



OPEN

## Ribosome selectivity and nascent chain context in modulating the incorporation of fluorescent non-canonical amino acid into proteins

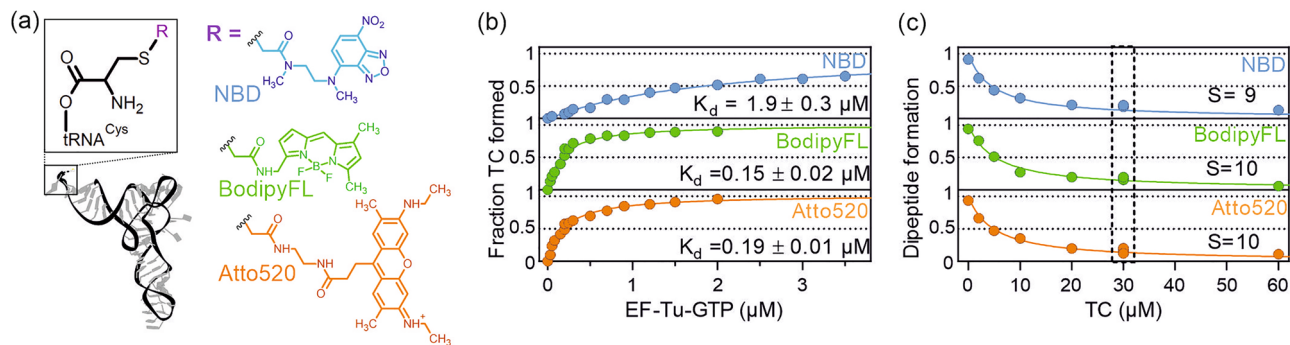
Michael Thommen<sup>1,2</sup>, Albena Draycheva<sup>1,2</sup> & Marina V. Rodnina<sup>1</sup>✉

Fluorescence reporter groups are important tools to study the structure and dynamics of proteins. Genetic code reprogramming allows for cotranslational incorporation of non-canonical amino acids at any desired position. However, cotranslational incorporation of bulky fluorescence reporter groups is technically challenging and usually inefficient. Here we analyze the bottlenecks for the cotranslational incorporation of NBD-, BodipyFL- and Atto520-labeled Cys-tRNA<sup>Cys</sup> into a model protein using a reconstituted in-vitro translation system. We show that the modified Cys-tRNA<sup>Cys</sup> can be rejected during decoding due to the reduced ribosome selectivity for the modified aa-tRNA and the competition with native near-cognate aminoacyl-tRNAs. Accommodation of the modified Cys-tRNA<sup>Cys</sup> in the A site of the ribosome is also impaired, but can be rescued by one or several Gly residues at the positions -1 to -4 upstream of the incorporation site. The incorporation yield depends on the steric properties of the downstream residue and decreases with the distance from the protein N-terminus to the incorporation site. In addition to the full-length translation product, we find protein fragments corresponding to the truncated N-terminal peptide and the C-terminal fragment starting with a fluorescence-labeled Cys arising from a StopGo-like event due to a defect in peptide bond formation. The results are important for understanding the reasons for inefficient cotranslational protein labeling with bulky reporter groups and for designing new approaches to improve the yield of fluorescence-labeled protein.

Fluorescence probes provide important insights into the structure and dynamics of macromolecules, but the repertoire of potential approaches for site-specific labeling is still limited<sup>1</sup>. In principle, incorporation of non-canonical amino acids (nc-aa) into recombinant proteins allows for residue-specific labeling with fluorescence reporters and methods for engineering such proteins have advanced rapidly in the past years<sup>2-4</sup>. The development of approaches for chemical aminoacylation of pdCpA dinucleotides and ligation to tRNA lacking two nucleotides at the 3'-end by T4 RNA ligase allowed for chemical charging of tRNA with nc-aa and showed that these chimeric nc-aa are incorporated into nascent peptides by the ribosome<sup>5,6</sup>. A general approach for incorporation of nc-aa in vivo became accessible by the expansion of the genetic code with orthogonal aminoacyl-tRNA synthetase-tRNA pairs that do not charge endogenous tRNA and, vice versa, are not substrates for the endogenous aminoacyl-tRNA synthetases, respectively<sup>7</sup>. Further progress came with the selection of mutant ribosomes with improved decoding efficiency of amber stop codon<sup>8</sup> or quadruplet codons<sup>9</sup> with the help of respective suppressor tRNAs. Ribozyme (Flexizyme)-mediated aminoacylation is an alternative method to prepare nc-aa-tRNA in vitro<sup>10,11</sup>. The combination of chemical and enzymatic aminoacylation methods made it possible to incorporate nc-aa site-specifically and with high yield<sup>12,13</sup>, providing the tools to synthesize designer proteins with desired properties. In the meantime, a large variety of nc-aa that were sterically restricted, modified in the backbone, or carried biophysical probes has been incorporated into proteins<sup>14-17</sup>. In some cases, in particular for *in-vitro* experiments, natural amino acids such as lysine, cysteine or N-terminal methionine can be used for specific cotranslational fluorescence labeling<sup>18-20</sup>, obviating the need to use genetic code expansion tools<sup>21-23</sup>. Many different fluorophores are well tolerated at the N-terminus of the nascent proteins. However, obtaining high

<sup>1</sup>Department of Physical Biochemistry, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany.

<sup>2</sup>These authors contributed equally: Michael Thommen and Albena Draycheva. ✉email: rodnina@mpinat.mpg.de



**Figure 1.** Ternary complex formation with RCys-tRNA and incorporation of RCys into dipeptides. (a) Chemical structures of fluorescent probes covalently attached to the amino acid side-chain thiol group to form RCys-tRNA<sup>Cys</sup>. (b) TC formation with RCys-tRNA<sup>Cys</sup> (10 nM) upon addition of increasing concentrations of EF-Tu-GTP monitored by the fluorescence change of the R probe (Figure S2). Data were normalized by setting the fluorescence intensity of RCys-tRNA<sup>Cys</sup> in the absence of EF-Tu-GTP to 0 and the saturating fluorescence intensity extrapolated by hyperbolic fitting to 1. Lines represent hyperbolic fits using a one site binding model (Material and Methods). The  $K_d$  value is reported with the standard error of the fit. (c) Formation of [<sup>3</sup>H] Met-RCys dipeptides on 70S IC (20 nM) in the presence of RCys-tRNA<sup>Cys</sup> (0.12 μM), EF-Tu-GTP (75 μM), and increasing concentration of *E. coli* aa-tRNA without Cys-tRNA<sup>Cys</sup>. [<sup>3</sup>H]Met-RCys peptides were separated by HPLC (Figure S3). Lines represent fits to a competition model yielding [TC(aa)]<sub>0.5</sub> as the concentration of TC required to reduce dipeptide formation by half of  $4.5 \pm 0.6 \mu\text{M}$ ,  $5.2 \pm 0.4 \mu\text{M}$ , and  $4.5 \pm 0.4 \mu\text{M}$  for NBD, BodipyFL, and Atto520, respectively. The selectivity *S* for the RCys-tRNA<sup>Cys</sup> relative to near-cognate aa-tRNAs indicated in the Figure was calculated taking into account the abundance of the near-cognate aa-tRNAs in the total tRNA pool (Methods). For the translation experiments in the next sections [TC(aa)] = 30 μM was kept constant (boxed).

cotranslational labeling efficiency at the internal positions remains a challenge and the potential differences in performance between nc-aa-tRNA and canonical aa-tRNA during the steps of decoding, peptide bond formation and translocation on the ribosome remain largely unexplored<sup>24</sup>.

Here we systematically explore the reasons for potentially low incorporation yield of fluorescence dyes into internal position of proteins using *in-vitro* translation approaches and Cys-tRNA as a scaffold for aa labeling. As fluorescence reporters, we used 7-nitrobenz-2-oxa-1,3-diazol-4-yl (NBD) and 4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene (BodipyFL), which were successfully used as fluorescence probes for cotranslational incorporation into proteins<sup>21–23,25–30</sup>. We also used Atto520 as it can be used for single-molecule studies<sup>31–35</sup>. The optimized *in-vitro* translation setup enabled us to identify limiting steps from decoding to peptide bond formation for fluorescent nc-aa-tRNA that were crucially modulated by the nascent chain context and length on the peptidyl-tRNA. We also identified a StopGo-like event resulting in formation of unlabeled truncated N-terminal and fluorescent-labeled C-terminal peptide products, which decreases the labeling yield of the full-length protein but may be developed into a useful tool to produce labeled peptide chains of defined stoichiometry.

## Results

**Delivery of aa-tRNA to the ribosome.** To identify the factors affecting the incorporation of fluorescent nc-aa into proteins, we used an *in-vitro* approach to introduce a fluorescence-labeled Cys into peptides in a fully reconstituted translation system<sup>30</sup>. We labeled the thiol group of enzymatically charged Cys-tRNA<sup>Cys</sup> with fluorescent probes (R) NBD, BodipyFL or Atto520, which are commonly used dyes for fluorescence labeling of proteins (Fig. 1a). RCys-tRNA<sup>Cys</sup> was purified by HPLC (Fig. S1) and added to the *in-vitro* translation system lacking native Cys-tRNA<sup>Cys</sup> (Materials and Methods). This approach allows for testing the cotranslational incorporation of a variety of probes into different nascent sequence contexts, in order to identify steps at which the incorporation fails and potentially circumvent such pitfalls.

