Building the genome of a minimal synthetic cell

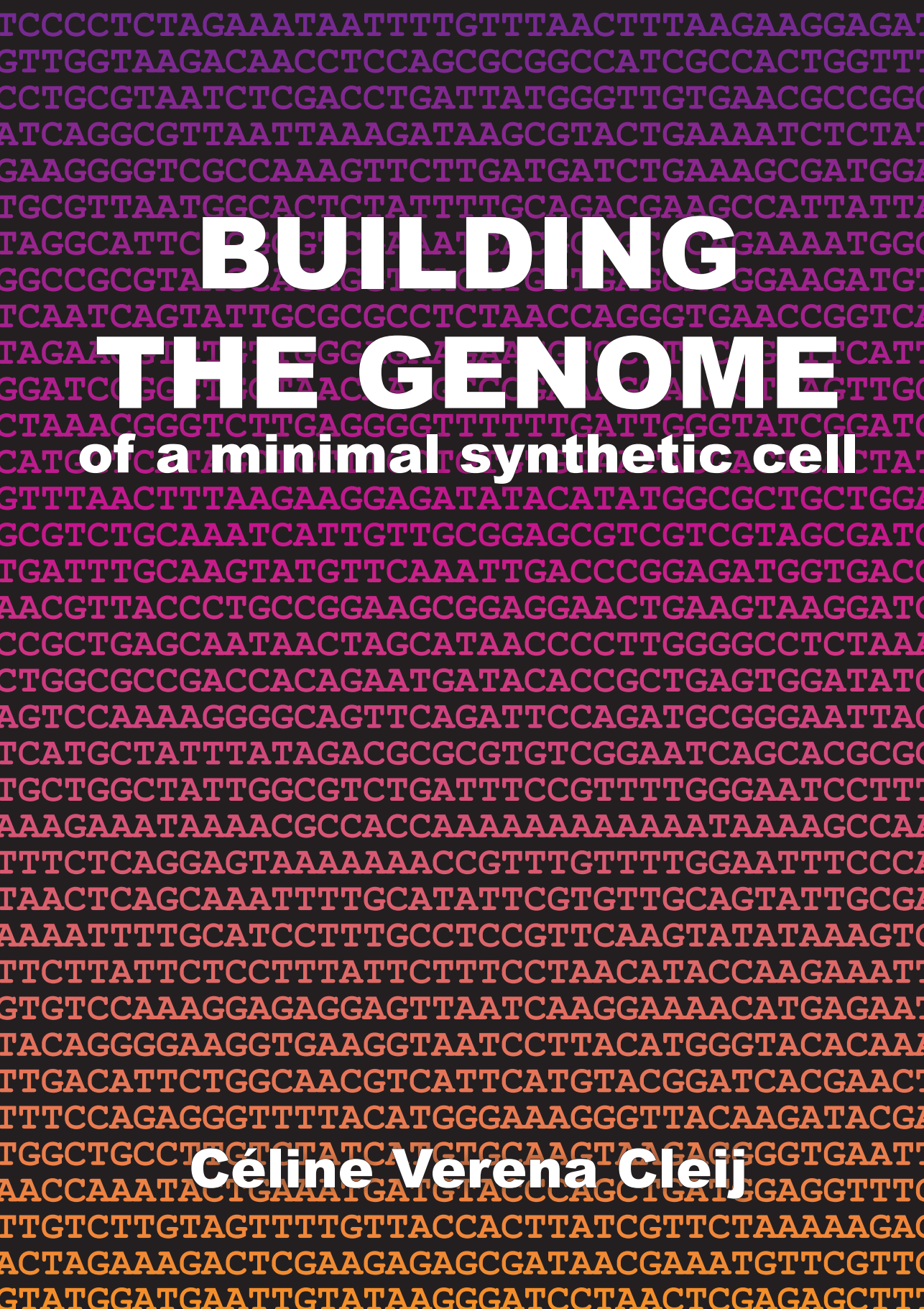
Cleij, C.V.

DOI
10.4233/uuid:bb76ad6b-e36e-4793-b1c9-54c0e6527ba7

Publication date
2025

Document Version
Final published version

Citation (APA)
Cleij, C. V. (2025). *Building the genome of a minimal synthetic cell*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:bb76ad6b-e36e-4793-b1c9-54c0e6527ba7


BUILDING
THE GENOME
of a minimal synthetic cell
Céline Verena Cleij

Building the genome of a minimal synthetic cell

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 18 September 2025 at 12:30

by

Céline Verena CLEIJ

Master of Science in Nanobiology,
Delft University of Technology & Erasmus University Rotterdam,
the Netherlands

born in Rotterdam, the Netherlands

This dissertation has been approved by the promotors.

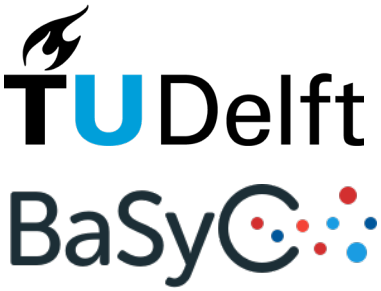
Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof.dr.ir. P.A.S. Daran-Lapujade	Delft University of Technology, promotor
Prof.dr. C.J.A. Danelon	Toulouse Biotechnology Institute, France, promotor

Independent members:

Prof.dr. G.H. Koenderink	Delft University of Technology
Prof.dr. B. Poolman	University of Groningen
Prof.dr M. Jules	INRAE, France
Dr. N.J.H.P. Claassens	Wageningen University & Research
Dr. C. Lartigue-Prat	INRAE, France
Prof.dr.ir. S.J.J. Brouns	Delft University of Technology, reserve member

The research presented in this thesis was performed at the Department of Bionanoscience and the Industrial Microbiology Section, Department of Biotechnology, Faculty of Applied Sciences, Delft University of Technology, The Netherlands. The project was funded by the Dutch Research Council (NWO) through the research programme BaSyC – Building A Synthetic Cell.



Keywords: minimal synthetic cell, synthetic genomics, minimal genome, DNA assembly, *Saccharomyces cerevisiae*, cell-free gene expression, PURE system

Cover:	Céline Verena Cleij
Layout:	Céline Verena Cleij
Printed by:	ProefschriftMaken
ISBN:	978-94-6384-821-3

An electronic copy of this dissertation is available at http://repository.tudelft.nl/

Table of contents

Summary

About 3.5 to 4 billion years ago, life on Earth emerged from non-living matter, giving rise to the first cellular organisms. Since then, countless cell divisions subjected to the process of evolution have shaped the complexity of life observed today. While scientists can modify existing cells for applications in medicine, food production and industrial biotechnology, the construction of a living cell from inanimate components remains a major scientific challenge. Rather than attempting to recreate the conditions under which life originally emerged, the bottom-up synthetic biology approach seeks to build a synthetic cell using complex biomolecules as building blocks, e.g., nucleic acids, amino acids and lipids. The research in this dissertation was conducted within the framework of the Building a Synthetic Cell (BaSyC) consortium, which aims to construct a minimal synthetic cell: a simple cell containing a minimal set of genes required and sufficient to exhibit the fundamental properties of life, namely self-maintenance, self-reproduction and evolvability. Achieving this goal will deepen our understanding of the essential principles underlying cellular life. This dissertation explores a crucial step towards the realization of a minimal synthetic cell: the *de novo* design and assembly of its genome.

The precise content of a prospective minimal synthetic genome is unknown, but it can be estimated using the three approaches described in **Chapter 1**: comparative genomics, genome reduction of existing microorganisms and a biochemical approach to define essential cellular processes. The latter allows the incorporation of genes from various sources, including extant cells and viruses, that can be rationally engineered or evolutionary optimized *in vitro* or *in vivo*. Essential cellular processes include, but are not limited to, DNA replication, transcription, translation, energy conservation, membrane synthesis and cell division. These functions are estimated to require 150 to 200 genes, resulting in a minimal genome size of approximately 150 to 200 kb. Besides a minimal genome, constructing a synthetic cell through a bottom-up approach requires assembling a compartment and a cytoplasm to kickstart gene expression. In the Danelon lab, the chosen synthetic cytoplasm is PURE system: a mixture of purified components supporting transcription, translation, and energy regeneration. Compartmentalization can be achieved by encapsulation of PURE system alongside the DNA genome in phospholipid vesicles known as liposomes. The choice of PURE system for gene expression leads to a high occurrence of repeated sequences on the genome, due to the limited availability of regulatory elements. Therefore, a suitable method for constructing minimal genomes should be compatible with a genome size exceeding 150 kb, and should allow modular assembly of transcription cassettes with homologous regulatory sequences. Additionally, the assembled constructs should be transferable to liposomes containing PURE system for *in vitro* expression. This dissertation explores *in vivo* assembly using the homologous recombination machinery of the yeast *Saccharomyces cerevisiae* to construct synthetic chromosomes (SynChrs).

Chapter 2 reviews the pivotal role of *S. cerevisiae* in synthetic genomics projects. This chapter compares DNA assembly in yeast to other existing *in vitro* and *in vivo* assembly

methods, and highlights the use of yeast as a platform for assembling entire microbial genomes. Challenges in genome assembly using yeast are identified, including SynChr extraction and purification, as well as misassembly caused by the presence of repeated sequences.

Chapter 3 systematically investigates how these challenges affect the assembly in yeast of SynChrs intended for expression in PURE system. A 67-kb test SynChr was successfully assembled from 20 fragments. One fragment encoded a fluorescent protein for expression in PURE system, nine were composed of non-coding DNA flanked by PURE regulatory sequences, and the other ten were intended for screening. To efficiently identify correctly assembled SynChrs, a screening pipeline based on auxotrophic, fluorescent, antibiotic resistance and chromogenic markers was implemented. Long-read sequencing confirmed that some assemblies were correct and that in others, repeated sequences had contributed to unintended recombination events via homologous recombination. The assembly efficiency was low (8%) and strongly reduced compared to a control SynChr that did not contain repeated PURE regulatory sequences. The test chromosome remained stable during propagation in yeast and was successfully extracted, though in picomolar quantities and contaminated with native yeast DNA. Synthesis of the fluorescent protein encoded on the SynChr was achieved in PURE system, albeit with lower expression levels than desired. This chapter concludes that further optimization of the SynChr concentration and purity is necessary to enable efficient expression of the numerous genes required in a minimal synthetic cell.

Building on these findings, **Chapter 4** describes the design and assembly of a synthetic minimal genome (SynMG1) which contains cellular modules for phospholipid biosynthesis, DNA replication and cell division. Fluorescent reporter genes were incorporated to monitor expression kinetics in PURE system. To enable replication *in vitro* by the DNA replication system from bacteriophage ϕ29, the SynMG1 design contained ϕ29 replication origins and an internal restriction site to generate a linearized variant of SynMG1 flanked with these replication origins. The same screening markers used for the test chromosome in Chapter 3 were incorporated. Moreover, with the goal of obtaining SynChr DNA in high quantities and purity, a bacterial artificial chromosome (BAC) backbone was added to enable transfer of the SynChr to *E. coli* for amplification. SynMG1, with a total size of 41 kb, was assembled from 14 fragments via homologous recombination in yeast and verified using long-read sequencing. In contrast, a larger 105-kb SynMG (SynMG2), which included a translation factor module, could not be assembled. Following extraction from yeast and amplification in *E. coli*, SynMG1 was isolated in higher quantity (nanomolar range) and purity as compared to direct purification from yeast. All encoded proteins were successfully synthesized in PURE system, as confirmed by liquid chromatography-mass spectrometry and fluorescence measurements. Furthermore, successful encapsulation and expression of SynMG1 in liposomes was demonstrated, although with a high liposome-to-liposome variability. Finally, preliminary results showed full-length replication of linearized SynMG1 by the ϕ29 DNA replication machinery.

Chapter 5 concludes this dissertation by presenting recommendations for future research. Strategies are proposed to improve SynChr construction in yeast, including enhancing the efficiency of assembling fragments with repeated sequences, optimizing the workflow from SynChr design to verification, and improving the transfer from

yeast to PURE system. Additionally, optimized Golden Gate assembly is discussed as an alternative to assembly in yeast. Finally, Chapter 5 presents strategies for *in vitro* characterization of SynChrs, for the design of larger synthetic minimal genomes, and for the evolutionary integration of the cellular functions encoded on the genome of the minimal cell.

Samenvatting

Ongeveer 3,5 tot 4 miljard jaar geleden ontstond er leven op Aarde uit niet-levende materie, wat leidde tot de eerste cellulaire organismen. Sindsdien hebben talloze celdelingen plaatsgevonden, die door het proces van evolutie hebben geresulteerd in de complexiteit van levensvormen zoals we die vandaag de dag kennen. De moderne wetenschap heeft de mogelijkheid om bestaande cellen te modificeren, met toepassingen op het gebied van geneeskunde, voedselproductie en industriële biotechnologie. Het bouwen van een levende cel vanaf de basis, uit levenloze bouwstenen, vormt echter een grote wetenschappelijke uitdaging. In plaats van te proberen de oorspronkelijke omstandigheden waaronder leven is ontstaan na te bootsen, richt de bottom-up benadering zich op het bouwen van een synthetische cel met bestaande complexe biomoleculen, zoals nucleïnezuren, aminozuren en lipiden. Dit proefschrift werd geschreven binnen het kader van het consortium Building a Synthetic Cell (BaSyC), dat als doel heeft een minimale synthetische cel te maken. Dat is een eenvoudige cel die het minimale aantal genen bevat dat nodig en voldoende is om de fundamentele eigenschappen van het leven te vertonen: zelfonderhoud, zelfreproductie en evolutie. Het bereiken van dit doel zal ons inzicht vergroten in de fundamentele principes die het leven op cellulair niveau mogelijk maken. Dit proefschrift onderzoekt een cruciale stap richting de realisatie van een minimale synthetische cel: het ontwerp en de vervaardiging van het genoom.

De precieze inhoud van een minimaal synthetisch genoom is nog onbekend, maar er kan een inschatting worden gemaakt aan de hand van de drie methoden die in **Hoofdstuk 1** worden beschreven: (i) het vergelijken van bestaande genomen om geconserveerde genen te ontdekken, (ii) genoomreductie van bestaande micro-organismen en (iii) een biochemische benadering om essentiële cellulaire processen te definiëren. Laatstgenoemde methode maakt het mogelijk om genen uit verschillende bronnen, waaronder bestaande cellen en virussen, te gebruiken. Die genen kunnen vervolgens geoptimaliseerd worden met een rationele benadering, of door middel van *in vitro* of *in vivo* evolutie. Essentiële cellulaire processen omvatten, maar zijn niet beperkt tot, DNA-replicatie, transcriptie, translatie, energieconservering, membraansynthese en celdeling. Er wordt geschat dat er 150 tot 200 genen nodig zijn om deze functies te realiseren, wat resulteert in een minimaal genoom van ongeveer 150 tot 200 kb. Voor het bouwen van een synthetisch cel moet er niet alleen een genoom gemaakt worden, maar ook een compartiment en een cytoplasma om de genexpressie op gang te brengen. In het laboratorium van Christophe Danelon wordt het PURE systeem gebruikt als synthetisch cytoplasma. Het PURE systeem bestaat uit gezuiverde componenten die transcriptie, translatie en energieherwinning mogelijk maken. Een fosfolipidemembraan kan dienen als compartiment, in de vorm van een liposoom: een blaasje dat het PURE systeem en het genoom omsluit. Het gebruik van het PURE systeem voor genexpressie resulteert in de veelvuldige aanwezigheid van herhaalde sequenties op het genoom, vanwege de beperkte beschikbaarheid van elementen voor genexpressieregulatie. Een methode om minimale genomen te bouwen moet dus geschikt zijn voor genoomgroottes van meer dan 150 kb, en moet in staat zijn om expressiecassettes met herhaalde sequenties aan elkaar te maken. Bovendien moet het mogelijk zijn om de gebouwde genomen

te verplaatsen naar liposomen die het PURE systeem bevatten voor genexpressie. Dit proefschrift onderzoekt of synthetische chromosomen (SynChrs) gebouwd kunnen worden met behulp van de gistsoort *Saccharomyces cerevisiae*. Deze gist beschikt over een systeem dat DNA-fragmenten via homologe recombinatie aaneenschakelt.

Hoofdstuk 2 bespreekt de sleutelrol van *S. cerevisiae* in synthetische genomica-projecten. Dit hoofdstuk vergelijkt homologe recombinatie in gist met andere methodes om *in vitro* of *in vivo* DNA fragmenten aan elkaar te koppelen. Ook belicht dit hoofdstuk het gebruik van gist om volledige microbiële genomen te bouwen. De uitdagingen bij het assembleren van genomen met behulp van gist worden geïdentificeerd, waaronder de extractie en zuivering van SynChrs uit gist, evenals misassemblage veroorzaakt door de aanwezigheid van herhaalde sequenties op de DNA fragmenten.

Hoofdstuk 3 onderzoekt systematisch wat het effect is van deze uitdagingen op de constructie in gist van een SynChr ontworpen voor expressie in het PURE systeem. De assemblage van 20 fragmenten tot een test-SynChr van 67 kb was succesvol. Één fragment codeerde voor een fluorescerend eiwit voor expressie in het PURE systeem. Negen fragmenten bestonden uit niet-coderend DNA, geflankeerd door sequenties die genexpressie in het PURE systeem reguleren. De overige tien fragmenten waren bedoeld voor screening. Voor de identificatie van correct gebouwde SynChrs werd een screeningspipeline geïmplementeerd, op basis van auxotrofie-, fluorescentie-, antibioticaresistentie- en chromogene markers. Door middel van long-read sequencing werd bevestigd dat sommige assemblages correct waren, terwijl in andere gevallen de herhaalde sequenties hadden geleid tot ongewenste recombinatie. De assemblage-efficiëntie was laag (8%) en sterk gereduceerd vergeleken met een controle-SynChr zonder herhaalde regulatiesequenties. Het testchromosoom bleef stabiel tijdens celdeling in gist en kon succesvol worden geïsoleerd, zij het in picomolaire hoeveelheden en met contaminatie van gist-DNA. In het PURE systeem kon het fluorescerende eiwit dat op het SynChr gecodeerd was geproduceerd worden, hoewel er minder eiwitten werden gemaakt dan gewenst. Dit hoofdstuk concludeert dat verdere optimalisatie van de concentratie en zuiverheid van SynChrs noodzakelijk is voor de efficiënte expressie van de vele genen die vereist zijn voor een minimale synthetische cel.

Op basis van deze bevindingen beschrijft **Hoofdstuk 4** het ontwerp en de assemblage van een synthetisch minimaal genoom (SynMG1), dat cellulaire modules bevat voor DNA-replicatie, celdeling en de biosynthese van fosfolipiden. Genen die coderen voor fluorescerende reporter-eiwitten werden toegevoegd om de expressiekinetiek in het PURE systeem te monitoren. Om *in vitro* replicatie mogelijk te maken door het DNA-replicatiesysteem van bacteriofaag ϕ29, bevatte het SynMG1-ontwerp ϕ29-replicatieoorsprongen en een interne restrictiesequentie. Deze restrictiesequentie is bedoeld om een lineaire variant van SynMG1 te genereren, geflankeerd door de replicatieoorsprongen. Dezelfde screeningsmarkers als in het testchromosoom van Hoofdstuk 3 werden opgenomen in het ontwerp. Daarnaast werd, met als doel SynChr-DNA in hoge hoeveelheden en zuiverheid te verkrijgen, een sequentie van een bacterieel artificieel chromosoom (BAC) toegevoegd. Dit maakt vermeerdering van het SynChr in *E. coli* mogelijk. SynMG1, met een totale grootte van 41 kb, werd geassembleerd uit 14 fragmenten via homologe recombinatie in gist en geverifieerd met long-read sequencing. Assemblage van SynMG2, een groter SynChr van 105 kb dat genen voor translatiefactoren bevatte, was niet succesvol. Na extractie uit gist en vermeerdering

in *E. coli*, werd SynMG1 geïsoleerd in hogere hoeveelheden (op nanomolaire schaal) en met een grotere zuiverheid dan bij directe zuivering uit gist. Door middel van vloeistofchromatografie-massaspectrometrie en fluorescentiemetingen werd bevestigd dat alle gecodeerde eiwitten succesvol geproduceerd waren in het PURE systeem. Bovendien was het mogelijk om liposomen te maken die SynMG1 bevatten, en werd expressie van SynMG1 in de liposomen aangetoond door fluorescentiemetingen. Er was echter een aanzienlijke variabiliteit tussen individuele liposomen. Tot slot toonden voorlopige resultaten aan dat de lineaire variant van SynMG1 volledig werd gerepliceerd door het ϕ29 DNA-replicatiesysteem.

Hoofdstuk 5 sluit dit proefschrift af met aanbevelingen voor toekomstig onderzoek. Er worden strategieën voorgesteld om de constructie van SynChrs in gist te verbeteren, met specifieke aandacht voor (i) het verhogen van de assemblage-efficiëntie van fragmenten met herhaalde sequenties, (ii) het stroomlijnen van de workflow van SynChr-ontwerp tot verificatie en (iii) het verhogen van zowel de opbrengst als de zuiverheid van uit gist geïsoleerde SynChrs voor expressie in het PURE systeem. Daarnaast wordt een alternatief voor assemblage in gist besproken: *in vitro* assemblage via een geoptimaliseerde Golden Gate-methode. Tot slot bespreekt Hoofdstuk 5 strategieën voor de *in vitro* karakterisering van SynChrs, het ontwerp van grotere synthetische minimale genomen en de evolutionaire integratie van de cellulaire functies die zijn gecodeerd op het genoom van de minimale cel.

Chapter 1

Introduction

It is intriguing to realize that all cells currently alive on Earth, estimated to total up to the staggering amount of ~10^{30} [1], emerged from pre-existing cells. As beautifully phrased by the German scientist Rudolf Virchow: *omnis cellula e cellula* [2], all cells come from cells, a discovery made by Robert Remak already in 1852 [3]. However, cells have not always existed on Earth. A transition from non-life to cellular life has taken place approximately 3.5–4 billion years ago [4, 5], followed by countless events of cell growth and division. Life as seen today is the outcome of increasing complexity during the next billions of years through the process of evolution. While it is fascinating to investigate how life has emerged from lifeless matter on early Earth, the conditions in which life emerged are currently unknown, which complicates the reconstruction of the historical steps that led to the origins of life [6]. Another exciting research field that aims to bridge the gap between lifeless molecules and living cells, is the field of synthetic cell research. Here, the goal is to construct a living cell starting from more complex biomolecules, not necessarily available on early Earth, optionally combined with synthetic components [7, 8].

Construction of a minimal synthetic cell

The bottom-up construction of a living synthetic cell is an ambitious endeavour that has seen increased popularity over the past decade, as exemplified by the numerous national and international research consortia and initiatives on this topic: MaxSynBio in Germany [9], BaSyC (https://www.basyc.nl/) and EVOLF (https://www.evolf.life/) in The Netherlands, SynCellEU in Europe (https://syntheticcell.eu/), Build-a-Cell in the USA [10] and recently SynCell Asia [11]. This research is fostered by recent technological developments in DNA synthesis, assembly and editing, sequencing and genomics, high-throughput molecular biology and screening equipment, microfluidics and modelling [12, 13]. The interest in synthetic cell construction is fueled by (i) curiosity about the fundamental principles of life, (ii) technology development that comes with synthetic cell research, and (iii) potential applications of synthetic cells. Through reconstitution of cellular processes, fundamental knowledge is gained on their components, interactions and functioning. Additionally, synthetic cell construction will inform on the essential requirements for life. Once the first bottom-up constructed synthetic cells are available, they can be used as platforms to study cellular processes (e.g., metabolic pathways, expression regulation) or system's level functionalities of living cells (e.g., evolution, adaptability, self-organization) in a simple background. Next to fundamental understanding, the construction of artificial cells will come with technology development that can be applied in related fields of biotechnology or medicine. Additionally, synthetic cells can possibly be equipped with useful features, such as the production of chemicals or medicines, and targeted delivery of medicines, or could function as biosensors or new biomaterials.

Although future applications of tailored synthetic cells are plenty, the first prototypes will be relatively simple. This dissertation was written within the framework of the BaSyC consortium, which aims to construct a minimal synthetic cell: one that has the minimal set of genes required and sufficient to sustain life under ideal laboratory conditions.

Essentials of a living cell

Defining the essential properties of cellular life can aid in the experimental design towards the construction of a minimal synthetic cell. However, the question "What is life?" is not one that has an unambiguous answer [14]. According to NASA, "Life is a self-sustaining chemical system capable of Darwinian evolution" [15]. Pier Luisi, one of the pioneers in the bottom-up construction of a minimal synthetic cell, defines three main properties of life: self-maintenance, self-reproduction and evolvability [16] (Figure 1.1A). Essential components to achieve these emerging functions are (i) an information carrier, (ii) catalysts and (iii) a compartment [17] (Figure 1.1B).

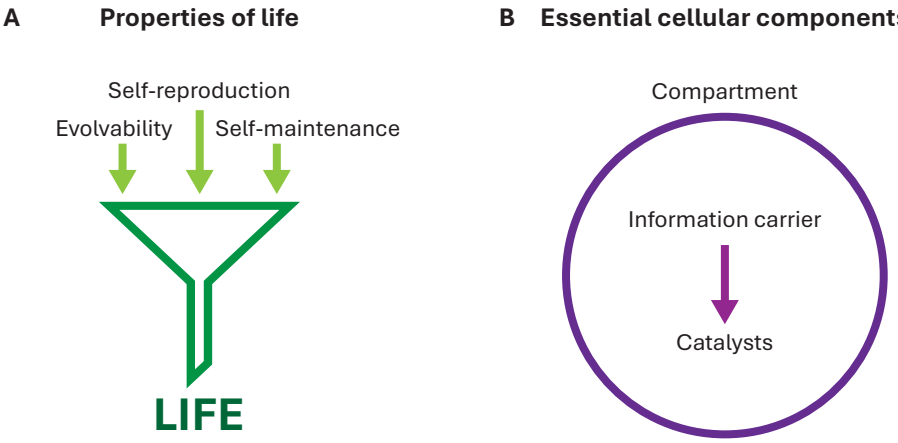


Figure 1.1: Properties and essential components of life. A) Properties that define life are self-maintenance, self-reproduction and evolvability. Figure adapted from [18]. **B)** Essential components of cells are an information carrier, catalysts and a compartment.

Possible information carriers are RNA, central player in the "RNA world hypothesis" of the origins of life, and DNA, the information carrier common to all extant cellular life [19]. Catalysis can be carried out by proteins (most common in extant life), RNA (ribozymes, with ribosomes as the most well-known example [20]) or, more recently discovered and not naturally occurring, DNA (deoxyribozymes) [21]. In synthetic cell construction efforts, the information carrier and catalysts present in extant life are the most straightforward to begin with. Therefore, the synthetic cell will have a DNA genome that can be transcribed into RNA and translated into proteins. Naturally occurring compartments are membranes composed of lipids, carbohydrates and proteins, but several synthetic options are available [22–24].

In the process of constructing a minimal synthetic cell, combining an information carrier, catalysts and a compartment should be taken as a starting point. Along the way, the cell will exhibit different degrees of "aliveness", similarly to the transition from non-life to life during the origins of life [6, 25]. It is also important to note that the complexity of the minimal cell is dependent on the availability of nutrients in the growth medium: the more nutrients available, the fewer biochemical reactions are required inside the minimal cell, thus allowing it to be simpler.

Determination of the minimal genome content

When building a minimal synthetic cell, one needs to determine which information needs to be encoded on the DNA genome. Numerous studies have estimated the minimal gene set to sustain life, based on one or multiple of the following three approaches [26]: (i) Comparative genomics of extant organisms, (ii) Genome reduction of simple organisms, (iii) Determination of the biochemical reactions necessary for life, based on the essential functions and components of a living cell as described in the previous section (Figure 1.2).

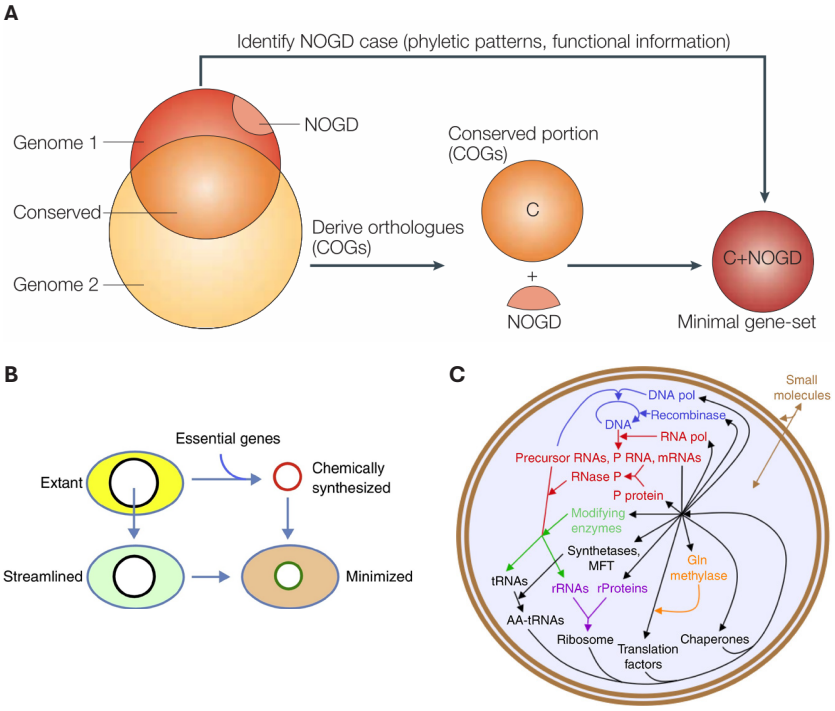


Figure 1.2: Three approaches to determine the minimal gene set. A) Comparative genomics. Two or more genomes are compared to find conserved genes. From phyletic patterns and functional information, genes that are subject to non-orthologous gene displacement (NOGD) are identified. The conserved genes are combined with the NOGD genes to form the minimal gene set. COGs = clusters of orthologous groups of proteins. Figure from [27]. **B)** Genome reduction. Genomes of extant organisms can be reduced top-down through stepwise deletion *in vivo* (streamlining). Bottom-up genome reduction can be achieved through identification of essential genes, chemical synthesis of a genome and implantation into a surrogate cytoplasm. Figure from [28]. **C)** Biochemistry. Essential biological macromolecules and pathways are identified and the corresponding coding genes are determined. Figure from [26].

Comparative genomics

Comparative genomics is the study of the evolutionary conservation of protein-coding genes by identifying orthologs in sequenced genomes, based on similarity of the corresponding protein sequences (Figure 1.2A). This field of study started in 1996 after the first full genome sequences were available for two parasitic bacterial species: the Gram-positive *Mycoplasma genitalium* and the Gram-negative *Haemophilus*

influenzae [29]. Conservation of genes in these two bacteria, that are likely more than 1.5 billion years separated from their common ancestor, are expected to be essential for cellular function. The authors Mushegian and Koonin found 240 orthologous genes. However, this set of genes was missing some enzymes in essential pathways, which can be explained by the phenomenon that the authors name "non-orthologous gene displacement": "the presence of non-orthologous genes for the same function in different organisms". After correction for non-orthologous gene displacement and removal of genes that the authors identified as parasite-specific, Mushegian and Koonin defined a minimal gene set of 256 genes. In 2003, Koonin repeated a similar analysis, based on all ca. 100 available sequenced genomes at the time. This analysis resulted in only 63 ubiquitous genes, mostly involved in translation and thus not a complete minimal gene set [27]. This result shows that non-orthologous gene displacement is a severe limitation to the comparative genomics approach, and knowledge of essential pathways is necessary to complement the conserved gene set with non-orthologous but essential genes.

Another drawback of the comparative genomics approach to arrive at a minimal gene set is the focus on naturally occurring systems in living cells: this leaves out the possibility to use synthetic or virus-derived systems to reconstitute cellular functions.

Genome reduction

Experimental determination of essential genes of a chosen organism can be achieved through knocking out one gene at a time. This approach was first taken by Itaya in 1995 through knockout of 79 randomly selected chromosomal loci in the bacterium *Bacillus subtilis*, of which only six were lethal [30]. Combination of 33 of the non-lethal knockouts resulted in viable cells, which shows that genome reduction of *B. subtilis* is possible. In 2003, a more complete assessment of gene essentiality in *B. subtilis* was carried out, which identified 271 essential genes [31]. Starting from a simpler organism, Hutchison et al. identified the essential genes of *Mycoplasma genitalium* through transposon mutagenesis [32]. *M. genitalium* has the smallest genome of any known free-living organism with a total size of 580 kb [33]. Based on the location of the inserted transposons in mutants grown together in mixed culture, 265 to 350 of the 480 protein-coding genes were suggested to be essential. Since this study did not characterize the mutants in pure culture, a new study was performed in 2006 [34] with clonal populations of mutants, which resulted in a prediction of 387 essential protein-coding genes and 43 RNA-coding genes. Surprisingly, 110 of the essential genes had unknown function or were hypothetical proteins. It should be noted that essential gene sets determined through single-gene knockouts are unlikely to be sufficient to sustain life, due to the phenomenon of synthetic lethality: sets of genes which can be individually knocked out, but for which simultaneous knockout is lethal [27].

Therefore, after identification of dispensable genes in single-gene knockouts, the next step is to use this information to generate an organism with a reduced genome, where all non-essential genes have been removed (reviewed in [28, 35–37]). Genome minimization can be done in two ways: top-down and bottom-up (Figure 1.2B). In the top-down approach, the starting point is a living organism, from which sections of the genome are serially deleted, based on the gene essentiality information gained through single-gene knockouts. In the bottom-up approach, the same information about gene

essentiality is used, but in this case to construct a genome starting from synthesized DNA oligos. This synthetic genome is then transplanted into a living cell and replaces the host genome, resulting in a minimal cell. The latter approach is highlighted in the next section. For both approaches, a reduced genome of a single organism is obtained, which is probably not the most minimal set of genes possible for each individual cellular module. Also, it will likely still contain genes of unknown function. Furthermore, this approach results in a biological cell with a minimized (synthetic) genome, rather than a fully bottom-up constructed minimal synthetic cell. The main difference is that the cytoplasm in which the minimized genome is initially expressed, is from a living cell. Even though the cytoplasm is gradually replaced by newly synthesized proteins encoded on the minimized genome during several rounds of replication, a living cell was needed to "kickstart" life from the minimized genome.

A special case: the JCVI minimal cell project

The first minimal cell that was constructed based on genome reduction of a living organism, is JCVI-syn3.0 [38]. Researchers of the John Craig Venter Institute (JCVI) had previously constructed a synthetic version of the 1.08-Mbp *Mycoplasma mycoides* genome, starting from chemically synthesized cassettes, through multiple rounds of assembly in yeast [39]. They transplanted the synthetic genome into a recipient *Mycoplasma capricolum* cell, after which the *M. capricolum* genome was lost and a cell was obtained that was controlled by the synthetic *M. mycoides* genome: JCVI-syn1.0. Next, using global transposon mutagenesis, genes of JCVI-syn1.0 were classified as essential, nonessential and, when deletion was not lethal but caused growth impairment, quasi-essential. After multiple design-build-test rounds, a new genome was designed, in which 428 JCVI-syn1.0 genes were deleted: most nonessential genes and some quasi-essential genes (Figure 1.3A). This genome was constructed from synthetic oligonucleotides through assembly *in vitro* and in yeast, and transplanted into an *M. capricolum* recipient cell, yielding a minimal cell with a genome smaller than any other autonomously replicating cell: JCVI-syn3.0 [38]. The 531-kb genome contains 438 protein-coding genes and 35 genes coding for RNA. Of these 473 genes, 149 genes could not be assigned a specific biological function, but for some a functional category could be determined. When dividing the 473 genes into four functional groups, 195 are involved in the expression of genomic information, 34 in the preservation of genomic information across generations, 84 genes encode membrane-related proteins, mostly involved in transport systems to import small molecules from the rich medium, and 81 genes are involved in cytosolic metabolism. The remaining 79 genes could not be assigned to a functional category. Phenotypically, JCVI-syn3.0 is different from JCVI-syn1.0 (Figure 1.3B): it has a growth rate of ca. 3 hours compared to 1 hour for JCVI-syn1.0, and forms segmented filaments and large vesicles during growth, while JCVI-syn1.0 grows as nonadherent spheroid cells [38].

When 19 additional genes from JCVI-syn1.0 were included in JCVI-syn3.0 and two other genes were removed, the spheroidal morphology was restored and the doubling time reduced to ca. 2 hours. This new strain was named JCVI-syn3A [40] and has a 543-kb genome, containing 493 genes of which 452 are protein-coding and 38 RNA-coding. Of the 452 protein-coding genes, 91 have no known specific biological function, which has later been reduced to 66 genes of unknown function through computational analysis [41].

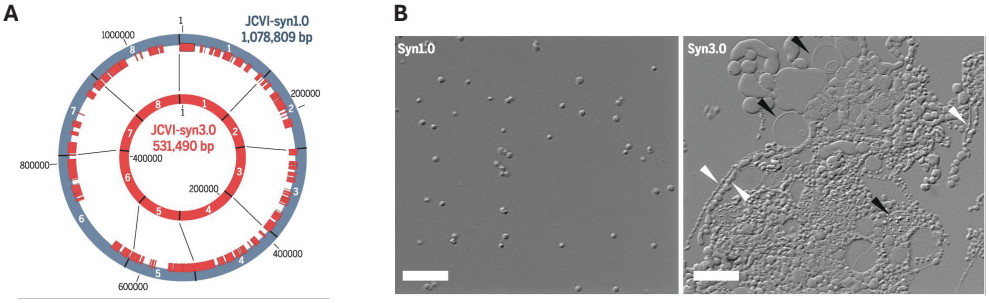


Figure 1.3: The JCVI project is an example of bottom-up genome minimization and resulted in JCVI-syn3:0. Figures adapted from [38]. **A)** JCVI-syn3.0 contains a 531-kb genome, obtained through bottom-up genome reduction of the 1.08-Mbp genome of JCVI-syn1.0. **B)** JCVI-syn1.0 (left) shows nonadherent spheroidal cell morphology. JCVI-syn3.0 (right) shows a different morphology with segmented filaments (white arrows) and large vesicles (black arrows).

Biochemistry

The third approach to estimate the minimal gene set to sustain life, is the biochemistry approach. One starts from the essential functions that the cell should have, looks into extant organisms for systems that fulfil this function, and then reconstitutes the biochemical reactions *in vitro*. This is the approach undertaken by BaSyC to build a bottom-up minimal cell.

The first study describing this approach focussed on the essential property of self-maintenance and did not take into account self-reproduction and evolvability [42]. Tomita et al. used their E-CELL software to construct a model of a hypothetical minimal cell containing 120 *Mycoplasma genitalium* genes and seven other genes, coding for transcription, translation, energy conservation and phospholipid synthesis. This hypothetic cell lacks genes for DNA replication and cell division, and can therefore not self-reproduce.

Luisi and coworkers [43] proposed three versions of a hypothetical minimal cell: (i) a minimal DNA cell based on *M. genitalium*, (ii) a 'simple-ribosome cell', which is the same as (i), but without ribosomal proteins that may not be essential for protein synthesis, and (iii) an extremely reduced cell, based on (ii), but with fewer polymerases and amino acids. Since (ii) and (iii) are based on a translation system without ribosomal proteins that is not known to extant life and that has not been experimentally reconstituted, version (i) is the most suitable to discuss in the context of this dissertation. The *M. genitalium*-based hypothetical DNA cell contains 146–150 genes, coding for DNA replication, transcription, translation and lipid synthesis. The growth medium should contain small molecules such as nucleotides, amino acids and substrates for lipid synthesis, which can permeate through the cell membrane, either through nonselective pores or via nonenzymatic facilitated diffusion by temporarily increasing the lipophilicity of these small molecules [44]. No genes for cell division are included, which would occur through physical forces as the cell membrane grows. With this minimal gene set, the cell should theoretically be capable of self-maintenance and self-reproduction.

Forster and Church [26] propose a similar minimal cell that consists of a lipid bilayer

vesicle in which DNA replication, transcription, RNA processing, translation, protein folding and post-translational modification take place (Figure 1.2C). Again, small molecules are not synthesized inside the cell, but imported through the lipid membrane. The proposed 113-kb genome contains 151 genes (113 protein-coding and 38 RNA-coding) that are not taken from *M. genitalium*, but instead from *E. coli* and viruses. No genes for membrane growth or cell division are included, as this minimal cell model relies on spontaneous vesicle growth through incorporation of lipids or fusion with other vesicles, and spontaneous division [17, 26].

As shown by the example of Forster and Church, the biochemistry approach allows the combination of systems from different organisms (eukaryotes, prokaryotes and viruses). For each biological function to be reconstituted, one can look into the systems of extant organisms that fulfil this function, and choose the most minimal mechanism. Additionally, this approach allows for the incorporation of synthetic modules or physical processes, for example to achieve membrane growth and division. In the bottom-up construction of a minimal cell, the systems encoded on the genome will be chosen through the biochemistry approach. Therefore, the genome is estimated to harbor approximately 150 genes, less than any extant organism on Earth (Figure 1.4).

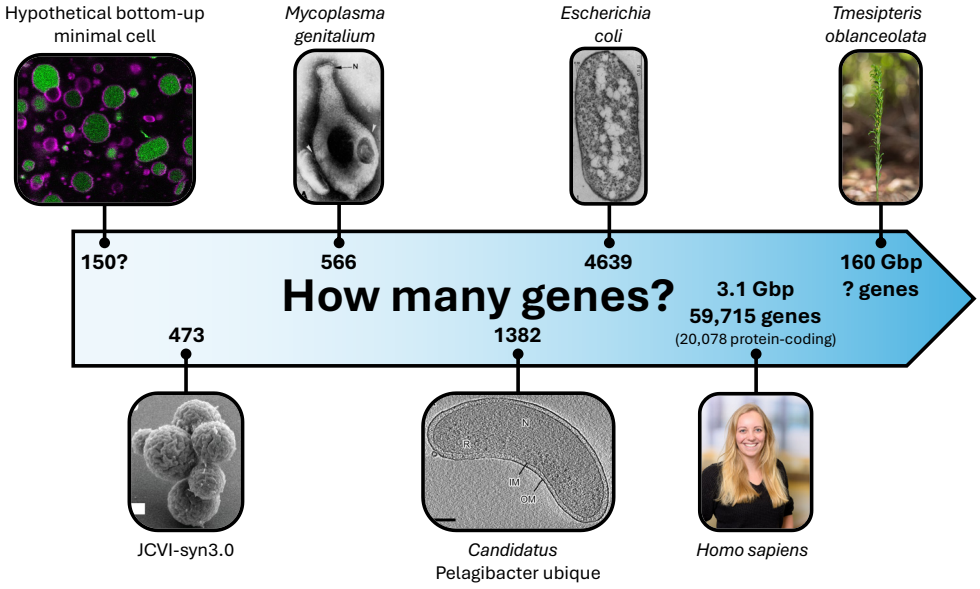


Figure 1.4: Gene counts of the hypothetical bottom-up minimal cell compared to extant organisms with particularly small or large genomes. The hypothetical bottom-up minimal cell is estimated to contain ca. 150 genes. JCVI-syn3.0 is currently the most minimal cell known to be alive. *Mycoplasma genitalium* possesses the smallest genome of any organism capable of growing independently. *Candidatus* Pelagibacter ubique has the smallest genome of all non-parasitic or -symbiotic organisms. *Escherichia coli* is the most well-studied prokaryote. The human genome size is included for reference. The largest known genome is that of the fern *Tmesipteris oblanceolata*. Information about gene counts was taken from: Hypothetical bottom-up minimal cell [26, 43], JCVI-syn3.0 [38], *M. genitalium*, *Ca.* Pelagibacter ubique, *E. coli* and *H. sapiens* from NCBI RefSeq (https://www.ncbi.nlm.nih.gov/refseq/), *T. oblanceolata* [45]. Images were taken from: Hypothetical bottom-up minimal cell (confocal microscopy image from our lab), JCVI-syn3.0 [38], *M. genitalium* [46], *Ca.* P. ubique [47], *E. coli* [48], *H. sapiens* (picture of me), *T. oblanceolata* (Hans Hillewaert).

A disadvantage of the biochemistry approach, is the anticipated difficulty of integrating subsystems from various origins. Each module may require different growth conditions, compete for substrates or produce (intermediate) products with inhibitory effects on the other subsystems in the cell.

Combined approaches

Several studies combined two or more of the approaches discussed in the previous sections, to arrive at a more substantiated or complete minimal set of genes.

In 2004, Gil and colleagues [49] combined the data from studies which had identified essential genes of five endosymbiotic bacteria, *B. subtilis*, *E. coli*, *M. genitalium*, *Mycoplasma pneumoniae, Staphylococcus aureus and Phytoplasma asteris*, through comparative genomics or by experimental determination of gene essentiality. In their resulting gene list, gaps in essential metabolic pathways were complemented. The proposed minimal genome contains 206 protein-coding genes involved in DNA metabolism (replication, repair, restriction and modification), transcription, translation (including tRNA and ribosome maturation and modification), RNA degradation, protein processing, folding and secretion, cell division, transport, energetic and intermediary metabolism (glycolysis, proton motive force generation, pentose phosphate pathway, lipid metabolism, biosynthesis of nucleotides and cofactors) and eight poorly characterized genes. The hypothetical cell would be grown in complex medium containing sugars, amino acids, free bases (adenine, guanine and uracil), cofactor precursors and lipid precursors.

A reviewed version of the abovementioned minimal gene set was published in 2014 by Rosario Gil [50], which includes RNA-coding genes and excludes poorly characterized genes. The new gene set is composed of 187 to 205 protein-coding genes and 35 to 38 RNA-coding genes.

In 2016, Ye and colleagues proposed a novel strategy that combines experimentally determined essential genes with comparative genomics and a biochemistry approach [51]: (i) starting from a database of experimentally determined essential genes from 15 bacterial species, (ii) orthologs were identified and retained when conserved among over half of the reference genomes (half-retaining), and (iii) the gene set was supplemented by genes necessary to construct a viable metabolic network. This resulted in a minimal set of 314 genes, which was compared with two previously identified sets [29, 49]: of the 141 genes that overlapped between the gene sets of Gil et al. and Mushegian & Koonin, 128 (91%) were also present in the gene set of Ye et al. The remaining 13 genes were added to the essential gene set of Ye et al., resulting in a total of 327 genes. Of the final gene set, 107 genes were not present in either of the previous two sets: 62 genes were identified through the half-retaining strategy and 45 genes through the reconstruction of a viable metabolic network. Both strategies were chosen to minimize the effect of non-orthologous gene displacement.

Finally, Rees-Garbutt et al. [52] modelled minimal gene sets that were found previously with any of the three approaches described above, by simulating genome edits on a whole-cell model of *M. genitalium*. All minimal gene sets produced a non-dividing cell *in silico*. The gene sets were “repaired” to restore *in silico* division, through reintroduc-

tion of essential genes from *M. genitalium*. The smallest minimal gene sets that could be obtained with this method consisted of 259 genes (protein-coding and RNA-coding), and were constructed through reintroduction of 128 genes to the gene set of Forster and Church [26], or through reintroduction of 119 genes to the set by Tomita and colleagues [42].

Essential modules for the minimal cell

From the studies described above, it is clear that it is difficult to define one single minimal gene set. However, we can identify cellular modules which are present in (almost) all gene sets described above: DNA replication, transcription, translation, energy conservation, membrane synthesis and cell division. Possibly, genes for DNA repair, RNA and protein modification, protein folding and transport are essential as well. For the bottom-up construction of the minimal synthetic cell, the genome content will be determined based on these essential modules. For each module, simple systems from extant organisms can be tested for *in vitro* reconstitution, or a synthetic system can be developed.

Bottom-up construction of a minimal synthetic cell

While the determination of the minimal genome content is a key design step, a living cell cannot be reduced to its sole genome. In contrast to the JCVI minimal cell that has been produced by insertion of a minimal synthetic genome into a host cell and deletion of the host genome, the bottom-up construction of a minimal cell requires the construction of a compartment and a cytoplasm to kickstart expression of the genome.

The cytoplasm

The cytoplasm of the minimal cell should contain the machinery for transcription and translation of the genome content. Two types of *in vitro* transcription-translation (IVTT) systems, also referred to as cell-free expression (CFE) or cell-free protein synthesis (CFPS) systems, are available: (i) systems based on cell extracts and (ii) systems reconstituted from purified components (reviewed recently in [53]). Extract-based CFE systems are produced through cell lysis, centrifugation to remove insoluble debris and addition of supplements such as energy sources and a suitable buffer. The cell extract can be obtained from a broad range of organisms, both prokaryotic and eukaryotic. In the context of a minimal cell, prokaryotic CFE systems, such as the well-characterized systems produced from *E. coli* cell lysates [54] or the recently developed JCVI-syn3A-based CFE system [55], would be most suitable. While extract-based CFE systems can achieve high protein yields (> 2–4 mg mL^{-1}, [53]) and are relatively easy to make in house, their use as cytoplasm for the minimal cell is limited. First, the content of cell extracts is not minimal, not completely defined and characterized, and not easily tuneable. It contains components that are unnecessary and sometimes even harmful for transcription and translation: RNase in the JCVI-syn3A-based system [55] and exonucleases that degrade linear DNA templates in *E. coli*-based systems [53]. More importantly, extract-based CFEs are incompatible with the aim of constructing a synthetic cell that can self-reproduce: since extracts are composed of a complex mixture of known and unknown biomolecules, it is impossible to encode the production of their components

on the genome, thereby eliminating the possibility of sustaining life over multiple cell cycles. As an alternative to cell extracts, CFE systems can be reconstituted from purified components. The first and most well-established CFE system composed of purified components is the Protein synthesis Using Recombinant Elements (PURE) system [56, 57]. This system contains all components necessary for transcription, translation, aminoacylation, and energy regeneration (Figure 1.5). Transcription is carried out by the RNA polymerase from bacteriophage T7. Most other components are derived from *E. coli*.

PURE system can be made in house using published protocols [56–58] and is commercially available from GeneFrontier, New England Biolabs and Creative Biolabs. Its composition is completely defined and tuneable. Although not fully achieved yet [59], in theory, PURE can be regenerated by encoding all enzymes and RNA on the minimal genome, and supplying small molecules in the medium. Current limitations of PURE system are its limited lifetime in batch reactions [60], production of truncated proteins [60, 61], and low total protein yield, which is about one to two orders of magnitude lower than what is required for PURE self-production [62, 63]. However, ongoing efforts to optimize PURE system are promising (reviewed in [59]).

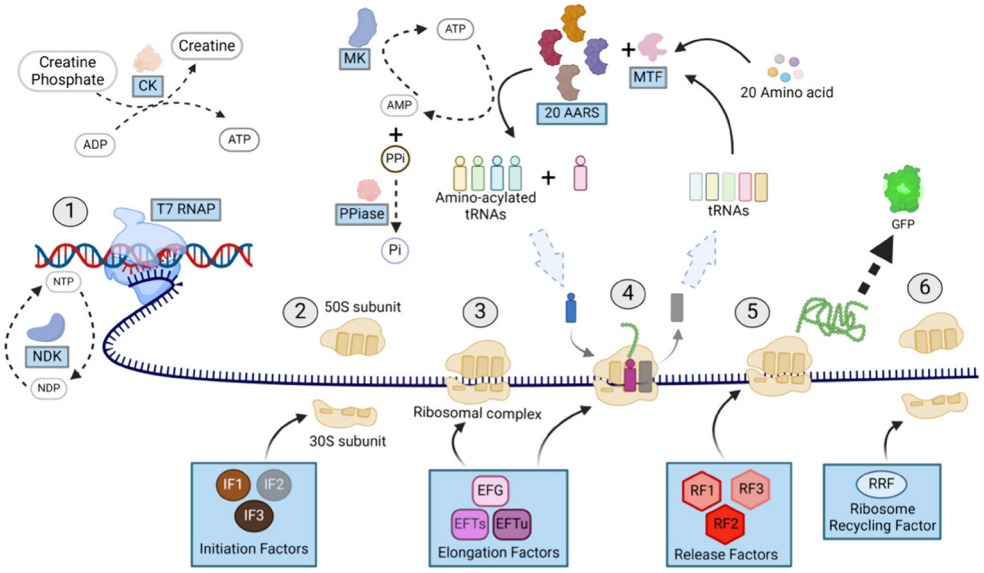


Figure 1.5: Schematic representation of all components and their function in PURE system. PURE system contains all components necessary for transcription (T7 RNAP & NTPs), translation (ribosomes, translation factors and GTP), aminoacylation (AARSs, MTF, tRNAs, amino acids and ATP) and energy regeneration (CK, NDK, MK, PPiase and creatine phosphate) in a suitable buffer. CK = creatine kinase, NDK = nucleoside-diphosphate kinase, MK = myokinase, PPiase = inorganic pyrophosphatase, AARS = aminoacyl-tRNA synthetase, MTF = methionyl-tRNA formyltransferase. Figure from [64].

The compartment

Next to a genome and a cytoplasm, a compartment is essential to build a cell, as it provides an evolutionary unit, plays important organizational roles, and regulates exchanges with the external environment. Natural compartments are enclosed by a membrane composed of lipids, proteins and carbohydrates. *In vitro*, compartmentalization can be achieved either with or without a membrane. Options to spatially confine molecules without membranes are hydrogels and coacervates. Although strategies for the growth and division of membrane-free compartments could be envisaged [65], membrane-based compartments offer a more suitable chassis for a synthetic cell. Vesicles can be produced with membranes made of lipids (droplets and liposomes), polymers (polymersomes), proteins (proteinosomes) or inorganic materials [22–24]. The use of inorganic materials and exotic polymers poses challenges on their production from a genome. Compared to water-in-oil or double-emulsion droplets, liposomes with embedded membrane proteins have the advantages that they resemble natural membranes and that the production of their constituents can be encoded on the genome. Additionally, they are compatible with encapsulated CFE systems. Therefore, most efforts towards the construction of a minimal cell rely on the use of liposomes.

Danelon lab approach

The Danelon lab, one of the two labs in which the research described in this dissertation was performed, aims to build a synthetic cell with a bottom-up approach, using PURE system for expression of DNA templates inside liposomes.

Reconstitution of essential cellular modules in this framework has been the focus of the lab over the past decade (Figure 1.6):

- DNA replication has been achieved through reconstitution of the protein-primed DNA replication system of bacteriophage ϕ29 inside liposomes [66]. Moreover, the 3.2-kb self-replicating DNA can undergo evolution *in vitro* [67].
- Thirty-two translation factor proteins have been expressed from a single plasmid in PURE system, an important step towards self-production of PURE system [61].
- Membrane synthesis has been realized by gene-encoded enzymes capable of phospholipid biosynthesis [68, 69].
- Liposomes could be deformed, in some cases constricted, with expressed cytoskeletal [70] or bacterial division proteins [71, 72].

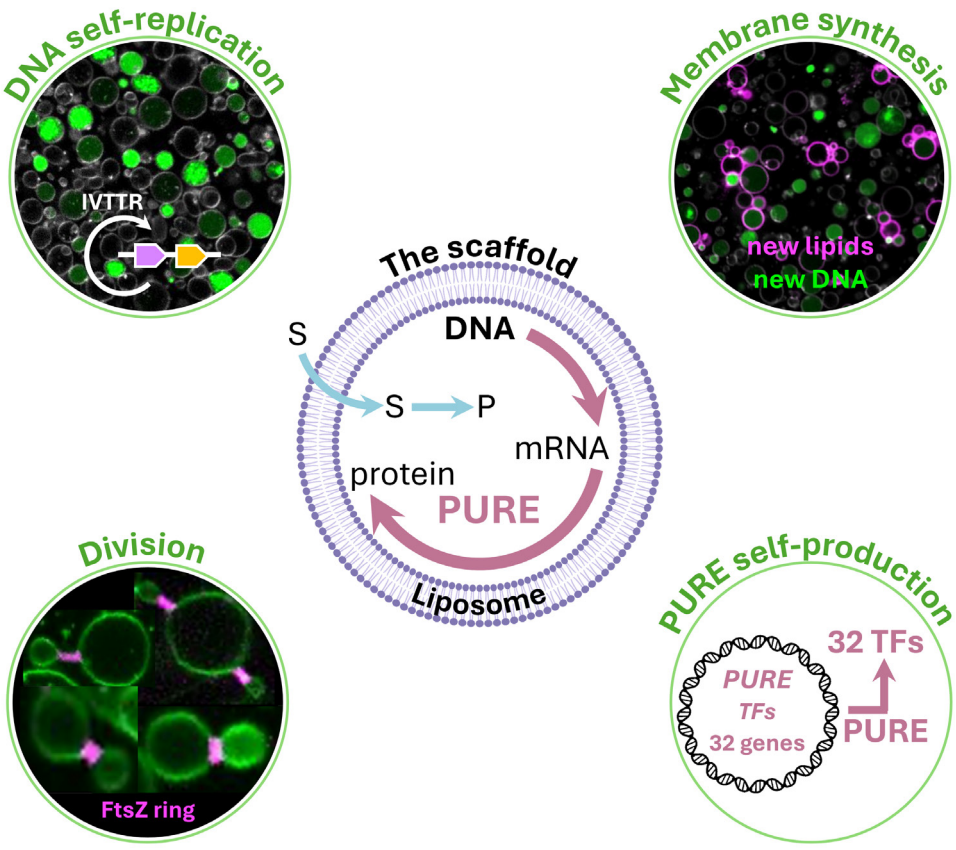


Figure 1.6: Essential cellular modules studied in the Danelon lab in the framework of gene-expressing liposomes. The DNA replication system of bacteriophage ϕ29 was reconstituted inside liposomes and self-replication of the encoding DNA template was achieved. Membrane synthesis was achieved through reconstitution of the Kennedy pathway from *E. coli*. Division proteins from *E. coli* were expressed inside liposomes and resulted in the formation of an FtsZ ring and constriction of the membrane, but no division. The first steps towards PURE self-production were taken by expression of 30 translation factors (encoded by 32 genes) in PURE system. IVTTR, *in vitro* transcription-translation-replication; S, substrate; P, product; TFs, translation factors.

Based on the systems studied in the Danelon lab for DNA replication, transcription, translation, energy conservation, membrane synthesis and cell division, the total number of genes in the minimal genome will be approximately 160 (Table 1.1).

Table 1.1: The potential content of a minimal genome, based on the modules studied in the Danelon lab

Module	Proposed system	Component(s)	Number of genes	Reference for (partial) *in vitro* reconstitution in PURE system or cell extract
DNA replication	ϕ29 phage protein-primed DNA replication system	DNA polymerase, terminal protein, DNA binding proteins	4	[66]
Transcription	PURE system	T7 RNA polymerase	1	[73, 74]
Translation	PURE system	Ribosome	54 protein-coding 3 rRNA-coding	[75–79]
		tRNAs	48 tRNA-coding	[80]
		tRNA synthetases & methionyl-tRNA formyltransferase	23	[61, 63, 81]
		Initiation factors	3	[61, 63, 81]
		Elongation factors	3	[61, 63, 81]
		Release factors & ribosome recycling factor	4	[61, 63, 81]
Energy conservation	PURE system	Energy regeneration enzymes	4	[63, 73]
Membrane synthesis	*E. coli* Kennedy pathway + PmtA from *Rhodobacer sphaeroides*	Enzymes for production of phospholipids PE, PG & PC	8	[69, 82]
Cell division	*E. coli* division machinery	Z-ring	2	[71, 83]
		Min system	3	[72, 83]
TOTAL			**160 genes** **109 protein-coding** **51 RNA-coding**	

For each of the modules described in Table 1.1, alternative or additional systems with the same functions can be employed in case the proposed mechanism does not suffice. Some suggestions for supplementary systems that have been (partially) reconstituted or proposed in the context of synthetic cell construction are listed in Table 1.2. Note that this list is not complete, but rather a list of suggestions.

Table 1.2: Suggestions for alternative or additional systems that can substitute or complement the systems proposed in Table 1.1

Module	Alternative or additional system	Component(s)	Number of genes	Reference for (partial) *in vitro* reconstitution in PURE system or cell extract
DNA replication	*E. coli*	Replication-cycle reaction (RCR)	25	With purified proteins: [84–86]
	Φ29 phage rolling-circle replication + Cre-loxP recombination	DNAP & Cre recombinase	2	[87, 88]
Transcription	SP6 phage transcription	SP6 RNA polymerase	1	[89]
	T3 phage transcription	T3 RNA polymerase	1	[90]
	E. coli transcription	*E. coli* RNA polymerase core enzyme without ω subunit	3	[91]
		Sigma factor	1	[91]
		Transcription elongation factors	2	[92]
Translation	Modified PURE system	Minimal set of tRNAs	21	[93]
Energy conservation	*L. lactis* arginine breakdown pathway for ATP synthesis	ArcA, ArcB, ArcC1, ArcD2	4	[94]
	Pyruvate-acetate pathway	Pox5, AckA, KatE	3	[95]
	RuBisCO-mediated ATP synthesis	RuBisCO	1	[96]
		Glycolytic enzymes	3	[96]
	Artificial photosynthesis	ATP synthase	8	[97, 98]
		Bacteriorhodopsin	1	[97]
Membrane synthesis	Vesicle fusion	-	0 additional	Reviewed in [99]
	Synthetic pathway for phospholipid production	FadD10	1	[100]
Cell division	Dynamin A combined with DNA nanostars	Dynamin A	1	[101]
	Crenarchaeota Cdv system	Cdv proteins	3–6	Hypothesized by [102–104]
	Actomyosin	Actin, Myosin and auxiliary protein(s)	Unknown	Reconstitution with purified proteins by [105, 106], reviewed in [107]
	Spontaneous deformation and scission through excess lipid biosynthesis	-	0 additional	Hypothesized by [17, 108]

Possibly, the modules mentioned in Table 1.1 and Table 1.2 are incomplete, or extra modules are necessary to achieve a full cell cycle. Post-transcriptional modification of tRNAs might be required, for which Hibi and colleagues used a set of nine *E. coli* enzymes to modify a minimal set of *in vitro* transcribed tRNAs [93]. Also, especially for proteins from non-prokaryotic origin, protein folding might need to be improved through the use of chaperones. A possible set of chaperones from *E. coli* origin includes DnaK, DnaJ, GrpE, GroEL, GroES and Trigger Factor, which would add another six genes to the minimal genome. Additionally, although liposomes can be naturally permeable to some small molecules [109], uptake of nutrients might need to be improved by the incorporation of a pore-forming protein, for which connexin-43 [109] or α-hemolysin [110] could be suitable candidates. Furthermore, the proposed gene set in Table 1.1 lacks genes involved in DNA repair, chromosome segregation, and regulation, all identified as essential modules of JCVI-syn3.0 [111], but not included in the proposed minimal genomes for self-reproducing cells determined with the biochemical approach [26, 43].

Having reconstituted some of the modules separately, the Danelon lab is now shifting its focus to integration of the modules through directed evolution [25, 112]. Integration comes with the need of combining multiple genetic modules on a single DNA template, and eventually constructing a full minimal genome. The approach to construct this genome is outlined in the next section.

Bottom-up construction of a synthetic minimal genome

Genome characteristics

Based on the systems described in Table 1.1, the number of genes on the minimal genome will be around 150–200, which corresponds to a genome size of approximately 200 kb. Expression of genes using PURE system in liposomes can be achieved from both linear and circular templates and thus both configurations can be envisaged for the minimal genome. The main determinant of the genome configuration will be the DNA replication system. Possible DNA replication mechanisms to amplify circular chromosomes are ϕ29 rolling-circle replication combined with recombination, and the *E. coli*-based replication-cycle reaction (Table 1.2). In case of a linear genome, replication could be carried out by a protein-primed DNA replication system like the one of the bacteriophage ϕ29, which is the system of choice in the Danelon lab [66]. This replication mechanism imposes a linear design with origins of replication at both ends of the synthetic genome.

