

Published in final edited form as:

Nat Chem. 2014 May ; 6(5): 393–403. doi:10.1038/nchem.1919.

## Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET

Kaihang Wang<sup>#2</sup>, Amit Sachdeva<sup>#2</sup>, Daniel J. Cox<sup>2</sup>, Nabil M. Wilf<sup>2</sup>, Kathrin Lang<sup>2</sup>, Stephen Wallace<sup>2</sup>, Ryan A. Mehl<sup>3</sup>, and Jason W. Chin<sup>2,\*</sup>

<sup>2</sup>Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK

<sup>3</sup>Oregon State University, Department of Biochemistry and Biophysics, 2011 ALS, Corvallis, OR97331

# These authors contributed equally to this work.

### Abstract

The ability to introduce different biophysical probes into defined positions in target proteins will provide powerful approaches for interrogating protein structure, function and dynamics. However, methods for site-specifically incorporating multiple distinct unnatural amino acids are hampered by their low efficiency. Here we provide a general solution to this challenge by developing an optimized orthogonal translation system that uses amber and evolved quadruplet-decoding transfer RNAs to encode numerous pairs of distinct unnatural amino acids into a single protein expressed in *Escherichia coli* with a substantial increase in efficiency over previous methods. We also provide a general strategy for labelling pairs of encoded unnatural amino acids with different probes via rapid and spontaneous reactions under physiological conditions. We demonstrate the utility of our approach by genetically directing the labelling of several pairs of sites in calmodulin with fluorophores and probing protein structure and dynamics by Förster resonance energy transfer.

### Introduction

Genetically programming the efficient site-specific incorporation of multiple unnatural amino acids into proteins will facilitate a range of applications from Förster resonance energy transfer (FRET) studies<sup>1, 2</sup> on protein conformational change and dynamics, to protein stapling<sup>3, 4</sup> for improving therapeutic proteins, and ultimately, to the encoded cellular synthesis and evolution of unnatural polymers<sup>5</sup>.

Encoding multiple distinct unnatural amino acids in cells requires (1) mutually orthogonal aminoacyl-transfer RNA (tRNA) synthetase/tRNA pairs that recognize distinct amino acids and (2) blank codons that can be assigned to new amino acids<sup>6-8</sup>. We previously

\*Correspondence: chin@mrc-lmb.cam.ac.uk.

**Author contributions:** JWC and KW conceived the project. JWC, KW and AS planned and designed experiments and wrote the manuscript, with input from other authors. RAM provided *MjTetPheRS* and compound **5** for pilot experiments. All other authors performed experiments or provided reagents.

demonstrated that the *Methanococcus janaschii* tyrosyl-tRNA synthetase (*MjTyrRS*)/*MjtRNA* pair and its active site variants, and the pyrrolysyl-tRNA synthetase (*PylRS*)/tRNA pair from certain *Methanosarcina* species and their active site variants are mutually orthogonal in their aminoacylation specificity<sup>9</sup>. We created orthogonal ribosome-mRNA pairs in which the orthogonal ribosome is specifically directed to an orthogonal message<sup>10</sup>. These ribosomes, unlike natural ribosomes, are not responsible for the synthesis of the proteome, and can be altered to perform new functions<sup>9, 11, 12</sup>. We evolved an orthogonal ribosome (ribo-Q1)<sup>9</sup> to decode quadruplet codons<sup>13, 14</sup> and amber codons, which are inefficiently decoded by natural ribosomes to provide a series of blank codons that can be used on orthogonal messages. By combining ribo-Q1 with an orthogonal message containing a quadruplet and amber codon, and *AzPheRS*\*/tRNA<sub>UCCU</sub> (a derivative of *MjTyrRS*/tRNA) and *PylRS*/tRNA<sub>CUA</sub> we were able to direct the incorporation of *p*-azidophenylalanine (**4**) and *N*6-[(2-propynyloxy)carbonyl]-L-lysine (**2**), and program the formation of a triazole crosslink<sup>9</sup>. Although ribo-Q1 provided a large increase in quadruplet-decoding efficiency, this increase was not exploited fully in the original system, in part because the *MjAzPheRS*\*/tRNA<sub>UCCU</sub> pair has suboptimal efficiency and specificity in directing the incorporation of **4**. Subsequent efforts to encode two distinct unnatural amino acids have used multiple stop codons to direct the incorporation of unnatural amino acids<sup>15-17</sup>. However, the efficiency of incorporation of amino acids in response to stop codons in natural translation is limited by competition with release factors<sup>11</sup>.

Here we evolve a series of Pyl tRNAs that provide optimal decoding of cognate quadruplet codons on ribo-Q1. We optimized orthogonal translation using the new tRNAs to provide a large increase in the efficiency of incorporating two distinct unnatural amino acids into a single polypeptide. We demonstrate the generality of our approach by encoding a matrix of unnatural amino acid pairs in proteins. Moreover, we demonstrate the first general route to the rapid and spontaneous labelling of proteins at genetically defined positions with two distinct probes under physiological conditions, and we demonstrate the utility of this approach for following conformational change in calmodulin (CaM) by FRET.

## Results

### Selection of quadruplet-decoding Pyl tRNA variants

*PylRS*, unlike *MjTyrRS* and its derivatives, does not recognize the anticodon stem loop of its cognate tRNA, Pyl tRNA<sub>CUA</sub><sup>18, 19</sup>. This may allow the anticodon stem loop of Pyl tRNA to be optimized for ribosomal decoding on ribo-Q1 without interfering with its aminoacylation by *PylRS*. We set out to select Pyl tRNA variants that are efficiently decoded on ribo-Q1. Previous work explored the selection of non-orthogonal tRNAs for decoding quadruplet codons on the natural ribosome<sup>20</sup>, and recently the selection of a tRNA, derived from Pyl tRNA, that decodes the AGGA codon on the natural ribosome has been reported<sup>21</sup>. We were interested in discovering (1) a number of extended anticodon/quadruplet codon pairs that are efficiently decoded on ribo-Q1 and (2) Pyl tRNA variants that are efficiently and specifically decoded on ribo-Q1.

We chose four quadruplet codons (AGGA, AGTA, TAGA and CTAG) as potential targets for decoding by evolved Pyl tRNA<sub>XXXX</sub> variants (in which the anticodon XXXX is

complementary to a specific codon) on ribo-Q1. AGGA was chosen because it has previously been used for unnatural amino acid incorporation<sup>9</sup>. The other codons were chosen because the two overlapping triplets embedded within the corresponding quadruplet anticodon are not recognition elements for natural synthetases in the cell (Supplementary Fig. S1), and we hypothesized that this would facilitate the selection of Pyl tRNA variants that are not mis-aminoacylated by endogenous synthetases.

To select for Pyl tRNA variants that are able to efficiently decode the selected quadruplet codons on ribo-Q1 we first created saturation mutagenesis libraries, Pyl tRNA(N8)<sub>XXXX</sub>, in which the sequence of eight nucleotides in the anticodon stem loop was varied to all possible combinations (Fig. 1a and 1b). Nucleotides 30-33 and 37-40 were randomized for tRNA(N8)<sub>U<sub>CUA</sub></sub> and 29-32 and 38-41 were randomized for Pyl tRNA(N8)<sub>U<sub>CCU</sub></sub>, Pyl tRNA(N8)<sub>U<sub>ACU</sub></sub>, and Pyl tRNA(N8)<sub>C<sub>UAG</sub></sub> (Fig. 1b). Each library contained at least 10<sup>8</sup> members, ensuring greater than 99.9% coverage.

We performed a two-step selection (Fig. 1c) to discover the desired tRNAs: a negative selection to remove Pyl tRNA(N8)<sub>XXXX</sub> library members that are substrates for endogenous synthetases (because all aminoacyl-tRNA synthetases, except for AlaRS, LeuRS and SerRS, recognize the anticodon stem loop of their cognate tRNA in *E. coli*<sup>22</sup>), and a positive selection to select for library members that are aminoacylated with an unnatural amino acid by PylRS and efficiently decoded on ribo-Q1 in response to a cognate quadruplet codon in the orthogonal mRNA.

To remove Pyl tRNA(N8)<sub>XXXX</sub> library members that are both substrates for endogenous synthetase and decoded on ribo-Q1 we created four negative selection reporter constructs (*O-barnase*<sub>3XXXX+45XXXX</sub>), each of which contain a barnase gene that is specifically translated by ribo-Q1 (Fig. 1c). Each of the four barnase genes contains two in-frame quadruplet codons (XXXX), where XXXX is the Watson-Crick complementary sequence to the specific anticodon of a tRNA(N8)<sub>XXXX</sub>. Cells contained ribo-Q1, *O-barnase*<sub>3XXXX+45XXXX</sub>, and PylRS/Pyl tRNA(N8)<sub>XXXX</sub>. If a Pyl tRNA(N8)<sub>XXXX</sub> library member is aminoacylated with a natural amino acid by an endogenous synthetase and can be efficiently decoded on *O-barnase*<sub>3XXXX+45XXXX</sub> by ribo-Q1, then full-length barnase will be produced, leading to cell death. Pyl tRNA(N8)<sub>XXXX</sub> library members that survive the selection are either not substrates for native synthetases and/or they are unable to decode their quadruplet codon on ribo-Q1.