One potential limitation to efficient incorporation of nc-aa into peptides is their compromised binding to EF-Tu<sup>22,36,37</sup>. To control for this effect, we monitored the formation of the ternary complex (TC) EF-Tu-GTP-RCys-tRNA<sup>Cys</sup> using the fluorescence intensity change of the R-group as a readout (Fig. 1b and Fig. S2). Cys-tRNA<sup>Cys</sup> labeled with BodipyFL or Atto520 exhibits slightly reduced binding affinity to EF-Tu-GTP (0.15 μM and 0.19 μM, respectively) compared to natural aa-tRNA (0.04–0.1 μM)<sup>36,38,39</sup>. The affinity for NBD-Cys-tRNA<sup>Cys</sup> is decreased by about tenfold (Fig. 1b). The reduced affinity of EF-Tu to NBD-Cys-tRNA<sup>Cys</sup> could potentially cause a bottleneck in NBD-Cys incorporation into peptides, but this effect can be compensated by the addition of sufficiently high concentration of EF-Tu-GTP into the *in-vitro* translation system.

In the next step, we monitored the incorporation of RCys into f[<sup>3</sup>H]Met-RCys dipeptide in the presence of increasing concentrations of other aa-tRNAs, thereby mimicking the *in-vitro* translation conditions. In the absence of extraneous aa-tRNAs, we observed efficient dipeptide formation for all three RCys-tRNA<sup>Cys</sup> variants (Fig. 1c). Addition of increasing concentrations of *E. coli* aa-tRNA (without Cys-tRNA<sup>Cys</sup>) decreases the yield of f[<sup>3</sup>H]Met-RCys dipeptides and concomitantly increases the yield of dipeptides formed between f[<sup>3</sup>H]Met and natural aa (Fig. S3). Because the aa-tRNA mixture does not contain cognate Cys-tRNA<sup>Cys</sup>, inhibition of RCys

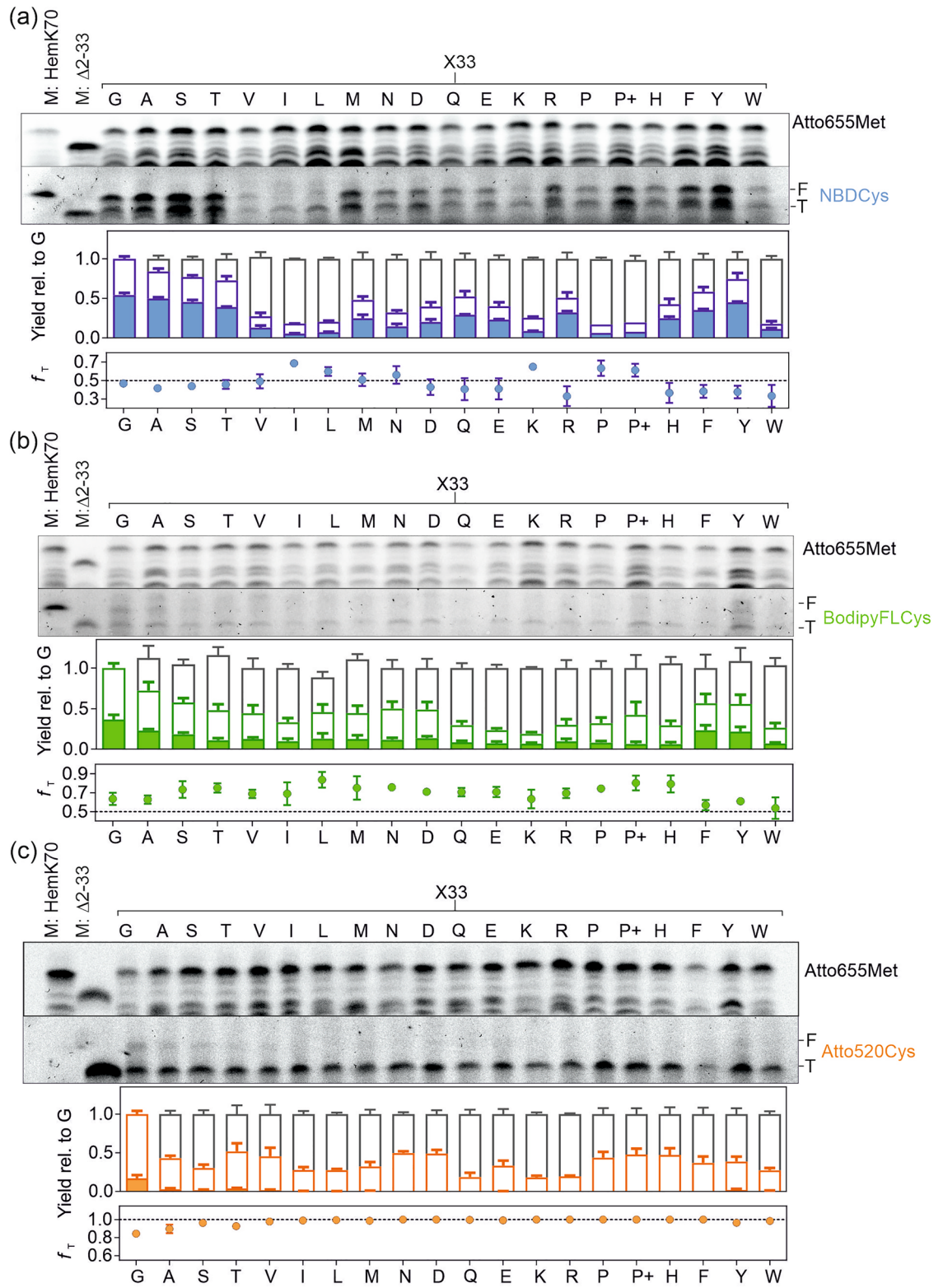
incorporation must be caused by the competition for binding to the Cys codon between the cognate RCys-tRNA<sup>Cys</sup> and near-cognate (or non-cognate) aa-tRNAs. Based on the rapid equilibrium competition model (Materials and Methods, Model 1), we calculated that the catalytic efficiency  $k_{\text{cat}}/K_M$  for the cognate RCys-tRNA<sup>Cys</sup> is 40 times higher than that for the bulk aa-tRNAs in the mixture<sup>40</sup>. If we consider near-cognate aa-tRNAs as main competitors and assume that the effect of non-cognate aa-tRNAs is negligible, the  $k_{\text{cat}}/K_M$  ratio would reduce to a value of only 9–10 because only a limited set of aa-tRNA are near-cognate to Cys-tRNA<sup>Cys</sup> (see Materials and Methods for further details)<sup>41</sup>. This value is by far smaller than the reported  $k_{\text{cat}}/K_M$  selectivity of 30–800 for the native cognate vs. near-cognate aa-tRNA at the initial selection stage<sup>42,43</sup>. Thus, the decoding selectivity for cognate nc-aa-tRNAs appears lower compared to natural aa-tRNA.

**Incorporation of RCys at position 34 of HemK.** In our previous work, we reported efficient incorporation of the large fluorescent probe Bodipy576/589 attached to the  $\epsilon$ -side-chain group of Lys-tRNA<sup>Lys</sup> into position 34 of the model protein HemK containing aa 1–70, HemK70<sup>22</sup>. To have a basis for comparison with the earlier results, we used the same model construct, but replaced Lys at position 34 with Cys to investigate the incorporation of RCys into the protein. The sequence context upstream of the incorporation site can modulate the incorporation efficiency of large nc-aa<sup>18,26</sup>. To systematically assess the influence of the sequence context on the incorporation of fluorescent probes of increasing size, we generated a library of constructs with all possible aa substitutions (except for Cys) at position 33 upstream of the incorporation site (Fig. 2). We used a fully reconstituted *E. coli in-vitro* translation system to synthesize HemK70 in the presence of RCys-tRNA<sup>Cys</sup> and a large excess of EF-Tu-GTP that is sufficient to overcome the reduced affinity of RCys-tRNA<sup>Cys</sup> to EF-Tu<sup>22,44</sup>. All experiments were carried out at single-turnover conditions, which ensured that neither the ribosome/mRNA recycling, nor recharging of aa-tRNA played a role in these experiments. In order to allow for normalization of differences in the overall translation efficiency and gel loading among the constructs, we also incorporated the fluorescent probe Atto655 at the N-terminus of HemK70 by performing translation initiation with Atto655-[<sup>3</sup>H]Met-tRNA<sup>Met</sup>, which does not affect translation<sup>45</sup>. HemK peptides were separated by SDS-PAGE and RCys incorporation was quantified by fluorescence imaging at the excitation/emission wavelength of the specific RCys (Fig. 2).

NBDCys is incorporated into full-length HemK70 with any aa at position 33 (X33; Fig. 2a), but the yield of the full-length (F) peptide carrying NBDCys depends on the identity of the X33 residue. Residues with small side-chains (Gly, Ala, Ser, Thr) and the aromatic residues Phe or Tyr allow for higher yield of NBDCys incorporation into the full-length peptide in comparison to Val, Ile, Leu, or Pro. A sizable fraction of NBDCys is found in the truncated translation product (T) that migrates at similar height as the HemK construct lacking residues 2 to 33 upstream of the incorporation site (M:  $\Delta 2$ -33; Fig. 2a–c), but higher than expected for the N-terminal HemK34 peptide (Fig. S4a). We note that sequence-dependent differences in electrophoretic mobility of such peptide fragments and the addition of fluorescence groups precludes unambiguous assignment of RCys-containing fragments. The fraction of T products is higher and the dependence on the upstream residue is stronger with BodipyFLCys than with NBDCys (Fig. 2b). Incorporation of Atto520Cys into the F product occurred only with Gly33 and to a minor extent with Ala33, whereas in all other cases only T was produced (Fig. 2c). Quantitative assessment using the fraction of truncated translation product relative to total incorporation ( $f_t = T/(T + F)$ ) reveals that RCys incorporation into F and T products is modulated differently, e.g., with NBDCys, Ile33 inhibited synthesis of F to a larger extent than T, whereas W33 is less inhibitory for F than for T (Fig. 2a). Addition of EF-P, a factor that accelerates peptide bond formation with Pro residues and enhances the incorporation of D- and  $\beta$ -amino acids<sup>44,46,47</sup>, slightly enhanced the yield of F and T (NBDCys and BodipyFLCys) and of T fragment (Atto520Cys). For all constructs, the incorporation of RCys into F and T products downstream of X33 was lower than into the original Gly33 construct (Fig. 2a–c).

