Repurposing the translation apparatus for synthetic biology

Benjamin J. Des Soye^{1,2,3,4†}, Jaymin R. Patel^{5,6†}, Farren J. Isaacs^{5,6*}, Michael C. Jewett^{1,2,3,4,7*}

¹ Interdisciplinary Biological Sciences Program, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA.

² Northwestern Institute on Complex Systems, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA.

³ Simpson Querrey Institute, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA.

⁴ Chemistry of Life Processes Institute, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA.

⁵ Systems Biology Institute, Yale University, West Haven, CT, 06516, USA,

⁶ Department of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06510, USA

⁷ Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA.

† Authors contributed equally to this work.

* To whom correspondence should be addressed:

Email: m-jewett@northwestern.edu, farren.isaacs@yale.edu

Keywords: synthetic biology, orthogonal translation systems, unnatural amino acids, genomically recoded organisms, cell-free biology

Submitted to *Current Opinion in Chemical Biology*

Abstract

The translation system (the ribosome and associated factors) is the cell's factory for protein synthesis. The extraordinary catalytic capacity of the protein synthesis machinery has driven extensive efforts to harness it for novel functions. For example, pioneering efforts have demonstrated that it is possible to genetically encode more than the 20 natural amino acids and that this encoding can be a powerful tool to expand the chemical diversity of proteins. Here, we discuss recent advances in efforts to expand the chemistry of living systems, highlighting improvements to the molecular machinery and genomically recoded organisms, applications of cell-free systems, and extensions of these efforts to include eukaryotic systems. The transformative potential of repurposing the translation apparatus has emerged as one of the defining opportunities at the interface of chemical and synthetic biology.

I. Background

Proteins represent a critical class of biomolecules, universally employed by *all* living organisms to fulfill essential structural, functional, and enzymatic roles necessary to support life. In nature, these polymers are composed generally of twenty natural amino acid (AA) building blocks, which can be combined in a near-infinite number of combinations to generate an impressive level of structural and functional diversity (Figure 1). However, many interesting chemistries cannot be accessed using only these natural building blocks; accordingly, for some time there has been an interest in the incorporation of non-standard amino acids (nsAAs) featuring novel functional sidegroups to expand the repertoire of protein functions.

Broadly speaking, nsAAs can be divided into two classes. *Synthetic nsAAs* are chemically synthesized, and can bear little resemblance to their naturally occurring counterparts. *Posttranslational modifications* (PTMs) are modified derivatives of canonical amino acids. In recent years, two distinct approaches for the incorporation of nsAAs into proteins have emerged (Figure 1). One such approach is *global suppression*. This method uses auxotrophic strains that are incapable of synthesizing a particular AA. When grown in the presence of a nsAA that bears close structural resemblance to the “missing” AA, the organism’s native translational machinery incorporates the nsAA instead^{1,2}. An alternative approach uses Orthogonal Translation Systems (OTSs) to genetically encode an nsAA of interest site-specifically by reassignment of codons, typically the amber stop codon (TAG) in a strategy known as *amber suppression*³.

To date, >150 nsAAs have been incorporated by OTSs into peptides⁴ for a wide range of applications including the introduction of bioorthogonal handles for protein tagging^{5,6}, alteration of protein stability^{7,8}, monitoring of protein localization, and genetic encoding of PTMs⁹⁻¹². As a result of these impressive efforts and the transformative potential to construct bio-based products beyond natural limits, expanding the genetic code has emerged as one of several major defining opportunities and points of synergy in chemical and synthetic biology.

This review focuses on recent developments in repurposing the translation system for novel functions, with a focus on codon reassignment. We first examine development of the molecular machinery at the heart of genetic code expansion. Next, we discuss nsAA incorporation in several contexts, including whole-genome recoding, prokaryotic and eukaryotic systems *in vivo* as well as *in vitro*. We end with a discussion of current challenges in the field and provide commentary on future opportunities.

II. Genetic code expansion using OTSs

Amber suppression seeks to “hijack” the amber translational stop codon (TAG), recoding it into a sense codon corresponding to a nsAA of interest. Generally, this is accomplished using a *suppressor tRNA* that has been mutated to decode the amber codon and an *aminoacyl-tRNA synthetase* (AARS) that has been mutated to accept the nsAA of interest and covalently load it onto the suppressor tRNA. These components are typically sourced from distant archeal species to ensure orthogonality to host translation machinery, undergoing directed evolution to improve their compatibility with a new nsAA and enable its site-specific incorporation into proteins.

