2010).
40. Song, J., Dong, F., Lilly, J. W., Stupar, R. M. & Jiang, J. Instability of bacterial artificial chromosome (BAC) clones containing tandemly repeated DNA sequences. Genome 44, 463–469 (2001).
41. St, L., Tj, G., Pj, S., Va, S. & Pt, D. Recombination between repeats in Escherichia coli by a recA-independent, proximity-sensitive mechanism. Molecular & general genetics : MGG 245, (1994).
42. Kobs, G. FluoroTectTM GreenLys in vitro Translation Labeling System.
43. Salas, M., Holguera, I., Redrejo-Rodríguez, M. & de Vega, M. DNA-Binding Proteins Essential for Protein-Primed Bacteriophage ϕ29 DNA Replication. Front Mol Biosci 3, 37 (2016).
44. Doerr, A. et al. Modelling cell-free RNA and protein synthesis with minimal systems. Phys. Biol. 16, 025001 (2019).
45. Godino, E., Restrepo Sierra, A. M. & Danelon, C. Imaging Flow Cytometry for High-Throughput Phenotyping of Synthetic Cells. ACS Synth. Biol. 12, 2015–2028 (2023).
46. Blanken, D., van Nies, P. & Danelon, C. Quantitative imaging of gene-expressing liposomes reveals rare favorable phenotypes. Phys. Biol. 16, 045002 (2019).
47. Hommelsheim, C. M., Frantzeskakis, L., Huang, M. & Ülker, B. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. Sci Rep 4, 5052 (2014).
48. Elías-Arnanz, M. & Salas, M. Bacteriophage ϕ29 DNA replication arrest caused by codirectional collisions with the transcription machinery. The EMBO Journal 16, 5775–5783 (1997).
49. Abil, Z., Restrepo Sierra, A. M. & Danelon, C. Clonal Amplification-Enhanced Gene Expression in Synthetic Vesicles. ACS Synth. Biol. 12, 1187–1203 (2023).
50. Iskakova, M. B., Szaflarski, W., Dreyfus, M., Remme, J. & Nierhaus, K. H. Troubleshooting coupled in vitro transcription–translation system derived from Escherichia coli cells: synthesis of high-yield fully active proteins. Nucleic Acids Res 34, e135 (2006).
51. Siegal-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V. & Murray, R. M. Gene Circuit Performance Characterization and Resource Usage in a Cell-Free "Breadboard". ACS Synth. Biol. 3, 416–425 (2014).
52. Chizzolini, F., Forlin, M., Cecchi, D. & Mansy, S. S. Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength. ACS Synth. Biol. 3, 363–371 (2014).
53. Du, L., Gao, R. & Forster, A. C. Engineering multigene expression in vitro and in vivo with small terminators for T7 RNA polymerase. Biotechnol Bioeng 104, 1189–1196 (2009).
54. Du, L., Villarreal, S. & Forster, A. C. Multigene expression in vivo: supremacy of large versus small terminators for T7 RNA polymerase. Biotechnol Bioeng 109, 1043–1050 (2012).
55. Zelcbuch, L. et al. Spanning high-dimensional expression space using ribosome-binding site combinatorics. Nucleic Acids Res 41, e98 (2013).
56. Mutalik, V. K. et al. Quantitative estimation of activity and quality for collections of functional genetic elements. Nat Methods 10, 347–353 (2013).
57. Shin, J. & Noireaux, V. An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. ACS Synth. Biol. 1, 29–41 (2012).
58. Karig, D. K., Iyer, S., Simpson, M. L. & Doktycz, M. J. Expression optimization and synthetic gene networks in cell-free systems. Nucleic Acids Res 40, 3763–3774 (2012).
59. Noise Minimization in Cell-Free Gene Expression. https://pubs.acs.org/doi/epdf/10.1021/acssynbio.3c00174 doi:10.1021/acssynbio.3c00174.
60. de Vega, M., Lázaro, J. M. & Salas, M. Improvement of ϕ29 DNA Polymerase Amplification Performance by Fusion of DNA Binding Motifs. in Rolling Circle Amplification (RCA) (ed. Demidov, V. V.) 11–24 (Springer International

Publishing, 2016). doi:10.1007/978-3-319-42226-8_2.
61. Doerr, A., Foschepoth, D., Forster, A. C. & Danelon, C. In vitro synthesis of 32 translation-factor proteins from a single template reveals impaired ribosomal processivity. Sci Rep 11, 1898 (2021).
62. Karig, D. K., Jung, S.-Y., Srijanto, B., Collier, C. P. & Simpson, M. L. Probing cell-free gene expression noise in femtoliter volumes. ACS Synth Biol 2, 497–505 (2013).
63. Chizzolini, F. et al. Cell-Free Translation Is More Variable than Transcription. ACS Synth. Biol. 6, 638–647 (2017).
64. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27, 946–950 (2009).
65. Salis, H. M. The ribosome binding site calculator. Methods Enzymol 498, 19–42 (2011).
66. Reis, A. C. & Salis, H. M. An Automated Model Test System for Systematic Development and Improvement of Gene Expression Models. ACS Synth. Biol. 9, 3145–3156 (2020).
67. Vezeau, G. E. & Salis, H. M. Tuning Cell-Free Composition Controls the Time Delay, Dynamics, and Productivity of TX-TL Expression. ACS Synth. Biol. 10, 2508–2519 (2021).
68. Carruthers, D. N. et al. Microbial production of high octane and high sensitivity olefinic ester biofuels. Biotechnology for Biofuels and Bioproducts 16, 60 (2023).
69. Fujii, R., Kitaoka, M. & Hayashi, K. Error-Prone Rolling Circle Amplification Greatly Simplifies Random Mutagenesis. in Directed Evolution Library Creation: Methods and Protocols (eds. Gillam, E. M. J., Copp, J. N. & Ackerley, D.) 23–29 (Springer, 2014). doi:10.1007/978-1-4939-1053-3_2.
70. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A. & Liu, C. C. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. Cell 175, 1946-1957.e13 (2018).
71. Tian, R. et al. Engineered bacterial orthogonal DNA replication system for continuous evolution. Nat Chem Biol 1–9 (2023) doi:10.1038/s41589-023-01387-2.
72. Tian, D. et al. Cell Sorting-Directed Selection of Bacterial Cells in Bigger Sizes Analyzed by Imaging Flow Cytometry during Experimental Evolution. International Journal of Molecular Sciences 24, 3243 (2023).
73. Pelletier, J. F., Glass, J. I. & Strychalski, E. A. Cellular mechanics during division of a genomically minimal cell. Trends in Cell Biology 32, 900–907 (2022).
74. Miyachi, R., Shimizu, Y. & Ichihashi, N. Transfer RNA Synthesis-Coupled Translation and DNA Replication in a Reconstituted Transcription/Translation System. ACS Synth. Biol. 11, 2791–2799 (2022).
75. Nitta, N. et al. Intelligent Image-Activated Cell Sorting. Cell 175, 266-276.e13 (2018).
76. Isozaki, A. et al. Intelligent image-activated cell sorting 2.0. Lab Chip 20, 2263–2273 (2020).
77. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. Nat Methods 9, 676–682 (2012).
78. Scott, A. et al. Cell-Free Phospholipid Biosynthesis by Gene-Encoded Enzymes Reconstituted in Liposomes. PLoS ONE 11, e0163058 (2016).

5

Imaging Flow Cytometry for High-throughput Phenotyping of Synthetic Cells

Abstract

The reconstitution of basic cellular functions in micrometer-sized liposomes has led to a surge of interest in the construction of synthetic cells. Microscopy and flow cytometry are powerful tools for characterizing biological processes in liposomes with fluorescence readouts. However, applying each method separately leads to a compromise between information-rich imaging by microscopy and statistical population analysis by flow cytometry. To address this shortcoming, we here introduce imaging flow cytometry (IFC) for high-throughput, microscopy-based screening of gene-expressing liposomes in laminar flow. We developed a comprehensive pipeline and analysis toolset based on a commercial IFC instrument and software. About 60 thousand of liposome events were collected per run starting from one microliter of the stock liposome solution. Robust population analysis was performed from individual liposome images based on fluorescence and morphological parameters. This allowed us to quantify complex phenotypes covering a wide range of liposomal states that are relevant for building a synthetic cell. The general applicability, current workflow limitations, and future prospects of IFC in synthetic cell research are finally discussed.

This chapter is based on an already published manuscript where I share co-first authorship with Elisa Godino: Godino, E.*, Restrepo Sierra, A. M.* & Danelon, C. Imaging Flow Cytometry for High-Throughput Phenotyping of Synthetic Cells. ACS Synth. Biol. 12, 2015–2028 (2023).

* Denotes equal contribution.

Introduction

Synthetic lipid vesicles, called liposomes, are widely used as biological membrane models for basic and applied research [1,2]. By virtue of their biocompatibility, non-immunogenicity and easy manufacturing, liposomes are successfully employed as pharmaceutical (nano) carriers [3]. Moreover, they are routinely utilized as bioreactors, diagnostic and biosensing tools, and as a proxy of cellular membranes to study a variety of biochemical and biophysical mechanisms [2]. In particular, giant vesicles with a diameter typically >1 µm provide a versatile cell-like platform for the reconstitution of various biological processes, such as DNA replication [4], cytokinesis using prokaryotic or eukaryotic protein systems [5-9], dynamic self-organization at the membrane [7,10], light-driven ATP production [11], and cell-cell communication mimicry [12]. The extended repertoire of reconstituted biological functions has now prompted synthetic biologists to envision the construction of an entire synthetic cell starting from liposomes as the chassis.

A major challenge to engineering artificial cells with advanced functionalities lies in the ability to detect complex phenotypes (hereafter referring to any measurable properties, e.g., protein concentration, localization, and liposome shape) in large populations of liposomes. This is critical to identify vesicles exhibiting desired properties from heterogeneous pools, as well as to optimize and compare protocols for the production of liposomes on the basis of quantifiable parameters (yield, size homogeneity, lamellarity, activity of internalized components). Appropriate analytical methods therefore need to generate information-rich data from individual liposomes for accurate phenotype identification, combined with high-speed screening for massive data collection from large populations of vesicles.

Owing to their large size, giant vesicles are routinely imaged at high spatial and temporal resolution by fluorescence microscopy techniques. Dynamical behaviours, such as biochemical pattern formation, such as membrane fluctuations and morphological changes can be visualized in real time at the single vesicle level. However, on the first hand, fluorescence imaging generally suffers from a low screening capability, and it remains challenging to convert single-liposome properties into a large dataset that enables extraction of rare phenotypes in a quantitative and statistically relevant manner [13]. On the other hand, flow cytometry is a powerful technology for high-throughput screening of giant vesicles [13,14]. Individual liposomes (or aggregates) diluted in suspension are sequentially analysed based on scattered light and fluorescence intensity signals. One drawback of flow cytometry is that data rely on a one-dimensional fluorescence signal, which severely limits the spectrum of features that can be investigated.

To alleviate the limitations inherent to conventional fluorescence microscopy and flow cytometry for high-throughput interrogation of complex liposome phenotypes, we propose here to combine their strengths by using imaging flow cytometry (IFC). By enabling the rapid acquisition of multispectral images of single cells in flow, IFC has gained popularity in a variety of cell biology-related disciplines, where the identification and quantification of

rare cellular phenotypes within heterogenous populations are important [15]. For example, IFC has been instrumental to study apoptosis in relation to alterations of nuclear morphology and structure [16], cell cycle progression based on chromatin condensation [17], protein and molecule translocation and/or co-localization in different cellular compartments [18,19], and cytoskeleton structures [20]. To our knowledge, IFC has been applied to giant vesicle suspensions in only one study that focused on the optimization protocol of liposome production with water-in-oil emulsion transfer methods [21].

In this work, we leverage the commercial IFC instrument ImageStream to characterize synthetic cell modules from liposome populations **(Fig. 1A,B)**. ImageStream is a benchtop imaging flow cytometer that enables multispectral acquisition of individual cells or objects in flow. We develop a comprehensive pipeline for liposome identification, selection of unilamellar vesicles, and multimodal analysis of image-based phenotypes using the built-in image processing software IDEAS. We finally discuss the current limitations and opportunities of IFC as an enabling technology to accelerate synthetic cell research.

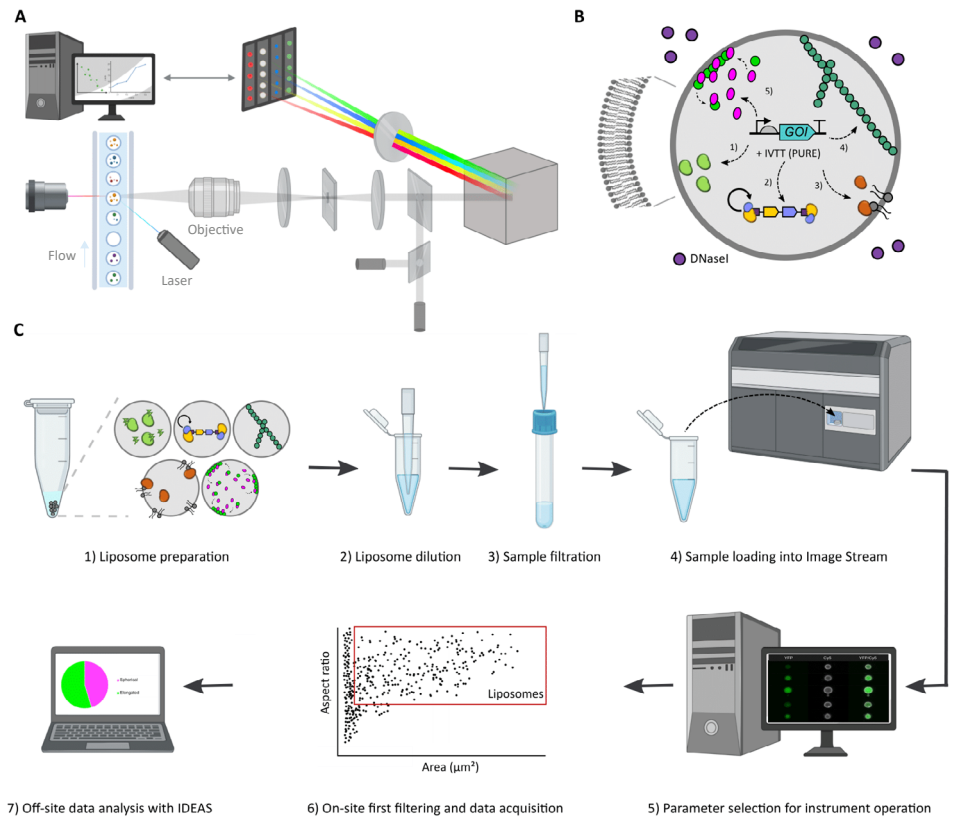


Fig. 1 Overview of the IFC workflow to characterize gene expressing liposomes. A) Schematic illustration of the primary components of ImageStream. Giant vesicles are hydrodynamically focused into a core stream and orthogonally illuminated. The emitted or scattered light is captured by the imaging objective, separated into multi-spectral bands via optical decomposition components, and projected on a charge-coupled detector. While recording, digital pictures

are displayed on the computer and stored for analysis. Image adapted from ref. [16]. **B)** Illustration of a liposome-based synthetic cell containing relevant biological processes that have separately been expressed from genes with PURE system: transcription-translation (1), DNA replication (2), phospholipid biosynthesis (3), and the formation of cytoskeletal structures (4). **C)** Overview of the workflow for sample treatment, data acquisition and image processing.

Results

Liposome detection pipeline

Samples consisted of phospholipid vesicles containing PURE system, a reconstituted transcription-translation machinery [22,23] to support gene expression in synthetic cells **(Fig. 1B)**. In most experiments, a small fraction of fluorescently-labelled phospholipids was included in the membrane composition for imaging. The glass bead-assisted lipid film swelling method was employed for liposome production (see Methods section) as it proved successful for expression of various cellular modules with PURE system. We collected for each sample around a million events, which were analysed in a batch mode by opening 100,000 recorded items at a time in IDEAS **(Fig. 1C)**. With "events" we here mean any objects whose fluorescence signal in the membrane dye channel crosses the detection threshold. Sample dilution (1 µL liposome stock solution was diluted 100 times) was set to record 800 to 1,000 events per second. With the instrument stabilization and calibration steps, measurement time for one sample was about 2 h.

We encountered that the raw file not only contained well-defined and isolated liposomes, but also aggregates of all sizes and shapes, lipid debris or small (<300 nm) vesicles, and left-over speed beads (employed to monitor sample flow, and ensure focus and core tracking). Therefore, the first step in the overall analysis pipeline consisted in setting up a gating strategy for selecting solely giant liposomes and excluding any other objects. At every step of the analysis, visual inspection of a subset of the images was carried out to assess the performance of different combinations of features (physical quantities, from now on written in italic) in a particular gating task. The IDEAS software offers a suite of integrated tools for high-content image analysis and data visualization. We started by defining an appropriate mask, i.e. the area of an image that will be used for further processing. Although the IDEAS software includes pre-loaded masks, it is possible to create customized ones by adjusting the channel and scalar values to better determine the section of the image that will be used for each feature computation. We decided to re-adjust the default mask to better encompass both the lumen and membrane of individual liposomes **(Fig. 2A)**. Then, we performed a comparative analysis over the *area* and *aspect ratio* features **(Fig. 2B)**. As the pixels are rendered into square micrometers (µm²), the area is given as the microns squared within the utilized mask. We determined that any object having a surface area lower than 40 µm² was not detected with sufficient resolution. The aspect ratio corresponds to the ratio between the minor axis and the major axis of each object and specifies how round an item is. The aspect ratio of circular objects equals one, while oblong structures have significatively lower values. Thus, all objects with an aspect ratio of less than 0.4

were excluded **(Fig. 2B,C)**. To further improve the identification of liposomes, we devised an additional three-step screening method for the selection for in-focus events. We first compared the *intensity* of the membrane signal vs the *area* of the mask **(Fig. 2D)**. Events with a small area and low membrane intensity were identified as debris or out of focus objects and were discarded from the analysis. Objects with a small area and high membrane intensity were discarded as dense lipid aggregates. Then, we used the *gradient RMS* feature to only select in-focus liposomes **(Fig. 2E)**. This feature assesses the sharpness of an image by identifying large variations in pixel intensity values. We found that all events with a gradient RMS lower than 18 corresponded to out-of-focus liposomes **(Fig. 2E,F)**. Next, we selected objects with high *H-homogeneity* values **(Fig. 2G)**, followed by a final step based on the *H-Correlation mean* and its *standard deviation* **(Fig. 2H)**. These H features establish a set of textures that describe the spatial relationships between the pixel values within the mask. We identified that low H-homogeneity with an elevated H-Correlation mean and a low standard deviation relate to aggregates. We confirmed the accuracy of the liposome selection pipeline by visually inspecting a large number of images from the final collection **(Fig. 2I)**. As staining liposomes with a membrane dye is not always possible or desired, we showed that a similar sequential gating approach as described above was also applicable to the brightfield images **(Fig. S1)**. Overall, from the one million events detected per sample, around 6% were classified as 'good' liposomes and were subjected to more advanced image processing to measure morphological features and fluorescence localization.

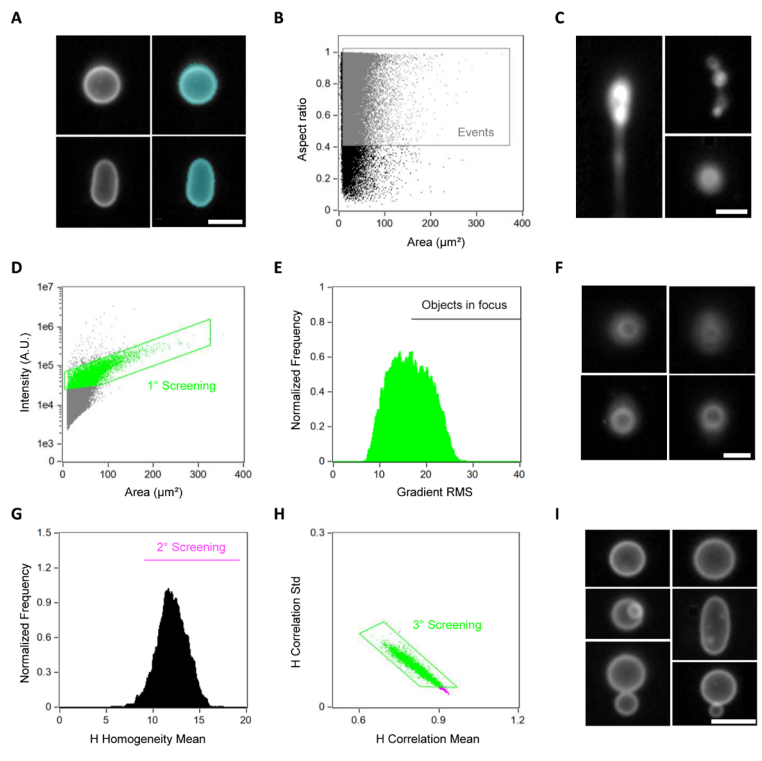


Fig. 2 Sequential gating pipeline for identification of liposomes. A) Images of two liposomes (left) acquired with ImageStream and their respective mask (right). The mask specifies the area to be used during the analysis and it was customized to select for the whole liposome. **B)** Data plot showing the *area* and *aspect ratio* features of collected events. Objects having a surface area higher than 40 μm² and an aspect ratio higher than 0.4 were gated as relevant events (in grey). **C)** Images of some aggregates with different sizes and shapes that were filtered out by the gating step in panel B **D)** Comparison of the *intensity* of the membrane signal vs the *area* of the mask. Selected data points are gated in green. **E)** Out-of-focus events were discarded based on the *gradient RMS* feature. Events with a *gradient RMS* higher than 18 were selected for further analysis. **F)** Examples of images showing out-of-focus liposomes that were discarded from the analysis. **G)** Second gating step aiming at selecting objects with high *H-homogeneity* values (in magenta). **H)** Final selection criterion based on the analysis of the *H-Correlation mean* and *standard deviation* functions. The selected objects (gated in green) correspond to the final library of liposomes, which is clean from undesired objects. **I)** Small gallery of liposome images that passed the gating pipeline. For all the images, the white colour represents the membrane dye signal. Scale bars are 7 μm.

