# Multiplexed genomic encoding of non-canonical amino acids for labeling large complexes

**Bijoy J. Desai**, **Ruben L. Gonzalez Jr.**[*]

Department of Chemistry, Columbia University, 3000 Broadway, MC3126, New York, NY 10027, USA

## Abstract

Stunning advances in the structural biology of multicomponent biomolecular complexes (MBCs) have ushered in an era of intense, structure-guided mechanistic and functional studies of these complexes. Nonetheless, existing methods to site-specifically conjugate MBCs with biochemical and biophysical labels are notoriously impracticable and/or significantly perturb MBC assembly and function. To overcome these limitations, we have developed a general, multiplexed method in which we genomically encode non-canonical amino acids (ncAAs) into multiple, structure-informed, individual sites within a target MBC; select for ncAA-containing MBC variants that assemble and function like the wildtype MBC; and site-specifically conjugate biochemical or biophysical labels to these ncAAs. As a proof-of-principle, we have used this method to generate unique single-molecule fluorescence resonance energy transfer (smFRET) signals reporting on ribosome structural dynamics that have thus far remained inaccessible to smFRET studies of translation.

## INTRODUCTION

Many fundamental cellular processes, including DNA repair, replication, transcription, messenger RNA (mRNA) processing and splicing, nuclear export, mRNA decay, translation, and protein degradation, are performed by MBCs. Continuing advances in X-ray crystallography and, more recently, cryogenic electron microscopy (cryo-EM) studies of MBCs are providing researchers with structural frameworks for most informatively positioning biochemical and biophysical probes and interpreting the resulting biochemical and biophysical data in terms of structure-based mechanistic models[1]. Unfortunately, however, it remains difficult and, in some cases, impossible to site-specifically label MBCs at defined positions, severely impeding mechanistic and functional studies. The primary reasons for this are that MBCs are composed of up to thousands of oftentimes essential

[*]To whom correspondence should be addressed: Ruben L. Gonzalez Jr., Department of Chemistry, Columbia University, 3000 Broadway, MC3126, New York, NY 10027, USA, Tel.: (212) 854-1096; Fax: (212) 932-1289; rlg2118@columbia.edu.

COMPETING INTERESTS

The authors declare no competing financial interests.

components, have tightly controlled component stoichiometries, and are typically assembled through intricate and highly regulated assembly pathways[2, 3]. Consequently, common methods for site-specifically labeling proteins, such as conjugation of a maleimide-derivatized reporter to a cysteine residue, are impracticable for MBCs that contain hundreds of native reactive residues, while the peptide tags often used in chemo-enzymatic labeling methods are usually limited to protein termini where they are least likely to inhibit assembly and/or function[4]. In addition, many labeling approaches involve production of a target protein from a genetic construct (*e.g.*, a plasmid) that removes the gene from its native genomic regulatory context, and this can perturb the component stoichiometry and/or cellular assembly process[5]. Attempts to overcome these issues by partially or fully *in vitro* reconstituting MBCs from recombinantly overexpressed, purified, and labeled components often results in compositionally and/or functionally heterogeneous MBC mixtures that exhibit impaired activities[5].

Here, we report an approach that integrates homologous recombination-based multiplexed genome engineering (MGE)[6], ncAA mutagenesis technology[7], and bioorthogonal chemistry[8] to rapidly generate numerous, fully functional MBC variants in which each variant can carry a biochemical or biophysical label at one or more defined target positions. Our approach combines the power of MGE to rapidly generate multiple, orthogonal codon mutations (Fig. 1a) with the specificity and modularity of bioorthogonal ncAA-based conjugation chemistry, while maintaining the genomic regulatory context, *in vivo* assembly pathway, and functional integrity of the target MBC (Fig. 1b–c).

To demonstrate the power of our approach, we chose to target the *Escherichia coli* ribosome – a 2.5 MDa, two-subunit MBC comprised of 55 ribosomal proteins and 3 ribosomal RNAs (rRNAs). In bacteria, assembly and maturation of the ribosome is a complex process that requires the action of ~100 cellular factors[2]. The importance of the genomic regulation and *in vivo* assembly and maturation process of this MBC is highlighted by the observation that ribosomes composed of a fully *in vitro* reconstituted small, or 30S, ribosomal subunit are only 34–50% active compared to ribosomes composed of 30S subunits purified from cells[5], a fact that has complicated interpretation of biophysical studies performed using ribosomes composed of 30S subunits that were site-specifically fluorophore-labeled *via* full *in vitro* reconstitution[9, 10].

Using the multiplexed method we have developed, we were able to rapidly generate ten mutant strains of *E. coli* that genomically encode an ncAA, *p*-azido-L-phenyl alanine (*p*-AzF), at ten different structure-informed target sites across eight ribosomal protein genes. *In vivo*-assembled ribosomes purified from these mutant strains carry *p*-AzF at defined position(s) that we then show can be bioorthogonaly conjugated with appropriately derivatized fluorophores. Collectively, these fluorophore-labeled ribosomes have enabled us to generate seven smFRET signals, five novel signals reporting on conformational dynamics of the ribosome that have thus far remained inaccessible to smFRET studies and two signals that serve as alternatives to previously reported signals that were difficult to interpret unambiguously. The results we present here highlight the power of our method for performing smFRET-based mechanistic studies of translation and provide a general

approach that enables an almost limitless number of biochemical and biophysical studies of other MBCs.

## RESULTS

### Design and selection of labeling positions

For site-specific labeling of the ribosome, we targeted thirteen different positions across nine ribosomal protein genes (Fig. 2, Table 1) such that fluorophores at those positions would enable smFRET studies of three ribosomal structural rearrangements that have thus far remained inaccessible to smFRET studies or have been studied using smFRET signals that could not be interpreted unambiguously: (*i*) intra-subunit rotation of the 'head' domain of the 30S subunit relative to the 'body' domain (*i.e.*, 'head swiveling', HS), (*ii*) movement of a translating ribosome along its mRNA template (*i.e.*, 'mRNA translocation', MT), and (*iii*) rotation of the large, or 50S, ribosomal subunit relative to the 30S subunit (*i.e.*, 'intersubunit rotation', IR) (Fig. 2). Specifically, direct monitoring of HS has thus far remained inaccessible to smFRET studies, while indirect monitoring of HS has been difficult to interpret due to the use of a transfer RNA (tRNA)-ribosome smFRET signal that convolutes movements of the tRNA with movements of the head domain[11]. Similarly, it has not yet been possible to monitor MT using smFRET. Monitoring of IR using smFRET has required full and/or partial *in vitro* reconstitution of one or both ribosomal subunits[9, 10, 12], the introduction of relatively large peptide tags for chemo-enzymatic labeling[13], and/or the engineering of relatively large RNA tags into the ribosomal RNA for hybridization of fluorophore-labeled DNA oligos[14, 15], approaches that in at least some cases can impair ribosome function[5, 9, 10] and/or result in contradictory conclusions[10, 12, 14].

To identify candidate labeling positions that would report on HS and MT, we used previously reported, high-resolution, X-ray crystallographic structures of ribosomal complexes to model HS and MT. For HS, we used structures of ribosomal complexes in which the head domain of the 30S subunits occupy the so-called 'non-swiveled' or 'swiveled' conformations[16, 17]. We then computationally identified pairs of amino acid residues consisting of one residue from the head domain and one from the body domain of the 30S subunit that would be predicted to result in an easily detected change in FRET efficiency ($E_{FRET}$) of > 0.2 between the non-swiveled and swiveled conformations. From this initial list of residue pairs, we selected for further consideration only those pairs in which both residues were predicted to be surface accessible as determined by visual inspection. Finally, residues from pairs under consideration were phylogenetically analyzed and only those pairs in which both residues exhibited less than 70% sequence identity were pursued experimentally (Fig. 2, Table 1, Methods). We used an analogous approach for MT, except we used a single high-resolution X-ray crystallographic structure in which the 3'- and 5' ends of the mRNA within the mRNA entry and exit channels of the 30S subunit, respectively, were ordered[18]. This structure allowed us to identify amino acid residues from ribosomal proteins in the body domain of the 30S that would result in a > 0.2 change in $E_{FRET}$ upon translocation of mRNA constructs harboring fluorophores at defined positions within their 3' or 5' ends. To identify candidate labeling positions that would report on IR, we used sites identified in previous studies[9, 10] (Fig. 2, Table 1, Methods).

### Engineering *E. coli* for MGE-based ncAA encoding

In order to site specifically label the *E. coli* ribosome at the selected thirteen positions, we decided to genomically encode the ncAA *p*-AzF using engineered UAG stop codons that are typically decoded by polypeptide release factor (RF) 1 during translation termination. The choice of *p*-AzF was driven by the fact that it undergoes a rapid and robust, bioorthogonal conjugation reaction with dibenzocyclooctyne (DBCO)-derivatized[19] labels under mild buffer and temperature conditions (Fig. 1c). To genomically encode *p*-AzF at engineered UAG stop codons in a manner that was free of potential competition from RF1-mediated translation termination, we chose to use the previously developed C321 A strain of *E. coli*[20] (Methods). In C321 A, all 321 naturally occurring, essential TAG stop codons in *E. coli* have been changed to TAA stop codons that are decoded by RF2 and, moreover, the gene encoding RF1, *prfA*, has been deleted. Additionally, C321 A carries a temperature-inducible lambda prophage for carrying out single-stranded, DNA-guided homologous recombination (*i.e.*, lambda Red ($\lambda_{Red}$)-mediated recombineering)[21], thus allowing MGE to be used to introduce codon mutations at the desired labeling positions. To tailor C321 A for *p*-AzF incorporation, we transformed it with a plasmid expressing an orthogonal nonsense suppressor tRNA$^{p\text{-}AzF}_{CUA}$ that recognizes UAG codons and a tRNA synthetase (RS) specific for amino-acylating this tRNA$^{p\text{-}AzF}_{CUA}$ with *p*-AzF[22, 23] (Fig 1b, Methods), and, to tailor it for ribosome work, we disrupted the *rna* gene encoding ribonuclease I (RNase I)[24–26] (Methods).