Another design criterium is the number of chromosomes for the synthetic cell: should the genome be made of a single DNA molecule or should it be multipartite (the genes are distributed on at least two molecules)? While a multipartite genome presents the advantage that each of its chromosomes is shorter, hence potentially easier to assemble, this design may require an active DNA segregation mechanism to efficiently partition all chromosomes in the daughter cells during division. Here again, the properties of the replication system may help rationalise the design. If the processivity of the ϕ29 DNA polymerase limits the replication efficiency, then a multipartite genome, with smaller chromosomes, is more suitable. Alternatively, if replication initiation is the bottleneck, a single chromosome should be favored. The maximum size for robust replication by

the ϕ29 apparatus in PURE system has not been determined yet. Therefore, at the start of this project, both design strategies were considered.

Finally, the occurrence of repeated sequences will be inevitable in a minimal genome for a PURE system-based synthetic cell. As transcription is controlled by the T7 RNA polymerase in the commercial kit, each gene should be cloned between a T7 promoter and a T7 terminator. Additionally, a ribosomal binding site (RBS) compatible with the *E. coli*-based translation machinery of PURE system should be present upstream of each gene. Polycistronic expression is possible, but protein synthesis levels of decrease with distance from the promoter [113, 114]. Of course, diversification of regulatory genetic elements, both to ease genome construction and to tune expression levels, could be achieved by employing additional RNA polymerases (Table 1.2), alternative terminators [115], and libraries of promoters [116] and RBS sequences [117]. Still, the minimal genome will likely have a high occurrence of repeats due to the limited repertoire of regulatory sequences.

Requirements for a genome construction method

Knowing the approximate content and characteristics of the minimal genome, a suitable method for its bottom-up construction should be chosen or developed. Aside from compatibility with a genome size of approximately 200 kb and the presence of repeated sequences, ideally, the method should enable modular assembly of transcription cassettes, facilitating the engineering of genome content and the production of genome libraries. Based on the estimated gene count of the minimal genome, 150–200 fragments should be assembled. The resulting constructs should be transferable to PURE system in sufficient amounts and purity to allow for protein expression.

Commonly used methods to assemble plasmids *in vitro* are Gibson assembly and Golden Gate cloning [118]. The former is limited in the number of fragments that can be assembled at once [119]. The latter, which is based on restriction-ligation, has been optimized to assemble 52 fragments into a 40-kb phage genome [120]. The authors of the study expect that 52 is the maximum number of fragments that can simultaneously be assembled with this method and assembly was facilitated by extremely stringent selection of correct assemblies [121]. Also based on restriction-ligation, assembly using the BioBrick standard was employed for the stepwise assembly of a 65-kb plasmid from 30 fragments inside a bacterial artificial chromosome (BAC) backbone [122]. BioBrick assembly is sequential and therefore time-consuming. Furthermore, all *in vitro* assembly methods rely on selection and amplification in *E. coli* [123], which is limited by unpredictable toxicity [124] or undesired recombination [122].

Because of the abovementioned limitations of *in vitro* assembly methods, most projects involving the construction of large genomes have relied instead on *in vivo* assembly using the yeast *Saccharomyces cerevisiae* (reviewed in Chapter 2 of this dissertation and summarized in the next section).

Assembly of a minimal genome in yeast

Assembly in yeast utilizes the homologous recombination machinery *of Saccharomyces cerevisiae*, capable of assembling overlapping linear DNA fragments into linear or circular constructs. This DNA assembly method meets most of the abovementioned requirements, especially regarding the size of the assembled genomes and modularity. Notably, *in vivo* assembly in the yeast *Saccharomyces cerevisiae* was employed for the construction of genomes in the JCVI minimal cell project with an assembly size of 1 Mbp [39], for the simultaneous assembly of up to 25 fragments [125] or for the assembly of more than 1000 fragments in sequential steps of *in vitro* and *in vivo* assembly [39]. In the lab of Pascale Daran-Lapujade, the second lab in which the research described in this dissertation was performed, simultaneous assembly of 44 fragments into a 100-kb chromosome was achieved [126]. Assembly in yeast was also applied on a large scale by the lab of Jason Chin, to assemble stretches of 10-kb recoded *E. coli* genome into ca. 100-kb BACs [127].

Encouraged by the potential of yeast to assemble large constructs from many fragments, we seek to investigate its capabilities to construct a minimal genome for the bottom-up synthetic cell. Compared to earlier studies involving assembly of large constructs in yeast, our application for the minimal cell comes with new challenges: (i) the occurrence of more repetitive sequences on the minimal genome than found in existing genomes and (ii) the chromosome transfer from yeast to PURE system.

Dissertation outline

The goal of this dissertation is to design and construct synthetic chromosomes for the PURE-based minimal cell.

In short, the following is discussed in the next chapters of this dissertation:

- In **Chapter 2**, we review the role of S*accharomyces cerevisiae* in synthetic genomics. We discuss the advantages and disadvantages of various *in vitro* and *in vivo* DNA assembly methods and highlight the ability of yeast to assemble large constructs from many fragments. We also discuss challenges in genome assembly using yeast.
- In **Chapter 3**, we explore the compatibility of synthetic chromosome assembly in yeast with expression in PURE system. We identify two main challenges: assembly of DNA with repeated regulatory sequences and extraction of the assembled chromosomes with sufficient yield and purity to allow cell-free expression. We design and assemble a mock synthetic chromosome, develop a pipeline to screen for correct assemblies, optimize a protocol for chromosome extraction from yeast, and show expression of a fluorescent reporter gene in PURE system.
- In **Chapter 4**, we design and construct in yeast a synthetic chromosome encoding multiple modules for the minimal cell. We develop a pipeline to transfer the chromosome from yeast to *E. coli*, achieving high DNA yield and purity after isolation. We show that multiple fluorescent markers under the control of two different promoters can be expressed from the chromosome in bulk PURE reactions and in liposomes. Mass spectrometry analysis reveals that all proteins were synthesized.

Additionally, preliminary experiments show successful encapsulation and expression of the chromosome in liposomes, and replication of the linearized 41-kb chromosome by the reconstituted ϕ29 DNA replication system.

- In **Chapter 5**, we give recommendations for future research regarding improvement of synthetic chromosome construction in yeast, strategies for *in vitro* characterization of assembled chromosomes, design of larger synthetic minimal genomes, and the integration of cellular functions encoded on the genome of the minimal cell.

References

1. Crockford,P.W., Bar On,Y.M., Ward,L.M., Milo,R. and Halevy,I. (2023) The geologic history of primary productivity. *Curr. Biol.*, **33**, 4741–4750. https://doi.org/10.1016/j.cub.2023.09.040
2. Virchow,R. (1855) Cellular-Pathologie. *Arch. für Pathol. Anat. und Physiol. und für Klin. Med.*, **8**.
3. Remak,R. (1852) Ueber extracellulare Entstehung thierischer Zellen und über die Vermehrung derselben durch Theilung. *Arch. für Anat. Physiol. und wissenschaftliche Med.*
4. Pearce,B.K.D., Tupper,A.S., Pudritz,R.E. and Higgs,P.G. (2018) Constraining the Time Interval for the Origin of Life on Earth. *Astrobiology*, **18**, 343–364. https://doi.org/10.1089/ast.2017.1674
5. Nisbet,E.G. and Sleep,N.H. (2001) The habitat and nature of early life. *Nature*, **409**, 1083–1091. https://doi.org/10.1038/35059210
6. Sutherland,J.D. (2017) Opinion: Studies on the origin of life—The end of the beginning. *Nat. Rev. Chem.*, **1**, 1–8. https://doi.org/10.1038/s41570-016-0012
7. Blain,J.C. and Szostak,J.W. (2014) Progress toward synthetic cells. *Annu. Rev. Biochem.*, **83**, 615–640. https://doi.org/10.1146/annurev-biochem-080411-124036
8. Guindani,C., Caire da Silva,L., Cao,S., Ivanov,T. and Landfester,K. (2022) Synthetic Cells: From Simple Bio-Inspired Modules to Sophisticated Integrated Systems. *Angew. Chemie Int. Ed.*, **61**, e202110855. https://doi.org/10.1002/anie.202110855
9. Schwille,P., Spatz,J., Landfester,K., Bodenschatz,E., Herminghaus,S., Sourjik,V., Erb,T.J., Bastiaens,P., Lipowsky,R., Hyman,A., et al. (2018) MaxSynBio: Avenues Towards Creating Cells from the Bottom Up. *Angew. Chemie Int. Ed.*, **57**, 13382–13392. https://doi.org/10.1002/anie.201802288
10. Frischmon,C., Sorenson,C., Winikoff,M. and Adamala,K.P. (2021) Build-a-cell: Engineering a synthetic cell community. *Life*, **11**, 1–7. https://doi.org/10.3390/life11111176
11. Fu,M., Li,X. and Zhao,W. (2024) The Asian Synthetic Cell Initiative: Highlights from the first SynCell Asia Workshop. *Natl. Sci. Rev.*, **nwae377**. https://doi.org/10.1093/nsr/nwae377
12. Elani,Y. (2021) Interfacing Living and Synthetic Cells as an Emerging Frontier in Synthetic Biology. *Angew. Chemie Int. Ed.*, **60**, 5602–5611. https://doi.org/10.1002/anie.202006941
13. Schwille,P. (2011) Bottom-Up Synthetic Biology: Engineering in a Tinkerer's World. *Science*, **333**, 4. https://doi.org/10.1126/science.121170
14. Luisi,P.L. (1998) About various definitions of life. *Orig. Life Evol. Biosph.*, **28**, 613–622. https://doi.org/10.1023/A:1006517315105
15. NASA Astrobiology (2025) About Life Detection. https://astrobiology.nasa.gov/research/life-detection/about/
16. Luisi,P.L., Ferri,F. and Stano,P. (2006) Approaches to semi-synthetic minimal cells: A review. *Naturwissenschaften*, **93**, 1–13. https://doi.org/10.1007/s00114-005-0056-z
17. Szostak,J.W., Bartel,D.P. and Luisi,P.L. (2001) Synthesizing life. *Nature*, **409**, 387–390. https://doi.org/10.1038/35053176
18. Adamski,P., Eleveld,M., Sood,A., Kun,Á., Szilágyi,A., Czárán,T., Szathmáry,E. and Otto,S. (2020) From self-replication to replicator systems en route to de novo life. *Nat. Rev. Chem.*, **4**, 386–403. https://doi.org/10.1038/s41570-020-0196-x
19. Alberts,B., Johnson,A., Lewis,J., Raff,M., Roberts,K. and Walter,P. (2002) The RNA World and the Origins of Life. In *Molecular Biology of the Cell, 4th edition*. Garland Science, New York.
20. Cech,T.R. (2000) The ribosome is a ribozyme. *Science*, **289**, 878–879. https://doi.org/10.1126/science.289.5481.878
21. Breaker,R.R. and Joyce,G.F. (2014) The expanding view of RNA and DNA function. *Chem. Biol.*, **21**, 1059–1065. https://doi.org/10.1016/j.chembiol.2014.07.008
22. Cho,E. and Lu,Y. (2020) Compartmentalizing cell-free systems: Toward creating life-like artificial cells and beyond. *ACS Synth. Biol.*, **9**, 2881–2901. https://doi.org/10.1021/acssynbio.0c00433
23. Gonzales,D.T., Suraritdechachai,S. and Tang,T.-Y.D. (2023) Compartmentalized Cell-Free Expression Systems for Building Synthetic Cells. In *Advances in Biochemical Engineering/Biotechnology*. https://doi.org/10.1007/10_2023_221
24. Maffeis,V., Heuberger,L., Nikoletić,A., Schoenenberger,C.A. and Palivan,C.G. (2024) Synthetic Cells Revisited: Artificial Cells Construction Using Polymeric Building Blocks. *Adv. Sci.*, **11**, 1–30. https://doi.org/10.1002/advs.202305837
25. Abil,Z. and Danelon,C. (2020) Roadmap to Building a Cell: An Evolutionary Approach. *Front. Bioeng. Biotechnol.*, **8**, 1–8. https://doi.org/10.3389/fbioe.2020.00927
26. Forster,A.C. and Church,G.M. (2006) Towards synthesis of a minimal cell. *Mol. Syst. Biol.*, **2**, 45. https://doi.org/10.1038/msb4100090

27. Koonin,E. V. (2003) Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.*, **1**, 127–136. https://doi.org/10.1038/nrmicro751
28. Martínez-García,E. and de Lorenzo,V. (2016) The quest for the minimal bacterial genome. *Curr. Opin. Biotechnol.*, **42**, 216–224. https://doi.org/10.1016/j.copbio.2016.09.001
29. Mushegian,A.R. and Koonin,E. V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *PNAS*, **93**, 10268–10273. https://doi.org/10.1073/pnas.93.19.10268
30. Itaya,M. (1995) An estimation of minimal genome size required for life. *FEBS Lett.*, **362**, 257–260. https://doi.org/10.1016/0014-5793(95)00233-Y
31. Kobayashi,K., Ehrlich,S.D., Albertini,A., Amati,G., Andersen,K.K., Arnaud,M., Asai,K., Ashikaga,S., Aymerich,S., Bessieres,P., et al. (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 4678–4683. https://doi.org/10.1073/pnas.0730515100
32. Hutchison,C.A., Peterson,S.N., Gill,S.R., Cline,R.T., White,O., Fraser,C.M., Smith,H.O. and Venter,J.C. (1999) Global transposon mutagenesis and a minimal mycoplasma genome. *Science*, **286**, 2165–2169. https://doi.org/10.1126/science.286.5447.2165
33. Fraser,C.M., Gocayne,J.D., White,O., Adams,M.D., Clayton,R.A., Fleischmann,R.D., Bult,C.J., Kerlavage,A.R., Sutton,G., Kelley,J.M., et al. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science*, **270**, 397–403. https://doi.org/10.1126/science.270.5235.397
34. Glass,J.I., Assad-Garcia,N., Alperovich,N., Yooseph,S., Lewis,M.R., Maruf,M., Hutchison,C.A., Smith,H.O. and Venter,J.C. (2006) Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 425–430. https://doi.org/10.1073/pnas.0510013103
35. Xu,X., Meier,F., Blount,B.A., Pretorius,I.S., Ellis,T., Paulsen,I.T. and Williams,T.C. (2023) Trimming the genomic fat: minimising and re-functionalising genomes using synthetic biology. *Nat. Commun.*, **14**, 1984. https://doi.org/10.1038/s41467-023-37748-7
36. Kurasawa,H., Ohno,T., Arai,R. and Aizawa,Y. (2020) A guideline and challenges toward the minimization of bacterial and eukaryotic genomes. *Curr. Opin. Syst. Biol.*, **24**, 127–134. https://doi.org/10.1016/j.coisb.2020.10.012
37. Kumar,N., Samant,S., Singh,K. and Reshamwala,S.M.S. (2023) Minimal Cells and Genome Minimization: Top-Down and Bottom-Up Approaches to Construct Synthetic Cells. In Singh,V., Show,P.L. (eds), *Biomanufacturing for Sustainable Production of Biomolecules*.pp. 17–44. https://doi.org/10.1007/978-981-19-7911-8_2
38. Hutchison,C.A., Chuang,R.Y., Noskov,V.N., Assad-Garcia,N., Deerinck,T.J., Ellisman,M.H., Gill,J., Kannan,K., Karas,B.J., Ma,L., et al. (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253. https://doi.org/10.1126/science.aad6253
39. Gibson,D.G., Glass,J.I., Lartigue,C., Noskov,V.N., Chuang,R.Y., Algire,M.A., Benders,G.A., Montague,M.G., Ma,L., Moodie,M.M., et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329**, 52–56. https://doi.org/10.1126/science.1190719
40. Breuer,M., Earnest,T.M., Merryman,C., Wise,K.S., Sun,L., Lynott,M.R., Hutchison,C.A., Smith,H.O., Lapek,J.D., Gonzalez,D.J., et al. (2019) Essential metabolism for a minimal cell. *Elife*, **8**, 1–75. https://doi.org/10.7554/ELIFE.36842
41. Bianchi,D.M., Pelletier,J.F., Hutchison,C.A., Glass,J.I. and Luthey-Schulten,Z. (2022) Toward the Complete Functional Characterization of a Minimal Bacterial Proteome. *J. Phys. Chem. B*, **126**, 6820–6834. https://doi.org/10.1021/acs.jpcb.2c04188
42. Tomita,M., Hashimoto,K., Takahashi,K., Shimizu,T.S., Matsuzaki,Y., Miyoshi,F., Saito,K., Tanida,S., Yugi,K., Venter,J.C., et al. (1999) E-CELL: Software environment for whole-cell simulation. *Bioinformatics*, **15**, 72–84. https://doi.org/10.1093/bioinformatics/15.1.72
43. Luisi,P.L., Oberholzer,T. and Lazcano,A. (2002) The notion of a DNA minimal cell: A general discourse and some guidelines for an experimental approach. *Helv. Chim. Acta*, **85**, 1759–1777. https://doi.org/10.1002/1522-2675(200206)85:6<1759::AID-HLCA1759>3.0.CO;2-7
44. Stillwell,W. and Rau,A. (1981) Primordial transport of sugars and amino acids via Schiff bases. *Orig. Life*, **11**, 243–254. https://doi.org/10.1007/BF00931390
45. Fernández,P., Amice,R., Bruy,D., Christenhusz,M.J.M., Leitch,I.J., Leitch,A.L., Pokorny,L., Hidalgo,O. and Pellicer,J. (2024) A 160 Gbp fork fern genome shatters size record for eukaryotes. *iScience*, **27**, 109889. https://doi.org/10.1016/j.isci.2024.109889
46. Tully,J.G., Taylor Robinson,D. and Rose,D.L. (1983) *Mycoplasma genitalium*, a new species from the human urogenital tract. *Int. J. Syst. Bacteriol.*, **33**, 387–396. https://doi.org/10.1099/00207713-33-2-387
47. Zhao,X., Schwartz,C.L., Pierson,J., Giovannoni,S.J., McIntosh,J.R. and Nicastro,D. (2017) Three-dimensional structure of the ultraoligotrophic marine bacterium '*Candidatus* Pelagibacter ubique'. *Appl. Environ. Microbiol.*, **83**, e02807-16. https://doi.org/10.1128/AEM.02807-16

48. Conti,S.F. and Gettner,M.E. (1962) Electron Microscopy of Cellular Division in *Escherichia coli*. *J. Bacteriol.*, **83**, 544–550. https://doi.org/10.1128/jb.83.3.544-550.1962
49. Gil,R., Silva,F.J., Peretó,J. and Moya,A. (2004) Determination of the Core of a Minimal Bacterial Gene Set. *Microbiol. Mol. Biol. Rev.*, **68**, 518–537. https://doi.org/10.1128/mmbr.68.3.518-537.2004
50. Gil,R. (2014) The Minimal Gene-Set Machinery. In Meyers,R.A. (ed), *Encyclopedia of Molecular Cell Biology and Molecular Medicine: Synthetic Biology*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 1–36. https://doi.org/10.1002/3527600906.mcb.20130079
51. Ye,Y.N., Ma,B.G., Dong,C., Zhang,H., Chen,L.L. and Guo,F.B. (2016) A novel proposal of a simplified bacterial gene set and the neo-construction of a general minimized metabolic network. *Sci. Rep.*, **6**, 1–10. https://doi.org/10.1038/srep35082
52. Rees-Garbutt,J., Rightmyer,J., Chalkley,O., Marucci,L. and Grierson,C. (2021) Testing Theoretical Minimal Genomes Using Whole-Cell Models. *ACS Synth. Biol.*, **10**, 1598–1604. https://doi.org/10.1021/acssynbio.0c00515
53. Hunt,A.C., Rasor,B.J., Seki,K., Ekas,H.M., Warfel,K.F., Karim,A.S. and Jewett,M.C. (2025) Cell-Free Gene Expression: Methods and Applications. *Chem. Rev.*, **125**, 91–149. https://doi.org/10.1021/acs.chemrev.4c00116
54. Garenne,D., Haines,M.C., Romantseva,E.F., Freemont,P., Strychalski,E.A. and Noireaux,V. (2021) Cell-free gene expression. *Nat. Rev. Methods Prim.*, **1**, 49. https://doi.org/10.1038/s43586-021-00046-x
55. Sakai,A., Jonker,A.J., Nelissen,F.H.T., Kalb,E.M., van Sluijs,B., Heus,H.A., Adamala,K.P., Glass,J.I. and Huck,W.T.S. (2023) Cell-Free Expression System Derived from a Near-Minimal Synthetic Bacterium. *ACS Synth. Biol.*, **12**, 1616–1623. https://doi.org/10.1021/acssynbio.3c00114
56. 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**, 751–5. https://doi.org/10.1038/90802
57. Shimizu,Y., Kanamori,T. and Ueda,T. (2005) Protein synthesis by pure translation systems. *Methods*, **36**, 299–304. https://doi.org/10.1016/j.ymeth.2005.04.006
58. Lavickova,B. and Maerkl,S.J. (2019) A Simple, Robust, and Low-Cost Method to Produce the PURE Cell-Free System. *ACS Synth. Biol.*, **8**, 455–462. https://doi.org/10.1021/acssynbio.8b00427
59. De Capitani,J. and Mutschler,H. (2023) The Long Road to a Synthetic Self-Replicating Central Dogma. *Biochemistry*, **62**, 1221–1232. https://doi.org/10.1021/acs.biochem.3c00023
60. Li,J., Zhang,C., Huang,P., Kuru,E., Forster-Benson,E.T.C., Li,T. and Church,G.M. (2017) Dissecting limiting factors of the Protein synthesis Using Recombinant Elements (PURE) system. *Translation*, **5**, e1327006. https://doi.org/10.1080/21690731.2017.1327006
61. 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**, 1–12. https://doi.org/10.1038/s41598-020-80827-8
62. Jewett,M.C. and Forster,A.C. (2010) Update on designing and building minimal cells. *Curr. Opin. Biotechnol.*, **21**, 697–703. https://doi.org/10.1016/j.copbio.2010.06.008
63. Wei,E. and Endy,D. (2021) Experimental tests of functional molecular regeneration via a standard framework for coordinating synthetic cell building. *bioRxiv*. https://doi.org/10.1101/2021.03.03.433818
64. Ganesh,R.B. and Maerkl,S.J. (2024) Towards Self-regeneration: Exploring the Limits of Protein Synthesis in the Protein Synthesis Using Recombinant Elements (PURE) Cell-free Transcription-Translation System. *ACS Synth. Biol.*, **13**, 2555–2566. https://doi.org/10.1021/acssynbio.4c00304
65. Slootbeek,A.D., van Haren,M.H.I., Smokers,I.B.A. and Spruijt,E. (2022) Growth, replication and division enable evolution of coacervate protocells. *Chem. Commun.*, **58**, 11183–11200. https://doi.org/10.1039/d2cc03541c
66. 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**, 1–12. https://doi.org/10.1038/s41467-018-03926-1
67. Abil,Z., María,A., Sierra,R., Stan,A.R., Châne,A. and Prado,A. (2024) Darwinian Evolution of Self-Replicating DNA in a Synthetic Protocell. *Nat. Commun.*, **15**, 9091. https://doi.org/10.1038/s41467-024-53226-0
68. Scott,A., Noga,M.J., De Graaf,P., Westerlaken,I., Yildirim,E. and Danelon,C. (2016) Cell-free phospholipid biosynthesis by gene-encoded enzymes reconstituted in liposomes. *PLoS One*, **11**, 1–23. https://doi.org/10.1371/journal.pone.0163058
69. Blanken,D., Foschepoth,D., Serrão,A.C. and Danelon,C. (2020) Genetically controlled membrane synthesis in liposomes. *Nat. Commun.*, **11**, 1–13. https://doi.org/10.1038/s41467-020-17863-5
70. Kattan,J., Doerr,A., Dogterom,M. and Danelon,C. (2021) Shaping liposomes by cell-free expressed

bacterial microtubules. *ACS Synth. Biol.*, **10**, 2447–2455. https://doi.org/10.1021/acssynbio.1c00278
71. 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**, 1–11. https://doi.org/10.1038/s42003-020-01258-9
72. 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**, 4969. https://doi.org/10.1038/s41467-019-12932-w
73. Libicher,K. and Mutschler,H. (2020) Probing self-regeneration of essential protein factors required for *in vitro* translation activity by serial transfer. *Chem. Commun.*, **56**, 15426–15429. https://doi.org/10.1039/d0cc06515c
74. Lavickova,B., Laohakunakorn,N. and Maerkl,S.J. (2020) A partially self-regenerating synthetic cell. *Nat. Commun.*, **11**, 1–11. https://doi.org/10.1038/s41467-020-20180-6
75. Aoyama,R., Masuda,K., Shimojo,M., Kanamori,T., Ueda,T. and Shimizu,Y. (2022) *In vitro* reconstitution of the *Escherichia coli* 70S ribosome with a full set of recombinant ribosomal proteins. *J. Biochem.*, **171**, 227–237. https://doi.org/10.1093/jb/mvab121
76. Tamaru,D., Amikura,K., Shimizu,Y., Nierhaus,K.H. and Ueda,T. (2018) Reconstitution of 30S ribosomal subunits in vitro using ribosome biogenesis factors. *RNA*, **24**, 1512–1519. https://doi.org/10.1261/rna.065615.118
77. Kosaka,Y., Miyawaki,Y., Mori,M., Aburaya,S., Fukuyama,M., Ueda,M. and Aoki,W. (2022) Reconstitution of ribosome self-replication outside a living cell. *bioRxiv*. https://doi.org/10.1101/2022.08.29.505692
78. Li,J., Haas,W., Jackson,K., Kuru,E., Jewett,M.C., Fan,Z.H., Gygi,S. and Church,G.M. (2017) Cogenerating Synthetic Parts toward a Self-Replicating System. *ACS Synth. Biol.*, **6**, 1327–1336. https://doi.org/10.1021/acssynbio.6b00342
79. Jewett,M.C., Fritz,B.R., Timmerman,L.E. and Church,G.M. (2013) *In vitro* integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol. Syst. Biol.*, **9**, 1–8. https://doi.org/10.1038/msb.2013.31
80. Cui,Z., Stein,V., Tnimov,Z., Mureev,S. and Alexandrov,K. (2015) Semisynthetic tRNA Complement Mediates *in Vitro* Protein Synthesis. *J. Am. Chem. Soc.*, **137**, 4404–4413. https://doi.org/10.1021/ja5131963
81. Libicher,K., Hornberger,R., Heymann,M. and Mutschler,H. (2020) In vitro self-replication and multicistronic expression of large synthetic genomes. *Nat. Commun.*, **11**, 904. https://doi.org/10.1038/s41467-020-14694-2
82. Blanken,D.M. (2021) Genetically encoded phospholipid production for autonomous synthetic cell proliferation. https://doi.org/10.4233/uuid:363f4643-0a68-4726-9945-e8daf6e0350c
83. Kohyama,S., Merino-Salomon,A. and Schwille,P. (2022) In vitro assembly, positioning and contraction of a division ring in minimal cells. *Nat. Commun.*, **13**, 6098. https://doi.org/10.1038/s41467-022-33679-x
84. Su'etsugu,M., Takada,H., Katayama,T. and Tsujimoto,H. (2017) Exponential propagation of large circular DNA by reconstitution of a chromosome-replication cycle. *Nucleic Acids Res.*, **45**, 11525–11534. https://doi.org/10.1093/nar/gkx822
85. Kaguni,J.M. and Kornberg,A. (1984) Replication initiated at the origin (oriC) of the *E. coli* chromosome reconstituted with purified enzymes. *Cell*, **38**, 183–190. https://doi.org/10.1016/0092-8674(84)90539-7
86. 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. https://doi.org/10.1021/acssynbio.0c00584
87. 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. https://doi.org/10.1021/acssynbio.1c00430
88. 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. *Sci. Rep.*, **8**, 13089. https://doi.org/10.1038/s41598-018-31585-1
89. Weise,L.I., Heymann,M., Mayr,V. and Mutschler,H. (2019) Cell-free expression of RNA encoded genes using MS2 replicase. *Nucleic Acids Res.*, **47**, 10956–10967. https://doi.org/10.1093/nar/gkz817
90. Niederholtmeyer,H., Stepanoèa,È. and Maerkl,S.J. (2013) Implementation of cell-free biological networks at steady state. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 15985–15990. https://doi.org/10.1073/pnas.1311166110
91. Asahara,H. and Chong,S. (2010) *In vitro* genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme. *Nucleic Acids Res.*, **38**, 1–10.

https://doi.org/10.1093/nar/gkq377
92. 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. https://doi.org/10.1021/acssynbio.6b00017
93. Hibi,K., Amikura,K., Sugiura,N., Masuda,K., Ohno,S., Yokogawa,T., Ueda,T. and Shimizu,Y. (2020) Reconstituted cell-free protein synthesis using in vitro transcribed tRNAs. *Commun. Biol.*, **3**, 1–11. https://doi.org/10.1038/s42003-020-1074-2
94. Pols,T., Sikkema,H.R., Gaastra,B.F., Frallicciardi,J., Śmigiel,W.M., Singh,S. and Poolman,B. (2019) A synthetic metabolic network for physicochemical homeostasis. *Nat. Commun.*, **10**, 4239. https://doi.org/10.1038/s41467-019-12287-2
95. Yadav,S., Perkins,A.J.P., Liyanagedera,S.B.W., Bougas,A. and Laohakunakorn,N. (2025) ATP Regeneration from Pyruvate in the PURE System. *ACS Synth. Biol.*, **14**, 247–256. https://doi.org/10.1101/2024.09.06.611674
96. Sugii,S., Hagino,K., Mizuuchi,R. and Ichihashi,N. (2023) Cell-free expression of RuBisCO for ATP production in the synthetic cells. *Synth. Biol.*, **8**. https://doi.org/10.1093/synbio/ysad016
97. Berhanu,S., Ueda,T. and Kuruma,Y. (2019) Artificial photosynthetic cell producing energy for protein synthesis. *Nat. Commun.*, **10**. https://doi.org/10.1038/s41467-019-09147-4
98. Matthies,D., Haberstock,S., Joos,F., Dötsch,V., Vonck,J., Bernhard,F. and Meier,T. (2011) Cell-free expression and assembly of ATP synthase. *J. Mol. Biol.*, **413**, 593–603. https://doi.org/10.1016/j.jmb.2011.08.055
99. Ivanov,I., Lira,R.B., Tang,T.Y.D., Franzmann,T., Klosin,A., da Silva,L.C., Hyman,A., Landfester,K., Lipowsky,R., Sundmacher,K., et al. (2019) Directed Growth of Biomimetic Microcompartments. *Adv. Biosyst.*, **3**, 1–9. https://doi.org/10.1002/adbi.201800314
100. Bhattacharya,A., Brea,R.J., Niederholtmeyer,H. and Devaraj,N.K. (2019) A minimal biochemical route towards de novo formation of synthetic phospholipid membranes. *Nat. Commun.*, **10**. https://doi.org/10.1038/s41467-018-08174-x
101. De Franceschi,N., Barth,R., Meindlhumer,S., Fragasso,A. and Dekker,C. (2024) Dynamin A as a one-component division machinery for synthetic cells. *Nat. Nanotechnol.*, **19**, 70–76. https://doi.org/10.1038/s41565-023-01510-3
102. Härtel,T. and Schwille,P. (2014) ESCRT-III mediated cell division in *Sulfolobus acidocaldarius* – A reconstitution perspective. *Front. Microbiol.*, **5**, 257. https://doi.org/10.3389/fmicb.2014.00257
103. Kattan,J. (2020) Biosynthesis of protein filaments for the creation of a minimal cell. https://doi.org/10.4233/uuid:724f70a9-efd9-4598-9a24-064aa77db509
104. Blanch Jover,A. and Dekker,C. (2023) The archaeal Cdv cell division system. *Trends Microbiol.*, **31**, 601–615. https://doi.org/10.1016/j.tim.2022.12.006
105. Reverte-López,M., Kanwa,N., Qutbuddin,Y., Jasnin,M. and Schwille,P. (2024) Self-organized spatial targeting of contractile actomyosin rings for synthetic cell division. *Nat. Commun.*, **15**, 10415. https://doi.org/10.1038/s41467-024-54807-9
106. Litschel,T., Kelley,C.F., Holz,D., Adeli Koudehi,M., Vogel,S.K., Burbaum,L., Mizuno,N., Vavylonis,D. and Schwille,P. (2021) Reconstitution of contractile actomyosin rings in vesicles. *Nat. Commun.*, **12**, 1–10. https://doi.org/10.1038/s41467-021-22422-7
107. Baldauf,L., van Buren,L., Fanalista,F. and Koenderink,G.H. (2022) Actomyosin-Driven Division of a Synthetic Cell. *ACS Synth. Biol.*, **11**, 3120–3133. https://doi.org/10.1021/acssynbio.2c00287
108. Nourian,Z., Scott,A. and Danelon,C. (2014) Toward the assembly of a minimal divisome. *Syst. Synth. Biol.*, **8**, 237–247. https://doi.org/10.1007/s11693-014-9150-x
109. Blanken,D., van Nies,P. and Danelon,C. (2019) Quantitative imaging of gene-expressing liposomes reveals rare favorable phenotypes. *Phys. Biol.*, **16**, 045002. https://doi.org/10.1088/1478-3975/ab0c62
110. 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. https://doi.org/10.1073/pnas.0408236101
111. Glass,J.I., Merryman,C., Wise,K.S., Hutchison III,C.A. and Smith,H.O. (2017) Minimal Cells — Real and Imagined. *Cold Spring Harb. Perspect. Biol.* https://doi.org/10.1101/cshperspect.a023861
112. Restrepo Sierra,A.M. (2024) Engineering Synthetic Cells through Module Integration and Evolution. https://doi.org/10.4233/uuid:2f2dfc76-5e29-4a04-84c1-3d337e3bf645
113. Kuruma,Y. and Ueda,T. (2015) The PURE system for the cell-free synthesis of membrane proteins. *Nat. Protoc.*, **10**, 1328–1344. https://doi.org/10.1038/nprot.2015.082
114. Chizzolini,F., Forlin,M., Cecchi,D. and Mansy,S.S. (2014) Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength. *ACS Synth. Biol.*, **3**,

363–371. https://doi.org/10.1021/sb4000977
115. Du,L., Villarreal,S. and Forster,A.C. (2012) Multigene expression in vivo: Supremacy of large versus small terminators for T7 RNA polymerase. *Biotechnol. Bioeng.*, **109**, 1043–1050. https://doi.org/10.1002/bit.24379
116. Komura,R., Aoki,W., Motone,K., Satomura,A. and Ueda,M. (2018) High-throughput evaluation of T7 promoter variants using biased randomization and DNA barcoding. *PLoS One*, **13**, 1–16. https://doi.org/10.1371/journal.pone.0196905
117. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950. https://doi.org/10.1038/nbt.1568
118. Eisenstein,M. (2020) How to build a genome. *Nature*. https://doi.org/10.1038/d41586-020-00511-9
119. Kok,S. De, Stanton,L.H., Slaby,T., Durot,M., Holmes,V.F., Patel,K.G., Platt,D., Shapland,E.B., Serber,Z., Dean,J., et al. (2014) Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth. Biol.*, **3**, 97–106. https://doi.org/10.1021/sb4001992
120. Pryor,J.M., Potapov,V., Bilotti,K., Pokhrel,N. and Lohman,G.J.S. (2022) Rapid 40 kb Genome Construction from 52 Parts through Data-optimized Assembly Design. *ACS Synth. Biol.*, **11**, 2036–2042. https://doi.org/10.1021/acssynbio.1c00525
121. Sikkema,A.P., Tabatabaei,S.K., Lee,Y.-J., Lund,S. and Lohman,G.J.S. (2023) High-Complexity One-Pot Golden Gate Assembly. *Curr. Protoc.*, **3**, e882. https://doi.org/10.1002/cpz1.882
122. Shepherd,T.R., Du,L., Liljeruhm,J., Samudyata, Wang,J., Sjödin,M.O.D., Wetterhall,M., Yomo,T. and Forster,A.C. (2017) *De novo* design and synthesis of a 30-cistron translation-factor module. *Nucleic Acids Res.*, **45**, 10895–10905. https://doi.org/10.1093/nar/gkx753
123. Hillson,N.J. (2011) DNA Assembly Method Standardization for Synthetic Biomolecular Circuits and Systems. In Koeppl,H., Setti,G., di Bernardo,M., Densmore,D. (eds), *Design and Analysis of Biomolecular Circuits*. Springer New York, NY, pp. 295–314. https://doi.org/10.1007/978-1-4419-6766-4
124. Warren,R.L., Freeman,J.D., Levesque,R.C., Smailus,D.E., Flibotte,S. and Holt,R.A. (2008) Transcription of foreign DNA in *Escherichia coli*. *Genome Res.*, **18**, 1798–1805. https://doi.org/10.1101/gr.080358.108
125. Gibson,D.G., Benders,G.A., Axelrod,K.C., Zaveri,J., Algire,M.A., Moodie,M., Montague,M.G., Venter,J.C., Smith,H.O. and Hutchison,C.A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 20404–20409. https://doi.org/10.1073/pnas.0811011106
126. Postma,E.D., Dashko,S., van Breemen,L., Taylor Parkins,S.K., van den Broek,M., Daran,J.-M. and Daran-Lapujade,P. (2021) A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **49**, 1769–1783. https://doi.org/10.1093/nar/gkaa1167
127. Fredens,J., Wang,K., de la Torre,D., Funke,L.F.H., Robertson,W.E., Christova,Y., Chia,T., Schmied,W.H., Dunkelmann,D.L., Beranek,V., et al. (2019) Total synthesis of *Escherichia coli* with a recoded genome. *Nature*, **569**, 514–518. https://doi.org/10.1038/s41586-019-1192-5

Chapter 2

Synthetic Genomics from a yeast perspective

Charlotte C. Koster*, Eline D. Postma*, Ewout Knibbe*, Céline Cleij*, Pascale Daran-Lapujade

* These authors have contributed equally to this work and share first authorship.

Synthetic Genomics focuses on the construction of rationally designed chromosomes and genomes and offers novel approaches to study biology and to construct synthetic cell factories. Currently, progress in Synthetic Genomics is hindered by the inability to synthesize DNA molecules longer than a few hundred base pairs, while the size of the smallest genome of a self-replicating cell is several hundred thousand base pairs. Methods to assemble small fragments of DNA into large molecules are therefore required. Remarkably powerful at assembling DNA molecules, the unicellular eukaryote *Saccharomyces cerevisiae* has been pivotal in the establishment of Synthetic Genomics. Instrumental in the assembly of entire genomes of various organisms in the past decade, the *S. cerevisiae* genome foundry has a key role to play in future Synthetic Genomics developments.

Adapted from the publication in *Frontiers in Bioengineering and Biotechnology* (2022), **10**, 869486. https://doi.org/10.3389/fbioe.2022.869486

Introduction

Synthetic Genomics (SG) is a recent Synthetic Biology discipline that focuses on the construction of rationally designed chromosomes and genomes. SG offers a novel approach to address fundamental biological questions by restructuring, recoding, and minimizing (parts of) genomes (as recently reviewed by [1]). SG is now spurring technological developments in academia and has a strong future potential in industry [2, 3]). Humankind's best microbial friend, the baker's yeast *Saccharomyces cerevisiae*, has played, and continues to play a key role in SG advances, both by enabling the construction of chromosomes for other hosts, and in the refactoring of its own genome. This mini review explores the reasons for this strategic positioning of *S. cerevisiae* in SG, surveys the main achievements enabled by this yeast and reflects on future developments.

Current limitations of genome assembly

While small-sized viral chromosomes were the first to be chemically synthetized, the breakthrough in the field of SG came with the synthesis and assembly of the 592 kilobase (kb) chromosome of *Mycoplasma genitalium* [4, 5]. The unicellular eukaryote *Saccharomyces cerevisiae* has made a key contribution to this famous milestone. To understand how this microbe, commonly used in food and beverages, contributes to the assembly of synthetic genomes, let us recapitulate how synthetic chromosomes can be constructed (Figure 2.1).

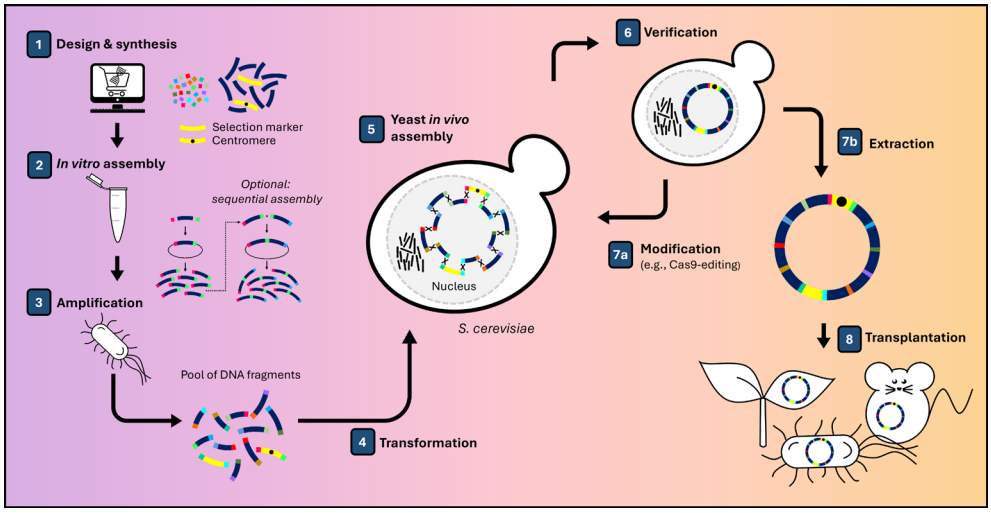


Figure 2.1: Simplified overview of chromosome construction using *Saccharomyces cerevisiae* for genome assembly and production.

It starts with the customized synthesis of short DNA molecules called oligonucleotides. Oligonucleotides are mostly synthetized via phosphoramidite chemistry, a 40 year-old method [6] that, despite decades of technological developments, struggles to deliver error-free oligonucleotides longer than 200 base pairs (bp). While the implementation of microarrays has substantially decreased the synthesis cost, it has not increased

oligo length, an achievement that requires new synthesis methods [7]. Enzymatic alternatives for DNA synthesis are under development [8, 9], but still have considerable shortcomings regarding automation and scalability that must be overcome before commercial scale can be considered (reviewed in [10–13]). Considering that a theoretical minimal genome would be around 113 kb long [14] and that the first fully synthesized genome of *M. genitalium* contains 583 kb [5], thousands of oligos must be stitched together to construct a complete synthetic genome. These DNA oligos can be assembled into longer DNA fragments owing to a plethora of *in vitro* methods (reviewed in [11, 15, 16]). A method that has gained tremendous popularity since its development is the homology-based Gibson isothermal assembly [17], devised to assemble the *M. genitalium* genome. As all *in vitro* methods, Gibson assembly is limited by the number of fragments that can reliably be stitched together in one reaction, usually around a dozen, requiring a stepwise assembly procedure of increasingly large genomic DNA constructs [18]. DNA must be recovered from the reaction, amplified and verified in each round, to allow further processing. Selection and amplification of correctly cloned DNA is routinely performed in *Escherichia coli,* however, maintenance of large constructs of exogenous DNA, especially from prokaryotic origins, in this bacterium is often limited by expression and toxicity of gene products [19]. *In vitro* alternatives for efficient and faithful selection and amplification of correctly assembled DNA are under development, but these are currently limited in length of amplified DNA and scalability [20–23]. While in principle stepwise *in vitro* assembly can lead to a DNA molecule of any size, and selection and amplification in *E. coli* worked well for DNA constructs up to 72 kb, *E. coli* had great difficulties maintaining quarter *M. genitalium* genomes, causing Gibson and colleagues to turn to baker's yeast [4, 5].

Saccharomyces cerevisiae as a genome foundry

S. cerevisiae seems a logical host for SG as it naturally maintains a 12 Mb genome consisting of 16 chromosomes ranging from 230 to 1500 kb in its haploid version, lives as polyploid in natural environments, and is extremely robust to changes in genome content and architecture [24]. The extreme robustness of *S. cerevisiae* to supernumerary, chimeric chromosomes, a key feature for SG, was already demonstrated in the late '80s [25, 26]. A second key feature of *S. cerevisiae* is its preference for homologous recombination (HR) to repair double-strand DNA breaks [27], a rare trait among eukaryotes. The ability of *S. cerevisiae* to efficiently and with high fidelity stitch together linear DNA molecules that present homologous regions as short as 40 bp [28] at their ends, was rapidly valorized for genetic manipulations and assembly of heterologous DNA. Renamed to *in vivo* assembly [29], this cloning technique (Figure 2.1) contributes to the remarkable genetic tractability and popularity of *S. cerevisiae* as model and industrial microbe. The combination of *S. cerevisiae*'s HR efficiency and fidelity, chromosome maintenance and propagation enabled the construction of the full *Mycoplasma* genome. Reflecting that "in the future, it may be advantageous to make greater use of yeast recombination to assemble chromosomes", this study propelled *S. cerevisiae* as powerful 'genome foundry' [5]. In the challenge to synthesize genomes, Ostrov and colleagues rightfully identified assembly of these long DNA constructs as 'the most critical hurdle' [10]. To date, *S. cerevisiae* has been key to assembling entire or partial genomes in most synthetic genome projects (Table 2.1).

Table 2.1: Overview of the contribution of *S. cerevisiae* in synthetic genomics by the assembly of large (>100 kb) DNA constructs

	Donor DNA	Number of fragments[1]	Approximate size of fragments[1,2]	Approximate size of final construct	Aim of yeast assembly	Reference
Viruses	Herpes simplex type 1	11	14 kb	152 kb	Assembly and modification of a viral genome, transfection and reconstitution in mammalian cells.	[30]
	Autographa californica nucleopolyhedrovirus	4	45 kb	145 kb	Assembly and modification of a viral genome, transfection and reconstitution in insect cells.	[31]
	Cytomegalovirus isolate Toledo	3	116 kb	230 kb	Assembly and modification of a viral genome, transfection and reconstitution in mammalian cells.	[32]
Prokaryotes	*Mycoplasma genitalium*	6	Up to 144 kb	592 kb	Assembly of a synthetic *M. genitalium* genome which could not be stably maintained in *E. coli*.	[5]
	Mycoplasma genitalium	25	17–35 kb	592 kb	Assembly of a synthetic *M. genitalium* genome from short fragments, exploring assembly capacity in yeast.	[4]
	Mycoplasma mycoides	11	100 kb	1 Mb	Assembly of a synthetic *M. mycoides* genome, transplantation into a recipient cell to create the first bacterial cell controlled by a synthesized genome.	[33]
	Mycoplasma pneumonia	2	10–816 kb	826 kb	Insertion of yeast regulatory elements in the full *M. pneumonia* genome to allow for cloning and engineering of the genome.	[34, 35]
	Mycoplasma hominis	2	5–665 kb	670 kb	Insertion of yeast regulatory elements in the full *M. hominis* genome to allow for cloning and engineering of the genome.	[36]
	Acholeplasma laidlawii	3[3]	121–897 kb	1.38 Mb	Exploring potential toxicity when assembling bacterial genomes in yeast.	[37]
	Escherichia coli	3	185–660 kb	1.03 Mb	Assembly of a minimal *E. coli* genome by Cas9-induced recombination of partial genomes.	[38]
	Escherichia coli	7–14	6–13 kb	100 kb	Assembly of recoded *E. coli* partial genomes, used to replace the *E. coli* genome by a recoded synthetic genome.	[39]
	Caulobacter crescentus	16	38–65 kb	785 kb	Assembly of a minimized and synthetic *C. crescentus* genome, recoded to be compatible with chemical DNA synthesis and transplanted in a recipient cell.	[40]
	Prechlorococcus marinus	2	580–675 kb	1.66 Mb	Exploring assembly capacity and DNA stability of exogenous genomes in yeast.	[41]
	Synechococcus elongatus	4	100–200 kb	454 kb	Exploring the ability to clone genomes with high G/C-content in yeast.	[42]

Algae	*Phaeodactylum tricornutum*	5	106–128 kb	497 kb	Assembly of DNA with a moderate G + C content as a case study for assembly and modification of eukaryotic chromosomes in yeast.	[43]
	Chlamydomonas reinhardtii chloroplast genome	6	34–129 kb	230 kb	Assembly of a partial *C. reinhardtii* chloroplast genome to create genetic diversity at multiple loci at once.	[44]
Yeasts	Yeast chromosome XII	33[4]	26–39 kb	976 kb	Assembly of a megabase synthetic yeast chromosome harboring the highly repetitive ribosomal DNA locus.	[45]
	Single-chromosome yeast	15[4]	230–1500 kb	11 Mb	Assembly of all sixteen *S. cerevisiae* chromosomes into a single chromosome.	[24]
	Yeast neochromosome	44	2.5 kb	100 kb	Assembly of a circular supernumerary *S. cerevisiae* neochromosome that can act as a platform for modular genome engineering.	[46]
	Yeast neochromosome for pathway engineering	43	2.5–5 kb	100 kb	Assembly of circular and linear supernumerary *S. cerevisiae* neochromosomes for expression of heterologous and essential metabolic pathways.	[47]
Other	Human *HPRT1* gene	13	3–83 kb	125 kb	Assembly of a synthetic human *HRPT1* gene and transplantation and expression in mammalian cells.	[48]
	Artificial data storage chromosome	5	40 kb	254 kb	Assembly of a *S. cerevisiae* artificial chromosome containing data-encoded DNA for digital data storage.	[49]

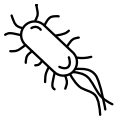
[1] In case of a sequential assembly, the fragment number and size of the last assembly is used.
[2] Short backbones containing regulatory elements such as CEN/ARS and markers not included.
[3] Initial assembly of the entire genome failed due to gene toxicity.
[4] Assembly was performed by stepwise integration in multiple rounds.

For instance, the entire 785 kb refactored *Caulobacter crescentus* (renamed *C. ethensis*) genome was assembled *in vivo* from 16 fragments [40], while the recoded *E. coli* genome was split over 10 fragments of 91 to 136 kb individually assembled in yeast, and then sequentially integrated in the *E. coli* chromosome to replace native segments [39] (Table 2.1). *In vivo* assembly also proved to be powerful in assembling and modifying genomes of organisms that are poorly amenable to genome editing; the rapid and faithful HR-based assembly of *S. cerevisiae* recently enabled the reconstruction of a synthetic SARS-CoV-2 genome in a single week [50], and has been shown to be a promising host for *in vivo* assembly and modification of other viral genomes [51] as well as the genomes of various pathogens [34] and even a 101 kb human gene, which was transplanted into mouse embryonic cells [48] (Table 2.1). Moreover, *S. cerevisiae* was selected for the construction of the first synthetic eukaryotic genome. The international Sc2.0 consortium, spearheaded by Jef Boeke, undertook less than ten years ago the daunting task of synthesizing recoded versions of the 16 yeast chromosomes. Via stepwise, systematic replacement of 30 to 40 kb (using ca. 12 DNA fragments of 2 to 4 kb) of the native yeast sequence, the consortium is close to the completion of the largest synthetic genome to date [52, 53], with the ambition to reshape and minimize the *S. cerevisiae* genome [54].

While *S. cerevisiae* is not the only microbial host available for the construction of (neo) chromosomes (Table 2.2), several key features make it superior to its bacterial alternatives *Bacillus subtilis* and *E. coli* as genome foundry: (i) *S. cerevisiae* has the natural ability to carry large amounts of DNA and therefore to host multiple exogenous bacterial genomes [34]; (ii) *E. coli* frequently struggles with toxicity caused by the expression of exogenous bacterial sequences [4, 19, 55], while *S. cerevisiae* is very robust to the presence of heterologous DNA from prokaryotic or eukaryotic origin [41]; (iii) *S. cerevisiae* can, in a single transformation, assemble many DNA oligonucleotides into (partial) genomes. *B. subtilis* can also maintain large exogenous DNA constructs, but requires a stepwise method for DNA assembly, in which each DNA part is integrated sequentially into the *B. subtilis* genome [56]. This approach is intrinsically more labor-intensive and time-consuming than *S. cerevisiae* single transformation assembly.

Surprised by the genetic tractability of *S. cerevisiae*, Gibson and colleagues wondered "how many pieces can be assembled in yeast in a single step?" [5]. Pioneering a SG approach for metabolic engineering based on modular, specialized synthetic chromosomes, Postma et al. probed this limit recently in our lab by constructing 100 kb artificial linear and circular neochromosomes from 44 DNA parts in a single transformation [46, 47]. The remarkable efficiency of *in vivo* assembly (36% of assemblies faithful to design) revealed that its limit has clearly not been reached yet, and that future systematic studies are required to evaluate the true potential of *S. cerevisiae* as a genome foundry. The supernumerary chromosomes were shown to stably maintain complete heterologous pathways as well as the yeast's central carbon metabolism, underlining the potential of yeast synthetic genomics in the development of optimized cell-factories. Once assembled, synthetic chromosomes could be easily edited in *S. cerevisiae* thanks to its efficient HR and rich molecular toolbox.

Table 2.2: Strengths and weaknesses of *in vivo* and in *vitro* approaches for DNA assembly in SG

	In vitro	*E. coli*	*B. subtilis*[1]	*S. cerevisiae*
				
One-pot assembly				
Number of fragments	Up to 20 [57]	Up to 6 [58]	2 [59]	Up to 44 [46]
Construct size	Up to 144 kb [5]	Up to ~180 kb [58]	Up to 200 kb [56]	Up to 1.66 Mb [41]
Compatible with heterologous genes	***	*	***	***
DNA yield	*	***	Unknown	**
Sequential assembly with most fragments	[18][2]	-	[59]	[33][2]
Number of assembled fragments	600	-	31	1078
Number of sequential HR events	4	-	18	3
Final construct size	16.3 kb	-	134.5 kb	1.08 Mb
Overall proficiency to assemble synthetic genomes	-[3]	*	**	***

[1] Assembly of fragments in *B. subtilis* is performed by integration into the host genome.
[2] Between rounds of sequential assembly, transformation of *E. coli* with the assemblies is conventional for selection and amplification of constructs.
[3] Requires *in vivo* amplification and selection in a microbial host.

Challenges in genome assembly using yeast

While *S. cerevisiae* is natively proficient for SG, several aspects of *in vivo* assembly in yeast are still far from optimal. Firstly, compared to bacterial alternatives, *S. cerevisiae* cells grow slowly with a maximum specific growth rate around 0.4–0.5 h^{-1} and are hard to disrupt due to their sturdy cell wall. Considering that large DNA constructs above a few hundred kilobases are sensitive to shear stress, chromosome extraction and purification from *S. cerevisiae* is possible, but remains tenuous and inefficient, leading to low DNA yields and potentially damaged chromosomes [60]. Secondly, the strength of *S. cerevisiae* can become its weakness, as the HR machinery can be overzealous and recombine any (short) DNA sequence with homology within or between the (neo) chromosomes, which may lead to misassemblies [36, 44]. Lastly, non-homologous end joining and microhomology-mediated end joining, DNA repair mechanisms that assemble pieces of DNA with no or minimal homology, are present in *S. cerevisiae* with low activity [61, 62], and can also cause misassemblies. Similar to how *E. coli* was engineered to become a lab tool for DNA amplification, these shortcomings could be alleviated by engineering *S. cerevisiae* into a more powerful genome foundry.

Are there future alternatives to *S. cerevisiae*? Naturally, *B. subtilis* and *E. coli* could also be engineered. However, considering the minute fraction of the vast microbial biodiversity that has been tested for genetic accessibility and DNA assembly, it is likely that microbes yet to be discovered are even better genome foundries. Environments

causing extreme DNA damage (high radiation, toxic chemicals, etc.) might be a source of HR-proficient organisms (e.g., [63, 64]) better suited for SG.

In a more distant future, *in vitro* alternatives might replace the need for live DNA foundries altogether, thereby accelerating and simplifying genome construction. However, this will require major technological advances in *in vitro* DNA assembly and amplification. Already substantial efforts have led to the development of methods for DNA amplification, such as rolling circle amplification by the phage ϕ29 DNA polymerase [65, 66], recently implemented for the amplification of a 116 kb multipartite genome [20] and the *in vitro* amplification of synthetic genomes using the *E. coli* replisome, which already demonstrated to be capable of amplification of 1 Mb synthetic genomes [22]. Targets for improvement of these methods are the maximal length of amplified DNA fragments, the yield of amplification, the need for restriction of the amplified, concatenated molecules or the formation of non-specifically amplified products. The development of an *in vitro* approach that can parallel the *in vivo* assembly capability of *S. cerevisiae* seems even more challenging. While an interesting avenue might be to reconstitute the HR DNA machinery of *S. cerevisiae in vitro*, it presents a daunting task considering that all players and their respective role have not been fully elucidated yet [62, 67]. Still, considering that highly complex systems such as the transcription and translation machineries have been successfully implemented *in vitro* and are commercially available [68], cell-free *S. cerevisiae* HR might become a reality in the coming years.

Outlook

Since the first genome synthesis in 2008, relatively few genomes have been synthetized. Low-cost, customizable construction of designer genomes, currently accessible for small viral, organellar or bacterial constructs, is still out of reach for large (eukaryotic) genomes. There are still numerous technical, financial, and computational hurdles that must be overcome on the road to microbial designer genomes, tailored to applications in bio-based industry. Here we reviewed why the yeast *S. cerevisiae* is a key organism in the field of SG, however, the spectrum of available hosts is expected to increase as research in SG advances. For example, a recent study shows improving the HR capacity of the industrially relevant yeast *Yarrowia lipolytica* could greatly expand the potential applications of SG in bio-based processes [69].

In the near future, SG is anticipated to contribute to various fields, such as a platform technology for industrial biotechnological processes [3, 47], as a new means for data storage [49] and for the development of new cell therapies and other medical applications, which is the ambition of the Genome Project-Write [70]. In parallel, worldwide bottom-up approaches endeavor to construct synthetic cells from scratch, such as the European consortia BaSyC (http://www.basyc.nl), MaxSynBio (https://www.maxsynbio.mpg.de) and the Synthetic cell initiative (http://www.syntheticcell.eu) and the US-based Build-a-cell initiative (http://buildacell.io) (reviewed in [71]). Looking further ahead, SG may even assist in understanding and engineering entire ecosystems by assembly of a metagenomes in a single cell [72]. SG, albeit still in its infancy and mostly limited to academic research, has bright days ahead, and *S. cerevisiae* is foreseen to remain a valuable, if not indispensable, SG tool for the coming decade.