To select for functional PylRS/Pyl tRNA(N8)<sub>XXXX</sub> pairs from the Pyl tRNA(N8)<sub>XXXX</sub> library members that survive the negative selection we created positive selection reporters that contains an orthogonal chloramphenicol acetyl transferase (*cat*) gene with an in-frame quadruplet codon (*O-cat*<sub>111XXXX</sub>)<sup>9, 20</sup> (Fig. 1c). Cells contained ribo-Q1, *O-cat*<sub>111XXXX</sub>, and PylRS/Pyl tRNA(N8)<sub>XXXX</sub>. In the presence of **1**, PylRS aminoacylated correctly transcribed, processed and folded Pyl tRNA(N8)<sub>XXXX</sub> library members with **1**, and the quadruplet codon in *cat* is decoded by ribo-Q1. The two-step selection should lead to chloramphenicol resistance only if Pyl tRNA(N8)<sub>XXXX</sub> is efficiently and selectively aminoacylated with **1** by PylRS and is efficiently decoded in response to the cognate quadruplet codon by ribo-Q1. Importantly, unlike selections for quadruplet-decoding tRNAs

on natural ribosomes, selection of quadruplet-decoding tRNAs directly on ribo-Q1 provides the possibility of selecting quadruplet-decoding tRNAs that are specific for the decoding centre of ribo-Q1, such tRNAs would further enhance the specificity of orthogonal translation.

We sequenced and phenotyped clones from each selection (Supplementary Fig. S2 and S3). On the basis of these initial experiments, the PylRS/evolved Pyl tRNA<sub>XXXX</sub> pairs giving the highest level of chloramphenicol resistance in cells containing ribo-Q1, *O-cat*<sub>111XXXX</sub>, in presence of **1** and lowest chloramphenicol resistance in the absence of **1** were characterized further.

The evolved Pyl tRNA<sub>UCUA</sub> contains a G–U pair in place of the parental G30–C40 pair, a G–C pair in place of the A31–U39 pair, and an A38U mutation. Although this library randomised eight positions, nucleotides at three positions (C32, U33 and A37), are identical to those found in the parent Pyl tRNA<sub>CUA</sub> (Fig. 2a). The evolved Pyl tRNA<sub>UCCU</sub>, contains a C29–G41 pair in place of the parental G29–C41 pair. G30C and C40A replace the parental G30–C40 pair, and the evolved sequence contains an A31U mutation. C32, A38 and U39 are conserved between the parental Pyl tRNA<sub>CUA</sub> and the evolved Pyl tRNA<sub>UCCU</sub>, these mutations are predicted to expand the anticodon loop and shorten the anticodon stem (Fig. 2a). In the evolved Pyl tRNA<sub>UACU</sub>, U29 and U41 replace the parental G29–C41 pair, a G31–C39 pair replaces the parental A31–U39 pair and A38 is mutated to U (Fig. 2a). In the evolved Pyl tRNA<sub>CUAG</sub>, C31 and C39 replace the A31–U39 base pair. These mutations are predicted to expand the anticodon loop (Fig. 2a).

### Enhanced unnatural amino acids incorporation at quadruplet codons

We next demonstrated the improved quadruplet decoding efficiency and specificity of the evolved Pyl tRNA<sub>XXXX</sub> variants, compared to the transplant Pyl tRNA<sub>XXXX</sub>, in which an extended quadruplet anticodon replaces CUA. Cells containing ribo-Q1, *O-cat*<sub>111XXXX</sub>, PylRS, **1**, and either evolved Pyl tRNA<sub>XXXX</sub> or transplant Pyl tRNA<sub>XXXX</sub> were plated on increasing concentrations of chloramphenicol (Fig. 2b) to assess the quadruplet decoding efficiency of each tRNA.

Cells that contained the evolved Pyl tRNA<sub>UCUA</sub> were able to survive on chloramphenicol concentrations of up to 450 µg ml<sup>-1</sup>, which is substantially greater than the 75 µg ml<sup>-1</sup> survival conferred by the transplant Pyl tRNA<sub>UCUA</sub>. This indicates that the evolved Pyl tRNA<sub>UCUA</sub> is more efficient at decoding the quadruplet codon TAGA than the transplant Pyl tRNA<sub>UCUA</sub> (Fig. 2b). Experiments in which the substrate for PylRS, **1**, is omitted (Fig. 2b) demonstrate that the evolved Pyl tRNA<sub>UCUA</sub> is specifically aminoacylated by PylRS and not by endogenous aminoacyl-tRNA synthetases.

The transplant Pyl tRNA<sub>UCCU</sub> and evolved Pyl tRNA<sub>UCCU</sub> both conferred chloramphenicol resistance up to 500 µg ml<sup>-1</sup> when cells were grown in the presence of **1**. However, the transplant Pyl tRNA<sub>UCCU</sub>, unlike the evolved Pyl tRNA<sub>UCCU</sub> conferred significant chloramphenicol resistance in the absence of **1** (Fig. 2b). This indicated that although both the transplant Pyl tRNA<sub>UCCU</sub> and evolved Pyl tRNA<sub>UCCU</sub> can decode the AGGA codon,

only the evolved Pyl tRNA<sub>UCCU</sub> is specifically recognized by PylRS and not by endogenous synthetases.

The transplant Pyl tRNA<sub>UACU</sub> and Pyl tRNA<sub>CUAG</sub> conferred minimal chloramphenicol resistance; in contrast the evolved Pyl tRNA<sub>UACU</sub> and Pyl tRNA<sub>CUAG</sub> were able to decode the quadruplet codons AGTA and CTAG and confer enhanced chloramphenicol resistance. Experiments in which **1** is omitted (Fig. 2b) demonstrate that the evolved Pyl tRNA<sub>UACA</sub> and Pyl tRNA<sub>CUAG</sub> are specifically aminoacylated by PylRS and not by endogenous aminoacyl-tRNA synthetases.

We next demonstrated that the improved quadruplet-decoding efficiency of the evolved PylRS/Pyl tRNA<sub>XXXX</sub> pairs on ribo-Q1 translates into an increased production of recombinant protein bearing an unnatural amino acid. To provide a benchmark for the incorporation achieved using the evolved Pyl tRNA<sub>XXXX</sub> systems, we expressed and purified a fusion protein of glutathione-S-transferase and calmodulin (GST-CaM) site-specifically incorporating **4** from cells containing ribo-Q1, *O-gst-cam*<sub>IAGGA</sub> (a fusion gene between *glutathione-S-transferase* (*gst*) and *cam* in which the first codon of *cam* is replaced with an AGGA codon), (Supplementary Fig. S4), and *MjAzPheRS*\*/tRNA<sub>UCCU</sub> (a synthetase/tRNA pair previously developed for incorporating **4** in response to AGGA codons)<sup>9</sup>. This led to approximately 13% full-length protein upon addition of **4**, with approximately 7% full-length protein produced in the absence of **4** (Supplementary Fig. S4).

To compare the incorporation achieved with the evolved Pyl tRNA<sub>XXXX</sub> and the corresponding transplant Pyl tRNA<sub>XXXX</sub>, we expressed GST-CaM from cells containing ribo-Q1, *O-gst-cam*<sub>IXXXX</sub> (XXXX is complementary to the tRNA anticodon used), amino acid **1**, and either transplant Pyl tRNA<sub>XXXX</sub> or evolved Pyl tRNA<sub>XXXX</sub> (Supplementary Fig. S4). The evolved Pyl tRNA<sub>UCUA</sub>, evolved Pyl tRNA<sub>UACU</sub> and evolved Pyl tRNA<sub>CUAG</sub> substantially increased the efficiency of incorporation of **1** in response to TAGA, AGTA, CTAG codons respectively (Supplementary Fig. S4).

Both the evolved and the transplant Pyl tRNA<sub>UCCU</sub> led to comparable read-through of the AGGA codon in *O-gst-cam*<sub>IAGGA</sub> upon addition of **1** (Supplementary Fig. S4). However, the transplant Pyl tRNA<sub>UCCU</sub> was substantially mis-aminoacylated in the absence of **1** (Supplementary Fig. 4). This is consistent with the phenotypic data (Fig. 2b) and may demonstrate the advantage of the negative selection step in selecting synthetase/tRNA pairs for unnatural amino acid dependent incorporation at quadruplet codons.

To demonstrate that a range of unnatural amino acids can be incorporated in response to quadruplet codons using the evolved tRNAs, we compared the efficiency of incorporation using chemically distinct unnatural amino acids (Fig. 2c and Supplementary Fig. 4). The efficiencies of incorporating **1**, **2** and **3** in response to quadruplet codons TAGA, AGGA, AGTA, and CTAG using their cognate evolved Pyl tRNA<sub>XXXX</sub> ranged from approximately 27% to 62%, and were all substantially greater than the unnatural amino acid incorporation efficiency using the *MjAzPhe RS*\*/tRNA<sub>UCCU</sub> pair<sup>9</sup> (Supplementary Fig. S4). Control experiments demonstrated that the unnatural amino acid incorporation efficiency was improved on providing multiple copies of PylRS (Supplementary Fig. S5), and was robust in

different strains of *E. coli* (data not shown). The efficiency and specificity of incorporating **1** in response to the quadruplet codons using our evolved quadruplet-decoding systems is greater than the efficiency and specificity we observe when using PylRS/tRNA anticodon variants and wild-type ribosome to decode the ochre (UAA), opal (UGA), or AGGA codons<sup>15, 17, 21</sup> at the same position in GST-CaM (Supplementary Figs. S4 & S5)

### Incorporating a matrix of unnatural amino acid pairs into proteins

Next, we optimized the incorporation of two distinct unnatural amino acids into proteins via ribo-Q1. Our starting system contains four plasmids and used ribo-Q1, *MjAzPheRS*\*/tRNA<sub>UCCU</sub>, and PylRS/tRNA<sub>CUA</sub> to read AGGA and amber codons on orthogonal messages (Fig. 3a and Supplementary Figs. S6 & S7).

