Spatially Confining Translation to Enable Optimized Genetic Code Expansion in Eukaryotes

---

Dissertation

submitted to the

Department of Chemistry, Pharmaceutical Science, Geography and Geosciences

at the Johannes Gutenberg University Mainz, Germany

for the degree of

Doctor of Natural Sciences

in Chemistry

presented by

Christopher Dieter Reinkemeier

born in Gütersloh, Germany

Heidelberg, 28.10.2020

Defense date: 29.01.2021
D77 (Dissertation Mainz)

Versicherung

für das Gesuch um Zulassung zur Promotion in dem Fachbereich 09

Hiermit versichere ich gemäß § 10 Abs. 3d der Promotionsordnung vom 24.07.2007

a) Ich habe die jetzt als Dissertation vorgelegte Arbeit selbst angefertigt und alle benutzten Hilfsmittel (Literatur, Apparaturen, Material) in der Arbeit angegeben.

b) Ich habe die jetzt als Dissertation vorgelegte Arbeit nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht.

c) Ich hatte weder die jetzt als Dissertation vorgelegte Arbeit noch Teile davon bei einer anderen Fakultät bzw. einem anderen Fachbereich als Dissertation eingereicht.

_____

Christopher Dieter Reinkemeier
Heidelberg, 28.10.2020

Abstract

The genetic code is the operating system of every living cell. It is executed through the central dogma of molecular biology and defines how the ribosome translates genetic information into a polypeptide.

The genetic code is conserved throughout all domains of life and encodes three stop codons as well as the incorporation of the canonical amino acids. The naturally occurring amino acids can be classified in four different chemical functionalities (nonpolar, polar, basic and acidic). Despite this simplicity the genetic code gives rise to the entire diversity seen across all kingdoms of life and it can only be imagined how life might look like if the ribosome could also incorporate noncanonical amino acids (ncAAs), covering the full spectrum of chemical functionalities, into proteins.

Genetic code expansion (GCE) is a powerful method to site-specifically incorporate ncAAs into proteins *in vivo*. To this end, typically an orthogonal aminoacyl-tRNA-synthetase/tRNA (RS/tRNA) suppressor pair is used to reassign a rare stop codon to be read as a sense codon. During the last decades of research GCE has been established to permit the genetic incorporation of a plethora of ncAAs, which have for example been used to enable site-specific protein labeling for super-resolution microscopy.

However, for imaging applications as well as for applications aiming to synthesize fully artificial polymers in eukaryotes, the current technology has at least three major limitations.

First, GCE is codon specific but it cannot distinguish the mRNA of the protein of interest from endogenous mRNAs, leading to recoding of untargeted codons in the transcriptome. Due to this activity, GCE can have adverse side effects or even be toxic. Second, only a few orthogonal RS/tRNA suppressor pairs have been established for eukaryotic systems and third, only two different stop codons can be suppressed at a time.

In this cumulative thesis I address the first problem, by developing synthetic membraneless organelles that allow to selectively translate only selected mRNAs with an expanded genetic code (**Chapters 2&3** and **Appendix II**). An alternative development using inducible expression systems, allowing to regulate the GCE components, is presented in **Chapter 2** and **Appendix IV**. I further develop multiple mutually orthogonally translating organelles to equip cells with multiple genetic codes, which represents a new way to obtain orthogonal RS/tRNA suppressor pairs and enables to multiple times reassign the same stop codon, thus solving the remaining two major contemporary limitation of GCE (**Chapter 5**).

Zusammenfassung

Der genetische Code ist das Betriebssystem jeder lebenden Zelle. Er wird durch das zentrale Dogma der Molekularbiologie ausgeführt und bestimmt wie das Ribosom genetische Information in eine Polypeptidsequenz übersetzt.

Er ist konserviert in allen Domänen des Lebens und codiert drei Stopcodons sowie den Einbau der kanonischen Aminosäuren. Die natürlich vorkommenden Aminosäuren können in vier unterschiedlich chemische Funktionalitäten klassifiziert werden (unpolar, polar, basisch und sauer). Trotz dieser Schlichtheit, führt der genetische Code zu der gesamten Diversität die in allen Königreichen des Lebens zu sehen ist und man kann sich nur vorstellen wie das Leben aussehen könnte, wenn das Ribosom nicht-kanonische Aminosäuren (ncAAs) in Protein einbauen könnte, die das gesamte Spektrum chemischer Funktionalitäten abdecken.

Die Erweiterung des genetischen Codes (GCE) ist eine wirksame Methode um ncAAs seitenspezifisch *in vivo* in Protein einzubauen. Dazu wird normalerweise ein orthogonales Aminoacyl-tRNA-synthetase/tRNA (RS/tRNA) Suppressionspaar verwendet, um ein seltenes Stopcodon als Sensecodon neu zu belegen. In den letzten Jahrzehnten wurde der Einbau einer Vielzahl von ncAAs mithilfe der GCE Forschung etabliert, welche zum Beispiel verwendet wurden, um seitenspezifische Proteinmarkierung für hochauflösende Mikroskopie zu ermöglichen.

Allerdings ist die GCE Technologie sowohl für Mikroskopie als auch für die Herstellung vollständig künstlicher Proteine in Eukaryoten durch mindestens drei große Probleme limitiert. Erstens, GCE ist codonspezifisch, kann aber nicht die mRNA des Zielproteins von anderen mRNAs unterscheiden, wodurch weitere Codons im Transkriptom unterdrückt werden können, was toxisch sein kann. Zweitens, nur wenige orthogonal RS/tRNA Paare sind für Eukaryoten etabliert, und drittens maximal zwei Stopcodons können gleichzeitig unterdrückt werden.

In dieser Arbeit habe ich das erste Problem adressiert, indem ich synthetische, membranlose Organellen entwickelt habe, die es erlauben nur ausgewählte mRNAs mit einem erweiterten genetischen Code zu übersetzen (**Chapter 2&3** und **Appendix II**). Eine alternative Entwicklung von induzierbaren Expressionssystemen, die es erlauben GCE Komponenten zu regulieren ist in **Chapter 2** und **Appendix IV** präsentiert. Des Weiteren, habe ich mehrere zueinander orthogonale Organellen entwickelt, die Zellen mit multiplen genetischen Codes ausstatten, was einen komplett neuen Weg darstellt um orthogonale RS/tRNA Suppressions-paare zu erhalten und es ermöglicht das gleiche Stopcodon mehrfach neu zu belegen, dadurch werden die verbleibenden zwei aktuellen Probleme der GCE Technologie gelöst (**Chapter 5**).

Publication list

1. Koehler C, Estrada Girona G, Reinkemeier CD & Lemke EA (2020). Inducible Genetic Code Expansion in Eukaryotes. ChemBioChem, doi: 10.1002/cbic.202000338
2. Reinkemeier CD & Lemke EA (2020). Raising the ribosomal repertoire. Nature Chemistry, 12, 503-504, doi:10.1038/s41557-020-0476-6
3. Reinkemeier CD*, Estrada Girona G* & Lemke EA (2019). Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. Science, Vol 363, Issue 6434, doi:10.1126/science.aaw2644

* These authors contributed equally.

These publications are part of this cumulative thesis and are summarized specifically in **Chapter 2** and attached as **Appendices II-IV**.

Acknowledgements

Progress in science depends on new techniques, new discoveries and new ideas, probably in that order.

Sydney Brenner

First, I would like to thank my supervisor Edward for his constant support and mentorship during the last years. I am grateful that he has been so excited about this PhD project and I would like to thank him for his may valuable ideas to make the project successful. On top I really appreciate that he gave me enough space to also explore my own ideas, making the experience of doing my PhD under his guidance truly rewarding.

I would also like to thank the members of my thesis advisory committee, Anne E., Martin B., Dirk S., Bernd B. and Kai J. for their great support and extremely valuable feedback for the project.

I am further indebted to the all the amazing current and former members of the Lemke group. First, I would like to thank my dear friends Christine and Aritra. Christine, not only for her lightning help in the lab, the fruitful collaborations and superb feedback on my writing, but also for emotional support when discussing the future, the past and the present. To Aritra, who is an inexhaustible source of knowledge with whom it is always fun to not only discuss science, but who is also a most generous and loyal friend.

I would like to thank Leonie, an awesome master student, who made great contributions to the further development of OT organelles, and with whom it was a great pleasure to work and share ideas. Gemma, for her great work and critical thinking at the beginning of the project, as well as Iker, for exciting and insightful discussions. I would also like to thank Nike, for interesting bioengineering project discussions. Mikhail, for questioning the fundamentals of synthetic organelle assembly, as well as Marius and Rajanya the newest members of the synthetic organelle team. I would like to thank Nataliia for her enthusiasm for side projects and click-labeling discussions. Paul, thanks for quick synthesis and chemical help, for the semi-appropriate jokes and sarcasm. I would also like to thank Sofya for showing me how to acquire and analyze super-resolution images and for interesting discussions about science and society. Piau Siong, the labs former great optical engineer. Joana for always being helpful and for keeping everything in order. Tom, for his critical thinking on organelles and protein

aggregation. Miao, with whom it is always fun to discuss the science of labeling and who is brave enough to try out OT organelles for advanced biological applications. Cathrin for the great organizing support. I would also like to thank Jun, Giulia, Giorgia, Sara and Hao for contributing to a great atmosphere in the group, as well as Daniel, Kai and Moritz for great coding work.

Thanks to Malte P. and Diana O. from the Flow Cytometry Core Facility, as well as Marko L. and Stefan T. from the Advanced Light Microscopy Facility for excellent help.

I would like to thank Benjamin B., for always being there when I need advice, for his quick and thorough feedback on my writing, as well as for engaging discussions about science and beyond. Furthermore, I would also like to thank all my friends from University and back home for making life more interesting and enjoyable.

I would like to thank my family for their unwavering and unconditional support, not only during the last four years of my studies but throughout my entire life. My grandmother Mathilde, who's energy and independence have always been a true inspiration. My grandfather Bernhard, who is a model of unrelenting strength and staunchness. My sister Lisa, who is always the first person I turn to, to discuss problems and challenges and who always has the best advice to overcome them. My niece Emma, an endless source of joy, and my brother in law Fabian. My parents, who truly made me who I am, with their tireless encouragement and reliability they always enabled me to pursue my dreams.

Finally, I would like to thank Isabell, my partner in life, who supports me throughout and endures my occasional obsession with ridiculous problems.

Table of contents

List of abbreviations

3-IF	3-iodophenylalanine
ASD	anti-Shine-Dalgarno
BSA	bovine serum albumin
C. elegans	*Caenorhabditis elegans*
DMNB-S	O-(4,5-Dimethoxy-2-nitrobenyl)-L-serine
dsDNA	double stranded DNA
E. coli	*Escherichia coli*
EB1	microtubule-end binding protein 1
ELP	elastin-like polypeptides
ERM	endoplasmic reticulum membrane
EWSR1	Ewing's sarcoma breakpoint region 1
FBS	fetal bovine serum
FFC	fluorescence flow cytometry
FISH	fluorescence *in situ* hybridization
FSC-A	forward scatter area
FUS	fused in sarcoma
GCE	genetic code expansion
GENESIS	genome stepwise interchange synthesis
GM	Golgi membrane
HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
hnRNP	heterogenous nuclear ribonucleoprotein
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IF	immunofluorescence
IRES	internal ribosomal entry site
K_D	dissociation constant
KIF13A	kinesin-like protein KIF13A
KIF16B	kinesin-like protein KIF16B
LAF-1	ATP-dependent RNA helicase LAF-1

LAT	linker for activation of T cells
LLPS	liquid-liquid phase separation
MCP	ms2 bacteriophage coat protein
mRNA	messenger RNA
MTOC	microtubule organizing center
ncAA	noncanonical amino acid
NPC	nuclear pore complex
OMM	outer mitochondrial membrane
OT organelle	orthogonally translating organelle
pBPA	4-benzoyl-L-phenylalanine
PBS	phosphate buffered saline
PCP	PP7 bacteriophage coat protein
PEI	polyethyleneimine
PLD	prion-like domain
PLK1	polo-like kinase 1
PM	plasma membrane
POI	protein of interest
PolII	RNA polymerase II
PolIII	RNA polymerase III
PylRS	pyrrolysyl aminoacyl-tRNA synthetase
RBD	RNA binding domain
REXER	replicon excision for enhanced genome engineering through programmed recombination
RF	release factor
RM	RNA motif
RNP	ribonucleoprotein
RRM	RNA recognition motif
rRNA	ribosomal RNA
RS	aminoacyl-tRNA synthetase
RT	room temperature
rtTA	reverse tetracycline-repressor protein

SCO	cyclooctyne lysine
SD	Shine-Dalgarno
SGs	stress granules
SIM	SUMO interacting motif
SLiMs	short linear motifs
SPD2	spindle-defective protein 2
SPD5	spindle-defective protein 5
SRM	super-resolution microscopy
SSC	saline sodium citrate
SSC-A	side scatter area
SSC-W	side scatter width
SSOs	semi-synthetic organisms
SUMO	small ubiquitin-like modifier
TAF15	TATA binding protein-associated factor 2N
TCR	T cell receptor
TDP43	TAR DNA-binding protein 43
TetO	tetracycline operator
TetR	tetracycline repressor protein
tRNA	transfer RNA
$tRNA^{Pyl}$	pyrrolysyl tRNA
UBP	unnatural base pair

Chapter 1 Introduction

Proteins are major biomolecules that execute the information stored in the genome and are involved in almost all biological processes. It is thus critical to understand how they function. One approach to analyze proteins is via classical *in vitro* biochemistry. Here, proteins are commonly purified and studied in isolation. However, proteins typically function in the crowded environment of a cell, which differs from neat *in vitro* conditions and it is therefore particularly interesting to study proteins in their native environment.

But how can proteins be visualized and interrogated directly in a cell?

Traditional methods, like for example immunofluorescence microscopy, are powerful to visualize proteins in fixed cells and offer a high signal to noise ratio, high resolution and exquisite molecular selectivity. However, these methods only give snapshots of protein localization at a defined time point. It would be even more interesting to study their dynamics over time. To do so, it is necessary to attach the fluorophore to the protein of interest (POI) directly in a living cell. The discovery of fluorescent proteins that can be genetically encoded and fused to the POI has been a major driver of this technology (*1*). However, the most commonly used fluorescent proteins are comparatively large and do not have optimal photophysical properties for example for super-resolution microscopy (SRM) techniques (*2*).

Therefore, several methods have been reported that enable to directly attach bright and photostable dyes to proteins *in vivo* for example self-labeling proteins like SNAP- (*3*), CLIP- (*4*) or HALO-tags (*5*). However, these are still several nanometers big. Some alternative *in vivo* methods that only use short peptide tags like FLAsH (*6*) and PRIME tags (*7*) to label the protein have also been reported, but these are still several amino acids large.

The most recent developments in SRM are pushing the optical resolution below 1 nm (*8*) and thus the size of the label on the POI becomes limiting. Therefore, the ideal solution would be to directly attach a fluorophore to the POI with single residue precision—a problem which appears to be tailored to be addressed by genetic code expansion (GCE) technology.

1.1 The central dogma of molecular biology and protein synthesis

GCE technologies engineer parts of the central dogma of molecular biology. Thus, I am going to explain how the central dogma works in this section and I will subsequently discuss, how it can be modified.

The central dogma of molecular biology describes how genomic information, stored as a nucleic acid, is transferred into a protein sequence (*9*). In eukaryotes, the genetic information

is stored as DNA, transcribed into RNA and then translated into proteins (**Figure 1-1A**). Initially, the central dogma was postulated as an unidirectional process, however the discovery of reverse transcription (*10*, *11*) used by many viruses has since amended the central dogma to include a reversed information flow from RNA to DNA.

The genetic code is an essential part of the central dogma since it determines how the sequence in the messenger RNAs (mRNAs) is translated into an amino acid sequence at the ribosome. The ribosome reads mRNAs in sets of triplet codons [consisting of three nucleotides, (*12*)], each corresponding to one amino acid. The ribosome utilizes aminoacylated transfer RNAs (tRNAs) to decode these triplet codons by base pairing of the anticodon loop of the tRNA with the respective codon in the mRNA. Thus, the ribosome itself does not recognize the codon in the mRNA: The actual readers of the genetic code are the aminoacyl-tRNA synthetases [RS, (*13*)]. These enzymes bind their cognate tRNA and aminoacylate it with their respective canonical amino acid and this tRNA is then blindly used by the ribosome (**Figure 1-1B**).

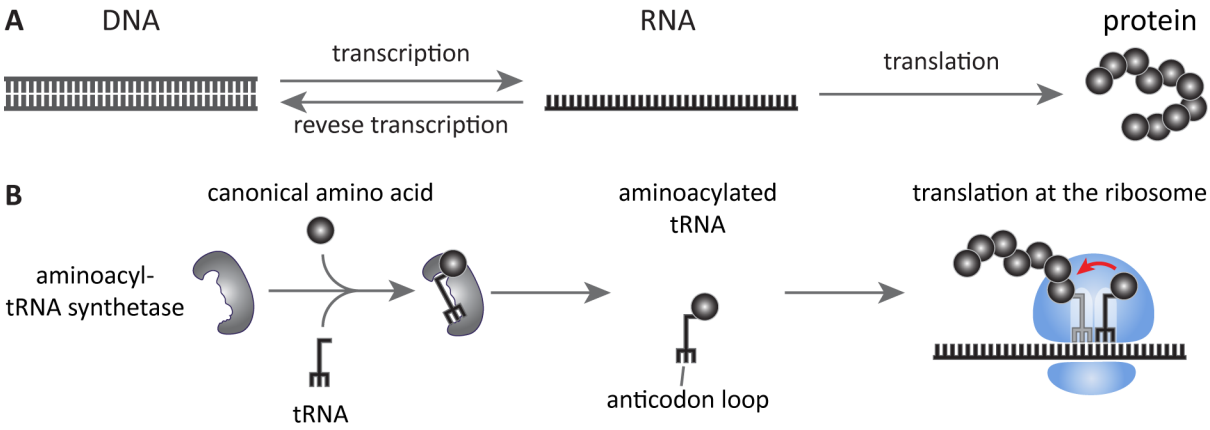


Figure 1-1. The central dogma of molecular biology. (**A**) The central dogma of molecular biology describes how information stored as DNA is transcribed into RNA and subsequently translated at the ribosome. Information can also partially flow in a reverse fashion from RNA to DNA. (**B**) Proteins are translated at the ribosome. To achieve this, aminoacyl-tRNA synthetases have to attach their respective canonical amino acids to their cognate tRNA. This tRNA has an anticodon loop that is complementary to a triplet codon in the mRNA and thus facilitates specific amino acid incorporation, decoding the genetic code.

What can be done with the genetic code? The genetic code is conserved across all domains of life and gives rise to the diversity of all living systems. With the four natural nucleobases it forms a total of 64 triplet codons, but it only encodes 20 canonical amino acids as well as three stop codons and is therefore highly redundant. The 20 canonical amino acids can be classified based on four chemical characteristics: nonpolar, polar, basic and acidic (**Figure 1-2**). This may seem to be a limitation. Even if the rare additions to the genetic code such as pyrrolysine (*14*) or selenocysteine (*15*) are taken into account, proteins can only access a tiny fraction of the chemical space (*16*). It is puzzling that the genetic code evolved to merely contain these four chemical characteristics. The abundance of protein co-factors (*17*, *18*) and

posttranslational modifications (*19*) however demonstrates that for some tasks this set is insufficient. It can thus only be imagined what might be possible if proteins could be composed of many more chemical functional groups and how life would look like with such a massively expanded genetic code.

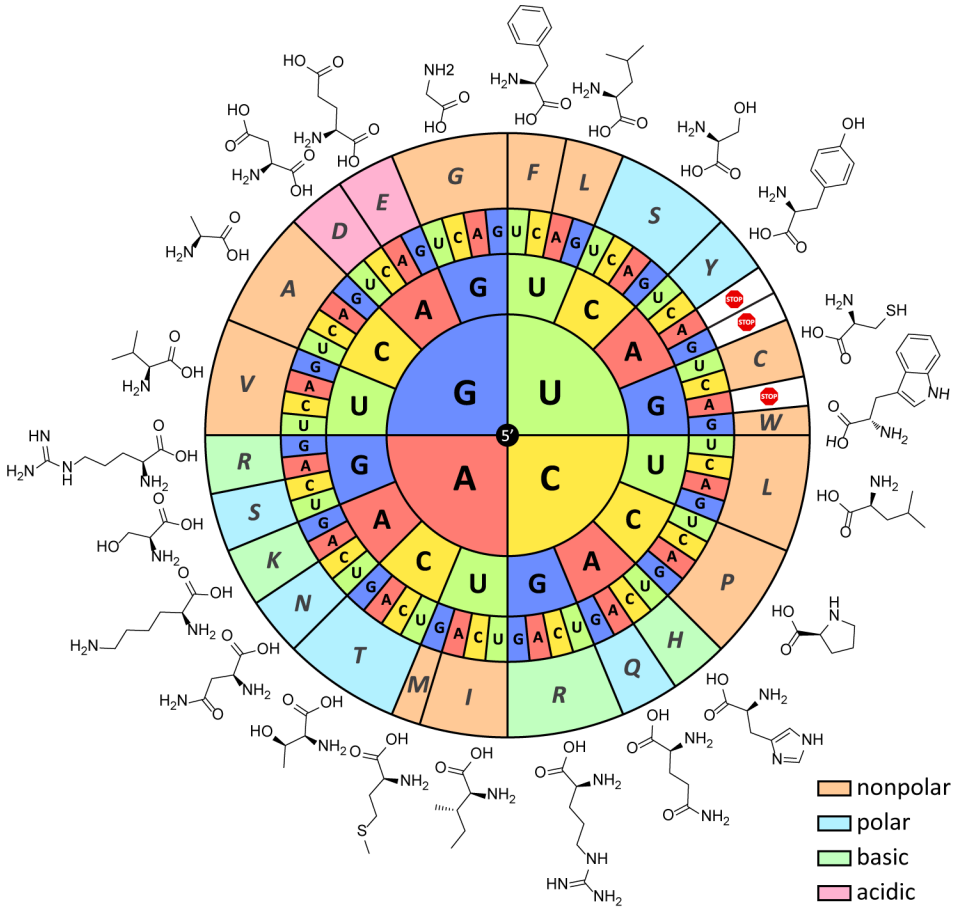


Figure 1-2. The genetic code is redundant and encodes a limited set of chemical functionalities. The genetic code consists of triplet codons made up of four natural nucleobases (adenine (A), cytosine (C), guanine (G) and uracil (U, in RNA)/thymine (T, in DNA). Thus, a total of 64 codons exists (inner three rings). These encode three stop codons and the 20 canonical amino acid (shown in the 1-letter code in the fourth ring, with their respective chemical structures adjacent to them), which can be classified into four chemical categories: nonpolar, polar, basic and acidic.

1.2 GCE—a powerful method for residue-specific protein modification

One approach to overcome the limitations of the accessible chemical space of proteins is GCE, which allows to co-translationally insert noncanonical amino acids (ncAAs) into proteins. Initial GCE approaches attached ncAAs to a suppressor tRNA biochemically *in vitro* (*20*, *21*). The anticodon of the tRNA is typically chosen to recognize a rare stop codon. Most commonly the amber stop codon (TAG) is chosen as it only terminates 7% of all *E. coli* proteins (*22*). The chosen stop codon is then inserted at a specific position into the mRNA of the POI and when this is then translated at the ribosome, this yields a site-specifically incorporated ncAA in the

protein. In this approach, the suppressor tRNA competes with the release factor (RF) machinery that would usually terminate translation.

Following these initial studies, the technology has been developed to enable the attachment of ncAAs to tRNAs *in vivo* (*23*). These applications rely on orthogonal RS/tRNA pairs. The orthogonal RS selectively attaches the ncAA to the suppressor tRNA, which then competes with the endogenous RF to facilitate site-specific ncAA incorporation in cells (**Figure 1-3**).

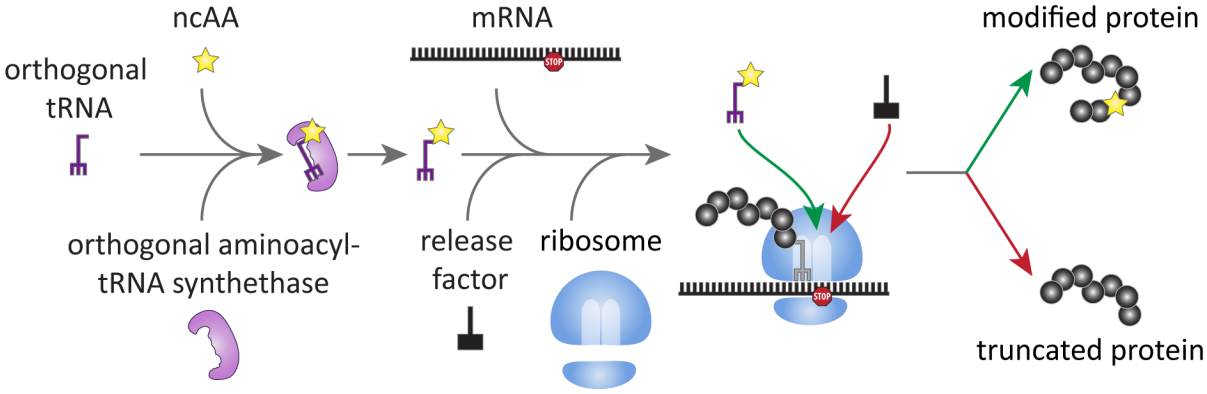


Figure 1-3. *In vivo* GCE mediated by orthogonal aminoacyl-tRNA synthetase/tRNA (RS/tRNA) pairs. The orthogonal RS binds its cognate tRNA that recognizes a stop codon, and aminoacylates it with the ncAA. When this stop codon, inserted into the mRNA, is encountered by the ribosome the suppressor tRNA can compete with the endogenous release factor. The release factor would lead to translation termination and thus yield a truncated protein (red arrow), while successful stop codon suppression would lead to ncAA incorporation and the synthesis of a modified full-length protein (green arrow).

Over the last decades, this technology has been successfully adapted for the incorporation of ncAAs into proteins in a variety of organisms including *Escherichia coli* [*E. coli* (*23*)], *Saccharomyces cerevisiae* (*24*), mammalian cells (*25–27*), and even multicellular organisms like *Caenorhabditis elegans* [*C. elegans* (*28*)], *Drosophila melanogaster* (*29*), *Danio rerio* (*30*) and *Mus musculus* (*31*).

In order for an RS/tRNA pair to be orthogonal it has to fulfill several requirements (*32*, *33*). First, the orthogonal RS should not recognize any canonical amino acid, and second it should not aminoacylate any endogenous tRNA. Third, the orthogonal tRNA should not be recognized by any endogenous tRNA-synthetase and fourth no endogenous tRNA-synthetase should recognize the ncAA (see **Figure 1-4**). Only if these conditions are fulfilled it is ascertained that the ncAA can be site-specifically incorporated into the POI.

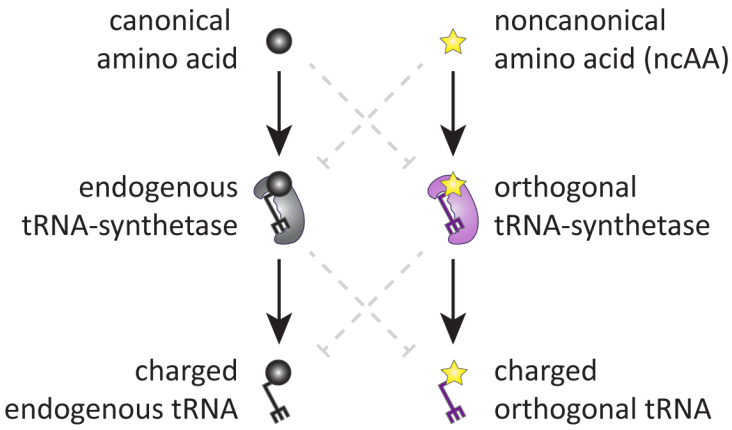


Figure 1-4. Orthogonal aminoacyl-tRNA-synthetase/tRNA suppressor pairs are essential for *in vivo* GCE. In order to be orthogonal an RS/tRNA pair has to fulfill several requirements. The orthogonal RS should not recognize any endogenous tRNAs and it should not aminoacylate it with any canonical amino acid. *Vice versa*, endogenous tRNA-synthetases should not bind and aminoacylate the orthogonal tRNA and should also not recognize the ncAA. (black arrows show direct interactions, grey dashed arrows indicate orthogonality).

While theoretically this seems to be a tangible problem, it has to be considered that every cell has at least 20 different endogenous RS/tRNA pairs and orthogonality needs to be achieved to considering all possible interactions. As a tRNA is a comparatively short molecule that interacts extensively with its cognate RS, engineering orthogonal RS/tRNA pairs *de novo* would be quite challenging. Therefore, in order to find a suitable orthogonal RS/tRNA pair for a particular host, typically evolutionary distinct RS/tRNA pairs are explored. For example several *E. coli* derived systems like the leucyl and tyrosyl RS/tRNA pairs are orthogonal in eukaryotes (*34*, *35*), while the tyrosyl RS/tRNA pair from the *archaeal Methanococcus janashii* is orthogonal in *E. coli* (*23*).

When an orthogonal RS/tRNA has been identified, it will usually at the beginning only permit to genetically incorporate its cognate canonical amino acid. In order to incorporate an ncAA of choice, the substrate specificity of the RS typically needs to be adapted in a second step. One powerful method to achieve this is directed evolution (*36*), which is particularly fast and convenient when *E. coli* can be used as a host for protein evolution (*36–38*).

Thus, a drawback of using *E. coli*-based synthetases for GCE in mammalian cells has long been the requirement to perform evolution in yeast cells (*35*), which is substantially more laborious and slower than performing evolution in *E. coli*. This problem was solved recently by developing an orthogonalized platform, in which an *E. coli* system can be used to evolve the endogenous *E. coli* synthetases for ncAA incorporation (*39*). The key idea here was that any endogenous *E. coli* RS/tRNA pair can be replaced with the corresponding yeast derived pair and subsequently the *E. coli* pair can be converted into an amber suppressor. This than allows to evolve this system for ncAA incorporation and as it is a *E. coli*-derived system it will most likely be orthogonal in eukaryotes. After its initial description this platform was further used to evolve multiple new systems for use in mammalian cells (*40–42*).

However, still this technology has one major drawback. Most endogenous RS/tRNA pairs have a very narrow substrate specificity, making evolution necessary for each new ncAA of interest, as well as an editing functionality, which can remove the attached ncAA (*43*). Additionally, most RSs also bind to the anticodon loop of the tRNA (*44*), hence they need to be further evolved in order to enable the reassignment of a codon of choice.

Due to these additional complications one orthogonal RS/tRNA pair stands out from all others. The pyrrolysyl tRNA-synthetase (PylRS) derived from methanogenic archaea has no editing functionality (*45*, *46*) and does not recognize the anticodon of its cognate $tRNA^{Pyl}$ (*47*). Moreover, it has inherently a broad substrate specificity and thus even the wild-type enzyme accepts several lysine derivatives (*48*, *49*). This particular enzyme type has been developed to encode more than 100 distinct amino acids with backbones as diverse as phenylalanine, lysine and histidine in response to various triplet as well as quadruplet codons (*46*, *50–52*). More recently even multiple mutually orthogonal PylRS variants were reported that allow to harness this powerful enzyme type to multiple times selectively incorporate distinct ncAAs (*53–57*). I further discuss the promises of these new technological developments for advanced polymer synthesis *in vivo* in **Chapter 2**, [see also **Appendix III** and (*32*)].

The PylRS based systems have been developed to accept various ncAAs with bioorthogonal reactive groups in their side chain (*58–63*). These can the subsequently be used after incorporation to attach small, photostable fluorescent dyes to the POI, which has for example been used to perform super-resolution microscopy (*64–66*).

1.2.1 Limitations of GCE—the necessity of orthogonal translation

With the growing availability of useful orthogonal RS/tRNA pairs additional limitations of GCE for *in vivo* studies become more relevant. In order to achieve site-specificity the anticodon of the tRNA is commonly chosen to decode a stop codon. While for the POI it can be ensured that it does not terminate on the same stop codon, this is not the case for endogenous mRNAs and hence the ncAA can be erroneously incorporated during the translation of several endogenous mRNAs. This cannot only lead to C-terminally extended proteins that might be toxic to the host, but it can also result in background in labeling experiments, which is problematic for imaging-based applications. This problem is aggravated in higher eukaryotes, as even their rarest stop codon is much more abundant. For example, while the amber codon (TAG) only terminates 7% of all proteins in *E. coli* (*22*), it appears at the end of more than 20% of all human proteins (*67*). Therefore, even suppression of the amber stop codon can lead to

substantial background incorporation as ribosomes cannot distinguish the mRNA of the POI from endogenous mRNAs (**Figure 1-5**).

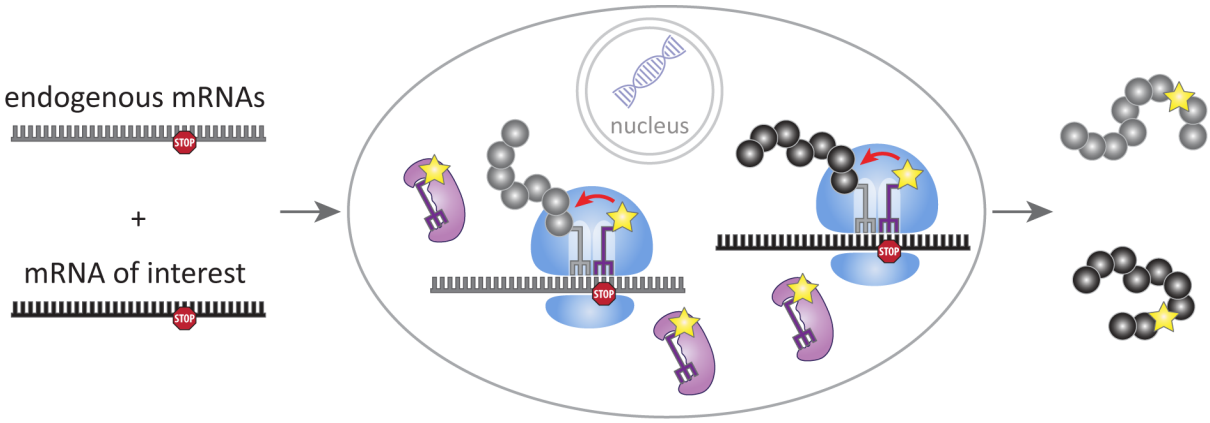


Figure 1-5. GCE can lead to unwanted ncAA incorporation into endogenous proteins. Besides the stop codon engineered into the mRNA of the POI (black) several endogenous mRNAs (grey) will naturally terminate on the same stop codon. As ribosomes are unable to distinguish them this leads to the incorporation of the ncAA into both endogenous proteins and the POI.

1.2.1.1 Orthogonal translation in E. coli

In *E. coli* the background incorporation problem has been addressed by three alternative approaches, which I will discuss in more detail in this section. Artificial base pairs have been added to the genetic code, orthogonal ribosomes have been developed that exclusively translate engineered mRNAs and entire *E. coli* genomes devoid of certain codons have been synthesized.

Artificial base pairs—adding new letters to the genetic alphabet

Canonically, the genetic code is written in four letters (adenine, thymine/uracil, guanine and cytosine) giving rise to 64 different triplet codons. By adding an additional letter pair to the genetic code, the number of possible triplet codons would increase to 216. If even two additional pairs would be added this number would increase to 512. These numbers would by far exceed the currently available pairs of orthogonal RS/tRNA pairs and would offer extensive capabilities for artificial polypeptide synthesis. Two major alternative approaches have been pursued by different laboratories to incorporate unnatural base pairs (UBPs) into nucleotide polymers.

One approach utilizes the same mechanisms that drive canonical base pairing between A:T or C:G. Native DNA is predominantly formed by size matching and hydrogen bonding (*68*). One small pyrimidine base always pairs with a larger purine. Additionally, A:T form two specific hydrogen bonds while C:G exhibit three. The hydrogen bond pattern here offers an opportunity for adding UBPs to nucleotide polymers that do not cross-react with the endogenous base pairs. In particular with three hydrogen bonds between a particular base pair the arrangement of donors and acceptors can in principle be designed such that specific and stable base pairing can

be achieved. However, while this theoretically seems straightforward in practice this can be rather complicated, as nucleotides are commonly subjected to tautomerization as well as protonation and deprotonation, which can affect base pairing and lead to loss of the UBP, hence often extensive optimization is required to ensure replicative fidelity (*69*, *70*). Pioneered by the laboratory of Steve A. Benner, this approach has been used to create six (*71*) and eight (*72*) letter DNA molecules that do not alter the overall structure of the DNA-duplex. For this approach it was shown that the artificial base pairs can be transcribed enzymatically (*73*) and even be used for ribosomal peptide synthesis *in vitro* (*74*). However, it was not yet shown that this approach can be used *in vivo* to store information in the form of DNA, transcribe it into RNA and translate this RNA into a protein.

The second approach pioneered by the Romesberg and Hirao laboratories is completely independent of hydrogen bonding and relies on hydrophobic interactions of synthetic nucleotides (*75*, *76*). This work was sparked by the observation that nucleotides devoid of Watson-Crick base pairing can be incorporated into DNA using DNA polymerases made by the Kool laboratory (*77*, *78*). Also this work required extensive optimization and was pursued using a medicinal chemistry like approach to develop a structure activity relationship and obtain an ideal UBP for replicative retention (*79–83*). Additionally, it has been shown that this kind of UBP can efficiently be transcribed into an RNA form (*84*, *85*). More importantly, this work already progressed to the stage of *in vivo* storage and retrieval of genetic information that allows to create semi-synthetic organisms (SSOs) with an expanded genetic alphabet. It was shown that UBPs can be specifically retained in plasmids in *E. coli* (*86*) that it can be transcribed into RNA and used for coding protein translation (*87*). Finally this can be combined with established GCE procedures to use it for selective ncAA incorporation at multiple sites [see **Figure 1-6A**, (*88*)]. Very recently researchers have also made progress to transfer this technology from *E. coli* hosts to mammalian cells (*89*). However, this will require further optimization to increase efficiency as the UBPs behave differently in mammalian cells.

Up to now, none of the UBPs have been shown to be functionally fully equivalent to the endogenous base pairs. Particularly for predominantly hydrophobic UBPs the sequence context is highly relevant, and it was shown that this can influence its retention in DNA (*90*). Moreover, the sequence context is highly relevant in the codon chosen to encode an amino acid with the UBP and determines how well and with what fidelity it can be used at the ribosome to synthesize proteins (*88*). Hence, so far only triplet codons containing one UBP have been used for ribosomal peptide synthesis in living SSOs. This substantially limits the number of new

codons that can be created, and it will be interesting to see if this technology can be developed to use an UBP equivalent to the natural base pairs.

Orthogonal ribosomes—RNA-specific translation

An alternative approach to achieve orthogonal translation would be to exclusively modify the genetic code for a specific mRNA. For example, to change the genetic code to utilize quadruplet instead of triplet codons. This is particularly interesting, as a triplet-based genetic code with the natural nucleotides only consists of 64 different codons, while a quadruplet codon based genetic code would offer 256 different combinations. Almost 20 years ago it was shown that tRNAs can be evolved to promote quadruplet codon decoding in mRNAs at a designated site resulting in a frameshift in an otherwise triplet codon based mRNA sequence (*91–93*).

However, the *E. coli* ribosome is quite inefficient in decoding quadruplet codons and hence it would be beneficial to evolve the ribosome (*94*). Unfortunately, evolving the ribosome in a living cell is quite challenging, as the ribosome is required for translation of the proteome and many mutated ribosomes might be insufficient for this task and hence lethal for the cell. Therefore, it is necessary to introduce an orthogonal ribosome that only translates chosen mRNAs.

In *E. coli* the ribosome selects its mRNA substrates predominantly through base pairing interactions between the Shine-Dalgarno (SD) sequence in the mRNA and the corresponding anti-Shine-Dalgarno (ASD) sequence in the 16S ribosomal RNA [rRNA, (*95*)]. Modifying these sequences can lead to changed mRNA-ribosome specificities (*96*). It has been shown that by evolving both the SD and ASD sequence it is possible to obtain mRNAs that are not at all translated by native ribosomes but still by corresponding orthogonal ribosomes with a modified ASD sequence (*97*). Meanwhile, the orthogonal ribosomes do not substantially translate endogenous mRNAs with native SD sequences. As these orthogonal ribosomes are now not required anymore for the synthesis of the proteome, they can be subjected to evolution to endow them with new functionalities (see **Figure 1-6B**). Initially, this has been applied to evolve ribosomes that more efficiently suppress amber codons by interfering with the interaction with the competing release factor 1 [RF1, (*98*)]. Subsequently, ribosomes were further developed to more efficiently decode quadruplet codons (*99*).

In these modified ribosomes, only the small ribosomal subunit can be evolved, as it contains the 16S rRNA, which was engineered for alternative mRNA selection. The large ribosomal subunit is shared with the native ribosomes and can hence not be modified.

In order to make the large ribosomal subunit accessible for evolution it would be necessary to ensure that the orthogonal small ribosomal subunit is specifically combined with the large ribosomal subunit. This was achieved by stapling the rRNAs of the large and small subunit together effectively producing one joined rRNA construct that allowed initial evolution experiments on the full ribosome (*100*, *101*). This technology was further optimized by modifying the linker sequence joining the 16S rRNA with the 23S rRNA to ensure that the small and large ribosomal subunits of the orthogonal stapled ribosomes always form a specific complex, while they do not cross-interact with the endogenous small or large ribosomal subunits (*102*).

These evolvable orthogonal ribosomes now allow to modify the translation process exclusively for selected mRNAs, which represents a major step towards the synthesis of fully artificial polypeptides in *E. coli*. However, this technology cannot be easily transferred to eukaryotic systems, as eukaryotic ribosomes recognize their mRNAs by different means.

In contrast to the SD-ASD sequence base pairing interaction, that directly recruits the small ribosomal subunit to mRNAs in *E. coli*, mammalian mRNAs are most commonly identified by their m^7G cap structure (*103*). To this end, usually multiple translation initiation factors bind the mRNA and subsequently recruit the eukaryotic small ribosomal subunit (*104*). As in this case no specific base pairing interaction exists a similar approach to engineer SD and ASD sequences is not possible. Therefore, it would be substantially more difficult to engineer orthogonal eukaryotic mRNA recognition.

Besides the canonical eukaryotic cap-based ribosome recruitment pathway also alternative translation initiation mechanisms have been described. For example, many viruses exhibit internal ribosomal entry sites (IRESs) that can recruit ribosomes also to uncapped mRNAs (*105*). IRESs utilize different mechanisms but usually the ribosome recruitment is mediated by a specific structural elements that promote interactions between the RNA and the ribosome and not by a simple and easy to engineer base pairing interaction (*105*). Interestingly, however there is growing evidence that for certain mRNAs specific base pairing interactions with 18S rRNA can exist and that this might even be involved in SD sequence-like translation initiation (*106–113*), but this is not as robustly established as the SD-ASD sequence mediated prokaryotic translation initiation mechanism (*105*).

With the growing understanding of eukaryotic translation initiation, the hope is that at some point it might be possible to develop orthogonal eukaryotic ribosomes by adopting a strategy similar to the one chosen in prokaryotes. However, the more complex and multi-facetted

translation initiation mechanisms present in eukaryotic cells make this substantially more challenging. An orthogonal eukaryotic ribosome would not only need to somehow specifically bind engineered mRNAs, but it would also be necessary to ensure that it does not participate in cap-based translation initiation. As this process is mediated by a plethora of translation initiation factors and their respective interaction with the ribosome it will be a formidable challenge to engineer this in an orthogonal fashion, which promises to be substantially more complicated than the approaches utilized to engineer ribosomes in *E. coli*.

Synthetic genomes—removing selected codons to freely reassign them

The third approach exploits the redundancy of the genetic code. As 64 native triplet codons encode the 20 canonical amino acids and 3 stop codons, multiple codons are redundant. If these were removed entirely from the genome it would be possible to reassign the freed-up codon to any desired noncanonical functionality. This approach heavily relies on genome engineering capabilities, as it is required to synthesize whole genomes.

Initial approaches were driven by a combination of multiplexed automated genome engineering (*114*) and conjugative assembly genome engineering (*115*) pioneered by the Church laboratory. These initial works focused on replacing the amber stop codon which occurs 321 times in the entire genome of *E. coli* (*115*, *116*) or a rare arginine codon in all essential genes, in which it only occurs at 123 instance (*117*). The replacement of the amber codon throughout the genome then allowed to delete the prokaryotic RF1, which is responsible for terminating translation in response to amber and ochre (TAA) codons. The remaining RF2 only recognizes ochre and opal (TGA) codons and would thus not compete with an amber suppressor-tRNA mediated ncAA incorporation (*118*). While successful these initial approaches had marked effects of the fitness of the recoded *E. coli*. This was most likely caused by off-target mutations and hence directed evolution was utilized to improve these recoded organisms (*119*).

These early efforts proofed that genome engineering is a powerful means to tailor the genetic code of a living organism. Still these approaches only created one free codon and for artificial polymer synthesis multiple free codons would be required. To this end, it would be particularly interesting to target redundant sense codon, but this is challenging for several reasons.

First, most sense codons do not only occur a few hundred times in the *E. coli* genome, but more commonly a few thousand times, which in turn demands much more efficient genome replacement technologies. Therefore, it would be beneficial if larger pieces of DNA could be replaced at a time. To replace DNA in *E. coli* often lambda red mediated homologous recombination is utilized (*120*), that allows to integrate linear double stranded DNA (dsDNA)

into the genome in a programmed fashion. This technology is very efficient for short sequences but has problems to efficiently integrate longer DNA stretches likely due to inefficient dsDNA delivery (*121*). To overcome this limitation Wang and coworkers developed replicon excision for enhanced genome engineering through programmed recombination [REXER, (*121*)]. In this method the recombineering step is separated from the bacterial transformation. In brief, the synthetic DNA designed to replace a complementary genomic fragment is introduced into the cell as a bacterial artificial chromosome (BAC). The synthetic fragment is framed by homologous recombination sites which are in turn flanked by CRISPR/Cas9 target sites. The cells express Cas9, lambda red and tracer RNA, which is one part of the RNA required to mediate Cas9 based DNA cutting (*122*, *123*). Upon addition of the spacer RNA, Cas9 linearizes the BAC, effectively excising the synthetic DNA in a linearized dsDNA form. This dsDNA is than used by lambda red for efficient recombineering into the hosts genome. As in this method introduction of dsDNA does not limit lambda red recombineering it is essentially size independent and has been used to effectively replace 100 kb fragments in the genome (*121*), thereby providing an improved method for genome editing.

Second, while synonymous codons of one amino acid are redundant considering the resulting protein, they are not necessarily functionally identical. Codon choice is involved in the regulation of translation (*124*), it can influence mRNA folding (*125*), protein synthesis speed (*126*, *127*), co-translational protein folding (*128*, *129*) and other processes. Hence, not all synonymous codon replacements are permissive and due to the limited *a priori* knowledge, it has to be determined empirically which synonymous codon replacement schemes are accepted by the cell. It would be extremely cost and labor intensive to perform this test directly on the whole genome, therefore it would be beneficial to test recoding schemes on a short, but meaningful model region. Wang and coworkers identified a 20 kb stretch in the *E. coli*, which contains several essential gene and plenty of their targeted codons (*121*). Using this test region, they analyzed several recoding schemes for the alanine, serine and leucine codons. Interestingly, while some are permissible others fail completely.

Having established a method that allows to faithfully replace 100 kb parts of the genome and identify likely working recoding schemes they then developed a method that allows to iteratively replace the genome, called genome stepwise interchange synthesis [GENESIS, (*121*)]. In this method 2 pairs of positive and negative selection markers are used which are iteratively replaced. First, one version of positive marker and negative marker in introduced into the cell. In the first step of genome replacement these are removed, and the second pair is

inserted. Cells are then screened for gain of the second positive marker and loss of the first negative one, which ensures targeted genome replacement (*121*). In the next step the second pair is again replaced to first one and cells are screened for presence of the first positive selection marker and absence of the second negative one.

With the GENESIS method that allows to iteratively replace 100 kb of DNA at a time it would take 40 steps to replace the entire 4 Mb genome of *E. coli*. Fredens and co-workers achieved this and fully synthesized an *E. coli* genome devoid of all amber codons and in which all TCG and TCA serine codons were exchanged to synonymous ones [see **Figure 1-6C**, (*130*)]. This then allowed to subsequently delete the corresponding tRNA genes as well as the gene encoding RF1 (*130*).

These massive genome recoding projects offer an interesting solution to create new codons for *E. coli*. Contrasting to this, eukaryotes have a much more complex and much larger genome. However, a consortium of synthetic biologist has taken up the challenge of synthesizing an entire *S. cerevisiae* genome, which will be substantially modified and is also planned to be devoid of amber codons (*131*). The size of the yeast genome is above 11 Mb and it will be very interesting to see what such a synthetic genome will be capable of (*132*). The human genome is above 3000 Mb big and synthesizing it *de novo* is still out of reach, even when the newest technological developments are taken into account. Hence, genome synthesis is not going to be a timely solution to create free codons for GCE in mammalian cells.

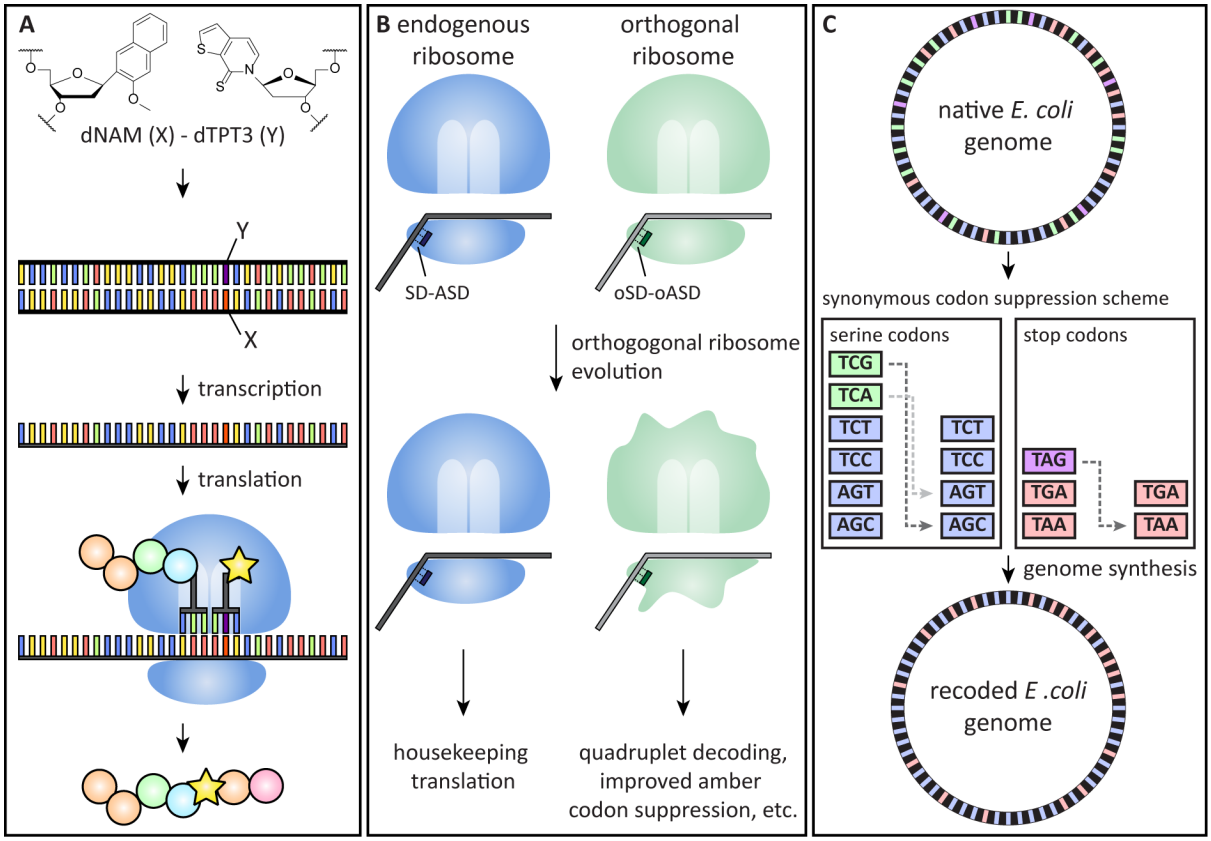


Figure 1-6. Orthogonal translation strategies in *E. coli*. (**A**) Adding unnatural base pairs (UBPs) to the genetic code is one way to achieve fully orthogonal translation. For example the UBPs dNAM and dTPT3 can be inserted into DNA, transcribed into mRNA and tRNA and subsequently be used by ribosomes for protein synthesis (*87*). (**B**) *E. coli* ribosomes recognize their mRNA via base pairing interactions between the Shine-Dalgarno (SD) sequence in the mRNA and the anti-Shine-Dalgarno (ASD) sequence in the 16S rRNA. This can be made orthogonal by changing the ASD to an orthogonal ASD (oASD), thus creating orthogonal ribosomes that only translate mRNAs with a corresponding orthogonal SD (oSD). These orthogonal ribosomes can subsequently be evolved for advanced capabilities like enhanced stop codon suppression (*98*) or optimized quadruplet codon decoding (*99*). (**C**) The genetic code is redundant, using multiple codons to encode the same ncAA. Thus, by changing all occurrences of a particular codon in the genome to a synonymous one it is subsequently possible to freely reassign it to an ncAA. This was for example done for the serine codons TCG and TCA as well as for the amber stop codon (*121*, *130*).

1.2.1.2 Hypothesis—establishing orthogonal translation in mammalian cells

GCE is a powerful method enabling site-specific fluorescent studies of proteins in mammalian cells to understand their functions under native conditions. Particularly for this application background incorporation into untargeted proteins is a major concern. Therefore, an orthogonal translation platform that allows to exclusively incorporate an ncAA into the POI would be crucial to allow this methodology to become widely utilized. Unfortunately, the three approaches that establish orthogonal translation in *E. coli* cannot be easily transferred to mammalian systems as described above.

What alternative approach could be used to selectively expand the genetic code for only specific mRNAs?

One feature that sets eukaryotic cells apart from bacterial cells is the presence of specialized organelles like nuclei, mitochondria or the endoplasmic reticulum. A key feature of these intracellular organelles is that they spatially separate specific functions, which allows cells to avoid interference between incompatible processes. I hypothesized that it might be possible to analogously create a synthetic organelle dedicated to protein synthesis. If this organelle would be orthogonal to the host's translation machinery it should be possible to enrich this organelle with a GCE system to synthesize proteins with advanced capabilities.

The first design principle that one could envision for such an organelle would be to encapsulate the translation process with a membrane (**Figure 1-7**). This however would be very challenging, as translation is a very complex process that requires hundreds of factors to work together, including the comparatively large ribosomes. All of these factors would have to be imported across the membrane in addition to the mRNA of the POI, the GCE system and the ncAA. Additionally, the modified POI would need to be exported after translation. This would most likely require a dedicated transport machinery similar to the nuclear pore complex (NPC) that regulates transport between the nucleus and the cytoplasm. The NPC itself is a gigantic macromolecular machine—in fact the largest nonpolymeric structure present in eukaryotic cells (*133–135*)—and it would be a challenging on its own to engineer it into a synthetic organelle to facilitate the required transport.

Therefore, a membrane enclosed approach seemed to be difficult. Intriguingly, there are many organelles in the cell that do not possess a traditional membrane boundary (*136*). Thus, I hypothesized that it might be possible to analogously create a protein engineering organelle in an open format.

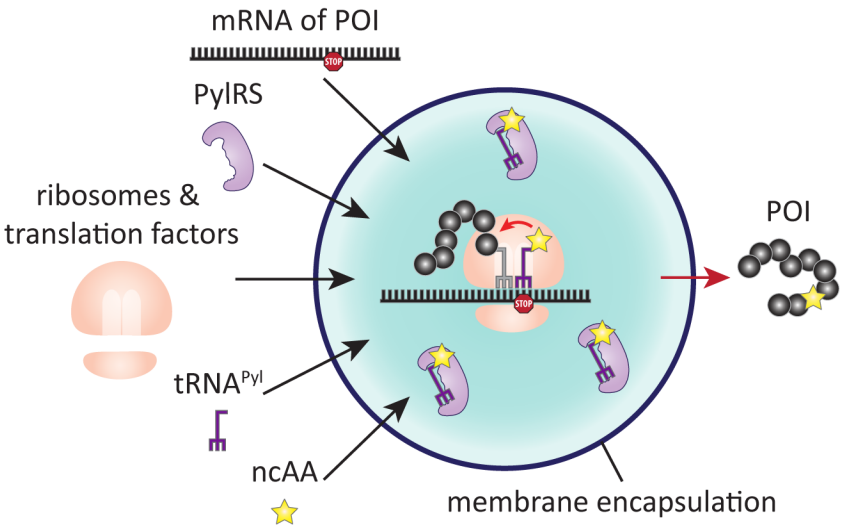


Figure 1-7. Creating a synthetic organelle for protein synthesis. One potential way to achieve mRNA specific orthogonal translation in a cell might be to create a dedicated, spatially separated organelle. If this would be transformed in a membrane enclosed compartment it would be necessary to import the GCE system, the ncAA, the mRNA of the POI as well as ribosomes and other translation factors (black arrows) and to subsequently export the modified POI (red arrow).

1.3 Biomolecular condensates—dynamic structures for complex tasks

What are these organelles without a membrane boundary and how are they formed?

Although membrane-enclosed organelles are the most prominent compartments in a cell and have for a long time been the focus of intense research, their membraneless relative has been known for quite some time.

The first membraneless organelle—later termed the nucleolus—has been described almost 200 years ago in the 1830s (*137*). The Cajal body, another membraneless organelle in the nucleus, has been described more than 100 years ago (*138*). Since then many more membraneless organelles have been described, including germ granules (*139*), stress granules (*140*), paraspeckles (*141*), Balbiani bodies (*142*) and others. These organelles are often composed of both nucleic acids and proteins, but they differ from traditional macromolecular complexes. Classical macromolecular complexes like ribosomes, can also consist of multiple proteins and RNAs, but they assume a fixed structure of a certain size with a defined stoichiometry. Contrary to this, membraneless organelles do not have a fixed structure or stoichiometry, they can grow and shrink and are often inherently dynamic (*143*).

The main feature of these organelles is that they are capable of concentrating molecules and that they are composed of various biological macromolecules, thus Banani and co-workers coined the term "biomolecular condensates" to describe all the different flavors of organelles that lack a traditional membrane boundary (*136*). Biomolecular condensates can exhibit characteristic liquid-like behaviors, indicating that they assume a highly dynamic liquid-like state in the aqueous environment of a cell (*144*). It has been shown that they can deform in response to shear flow, drip, fuse and relax to spherical shapes as well as rapidly rearrange their contents as has been shown by fluorescence recovery after photobleaching experiments (*144*). Additionally, long-term they can coalesce and undergo Ostwald ripening (*145*). These biomolecular condensates can not only assume liquid-like states, but they can for example also solidify and form gels or fibers (*146*).

1.3.1 Liquid-liquid phase separation

For a long time, it has been enigmatic how these condensates form and despite their lack of a defined structure can perform specific tasks. However, in the last 10 years (*144*) it was unraveled that biomolecular condensates are often formed by a process called liquid-liquid phase separation (LLPS, see **Figure 1-8** for a schematic overview). LLPS describes the process how molecules can transition from a soluble or mixed state into a phase separated state via liquid-liquid demixing. When molecules undergo a phase separation process, they separate into

a dense and a dilute phase. Within the dense phase, the concentration of the phase separated molecule can be orders of magnitude higher than in the remaining dilute phase. In case of a liquid material state, molecules from the dense phase can rapidly and dynamically exchange with the surrounding dilute phase, crossing the phase boundary. A phase transition occurs when the respective molecule exceeds it soluble concentration and it is thus energetically favorable when the molecules separate into a dense and dilute phase.

Taking the thermodynamics of this process into account, it is crucial that the molecule interacts with itself and that is energetically favorable over interacting with the solvent, resulting in a negative enthalpy upon phase separation (*147*, *148*). Additionally, this interaction needs to be energetically sufficiently favorable to compensate for the reduction in entropy upon phase separation, so that in total the free energy is minimized.

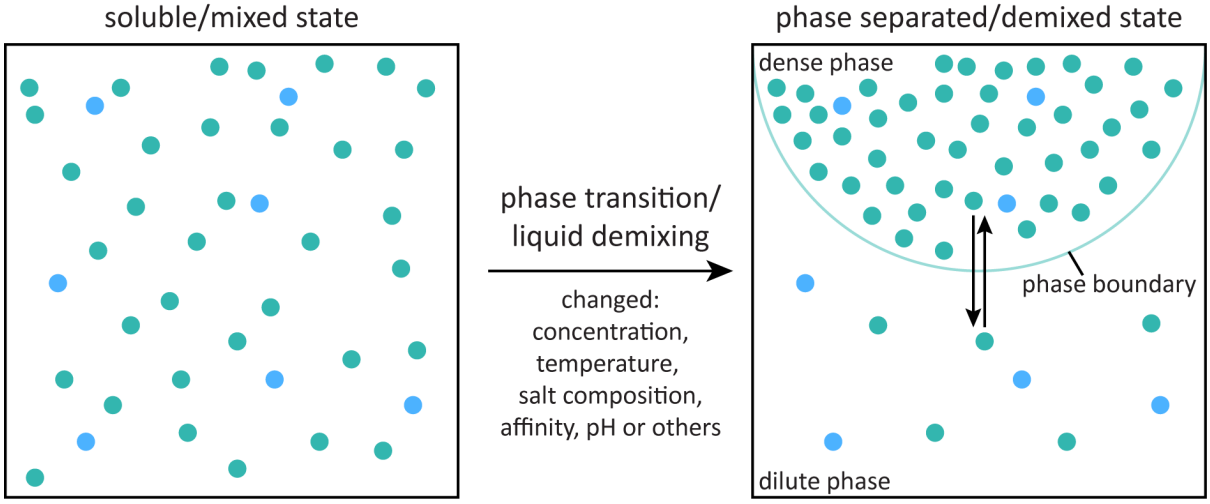


Figure 1-8. Liquid-liquid phase separation. Certain molecules (green dots) can undergo a phase transition when they exceed a concentration threshold. They thus separate into a dense and dilute phase. As long as the material state remains liquid and dynamic, they can rapidly exchange between the dense and dilute phase, crossing the phase boundary (see black arrows). Noninteracting molecules (blue dots) do not change their concentration and are therefore distributed evenly between both phases.

1.3.2 Phase separation in biological systems

In the crowded environment of a cell, biomolecular condensates are typically composed of multiple components, including tens to hundreds of different proteins (*149*, *150*). These components are not all equal, as it has been shown that some are necessary for phase separation (*151–153*), while others are dispensable and their partition into condensates can vary depending on the state of the cell (*154*, *155*). To describe these components Banani et al. introduced the concept of scaffolds and clients (*156*). Scaffolds are the core part of the condensate and typically exhibit a high valency. Clients can use the same types of interactions utilized by the scaffold but have a lower valency. Therefore, clients associate with the

condensate much more dynamically and can be recruited transiently. Importantly, clients can only partition into the condensate if free binding sites are available.

1.3.2.1 Multivalency in phase separating systems

Banani et al. developed a simple model system that is based on the specific interaction of the small ubiquitin-like modifier (SUMO) and the SUMO interacting motif (SIM), which form a specific complex (*156*). By using oligomers of SUMO and SIM they created phase separated droplets *in vitro* and *in vivo*. These oligomers represented the phase separation scaffold and they could control if either SUMO or SIM was present in access and thus create droplets that have respectively free SUMO or SIM interaction sites. To study client recruitment they used monovalent SIM or SUMO variants, tagged with fluorescent proteins and could observe that a SIM-client is only recruited to a droplet that has a SUMO excess, while a SUMO-client is only recruited to a droplet with a SIM excess (*156*). Importantly, they could confirm this behavior *in vivo*, indicating that availability of binding sites can govern partition into biomolecular condensates. This model system is purely based on simple bimolecular interactions but demonstrates that multivalency can be sufficient to drive LLPS.

A common trend for biological phase separating systems however is that the involved biomolecules are inherently multivalent and use multiple modes of interaction. It has been shown that essentially all types of molecular interactions can create phase separating systems. These include electrostatic, dipole-dipole, hydrophobic, cation-π and π-π interactions (*157*, *158*). Due to the nature of these different interactions and the competition between enthalpic and entropic contributions the soluble concentration of a biomolecule is influenced by many parameters including not only its concentration *per se*, but also temperature, salt composition, affinity, pH and other factors (*136*, *143*, *146*).

1.3.2.2 Intrinsically disordered proteins in phase separation

Intrinsically disordered proteins (IDPs) often play a prominent role in phase separating systems (*159*, *160*). A hallmark of IDPs is that they lack a predetermined three dimensional structure and thus typically exist in a vast conformational assembly (*161–164*). Many proteins that can be classified as IDPs only have a region within their sequence that is intrinsically disordered (IDR, intrinsically disordered region). Some of these sequences are prone to form ß-amyloids found in prions and are therefore called prion-like domains [PLDs, (*165*)]. In eukaryotes it is estimated that around 30% of the entire proteome contain IDRs (*166*).

IDPs are commonly composed of a limited variety of amino acids and are particularly often enriched in flexible, polar and charged amino acids, including glycine, serine, glutamine,

glutamate, proline, lysine and arginine (*167–169*). Additionally, they often lack hydrophobic amino acids. IDPs that participate in phase separation processes however often contain tyrosine or phenylalanine residues to enable π-π or cation-π interactions (*158*).

For the interaction of IDPs with their binding partners at least two distinct mechanism are known. The two most prominent ones are coupled folding-binding or fuzzy interactions (*163*). In the first variant, the IDP only interacts with a binding partner in a defined configuration (*170*), in the second, the IDP maintains disorder throughout (*171*). For example, for the transport machinery between the nucleus and the cytoplasm both mechanisms have been described. For the FG-nucleoporin 214 a coupled folding-binding mechanism was observed for its interactions with the nuclear transport machinery (*172*)—FG-nucleoporins are IDPs enriched in phenylalanine and glycine repeats that fill the central channel of the nuclear pore complex (*173*). Many other FG-nucleoporins, like nucleoporin 153, interact with nuclear transport receptors via fuzzy interactions (*135*, *174*).

To interact with their binding partners IDPs often utilize short linear motifs (SLiMs). SLiMs are short linear amino acid sequences of about 2-8 amino acids and due to their relatively small size they typically participate in rather weak interactions (*165*, *175*, *176*)]. As many SLiMs can be concatenated in one IDP, interspersed by flexible linkers, they can easily provide the multivalency required for phase separation (*177*).