Directed evolution is the most widely-used approach for the generation of novel OTS components¹³⁻¹⁵ (Figure 2). These efforts start with the selection of a scaffold AARS/tRNA pair. To date, several AARS/tRNA pairs have been used in the creation of new OTSs. The *Methanocaldococcus jannaschii* TyrRS/tRNA^{Tyr} pair is arguably the most common pair used, but is generally limited to aromatic amino acids and is not orthogonal in eukaryotes^{4,15,16}. The PylRS/tRNA^{Pyl} pair from *Methanosarcina* species (*M. mazei*, *M. barkeri*) has shown compatibility with eukaryotic systems^{16,17}, and is an especially attractive starting point for evolution as the native PylRS natively demonstrates polysubstrate specificity¹⁸ and tRNA^{Pyl} natively decodes the amber codon¹⁹. Other starting components have included the o-phosphoserine (Sep)RS from *Methanococcus maripaludis*⁹ and the TrpRS/tRNA^{Trp} pair from *Saccharomyces cerevisiae*²⁰. The TyrRS/tRNA^{Tyr} and LeuRS/tRNA^{Leu} pairs from *Escherichia coli* have also been used in OTS development for use in higher organisms^{21,22}. After selecting a scaffold pair, crystal structural data is commonly used to identify specific residues in the AARS that interact with the amino acid and the tRNA^{13,23}. These residues are randomized to create a library of mutant AARS variants *in vitro* and subsequently transformed *in vivo*. Finally, alternating rounds of selection identify mutant variants that are both functional and orthogonal to native machinery^{13,15}. Negative selections eliminate variants that can charge the o-tRNA with native amino acids based on synthesis of a toxic gene (e.g., barnase) in the absence of the cognate nsAA. Positive selections in the presence of the nsAA isolate variants that can charge the nsAA of interest onto the o-tRNA based on the suppression-dependent synthesis of a selectable marker (e.g., antibiotic resistance gene) or reporter (e.g., GFP). By subjecting “winners” to alternating rounds of these screens, a nsAA-specific AARS/tRNA pair can be identified.

While most new OTSs have focused on the generation of new nsAA-AARS-tRNA pairs, some efforts have expanded OTSs to include additional components. In an exemplary study, incorporation of Sep was *only possible* after the development of a Sep-specific elongation factor⁹. More recently, researchers were able to increase incorporation efficiency of selenocysteine (Sec) to >90% by engineering an elongation factor optimized for Sec²⁴. NsAA incompatibility with the ribosome may also be an impediment to incorporation²⁵. Platforms for ribosomal engineering and evolution will be integral to the elucidation and optimization of ribosome/nsAA interactions. Two recent approaches permit construction of modified ribosomes by decoupling organism fitness from ribosome function. In the first, researchers engineered a 30S subunit that is made orthogonal to natural components by a mutant 16S rRNA; this orthogonal subunit can be mutated to evolve novel function without impairing host viability^{26,27}. In a parallel approach, a system for the *in vitro* assembly of functional ribosomes has been developed²⁸⁻³⁰.

III. Prokaryotic strain engineering tailored for genetic code expansion

Historically, nsAA incorporation via natural codon suppression has been significantly limited by native translational components that have evolved essential function to faithfully decode all codons within an open reading frame. In the case of amber suppression, *release factors* are a class of proteins responsible for facilitating the termination of translation in response to ribosomal stalling at a stop codon. Competition

between loaded suppressor tRNAs and release factor proteins at amber codons meant to encode nsAAs severely hinders successful suppression, with release factor activity resulting in the premature truncation of most of the protein product³¹. Resultant yields of the target nsAA-containing protein under these conditions are very low, especially for proteins containing multiple instances of the nsAA³². Conversely, the presence of OTS on high copy plasmids in the presence of high concentrations of nsAA *in vivo* can also drive the incorporation of nsAAs at >300 amber codons that terminate native genes, resulting in cellular toxicity³³.

Consequently, much effort has gone into the elimination of release factor activity to improve nsAA incorporation. Initial attempts to outright delete the essential gene *prfA* that encodes release factor 1 (RF1) were stymied by cell inviability. Several early approaches, including release factor engineering³⁴ and supplementation *in trans* with partially-recoded versions of the amber-dependent essential genes³⁵ permitted subsequent removal of *prfA* from the organism. More recent efforts have recoded the genome of *E. coli*. In such efforts, researchers edited the amber-dependent essential genes to terminate instead with the synonymous ochre codon (TAA) and deleted *prfA* from the organism. Incorporation of Sep was greatly enhanced in one such RF1-deficient recoded strain³⁶. More recently, a recoded RF1-deficient strain was used in the preparation of cell lysates to improve the incorporation of the nsAA p-propargyloxy-L-phenylalanine (pPaF) *in vitro*^{32,37}.

These strain engineering efforts culminated with the recently-reported completion of the first completely genomically recoded organism (GRO), an *E. coli*-derived strain that lacks all amber codons and RF1^{33,38}. In this study, the authors systematically reassigned *all* 321 native instances of the amber (TAG) codon to the ochre (TAA) codon and deleted the *prfA* gene (Figure 3). In the resulting strain (C321ΔA) the amber codon is orthogonal and unrecognized by the remaining translational machinery, freeing it for use as a dedicated codon for nsAA incorporation. This strain has demonstrated improved properties for incorporation of nsAAs³³ and has more recently been engineered to depend on nsAAs as a 21st synthetic biochemical building block.^{39,40}

Beyond amber suppression, some effort has been made to access additional codons for nsAA assignment, such as the “ochre” (TAA)⁴¹ and “opal” (TGA) stop codons⁴². In a parallel approach, researchers instead sought to access non-natural *quadruplet* codons to encode nsAAs^{43,44}. Combined with amber suppression, quadruplet suppression has enabled the incorporation of multiple distinct nsAAs into a single peptide⁴⁵⁻⁴⁷.

