REVIEW

Genetic Code Expansion and Optoproteomics

Yuting Chen[a], Linjie Lu[b], and Shixin Ye[b,c,d,*]

[a]Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University (ECNU), Shanghai, China; [b]Université Pierre-et-Marie-Curie, Laboratory of Computational and Quantitative Biology (LCQB), Institute of Biology Paris-Seine, Paris, France; [c]Institut National de la Santé et de la Recherche Médicale, Paris, France; [d]Centre National de la Recherche Scientifique, Paris, France

Nature has invented photoreceptor proteins that are involved in sensing and response to light in living organisms. Genetic code expansion (GCE†) technology has provided new tools to transform light insensitive proteins into novel photoreceptor proteins. It is achieved by the site-specific incorporation of unnatural amino acids (Uaas) that carry light sensitive moieties serving as "pigments" that react to light via photo-decaging, cross-linking, or isomerization. Over the last two decades, various proteins including ion channels, GPCRs, transporters, and kinases have been successfully rendered light responsive owing to the functionalities of Uaas. Very recently, Cas9 protein has been engineered to enable light activation of genomic editing by CRISPR. Those novel proteins have not only led to discoveries of dynamic protein conformational changes with implications in diseases, but also facilitated the screening of ligand-protein and protein-protein interactions of pharmacological significance. This review covers the genetic editing principles for genetic code expansion and design concepts that guide the engineering of light-sensitive proteins. The applications have brought up a new concept of "optoproteomics" that, in contrast to "optogenetics," aims to combine optical methods and site-specific proteomics for investigating and intervening in biological functions.

INTRODUCTION

An Innovative Molecular Biology Tool Based on Site-directed Mutagenesis

Proteins are essential biomolecules for cellular and physiological functions. To understand how a protein catalyzes a biological reaction requires techniques to specifically manipulate the protein and detect associated functional signals. Since the 1980s, the development of site-directed unnatural amino acid mutagenesis [1,2] has evolved as a powerful approach to correlate structure-activity relationship by introducing novel functionality into proteins. Similar to conventional site-directed mutagenesis, which allows one to change a specific amino acid in a protein to any of the other 20 naturally occurring amino acids, Uaa mutagenesis unleashes the limit of amino acid repertoire to enable introduction of unique chemical groups carried by the Uaa. Varieties of Uaas containing side-chains that serve as photoactive, spectroscopic, and redox probes have been successfully introduced into proteins, which facilitate the development of novel detection methods [3-9].

*To whom all correspondence should be addressed: Shixin Ye-Lehmann, Laboratory of Computational and Quantitative Biology, Institute of Biology Paris-Seine, University of Pierre and Marie Curie, Bldg. C, 3rd floor, Room 311a, 4 Place Jussieu, 75006 Paris, France, Tel: 33.(0)1.44.27.60.57, Email: shixin.ye-lehmann@upmc.fr.

†Abbreviations: GCE, genetic code expansion; Uaa, unnatural amino acid; aaRS, aminoacyl-tRNA synthetase; AcK, Nε-acetyl-lysine; AzF, p-azido-L-phenylalanine; GPCR, G protein-coupled receptor; ChR, channelrhodopsins; Tyr(ONB), caged-Tyr; nAChR, nicotinic acetylcholine receptor; Cmn, caged serine; Bpa, p-benzyol-L-phenylalanine; iGluR, ionotropic glutamate receptor; AMPAR, AMPA receptors; NMDAR, NMDA receptors; LBD, ligand-binding domain; NTD, N-terminal domain; P_o, open probability; PSAA, photo-switchable amino acids.

Keywords: Genetic code expansion, optoproteomics, optochemical genetics, ion channels, kinases interactions, unnatural amino acids

Table 1. Five commonly used orthogonal aaRS/tRNA pairs.

aaRS/tRNA	Origin	Orthogonal organisms
*Mj*TyrRS/tRNA	*Methanococcus jannaschii*	*E. coli*
*Ec*TyrRS/*Bst*tRNA	*E. coli* / *Bacillus stearothermophilus*	Eukaryotic cells
*Ec*TyrRS/*Ec*tRNA	*E. coli*	Eukaryotic cells
*Ec*LeuRS/tRNA	*E. coli*	Eukaryotic cells
*Mm*PylRS/tRNA	*Methanosarcina mazei*	*E. coli*, Eukaryotic cells

The common strategy of Uaa mutagenesis uses the amber stop codon suppression methodology, which involves the translation of an amber stop codon inserted into the gene of the targeted protein by a suppressor tRNA aminoacylated with a desired Uaa (Figure 1a). This strategy is later referred to as genetic code expansion, GCE [10], a concept based on assigning unnatural amino acids to the genetic code by using artificial coding rules. The main challenge is the generation of a Uaa aminoacylated suppressor tRNA. Three approaches have been developed (Figure 1b): 1) semi-chemical synthesis [1,2], 2) ribozyme [3,11], and 3) engineered aminoacyl-tRNA synthetase (aaRS) [4,10,12]. The first approach requires multiple steps of synthesis whereas the other two approaches are achieved through enzymatic reactions catalyzed by either ribozyme or aaRS. Although the main advantage of enzymatic reactions is that the Uaa aminoacylated tRNA can be generated *in situ*, the first approach offers an unmatched ability to aminoacylate any type of Uaa that would otherwise require specific design and screening of the ribozymes and aaRS.

Despite the fact that Uaa mutagenesis has evolved gradually into discovery tools, compared to many well-established molecular biology tools such as restriction enzymes [13], molecular cloning [14,15], site-directed mutagenesis [16-18], PCR [19,20], and genetically encoding GFP [21,22], most efforts are currently focused on technical developments for the reprogramming of the genetic code (Figure 2). Since the successful demonstration of Uaa mutagenesis via semi-chemical synthesis of suppressor tRNA [1,2], many concepts have been later shown to be feasible. When injecting the suppressor tRNA into *Xenopus* oocytes [23], it was possible to generate proteins site-specifically incorporating the Uaa, making it the first demonstration of *in vivo* Uaa mutagenesis. Later on, reassignment of quadruplet, synonymous codons, and codons containing artificial bases were proven to be feasible [24,25]. The development of engineered aaRS/tRNA pairs to recognize varieties of Uaas further advanced the *in vivo* Uaa mutagenesis [10,12]. Engineering more complex translational apparatus such as the ribosome has also been achieved [26], making it an attractive solution to reprogram the *Escherichia coli* genome to tolerate incorporation of multiple Uaas, although the applications of these orthogonal ribosomes in solving biological questions are still not fully demonstrated. The recent methodology development in creating transgenic animals including *Caenorhabditis elegans* [27,28], Drosophila [29] and mice [30,31] harboring aaRS/tRNA pairs for Uaa metabolism will further provide new biological insights into protein conformation changes, protein interactions, elementary processes in signal transduction, and the role of post-translational modifications *in vivo*.

