An in vitro tag-and-modify protein sample generation method for single-molecule fluorescence resonance energy transfer

Received for publication, April 17, 2017, and in revised form, July 25, 2017. Published, Papers in Press, July 28, 2017, DOI 10.1074/jbc.M117.791723

Kambiz M. Hamadani1,†, Jesse Howe5, Madeleine K. Jensen*, Peng Wu†, Jamie H. D. Cate1,‡,§,∥ and Susan Marqusee†,‡,∥

From the 1Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, the 4California Institute for Quantitative Biosciences and the Departments of 4Molecular and Cell Biology and 4Chemistry, University of California, Berkeley, California 94720, the 5Department of Chemistry and Biochemistry, California State University, San Marcos, California 92096, and the 1Department of Chemical Physiology, Scripps Research Institute, La Jolla, California 92037

Edited by Wolfgang Peti

Biomolecular systems exhibit many dynamic and biologically relevant properties, such as conformational fluctuations, multistep catalysis, transient interactions, folding, and allosteric structural transitions. These properties are challenging to detect and engineer using standard ensemble-based techniques. To address this drawback, single-molecule methods offer a way to access conformational distributions, transient states, and asynchronous dynamics inaccessible to these standard techniques. Fluorescence-based single-molecule approaches are parallelizable and compatible with multiplexed detection; to date, however, they have remained limited to serial screens of small protein libraries. This stems from the current absence of methods for generating either individual dual-labeled protein samples at high throughputs or protein libraries compatible with multiplexed screening platforms. Here, we demonstrate that by combining purified and reconstituted in vitro translation, quantitative unnatural amino acid incorporation via AUG codon reassignment, and copper-catalyzed azide-alkyne cycloaddition, we can overcome these challenges for target proteins that are, or can be, methionine-depleted. We present an in vitro parallelizable approach that does not require laborious target-specific purification to generate dual-labeled proteins and ribosome-nascent chain libraries suitable for single-molecule FRET-based conformational phenotyping. We demonstrate the power of this approach by tracking the effects of mutations, C-terminal extensions, and ribosomal tethering on the structure and stability of three protein model systems: barnase, spectrin, and T4 lysozyme. Importantly, dual-labeled ribosome-nascent chain libraries enable single-molecule co-localization of genotypes with phenotypes, are well suited for multiplexed single-molecule screening of protein libraries, and should enable the in vitro directed evolution of proteins with designer single-molecule conformational phenotypes of interest.

Biomolecular systems exhibit important dynamic properties, such as conformational fluctuations, multistep catalysis, transient interactions, folding, and allosteric structural transitions that are challenging to detect and engineer using standard ensemble-based techniques. Although single-molecule bio-physics has proven to be a powerful approach to decode these dynamic processes, coupling such single-molecule studies with high-throughput screening and in vitro directed evolution remains a challenge (1, 2). Fluorescence-based single-molecule detection has made considerable progress toward these goals (3–8); however, its application in structure-based protein biophysics and directed evolution is limited by our inability to first generate and then screen large libraries of dye-labeled proteins.

Creating such libraries has been challenging because single-molecule detection methods, such as fluorescence resonance energy transfer (smFRET),4 require compositionally homogeneous5 and often dual site-specifically labeled samples (9). “Tag-and-modify” approaches, which combine unnatural amino acid (UAA) incorporation via genetic code expansion and click chemistry–based dye conjugation, facili-

†To whom correspondence may be addressed. E-mail: khamadani@berkeley.edu.
‡To whom correspondence may be addressed. E-mail: jcate@lbl.gov.
§To whom correspondence may be addressed. E-mail: marqusee@berkeley.edu.

4The abbreviations used are: smFRET, single-molecule FRET; IVT, in vitro translation; prIVT, purified and reconstituted IVT; UAA, unnatural amino acid; CuAAC, copper-catalyzed azide-alkyne cycloaddition; HPG, homopropargylglycine; NTD, N-terminal domain; T4L, T4 lysozyme; Eprox, proximity ratio; S, stoichiometry; ALEX, alternating laser excitation; μsALEX, microsecond ALEX; D, donor; A, acceptor; F, folded; U, unfolded; PTC, peptidyltransferase center; PDA, probability distribution analysis; Bicine, N,N-bis(2-hydroxyethyl)glycine; PDB, Protein Data Bank; BTTES, Bis(tertbutyl)-tris(triazolylmethyl)amine-ethane sulfonic acid; BTTP, Bis(tertbutyl)-tris(triazolylmethyl)amine-propanol.
5Compositional homogeneity refers to the degree of (stereo)chemical uniformity within an ensemble of fluorescently labeled sample molecules. The fidelity of protein synthesis, the efficiency and specificity of UAA tag incorporation, the extent of labeling specificity (site-specific versus residue/group-specific), peptide bond cis-trans isomerization, and even dye labeling regiospecificity or enantiospecificity can all influence sample compositional homogeneity. Acceptable levels of sample compositional homogeneity differ for qualitative and quantitative FRET applications. For the latter, even persistent conformational (as opposed to compositional) sources of heterogeneity, such as dye rotational anisotropy within the donor fluorescence lifetime, can be problematic.
tate site-specific protein labeling (10, 11). However, these approaches generally rely on high-yield cell-based expression, have limited nonsense suppression efficiencies (~1–30%) (12), are subject to nonspecific UAA tagging of host proteins (13, 14), and thus require affinity purification in order to achieve compositional homogeneity (15) (Fig. 1A). Some of these issues have been mitigated by quantitative enzymatic aldehyde tagging and an optimized oxime ligation reaction (16); this approach, however, still remains subject to the various limitations of cell-based cloning and expression.

Here, we overcome the constraints of cell-based approaches using a purified and reconstituted in vitro translation (prIVT) system (17) to establish a high-throughput and cell-free method to generate fluorescently labeled protein samples for smFRET experiments (Fig. 1B). By using a cell-free approach, we avoid the laborious target-specific purification methods required in cell-based cloning and expression approaches. By using PCR to make prIVT templates, we can express UAA-tagged proteins in a parallelizable manner and in quantities appropriate for single-molecule studies, limiting cost and improving throughput. In addition, prIVT systems are free from ribonucleases and proteases, making them well suited to the generation of monovalent genotype/phenotype-linked ribosome display (18) or mRNA display (19) libraries. Most importantly, prIVT systems permit user-controlled elimination of undesirable translation processes, such as peptide release, misincorporation, or off-target expression and UAA incorporation, thereby enabling the quantitative incorporation of UAA tags selectively into desired targets. Our approach is compatible with a number of UAA-tagging strategies ranging from residue-specific sense codon reassignment (20) to site-specific genetic code expansion (21) and genetic code reprogramming (22–24). Here, we employed a simple but readily accessible metabolic AUG codon reassignment approach (25) to achieve quantitative alkyne-UAA incorporation using a commercial prIVT system (PURExpress®). With this method, we generated quantitatively alkyne-tagged released proteins (Fig. 1B, top branch) and ribosome-bound nascent chains (RNCs; Fig. 1B, bottom branch). We then tested a small library of copper-catalyzed azide-alkyne cycloaddition reactions (CuAAC) for efficient and specific dual labeling of these submicromolar targets at mg/ml concentrations of background biomass.

