Visualizing the complexity of proteins in living cells with genetic code expansion

Kanokpol Aphicho†, Narongyot Kittipanukul†, Chayasith Uttamapinant*

School of Biomolecular Science and Engineering, Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong, Thailand

†These authors contributed equally to this work
*Corresponding author: chayasith.u@vistec.ac.th

Abstract

Genetic code expansion has emerged as an enabling tool to provide insight into functions of understudied proteinogenic species such as small proteins and peptides, and to probe protein biophysics in the cellular context. Here we discuss recent technical advances and applications of genetic code expansion in cellular imaging of complex mammalian protein species, along with considerations and challenges upon using the method.

Keyword

Genetic code expansion; unnatural amino acid; fluorescence imaging

Introduction

Increasing demands to probe protein structure, dynamics and interactions in the cellular context have driven the development of new imaging modalities, probes, and tagging tools of cellular proteins. Advances in instrumentation [1-3] and optical probes [4-6] have pushed the spatial resolution of biological imaging down to molecular resolution. The bottleneck of many cellular protein imaging studies now lies in tagging technologies [7], almost all of which employ potentially perturbative protein or peptide appendages. Properties of such tags on proteins become critical when studying: microproteins, peptides, and macromolecular assemblies, where a large tag is not tolerated; intrinsically disordered proteins, where a tag can affect disorder degrees or complex formation; or physical properties of proteins such as folding states, conformational changes, or stoichiometry within complexes, all of which are affected by the tag. Demanding imaging modalities such as single-molecule and super-resolution techniques further require a minimally sized tag—to best capture actual dimensions of biological structures—coupled to bright, photostable organic fluorophores for best image resolution.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

Genetic code expansion (GCE)—in which a single amino acid on a protein is substituted with an unnatural amino acid (UAA) through an orthogonal aminoacyl-tRNA synthetase (aaRS)/ tRNA pair—can be used to site-specifically label proteins with diverse probes including fluorophores, without adding heavy protein/peptide bulk. Continual improvements to the technology [8,9] have resulted in the discovery of hundreds of new amino acid substrates for tRNA synthetases [10]; new codons and optimal cognate tRNAs for multisite incorporation of UAAs [11,12]; and maximal protein production through RNA optimizations [11] and engineering of cellular machineries [13,14]. These efforts enable construction of engineered cells capable of producing unnatural proteins for diverse *in vitro* uses, with expanded functionalities to interrogate biological processes.

Despite its potential, GCE is used much less frequently as a protein tagging tool for cell biological studies. In this review, we discuss recent technical advances and applications of genetic code expansion to cellular protein studies; analyze fundamental and technical challenges in expanding the utility of genetic code expansion to mammalian cell biological studies; and envision further improvements to the technology and applications.

Components of genetic code expansion for protein labeling

GCE for protein labeling consists of the following minimal exogenous parts: the UAA of choice; the aaRS specific for the UAA, and its cognate tRNA; and the gene of interest bearing a blank codon (typically not designated for decoding by cellular tRNAs) at a user-defined site, serving as the UAA incorporation position (Figure 1a). The amber stop codon is used as a blank codon in typical GCE experiments. Protein labeling can be performed in one step if the UAA already contains sufficient properties (e.g. fluorescent amino acids [15]) and if aaRS variants have been identified for the UAAs. To access diverse probe/fluorophore structures, a two-step labeling strategy in which the UAA bears a chemically derivatizable functional group can be employed (Figure 1b). In two-step labeling experiments, additional parts to consider are the choice of derivatizable reaction partners (related to the choice of UAA), and the choice of fluorophore.

Beyond exogenous parts, since genetic code expansion tags proteins co-translationally, it is subject to interference from complex endogenous pathways and components associated with protein translation. To produce sensitive and specific protein labeling, exogenous and endogenous factors—related to genetic code expansion and mammalian translation respectively—must be considered.

Genetic code expansion systems for cellular imaging studies

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

In the simplest form, proteins can be labeled through direct incorporation of fluorescent amino acids such as (acetylnaphthalenyl)amino amino propanoic (Anap) [16], coumarin-derived amino acid [17], dansylalanine [18], and acridonylalanine (ACD) [19,20] (Figure 2a). Anap fluorescence is environmentally sensitive and can be used to probe conformational changes of proteins [21]. ACD has superior brightness and photostability among encodable fluorescent UAAs [15] , is suitable for fluorescence lifetime imaging (FLIM) as shown on insulin receptors [19], and has a photoactivatable thioketone variant [22].

In the more versatile two-step labeling, UAAs with a clickable functional handle can be used (Figure 2b). The field has converged on UAA reaction partners for inverse-electron-demand Diels-Alder (DA) cycloaddition [23,24] due to its fast reaction kinetics (up to 10^5 M^{-1} s^{-1} [25])) and biocompatibility. Lysine derivatives bearing *endo* bicyclononyne (BCNK) and axial *trans*-cyclooct-2-ene (TCO^AK) are at present the best DA-UAAs for protein labeling: BCNK reacts ~10-fold more slowly to tetrazines than TCO^AK but the latter suffers from higher instability in cells and unwanted β-elimination which reverts fluorophore conjugation [26]. Both amino acids benefit from fluorogenic coupling to tetrazine-fluorophores and are efficiently incorporated to mammalian proteins via *Methanosarcina* pyrrolysyl-tRNA synthetase (PylRS)/$tRNA_{CUA}$ (PylT) pair (particularly the combination of the Y306A/Y384F double mutant, or $PylRS^{AF}$ [27,28], whose activity can be boosted upon fusion to a nuclear export sequence [29]; and the engineered pyrrolysyl $tRNA^{M15}$ [30]). Recent comparative studies of BCNK, TCO^AK, and other UAAs for mammalian intracellular protein labeling by the Lemke and Hang groups have confirmed superior properties of these two UAAs [31,32]. Reinkemeier and Lemke additionally discovered equatorial *trans*-cyclooct-4-ene-lysine to have excellent properties for protein labeling, and engineered an optimal synthetase for its incorporation [32]. Beyond dienophile UAAs, tetrazine-based UAAs with exceptionally fast kinetics such as Tet-v3.0Bu can be used [25] and coupled to photoactivatable BCN (photo-DMBO [33]) for light-activated labeling. Tet-v3.0Bu has superior stability compared to dienophile UAAs, but its labeling is not fluorogenic and may suffer from lower signal-to-noise from off-target sticking of probes to the cellular interior.

Diverse fluorophores can be linked to tetrazines and fluorogenically labeled on cellular proteins bearing dienophile UAAs, but the degree of fluorescence increase depends on the composite nature of the fluorophore-tetrazine conjugate. Continual developments of long-wavelength dyes for minimally phototoxic imaging and tetrazine-based quenching have yielded increasingly well-performing conjugates, particularly those based on silicon rhodamine (SiR, Figure 2c). Developed by the Johnsson group, SiR is intrinsically fluorogenic due to its preference to adopt a non-fluorescent lactone form prior to conjugation to proteins [6].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

Installation of an internal hydroxymethyl nucleophile modulates the lactone-zwitterionic equilibrium of the SiR scaffold and creates hydroxymethyl silicon rhodamine (HMSiR) [34], a self-blinking dye suitable for super-resolution imaging. While SiR and HMSiR can be coupled to traditional 3-aminobenzyl-tetrazine via an amide linkage [35], the Wombacher group directly grafted the 6-methyl-3-phenyl-tetrazine moiety onto the phenyl ring of HMSiR, increasing tetrazine-based quenching efficiency and produced highly fluorogenic Me-tet-fHMSiR [36]. Further linker optimizations to increase quenching via co-planar stacking and Dexter electron transfer yielded HD65(x) dye series with up to 50-fold fluorescence turn-on [37]. Coupling of GCE labeling with BCNK to self-blinking, highly fluorogenic HDdyes enabled wash-free and super-resolution imaging of cellular proteins under low-laser, minimally toxic imaging conditions. A near-IR variant of HMSiR, called Yale676sb, was also developed [38], allowing multicolor super-resolution imaging in combination with HMSiR, with potential extension to imaging *in vivo*.