References

1. Coradini,A.L.V., Hull,C.B. and Ehrenreich,I.M. (2020) Building genomes to understand biology. *Nat. Commun.*, **11**, 1–11. https://doi.org/10.1038/s41467-020-19753-2
2. Zhang,W., Mitchell,L.A., Bader,J.S. and Boeke,J.D. (2020) Synthetic Genomes. *Annu. Rev. Biochem.*, **89**, 77–101. https://doi.org/10.1146/annurev-biochem-013118-110704
3. Schindler,D. (2020) Genetic engineering and synthetic genomics in yeast to understand life and boost biotechnology. *Bioengineering*, **7**, 1–18. https://doi.org/10.3390/bioengineering7040137
4. Gibson,D.G., Benders,G.A., Axelrod,K.C., Zaveri,J., Algire,M.A., Moodie,M., Montague,M.G., Venter,J.C., Smith,H.O. and Hutchison,C.A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 20404–20409. https://doi.org/10.1073/pnas.0811011106
5. Gibson,D.G., Benders,G.A., Andrews-Pfannkoch,C., Denisova,E.A., Baden-Tillson,H., Zaveri,J., Stockwell,T.B., Brownley,A., Thomas,D.W., Algire,M.A., et al. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, **319**, 1215–1220. https://doi.org/10.1126/science.1151721
6. Beaucage,S.L. and Caruthers,M.H. (1981) Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.*, **22**, 1859–1862. https://doi.org/10.1016/S0040-4039(01)90461-7
7. Hughes,R.A. and Ellington,A.D. (2017) Synthetic DNA synthesis and assembly: Putting the synthetic in synthetic biology. *Cold Spring Harb. Perspect. Biol.*, **9**. https://doi.org/10.1101/cshperspect.a023812
8. Lee,H., Wiegand,D.J., Griswold,K., Punthambaker,S., Chun,H., Kohman,R.E. and Church,G.M. (2020) Photon-directed multiplexed enzymatic DNA synthesis for molecular digital data storage. *Nat. Commun.*, **11**, 1–9. https://doi.org/10.1038/s41467-020-18681-5
9. Lee,H.H., Kalhor,R., Goela,N., Bolot,J. and Church,G.M. (2019) Terminator-free template-independent enzymatic DNA synthesis for digital information storage. *Nat. Commun.*, **10**, 1–12. https://doi.org/10.1038/s41467-019-10258-1
10. Ostrov,N., Beal,J., Ellis,T., Benjamin Gordon,D., Karas,B.J., Lee,H.H., Lenaghan,S.C., Schloss,J.A., Stracquadanio,G., Trefzer,A., et al. (2019) Technological challenges and milestones for writing genomes. *Science*, **366**, 310–312. https://doi.org/10.1126/science.aay0339
11. Paul,S.S., Trabelsi,H., Yaseen,Y., Basu,U., Altaii,H.A. and Dhali,D. (2021) Advances in long DNA synthesis. In Singh,V. (ed), *Microbial Cell Factories Engineering for Production of Biomolecules*. Elsevier Inc., pp. 21–36. https://doi.org/10.1016/b978-0-12-821477-0.00014-3
12. Hao,M., Qiao,J. and Qi,H. (2020) Current and emerging methods for the synthesis of single-stranded DNA. *Genes (Basel).*, **11**. https://doi.org/10.3390/genes11020116
13. Eisenstein,M. (2020) Enzymatic DNA synthesis enters new phase. *Nat. Biotechnol.*, **38**, 1113–1115. https://doi.org/10.1038/s41587-020-0695-9
14. Forster,A.C. and Church,G.M. (2006) Towards synthesis of a minimal cell. *Mol. Syst. Biol.*, **2**, 45. https://doi.org/10.1038/msb4100090
15. Chao,R., Yuan,Y. and Zhao,H. (2015) Recent advances in DNA assembly technologies. *FEMS Yeast Res.*, **15**, 1–9. https://doi.org/10.1111/1567-1364.12171
16. Casini,A., Storch,M., Baldwin,G.S. and Ellis,T. (2015) Bricks and blueprints: Methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.*, **16**, 568–576. https://doi.org/10.1038/nrm4014
17. 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. https://doi.org/10.1038/nmeth.1318
18. Gibson,D.G., Smith,H.O., Hutchison,C.A., Venter,J.C. and Merryman,C. (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods*, **7**, 901–903. https://doi.org/10.1038/nmeth.1515
19. Karas,B.J., Suzuki,Y. and Weyman,P.D. (2015) Strategies for cloning and manipulating natural and synthetic chromosomes. *Chromosom. Res.*, **23**, 57–68. https://doi.org/10.1007/s10577-014-9455-3
20. Libicher,K., Hornberger,R., Heymann,M. and Mutschler,H. (2020) In vitro self-replication and multicistronic expression of large synthetic genomes. *Nat. Commun.*, **11**, 904. https://doi.org/10.1038/s41467-020-14694-2
21. 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**, 1–12. https://doi.org/10.1038/s41467-018-03926-1
22. Mukai,T., Yoneji,T., Yamada,K., Fujita,H., Nara,S. and Su'etsugu,M. (2020) Overcoming the Challenges of Megabase-Sized Plasmid Construction in *Escherichia coli*. *ACS Synth. Biol.*, **9**, 1315–1327. https://

doi.org/10.1021/acssynbio.0c00008
23. Su'etsugu,M., Takada,H., Katayama,T. and Tsujimoto,H. (2017) Exponential propagation of large circular DNA by reconstitution of a chromosome-replication cycle. *Nucleic Acids Res.*, **45**, 11525–11534. https://doi.org/10.1093/nar/gkx822
24. Shao,Y., Lu,N., Wu,Z., Cai,C., Wang,S., Zhang,L.L., Zhou,F., Xiao,S., Liu,L., Zeng,X., et al. (2018) Creating a functional single-chromosome yeast. *Nature*, **560**, 331–335. https://doi.org/10.1038/s41586-018-0382-x
25. Burke,D.T., Carle,G.F. and Olson,M. V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. 1987. *Science*, **236**, 806–812. https://doi.org/10.1126/science.3033825
26. Larionov,V., Kouprina,N., Graves,J., Chen,X.N., Korenberg,J.R. and Resnick,M.A. (1996) Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proc. Natl. Acad. Sci. U. S. A.*, **93**, 491–496. https://doi.org/10.1073/pnas.93.1.491
27. Kunes,S., Botstein,D. and Fox,M.S. (1985) Transformation of yeast with linearized plasmid DNA. Formation of inverted dimers and recombinant plasmid products. *J. Mol. Biol.*, **184**, 375–387. https://doi.org/10.1016/0022-2836(85)90288-8
28. Noskov,V.N., Koriabine,M., Solomon,G., Randolph,M., Barrett,J.C., Leem,S.H., Stubbs,L., Kouprina,N. and Larionov,V. (2001) Defining the minimal length of sequence homology required for selective gene isolation by TAR cloning. *Nucleic Acids Res.*, **29**, 32. https://doi.org/10.1093/nar/29.6.e32
29. Gibson,D.G. (2009) Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res.*, **37**, 6984–6990. https://doi.org/10.1093/nar/gkp687
30. Oldfield,L.M., Grzesik,P., Voorhies,A.A., Alperovich,N., MacMath,D., Najera,C.D., Chandra,D.S., Prasad,S., Noskov,V.N., Montague,M.G., et al. (2017) Genome-wide engineering of an infectious clone of herpes simplex virus type 1 using synthetic genomics assembly methods. *Proc. Natl. Acad. Sci. U. S. A.*, **114**, E8885–E8894. https://doi.org/10.1073/pnas.1700534114
31. Shang,Y., Wang,M., Xiao,G., Wang,X., Hou,D., Pan,K., Liu,S., Li,J., Wang,J., Arif,B.M., et al. (2017) Construction and Rescue of a Functional Synthetic Baculovirus. *ACS Synth. Biol.*, **6**, 1393–1402. https://doi.org/10.1021/acssynbio.7b00028
32. Vashee,S., Stockwell,T.B., Alperovich,N., Denisova,E.A., Gibson,D.G., Cady,K.C., Miller,K., Kannan,K., Malouli,D., Crawford,L.B., et al. (2017) Cloning, Assembly, and Modification of the Primary Human Cytomegalovirus Isolate Toledo by Yeast-Based Transformation-Associated Recombination. *mSphere*, **2**. https://doi.org/10.1128/mspheredirect.00331-17
33. Gibson,D.G., Glass,J.I., Lartigue,C., Noskov,V.N., Chuang,R.Y., Algire,M.A., Benders,G.A., Montague,M.G., Ma,L., Moodie,M.M., et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329**, 52–56. https://doi.org/10.1126/science.1190719
34. Benders,G.A., Noskov,V.N., Denisova,E.A., Lartigue,C., Gibson,D.G., Assad-Garcia,N., Chuang,R.Y., Carrera,W., Moodie,M., Algire,M.A., et al. (2010) Cloning whole bacterial genomes in yeast. *Nucleic Acids Res.*, **38**, 2558–2569. https://doi.org/10.1093/nar/gkq119
35. Ruiz,E., Talenton,V., Dubrana,M.-P., Guesdon,G., Lluch-Senar,M., Salin,F., Sirand-Pugnet,P., Arfi,Y. and Lartigue,C. (2019) CReasPy-Cloning: A Method for Simultaneous Cloning and Engineering of Megabase-Sized Genomes in Yeast Using the CRISPR-Cas9 System. *ACS Synth. Biol.*, **8**, 2547–2557. https://doi.org/10.1021/acssynbio.9b00224
36. Rideau,F., Le Roy,C., Descamps,E.C.T., Renaudin,H., Lartigue,C. and Bébéar,C. (2017) Cloning, Stability, and Modification of *Mycoplasma hominis* Genome in Yeast. *ACS Synth. Biol.*, **6**, 891–901. https://doi.org/10.1021/acssynbio.6b00379
37. Karas,B.J., Tagwerker,C., Yonemoto,I.T., Hutchison,C.A. and Smith,H.O. (2012) Cloning the *Acholeplasma laidlawii* PG-8A genome in *Saccharomyces cerevisiae* as a Yeast Centromeric Plasmid. *ACS Synth. Biol.*, **1**, 22–28. https://doi.org/10.1021/sb200013j
38. Zhou,J., Wu,R., Xue,X. and Qin,Z. (2016) CasHRA (Cas9-facilitated Homologous Recombination Assembly) method of constructing megabase-sized DNA. *Nucleic Acids Res.*, **44**. https://doi.org/10.1093/nar/gkw475
39. Fredens,J., Wang,K., de la Torre,D., Funke,L.F.H., Robertson,W.E., Christova,Y., Chia,T., Schmied,W.H., Dunkelmann,D.L., Beranek,V., et al. (2019) Total synthesis of *Escherichia coli* with a recoded genome. *Nature*, **569**, 514–518. https://doi.org/10.1038/s41586-019-1192-5
40. Venetz,J.E., Medico,L. Del, Wölfle,A., Schächle,P., Bucher,Y., Appert,D., Tschan,F., Flores-Tinoco,C.E., Van Kooten,M., Guennoun,R., et al. (2019) Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. *Proc. Natl. Acad. Sci. U. S. A.*, **116**, 8070–8079. https://doi.org/10.1073/pnas.1818259116
41. Tagwerker,C., Dupont,C.L., Karas,B.J., Ma,L., Chuang,R.Y., Benders,G.A., Ramon,A., Novotny,M.,

Montague,M.G., Venepally,P., et al. (2012) Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic Acids Res.*, **40**, 10375–10383. https://doi.org/10.1093/nar/gks823
42. Noskov,V.N., Karas,B.J., Young,L., Chuang,R.Y., Gibson,D.G., Lin,Y.C., Stam,J., Yonemoto,I.T., Suzuki,Y., Andrews-Pfannkoch,C., et al. (2012) Assembly of large, high G+C bacterial DNA fragments in yeast. *ACS Synth. Biol.*, **1**, 267–273. https://doi.org/10.1021/sb3000194
43. Karas,B.J., Molparia,B., Jablanovic,J., Hermann,W.J., Lin,Y.C., Dupont,C.L., Tagwerker,C., Yonemoto,I.T., Noskov,V.N., Chuang,R.Y., et al. (2013) Assembly of eukaryotic algal chromosomes in yeast. *J. Biol. Eng.*, **7**, 1–12. https://doi.org/10.1186/1754-1611-7-30
44. O'Neill,B.M., Mikkelson,K.L., Gutierrez,N.M., Cunningham,J.L., Wolff,K.L., Szyjka,S.J., Yohn,C.B., Redding,K.E. and Mendez,M.J. (2012) An exogenous chloroplast genome for complex sequence manipulation in algae. *Nucleic Acids Res.*, **40**, 2782–2792. https://doi.org/10.1093/nar/gkr1008
45. Zhang,W., Zhao,G., Luo,Z., Lin,Y., Wang,L., Guo,Y., Wang,A., Jiang,S., Jiang,Q., Gong,J., et al. (2017) Engineering the ribosomal DNA in a megabase synthetic chromosome. *Science*, **355**, eaaf3981. https://doi.org/10.1126/science.aaf3981
46. Postma,E.D., Dashko,S., van Breemen,L., Taylor Parkins,S.K., van den Broek,M., Daran,J.-M. and Daran-Lapujade,P. (2021) A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **49**, 1769–1783. https://doi.org/10.1093/nar/gkaa1167
47. Postma,E.D., Hassing,E.J., Mangkusaputra,V., Geelhoed,J., de la Torre,P., van den Broek,M., Mooiman,C., Pabst,M., Daran,J.M. and Daran-Lapujade,P. (2022) Modular, synthetic chromosomes as new tools for large scale engineering of metabolism. *Metab. Eng.*, **72**, 1–13. https://doi.org/10.1016/j.ymben.2021.12.013
48. Mitchell,L.A., McCulloch,L.H., Pinglay,S., Berger,H., Bosco,N., Brosh,R., Bulajić,M., Huang,E., Hogan,M.S., Martin,J.A., et al. (2021) *De novo* assembly and delivery to mouse cells of a 101 kb functional human gene. *Genetics*. https://doi.org/10.1093/genetics/iyab038
49. Chen,W., Han,M., Zhou,J., Ge,Q., Wang,P., Zhang,X., Zhu,S., Song,L. and Yuan,Y. (2021) An artificial chromosome for data storage. *Natl. Sci. Rev.*, **8**, 1–9. https://doi.org/10.1093/nsr/nwab028
50. Thi Nhu Thao,T., Labroussaa,F., Ebert,N., V'kovski,P., Stalder,H., Portmann,J., Kelly,J., Steiner,S., Holwerda,M., Kratzel,A., et al. (2020) Rapid reconstruction of SARS-CoV-2 using a synthetic genomics platform. *Nature*, **582**, 561–565. https://doi.org/10.1038/s41586-020-2294-9
51. Vashee,S., Arfi,Y. and Lartigue,C. (2020) Budding yeast as a factory to engineer partial and complete microbial genomes. *Curr. Opin. Syst. Biol.*, **24**, 1–8. https://doi.org/10.1016/j.coisb.2020.09.003
52. Pretorius,I.S. and Boeke,J.D. (2018) Yeast 2.0—connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Res.*, **18**, 1–15. https://doi.org/10.1093/femsyr/foy032
53. Eisenstein,M. (2020) How to build a genome. *Nature*. https://doi.org/10.1038/d41586-020-00511-9
54. Dai,J., Boeke,J.D., Luo,Z., Jiang,S. and Cai,Y. (2020) Sc3.0: Revamping and minimizing the yeast genome. *Genome Biol.*, **21**, 1–4. https://doi.org/10.1186/s13059-020-02130-z
55. Sorek,R., Zhu,Y., Creevey,C.J., Pilar Francino,M., Bork,P. and Rubin,E.M. (2007) Genome-Wide Experimental Determination of Barriers to Horizontal Gene Transfer. *Science*, **318**, 1449–1453. https://doi.org/10.1126/science.1147112
56. Itaya,M., Sato,M., Hasegawa,M., Kono,N., Tomita,M. and Kaneko,S. (2018) Far rapid synthesis of giant DNA in the *Bacillus subtilis* genome by a conjugation transfer system. *Sci. Rep.*, **8**, 1–6. https://doi.org/10.1038/s41598-018-26987-0
57. Chandran,S. (2017) Rapid assembly of DNA via ligase cycling reaction (LCR). *Methods Mol. Biol.*, **1472**, 105–110. https://doi.org/10.1007/978-1-4939-6343-0_8
58. Kostylev,M., Otwell,A.E., Richardson,R.E. and Suzuki,Y. (2015) Cloning should be simple: *Escherichia coli* DH5α-mediated assembly of multiple DNA fragments with short end homologies. *PLoS One*, **10**, 1–15. https://doi.org/10.1371/journal.pone.0137466
59. Itaya,M., Fujita,K., Kuroki,A. and Tsuge,K. (2008) Bottom-up genome assembly using the *Bacillus subtilis* genome vector. *Nat. Methods*, **5**, 41–43. https://doi.org/10.1038/nmeth1143
60. Blount,B.A., Driessen,M.R.M. and Ellis,T. (2016) GC preps: Fast and easy extraction of stable yeast genomic DNA. *Sci. Rep.*, **6**, 1–4. https://doi.org/10.1038/srep26863
61. Lee,K., Ji,J.H., Yoon,K., Che,J., Seol,J.H., Lee,S.E. and Shim,E.Y. (2019) Microhomology selection for Microhomology mediated end joining in *Saccharomyces cerevisiae*. *Genes (Basel).*, **10**, 1–14. https://doi.org/10.3390/genes10040284
62. Ranjha,L., Howard,S.M. and Cejka,P. (2018) Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. *Chromosoma*, **127**, 187–214.

https://doi.org/10.1007/s00412-017-0658-1
63. Albarracín,V.H., Pathak,G.P., Douki,T., Cadet,J., Borsarelli,C.D., Gärtner,W. and Farias,M.E. (2012) Extremophilic *Acinetobacter* Strains from High-Altitude Lakes in Argentinean Puna: Remarkable UV-B Resistance and Efficient DNA Damage Repair. *Orig. Life Evol. Biosph.*, **42**, 201–221. https://doi.org/10.1007/s11084-012-9276-3
64. Sato,T., Takada,D., Itoh,T., Ohkuma,M. and Atomi,H. (2020) Integration of large heterologous DNA fragments into the genome of *Thermococcus kodakarensis*. *Extremophiles*, **24**, 339–353. https://doi.org/10.1007/s00792-020-01159-z
65. Dean,F.B., Nelson,J.R., Giesler,T.L. and Lasken,R.S. (2001) Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.*, **11**, 1095–1099. https://doi.org/10.1101/gr.180501
66. Lau,Y.H., Stirling,F., Kuo,J., Karrenbelt,M.A.P., Chan,Y.A., Riesselman,A., Horton,C.A., Schäafer,E., Lips,D., Weinstock,M.T., et al. (2017) Large-scale recoding of a bacterial genome by iterative recombineering of synthetic DNA. *Nucleic Acids Res.*, **45**, 6971–6980. https://doi.org/10.1093/nar/gkx415
67. Kwon,Y., Daley,J.M. and Sung,P. (2017) Reconstituted System for the Examination of Repair DNA Synthesis in Homologous Recombination. In Eichman,B.F. (ed), *Methods in Enzymology*. Elsevier Inc., Vol. 591, pp. 307–325. https://doi.org/10.1016/bs.mie.2017.03.021
68. 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**, 751–5. https://doi.org/10.1038/90802
69. Guo,Z.-P., Borsenberger,V., Croux,C., Duquesne,S., Truan,G., Marty,A. and Bordes,F. (2020) An artificial chromosome ylAC enables efficient assembly of multiple genes in *Yarrowia lipolytica* for biomanufacturing. *Commun. Biol.*, **3**. https://doi.org/10.1038/s42003-020-0936-y
70. Boeke,J.D., Church,G., Hessel,A., Kelley,N.J., Arkin,A., Cai,Y., Carlson,R., Chakravarti,A., Cornish,V.W., Holt,L., et al. (2016) The Genome Project–Write. *Science*, **353**, 126–127. https://doi.org/10.1126/science.aaf6850
71. Mutschler,H., Robinson,T., Tang,T.-Y.D. and Wegner,S. (2019) Special Issue on Bottom-Up Synthetic Biology. *ChemBioChem*, **20**, 2533–2534. https://doi.org/10.1002/cbic.201900507
72. Belda,I., Williams,T.C., de Celis,M., Paulsen,I.T. and Pretorius,I.S. (2021) Seeding the idea of encapsulating a representative synthetic metagenome in a single yeast cell. *Nat. Commun.*, **12**, 1–8. https://doi.org/10.1038/s41467-021-21877-y

Chapter 3

Synthetic chromosome assembly in yeast for cell-free protein synthesis

De novo design and assembly of a DNA genome represents a major challenge for the construction of a minimal synthetic cell. We have explored homologous recombination in the yeast *Saccharomyces cerevisiae* as a method for the assembly of synthetic chromosomes from multiple DNA fragments. The compatibility of this approach with the constraints imposed by the utilization of a reconstituted transcription-translation system (PURE) for protein production was tested. Two challenges were identified: (i) the assembly of sequences with repeats arising from the limited choice of regulatory elements in PURE system and (ii) isolation of the assembled synthetic chromosome from yeast, which results in DNA with low concentration and purity. We showed that assembly in yeast of 20 DNA fragments, of which ten contain repeats consisting of PURE regulatory sequences and ten contain screening markers, into a 67-kb centromeric chromosome is possible, albeit with low efficiency. We established a screening pipeline to strategically identify correct assemblies by using auxotrophic, fluorescent, antibiotic resistance and chromogenic markers. Assemblies were verified by Nanopore sequencing, providing insights into the role of repeats in unintended recombination events. The model synthetic chromosome was stable during propagation in yeast for at least 32 generations, could be extracted, and used as template for *yfp* expression in PURE system. Further improvement of the chromosome extraction protocol from yeast will be needed to increase cell-free protein production. Finally, suggestions for design optimization are presented, which is necessary for efficient expression of the numerous genes required in a PURE system-based synthetic cell.

Introduction

A grand challenge in synthetic biology is the construction of a functioning cell from lifeless components [1–3]. The bottom-up assembly of a synthetic cell will likely have a profound impact on our understanding of life's fundamental principles, as well as on the development of new biotechnologies. A major conceptual and technical challenge toward building a synthetic cell resides in the design and assembly of a minimal synthetic chromosome that would encode the genetic information and could be inherited for cell perpetuation. The exact gene content, organization, and DNA size that would support life in its simplest form are currently unknown. Estimates range from the theoretical approximations of 150–400 protein-coding genes [4–8] to an upper limit given by the 531-kb genome of the JCVI-syn 3.0 cell, containing 473 genes (of which 438 are protein-coding), which has been designed through genome reduction of *Mycoplasma mycoides* [9]. Importantly, a prospective synthetic cell genome will likely contain natural DNA parts derived from various species, as well as *de novo* designed elements [10].

A suitable method for the construction of a synthetic cell genome should thus enable the assembly of >100-kb DNA from multiple fragments, and preferably allow for the modular, one-pot assembly from individual transcription cassettes to ease testing a variety of genome designs. Despite recent successes to develop *in vitro* DNA assembly methods [11, 12] (reviewed in [13]), this approach is currently not compatible with the large genome size and number of transcription units. Moreover, *in vitro* methods require transformation of *E. coli* for selection and amplification [14], which is limited by the drop in transformation efficiency with increasing DNA size [15] and by the unpredictable toxicity of heterologous genes due to unwanted transcription [16].

The most well-established method for one-pot assembly of multiple fragments into large constructs is *in vivo* assembly in the yeast *Saccharomyces cerevisiae*. This homology-based assembly approach was used for the construction of several genomes in the JCVI *Mycoplasma* project [9, 17, 18]. It is capable of one-pot assembly of 44 fragments [19], and is suitable for the assembly of chromosomes up to 1.66 Mb [20] (reviewed in [21]). By including homologous overhangs in the assembly fragments, the homology-directed repair (HDR) machinery of *S. cerevisiae* can be employed to stitch linear DNA fragments upon transformation (Figure 3.1A). Moreover, the use of synthetic homology regions (SHRs) enables modularity of the assembly design [22].

Important to consider when designing the synthetic genome, are the genetic elements controlling transcription and translation and the cell-free protein synthesis system that operates in the minimal cell. A variety of cell-free gene expression systems are available [23, 24]. Protein Synthesis Using Recombinant Elements (PURE) system [25, 26] is particularly relevant in the synthetic cell framework because of its reconstituted nature, well-defined composition, and minimal number of proteins and cofactors. PURE system has successfully been utilized for expression of several proteins involved in key cellular processes, such as DNA replication [27], phospholipid synthesis for membrane growth [28, 29], division [30–32], and regeneration of some parts of the translation system [33–36]. In the commercial version of PURE system, transcription is controlled by the T7 RNA polymerase, while translation relies on *E. coli* ribosomes and factors. As we consider PURE system to be at the core of the design strategy for building a synthetic cell, the DNA chromosome should harbor the appropriate regulatory se-

quences, i.e., T7-based transcriptional promoter and terminator, and *E. coli* ribosome binding site (RBS). Although additional regulatory sequences could be envisaged (see Discussion), the limited choice of regulatory genetic elements will inevitably result in a high occurrence of repeats in the synthetic genome, which might mislead HDR and produce unwanted assemblies of the DNA fragments in yeast [37–41] (Figure 3.1B). However, the capability of *S. cerevisiae* to correctly assemble DNA parts containing heterologous sequences with multiple repeats remains unexplored.

In this study, we establish yeast as an assembly platform to construct minimal genomes for PURE-based synthetic cells. We designed cassettes containing PURE regulatory sequences at both ends, and included SHRs through PCR. After transformation of yeast with 20 DNA fragments and screening for correct assemblies using selection markers, the 67-kb synthetic chromosome was isolated from yeast and directly assayed for expression in PURE system (Figure 3.1A).

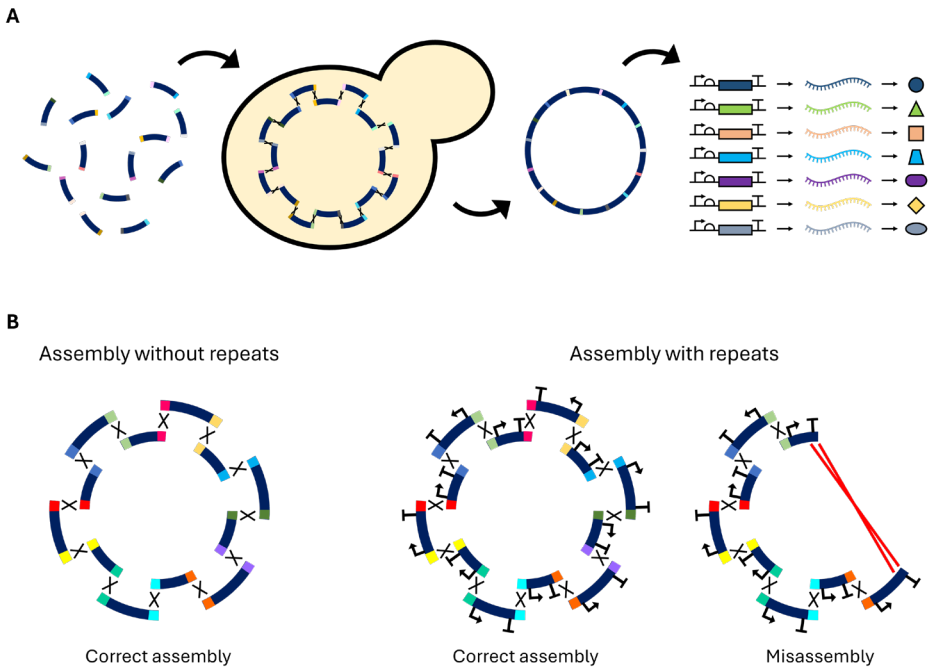


Figure 3.1: Assembly in yeast of a minimal genome for PURE-based synthetic cells. A) DNA fragments are obtained through PCR amplification of cassettes using primers with SHR overhangs. *S. cerevisiae* is transformed and the fragments are assembled through HDR. The synthetic chromosome is isolated from yeast and assayed for expression in PURE system. **B)** Without repeated sequences, assembly is expected to be correct. Repeated sequences present in the DNA fragments (promoters, 5' UTRs and terminators) are expected to cause misassembly in yeast, but correct assemblies might still occur.

Results

Design of a synthetic chromosome for expression in PURE system

A synthetic chromosome (named SynChr^{PURE}) was designed for expression in PURE system following three main requirements: (i) the presence of at least one transcription unit leading to expression in PURE system, (ii) the ability to be assembled and maintained in yeast, and (iii) the possibility for easy screening of yeast clones with correctly assembled chromosomes. In addition, SynChr^{PURE} was equipped with an *E. coli* replication origin, an optional feature in case amplification of the chromosome would be required.

The designed SynChr^{PURE} includes ten transcriptional units. Each cassette starts with the same 139-bp sequence containing a T7 promoter and an *E. coli* RBS, and ends with the same 119-bp sequence containing a T7 terminator, hereafter called 'PURE repeats' (Figure 3.2). One of these cassettes contains the reporter gene *yfp* for characterization of protein synthesis in PURE system. The other nine cassettes were designed as mock transcription units with standardized DNA parts. They do not lead to cell-free expressed proteins in PURE and were designed to be non-coding in both *S. cerevisiae* and *E. coli* to avoid expression of potentially toxic proteins. They consist of 5-kb genomic sequences from the plant *Arabidopsis thaliana* flanked by PURE repeats. Expression of plant DNA is not expected in yeast [42] and, while transcription and translation may not totally be eliminated in *E. coli*, this design should keep it to a minimum [16, 43–47]. The *A. thaliana* DNA fragments were checked for the absence of T7 promoters to prevent transcription in PURE system.

To simplify the screening pipeline for identification of yeast strains harboring correctly assembled chromosomes, a series of nine genetic selection and screening markers were included as DNA parts interspacing the PURE transcription cassettes (Figure 3.2): *hphNT1* conferring resistance to hygromycin, *crtYB*, *crtE,* and *crtI* leading to β-carotene biosynthesis identifiable by the orange coloring of colonies in the presence of all three genes, *mTurqoise2* and *mRuby2* encoding fluorescent proteins, and three auxotrophic markers *URA3*, *LEU2,* and *HIS3* enabling growth in the absence of uracil, leucine, and histidine, respectively.

For replication and segregation of SynChr^{PURE} in yeast, a centromeric origin and an autonomously replicating sequence were added to the design as a single fragment (CEN6/ARS4). We therefore expected SynChr^{PURE} to be present in a single copy per cell on average. Alternatively, we explored a design incorporating a 2μ replication origin (SynChr$^{PURE_2\mu}$) to increase the SynChr copy number and to allow higher yields during SynChr extraction. Finally, a bacterial ColE1-type replication origin and the *bla* antibiotic resistance gene (conferring resistance to ampicillin) were added on the same DNA fragment as the yeast origin of replication, giving the possibility to maintain and amplify the chromosome in *E. coli*.

Summarizing, the *de novo* designed SynChr^{PURE} comprises ten fragments with PURE expression cassettes (PF1–10), nine screening markers scattered over the chromosome

between the PURE fragments, and a fragment containing genetic elements for the replication and propagation in yeast and in *E. coli*, giving an expected size of about 69 kb (Figure 3.2). All 20 DNA parts were flanked by 60-bp SHR sequences overlapping the adjacent fragments to promote the assembly of SynChr^{PURE} in yeast via HDR. A control synthetic chromosome (SynChr^{control}) was also designed, differing only in the absence of PURE repeats, with a total size of 66 kb (Figure 3.2). Similarly, a control SynChr with 2μ origin was designed (SynChr$^{control_2\mu}$).

All designed SynChr maps are available in Supplementary Data.

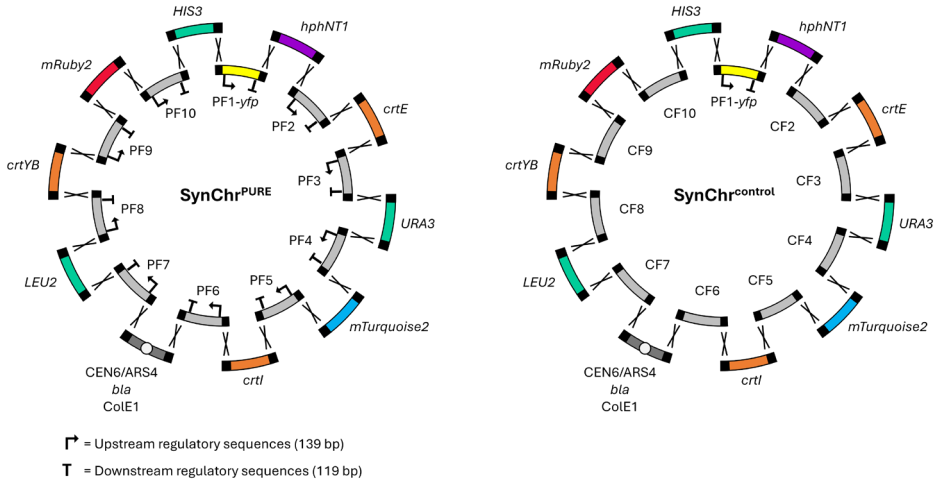


Figure 3.2: Design of a synthetic chromosome for expression in PURE system. Design of SynChr^{PURE} (left) and SynChr^{control} (right), differing only by the absence of PURE repeats in CF2–CF10. PF = PURE fragment. CF = control fragment.

Screening pipeline and selection of positive clones

Long-read DNA sequencing is the most accurate method to determine whether assembled chromosomes are faithful to the *in silico* design, but it has a low throughput and is costly when screening hundreds of clones. Therefore, we established a pipeline to screen, prior to sequencing, for yeast strains harboring chromosomes with all fragments (Figure 3.3). The screening pipeline was designed to exclude PCR verification for two reasons: the inherent risk of false negative results, and in case of low assembly efficiency, colony screening by PCR would be excessively labor-intensive.

After transformation and assembly in yeast, plating on YPD with hygromycin selected for strains with chromosomes carrying the *hphNT1* fragment (Figure 3.3, step 1). Visual inspection of the colonies after three days further revealed the presence of all three carotenoid genes (Figure 3.3, step 2), as their simultaneous presence leads to an orange coloring of the colonies [48]. Resuspension of yeast colonies and fluorescence analysis by flow cytometry identified clones expressing the *mRuby2* and *mTurquoise2* genes (Figure 3.3, step 3). Finally, plating on minimal synthetic medium selected for colonies harboring all three auxotrophic markers (Figure 3.3, step 4). For additional information, plating on media lacking single nutrients could be used to identify the

absence of specific auxotrophic markers. The entire screening pipeline was designed to test hundreds of clones within six days for the presence or absence of the nine markers. This would allow identification of correct assemblies with occurrence frequencies as low as 0.3–1%. Positive clones were sequenced using Nanopore technology (Figure 3.3, step 5).

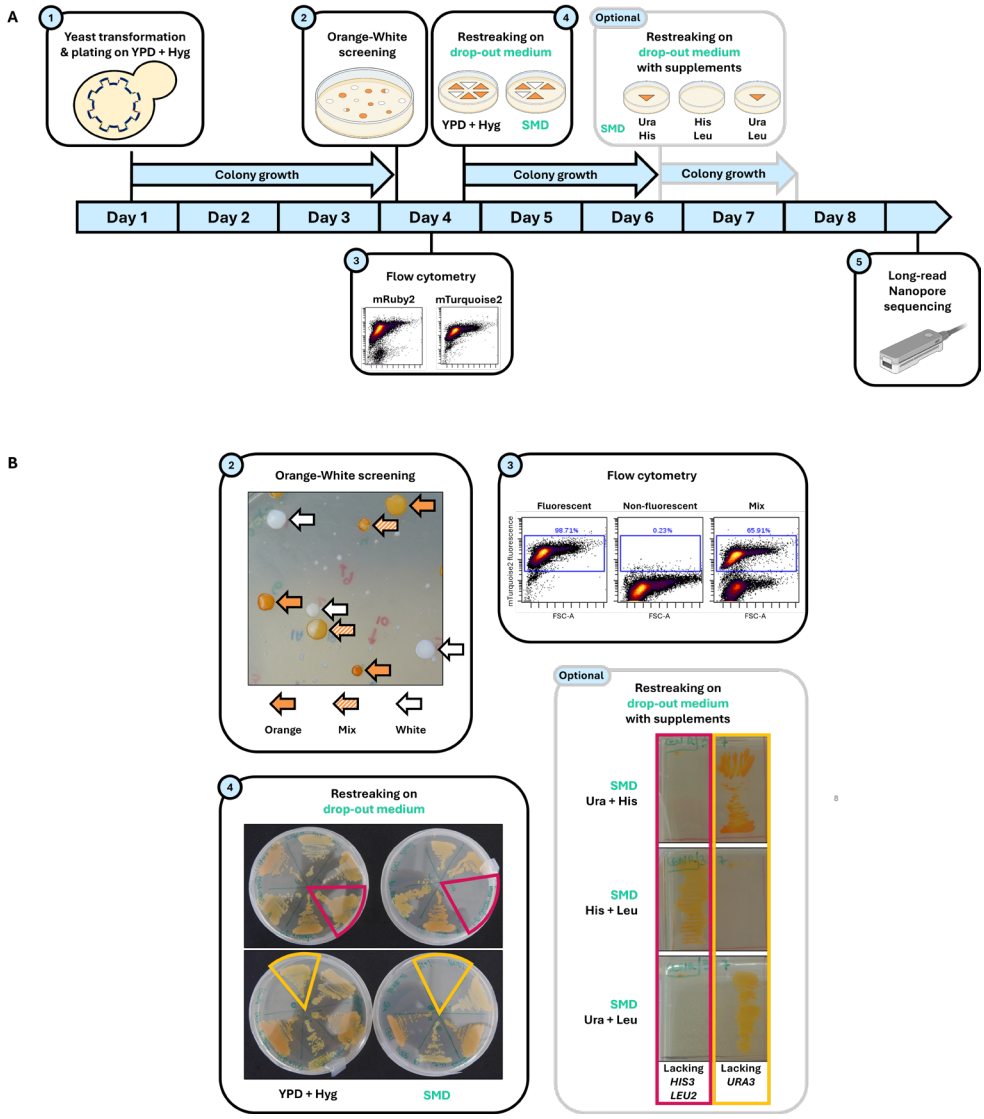


Figure 3.3: Screening pipeline to select clones with correct assemblies. A) Step 1: Transformation of yeast with DNA fragments and plating on selective agar plates containing YPD and hygromycin. Visible colonies are formed within three days. Step 2: Orange-White screening by visual inspection, based on β-carotene production. Step 3: mRuby2 and mTurquoise2 detection by flow cytometry. Step 4: Detection of auxotrophies by restreaking on YPD + Hyg and SMD. Growth is visible within two days. Optional step: in case of absence of growth on SMD, specific auxotrophies can be identified by restreaking on various drop-out media followed by growth for two days. Step 5: Positively screened strains were selected for long-read Nanopore sequencing. **B)** Panels with classification examples for steps 2, 3, 4 and optional screening step.

Three independent yeast transformations were conducted with the 20 assembly fragments for both the SynChr^{PURE} and SynChr^{control}. The total number of colonies over three transformations was considered rather than individual transformations to report the assembly efficiency, as this mitigates expected day-to-day variation between independent transformations. We obtained a total of 53 SynChr^{PURE} and 29 SynChr^{control} colonies for screening step 1 (Figure 3.4A). Visual inspection for the presence of carotenoid marker genes (step 2) showed that for SynChr^{PURE}, 25% of the colonies (13 out of 53 colonies) were orange, while 30% were white (16 colonies) and 45% (24 colonies) displayed mixed colors (Figure 3.4A and B). For the SynChr^{control}, 90% (26 out of 29) colonies were orange, 7% (2 colonies) were white, and a single colony (3%) displayed a mixed phenotype (Figure 3.4A and B). The results confirm the expected higher assembly efficiency in the absence of PURE repeats.

Mixed phenotypes reveal the presence of mixed populations that might result from different events. Upon plating after transformation, several adjacent cells could form a single colony. Alternatively, colonies could result from single cells that harbor multiple chromosomes with different configurations that segregate into separate daughter cells during division. Finally, mixed population could result from chromosome instability and recombination upon cell propagation. The possible occurrence of chromosome instability was investigated as described in the section "The synthetic chromosomes are stable during growth". Irrespective of the underlying events, the presence of PURE repeats and resulting undesired recombination events increased the frequency of colonies with mixed phenotype compared to SynChr^{control} colonies.

Through the following screening steps, five out of the 13 orange SynChr^{PURE} colonies (representing 9% of total colony count in step 1) were positively screened for both fluorescent and auxotrophic markers (Figure 3.4B), suggesting correct assemblies. For the SynChr^{control}, 25 out of the 26 orange colonies carried the two fluorescent and the three auxotrophic markers (representing 86% of the total colonies from step 1) (Figure 3.4B). Data from individual transformations are reported in Figure S3.1. These findings suggest that correct assembly of SynChr^{PURE} may be possible, albeit with a lower efficiency than that of SynChr^{control}, indicating that PURE repeats cause undesired recombination events.

Finally, we explored the potential of the 2µ replication origin by performing a 20-fragment assembly for both SynChr$^{PURE_2\mu}$ and SynChr$^{control_2\mu}$. For SynChr$^{control_2\mu}$, only one of the resulting 6 colonies was orange (Figure S3.2A and C). Further screening of the orange colony for the presence of *mRuby2* and *mTurquoise2* fragments revealed a mixture of phenotypes with and without fluorescence and the colony was therefore classified as incorrect (Figure S3.2B). For SynChr$^{PURE_2\mu}$, none of the 12 colonies displayed an orange phenotype (Figure S3.2A and C). Due to the low assembly efficiency revealed by this preliminary data, further work on the 2µ version was discontinued.

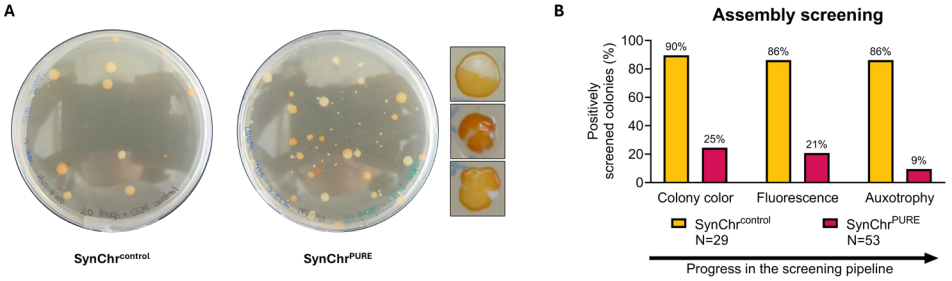


Figure 3.4: Screening for SynChrs with correct assembly. A) Example of plates with YPD and hygromycin after transformation. SynChr^{control} (left) and SynChr^{PURE} (right). Zoom-in images show colonies with mixed phenotype regarding the presence of carotenoid genes. **B)** Percentage of positively screened colonies at each screening step, compared to step 1. "Positive" means that the clone carries the markers of the inspected screening step and all previously screened markers. Colonies that grew on YPD with hygromycin plates were classified as positive when displaying an orange color, when showing fluorescence for both mTurquoise2 and mRuby2, and when growth on SMD was visible. Aggregated data from three independent transformations are displayed.

Long-read DNA sequencing confirms the correct assembly of SynChr^{PURE}

The five strains identified by the screening pipeline as harboring SynChr^{PURE} with all expected marker fragments were named IMF50 to IMF54 and sequenced by long-read Nanopore sequencing (Table S3.3). Four out of the five strains showed correct SynChr configurations (IMF50, IMF51, IMF52 and IMF54) (Table S3.3). In strain IMF53, recombination events occurred between PURE repeats as revealed by the raw reads, but no consensus sequence could be determined. In-depth inspection of the sequences showed that IMF51 and IMF54 were near-perfect replica of the *in silico* design (differing only by five point mutations for IMF51 and seven point mutations for IMF54). The total expected size of the SynChr^{PURE} was 69 kb, however, due to variations in DNA fragment size between the *A. thaliana* reference sequence used for the design and the final size in the template plasmids, the size of the constructed SynChr^{PURE} was 67 kb. The SynChrs of IMF50 and IMF52 showed some minor deviations that were most likely not caused by the presence of PURE repeats (Figure S3.3). IMF50 missed a 1,992-bp sequence in the middle of the fragment with CEN6/ARS4, leading to the absence of ARS4, *bla* and ColE1 (Figure S3.4B). The missing sequence does not contain similarity with other parts of the SynChr^{PURE}, indicating that unexpected homologous recombination events were not responsible for misassembly. IMF52 harbored an insertion of a 1,780-bp sequence in the terminator region downstream of PF9, corresponding to a plasmid backbone used as template during PCR, which suggests that this backbone was a contaminant in the transformation mix.

To explore if PURE repeats were involved in misassemblies of SynChr^{PURE}, six strains that did not pass the screening due to the absence of at least one marker were sequenced (strains IMF55, IMF56, IMF59, IMF61, IMF62 and IMF64, Figure 3.5, Table S3.3). Based on the phenotype observed during the different screening stages, configurations could be predicted for the misassembled SynChrs (Figure 3.5). For white colonies, no prediction

could be made as to which of the three markers for β-carotenoid were missing. Four out of the six strains (IMF55, IMF56, IMF561 and IMF62) showed the predicted configuration. Recombination had occurred between either promoter or terminator regions. The strain IMF55, similarly to IMF50, missed part of the CEN6/ARS4_*bla*_ColE1 fragment (2,076 bp, Figure S3.4C). For IMF59, a duplication of PF10 and the *HIS3* fragment had occurred, which was inserted in the promoter region upstream PF9 (Figure 3.5), likely due to homologous recombination between PURE repeats. For IMF64, no consensus sequence could be determined because of a low depth of sequencing reads. These results show that for most sequenced chromosomes, PURE repeats were responsible for misassemblies. Notably, recombination between SHRs is more frequent than between PURE repeats in the selected strains. Moreover, our screening pipeline successfully predicts chromosome configurations, even though it does not detect duplications or insertions, as expected.

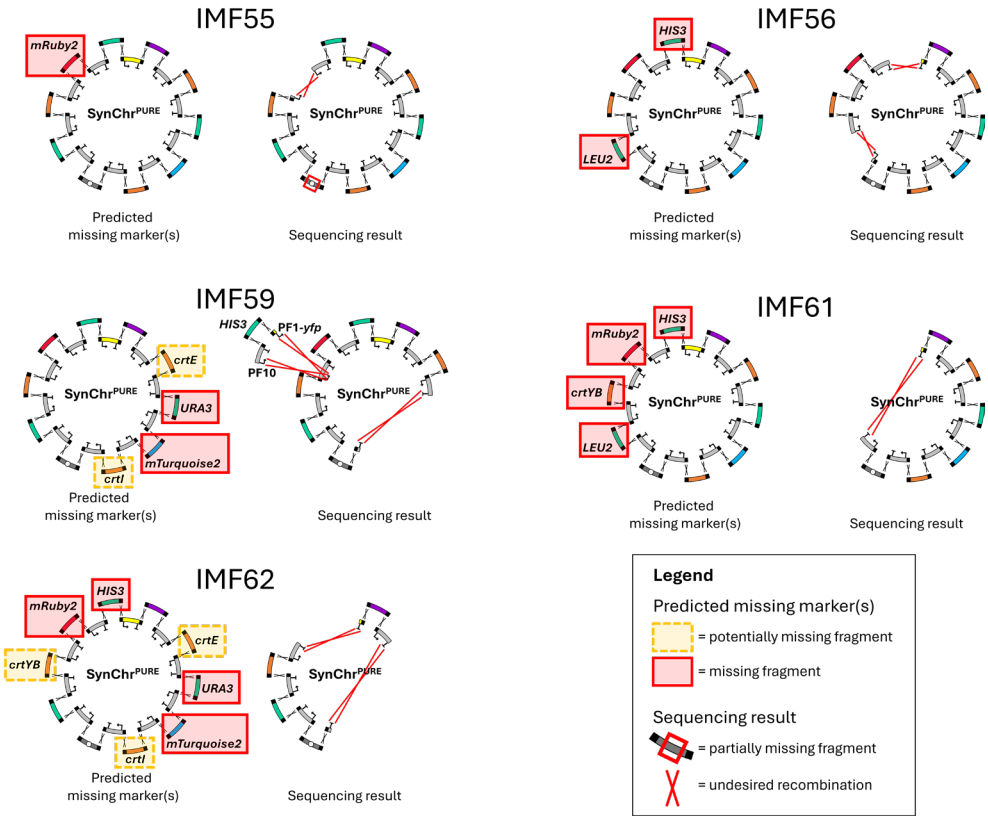


Figure 3.5: Predicted versus sequenced SynChr^{PURE} configurations for strains missing at least one marker. On the left the predicted missing markers. On the right the SynChr^{PURE} configuration as revealed by sequencing.

A single strain harboring SynChr^{control} (IMF49) with all expected marker fragments was sequenced. Based on previous studies reporting a relatively high assembly efficiency (36%) for 44 fragments [19], an even higher efficiency was expected for SynChr^{control} from 20 fragments, eliminating the need for sequencing multiple strains. The sequencing results confirmed the presence of all markers and DNA fragments with the expected configuration (Figure S3.3). However, a 89-bp insert was present after the first SHR of CF4 and, similarly to what was observed for IMF50 and IMF55, part of the CEN6/ARS4_*bla*_ColE1 fragment (1,891 bp) showed approximately half the sequencing depth of the rest of the SynChr^{control} (Figure S3.4A).

Overall, Nanopore sequencing confirmed that *S. cerevisiae* is capable of successfully assembling SynChrs from at least ten fragments with repeats, albeit with a relatively low assembly efficiency (8%). The established screening pipeline allowed us to select positive clones with the desired configuration of synthetic chromosomes, alleviating the need for extensive sequencing.

The synthetic chromosomes are stable during growth

An important requirement for *S. cerevisiae* to become a genome foundry for PURE-based synthetic cells is the stability of SynChrs in yeast cells upon assembly. The high incidence of colonies with mixed phenotypes (45%) for the SynChr^{PURE} assembly might be caused by post-assembly homologous recombination events between PURE repeats. As HDR is primarily active during cell division (S and G2 phases of the mitotic cell cycle [49]), the risk of recombination events caused by the presence of repeats might increase during cell propagation. Many generations are required for completion of the workflow, from plating of the transformation mix, to screening, storing and finally propagating for SynChr extraction. The stability of SynChrs during propagation was therefore explored. Strain IMF54 harboring a correctly assembled SynChr^{PURE} and strain IMF49 carrying the SynChr^{control} were grown in liquid medium, and subjected to four serial transfers (Figure 3.6A).

Two different media were used to exert different levels of selection pressure for the maintenance of integer SynChrs. Growth on SMD without supplement selected for the presence of all three auxotrophic markers *HIS3*, *LEU2* and *URA3*, while survival on SMD supplemented with uracil and histidine only required the presence of a single marker (*LEU2*, Figure 3.2). The integrity of SynChrs was verified by measuring mRuby2 and mTurquoise2 fluorescence by flow cytometry, and by plating combined with visual inspection for β-carotene production after 44 and 88 hours of growth in culture (ca. 16 and 32 generations, respectively) (Figure 3.6A). For both strains (IMF49 and IMF54), the percentage of cells with mTurquoise2 fluorescence remained above 98% after 88 hours, in both selection conditions (Figure 3.6B). The percentage of cells showing fluorescence was not significantly affected by the difference in selection pressure (two-way ANOVA with Post-Hoc Tukey-Kramer, $p > 0.005$). Similar results were obtained after 44 hours and for mRuby2 (Figure S3.5A–C). On plates, only 0 to 3 white colonies were observed on a total of more than 500 colonies per plate (Figure 3.6C and Figure S3.5D–F). Overall, our results indicate that SynChrs are stable during propagation in yeast and that the presence of PURE repeats does not hamper stability. We therefore speculate that the population heterogeneity shown by the mixed colors of 45% of the

transformants may have occurred earlier during chromosome assembly or may be due to the fact that two different colonies converged on the same physical location on the plate.

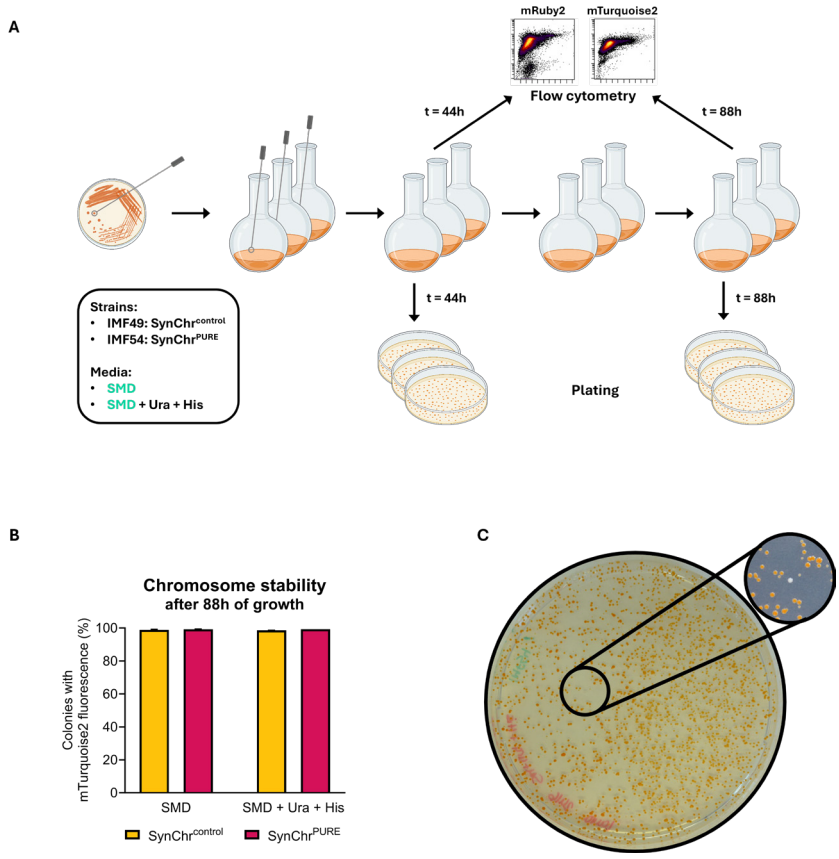


Figure 3.6: SynChrs are stable during strain propagation. A) Experimental set-up to test SynChr stability. Strains IMF49 (SynChr^{control}) and IMF54 (SynChr^{PURE}) were propagated for 4 days (88h) in triplicates in 100 mL high-selection medium SMD and low-selection medium SMD + Ura + His. Media were refreshed every day. Samples were taken on day 2 (after 44h) and day 4 (after 88h), and tested for fluorescence by flow cytometry and for orange color by plating (in the same selective medium as the liquid cultures). **B)** Flow cytometry results for mTurquoise2 after 4 days (88h) of growth in culture, in both media conditions, for both strains. **C)** Representative plate showing a very low incidence of white colonies.

Synthetic chromosomes isolated from *S. cerevisiae* can be expressed in PURE system

Next, we investigated the ability of SynChr^{PURE} to serve as template for *yfp* expression in PURE system. Several iterations of protocol improvement were necessary to extract sufficient SynChr^{PURE} from yeast and achieve detectable levels of YFP. All attempted protocols are detailed in Methods. We noticed that careful yeast spheroplasting to avoid osmotic lysis, followed by column purification using a kit designed for large-construct isolation, was essential to extract sufficient amounts of SynChr^{PURE}. Two isolations were

performed using the final protocol, which differed in starting yeast cell count (ca. four times more for the second isolation). Data from the first isolation are reported in Figure S3.6, while data from the second isolation are presented in Figure 3.7. Total concentration of DNA was measured by Qubit (Figure 3.7A). The concentration of SynChr^{PURE} was determined by qPCR using primers targeting specific regions on the chromosome (Figure 3.7B).

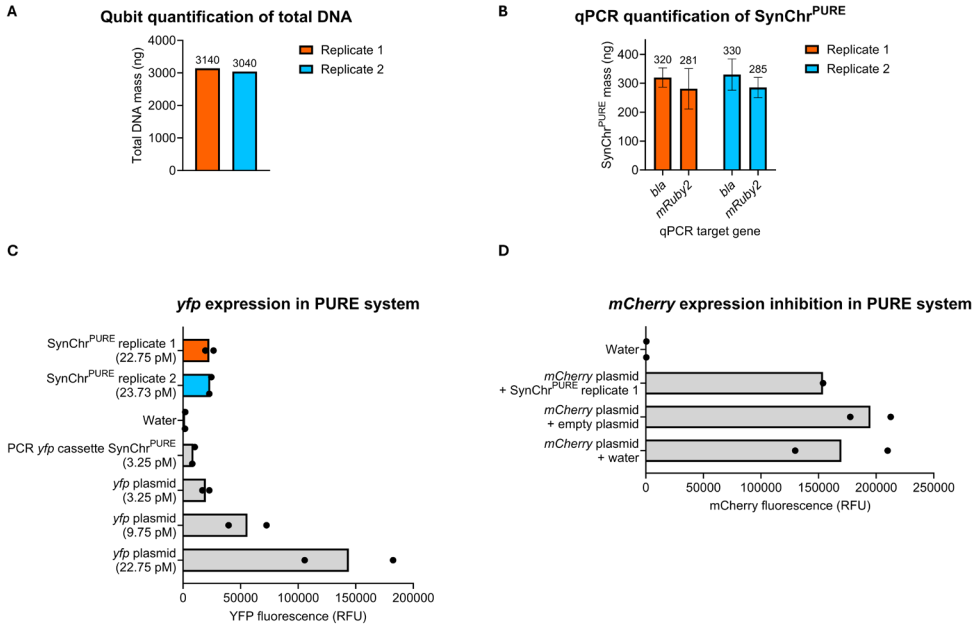


Figure 3.7: SynChr^{PURE} can be isolated from *S. cerevisiae* and shows expression of *yfp* in PURE system. Data from the second isolation are presented. **A)** Qubit quantification of total isolated DNA mass (ng) in 100 µL. **B)** qPCR quantification of SynChr^{PURE} mass (ng) in 100 µL, quantified by primer sets targeting *bla* and *mRuby2* genes. Mean and standard deviation from three technical replicates are plotted. **C)** YFP fluorescence measured after 16 h incubation of various DNA templates in PURE system. Individual data points represent averages of three fluorescence measurements from two independent PURE reactions. Final DNA template concentrations in the assembled PURE reaction are indicated between brackets. SynChr^{PURE} = SynChr^{PURE} isolated from *S. cerevisiae* strain IMF54. Water = No DNA template (negative control). *yfp* plasmid = Plasmid containing a *yfp* gene, isolated from *E. coli* (positive control). PCR *yfp* cassette SynChr^{PURE} = *yfp* cassette amplified by PCR from SynChr^{PURE} isolated from *S. cerevisiae* strain IMF54 (positive control). **D)** mCherry fluorescence measured after 16 h incubation of a diluted *mCherry* plasmid in PURE system. Individual data points represent averages of three fluorescence measurements from two independent PURE reactions. Water = No DNA template (negative control). *mCherry* plasmid + SynChr^{PURE} = 0.5 µL of *mCherry* plasmid (final concentration: 1 nM) isolated from *E. coli*, diluted in 2.75 µL SynChr^{PURE} sample. *mCherry* plasmid + water = 0.5 µL of *mCherry* plasmid (final concentration: 1 nM) isolated from *E. coli*, diluted in 2.75 µL water (positive control). *mCherry* plasmid + empty plasmid = 0.5 µL of *mCherry* plasmid (final concentration: 1 nM) isolated from *E. coli*, diluted in 2.75 µL non-coding plasmid.

About 10% of total DNA isolated from strain IMF54 corresponded to SynChr^{PURE} (Figure 3.7A and B). Approximately 300 ng of SynChr^{PURE} DNA could be extracted, which was sufficient for assaying expression of the encoded *yfp* gene in PURE system. Starting from about 20 pM of SynChr^{PURE} template, clear production of YFP was detected by spectrofluorometry after 16h of incubation, compared to the negative control containing water

instead of DNA (Figure 3.7C). We also performed a series of control reactions using two reference templates: a plasmid isolated from *E. coli* coding for the YFP reporter protein (added at various concentrations, one of which similar to the SynChr^{PURE} concentration) and a *yfp* transcriptional unit that was PCR-amplified from SynChr^{PURE} (3.25 pM final concentration). Both control reactions revealed a higher relative yield of synthesized YFP per DNA template than with SynChr^{PURE}. As the lower yield may result from inhibitory effects from the SynChr extract (e.g., the presence of native yeast DNA), we mixed a plasmid encoding mCherry with the SynChr^{PURE} extract and measured its fluorescence after 16 h (Figure 3.7D). No significant drop of mCherry signal could be measured as compared to the control condition, where the SynChr^{PURE} solution was substituted with water or a non-coding plasmid isolated from *E. coli*, indicating that the SynChr extract did not visibly inhibit expression in PURE system. Most likely, the presence of the nine other transcription units present in SynChr^{PURE} (vs. a single transcription unit in the reference templates) may lower the concentration of free T7 RNA polymerase, thus reducing the amounts of available resources for *yfp* transcription. Additionally, the lack of accuracy in determining the concentration and purity of SynChr^{PURE} could impact the results.

Discussion

S. cerevisiae is an emerging DNA assembly platform for synthetic genomics due to its powerful homologous recombination capability. However, the application of *S. cerevisiae* for the construction of *de novo* designed chromosomes used for gene expression in minimal cell-free systems remains unexplored. This novel application comes with two challenges: (i) the presence of repeated sequences on the chromosome, and (ii) the transfer of the assembled chromosomes from yeast to cell-free systems. The first challenge arises from the limited set of regulatory sequences that is available for T7 RNA polymerase-based expression in PURE system. Therefore, many repeats will be present on the chromosomes, which are known to cause misassembly in yeast [37–41]. The second challenge is the isolation of intact, pure chromosomes from yeast in high concentration, necessary for cell-free expression. Existing protocols and kits for isolation of circular chromosomes or plasmids from yeast have been developed for subsequent transformation of a living host or for amplification by PCR, but these applications do not require chromosomes in high quantity or purity [50, 51].

Assembly in yeast of a 67-kb synthetic chromosome from 20 DNA fragments, of which ten contained repeated regulatory sequences, was successful, albeit with low efficiency. The screening pipeline could be carried out in just a week and served to identify incorrect assemblies, as well as potentially correct chromosome configurations. The effectiveness of the pipeline was demonstrated for a 20-fragment assembly. However, it will be lower as the number of fragments (genes) increases, exceeding the number of available markers. Moreover, including many markers in the final design will 'unnecessarily' increase the size of the minimal genome. Therefore, a great number of assemblies will have to be sequenced via whole-genome sequencing, which has lately become more affordable.

Although HR events occurred more frequently between SHRs than between PURE repeats, scaling up the number of genes to build a synthetic cell (>150 genes) or com-

plex metabolic pathways for cell-free production will inevitably increase the rate of undesired recombination events, further reducing the fraction of correctly assembled SynChrs. Possible solutions to increase the yield of correctly assembled chromosomes can be envisaged, either separately or in combination: (i) precloning of multiple expression cassettes into one fragment (using conventional molecular cloning methods), prior to the final assembly in yeast, (ii) reducing the degree of similarity between PURE repeats [52], using existing libraries of T7 promoters [53] and RBS sequences [54] that can simultaneously serve to regulate expression levels, (iii) moving the PURE repeats further inward of the fragments, to be overlooked by the HDR machinery, which predominantly targets the fragment ends, and (iv) altering the processivity of the nucleases Exo1 and Sgs1-Dna2 to shorten the ssDNA overhang, excluding the PURE repeats from involvement in recombination [55].

A more radical design strategy to overcome the lack of diversity in regulatory sequences that is inherent when using the commercial T7 RNA polymerase-based PURE system is to employ an alternative RNA polymerase, such as the *E. coli* RNA polymerase. While this will surely expand the number of regulatory elements available, auxiliary transcription factors were required by the *E. coli* RNA polymerase to achieve transcription rates comparable to those of the T7 RNA polymerase [56]. Including these factors, along with all protein subunits of the *E. coli* RNA polymerase, will certainly increase the complexity of the synthetic genome, which is not desired in the context of a minimal cell. RNA polymerases from bacteriophages SP6 and T3 can also be supplemented in the purified form or co-expressed in PURE system, bringing more diversity in promoter sequences [28, 57].

The strain harboring the correctly assembled SynChr must be cultured for one or several days to obtain sufficient cells for chromosome isolation. It is therefore important that the chromosome remains stable during propagation and does not undergo undesired rearrangements due to the presence of PURE repeats. We demonstrated that our PURE SynChr was stably maintained in *S. cerevisiae* during 88 h, which corresponds to approximately 32 generations. Considering that SynChr isolation from yeast was carried out after approximately 13 generations, we can assume that the SynChr^{PURE} is faithful to the design before isolation. In the event that increasing the number of PURE repeats when scaling up the number of genes leads to chromosome instability, we propose using a conditional recombination-deficient strain [58]. Using this strategy, chromosome assembly will be conducted through homologous recombination, which will be later switched off during strain propagation. If a DSB occurs in a repeated sequence during propagation, it will not be repaired, resulting in cell death provided a selection marker is introduced.

SynChr^{PURE} extracted from yeast can directly serve as template for cell-free protein synthesis in PURE system (Figure 3.7C). However, the low quantity of isolated DNA limits the amount of produced protein. A possible solution could be to increase the copy number of SynChr in yeast through replacement of the CEN6/ARS4 cassette by the replication origin of the native 2µ plasmid of *S. cerevisiae*. However, our preliminary data regarding assembly of the SynChr with a 2µ replication origin do not seem promising, as we did not obtain any correct clones (Figure S3.2). Limiting factors in this strategy could be the DNA size limit for replication by the 2µ system (the size of the native 2µ plasmid is ca. 6.3 kb [59]) and the possibility of multiple SynChr configurations being

present in a single yeast cell, thereby complicating screening and isolation. Another method to generate larger amounts of purified template for expression in PURE system is the introduction of an amplification step in *E. coli* of the SynChr extracted from yeast. This requires the presence of an *E. coli* backbone that can tolerate large inserts and that maintains the SynChr at a low copy number to avoid recombination in *E. coli* (i.e., a Bacterial Artificial Chromosome (BAC) backbone). Due to unpredictable toxicity of heterologous sequences in *E. coli* [16], this strategy may not be universally suitable for all SynChr designs.

In conclusion, we established a pipeline for the *de novo* design of synthetic chromosomes, their assembly in *S. cerevisiae*, screening of transformants, chromosome extraction, and cell-free expression in PURE system. The construction of complex synthetic genomes encoding functions for artificial cells will require the development of standardized, robust pipelines with automated workflows to accelerate the success of assembly.

Methods

Strains and culture conditions

Saccharomyces cerevisiae strains used and constructed in this study (Table S3.1) are derived from the CEN.PK lineage [60]. Strains were propagated in complex medium (Yeast extract Peptone Dextrose, YPD) containing 10 g L^{-1} Bacto yeast extract (Gibco, Thermo Fisher Scientific, Waltham, MA, USA), 20 g L^{-1} Bacto peptone (Gibco) and 20 g L^{-1} glucose or in synthetic medium (Synthetic Medium Dextrose, SMD) containing 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} $MgSO_4 \cdot 7H_2O$, 5 g L^{-1} $(NH_4)_2SO_4$, 20 g L^{-1} glucose, trace elements and vitamins [61]. Both media were initially prepared without glucose, set to pH 6.0 using 2 M KOH and sterilized (20 min at 110 °C for YP, 20 min at 121 °C for SM). Glucose was sterilized separately at 110 °C for 20 min before addition to the sterilized media. Filter-sterilized vitamins were added to SMD after media sterilization. For selection based on the dominant marker *hphNT*1, YPD was supplemented with hygromycin B (Hyg) to 200 mg L^{-1}. For selection based on auxotrophic markers, SMD was supplemented with separately sterilized solutions of uracil (Ura) to 150 mg L^{-1}, histidine (His) to 125 mg L^{-1} and/or leucine (Leu) to 500 mg L^{-1}. Liquid yeast cultures were grown aerobically at 30 °C and 200 rpm in an Innova incubator shaker (New Brunswick Scientific, Edison, NJ, USA).

Escherichia coli strains XL1-Blue (Agilent, Santa Clara, CA, USA), TOP10 (Thermo Fisher Scientific) or DH5-α (New England Biolabs, Ipswich, MA, USA) were used for molecular cloning and plasmid propagation. *E. coli* was grown in Lysogeny Broth (LB) medium containing 5 g L^{-1} Bacto yeast extract (Gibco), 10 g L^{-1} Bacto tryptone (Gibco) and 5 g L^{-1} NaCl, supplemented with 50–100 mg L^{-1} ampicillin (Amp), 50 mg L^{-1} kanamycin (Kan) or 25 mg L^{-1} chloramphenicol (Cam) if necessary. Cultivation was performed at 37 °C and 250 rpm in an Innova 4000 shaker (New Brunswick Scientific), unless stated otherwise.