**A steric model based on linear free energy relationships.** To gain further insights into the effect of aa at position 33, we used linear free energy relations (LFER)<sup>48</sup> to correlate the relative incorporation levels for G33X constructs (except for G33P) with aa descriptors for the ground-state properties of the upstream residues. The underlying assumption of this approach is that the yields of products result from the kinetic partitioning governed by the respective rate-limiting steps, which justifies the use of the LFER approach. Mei *et al.*<sup>49</sup> reported a set of 50 aa descriptors composed of 18 hydrophobic, 17 steric, and 15 descriptors for electronic properties (Table S1). The linear correlations with the aa descriptors were ranked according to the  $R^2$  value, which indicates how much of the variation in the experimental dataset is captured by the regression model (Fig. S5a–c). In all cases, the steric descriptors exhibited the strongest correlations, whereas the hydrophobic and electronic descriptors did not capture the experimental trends. The steric descriptors are highly correlated among each other (Fig. S4d), which precludes their ranking based on  $R^2$ . However, the steric descriptors can be divided into Taft-type descriptors that parametrize the steric hindrance at the reaction center and those that are related to the overall volume of the side chain<sup>50</sup>. We consistently found a strong correlation with the van-der-Waals volume of the side chain descriptor V35 or the length of the side chain descriptor V23, whereas the correlation with the Taft-type descriptors V21 and V24 for all products was weaker (Fig. S4a–c). For NBDCys incorporation into F product all correlations yielded a lower  $R^2$  value. This analysis suggests an important role of the overall steric volume of the upstream residue in modulating the incorporation levels of RCys. The incorporation levels calculated based on the coefficients obtained from LFER, e.g., the length of the amino acid side chain, replicate many features of the experimental data (Fig. S4e–g).

The slope of the linear correlation can be interpreted as sensitivity of RCys incorporation towards the given steric descriptor. The absolute values of  $b_T$  and  $b_F$  depend on the descriptor, e.g. the absolute values for the descriptor V35 representing van-der-Waals volume are larger than V23 for the length of the side chain. Therefore,



**◀Figure 2.** Incorporation of RCys at codon 34 of HemK70: Dependence on the identity of the upstream aa. (a) NBDcys incorporation. Upper gel, fluorescence scans of peptide products detecting Atto655-Met1 (top scan) and NBDcys34 (bottom scan) incorporation. Size marker M: HemK70 was prepared by mixing standards prepared individually by initiating HemK70 translation with N-terminal Atto655Met1 or BodipyFLMet1 (Materials and Methods). Size marker M:  $\Delta 2$ -33 was obtained by translating a construct lacking aa 2 to 33 of HemK70 with RCys incorporated immediately after Atto655-Met1 (see also Fig. 4). The incorporation of NBDcys into full length (F) and the truncated (T) HemK70 (bar graph) was quantified relative to the construct with Gly33 using the Atto655 fluorescence in the F product for normalization (Materials and Methods). Low-molecular weight bands in the Atto655 channel represent truncated peptides that appear at limiting concentration of total aa-tRNA. These bands were not considered further, as their mobility is significantly lower than that of the RCys-labeled T product. Stacked bars indicate contributions of T (color-coded open bars) and F (color-coded filled bars); open bars with gray border represent the fraction of peptides that did not incorporate RCys. The fraction of T product ( $f_T$ ; bottom panel) shows the relative band intensities of T and F products. The construct with Pro33 was translated in the absence (P-) or presence (P+) of EF-P (3  $\mu$ M). (b) and (c) same as in (a) but for BodipyFL and Atto520, respectively. Error bars indicate the SEM values for  $n=5$  for NBDcys and BodipyFLcys or  $n=4$  for Atto520cys.

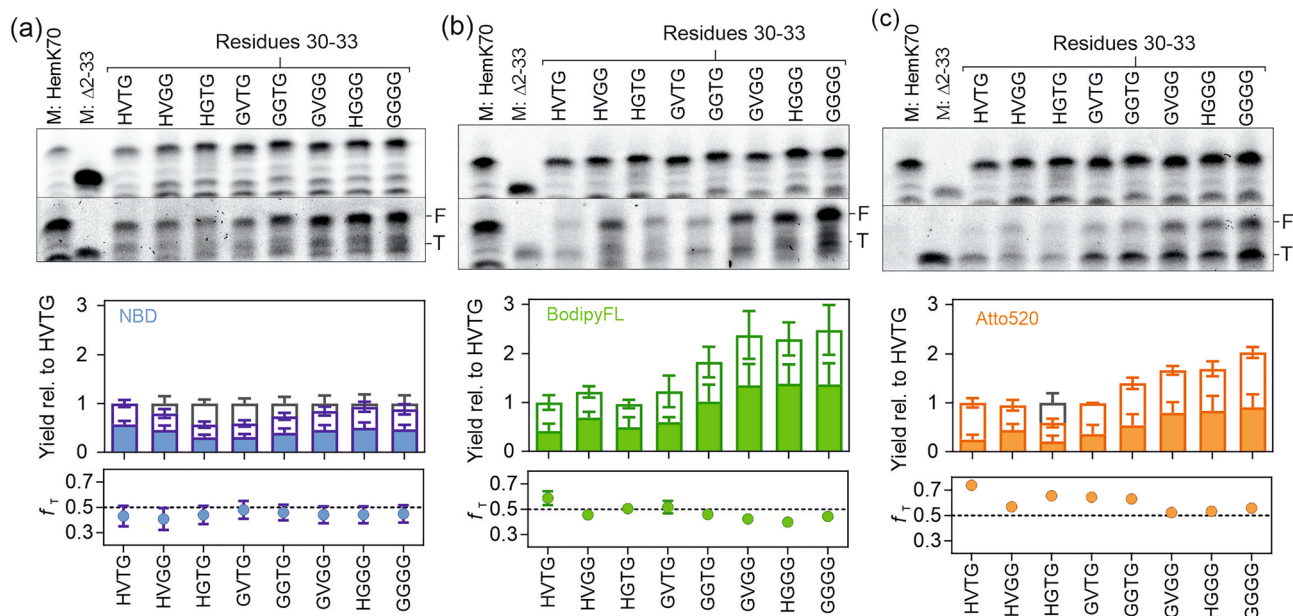
the slopes obtained for different descriptors cannot be compared to one another. However, each descriptor can be used to compare between T and F formation as well as between different R substituents. For the incorporation of RCys into T products, the slopes  $b_T$  for the same descriptor are similar for all RCys, indicating a similar dependence of the rate-limiting step on the properties of the X33 residue. For NBD and BodipyFL the  $b_F/b_T$  ratio is close to 1, suggesting a similar sensitivity also for F formation. However, for Atto520cys, the dependence on steric contribution is significantly higher for F than for T (Table S2), which suggests that the rate-limiting steps are different for the two products.

Testing the correlations of  $\log(F)$  and  $\log(T)$  with the *E. coli* codon usage frequency and tRNA abundance yielded significantly smaller  $R^2$  values compared to the steric descriptors (Fig. S5h,i). The yield of the F and T products detected by the RCys fluorescence normalized to the N-terminal Atto655 probe show only a weak dependence on the codon used to encode residue X33. Lastly, in the case of G, R, H, E, F, Y, and K at position 33, the isoacceptor aa-tRNA binds with a similar stability to the P site<sup>51</sup>. Thus, codon-dependent interactions in the P site have only minor, if any, contribution to the yield of F and T products.

**Multiple Gly residues upstream of the incorporation site.** The data in Fig. 2 show that a Gly upstream of the nc-aa incorporation site allowed for the highest RCys incorporation level as well as for the highest fraction of F product among the constructs. An enhanced incorporation of large fluorescent nc-aa was also reported downstream of Gly or Gly-Gly motive in other proteins<sup>19,26</sup>. To systematically investigate whether multiple upstream Gly residues might further enhance the incorporation of RCys, we generated an additional set of mutations in the native HVTG sequence containing Gly substitutions at different positions -1 to -4 relative to the incorporation site. For NBDcys, we did not observe a significant change in the formation of F or T products (Fig. 3a). In contrast, incorporation of BodipyFLcys was much more efficient in the presence of two additional upstream Gly residues in the GVGG, GGTG, or HGGG sequence context. Including a Gly residue at position -4 (GGGG) did not have an additional effect (Fig. 3b). Similarly, constructs with upstream Gly substitutions increased the incorporation of Atto520cys, albeit not to the same extent as with BodipyFLcys (Fig. 3c). For both BodipyFLcys and Atto520cys, the upstream Gly substitutions stimulated formation of both F and T products, resulting in small changes in  $f_T$ . A Gly at position -2 appears to be important, as direct comparison of Gly32 vs. Thr32 constructs consistently showed a higher fraction of the F product for the Gly32 containing sequence. For BodipyFLcys and Atto520cys, the fraction of the T product slightly decreased with two or more consecutive upstream Gly residues.