Morphometric analysis of liposomes

Quantification of liposome morphological features, such as size, shape, lamellarity and dispersity, is essential for comparing different methods for liposome production, as well as for understanding the interplay between the inner biochemical processes and membrane mechanics. The histogram of liposome sizes shows a *diameter* of 7.2 ± 1.7 μm (mean and standard deviation), with the majority of the liposomes (91%) having a diameter lower than 10 μm **(Fig. 3A)**. The spatial resolution was good enough to clearly distinguish the membrane and subliposomal structures in vesicles bigger than 3 μm in diameter. We then examined the *circularity* feature to quantify any deviation from a spherical shape **(Fig. 3B,C)**. The circularity feature, which determines how much the analyzed mask deviates from a circle, is calculated by dividing the average distance between the object's edge and its center by the variance of such a distance. As a result, the more an object is circular, the smaller the variance and thus the higher the circularity value. We considered liposomes with a circularity higher than 10 as spherical, representing ~45% of the population of selected liposomes (62,408 events).

To identify multilamellar and multivesicular liposomes within the spherical liposome population we used the IDEAS tool called *"feature finder wizard"*. This tool assists the user at identifying the optimal features that best describe a particular phenotypic trait. To employ the *wizard*, one needs to first select individual event images (at least 25) for each of the phenotypes of interest. From this training data set, the software then suggests the appropriate features that best differentiate the predetermined/pre-defined phenotype categories. We manually classified liposomes in two distinct categories, unilamellar and multilamellar, based on visual inspection of some images, and performed the automated analysis. The output graph displayed the *compactness* and *Max Pixel* features **(Fig. 3D)**. Compactness measures the degree to which an object is packed together. The higher the value, the more condensed the object. The Max Pixel feature is the largest intensity value obtained from the background subtracted pixels from the input mask. When applying this pair of features to analyse the collection of spherical liposome images, we obtained that

~58% of the classified vesicles were unilamellar **(Fig. 3E, J)**, while the rest exhibited internal membrane structures **(Fig. 3F).** Therefore, from the one million of events collected in total per sample, ~1.6% (6 × 0.45 × 0.58) corresponds to spherical and unilamellar liposomes, that is 16,000 vesicles. It should be mentioned that, despite the capability of the wizard analysis tool to generate two distinct clusters of data points, the method is not perfect and both false positive and false negative events were detected, along with liposomes that were left unclassified. Utilization of the upgraded version of the IDEAS software provided with machine learning algorithms may solve this issue (see Discussion).

We finally sought to classify non-spherical liposomes as rod-shaped or doublets (two liposomes attached together) **(Fig. 3G)**, and to quantify their abundance. To distribute liposomes in either of the two categories, we again employed the *feature finder wizard*. The generated data are a scatter plot of the texture features *H-entropy std* and *H-correlation std* **(Fig. 3H)**. When applied to the population of non-spherical liposomes, the analysis predicted that ~15% of liposomes were rode-shaped and ~69% were doublets **(Fig. 3I,K)**.

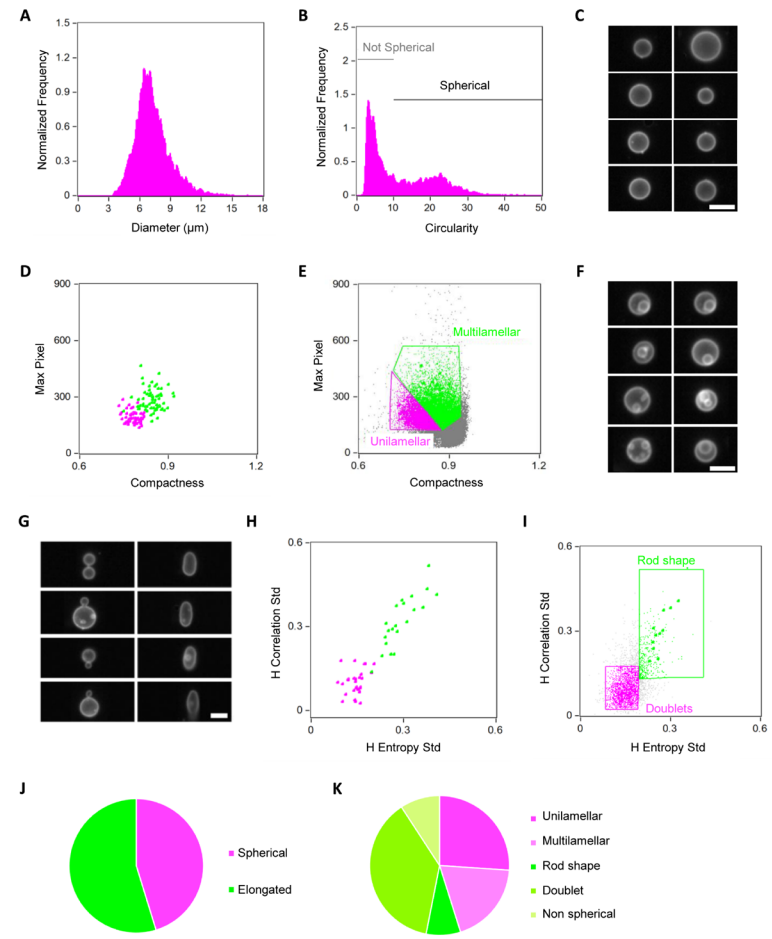


Fig. 3 Characterization of liposome morphology. A) Distribution of liposome diameter. **B)** Shape analysis using the *circularity* feature. Objects with a high circularity value were gated as spherical liposomes. **C)** Image gallery of some liposomes classified as spherical. **D)** Two-dimensional plot of the *compactness* and *max pixel* features as output from the *feature finder wizard tool* performed to distinguish liposomes having intravesicular membrane structures. **E)** Population analysis (*compactness* vs *max pixel*) applied to the spherical liposomes. Multilamellar/multivesicular and unilamellar liposomes are gated in green and magenta, respectively. **F)** Image gallery of representative liposomes classified as multilamellar/multivesicular. **G)** Image gallery displaying two different phenotypes present in the subpopulation of non-spherical liposomes gated in panel (B): doublets which correspond to two liposomes attached together (left) and rod-shape liposomes (right). **H)** Two-dimensional plot of the *H entropy std* and *H correlation std* features as the outcome of the *feature finder wizard* tool analysis for discriminating between the two subpopulations of non-spherical liposomes shown in panel (G). **I)** Population analysis (*H entropy std* vs *H correlation std*) applied to the non-spherical liposomes. Rod-shaped and doublet liposomes are gated in green and magenta, respectively. **J, K)** Graphical representations of the statistics obtained from the morphometric analysis carried out on a set of over 70,000 liposomes from one sample. Non spherical liposomes that did not belong to the categories 'rod shape' or 'doublet' were classified as 'non spherical'. For all the images, the white color represents the membrane dye signal. Scale bars are 7 µm.

The results so far demonstrate the power of IFC in acquiring large collections of liposome images and providing statistical population analysis of morphological properties. Next, we applied IFC capabilities for quantitative analysis of synthetic cell modules whose gene-encoded functionalities lead to distinct liposomal phenotypes.

Lumen localization reporter of transcription-translation

A linear DNA encoding the yellow fluorescent protein (YFP) was expressed in liposomes [13] The sample was run into ImageStream and liposomes were identified as described above. To select YFP-expressing liposomes, we generated a fitting mask and plotted the histogram of the *intensity* measured in the 488-nm channel **(Fig. 4A)**. After applying an intensity threshold above which liposomes were classified as expressing, we found that ~60% of liposomes (including nonspherical and multilamellar ones), exhibit YFP signal in their lumen. We estimate the lower limit of detection of freely diffusing YFP in the lumen by IFC to be ~500 nM for accurate quantitation, similar as we observed with confocal imaging of glass surface-immobilized vesicles [13]. This corresponds to ~10,000 molecules in a 4-µm diameter liposome. Intensity values span across an order of magnitude indicating that transcription-translation efficiency can substantially vary between liposomes. This heterogeneity in gene expression levels holds over the entire range of liposomes sizes, with no strong correlation between YFP signal and vesicle size, as shown when plotting the *intensity* of YFP as a function of the *area* of liposomes **(Fig. 4B,C)**. Similar observations have already been reported using confocal fluorescence microscopy [13]. The main advantages of IFC are that liposomes are imaged in suspension (not in contact to a glass surface), the screening throughput is higher allowing us to perform more accurate statistical population analysis, and that a gallery of individual liposome images is generated in real-time **(Fig. 4D,E)**.

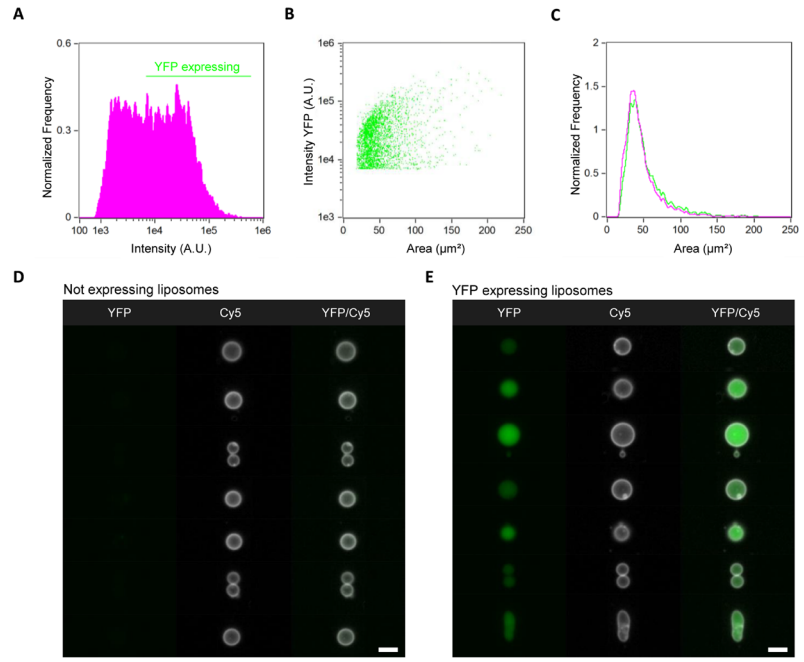


Fig. 4 IFC data analysis of the transcription-translation module. A) Histogram of the internal YFP *intensity* and definition of a threshold to distinguish non-expressing from gene expressing liposomes. **B)** Scatter plot of the liposomes *area* vs the *intensity* of expressed YFP. **C)** Comparison of the size distribution between YFP-expressing liposomes (green) and non-expressing liposomes (magenta). **D, E)** Gallery of representative images of non-expressing liposomes in panel (D), and YFP-expressing liposomes in panel (E). Membrane dye (Cy5) signal is coloured in white, YFP is in green. The analysis was performed on a set of over 70,000 liposomes from one sample. Scale bars are 7 µm.

DNA replication coupled to transcription-translation

We recently designed and implemented in PURE system a self-replicating DNA based on the essential replication proteins of bacteriophage Phi29 [4]. Cell-free expression of the replicator DNA in liposomes yielded exponential amplification of DNA, which led to increased fluorescence of a nucleic acid intercalating dye, as observed by confocal microscopy [4]. Fluorescence was not always evenly distributed in the vesicle lumen, but localized into bright spots (or replication 'blobs'), suggesting aggregation of concentrated DNA in the presence of some PURE compounds (e.g., spermidine, inorganic phosphate, Mg^{2+}) [4]. No statistical population analysis was performed in our previous study because of significant background from the acridine orange DNA binding dye and the limited number of imaged liposomes.

Herein, we exploited the assets of IFC, combined with dsGreen as a lower-background DNA dye, to quantitatively assay in-liposome DNA replication. From the collected single-liposome images, we computed the histogram of dsGreen *intensity* and found that ~36% of liposomes were active for DNA replication (i.e. dsGreen signal was higher than background-

corrected threshold) **(Fig. 5A)**. We visually identified three distinct phenotypes based on the intensity and lumen localization of dsGreen fluorescence signal: no or weak dsGreen signal **(Fig. 5D)**, homogeneous intensity of dsGreen within the lumen **(Fig. 5E)**, and (usually one) replication blob **(Fig. 5F)**, the latter two phenotypes corresponding to successful DNA amplification. To distinguish these two visual phenotypes, we screened several IDEAS features and found that *std dev* and *H-Homogeneity mean* offered the best combination to discriminate them with high accuracy **(Fig. 5B)**. The *std dev* feature describes the general distribution of pixel intensities by computing the standard deviation in the defined mask. A greater std dev value implies a higher texture. Liposomes with high std dev and low H-Homogeneity mean typically display a bright replication blob **(Fig. 5B)**, representing ~34% of the vesicle population **(Fig. 5G)**. Moreover, we discovered that the liposomes exhibiting a replication blob had a relatively smaller area (size) compared to those having a homogeneously distributed dsGreen signal **(Fig. 5C)**. Further investigations are needed to clarify which factors trigger the formation of DNA condensation. These results demonstrate that IFC is capable to reveal and provide statistical analysis of phenotypic heterogeneity at sub liposomal resolution.

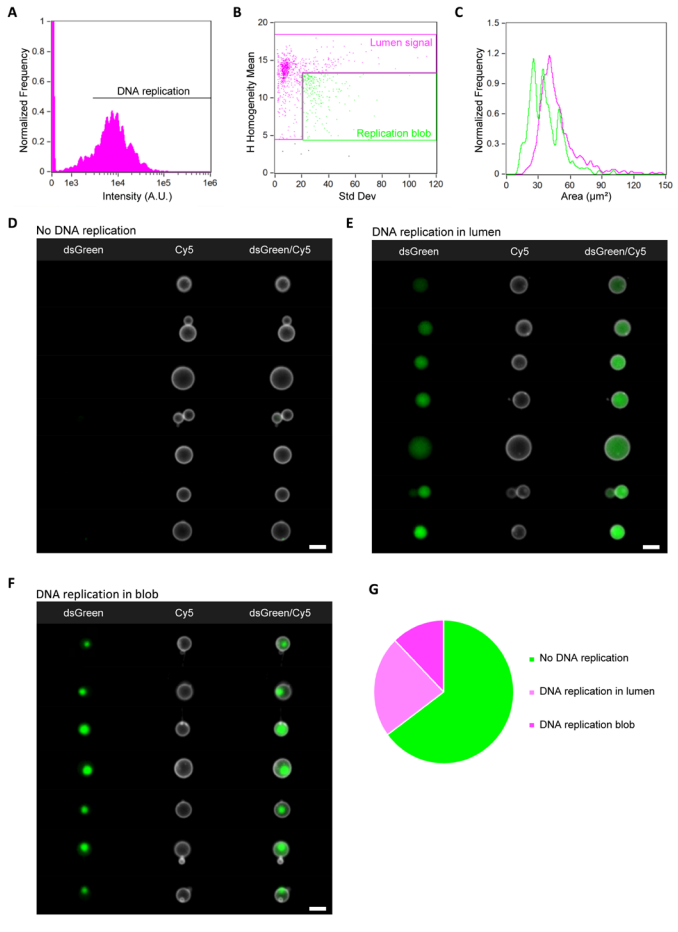


Fig. 5 IFC data analysis of the DNA replication module. A) Histogram of the intensity of the DNA-binding dye dsGreen in liposomes. Successful DNA replication is reflected by high intensity values. **B)** Scatter diagram of the *H-homogeneity mean* vs *std dev* features to classify images with respect to the distribution of dsGreen fluorescence in the vesicle lumen. Liposomes with homogeneously distributed dsGreen are gated in magenta, while vesicles exhibiting a "replication blob" are gated in green. **C)** Histograms of liposome size (area) for the two subpopulations defined in panel (B). Color coding is the same as in panel (B). Liposomes with a replication blob are on average smaller than those with an even intraluminal signal. **(D–F)** Gallery of representative images of liposomes with no DNA replication in panel (D), increased DNA amount with homogeneous spatial localization in panel (E), and increased DNA amount with formation of condensates in panel (F). **G)** Graphical representation of the statistics obtained from the analysis of over 60,000 liposomes from one sample. For all the images, the membrane dye (Cy5) signal is colored in white and DNA-bound dsGreen is in green. Scale bars are 7 µm.

Protein self-organization into bacterial microtubules

Bacterial microtubules are protein filaments composed of polymerized tubulins BtubA and BtubB from *Prosthecobacter* cells [24,25]. Expression of the genes *btubA* and *btubB* in PURE system produces BtubA/B microtubules on flat membranes in inside liposomes [26]. In a synthetic cell, such cytoskeletal structures could play a role in membrane stabilization, polarization or internal trafficking processes. Liposomes with expressed bacterial microtubules were analysed by IFC. A trace amount of purified BtubA/B labelled with the fluorophore AlexaFluor-488 was co-encapsulated for visualization. The membrane signal was employed to select liposomes as described above. We discarded from the analysis all the vesicles that displayed a fluorescence *intensity* value of AlexaFluor-488 lower than a certain threshold **(Fig. 6A)**. Then, a wide range of textural features were explored individually or in combination to distinguish the liposomes with protein filaments. The *std dev* function (for detecting inner inhomogeneity) and the *contrast* feature (which assesses an image's sharpness by identifying big variations in pixel values) proved to be the best combination to detect microtubules, with high std dev and high contrast values **(Fig. 6B)**. An image library of the two sub-populations of liposomes with or without self-organized bacterial microtubules is shown in **Fig. 6C,D**. Statistical population analysis revealed that around 4% of liposomes exhibited detectable BtubA/B filaments.

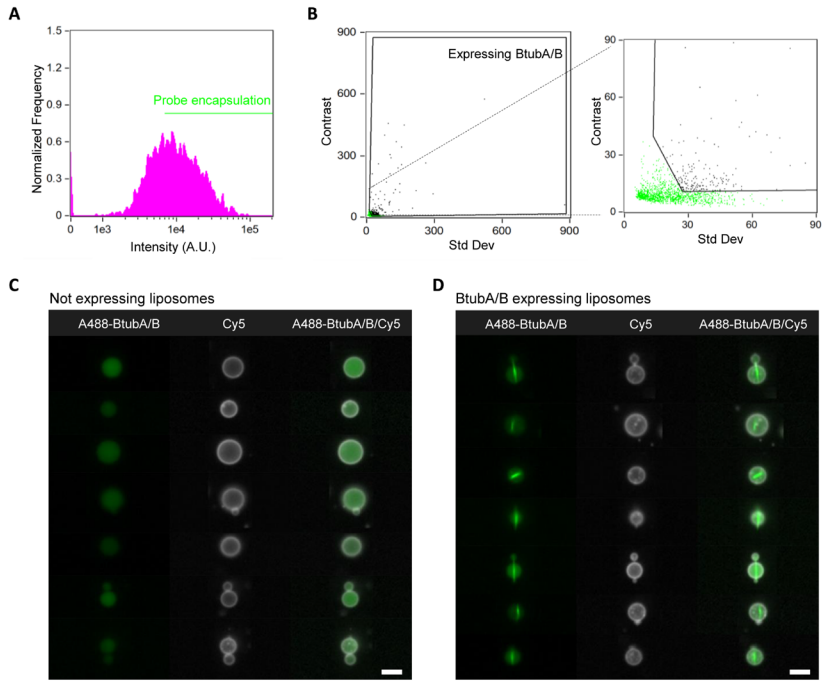


Fig. 6 IFC data analysis of the bacterial microtubule module. A) Histogram of the intensity of encapsulated A488-BtubA/B. Liposomes with an above-threshold intensity value were selected. **B)** Liposomes with cytoskeletal structures were identified by plotting the *std dev* vs the *contrast* feature (black gate). A zoom-in image is shown on the side to better visualize the two populations and the applied gating. **C, D)** Gallery of representative images of liposomes with encapsulated free tubulin but no filaments in panel (C), or with characteristic microtubules of expressed BtubA/B in panel (D). For all the images, the membrane dye (Cy5) signal is coloured in white and A488-BtubA/B is in green. The analysis was performed on a set of over 35,000 liposomes from one sample. Scale bars are 7 µm.