1.3.2.3 The sticker-spacer model of phase separation

One hypothesis explaining how multivalent protein and RNA interactions can facilitate phase separation is the sticker-spacer model proposed by Pappu and co-workers (*178*). The key idea in this model is that phase separating biomolecules can be described as associative polymers, which are polymers with attractive or interacting groups (*179*).

The stickers describe the attractive group and in case of biological systems can comprise any of the above described possible protein and RNA interactions. The spacers are in contrast inert linkers between the stickers, which should not interact. In the sticker-spacer model, the stickers ensure cross-linking between different polymers and can therefore drive phase separation. This model can not only be used to describe the phase separation process itself, but also how a liquid condensate can transition into a gel state by a process called percolation, which describes that at a certain concentration the non-covalent interaction network can span the entire structure (*177*, *178*). Recently, collaborative work by the Hyman, Alberti and Pappu laboratories has demonstrated this model for the fused in sarcoma (FUS) protein family (*180*).

1.3.2.4 The FUS protein family

The FUS protein family is named after the protein FUS, its earliest described member (*181*). All FUS family proteins share a similar domain structure with one IDR or PLD that is combined with a RNA binding domain (RBD), which contains at least one RNA binding motif (*182*). The FUS family contains the FET proteins [FUS, Ewing's sarcoma breakpoint region 1 (EWSR1), TATA binding protein-associated factor 2N (TAF15), (*183*)], heterogenous nuclear ribonucleoprotein [hnRNP, (*184*)] variants as well as TAR DNA-binding protein 43 [TDP-43, (*185*)] and others (*180*).

For the FET family, hnRNP and TDP43 it has been shown that they all can phase separate (*186–189*). For several of these proteins the PLDs can phase separate on their own *in vitro* (*186*, *187*, *190*, *191*). Therefore, the PLDs of these proteins have been considered to be particularly important for phase separation. But, the PLD of FUS on its own phase separates only at relatively high concentrations or in the presence of molecular crowders (*187*, *192*).

In contrast to this the respective full-length protein phase separates at a two orders of magnitude lower concentration, indicating an important role of the RBD in the phase separation process (*180*). Similar observations were made for EWSR1 (*180*). Strikingly, the FET proteins have RBDs that are particularly enriched in arginine residues, while their PLDs are tyrosine-rich (*180*). The network interactions between the arginine and tyrosine residues is crucial to determine the saturation concentration and is distinct from pure cation-π interactions, as a replacement of tyrosine with phenylalanine or arginine with lysine reduces phase separation propensity (*180*). In addition electrostatic interactions of negatively charged amino acids in the PLDs with the positively charged RBDs further promote phase separation (*180*). This is in good agreement with studies showing the relevance of both electrostatic and aromatic interactions for other phase separated systems (*193*, *194*).

Now taking the theoretical sticker and spacer framework into account, the tyrosine and arginine residues can be considered as the adhesive sticker moiety. In the FET protein family the spacers are formed by glycine, serine and glutamine motifs, which regulate the material properties of the condensate, for example glycine residues are important to retain a liquid-like state (*180*).

Similarly, for hnRNPA1 a sticker-spacer model was recently used to quantitatively describe its phase separation behavior (*195*). Interestingly, in this example it was shown that the spacing between stickers is of high importance to determine the material properties of the condensed phase. Evenly spaced stickers in the PLD (e.g. evenly spaced aromatic residues) promote a

liquid-like state, meanwhile clustered stickers lead to a pronounced amorphous aggregation phenotype (*195*).

In summary the sticker-spacer model can be used to explain and understand the phase transition process of FUS family proteins. The nature and abundance of stickers mostly determines the critical concentration for phase separation (*180*, *195*), which is in good agreement with the established role of multivalency to promote phase separation (*156*). The spacers in between the stickers do not affect the critical concentration, but their size and composition modulates the material properties of the resulting condensate (*180*, *195*). Therefore, this body of work opens up interesting future possibilities to understand and modify the phase separation behavior of FUS family protein based biomolecular condensates.

1.3.3 RNA and protein granules that regulate cellular biochemistry

Up to now we explored how phase separation with biological macromolecules occurs with a particular focus on multivalency, IDPs and the FUS protein family. Next, I am going to dive into the physiological relevance of phase separation processes and what kind of molecular processes can be regulated by it. To this end, I am going to discuss three specific examples of biomolecular condensates, namely stress granules, the *C. elegans* centrosome as well as receptor clustering on membranes.

1.3.3.1 Stress granules

Stress granules (SGs) are ribonucleoprotein (RNP) granules that can form in a cell as a response to various types of stress, including oxidative stress, heat shocks, osmolar stress and viral infection (*196*). They got particular prominence in recent years as proteins involved in their formation have been implicated in aberrant transitions at the core of severe neuropathic diseases like amyotrophic lateral sclerosis (*188*).

SGs typically contain polyadenylated mRNAs that are not actively translated as well as preinitiation complexes including the small ribosomal subunit, but no actively translating ribosomes (*197*). In fact, if polysome disassembly is inhibited SGs do not form (*198*), indicating that an increase of free cytoplasmic RNA is important. Concomitant with their high RNA content (*199*), SGs contain many RNA binding proteins like FET family proteins, G3BP1, G3BP2, hnRNPA1 and others (*200–202*).

How do these heterogenous SGs, with hundredths of different RNAs and proteins actually assembly *in vivo*?

On the one hand, RNA self-association has been shown to contribute to SG assembly (*203*), on the other hand, many of the associated RNA binding proteins have been implicated as

potentially being essential for SG assembly, including CSDE1, G3BP1, G3BP2, HDAC6, PPRC2C, TIA1 and UBAP2L (*201*, *202*, *204–206*). Therefore, for a long time it has not been clear how SG essentially assemble and which factors are indispensable until recently complementary work from three different laboratories has converged to thoroughly define the biogenesis of SGs and their core assembly components (*207–209*).

Of the many different RNA binding proteins only G3BP1 or G3BP2 are essential for SG formation, and only in cell in which these two proteins are knocked-out no SG formation can be observed (*208*). G3BP1 and G3BP2 have a high sequence similarity and seem to be functionally similar (*208*), although G3BP2 exhibits a higher valency and has a slightly higher phase separation propensity (*209*). For the remainder of the section I am going to refer to both G3BP1 and G3BP2 collectively as G3BP for simplicity.

G3BP does not phase separate under physiological conditions on its own but phase separation can be triggered by molecular crowding agents (*208*, *209*). However, in presence of long single stranded RNA G3BP robustly phase separates (*208*). Phase separation is independent of the RNA sequence but correlates with RNA length (*208*). Besides the requirement of single strandedness it has also been shown that the RNA needs to be unfolded to trigger G3BP phase separation (*209*).

To bind to RNA G3BP has two distinct RBDs at its C-terminus, one folded RNA recognition motif (RRM) and one RNA binding IDR. Further to the N-terminus G3BP contains two additional IDRs, one very acidic and one with a slightly positive charge (*207*, *208*). At the N-terminus G3BP has a dimerizing NTF2 domain. Work by the Taylor, Brangwynne and Alberti laboratories dissected the relevance of all these factors to get a more detailed understanding of SG assembly (*207–209*).

Dimerization via the NTF2 domain is important for phase separation to increase the overall valency of the protein, it however is not specific to NTF2 and can be replaced with other dimerization moieties (*208*), however this requires higher concentrations indicating that NTF2 also leads to higher order assemblies by interacting with UBAP2L, another SG protein (*207*). The two RBDs at the C-terminus also serve to increase valency by binding RNAs multiple times, therefore just removing one of those significantly impairs phase separation (*208*).

What do the remaining two IDRs contribute? It is commonly assumed that IDRs promote phase separation in general, this is however not true for the IDRs in G3BP. Strikingly the acidic IDR turns out to have an autoinhibitory function on phase separation which is regulated by the slightly basic IDR, this is apparent as by deleting the acidic IDR, SGs form spontaneously,

while in absence of the basic IDR SGs do not form at all (*207–209*). This is an interesting observation illustrating that the mere occurrence of IDRs in phase separating systems does not necessarily proof a biophysical causality and thus needs to be considered under the relevant circumstances (*210*). Noteworthy, other proteins can be subsequently recruited to assembled SGs and can exhibit modulatory functions (*209*), for example by influencing G3BP/RNA phase separation in a positively or negatively cooperative manner (*208*).

One additional interesting observation about the physiological significance of SGs is that they themselves do not inhibit translation as a stress response, which agrees with previous observations (*206*, *211*), but in fact sequester the released mRNAs and thus prevent them from aggregating, which might be important for restarting translation once the stress factor has passed (*209*).

In summary SGs are particularly interesting examples of biomolecular condensates, which combine many assembly factors, like multivalency and IDRs, but also nicely stresses that not all IDRs behave the same. The putative RNA chaperoning functionality of SGs moreover represents a great example illustrating the physiological relevance of LLPS.

1.3.3.2 The C. elegans centrosome

Phase separation processes have also been implicated in organizing components involved in cell division (*212*, *213*). For example, several components involved in mitotic spindles—a structure ensuring the proper segregation of chromosomal DNA to the emerging daughter cells during mitosis (*214–216*)—have been shown to undergo phase separation (*217–219*). This is interesting as self-organization has long been considered to be relevant for mitotic spindle assembly (*220*) and LLPS seems to be predisposed to aid in such a process.

In this section I am going to further discuss one specific example of phase separated mitotic structures—the *C. elegans* centrosome.

The centrosome is a prominent micrometer-sized membraneless organelle, that is the major microtubule organizing center (MTOC) in most cells and can be found at the poles of mitotic spindles (*221*, *222*). The centrosome is composed of two centrioles that are surrounded by the pericentriolar material [PCM, (*223*)]. In *C. elegans* the PCM is composed of several proteins including polo-like kinase 1 (PLK1), the spindle-defective proteins 2 and 5 (SPD2 and SPD5) and others (*224–227*).

SPD5 is a protein enriched in coiled-coil structures that has been shown to self-assemble to form a network that can subsequently recruit additional PCM components (*228*). Moreover, SPD5 can even form phase separated, spherical condensates *in vitro* that are initially liquid but

rapidly solidify (*229*). Intriguingly, SPD5 can achieve this without any IDR as it only contains nine predicted coiled-coil domains making up almost half the protein (*229*). Other PCM proteins can co-partition into the SPD5 droplet and they can collectively concentrate tubulin and promote microtubule polymerization (*229*).

The PCM in *C. elegans* has important mechanical functions during the cell cycle, as the centrosome needs to withstand spindle pulling forces. Hence, it is interesting that the material properties of centrosomes are changed during the cell cycle (*230*, *231*). At the onset of mitosis the PCM becomes hard but pliable due to actions of PLK1 and SPD2—PLK1 phosphorylates SPD5 conferring strength to the centrosome, while SPD2 binding increases pliability (*231*). However, in anaphase PLK1 and SPD2 leave the centrosome and protein phosphatase 2A removes PLK1 mediated SPD5 phosphorylations, which progressively weakens the PCM until it can be pulled apparat by microtubule mediated forces (*231*).

In summary the *C. elegans* centrosome is a beautiful example for phase separation mediated by structured proteins and the dynamics of the material properties of biomolecular condensates, which can have important implications for biological function (*229–231*).

1.3.3.3 Biomolecular condensates on membranes

Many biomolecular condensates operate freely in the cyto- or nucleoplasm in a membraneless fashion. However, it is increasingly becoming clear that condensates on membrane surfaces can also play important roles in the cell. They for example can participate in membrane signaling (*232*), by e.g. organizing membrane associated actin polymerization (*233*, *234*) or they can by mediate the formation of tight junctions at cell-cell contact sites (*235*, *236*).

How does this process differ from in solution phase separation processes?

Biomolecular condensates on membranes commonly also form via multivalent interactions (*234*, *237*), similar to phase separation processes described above. However, while membraneless organelles usually form via three-dimensional phase separation, clustering on membranes is inherently a two-dimensional process (*232*). Therefore, proteins can phase separate at much lower absolute concentrations (*237*), which can be caused by the effective increase in local concentration due to the reduced dimensionality. Due to this it is possible that phase separation on membranes can occur, while the overall concentration is low enough to avoid phase separation in solution (*232*).

Due to the location at the surface of the cell, membrane associated biomolecular condensates are particularly amenable to regulate the transfer of information from the cell exterior into the cell.

This was for example convincingly shown for T cell receptor (TCR) signaling (*238*, *239*). Intriguingly, multivalency is generated by phosphorylation of tyrosine residues in the important adapter protein LAT (linker for activation of T cells). Downstream effector proteins bind phosphorylated LAT forming a liquid-like, phase separated network on the membrane (*238*). The condensate subsequently regulates mitogen-activated protein kinase (MAPK), as well as actin polymerization (*238*). The composition of the TCR associated condensate can also locally and temporarily change and thereby influence the interaction of LAT with the actin cytoskeleton (*240*). An interesting observation in this signaling system was also that the dwell time for MAPK pathway proteins on the membrane is increased due to phase separation, which in turn leads to an increased activation (*241*).

Analogously, for nephrin—a transmembrane protein that is important for the formation of the glomerular filtration barrier in the kidney (*242*)—it was also shown that it can form liquid condensates on the plasma membrane due to phosphorylation and subsequently interact with the actin cytoskeleton (*243*). Also in these condensates the dwell time of actin interacting proteins is increased, which increases their the specific activity leading to an increased actin polymerization (*233*).

Together, these two specific examples show that liquid condensates on membrane surfaces can have important functional consequences. Additionally, they could enable kinetic proofreading of signaling information to ensure robust and reliable signaling (*233*, *241*).

As long as the receptor is not phosphorylated, the signaling effector proteins can diffuse quickly and therefore do not reside for extended periods of time on the membrane, which makes their activation unlikely. Upon receptor activation and phosphorylation, phase separation on the membrane is triggered. This increases the dwell time of the downstream signaling proteins, which leads to their activation. Hence, this mechanism prevents spontaneous signaling and ensures that robust receptor signaling only occurs under induced conditions (*233*, *241*). Although it so far has only been shown for a few examples, it is conceivable that this mechanism could be generalizable and might be involved in many naturally occurring signaling processes.

1.3.4 Synthetic biomolecular condensates

Up to now we discussed naturally occurring biomolecular condensates. However, the growing knowledge about biological phase separation has also sparked several efforts to engineer this process in a synthetic biology approach. These efforts can be broadly divided into two classes

i) tools that allow to study natural phase separating processes and ii) the development of synthetic phase separating systems for biological and biotechnological applications.

1.3.4.1 Synthetic tools to study phase separation in biological systems

Employing synthetic approaches to study biological phase separation directly in living cells promises to lead to new discoveries, as simplified perturbation experiments can be used to dissect and isolate more complex processes (*244*, *245*). To this end, the Brangwynne and Toettcher laboratories pioneered optogenetic tools that can be used to control phase separation *in vivo* (*246–250*).

In the first version of the technology light irradiation triggers oligomerization of an optogenetically controlled domain (*251*), which is fused to IDRs of naturally occurring phase separating proteins (*248*). Depending on intensity and length of light irradiation this can than lead to phase separation and thereby allows to study the behavior of the IDR *in vivo*. This technology was combined with CRISPR-Cas9 technology—allowing to target defined genomic regions by base pairing of guide RNA (*123*, *252*)—to tether LLPS to specific genomic loci and study their functionality (*250*). One key parameter of the system is the oligomerization of the optogenetic domain, which increases the valency and thus triggers phase separation, in analogy to naturally occurring phase separating processes (*207–209*).

As oligomerization in this case however is not perfectly quantitatively defined the technology was further optimized by using an optogenetic dimerizer (*253*). One part of the dimerizer is fused to an oligomeric core that has a defined stoichiometry (*254*), while the other half is fused to the IDR of interest. Upon light irradiation the IDR will then be tethered to the oligomeric core and can then undergo phase separation (*249*). With this technology it was now possible to map full phase diagrams and thereby fully measure LLPS systems (*249*).

Both of these technologies induce phase separation upon light irradiation and if the light is applied in a certain pattern, phase separation will be restricted to certain areas, which is of interest to understand polarized phase separation processes.

Unfortunately, in both instances blue light is needed to trigger phase separation. Blue light irradiation can be toxic for cells and therefore these experiments cannot be performed infinitely long. Long term experiments would however be interesting to study spatial memory and developmental processes.

To overcome this limitation Dine and co-workers employed an inverse strategy (*246*). They construct a phase separation system based on two proteins, that oligomerize in the absence of light but can dissociate upon light irradiation (*255*). In analogy to above, they fuse IDRs to

these oligomerizing proteins, which in turn robustly phase separate in the dark, but dissolve upon light irradiation. The authors use this system to dissolve phase separated structures in a defined area of the cell, intriguingly even after light removal the pattern persists (*246*). This indicates that phase separation can be used to establish a spatial memory, which might be relevant to for example confer polarization to cells.

In summary, these elegant synthetic approaches to engineer phase separation and thereby discover new mechanisms that are relevant for the formation and function of biomolecular condensates are particularly insightful and it is interesting to see to what new discoveries they will lead.

1.3.4.2 Bioengineering with phase separation

The aforementioned studies represent tools to study phase separation in cells. Phase separation is inherently capable to drive self-assembly which can have broad implications for biotechnology. Therefore, it is interesting to apply phase separation to specific technological problems as well as to engineer phase separation and create new phase separating systems to be able to control and regulate the engineered processes.

Engineering with naturally phase separating proteins

The knowledge about native phase separating proteins is constantly growing and therefore they represent a good starting point for phase separation-based bioengineering applications. For example, the optogenetic systems described above have been engineered to assemble metabolic clusters (*247*). Due to their assembly, the metabolic flux and thereby the small molecule product is changed. As this approach is light controlled it establishes a light switch for metabolic pathways. Particularly, for biotechnological production processes this could be interesting, as it would allow to rapidly switch the desired production outcome.

Similarly, recently the IDR region of the *C. elegans* P-granule protein LAF-1—an ATP-dependent RNA helicase, which has been shown to phase separate (*256*)—has been used for engineering purposes (*257*, *258*). While the IDR itself hardly phase separates, it can be engineered to do so more readily by concatenating multiple copies of it. The number of consecutive copies determines the respective saturation concentration and importantly this process can be further customized (*257*). These LAF-1 based synthetic organelles allow to recruit cargos based on small signaling peptides and to release proteins based on protease mediated cleavage, thus allowing to specifically regulate the condensate composition (*257*). This technology was further refined by adding solubility tags and a photocleavable domain, thereby getting optical control of phase separation (*258*). For this system it however needs to

be noticed that the photoconversion is irreversible, which sets it apart from the aforementioned optogenetic tools.

Engineering artificial phase separating systems

If naturally phase separating systems are used for engineering purposes, one concern is that they could interact with endogenous phase separated systems and thus interfere with the normal physiology of the cell. Therefore, several efforts in this direction have been made. Essentially one of the first major landmark papers I discussed in **Section 1.3.2.1**, studies the effects of multivalency on phase separation with naturally non-phase separating biomolecules (*156*). In this section I will discuss two more recent examples based on artificial IDRs (*259*) and selective multivalency networks (*260*).

As discussed in section **1.3.2.2**, IDRs have been shown to be crucial for many phase separation processes. Thus, it is not surprising that they have been explored for synthetic condensates. Elastin-like polypeptides (ELPs) are stereotypical IDPs composed of multiple repeats of a pentapeptide sequence [valine-proline-glycine-any amino acid-glycine, (*261*)]. These, have for example been used to program phase separation processes *in vitro* and they have even been shown to enable the formation of multilayered phase separated systems (*259*). Depending on their application these ELPs can be tuned to assume different shapes (*262*), however they are typically not used as tools in cells.

In yeast multivalent networks were recently designed to phase separate (*260*). This system uses a dimerizing and a tetramerizing building block. Importantly, the dimerizing and tetramerizing parts can interact over a specific protein-protein binding domain. By tuning the interaction strength of the binding domain it is possible to adapt the phase separating behavior of the network (*260*). Due to this modular nature this system offers quite some potential for further engineering, but so far it was mostly used to study multimeric phase separation *in vivo* and to analyze how translation can localize mRNAs in the cell (*260*). Hence, it will be interesting to see how this technology is going to be further utilized for synthetic biology approaches in the future.

1.4 Objectives—Optimized GCE tools for eukaryotes

GCE is one of the most powerful techniques that enables to site-specifically modify proteins in living cells. Particular in combination with biorthogonal chemistry that subsequently enables further modification of the POI this technology has huge promises. It could for example be used exquisitely to study proteins in their native environment or to create new advanced, fully artificial biopolymers.

To this end, it would be particularly important that multiple different ncAAs could be simultaneously incorporated into the POI. However, to do this in eukaryotes, GCE technology is fundamentally limited by three problems. First, GCE technology lacks mRNA specificity, which can lead to background incorporation and toxicity. Second, only a limited number of orthogonal RS/tRNA pairs are useful in mammalian cells and new ones need to be discovered to allow the synthesis of truly artificial polypeptides. Third, as only two stop codons can be suppressed at a time, the technology currently cannot incorporate more than two distinct ncAAs into proteins specifically, while still faithfully terminating protein translation. In this thesis I am going to address these three problems to optimize GCE for mammalian applications.

1.4.1 Developing mRNA specific GCE in in mammalian cells

Establishing orthogonal translation in mammalian cells is challenging, particularly as the approaches developed for bacteria cannot be easily transferred. I hypothesize that it should be possible to create an artificial organelle dedicated to protein engineering. In this organelle, the genetic code should be selectively expanded for the POI, ensuring specific ncAA incorporation with minimal perturbation of housekeeping translation.

1.4.2 Developing inducible GCE to avoid toxicity

Complementary, to avoid toxic effects of GCE components during long term expression experiments, inducible expression systems that allow to regulate the expression of not only the RS but also the orthogonal tRNA will be developed.

1.4.3 Developing mutually orthogonal GCE systems

I hypothesize that it should be possible to equip a cell with multiple, mutually orthogonal synthetic organelles, which each execute distinct orthogonal genetic codes. By achieving this I would solve two major problems of GCE technology. First, this would enable to multiple times reuse the same RS/tRNA pair, thus establishing a completely novel way to create orthogonal RS/tRNA pairs. Second, with multiple mutually orthogonal synthetic organelles it should be possible to reassign the same codon multiple times to distinct ncAAs, which would provide new orthogonal codons by means of this new concept of localized translation.

Chapter 2 Summaries of published manuscripts

2.1 Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes

This work was published as:

Reinkemeier CD*, Estrada Girona G* & Lemke EA (2019). Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. Science, Vol 363, Issue 6434, doi:10.1126/science.aaw2644, * These authors contributed equally

The article is can be found in Appendix II.

This sentence specifies my contributions to the manuscript:

E.A.L. conceived the project. C.D.R. and G.E.G. designed and performed all experiments. C.D.R., G.E.G., and E.A.L. analyzed all of the data and cowrote the manuscript.

All data shown in this manuscript, except the ones in Fig. S8B&D, were acquired by me.

Summary

In this work I combined two assembler classes—an assembler is defined as anything that forms an organelle in a cell—to generate orthogonally translating (OT) organelles in living cells. In particular, I use phase separating proteins (FUS/EWSR1 or SPD-5) as well as microtubule motor proteins (KIF13A or KIF16B) as assemblers. These are then fused to an RNA recruitment system (MCP, ms2 bacteriophage coat protein), which selectively binds to a specific RNA secondary structure (ms2-loops), as well as a genetic code expansion system (PylRS). The OT organelles highly enrich mRNA::ms2 and $tRNA^{Pyl}$, effectively depleting it from the remainder of the cell. Subsequently, even cellular ribosomes are recruited to perform orthogonal translation within the organelle.

Thereby, OT organelles enable site- and mRNA-specific ncAA incorporation, decoding one specific codon exclusively in the mRNA of choice. I show that this works selectively for all three stop codons and with different ncAAs. Moreover, this can be done for multiple stop codons in one protein and in a variety of proteins ranging from simple fluorescent proteins to nucleoporins and even transmembrane proteins that need to be inserted into them membrane during translation at the endoplasmic reticulum.

My results demonstrate a simple but effective approach to generate OT organelles in eukaryotic cells. These semi-synthetic cells are the first ones that are capable of simultaneously executing two distinct genetic codes, one canonical one for untargeted proteins in the cytoplasm and one

expanded genetic code for chosen mRNAs inside of the synthetic organelles, thereby providing a route towards customized orthogonal translation and protein engineering.

2.2 Raising the ribosomal repertoire

This work was published as:

Reinkemeier CD & Lemke EA (2020). Raising the ribosomal repertoire. Nature Chemistry, 12, 503-504, doi:10.1038/s41557-020-0476-6

The article is can be found in Appendix III.

This sentence specifies my contributions to the manuscript:

C.D.R. and E.A.L. cowrote the manuscript.

Summary

In this article I discuss a recent study published by the Chin lab (*54*) that develops multiple mutually orthogonal PylRS based genetic code expansion systems. This is of particular interest, as for the synthesis of truly artificial polymers the number of available orthogonal RS/tRNA pairs is limiting.

Prior to this work, in *E. coli* only a handful of orthogonal RS/tRNA pairs have been reported of which mainly the *Methanococcus janashii* tyrosyl RS/tRNA pair as well as the PylRS/$tRNA^{Pyl}$ pairs derived from methanogenic archaea have been widely used for GCE applications.

Arguably the most useful pair is the PylRS/$tRNA^{Pyl}$, as I discussed above (see **Section 1.2**). Before the publication by Chin and co-workers only two mutually orthogonal PylRS have been reported (*53*, *57*). The earliest described PylRS variants (e.g. from *Methanosacina mazei*) that were used for GCE contain an N-terminal domain that is important for tRNA recognition. Interestingly, some PylRS enzymes lack an N-terminal domain and could thus be developed to be orthogonal to the PylRS variants containing an N-terminal tRNA binding domain (*53*, *57*). In their newly published work Chin and co-workers identify a second class of PylRS that also lack an N-terminal tRNA binding domain (*54*). Based on these three existing, naturally diverse classes they utilize directed evolution to create three pairs that are mutually orthogonal.

This now allows to utilize this powerful enzyme family to site-specifically incorporate three distinct ncAAs into protein *in vivo*.

2.3 Inducible Genetic Code Expansion in Eukaryotes

This work was published as:

Koehler C, Estrada Girona G, Reinkemeier CD & Lemke EA (2020). Inducible Genetic Code Expansion in Eukaryotes. ChemBioChem, 21, 1-5, doi:10.1002/cbic.202000338

The article is can be found in Appendix IV.

The following sentence specifies my contributions to this manuscript:

For this work I performed flow cytometry experiments and cloning work. Additionally, I co-wrote manuscript.

Summary

One limitation of GCE technologies is the potential toxicity due to constant expression of the suppressor RS/tRNA pair (in this case PylRS/tRNA^{Pyl}). In particular this is a concern for long term expression experiments as would be the case for the generation of stable amber suppressing cell lines, which would be of high interest for biotechnological applications. This problem is further aggravated by the requirement of high tRNA expression levels to effectively compete with the endogenous release factor machinery, in order to achieve sufficient amber suppression levels.

It can be hypothesized that by inducing tRNA^{Pyl} as well as PylRS expression, it should be possible to avoid toxic interferences with the normal physiology of the host.

mRNAs can be efficiently transcribed by RNA-polymerase II (PolII), while tRNAs need RNA-polymerase III (PolIII) for efficient transcription. Thus, respective inducible promoters are needed to achieve inducibility of the GCE system.

Both the TREx- as well as the Tet-On system were explored. These are well defined inducible expression systems, and they were tested in several PolII- and PolIII-based promoter designs with flanking tetracycline operator (TetO) signals.

In the Tet-On system the reverse tetracycline-repressor protein (rtTA) can only bind to the TetO sequence in the DNA upon the addition of the small molecule doxycycline. Upon binding to the TetO sequence it recruits PolII for transcription but not PolIII. Therefore, this system allows to induce the expression of the PylRS but not of tRNA^{Pyl}, in fact due to the binding of rtTA to the PolIII-promoter used for tRNA^{Pyl} expression, addition of doxycycline can even reduce transcription levels and lower GCE efficiency.

In contrast to this, in the TREx-system the tetracycline repressor protein (TetR) constitutively binds to the TetO sequence in the DNA and thereby blocks the access of all RNA-polymerases.

Upon tetracycline addition TetR dissociates from the DNA making the promoter accessible. Subsequently, PolII can transcribe the PylRS gene and PolIII can transcribe $tRNA^{Pyl}$.

In summary this work demonstrates that the Tet-On system can be used to regulate PylRS expression to construct an inducible GCE system, however with the TREx-system it is possible to control both PylRS and $tRNA^{Pyl}$ expression, which allows to obtain an on-demand GCE system that should help to alleviate toxicity concerns for eukaryotic applications.

Chapter 3 Microtubule-based orthogonally translating organelles in mammalian cells

3.1 Introduction

I recently developed a membraneless approach that enables orthogonal translation of selected mRNAs with an expanded genetic code in eukaryotes (*263*), this work is also part of this thesis (see **Chapter 2** and **Appendix II**). This first OT organelle version is comprised of i) the *Methanosarcina mazei* derived $tRNA^{Pyl}$/PylRS system; ii) an mRNA recruitment system in which two ms2-loops are inserted into the 3'UTR of the mRNA of the POI, which are bound specifically by the MCP protein (*264*); iii) an assembler component that drives organelle formation within the cell. As assemblers I for example used the proteins FUS and EWSR1 that both contain PLDs and have been shown to phase separate into liquid, gel or solid states (*188*, *189*), termed **P1**. Additionally, I used hyperactive truncations of microtubule motor proteins, like $KIF16B_{1\text{-}400}$ (termed **K2**) that constitutively move towards the microtubule-plus end (*265*). The so far best working orthogonally translating (OT) organelle **K2::P1** is built by co-expressing $KIF16B_{1\text{-}400}$::FUS::PylRS with $KIF16B_{1\text{-}400}$::EWSR1::MCP (*263*).

In this chapter, I will show how I expanded this microtubule-based technology, by demonstrating that the design of OT organelles can be simplified by directly fusing the assemblers to an MCP::PylRS fusion, and by showing that the microtubule motor domain can be replaced by a microtubule-end binding protein like EB1 (*266–268*) to facilitate OT organelle formation. Intriguingly, the EB1 based synthetic organelles assume a substantially different and much thinner structure and appear to coat the microtubules, indicating that OT organelles can also be formed in a much more minimalistic fashion.

3.2 Results

3.2.1 Design of improved OT organelles

To simplify and improve OT organelles I explored a direct fusion of both functional entities (MCP and PylRS respectively) to the previously established assembler combination of KIF16B and FUS. Additionally, I also replaced the kinesin motor domain with EB1 and tested it with and without the phase separation based assemblers (FUS/EWSR1, **Figure 3-1A**, shows an overview of all tested assembler strategies). I chose to use EB1 as a replacement for the microtubule motor domain for two reasons. First, the previous approach used microtubule-plus end directed motor proteins, as EB1 preferentially binds to microtubule-plus ends it should

thus confer a similar localization to the OT organelles. Second, recently the microtubule associated protein TPX2 was shown to phase separate (*269*), also EB1 was implicated in potentially forming condensates with another microtubule associated protein [Tau, (*270*)]. Therefore, I hypothesized that EB1 might combine both the phase separation potential as well as the localization information required for OT organelle assembly.

In particular I designed the following four new systems (I use "•" do indicate co-expression of two constructs):

dK2::P1 = KIF16B$_{1\text{-}400}$::FUS::MCP::PylRS

EB1 = EB1::PylRS•EB1::MCP

EB1::P1 = EB1::FUS::PylRS•EB1::EWSR1::MCP

dEB1::P1 = EB1::FUS::MCP::PylRS

I then tested these new constructs in an FFC based assay to analyze selectivity. For this assay I use a dual color reporter in which mCherry and egfp mRNA are expressed from one plasmid with an amber stop codon at a permissive site (mCherry^{190TAG}, egfp^{39TAG}, mRNAs are denoted in lower case) and the mCherry mRNA is tagged with ms2-loops in the 3'UTR (**Figure 3-1B**). In case of cytoplasmic GCE one would expect to produce both full-length EGFP and mCherry, resulting in a diagonal population in FFC analysis. If an OT organelle is working selectively, meaning that the amber codon is only suppressed for the targeted mRNA::ms2, one would expect to produce only full-length mCherry, resulting in a vertical mCherry positive population in FFC. The amber stop codon in the egfp mRNA, translated in the cytoplasm would lead to translation termination as it cannot be suppressed by tRNA^{Pyl}. From the FFC data I then quantify selectivity and efficiency to compare the different OT organelles. I define fold change selectivity of a given system as the ratio of the mean mCherry signal divided by the mean EGFP signal, divided by the ratio of the mean mCherry to mean EGFP signal of the cytoplasmic PylRS system. The relative efficiency is defined as the mean mCherry signal of a system divided by the mean mCherry signal of the cytoplasmic GCE system.

3.2.2 Direct fusions of MCP and PylRS to assemblers enable selective orthogonal translation

Cytoplasmic GCE expectedly leads to expression of both full-length EGFP and mCherry, while the OT$^{K2::P1}$ (**K2::P1** = KIF16B$_{1\text{-}400}$::FUS::PylRS•KIF16B$_{1\text{-}400}$::EWSR1::MCP) system leads to predominant production of full-length mCherry (6.6 fold change selectivity, 30% efficiency, **Figure 3-1C,D**). The OT$^{dK2::P1}$ system also permits highly selective mCherry production, which tends to be slightly more selective (7.1 fold change selectivity), but a bit less efficient

(21% efficiency). Replacing the kinesin motor domain with EB1 also permits selective GCE. Here I observe that already the OT^{EB1} system leads to a selectivity of about 2.6 fold, however this is system also leads to a marked drop in efficiency to 11%. Adding a phase separation based assembler to this (OT$^{EB1::P1}$) improves selectivity to 3.4 fold and restores efficiency to 39% of the cytoplasmic GCE system. For the EB1 based system I can observe a substantial increase in selectivity by using a direct genetic fusion of MCP and PylRS (OT$^{dEB1::P1}$), resulting in a fold change selectivity of 4.9 and an efficiency of 22%.

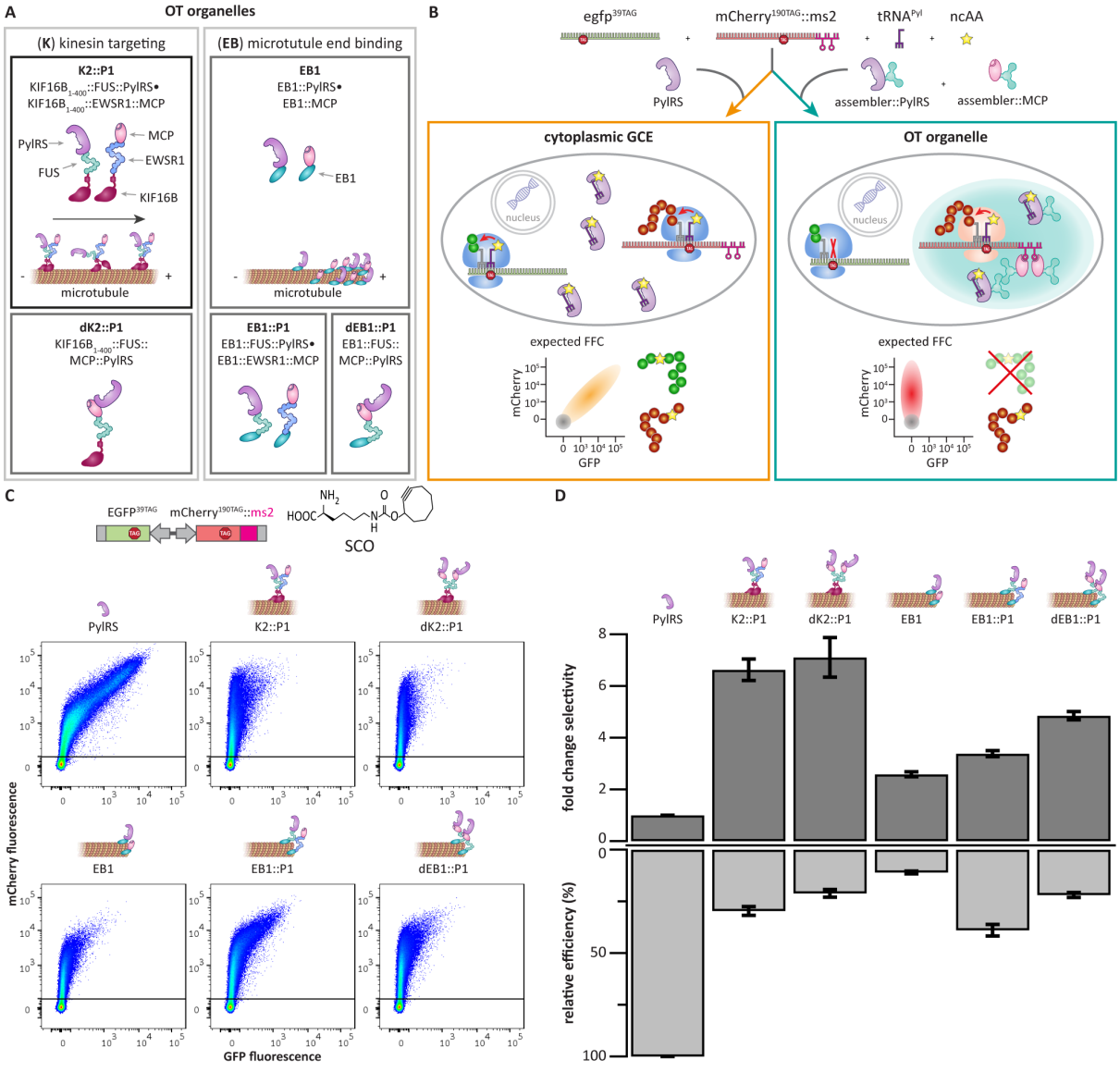


Figure 3-1 Synthetic organelles on the microtubule cytoskeleton permit selective orthogonal GCE. (**A**) Schematic representation of orthogonally translating (OT) organelles used in this study. K2::P1, fusion of a hyperactive truncation of kinesin KIF16B (KIF16B$_{1-400}$) with FUS/EWSR1 and PylRS/MCP (as described previously (*263*)); dK2::P1, direct fusion of KIF16B$_{1-400}$ with FUS, MCP and PylRS. EB1, fusion of the microtubule-end binding protein 1 (EB1) with PylRS and MCP; EB1::P1, combination of EB1 with FUS/EWSR1 fused to PylRS and MCP; and dEB1::P1, direct fusion of EB1 with FUS, MCP and PylRS. (**B**) Schematic overview of fluorescence flow cytometry (FFC) based assay to analyze orthogonal translation. egfp and mCherry mRNA with stop codons at permissive sites are co-expressed with tRNA^{Pyl} and either an OT organelle or a cytoplasmic PylRS variant. mCherry mRNA is tagged with ms2-loops and is hence recruited selectively into

synthetic organelles. In presence of a cytoplasmic PylRS system ribosomes cannot distinguish the two mRNAs leading to amber codon suppression in both egfp and mCherry, hence both full-length proteins are produced resulting in a diagonal EGFP and mCherry positive population in FFC. In presence of a working OT organelle the amber codon is only suppressed in the recruited, and locally translated mCherry mRNA, while it is processed canonically in the egfp mRNA leading to translation termination. Therefore, a working OT organelle leads to exclusive full-length mCherry production resulting in a vertical mCherry positive population. (**C**) FFC analysis of HEK293T cells co-expressing the indicated dual reporter, $tRNA^{Pyl}$ as well as the indicated GCE system in presence of a lysine based ncAA with a cyclooctyne modified side chain (SCO (*63*)). The dot plots show the sum of four independent experiments, axis indicate mCherry and GFP fluorescence in arbitrary units. Controls without the MCP protein present are shown in **Supplementary Figure 3-1**; **Supplementary Figure 3-2** shows an inverted reporter, in which the egfp mRNA is tagged with ms2-loops. (**D**) Bar graph corresponding to OT organelles shown in (C). Dark grey bars show the fold change selectivity, defined as the ratio of the mean mCherry divided by the mean GFP signal of a given system normalized to the ratio of the mCherry to GFP signal for the cytoplasmic GCE system. Light grey bars show the relative efficiency of a given system, defined as the mean mCherry signal divided by the mean mCherry signal of the cytoplasmic PylRS system. Shown is the mean of four independent experiments, error bars represent the SEM.

I verified that this selectivity depends on the ms2-MCP interaction by omitting the MCP construct in the FFC experiments and observed that without MCP full-length mCherry production is not enhanced (**Supplementary Figure 3-1**). Also when I invert the reporter and attach the ms2-loops to the egfp instead of the mCherry mRNA I expectedly observe predominant full-length EGFP production (**Supplementary Figure 3-2**).

3.2.3 EB1 and K2 based OT organelles form along microtubules

A key assembly component for both types of OT organelles is the localization to the microtubule cytoskeleton. To confirm that this is in fact the dominant localization of the OT organelles I use immunofluorescence (IF) stainings to visualize microtubules and PylRS (**Figure 3-2A**, **Supplementary Figure 3-3**). For the K2 based organelles ($OT^{K2::P1}$ and $OT^{dK2::P1}$) I observe that in most cells one large organelle is formed. This enriches microtubules and rearranges the microtubule cytoskeleton to some extent. In contrast for all EB1 based synthetic organelles (OT^{EB1}, $OT^{EB1::P1}$ and $OT^{dEB1::P1}$) I do not observe a major alteration of the microtubule cytoskeleton. Instead the organelles form along microtubule fibers and spread out substantially throughout the cell.

3.2.4 EB1 and K2 based OT organelles enrich $tRNA^{Pyl}$, mRNA::ms2 and ribosomes

Both EB1 and K2 based OT organelles perform selective orthogonal translation of only targeted mRNAs. In order to achieve orthogonal translation with an expanded genetic code, $tRNA^{Pyl}$ needs to be depleted from the cytoplasm and enriched it in the organelle. To analyze if this is in fact the case, I performed IF and fluorescence *in situ* hybridization (FISH) stainings. I co-stained PylRS and $tRNA^{Pyl}$ in cells transfected with the different organelle variants and observed that as expected $OT^{K2::P1}$ forms a large synthetic organelle that highly enriches $tRNA^{Pyl}$ and depletes it from the cytoplasm (**Figure 3-2A**, **Supplementary Figure 3-4**). Also the new

K2 based organelle variant OT$^{dK2::P1}$ forms a similar large synthetic organelles that fills up a substantial part of the cytoplasm and highly enriches tRNA^{Pyl}. In contrast to these large organelles observed when a kinesin motor protein is used as an assembler, using a microtubule-end binding protein as an assembler leads to a substantially different shape. For OT^{EB1}, OT$^{EB1::P1}$ and OT$^{dEB1::P1}$ organelles PylRS exhibit a fibrous shape in the cells cytoplasm (**Figure 3-2A** and **Supplementary Figure 3-3**), importantly however still tRNA^{Pyl} is highly enriched.

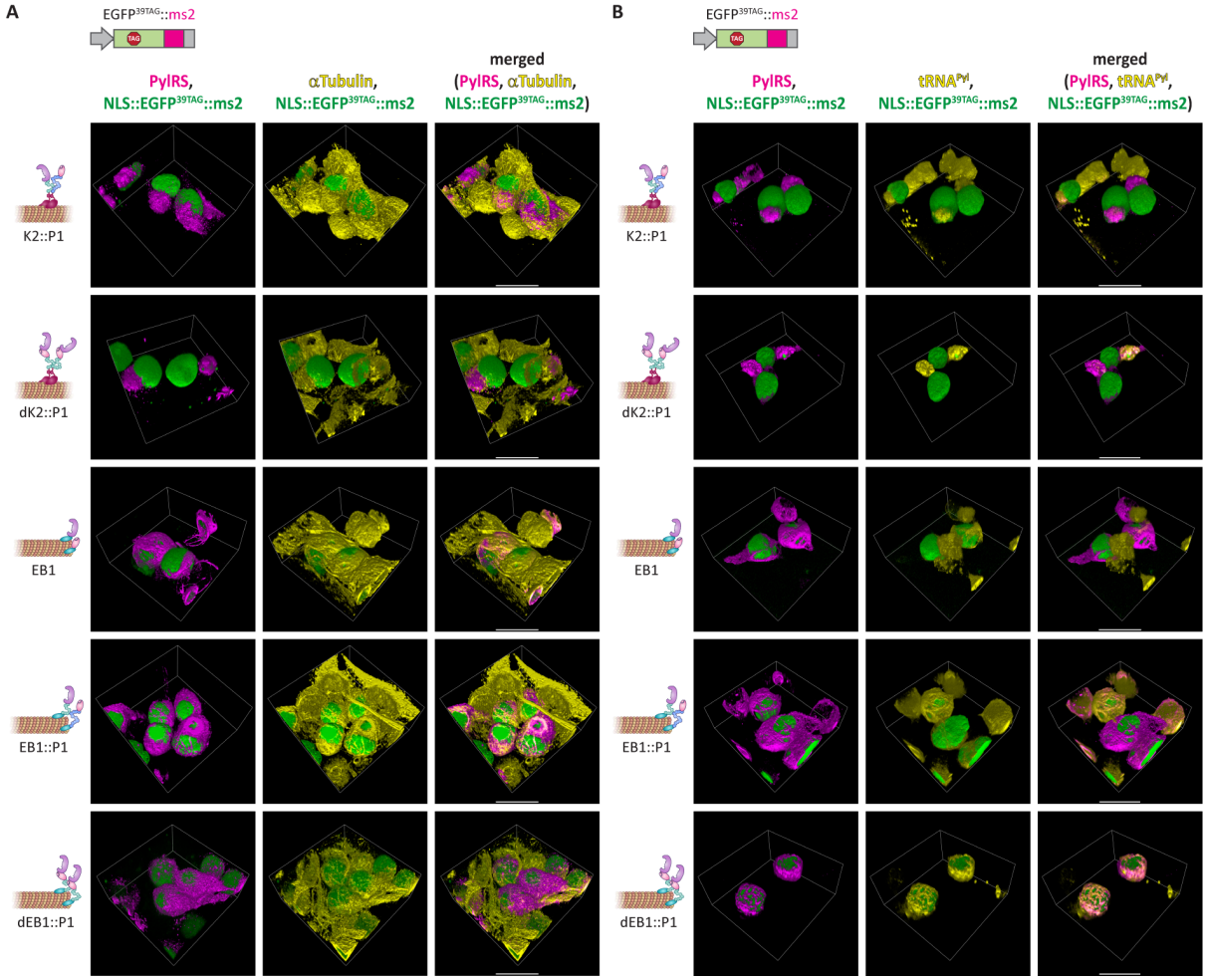


Figure 3-2. Synthetic organelles form along microtubules and highly enrich tRNA^{Pyl}. (**A**) 3D reconstructions of IF stainings against PylRS (magenta) and αTubulin (yellow) in cells expressing tRNA^{Pyl} and the indicated synthetic organelles in presence of the ncAA SCO. Cells were additionally transfected with a construct encoding EGFP^{39TAG}::ms2 (green) to identify amber suppressing cells. Additional images are shown in **Supplementary Figure 3-3**, to confirm that microtubule based synthetic organelles co-localize with αTubulin. (**B**) 3D reconstructions of combined IF and FISH stainings against PylRS (magenta) and tRNA^{Pyl} (yellow) of cells expressing EGFP^{39TAG}::ms2 (shown in green), the indicated synthetic organelles and tRNA^{Pyl}. Additional images confirming high local enrichment of tRNA^{Pyl} in OT organelles are shown in **Supplementary Figure 3-4**. (Scale bars = 20 μm)

Additionally, to enable local translation it is necessary, that also mRNA::ms2 as well as ribosomes can be found in the synthetic organelles. To analyze this I stained cells expressing the five different OT organelle variants for mRNA::ms2 and PylRS (FISH combined with IF), or for the ribosomal protein RPL26L1 and PylRS (IF). In agreement with the above described

imaging experiments I observed similarly shaped organelles, which also show co-localized mRNA::ms2 as well as ribosomes. (**Figure 3-3**, **Supplementary Figure 3-4, Supplementary Figure 3-5**).

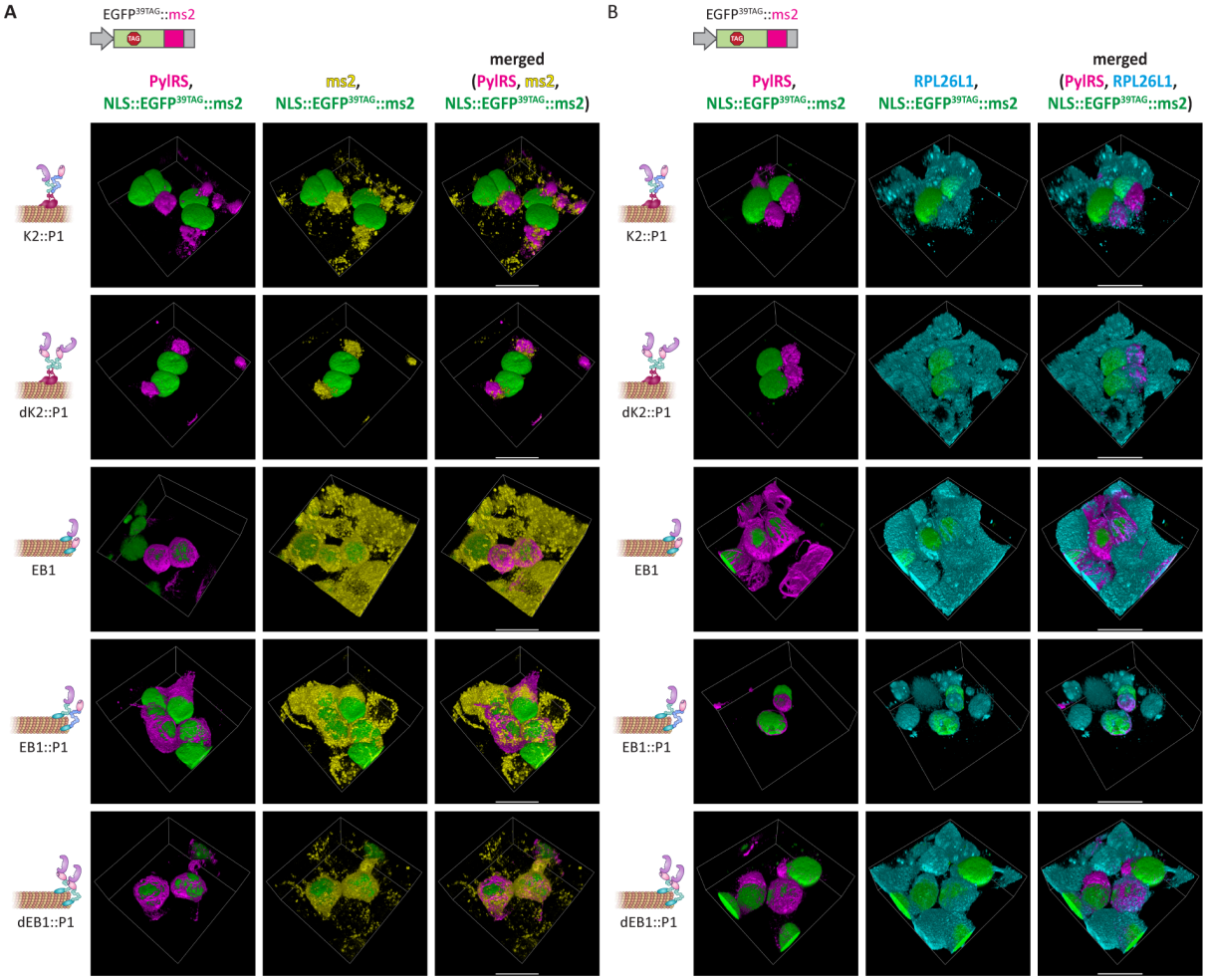


Figure 3-3. Synthetic organelles enrich mRNA-ms2 and even cellular ribosomes can access the synthetic organelles. (**A**) 3D reconstructions of IF and FISH stainings against PylRS (magenta) and mRNA::ms2 (yellow) in cells expressing $tRNA^{Pyl}$ and the indicated synthetic organelles as well as $EGFP^{39TAG}$::ms2 in presence of SCO. The EGFP signal is shown in green to identify amber suppressing cells. Additional images are shown in **Supplementary Figure 3-5**, confirming recruitment of mRNA::ms2 into synthetic organelles. (**B**) 3D reconstructions of IF staining against PylRS (magenta) and RPL26L1 (cyan) of HEK293T cells expressing the indicated synthetic organelles, $tRNA^{Pyl}$ and $EGFP^{39TAG}$::ms2 (shown in green). These stainings confirm that ribosomes can access synthetic organelles, indicating that orthogonal translation potentially can take place within them. Additional images confirming RPL26L1 co-localization are shown in **Supplementary Figure 3-6**. (Scale bars = 20 μm)

3.3 Discussion

Creating synthetic organelles *in vivo* is a powerful way to bottom-up equip cells with new functionalities. The synthetic organelles I present in this study all perform selective orthogonal translation of selected mRNAs with an expanded genetic code. To construct them I combined two organizing principles phase separation domains as well as targeting to microtubule-plus ends. Both active transport using motor domains as well as microtubule-end binding proteins

represent a powerful platform to construct OT organelles on the microtubule cytoskeleton. Although they assume substantially different shapes *in vivo* both effectively equip cells with a second, orthogonal genetic code, allowing to selectively and efficiently incorporate ncAAs exclusively into selected proteins. The kinesin based synthetic organelles exhibit a higher selectivity, which is in agreement with the higher local concentration of $tRNA^{Pyl}$ observed in FISH stainings. All EB1 based organelles distribute a bit broader throughout the cell, which can explain the comparatively lower selectivity. In contrast to the kinesin based organelles they however have no detectable influence on the microtubule cytoskeleton which might be beneficial for the viability of the host. Particularly, $OT^{dEB1::P1}$ shows a good selectivity towards targeted mRNA::ms2 with a less pronounced change in the microtubule cytoskeleton.

By directly fusing the effector proteins MCP and PylRS to the assembler moieties it is possible to increase selectivity particularly for EB1 based synthetic organelles, indicating a higher stringency of organelle assembly. This however comes at the cost of a reduction in GCE efficiency. Hence, depending on the application either the separate and more efficient ($OT^{K2::P1}$ or $OT^{EB1::P1}$) or the directly fused and more selective organelle constructs ($OT^{dK2::P1}$ or $OT^{dEB1::P1}$) might be more suitable.

In summary my results stress that the microtubule cytoskeleton can serve as a strong platform for synthetic organelle design *in vivo*. These organelles can perform tasks as complex as orthogonal translation despite their broadly distinct shapes. Hence, I expect that this platform has a huge potential for future organelle engineering approaches in living cells.

3.4 Material and methods

3.4.1 Cell culture

HEK293T cells (ATCC CRL-3216) were maintained in Dulbecco's modified Eagle's medium (DMEM, Gibco 41965-039) supplemented with 1% penicillin-streptomycin (Sigma-Aldrich P0781), 1% L-Glutamine (Sigma-Aldrich G7513), 1% sodium pyruvate (Life Technologies 11360), and 10% fetal bovine serum (FBS, Sigma-Aldrich F7524). Cells were cultured at 37 °C in a 5% CO_2 atmosphere and passaged every 2-3 days up to 20 passages. For transient transfection experiments cells were seeded 15-20 h prior to transfection at a density resulting in 70-80% confluency at the time of transfection. Flow cytometry was performed using 24-well plates with plastic bottom (Nunclon Delta Surface ThermoScientific). Immunofluorescence labeling and FISH were performed on 24-well plates with glass bottom (Greiner Bio-One) or four-well cell imaging coverglasses (Eppendorf).

3.4.2 Constructs and cloning

Dual-color reporter: The dual color fluorescent protein reporters was cloned in a pBI-CMV1 vector (Clontech 631630), with ms2 tagged fluorescence protein (mRNA) version in one multiple cloning site and ms2 free version in the other. $EGFP^{39TAG}$ or $mCherry^{190TAG}$ were used as N-terminal fusions with nuclear localization sequences (NLS) as described previously (*263*). $NLS::EGFP^{39TAG}$::ms2 reporter: $NLS::EGFP^{39TAG}$ was cloned with two copies of ms2-loops into the pBI-CMV1 vector as a reporter for successful amber suppression for imaging experiments (*263*).

OT organelle constructs: $tRNA^{Pyl}$ was cloned under the control of a human U6 promoter in a pUC57 plasmid, and all other constructs were under CMV promoters cloned in a pcDNA3.1 (Invitrogen V86020) vector. MCP protein was cloned from the addgene plasmid #31230, FUS from the Addgene plasmid #26374 and EWSR1 from the addgene plasmid #26377. In all FUS fusions, amino acids 1-478 were used, in all EWSR1 fusions, amino acids 1-628. In all PylRS fusions the previously reported efficient $NES::PylRS^{AF}$ (Y306A, Y384F) sequence was used (*64*). EB1 and $KIF16B_{1\text{-}400}$ were cloned from human cDNA and inserted into pcDNA3.1 via restriction cloning. EB1 and $KIF16B_{1\text{-}400}$ fusions with MCP, $PylRS^{AF}$, EWSR1::MCP, $FUS::PylRS^{AF}$, $FUS::MCP::PylRS^{AF}$ were assembled via Gibson assembly (*271*).

3.4.3 Transfections and used ncAAs

Transfections of HEK293T cells were performed with polyethyleneimine (PEI, Sigma-Aldrich 408727) using 3 μg PEI per 1 μg DNA (1200 ng total DNA for 24-well plates, 1680 ng DNA for four-well cell imaging cover glasses, diluted in DMEM without Phenol Red, Gibco11880-028).

For amber suppression system tests, cells were transfected at a ratio of a 1:1:1:1 with POI^{TAG} vectors, $tRNA^{Pyl}$, PylRS assembler fusions and MCP assembler fusions or assembler::MCP::PylRS fusion (a mock plasmid was used to keep the total plasmid amount constant). 4-6 hours after transfection the medium was swapped to fresh one containing ncAA supplemented with 25 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES, pH = 7.25). HEK293T cells were analyzed one day after transfection.

Stock and working solutions for the used ncAA was prepared as described in previous work (*272*). SCO (cyclooctyne lysine, SiChem SC-8000) was used at a final concentration of 250 μM. SCO is efficiently recognized by $PylRS^{AF}$ (Y306A, Y384F) (*63*).

3.4.4 Fluorescence flow cytometry

HEK293T cells were harvested 1 day after transfection by removing the medium, resuspending the cells in 1xPBS (phosphate buffered saline) and passing them through 100 μm nylon mesh. Data acquisition was performed in an LSRFortessa SORP Cell Analyzer (BD). Analysis was done using the FlowJo software (BD). Cells were first gated by cell type (using FSC-A x SSC-A parameters) and then by single cell (SSC-A x SSC-W). The workflow of cell gating is shown in **Supplementary Figure 3-7**. Each shown FFC plot is the sum of three independent biological replicates from which the mean and SEM were calculated. Lastly, fluorescence was acquired in the 488-530/30 channel for GFP signal and in the 561-610/20 channel for mCherry signal. Bar graphs were generated using IgorPro software.

3.4.5 IF labeling, FISH and confocal imaging

Immunofluorescence (IF): For immunolabeling experiments, cells were rinsed with DMEM without phenol red, fixed in 2% paraformaldehyde in 100 mM HEPES, 50 mM EGTA (ethylene glycol-bis-(2-aminoethylether)-N,N,N',N'-tetraacetic acid) 10 mM $MgSO_4$, 0.2% Triton-x-100 at room temperature (RT) for 10 minutes (*273*).

Subsequently, cells were rinsed with PBS and permeabilized with 0.5% Triton-x-100 solution in 1xPBS for 15 minutes at RT and rinsed twice prior to blocking. Samples were blocked in 3% bovine serum albumin (BSA, Sigma A7906) in 1xPBS for 90 minutes at RT, after which incubation with the primary antibody was done overnight at 4 °C in blocking solution [Ab_{PylRS} (1 μg/mL (*64*)), $Ab_{\alpha Tubulin}$ (Abcam ab89984, 2 μg/mL) and/or $Ab_{RPL26L1}$ (Abcam ab137046, 1:200)]. The next day, cells were rinsed with PBS and incubated with secondary antibody (Abcam ab175671; ThermoFisher A-21246, A-21449, at 2 μg/mL in blocking solution) for 60 minutes at RT. Then, cells were rinsed with PBS and fresh PBS was added for imaging.

Fluorescence in situ hybridization (FISH): FISH experiments were performed one day after transfection analogously to described previously (*64*). Briefly, the hybridization protocol was adapted for 24 well plates from Pierce et al. (*274*).

For imaging of $tRNA^{Pyl}$, the hybridization probe (5'-(Cy5)-CTAACCCGGCTGAAC-GGATTTAGAGTCCATTCGATC-3') was used at 0.5 μM (hybridization at 37 °C, overnight). For imaging of mRNA::ms2, the hybridization probe for ms2 (5'-(Alexa647)-CTGCAGACATGGGTGATCCTCATGTTT-TCTA) was used at 0.5 μM. After four washes with saline sodium citrate buffer (SSC) and one wash with Tris-HCl•NaCl buffer (TN), cells were incubated for 1.5 h at RT in 3% BSA prior to immunofluorescence labeling. Cells were incubated with primary antibodies for at 4 °C overnight, rinsed with PBS and incubated with

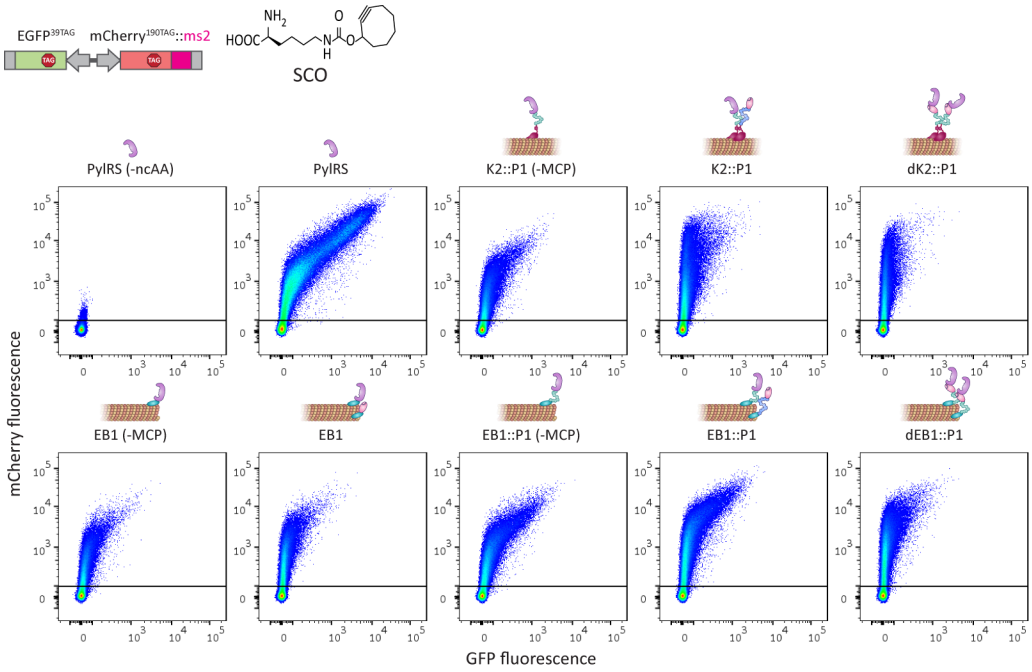
secondary antibodies for 1 h at RT (antibodies described above). Finally, cells were rinsed with PBS and fresh PBS was added for imaging.

Imaging: Confocal images were acquired on an Olympus Fluoroview FV3000 microscope using 405 nm (for Alexa405), 488 nm (GFP) and 640 nm (for Alexa 647 or Cy5) lasers for excitation with a 60x/1.40 oil immersion objective for acquisition. Images were processed using FIJI software.

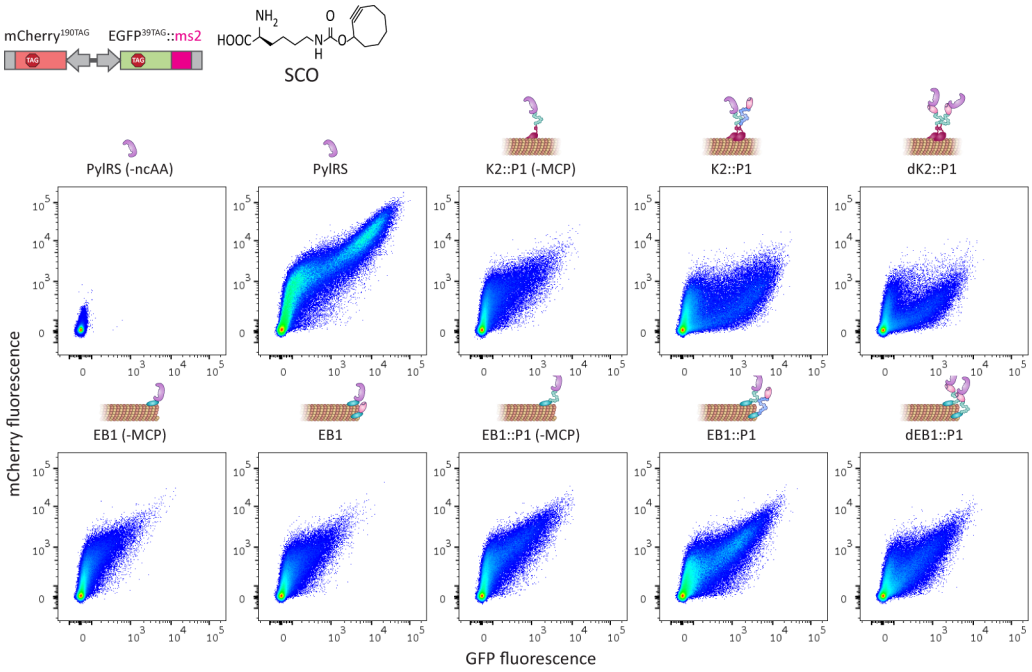
3D Reconstruction: 3D reconstructions were made using the arivis Vision4D software (arivis AG).