Solid media were prepared by adding 20 g L^{-1} Bacto agar (Gibco) to the medium prior to heat sterilization. All *S. cerevisiae* and *E. coli* strains were stored at −80 °C in aliquots of overnight grown culture supplemented with sterile 30% (by volume) glycerol.

PCR

Amplification of *Arabidopsis thaliana* genomic DNA (gDNA) fragments was performed using the LongRange PCR Kit (Qiagen, Venlo, The Netherlands) according to the manufacturer's instructions. All other DNA fragments used for cloning and expression were obtained using the Phusion High-Fidelity DNA Polymerase (Thermo Fisher Scientific) according to the manufacturer's instructions, with a reduced primer concentration of 0.2 µM (except for cloning of G131, where a primer concentration of 0.5 µM was used). If amplification was unsuccessful, final primer concentration was reduced further to 20 nM, 5% DMSO was added and initial denaturation was prolonged to 3 min to reduce primer dimer formation. Diagnostic colony PCR was performed using the DreamTaq PCR Master Mix (Thermo Fisher Scientific) according to the manufacturer's protocol, downscaled to 10 µL reaction volume, with 10 min initial incubation at 95 °C for cell lysis and DNA release.

PAGE-purified primers (Sigma-Aldrich, Merck, Darmstadt, Germany) were used to generate fragments for assembly in yeast. All other primers were desalted or HPLC-purified and purchased from Sigma-Aldrich (IMB primers), Ella Biotech (Fürstenfeldbruck, Germany) (ChD primers) or Biolegio (Nijmegen, The Netherlands) (ChD primers). All primers are listed in Tables S3.8–S3.15.

PCR fragment size was validated by agarose gel electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer. PCR products used for cloning were treated with DpnI (Thermo Fisher Scientific or New England Biolabs) and purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific), the ZymoClean Gel DNA Recovery kit (Zymo Research, Irvine, CA, USA) or the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA) according to manufacturer's instructions.

In vitro plasmid assembly

Gibson assembly was performed using the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs) according to the manufacturer's protocol, downscaled to 5 µL and with an increased incubation time of 60 min. Golden Gate assembly was performed according to the protocol described in [62] with adjustments. For level 0 assemblies, insert concentration was increased to 40–150 fmol at an entry vector concentration of 20 fmol, and 30 cycles of digestion and ligation were performed. For level 1 assemblies, digestion in each cycle was carried out at 37 °C for 3 min. For all other Golden Gate reactions, the reaction volume was downscaled to 5 µL and all digestion steps were performed at 37 °C.

E. coli transformation

5 µL of Gibson or Golden Gate reaction mixture was mixed with 50 µL XL1-Blue chemically competent *E. coli* cells (Agilent), prepared in-house according to the protocol described in [63]. For G131 cloning, 1 µL of PCR product was mixed with 50 µL One Shot TOP10 chemically competent *E. coli* (Thermo Fisher Scientific). Cells were incubated for 5 min on ice before transformation via heat shock for 45 s at 42 °C, followed by incubation on ice for 2 min and resuspension in 450 µL Super Optimal broth with Catabolic repression (SOC), containing 5 g L^{-1} Bacto yeast extract (Gibco), 20 g L^{-1} Bacto tryptone

(Gibco), 0.58 g L^{-1} NaCl, 0.19 g L^{-1} KCl, 2 g L^{-1} $MgCl_2 \cdot 6H_2O$, 2.46 g L^{-1} $MgSO_4 \cdot 7H_2O$ and 3.6 g L^{-1} glucose. After incubation for 1 h at 37 °C and 250 rpm, cells were plated on LB agar plates with appropriate antibiotics and grown overnight at 37 °C.

Plasmid and SynChr isolation from *E. coli* and verification

Plasmid isolation from *E. coli* was performed using the PureYield Plasmid Miniprep System (Promega) or the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific). Elution was done in 25 µL water.

When necessary, restriction analysis of plasmids was carried out using FastDigest enzymes (Thermo Fisher Scientific) according to the manufacturer's protocol, scaled down to 10 µL and with a prolonged incubation time of 30 min. Restriction products were visualized using agarose gel electrophoresis on a 1% (w/v) agarose gel in 1× TAE buffer.

Plasmids and SynChrs isolated from *E. coli* were sequence-verified using Sanger sequencing (Macrogen Europe, Amsterdam, The Netherlands) or Oxford Nanopore technology (Plasmidsaurus, Eugene, OR, USA).

S. cerevisiae transformation

S. cerevisiae strain CEN.PK102-12A was used for transformation with the high-efficiency lithium acetate/single-stranded carrier DNA/polyethylene glycol method [64] with adaptations. After washing the cells in 25 mL water, resuspension was done in 1 mL 0.1 M lithium acetate (LiAc), and after removal of the supernatant, the cell pellet was resuspended in 0.1 M LiAc to a total volume of 500 µL resuspended cells. Per transformation, 50 µL of resuspended cells were pelleted and the transformation mix was added after removal of the supernatant. An adapted composition of the transformation mix was used (240 µL 50% (w/v) PEG 3350, 36 µL 1M LiAc, 25 µL 2 mg mL^{-1} single-stranded carrier DNA and 50 µL DNA in water), an extra 30 min incubation at 30 °C was performed before heat shock at 42 °C for 30 min, and the cells were incubated for 1–2 h at 30 °C in 1 mL YPD before plating on YPD + Hyg agar plates.

S. cerevisiae gDNA isolation

Genomic DNA from *S. cerevisiae* was extracted for whole-genome sequencing using the Qiagen Blood & Cell Culture Kit with 100/G Genomic-tips (Qiagen) according to the manufacturer's instructions for yeast.

DNA analysis

DNA purity was determined using the NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific) and DNA concentration was measured with NanoDrop or Qubit 2.0 Fluorometer (Invitrogen, Thermo Fisher Scientific) using the Qubit dsDNA Broad Range Assay kit (Invitrogen). Samples containing plasmids or SynChrs isolated from yeast were additionally analyzed by quantitative PCR (qPCR) on the QuantStudio 5 Real-Time PCR system (Applied Biosystems, Thermo Fisher Scientific) using the PowerUp SYBR Green Master Mix (Applied Biosystems), according to the supplier's

instructions. Data were analyzed with QuantStudio Design & Analysis software v1.5.1 (Applied Biosystems). Plasmids purified from *E. coli* and quantified using Qubit were used to prepare a standard curve of known concentrations ranging from 100 fM to 1 nM (Table S3.6). Primers used for qPCR are listed in Table S3.14.

Plasmid construction

All plasmids used in this study are listed in Tables S3.4–S3.7.

PURE cassette plasmids

Plasmids consisting of non-coding DNA fragments originating from *A. thaliana* flanked by PURE repeats (Table S3.4) were assembled via Golden Gate according to the yeast toolkit principle [62]. Part plasmids (level 0) were constructed by PCR amplification of the target region with primers containing part type-specific overhangs and assembled into the entry vector pYTK001 (*gfp* dropout) by Golden Gate cloning. Twelve part plasmids were assembled: a T7 promoter-lacO-g10L RBS-T7 tag sequence (*pT7*, amplified from G149) as a type 2 part plasmid, a *yfp* gene (amplified from G365) and nine 5-kb *A. thaliana* chunks (C2–C10) as type 3 part plasmids, and a T7 terminator (*tT7*, amplified from G131) as a type 4 part plasmid. All primers used for PCR amplification to create PURE cassette plasmids are listed in Table S3.8. To construct the *A. thaliana* chunk part plasmids, gDNA from *A. thaliana* ecotype Columbia (Col-0) was donated by Emma Barahona and Alvaro Eseverri (Center for Plant Biotechnology and Genomics, Madrid, Spain). Nine 5-kb fragments ("chunks") that did not contain any BsmBI and BsaI sites were selected from the *A. thaliana* Col-0 reference genome (TAIR10) *in silico*. DNA chunks amplified from *A. thaliana* gDNA were obtained using the LongRange PCR kit in two sequential PCR reactions to ensure sufficient DNA yield. 2 µL of the product of the first PCR reaction with external primers was used as template for the second PCR with internal primers, which resulted in 5-kb fragments flanked with type 3 overhangs for Golden Gate cloning. For chunk 5, only the PCR with internal primers was required. *E. coli* XL1-Blue was transformed with Golden Gate part plasmid assemblies and transformants were plated on LB + Cam agar plates. For each assembly, two to eight white colonies (indicating absence of the *gfp* gene) were verified by colony PCR using primer pair 14036/19265 (Table S3.11). Level 0 plasmids containing C10 and *tT7* (pGGKp363 and pGGKp348, respectively, Table S3.4) were verified by long-read whole-plasmid sequencing using Nanopore technology by Plasmidsaurus (Eugene, OR, USA).

For the assembly of *pT7*-chunk-*tT7* cassette plasmids (level 1), six level 0 part plasmids were combined in a BsaI Golden Gate assembly reaction: (i) pYTK002 (LS connector), (ii) pGGKp346 (*pT7*), (iii) pGGKp347 (*yfp*), pGGKp350 (chunk C5), pGGKp357 (chunk C2), pGGKp358 (chunk C3) or pGGKp361–63 (chunk C7, C9 and C10), (iv) pGGKp348 (*tT7*), (v) pYTK067 (R1 connector), and (vi) pYTK095 (*gfp* dropout). *E. coli* XL1-Blue was transformed with the resulting assemblies and plated on LB + Amp agar plates. Correct assemblies were verified by colony PCR of two to eight white colonies using primer pair 14776/19088 (Table S3.11). The start and end of the cassettes were confirmed by Sanger sequencing using primers 10320 and 10325 (Table S3.12) to ensure that all cassettes contained the same PURE repeats.

Due to difficulties in the level 1 Golden Gate cloning of chunks C4, C6 and C8, cas-

sette plasmids with these chunks were constructed via Gibson assembly. pUD1251 excluding the *yfp* gene was PCR amplified using primer pair 19525/19526. The chunks were amplified from their corresponding level 0 plasmids with primers containing 20-bp homology flanks to the pUD1251 backbone (Table S3.8). After DpnI digestion and purification using the GeneJET PCR Purification kit, Gibson assembly was carried out as described earlier. *E. coli* XL1-Blue was transformed with the plasmids and plated on LB + Amp agar plates. Assemblies were verified via diagnostic colony PCR using primer pair 14776/19088 (Table S3.11) and via Sanger sequencing with primers 10320 and 10325 (Table S3.12).

Level 1 plasmids containing C2–C10 (Table S3.4) were verified by long-read whole-plasmid sequencing using Nanopore technology by Plasmidsaurus.

crt expression cassette plasmids

Plasmids containing the β-carotene biosynthesis genes under control of a *S. cerevisiae* promoter and terminator, namely pUD1248 (*pPGK1-crtYB-tPGK*), pUD1249 (*pHHF2-crtE-tADH1)* and pUD1250 (*pTDH3-crtI-tTDH1*), were constructed via Gibson assembly (Table S3.5). Individual parts were amplified using Phusion PCR with primers containing 20-bp overlapping overhangs for assembly. Template plasmids and primers for PCR amplification are listed in Table S3.5 and Table S3.9, respectively. Plasmid pUDE269 was used as template for amplification of *crtYB*, *crtI*, and *crtE*. As pUDE269 contains these genes in a polycistronic cassette, PCR primers were designed to add the necessary start and stop codons: a stop codon for *crtYB* and *crtI* and a start codon for *crtI* and *crtE*. The PCR products were DpnI-digested, purified and assembled in a Gibson assembly reaction. *E. coli* XL1-Blue was transformed with the assemblies and plated onto LB + Cam agar plates. Assembled plasmids were verified via restriction analysis using DraI and Sanger sequencing. Sequencing primers are listed in Table S3.12.

Other plasmids

Plasmid G131 (Table S3.4) was constructed by site-directed mutagenesis PCR on plasmid G146 to remove the DraI restriction site, using primer pair 640/641 (Table S3.10) and subsequent circularization through homologous recombination in *E. coli* TOP10. The plasmid was verified by restriction analysis using EcoRI, XbaI and DraI and by Sanger sequencing with primers 106, 107, 337, 338, 339 and 340 (Table S3.12).

Plasmids pUD1226, pUDE1110, pUDE1039 and pUDE1217 (Table S3.5) were constructed via Golden Gate assembly and transformation into *E. coli* XL1-Blue. pUD1226 was constructed with plasmids pYTK013 (*pTEF1*), pGGKp304 (*ymNeongreen*) and pYTK054 (*tPGK1*) in the pGGKd015 backbone. pUDE1110 was constructed with plasmids pYTK009 (*pTDH3*), pGGKp303 (*ymTurquoise2*) and pYTK053 (*tADH1*) in the pGGKd034 backbone. Plasmids pUD1226 and pUDE1110 were checked by colony PCR using primers 10320 & 10325 (Table S3.11). pUDE1039 was constructed with plasmids pYTK018 (*pALD6*), pGGKp302 (*ymScarletI*) and pYTK051 (*tENO1*) in the pGGKd017 backbone and verified by colony PCR using primer pairs 10320/10335 and 10320/2442 (Table S3.11). pUDE1217 was constructed from four PCR products containing BsmBI restriction sites, amplified from pUD1226, pUDE1110, pUDE1039 and pGGKd71 with primer pairs 19059/19060, 19061/19062, 19063/19064 and 19065/19066, respectively

(Table S3.10). pUDE1217 was verified by restriction analysis using NcoI and SspI and Sanger sequencing with primers 865, 1779 and 10901 (Table S3.12).

Assembly of SynChrs in yeast

The synthetic chromosomes constructed in this study (Table S3.2) were assembled in *S. cerevisiae* CEN.PK102-12A (Table S3.1) from 20 DNA fragments (Table S3.17): a *yfp* cassette for expression in PURE system (PF1), nine 5-kb *A. thaliana* chunks with or without PURE repeats (PF2–10 and CF2–10), a yeast replication origin (CEN6/ARS4 or 2μ) combined with an *E. coli* replication origin (ColE1) and antibiotic marker (*bla*), and nine yeast markers (*mRuby2*, *mTurquoise2*, *HIS3*, *URA3*, *LEU2*, *crtE*, *crtI*, *crtYB* and *hphNT1*). Chunk 2 harbored an I-SceI restriction site, allowing the linearization for screening purposes. PCR primers for amplification of these fragments (Table S3.13) contained 60-bp SHR overhangs (Table S3.16) for assembly by homologous recombination. DNA fragments were pooled into a mix containing 100 fmol of the *hphNT1* and yeast replication origin fragments and 150 fmol of the other 18 DNA fragments. After transformation, cells were plated on YPD + Hyg plates and incubated for three days at 30 °C. Since the *hphNT1* fragment, conferring resistance to hygromycin B, is located on the SynChr opposite of the yeast replication origin, plating on YPD + Hyg plates will select for transformants which have taken up those two fragments and any DNA that can connect them. Transformants were screened based on β-carotene production by visual classification of orange and white colonies, fluorescence by flow cytometry, auxotrophy by restreaking on dropout media and were sequenced by long-read sequencing. Positively and negatively screened strains were stocked for SynChr^{control} (IMF49) and SynChr^{PURE} (IMF50–IMF64) (Table S3.3).

Assembly screening

Fluorescence detection by flow cytometry

Colonies containing SynChrs assembled in yeast were checked for *mTurquoise2* and *mRuby2* expression by flow cytometry using the BD FACSAria II Cell Sorter with the BD FACSDiva software (BD Biosciences, Franklin Lakes, NJ, USA). A Cytometer Setup and Tracking (CS&T) cycle was run before every experiment according to the manufacturer's instructions. Each colony was resuspended in 500 µL Isoton II Diluent (Beckman Coulter, Woerden, The Netherlands), analyzed with the 70 µm nozzle and at least 30,000 events were recorded. mTurquoise2 and mRuby2 fluorescence were measured using a 445 nm excitation laser with a 525/50 nm bandpass emission filter or a 561 nm laser with a 582/15 nm bandpass filter, respectively. Data were analyzed using the Cytobank software (Beckman Coulter). Gates to determine the presence of fluorescence were based on the negative control strain CEN.PK113-7D and positive control IMF2 (Table S3.1).

Auxotrophy detection by restreaking on dropout media

The presence of the *URA3*, *LEU2* and *HIS3* fragments in the assembled SynChrs was checked by colony restreaking on dropout media. Colonies were picked from YPD + Hyg agar plates and restreaked on SMD agar plates and YPD + Hyg agar plates. Plates were

incubated for two days at 30 °C. In the absence of growth on SMD, a second restreak was performed from the YPD + Hyg plate onto three SMD agar plates: (i) SMD + uracil + histidine, (ii) SMD + uracil + leucine, (iii) SMD + histidine + leucine.

Long-read sequencing

To determine the sequence of the SynChrs assembled in yeast, whole genome long-read sequencing was performed in-house using MinION technology (Oxford Nanopore Technologies, Oxford, UK). Genomic DNA was prepared and checked for purity and quantity using NanoDrop and Qubit as described in previous sections. DNA length and integrity were analyzed using a genomic DNA ScreenTape assay (Agilent) in the TapeStation 2200 (Agilent). Libraries of two to four barcoded samples were prepared with the ligation sequencing kit SQK-LSK109 and expansion kit EXP-NBD104 (Oxford Nanopore Technologies) according to the manufacturer's recommendations. Three FLO-MIN106 (R9.4.1) flow cells were used to sequence the following strains (Table S3.3): (i) IMF49–IMF52, (ii) IMF53–IMF56, (iii) IMF59, IMF61, IMF62 and IMF64. For IMF53–IMF56, additional I-SceI digestion (Thermo Fisher Scientific) to linearize the SynChrs was performed according to the manufacturer's instructions and the DNA was cleaned-up using AMPure XP beads (Beckman Coulter). Due to low sequencing read depth, strains IMF51 and IMF54 were additionally sequenced together on a FLO-MIN106 (R9.4.1) flow cell with adaptations to the library preparation to improve read length: fresh genomic DNA (five days old) was used, ligation and end repair times were extended to 10 min, adapter ligation was extended to 30 min, incubation time with magnetic beads at each step was extended to 10 min and elution from the beads was performed at 37 °C for 10 min. To increase output for IMF51 and IMF54, after loading one-third of the library and sequencing for 21 h, the flow cell was washed with kit EXP-WSH004 according to the manufacturer's instructions and the other two-third of the library was loaded and sequenced for 45.5 h.

Guppy (GPU v6.3.4, Oxford Nanopore Technologies) was used for basecalling and demultiplexing. The resulting FASTQ files were filtered on length (> 200 bp for the first three sequencing runs, > 1 kb for the second run of IMF51 and IMF54) and mapped using minimap2 (parameters: *-ax map-ont*) [65] to a reference FASTA file containing the sequence of CENPK113-7D [66] concatenated with the designed SynChr sequence. The mapped reads were sorted and indexed using SAMTools [67] and visualized with the Integrated Genomics Viewer (IGV, version 2.11.9) [68]. An annotation (.gff) file of the designed SynChr was created by exporting the annotations from SnapGene (version 7.0.2, GSL Biotech, San Diego, CA, USA) and using the R package labtools (version 0.1.0, function: *make_gff_from_snap*) [69] in R Statistical Software (version 4.1.2) [70]. *De novo* assembly of the reads was performed using Flye version 2.7.1-b1673 (parameters: *--plasmids –nano-raw*) [71] or Canu version 2.2 (parameters: *genomeSize=12m, minReadLength=200, minOverlapLength=50, -nanopore-raw*) [72] and the assembled contig corresponding to the SynChr was identified with the MUMmer package [73], using NUCmer (parameter: *-maxmatch*) and delta-filter (parameter: *-q*). If necessary, a consensus SynChr sequence was assembled in SnapGene using information from the Flye and Canu assemblies and raw reads.

SynChr stability

A single colony of IMF49 or IMF54 was used to inoculate three individual 50-mL centrifuge tubes containing 20 mL SMD or SMD supplemented with uracil and histidine. Cultures were grown overnight at 30 °C while shaking. For four days, the cells were kept at exponential phase by transferring to fresh medium every day. Cultures after 44 and 88 h were subjected to flow cytometry as described in the section "Assembly screening". Additionally, 1 µL of culture was diluted in 999 µL of sterile water and 50 µL of this dilution was plated on SMD or SMD + uracil + histidine agar plates. Plates were incubated at 30 °C for two days.

SynChr isolation from yeast for cell-free expression

Two isolations of SynChr^{PURE} were performed (isolation 1 without replicates, isolation 2 with technical duplicates), which differed in starting yeast cell count. IMF54 was grown from glycerol stock in SMD until a final ODV (ODV = OD_{660} × Volume [mL]) of 1070 for isolation 1 and ODV of 9000 for isolation 2 (4500 per replicate). For isolation 1, cells were harvested by centrifugation at 4000 ×*g* for 5 min at 4 °C and the cell pellet was washed by resuspension in 4 mL TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). Washed cells were harvested by centrifugation at 4000 ×*g* for 5 min at 4 °C and resuspended in 4 mL Buffer Y1 (1 M sorbitol, 100 mM EDTA, 14 mM β-mercaptoethanol, pH 8.0, prepared in house or taken from Genomic DNA Buffer Set, Qiagen). To digest the cell walls and obtain spheroplasts, 12.5 mg Zymolyase 20T (AMSBIO, Alkmaar, The Netherlands), dissolved in 250 µL Milli-Q, was added and mixed with the cells through inversion. For isolation 2, the isolation of the SynChrs was performed in duplicate, starting from the same cell culture. Cells of both replicates were harvested together by centrifugation and washed in 33.6 mL TE buffer, harvested again and resuspended in 33.6 mL Buffer Y1, after which the cells were divided into two tubes to do the SynChr isolation in duplicate. To each duplicate, 52.6 mg Zymolyase 20T resuspended in 1051 µL Milli-Q was added.

The cells were incubated in a water bath at 30 °C for at least 1 h with regular mixing. Efficiency of spheroplasting was measured by diluting 5 µL of cells before and after spheroplasting in 995 µL demiwater and measuring the OD_{660}. The low osmotic pressure of water will result in bursting of the spheroplasts. Once the OD_{660} was reduced by more than 80%, the spheroplasts were placed on ice and harvested by centrifugation at 4000 ×*g* for 15 min at 4 °C. The spheroplast pellet was resuspended in 1 mL 1 M sorbitol (isolation 1) or 2 mL 1 M sorbitol (isolation 2). SynChr^{PURE} was isolated from the spheroplasts using the Qiagen Large-Construct Kit following the manufacturer's instructions, starting from the addition of 19 mL (isolation 1) or 18 mL (isolation 2) Buffer P1 to have a total volume of 20 mL resuspended spheroplasts. The following adjustments were made to the manufacturer's instructions: After elution in 15 mL Buffer QF, the eluate was stored overnight at 4 °C and the next morning, the DNA was precipitated with isopropanol in the presence of 15 µL 15 mg mL^{-1} GlycoBlue Coprecipitant (Invitrogen) and 1.2 mL 3M sodium acetate (pH 5.2, Sigma-Aldrich) to improve visibility of the DNA pellet. The final isolated SynChr^{PURE} DNA was dissolved in 100 µL 10 mM Trizma hydrochloride solution (pH 8.5, Sigma-Aldrich).

SynChr^{PURE} isolation from IMF54 was attempted with three other protocols, which failed to result in *yfp* expression in PURE system. The protocols are described below.

(i) Cells with an ODV of 1912 (956 per replicate) were harvested together for both replicates through centrifugation at 4000 ×*g* for 5 min at 4 °C, resuspended in 20 mL 0.1 M Tris-HCl buffer (pH 8.0, Thermo Scientific, Thermo Fisher Scientific) and centrifuged again at 4000 ×*g* for 5 min at 4 °C. The cell pellet was resuspended in 2185 µL Tris-DTT solution (0.1 M Tris-HCl, 10 mM dithiothreitol (Roche, Sigma-Aldrich), prepared fresh before use) and divided into two tubes to do the SynChr isolation in duplicate. RNase A solution (20 mg mL^{-1}, Sigma-Aldrich) was added into a final concentration of 100 µg mL^{-1} and the cells were incubated for 10 min at 30 °C in a shaking incubator at 200 rpm. Zymolyase 20T (13.65 mg) was dissolved in 136.5 µL Tris-DTT solution and added to the cells, followed by incubation for 15 min at 30 °C in a shaking incubator at 200 rpm. Efficiency of spheroplasting was determined as described above and the spheroplasts were placed on ice once the OD_{660} was reduced by more than 80%. Spheroplasts were not harvested through centrifugation, but directly divided over multiple tubes (250 µL per tube). The SynChr^{PURE} was isolated using the GeneJET miniprep kit (Thermo Fisher Scientific) starting from the addition of 250 µL lysis buffer, following the manufacturer's instructions with adaptations. After centrifugation of the cell debris and chromosomal DNA, the supernatant was not clear and was therefore transferred to a new tube and centrifuged again for 5 min at 15000 ×*g*, after which the resulting supernatant was loaded onto the column. At this step, the supernatants from all tubes were combined onto one column per replicate. SynChr^{PURE} was eluted from the column by addition of 25 µL prewarmed Milli-Q (55 °C) and incubation for 10 min at RT before 2 min centrifugation.

(ii) An adaptation of protocol (i) was performed, where lysis of spheroplasts by osmotic shock was avoided through changing the spheroplasting buffer from Tris-DTT solution to Buffer Y1. The SynChr^{PURE} was isolated from the intact spheroplasts using the GeneJET miniprep kit, starting from addition of resuspension buffer. Cells were harvested at an ODV of 2140 (1070 per duplicate) by centrifugation at 4000 ×*g* for 5 min at 4 °C and the cell pellet was washed by resuspension in 8 mL TE buffer. Washed cells were harvested by centrifugation at 4000 ×*g* for 5 min at 4 °C and resuspended in 8 mL Buffer Y1. Zymolyase 20T (25 mg dissolved in 500 µL Milli-Q) was added and mixed with the cells through inversion. The cells were incubated in a water bath at 30 °C for 1 h 25 min with regular mixing. Spheroplasts were placed on ice and harvested by centrifugation at 4000 ×*g* for 10 min at 4 °C. The spheroplast pellet was resuspended in resuspension buffer from the GeneJET miniprep kit to have a total volume of 2.5 mL and was divided over 10 tubes (250 µL per tube, 5 tubes per duplicate). The SynChr^{PURE} was isolated in duplicate from the resuspended spheroplasts using the GeneJET miniprep kit, following manufacturer's instructions from the addition of 250 µL lysis buffer to each tube. After centrifugation to pellet the cell debris and chromosomal DNA, each supernatant was transferred to a new tube and centrifuged again for 5 min at 15000 ×*g*. For each replicate, five supernatants were combined onto one column. SynChr^{PURE} was eluted from the column by addition of 25 µL prewarmed Milli-Q (55 °C) and incubation for 10 min at RT before 2 min centrifugation.

(iii) The published protocol by Noskov and colleagues [50] was attempted twice with technical duplicates.

Cell-free gene expression

In vitro transcription-translation was carried out using PURE*frex*2.0 (GeneFrontier Corporation, Kashiwa, Japan), following manufacturer's instructions for storage and handling. Reaction mixtures containing 5 µL solution I (buffer), 0.5 µL solution II (enzymes), 1 µL solution III (ribosomes), 5 U SUPERase·In (Thermo Fisher Scientific), DNA template and Milli-Q water were prepared to a final volume of 10 µL. Positive control templates for *yfp* expression were plasmids G76 (isolation 1) and pUD1251 (isolation 2), isolated from *E. coli* (Table S3.7), and PCR products of the *yfp* cassette (isolation 2 only), amplified using Phusion polymerase and primer pair 19274/19277 (Table S3.15) from the SynChr^{PURE} isolated from IMF54. Plasmid G28 (Table S3.7) was used to determine PURE expression inhibition by the SynChr^{PURE} sample (isolation 1 and 2). The *mCherry* gene encoded on G28 was expressed in PURE system in the presence of SynChr^{PURE} sample, water (positive control) or non-coding plasmid DNA (positive control). Plasmid G28 (0.5 µL, resulting in 1 nM final plasmid concentration), diluted in 2.75 µL SynChr^{PURE} isolated from IMF54, 2.75 µL water or 2.75 µL 30 ng µL^{-1} non-coding plasmid pUDC192, was added to the reaction mixture. Reactions were assembled in PCR tubes and transferred to a black 384-well flat µClear bottom microplate (Greiner Bio-One, Kremsmünster, Austria), which was sealed with a highly transparent film (Sarstedt, Nümbrecht, Germany). The plate was incubated at 37 °C for 16 h and fluorescence was measured at end-point in triplicates in an Infinite 200 Pro plate reader (Tecan, Männedorf, Switzerland) using bottom and top measurements (isolation 1) or a CLARIOstar plate reader (BMG LABTECH, Ortenberg, Germany) using bottom measurement (isolation 2). YFP fluorescence was measured with excitation at 488 nm (Infinite 200 Pro) or 497 nm with 15 nm bandwidth (CLARIOstar) and emission at 527 nm (Infinite 200 Pro) or 540 nm with 20 nm bandwidth (CLARIOstar). For mCherry, fluorescence was detected with excitation at 570 nm (Infinite 200 Pro) or 559 nm with 20 nm bandwidth (CLARIOstar) and emission at 620 nm (Infinite 200 Pro) or 630 nm with 40 nm bandwidth (CLARIOstar). For measurements in the Infinite 200 Pro, a gain of 80% was used. For measurements in the CLARIOstar, the focus and gain were adjusted to a well containing a PURE*frex*2.0 reaction solution with 70 pM pUD1251 (YFP) or 1 nM G28 (mCherry), that was pre-incubated for five hours at 37 °C in a thermocycler. The gains used in the CLARIOstar were 1852 (YFP) and 2090 (mCherry) and the focus height 3.1 mm.

Acknowledgements

We thank the following people for constructing and sharing plasmids: Quinte Smitskamp (pUDE1217), David Foschepoth (G131), Nicole Bennis (pUDE1110), Matic Kostanjšek (pUD1226) and Jasmijn Hassing (pUDE1039).

We thank Álvaro Eseverri and Emma Barahona (Center for Plant Biotechnology and Genomics, Spain) for providing genomic DNA of *Arabidopsis thaliana*, Marijke Luttik for technical assistance with FACS analysis and the development of the SynChr isolation protocol, Bastiaan van Dijk for help with developing SynChr isolation protocol, Clara Carqueija Cardoso for help with library preparation for long-read sequencing and Marcel van den Broek for help with sequencing data analysis.

I would like to thank Laura Sierra Heras for her excellent work on designing and con-

structing the SynChrs described in this chapter, as well as for her contributions to the writing. Her enthusiasm, perseverance and great laboratory work contributed greatly to the progress of this project.

Images in Figures 3.1, 3.3 and 3.6 were (partially) created with BioRender.com.

This project was financially supported by The Netherlands Organization for Scientific Research (NWO/OCW) via the “BaSyC – Building a Synthetic Cell” Gravitation Grant (024.003.019) and by Agence Nationale de la Recherche (ANR-22-CPJ2-0091-01).

Supplementary Figures

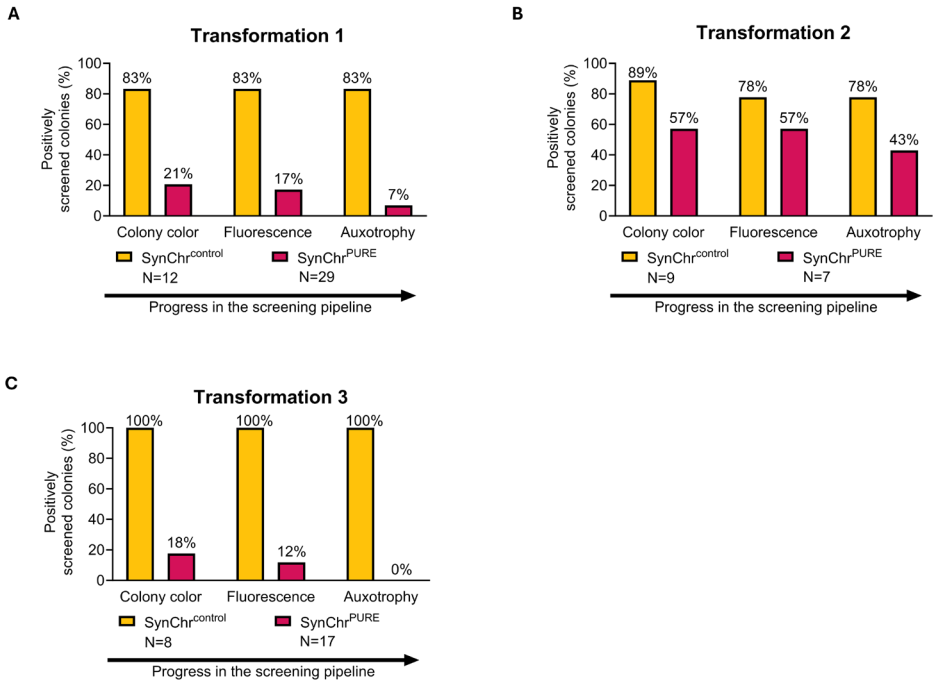


Figure S3.1: Percentage of positively screened colonies for three independent transformations. "Positive" means that the clone carries the markers of the inspected screening step and all previously screened markers. Colonies that grew on YPD with hygromycin plates were classified as positive when displaying an orange color, when showing fluorescence for both mTurquoise2 and mRuby2, and when growth on SMD was visible. **A)** Transformation 1. **B)** Transformation 2. **C)** Transformation 3.

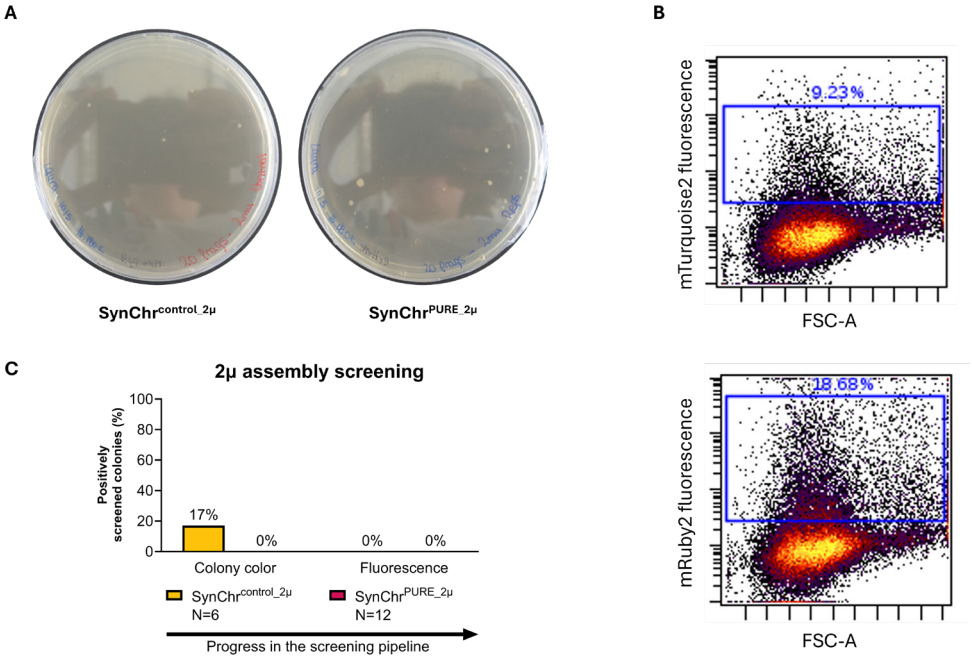


Figure S3.2: Screening for correctly assembled SynChrs with 2µ origin. A) Plates with YPD and hygromycin after transformation. SynChr$^{control_2\mu}$ (left) and SynChr$^{PURE_2\mu}$ (right). **B)** FACS screening for mTurquoise2 and mRuby2 fluorescence of the only orange colony for SynChr$^{control_2\mu}$. Events within the blue frame are classified as fluorescent. **C)** Percentage of positively screened colonies at each screening step, compared to step 1. "Positive" means that the clone carries the markers of the inspected screening step and all previously screened markers. Colonies that grew on YPD with hygromycin plates were classified as positive when displaying an orange color and when showing fluorescence for both mTurquoise2 and mRuby2. Data from one transformation are displayed.

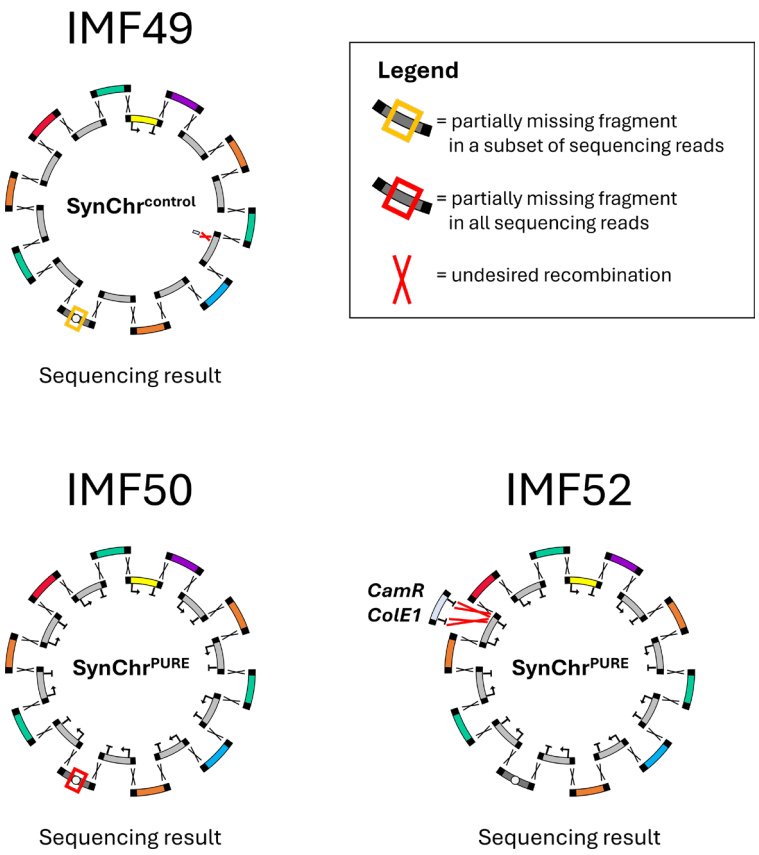


Figure S3.3: Sequenced SynChr configurations for strains IMF49, IMF50 and IMF52. The SynChr^{control} of IMF49 showed approximately half the sequencing depth for part of the CEN6/ARS4_bla_ColE1 fragment (1,891 bp) compared to the rest of the SynChr^{control} and has a 89-bp insert in CF4. IMF50 misses a 1,992-bp sequence in the middle of the fragment with CEN6/ARS4. IMF52 contains an insertion of a 1,780-bp sequence in the terminator region downstream of PF9, corresponding to a plasmid backbone (containing a chloramphenicol resistance gene and ColE1-type origin of replication), used as template during PCR.

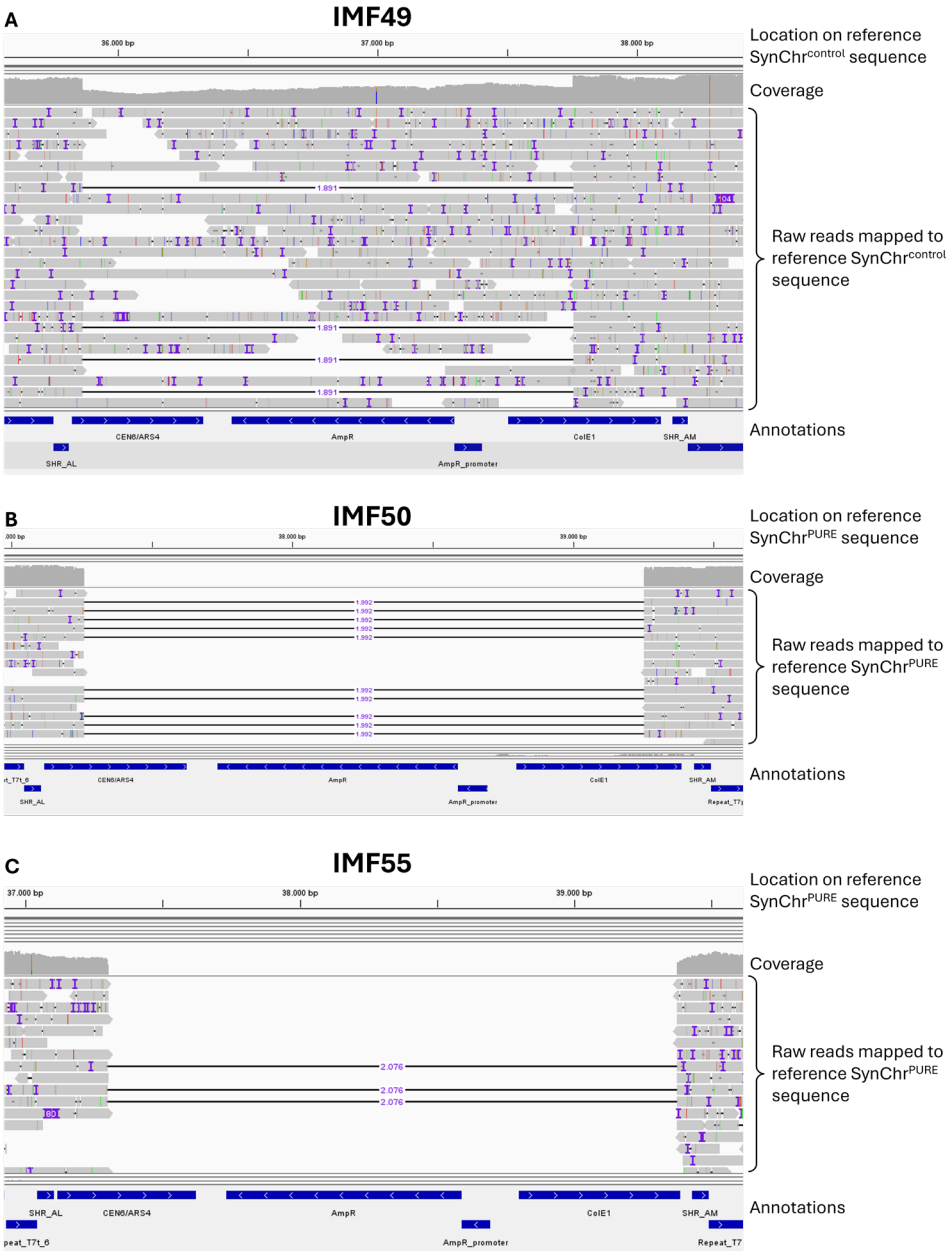
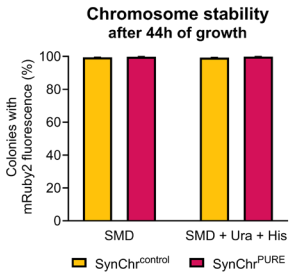
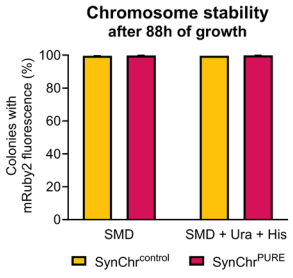


Figure S3.4: Raw reads mapped to the reference SynChr sequence, showing gaps in the CEN6/ARS4_bla_ColE1 fragment. A) SynChr^{control} from strain IMF49. A subset of the raw reads contains an 1891-bp gap from the beginning of CEN6/ARS4 until halfway ColE1. **B)** SynChr^{PURE} from strain IMF50. All raw reads contain a 1992-bp gap from approximately one-third of CEN6/ARS4 until approximately three-quarters of ColE1. **C)** SynChr^{PURE} from strain IMF55. All raw reads contain a 2076-bp gap from halfway CEN6/ARS4 until the end of ColE1. Annotation AmpR = *bla* gene.

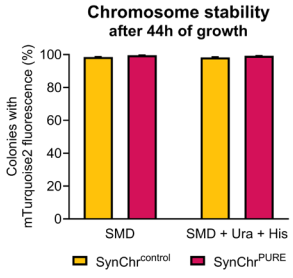
A



B



C



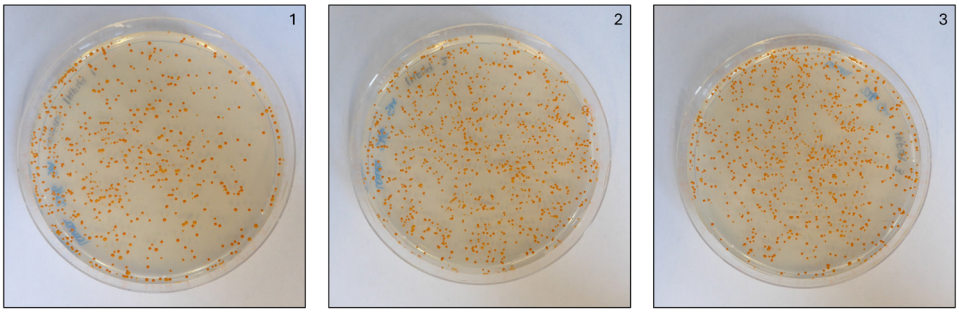
D

	Number of white colonies after 44 h of growth				Number of white colonies after 88 h of growth			
	SynChr^control		SynChr^PURE		SynChr^control		SynChr^PURE	
	SMD	SMD + Ura + His	SMD	SMD + Ura + His	SMD	SMD + Ura + His	SMD	SMD + Ura + His
Plate 1	0	0	1	0	0	0	1	1
Plate 2	0	0	0	1	0	0	0	0
Plate 3	0	0	0	1	3	0	0	1

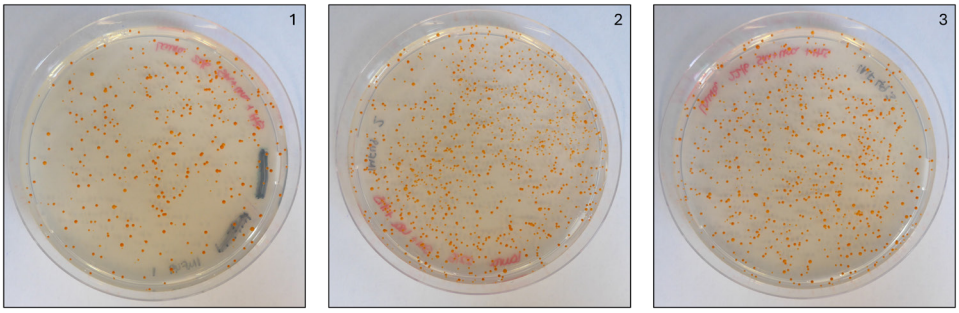
Figure continues on the next two pages.

E

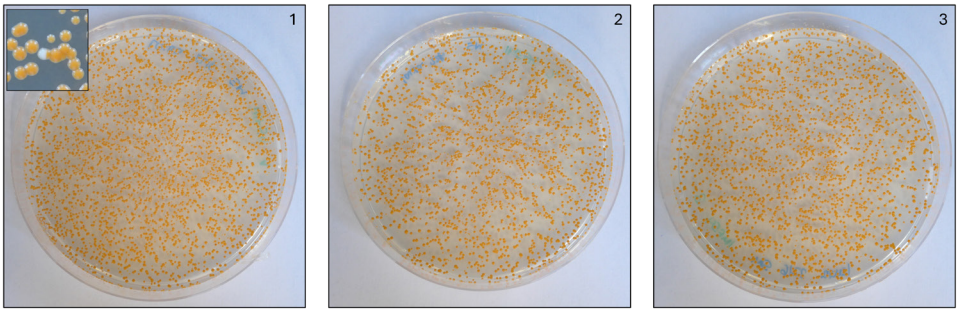
SynChr^{control}, SMD, 44h



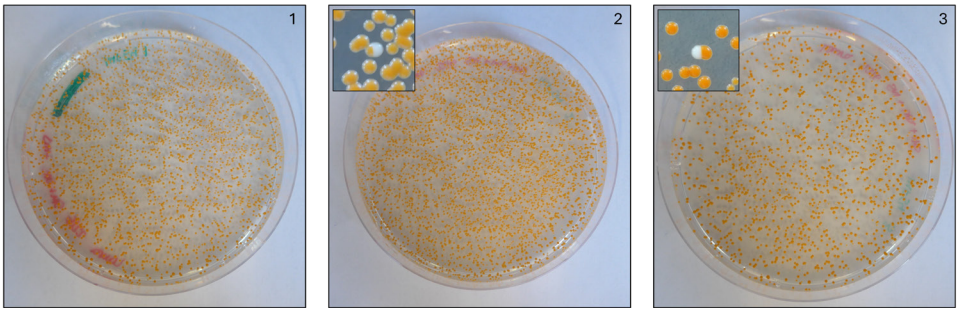
SynChr^{control}, SMD + Ura + His, 44h



SynChr^{PURE}, SMD, 44h

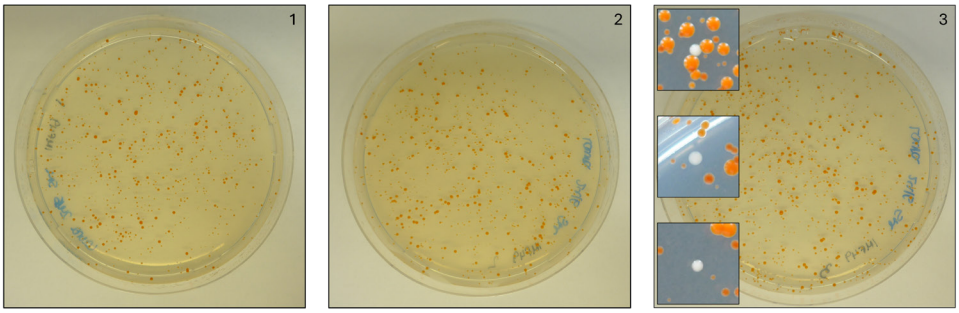


SynChr^{PURE}, SMD + Ura + His, 44h

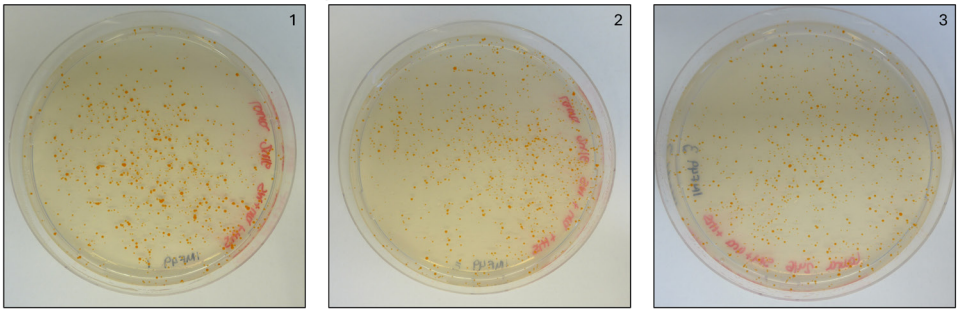


F

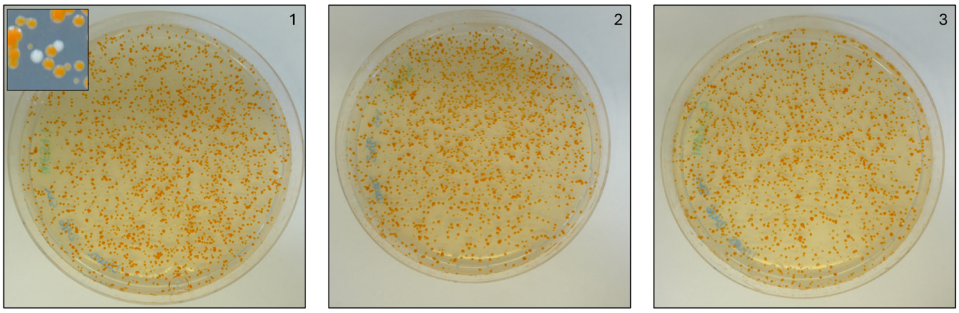
SynChr^{control}, SMD, 88h



SynChr^{control}, SMD + Ura + His, 88h



SynChr^{PURE}, SMD, 88h



SynChr^{PURE}, SMD + Ura + His, 88h

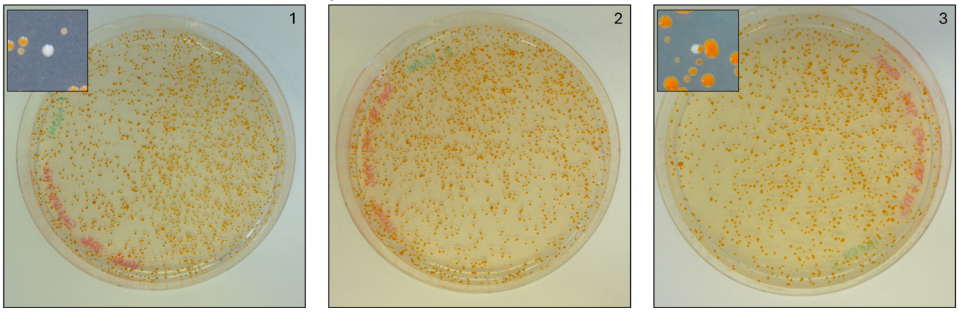


Figure S3.5: Stability of SynChr^{control} and SynChr^{PURE}. A) – C) Flow cytometry results for mRuby2 and mTurquoise2 fluorescence after 2 and 4 days (44 h and 88 h, respectively) of growth in culture, in both media conditions (SMD and SMD + Ura + His), for both strains IMF49 (SynChr^{control}) and IMF54 (SynChr^{PURE}). See Figure 3.6B for mTurquoise2 fluorescence after 88 h. **A)** mRuby2, 44 h. **B)** mRuby2, 88 h. **C)** mTurquoise2, 44 h. **D)** Number of white colonies per plate after 44 h or 88 h of growth in triplicates. **E)** Plates showing low incidence of white colonies after 44h of growth in both media conditions for both strains in triplicates. Insets show the only white colonies of each plate. **F)** Plates showing low incidence of white colonies after 88h of growth in both media conditions for both strains in triplicates. Insets show only white colonies of each plate.

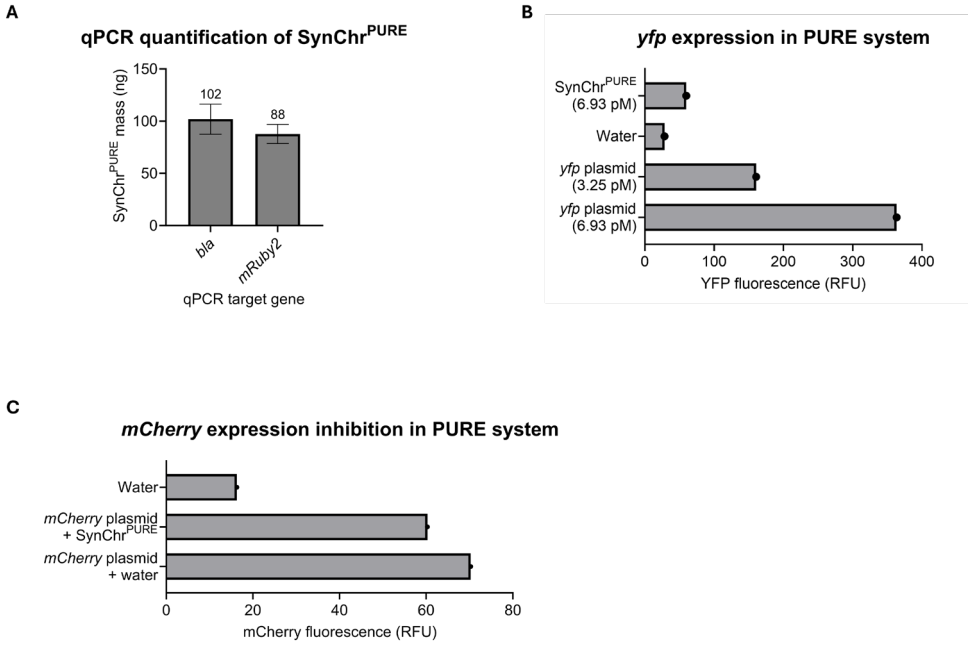


Figure S3.6: SynChr^{PURE} can be isolated from *S. cerevisiae* and shows expression of *yfp* in PURE system. Data from first isolation. **A)** qPCR quantification of SynChr^{PURE} mass (ng) in 100 µL, quantified by primer sets targeting *bla* and *mRuby2* genes. Mean and standard deviation from three technical replicates are plotted. **B)** YFP fluorescence measured after 16 h incubation of various DNA templates in PURE system. Individual data points represent averages of three fluorescence measurements. Final DNA template concentrations in the assembled PURE reaction are indicated between brackets. SynChr^{PURE} = SynChr^{PURE} isolated from *S. cerevisiae* strain IMF54. Water = No DNA template (negative control). *yfp* plasmid = Plasmid containing a *yfp* gene, isolated from *E. coli* (positive control). **C)** mCherry fluorescence measured after 16 h incubation of a diluted *mCherry* plasmid in PURE system. Individual data points represent averages of three fluorescence measurements. Water = No DNA template (negative control). *mCherry* plasmid + SynChr^{PURE} = 0.5 µL of *mCherry* plasmid (final concentration: 1 nM) isolated from *E. coli*, diluted in 2.75 µL SynChr^{PURE} sample. *mCherry* plasmid + water = 0.5 µL of *mCherry* plasmid (final concentration: 1 nM) isolated from *E. coli*, diluted in 2.75 µL water (positive control).

Supplementary Tables

Table S3.1: *S. cerevisiae* strains used in this study

Strain	Relevant genotype	Source
Transformation host for assembly in yeast		
CEN.PK102-12A	*MATa ura3-52 his3-1 leu2-3*	[60]
Negative and positive controls for flow cytometry		
CEN.PK113-7D	*MATa URA3 HIS3 LEU2*	[60]
IMF2	*MATa ura3-52 his3-1 leu2-3 NeoChr2(pCCW12-mRuby2- tENO1 pTEF2-mTurquoise2-tSSA1)*	[19]
Strains containing SynChrs assembled in yeast		
IMF49	*MATa ura3-52 his3-1 leu2-3 SynChr^{control}*	This study
IMF50–IMF64	*MATa ura3-52 his3-1 leu2-3 SynChr^{PURE}*	This study

Table S3.2: Configurations of synthetic chromosomes designed for assembly in yeast

Name	Size SynChr design[1]	Replication origin	Control / PURE	SynChr configuration[2,3,4]
SynChr^{control}	62970 bp	CEN6/ARS4	Control	$_{AA}$ *pT7-yfp-tT7* $_{AB}$ *pTEF-hphNT1-tTEF* $_{AC}$ C2 $_{AD}$ *pHHF2-crtE-tADH1* $_{AE}$ C3 $_{AW}$ *pURA3-URA3-tURA3* $_{AG}$ C4 $_{AH}$ *pTEF2-mTurquoise2-tSSA1* $_{AI}$ C5 $_{AJ}$ *pTDH3-crtI-tTDH* $_{AK}$ C6 $_{AL}$ CEN6/ARS4 *bla* ColE1 $_{AM}$ C7 $_{AN}$ *pLEU2-LEU2-tLEU2* $_{AO}$ C8 $_{AP}$ *pPGK1-crtYB-tPGK1* $_{CE}$ C9 $_{AR}$ *pCCW12-mRuby2- tENO1* $_{AS}$ C10 $_{AT}$ *pHIS3- HIS3-tHIS3*
SynChr^{PURE}	67426 bp	CEN6/ARS4	PURE	$_{AA}$ *pT7-yfp-tT7* $_{AB}$ *pTEF-hphNT1-tTEF* $_{AC}$ *pT7-*C2-*tT7* $_{AD}$ *pHHF2- crtE-tADH1* $_{AE}$ *pT7-*C3-*tT7* $_{AW}$ *pURA3-URA3-tURA3* $_{AG}$ *pT7-*C4-*tT7* $_{AH}$ *pTEF2-mTurquoise2-tSSA1* $_{AI}$ *pT7-*C5-*tT7* $_{AJ}$ *pTDH3-crtI-tTDH* $_{AK}$ *pT7-*C6-*tT7* $_{AL}$ CEN6/ARS4 *bla* ColE1 $_{AM}$ *pT7-*C7-*tT7* $_{AN}$ *pLEU2-LEU2-tLEU2* $_{AO}$ *pT7-*C8-*tT7* $_{AP}$ *pPGK1-crtYB-tPGK1* $_{CE}$ *pT7-*C9-*tT7* $_{AR}$ *pCCW12- mRuby2-tENO1* $_{AS}$ *pT7-*C10-*tT7* $_{AT}$ *pHIS3- HIS3-tHIS3*
SynChr$^{control_2\mu}$	63459 bp	2µ	Control	$_{AA}$ *pT7-yfp-tT7* $_{AB}$ *pTEF-hphNT1-tTEF* $_{AC}$ C2 $_{AD}$ *pHHF2-crtE-tADH1* $_{AE}$ C3 $_{AW}$ *pURA3-URA3-tURA3* $_{AG}$ C4 $_{AH}$ *pTEF2-mTurquoise2-tSSA1* $_{AI}$ C5 $_{AJ}$ *pTDH3-crtI-tTDH* $_{AK}$ C6 $_{AL}$ 2µ *bla* ColE1 $_{AM}$ C7 $_{AN}$ *pLEU2-LEU2-tLEU2* $_{AO}$ C8 $_{AP}$ *pPGK1-crtYB-tPGK1* $_{CE}$ C9 $_{AR}$ *pCCW12-mRuby2- tENO1* $_{AS}$ C10 $_{AT}$ *pHIS3-HIS3-tHIS3*
SynChr$^{PURE_2\mu}$	67915 bp	2µ	PURE	$_{AA}$ *pT7-yfp-tT7* $_{AB}$ *pTEF-hphNT1-tTEF* $_{AC}$ *pT7-*C2-*tT7* $_{AD}$ *pHHF2- crtE-tADH1* $_{AE}$ *pT7-*C3-*tT7* $_{AW}$ *pURA3-URA3-tURA3* $_{AG}$ *pT7-*C4-*tT7* $_{AH}$ *pTEF2-mTurquoise2-tSSA1* $_{AI}$ *pT7-*C5-*tT7* $_{AJ}$ *pTDH3-crtI-tTDH* $_{AK}$ *pT7-*C6-*tT7* $_{AL}$ 2µ *bla* ColE1 $_{AM}$ *pT7-*C7-*tT7* $_{AN}$ *pLEU2-LEU2-tLEU2* $_{AO}$ *pT7-*C8-*tT7* $_{AP}$ *pPGK1-crtYB-tPGK1* $_{CE}$ *pT7-*C9-*tT7* $_{AR}$ *pCCW12-mRuby2-tENO1* $_{AS}$ *pT7-*C10-*tT7* $_{AT}$ *pHIS3-HIS3-tHIS3*

[1] Designed SynChr size after correction for Chunk 10 (designed size: 5000 bp), based on sequencing of template plasmids pGGKp363 (3310 bp gap in Chunk 10) and pUD1279 (1174 bp gap in Chunk 10).
[2] SHRs are differently annotated than in [22] and are listed in Table S3.16. SHRs are annotated in subscript between the genetic fragments that they join together.
[3] *pT7 = T7 promoter - lacO - g10L RBS - T7 tag*
[4] *tT7 = T7 terminator*

Table S3.3: Stocked strains containing SynChrs assembled in yeast

Strain	SynChr	Orange / White	Absent markers[1]	Sequenced	Correct configuration
IMF49	SynChr^{control}	Orange	-	Yes	Yes[2]
IMF50	SynChr^{PURE}	Orange	-	Yes	Yes[3]
IMF51	SynChr^{PURE}	Orange	-	Yes	Yes
IMF52	SynChr^{PURE}	Orange	-	Yes	No
IMF53	SynChr^{PURE}	Orange	-	Yes	No
IMF54	SynChr^{PURE}	Orange	-	Yes	Yes
IMF55	SynChr^{PURE}	Orange	*mRuby2*	Yes	No
IMF56	SynChr^{PURE}	Orange	*HIS3, LEU2*	Yes	No
IMF57	SynChr^{PURE}	White	*LEU2*	No	No
IMF58	SynChr^{PURE}	White	-	No	No
IMF59	SynChr^{PURE}	White	*URA3, mTurquoise2*	Yes	No
IMF60	SynChr^{PURE}	White	*HIS3, mRuby2*	No	No
IMF61	SynChr^{PURE}	White	*LEU2, HIS3, mRuby2*	Yes	No
IMF62	SynChr^{PURE}	White	*URA3, HIS3, mRuby2, mTurquoise2*	Yes	No
IMF63	SynChr^{PURE}	White	*URA3, HIS3, LEU2, mRuby2, mTurquoise2*	No	No
IMF64	SynChr^{PURE}	Red[4]	*HIS3*	Yes	No
IMF65	SynChr$^{control_2\mu}$	Orange	*mRuby2* (mix), *mTurquoise2* (mix)	No	No
IMF68	SynChr$^{PURE_2\mu}$	White	*mRuby2, mTurquoise2*	No	No
IMF70	SynChr$^{PURE_2\mu}$	White	*URA3, mRuby2* (mix), *mTurquoise2* (mix)	No	No

[1] *URA3, HIS3, LEU2*, *mRuby2* and/or *mTurquoise2*.
[2] All fragments are present in the correct order, but an insert of 89 bp is present in CF4, part of the CEN6/ARS4_*bla*_ColE1 fragment shows lower sequencing coverage and Chunk 10 contains a 4111 bp gap instead of a 1174 or 3310 bp gap in the pGGKp363 template plasmid.
[3] All fragments are present in the correct order, but part of the CEN6/ARS4_*bla*_ColE1 fragment is missing.
[4] IMF64 showed a red color instead of orange.