IV. Cell-free systems for genetic code expansion

Cell-free protein synthesis (CFPS) systems offer another approach for the incorporation of nsAAs (Figure 3). CFPS is the *in vitro* synthesis of proteins without using intact cells⁴⁸⁻⁵⁰. A lack of physical boundaries permits precise manipulation of reaction contents, simplifies product purification, and enables efficient incorporation of bulky/charged nsAAs that typically exhibit poor membrane permeability *in vivo* (*i.e.*, pPaF)⁵¹. Additionally, cell viability is no longer a constraint – reactions may be supplemented directly with purified OTS components, which have known toxicity effects *in vivo*⁵². These advantages, when combined with recent improvements to chassis strains and energy regeneration systems that have enabled high-yielding (>1 g/L)⁴⁹ and

long-lasting (>10 h batch mode)³⁷ synthesis, have led to increased interest in the use of CFPS for nsAA incorporation into proteins^{32,37,51,53-55}.

One particularly exciting area for CFPS is in genetic code reassignment. For instance, translation systems reconstituted from purified components⁵⁶ or crude cell extracts depleted of native tRNAs⁵⁷ can be selectively supplied with purified tRNAs to essentially create a custom genetic code with select sense codons left “blank” for nsAA reassignment^{58,59}. Alternatively, in a recent effort a 50S ribosomal subunit was developed with mutations at the peptidyl transferase center that abolish its ability to use native tRNAs for translation⁶⁰. The use of this ribosome together with tRNAs containing compensatory point mutations enables arbitrary rewriting of the genetic code, with select native sense codons reserved for nsAA encoding. Finally, *in vitro* nsAA-incorporation efforts uniquely benefit from the use of self aminoacylating tRNAs featuring *flexizyme* ribozymes to extend the genetic code without the need for laborious nsAA-AARS-tRNA cognate pair development⁶¹.

V. Genetic code expansion in eukaryotic systems

Amber suppression has been adapted to eukaryotic systems as a tool for the interrogation of cellular biology. Engineered orthogonal AARS/tRNA pairs have been used to genetically encode nsAAs that modulate posttranslational modifications by photocaging lysine⁶² and serine⁶³ residues, promote chromatin condensation with crosslinking residues⁶⁴, introduce chemical handles for fluorescent protein labeling and live cell imaging⁶⁵, among many other applications. Most often, OTS components are scaffolded on *M. mazei* PylRS/tRNA_{Pyl}, *E. coli* TyrRS/tRNA_{Tyr} or *E. coli* LeuRS/tRNA_{Leu} pairs, engineered in *E. coli* or *S. cerevisiae*²¹, and then ported to mammalian vectors.

In addition to unicellular organisms and tissue culture, amber suppression with engineered PylRS/tRNA_{Pyl} derivatives has been demonstrated in the nematode *Caenorhabditis elegans*⁶⁶. Similar work has been successful in the fly *Drosophila melanogaster*, where nsAAs could be incorporated both site-specifically and tissue-specifically into proteins⁶⁷.

As with prokaryotic systems, competition with release factors remains a major barrier for efficient amber suppression in eukaryotes. Though orthogonal ribosomes or genome-wide codon reassignment have not yet been demonstrated in these systems, efforts are being made in this direction. Recently, a *de novo* synthesis of *S. cerevisiae* chromosome III included TAG->TAA stop codon reassignments⁶⁸. As this effort is extended to the remaining chromosomes, the resulting synthetic strain will lack native TAG codons, motivating the need to engineer the specificity of the single eukaryotic release factor to exclude amber recognition.

An additional challenge is native quality control machinery. Nonsense mediated decay (NMD) surveys transcripts for nonsense codons excessively distal from the 3' end, triggering degradation in response to their presence⁶⁹. Addressing this limitation, the pathway has been knocked out in *C. elegans* to boost amber suppression efficiency⁶⁶.

A final consideration in these systems is the expression of the orthogonal tRNA, as eukaryotic RNA polymerase III is typically recruited to intragenic A and B-Box

promoter sequences. One solution has been to identify orthogonal AARS/tRNA pairs containing A and B-like elements. For instance, *E. coli*'s TyrRS/tRNA_{Tyr} is orthogonal in *S. cerevisiae* and contains its consensus A and B-box²¹. However, OTSs scaffolded on PylRS/tRNA_{Pyl}, lack such intragenic sequences. Instead these tRNAs have been successfully expressed by co-opting dicistronic tRNA scaffolds such as tRNA^{ARG}_{UCU}⁷⁰, or by using the extragenic U6 RNA polymerase III promoter⁷¹.

VI. Current challenges and future outlook

Looking forward, three major challenges define the trajectory of this field: the development of more efficient OTSs, accession of more open coding channels to enable multi-site incorporation of multiple nsAAs, and the generation of OTSs engineered for more exotic nsAAs.