3.5 Supplementary figures

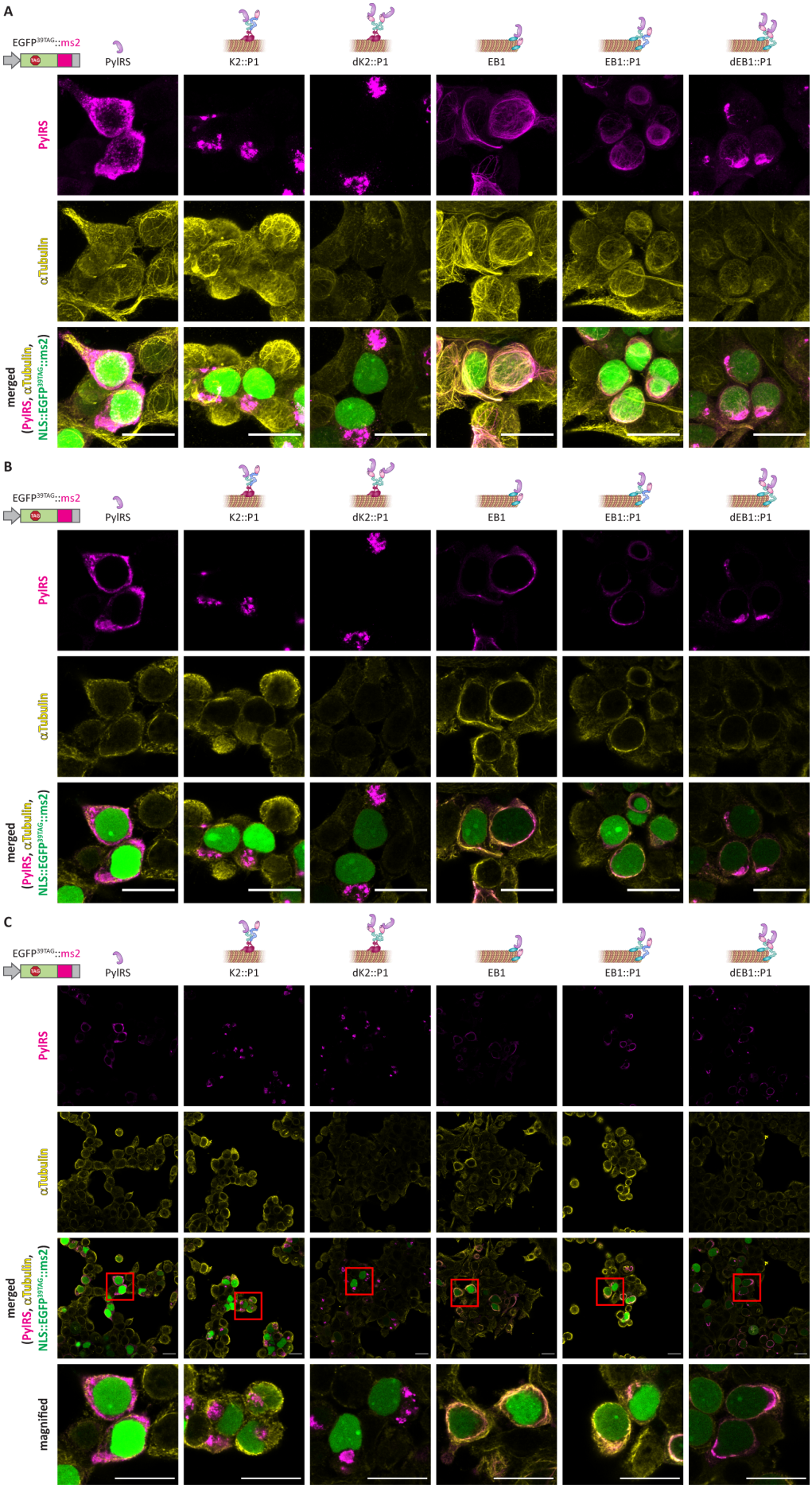
Supplementary Figure 3-1. Orthogonal translation in synthetic organelles depends on ms2-MCP based mRNA. Corresponding to **Figure 3-1**. FFC analysis of HEK293T cells co-expressing the indicated dual reporter, $tRNA^{Pyl}$ as well as the indicated GCE system in presence of the ncAA SCO (Upper left dot plot shows cytoplasmic PylRS system without ncAA, resulting in a lack of fluorescence). The dot plots show the sum of four independent experiments, axis indicate mCherry and GFP fluorescence in arbitrary units. If the MCP protein is omitted full-length mCherry production is not enhanced.



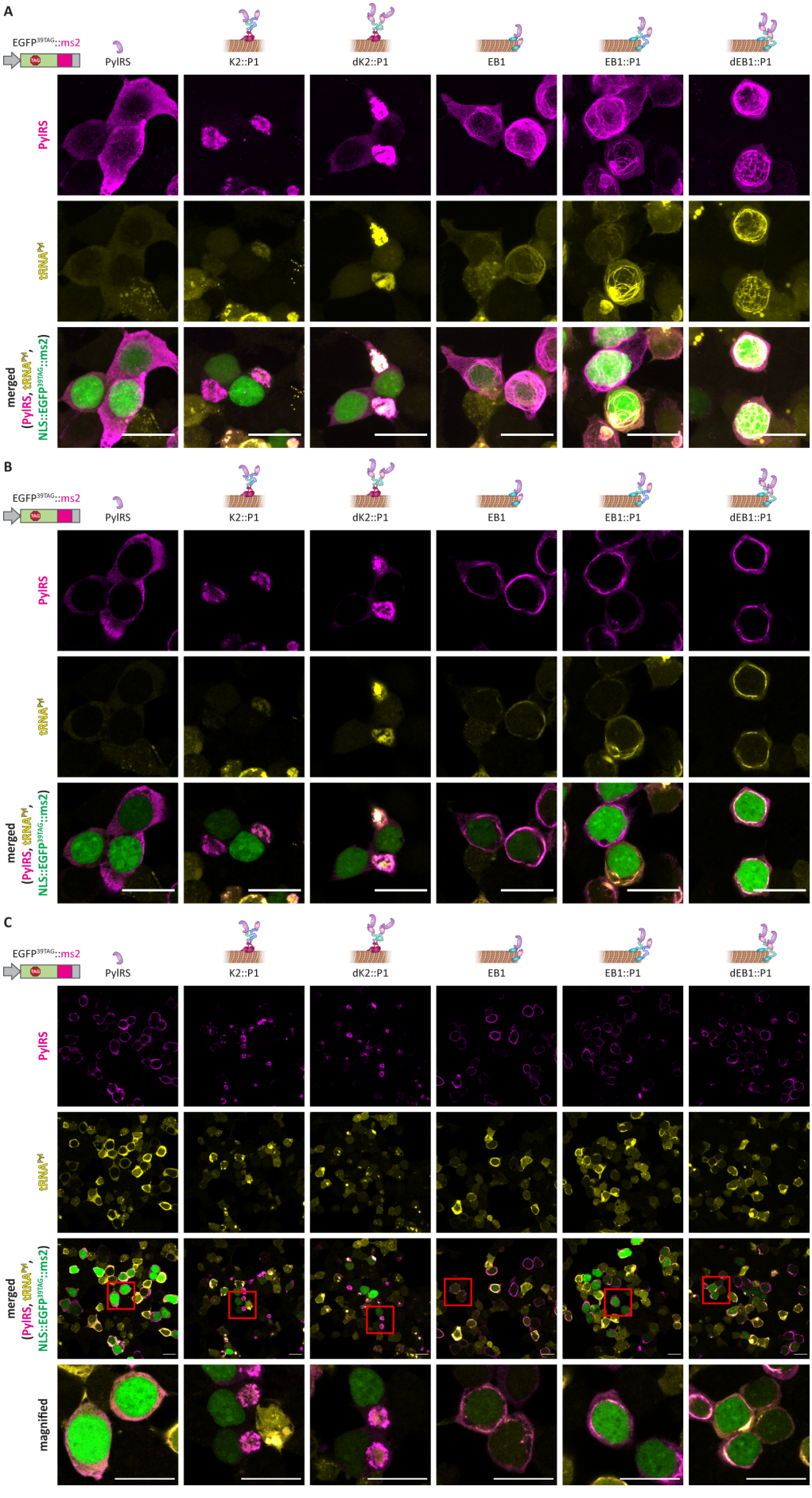
Supplementary Figure 3-2. Inverting the dual color FFC reporter expectedly inverts selectivity. FFC analysis of HEK293T cells co-expressing the indicated dual reporter, $tRNA^{Pyl}$ as well as the indicated GCE system in presence of the ncAA SCO (Upper left dot plot shows cytoplasmic PylRS system without ncAA, resulting in a lack of fluorescence). The dot plots show the sum of four independent experiments, axis indicate mCherry and GFP fluorescence in arbitrary units. When the ms2-loops are fused to the egfp mRNA instead of the mCherry mRNA OT organelles expectedly predominantly produce the targeted full-length EGFP protein. Also here omission of the MCP protein leads to a loss of selectivity.



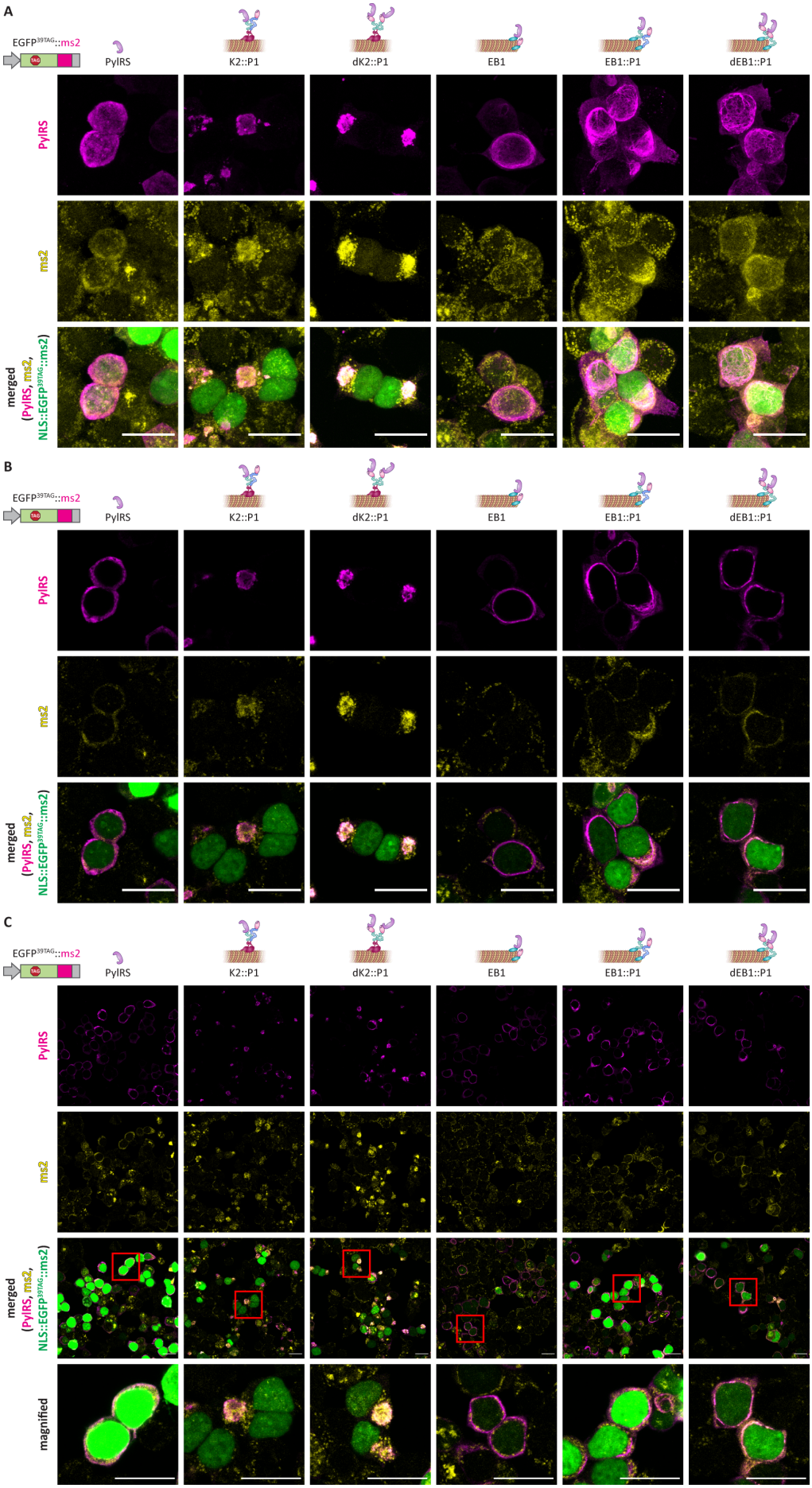
Supplementary Figure 3-3. Synthetic organelles form along microtubules. IF stainings against PylRS (magenta) and αTubulin (yellow) in cells expressing $tRNA^{Pyl}$ and the indicated synthetic organelles in presence of the ncAA SCO. Cells were additionally transfected with a construct encoding $EGFP^{39TAG}$::ms2 (green) to identify amber suppressing cells. (**A,B**) corresponding to **Figure 3-2 A**. In (A) a maximum intensity Z-projection is shown and in (B) one selected slice of the Z-stack (Top to bottom: PylRS (magenta), αTubulin (yellow), merged (NLS::$EGFP^{39TAG}$::ms2 in green). In (**C**) a larger field of view is shown highlighting that synthetic organelles appear similar in all transfected cells (Top to bottom: PylRS (magenta), αTubulin (yellow), merged (NLS::$EGFP^{39TAG}$::ms2 in green), and magnified corresponding to the red box in the merged view). In all cases microtubule associated synthetic organelles co-localize with α-Tubulin. (Scale bars = 20 μm)


A
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
αTubulin
merged (PylRS, αTubulin, NLS::EGFP397TAG::ms2)
B
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
αTubulin
merged (PylRS, αTubulin, NLS::EGFP397TAG::ms2)
C
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
αTubulin
merged (PylRS, αTubulin, NLS::EGFP397TAG::ms2)
magnified

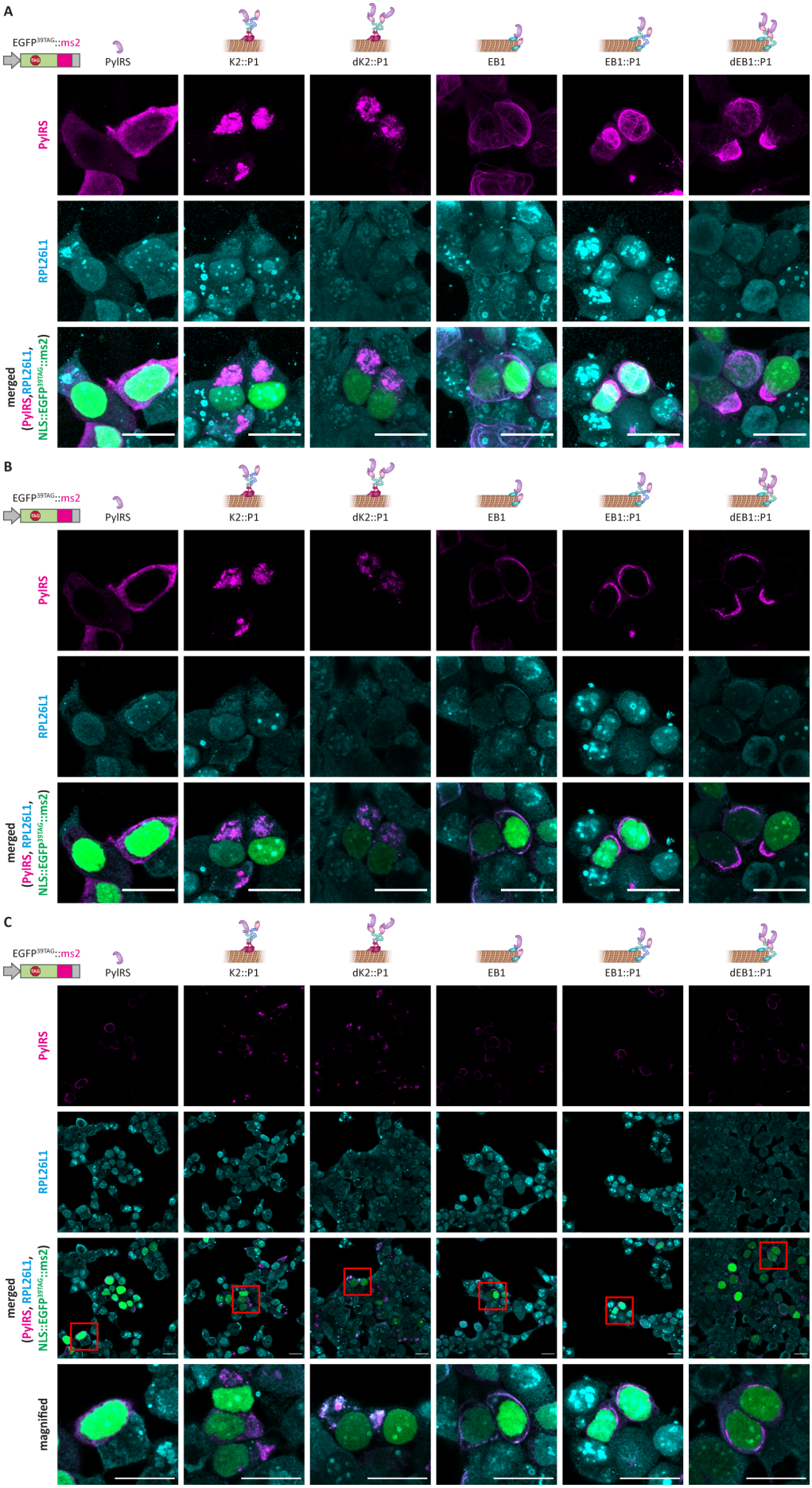
Supplementary Figure 3-4. Synthetic organelles highly enrich tRNA^{Pyl}. IF and FISH stainings against PylRS (magenta) and tRNA^{Pyl} (yellow) in cells expressing tRNA^{Pyl} and the indicated synthetic organelles in presence of the ncAA SCO. Cells were additionally transfected with a construct encoding EGFP^{39TAG}::ms2 (green) to identify amber suppressing cells. (**A,B**) corresponding to **Figure 3-2 B**. In (A) a maximum intensity Z-projection is shown and in (B) one selected slice of the Z-stack (Top to bottom: PylRS (magenta), tRNA^{Pyl} (yellow), merged (NLS::EGFP^{39TAG}::ms2 in green). (**C**) Shown is a larger field of view highlighting that OT organelles appear similar in all transfected cells (Top to bottom: PylRS (magenta), tRNA^{Pyl} (yellow), merged (NLS::EGFP^{39TAG}::ms2 in green), and magnified corresponding to the red box in the merged view). In all cases tRNA^{Pyl} is highly enriched in synthetic organelles, however it is a bit broader distributed in the EB1 based synthetic organelles, which can potentially explain the lower selectivity compared to the **K2** based synthetic organelles. (Scale bars = 20 μm)


A
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
tRNAPyl
merged (PylRS, tRNAPyl, NLS::EGFP397TAG::ms2)
B
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
tRNAPyl
merged (PylRS, tRNAPyl, NLS::EGFP397TAG::ms2)
C
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
tRNAPyl
merged (PylRS, tRNAPyl, NLS::EGFP397TAG::ms2)
magnified

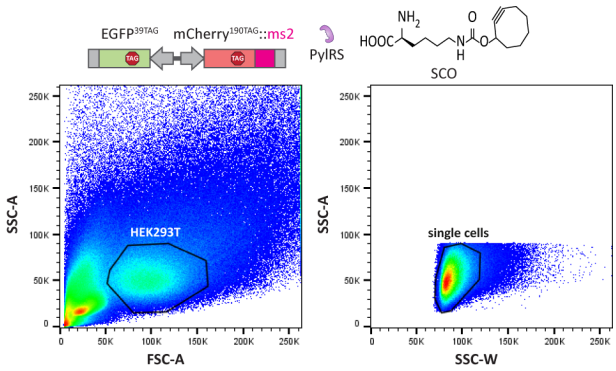
Supplementary Figure 3-5. Synthetic organelles enrich mRNA::ms2. IF and FISH stainings against PylRS (magenta) and mRNA::ms2 (yellow) in cells expressing $tRNA^{Pyl}$ and the indicated synthetic organelles in presence of the ncAA SCO. Cells were additionally transfected with a construct encoding $EGFP^{39TAG}$::ms2 (green) to identify amber suppressing cells. (**A,B**) corresponding to **Figure 3-3 A**. In (A) a maximum intensity Z-projection is shown and in (B) one selected slice of the Z-stack (Top to bottom: PylRS (magenta), mRNA::ms2 (yellow), merged (NLS::$EGFP^{39TAG}$::ms2 in green). (**C**) Shown is a larger field of view highlighting that OT organelles appear similar in all transfected cells (Top to bottom: PylRS (magenta), mRNA::ms2 (yellow), merged (NLS::$EGFP^{39TAG}$::ms2 in green), and magnified corresponding to the red box in the merged view). In all cases mRNA::ms2 is enriched in synthetic organelles. (Scale bars = 20 μm)


A
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
ms2
merged (PylRS, ms2, NLS::EGFP397TAG::ms2)
B
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
ms2
merged (PylRS, ms2, NLS::EGFP397TAG::ms2)
C
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
ms2
merged (PylRS, ms2, NLS::EGFP397TAG::ms2)
magnified

Supplementary Figure 3-6. Even ribosomes have access to OT organelles. IF stainings against PylRS (magenta) and RPL26L1 (cyan) in cells expressing $tRNA^{Pyl}$ and the indicated synthetic organelles in presence of the ncAA SCO. Cells were additionally transfected with a construct encoding $EGFP^{39TAG}$::ms2 (green) to identify amber suppressing cells. (**A,B**) corresponding to **Figure 3-3 B**. In (A) a maximum intensity Z-projection is shown and in (B) one selected slice of the Z-stack (Top to bottom: PylRS (magenta), RPL26L1 (cyan), merged (NLS::$EGFP^{39TAG}$::ms2 in green). In (**C**) a larger field of view (single plane) is shown highlighting that OT organelles are similar in all transfected cells. Note that as RPL26L1 is an endogenous protein it is also visible in untransfected cells. (Top to bottom: PylRS (magenta), RPL26L1 (cyan), merged (NLS::$EGFP^{39TAG}$::ms2 in green), and magnified corresponding to the red box in the merged view). In all cases mRNA::ms2 is enriched in synthetic organelles. (Scale bars = 20 μm)


A
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
RPL26L1
merged (PylRS,RPL26L1, NLS::EGFP397TAG::ms2)
B
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
RPL26L1
merged (PylRS, RPL26L1, NLS::EGFP397TAG::ms2)
C
EGFP397TAG::ms2
PylRS
K2::P1
dK2::P1
EB1
EB1::P1
dEB1::P1
PylRS
RPL26L1
merged (PylRS, RPL26L1, NLS::EGFP397TAG::ms2)
magnified

Supplementary Figure 3-7. Scatter values are used to sequentially identify single HEK293T cells in FFC. HEK293T cells were first identified based on side scatter area (SSC-A) and forward scatter area (FSC-A) values. Subsequently, single cells are identified using SSC-A and the side scatter width (SSC-W) parameters. In this figure one representative example is shown in which HEK293T cells were transfected with a cytoplasmic PylRS system, the indicated double reporter and $tRNA^{Pyl}$ in presence of the ncAA SCO. Cells that passed the first gate (left panel) are subsequently gated for single cells (right panel).



Chapter 4 Synthetic film-like organelles to create eukaryotic cells with three genetic codes

4.1 Introduction

Up to now I developed synthetic organelles based on phase separation and microtubule-tip association [**Appendix II**, **Chapter 3**, (*263*)]. These equip cells with mRNA specific GCE, solving the first mentioned major limitation of the technology. Complementary to this, the inducible GCE system solves potential long term toxicity issues [(*275*) and **Appendix IV**]. The two remaining challenges are the limited number of codons that can be reassigned without altering host functionality and canonical translation, and the dearth of orthogonal RS/tRNA pairs.

These are inherently difficult to address in eukaryotes. The scarcity of free codons has been addressed in *E. coli* by various means. These have been created using whole genome synthesis (*121*, *130*), unnatural base pairs [UBPs (*87*)], or orthogonal ribosomes (*276*), as also discussed in more detail in **Section 1.2.1.1**. However, transferring those approaches to eukaryotes is challenging, as eukaryotes have much larger and more complex genomes and use a different mechanism of mRNA recognition, and so far only for the UBP approach progress towards implementation in eukaryotic systems has been made (*89*).

Over recent decades, a number of orthogonal RS/tRNA pairs have been discovered that can be used with varying degrees of success in eukaryotes, thus addressing the third of the aforementioned limitations (*24*, *34*, *39*, *41*, *42*, *55*, *277*, *278*). More recently, advanced evolutionary and rational design strategies have been used to evolve multiple orthogonal versions of PylRS, which doubled this repertoire at least for *E. coli* (*53*, *54*). This demonstrates that using the classical approach of obtaining orthogonal RS/tRNA pairs through evolution can be quite powerful, but has its limitations as the discovery rate of new useful pairs is quite low. Therefore, novel alternative strategies for generating orthogonal tRNA/RS pairs would be extremely useful to complement the existing pairs in order to achieve the ribosomal synthesis of truly artificial polymers with a plethora of new functionalities.

I hypothesize that it should be possible to solve both limitations at once by introducing multiple independent OT organelles in one cell.

4.1.1 How to fit multiple OT organelles in one cell?

Unfortunately, the organelles I created so far fill up a substantial part of the cell and due to spatial restrictions, it is thus unlikely that it will be possible to fit two of these into one cell. Alternatively, I propose that by creating multiple OT film-like organelles I can overcome both challenges. Having multiple OT organelles should enable me to reuse the same stop codon to incorporate distinct ncAAs into different proteins *in vivo* without interfering with the host's other translational duties. Furthermore, I expect to prevent cross reactions between different tRNA/RS pairs by creating spatial orthogonality between them—a completely novel way to obtain orthogonal pairs.

To create multiple, mutually orthogonal OT organelles in one cell, their orthogonality should be threefold (**Figure 4-1A**). First, the organelles should form independently and without mixing. Second, each OT organelle needs to recruit a specific subset of mRNAs. Third, each RS variant in a respective organelle should selectively utilize a distinct ncAA. I aimed to assign one stop codon (amber, TAG) to a different ncAA in each organelle, hence the different full-length proteins of interest (POIs) are produced only in the presence of their programmed corresponding ncAA (**Figure 4-1B,C**).

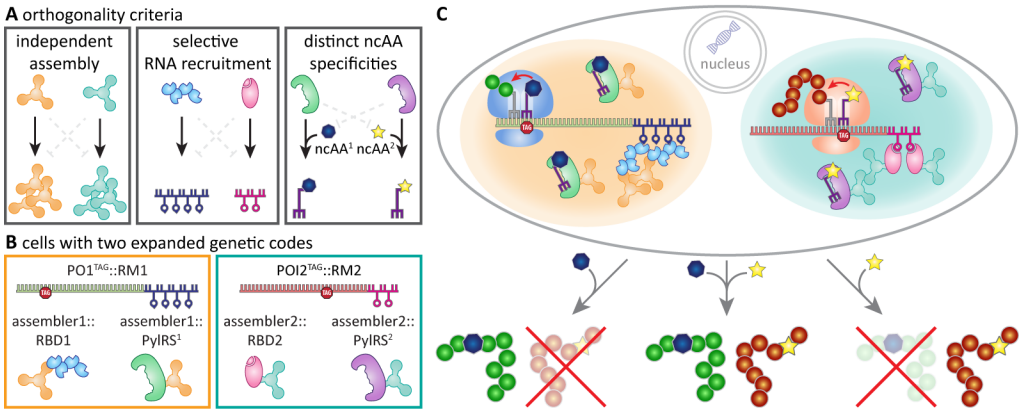


Figure 4-1 Multiple orthogonal translating organelles to equip cells with multiple orthogonal genetic codes. (**A**) Schematic overview of orthogonality criteria. In order to construct multiple orthogonal organelles, their assembly, RNA recruitment, and substrate specificity need to be orthogonal. (**B**, **C**) Combining two fully orthogonal OT organelles in one cell equips the host with a total of three genetic codes: the canonical one in the cytoplasm and two that are mutually orthogonal in the organelles. Only selected mRNAs are translated with the respective expanded genetic code and hence only selected proteins are equipped with the specific ncAAs.

4.1.2 Design of OT film-like organelles

In a three-step retrosynthetic approach I addressed each of the orthogonality criteria sequentially. As a first step, I needed to construct multiple spatially distinct OT organelles. The previously developed micron-sized synthetic organelles were similar in size to a whole cell nucleus and thus occupied a substantial portion of the cytoplasm (*263*). Therefore, it would be challenging to fit two of these organelles into one cell and ensure that they are immiscible.

I thus aimed to construct OT organelles as film-like membrane associated condensates on the cytoplasmic surfaces of cellular membranes. To achieve this, I combined a phase-separation domain with a membrane localization domain and fused them directly to the effector proteins (I use "::" to denote genetic fusion and "•" for co-expression). I used the proteins FUS and the EWSR1 as phase-separating proteins, for which I previously demonstrated that they can serve as assemblers (serving as a glue) for synthetic organelles *in vivo* (*263*). Also, I localized the condensates to the plasma membrane (PM) using the N-terminal domain of the rodent LCK tyrosine kinase (*279*), the outer mitochondrial membrane (OMM) with the N-terminal domain of human TOM20 (*280*), the Golgi membrane (GM) using the N-terminal domain of human EBAG9 (*281*) and the endoplasmic reticulum membrane (ERM) using the N-terminal domain of human cytochrome P450 2C1[(*282*), **Figure 4-2C–F**].

4.2 Results

4.2.1 OT film-like organelles on cellular membranes enable orthogonal translation

I first used the previously established dual-color reporter to assess orthogonal translation using FFC. In this reporter, both EGFP^{39TAG} and mCherry^{190TAG} are expressed independently but from one plasmid with amber stop codons at the indicated permissive sites. Only the mCherry mRNA is tagged with ms2-loops. In the case of cytoplasmic GCE there is no distinction possible and the amber codon is suppressed in both EGFP and mCherry, leading to a diagonal population in FFC. In contrast, if an OT organelle is working selectively, only mCherry is translated with an expanded genetic code, whereas the untargeted EGFP is terminated once the amber codon is encountered, leading to a vertical mCherry-positive population in FFC analysis (**Figure 4-2A**). Based on the flow cytometry data I calculated the fold change in selectivity and the relative efficiency of each system. As before, the fold change in selectivity is defined as the mean mCherry signal divided by the mean EGFP signal of a given system normalized to the respective ratio for the cytoplasmic GCE system (PylRS). Analogously, the relative efficiency is defined as the mean mCherry signal of a given system normalized to the mean mCherry signal of cytoplasmic PylRS.

Gratifyingly, all four membrane-localized film-like organelles facilitate highly selective orthogonal translation of the targeted mRNAs::ms2 (**Figure 4-2G**). I observed selectivities of 9- to 12-fold and achieved efficiencies of up to 45% (**Figure 4-2I**). The best of these systems compares favorably to the previously established OT organelle systems (*263*). Using IF stainings as well as FISH, I was able to confirm that also these organelles highly enrich

$tRNA^{Pyl}$, mRNA::ms2 as well as cellular ribosomes (**Supplementary Figure 4-1** and **Supplementary Figure 4-2**). I performed super-resolution microscopy (*2*) and was able to detect a high degree of localization of $tRNA^{Pyl}$ to a very thin layer and almost no $tRNA^{Pyl}$ in the remaining cytoplasm (**Supplementary Figure 4-3**).

To determine if the four membrane associated OT organelles can act independently of each other, I created constructs in which PylRS and MCP are not fused together but individually fused to both the respective membrane targeting domain and to a phase-separation-based assembler (FUS and EWSR1, respectively; **Supplementary Figure 4-4A**). I then co-expressed each membrane associated PylRS variant, either without an MCP construct or with the MCP protein targeted to either of the four membranes and assayed for selective orthogonal translation by FFC. I observed that the system targeted to the OMM functions completely independent of the others and that the PMP and ERMP systems are mutually independent. Both the ERMP and PMP systems appeared to mix with the GMP system—I observed translation of the targeted $mCherry^{190TAG}$::ms2 mRNA when PylRS is localized to the GM and the MCP protein is localized to either the PM or ERM and also *vice versa*, when MCP is targeted to the GM and the synthetase is targeted to the PM or ERM (**Supplementary Figure 4-4B**, **Supplementary Figure 4-5** for corresponding imaging).

For the MCP based GCE systems I also tested if I could omit the phase separation domain (**Supplementary Figure 4-6A**). I observed, that particularly for the plasma membrane targeted system also in the absence of phase separation selective orthogonal translation occurs. However, in all cases addition of a phase separating moiety aids translation selectivity. For all membrane-based systems I observed higher amber suppression yields upon addition of the phase separation domain to the PylRS construct, for the ERM and PM targeted system addition of EWSR1 to the MCP moiety seems to be less important (**Supplementary Figure 4-6B**).

4.2.2 λ_{N22} peptides enable to selectively translate boxB tagged mRNAs

The second step was to recruit distinct subsets of mRNAs into each respective organelle. To this end, I used λ_{N22} peptides that selectively bind boxB-loops (*283*). I elaborated on the dual-color reporter mentioned above by tagging the mCherry mRNA with four boxB-loops. I also fused four copies of λ_{N22} to the organelle components, and, similar to the MCP–ms2 system, observed highly selective orthogonal translation of the targeted mRNA with all four different membrane associated systems (**Figure 4-2H**). Interestingly, using the boxB–λ_{N22} approach, I observed even higher selectivity for all membrane associated OT organelles (12–17-fold changes) and achieved efficiencies of up to 80% (**Figure 4-2I**).

For experiments analogous to those with MCP–ms2-based synthetic organelles, I generated constructs in which the λ_{N22} peptides and PylRS were individually fused to the assembler components and performed alternating co-expression experiments. Again, I observed that the OMMP system functions independently of all other organelles and that the ERMP and PMP OT organelles are orthogonal, whereas the GMP system mixes with the ERMP and PMP systems (**Supplementary Figure 4-7**). I performed IF and FISH stainings to characterize the composition of the organelles and observed strong enrichment of PylRS, $tRNA^{Pyl}$, mRNA::boxB, and ribosomes (**Supplementary Figure 4-8**, **Supplementary Figure 4-9** and **Supplementary Figure 4-10**).

I also probed the orthogonality of these systems by co-expressing an OT organelle possessing MCP or λ_{N22} peptides with the boxB and ms2 reporter, respectively. I observed selective GCE of the targeted mRNA only for the correct combinations with the corresponding RNA recruitment domain (MCP–ms2 and λ_{N22}–boxB; **Supplementary Figure 4-11**), which demonstrates that I can selectively recruit RNAs into the OT organelles.

Overall, the ERMP, PMP, and OMMP systems are most efficient and selective for both MCP–ms2 and λ_{N22}–boxB approaches, hence I proceeded to develop these into fully orthogonal systems.

Having established multiple spatially orthogonal OT organelles within the cell, I proceeded with step 3: giving orthogonality to their ncAA selectivity. To this end, I utilized specific PylRS point mutants that incorporate different ncAAs. In particular, I used $PylRS^{AF}$ (Y306A, Y384F) (*64*, *65*) to incorporate a bulky lysine derivative and $PylRS^{AA}$ (N346A, C348A) (*284*) to encode a phenylalanine derivative (**Supplementary Figure 4-12**).

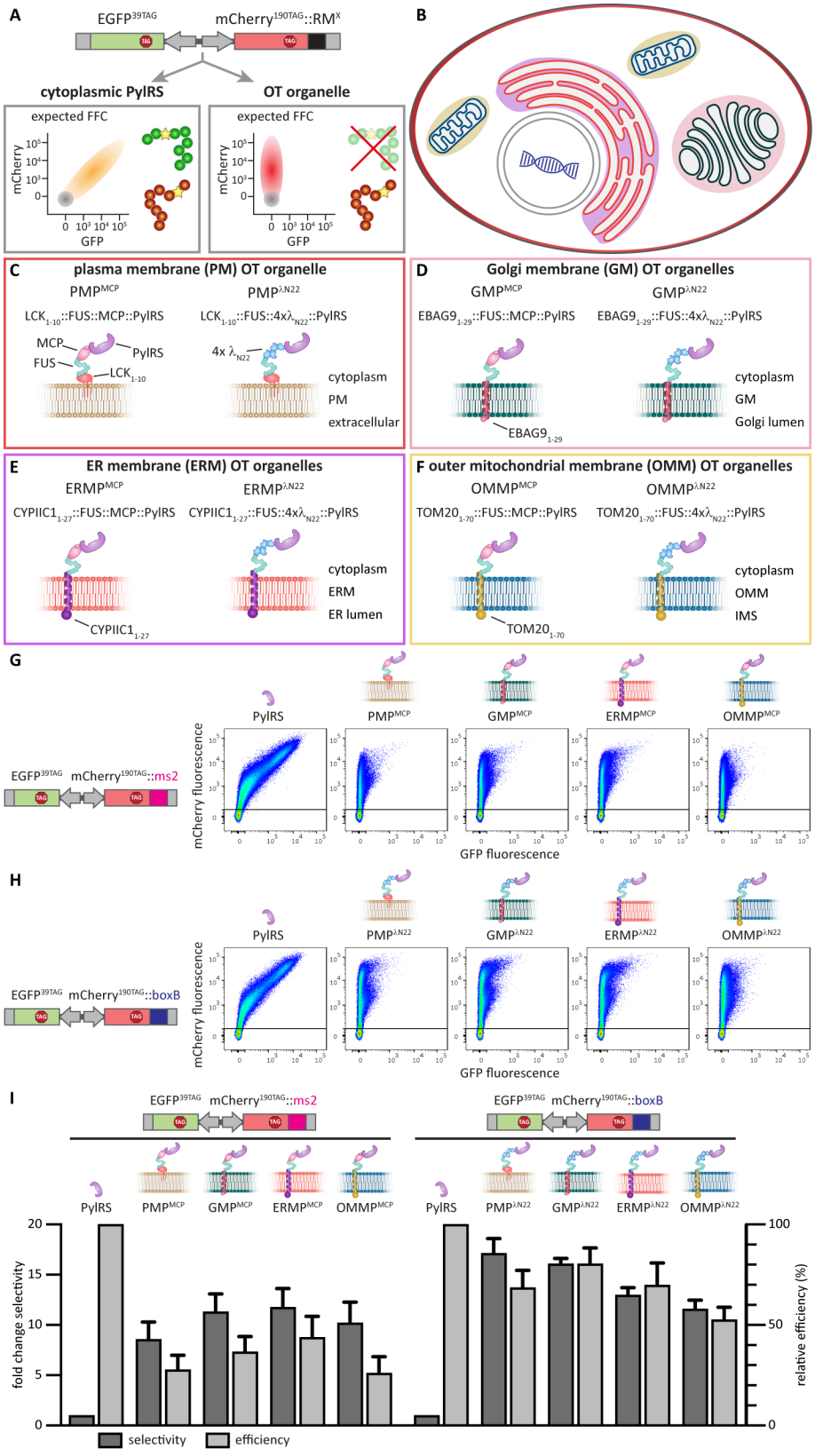


Figure 4-2. Synthetic organelles on intracellular membranes enable highly selective orthogonal translation. (**A**) Schematic representation of dual-color reporters. mRNAs encoding EGFP and mCherry with amber stop codons at permissive sites are expressed from one plasmid. The mCherry mRNA is tagged with specific RNA motifs (RM). In the case of cytoplasmic genetic code expansion, both full-length EGFP and mCherry should be produced, leading to an approximate diagonal in FFC, shown schematically in orange. If the OT organelle is working selectively, only mCherry will be produced, resulting in an exclusively mCherry-positive population in FFC (drawn as a red vertical population). Untransfected cells are represented as a gray circle. (**B**) Schematic

representation of membranes targeted for creating OT organelles. (**C–F**) Overview of synthetic organelle classes. To create OT organelles on the cytoplasmic surface of cellular membranes, PylRS and MCP or λ_{N22} peptides were fused to the phase-separation-based assembler fused in sarcoma (FUS) as well as to membrane targeting domains. In particular, the N-terminal fragment of rodent LCK tyrosine kinase ($LCK_{1–10}$) was used for plasma membrane (PM) targeting, the N-terminal domain of the human $EBAG9_{1–30}$ for Golgi membrane (GM) targeting, the N-terminal domain of human cytochrome P450 2C1 ($CYPIIC1_{1–27}$) for targeting to the membrane of the endoplasmic reticulum (ER), and the N-terminal domain of human $TOM20_{1–70}$ was used for targeting to the outer mitochondrial membrane (OMM). (**G, H**) FFC analysis of OT organelle selectivity. HEK293T cells expressing the indicated reporter and GCE systems as well as $tRNA^{Pyl}$ in the presence of the cyclooctyne-based lysine ncAA SCO. Shown is the sum of at least three independent experiments. (**I**) Bar plot corresponding to FFC experiments shown in G, H. The dark gray bars show the fold-change selectivity, defined as the ratio of the mean fluorescence intensities of mCherry versus EGFP, normalized to the respective cytoplasmic GCE control (PylRS). The light gray bars represent the relative efficiency as defined by the mean fluorescence intensity of mCherry for each condition divided by the cytoplasmic GCE control (PylRS). Bar graphs show the mean value of at least three independent experiments; error bars represent the standard deviation.

4.2.3 Two OT film-like organelles to selectively incorporate two distinct ncAAs

Finally, I needed to combine these three individual solutions to obtain fully orthogonal GCE systems. To assess orthogonality, I developed a dual-color fluorescence reporter in which $EGFP^{39TAG}$ mRNA is fused to boxB-loops and $mCherry^{190TAG}$ mRNA is fused to ms2-loops. With this reporter, a λ_{N22}-based synthetic organelle should exclusively produce full-length $EGFP^{39TAG}$, whereas an MCP-based synthetic organelle should only incorporate the ncAA into $mCherry^{190TAG}$. In particular I used a OMM-targeted, λ_{N22}-based systems with $PylRS^{AF}$ or $PylRS^{AA}$ ($OMMP^{\lambda N22,AF}$ and $OMMP^{\lambda N22,AA}$), a PM-targeted, MCP-based system with $PylRS^{AA}$ ($PMP^{MCP,AA}$), as well as a ER-targeted, MCP-based system with $PylRS^{AF}$ ($ERMP^{MCP,AF}$). Using the double tagged reporter, the expected selectivity can be observed for all tested systems. In presence of the MCP-based OT organelles an mCherry to GFP ratio of 12 ($PMP^{MCP,AA}$) and 16 ($ERMP^{MCP,AF}$) can be observed, while for the λ_{N22}-based systems EGFP is 8- and 11-fold higher expressed than mCherry ($OMMP^{\lambda N22,AA}$ and $OMMP^{\lambda N22,AF}$, **Supplementary Figure 4-13**). Unfortunately, if I simply combined $PMP^{MCP,AA}$ with $OMMP^{\lambda N22,AF}$, I observed substantial crosstalk between them, as evident from the production of both full-length EGFP and mCherry, even if only the ncAA for one system was added (**Supplementary Figure 4-14A,B**). Similarly, I observe a loss of selectivity when I combine $ERMP^{MCP,AF}$ with $OMMP^{\lambda N22,AA}$.

A possible explanation for this effect is that one or more of the organelle components are "leaky" and can shuttle between the two organelles. I demonstrated that the recruitment of mRNA is independent between the different recruiting systems, so that cannot cause loss of selectivity. Therefore, either the ncAA-charged tRNA shuffles between the organelles or PylRS is partially mislocalized. As I used PylRS variants with orthogonal substrate specificities, I deemed it unlikely that $tRNA^{Pyl}$ charged in one organelle would have affinity for the other organelle. Hence, I hypothesized that the most likely cause of selectivity loss would be the mixing of the PylRS component between different organelles. The organelle assembly

per se is unlikely to be the cause of this, as I specifically screened for independence of the organelle assembly and was able to demonstrate that the OMMP system is built orthogonal to the PMP and ERMP systems (**Supplementary Figure 4-4** and **Supplementary Figure 4-7**).
However, PylRS is known to be active as a dimer (*46*, *285*), and dimerization prior to proper membrane localization might mistarget some synthetase molecules. To investigate this, I performed IF stainings against two synthetase variants—one targeted to the PM or ERM and the other to the OMM—and I observed that indeed the PM- or ERM targeted synthetases mix with the OMM-targeted synthetases (**Supplementary Figure 4-14C**). This is particularly pronounced for the PM-targeted synthetase, of which a substantial fraction relocalizes to the OMM (**Supplementary Figure 4-14C**). To overcome this, I hypothesized that intramolecular homodimerization of PylRS variants might be kinetically favored over unwanted intermolecular heterodimerization. Hence, I fused two identical copies of PylRS directly with FUS as a linker and targeted them to the various membranes (**Figure 4-3A**).
Using this approach, I obtained two particularly well-working orthogonal systems that comprised an OMM-targeted λ_{N22}-based OT organelle combined with either a PM- or ERM-targeted MCP-based system. The OT organelle at the OMM contained either the PylRS^{AF} or the PylRS^{AA} mutant (termed OMM-λ_{N22}-OT^{v2,AF} or OMM-λ_{N22}-OT^{v2,AA}); the OT organelle at the PM was constructed with PylRS^{AA} (PM-MCP-OT^{v2,AA}); the one at the ERM used PylRS^{AF} (ERM-MCP-OT^{v2,AF}).
I first used the dual-color reporter in which EGFP^{39TAG} is fused to boxB-loops and mCherry^{190TAG} is fused to ms2-loops and observed that using the λ_{N22}-based, OMM-targeted systems full-length EGFP is predominantly produced (EGFP/mCherry ratios of 18 (OMM-λ_{N22}-OT^{v2,AF}) and 8 OMM-λ_{N22}-OT^{v2,AA}, normalized to a cytoplasmic GCE system, **Supplementary Figure 4-15**). Contrary to this, using the MCP-based systems targeted to the ERM or PM, mCherry is predominantly produced (mCherry/GFP ratios of 36 (ERM-MCP-OT^{v2,AF}) and 17 (PM-MCP-OT^{v2,AA}, **Supplementary Figure 4-15**).

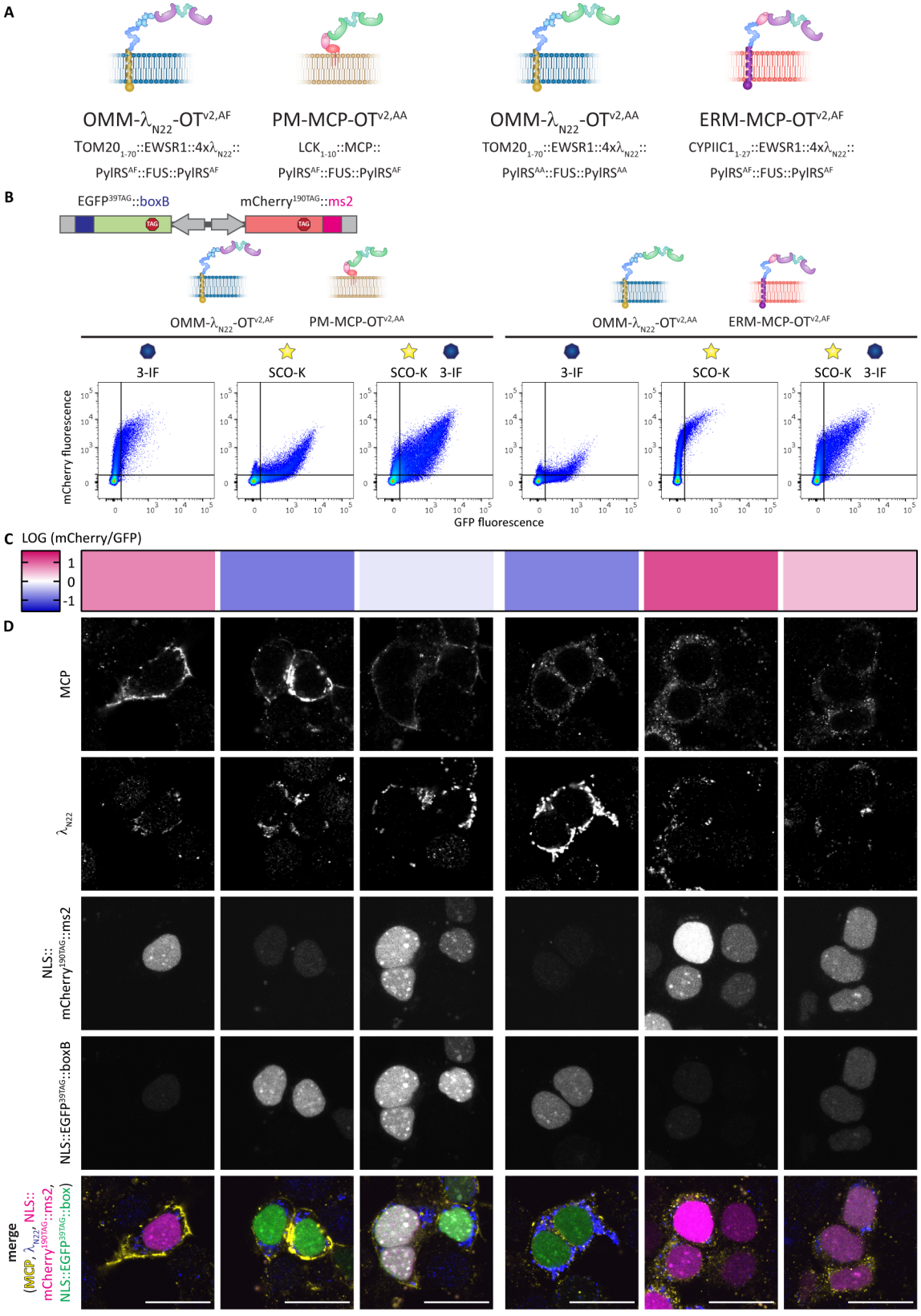


Figure 4-3. Fusing two copies of PylRS creates fully orthogonal systems. (**A**) Schematic overview of the second generation of OT film-like organelles. Constitutive PylRS dimers (PylRS^{AF} or PylRS^{AA}) are fused using FUS and they are subsequently targeted to the respective membrane localizations by fusing them to TOM20$_{1\text{-}70}$::EWSR1::λ_{N22}, LCK$_{1\text{-}10}$::MCP or CYPIIC1$_{1\text{-}27}$::EWSR1::MCP. (**B**) HEK293T cells expressing the double tagged reporter (EGFP^{39TAG}::boxB, mCherry^{190TAG}::ms2) together with tRNA^{Pyl} and the optimized GCE systems (400 ng of reporter and tRNA plasmid, co-transfected with 100 ng of the bicistronic OT organelle construct and 300 ng mock plasmid). Experiments were performed in the presence of the indicated ncAAs. Shown

is the sum of three independent experiments. When OMM-λ_{N22}-OT^{v2,AF} is combined with PM-MCP-OT^{v2,AA} full-length EGFP^{39TAG} expression can be observed in presence of the ncAA SCO, full-length mCherry^{190TAG} expression in the presence of the ncAA 3-IF and production of both when both ncAAs are present. Similarly, when OMM-λ_{N22}-OT^{v2,AA} is combined with ERM-MCP-OT^{v2,AF}, I expectedly observe the inverse behavior, such that EGFP^{39TAG} is produced in presence of 3-IF, mCherry^{190TAG} in presence of SCO, and both proteins in presence of both ncAAs. (**C**) Heatmap showing the ratio of the mCherry signal GFP signal corresponding to the conditions in (B), normalized to a cytoplasmic GCE system (see **Supplementary Figure 4-13**). Red values indicate a positive, logarithmic mCherry to GFP ratio, blue values indicate a negative logarithmic value, thus an excess of GFP over mCherry. (**D**) IF staining against MCP and λ_{N22} corresponding to the conditions in (B). For all systems the optimized film-like OT organelles can be observed at their expected localization. Top to bottom: MCP, λ_{N22}, mCherry, EGFP and merged (MCP in yellow, λ_{N22} in blue, mCherry in magenta, EGFP in green). (Scale bars = 20 μm)

I then tested the combined systems, in which always one optimized OMM-targeted system is combined with a corresponding optimized ERM- or PM-targeted system. Using the dual-color reporter with both RMs, I observed that in the presence of the ncAA for the OMM-targeted, λ_{N22}-based system, EGFP is selectively produced [EGFP/mCherry ratio of 6-7 (normalized to a cytoplasmic GCE system)], whereas if the ncAA specific to the PM- or ERM-targeted MCP-based systems were used, mCherry fluorescence is predominated (mCherry/EGFP ratios of 6 and 14, respectively). Only if both ncAAs were present I observed production of both EGFP and mCherry (**Figure 4-3B,C**), corresponding to mCherry/EGFP ratios of 0.7 (for the combination of OMM-λ_{N22}-OT^{v2,AF} and PM-MCP-OT^{v2,AA}) and 2.7 (for OMM-λ_{N22}-OT^{v2,AA} and ERM-MCP-OT^{v2,AF}). I performed IF stainings for these systems, and observed that the internally linked constructs do not mix and can exclusively be observed at their intended localization (**Figure 4-3D**).

To further corroborate that this new double orthogonal system was working selectively, I developed an imaging-based reporter in which the nucleoporin 153 (Nup153) is fused to EGFP and boxB-loops and the histone H2B is fused to mCherry and ms2-loops; both constructs contain an amber codon at a permissive site. I co-expressed this reporter with the different optimized organelles and analyzed cells using fluorescence microscopy. For all tested constructs I observed that full-length Nup153::EGFP^{149TAG} was produced only in the presence of the ncAA specific for the λ_{N22}-based OT organelle; full-length H2B::mCherry^{190TAG} was produced only in the presence of the ncAA for the MCP-based organelle; and only in the presence of both ncAAs both proteins were produced (**Figure 4-4**). Importantly, these proteins can be observed at their expected subcellular localization (Nup153 at the nuclear rim, H2B in the nucleus).

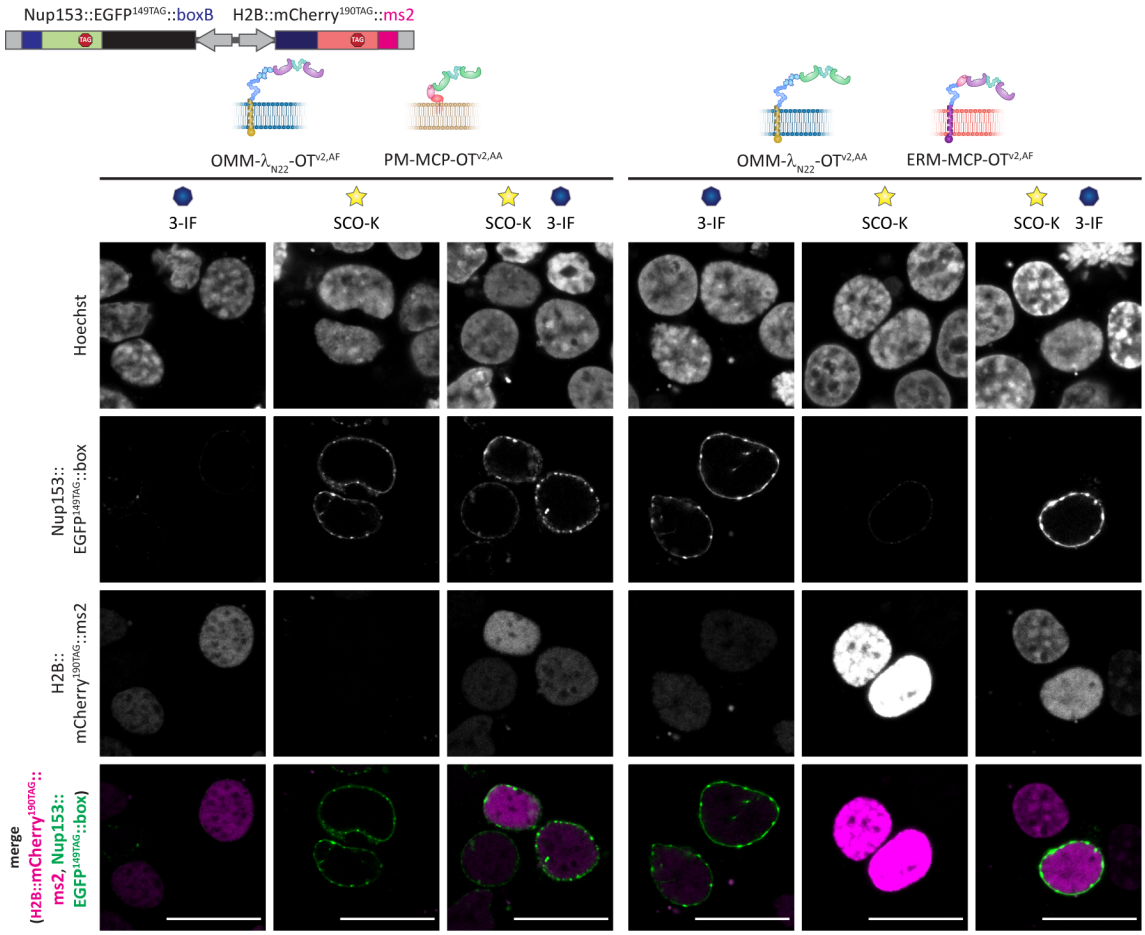


Figure 4-4. OT film-like organelles enable the selective expression of histones and nucleoporins. HEK293T cells expressing the double tagged reporter (Nup153::EGFP^{149TAG}::boxB, H2B::mCherry^{190TAG}::ms2) together with tRNA^{Pyl} and the optimized GCE systems (400 ng of reporter and tRNA plasmid, co-transfected with 400 ng of the bicistronic OT organelle construct). Experiments were performed in the presence of the indicated ncAAs. When OMM-λ_{N22}-OT^{v2,AF} is combined with PM-MCP-OT^{v2,AA} full-length Nup153::EGFP^{149TAG} expression can be observed in presence of SCO, full-length H2B::mCherry^{190TAG} expression in the presence of 3-IF and production of both when both ncAAs are present. When OMM-λ_{N22}-OT^{v2,AA} is combined with ERM-MCP-OT^{v2,AF}, I observed that Nup153::EGFP^{149TAG} is produced in presence of 3-IF, H2B::mCherry^{190TAG} in presence of SCO, and both proteins in presence of both ncAAs. (Scale bars = 20 μm)

4.3 Discussion

It has long been known that cells use their membranes to organize various functionalities and to process signaling events. By analogy, membraneless organelles have been identified as a means to organize cellular functionality (*136*, *286*, *287*). Only more recently has the combination of both come into focus, and membrane associated condensates have been found to play roles in the organization of receptor signaling and membrane associated actin polymerization (*233*, *237*, *238*, *240*, *243*).

In this chapter I show that this narrow interface can also be exploited as a powerful bioengineering platform that straightforwardly allows modification of a process as complex as protein translation, which requires hundreds of factors to work together. Using this platform, I can effectively equip a cell with multiple orthogonal genetic codes by developing a

revolutionary way to create independent and functionally orthogonal enzymes in one cell by spatial organization. The organelles I built are capable of enforcing functional orthogonality even between enzymes that only differ in their substrate binding pocket. This demonstrates how powerful the creation of unique chemical microenvironments in one cell can be to equip a cell with new functionalities, and I apply these systems to specifically modify different proteins with different ncAAs.

This revolutionary technology of OT film-like organelles has not only immense implications for the future design of artificial, functional organelles *in vivo* but this new technology shows how powerful spatial organization principles are for creating novel enzyme functionalities. Membrane associated phase separation has now been observed in a few studies. Thin-film formation dramatically extends cellular signaling possibilities beyond what can be achieved by membrane binding or phase separation on its own. In this case, the language of protein synthesis can be changed within a few nanometers. This enables me to create new subtypes of ribosomal activity in the cytoplasm going beyond the known groups of ER-associated ribosomes and cytoplasmic ribosomes. To what extent translation might be different in other cases in which ribosomes have nonrandom distribution in the cytoplasm might be an interesting subject for future studies.

It is possible that natural biological systems might have used similar principles during evolution of ever more complex functionalities. Hence, these new insights could have broad implications for understanding how the vast complexity of living matter arose.

4.4 Material and methods

4.4.1 Cell culture

HEK293T cells (ATCC CRL-3216) were maintained in DMEM (Life Technologies 41965-039) supplemented with 1% penicillin-streptomycin (Sigma P0781), 1% L-Glutamine (Sigma G7513), 1% sodium pyruvate (Life Technologies 11360), and 10% FBS (Sigma F7524). Cells were cultured at 37 °C in a 5% CO_2 atmosphere and passaged every 2-3 days up to 20 passages. In all cases, cells were seeded 15-20 h prior to transfection at a density resulting in 70-80% confluency at the time of transfection. Flow cytometry was performed using 24-well plates with plastic bottom (Nunclon Delta Surface ThermoScientific). Immunofluorescence labeling and FISH were performed on 24-well plates with glass bottom (Greiner Bio-One), four-well chambered Lab-Tek #1.0 borosilicate coverglass (ThermoFisher), four-well cell imaging coverglass (Eppendorf) or glass bottom 4 well μ-slides (ibidi).

4.4.2 Constructs, cloning and mutagenesis

Dual-color reporters: The dual fluorescent protein reporters were cloned in a pBI-CMV1 vector (Clontech 631630) as described for the ms2 based reporters before (*263*). Similar reports were constructed with four boxB-loops attached to the mCherry mRNA (*283*).

Double RNA-loops tagged reports: One multiple cloning site (MCS) of pBI-CMV1 was tagged with four boxB-loops and the other with two ms2-loops. Subsequently, mCherry^{190TAG} and EGFP^{39TAG} were inserted into the MCSs. For imaging experiments similar double tagged pBIs were used, in which only the EGFP^{39TAG} gene was inserted into the ms2 or boxB-loops tagged MCS. EGFP^{39TAG} or mCherry^{190TAG} were used as N-terminal fusions with nuclear localization sequences (NLS).

Analogously, also Nup153::EGFP^{149TAG} was inserted into one MCS via restriction cloning, while and H2B:mCherry^{190TAG} was inserted into the other via Gibson assembly (*271*). The template for H2B was a kind gift from the Ellenberg laboratory.

The iRFP::GFP^{39TAG} construct, used in **Supplementary Figure 4-12** was published before (*64*).

OT organelle constructs: tRNA^{Pyl} was cloned under the control of a human U6 promoter, and all other constructs were under CMV promoters cloned in a pcDNA3.1 vector (Invitrogen V86020) as described previously (*263*). The template for the 4xλ_{N22} peptides was a kind gift from the Ellenberg laboratory (*283*). They were generally used with an N-terminally fused c-Myc tag (EQKLISEEDL), MCP constructs were fused to an HA-tag (YPYDVPDYA) unless otherwise indicated.

The MCP protein was cloned from the addgene plasmid #31230, FUS from the addgene plasmid #26374 and EWSR1 from the addgene plasmid #26377. In all FUS fusions, amino acids 1-478 were used, in all EWSR1 fusions, amino acids 1-628. In all PylRS fusions the previously reported efficient NES::PylRS^{AF} (Y306A, Y384F) was used. Alternatively when specifically indicated NES::PylRS^{AA} (N346A, C348A) was used.

The murine LCK$_{1-10}$ sequence [MGCVCSSNPE, (*279*)], was fused to PylRS^{AF}, FUS::PylRS^{AF}, MCP or EWSR1::MCP (*263*) via restriction cloning and to EWSR1::4xλ_{N22} via Gibson assembly. To create LCK$_{1-10}$::FUS::MCP::PylRS^{AF} and LCK$_{1-10}$::FUS::4xλ_{N22}::PylRS^{AF} MCP and 4xλ_{N22} respectively were inserted into LCK$_{1-10:}$:FUS::PylRS^{AF} via restriction cloning.

The human CYPIIC1$_{1-27}$ sequence [MDPVVVLGLCLSCLLLLSLWKQSYGGG, (*282*), CYPIIC1$_{1-27}$] was fused to PylRS^{AF}, FUS::PylRS^{AF}, MCP or EWSR1::MCP and EWSR1::4xλ_{N22} via restriction cloning. To create CYPIIC1$_{1-27}$::FUS::MCP::PylRS^{AF} and

CYPIIC1$_{1\text{-}27}$::FUS::4xλ_{N22}::PylRS^{AF}, MCP and 4xλ_{N22} respectively were inserted into CYPIIC1$_{1\text{-}27}$::FUS::PylRS^{AF} via restriction cloning.

TOM20$_{1\text{-}70}$ was cloned from human cDNA and inserted into pcDNA3.1 via restriction cloning. It was subsequently fused to PylRS^{AF}, FUS::PylRS^{AF}, EWSR1::MCP, EWSR1::4xλ_{N22}, MCP (without an HA-tag) and FUS::4xλ_{N22}::PylRS^{AF} via Gibson assembly. To create TOM20$_{1\text{-}70}$::FUS::MCP::PylRS^{AF} MCP was inserted into a TOM20$_{1\text{-}70}$::FUS::PylRS^{AF} containing construct via restriction cloning.

Full-length EBAG9 was cloned from human cDNA and inserted into pcDNA3.1 via restriction cloning. Subsequently EBAG9$_{1\text{-}29}$ was fused to PylRS^{AF}, FUS::PylRS^{AF}, MCP, EWSR1::MCP or EWSR1::4xλ_{N22} via Gibson assembly. To create EBAG9$_{1\text{-}29}$::FUS::MCP::PylRS^{AF} and EBAG9$_{1\text{-}29}$::FUS::4xλ_{N22}::PylRS^{AF}, MCP and 4xλ_{N22} were inserted into the EBAG9$_{1\text{-}29}$::FUS::PylRS^{AF} construct via restriction cloning.

To obtain LCK$_{1\text{-}10}$::FUS::MCP::PylRS^{AA} and TOM20$_{1\text{-}70}$::FUS::4xλ_{N22}::PylRS^{AA}, the PylRS variant was exchanged in the corresponding PylRS^{AF} based constructs via restriction cloning.

Internally linked PylRS dimers: First, OMM-λ_{N22}-OT^{v2,AF} (TOM20$_{1\text{-}70}$::EWSR1::4xλ_{N22}::PylRS^{AF}::FUS::PylRS^{AF}), OMM-λ_{N22}-OT^{v2,AA} (TOM20$_{1\text{-}70}$::EWSR1::4xλ_{N22}::PylRS^{AA}::FUS::PylRS^{AA}) and ERM-MCP-OT^{v2,AF} (CYPIIC1$_{1\text{-}27}$::EWSR1::MCP::PylRS^{AF}::FUS::PylRS^{AF}) were assembled via Gibson assembly in pcDNA3.1 vectors. PM-MCP-OT^{v2,AA} (LCK$_{1\text{-}10}$::MCP::PylRS^{AA}::FUS::PylRS^{AA}) was subsequently created via restriction cloning. Then OMM-λ_{N22}-OT^{v2,AF}, OMM-λ_{N22}-OT^{v2,AA}, ERM-MCP-OT^{v2,AF} and OMM-λ_{N22}-OT^{v2,AA} were inserted into one of the two multiple cloning sites in a bicistronic amber suppression vector via restriction cloning [pBI-CMV1with tRNA^{Pyl} under the control of a hU6 promoter as described previously (*263*)]. For bicistronic constructs ERM-MCP-OT^{v2,AF} was inserted into the free multiple cloning site of the pBI-CMV1 with OMM-λ_{N22}-OT^{v2,AA} and PM-MCP-OT^{v2,AA} into the one of the OMM-λ_{N22}-OT^{v2,AF}.