Considerations and design tools for GCE in mammalian cells

One roadblock to wider use of GCE for protein labeling is the need to screen multiple UAA incorporation positions (we typically do 3-6 sites to start) to identify an optimal label site for each protein of interest. Such an exercise is necessary not only to maximize protein yields, but to ensure proper protein structure and function. In addition to avoiding potentially functionally and structurally crucial residues as the UAA placement site, we factor in the following considerations. First, we try to select surface-exposed positions of which the native amino acid has similar polarity as the desired UAA, to minimize drastic changes to surface charges. Second, we monitor if there is UAA-independent synthesis of the full-length protein, which indicates interference from near-cognate native tRNAs and incorporation of native amino acids. Such a phenomenon is rare in our experience and is highly sequence context-dependent (context effects of genetic code expansion in bacteria were recently reviewed [39]) (Figure 3). Third, we avoid placing the UAA near the very N-terminus of the protein, as the ribosome is more prone to stalling during translation initiation, affecting protein yield [40]. Fourth, we would watch out for translation re-initiation events [41], in which the mammalian ribosome dissociates upon encountering the placed amber codon and re-assembles at a downstream methionine codon with suitable sequence contexts. Such translation re-initiation generates undesired N-truncated by-products. Interestingly, transcripts with premature stop codons can evade nonsense-mediated mRNA decay and become more stable (and potentially give higher protein yields) if translation can be re-initiated downstream [41]. Lastly, we would avoid placing the UAA near the C-terminus of the protein, as the genetic code translation machinery may not be able to compete with translation termination machineries positing at the 3'-UTR of the mRNA [42]. The resulting near full-length C-truncated by-products can have

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

dominant negative effects [43], especially if the protein of interest is multimeric in its functional state.

Beyond empirical optimizations, recent efforts have been made to computationally identify an optimal UAA placement site and to develop a uniform UAA incorporation site at the N- and C-terminus of proteins. In the former, a linear regression algorithm called 'Identification of Permissive Amber Site for Suppression' (iPASS) was developed by the Lang and Bultmann groups based on proteomic analysis of mammalian translation readthrough events, and can be used to predict relative UAA incorporation efficiencies of given sequence contexts [44]. In the latter, the Elsässer group created a single-residue tag called 'Single residue Terminal Labeling' (STELLA) via traceless removal of ubiquitin and intein tags placed at the N- and C-terminal of the desired protein, respectively [45]. Such minimal yet generalizable tags were used to image multiple microproteins.

Progress toward efficient multicolor GCE labeling in cells

Discoveries and engineering of mutually orthogonal pairs of aaRSs and cognate tRNAs have enabled encoding of multiple UAAs in a cellular protein of interest. To install multiple probes onto proteins, one also needs multiple blank codons and mutually orthogonal bio-orthogonal reactions of which one reaction partner is encodable as an amino acid by GCE. The exciting challenge for the field is to develop these mutually orthogonal parts—blank codons, aaRS/tRNA pairs, UAAs, and their bio-orthogonal derivatizations—to be as efficient as those of the optimized systems for single-label GCE.

While the opal (TGA) and ochre (TAA) stop codons can be used in conjunction with the amber (TAG) codon for multisite GCE [46-49], there are trade-offs between these two second-choice stop codons. Ochre is used less frequently as a stop sign in the human/mammalian genome compared to opal so one expects lower off-target suppression with ochre, but ochre-targeting tRNAs exhibit cross-reactivity and can suppress amber codons [49]. Beyond canonical stop codons, the suppression efficiencies on quadruplet codons in eukaryotes have been continuously improved [50,51]. In recent work, the Greiss group developed an engineered pylT with optimized quadruplet decoding and functionality in eukaryotes, and demonstrated efficient quadruplet decoding in *Caenorhabditis elegans* [52].

To target two UAAs to two distinct codons on one or more proteins, two mutually orthogonal aaRS/tRNA pairs are required. Several aaRS/tRNA systems are known to be orthogonal with respect to endogenous translation machineries in mammalian cells. Among these systems, the tyrosyl- and leucyl-tRNA synthetase and tRNA systems from *Escherichia coli* were

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

demonstrated to be mutually orthogonal to the PylRS/PylT system from *Methanosarcina* [48,49]. While the two *E. coli* aaRS/tRNA systems currently incorporate a small repertoire of UAAs bearing cell-compatible functional handles and have limited use for live-cell protein imaging, recent advances in creating an *E. coli*-based directed evolution system by the Chatterjee group ([53] and Grasso KT et al., bioRxiv doi:https://doi.org/10.1101/2021.11.28.470256)—in which the native tyrosyl-tRNA synthetase/tRNA pair is replaced with an archaeal counterpart—hold promise to expand the substrate diversity of aaRS/tRNA systems from *E. coli* for GCE labeling. Recently, multiple mutually orthogonal PylRS/PylT systems lacking N-terminal domains (ΔNPylRS/PylT) were discovered and engineered by the Chin and Elsässer groups. ΔNPylRS/PylT systems from *Methanomethylophilus alvus* [54-56] and *Methanogenic archaeon ISO4-G1* [46,54], in particular, are shown to be orthogonal in mammalian cells and mutually orthogonal to *Methanosarcina* PylRS/PylT. Moreover, polyspecific mutations discovered on *Methanosarcina* PylRS can be transferred to ΔNPylRSs, enabling incorporation of bulkier UAAs useful for protein labeling such as BCNK and $TCO^{A}K$. Continual efforts to improve the efficiency of ΔNPylRSs, which has intrinsically weaker affinities for their ΔNPylTs due to the lack of tRNA-binding N-terminal domains, can enable their routine use in multicolor GCE labeling.

Functional handles for diverse bio-orthogonal reactions can be encoded as UAAs for GCE labeling, and mutual orthogonality of some chemistry pairs was demonstrated in multicolor, live-cell labeling on proteins [46,47,57,58]. Multicolor GCE labeling will still benefit from more bio-orthogonal reactions that are fast and function inside living cells, in the similar vein as the inverse-electron-demand Diels-Alder reaction. Exciting efforts are underway to develop mutually orthogonal, kinetically efficient chemistries; these advances are recently reviewed [59,60]. One notable example is the retro-Cope elimination reaction between N,N-dialkylhydroxylamines and cyclooctynes developed by the Kim group [61]. The reaction is fast (rates up to 84 M^{-1} s^{-1}), compatible with intracellular labeling, and orthogonal to the *trans*-cyclooctene/tetrazine Diels-Alder reaction. Beyond reactivity and orthogonality, optimizing properties of UAAs and their click partners so they are well-behaved in cells is of practical importance. For instance, the Kele and Lemke groups created a hydrophilic *trans*-cyclooctene UAA, DOTCO-lysine, which can be washed out quickly from cells, permitting cellular imaging with lower background and potentially allowing monitoring of proteins with fast turnover [62].

Fluorescence imaging applications of GCE in mammalian cells

Recent studies have shown the utility of GCE labeling in elucidating cellular functions of small proteins (typically <100 amino acids in length) (Figure 4a). Hundreds of such microproteins

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

were discovered via genomic tools and detected with proteomic technologies [64,65], yet few were functionally characterized. Many microproteins are intrinsically disordered [66] and elicit their functions through complexation with other proteins and biomolecules, necessitating that their characterization studies be performed non-invasively in cells.

To this end, the Hang group used TCO^{A}K and BODIPY-FL to fluorescently track a 15-kD vesicle-associated protein IFITM3, a protein known to modulate viral and bacterial infection in vertebrates [31,67]. Live-cell tracking of IFITM3 reveals clear interactions of IFITM3-containing vesicles with viral particles, in a virus-specific manner, upon viral entry; and how site-specific S-palmitoylation of IFITM3 affects this interaction and downstream antiviral activity [67]. In another study, the Slavoff lab and our group used BCNK and SiR to label Alt-RPL36, a 16-kD protein translated from an alternative open reading frame of an mRNA encoding a human ribosomal protein RPL36 [68]. Live-cell imaging of Alt-RPL36 showed it partially co-localize to the cytosolic face of the endoplasmic reticulum, consistent with its phosphorylation-dependent interaction with a lipid transporter TMEM24. Beyond these two studies, the aforementioned STELLA tag should be broadly applicable to biological characterizations of microproteins.