Re-localization of Min proteins to the membrane

The Min system, which comprises the proteins MinC, MinD, and MinE, is primarily responsible for the spatial organization of the division site in *E. coli* [27]. The Min proteins self-organize at the inner surface of the cytoplasmic membrane and oscillate between the two cell poles in a dynamic manner. MinD and MinE drive the oscillations, while MinC travels on the waves by interacting with MinD [28]. Reconstitution of Min protein dynamics in liposomes has already been accomplished using PURE system [7,10], which may assist binary fission in a prospective synthetic cell.

We challenged IFC to detect the re-localization of eGFP-MinC from the lumen to the membrane, where it is recruited by cell-free expressed MinD. During the liposome identification procedure, we noticed a high percentage of events corresponding to liposomes with inner membrane structures. Thus, we decided to include a more stringent image processing step for accurate selection of unilamellar liposome. A fourth gating using *H-Correlation* and *H-Contrast standard deviations* in the membrane dye channel was applied

to discard events with high local intensity variations **(Fig. 7A)**. Next, discrimination of liposomes with eGFP-MinC located exclusively in the lumen or also on the membrane (i.e. expressing MinD), was carried out with the *bright detail similarity R3* feature **(Fig. 7B)**. This function compares the bright details of two images and can be used to measure signal co-localization. The feature computes the log-transformed Pearson's correlation coefficient of small bright regions (3 pixels radius) inside the mask provided for the two input images. Here, the membrane dye and the eGFP-MinC images were chosen as the two inputs. Events with a bright detail similarity greater than 2.5 were considered as correlated (eGFP-MinC also localized at the membrane) and the liposomes were classified as MinD-expressing, which accounted for approximately 54% of the total population. An image gallery of liposomes classified as MinD-expressing or not is displayed in **Fig. 7C,D.**

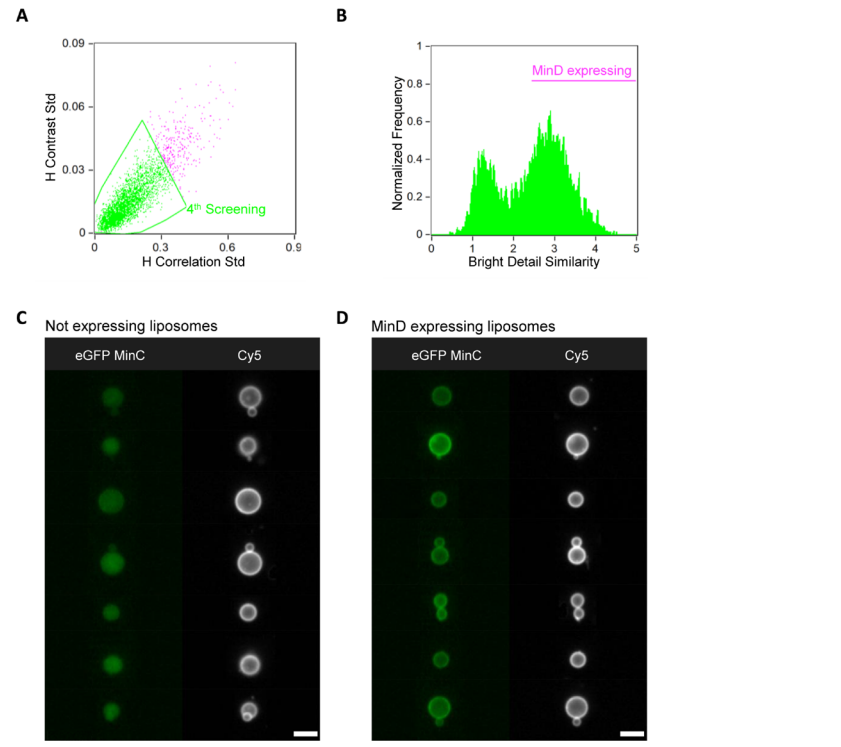


Fig. 7 IFC data analysis of the Min self-organization module. A) Scatter diagram of the parameters *H-Correlation std* and *H-Contrast std* used as an additional image processing step for the identification of higher quality liposomes. The membrane dye (Cy5) signal was used for the analysis. Gated events considered as good liposomes are coloured in green. **B)** Histogram of the *bright detail similarity R3* feature (using as inputs the two images in the Cy5 and eGFP-MinC channels) from the liposomes gated in panel (A). Gated liposomes with active Min proteins exhibit membrane localization of the eGFP-MinC signal upon binding to expressed MinD. **C, D)** Gallery of representative images of liposomes with an inactive (lumen localization of eGFP-MinC in panel C) or active (membrane localization of eGFP-MinC in panel D) MinD-MinC system. For all the images, the membrane dye (Cy5) signal is coloured in white and eGFP-MinC is in green. The analysis was performed on a set of 40,000 liposomes from one sample. Scale bars are 7 µm.

PssA-catalysed synthesis of PS lipid

The DNA-encoded production and incorporation of membrane constituents is essential for sustainable growth of synthetic cells. Enzymes from the *E. coli* Kennedy pathway for phospholipid biosynthesis were expressed in PURE system and newly synthesized lipids were inserted in the liposome membrane [29]. One of these enzymes, CDP-diacylglycerol-serine O-phosphatidyltransferase (PssA), converts cytidine diphosphate (CDP) lipid headgroup into phosphatidylserine (PS) using L-serine as a co-substrate. Liposomes containing a small fraction of CDP-DAG lipids in the membrane were formed, the *pssA* gene was expressed with co-encapsulated PURE system and L-serine, and PS-containing vesicles were stained using the fluorescent probe eGFP-LactC2 prior to the analysis with ImageStream. Recruitment of the externally added eGFP-LactC2 to the liposome membrane indicates internal expression of PssA and concomitant production of PS. We assume that flip-flop of PS from the inner to the outer membrane leaflet precedes binding to eGFP-LactC2, a process that may be less energetically unfavorable in PURE-containing liposomes compared to simple buffer conditions. Since we noticed mild unspecific binding of the eGFP-LactC2 probe to the liposome membrane, we decided to also run a negative control sample, where the *pssA* gene was omitted. We plotted the histograms of fluorescence intensity of eGFP-LactC2 and defined a threshold value above which liposomes were classified as PS-containing with a minimal number of false positives **(Fig. 8A)**. An image gallery of the two sub-populations of liposome images is shown in **Fig. 8B,C**. Under these conditions, about 62% of the liposomes successfully converted CDP-DAG lipids into PS by the internally expressed PssA enzyme.

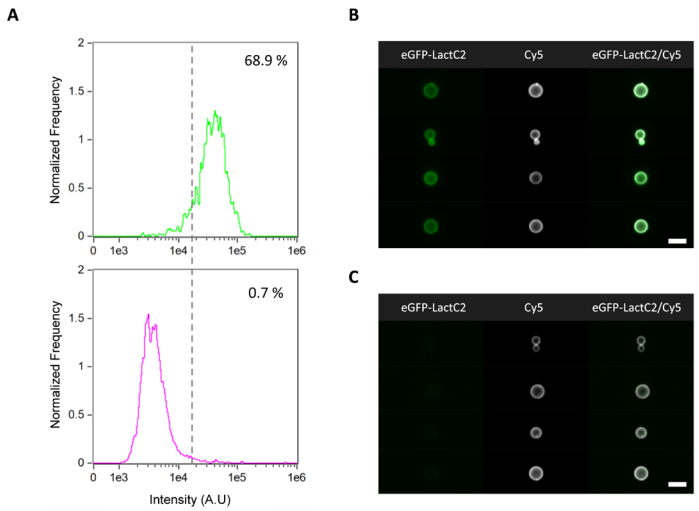


Fig. 8 IFC data analysis of the PssA-catalysed phosphatidylserine biosynthesis module. A) Histogram of the fluorescence intensity of the liposome membrane-recruited eGFP-LactC2 (PS-specific probe) in a positive sample (+DNA, top) and in a control negative sample (−DNA, bottom). The control sample was run in order to exclude false positive events resulting from mild unspecific binding of the probe. The applied intensity threshold is indicated by the vertical dashed line and the percentage of PS-containing liposomes is appended in each graph. **B, C)** Gallery of representative images of liposomes with PS-production in panel (B) and no production in panel (C). For all the images, the membrane dye (Cy5) signal is

coloured in white and eGFP-LactC2 signal is in green. The analysis was performed on a set of over 50,000 liposomes from one sample. Scale bars are 7 µm.

Discussion

We introduced IFC for high-throughput imaging of liposomes in laminar flow. The general applicability of IFC in synthetic cell research was demonstrated by assaying diverse populations of liposomes with gene-encoded functional modules. A comprehensive workflow was developed to collect and process up to one million of images of single vesicles, allowing for the evaluation of fluorescence and morphological parameters with statistical analysis of subpopulations, and for sample comparison with no selection bias. The commercial IDEAS software provides over 200 features, covering a wide range of liposome phenotypes that can be quantified. The capability to retrieve images of individual events from plotted data strongly reduces the risks to include false-positive or false-negative events in the analysis.

From the 1 million events detected per sample, about 6% were classified as "good" liposomes (60,000 liposomes on average across all the samples assayed in the study). Besides the stringent gating chosen, it is possible that some liposomes get disrupted during sample filtration before loading to ImageStream or inside the flow device. It is important to note that the samples consisted of a diluted solution of only 1 µL of liposome suspension taken from the 20-µL PURE solution used for lipid film swelling. Scaling up the liposome suspension volume or concentration would straightforwardly increase the number of events, which may be relevant for statistical analysis of very low abundance phenotypes. Data analysis pipelines can easily be saved as templates for the analysis of multiple batches from 1,000 to 100,000 events each. Data processing is then rapid and not excessively computationally demanding, even for complex phenotypes as reported here. To increase the fraction of good liposomes over debris and aggregates, a stricter gating could be performed during image acquisition, for instance by applying the first step **(Fig. 2B)** – or first two steps **(Fig. 2E)** – in our current postacquisition image analysis pipeline. Because the user cannot adjust the channel masks during acquisition, we recommend to perform this step again during offline data analysis anyway.

Further expansion of imaging capabilities with more than two colors (488 and 642 nm laser lines are provided in the basic ImageStream configuration) is also possible with additional wavelengths (405, 561, and 592 nm are also available), which can be useful for multispectral analysis of complex phenotypes. Furthermore, increased analysis power and workflow simplification on the IDEAS software are now possible with a new machine learning module (https://www.luminexcorp.com/imagestreamx-mk-ii/#software).

Many protocols for the formation of giant liposomes have been reported in the literature (we here limit the citations to articles, in which cell-free gene expression was demonstrated [30–36]), each laboratory often having its preferred method based on available equipment, experimental or biological constraints, inclination to microfluidic approaches or not, etc. It

would however be relevant for the synthetic cell community to be offered some guidelines to choose the most appropriate methodology for liposome preparation (including lipid composition and other input parameters) on the basis of objective performance metrics with robust population statistics. In this context, IFC represents a technology of choice for high-throughput screening and quantitative analysis of different liposome samples. The presented workflow provides a generic template for analyzing liposome samples prepared with any methodologies. Other types of synthetic cell chassis, such as peptide vesicles [37], polymersomes [38], and polymer microcapsules containing a clay-hydrogel [39], could also be analyzed with IFC. In that case, users can define other sets of IDEAS features and adjust the gating stringency for some of the steps as this may better fit their purposes.

Leveraging IFC with physical sorting of liposome subpopulations will open the door to directed evolution of synthetic cells [40]. The recent technological breakthrough in image-activated cell sorting [41] represents a milestone toward imaging-based selection of liposomes exhibiting desired phenotypic traits, their selection for further analytical investigations, and enrichment of genetic variants conferring a higher degree of aliveness. Finally, as IFC is an on-chip technology, it could be combined with microfluidic production of liposomes, creating a completely automated platform for synthetic cell generation and analysis.

Materials and Methods

Purified proteins

eGFP-MinC was purified according to published protocols [40]. Protein concentration was determined by Bradford assay and by measuring eGFP absorbance. BtubA/B was purified and labelled with AlexaFluor-488 as previously described [26]. Concentration of purified bacterial tubulin was determined by absorbance measurement at 280 nm (extinction coefficient 103,754.2 $M^{-1}cm^{-1}$). Purified Phi29 DNA-binding proteins were produced as described in ref [41] and ref [42]. Purified eGFP-LactC2 was prepared as described in ref [29].

DNA constructs

All DNA templates expressed in PURE system were linear products of polymerase chain reactions (PCR) from a parental plasmid. Constructs containing the *minD*, *btubA* or *btubB* gene were prepared as previously reported [10,26]. Forward and reverse primers ChD709 and ChD757, respectively annealing to the T7 promoter and T7 terminator sequences, were used for PCR. The *yfp*, *pssA*, and *p2-p3* (self-replicating DNA) expressing plasmids were sub-cloned into ori-containing vectors via Gibson assembly. All the plasmids were cloned into *E. coli* Top10 chemically competent cells. Individual colonies were outgrown in LB/ampicillin (50 µg/mL). Plasmids were extracted using the PURE Yield Plasmid Miniprep kit (Promega) and sent for Sanger sequencing confirmation at Macrogen Europe B.V. Linear fragments were obtained from PCR amplification of the transcription cassette in sequence-verified plasmids using Phusion High-Fidelity DNA Polymerase (NEB) with the forward and reverse primers ChD491 and ChD492, respectively. The amplified PCR fragments

were purified using QIAquick PCR purification kit (Qiagen). For the purification of *p2-p3* linear DNA, RNeasy MinElute Cleanup columns (Qiagen) were utilized instead of DNA columns provided with the kit. The general QIAquick manufacturer protocol was modified by having a longer pre-elution buffer drying step (at least 4 min at 10,000 g with open columns), and a longer column incubation step (at least 5 min) with ultrapure water (20-30 µL of Merck Milli-Q water) prior to the final DNA elution. The purified DNA was quantified by Nanodrop 2000c spectrophotometer (Isogen Life Science) and further analysed for size and purity with DNA gel electrophoresis. Purified DNA fragments were stored at −20 °C. Sequences of the primers can be found on **Table S1**. DNA sequences from the utilized linear constructs can be found in https://pubs.acs.org/doi/10.1021/acssynbio.3c00074.

Lipids

1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE), 1,2-dioleoyl-sn-glycero-3phosphoglycerol (DOPG), 1',3'-bis[1,2-dioleoyl-sn-glycero-3-phospho]-glycerol (18:1 cardiolipin), 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[biotinyl(polyethylene glycol)-2000 (DSPE-PEG-biotin), 1,2-dioleoyl-sn-glycero-3-(cytidine diphosphate) (CDP-DAG), and DOPE-Cy5 were from Avanti Polar Lipids.

Preparation of lipid-coated beads

The glass bead-assisted lipid film swelling method was utilized for liposome production [34]. Lipid-coated microbeads provide a large lipid film surface area, thus a high yield of liposomes even when starting from microliter swelling solution. Moreover, the method is solvent-free and compatible with a large variety of natural and functionalized lipids. Two different lipid mixtures were prepared. Mixture 1 was used in samples for assaying liposome morphology, YFP, DNA replication, tubulin, and Min proteins, and it consisted of DOPC (50 mol %), DOPE (36 mol %), DOPG (12 mol %), 18:1 cardiolipin (2 mol %), DSPE-PEG-biotin (1 mass%), and DOPE-Cy5 (0.5 mass%) for a total mass of 2 mg. Mixture 2 was used in samples for assaying PS biosynthesis and it contained DOPC (47.5 mol %), DOPE (34.2 mol %), DOPG (11.4 mol %), 18:1 cardiolipin (1.9 mol %), 1,2-dioleoyl-sn-glycero-3-(cytidine diphosphate) (5 mol %), DSPE-PEG-biotin (1 mass%), and DOPE-Cy5 (0.5 mass%) for a total mass of 2 mg. For both mixtures, lipids dissolved in chloroform were mixed in a 5 ml round-bottom glass flask. Methanol containing 100 mM rhamnose was added to the lipid solution with a chloroform-to-methanol volume ratio of 2.5:1. Then, 600 mg of 212–300 µm glass beads (acid washed, Sigma-Aldrich) was poured to the lipid-rhamnose solution, and the organic solvent was removed by rotary evaporation at 200 mbar for 2 h at room temperature, followed by overnight desiccation. Lipid-coated beads were stored under argon at −20 °C until use.

Production of gene expressing liposomes

Twenty microliters of PURE*frex*2.0 (GeneFrontier, Japan) reaction mixtures were assembled

on ice in a 1.5 mL Eppendorf tubes according to the supplier's recommendations. The exact composition was adjusted to the specific biological module to be reconstituted.

- YFP expression: PURE*frex*2.0 and 4 nM of *yfp* DNA.
- DNA replication: PURE*frex*2.0, 20 mM ammonium sulphate, 300 µM dNTPs, 750 µg mL^{-1} purified SSB, 210 µg mL^{-1} purified DSB, 1.2 units $µL^{-1}$ of Superase-In RNase inhibitor (Ambion), and 4 nM of *p2-p3* DNA.
- Bacterial tubulin: PURE*frex*2.0, 1 µL DnaK mix (GeneFrontier), 100 nM Atto488-BtubA/B, 3.75 nM of *btubA* and 2.5 nM of *btubB* DNA.
- Min system: PURE*frex*2.0, 1 µL DnaK mix, 2.5 mM ATP, 0.5 µM purified eGFP-MinC, 5 nM of *minD* DNA.
- PS synthesis: PURE*frex*2.0 and 4 nM of *pssA* DNA.

About 10 mg of lipid-coated beads was transferred to the pre-assembled PURE*frex*2.0 reaction solution and liposomes were formed by natural swelling of the lipid film. The tubes were gently rotated on an automatic tube rotator (VWR) at 4 °C along its axis for 30 min. The samples were then subjected to four freeze-thaw cycles by alternating short incubations in liquid nitrogen and in ice. Using a cut pipette tip, 10 µL of the liposome suspension was harvested, by paying attention to not collect glass beads, and transferred to a PCR tube, where it was mixed with 1 µL of DNase I (0.07U $µL^{-1}$) (Thermo Scientific) to prevent gene expression outside liposomes. Samples were incubated in a Thermal Cycler (C1000 Touch, Biorad) at 30 °C (for DNA replication) or 37 °C (for all the other conditions) for 2.5-3 hours (bacterial tubulin assay), 3 hours (Min system assay), or 16 hours for the other cellular modules.

Sample preparation for IFC measurements

Liposome solution of 1 µL was diluted in 100 µL of buffer (20 mM HEPES-KOH pH 7.6, 180 mM potassium glutamate, 14 mM magnesium acetate). To remove any remaining glass beads from the liposome suspension, the diluted sample was gently filtered through a cell-strainer cap (35 µm nylon) and collected into 5 mL round bottom polystyrene test tubes (Falcon). An additional staining step was performed in some samples prior running IFC experiments. To assay DNA replication, dsGreen (Lumiprobe) dye was supplemented at a 1:100,000 dilution factor of the stock concentration. For PS detection, purified eGFP-LactC2 was added to a final concentration of 316 nM. Samples were incubated for 30-60 min at room temperature before loading into ImageStream.

Acquisition and analysis of IFC data

All samples were analysed with the Amnis ImageStream Mk II and INSPIRE acquisition software (201.1.0.724) (Luminex Corporation). The following laser power settings were used:

- Morphology analysis, 150 for 488 nm / 50 for 642 nm / 3 for 785 nm,

- YFP expression, 40 for 488 nm / 50 for 642 nm / 3 for 785 nm,
- DNA replication, 15 for 488 nm / 50 for 642 nm / 3 for 785 nm,
- Bacterial tubulin, 120 for 488 nm / 40 for 642 nm
- Min system, 200 for 488 nm / 50 for 642 nm / 3 for 785 nm.
- PS synthesis: 200 for 488 nm / 160 for 642 nm / 3 for 785 nm.

The 60X magnification objective was employed, focus was set to automatic mode, and fluidics were set to low speed and high sensitivity. One million events were collected for each sample, which corresponds to an estimated volume of 40 to 50 µL of the diluted liposome suspension. All data were analysed with Amnis IDEAS 6.2 analysis software (Luminex). The displayed images are representative of the whole sample. Scatterplots are from 100,000 recorded items, as we analysed 1 million events by opening them in sequential batch mode.

Acknowledgements

We thank Duco Blanken, Mats van Tongeren, and Jard Mattens for running pilot experiments with ImageStream. We are grateful to Peter Rhein and Erwin Swart (Luminex Corporation) for helping us to set up preliminary measurements and for fruitful discussions. Images in Fig. 1 and the graphical abstract were created with BioRender.com. This work was financially supported by The Netherlands Organization for Scientific Research (NWO/OCW) via the "BaSyC – Building a Synthetic Cell" Gravitation Grant (024.003.019).