GENETIC CODE EXPANSION

Orthogonal tRNA/RS Pairs

The most widely used approach utilizes the orthogonal aaRS/tRNA pairs delivered or incorporated into the host organisms for decoding the stop codon (Figure 3). The binding pocket of aaRS can be altered in order to recognize a Uaa with higher specificity than the natural substrate. The suppressor tRNA can recognize the blank codon, commonly the amber stop codon. Table 1 summarizes the available orthogonal tRNA/RS pairs screened by rational mutagenesis in *E. coli* to encode structurally distinct amino acids. These pairs were derived from natural aaRS/tRNAs in archaea or bacteria which are orthogonal to those in the host organisms. For example, The *Methanococcus jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS) pair is orthogonal in *E. coli* but not eukaryotic cells. *Ec*TyrRS/*Bst*tRNA pair, *Ec*TyrRS/*Ec*tRNA and *Ec*LeuRS/tRNA pairs are orthogonal in eukaryotic cells but cannot be reused in *E. coli*. Comparing the two TyrRS/tRNA pairs used in eukaryotic cells, *Bst*tRNA contains intact internal A-and B-box promoters for expression in mammalian cells, whereas the *Ec*tRNA requires external H1 and U6 promoters. Key residues in the substrate amino acid binding pocket are mutated in order to recognize varieties of Uaas [4,10,12]. Overall the *Pyl*RS/tRNA pair is arguably the most useful pair for GCE, because *Pyl*RS is orthogonal in all kinds of cellular systems to encode pyrrolysine in response to the amber stop codon [32,33]. In addition, the plasticity of the *Pyl*RS makes it ideal to engineer variants to recognize varieties of Uaas.
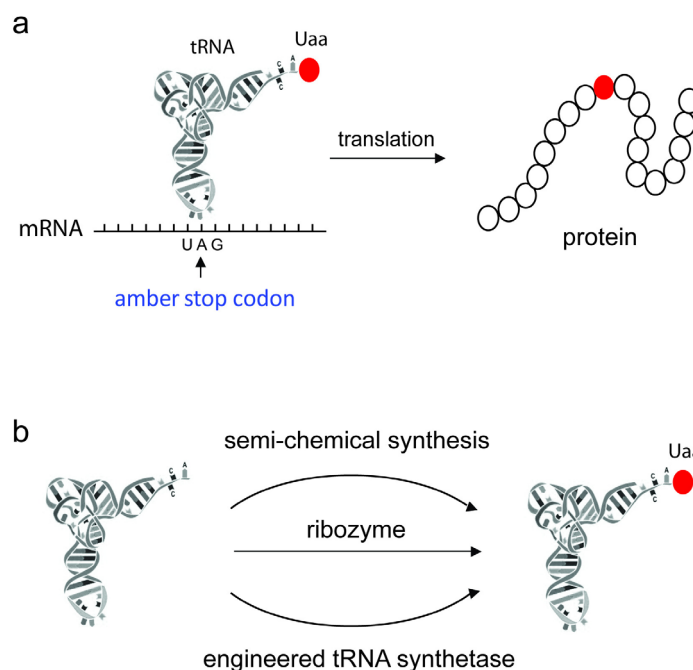
Figure 1. Schematic illustration of Uaa incorporation through the amber stop codon suppression. **a**. Engineered suppressor tRNA carrying a desired Uaa (red dot) recognizes the UAG stop codon in the mRNA, which allows read-through of protein translation. The protein product has the Uaa incorporated at the specific site. **b**. Three routes to generate Uaa aminoacylated suppressor tRNA: 1) semi-chemical synthesis, 2) ribozyme, 3) engineered aminoacyl tRNA synthetase.

Unnatural Amino Acids

More than 100 different Uaas have been designed and incorporated into proteins [5-9,34]. A large fraction of them have been further exploited by screening aaRS/tRNA pairs, which could eventually lead to applications *in vivo*. Uaas containing spectroscopic probes, post-translational modifications, metal chelators, photo affinity labels, and other chemical moieties have been selectively incorporated into proteins with good fidelity and efficiency. They have been used to study biological processes and address the emerging challenges for structure-activity relationships of proteins that are inaccessible to other methods.

Genome Editing Tools to Introduce New Coding Rules

The simplest way to provide orthogonal tRNA/aaRS pairs for a cellular system is by the delivery of the genes encoded in plasmid vectors, which are under the control of appropriate promoters for transient expression into the host cells, such as *E. coli*, yeast, oocytes, mammalian cells (HEK293, Hela, CHO cells), or primary cultured cells.

Implementing GCE to study proteins in whole animals is challenging, and several model systems have been established. These demonstrations have tackled two major challenges: (i) inefficient gene delivery of suppressor tRNA, orthogonal synthetase, and protein of interest to the animals; and (ii) insufficient bioavailability of the Uaa at the desired tissue or cell type to enable acceptable level of mutant expression. It was first established in simple animals *C. elegans* [27,28], *D. melanogaster* [29], and recently in vertebrate animals including zebrafish [30,35] and mice [30,31]. *In utero* electroporation of plasmid DNAs coupled to direct injection of the Uaa into the embryonic brain provided a first possible method [36]. Gene transfer using lentiviral [30,31], adeno-associated viral (AAV) [37] and baculoviral vectors [38] offer an alternative way for efficient delivery of the Uaa genetic machinery in mammalian cells, tissues, and brains of living mice. Recently, GCE has been successfully developed in mice with a heritable expanded genetic code [30,31] – *i.e.* animals with tRNA/aaRS genes integrated into their genome. They provide opportunities to study human physiology and disease *in vivo*, because the mouse genome is more than 99 percent similar to that of humans and it contains most human gene counterparts or functionally related genes [39]. Meanwhile, the mouse has a short life span and is easy to breed and handle in the laboratory. More importantly, the mouse genome is well-known, which makes it possible to generate custom-made mutant mouse strains for *in vivo* studies of
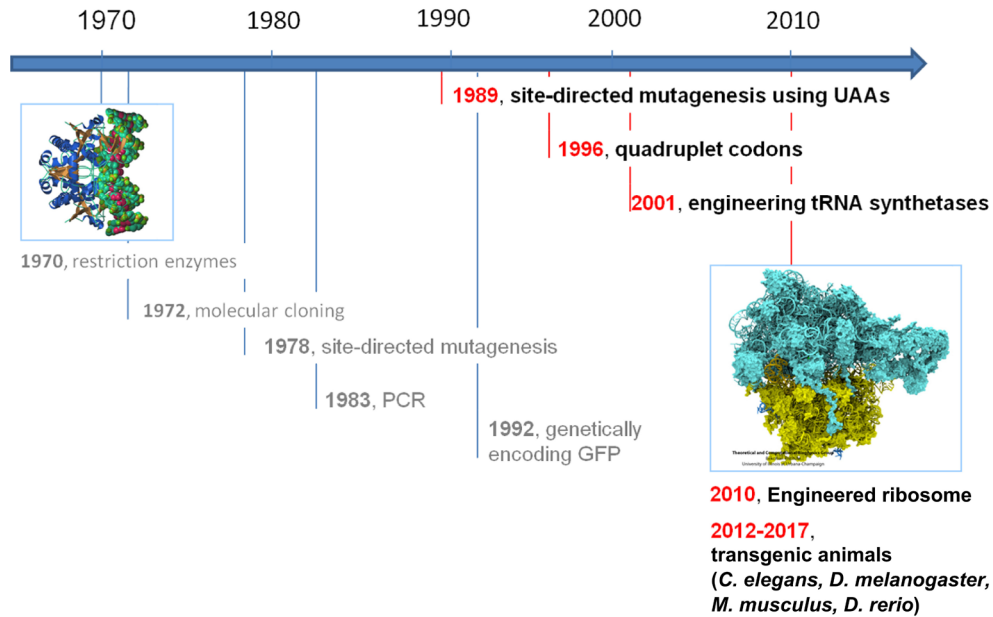
Figure 2. Timeline of the technical developments of genetic code expansion (black font) and other molecular biology tools (grey font).