Non-dividing cells and tissues such as primary neurons and organotypic brain slices with dense, interesting ultrastructural organizations would benefit from GCE labeling (Figure 4b), but it is challenging to perform GCE, which needs high transgene expression for UAA incorporation, in these delicate settings. The Choquet and Sauer groups used an inducible expression system to temporarily increase PylRS expression while mitigating long-term toxicity, and performed GCE labeling on transmembrane AMPAR regulatory proteins (TARPs), which regulates AMPA receptors (AMPARs), via TCO^{A}K and diverse membrane-impermeable fluorophores [69]. GCE-tagged TARPs allow access to sterically occluded epitopes inaccessible to antibodies as well as FRET-based monitoring of their interactions with AMPARs. Labeling on TARPs can be performed in dissociated hippocampal neurons at sufficiently high labeling density and specificity for super-resolution dSTORM imaging, allowing analysis of the distribution of TARPs in individual dendritic spines. Upon coupling to single-cell electroporation, TARPs in hippocampal slice cultures can be visualized via GCE. Along a similar goal of extending GCE to neuronal studies, the Nikić-Spiegel group optimized GCE labeling of neurofilament light chain (NFL) with TCO^{A}K in mouse cortical neurons, and developed pulse-chase labeling with two fluorophores (BODIPY-FL and SiR) to examine neurofilament growth (Arsić A et al., bioRxiv doi: 10.1101/2021.01.14.426692). In the same study, the amber codon was introduced to the endogenous NFL gene via CRISPR, allowing GCE labeling of NFL produced from its genomic locus in primary neurons. Improvements to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

amber suppression and transfection efficiencies in primary cells will further increase signal obtained from such challenging labeling experiments.

GCE can also be used to image small peptides generated from proteolytic processing, but requires a composite labeling approach in order to distinguish labels on the processed peptides vs those on precursor proteins (Figure 4c). Our group used a molecular beacon labeling strategy, in which a fluorophore is inserted in the peptide-containing segment of the precursor protein via GCE, while a fluorescence quencher is attached to another site within the protein via enzyme-mediated labeling. Fluorescence is effectively quenched while the precursor protein is intact but becomes detectable once the peptide is proteolytically released from its precursor. We used such a strategy to track amyloid-β peptides (labeled with BCNK and cy5 via GCE) as they are processed from amyloid precursor protein (APP, labeled with QSY21 quencher via HaloTag) during endocytosis [70].

The small size and ability to designate the incorporation site of GCE labels lends themselves to biophysical studies that probe conformation changes of proteins *in vitro*, and recently, in cells (Figure 4d). For instance, the fluorescent amino acid Anap displays solvatochromic shifts in polar environments and is a well-established tool for probing protein dynamics. Anap was recently coupled to spectral imaging by the Cornish and Kass groups to monitor dynamics of sodium channels in live HEK cells [71]. The Ashcroft group labeled ATP-sensitive potassium channels with Anap and monitored nucleotide binding to the channel via Förster Resonance Energy Transfer (FRET) with fluorescent nucleotides in unroofed HEK membranes [21]. FRET-based, cell-compatible conformational sensors with other suitable fluorophores were also created. A recent example is the conformational EGFR indicator developed by the Bastiaens and Chin groups [72]. Here, EGFR is doubly labeled at strategic sites with mCitrine FRET donor and ATTO590 FRET acceptor; labeling of the latter was accomplished via BCNK. The sensor, called '(FRET)-based conformational EGFR indicator' (CONEGI), was used in FLIM-FRET experiments in live HEK cells to probe EGF-independent conformation changes and resulting autoactivation of EGFR.

Outlook for improvements to GCE and its applications

GCE labeling has two notable weaknesses, which become prominent when the UAA incorporation to desired proteins is not working well. First, leftover amber-suppressor tRNAs bearing UAAs can be fluorescently derivatized (e.g. with tetrazine probes if BCNK or $TCO^{A}K$ is used), creating off-protein labeling background [73]. Second, an UAA can be incorporated in response to endogenous amber codons [44,74], creating labeling background on other proteins. While optimizations can minimize these sources of background, future work should

aim to eradicate them. Background from aminoacylated tRNAs (i.e. pyrrolysyl tRNA) likely stems from their tight association with cognate aaRS's (i.e. PylRS), which also limits the enzyme turnover. Enzyme engineering of aaRS's to improve turnover should simultaneously improve GCE efficiency and reduce tRNA-dependent labeling background. Approaches to enhance translation of selective mRNAs, such as spatial confinement of desired mRNAs and GCE machineries through liquid-liquid phase separation [75,76], are exciting prospects to concurrently increase GCE efficiency while reducing translation readthrough on endogenous mRNAs.

Continual improvements to the signal-to-noise of labeling by GCE may allow access to imaging experiments impossible with other tagging techniques. Single-molecule FRET studies to probe structural dynamics of cellular proteins [77] would benefit from non-invasive, site-specific tagging of multiple bright, photostable fluorophores offered by mutually orthogonal and highly active GCE machineries [46]. In combination with genome editing tools such as CRISPR to introduce a suitable blank triplet or quadruplet codon at desired sites within the genome, GCE with efficiency matching natural translation may permit sensitive, non-invasive labeling of proteins expressed from endogenous genomic loci, under native regulatory mechanisms.

Conflict of interest statement

The authors declare no competing interests.

Acknowledgements

The authors were supported by a start-up grant from VISTEC, Wellcome Trust [217249/Z/19/Z], Swiss National Science Foundation [IZSTZ0_193915], and Thailand Science Research and Innovation (Global Partnership Program from Program Management Unit-B), all to C.U. K.A. and N.K. were supported by research assistant and student funds from VISTEC respectively. The authors thank Prof. Robert Robinson (Okayama University, Japan) for helpful feedback on the manuscript.

References and recommended reading

Papers of particular interest, published within the period of review (2019-2021), have been highlighted as:

* of special interest

** of outstanding interest

1. Bates M, Huang B, Dempsey GT, Zhuang X: **Multicolor super-resolution imaging with photo-switchable fluorescent probes**. *Science* 2007, **317**:1749-1753. https://doi.org/10.1126/science.1146598
2. Willig KI, Rizzoli SO, Westphal V, Jahn R, Hell SW: **STED microscopy reveals that synaptotagmin remains clustered after synaptic vesicle exocytosis**. *Nat* 2006, **440**:935-939. https://doi.org/10.1038/nature04592
3. Shroff H, Galbraith CG, Galbraith JA, Betzig E: **Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics**. *Nat Methods* 2008, **5**:417-423. https://doi.org/10.1038/nmeth.1202
4. Grimm JB, English BP, Chen J, Slaughter JP, Zhang Z, Revyakin A, Patel R, Macklin JJ, Normanno D, Singer RH, et al.: **A general method to improve fluorophores for live-cell and single-molecule microscopy**. *Nat Methods* 2015, **12**:244-250, 243 p following 250. https://doi.org/10.1038/nmeth.3256
5. Grimm JB, Tkachuk AN, Xie L, Choi H, Mohar B, Falco N, Schaefer K, Patel R, Zheng Q, Liu Z, et al.: **A general method to optimize and functionalize red-shifted rhodamine dyes**. *Nat Methods* 2020, **17**:815-821. https://doi.org/10.1038/s41592-020-0909-6
6. Lukinavicius G, Umezawa K, Olivier N, Honigmann A, Yang G, Plass T, Mueller V, Reymond L, Correa IR, Jr., Luo ZG, et al.: **A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins**. *Nat Chem* 2013, **5**:132-139. https://doi.org/10.1038/nchem.1546
7. Zhang G, Zheng S, Liu H, Chen PR: **Illuminating biological processes through site-specific protein labeling**. *Chem Soc Rev* 2015, **44**:3405-3417. https://doi.org/10.1039/c4cs00393d
8. de la Torre D, Chin JW: **Reprogramming the genetic code**. *Nat Rev Genet* 2021, **22**:169-184. https://doi.org/10.1038/s41576-020-00307-7
9. Shandell MA, Tan Z, Cornish VW: **Genetic Code Expansion: A Brief History and Perspective**. *Biochemistry* 2021, **60**:3455-3469. https://doi.org/10.1021/acs.biochem.1c00286
10. Young DD, Schultz PG: **Playing with the Molecules of Life**. *ACS Chem Biol* 2018, **13**:854-870. https://doi.org/10.1021/acschembio.7b00974
11. Dunkelmann DL, Oehm SB, Beattie AT, Chin JW: **A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design**. *Nat Chem* 2021, **13**:1110-1117. https://doi.org/10.1038/s41557-021-00764-5
12. Dunkelmann DL, Willis JCW, Beattie AT, Chin JW: **Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids**. *Nat Chem* 2020, **12**:535-544. https://doi.org/10.1038/s41557-020-0472-x
13. Schmied WH, Elsasser SJ, Uttamapinant C, Chin JW: **Efficient multisite unnatural amino acid incorporation in mammalian cells via optimized pyrrolysyl tRNA synthetase/tRNA expression and engineered eRF1**. *J Am Chem Soc* 2014, **136**:15577-15583. https://doi.org/10.1021/ja5069728
14. Bianco A, Townsley FM, Greiss S, Lang K, Chin JW: **Expanding the genetic code of Drosophila melanogaster**. *Nat Chem Biol* 2012, **8**:748-750. https://doi.org/10.1038/nchembio.1043
15. Cheng Z, Kuru E, Sachdeva A, Vendrell M: **Fluorescent amino acids as versatile building blocks for chemical biology**. *Nat Rev Chem* 2020, **4**:275-290. https://doi.org/10.1038/s41570-020-0186-z
16. Puljung MC: **ANAP: A versatile, fluorescent probe of ion channel gating and regulation**. *Methods Enzymol* 2021, **654**:49-84. https://doi.org/10.1016/bs.mie.2021.01.048
17. Wang J, Xie J, Schultz PG: **A genetically encoded fluorescent amino acid**. *J Am Chem Soc* 2006, **128**:8738-8739. https://doi.org/10.1021/ja062666k