4.4.3 Transfections and used ncAAs

Transfections of HEK293T cells were performed using 3 μg PEI (Sigma-Aldrich 408727) per 1 μg DNA (1200 ng total DNA for 24-well plates and for the ms2 FISH experiment in four-well imaging cover glasses (Eppendorf); 1680 ng DNA for four-well cell imaging cover glasses (Eppendorf), or 4 well μ-slides (ibidi), diluted in DMEM without Phenol Red, Gibco11880-028).

For amber suppression system tests, cells were transfected at a ratio of a 1:1:1:1 with POI^{TAG} vectors, tRNA^{Pyl}, and OT organelle plasmids (a mock plasmid was used to adjust the total

plasmid amount to be constant). 4-6 hours after transfection the medium was exchanged to fresh one containing ncAA supplemented with 25 mM HEPES (pH = 7.25). Cells were analyzed one day after transfection.

For dual organelle FFC experiments, cells were transfected at a ratio of 4:4:1:3 with the dual tagged reporter, $tRNA^{Pyl}$, bicistronic OT organelle vector and a mock plasmid. Alternatively, they were transfected at a ratio of 4:4:1:1:2 with the dual tagged reporter, $tRNA^{Pyl}$, and two individual OT organelle vectors and a mock plasmid.

For dual organelle imaging experiments, cells were transfected at a ratio of 1:1:1 with the dual tagged reporter with H2B::mCherry and Nup153::EGFP, $tRNA^{Pyl}$ and a bicistronic OT organelle vector.

Stock solutions for all the used ncAAs were prepared as described in previous work (*288*). SCO (cyclooctyne lysine, SiChem SC-8000) was used at a final concentration of 250 μM; 3-iodophenylalanine (Chem-Impex International Inc., 14352) was used at a final concentration of 1 mM. SCO is efficiently recognized by $PylRS^{AF}$ (Y306A, Y384F) while 3-iodophenylalanine is recognized by $PylRS^{AA}$ (C346A, N348A).

4.4.4 Fluorescence flow cytometry

HEK293T cells were harvested after 1 day by removing the medium, resuspending the cells in 1xPBS and passing them through 100 μm nylon mesh. Transfections for flow cytometry were performed with four plasmids (reporter plasmid, $tRNA^{Pyl}$, the wanted version of synthetase and an MCP fusion or a mock plasmid) at a 1:1:1:1 ratio with 1.2 μg total DNA. Medium was exchanged for fresh medium containing ncAA 4-6 h post-transfection and left until the time of harvesting.

Data acquisition was performed in an LSRFortessa SORP Cell Analyzer (BD). Analysis was done using the FlowJo software (FlowJo). Cells were first gated by cell type (using FSC-A x SSC-A parameters) and then by single cell (SSC-A x SSC-W). The workflow of cell gating is shown in **Supplementary Figure 4-16**. Lastly, fluorescence was acquired in the 488-530/30 channel for GFP signal, in the 561-610/20 channel for mCherry signal and in the 640-730/45 channel for the iRFP signal. Bar plots were generated using Prism 8 software (GraphPad).

4.4.5 IF labeling, FISH and confocal imaging

IF: For immunolabeling experiments, cells were rinsed with PBS, fixed in 2% paraformaldehyde in 1xPBS at room temperature (RT) for 10 minutes and rinsed with PBS. Subsequently, cells were permeabilized with 0.5% Triton-x-100 solution in 1xPBS for 15 minutes at RT and rinsed twice prior to blocking. Samples were blocked in 3% BSA in

1xPBS for 90 minutes at RT, after which incubation with the primary antibody {Ab_{PylRS} [1 μg/mL (*65*)], Ab_{HA} (Sigma H9658, 1:2000), Ab_{Myc} (Abcam ab9106, 0.5 μg/mL) and/or $Ab_{RPL26L1}$ (Abcam ab137046, 1:200)} was done overnight at 4 °C in blocking solution. The next day, cells were rinsed with PBS and incubated with secondary antibody (Abcam ab175671 or ab175673; ThermoFisher A-21246, A32728 and/or A-31553, at 2 μg/mL in blocking solution) for 60 minutes at RT. Finally, cells were rinsed twice with PBS.

If only DNA was stained, cells were fixed and permeabilized as described above and subsequently stained with Hoechst 33342 (Sigma-Aldrich B2261) at 1 mg/mL in 1xPBS for 10 min at RT. Then, cells were rinsed with PBS and fresh PBS was added for imaging.

FISH: FISH experiments were performed one day after transfection analogously to described previously (*64*). Briefly, the hybridization protocol was adapted from Pierce et al. (*274*).

For imaging of only $tRNA^{Pyl,}$ the hybridization probe (5'-(Alexa647)-CTAACCCG-GCTGAACGGATTTAGAGTCCATTCGATC-3') was used at 0.5 μM. After four washes with SSC and one wash with Tris-HCl•NaCl buffer (TN), cells were incubated for 1 h at RT in 3% BSA prior to standard immunofluorescence labeling as described above.

For imaging of mRNA:: The hybridization probe for MS2 (5'-(Alexa647)-CTGCAGACATGGGTGATCCTCATGTTTTCTA) was used at 1 μM. The hybridization probe for GFP (5'-(Cy5)-ATCTTGAAGTTGGCCTTGATGCCGTTCTTCTGCTT) was used at 0.5 μM After the SSC washes, cells were incubated for 1 h at RT in 3% BSA prior to standard immunofluorescence labeling as described above.

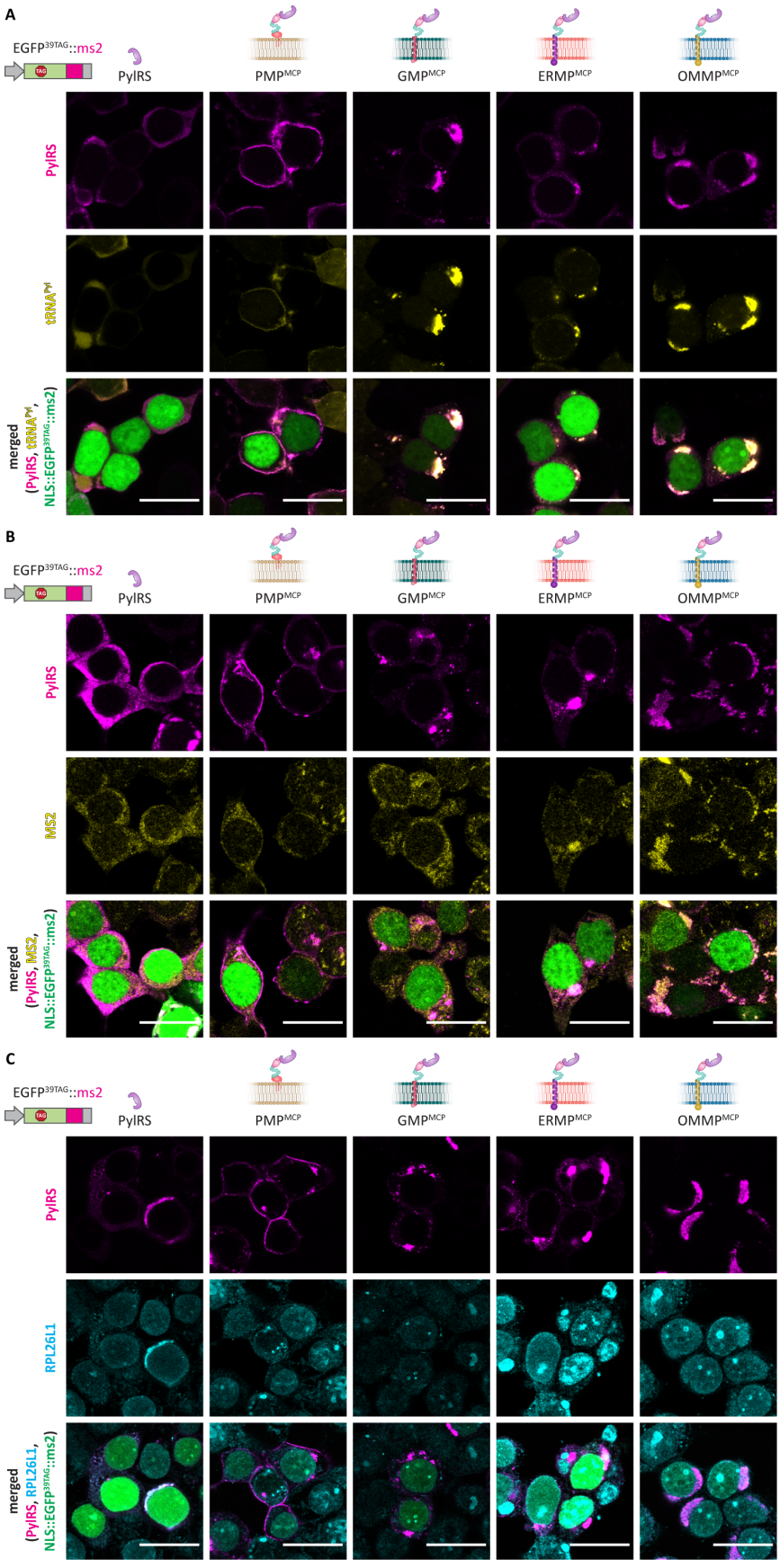
Confocal images were acquired on an Olympus Fluoroview FV3000 microscope using 405 nm (Alexa405), 488 nm (GFP), 594 nm (mCherry) and 640 nm (for Alexa 647, Cy5) lasers for excitation with a 60x/1.40 oil immersion objective. Alternatively, they were acquired on a Leica SP8 STED 3X microscope using the 405 nm (for Alexa405), 488 nm (for GFP), and 647 nm (for Alexa647, Cy5) laser lines for excitation, using a 63x/1.40 oil immersion objective. Images were processed using FIJI software.

STORM images of Alexa647 labeled $tRNA^{Pyl}$ were acquired on a Leica GSDIM microscope with a 160x/1.43 oil immersion objective using a 642 nm laser for excitation. Images were acquired in a glucose oxidase-based oxygen scavenging system buffer [40 μg/mL catalase (Sigma C3155), 0.5 mg/mL glucose oxidase (Sigma G7141), 10% glucose, 10 mM ß-mercaptoethylamine (Sigma 411000), 50 mM Tris-HCl (pH 8), similar to (*289*)]. Images were transformed into TIF files using FIJI software and subsequently analyzed using a customized Igor Pro software.

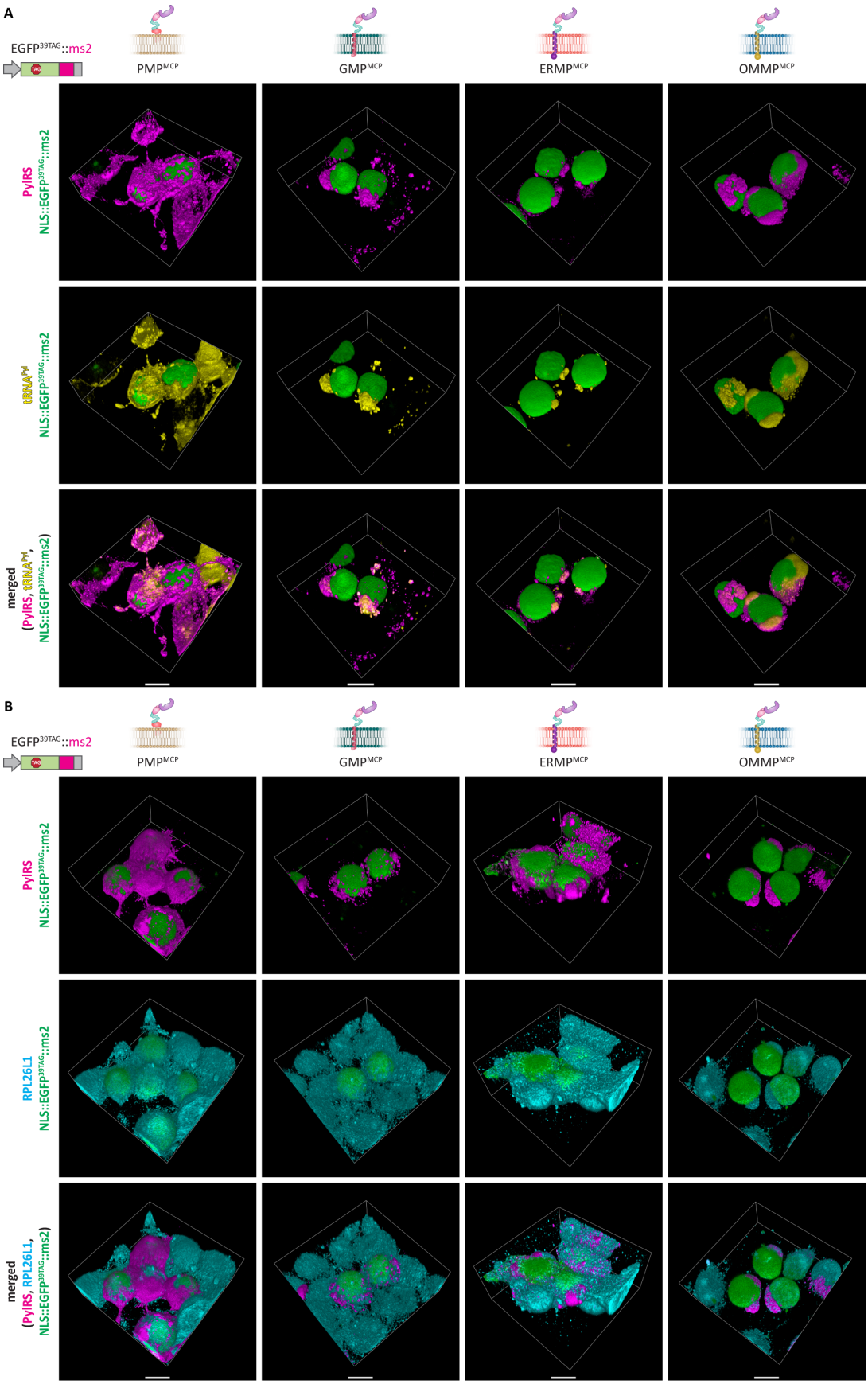
3D Reconstruction: 3D reconstructions were made using the arivis Vision4D software (arivis AG).

4.5 Supplementary figures

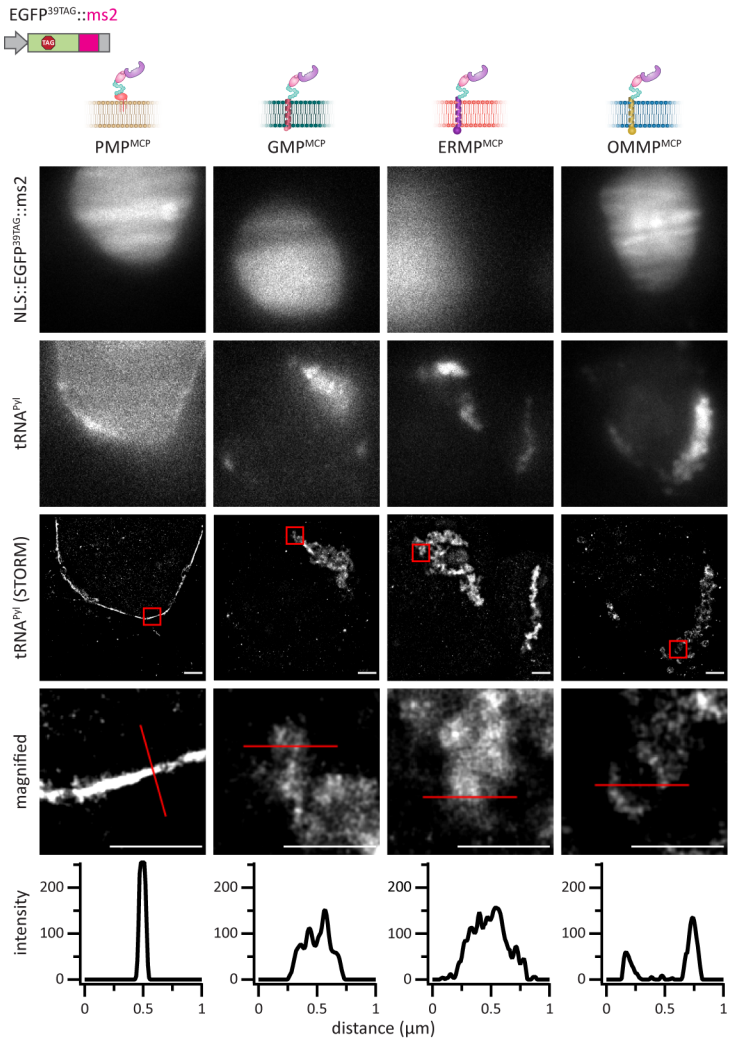
Supplementary Figure 4-1. MCP-based synthetic organelles enrich tRNA^{Pyl}, mRNA::ms2 as well as cellular ribosomes. (A-C) HEK293T cells expressing the indicated OT organelle, tRNA^{Pyl} and a construct encoding EGFP^{39TAG}::ms2 in presence of the ncAA SCO. All images show a selected plain of a z-stack. (**A**) IF staining against PylRS and FISH staining of tRNA^{Pyl} show strong recruitment of tRNA^{Pyl} into OT organelles. Top to bottom: PylRS (magenta), tRNA^{Pyl} (yellow), merged (NLS::EGFP^{39TAG}::ms2 in green). (**B**) IF staining against PylRS and FISH staining of mRNA::ms2 show recruitment of ms2-tagged mRNA into membrane associated OT organelles. Top to bottom: PylRS (magenta), mRNA::ms2 (yellow), merged (NLS::EGFP^{39TAG}::ms2 in green). (**C**) IF stainings against PylRS and RPL26L1 show that endogenous ribosomes are recruited into film-like OT organelles. Note that due to the abundance of ribosomes, a substantial staining outside of the OT organelles can be observed. Top to bottom: PylRS (magenta), RPL26L1 (cyan), merged (NLS::EGFP^{39TAG}::ms2 in green). (Scale bars = 20 μm)


A
EGFP39TAG::ms2
PylRS
PMPMCP
GMPMCP
ERMPMCP
OMMPMCP
PylRS
tRNAPyl
merged (PylRS, tRNAPyl, NLS::EGFP39TAG::ms2)
B
EGFP39TAG::ms2
PylRS
PMPMCP
GMPMCP
ERMPMCP
OMMPMCP
PylRS
MS2
merged (PylRS, MS2, NLS::EGFP39TAG::ms2)
C
EGFP39TAG::ms2
PylRS
PMPMCP
GMPMCP
ERMPMCP
OMMPMCP
PylRS
RPL26L1
merged (PylRS, RPL26L1, NLS::EGFP39TAG::ms2)

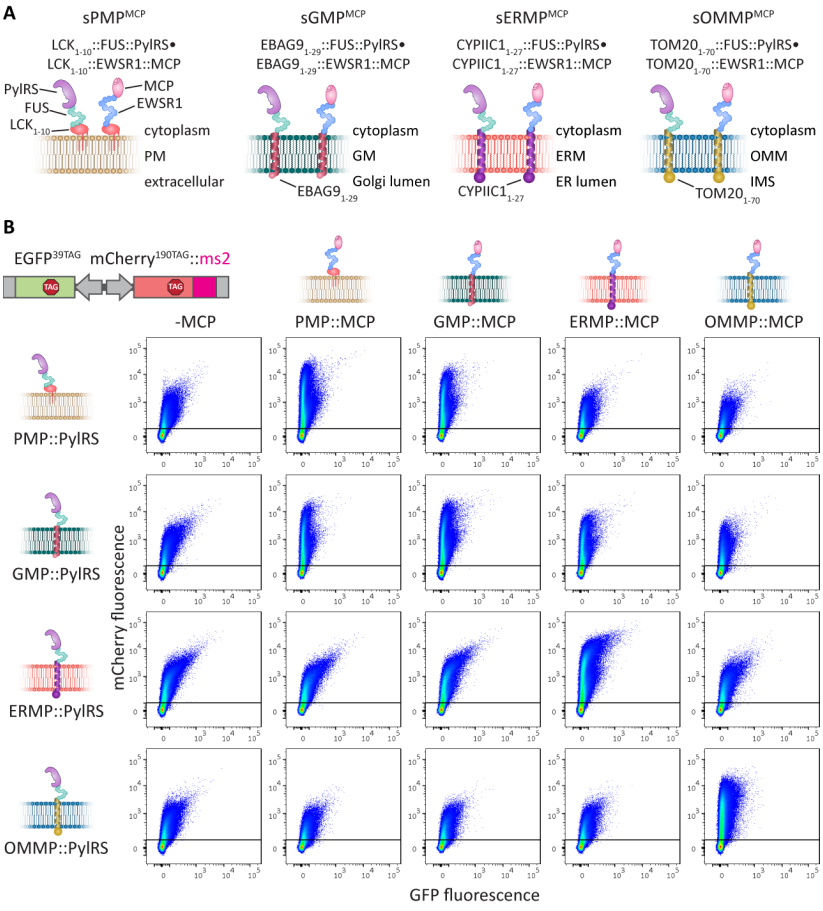
Supplementary Figure 4-2. 3D reconstructions of MCP-based synthetic organelles illustrates efficient tRNA^{Pyl} and ribosomes recruitment. HEK293T cells expressing the indicated OT organelle, tRNA^{Pyl} and a EGFP^{39TAG}::ms2 in presence of the ncAA SCO (corresponding to **Supplementary Figure 4-1**). (**A**) IF staining against PylRS and FISH staining of tRNA^{Pyl} show strong tRNA^{Pyl} recruitment. Top to bottom: PylRS (magenta), tRNA^{Pyl} (yellow), merged (NLS::EGFP^{39TAG}::ms2 shown in green in all panels). (**B**) IF stainings against PylRS and RPL26L1 show endogenous ribosomes in MCP-based OT organelles. Note that due to the abundance of ribosomes, a substantial staining outside of the OT organelles can be observed. Top to bottom: PylRS (magenta), RPL26L1 (cyan), merged (NLS::EGFP^{39TAG}::ms2 in green in all panels). (Scale bars = 10 μm)


A
EGFP39TAG::ms2
PMPMCP
GMPMCP
ERMPMCP
OMMPMCP
PylRS
NLS::EGFP39TAG::ms2
tRNAPyl
NLS::EGFP39TAG::ms2
merged
(PylRS, tRNAPyl,
NLS::EGFP39TAG::ms2)
B
EGFP39TAG::ms2
PMPMCP
GMPMCP
ERMPMCP
OMMPMCP
PylRS
NLS::EGFP39TAG::ms2
RPL26L1
NLS::EGFP39TAG::ms2
merged
(PylRS, RPL26L1,
NLS::EGFP39TAG::ms2)

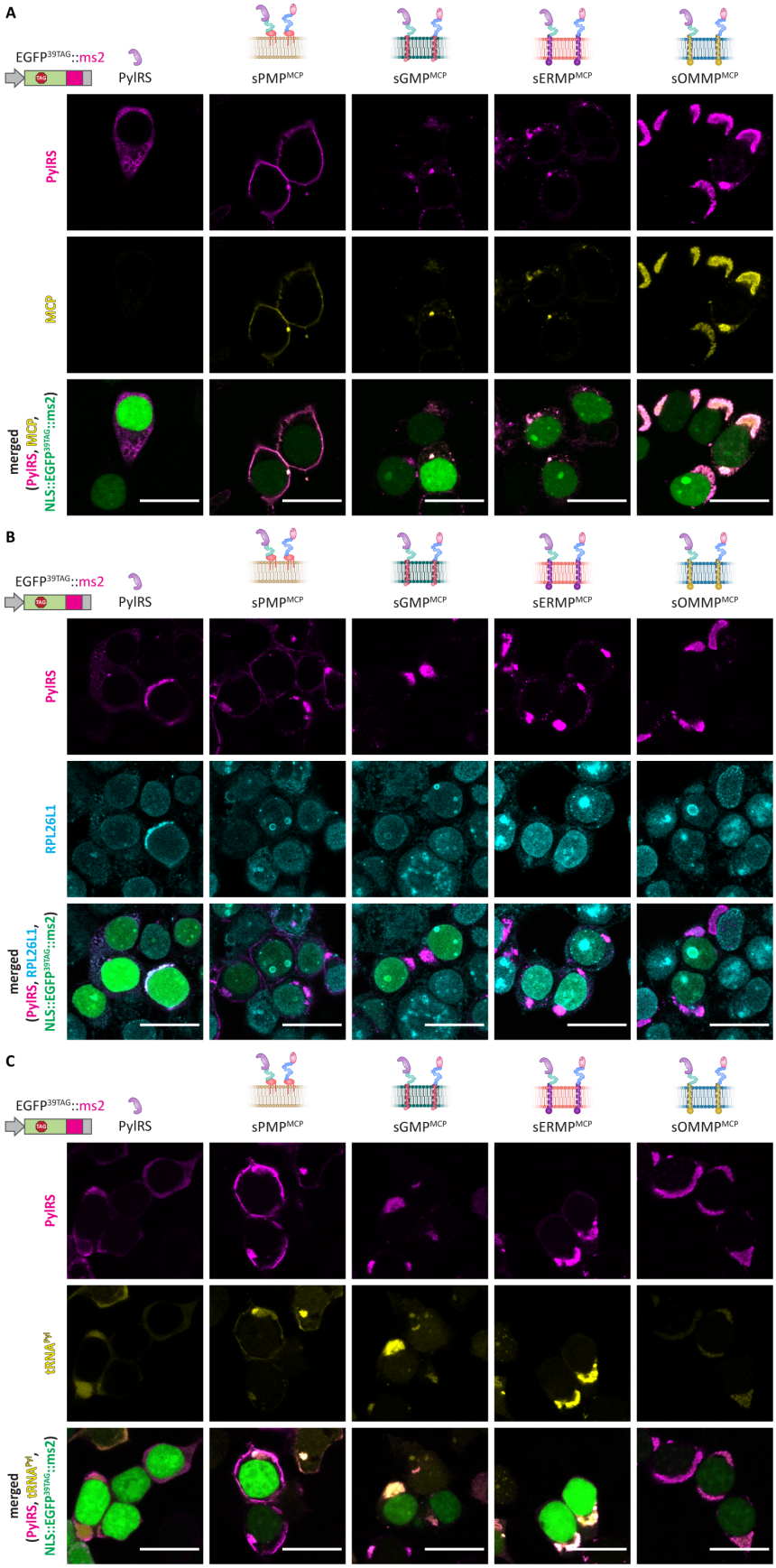
Supplementary Figure 4-3. Super-resolution microscopy shows high enrichment of tRNA^{Pyl} to a very thin layer at OT organelles. HEK293T cells expressing the indicated OT organelle, tRNA^{Pyl} and a construct encoding EGFP^{39TAG}::ms2 in presence of SCO. (A) FISH staining of tRNA^{Pyl} show strong tRNA^{Pyl} recruitment. Particularly for the PMP^{MCP} OT organelle it can be seen that tRNA^{Pyl} is restricted to a thin film of only a few nanometers at the periphery of the cell. Top to bottom: NLS::EGFP::ms2 (TIRF), tRNA^{Pyl} (TIRF), tRNA^{Pyl}-STORM, magnified (corresponding to red box in tRNA^{Pyl}-STORM) and a line scan corresponding to the red line in the magnification. (Scale bars = 1 μm)



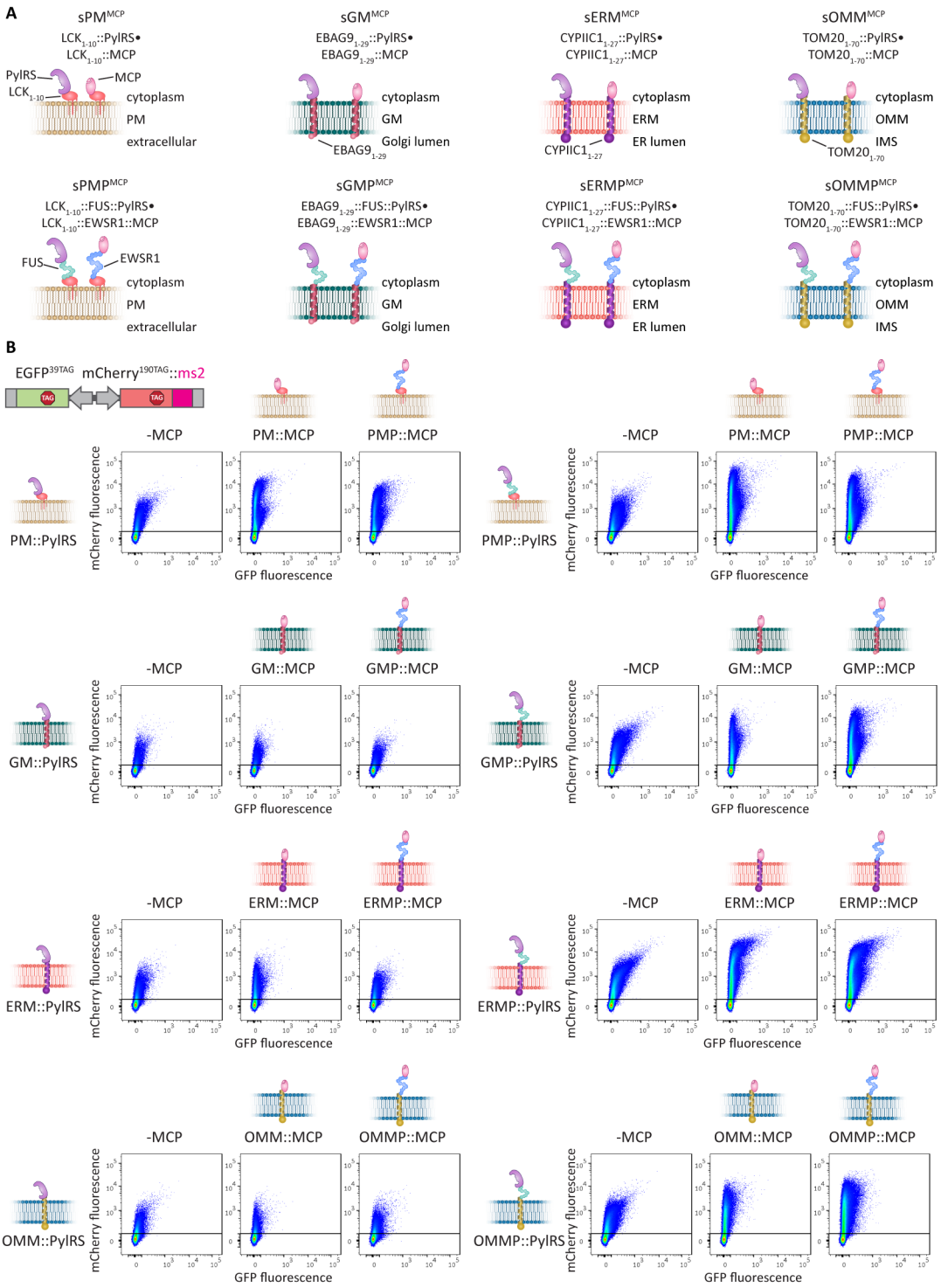
Supplementary Figure 4-4. Mitochondrial membrane targeting is orthogonal to all others, while also plasma membrane and ER membrane targeting are independent. To test independence of the different membrane systems I created constructs that target PylRS or MCP (with FUS or EWSR1 respectively) individually to the four different membrane localizations. (**A**) Overview cartoon of the used constructs. (**B**) FFC analysis of separate OT organelle constructs. HEK293T cells expressing the indicated dual color reporter, $tRNA^{Pyl}$, a respective PylRS OT organelle construct as well as either no MCP protein (first column) or the MCP protein targeted to each respective membrane. For PMP::PylRS mCherry::ms2 expression is enhanced for co-expression with PMP::MCP or GMP::MCP. For GMP::PylRS mCherry::ms2 expression is enhanced in presence of PMP::MCP, GMP::MCP and ERMP::MCP. For ERMP::PylRS the co-expression of GMP::MCP as well as ERMP::MCP enhances mCherry::ms2 expression. For OMMP::PylRS only the co-expression of OMMP::MCP leads to an increase in targeted mCherry::ms2 expression. Thus, the OMMP system is fully orthogonal to all others, while additionally the PMP and ERMP systems are orthogonal. Shown is the sum of at least three independent experiments.



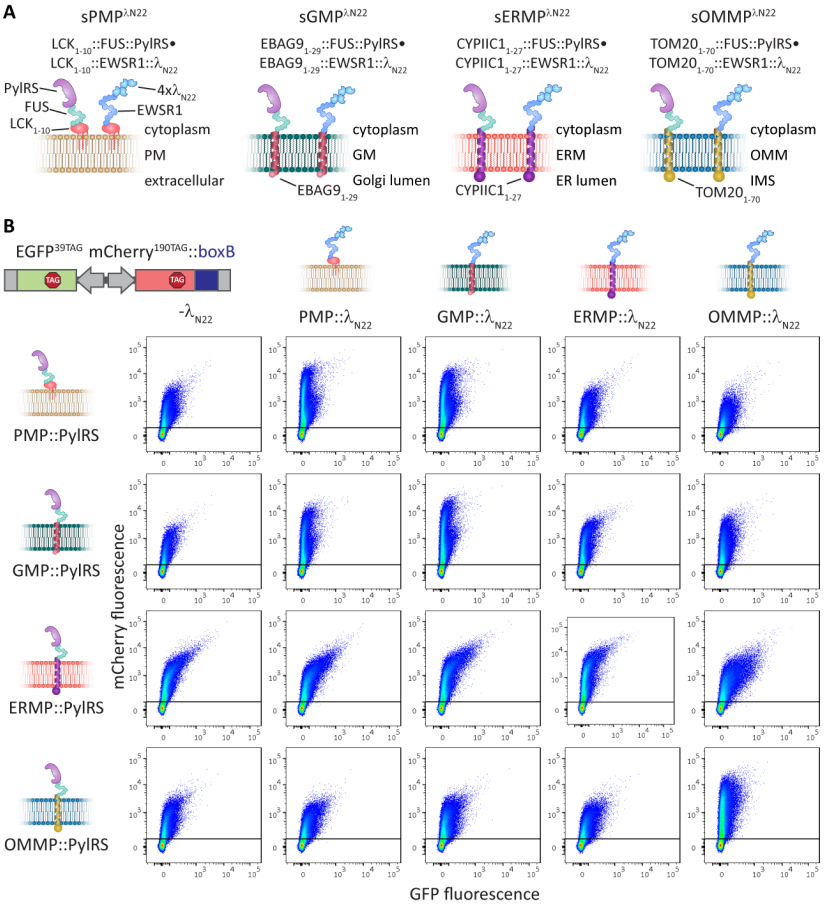
Supplementary Figure 4-5. Separate OT organelle systems co-localize MCP and PylRS and also recruit ribosomes and tRNA^{Pyl}. (A-C) HEK293T cells expressing the indicated OT organelle, tRNA^{Pyl} and a construct encoding EGFP^{39TAG}::ms2 in presence of the ncAA SCO. All images show a selected plain of a z-stack. (**A**) IF staining against PylRS and MCP (with an anti-HA antibody) show co-localization in OT organelles. Top to bottom: PylRS (magenta), MCP (yellow), merged (NLS::EGFP^{39TAG}::ms2 in green). (**B**) IF stainings against PylRS and RPL26L1 show that endogenous ribosomes are recruited into OT film-like organelles. Note that as RPL26L1 is an endogenous protein, also untransfected cells are stained. (Top to bottom: PylRS (magenta), RPL26L1 (cyan), merged (NLS::EGFP^{39TAG}::ms2 in green). (**C**) IF staining against PylRS and FISH staining of tRNA^{Pyl} show strong tRNA^{Pyl} recruitment into membrane associated OT organelles. Top to bottom: PylRS (magenta), tRNA^{Pyl} (yellow), merged (NLS::EGFP^{39TAG}::ms2 in green). (Scale bars = 20 μm)


A
EGFP39TAG::ms2
PylRS
sPMPMCP
sGMPMCP
sERMPMCP
sOMMPMCP
PylRS
MCP
merged (PylRS, MCP, NLS::EGFP39TAG::ms2)
B
EGFP39TAG::ms2
PylRS
sPMPMCP
sGMPMCP
sERMPMCP
sOMMPMCP
PylRS
RPL26L1
merged (PylRS, RPL26L1, NLS::EGFP39TAG::ms2)
C
EGFP39TAG::ms2
PylRS
sPMPMCP
sGMPMCP
sERMPMCP
sOMMPMCP
PylRS
tRNAPyl
merged (PylRS, tRNAPyl, NLS::EGFP39TAG::ms2)

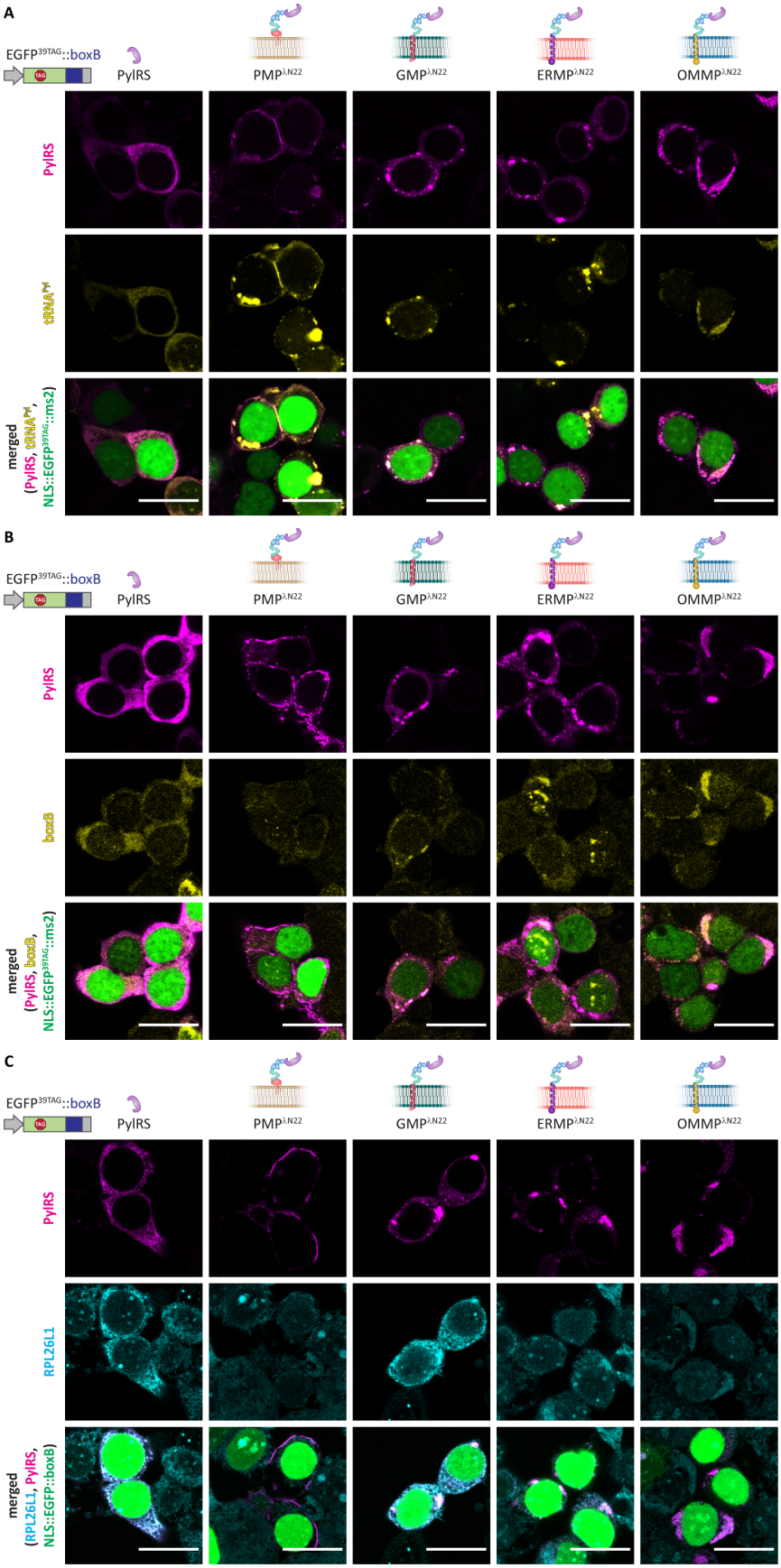
Supplementary Figure 4-6. Not all membrane-based systems require a phase separation domain. To test if phase separation is required for the different membrane systems I created constructs that target PylRS or MCP with or without FUS/EWSR1 individually to the four different membranes. (**A**) Overview cartoon of the used constructs. (**B**) FFC analysis of separate OT organelles with and without phase separation domains. HEK293T cells expressing the indicated dual color reporter, $tRNA^{Pyl}$, a respective PylRS construct without (left panel) or with phase separation domain (right panel), as well as either no MCP protein (first column) or a membrane targeted MCP protein without (second column) or with phase separation domain (third column). Only the PM targeted PylRS variant is efficient without phase separation domain, however it is even more efficient if FUS is added. MCP does not necessarily need to be targeted to the membrane with a phase separation domain for all systems, however particularly for GM and OMM targeting the efficiency is increased by addition of EWSR1 to the MCP variant.


A
sPM^MCP
LCK1-10::PylRS•
LCK1-10::MCP
PylRS
LCK1-10
MCP
cytoplasm
PM
extracellular
sGM^MCP
EBAG91-29::PylRS•
EBAG91-29::MCP
cytoplasm
GM
Golgi lumen
EBAG91-29
sERM^MCP
CYPIIC11-27::PylRS•
CYPIIC11-27::MCP
cytoplasm
ERM
ER lumen
CYPIIC11-27
sOMM^MCP
TOM201-70::PylRS•
TOM201-70::MCP
cytoplasm
OMM
IMS
TOM201-70
sPMP^MCP
LCK1-10::FUS::PylRS•
LCK1-10::EWSR1::MCP
FUS
EWSR1
cytoplasm
PM
extracellular
sGMP^MCP
EBAG91-29::FUS::PylRS•
EBAG91-29::EWSR1::MCP
cytoplasm
GM
Golgi lumen
sERMP^MCP
CYPIIC11-27::FUS::PylRS•
CYPIIC11-27::EWSR1::MCP
cytoplasm
ERM
ER lumen
sOMMP^MCP
TOM201-70::FUS::PylRS•
TOM201-70::EWSR1::MCP
cytoplasm
OMM
IMS
B
EGFP39TAG
mCherry190TAG::ms2
TAG
TAG
-MCP
PM::MCP
PMP::MCP
PM::PylRS
PMP::PylRS
GM::MCP
GMP::MCP
GM::PylRS
GMP::PylRS
ERM::MCP
ERMP::MCP
ERM::PylRS
ERMP::PylRS
OMM::MCP
OMMP::MCP
OMM::PylRS
OMMP::PylRS
mCherry fluorescence
GFP fluorescence

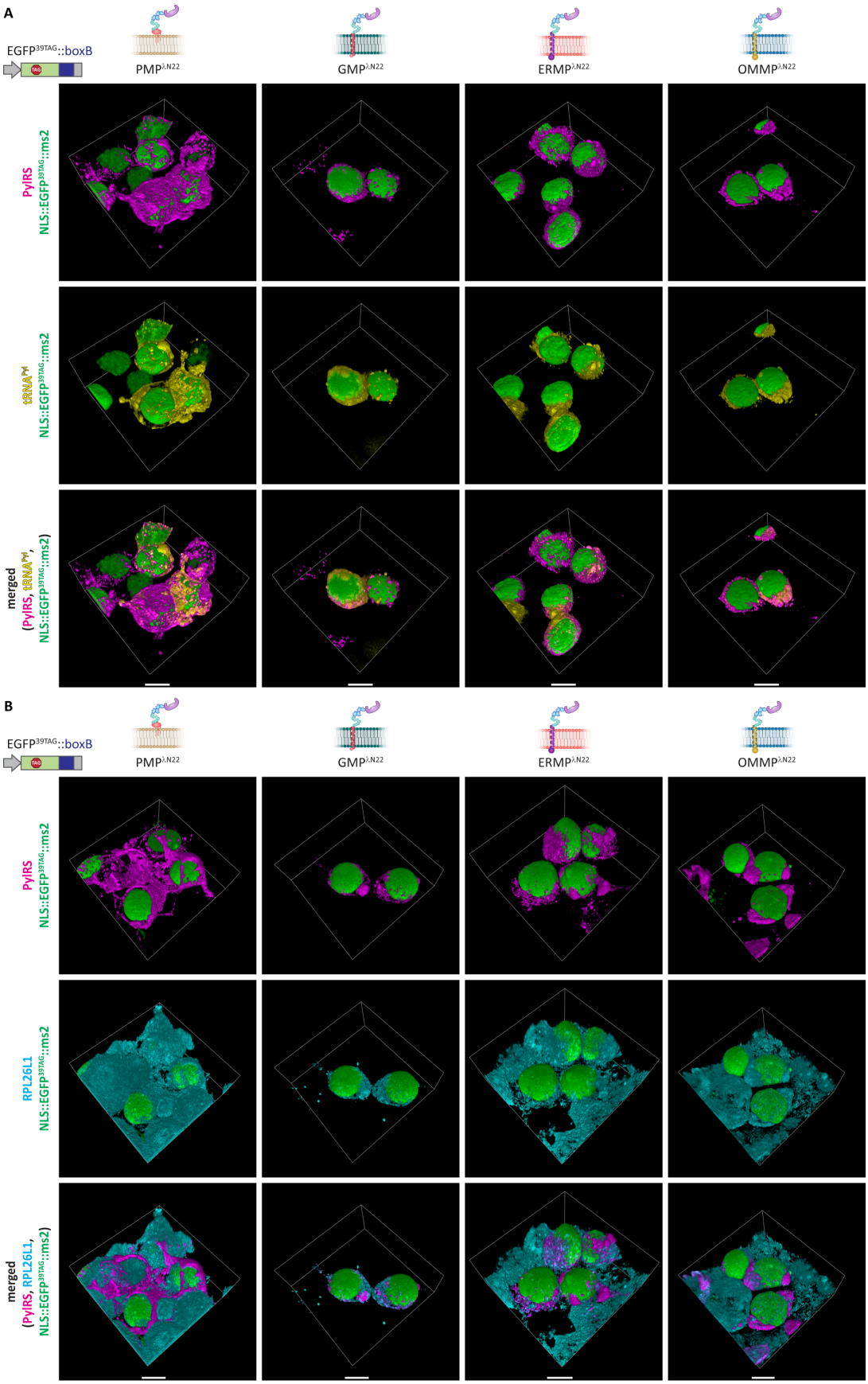
Supplementary Figure 4-7. Also, for boxB-λ_{N22} OT organelles OMM targeting is orthogonal to all others, while PM and ERM targeting are also independent. To test independence of the different membrane systems I created constructs that target PylRS or λ_{N22}-peptides (with FUS or EWSR1 respectively) to the four different membranes. (**A**) Overview cartoon of the used constructs. (**B**) FFC analysis of separate OT organelle constructs. HEK293T cells expressing the indicated dual color reporter, $tRNA^{Pyl}$, a respective PylRS OT organelle construct as well as either no λ_{N22}-peptides (first column) or the λ_{N22}-peptides targeted to each respective membrane. For PMP::PylRS mCherry::boxB expression is enhanced in presence PMP::λ_{N22} or GMP::λ_{N22}. For GMP::PylRS mCherry::boxB expression is enhanced by PMP::λ_{N22}, GMP::$\lambda_{N22}$and ERMP::λ_{N22}. For ERMP::PylRS the co-expression of GMP::λ_{N22} as well as ERMP::λ_{N22} enhances mCherry::boxB expression. For OMMP::PylRS only OMMP::λ_{N22} leads to an increase in targeted mCherry::boxB expression. Thus, as for the ms2-MCP based systems the OMMP system is fully orthogonal to all others, while additionally the PMP and ERMP systems are orthogonal. Shown is the sum of at least three independent experiments.


A
sPMP$^{\lambda N22}$
LCK$_{1-10}$::FUS::PylRS•
LCK$_{1-10}$::EWSR1::λ_{N22}
PylRS
FUS
LCK$_{1-10}$
4xλ_{N22}
EWSR1
cytoplasm
PM
extracellular
sGMP$^{\lambda N22}$
EBAG9$_{1-29}$::FUS::PylRS•
EBAG9$_{1-29}$::EWSR1::λ_{N22}
cytoplasm
GM
Golgi lumen
EBAG9$_{1-29}$
sERMP$^{\lambda N22}$
CYPIIC1$_{1-27}$::FUS::PylRS•
CYPIIC1$_{1-27}$::EWSR1::λ_{N22}
cytoplasm
ERM
ER lumen
CYPIIC1$_{1-27}$
sOMMP$^{\lambda N22}$
TOM20$_{1-70}$::FUS::PylRS•
TOM20$_{1-70}$::EWSR1::λ_{N22}
cytoplasm
OMM
IMS
TOM20$_{1-70}$
B
EGFP^{39TAG} mCherry^{190TAG}::boxB
TAG
TAG
-λ_{N22}
PMP::λ_{N22}
GMP::λ_{N22}
ERMP::λ_{N22}
OMMP::λ_{N22}
PMP::PylRS
GMP::PylRS
ERMP::PylRS
OMMP::PylRS
mCherry fluorescence
GFP fluorescence

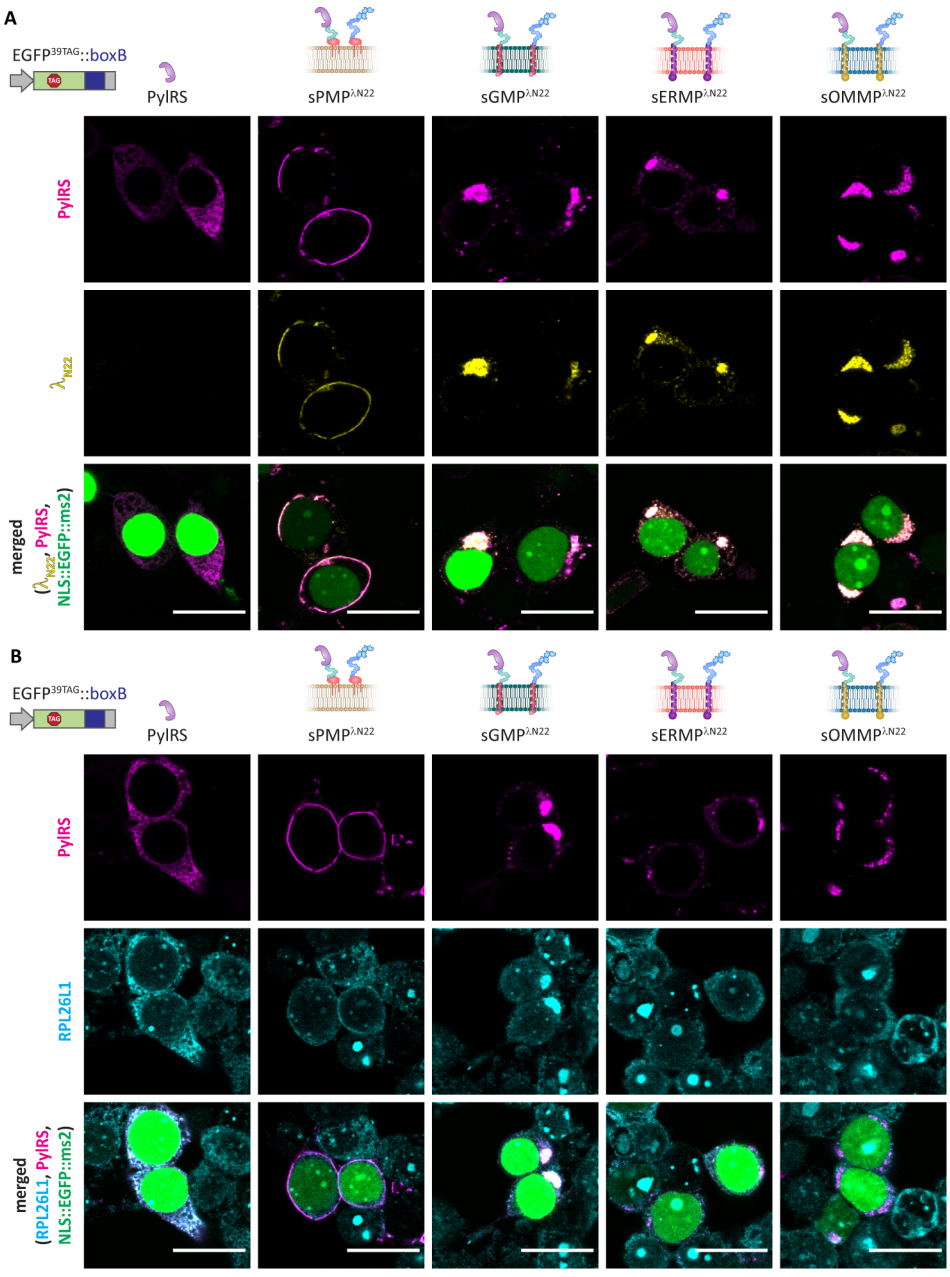
Supplementary Figure 4-8. λ_{N22}-based synthetic organelles enrich $tRNA^{Pyl}$, mRNA::boxB as well as cellular ribosomes. (A-C) HEK293T cells expressing the indicated OT organelle, $tRNA^{Pyl}$ and a construct encoding $EGFP^{39TAG}$::boxB in presence of the ncAA SCO. All images show a selected plain of a z-stack. (**A**) IF staining against PylRS and FISH staining of $tRNA^{Pyl}$ show strong recruitment of $tRNA^{Pyl}$ into OT organelles. Top to bottom: PylRS (magenta), $tRNA^{Pyl}$ (yellow), merged (NLS::$EGFP^{39TAG}$::ms2 in green). (**B**) IF staining against PylRS and FISH staining of mRNA::boxB show recruitment of boxB-tagged mRNA into λ_{N22}-based OT organelles. Top to bottom: PylRS (magenta), mRNA::boxB(yellow), merged (NLS::$EGFP^{39TAG}$::ms2 in green). (**C**) IF stainings against PylRS and RPL26L1 show that endogenous ribosomes are recruited into membrane associated OT organelles. Note that due to the abundance of ribosomes, a substantial staining outside of the OT organelles can be observed. Top to bottom: PylRS (magenta), RPL26L1 (cyan), merged (NLS::$EGFP^{39TAG}$::ms2 in green). (Scale bars = 20 μm)


A
EGFP39TAG::boxB
PylRS
PMPλN22
GMPλN22
ERMPλN22
OMMPλN22
PylRS
tRNAPyl
merged (PylRS, tRNAPyl, NLS::EGFP39TAG::ms2)
B
EGFP39TAG::boxB
PylRS
PMPλN22
GMPλN22
ERMPλN22
OMMPλN22
PylRS
boxB
merged (PylRS, boxB, NLS::EGFP39TAG::ms2)
C
EGFP39TAG::boxB
PylRS
PMPλN22
GMPλN22
ERMPλN22
OMMPλN22
PylRS
RPL26L1
merged (RPL26L1, PylRS, NLS::EGFP::boxB)

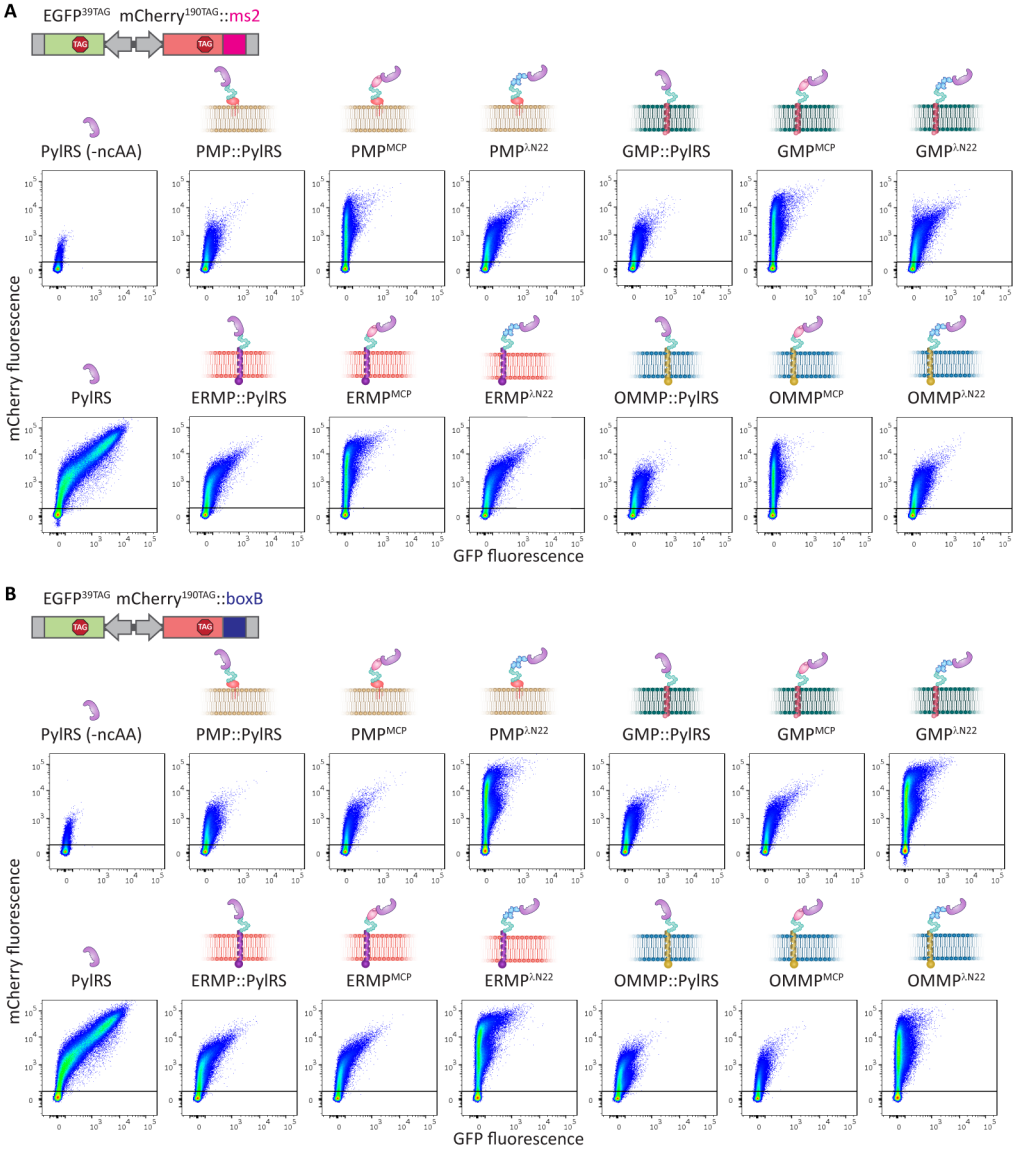
Supplementary Figure 4-9. 3D reconstructions of λ_{N22}-based synthetic organelles illustrate $tRNA^{Pyl}$ and ribosomes recruitment. HEK293T cells expressing the indicated OT organelle, $tRNA^{Pyl}$ and a construct encoding $EGFP^{39TAG}$::boxB in presence of the ncAA SCO (corresponding to **Supplementary Figure 4-8**). (**A**) IF staining against PylRS and FISH staining of $tRNA^{Pyl}$ show strong $tRNA^{Pyl}$ recruitment. Top to bottom: PylRS (magenta), $tRNA^{Pyl}$ (yellow), merged (NLS::$EGFP^{39TAG}$::boxB shown in green in all panels). (**B**) IF stainings against PylRS and RPL26L1 show endogenous ribosomes in OT film-like organelles. Note that due to the abundance of ribosomes, a substantial staining outside of the OT organelles can be observed. Top to bottom: PylRS (magenta), RPL26L1 (cyan), merged (NLS::$EGFP^{39TAG}$::boxB in green in all panels). (Scale bars = 10 μm)


A
EGFP39TAG::boxB
PMPλN22
GMPλN22
ERMPλN22
OMMPλN22
PylRS
NLS::EGFP39TAG::ms2
tRNAPyl
NLS::EGFP39TAG::ms2
merged
(PylRS, tRNAPyl,
NLS::EGFP39TAG::ms2)
B
EGFP39TAG::boxB
PMPλN22
GMPλN22
ERMPλN22
OMMPλN22
PylRS
NLS::EGFP39TAG::ms2
RPL26L1
NLS::EGFP39TAG::ms2
merged
(PylRS, RPL26L1,
NLS::EGFP39TAG::ms2)

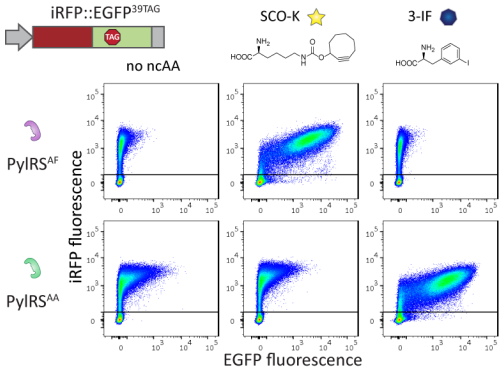
Supplementary Figure 4-10. Separate λ_{N22}-based synthetic organelles co-localize PylRS and λ_{N22}-peptides as well as ribosomes. (A-B) HEK293T cells expressing the indicated OT organelle, $tRNA^{Pyl}$ and a construct encoding $EGFP^{39TAG}$::boxB in presence of the ncAA SCO. All images show a selected plain of a z-stack. (**A**) IF staining against PylRS and λ_{N22} (with an anti-Myc antibody) show co-localization in OT organelles. Top to bottom: PylRS (magenta), λ_{N22} (yellow), merged (NLS::$EGFP^{39TAG}$::boxB in green). (**B**) IF stainings against PylRS and RPL26L1 show that endogenous ribosomes are recruited into membrane associated OT organelles. Note that due to the abundance of ribosomes, a substantial staining outside of the OT organelles can be observed. Top to bottom: PylRS (magenta), RPL26L1 (cyan), merged (NLS::$EGFP^{39TAG}$::boxB in green). (Scale bars = 20 μm)



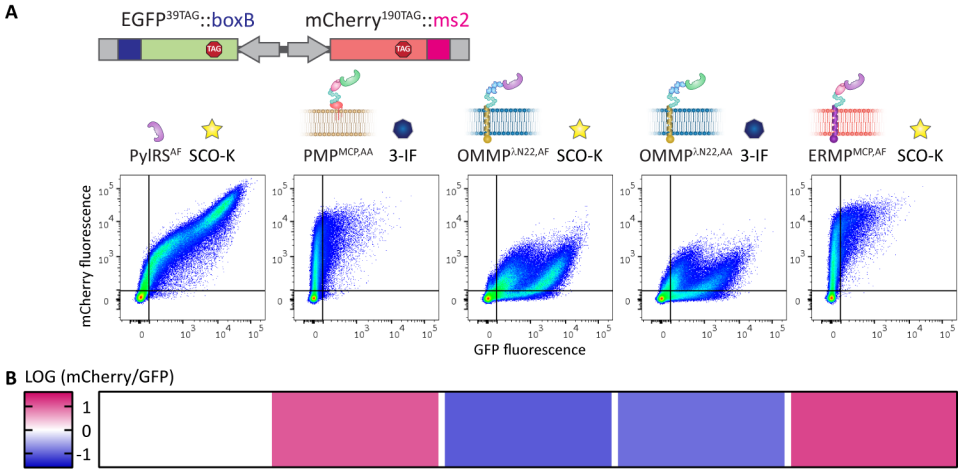
Supplementary Figure 4-11. BoxB- and ms2-based mRNA targeting are mutually orthogonal. To test if mRNAs tagged with boxB- or ms2-loops are selectively recruited to their respective organelles, the two different dual color reporters were co-expressed with membrane localized OT organelles that either had no RNA recruitment domain, MCP or λ_{N22}-peptides. **(A,B)** Besides the indicated OT organelles cells were also expressing $tRNA^{Pyl}$ and incubated with the ncAA SCO. As a control cells were transfected with a construct encoding a cytoplasmic PylRS variant. (**A**) FFC analysis for the ms2-reporter reveals selective mCherry expression only in presence of the MCP-based OT organelles. (**B**) Analogously, boxB-tagged mCherry is only selectively expressed in presence of λ_{N22}-based OT organelles.



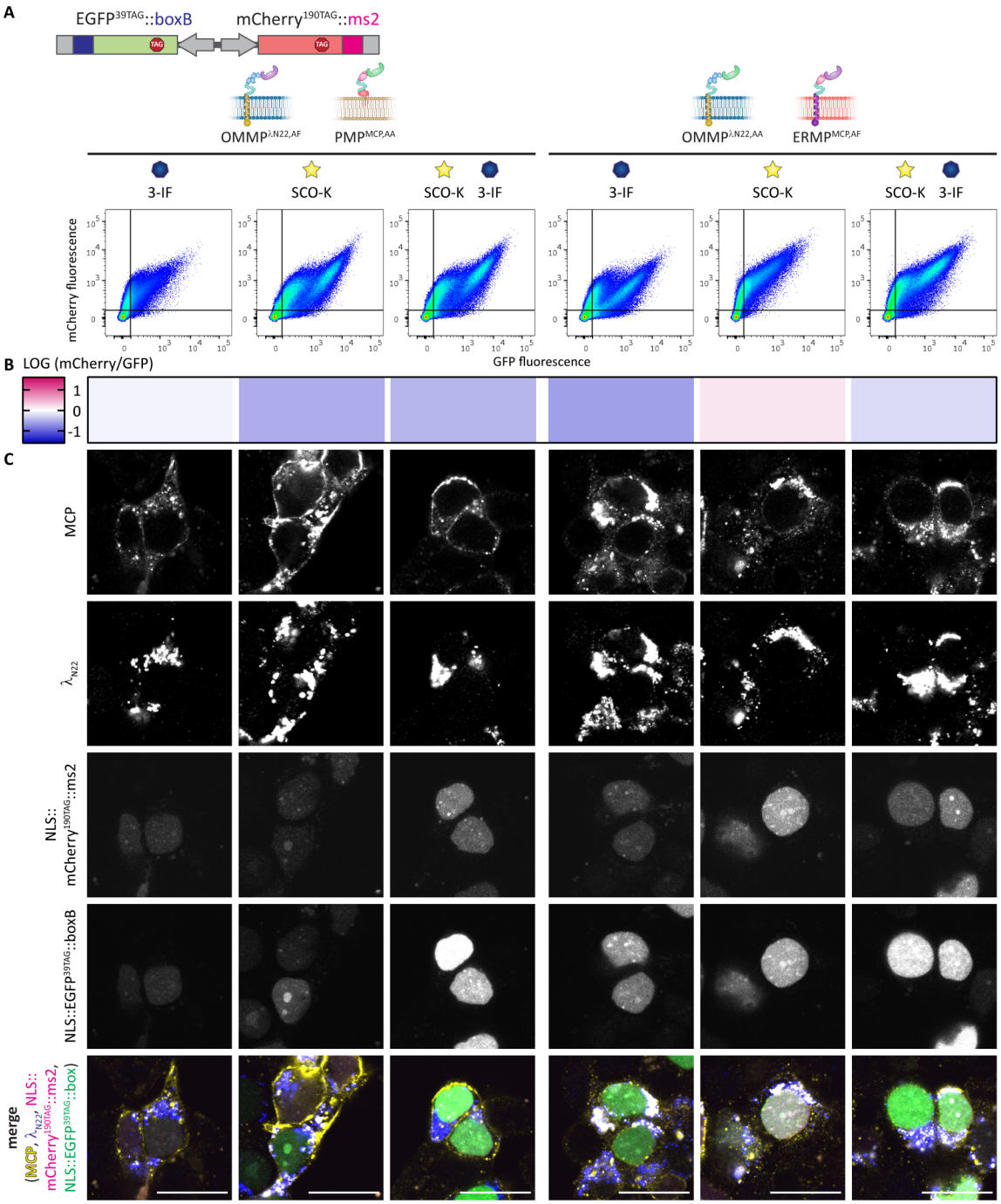
Supplementary Figure 4-12. PylRS^{AF} and PylRS^{AA} exhibit orthogonal substrate specificities. FFC analysis of cells expressing an iRFP::EGFP^{39TAG} reporter, tRNA^{Pyl} and PylRS^{AF} or PylRS^{AA}. iRFP serves as a transfection control, while EGFP fluorescence reports on successful amber codon suppression. Cells were incubated either with no ncAA, with SCO (250 μM) or two aromatic amino acids 3-IF (1 mM). For PylRS^{AF} robust EGFP expression can be observed in presence of SCO, for PylRS^{AA} EGFP expression is only enhance in presence of the 3-IF.