specific genes and providing excellent models for various human diseases [40]. Now mouse strains allowing incorporation of the photocrosslinker AzF [30] and the post-translationally modified Nε-acetyl-lysine AcK [31] have become available. We expect this strategy to be extended to other tRNA/aaRS pairs in a custom-made manner, including proteins of interest containing amber stop mutation, which could be stably integrated into the genome for physiological studies.

Uaa delivery in the mice can be achieved by intraperitoneal injection, or direct injection into target tissues for site-specific induction of Uaa-engineered protein expression [31]. Uaa supplementation in the mouse drinking water provides a convenient option, even when targeting brain proteins as recently shown with an AcK [37]. In zebrafish, incubation of live animals in water containing AzF was shown to be sufficient for Uaa delivery [30,35].

APPLICATIONS IN OPTOGENETICS AND OPTOPROTEOMICS

A toolbox of chemical biology approaches centered on GCE technology has enabled the incorporation of Uaas serving as unique physical and chemical probes into proteins to correlate protein structure activity relationship. There are many excellent reviews on a variety of applications of Uaas for studies of ligand-gated ion channels [5,41,42] and GPCRs are excellently reviewed [41,43,44-46]. Our understanding of functional properties of those important receptors is increasing

rapidly due to the combination of X-ray crystal structure characterization and GCE applications in recent years.

In the past decade, optogenetics based on natural light-sensitive proteins has evolved as a powerful strategy to study cellular functions. In that technology, channelrhodopsin (ChR, a light-driven ion channel found in archaebacteria) is genetically encoded into cellular systems such as neurons to enable the regulation of their function by light induced activation of ChR [47]. Despite the power of manipulation of biological functions with cell specificity, the regulation of a specific protein function by light remains a major challenge. The approach we call "optoproteomics," combining optical methods and protein-specificity for investigating and intervening cellular functions, provides a general solution to regulating physiological function by light.

Here we will focus on the applications of GCE technology to reprogram proteins in response to light. Light is a unique signal input with three major advantages for studying biological functions: 1) it has high temporal and spatial resolution to control a biological reaction; 2) reaction kinetics are significantly faster and more specific than small molecule based controls that rely on molecule diffusion; 3) cooperative mechanisms can be elucidated by combining controls by light and small molecules. The design concept is to introduce a light-sensitive molecule into proteins serving as a switch to regulate protein function. Light-sensitive Uaas can bestow proteins with such photoswitches, inducing a specific functional change in the protein upon illumination (Figure 4a). To date, three types of light-sensitive Uaas have been exploited to
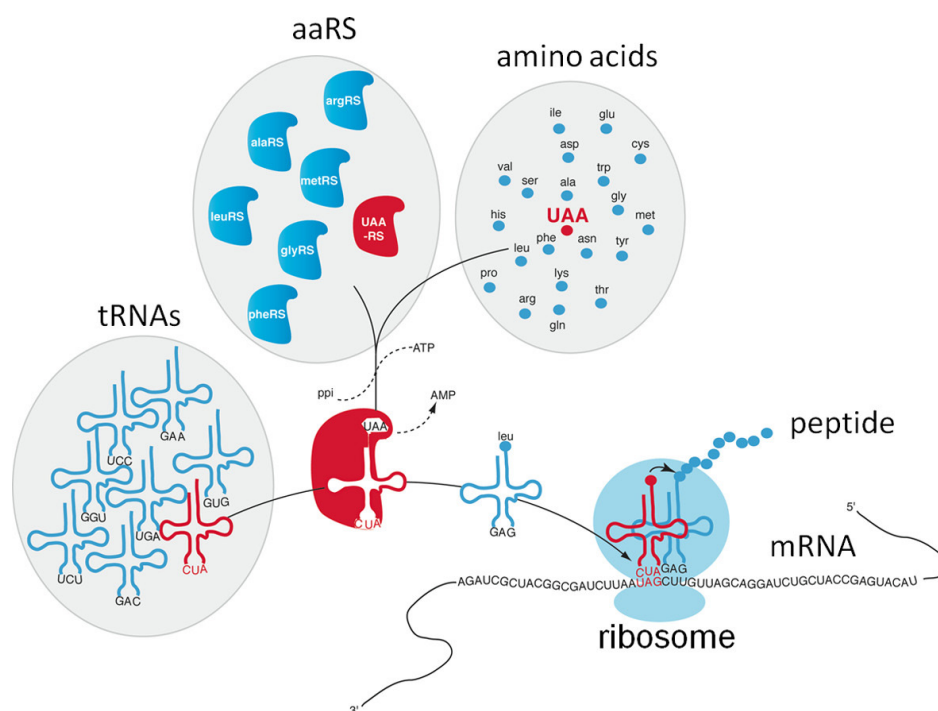
Figure 3. Schematic illustration of genetic code expansion utilizing the engineered aminoacyl tRNA synthetase/ tRNA pairs (colored red). The orthogonal aaRS recognizes the Uaa and aminoacylates the suppressor tRNA *in situ*, which participates in the protein translations to incorporate Uaa at a site where amber stop codon is inserted into the mRNA. Endogenous aaRS/tRNAs and amino acids are colored blue.

transform light-insensitive proteins to respond to light: 1) caged-Uaas which release a photolabile moiety from the side chain upon light activation; 2) photo-cross-linking Uaas, which cross-link to a nearby molecule upon light activation; 3) photo-switchable Uaas, which toggle in between two conformations reversibly upon illumination with different wavelengths (Figure 4b).

Caged Amino Acids

This category of Uaas has side chains protected by a photochemically removable group (Figure 4b). Upon light activation, the protection group – or the cage – is released to generate the uncaged side chain. With a caged-Uaa introduced at a key functional site in the protein, protein functions can be activated due to the light stimulations.