To our knowledge, this is the first time a prIVT-based tag-and-modify method has been used for smFRET sample generation. This approach avoids the in vivo cloning/expression and purification steps, which traditionally limit the generation of dual-labeled protein samples for smFRET. In the case of RNC samples, it also affords monovalent genotype-phenotype-linked protein libraries, which enable molecular barcoding and multiplexing of fluorescence-based single-molecule screens (Fig. 1C). This novel approach should be generalizable to other protein systems with the limitation that it relies on engineering and an optimized oxime ligation reaction (16); this approach, however, still remains subject to the various limitations of cell-based cloning and expression.

Results

Dye attachment via ligand-assisted and Cu⁺-catalyzed azide-alkyne cycloaddition

A major challenge in developing a high-throughput route to smFRET sample generation is the need to alleviate the bottleneck associated with target-specific purification steps. Existing tag-and-modify approaches rely on target purification to resolve the compositional heterogeneity that results during inefficient and nonspecific in vivo UAA incorporation (Fig. 1A). Whereas combinations of sense or nonsense suppression with prIVT enable target- and site-specific as well as quantitative UAA incorporation (21, 26), low target yields and the challenges of labeling unpurified targets still preclude efficient and selective dye conjugation (Fig. 1B). Here we use a combination of established prIVT and metabolic codon reassignment methods and focus on identifying a labeling chemistry with the required sensitivity and specificity for our application.

Our high-throughput approach of using prIVT followed by dye labeling demands selective dye attachment in the presence of diverse prIVT components. Given the target and background biomass concentrations expected from unpurified prIVT reactions (~0.1–0.3 μM target protein, ~2 mg/ml background protein, and ~5 mg/ml background RNA) and no more than a 20-fold excess of free dye over target, we estimate that an on-target rate constant of $k_{on} > 100 \text{ M}^{-1} \text{s}^{-1}$, a 10,000-fold kinetic selectivity against off-target labeling of biological nucleophiles (i.e. $k_{SI}<0.01 \text{ M}^{-1} \text{s}^{-1}$), and minimal nonspecific adsorption of excess free dye during free dye removal will be required for single-molecule fluorescence applications. An evaluation of currently available ligation schemes (supplemental Table S1) shows that ligand-assisted CuAAC has the needed sensitivity and specificity for our applications. Ligand-assisted CuAAC is also regioselective and results in a short, flexible target–probe linker, an advantage for smFRET analyses.

Several ligand-assisted CuAAC reactions theoretically meet the above rate and selectivity criteria. However, initial rates in model CuAAC reactions are often poor predictors of percent completion in bioconjugations due to inhibitory off-pathway copper center aggregation and oxidative ligand inactivation (27–30). Bis(tert-butyl)-modified ligands (Fig. 2A, BTTES and BTTP) minimize copper center aggregation and outperform other ligand classes when accounting for both initial rate and robust completion (31–33). Under aerobic conditions, copper- or dehydroascorbate-mediated oxidative damage can be problematic for sensitive targets, such as RNCs (30, 34, 35). We avoided these issues by labeling under anaerobic conditions (i.e. <10 ppm O₂), which have previously been shown to prevent damage to such highly sensitive biomolecules (36, 37). To achieve quantitative UAA incorporation in a generally accessible manner, we used commercial prIVT technology (PURExpress® Δ(aa, tRNA) from New England Biolabs) and residue-specific sense codon reassignment of methionine by its redox-stable structural analogue homopropargylglycine (HPG; Fig. 2A) (25). Importantly, HPG is an efficient substrate for dual-labeled protein samples for smFRET.
In vitro tag-and-modify sample generation for smFRET

Figure 1. Comparison between cell-based (A) and prIVT (B) tag-and-modify sample generation (A and B) and screening (C) strategies for smFRET protein biophysics. A, cell-based expression is scalable and provides nanomoles of highly heterogeneous product that must be purified prior to labeling. Large yields and target purification relax the sensitivity (on-target rate constant, \( k_{on} \geq 1 \text{ M}^{-1} \text{s}^{-1} \)) and specificity (off-target rate constant, \( k_{off} < 0.01 \text{ M}^{-1} \text{s}^{-1} \)) requirements imposed on labeling and enable the use of many different chemicals for dye attachment. Unfortunately, target purification limits sample generation throughput, whereas cell-based expression is incompatible with monovalent genotype-phenotype–linked library generation as required for multiplexed detection thereby. B, a prIVT approach yields only picomoles of product, but because translation can be easily controlled to achieve quantitative and entirely target-specific UAA incorporation, target purification is unnecessary. Although smaller yields impose stricter sensitivity requirements on labeling (on-target rate constant (\( k_{on} \)) \( > 100 \text{ M}^{-1} \text{s}^{-1} \)), if a chemistry can be found that meets these demands without compromising specificity (off-target rate constant (\( k_{off} \)) \( < 0.01 \text{ M}^{-1} \text{s}^{-1} \)), then higher throughputs and multiplexed screening are enabled. C, there are important trade-offs between serial, parallel, and multiplexed smFRET screening approaches: serial confocal screens offer the highest spatiotemporal resolution at the expense of lower throughputs and more stringent sample generation requirements (e.g. dual-site-specific labeling is required); parallel confocal screening offers enhanced screening rates at slightly lower spatiotemporal resolution and without much multiplexing capability; parallelization via wide-field total internal reflection fluorescence (TIRF) imaging also increases screening rates at the expense of spatiotemporal resolution but again without enabling highly multiplexed sample screening; and the monovalent genotype-phenotype linkage of RNC libraries allows the colocализed single-molecule detection of both the genotype and the phenotype of a given library member and thereby enables one-pot sample multiplexing (e.g. using zero-mode waveguides (ZMW) and single-molecule real-time nucleic acid sequencing/genotyping). Arrow thickness, throughput. UAA tags are shown as gray spheres. Donor and acceptor dyes are shown as blue and red dots.
suggest that sample oxidation/degradation is negligible under anaerobic labeling conditions in agreement with previous findings (36, 37, 50–52). Finally, for temperature- or ascorbate-sensitive targets, we also confirmed that labeling at 4 °C or in the presence of the protective agent aminoguanidine (AG) did not significantly compromise reaction completion (Fig. 2F).

**smFRET on in vitro generated proteins and ribosome-bound nascent chains**

Using this optimized prIVT tag-and-modify protocol (for details, see “Experimental procedures”), we generated libraries of statistically dual-labeled released and SecM-stalled (40) RNC variants of barnase (41), T4L (43), and the R16 domain of chicken/H9251-spectrin (42) (supplemental Table S2). We benchmarked the UAA incorporation and dye labeling efficiency and fidelity with smFRET-ALEX, which is particularly well suited for samples generated by statistical labeling. Each 12.5-μl prIVT reaction provided sufficient product (~10–50 pmol) for multiple labeling reactions with different combinations or relative concentrations of dyes (see below). Labeled samples were serially screened under a variety of solution conditions using one of two diffusion-based confocal smFRET microscopes with microsecond alternating laser excitation (μsALEX) capabilities (Fig. 1C (top) and supplemental Fig. S1) to separate donor-only (D-only) and acceptor-only (A-only) subpopulations from dual-labeled (DA or AD) species (53, 54). 2D EPR-S histograms from 5–20 min of data acquisition illustrate the sample quality achievable using this approach (Figs. 3–6). D-only bursts appear at low EPR and high S values (top left corner of each histogram), whereas A-only bursts are dominated by shot noise along the EPR axis and thus appear as a series of narrow vertical lines at low S values (bottom of each histogram) (55). For each set of samples, we carried out a control translation/labeling reaction using a single-AUG mRNA template. If either HPG incorporation or dye labeling lacks specificity, we would expect the single-AUG template to yield some dual-labeled bursts. If HPG incorporation or labeling is inefficient, we would expect occurred. Thus, statistical dual-labeling of single-tagged and dual-tagged templates with analysis of smFRET EPR-S 2D histograms represents a stringent test of the combined efficiency and specificity of both UAA incorporation and dye labeling.
Figure 4. Spectrin R16 smFRET-µsALEX EPR-S histograms (released proteins). A, native structure of spectrin R16-R17a (PDB entry 1U4Q, 128 residues). Labeling site pairs 1–36 (62 Å), 1–39 (63 Å), 1–92 (33 Å), and 1–99 (41 Å) explored in this work are indicated. B, dual-tag/AUG released templates yield dual-labeled subpopulations consistent with highly efficient UAA incorporation and dye attachment. C, collapse of the dual-labeled subpopulations from D and E along the EPR axis (black) indicates (EPR) values qualitatively consistent with the inter-dye distances expected for natively folded spectrin (A). Red lines, probability distribution analysis fits of the data assuming a single conformational state, no bleaching, and no background. In vitro tag-and-modify sample generation for smFRET 15640 J. Biol. Chem. (2017) 292(38) 15636 –15648