To improve the system, we first created a three-plasmid system using the same elements (Fig. 3b, c and Supplementary Fig S6, & S7). This increased the fraction of full-length protein produced, but also resulted in a high level of unnatural amino acid independent read-through of the AGGA codon. To increase the fidelity and efficiency of the system for incorporating two distinct unnatural amino acids we used the evolved Pyl tRNA<sub>UACU</sub>, developed in this work, in combination with *MjAzPheRS*/tRNA<sub>CUA</sub><sup>23</sup>. This led to a further substantial increase in the efficiency and specificity of the system, and double incorporation efficiencies in excess of 20% (Figure 3d, and Supplementary Fig. S6, & S7). The efficiency of double incorporation using evolved Pyl tRNA<sub>UCCU</sub> or evolved Pyl tRNA<sub>CUAG</sub> was comparable to the efficiency in experiments using evolved Pyl tRNA<sub>UACU</sub> (Supplementary Fig. S7). By combining the *MjAzPheRS*\*/tRNA<sub>UCCU</sub> pair and the PylRS/tRNA<sub>UACU</sub> pair with ribo-Q1, we directed the cellular incorporation of multiple distinct unnatural amino acids in response to multiple distinct quadruplet codons (Fig. 3e and Supplementary Fig. S7).

To demonstrate the generality of the approach for incorporating chemically diverse pairs of unnatural amino acids into proteins, we incorporated every combination of three distinct unnatural PylRS substrates and four distinct substrates for *MjTyrRS* variants (Figure 4 and Supplementary Fig. S8) using the orthogonal translation system described in Fig. 3d. The resulting matrix of 12 distinct pairs of unnatural amino acids, including unique chemical handles (azides<sup>23</sup>, alkenes<sup>24</sup>, alkynes<sup>25, 26</sup>, tetrazines<sup>27</sup>) and a photocrosslinker (benzophenone<sup>28</sup>) were efficiently and specifically incorporated into GST-CaM, with full length protein yields from 1.5-2 mg L<sup>-1</sup> of culture (for comparison the yield of wild-type GST-CaM was approximately 4 mg L<sup>-1</sup> of culture). There is some read-through of the codons used to direct unnatural amino acid incorporation in the absence of unnatural amino acid, as previously described<sup>9, 11</sup>. However, electrospray ionization mass spectrometry (ESI-MS) and matrix-assisted laser desorption mass spectrometry (MALDI-MS) (Supplementary Fig. S8) conclusively demonstrated the genetically programmed incorporation of both added unnatural amino acids, without measurable mis-incorporation of natural amino acids.

We compared the efficiency of incorporating **1** and **4** in response to two stop codons on the natural ribosome<sup>15, 17</sup> and to (using an *MjTyrRS* variant specific for **4** (ref. 23) PylRS/tRNA<sub>UUA</sub> and PylRS/tRNA<sub>UCA</sub> to direct the incorporation of **1** in response the ochre or opal codon) to our system. Double incorporation in response to two stop codons was not

detectable in GST-CaM (Supplementary Fig. S7), consistent with the low efficiency with which both PylRS/tRNA<sub>UUA</sub> and PylRS/tRNA<sub>UCA</sub> decode their cognate stop codons (Supplementary Fig. S5). Comparison of the experiments using two stop codons with the experiments using our optimized system with amber and quadruplet codons decoded on the evolved orthogonal ribosome (Supplementary Fig. S7) reveals the advantages of the approach reported herein in GST-CaM.

### Programming mutually orthogonal chemical handles into proteins

The ability to program multiple unnatural amino acids into proteins raises the possibility of programming pairs of mutually orthogonal functional groups into proteins that can be specifically and quantitatively labelled with a range of probes under conditions that maintain the native fold of proteins (physiological pH, temperature and pressure). This would facilitate a range of applications, including FRET to study protein structure, conformation and dynamics<sup>1</sup>. Efforts to double-label proteins at genetically encoded unnatural amino acids have relied on encoding ketones to react with alpha-effect nucleophiles, and encoding azides to react with strained alkynes<sup>16, 17</sup>. Unfortunately, ketone labelling is very slow (with rate constants (*r*) of approximately 10<sup>-4</sup> M<sup>-1</sup>s<sup>-1</sup>) and commonly requires low pH, which makes it incompatible with the folding and stability of many proteins. Moreover genetically encoded azides are susceptible to *in vivo* reduction (Figure 4)<sup>29</sup>, which makes their quantitative labelling problematic.

Inverse electron-demand Diels-Alder reactions allow the spontaneous and rapid coupling of a number of strained alkenes or alkynes and tetrazines<sup>30, 31</sup>, and provide a promising route to rapid and quantitative labelling of proteins<sup>27, 32-36</sup>. The rates of these reactions depend on the tetrazine substituents and the strain/sterics of the alkene or alkyne<sup>37, 38</sup>. We reasoned that it may be possible to tune the reactivity of tetrazines and strained alkenes in encoded amino acids such that they do not react with each other in the protein, but can be specifically and rapidly labelled with added probes.

As a model for reaction between **5** and **8**, we investigated the rate of reaction between 5-norbornene-2-ol (**8a**) and the t-butyloxycarbonyl (Boc) protected amino acid containing a tetrazine (**5a**) that is deactivated in inverse electron-demand Diels-Alder reactions by virtue of electron donation from the aniline moiety (Supplementary Fig. S9). The reaction is very slow, and we estimate that the rate constant for this reaction is approximately 10<sup>-5</sup> M<sup>-1</sup>s<sup>-1</sup> (Supplementary Fig. S9). To demonstrate that we can genetically encode **5** and **8** in a single polypeptide, we grew cells containing *O-gst-cam<sub>1TAG+149AGTA</sub>*, ribo-Q1, *MjTetPheRS* (an *MjTyrRS* variant specific for **5**)/tRNA<sub>CUA</sub>, NorKRS (a PylRS variant<sup>39</sup> for the incorporation of **8**)/evolved Pyl tRNA<sub>UACU</sub> in the presence of **5** and **8**. The production of full-length protein (CaM**5**<sub>1</sub>-**8**<sub>149</sub>) was dependent on the addition of both **5** and **8** (Fig. 5a).

CaM**5**<sub>1</sub>-**8**<sub>149</sub> migrated as the linear protein in SDS-PAGE (Fig. 5a) suggesting that **5** and **8** do not react, even when they are placed in proximity by the protein structure. The incorporation of **5** and **8** in the protein without cyclization was further confirmed by ESI-MS (Supplementary Fig. S10). These results demonstrate that **5** and **8** do not react even when placed at positions in a protein that facilitate the proximity-accelerated Cu (I) catalysed azide-alkyne cycloaddition between encoded amino acids (Fig. 5a), which is consistent with

the rate of reaction between **5** and **8** being orders of magnitude slower than rates reported for Cu(I)-catalysed azide-alkyne cycloadditions<sup>40</sup>.

To demonstrate that amino acid **5a** can be conjugated efficiently to bicyclononyne probes, we investigated the rate of the reaction between **5a** and **10a** (Fig. 5b). This reaction proceeded with a rate constant of  $0.314 \pm 0.010 \text{ M}^{-1}\text{s}^{-1}$  in MeOH/H<sub>2</sub>O (55/45) at room temperature (Fig. 5b and Supplementary Fig. S9). Previously we have demonstrated that the rate constant for the reaction of **8a** with the activated tetrazine **9a** (Fig. 5b) is  $0.94 \pm 0.008 \text{ M}^{-1}\text{s}^{-1}$  in MeOH/H<sub>2</sub>O (5/95) at room temperature (for comparison this rate constant is  $0.47 \pm 0.070 \text{ M}^{-1}\text{s}^{-1}$  in MeOH/H<sub>2</sub>O (55/45))<sup>33</sup>. These experiments demonstrate that the tetrazine and norbornene functionalities embedded in **5** and **8** can be labelled with strained alkynes and appropriately activated tetrazines, respectively.

We next synthesized BODIPY-TMR-X bicyclononyne conjugate **10** and demonstrated that it can be used to label **5**- bearing CaM. The on-protein rate constant for this reaction is  $1.00 \pm 0.13 \text{ M}^{-1}\text{s}^{-1}$ . ESI-MS demonstrates the quantitative single labelling of the protein (Fig. 5c). Similarly, we synthesized BODIPY-FL tetrazine conjugate **9** and demonstrated that it can be used to rapidly label **8**- bearing CaM. The on-protein rate constant for this reaction is  $1.22 \pm 0.28 \text{ M}^{-1}\text{s}^{-1}$ . ESI-MS demonstrated the quantitative single labelling of the protein (Fig. 5d). Control experiments demonstrate that bearing **8**-CaM is not labelled with **10**, and bearing **5**-CaM is not labelled with **9**. These experiments demonstrate that the rate constants measured for small-molecule labelling translate into efficient protein labelling and that the labelling reagents are specific for their targeted unnatural amino acid in proteins. The rate constants for the reaction between **5** and **10**, and between **8** and **9** are two to four orders of magnitude faster than ketone labelling reactions, label metabolically stable unnatural amino acids at physiological pH, do not use transition-metal catalysis, and generate nitrogen gas as the only by-product, making them compatible with most proteins.

### Site-specific dual labelling of proteins under physiological conditions

FRET probes have been introduced via *in vitro* translation<sup>2</sup>, which allows access to small amounts of protein labelled with particular fluorophores. Native chemical ligation has been used to introduce FRET probes into proteins<sup>41</sup>, and this has been combined with unnatural amino acid mutagenesis to introduce a limited range of short-wavelength probes<sup>42</sup>. These approaches generally limit the introduction of at least one of the probes to the termini of proteins.

We demonstrated that the reactions we have characterized can be used to rapidly and spontaneously label proteins with two distinct fluorophores for FRET studies (Fig. 6). We chose BODIPY-FL (donor) and BODIPY-TMR-X (acceptor) as FRET probes because of the significant overlap between their emission and excitation spectra. We labelled **8** at position 149 in CaM**5**<sub>1</sub>-**8**<sub>149</sub> with a BODIPY-FL activated tetrazine conjugate **9**, and labelled **5** at position 1 in CaM**5**<sub>1</sub>-**8**<sub>149</sub> with a BODIPY-TMR-X bicyclononyne conjugate **10**, to create a doubly labelled protein (CaM**5**-**10**<sub>1</sub>-**8**-**9**<sub>149</sub>). The labelling with each fluorophore was quantitative as judged by SDS-PAGE, fluorescence gel imaging and ESI-MS (Supplementary Fig. S10).