Tyr(ONB)

Early experiments incorporating caged-Tyr Tyr(ONB) into proteins demonstrated the feasibility of time-resolved studies on protein function by decaging via photolysis [48]. Tyrosine, serine, threonine are substrates of a diversity of kinases, and phosphorylation of these residues plays a key role in the regulation of a wide range of pathways. For example, kinase-dependent modulation of ion channels and neuronal receptors have been demonstrated to play key roles in synaptic plasticity,

the changes in the structure and composition of a synapse that are associated with learning and memory. Tyr(ONB) along with caged serine and threonine, hold special promise as general tools to regulate a specific tyrosine phosphorylation target in signaling pathways. Tyr(ONB) was first incorporated into the alpha subunit of mouse muscle nicotinic acetylcholine receptor (nAChR), which belongs to the pentameric ligand-gated ion channel family [48]. When introduced at different tyrosine positions (93Y, 127Y, 198Y), three types of light-induced kinetics were observed, providing the first demonstration of the light-control of protein function through caged Uaas. In particular, the slower kinetics observed for 93Tyr(ONB) and 198Tyr(ONB) could correspond to conformational changes of the receptor upon decaging, although no further information was available about how the state of the phosphorylation affected the receptor function. In a subsequent study with Tyr(ONB) introduced at the conserved tyrosine site in the K$^+$-channel Kir 2.1, photo-decaging led to a ~30 percent decrease in K$^+$-channel activity when the kinase v-src was also expressed in the oocyte [49]. Detailed electrophysiology studies revealed that decaging initiated two kinase-dependent pathways. One involved direct modulation of the channel presumably due to the conformational changes associated with phosphorylation. The other involved kinase-initiated endocytosis of receptors, which diminished the observed
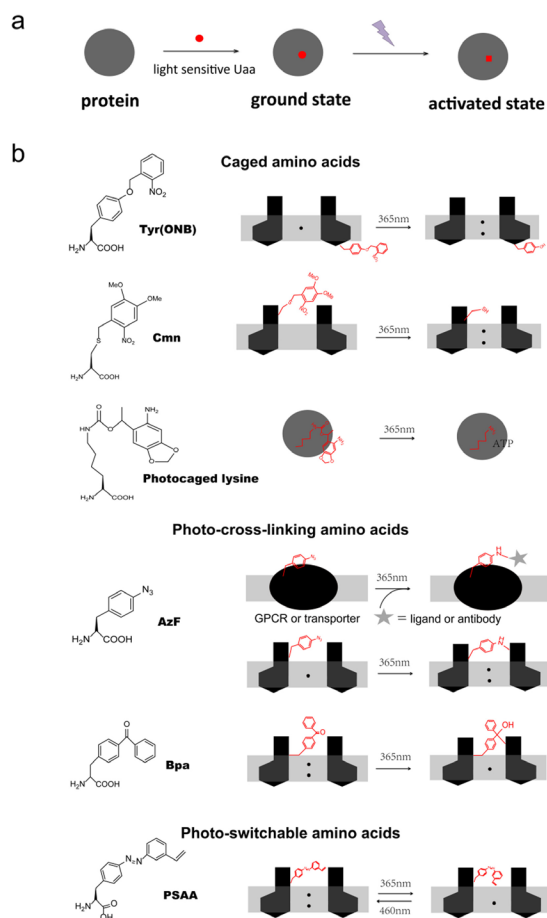
Figure 4. Schematic illustration of engineering photo-responsive proteins with light-sensitive Uaas. **a**. Inserting a light-sensitive Uaa into a protein can transform a protein to respond to light. Upon light stimulation, protein changes from the ground state to the activated state due to the photo-responses of the Uaa, which leads to a functional read-out. **b**. Three main categories of light responsive Uaas: caged amino acids, photo-cross-linking amino acids, and photo-switchable amino acids. When these Uaas are introduced into ion channels (black channels) or other proteins (grey circles), various biological responses have been observed.

K$^+$ current. Dissecting these two competing pathways would have been very difficult using conventional tools. When coupled with caged serine, threonine, and phospho-amino acids, these strategies have the potential for providing new insights into phosphorylation pathways of biological signaling.

Photocaged Lysine

Although powerful, the cases of incorporating Tyr(ONB) relied on semi-chemical synthesis of tRNA, which limits applications *in vivo*. Engineering an aaRS/tRNA pair that accepts light-activatable amino acids as substrates provides an attractive solution to achieve *in vivo* applications. The first such application established was incorporation of photocaged lysine for creating kinases that can be activated in living mammalian cells by light [50,51]. These studies made replaced the conserved lysine in the ATP binding pocket of the catalytic domain with caged lysine. The crystal structure of the MEPK/ATP complex reveals that the lysine anchors and orients ATP for catalytic functions. By replacing the lysine with caged lysine, the kinase activity is almost completely blocked due to the inability to bind ATP for kinase function. Upon light activation, the cage is released and lysine is restored. This approach allows the design of subnetworks controlled solely by light to study the kinetics of individual steps in signaling cascades and the dissection of temporal regulation. This methodology provides high spatial and temporal resolution that is inaccessible to other methodologies. Due to the substantial conservation of the targeted lysine residue (present in 95 percent of human protein kinases), the light-activation method reported here could be generally and readily applicable to creating photo-activated versions of other protein kinases. Moreover, by applying the lysine photocaging to each kinase in a pathway, precise quantitative insights
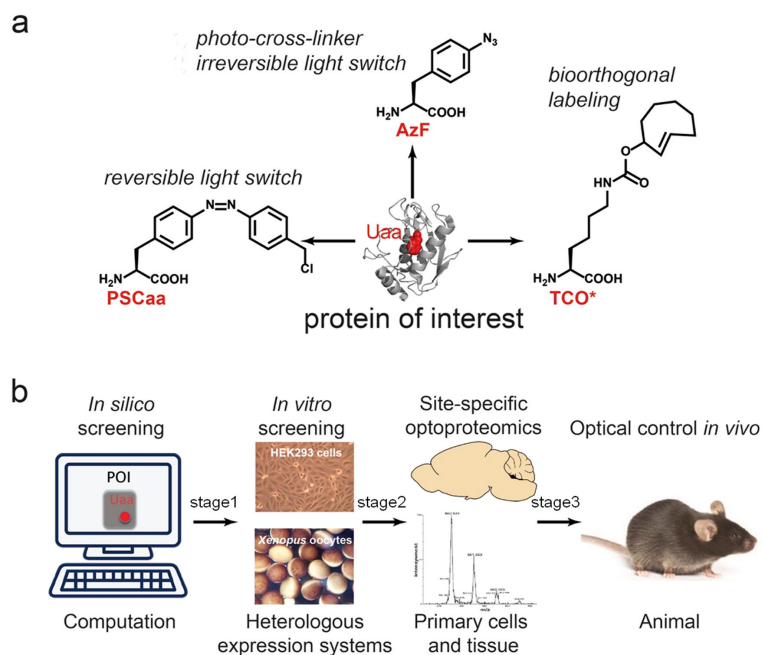
Figure 5. Concept of optoproteomics. **a**. Introducing various Uaas to render a protein of interest to be light sensitive by site-specific incorporation of Uaas. **b**. Engineering a light-sensitive protein with *in silico* screening based on the atomic structures, *in vitro* functional screening, and system-wide investigation of protein specific response in tissues and *in vivo*.