18. Summerer D, Chen S, Wu N, Deiters A, Chin JW, Schultz PG: **A genetically encoded fluorescent amino acid**. *Proc Natl Acad Sci U S A* 2006, **103**:9785-9789. https://doi.org/10.1073/pnas.0603965103
19. Jones CM, Robkis DM, Blizzard RJ, Munari M, Venkatesh Y, Mihaila TS, Eddins AJ, Mehl RA, Zagotta WN, Gordon SE, et al.: **Genetic encoding of a highly photostable, long lifetime fluorescent amino acid for imaging in mammalian cells**. *Chem Sci* 2021, **12**:11955-11964. https://doi.org/10.1039/d1sc01914g
20. Hostetler ZM, Cory MB, Jones CM, Petersson EJ, Kohli RM: **The Kinetic and Molecular Basis for the Interaction of LexA and Activated RecA Revealed by a Fluorescent Amino Acid Probe**. *ACS Chem Biol* 2020, **15**:1127-1133. https://doi.org/10.1021/acschembio.9b00886
21. Puljung M, Vedovato N, Usher S, Ashcroft F: **Activation mechanism of ATP-sensitive K(+) channels explored with real-time nucleotide binding**. *Elife* 2019, **8**. https://doi.org/10.7554/eLife.41103
22. Tang J, Yu C, Loredo A, Chen Y, Xiao H: **Site-Specific Incorporation of a Photoactivatable Fluorescent Amino Acid**. *ChemBioChem* 2021, **22**:501-504. https://doi.org/10.1002/cbic.202000602
23. Lang K, Davis L, Wallace S, Mahesh M, Cox DJ, Blackman ML, Fox JM, Chin JW: **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* 2012, **134**:10317-10320. https://doi.org/10.1021/ja302832g
24. Blackman ML, Royzen M, Fox JM: **Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity**. *J Am Chem Soc* 2008, **130**:13518-13519. https://doi.org/10.1021/ja8053805
25. Jang HS, Jana S, Blizzard RJ, Meeuwsen JC, Mehl RA: **Access to Faster Eukaryotic Cell Labeling with Encoded Tetrazine Amino Acids**. *J Am Chem Soc* 2020, **142**:7245-7249. https://doi.org/10.1021/jacs.9b11520
26. Carlson JCT, Mikula H, Weissleder R: **Unraveling Tetrazine-Triggered Bioorthogonal Elimination Enables Chemical Tools for Ultrafast Release and Universal Cleavage**. *J Am Chem Soc* 2018, **140**:3603-3612. https://doi.org/10.1021/jacs.7b11217
27. Borrmann A, Milles S, Plass T, Dommerholt J, Verkade JM, Wiessler M, Schultz C, van Hest JC, van Delft FL, Lemke EA: **Genetic encoding of a bicyclo[6.1.0]nonyne-charged amino acid enables fast cellular protein imaging by metal-free ligation**. *ChemBioChem* 2012, **13**:2094-2099. https://doi.org/10.1002/cbic.201200407
28. Yanagisawa T, Kuratani M, Seki E, Hino N, Sakamoto K, Yokoyama S: **Structural Basis for Genetic-Code Expansion with Bulky Lysine Derivatives by an Engineered Pyrrolysyl-tRNA Synthetase. *Cell Chem Biol* 2019, **26**:936-949.e13. https://doi.org/10.1016/j.chembiol.2019.03.008

This work provided structural basis for the polyspecificity of *Methanosarcina Mazei* pyrrolysyl-RNA synthetase bearing Y306A/Y384F mutations, which is used ubiquitously for incorporation of amino acids useful for protein labeling, including BCNK and TCOAK.

29. Nikic I, Estrada Girona G, Kang JH, Paci G, Mikhaleva S, Koehler C, Shymanska NV, Ventura Santos C, Spitz D, Lemke EA: **Debugging Eukaryotic Genetic Code Expansion for Site-Specific Click-PAINT Super-Resolution Microscopy**. *Angew Chem Int Ed Engl* 2016, **55**:16172-16176. https://doi.org/10.1002/anie.201608284
30. Serfling R, Lorenz C, Etzel M, Schicht G, Bottke T, Morl M, Coin I: **Designer tRNAs for efficient incorporation of non-canonical amino acids by the pyrrolysine system in mammalian cells**. *Nucleic Acids Res* 2018, **46**:1-10. https://doi.org/10.1093/nar/gkx1156

31. Peng T, Hang HC: **Site-Specific Bioorthogonal Labeling for Fluorescence Imaging of Intracellular Proteins in Living Cells**. *J Am Chem Soc* 2016, **138**:14423-14433. https://doi.org/10.1021/jacs.6b08733

*32. Reinkemeier CD, Koehler C, Sauter PF, Shymanska NV, Echalier C, Rutkowska A, Will DW, Schultz C, Lemke EA: **Synthesis and Evaluation of Novel Ring-Strained Noncanonical Amino Acids for Residue-Specific Bioorthogonal Reactions in Living Cells**. *Chemistry* 2021, **27**:6094-6099. https://doi.org/10.1002/chem.202100322

This paper describes the syntheses and evaluation of new trans-cyclooctene structures whose educts with tetrazines should be more stable in cells. New structures were carefully benchmarked against established Diels-Alder unnatural amino acids, using flow cytometry-based FRET measurements

33. Mayer SV, Murnauer A, von Wrisberg MK, Jokisch ML, Lang K: **Photo-induced and Rapid Labeling of Tetrazine-Bearing Proteins via Cyclopropenone-Caged Bicyclononynes**. *Angew Chem Int Ed Engl* 2019, **58**:15876-15882. https://doi.org/10.1002/anie.201908209

34. Uno SN, Kamiya M, Yoshihara T, Sugawara K, Okabe K, Tarhan MC, Fujita H, Funatsu T, Okada Y, Tobita S, et al.: **A spontaneously blinking fluorophore based on intramolecular spirocyclization for live-cell super-resolution imaging**. *Nat Chem* 2014, **6**:681-689. https://doi.org/10.1038/nchem.2002