The fluorescence emission spectra of CaM5-10<sub>1</sub>-8-9<sub>149</sub> shows two distinct peaks when the sample is excited at 485 nm: one peak corresponds to emission from BODIPY-FL (donor, 515 nm), and the other peak corresponds to emission from BODIPY-TMR-X (acceptor, 570 nm) (Supplementary Fig. S10). In contrast, the fluorescence spectra of singly labelled protein with either BODIPY-FL (donor) at position 149 (CaM5<sub>1</sub>-8-9<sub>149</sub>) or BODIPY-TMR-X (acceptor) at position 1 (CaM5-10<sub>1</sub>-8<sub>149</sub>) shows minimal emission at 570 nm when the samples are excited at 485 nm (Supplementary Fig. S10). These data are consistent with FRET between the donor and acceptor fluorophores in CaM5-10<sub>1</sub>-8-9<sub>149</sub>.

### Using FRET to follow protein conformational change

The observed FRET signal for CaM5-10<sub>1</sub>-8-9<sub>149</sub> is sensitive to increasing concentrations of urea (Fig. 6c and Supplementary Fig. S10); the donor fluorescence signal increases and the acceptor fluorescence signal decreases as the urea concentration is increased. The observed decrease in FRET is consistent with the fluorophores moving further apart in the denatured state and this experiment demonstrates that protein conformational changes can be followed with our protein-labelling strategy.

CaM senses calcium in the cell and modifies its interactions with protein-binding partners<sup>43</sup>. CaM mediates many important processes, including short- and long-term memory, and the immune response<sup>44</sup>. CaM contains four EF-hand motifs, each of which binds one calcium ion<sup>45, 46</sup>. There are two EF-hand motifs in the N-terminal domain and two EF-hand motifs in the C-terminal domain<sup>45, 46</sup>. The two EF-hands within each domain bind calcium ions cooperatively, although the N- and C-terminal domains bind calcium independently<sup>47</sup>.

Structures of CaM have been solved in the presence of four calcium ions and in the absence of calcium ions<sup>46, 48, 49</sup>. Two distinct structures have been reported in the presence of four calcium ions<sup>45, 46</sup>, and single molecule solution measurements demonstrate that the protein is highly dynamic<sup>50</sup>. Moreover, much less is known about the structure of CaM with one, two or three calcium ions bound, though intermediate states of calcium binding are physiologically relevant in modulating the selectivity of CaM's interactions with binding partners and signal propagation as a function of cellular calcium.

Although cellular calcium sensors based on FRET between fluorescent protein fusions, such as Chameleon, take advantage of global structural transitions in CaM that result from binding four calcium ions and movements of whole domains with respect to each other, it has been challenging to observe directly the local conformational change within domains as a function of calcium binding at individual sites because it is challenging to quantitatively install pairs of small molecule fluorophores at defined sites in the protein. These local changes are important because they provide the molecular basis for understanding cooperative binding of calcium to CaM and because of their biological significance.

To investigate local conformational changes in the N-terminal domain of CaM as a function of calcium concentration, we labelled CaM5<sub>1</sub>-8<sub>40</sub> (Fig. 6b & Supplementary Fig. S11) with **9** and **10**. Quantitative labelling of CaM was confirmed by fluorescence gel imaging and ESI-MS (Fig. 6b & Supplementary Fig. S11). The fluorescence spectra of the doubly labelled protein (CaM5-10<sub>1</sub>-8-9<sub>40</sub>) and singly labelled proteins (CaM5-10<sub>1</sub>-8<sub>40</sub> and

CaM5<sub>1-8-9</sub><sub>40</sub>) are consistent with FRET between the donor and acceptor fluorophores in CaM5-10<sub>1-8-9</sub><sub>40</sub> (Supplementary Fig. S11).

We observe two discrete transitions of comparable magnitude in the FRET spectra of CaM5-10<sub>1-8-9</sub><sub>40</sub> as a function of calcium concentration (Fig. 6d, and Supplementary Fig. S11). The first transition ( $K_1 = 1.125 \pm 0.34 \times 10^{-6}$  M) is close to the binding constants reported for the first three calcium ions binding to CaM in recent isothermal titration calorimetry (ITC) measurements<sup>51</sup> ( $K_{D1} = 3.69 \pm 0.53 \times 10^{-6}$  M,  $K_{D2} = 7.09 \pm 1.206 \times 10^{-5}$ ,  $K_{D3} = 5.21 \pm 0.841 \times 10^{-6}$ ) and within the range reported in the literature<sup>52-54</sup>. As calcium binding at each EF-hand motif within the N- or C-terminal domain of CaM is cooperative, but binding of calcium to the distinct domains of CaM is thermodynamically independent<sup>47</sup>, it seems probable that the first transition we observe with probes in the N-terminal domain results from the binding of the third calcium ion to CaM: the first binding event in the N-terminal domain. However, we cannot rule out contributions to the FRET change we observe from conformational changes in the N-terminal domain that result from calcium binding to the C-terminal EF hand motifs of CaM ( $K_{D1}$  and  $K_{D2}$ ). The first conformational change of the N-terminal domain may set up the protein to bind the fourth calcium ion.

The second transition ( $K_2 = 1.711 \pm 0.659 \times 10^{-3}$  M) is close to binding constants reported recently for binding the fourth calcium ion to CaM (ITC study,  $K_{D4} = 0.44 \pm 0.0465 \times 10^{-3}$  M), and the range of literature values for  $K_{D4}$ <sup>52-54</sup>, which suggests that this conformational change may result from binding the fourth and final calcium ion<sup>51</sup>.

Our data report on local conformational changes within a domain as a result of calcium-binding events at distinct sites, and allow us to observe discrete structural transitions as a result of sequential ligand-binding events at distinct sites. The FRET changes we observe are reproducibly small (Supplementary Fig. S11), consistent with one of the two X-ray structures of four calcium ions bound to CaM<sup>45, 46</sup>. However, FRET measurements depend on both probe separation and assumptions about probe orientation, our data do not distinguish unambiguously between the divergent structures of CaM bound to four calcium ions, which single-molecule studies demonstrate from part of an ensemble of structures in solution<sup>50</sup>.

## Discussion

We evolved a series of quadruplet decoding Pyl tRNA<sub>XXXX</sub> that efficiently and specifically read their cognate quadruplet codons on messages decoded by ribo-Q1. In combination with PylRS or its evolved variants, these tRNAs direct the efficient and specific incorporation of diverse unnatural amino acids into proteins. We created an optimized orthogonal translation system that takes advantage of the new quadruplet-decoding tRNAs and directs the efficient incorporation of a matrix of pairs of unnatural amino acids in *E. coli*. By carefully characterizing the kinetics of inverse electron-demand Diels-Alder reactions we designed a strategy for encoding a tetrazine and a norbornene that do not react with each other in proteins, but can be labelled rapidly with complementary bicyclononynes and electron-deficient tetrazines. This provides the first modular route to install diverse pairs of probes into proteins at essentially any genetically controlled pair of sites in proteins at physiological

temperature, pressure and pH. The reactions proceed spontaneously without generating toxic side products or using toxic catalysts. We demonstrate the utility of our strategy by site-specifically labelling CaM with FRET probes and reveal that this allows local conformational changes within a domain that result from sequential ligand-binding events at distinct sites to be observed. The modular approach we have developed will find utility in labelling proteins with a range of fluorophores, both for ensemble and single-molecule FRET experiments and for other biophysical measurements. Future work will focus on further extending and applying the approaches we have developed to address outstanding challenges in defining the molecular changes that underpin ligand binding, cooperativity, allostery and catalysis.

## Methods

Protocols for chemical synthesis of amino acids, and fluorescent labelling probes as well as protocols for cloning, protein expression, protein labelling, mass spectrometry, determination of rate constants, and fluorescence measurements are provided in the Supplementary material along with detailed descriptions of all experimental procedures. The PylRS/tRNA pairs are derived from *M.barkeri* unless otherwise stated.

### Selection of Pyl tRNA(N8)<sub>XXXX</sub> for enhanced quadruplet decoding in the ribo-Q1 orthogonal mRNA system

We transformed each PyltRNA(N8)<sub>XXXX</sub> library into GeneHogs *E. coli* cells bearing the matching *O-barnase*<sub>3XXXX+45XXXX</sub> plasmid and RSF ribo-Q1 plasmid. The transformed cells were recovered in 1 ml SOB medium containing 2% glucose for 1 hour at 37°C. The transformation was used to inoculate 200 ml of LB-GKST (LB media with 2% glucose, 12.5 µg ml<sup>-1</sup> kanamycin, 18.8 µg ml<sup>-1</sup> spectinomycin, and 6.25 µg ml<sup>-1</sup> tetracycline) and incubated overnight (37°C, 250 rpm, 16 h). The overnight culture (2ml) was pelleted by centrifugation (3,000g) and washed with LB to remove glucose, prior to diluting to OD<sub>600</sub>=0.1 in LB-KST (LB media with 12.5 µg ml<sup>-1</sup> kanamycin, 18.8 µg ml<sup>-1</sup> spectinomycin, and 6.25 µg ml<sup>-1</sup> tetracycline), and incubated (37°C, 250 rpm, 2-3 h) until OD<sub>600</sub> reached 0.5. Isopropyl β-D-1thiogalactopyranoside (IPTG (1 mM final concentration)) was added to induce the expression of ribo-Q1 ribosomal RNA (rRNA) at OD<sub>600</sub>≈0.5. The culture was incubated (37°C, 250 rpm) until the OD<sub>600</sub> reached approximately 2. Aliquots (500 µl) of the culture were serially diluted and plated on LB-KST agar plate supplemented with 1 mM IPTG. The plates were incubated at 37°C for 16 hours. Surviving colonies were scraped from the plate and plasmid DNA was extracted by Qiagen Miniprep Kit to make the post-negative selection pool.