**Dependence of RCys incorporation on the upstream sequence length.** Lastly, we tested how the length of the nascent peptide upstream of the labeling site affects RCys incorporation. We generated a set of constructs with amino acid deletions upstream of the HGGG sequence preceding the RCys34 incorporation site (Fig. 4a). Notably, the critical -4 to -1 residues remain at the same location in the exit tunnel at the time when RCys is incorporated and the differences between the nascent chain complexes are only in the distal N-terminal part of the nascent chain.

A single major peptide product carrying the N-terminal Atto655-Met1 appears when the nascent chain upstream of Cys34 is up to 10 aa in length ( $\Delta 2$ -24) (Fig. 4b-d, top gel scans). The gel scans for RCys labels also show a corresponding peptide. Although the electrophoretic mobility of the product hardly changes with size for the increase in upstream sequence length from 1 to 10 aa (total peptide size of 35 to 45 aa), we note that a translation stop immediately after RCys incorporation would result in very short peptides (2 to 11 aa) that would be probably too small to detect on gels. Thus, RCys incorporation most likely does not inhibit translation of the mRNA sequence downstream of codon 34. As the sequence upstream of Cys34 expands from 12 to 28 aa, the mobility of the upper peptide band changes in both the Atto655-Met1 and RCys34 channels, as expected for a chain length increasing from 47 aa in the  $\Delta 2$ -22 construct to 65 aa for  $\Delta 2$ -6 in total (Fig. 4b-d). With these constructs, we observe pronounced ribosome pausing manifested by appearance of discrete bands marked by a bracket in Fig. 4b-d below the respective full-length Atto655-Met1 band. The mobility of the peptides arising from pausing also depends on the sequence length upstream of HGGC, suggesting that these peptides represent N-terminal peptide fragments. Although it is not clear why these fragments arise, they are notably shorter than the control  $\Delta 2$ -33 peptide (38 aa) and are most likely drop-off products of translation prior to codon 34.



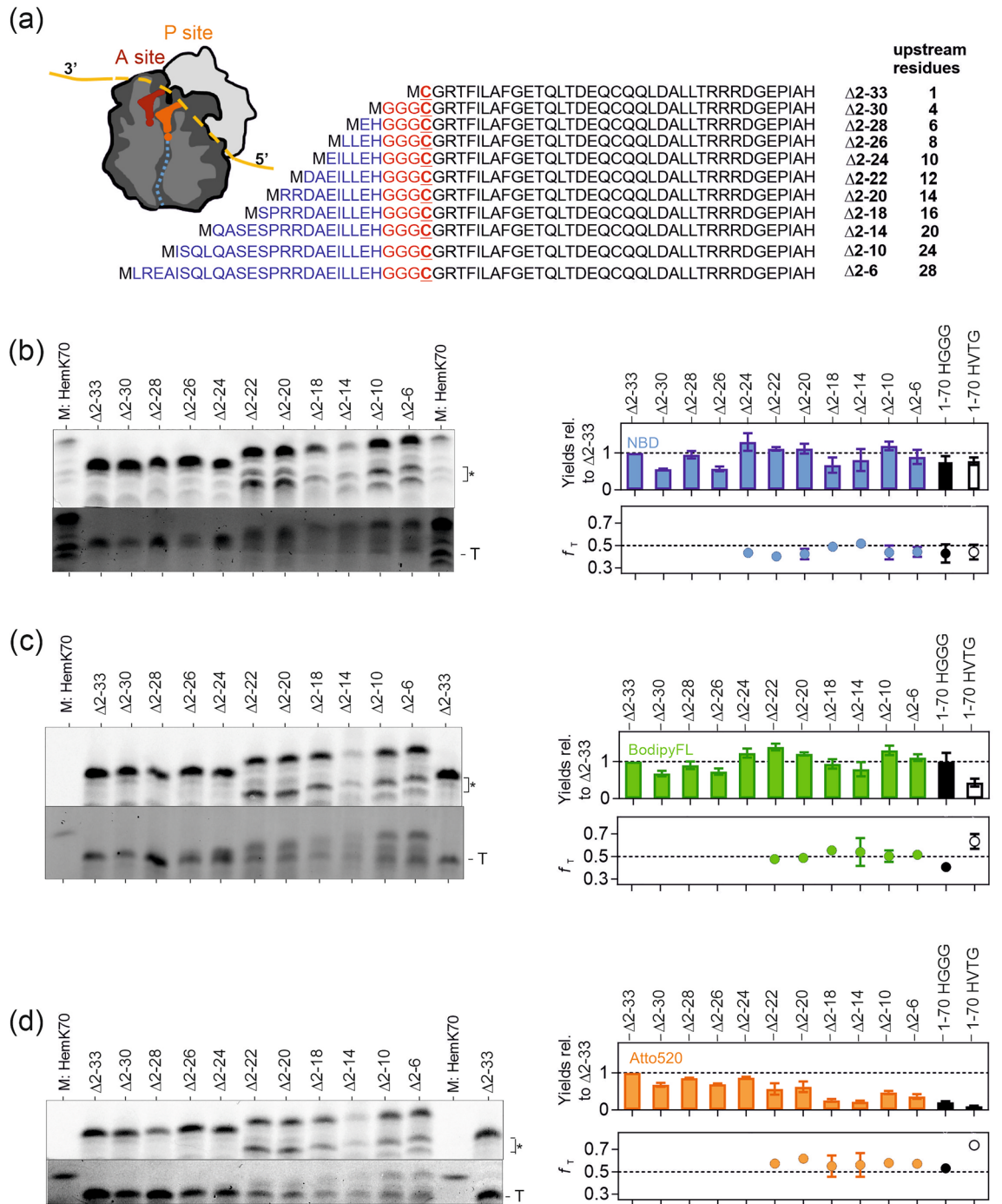
**Figure 3.** Effect of upstream residues  $-4$  to  $-2$  on RCys incorporation. Separation and quantification of translation products containing Atto655-Met1 and NBDCys34 was performed as described for Fig. 2. M: HemK70 marker was prepared from peptides containing N-terminal BodipyFLMet1 and Atto655Met1. The yield of F (filled bars) and T (open bars) products are shown relative to the native sequence context, HVTG. Error bars indicate the SEM values for  $n=5$  for NBDCys and  $n=6$  for BodipyFLCys and Atto520Cys. (a) NBD. (b) BodipyFL. (c) Atto520.

If translation stalled after the incorporation of RCys into the nascent peptide, the respective peptides (2 to 34 aa length) should be also found in this region, but such bands are not detectable in the RCys channel (lower panels). Instead, RCys is present in a band that migrates at a similar height as the  $\Delta 2-33$  peptide independent of the overall construct length as well as in high molecular weight bands that migrate dependent on the construct length. The RCys-containing band that is independent of construct length does not give rise to a band in the Atto655-Met1 channel and thus most likely represents the C-terminal peptide fragment that starts from RCys, whereas the top band entails the double-labeled peptides of respective length together with those labeled with Atto655-Met1 alone. The  $f_T$  value for the  $\Delta 2-22$  to  $\Delta 2-6$  constructs depends on the dye, e.g. it is 0.45–0.51 for NBD, 0.41–0.55 for BodipyFL and 0.53–0.61 for Atto520, but is essentially independent of the construct length (Fig. 4b–d, right panels).

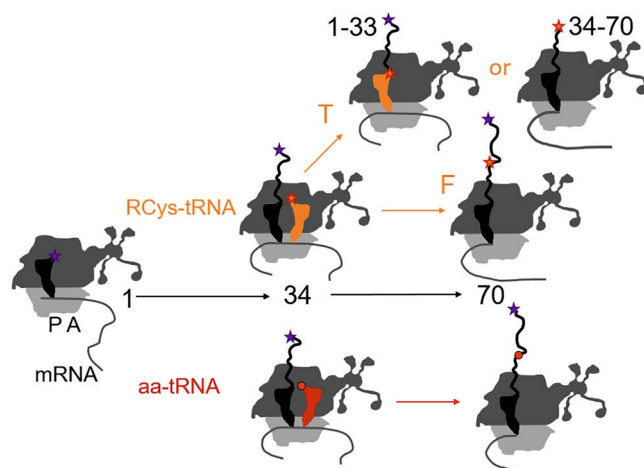
To compare the relative incorporation levels of RCys into constructs of different length, RCys fluorescence was normalized to that of the shortest construct  $\Delta 2-33$  where RCys is incorporated as second residue adjacent to Atto655-Met1, which allows for the comparison to the absolute incorporation levels of RCys established in Fig. 1c (see Methods). For NBDCys and BodipyFLCys, the total incorporation into the full-length and truncated translation products did not change considerably (i.e., by more than a factor of 2) upon truncation of the N-terminal sequence in the HGGG context (Fig. 4b,c, right panels; the only exception is  $\Delta 2-14$  where the intensity for Atto655-Met1 and all RCys is reduced for unknown reasons). In contrast, the incorporation of Atto520Cys in F and T gradually decreased with increasing length of the peptide sequence upstream of the HGGG sequence (Fig. 4d, right panel). With longer upstream polypeptide sequences, the HGGG motif supported very little Atto520Cys incorporation, whereas the  $f_T$  ratio remained constant, suggesting that the presence of a longer nascent chain did not affect the partitioning into F and T products but rather decreased the overall incorporation of Atto520Cys, potentially by favoring incorporation of the natural aa.