Supplementary information

Name	Sequence (5'→ 3')
ChD491	AAAGTAAGCCCCCACCCTCACATG
ChD492	AAAGTAGGGTACAGCGACAACATACAC
ChD709	CAAAAAACCCCTCAAGACCCGTTTAGAGG
ChD757	TAATACGACTCACTATAGGG

Table. S1 List of primers used in this study

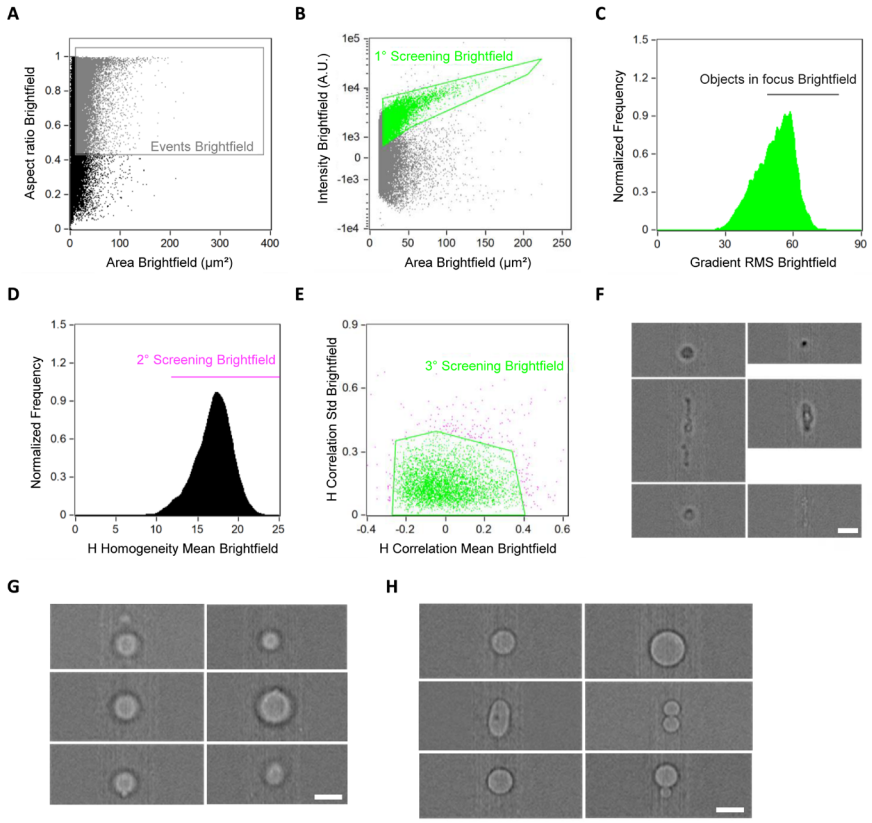


Fig. S1 Liposome identification using the brightfield channel. The brightfield channel was used to identify liposomes that did not have a membrane signal. The pipeline is similar to the one reported in Fig. 2. A mask fitting the area to be used during the analysis in the brightfield channel was created. Identifying liposomes via the brightfield channel was more challenging than using a membrane dye signal. Structural characterization was less precise and stricter cut-off values were needed for the selection of actual liposomes. This stringency can, however, discard some "good" liposomes. We always recommend visual inspection of all gated populations for validation. **A)** Scatter diagram of the *area* and *aspect ratio* features. Objects having a surface area > 30 µm² and an aspect ratio > 0.4 were gated as relevant events (grey gate). **B)** Scatter plot of the *intensity* of the brightfield signal and the *area* of the mask (green gate). C) In-focus events were selected based on the *gradient RMS* feature. The events having a gradient RMS > 4.5 were selected for further analysis (black gate). **D)** Selection of liposomes with high *H-homogeneity* values (magenta gate). **E)** Scatter diagram of the *H-Correlation mean* and *standard deviation*. The selected objects (green gate) correspond to the final liposome population, which is cleared from undesired events. **F)** Images of debris, aggregates with different sizes and shapes that were eliminated. **G)** Images of out-of-focus liposomes that were excluded from the analysis. **H)** Gallery of liposome images that passed the gating steps. Analysis in the brightfield channel was performed on a small set of 8,000 liposomes from one sample. Scale bars are 7 µm.

References

1. Walde, P., Cosentino, K., Engel, H. & Stano, P. Giant vesicles: preparations and applications. Chembiochem 11, 848–865 (2010).
2. Pick, H., Alves, A. C. & Vogel, H. Single-vesicle assays using liposomes and cell-derived vesicles: from modeling complex membrane processes to synthetic biology and biomedical applications. Chem. Rev. 118, 8598–8654 (2018).
3. Daraee, H., Etemadi, A., Kouhi, M., Alimirzalu, S. & Akbarzadeh, A. Application of liposomes in medicine and drug delivery. Artificial Cells, Nanomedicine, and Biotechnology 44, 381–391 (2016).
4. Van Nies, P. et al. Self-replication of DNA by its encoded proteins in liposome-based synthetic cells. Nat. Commun. 9, 1583 (2018).
5. Furusato, T. et al. De novo synthesis of basal bacterial cell division proteins FtsZ, FtsA, and ZipA inside giant vesicles. ACS Synth. Biol. 7, 953–961 (2018).
6. Godino, E. et al. Cell-free biogenesis of bacterial division proto-rings that can constrict liposomes. Commun. Biol. 3, (2020).
7. Kohyama, S., Merino-Salomón, A. & Schwille, P. In vitro assembly, positioning and contraction of a division ring in minimal cells. Nat Commun. 13, 6098 (2022).
8. Godino, E. & Danelon, C. Gene-directed FtsZ ring assembly generates constricted liposomes with stable membrane necks. Adv. Biol. 2, e2200172 (2023).
9. Litschel, T., Kelley, C. F., Holz, D., Adeli Koudehi, M., Vogel, S. K., Burbaum, L., Mizuno, N., Vavylonis, D. & Schwille, P. Reconstitution of contractile actomyosin rings in vesicles. Nat. Commun. 12, 2254 (2021).
10. Godino, E. et al. De novo synthesized Min proteins drive oscillatory liposome deformation and regulate FtsA-FtsZ cytoskeletal patterns. Nat. Commun. 10, 4969 (2019).
11. Berhanu, S., Ueda, T. & Kuruma, Y. Artificial photosynthetic cell producing energy for protein synthesis. Nat. Commun. 10, 1325 (2019).
12. Chakraborty, T. & Wegner, S. V. Cell to cell signaling through light in artificial cell communities: Glowing predator lures prey. ACS Nano 15, 9434–9444 (2021).
13. Blanken, D., van Nies, P. & Danelon, C. Quantitative imaging of gene-expressing liposomes reveals rare favorable phenotypes. Phys. Biol. 16, 045002 (2019).
14. Nishimura, K. et al. Population analysis of structural properties of giant liposomes by flow cytometry. Langmuir 25, 10439–10443 (2009).
15. Barteneva, N. S., Fasler-Kan, E. & Vorobjev, I. A. Imaging flow cytometry: coping with heterogeneity in biological systems. J. Histochem. Cytochem. 60, 723–733 (2012).
16. George, T. C. et al. Distinguishing modes of cell death using the ImageStream multispectral imaging flow cytometer. Cytometry. A 59, 237–245 (2004).
17. Blasi, T. et al. Label-free cell cycle analysis for high-throughput imaging flow cytometry. Nat. Commun. 7, 10256 (2016).
18. George, T. C. et al. Quantitative measurement of nuclear translocation events using similarity analysis of multispectral cellular images obtained in flow. J. Immunol. Methods 311, 117–129 (2006).
19. Beum, P. V. et al. Quantitative analysis of protein co-localization on B cells opsonized with rituximab and complement using the ImageStream multispectral imaging flow cytometer. J. Immunol. Methods 317, 90–99 (2006).
20. Konstantinidis, D. G. et al. Signaling and cytoskeletal requirements in erythroblast enucleation. Blood 119, 6118–6127 (2012).
21. Matsushita-Ishiodori, Y., Hanczyc, M. M., Wang, A., Szostak, J. W. & Yomo, T. Using imaging flow cytometry to quantify and optimize giant vesicle production by water-in-oil emulsion transfer methods. Langmuir 35, 2375–2382 (2019).
22. Shimizu, Y. et al. Cell-free translation reconstituted with purified components. Nat. Biotechnol. 19, 751–755 (2001).
23. Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. Methods 36, 299–304 (2005).
24. Sontag, C. A., Staley, J. T. & Erickson, H. P. In vitro assembly and GTP hydrolysis by bacterial tubulins BtubA and BtubB. J. Cell Biol. 169, 233–238 (2005).
25. Pilhofer, M., Ladinsky, M. S., McDowall, A. W., Petroni, G. & Jensen, G. J. Microtubules in bacteria: Ancient tubulins build a five-protofilament homolog of the eukaryotic cytoskeleton. PLoS Biol. 9, e1001213 (2011).
26. Kattan, J., Doerr, A., Dogterom, M. & Danelon, C. Shaping liposomes by cell-free expressed bacterial microtubules. ACS Synth. Biol. 10, 2447–2455 (2021).
27. Rowlett, V. W. & Margolin, W. The Min system and other nucleoid-independent regulators of Z ring positioning.

Front. Microbiol. 6, 478 (2015).
28. Rowlett, V. W. & Margolin, W. The bacterial Min system. Curr. Biol. 23, R553–R556 (2013).
29. Blanken, D., Foschepoth, D., Serrão, A. C. & Danelon, C. Genetically controlled membrane synthesis in liposomes. Nat. Commun. 11, 4317 (2020).
30. Nomura, S. I. M. et al. Gene expression within cell-sized lipid vesicles. ChemBioChem 4, 1172–1175 (2003).
31. Noireaux, V. & Libchaber, A. A vesicle bioreactor as a step toward an artificial cell assembly. Proc. Natl. Acad. Sci. U. S. A. 101, 17669–17674 (2004).
32. Saito, H., Kato, Y., Le Berre, M., Yamada, A., Inoue, T., Yosikawa, K. & Baigl, D. Time-resolved tracking of a minimum gene expression system reconstituted in giant liposomes. Chembiochem. 10, 1640–1653 (2009).
33. Sunami, T., Matsuura, T., Suzuki, H. & Yomo, T. Synthesis of functional proteins within liposomes. Methods Mol Biol. 607, 243–256 (2010).
34. Nourian, Z., Roelofsen, W. & Danelon, C. Triggered gene expression in fed-vesicle microreactors with a multifunctional membrane. Angew. Chemie Int. Ed. 51, 3114–3118 (2012).
35. Van de Cauter, L., Fanalista, F., Van Buren, L., De Franceschi, N., Godino, E., Bouw, S., et al. Optimized cDICE for efficient reconstitution of biological systems in giant unilamellar vesicles. ACS Synth. Biol. 10, 1690–1702 (2021).
36. Gonzales, D. T., Yandrapalli, N., Robinson, T., Zechner, C., and Tang, T. Y. D. Cell-free gene expression dynamics in synthetic cell populations. ACS Synth. Biol. 11, 205–215 (2022).
37. Vogele, K., Frank, T., Gasser, L., Goetzfried, M. A., Hackl, M. W., Sieber, S. A., et al. Towards synthetic cells using peptide-based reaction compartments. Nat. Commun. 9, 3862 (2018).
38. Seo, H. & Lee, H. Spatiotemporal control of signal-driven enzymatic reaction in artificial cell-like polymersomes. Nat. Commun. 13, 5179 (2022).
39. Niederholtmeyer, H., Chaggan, C. & Devaraj, N. K. Communication and quorum sensing in non-living mimics of eukaryotic cells. Nat. Commun. 9, 5027 (2018).
40. Loose, M., Fischer-Friedrich, E., Herold, C., Kruse, K. & Schwille, P. Min protein patterns emerge from rapid rebinding and membrane interaction of MinE. Nat. Struct. Mol. Biol. 18, 577–583 (2011).
41. Soengas, M. S., Gutiérrez, C. & Salas, M. Helix-destabilizing activity of phi 29 single-stranded DNA binding protein: effect on the elongation rate during strand displacement DNA replication. J. Mol. Biol. 253, 517–529 (1995).
42. Mencía, M., Gella, P., Camacho, A., de Vega, M. & Salas, M. Terminal protein-primed amplification of heterologous DNA with a minimal replication system based on phage Phi29. Proc. Natl. Acad. Sci. U. S. A. 108, 18655–60 (2011).

6

Future Endeavours in the Semi-rational Engineering of Synthetic Cells

Abstract

After quite a journey of billions of years and 166 pages, our arrival here was not a straight line. Evolution, driving optimization and survival, sculpted life through countless cycles of changes, adaptations, and resilience. Did we craft a living entity already? This ultimate goal is still on the table. However, throughout the chapters of this thesis, we aimed to contribute a valuable piece towards understanding life, and the eventual creation of a functional synthetic cell from scratch. In particular, we hope we convinced you, the reader (and synthetic cell enthusiast), about the power of evolution and the important role of module integration and coordination for creating a self-sufficient synthetic cell. In this final chapter, we want to provide the reader with insights and initial findings into what we consider would be interesting to keep exploring ahead. We discuss the semi-rational engineering approach to build up a robust evolutionary campaign for upcoming module integration or individual module optimization efforts. We include possible strategies for (i) building/improving DNA templates for evolution, including the introduction of genetic diversity, and (ii) screening, sorting, and recovering the most suitable variants. Finally, we apply some of this framework to the evolution of *DNArep-PLsyn* integrated modules, an immediate next step for what we presented in chapter 4 of this thesis.

The iterative design-build-test-learn (DBTL) process, commonly used for engineering any extant biological organism [1], also aligns well for the engineering of minimal bacterial cells [2]. However, dependence on a standard DBTL cycle only based on rational engineering strategies remains inadequate for creating a fully functional living entity from the bottom up. Therefore, integrating evolution as an engineering tool within the DBTL framework holds great potential to propel synthetic cell development.

In general, the main steps of an evolutionary campaign encompass creating genetic diversity and identifying preferred variants using a screening and selection process. However, in the realm of synthetic cell engineering, these latter steps are not the only primary ones to consider. As bottom-up synthetic cell crafters, we position the starting point and conditions for evolution. For now, we are the ones in control, defining (i) which bio modules are essential to explore and/or integrate, and from which organisms they come from, (ii) what the nature and sequence of our synthetic cell genome are (i.e., RNA or DNA) and how to build it, (iii) how to introduce genetic diversity into our genome to drive evolution, (iv) what do we want to evolve for, and (v) how to select and extract genetic information from the variants that we want. This flexibility, although somehow overwhelming, is actually convenient. Having some rational engineering choices can greatly assist in constructing a robust in vitro evolution platform.

The rational design behind synthetic cell evolution

Choosing the starting point

The first rational decision is selecting the starting point. This involves determining which module(s) should undergo evolution, and how their information should be encoded within a DNA construct. Already tested bio modules, either individually [3–6] or in conjunction (chapter 4 of this thesis), can serve as an ideal starting point for this process. However, the initial phase could also involve more suitable biological modules, strategically modified to initiate an evolutionary campaign. Making rational adjustments on the DNA design or the synthetic cell environment, to some degree, can help find a stronger basis to produce genetic diversity and initiate evolution. Now, let's explore some key aspects to ponder for the semi-rational engineering of a synthetic cell:

1. *Choosing the right genome design and assembly method.* When exploring synthetic cell individual modules or the integration of a few, selecting the right DNA template, including gene count, protein encoding sequence, orientation, regulatory mechanisms, DNA assembly approach, and more, has shown to be essential in achieving functional bio modules (chapter 4) capable of undergoing evolution (chapter 3). With the future holding integration, evolution, and an expected synthetic cell genome of probably more than 100 kb [7,8] , it is essential to refine our genome design and construction strategies to what comes ahead. On the scope of what we have learned so far: (i) multiple repetitive elements on a single DNA template can pose strong challenges for DNA self-replication processes and DNA cloning methods (chapter 3 and chapter 4), and (ii) gene-expression profiles can greatly influence the activity of protein machineries and phenotype occurrence (chapter 4).

Starting with the influence of repetitive elements on creating functional DNA templates, one approach to address this issue involves leveraging only in vitro assembly methods. Yet, as observed with the *DNArep-PLsyn* genome in chapter 4 of this thesis, in vitro techniques can introduce undesirable assembly variability. Looking ahead to the integration of more modules into the same genome, in vitro-only techniques would struggle to consistently build a long DNA genome (>100 kb). Managing both repetitive elements and template length/complexity issues might be more effectively handled if choosing the right organism and cloning vector for the in vitro assembled templates. Utilizing low copy number vectors, linear vectors, or Bacterial Artificial Chromosomes (BAC) alongside specific bacterial strains [9–13], such as those with deficient recombination [9,14], can be a successful arrangement to avoid cloning fails. For even longer genomes, recent BAC stepwise insertions [15] or yeast recombination assembly can be suitable and appear as efficient alternatives [16,17]. Koster et al., illustrate in [17] how yeasts could help build rationally designed synthetic genomes for synthetic cell development. Yet, when implementing yeast assembly strategies, subsequent efforts should still depict effective ways to cope with possible repetitive elements, isolate the DNA, and utilize it as template for gene expression within a synthetic cell framework.

Switching to the effects that gene-expression profiles can have on bio module activity (chapter 4), next-generation DNA templates could have a stronger control over gene expression by lessening the regulatory elements with an operon design [18], or with alternative regulatory sequences (i.e., SP6 or T3 promoters) [5,19]. Nicely, these strategies could also help to reduce the quantity of repetitive elements, aiding with the DNA assembly of current and upcoming larger genomes [20]. Additionally, these new designs can be complemented with genetic- (or light-) driven circuits for a stronger temporal control over gene expression [19,21–24]. On this note, previous work has showcased multiple stage cascades, AND gates, and negative feedback loops in cell-free systems and within liposome compartments utilizing core *E.coli* sigma factors and two RNA polymerases (T3 and T7 RNAP) [25]. Moreover, Kim et al., implemented a simpler scheme with an RNA-regulated genetic circuit utilizing T7 RNA polymerase (RNAP) and *E. coli* ribonuclease H (RNase H) in a cell-free system framework [26]. Finally, Baumshlager et al., presented an interesting alternative with a light-inducible T7 RNAP (Opto-T7RNAPs) for a dynamic control over gene expression [27].

2. *Improve or minimize gene-encoded protein machineries.* As synthetic cell builders, choosing the right machineries to perform the essential life function is crucial. We now have a great repertoire of protein machineries that can perform well enough to kick start an evolutionary campaign (chapter 3). Nonetheless, upcoming evolution and module

integration efforts could also benefit from choosing protein alternatives (i.e, protein homologs, or already engineered variants) that could outperform current protein systems. For instance, Salas et al., reported the enhancement in the replication activity of Phi29 DNAP by incorporating helix-hairpin-helix [$(HhH)_2$] DNA binding domains at DNAPs C-terminus [28] **(Fig. 1a)**. As DNAP[$(HhH)_2$] shows stronger DNA binding [28], future work could test the implementation of DNAP[$(HhH)_2$] version for more complex self-replicators with added bio modules (i.e., *DNArep-PLsyn* from chapter 4, and upcoming ones), and/or test the role of SSB and DSB auxiliary proteins for full-length self-replication activity.

Besides adding specific domains that can confer a functional advantage, evaluating protein alternatives from other organisms can also be part of the rational improvement of our current bio modules. For instance, the phosphatidylserine synthase (PssA) protein from *B. subtilis* is a smaller protein (177 aa), and can be a 'simpler' alternative to the *E. coli* protein variant that we now use (451 aa) [29,30]. As already illustrated in chapter 4, PssA catalyses the synthesis of phosphatidylserine (PS), an intermediate step for phosphatidylethanolamine in the Kennedy Pathway, utilizing CDP-DAG and L-serine as precursors. As a preliminary test of *B. subtilis* PssA performance on our synthetic cell context, we codon optimized the wild type DNA sequence [31] (Twist Biosciences, USA) and compared both PssA versions within CDP-DAG containing vesicles. Interestingly, we found that codon optimized *B.Subtills* PssA has a similar PS production activity than our *E.coli* PssA version, with an apparent higher production in the first hours of sample incubation (0-3 h) **(Fig. 1b)**. With additional biological repeats, follow-up assays could delve into testing *B. subtilis* PssA within our complete Kennedy pathway context. Implementing a nearly three times smaller PssA could serve as a backup for future DNA library designs with fewer DNA variants to explore.