The negative selection was followed by a positive selection. The post negative selection Pyl tRNA(N8)<sub>XXXX</sub> pools were cleansed of non pCDF PylRS/tRNA<sub>XXXX</sub> plasmids by digestion with SacII to remove the *O-barnase*<sub>3XXXX+45XXXX</sub> and pRSF ribo-Q1 plasmid. The cleansed post-negative selection Pyl tRNA(N8)<sub>XXXX</sub> pools were electroporated into GeneHog cells bearing the matching *O-cat*<sub>111XXXX</sub> and pRSF ribo-Q1 plasmids. The transformed cells were recovered in 1 ml SOB medium containing 2% glucose for one hour at 37°C. The transformation was used to inoculate 200 ml of LB-GKST and incubated overnight (37°C,

250 rpm, 16 h). The overnight culture (2 ml) was pelleted by centrifugation (3,000g) and washed with LB to remove glucose, prior to diluting to  $OD_{600}=0.1$  in LB-KST, and incubated (37°C, 250 rpm, 2-3 h) until  $OD_{600}$  reached 0.5. IPTG (1 mM final concentration) and **1** (8 mM final concentration) were added to induce the expression of ribo-Q1 rRNA and facilitate aminoacylation of Pyl tRNA(N8)<sub>XXXX</sub> library members by PylRS at  $OD_{600}\approx 0.5$ . The culture was incubated (37°C, 250 rpm) until the  $OD_{600}$  reached approximately 2. Aliquots (500  $\mu$ l) of the culture were serially diluted and plated on LB-KST agar plates supplemented with 1mM IPTG, 8mM **1**, and chloramphenicol of different concentrations (100  $\mu$ g ml<sup>-1</sup>, 150  $\mu$ g ml<sup>-1</sup>, 200  $\mu$ g ml<sup>-1</sup>). The plates were incubated at 37°C for 40 hours.

To separate the selected pCDF PylRS/tRNA(N8)<sub>XXXX</sub> plasmid from the *O-cat*<sub>111XXXX</sub> and pRSF ribo-Q1 plasmids, total plasmid DNA from selected clones was purified, digested with SacII, transformed into DH10B cells and plated on LB agar supplemented with spectinomycin (75  $\mu$ g ml<sup>-1</sup>). Individual colonies from the spectinomycin plates were incubated overnight in LB media supplemented with spectinomycin (75  $\mu$ g ml<sup>-1</sup>) and replica plated onto three sets of LB agar plates supplemented with spectinomycin (75  $\mu$ g ml<sup>-1</sup>), or tetracycline (25  $\mu$ g ml<sup>-1</sup>), or kanamycin (50  $\mu$ g ml<sup>-1</sup>). After 16 h, colonies were observed only on spectinomycin plates but not on tetracycline or kanamycin plates. The DNA from separated pCDF PylRS/tRNA(N8)<sub>XXXX</sub> clones was purified, confirmed by restriction digestion and DNA sequencing.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

Medical Research Council (U105181009, UD99999908). We thank Sew-Peak Chew (MRC-LMB Mass Spectrometry) for obtaining MALDI data.

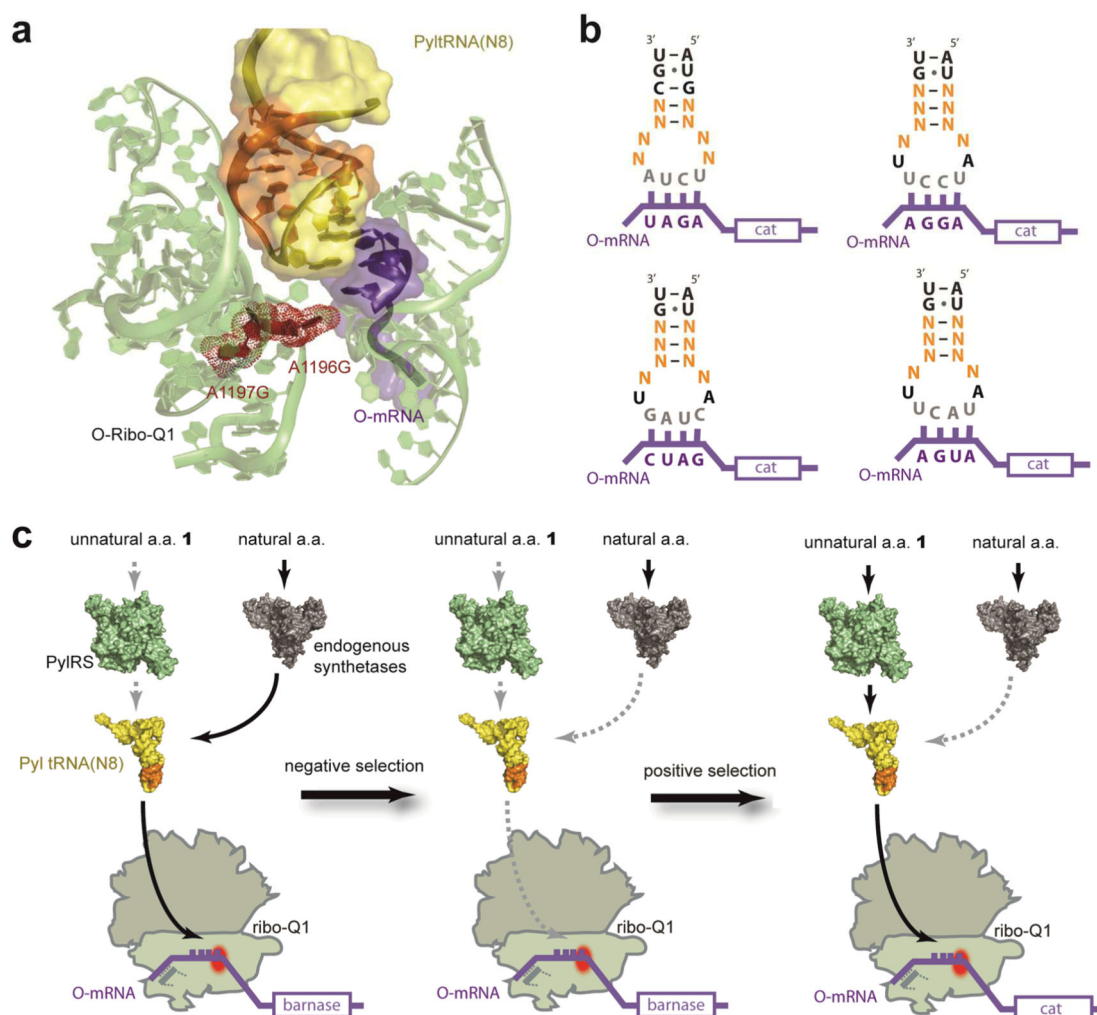
## References

1. Zhang J, Campbell RE, Ting AY, Tsien RY. Creating new fluorescent probes for cell biology. *Nature Reviews Molecular Cell Biology*. 2002; 3:906–918.
2. Kajihara D, et al. FRET analysis of protein conformational change through position-specific incorporation of fluorescent amino acids. *Nat Methods*. 2006; 3:923–929. [PubMed: 17060916]
3. Scott CP, Abel-Santos E, Wall M, Wahnon DC, Benkovic SJ. Production of cyclic peptides and proteins in vivo. *Proc Natl Acad Sci U S A*. 1999; 96:13638–13643. [PubMed: 10570125]
4. Li P, Roller PP. Cyclization strategies in peptide derived drug design. *Curr Top Med Chem*. 2002; 2:325–341. [PubMed: 11944823]
5. Wang K, Schmied WH, Chin JW. Reprogramming the genetic code: from triplet to quadruplet codes. *Angew Chem Int Ed Engl*. 2012; 51:2288–2297. [PubMed: 22262408]
6. Davis L, Chin JW. Designer proteins: applications of genetic code expansion in cell biology. *Nature Publishing Group*. 2012; 13:168–182.
7. Chin JW. Molecular biology. Reprogramming the genetic code. *Science*. 2012; 336:428–429.
8. Chin JW. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu Rev Biochem*. 2014; 83 doi: 10.1146/annurev-biochem-060713-035737.
9. 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; 464:441–444. [PubMed: 20154731]