## Discussion

Our data on RCys incorporation are consistent with a simple model that predicts three likely outcomes of translation at codon Cys34 (Fig. 5). When codon 34 arrives at the A site of the ribosome, the P site is occupied with the peptidyl-tRNA, e.g. peptidyl-Gly-tRNA<sup>Gly</sup> in case of the native HemK mRNA sequence. The ribosome then accepts either the cognate RCys-tRNA<sup>Cys</sup> or a near-cognate native aa-tRNA owing to the low selectivity of the ribosome for RCys-tRNA<sup>Cys</sup>. Misreading of an internal codon by a near-cognate aa-tRNA in most cases does not halt translation; as a result, full-length peptides are produced. Our estimations of the ribosome selectivity for the RCys- vs. other aa-tRNAs suggest that depending on the relative concentrations of labeled and unlabeled aa-tRNAs, a large fraction of ribosomes may favor a near-cognate aa-tRNA, thereby reducing the yield of fluorescence-labeled product (Fig. 1c). Notably, the selectivity for the Bodipy576/589-labeled Lys-tRNA<sup>Lys</sup> appears to be somewhat higher, as we were able to achieve 70% incorporation efficiency of Bodipy576/589 at codon 34 of HemK<sup>52</sup>. Using genetically engineered tRNAs, e.g., with a tetranucleotide anticodon, would likely increase the selectivity for the respective cognate nc-aa-tRNA, as there are no natural cognate or near-cognate tRNAs



**Figure 4.** Upstream nascent peptide length effect. (a) Sequence of the HemK constructs. All construct except for  $\Delta 2-33$  contained a GGG sequence upstream of the Cys34 codon. The structural cartoon shows the nascent polypeptide attached to the peptidyl-tRNA in the P site prior to peptide bond formation with RCys-tRNA<sup>Cys</sup> in the A site. (b) Left panel: Visualization of nascent chains containing Atto655Met1 (top gel) or NBD-Cys (bottom gel) by SDS-PAGE as described in Fig. 2. An asterisk indicates short peptide products resulting from ribosome stalling prior to codon 34. Right panel: The sum of F and T products (NBD-Cys signal) for different upstream nascent chain lengths, shown relative to that for the shortest construct  $\Delta 2-33$  (top panel) and the  $f_T$  values (bottom panel). The values for HemK 1–70 HGGG (from Fig. 3) and HemK 1–70 HVTG (from Fig. 2) are shown as black or white symbols, respectively. (c) and (d) are same as (b) for BodipyFL-Cys and Atto-520-Cys, respectively, with the exception that marker M:HemK70 contained only peptides with the N-terminal BodipyFL.



**Figure 5.** RCys incorporation scenarios. Synthesis of HemK70 begins with 70S initiation complex, in the present experiments containing Atto655-Met1-tRNA<sup>Met</sup> in the P site. At position 34, RCys-tRNA<sup>Cys</sup> (orange) or a competing aa-tRNA (red) is incorporated and translation elongation can proceed to full-length HemK70 or result in formation of truncated products. If the ribosome stalls after peptide bond formation, the resulting T product contains RCys in the N-terminal fragment. Alternatively, stalling might occur before peptide bond formation between X33 and RCys34; the following translocation of RCys-tRNA<sup>Cys</sup> to the P site then results in labeling of the C-terminal fragment, akin to the products of the StopGo translation in eukaryotes.

for the respective four-letter codon. However, in this case an occasional misreading of a three-letters codon by a native aa-tRNA may result in ribosome frameshifting, thereby abolishing synthesis of the desired polypeptide<sup>53</sup>.

Successful accommodation of RCys-tRNA in the A site completes the decoding phase, but does not necessarily result in continuing translation (Fig. 5). Only a fraction of ribosomes proceeds to synthesize the full-length protein, as indicated by the  $f_T$  fraction of the truncated relative to the full-length peptides. In principle, the observed truncated products may arise if ribosomes stop elongation immediately after incorporation of RCys<sup>54–56</sup>; in this case, the N-terminal 1–34 aa HemK fragment would be RCys-labeled at its C-terminus. However, we find truncated peptides that do not carry the N-terminal label (Fig. 4), indicating that they lack the N-terminal part. One possible scenario is that a defect in peptide bond formation may arrest ribosomes in a state with a peptidyl-tRNA in the P site and RCys-tRNA<sup>Cys</sup> in the A site. At this point in translation, the peptide moiety must be released from the P-site tRNA to allow translocation of RCys-tRNA<sup>Cys</sup>, because a peptidyl-tRNA with a long nascent chain is unlikely to translocate due to interactions of the nascent peptide with the exit tunnel (see also below). The subsequent tRNA translocation would then release deacylated tRNA, whereas ribosomes would continue translation of the HemK mRNA producing the 34–70 aa peptide that is labeled at the N-terminus with RCys (Fig. 5). Similar observations have been reported for the incorporation of other large fluorescent probes, D-aa, and aromatic oligoamines<sup>18,57,58</sup>.

Large nc-aa were reported to be more efficiently incorporated downstream of aa residues with small side chains such as Gly residues<sup>19,26</sup>. In agreement with this notion, we find that the aa context upstream of Cys34 has a strong effect on the yield of RCys incorporation and Gly is the best penultimate residue; other aa can be tolerated depending on the nature of R. With NBD, a number of other aa are tolerated at the upstream position, whereas with BodipyFL and particularly Atto520, Gly is the only residue that allows for significant RCys incorporation. This effect is unlikely to be due to the competition between RCys-tRNA<sup>Cys</sup> and near-cognate aa-tRNAs at the initial selection phase of decoding, which is dominated by codon recognition on the small ribosomal subunit. In contrast, the accommodation of RCys-tRNA<sup>Cys</sup> in the A site of the large ribosomal subunit may depend on the nature of the side chain occupying the P site, e.g. a bulky side chain could provide a steric hindrance to positioning of RCys in the A site and favor rejection of RCys-tRNA prior to peptide bond formation. In fact, our LFER analysis (Fig. S5) indicates that the efficiency of RCys incorporation depends on steric properties of the X33 side chain such as the side chain length and van-der-Waals radius. Furthermore, after RCys-tRNA accommodation only a fraction of it appears in full-length peptides, indicating additional defects at the step of peptidyl transfer (Fig. 5). This notion is particularly true for Atto520Cys, which shows a higher sensitivity to the steric properties of X33 residues for synthesis of F protein compared to the T fragment (Table S2).

In addition to Gly33 at the -1 position, Gly residues at positions -2 to -4 further enhance the incorporation of BodipyFLCys and Atto520Cys (Fig. 3). Stretches of consecutive Gly residues may make the nascent peptide more flexible and leave more space inside the ribosomal exit tunnel to accommodate bulky nc-aa and facilitate better positioning of RCys for catalysis. With BodipyFLCys and Atto520Cys, the simultaneous increase in RCys incorporation into both F and T (Fig. 2) might indicate that the upstream Gly residues inside the ribosome exit tunnel modulate the RCys-tRNA accommodation in the A site (Fig. 5). In addition, the decrease in the  $f_T$  value indicates that synthesis of the full-length protein is enhanced, suggesting that multiple Gly residues have a favorable effect on peptide bond formation. Thus, protein engineering to accommodate Gly-rich stretches upstream of the incorporation site might be a generic solution to increase the incorporation of large nc-aa by



decreasing the steric hindrance in the peptidyl transferase center. Of note, by analogy to the enhancement of large nc-aa incorporation, a Gly residue in the P site also allows for flexible orientation of antibiotics chloramphenicol and oxazolidinone in the A site of the peptidyl transferase center in a manner that overcomes the inhibition by promoting drug displacement by the accommodating A-site tRNA<sup>59</sup>.

Our data also show that RCys incorporation strongly depends on the length of the nascent peptide upstream of the incorporation site. Large nc-aa can be more readily incorporated at N-terminal positions of the nascent chain than at the middle position in the protein<sup>24</sup>. Even for the largest probe Atto520, we observe an efficient labeling at the second position of the nascent chain, which decreases dramatically with the chain length (Fig. 4). Interactions of the nascent chain with the exit tunnel may induce a conformation of the peptidyl transferase center that disfavors nc-aa-tRNA accommodation<sup>60</sup> or reduce the mobility of the peptidyl-tRNA thereby preventing sampling of the conformation that allow peptide bond formation<sup>61,62</sup>. In contrast, for canonical aa the presence of a longer nascent chain attached to the peptidyl-tRNA can improve the positioning of the P-site substrate. This is manifested by the acceleration of peptide bond formation with nascent peptide length as well as restoration of translation elongation for P-site tRNA lacking the 2'OH group on the terminal adenine A76<sup>63,64</sup>.