*35. Beliu G, Kurz AJ, Kuhlemann AC, Behringer-Pliess L, Meub M, Wolf N, Seibel J, Shi ZD, Schnermann M, Grimm JB, et al.: **Bioorthogonal labeling with tetrazine-dyes for super-resolution microscopy**. *Commun Biol* 2019, **2**:261. https://doi.org/10.1038/s42003-019-0518-z

This paper describes the syntheses and evaluation of new trans-cyclooctene structures whose educts with tetrazines should be more stable in cells. New structures were carefully benchmarked against established Diels-Alder unnatural amino acids, using flow cytometry-based FRET measurements

36. Werther P, Yserentant K, Braun F, Kaltwasser N, Popp C, Baalmann M, Herten DP, Wombacher R: **Live-Cell Localization Microscopy with a Fluorogenic and Self-Blinking Tetrazine Probe**. *Angew Chem Int Ed Engl* 2020, **59**:804-810. https://doi.org/10.1002/anie.201906806

*37. Werther P, Yserentant K, Braun F, Grussmayer K, Navikas V, Yu M, Zhang Z, Ziegler MJ, Mayer C, Gralak AJ, et al.: **Bio-orthogonal Red and Far-Red Fluorogenic Probes for Wash-Free Live-Cell and Super-resolution Microscopy**. *ACS Cent Sci* 2021, **7**:1561-1571. https://doi.org/10.1021/acscentsci.1c00703

This paper describes the latest generation of blinking, fluorogenic rhodamines and silicon rhodamines, collectively called HDyes. Optimizations of proximal quenching resulted in orange/red fluorophore-tetrazine conjugates with the best fluorescene turn-on properties to date. Applications include in wash-free and super-resolution imaging via GCE labeling.

38. Tyson J, Hu K, Zheng S, Kidd P, Dadina N, Chu L, Toomre D, Bewersdorf J, Schepartz A: **Extremely Bright, Near-IR Emitting Spontaneously Blinking Fluorophores Enable Ratiometric Multicolor Nanoscopy in Live Cells**. *ACS Cent Sci* 2021, **7**:1419-1426. https://doi.org/10.1021/acscentsci.1c00670

39. Chemla Y, Ozer E, Algov I, Alfonta L: **Context effects of genetic code expansion by stop codon suppression**. *Curr Opin Chem Biol* 2018, **46**:146-155. https://doi.org/10.1016/j.cbpa.2018.07.012

40. Segal I, Nachmias D, Konig A, Alon A, Arbely E, Elia N: **A straightforward approach for bioorthogonal labeling of proteins and organelles in live mammalian cells, using a short peptide tag**. *BMC Biol* 2020, **18**:5. https://doi.org/10.1186/s12915-019-0708-7
41. Cohen S, Kramarski L, Levi S, Deshe N, Ben David O, Arbely E: **Nonsense mutation-dependent reinitiation of translation in mammalian cells**. *Nucleic Acids Res* 2019, **47**:6330-6338. https://doi.org/10.1093/nar/gkz319
42. Eberle AB, Stalder L, Mathys H, Orozco RZ, Muhlemann O: **Posttranscriptional gene regulation by spatial rearrangement of the 3' untranslated region**. *PLoS Biol* 2008, **6**:e92. https://doi.org/10.1371/journal.pbio.0060092
43. Lu T, Ting AY, Mainland J, Jan LY, Schultz PG, Yang J: **Probing ion permeation and gating in a K+ channel with backbone mutations in the selectivity filter**. *Nat Neurosci* 2001, **4**:239-246. https://doi.org/10.1038/85080
*44. Bartoschek MD, Ugur E, Nguyen TA, Rodschinka G, Wierer M, Lang K, Bultmann S: **Identification of permissive amber suppression sites for efficient non-canonical amino acid incorporation in mammalian cells**. *Nucleic Acids Res* 2021, **49**:e62. https://doi.org/10.1093/nar/gkab132

This paper describes the development of iPASS, which can computionally predict genetic code expansion efficiencies based on sequence contexts surrounding the desired incorporation site.

*45. Lafranchi L, Schlesinger D, Kimler KJ, Elsasser SJ: **Universal Single-Residue Terminal Labels for Fluorescent Live Cell Imaging of Microproteins**. *J Am Chem Soc* 2020, **142**:20080-20087. https://doi.org/10.1021/jacs.0c09574

This paper introduced STELLA, a single-amino acid label based on genetic code expansion and tracelessly cleaved ubiquitin and intein tags. The technique was used to image multiple microproteins and a number of SARS-CoV-2 ORFs.

*46. Meineke B, Heimgartner J, Eirich J, Landreh M, Elsasser SJ: **Site-Specific Incorporation of Two ncAAs for Two-Color Bioorthogonal Labeling and Crosslinking of Proteins on Live Mammalian Cells**. *Cell Rep* 2020, **31**:107811. https://doi.org/10.1016/j.celrep.2020.107811

This paper engineered the the PylRS/PylT pair from methanogenic archaeon ISO4-G1 first characterized for GCE by Willis and Chin and improved its orthogonality. The paper also demonstrates dual labeling of Notch and GPC receptors with two mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs

47. Bednar RM, Jana S, Kuppa S, Franklin R, Beckman J, Antony E, Cooley RB, Mehl RA: **Genetic Incorporation of Two Mutually Orthogonal Bioorthogonal Amino Acids That Enable Efficient Protein Dual-Labeling in Cells**. *ACS Chem Biol* 2021, **16**:2612-2622. https://doi.org/10.1021/acschembio.1c00649
48. Zheng Y, Mukherjee R, Chin MA, Igo P, Gilgenast MJ, Chatterjee A: **Expanding the Scope of Single- and Double-Noncanonical Amino Acid Mutagenesis in Mammalian Cells Using Orthogonal Polyspecific Leucyl-tRNA Synthetases**. *Biochemistry* 2018, **57**:441-445. https://doi.org/10.1021/acs.biochem.7b00952
49. Zheng Y, Addy PS, Mukherjee R, Chatterjee A: **Defining the current scope and limitations of dual noncanonical amino acid mutagenesis in mammalian cells**. *Chem Sci* 2017, **8**:7211-7217. https://doi.org/10.1039/c7sc02560b
50. Niu W, Schultz PG, Guo J: **An expanded genetic code in mammalian cells with a functional quadruplet codon**. *ACS Chem Biol* 2013, **8**:1640-1645. https://doi.org/10.1021/cb4001662

51. Chen Y, Wan Y, Wang N, Yuan Z, Niu W, Li Q, Guo J: **Controlling the Replication of a Genomically Recoded HIV-1 with a Functional Quadruplet Codon in Mammalian Cells**. *ACS Synth Biol* 2018, **7**:1612-1617. https://doi.org/10.1021/acssynbio.8b00096
52. Xi Z, Davis L, Baxter K, Tynan A, Goutou A, Greiss S: **Using a quadruplet codon to expand the genetic code of an animal**. *Nucleic Acids Res* 2021. https://doi.org/10.1093/nar/gkab1168
53. Italia JS, Latour C, Wrobel CJJ, Chatterjee A: **Resurrecting the Bacterial Tyrosyl-tRNA Synthetase/tRNA Pair for Expanding the Genetic Code of Both E. coli and Eukaryotes**. *Cell Chem Biol* 2018, **25**:1304-1312.e5. https://doi.org/10.1016/j.chembiol.2018.07.002
54. Willis JCW, Chin JW: **Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs**. *Nat Chem* 2018, **10**:831-837. https://doi.org/10.1038/s41557-018-0052-5
55. Meineke B, Heimgartner J, Lafranchi L, Elsasser SJ: **Methanomethylophilus alvus Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/Aminoacyl-tRNA Synthetase Pairs in Mammalian Cells**. *ACS Chem Biol* 2018, **13**:3087-3096. https://doi.org/10.1021/acschembio.8b00571
56. Beranek V, Willis JCW, Chin JW: **An Evolved Methanomethylophilus alvus Pyrrolysyl-tRNA Synthetase/tRNA Pair Is Highly Active and Orthogonal in Mammalian Cells**. *Biochemistry* 2019, **58**:387-390. https://doi.org/10.1021/acs.biochem.8b00808
57. Nikic I, Plass T, Schraidt O, Szymanski J, Briggs JA, Schultz C, Lemke EA: **Minimal tags for rapid dual-color live-cell labeling and super-resolution microscopy**. *Angew Chem Int Ed Engl* 2014, **53**:2245-2249. https://doi.org/10.1002/anie.201309847
58. Szatmari A, Cserep GB, Molnar TA, Soveges B, Biro A, Varady G, Szabo E, Nemeth K, Kele P: **A Genetically Encoded Isonitrile Lysine for Orthogonal Bioorthogonal Labeling Schemes**. *Molecules* 2021, **26**:4988. https://doi.org/10.3390/molecules26164988
59. Smeenk M, Agramunt J, Bonger KM: **Recent developments in bioorthogonal chemistry and the orthogonality within**. *Curr Opin Chem Biol* 2021, **60**:79-88. https://doi.org/10.1016/j.cbpa.2020.09.002
60. Hu Y, Schomaker JM: **Recent Developments and Strategies for Mutually Orthogonal Bioorthogonal Reactions**. *ChemBioChem* 2021, **22**:3254-3262. https://doi.org/10.1002/cbic.202100164
*61. Kang D, Kim J: **Bioorthogonal Retro-Cope Elimination Reaction of N,N-Dialkylhydroxylamines and Strained Alkynes**. *J Am Chem Soc* 2021, **143**:5616-5621. https://doi.org/10.1021/jacs.1c00885