Engineered OTSs suffer poor enzymatic efficiencies relative to native translational machinery²⁵. Overall, this represents a problem of insufficient AARS evolution, as typically only 6-8 residues proximal to the amino acid binding pocket are mutagenized. Further diversification is limited by library size constraints – the randomization of 6 residues results in a library of nearly 10⁸ members, quickly saturating standard screening techniques. The use of computational protein modeling may enable more rational engineering, guided by *in silico* binding predictions⁷². Further, during evolution OTSs are screened for the ability to suppress just a few codons in selectable markers, whereas native translational machinery faces a load of thousands of codons. Increasing the load on OTSs may promote the selection of more catalytically active enzymes. Amber suppression of native essential genes has proven to be an effective strategy to tie strain viability to nsAA incorporation^{39,40}, offering a potential route to drive protein evolution under more stringent selective conditions.

Multi-site incorporation of a single nsAA has remained elusive due to the use of codons that have their cognate translation components (e.g. RF1) out-competing the incorporation of the nsAA, as well as the impaired activity of OTS. These limitations are compounded in efforts to achieve multi-site incorporation of *multiple distinct* nsAAs. This pursuit will require advances in our ability to suppress multiple codons simultaneously, and in the development of mutually-orthogonal OTSs. *In vivo*, the field is currently limited to suppressing two distinct codons simultaneously. The use of nonstandard nucleotide bases⁷³ may enable crossing this barrier in the near future. Alternatively, radical recoding strategies that extend off the GRO lacking RF1 may provide a route to more codons. Even with more codons, new advances in orthogonal OTSs are needed. OTSs have been selected against interactions with *native* translational machinery, but not necessarily against other OTSs. This raises the potential for OTS cross-reactions when expressed simultaneously. Work is being done to develop orthogonal tRNA acceptor stems⁷⁴ and additional AARS/tRNA scaffolds to mitigate this cross-orthogonality.

Finally, certain biological constraints limit the scope of nsAA diversity. These include limitations on cell membrane permeability and steric incompatibility with the ribosome and other translational components such as elongation factors. Further

engineering these components with cell-free systems or synthetic ribosomes is necessary for increasing the available nsAA chemical space.

Overcoming some of these technological barriers will better enable the creative potential of nonstandard protein development in producing the next wave of highly functionalized biomaterials with broad applications in medicine, materials science and biotechnology.

Acknowledgements

MCJ gratefully acknowledges the National Science Foundation (MCB-0943393), the Office of Naval Research (N00014-11-1-0363), the DARPA YFA Program (N66001-11-1-4137), the Army Research Office (W911NF- 11-1-0445), the NSF Materials Network Grant (DMR - 1108350), the DARPA Living Foundries Program (N66001-12-C-4211), the David and Lucile Packard Foundation (2011-37152), ARPA-E (DE-AR0000435), and the Chicago Biomedical Consortium with support from the Searle Funds at the Chicago Community Trust for support. BJD is an NSF Graduate Fellow. FJI gratefully acknowledges support from the Defense Advanced Research Projects Agency (N66001-12-C-4020, N66001-12-C-4211); Department of Energy (152339.5055249.100); Gen9, Inc., DuPont Inc. and the Arnold and Mabel Beckman Foundation.

References

- 1 van Hest, J. C. M., Kiick, K. L. & Tirrell, D. A. Efficient incorporation of unsaturated methionine analogues into proteins in vivo. *J Am Chem Soc* **122**, 1282-1288, doi:Doi 10.1021/Ja992749j (2000).
- 2 Ngo, J. T. & Tirrell, D. A. Noncanonical amino acids in the interrogation of cellular protein synthesis. *Accounts of chemical research* **44**, 677-685, doi:10.1021/ar200144y (2011).
- 3 Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annual review of biochemistry* **83**, 379-408, doi:10.1146/annurev-biochem-060713-035737 (2014).
- 4 Dumas, A. e., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment – Expanding the chemistry in biology. *Chemical Science* **6**, 50-69 (2014).
- 5 Chin, J. W. *et al.* Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. *J Am Chem Soc* **124**, 9026-9027, doi:Doi 10.1021/Ja027007w (2002).
- 6 Yang, H., Srivastava, P., Zhang, C. & Lewis, J. C. A General Method for Artificial Metalloenzyme Formation through Strain-Promoted Azide-Alkyne Cycloaddition. *Chembiochem* **15**, 223-227, doi:DOI 10.1002/cbic.201300661 (2014).
- 7 Cho, H. *et al.* Optimized clinical performance of growth hormone with an expanded genetic code. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 9060-9065, doi:10.1073/pnas.1100387108 (2011).
- 8 Tada, S. *et al.* Genetic PEGylation. *PloS one* **7**, e49235, doi:10.1371/journal.pone.0049235 (2012).
- 9 Park, H. S. *et al.* Expanding the genetic code of Escherichia coli with phosphoserine. *Science* **333**, 1151-1154, doi:10.1126/science.1207203 (2011).
- 10 Lee, S. *et al.* A facile strategy for selective incorporation of phosphoserine into histones. *Angewandte Chemie* **52**, 5771-5775, doi:10.1002/anie.201300531 (2013).
- 11 Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nature chemical biology* **4**, 232-234, doi:10.1038/nchembio.73 (2008).
- 12 Ai, H. W., Lee, J. W. & Schultz, P. G. A method to site-specifically introduce methyllysine into proteins in E. coli. *Chemical communications* **46**, 5506-5508, doi:10.1039/c0cc00108b (2010).
- 13 Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500, doi:10.1126/science.1060077 (2001).
- 14 Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. *Annual review of biophysics and biomolecular structure* **35**, 225-249, doi:10.1146/annurev.biophys.35.101105.121507 (2006).
- 15 Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nature biotechnology* **20**, 1044-1048, doi:10.1038/nbt742 (2002).
- 16 Nozawa, K. *et al.* Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163-1167, doi:10.1038/nature07611 (2009).
- 17 Fekner, T. & Chan, M. K. The pyrrolysine translational machinery as a genetic-code expansion tool. *Current opinion in chemical biology* **15**, 387-391, doi:10.1016/j.cbpa.2011.03.007 (2011).
- 18 Wan, W., Tharp, J. M. & Liu, W. R. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochimica et biophysica acta* **1844**, 1059-1070, doi:10.1016/j.bbapap.2014.03.002 (2014).
- 19 Li, W. T. *et al.* Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs. *Journal of molecular biology* **385**, 1156-1164, doi:10.1016/j.jmb.2008.11.032 (2009).