into the kinetics of kinase networks, and the identification of substrates of individual kinases, become possible. Such developments are attractive because signal transduction pathways use cascades of protein and lipid kinase-mediated phosphorylation to elicit specific responses to distinct extracellular stimuli. Kinase pathways are extremely complex, dynamic, multistep processes with cross-talk, feed-back and feed-forward steps that are hard to differentiate. Light activatable kinases provide unique tools to uncover the architecture of pharmacologically important, time-dependent, adaptive control processes, such as feedback inhibition. Recently, photo-activatable kinases have been applied in living zebrafish embryos [35].

Besides kinases, the caged lysine has been introduced into Cas9 protein making a light-activated Cas9 in mammalian cells, an attractive CRISPR/Cas9 system to control exogenous and endogenous gene functions by light [52]. These discoveries provide new tools for the optochemical control of protein function and gene editing in mammalian cells and tissues.

Caged Cysteine and Caged Serine (Cmn)

Many proteins use cysteine for a broad range of functional roles through the intrinsic properties of the thiol group. When a caged cysteine was introduced into the substrate-binding pocket of *Renilla* luciferase, the bulky caging group caused a complete loss of enzyme activity [53]. Upon brief exposure to UV light which removes the cage and restores the coelenterazine binding the activity was rescued (though not reaching the wild-type level), thus showcasing the applicability of the caged cysteine controlling protein activity in live human cells. The incomplete activity restoration may be due to the decaging not reaching 100 percent.

Cmn has also been used to generate a photoactivatable Kir2.1 [36], by introducing Cmn into a specific site in the transmembrane region. The caged-serine introduced at C149 site had the most pronounced light sensitivity. Before light treatment, the Kir2.1 channel showed minimal activation due to Cmn occluding the pore and blocking the channel. After UV illumination, the cage is released to generate a serine side chain which is similar to the native cysteine, restoring the outward K^+ activity. This light activatable Kir2.1 was screened in mammalian HEK293 cells, and further implemented in cultured rat hippocampal neurons as well as embryonic mouse neocortex to achieve light-activated suppression of neuronal firing specific to Kir2.1, demonstrating the proof-of-concept of optical control of protein function using genetically encoded Uaas.

Photo-cross-linking Amino Acids

This category of Uaas has side chains that become reactive when exposed to light (Figure 4b). They will conjugate to any one of several common functional

groups in proteins upon activation. This feature makes them particularly useful in capturing protein-protein or protein-ligand interactions, and has been widely exploited in the studies of signaling pathways involving G-protein coupled receptors (GPCRs) [43,44], the most popular membrane protein family for drug development. Compared to the long history of engineering light-activatable proteins using caged amino acids, the development of light-sensitive proteins with photo-cross-linking amino acids is a more recent phenomenon.

A series of pioneering studies focused on encoding photo-cross-linking Uaas AzF and Bpa into GPCRs [43-47]. The encoding of AzF and Bpa takes advantage of an *E. coli* TyrRS and *Bacillus stearothermophilus* Tyr-tRNA$_{CUA}$ pair for incorporating tyrosine analogs into proteins in mammalian cells at high efficiency [54]. High efficiency was achieved by expression of the suppressor Tyr-tRNA$_{CUA}$ of bacterial origin using conserved promoter sequences of the mammalian Tyr-tRNAs [55,56]. The challenge of directly observing photo-cross-linking was first overcome to observe a photo-induced GPCR-ligand complex. GPCRs such as chemokine receptors containing AzF or Bpa introduced at specific sites of the ligand-binding pocket can covalently conjugate their cognate ligands such as T140 (an HIV-1 coreceptor blocker) [57]. Many other GPCRs and their interacting partners other than ligands (such as antibody epitopes) [58-61] have been investigated and characterized using this approach, providing novel biological insights. In addition to GPCRs, complexes between serotonin transporter and two antidepressant drugs have been systematically studied, precisely defining the high-affinity drug-binding site in this important drug target for the treatment of depression [62].

Inspired by the works on GPCRs, AzF and Bpa were also incorporated to create light-sensitive ionotropic glutamate receptor (iGluR) family proteins, including AMPA receptors [63] and NMDA receptors [64-66]. These receptors mediate excitatory synaptic transmissions and are made of a tetrameric assembly of subunits. Sites were chosen at subunit interface revealed by the X-ray crystal structures. The incorporation of Bpa into the AMPA receptor GluA2 at two sites (S729 or G725) located at subunit interfaces between ligand-binding domains (LBD), allowed UV-induced state-dependent photo-inactivation. When pairing GluA2-AzF with GluA1, there was photo-inactivation although it was less than that of homomeric GluA2 receptors [63]. Overall, these AMPA mutants undergo the same desensitization mechanism for light inactivation by Bpa or AzF.

In contrast to AMPA receptors, surprisingly, light-sensitive NMDA receptors with different mechanisms have been identified by the incorporation of AzF at the subunit interface. This interface is at N-terminal domains (NTD), which are extracellular domains adjacent to the LBD where multiple allosteric modulator binding sites have been identified. Two Glu2B AzF mutants in residues at the NTD interface, where the synthetic compound ifenprodil binds (revealed by X-ray crystal structures), showed opposite light-dependent effects [65,66]. GluN1-Y109AzF/GluN2B demonstrated UV-induced inactivation [65], whereas GluN1/GluN2B-F114AzF demonstrated UV-induced potentiation [66]. Why would introducing AzF at the same interface, one in the GluN1 subunit, and the other one in the GluN2B subunit elicit completely different light-induced modulations? It turns out that in the case of GluN1-Y109AzF/GluN2B, UV induced intra-subunit crosslinking within GluN1 changed the receptors from the wild-type conformation to a low channel-open-probability (P$_o$) state. Whereas for the GluN1/GluN2B-F114AzF, UV induced inter-subunit between GluN1/GluN2B restored the receptors from a low P$_o$ to the wild-type state. Effects on ion conductivity and gating by the natural ligands or allosteric modulators (either inhibitors such as Zn^{2+} and ifenprodil, or potentiators such as spermine) can then be compared to the wild type receptors or the uncrosslinked mutants. Therefore, those light-sensitive NMDARs serve as novel "photoreceptors" to probe receptor functions, providing new platforms to uncover novel mechanisms of allosteric regulation. Such design concept through interfacial crosslinking has recently been extended to potassium channels [67,68]. These studies highlight the pharmacological mechanisms of receptors that make them promising drug targets for the treatment of diseases.