### Multiplexed genomic encoding of ncAAs

Starting with our tailored C321 A strain, we performed six to eight rounds of $\lambda_{Red}$-mediated MGE using a multiplexed pool of oligonucleotides targeting the thirteen labeling sites for mutation to TAG stop codons in the ribosomal protein genes in the *E. coli* genome (Supplementary Table 1, Methods). All rounds of MGE were performed in the presence of 1 mM *p*-AzF such that incorporation of the ncAA into the MBC occurs immediately upon introduction of the mutation into the genome. The resulting cell population was screened using multiplexed, allele-specific colony (MASC)-PCR[27] (Methods) such that we could identify strains carrying each mutation (Supplementary Fig. 1). Once identified, the presence of each mutation was further confirmed by Sanger sequencing (Supplementary Fig. 2). Using this approach, we were able to isolate ten strains in which each strain carried either one or two TAG mutations at locations corresponding to ten of the thirteen positions we originally targeted. We hypothesized that strains carrying TAG mutations at the remaining three positions could not be isolated because these mutations conferred significant fitness disadvantages. Supporting this hypothesis, the percent enrichment of mutation at each of the targeted positions ranged between two- to ten-fold lower than a mutation of a similar size in a non-essential gene (Table 1). These observations are consistent with our expectation that, during the rounds of MGE, our approach selects for MBCs with functionally permissible mutations.

To demonstrate the broad applicability of our approach, we repeated our experiments using the EcNR2 strain of *E. coli*, a standard *E. coli* MG1655 strain containing all 321 naturally occurring, essential TAG stop codons and intact *prfA* and *rna* genes[6] (Methods). For the experiments in EcNR2, we used a limited subset of oligonucleotides targeting sites that had

exhibited high (L9 N11, where L9 denotes protein L9 in the 50S subunit and N11 denotes the asparagine at residue position 11), intermediate (S18 R8, where S18 denotes protein S18 in the 30S subunit), and low (S18 Q75) percent enrichments in the C321 A strain. The results of these experiments showed we could isolate all of the target mutations in EcNR2, demonstrating the efficacy of our approach in any standard *E. coli* strain (Supplementary Table 2). Furthermore, the percent enrichments of the targets showed the same trend as in C321 A, suggesting the mutations exert similar selective pressures on the two strains.

### Bioorthogonal labeling of ncAA-containing ribosomes

To demonstrate that the isolated, mutant *E. coli* strains were able to assemble ribosomes with site-specific incorporation of *p*-AzF that can be efficiently labeled, we first purified ribosomes from three double-mutant strains (HS1, HS2, and IR1), and one single-mutant strain (MT1) (Methods). To ensure high-efficiency labeling, we separated the purified ribosomes into 30S and 50S subunits. The subunits were then labeled with DBCO-derivatized Cy3 FRET donor and/or Cy5 FRET acceptor fluorophores (Methods). The ribosomal proteins from each subunit were separated on an SDS-PAGE gel, and the gel was imaged using a fluorescence gel scanner (Fig. 3). In these scans, fluorescence was observed exclusively from bands corresponding to ribosomal protein(s) whose gene(s) contained the TAG mutation(s). Moreover, fluorescence scanning of an SDS-PAGE gel containing ribosomal proteins obtained from the 30S and 50S subunits of a C321 A strain that did not contain any TAG mutations, but that was otherwise grown in the presence of *p*-AzF and used to purify and fluorophore-label the 30S and 50S subunits in a manner identical to that of the strains containing TAG mutations did not exhibit any fluorescence (Extended Data Fig. 1). This observation demonstrates the site-specific nature of *p*-AzF incorporation into the ribosomal proteins of the mutant strains.

As expected, 30S subunits purified from the HS1 and HS2 double-mutant strains and labeled with a mixture of DBCO-derivatized Cy3 and Cy5 fluorophores contained an equal amount of Cy3 and Cy5 at each labeling site (Fig. 3, Middle and Right Panels, Lanes 4 and 5). Such a mixture of labeled products can be separated using chromatographic methods[28] or, in the case of single-molecule fluorescence experiments, analyzed using computational methods[29].

To measure the labeling efficiency of each labeled position, we quantified the fluorescence intensity of each labeled protein band, and compared it to a standard curve generated using known quantities of a fluorophore-labeled protein standard (Supplementary Fig. 3, Methods). We found that the labeling efficiency was different for each targeted position, and ranged from a high of 96 % for labeling at S6 D41 to a low of 15 % for S7 G112 (Supplementary Table 3). While it's possible that the lower labeling efficiencies could be due to low *p*-AzF incorporation caused by background suppression of the UAG mutation by near-cognate aa-tRNAs (*e.g.*, Tyr-tRNA[Tyr]), such background suppression has been found to be quite low in C321 A-based strains in presence of *p*-AzF[22] and we therefore suspect it is more likely due to incomplete fluorophore labeling arising from the potentially low solvent exposure of the targeted residue.

## smFRET experiments using fluorophore-labeled ribosomes

To demonstrate the utility of our labeled ribosomes for smFRET studies, we performed smFRET experiments using wide-field, total internal reflection fluorescence (TIRF) microscopy (Methods). In these experiments, the Cy3- and/or Cy5-labeled ribosomes described above and any required tRNAs and/or translation factors were used to assemble ribosomal initiation or elongation complexes on a 5'-biotinylated mRNA that was either unlabeled or Cy5-labeled by hybridizing a 5'-Cy5-labeled DNA oligonucleotide just upstream of the ribosome binding site of the mRNA. The resulting complexes were then tethered to the surface of a polyethylene glycol (PEG)/biotin-PEG-derivatized quartz microfluidic flow-cell using a biotin-streptavidin-biotin bridge (Fig. 4, Top Panel). As previously described, tethered complexes were illuminated using the evanescent wave generated by the 532 nm laser component of the TIRF microscope, and Cy3 and Cy5 fluorescence emissions from individual ribosomes were collected, wavelength separated, and imaged using the optical and detector components of the TIRF microscope[29, 30]. We imaged complexes harboring the HS1, MT1, and IR1 smFRET signals and, in each case, we observed Cy3 and Cy5 florescence intensity *versus* time trajectories that exhibited anti-correlated Cy3 and Cy5 fluorescence intensity changes followed by single-step photobleaching of the Cy3 and Cy5 fluorophore, demonstrating that the complexes contained single Cy3 and Cy5 fluorophores that were positioned so as to generate detectable $E_{\mathrm{FRET}}$s (Fig. 4, Middle and Bottom Panels).

For an initiation complex harboring the HS1 smFRET signal, the $E_{\mathrm{FRET}}$ *versus* time trajectories reveal that the head domain of the 30S subunit exists in two $E_{\mathrm{FRET}}$ states (Fig. 4a, Extended Data Fig. 2a and b). The observed $E_{\mathrm{FRET}}$ values associated with these two $E_{\mathrm{FRET}}$ states, ~0.5 and ~0.7, are consistent with $E_{\mathrm{FRET}}$ values predicted using distances taken from X-ray crystallographic structures of ribosomal complexes in which the head domain of the 30S subunit is in the non-swiveled or swiveled conformations (~0.6 for the 50 Å distance between our labeling sites in the non-swiveled conformation and ~0.9 for 39 Å distance between our labeling sites in the swiveled conformation) (Table 2)[17]. This result reveals that, within the context of an initiation complex, the head domain of the 30S subunit can sample at least two conformational states. These observations strongly suggest that these states play an important role in the translation initiation pathway[31, 32] and set the stage for smFRET studies aimed at characterizing the contributions that these states might make to the mechanism and regulation of initiation.

For an initiation complex harboring the MT1 smFRET signal, the $E_{\mathrm{FRET}}$ *versus* time trajectories reveal that the segment of mRNA just upstream of the mRNA binding channel samples a single $E_{\mathrm{FRET}}$ state (Fig. 4b, Extended Data Fig. 2c). The $E_{\mathrm{FRET}}$ value observed for this $E_{\mathrm{FRET}}$ state, ~0.8, is consistent with the ~0.9 $E_{\mathrm{FRET}}$ value predicted using an ~36 Å estimate of the distance between our labeling sites on the mRNA and near the exit of the mRNA binding channel (Table 2). This observation strongly suggests that the segment of mRNA just upstream of the mRNA binding channel either adopts a single conformation or, alternatively, samples more than one conformation on a timescale that is at least an order of magnitude faster than the time resolution of the electron multiplying charge-coupled device (EMCCD) camera detector on our TIRF microscope.