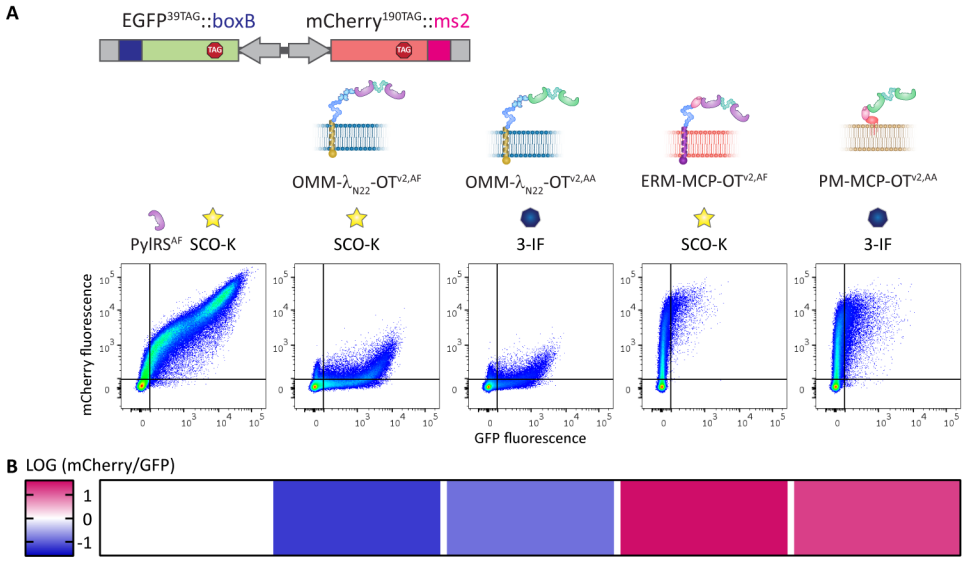
Supplementary Figure 4-13. Selective GCE of boxB- and ms2-tagged mRNA in presence of λ_{N22}- or MCP-based OT organelles. (**A**) FFC analysis of HEK293T cells expressing the double tagged reporter (EGFP^{39TAG}::boxB, mCherry^{190TAG}::ms2) together with tRNA^{Pyl} and the indicated GCE systems (400 ng of reporter and tRNA plasmid, co-transfected with 100 ng of the OT organelle construct and 300 ng mock plasmid). Experiments were performed in the presence of the indicated ncAAs. Shown is the sum of three independent experiments (except for ERMP^{MCP,AF}, where only the sum of two independent experiments is shown). In presence of MCP-based OT organelles predominant mCherry expression can be observed, while in the presence of λ_{N22}-based OT organelles mostly full-length EGFP is produced (**B**) Heatmap showing the ratio of the mCherry signal GFP signal corresponding to the conditions in (A), normalized to a cytoplasmic GCE system. Red values indicate a positive, logarithmic mCherry to GFP ratio, blue values indicate a negative logarithmic value, thus an excess of GFP over mCherry.



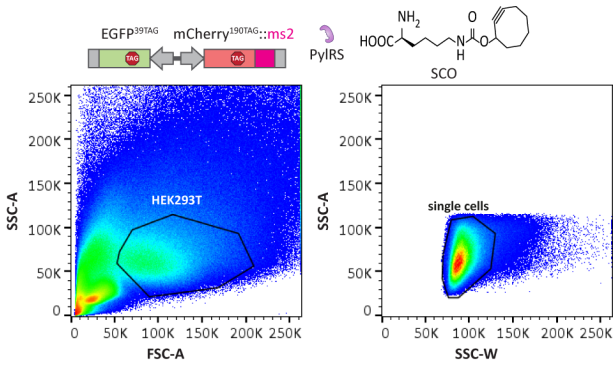
Supplementary Figure 4-14. Combining two OT film-like organelles leads to mixing and substantial selectivity loss. (**A**) FFC analysis of HEK293T cells expressing the double tagged reporter (EGFP^{39TAG}::boxB, mCherry^{190TAG}::ms2) together with tRNA^{Pyl} and the indicated GCE systems (400 ng of reporter and tRNA plasmid, co-transfected with 100 ng of the OT organelle construct and 300 ng mock plasmid). Experiments were performed in the presence of the indicated ncAAs. Shown is the sum of three independent experiments. When OMMP$^{\lambda N22,AF}$ is combined with PMP^{MCP,AA} full-length EGFP^{39TAG} and mCherry^{190TAG} expression can be observed, independent of the added ncAA. Similarly, when OMMP$^{\lambda N22,AA}$ is combined with ERMP^{MCP,AF}, substantial expression of both full-length EGFP^{39TAG} and mCherry^{190TAG} can be observed under all conditions. (**B**) Heatmap showing the ratio of the mCherry to GFP signal corresponding to the conditions in (A), normalized to a cytoplasmic GCE system (see **Supplementary Figure 4-13**). Red values indicate a positive, logarithmic mCherry to GFP ratio, blue values indicate a negative logarithmic value, thus an excess of GFP over mCherry. (**C**) IF staining against MCP and λ_{N22} corresponding to the conditions in (A). Particularly, for PMP^{MCP,AA} a substantial re-localization to mitochondria can be observed. Also, ERMP^{MCP,AF} and OMMP$^{\lambda N22,AA}$ clearly mix. Top to bottom: MCP, λ_{N22}, mCherry, EGFP and merged (MCP in yellow, λ_{N22} in blue, mCherry in magenta, EGFP in green). (Scale bars = 20 μm)



Supplementary Figure 4-15. Selective GCE of boxB- and ms2-tagged mRNA in presence of optimized λ_{N22}- or MCP-based OT organelles. (**A**) FFC analysis of HEK293T cells expressing the double tagged reporter (EGFP^{39TAG}::boxB, mCherry^{190TAG}::ms2) together with tRNA^{Pyl} and the indicated GCE systems (400 ng of reporter and tRNA plasmid, co-transfected with 100 ng of the OT organelle construct and 300 ng mock plasmid). Experiments were performed in the presence of the indicated ncAAs. Shown is the sum of three independent experiments. In presence of MCP-based OT organelles predominant mCherry expression can be observed, while in the presence of λ_{N22}-based OT organelles mostly full-length EGFP is produced (**B**) Heatmap showing the ratio of the mCherry signal GFP signal corresponding to the conditions in (A), normalized to a cytoplasmic GCE system. Red values indicate a positive, logarithmic mCherry to GFP ratio, blue values indicate a negative logarithmic value, thus an excess of GFP over mCherry.



Supplementary Figure 4-16. FFC gating scheme to identify single HEK293T cells. HEK293T cells were first identified based on SSC-A and FSC-A values. Subsequently, single cells are identified using SSC-A and SSC-W parameters. In this figure one representative example is shown. Here, HEK293T cells were transfected with a cytoplasmic PylRS system, the indicated ms2-based double reporter and tRNA^{Pyl} in presence of the ncAA SCO. Cells that passed the first gate (left panel) are subsequently gated for single cells (right panel).



Chapter 5 Discussion

In this thesis I developed unprecedented approaches to engineer synthetic organelles for orthogonal translation of proteins.

Initially, using microtubule associated synthetic organelles, I developed the first orthogonal translation strategy in mammals that expands the genetic code exclusively for selected mRNAs, effectively creating the first eukaryotic cells with two genetic codes.

Furthermore, I developed multiple membrane associated synthetic organelles, which allow to synthetically control gene expression by two additional layers. First, the organelles demonstrate that it is possible to equip mammalian with multiple orthogonal genetic codes. Second, by sequestering enzymes into different subcellular compartments, it is possible to make them functionally orthogonal based on spatial separation, which is a completely new way to create orthogonal enzymes.

5.1 Microtubule-associated membraneless organelles for orthogonal translation

The first approach towards customized orthogonal translation was inspired by LLPS principles in the cytoplasm. To create highly selective OT organelles, phase separating moieties were combined with spatial targeting to the microtubule cytoskeleton, since it was not sufficient to just use one assembler component (*263*).

If only the phase separation functionality is used GCE is still efficient (50% efficiency), but it is not particularly selective (1.2-fold selectivity). If only the kinesin motor protein is used selectivity is slightly higher (up to 2-fold), but efficiency is substantially reduced [below 20%, **Appendix II**, **Fig. 2**, (*263*)]. This is likely caused by an insufficient spatial enrichment if only one targeting domain is used.

The OT organelles are assayed for selectively expanding the genetic code of only recruited mRNAs. The key molecule they need to sequester for this is the small $tRNA^{Pyl}$. If phase separation or kinesin motor domains are used individually to construct OT organelles, IF and FISH stainings revealed a broad distribution of PylRS and $tRNA^{Pyl}$ throughout the cell [**Appendix II**, **Fig. 4A**, (*263*)]. This is likely insufficient to prevent ribosomes in the cytoplasm to access the suppressor $tRNA^{Pyl}$ and thus mRNA that are not specifically targeted can also be translated with an expanded genetic code. Only when kinesin motor proteins and phase separation domains are combined a strong local enrichment of $tRNA^{Pyl}$, PylRS and mRNA::ms2 can be observed [**Appendix II**, **Fig. 4A**, **Fig. S6-S8**, (*263*)]. This then allows to

exclusively expand the genetic code for the recruited mRNA::ms2 inside or near the OT organelle [selectivity up to 10-fold, efficiency approximately 50%, **Appendix II**, **Fig. 2**, (*263*)]. Orthogonal translation could either take place on the surface of or within the OT organelle. To get insights into this mechanism I stained for ribosomal proteins and PylRS in cells expressing OT organelles and observed that ribosomes strongly enrich throughout the entire organelle [**Appendix II**, **Fig. 4B**, **C**, **Fig. S8**, (*263*)]. It is thus likely, that translation happens within the organelle. To definitively proof this, it would however be necessary to directly image translation, which could for example be done using SunTag- or MoonTag-based imaging of the translation of individual mRNAs (*290–292*).

5.1.1 PylRS-based OT organelles efficiently sequester tRNA^{Pyl}

One striking observation in these experiments is that OT organelles comprised of a kinesin motor protein and a phase separating moiety can highly efficiently recruit tRNA^{Pyl} and effectively deplete it from the remaining cytoplasm. For a small molecule like a tRNA this is quite surprising as it should rapidly diffuse throughout the cell. In FISH experiments, for the kinesin based OT organelles [see **Appendix II, Fig. 4**, (*263*)], but also particularly for the OT film-like organelles, tRNA^{Pyl} cannot be detected in the cytoplasm (**Supplementary Figure 4-1**, **Supplementary Figure 4-2** and **Supplementary Figure 4-3**), which is probably the reason why these organelles can selectively reassign stop codons only for recruited mRNAs.

This is even more striking as when I tried to construct OT organelles with other orthogonal RS systems [like *E. coli* LeuRS (*293*, *294*), *E. coli* TrpRS (*39*), *E. coli* TyrRS (*295–297*) or *M. alvus* PylRS (*55*)], I did not observe selective orthogonal translation of the recruited mRNA::ms2 (see **Figure 5-1** for OMMP organelles with *E. coli* LeuRS, *E. coli* TyrRS and *M. alvus* PylRS)

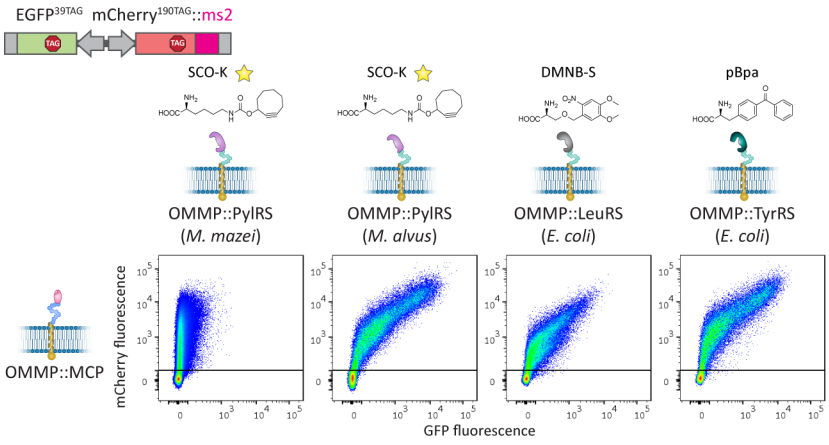


Figure 5-1. OT organelles are only selective with *M. mazei PylRS*. FFC analysis of HEK293T cells expressing the indicated ms2-based reporter, OMMP::MCP, a OMM-targeted RS and corresponding amber suppressor tRNAs. Experiments were performed in presence of the indicated ncAAs. Neither *M. alvus* PylRS, nor the *E. coli* TyrRS or LeuRS enable selective expression of mCherry::ms2 (see **Appendix I** for a detailed methods description).

What makes *M. mazei* PylRS special and how can it capture its tRNA in a synthetic organelle?

One possible explanation is that if PylRS has an extraordinarily high affinity for its cognate $tRNA^{Pyl}$ and might recruit it particularly strong. However, *M. mazei* PylRS variants have a dissociation constant (K_D) of 0.8 μM for its cognate $tRNA^{Pyl}$ (*298*), similar to the one observed for wild-type *M. barkeri* PylRS of approximately 0.2 μM (*47*, *299*). This is similar to the K_D measured for *E. coli* LeuRS at 0.2 μM (*300*) and *E. coli* TyrRS at 0.3 μM (*301*). The K_D of *E. coli* TrpRS is slightly higher at 2 μM (*302*). However, they are all in the same order of magnitude and it is thus unlikely that the K_D can explain the distinct selectivity.

A potential alternative would be that the aminoacylation activity of PylRS could be extremely slow. PylRS has an aminoacylation rate constant of 0.008 s^{-1} (*303*), meaning that one PylRS molecule takes more than two minutes to aminoacylate one $tRNA^{Pyl}$ molecule. In contrast, the aminoacylation rate constants for canonical RSs is typically 1000-fold higher (*300*, *304–307*), for example for *E. coli* TyrRS it is 0.7–11 s^{-1} (*301*, *307*), *E. coli* LeuRS it is 6 s^{-1} (*300*), for *E. coli* TrpRS it is 13-44 s^{-1} (*302*, *308*), which means that these endogenous tRNA synthetases aminoacylate their cognate tRNAs 100-5000 fold faster than PylRS and a tRNA molecule will be aminoacylated by its cognate synthetase in only 22 ms to 1 s. This striking difference makes sense from an evolutionary point of view, as PylRS typically only needs to provide the aminoacylated tRNAs for approximately 50 codons in the *Methanosarcina* genome, while for example the *E. coli* LeuRS needs to provide the substrate for 150.000 codons (*303*).

In the context of the synthetic organelle it can thus be hypothesized that *M. mazei* PylRS is sufficiently slow to permit the locally translating ribosomes to consume all aminoacylated tRNAs before they can diffuse out into the cytoplasm. In contrast the faster RS systems produce aminoacylated tRNAs so quickly that they cannot be translationally consumed, which enables them to diffuse out of the organelle and suppress stop codons in untargeted mRNAs. This can likely explain why only *M. mazei* PylRS enables orthogonal translation.

5.1.2 OT organelles are highly adaptable

In contrast to the orthogonal translation approaches developed for *E. coli*, which I described in more detail in **Section 1.2.1.1**, in this approach only spatial proximity defines how the genetic code is expanded. Therefore, it is straightforward to change the targeted codon by simple mutagenesis of the anticodon of the suppressor tRNA and of the mRNA::ms2 [**Appendix II**, **Fig. 2**, **Fig. S1**, (*263*)]. Furthermore, this technology can be adapted to a variety of proteins including even transmembrane proteins that need to be inserted into the ER membrane during translation, highlighting the versatility of synthetic organelles based on phase separation and kinesin targeting [**Appendix II**, **Fig. 3**, **Fig. S5**, (*263*)].

Even though it is possible to target other stop codons, using the OT organelle approach it is currently only possible to selectively incorporate one ncAA into one POI. To incorporate more than one ncAA into the same POI, it would be necessary to reassign multiple codons within one organelle. Unfortunately, the OT organelles developed in this study can only use *M. mazei* PylRS for orthogonal translation and all other tested RSs fail to confer selectivity (**Figure 5-1**), thus only one codon can be reassigned.

I hypothesize that the reason for this is that the other orthogonal RSs have much higher catalytic rates, which can oversaturate the OT organelle with aminoacylated tRNAs, which can then not be consumed by the local ribosomes (see **Section 5.1.1**). Therefore, it might be possible to overcome this limitation by evolving orthogonal RSs to have a lower catalytic rate. However, this would need innovative evolution approaches.

Typically, directed evolution approaches applied to orthogonal RS/tRNA pairs combine positive and negative selection markers that are only produced upon successful stop codon suppression (*309*). The positive selection marker is used in presence of the ncAA of choice to select active RS variants, while the negative selection marker is used in absence of the ncAA to exclude unspecific RS variants that use endogenous amino acids as substates. As the expression of a positive selection marker—that is required for cell survival—is proportional to the efficiency of stop codon suppression, these selection schemes favor more active RS variants. For making GCE in OT organelles selective the opposite would probably be necessary.

To achieve this, it might be possible to exploit the selectivity of working OT organelles. For example, if a positive selection marker would be targeted to the organelle, while the negative selection marker is not, cells would only express the right combination if GCE is selective. Such a screen could for example be directly performed in mammalian cells using the dual color reporter I developed in this study in combination with established evolution procedures using for example adenovirus systems (*310*). Alternatively, it might be possible to adapt the synthetic organelle approach to yeast cells using established artificial phase separating systems (*260*), which would then allow to perform evolution in yeast (*311*). It might even be possible to adapt OT organelles to perform selective GCE in *E. coli*, as it was recently shown that they can also be equipped with synthetic, membraneless organelles (*312*), which would then allow to harness the full power of *E. coli*-based directed evolution.

5.1.3 Active transport is not required for OT organelle assembly

The first generation of microtubule-associated organelles was based on combining phase separation domains and kinesin motor proteins that constantly move towards the microtubule-plus end (*263*). To understand, if the active transport is required, I replaced the kinesin motor domain with EB1, a microtubule-tip binding protein as discussed in **Chapter 3**. Intriguingly, this also enables selective orthogonal translation, demonstrating that active transport is dispensable.

Kinesin-based OT organelles form a micrometer-sized, roughly spherical structure in the cell, filling up a substantial part of the cytoplasm, concomitant with a rearrangement of the microtubule cytoskeleton [**Figure 3-2**, **Appendix II**, **Fig. S8**, (*263*)]. In contrast, EB1-based OT organelles coat microtubules in a film-like fashion without rearranging the microtubule cytoskeleton.

This indicates that even much thinner OT organelles can perform orthogonal translation, which is particularly remarkable taking into account the small size of the tRNA, once again stressing how efficiently PylRS can recruit and sequester $tRNA^{Pyl}$.

5.2 Membrane associated synthetic organelles to equip cells with multiple genetic codes

In **Chapter 4** I developed membrane associated OT organelles that simultaneously address two major limitations of the GCE technology.

GCE would be particularly powerful if it would enable to incorporate multiple ncAAs into proteins *in vivo*. However, in mammalian cells this technology is fundamentally limited by the availability of codons that can be reassigned and by the dearth of orthogonal RS/tRNA pairs.

I hypothesized that equipping a cell with multiple OT organelles should address both problems simultaneously. First, it should be possible to multiple times reassign the same stop codon. Second, it should be possible to reuse the same RS multiple times, which would represent a completely new way to create functionally orthogonal enzymes.

The microtubule-based synthetic organelles did not seem to be ideally suited for this particular task, as they fill a substantial area in the cell's cytoplasm. It would thus be difficult to engineer multiple of these organelles into one cell, while still ensuring that they are immiscible, which is crucial to ensure that each organelle operates with a unique, orthogonal genetic code.

Besides phase separation, the targeting towards the cell periphery is a crucial element of the microtubule-based OT organelle design. A major element of the cell periphery is the plasma membrane and I thus hypothesized, that it might be possible to tether the OT organelle directly

to the membrane and that this should potentially serve as a new platform for selective, orthogonal translation. Gratifyingly, this is the case and it does not only work on the plasma membrane but on all intracellular membranes I tested, i.e. the ER, Golgi and outer mitochondrial membrane can serve as an orthogonal translation platform (see **Figure 4-2**).

In **Chapter 4** I also showed that instead of the ms2-MCP RNA recruitment system the λ_{N22}-based RNA recruitment system can enable selective, orthogonal translation of targeted mRNA::boxB. This system is even more efficient and selective, as it reaches an up to 17 fold selectivity with an efficiency of up to 80%, while the best tested ms2-MCP based systems only reach a selectivity of 12 fold and an efficiency of 45% (see **Figure 4-2)**.

Of the membrane associated OT film-like organelles several are functionally fully independent and in particular the OMMP OT organelle is independent from all others (**Supplementary Figure 4-4** and **Supplementary Figure 4-7**).

However, when initially combined with an ERMP or PMP OT organelle system I observed substantial cross reactivity (**Supplementary Figure 4-14**). A likely cause of this is that PylRS is active as a dimer and if dimerization precedes proper membrane localization this could cause the observed leakiness.

To overcome this, I hypothesized that in order to overwrite undesired intermolecular heterodimerization it might be possible to express PylRS as a constitutive intramolecular homodimer. To this end, I fused two copies of PylRS with the IDP FUS as a long flexible linker in between. This system substantially reduces cross reactivity and by combining this improved OMMP OT organelle system with either an ERMP or PMP system it is possible to obtain fully independent OT organelles. These allow to selectively reassign the amber codon two times in the two independent OT organelles to two distinct ncAAs.

This development represents a unique and novel way to make PylRS enzymes independent from each other without the requirement of engineering the molecular interaction surface.

In this thesis I demonstrated that this principle can be used to equip a cell with two additional orthogonal genetic codes. On top I showed that at least three membrane localizations are mutually orthogonal (OMM, PM and ERM). It should thus in principle be possible to equip a cell with three independent OT film-like organelles to equip a cell with three orthogonal genetic codes.

Besides the independent OT organelle assembly, a third independent RNA recruitment system would then additionally be necessary. In preliminary experiments that are not part of this thesis I was able to show that, analogously to ms2-MCP and boxB-λ_{N22}, stop codon containing

mRNAs can also be tagged with pp7-loops, which can be bound specifically by the PP7 bacteriophage coat protein [PCP (*313*)]. By equipping OT organelles with the PCP moiety, it is then possible to selectively translated pp7-tagged mRNAs with an expanded genetic code.
In order to harness this system to enable the genetic incorporation of three distinct ncAAs selectively, it would be necessary to additionally use three PylRS active site-mutants that have orthogonal substrate specificities. For alternative PylRS systems that lack an N-terminal domain this was recently shown in *E. coli* (*54*). However, in this study three unreactive ncAAs were used that cannot be further modified (*54*). It would be particularly interesting to instead develop synthetase variants for three different reactive ncAAs that would allow a triple color labeling approach. This, however, would require on one the hand, three mutually orthogonal click-chemistries that can be genetically encoded, and on the other hand, PylRS active site mutants that selectively recognize them. Therefore, in order to achieve this, additional evolution of PylRS would be required.

5.3 Towards the synthesis of fully artificial polymers in eukaryotes

In order to enable the synthesis of fully artificial polymers using eukaryotic ribosomes it would be necessary to have a high number of codons that can be reassigned, as well as a matching number of orthogonal RS/tRNA pairs.
In this thesis, I showed that OT film-like organelles can be used to multiple times reassign the same codon and to reuse the same type of orthogonal RS/tRNA (see **Chapter 4**). This so far only enables to incorporate two ncAAs into two distinct proteins, but not to incorporate two ncAAs into one protein. To directly enable this, it would be necessary to equip one organelle with multiple orthogonal RS systems, this is however so far not possible, as discussed in **Section 5.1.1** and **5.1.2**.
The current approach could in principle be combined with intein mediated protein-trans splicing technologies to overcome this limitation (*314*, *315*). Using inteins it would for example be possible to synthesize two parts of a POI separately in two distinct OT organelles to equip them with two distinct ncAAs. After translation these could then be spontaneously ligated to one full length protein with two specifically inserted ncAAs via intein mediated protein splicing. Even if multiple ncAAs could be inserted into one protein using synthetic organelles, such an approach might still be interesting to further amplify the protein engineering capabilities.

5.4 Biomolecular condensates can perform highly complex tasks

Besides representing a powerful and promising synthetic biology tool, the OT organelles I developed in this thesis also demonstrate the potential of biomolecular condensates in cells in general.

Several studies meanwhile elucidated the physiological function of native biomolecular condensates, for example in RNA chaperoning during stress (*207–209*), in organizing cell division components (*213*) and in mediating membrane signaling [(*233*), for a more detailed overview see **Section 1.3.3**].

The work presented in this thesis demonstrate that biomolecular condensates enable very complex processes, and can even host protein translation, one of the most complex processes taking place in cells.

Previous *in vitro* studies showed that the dimensionality reduction from 3D to 2D can lower critical concentrations for LLPS (*234*, *237*). The OT organelles developed in this study achieve their high degree of selectivity not just by phase separation, but by combining this with a spatial enrichment, which follows the logic of the described 2D phase separation processes, which have now also been presented in the literature (*232*).

Although LLPS in solution on its own is a powerful means to create local concentration gradients, this is insufficient to create high enough local concentrations to permit selective orthogonal translation [see **Appendix II**, **Fig. 2**, (*263*)]. But this can be overcome by adding a spatial enrichment component, adding either microtubule or membrane targeting in addition to the phase separation modality. It is possible that such a two layered assembly pathway could also play important roles in many native cellular processes.

5.5 Summary

In summary, in this thesis I developed the first eukaryotic, orthogonal translation platform. It is based on spatial separation instead of the development of orthogonal molecular interactions. I further used this principle to not only develop a system that is orthogonal to the canonical cytoplasmic translation, but also to develop multiple synthetic systems that are orthogonal to each other and to the endogenous pathway.

This could be particularly interesting for understanding evolutionary principles of compartmentalization endowing cells with new functions and adaptations. Spatial separation could be complementary to molecular evolution: Enzymes could first be distributed to distinct subcellular localizations and function independent from each other. Subsequently, their biochemical functionality could be diversified to evolve new functionalities.

Bibliography

1. R. Y. Tsien, The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
2. M. J. Rust, M. Bates, X. W. Zhuang, Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat Methods.* **3**, 793–795 (2006).
3. A. Keppler *et al.*, A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21**, 86–89 (2003).
4. A. Gautier *et al.*, An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **15**, 128–136 (2008).
5. G. V Los *et al.*, HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373–382 (2008).
6. B. A. Griffin, S. R. Adams, R. Y. Tsien, Specific covalent labeling of recombinant protein molecules inside live cells. *Science (80-.).* **281**, 269–272 (1998).
7. C. Uttamapinant *et al.*, A fluorophore ligase for site-specific protein labeling inside living cells. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10914–10919 (2010).
8. K. C. Gwosch *et al.*, MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells. *Nat. Methods.* **17**, 217–224 (2020).
9. F. H. C. Crick, On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–163 (1958).
10. H. M. Temin, S. Mizutani, Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus. *Nature.* **226**, 1211–1213 (1970).
11. D. Baltimore, Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. *Nature.* **226**, 1209–1211 (1970).
12. F. H. C. Crick, L. Barnette, S. Brenner, R. J. Watts-Tobin, General Nature of the Genetic Code for Proteins. *Nature.* **192**, 1227–1232 (1961).
13. C. R. Woese, G. J. Olsen, M. Ibba, D. Söll, Aminoacyl-tRNA Synthetases, the Genetic Code, and the Evolutionary Process. *Microbiol. Mol. Biol. Rev.* **64**, 202–236 (2000).
14. G. Srinivasan, C. M. James, J. A. Krzycki, Pyrrolysine encoded by UAG in archaea: Charging of a UAG-decoding specialized tRNA. *Science (80-.).* **296**, 1459–1462 (2002).
15. W. Leinfelder, E. Zehelein, M. Mandrandberthelot, A. Bock, Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature.* **331**, 723–725 (1988).
16. C. M. Dobson, Chemical space and biology. *Nature.* **432**, 824–828 (2004).
17. J. D. Fischer, G. L. Holliday, J. M. Thornton, The CoFactor database: Organic cofactors in enzyme catalysis. *Bioinformatics.* **26**, 2496–2497 (2010).
18. C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday, J. M. Thornton, Metal-MACiE: A database of metals involved in biological catalysis. *Bioinformatics.* **25**, 2088–2089 (2009).
19. H. Ryšlavá, V. Doubnerová, D. Kavan, O. Vaněk, Effect of posttranslational modifications on enzyme function and assembly. *J. Proteomics.* **92**, 80–109 (2013).
20. J. A. Ellman, D. Mendel, P. G. Schultz, Site-specific incorporation of novel backbone

structures into proteins. *Science (80-.).* **255**, 197–200 (1992).

21. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins. *Science (80-.).* **244**, 182–188 (1989).

22. K. Wals, H. Ovaa, Unnatural amino acid incorporation in E. coli: Current and future applications in the design of therapeutic proteins. *Front. Chem.* **2**, 1–12 (2014).

23. L. Wang, A. Brock, B. Herberich, P. G. Schultz, Expanding the genetic code of Escherichia coli. *Science (80-.).* **292**, 498–500 (2001).

24. J. W. Chin *et al.*, An expanded eukaryotic genetic code. *Science (80-.).* **301**, 964–967 (2003).

25. T. Mukai *et al.*, Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* **371**, 818–822 (2008).

26. P. R. Chen *et al.*, A facile system for encoding unnatural amino acids in mammalian cells. *Angew. Chemie - Int. Ed.* **48**, 4052–4055 (2009).

27. A. Gautier *et al.*, Genetically encoded photocontrol of protein localization in mammalian cells. *J. Am. Chem. Soc.* **132**, 4086–4088 (2010).

28. S. Greiss, J. W. Chin, Expanding the genetic code of an animal. *J. Am. Chem. Soc.* **133**, 14196–14199 (2011).

29. A. Bianco, F. M. Townsley, S. Greiss, K. Lang, J. W. Chin, Expanding the genetic code of Drosophila melanogaster. *Nat. Chem. Biol.* **8**, 748–750 (2012).

30. J. Liu, J. Hemphill, S. Samanta, M. Tsang, A. Deiters, Genetic Code Expansion in Zebrafish Embryos and Its Application to Optical Control of Cell Signaling. *J. Am. Chem. Soc.* **139**, 9100–9103 (2017).

31. S. Han *et al.*, Expanding the genetic code of Mus musculus. *Nat. Commun. Publ. online 21 Febr. 2017; | doi10.1038/ncomms14568.* **8**, 14568 (2017).

32. C. D. Reinkemeier, E. A. Lemke, Raising the ribosomal repertoire. *Nat. Chem.* **12**, 503–504 (2020).

33. J. W. Chin, Expanding and reprogramming the genetic code. *Nature.* **550**, 53–60 (2017).

34. E. A. Lemke, D. Summerer, B. H. Geierstanger, S. M. Brittain, P. G. Schultz, Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* **3**, 769–772 (2007).

35. J. W. Chin *et al.*, An expanded eukaryotic genetic code. *Science (80-.).* **301**, 964–967 (2003).

36. F. H. Arnold, The nature of chemical innovation: New enzymes by evolution. *Q. Rev. Biophys.* **48**, 404–410 (2015).

37. D. I. Bryson *et al.*, Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).

38. D. D. Young, P. G. Schultz, Playing with the Molecules of Life. *ACS Chem. Biol.* **13**, 854–870 (2018).

39. J. S. Italia *et al.*, An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat. Chem. Biol.* **13**, 446–450 (2017).

40. J. S. Italia *et al*., Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites. *J. Am. Chem. Soc.* **141**, 6204–6212 (2019).

41. J. S. Italia *et al*., Genetically encoded protein sulfation in mammalian cells. *Nat. Chem. Biol.* **16**, 379–382 (2020).

42. J. S. Italia, C. Latour, C. J. J. Wrobel, A. Chatterjee, Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes. *Cell Chem. Biol.* **25**, 1304-1312.e5 (2018).

43. M. Ibba, D. Söll, Aminoacyl-tRNA synthesis. *Annu. Rev. Biochem.* **69**, 617–650 (2000).

44. R. Giegé, M. Sissler, C. Florentz, Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **26** (1998), doi:10.1093/nar/26.22.5017.

45. Y. L. J. Pang, K. Poruri, S. A. Martinis, tRNA synthetase: tRNA aminoacylation and beyond. *Wiley Interdiscip. Rev. RNA.* **5**, 461–480 (2014).

46. W. Wan, J. M. Tharp, W. R. Liu, Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta.* **1844**, 1059–1070 (2014).

47. A. Ambrogelly *et al*., Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 3141–3146 (2007).

48. C. R. Polycarpo *et al*., Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **580**, 6695–6700 (2006).

49. T. Fekner, X. Li, M. M. Lee, M. K. Chan, A pyrrolysine analogue for protein click chemistry. *Angew. Chemie - Int. Ed.* **48**, 1633–1635 (2009).

50. W. Wan *et al*., A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in Escherichia coli. *Angew. Chemie.* **122**, 3279–3282 (2010).

51. W. Niu, P. G. Schultz, J. Guo, An expanded genetic code in mammalian cells with a functional quadruplet codon. *ACS Chem. Biol.* **8**, 1640–1645 (2013).

52. H. Xiao *et al*., Genetic incorporation of histidine derivatives using an engineered pyrrolysyl-tRNA synthetase. *ACS Chem. Biol.* **9**, 1092–1096 (2014).

53. J. C. W. Willis, J. W. Chin, Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. *Nat. Chem.* **10**, 831–837 (2018).

54. D. L. Dunkelmann, J. C. W. Willis, A. T. Beattie, J. W. Chin, Engineered triply orthogonal pyrrolysyl–tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids. *Nat. Chem.* **12**, 535–544 (2020).

55. V. Beránek, J. C. W. Willis, J. W. Chin, An Evolved Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase/tRNA Pair Is Highly Active and Orthogonal in Mammalian Cells. *Biochemistry.* **58**, 387–390 (2019).

56. B. Meineke, J. Heimgärtner, J. Eirich, M. Landreh, S. J. Elsässer, Site-Specific Incorporation of Two ncAAs for Two-Color Bioorthogonal Labeling and Crosslinking of Proteins on Live Mammalian Cells. *Cell Rep.* **31**, 107811 (2020).

57. B. Meineke, J. Heimgärtner, L. Lafranchi, S. J. Elsässer, Methanomethylophilus alvus Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/Aminoacyl-tRNA Synthetase Pairs in Mammalian Cells. *ACS Chem. Biol.* **13**, 3087–3096 (2018).

58. T. Plass *et al*., Amino Acids for Diels-Alder Reactions in Living Cells. *Angew. Chemie - Int. Ed.* **51**, 4166–4170 (2012).

59. K. Lang *et al*., Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic diels-alder reactions. *J. Am. Chem. Soc.* **134**, 10317–10320 (2012).

60. H. S. Jang, S. Jana, R. J. Blizzard, J. C. Meeuwsen, R. A. Mehl, Access to Faster Eukaryotic Cell Labeling with Encoded Tetrazine Amino Acids. *J. Am. Chem. Soc.* **142**, 7245–7249 (2020).

61. Y. Kurra *et al*., Two Rapid Catalyst-Free Click Reactions for in Vivo Protein Labeling of Genetically Encoded Strained Alkene/Alkyne Functionalities. *Bioconjug. Chem.* **25**, 1730–1738 (2014).

62. Y. J. Lee *et al*., Genetically encoded unstrained olefins for live cell labeling with tetrazine dyes. *Chem. Commun.* **50**, 13085–13088 (2014).

63. T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, Genetically Encoded Copper-Free Click Chemistry. *Angew. Chemie - Int. Ed.* **50**, 3878–81 (2011).

64. I. Nikic *et al*., Debugging Eukaryotic Genetic Code Expansion for Site-Specific Click-PAINT Super-Resolution Microscopy. *Angew. Chemie - Int. Ed.* **55**, 16172–16176 (2016).

65. I. Nikic *et al*., Minimal Tags for Rapid Dual-Color Live-Cell Labeling and Super-Resolution Microscopy. *Angew. Chemie - Int. Ed.* **53**, 2245–2249 (2014).

66. C. Uttamapinant *et al*., Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins. *J. Am. Chem. Soc.* **137**, 4602–4605 (2015).

67. J. Sun, M. Chen, J. Xu, J. Luo, Relationships among stop codon usage bias, its context, isochores, and gene expression level in various eukaryotes. *J. Mol. Evol.* **61**, 437–444 (2005).

68. J. D. Watson, F. H. C. Crick, Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*. **171**, 737–738 (1953).

69. N. B. Karalkar, S. A. Benner, The challenge of synthetic biology. Synthetic Darwinism and the aperiodic crystal structure. *Curr. Opin. Chem. Biol.* **46** (2018), pp. 188–195.

70. A. M. Sismour, S. A. Benner, The use of thymidine analogs to improve the replication of an extra DNA base pair: A synthetic biological system. *Nucleic Acids Res.* **33**, 5640–5646 (2005).

71. L. Zhang *et al*., Evolution of functional six-nucleotide DNA. *J. Am. Chem. Soc.* **137**, 6734–6737 (2015).

72. S. Hoshika *et al*., Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science (80-.).* **363**, 884–887 (2019).

73. N. A. Leal *et al*., Transcription, reverse transcription, and analysis of RNA containing artificial genetic components. *ACS Synth. Biol.* **4**, 407–413 (2015).

74. J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature*. **356**, 537–539 (1992).

75. K. Hamashima, M. Kimoto, I. Hirao, Creation of unnatural base pairs for genetic

alphabet expansion toward synthetic xenobiology. *Curr. Opin. Chem. Biol.* **46**, 108–114 (2018).

76. V. T. Dien, S. E. Morris, R. J. Karadeema, F. E. Romesberg, Expansion of the genetic code via expansion of the genetic alphabet. *Curr. Opin. Chem. Biol.* **46** (2018), pp. 196–202.

77. S. Moran, R. X. F. Ren, S. Rumney IV, E. T. Kool, Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication. *J. Am. Chem. Soc.* **119**, 2056–2057 (1997).

78. S. Moran, R. X. F. Ren, E. T. Kool, A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 10506–10511 (1997).

79. A. W. Feldman, F. E. Romesberg, Expansion of the Genetic Alphabet: A Chemist's Approach to Synthetic Biology. *Acc. Chem. Res.* **51**, 394–403 (2018).

80. Y. J. Seo, G. T. Hwang, P. Ordoukhanian, F. E. Romesberg, Optimization of an unnatural base pair toward natural-like replication. *J. Am. Chem. Soc.* **131**, 3246–3252 (2009).

81. A. M. Leconte *et al.*, Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *J. Am. Chem. Soc.* **130**, 2336–2343 (2008).

82. T. Lavergne *et al.*, Expanding the scope of replicable unnatural DNA: Stepwise optimization of a predominantly hydrophobic base pair. *J. Am. Chem. Soc.* **135**, 5408–5419 (2013).

83. K. Dhami *et al.*, Systematic exploration of a class of hydrophobic unnatural base pairs yields multiple new candidates for the expansion of the genetic alphabet. *Nucleic Acids Res.* **42**, 10235–10244 (2014).

84. Y. J. Seo, S. Matsuda, F. E. Romesberg, Transcription of an expanded genetic alphabet. *J. Am. Chem. Soc.* **131**, 5046–5047 (2009).

85. T. Chen, F. E. Romesberg, Polymerase Chain Transcription: Exponential Synthesis of RNA and Modified RNA. *J. Am. Chem. Soc.* **139**, 9949–9954 (2017).

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

87. Y. Zhang *et al.*, A semi-synthetic organism that stores and retrieves increased genetic information. *Nature*. **551**, 644–647 (2017).

88. E. C. Fischer *et al.*, New codons for efficient production of unnatural proteins in a semisynthetic organism. *Nat. Chem. Biol.* **16**, 570–576 (2020).

89. A. X. Z. Zhou, K. Sheng, A. W. Feldman, F. E. Romesberg, Progress toward Eukaryotic Semisynthetic Organisms: Translation of Unnatural Codons. *J. Am. Chem. Soc.* **141**, 20166–20170 (2019).

90. Y. Zhang *et al.*, A semisynthetic organism engineered for the stable expansion of the genetic alphabet. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 1317–1322 (2017).

91. T. J. Magliery, J. C. Anderson, P. G. Schultz, Expanding the genetic code: Selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in Escherichia coli. *J. Mol. Biol.* **307**, 755–769 (2001).

92. J. C. Anderson, P. G. Schultz, Adaptation of an Orthogonal Archaeal Leucyl-tRNA and

Synthetase Pair for Four-base, Amber, and Opal Suppression. *Biochemistry*. **42**, 9598–9608 (2003).

93. J. C. Anderson *et al*., *Proc. Natl. Acad. Sci*., in press, doi:10.1021/cb4001662.

94. I. A. Chen, M. Schindlinger, Quadruplet codons: One small step for a ribosome, one giant leap for proteins. *BioEssays*. **32**, 650–654 (2010).

95. J. Shine, L. Dalgarno, The 3' terminal sequence of Escherichia coli 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 1342–1346 (1974).

96. A. Hui, H. A. de Boer, Specialized ribosome system: preferential translation of a single mRNA species by a subpopulation of mutated ribosomes in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 4762–4766 (1987).

97. O. Rackham, J. W. Chin, A network of orthogonal ribosome·mRNA pairs. *Nat. Chem. Biol*. **1**, 159–166 (2005).

98. K. Wang, H. Neumann, S. Y. Peak-Chew, J. W. Chin, Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol*. **25**, 770–777 (2007).

99. H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*. **464**, 441–444 (2010).

100. S. D. Fried, W. H. Schmied, C. Uttamapinant, J. W. Chin, Ribosome subunit stapling for orthogonal translation in E. coli. *Angew. Chemie - Int. Ed*. **54**, 12791–12794 (2015).

101. C. Orelle *et al*., Protein synthesis by ribosomes with tethered subunits. *Nature*. **524**, 119–124 (2015).

102. W. H. Schmied *et al*., Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature*. **564**, 444–448 (2018).

103. N. Sonenberg, A. G. Hinnebusch, Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell*. **136**, 731–745 (2009).

104. R. J. Jackson, C. U. T. Hellen, T. V. Pestova, The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol*. **11**, 113–127 (2010).

105. J. Deforges, N. Locker, B. Sargueil, mRNAs that specifically interact with eukaryotic ribosomal subunits. *Biochimie*. **114**, 48–57 (2015).

106. F. Martin *et al*., Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. *Nat. Commun*. **7**, 1–7 (2016).

107. D. Matsuda, V. P. Mauro, Base pairing between hepatitis C virus RNA and 18S rRNA is required for IRES-dependent translation initiation in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15385–15389 (2014).

108. C. Y. M. Hu, P. Tranque, G. M. Edelman, V. P. Mauro, rRNA-complementarity in the 5′ untranslated region of mRNA specifying the Gtx homeodomain protein: Evidence that base-pairing to 18S rRNA affects translational efficiency. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1339–1344 (1999).

109. S. A. Chappell, G. M. Edelman, V. P. Mauro, A 9-nt segment of a cellular mRNA can function as an internal ribosome entry site (IRES) and when present in linked multiple

copies greatly enhances IRES activity. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 1536–1541 (2000).

110. J. Dresios, S. A. Chappell, W. Zhou, V. P. Mauro, An mRNA-rRNA base-pairing mechanism for translation initiation in eukaryotes. *Nat. Struct. Mol. Biol.* **13**, 30–34 (2006).

111. Z. Meng, N. L. Jackson, O. D. Shcherbakov, H. Choi, S. W. Blume, The human IGF1R IRES likely operates through a shine-dalgarno-like interaction with the G961 loop (E-site) of the 18S rRNA and is kinetically modulated by a naturally polymorphic polyU loop. *J. Cell. Biochem.* **110**, 531–544 (2010).

112. P. Panopoulos, V. P. Mauro, Antisense masking reveals contributions of mRNA-rRNA base pairing to translation of Gtx and FGF2 mRNAs. *J. Biol. Chem.* **283**, 33087–33093 (2008).

113. A. Bastide, A. David, Interaction of rRNA with mRNA and tRNA in translating mammalian ribosome: Functional implications in health and disease. *Biomolecules*. **8** (2018), doi:10.3390/biom8040100.

114. H. H. Wang *et al.*, Programming cells by multiplex genome engineering and accelerated evolution. *Nature*. **460**, 894–898 (2009).

115. F. J. Isaacs *et al.*, Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science (80-.).* **333**, 348–353 (2011).

116. M. J. Lajoie *et al.*, Genomically Recoded Organisms Expand Biological Functions. *Science (80-.).* **342**, 357–360 (2013).

117. M. G. Napolitano *et al.*, Emergent rules for codon choice elucidated by editing rare arginine codons in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E5588–E5597 (2016).

118. E. Scolnick, R. Tompkins, T. Caskey, M. Nirenberg, Release factors differing in specificity for terminator codons. *Proc. Natl. Acad. Sci. U. S. A.* **61**, 768–774 (1968).

119. T. M. Wannier *et al.*, Adaptive evolution of genomically recoded Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3090–3095 (2018).

120. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 6640–6645 (2000).

121. K. Wang *et al.*, Defining synonymous codon compression schemes by genome recoding. *Nature*. **539**, 59–64 (2016).

122. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).

123. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337**, 816–821 (2012).

124. G. W. Li, E. Oh, J. S. Weissman, The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*. **484**, 538–541 (2012).

125. G. Kudla, A. W. Murray, D. Tollervey, J. B. Plotkin, Coding-sequence determinants of expression in escherichia coli. *Science (80-.).* **324**, 255–258 (2009).

126. J. F. Curran, M. Yarus, Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. *J. Mol. Biol.* **209**, 65–77 (1989).

127. M. A. Sørensen, S. Pedersen, Absolute in vivo translation rates of individual codons in Escherichia coli. The two glutamic acid codons GAA and GAG are translated with a threefold difference in rate. *J. Mol. Biol.* **222**, 265–280 (1991).

128. G. Zhang, M. Hubalewska, Z. Ignatova, Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nat. Struct. Mol. Biol.* **16**, 274–280 (2009).

129. C. Kimchi-Sarfaty *et al.*, A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science (80-.).* **315**, 525–528 (2007).

130. J. Fredens *et al.*, Total synthesis of Escherichia coli with a recoded genome. *Nature.* **569**, 514–518 (2019).

131. S. M. Richardson *et al.*, Design of a synthetic yeast genome. *Science (80-.).* **355**, 1040–1044 (2017).

132. W. Zhang, L. A. Mitchell, J. S. Bader, J. D. Boeke, Synthetic Genomes. *Annu. Rev. Biochem.* **89**, 77–101 (2020).

133. B. Hampoelz, A. Andres-Pons, P. Kastritis, M. Beck, Structure and Assembly of the Nuclear Pore Complex. *Annu. Rev. Biophys.* **48**, 515–536 (2019).

134. S. Otsuka, J. Ellenberg, Mechanisms of nuclear pore complex assembly – two different ways of building one molecular machine. *FEBS Lett.* **592**, 475–488 (2018).

135. I. V. Aramburu, E. A. Lemke, Floppy but not sloppy: Interaction mechanism of FG-nucleoporins and nuclear transport receptors. *Semin. Cell Dev. Biol.* **68**, 34–41 (2017).

136. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).

137. T. Pederson, The nucleolus. *Cold Spring Harb. Perspect. Biol.* **3**, 1–15 (2011).

138. J. G. Gall, The centennial of the Cajal body. *Nat. Rev. Mol. Cell Biol.* **4**, 975–980 (2003).

139. S. Strome, R. Lehmann, Germ versus soma decisions: Lessons from flies and worms. *Science (80-.).* **316**, 392–393 (2007).

140. P. Anderson, N. Kedersha, Stress granules: the Tao of RNA triage. *Trends Biochem. Sci.* **33**, 141–150 (2008).

141. A. H. Fox *et al.*, Paraspeckles: A novel nuclear domain. *Curr. Biol.* **12**, 13–25 (2002).

142. M. E. Pepling, J. E. Wilhelm, A. L. O'Hara, G. W. Gephardt, A. C. Spradling, Mouse oocytes within germ cell cysts and primordial follicles contain a Balbiani body. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 187–192 (2007).

143. Y. Shin, C. P. Brangwynne, Liquid phase condensation in cell physiology and disease. *Science (80-.).* **357** (2017), doi:10.1126/science.aaf4382.

144. C. P. Brangwynne *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (80-.).* **324**, 1729–1732 (2009).

145. J. Berry *et al.*, RNA transcription modulates phase transition-driven nuclear body assembly. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E5237–E5245 (2015).

146. S. Alberti, A. A. Hyman, Are aberrant phase transitions a driver of cellular aging? *Bioessays.* **38**, 959–968 (2016).

147. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-liquid phase separation in biology. *Annu.*

Rev. Cell Dev. Biol. **30** (2014), pp. 39–58.

148. E. W. Martin, T. Mittag, Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry*. **57**, 2478–2487 (2018).

149. K. wing Fong *et al.*, Whole-genome screening identifies proteins localized to distinct nuclear bodies. *J. Cell Biol.* **203**, 149–164 (2013).

150. J. R. Buchan, R. Parker, Eukaryotic Stress Granules: The Ins and Outs of Translation. *Mol. Cell.* **36**, 932–941 (2009).

151. M. Hanazawa, M. Yonetani, A. Sugimoto, PGL proteins self associate and bind RNPs to mediate germ granule assembly in C. elegans. *J. Cell Biol.* **192**, 929–937 (2011).

152. A. M. Ishov *et al.*, PML is critical for ND10 formation and recruits the PML-interacting protein Daxx to this nuclear structure when modified by SUMO-1. *J. Cell Biol.* **147**, 221–233 (1999).

153. C. M. Clemson *et al.*, An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol. Cell.* **33**, 717–726 (2009).

154. T. Grouši *et al.*, Robust heat shock induces eIF2α-phosphorylation-independent assembly of stress granules containing eIF3 and 40S ribosomal subunits in budding yeast, Saccharomyces cerevisiae. *J. Cell Sci.* **122**, 2078–2088 (2009).

155. G. Dellaire, C. H. Eskiw, H. Dehghani, R. W. Ching, D. P. Bazett-Jones, Mitotic accumulations of PML protein contribute to the re-establishment of PML nuclear bodies in G1. *J. Cell Sci.* **119**, 1034–1042 (2006).

156. S. F. Banani *et al.*, Compositional Control of Phase-Separated Cellular Bodies. *Cell.* **166**, 651–663 (2016).

157. S. Boeynaems *et al.*, Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol.* **28**, 420–435 (2018).

158. C. P. Brangwynne, P. Tompa, R. V. Pappu, Polymer physics of intracellular phase transitions. *Nat. Phys.* **11**, 899–904 (2015).

159. V. N. Uversky, Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr. Opin. Struct. Biol.* **44**, 18–30 (2017).

160. M. Kato *et al.*, Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell.* **149**, 753–767 (2012).

161. H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).

162. P. Tompa, Intrinsically disordered proteins: A 10-year recap. *Trends Biochem. Sci.* **37**, 509–516 (2012).

163. P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).

164. C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584 (2014).

165. L. Malinovska, S. Kroschwald, S. Alberti, Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochim. Biophys. Acta - Proteins Proteomics.* **1834**, 918–931 (2013).

166. J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, D. T. Jones, Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **337**, 635–645 (2004).

167. S. Vucetic, C. J. Brown, A. K. Dunker, Z. Obradovic, Flavors of protein disorder. *Proteins Struct. Funct. Genet.* **52**, 573–584 (2003).

168. P. Romero *et al.*, Sequence complexity of disordered protein. *Proteins Struct. Funct. Genet.* **42**, 38–48 (2001).

169. V. N. Uversky, A. K. Dunker, Understanding protein non-folding. *Biochim. Biophys. Acta - Proteins Proteomics.* **1804**, 1231–1264 (2010).

170. H. J. Dyson, P. E. Wright, Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60 (2002).

171. P. Tompa, M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* **33**, 2–8 (2008).

172. P. S. Tan *et al.*, Two Differential Binding Mechanisms of FG-Nucleoporins and Nuclear Transport Receptors. *Cell Rep.* **22**, 3660–3671 (2018).

173. L. J. Terry, S. R. Wente, Flexible gates: Dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport. *Eukaryot. Cell.* **8**, 1814–1827 (2009).

174. S. Milles *et al.*, Plasticity of an Ultrafast Interaction between Nucleoporins and Nuclear Transport Receptors. *Cell.* **163**, 734–745 (2015).

175. N. E. Davey *et al.*, Attributes of short linear motifs. *Mol. Biosyst.* **8**, 268–281 (2012).

176. R. Pancsa, M. Fuxreiter, Interactions via intrinsically disordered regions: What kind of motifs? *IUBMB Life.* **64**, 513–520 (2012).

177. T. S. Harmon, A. S. Holehouse, M. K. Rosen, R. V. Pappu, Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *Elife.* **6** (2017), doi:10.7554/eLife.30294.

178. J. M. Choi, A. S. Holehouse, R. V. Pappu, Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu. Rev. Biophys.* **49**, 107–1033 (2020).

179. M. Rubinstein, A. V Dobrynin, Solutions of Associative Polymers. *Trends Polym. Sci.* **5**, 181–186 (1997).

180. J. Wang *et al.*, A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell.* **174**, 688-699.e16 (2018).

181. A. Crozat, P. Åman, N. Mandahl, D. Ron, Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. *Nature.* **363**, 640–644 (1993).

182. O. D. King, A. D. Gitler, J. Shorter, The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. *Brain Res.* **1462**, 61–80 (2012).

183. M. K. Andersson *et al.*, The multifunctional FUS, EWS and TAF15 proto-oncoproteins show cell type-specific expression patterns and involvement in cell spreading and stress response. *BMC Cell Biol.* **9**, 37 (2008).

184. G. Dreyfuss, M. J. Matunis, S. Pinol-Roma, C. G. Burd, hnRNP Proteins and the Biogenesis of mRNA. *Annu. Rev. Biochem.* **62**, 289–321 (1993).

185. I. F. Wang, L. S. Wu, C. K. J. Shen, TDP-43: an emerging new player in neurodegenerative diseases. *Trends Mol. Med.* **14**, 479–485 (2008).

186. A. Molliex *et al*., Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization. *Cell*. **163**, 123–133 (2015).

187. K. A. Burke, A. M. Janke, C. L. Rhine, N. L. Fawzi, Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol*. *Cell*. **60**, 231–241 (2015).

188. A. Patel *et al*., A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell*. **162**, 1066–1077 (2015).

189. M. Altmeyer *et al*., Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat*. *Commun*. **6**, 8088 (2015).

190. T. W. Han *et al*., Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies. *Cell*. **149**, 768–779 (2012).

191. Y. Lin, D. S. W. Protter, M. K. Rosen, R. Parker, Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol*. *Cell*. **60**, 208–219 (2015).

192. D. T. Murray *et al*., Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell*. **171**, 615-627.e16 (2017).

193. Y. Lin, S. L. Currie, M. K. Rosen, Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. *J*. *Biol*. *Chem*. **292**, 19110–19120 (2017).

194. C. W. Pak *et al*., Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol*. *Cell*. **63**, 72–85 (2016).

195. E. W. Martin *et al*., Valence and patterning of aromatic residues determine the phase behavior of prion-like domains Downloaded from. *Science (80-.)*. **367**, 694–699 (2020).

196. P. Anderson, N. Kedersha, Stress granules. *Curr*. *Biol*. **19**, R397–R398 (2009).

197. P. Ivanov, N. Kedersha, P. Anderson, Stress granules and processing bodies in translational control. *Cold Spring Harb*. *Perspect*. *Biol*. **11**, a032813 (2019).

198. N. Kedersha *et al*., Dynamic shuttling of TIA-1 accompanies the recruitment of mRNA to mammalian stress granules. *J*. *Cell Biol*. **151**, 1257–1268 (2000).

199. A. Khong *et al*., The Stress Granule Transcriptome Reveals Principles of mRNA Accumulation in Stress Granules. *Mol*. *Cell*. **68**, 808-820.e5 (2017).

200. S. Jain *et al*., ATPase-Modulated Stress Granules Contain a Diverse Proteome and Substructure. *Cell*. **164**, 487–498 (2016).

201. S. Markmiller *et al*., Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. *Cell*. **172**, 590-604.e13 (2018).

202. J. Y. Youn *et al*., High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol*. *Cell*. **69**, 517-532.e11 (2018).

203. B. Van Treeck *et al*., RNA self-assembly contributes to stress granule formation and defining the stress granule transcriptome. *Proc*. *Natl*. *Acad*. *Sci*. *U*. *S*. *A*. **115**, 2734–2739 (2018).

204. N. Gilks *et al*., Stress granule assembly is mediated by prion-like aggregation of TIA-1. *Mol*. *Biol*. *Cell*. **15**, 5383–5398 (2004).

205. S. Kwon, Y. Zhang, P. Matthias, The deacetylase HDAC6 is a novel critical component

of stress granules involved in the stress response. *Genes Dev*. **21**, 3381–3394 (2007).

206. N. Kedersha *et al*., G3BP-Caprin1-USP10 complexes mediate stress granule condensation and associate with 40S subunits. *J. Cell Biol*. **212**, 845–60 (2016).

207. D. W. Sanders *et al*., Competing Protein-RNA Interaction Networks Control Multiphase Intracellular Organization. *Cell*. **181**, 306-324.e28 (2020).

208. P. Yang *et al*., G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell*. **181**, 325-345.e28 (2020).

209. J. Guillé N-Boixet *et al*., RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation Article RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell*. **181**, 346–361 (2020).

210. J. A. Riback *et al*., Stress-Triggered Phase Separation Is an Adaptive , Evolutionarily Tuned Response Article Stress-Triggered Phase Separation Is an Adaptive , Evolutionarily Tuned Response. *Cell*. **168**, 1028-1040.e19 (2017).

211. N. Bley *et al*., Stress granules are dispensable for mRNA stabilization during cellular stress. *Nucleic Acids Res*. **43**, 26 (2015).

212. J. Y. Ong, J. Z. Torres, Phase Separation in Cell Division. *Mol. Cell*. **80**, 9–20 (2020).

213. J. B. Woodruff, Assembly of Mitotic Structures through Phase Separation. *J. Mol. Biol*. **430**, 4762–4772 (2018).

214. S. Petry, Mechanisms of Mitotic Spindle Assembly. *Annu. Rev. Biochem*. **85**, 659–683 (2016).

215. I. Schneider, P. Lénárt, Chromosome Segregation: Is the Spindle All About Microtubules? *Curr. Biol*. **27**, R1168–R1170 (2017).

216. J. Reichmann *et al*., Dual-spindle formation in zygotes keeps parental genomes apart in early mammalian embryos. *Science (80-.).* **361**, 189–193 (2018).

217. M. R. King, S. Petry, Phase separation of TPX2 enhances and spatially coordinates microtubule nucleation. *Nat. Commun*. **11**, 1–13 (2020).

218. J. E. Park *et al*., Phase separation of Polo-like kinase 4 by autoactivation and clustering drives centriole biogenesis. *Nat. Commun*. **10**, 1–19 (2019).

219. H. Jiang *et al*., Phase Transition of Spindle-Associated Protein Regulate Spindle Apparatus Assembly. *Cell*. **163**, 108–122 (2015).

220. E. Karsenti, I. Vernos, The Mitotic Spindle: A Self-Made Machine. *Science (80-.).* **294**, 543–547 (2001).

221. P. T. Conduit, A. Wainman, J. W. Raff, Centrosome function and assembly in animal cells. *Nat. Rev. Mol. Cell Biol*. **16**, 611–624 (2015).

222. T. Mitchison, M. Kirschner, Microtubule assembly nucleated by isolated centrosomes. *Nature*. **312**, 232–237 (1984).

223. J. Fu, I. M. Hagan, D. M. Glover, The centrosome and its duplication cycle. *Cold Spring Harb. Perspect. Med*. **5** (2015), doi:10.1101/cshperspect.a015800.

224. C. A. Kemp, K. R. Kopish, P. Zipperlen, J. Ahringer, K. F. O'Connell, Centrosome maturation and duplication in C. elegans require the coiled-coil protein SPD-2. *Dev. Cell*. **6**, 511–523 (2004).

225. M. Decker *et al*., Limiting amounts of centrosome material set centrosome size in C. elegans embryos. *Curr. Biol.* **21**, 1259–1267 (2011).

226. L. Pelletier *et al*., The Caenorhabditis elegans centrosomal protein SPD-2 is required for both pericentriolar material recruitment and centriole duplication. *Curr. Biol.* **14**, 863–873 (2004).

227. D. R. Hamill, A. F. Severson, J. C. Carter, B. Bowerman, Centrosome maturation and mitotic spindle assembly in C. elegans require SPD-5, a protein with multiple coiled-coil domains. *Dev. Cell.* **3**, 673–684 (2002).

228. J. B. Woodruff *et al*., Regulated assembly of a supramolecular centrosome scaffold in vitro. *Science (80-.).* **348**, 808–812 (2015).

229. J. B. Woodruff *et al*., The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin. *Cell.* **169**, 1066-1077.e10 (2017).

230. S. J. Enos, M. Dressler, B. F. Gomes, A. A. Hyman, J. B. Woodruff, Phosphatase PP2A and microtubule-mediated pulling forces disassemble centrosomes during mitotic exit. *Biol. Open.* **7** (2018), doi:10.1242/bio.029777.

231. M. Mittasch *et al*., Regulated changes in material properties underlie centrosome disassembly during mitotic exit. *J. Cell Biol.* **219**, 1–17 (2020).

232. L. B. Case, J. A. Ditlev, M. K. Rosen, Regulation of Transmembrane Signaling by Phase Separation. *Annu. Rev. Biophys.* **48**, 465–494 (2019).

233. L. B. Case, X. Zhang, J. A. Ditlev, M. K. Rosen, Stoichiometry controls activity of phase-separated clusters of actin signaling proteins. *Science (80-.).* **363**, 1093–1097 (2019).

234. P. Li *et al*., Phase transitions in the assembly of multivalent signalling proteins. *Nature*. **483**, 336–340 (2012).

235. C. Schwayer *et al*., Mechanosensation of Tight Junctions Depends on ZO-1 Phase Separation and Flow. *Cell.* **179**, 937-952.e18 (2019).

236. O. Beutel, R. Maraspini, K. Pombo-García, C. Martin-Lemaitre, A. Honigmann, Phase Separation of Zonula Occludens Proteins Drives Formation of Tight Junctions. *Cell.* **179**, 923-936.e11 (2019).

237. S. Banjade, M. K. Rosen, Phase transitions of multivalent proteins can promote clustering of membrane receptors. *Elife.* **3** (2014), doi:10.7554/eLife.04123.

238. X. Su *et al*., Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science (80-.).* **352**, 595–599 (2016).

239. W. Y. C. Huang *et al*., Phosphotyrosine-mediated LAT assembly on membranes drives kinetic bifurcation in recruitment dynamics of the Ras activator SOS. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 8218–8223 (2016).

240. J. A. Ditlev *et al*., A composition-dependent molecular clutch between T cell signaling condensates and actin. *Elife.* **8** (2019), doi:10.7554/eLife.42695.

241. W. Y. C. Huang *et al*., A molecular assembly phase transition and kinetic proofreading modulate Ras activation by SOS. *Science (80-.).* **363**, 1098–1103 (2019).

242. C. E. Martin, N. Jones, Nephrin signaling in the podocyte: An updated view of signal regulation at the slit diaphragm and beyond. *Front. Endocrinol. (Lausanne).* **9**, 1–12 (2018).

243. S. Kim, J. M. Kalappurakkal, S. Mayor, M. K. Rosen, Phosphorylation of nephrin induces phase separated domains that move through actomyosin contraction. *Mol. Biol. Cell.* **30**, 2996–3012 (2019).

244. G. Paci, E. A. Lemke, Shining a Light on Phase Separation in the Cell. *Cell.* **168**, 11–13 (2017).

245. D. Bracha, M. T. Walls, C. P. Brangwynne, Probing and engineering liquid-phase organelles. *Nat. Biotechnol.* **37**, 1435–1445 (2019).

246. E. Dine, A. A. Gil, G. Uribe, C. P. Brangwynne, J. E. Toettcher, Protein Phase Separation Provides Long-Term Memory of Transient Spatial Stimuli. *Cell Syst.* **6**, 655-663.e5 (2018).

247. E. M. Zhao *et al.*, Light-based control of metabolic flux through assembly of synthetic organelles. *Nat. Chem. Biol.* **15**, 589–597 (2019).

248. Y. Shin *et al.*, Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets. *Cell.* **168**, 159-171.e14 (2017).

249. D. Bracha *et al.*, Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell.* **175**, 1467-1480.e13 (2018).

250. Y. Shin *et al.*, Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell.* **175**, 1481-1491.e13 (2018).

251. L. J. Bugaj, A. T. Choksi, C. K. Mesuda, R. S. Kane, D. V. Schaffer, Optogenetic protein clustering and signaling activation in mammalian cells. *Nat. Methods.* **10**, 249–252 (2013).

252. L. S. Qi *et al.*, Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell.* **152**, 1173–1183 (2013).

253. G. Guntas *et al.*, Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 112–117 (2015).

254. G. Bellapadrona, M. Elbaum, Supramolecular protein assemblies in the nucleus of human cells. *Angew. Chemie - Int. Ed.* **53**, 1534–1537 (2014).