Photoswitchable Amino Acids (PSAAs)

PSAAs containing azobenzene moiety can reversibly change conformation under two wavelengths of light (Figure 4b). This reversibility is unique among other optoproteomic approaches discussed above. The concept was initially demonstrated in horseradish peroxidase [69], catabolite activator protein [70], and calmodulins [71] by incorporating various PSAAs. Recently photoswitchable NMDARs were generated by incorporating PSAAs [72]. This study proved for the first time the feasibility to convert neuronal receptors to respond to light with reversible control by two different wavelengths of light. PSAAs incorporated at several positions within the multi-domain receptors endowed robust light-induced modulations. They demonstrated different function switching mechanisms. When introduced at the GluN1 clamshell hinge region, PSAA photoisomerization enabled the control of receptor's glycine sensitivity and activation efficacy. When inserted into the site in the pore domain which contributes to a conserved transmembrane cavity, PSAA photoisomerization impacted both gating and permeation properties due to the conformational

changes of the membrane helix M3. This work demonstrates not only a unique approach to understand biophysical properties of transmembrane pore sites whose conformational changes during receptor activity remain poorly understood, but also unveiled the contribution of specific side chains to distinct receptor properties such as agonist sensitivity, channel open probability, and permeation.

Optical control of protein function provides excellent spatial-temporal resolution for studying proteins functions *in situ*. The general advantage of the GCE approach is that it provides molecular selectivity and an acute modulation of protein functions inaccessible to other methodologies. Combined with rapid developments in structural biology, the goal is to gain functional information in a protein's native environment and to develop new therapeutic strategies [73-76]. The development of light-controllable ion channels with genetic encodability, as reviewed above, should be valuable to manipulate receptor signaling and unmask specific sidechain roles in physiological function. We expect these principles to be generally applied to other proteins, enabling optical investigation of a range of receptors, ion channels, and kinases both in recombinant and native systems.

Pros and Cons

Complementing other approaches which rely on introducing photo-sensitive ligands to engineer light-sensitive receptors [74-76], the GCE directly inserts the light-sensitive moiety into the protein at any site using a stop codon suppression system. Using Uaa-based optoproteomics to control protein function offers several key advantages over existing methodologies: 1) The genetic encodability affords high molecular (*i.e.* subunit) specificity and cell-type targeting. 2) Varieties of photo-switching mechanisms can be identified. By targeting allosteric mechanisms using photo-cross-linkers and PSAA, rather than direct orthosteric interactions as for photoswitchable tethered ligands, it provides the ability to fine-tune receptor activity without interfering with its natural activation. 3) Different types of light functionalities can be installed into the same protein using a single site.

However, there are many challenges to tackle for the *in vivo* applications. The low efficiency of mutant expression due to the inefficient incorporation of bulky Uaas such as caged Uaas and PSAA, and low efficiency of amber suppression and *in vivo* have to be kept in mind. Several strategies have been implemented to increase the efficiencies, such as optimization of tRNA gene design [35,55,77], development of controlled delivery of orthogonal pairs via viral vectors [37,38], and application of siRNA to down-regulate mammalian Upf2, which is the key regulator of nonsense mediated mRNA decay pathway [31].

Another major challenge is to identify the proper site to introduce the light functionality for a given protein. Proteins differ from one another primarily in their sequence of amino acids, which is dictated by the nucleotide sequence of their genes, and results in protein folding into a specific three-dimensional structure that determines its activity. We and others have shown that light-sensitive Uaas introduced at conformation sensitive sites can generate light-responsiveness. However, previously sites were often chosen based on previous functional studies using conventional site-directed mutagenesis in combination with empirical analyses of available X-ray crystal structures that are static. Functional mutants can be validated by conducting cell-based screening assays. In practice there is often a low success rate, with more than half of the designed mutants being light insensitive, likely due to the sheer complexity of the protein structure-function relationship. To increase the success rate, prediction methodologies could be improved using molecular modeling [78]. Recent progress in system-wide protein computational analyses has made a profound impact in elucidating sequence-structure relationship with significantly improved accuracy. For example, structural based protein design algorithms can predict conformational changes with high accuracy. For proteins without high-resolution structures, ancestral sequence reconstruction analysis can provide insights into the specific amino acids in a given protein sequence that contribute to the desirable properties, such as increased flexibility and conservation. Sequence coevolution analysis enables the prediction of protein–protein interactions at the precision of amino acids. It is therefore conceivable to apply these analyses to design light-responsive proteins on a large-scale with increased prediction precision. When implementing a designed protein *in vivo* to control functionality with light through optoproteomics, as illustrated in Figure 5, we can discover the physiological effect of a targeted protein.

CONCLUSIONS AND OUTLOOK

Advances in GCE have opened up new landscapes in optoproteomics. Designed light sensitive proteins provide an expanding palette (Figure 4) for specific electrical or biochemical effector function control through well-defined protein functional changes. In some sense, optogenetics could be considered a branch of optoproteomics due to the utilization of naturally occurring microbial channelrhodopsins to mediate neuronal and physiological activities in organisms of higher complexity. The temporal precision of the use of light along with the specific protein functional change enabled by GCE creates a unique link between a molecular event defined precisely through a

change of a single amino acid side-chain in the protein and its physiological consequences. To realize the goal of *in vivo* light control, advances in light targeting/readout technologies offered by optogenetics will provide solid foundation. In the future, models of genetic diseases caused to specific proteins carrying amber stop codon mutations [79] may be established and explored by optoproteomics. Firstly, the read-through of the stop codon could lead to a phenotypic disease rescue; and secondly, the light could introduce a proteomic change to enhance the therapeutics. Such efforts will undoubtedly reveal more unexpected and exciting roles for nonsense mutations in human diseases.

REFERENCES

1. Heckler T, Chang L, Zama Y, Naka T, Hecht S. Preparation of '2,('3)-O-acyl-pCpA derivatives as substrates for T4 RNA ligase-mediated "chemical aminoacylation". Tetrahedron. 1984; 40:87–94.
2. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. A general method for site-specific incorporation of unnatural amino acids into proteins. Science. 1989; 244:182.
3. Hohsaka T, Sisdo M. Incorporation of non-natural amino acids into proteins. Curr Opin Chem Biol. 2002;6:809-15.
4. Link AJ, Mock ML, Tirrell DA. Non-canonical amino acids in protein engineering. Curr Opin Biotechnol. 2003.
5. Beene DL, Dougherty DA, Lester HA. Unnatural amino acid mutagenesis in mapping ion channel function. Curr Opin Neurobiol. 2003;13:264–270.
6. Liu CC and Schultz PG. Adding new chemistries to the genetic code. Annu Rev Biochem. 2010.
7. Davis L and Chin JW. Designer proteins: applications of genetic code expansion in cell biology. Nat Rev Mol Cell Biol. 2012;13:168-82.
8. Chin JW. Expanding and reprogramming the genetic code of cells and animals. Biochem, 2014;83:379–408.
9. Xiao H and Schultz PG. At the interface of chemical and biological synthesis: an expanded genetic code. Cold Spring Harb Perspect Biol. 2016;8(9):a023945.
10. L Wang, Brock A, Herberich B, Schultz PG. Expanding the Genetic Code of Escherichia coli. Science. 2001;292:498-500.
11. Bessho Y, Hodgson DRW, Suga H. A tRNA aminoacylation system for non-natural amino acids ased on a programmable ribozyme. Nat Biotechnol. 2002;20:723-728.
12. David R. Liu, Thomas J. Magliery, Miro Pastrnak, and Peter G. Schultz. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. Proc Natl Acad Sci U S A. 1997;94:10092-7.
13. Kelly TJ, Jr, Smith HO. A restriction enzyme from He-