For an elongation complex harboring the IR1 smFRET signal, the $E_{\text{FRET}}$ *versus* time trajectories reveal that the ribosome undergoes stochastic, thermally activated fluctuations between at least two $E_{\text{FRET}}$ states (Fig 4c, Extended Data Fig. 2d). Notably, the $E_{\text{FRET}}$ values observed for these two $E_{\text{FRET}}$ states, ~0.7 and ~0.5, are inconsistent with the $E_{\text{FRET}}$ values predicted using distances taken from X-ray crystallographic structures of ribosomal complexes in which the ribosome is in the so-called 'non-rotated' or 'rotated' subunit orientations (~0.4 for the 59 Å distance between our labeling sites in the non-rotated subunit orientation and ~0.2 for the 68 Å distance between our labeling sites in the rotated subunit orientation)[33] (Table 2). Nonetheless, our observed $E_{\text{FRET}}$ values are in very close agreement with those previously observed by two different groups using the analogous smFRET signal[9, 10, 12]. Moreover, the $E_{\text{FRET}}$ values observed by us here and by others using the analogous smFRET signal[10, 12] predict that transitions between the non-rotated and rotated subunit orientations should result in a change in distance between our labeling sites of ~10 Å. This is a result that is in excellent agreement with the change in distance of ~9 Å that is observed when comparing the X-ray crystallographic structures of the non-rotated and rotated subunit orientations of the ribosome[33]. Furthermore, the timescale of the transitions we observe between our two $E_{\text{FRET}}$ states is also consistent with that observed previously using the analogous smFRET signal[10, 12]. Collectively, these observations strongly suggest that the previously reported full *in vitro* reconstitution of the 30S subunit[5] resulted in a subpopulation of subunits that were fully active and a subpopulation that were fully inactive and, furthermore, that the fully inactive subpopulation did not significantly contribute to the previously reported smFRET results[10].

## DISCUSSION

Here we have demonstrated how a combination of MGE, ncAA mutagenesis, and bioorthogonal chemistry can be used to quickly produce a large set of fully functional, site-specifically labeled MBC variants. The use of MGE enables rapid and high-throughput mutagenesis of numerous positions within an MBC. In addition, MGE preserves the genomic context of the targeted genes, thereby maintaining proper genomic regulation of MBC components and ensuring correct *in vivo* assembly of the MBC. Moreover, in the case of essential MBCs, MGE requires each mutant strain to compete against all other mutant strains and against the wildtype strain, thereby selecting for mutant MBCs that assemble and function with near-wildtype activities.

The use of ncAA mutagenesis enables a unique chemical moiety to be site-specifically incorporated into a specific protein component of an MBC. Compared to the relatively large peptide recognition tags that must be inserted into target proteins for chemo-enzymatic labeling, substitution of canonical amino acids with an ncAAs is a much less structurally and functionally perturbative change. Moreover, chemo-enzymatic tags must be labeled *via* sterically demanding enzymatic reactions, whereas ncAAs can be labeled using more sterically permissive small-molecule chemical reactions. Thus, relative to chemo-enzymatic methods, ncAA mutagenesis allows efficient labeling of virtually any position in an MBC. Moreover, implementing ncAA mutagenesis using ncAAs that can undergo bioorthogonal conjugation chemistry enables MBCs to be labeled with a wide variety of biochemical or biophysical probes that enable a large array of biochemical and biophysical applications

(Supplementary Table 4). Thus, by uniquely integrating MGE, ncAA mutagenesis, and bioorthogonal chemistry, we have developed a method for rapidly generating a series of mutant strains in which each strain produces MBCs that can be differentially and, if needed, combinatorially labeled with almost any biochemical or biophysical probe of interest without significantly perturbing the genomic regulation, *in vivo* assembly, structure, or function of the labeled MBCs.

Using this method, we were able to readily generate a set of *E. coli* mutant strains capable of site-specifically incorporating a bioorthogonal ncAA, *p*-AzF, at one or two of ten different amino acid positions in eight ribosomal proteins without perturbing the *in vivo* assembly or function of the ribosome. Each of these ribosome variants could then be labeled with Cy3 and/or Cy5 to generate five novel and two previously reported smFRET signals. Given the difficulty of *a priori* predicting the labeling efficiency of a particular target residue, a major advantage of our method is that it is based on MGE, an approach that allows rapid and facile screening for other, comparable labeling positions with higher labeling efficiencies. Collectively, the smFRET signals we have developed report on functionally important structural dynamics of the ribosome that have not previously been possible to observe using smFRET or that have been investigated using smFRET signals that have been challenging to interpret.

Using four of the seven smFRET signals we developed, we demonstrated how our approach enables powerful smFRET studies of the HS, MT, and IR dynamics of the ribosome (Fig. 4). In addition to studies of the role that HS plays during translation initiation (Fig. 4a), our HS1 and HS2 smFRET signals will enable future investigations aimed at determining the contributions that HS makes to the mechanism of late steps in the translocation of the ribosome along the mRNA during translation elongation[17, 34], translation termination[35], ribosome recycling[35], and translational quality control[36]. Likewise, our newly developed MT1 smFRET signal will not only allow us to monitor the kinetics of initiation complex recruitment to an mRNA during translation initiation (Fig. 4b), but also of stepwise extrusion of mRNA from the exit of the mRNA binding channel as the ribosome undergoes multiple rounds of translocation during translation elongation. Similarly, our IR1 smFRET signal (Fig. 4c) will allow us to unambiguously characterize the contributions that IR makes to the mechanisms and regulation of all stages of translation.

Given recent advances in the use of MGE and ncAAs in bacterial[37], yeast[38], and mammalian cells[39], our approach can be readily expanded for site-specific labeling of both bacterial and eukaryotic MBCs. Of particular relevance are the recent TALEN-[40], CRISPR-[39], and zinc-finger nuclease-based MGE methods that have been developed for use in eukaryotic model systems[41] and that could facilitate expansion of our approach into eukaryotic cells for the labeling of eukaryotic MBCs. Such an expansion would represent an important next step for the evolution of our approach, as it would facilitate labeling of MBCs that can either be most easily or exclusively expressed in eukaryotic cells, but that nonetheless comprise MBCs of exceptional importance. This includes MBCs such as eukaryotic ribosomes, nucleosomes, spliceosomes, proteasomes, G-protein-coupled receptors (GPCRs), and many others that can collectively serve as therapeutic drug targets for a large number of human diseases. In addition, combining MGE with co-selection or counter-selection methods[42] should further

improve the ease, efficiency, and success of our approach. Likewise, we expect that future extensions of the approach we describe here will include incorporation of a second, chemically distinct ncAA that can be encoded by a different codon using an additional orthogonal tRNA-RS pair (*e.g.* the pyrrolysine tRNA-RS pair[43]). Indeed, emerging improvements in ncAA mutagenesis technology to recode sense codons and quadruplet codons (*i.e.*, *via* +1 frameshifting) and innovations in novel bioorthogonal chemistries should allow future development of the approach described here to include even more greatly expanded, multi-site-specific bioorthogonal labeling of MBCs.

## ONLINE METHODS

### Strains and Plasmids

The C321 ΔA strain was a kind gift from George Church[20]. It is an *E. coli* MG1655 strain in which all 321 naturally occurring essential TAG stop codons have been changed to TAA stop codons and the gene encoding RF1, *prfA*, has been deleted. Its exact geneotype is: Δ(ybhB-bioAB)::[lcI857 N(cro-ea59)::tetR-bla] ΔprfA Δ mutS::zeoR. To render the C321 ΔA strain RNase I deficient, we performed four homologous recombination cycles using a mutagenic oligonucleotide that targeted the ninth codon of the gene encoding RNase I, *rna*, for mutation to a TAA stop codon using a previously described method[6, 27]. Strains containing the analogous mutation in *rna* are standard tools in biochemical and biophysical studies of ribosomes, translation, and translational control[24–26]. The pEvol-pAzFRS.2.t1 plasmid encoding an orthogonal non-sense suppressor tRNA$^{p-AzF}_{CUA}$ that recognizes UAG codons and the tRNA RS that specifically amino-acylates this tRNA$^{p-AzF}_{CUA}$ with *p*-AzF was a kind gift from Farren Isaacs[22]. For the experiments described here, we transformed the RNase I-deficient C321 ΔA strain with pEvol-pAzFRS.2.t1. The EcNR2 strain, which is an *E. coli* MG1655 strain containing all 321 naturally occurring, essential TAG codons as well as the *prfA* and *rna* genes encoding RF1 and RNase I, respectively, was a kind gift from Harris Wang. For the experiments described here, we transformed the EcNR2 strain with pULTRA-pAzFRS.2.t1.

### Structure- and sequence-guided design and selection of labeling positions

All structural modeling was performed using the PyMOL Molecular Graphics System v2.2 (Schrödinger, LLC)[46]. To identify labeling positions that would report on HS, we used high-resolution structures representing translation elongation intermediates that capture the head domain of the 30S subunit in its non-swiveled and swiveled conformations (PDB IDs 5IBB[16] and 4W29[17] respectively). For each structure, we visually identified the amino acid residues from ribosomal proteins S2, S3, S7, S9, S10, S13, S14 and S19 in the head domain of the 30S subunit that were surface accessible. Then, for each structure, we used PyMOL selection commands to identify all of the amino acid residues from ribosomal proteins S4-S6, S8, S11, S12, S15-S18, and S20 in the body domain of the 30S subunit whose C$_\alpha$s were within 38–69 Å of the C$_\alpha$ of one of the surface-accessible residues identified in ribosomal proteins S2, S3, S7, S9, S10, S13, S14 and S19 in the head domain of the 30S subunit. The 38–69 Å range of distances that was used to identify these pairs of residues corresponds to Cy3-Cy5 distances that fall within the quasi-linear regime of the $E_{FRET}$ *versus* Cy3-Cy5 distance curve[45]. Residue pairs in which the identified residue in the body domain of the

30S subunit was not surface accessible by visual inspection were then excluded from further analysis. Only the residue pairs that met all of these criteria in both structures were selected for further analysis. For each structure, we next calculated the predicted $E_{FRET}$ for each selected residue pair using the equation $E_{FRET} = 1 / (1 + (R/R_0)^6)$, where R was set to the $C_\alpha$-$C_\alpha$ distance in Å between each residue in the residue pair (*i.e.*, an estimate of the expected Cy3-Cy5 distance in the corresponding smFRET experiment) and $R_0$ was 55 Å (*i.e.*, the Förster radius for the Cy3-Cy5 FRET donor-acceptor pair[45]). For each residue pair, we then calculated the predicted $|\Delta E_{FRET}|$ using the equation $|\Delta E_{FRET}| = |E_{FRET}$ (non-swiveled) $- E_{FRET}$ (swiveled)|, where $E_{FRET}$ (non-swiveled) and $E_{FRET}$ (swiveled) are the predicted $E_{FRET}$s calculated for the high-resolution structures that capture the 30S head domain in non-swiveled and swiveled conformation expected during canonical translation elongation[17], respectively. Finally, each residue in the residue pairs for which the predicted $|\Delta E_{FRET}|$ was ≥ 0.2 was phylogenetically analyzed. Each residue from residue pairs for which both residues had < 70% sequence identity were then used to generate the final list of labeling positions. The final list of labeling positions can be found in Table 1 and their locations within the ribosome can be found in Fig. 2.