To reduce the number of printed pages, Supplementary Tables S3.4–S3.17 are available online from the 4TU.ResearchData repository, as described in the next section Supplementary data. Any references mentioned in the Supplementary Tables are included in the section References. A brief description of the content of these tables can be found in the table titles below.

Table S3.4: Plasmids used to create PURE cassette fragments for assembly of SynChrs in yeast

Table S3.5: Plasmids used to create marker fragments for assembly of SynChrs in yeast

Table S3.6: Plasmid used for qPCR standard curve

Table S3.7: Plasmids used for PURE expression

Table S3.8: Primers used to construct the PURE cassette plasmids

Table S3.9: Primers used to construct the *crt* expression cassette plasmids

Table S3.10: Other primers used for plasmid construction

Table S3.11: Diagnostic primers for *E. coli* colony PCR

Table S3.12: Diagnostic primers for Sanger sequencing

Table S3.13: Primers used to generate the fragments for SynChr assembly in yeast

Table S3.14: Primers used for qPCR

Table S3.15: Primers used to amplify the *yfp* cassette from SynChrs for PURE expression

Table S3.16: SHR sequences used in this study

Table S3.17: Fragments for assembly of SynChrs in yeast

Supplementary data

Supplementary data is available online from the 4TU.ResearchData repository at https://doi.org/10.4121/feb7423b-8194-4d99-89d8-593023e06473 or via the QR code below.

Links to: https://doi.org/10.4121/feb7423b-8194-4d99-89d8-593023e06473

The following data is included:

- Designed maps of SynChr^{PURE}, SynChr^{control}, SynChr$^{PURE_2\mu}$ and SynChr$^{control_2\mu}$ in GenBank format
- Sequencing data after total DNA isolation from *S. cerevisiae*
  - » Raw Nanopore sequencing reads, in FASTQ format
  - » Consensus sequences of SynChr variants, in GenBank format
- Supplementary Tables S3.4–S3.17 in Excel format

References

1. Luisi,P.L. and Stano,P. eds. (2011) The Minimal Cell. Springer Dordrecht. https://doi.org/10.1007/978-90-481-9944-0
2. Rothschild,L.J., Averesch,N.J.H., Strychalski,E.A., Moser,F., Glass,J.I., Perez,R.C., Yekinni,I.O., Rothschild-mancinelli,B., Kingman,G.A.R., Wu,F., et al. (2024) Building Synthetic Cells - From the Technology Infrastructure to Cellular Entities. *ACS Synth. Biol.*, **13**, 974–997. https://doi.org/10.1021/acssynbio.3c00724
3. Alberts,B. (2011) A grand challenge in biology. *Science*, **333**, 1200. https://doi.org/10.1126/science.1213238
4. Mushegian,A.R. and Koonin,E. V. (1996) A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *PNAS*, **93**, 10268–10273. https://doi.org/10.1073/pnas.93.19.10268
5. Gil,R., Silva,F.J., Peretó,J. and Moya,A. (2004) Determination of the Core of a Minimal Bacterial Gene Set. *Microbiol. Mol. Biol. Rev.*, **68**, 518–537. https://doi.org/10.1128/mmbr.68.3.518-537.2004
6. Glass,J.I., Assad-Garcia,N., Alperovich,N., Yooseph,S., Lewis,M.R., Maruf,M., Hutchison,C.A., Smith,H.O. and Venter,J.C. (2006) Essential genes of a minimal bacterium. *Proc. Natl. Acad. Sci. U. S. A.*, **103**, 425–430. https://doi.org/10.1073/pnas.0510013103
7. Kobayashi,K., Ehrlich,S.D., Albertini,A., Amati,G., Andersen,K.K., Arnaud,M., Asai,K., Ashikaga,S., Aymerich,S., Bessieres,P., et al. (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U. S. A.*, **100**, 4678–4683. https://doi.org/10.1073/pnas.0730515100
8. Koonin,E. V. (2000) How Many Genes Can Make a Cell: The Minimal-Gene-Set Concept. *Annu. Rev. Genomics Hum. Genet.*, **1**, 99–116.
9. Hutchison,C.A., Chuang,R.Y., Noskov,V.N., Assad-Garcia,N., Deerinck,T.J., Ellisman,M.H., Gill,J., Kannan,K., Karas,B.J., Ma,L., et al. (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253. https://doi.org/10.1126/science.aad6253
10. Abil,Z. and Danelon,C. (2020) Roadmap to Building a Cell: An Evolutionary Approach. *Front. Bioeng. Biotechnol.*, **8**, 1–8. https://doi.org/10.3389/fbioe.2020.00927
11. Shepherd,T.R., Du,L., Liljeruhm,J., Samudyata, Wang,J., Sjödin,M.O.D., Wetterhall,M., Yomo,T. and Forster,A.C. (2017) *De novo* design and synthesis of a 30-cistron translation-factor module. *Nucleic Acids Res.*, **45**, 10895–10905. https://doi.org/10.1093/nar/gkx753
12. Pryor,J.M., Potapov,V., Bilotti,K., Pokhrel,N. and Lohman,G.J.S. (2022) Rapid 40 kb Genome Construction from 52 Parts through Data-optimized Assembly Design. *ACS Synth. Biol.*, **11**, 2036–2042. https://doi.org/10.1021/acssynbio.1c00525
13. Casini,A., Storch,M., Baldwin,G.S. and Ellis,T. (2015) Bricks and blueprints: Methods and standards for DNA assembly. *Nat. Rev. Mol. Cell Biol.*, **16**, 568–576. https://doi.org/10.1038/nrm4014
14. Hillson,N.J. (2011) DNA Assembly Method Standardization for Synthetic Biomolecular Circuits and Systems. In Koeppl,H., Setti,G., di Bernardo,M., Densmore,D. (eds), *Design and Analysis of Biomolecular Circuits*. Springer New York, NY, pp. 295–314. https://doi.org/10.1007/978-1-4419-6766-4
15. Sheng,Y., Mancino,V. and Birren,B. (1995) Transformation of *Escherichia coli* with large DNA molecules by electroporation. *Nucleic Acids Res.*, **23**, 1990–1996. https://doi.org/10.1093/nar/23.11.1990
16. Warren,R.L., Freeman,J.D., Levesque,R.C., Smailus,D.E., Flibotte,S. and Holt,R.A. (2008) Transcription of foreign DNA in *Escherichia coli*. *Genome Res.*, **18**, 1798–1805. https://doi.org/10.1101/gr.080358.108
17. Gibson,D.G., Benders,G.A., Andrews-Pfannkoch,C., Denisova,E.A., Baden-Tillson,H., Zaveri,J., Stockwell,T.B., Brownley,A., Thomas,D.W., Algire,M.A., et al. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science*, **319**, 1215–1220. https://doi.org/10.1126/science.1151721
18. Gibson,D.G., Benders,G.A., Axelrod,K.C., Zaveri,J., Algire,M.A., Moodie,M., Montague,M.G., Venter,J.C., Smith,H.O. and Hutchison,C.A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl. Acad. Sci. U. S. A.*, **105**, 20404–20409. https://doi.org/10.1073/pnas.0811011106
19. Postma,E.D., Dashko,S., van Breemen,L., Taylor Parkins,S.K., van den Broek,M., Daran,J.-M. and Daran-Lapujade,P. (2021) A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **49**, 1769–1783. https://doi.org/10.1093/nar/gkaa1167
20. Tagwerker,C., Dupont,C.L., Karas,B.J., Ma,L., Chuang,R.Y., Benders,G.A., Ramon,A., Novotny,M., Montague,M.G., Venepally,P., et al. (2012) Sequence analysis of a complete 1.66 Mb *Prochlorococcus marinus* MED4 genome cloned in yeast. *Nucleic Acids Res.*, **40**, 10375–10383.

https://doi.org/10.1093/nar/gks823
21. Koster,C.C., Postma,E.D., Knibbe,E., Cleij,C. and Daran-Lapujade,P. (2022) Synthetic Genomics From a Yeast Perspective. *Front. Bioeng. Biotechnol.*, **10**, 869486. https://doi.org/10.3389/fbioe.2022.869486
22. Kuijpers,N.G.A., Solis-Escalante,D., Bosman,L., van den Broek,M., Pronk,J.T., Daran,J.M. and Daran-Lapujade,P. (2013) A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. *Microb. Cell Fact.*, **12**, 1–13. https://doi.org/10.1186/1475-2859-12-47
23. 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. https://doi.org/10.1038/s41576-019-0186-3
24. Garenne,D., Haines,M.C., Romantseva,E.F., Freemont,P., Strychalski,E.A. and Noireaux,V. (2021) Cell-free gene expression. *Nat. Rev. Methods Prim.*, **1**, 49. https://doi.org/10.1038/s43586-021-00046-x
25. 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**, 751–5. https://doi.org/10.1038/90802
26. Shimizu,Y., Kanamori,T. and Ueda,T. (2005) Protein synthesis by pure translation systems. *Methods*, **36**, 299–304. https://doi.org/10.1016/j.ymeth.2005.04.006
27. 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**, 1–12. https://doi.org/10.1038/s41467-018-03926-1
28. Blanken,D., Foschepoth,D., Serrão,A.C. and Danelon,C. (2020) Genetically controlled membrane synthesis in liposomes. *Nat. Commun.*, **11**, 1–13. https://doi.org/10.1038/s41467-020-17863-5
29. Eto,S., Matsumura,R., Shimane,Y., Fujimi,M., Berhanu,S., Kasama,T. and Kuruma,Y. (2022) Phospholipid synthesis inside phospholipid membrane vesicles. *Commun. Biol.*, **5**, 1–11. https://doi.org/10.1038/s42003-022-03999-1
30. 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**, 4969. https://doi.org/10.1038/s41467-019-12932-w
31. 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**, 1–11. https://doi.org/10.1038/s42003-020-01258-9
32. Kohyama,S., Merino-Salomon,A. and Schwille,P. (2022) In vitro assembly, positioning and contraction of a division ring in minimal cells. *Nat. Commun.*, **13**, 6098. https://doi.org/10.1038/s41467-022-33679-x
33. 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**, 1–12. https://doi.org/10.1038/s41598-020-80827-8
34. Hagino,K. and Ichihashi,N. (2023) In Vitro Transcription/Translation-Coupled DNA Replication through Partial Regeneration of 20 Aminoacyl-tRNA Synthetases. *ACS Synth. Biol.*, **12**, 1252–1263. https://doi.org/10.1021/acssynbio.3c00014
35. Jewett,M.C., Fritz,B.R., Timmerman,L.E. and Church,G.M. (2013) *In vitro* integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Mol. Syst. Biol.*, **9**, 1–8. https://doi.org/10.1038/msb.2013.31
36. Ganesh,R.B. and Maerkl,S.J. (2024) Towards Self-regeneration: Exploring the Limits of Protein Synthesis in the Protein Synthesis Using Recombinant Elements (PURE) Cell-free Transcription-Translation System. *ACS Synth. Biol.*, **13**, 2555–2566. https://doi.org/10.1021/acssynbio.4c00304
37. O'Neill,B.M., Mikkelson,K.L., Gutierrez,N.M., Cunningham,J.L., Wolff,K.L., Szyjka,S.J., Yohn,C.B., Redding,K.E. and Mendez,M.J. (2012) An exogenous chloroplast genome for complex sequence manipulation in algae. *Nucleic Acids Res.*, **40**, 2782–2792. https://doi.org/10.1093/nar/gkr1008
38. Rideau,F., Le Roy,C., Descamps,E.C.T., Renaudin,H., Lartigue,C. and Bébéar,C. (2017) Cloning, Stability, and Modification of *Mycoplasma hominis* Genome in Yeast. *ACS Synth. Biol.*, **6**, 891–901. https://doi.org/10.1021/acssynbio.6b00379
39. Hohnholz,R., Pohlmann,K.J. and Achstetter,T. (2017) Impact of plasmid architecture on stability and yEGFP3 reporter gene expression in a set of isomeric multicopy vectors in yeast. *Appl. Microbiol. Biotechnol.*, **101**, 8455–8463. https://doi.org/10.1007/s00253-017-8558-0
40. Gibson,D.G. (2011) Gene and genome construction in yeast. *Curr. Protoc. Mol. Biol.*, **94**, 3.22.1-3.22.17. https://doi.org/10.1002/0471142727.mb0322s94
41. Kouprina,N. and Larionov,V. (2003) Exploiting the yeast *Saccharomyces cerevisiae* for the study of the organization and evolution of complex genomes. *FEMS Microbiol. Rev.*, **27**, 629–649. https://doi.

org/10.1016/S0168-6445(03)00070-6
42. Tuite,M.F. (1991) Expression of Heterologous Genes. In Tuite,M.F., Oliver,S.G. (eds), *Saccharomyces*. *Biotechnology Handbooks, vol 4*. Springer, Boston, MA. pp. 169–212. https://doi.org/10.1007/978-1-4899-2641-8_6
43. Lewin,A., Tran,T.T., Jacob,D., Mayer,M., Freytag,B. and Appel,B. (2004) Yeast DNA sequences initiating gene expression in *Escherichia coli*. *Microbiol. Res.*, **159**, 19–28. https://doi.org/10.1016/j.micres.2004.01.006
44. Jopcik,M., Bauer,M., Moravcikova,J., Boszoradova,E., Matusikova,I. and Libantova,J. (2013) Plant tissue-specific promoters can drive gene expression in *Escherichia coli*. *Plant Cell. Tissue Organ Cult.*, **113**, 387–396. https://doi.org/10.1007/s11240-012-0278-7
45. Jacob,D., Lewin,A., Meister,B. and Appel,B. (2002) Plant-specific promoter sequences carry elements that are recognised by the eubacterial transcription machinery. *Transgenic Res.*, **11**, 291–303. https://doi.org/10.1023/A:1015620016472
46. Lewin,A., Jacob,D., Freytag,B. and Appel,B. (1998) Gene expression in bacteria directed by plant-specific regulatory sequences. *Transgenic Res.*, **7**, 403–411. https://doi.org/10.1023/A:1008876826415
47. Lewin,A., Mayer,M., Chusainow,J., Jacob,D. and Appel,B. (2005) Viral promoters can initiate expression of toxin genes introduced into *Escherichia coli*. *BMC Biotechnol.*, **5**, 1–9. https://doi.org/10.1186/1472-6750-5-19
48. Verwaal,R., Wang,J., Meijnen,J.P., Visser,H., Sandmann,G., van den Berg,J.A. and van Ooyen,A.J.J. (2007) High-level production of beta-carotene in *Saccharomyces cerevisiae* by successive transformation with carotenogenic genes from *Xanthophyllomyces dendrorhous*. *Appl. Environ. Microbiol.*, **73**, 4342–4350. https://doi.org/10.1128/AEM.02759-06
49. Mathiasen,D.P. and Lisby,M. (2014) Cell cycle regulation of homologous recombination in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.*, **38**, 172–184. https://doi.org/10.1111/1574-6976.12066
50. Noskov,V.N., Chuang,R.Y., Gibson,D.G., Leem,S.H., Larionov,V. and Kouprina,N. (2011) Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. *Nat. Protoc.*, **6**, 89–96. https://doi.org/10.1038/nprot.2010.174
51. Singh,M. V. and Weil,P.A. (2002) A method for plasmid purification directly from yeast. *Anal. Biochem.*, **307**, 13–17. https://doi.org/10.1016/S0003-2697(02)00018-0
52. Datta,A., Hendrix,M., Lipsitch,M. and Jinks-Robertson,S. (1997) Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 9757–9762. https://doi.org/10.1073/pnas.94.18.9757
53. Komura,R., Aoki,W., Motone,K., Satomura,A. and Ueda,M. (2018) High-throughput evaluation of T7 promoter variants using biased randomization and DNA barcoding. *PLoS One*, **13**, 1–16. https://doi.org/10.1371/journal.pone.0196905
54. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950. https://doi.org/10.1038/nbt.1568
55. Mimitou,E.P. and Symington,L.S. (2008) Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature*, **455**, 770–774. https://doi.org/10.1038/nature07312
56. 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. https://doi.org/10.1021/acssynbio.6b00017
57. 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. https://doi.org/10.1016/j.febslet.2004.09.046
58. Kouprina,N., Nikolaishvili,N., Graves,J., Koriabine,M., Resnick,M.A. and Larionov,V. (1999) Integrity of human YACs during propagation in recombination-deficient yeast strains. *Genomics*, **56**, 262–273. https://doi.org/10.1006/geno.1998.5727
59. Chan,K.M., Liu,Y.T., Ma,C.H., Jayaram,M. and Sau,S. (2013) The 2 micron plasmid of *Saccharomyces cerevisiae*: A miniaturized selfish genome with optimized functional competence. *Plasmid*, **70**, 2–17. https://doi.org/10.1016/j.plasmid.2013.03.001
60. Entian,K.D. and Kötter,P. (2007) 25 Yeast Genetic Strain and Plasmid Collections. *Methods Microbiol.*, **36**, 629–666. https://doi.org/10.1016/S0580-9517(06)36025-4
61. Verduyn,C., Postma,E., Scheffers,W.A. and van Dijken,J.P. (1992) Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, **8**, 501–517. https://doi.org/10.1002/yea.320080703
62. Lee,M.E., DeLoache,W.C., Cervantes,B. and Dueber,J.E. (2015) A Highly Characterized Yeast Toolkit for

Modular, Multipart Assembly. *ACS Synth. Biol.*, **4**, 975–986. https://doi.org/10.1021/sb500366v
63. Inoue,H., Nojima,H. and Okayama,H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23–28. https://doi.org/10.1016/0378-1119(90)90336-P
64. Gietz,R.D. and Woods,R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.*, **350**, 87–96. https://doi.org/10.1016/S0076-6879(02)50957-5
65. Li,H. (2018) Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191
66. Salazar,A.N., de Vries,A.R.G., van den Broek,M., Wijsman,M., Cortés,P. de la T., Brickwedde,A., Brouwers,N., Daran,J.M.G. and Abeel,T. (2017) Nanopore sequencing enables near-complete *de novo* assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res.*, **17**, 1–11. https://doi.org/10.1093/femsyr/fox074
67. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
68. Thorvaldsdóttir,H., Robinson,J.T. and Mesirov,J.P. (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192. https://doi.org/10.1093/bib/bbs017
69. The Gresham Lab (2022) labtools: Useful Tools for the Gresham Lab.
70. R Core Team (2021) R: A Language and Environment for Statistical Computing.
71. Kolmogorov,M., Yuan,J., Lin,Y. and Pevzner,P.A. (2019) Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.*, **37**, 540–546. https://doi.org/10.1038/s41587-019-0072-8
72. Koren,S., Walenz,B.P., Berlin,K., Miller,J.R., Bergman,N.H. and Phillippy,A.M. (2017) Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Res.*, **27**, 722–736. https://doi.org/10.1101/gr.215087.116
73. Kurtz,S., Phillippy,A., Delcher,A.L., Smoot,M., Shumway,M., Antonescu,C. and Salzberg,S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biol.*, **5**. https://doi.org/10.1186/gb-2004-5-2-r12
74. Scott,A., Noga,M.J., De Graaf,P., Westerlaken,I., Yildirim,E. and Danelon,C. (2016) Cell-free phospholipid biosynthesis by gene-encoded enzymes reconstituted in liposomes. *PLoS One*, **11**, 1–23. https://doi.org/10.1371/journal.pone.0163058
75. Abil,Z., Restrepo Sierra,A.M. and Danelon,C. (2023) Clonal Amplification-Enhanced Gene Expression in Synthetic Vesicles. *ACS Synth. Biol.*, **12**, 1187–1203. https://doi.org/10.1021/acssynbio.2c00668
76. Beekwilder,J., van Rossum,H.M., Koopman,F., Sonntag,F., Buchhaupt,M., Schrader,J., Hall,R.D., Bosch,D., Pronk,J.T., van Maris,A.J.A., et al. (2014) Polycistronic expression of a β-carotene biosynthetic pathway in *Saccharomyces cerevisiae* coupled to β-ionone production. *J. Biotechnol.*, **192**, 383–392. https://doi.org/10.1016/j.jbiotec.2013.12.016
77. Hassing,E.J., Buijs,J., Blankerts,N., Luttik,M.A., Hulster,E.A. d., Pronk,J.T. and Daran,J.M. (2021) Elimination of aromatic fusel alcohols as by-products of *Saccharomyces cerevisiae* strains engineered for phenylpropanoid production by 2-oxo-acid decarboxylase replacement. *Metab. Eng. Commun.*, **13**, e00183. https://doi.org/10.1016/j.mec.2021.e00183
78. Hassing,E.J., de Groot,P.A., Marquenie,V.R., Pronk,J.T. and Daran,J.M.G. (2019) Connecting central carbon and aromatic amino acid metabolisms to improve *de novo* 2-phenylethanol production in *Saccharomyces cerevisiae*. *Metab. Eng.*, **56**, 165–180. https://doi.org/10.1016/j.ymben.2019.09.011
79. Botman,D., de Groot,D.H., Schmidt,P., Goedhart,J. and Teusink,B. (2019) *In vivo* characterisation of fluorescent proteins in budding yeast. *Sci. Rep.*, **9**, 1–14. https://doi.org/10.1038/s41598-019-38913-z
80. Randazzo,P., Bennis,N.X., Daran,J.M. and Daran-Lapujade,P. (2021) gEL DNA: A Cloning-and Polymerase Chain Reaction-Free Method for CRISPR-Based Multiplexed Genome Editing. *CRISPR J.*, **4**, 896–913. https://doi.org/10.1089/crispr.2020.0028
81. Wronska,A.K., Haak,M.P., Geraats,E., Bruins Slot,E., van den Broek,M., Pronk,J.T. and Daran,J.-M. (2020) Exploiting the Diversity of Saccharomycotina Yeasts To Engineer Biotin-Independent Growth of *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.*, **86**, e00270-20.
82. Sarabipour,S., King,C. and Hristova,K. (2014) Uninduced high-yield bacterial expression of fluorescent proteins. *Anal. Biochem.*, **449**, 155–157. https://doi.org/10.1016/j.ab.2013.12.027
83. 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**. https://doi.org/10.1088/1478-3975/aaf33d

Chapter 4

De novo design and assembly of minimal genomes for the synthetic cell

The bottom-up construction of a minimal synthetic cell requires the *de novo* design and assembly of its DNA genome. This chapter presents the design and construction of a synthetic minimal genome (SynMG1), harboring cellular modules for phospholipid biosynthesis, DNA replication, and cell division, intended for *in vitro* expression using PURE system. Fluorescent reporter genes were incorporated to monitor expression kinetics. The design included replication origins that, upon SynMG linearization, should enable replication by the protein-primed ϕ29 DNA replication system. SynMG1 was assembled in yeast via homologous recombination from pre-assembled fragments containing cellular modules. Yeast marker fragments were added between coding fragments to facilitate screening of correct assemblies. Moreover, we included a BAC backbone to enable transfer from yeast to *E. coli* for SynMG amplification. Correct assembly of the 41-kb SynMG1 from 14 fragments was verified by long-read sequencing. In contrast, attempts to construct a larger 105-kb SynMG (SynMG2), which included a translation factor module, were unsuccessful. SynMG1, isolated from *E. coli*, was expressed in PURE system and all encoded proteins were detected by liquid chromatography-mass spectrometry and fluorescence measurements. Furthermore, successful encapsulation and expression of SynMG1 in liposomes was demonstrated, as well as full-length replication of linearized SynMG1 by the ϕ29 DNA replication machinery.

Introduction

A major challenge for the bottom-up construction of synthetic cells is the design and assembly of a DNA genome encoding all cellular functions. Even the simplest self-replicating organisms contain several hundreds of genes organized in a single chromosome. Although the minimum number of genes that would be sufficient to support life is under debate, a prospective synthetic genome will likely contain between 150 and 500 genes, with a total size exceeding 150 kb. The lower bound is set by the conserved core of essential genes revealed from a comparative genome analysis and a biochemical approach to determine essential cellular functions ([1–3], reviewed in Chapter 1), while the upper range corresponds to the size of the JCVI-syn3.0 genome [4]. In this latter study, a minimal bacterial genome was constructed from shorter DNA fragments through a combination of *in vitro* assembly and assembly by homologous recombination in the yeast *Saccharomyces cerevisiae*. This shows the feasibility of stepwise assembly of genome-sized DNA. However, this minimal genome was largely based on the native genome design of *Mycoplasma mycoides*, maintaining much of its gene order and regulatory sequences.

Our vision of a synthetic cell genome is an assembly of genetic modules essential for life, irrespective of their origin. The minimal genome will therefore consist of genes and regulatory elements derived from different organisms, from engineering efforts or from *in vitro* or *in vivo* evolution. Modules are selected based on their simplicity, prioritizing systems that achieve a given function with the fewest necessary genes. Additional selection criteria include the current level of characterization and compatibility with *in vitro* transcription-translation. The core decoding machinery of our envisioned minimal genome is the PURE system, with all its components for transcription, translation, aminoacylation and energy regeneration. This approach to the construction of a minimal cell comes with new challenges, not only for the selection of the optimal gene set, but also in the *de novo* design and construction of artificial genomes. For instance, the use of T7 promoter and canonical prokaryotic RBSs for gene expression leads to a higher occurrence of repeated sequences as compared to existing natural or synthetic bacterial genomes.

Important milestones toward the completely *de novo* design and assembly of a chromosome for a synthetic cell are the synthesis of a 15-kb DNA construct containing all 21 genes that encode the proteins of the *E. coli* 30S ribosomal subunit [5] and the construction of a 30-cistron translation factor module, named pTFM1, encoding 30 of the 31 translation factors of PURE system, with only EF-Tu missing [6]. Although successful and relatively cheap, the assembly of pTFM1 by BioBrick cloning was labor-intensive, posing a challenge for scaling up the number of genes.

Cell-free expression of pTFM1 [7] and of the combined pLD plasmids (each containing a subset of the translation factor module of pTFM1) [8] was realized in PURE system. All proteins encoded by the plasmids were detected by liquid chromatography-coupled mass spectrometry (LC-MS). Notably, approximately half of these proteins could be expressed in amounts equal to or greater than their initial levels in PURE system [7, 8]. These results represent a major step toward the self-production of the PURE machinery. The next challenge lies in increasing the genetic complexity of the synthetic chromosome by introducing genes that support other essential cellular functions, such

as membrane growth, division, and DNA replication. When expressed in PURE system, this more 'complete' genome would then generate the transcriptome and proteome of the synthetic cell.

Herein is reported the *de novo* design and successful assembly in *S. cerevisiae* of a synthetic minimal genome (SynMG), harboring 15 genes for expression in PURE system. The following cellular modules are included: phospholipid biosynthesis by the *E. coli* Kennedy pathway, DNA replication with the phage ϕ29 replication apparatus, and cell division with *E. coli* proteins. Furthermore, it includes two fluorescent reporter genes for easy monitoring of PURE expression kinetics, and origins of replication compatible with the protein-primed ϕ29 DNA replication system. To streamline the assembly-to-characterisation pipeline, screening and selection markers were incorporated between each coding segment, as described in Chapter 3. Moreover, since the quantity and purity of SynChr DNA isolated from yeast is a bottleneck for expression in PURE system (Chapter 3), the chromosome design included a BAC backbone to enable shuttling to *E. coli* for amplification and isolation with high DNA yield. Purified SynMG was expressed in PURE system and the presence of encoded proteins was checked by LC-MS and fluorescence. Encapsulation and expression of SynMG1 in liposomes was explored, as well as replication of the linearized SynMG by the ϕ29 DNA replication machinery.

Results

Design of a minimal synthetic cell genome

Two different minimal genomes encoding multiple cellular modules named SynMG1 and SynMG2 were designed (Figure 4.1). The modules were previously characterized when individually expressed in PURE system. The membrane growth module consisted of genes encoding enzymes of the *E. coli* Kennedy pathway for phospholipid synthesis: *plsB*, *plsC*, *cdsA*, *pssA*, *psd*, *pgsA* and *pgpA* [9]. Two cell division modules were included: (i) genes *minD* and *minE* encoding two *E. coli* proteins of the Min system, which can localize the division machinery and deform liposomes [10, 11], and (ii) genes *ftsZ* and *ftsA*, encoding *E. coli* divisome proteins which together can constrict liposomes [12]. FtsZ was fused to the fluorescent mVenus protein for easy visualization during *in vitro* experiments. The genes were organized into an operon to mimic the FtsZ and FtsA protein ratio found in *E. coli* cells. The functionality of this module for liposome constriction was previously validated in the Christophe Danelon lab (Federico Ramirez Gomez, unpublished). The DNA replication module consisted of the *p2* and *p3* genes, encoding DNAP and TP, the two most important proteins of the protein-primed DNA replication system from bacteriophage ϕ29 [13]. These four modules formed SynMG1. A second, larger minimal genome named SynMG2 was constructed by extension of the gene set of SynMG1 to include a translation module. This additional module consisted of 32 translation factor genes from *E. coli* encoding 20 aminoacyl-tRNA synthetases (*alaS*, *argS*, *asnS*, *aspS*, *cysS*, *glnS*, *gltX*, *glyQ*, *glyS*, *hisS*, *ileS*, *leuS*, *lysS*, *metG*, *pheS*, *pheT*, *proS*, *serS*, *thrS*, *trpS*, *tyrS*, *valS*), methionyl-tRNA formyltransferase (*fmt*), initiation factors (*infA*, *infB*, *infC*), elongation factors (*fusA*, *tsf*), release factors (*prfA*, *prfB*, *prfC*) and a ribosome recycling factor (*frr*) [6, 7]. The genes *pheS* and *pheT*, encoding two subunits of phenylalanyl-tRNA synthetase, were combined in an operon, as well as

the genes *glyQ* and *glyS*, encoding subunits of glycyl-tRNA synthetase. Expression of all minimal cell genes was under the control of a T7 promoter (*pT7*), with the exception of the phospholipid synthesis genes *pgsA* and *pgpA* placed under the control of an SP6 promoter (*pSP6*) [9]. SP6 RNAP can be added to PURE system and is expected to have orthogonal activity to T7 RNAP [9]. The presence of SP6 and T7 promoters enabled separate activation of the two branches of the phospholipid synthesis pathway. Additionally, both SynMG designs included fluorescent markers for direct readout of protein synthesis in PURE system: *mCherry*, controlled by *pT7*, and *eYFP*, controlled by *pSP6*. The fusion protein FtsZ:mVenus could also be used as fluorescent readout under *pT7* control. A *spinach* RNA sequence was introduced downstream the *eYFP* gene, which allowed monitoring of transcription [14]. A requirement for the chosen DNA replication machinery is a linear DNA configuration flanked by the ϕ29 origins of replication. Therefore, both synthetic chromosomes (SynChrs) included a fragment containing the two origins of replication with an internal PmeI restriction site for linearization. The requirement of a recognition site for PmeI added a design constraint to the synthetic genome, i.e., the absence of additional PmeI recognition sites. While SynMG1 was devoid of PmeI sites, SynMG2 harbored one in the *fusA* gene, which will have to be removed, for instance through CRISPR/Cas9 editing of the SynChr in yeast.

The regulatory sequences (promoter, RBS, terminator) of most genes are similar and therefore repeated multiple times across the SynChrs, which is expected to cause unwanted recombination during SynChr assembly in yeast (Chapter 3). To minimize the number of transformants that needed to be sequenced for identification of correct SynChr assemblies, yeast markers (auxotrophic, fluorescent, antibiotic resistance and chromogenic markers) were included between each minimal cell module (Chapter 3). To ensure maintenance in yeast, the designs included a CEN6 centromere and an ARS4 replication origin, along with one (SynMG1: ARS417) or two (SynMG2: ARS417 and ARS1) additional autonomously replicating sequences (ARS), providing a replication origin every 30–40 kb [15]. The CEN6/ARS4 fragment additionally contained an I-SceI restriction site for selective linearization of the SynChr and a landing pad with synthetic gRNA target sites [16, 17] to allow for SynChr engineering by CRISPR/Cas9. Due to the low SynChr concentration and purity obtained when isolating SynChrs from yeast (Chapter 3), an amplification step in *E. coli* was envisioned. The SynChrs of Chapter 3, which contained a high-copy number ColE1-like origin of replication, could not be stably maintained in *E. coli* (results not shown). This was likely due to the incompatibility of the replication origin with the large SynChr size and to undesired (intermolecular) recombination at the repeated regulatory sequences of SynChr^{PURE}. Large inserts can be stably maintained in bacterial artificial chromosomes (BACs), because their low copy number alleviates metabolic burden, limits intermolecular recombination, and provides increased tolerance for toxic sequences [18, 19]. Therefore, a pCC1BAC backbone (Epicentre) was included in the designs of SynMG1 and SynMG2. Constructs with a pCC1BAC backbone are maintained at a single copy, but the optional activation of a second origin through *trfA* expression in *E. coli* strain EPI300—induced by L-arabinose—raises the copy number to approximately 10–20, enhancing DNA yield and purity upon isolation (Epicentre, [20]).

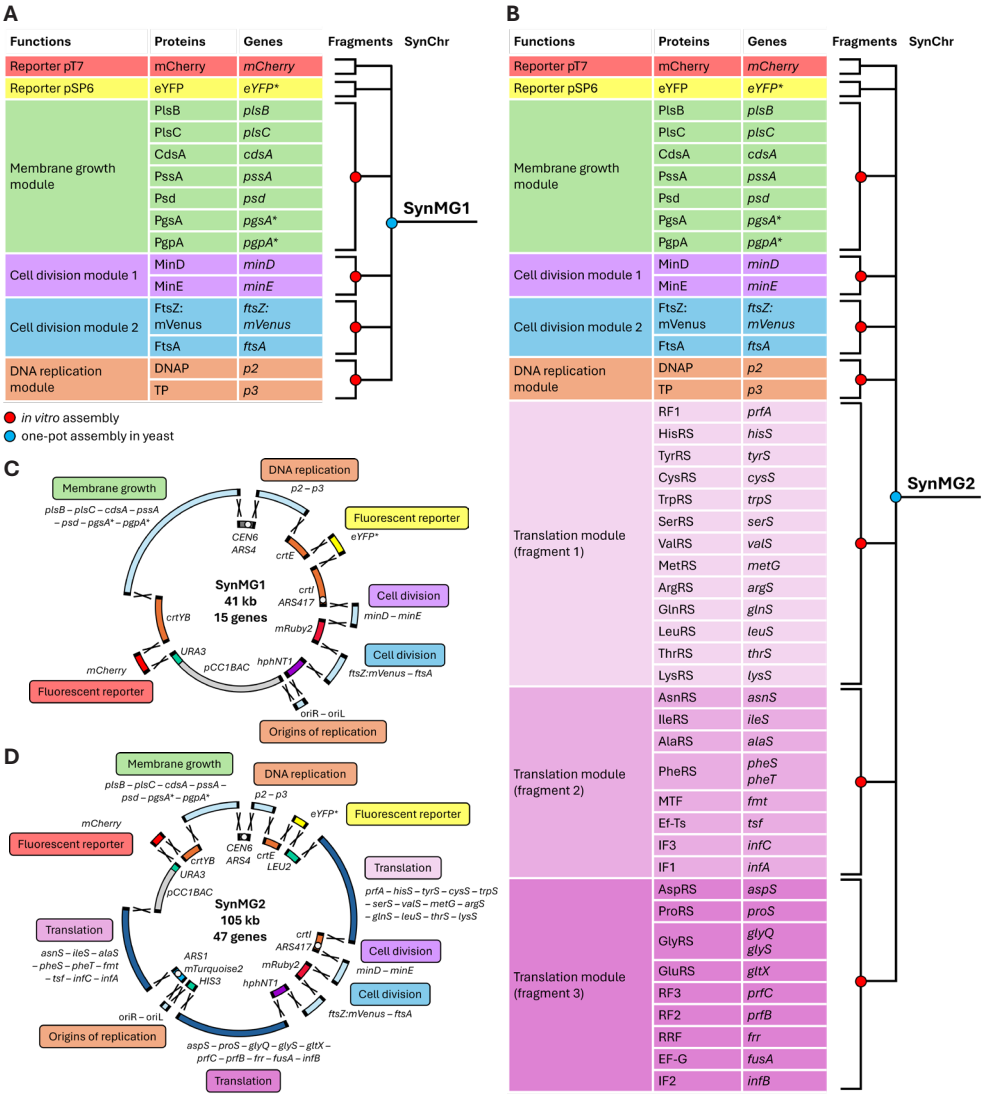


Figure 4.1: Design of two synthetic minimal genomes. A) Genes and encoded proteins on SynMG1. Genes involved in each cellular module were previously cloned together on a plasmid *in vitro* (Table S4.5). Fragments were amplified from these plasmids by PCR and assembled in yeast to obtain SynMG1. **B)** Genes and encoded proteins on SynMG2. The same assembly strategy was used as for SynMG1. SynMG2 contains all modules of SynMG1, with the addition of a translation module split on three assembly fragments. **C)** Design of SynMG1 encoding 15 genes for expression in PURE system and with a total size of 41 kb. **D)** Design of SynMG2 containing 47 genes for expression in PURE system and with a total size of 105 kb. In both designs, a marker for screening and selection in yeast or *E. coli* was introduced between each fragment harboring a synthetic cell module. Genes indicated with an asterisk are controlled by *pSP6*, all other genes by *pT7*.

The SynMG1 design therefore comprised 14 assembly fragments: seven fragments containing minimal cell modules or reporters for expression in PURE system, which have previously been constructed by *in vitro* assembly (Table S4.5), and seven marker fragments for propagation, screening and selection in yeast and *E. coli*. All fragments were flanked by 60-bp SHR sequences for assembly in yeast through homologous recombination [21]. The final SynChr was 41 kb in size and included 15 genes for expression in PURE system (Figure 4.1A and C). SynMG2, which differs from SynMG1 by the addition of three translation module fragments and three yeast marker fragments—bringing the total to 20 assembly fragments—was 105 kb in size and contained 47 genes for expression in PURE system (Figure 4.1B and D). The full content of both SynChr designs is described in Table S4.2 and provided in Supplementary Data. Assembly fragments are listed in Table S4.13 and a list of proteins encoded on the SynChrs for expression in PURE system is provided in Table S4.14.

SynMG1, but not SynMG2, could be correctly assembled in yeast

All fragments for assembly were obtained by PCR and cleaned up from gel, except for the three translation module fragments of SynMG2 that could not be gel-purified due to their large size (15, 19 and 25 kb).

A single transformation was carried out using the 14 assembly fragments for SynMG1, while three transformations were performed for SynMG2 with its 20 assembly fragments, which varied in the treatment of the large translation fragments and the amount of each fragment (see Methods). The SynMG1 transformation and two out of the three SynMG2 transformations resulted in orange colonies on the transformation plates (Figure S4.1). Screening and selection of SynChrs was performed using the pipeline described in Chapter 3. After total DNA isolation from yeast and long-read sequencing, three versions of SynMG1 (SynMG1.1, SynMG1.2 and SynMG1.3, from yeast strains IMF82, IMF85 and IMF87, respectively) and two versions of SynMG2 (SynMG2.1 and SynMG2.2, from yeast strains IMF95 and IMF96, respectively) were selected for further characterization (Table S4.3).

SynMG1.1 contained all fragments in the correct configuration, whereas SynMG1.2 lacked a single gene (*pgpA*). SynMG1.3 consisted of a mixture of correctly assembled SynChrs and SynChrs missing *mCherry* and the lipid synthesis genes. SynMG1.1, SynMG1.2 and SynMG1.3 all showed an unexpectedly high occurrence of point mutations in the lipid synthesis segment.

None of the versions of SynMG2 were correctly configured. SynMG2.1 was missing 25 out of 32 translation factor genes, as well as *minD* and *minE*. SynMG2.2 missed 28 out of 32 translation factor genes. Therefore, SynMG2 variants were not considered for further analysis.

An overview of all relevant mutations in SynMG1 and SynMG2 variants is given in Supplementary Data.

Sequence analysis revealed that incorrect configurations were caused by homologous recombination between repeated regulatory sequences, as previously observed (Chap-

ter 3). Most point mutations likely originated from the polymerase used in PCR. The lipid synthesis segment in SynMG1 and the pCC1BAC and translation segments in SynMG2, enriched for point mutations, were all amplified by the UltraRun LongRange polymerase (Table S4.13). These segments showed a higher incidence of point mutations than the other ones that were amplified by the high-fidelity Phusion or KOD Xtreme polymerases.

Amplification in *E. coli* enables the isolation of SynMG1 at nanomolar concentrations

Total DNA was isolated from yeast strains carrying SynChrs and used to transform *E. coli* strain EPI300 for SynChr amplification (Figure 4.2A). Transformation was successful for all selected SynMG1 variants. Single colonies from the transformation plates were checked for the presence of pCC1BAC, *mRuby2* and CEN6/ARS4 fragments by colony PCR, and positive colonies were cultured without high-copy number induction. DNA was isolated from *E. coli* and sequenced, revealing a mixture of complete SynChrs and SynChrs missing sequences between repeats (histograms with read lengths shown in Figure 4.2B). SynChrs may have been unstable during transformation of *E. coli* or subsequent recovery (as also observed in [6]), leading to colonies on the transformation plate consisting of a mixed population of cells with different SynChr configurations. Therefore, cells from the transformation plates were streaked to obtain single colonies that were checked by PCR and used to inoculate overnight cultures. Since SynChrs are assumed to be stable during colony growth after streaking [6], these single colonies were expected to contain either recombined or full SynChrs, but not a mixture. This workflow ensured that cultures inoculated from a resuspended colony, which was confirmed as positive by colony PCR, contained a single SynChr configuration. Sequencing confirmed single SynChr configurations, reflected by a single high peak in the histogram in Figure 4.2C. Glycerol stocks were prepared from the cultures used for this SynChr isolation.

DNA concentrations measured by Qubit ranged from ca. 200 to 450 ng μL^{-1} (ca. 8–16 nM) in 50 µL. Nanodrop analysis demonstrated $A_{260/280}$ ratios of 1.8–1.9 (within the desired range) and $A_{260/230}$ ratios of 1.6–1.8 a (slightly below the preferred 1.8–2.2 range). Sequencing revealed *E. coli* genomic DNA contamination at 51% for SynMG1.1, 53% for SynMG1.2 and 35% for SynMG1.3, likely due to the low SynChr copy number [20]. Consequently, Qubit overestimated SynChr DNA concentration by a factor two to three.

A preliminary experiment (without replicates) showed that high-copy number induction of SynMG1.1 with L-arabinose solution did not trigger recombination and yielded higher SynChr DNA concentrations (> 1000 ng μL^{-1}) with reduced genomic DNA contamination (9%). For this isolation, cells were grown from glycerol stock, streaked to obtain single colonies, and verified by colony PCR before being cultured for SynChr isolation. This process ensured that any potential recombination caused by cell recovery from storage was detected, allowing for SynChr isolation only from cells with non-recombined SynChrs.

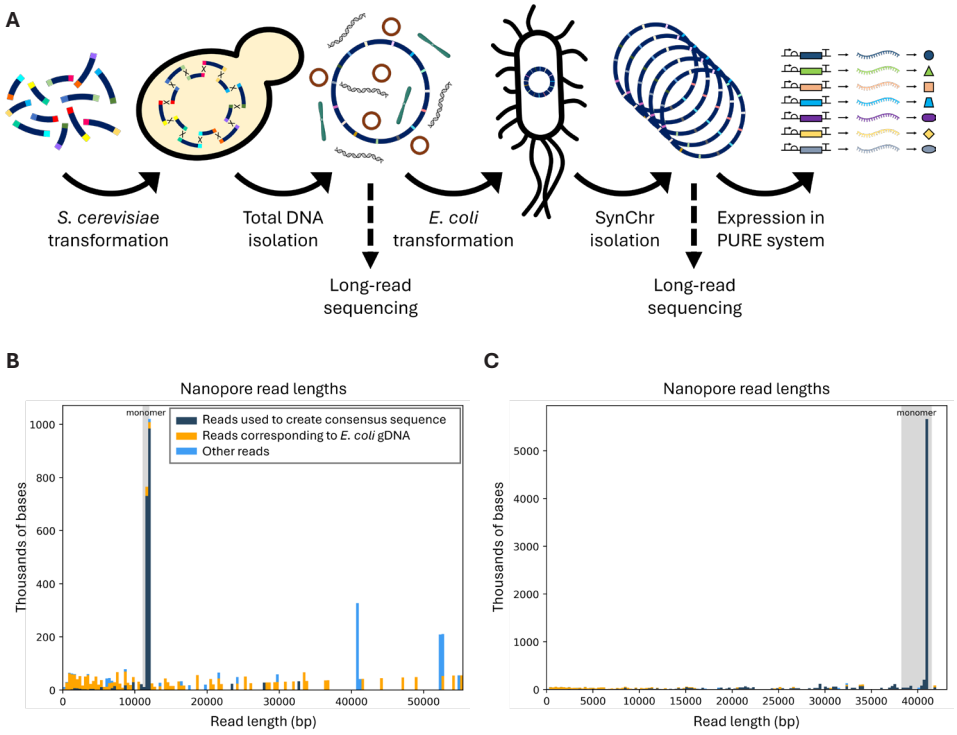


Figure 4.2: *E. coli* was used as an amplification host for SynChrs isolated from yeast before expression in PURE system. A) Pipeline from assembly in yeast to expression in PURE system, involving an intermediate step for amplification of the SynChr in *E. coli*. After transformation of yeast with assembly DNA fragments, the SynChr was assembled by homologous recombination. Total DNA was extracted from yeast, verified by long-read sequencing, and *E. coli* was transformed, which selectively amplified the SynChr. SynChr DNA was isolated from *E. coli*, checked by long-read sequencing, and tested for expression in PURE system. **B)** Representative histogram showing read lengths of SynChrs isolated from *E. coli*, when grown directly from the transformation plate. Multiple dark and light blue peaks are visible (at approximately 11, 41 and 53 kb), corresponding to multiple SynChr configurations (recombined and non-recombined). **C)** Representative histogram showing read lengths of SynChrs isolated from *E. coli*, when streaked for single colonies. A single dominant dark blue peak is visible at 41 kb, corresponding to the non-recombined SynChr.

Successful expression of fluorescent markers under T7 or SP6 promoters from SynMG1 in bulk PURE system

To assess the expression of the three fluorescent proteins encoded by SynMG1, bulk PURE reactions were performed for all three SynMG1 variants. Production of mCherry and mVenus controlled by *pT7*, and eYFP controlled by *pSP6*, was tested using reactions containing either T7 RNAP or SP6 RNAP (Figure 4.3A). Fluorescence kinetics were measured during 16 h. As positive controls, plasmids encoding individual fluorescent markers—originally used as templates for PCR to obtain SynChr assembly fragments—were used. All DNA templates were added at 1 nM final concentration (for SynChrs, the concentrations measured by Qubit were taken, not corrected for *E. coli* genomic DNA contamination). Negative controls consisted of PURE reactions without DNA. Each condition was tested in three independent replicates.

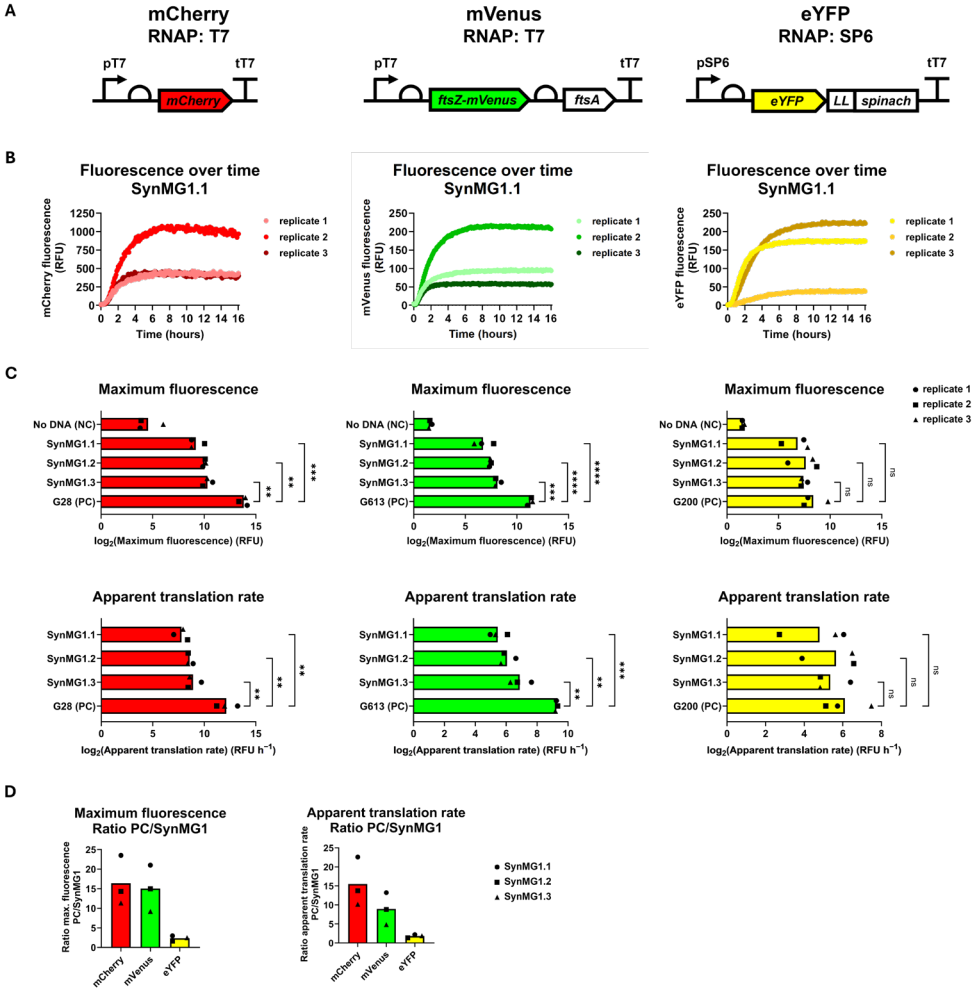


Figure 4.3: Expression kinetics of fluorescent reporter proteins encoded on SynMG1. A) Schematic representation of expression cassettes of *mCherry*, *mVenus* and *eYFP*. *mCherry* and *mVenus* are controlled by *pT7*, *eYFP* by *pSP6*. *LL* = long linker. **B)** Fluorescence measurements over 16 h of expression of SynMG1.1 in PURE system. See Figure S4.2 for graphs of other SynMG1 variants and control templates. **C)** Kinetic parameters of *mVenus*, *mCherry* and *eYFP* expression in PURE system: maximum fluorescence (RFU) and apparent translation rate (RFU h^{-1}). All DNA templates were added at 1 nM final concentration. Bars indicate mean values across three replicates. PC, positive control; NC, negative control. Asterisks denote statistically significant differences from a two-way ANOVA with Tukey's multiple comparisons test to compare each SynMG1 variant with the positive control sample (**, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$), ns, not significant. **D)** Ratio of maximum fluorescence and apparent translation rate between the positive control plasmid (PC) and SynMG1 variants. Data points represent averages of three biological replicates.

All fluorescent proteins were successfully expressed from all versions of SynMG1 (Figure 4.3B and Figure S4.2).

Orthogonality of the SP6 RNAP with the T7 promoters on SynMG1 was confirmed by the absence of mCherry expression in the SP6 RNAP samples (Figure S4.3). Orthogonality of the T7 RNA polymerase with the SP6 promoters on SynMG1 could not be determined

by fluorescence measurements because of the overlapping spectra of mVenus and eYFP.

Since SynMG1 should be linearized before it can be replicated by the ϕ29 DNA replication machinery, the effect of SynChr configuration (linear or circular) on expression was tested for all three SynMG1 variants. Preliminary data showed that linearization did not notably increase or decrease expression levels in PURE system (Figure S4.4).

Two kinetic parameters of *mVenus*, *mCherry* and *eYFP* expression were further analyzed: the apparent translation rate and maximum fluorescence (Figure 4.3C). The apparent translation rate (RFU h^{-1}, corresponding to the maximum slope in the linear regime of expression) was estimated using sigmoidal fitting on the fluorescence data [22] (Figure S4.2). Maximum fluorescence (RFU, used as a proxy for the total amount of fluorescent protein produced) was determined by identifying the 100-minute time window with the highest mean fluorescence.

DNA templates and replicates were analyzed as potential sources of variation in the dataset. Since each set of experiments was prepared using a shared PURE master mix for all DNA templates, any variation due to master mix preparation needed to be ruled out. Replicates were a non-significant source of variation in maximum fluorescence and apparent translation rates across the tested conditions (two-way ANOVA on $\log_2$-transformed values, $p > 0.05$). In contrast, DNA templates significantly influenced all datasets ($p \leq 0.01$), except for the apparent translation rate of eYFP ($p > 0.05$).

Pairwise comparisons were conducted for DNA templates (Figure 4.3C). Maximum fluorescence values from SynMG1 variants were significantly higher than the negative control lacking a DNA template, demonstrating successful reporter gene expression (Tukey's multiple comparisons test, $p \leq 0.01$). Positive control templates for mCherry and mVenus exhibited significantly higher maximum fluorescence than SynMG1 variants ($p \leq 0.01$). However, for eYFP, the difference in maximum fluorescence between SynMG1 variants and positive control was not significant ($p > 0.05$). Variation in maximum fluorescence among SynMG1 variants was generally not significant ($p > 0.05$), except for mVenus, where maximum fluorescence from SynMG1.3 was significantly higher than from SynMG1.1 ($p \leq 0.05$). When grouping maximum fluorescence data of all three SynMG1 variants and comparing to control plasmids (Figure 4.3D), a difference was visible between mCherry and mVenus (controlled by *pT7*) and eYFP (controlled by *pSP6*). On average, the maximum mCherry fluorescence in the control sample was 16 times higher than in the samples with SynMG1 variants, and mVenus fluorescence was 15-fold higher for the control. In contrast, for eYFP, the control plasmid exhibited only a twofold higher maximum fluorescence.

Apparent translation rates (RFU h^{-1}) were significantly higher for positive control templates of mVenus and mCherry compared to SynMG1 variants ($p \leq 0.01$). However, no significant difference was observed in the apparent eYFP translation rate between SynMG1 variants and the positive control ($p > 0.05$). Variation in apparent translation rates among SynMG1 variants was not significant for any of the three fluorescent proteins ($p > 0.05$). When comparing apparent translation rates between SynMG1 variants and control plasmids (Figure 4.3D), mCherry was translated on average 15× faster from the control template, while the ratio between the control and SynMG1 variants was on average nine for mVenus and two for eYFP.

Summarizing, these results confirmed the expression of the three reporter genes from SynMG1 templates, and that transcription was RNA polymerase-specific. Expression kinetics and final yields were similar for all three variants, albeit lower than for the control templates for T7RNAP-driven expression.

Mass spectrometry confirms expression of all proteins encoded on SynMG1

SynMG1 coded for 15 proteins, including the fluorescent proteins previously detected in the fluorescence assays. Expression in PURE of all (fluorescent and non-fluorescent) proteins from SynMG1 was assessed by mass spectrometry analysis of the same samples used for fluorescence measurements. For comparison with protein expression of individual modules, control samples were prepared using PURE system with T7 RNAP and plasmids (1 nM) carrying the genes of the individual modules.

Peptide ion intensities detected by mass spectrometry were summed for each protein to calculate raw abundance values, and relative protein abundances were estimated by Median Ratio Fitting, followed by least-squares regression and rescaling to the total sum of ion abundances across runs (Methods). For PURE reactions ran in the presence of T7 RNAP, proteolytic peptides could be detected for all proteins under *pT7* for all SynMG1 variants, albeit with varying coverage levels (Figure 4.4A). The percent coverage of CdsA, a transmembrane protein, is particularly low, while the coverage of most proteins expressed from *pT7* is over 50%. As expected, proteins PgsA and PgpA, under control of *pSP6*, were not detected in these conditions. Principal component analysis of relative protein abundances from biological replicates showed clustering of all replicates of each condition and of the three SynMG1 variants, indicating that relative abundance profiles were consistent and reproducible (Figure 4.4B).

The orthogonality between T7 and SP6 RNA polymerases and their corresponding promoters was further demonstrated by statistically significant changes in protein abundance (Figure 4.4C). Specifically, proteins under control of *pSP6* (PgsA and PgpA) were only expressed with SP6 RNAP, while proteins controlled by *pT7* (all other proteins) were significantly more abundant in samples with T7 RNAP.

Raw abundance values of proteins within each module were compared across SynMG1 variants and control plasmids for samples with T7 RNAP (Figure 4.4D). Overall, abundance was higher for control samples than for SynMG1 variants, as expected from the fewer number of encoded genes. However, this difference was not significant for all proteins (two-way ANOVA on $\log_2$-transformed values with post-hoc Tukey's multiple comparisons test). Additionally, relative protein abundances were plotted for each SynMG1 variant against three control plasmids (Figure S4.5), which corroborates the higher expression levels in samples containing control samples.

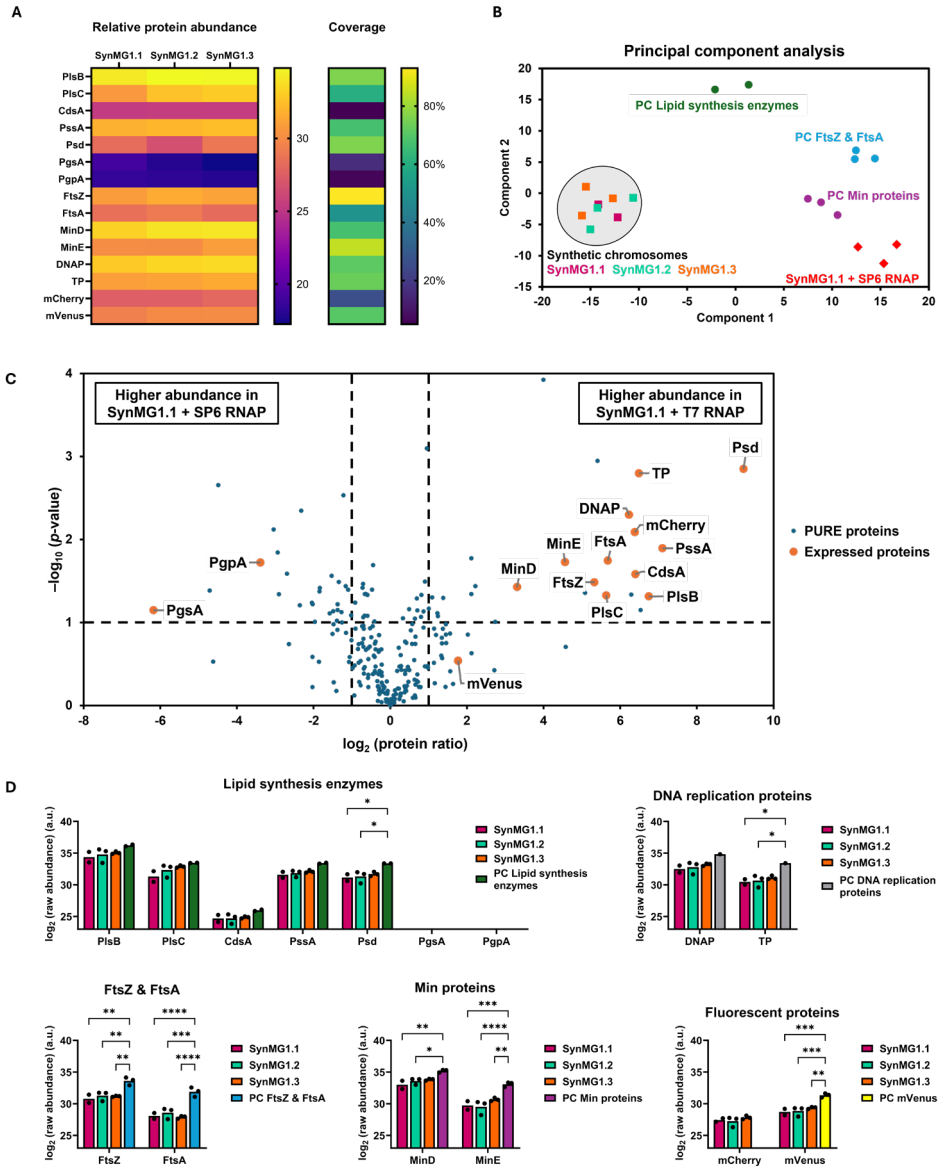


Figure 4.4: Mass spectrometry reveals protein synthesis from the SynMG1 variants. A) Relative abundance of the SynMG1-encoded proteins synthesized in PURE system using T7 RNAP. The percentage of protein sequence coverage was calculated by dividing the number of amino acids in all detected peptides by the total number of amino acids in the entire protein. **B)** Principal component analysis of relative protein abundance. **C)** Volcano plot displaying the change in relative protein abundance between SynMG1.1 transcribed either by T7 RNAP or SP6 RNAP. Vertical lines indicate a 2-fold increase. The horizontal line indicates a *p*-value = 0.1 from a two-tailed *t*-test. **D)** Raw abundance of the proteins synthesized with T7 RNAP, classified by functional modules. Bars indicate mean values. Individual values (*n*) are plotted as dots. Asterisks denote statistically significant differences per protein between DNA templates from a two-way ANOVA with Tukey's multiple comparisons test (*, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ****, $p \leq 0.0001$), all other differences are non-significant. *n* = 1 for the control plasmid for DNA replication proteins, *n* = 2 for SynMG1.1 with T7 RNAP and the control plasmid for lipid synthesis enzymes, and *n* = 3 for all other conditions. PC, positive control.

SynMG1 can be expressed inside liposomes

To emulate gene expression in a synthetic cell, PURE reactions with 1 nM SynMG1.1 were encapsulated in liposomes, and gene expression in the presence of T7 RNAP or SP6 RNAP was monitored by fluorescence confocal microscopy (Figure 4.5A–C). At this DNA concentration, about 20 copies of SynGM1.1 are expected on average per liposome with a diameter of 4 μm if one assumes a Poisson distribution [23]. Fluorescence signals from the mVenus, mCherry and eYFP proteins were observed in some liposomes (Figure 4.5B and C), demonstrating that compartmentalized expression of a 15-gene genome is feasible. A fraction of liposomes containing T7 RNA polymerase exhibited both mCherry and mVenus signals (Figure 4.5C, left).