14. Jackson DA, Symons RH, Berg P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. Proc Natl Acad Sci U S A. 1972;69:2904-9.
15. Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci U S A. 1973;70:3240-4.
16. Flavell RA; Sabo DL; Bandle EF, Weissmann C. Site-directed mutagenesis: effect of an extracistronic mutation on the in vitro propagation of bacteriophage Qbeta RNA. Proc Natl Acad Sci U S A. 1975;72:367–371.
17. Müller W, Weber H, Meyer F, Weissmann C. Site-directed mutagenesis in DNA: Generation of point mutations in cloned β globin complementary DNA at the positions corresponding to amino acids 121 to 123. J Mol Biol. 1978;124:343–358.
18. Hutchison CA, Phillips S, Edgell MH, Gillam S, Jahnke P, Smith M. Mutagenesis at a specific position in a DNA sequence. J Biol Chem. 1978;253:6551–60.
19. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol. 1986;51:263-273.
20. Hemsley A, Arnheim N, Toney MD, Cortopassi G, Galas J. A simple method for site-directed mutagenesis using the polymerase chain reaction. Nucleic Acids Res. 1989;17:6545–6551.
21. Prasher, DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the Aequorea victoria green-fluorescent protein. Gene. 1992;111:229–233.
22. Tsien RY .The green fluorescent protein. Annu Rev Biochem. 1998;67:509-44.
23. Saks ME, Sampson JR, Nowak MW, Kearney PC, Du F, Abelson JN, et al. An engineered Tetrahymena tRNA-Gln for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression. J Bio Chem. 1996;271:23169-75.
24. Hohsaka T, Ashizuka Y, Sisido M. Incorporation of two nonnatural amino acids into proteins through extension of the genetic code. Nucleic Acids Symp Ser. 1999;42:79-80.
25. Moore B, Persson BC, Nelson CC, Gesteland RF, Atkins JF. Quadruplet codons: implications for code expansion and the specification of translation step size. J Mol Biol. 2000;298:195-209.
26. Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature. 2010;464(7287):441-4.
27. Greiss S, Chin JW Expanding the genetic code of an animal. J Am Chem Soc. 2011;603 133:14196–14199.
28. Parrish AR, She X, Xiang Z, Coin I, Shen Z, Briggs SP, et al. Expanding the genetic code of Caenorhabditis elegans using bacterial aminoacyl-tRNA synthetase/tRNA pairs. ACS Chem Biol. 2012;7:1292–1302.
29. Bianco A, Townsley FM, Greiss S, Lang K, Chin JW. Expanding the genetic code of Drosophila melanogaster. Nat Chem Biol. 2012;8:748–750.
30. Chen Y, Ma J, Lu W, Tian W, Yuan C, Volovitch M , et

al. Heritable expansion of the genetic code in mouse and zebrafish. Cell Res. 2017;27:294.

31. Han S, Yang A, Lee S, Lee H, Park C, Park H. Expanding the genetic code of Mus musculus. Nat Commun. 2017;8:14568.

32. Srinivasan G, James CM, Krzycki JA., Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science. 2002;296:1459-62.

33. Ibba M and Soll D. Genetic code: introducing pyrrolysine. Curr Biol. 2002;12:R464-6.

34. Neumann-Staubitz P, Neumann H. The use of unnatural amino acids to study and engineer protein function. Curr Opin Struct Biol. 2016;38:119–128.

35. Liu J, Hemphill J, Samanta S, Tsang M, Deiters A. Genetic Code Expansion in Zebrafish Embryos and Its Application to Optical Control of Cell Signaling. J Am Chem Soc. 2017;139(27):9100-9103.

36. Kang JY, Kawaguchi D, Coin I, Xiang Z, O'Leary DD, Slesinger PA, et al. In Vivo Expression of a Light-Activatable Potassium Channel Using Unnatural Amino Acids. Neuron. 2013;80:358-370.

37. Ernst RJ, Krogager TP, Maywood ES, Zanchi R, Beránek V, Elliott TS, et al. Genetic code expansion in the mouse brain. Nat Chem Biol. 2016;12(10):776-778.

38. Zheng Y, Lewis Jr TL, Igo P, Polleux F, Chatterjee A. Virus-enabled optimization and delivery of the genetic machinery for efficient unnatural amino acid mutagenesis in mammalian cells and tissues. ACS Synthetic Biology. 2017;6(1):13-18.

39. Rosenthal N, Brown S. The mouse ascending: perspectives for human-disease models. Nat Cell Biol. 2007;9:993–999.

40. Schvartzman JM, Sotillo R, Benezra R. Mitotic chromosomal instability and cancer: mouse modelling of the human disease. Nat Rev Cancer. 2010;2:102–115.

41. Pless SA, Ahern CA. Unnatural amino acids as probes of ligand-receptor interactions and their conformational consequences. Annu Rev Pharmacol Toxicol. 2013;53:211-29.

42. Dougherty DA, Van Arnam EB. In vivo incorporation of non-canonical amino acids by using the chemical aminoacylation strategy: a broadly applicable mechanistic tool. Chembiochem. 2014;15:1710-20.

43. Daggett KA, Sakmar TP. Site-specific in vitro and in vivo incorporation of molecular probes to study G-protein-coupled receptors. Curr Opin Chem Biol. 2011;3:392-8.

44. Huber T, Sakmar TP. Chemical biological methods for investigating G protein-coupled receptor signaling. Chemistry & Biology. 2014;1224-1237.

45. Tian H, Fürstenberg A, and Huber T. Labeling and Single-Molecule Methods to Monitor G Protein-Coupled Receptor Dynamics. Chem Rev. 2017;186-245.

46. Tian M, Wang Q, Yuan C, Ye S. Structure and Function Studies of GPCRs by Site-Specific Incorporation of Unnatural Amino Acids. Springer International Publishing AG. 2017.

47. Deisseroth K. Optogenetics: 10 years of microbial opsins in neuroscience. Nat Neurosci. 2015;9:1213-25.

48. Miller JC1, Silverman SK, England PM, Dougherty DA, Lester HA. Flash decaging of tyrosine sidechains in an ion channel. Neuron. 1998;4:619-24.