To identify labeling positions that would report on MT, we used a single high-resolution structure in which the 3'- and 5' ends of the mRNA within the mRNA entry and exit channels of the 30S subunit, respectively, were ordered (PDB ID 4V4J)[18]. Following a procedure analogous to that described in the previous paragraph, we used this structure to visually identify surface-accessible amino acid residues in the ribosomal proteins in the body domain of the 30S subunit whose $C_\alpha$s were within 38–69 Å of the 5'-hydroxyl O of the 5' nucleotide or the 3'-hydroxyl O of the 3' nucleotide of the mRNA. Ribosomal protein $C_\alpha$s within this distance range of these mRNA nucleotide Os would be expected to exhibit a predicted $E_{FRET}$ of > 0.2 when the $C_\alpha$ is labeled with Cy3 and the mRNA is labeled with Cy5. Once again, only residues exhibiting < 70% sequence identity were used to generate the final list of labeling positions (Fig. 2, Table 1).

### Multiplexed genome engineering cycles

Thirteen oligonucleotides targeting different codon sites in ribosomal protein genes for mutation to TAG codons were designed as previously described[6, 27], and purchased from Integrated DNA Technologies Inc. (Supplementary Table 1). MGE cycles, each of which take 3–5 hours, were performed using the RNase I-deficient C321 ΔA or EcNR2 strains that had been transformed with the pEvol-pAzFRS.2.t1 or pULTRA-pAzFRS.2.t1 plasmids, respectively, and following a previously published procedure[6, 27]. All of the media used for MGE contained 1 mM *p*-AzF and either 0.2 % arabinose (for experiments with the RNase I-deficient C321 ΔA strain) or 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) (for experiments with the EcNR2 strain). Cultures at the end of 6–8 MGE cycles were used to screen for mutant strains using MASC-PCR (*vide infra*). If a double mutant was needed, as in the case of generating the HS signals, one of the successfully engineered single-mutant strains was subjected to single-plexed genome engineering using the oligonucleotide targeting the second position.

### Screening for mutants and calculation of percent enrichment

MASC-PCR was used to screen for mutant strains in the pool of cells present in the culture after 6–8 MGE cycles, using previously published methods[6, 27]. Briefly, several dilutions of the culture after MGE cycles 6–8 were plated on LB-agar plates containing the appropriate antibiotics, 1 mM *p*-AzF, and either 0.2% (w/v) arabinose (for experiments with the RNase I-deficient C321 A strain) or 1 mM IPTG (for experiments with the EcNR2 strain) so as to obtain single, well-distributed colonies. 48 to 96 colonies were then selected and screened for the presence of each mutation using MASC-PCR. To perform the MASC-PCR, we used two pools of PCR forward primers where 3' ends of each primer in the pool either annealed to wildtype or mutant sequence in the genome, and the reverse primers were designed to produce PCR products of distinct, pre-defined lengths for each targeted position. Colonies that tested positive for mutations in the MASC-PCR screening were sequenced by Sanger sequencing (Genewiz) in order to confirm the presence and identity of the mutation. The percent enrichment of a particular mutation was calculated by dividing the number of confirmed mutant colonies by the total number of colonies screened and multiplying the result by 100.

### Ribosome purification and labeling

70S ribosomes were purified using previously established procedures[47] with slight modifications. 1 mL culture of C321 A wt or isolated mutant strain was used to inoculate 1L of 2× Yeast Tryptone media (2× YT) containing 1mM *p*-AzF, and 0.2 % arabinose. The culture was grown to mid-log phase (OD 0.5), and the cells were harvested by centrifugation at 5000×g for 10 min. The actively translating 70S ribosomes and polysomes were then purified from the harvested cells as previously described[47].

In order to efficiently label the ribosomal subunits with DBCO-derivatized Cy3 and/or Cy5, the purified 70S ribosomal pellet was gently resuspended in Low Magnesium Buffer (20 Tris(hydroxymethyl)aminomethane (Tris) acetate (OAc) at a pH at room temperature (i.e.,~25 °C) ($pH_{RT}$) = 7.5, 60 mM ammonium chloride ($NH_4Cl$), 1 mM magnesium acetate ($Mg(OAc)_2$), 0.5 mM ethylenediaminetetraacetic acid (EDTA), and 6 mM 2-mercaptoethanol (β-ME)) such that the 70S ribosomes would dissociate into their component 30S and 50S subunits and the resulting solution was diluted such that the final concentrations of 30S and 50S subunits was 1 μM each. DBCO-derivatized Sulfo-Cy3 and/or DBCO-derivatized Sulfo-Cy5 (Lumiprobe) were then added to the diluted 30S and 50S subunits to a final concentration of 10 μM. The labeling reaction was carried out at 4 °C for 12 hours in the dark. The labeling reaction was then dialyzed against Low Magnesium Buffer to remove excess, unconjugated fluorophores. Labeled 30S and/or 50S subunits were then purified from the labeling reactions as previously reported using sucrose gradient ultracentrifugation in Low Magnesium Buffer[30]. Labeling efficiencies were calculated by interpolating the cumulative fluorescence intensity of each labeled protein band from an SDS-PAGE gel of labeled 30S and/or 50S subunits from linear standard curves generated using Cy3- or Cy5-labeled protein standards of defined quantities and labeling efficiencies. Specifically, the Cy3- and Cy5-labled protein standards we used were L1 Q18C-Cy3 (L1 [Cy3]) and RF1 S167C-Cy5 (RF1 [Cy5]), respectively, that were isolated, labeled, and

purified so as to exhibit 100 % Cy3- or Cy5 labeling efficiencies as previously described[29, 48] (Supplementary Fig. 3).

### Preparation of mRNAs, tRNAs, and translation factors

Ribosomal complexes for smFRET experiments using the HS and IR smFRET signals were assembled on a previously described[49, 50], 5'-biotinylated model mRNA that is a variant of the mRNA encoding bacteriophage T4 gene product 32 (Bio-mRNA). Bio-mRNA was chemically synthesized (IDT) (Supplementary Table 5). Ribosomal complexes for smFRET experiments using the MT smFRET signal were assembled on a previously described[47], non-biotinylated model mRNA consisting of the first twenty codons of bacteriophage T4 gene product 32 (NonBio-mRNA) that was hybridized to a previously described[47], 5'-Cy5-labeled, 3'-biotinylated DNA oligonucleotide (Cy5-DNA-Bio) (Supplementary Table 5). NonBio-mRNA was synthesized by *in vitro* transcription using bacteriophage T7 RNA polymerase, as previously described[30], and Cy5-DNA-Bio was chemically synthesized (IDT). Hybridization of NonBio-mRNA with Cy5-DNA-Bio was performed as previously described[30]. The hybridized NonBio-mRNA:Cy5-DNA-Bio was purified away from excess, unhybridized Cy5-DNA-Bio using size-exclusion chromatography. *E. coli* initiator, formylmethionine-specific tRNA (tRNA$^{fMet}$) was purchased from MP Bio and was aminoacylated and formylated using methionyl-tRNA synthetase and methionyl-tRNA formyltranferase, respectively, using previously published procedures[47]. *E. coli* translation initiation factors (IFs) 1, 2, and 3 were expressed and purified using established protocols as described elsewhere[30].

### Preparation of ribosomal complexes for smFRET experiments

To perform smFRET experiments using the HS smFRET signal, we attempted to prepare a 30S IC similar to one that has been described as exhibiting HS dynamics in a recently published cryogenic electron microscopy (cryo-EM) study[51]. To do this, we adapted general protocols that we have previously developed for the preparation of 30S ICs[28, 52]. Specifically, we incubated 0.6 μM 30S subunits labeled with Cy3 and Cy5 (30S HS1 [Cy3/Cy5]), 1.8 μM Bio-mRNA (Supplementary Table 5), 0.9 μM fMet-tRNA$^{fMet}$, 0.9 μM IF1, and 0.6 μM IF3 in Tris-Polymix Buffer (50 mM Tris-OAc at pH$_{RT}$ = 7.5, 100 mM KCl, 5 mM ammonium acetate (NH$_4$OAc), 5 mM Mg(OAc)$_2$, 0.1 mM EDTA, 1 mM guanosine triphosphate (GTP), 5 mM putrescine-HCl, 1 mM spermidine-free base, and 6 mM β-ME) for 10 minutes at 37 °C and subsequently transferred the reaction to ice for an additional 5 minutes. The complexes were divided into 1 μL aliquots and flash-frozen by immersing the tube in liquid nitrogen. The aliquots were stored at –80 °C for future use[28, 52]. Given these reaction conditions, the fact that IF1 and IF3 were maintained at 1 μM and 25 nM, respectively, in the buffers used throughout all complex formation, dilution, tethering, and imaging steps (*vide infra*), and the results of previous biochemical and smFRET work by us and others[49], we expect this reaction to predominantly yield the IF1- and IF3-containing 30S IC that is schematized in Fig. 4a (Top Panel).