The formation of T products that entail the C-terminal part of the protein and start from the fluorescence labeled RCys (Fig. 4) is reminiscent of the StopGo translation (also known as ribosome “skipping” or “stop-carry on”) in eukaryotic cells mediated by the viral 2A sequence<sup>65,66</sup>. Specific elements of the 2A nascent peptide promote ribosome stalling before incorporation of Pro-tRNA<sup>Pro</sup> into the nascent peptide. Presumably, the release of the nascent peptide from the P-site tRNA is promoted by translation termination factors 1 and 3. Then, Pro-tRNA<sup>Pro</sup> is translocated from the A site to the P site and the ribosome continues protein synthesis. StopGo translation has not been described in prokaryotes, but StopGo-like translation events were observed in bacterial cell-free translation systems upon incorporation of nc-aa with bulky side chain structures, such as large fluorescent probes or bulky aromatic oligo amides<sup>18,58,67</sup>, D-Trp and D-Arg<sup>57</sup> and L-β-homoglycine-Trp<sup>68</sup>. It is unclear what happens to the P-site peptidyl-tRNA in the bacterial system, as termination factors are not present in our reconstituted translation assay and thus the mechanism must differ from that of the authentic StopGo. In principle, peptidyl-tRNA could dissociate from the ribosome, but the rate of peptidyl-tRNA drop-off decreases dramatically when peptides grow more than 4 to 6 aa in length<sup>69,70</sup>. Importantly, a similar level of T product was formed upon shortening the upstream nascent chain length from 28 to 12 residues (Fig. 4), indicating that a shorter peptide—which is presumably more prone to drop-off—does not stimulate the StopGo-like event. Similarly, a high level of a StopGo-like translation product was reported upon incorporation of TAMRA fluorescent probe for a nascent chain with 22 aa upstream of the StopGo site<sup>18</sup>. These findings argue against a peptidyl-tRNA drop-off event during the formation of truncated translation products while incorporating large nc-aa. Alternatively, peptidyl-tRNA might undergo hydrolysis on the ribosome forming a pretranslocation complex with a deacylated tRNA in the P site and aa-tRNA in the A site. Such ribosomes can perform translocation, albeit slowly<sup>71,72</sup>. Spontaneous peptidyl-tRNA hydrolysis is typically much slower than the steps of the on-going translation elongation<sup>73</sup>. However, a long pause due to delayed accommodation of RCys-tRNA<sup>Cys</sup> may open the access of water to the catalytic site and promote hydrolysis. Thus, incorporation of large non-canonical fluorescence probes may be inefficient not only because they compete with highly abundant natural aa-tRNAs, but also due to inherent defects in accommodation and peptide bond formation that can reprogram the peptidyl transferase center from peptide bond formation to non-canonical termination activity. Screening the 23S rRNA libraries focusing on residues 2057–2063 and 2492–2507 at the peptidyl transferase center yielded ribosome variants that enhanced the incorporation of β-aa, dipeptides, dipeptidomimetic, and cyclic amino acids<sup>74,75</sup>. As some dipeptidomimetics approach the molecular weight of NBD-Cys and BodipyFL-Cys, it would be interesting to test whether such mutant ribosomes avoid premature termination. Our results also suggest that nc-aa with large or bulky side chains have a potential to act as 2A sequence analogs for coexpression of multiple proteins in bacteria from a single open reading frame, thereby pointing at an unexpected biotechnological potential of nc-aa.

## Materials and methods

**Buffers and reagents.** All chemicals were purchased from Merck and Sigma Aldrich. Radioactive amino acids were from Hartmann Analytic. Iodoacetamide-NBD amide (NBD) and BodipyFL-C<sub>1</sub>-iodoacetamide (BodipyFL) were purchased from Thermofisher Scientific. Atto520-iodoacetamide and Atto655-NHS were obtained from Atto-Tec. Buffer A: 50 mM Tris-HCl, pH 7.5, 70 mM NH<sub>4</sub>Cl, 30 mM KCl, 7 mM MgCl<sub>2</sub>. Buffer B: 50 mM Tris-HCl, pH 7.5, 70 mM NH<sub>4</sub>Cl, 30 mM KCl, 3.5 mM MgCl<sub>2</sub>, 0.5 mM spermidine, 8 mM putrescine, 2 mM DTT. Buffer C: 100 mM HEPES pH 7.5, 15 mM KCl, 7 mM MgCl<sub>2</sub>. Buffer D: 20 mM NH<sub>4</sub>OAc-HOAc pH 4.5, 10 mM MgCl<sub>2</sub>, 400 mM NaCl, 5% EtOH (v/v). Buffer E: 20 mM NH<sub>4</sub>OAc pH 4.5, 10 mM MgCl<sub>2</sub>, 400 mM NaCl, 40% EtOH (v/v).

**Ribosomes, translation factors and total aa-tRNA without Cys-tRNA<sup>Cys</sup>.** Ribosomes from *E. coli* MRE 600, initiation factors (IF1, IF2, IF3), and elongation factors (EF-G, EF-P, EF-Tu) were prepared as described<sup>44,76–78</sup>. Total tRNA from *E. coli* MRE 600 (Roche) was aminoacylated with 19 amino acids omitting L-Cys<sup>79</sup>.

**Preparation of RCys-tRNA<sup>Cys</sup>.** *E. coli* tRNA<sup>Cys</sup><sub>GCA</sub> was subcloned into pUC19, transcribed in vitro, and purified<sup>30</sup>. Aminoacylation was carried out in buffer C with DTT (5 mM), ATP (2 mM), L-Cys (150 μM), tRNA<sup>Cys</sup> (35 A<sub>260</sub> units mL<sup>-1</sup>), and S-100 enzymes (6%) for 45 min at 37 °C. Labeling of Cys-tRNA<sup>Cys</sup> (30 μM) was performed with different fluorescent probes (1 mM) in 50 mM HEPES pH 7.5, 50% DMF for 45 min at 25 °C<sup>30</sup>. Unreacted BodipyFL and Atto520 was removed by extraction with 50:50 phenol:chloroform followed by ethanol precipitation. Excess of the more hydrophilic NBD was removed by size-exclusion chromatography using a G-25

NAP-5 column (GE Healthcare). Labeled aa-tRNA was separated by HPLC using a LiChrospher WP300 RP-18 column and a step-gradient from buffer D to buffer E at a flow rate of 0.5 ml min<sup>-1</sup>: 15 min, 0% buffer E; 2 min, 0–30% buffer E; 20 min, 30% buffer E constant; 30 min, 30–100% buffer E. The concentration of RCys-tRNA<sup>Cys</sup> was determined by the absorbance of the R group.

**Preparation of Atto655-<sup>[3H]</sup>Met-tRNA<sup>fMet</sup>.** [<sup>3H</sup>]Met-tRNA<sup>fMet</sup> was prepared as described<sup>76</sup>. To obtain Atto655-<sup>[3H]</sup>Met-tRNA<sup>fMet</sup>, [<sup>3H</sup>]Met-tRNA<sup>fMet</sup> (50 μM) was reacted with Atto655-NHS (2 mM) in 100 mM NaOAc pH 5.0, 80% DMSO for 18 h at 4 °C. Unreacted fluorophore was removed by extraction with 50:50 phenol:chloroform followed by ethanol precipitation. Atto655-<sup>[3H]</sup>Met-tRNA<sup>fMet</sup> was purified by HPLC as described for RCys-tRNA<sup>Cys</sup>.

**mRNA Constructs.** HemK cloned into pET24a (Novagen) was used as template for site-directed mutagenesis<sup>52</sup> using oligonucleotide primers listed in Table S3. In order to minimize bias by codon usage and wobble decoding, the codon cognate to the most abundant isoaccepting tRNA pairing with Watson–Crick geometry was selected for substitution at position 33 of HemK<sup>41</sup>. For constructs with additional Gly substitutions codons GGU and GGC were used that are decoded by the major isoacceptor tRNA<sup>Gly3</sup>. mRNA used for RCys incorporation of RCys or translation of size-marker were prepared as described<sup>44</sup>.

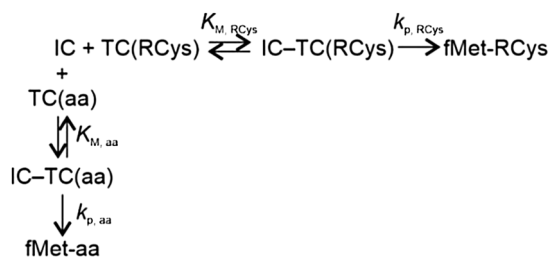
**Determination of RCys-tRNA<sup>Cys</sup> Affinity to EF-Tu–GTP.** In order to form the binary complex EF-Tu–GTP, EF-Tu–GDP (40 μM) was incubated with GTP (2 mM), phosphoenol pyruvate (PEP) (6 mM), and pyruvate kinase (0.1 mg ml<sup>-1</sup>) in buffer B at 37 °C for 15 min. MgCl<sub>2</sub> was added to compensate for GTP- and PEP-complexed Mg<sup>2+</sup><sup>80</sup>. EF-Tu–GTP was added to RCys-tRNA<sup>Cys</sup> (10 nM) and incubated for 2 min at 37 °C and the fluorescence change monitored in the Fluorolog spectrofluorimeter (Horiba Jobin Yvon). NBD emission was monitored at 525 nm after excitation at 480 nm. BodipyFL emission was monitored at 510 nm after excitation at 470 nm. Atto520 emission was monitored at 545 nm after excitation at 500 nm. The fluorescence change ΔF was fitted to a hyperbolic binding model:

$$\Delta F = F_0 + \Delta F_{max} \frac{[EF - Tu]}{K_d + [EF - Tu]}$$

where  $F_0$  as initial fluorescence,  $\Delta F_{max}$  the maximum fluorescence change and  $K_d$  is the equilibrium dissociation constant. The respective values are reported in Fig. 1b.

**70S initiation complex formation.** Initiation complexes were formed by incubating 70S ribosomes (1.5 μM) with initiation factors IF1, IF2, IF3 (1.5 μM each), mRNA (3 μM), GTP (1 mM) and either f [<sup>3H</sup>]Met-tRNA<sup>fMet</sup> (1 μM) or Atto655-<sup>[3H]</sup>Met-tRNA<sup>fMet</sup> in buffer A for 60 min. Initiation complex formation was monitored by nitrocellulose filtration and radioactivity counting.