- 20 Ellefson, J. W. *et al.* Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nature biotechnology* **32**, 97-101, doi:10.1038/nbt.2714 (2014).
- 21 Chin, J. W. *et al.* An expanded eukaryotic genetic code. *Science* **301**, 964-967, doi:10.1126/science.1084772 (2003).
- 22 Ai, H. W., Shen, W., Brustad, E. & Schultz, P. G. Genetically encoded alkenes in yeast. *Angewandte Chemie* **49**, 935-937, doi:10.1002/anie.200905590 (2010).
- 23 Zhang, Y., Wang, L., Schultz, P. G. & Wilson, I. A. Crystal structures of apo wild-type M. jannaschii tyrosyl-tRNA synthetase (TyrRS) and an engineered TyrRS specific for O-methyl-L-tyrosine. *Protein science : a publication of the Protein Society* **14**, 1340-1349, doi:10.1110/ps.041239305 (2005).
- 24 Haruna, K., Alkazemi, M. H., Liu, Y., Soll, D. & Englert, M. Engineering the elongation factor Tu for efficient selenoprotein synthesis. *Nucleic acids research* **42**, 9976-9983, doi:10.1093/nar/gku691 (2014).
- 25 O'Donoghue, P., Ling, J., Wang, Y. S. & Soll, D. Upgrading protein synthesis for synthetic biology. *Nature chemical biology* **9**, 594-598, doi:10.1038/nchembio.1339 (2013).
- 26 Rackham, O. & Chin, J. W. A network of orthogonal ribosome x mRNA pairs. *Nature chemical biology* **1**, 159-166, doi:10.1038/nchembio719 (2005).
- 27 Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature biotechnology* **25**, 770-777, doi:10.1038/nbt1314 (2007).
- 28 Jewett, M. C., Fritz, B. R., Timmerman, L. E. & Church, G. M. In vitro integration of ribosomal RNA synthesis, ribosome assembly, and translation. *Molecular systems biology* **9**, 678, doi:10.1038/msb.2013.31 (2013).
- 29 Fritz, B. R. & Jewett, M. C. The impact of transcriptional tuning on in vitro integrated rRNA transcription and ribosome construction. *Nucleic acids research* **42**, 6774-6785, doi:10.1093/nar/gku307 (2014).
- 30 Liu, Y., Fritz, B. R., Anderson, M. J., Schoborg, J. A. & Jewett, M. C. Characterizing and Alleviating Substrate Limitations for Improved in vitro Ribosome Construction. *ACS synthetic biology*, doi:10.1021/sb5002467 (2014).
- 31 Hong, S. H., Kwon, Y. C. & Jewett, M. C. Non-standard amino acid incorporation into proteins using Escherichia coli cell-free protein synthesis. *Frontiers in chemistry* **2**, 34, doi:10.3389/fchem.2014.00034 (2014).
- 32 Hong, S. H. *et al.* Cell-free protein synthesis from a release factor 1 deficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation. *ACS synthetic biology* **3**, 398-409, doi:10.1021/sb400140t (2014).
- 33 Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342**, 357-360, doi:10.1126/science.1241459 (2013).
- 34 Johnson, D. B. *et al.* RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nature chemical biology* **7**, 779-786, doi:10.1038/nchembio.657 (2011).
- 35 Mukai, T. *et al.* Codon reassignment in the Escherichia coli genetic code. *Nucleic acids research* **38**, 8188-8195, doi:10.1093/nar/gkq707 (2010).
- 36 Heinemann, I. U. *et al.* Enhanced phosphoserine insertion during Escherichia coli protein synthesis via partial UAG codon reassignment and release factor 1 deletion. *FEBS letters* **586**, 3716-3722, doi:10.1016/j.febslet.2012.08.031 (2012).
- 37 Hong, S. H. *et al.* Improving Cell-Free Protein Synthesis through Genome Engineering of Escherichia coli Lacking Release Factor 1. *Chembiochem*, doi:10.1002/cbic.201402708 (2015).