To perform smFRET experiments using the MT smFRET signal, we attempted to prepare a 30S IC similar to one that we have described in a recent smFRET study[49], that would serve as a preliminary proof-of-concept model for MT signal. To do this, we again adapted general

protocols that we have previously developed for the preparation of 30S ICs[28, 49, 52]. Specifically, we incubated 0.6 μM 30S subunits labeled with Cy3 (30S MT1 [Cy3]), 1.8 μM NonBio-mRNA:Cy5-DNA-Bio (Supplementary Table 5), 0.9 μM fMet-tRNA$^{fMet}$, 0.9 μM IF1, and 0.9 μM IF2 in Tris-Polymix Buffer for 10 minutes at 37 °C and subsequently transferred the reaction to ice for an additional 5 minutes. The complexes were divided into 1 μL aliquots and flash-frozen by immersing the tubes in liquid nitrogen. The aliquots were stored at –80 °C for future use[28, 52]. Given these reaction conditions, the fact that IF1 and IF2 were maintained at 1 μM and 25 nM, respectively, in the buffers used throughout all complex formation, dilution, tethering, and imaging steps, and the results of previous smFRET studies[49], we expect this reaction to predominantly yield the IF1-, IF2-, and fMet-tRNA$^{fMet}$-containing 30S IC that is schematized in Fig 4b (Top Panel).

To perform smFRET experiments using the IR smFRET signal, we attempted to prepare a 70S pre-translocation complex mimic lacking a peptidyl-tRNA in the ribosomal aminoacyl-tRNA binding site (*i.e.*, a 70S PRE$^{-A}$ complex) similar to one that has been described as exhibiting IR dynamics in a previously published smFRET study[10]. To do this, we adapted general protocols that we have previously developed for the non-enzymatic preparation of 70S PRE$^{-A}$ complexes[50]. Specifically, we incubated 15 pmol 30S subunits labeled with Cy5 (30S IR1 [Cy5]), 30 pmol Bio-mRNA (Supplementary Table 5), and 20 pmol deacylated tRNA$^{fMet}$ in 30 μL of 70S PRE$^{-A}$ Assembly Buffer (50 mM Tris hydrochloride (Tris-HCl) at pH$_{RT}$ = 7.5, 70 mM NH$_4$OAc, 30 mM KCl, 6 mM β-ME, and 7 mM MgCl$_2$) for 10 minutes at 37 °C, at which point 10 pmol 50S subunits labeled with Cy3 (50S IR1 [Cy3]) were added and the reaction incubated for additional 20 minutes at 37 °C. The reaction was then placed on ice for 5 minutes and diluted to a final volume of 100 μL with Tris-Polymix Buffer that had been adjusted to 20 mM Mg(OAc)$_2$. The reaction was carefully layered on top of 10–40% (w/v) sucrose gradient made in Tris-Polymix Buffer adjusted to 20 mM Mg(OAc)$_2$ and purified by density gradient ultracentrifugation as previously described[29, 50]. The complexes were divided into 25 μL aliquots and flash-frozen by immersing the tubes in liquid nitrogen. The aliquots were stored at –80 °C for future use. Given these reaction conditions and the results of previous smFRET studies[50], we expect this reaction to predominantly yield the 70S PRE$^{-A}$ that is schematized in Fig. 4c (Top Panel).

### TIRF-based smFRET experiments

Ribosomal complexes were diluted to ~100–500 pM in Tris-Polymix Buffer adjusted to 5 mM Mg(OAc)$_2$ for smFRET experiments using the HS1 and MT1 smFRET signals and to 15 mM Mg(OAc)$_2$ for smFRET experiments using the IR1 smFRET signal and supplemented with 1 μM IF1 and 25 nM IF3 for smFRET experiments using the HS1 smFRET signal and 1 μM IF1 and 25 nM IF2 for smFRET experiments using the MT1 smFRET signal. The complexes were tethered to the PEG/Biotin-PEG-derivatized surfaces of our quartz microfluidic flow-cells *via* a biotin-streptavidin-biotin bridge by incubating the diluted complexes in our flow-cells for 5 minutes, after which unbound complexes were flushed out of the flow-cells using the same Tris-Polymix Buffers that were used to dilute the complexes, but that had been further supplemented with an oxygen-scavenging system (5 mM protocatechuic acid and 10 nM protocatechuate-3,4-dioxygenase) and a triplet state quencher cocktail (1 mM cyclooctatetraene and 1 mM nitrobenzoic acid).

The tethered complexes were then imaged using a laboratory-built, wide-field, prism-based TIRF microscope with a diode-pumped solid-state 532 nM laser (Laser Quantum GEM532) as an excitation source for Cy3. Fluorescence emissions from Cy3 and Cy5 were collected through a 60× magnification, water-immersion objective with a numerical aperture of 1.2 (Nikon), wavelength separated using a Dual-View image-splitting device (Photometrics), and imaged using an water-cooled, electron-multiplying charged coupled device (EMCCD) camera (Andor iXon Ultra 888) operating with 2× binning. 600-frame movies were collected at a time resolution of 0.1 second using μManager[53].

TIRF movies were analyzed using custom-written software (manuscript in preparation; Jason Hon, Colin Kinz-Thompson, R.L.G.). First, fluorophores were identified by locating local maxima pixels in the movie and classifying them into either 'fluorophore' or 'background' classes. The Cy3 and Cy5 imaging channels were then aligned by applying a polynomial transformation that had been separately computed using a control image of fiducial markers. Using the aligned Cy3 and Cy5 imaging channels, we next fit each Cy3 and Cy5 fluorophore in each image of the movie to a 2D Gaussian in order to estimate the Cy3 and Cy5 fluorescence intensity *versus* time trajectories for each identified and aligned pair of Cy3 and Cy5 fluorophores. For each time point, Cy5 fluorescence intensity values were corrected for Cy3 bleedthrough by subtracting 5% of the Cy3 florescence intensity value in the corresponding Cy3 fluorescence intensity trajectory. $E_{FRET}$ *versus* time trajectories were then generated by using the Cy3 fluorescence intensity trajectories and bleedthrough-corrected Cy5 fluorescence intensity trajectories to calculate the $E_{FRET}$ value at each time point of the corresponding $E_{FRET}$ trajectory. The $E_{FRET}$ values were calculated by dividing the Cy5 fluorescence intensity ($I_{Cy5}$) by the sum of the Cy3 and Cy5 fluorescence intensities ($I_{Cy5} + I_{Cy3}$), as previously described[54]. Visual inspection was then used to select only those $E_{FRET}$ trajectories for which the corresponding Cy3 and Cy5 fluorescence intensity trajectories exhibited single-step photobleaching of both the Cy3 and Cy5 fluorophores, thereby confirming that the $E_{FRET}$ trajectory originated from a pair of single donor and acceptor fluorophores and not a single donor fluorophore with no corresponding acceptor fluorophore or a pair of donor fluorophores. Idealized $E_{FRET}$ *versus* time trajectories were generated using Bayesian inference-based hidden Markov modeling with 3 state- (HS1 and IR1 data) or 2 state- (MT1 data) models, as previously described[55].

### Code Availability

The code used to analyze the TIRF movies in this study is associated with a manuscript in preparation (Jason Hon, Colin Kinz-Thompson, R.L.G.), and is consequently available from the corresponding author upon request.

### Data Availability

With the exception of the smFRET data, all other data supporting the findings of this study are presented within this article. Due to the lack of a public repository for smFRET data, the smFRET data supporting the findings of this study are available from the corresponding author upon request.

# Extended Data
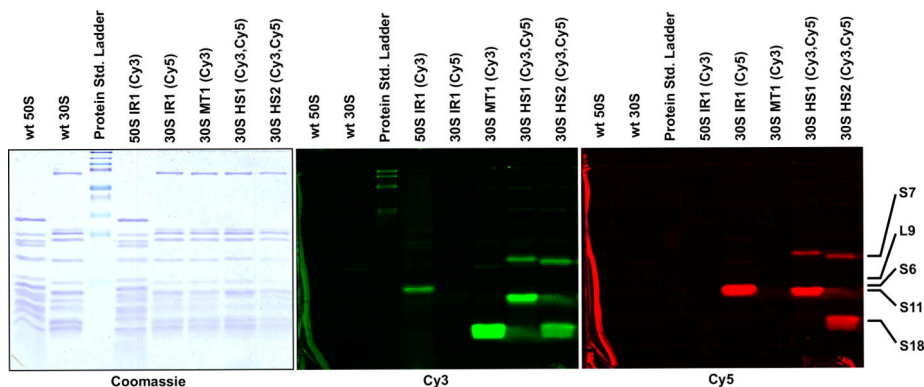


**Extended Data Figure 1. Site-specific Cy3 and/or Cy5 labeling of ribosomes purified from wildtype *versus* genomic mutant strains.**