Figure 3. Barnase smFRET-µsALEX EPR-S histograms. A, illustrated EPR-S histogram with donor-only (S = 1), dual-labeled (S = 0.5), and acceptor-only (S = 0) populations indicated. B, native barnase structure (PDB entry 1BRN, 110 residues) and labeling site pairs (1–44, 16 Å) and (1–66, 32 Å) chosen in this work. C, a single-tag/AUG barnase template yields only D-only and A-only subpopulations consistent with high-fidelity translation, specific UAA incorporation, and specific labeling. D and E, dual-tag/AUG released templates yield dual-labeled subpopulations consistent with highly efficient UAA incorporation and dye attachment. F, collapse of the dual-labeled subpopulations from D and E along the EPR axis (black) indicates (EPR) values qualitatively consistent with the inter-dye distances expected for natively folded barnase (B). Red lines, probability distribution analysis fits of the data assuming a single conformational state, no bleaching, and no background. G–K, EPR-S histograms of dual-tag/AUG barnase 1–44 RNC templates extruded to different extents from the exit tunnel of the ribosome. L, collapse of the dual-labeled subpopulations from G–K along the EPR axis (black) show an initially collapsed non-native state (EPR = 0.89), which shifts to lower EPR values (EPR = 0.82–0.75) as the last few residues of barnase are added and then extruded from the ribosome with a glycine–serine spacer. Constructs are illustrated above each panel. SPTC, the number of residues between the C terminus of the construct and the PTC of the ribosome. For released barnase samples (C–E), the observed dual-labeled population was assigned to the folded state (F in D and E).

In vitro tag-and-modify sample generation for smFRET 15640 J. Biol. Chem. (2017) 292(38) 15636 –15648
the dual-AUG mRNA templates to yield less or no dual-labeled products.

In nearly all cases, single-AUG mRNA templates yielded singly labeled (i.e. D-only or A-only) products (Figs. 3C, 4B, 5A, and 6B), whereas dual-AUG templates yielded significant dual-labeled/FRET-active subpopulations consistent with high efficiency incorporation of specific amino acids and specific labeling. B–G, dual-tag/AUG templates yield dual-labeled populations consistent with primarily collapsed and unfolded nascent chain conformations on the ribosome. D, released spectrin 1–39, in contrast, exhibits a low-FRET distribution consistent with the folded state. E–G, RNC constructs labeled at positions 1–92 and 1–99 fail to resolve the folded and unfolded states. H–K, sucrose-pelleted RNPs resulted in considerable sample aggregation, as evidenced by bridges between the D-only and A-only subpopulations. L, collapse of the dual-labeled subpopulations from B–G along the EPR axis (black histograms) indicate that the ribosome prevents the folding of nascent polypeptides. Red lines are fits of the experimental burst size distributions using PDA models including either one or two conformational states, no photobleaching, potential rapid interconversion between states (i.e. line broadening), and no background counts.

Figure 5. Spectrin R16 smFRET-µsALEX EPR histograms (RNC samples). A, a single-tag/AUG template yields only D-only and A-only subpopulations consistent with high-fidelity translation, specific UAA incorporation, and specific labeling. B and C, spectrin 1–36 and spectrin 1–39 RNC constructs exhibit high-EPR populations consistent with primarily collapsed and unfolded nascent chain conformations on the ribosome. D, released spectrin 1–39, in contrast, exhibits a low-EPR distribution consistent with the folded state. E–G, RNC constructs labeled at positions 1–92 and 1–99 fail to resolve the folded and unfolded states. H–K, sucrose-pelleted RNPs resulted in considerable sample aggregation, as evidenced by bridges between the D-only and A-only subpopulations. L, collapse of the dual-labeled subpopulations from B–G along the EPR axis (black histograms) indicate that the ribosome prevents the folding of nascent polypeptides. Red lines are fits of the experimental burst size distributions using PDA models including either one or two conformational states, no photobleaching, potential rapid interconversion between states (i.e. line broadening), and no background counts.

the dual-AUG mRNA templates to yield less or no dual-labeled products.

In nearly all cases, single-AUG mRNA templates yielded singly labeled (i.e. D-only or A-only) products (Figs. 3C, 4B, 5A, and 6B), whereas dual-AUG templates yielded significant dual-labeled/FRET-active subpopulations (Figs. 3–6 and supplemental Figs. S2–S58) that were easily detected within short acquisition times (i.e. ~10 min) using a simple non-parallelized and non-multiplexed confocal smFRET-ALEX screening platform (Fig. 1C and supplemental Fig. S1). These results indicate that most of the labeling sites/constructs were indeed surface-accessible as predicted. They also indicate that both HPG tagging and CuAAC dye labeling were highly efficient and specific for smFRET applications. As expected, the labeling reaction rates at different sites, for different constructs, or with different dyes were not always identical. For instance, the Atto647N-azide acceptor (+1 net charge) was more reactive toward negatively charged RNC constructs than either Alexa-488 azide (~2 net charge) or Alexa 647-azide (~3 net charge). At a 1:1 ratio of Alexa-488 to Atto647N during labeling, most detected events were A-only species, and in some cases, we were not able to obtain a large enough dual-labeled population for smFRET analysis with short (i.e. <10 min/sample) acquisition times. We compensated for this effect by labeling all RNC targets with a 2–3-fold molar excess of Alexa-488 over Atto647N to equalize the reaction rates of the donor and acceptor dyes with the target. This, however, still does not ensure either that labeling is complete at both sites or that each site has an equal reactivity with both dyes. Both of these requirements are necessary to obtain an idealized 1:2:1 ratio of D-only, dual-labeled, and A-only subpopulations (see additional comments in the supplemental materials). Finally, without direct verification of the HPG incorporation efficiencies at each site (e.g. using mass spectrometry and much larger-scale and hence prohibitively expensive prVIT reactions) it is impossible to deconvolute inefficient labeling from inefficient HPG incorporation.
This limitation, unfortunately, prevents the extraction of proper labeling efficiencies from smFRET EPR-S histogram data without making certain assumptions. For example, because the catalytic efficiency for HPG charging by methionine-tRNA synthetase is roughly 500-fold lower than for methionine (25), even low nanomolar amounts of methionine in our 12.5-μl prIVT reactions would noticeably affect the relative levels of dual-tagged and labeled DA or AD products relative to single-labeled D-only or A-only subpopulations.