- 38 Isaacs, F. J. *et al.* Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348-353, doi:10.1126/science.1205822 (2011).
- 39 Rovner, A. J. *et al.* Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**, 89-93, doi:10.1038/nature14095 (2015).
- 40 Mandell, D. J. *et al.* Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55-60, doi:10.1038/nature14121 (2015).
- 41 Wan, W. *et al.* A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. *Angewandte Chemie* **49**, 3211-3214, doi:10.1002/anie.201000465 (2010).
- 42 Odoi, K. A., Huang, Y., Rezenom, Y. H. & Liu, W. R. Nonsense and sense suppression abilities of original and derivative *Methanosarcina mazei* pyrrolysyl-tRNA synthetase-tRNA(Pyl) pairs in the *Escherichia coli* BL21(DE3) cell strain. *PLoS one* **8**, e57035, doi:10.1371/journal.pone.0057035 (2013).
- 43 Anderson, J. C. *et al.* An expanded genetic code with a functional quadruplet codon. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 7566-7571, doi:10.1073/pnas.0401517101 (2004).
- 44 Chatterjee, A., Lajoie, M. J., Xiao, H., Church, G. M. & Schultz, P. G. A bacterial strain with a unique quadruplet codon specifying non-native amino acids. *Chembiochem : a European journal of chemical biology* **15**, 1782-1786, doi:10.1002/cbic.201402104 (2014).
- 45 Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441-444, doi:10.1038/nature08817 (2010).
- 46 Wang, K. *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nature chemistry* **6**, 393-403, doi:10.1038/nchem.1919 (2014).
- 47 Sachdeva, A., Wang, K., Elliott, T. & Chin, J. W. Concerted, rapid, quantitative, and site-specific dual labeling of proteins. *J Am Chem Soc* **136**, 7785-7788, doi:10.1021/ja4129789 (2014).
- 48 Jewett, M. C., Calhoun, K. A., Voloshin, A., Wu, J. J. & Swartz, J. R. An integrated cell-free metabolic platform for protein production and synthetic biology. *Molecular systems biology* **4**, 220, doi:10.1038/msb.2008.57 (2008).
- 49 Caschera, F. & Noireaux, V. Synthesis of 2.3 mg/ml of protein with an all *Escherichia coli* cell-free transcription-translation system. *Biochimie* **99**, 162-168, doi:10.1016/j.biochi.2013.11.025 (2014).
- 50 Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: applications come of age. *Biotechnology advances* **30**, 1185-1194, doi:10.1016/j.biotechadv.2011.09.016 (2012).
- 51 Bundy, B. C. & Swartz, J. R. Site-specific incorporation of p-propargyloxyphenylalanine in a cell-free environment for direct protein-protein click conjugation. *Bioconjugate chemistry* **21**, 255-263, doi:10.1021/bc9002844 (2010).
- 52 Nehring, S., Budisa, N. & Wiltschi, B. Performance analysis of orthogonal pairs designed for an expanded eukaryotic genetic code. *PLoS one* **7**, e31992, doi:10.1371/journal.pone.0031992 (2012).
- 53 Loscha, K. V. *et al.* Multiple-site labeling of proteins with unnatural amino acids. *Angewandte Chemie* **51**, 2243-2246, doi:10.1002/anie.201108275 (2012).
- 54 Gerrits, M. *et al.* in *Cell-Free Protein Expression* (eds W. Kudlicki, F. Katzen, & R. Benner) 166-180 (Landes Bioscience, 2007).
- 55 Goerke, A. R. & Swartz, J. R. High-level cell-free synthesis yields of proteins containing site-specific non-natural amino acids. *Biotechnology and bioengineering* **102**, 400-416, doi:10.1002/bit.22070 (2009).