This paper introduces the retro-Cope elimination of N,N-dialkylhydroxylamines and strained alkynes as a new bio-orthogonal reaction. The reaction has several desirable attributes: fast reaction rates; synthetic accessibility; intracellular compatibility; and orthogonality to inverse-electron-demand Diels-Alder reaction.

62. Kozma E, Nikic I, Varga BR, Aramburu IV, Kang JH, Fackler OT, Lemke EA, Kele P: **Hydrophilic trans-Cyclooctenylated Noncanonical Amino Acids for Fast Intracellular Protein Labeling**. *ChemBioChem* 2016, **17**:1518-1524. https://doi.org/10.1002/cbic.201600284
63. Cao X, Khitun A, Na Z, Dumitrescu DG, Kubica M, Olatunji E, Slavoff SA: **Comparative Proteomic Profiling of Unannotated Microproteins and Alternative Proteins in Human Cell Lines**. *J Proteome Res* 2020, **19**:3418-3426. https://doi.org/10.1021/acs.jproteome.0c00254
64. Slavoff SA, Mitchell AJ, Schwaid AG, Cabili MN, Ma J, Levin JZ, Karger AD, Budnik BA, Rinn JL, Saghatelian A: **Peptidomic discovery of short open reading frame-**

encoded peptides in human cells. *Nat Chem Biol* 2013, **9**:59-64. https://doi.org/10.1038/nchembio.1120

65. Schlesinger D, Elsasser SJ: **Revisiting sORFs: overcoming challenges to identify and characterize functional microproteins**. *FEBS J* 2021. https://doi.org/10.1111/febs.15769

66. Spence JS, He R, Hoffmann HH, Das T, Thinon E, Rice CM, Peng T, Chandran K, Hang HC: **IFITM3 directly engages and shuttles incoming virus particles to lysosomes. *Nat Chem Biol* 2019, **15**:259-268. https://doi.org/10.1038/s41589-018-0213-2

This paper demonstrated the power of genetic code expansion in driving biological studies of IFITM3, a small protein intractable to labeling with other means. The authors discovered, through fluorescent tagging of IFITM3 via genetic code expansion, that IFITM3 vesicles fuse with invading viral particles and direct the pathogenic cargo to lysosomes. Such mode of trafficking was found to depend on virus types and site-specific lipidation on IFITM3.

*67. Cao X, Khitun A, Luo Y, Na Z, Phoodokmai T, Sappakhaw K, Olatunji E, Uttamapinant C, Slavoff SA: **Alt-RPL36 downregulates the PI3K-AKT-mTOR signaling pathway by interacting with TMEM24**. *Nat Commun* 2021, **12**:508. https://doi.org/10.1038/s41467-020-20841-6

This paper used genetic code expansion to probe the subcellular localization and dynamics of Alt-RPL36, a new protein translated from an alternative open reading frame of RPL36 mRNA. Alt-RPL36 was found to localize partially to the ER, where it interacts with TMEM24, and regulates PIP2 levels and the PI3K-AKT-mTOR pathway.

68. Bessa-Neto D, Beliu G, Kuhlemann A, Pecoraro V, Doose S, Retailleau N, Chevrier N, Perrais D, Sauer M, Choquet D: **Bioorthogonal labeling of transmembrane proteins with non-canonical amino acids unveils masked epitopes in live neurons. *Nat Commun* 2021, **12**:6715. https://doi.org/10.1038/s41467-021-27025-w

Several advances were introduced in this paper: optimizations of GCE labeling for challenging venues (cultured neurons and organotypic brain slices); demonstration of the advantage of the small tag size in accessing hidden epitopes within proteins; applications to super-resolution imaging; and nanoscale distribution of different AMPAR auxiliary units in neuronal processes.

*69. Sappakhaw K, Jantarug K, Slavoff SA, Israsena N, Uttamapinant C: **A Genetic Code Expansion-Derived Molecular Beacon for the Detection of Intracellular Amyloid-beta Peptide Generation**. *Angew Chem Int Ed Engl* 2021, **60**:3934-3939. https://doi.org/10.1002/anie.202010703

This work combined genetic code expansion with FRET-based quenching to create a molecular beacon reporter for amyloid-β peptides and demonstrates their trafficking along the endocytic pathway. The approach could be applied to studying other proteolytic processed products such as neuropeptides.

*70. Shandell MA, Quejada JR, Yazawa M, Cornish VW, Kass RS: **Detection of Nav1.5 Conformational Change in Mammalian Cells Using the Noncanonical Amino Acid ANAP**. *Biophys J* 2019, **117**:1352-1363. https://doi.org/10.1016/j.bpj.2019.08.028

This paper used spectral imaging to detect emission shifts of Anap fluorescent amino acid in response to conformational changes of sodium channels in live cells.

71. Baumdick M, Gelleri M, Uttamapinant C, Beranek V, Chin JW, Bastiaens PIH: **A conformational sensor based on genetic code expansion reveals an autocatalytic component in EGFR activation**. *Nat Commun* 2018, **9**:3847. https://doi.org/10.1038/s41467-018-06299-7
72. Uttamapinant C, Howe JD, Lang K, Beranek V, Davis L, Mahesh M, Barry NP, Chin JW: **Genetic code expansion enables live-cell and super-resolution imaging of site-specifically labeled cellular proteins**. *J Am Chem Soc* 2015, **137**:4602-4605. https://doi.org/10.1021/ja512838z
73. Kramarski L, Arbely E: **Translational read-through promotes aggregation and shapes stop codon identity**. *Nucleic Acids Res* 2020, **48**:3747-3760. https://doi.org/10.1093/nar/gkaa136
74. Reinkemeier CD, Girona GE, Lemke EA: **Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes. *Science* 2019, **363**:eaaw2644. https://doi.org/10.1126/science.aaw2644

This paper introduced liquid-liquid phase separation as an approach to create an orthogonal translation system in mammalian cells, allowing selective translation and unnatural amino acid incorporation on desired mRNAs.