49. Tong Y, Brandt GS, Li M, Shapovalov G, Slimko E, Kar-

schin A, et al. Tyrosine decaging leads to substantial membrane trafficking during modulation of an inward rectifier potassium channel. J Gen Physiol. 2001;117:103-18.

50. Gautier A, Nguyen DP, Lusic H, An W, Deiters A, Chin JW. Genetically encoded photocontrol of protein localization in mammalian cells. J Am Chem Soc. 2010;132(12):4086-8.

51. Gautier A, Deiters A, Chin JW. Light-activated kinases enable temporal dissection of signaling networks in living cells. J Am Chem Soc. 2011;133:2124-7.

52. Hemphill J, Borchardt EK, Brown K, Asokan A, Deiters A. Optical control of CRISPR/Cas9 gene editing. J Am Chem Soc. 2015;137:5542.

53. Uprety R, Luo J, Liu J, Naro Y, Samanta S, Deiters A.. Genetic encoding of caged cysteine and caged homocysteine in bacterial and mammalian cells. ChemBioChem. 2014;15:1793-9.

54. Ye S, Köhrer C, Huber T, Kazmi M, Sachdev P, Yan EC, et al. Site-specific incorporation of keto amino acids into functional G protein-coupled receptors using unnatural amino acid mutagenesis. J Biol Chem. 2008;283:1525-33.

55. Sakamoto K, Hayashi A, Sakamoto A, Kiga D, Nakayama H, Soma A, et al. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. Nucleic Acids Res. 2002;30:4692-4699.

56. Köhrer C, Xie L, Kellerer S, Varshney U, RajBhandary UL. Import of amber and ochre suppressor tRNAs into mammalian cells: a general approach to site-specific insertion of amino acid analogues into proteins. Proc Natl Acad Sci U S A. 2001;98:14310-5.

57. Grunbeck A, Sakmar TP. Probing G protein-coupled receptor-ligand interactions with targeted photoactivatable cross-linkers. Biochemistry. 2013;52:8625-32.

58. Grunbeck A, Huber T, Abrol R, Trzaskowski B, Goddard WB, Sakmar TP. Genetically Encoded Photo-cross-linkers Map the Binding Site of an Allosteric Drug on a G Protein-Coupled Receptor. ACS Chem Biol. 2012;7(6):967–972.

59. Ray-Saha S, Huber T, and Sakmar TP. Antibody Epitopes on G Protein-Coupled Receptors Mapped with Genetically Encoded Photoactivatable Cross-Linkers. Biochemistry. 2014;53(8):1302-1310.

60. Koole C, Reynolds CA, Mobarec JC, Hick C, Sexton PM, Sakmar TP. Genetically encoded photocross-linkers determine the biological binding site of exendin-4 peptide in the N-terminal domain of the intact human glucagon-like peptide-1 receptor (GLP-1R). J Biol Chem. 2017;292(17):7131-7144.

61. Coin I, Katritch V, Sun T, Xiang Z, Siu FY, Beyermann M, et al. Genetically encoded chemical probes in cells reveal the binding path of urocortin-I to CRF class B GPCR. Cell. 2013;155:1258-1269.

62. Rannversson H, Andersen J, Sørensen L, Bang-Andersen B, Park M, Huber M, et al. Genetically encoded photocrosslinkers locate the high-affinity binding site of antidepressant drugs in the human serotonin transporter. Nat Commun. 2016;7:11261.

63. Klippenstein V, Ghisi V, Wietstruk M, Plested AJ. Photoinactivation of glutamate receptors by genetically encoded unnatural amino acids. J Neurosci. 2014;34:980-91.

64. Ye S, Riou M, Carvalho S, Paoletti P. Expanding the genetic code in Xenopus laevis oocytes. Chembiochem. 2013;14:230-5.

65. Zhu S, Riou M, Yao CA, Carvalho S, Rodriguez PC, et al. Genetically encoding a light switch in an ionotropic glutamate receptor reveals subunit-specific interfaces. Proc Natl Acad Sci USA. 2014;111:6081-6.

66. M Tian and S Ye. Allosteric regulation in NMDA receptors revealed by the genetically encoded photo-cross-linkers. Sci Rep. 2016;6:34751.

67. Westhoff M, Murray CI, Eldstrom J, Fedida D. Photo-Cross-Linking of IKs Demonstrates State-Dependent Interactions between KCNE1 and KCNQ1. Biophys J. 2017;113(2):415-425.

68. Zaydman MA, Kasimova MA, McFarland K, Beller Z, Hou P, Kinser HE, et al. Domain–domain interactions determine the gating, permeation, pharmacology, and subunit modulation of the IKs ion channel. Elife. 2014;3:e03606.

69. Muranaka N, Hohsaka T, Sisido M. Photoswitching of peroxidase activity by position-specific incorporation of a photoisomerizable non-natural amino acid into horseradish peroxidase. FEBS Lett. 2002;510:10-2.

70. Bose M, Groff D, Xie J, Brustad E, Schultz PG. The Incorporation of a Photoisomerizable Amino Acid into Proteins in E. coli. J Am Chem Soc. 2006;128:388–389.

71. Hoppmann C, Lacey VK, Louie GV, Wei J, Noel JP, Wang L. Genetically encoding photoswitchable click amino acids in Escherichia coli and mammalian cells. Angew Chem Int Ed Engl. 2014;53:3932-3936.

72. Klippenstein V, Hoppmann C, Ye S, Wang L, Paoletti P. Optocontrol of glutamate receptor activity by single side-chain photoisomerization. Elife. 2017;6:e25808.

73. Fehrentz T, Schönberger M, Trauner D. Optochemical genetics. Angew Chem Int Ed Engl. 2011;50:12156-82.

74. Reiner A, Isacoff EY. Photoswitching of cell surface receptors using tethered ligands. Methods Mol Biol. 2014;1148:45-68.

75. Banghart MR, Volgraf M, Trauner D. Engineering light-gated ion channels. Biochemistry. 2006;45:15129-41.

76. Kramer RH, Fortin DL, Trauner D. New photochemical tools for controlling neuronal activity. Curr Opin Neurobiol. 2009;5:544-52.

77. Elsässer SJ, Ernst RJ, Walker OS, Chin JW. Genetic code expansion in stable cell lines enables encoded chromatin modification. Nat Methods. 2016;13(2):158-64.

78. Arpino JA, Baldwin AJ, McGarrity AR, Tippmann EM, Jones DD. In-frame amber stop codon replacement mutagenesis for the directed evolution of proteins containing non-canonical amino acids: identification of residues open to bio-orthogonal modification. PLoS One. 2015 May 26;10(5):e0127504.

79. Keeling KM, Xue X, Gunn G, Bedwell DM. Therapeutics based on stop codon readthrough. Annu Rev Genomics Hum Genet. 2014;371-94.