- 56 Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nature biotechnology* **19**, 751-755, doi:10.1038/90802 (2001).
- 57 Ahn, J.-H. *et al.* Preparation Method for *Escherichia coli* S30 Extracts Completely Dependent upon tRNA Addition to Catalyze Cell-free Protein Synthesis. *Biotechnology and Bioprocess Engineering* **11**, 420-424 (2006).
- 58 Forster, A. C. *et al.* Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 6353-6357, doi:10.1073/pnas.1132122100 (2003).
- 59 Cui, Z., Stein, V., Tnimov, Z., Mureev, S. & Alexandrov, K. Semisynthetic tRNA Complement Mediates in Vitro Protein Synthesis. *Journal of the American Chemical Society* **137**, 4404-4413, doi:10.1021/ja5131963 (2015).
- 60 Terasaka, N., Hayashi, G., Katoh, T. & Suga, H. An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center. *Nature chemical biology* **10**, 555-557, doi:10.1038/nchembio.1549 (2014).
- 61 Ohuchi, M., Murakami, H. & Suga, H. The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Current opinion in chemical biology* **11**, 537-542, doi:10.1016/j.cbpa.2007.08.011 (2007).
- 62 Gautier, A., Deiters, A. & Chin, J. W. Light-activated kinases enable temporal dissection of signaling networks in living cells. *J Am Chem Soc* **133**, 2124-2127, doi:10.1021/ja1109979 (2011).
- 63 Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nature chemical biology* **3**, 769-772, doi:10.1038/nchembio.2007.44 (2007).
- 64 Wilkins, B. J. *et al.* A cascade of histone modifications induces chromatin condensation in mitosis. *Science* **343**, 77-80, doi:10.1126/science.1244508 (2014).
- 65 Uttamapinant, C. *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, doi:10.1021/ja512838z (2015).
- 66 Greiss, S. & Chin, J. W. Expanding the genetic code of an animal. *J Am Chem Soc* **133**, 14196-14199, doi:10.1021/ja2054034 (2011).
- 67 Bianco, A., Townsley, F. M., Greiss, S., Lang, K. & Chin, J. W. Expanding the genetic code of *Drosophila melanogaster*. *Nature chemical biology* **8**, 748-750, doi:10.1038/nchembio.1043 (2012).
- 68 Annaluru, N. *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55-58, doi:10.1126/science.1249252 (2014).
- 69 Longman, D., Plasterk, R. H., Johnstone, I. L. & Caceres, J. F. Mechanistic insights and identification of two novel factors in the *C. elegans* NMD pathway. *Genes & development* **21**, 1075-1085, doi:10.1101/gad.417707 (2007).
- 70 Hancock, S. M., Uprety, R., Deiters, A. & Chin, J. W. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. *J Am Chem Soc* **132**, 14819-14824, doi:10.1021/ja104609m (2010).
- 71 Ma, H., Zhang, J. & Wu, H. Designing Ago2-specific siRNA/shRNA to Avoid Competition with Endogenous miRNAs. *Molecular therapy. Nucleic acids* **3**, e176, doi:10.1038/mtna.2014.27 (2014).
- 72 Meiler J., B. D. ROSETTALIGAND: Protein–small molecule docking with full side-chain flexibility. *Proteins*, doi:10.1002/prot.21086 (2006).
- 73 Malyshev D., D. K., Lavergne T., Chen T., Dai N., Foster J., Corrêa I., Romesberg F. A semi-synthetic organism with an expanded genetic alphabet. *Nature*, doi:10.1038/nature13314 (2014).

- 74 Neumann H. , S. A. L., Chin J.W. De Novo Generation of Mutually Orthogonal Aminoacyl-tRNA Synthetase/tRNA Pairs. *J. Am. Chem. Soc.*, doi:10.1021/ja9068722 (2010).

- of special interest
- of outstanding interest

- Annaluru, N. *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55-58, doi:10.1126/science.1249252 (2014).

In an impressive triumph of DNA synthesis and assembly, these authors build a functional synthetic yeast chromosome. Importantly for nsAA incorporation efforts, the synthetic chromosome has undergone global UAG/UAA reassignment. This work thus lays the foundation for a recoded yeast strain, which has the potential to be as transformative in the area of eukaryotic amber suppression as the recoded E coli strain has been in prokaryotic amber suppression.

- Bundy, B. C. & Swartz, J. R. Site-specific incorporation of p-propargyloxyphenylalanine in a cell-free environment for direct protein-protein click conjugation. *Bioconjugate chemistry* **21**, 255-263, doi:10.1021/bc9002844 (2010).

This work highlights one of the unique advantages of using cell-free systems for nsAA incorporation. The authors exploit the barrier-free nature of these systems to improve incorporation of a nsAA that suffers from low bioavailability in vivo due to its poor water solubility, surpassing previous in vivo yields by 27-fold.

- Greiss, S. & Chin, J. W. Expanding the genetic code of an animal. *Journal of the American Chemical Society* **133**, 14196-14199, doi:10.1021/ja2054034 (2011).

With success in both prokaryotes and unicellular eukaryotes, nsAA incorporation in multicellular organisms was a logical next step. This paper is the first reported instance of nsAA incorporation via amber suppression in an animal, setting the stage for the development of molecular tools for the study of complex processes in whole organisms.

- Hong, S. H. *et al.* Improving Cell-Free Protein Synthesis through Genome Engineering of Escherichia coli Lacking Release Factor 1. *Chembiochem : a European journal of chemical biology*, doi:10.1002/cbic.201402708 (2015).

Building on earlier success in nsAA incorporation using cell-free systems, these authors rationally modify a partially-recoded E. coli strain to improve productivity in vitro via the knockout of putative negative effectors of cell-free protein synthesis. Their improved strain was able to synthesize 550 mg/L of nsAA-containing protein in batch mode, a significant improvement over the previous state-of-the-art.

- Lajoie, M. J. et al. Genomically recoded organisms expand biological functions. *Science* **342**, 357-360, doi:10.1126/science.1241459 (2013).

Through painstaking genomic engineering, these authors replaced every endogenous instance of the amber (UAG) stop codon in the genome of E. coli MG1655 with the synonymous ochre (UAA) stop codon and deleted prfA (RF-1) from the host genome. The amber codon is orthogonal to the remaining translational machinery in this strain, which promises to be transformative in the area of nsAA incorporation via amber suppression.