75. Reinkemeier CD, Lemke EA: **Dual film-like organelles enable spatial separation of orthogonal eukaryotic translation**. *Cell* 2021, **184**:4886-4903.e21. https://doi.org/10.1016/j.cell.2021.08.001
76. Asher WB, Geggier P, Holsey MD, Gilmore GT, Pati AK, Meszaros J, Terry DS, Mathiasen S, Kaliszewski MJ, McCauley MD, et al.: **Single-molecule FRET imaging of GPCR dimers in living cells**. *Nat Methods* 2021, **18**:397-405. https://doi.org/10.1038/s41592-021-01081-y

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

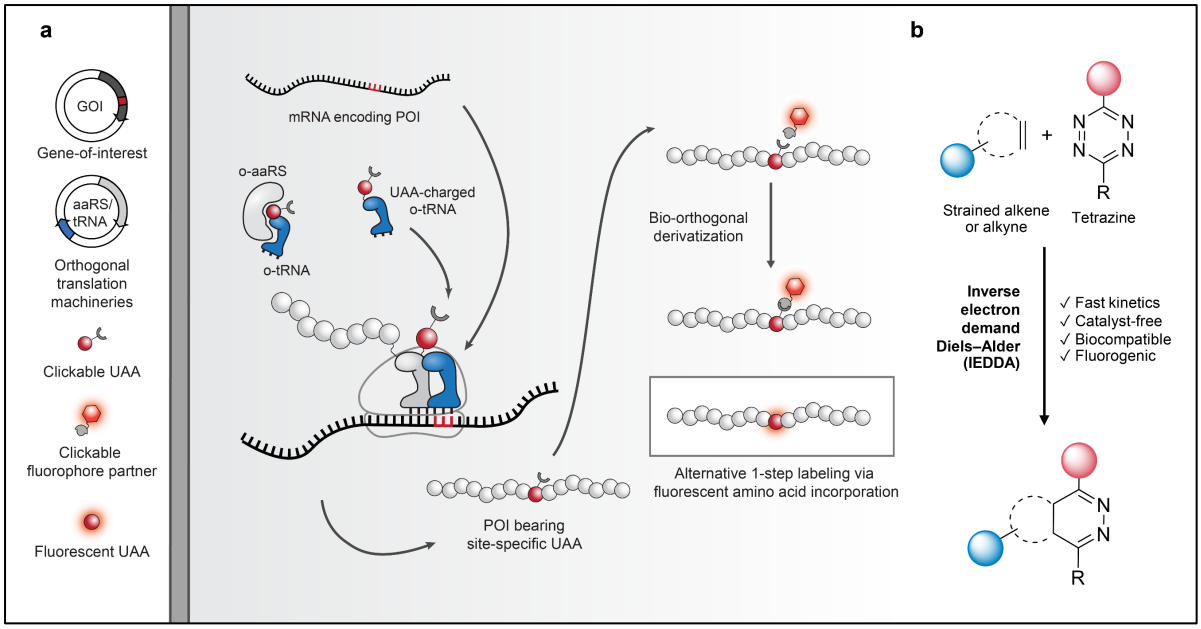

Figure 1. Genetic code expansion for cellular protein imaging. (**a**) Exogenous components of genetic code expansion are shown on the left. An unnatural amino acid (UAA, clickable or intrinsically fluorescent) supplied to the growth media is charged onto an orthogonal, amber-suppressor tRNA (o-tRNA) by an orthogonal aminoacyl-tRNA synthetase (o-aaRS). o-tRNA is decoded by the ribosome in response to an introduced amber codon placed in a gene of interest, allowing site-specific incorporation of the UAA into the protein of interest (POI). The clickable UAA is further derivatized with a fluorophore via bio-orthogonal chemistry. (**b**) Inverse-electron-demand Diels-Alder (IEDDA) cycloaddition as a tool for cellular protein labeling.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

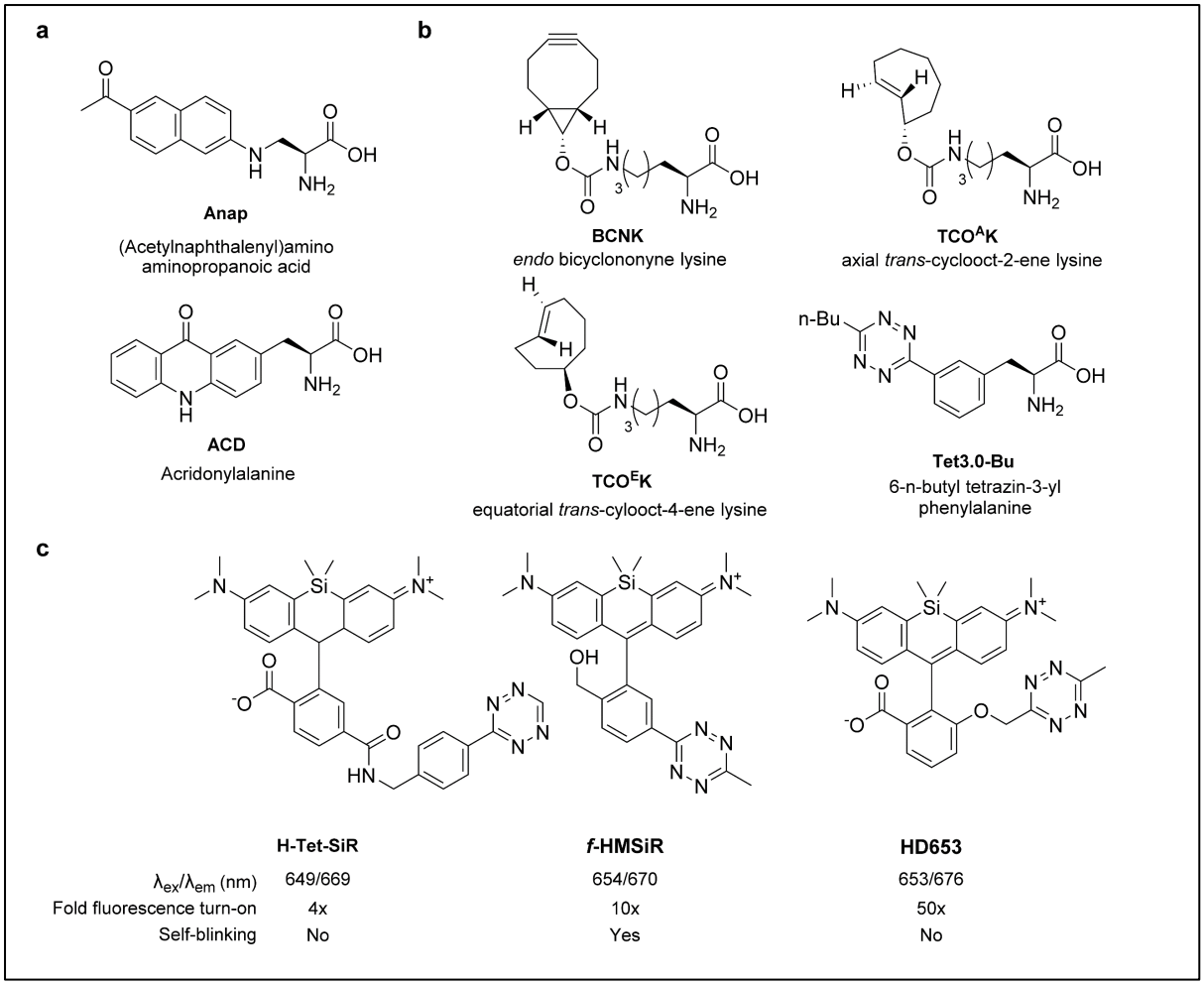

Figure 2. Unnatural amino acids and fluorophore conjugates for GCE labeling. (**a**) Fluorescent UAAs Anap and acridonylalanine. (**b**) Useful clickable UAAs for IEDDA. (**c**) Evolution of silicon rhodamine-tetrazine conjugates.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