10. Rackham O, Chin JW. A network of orthogonal ribosome x mRNA pairs. *Nat Chem Biol.* 2005; 1:159–166. [PubMed: 16408021]
11. Wang K, Neumann H, Peak-Chew SY, Chin JW. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nature biotechnology.* 2007; 25:770–777.
12. An W, Chin JW. Synthesis of orthogonal transcription-translation networks. *P Natl Acad Sci USA.* 2009; 106:8477–8482.
13. Atkins JF, Bjork GR. A Gripping Tale of Ribosomal Frameshifting: Extragenic Suppressors of Frameshift Mutations Spotlight P-Site Realignment. *Microbiol Mol Biol R.* 2009; 73:178. –+
14. Stahl G, McCarty GP, Farabaugh PJ. Ribosome structure: revisiting the connection between translational accuracy and unconventional decoding. *Trends Biochem Sci.* 2002; 27:178–183. [PubMed: 11943544]
15. Wan W, et al. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. *Angew Chem Int Ed Engl.* 2010; 49:3211–3214. [PubMed: 20340150]
16. Wu B, Wang Z, Huang Y, Liu WR. Catalyst-free and site-specific one pot dual-labeling of a protein directed by two genetically incorporated noncanonical amino acids. *Chembiochem : a European journal of chemical biology.* 2012; 13:1405–1408. [PubMed: 22628069]
17. Chatterjee A, Sun SB, Furman JL, Xiao H, Schultz PG. A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in *Escherichia coli*. *Biochemistry.* 2013
18. Ambrogelly A, et al. Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc Natl Acad Sci U S A.* 2007; 104:3141–3146. [PubMed: 17360621]
19. Jiang R, Krzycki JA. PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine. *J Biol Chem.* 2012; 287:32738–32746. [PubMed: 22851181]
20. Magliery TJ, Anderson JC, Schultz PG. Expanding the genetic code: Selection of efficient suppressors of four-base codons and identification of “shifty” four-base codons with a library approach in *Escherichia coli*. *J Mol Biol.* 2001; 307:755–769. [PubMed: 11273699]
21. Niu W, Schultz PG, Guo J. An Expanded Genetic Code in Mammalian Cells with a Functional Quadruplet Codon. *ACS Chem Biol.* 2013
22. Beuning PJ, Musier-Forsyth K. Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers.* 1999; 52:1–28. [PubMed: 10737860]
23. Chin JW, et al. Addition of p-Azido- l-phenylalanine to the Genetic Code of *Escherichiacoli*. *Journal of the American Chemical Society.* 2002; 124:9026–9027. [PubMed: 12148987]
24. Yanagisawa T, et al. Multistep Engineering of Pyrrolysyl-tRNA Synthetase to Genetically Encode Nε-(o-Azidobenzyloxycarbonyl) lysine for Site-Specific Protein Modification. *Chemistry & Biology.* 2008; 15:1187–1197. [PubMed: 19022179]
25. Nguyen DP, et al. Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins viaa Pyrrolysyl-tRNA Synthetase/tRNA CUAPair and Click Chemistry. *Journal of the American Chemical Society.* 2009; 131:8720–8721. [PubMed: 19514718]
26. Deiters A, Schultz PG. In vivo incorporation of an alkyne into proteins in *Escherichia coli*. *Bioorganic & Medicinal Chemistry Letters.* 2005; 15:1521–1524. [PubMed: 15713420]
27. Seitchik JL, et al. Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific in VivoBioorthogonal Ligation with trans-Cyclooctenes. *Journal of the American Chemical Society.* 2012; 134:2898–2901. [PubMed: 22283158]
28. Chin JW, Martin AB, King DS, Wang L, Schultz PG. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proceedings of the National Academy of Sciences.* 2002; 99:11020–11024.
29. Sasmal PK, et al. Catalytic azide reduction in biological environments. *ChemBioChem.* 2012; 13:1116–1120. [PubMed: 22514188]
30. Blackman ML, Royzen M, Fox JM. Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *Journal of the American Chemical Society.* 2008; 130:13518–13519. [PubMed: 18798613]
31. Devaraj NK, Weissleder R. Biomedical Applications of Tetrazine Cycloadditions. *Acc Chem Res.* 2011; 44:816–827. [PubMed: 21627112]

32. Lang K, Chin JW. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* 2014 in press.
33. Lang K, et al. Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem.* 2012; 4:298–304. [PubMed: 22437715]
34. Lang K, et al. Genetic Encoding of Bicyclononynes and trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels–Alder Reactions. *Journal of the American Chemical Society.* 2012; 134:10317–10320. [PubMed: 22694658]
35. Plass T, et al. Amino Acids for Diels–Alder Reactions in Living Cells. *Angewandte Chemie International Edition.* 2012; 51:4166–4170.
36. Borrmann A, et al. 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. [PubMed: 22945333]
37. Balcar J, Chrisam G, Huber FX, Sauer J. Reaktivitaet von Stickstoff-Heterocyclen gegeneber Cyclooctin als Dienophil. *Tetrahedron Letters.* 1983; 24:1481–1484.
38. Thalhammer F, Wallfaher U, Sauer J. Reaktivitaet einfacher offenkettiger und cyclischer Dienophile bei Diels–Alder Reaktionen mit inversem Elektronenbedarf. *Tetrahedron Letters.* 1990; 31:6851–6854.
39. Kaya E, et al. A Genetically Encoded Norbornene Amino Acid for the Mild and Selective Modification of Proteins in a Copper-Free Click Reaction. *Angewandte Chemie International Edition.* 2012; 51:4466–4469.
40. Lang K, Chin JW. Bioorthogonal reactions for labeling proteins. *ACS Chem Biol.* 2014; 9:16–20. [PubMed: 24432752]
41. Hellstrand E, et al. Forster resonance energy transfer studies of calmodulin produced by native protein ligation reveal inter-domain electrostatic repulsion. *FEBS J.* 2013; 280:2675–2687. [PubMed: 23552119]
42. Wissner RF, Batjargal S, Fadzen CM, Petersson EJ. Labeling proteins with fluorophore/thioamide Forster resonant energy transfer pairs by combining unnatural amino acid mutagenesis and native chemical ligation. *J Am Chem Soc.* 2013; 135:6529–6540. [PubMed: 23594264]
43. Hoeflich KP, Ikura M. Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell.* 2002; 108:739–742. [PubMed: 11955428]
44. Carafoli E. Calcium signaling: a tale for all seasons. *Proc Natl Acad Sci U S A.* 2002; 99:1115–1122. [PubMed: 11830654]
45. Taylor DA, Sack JS, Maune JF, Beckingham K, Quijcho FA. Structure of a recombinant calmodulin from *Drosophila melanogaster* refined at 2.2-Å resolution. *J Biol Chem.* 1991; 266:21375–21380. [PubMed: 1939171]
46. Fallon JL, Quijcho FA. A closed compact structure of native Ca(2+)-calmodulin. *Structure.* 2003; 11:1303–1307. [PubMed: 14527397]
47. Linse S, Helmersson A, Forsen S. Calcium binding to calmodulin and its globular domains. *J Biol Chem.* 1991; 266:8050–8054. [PubMed: 1902469]
48. Babu YS, et al. Three-dimensional structure of calmodulin. *Nature.* 1985; 315:37–40. [PubMed: 3990807]
49. Chou JJ, Li S, Klee CB, Bax A. Solution structure of Ca(2+)-calmodulin reveals flexible hand-like properties of its domains. *Nat Struct Biol.* 2001; 8:990–997. [PubMed: 11685248]
50. Johnson CK. Calmodulin, conformational states, and calcium signaling. A single-molecule perspective. *Biochemistry.* 2006; 45:14233–14246.
51. Wu G, Gao Z, Dong A, Yu S. Calcium-induced changes in calmodulin structural dynamics and thermodynamics. *Int J Biol Macromol.* 2012; 50:1011–1017. [PubMed: 22387072]
52. Porumb T. Determination of calcium-binding constants by flow dialysis. *Anal Biochem.* 1994; 220:227–237. [PubMed: 7978263]
53. Haiech J, Klee CB, Demaille JG. Effects of cations on affinity of calmodulin for calcium: ordered binding of calcium ions allows the specific activation of calmodulin-stimulated enzymes. *Biochemistry.* 1981; 20:3890–3897. [PubMed: 7272283]

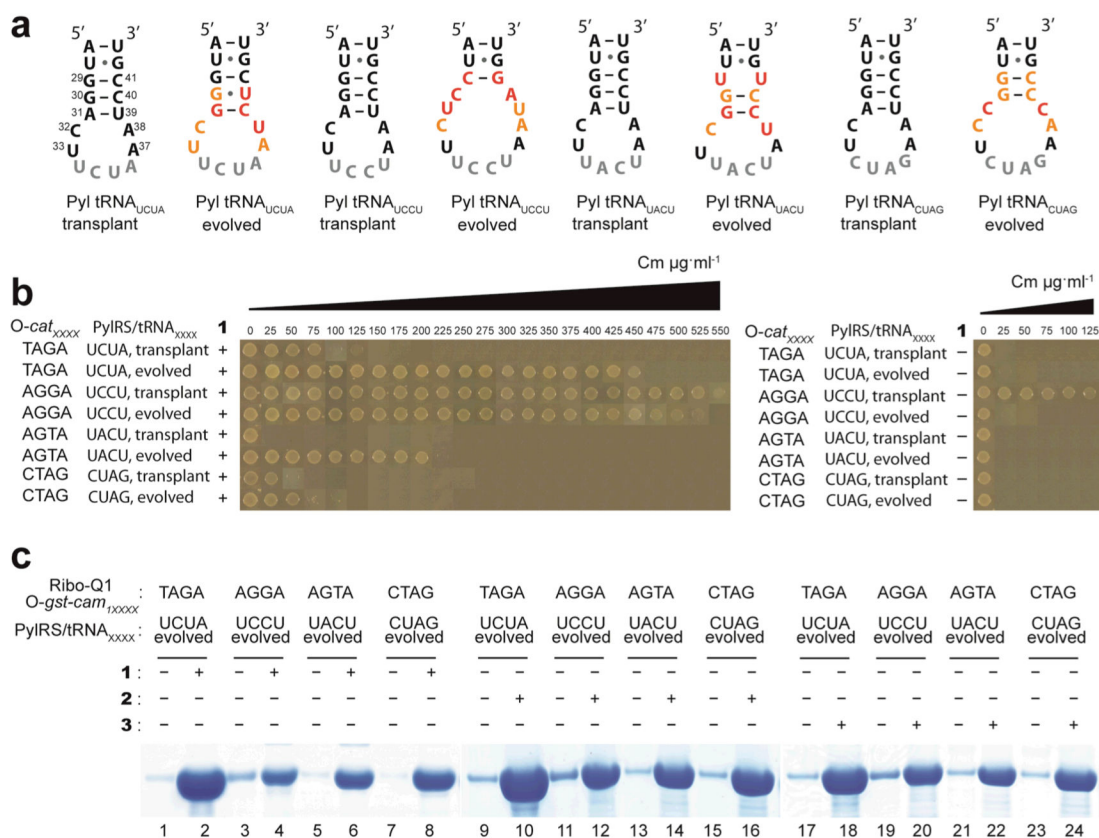
54. Stefan MI, Edelstein SJ, Le Novere N. An allosteric model of calmodulin explains differential activation of PP2B and CaMKII. *Proc Natl Acad Sci U S A.* 2008; 105:10768–10773. [PubMed: 18669651]



**Figure 1. Optimizing Pyl tRNA(N8)<sub>XXXX</sub> for incorporating unnatural amino acids in response to quadruplet codons decoded by ribo-Q1**