Flow cytometry was performed to quantify the number of liposomes with fluorescence signal and the distribution of fluorescence intensity within the liposome population (Figure 4.5D). The percentage of liposomes containing SynMG1.1 that showed fluorescence was 64% for mVenus (upper and lower right quadrants), 1% for mCherry (upper left and right quadrants), and 3% for eYFP (upper and lower right quadrants), which is lower compared to the control templates. Control liposomes containing plasmids with a single fluorescent protein showed fluorescence in 89% of liposomes for mVenus, 20% for mCherry, and 69% for eYFP. In the SynMG1.1 sample with T7 RNAP, 1% of liposomes exhibited both mVenus and mCherry signals (upper right quadrant). These results corroborate the conclusion from confocal microscopy analysis that compartmentalization and expression of SynMG1 is feasible, although with large heterogeneity in expression levels across liposomes.

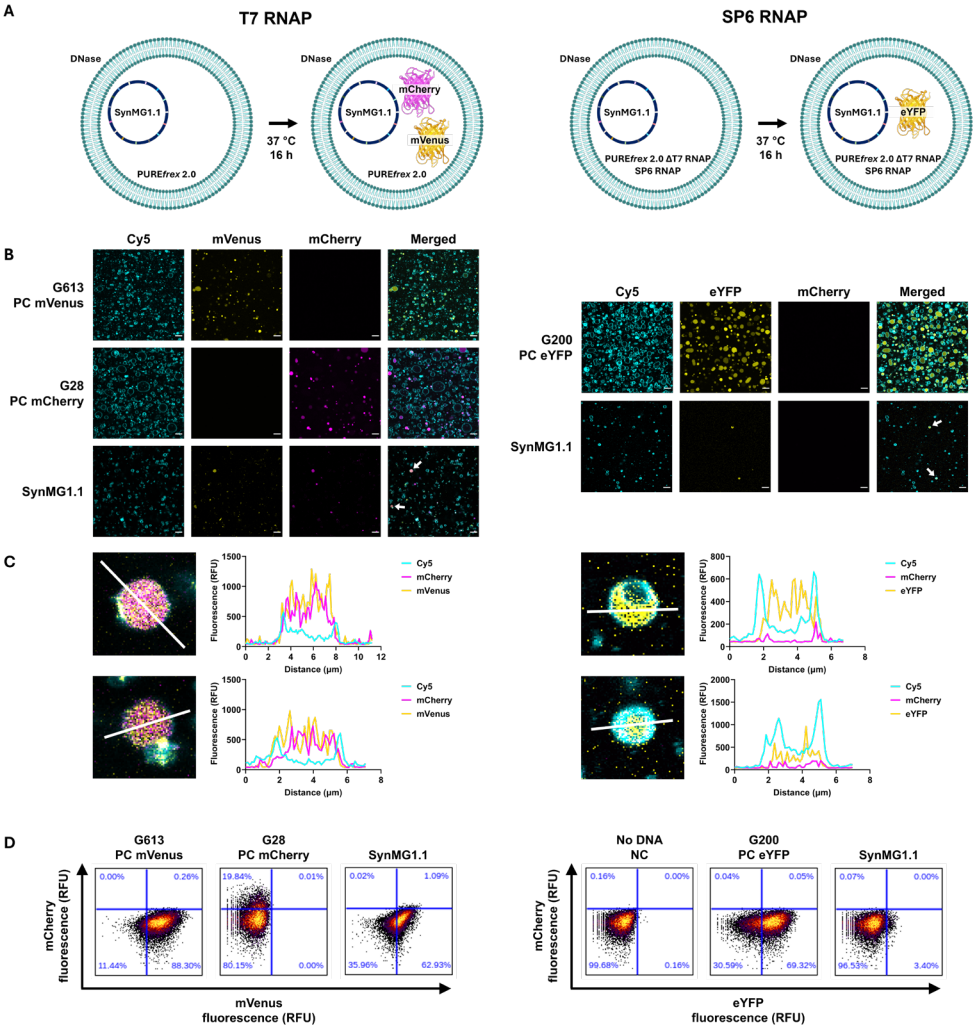


Figure 4.5: In-liposome expression of fluorescent reporter genes from SynMG1.1. A) Schematic illustration of SynMG1 expression in liposomes, directed by T7 RNAP (left) or SP6 RNAP (right). The circular SynMG1.1 was encapsulated together with PURE system. DNase was added to prevent expression outside liposomes. Membrane dye: Cy5. Expected markers to be produced: mVenus and mCherry (left), eYFP (right). **B)** Confocal microscopy images of liposomes containing either T7 RNAP (left) or SP6 RNAP (right) directing expression from control plasmids (mCherry, G28; mVenus, G613; G200, eYFP) or SynMG1.1. Liposomes indicated with a white arrow are highlighted in C). Scale bar is 10 µm. **C)** Highlighted liposomes showing mCherry and mVenus (left) or eYFP signal (right) in the lumen. Fluorescence intensity profiles of Cy5, mCherry and mVenus or eYFP signals were measured along the white line. **D)** Flow cytometry data of liposomes from B). Gates were defined based on a negative control sample (no DNA). PC, positive control; NC, negative control.

Linear SynMG1 can be replicated by the ϕ29 DNA replication machinery

To be self-replicating, minimal cells require a machinery able to replicate large DNA molecules of a few hundred kilobases. The *in vitro* reconstituted ϕ29 system for protein-primed replication has previously been tested with synthetic DNA templates up to 10 kb [24] and with the 19.3-kb ϕ29 genome itself [13], but it has not been tested yet for larger synthetic templates such as the 41-kb long SynMG1. SynMG1.1 was digested with PmeI to generate a linear template flanked by the replication origins oriL and oriR (Figure 4.6A). It was added to PURE system devoid of ribosomes, preventing protein synthesis from SynMG1.1, supplemented with purified DNA replication proteins DNAP, TP, SSB and DSB. (Figure 4.6B). A synthetic 11.6-kb template was included as positive control. To distinguish newly synthesized DNA from the initial template, dTTP was partly substituted by fluorescein-dUTP and the reaction products were visualized on agarose gel. Full-length replication of both DNA templates was confirmed by the presence of bands of the expected size in the fluorescein channel (Figure 4.6C). This result supports the capability of the reconstituted ϕ29 system for replicating multigene DNA templates of up to 41 kb, even in the presence of transcribing T7 RNAP [25].

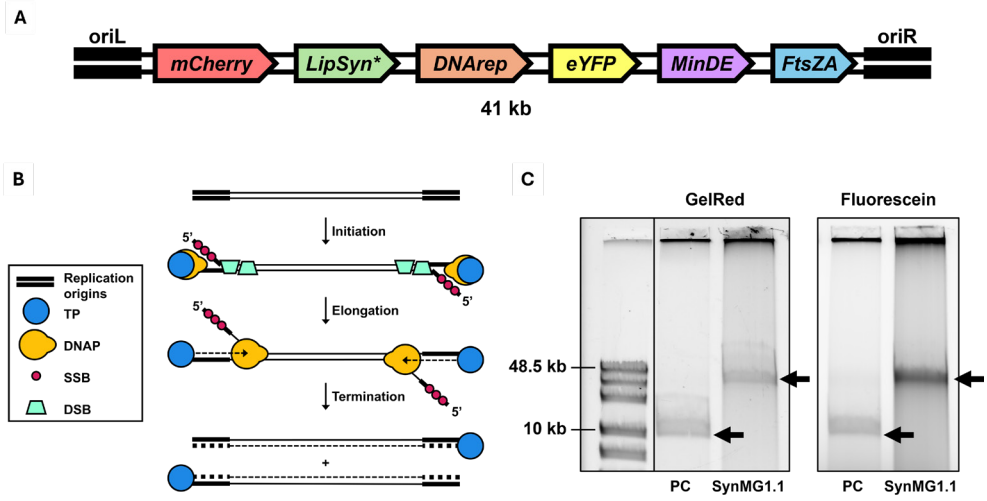


Figure 4.6: Protein-primed replication of linearized SynMG1. A) Schematic representation of SynMG1 linearized by PmeI. The order and direction of the minimal cell modules, including fluorescent reporters, is shown. oriL, left origin of replication; LipSyn, lipid synthesis enzymes; DNArep, DNA replication proteins; MinDE, MinD and MinE; FtsZA, FtsZ:mVenus and FtsA; oriR, right origin of replication. *The two lipid synthesis genes under pSP6 control (*pgsA* and *pgpA*) are encoded in the opposite direction. **B)** Schematic illustration of the protein-primed DNA replication system of bacteriophage ϕ29, requiring a linear template flanked with replication origins. TP, terminal protein; DNAP, DNA polymerase; SSB, single-stranded DNA binding protein; DSB, double-stranded DNA binding protein. Illustration adapted from [13, 24]. **C)** Full-length replication of SynMG1.1 visualized on agarose gel. The GelRed channel shows all DNA, the fluorescein channel shows newly synthesized DNA. Arrows point at the bands corresponding to the template sizes (PC, 11.6 kb; SynMG1.1, 41 kb). PC, positive control.

Discussion

A 15-cistron synthetic minimal genome (SynMG1) encoding multiple (viral and bacterial) proteins involved in synthetic cell modules was successfully assembled in yeast and amplified by *E. coli*. Three SynMG1 variants were isolated and sequence-verified by Oxford Nanopore technology, although a few point mutations were identified. *In vitro* transcription and translation using PURE system enabled the one-pot synthesis of the 15 encoded proteins, demonstrating the feasibility to design, assemble, and express large multigene constructs encoding cellular functions for a synthetic cell. Additionally, preliminary experiments demonstrated successful encapsulation and expression of SynMG1 in liposomes and full-length replication of linear SynMG1 by the protein-primed DNA replication system of ϕ29.

We attempted to expand the minimal genome content by assembling a 105-kb SynChr containing the 47 genes of the combined SynMG1 and pTFM1 plasmid [6]: SynMG2. Assembly in yeast of the 20 DNA fragments led to unintended recombination events, resulting in truncated versions of SynMG2 missing part of the translation factor genes. Current efforts aim at improving the assembly efficiency (see Chapter 5), which will be indispensable to scale up the number of genes (e.g., to encode the 54 ribosomal proteins, 3 ribosomal RNAs, 21 or 48 tRNAs, and 20 aminoacyl-tRNA synthetases), and empowering synthetic cells with more autonomy.

It would be preferred to transfer SynChrs directly from yeast to PURE system without intermediate amplification in *E. coli*, to circumvent possible problems regarding sequence toxicity [26, 27] or recombination in *E. coli* [6] and to streamline the pipeline from SynChr assembly to characterization. However, as shown in Chapter 3, currently, isolation from yeast leads to SynChrs with suboptimal concentration and purity for expression in PURE system. We found that the use of a pCC1BAC backbone with a tunable copy number was suitable for maintenance of SynMG1 in *E. coli* and subsequent isolation, even at low copy number (i.e., no induction). Specifically, 8–16 nM SynMG1 with 35–55% *E. coli* genomic DNA contamination could be obtained at a single SynChr copy, which could be increased to over 35 nM with ca. 9% contamination when the high-copy number origin was activated. In a standard PURE*frex* reaction, 3.5 µL of the 10 µL total volume is available for the DNA template. Consequently, preparing a bulk PURE*frex* reaction at the recommended DNA concentration of 1–2 nM (optimized for single-gene constructs) requires approximately 2–6 nM of SynChr DNA. Alternatively, an initial SynChr concentration of 150 pM is needed to prepare a PURE*frex* reaction with 50 pM DNA template, which is sufficient to achieve, on average, the encapsulation of a single DNA molecule per liposome [23].

Fluorescence readouts confirmed that the three encoded fluorescent proteins could be expressed from SynMG1 in an active form (Figure 4.3), also when the reaction was compartmentalized inside liposomes (Figure 4.5). The latter experiment was only performed once and should be repeated for further analysis of encapsulation efficiency, expression levels, vesicle-to-vesicle heterogeneity and the effect of the DNA template on liposome sample quality.

All encoded proteins could be detected by mass spectrometry (Figure 4.4). Moreover, the fluorescence and mass spectrometry results showed orthogonality between the T7

and SP6 promoters (Figure 4.4C and Figure S4.3).

Although the regulatory sequences of most genes are similar, it is not expected that expression levels are equal, because of some differences in regulatory sequences and the importance of the coding sequence itself for expression levels [28]. Although the eleven T7 promoter sequences that control the expression of 12 genes were identical, variation in protein abundance could be expected. The genes *ftsZ:mVenus* and *ftsA* were together in an operon, controlled by a single *pT7*, which is expected to result in lower expression of *ftsA* than *ftsZ:mVenus* [29]. The three SP6 promoter sequences that control the expression of *eYFP*, *pgpA* and *pgsA* were identical. RBS sequences were the same for all 15 genes, but 5' UTR sequences varied, which is expected to influence protein expression levels [28]. Also, for some genes (*eYFP*, *minD*, *minE*, *ftsZ:mVenus*, *cdsA*, *pssA*, *psd*), a lac operator sequence was present between the promoter and RBS, which may result in a decrease in protein yield (as described in the PURE*frex* 2.0 manual). Transcription of the *p2* gene was terminated by two VSV terminators, whereas transcription of all other genes was terminated by a T7 terminator, and 3' UTR sequences varied between genes, which may influence protein expression levels as well [28]. Accordingly, abundance varied between produced proteins (Figure 4.4). Additionally, expression levels were generally higher from control plasmids than from SynMG1, although the difference was not always significant (Figure 4.4D). This could be explained by the lower number of genes expressed in control samples than SynMG1 samples, therefore a higher availability of transcription and translation resources per gene. Interestingly, when looking at fluorescence readout, the reporters under *pT7* control (mVenus and mCherry) were expressed on average at 15–16× higher levels from the control plasmid than from the SynMG1 variants, while eYFP (under *pSP6* control) was expressed only 2× more from the control. This difference can be explained by the fact that the control plasmids contained a single expression cassette, while SynMG1 contained 11 expression cassettes under *pT7* control and only three controlled by *pSP6*.

The functionality of proteins encoded on SynMG1 remains to be investigated. The absolute or relative amounts of proteins might be too low or imbalanced to support a function, or some proteins for which the coverage of proteolytic peptides was not complete could have undergone impaired translation [7].

Genome-wide engineering can be used to further tune the expression of the synthetic cell modules. This could be realized by targeting the coding or regulatory sequences to improve the rate of transcription or translation of specific genes, or by changing the order or orientation of some cistrons through combinatorial assembly. Genome modifications may directly be implemented in yeast with CRISPR-based tools [30] or inside *E. coli* using recombineering methods [31]. This will create genetic diversity, which, when combined with screening or selection of the genome variants, will establish the basis for directed evolution of system's level functions.

Finally, the genome design must be compatible with the DNA replication mechanism in the minimal synthetic cell. A preliminary experiment demonstrated full-length replication of the linearized 41-kb SynMG1.1 by purified ϕ29 DNA replication proteins. Future experiments should determine the replication fold using quantitative PCR and examine potential interference between transcription and replication caused by collisions of RNA polymerase and DNA polymerase on the DNA template [25]. Furthermore, the

ability of the ϕ29 machinery to replicate longer DNA templates should be assessed. Collectively, these experiments will provide valuable insights into minimal genome design.

Methods

Strains and culture conditions

Saccharomyces cerevisiae strains used and constructed in this study (Table S4.1) were derived from the CEN.PK family [32]. Media used for strain propagation were YPD (Yeast extract Peptone Dextrose), a complex medium composed of 10 g L^{-1} Bacto yeast extract (Gibco, Thermo Fisher Scientific, Waltham, MA, USA), 20 g L^{-1} Bacto peptone (Gibco) and 20 g L^{-1} glucose, or SMD (Synthetic Medium Dextrose), containing 5 g L^{-1} $(NH_4)_2SO_4$, 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} $MgSO_4 \cdot 7H_2O$, 20 g L^{-1} glucose, trace elements and vitamins [33]. Glucose (sterilized at 110 °C for 20 min) and filter-sterilized vitamins were added after setting the pH to 6.0 with 2 M KOH and sterilization for either 20 min at 110 °C for YP or 20 min at 121 °C for SM. Solid media were prepared by addition of 20 g L^{-1} Bacto agar (Gibco) prior to sterilization. To select for the presence of the dominant marker *hphNT1*, YPD was supplemented with hygromycin B (Hyg) to 200 mg L^{-1}. Yeast strains were grown aerobically in 100 mL liquid medium in 500 mL shake flasks at 30 °C and 200 rpm in an Innova 44 incubator shaker (New Brunswick Scientific, Edison, NJ, USA) or on solid media plates at 30 °C.

Escherichia coli strains XL1-Blue (Agilent, Santa Clara, CA, USA), TOP10 (Thermo Fisher Scientific) or DH5α (Thermo Fisher Scientific) were used for molecular cloning and propagation of conventional plasmids. *E. coli* strains DH10B (Invitrogen, Thermo Fisher Scientific) and EPI300 (LGC Biosearch Technologies, Hoddesdon, UK) were used for the construction and propagation of constructs with the CopyControl pCC1BAC vector backbone [34]. SynChrs that were assembled in *S. cerevisiae* contained a pCC1BAC vector backbone and were propagated in EPI300 after isolation from yeast. *E. coli* strains were grown in Lysogeny Broth (LB) medium containing 5 g L^{-1} Bacto yeast extract, 10 g L^{-1} Bacto tryptone (Gibco) and 5 g L^{-1} NaCl, supplemented with 50–100 mg L^{-1} ampicillin (Amp), 50 mg L^{-1} kanamycin (Kan) or chloramphenicol (Cam, 25 mg L^{-1} for high-copy number plasmids, 12.5 mg L^{-1} for constructs with pCC1BAC backbone) if required. Solid LB medium was prepared by addition of 20 g L^{-1} Bacto agar. *E. coli* strains were grown at 37 °C in 10 mL liquid medium in 50 mL centrifuge tubes at 250 rpm in an Innova 4000 incubator shaker (New Brunswick Scientific), in 100 mL liquid medium in 500 mL shake flasks at 200 rpm in an Innova 44 incubator shaker or on solid media plates at 37 °C.

S. cerevisiae and *E. coli* strains were prepared for storage by addition of sterile glycerol (30% (by volume) final concentration) to an overnight grown culture and were stored in 1 mL aliquots at −80 °C.

PCR and clean-up

Amplification of DNA fragments for plasmid construction was performed using KOD Xtreme Hot Start DNA Polymerase (Sigma-Aldrich, Merck, Darmstadt, Germany) according to the manufacturer's protocols, or using Phusion High-fidelity DNA Polymerase (Thermo Fisher Scientific) according to the manufacturer's instructions or with a reduced primer concentration of 0.2 µM. Fragments for assembly in yeast (Table S4.13) were obtained with Phusion, KOD Xtreme or UltraRun LongRange PCR kit (Qiagen, Venlo, The Netherlands), following the manufacturer's instructions except that we lowered the primer concentration to 0.2 µM for Phusion reactions and scaled up the reaction volume to 50 µL for UltraRun LongRange reactions. In case of unsuccessful amplification, primer-dimer formation was reduced by lowering the primer concentration tenfold, addition of 5% DMSO, and performing an initial denaturation step for 3 min and one in the cycling stage for 20 s. Verification of constructed plasmids in *E. coli* was done through colony PCR using GoTaq DNA polymerase (Promega, Madison, WI, USA) according to the manufacturer's instructions, directly using a resuspended colony or with the addition of an extra lysis step. In the latter case, a colony was resuspended in 50 µL Milli-Q and cells were lysed at 95 °C for 5 min, after which 2.5 µL of lysed cells was used in the PCR reaction. Alternatively, plasmids and SynChrs in *E. coli* were verified by colony PCR using DreamTaq PCR Master Mix (Thermo Fisher Scientific) according to the manufacturer's instructions, downscaled to 10 µL reaction volume, using a resuspended colony and with an initial incubation at 95 °C for 10 min to lyse the cells and release the DNA.

Primers used to generate fragments for assembly in yeast were PAGE-purified and purchased from Sigma-Aldrich. All other primers were desalted when shorter than 30 nt and HPLC-purified or PAGE-purified when longer. IMB primers were purchased from Sigma-Aldrich, ChD primers from Ella Biotech (Fürstenfeldbruck, Germany) or Biolegio (Nijmegen, The Netherlands). All primers are listed in Tables S4.8–S4.11.

PCR product size was analyzed by agarose gel electrophoresis on a 0.6–1% agarose gel in 1× TAE buffer, with the GeneRuler DNA Ladder Mix (Thermo Fisher Scientific), BenchTop 1kb DNA Ladder (Promega) or Quick-Load 1 kb Extend ladder (New England Biolabs) as reference. Fragments used for plasmid construction were treated with DpnI (Thermo Fisher Scientific or New England Biolabs) to remove the parental plasmid, by addition of 1 µL DpnI enzyme directly to the 50 µL PCR sample and incubation at 37 °C for 30 min. Purification of PCR products used for cloning was done using InnuPREP PCRpure Kit (IST Innuscreen GmbH, Berlin, Germany) according to the manufacturer's instructions with additional centrifugation for 2 min after discarding the binding buffer, using the Wizard SV Gel Clean-Up System (Promega) following the manufacturer's protocol or using the GeneJET PCR purification kit (Thermo Fisher Scientific) according to the manufacturer's instructions. In case of undesired PCR side products, purification from agarose gel was performed using ZymoClean Gel DNA Recovery kit (Zymo Research, Irvine, CA, USA) or the Wizard SV Gel and PCR Clean-Up System (Promega) according to the manufacturer's instructions. For all kits, DNA was eluted after 5–10 min incubation with 50–60 °C Milli-Q.

All fragments for assembly in yeast (Table S4.13) were obtained by pooling eight to eleven 50 µL PCR reactions, treatment with DpnI and purification from gel to minimize the

chance of transforming yeast with parental plasmids or side products. Extraction from gel was done using the ZymoClean Gel DNA Recovery kit according to manufacturer's instructions with the following modifications: the columns were incubated with the melted agarose solution for at least 30 min before centrifugation, three washes were performed with 2 min incubation of the wash buffer, residual ethanol was removed by 2 min centrifugation of the column after removal of the flowthrough in the last wash step, and the column was incubated for 5–10 min with 50–60 °C Milli-Q prior to elution. Alternatively, ZymoClean Large Fragment DNA Recovery Kit (Zymo Research) was used according to the manufacturer's instructions with the modifications described above, or the QIAquick Gel Extraction Kit (Qiagen) was used according to manufacturer's protocol with an additional wash step with Buffer PE, incubation of the column with Buffer PE for 2 min prior to centrifugation, and 5 min incubation with 50–60 °C Milli-Q before elution of the DNA.

Attempts to extract fragments BG_pLD1_BE, AD_pLD2_AT and AP_pLD3_AR from gel with the ZymoClean Large Fragment DNA Recovery Kit failed, possibly due to retention of a large fraction of the PCR products in the well of the gel. Instead, after DpnI digestion, 50 µL out of 400 µL pooled PCR product was purified with the InnuPREP PCRpure Kit according to the manufacturer's protocol, with additional 5 min incubation of the sample with binding buffer on the spin filter before centrifugation, additional 2 min centrifugation after discarding the binding buffer, and 10 min incubation with 50–60 °C Milli-Q prior to elution. Because of a high percentage of DNA loss with this method, the remaining volume of sample was treated differently: 150 µL was not cleaned up further after DpnI treatment and the remaining 200 µL was precipitated to remove proteins and concentrate the DNA. Precipitation was performed by mixing the sample with 0.1× sample volume 3 M sodium acetate (pH 5.2, Sigma-Aldrich) and 2.5× sample volume ice-cold 100% ethanol and incubation at −80 °C for 2 h 20 min. The sample was centrifuged at 4 °C and 21,100 ×*g* for 30 min. The pellet was washed with 1 mL ice-cold 70% ethanol, centrifuged again at 4 °C and 21,100 ×*g* for 50 min, and air-dried at 37 °C for 40 min. The pellet was attempted to be dissolved in 30 µL Milli-Q at 4 °C for 48 h, but because it was still intact, the samples were put at 60 °C for 2 h, 90 µL extra Milli-Q was added, and the samples were vortexed and pipetted up and down to resuspend the pellet.

In vitro plasmid assembly

Small insertions or mutations were introduced in plasmids using site-directed mutagenesis PCR, where the mutation or insertion was included in the overhang of the primer. The PCR product was circularized in one of the following two ways: (i) Phosphorylated primers were used during PCR and after DpnI digestion and optionally cleanup, the PCR product was self-ligated by T4 ligase (Thermo Fisher Scientific), following the manufacturer's instructions. (ii) Primers without phosphorylation were used during PCR and after DNA extraction from gel, the fragment was circularized by Gibson assembly, following the third protocol described below.

Gibson assembly of one or multiple linear fragments was performed in one of the following three ways. (i) Using the NEBuilder HiFi Master Mix (New England Biolabs) according to the manufacturer's protocol, downscaled to 10 µL reaction volume and

with an incubation of 1 h at 50 °C. (ii) Using the Gibson Assembly Master Mix (New England Biolabs) according to the manufacturer's protocol. (iii) Using a protocol adapted from [35]: A 5× ISO buffer (0.5M Tris-HCl pH 7.5, 50 mM $MgCl_2$, 1 mM of each of the four dNTPs, 50 mM DTT, 25% PEG-8000, 5 mM NAD) was prepared according to the protocol in [35]. An adapted Master Mix was assembled, containing 320 µL 5× ISO buffer, 0.64 µL 10 U $µL^{-1}$ T5 exonuclease, 20 µl 2 U $µL^{-1}$ Phusion Polymerase, 160 µL 40 U $µL^{-1}$ Taq ligase and 700 µL Milli-Q. One 20 µL Gibson Assembly reaction consisted of 15 µL Master Mix with 100 ng of the linear PCR product in Milli-Q. The reaction was incubated at 50 °C for 60 min.

E. coli transformation with plasmids and SynChrs

E. coli was transformed with *in vitro* assembled plasmids for propagation and storage, following one of the protocols described below.

E. coli XL1-Blue cells (Agilent), made chemically competent in-house according to the protocol described in [36], were used for transformation of ligated site-directed mutagenesis PCR products. A mixture of 5 µL ligation product and 50 µL cells was incubated on ice for 5 min. Cells were transformed via heat shock for 45 s at 42 °C, followed by incubation on ice for 2 min and resuspension in 450 µL Super Optimal broth with Catabolic repression (SOC), containing 5 g L^{-1} Bacto yeast extract, 20 g L^{-1} Bacto tryptone, 0.58 g L^{-1} NaCl, 0.19 g L^{-1} KCl, 2 g L^{-1} $MgCl_2 \cdot 6H_2O$, 2.46 g L^{-1} $MgSO_4 \cdot 7H_2O$ and 3.6 g L^{-1} glucose. The cells were incubated for 1 h at 37 °C and 250 rpm, plated on LB agar plates with appropriate antibiotics and grown overnight at 37 °C.

Chemically competent *E. coli* TOP10 cells (Thermo Fisher Scientific) were used for transformation of site-directed mutagenesis PCR products that were circularized by Gibson assembly. To 50 µL of cells, 5 µL of Gibson assembly mix was added and the cells were incubated for 10 min on ice. Heat shock was performed as described above, but replacing 450 µL SOC by 1 mL LB medium.

Chemically competent *E. coli* DH5α cells (Thermo Fisher Scientific) were used for transformation of Gibson Assembly products obtained using the Gibson Assembly Master Mix. The transformation protocol was the same as for TOP10 cells.

E. coli DH10B cells (Invitrogen) were used for electroporation of NEBuilder Hifi assembly products containing a pCC1BAC backbone. The cells were made electrocompetent in-house by growing them in 300 mL LB without NaCl until an OD_{600} of 1.6, incubation on ice for 10 min, and performing multiple rounds of centrifugation for 10 min at 4000 ×*g* and 4 °C followed by resuspension of the cell pellet in 150 mL, 40 mL, 15 mL and 600 µL ice-cold 10% glycerol, respectively. The electrocompetent cells were stored in 50 µL aliquots at −80 °C until use. For transformation, 2 µL of NEBuilder Hifi assembly product was mixed with 50 µL electrocompetent cells and electroporation was performed with the MicroPulser Electroporator (Bio-Rad Laboratories, Hercules, CA, USA) according to the manufacturer's instructions, using 0.2 cm gap Gene Pulser/MicroPulser Electroporation cuvettes (Bio-Rad Laboratories).

E. coli TransforMax EPI300 cells (LGC Biosearch technologies) were used as host for the propagation and storage of SynChrs assembled in yeast, which contain a pCC1BAC backbone. Following extraction of total DNA from yeast, 1 µL DNA was mixed with 50

µL cells and electroporation was performed as described for DH10B. Since the SynChr is the only DNA in the total DNA of yeast that contains a pCC1BAC backbone, it was expected that transformation of *E. coli* results in selective amplification of SynChrs.

Plasmid and SynChr isolation from *E. coli* and verification

Plasmid isolation from *E. coli* was performed using the PureYield Plasmid Miniprep System (Promega) or the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific), according to the manufacturer's instructions. Elution was done in 25 µL 50–60 °C Milli-Q.

Plasmids isolated from *E. coli* were verified by one or multiple of the following methods: (i) colony PCR as described before, (ii) restriction analysis using FastDigest enzymes (Thermo Fisher Scientific) or CutSmart enzymes (New England Biolabs), according to the manufacturer's protocol, and visualization with agarose gel electrophoresis, (iii) Sanger sequencing by Macrogen Europe (Amsterdam, The Netherlands) or (iv) whole-plasmid sequencing using Oxford Nanopore technology by Plasmidsaurus (Eugene, OR, USA). For plasmids pLD1, pLD2 and pLD3, restriction analysis was performed after each new plasmid isolation from glycerol stock, following the recommendations of the Forster lab [6].

SynChr isolation from EPI300 *E. coli* cells was performed using the NucleoBond Xtra Midi kit (Macherey-Nagel, Düren, Germany). No induction of the high-copy number origin on the pCC1BAC backbone was performed, so the SynChr was expected to be maintained at one copy per cell. After cultures were grown directly from colonies on the transformation plate, the resulting isolated DNA consisted of a mixture of recombined and non-recombined SynChrs, as determined by whole-plasmid sequencing. Therefore, extra care was taken to ensure that a single species of SynChr was isolated. Colonies from the transformation plate were streaked to obtain single colonies. Single colonies were resuspended in 15 µL Milli-Q and 1 µL was used per colony PCR reaction to check for the presence of three marker fragments containing the pCC1BAC backbone, *mRuby2* gene and CEN6/ARS4 sequence, using primer pairs 15812/20422, 20333/2306 and 20641/3232, respectively (Table S4.9). The same resuspended colony was used to inoculate 10 mL LB with 12.5 µg mL^{-1} chloramphenicol and the culture was incubated overnight at 37 °C without shaking. The next morning, the 10 mL culture was used to inoculate 300 mL LB with 12.5 µg mL^{-1} chloramphenicol, which was incubated overnight at 37 °C and 200 rpm. Of the final culture, 1–2 mL was used to prepare a glycerol stock. The remaining culture volume was utilized for SynChr isolation with the NucleoBond Xtra Midi kit, starting from an OD_{600} × culture volume (mL) of 800. The manufacturer's instructions were used for purification of low-copy plasmids with the following modifications: extra care was taken to mix gently after addition of Buffer NEU to prevent formation of clouds and thereby clogging the filter-column combination, 15 µL GlycoBlue (Invitrogen) was added after elution with Buffer Elu to ease visualization of the DNA pellet, centrifugation after addition of isopropanol was extended to 45 min, and the DNA pellet was dissolved by addition of 50 µL 50–60 °C Milli-Q and incubation overnight at 4 °C.

Another isolation was performed with high-copy number induction. Single colonies were obtained from the glycerol stock by streaking directly on an agar plate. Colonies were checked by colony PCR and grown overnight in LB with chloramphenicol as de-

scribed before. The next morning, 9 mL of overnight culture was used to inoculate 291 mL LB with 12.5 µg mL^{-1} chloramphenicol and 300 µL of CopyControl solution (LGC Biosearch Technologies) was added. After overnight incubation at 37 °C and 200 rpm, SynChr isolation was performed as described before.

SynChrs sequenced by Plasmidsaurus as big plasmid were analyzed by looking at (i) the consensus sequence, (ii) the histogram with read lengths, and (iii) the raw reads. The consensus sequence was aligned in SnapGene (version 7.0.2, GSL Biotech, San Diego, CA, USA) to the designed SynChr sequence to check for deletions and point mutations. The histogram was analyzed to determine whether a mixture or a single configuration of SynChrs was present. For analysis of the raw reads, a reference FASTA file containing the sequence of CENPK113-7D [37] concatenated with the designed SynChr sequence was prepared. An annotation (.gff) file of the designed SynChr was made by exporting the annotations from SnapGene and using the R package labtools (version 0.1.0, function: *make_gff_from_snap*) [38] in R Statistical Software (version 4.1.2) [39]. The raw reads were mapped with minimap2 (parameters: *-ax map-ont*) [40] to the reference FASTA file and were sorted and indexed using SAMtools [41]. Visualization of the mapped reads was done in the Integrated Genomics Viewer (IGV, version 2.11.9) [42].

Raw reads, consensus sequences and an overview of mutations in the SynMG variants are provided in Supplementary Data.

S. cerevisiae transformation

S. cerevisiae transformation was performed using the high-efficiency lithium acetate/single-stranded carrier DNA/polyethylene glycol method [43] with the following modifications. After washing the cells with 25 mL of sterile water, cells were resuspended in 1 mL 0.1 M lithium acetate (LiAc), the cells were pelleted, the supernatant was removed and 0.1 M LiAc was added to a total volume of 500 µL. The cells were resuspended and divided in 50 µL aliquots per transformation, pelleted again, the supernatant was removed and the transformation mix was added on top of the cell pellet. The transformation mix contained 240 µL 50 % PEG 3350, 36 µL 1M LiAc, 25 µL 2 mg mL^{-1} single-stranded carrier DNA and 50 µL DNA in water (resulting in a total transformation mix volume of 351 µL), extra 30 min incubation at 30 °C was performed before heat shock for 30 min at 42 °C and the cells were incubated 1–2 h at 30 °C in 1 mL YPD before plating on YPD + Hyg agar plates.

Total DNA extraction from *S. cerevisiae*

Total DNA extraction from *S. cerevisiae* for whole-genome sequencing and transformation of *E. coli* for SynChr propagation was performed using the Qiagen Blood & Cell Culture Kit with 100/G Genomic-tips (Qiagen) according to the manufacturer's protocol for yeast samples.

DNA analysis

The NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific) was used for determination of DNA purity and concentration. DNA concentration was additionally measured with a Qubit 2.0 or 4.0 Fluorometer (Invitrogen, Thermo Fisher Scientific) using the Qubit dsDNA Broad Range Assay kit (Invitrogen).

Plasmid construction in *E. coli*

All plasmids used and constructed in this study are listed in Tables S4.4–S4.6. Primers used for plasmid construction and verification by colony PCR or Sanger sequencing are listed in Tables S4.8–S4.10.

Plasmids pUDC436, pUD1386 and pUD1394 (Table S4.4) were constructed through site-directed mutagenesis PCR with phosphorylated primers, self-ligation with T4 DNA ligase, and heat-shock transformation of *E. coli* XL1-Blue. Plasmid pUDC436 was obtained through insertion of the ARS1 sequence downstream of the *mTurquoise2* expression cassette in plasmid pUDC192 with primer pair 20431/20432. The plasmid was verified by colony PCR using primer pairs 20433/20434, 20433/20436 and 20435/20434, and its sequence was confirmed by whole-plasmid sequencing. Plasmid pUD1386 was constructed by insertion of the ARS417 sequence downstream of the *crtI* expression cassette in pUD1250 using primers 20425/20426. Verification was done by colony PCR with primer pairs 20427/20428, 20427/20430 and 20429/20428, and by whole-plasmid sequencing. Plasmid pUD1394 was designed to include an I-SceI cut site and a landing pad upstream of the CEN6/ARS4 sequence of pYTK081. The I-SceI cut site was included to allow linearization of the SynChrs and subsequent visualization on a CHEF gel. The landing pad was introduced to allow easy CRISPR/Cas9 editing of the SynChrs after assembly in yeast, and consists of two gRNA target sites ("Cas9 Target Site 9 (T9)" from [16] and "sTarget#2" from [17]). The first site is flanked by two 60-bp non-coding sequences without homology to the yeast genome. The second gRNA target site is located downstream of the right flank, and is directly followed by the CEN6/ARS4 sequence, to allow exchange of the centromeric region. pUD1394 was constructed using primer set 20522/20523 for site-directed mutagenesis PCR on pYTK081. The insertion was verified by colony PCR with primer set 19934/20428 and the plasmid sequence was confirmed with whole-plasmid sequencing.

Plasmids G162, G197 and G200 (Table S4.5) were constructed through site-directed mutagenesis PCR, Gibson assembly to circularize the linear product following the protocol adapted from [35], and heat-shock transformation of *E. coli* TOP10. Plasmid G162 was constructed by replacing the DraI cut site in pETORPHI by a PmeI cut site using primer set 683/684. The modification was confirmed by Sanger sequencing using primer 288 and the plasmid sequence was determined by whole-plasmid sequencing. Plasmid G197 was obtained by insertion of a *lacO* site upstream of the *yfp* gene of plasmid G76 using primer pair 715/716, and verified by Sanger sequencing with primer 365. Plasmid G200 was constructed through replacement of *pT7* in G197 by *pSP6*, using primer set 719/720. Verification of the promoter sequence was done by Sanger sequencing with primer 365 and the plasmid was sequenced using whole-plasmid sequencing.

Plasmid pUD1387 (Table S4.4) was constructed through Gibson assembly using NEBuilder HiFi Master Mix and electroporation of *E. coli* DH10B. The *pURA3-URA3-tURA3* fragment was amplified from pYTK074 using primers 20419/20420 and was inserted into the pCC1BAC backbone, amplified from pCC1BAC-lacZα using primers 20417/20418. The constructed plasmid was verified by colony PCR with primer sets 20421/20422, 20421/20424 and 20423/20422, and by whole-plasmid sequencing.

Plasmids G607 and G613 (Table S4.5) were constructed through Gibson assembly using the Gibson Assembly Master Mix and heat-shock transformation of *E. coli* DH5α. Plasmid G607 was designed to encode a fusion protein of FtsZ with the fluorescent reporter mVenus. For construction of G607, plasmid G379 containing *pT7-ftsZ-tT7* was linearized by PCR using primers 1491/1492, thereby splitting the *ftsZ* coding sequence between glycine 55 and glutamine 56 and inserting the linkers GSTLE and LEGST downstream and upstream of each respective residue [44]. The *mVenus* coding sequence was amplified from pWKD014 using primers 1493/1494, thereby flanking the fragment with GSTLE and LEGST linkers on its 5' and 3' ends, respectively. After Gibson assembly of the two fragments, the resulting G607 plasmid was verified by colony PCR using primer set 770/797. Plasmid G613 was designed to contain an operon encoding the fusion protein FtsZ-mVenus and FtsA. The *ftsA* fragment was amplified from G385 with primers 709/1508 and the resulting PCR product included a 15-bp spacer sequence with 30% GC content, generated via the Random DNA generator provided by the Maduro Lab (https://faculty.ucr.edu/~mmaduro/random.htm), upstream of the T7 gene 10 leader sequence and RBS of the expression cassette. Plasmid G607, containing *pT7-ftsZ:mVenus-tT7*, was linearized by PCR using primers 452/1511 and contained the same 15-bp spacer sequence downstream of the *ftsZ:mVenus* stop codon. After Gibson assembly of the two fragments, correct insertion was verified by colony PCR with primer set 984/1211. The sequence of G613 was verified by whole-plasmid sequencing.

Assembly of SynChrs in yeast

Two synthetic chromosomes were constructed in this study (Table S4.2): SynMG1 and SynMG2. Strain CEN.PK113-5D was used for assembly of SynMG1 from 14 fragments and strain CEN.PK102-12A for assembly of SynMG2 from 20 fragments. Host strains are listed in Table S4.1 and assembly fragments in Table S4.13.

Fragments for SynMG1 assembly include six fragments containing genes for expression in PURE system: the DNA replication genes *p2* and *p3*, the Min system genes *minD* and *minE*, the division protein genes *ftsZ:mVenus* and *ftsA*, the phospholipid synthesis genes *plsB*, *plsC*, *cdsA*, *pssA*, *psd*, *pgsA* and *pgpA*, and two fragments with the fluorescent marker genes *eYFP-spinach* and *mCherry*. All genes are under the control of a T7 promoter, except *pgsA*, *pgpA* and *eYFP-spinach*, which are controlled by an SP6 promoter. The genes *ftsZ:mVenus* and *ftsA* are combined in an operon. Additionally, a fragment containing the oriR and oriL sequences with an internal PmeI restriction site is included, which allows linearization of the chromosome and replication with the ϕ29 DNA replication machinery. The remaining seven fragments contain sequences for amplification, screening and selection in *S. cerevisiae*: a yeast centromere and replication origin (CEN6/ARS4), and yeast screening and selection markers (*crtE*, *crtI*, *crtYB*,

mRuby2, *hphNT1* and *URA3)*. The CEN6/ARS4 fragment contains a landing pad and I-SceI restriction site for editing and linearization of the chromosome. The *crtI* fragment contains an additional yeast replication origin: ARS417. The *URA3* gene is combined on one fragment with a pCC1BAC backbone containing an antibiotic marker (*cat*) for amplification and selection in *E. coli*. SHR overhangs of 60 bp (Table S4.12) were included in the PCR primers used to amplify these fragments (Table S4.11) to allow assembly in yeast by homologous recombination.

Assembly of SynMG2 was done with the same 14 fragments used for SynMG1 construction (differing in SHR sequence when necessary), and six additional fragments. Three fragments contain genes encoding for the translation factors of PURE system (with the exception of EF-Tu) and were amplified from the plasmids pLD1 (*prfA*, *hisS*, *tyrS*, *cysS*, *trpS*, *serS*, *valS*, *metG*, *argS*, *glnS*, *leuS*, *thrS* and *lysS*), pLD2 (*asnS*, *ileS*, *alaS*, *pheS*, *pheT*, *fmt*, *tsf*, *infC* and *infA*) and pLD3 (*aspS*, *proS*, *glyQ*, *glyS*, *gltX*, *prfC*, *prfB*, *frr*, *fusA* and *infB*) [6]. All translation factor genes are controlled by a T7 promoter. The genes *pheS* and *pheT* are combined in an operon, as well as the genes *glyQ* and *glyS*. The other three fragments contain extra markers for screening and selection in yeast (*LEU2*, *HIS3* and *mTurquoise2*). The *mTurquoise2* gene and the additional yeast replication origin ARS1 were combined on the same fragment.

All assembly fragments were cleaned up by DpnI digestion and gel extraction, as described in the section “PCR and clean-up”, with the exception of BG_pLD1_BE, AD_pLD2_AT and AP_pLD3_AR used for SynMG2 assembly, for which only DpnI digestion was performed. A mix of DNA fragments was prepared, containing 100 fmol of the CEN6/ARS4 and *hphNT1* fragments and 150 fmol of all other fragments. For SynMG1, this fitted within the standard 50 µL volume used for transformation. For SynMG2, the total volume of pooled fragments was 78.17 µL. To ensure that the concentrations of the components in the transformation mix remained unchanged, all component solutions were scaled up to have a total transformation mix volume of 548.76 µL instead of 351 µL, and 78.17 µL of cells were used instead of 50 µL. After transformation, cells were plated on YPD + Hyg pates and incubated for three days at 30 °C. Plating on YPD + Hyg plates selects for transformants which have taken up both the *hphNT1* fragment, conferring resistance to hygromycin B, and the CEN6/ARS4 fragment, placed on opposite locations on the SynChrs. Transformants were screened based on (i) β-carotene production, visible on plate as orange colonies, (ii) fluorescence measured by flow cytometry, and (iii) auxotrophy by streaking on SMD. The sequence of the SynChrs was determined by long-read sequencing of total DNA extracted from yeast.

For SynMG1, all stocked yeast strains (Table S4.3) were obtained from one transformation with the abovementioned method.

For SynMG2, the transformation described above (transformation 1) did not result in any strains testing positive for all yeast markers and only IMF95 was stocked (Table S4.3). Two additional transformations were performed with modifications. For one (transformation 2), the transformation mix consisted of 100 fmol of all fragments and the components of the transformation mix were not rescaled despite a total DNA volume of 58.52 µL. No colonies grew on the transformation plates. For the other (transformation 3), the transformation mix consisted of 100 fmol of all fragments, the DpnI-digested fragments BG_pLD1_BE, AD_pLD2_AT and AP_pLD3_AR were precipitated before use,

and the total DNA volume did not exceed 50 µL. The transformation plate contained several colonies, of which two were stocked as IMF96 and IMF97 (Table S4.3).

Assembly screening and sequencing

Fluorescence detection by flow cytometry

Yeast colonies containing assembled SynChrs were checked for mRuby2 (SynMG1 and SynMG2 colonies) and mTurquoise2 (SynMG2 colonies) fluorescence by flow cytometry using the BD FACSAria II Cell Sorter with the BD FACSDiva software (BD Biosciences, Franklin Lakes, NJ, USA). Before every experiment, a cytometer Setup and Tracking (CS&T) cycle was run according to the manufacturer's instructions. Each colony was resuspended in 500 µL synthetic medium without glucose and vitamins for direct measurement, or in 500 µL complete SMD followed by 4 h incubation at 30 °C at 200 rpm to wake up the cells and ensure active expression of the fluorescent markers before measurement. The 70 µm nozzle was used and at least 50,000 events per sample were recorded. A 561 nm laser with a 582/15 nm bandpass emission filter was used for detection of mRuby2, and a 445 nm excitation laser with a 525/50 nm bandpass filter for mTurquoise2 detection. Data analysis was performed with the Cytobank software (Beckman Coulter). The negative control strain CEN.PK113-7D and positive control strain IMF2 (Table S4.1) were used to determine the gates to distinguish fluorescent cells from non-fluorescent ones.

Auxotrophy detection by restreaking on SMD

The presence of the *URA3* fragment (SynMG1 and SynMG2) and *LEU2* and *HIS3* fragments (SynMG2) was determined by picking colonies from the YPD + Hyg transformation plate and streaking them on SMD. No growth on SMD indicated absence of at least one auxotrophic marker.

Long-read sequencing

Yeast strains testing positive for (almost) all markers were selected for long-read sequencing. Total DNA was extracted from yeast and sent to Plasmidsaurus to be sequenced using the "Big bacterial genome" service. The resulting .fna file containing all contigs was opened in SnapGene and the contig corresponding to the SynChr was identified by detection of common features. Some SynChr contigs were predicted to be linear by the analysis pipeline of Plasmidsaurus (involving *de novo* assembly of reads using Flye), which is expected to be incorrect. Whenever doubts were present about the accuracy of the consensus sequence, raw reads were analyzed as described for SynChrs isolated from *E. coli*.

Raw reads of the total extracted DNA, consensus sequences of SynMG variants and an overview of mutations in the SynMG variants are provided in Supplementary Data.

After transfer of the SynChrs from *S. cerevisiae* to *E. coli* and isolation from *E. coli*, the SynChrs were again sequenced by Plasmidsaurus, as described in the section "Plasmid and SynChr isolation from *E. coli* and verification".

Cell-free gene expression

In vitro transcription-translation of SynChrs isolated from EPI300 *E. coli* was performed using PURE*frex*2.0 (GeneFrontier Corporation, Kashiwa, Japan), following the manufacturer's instructions for storage and handling. Cell-free expression was performed in bulk (test tube) reactions and in liposomes.

Bulk reactions

Bulk reactions were prepared in standard PURE*frex* 2.0, containing T7 RNAP for transcription, or in custom PURE*frex* 2.0 without T7 RNAP and with addition of SP6 RNAP (Promega). PURE reactions with T7 RNAP were prepared by mixing 5 µL solution I (buffer), 0.5 µL solution II (enzymes), 1 µL solution III (ribosomes), 5 U SUPERase·In (Thermo Fisher Scientific), DNA template and Milli-Q in a final volume of 10 µL. PURE reactions with SP6 RNAP were prepared by mixing 5 µL solution I, 0.5 µL solution II ΔT7 RNAP, 20 U SP6 RNAP, 1 µL solution III, 5 U SUPERase·In, DNA template and Milli-Q in a final volume of 10 µL. DNA templates were measured by Qubit and added in a final concentration of 1 nM (SynChrs and control plasmids) or 0.1 nM (control plasmids). Plasmids G28, G200, G276, G396, G435 and G613 (Table S4.6) were used as control plasmids for expression of individual modules. All reactions were performed in biological triplicate. To assess whether a circular or linear SynChr configuration had an effect on expression levels in PURE system, SynChrs were linearized by I-SceI (Thermo Fisher Scientific) in a restriction reaction at 3.9 nM final DNA concentration, following the manufacturer's protocol. Control samples containing circular SynChrs were prepared in the same way, without addition of the I-SceI enzyme. Linearization was verified by comparison of the linearized and circular SynChr samples on a 0.6% agarose gel in 1× TAE buffer, with the GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific) as reference. Bulk reactions with T7 RNAP were prepared as described above with linear or circular SynChrs at 1 nM final concentration, or with circular control plasmids G28 and G613 at 1 nM or 0.1 nM final concentration. No replicates were performed.

Reactions were assembled in PCR tubes and incubated at 37 °C for 16 h, either in PCR tubes in a thermocycler (samples containing control plasmids G276, G396, G435 or G613) or after transfer to a 384-well plate in a spectrofluorometer (samples containing SynChrs or control plasmids G28, G200 or G613) as described below. After incubation, samples were flash-frozen in liquid nitrogen and stored at −80 °C.

Spectrofluorometry

Bulk reactions containing SynChrs or control plasmids G28, G200 or G613 were transferred to a black 384-well flat µClear bottom microplate (Greiner Bio-One, Kremsmünster, Austria), which was sealed with a highly transparent film (Sarstedt, Nümbrecht, Germany). The plate was incubated at 37 °C for 16 h in a Synergy H1 Multi-Mode Microplate Reader (Agilent) and bottom fluorescence was measured every 5 min using an excitation bandwidth of 500/20 nm and emission bandwidth of 539/20 nm for detection of eYFP and mVenus fluorescence, and with excitation at 579/20 nm and emission at 616/20 nm for mCherry detection. The read height was set to 9.5 mm and a gain of 100 was used for mCherry and 50 for eYFP and mVenus.

Fluorescence data were analyzed using a custom Python script that automated processing and plotting of kinetic curves (Supplementary Data). Gene expression kinetics were modeled using a sigmoid function based on [22] to fit the experimental data, allowing extraction of the maximum apparent translation rate (RFU h^{-1}). Maximum fluorescence (RFU) was determined using a sliding window approach, identifying the 100-min time period with the highest average fluorescence. The maximum fluorescence was then calculated by averaging the fluorescence values of the 20 data points within this window. A two-way ANOVA with post-hoc Tukey testing was performed on $\log_2$-transformed maximum fluorescence and apparent translation rate values to analyze variation across replicates and DNA templates.

LC-MS

Sample preparation for label-free proteomics analysis

Bulk reactions containing SynChrs (T7 RNAP and SP6 RNAP samples) or control plasmids G276, G396, G435 or G613 (T7 RNAP samples only) were analyzed by liquid chromatography mass spectrometry (LC-MS). Ten microliter of protein samples were processed for trypsin digestion by addition of ten microliter of trypsin (500 ng) in ammonium bicarbonate (100 mM) to each sample for overnight digestion. The reaction was stopped by adding 1.5 µL 1% trifluoroacetic acid (TFA).

NanoLC-MS/MS analysis of proteins

Tryptic peptides were analyzed by nano-LC coupled to tandem MS, using an UltiMate 3000 system (NCS-3500RS Nano/Cap System, Thermo Fisher Scientific) coupled to an Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher Scientific). Five microliter of sample was injected on a C18 precolumn (300 µm inner diameter × 5 mm, Thermo Fisher Scientific) in a solution consisting of 2% acetonitrile and 0.05% TFA, at a flow rate of 20 µL min^{-1}. After 5 min of desalting, the precolumn was switched online with the analytical C18 column (75 µm inner diameter × 50 cm; in-house packed with Reprosil C18) equilibrated in 95% solvent A (5% acetonitrile, 0.2% formic acid) and 5% solvent B (80% acetonitrile, 0.2% formic acid). Peptides were eluted using a 10–50% gradient of solvent B over 105 min at a flow rate of 300 nL min^{-1}. The mass spectrometer was operated in data-dependent acquisition mode with the Xcalibur software. MS survey scans were acquired with a resolution of 70,000 and an AGC target of 3×10^6. The ten most intense ions were selected for fragmentation by high-energy collision-induced dissociation, and the resulting fragments were analyzed at a resolution of 17,500 using an AGC target of 1×10^5 and a maximum fill time of 50 ms. Dynamic exclusion was used within 30 s to prevent repetitive selection of the same peptide.

Bioinformatics analysis of MS raw files

Raw MS files were processed with the Mascot software (version 2.7.0) for database search and Proline for label-free quantitative analysis (version 2.1.2). Data were searched against *E. coli* entries of the UniProtKB protein database release Swiss-Prot 2019_11 (23,135 entries) and homemade database (built with FASTA sequences of expected proteins mVenus from *Aequorea victoria*, DNAP and TP from *Bacillus* phage ϕ29 and mCherry from *Anaplasma marginale*). Carbamidomethylation of cysteines

was set as a fixed modification, whereas oxidation of methionine was set as variable modifications. Specificity of trypsin/P digestion was set for cleavage after K or R, and two missed trypsin cleavage sites were allowed. The mass tolerance was set to 10 ppm for the precursor and to 20 mmu in tandem MS mode. Minimum peptide length was set to 7 amino acids, and identification results were further validated in Proline by the target decoy approach using a reverse database at both a PSM and protein false-discovery rate of 1%. For label-free relative quantification of proteins across biological replicates and conditions, cross-assignment of peptide ion peaks was enabled within each group with a match time window of 1 min, after alignment of the runs with a tolerance of +/- 600 s. Raw abundance values were visualized as bar plots. A two-way ANOVA with post-hoc Tukey testing was performed on $\log_2$-transformed raw abundance values to analyze variation across DNA templates and proteins.

A matrix of abundance ratios was generated for each pair of runs using Median Ratio Fitting, based on ion abundances for each protein. For each pairwise comparison, the median of the ion abundance ratios was then calculated and used to represent the protein ratio between these two runs. A least-squares regression was performed to estimate the relative abundance of the protein across all runs in the dataset. Finally, these abundance values were rescaled to the total sum of the ion abundances across runs.

To assess differences in protein abundance between biological groups, a two-tailed Student's *t*-test (equal variances) was performed on $\log_2$-transformed abundance values. Significance level was set at $p = 0.1$ and ratios were considered relevant if higher than +/- 2. Results were visualized using volcano plots.

Lastly, principal component analysis (PCA) was conducted on the rescaled abundance values to identify potential outliers.

LC-MS data underlying Figure 4.4 and Figure S4.5 is provided in Supplementary Data.

Liposomes

Liposome preparation

Lipid-coated beads were prepared as described in [13, 22] with minor modifications. Glass beads (0.6 g, Sigma-Aldrich) were coated with 2 mg of 50 mol% DOPC, 36 mol% DOPE, 12 mol% DOPG, 2 mol% 18:1 cardiolipin, 0.05% (by mass) DOPE-Cy5, and 1% (by mass) DSPE-PEG-biotin. All lipids were from Avanti Polar Lipids.

PURE*frex* 2.0 reactions with T7 RNAP or SP6 RNAP were prepared in 1.5 mL tubes as described above for bulk reactions, but scaled up to 20 µL total volume and at a final DNA concentration of 1 nM for SynChrs and control plasmids G28, G200 or G613. Lipid-coated beads were desiccated for at least 30 min before use and 11–12 mg of beads was added to each PURE sample. The samples were gently rotated in an automatic tube rotator (VWR) for 1 h at 4 °C to allow liposome swelling, followed by four cycles of freezing by dipping in liquid nitrogen for 10 s and thawing for 5–10 min on ice. About 12 µL of the upper solution in each liposome sample was transferred to a PCR tube containing 0.5 µL DNase I (2 U μL^{-1}, New England Biolabs), using a cut pipette tip, and the solution was gently pipetted up and down twice for mixing. Samples were incubated in a thermocycler for 16 h at 37 °C.

Confocal microscopy

Custom-made glass chambers were washed 3× with 10 µL Milli-Q, functionalized with BSA (Sigma-Aldrich, 1 mg mL^{-1} in Milli-Q) by incubation for 10 min and washed again 2× with 10 µL Milli-Q. To each well, 10 µL homemade PURE buffer (180 mM potassium glutamate, 14 mM magnesium acetate, 20 mM HEPES-KOH at a pH of 7.6) was added. Liposome samples were diluted by mixing 5 µL homemade PURE buffer with 7 µL liposome sample using a cut pipette tip and slowly pipetting up and down three times. All PURE buffer was pipetted out of the glass chambers and immediately 12 µL of diluted liposome sample was added. Chambers were sealed using a silicone spacer sheet and a cover slip. Microscopy was carried out on the A1R Laser scanning confocal microscope (Nikon) with an SR Apo TIRF 100× oil immersion objective with a pinhole size of 42.1 µm (1 A.U. for 405 nm). Imaging was performed using a 488 nm excitation laser with 520/50 nm bandpass emission filter for mVenus and eYFP, a 561 nm excitation laser with 595/50 nm bandpass emission filter for mCherry and a 640 nm laser with 700/75 nm bandpass emission filter for Cy5. Laser power and photomultiplier tube (PMT) voltages were adjusted per sample. Image acquisition was performed at room temperature with the software NIS-Elements version 5.30.07 (Nikon).

Confocal images were processed using Fiji (ImageJ version 1.53c) [45, 46]. Brightness was adjusted for each channel separately to improve visibility of the fluorescence signal and the three channels were combined into a composite image.

Flow cytometry

Liposome samples were prepared for flow cytometry by mixing 2 µL liposome sample with 300 µL homemade PURE buffer using a cut pipette tip. To remove possible remaining glass beads, the solution was filtered through a 35-µm mesh of a cell-strainer cap from a 5-mL round-bottom polystyrene test tube (Falcon, Corning, NY, USA). Samples were measured on a BD FACSMelody Cell Sorter (BD Biosciences) using a 100 µm nozzle. A 488 nm excitation laser with 527/32 nm bandpass emission filter was used for detection of mVenus and eYFP, and a 561 nm excitation laser with 613/18 nm bandpass emission filter for detection of mCherry. PMT voltages were kept constant across samples and 20,000 events were recorded per sample. Data analysis was performed with the Cytobank software to filter out possible aggregates and liposome debris, as previously described [23]. Gates to distinguish fluorescent liposomes from non-fluorescent ones were determined using a liposome sample that did not contain DNA template.

DNA replication

DNA replication of SynMG1 linearized by PmeI was performed using purified ϕ29 DNA replication proteins in an adjusted PURE*frex* 2.0 reaction, supplemented with the required substrates and cofactors for DNA replication. The concentration of tRNAs and NTPs was changed in PURE*frex* 2.0 (see the modified composition below) to allow DNA replication, and no ribosomes were added to prevent translation. Linear DNA templates used in DNA replication experiments are listed in Table S4.15. Replication fold was analyzed by quantitative PCR (qPCR) and full-length replication was verified by agarose gel electrophoresis.

Preparation of DNA templates

SynMG1.1 isolated from EPI300 *E. coli* (pUDF006) and the control plasmid G363 were linearized by PmeI (Thermo Fisher Scientific) following the manufacturer's instructions, at a final DNA concentration of 100 ng μL^{-1}. Linearization was verified by comparison of the linearized samples with the circular templates on a 0.6% agarose gel in 1× TAE buffer, with the Quick-Load 1 kb Extend ladder as reference. Linear SynMG1 and G363 were not purified from the restriction reaction. DNA concentrations were quantified by Qubit.

Purified ϕ29 DNA replication proteins

Purified ϕ29 DNA replication proteins were produced as described in [13]. Stock concentrations and storage buffers were: DNAP (420 ng μL^{-1} in 25 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 0.5 mM EDTA, 3.5 mM β-mercaptoethanol (BME), 0.025% Tween 20, 50% glycerol), TP (320 ng μL^{-1} in 50 mM Tris-HCl (pH 7.5), 0.5 M NaCl, 1 mM EDTA, 7 mM BME, 50% glycerol), SSB (10 mg mL^{-1} in 50 mM Tris-HCl (pH 7.5), 60 mM $(NH_4)_2SO_4$, 1 mM EDTA, 7 mM BME, 50% glycerol), DSB (10 mg mL^{-1} in 50 mM Tris-HCl (pH 7.5), 450 mM $(NH_4)_2SO_4$, 1 mM EDTA, 7 mM BME, 50% glycerol). The proteins were aliquoted and stored at −80 °C.

Reaction conditions

DNA replication reactions were prepared by mixing 5 µL custom PURE*frex* 2.0 solution I (ΔtRNAs, ΔNTPs), 0.33× dNTPs, 0.33× tRNAs, 1 µL PURE*frex* 2.0 solution II, 12 U SUPERase·In, 20 mM $(NH_4)_2SO_4$, 3 ng μL^{-1} DNAP, 3 ng μL^{-1} TP, 375 ng μL^{-1} SSB, 105 ng μL^{-1} DSB, 300 µM dCTP, 300 µM dGTP, 300 µM dATP, 180 µM dTTP, 120 µM fluorescein-12-dUTP (Thermo Fisher Scientific), 100 ng DNA template and Milli-Q in a final volume of 20 µL. Reactions were assembled in PCR tubes and incubated at 30 °C in a thermocycler for 16 h.

Visualization on agarose gel

After incubation, 10 µL of each sample was transferred to a new PCR tube and incubated with 0.2 mg mL^{-1} RNase A (Promega) at 30 °C in a thermocycler for 1.5 h to remove RNA. The reaction was quenched by addition of 6 µL STOP solution (30 mM EDTA, 0.3% SDS) and proteins were removed using 1 mg mL^{-1} Proteinase K (Thermo Fisher Scientific) and 1 h incubation at 50 °C in a thermocycler. The samples were stored in the dark at 4 °C until visualization on gel.

The samples were analyzed by agarose gel electrophoresis on a 0.6% agarose gel without DNA stain in 1× TAE buffer, with the Quick-Load 1 kb Extend ladder as reference. The gel was run for 90 min at 60 V and visualized on a fluorescence gel imager (Typhoon, Amersham Biosciences) using a 488 nm excitation laser with 525/20 nm bandpass emission filter to detect fluorescein-labeled DNA. Post-staining was done in GelRed Nucleic Acid Stain (Millipore, Merck, Burlington, MA, USA), according to the manufacturer's instructions, and the gel was washed twice for 5 min in Milli-Q. Total DNA was visualized on the fluorescence gel imager using a 532 nm excitation laser with a 570/20 nm bandpass filter to detect GelRed.

Acknowledgements

We thank the following people for constructing and sharing plasmids: Kristel Doets from Wageningen University and Research (pCC1BAC-lacZα), Ramon van de Valk and Liedewij Laan (pWKD014), Elynor Moore (pUD1394), Sophie van der Horst (G162, G197 and G200) and Federico Ramírez Gómez (G607 and G613).

We thank Laura Sierra Heras for help with the assembly primer design, Marcel van den Broek for help with sequencing data analysis, Yannick Bernard-Lapeyre for calculation of the maximum fluorescence and apparent translation rates from fluorescence data, Alexandre Stella and Etienne Bancal (ProteoToul Services, Institut de Pharmacologie et de Biologie Structurale, Toulouse) for performing and analyzing the LC-MS experiment, and Marijn van den Brink for help with liposome preparation, confocal microscopy and flow cytometry.

I would like to thank Ellen Zwiers for her great work on isolation and characterization of the SynChrs described in this chapter. Her enthusiasm, eagerness to learn and efficient laboratory work contributed greatly to the progress of this project.