Future research could also delve into alternative, yet simpler complete protein programs. For instance, implementing a shorter pathway for phospholipid (PL) synthesis than our current Kennedy pathway could enhance protein and PL production yields. A smaller pathway could decrease resource consumption for imminent module integration efforts, and facilitate upcoming evolutionary campaigns. In this context, future efforts could explore the functionalities of the lysophospholipid acyltransferases (LPLAT) family from the Land's cycle phospholipid remodelling pathway. LPLATs transfer an acyl group from acyl-CoA or acyl-ACP to an already head-modified lysophospholipid, requiring only one step for producing PS or PE. An interesting single-protein candidate to explore is the *E.coli* acyl-acyl carrier protein (ACP) synthetase (Aas) [32]. Interestingly, this enzyme is known to possess two activities: acyl-ACP synthetase and 2-acyl-glycerophosphoethanolamine acyltransferase. Thus, a free fatty acid and a lysophospholipid are required to produce PE **(Fig. 1c)**. In an attempt to kickstart this project, we already demonstrated that Aas protein can be effectively produced in the PURE system with GreenLys protein labelling **(Fig. 1d)**. Upcoming work should then

delve into protein activity within liposome compartments, growth assays, and module integration efforts.

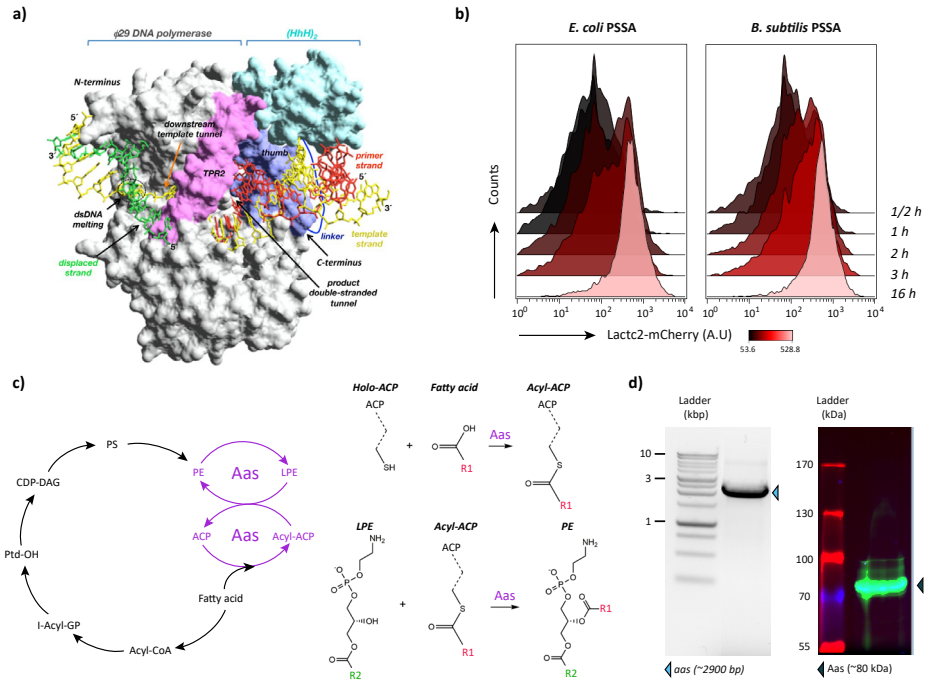


Fig. 1. a) Structural model taken from [28] of a Phi29 DNAP (grey) fused with an $(HhH)_2$ domain (cyan). **b)** Flow cytometry data of PS producing liposomes PURE-expressing PssA's protein versions from both *E. coli* and *B. subtilis*. **c)** On the left, schematic representation of the Kennedy Pathway (black) compared to the shortcut pathway (purple) proposed for the production of PE using the Aas protein from *E. coli*. On the right, schematic schemes of the Aas catalysed reactions: fatty acid transfer to Holo-ACP and LPE to produce PE. **d)** Gel electrophoresis of an Aas expressing DNA construct (~2900 bp) (left), and SDS-PAGE analysis of a GreenLys supplemented bulk IVTT reaction expressing the *aas* gene (right).

3. *Fine tune reaction components.* Finally, when exploring single or joint modules, rationally adjusting some reaction components is essential for setting up a feasible reaction background to kickstart evolution, reach higher-end functionalities, and eventually autonomy. This could either be with addition, or ideally, reduction of externally added components to the reaction. For example, before starting the intermittent evolution of a minimal self-replicator (chapter 3), we showed that compartmentalized DNA self-replication (with a starting concentration at λ = 0.2) remained unaffected after removing the purified double-stranded binding protein (DSB) from the reaction components. Curious about this result, we decided to also test DSB removal for the DNA self-replication module of *DNArep-PLsyn* genome (chapter 4). Under DNA replication conditions only (without the addition of PLsyn substrates), and a starting DNA concentration of 500 pM, we observed similar amplification levels of the *DNArep-PLsyn* genome with or without DSB **(Fig. S1a,b)**. Future work could delve further into this by performing additional replicates, including PLsyn substrates, and attempting to

recover the full-length DNA from the liposome suspension (conditions tested already in chapter 4 of this thesis). Possibly, reactions with no DSB added could be a suitable starting point for the evolution of DNArep-PLsyn genome.

In addition to removing components, the performance of gene expression in PURE or certain bio modules could be enhanced by introducing new or adjusting current concentrations of reaction components. For instance, if implementing alternative promoters for better control over gene expression, or if needing better protein folding or activity on a specific bio module, the PURE system reaction can easily be supplemented with commercially available RNA polymerases, chaperons (i.e., DnaK mix), and substrates/cofactors (i.e., GTP, ATP, CoA, and more). Moreover, when acquiring a commercial PURE system, its components can also be tailored in consultation with the manufacturer.

Moreover, in a recent work, Bartelds et al., reported that the intrinsic stochasticity of protein expression within cell-like compartments could be ameliorated upon addition of purified MazF, which enhances the degradation rate of mRNA up to 18-fold [33]. This could be useful for module integration efforts and evolutionary campaigns, where a high noise could lead to low phenotype occurrence (chapter 4), and/or unwanted false-positives/negatives conducting to a 'non-convergent' evolution process. Moreover, Sheahan et al. noted higher PURE protein production yields when they added purified ribosomal protein S1 to the reaction [34]. This protein, involved in translation initiation, aids recruitment of the mRNA to the 30S ribosomal subunit, thus helping translation efficiency. Finally, recent developments on active learning workflows can accelerate the exploration of reaction components (nature and concentration) at large scale with a specific optimization target [35].

There are several aspects that we haven't mentioned here that could be explored to optimize both individual modules and their integrated versions. Nonetheless, as also discussed in the introductory chapter of this thesis, we do not see rational engineering as the sole tool for creating a synthetic cell. There are so many parameters to explore, and the sequence-structure-function relationship is still so poorly understood that we cannot exclusively rely on a rational engineering approach. It serves as a valuable starting point, yet the construction paradigm must require an element of randomness. This is where evolution comes into play [36] for synthetic cell development.

Genetic diversification, essential for any evolution campaign (chapter 3) can be attained by traditional techniques like random mutagenesis with an error-prone DNA polymerase, in vivo mutagenesis, random DNA shuffling or recombination for generating new random chimeric proteins [37–41]. These are long-standing interesting approaches, which we still think are valuable for the optimization of one or a few biological modules (as illustrated in chapter 3 with the evolution of the DNA replication machinery). Yet, when it comes to larger protein complexes and multiple functionalities, only relying on conventional mutagenesis

techniques for introducing genetic diversity might be insufficient. The exploration space and library size can become incredibly large, with high quantity of inactive or mediocre performing variants present. Finding optimal performing variants can thus become time-consuming, expensive, and possibly unfeasible.

To address the overwhelming increase of sequence-space upon module integration, a useful strategy is to prioritize an 'intelligent' DNA diversification approach for building smaller but smarter DNA libraries. By employing computational tools, rational design strategies, and knowledge-driven approaches, a more compact library can be tailored to encompass a diverse, yet purpose-driven selection of variants [42–48]. This thorough curation can (i) optimize resources, (ii) cope with the library size capabilities of our synthetic cell framework, and (iii) enhance the probability of identifying superior traits or functionalities within the constrained variant pool. Next, the decision lies on whether the target is the protein encoding sequence(s) or/and the gene-expression regulation space. This decision is conditional to the targeted machinery and optimization objective. On the one hand, if focusing on the protein space, we can explore the protein encoding sequence (CDS) to target protein(s) kinetic parameters, stability, and/or binding sites. On the other hand, if focusing on protein production dynamics, both CDS and regulatory element DNA sequences can be explored [49]. However, considering the large size of the protein encoding sequence, modular regulatory elements (such as promoters, terminators, and RBSs) are generally preferred as gene-expression engineering targets [48,50,51].

Starting with the protein encoding sequence, machine-learning (ML) based algorithms/tools are now dictating the structure of concise and strategic libraries [52]. Some of these tools can be efficient at predicting protein structures from the amino acid sequence or protein homologs, while some others need to be fed with an existing protein structure, which is not always ensured, especially for membrane proteins [47,53–56]. Fortunately, tools like AlphaFold, which have proved their strong accuracy on building sequence-structure models can be a great aid to either feed protein structures to library generating tools, or for estimating structural effects of user-defined mutations [57,58]. Recently, Yang et al., also reported how machine-learning can utilize the pool of unimproved variants from directed evolution campaigns to build accurate sequence-function models, and depict better variants to screen [44]. Clearly, AI-driven tools are great helpers to further explore for choosing the right library design when focusing on the protein encoding sequence. Continuing with a library design for optimizing gene expression, both promoters and *E.coli* RBSs have been studied in vivo for controlling/connecting genetic circuits and adjusting metabolic fluxes to increase product production yields [59,60]. Yet, in our synthetic cell framework (PURE system), where translation has been identified as a primary limiting step [61,62], the RBS region (16 bp) surrounding a conserved shine Dalgarno (SD) sequence, can be an ideal target for directing gene-expression-focused evolution. Synthetic biology's drive to enhance biological product yields through metabolic flux optimisation, has resulted in the development of great tools to help design compact but smart RBS libraries. Though designed primarily for in vivo use,

these tools might extend their utility to our synthetic cell context. An already well-known tool is the Salis Lab RBS calculator, developed in early 2000s for different bacterial strains [59,63]. This online available software, called 'De Novo DNA' helps design RBS libraries with varied translation rates by exploring different RBS and CDS sequences. In 2021, Vezeau and Salis highlighted the RBS calculator's ability to forecast translation in cell-free extracts, unravelling its potential use for optimizing synthetic cell bio modules synthesized within a cell-free environment [62].

Designed DNA libraries, whether targeting the protein encoding sequence or its regulatory elements only, can be obtained by different means. Perhaps the most straightforward one can be to introduce mutations by PCR amplification with primers containing the specific mutation/DNA change, or with degenerative primers [64,65]. However, these strategies are usually implemented on a continuous DNA region, and on relatively small DNAs when compared to what a synthetic cell genome could be in the future (>100 kb) [66,67]. While sequential PCRs could partially address these limitations, future efforts might as well explore alternative mutagenesis techniques that handle better the size of our envisioned synthetic cell genome, and the introduction of mutations at distinct locations. Multiplex Automated Genome Engineering (MAGE), developed by Wang et al., in 2009, can be a robust method for this purpose [68]. Already explored for bacterial genomes, artificial chromosomes (BAC), and plasmids, MAGE introduces ss/dsDNA with degenerate bases into bacterial cells for the simultaneous modification of targeted DNA locations [69]. The mutagenesis efficiencies can widely vary depending on the nature and quantity of the mutations, DNA oligo design, and intrinsic DNA repair mechanism from the organism. Nonetheless, previous and current research continues to explore alternatives to enhance MAGE efficiencies by increasing the number of MAGE cycles [68,70], introducing CRISPR aids (CRMAGE) [71], or by integrating all necessary engineering components into a broad-host-vector (pORTMAGE) [72]. Clearly, MAGE has become a widely adopted method for in vivo metabolic engineering and holds great promise as a valuable tool for synthetic cell engineering.

Choosing right on screening, selection, and DNA isolation strategies.

Having a robust genetic diversification strategy is crucial for an efficient in vitro evolutionary campaign. Nonetheless, it has to be well accompanied by robust phenotype screening and selection schemes. The first essential step is to establish a clear genotype to phenotype link, as it enables the correct identification and isolation of variants with desired traits among the pool of all diverse candidates. In our synthetic cell framework, we proved that a genotype to phenotype link could be well established by DNA dilution such that the quantity of DNA molecules per liposome is one or less (λ = 0.2 or 1), which corresponds to a starting DNA concentration of 10 pM (λ = 0.2) or 50 pM (λ = 1) (chapter 2 and 3). If some phenotypes of interest only appear at a higher DNA concentration due to higher protein expression needs, CADGE could be a great help for clonally increasing the DNA concentration per liposome, while enabling higher protein expression and phenotype occurrence and/or performance

(chapter 2).

Moving forward, future efforts could focus on enhancing our existing abilities in vesicle production and refining screening/sorting methods. In such a way, the smart reduction of our DNA library size combined with a scaled-up vesicle production could lead to an efficient coverage of all DNA variants within only one or a couple of in vitro evolution rounds. Although scaling up liposome production is feasible, it is not easily manageable with nearly all liposome production methods. Certain techniques exhibit greater proficiency in generating a large number of liposomes per PURE reaction (typically 20 µl) within a specific time frame. For instance, the high-throughput production of monodisperse giant vesicles is often associated with droplet-based microfluidic technologies [73,74]. However, this method frequently employs organic solvents (e.g., chloroform, hexane, octanol), oil, and materials that can adversely affect cell-free expression and the performance of certain bio modules upon transient exposure (i.e., membrane-incorporated proteins). As an immediate, and probably sufficient alternative (considering the decrease in library sizes from a smart design), we can explode our current gentle swelling liposome production method. We could aim at multiplexing our current production strategy by scaling up in parallel the quantity of microliter-sized reactions (usually max 20 µl), or by changing our actual settings to have a larger surface area for lipid swelling. This can come in hand with the optimization of the amount of lipids added to coat the glass beads or the selected alternative surface.

Given a robust core on genome design, DNA diversity, genotype to phenotype link, and vesicle production, the screening and selection strategies must be powerful enough to identify desired synthetic cell variants. Traditionally, whether in cellulo or within in vitro compartmentalized reactions, fluorescence-activated cell sorting (FACS) has been the main methodology to enable the rapid assessment and isolation of variants displaying improved properties. However, FACS relies on low-resolution data (i.e., light scattering and fluorescence intensity signals), which limits its application to advanced synthetic cell functions [75]. When envisioning the evolution of more complex phenotypes, and/or upon bio module integration, FACS would not be enough. Developing complex functions in vitro will probably require a multidimensional image-guided screening and selection approach. To address this limitation for upcoming evolutionary campaigns, recently developed technologies, such as Intelligent image-activated cell-sorting (IACS), can be a great tool to explore [76,77]. This technology, initially developed by the Goda Lab for human cells and small organisms, combines high-throughput cell microscopy with intelligent image-based cell screening and sorting (all within 32 ms). Recognizing the potential of this technology for synthetic cell engineering, we decided to assess if IACS could serve as the primary screening and selection platform for our in vitro evolution campaigns. As detailed in [76], the IACS utilizes four optical interrogation points (OI1-OI4), generating forward scatter (FSC) signals, higher than a predetermined threshold, needed for an accurate cell recognition, imaging, and precise sorting. Interestingly, although not entirely surprising, preliminary runs on the IACS with PURE containing vesicles showed very poor optical detection at

any interrogation point. This is probably due to the low refractive index, hence low FSC signals, from the synthetic vesicles which do not (yet) contain multiple/complex cellular structures as actual cells do. Therefore, utilizing IACS for our synthetic cell evolutionary campaigns will require to increase the refractive index of the liposome lumen for generating detectable FSC values. To do so, we are currently investigating the co-encapsulation of Iodixanol (commercially available as OptiPrep™) [78], trehalose/glucose sugar gradients, or polystyrene and silica nanoparticles. For now, we have run pilot experiments on the IACS with liposomes containing homemade PURE buffer or the complete PURE system with either OptiPrep or trehalose. Interestingly, despite variabilities from run to run, we found that both OptiPrep and a trehalose/glucose gradient can help increase the amount of detected liposome events on all interrogation points (OI1-OI2-OI3), with a preference for OptiPrep **(Fig. S2a)**. However, detected events were quite sporadic when compared to the standard capabilities of the IACS [76]. Moreover, both OptiPrep and trehalose appear to hinder gene expression in PURE, with OptiPrep exhibiting a more pronounced effect **(Fig. S2b)**. Despite this limitation, we were able to detect and collect images from bacterial microtubule formation inside our synthetic cell context **(Fig. S2c)**. For this, we encapsulated high concentrations of fluorescently labelled BtubA and BtubB, along with the complete PURE system and trehalose at 1 M. Upcoming interesting experiments could utilize this collected data as a training set for machine learning models to define a sorting function based on microtubule features (i.e., with mock DNA libraries as in chapter 2). These experiments could help evaluate vesicle screening and selection capabilities of IACS, while helping devise ideal DNA recovery and quantification strategies for improving screening and sorting efficiency. Moreover, future research could also explore alternative approaches for augmenting liposome refractive index and FSC signal without compromising gene-expression. With this, IACS can undergo further testing to screen and sort gene-expressing vesicles that encode and display either single or integrated synthetic cell modules. This line of research will be pursued in collaboration with the Goda lab from the University of Tokyo, where I stayed in the summer 2022, during which we collected the data shown in **(Fig. S4a,c)**.

The immediate evolutionary case: evolution of DNArep-PLsyn

Having reviewed various options for performing evolutionary campaigns on synthetic cells, let's now explore possible experimental milestones for evolving DNArep-PLsyn integrated modules, already presented in chapter 4 of this thesis **(Fig. 2)**. Our immediate optimization goal is to increase the occurrence of liposomes displaying joint DNArep-PLsyn phenotypes. Further on, our evolutionary campaign can focus on enhancing DNArep and PLsyn efficiency to achieve significant vesicle growth.

The DNA template

Our current *DNArep-PLsyn* genome design is an interesting initial stage for evolution, as it enables joint-phenotype occurrence on hundreds of liposomes from a few microliters

(chapter 4). Nonetheless, it relies for now on an in vitro assembly strategy, which could potentially lead to unwanted variabilities in phenotype performance along the evolutionary campaign. Upcoming work could help tackle this by attempting other DNA cloning strategies that could better cope with template complexity and quantity of repetitive elements. For instance, different bacterial strains and low-copy number vector designs, may be considered [9]. Alternatively, or in parallel, we could also focus on exploring a different *DNArep-PLsyn* genome design with other regulatory elements for avoiding possible recombination events when cloning. On this note, earlier work in our lab by former postdoc David Foschepoth has shown successful cloning of a *DNArep-PLsyn* template utilizing an SP6 promoter instead of a T7 for controlling the transcription of the *DNArep* machinery. Here, a PmeI restriction site was added between the two origins of replication, such that the linear DNA template for IVTTR can be produced by enzymatic digestion instead of PCR. Future work could focus on finding the right balance for both SP6 and T7 co-expression, such that *DNArep-PLsyn* self-replication can be accomplished with a SP6 driven DNArep module. Moreover, self-replication could be enhanced by implementing the previously discussed DNAP[$(HhH)_2$] variant instead of the WT DNAP, or by integrating on DNAP some of the enriched mutations from our replicator evolution in Chapter 3 (i.e., DNAP(S79G) or DNAP(A80T)). These rational and reverse engineering approaches could be useful to either ameliorate replication on our current T7-based *DNArep-PLsyn* replicator or kickstart *DNArep-PLsyn* self-replication on a SP6-driven DNArep module.

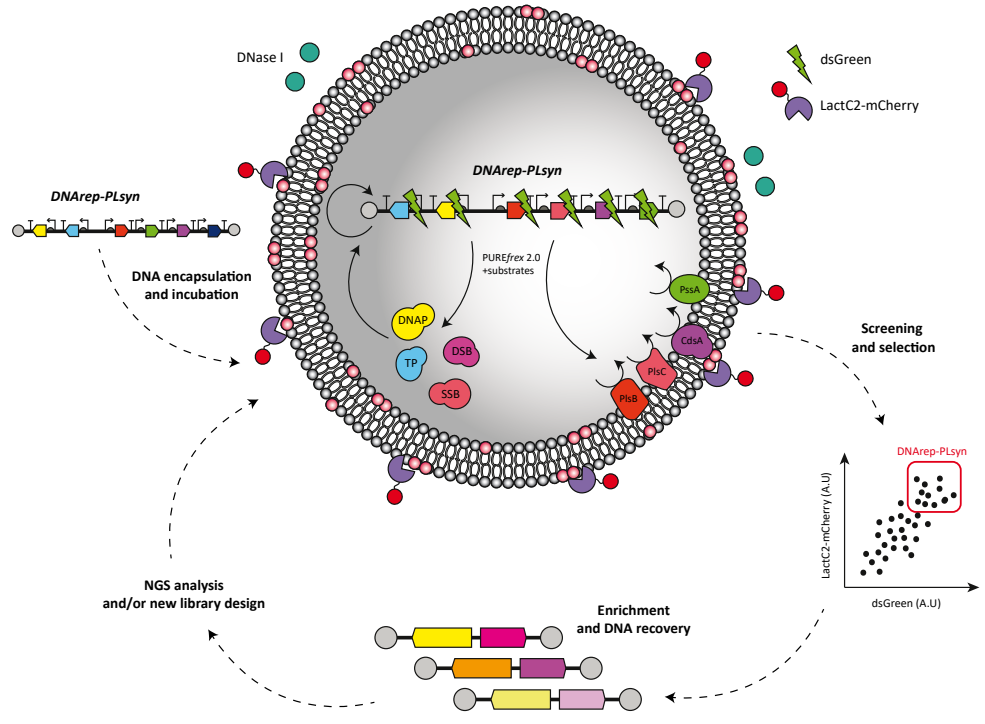


Fig. 2 Schematic illustration of an intermittent evolutionary campaign for DNArep-PLsyn integrated modules. Both

LactC2-mCherry and dsGreen labelling probes (for PL production and DNA self-replication, respectively), can serve for liposome screening and sorting with FACS.