255. H. Yuan, C. E. Bauer, PixE promotes dark oligomerization of the BLUF photoreceptor PixD. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11715–11719 (2008).

256. S. Elbaum-Garfinkle *et al.*, The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7189–7194 (2015).

257. B. S. Schuster *et al.*, Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* **9** (2018), doi:10.1038/s41467-018-05403-1.

258. E. H. Reed, B. S. Schuster, M. C. Good, D. A. Hammer, SPLIT: Stable Protein Coacervation Using a Light Induced Transition. *ACS Synth. Biol.* **9**, 500–507 (2020).

259. J. R. Simon, N. J. Carroll, M. Rubinstein, A. Chilkoti, G. P. López, Programming molecular self-assembly of intrinsically disordered proteins containing sequences of low complexity. *Nat. Chem.* **9**, 509–515 (2017).

260. M. Heidenreich *et al.*, Designer protein assemblies with tunable phase diagrams in living cells. *Nat. Chem. Biol.* **16**, 939–945 (2020).

261. D. E. Meyer, A. Chilkoti, Purification of recombinant proteins by fusion with thermally-responsive polypeptides. *Nat. Biotechnol.* **17**, 1112–1115 (1999).

262. S. Saha, S. Banskota, S. Roberts, N. Kirmani, A. Chilkoti, Engineering the Architecture of Elastin-Like Polypeptides: From Unimers to Hierarchical Self-Assembly. *Adv. Ther.* **3**, 1900164 (2020).

263. C. D. Reinkemeier, G. E. Girona, E. A. Lemke, Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science (80-.).* **363**, eaaw2644 (2019).

264. E. Bertrand *et al.*, Localization of ASH1 mRNA particles in living yeast. *Mol. Cell.* **2**, 437–45 (1998).

265. V. Soppina *et al.*, Dimerization of mammalian kinesin-3 motors results in superprocessive motion. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 5562–7 (2014).

266. J. S. Tirnauer, S. Grego, E. D. Salmon, T. J. Mitchison, EB1-microtubule interactions in Xenopus egg extracts: Role of EB1 in microtubule stabilization and mechanisms of targeting to microtubules. *Mol. Biol. Cell.* **13**, 3614–3626 (2002).

267. S. C. Schuyler, D. Pellman, Microtubule "plus-end-tracking proteins": The end is just the beginning. *Cell.* **105** (2001), pp. 421–424.

268. J. S. Tirnauer, B. E. Bierer, EB1 proteins regulate microtubule dynamics, cell polarity, and chromosome stability. *J. Cell Biol.* **149** (2000), pp. 761–766.

269. M. R. King, S. Petry, Phase separation of TPX2 enhances and spatially coordinates microtubule nucleation. *Nat. Commun.* **11**, 1–13 (2020).

270. X. Zhang *et al.*, *bioRxiv*, in press, doi:10.1101/2020.05.04.076968.

271. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods.* **6**, 343–5 (2009).

272. I. Nikic, J. H. Kang, G. E. Girona, I. V Aramburu, E. A. Lemke, Labeling proteins on live mammalian cells using click chemistry. *Nat Protoc.* **10**, 780–791 (2015).

273. M. Schuh, J. Ellenberg, Self-Organization of MTOCs Replaces Centrosome Function during Acentrosomal Spindle Assembly in Live Mouse Oocytes. *Cell.* **130**, 484–498 (2007).

274. J. B. Pierce, S. C. Chafe, M. B. K. Eswara, G. Van der Merwe, D. Mangroo, in *Methods in Cell Biology* (Academic Press Inc., 2014), vol. 122, pp. 415–436.

275. C. Koehler, G. Estrada Girona, C. D. Reinkemeier, E. A. Lemke, *ChemBioChem*, in press, doi:10.1002/cbic.202000338.

276. H. Neumann, A. L. Slusarczyk, J. W. Chin, De Novo Generation of Mutually Orthogonal Aminoacyl-tRNA Synthetase/ tRNA pairs. *J. Am. Chem. Soc.* **132**, 2142–2144 (2010).

277. V. Beránek *et al.*, Genetically Encoded Protein Phosphorylation in Mammalian Cells. *Cell Chem. Biol.* **0** (2018), doi:10.1016/j.chembiol.2018.05.013.

278. D. Cervettini *et al.*, Rapid discovery and evolution of orthogonal aminoacyl-tRNA synthetase–tRNA pairs. *Nat. Biotechnol.* **38**, 989–999 (2020).

279. P. Zlatkine, B. Mehul, A. I. Magee, Retargeting of cytosolic proteins to the plasma membrane by the Lck protein tyrosine kinase dual acylation motif. *J. Cell Sci.* **110**, 673–

679 (1997).

280. F. Stavru, G. Nautrup-Pedersen, V. C. Cordes, D. Görlich, Nuclear pore complex assembly and maintenance in POM121- and gp210-deficient cells. *J. Cell Biol.* **173**, 477–483 (2006).

281. A. Engelsberg *et al.*, The Golgi protein RCAS1 controls cell surface expression of tumor-associated O-linked glycan antigens. *J. Biol. Chem.* **278**, 22998–23007 (2003).

282. V. Hung *et al.*, Proteomic mapping of cytosol-facing outer mitochondrial and ER membranes in living human cells by proximity biotinylation. *Elife.* **6** (2017), doi:10.7554/eLife.24463.

283. N. Daigle, J. Ellenberg, LambdaN-GFP: an RNA reporter system for live-cell imaging. *Nat. Methods.* **4**, 633–636 (2007).

284. Y. S. Wang *et al.*, Genetic incorporation of twelve meta -substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. *ACS Chem. Biol.* **8**, 405–415 (2013).

285. J. M. Kavran *et al.*, Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11268–11273 (2007).

286. D. M. Mitrea, R. W. Kriwacki, Phase separation in biology; Functional organization of a higher order Short linear motifs - The unexplored frontier of the eukaryotic proteome. *Cell Commun. Signal.* **14** (2016), doi:10.1186/s12964-015-0125-7.

287. S. Alberti, The wisdom of crowds: Regulating cell function through condensed states of living matter. *J. Cell Sci.* **130**, 2789–2796 (2017).

288. I. Nikic, J. H. Kang, G. E. Girona, I. V. Aramburu, E. A. Lemke, Labeling proteins on live mammalian cells using click chemistry. *Nat. Protoc.* **10**, 780–791 (2015).

289. A. Szymborska *et al.*, Nuclear pore scaffold structure analyzed by super-resolution microscopy and particle averaging. *Science (80-.).* **341**, 655–658 (2013).

290. M. E. Tanenbaum, L. A. Gilbert, L. S. Qi, J. S. Weissman, R. D. Vale, A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell.* **159**, 635–646 (2014).

291. X. Yan, T. A. Hoek, R. D. Vale, M. E. Tanenbaum, Dynamics of Translation of Single mRNA Molecules in Vivo. *Cell.* **165**, 976–989 (2016).

292. S. Boersma *et al.*, Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding. *Cell.* **178**, 458-472.e19 (2019).

293. N. Wu, A. Deiters, T. A. Cropp, D. King, P. G. Schultz, A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **126**, 14306–14307 (2004).

294. E. A. Lemke, D. Summerer, B. H. Geierstanger, S. M. Brittain, P. G. Schultz, Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.* **3**, 769–772 (2007).

295. T. S. Young, I. Ahmad, A. Brock, P. G. Schultz, Expanding the genetic repertoire of the methylotrophic yeast Pichia pastoris. *Biochemistry.* **48**, 2643–2653 (2009).

296. W. Liu, A. Brock, S. Chen, S. Chen, P. G. Schultz, Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat. Methods.* **4**, 239–244 (2007).

297. W. Wang *et al.*, Quantitative analysis of T Cell receptor complex interaction sites using

genetically encoded photo-cross-linkers. *ACS Chem. Biol.* **9**, 2165–2172 (2014).

298. J. M. Ho *et al.*, Efficient Reassignment of a Frequent Serine Codon in Wild-Type Escherichia coli. *ACS Synth. Biol.* **5**, 163–171 (2016).

299. S. Herring *et al.*, The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. *FEBS Lett.* **581**, 3197–3203 (2007).

300. M. Tan *et al.*, The Yin and Yang of tRNA: Proper binding of acceptor end determines the catalytic balance of editing and aminoacylation. *Nucleic Acids Res.* **41**, 5513–5523 (2013).

301. R. Jakes, A. R. Fersht, Tyrosyl-tRNA Synthetase from Escherichia coli. Stoichiometry of Ligand Binding and Half-of-the-Sites Reactivity in Aminoacylation. *Biochemistry.* **14**, 3344–3350 (1975).

302. M. Ibba, S. Sever, M. Prætorius-Ibba, D. Söll, Transfer RNA identity contributes to transition state stabilization during aminoacyl-tRNA synthesis. *Nucleic Acids Res.* **27**, 3631–3637 (1999).

303. L. T. Guo *et al.*, Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 16724–16729 (2014).

304. F. Martin, S. Barends, G. Eriani, Single amino acid changes in AspRS reveal alternative routes for expanding its tRNA repertoire in vivo. *Nucleic Acids Res.* **32**, 4081–4089 (2004).

305. J. F. Chen, N. N. Guo, T. Li, E. D. Wang, Y. L. Wang, CP1 domain in Escherichia coli leucyl-tRNA synthetase is crucial for its editing function. *Biochemistry.* **39**, 6726–6731 (2000).

306. M. Ibba, P. Kast, H. Hennecke, Substrate Specificity Is Determined by Amino Acid Binding Pocket Size in Escherichia coli Phenylalanyl-tRNA Synthetase. *Biochemistry.* **33**, 7107–7112 (1994).

307. F. Hamano-Takaku *et al.*, A mutant Escherichia coli tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine. *J. Biol. Chem.* **275**, 40324–40328 (2000).

308. R. Zúñiga, J. Salazar, M. Canales, O. Orellana, A dispensable peptide from Acidithiobacillus ferrooxidans tryptophanyl-tRNA synthetase affects tRNA binding. *FEBS Lett.* **532**, 387–390 (2002).

309. J. Xie, P. G. Schultz, A chemical toolkit for proteins - An expanded genetic code. *Nat. Rev. Mol. Cell Biol.* **7** (2006), pp. 775–782.

310. C. M. Berman *et al.*, An Adaptable Platform for Directed Evolution in Human Cells. *J. Am. Chem. Soc.* **140**, 18093–18103 (2018).

311. S. Bratulic, A. H. Badran, Modern methods for laboratory diversification of biomolecules. *Curr. Opin. Chem. Biol.* **41**, 50–60 (2017).

312. S. P. Wei *et al.*, Formation and functionalization of membraneless compartments in Escherichia coli. *Nat. Chem. Biol.* **16**, 1143–1148 (2020).

313. B. Wu, J. A. Chao, R. H. Singer, Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. *Biophys. J.* **102**, 2936–2944 (2012).

314. J. A. Gramespacher, A. J. Stevens, D. P. Nguyen, J. W. Chin, T. W. Muir, *J. Am. Chem.*

Soc., in press, doi:10.1021/jacs.7b02618.

315. N. H. Shah, T. W. Muir, Inteins: nature's gift to protein chemists. *Chem. Sci*. **5** (2014), pp. 446–461.

Appendix I Supplementary methods

In this section I am describing the methods corresponding to **Figure 5-1**.

Constructs and cloning

The dual color reporter used in this study is described in (*263*) and **Section 3.4.2**.

M. alvus $tRNA^{Pyl}$ (anticodon CUA) under the control of a human U6 promoter as well as *M. alvus* PylRS (Y126A, Y206F) were ordered from Genewiz. *M. alvus* PylRS was subsequently fused to $TOM20_{1-70}$::FUS via restriction cloning.

E. coli $tRNA^{Leu}$ (anticodon CUA) was cloned via restriction cloning under the control of a human U6 promoter. *E. coli* LeuRS (Q2E, M40G, L41Q, T252A, Y499L, Y527G, H537F) was fused to $TOM20_{1-70}$::FUS via restriction cloning.

Bacillus stearothermophilus $tRNA^{Tyr}$ (anticodon CUA) was cloned via restriction cloning under the control of a human U6 promoter, it was used as a cognate tRNA for *E. coli* TyrRS. *E. coli* TyrRS (Y37G, D182G, F183Y, L186M) was fused to $TOM20_{1-70}$::FUS via restriction cloning.

Cell culture and FFC experiments

Cell culture conditions were identical to **Section 4.4.1**.

Transfections and FFC experiments were identical to **Section 4.4.3** and **4.4.4**. For the *M. alvus* PylRS, *E. coli* LeuRS and *E. coli* TyrRS experiments the medium was not supplemented with 25 mM HEPES after ncAA addition. O-(4,5-Dimethoxy-2-nitrobenyl)-L-serine (DMNB-S, TOCRIS, 6315) and 4-benzoyl-L-phenylalanine (pBPA, Iris Biotech, HAA6010) were used at a final concentration of 250 μM, 100 mM stock solutions were prepared as described in (*288*).

For OMMP *M. mazei* PylRS the sum of three independent experiments is shown (identical to **Supplementary Figure 4-4**). For the other three synthetases the data from only one experiment are shown.

Appendix II Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes

This work was published as:

Reinkemeier CD*, Estrada Girona G* & Lemke EA (2019). Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. Science, Vol 363, Issue 6434, doi:10.1126/science.aaw2644, * These authors contributed equally.

RESEARCH ARTICLE SUMMARY

SYNTHETIC BIOLOGY

Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes

Christopher D. Reinkemeier*, Gemma Estrada Girona*, Edward A. Lemke†

INTRODUCTION: The ability to engineer translation of noncanonical (unnatural) amino acids (ncAAs) site-specifically into proteins in living cells greatly expands the chemical space that can be used to control, tailor, and study cellular function. However, translation is a complex multistep process in which at least 20 different aminoacylated tRNAs, their cognate tRNA synthetases, ribosomes, and other factors need to act in concert to synthesize a polypeptide chain encoded by an mRNA transcript. To minimize interference with the host machinery, we aimed to engineer fully orthogonal translation into eukaryotes: to encode a new functionality in response to a specific codon in only one targeted mRNA, leading to site-specific ncAA incorporation only into the selected protein of choice. Although codon specificity can be achieved with genetic code expansion (GCE), this technology relies on using an orthogonal tRNA/tRNA synthetase pair (one that does not cross-react with any of the endogenous pairs) to reprogram a stop codon. Most commonly, the Amber (UAG) stop codon is used (20% abundance in human cells), and in principle, stop codon suppression can happen for every cytoplasmic mRNA that terminates naturally on this codon. Here, we present a strategy to generate a distinctly expanded genetic code for only selected mRNAs.

RATIONALE: We hypothesized that it should be possible to create an orthogonal translation system by spatially enriching the key components of the GCE machinery in an orthogonally translating (OT) synthetic designer organelle and by targeting a specific mRNA to it. In order to perform protein translation, such an OT organelle would need to be readily accessible to the entire translational machinery of the host, thus precluding membrane encapsulation. Inspired by the concept of phase separation, which is used by cells to concentrate specific proteins and RNA locally, we hypothesized that it might be possible to use this principle to create such membraneless OT organelles. In our design, only a spatially distinct set of ribosomes associated with OT organelles can use the aminoacylated suppressor tRNA and thus will decode Amber codons only in the selected mRNA translated by the OT organelle, leading to a protein containing the ncAA.

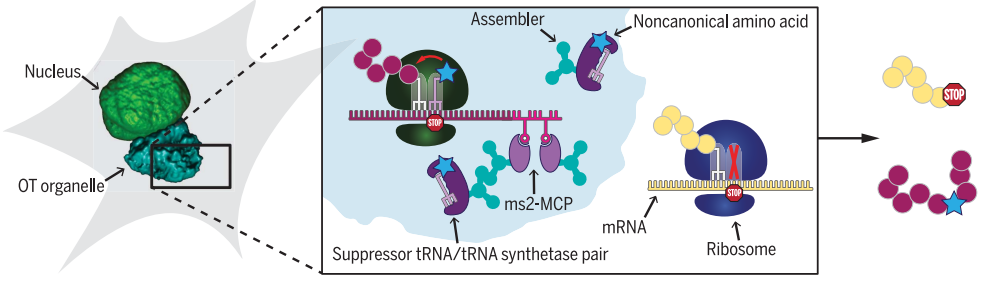
RESULTS: To bring the modified suppressor tRNA and the translated mRNA of choice in close proximity to each other, we used different strategies to generate highly concentrated assemblies and spatial separation inside cells: (i) proteins undergoing phase separation in cells [fused-in sarcoma (FUS), Ewing sarcoma breakpoint region 1 (EWSR1), and spindle-defective protein 5 (SPD5), which contain long intrinsically disordered domains] and (ii) kinesin motor proteins, which spatially enrich at microtubule plus ends (KIF13A and KIF16B). We fused each of these to the suppressor tRNA synthetase as well as an RNA-binding domain major capsid protein (MCP) that binds to a specific RNA motif (ms2 loops) engineered into the untranslated region of the mRNA of choice, forming an ms2-MCP complex. Each of these approaches yielded the desired local enrichment and preferential stop codon suppression of the mRNA tagged with ms2 loops. However, by far the best performing system was a combination of phase and spatial separation, which typically formed a micrometer-sized organelle-like structure per cell. Cells that contained this organelle efficiently and selectively performed Amber suppression of only the targeted mRNA. We were able to demonstrate the utility and robustness of these OT organelles by selectively decoding any of the three stop codons in a variety of proteins with different ncAA functionalities in two different mammalian cell lines.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaw2644

CONCLUSION: Our results show how to combine phase and spatial separation inside cells to allow the concentration of a custom designed task into a distinct specially designed membraneless organelle. We successfully demonstrated that specific and selective protein translation could be achieved within these OT organelles, which allowed the introduction of noncanonical functionalities into proteins in a codon-specific and mRNA-selective manner. The system only requires engineering five components into the cell and can be reprogrammed to other stop codons in a single step. We expect this concept to be a scalable platform for further organelle engineering and to provide a route toward generation of semisynthetic eukaryotic cells and organisms. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: edlemke@uni-mainz.de
Cite this article as C. D. Reinkemeier *et al.*, *Science* **363**, eaaw2644 (2019). DOI: 10.1126/science.aaw2644



Membraneless OT organelles enable mRNA-specific GCE in eukaryotes. OT organelles are designed organelles enriched in a suppressor tRNA/tRNA synthetase pair and a specific mRNA binding domain (MCP) by means of using an assembler protein (such as FUS and/or KIFs). A spatially distinct set of ribosomes associated with the OT organelle preferentially translates recruited mRNAs tagged with ms2 loops to yield the selected protein with the targeted site-specific noncanonical functionality.

RESEARCH ARTICLE

SYNTHETIC BIOLOGY

Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes

Christopher D. Reinkemeier[1,2,3]*, Gemma Estrada Girona[3]*, Edward A. Lemke[1,2,3]†

Nature regulates interference between cellular processes—allowing more complexity of life—by confining specific functions to organelles. Inspired by this concept, we designed an artificial organelle dedicated to protein engineering. We generated a membraneless organelle to translate only one type of messenger RNA—by recruiting an RNA-targeting system, stop codon–suppression machinery, and ribosomes—by means of phase separation and spatial targeting. This enables site-specific protein engineering with a tailored noncanonical function in response to one specific codon in the entire genome only in the protein of choice. Our results demonstrate a simple yet effective approach to the generation of artificial organelles that provides a route toward customized orthogonal translation and protein engineering in semisynthetic eukaryotic cells.

The ability to engineer orthogonal (non–cross-reactive) translation site-specifically into living cells enables the introduction of new functionalities into proteins. However, this is a herculean task because translation is a complex multistep process in which at least 20 different aminoacylated tRNAs, their cognate tRNA synthetases (RSs), ribosomes, and diverse other factors work in concert to synthesize a polypeptide chain from the RNA transcript. An ideal orthogonal system would show no cross-reactivity with factors of the host machinery, minimizing its impact on the housekeeping translational activity and normal physiology of the cell.

Toward this goal, genetic code expansion (GCE) is a method that enables reprogramming of a specific codon. With GCE, an orthogonal suppressor RS can aminoacylate its cognate suppressor tRNA with noncanonical amino acids (ncAAs). These ncAAs are typically custom designed and harbor chemical functionalities that can, for example, enable protein function to be photocontrolled, encode posttranslational modifications, or allow the introduction of fluorescent labels for microscopy studies by using click chemistry. To introduce ncAAs site-specifically into a protein of interest (POI), the anticodon loop of the tRNA is chosen to decode and thus suppress one of the stop codons [(*1*–*3*), reviews]. To minimize the impact on the host cell machinery, the Amber stop codon (TAG) is often used, owing to its particularly low abundance in *Escherichia coli*, to terminate endogenous proteins (<10%). Nevertheless, in principle any Amber codon in the transcriptome can be suppressed, potentially leading to unwanted modification of nontargeted host proteins. If ncAA-modified proteins are recombinantly produced for in vitro applications, this background incorporation can be largely ignored as long as the yields of purified full-length protein are acceptable. However, the challenge is different if the host is considered more than just a bioreactor vessel that can be sacrificed for its protein. In order to study the function of a host-cell POI in situ, the physiological condition of that host cell is an important factor. In that context, minimization of background incorporation of the ncAA is required to ensure well-controlled experiments.

At least three elegant approaches have been developed to enable orthogonal translation in *E. coli*—that is, to decode a specific codon only for the RNA of the POI and not the entire genome. (i) Orthogonal ribosomes that recognize RNA with a specific Shine-Dalgarno sequence (*4*–*6*) have been developed to enhance codon specificity, to site-specifically encode an ncAA into a POI. (ii) Recently, genome engineering has advanced to the stage that *E. coli* strains can be depleted of selected native codons (*7*–*10*), providing a genetically clean (such as Amber codon–free) host background for selective decoding of specific codons only in the POI. (iii) Distinct noncanonical codons have been designed by using an artificial base pair encoded only in the coding sequence of the POI. This lowers the risk of nonspecific decoding in other parts of the genome (*11*). However, because of genome complexity, it is not straightforward to transfer these orthogonal translation approaches to eukaryotes [(*12*), review], in which the Amber codon is highly abundant (~20% in human cells).

We hypothesized that it is possible to create an orthogonal translation system by spatially enriching specific components of the GCE machinery in an artificial orthogonally translating (OT) organelle/droplet/aggregate/condensate/dense phase. For an OT organelle to translate only the mRNA of the POI, it should be readily accessible to the translational machinery (>100 different biomolecules, such as canonical aminoacylated tRNAs, translation factors, and ribosomal subunits) and thus cannot be further easily membrane-encapsulated inside the cell. Another requirement is that the small cognate suppressor tRNA localizes efficiently to the OT organelle and is depleted from the rest of the cytoplasm.

The idea to create such an OT organelle was inspired by the concept of phase separation, which can generate high local concentrations of proteins and RNAs in cells (*13*, *14*). Recently, phase separation has gained attention owing to the discovery of its prevalence in cell biology and its role in the formation of specialized organelles such as nucleoli, stress/RNA granules, and Balbiani bodies [(*15*), review]. Although our understanding of the design and functional principles of these organelles is emerging, it has been established that they are membraneless and thus are in direct contact and exchange with the surrounding cytoplasm and/or nucleoplasm. Despite lacking a membrane, these organelles can efficiently perform complex tasks, such as transcription in the nucleolus.

We aimed to create a new OT organelle in a living mammalian cell and envisioned use of a strategy in which we selectively target the RS and the mRNA of a POI to a spatially distinct site in the cytoplasm. We found that a combination of phase separation with spatial targeting by using motor proteins yields an organelle-like structure enriched in RS and mRNA, to which the cognate suppressor tRNA and ribosomes effectively copartition. This affords a set of spatially distinct ribosomes, forming an OT system that preferentially translates only our tagged mRNA, which enables site-specific recoding of a stop codon only in this mRNA. We show for a variety of proteins, including membrane proteins, that we can incorporate site-specific noncanonical functions only into a POI, whereas other mRNAs in the cytoplasm that contain the same stop codon are not translated efficiently.

Design of OT organelles

Our synthetic designer OT organelle (Fig. 1) is engineered with the following components.

(i) An mRNA-targeting system in which two ms2 RNA stem loops are fused to the mRNA of choice, creating an mRNA::ms2 fusion coding for the POI. We denoted DNA in italics. The ms2 loops bind specifically to the phage-derived major capsid protein (MCP) (*16*), which will thus form a stable and specific mRNA::ms2–MCP complex in cells. The ms2 loops were always fused to the 3′ untranslated region (3′UTR) of the mRNA, which ensures translation to yield a scarless final POI.

[1]Biocentre, Departments of Biology and Chemistry, Pharmacy and Geosciences, Johannes Gutenberg–University Mainz, Hanns-Dieter-Hüsch-Weg 15, 55128 Mainz, Germany. [2]Institute of Molecular Biology, Ackermannweg 4, 55128 Mainz, Germany. [3]Structural and Computational Biology Unit and Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany.
*These authors contributed equally to this work.
†Corresponding author. Email: edlemke@uni-mainz.de

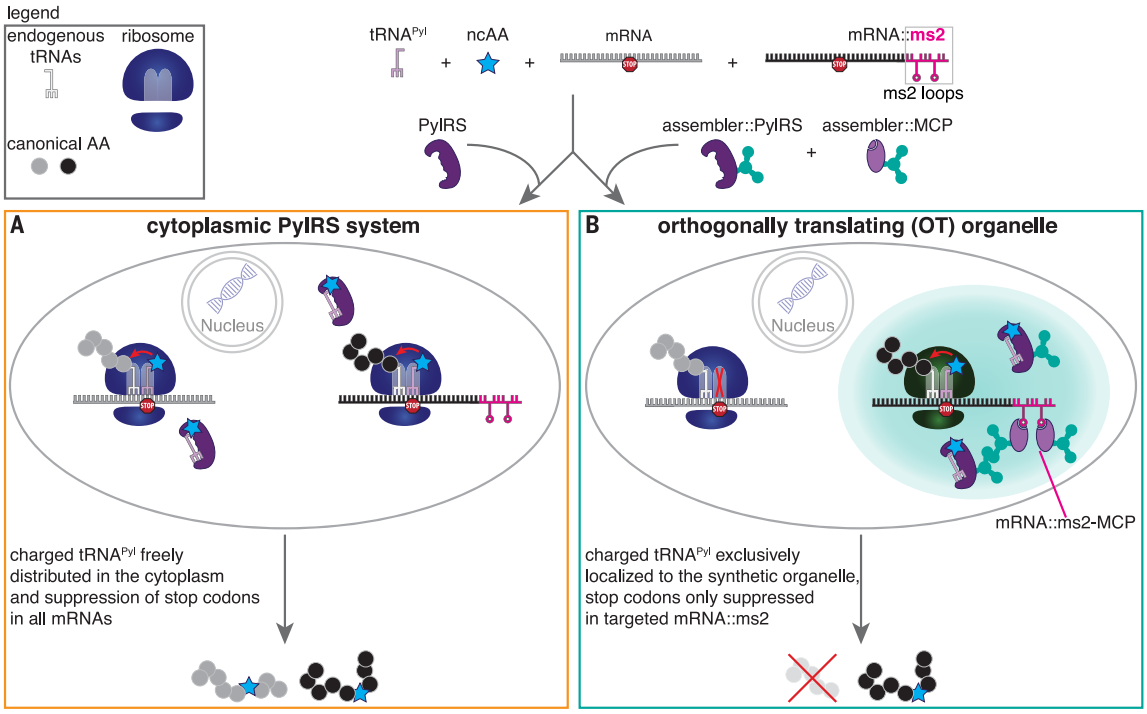


Fig. 1. Spatial separation of the necessary components to enable orthogonal translation to decode a specific stop codon in a specifically tagged mRNA. (**A**) Expression of the synthetase PylRS leads to aminoacylation of its cognate stop codon suppressor tRNA^{Pyl} with a custom designed ncAA. This leads to site-specific ncAA incorporation whenever the respective stop codon occurs in the mRNA of the POI. Given that many endogenous mRNAs terminate on the same stop codon, using this approach in the cytoplasm potentially leads to misincorporation of the ncAA into unwanted proteins. (**B**) To avoid this, we propose to spatially enrich all components to an OT organelle, including the mRNA of the POI, the aminoacyl-tRNA synthetase, the tRNA, and ribosomes through the use of "assemblers." Aminoacylated tRNA^{Pyl} should only be available in direct proximity of the OT organelle, so that only here stop codon suppression can occur. The corresponding stop codon in mRNAs that are not targeted to the OT organelle should not get translated. Whereas in (A) GCE is stop codon specific, in (B), it is stop codon– and mRNA-specific.

(ii) A tRNA/RS suppressor pair. We chose the orthogonal tRNA/RS pair from the *Methanosarcina mazei* pyrrolysyl system (tRNA^{Pyl}/PylRS) because it has enabled the encoding of more than 100 ncAAs with diverse functionalities into proteins by using GCE in a multitude of cell types and species, including *E. coli*, mammalian cells, and even living mice [(*1*–*3*), reviews].

(iii) The assembler, the key component required to form an OT organelle. The purpose of the assembler is to create a dense phase or condensate, in which the mRNA::ms2–MCP complex is brought into close proximity of the tRNA^{Pyl}/PylRS pair.

The simplest assembler strategy we tested is the bimolecular fusion of MCP::PylRS (termed **B**) (Fig. 2A). In addition, we tested strategies in which we expected to yield much larger assemblies. All of those assembly systems are composed of an assembler fusion to PylRS coexpressed with an assembler fusion to MCP. We expected assembler::PylRS•assembler::MCP to form large aggregates (we denote coexpression with a center dot, "•"). One tested assembly strategy was based on phase separation of proteins, and one was based on the assembly of kinesins, which we abbreviate here as **P** and **K**, respectively (Fig. 2A). Furthermore, for each **P** and **K** approach, we tested two different molecular designs: **P1**, **P2** and **K1**, **K2**, respectively.

P1

Previous studies have established the capacity of the proteins fused-in sarcoma (FUS) and Ewing sarcoma breakpoint region 1 (EWSR1) to form mixed dropletlike structures by means of phase separation. They both contain a prion-like disordered domain that facilitates phase separation into liquid, gel, and solid states (*17*, *18*). In a phase-separated state, these proteins are locally highly concentrated (approximately orders of magnitude) compared with the remaining soluble fraction in the cell. FUS was fused to PylRS, and EWSR1 fused to MCP, and we speculated that this would lead to the formation of droplets in which MCP and PylRS are highly enriched. **P1** is denoted FUS::PylRS•EWSR1::MCP.

P2

The *Caenorhabditis elegans* protein spindle-defective protein 5 (SPD5) has recently been shown to phase-separate into particularly large (several micrometer-sized) droplets (*19*). In a phase-separated state, SPD5 is locally highly concentrated compared with the remaining soluble fraction in the cytoplasm (by orders of magnitude). We speculate that a protein fused to SPD5 will condense into droplets. Similarly to FUS-EWSR1 droplets, we speculated that PylRS fused to SPD5 and MCP fused to SPD5 will be highly enriched. **P2** is denoted SPD5::PylRS•SPD5::MCP.

K1

Certain kinesin truncations constitutively move toward microtubule-plus ends in living cells (*20*). One such truncated kinesin is KIF13A$_{1–411,\Delta P390}$, and we speculated that PylRS and MCP respectively fused to this kinesin truncation and coexpressed would be locally enriched, owing to spatial targeting to microtubule-plus ends. **K1** is denoted KIF13A$_{1-411,\Delta P390}$::PylRS•KIF13A$_{1-411,\Delta P390}$::MCP. (Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; F, Phe; N, Asn; P, Pro; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example, Y306A indicates that tyrosine at position 306 was replaced by alanine).

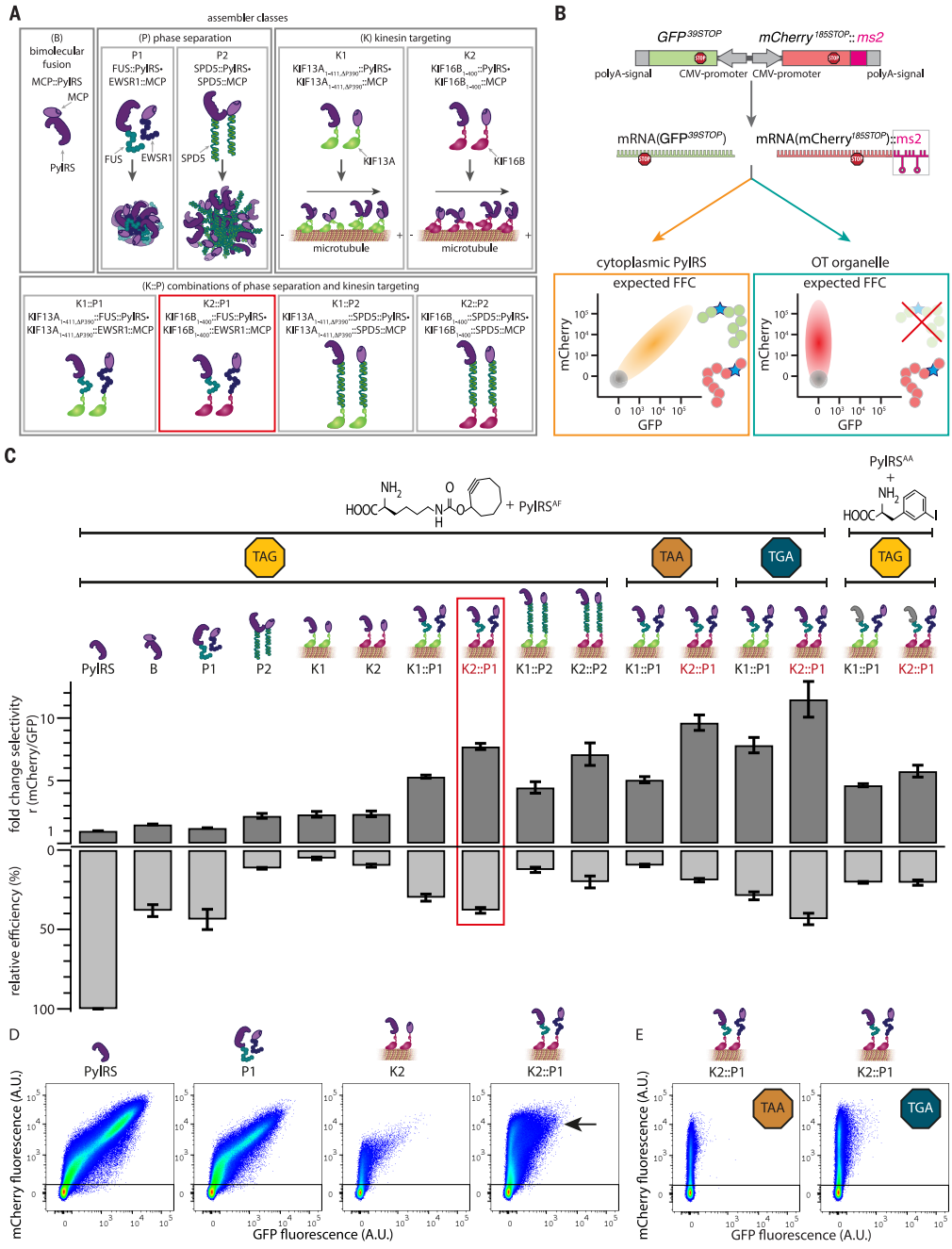


Fig. 2. Local enrichment by means of phase separation is a means to generate OT organelles. (**A**) Schematic representation of different assembler classes. **B**, bimolecular MCP::PylRS fusion; **P1**, fusions to FUS and EWSR1; **P2**, SPD5; **K1**, truncation of kinesin KIF13A ($KIF13A_{1\text{-}411,\Delta P390}$); **K2**, truncation of kinesin KIF16B ($KIF16B_{1\text{-}400}$) and combinations thereof (**K1::P1**, **K1::P2**, **K2::P1**, and **K2::P2**). (**B**) Schematic representation of the dual-color reporter. mRNAs encoding the fluorescent proteins GFP and mCherry, containing stop codons at permissive sites, are expressed from one plasmid, each with its own CMV promoter, ensuring a constant ratio of mRNA throughout each experiment. The mRNA of the mCherry reporter is tagged with two ms2 stem loops, mCherry::ms2. In the presence of ncAA and $tRNA^{Pyl}$, in the case of cytoplasmic PylRS, both GFP^{39STOP} and $mCherry^{185STOP}$ are produced, leading to a diagonal (schematically drawn in orange) in FFC analysis. However, under the same conditions, orthogonal translation in OT organelles should enable selective stop codon suppression of mCherry::ms2 mRNA, resulting in an mCherry-positive and GFP-negative population (drawn schematically as a red vertical population). In both schemes, nontransfected HEK293T cells, which are also detected with FFC, are represented by a gray circle. (**C**) For all experiments, the indicated constructs were coexpressed with $tRNA^{Pyl}$ (anticodon corresponding to the indicated codon) and the dual reporter (GFP^{39STOP}, $mCherry^{185STOP}$::*ms2*). GCE was performed in presence of the indicated ncAAs, and cells were analyzed by means of FFC. The dark gray bars (normalized to cytoplasmic PylRS) represent the fold change in the ratios *r* of the mean fluorescence intensities of mCherry versus GFP (derived from FFC) (Fig. 2, D and E, and fig. S1) for all the systems tested in this study. The light gray bars represent the relative efficiency as defined by the mean fluorescence intensity of mCherry for each condition divided by cytoplasmic PylRS control (derived from FFC) (Fig. 2, D and E, and fig. S1). Shown are the mean values of at least three independent experiments; error bars represent the SEM. The red box highlights the best performing OT organelle ($OT^{K2::P1}$). (**D**) FFC analysis of the dual-color reporter expressed with the four indicated systems in transfected HEK293T cells and $tRNA^{Pyl}$ in the presence of the ncAA SCO, a lysine derivative with a cyclooctyne side chain. Highly selective and efficient orthogonal translation was observed for the OT organelle (the black arrow indicates a bright, highly mCherry-positive population). Shown in the dot plots are the sums of at least three independent experiments. Axes indicate fluorescence intensity in arbitrary units (all FFC plots are summarized in fig. S1). (**E**) FFC plots for the OT organelle selectively translating Opal and Ochre codons only of recruited $mCherry^{185TGA}$::ms2 and $mCherry^{185TAA}$::ms2 mRNA, respectively (corresponding cytoplasmic PylRS controls for those stop codons are provided in fig. S1).

K2

By analogy to **K1**, we also tested the truncated kinesin KIF16B$_{1–400}$. **K2** is denoted KIF16B$_{1–400}$::PylRS•KIF16B$_{1–400}$::MCP.

We also tested whether combinations of these systems would lead to efficient OT organelles:

K1::P1 = KIF13A$_{1–411,\Delta P390}$::FUS::PylRS•KIF13A$_{1–411,\Delta P390}$::EWSR1::MCP

K2::P1 = KIF16B$_{1–400}$::FUS::PylRS•KIF16B$_{1–400}$::EWSR1::MCP

K1::P2 = KIF13A$_{1–411,\Delta P390}$::SPD5::PylRS•KIF13A$_{1–411,\Delta P390}$::SPD5::MCP

K2::P2 = KIF16B$_{1–400}$::SPD5::PylRS•KIF16B$_{1–400}$::SPD5::MCP

In order to evaluate these assemblers for facilitating functional orthogonal translation of an ms2-tagged mRNA, we designed a dual-reporter construct, in which green fluorescent protein (GFP) and mCherry mutants are simultaneously expressed from two different expression cassettes from one plasmid, ensuring that the mRNA ratio between them is constant across all experiments. Stop codons were introduced at permissive sites into GFP at position 39 (GFP^{39STOP}) and into mCherry at position 185 (mCherry^{185STOP}) (Fig. 2B). Only if stop codon suppression is successful will the corresponding GFP or mCherry be produced. Transfected cells were analyzed by means of fluorescence flow cytometry (FFC); settings were adjusted so that an approximate diagonal results in the FFC plots if GFP and mCherry are expressed from this plasmid by using the conventional cytoplasmic PylRS system, which cannot differentiate mRNAs. A selective and functional OT organelle should selectively express mCherry only if the ms2 loops are fused to the 3′UTR of the mCherry mRNA, leading to appearance of a vertical line in the cytometry plot (Fig. 2B). Unless otherwise reported, all experiments were performed in the presence of tRNA^{Pyl} and the ncAA SCO, a widely used and well-characterized lysine derivative, the side chain of which carries a cyclooctyne that can be used in a variety of click-chemistry reactions to install diverse chemical groups onto the protein. As previously reported, this ncAA is efficiently encoded by a Y306A, Y384F double mutant of PylRS (for simplicity we refer to this PylRS^{AF} mutant as PylRS unless otherwise specified) (*21*–*23*). Omission of the ncAA serves as a standard negative control and leads to no expression of GFP or mCherry (fig. S1).

We evaluated the performance of each OT system according to their selectivity and relative efficiency. We define selectivity as the ratio r of the mean mCherry FFC signal divided by the mean GFP signal. Final values are expressed as fold selectivity relative to that of cytoplasmic PylRS. We define relative efficiency as the mean mCherry signal of each system divided by the mean mCherry signal of the cytoplasmic PylRS system, which serves as the reference (here defined as 100%). All results on selectivity (Fig. 2C, dark gray positive bars) and efficiency (Fig. 2C, light gray negative bars) are summarized in the bar plot in Fig. 2C, and all corresponding FFC dot plots are summarized in fig. S1, whereas selected FFC data are also shown in Fig. 2, D and E.

Combining two assembler strategies yields highly selective and efficient OT systems

The conceptually simplest assembly strategy **B** (MCP fused to PylRS) showed only a minor selectivity gain of about 1.5-fold, which is concomitant with a 60% decrease in efficiency (Fig. 2C and fig. S1). The OT system **P1** (based on phase separation of FUS/EWSR1) performed similarly in terms of selectivity gain. In addition, a 50% decrease in efficiency was measured (Fig. 2, C and D). The **P2** system (based on SPD5) showed an approximate twofold selectivity gain accompanied by an almost 90% decrease in efficiency (Fig. 2C and fig. S1). Analogously, we tested the kinesin-based assembly strategy and observed for **K1** a twofold selectivity increase, with an efficiency decrease of ~90% (Fig. 2C and fig. S1). The **K2** system behaved similarly (Fig. 2, C and D). In total, the selectivity gains were small but robustly detected, indicating that bringing the ncAA aminoacylation activity (the tRNA^{Pyl}/PylRS in the presence of ncAA) in direct proximity of the target mRNA represents a pathway to more selective codon suppression.

Next, we tested the assembler combination strategies (**K1::P1**, **K2::P1**, **K1::P2**, and **K2::P2**). For all combinations, we observed at least fivefold selectivity gain, indicating orthogonal translation [the observed selectivity effect is robust across a titration of Amber suppression efficiencies (fig. S2)]. The best performing system on the basis of the fusion of FUS/EWSR1 with KIF16B$_{1-400}$, **K2::P1**, exhibited a selectivity of eightfold and 40% efficiency (Fig. 2C, red box). This was also directly obvious from the FFC data, in which the bright, mCherry-positive cell population was clearly retained, whereas GFP expression was minimal (Fig. 2D, black arrow).

To validate that the observed selectivity gain is specific to the ms2-MCP interaction, we further characterized OT organelles by expressing the RS assembler fusion of each OT system without MCP (figs. S1 and S3). As expected, no selective orthogonal translation of ms2-tagged mRNA was observed in those cases. Additionally, we performed a reporter inversion by moving the ms2 loops from the mCherry to the GFP cassette in the dual-color reporter, which as expected inverted selectivity of the system toward dominant GFP expression (fig. S3B). This establishes that the OT system acts selectively on the ms2-tagged RNA.

GCE can also be used to introduce multiple ncAAs into the same POI [(*1*–*3*), reviews]. However, only very few publications report on more than one—that is, two- or three-codon suppression in the same protein in eukaryotes—because yields typically suffer compared with single-codon suppression (*24*–*26*). Even dual- and triple-Amber proteins were still suppressed with the OT organelle (fig. S4).

To ensure that other ncAAs also can be incorporated by the OT system, we tested another structurally different ncAA (3-iodophenylalanine), which is a phenylalanine derivative instead of a lysine derivative and is encoded by a different PylRS mutant (N346A and C348A) (*27*). Consistent results were also observed for this system (Fig. 2C and fig. S1).

Because Opal (TGA) and Ochre (TAA) codons are highly abundant in eukaryotic genomes (~52% Opal, ~28% Ochre in the human genome), the Amber codon is by far the most used for GCE in eukaryotes. In addition, genomic approaches to orthogonal translation by removing these codons in the entire eukaryotic genome would be even more challenging than for the Amber codon and are currently beyond the state of the art. However, in the OT system, a simple mutation in the anticodon loop of the tRNA^{Pyl}, as well as in the respective codon in the mRNA::ms2, should allow orthogonal translation of these codons. FFC analysis revealed that the OT organelles provide freedom of choice with respect to the stop codon (Fig. 2, C and E, and fig. S1). In fact, Opal suppression showed an 11-fold selectivity increase at 50% efficiency (slightly outperforming Amber suppression). Ochre suppression still showed fivefold selectivity increase, with 20% efficiency.

The OT$^{K2::P1}$ organelle enables orthogonal translation of proteins of various cellular compartments

To visualize the power of the OT$^{K2::P1}$ organelle (our best performing Amber suppression OT organelle in terms of selectivity and efficiency) beyond "simple" reporters, we next aimed to show differential expression of human nucleoporin 153 (Nup153) versus cytoskeletal vimentin. Nup153 locates to the nuclear pore complex and is more than 1500 amino acids long. Hence, its mRNA is approximately sixfold larger than those of the fluorescent protein reporters used above. We used a previously described C-terminal GFP fusion, with an Amber mutation (Nup153::GFP^{149TAG}) that gave rise to a characteristic nuclear envelope stain in confocal images only if Amber suppression was successful (*28*). Nup153::GFP^{149TAG} was then tagged with two ms2 loops (*NUP153::GFP^{149TAG}::ms2*) and coexpressed from the same plasmid with vimentin containing an Amber codon at position 116 fused to mOrange (*VIM^{116TAG}::mOrange*). Expression in human embryonic kidney (HEK) 293T cells resulted in production of both proteins in the presence of the cytoplasmic PylRS showing the characteristic nuclear envelope and cytoskeletal staining, respectively. In the presence of the OT$^{K2::P1}$ organelle, only Nup153::GFP was visible (Fig. 3A, selective nuclear rim stain). Consistent results were also observed in *Cercopithecus aethiops* kidney (COS-7) cells (fig. S5). Swapping the ms2 loops to vimentin inverted the effect, so that only Vimentin^{116TAG}::mOrange was visible (further experiments for COS-7 and for HEK293T cell experiments are shown in Fig. 3B and fig. S5). This showed that the OT$^{K2::P1}$ worked for dramatically different mRNAs.

Next, we asked whether it would be possible to selectively express transmembrane proteins by

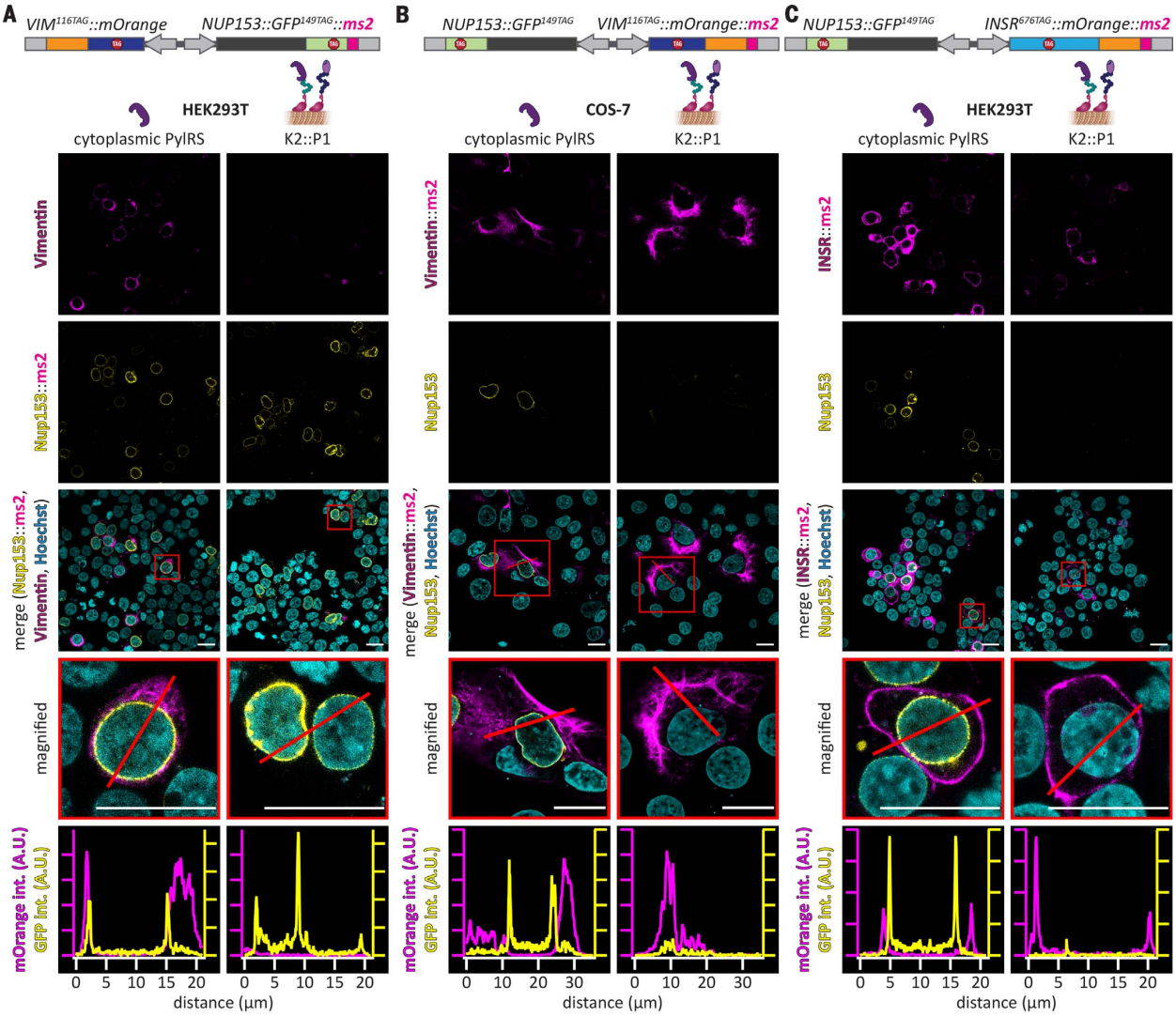


Fig. 3. A versatile OT organelle for selective and efficient orthogonal translation. (**A** to **C**) Confocal images of cells transfected with constructs encoding PylRS (left column) or the OT$^{K2::P1}$ organelle (right column) for different protein pairs. SCO and tRNA^{Pyl} were present in all cases. (A) HEK293T cells were transfected with *NUP153::GFP*^{149TAG}*::ms2* and *VIM*^{116TAG}*::mOrange*. (B) *VIM*^{116TAG}*::mOrange::ms2* and *NUP153::GFP*^{149TAG} transfected in COS-7 cells. More representative examples for (A) using COS-7 cells and for (B) using HEK293T cells are shown in fig. S5. Shown from top to bottom are Vimentin^{116TAG}::mOrange (magenta, characteristic cytoskeletal stain), Nup153::GFP^{149TAG} (yellow, characteristic nuclear envelope stain), overlay with Hoechst (cyan, nuclear stain), magnified images of representative cells (red boxes), and line profiles for the mOrange and GFP channel (red line, magenta and yellow curves, respectively), to highlight that only the ms2-tagged mRNA yields its respective expressed protein if the OT$^{K2::P1}$ organelle is present. Scale bars, 20 μm. (C) HEK293T cells were transfected with *INSR*^{676TAG}*::mOrange::ms2* and *NUP153::GFP*^{149TAG}. Shown from top to bottom are INSR^{676TAG}::mOrange (magenta, characteristic plasma membrane stain), Nup153::GFP^{149TAG} (yellow), overlay with Hoechst (cyan), magnified images of representative cells (red boxes), and line profiles for the mOrange and GFP channel (red line, magenta and yellow profiles, respectively), demonstrating selective translation of insr^{676TAG}::mOrange::ms2 mRNA by the OT$^{K2::P1}$ organelle. Scale bars, 20 μm.

using our synthetic OT$^{K2::P1}$ organelle. Membrane protein expression represents another layer of translational complexity because ribosomes need to bind the endoplasmic reticulum (ER) during translation, where the proteins are cotranslationally inserted into the membrane. To this end, we used a fusion of insulin receptor 1 with an Amber codon at position 676 with mOrange (INSR^{676TAG}::mOrange), which locates to the plasma membrane and gives rise to a characteristic plasma membrane stain in HEK293T cells (*21*). This construct was tagged with ms2 loops in the 3′UTR and cloned with Nup153::GFP^{149TAG} into one dual-cassette plasmid. We then expressed it in HEK293T cells either in the presence of the cytoplasmic PylRS system or in the presence of the OT$^{K2::P1}$ organelle (Fig. 3C). In the presence of the synthetic OT$^{K2::P1}$ organelle, we observed selective expression of the ms2-tagged protein and the expected plasma membrane localization of INSR^{676TAG}::mOrange, indicating the potential of our organelle to participate in even more complex membrane-associated translational processes.

The OT organelle functions by recruiting ribosomes and tRNA^{Pyl}

The above experiments demonstrated the functionality of our synthetic OT$^{K2::P1}$ organelle. Next, we aimed to study the spatial distribution of the different systems in the cell to understand the basic working mechanism of the OT organelle.

To assess the spatial distribution of PylRS in cells, we used immunofluorescence (IF). We also used fluorescence in situ hybridization (FISH) against tRNA^{Pyl}. In contrast to the dual-color reporter used in the FFC experiments above, in all IF/FISH experiments we used a single-color NLS::GFP^{39TAG} reporter fused to ms2 loops to identify cells active in Amber suppression (this yields a green nucleus if Amber suppression is successful). IF and FISH stainings showed that in contrast to cytoplasmic PylRS, the **P1** system formed small, intracellular assembler::PylRS droplets (Fig. 4A). This indicated the occurrence of phase separation. The tRNA^{Pyl} colocalized well with assembler::PylRS droplets (Fig. 4A), indicating that it could nicely partition into the assembler::PylRS phase. Additional stainings show further colocalization with assembler::MCP (fig. S6). Compared with **P1**, the **P2** system (figs. S6 and S7) showed larger but still multiple dispersed dropletlike structures. Phase separation of proteins is based on exceeding the critical concentration up to which proteins are fully soluble in the cytoplasm. However, a soluble species coexists with the phase-separated species (*29*) that can contribute to stop codon suppression outside the droplet. The components of the **K1** (fig. S7) and **K2** (Fig. 4A and fig. S6) systems were mostly observed distributed across the cytoplasm, likely because of binding to the microtubule cytoskeleton, which appears rather distributed throughout the cytoplasm (fig. S8). Because small tRNAs can diffuse rapidly, we believe that a critical factor for the design of an OT organelle is how well the tRNA^{Pyl} is confined to few sites in the cell and thus spatially separated and sequestered from the rest of the cytoplasm. None of the **P1**, **P2**, **K1**, or **K2** systems displaying high selectivity (Fig. 2C) is consistent with the observation that the systems showed a rather dispersed distribution in the cell.

However, if we combined both assembler strategies (**K1::P1**, **K2::P1**, **K1::P2**, and **K2::P2**), we observed the formation of large micrometer-sized, organelle-like structures in the cytoplasm, which were in most cases localized to few or even a single position per cell. Association of the OT organelle with the microtubule cytoskeleton was also observed (fig. S8). As shown for **K2::P1** in Fig. 4A and figs. S6 and S8, mRNA::ms2, tRNA^{Pyl}, assembler::PylRS, and assembler::MCP all colocalize to organelle-like structures (other combined assemblers are shown in figs. S6 and S7). The combination of the two assembler strategies—that is, phase separation paired with spatial targeting by kinesin truncations—yielded the best confinement as determined with FISH and IF and the highest selectivity increase.

This is consistent with our hypothesis that the higher spatial segregation and thus higher local concentration of aminoacylated tRNA^{Pyl} and mRNA correlates with higher selectivity. This effectively translates into a higher partition coefficient of tRNA^{Pyl} into the droplet and thus depletion of tRNA^{Pyl} from the cytoplasm, yielding a high concentration gradient between cytoplasm and OT organelle (fig. S7).

We next performed staining for ribosomes to see whether they colocalize to the OT$^{K2::P1}$ organelle. IF staining of the ribosomal protein RPL26L1 showed its distribution throughout the cell but also an enrichment at the OT$^{K2::P1}$ organelle [Fig. 4, B and C, two-dimensional (2D) projection and a 3D reconstruction; movie S1; and fig. S8), demonstrating partial ribosome recruitment, tentatively owing to binding to mRNA::ms2 during translation. We conclude that only ribosomes sufficiently immersed into the tRNA^{Pyl} gradient can perform codon suppression efficiently. This also visualizes the mobility of a set of ribosomes in the cell. High mobility can also explain why we were even able to express the membrane protein INSR (Fig. 3C) because for orthogonal translation of a membrane protein, two things must happen either sequentially or at the same time: The translating ribosome needs to interact with the ER and with the OT organelle. Together, this body of evidence strongly suggests that selective orthogonal translation happens within close proximity of the OT organelles, potentially even inside the organelle, by a set of recruited ribosomes that are near or fully immersed into a concentrated pool of tRNA^{Pyl}. tRNA^{Pyl} itself is recruited to the OT$^{K2::P1}$ organelle because of its affinity for assembler::PylRS and can readily copartition into the droplet to be aminoacylated with its cognate ncAA, whereas assembler::MCP recruits ms2-tagged mRNA. This in turn attracts ribosomes to migrate to the dense phase formed by the dual assembler system (**K2::P1** = KIF16B::FUS::PylRS and KIF16B::EWSR1::MCP), which maintains access to other translation factors for translation to function (Fig. 4D). Ribosomes elsewhere in the cytoplasm that are not exposed to tRNA^{Pyl} perform their canonical function to terminate translation whenever they encounter a stop codon.

Our route to enable orthogonal translation required only five extra components and represents an important step toward generating semisynthetic eukaryotic organisms that can potentially follow what has been dubbed the "orthogonal central dogma" (*30*). The OT organelle also represents a general strategy for tailoring complex functions in eukaryotes by mimicking the evolutionary concept to build distinct, but membraneless, organelles inside eukaryotic cells. Proteins such as FUS, EWSR1, and SPD5 have many vital functions in the cell, and we cannot exclude the possibility that the lower expression yield (half, in many cases) we observed is also due to overexpression of components of the OT organelle. The need to combine two different assembly strategies (phase separation–based assemblers with spatial targeting by using kinesin truncations) puts potentially an additional burden on the cell. However, our understanding of phase separation is developing dramatically, such as the identification of amino acid sequences that can be used to make layered droplets (*31*). We can thus expect future versions of this technology to afford even better selectivity, efficiency, and the ability to bestow additional functions on the OT organelle or for constructing other novel organelles with new functions.

Materials and methods

Cell culture

HEK293T cells (ATCC CRL-3216) and COS-7 cells (Sigma-Aldrich 87021302) were maintained in Dulbecco's modified Eagle's medium (DMEM, Gibco 41965-039) supplemented with 1% penicillin-streptomycin (Sigma-Aldrich P0781), 1% L-Glutamine (Sigma-Aldrich G7513), 1% sodium pyruvate (Life Technologies 11360), and 10% FBS (Sigma-Aldrich F7524). Cells were cultured at 37°C in a 5% CO_2 atmosphere and passaged every 2-3 days up to 20 passages.

In all cases, cells were seeded 15-20 hours prior to transfection at a density resulting in 70-80% confluency at the time of transfection. Flow cytometry was performed using 24-well plates with plastic bottom (Nunclon Delta Surface ThermoScientific). IF labeling and FISH were performed on 24-well plates with glass bottom (Greiner Bio-One) or four-well chambered Lab-Tek #1.0 borosilicate coverglass (ThermoFisher).

Constructs, cloning, and mutagenesis

Dual-color reporters: The dual fluorescent protein reporters were cloned in a pBI-CMV1 vector (Clontech 631630), with ms2 tagged fluorescence protein (mRNA) version in one multiple cloning site and ms2 free version in the other. *GFP^{39TAGI}* or *mCherry^{185TAG}* were used as N-terminal fusions with nuclear localization sequences (NLS). Similar reporters for Ochre and Opal suppression were prepared (with *GFP^{39TAA}*, *mCherry^{185TAA}* and *GFP^{39TGA}*, *mCherry^{185TGA}*, respectively).

NLS::GFP^{39TAG}::ms2 reporter: *NLS::GFP^{39TAG}* was cloned with two copies of ms2 loops into the pBI-CMV1 vector as a reporter for successful Amber suppression for imaging experiments.

Double and triple Amber *GFP*: Position 149 and subsequently position 182 of *GFP* were mutated to TAG to obtain pBI-CMV constructs with multiple amber codons, these constructs did not have *mCherry* in the second multiple cloning site.

OT organelle constructs: Pyrrolysyl tRNA was cloned under the control of a human U6 (hU6) promoter, and all other constructs were under CMV (cytomegalovirus) promoters cloned in the pcDNA3.1 (Invitrogen V86020) vector. *MCP* was cloned from the Addgene plasmid #31230 and *FUS* from the Addgene plasmid #26374. In all *FUS* fusions, amino acids 1-478 were used, replacing the C-terminal NLS region by a Flag-tag. In all pyrrolysine synthetase fusions the previously reported efficient *NES::PylRS^{AF}* (Y306A, Y384F) sequence was used (*21*, *28*). *NES::PylRS^{4A}* (N346A, C348A) was cloned via site-directed mutagenesis starting from *NES::PylRS^{WT}*. The *SPD5* gene was ordered from Genewiz and fused

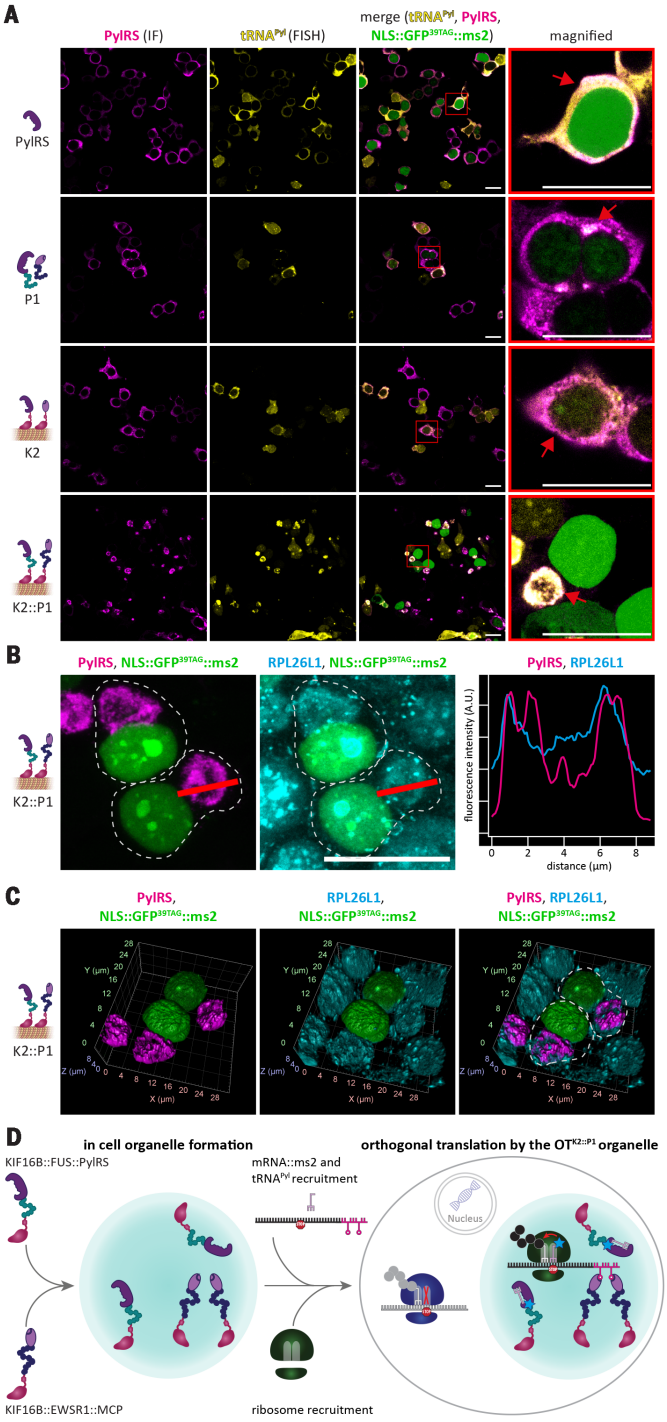


Fig. 4. OT organelles enrich tRNA^{Pyl} and ribosomes for orthogonal translation. (**A**) IF and FISH imaging of HEK293T cells expressing tRNA^{Pyl} and the indicated system. For simplicity and to direct the eye, in all imaging experiments, a simple NLS::GFP^{39TAG}::ms2 (nuclear staining) was used instead of the dual-color reporter. Green-colored nuclei report on faithful Amber suppression of the cells (shown only in overlay). Shown from left to right are IF against PylRS in magenta, FISH against tRNA^{Pyl} in yellow, overlay, and magnified images of representative cells (red boxes and red arrows highlight representative structures). Scale bars, 20 µm. (**B**) Maximum intensity *Z*-projection of IF image Z-stacks of HEK293T cells transfected with constructs encoding OT$^{K2::P1}$ organelle and NLS::GFP^{39TAG}::ms2 in the presence of SCO and tRNA^{Pyl}. Shown from left to right are IF against PylRS (magenta), IF against RPL26L1 (cyan), merge (NLS::GFP^{39TAG}::ms2 is shown in green in all images), and line profiles for the PylRS and RPL26L1 channels (red line, magenta and cyan curves, respectively). Scale bar, 20 µm. (**C**) 3D reconstructions of IF images corresponding to those in (B) and movie S1. Shown from left to right are IF against PylRS (magenta), IF against RPL26L1 (cyan), and merge [NLS::GFP^{39TAG}::ms2 in green; gray dashed lines highlight approximate cell boundaries in (B) and (C)]. RPL26L1 staining of the OT organelle demonstrates partial recruitment of ribosomes, which appear to be immersed into the organelle. (**D**) Working model of OT$^{K2::P1}$-organelle–enabled orthogonal translation. KIF16B::FUS::PylRS and KIF16B::EWSR1::MCP form a spatially separated organelle inside a living cell. PylRS recruits tRNA^{Pyl} and largely depletes its availability in the cytoplasm, whereas MCP recruits ms2-tagged mRNA. Ribosomes and other translation factors are recruited to the organelle for orthogonal translation. Because the charged tRNA^{Pyl} is now in close proximity to only the recruited mRNA of the POI and the spatially distinct set of ribosomes, the selected stop codon can only be translated in the immediate vicinity of the synthetic OT organelle. Meanwhile, all other mRNAs that are not recruited to the OT organelle are subject to normal translational processing of the host machinery and available ribosomes in the remaining cytoplasm—that is, a stop codon will terminate translation. The assembly order of the OT organelle is unknown and is only shown here with arrows for illustrative purposes.

to *MCP* and *PylRS^{AF}* via restriction cloning. *KIF13A$_{1-411}$* and *KIF16B$_{1-400}$* were cloned from human cDNA and inserted into pcDNA3.1 via restriction cloning. P390 of *KIF13A$_{1-411}$* was removed via site-directed mutagenesis. *KIF13A$_{1-411\Delta P390}$* and *KIF16B$_{1-400}$* fusions with *MCP*, *PylRS^{AF}*, *EWSR1::MCP*, *FUS::PylRS^{AF}*, *FUS::PylRS^{AA}*, *SPD5::MCP*, and *SPD5::PylRS^{AF}* were assembled via Gibson assembly (*32*).