Despite these various potential problems with quantifying labeling efficiencies using smFRET EPR-S histograms, we found that in the large majority of targets tagged at previously verified surface-accessible sites, we were able to obtain dual-labeled populations sufficient for rapid smFRET screening. Collapse of the EPR-S histograms along the stoichiometry axis yields lower bounds on the combined UAA incorporation/dye-labeling efficiencies that are obtained for each construct assuming equal reactivity of either dye at either site (see supplemental figures).

Notably, sample aggregation can also cause single-tagged templates to yield dual-labeled bursts. For example, sucrose-pelleted RNC samples exhibited anomalously large burst size distributions and bridged EPR-S histograms even at confocal detection volume occupancy probabilities of <0.1, suggesting that sucrose pelleting can induce significant sample aggregation (Fig. 5, H–K).

smFRET data on all of our samples were in qualitative agreement with expectations based on previous structural and energetic studies of these model systems. Dual-labeled samples
In vitro tag-and-modify sample generation for smFRET

released from the ribosome either naturally (Figs. 3 (C–E) and 4 (B–L)) or via RNase A/EDTA cleavage of RNCs (Figs. 4 (M and N) and 6S) yielded $E_{\text{PR}}$ values qualitatively consistent with the native structure of each model protein examined (Figs. 3B, 4A, and 6A). However, because at low denaturant concentrations unfolded proteins tend to shift to high-$E_{\text{FRET}}$ values (56) (and shorter inter-dye distances (57)) that overlap with the high-$E_{\text{FRET}}$ signals of most small folded globular proteins, we were unable to uniquely assign $E_{\text{PR}}$ subpopulations to either an unfolded (U) or folded (F) state for barnase or T4L NTD. In contrast, upon folding, the 1–39 and 1–36 labeling sites of spectrin R16 should separate by about 62 Å (Fig. 4A), enabling resolution of U ($E_{\text{PR}}$ $\sim$ 0.8–0.85) and F ($E_{\text{PR}}$ $\sim$ 0.4–0.5) even at low denaturant concentrations (Fig. 4, C and K). Such resolution of subpopulations using smFRET together with denaturant titration screens allows the structural and thermodynamic characterization of the underlying energy landscape for folding with great spatial and temporal precision (58).

As an example of the information that such samples can provide, we examined the relative populations of the low (folded) and high (unfolded) FRET states of several variants of the well-studied protein domain spectrin R16 (42) at low denaturant concentrations. A C-terminal α-helical extension (EK peptide) (59) stabilized spectrin R16 relative to its native C-terminal fusion context (Fig. 4, C versus E). Consistent with previous studies, an L97A mutation significantly destabilized spectrin R16 (Fig. 4, E versus F). A G105A mutation, which we thought might stabilize helix C, had no significant effect on spectrin R16 stability at low denaturant concentrations (Fig. 4, E versus G). Single proline or glycine insertions between the EK peptide extension and spectrin R16 also had little effect on the equilibrium between U and F at low denaturant (Fig. 4, G versus H, E versus I, and G versus J). Previous attempts to disrupt interdomain folding cooperativity in spectrin employed three consecutive proline residues (60). Unfortunately, the lack of EF-P in commercial prlVT systems results in drastically decreased yields of full-length protein, and we were unsuccessful in generating such constructs. A comparison of naturally released versus SecM-stalled and subsequently RNase A/EDTA-released spectrin R16 samples suggests that the 17-residue SecM-stalling peptide also does not appreciably perturb spectrin R16 structure or stability (Fig. 4, C versus M and D versus N). It is important to note that these constructs (as well as many others not shown; see supplemental Table S2 and Figs. S2–S58 were cloned, expressed, labeled, and screened in a matter of days using standard benchtop methods and commercially available prlVT systems, UAAAs, and click chemistry reagents. We also verified that much higher sample generation throughputs can be achieved using a 96-well sample generation platform for PCR template generation, prlVT, and all desalting/labeling steps. Using our standard benchtop methods for sample generation, smFRET quickly becomes the rate-limiting bottleneck. Thus, testing all of our constructs under a range of solution conditions, as required for a quantitative thermodynamic assessment of the underlying folding energy landscape, is beyond the scope of this current work (work in progress).

To demonstrate smFRET detection of conformational phenotypes from monovalent genotype-phenotype–linked RNC libraries, we generated and screened a series of SecM-stalled RNC variants of the three model proteins and explored the effects of the ribosome on nascent chain structure and folding (61–65). Figs. 3, 5, and 6 demonstrate the feasibility of such studies using a series of barnase, spectrin R16, and T4L NTD RNC constructs, respectively, which have been incrementally extruded from the ribosome exit tunnel by adding native C-terminal residues, EK peptides, and/or glycine–serine linkers upstream of the SecM-stalling sequence. Fig. 3 (G–I) demonstrates the changes in nascent chain $E_{\text{PR}}$–$S$ distributions for barnase as more native residues are added to the C terminus. Fig. 3 (J and K) shows the effects of extruding barnase out from the ribosome exit tunnel using glycine–serine linkers so that its C terminus is 27 or 47 residues from the peptidyltransferase center (PTC) of the ribosome. Fig. 3 (F and L) shows normalized $E_{\text{PR}}$ collapses of the dual-labeled subpopulations from Fig. 3. Single-state probability distribution analysis (PDA) fits to the data are also shown in red. Nascent barnase 1–44 Δ95 (Fig. 3G) appears highly collapsed on the ribosome. Upon extrusion, barnase shifts to lower $E_{\text{PR}}$ values; however, it is difficult to say whether barnase is fully folded or not when its C terminus is separated by 47 residues from the PTC, because we cannot resolve the folded and unfolded states. Fig. 6 (B–T) shows similar results for RNC variants of the NTD of T4L. These data demonstrate proof of principle for smFRET-based structural phenotyping of monovalent genotype-phenotype–linked protein libraries suitable for multiplexed single-molecule detection (66). The difference in the $E_{\text{PR}}$ distributions, both between the various partially extruded variants of a given protein and in the $E_{\text{PR}}$ distributions of full-length proteins off versus on the ribosome illustrates the power as well as the limitations of this approach for uncovering the effects of the ribosome on nascent chain structure and dynamics (67). It is important to note that for many of these constructs, the unambiguous assignment and resolution of U and F subpopulations and the observation of distinct folding transitions will require more advanced (i.e. site-specific and dual-internal) labeling schemes and the application of advanced analytical tools capable of achieving quantitative transfer of $E_{\text{PR}}$ distributions into inter-dye distance distributions (6, 57, 68).

In the case of spectrin R16, where U and F subpopulations are resolvable by smFRET off of the ribosome (see above), the effects of the ribosome on nascent chain structure and folding (64, 65) are easily observed. Whereas released spectrin R16 1–36 and 1–39 yielded low-$E_{\text{PR}}$ natively folded subpopulations in the absence of ribosomal tethering (Figs. 4 (C, K, and M) and 5D), RNC variants yielded high-$E_{\text{PR}}$ subpopulations consistent with collapsed U conformations (Fig. 5, B and C). PDA (55, 69) of the various dual-labeled released proteins and RNC samples listed above indicate minimal sample-induced $E_{\text{PR}}$ heterogeneity and histogram broadening, suggesting that statistically labeled samples can provide useful qualitative information about interresidue distance changes. Fluorescence correlation spectroscopy-based molecular brightness analysis (70) on selected singly labeled control samples indicated only minor changes in donor and acceptor quantum yield upon release of nascent chains from the ribosome (data not shown), thereby justifying comparisons of the $E_{\text{PR}}$ distributions of RNCs with...
In vitro tag-and-modify sample generation for smFRET

those of released samples. In addition, because RNCs and released proteins have significantly different diffusion coefficients, the integrity of RNC samples can be directly monitored using fluorescence correlation spectroscopy (67, 71).