To introduce genetic diversity on the *DNArep-PLsyn* genome, either with an SP6 or complete T7 regulation (chapter 4), we have considered two complementary approaches. Firstly, we envisaged smart RBSs libraries on the *DNArep* and/or *PLsyn* modules from the *DNArep-PLsyn* genome, having mutations introduced with MAGE as the main technique [68,70]. Utilizing the RBS calculator from the Salis lab, we've already started on this quest by designing small and smart libraries targeting the *DNArep* module covering a wide range of predicted initiation translation rates for both *p2* and *p3* genes encoding for DNAP and TP proteins **(Fig. 3a)(Table S1)**. Besides the implementation of *DNArep* RBS libraries, upcoming efforts can also focus on utilizing *PLsyn* RBS libraries to explore a better gene-expression balance for higher joint phenotype occurrence, and possibly higher and/or faster overall production of PS. Secondly, we've considered the in situ introduction of mutations upon DNA self-replication and PCR DNA recovery, as explored for a minimal self-replicator on chapter 3. This could be accomplished with a WT DNAP or a DNAP mutator, such as DNAP(F62Y), which both have proven to be effective at introducing diversity across a self-replicator evolutionary campaign (chapter 3).

Screening and selection of best variants

The screening and selection of high performing vesicles can be accomplished by FACS, after labelling DNA with dsGreen fluorescent probe, and PS production with LactC2-mCherry. The gate for sorting the best performing variants can correspond to the top 1% (or lower) vesicles displaying joint-phenotypic traits (ROI 2 in chapter 4) under a low starting DNA concentration (10-50 pM) to ensure a strong genotype to phenotype link **(Fig. 2)**. On the latter, preliminary results have already shown that for a whole T7 based *DNArep-PLsyn* genome, 100 pM is good enough to produce both DNArep and PLsyn phenotypes **(Fig. 3b,c)**, which sets a promising ground for lower DNA concentrations (50 pM or 10 pM). DNA recovery after sorting can be attempted as previously illustrated in chapter 4 for the *DNArep-PLsyn* genome. Considering that starting DNA concentrations are ten times lower than what was utilized in chapter 4, an active DNA replication can greatly facilitate the recovery of DNA (as shown for orthogonal DNA amplification on chapter 2). If, despite active DNA replication, recovery becomes troublesome, we can vary the selection approach between rounds, avoiding exclusive dependence on FACS as the primary screening and selection method. For instance, the first couple of evolutionary rounds could mainly focus on enriching for better DNA self-replication under joint-module reaction conditions. Selection can rely on DNA self-amplification efficiencies as performed in chapter 3 for a minimal self-replicator, and we enrich and select for self-replicators capable of replicating the entire *DNArep-PLsyn* template by PCR recovery in between rounds. On this note, the *DNArep-PLsyn* PCR recovery from chapter 4, and preliminary liposome experiments with a longer *DNArep-PLsyn* DNA **(Fig. 3d)** have demonstrated the feasibility to recover

the complete genome from a liposome suspension when starting with higher DNA concentrations, which could be accomplished by an active DNA replication machinery. With this initial replication-based screening and selection strategy, we do not rely on FACS to kickstart the evolutionary campaign, but we still expose the DNArep module into the environmental conditions of coupled DNArep and PL production.

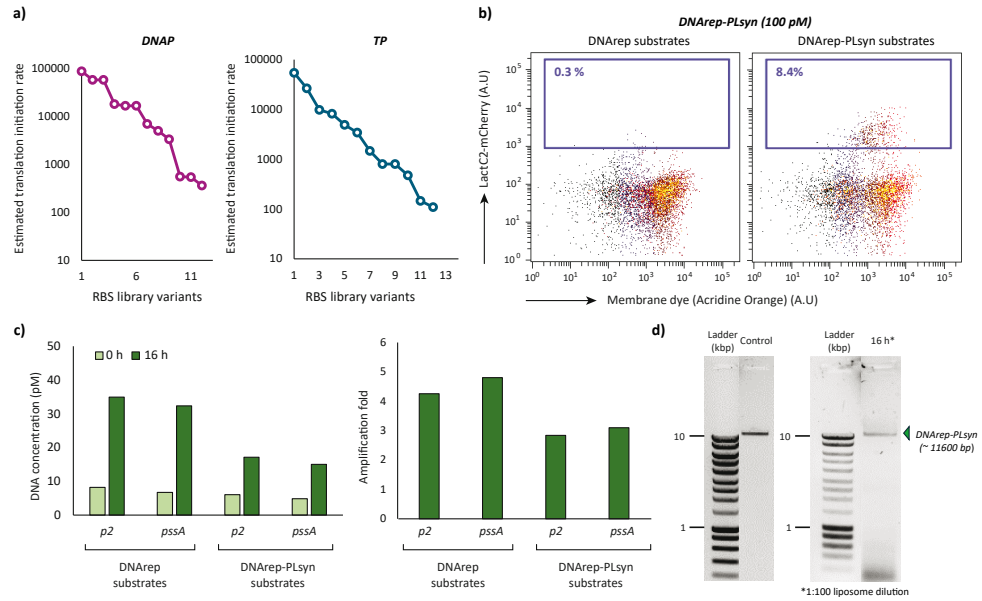


Fig. 3 a) Estimated translation initiation rate profile for an RBS generated library controlling both *p2* and *p3* genes. This library was designed with De Novo DNA software [59,63] and the DNA sequences can be found on **(Table. S1)**. **b)** Flow cytometry scatter plots displaying PS producing liposomes from encapsulated *DNArep-PLsyn* genome under a 100 pM DNA concentration and added substrates for both: PLsyn-DNArep, and DNArep only bio modules. **c)** Absolute DNA quantification by qPCR and calculated amplification fold (16 h DNA concentration / 0 h DNA concentration) from panel (b) samples with DNArep-PLsyn and DNArep only added substrates. qPCR target regions (~200 bp) are from *p2* and *pssA* genes. **d)** Agarose gel electrophoresis of a PCR recovered *DNArep-PLsyn* genome (~11600 bp and SP6 regulated *DNArep* module) from a 1:100 diluted liposome sample. The 'control' lane corresponds to the PmeI linearized version of the *DNArep-PLsyn* cloned construct.

In the future, our DNArep-PLsyn system could be utilized to explore vesicle membrane growth as an optimization target. Despite the reliability of PS production and the associated LactC2-mCherry labelling as a measure of PL biosynthesis (chapter 2 and 4), it remains an indirect method and might be inaccurate to detect high PS content as a result of gene-encoded liposome growth [5]. Moreover, PS is still a precursor to the most abundant phospholipids in an *E.coli*-based membrane composition: phosphatidylethanolamine (PE) and phosphatidylglycerol (PG). To eventually aim for a significant gene-encoded liposome growth, our current four-protein Kennedy Pathway on *DNArep-PLsyn* genome could be complemented with the proteins needed to produce PE and PG, which have already been reconstituted in liposomes [5]. Then, the incorporation of internally synthesized PE and/

or PG phospholipids leading to liposome growth can possibly be monitored by Förster resonance energy transfer (FRET). As an example, Vogele et al. already employed a Cy3-Cy5 FRET pair to showcase the growth of peptide vesicles. By monitoring the decrease in the acceptor signal (Cy5) and the increase in donor fluorescence (Cy3), they were able to identify membrane expansion resulting from the integration of gene-expressed elastin-like peptides into the membrane [79].

General conclusion

Overall, we hope that this thesis serves as an inspiration for future research on the development of synthetic cells using in vitro evolution. Specially, we hope that we've convinced the synthetic cell enthusiast readers to indulge into module integration efforts, high-throughput screening/characterization techniques, and implementation of evolution as a great tool for the optimization of single or joint modules in a synthetic cell context. Four years, and now 180 pages later, we've showcased the enhancement of in vitro compartmentalized directed evolution platforms with clonal DNA amplification strategies, demonstrated the evolution of a minimal DNA self-replicator, integrated a functional central dogma (transcription-translation-replication) and membrane biosynthesis within synthetic cell compartments, implemented high-throughput imaging strategies for characterizing synthetic cell phenotypes, and finally discussed future endeavours that combine rational design and evolutionary strategies for synthetic cell development. As we wrap up this final section, we encourage the synthetic cell community to stay creative, look into nature's cues, and continue exploring the amazing power of evolution for crafting artificial life from the ground up.

Materials and methods

Buffers and solutions

All buffers were made with MilliQ grade water with 18.2 MΩ resistivity (Millipore, USA). All chemicals were purchased from Sigma-Aldrich unless indicated otherwise.

DNA constructs

DNArep-PLsyn genome with a T7-based *DNArep* and *PLsyn* modules was assembled in vitro as described already in chapter 4 of this thesis. *DNArep-PLsyn* genome with an SP6 regulated *DNArep* module, and *DNArep* DNA derived from G363 and G435 plasmids, respectively and as described in chapter 4. Plasmids G322 (harbouring a *gfp* gene **(Fig. S2b)**), and G368 (harbouring *E. coli pssA*) were constructed with Gibson Assembly by subcloning the gene of interest (*gfp* of *pssA*) into a Phi29 origins-flanked pUC57 vector. Plasmids G616 (harbouring codon optimized *pssA* from *B. subtilis*) and G401 (harbouring codon optimized *aas* from *E. coli*) were synthesized with Twist Bioscience (USA) into a high copy Phi29 origins-flanked vector. Linear DNA fragments containing either *pssA* or *aas*-expressing constructs were prepared by PCR with primers 491 and 492 ChD as also described in previous chapters (2-4). DNA sequences are available upon request.

Purification of SSB, DSB, LactC2-mCherry, and BtubA&B

Purified SSB, DSB, and LactC2-mCherry proteins were produced and stored in −80 °C as previously described in chapter 2-4. BtubA&B purified proteins were produced and stored as described in ref [4].

IVTT reactions and liposome encapsulation.

Bulk IVTT reactions and GreenLys SDS-PAGE assays were performed as illustrated in previous chapters. For in-liposome reactions, lipid coated beads and liposome encapsulation were carried out as indicated in the previous chapters for specific phenotypic trait(s) such as PS production or DNArep-PLsyn joint phenotypes with minor modifications (chapter 2-4). PssA assays **(Fig 1)** were performed at 37 °C with a linear ori-flanked *pssA* construct at 50 pM concentration and CDP-DAG containing lipid-coated beads (chapter 2). DNArep-PLsyn assays **(Fig. 3)** were performed with substrates for DNArep and/or PLsyn and incubated at 37 °C for joint DNArep-PLsyn phenotypes, and at 30 °C for DNArep only. The *DNArep-PLsyn* genome concentration was kept 500 pM unless otherwise indicated (i.e., experiments with 100 pM **(Fig. 2 b)**). Liposome-protocol modifications were introduced on IACS preliminary experiments. Lipid coated beads were generally prepared with our standard lipid composition (chapter 2-5) and no membrane label. Swelling solutions were often complemented with either trehalose or OptiPrep at indicated concentrations **(Fig. S2)**. To achieve such a high sugar concentration, a highly concentrated trehalose stock (1.3-5 M) was freshly prepared every time, heated up to fully dissolve, and let to cool down before adding it to the liposomes swelling solution. Gene-expression-inhibition assays **(Fig. S2b)** were performed with Texas-red labelled lipid coated beads and G322 plasmid, encoding for GFP, as DNA template (5 nM). Refractive-index assays, performed with PB and FITC-containing liposomes **(Fig S2a)**, had a liposome-swelling time of 2 hours with manual tube-tumbling every 15-20 minutes. IACS-imaging assays with PURE-containing liposomes and co-encapsulated purified bacterial tubulin (488 BtubA&B (2.7 µM) + unlabelled BtubA&B (0.9 µM)) **(Fig S2 c)**, were manually and continuously tumbled during ~30 minutes on ice for liposome swelling.

Flow cytometry

Flow cytometry runs were performed as indicated in previous chapters for the specific phenotype or expressed protein to be detected, with minor modification on some sample preparations. For PssA analysis **(Fig. 1 b)**, the liposome suspension was diluted in a 1:100 ratio with a freshly prepared LactC2-mCherry-PB solution to obtain a final LactC2-mCherry concentration of ~320 nM and a minimal final volume of 150 µl. For DNArep-PLsyn experiments with 100 pM genome concentrations **(Fig. 2 b)**, the liposome suspension was diluted in a 1:100 ratio in a PB solution with both LactC2-mCherry and Acridine Orange (A.O) for a final volume of 150 µl and final concentrations of 300 nM for LactC2-mCherry and 6 µM of A.O. Flow cytometry data analysis was performed as indicated in previous

chapters (Chapter 2-4).

IACS

IACS experiments were performed with Mika Hayashi (PhD student from Goda Lab) following the specifications of the instrument as illustrated in ref [76]. Before every run, the liposome suspension was diluted in a 1:300 ratio with a phosphate-saline buffer (PBS) for OptiPrep assays or glucose-PBS solution for trehalose assays. The glucose concentration was adjusted depending on the trehalose concentration utilized to prepare the liposomes.

Recovery of DNArep-PLsyn DNA from liposomes.

A 1 nM *DNArep-PLsyn* (SP6 regulated DNArep) liposome suspension was diluted in MilliQ water at a 1:100 ratio. A standard 10-20 µl PCR reaction was prepared with final concentrations of 1X Platinum™ SuperFi™ buffer (Thermo Fisher), 20-40% of liposome:MilliQ dilution, 200 µM of dNTPs (Thermo Fisher), 200 nM of each primer (491 and 492 ChD), and 0.04 U/µl of Platinum SuperFi DNA Polymerase. The PCR reaction was incubated in a thermal cycler programmed to follow: 30 sec at 98 °C, 25-32 cycles of (98 °C for 10 sec, 64 °C for 15-20 sec, 72 °C for 6 min), and 72 °C for 5 min. PCR products were then analysed by agarose gel electrophoresis.

Acknowledgements

We thank Brent van Schagen for his bachelor end project work on evaluating *E.coli* and *B. subtilis* PssA protein variants. We are grateful to Marijn van den Brink for conducting pilot FACS experiments to assess the impact on gene expression when trehalose and OptiPrep reagents are co-encapsulated within liposomes. We extend our gratitude to Akihiro Isozaki, Mika Hayashi, Tsubasa Kobayashi, and Professor Keisuke Goda for a great welcoming and guidance to perform pilot liposome-IACS experiments in Tokyo during summer 2022. This summer-scientific internship was funded by the European Molecular Biology Organization (EMBO) as part of their Scientific Exchange Grants fund.

Supplementary Information

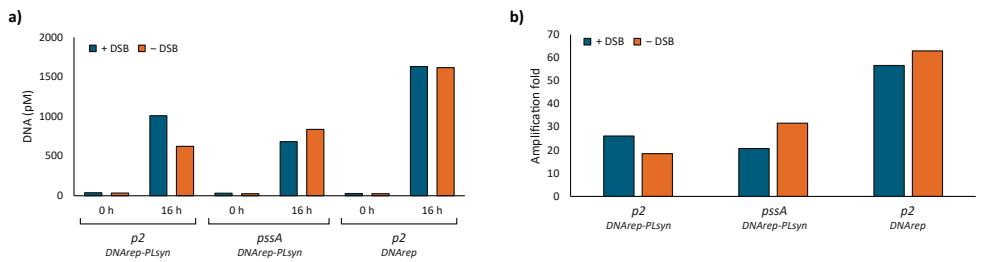


Fig S1. a) Absolute DNA quantification by qPCR on *DNArep-PLsyn* and *DNArep* containing liposomes with DNArep substrates and with or without DSB protein. **b)** Calculated amplification folds (16 h DNA concentration / 0 h DNA concentration) from panel (a) samples. qPCR target regions (~200 bp) are from *p2* and *pssA* genes.

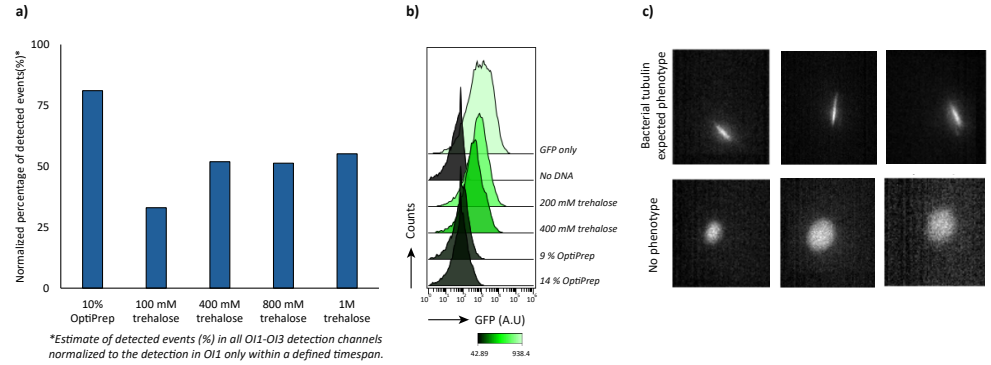


Fig. S2 a) Estimate of detected events in all three IACS detection points (OI1-OI3) with variable trehalose and OptiPrep concentrations within PB-FITC containing liposomes. **b)** Flow cytometry data of PURE containing liposomes expressing GFP with or without co-encapsulated trehalose and OptiPrep. **c)** Recorded IACS images from PURE containing liposomes with co-encapsulated trehalose (1 M) and high concentrations of purified 488-labelled bacterial tubulin (Atto488-BtubA&B).

	DNAP	
	GGGAGACCACAACGGUUUCCCUCUAGAAAUAAUUUUGUUUAACUUUAADAMGGAGSUAUACAU	
# of variants	**DNA sequence of each variant**	**translation Initiation rate**
1	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAATAAGGAGGTATACAT	86579,07
2	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAAAAGGAGGTATACAT	57484,61
3	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGGTATACAT	57484,61
4	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAATACGGAGGTATACAT	17919,70
5	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAAACGGAGGTATACAT	16525,30
6	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGACGGAGGTATACAT	16525,30
7	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAATAAGGAGCTATACAT	6933,09
8	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGCTATACAT	4991,68
9	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAAAAGGAGCTATACAT	3329,20
10	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAATACGGAGCTATACAT	555,19
11	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAGACGGAGCTATACAT	542,84
12	GGGAGACCACAACGGTTTCCCTCTAGAAATAATTTTGTTTAACTTTAAAACGGAGCTATACAT	362,04
	TP	
	GGGCCCUCUGGAGACACCAGAGGGUUUACAUGUUUAUUUGUUUAACUUUAAGAAGSRGBUUGACUA	
# of variants	**DNA sequence of each variant**	**translation Initiation rate**
1	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGGAGGTTGACTA	54293,59
2	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGGGGGTTGACTA	26544,83
3	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGGAGTTTGACTA	9818,17
4	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGGGGTTTGACTA	8223,96
5	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGGAGCTTGACTA	4909,47
6	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGGGGCTTGACTA	3425,10
7	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGCAGGTTGACTA	1476,31
8	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGCAGCTTGACTA	800,50
9	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGCAGTTTGACTA	800,50
10	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGCGGGTTGACTA	477,08
11	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGCGGTTTGACTA	145,68
12	GGGCCCTCTGGAGACACCAGAGGGTTTACATGTTTATTTGTTTAACTTTAAGAAGCGGCTTGACTA	109,51
custom one (13)	GACACCAGAGGGUUUACAUGUUUAUUUGUUUAACUUUAAGAAGGAGGUUUACUA	110312,50

Table S1. DNA sequences from the designed RBS variants shown on **(Fig. 3a)** for both *p2* and *p3* genes.