Constructs for differential imaging experiments (Fig. 3): To selectively express Nup153::GFP^{149TAG} and Vimentin^{116TAG}::mOrange, one gene was first inserted with *ms2* loops into pBI-CMV1 (*21*). Subsequently, the other gene was inserted without ms2 loops. *INSR^{676TAG}::mOrange* was fused to *ms2* loops by replacing *VIM^{116TAG}::mOrange* in the pBI vector bearing *NUP153::GFP^{149TAG}* and *VIM^{116TAG}::mOrange::ms2*

to yield a bicistronic vector with *INSR*^{676TAG}*::mOrange* in one and *NUP153::GFP*^{149TAG} in the other cassette.

Multicistronic Amber suppression vectors: For ease of experiments with the OT$^{K2::P1}$ organelle we generated multicistronic vectors harboring all necessary components. To assemble multicistronic Amber suppression vectors, first one copy of *tRNA*^{Pyl} under the control of a hU6 promoter was inserted into the pBI-CMV1 vector via Gibson assembly. Subsequently, first *KIF16B::FUS::PylRS*^{AF} and finally *KIF16B::EWSR1::MCP* were inserted via Gibson assembly. Alternatively, a previously published pcDNA3.1 based construct (*21*) expressing NES::PylRS^{AF} under a CMV promoter and tRNA^{Pyl} under a hU6 promoter was used. Alternatively, *hU6-tRNA*^{Pyl} and *NES::PylRS*^{AF} or the components of OT$^{K2::P1}$ were inserted into a pDonor vector (GeneCopoeia). These constructs were used for all experiments in COS-7 and for ribosome/tubulin imaging experiments.

OT$^{K2::P1}$ construct tagging with iRFP (fig. S8): To exclude the possibility of staining artefacts we replaced the *KIF16B::FUS::PylRS*^{AF} in the multicistronic pBI-CMV1 vector with a *KIF16B::iRFP::FUS::PylRS*^{AF} fusion via Gibson assembly. The final construct additionally encodes KIF16B::EWSR1::MCP and tRNA^{Pyl}.

Transfections and used ncAAs

Transfections of HEK293T cells were performed with polyethylenimine (PEI, Sigma-Aldrich 408727) using 3 µg PEI per 1 µg DNA (1200 ng total DNA, diluted in DMEM without Phenol Red, Gibco11880-028). COS-7 cells were transfected using the JetPrime reagent (PeqLab) according to the manufacturer's recommendations at a ratio of 1:2 (1000 ng total DNA).

For Amber suppression system tests, cells were transfected at a ratio of a 1:1:1:1 with POI^{TAG} vectors, tRNA^{Pyl}, PylRS assembler fusions and MCP assembler fusions or mock constructs. 4-6 hours after transfection, the medium was swapped to fresh one containing ncAA. HEK293T cells were analyzed one day after transfection, while COS-7 cells were processed after two days.

Stock and working solutions for all the used ncAAs were prepared as described in previous work (*33*). SCO (cyclooctyne lysine, SiChem SC-8000) was used at a final concentration of 250 µM, while 3-Iodophenylalanine (Chem-Impex International Inc.) was used at a final concentration of 1 mM. SCO is efficiently recognized by PylRS^{AF} (Y306A, Y384F) (*23*) while 3-Iodophenylalanine is recognized by PylRS^{AA} (C346A, N348A) (*27*).

Flow cytometry

HEK293T cells were harvested 1 day after transfection by removing the medium, resuspending the cells in 1xPBS (phosphate buffered saline) and passing them through 100 µm nylon mesh.

Data acquisition was performed on an LSRFortessa SORP Cell Analyzer (BD). Analysis was done using the FlowJo software (FlowJo). Cells were first gated by cell type (using FSC-A x SSC-A parameters) and then by single cell (SSC-A x SSC-W). The workflow of cell gating is shown in Fig. S9. Each shown FFC plot is the sum of three independent biological replicates from which the mean and SEM were calculated. At least 130000 single cells were analyzed per condition. Lastly, fluorescence was acquired in the 488-530/30 channel for GFP signal and in the 561-610/20 channel for mCherry signal.

IF labeling, FISH, and confocal imaging

IF: For immunolabeling experiments, cells were rinsed with PBS, fixed in 2% paraformaldehyde in 1xPBS at room temperature (RT) for 10 min.

Alternatively, if cells were stained for α-Tubulin, they were rinsed with DMEM without Phenol Red and subsequently fixed in a buffer to preserve microtubule structures (100 mM 1,4-Piperazinediethanesulfonic acid (PIPES), 1 mM $MgCl_2$, 0.1 mM $CaCl_2$, 0.1% Triton-x-100 and 2% PFA; pH = 7) for 10 min at RT.

Subsequently, cells were rinsed with PBS and permeabilized with 0.5% Triton-x-100 solution in 1xPBS for 15 min at RT and rinsed twice prior to blocking. Samples were blocked in 3% BSA in 1xPBS for 90 min at RT, after which incubation with the primary antibody was done overnight at 4°C in blocking solution (Ab_{PylRS} (1 µg/mL (*21*)), Ab_{MCP} (Merck ABE76-I, 1:333), $Ab_{α\text{-}Tubulin}$ (Sigma-Aldrich T6199, 2 µg/mL) and/or $Ab_{RPL26L1}$ (Abcam ab137046, 1:200)). The next day, cells were rinsed with PBS and incubated with secondary antibody (ThermoFisher A-21471, A-31553 and/or A-21246, at 2 or 4 µg/mL in blocking solution) for 60 min at RT. Then, cells were rinsed with PBS and fresh PBS was added for imaging.

If only DNA was stained, cells were fixed and permeabilized the same way prior to staining with Hoechst 33342 (Sigma-Aldrich B2261) at 1 µg/mL in 1xPBS for 10 min at RT. Subsequently, cells were rinsed with PBS and fresh PBS was added for imaging.

FISH: FISH experiments were performed one day after transfection analogously to described previously (*21*). Briefly, the hybridization protocol was adapted for 24-well plates from Pierce *et al.* (*34*). In general, IF stainings appear crisper then FISH stainings.

For imaging of only tRNA^{Pyl}, the hybridization probe (5′-(Cy5)-CTAACCCGGCTGAACGGATTTAG-AGTCCATTCGATC-3′) was used at 0.25 µM (hybridization at 37°C, overnight). After four washes with saline sodium citrate buffer (SSC) and one wash with Tris-HCl•NaCl buffer (TN), cells were incubated for 1 hour at RT in 3% BSA prior to IF labeling. Cells were incubated with primary antibodies for 2 hours at RT, rinsed with PBS and incubated with secondary antibodies for 2 hours at RT (antibodies described above). Finally, cells were rinsed with PBS and fresh PBS was added for imaging.

For imaging of both tRNA^{Pyl} and mRNA::ms2, the hybridization probe for tRNA^{Pyl} (5′-(DIG)-CTAACCCGGCTGAACGGATTTAGAGTCCATTC-GATC-3′) was used at 0.16 µM, and the probe for ms2 (5′-(Alexa647)-CTGCAGACATGGGTGATCCTCA-TGTTTTCTA) was used at 0.75 µM. After the SSC washes, cells were incubated for 1 hour at RT in blocking buffer (0.1 M TrisHCl, 150 mM NaCl, 1x blocking reagent (Sigma-Aldrich 11096176001). Then, cells were incubated with anti-digoxigenin-fluorescein antibody (Sigma-Aldrich 11207741910) at a 1:200 dilution in blocking buffer overnight at 4°C. The next day, 3 washes of 5 min were done in Tween buffer (0.1 M TrisHCl, 150 mM NaCl, 0.5% Tween20), before cells were rinsed with PBS and fresh PBS was added for imaging.

Imaging: Confocal images were acquired on a Leica SP8 STED 3X microscope using the 405 nm (for Hoechst), 488 nm (for fluorescein, GFP), 548 nm (mOrange), 594 nm (for Alexa594) and 647 nm (for Alexa647, Cy5) laser lines for excitation. For HEK293T and COS-7 cells a 63x/1.40 oil immersion objective was used. IF images with ribosomes, microtubules, and/or iRFP were taken on an Olympus Fluoroview FV3000 microscope using 405 nm (Alexa405), 488 nm (GFP), 594 nm (for Alexa594), and 640 nm (for Alexa 647 and iRFP) lasers for excitation with a 60x/1.40 oil immersion objective for acquisition. Images were processed using FIJI software.

3D Reconstruction

3D reconstructions in Fig. 4C and corresponding movie S1 were made by using the arivis Vision4D software (arivis AG).

REFERENCES AND NOTES

1. C. C. Liu, P. G. Schultz, Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010). doi: 10.1146/annurev.biochem.052308.105824; pmid: 20307192
2. E. A. Lemke, The exploding genetic code. *ChemBioChem* **15**, 1691–1694 (2014). doi: 10.1002/cbic.201402362; pmid: 25079784
3. J. W. Chin, Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017). doi: 10.1038/nature24031; pmid: 28980641
4. H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444 (2010). doi: 10.1038/nature08817; pmid: 20154731
5. C. Orelle *et al.*, Protein synthesis by ribosomes with tethered subunits. *Nature* **524**, 119–124 (2015). doi: 10.1038/nature14862; pmid: 26222032
6. S. D. Fried, W. H. Schmied, C. Uttamapinant, J. W. Chin, Ribosome Subunit Stapling for Orthogonal Translation in *E. coli*. *Angew. Chem.* **127**, 12982–12985 (2015). doi: 10.1002/anie.201506311; pmid: 27570300
7. F. J. Isaacs *et al.*, Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348–353 (2011). doi: 10.1126/science.1205822; pmid: 21764749
8. M. J. Lajoie *et al.*, Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013). doi: 10.1126/science.1241459; pmid: 24136966
9. N. Ostrov *et al.*, Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016). doi: 10.1126/science.aaf3639; pmid: 27540174
10. K. Wang *et al.*, Defining synonymous codon compression schemes by genome recoding. *Nature* **539**, 59–64 (2016). doi: 10.1038/nature20124; pmid: 27776354
11. Y. Zhang *et al.*, A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **551**, 644–647 (2017). doi: 10.1038/nature24659; pmid: 29189780
12. D. B. Thompson *et al.*, The future of multiplexed eukaryotic genome engineering. *ACS Chem. Biol.* **13**, 313–325 (2018). doi: 10.1021/acschembio.7b00842; pmid: 29241002
13. C. P. Brangwynne *et al.*, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009). doi: 10.1126/science.1172046; pmid: 19460965
14. P. Li *et al.*, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012). doi: 10.1038/nature10879; pmid: 22398450

15. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014). doi: 10.1146/annurev-cellbio-100913-013325; pmid: 25288112
16. E. Bertrand *et al*., Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998). doi: 10.1016/S1097-2765(00)80143-4; pmid: 9809065
17. M. Altmeyer *et al*., Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat. Commun.* **6**, 8088 (2015). doi: 10.1038/ncomms9088; pmid: 26286827
18. A. Patel *et al*., A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015). doi: 10.1016/j.cell.2015.07.047; pmid: 26317470
19. J. B. Woodruff *et al*., The centrosome is a selective condensate that nucleates microtubules by concentrating tubulin. *Cell* **169**, 1066–1077.e10 (2017). doi: 10.1016/j.cell.2017.05.028; pmid: 28575670
20. V. Soppina *et al*., Dimerization of mammalian kinesin-3 motors results in superprocessive motion. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5562–5567 (2014). doi: 10.1073/pnas.1400759111; pmid: 24706892
21. I. Nikić *et al*., Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy. *Angew. Chem.* **53**, 2245–2249 (2014). doi: 10.1002/anie.201309847; pmid: 24474648
22. T. Plass *et al*., Amino acids for Diels-Alder reactions in living cells. *Angew. Chem.* **51**, 4166–4170 (2012). doi: 10.1002/anie.201108231; pmid: 22473599
23. T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, Genetically encoded copper-free click chemistry. *Angew. Chem.* **50**, 3878–3881 (2011). doi: 10.1002/anie.201008178; pmid: 21433234
24. H. Xiao *et al*., Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angew. Chem.* **52**, 14080–14083 (2013). doi: 10.1002/anie.201308137; pmid: 24353230
25. W. H. Schmied, S. J. Elsässer, C. Uttamapinant, J. W. Chin, Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. *J. Am. Chem. Soc.* **136**, 15577–15583 (2014). doi: 10.1021/ja5069728; pmid: 25350841
26. Z. Zhang *et al*., Construction of an inducible stable cell line for efficient incorporation of unnatural amino acids in mammalian cells. *Biochem. Biophys. Res. Commun.* **489**, 490–496 (2017). doi: 10.1016/j.bbrc.2017.05.178; pmid: 28576486
27. Y. S. Wang *et al*., Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. *ACS Chem. Biol.* **8**, 405–415 (2013). doi: 10.1021/cb300512r; pmid: 23138887
28. I. Nikić *et al*., Debugging eukaryotic genetic code expansion for site-specific click-PAINT super-resolution microscopy. *Angew. Chem.* **55**, 16172–16176 (2016). doi: 10.1002/anie.201608284; pmid: 27804198
29. S. F. Banani *et al*., Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016). doi: 10.1016/j.cell.2016.06.010; pmid: 27374333
30. C. C. Liu, M. C. Jewett, J. W. Chin, C. A. Voigt, Toward an orthogonal central dogma. *Nat. Chem. Biol.* **14**, 103–106 (2018). doi: 10.1038/nchembio.2554; pmid: 29337969
31. J. R. Simon, N. J. Carroll, M. Rubinstein, A. Chilkoti, G. P. López, Programming molecular self-assembly of intrinsically disordered proteins containing sequences of low complexity. *Nat. Chem.* **9**, 509–515 (2017). doi: 10.1038/nchem.2715; pmid: 28537592
32. D. G. Gibson *et al*., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009). doi: 10.1038/nmeth.1318; pmid: 19363495
33. I. Nikić, J. H. Kang, G. E. Girona, I. V. Aramburu, E. A. Lemke, Labeling proteins on live mammalian cells using click chemistry. *Nat. Protoc.* **10**, 780–791 (2015). doi: 10.1038/nprot.2015.045; pmid: 25906116
34. J. B. Pierce, S. C. Chafe, M. B. Eswara, G. van der Merwe, D. Mangroo, Strategies for investigating nuclear-cytoplasmic tRNA dynamics in yeast and mammalian cells. *Methods Cell Biol.* **122**, 415–436 (2014). doi: 10.1016/B978-0-12-417160-2.00019-9; pmid: 24857741

ACKNOWLEDGMENTS

We thank I. Schneider as well as all members of the Lemke laboratory, in particular C. Koehler and J. Caria, for helpful discussions. We thank the European Molecular Biology Laboratory (EMBL) flow cytometry core facility and the advanced light microscopy facility for expert assistance. **Funding:** The Lemke laboratory acknowledges generous support from European Research Council (ERC) SMPFv2.0, SPP1623, and SFB1129 (Projektnummer 240245660 funded by Deutsche Forschungsgemeinschaft) and the Gutenberg Research College (GRC). **Author contributions:** E.A.L. conceived the project. C.D.R. and G.E.G. designed and performed all experiments. C.D.R., G.E.G., and E.A.L. analyzed all of the data and cowrote the manuscript. **Competing interests:** C.D.R., G.E.G., and E.A.L. have filed a patent application on OT organelle technology (EP 19157257.7). **Data and materials availability:** All data are available in the main text or the supplementary materials. All plasmids can be obtained upon reasonable request via an EMBL materials transfer agreement (free of charge for noncommercial purposes).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6434/eaaw2644/suppl/DC1
Figs. S1 to S9
Movie S1

4 December 2018; accepted 7 February 2019
10.1126/science.aaw2644

Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes

Christopher D. Reinkemeier, Gemma Estrada Girona and Edward A. Lemke

Science **363** (6434), eaaw2644.
DOI: 10.1126/science.aaw2644

How to make an organelle in eukaryotes

A key step in the evolution of complex organisms like eukaryotes was the organization of specific tasks into organelles. Reinkemeier *et al.* designed an artificial, membraneless organelle into mammalian cells to perform orthogonal translation. In response to a specific codon in a selected messenger RNA, ribosomes confined to this organelle were able to introduce chemical functionalities site-specifically, expanding the canonical set of amino acids. This approach opens possibilities in synthetic cell engineering and biomedical research.

Science, this issue p. eaaw2644

www.sciencemag.org/content/363/6434/eaaw2644/suppl/DC1

Supplementary Material for

Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes

Christopher D. Reinkemeier, Gemma Estrada Girona, Edward A. Lemke*

*Corresponding author. Email: edlemke@uni-mainz.de

Published 29 March 2019, *Science* **363**, eaaw2644 (2017)
DOI: 10.1126/science.aaw2644

This PDF file includes:

Figs. S1 to S9

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/363/6434/eaaw2644/suppl/DC1)

Movie S1

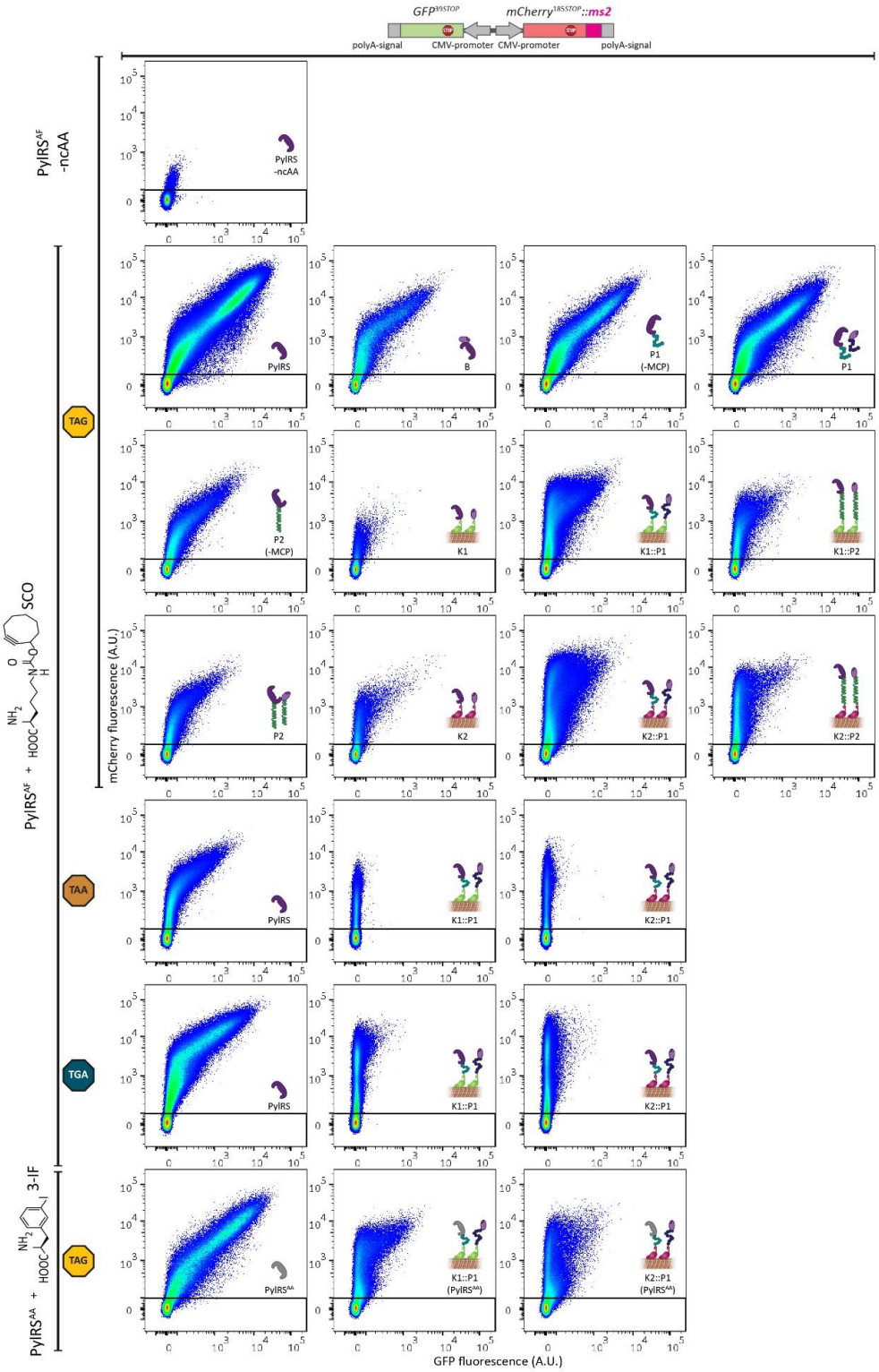

GFP39STOP
mCherry185STOP::ms2
polyA-signal
CMV-promoter
CMV-promoter
polyA-signal
PylRS-ncAA
PylRS
B
P1 (-MCP)
P1
P2 (-MCP)
K1
K1::P1
K1::P2
P2
K2
K2::P1
K2::P2
PylRS
K1::P1
K2::P1
PylRS
K1::P1
K2::P1
PylRS
K1::P1 (PylRS)
K2::P1 (PylRS)
mCherry fluorescence (A.U.)
GFP fluorescence (A.U.)
PylRS-ncAA
PylRS + SCO
PylRS + 3-IF
TAG
TAA
TGA
TAG

Fig. S1. FFC dot plot data corresponding to the bar graph in Fig. 2.

FFC of HEK293T cells expressing the dual-color reporter (mCherry^{185TAG} tagged with ms2 loops and GFP^{39TAG}), tRNA^{Pyl} and the indicated system. Plots show the sum of at least three independent experiments and the axes represent fluorescence intensities in arbitrary units. For transfections with PylRS$^{Y306A, Y384F}$ (PylRS^{AF}) encoding constructs, the ncAA SCO (cyclooctynyl lysine) was added at a concentration of 250 µM, for PylRS$^{N346A, C348A}$ (PylRS^{AA}) 3-iodophenylalanine was added at a concentration of 1 mM (structures indicate used ncAAs). In addition also Opal (TGA) and Ochre (TAA) stop codons were tested by performing experiments with GFP^{39TGA}•mCherry^{185TGA}::ms2 or GFP^{39TAA}•mCherry^{185TAA}::ms2 respectively (tRNA^{Pyl} in these cases with the corresponding anticodons UCA or UUA). In the absence of ncAA, only very low background expression of mCherry and GFP is detected (for this minor background population, the efficiency was typically lower than 3% and the selectivity < 1.7 fold). In row 2 and 3 also data sets are show that did not contain MCP (corresponding to bar graph in maintext Fig. 2). Those serve as a negative control to validate that any observed selectivity effect is dependent on the interaction of MCP with ms2.

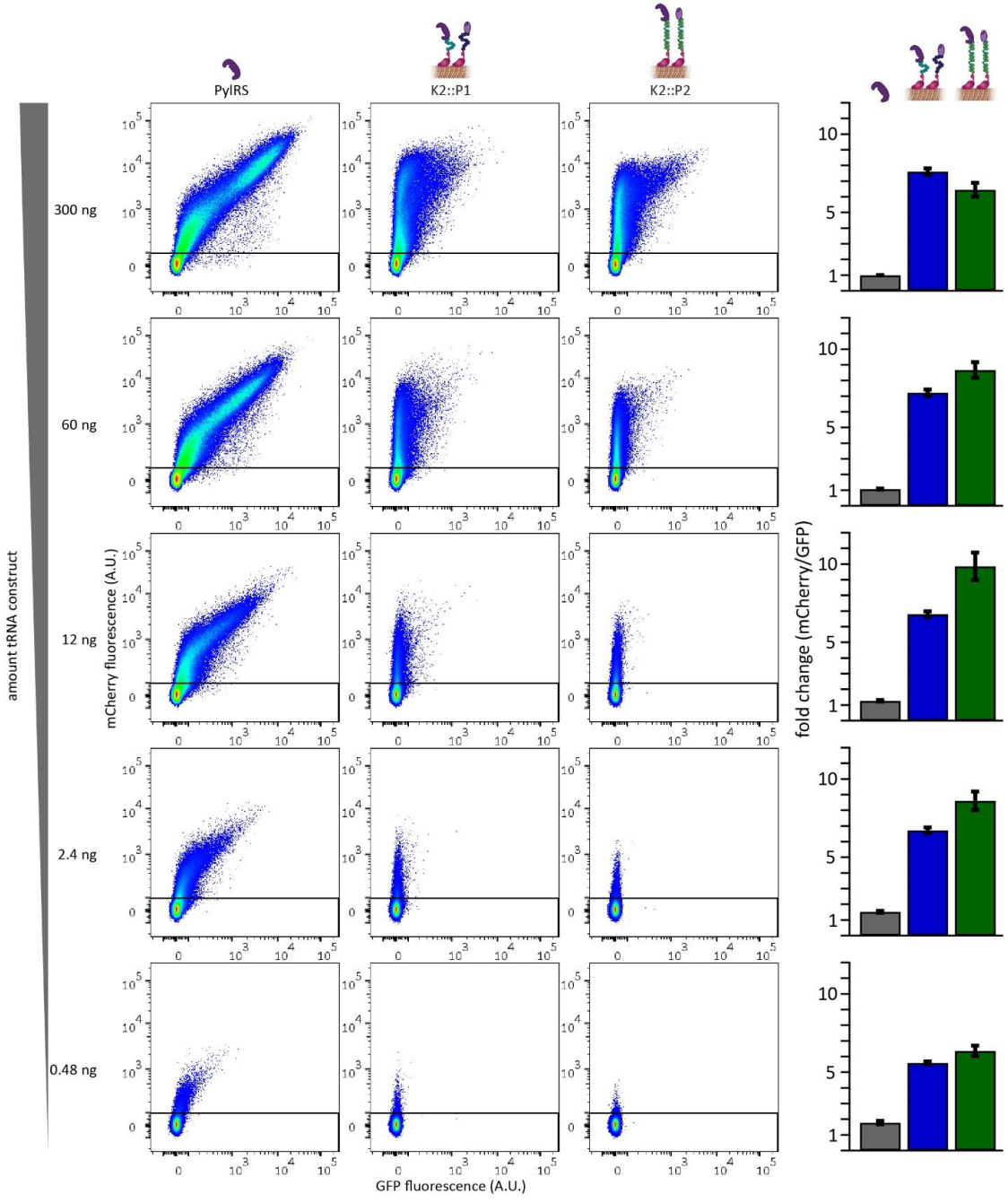


Fig. S2. The OT organelles function across a large dynamic working range of Amber suppression efficiency.

The concentration of $tRNA^{Pyl}$ is known to be very important for Amber suppression *(21).* Here we titrated the amount of *$tRNA^{Pyl}$* (ng of plasmid DNA given) to determine that the observed OT selectivity effect is robust across a large concentration and efficiency range of

Amber suppression. Shown are the corresponding FFC data as well as the bar plot analysis for the tested cytoplasmic PylRS (grey bar), the $OT^{K2::P1}$ (blue bar) and the $OT^{K2::P2}$ (green bar) systems. To this end, HEK293T cells were transfected with the dual-color reporter (*mCherry*^{185TAG} tagged with ms2 loops and *GFP*^{39TAG}) and the indicated genetic code expansion system (300 ng each plasmid). Each system was tested against a range of $tRNA^{Pyl}$ encoding constructs, while the total amount of DNA was adjusted to be kept constant by addition of a mock plasmid. All experiments were performed in the presence of SCO. Dot plots show the sum of three independent experiments and axes indicate fluorescence intensity in arbitrary units. The bar plots show the ratio of mCherry to GFP signal normalized to the transfection with the cytoplasmic *PylRS* and 300 ng *tRNA*^{Pyl} construct. Shown are mean values with SEM of three independent experiments. The selectivity gain of the two best performing OT organelles is robust (>5.5 and <10 fold) even across a three order of magnitude change in *tRNA*^{Pyl} amount.

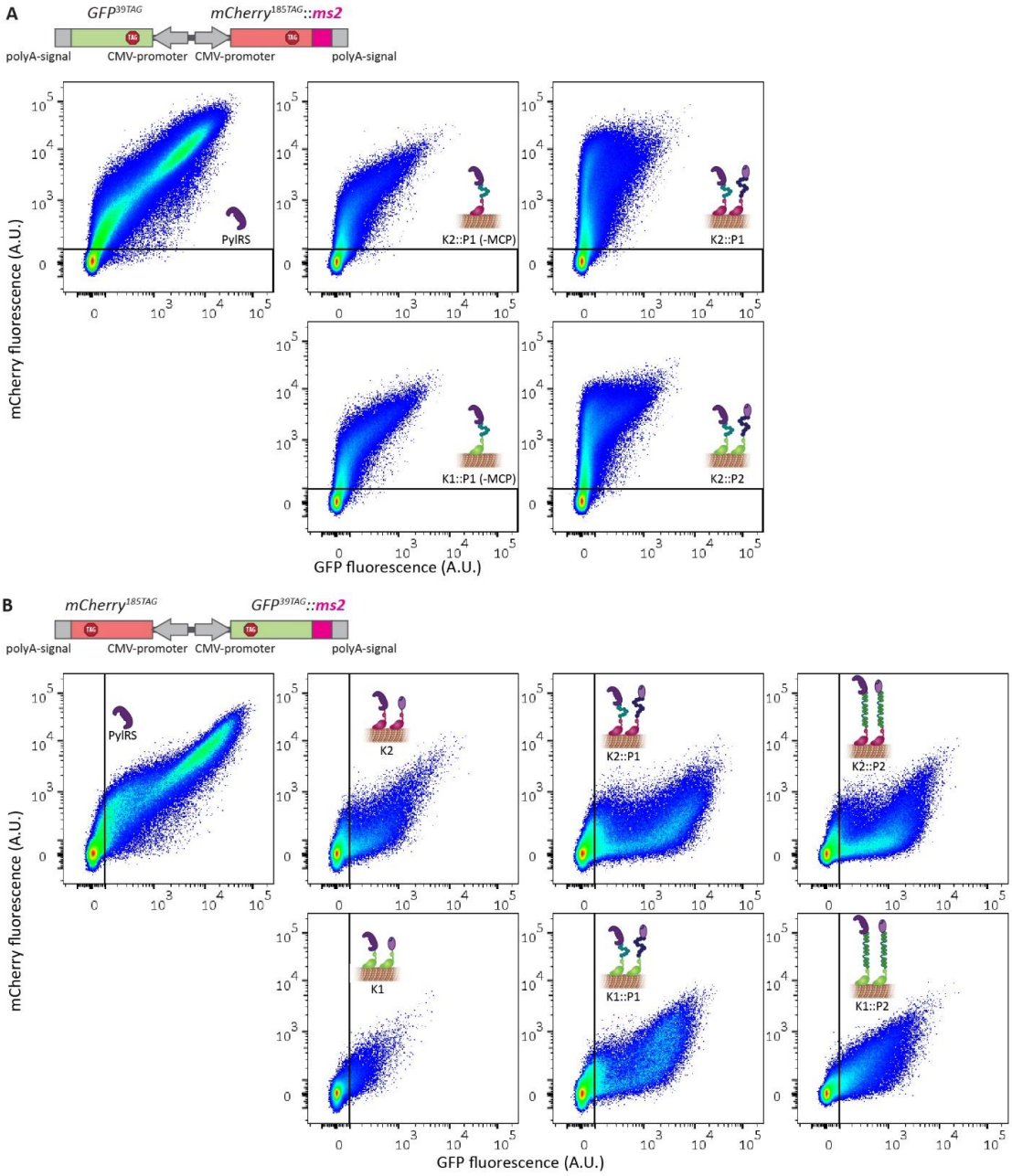


Fig. S3. Additional validations that the OT organelle selectivity is dependent on the complex formation of MCP with ms2 loops.

(A) FFC analysis of HEK293T cells expressing the dual-color reporter (mCherry^{185TAG} tagged with ms2 loops and GFP^{39TAG}), tRNA^{Pyl} and the indicated system. SCO was present in all cases. The result clearly shows that proper function (i.e. selective expression of only the ms2-

tagged mCherry^{185TAG}) of the OT organelles is dependent on the presence of MCP, since in the absence of MCP the FFC plots show a diagonal distribution similar to the cytoplasmic PylRS system. **(B)** FFC analysis of HEK293T cells expressing an inverted dual-color reporter (relative to main text, since here GFP^{39TAG} was tagged with ms2 loops instead of the mCherry^{185TAG}), tRNA^{Pyl} and the indicated system. In the presence of the OT organelles (**K2::P1**, **K2::P2**, **K1::P1** and **K1::P2**), now GFP is preferentially expressed. The result demonstrates that swapping the ms2-tag on the dual-color reporter swaps also the FFC signal. All FFC plots show the sum of at least three independent experiments, the axes represent fluorescence intensity in arbitrary units.

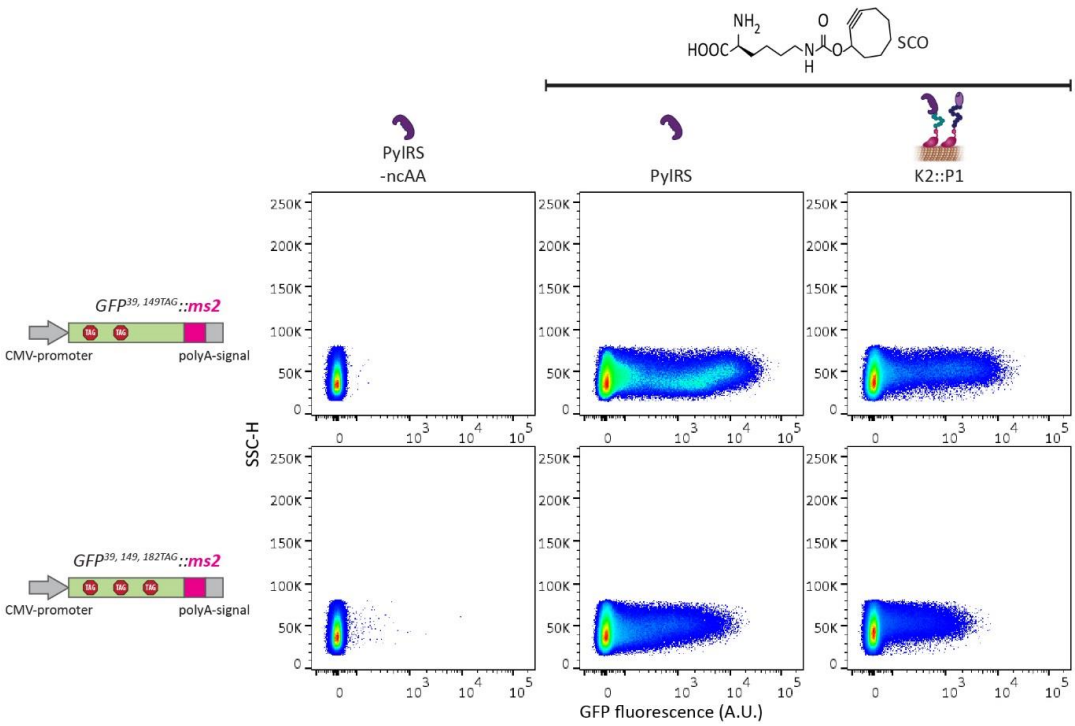


Fig. S4.The OT$^{K2::P1}$ organelle permits suppression of multiple Amber codons.

HEK293T cells were transfected with constructs encoding tRNA^{Pyl} together with either cytoplasmic PylRS or the OT$^{K2::P1}$ organelle and either a construct encoding the single color reporter GFP$^{39TAG, 149TAG}$::ms2 (dual Amber construct) or GFP$^{39TAG, 149TAG, 182TAG}$::ms2 (triple Amber construct) and analyzed via FFC. Plots show the sum of at least three independent experiments. Axes indicate side scatter height (SSC-H) vs GFP fluorescence intensity in arbitrary units. The data clearly show that like the cytoplasmic PylRS, the OT$^{K2::P1}$ organelle is capable of efficiently suppressing multiple Amber codons in one construct (~ 47% for dual Amber and ~ 53% for triple Amber relative to the corresponding cytoplasmic PylRS system for all cells with GFP fluorescence A.U. units >10^2).

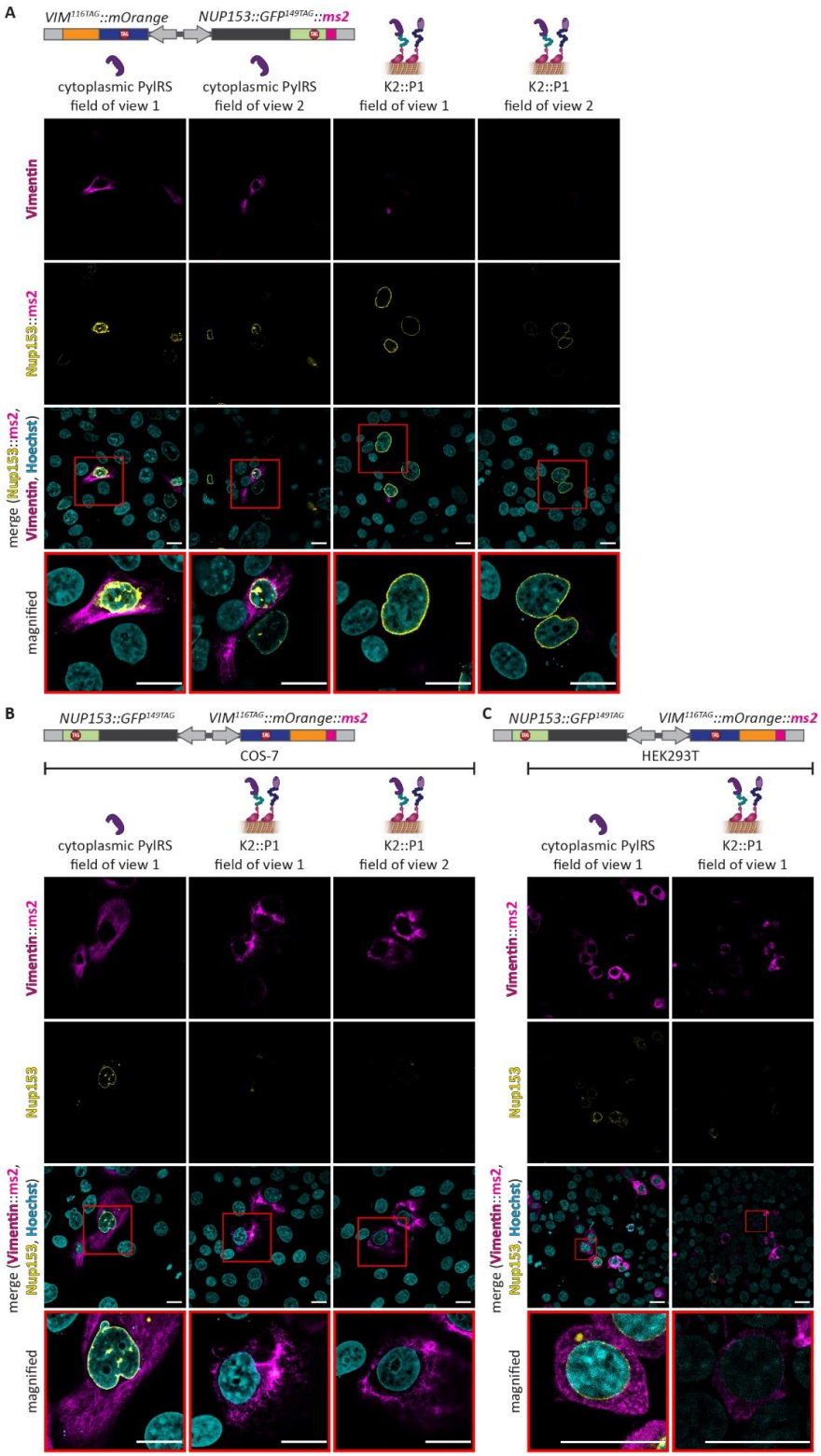

A
VIM116TAG::mOrange
NUP153::GFP149TAG::ms2
cytoplasmic PylRS field of view 1
cytoplasmic PylRS field of view 2
K2::P1 field of view 1
K2::P1 field of view 2
Vimentin
Nup153::ms2
merge (Nup153::ms2, Vimentin, Hoechst)
magnified
B
NUP153::GFP149TAG
VIM116TAG::mOrange::ms2
COS-7
cytoplasmic PylRS field of view 1
K2::P1 field of view 1
K2::P1 field of view 2
Vimentin::ms2
Nup153
merge (Vimentin::ms2, Nup153, Hoechst)
magnified
C
NUP153::GFP149TAG
VIM116TAG::mOrange::ms2
HEK293T
cytoplasmic PylRS field of view 1
K2::P1 field of view 1
Vimentin::ms2
Nup153
merge (Vimentin::ms2, Nup153, Hoechst)
magnified

Fig. S5. The versatile OT$^{K2::P1}$ organelle enables selective and efficient orthogonal translation of a variety of proteins (this figure shows the complementary HEK293T or COS-7 experiments to maintext Fig. 3).

(A-C) Confocal images of cells transfected with constructs encoding PylRS or the OT$^{K2::P1}$ organelle for different protein pairs. SCO and tRNA^{Pyl} were present in all cases. **(A)** COS-7 cells were transfected with *NUP153::GFP^{149TAG}::ms2* and *VIM^{116TAG}::mOrange*. **(B,C)** *VIM^{116TAG}::mOrange::ms2* and *NUP153::GFP^{149TAG}* transfected in COS-7 cells (**(B)**, analog to maintext Fig. 3B) and HEK293T cells **(C)**. Top to bottom: Vimentin^{116TAG}::mOrange (magenta), Nup153::GFP^{149TAG} (yellow), overlay with Hoechst (cyan) and magnified images of representative cells (see red boxes; scale bars, 20 µm). The experiments clearly show for a variety of proteins and two different cell types, that in the presence of the OT$^{K2::P1}$ organelle, only the ms2 carrying mRNA is translated, while the cytoplasmic PylRS system always translates both.

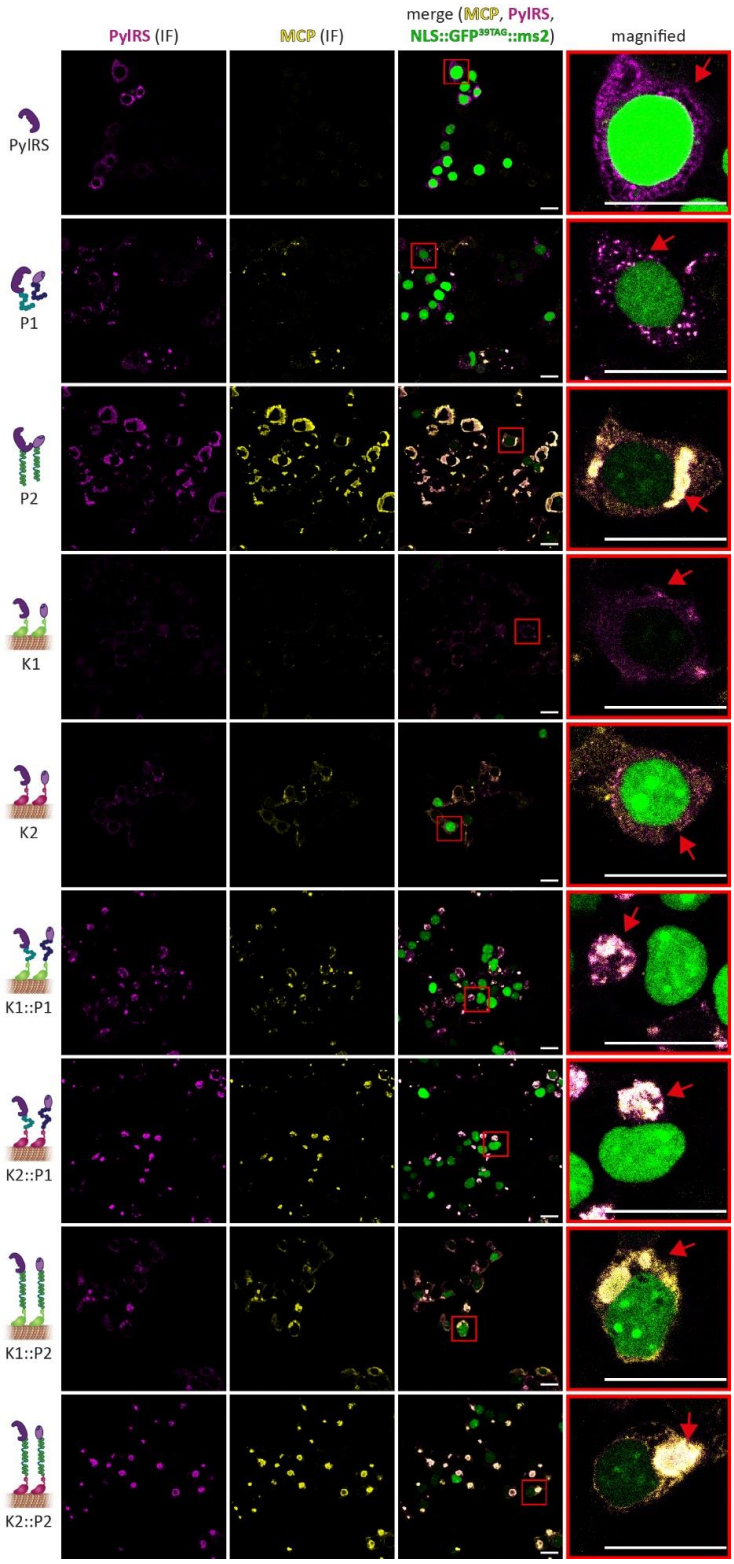

PylRS (IF)
MCP (IF)
merge (MCP, PylRS, NLS::GFP39TAG::ms2)
magnified
PylRS
P1
P2
K1
K2
K1::P1
K2::P1
K1::P2
K2::P2

Fig. S6. Different dual imaging combinations to verify that assembler::MCP always colocalizes with assembler::PylRS.

(**A**) In maintext Fig. 4 we showed that $tRNA^{Pyl}$ (visualized by FISH) and PylRS (visualized by IF) colocalize well. Complementary to this, here we show that also PylRS and MCP fused to the assemblers colocalize. IF of HEK293T cells expressing $NLS::GFP^{39TAG}$::ms2 with $tRNA^{Pyl}$ and the indicated system. Note, that in general IF stainings appear crisper than FISH stainings. Left to right: PylRS (magenta), MCP (yellow), merge ($NLS::GFP^{39TAG}$::ms2 in green) and magnified image (see red box). $NLS::GFP^{39TAG}$ yields nuclear fluorescence only if Amber suppression was successful, and helps here to identify cells in which transfection with all necessary plasmids occurred. All magnified zooms were centered on GFP positive cells. However, color settings were kept constant through this work for consistency and GFP expression differed between tested system (and cells), consistent with the observed Amber suppression efficiencies for the different systems, which in some cases were very low. Scale bars are 20 μm and in all magnified images red arrows point to representative structures as discussed in maintext Fig. 4.

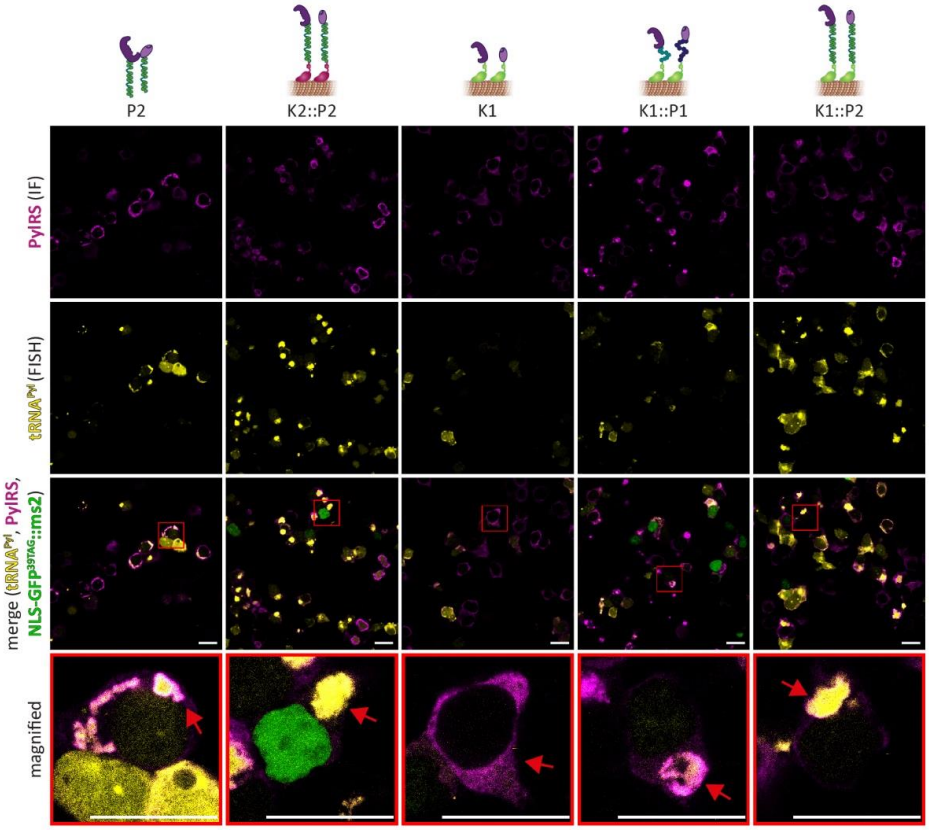


Fig. S7. PylRS and tRNA^{Pyl} colocalize for different KIF and SPD5 assembler systems (complementary to maintext Fig. 4).

In maintext Fig. 4 we showed colocalization of tRNA^{Pyl} and PylRS for assembler systems **P1**, **K2** and **K2::P1**. Here we show that they also colocalize for **P2**, **K2::P2**, **K1**, **K1::P1** and **K1::P2**. IF and FISH imaging of HEK293T cells expressing tRNA^{Pyl}, NLS::GFP^{39TAG}::ms2 and the indicated system. Top to bottom: IF against PylRS (magenta), FISH against tRNA^{Pyl} (yellow), merge (NLS::GFP^{39TAG}::ms2 in green) and magnified images of representative cells (see red boxes, red arrows highlight representative structures as discussed in maintext Fig. 4, scale bars, 20 µm). All magnified zooms were centered on GFP positive cells. However, color settings were kept constant through this work for consistency and GFP expression differed between tested system (and cells), consistent with the observed Amber suppression efficiencies for the different systems, which in some cases were very low.

The colocalization analysis is in line with the working hypothesis that selectivity is highly dependent on the partition coefficient of tRNA^{Pyl} into the OT organelle. For the three systems where droplet appearance is clearly visible (**P1**, **P2**, **K2::P1**), a qualitative analysis of the relative intensity ratio (IR) of droplets vs. cytoplasm (here used as a qualitative proxy for the partition coefficient; IR = intensity of droplets/intensity in cytoplasm) correlates well with the observed selectivity yielding IR(**K2::P1**) >> IR(**P2**) ~ IR(**P1**). However, despite the general challenges of

quantitative imaging in cells to establish exact concentrations of proteins *(29)*, the fact that we compare droplets of very different size make absolute quantitation very difficult. In addition, all those system show some selectivity, indicative of a depletion of $tRNA^{Pyl}$ from the cytoplasm to varying degrees. The remaining $tRNA^{Pyl}$ levels in the cytoplasm are hardly distinguishable from background fluorescence, so that we found any form of quantitation to be highly dependent on set parameters (threshold, background correction etc.). The trend, however was robust IR(**K2::P1**) >> IR(**P2**) ~ IR(**P1**). We suggest that in the future quantitative technologies such as in cell fluorescence correlation spectroscopy or quantitative super resolution microscopy could be used to measure the concentration of compounds in the droplet directly.

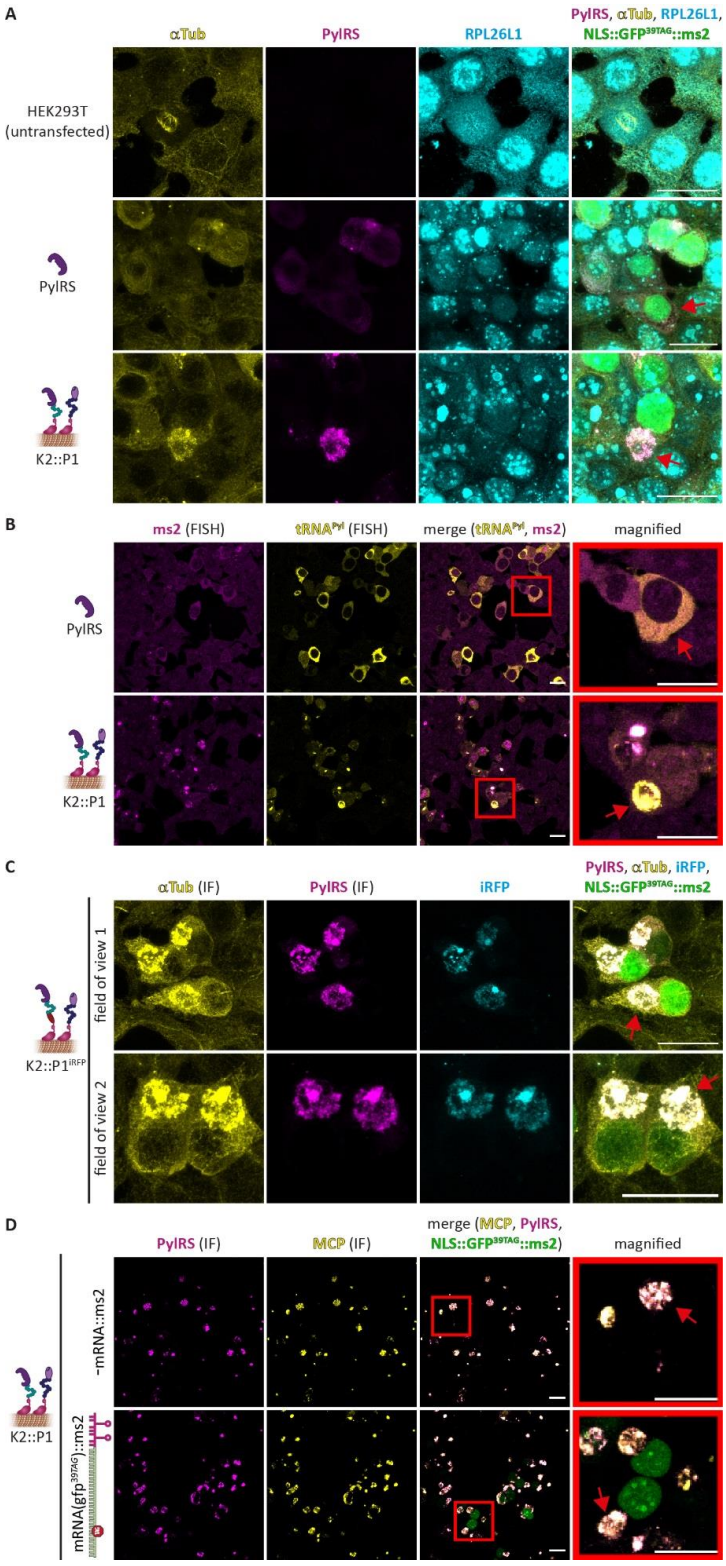

A
αTub
PylRS
RPL26L1
PylRS, αTub, RPL26L1,
NLS::GFP39TAG::ms2
HEK293T
(untransfected)
PylRS
K2::P1
B
ms2 (FISH)
tRNAPyl (FISH)
merge (tRNAPyl, ms2)
magnified
PylRS
K2::P1
C
αTub (IF)
PylRS (IF)
iRFP
PylRS, αTub, iRFP,
NLS::GFP39TAG::ms2
field of view 1
field of view 2
K2::P1iRFP
D
PylRS (IF)
MCP (IF)
merge (MCP, PylRS,
NLS::GFP39TAG::ms2)
magnified
-mRNA::ms2
mRNA(gfp39TAG)::ms2
K2::P1

Fig. S8. RPL26L1, α-Tubulin and mRNA::ms2 colocalize with $OT^{K2::P1}$ organelle, its structure is not changed by the presence of mRNA::ms2.

(**A**) Maximum intensity Z-projection of IF of either untransfected HEK293T cells or HEK293T cells transfected with $tRNA^{Pyl}$, NLS::GFP^{39TAG}::ms2 and PylRS or $OT^{K2::P1}$ encoding constructs. Left to right: α-Tubulin (yellow), PylRS (magenta), RPL26L1 (cyan) and merge (NLS::GFP^{39TAG}::ms2 in green, scale bars 20 µm). This image series shows that microtubules and ribosomes colocalize in the $OT^{K2::P1}$ organelle.

(**B**) In Fig. 4, S6 and S7 we have shown that $tRNA^{Pyl}$, PylRS, MCP colocalize to OT organelles. Here we use FISH against mRNA::ms2 to visualize that also our targeted mRNA was recruited to the $OT^{K2::P1}$ organelle. FISH was performed in HEK293T cells expressing the dual-color reporter mRNAs (GFP^{39TAG}, $mCherry^{185TAG}$::ms2) as well as $tRNA^{Pyl}$ and PylRS or the $OT^{K2::P1}$ organelle (here, no ncAA was present to avoid formation of fluorescent proteins interfering with image analysis). Left to right ms2 (magenta), $tRNA^{Pyl}$ (yellow), merge and magnified images (see red box).

(**C**) Maximum intensity Z-projection of IF of HEK293T cells expressing with $tRNA^{Pyl}$, NLS::GFP^{39TAG}::ms2 and $OT^{K2::P1}$, in which KIF16B::FUS::PylRS is genetically tagged with iRFP. Left to right: α-Tubulin (yellow), PylRS (magenta), iRFP (cyan) and merge (NLS::GFP^{39TAG}::ms2 in green). This image series shows that the hollow appearance of the $OT^{K2::P1}$ organelle cannot be attributed to a staining artefact. We believe that the hollow appearance is rather linked to scaffolding to the Tubulin network, but not necessarily of functional relevance.

(**D**) IF of HEK293T cells transfected with constructs encoding $tRNA^{Pyl}$, $OT^{K2::P1}$ and either a mock plasmid or a plasmid encoding NLS::GFP^{39TAG}::ms2. Left to right: PylRS (magenta), MCP (yellow), merge (NLS::GFP^{39TAG}::ms2 in green) and magnified image (see red box). This image series shows that in $OT^{K2::P1}$ droplet-like appearance and colocalization of MCP and PylRS was detectable in cells expressing mRNA::ms2 and in cells without, indicating, that droplet formation did not depend on RNA recruitment.

(**A-D**) All scale bars are 20 µm and red arrows point to representative structures.

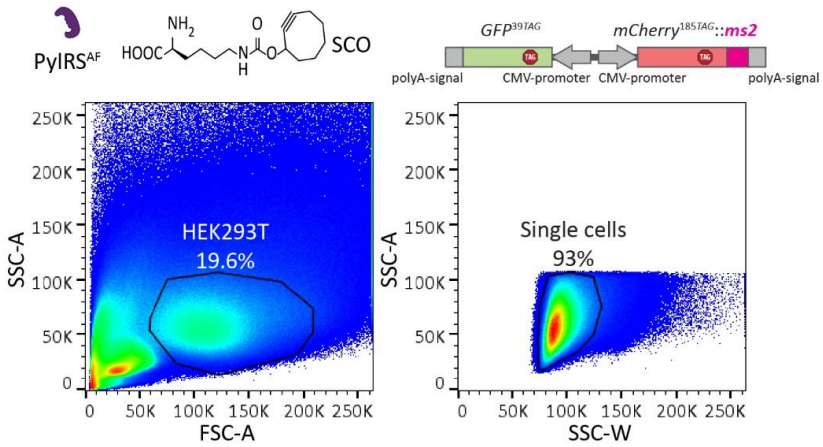


Fig. S9. Single HEK293T cells were sequentially identified based on scatter values.

As a representative example HEK293T cells expressing PylRS, $tRNA^{Pyl}$ and the dual-color reporter ($mCherry^{185TAG}$::ms2 and GFP^{39TAG}) are shown (SCO was present in this experiment). Cells were first gated by cell type using forward scatter area (FSC-A) and side scatter area (SSC-A) parameters. Subsequently, single cells were identified based on SSC-A and side scatter width (SSC-W). Percentage next to gate name is based on the parent population (all data for SSC-A vs. FSC-A, HEK293T for SSC-A vs SSC-W). Cells passing the first gate (left panel) are further gated in the second (right) panel. All FFC data shown in this study were subjected to this analysis.

Movie 1: Complementary to maintext Fig. 4C here we show a movie of the 3D reconstruction.

3D reconstructions of IF images corresponding to those in maintext Fig. 4B,C. NLS::GFP^{39TAG}::ms2 (green), IF against RPL26L1 (cyan) and IF against PylRS (magenta). RPL26L1 staining inside the OT organelle demonstrates recruitment of ribosomes.

References and Notes

1. C. C. Liu, P. G. Schultz, Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010). doi:10.1146/annurev.biochem.052308.105824 Medline

2. E. A. Lemke, The exploding genetic code. *ChemBioChem* **15**, 1691–1694 (2014). doi:10.1002/cbic.201402362 Medline

3. J. W. Chin, Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017). doi:10.1038/nature24031 Medline

4. H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444 (2010). doi:10.1038/nature08817 Medline

5. C. Orelle, E. D. Carlson, T. Szal, T. Florin, M. C. Jewett, A. S. Mankin, Protein synthesis by ribosomes with tethered subunits. *Nature* **524**, 119–124 (2015). doi:10.1038/nature14862 Medline

6. S. D. Fried, W. H. Schmied, C. Uttamapinant, J. W. Chin, Ribosome Subunit Stapling for Orthogonal Translation in *E. coli*. *Angew. Chem.* **127**, 12982–12985 (2015). doi:10.1002/anie.201506311 Medline

7. F. J. Isaacs, P. A. Carr, H. H. Wang, M. J. Lajoie, B. Sterling, L. Kraal, A. C. Tolonen, T. A. Gianoulis, D. B. Goodman, N. B. Reppas, C. J. Emig, D. Bang, S. J. Hwang, M. C. Jewett, J. M. Jacobson, G. M. Church, Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348–353 (2011). doi:10.1126/science.1205822 Medline

8. M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, F. J. Isaacs, Genomically recoded organisms expand biological functions. *Science* **342**, 357–360 (2013). doi:10.1126/science.1241459 Medline

9. N. Ostrov, M. Landon, M. Guell, G. Kuznetsov, J. Teramoto, N. Cervantes, M. Zhou, K. Singh, M. G. Napolitano, M. Moosburner, E. Shrock, B. W. Pruitt, N. Conway, D. B. Goodman, C. L. Gardner, G. Tyree, A. Gonzales, B. L. Wanner, J. E. Norville, M. J. Lajoie, G. M. Church, Design, synthesis, and testing toward a 57-codon genome. *Science* **353**, 819–822 (2016). doi:10.1126/science.aaf3639 Medline

10. K. Wang, J. Fredens, S. F. Brunner, S. H. Kim, T. Chia, J. W. Chin, Defining synonymous codon compression schemes by genome recoding. *Nature* **539**, 59–64 (2016). doi:10.1038/nature20124 Medline

11. Y. Zhang, J. L. Ptacin, E. C. Fischer, H. R. Aerni, C. E. Caffaro, K. San Jose, A. W. Feldman, C. R. Turner, F. E. Romesberg, A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* **551**, 644–647 (2017). doi:10.1038/nature24659 Medline

12. D. B. Thompson, S. Aboulhouda, E. Hysolli, C. J. Smith, S. Wang, O. Castanon, G. M. Church, The future of multiplexed eukaryotic genome engineering. *ACS Chem. Biol.* **13**, 313–325 (2018). doi:10.1021/acschembio.7b00842 Medline

13. C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoege, J. Gharakhani, F. Jülicher, A. A. Hyman, Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009). doi:10.1126/science.1172046 Medline

14. P. Li, S. Banjade, H.-C. Cheng, S. Kim, B. Chen, L. Guo, M. Llaguno, J. V. Hollingsworth, D. S. King, S. F. Banani, P. S. Russo, Q.-X. Jiang, B. T. Nixon, M. K. Rosen, Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012). doi:10.1038/nature10879 Medline

15. A. A. Hyman, C. A. Weber, F. Jülicher, Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014). doi:10.1146/annurev-cellbio-100913-013325 Medline

16. E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer, R. M. Long, Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998). doi:10.1016/S1097-2765(00)80143-4 Medline

17. M. Altmeyer, K. J. Neelsen, F. Teloni, I. Pozdnyakova, S. Pellegrino, M. Grøfte, M.-B. D. Rask, W. Streicher, S. Jungmichel, M. L. Nielsen, J. Lukas, Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat. Commun.* **6**, 8088 (2015). doi:10.1038/ncomms9088 Medline

18. A. Patel, H. O. Lee, L. Jawerth, S. Maharana, M. Jahnel, M. Y. Hein, S. Stoynov, J. Mahamid, S. Saha, T. M. Franzmann, A. Pozniakovski, I. Poser, N. Maghelli, L. A. Royer, M. Weigert, E. W. Myers, S. Grill, D. Drechsel, A. A. Hyman, S. Alberti, A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077 (2015). doi:10.1016/j.cell.2015.07.047 Medline

19. J. B. Woodruff, B. Ferreira Gomes, P. O. Widlund, J. Mahamid, A. Honigmann, A. A. Hyman, The centrosome is a selective condensate that nucleates microtubules by concentrating tubulin. *Cell* **169**, 1066–1077.e10 (2017). doi:10.1016/j.cell.2017.05.028 Medline

20. V. Soppina, S. R. Norris, A. S. Dizaji, M. Kortus, S. Veatch, M. Peckham, K. J. Verhey, Dimerization of mammalian kinesin-3 motors results in superprocessive motion. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5562–5567 (2014). doi:10.1073/pnas.1400759111 Medline

21. I. Nikić, T. Plass, O. Schraidt, J. Szymański, J. A. Briggs, C. Schultz, E. A. Lemke, Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy. *Angew. Chem.* **53**, 2245–2249 (2014). doi:10.1002/anie.201309847 Medline

22. T. Plass, S. Milles, C. Koehler, J. Szymański, R. Mueller, M. Wiessler, C. Schultz, E. A. Lemke, Amino acids for Diels-Alder reactions in living cells. *Angew. Chem.* **51**, 4166–4170 (2012). doi:10.1002/anie.201108231 Medline

23. T. Plass, S. Milles, C. Koehler, C. Schultz, E. A. Lemke, Genetically encoded copper-free click chemistry. *Angew. Chem.* **50**, 3878–3881 (2011). doi:10.1002/anie.201008178 Medline

24. H. Xiao, A. Chatterjee, S. H. Choi, K. M. Bajjuri, S. C. Sinha, P. G. Schultz, Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. *Angew. Chem.* **52**, 14080–14083 (2013). doi:10.1002/anie.201308137 Medline

25. W. H. Schmied, S. J. Elsässer, C. Uttamapinant, J. W. Chin, Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1. *J. Am. Chem. Soc.* **136**, 15577–15583 (2014). doi:10.1021/ja5069728 Medline

26. Z. Zhang, H. Xu, L. Si, Y. Chen, B. Zhang, Y. Wang, Y. Wu, X. Zhou, L. Zhang, D. Zhou, Construction of an inducible stable cell line for efficient incorporation of unnatural amino acids in mammalian cells. *Biochem. Biophys. Res. Commun.* **489**, 490–496 (2017). doi:10.1016/j.bbrc.2017.05.178 Medline

27. Y. S. Wang, X. Fang, H.-Y. Chen, B. Wu, Z. U. Wang, C. Hilty, W. R. Liu, Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-tRNA synthetase mutant. *ACS Chem. Biol.* **8**, 405–415 (2013). doi:10.1021/cb300512r Medline

28. I. Nikić, G. Estrada Girona, J. H. Kang, G. Paci, S. Mikhaleva, C. Koehler, N. V. Shymanska, C. Ventura Santos, D. Spitz, E. A. Lemke, Debugging eukaryotic genetic code expansion for site-specific click-PAINT super-resolution microscopy. *Angew. Chem.* **55**, 16172–16176 (2016). doi:10.1002/anie.201608284 Medline

29. S. F. Banani, A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker, M. K. Rosen, Compositional control of phase-separated cellular bodies. *Cell* **166**, 651–663 (2016). doi:10.1016/j.cell.2016.06.010 Medline

30. C. C. Liu, M. C. Jewett, J. W. Chin, C. A. Voigt, Toward an orthogonal central dogma. *Nat. Chem. Biol.* **14**, 103–106 (2018). doi:10.1038/nchembio.2554 Medline

31. J. R. Simon, N. J. Carroll, M. Rubinstein, A. Chilkoti, G. P. López, Programming molecular self-assembly of intrinsically disordered proteins containing sequences of low complexity. *Nat. Chem.* **9**, 509–515 (2017). doi:10.1038/nchem.2715 Medline

32. D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison 3rd, H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009). doi:10.1038/nmeth.1318 Medline

33. I. Nikić, J. H. Kang, G. E. Girona, I. V. Aramburu, E. A. Lemke, Labeling proteins on live mammalian cells using click chemistry. *Nat. Protoc.* **10**, 780–791 (2015). doi:10.1038/nprot.2015.045 Medline

34. J. B. Pierce, S. C. Chafe, M. B. Eswara, G. van der Merwe, D. Mangroo, Strategies for investigating nuclear-cytoplasmic tRNA dynamics in yeast and mammalian cells. *Methods Cell Biol.* **122**, 415–436 (2014). doi:10.1016/B978-0-12-417160-2.00019-9 Medline

Appendix III Raising the ribosomal repertoire

This work was published as:

Reinkemeier CD & Lemke EA (2020). Raising the ribosomal repertoire. Nature Chemistry, 12, 503-504, doi:10.1038/s41557-020-0476-6

SYNTHETIC BIOLOGY

Raising the ribosomal repertoire

Three versatile and mutually orthogonal tRNA/aminoacyl-tRNA synthetase pairs have been developed. Collectively, these pairs enable the site-specific incorporation of three different non-canonical amino acids into a protein that can still be terminated faithfully by a natural stop codon.