RNC samples could be stored for about a week at 4 °C in high-magnesium buffer without appreciable peptidyl-tRNA bond degradation or RNC disassembly. Two-month-old RNC samples stored at 25 °C, in contrast, were mostly disassembled (Fig. 5, D and F). In-gel fluorescence monitoring of the peptidyl-tRNA bond integrity also suggested that RNCs were stable for about 30 min in RNC buffer at 25 °C in 2 M GdmCl (data not shown), providing an upper limit on the GdmCl concentrations and equilibration times that can be used for equilibrium denaturation of RNC samples (67). Finally, labeling of selected samples with different dye pairs (e.g. Alexa488/Alexa647 versus Atto488/Atto647N) yielded results qualitatively similar to those described above (data not shown), suggesting that dye photophysics, anisotropy, and perturbation are not significant sources of artifacts at our level of qualitative analysis. On the whole, these data highlight the advantages of our prIVT tag-and-modify approach for making large libraries of dual-labeled proteins and RNCs for smFRET-based monitoring of co-translational folding or nascent chain conformation on the ribosome. The generation of such RNC samples also opens the doors to multiplexed single-molecule screening of much larger protein libraries by harnessing the monovalent genotype-phenotype linkage that they provide.

Discussion

The detection of transient conformational states and dynamics from individual proteins using single-molecule fluorescence methods is currently limited in several ways. First, screening biomolecular ensembles in a statistically meaningful way one molecule at a time is an inherently slow process. Whereas this issue has been resolved for nucleic acid sequencing by parallelizing and multiplexing detection (5, 72, 73), it remains unresolved for protein conformational phenotyping. Second, generating large libraries of dual-labeled proteins for such screens remains challenging. Third, smFRET applications requiring high spatiotemporal resolution place strict demands on sample compositional homogeneity (9). The development of in vivo tag-and-modify genetic code expansion approaches enables dual site-specific labeling (10, 11) and can provide large amounts of highly homogeneous protein samples suitable for low-throughput/high-resolution screening. Importantly, the sensitivity and specificity requirements for dye coupling are relaxed as target expression yields and purity levels increase, respectively. Thus, high-yield in vivo expression and target affinity purification have made it possible to demonstrate the utility of a wide range of click chemistries for dual site-specific tag-and-modify approaches to smFRET sample generation (Fig. 1A). Unfortunately, when expression yields are low or sample generation and screening throughput rather than compositional homogeneity and spatiotemporal resolution are of primary interest, there are currently few options (16). The inefficient and nonspecific nature of suppressor-mediated in vivo UAA incorporation exacerbates the problem by further necessitating target-specific purification, limiting sample generation throughputs, and precluding generation of the monovalent genotype-phenotype–linked libraries required for multiplexed single-molecule screens of proteins (66) (Fig. 1C).

Here, instead of addressing the issues of inefficient and nonspecific UAA incorporation and labeling indirectly through purification and high-yield in vivo expression, we address these issues directly, thus eliminating the need for high-yield in vivo expression and target-specific purification (Fig. 1B). We used prIVT expression together with residue-specific sense codon reassignment to quantitatively incorporate an alkylene-bearing UAA (HPG) into proteins with absolute target specificity. We identified a ligand-accelerated CuAAC reaction that can overcome the low yields and high background biomass levels present in unpurified prIVT reactions and thus allow statistical dual labeling of proteins and RNC samples for qualitative smFRET screens. Although not generally site-specific and limited to N-terminally tagged and methionine-depleted constructs, this highly accessible prIVT-based tag-and-modify approach can easily be extended by using nascent chain-processing enzymes (74), kinetic labeling schemes (75), or more involved site-specific suppressor-mediated UAA incorporation methods (23, 76) to afford fully flexible dual site-specific dye attachment as required for quantitative smFRET studies. Furthermore, at the expense of a slightly larger dye linker, complete labeling can also be achieved faster, at <100 μM copper or with lower amounts of excess dye using commercially available picolyl-azide dyes (77).

We reiterate that there is an inherent trade-off between screening throughput and spatiotemporal resolution in smFRET applications. The quantitative conversion of $E_{\text{PR}}$ or $E_{\text{FRET}}$ values into inter-dye distances requires control experiments and analyses (e.g. $k^2$ simulations, quantum yield determinations, instrument detection efficiency controls), which also limit throughput when the highest spatial resolutions are required. The simple approach outlined here is therefore best suited for rapid initial qualitative screening of large libraries. This approach can achieve sample generation rates 2–3 orders of magnitude greater than existing methods. It also reduces dye consumption by about 500-fold, thereby dramatically reducing the cost per construct. Finally, each 12.5-μl prIVT reaction provides enough sample for 5–10 different labeling reactions (e.g. with different dye pairs) and thousands of individual smFRET screens under different solution conditions. Using a hierarchical approach, one can use such high-throughput qualitative screens to identify the most informative constructs (e.g. those with surface-accessible tag sites that resolve the subpopulations of primary interest) and then use dual-site-specific labeling methods and higher spatiotemporal screens on a smaller subset of the most informative constructs.

We have successfully applied our method to a host of different protein constructs, including destabilized, intrinsically disordered, and ribosome-bound proteins, many of which proved challenging to express, purify, label, and/or structurally characterize using traditional methods. Finally, the generation of dual-labeled monovalent genotype-phenotype–linked libraries...
(e.g. FRET-labeled RNC complexes; Fig. 1B) enables highly parallelized as well as multiplexed smFRET-based screens (Fig. 1C).

The increasing accessibility of single-molecule fluorescence microscopy systems, data collection and analysis software, and user facilities worldwide, both with and without single-molecule sequencing capabilities, has slowly lifted the major instrumentation barrier to entry for many single-molecule protein biophysics applications. The sample generation methods described here should help resolve some of the sample generation barriers that also prevent prospective users from using these powerful and as yet untapped tools and instruments.

Finally, in screening various click chemistries for our purpose, it became apparent that the trade-off between selectivity and sensitivity is not always fully characterized in the literature. We therefore suggest that the labeling of prIVT-generated RNCs may provide a useful tool for benchmarking newly developed click chemistries based on their sensitivity, selectivity, and bio-orthogonality in complex bioconjugation reactions.