SDS-PAGE analysis of ribosomal proteins derived from 30S or 50S subunits isolated from the wildtype (wt) strain (Lanes 1 and 2) or the IR1, MT1, HS1, and HS2 mutant strains (Lanes 4–8) and reacted with DBCO-derivatized Cy3 and/or Cy5 fluorophores as shown in Fig. 3. Left panel shows visible light scan of Coomassie-stained gel. Middle and right panels show fluorescence emission scans of pre-Coomassie-stained gel using excitation wavelengths of 532 nm for Cy3 (Middle Panel) and 635 nm for Cy5 (Right Panel). The position at which each labeled ribosomal protein is expected to run on the SDS-PAGE gel was determined using a standard protein molecular weight ladder, and is indicated along the right side of the figure. The experiment shown here was replicated a total of three times, with similar results each time.
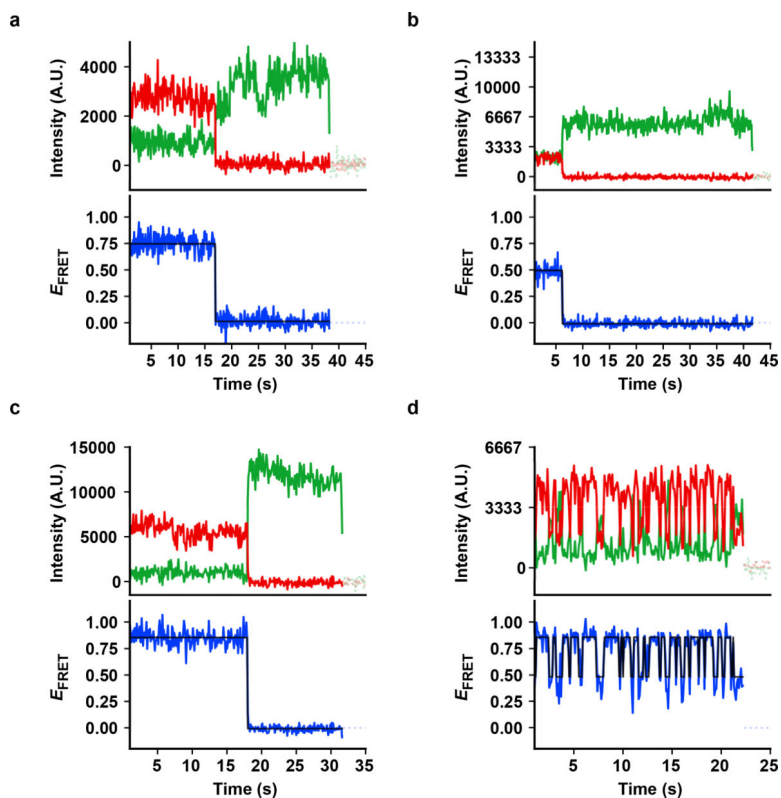
**Extended Data Figure 2. Idealized $E_{\text{FRET}}$ *versus* time trajectories generated using Bayesian inference-based hidden Markov modeling.**

Representative Cy3- and Cy5 fluorescence intensity *versus* time trajectories (Top Sub-Panel) and corresponding $E_{\text{FRET}}$ *versus* time trajectories (Bottom Sub-Panel) for smFRET experiments performed on ribosomal complexes assembled using Cy3- and/or Cy5-labeled 30S and/or 50S subunits isolated from the (a and b) HS1, (c) MT1, and (d) IR1 strains as shown in Fig. 4. In the Cy3 and Cy5 fluorescence intensity *versus* time trajectories, the Cy3 and Cy5 fluorescence intensities are shown as green and red curves, respectively. In the $E_{\text{FRET}}$ *versus* time trajectories, the $E_{\text{FRET}}$ is shown as blue curves. The idealized $E_{\text{FRET}}$ versus time trajectory is shown as black lines.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENTS

## REFERENCES

1. Frank J & Gonzalez RL Jr. Structure and dynamics of a processive Brownian motor: the translating ribosome. Annu Rev Biochem 79, 381–412 (2010). [PubMed: 20235828]

2. Shajani Z, Sykes MT & Williamson JR Assembly of bacterial ribosomes. Annu Rev Biochem 80, 501–526 (2011). [PubMed: 21529161]

3. Shoji S, Dambacher CM, Shajani Z, Williamson JR & Schultz PG Systematic chromosomal deletion of bacterial ribosomal protein genes. J Mol Biol 413, 751–761 (2011). [PubMed: 21945294]

4. Lotze J, Reinhardt U, Seitz O & Beck-Sickinger AG Peptide-tags for site-specific protein labelling in vitro and in vivo. Mol Biosyst 12, 1731–1745 (2016). [PubMed: 26960991]

5. Culver GM & Noller HF Efficient reconstitution of functional Escherichia coli 30S ribosomal subunits from a complete set of recombinant small subunit ribosomal proteins. RNA 5, 832–843 (1999). [PubMed: 10376881]

6. Wang HH et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460, 894–898 (2009). [PubMed: 19633652]

7. Lang K & Chin JW Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. Chem Rev 114, 4764–4806 (2014). [PubMed: 24655057]

8. Lang K & Chin JW Bioorthogonal reactions for labeling proteins. ACS Chem Biol 9, 16–20 (2014). [PubMed: 24432752]

9. Ermolenko DN et al. Observation of intersubunit movement of the ribosome in solution using FRET. J Mol Biol 370, 530–540 (2007). [PubMed: 17512008]

10. Cornish PV, Ermolenko DN, Noller HF & Ha T Spontaneous intersubunit rotation in single ribosomes. Mol Cell 30, 578–588 (2008). [PubMed: 18538656]

11. Wasserman MR, Alejo JL, Altman RB & Blanchard SC Multiperspective smFRET reveals rate-determining late intermediates of ribosomal translocation. Nat Struct Mol Biol 23, 333–341 (2016). [PubMed: 26926435]

12. Sharma H et al. Kinetics of Spontaneous and EF-G-Accelerated Rotation of Ribosomal Subunits. Cell Rep 16, 2187–2196 (2016). [PubMed: 27524615]

13. Wang L et al. Allosteric control of the ribosome by small-molecule antibiotics. Nat Struct Mol Biol 19, 957–963 (2012). [PubMed: 22902368]

14. Marshall RA, Dorywalska M & Puglisi JD Irreversible chemical steps control intersubunit dynamics during translation. Proc Natl Acad Sci U S A 105, 15364–15369 (2008). [PubMed: 18824686]

15. Dorywalska M et al. Site-specific labeling of the ribosome for single-molecule spectroscopy. Nucleic Acids Res 33, 182–189 (2005). [PubMed: 15647501]

16. Rozov A, Westhof E, Yusupov M & Yusupova G The ribosome prohibits the G*U wobble geometry at the first position of the codon-anticodon helix. Nucleic Acids Res 44, 6434–6441 (2016). [PubMed: 27174928]

17. Zhou J, Lancaster L, Donohue JP & Noller HF How the ribosome hands the A-site tRNA to the P site during EF-G-catalyzed translocation. Science 345, 1188–1191 (2014). [PubMed: 25190797]

18. Korostelev A et al. Interactions and dynamics of the Shine Dalgarno helix in the 70S ribosome. Proc Natl Acad Sci U S A 104, 16840–16843 (2007). [PubMed: 17940016]

19. Jewett JC, Sletten EM & Bertozzi CR Rapid Cu-free click chemistry with readily synthesized biarylazacyclooctynones. J Am Chem Soc 132, 3688–3690 (2010). [PubMed: 20187640]

20. Lajoie MJ et al. Genomically recoded organisms expand biological functions. Science 342, 357–360 (2013). [PubMed: 24136966]

21. Yu D et al. An efficient recombination system for chromosome engineering in Escherichia coli. Proc Natl Acad Sci U S A 97, 5978–5983 (2000). [PubMed: 10811905]

22. Amiram M et al. Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. Nat Biotechnol 33, 1272–1279 (2015). [PubMed: 26571098]

23. Chin JW et al. Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. J Am Chem Soc 124, 9026–9027 (2002). [PubMed: 12148987]

24. Wade HE & Robinson HK The inhibition of ribosomal ribonuclease by bacterial ribosomes. Biochem J 97, 747–753 (1965). [PubMed: 5326718]

25. Kurylo CM et al. Genome Sequence and Analysis of Escherichia coli MRE600, a Colicinogenic, Nonmotile Strain that Lacks RNase I and the Type I Methyltransferase, EcoKI. Genome Biol Evol 8, 742–752 (2016). [PubMed: 26802429]

26. Cammack KA & Wade HE The sedimentation behaviour of ribonuclease-active and -inactive ribosomes from bacteria. Biochem J 96, 671–680 (1965). [PubMed: 5324303]

27. Wang HH & Church GM Multiplexed Genome Engineering and Genotyping Methods: Applications for Synthetic Biology and Metabolic Engineering. Method Enzymol 498, 409–426 (2011).