Images in Figures 4.2 and 4.5 were (partially) created with BioRender.com.

This project was financially supported by The Netherlands Organization for Scientific Research (NWO/OCW) via the “BaSyC – Building a Synthetic Cell” Gravitation Grant (024.003.019) and by Agence Nationale de la Recherche (ANR-22-CPJ2-0091-01).

Supplementary Figures

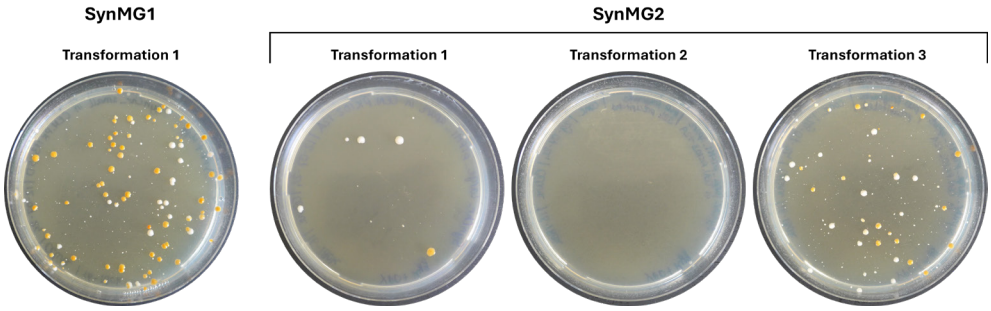


Figure S4.1: Plates after yeast transformation with assembly fragments for SynMG1 and SynMG2.

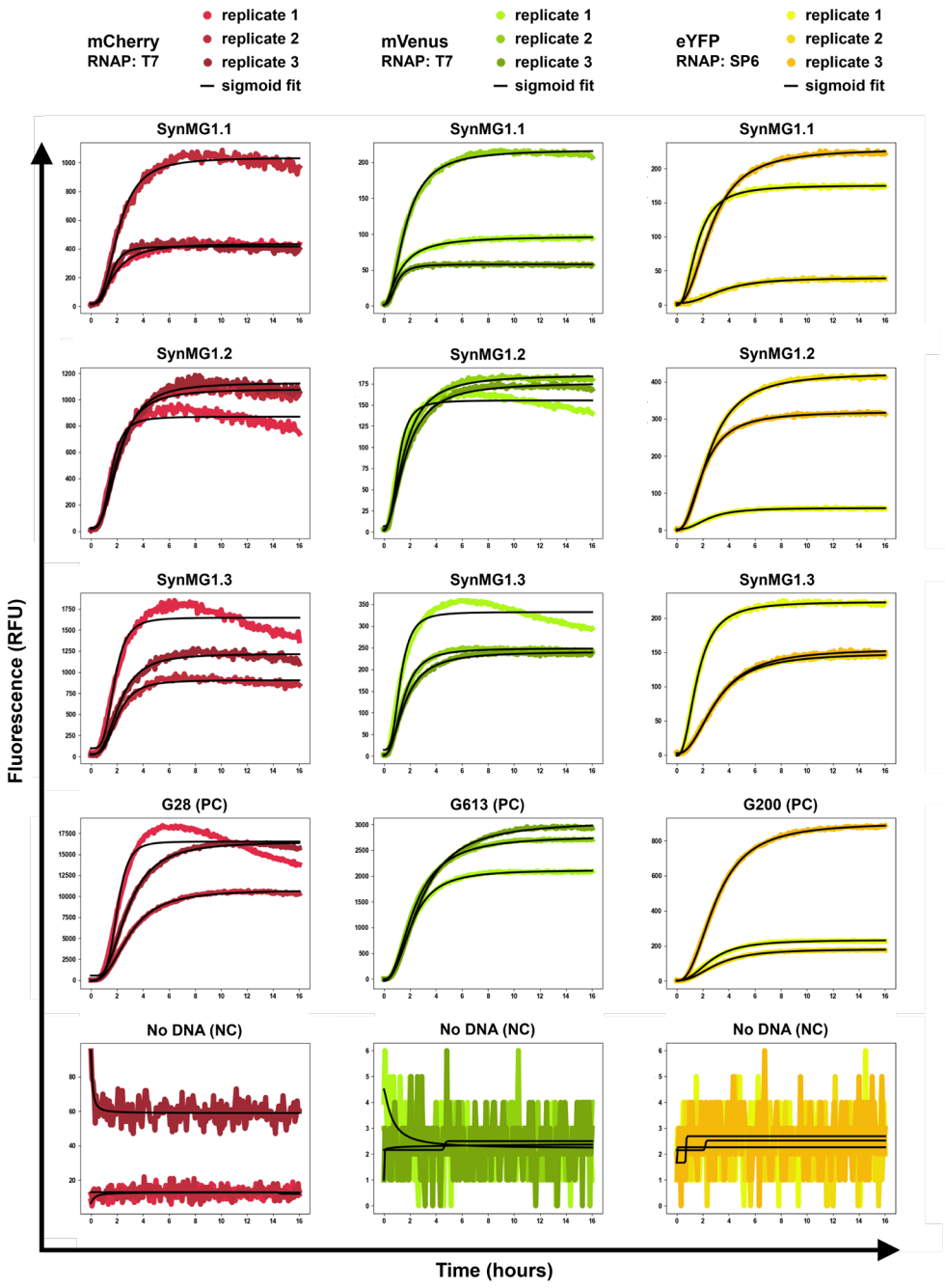


Figure S4.2: Fluorescence measurements over 16 h of expression with PURE system in bulk reactions. The DNA templates (SynMG1 variants or control DNAs), as well as the fluorescent reporter proteins, are indicated. Black lines show sigmoid fitting for calculation of the apparent translation rate (maximum slope). Minimum and maximum of y-axis differ per sample for better visualization of sigmoid fitting.

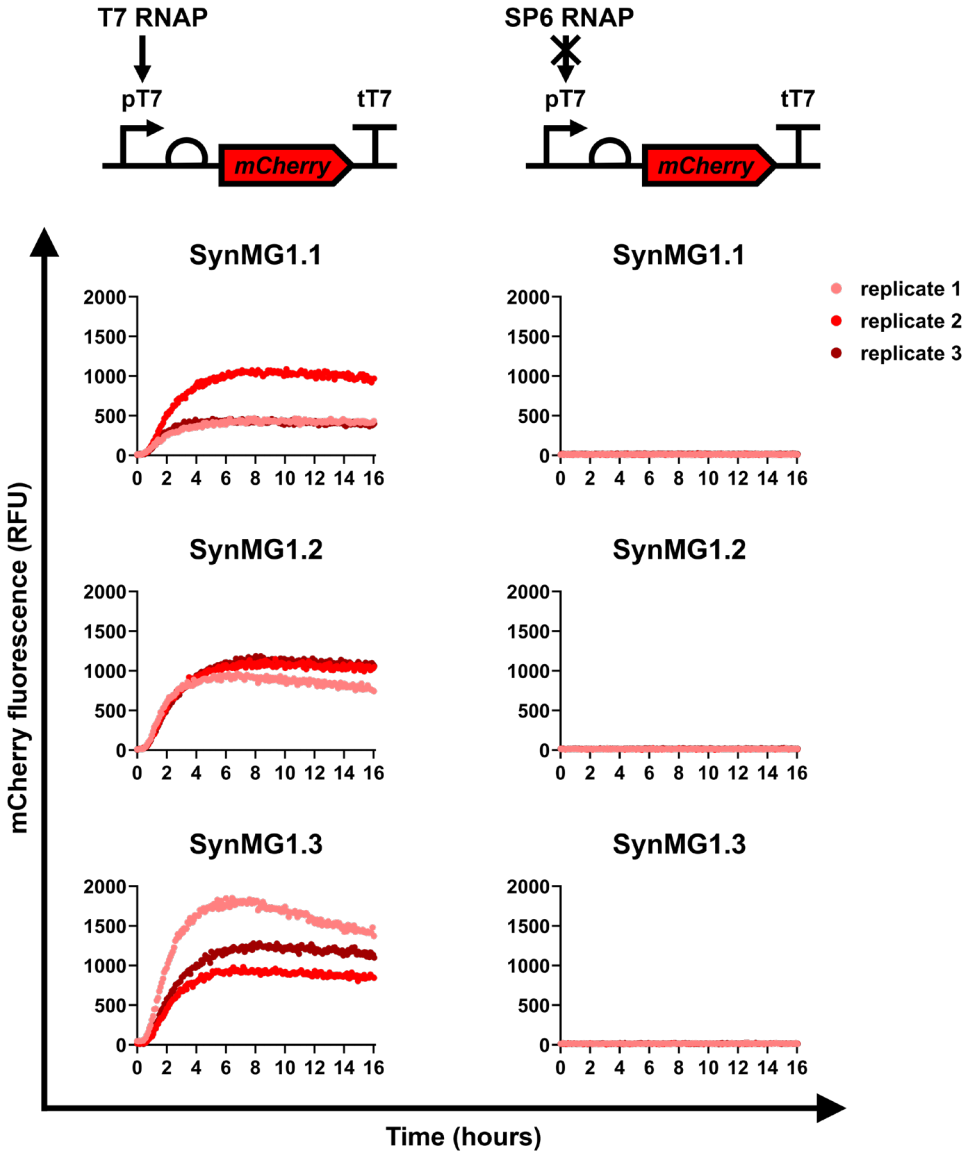


Figure S4.3: Orthogonality of SP6 RNAP with the T7 promoter. For comparison, y-axis minimum and maximum values are the same for all samples.

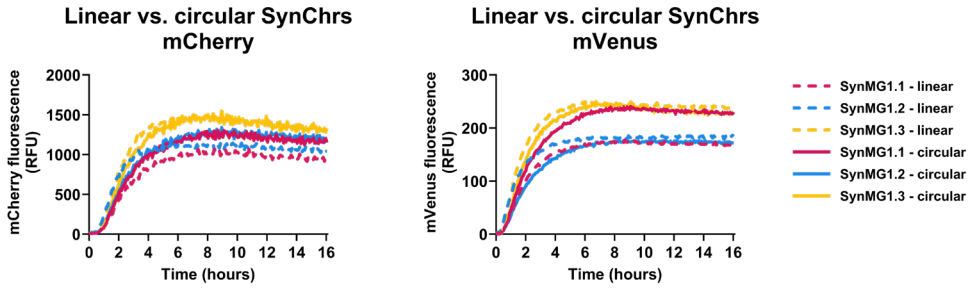


Figure S4.4: Fluorescence measurements for expression of linear versus circular SynChrs with PURE system in bulk reactions. Left: mCherry fluorescence. Right: mVenus fluorescence. All reactions were carried out with T7 RNAP.

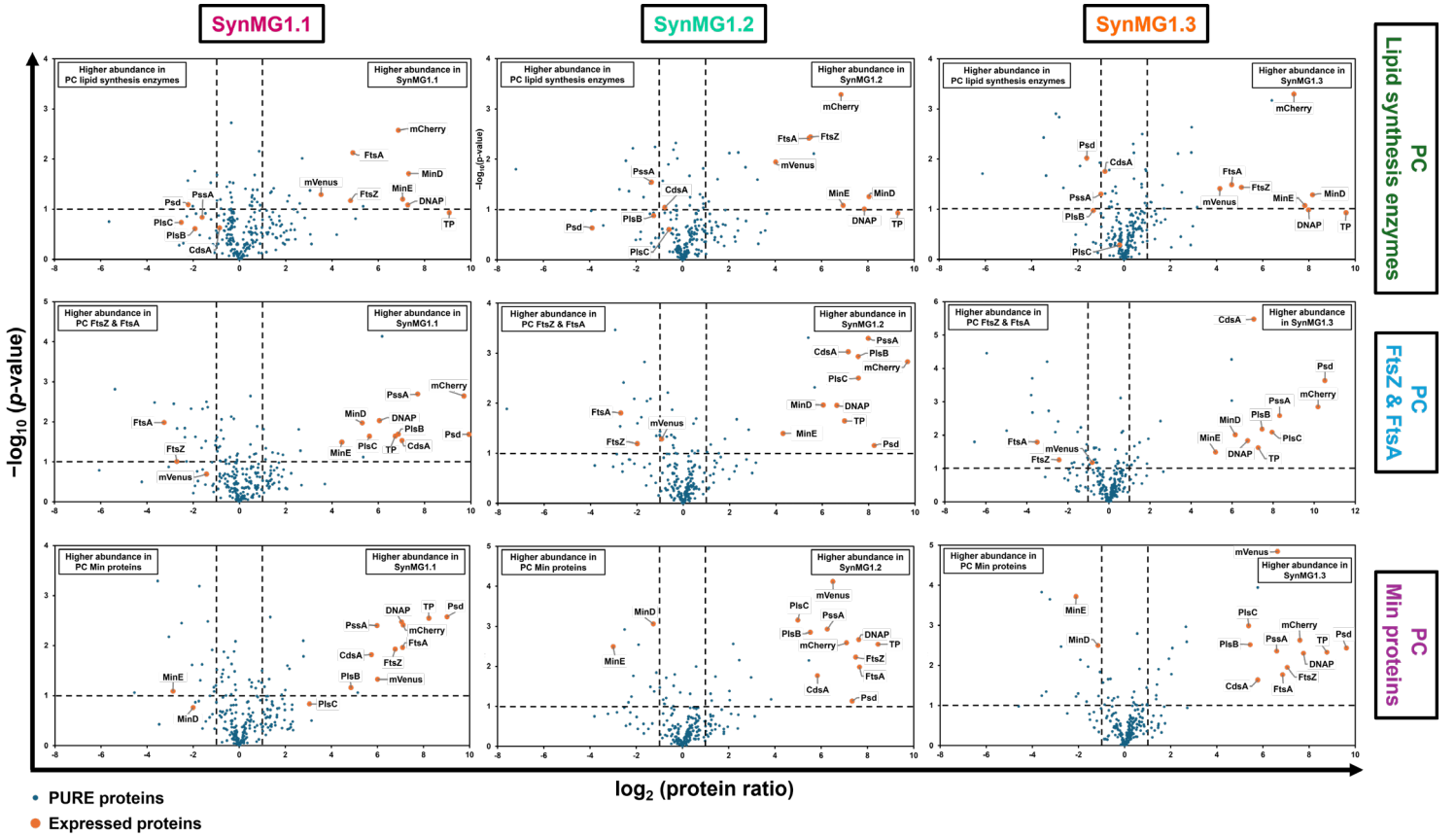


Figure S4.5: Volcano plots displaying the change in relative protein abundance between SynMG1 variants and control plasmids expressed with T7 RNAP. Vertical lines indicate a 2-fold increase. The horizontal line indicates a *p*-value = 0.1 from a two-tailed *t*-test. Values are calculated from independent samples (*n*). *n* = 2 for SynMG1.1 and the control plasmid for lipid synthesis enzymes, and *n* = 3 for all other conditions. Because *n* = 1 for the control plasmid for DNA replication proteins, no volcano plot could be generated for this condition. PC, positive control.

Supplementary Tables

Table S4.1: *S. cerevisiae* strains used in this study

Strain	Relevant genotype	Source
Transformation hosts for assembly in yeast		
CEN.PK113-5D	*MATa ura3-52 HIS3 LEU2*	[32]
CEN.PK102-12A	*MATa ura3-52 his3-1 leu2-3*	[32]
Negative and positive controls for flow cytometry		
CEN.PK113-7D	*MATa URA3 HIS3 LEU2*	[32]
IMF2	*MATa ura3-52 his3-1 leu2-3 NeoChr2(pCCW12-mRuby2- tENO1 pTEF2-mTurquoise2-tSSA1)*	[47]
Strains containing SynChrs assembled in yeast		
IMF75–IMF94	*MATa ura3-52 HIS3 LEU2* SynMG1	This study
IMF95–IMF97	*MATa ura3-52 his3-1 leu2-3* SynMG2	This study

Table S4.2: Configurations of synthetic chromosomes designed for assembly in yeast

Name	Size	Configuration[1,2,3,4,5]
SynMG1	41098 bp	$_{AA}$ CEN6/ARS4 $_{AB}$ *pT7-p2-tVSV-tVSV pT7-p3-tT7* $_{AC}$ *pHHF2-crtE-tADH1* $_{AW}$ *pSP6-eYFP-LL-spinach-tT7* $_{AE}$ *pTDH3-crtI-tTDH* ARS417 $_{AF}$ *pT7-minD-tT7 pT7-minE-tT7* $_{BA}$ *pCCW12-mRuby2-tENO1* $_{BD}$ *pT7-ftsZ:mVenus-ftsA-tT7* $_{AG}$ *pTEF-hphNT1-tTEF* $_{AH}$ oriR-oriL $_{AJ}$ *cat redF* oriV ori2 *repE* incC *sopA sopB* sopC *pURA3-URA3-tURA3* $_{AK}$ *pT7-mCherry-tT7* $_{AL}$ *pPGK1-crtYB-tPGK1* $_{AM}$ *pT7-plsB-tT7 pT7-plsC-tT7 pT7-cdsA-tT7 pT7-pssA-tT7 pT7-psd-tT7 pSP6-pgsA-tT7 pSP6-pgpA-tT7*
SynMG2	104899 bp	$_{AA}$ CEN6/ARS4 $_{AB}$ *pT7-p2-tVSV-tVSV pT7-p3-tT7* $_{AC}$ *pHHF2-crtE-tADH1* $_{AW}$ *pSP6-eYFP-LL-spinach-tT7* $_{AE}$ *pLEU2-LEU2-tLEU2* $_{BG}$ *pT7-prfA-tVSV-tT7 pT7-hisS-tVSV-tT7 pT7-tyrS-tVSV-tT7 pT7-cysS-tVSV-tT7 pT7-trpS-tVSV-tT7 pT7-serS-tVSV-tT7 pT7-valS-tVSV-tT7 pT7-metG-tVSV-tT7 pT7-argS-tVSV-tT7 pT7-glnS-tVSV-tT7 pT7-leuS-tVSV-tT7 pT7-thrS-tT7 pT7-lysS-tVSV-tT7* $_{BE}$ *pTDH3-crtI-tTDH ARS417* $_{AF}$ *pT7-minD-tT7 pT7-minE-tT7* $_{BA}$ *pCCW12-mRuby2-tENO1* $_{BD}$ *pT7-ftsZ:mVenus-ftsA-tT7* $_{AG}$ *pTEF-hphNT1-tTEF* $_{AP}$ *pT7-aspS-tVSV-tT7 pT7-proS-tVSV-tT7 pT7-glyQ-glyS-tVSV-tT7 pT7-gltX-tVSV-tT7 pT7-prfC-tVSV-tT7 pT7-prfB-tVSV-tT7 pT7-frr-tVSV-tT7 pT7-fusA-tVSV-tT7 pT7-infB-tVSV-tT7* $_{AR}$ *pHIS3- HIS3-tHIS3* $_{AS}$ oriR-oriL $_{AX}$ *pTEF2-mTurquoise2-tSSA1* ARS1 $_{AD}$ *pT7-asnS-tVSV-tT7 pT7-ileS -tVSV-tT7 pT7-alaS-tVSV-tT7 pT7-pheS-pheT-tVSV-tT7 pT7-fmt-tVSV-tT7 pT7-tsf-tVSV-tT7 pT7-infC-tT7 pT7-infA-tT7* $_{AT}$ *cat redF* oriV ori2 *repE* incC *sopA sopB* sopC *pURA3-URA3-tURA3* $_{AK}$ *pT7-mCherry-tT7* $_{AL}$ *pPGK1-crtYB-tPGK1* $_{AM}$ *pT7-plsB-tT7 pT7-plsC-tT7 pT7-cdsA-tT7 pT7-pssA-tT7 pT7-psd-tT7 pSP6-pgsA-tT7 pSP6-pgpA-tT7*

[1] SHRs are differently annotated than in [21] and are listed in Table S4.13. SHRs are annotated in subscript between the genetic fragments that they join together.
[2] *pT7* = T7 promoter + RBS, *tT7* = T7 terminator, *tVSV* = VSV terminator, *pSP6* = SP6 promoter, *LL* = long linker.
[3] CEN6/ARS4 fragment also contains an I-SceI cut site and a landing pad.
[4] oriR-oriL fragment contains a PmeI cut site between the ori sequences.
[5] *lacO* sites and T7 tags are present in some expression cassettes, but not mentioned here.

Table S4.3: Stocked yeast and *E. coli* strains containing SynChrs

Yeast strain	SynChr design	All screening markers present?	Sequenced	Correct configuration[1]	Corresponding *E. coli* strain	Difference in SynChr configuration in yeast vs. *E. coli*	Name of SynChr version
IMF75	SynMG1	Yes	Yes	No, duplication of fragments AG_HygR_AH AH_oriRL_AJ AJ_pCC1BAC-URA3_AK AK_mCherry_AL	pUDF005	Three configurations present in *E. coli*: i) recombined to only contain fragments AG_HygR_AH AH_oriRL_AJ AJ_pCC1BAC-URA3_AK AK_mCherry_AL ii) correct configuration iii) same duplication as in yeast	-
IMF76	SynMG1	Yes	No	Unknown	-	-	-
IMF77	SynMG1	Yes	No	Unknown	-	-	-
IMF78	SynMG1	Yes	No	Unknown	-	-	-
IMF79	SynMG1	Yes	No	Unknown	-	-	-
IMF80	SynMG1	Yes	No	Unknown	-	-	-
IMF81	SynMG1	Yes	No	Unknown	-	-	-
IMF82	SynMG1	Yes	Yes	Yes[2]	pUDF006	None	SynMG1.1
IMF83	SynMG1	Yes	No	Unknown	-	-	-
IMF84	SynMG1	Yes	No	Unknown	-	-	-
IMF85	SynMG1	Yes	Yes	No, misses *pgpA* gene	pUDF007	None	SynMG1.2
IMF86	SynMG1	Yes	No	Unknown	-	-	-
IMF87	SynMG1	Yes	Yes	Yes[3]	pUDF008	None	SynMG1.3
IMF88	SynMG1	Yes	No	Unknown	-	-	-
IMF89	SynMG1	Yes	No	Unknown	-	-	-
IMF90	SynMG1	Yes	No	Unknown	-	-	-
IMF91	SynMG1	Yes	No	Unknown	-	-	-
IMF92	SynMG1	Yes	No	Unknown	-	-	-
IMF93	SynMG1	Yes	No	Unknown	-	-	-
IMF94	SynMG1	Yes	No	Unknown	-	-	-

IMF95	SynMG2	No (white colony, mRuby2 mix)	Yes	No, missing (parts of) fragments BG_pLD1_BE BE_crtI-ARS417_AF AF_Min_BA AP_pLD3_AR AD_pLD2_AT	pUDF009	SynChr in *E. coli* additionally misses fragments AA_CENARS_AB AB_DNArep_AC AL_crtYB_AM AM_PLS_AA	SynMG2.1
IMF96	SynMG2	Yes	Yes	No, missing (parts of) fragments BG_pLD1_BE AP_pLD3_AR AD_pLD2_AT	-	-	SynMG2.2
IMF97	SynMG2	Yes	Yes	No, completely different configuration	-	-	-

[1] With "correct configuration", the presence of all complete fragments in the correct order is meant. Point mutations are not considered here. See Table S4.14 for a description of the content of each fragment.
[2] The *mCherry* gene misses from the consensus sequence of the SynChr isolated from yeast, but this is expected to be a sequencing data analysis error. The *mCherry* gene is present in the SynChr isolated from *E. coli*.
[3] The raw reads after isolation from yeast show a mixture of recombined and non-recombined SynChrs. The *E. coli* stock only contains non-recombined SynMG1.3.

To reduce the number of printed pages, Supplementary Tables S4.4–S4.12 and S4.15 are available online from the 4TU.ResearchData repository, as described in the next section Supplementary data. Any references mentioned in the Supplementary Tables are included in the section References. A brief description of the content of these tables can be found in the table titles below.

Table S4.4: Plasmids used to create marker fragments for assembly of SynChrs in yeast

Table S4.5: Plasmids used to create minimal cell fragments for assembly of SynChrs in yeast

Table S4.6: Plasmids used for protein expression in PURE system

Table S4.7: Plasmid used for the DNA replication experiment

Table S4.8: Primers used to construct plasmids

Table S4.9: Diagnostic primers for *E. coli* colony PCR

Table S4.10: Diagnostic primers for Sanger sequencing

Table S4.11: Primers used to generate the fragments for SynChr assembly in yeast

Table S4.12: SHR sequences used in this study

Table S4.15: Linear templates used for DNA replication

Table S4.13: Fragments for assembly of SynChrs in yeast

Fragment	Relevant characteristics	Size (bp)[1]	Template[2]	Primer fw[3]	Primer rv[3]	PCR polymerase[4]	Used for SynChr
Marker fragments							
AA_CENARS_AB	I-SceI cut site Landing pad CEN6/ARS4	832	pUD1394	20524	20446	KOD Xtreme	SynMG1 SynMG2
AC_crtE_AW	*pHHF2-crtE-tADH1*	2173	pUD1249	20447	20448	KOD Xtreme	SynMG1 SynMG2
AE_LEU2_BG	*pLEU2-LEU2-tLEU2*	2050	pYTK075	20449	20890	Phusion	SynMG2
AE_crtI-ARS417_AF	*pTDH3-crtI-tTDH* ARS417	2875	pUD1386	20452	20453	Phusion	SynMG1
BE_crtI-ARS417_AF	*pTDH3-crtI-tTDH* ARS417	2875	pUD1386	20891	20453	KOD Xtreme	SynMG2
BA_mRuby2_BD	*pCCW12-mRuby2-tENO1*	1785	pUDC191	20894	20895	KOD Xtreme	SynMG1 SynMG2
AG_HygR_AH	*pTEF-hphNT1-tTEF*	1741	pYTK079	20456	20457	Phusion	SynMG1
AG_HygR_AP	*pTEF-hphNT1-tTEF*	1741	pYTK079	20456	20458	Phusion	SynMG2
AR_HIS3_AS	*pHIS3-HIS3-tHIS3*	1243	pYTK076	20459	20460	KOD Xtreme	SynMG2
AX_mTurquoise2-ARS1_AD	*pTEF2-mTurquoise2-tSSA1* ARS1	1886	pUDC436	20461	20462	KOD Xtreme	SynMG2
AJ_pCC1BAC-URA3_AK	*cat redF* oriV ori2 *repE* incC *sopA sopB* sopC *pURA3-URA3-tURA3*	8672	pUD1387	20463	20464	KOD Xtreme	SynMG1
AT_pCC1BAC-URA3_AK	*cat redF* oriV ori2 *repE* incC *sopA sopB* sopC *pURA3-URA3-tURA3*	8672	pUD1387	20465	20464	UltraRun LR	SynMG2
AL_crtYB_AM	*pPGK1-crtYB-tPGK1*	3100	pUD1248	20466	20893	KOD Xtreme	SynMG1 SynMG2
Minimal cell fragments							
AB_DNArep_AC	*pT7-p2-tVSV-tVSV* *pT7-p3-tT7*	2972	G435	20468	20469	Phusion	SynMG1 SynMG2
AW_YFP_AE	*pSP6-eYFP-LL-spinach-tT7*	1308	G200	20470	20471	KOD Xtreme	SynMG1 SynMG2

Fragment	Relevant characteristics	Size (bp)[1]	Template[2]	Primer fw[3]	Primer rv[3]	PCR polymerase[4]	Used for SynChr
BG_pLD1_BE	*pT7-prfA-tVSV-tT7* *pT7-hisS-tVSV-tT7* *pT7-tyrS-tVSV-tT7* *pT7-cysS-tVSV-tT7* *pT7-trpS-tVSV-tT7* *pT7-serS-tVSV-tT7* *pT7-valS-tVSV-tT7* *pT7-metG-tVSV-tT7* *pT7-argS-tVSV-tT7* *pT7-glnS-tVSV-tT7* *pT7-leuS-tVSV-tT7* *pT7-thrS-tT7* *pT7-lysS-tVSV-tT7*	25314	pLD1	20896	20897	UltraRun LR	SynMG2
AF_Min_BA	*pT7-minD-tT7* *pT7-minE-tT7*	1741	G396	20904	20899	KOD Xtreme	SynMG1 SynMG2
BD_FtsZA_AG	*pT7-ftsZ:mVenus-ftsA-tT7*	3602	G613	20902	20903	KOD Xtreme	SynMG1 SynMG2
AP_pLD3_AR	*pT7-aspS-tVSV-tT7* *pT7-proS-tVSV-tT7* *pT7-glyQ-glyS-tVSV-tT7 pT7-gltX-tVSV-tT7* *pT7-prfC-tVSV-tT7* *pT7-prfB-tVSV-tT7* *pT7-frr-tVSV-tT7* *pT7-fusA-tVSV-tT7* *pT7-infB-tVSV-tT7*	18548	pLD3	20476	20477	UltraRun LR	SynMG2
AH_oriRL_AJ	oriR PmeI cut site oriL	538	G162	20478	20479	KOD Xtreme	SynMG1
AS_oriRL_AX	oriR PmeI cut site oriL	538	G162	20480	20481	Phusion	SynMG2

Fragment	Relevant characteristics	Size (bp)[1]	Template[2]	Primer fw[3]	Primer rv[3]	PCR polymerase[4]	Used for SynChr
AD_pLD2_AT	*pT7-asnS-tVSV-tT7* *pT7-ileS -tVSV-tT7* *pT7-alaS-tVSV-tT7* *pT7-pheS-pheT-tVSV-tT7 pT7-fmt-tVSV-tT7* *pT7-tsf-tVSV-tT7* *pT7-infC-tT7* *pT7-infA-tT7*	15120	pLD2	20482	20483	UltraRun LR	SynMG2
AK_mCherry_AL	*pT7-mCherry-tT7*	1195	G28	20484	20485	KOD Xtreme	SynMG1 SynMG2
AM_PLS_AA	*pT7-plsB-tT7* *pT7-plsC-tT7* *pT7-cdsA-tT7* *pT7-pssA-tT7* *pT7-psd-tT7* *pSP6-pgsA-tT7* *pSP6-pgpA-tT7*	9415	G276	20900	20901	UltraRun LR	SynMG1 SynMG2

[1] The size of the fragments includes the SHR sequences.
[2] See Table S4.4 & Table S4.5.
[3] See Table S4.11.
[4] KOD Xtreme = KOD Xtreme Hot Start DNA Polymerase, Phusion = Phusion High-Fidelity DNA Polymerase, UltraRun LR = UltraRun LongRange PCR Kit

Table S4.14: Fragments for assembly of SynChrs in yeast

Minimal cell module	System	Protein	Gene(s)	Promoter	Description
DNA replication	Protein-primed DNA replication system from ϕ29	DNAP	*p2*	*pT7*	DNA polymerase
		TP	*p3*	*pT7*	Terminal protein that primes the replication reaction
Phospholipid biosynthesis	Kennedy pathway from *E. coli*	PlsB	*plsB*	*pT7*	Glycerol-3-phosphate acyl transferase that converts glycerol-3-phosphate (G3P) into lysophosphatidic acid (LPA) using fatty acyl coenzyme A (acyl-CoA) as a fatty acid donor
		PlsC	*plsC*	*pT7*	LPA acyl transferase that converts LPA into phosphatidic acid (PA) using acyl-CoA as a fatty acid donor
		CdsA	*cdsA*	*pT7*	Integral membrane protein that catalyzes the formation of CDP-diacylglycerol (CDP-DAG) through activation of PA using CTP
		PssA	*pssA*	*pT7*	CDP-DAG-serine O-phosphatidyltransferase that catalyzes the formation of phosphatidylserine (PS) from CDP-DAG and L-serine
		Psd	*psd*	*pT7*	Phosphatidylserine decarboxylase that catalyzes the decarboxylation of PS into phosphatidylethanolamine (PE)
		PgsA	*pgsA*	*pSP6*	CDP-DAG-G3P 3-phosphatidyltransferase that converts CDP-DAG and G3P into phosphatidylglycerol phosphate (PGP)
		PgpA	*pgpA*	*pSP6*	Phosphatidylglycerophosphatase A that catalyzes the dephosphorylation of PGP into phosphatidylglycerol (PG)
Cell division	Constricting Z-ring from *E. coli*	FtsZ:mVenus	*ftsZ: mVenus*	*pT7*	Fusion protein of FtsZ, a tubulin-related GTPase that can polymerase into protofilaments, and mVenus, a fluorescent reporter protein
		FtsA	*ftsA*	Operon with *ftsZ: mVenus*	Actin homolog that can anchor FtsZ to the membrane
	Min system from *E. coli*	MinD	*minD*	*pT7*	Regulates spatiotemporal formation of Z-ring together with MinE
		MinE	*minE*	*pT7*	Regulates spatiotemporal formation of Z-ring together with MinD
-	Fluorescent reporters	mCherry	*mCherry*	*pT7*	Red fluorescent reporter
		eYFP	*eYFP*	*pSP6*	Yellow fluorescent reporter

Minimal cell module	System	Protein	Gene(s)	Promoter	Description
Translation	PURE system translation factors[1]	RF1	*prfA*	*pT7*	Release factor 1
		HisRS	*hisS*	*pT7*	Histidyl-tRNA synthetase
		TyrRS	*tyrS*	*pT7*	Tyrosyl-tRNA synthetase
		CysRS	*cysS*	*pT7*	Cysteinyl-tRNA synthetase
		TrpRS	*trpS*	*pT7*	Tryptophanyl-tRNA synthetase
		SerRS	*serS*	*pT7*	Seryl-tRNA synthetase
		ValRS	*valS*	*pT7*	Valyl-tRNA synthetase
		MetRS	*metG*	*pT7*	Methionyl-tRNA synthetase
		ArgRS	*argS*	*pT7*	Arginyl-tRNA synthetase
		GlnRS	*glnS*	*pT7*	Glutaminyl-tRNA synthetase
		LeuRS	*leuS*	*pT7*	Leucyl-tRNA synthetase
		ThrRS	*thrS*	*pT7*	Threonyl-tRNA synthetase
		LysRS	*lysS*	*pT7*	Lysyl-tRNA synthetase
		AsnRS	*asnS*	*pT7*	Asparaginyl-tRNA synthetase
		IleRS	*ileS*	*pT7*	Isoleucyl-tRNA synthetase
		AlaRS	*alaS*	*pT7*	Alanyl-tRNA synthetase
		PheRS	*pheS, pheT*	*pT7*	Phenylalanyl-tRNA synthetase
		MTF	*fmt*	*pT7*	Methionyl-tRNA formyltransferase
		EF-Ts	*tsf*	*pT7*	Elongation factor Ts
		IF3	*infC*	*pT7*	Initiation factor 3
		IF1	*infA*	*pT7*	Initiation factor 1
		AspRS	*aspS*	*pT7*	Aspartate-tRNA synthetase
		ProRS	*proS*	*pT7*	Prolyl-tRNA synthetase
		GlyRS	*glyQ, glyS*	*pT7*	Glycyl-tRNA synthetase
		GluRS	*gltX*	*pT7*	Glutamyl-tRNA synthetase
		RF3	*prfC*	*pT7*	Release factor 3
		RF2	*prfB*	*pT7*	Release factor 2
		RRF	*frr*	*pT7*	Ribosome recycling factor
		EF-G	*fusA*	*pT7*	Elongation factor G
		IF2	*infB*	*pT7*	Initiation factor 2

[1] Encoded on SynMG2 only, not on SynMG1.

Supplementary data

Supplementary data is available online from the 4TU.ResearchData repository at https://doi.org/10.4121/ad21c652-ad75-4a99-a09a-46c7d8f383d6 or via the QR code below.

Links to: https://doi.org/10.4121/ad21c652-ad75-4a99-a09a-46c7d8f383d6

The following data is included:

- Designed maps of SynMG1 and SynMG2, in GenBank format
- Sequencing data of SynMG1.1, SynMG1.2, SynMG1.3, SynMG2.1 and SynMG2.2 after total DNA isolation from *S. cerevisiae*
  - » Raw Nanopore sequencing reads provided by Plasmidsaurus, in FASTQ format
  - » Consensus sequences provided by Plasmidsaurus and annotated manually in SnapGene, in GenBank format
- Sequencing data of SynMG1.1, SynMG1.2, SynMG1.3 and SynMG2.1 after SynChr isolation from *E. coli*
  - » Raw Nanopore sequencing reads provided by Plasmidsaurus, in FASTQ format
  - » Consensus sequences provided by Plasmidsaurus and annotated manually in SnapGene, in GenBank format
- Tables with an overview of relevant mutations SynMG1.1, SynMG1.2, SynMG1.3, SynMG2.1 and SynMG2.2 compared to the designed maps, in Excel format
- Tables raw LC-MS data, in Excel format
- Supplementary Tables S4.4–S4.12 and S4.15 in Excel format

The script for analysis of bulk fluorescence measurements is available from https://github.com/DanelonLab.

References

1. Luisi,P.L., Oberholzer,T. and Lazcano,A. (2002) The notion of a DNA minimal cell: A general discourse and some guidelines for an experimental approach. *Helv. Chim. Acta*, **85**, 1759–1777. https://doi.org/10.1002/1522-2675(200206)85:6<1759::AID-HLCA1759>3.0.CO;2-7
2. Forster,A.C. and Church,G.M. (2006) Towards synthesis of a minimal cell. *Mol. Syst. Biol.*, **2**, 45. https://doi.org/10.1038/msb4100090
3. Gil,R. (2014) The Minimal Gene-Set Machinery. In Meyers,R.A. (ed), *Encyclopedia of Molecular Cell Biology and Molecular Medicine: Synthetic Biology*. Wiley-VCH Verlag GmbH & Co. KGaA, pp. 1–36. https://doi.org/10.1002/3527600906.mcb.20130079
4. Hutchison,C.A., Chuang,R.Y., Noskov,V.N., Assad-Garcia,N., Deerinck,T.J., Ellisman,M.H., Gill,J., Kannan,K., Karas,B.J., Ma,L., et al. (2016) Design and synthesis of a minimal bacterial genome. *Science*, **351**, aad6253. https://doi.org/10.1126/science.aad6253
5. Tian,J., Gang,H., Sheng,N., Zhou,X., Gulari,E., Gao,X. and Church,G. (2004) Accurate multiplex gene synthesis from programmable DNA microchips. *Nature*, **432**, 1050–1054. https://doi.org/10.1038/nature03151
6. Shepherd,T.R., Du,L., Liljeruhm,J., Samudyata, Wang,J., Sjödin,M.O.D., Wetterhall,M., Yomo,T. and Forster,A.C. (2017) *De novo* design and synthesis of a 30-cistron translation-factor module. *Nucleic Acids Res.*, **45**, 10895–10905. https://doi.org/10.1093/nar/gkx753
7. 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**, 1–12. https://doi.org/10.1038/s41598-020-80827-8
8. Libicher,K., Hornberger,R., Heymann,M. and Mutschler,H. (2020) In vitro self-replication and multicistronic expression of large synthetic genomes. *Nat. Commun.*, **11**, 904. https://doi.org/10.1038/s41467-020-14694-2
9. Blanken,D., Foschepoth,D., Serrão,A.C. and Danelon,C. (2020) Genetically controlled membrane synthesis in liposomes. *Nat. Commun.*, **11**, 1–13. https://doi.org/10.1038/s41467-020-17863-5
10. Godino,E., Doerr,A. and Danelon,C. (2022) Min waves without MinC can pattern FtsA-anchored FtsZ filaments on model membranes. *Commun. Biol.*, **5**, 1–13. https://doi.org/10.1038/s42003-022-03640-1
11. 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**, 4969. https://doi.org/10.1038/s41467-019-12932-w
12. 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**, 1–11. https://doi.org/10.1038/s42003-020-01258-9
13. 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**, 1–12. https://doi.org/10.1038/s41467-018-03926-1
14. van Nies,P., Canton,A.S., Nourian,Z. and Danelon,C. (2015) Monitoring mRNA and protein levels in bulk and in model vesicle-based artificial cells. In Burke-Aguero,D.H. (ed), *Methods in Enzymology*. Academic Press, Vol. 550, pp. 187–214. https://doi.org/10.1016/bs.mie.2014.10.048
15. Dhar,M.K., Sehgal,S. and Kaul,S. (2012) Structure, replication efficiency and fragility of yeast ARS elements. *Res. Microbiol.*, **163**, 243–253. https://doi.org/10.1016/j.resmic.2012.03.003
16. Bourgeois,L., Pyne,M.E. and Martin,V.J.J. (2018) A Highly Characterized Synthetic Landing Pad System for Precise Multicopy Gene Integration in Yeast. *ACS Synth. Biol.*, **7**, 2675–2685. https://doi.org/10.1021/acssynbio.8b00339
17. Baek,S., Utomo,J.C., Lee,J.Y., Dalal,K., Yoon,Y.J. and Ro,D.K. (2021) The yeast platform engineered for synthetic gRNA-landing pads enables multiple gene integrations by a single gRNA/Cas9 system. *Metab. Eng.*, **64**, 111–121. https://doi.org/10.1016/j.ymben.2021.01.011
18. Shizuya,H., Birren,B., Kim,U.J., Mancino,V., Slepak,T., Tachiiri,Y. and Simon,M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 8794–8797. https://doi.org/10.1073/pnas.89.18.8794
19. Shizuya,H. and Kouros-Mehr,H. (2001) The development and applications of bacterial artificial chromosome cloning system. *Keio J. Med.*, **50**, 26–30. https://doi.org/10.2302/kjm.50.26
20. Wild,J., Hradecna,Z. and Szygbalski,W. (2002) Conditionally amplifiable BACs: Switching from single-copy to high-copy vectors and genomic clones. *Genome Res.*, **12**, 1434–1444. https://doi.org/10.1101/gr.130502

21. Kuijpers,N.G.A., Solis-Escalante,D., Bosman,L., van den Broek,M., Pronk,J.T., Daran,J.M. and Daran-Lapujade,P. (2013) A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. *Microb. Cell Fact.*, **12**, 1–13. https://doi.org/10.1186/1475-2859-12-47
22. Blanken,D., van Nies,P. and Danelon,C. (2019) Quantitative imaging of gene-expressing liposomes reveals rare favorable phenotypes. *Phys. Biol.*, **16**, 045002. https://doi.org/10.1088/1478-3975/ab0c62
23. Abil,Z., Restrepo Sierra,A.M. and Danelon,C. (2023) Clonal Amplification-Enhanced Gene Expression in Synthetic Vesicles. *ACS Synth. Biol.*, **12**, 1187–1203. https://doi.org/10.1021/acssynbio.2c00668
24. Restrepo Sierra,A.M., Gomez,F.R., van Tongeren,M., Heras,L.S. and Danelon,C. (2025) A synthetic cell with integrated DNA self-replication and membrane biosynthesis. *bioRxiv*. https://doi.org/10.1101/2025.01.14.632951
25. van Nies,P. (2017) Virus-inspired DNA replication coupled with gene expression in a minimal cell framework. https://doi.org/10.4233/uuid
26. Sorek,R., Zhu,Y., Creevey,C.J., Pilar Francino,M., Bork,P. and Rubin,E.M. (2007) Genome-Wide Experimental Determination of Barriers to Horizontal Gene Transfer. *Science*, **318**, 1449–1453. https://doi.org/10.1126/science.1147112
27. Lamberte,L.E., Baniulyte,G., Singh,S.S., Stringer,A.M., Bonocora,R.P., Stracy,M., Kapanidis,A.N., Wade,J.T. and Grainger,D.C. (2017) Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nat. Microbiol.*, **2**, 1–9. https://doi.org/10.1038/nmicrobiol.2016.249
28. Chizzolini,F., Forlin,M., Yeh Martín,N., Berloffa,G., Cecchi,D. and Mansy,S.S. (2017) Cell-Free Translation Is More Variable than Transcription. *ACS Synth. Biol.*, **6**, 638–647. https://doi.org/10.1021/acssynbio.6b00250
29. Chizzolini,F., Forlin,M., Cecchi,D. and Mansy,S.S. (2014) Gene position more strongly influences cell-free protein expression from operons than T7 transcriptional promoter strength. *ACS Synth. Biol.*, **3**, 363–371. https://doi.org/10.1021/sb4000977
30. Mans,R., van Rossum,H.M., Wijsman,M., Backx,A., Kuijpers,N.G.A., van den Broek,M., Daran-Lapujade,P., Pronk,J.T., van Maris,A.J.A. and Daran,J.M.G. (2015) CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.*, **15**, 1–15. https://doi.org/10.1093/femsyr/fov004
31. van den Brink,M., Althuis,T.Y., Danelon,C. and Claassens,N.J. (2024) MOSAIC: a highly efficient, one-step recombineering approach to plasmid editing and diversification. *bioRxiv*.
32. Entian,K.D. and Kötter,P. (2007) 25 Yeast Genetic Strain and Plasmid Collections. *Methods Microbiol.*, **36**, 629–666. https://doi.org/10.1016/S0580-9517(06)36025-4
33. Verduyn,C., Postma,E., Scheffers,W.A. and van Dijken,J.P. (1992) Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, **8**, 501–517. https://doi.org/10.1002/yea.320080703
34. Wild,J. and Szybalski,W. (2004) Copy-control pBAC/oriV vectors for genomic cloning. *Methods Mol. Biol.*, **267**, 145–154. https://doi.org/10.1385/1-59259-774-2:145
35. 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. https://doi.org/10.1038/nmeth.1318
36. Inoue,H., Nojima,H. and Okayama,H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*, **96**, 23–28. https://doi.org/10.1016/0378-1119(90)90336-P
37. Salazar,A.N., de Vries,A.R.G., van den Broek,M., Wijsman,M., Cortés,P. de la T., Brickwedde,A., Brouwers,N., Daran,J.M.G. and Abeel,T. (2017) Nanopore sequencing enables near-complete de novo assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. *FEMS Yeast Res.*, **17**, 1–11. https://doi.org/10.1093/femsyr/fox074
38. The Gresham Lab (2022) labtools: Useful Tools for the Gresham Lab.
39. R Core Team (2021) R: A Language and Environment for Statistical Computing.
40. Li,H. (2018) Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, **34**, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191
41. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
42. Thorvaldsdóttir,H., Robinson,J.T. and Mesirov,J.P. (2013) Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.*, **14**, 178–192. https://doi.org/10.1093/bib/bbs017

43. Gietz,R.D. and Woods,R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.*, **350**, 87–96. https://doi.org/10.1016/S0076-6879(02)50957-5
44. Moore,D.A., Whatley,Z.N., Joshi,C.P., Osawa,M. and Erickson,H.P. (2017) Probing for binding regions of the FtsZ protein surface through site-directed insertions: Discovery of fully functional FtsZ-fluorescent proteins. *J. Bacteriol.*, **199**. https://doi.org/10.1128/JB.00553-16
45. Schindelin,J., Arganda-Carreras,I., Frise,E., Kaynig,V., Longair,M., Pietzsch,T., Preibisch,S., Rueden,C., Saalfeld,S., Schmid,B., et al. (2012) Fiji: An open-source platform for biological-image analysis. *Nat. Methods*, **9**, 676–682. https://doi.org/10.1038/nmeth.2019
46. Rueden,C.T., Schindelin,J., Hiner,M.C., DeZonia,B.E., Walter,A.E., Arena,E.T. and Eliceiri,K.W. (2017) ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, **18**, 1–26. https://doi.org/10.1186/s12859-017-1934-z
47. Postma,E.D., Dashko,S., van Breemen,L., Taylor Parkins,S.K., van den Broek,M., Daran,J.-M. and Daran-Lapujade,P. (2021) A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **49**, 1769–1783. https://doi.org/10.1093/nar/gkaa1167
48. Lee,M.E., DeLoache,W.C., Cervantes,B. and Dueber,J.E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.*, **4**, 975–986. https://doi.org/10.1021/sb500366v
49. Sarabipour,S., King,C. and Hristova,K. (2014) Uninduced high-yield bacterial expression of fluorescent proteins. *Anal. Biochem.*, **449**, 155–157. https://doi.org/10.1016/j.ab.2013.12.027
50. 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. U. S. A.*, **108**, 18655–18660. https://doi.org/10.1073/pnas.1114397108
51. Godino,E. (2022) Expression of a gene-encoded FtsZ-based minimal machinery to drive synthetic cell division. https://doi.org/10.4233/uuid

Chapter 5

Discussion and outlook

An essential aspect of the bottom-up approach to building a minimal cell is the design and construction of its genome. Importantly, the genome design and assembly method should be compatible with the expression system envisaged for protein synthesis, in this case PURE system. In this dissertation, we explored homologous recombination in yeast as a method for the modular assembly of synthetic chromosomes (SynChrs) from multiple expression cassettes. Starting with test chromosomes designed to assess the efficiency of DNA assembly and extraction from yeast (Chapter 3), we then proceeded to design and build SynChrs containing gene sets relevant for a minimal cell and we demonstrated their successful expression in PURE system (Chapter 4).

In this chapter, I give recommendations for improving SynChr construction in yeast, specifically for enhancing the assembly efficiency of fragments containing repeated sequences, optimizing the pipeline to design, assemble and screen SynChrs and improving the transfer from yeast to PURE system. Additionally, I briefly discuss an alternative to assembly in yeast. Finally, I discuss strategies for characterizing SynChrs *in vitro*, designing larger synthetic genomes and integrating the functionalities of a minimal cell that are encoded on its genome.

Construction of a synthetic chromosome for the minimal cell

Assembly in yeast was explored as a method to construct a chromosome for minimal cells. Two major challenges were identified: (i) assembly of DNA fragments with repeated regulatory sequences for expression in PURE system ("PURE repeats") and (ii) transfer of the assembled chromosome from yeast to PURE system.

Assembly of DNA fragments with repeats in yeast

Assembly of DNA fragments in *Saccharomyces cerevisiae* relies on homologous recombination of double-stranded DNA (dsDNA) fragments flanked by homologous sequences. While not explicitly mentioned in studies describing DNA assembly in yeast, single-strand annealing (SSA) is the most probable underlying pathway. Before envisioning possible strategies to improve assembly efficiency of dsDNA fragments, I here describe the basics of SSA in the context of yeast physiology. SSA is a mechanism for repair of double-strand breaks (DSBs) in the yeast genome, which is employed when repeats are located close to the DSB and no sister chromatid is available [1, 2]. The DSB ends are processed by end resection, resulting in single-stranded DNA (ssDNA) ends revealing the repeats, which can be annealed together (Figure 5.1A). Any non-homologous 3' ssDNA ends are cleaved off, existing gaps are filled by a DNA polymerase and a DNA ligase completes the repair [2]. During this process, one of the repeats and the sequence between the repeats (if present) are lost. Contrarily to the classical DSB repair pathway, strand invasion catalyzed by RAD51 is not necessary for SSA. Disruption of RAD51 leads to an increase in the frequency of SSA, while reducing the occurrence of homology-directed repair involving strand invasion [1–3]. During DNA assembly in yeast, the designed overlapping sequences at the ends of the assembly fragments—SHRs [4] in this dissertation—are the repeats involved in SSA. However, other repeated sequences (PURE repeats) are present on the assembly fragments of the SynChrs as

well, both near the fragment ends, as well as internally. Recombination of these extra repeats results in misassembled SynChrs. To understand whether repeats located farther from the fragment ends could be involved in SSA as well, we can take a closer look at the mechanism of end resection (Figure 5.1A). End resection is proposed to start by an incision made by the Mre11 protein <100 bp away from the DSB [5]. Mre11 has 3'-5' exonuclease activity, which degrades the DNA from the incision back toward the DSB break. The exonucleases Exo1 and Sgs1-Dna2 are recruited by Mre11. With their 5'-3' exonuclease activity, the 3' ssDNA overhang is enlarged in the direction away from the DSB. This can result in an ssDNA overhang of multiple kilobases in size [6]. Any repeats located within this ssDNA overhang can be involved in recombination, resulting in misassembly of SynChrs at repeated sequences (Figure 5.1B). With the underlying mechanism in mind, we can discuss strategies to optimize assembly efficiency of SynChrs with repeats in yeast.

In Chapter 3, we evaluated the assembly efficiency of ten fragments flanked by PURE repeats ($SynChr^{PURE}$) and compared it to an equivalent assembly without repeats ($SynChr^{control}$). For $SynChr^{PURE}$, 9% of colonies contained all screening and selection markers, and sequencing confirmed a correct configuration in 8% of colonies. In contrast, $SynChr^{control}$ showed a much higher assembly efficiency, with 86% of colonies containing all markers. In Chapter 4, we attempted to assemble seven (SynMG1) or ten (SynMG2) fragments, containing 15 (SynMG1) or 47 (SynMG2) genes for expression in PURE system. While the use of a plethora of markers—one between each fragment with repeats—allowed the identification of correct assemblies for $SynChr^{PURE}$ and SynMG1, no correct assemblies could be obtained for SynMG2. As expected, many incorrect assemblies were obtained due to recombination at the PURE repeats. A minimal genome for a bottom-up synthetic cell is expected to contain over 150 genes (Chapter 1), requiring the assembly of a much greater number of fragments, many of which will contain repeated sequences. This will further reduce the fraction of correct assemblies. In practice, correct SynChrs can be identified when approximately 0.1–1% of colonies contain the designed construct. Therefore, improving the assembly efficiency of fragments with repeated sequences in yeast is essential.

Firstly, the total number of assembled fragments is important for assembly efficiency [7] and can be reduced in two ways: (i) by reducing the number of marker fragments and (ii) by combining expression cassettes on one template prior to assembly in yeast. The former strategy necessitates a different screening pipeline, for example based on (q)PCR [8] to verify the presence of all fragments. Manual screening in this way is labor-intensive and, therefore, automated equipment is essential, as further elaborated on below. The latter strategy is the one employed in Chapter 4. While *in vitro* cloning of template plasmids that combine multiple expression cassettes reduced the number of fragments for assembly, this strategy also posed several disadvantages. Firstly, in the assembly of SynMG1 and SynMG2, the longest fragments (encoding the phospholipid synthesis pathway and translation factors) were most involved in undesired recombination. This could be caused by the high number of PURE repeats on these fragments, but also by the fragment size itself: increased fragment size comes with more difficulties in obtaining a clean sample of the fragment. The error rate of the polymerase employed to amplify fragments with sizes of 10–30 kb (UltraRun LongRange PCR Kit, Qiagen, reported fidelity "10× higher than Taq") resulted in frequent point mutations.

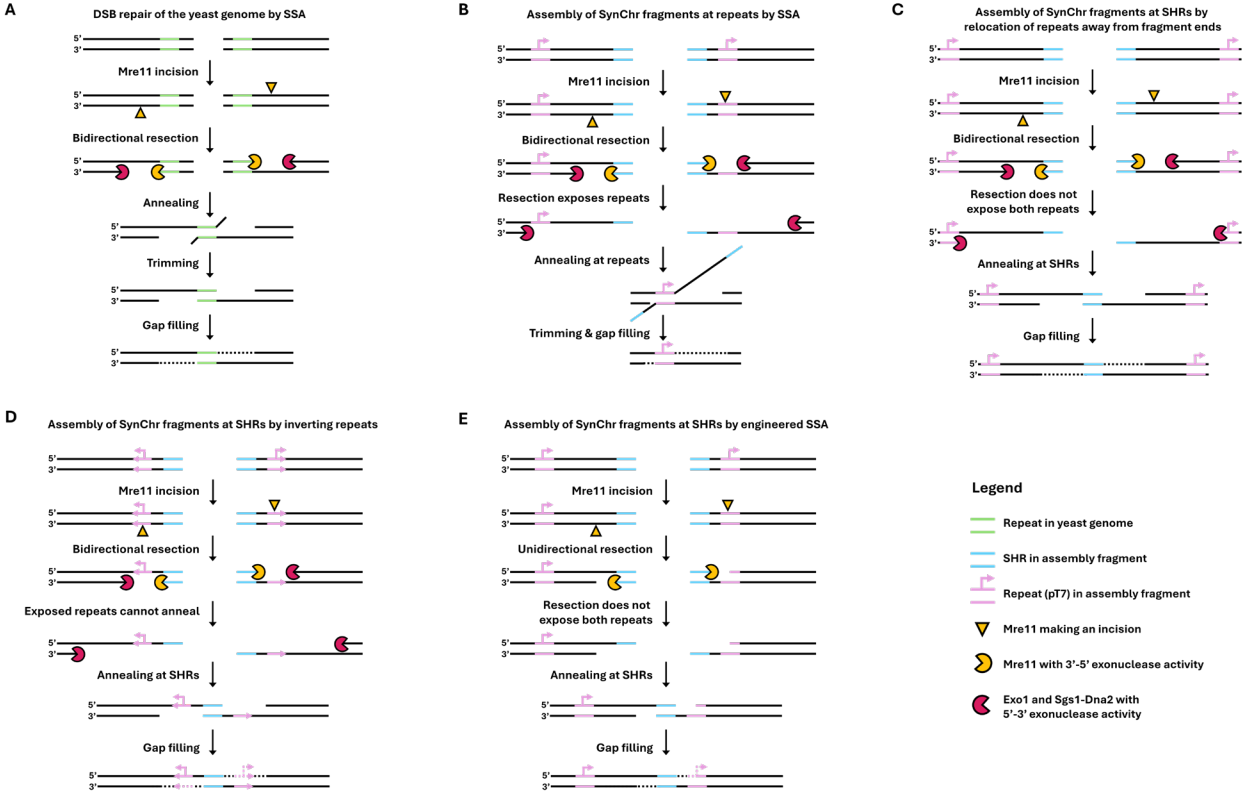


Figure 5.1: Schematic showing the mechanism of single-strand annealing (SSA) and strategies to prevent recombination at repeats during SynChr assembly. A) SSA mechanism to repair DSBs in the yeast genome, when repeats are located closed to the DSB. **B)** SSA mechanism for SynChr assembly that may cause assembly at repeated sequences such as pT7 instead of at SHRs, through exposure of the repeats during end resection. **C, D, E)** Strategies to direct assembly towards SHRs instead of repeats. **C)** Relocation of repeats farther away from fragment ends may prevent exposure of the repeats in the ssDNA overhang after end resection. **D)** Inversion of repeats may prevent annealing after exposure by end resection. **E)** Engineered SSA without Exo1 and Sgs1-Dna2 may prevent exposure of both repeats in the ssDNA overhangs.

Furthermore, extraction of long PCR products from gel is hindered by poor DNA migration and by the cut-off size of around 10 kb for most DNA clean-up columns. As an alternative strategy, restriction digestion of template plasmids combining multiple expression cassettes could be used to produce the fragment of interest, thereby preventing the occurrence of PCR-born point mutations. If the plasmid backbone does not have to be removed from the sample prior to assembly in yeast, this strategy would also overcome problems caused by extraction and clean-up from gel.

Secondly, to reduce the abundance and similarity of repeated sequences and consequently lower the incidence of undesired recombination events, the regulatory sequences of each expression cassette can be diversified [9]. The design of expression cassettes in this dissertation contained numerous nearly identical copies of the promoter, RBS and terminator regions. While using the same regulatory sequences simplifies the construction of expression cassettes, as the same plasmid backbone can be used to clone regulatory sequences upstream and downstream of each gene, repeated sequences complicate assembly in yeast and maintenance in *E. coli* (Chapter 4). Furthermore, different genes might need different expression levels in minimal cells, as well as temporal tuning during cell growth and division, which will require diversification of the regulatory sequences. Many options for diversification are available: libraries of T7 promoters [10], alternative promoter-polymerase pairs from bacteriophages (T7 variants [11], SP6 [12, 13] or T3 [14]) or from *E. coli* [14–16], RBS libraries [17] and alternative terminators [18, 19].

Thirdly, the location of repeated sequences on the assembly fragments might influence the rate of undesired recombination in yeast. In our initial design, SynChr^{PURE}, PURE repeats were only located on the fragment ends, directly next to the SHRs. In the designs of SynMG1 and SynMG2, repeats were present both at fragment ends, as well as internally. While we did not systematically investigate the effect of the location of the repeats on recombination, we did identify SynMG1 and SynMG2 variants which are the products of recombination events that occurred at internal repeats. With SSA as the main mechanism for assembly of DNA fragments, it is expected that repeats at fragment ends are more involved in undesired recombination than internal repeats: when end resection continues past the SHRs, PURE repeats are revealed and available for annealing (Figure 5.1B). In that case, locating PURE repeats away from fragment ends might reduce unwanted recombination (Figure 5.1C). This can be achieved by designing assembly fragments which start and end in the middle of a gene, instead of outside an expression cassette. A downside to this approach is the reduction in modularity compared to the SHR-based approach: fragments cannot easily be interchanged [4]. Alternatively, the direction of expression cassettes on the fragment flanks might matter. Flipping part of the outermost expression cassettes might reduce recombination, due to inability to anneal inverted PURE repeats revealed by end resections (Figure 5.1D). However, internal repeats also contributed to undesired recombination during assembly of SynMG1 and SynMG2. This might have been caused by the presence of damaged DNA fragments, or SSA might not be the only mechanism involved in recombination during SynChr assembly. In the latter case, repeats revealed by end resection might invade other assembly fragments at internal repeats and cause recombination with the canonical pathway for DSB repair. A possible strategy to eliminate recombination of internal repeats would be to engineer the *S. cerevisiae* host to knock out *rad51* [1–3].

Combined with relocation of PURE repeats away from fragment ends, this strategy might increase assembly efficiency of SynChrs. However, knockout of *rad51* leads to accumulation of unrepaired DSBs in the yeast genome and therefore has a high impact on cellular physiology [20], with unknown consequences for the suitability of this mutant as SynChr foundry. Additionally, it should be noted that, in case *E. coli* is required for amplification of the SynChrs, instability of the SynChr in *E. coli* might occur regardless of the location of the repeats on the assembly fragments (see paragraph "Transfer of assembled SynChrs from yeast to PURE system").

Fourthly, the homologous recombination machinery of *S. cerevisiae* could be engineered to alter the processivity of the nucleases involved in end resection, thereby reducing the likelihood that PURE repeats are involved in recombination. The exonucleases Exo1 and Sgs1-Dna2 are involved in producing long ssDNA overhangs during SSA [6]. When these exonucleases are knocked out (individually or double knockouts of *exo1* and *sgs1* or *dna2* and *sgs1*), shorter ssDNA tails are created [6, 21]. Shorter ssDNA overhangs that do not reveal the PURE repeat located next to the SHR might reduce the involvement of these PURE repeats in undesired recombination (Figure 5.1E). However, these exonucleases are also involved in the canonical DSB repair pathway, and consequently, mutants might exhibit similar physiological changes as for the *rad51* knockout described above [20].

Lastly, the choice of selection pressure upon transformation with the assembly mix might matter for assembly efficiency. In this dissertation, selection on the transformation plates was done through the presence of the antibiotic hygromycin. It should be investigated whether the use of an auxotrophic marker instead of a dominant antibiotic marker for initial selection could promote correct assembly. After transformation, selection by antibiotics might occur before the yeast cell has had enough time to assemble all fragments and express sufficient resistance marker, resulting in cell death. Auxotrophic selection does not kill cells immediately, but rather prevents growth of cells which have not assembled the auxotrophic marker. Therefore, more time is available to assemble all fragments, possibly resulting in a higher assembly efficiency.

Improvement of the pipeline to design, assemble and screen SynChrs

Design, fragment generation, assembly and screening of SynChrs were performed manually in this work. Automation of one or multiple steps in the pipeline would speed up the process, reduce the amount of consumables per assembly, and allow for the construction of multiple designs or even libraries of SynChrs simultaneously.

Primer design was done by manual identification of unique primer binding sites with desired annealing temperature, addition of an SHR tag and checking for primer dimers using online tools. The final design was checked by simulating a Gibson assembly in SnapGene. This manual process can take multiple hours, but it can be easily automated. Existing tools like NEBuilder (https://nebuilder.neb.com/) could be utilized, where SHRs need to be added manually as custom spacers. In collaboration with Laura Sierra Heras, a script to design primers with SHRs for assembly in yeast was written. While this script has now been shown to work well for other assemblies (unpublished), it

was not suitable yet for the SynChr designs during this doctoral work. Once this tool is finalized and available online, it will be of great benefit for DNA assembly in yeast.