**Experimental procedures**

**DNA and mRNA templates for prIVT reactions**

Full-length genes for cysteine-free T4L (43), the catalytically inactive H102A mutant of barnase (41), and the R16 domain of chicken α-spectrin (42) were subcloned into pET-LIC-(2A-T) (Addgene) containing a C-terminal SecM-stalling sequence (FSTPVWISQAGIRAGPQ) (49, 78). Neither barnase nor the NTD of T4L (residues 12–74) has endogenous methionines. One endogenous methionine in spectrin R16 was mutated to alanine (M15A) using standard protocols. Large modifications (C-terminal extensions or truncations) were introduced via traditional plasmid-based cloning and verified by sequencing. Smaller modifications (e.g. introduction of internal AUG codons and other such point mutations) were generally made via overlap extension PCR. Linear DNA templates for *in vitro* transcription reactions were generated via two-step nested PCR to add a T7 promoter, a stable mRNA hairpin structure (GGGAGACCACAAGGGGUUCCCA), an e enhancer element (UUAACCUUUA), and a strong ribosome-binding site (AGAGGAGA) to the 5′-UTR. PCR products were ethanol-precipitated, resuspended in RNase-free 10 mM Tris-Cl (pH 7.6), quantified, and diluted to 4 μM before storage at −20 °C. mRNA templates were generated from these linear DNA templates using standard T7 RNA polymerase *in vitro* transcription protocols. mRNAs were then ethanol-precipitated, quantified, and diluted to 20 μM in mRNA storage buffer (10 mM KOAc (pH 4.5) prior to flash-freezing and storage at −80 °C.

**prIVT reactions**

Commercial PURExpress® Δ(aa, tRNA) (New England Bio-labs) reactions (12.5 μl each) were set up according to the manufacturer’s instructions. The reactions included presynthesized mRNAs rather than DNA templates and an amino acid mix that, when diluted to its final working concentration, had 0.3 mM of each amino acid (except methionine) and 0.3 mM HPG. mRNAs were heat-denatured at 65 °C for 3 min and then quenched on ice before addition to the IVT master mix at a final concentration of 2.8 μM to initiate translation. Reactions were placed in a 37 °C incubator for 45 min or 2 h to generate single-turnover stalled RNCs or multiturnover released proteins, respectively. Because properly SecM-stalled RNCs are puromycin-insensitive (49), prematurely stalled RNC products due to polysome formation could be eliminated by a 5-min, 37 °C, 1 mM puromycin treatment following all single-turnover prIVTs, but this was generally unnecessary because we usually used a large excess of mRNA so that few polysomes were generated.

**Nascent chain product quantification**

prIVT reactions were carried out as described above except that radioactive 50 μM 14C-Phe (100 Ci/mol) was used instead of 0.3 mM non-radioactive Phe. Following each IVT reaction, 14C-Phe-tRNAs and peptidyl-tRNAs were degraded using RNase A/EDTA. Acid-insoluble (i.e. proteinaceous) radioactivity was then quantified by TCA precipitation followed by liquid scintillation counting on a Tri-Carb 2700TR analyzer (Packard/PerkinElmer Life Sciences).

**Fluorescence labeling**

prIVT reactions were carried out as described above and quenched with 1 volume (12.5 μl) of ice-cold 2X RNC stabilization buffer and desalted to remove excess HPG using P30 (RNC samples) or P6 (released protein samples) MicroBioSpin columns (Bio-Rad) equilibrated in 1X RNC buffer (20 mM Hicin (pH 7.0), 50 mM Mg(OAc)2, 75 mM NH4OAc, 120 mM KOAc, 0.05% Tween 20). Some RNC samples were also loaded onto 70-μl 1 M sucrose cushions in RNC buffer and spun for 75 min at 90,000 rpm in a TLA100 rotor. Desalted or pelleted samples along with stocks of either Alexa488-azide and Alexa647-azide (Thermo Fisher Scientific) or Atto488-azide and Atto647N-azide (ATTO-TEC GmbH) were brought into a vinyl anaerobic (<10 ppm O2) chamber (Coy Laboratory Products) and transferred into deoxygenated tubes. Donor and acceptor azido-dyes were added to 10 μl of each HPG-tagged target to a final total concentration of 5–10 μM (~10–20-fold excess). Samples were then deoxygenated for ~1 h. To initiate the CuAAC reaction, equal volumes of 10 mM CuSO4 and either 20 mM BTTES (31) or 20 mM BTTP (33) ligand were mixed together and then added to the deoxygenated samples to a final concentration of 0.5 mM copper (1 mM ligand). Finally, a preweighed and anaerobically stored dry aliquot of ascorbic acid was dissolved in deoxygenated double-distilled H2O to 10 mM and then added to a final concentration of 1 mM to initiate the reaction. After 1–2 h, 1 μl of the reaction containing a total of ~5–10 pmol of each dye was removed and saved as a control for in-gel fluorescence quantification of the labeling efficiency (see below). The rest of the reaction was brought up to ~30 μl with RNC buffer. Unreacted free dyes, copper, ascorbate, and ligand were then removed using a Micro Bio-Spin P6 or P30 size exclusion column (Bio-Rad) as per the manufacturer’s instructions prior to bringing the sample out from the anaerobic chamber. An A260 measurement of RNC samples was then used to quantify the efficiency of ribosome recovery during sample processing (typically >80% for all steps combined).
In vitro tag-and-modify sample generation for smFRET

In-gel fluorescence quantification of dye labeling efficiency

All of the dye conjugates used in the present study migrate near the dye front in the BisTris-MES (pH 6.5) SDS-PAGE system used. A Typhoon Trio gel scanner (GE Healthcare) together with ImageQuant TL (GE Healthcare) or ImageJ (National Institutes of Health) software was used to integrate product band intensities \( I_p \) as well as the free dye band intensity \( I_d \) in the control lane of each gel (e.g. Fig. 2D, lane 1). Radioactivity measurements (see above) were used to determine the yield of a given prfV reaction and thus the picomoles of tagged protein or NCs loaded in each lane. We then estimated the labeling efficiency as follows.

\[
\text{Efficiency} = \frac{\text{pmol of dye}}{\text{pmol of HPG}} = \frac{I_p \times (\text{pmol of free dye})/I_d}{(\text{pmol of NC}) \times (\text{no. of AUGs})}
\] (Eq. 1)

Diffusion-based smFRET-µsALEX

Supplemental Fig. S1 illustrates the basic elements of the two smFRET-µsALEX microscopes that were used for the present study. In setup A, the 488-nm line of an argon ion laser (Midwest Laser Products) and a 635-nm diode laser (Coherent) were combined using a dichroic mirror (D1: 600dcxr, Chroma). In setup B, a multiline (488 nm-568 nm-647 nm) argon-krypton mixed gas laser (Melles Griot) were combined using a dichroic mirror (D1: z488/633rpc in setup A, ZT488/640rpc-UF1 in setup B, Chroma), and underfilled \( (\lambda \sim 3) \) into the back aperture of an infinity-corrected UplanS apochromat 60 × 1.2 numerical aperture water immersion objective (Olympus America), thereby defining the two excitation volumes of the smFRET-µsALEX microscope system. In setup A, the objective was mounted onto a custom-made microscope body, whereas setup B used an Olympus IX-71 microscope body. The input laser powers during single-molecule data acquisition were \( \sim 50–100 \) microwatts for the 488-nm line and \( 15–30 \) microwatts for the 635/647-nm lines. Emitted bursts of fluorescence from freely diffusing labeled species were collected by the same objective, focused through the tube lens onto a 100-μm pinhole, collimated, spectrally separated into donor and acceptor emission paths by the emission dichroic (D3: 630dcxr in setup A, T635lpxr in setup B, Chroma), and refocused onto the active areas of two single-photon avalanche photodiode detectors (PerkinElmer Optoelectronics). The output from each detector was routed to a counter/timer board (PCI-6602, National Instruments) enabling 12.5 ns resolution time-stamping of each photon as described previously (54). Control signals were also sent to the acousto-optic tunable filter driver to alternate at a 25-μs periodicity between the donor and acceptor excitation beams. For smFRET-ALEX data acquisition, RNC samples were diluted in RNC buffer to \( \sim 100 \) pm ribosomes \( \sim 20 \) pm RNCs. Released proteins were diluted to \( \sim 20 \) pm in PM buffer. Single-molecule data sets were acquired for 10–20 min/sample. No oxygen scavengers or coupled reducing and oxidizing system reagents (79) were employed.