References

1. Kitano, S., Lin, C., Foo, J. L. & Chang, M. W. Synthetic biology: Learning the way toward high-precision biological design. PLoS Biol 21, e3002116 (2023).
2. Hutchison, C. A. et al. Design and synthesis of a minimal bacterial genome. Science 351, aad6253 (2016).
3. Godino, E. et al. De novo synthesized Min proteins drive oscillatory liposome deformation and regulate FtsA-FtsZ cytoskeletal patterns. Nat Commun 10, 4969 (2019).
4. Kattan, J., Doerr, A., Dogterom, M. & Danelon, C. Shaping Liposomes by Cell-Free Expressed Bacterial Microtubules. ACS Synth. Biol. 10, 2447–2455 (2021).
5. Blanken, D., Foschepoth, D., Serrão, A. C. & Danelon, C. Genetically controlled membrane synthesis in liposomes. Nat Commun 11, 4317 (2020).
6. Godino, E. et al. Cell-free biogenesis of bacterial division proto-rings that can constrict liposomes. Commun Biol 3, 1–11 (2020).
7. Mushegian, A. R. & Koonin, E. V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proceedings of the National Academy of Sciences 93, 10268–10273 (1996).
8. Forster, A. C. & Church, G. M. Towards synthesis of a minimal cell. Molecular Systems Biology 2, 45 (2006).
9. Shepherd, T. R. et al. De novo design and synthesis of a 30-cistron translation-factor module. Nucleic Acids Research 45, 10895–10905 (2017).
10. Chakiath, C. S. & Esposito, D. Improved recombinational stability of lentiviral expression vectors using reduced-genome Escherichia coli. BioTechniques 43, 466–470 (2007).
11. Godiska, R. et al. Linear plasmid vector for cloning of repetitive or unstable sequences in Escherichia coli. Nucleic Acids Research 38, e88 (2010).
12. Wood, W. N., Smith, K. D., Ream, J. A. & Lewis, L. K. Enhancing yields of low and single copy number plasmid DNAs from E. coli cells. J Microbiol Methods 133, 46–51 (2017).
13. Rouches, M. V., Xu, Y., Cortes, L. B. G. & Lambert, G. A plasmid system with tunable copy number. Nat Commun 13, 3908 (2022).
14. Kurnit, D. M. Escherichia coli recA deletion strains that are highly competent for transformation and for in vivo phage packaging. Gene 82, 313–315 (1989).
15. Zürcher, J. F. et al. Continuous synthesis of E. coli genome sections and Mb-scale human DNA assembly. Nature 619, 555–562 (2023).
16. Guesdon, G. et al. Combining Fusion of Cells with CRISPR-Cas9 Editing for the Cloning of Large DNA Fragments or Complete Bacterial Genomes in Yeast. ACS Synth. Biol. 12, 3252–3266 (2023).
17. Koster, C. C., Postma, E. D., Knibbe, E., Cleij, C. & Daran-Lapujade, P. Synthetic Genomics From a Yeast Perspective. Frontiers in Bioengineering and Biotechnology 10, (2022).
18. Chizzolini, F., Forlin, M., Cecchi, D. & Mansy, S. S. Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength. ACS Synth. Biol. 3, 363–371 (2014).
19. Niederholtmeyer, H., Stepanova, V. & Maerkl, S. J. Implementation of cell-free biological networks at steady state. Proc. Natl. Acad. Sci. U.S.A. 110, 15985–15990 (2013).
20. Hossain, A. et al. Automated design of thousands of nonrepetitive parts for engineering stable genetic systems. Nat Biotechnol 38, 1466–1475 (2020).
21. Noireaux, V., Bar-Ziv, R. & Libchaber, A. Principles of cell-free genetic circuit assembly. Proceedings of the National Academy of Sciences 100, 12672–12677 (2003).
22. Tickman, B. I. et al. Multi-layer CRISPRa/i circuits for dynamic genetic programs in cell-free and bacterial systems. Cell Syst 13, 215-229.e8 (2022).
23. Voyvodic, P. L. et al. Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. Nat Commun 10, 1697 (2019).
24. Greco, F. V., Pandi, A., Erb, T. J., Grierson, C. S. & Gorochowski, T. E. Harnessing the central dogma for stringent multi-level control of gene expression. Nat Commun 12, 1738 (2021).

25. Shin, J. & Noireaux, V. An E. coli Cell-Free Expression Toolbox: Application to Synthetic Gene Circuits and Artificial Cells. ACS Synth. Biol. 1, 29–41 (2012).
26. Kim, J., White, K. S. & Winfree, E. Construction of an in vitro bistable circuit from synthetic transcriptional switches. Mol Syst Biol 2, 68 (2006).
27. Baumschlager, A., Aoki, S. K. & Khammash, M. Dynamic Blue Light-Inducible T7 RNA Polymerases (Opto-T7RNAPs) for Precise Spatiotemporal Gene Expression Control. ACS Synth. Biol. 6, 2157–2167 (2017).
28. de Vega, M., Lázaro, J. M. & Salas, M. Improvement of Phi29 DNA Polymerase Amplification Performance by Fusion of DNA Binding Motifs. in Rolling Circle Amplification (RCA) (ed. Demidov, V. V.) 11–24 (Springer International Publishing, Cham, 2016). doi:10.1007/978-3-319-42226-8_2.
29. Matsumoto, K. Phosphatidylserine synthase from bacteria. Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism 1348, 214–227 (1997).
30. Okada, M., Matsuzaki, H., Shibuya, I. & Matsumoto, K. Cloning, sequencing, and expression in Escherichia coli of the Bacillus subtilis gene for phosphatidylserine synthase. J Bacteriol 176, 7456–7461 (1994).
31. What does the Twist Codon Optimization tool do? | Twist Bioscience. https://www.twistbioscience.com/faq/using-your-twist-account/what-does-twist-codon-optimization-tool-do.
32. Jackowski, S., Jackson, P. D. & Rock, C. O. Sequence and function of the aas gene in Escherichia coli. J. Biol. Chem. 269, 2921–2928 (1994).
33. Bartelds, M. W. et al. Noise Minimization in Cell-Free Gene Expression. ACS Synth. Biol. 12, 2217–2225 (2023).
34. Sheahan, T. & Wieden, H.-J. Ribosomal Protein S1 Improves the Protein Yield of an In Vitro Reconstituted Cell-Free Translation System. ACS Synth. Biol. 11, 1004–1008 (2022).
35. Pandi, A. et al. A versatile active learning workflow for optimization of genetic and metabolic networks. Nat Commun 13, 3876 (2022).
36. Abil, Z. & Danelon, C. Roadmap to Building a Cell: An Evolutionary Approach. Frontiers in Bioengineering and Biotechnology 8, (2020).
37. Chusacultanachai, S. & Yuthavong, Y. Random mutagenesis strategies for construction of large and diverse clone libraries of mutated DNA fragments. Methods Mol Biol 270, 319–334 (2004).
38. Volkov, A. A. & Arnold, F. H. [26] Methods for in vitro DNA recombination and random chimeragenesis. in Methods in Enzymology (eds. Thorner, J., Emr, S. D. & Abelson, J. N.) vol. 328 447–456 (Academic Press, 2000).
39. Labrou, N. E. Random mutagenesis methods for in vitro directed enzyme evolution. Curr Protein Pept Sci 11, 91–100 (2010).
40. Sellés Vidal, L., Isalan, M., Heap, J. T. & Ledesma-Amaro, R. A primer to directed evolution: current methodologies and future directions. RSC Chem Biol 4, 271–291.
41. Coco, W. M. et al. DNA shuffling method for generating highly recombined genes and evolved enzymes. Nat Biotechnol 19, 354–359 (2001).
42. Saito, Y. et al. Machine-Learning-Guided Library Design Cycle for Directed Evolution of Enzymes: The Effects of Training Data Composition on Sequence Space Exploration. ACS Catal. 11, 14615–14624 (2021).
43. Yang, J. et al. DeCOIL: Optimization of Degenerate Codon Libraries for Machine Learning-Assisted Protein Engineering. ACS Synth. Biol. 12, 2444–2454 (2023).
44. Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. Nat Methods 16, 687–694 (2019).
45. Chica, R. A., Doucet, N. & Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Current Opinion in Biotechnology 16, 378–384 (2005).
46. Chu, H. Y. & Wong, A. S. L. Facilitating Machine Learning-Guided Protein Engineering with Smart Library Design and Massively Parallel Assays. Advanced Genetics 2, 2100038 (2021).
47. Sumbalova, L., Stourac, J., Martinek, T., Bednar, D. & Damborsky, J. HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information. Nucleic Acids Res 46, W356–W362 (2018).

48. Jeschek, M., Gerngross, D. & Panke, S. Rationally reduced libraries for combinatorial pathway optimization minimizing experimental effort. Nat Commun 7, 11163 (2016).
49. Espah Borujeni, A. & Salis, H. M. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. J. Am. Chem. Soc. 138, 7016–7023 (2016).
50. Calvopina-Chavez, D. G., Gardner, M. A. & Griffitts, J. S. Engineering efficient termination of bacteriophage T7 RNA polymerase transcription. G3 Genes|Genomes|Genetics 12, jkac070 (2022).
51. Kralicek, A. A Cell-Free Expression Screen to Identify Fusion Tags for Improved Protein Expression. in Cell-Free Protein Synthesis: Methods and Protocols (eds. Alexandrov, K. & Johnston, W. A.) 35–54 (Humana Press, Totowa, NJ, 2014). doi:10.1007/978-1-62703-782-2_3.
52. Mardikoraem, M. & Woldring, D. Machine Learning-driven Protein Library Design: A Path Toward Smarter Libraries. Methods Mol Biol 2491, 87–104 (2022).
53. Kaufmann, K. W., Lemmon, G. H., DeLuca, S. L., Sheehan, J. H. & Meiler, J. Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You. Biochemistry 49, 2987–2998 (2010).
54. Damborsky, J. & Brezovsky, J. Computational tools for designing and engineering enzymes. Curr Opin Chem Biol 19, 8–16 (2014).
55. Verma, R., Schwaneberg, U. & Roccatano, D. Computer-Aided Protein Directed Evolution: a Review of Web Servers, Databases and other Computational Tools for Protein Engineering. Comput Struct Biotechnol J 2, e201209008 (2012).
56. Patsch, D., Eichenberger, M., Voss, M., Bornscheuer, U. T. & Buller, R. M. LibGENiE – A bioinformatic pipeline for the design of information-enriched enzyme libraries. Computational and Structural Biotechnology Journal 21, 4488–4496 (2023).
57. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).
58. Varadi, M. et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research 50, D439–D444 (2022).
59. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27, 946–950 (2009).
60. LaFleur, T. L., Hossain, A. & Salis, H. M. Automated model-predictive design of synthetic promoters to control transcriptional profiles in bacteria. Nat Commun 13, 5159 (2022).
61. Doerr, A., Foschepoth, D., Forster, A. C. & Danelon, C. In vitro synthesis of 32 translation-factor proteins from a single template reveals impaired ribosomal processivity. Sci Rep 11, 1898 (2021).
62. Vezeau, G. E. & Salis, H. M. Tuning Cell-Free Composition Controls the Time Delay, Dynamics, and Productivity of TX-TL Expression. ACS Synth. Biol. 10, 2508–2519 (2021).
63. Salis, H. M. The ribosome binding site calculator. Methods Enzymol 498, 19–42 (2011).
64. Steffens, D. L. & Williams, J. G. K. Efficient Site-Directed Saturation Mutagenesis Using Degenerate Oligonucleotides. J Biomol Tech 18, 147–149 (2007).
65. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnology 8, 91 (2008).
66. Zhang, K. et al. A high-efficiency method for site-directed mutagenesis of large plasmids based on large DNA fragment amplification and recombinational ligation. Sci Rep 11, 10454 (2021).
67. Hallak, L. K. et al. Efficient method for site-directed mutagenesis in large plasmids without subcloning. PLoS One 12, e0177788 (2017).
68. Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460, 894–898 (2009).
69. Wannier, T. M. et al. Recombineering and MAGE. Nat Rev Methods Primers 1, 1–24 (2021).
70. Wang, H. H. & Church, G. M. Multiplexed Genome Engineering and Genotyping Methods. in Methods in Enzymology vol. 498 409–426 (Elsevier, 2011).
71. Ronda, C., Pedersen, L. E., Sommer, M. O. A. & Nielsen, A. T. CRMAGE: CRISPR Optimized MAGE Recombineering. Sci Rep 6, 19452 (2016).
72. Nyerges, Á. et al. A highly precise and portable genome engineering method allows comparison of mutational effects across bacterial species. Proceedings of the National Academy of Sciences 113, 2502–2507 (2016).
73. Seo, H. & Lee, H. Recent developments in microfluidic synthesis of artificial cell-like polymersomes and liposomes for functional bioreactors. Biomicrofluidics 15, 021301 (2021).
74. Martino, C. & deMello, A. J. Droplet-based microfluidics for artificial cell generation: a brief review. Interface Focus 6, 20160011 (2016).
75. Godino, E., Restrepo Sierra, A. M. & Danelon, C. Imaging Flow Cytometry for High-Throughput Phenotyping of Synthetic Cells. ACS Synth. Biol. 12, 2015–2028 (2023).
76. Nitta, N. et al. Intelligent Image-Activated Cell Sorting. Cell 175, 266-276.e13 (2018).
77. Isozaki, A. et al. Intelligent image-activated cell sorting 2.0. Lab Chip 20, 2263–2273 (2020).
78. Boothe, T. et al. A tunable refractive index matching medium for live imaging cells, tissues and model organisms. eLife 6, e27240.
79. Vogele, K. et al. Towards synthetic cells using peptide-based reaction compartments. Nat Commun 9, 3862 (2018).

Summary

Life! what a great scientific drive to get inspired by it and try to craft a simple (for now) form of it, a synthetic cell. Motivated by Nature's evolutionary process, which has shaped all life forms over billions of years, our goal in Christophe Danelon's lab is to construct an autonomous synthetic cell from the bottom-up using laboratory evolution as an engineering tool. This dissertation contributes to the objective by showing first steps in the integration of biological modules, and developing new methods for in vitro evolution, screening and selection of synthetic cell models.

In the introductory chapter **(chapter 1)**, we present to the reader the long-standing scientific curiosity for life's principles and the notion of a synthetic cell. We illustrate how the intricacy of living processes can be approached from different perspectives, and how the synthetic biology field has embraced the journey to create a minimal artificial cell. Some of the great advances that the synthetic cell community has had so far are showcased with examples taken from the bottom-up and top-down strategies. With this overview, we explain our DNA-based approach for crafting synthetic cells with integrated functionalities, and emphasize the importance to now look at module integration and evolution as powerful next steps.

In the first experimental chapter of this dissertation **(chapter 2)**, we tackle the challenge of the low phenotypic output imposed by the low input DNA concentrations required in current in vitro evolution assays. To do so, we establish CADGE, an isothermal-based DNA amplification strategy for the orthogonal replication of linear protein-coding dsDNA templates during simultaneous transcription and translation. We show CADGE effectiveness on (i) enhancing protein production and phenotypic output of a few tested soluble and membrane-associated proteins, (ii) enabling the recovery of liposome-encapsulated DNA from low (clonal) DNA concentrations, and (iii) facilitating the enrichment of a DNA variant from a mock gene library. Lastly, we illustrate the significant potential of CADGE in the realms of in vitro protein engineering and synthetic cell research.

Inspired by the robust outcomes that evolution can offer, i.e., life itself, but also the many successful examples of protein engineering using laboratory evolution, we introduce in **chapter 3** a synthetic cell evolutionary framework based on a minimal dsDNA self-replicator. We present two different in-liposome evolutionary set-ups: intermittent and a (semi)continuous, and demonstrate that both settings (i) enable persistent DNA self-replication from low starting DNA concentrations (clonal conditions), (ii) introduce in situ DNA diversity to kickstart evolution, and (iii) lead to enriched variants with improved replication ability. Within only a few evolutionary rounds, we show the emergence and

persistence of fitter full-length self-replicators. We identify and characterize some of the fixed mutations, offering valuable insights to consider for designing and evolving future synthetic genomes.

Recognizing module integration as a pivotal next step for constructing a synthetic cell, **chapter 4** presents results on the combination of DNA replication, membrane biosynthesis, and transcription-translation, all within single liposome compartments. We demonstrate minimal cross-talk effects from each other's substrates or intermediary compounds. However, we find that co-expression of the two genetic modules, DNA replication and phospholipid biosynthesis, negatively influences DNA amplification yields and the overall occurrence of liposomes displaying both phenotypes. Finally, we discuss potential optimization paths, such as in vitro evolution, to expedite and enhance future module integration attempts.

In our final experimental chapter **(chapter 5)**, we transition to the high-throughput characterization of synthetic cells, which we believe is a powerful technology for engineering synthetic cells. We illustrate the applicability of Imaging Flow Cytometry (IFC) as a great microscopy-based tool for screening large populations of gene expressing liposomes. We present user-friendly pipelines with a commercial IFC instrument and software, which enables the quantitative analysis of a wide range of synthetic cell phenotypic traits. Within only a few hours, more than 60 thousand liposome images can be collected and computationally analyzed to obtain statistically significant parameters about their morphological features. We close this chapter by discussing the great applicability of IFC in synthetic cell research, their current limitations, and ideas for next-generation instruments.

We finalize this dissertation by envisioning the upcoming endeavors of synthetic cell research in **chapter 6**. In particular, we discuss a semi-rational evolutionary approach and reflect on how to build DNA templates with desired genetic diversity, and how to screen, sort, and recover the most suitable variants for characterization and reverse engineering. Within the overview of options to explore, we present some preliminary experimental results on (i) alternative proteins and/or pathways for phospholipid synthesis, (ii) the construction of DNA constructs and smart libraries, and (iii) an intelligent image-based screening and selection strategy for future synthetic cell evolutionary campaigns. Additionally, and also accompanied with some preliminary data, we discuss our immediate next steps on the semi-rational evolution of the integrated DNA replication-lipid synthesis modules, already presented in chapter 4. We end chapter 6 and now this thesis by reinforcing the idea that we, as synthetic cell crafters, will probably not create an exact replica of an existing (simplest) cell, but should still use nature's cues and longstanding evolutionary trajectory as a great inspiration to build synthetic life.

Samenvatting

Het leven! Wat een geweldige wetenschappelijke drijfveer om door geïnspireerd te raken – geïnspireerd om er een (voorlopig) eenvoudige vorm van te creëren: een synthetische cel. Het natuurlijke proces van evolutie, dat alle levensvormen op aarde gedurende miljarden jaren heeft gevormd, heeft het lab van Christophe Danelon geïnspireerd om bottom-up autonome synthetische cellen te construeren met behulp van in vitro evolutie. Deze dissertatie draagt bij aan dit doel door de eerste stappen te zetten in de integratie van biologische modules, en door het ontwikkelen van nieuwe methoden voor in vitro evolutie, screening en selectie van synthetische celmodellen.

In het inleidende hoofdstuk bespreken we de langdurige wetenschappelijke nieuwsgierigheid naar de fundamentele eigenschappen van het leven en het concept van een synthetische cel **(hoofdstuk 1)**. We illustreren hoe de complexiteit van levende processen vanuit verschillende perspectieven kan worden benaderd, en hoe het veld van de synthetische biologie is begonnen aan de bouw van een minimale kunstmatige cel. Aan de hand van voorbeelden uit de bottom-up en top-down strategieën beschrijven wij enkele grote vooruitgangen die de synthetische cel gemeenschap tot nu toe heeft geboekt. Tenslotte leggen wij onze benadering voor het creëren van synthetische cellen uit: gebaseerd op DNA en met geïntegreerde functies. Ook benadrukken we het belang om nu naar module-integratie en evolutie te kijken als volgende stappen.

In het eerste experimentele hoofdstuk van deze dissertatie **(hoofdstuk 2)** bespreken we het probleem dat een lage input DNA concentratie, die essentieel is voor in vitro evolutie experimenten, resulteert in een lage fenotypische output. Om dit probleem op te lossen, presenteren wij CADGE: een isothermische methode voor de orthogonale replicatie van lineair DNA waarbij tegelijkertijd transcriptie en translatie kunnen plaatsvinden. Wij laten zien dat CADGE (i) de eiwitproductie en fenotypische output verbetert van zowel enkele oplosbare eiwitten als enkele membraaneiwitten, (ii) het mogelijk maakt om DNA, ingekapseld in liposomen bij een zeer lage concentratie, te amplificeren voor verdere karakterisering, en (iii) de verrijking van een DNA-variant uit een testlibrary vergemakkelijkt. Ten slotte illustreren we het potentieel van CADGE op het gebied van in vitro eiwitengineering en synthetische celonderzoek.

Geïnspireerd door de potentie van evolutie, aangetoond door het leven zelf maar ook door de vele succesvolle voorbeelden van evolutionaire eiwitengineering, introduceren we in **hoofdstuk 3** een framework voor de evolutie van een minimale zelfreplicerend dubbelstrengs DNA molecuul. We presenteren twee verschillende manieren voor evolutie in liposomen: onderbroken en (semi)continue evolutie. We tonen aan dat beide manieren

(i) aanhoudende DNA-zelfreplicatie mogelijk maken met lage start-DNA-concentraties, (ii) in situ DNA-diversiteit introduceren om evolutie op gang te brengen, en (iii) leiden tot verrijkte varianten die beter kunnen repliceren. Binnen slechts enkele evolutionaire rondes laten we de opkomst en stabiliteit zien van fittere zelfreplicatoren waarbij de lengte van het DNA wordt behouden. We identificeren en karakteriseren enkele van de gefixeerde mutaties, die waardevolle inzichten bieden voor het ontwerpen en evolueren van toekomstige synthetische genomen.