**Dipeptide formation.** Ternary complexes were formed in buffer B by incubating EF-Tu–GTP (75 μM) with RCys-tRNA<sup>Cys</sup> (120 nM) (TC(RCys)), with or without total aa-tRNA lacking Cys-tRNA<sup>Cys</sup> (0–60 μM) (TC(aa)) for 2 min at 37 °C. Dipeptide formation was initiated by the addition of TC to IC (20 nM) containing 70S ribosomes, f [<sup>3H</sup>]Met-tRNA<sup>fMet</sup> and mRNA HemK70 Δ2-33 K34C. The reaction was carried out for 15 min at 37 °C. The antibiotic viomycin (200 μM) was added to the dipeptide reactions in order to stop translocation after a single round of peptide bond formation. The samples were flash-frozen, then thawed in the presence of 0.4 mg mL<sup>-1</sup> RNaseA (Thermo Scientific) and digested for 30 min at 37 °C. The pH value was adjusted by the addition of 1/10 volume of ice acetic acid. Dipeptides were separated by HPLC using a LiChrospher 100 RP-8 column and a gradient from 0 to 65% ACN in 0.1% TFA: 5 min, 0% ACN; 30 min, 0 to 65% ACN; 5 min, 65% ACN. The selectivity (S) of TC(RCys) for binding to the Cys codon relatively to the near-cognate aa-tRNAs were estimated based on the rapid equilibrium competition model:



**Model 1. Rapid equilibrium competition model for incorporation of RCys and natural aa into dipeptides.**  $K_M$  denotes the Michaelis–Menten constant and  $k_p$  the rate of peptide bond formation. The fraction  $f_{\text{RCys}}$  represents the fraction of ribosomes that incorporated RCys into the dipeptide.  $[\text{TC(aa)}]_{0.5}$  corresponds to the concentration of total TCs required to reduce  $f_{\text{RCys}}$  to 0.5 as determined in Fig. 1c.

For a mixture of TC(RCys) and TC(aa), the fraction of IC that formed the dipeptide  $f_{\text{RCys}}$  is determined by the Eq. 1:

$$f_{RCys} = \frac{k_{p,RCys} \frac{[TC(RCys)]}{K_{M,RCys}}}{k_{p,RCys} \frac{[TC(RCys)]}{K_{M,RCys}} + k_{p,aa} \frac{[TC(aa)]}{K_{M,aa}}}, \quad (1)$$

where  $k_{p,RCys}$  and  $k_{p,aa}$  are incorporation rates of RCys and aa, respectively, into the dipeptide and  $K_{M,RCys}$  and  $K_{M,aa}$  are Michaelis–Menten constants for the TC(RCys) and TC(aa), respectively. The terms in square brackets are the respective concentrations.  $[TC(aa)]_{\text{near-cognate}}$  was calculated from the total aa-tRNA TC(aa) concentration assuming that aa-tRNAs near-cognate to the Cys codon (tRNA<sup>Arg2</sup>, tRNA<sup>Gly3</sup>, tRNA<sup>Phe</sup>, tRNA<sup>Ser3</sup>, tRNA<sup>Sec</sup>, tRNA<sup>Ser5</sup>, tRNA<sup>Tyr1</sup>, tRNA<sup>Tyr2</sup>, and tRNA<sup>Tyr3</sup>) constitute about 24% of total *E. coli* tRNA<sup>41</sup>. From, Eq. 1, at the concentration of total TC required to reduce  $f_{RCys}$  by half ( $[TC(aa)]_{0.5}$ ),

$$k_{p,RCys} \frac{[TC(RCys)]}{K_{M,RCys}} = k_{p,aa} \frac{[TC(aa)]_{0.5}}{K_{M,aa}}, \quad (2)$$

and the selectivity S for the RCys incorporation, which is proportional to  $\frac{k_{p,RCys}}{K_{M,RCys}} \frac{K_{M,aa}}{k_{p,aa}}$  equals to

$$S = \frac{[TC(aa)]_{0.5}}{[TC(RCys)]}. \quad (3)$$

The respective S values calculated from  $[TC(aa)]_{\text{near-cognate}]_{0.5}$  are reported in Fig. 1c.

**In vitro translation and analysis.** Translation in the fully reconstituted single-turnover in vitro translation system was carried out as described<sup>22,44</sup>. 70S IC (20 nM) containing Atto655-<sup>[3H]</sup>Met-tRNA<sup>Met</sup> was mixed with EF-G (3 μM) and TCs formed from EF-Tu-GTP (75 μM), total aa-tRNA without Cys-tRNA<sup>Cys</sup> (30 μM), and RCys-tRNA<sup>Cys</sup> (120 nM) in buffer B and incubated for 15 min at 37 °C. To study the incorporation of Bodipy-FLCys or Atto520Cys, reactions were quenched by the addition of Laemmli sample buffer. Peptidyl-tRNA and aa-tRNA were subjected to aminolysis by 1.5 M NH<sub>2</sub>OH for 16 h at 37 °C<sup>81</sup>. To study NBDCys incorporation, the reactions were stopped by flash freezing, thawed in the presence of 0.4 mg mL<sup>-1</sup> RNaseA and digested for 30 min at 37 °C. Nascent polypeptides were resolved by Tris-tricine SDS-PAGE using 4% stacking, 10% spacer and 16.5% separation gels (49.5% T, 3% C) and scanned on a FLA-9000 fluorescence imager (Fujifilm Lifescience). Fluorescence of NBDCys and BodipyFLCys was excited at 473 nm and monitored after passing a 510 nm long pass filter. Fluorescence of Atto520Cys was excited at 532 nm and monitored after passing a 575 nm long pass filter. Fluorescence of Atto655 was excited at 685 nm and monitored after passing a 700 nm band pass filter. All scans were saved as 16 bit grayscale image. For the full-sized gel images in Fig. S6–S9, a constant linear contrast 270 to 6868 was applied to all 16 bit images. Size markers (highlighted as M) was generated by *in-vitro* translation of HemK70 mRNA (or Δ2-33 construct) using BodipyFL-<sup>[3H]</sup>Met-tRNA<sup>Met</sup> or Atto655-<sup>[3H]</sup>Met-tRNA<sup>Met</sup> as initiator tRNA<sup>44</sup>. The two fluorescence-labeled markers were produced individually, mixed, and used as SDS-PAGE markers that gave signals in both Atto655 and NDB/BodipyFL channels. Band intensities of all products was quantified using ImageJ (Version 1.51 23)<sup>82</sup>. Background subtraction was performed globally for all samples on the same gel. To account for translation and gel loading variations, the intensity of F and T bands was corrected by the band intensity of full-length HemK70 in the Atto655 channel and all resulting values further normalized to the RCys signal of the G33 construct. The relative fraction of the full-length product that does not contain RCys is calculated as 1 – (F + T) after the normalization. The fraction of truncated product  $f_T$  was calculated as T/(F + T). Band intensities of peptides with N-terminal truncations (Fig. 4) were normalized to the intensity of the Δ2-33 construct.

**Linear regression for correlation with amino acid descriptors.** Linear regression was performed using the equation  $y = a + b \times V$  with  $y$  as log<sub>10</sub> of the RCys incorporation into F or T,  $a$  as off-set,  $b$  as coefficient (slope), and  $V$  as aa descriptors from Table S1 that can also be retrieved in part from the AA index database<sup>83</sup>. Abundance of *E. coli* tRNA was from ref<sup>41</sup> and codon usage frequencies from the Kazusa codon usage database (<https://www.kazusa.or.jp/codon/>). Linear regressions and correlations between aa descriptors not contained in the AA index database were analyzed with the R software package (version 3.5.3).

Received: 11 March 2022; Accepted: 18 July 2022

Published online: 27 July 2022

## References

- Specht, E. A., Braselmann, E. & Palmer, A. E. A critical and comparative review of fluorescent tools for live-cell imaging. *Annu. Rev. Physiol.* **79**, 93–117. <https://doi.org/10.1146/annurev-physiol-022516-034055> (2017).
- Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60. <https://doi.org/10.1038/nature24031> (2017).
- d'Aquino, A. E., Kim, D. S. & Jewett, M. C. Engineered ribosomes for basic science and synthetic biology. *Annu. Rev. Chem. Biomol. Eng.* **9**, 311–340. <https://doi.org/10.1146/annurev-chembioeng-060817-084129> (2018).
- Katoh, T. & Suga, H. Engineering translation components improve incorporation of exotic amino acids. *Int. J. Mol. Sci.* <https://doi.org/10.3390/ijms20030522> (2019).
- Heckler, T. G. *et al.* T4 RNA ligase mediated preparation of novel “chemically misacylated” tRNAPheS. *Biochemistry* **23**, 1468–1473 (1984).