Data analysis consisted of first defining the donor and acceptor laser excitation windows of the alternation cycle. Next, a sliding window burst search algorithm was applied to the sum of all photons detected within either the donor or acceptor laser excitation windows of the alternation period. The criteria for defining a burst were as follows: 1) the interphoton delay time must be \( <1 \) ms, and 2) at least 20 consecutive photons must meet the first criterion. PDA analysis was carried out as described previously (55). Additional details on how the \( E_{\text{pte-S}} \) histograms were plotted and quantified are given in the supplemental materials.

Author contributions—K. M. H., S. M., and J. H. D. C. designed experiments. K. M. H. and J. H. carried out experiments and analyzed the data. P. W. provided the bis(tert-buty)tris-triazolyl ligands and suggestions for their use. K. M. H., J. H. D. C., M. K. J., and S. M. wrote the manuscript.

Acknowledgments—We thank Bruno Martinez, Avi Samelson, Sabriya Rosenmond, and Adrian Nickson for various DNA constructs; Ted Laurence and Shimon Weiss for single-molecule data acquisition and analysis software; Ian Liphardt for hardware used to build the single-molecule fluorescence microscope at UC Berkeley; Michael H. Schmidt for hardware used to build the single-molecule fluorescence microscope at CSU San Marcos; ANDbio LLC for assistance in finding cost-effective components used to build the single-molecule fluorescence microscope at CSU San Marcos; and Arthur E. Johnson, Jennifer Clark, Ben Schuler, Antonija Burcul, Bruno Martinez, and the entire Marquese and Cate laboratories for helpful comments and suggestions.

References

1. Eigen, M., and Rigler, R. (1994) Sorting single molecules: application to diagnostics and evolutionary biotechnology. Proc. Natl. Acad. Sci. U.S.A. 91, 5740–5747
2. Ha, T. (2014) Single-molecule methods leap ahead. Nat. Methods 11, 1015–1018
3. Jain, A., Liu, R., Ramani, B., Arauz, E., Ishitsuka, Y., Ragunathan, K., Park, J., Chen, J., Xiang, Y. K., and Ha, T. (2011) Probing cellular protein complexes using single-molecule pull-down. Nature 473, 484–488
4. Levene, M. J., Korphach, J., Turner, S. W., Foquet, M., Craighead, H. G., and Webb, W. W. (2003) Zero-mode waveguides for single-molecule analysis at high concentrations. Science 299, 682–686
5. Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibilo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., et al. (2009) Real-time DNA sequencing from single polymerase molecules. Science 323, 133–138
6. Kalinin, S., Peukert, T., Sindbert, S., Rothwell, P. J., Berger, S., Restle, T., Goody, R. S., Gohlke, H., and Seidel, C. A. (2012) A toolkit and benchmark study for FRET-restrained high-precision structural modeling. Nat. Methods 9, 1218–1225
7. Kim, S., Streets, A. M., Lin, R. R., Quake, S. R., Weiss, S., and Majumdar, D. S. (2011) High-throughput single-molecule opticalfluuidic analysis. Nat. Methods 8, 242–245
8. Tyagi, S., VanDelinder, V., Banterle, N., Fuertes, G., Milles, S., Agez, M., and Lemke, E. A. (2014) Continuous throughput and long-term observation of single-molecule FRET without immobilization. Nat. Methods 11, 297–300
9. Lerner, E., Hilzenrat, G., Amir, D., Tauber, E., Garini, Y., and Haas, E. (2013) Preparation of homogeneous samples of double-labelled protein...
suitable for single-molecule FRET measurements. Anal. Bioanal. Chem. 405, 5983–5991.

10. Spicer, C. D., and Davis, B. G. (2014) Selective chemical protein modification. Nat. Commun. 5, 4740.

11. Brustad, E. M., Lemke, E. A., Schultz, P. G., and Deniz, A. A. (2008) A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. J. Am. Chem. Soc. 130, 17664–17665.

12. Wang, K., Sachdeva, A., Cox, D. J., Wulf, N. M., Lang, K., Wallace, S., Meh, R. A., and Chin, J. W. (2014) Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labeling and FRET. Nat. Chem. 6, 393–403.

13. Lajoie, M. J., Rovner, A. J., Goodeman, D. B., Aerni, H. R., Haimovich, A. D., Kuznetsov, G., Mercer, J. A., Wang, H. H., Carr, P. A., Mosberg, J. A., Rohland, N., Schultz, P. G., Jacobson, J. M., Rinehart, J., Church, G. M., and Isaacs, F. J. (2013) Genomically recoded organisms expand biological functions. Science 342, 357–360.

14. Wang, K., Neumann, H., Peak-Chew, S. Y., and Chin, J. W. (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. Nat. Biotechnol. 25, 770–777.

15. Lemke, E. A. (2011) Site-specific labeling of proteins for single-molecule FRET measurements using genetically encoded ketone functionalities. Methods Mol. Biol. 751, 3–15.

16. Shi, X., Jung, Y., Lin, L. J., Liu, C., Wu, C., Cann, I. K. O., and Ha, T. (2012) Quantitative fluorescence labeling of aldehyde-tagged proteins for single-molecule imaging. Nat. Methods 9, 499–503.

17. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001) Cell-free translation reconstituted with purified components. Nat. Biotechnol. 19, 751–755.

18. Ohashi, H., Shimizu, Y., Ying, B. W., and Ueda, T. (2007) Efficient protein selection based on ribosome display system with purified components. Biochem. Biophys. Res. Commun. 352, 270–276.

19. Passiourea, T., Katoh, T., Goto, Y., and Suga, H. (2014) Selection-based discovery of druglike macrocyclic peptides. Annu. Rev. Biochem. 83, 727–752.

20. Josephson, K., Hartman, M. C. T., and Szostak, J. W. (2005) Ribosomal synthesis of unnatural peptides. J. Am. Chem. Soc. 127, 11727–11735.

21. Hong, S. H., Ntai, I., Haimovich, A. D., Kelleher, N. L., Isaacs, F. J., and Jeffew, M. C. (2014) Cell-free protein synthesis from a release factor 1 deficient Escherichia coli activates efficient and multiple site-specific nonstandard amino acid incorporation. ACS Synth. Biol. 3, 398–409.

22. Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006) A highly flexible tRNA acylation method for non-natural polypeptide synthesis. Nat. Methods 3, 357–359.

23. Terasaka, N., Iwane, Y., Geiermann, A. S., Goto, Y., and Suga, H. (2015) Recent developments of engineered translational machineries for the incorporation of non-canonical amino acids into polypeptides. Int. J. Mol. Sci. 16, 6513–6531.

24. Forster, A. C., Tan, Z., Nalam, M. N., Lin, H., Qu, H., Cornish, V. W., and Terasaka, N., Iwane, Y., Geiermann, A. S., Goto, Y., and Suga, H. (2015) Synthesis of unnatural peptides. Annu. Rev. Biochem. 84, 344–366.

25. Baase, W. A., Liu, L., Tronrud, D. E., and Matthews, B. W. (2010) Lessons from the lysozyme of phage T4. Proteins Sci. 19, 631–641.