Christopher D. Reinkemeier and Edward A. Lemke

The genetic code is the universal rule book determining how proteins are synthesized at the ribosome. It is conserved throughout all domains of life, yet despite its simplicity it gives rise to the diversity seen across biology. It is, however, limited to encode canonical amino acids and therefore, all proteins are composed of a small set of chemical functionalities. We can only imagine what might be possible if the entire repertoire of chemical functionality could be explored.

Genetic code expansion enables site-specific incorporation of non-canonical amino acids (ncAAs) into proteins by the ribosome, and has been developed to encode over 200 diverse ncAAs during the past few decades of research. This technology depends on orthogonal transfer RNA/aminoacyl-tRNA synthetase (tRNA/RS) pairs, and blank codons that can be inserted into the protein coding sequence to achieve site-specificity. An orthogonal tRNA/RS pair has to fulfil the following three criteria: the tRNA should only be recognized by the orthogonal RS and not by an endogenous one; the orthogonal RS may not bind any endogenous tRNA; and the orthogonal RS should only bind the ncAA and no canonical amino acid. The anticodon of the orthogonal tRNA is typically mutated to decode a rare amber stop codon that is introduced into the mRNA of the protein of interest at a chosen site. However, even rare amber stop codons still occur many times in the transcriptome, and background

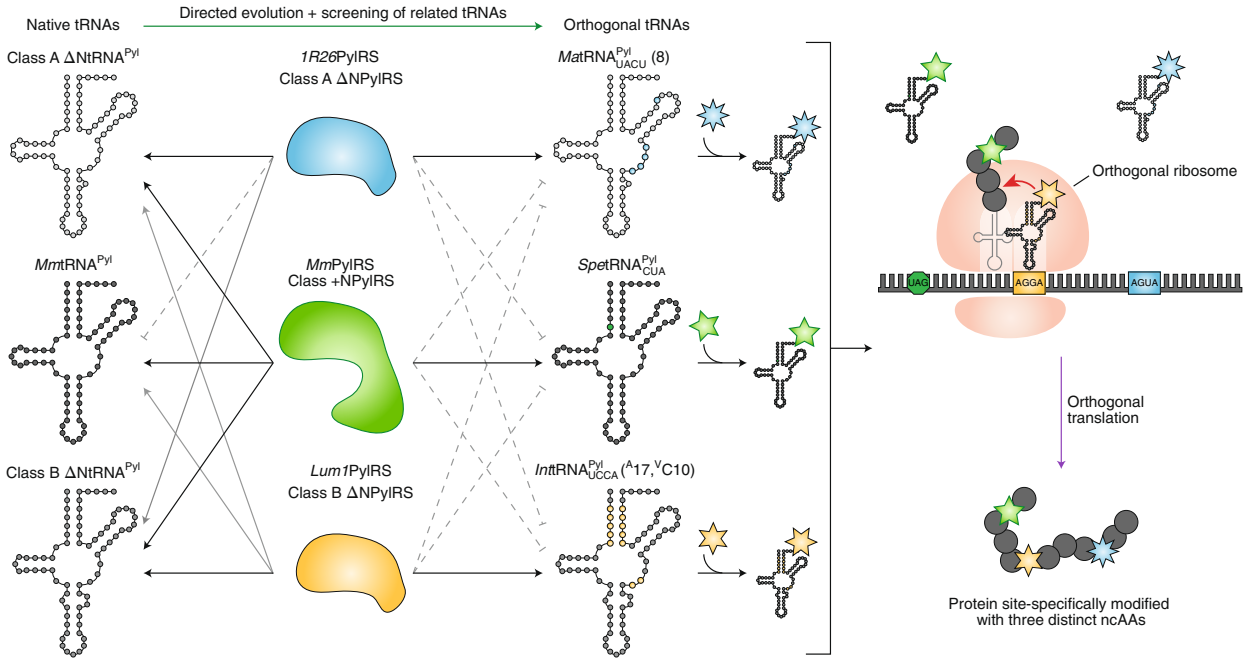


Fig. 1 | Mutually orthogonal tRNA^{Pyl}/PylRS systems enable advanced protein engineering. Combining three orthogonal tRNA^{Pyl}/PylRS based systems with orthogonal quadruplet decoding ribosomes enables the selective incorporation of three distinct ncAAs into one faithfully terminated protein. PylRS enzymes are classified into three separate classes: one containing an N-terminal domain that binds the variable loop of tRNA^{Pyl} (PylRS shown in green), and two functionally distinct ΔNPylRS classes (shown in blue and orange). By exploring natural diversity and combining this with directed evolution of the tRNA^{Pyl} variants, these three pairs were made orthogonal to each other (black arrows indicate amino acylation activity, grey arrows indicate residual amino acylation, blunt arrows indicate orthogonality). When combined with specific active site mutations and orthogonal ribosomes these pairs enable the incorporation of three different ncAAs (indicated by blue, green and orange stars, respectively) into a protein by decoding an amber stop codon, as well as two different quadruplet codons.

incorporations into untargeted proteins that naturally terminate on the amber codon are always a concern — which might not be the case when proteins are removed from the cell and subsequently purified.

For synthesizing truly artificial proteins, with multiple different ncAAs, the number of blank codons and orthogonal tRNA/RS pairs is limiting. In *E. coli* multiple blank codons have been established by three independent approaches. Firstly, artificial base pairs can be inserted into the genome, transcribed into RNA and decoded at the ribosome[1]; secondly, orthogonal ribosomes can efficiently translate quadruplet codons[2]; and thirdly, recent advances in genome engineering make it possible to remove entire codons from *E. coli* genomes, which can then be freely reassigned[3]. These free codons now need to be matched by the corresponding numbers of orthogonal tRNA/RS pairs if one aims to expand on the existing genetic code.

Despite substantial efforts, up to now only a handful of orthogonal tRNA/RS pairs have been discovered. Two additional practical limitations for most orthogonal tRNA/RS pairs are the difficulties of incorporating diverse ncAAs by adapting the substrate specificity of the RS and decoding different codons by changing the anticodon of the tRNA, as many RSs have a narrow substrate range and bind to the anticodon of their cognate tRNA. The most widely used example of this category is the $tRNA^{Tyr}$/tyrosyl tRNA synthetase pair derived from *Methanocaldococcus jannashii* (*Mj*), which has been evolved to incorporate many different ncAAs. It is also possible to obtain mutually orthogonal *Mj*TyrRS pairs by directed evolution of both the tRNA and the RS[4].

In contrast to most other orthogonal tRNA/RS pairs, the pyrrolysyl tRNA/tRNA synthetase pairs derived from methanogenic archaea bacteria [$tRNA^{Pyl}$/PylRS from *Methanosarcina mazei (Mm)* or *-barkeri (Mb)*] do not bind the anticodon loop of the tRNA, so the anticodon can easily be chosen to reprogram any codon. Additionally, PylRS systems naturally have a broad substrate specificity and the wild-type itself does not bind any canonical amino acid. This, and the fact that PylRS is orthogonal in both eukaryotes and bacteria, have made it arguably one of the most useful genetic code expansion tools. Nevertheless, having more $tRNA^{Pyl}$/PylRS pairs that are mutually orthogonal to each other would dramatically increase the power of the method, as this would enable researchers to harness the diversity of active site mutants to encode various ncAAs and to easily change the target codon. Writing in *Nature Chemistry*, a team led by Jason W. Chin now reports[5] an elegant method to use multiple different $tRNA^{Pyl}$/PylRS pairs to encode multiple ncAAs into proteins in *E. coli*.

Recently, PylRS variants that lack the N-terminal domain found in the closely related *Mm*- or *Mb*PylRS pairs were discovered and evolved to be orthogonal to *Mm*PylRS[6,7]. The team has now expanded on this approach by identifying a third class of PylRS enzymes and developing a system with three mutually orthogonal $tRNA^{Pyl}$/PylRS variants. Intriguingly, based on sequence data and functionality the $tRNA^{Pyl}$/PylRS systems can be classified in three distinct classes: one class having an N-terminal domain binding to the variable loop (+N class) and two functionally distinct classes without the N-terminal domain, termed class A and B. Although these three native classes already show some degree of orthogonality, there is still substantial cross reactivity, hence the team looked to optimize them to create three fully orthogonal systems.

In particular, they needed to develop (1) a tRNA that is a substrate for *Mm*PylRS but orthogonal to both ΔNPylRS classes, (2) a tRNA that is a substrate for ΔNPylRS class A but not for *Mm*PylRS or ΔNPylRS class B; and (3) a tRNA that is exclusively a substrate for ΔNPylRS class B. To discover a tRNA for *Mm*PylRS the team tested evolutionarily related $tRNA^{Pyl}$ variants belonging to the +NPylRS class. This enabled them to identify that $tRNA^{Pyl}$ from *Methanosarcina spelaei* (*Spet*RNA^{Pyl}) — although only having one point mutation in the acceptor stem loop compared to *Mm*$tRNA^{Pyl}$ — is completely orthogonal to both ΔNPylRS classes, while retaining activity with *Mm*PylRS.

In their previous work the team demonstrated that extending the variable loop of the class A Δ$NtRNA^{Pyl}$ can establish orthogonality to *Mm*PylRS[6]. By screening several variable loop mutants they could identify some that are also orthogonal to class B ΔNPylRS. To obtain three independent, fully orthogonal systems, however, they still needed to discover a $tRNA^{Pyl}$ that is exclusively a class B ΔNPylRS substrate. All native $tRNA^{Pyl}$ variants for class B ΔNPylRS variants are also substrates for class A ΔNPylRS variants and *Mm*PylRS. Therefore, Chin and co-workers performed directed evolution to eliminate cross reactivity with the other synthetase classes. To this end, they used two complementary strategies: in an approach that was analogous to their previous work they created a library with an extended variable loop; additionally, they screened a library with a randomized acceptor stem loop. In both libraries they discovered better hits than the parent $tRNA^{Pyl}$, but none with sufficient orthogonality. However, they found that when they combined mutations from both approaches, the resulting $tRNA^{Pyl}$ variants were fully orthogonal. Using this set of three mutually orthogonal $tRNA^{Pyl}$/PylRS systems, and a previously known orthogonal ribosome that efficiently decodes quadruplet codons, Chin and co-workers were then able to site-specifically incorporate three different ncAAs into one protein that can still be terminated faithfully by a natural stop codon (see Fig. 1).

The discovery of these new $tRNA^{Pyl}$/PylRS classes, and their development to orthogonal genetic code expansion systems, is an important step towards the genetically encoded synthesis of fully artificial polymers. The past 20 years have delivered roughly five orthogonal RS systems to expand the genetic code, which have found applications in many laboratories. Recently, another elegant method utilizing genomic data and a high-throughput assay to explore tRNA orthogonality has been described, which effectively added another five orthogonal pairs[8]. Together, the past four weeks have more than doubled the number of orthogonal tRNA/RS pairs, and we can expect that the number of applications will do so as well. Currently, the development of tools for work in *E. coli* is much further advanced than in eukaryotes, but it is becoming easier to transfer genetic code expansion approaches to study eukaryotes[9,10], and we will see another wave of synthetic biology revolutionizing biochemistry and cellular biology. ❐

Christopher D. Reinkemeier[1,2,3] and Edward A. Lemke[1,2,3]✉

[1]*Biocentre, Departments of Biology and Chemistry, Johannes Gutenberg–University Mainz, Mainz, Germany.* [2]*Institute of Molecular Biology, Mainz, Germany.* [3]*Structural and Computational Biology Unit and Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany.*
✉*e-mail: edlemke@uni-mainz.de*

Published online: 29 May 2020
https://doi.org/10.1038/s41557-020-0476-6

References
1. Zhang, Y. et al. *Nature* **551**, 644–647 (2017).
2. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. *Nature* **464**, 441–444 (2010).
3. Fredens, J. et al. *Nature* **569**, 514–518 (2019).
4. Neumann, H., Slusarczyk, A. L. & Chin, J. W. *J. Am. Chem. Soc.* **132**, 2142–2144 (2010).
5. Dunkelmann, D. L., Willis, J. C., Beattie, A. T., Chin, J. W. *Nat. Chem.* https://doi.org/10.1038/s41557-020-0472-x (2020).
6. Willis, J. C. W. & Chin, J. W. *Nat. Chem.* **10**, 831–837 (2018).
7. Meineke, B., Heimgärtner, J., Lafranchi, L. & Elsässer, S. J. *ACS Chem. Biol.* **13**, 3087–3096 (2018).
8. Cervettini, D. et al. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-020-0479-2 (2020).
9. Reinkemeier, C. D., Girona, G. E. & Lemke, E. A. *Science* **363**, eaaw2644 (2019).
10. Richardson, S. M. et al. *Science* **355**, 1040–1044 (2017).

Appendix IV Inducible genetic code expansion in eukaryotes

This work was published as:

Koehler C, Estrada Girona G, Reinkemeier CD & Lemke EA (2020). Inducible Genetic Code Expansion in Eukaryotes. ChemBioChem, 21, 1-5, doi:10.1002/cbic.202000338

Inducible Genetic Code Expansion in Eukaryotes

Christine Koehler,[a, b, c, d] Gemma Estrada Girona,[c] Christopher D. Reinkemeier,[a, b, c] and Edward A. Lemke*[a, b, c]

Genetic code expansion (GCE) is a versatile tool to site-specifically incorporate a noncanonical amino acid (ncAA) into a protein, for example, to perform fluorescent labeling inside living cells. To this end, an orthogonal aminoacyl-tRNA-synthetase/tRNA (RS/tRNA) pair is used to insert the ncAA in response to an amber stop codon in the protein of interest. One of the drawbacks of this system is that, in order to achieve maximum efficiency, high levels of the orthogonal tRNA are required, and this could interfere with host cell functionality. To minimize the adverse effects on the host, we have developed an inducible GCE system that enables us to switch on tRNA or RS expression when needed. In particular, we tested different promotors in the context of the T-REx or Tet-On systems to control expression of the desired orthogonal tRNA and/or RS. We discuss our result with respect to the control of GCE components as well as efficiency. We found that only the T-REx system enables simultaneous control of tRNA and RS expression.

Genetic code expansion (GCE) is a powerful method to site-specifically introduce noncanonical amino acids (ncAAs) into proteins *in vivo*. In order to achieve this, most commonly an orthogonal aminoacyl-tRNA-synthetase/tRNA (RS/tRNA) pair is used. Usually, the anticodon of the orthogonal tRNA is chosen to recognize the amber stop codon (TAG). The orthogonal RS aminoacylates the tRNA with the ncAA, and the tRNA then suppresses the amber codon to site-specifically incorporate the ncAA into the growing peptide chain.[1] The archaea derived PylRS/tRNA^{Pyl} pairs are among the most popular systems as it is orthogonal in both *Escherichia coli* and eukaryotic systems and over 100 different ncAAs have become available for incorporation with this RS/tRNA pair.[2]

To control the expression of synthetase and tRNA, we developed an inducible amber suppression system that gives us the opportunity to switch on GCE technology on demand. We used the T-REx and Tet-On systems to regulate not only the synthetase gene, but also different kinds of tRNA constructs.

RNA polymerase II (Pol II) transcribes protein-coding genes into mRNAs and hence controls the expression levels of the synthetase. tRNA molecules, instead, are transcribed by RNA polymerase III (Pol III). Eukaryotic tRNAs contain internal promoter regions (i.e. A and B boxes); however, the archaeal tRNA^{Pyl} lacks these and therefore requires an external Pol III-specific promoter (e.g., U6 or H1) or a bicistronic expression cassette (derived from Val or Arg tRNA) for efficient transcription.[2b,3] Major components of a U6 or an H1 promoter are, besides the TATA box, the distal and proximal sequence elements (DSE and PSE), which are important for gene expression.[4] It is assumed that high levels of tRNAs are needed for efficient amber suppression.[5] Accumulation of RS and/or suppressor tRNA could have adverse effects on the host machinery, as this can lead to mischarging of tRNA and/or read-through of natural stop codons.[6]

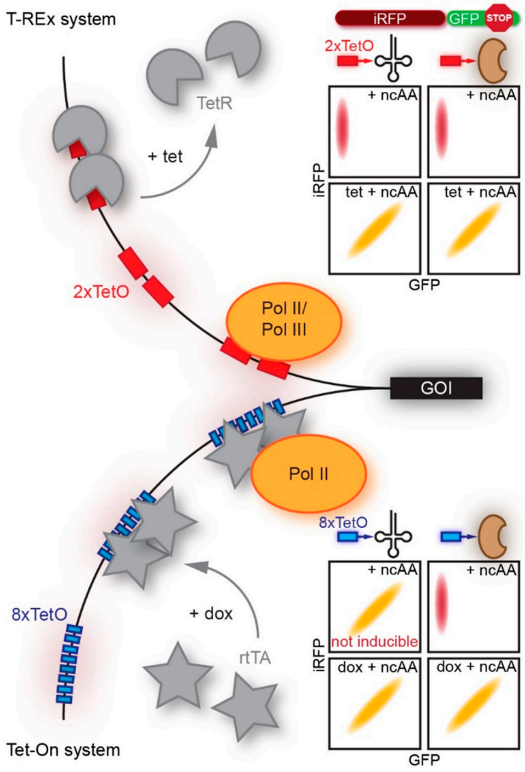
A better control of amber suppression is demanded, as it can help to minimize crosstalk with the host machinery. We developed an inducible GCE method by combining amber suppression technology with known tetracycline inducible systems (Scheme 1). Both the T-REx and Tet-On systems rely on a tet-responsive promoter (Ptet-1 based on the Tn10-specified tetracycline-resistance operon of *E. coli*[7]) in combination with a regulatory element, that is, the tetracycline repressor protein (TetR) or an evolved reverse TetR (rtTA), respectively. This promoter contains two TetO signals, O1 and O2. The TetO2 signal was further used to develop a tetracycline inducible system in eukaryotes (for simplicity we refer to this as TetO). In the T-REx system, TetR binds to the 2xTetO promoter sequence and blocks protein expression. Upon tetracycline (tet) addition TetR undergoes structural changes, unbinds the promoter and transcription can occur.[7,8]

In the Tet-On system, the TetR protein is fused to the transcription activation domain of the herpes simplex virus VP16. Introduction of several mutations lead to the reverse tetracycline-controlled transactivator (rtTA), which binds the promoter region and induces gene expression only in presence of doxycycline (dox).[9] To improve the Tet-On system, several attempts have been undertaken, for example, viral evolution

[a] C. Koehler, C. D. Reinkemeier, Prof. Dr. E. A. Lemke
Biocentre, Johannes-Gutenberg University Mainz
55128 Mainz (Germany)
E-mail: edlemke@uni-mainz.de
[b] C. Koehler, C. D. Reinkemeier, Prof. Dr. E. A. Lemke
Institute of Molecular Biology gGmbH
55128 Mainz (Germany)
[c] C. Koehler, G. Estrada Girona, C. D. Reinkemeier, Prof. Dr. E. A. Lemke
Structural and Computational Biology Unit and Cell Biology and Biophysics Unit
European Molecular Biology Laboratory
Meyerhofstraße 1, 69117 Heidelberg (Germany)
[d] C. Koehler
ARAXA Biosciences GmbH
Meyerhofstraße 1, 69117 Heidelberg (Germany)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/cbic.202000338

This article is part of a Special Collection on Xenobiology. To view the complete collection, visit our homepage



Scheme 1. Overview of T-REx and Tet-On systems in combination with GCE technology. Top: Upon addition of tetracycline (tet), the tet-inducible repressor protein (TetR) unbinds the TetO signals facilitating polymerase (Pol II and Pol III) binding to the promoter. Production of tRNA and translation of the synthetase are inducible in the T-REx system, illustrated by the flow cytometry schemes. The full-length reporter can only be expressed in the presence of tet and ncAA (yellow diagonal ellipse), but not upon addition of ncAA alone (red vertical ellipse). The promoter sequence, containing a 2xTetO signal (red box with arrow) leads to expression of the tRNA or synthetase (shown in brown). Bottom: In the Tet-On system, the reverse Tet-repressor protein (rtTA) can bind to the TetO signals in response to doxycycline (dox), enhancing the binding of Pol II; this enables inducible translation of the synthetase gene, but not a controllable production of tRNA, as illustrated by the flow cytometry schemes. The yellow ellipse represents full-length production of the reporter, whereas the red vertical ellipse illustrates the expression of iRFP alone. An 8xTetO-promoter (blue box with arrow) is positioned in front of the tRNA and synthetase gene (shown in brown).

studies helped to find novel rtTA variants, like Tet-On® 3G (Clontech® Laboratories, Inc.), with increased transcriptional activities and dox-sensitivity.[10] Other studies identified several variants of the promoter Ptet-1 that decrease background expression and increase induced expression levels.[11]

First, we tested the T-REx and Tet-On systems in combination with amber suppression technology, using U6- and H1-driven tRNA expression cassettes. We cloned the tRNA-synthetase, PylRS^{AF} (AF referring to a variant of PylRS 306A 384F) from *Methanosarcina mazei* (Mm) downstream of a nuclear export signal (NES),[6f] under the control of a CMV tet-inducible promoter for the T-REx system (2xTetO) or the CMV dox-inducible promoter for the Tet-On system (8xTetO). To investigate both systems, we followed the expression of a reporter gene, which contains a nuclear localization signal (NLS) upstream of iRFP fused to GFP, harboring the amber stop codon (iRFP-GFP^{Y39TAG}) with fluorescence flow cytometry (FFC). In this reporter, an iRFP signal reports on transfection, whereas full-length GFP and thus green fluorescence is only produced upon successful amber codon suppression by incorporation of a noncanonical amino acid such as *N*-(*tert*-Butoxycarbonyl)-L-lysine (BocK). We designed six different U6- or H1-driven tRNA expression cassettes[12] (Figure 1a). The first set carries TetO sequences surrounding the TATA box for both promoters, and we call these U6-TetO and H1-TetO. The 8xTetO-U6 and 8xTetO-H1 constructs have an additional stretch of 8xTetO sequences upstream of the DSE but no TetO sequences flanking the TATA box. Lastly, the 8xTetO-U6-TetO or 8xTetO-H1-TetO contain both the 8xTetO sequences as well as the two TetO sequences flanking the TATA box.

We tested the inducibility of these constructs via functional amber suppression readout by FFC in the Flp-In™ T-REx™ 293 cell line (Invitrogen), which harbors a stably integrated TetR gene, as well as in HEK293T cells transfected with the Tet-On system using a noninducible NES-PylRS^{AF} as a control and our reporter iRFP-GFP^{Y39TAG}. In the case of the T-REx system, we observe that all constructs containing the TetO sequences flanking the TATA box are inducible by tetracycline (U6-TetO, 8xTetO-U6-TetO, H1-TetO and 8xTetO-H1-TetO). By calculating the ratio of the geometric mean of GFP to iRFP, we show that for these constructs the expression of the full-length reporter, iRFP-GFP^{Y39TAG}, can only be achieved in the presence of tet and BocK, but not with the addition of only BocK. However, if the two TetO sequences surrounding the TATA box are not present (8xTetO-U6 and 8xTetO-H1), tRNA expression is always on and tetracycline independent (Figure 1b). Further controls comparing the data to conventional U6 promoters without any TetO signal are shown in Figure S1 in the Supporting Information. For these constructs, the reporter construct can also be expressed by only adding BocK. In general, higher amber suppression efficiency was observed for U6 versus H1 promoters in line with the known expression strength of those promotors.[13] Next, we tested whether we could also induce the expression of the synthetase gene, and therefore cloned NES-PylRS^{AF} into a plasmid harboring a CMV promoter with 2xTetO sequences (Figure 1c and d). We analyzed the expression of the reporter gene by FFC and, as expected, observed that the synthetase gene is also controllable by tet with this system (Figure S2).

In contrast, the Tet-On system does not allow for inducible U6- or H1-driven tRNA expression (Figures 2 and S3). None of the tested tRNA expression cassettes showed a dox-dependent inducibility but rather a negative influence on the expression level can be seen when dox is present. Differences in the associated mechanisms can explain this lack of inducibility. In the T-REx system, TetR is binding to the promoter region blocking the transcription only when tet is absent. However, in the case of the Tet-On system, rtTA binds the promoter region,

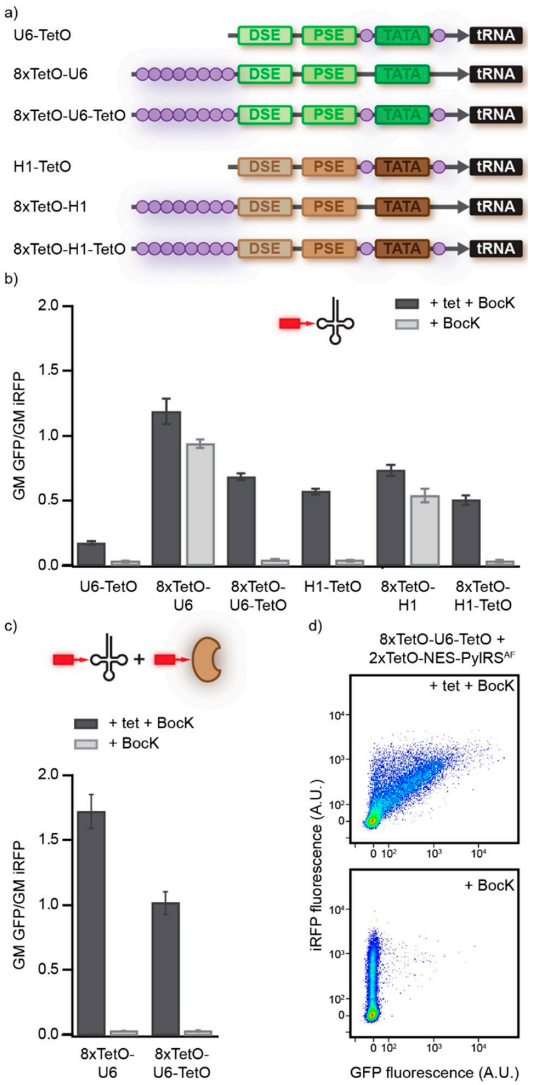


Figure 1. The inducible T-REx system. a) Six different Pol III promoter sequences based on the U6 or H1 promoter are shown. TetO signals (violet circles) are added to specific sites in the promoter. U6-TetO contains two TetO signals, one before and one after the TATA box. 8xTetO-U6 harbors eight TetO signals in front of the distal and proximal sequence elements (DSE and PSE). The 8xTetO-U6-TetO includes the 8xTetO signals before the DSE element and the two TetO signals framing the TATA box. The H1 sequences contain the same pattern as the U6 sequences, but the backbone originating from the H1 promoter. b)–d) HEK Flp-In T-REx 293 cells expressing reporter protein iRFP-GFP^{Y39ncAA} using the T-REx system. b) Measuring the geometric mean (GM) of iRFP and GFP by fluorescence flow cytometry (FFC) using the reporter gene (iRFP-GFP^{Y39ncAA}), the synthetase under CMV promoter in combination with the different Pol III promoters (U6 and H1) with tet and BocK (dark gray) or only with BocK (light gray). The bar plot visualizes how inducible the different Pol III promoters are. c) Bar plot resulting from FFC measurements of the reporter protein using the inducible synthetase gene with and without an inducible tRNA gene. The error bars in b) and c) indicate the standard error of the mean and are calculated from at least three independent measurements. d) FFC data of reporter gene expressed with inducible synthetase construct (2xTetO-CMV-NES-PylRS^{AF}) together with noninducible 8xTetO-U6-TetO-tRNA construct. The upper panel shows expression with tet and BocK, the lower just with BocK.

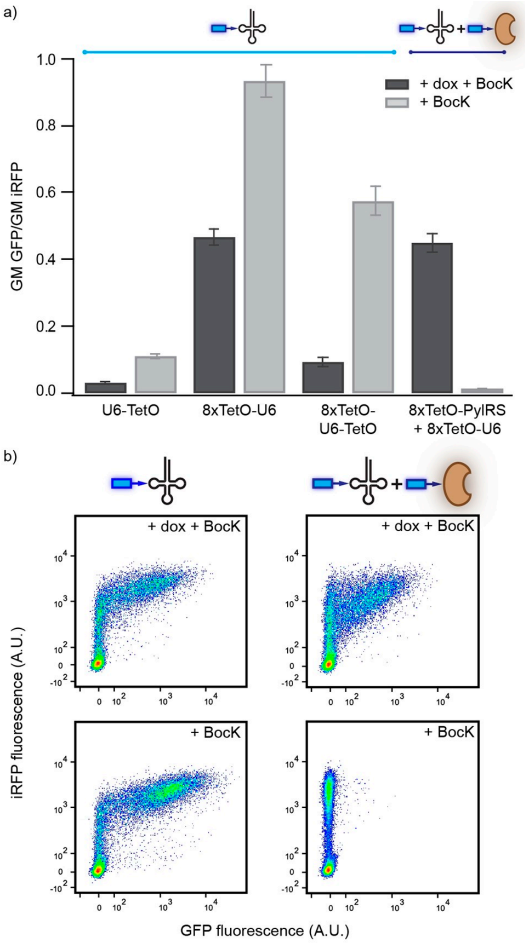


Figure 2. Inducibility of the Tet-On system in HEK293T cells. a) Bar plot illustrating the ratio between the geometric mean (GM) of GFP and iRFP resulting from FFC data measuring expression of iRFP-GFP^{Y39ncAA} with synthetase under CMV promoter in combination with the different Pol III-driven promoter constructs (U6) with and without dox in presence of BocK. The last two columns show the data for the Tet-On system in the presence of the 8xTetO-CMV-NES-PylRS^{AF} construct. None of the three U6 promoters is inducible through the Tet-On system, whereas synthetase expression can be induced by using the Tet-On system. The error bars indicate the standard error of the mean and are calculated from at least three independent measurements. b) FFC data showing expression of iRFP-GFP^{Y39ncAA}. Left: data from the 8xTetO-U6 construct with and without dox (top and bottom, respectively) in the presence of BocK. Right: the inducibility of the GCE system when using the 8xTetO-CMV-NES-PylRS^{AF} construct together with the 8xTetO-U6 construct.

resulting in an activation of transcription in the presence of dox. We speculate that the binding of rtTA to the promoter disturbs the binding of Pol III to the U6 or H1 promoter and the transcription of the tRNA gene cannot occur. As shown in Figure 2, the Tet-On system only allows for control of PylRS expression (8xTetO-CMV-NES-PylRS^{AF}, for more detail see Figure S4).

In summary, we have designed an inducible GCE technology to control the expression level of tRNA as well as of the required synthetase.

We compared different tRNA expression cassettes, derived from U6 and H1 promoter sequences in combination with Tet-On and T-REx systems. We show that for the T-REx system inducibility of the tRNA can be achieved for four Pol III promoter sequences and for the Pol II-dependent promoter of the synthetase gene (Figure 1). On the other hand, it is not possible to induce tRNA expression with the Tet-On system with our current set of tested Pol III promoters, and we speculate that this is caused by an inability of Pol III to bind the promoter sequence as long as rtTA is bound. Only the synthetase translation can be controlled in this system because Pol II is recruited by rtTA and hence can bind to the promoter sequence and transcribe the synthetase gene (Figure 2).

With the T-REx system it is now possible to directly switch on tRNA expression in mammalian cells together with or independently of the RS whenever needed. One disadvantage of the T-REx system is the need for TetR to be present and bound to its promotor before gene expression can be controlled by tet. This system suffers from high background expression in transient transfection because all plasmids are introduced to the cells simultaneously, whereas the TetR protein should be produced first in order to bind the promoter and block Pol II- or Pol III-based transcription. Therefore, the T-REx system is only useful if the TetR protein is stably integrated into the cell line, as in HEK293T FlpIn T-REx 293 cells and not suitable in a transient transfection using for example HEK293T cells (data not shown). However, various genome engineering tools to make stable cell lines expressing proteins via Pol II have now become available. In contrast, achieving high-yielding tRNA expression in stable cell lines is still a huge challenge.[14] Methods like CRISPR and even transposons only introduce a few copies into the genome, whereas in transient transfections, easily 100s of tRNA genes are transferred into the cell.

Once this challenge is addressed, we expect the T-REx system to be useful to generate stable mammalian cell lines for amber suppression with minimized impact of the GCE machinery on housekeeping and physiological function of the cell.

Acknowledgements

The authors thank all the members of the Lemke lab for constructive discussions. We acknowledge funding support by ERC SMPFv2.0, the Gutenbergforschungskolleg and SFB1129 and SPP1623 of the DFG (German Research Foundation, Project no. 240245660, 223208190). Open access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

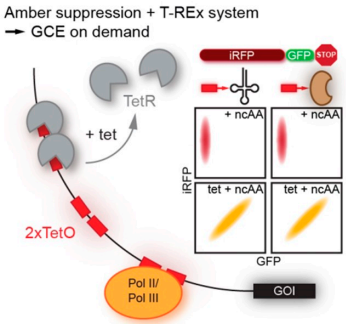
Keywords: amber suppression · PylRS · Tet-On · T-REx · unnatural amino acid

[1] a) C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444; b) J. W. Chin, *Annu. Rev. Biochem.* **2014**, *83*, 379–408; c) W. Wan, J. M. Tharp, W. R. Liu, *Biochim. Biophys. Acta* **2014**, *1844*, 1059–1070; d) J. S. Italia, Y. Zheng, R. E. Kelemen, S. B. Erickson, P. S. Addy, A. Chatterjee, *Biochem. Soc. Trans.* **2017**, *45*, 555–562.
[2] a) P. R. Chen, D. Groff, J. Guo, W. Ou, S. Cellitti, B. H. Geierstanger, P. G. Schultz, *Angew. Chem. Int. Ed.* **2009**, *48*, 4052–4055; *Angew. Chem.* **2009**, *121*, 4112–4115; b) T. Mukai, T. Kobayashi, N. Hino, T. Yanagisawa, K. Sakamoto, S. Yokoyama, *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818–822; c) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232–234; d) E. A. Lemke, *ChemBioChem* **2014**, *15*, 1691–1694.
[3] S. M. Hancock, R. Uprety, A. Deiters, J. W. Chin, *J. Am. Chem. Soc.* **2010**, *132*, 14819–14824.
[4] P. Carbon, S. Murgo, J. P. Ebel, A. Krol, G. Tebb, L. W. Mattaj, *Cell* **1987**, *51*, 71–79.
[5] a) A. Garcia, G. Roy, C. Kiefer, S. Wilson, M. Marelli, *PLoS One* **2019**, *14*, e0216356; b) W. H. Schmied, S. J. Elsasser, C. Uttamapinant, J. W. Chin, *J. Am. Chem. Soc.* **2014**, *136*, 15577–15583; c) N. M. Reynolds, O. Vargas-Rodriguez, D. Soll, A. Crnkovic, *Biochim. Biophys. Acta* **2017**, *1861*, 3001–3008.
[6] a) H. J. Park, U. L. RajBhandary, *Mol. Cell. Biol.* **1998**, *18*, 4418–4425; b) J. Guo, C. E. Melancon III, H. S. Lee, D. Groff, P. G. Schultz, *Angew. Chem. Int. Ed.* **2009**, *48*, 9148–9151; *Angew. Chem.* **2009**, *121*, 9312–9315; c) H. Tian, D. Deng, J. Huang, D. Yao, X. Xu, X. Gao, *Biochimie* **2013**, *95*, 881–888; d) A. Bhattacharya, C. Kohrer, D. Mandal, U. L. RajBhandary, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6015–6020; e) C. D. Reinkemeier, G. E. Girona, E. A. Lemke, *Science* **2019**, *363*; f) I. Nikic, G. Estrada Girona, J. H. Kang, G. Paci, S. Mikhaleva, C. Koehler, N. V. Shymanska, C. Ventura Santos, D. Spitz, E. A. Lemke, *Angew. Chem. Int. Ed.* **2016**, *55*, 16172–16176; *Angew. Chem.* **2016**, *128*, 16406–16410.
[7] M. Gossen, H. Bujard, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 5547–5551.
[8] A. T. Das, L. Tenenbaum, B. Berkhout, *Curr. Gene Ther.* **2016**, *16*, 156–167.
[9] M. Gossen, S. Freundlieb, G. Bender, G. Muller, W. Hillen, H. Bujard, *Science* **1995**, *268*, 1766–1769.
[10] a) A. T. Das, X. Zhou, M. Vink, B. Klaver, K. Verhoef, G. Marzio, B. Berkhout, *J. Biol. Chem.* **2004**, *279*, 18776–18782; b) X. Zhou, M. Vink, B. Klaver, B. Berkhout, A. T. Das, *Gene Ther.* **2006**, *13*, 1382–1390.
[11] R. Loew, N. Heinz, M. Hampf, H. Bujard, M. Gossen, *BMC Biotechnol.* **2010**, *10*, 81.
[12] H. Zhou, C. Huang, X. G. Xia, *Biochim. Biophys. Acta* **2008**, *1779*, 773–779.
[13] P. I. Makinen, J. K. Koponen, A. M. Karkkainen, T. M. Malm, K. H. Pulkkinen, J. Koistinaho, M. P. Turunen, S. Yla-Herttuala, *J. Gene Med.* **2006**, *8*, 433–441.
[14] a) S. J. Elsasser, R. J. Ernst, O. S. Walker, J. W. Chin, *Nat. Methods* **2016**, *13*, 158–164; b) S. Shao, M. Koh, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 8845–8849; c) C. Koehler, P. F. Sauter, M. Wawryszyn, G. E. Girona, K. Gupta, J. J. Landry, M. H. Fritz, K. Radic, J. E. Hoffmann, Z. A. Chen, J. Zou, P. S. Tan, B. Galik, S. Junttila, P. Stolt-Bergner, G. Pruneri, A. Gyenesei, C. Schultz, M. B. Biskup, H. Besir, V. Benes, J. Rappsilber, M. Jechlinger, J. O. Korbel, I. Berger, S. Braese, E. A. Lemke, *Nat. Methods* **2016**, *13*, 997–1000; d) B. Shen, Z. Xiang, B. Miller, G. Louie, W. Wang, J. P. Noel, F. H. Gage, L. Wang, *Stem Cells* **2011**, *29*, 1231–1240.

Manuscript received: May 28, 2020
Revised manuscript received: June 25, 2020
Accepted manuscript online: June 29, 2020
Version of record online: ■■, ■■■

COMMUNICATIONS

Genetic code expansion on demand by combining the amber suppression technology with the T-REx system. Amber suppression is regulated by the T-REx system to switch on the tRNA production and/or the synthetase translation only when needed, thus minimizing the risk of adverse effects for the cells resulting from over-production of unused tRNA or synthetase.



*C. Koehler, G. Estrada Girona, C. D. Reinkemeier, Prof. Dr. E. A. Lemke**

ChemBioChem

Supporting Information

Inducible Genetic Code Expansion in Eukaryotes

Christine Koehler, Gemma Estrada Girona, Christopher D. Reinkemeier, and Edward A. Lemke*

Method section

Cell culture

Transient transfections

Cloning of constructs

Fluorescence flow cytometry (FFC)

Supplementary Figures

Supporting Figure S1: FFC analysis of different U6-or H1-based promoter sequences in the T-REx system

Supporting Figure S2: FFC analysis of the inducible synthetase construct in the T-REx system

Supporting Figure S3: FFC analysis of different U6-promoter sequences in the Tet-On system

Supporting Figure S4: FFC analysis of the inducible synthetase construct in the Tet-On system

References

Method section

Cell culture

HEK293T and Flp-In™ T-REx™ 293 (Invitrogen) cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco 41965-039) supplemented with 10% FBS (Sigma-Aldrich F7524), 1% penicillin-streptomycin (Sigma-Aldrich P0781), 1% L-Glutamine (Sigma-Aldrich G7513) and 1% sodium pyruvate (Life Technologies 11360) at 37 °C and 5% CO_2. The cells were passaged every 2-3 days at 1:10 or 1:20 ratios. The Flp-In T-REx 293 cells contain the TetR gene stably integrated downstream of a CMV promoter sequence. For fluorescence flow cytometry (FFC), cells were seeded 16 hours prior to transfections in 24-well plates (Nunclon Delta Surface ThermoFisher Scientific) at a cell density of 220000 cells/mL, 500 µL/well in the case of HEK293T cells and 250000 cell/mL, 500 µL/well for Flp-In T-REx 293 cells.

Transient transfections

For each well of a 24-well plate 1500 ng of total DNA was used. Equal amounts of the desired plasmids were mixed in 50 µL DMEM without phenol red. Polyethylenimine (PEI, 1 µg/µL, Sigma 408727-100ml) was added in a ratio of 1:3 (DNA:PEI). The mixture was vortexed for 10 seconds, spun down for 5 seconds and incubated for 15 minutes at room temperature before it was added drop wise to the well. Master mixes were prepared wherever adequate. After 4 hours, the medium was aspirated and fresh medium containing 250 µM of the required ncAAs and 1 µg/mL tet or dox was pipetted on the cells. After 48 hours, the cells were analyzed by fluorescence flow cytometry.

Cloning of constructs

The reporter gene, NLS-iRFP-GFP^{Y39TAG} was cloned into a pCI plasmid (Promega, E1731) as previously described[1]. GFP contains an amber stop codon (TAG) at position Y39. The synthetase gene (*Methanosarcina mazei* PylRS^{AF}) was cloned upstream of a CMV promoter with an N-terminal nuclear export signal (NES) and contains two mutations (Y306A and Y384F)[1].

We designed six different tRNA expression constructs, originating from human U6 or H1 promoter sequences (as shown in Figure 1 a in the main text)[2]. The first construct, U6-TetO contained two tet-operator (TetO2) signals (5' tccctatcagtgatagaga 3'), one before and one after the TATA sequence. The same logic also applied for the H1-TetO promoter sequence. For simplicity, we refer to this TetO2 sequences as "TetO" throughout the paper. The 8xTetO-U6 or 8xTetO-H1 promoter sequences comprised eight consecutive TetO signals separated by small linker regions in front of the distal sequence element (DSE). The combination of both constructs resulted in the 8xTetO-U6-TetO or 8xTetO-H1-TetO promoter sequences. These promoters contained two TetO signals framing the TATA box as well as the eight TetO repeats before the DSE element. In addition, a U6-tRNA cassette was cloned, which did not contain any TetO signal.

For the T-REx system the plasmid for the inducible PylRS construct contained two TetO2 signals directly after the CMV promoter upstream of the NES-PylRS^{AF} gene[3]. Eight TetO2 signals were required for the Tet-On system, which was integrated into a CMV promoter upstream of NES-PylRS^{AF} resulting into the plasmid pcDNA3.1_8xTetO-NES-PylRS^{AF}[4].

The plasmid containing the TetOn3G gene had a pcDNA3.1-Zeo backbone and contained a CMV promoter upstream of the TetOn3G gene.

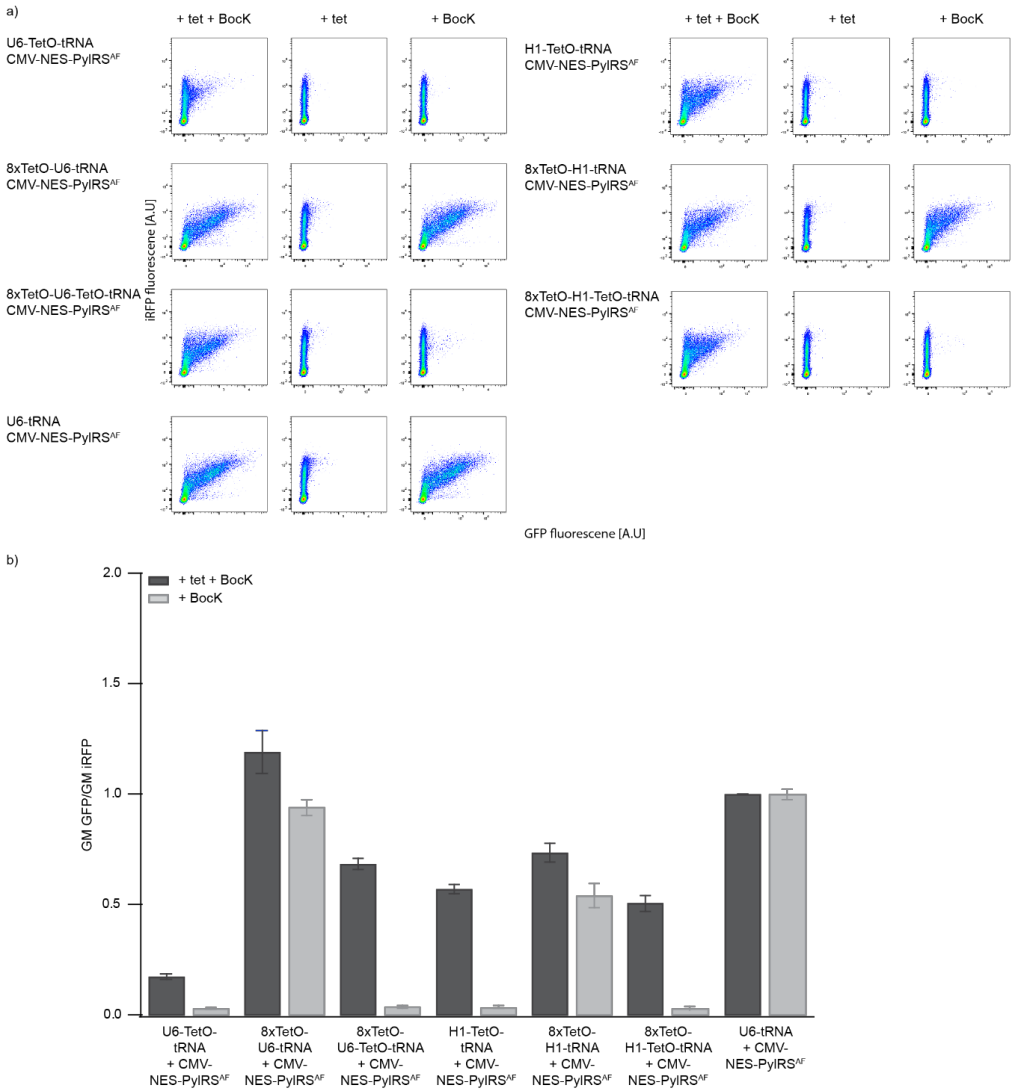
Fluorescence flow cytometry (FFC) and data analysis

Transient transfections were analyzed after 48 hours incubation with a flow cytometer analyzer (LSRFortessa™, BD Biosciences) using the 488 nm laser with a 530-30 filter and the 640 nm laser with a 730-45 filter. After gating for live cells and single cells using FlowJo software (BD Biosciences), the cells were divided into transfected and untransfected cells based on the iRFP signal. For the transfected cell population, the geometric mean of GFP and iRFP was determined. The ratio of the geometric mean of GFP to iRFP of the different tRNA cassettes was normalized to the ratio obtained from the U6-tRNA measurement, which was measured using the CMV-NES-PylRS^{AF} construct.

Supplementary Figures

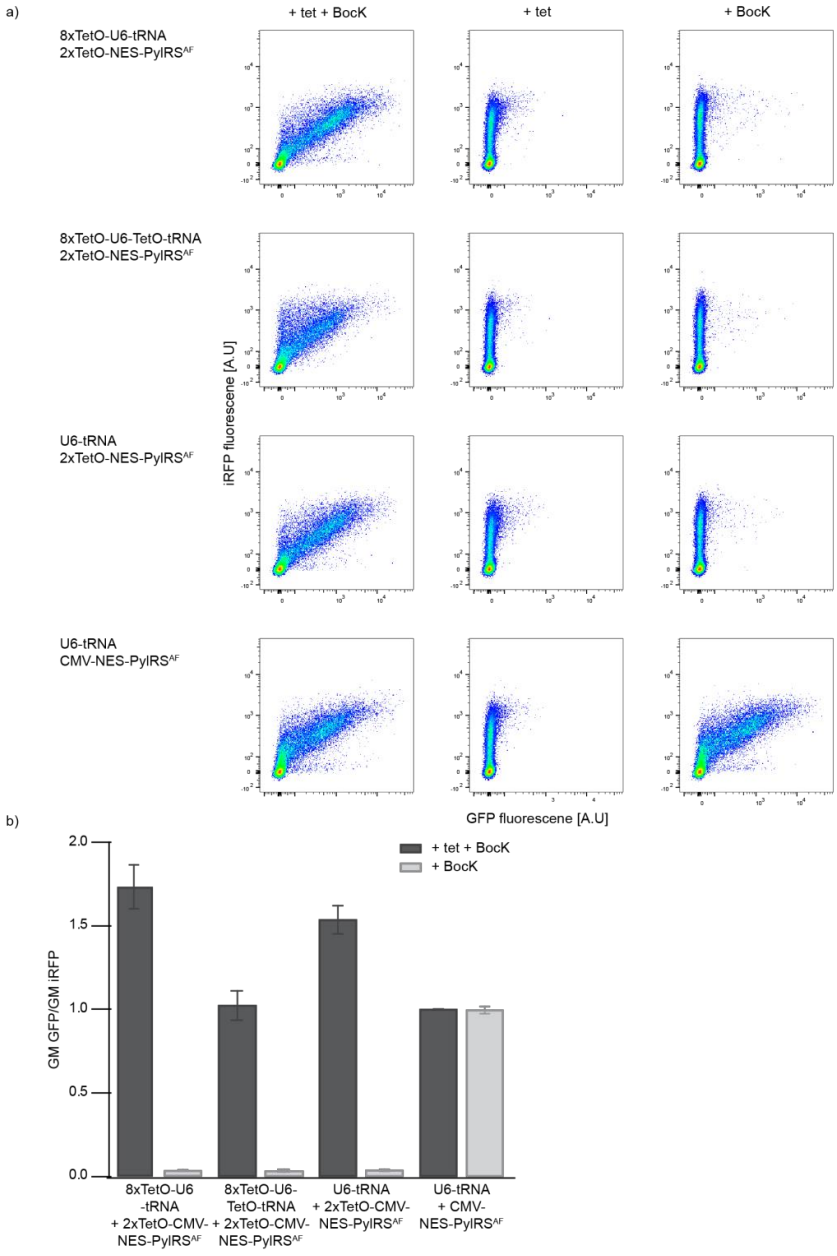
Supporting Figure S1 (corresponding to maintext Figure 1): FFC analysis of different U6- or H1-based promoter sequences in the T-REx system.

Analysis of the different U6- and H1-based promoter sequences as shown in Scheme 1 with different numbers of TetO-repeats at different positions in the promoter sequence, or without any TetO signal, in the case of U6. a) Shown are representative FFC analysis of NLS-iRFP-GFP^{Y39TAG} with tet and BocK expressed in Flp-In T-REx 293 cells, with only tet or with only BocK present in the medium. Four out of the seven promoters are inducible via the T-REx system, which is dependent on the addition of tet to the medium. b) The ratio of the geometric mean (GM) of GFP to iRFP of the U6-tRNA construct in presence of tet and BocK was used to normalize the obtained data for the other constructs. Shown are the mean values of at least three independent experiments, error bars represent the standard error of the mean (SEM).



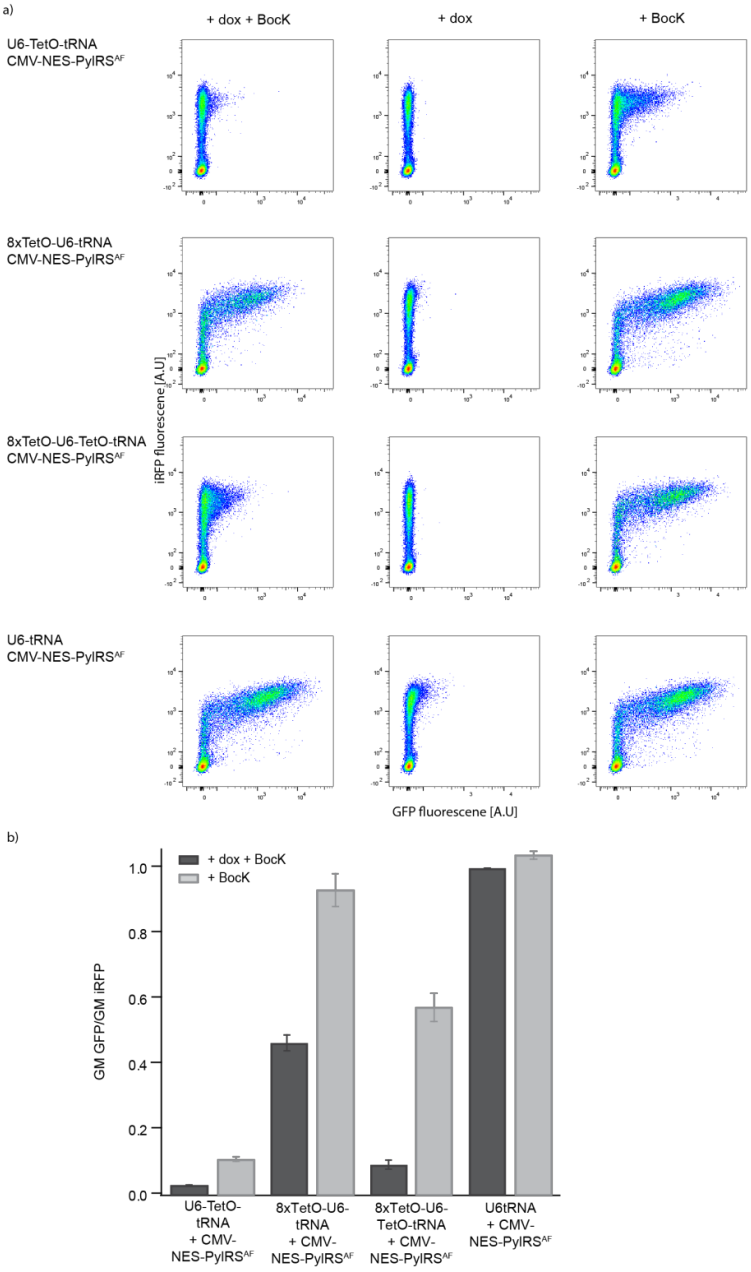
Supporting Figure S2 (corresponding to maintext Figure 1): FFC analysis of the inducible synthetase construct in the T-REx system.

a) Representative FFC experiments of the inducible synthetase construct (2xTetO-NES-PylRS^{AF}) for the T-REx system. Flp-In T-REx 293 cells containing the reporter gene in combination with different tRNA cassettes, as indicated in the figure. First panel: 8xTetO-U6-tRNA in combination with 2xTetO-NES-PylRS^{AF}, second panel: 8xTetO-U6-TetO-tRNA with 2xTetO-NES-PylRS^{AF}, third panel: U6-tRNA with 2xTetO-NES-PylRS^{AF} and fourth panel: U6-tRNA with CMV-NES-PylRS^{AF} as control. The three two panels show an inducible GCE system demonstrating the inducibility of the synthetase gene under the 2xTetO promoter sequences for the T-REx system, which is tet dependent. b) The ratios of the geometric mean (GM) of GFP to iRFP for the different U6-constructs are shown in the barplot.The ratio of the geometric mean (GM) of GFP to iRFP of the U6-tRNA construct in presence of tet and BocK was used to normalize the obtained data for the other constructs. Shown are the mean values of at least three independent experiments, error bars represent the standard error of the mean (SEM).



Supporting Figure S3 (corresponding to maintext Figure 2): FFC analysis of different U6- promoter sequences in the Tet-On system.

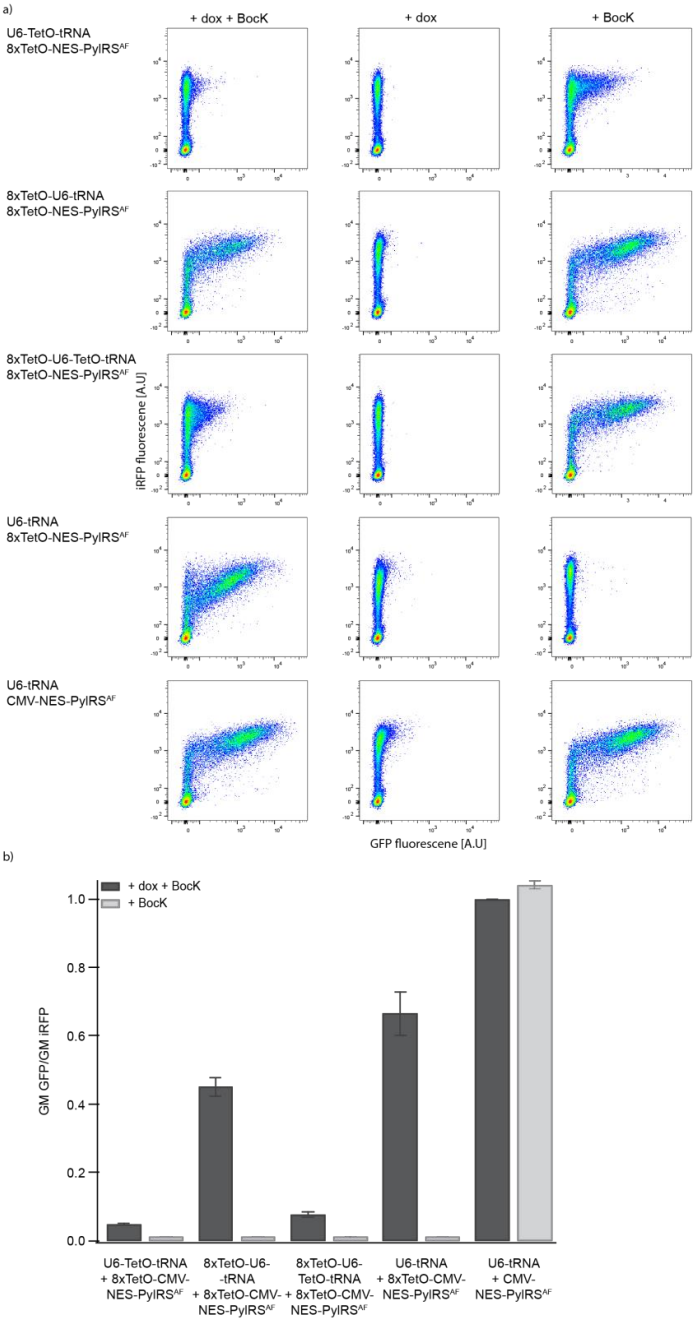
a) Representative FFC analysis of reporter gene NLS-iRFP-GFP^{Y39TAG} in HEK293T cells for the Tet-On system with different U6-promoter constructs. None of the three constructs show dox dependent expression and GCE induction. In case of the U6-TetO construct, there is even a negative effect of translational efficiency upon addition of dox. b) The ratios of the geometric mean (GM) of GFP to iRFP for the different U6-constructs are shown in the barplot. The GM of GFP to iRFP of the U6-tRNA construct in presence of dox and BocK was used to normalize the obtained data for the other constructs. Shown are the mean values of at least three independent experiments, error bars represent the standard error of the mean (SEM).



.

Supporting Figure S4 (corresponding to maintext Figure 2): FFC analysis of the inducible synthetase construct in the Tet-On system.

a) Representative FFC analysis of reporter gene NLS-iRFP-GFP^{Y39TAG} in in HEK293T cells for the Tet-On system with different U6-promoter constructs with and without an inducible synthetase construct. The upper four panels show the FFC data of different tRNA constructs (as indicated in the figure) with the inducible 8xtetO-NES-PylRS^{AF} construct, the lowest one the U6-tRNA construct in combination with CMV-NES-PylRS^{AF}. b) The ratios of the geometric mean (GM) of GFP to iRFP for the different U6-constructs are shown in the barplot. The GM of GFP to iRFP of the U6-tRNA construct in presence of dox and BocK was used to normalize the obtained data for the other constructs. Shown are the mean values of at least three independent experiments, error bars represent the standard error of the mean (SEM).



Supplementary References

[1] I. Nikic, G. Estrada Girona, J. H. Kang, G. Paci, S. Mikhaleva, C. Koehler, N. V. Shymanska, C. Ventura Santos, D. Spitz, E. A. Lemke, *Angew Chem Int Ed Engl* **2016**, *55*, 16172-16176.

[2] H. Zhou, C. Huang, X. G. Xia, *Biochim Biophys Acta* **2008**, *1779*, 773-779.

[3] M. Gossen, H. Bujard, *Proc Natl Acad Sci U S A* **1992**, *89*, 5547-5551.

[4] a) M. Gossen, S. Freundlieb, G. Bender, G. Muller, W. Hillen, H. Bujard, *Science* **1995**, *268*, 1766-1769; b) X. Zhou, M. Vink, B. Klaver, B. Berkhout, A. T. Das, *Gene Ther* **2006**, *13*, 1382-1390; cA. T. Das, X. Zhou, M. Vink, B. Klaver, K. Verhoef, G. Marzio, B. Berkhout, *J Biol Chem* **2004**, *279*, 18776-18782.

Appendix V Curriculum vitae

PERSONAL DETAILS

Address	Albert-Mays Str. 1 69115 Heidelberg Germany
Citizenship	German

EDUCATION

09/2016-Present	**European Molecular Biology Laboratory Heidelberg, Structural and Computational Biology Unit, Germany**
10/2014-06/2016	**University of Konstanz, Germany** M.Sc. Life Science, with distinction (grade: 1.0)
10/2011-09/2014	**University of Konstanz, Germany** B.Sc. Life Science, with distinction (grade: 1.2)

AWARDS

10/2020	International Birnstiel Award for Doctoral Research in Molecular Life Sciences
02/2020	BPS Travel Award 2020
12/2019	ASCB Travel Grant
09/2019	Best poster award EMBO Workshop 'Creating is Understanding: Synthetic Biology Masters Complexity'
10/2014	Prize for the best Bachelor degree in 'Life Science' by the Alumni Association of the University of Konstanz
04/2012-06/2016	Scholarship by the 'Konrad-Adenauer-Stiftung'

INVITED TALKS

09/2020 Opening lecture 3rd BioHackathon, Vilnius (Lithuania)

FULL PUBLICATION LIST

Koehler C, Estrada Girona G, Reinkemeier CD & Lemke EA (2020). Inducible Genetic Code Expansion in Eukaryotes. ChemBioChem, doi:10.1002/cbic.202000338

Reinkemeier CD & Lemke EA (2020). Raising the ribosomal repertoire. Nature Chemistry, 12, 503-504, doi:10.1038/s41557-020-0476-6

Reinkemeier CD*, Estrada Girona G* & Lemke EA (2019). Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. Science, Vol 363, Issue 6434, doi:10.1126/science.aaw2644

Beránek V*, Reinkemeier CD*, Zhang MS, Liang AD, Kym G & Chin JW (2018). Genetically Encoded Protein Phosphorylation in Mammalian Cells. Cell Chemical Biology, Vol 25, Issue 9, pp.1067-1074, doi:10.1016/j.chembiol.2018.05.013

Garcia Castro M, Kremer L, Reinkemeier CD, Unkelbach C, Strohmann C, Ziegler S, Ostermann C & Kumar K (2015). De novo branching cascades for structural and functional diversity in small molecules. Nature Communications, 6(3): 6516, pp. 1-13. doi:10.1038/ncomms7516.

* These authors contributed equally.