28. Elvekrog MM & Gonzalez RL Jr. Conformational selection of translation initiation factor 3 signals proper substrate selection. Nat Struct Mol Biol 20, 628–633 (2013). [PubMed: 23584454]

29. Fei J, Kosuri P, MacDougall DD & Gonzalez RL Jr. Coupling of ribosomal L1 stalk and tRNA dynamics during translation elongation. Mol Cell 30, 348–359 (2008). [PubMed: 18471980]

30. Fei J et al. A highly purified, fluorescently labeled in vitro translation system for single-molecule studies of protein synthesis. Methods Enzymol 472, 221–259 (2010). [PubMed: 20580967]

31. Milon P, Maracci C, Filonava L, Gualerzi CO & Rodnina MV Real-time assembly landscape of bacterial 30S translation initiation complex. Nat Struct Mol Biol 19, 609–615 (2012). [PubMed: 22562136]

32. Antoun A, Pavlov MY, Lovmar M & Ehrenberg M How initiation factors tune the rate of initiation of protein synthesis in bacteria. EMBO J 25, 2539–2550 (2006). [PubMed: 16724118]

33. Tourigny DS, Fernandez IS, Kelley AC & Ramakrishnan V Elongation factor G bound to the ribosome in an intermediate state of translocation. Science 340, 1235490 (2013). [PubMed: 23812720]

34. Ratje AH et al. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature 468, 713–716 (2010). [PubMed: 21124459]

35. Graf M et al. Visualization of translation termination intermediates trapped by the Apidaecin 137 peptide during RF3-mediated recycling of RF1. Nat Commun 9, 3053 (2018). [PubMed: 30076302]

36. Ramrath DJ et al. The complex of tmRNA-SmpB and EF-G on translocating ribosomes. Nature 485, 526–529 (2012). [PubMed: 22622583]

37. Adiego-Perez B et al. Multiplex genome editing of microorganisms using CRISPR-Cas. FEMS Microbiol Lett 366 (2019).

38. DiCarlo JE et al. Yeast oligo-mediated genome engineering (YOGE). ACS Synth Biol 2, 741–749 (2013). [PubMed: 24160921]

39. Cong L et al. Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823 (2013). [PubMed: 23287718]

40. Reyon D et al. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30, 460–465 (2012). [PubMed: 22484455]

41. Thompson DB et al. The Future of Multiplexed Eukaryotic Genome Engineering. ACS Chem Biol 13, 313–325 (2018). [PubMed: 29241002]

42. Carr PA et al. Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection. Nucleic Acids Res 40, e132 (2012). [PubMed: 22638574]

43. Ambrogelly A et al. Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. Proc Natl Acad Sci U S A 104, 3141–3146 (2007). [PubMed: 17360621]

44. Selmer M et al. Structure of the 70S ribosome complexed with mRNA and tRNA. Science 313, 1935–1942 (2006). [PubMed: 16959973]

45. Murphy MC, Rasnik I, Cheng W, Lohman TM & Ha T Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. Biophys J 86, 2530–2537 (2004). [PubMed: 15041689]

46. DeLano WL Pymol: An open-source molecular graphics tool. CCP4 Newsletter on Protein Crystallography 40, 82–92 (2002).

47. Blanchard SC, Gonzalez RL, Kim HD, Chu S & Puglisi JD tRNA selection and kinetic proofreading in translation. Nat Struct Mol Biol 11, 1008–1014 (2004). [PubMed: 15448679]

48. Sternberg SH, Fei J, Prywes N, McGrath KA & Gonzalez RL Jr. Translation factors direct intrinsic ribosome dynamics during translation termination and ribosome recycling. Nat Struct Mol Biol 16, 861–868 (2009). [PubMed: 19597483]

49. Wang J, Caban K & Gonzalez RL Jr. Ribosomal initiation complex-driven changes in the stability and dynamics of initiation factor 2 regulate the fidelity of translation initiation. J Mol Biol 427, 1819–1834 (2015). [PubMed: 25596426]

50. Fei J, Richard AC, Bronson JE & Gonzalez RL Jr. Transfer RNA-mediated regulation of ribosome dynamics during protein synthesis. Nat Struct Mol Biol 18, 1043–1051 (2011). [PubMed: 21857664]

51. Hussain T, Llacer JL, Wimberly BT, Kieft JS & Ramakrishnan V Large-Scale Movements of IF3 and tRNA during Bacterial Translation Initiation. Cell 167, 133–144 e113 (2016). [PubMed: 27662086]

52. Caban K, Pavlov M, Ehrenberg M & Gonzalez RL Jr. A conformational switch in initiation factor 2 controls the fidelity of translation initiation in bacteria. Nat Commun 8, 1475 (2017). [PubMed: 29133802]

53. Edelstein AD et al. Advanced methods of microscope control using muManager software. J Biol Methods 1, e10 (2014). [PubMed: 25606571]

54. Ning W, Fei J & Gonzalez RL Jr. The ribosome uses cooperative conformational changes to maximize and regulate the efficiency of translation. Proc Natl Acad Sci U S A 111, 12073–12078 (2014). [PubMed: 25085895]

55. Bronson JE, Fei J, Hofman JM, Gonzalez RL Jr. & Wiggins CH Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. Biophys J 97, 3196–3205 (2009). [PubMed: 20006957]
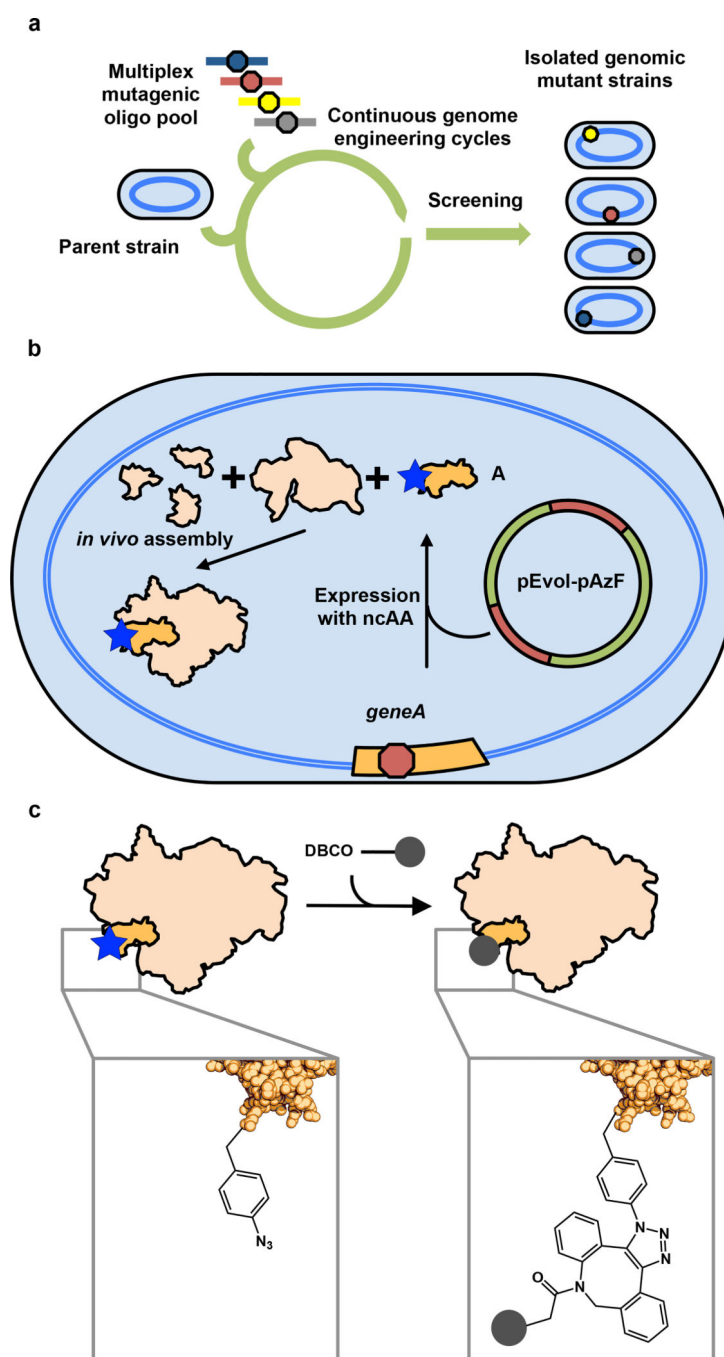
**Figure 1. General overview of multiplexed, bioorthogonal labeling of MBCs using genomically encoded ncAAs.**

(a) Iterative MGE cycles introduce orthogonal codon mutations (hexagons) at specific genomic positions in genes encoding MBC proteins in a background of wildtype parent strain. (b) *In vivo* expression and assembly of the MBC in each mutant strain, including the MBC protein carrying the incorporated ncAA (blue star), is achieved by performing the MGE cycles in the presence of a plasmid expressing a ncAA-specific, orthogonal tRNA-tRNA synthetase pair and in the presence of the ncAA in the growth media such that each resulting mutant strain can assemble MBCs carrying the ncAA at one or more of the targeted