26. Llinás, M., Gillespie, B., Dahlquist, F. W., and Marqusee, S. (1999) The energetics of T4 lysozyme reveal a hierarchy of conformations. Nat. Struct. Biol. 6, 1072–1078.

27. Llinás, M., and Marqusee, S. (1998) Subdomain interactions as a determinant in the folding and stability of T4 lysozyme. Protein Sci. 7, 96–104.

28. Back, J. W., David, O., Kramer, G., Masson, G., Kasper, P. T., de Koning, L. J., de Jong, L., van Maarseveen, J. H., and de Koster, C. G. (2005) Mild and chemoselective peptide-bond cleavage of peptides and proteins at azido homoalanine. Angew. Chem. Int. Ed. Engl. 44, 7946–7950.

29. Zhang, J., Leong, K. W., Johansson, M., and Ehrenberg, M. (2013) Accuracy of initial codon selection by aminoaoyl-tRNAs on the mRNA-programmed bacterial ribosome. Proc. Natl. Acad. Sci. U.S.A. 110, 9602–9607.

30. Johansson, M., Zhang, J., and Ehrenberg, M. (2012) Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. Proc. Natl. Acad. Sci. U.S.A. 109, 131–136.

31. Muto, H., Nakatogawa, H., and Ito, K. (2006) Genetically encoded but bioorthogonal propargylated L-phenylalanine in a cell-free environment for direct protein-protein click conjugation. Bioconjug. Chem. 17, 255–263.

32. Kapanidis, A. N., Lee, N. K., Laurence, T. A., Doose, S., Margeat, E., and Weiss, S. (2004) Fluorescence-aided molecule sorting: analysis of struct-

In vitro tag-and-modify sample generation for smFRET
In vitro tag-and-modify sample generation for smFRET

ture and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8936–8941

54. Hamadani, K. M., and Weiss, S. (2008) Nonequilibrium single molecule protein folding in a coaxial mixer. *Biophys. J.* **95**, 352–365

55. Nir, E., Michalet, X., Hamadani, K. M., Laurence, T. A., Neuhauer, D., Kovchevich, Y., and Weiss, S. (2006) Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. *J. Phys. Chem. B* **110**, 22103–22124

56. Yoo, T. Y., Meisburger, S. P., Hinshaw, J., Pollack, L., Haran, G., Gosnick, T. R., and Plaxco, K. (2012) Small-angle X-ray scattering and single-molecule FRET spectroscopy produce highly divergent views of the low-decarboxylase unfolded state. *J. Mol. Biol.* **418**, 226–236

57. Borgia, A., Zheng, W., Buhholzer, K., Borgia, M. B., Schuler, A., Hofmann, H., Soranno, A., Nettels, D., Gast, K., Grishaev, A., Best, R. B., and Schuler, B. (2016) Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. *J. Am. Chem. Soc.* **138**, 11714–11726

58. Nettels, D. and Schuler, B. (2011) Single-molecule FRET of protein-folding dynamics. In *Single-molecule Biophysics: Experiment and Theory*, Vol. 146 (Komatsuzaki, T., Kawakami, M., Takahashi, S., Yang, H., and Silbey, R. J., eds) John Wiley & Sons, Inc., Hoboken, NJ 10.1002/9781118131374.ch2

59. Marqusee, S., and Baldwin, R. L. (1987) Helix stabilization by Glu–… Lys+ salt bridges in short peptides of Denovo design. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8898–8902

60. Batey, S., Scott, K. A., and Clarke, J. (2006) Complex folding kinetics of a multidomain protein. *Biophys. J.* **90**, 2120–2130

61. Woolhead, C. A., McCormick, P. J., and Johnson, A. E. (2004) Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116**, 725–736

62. Lu, J., and Deutsch, C. (2005) Folding zones inside the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **12**, 1123–1129

63. Ellis, J. P., Bakke, C. K., Kirchdoerffer, R. N., Jungbauer, L. M., and Cavagnero, S. (2008) Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chem. Biol.* **3**, 555–566

64. Hoffmann, A., Becker, A. H., Zachmann-Brand, B., Deuerling, E., Bukau, B., and Kramer, G. (2012) Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol. Cell* **48**, 63–74

65. Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I., Jr., and Bustamante, C. (2011) The ribosome modulates nascent protein folding. *Science* **334**, 1723–1727

66. Gu, L., Li, C., Aach, J., Hill, D. E., Vidal, M., and Church, G. M. (2014) Multiplex single-molecule-interaction profiling of DNA-barcoded proteins. *Nature* **515**, 554–557

67. Samelson, A. J., Jensen, M. K., Soto, R. A., Cate, J. H., and Marqusee, S. (2016) Quantitative determination of ribosome nascent chain stability. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 13402–13407

68. Best, R. B., Hofmann, H., Nettels, D., and Schuler, B. (2015) Quantitative interpretation of FRET experiments via molecular simulation: force field and validation. *Biophys. J.* **108**, 2721–2731

69. Antonik, M., Felekyan, S., Gaiduk, A., and Seidel, C. A. M. (2006) Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B* **110**, 6970–6978

70. Kempe, D., Schöne, A., Fitter, J., and Gabb, M. (2015) Accurate fluorescence quantum yield determination by fluorescence correlation spectroscopy. *J. Phys. Chem. B* **119**, 4668–4672

71. Riedel, C., Gabizon, R., Wilson, C. A. M., Hamadani, K., Tsekouras, K., Marqusee, S., Pressé, S., and Bustamante, C. (2015) The heat released during catalytic turnover enhances the diffusion of an enzyme. *Nature* **517**, 227–230

72. Volfan, I. D., Tsai, Y. C., Clark, T. A., Wegener, J., Dai, Q., Yi, C., Pan, T., Turner, S. W., and Korlach, J. (2013) Analysis of RNA base modification and structural rearrangement by single-molecule real-time detection of reverse transcription. *Nano-biotechnology* **11**, 8

73. Ingargiola, A., Panzeri, F., Sarkosh, N., Guliniatti, A., Rech, I., Ghioni, M., Weiss, S., and Michalet, X. (2013) 8-spot smFRET analysis using two 8-pixel SPAD arrays. *Proc. SPIE* 10.1117/12.2003704

74. Wang, A., Winblade Nairn, N., Johnson, R. S., Tirrell, D. A., and Grabstein, K. (2008) Processing of N-terminal unnatural amino acids in recombinant human interferon-β in *Escherichia coli*. *Chembiochem* **9**, 324–330

75. Ratner, V., Kahana, E., Eichler, M., and Haas, E. (2002) A general strategy for site-specific double labeling of globular proteins for kinetic FRET studies. *Bioconjug. Chem.* **13**, 1163–1170

76. Xiao, H., and Schultz, P. G. (2016) At the interface of chemical and biological synthesis: an expanded genetic code. *Cold Spring Harb. Perspect. Biol.* 10.1101/cshperspect.a023945

77. Jiang, H., Zheng, T., Lopez-Aguilar, A., Feng, L., Kopp, F., Marlow, F. L., and Wu, P. (2014) Monitoring dynamic glycosylation in vivo using super-sensitive click chemistry. *Bioconjug. Chem.* **25**, 689–706

78. Bhushan, S., Hoffmann, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., Wilson, D. N., and Beckmann, R. (2011) SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol.* **9**, e1000581

79. Vogelsang, J., Kasper, R., Steinhauser, C., Person, B., Heilemann, M., Sauer, M., and Tinnefeld, P. (2008) A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chem. Int. Ed. Engl.* **47**, 5465–5469