- Mandell, D. J. et al. Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518**, 55-60, doi:10.1038/nature14121 (2015).

Guided by computational molecular modeling, these authors redesigned essential enzymes in a genomically-recoded organism to require an nsAA for proper translation, folding, and function. The resulting genetically modified organisms require the nsAA in order to proliferate, and demonstrate unprecedented robustness against biocontainment escape mechanisms.

- Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441-444, doi:10.1038/nature08817 (2010).

Accessing additional codons for the incorporation of multiple distinct nsAAs into a single peptide is one of the great challenges still facing the field of nsAA incorporation. This work provides a brilliant approach, tapping into non-natural quadruplet codons for nsAA encoding.

- Nozawa, K. et al. Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* **457**, 1163-1167, doi:10.1038/nature07611 (2009).

Orthogonal in many prokaryotic and eukaryotic translation systems, the discovery and characterization of the pyrrolysine PylRS/tRNA^{Pyl} pair was a tremendous boon to nsAA incorporation efforts. This work elucidates the molecular underpinnings of the orthogonality of PylRS/tRNA^{Pyl} in its various hosts.

- Park, H. S. et al. Expanding the genetic code of Escherichia coli with phosphoserine. *Science* **333**, 1151-1154, doi:10.1126/science.1207203 (2011).

By adding phosphoserine to the protein engineer's toolkit, this work facilitated study of one of the most important and ubiquitous classes of PTMs in all of protein biology. This work also highlights an important consideration for future OTS development with the first instance of a nsAA-specific engineered elongation factor.

- Rovner, A. J. et al. Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518**, 89-93, doi:10.1038/nature14095 (2015).

By adding orthogonal translation components to the chromosome of a genomically-recoded organism and installing amber codons into the coding regions of 22 essential genes, these authors were able to generate an organism that is entirely dependent on nsAAs in order to live. This has important applications in biocontainment as well as the evolution of more efficient OTS components.

- Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature biotechnology* **25**, 770-777, doi:10.1038/nbt1314 (2007).

This important manuscript demonstrates the feasibility of using orthogonal ribosomes to evolve novel function without impacting host viability, providing critical guidance for potential future projects that attempt to reconcile nsAA incorporation bottlenecks based in ribosomal incompatibility.

Figure 1: Methods for Genetic Code Expansion **a.** Two general paradigms exist for the genetic incorporation of nonstandard amino acids into proteins contrasted with the natural process of encoding the canonical amino acids. The close analog methodology complements a natural amino acid auxotrophy with a close nonstandard analog, enabling global protein labeling by native translational machinery. The orthogonal translation methodology introduces orthogonal translational machinery engineered to charge an orthogonal tRNA with a nonstandard amino acid, enabling site-specific targeted genetic incorporation. Certain nsAAs may require additional mutations in the elongation factor or the ribosome. **b.** For targeted genetic incorporation, amber suppression is the most widely used technique. Competition with release factors limits efficiency, and methods are discussed to overcome this. **c.** Quadruplet suppression can be performed with appreciable efficiency with the use of an engineered orthogonal 16S ribosomal subunit²⁷. **d.** Sense codons can be completely reassigned by using an orthogonal 23S ribosomal subunit, engineered to accept a synthetic set of tRNAs⁶⁰.

Figure 2: General methodology for engineering orthogonal translation systems for novel NSAAs. **a.** Scaffold orthogonal AARS/tRNA pairs are selected from distant organisms. Residues in the amino acid binding pocket of the AARS are randomized. **b.** A negative selection is performed to exclude AARS variants capable of charging natural amino acids. In the absence of the nsAA, clones are selected by the inability to suppress a toxic gene. **c.** A positive selection is performed to retrieve AARS variants capable of charging the nsAA, selecting for the ability to suppress an essential gene, such as an antibiotic resistance marker. **d.** The selected orthogonal translation system consists of the orthogonal AARS, tRNA, and nsAA.

Figure 3: Genomically recoded organisms (GROs), Strain Engineering and *in vitro* nsAA translation. A genomically recoded organism reassigns the TAG stop codon to an open sense coding channel for enhanced nsAA amber suppression. Ribosomal engineering promotes other classes of codon suppression. These engineered strains can be used *in vivo* and, via cell lysate, to develop *in vitro* translation systems with open coding channels for the production of nsAA-containing proteins.

a. Genetic Encoding of NSAAs **b. Amber Suppression** **c. Quadruplet Suppression** **d. Sense Codon Reassignment**

a. Randomization of Amino Acid Binding Pocket

b. Negative Selection to isolate orthogonal aaRS variants

c. Positive Selection to select for efficient aaRS/nsAA pairs

d. Evolved orthogonal aaRS/tRNA

Codon Reassignment

Genomically Recoded Organism: Conversion of TAG to Sense Codon for nsAA incorporation

vs

Natural Translation

Ribosome Refactoring

Orthogonal aaRS/tRNA

Chemically-charged or
self-aminoacylated tRNAs

Cell-Free Protein Synthesis (CFPS)