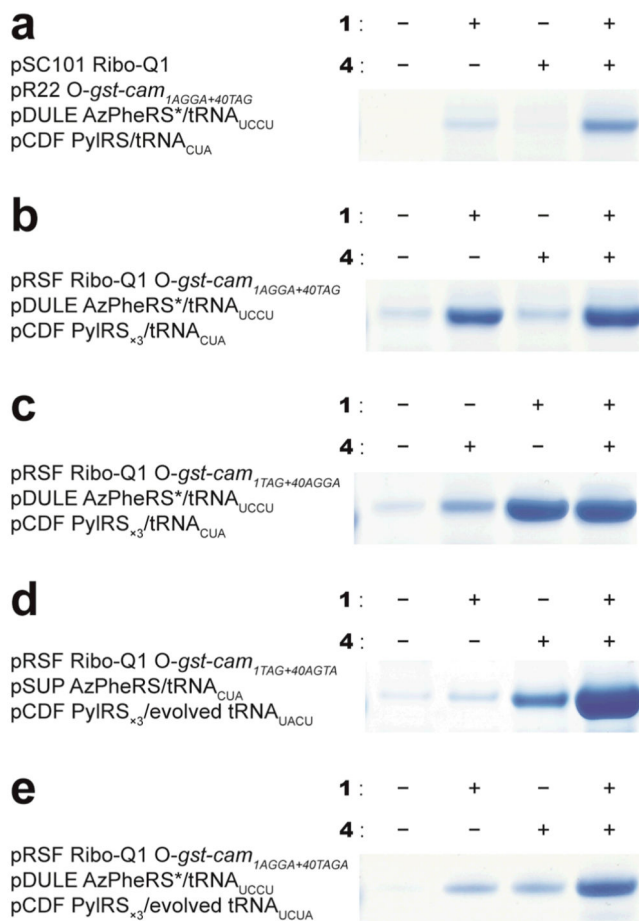
**a.** Nucleotides that are targeted for mutagenesis in the anticodon stem loop of Pyl tRNA(N8)<sub>XXXX</sub> are represented in orange on tRNA bound to the ribosome, and the rest of the tRNA is in yellow. rRNA in pale green with ribo-Q1 mutation sites in red, and mRNA in purple. The structural image is based on PDB ID 2J00, created with Pymol ([www.pymol.org](http://www.pymol.org)). **b.** The anticodon stem loop of Pyl tRNA(N8)<sub>XXXX</sub>. The nucleotides in orange are randomized in each library. Codon sequences and mRNAs are in purple, and anticodons in grey. **c.** Two-step selection procedure for identifying specific and efficient Pyl tRNA(N8)<sub>XXXX</sub> library members in orthogonal translation. A negative *O-barnase* selection followed by a positive *O-cat* selection. Negative selection in the absence of unnatural amino acids eliminates Pyl tRNA(N8)<sub>XXXX</sub> library members that are mis-aminoacylated with natural amino acids by endogenous aminoacyl synthetases. Subsequent positive selection enriches evolved Pyl tRNA(N8)<sub>XXXX</sub> library members that are aminoacylated with the added unnatural amino acid by PyIRS and efficiently decoded at quadruplet codons by ribo-Q1. a.a., amino acid; O-mRNA, orthogonal mRNA; cat, chloramphenicol acetyl transferase.





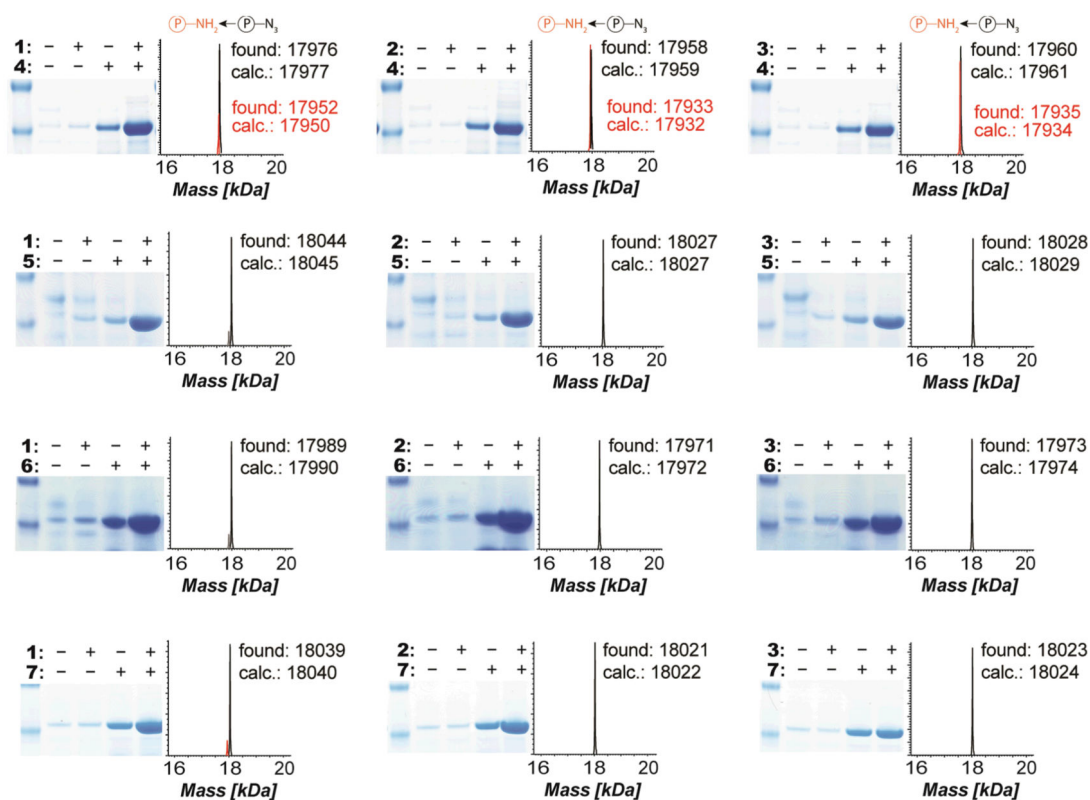
**Figure 2. Evolved Pyl tRNA(N8)<sub>XXXX</sub> direct the efficient unnatural amino acids incorporation in response to quadruplet codons decoded by ribo-Q1**

**a.** The selected anticodon stem-loop sequences of evolved Pyl tRNA<sub>XXXX</sub>, and the corresponding transplant sequences. The anticodons are in gray, and the nucleotides mutated in the library are shown in colour. Positions where parental sequence is selected are in orange, and positions where new nucleotides are selected are in red. **b.** Evolved Pyl tRNA<sub>XXXX</sub> substantially enhance the incorporation of unnatural amino acids by ribo-Q1 in response to quadruplet codons when compared to the corresponding transplant Pyl tRNA<sub>XXXX</sub>. The unnatural amino acid-dependent decoding of quadruplet codons in the *O-cat*<sub>111XXXX</sub> was measured by survival on increasing concentrations of chloramphenicol (Cm). **c.** Diverse unnatural amino acids are efficiently incorporated in recombinant proteins in response to quadruplet codons using PylRS/Pyl tRNA<sub>XXXX</sub> with orthogonal translation. Full gels are given in Supplementary Fig. S4.



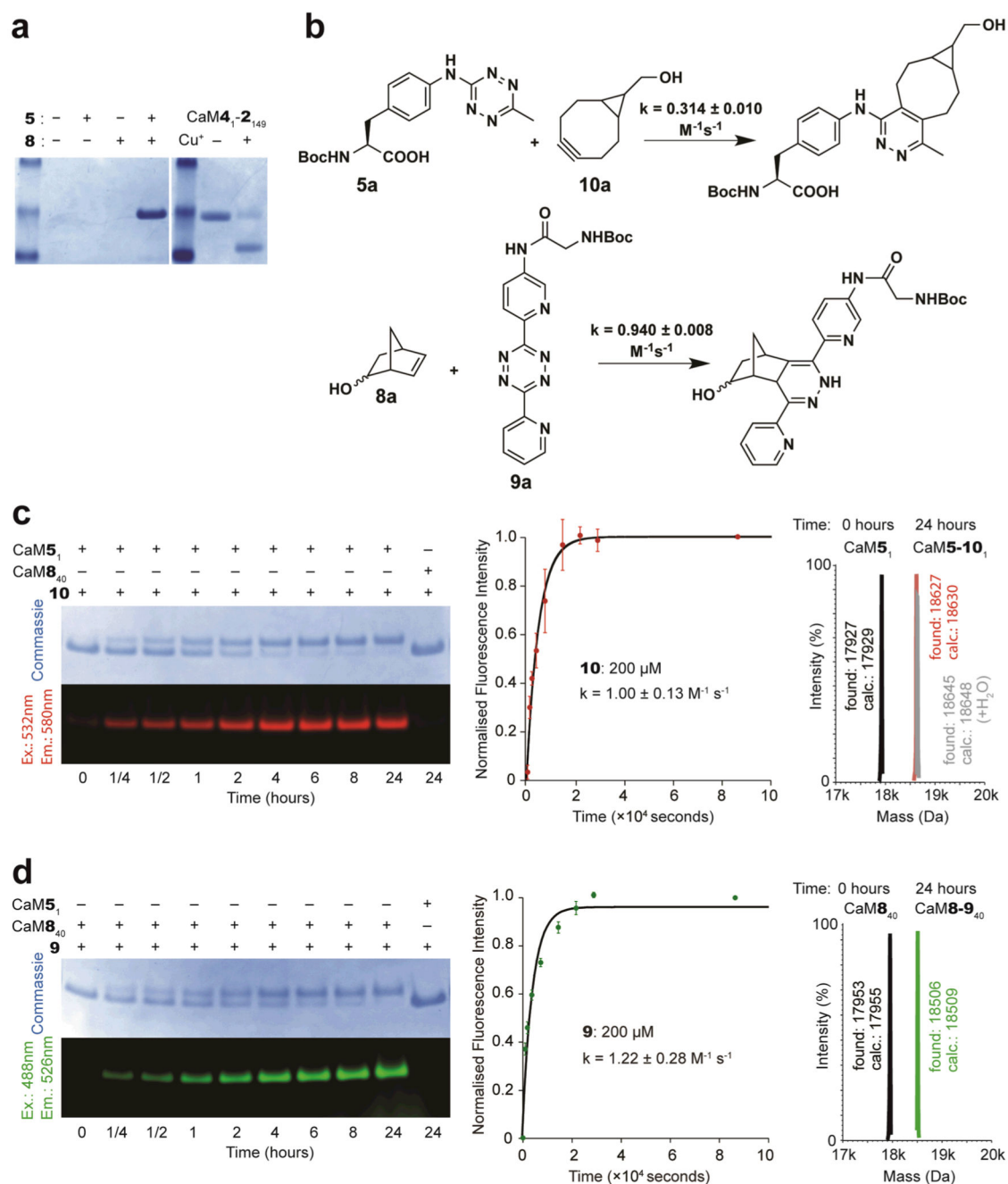
**Figure 3. Efficient incorporation of multiple distinct unnatural amino acids into a single polypeptide**