Generation of assembly fragments by PCR is the most time-consuming step in the assembly pipeline. PCR conditions had to be optimized and once the correct settings were found, multiple reactions had to be run to obtain enough product, and bands had to be excised from agarose gels by hand and cleaned up. The PCR step could be simplified by the use of a more performant DNA polymerase that is fast, robust and efficient at a single annealing temperature regardless of the melting temperature of the primers. A promising candidate that fulfills these requirements is the repliQa HiFi ToughMix from Quantabio. Additionally, automated PCR systems could be used for assembling and running the PCR reactions. If amplification is highly specific, clean-up from gel is not necessary and PCR products can directly be purified. This step could be automated by the use of liquid handling robots as well.

Alternatively, gene synthesis services could be employed to obtain linear assembly fragments. However, the costs are higher than PCR on an already in-house plasmid or genome, especially when assembly fragments are large.

In this dissertation, DNA quantification of template plasmids and assembly fragments was done manually on a Qubit fluorometer. Instead, DNA concentration measurements can be performed at high-throughput with a fluorescence plate reader, after mixing the DNA fragments with a fluorescent dye that selectively binds dsDNA (e.g., Quant-iT assay kits, Invitrogen), using a liquid handling robot.

After obtaining all assembly fragments, yeast transformation can be automated with the use of liquid handling robots [22]. Preparation of the DNA mix from tens of fragments could be done with a low-volume dispenser. After transformation, a colony picking robot could be employed for streaking to obtain single colonies. No extensive suite of screening and selection markers is necessary, because the presence of all fragments can be detected at high-throughput by qPCR [8]. This approach requires the inclusion of qPCRTags in the SynChr design, or analysis of naturally occurring qPCR primer binding sites.

Finally, strains testing positive for all fragment junctions should be sequenced. Due to the high prevalence of repeats, long-read sequencing is preferred over short-read sequencing to simplify data analysis. In our experience, outsourcing this to a commercial provider is cheaper and faster than sequencing in house. Sequencing can be done efficiently in house if high-throughput sequencing equipment, data analysis and storage facilities, and dedicated technical support are available. Notably, until SynChr isolation from yeast is optimized, whole-genome sequencing of yeast strains carrying the assemblies remains necessary. In our hands, sufficient read depth could only be achieved by limiting multiplexing to a maximum of four samples per Nanopore MinION flow cell. Improving SynChr extraction from yeast would significantly reduce sequencing costs by allowing for multiplexing of more samples, and reducing data volume and analysis time.

High-throughput automation equipment comes with a high price tag, and is not available in most academic labs. Therefore, the establishment of biofoundries—specialized facilities with such equipment and dedicated support staff—is important for the

accessibility of automation technology and, ultimately, the advancement of synthetic genomics [23].

Transfer of assembled SynChrs from yeast to PURE system

After identification of correctly assembled SynChrs by sequencing, the SynChrs need to be isolated from yeast for expression in PURE system. It is important that the SynChrs stay intact during isolation, and are obtained in high purity and concentration. Ideally, SynChrs can directly be transferred from yeast to PURE system, without intermediate amplification in *E. coli*. This would simplify the pipeline and circumvent any problems with toxicity or stability in *E. coli*.

The recommended DNA concentration to use in the commercial PURE*frex* kit (GeneFrontier) is 2 nM, but a concentration of 50 pM is enough to encapsulate approximately one DNA molecule per liposome (assuming a diameter of 4 μm) [24], and is sufficient for co-expression of the DNA replication and phospholipid synthesis modules [25]. In a standard PURE*frex* reaction, 3.5 μL out of 10 μL are available for DNA template, requiring the initial DNA concentration to be at least ca.150 pM to have one chromosome per liposome. An estimation can be made for the maximum SynChr concentration that can be obtained after isolation from yeast. The average copy number of a chromosome with centromeric origin is one [7]. The starting cell count for a midiprep isolation is approximately 7.0×10^9 (according to the Qiagen Genomic DNA handbook), from which theoretically 7.0×10^9 SynChr molecules could be isolated, corresponding to 1.16×10^{-14} moles. Elution volume is typically 100 μL, resulting in a final SynChr concentration of ca. 120 pM. This means that, in the optimal case that all SynChr DNA can be isolated, the minimum concentration required for encapsulation in liposomes is not reached. Therefore, next to efforts to improve DNA yields by scaling up the number of cells or improving existing extraction protocols, as attempted in Chapter 3 (resulting in a SynChr concentration of 60–80 pM), it is recommended to investigate how the copy number of the SynChr in yeast can be increased. An alternative replication origin with high copy number is that of the native 2μ plasmid. However, preliminary data in Chapter 3 suggest that it is not suitable for SynChrs, due to an increased number of misassemblies. Ongoing efforts in the Pascale Daran-Lapujade lab are therefore focusing on strategies to increase copy number of SynChrs after assembly.

Purity of SynChrs can be affected by carryover of chemicals from the extraction method, or the presence of yeast proteins, RNA or DNA. Most extraction protocols include steps to remove proteins and RNA by enzymatic degradation, using proteinase and RNase, respectively. Yeast DNA, in *S. cerevisiae* consisting of 12-Mb genomic DNA [26], present as 16 linear chromosomes, multicopy 86-kb mitochondrial DNA [27], present in circular and linear form [28], and a multicopy circular 6-kb 2μ plasmid [29], can be removed in multiple ways. In protocols for isolation of plasmids or SynChrs based on alkaline extraction [30, 31], genomic DNA is removed by denaturation under alkaline conditions and subsequent precipitation upon neutralization. Smaller circular DNA, including SynChrs and the native 2μ plasmid, is renatured upon neutralization and subsequently purified using a column. However, the alkaline lysis-based protocol used for SynChr isolation in Chapter 3 resulted in a DNA sample with a low SynChr content of approximately 10%, indicating substantial contamination with host DNA.

Additionally or alternatively to alkaline lysis for genomic DNA removal, exonuclease can be utilized to degrade linear genomic and mitochondrial DNA, and circular DNA can be separated from linear DNA by topological trapping of the circular DNA in agarose plugs [32, 33]. However, we did not succeed in isolating SynChr DNA using the corresponding protocol, and the authors described even further reduction of SynChr yield when the topological trapping procedure is included. Two alternative approaches to selectively isolate SynChrs in high purity are currently being explored in the Pascale Daran-Lapujade lab: (i) *in vivo* reduction of non-SynChr DNA prior to SynChr isolation by utilizing an *S. cerevisiae* strain lacking mitochondria and native 2µ plasmids, combined with induction of host chromosome degradation before DNA extraction, and (ii) selective SynChr isolation *in vitro* by "fishing" the SynChrs out of cell lysate or isolated total DNA, using a sequence-specific probe [34].

Lastly, the integrity of the isolated SynChrs is important for their functionality as minimal genomes in liposomes. Therefore, gentle enzymatic removal of the yeast cell wall is preferred over cell wall disruption by physical methods. Furthermore, once the cells are lysed, large DNA molecules are prone to shearing by vortexing or pipetting. To avoid shearing, wide-bore pipette tips can be used, or DNA isolation can be performed inside agarose plugs [33].

In case isolation from yeast cannot be optimized sufficiently to obtain SynChrs in high purity and concentration, utilization of *E. coli* as intermediate amplification host should still be considered. Transfer of the SynChr from yeast to *E. coli* addresses issues of purity and concentration. *E. coli* selectively amplifies the SynChr, not the cotransformed yeast DNA, due to absence of replication origins and antibiotic resistance markers on the native yeast DNA. Commercial kits for isolation of BACs from *E. coli* with high yields are readily available (e.g., the NucleoBond Xtra Midi kit from Macherey-Nagel). While the maximum insert size for BACs is uncertain, the successful transformation and maintenance of 300–500 kb BACs suggest that *E. coli* could be used effectively to amplify SynChrs comparable in size to that of the estimated minimal genome [35, 36]. Stability and unpredictable toxicity, however, are limiting factors in the use of *E. coli* as SynChr amplification host. In Chapter 4, we found that multiple SynChr configurations were present after isolation from *E. coli*, due to recombination of the PURE repeats. Still, clonal populations of the full-sized SynChr could be obtained after culturing single colonies that were identified as positive clones by colony PCR. A possibility to decrease the chance of recombination, is to spread out the components of the pCC1BAC backbone over the SynChr design. Toxicity of heterologous sequences in *E. coli* is a known problem and is, unfortunately, unpredictable [37–39]. Both stability and toxicity of SynChrs in *E. coli* can be reduced, but not eliminated, by the use of single-copy number BAC backbones.

An alternative to assembly in yeast

Given the challenges that must be addressed before assembly in yeast can become a versatile tool for constructing minimal genomes, alternative assembly methods should be considered. One approach is optimized Golden Gate assembly [40, 41]. Pryor and colleagues showed successful one-pot assembly of 52 fragments into a 40-kb phage genome. The ability to assemble a high number of fragments simultaneously makes this

a promising method for minimal genome construction. However, the authors note that assembly efficiency declines rapidly as fragment number increases, recommending a maximum fragment number of 40 for optimal results [40]. The maximum size of constructs that can be assembled with this strategy was not specified. Unlike homology-based methods, Golden Gate relies on restriction-ligation, which may reduce issues related to repeated sequences during assembly. However, since *E. coli* is used for amplification of the *in vitro* assembled DNA, repeated sequences might still hinder stable maintenance of the construct and unpredictable toxicity remains a concern [40]. Another advantage of this method is that assembly fragments can be obtained from sequence-verified template plasmids by restriction, eliminating the need for PCR. This minimizes the risk of point mutations introduced by PCR polymerases, while removing the necessity of optimizing PCR conditions. On the other hand, the preparation of assembly fragments requires site-directed mutagenesis to remove internal restriction sites, or fragments need to be ordered as synthetic DNA from a vendor.

The optimal strategy for constructing minimal genomes has yet to be determined. Further research is necessary to improve *S. cerevisiae* as a genome foundry and to deepen our understanding of the mechanisms underlying toxicity and stability in *E. coli* [38, 39]. Alternatively, other host systems could be explored, e.g. *Bacillus subtilis* ([42] and Chapter 2).

Characterization of a synthetic chromosome for the minimal cell

After successful assembly and isolation of SynChrs, their expression was tested in PURE system. In Chapter 4, we demonstrated synthesis of all SynChr-encoded proteins in bulk PURE reactions, using fluorescence measurements of marker proteins and mass spectrometry. Bulk reactions serve as an initial test to assess DNA concentration and purity for expression in PURE system, but should be followed by characterization in liposomes to evaluate module functionality in a minimal cell-like environment. Stochastic encapsulation of reaction components and a diversity of liposome sizes typically lead to large phenotypic heterogeneity. Interestingly, some liposomes may produce more proteins or have a longer expression lifespan than bulk reactions [43]. Therefore, we encapsulated SynMG1 in liposomes and confirmed the expression of fluorescent markers. However, SynChr expression in liposomes was performed only once and showed low liposome quality and weak fluorescence signals, necessitating replication of this experiment with freshly isolated SynChr DNA. Further analysis using mass spectrometry on fluorescent liposome populations sorted by FACS will provide more insights into expression levels under minimal cell-like conditions, providing new insights for re-designing the SynChr with optimized concentrations of the synthesized proteins. Additionally, single-liposome proteomics may become feasible in the near future [44]. Next to determination of relative protein synthesis levels, absolute quantitative proteomics analysis should be performed on liposome samples to determine whether protein levels reach the required thresholds for module functionality in the minimal cell. Finally, high-throughput mass spectrometry of PURE samples is under investigation in the Christophe Danelon lab for assaying the expression of SynChrs under many different conditions.

The functionality of all individual modules encoded on SynMG1 of Chapter 4 has previously been characterized in PURE system in liposomes [45–48]. Similar approaches will be employed to characterize their activity when co-expressed from the SynChr. Attention will be paid to avoid spectral overlap of the fluorescent reporters for the different modules.

Design of larger synthetic chromosomes for a minimal cell

A 41-kb synthetic chromosome encoding genes for DNA replication, phospholipid synthesis and cell division was constructed in Chapter 4 (SynMG1). This SynChr can be seen as the starting point for the design and construction of larger and more complete SynChrs for the minimal cell. Firstly, SynMG1 can be engineered to include more genes, for instance by CRISPR-mediated integration into its landing pad [7]. Potential first genes for integration are the PURE translation factor genes included on the plasmid pTFM1 [49], resulting in an equivalent to SynMG2. Additional genes missing from the current modules could be included as well: *minC* of the Min system [46], and *p5* and *p6* of the DNA replication system [48]. Secondly, the knowledge gained in this study can be applied to the design and assembly in yeast of new SynChrs encoding additional or alternative modules.

A key consideration in the design of new SynChrs is expression regulation, which can be included in the form of diversified regulatory elements (promoter libraries, RBS libraries, additional polymerases, operons) or genetic circuits to tune expression levels and timing. Therefore, ongoing work in the Danelon lab focuses on diversification of RBS sequences and development of genetic circuits.

Another consideration is the configuration of the minimal genome, which is dictated by the DNA replication machinery of choice. Preliminary results in Chapter 4 demonstrated that the protein-primed ϕ29 DNA replication machinery can replicate the linearized 41-kb SynChr. Combined with the successful reconstitution of this system in liposomes [48] and its ability to self-replicate a 9.6-kb template encoding two minimal cell modules [25], these results show that the ϕ29 machinery is a promising DNA replication system for the minimal cell. Future research should assess its ability to replicate even longer templates, which are simultaneously being transcribed by RNAP.

Genome configuration could also have an impact on expression levels in PURE system and encapsulation efficiency in liposomes. Expression from both circular and linearized SynMG1 was tested in bulk PURE reactions and the preliminary results are shown in Chapter 4. The tested SynChr variants did not show significant differences in fluorescence levels between the circular and linear configuration. The difference in encapsulation efficiency in liposomes was not tested and should be assessed in future research.

In this dissertation, SynChrs were assembled in circular configuration in yeast and were subsequently cut with the PmeI restriction enzyme *in vitro* to obtain a linear DNA flanked by origins of replication for replication by the reconstituted ϕ29 DNA replication machinery. This approach necessitates the removal of any PmeI recognition sites in

the assembly fragments. SynChrs could also be assembled in a linear configuration in yeast [50], but this requires the presence of a synthetic telomere at both DNA ends, which should be removed—e.g., by restriction digestion—to reveal the ϕ29 origins of replication prior to DNA replication.

Lastly, genome organization—whether as a monopartite or multipartite genome—is another key design consideration for the minimal cell. Again, this decision depends on the chosen DNA replication mechanism, considering which of the reaction steps, processivity or replication initiation, limits efficiency. The ϕ29 DNAP exhibits high processivity [51, 52], but its initiation efficiency is suboptimal [53]. If it can replicate DNA templates of the estimated minimal genome size, a monopartite genome would be preferable, as a single chromosome simplifies DNA segregation during cell division. However, if the maximum template size for ϕ29 DNA replication is smaller than the expected minimal genome size and replication initiation is not a bottleneck, a multipartite genome should be considered. Alternatively, other DNA replication systems could be explored, as suggested in Chapter 1 (Table 1.2). However, these systems each present their own challenges. For instance, the replication-cycle reaction based on the *E. coli* replication machinery involves many more proteins (14, encoded by 25 genes [54]) than the ϕ29 DNA replication system, increasing the complexity of the minimal cell. Additionally, a DNA replication mechanism based on ϕ29 rolling-circle amplification combined with Cre-loxP recombination would alter the genome configuration from circular to mostly linear upon replication [55], which may limit its suitability to sustain genome replication across multiple cell divisions. Hence, in light of the present results and discussion, we believe that a linear, single SynChr replicated by the protein-primed ϕ29 apparatus is currently the most promising design strategy to engineer a replicating synthetic genome.

Evolutionary integration of functionalities in a minimal cell

Rational design of DNA templates and reaction conditions to optimize individual modules has not always resulted in the desired functionality: the phospholipid synthesis module produces insufficient lipids for detectable liposome growth [45] and the division machinery has been capable of constriction, but not division [56]. Furthermore, although rational design has led to concomitant activity for DNA replication and phospholipid synthesis [25], integration of more modules through a rational approach might be difficult due to the lack of understanding of the biological processes to be reconstituted and the vast combination of experimental conditions that could be attempted. Challenges that may impact co-functionality include: (i) insufficient protein synthesis due to competition for shared transcription and translation resources, (ii) inhibitory effects of substrates, cofactors or (intermediate) reaction (by)products from one module on another, (iii) inefficient encapsulation of DNA, PURE system and all necessary substrates and cofactors, (iv) differences in optimal reaction temperatures across modules, (v) requirements for timing gene expression to enable the coordination between different modules, and (vi) simultaneous occupancy of DNA by multiple DNA-binding proteins: RNAP, DNAP and DSB.

Evolution can be used as an engineering tool to optimize individual proteins (e.g., [57]), single modules [58], or integration of multiple modules [59]. *In vitro* evolution of minimal cell-like systems requires rounds of (i) genetic diversification and (ii) variant selection. In both cases, it is of importance that liposomes contain only one DNA copy to ensure a genotype-phenotype link [58]. Continuous evolution would involve rounds of DNA replication and mutagenesis inside liposomes, liposome division, and selection of those liposomes with desired phenotypes. Discontinuous evolution would encompass the disruption of selected liposomes for DNA release, amplification and diversification of the extracted DNA, and subsequent re-encapsulation in liposomes [58, 59].

Genetic diversification can be done on multiple levels: by point mutations targeting regulatory or coding sequences, or on a larger scale by varying the order, direction and presence of expression cassettes. In discontinuous evolution mode, point mutations can possibly be introduced in SynMG1 through MOSAIC [60]: a protocol for generation of combinatorial plasmid libraries which has previously been used to target RBS sequences. Two challenges that may arise when applying MOSAIC to SynMG1 are: (i) recommended co-transformation of *E. coli* with oligos and SynChrs might be complicated by recombination of SynChrs upon transformation, and (ii) the low editing efficiency for plasmids with the RK2 (oriV) origin, which is the "high-copy" origin of the pCC1BAC backbone of the SynChr that is activated by arabinose induction. Alternatively, in continuous evolution, point mutations could be introduced by an error-prone variant of the ϕ29 DNAP [59, 61]. This would result in random mutagenesis of both regulatory and coding sequences.

Larger-scale genetic diversification could be introduced in future SynChrs through combinatorial assembly in yeast or SCRaMbLing of the SynChrs after assembly. The former approach requires the generation of a fragment library for each "position" on the SynChr by flanking a diverse set of fragments with the same SHR pair. The latter approach requires the incorporation of loxPsym sequences to allow SCRaMbLing in yeast [62] or *in vitro* [63]. The effect of repeated sequences on the feasibility of these methods should be assessed, as well as the expected library size and its compatibility with the available screening and selection methods of liposomes.

Variant selection of liposomes can be based on protein synthesis levels or on module functionality. Protein synthesis-based selection may be done by quantification of the protein content with mass spectrometry either from bulk reaction samples in 384-well plates or similar to the single-cell proteomics approach described in [64]. Variants with desired relative or absolution protein concentrations could be selected for subsequent rounds of genetic diversification and selection. The library size that can be screened with this method is rather low, due to the maximum throughput and costs of mass spectrometry. Alternatively, selection based on module activity could be done by phenotypic sorting of liposomes. This requires the fluorescence labelling of the involved proteins or reaction products, which may be limited by the availability of lasers and proteins or dyes with distinct fluorescence spectra. Three promising approaches can be taken. Firstly, fluorescence-activated cell sorting (FACS) can be used for sorting of liposomes with the desired fluorescence signals. It is a high-throughput method that has been successfully applied to liposome screening and sorting [24, 43]. However, detection of spatial or temporal localization—essential for assessing the functionality of dynamic processes (e.g., the Min system) or localized components

(e.g., the constriction ring)—is not possible. Secondly, an image-based method for liposome selection is currently being developed in the Christophe Danelon lab. This approach involves fluorescence microscopy imaging, followed by tagging of liposomes exhibiting the desired phenotypes via fluorescent protein activation and subsequent sorting of fluorescent liposomes using FACS. With this method, it is possible to assess liposome morphology, spatial localization of proteins and dynamic behaviour, albeit with lower throughput than sorting based solely on FACS. Thirdly, a combination of the above-mentioned methods is a promising approach to variant selection: intelligent image-activated cell sorting (iIACS) [65, 66]. The iIACS system integrates a high-throughput fluorescence microscope with a cell sorter that sorts cells based on image analysis by a deep neural network. This approach combines the high throughput of FACS with the spatial detection capabilities of microscopy. However, the low refractive index of liposomes containing PURE system complicates accurate detection based on forward scatter signals [67], currently limiting the feasibility of this approach for selection of liposomes with desired phenotypes.

Application of evolutionary approaches to liposomes containing SynChrs will enable the optimization and stepwise integration of new functions that can be linked to a fluorescent reporter, or that enhance DNA replication efficiency or expression in PURE system.

Final conclusion

With the methodologies developed in this dissertation for design, assembly, screening, isolation, and characterization of synthetic chromosomes for the minimal cell, as well as the constructed SynChr prototypes, we have paved the way for future engineering of minimal genomes and integration of functional modules in synthetic cells. I foresee that the combination of the expanding toolsets for synthetic genomics with synthetic cell-related technologies (e.g., liposomes, PURE system, active learning-assisted directed evolution, automated experimentation) will soon bring the first synthetic cell to reality.

References

1. Mathiasen,D.P. and Lisby,M. (2014) Cell cycle regulation of homologous recombination in *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.*, **38**, 172–184. https://doi.org/10.1111/1574-6976.12066
2. Bhargava,R., Onyango,D.O. and Stark,J.M. (2016) Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet.*, **32**, 566–575. https://doi.org/10.1016/j.tig.2016.06.007
3. Jasin,M. and Rothstein,R. (2013) Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.*, **5**, 1–18. https://doi.org/10.1101/cshperspect.a012740
4. Kuijpers,N.G.A., Solis-Escalante,D., Bosman,L., van den Broek,M., Pronk,J.T., Daran,J.M. and Daran-Lapujade,P. (2013) A versatile, efficient strategy for assembly of multi-fragment expression vectors in *Saccharomyces cerevisiae* using 60 bp synthetic recombination sequences. *Microb. Cell Fact.*, **12**, 1–13. https://doi.org/10.1186/1475-2859-12-47
5. Bazzano,D., Lomonaco,S. and Wilson,T.E. (2021) Mapping yeast mitotic 5' resection at base resolution reveals the sequence and positional dependence of nucleases *in vivo*. *Nucleic Acids Res.*, **49**, 12607–12621. https://doi.org/10.1093/nar/gkab597
6. Zhu,Z., Chung,W.H., Shim,E.Y., Lee,S.E. and Ira,G. (2008) Sgs1 Helicase and Two Nucleases Dna2 and Exo1 Resect DNA Double-Strand Break Ends. *Cell*, **134**, 981–994. https://doi.org/10.1016/j.cell.2008.08.037
7. Postma,E.D., Dashko,S., van Breemen,L., Taylor Parkins,S.K., van den Broek,M., Daran,J.-M. and Daran-Lapujade,P. (2021) A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **49**, 1769–1783. https://doi.org/10.1093/nar/gkaa1167
8. Mitchell,L.A., Phillips,N.A., Lafont,A., Martin,J.A., Cutting,R. and Boeke,J.D. (2015) qPCRTag analysis - A high throughput, real time PCR assay for Sc2.0 genotyping. *J. Vis. Exp.*, **2015**, 1–7. https://doi.org/10.3791/52941
9. Datta,A., Hendrix,M., Lipsitch,M. and Jinks-Robertson,S. (1997) Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc. Natl. Acad. Sci. U. S. A.*, **94**, 9757–9762. https://doi.org/10.1073/pnas.94.18.9757
10. Komura,R., Aoki,W., Motone,K., Satomura,A. and Ueda,M. (2018) High-throughput evaluation of T7 promoter variants using biased randomization and DNA barcoding. *PLoS One*, **13**, 1–16. https://doi.org/10.1371/journal.pone.0196905
11. Meyer,A.J., Ellefson,J.W. and Ellington,A.D. (2015) Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for *in Vivo* or *in Vitro* Synthetic Circuitry. *ACS Synth. Biol.*, **4**, 1070–1076. https://doi.org/10.1021/sb500299c
12. 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. https://doi.org/10.1016/j.febslet.2004.09.046
13. Weise,L.I., Heymann,M., Mayr,V. and Mutschler,H. (2019) Cell-free expression of RNA encoded genes using MS2 replicase. *Nucleic Acids Res.*, **47**, 10956–10967. https://doi.org/10.1093/nar/gkz817
14. Niederholtmeyer,H., Stepanoèa,È. and Maerkl,S.J. (2013) Implementation of cell-free biological networks at steady state. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 15985–15990. https://doi.org/10.1073/pnas.1311166110
15. Asahara,H. and Chong,S. (2010) *In vitro* genetic reconstruction of bacterial transcription initiation by coupled synthesis and detection of RNA polymerase holoenzyme. *Nucleic Acids Res.*, **38**, 1–10. https://doi.org/10.1093/nar/gkq377
16. 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. https://doi.org/10.1021/acssynbio.6b00017
17. Salis,H.M., Mirsky,E.A. and Voigt,C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.*, **27**, 946–950. https://doi.org/10.1038/nbt.1568
18. Du,L., Villarreal,S. and Forster,A.C. (2012) Multigene expression in vivo: Supremacy of large versus small terminators for T7 RNA polymerase. *Biotechnol. Bioeng.*, **109**, 1043–1050. https://doi.org/10.1002/bit.24379
19. Calvopina-Chavez,D.G., Gardner,M.A. and Griffitts,J.S. (2022) Engineering efficient termination of bacteriophage T7 RNA polymerase transcription. *G3 Genes, Genomes, Genet.*, **12**, 3–8. https://doi.org/10.1093/g3journal/jkac070
20. Holland,C.L., Weis,M.F., England,C.J., Berry,A.M., Hall,P.D. and Lewis,L.K. (2023) Deficiency in

homologous recombination is associated with changes in cell cycling and morphology in *Saccharomyces cerevisiae*. *Exp. Cell Res.*, **430**, 113701. https://doi.org/10.1016/j.yexcr.2023.113701
21. Mimitou,E.P. and Symington,L.S. (2008) Sae2, Exo1 and Sgs1 collaborate in DNA double-strand break processing. *Nature*, **455**, 770–774. https://doi.org/10.1038/nature07312
22. Liu,G., Lanham,C., Buchan,J.R. and Kaplan,M.E. (2017) High-throughput transformation of *Saccharomyces cerevisiae* using liquid handling robots. *PLoS One*, **12**, 1–15. https://doi.org/10.1371/journal.pone.0174128
23. Stephenson,A., Lastra,L., Nguyen,B., Chen,Y.J., Nivala,J., Ceze,L. and Strauss,K. (2023) Physical Laboratory Automation in Synthetic Biology. *ACS Synth. Biol.*, **12**, 3156–3169. https://doi.org/10.1021/acssynbio.3c00345
24. Abil,Z., Restrepo Sierra,A.M. and Danelon,C. (2023) Clonal Amplification-Enhanced Gene Expression in Synthetic Vesicles. *ACS Synth. Biol.*, **12**, 1187–1203. https://doi.org/10.1021/acssynbio.2c00668
25. Restrepo Sierra,A.M., Gomez,F.R., van Tongeren,M., Heras,L.S. and Danelon,C. (2025) A synthetic cell with integrated DNA self-replication and membrane biosynthesis. *bioRxiv*. https://doi.org/10.1101/2025.01.14.632951
26. Goffeau,A., Barrell,B.G., Bussey,H., Davis,R.W., Dujon,B., Feldmann,H., Galibert,F., Hoheisel,J.D., Jacq,C., Johnston,M., et al. (1996) Life with 6000 Genes. *Science*, **274**, 546–567.
27. Foury,F., Roganti,T., Lecrenier,N. and Purnelle,B. (1998) The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett.*, **440**, 325–331. https://doi.org/10.1016/S0014-5793(98)01467-7
28. Chen,X.J. and Clark-Walker,G.D. (2018) Unveiling the mystery of mitochondrial DNA replication in yeasts. *Mitochondrion*, **38**, 17–22. https://doi.org/10.1016/j.mito.2017.07.009
29. Rizvi,S.M.A., Prajapati,H.K. and Ghosh,S.K. (2018) The 2 micron plasmid: a selfish genetic element with an optimized survival strategy within *Saccharomyces cerevisiae*. *Curr. Genet.*, **64**, 25–42. https://doi.org/10.1007/s00294-017-0719-2
30. Devenish,R.J. and Newlon,C.S. (1982) Isolation and characterization of yeast ring chromosome III by a method applicable to other circular DNAs. *Gene*, **18**, 277–288. https://doi.org/10.1016/0378-1119(82)90166-4
31. Singh,M. V. and Weil,P.A. (2002) A method for plasmid purification directly from yeast. *Anal. Biochem.*, **307**, 13–17. https://doi.org/10.1016/S0003-2697(02)00018-0
32. Noskov,V.N., Chuang,R.Y., Gibson,D.G., Leem,S.H., Larionov,V. and Kouprina,N. (2011) Isolation of circular yeast artificial chromosomes for synthetic biology and functional genomics studies. *Nat. Protoc.*, **6**, 89–96. https://doi.org/10.1038/nprot.2010.174
33. Lartigue,C., Glass,J.I., Alperovich,N., Pieper,R., Parmar,P.P., Hutchison,C.A., Smith,H.O. and Venter,J.C. (2007) Genome transplantation in bacteria: Changing one species to another. *Science*, **317**, 632–638. https://doi.org/10.1126/science.1144622
34. Holub,M. (2024) Single-Chromosome Biophysics. https://doi.org/10.4233/uuid:08912541-02f4-44af-b8a6-c3b7a3647f17
35. Karas,B.J., Molparia,B., Jablanovic,J., Hermann,W.J., Lin,Y.C., Dupont,C.L., Tagwerker,C., Yonemoto,I.T., Noskov,V.N., Chuang,R.Y., et al. (2013) Assembly of eukaryotic algal chromosomes in yeast. *J. Biol. Eng.*, **7**, 1–12. https://doi.org/10.1186/1754-1611-7-30
36. Shizuya,H., Birren,B., Kim,U.J., Mancino,V., Slepak,T., Tachiiri,Y. and Simon,M. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. U. S. A.*, **89**, 8794–8797. https://doi.org/10.1073/pnas.89.18.8794
37. Warren,R.L., Freeman,J.D., Levesque,R.C., Smailus,D.E., Flibotte,S. and Holt,R.A. (2008) Transcription of foreign DNA in *Escherichia coli*. *Genome Res.*, **18**, 1798–1805. https://doi.org/10.1101/gr.080358.108
38. Sorek,R., Zhu,Y., Creevey,C.J., Pilar Francino,M., Bork,P. and Rubin,E.M. (2007) Genome-Wide Experimental Determination of Barriers to Horizontal Gene Transfer. *Science*, **318**, 1449–1453. https://doi.org/10.1126/science.1147112
39. Lamberte,L.E., Baniulyte,G., Singh,S.S., Stringer,A.M., Bonocora,R.P., Stracy,M., Kapanidis,A.N., Wade,J.T. and Grainger,D.C. (2017) Horizontally acquired AT-rich genes in *Escherichia coli* cause toxicity by sequestering RNA polymerase. *Nat. Microbiol.*, **2**, 1–9. https://doi.org/10.1038/nmicrobiol.2016.249
40. Sikkema,A.P., Tabatabaei,S.K., Lee,Y.-J., Lund,S. and Lohman,G.J.S. (2023) High-Complexity One-Pot Golden Gate Assembly. *Curr. Protoc.*, **3**, e882. https://doi.org/10.1002/cpz1.882
41. Pryor,J.M., Potapov,V., Bilotti,K., Pokhrel,N. and Lohman,G.J.S. (2022) Rapid 40 kb Genome Construction from 52 Parts through Data-optimized Assembly Design. *ACS Synth. Biol.*, **11**, 2036–2042. https://doi.org/10.1021/acssynbio.1c00525
42. Karas,B.J., Suzuki,Y. and Weyman,P.D. (2015) Strategies for cloning and manipulating natural and

synthetic chromosomes. *Chromosom. Res.*, **23**, 57–68. https://doi.org/10.1007/s10577-014-9455-3
43. Blanken,D., van Nies,P. and Danelon,C. (2019) Quantitative imaging of gene-expressing liposomes reveals rare favorable phenotypes. *Phys. Biol.*, **16**, 045002. https://doi.org/10.1088/1478-3975/ab0c62
44. Mansuri,M.S., Williams,K. and Nairn,A.C. (2023) Uncovering biology by single-cell proteomics. *Commun. Biol.*, **6**, 1–6. https://doi.org/10.1038/s42003-023-04635-2
45. Blanken,D., Foschepoth,D., Serrão,A.C. and Danelon,C. (2020) Genetically controlled membrane synthesis in liposomes. *Nat. Commun.*, **11**, 1–13. https://doi.org/10.1038/s41467-020-17863-5
46. 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**, 4969. https://doi.org/10.1038/s41467-019-12932-w
47. 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**, 1–11. https://doi.org/10.1038/s42003-020-01258-9
48. 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**, 1–12. https://doi.org/10.1038/s41467-018-03926-1
49. Shepherd,T.R., Du,L., Liljeruhm,J., Samudyata, Wang,J., Sjödin,M.O.D., Wetterhall,M., Yomo,T. and Forster,A.C. (2017) *De novo* design and synthesis of a 30-cistron translation-factor module. *Nucleic Acids Res.*, **45**, 10895–10905. https://doi.org/10.1093/nar/gkx753
50. Annaluru,N., Muller,H., Mitchell,L.A., Ramalingam,S., Stracquadanio,G., Richardson,S.M., Dymond,J.S., Kuang,Z., Scheifele,L.Z., Cooper,E.M., et al. (2014) Total Synthesis of a Functional Designer Eukaryotic Chromosome. *Science*, **344**, 55–59.
51. Blanco,L., Bernad,A., Lázaro,J.M., Martín,G., Garmendia,C. and Salas,M. (1989) Highly Efficient DNA Synthesis by the Phage ϕ 29 DNA Polymerase. *J. Biol. Chem.*, **264**, 8935–8940. https://doi.org/10.1016/s0021-9258(18)81883-x
52. 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**, 1035–1036. https://doi.org/10.1016/j.molcel.2004.12.006
53. Esteban,J.A., Salas,M. and Blanco,L. (1993) Fidelity of ϕ29 DNA polymerase. Comparison between protein-primed initiation and DNA polymerization. *J. Biol. Chem.*, **268**, 2719–2726. https://doi.org/10.1016/s0021-9258(18)53833-3
54. Su'etsugu,M., Takada,H., Katayama,T. and Tsujimoto,H. (2017) Exponential propagation of large circular DNA by reconstitution of a chromosome-replication cycle. *Nucleic Acids Res.*, **45**, 11525–11534. https://doi.org/10.1093/nar/gkx822
55. 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. https://doi.org/10.1021/acssynbio.1c00430
56. Godino,E. and Danelon,C. (2023) Gene-Directed FtsZ Ring Assembly Generates Constricted Liposomes with Stable Membrane. *Adv. Biol.*, **7**, 2200172. https://doi.org/10.1002/adbi.202200172
57. Yu,T., Boob,A.G., Singh,N., Su,Y. and Zhao,H. (2023) *In vitro* continuous protein evolution empowered by machine learning and automation. *Cell Syst.*, **14**, 633–644. https://doi.org/10.1016/j.cels.2023.04.006
58. Abil,Z., María,A., Sierra,R., Stan,A.R., Châne,A. and Prado,A. (2024) Darwinian Evolution of Self-Replicating DNA in a Synthetic Protocell. *Nat. Commun.*, **15**, 9091. https://doi.org/10.1038/s41467-024-53226-0
59. Abil,Z. and Danelon,C. (2020) Roadmap to Building a Cell: An Evolutionary Approach. *Front. Bioeng. Biotechnol.*, **8**, 1–8. https://doi.org/10.3389/fbioe.2020.00927
60. van den Brink,M., Althuis,T.Y., Danelon,C. and Claassens,N.J. (2024) MOSAIC: a highly efficient, one-step recombineering approach to plasmid editing and diversification. *bioRxiv*.
61. De Vega,M., Lázaro,J.M., Salas,M. and Blanco,L. (1998) Mutational analysis of ϕ29 DNA polymerase residues acting as ssDNA ligands for 3'-5' exonucleolysis. *J. Mol. Biol.*, **279**, 807–822. https://doi.org/10.1006/jmbi.1998.1805
62. Dymond,J.S., Richardson,S.M., Coombes,C.E., Babatz,T., Muller,H., Annaluru,N., Blake,W.J., Schwerzmann,J.W., Dai,J., Lindstrom,D.L., et al. (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature*, **477**, 471–476. https://doi.org/10.1038/nature10403
63. Wu,Y., Zhu,R.Y., Mitchell,L.A., Ma,L., Liu,R., Zhao,M., Jia,B., Xu,H., Li,Y.X., Yang,Z.M., et al. (2018) In vitro

DNA SCRaMbLE. *Nat. Commun.*, **9**, 1935. https://doi.org/10.1038/s41467-018-03743-6
64. Ye,Z., Sabatier,P., van der Hoeven,L., Lechner,M.Y., Phlairaharn,T., Guzman,U.H., Liu,Z., Huang,H., Huang,M., Li,X., et al. (2025) Enhanced sensitivity and scalability with a Chip-Tip workflow enables deep single-cell proteomics. *Nat. Methods*. https://doi.org/10.1038/s41592-024-02558-2
65. Isozaki,A., Mikami,H., Tezuka,H., Matsumura,H., Huang,K., Akamine,M., Hiramatsu,K., Iino,T., Ito,T., Karakawa,H., et al. (2020) Intelligent image-activated cell sorting 2.0. *Lab Chip*, **20**, 2263–2273. https://doi.org/10.1039/d0lc00080a
66. Nitta,N., Sugimura,T., Isozaki,A., Mikami,H., Hiraki,K., Sakuma,S., Iino,T., Arai,F., Endo,T., Fujiwaki,Y., et al. (2018) Intelligent Image-Activated Cell Sorting. *Cell*, **175**, 266-276.e13. https://doi.org/10.1016/j.cell.2018.08.028
67. Restrepo Sierra,A.M. (2024) Engineering Synthetic Cells through Module Integration and Evolution. https://doi.org/10.4233/uuid:2f2dfc76-5e29-4a04-84c1-3d337e3bf645

Acknowledgements

Welcome to the most well-read chapter of my dissertation! In this chapter, I would like to thank everyone who, in one way or another, contributed to my enjoyable PhD time, both inside and outside the lab.

Christophe, my fascination with the research you do began already in my first year of studying Nanobiology, when you gave an inspiring guest seminar. This led to my application for a BEP position in your lab three years later, when I wrote you an email stating that the project *"assembly of synthetic minimal genomes seemed the most interesting to me"*. Little did I know that this would become my PhD topic instead! During my time as a BEP student working with Jonás on the beautiful Min project, I got to know you as a truly kind person, who is highly involved in everyone's projects, and deeply passionate and knowledgeable about minimal cells. After my enjoyable BEP experience and after taking your course on the origins of life (my favorite course of the Master's programme!), I did not hesitate to contact you for a PhD position. I was super happy when you offered me a position together with Pascale! I admire your genuine care for the well-being of your lab members, your availability both in person and online, your attention for experimental detail, your excitement whenever I shared promising results, and your optimistic mindset in case of not-so-promising results. I would like to thank you for leading a fun and supportive lab group, for all your help with writing, for your belief in me, and for your limitless support throughout the project. And of course, also for inviting me over to the beautiful city of Toulouse (three times!). I am sure you will continue your excellent scientific work and inspire future generations of scientists who are lucky to work with you.

Pascale, thank you for welcoming me at IMB and introducing me to the wonderful field of synthetic genomics. You made me feel right at home in the first months of my PhD, when we discussed everything from personal interests to science to practical matters in our weekly meetings. Throughout my PhD, you created a safe environment to talk openly about how things were going, even in times when I found it difficult to admit that I was struggling with motivation or progress. I really appreciated your help in setting priorities and deadlines, and in structuring the writing process during our dedicated meetings. Thank you for believing in me and for your encouragement, not only about my project and future career, but even while running my first 5k race! I'm also grateful for your efforts as director of the FGS. I knew that the well-being of the PhDs in the faculty was in good hands during your term, in which you implemented the PhD-in-4 procedure to protect PhDs and help them move forward in both their professional and personal lives after their contracts ended. I'm very happy to know that the genome construction efforts at IMB are continuing, and that your synthetic genomics research line is growing, with both awesome new equipment and great people!

I would also like to thank the other members of my PhD committee: Prof.dr. **Gijsje Koenderink**, Dr. **Nico Claassens**, Prof.dr. **Bert Poolman**, Dr. **Carole Lartigue-Prat**, Prof.dr. **Matthieu Jules** and reserve member Prof.dr.ir. **Stan Brouns**. Thank you for taking the time to critically read my dissertation. I'm looking forward to discussing the

contents with you during my defense! In particular, I would like to thank **Gijsje** for the feedback during my Go/No go meeting and the occasional chats and encouragements in the hallways of BN, **Nico** for the fun discussions during BaSyC meetings with and for inviting me to join your team outing on the day of the MAGE workshop in Wageningen and **Stan** for the supervision during my MEP, introducing me into the fascinating world of phages and anti-phage defense systems.

Indispensable to my project, both scientifically and for the great energy they brought, were the talented and fun students I had the pleasure of supervising. **Laura**, I was super lucky to have you as my first student. You greatly contributed to the project with your hard work, perseverance and efficiency and you required very little guidance. It was a lot of fun to work with you and I was happy that you wanted to continue on the synthetic cell project, now as a PhD in Toulouse! You're always willing to help others and play a big role in both the lab's spirit and scientific progress. The design and assembly of pSmallER, pLarge and pHUGE (and pHUGER, pHUGEST, p...?) is certainly in super capable hands with you. And thank you for being my survival buddy in the Pyrenees, I'm happy to know that I wasn't the only one mentally planning our helicopter rescue ;). **Bastiaan**, you are a super social person with genuine interest in other lab members and a lot of enthusiasm for presenting and teaching. Thank you for your contributions to the plasmid and synthetic chromosome isolation from yeast, which resulted in the final protocol as described in this dissertation. I also admire your activities outside your studies: inspiring a new generation of students and standing up for a more humane world. **Rebecca**, you took on a challenging project in which you were the first to try *in vitro* replication of plasmids and chromosomes isolated from yeast. You are eager to learn and took a deep dive into the theoretical background of your project. I'm impressed by your computational skills and beautiful presentations. I was happy to hear that after a Master's in Toulouse, you're continuing your academic career with a PhD in Leiden! I am sure you will continue to do great scientific work. **Ellen**, I am very grateful that you joined the project in the last stage of my PhD. I don't know what would've become of this dissertation if it weren't for your hard work and great results. You joined the project with a lot of enthusiasm, showed high independence and efficiency in the lab, and are just a very "gezellig" person in general. It was fun to travel to Toulouse together and I wish you lots of fun on your current (and future?) international adventures.

My scientific journey into the world of minimal cells started during my BEP in the ChD group. To all the people that were around at the time: you made it such an enjoyable experience that I came back for my PhD! Especially **Jonás**: thank you for being an amazing supervisor and for your dedication to the Min project. We were a great team, both in the lab and at the kicker table! **Duco**, **David**, **Alicia** and **Johannes**, I loved your humor and gezelligheid.

I'd like to continue with thanking the amazing ChD lab members at the time of my PhD. **Ilja**, **Elisa**, **Ana**, **Marijn**, **Fede** and **Amélie**: thank you for sticking together and looking out for each other during the challenging period that we faced. **Andreea**, I will remember your passion for science and the good moments that we shared. **Ilja**, thank you for welcoming me during both my BEP and PhD and showing me around the lab. You were a great support during the difficult period that we went through. I was happy to still have you around when you transferred to IMB. Thank you for always being available for any molbio-related questions. **Elisa**, we shared our excitement for the Min system, and I

was happy that you were still around when I joined for my PhD. I admire your dedication and appreciated both the science- and non-science related chats that we shared. **Ana**, you brighten up the room with your enthusiasm and positive attitude. Any plans for food, sports or trips are always welcomed or initiated by you with great excitement. You are a dedicated and collaborative scientist and always willing to help others. Thank you for caring so much about your colleagues and friends, and keep up the positive energy! **Marijn**, you are a caring friend and great scientist. You're always there to help, to chat or to plan fun activities together. I admire your creativity both in- and outside the lab and love to listen to your stories about amazing hikes or other outdoor adventures. I'm very happy to have you as my paranymph! **Fede**, you are one of the most social people I know. You care a lot about the people around you and are always in for fun. Your dedication to your work is admirable and I am sure you will come far with your perseverance and collaborative spirit. Thank you for introducing me to tacos al pastor, tacos de lengua, and carne en su jugo, and many other Mexican dishes that I forgot the name of ;). And I'm sure my PhD movie is in great hands with you as paranymph (you will wear a dress, right?)! **Ana**, **Marijn**, **Fede**: we shared a million good moments together. I thoroughly enjoyed our trips to Toulouse (including exciting food and hikes) and the USA, countless dinners at one of our places or nice restaurants, the sports events that we joined (running, swimming, lacrosse) and making (PhD) movies together. I am sure we will share many more fun experiences. I'd also like to thank the students that joined the ChD lab and that both contributed to the science and atmosphere of the lab: **Amélie**, **Kelly**, **Marlena**, **Brent**, **Haroun**, **Rosalie**, **Anna**, **Jesse** and **Enrica**.

Halfway through my PhD, the ChD lab partly relocated to Toulouse, where a great bunch of people joined the group. **Laura**, **Yannick** and **Thibault**: I am happy that you're continuing the awesome science of the ChD group. It was fun to join SynCell together and I was happy to work with you during my short postdoc that included a 5-week stay in Toulouse. You bring a great atmosphere and I love the memes! :)

When I joined IMB for my PhD, I found a great group of people with a lot of social interaction who are always willing to help each other out. To start with, I would like to thank the indispensable staff that make IMB run smoothly, practically (in the lab, *in silico* and in the offices) and socially. **Erik**, you made me feel very welcome from the start, and I loved teaching BBT1 together. You are the social motor of IMB, always initiating and joining fun activities and caring about the well-being of others. I have never touched a bioreactor, but you were always willing to help me out with other practical questions. **Marijke**, you are a very kind person and always open to any questions or support. Thank you for helping me out with FACS and DNA isolation from yeast, but most importantly: thank you for fostering a healthy social environment at IMB. **Clara**, thank you for taking care of the most important lab at IMB and for helping me out with any practical questions. I admire your dedication to both your work and your family, and I appreciate the positive feedback that you gave me. **Marcel vdB**, your help with sequencing data analysis was indispensable. You patiently taught me to work with the command line and to understand each step of the analysis pipeline. I also enjoyed the chats we had during any IMB activities after work. **Michelle** and **Susan**, thank you for all your administrative support and your initiatives to send cards to IMB members on sad or happy occasions. **Jannie**, **Gea** and **Apilena**, thank you for taking care of all the laboratory supplies and glassware. Your friendliness and willingness to help out,

both during stressful experimental days and in the preparation of teaching or science communication activities, are much appreciated.

I'd also like to thank the PIs of IMB that contributed to the great work atmosphere at IMB and that gave valuable input during work discussions: **Jack**, **Jean-Marc**, **Robert**, **Walter**, **Rinke**, **Djordje** and **Marcel VL**. **Jack**, thank you for your dedication to the functioning and atmosphere of the department. I appreciated your wise advice over the years, for example during meetings with the PhD committee or when you shared your personal story with all PhDs of BT. **Jean-Marc**, thank you both for your valuable scientific input during the SynBio meetings, and for your efforts as group leader of IMB.

Nicole, you're a very "gezellige meid" and a good friend. Your work ethic and perseverance are admirable, as well as your creativity and dedication to sports. I was honored to be your paranymph together with Marieke and to attend your beautiful wedding! All other work- and non-work related things we did together were a lot of fun as well, and I'm sure we'll experience many more nice moments together! **Charlotte**, thank you for showing me around in the first months of my PhD. I'm impressed by your perseverance and beautiful scientific illustrations. Your dedication to science communication has been an inspiration to me. **Rozanne**, from the first time that I met you as my new office mate, I found you a very likable, easy-going and fun person. You have taken some great initiatives like the orphan BBQ, suppen in the Delftse Hout and pumpkin carving, which I all really enjoyed! I hope you'll never stop being as excited about little lambs as I am :). **Elynor**, it was fun to work together with you during the first months of your postdoc and nice to have someone to share the struggle of isolating synthetic chromosomes from yeast. I appreciate your kindness and interest in others' well-being, and I really enjoyed going to Toulouse and to BaSyC events together. **Marieke**, you're gezellig, down-to-earth and always thinking along with other people's projects. I really enjoyed preparing Nicole's defense together with you. Lots of luck with the finalization of your PhD! **Denzel**, thank you for the gezelligheid during the many IMB outings we joined and for the fun we had brewing IMBeer. I wish you all the best with the finalization of your PhD and enjoy fatherhood! **Aafke**, **Sophie** and **Sanne**, you were all very welcoming and friendly when I joined IMB and I love your creativity! **Rosalie**, **Effie**, **Miriam**, you bring positive energy and I appreciate your directness. Thank you for your inputs during SynBio meetings. **Miriam**, I'm happy to see you share my enthusiasm for science communication! **Tobi**, you bring the party spirit and are always up for fun. I love your enthusiasm and listening to your stories. **Sagarika**, it was fun brewing IMBeer together and having you around at the many IMB gatherings. Good luck with the final bits of your PhD! **Flip**, I appreciate your directness and it was fun to talk to you at the many IMB gatherings. **Annemieke**, **Dani** and **Timon**: I'm happy I got to know you and your enthusiasm during my short postdoc. I'm confident that the genome of our minimal cell is in good hands with you. Good luck with all of your projects!

The list of lovely people at IMB is endless and I would like to thank **everyone at IMB** who contributed to the fun and the welcoming environment and who helped me out in one way or another. I will always remember the "gezellige" moments we shared: the countless cake times, lunch breaks and Friday drinks, lab outings online and in de Biesbosch, brewing IMBeer together for the fermentation festival, barbecues, game nights, pumpkin carving, organizing a lab workshop for the TU Delft science day, swimming in de Schie to raise money for cancer research, going out for an escape room, jeu de

boules, or karaoke and celebrating Sinterklaas with beautiful surprises.

I would also like to thank **everyone at BN** with whom I shared nice lunch breaks, went on fun outings and who helped me out in the lab. Some people I'd like to mention in particular: **Beatriz**, you're a sunshine and I really enjoyed our time together in Toulouse and at BaSyC meetings. **Essie**, you are such a cheerful presence in the department and I'd like to thank you for the fun chats we had. **Marianne**, **Jaïrus**, **Fede** and **Marijn**: it was nice organizing the alumni career day together and I think it was a big success! **Jan**, you are indispensable to BN and I'm grateful for the many times you helped me out. **MSR staff**, especially **Anke**, thank you for taking care of the glassware, media, packages and everything else. **Tracey** and the other ladies at the **BN secretariat**, thank you for your administrative support and the organization of many nice social events during Quo Vadis.

I would like to extend my thanks to the members of the **BaSyC** consortium inside and outside of the TU Delft. I enjoyed the meetings, workshops and conferences that we attended together and I have learned a lot from you!

One of the "duties" of a PhD student is teaching. This was actually one of my favorite elements, thanks to the lovely people that I taught together with. Thank you **Pascale** for putting me on the BBT1 team. I had a very good time together with **Erik**, **Jannie**, **Zita**, **Ben** and others. I'm still amazed by the light-emitting bacteria that we isolated. Also thank you **Greg** for having me as TA for Metabolic Systems Biology, I really enjoyed teaching the course with you and I admire your genuine interest in and patience with students.

During my first year, I joined two committees with whom I co-organized activities and had a good time. Thank you to **Timmy**, **Sam**, **Mariana** and **Susan** the fun moments we shared as BT PhD committee! No one can beat our promotion material, and despite the "interesting" opinion of the location on the corona measures, I think the BT symposium was a big success. **Martijn** and **Alberto**, we were a great team as new "board" of the AS PhD council. And **Martijn**, thank you so much for encouraging me to join the Hoe?Zo! show!

Speaking of the Hoe?Zo! show: joining this was one of the best decisions of my PhD. A big thank you to **Barbara**, **Lennart** and **Boy** for initiating this project and the organization. To all other PhDs that I shared the stage with: I had so much fun with you and I'm curious to hear about your future science communication activities.

Gelukkig waren er ook veel lieve mensen buiten werk die ervoor hebben gezorgd dat ik erg van mijn PhD tijd heb genoten, door de vele weekendjes weg, vakanties, lacrossetrainingen, wandelingen, etentjes, lacrosse- en hardloopwedstrijden, spelletjesavonden, feestjes, en andere uitjes die we samen hebben beleefd. **Alexandra**, **Sophia**, **Nina**, **Lotte**, **Anna**, **Ilse**, **Oud Huize Gekko**, **(oud-)Diamantjes**, **Totem** en **(k)Oele**: dank jullie wel voor alle gezellige momenten en goede gesprekken. Ik voel me heel gelukkig met zulke fijne vrienden!

Ook heb ik een hele lieve en betrokken schoonfamilie. **Geert** en **Marijke**, wat heb ik een geluk met zulke schoonouders. Jullie hebben me met open armen in de familie ontvangen. Ik waardeer alle gezellige etentjes, fantastische vakanties, goede gesprekken, jullie geduld met mij op de piste, en jullie interesse in mijn PhD enorm. **Dani** en

Pat, dank jullie wel voor de gezelligheid en leuke vakanties samen. **Carla**, dankjewel voor alle gezellige etentjes over de jaren en alle krantenberichten over stamcellen, fagen, embryo-onderzoek en synthetische cellen die je naar me hebt gestuurd. Nadat ik helaas last-minute een etentje moest afzeggen omdat mijn labwerk uitliep, was je na het weekend zó benieuwd of mijn experiment was gelukt, dat je de eerste was die ik een foto van mijn oranje en witte gisten stuurde. Dit waren de gisten die uiteindelijk het synthetisch chromosoom bleken te bevatten waar hoofdstuk 4 over gaat! Ik heb veel bewondering voor hoe jij in het leven staat.

Natuurlijk was ik nooit tot dit punt gekomen zonder mijn allerliefste familie. **Papa** en **mama**, jullie hebben mij altijd onvoorwaardelijk gesteund, zijn trots op alles wat ik doe en zijn de beste ouders die ik me kan bedenken. **Mama**, jij bent er altijd voor een luisterend oor, een knuffel, gezelligheid, een aanmoediging of een attent berichtje. Ik weet dat ik alles met jou kan delen, dat je altijd klaar zal staan en dat ik straks tijdens mijn verdediging jouw trotste gezicht op de eerste rij zal zien en zal weten dat alles goed komt. **Papa**, als kind kon ik bij jou terecht met al mijn vragen over het ontstaan van vulkanen, de evolutie van de mens en de oorsprong van de fossielen die we vonden. Mijn keuze voor een wetenschappelijke studie en promotietraject waren dan ook een logisch vervolg van jouw geduldige uitleg en enthousiasme voor de historie en werking van de wereld om ons heen. Je interesse in mijn onderzoek, aanmoedigingen voor het hardlopen en de vele keren dat je me met de auto thuis hebt gebracht waardeer ik enorm.

Mervin, mijn stoere grote broer met zijn heerlijke kookkunsten en droge grapjes. Ik kan altijd rekenen op jou als trotse broer bij belangrijke gebeurtenissen. **Lieanne**, ik vind het heel gezellig als je bij familiegelegenheden bent en ik waardeer je creatieve kaartjes enorm. **Diane**, mijn stoere grote zus en grote voorbeeld. Wat was ik vereerd toen je me vroeg als ceremoniemeester van de bruiloft van Sylvain en jou en als paranimf tijdens je promotie. Ik ben heel blij met zo'n lieve zus als jij! **Sylvain**, secretly I do appreciate your bad puns and I'm happy to have you as my brother-in-law. **Lukas** en **Enora**, wat was ik ontzettend blij toen ik hoorde dat ik een neefje en nichtje zou krijgen. Wat een lieve en gezellige kinderen zijn jullie! En wie weet, ook kleine wetenschappers in de dop ;).

Lieve **Rik**, al meer dan 10 jaar ben ik heel gelukkig met jou. Tijdens mijn PhD was ik niet altijd even gezellig, als ik weer eens laat uit werk kwam door een uitgelopen experiment, 's avonds tot de vroege uurtjes bezig was met schrijven, of niet veel los liet over de dingen die me dwars zaten. Maar jij was er altijd met een luisterend oor en om me op te vrolijken met taart, een lief liedje of een bezoekje aan de schaapjes. Ik weet dat met jou aan mijn zijde alles goed komt. Dankjewel voor alle steun en aanmoedigingen en alle geweldige kleine en grote avonturen die we samen hebben beleefd. Ik heb ontzettend veel zin in de rest van ons leven samen!

Curriculum vitae

Céline Verena Cleij was born on June 16th, 1996, in Rotterdam, the Netherlands. She attended pre-university education from 2008 to 2014 at het Erasmiaans Gymnasium in Rotterdam, where she developed her interests in languages (French, Spanish, Latin, Mandarin, English) and science (physics, chemistry, biology and mathematics). During her time at het Erasmiaans, she competed in the 6th Chinese Bridge, a Chinese Proficiency Competition for Foreign College Students, obtained her Cambridge C2 Proficiency in English and taught weekly Spanish lessons at Public Elementary School 'Het Landje'.

She moved to Delft in 2014 to pursue her passion for science in a Bachelor's Nanobiology, a joint programme at Delft University of Technology and Erasmus University Rotterdam. During her Bachelor's, she was an active member of her lacrosse club and was responsible for its operations and activities as president during her board year. She pursued her long-time interest in human evolution, sparked by her father, by completing her minor at Leiden University on this topic. Her first steps into the field of synthetic cells were taken during her Bachelor's End Project in the Christophe Danelon lab at the Delft University of Technology, under the supervision of Jonás Noguera López. Her project focused on reconstituting the Min system *in vitro*, which looked beautiful under the TIRF microscope. She obtained her BSc degree in 2018.

Céline continued her studies with a Master's degree in Nanobiology at Delft University of Technology and Erasmus University Rotterdam. During her first year, she was a member of the organizing committee for the Applied Sciences Connect career event of the Faculty of Applied Sciences. In her second year, she moved to Montpellier, France, to conduct her Master's internship at Ciloa, a small company that utilizes exosomes for therapeutic applications. She took the opportunity to improve her proficiency in French by attending a two-week summer school at Institute Linguistique Adenet. Céline returned to Delft for her Master's End Project in Stan Brouns' group, under the supervision of Franklin Nobrega, where she investigated an anti-CRIPSR protein from an *Acinetobacter baumannii* prophage. She obtained her MSc degree in 2020.

Inspired by her Bachelor End Project and the fascinating Master's course on the Origins of Life, Céline returned to synthetic cell research in 2021 for her PhD project, of which the results are presented in this dissertation. Her PhD work was performed at the Faculty of Applied Sciences, in the groups of Christophe Danelon at the Department of Bionanoscience and Pascale Daran-Lapujade at the Department of Biotechnology, section Industrial Microbiology. During her time as a PhD candidate, Céline was a member of the PhD council of the Faculty of Applied Sciences and the PhD committee of the Department of Biotechnology, with whom she co-organized the annual departmental symposium. For the Department of Bionanoscience, she co-organized an alumni career day. She enjoyed teaching the first year's practical course on Basic

Biotechnological Techniques and the Master's course on Metabolic Systems Biology. Her interest in science communication grew as she the co-organized a microbiology workshop for the university's Science Day, and took part in the Hoe?Zo! show, a science theatre show aimed at fostering the curiosity of primary school children by allowing them to ask scientists their burning scientific questions.

List of publications

Cleij C, Sierra Heras L, Zwiers E, Stella A, Bancal E, Daran-Lapujade P, Danelon C. Assembly and cell-free expression of a partial genome for a synthetic cell. *Manuscript in preparation for submission*.

Koster CC*, Postma ED*, Knibbe E*, **Cleij C***, Daran-Lapujade P. Synthetic Genomics From a Yeast Perspective. Frontiers in Bioengineering and Biotechnology. 2022;10(March):869486. https://doi.org/10.3389/fbioe.2022.869486. **These authors have contributed equally to this work and share first authorship*.

Godino E, Noguera López J, Foschepoth D, **Cleij C**, Doerr A, Ferrer Castellà C, Danelon C. De novo synthesized Min proteins drive oscillatory liposome deformation and regulate FtsA-FtsZ cytoskeletal patterns. Nature Communications. 2019;10(1):4969. https://doi.org/10.1038/s41467-019-12932-w


TTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACA
ACATATGGCACGCATTATTGTTGTTACTTCGGGCAAAGGGG
CCCAGAAGGGAAAGAAAACTGTCGTGATAGATTTTGATATC
GTTTACGATTTCGTCAACGTCATTCAGGGCGATGCAACGCT
TCTGCCGGCATCGCAAACACGCGATAAAGATGCCCTCACCC
TGAATTTATCGTTTGTGACTCCCCGGCAGGGATTGAAACC
ACCAACCCGGAAGTCTCCTCAGTACGCGACTCTGACCGTAT
AGAGCCTATTAAAGAGCACCTGCTGTTAACGCGCTATAACC
TGGAGATCCTGCGCATCAAACTCGTCGGCGTGATCCCAGAG
CTCGACATTAACGCCGATGCGGGTAAAGCCTACGCAGATAC
AGAAGAGAAGAAAGGCTTCCTCAAACGCTTGTTCGGAGGAT
CCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGC
GGCCCGTCGACTGCAGAGGCCTGCATGCAAGCTTGGCGTA
GGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATT
TCTTTCTGAGCCGTAAGAAAAACACCGCGAACATCGCGAAA
AACCGCACTACCTGCCGCAGCTGCGTAAAGATATCCTGGA
TCAGCTGGAACAAAAGGACGGTGATATCAGCATTCTGGAGC
CTGCTAACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGC
GTCTTGAGGGGTTTTTTGCTGCCCGCTTTCCAGTCGGGAA
CGAGATGTCAAGAGACTTAACGCACCCATGAACCACACGG
TGCTGCCACCCTCACCTCACTAACGCTGCGGTGTGCGGATA
AACCAAATGGGAAAATCGGAATGGGTCCAGAACTGCTTTG
GCGCGCCCCTCTCAAAACTCCGCACAAGTCCCAGAAAGCG
TCGAAGCGTGGGTGGTAGGCCCTGGATTATCCCGTACAAG
TCGCGGCCACCTACGCCGCTATCTTTGCAACAACTATCTGC
TGGGAGTCTTACTTCCAACATAACGGCAGAAAGAAATGTG
ATGCTTGATAATCTTTCTTTCCATCCTACATTGTTCTAAT
TCTTCTGTCATTCGCTTAAACACTATATCAATAAAGATCTA
AGTTGTCATGGAGGGCTCCGTTAATGGTCACCAATTCAAG
TGAGAATTAAAGTAATTGAAGGCGGACCACTACCATTTGC
ATCAAGTACCCTAAAGGTATACCAGACTTTTTCAAGCAAT
TGGGGGTGTCGTCACAGTTATGCAAGATACTTCATTAGAA
CTTCTAACGGACCTGTGATGCAGAAAAAGACCAAAGGTTG
GGCTACACACACATGGCGCTTAAAGTTGATGGTGGAGGTC
TGGCAATATCAAAATGCCAGGAATACATGCTGTAGACCAC
AGAGAGAGCATGCCGTAGCCAAATTTGCTGGCTTAGGCGG