positions. In our case, we have used *p*-AzF as the ncAA and the pEvol-pAzFRS.2.t1 plasmid to express the corresponding, orthogonal tRNA-tRNA synthetase pair. (c) ncAAs incorporated into MBCs purified from successfully selected mutant strains can be conjugated to an appropriately derivatized label or reporter (dark-grey) using bioorthogonal chemistry. In our case, we have used the strain-promoted, azide-alkyne, bioorthogonal conjugation reaction of *p*-AzF with DBCO-derivatized Cy3 and/or Cy5 fluorophores.
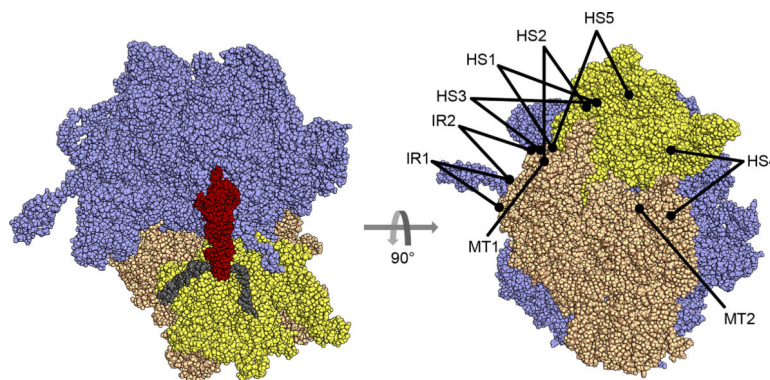
**Figure 2. Locations of amino acid residues in ribosomal proteins that were targeted for labeling.**
The locations of amino acid residues in ribosomal proteins that were targeted for labeling are denoted by the black filled circles at the ends of the black lines on the ribosomal complex shown in the left panel and are labeled according to the ribosome dynamics that they are expected to report on: intrasubunit rotation of the head domain of the 30S subunit relative to the body domain (*i.e.*, 'head swiveling', HS); movement of a translating ribosome along its mRNA template (*i.e.*, 'mRNA translocation', MT); and rotation of the 50S subunit relative to the 30S subunit (*i.e.*, 'intersubunit rotation', IR). The structure shown here is that of an atomic-resolution, X-ray crystallographic structure of a *Thermus thermophilus* ribosomal complex (PDB ID: 5IBB) that is shown as a space-filling model. The head domain of the 30S subunit is shown in yellow, the body domain of the 30S subunit is shown in tan, the 50S subunit is shown in light blue, the P site-bound tRNA is shown in dark red, and the mRNA is shown in grey.
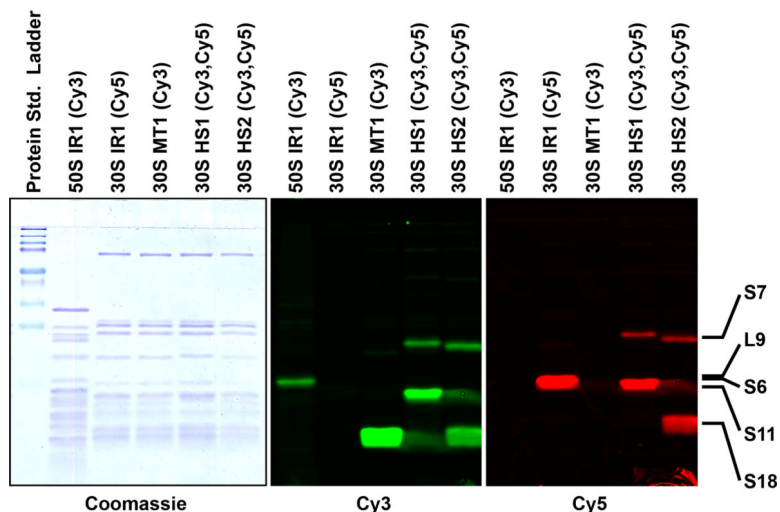
**Figure 3. Site-specific Cy3 and/or Cy5 labeling of ribosomes purified from genomic mutant strains.**

SDS-PAGE analysis of ribosomal proteins derived from 30S or 50S subunits isolated from the IR1, MT1, HS1, and HS2 mutant strains and reacted with DBCO-derivatized Cy3 and/or Cy5 fluorophores. Left panel shows visible light scan of Coomassie-stained gel. Middle and right panels show fluorescence emission scans of pre-Coomassie-stained gel using excitation wavelengths of 532 nm for Cy3 (Middle Panel) and 635 nm for Cy5 (Right Panel). The position at which each labeled ribosomal protein is expected to run on the SDS-PAGE gel was determined using a standard protein molecular weight ladder, and is indicated along the right side of the figure. The experiment shown here was replicated a total of three times, with similar results each time.
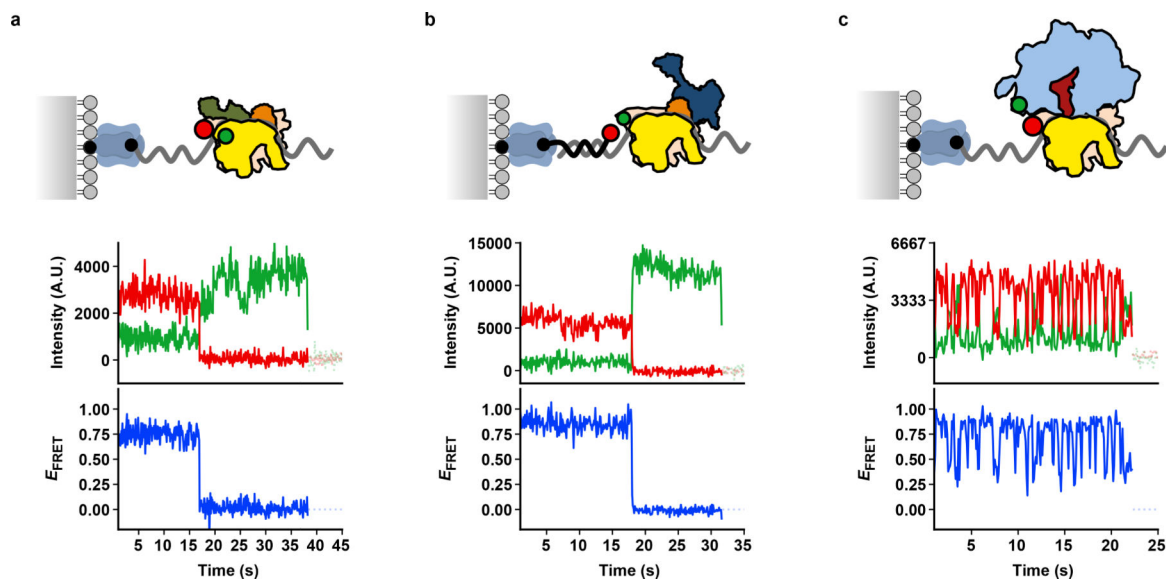
**Figure 4. smFRET experiments using Cy3- and Cy5-labled ribosomal complexes assembled using ribosomes isolated from mutant strains.**

Schematics (Top Panel), representative Cy3 and Cy5 fluorescence intensity *versus* time trajectories (Center Panel), and corresponding $E_{FRET}$ *versus* time trajectories (Bottom Panel) for smFRET experiments performed on ribosomal complexes assembled using Cy3- and/or Cy5-labeled 30S and/or 50S subunits isolated from the (a) HS1, (b) MT1, and (c) IR1 mutant strains. In the schematics, the surface of the microfluidic flow-cells are shown as a grey surface, passivating PEG molecules are shown as grey spheres, biotinylated-PEGs are shown as black spheres, streptavidin is shown in blue-grey, mRNAs are shown as grey curves, hybridizing DNA oligonucleotides are shown as black curves, biotins at the 5' end of the mRNAs or the 3' end of the DNAs are shown as black spheres, the head domain of the 30S subunits is shown in yellow, the body domains of the 30S subunits are shown in tan, the 50S subunits are shown in light blue, the deacylated tRNA$^{fMet}$ is shown in dark red, IF1 is shown in orange, IF2 is shown in dark blue, IF3 is shown in dark green, the Cy3 fluorophores are shown as green spheres, and the Cy5 fluorophores are shown as red spheres. In the Cy3 and Cy5 fluorescence intensity *versus* time trajectories, the Cy3 and Cy5 fluorescence intensities are shown as green and red curves, respectively. In the $E_{FRET}$ *versus* time trajectories, the $E_{FRET}$ is shown as blue curves.

**Table 1.**

Percent enrichment of mutants after the specified number of MGE cycles

| smFRET signal | Target for mutation | Number of bp mutated | Number of MGE cycles | Percent enrichment[1] |
|---|---|---|---|---|
| HS1 | S7 G112 | 3 | 8 | 1.0 |
| | S11 A102 | 3 | 6 | 1.0 |
| HS2 | S7 K131 | 2 | 8 | 2.1 |
| | S18 R8 | 3 | 6 | 2.1 |
| HS3 | S7 G112 | 3 | 8 | 1.0 |
| | S18 Q75 | 1 | 8 | 0 |
| HS4 | S19 Q56 | 1 | 8 | 3.1 |
| | S12 K108 | 2 | 6 | 4.6 |
| HS5 | S13 D11 | 2 | 6 | 0 |
| | S11 R106 | 3 | 8 | 0 |
| MT1 | S18 R8 | 3 | 6 | 2.1 |
| MT2 | S5 E10 | 2 | 8 | 1.0 |
| IR1 | L9 N11 | 2 | 8 | 4.2 |
| | S6 D41 | 2 | 8 | 1.0 |
| IR2 | L9 N11 | 2 | 8 | 4.2 |
| | S11 E76 | 2 | 8 | 1.0 |
| Control | RNase I A9 | 3 | 4 | 10.4 |

[1] Percent enrichment was measured by dividing the number of mutant colonies detected by MASC-PCR and confirmed via Sanger sequencing by the total number of colonies that were screened by MASC-PCR and multiplying the result by 100.

**Table 2.**

Predicted distances, predicted $E_{FRET}$ values, and observed $E_{FRET}$ values.

| smFRET signal | Labeling positions | Conformation | Predicted distance (Å)[1] | Predicted $E_{FRET}$ value[2] | Observed $E_{FRET}$ value |
|---|---|---|---|---|---|
| HS1 | S7 G112 | Non-swiveled | ~50 | ~0.6 | ~0.5 |
|  | S11 A102 | Swiveled | ~39 | ~0.9 | ~0.7 |
| MT1 | S18 R8 DNA Oligo | mRNA exit | ~36 | ~0.9 | ~0.8 |
| IR1 | S6 D41 | Non-rotated | ~59 | ~0.4 | ~0.7 |
|  | L9 N11 | Rotated | ~68 | ~0.2 | ~0.5 |

[1]The distances for the non-swiveled and swiveled conformations of the HS1 smFRET signal were calculated using coordinates taken from PBD entries 5IBB[16] and 4W29[17]. The distance for the mRNA conformation of the MT1 smFRET signal was calculated using coordinates taken from PBD entry 4V4J[18]. The distances for the non-rotated and rotated subunit orientations of the IR1 smFRET signal were calculated using coordinates taken from PBD entries 4V51[44] and 4V9H[33].

[2]Predicted $E_{FRET}$ values were calculated using the equation $E_{FRET} = 1 / (1 + (R/R_0)^6)$, where R is distance in angstrom (Å) between the donor and acceptor and $R_0$ is the Förster radius, which is ~55 Å for the Cy3-Cy5 pair[45].