**a.** Site-specific incorporation of **1** and **4**. AGGA replaces the 1<sup>st</sup> codon and UAG replaces the 40<sup>th</sup> codon in the *cam* open reading frame of *O-gst-cam* to make *O-gst-cam*<sub>IAGGA+40TAG</sub>. Decoding of both the AGGA and TAG codons by ribo-Q1 produces full length Gst-CaM, and failure to decode these codons leads to premature termination of the polypeptide. **b.** and **c.** The site-specific incorporation efficiency of **1** and **4** is improved by reducing the number of plasmids. **d.** The newly evolved PylRS/tRNA<sub>UACU</sub> pair substantially increases the efficiency of double incorporation. **e.** Incorporating two distinct unnatural amino acids using two distinct quadruplet codons. Full gels are given in Supplementary Fig. S7.



**Figure 4. Efficient incorporation of a matrix of pairs of unnatural amino acids, including photocrosslinkers and chemical handles (azides, alkenes, alkynes, tetrazines) demonstrates generality**

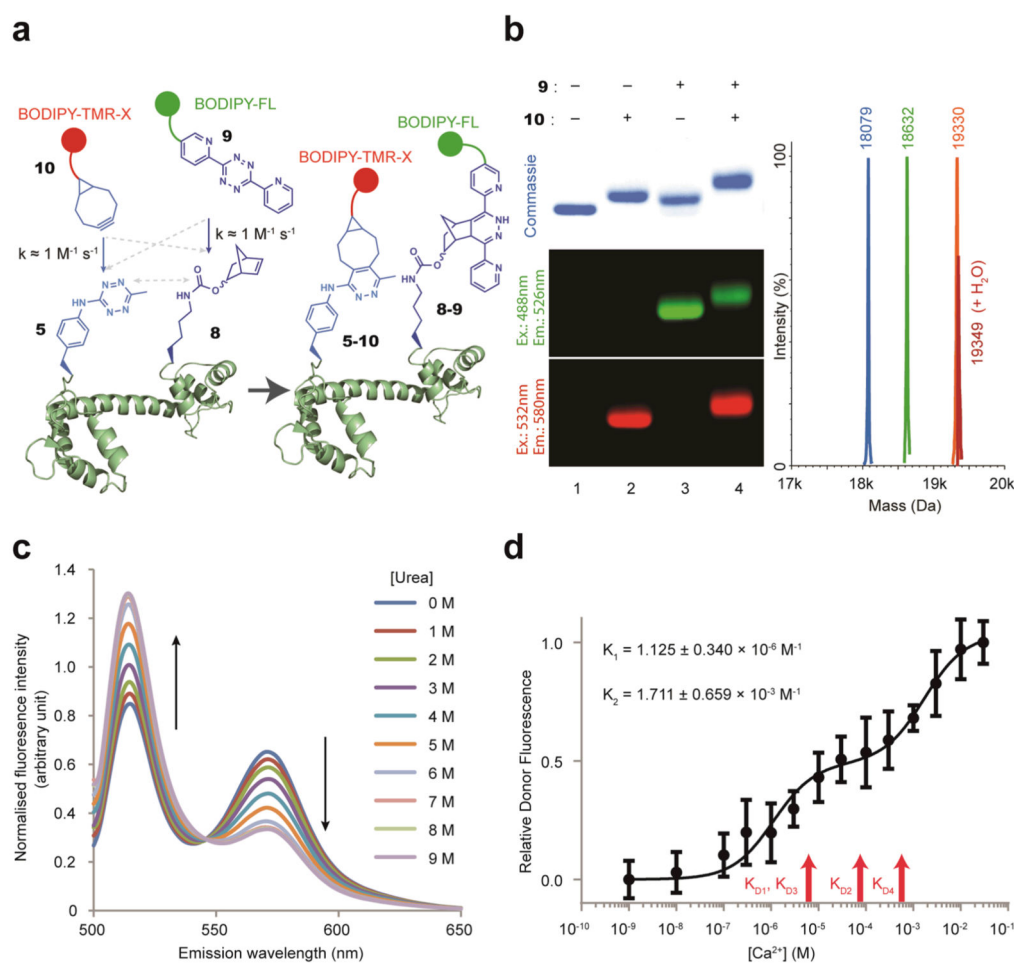
Cells contained *O-gst-cam*<sub>ITAG+40AGTA</sub> and ribo-Q1 expressed from an RSF plasmid, the pSUP *MjAzPheRS/tRNA<sub>CUA</sub>* plasmid (or a variant specific for the relevant substrate) and the pCDF *PylRS/evolved tRNA<sub>UACU</sub>* plasmid. All combinations of *PylRS* substrates (**1-3**) and *MjTyrRS* active site variant substrates (**4-7**) were incorporated in 3×4 matrix. We further confirmed the incorporation of distinct unnatural amino acids by ESI and MALDI mass spectrometry (Supplementary Fig. S8). Full gels are given in Supplementary Fig. S8. We observed an additional peak for protein samples with **4** corresponding to the reduction of the azide to an amine.



**Figure 5.** 5 and 8 do not react with each other in a protein, but can be efficiently labelled with 10 and 9

**a.** 5 and 8 do not react when placed in proximity within a protein. CaM5<sub>1</sub>-8<sub>149</sub> was purified from cells bearing pRSF ribo-Q1 O-*gst-cam*<sub>1TAG+149AGTA</sub>, pSUP *Mj*TetPheRS/tRNA<sub>CUA</sub>, pCDF NorKRS<sub>x3</sub>/evolved tRNA<sub>UACU</sub>. CaM4<sub>1-2</sub><sub>149</sub> undergoes a Cu(I) catalyzed click reaction to cyclize the protein (right gel panel). CaM5<sub>1</sub>-8<sub>149</sub> does not cyclize, as judged by mobility shift (compare left and right panels) and ESI-MS (Supplementary Fig. S10). **b.** Rate constants for the indicated reactions. **c.** CaM5<sub>1</sub> and CaM8<sub>40</sub> were incubated with 100 molar equivalents of 10 at 25°C. Only CaM5<sub>1</sub> was labelled with 10, yielding CaM5-10<sub>1</sub>.

Labelling was visualized using a Typhoon Imager and the resulting time-dependent fluorescence was used to calculate the on-protein labelling rate constant. All measurements were repeated twice and the error bars represent the standard deviation. ESI-MS confirmed that protein labelling is quantitative. **d.** CaM5<sub>1</sub> and CaM8<sub>40</sub> were incubated with 100 molar equivalents of **9** at 25°C. Only CaM8<sub>40</sub> was labelled with **9**, which yielded CaM8-**9**<sub>40</sub>. The labelling reaction was analysed as described in **c.** ESI-MS confirmed that labelling was quantitative.



**Figure 6. Site-specific double labelling of CaM with a FRET pair to follow changes in protein conformation**

**a.** Strategy for protein double labelling via inverse electron-demand Diels-Alder reactions.

**b.** Quantitative and site-specific double labelling of CaM<sub>51-840</sub>. Purified CaM<sub>51-840</sub> (**Lane 1**) was labelled with 100 equivalents (200 $\mu$ M) of **10**, which yielded CaM<sub>5-101-840</sub> (**Lane 2**) or 100 equivalents **9**, which yielded CaM<sub>51-8-940</sub> (**Lane 3**). CaM<sub>51-8-940</sub> was labelled with 100 equivalents of **10**, which yielded CaM<sub>5-101-8-940</sub> (**Lane 4**). Labelling was visualised by fluorescence imaging and led to a mobility shift. All labelling reactions were quantitative, as confirmed by ESI-MS. CaM<sub>51-840</sub> (blue peak; calculated Mass=18081 Da, observed Mass=18079 Da), CaM<sub>51-8-940</sub> (green peak; calculated Mass=18635, observed Mass=18632), CaM<sub>5-101-8-940</sub> (orange peak; calculated Mass=19336, observed Mass=19330). **c.** Fluorescence spectra of CaM<sub>5-101-8-9149</sub> (following donor excitation at 485 nm) in the presence of increasing concentrations of urea. **d.** The relative donor-fluorescence intensity from doubly labelled CaM<sub>5-101-8-940</sub> as a function of Ca<sup>2+</sup> concentration. All measurements were repeated at least six times and the error bars represent the standard deviation.  $K_1$ ,  $K_2$  are for the observed transitions.  $R^2=0.9005$ .  $K_{D1}$ ,  $K_{D2}$ ,  $K_{D3}$  and  $K_{D4}$  are the reported  $K_D$  (dissociation constants) values for sequential Ca<sup>2+</sup> binding<sup>51</sup>.