Omdat module-integratie een cruciale volgende stap is in de constructie van een synthetische cel, combineren we in **hoofdstuk 4** drie modules in één liposoom: DNA-replicatie, synthese van membraanfosfolipiden en transcriptie-translatie. We tonen aan dat de crosstalk tussen de substraten en tussenproducten van de verschillende modules minimaal is. Toch zien we dat de co-expressie van de twee modules DNA-replicatie en fosfolipidesynthese een negatieve invloed heeft op de mate van DNA-replicatie en op de hoeveelheid liposomen die beide fenotypen laten zien. Tot slot bespreken we opties voor optimalisatie, zoals in vitro evolutie, om toekomstige pogingen tot module-integratie te versnellen en te verbeteren.

In ons laatste experimentele hoofdstuk **(hoofdstuk *5*)** maken we de overstap naar de high-throughput karakterisering van synthetische cellen. We illustreren de toepasbaarheid van Imaging Flow Cytometry (IFC), een handig instrument dat gebruik maakt van microscopie voor het screenen van grote populaties liposomen waarin genexpressie plaatsvindt. We presenteren gebruiksvriendelijke workflows met een commercieel IFC-instrument en commerciële software, waarmee de kwantitatieve analyse van een breed scala aan fenotypische eigenschappen van synthetische cellen mogelijk is. Binnen slechts enkele uren kunnen meer dan 60 duizend liposoomafbeeldingen worden verzameld en geanalyseerd. Zo kunnen statistisch significante parameters over de morfologische kenmerken van de liposomen verkregen worden. We sluiten dit hoofdstuk af met het bespreken van de grote toepasbaarheid van IFC in synthetisch celonderzoek, de huidige beperkingen en ideeën voor toekomstige instrumenten.

Tot slot bespreken we de toekomstige onderzoeksrichtingen voor synthetische celonderzoek in **hoofdstuk 6**. In het bijzonder bespreken we een semi-rationele evolutionaire aanpak, en behandelen we hoe DNA met gewenste genetische diversiteit kan worden gesynthetiseerd. Ook bespreken we methoden voor het screenen en sorteren van de fitste varianten, gevolgd door het amplificeren van de gesorteerde DNA-templates voor verdere karakterisering en reverse engineering. Voor enkele van de gesuggereerde methoden presenteren we voorlopige experimentele resultaten. Dit zal gaan over: (i) alternatieve eiwitten en/of routes voor fosfolipidesynthese, (ii) de synthese van DNA-constructen en smart libraries, en (iii) een intelligente, op afbeelding gebaseerde screening- en selectiestrategie voor toekomstige evolutionaire experimenten met synthetische cellen. Als laatste bespreken we onze directe volgende stappen in de semi-rationele evolutie van de geïntegreerde modules voor DNA-

replicatie en lipidesynthese, gesterkt met enkele voorlopige resultaten die voortbouwen op de resultaten in hoofdstuk 4. We sluiten hoofdstuk 6 en deze scriptie af door te benadrukken dat wij, als makers van synthetische cellen, hoogstwaarschijnlijk geen exacte replica van een bestaande (eenvoudigste) cel zullen creëren. Wel kunnen we de natuur en haar langdurige evolutionaire traject gebruiken als inspiratie om synthetisch leven te bouwen.

Resumen

¡La vida!, que gran inspiración científica para tratar de entender cómo funciona y por qué no, tratar de crear una forma simple (por ahora) de ella, ¡Una célula sintética! En el grupo de investigación del profesor Christophe Danelon, ese es nuestro objetivo. Crear una célula sintética que pueda llamarse 'viva' comenzando por macromoléculas inertes pero funcionales, y utilizando evolución como una excelente herramienta. En particular, esta tesis muestra la importancia de la integración de maquinarias esenciales para la vida en la construcción de una célula sintética, e ilustra posibles métodos evolutivos para optimizar esta integración.

En el capítulo introductorio **(capítulo 1)**, ilustramos cómo la curiosidad científica por los principios de la vida nos ha llevado ha abordar la investigación de complejos procesos vitales desde diferentes perspectivas y cómo el campo de la biología sintética ha embarcado el viaje de crear una célula artificial. Con un panorama general y la noción de lo que es una célula sintética, explicamos nuestro particular enfoque basado en ADN para crear células artificiales con diferentes funcionalidades y de nuevo enfatizamos la importancia de ahora enfocarse en la integración de módulos biológicos.

En el primer capítulo experimental de esta tesis **(capítulo 2)**, abordamos el problema de baja producción fenotípica en campañas evolutivas in vitro dentro de liposomas, impuesto por bajas concentraciones iniciales de ADN, necesarias para mantener un enlace entre el genotipo y el fenotipo. Para sobrepasar este problema establecemos CADGE, una estrategia de amplificación isotérmica de ADN en donde un ADN linear (y de doble cadena) que codifica una proteína de interés puede ser replicado de manera clonal durante transcripción y traducción simultáneas. Nuestros resultados muestran la efectividad de nuestro método CADGE para (i) mejorar la producción de proteína y el fenotipo asociado para algunas proteínas solubles o vinculadas a la membrana, (ii) permitir la recuperación de ADN encapsulado en liposomas a partir de concentraciones bajas (clonales) de ADN, y (iii) facilitar el enriquecimiento de una variante de ADN sobre una biblioteca de genes simulada. Por último, ilustramos el potencial de CADGE en los ámbitos de la ingeniería de proteínas in vitro y la investigación de células sintéticas.

Inspirados por los maravillosos resultados que la evolución puede ofrecer, la vida misma (claro está), pero también los muchos ejemplos exitosos de ingeniería de proteínas en el laboratorio, presentamos en el **capítulo 3** un marco evolutivo para células sintéticas basado en un replicador de ADN de doble cadena. Presentamos dos configuraciones evolutivas dentro de los liposomas: intermitente y (semi)continua, y demostramos que ambas configuraciones (i) permiten una autorreplicación persistente de DNA bajo

condiciones clonales (una molécula de ADN por liposoma), (ii) permiten la introducción de diversidad genética en el ADN del replicador para iniciar la evolución y (iii) conducen al enriquecimiento de variantes de ADN con capacidad de replicación mejorada, en solo unas cuantas rondas evolutivas. Finalmente, identificamos y caracterizamos algunas de las mutaciones encontradas luego de la evolución, proporcionando información interesante a tener en cuenta para el diseño y evolución de futuros genomas sintéticos.

Reconociendo la integración de módulos biológicos como un paso crucial para construir una célula sintética, el **capítulo 4** presenta resultados sobre la combinación de tres procesos esenciales: replicación de ADN, biosíntesis de membranas y transcripción-traducción, todo dentro de compartimentos individuales de liposomas. Demostramos efectos mínimos de interferencia entre maquinaria y/o sustratos, pero encontramos que la producción simultanea de los dos módulos genéticos (replicación de ADN y síntesis de fosfolípidos), influye negativamente en los rendimientos de amplificación de ADN y en la ocurrencia general de liposomas que muestran ambos fenotipos. Finalmente, discutimos posibles caminos de optimización, como la evolución in vitro, para acelerar y mejorar los intentos futuros de integración de módulos biológicos.

En nuestro último capítulo experimental **(capítulo *5*)**, pasamos a la caracterización de células sintéticas con tecnologías de alto rendimiento. Específicamente, ilustramos la aplicabilidad de la Citometría de Flujo de Imágenes (IFC) como una gran herramienta de caracterización para evaluar con microscopía grandes poblaciones de liposomas con fenotipos derivados de expresión genética interna. Ilustramos que el procesamiento de muestras y de datos se puede realizar con un instrumento comercial y de forma amigable para el usuario. En particular, utilizamos el software proveído con el instrumento para analizar datos de diferentes maneras y lograr realizar análisis cuantitativos en una amplia gama de rasgos fenotípicos de células sintéticas. En solo unas pocas horas, se pueden recopilar y analizar computacionalmente más de 60 mil imágenes de liposomas para obtener parámetros estadísticamente significativos sobre sus características morfológicas. Cerramos este capítulo discutiendo la gran aplicabilidad de IFC en la investigación de células sintéticas, sus limitaciones actuales e ideas para futuros instrumentos.

Finalizamos esta tesis imaginando futuros objetivos de investigación en el campo de células sintéticas **(capítulo 6)**. En particular, discutimos un enfoque evolutivo semi racional y reflexionamos sobre cómo construir futuros genomas sintéticos con diversidad genética deseada. Nos enfocamos también en discutir estrategias para evaluar, clasificar y recuperar los genomas más adecuados para su caracterización. Dentro del panorama de opciones a explorar, presentamos algunos resultados preliminares sobre (i) proteínas y/o rutas alternativas para la síntesis de fosfolípidos, (ii) la construcción de genomas y bibliotecas ‘inteligentes’, y (iii) una estrategia ‘inteligente’ para la evaluación y selección de células sintéticas basada en citometría de flujo de imágenes. Adicionalmente, y también acompañados de algunos datos preliminares, discutimos los pasos a seguir para la evolución

semi racional de los módulos integrados en el capítulo 4 de esta tesis: replicación de ADN y síntesis de fosfolípidos. Concluimos el capítulo 6 y ahora esta tesis reforzando la idea de que nosotros, como creadores de células sintéticas, probablemente no crearemos una réplica exacta de una célula existente (así sea la más simple), pero igual podemos inspirarnos en la naturaleza y su larga trayectoria evolutiva para algún día lograr construir vida sintética.

Acknowledgements

I would like to start this section by thanking my PhD supervisor, **Christophe Danelon**. I still remember the first time that I encountered your work when I was doing my Master degree back in Saudi. Who would have thought that this initial encounter would drive me to work on such great science and with such great people for almost five years. Christophe, you are an admiring person, supervisor, and scientist. Thank you for welcoming me in your awesome research group, for your great and patient guidance, inspiring ideas, and for your always motivating attitude towards good and bad results or situations. Thank you for fostering an environment where mistakes are embraced as part of the learning process and exploring new scientific ideas and other passions (e.g. drawing and exploring other cultures) are welcomed and encouraged. I am grateful that you always care to choose the right people to keep a great working environment, which I always dared to call a second family for me. Definitely, my PhD experience wouldn't have been so enriching personally and scientifically if I did not have you as a supervisor, and I will always be thankful for that. I really hope that the academic community will have more advisors like you, driven by trueful scientific passion and with a deep commitment to nurturing their students.

I would like to also express my gratitude to **Gijsje Koenderink**, my now second promotor here in Delft. Thank you for adopting me in your group with such open arms, for giving me insightful comments in your feedback for this thesis, and for always having encouraging words to keep going. Your great attitude and scientific curiosity is contagious and very much appreciated!

Before moving forward with all the great people during my PhD, I would like to go back to my Master and Bachelor, and thank all the great people that inspired me back then to be driven by curiosity, and explore a scientific career. I thank my advisors in Cornell, **Adam Bogdanove** and **Morgan Carter**. Your great guidance and scientific work confirmed for me how awesome science could be, and inspired me to continue with a scientific life journey. I continue with my great advisors in KAUST. Even though I thanked you already in my Master thesis, I think it deserves to be done again. **Stefan**, thank you for hosting me in your group and for giving me the opportunity to immerse in an incredible unexplored project that taught me so much on fully embracing research, on having patience, and on troubleshooting. It helped me a lot during my PhD. **Raik**, yes, I will thank you again of course. You are an open encyclopedia, full of knowledge. Thank you for sharing it with your students, including me. I admire your passion for science, creativity, and willingness to explore crazy scientific ideas. Thank you for all the extensive and fruitful discussions about science and almost any topic, and thank you for your patience and willingness to continue

working to publish our work even some years after finishing it.

Now, to my colleagues and friends from the PhD, it wouldn't have been such a fruitful journey if it wasn't for you all. **Zhanar,** thank you for receiving me in Christophe's lab with open arms to work with you. I enjoyed a lot our scientific discussions back then in the office and lab. I am grateful for your always useful advice on how to shape projects, present, and write science. I admire your perseverance and passion for science and academia, and I am very happy to have worked with you throughout these PhD years. **Elisa,** you are an inspiring scientist and friend. I greatly enjoyed our time working together (in iGEM and in our IFC project), but also our time in the office or lab embarking in scientific discussions about each other's projects. Apart from science, I'm grateful for your regular reminders to take care of myself and focus on my well-being. Your words of encouragement were always there to uplift me whenever I had tough days, and I will always be thankful for that. **Duco, David** and **Ilja,** thank you for welcoming me in the PhD with such great attitude and lots of fun. **Duco,** thanks for teaching me how to make liposomes on my very first day in the lab, for all the LC-MS and MATLAB teaching sessions, and for some great Dutch lunch lessons (that I now forgot but are still appreciated). **David,** thank you for your always lively attitude in the lab and the office with nice Coldplay and ABBA songs in background. 'Muchos cool' results came out of such positive environment, and I thank you for that. **Ilja,** thank you for keeping the lab running so nicely, for taking care of us, and for always being open to help me whenever I needed it until the almost very end of my PhD.

With Elisa and Ilja still included, I now include Fede, Marijn, Celine and Laura. Words cannot convey how grateful I am that we were together to go through the challenges we faced. Looking back, I see us thrive as a family with immense care and support for each other, and I will always be grateful for that. **Andreea**, I am grateful for the good moments we shared together. I will always remember them. **Fede,** it was first great to have you as an iGEM student but it was even better to have you as a master student and now as a colleague and friend. I really enjoyed having endless scientific discussions in 'Spanglish' in the lab, in the office, on WhatsApp, on Zoom. You name the media. Thank you for your great committement with our project, for always being an assertive student, with a curious mind and inspiring passion for science. Besides science, I am greateful for your openness to always share a word when needed, and for caring about people as much as I do, even if sometimes could be frustrating. I am happy that you are still with us on your PhD, and I can already see the great science that lies ahead in your journey. **Marijn,** you are such a great person, friend and scientist. I am so happy that you chose to return for doing your PhD here with us, and to see you rocking' it in your research. Thank you for the countless support when I struggle in different things, and for always being there to discuss science or share a word on anything! I am thrilled to leave pMAR3 in such great hands, and I will be waiting for you to create our 'Sciency' store/café with great almond croissants, coffee, artsy workshops, and much more! It will be a great success :). **Celine,** you are one of the happiest great scientists and friends I have! thank you for having such a contagious cheerful and curious attitude

towards science and anything in life. Thank you for always being there to listen, discuss, have answers, and give nice feedback with such a great and helpful disposition. Enjoy the rest of your PhD ride, you are doing and will do great science! **Laura,** I am so grateful to have met you in iGEM! I was so happy to hear that you were going to join us as a Master student with Celine, but even happier when I heard that you were coming back to work with me for a while, and now in Toulouse as a PhD, and as a long-distance friend. Thank you for your constant clever attitude towards science and so many things in life, and for your willingness to always help me and others on our projects. You are an inspiring person and scientist, I can already see the great science coming ahead from your hands. Now, I would like to finish this sub-section by also thanking you all for always being open to share great experiences outside the lab and office, for being great friends and 'compinches' for doings sports, traveling together, going to dip ourselves in cold north-sea waters, tasting great food in our potluck dinners or in nice restaurants, and much more! we did feed well our hearts and bellies during these years, and I thank you all for that.

I would like to also acknowledge all the great students that I had the chance to supervise during my PhD: having mentioned Fede already, I now add **Judith, Fleur, Brent, Anna, and all the iGEM teams: Phocus, AptaVita, and SPYKE.** I really enjoyed the time working with you on your projects, as well as watching you shine on the iGEM Jamboree and/or graduation talks and evaluations. Regarding iGEM, I also want to acknowledge **Essie, Elisa, Martin, Justas, Charlotte, and Britte** for the great time we spent together co-supervising iGEM. **Essie,** you are a great person with a beautiful energy and always a smile to share! I am grateful for our time exploring Paris, discussing about iGEM at BN, or thinking about a whole new (and maybe better?) Synthetic Biology community (iConfuse) with Justas.

Besides the people I've mentioned above, I am grateful to have met many more fantastic people in BN: **Amelie, Rosalie, Marlena, Kelly, Leila, Daniel, Luci, Gerard, SaFyre, Jeffrey, Justas, Wayne, Martin, Sandro, Alberto, Reza, Beatriz, Ramon, Marieke, Hidde, Christos, Adja,...** and so many more. Thank you all for always being so friendly and open to share a word, even though I was (sometimes) too shy to say hi or stay talking for long.

To all my other friends here in the Netherlands and abroad: **Lau, Leo, Juanda, Nadine, Lore, Turru, Mile, Aleja, Juanri, Saris, Escarlet...**, and many more. Thank you for always being present to talk in both happy and stressful spikes throughout these years. I am really thankful for the constant support you've given me as the wonderful friends you are.

To my **Suki!,** I am delighted and grateful that we finally crossed paths here. Without your love and support, I think I wouldn't have arrived here in 'all pieces'. Thank you for being by my side and for taking such good care of me and our home during this last phase of the PhD. Thank you for your patience and always encouraging words, for having a contagious serenity whenever I was thinking the worst could happen, and for helping bring my motivation back up whenever intrusive thoughts brought it down. You were always there to help declutter my head from different things, to keep a spotless home, and to bring calm

accompanied by nice cappuccinos. I am deeply grateful to have such an amazing person as a life partner.

Finalmente, quiero dar gracias a mi familia. ¡Qué habría sido de mi y de este doctorado sin ustedes! Gracias por ser un ejemplo a seguir desde pequeña, por siempre cuidar de mí, regañarme, inculcarme la importancia del estudio, de los idiomas y de explorar el mundo. Gracias por su incondicional apoyo en mis decisiones de vida y por siempre darme los mejores consejos mientras yo exploraba y decidía los caminos a seguir. Gracias Mamá por darme siempre el mejor ejemplo de vida, gracias por tu invaluable perseverancia y fortaleza ante múltiples situaciones. Gracias por tu infaltable apoyo, palabras de motivación, y por siempre rellenarme de energía desde la distancia y en Colombia. Gracias Papá por tu apoyo, paciencia, y por siempre tener la disposición para ayudarme con todo lo que necesitara. A mi hermano Jota, gracias por ser un modelo a seguir desde pequeña, por tu admirable energía, conocimiento, y ejemplo de perseverancia por lo que te propones en la vida. Maria, Susi y Manuel, ¡qué felicidad tenerlos en mi vida! Gracias por todos los momentos compartidos, por las risas, tintos, montadas en bici, sesiones de pintura, partidos de fútbol, y palabras de apoyo que siempre me recargaron estos años para volver a continuar en el doctorado. Agradezco a la vida por darme esta maravillosa familia que me ha acompañado y respaldado en cada etapa de esta aventura que hoy culmina.

Ana María Restrepo Sierra

March, 2024. Delft, The Netherlands.

Curriculum Vitae

Ana María Restrepo Sierra

01-06-1993	Born in Medellín, Colombia.
2004-2010	Secondary Education Bethlemitas School, Medellín, CO.
2011-2016	B. Sc. in Biological Engineering National University of Colombia, Medellín, CO.
2016	Research Internship School of Integrative Plant Science, Cornell University, Ithaca, USA.
2017-2019	M. Sc. in Bioscience King Abdullah University of Science and Technology, Thuwal, SA.
2022	Research Internship Department of Chemistry, University of Tokyo, Tokyo, JP.
2019-2023	Ph. D. in Bionanoscience Delft University of Technology, Delft, NL. *Supervisor:* Prof. dr. C.J.A. Danelon *Promotors:* Prof. dr. C.J.A. Danelon and Prof. dr. G.H. Koenderink

List of publications

Restrepo Sierra, A. M., Ramírez Gómez, F., Van Tongeren, M. Synthetic Cells with Integrated DNA Self-replication and Phospholipid Biosynthesis. *in preparation.*

Abil Z.*, **Restrepo Sierra, A. M.***, *et al*. Adaptive Evolution of Self-Replicating DNA in a Synthetic Protocell. *in preparation.*

Godino, E.*, **Restrepo Sierra, A. M.***, & Danelon, C. (2023). Imaging Flow Cytometry for High-Throughput Phenotyping of Synthetic Cells. ACS Synthetic Biology, 12(7), 2015–2028.

Abil, Z.*, **Restrepo Sierra, A. M.***, & Danelon, C. (2023). Clonal Amplification-Enhanced Gene Expression in Synthetic Vesicles. ACS Synthetic Biology, 12(4), 1187–1203.

Restrepo Sierra, A. M., Arold, S. T., & Grünberg, R. (2022). Efficient multi-gene expression in cell-free droplet microreactors. Plos One, 17(3), e0260420.

Carter, M. E., Helm, M., Chapman, A. V. E., Wan, E., **Restrepo Sierra, A. M.**, Innes, R. W., Bogdanove, A. J., & Wise, R. P. (2019). Convergent Evolution of Effector Protease Recognition by Arabidopsis and Barley. Molecular Plant-Microbe Interactions®, 32(5), 550–565.

* Denotes equal contribution.