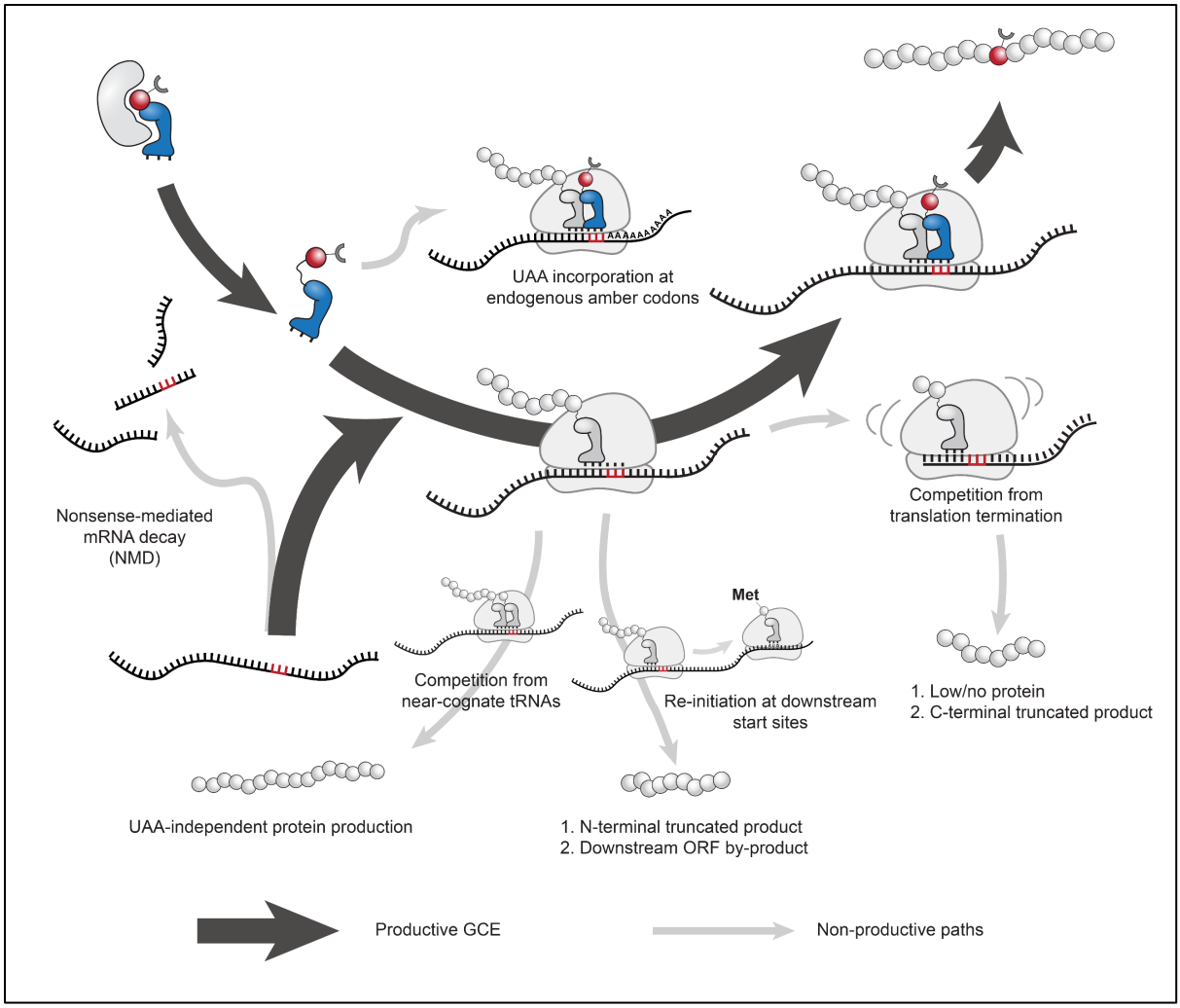

Figure 3. Desired path for productive genetic code expansion and non-productive paths, which affect signal-to-noise of protein labeling.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 21 December 2021 doi:10.20944/preprints202112.0331.v1

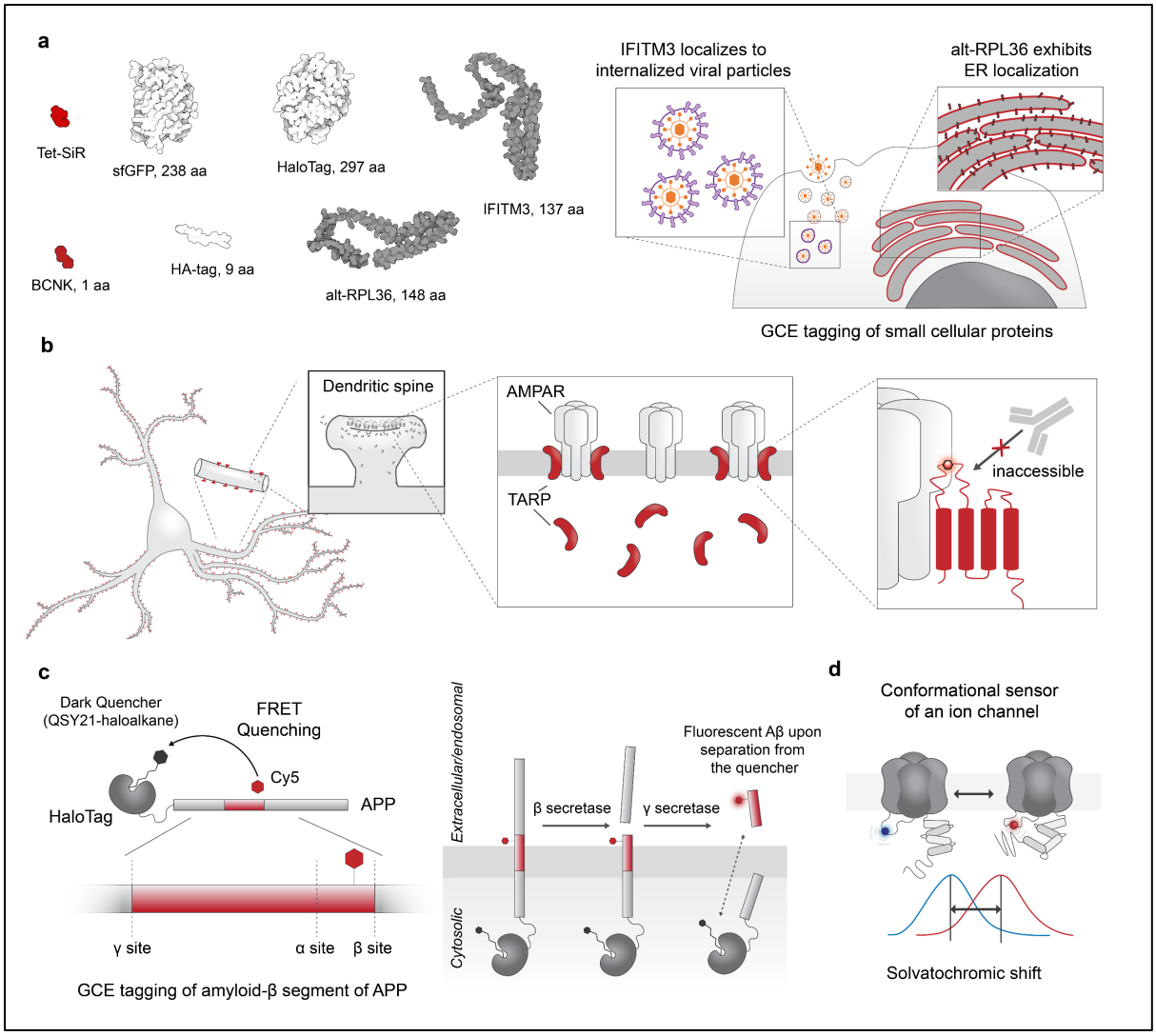

Figure 4. Cellular imaging applications of genetic code expansion. (**a**) Dynamic imaging of small cellular proteins IFITM3 [67] and alt-RPL36 [68]. Left, size of labeling tags—from small molecules BCNK and tetrazine-SiR to peptide and protein tags—compared to that of small proteins IFITM3 and alt-RPL36. Through GCE, interactions of IFITM3 vesicles with internalized viral particles can be tracked, and subcellular localization of alt-RPL36 revealed. (**b**) Super-resolution imaging of nanoscale organizations. Here GCE enables tagging and dSTORM imaging of transmembrane AMPA receptor regulatory proteins (TARPs), reveals different organizations of TARP members in dendritic spines, and preserves their modulation activity of AMPAR [69]. Antibodies cannot recognize TARPs due to steric occlusion. (**c**) Molecular-beacon labeling of amyloid-β (Aβ) peptides [70]. GCE is used to install a fluorescent dye cy5 into the Aβ-containing segment of amyloid precursor protein (APP). QSY21 fluorescence quencher suppresses cy5 fluorescence while APP is intact; quenching is relieved upon secretase-mediated processing of APP to generate Aβ. (**d**) Conformational sensors of a

 doi:10.20944/preprints202112.0331.v1

voltage-gated sodium channel $Na_v1.5$ based on solvatochromic shifts of Anap fluorescent amino acid [71].