6. Baldini, G., Martoglio, B., Schachenmann, A., Zugliani, C. & Brunner, J. Mischarging Escherichia coli tRNA<sup>Phe</sup> with L-4'-[3-(trifluoromethyl)-3H-diazirin-3-yl]phenylalanine, a photoactivatable analogue of phenylalanine. *Biochemistry* **27**, 7951–7959 (1988).
7. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498–500. <https://doi.org/10.1126/science.1060077> (2001).
8. Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* **25**, 770–777. <https://doi.org/10.1038/nbt1314> (2007).
9. Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441–444. <https://doi.org/10.1038/nature08817> (2010).
10. Goto, Y., Katoh, T. & Suga, H. Flexizymes for genetic code reprogramming. *Nat. Protoc.* **6**, 779–790. <https://doi.org/10.1038/nprot.2011.331> (2011).
11. Ohuchi, M., Murakami, H. & Suga, H. The flexizyme system: A highly flexible tRNA aminoacylation tool for the translation apparatus. *Curr. Opin. Chem. Biol.* **11**, 537–542. <https://doi.org/10.1016/j.cbpa.2007.08.011> (2007).
12. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188 (1989).
13. Hohsaka, T., Ashizuka, Y., Murakami, H. & Sisido, M. Incorporation of nonnatural amino acids into streptavidin through in vitro frame-shift suppression. *J. Am. Chem. Soc.* **118**, 9778–9779. <https://doi.org/10.1021/ja9614225> (1996).
14. Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C. J. & Schultz, P. G. Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. *Methods Enzymol.* **202**, 301–336 (1991).
15. Cornish, V. W. *et al.* Site-specific incorporation of biophysical probes into proteins. *Proc. Natl. Acad. Sci. USA* **91**, 2910–2914. <https://doi.org/10.1073/pnas.91.8.2910> (1994).
16. Ellman, J. A., Mendel, D. & Schultz, P. G. Site-specific incorporation of novel backbone structures into proteins. *Science* **255**, 197–200. <https://doi.org/10.1126/science.1553546> (1992).
17. Mendel, D., Ellman, J. & Schultz, P. G. Protein-biosynthesis with conformationally restricted amino-acids. *J. Am. Chem. Soc.* **115**, 4359–4360. <https://doi.org/10.1021/ja00063a063> (1993).
18. Abe, R., Shiraga, K., Ebisu, S., Takagi, H. & Hohsaka, T. Incorporation of fluorescent non-natural amino acids into N-terminal tag of proteins in cell-free translation and its dependence on position and neighboring codons. *J. Biosci. Bioeng.* **110**, 32–38. <https://doi.org/10.1016/j.jbiosc.2010.01.003> (2010).
19. Ezure, T. *et al.* A cell-free translocation system using extracts of cultured insect cells to yield functional membrane proteins. *PLoS ONE* **9**, e112874. <https://doi.org/10.1371/journal.pone.0112874> (2014).
20. Hohsaka, T. Incorporation of unnatural non-alpha-amino acids into the N-terminus of proteins in a cell-free translation system. *Methods Mol. Biol.* **794**, 229–239. [https://doi.org/10.1007/978-1-61779-331-8\\_14](https://doi.org/10.1007/978-1-61779-331-8_14) (2012).
21. Gubbens, J., Kim, S. J., Yang, Z., Johnson, A. E. & Skach, W. R. In vitro incorporation of nonnatural amino acids into protein using tRNA(Cys)-derived opal, ochre, and amber suppressor tRNAs. *RNA* **16**, 1660–1672. <https://doi.org/10.1261/rna.2024810> (2010).
22. Mittelstaet, J., Konevega, A. L. & Rodnina, M. V. A kinetic safety gate controlling the delivery of unnatural amino acids to the ribosome. *J. Am. Chem. Soc.* **135**, 17031–17038. <https://doi.org/10.1021/ja407511q> (2013).
23. Woolhead, C. A., McCormick, P. J. & Johnson, A. E. 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. [https://doi.org/10.1016/S0092-8674\(04\)00169-2](https://doi.org/10.1016/S0092-8674(04)00169-2) (2004).
24. Hohsaka, T. Site-specific incorporation of fluorescent nonnatural amino acids into proteins and its application to fluorescence analysis of proteins. *Mol. Sci. Fluct. Toward Biol. Funct.* **1**, 105–123. [https://doi.org/10.1007/978-4-431-55840-8\\_5](https://doi.org/10.1007/978-4-431-55840-8_5) (2016).
25. Alder, N. N., Jensen, R. E. & Johnson, A. E. Fluorescence mapping of mitochondrial TIM23 complex reveals a water-facing, substrate-interacting helix surface. *Cell* **134**, 439–450. <https://doi.org/10.1016/j.cell.2008.06.007> (2008).
26. Iijima, I. & Hohsaka, T. Position-specific incorporation of fluorescent non-natural amino acids into maltose-binding protein for detection of ligand binding by FRET and fluorescence quenching. *ChemBioChem* **10**, 999–1006. <https://doi.org/10.1002/cbic.200800703> (2009).
27. Koubek, J., Chen, Y. R., Cheng, R. P. & Huang, J. J. T. Nonorthogonal tRNA(Amber)(cys) for protein and nascent chain labeling. *RNA* **21**, 1672–1682. <https://doi.org/10.1261/rna.051805.115> (2015).
28. Perronet, K. *et al.* Single molecule fluorescence detection of BODIPY-FL molecules for monitoring protein synthesis. *J. Lumin.* **127**, 264–268. <https://doi.org/10.1016/j.jlumin.2007.02.051> (2007).
29. Sadoine, M., Cerminara, M., Gerrits, M., Fitter, J. & Katranidis, A. Cotranslational incorporation into proteins of a fluorophore suitable for smFRET studies. *ACS Synth. Biol.* **7**, 405–411. <https://doi.org/10.1021/acssynbio.7b00433> (2018).
30. Buhr, F. *et al.* Synonymous codons direct cotranslational folding toward different protein conformations. *Mol. Cell* **61**, 341–351. <https://doi.org/10.1016/j.molcel.2016.01.008> (2016).
31. Dempsey, G. T., Vaughan, J. C., Chen, K. H., Bates, M. & Zhuang, X. W. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods* **8**, 1027. <https://doi.org/10.1038/Nmeth.1768> (2011).
32. Jeong, H. J. *et al.* Detection of vimentin serine phosphorylation by multicolor Quenchbodies. *Biosens. Bioelectron.* **40**, 17–23. <https://doi.org/10.1016/j.bios.2012.06.030> (2013).
33. Lymperopoulos, K., Kiel, A., Seefeld, A., Stohr, K. & Herten, D. P. Fluorescent probes and delivery methods for single-molecule experiments. *ChemPhysChem* **11**, 43–53. <https://doi.org/10.1002/cphc.200900359> (2010).
34. Ohashi, H. *et al.* Insight into the working mechanism of quenchbody: Transition of the dye around antibody variable region that fluoresces upon antigen binding. *Bioconj. Chem.* **27**, 2248–2253. <https://doi.org/10.1021/acs.bioconjchem.6b00217> (2016).
35. Winterfeld, S., Ernst, S., Borsch, M., Gerken, U. & Kuhn, A. Real time observation of single membrane protein insertion events by the Escherichia coli insertase YidC. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0059023> (2013).
36. Jeong, K. W., Pavlov, M. Y., Kwiatkowski, M., Forster, A. C. & Ehrenberg, M. Inefficient delivery but fast peptide bond formation of unnatural L-aminoacyl-tRNAs in translation. *J. Am. Chem. Soc.* **134**, 17955–17962. <https://doi.org/10.1021/ja3063524> (2012).
37. Nakata, H., Ohtsuki, T., Abe, R., Hohsaka, T. & Sisido, M. Binding efficiency of elongation factor Tu to tRNAs charged with non-natural fluorescent amino acids. *Anal. Biochem.* **348**, 321–323. <https://doi.org/10.1016/j.ab.2005.08.008> (2006).
38. Chapman, S. J., Schrader, J. M. & Uhlenbeck, O. C. Histidine 66 in Escherichia coli elongation factor tu selectively stabilizes aminoacyl-tRNAs. *J. Biol. Chem.* **287**, 1229–1234. <https://doi.org/10.1074/jbc.M111.294850> (2012).
39. Ott, G., Faulhammer, H. G. & Sprinzl, M. Interaction of elongation factor Tu from Escherichia coli with aminoacyl-tRNA carrying a fluorescent reporter group on the 3' terminus. *Eur. J. Biochem.* **184**, 345–352. <https://doi.org/10.1111/j.1432-1033.1989.tb15025.x> (1989).
40. Segel, I. H. *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems.* (Wiley, 1993).
41. Dong, H., Nilsson, L. & Kurland, C. G. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J. Mol. Biol.* **260**, 649–663. <https://doi.org/10.1006/jmbi.1996.0428> (1996).
42. Gromadski, K. B., Daviter, T. & Rodnina, M. V. A uniform response to mismatches in codon-anticodon complexes ensures ribosomal fidelity. *Mol. Cell* **21**, 369–377. <https://doi.org/10.1016/j.molcel.2005.12.018> (2006).
43. Gromadski, K. B. & Rodnina, M. V. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol. Cell* **13**, 191–200 (2004).

44. Doerfel, L. K. *et al.* EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science* **339**, 85–88. <https://doi.org/10.1126/science.1229017> (2013).
45. Liutkute, M., Maiti, M., Samatova, E., Enderlein, J. & Rodnina, M. V. Gradual compaction of the nascent peptide during cotranslational folding on the ribosome. *Elife* <https://doi.org/10.7554/eLife.60895> (2020).
46. Katoh, T. & Suga, H. Ribosomal incorporation of consecutive beta-amino acids. *J. Am. Chem. Soc.* **140**, 12159–12167. <https://doi.org/10.1021/jacs.8b07247> (2018).
47. Katoh, T., Wohlgenuth, I., Nagano, M., Rodnina, M. V. & Suga, H. Essential structural elements in tRNA(Pro) for EF-P-mediated alleviation of translation stalling. *Nat. Commun.* **7**, 11657. <https://doi.org/10.1038/ncomms11657> (2016).
48. Fersht, A. R. *Structure and Mechanism in Protein Science*. 9th edn, (W.H. Freeman and Company, 1998).
49. Mei, H., Liao, Z. H., Zhou, Y. & Li, S. Z. A new set of amino acid descriptors and its application in peptide QSARs. *Biopolymers* **80**, 775–786. <https://doi.org/10.1002/bip.20296> (2005).
50. Fauchere, J. L., Charton, M., Kier, L. B., Verloop, A. & Pliska, V. Amino acid side chain parameters for correlation studies in biology and pharmacology. *Int. J. Pept. Prot. Res.* **32**, 269–278. <https://doi.org/10.1111/j.1399-3011.1988.tb01261.x> (1988).
51. Fahlman, R. P., Dale, T. & Uhlenbeck, O. C. Uniform binding of aminoacylated transfer RNAs to the ribosomal A and P sites. *Mol. Cell.* **16**, 799–805. <https://doi.org/10.1016/j.molcel.2004.10.030> (2004).
52. Holtkamp, W. *et al.* Cotranslational protein folding on the ribosome monitored in real time. *Science* **350**, 1104–1107. <https://doi.org/10.1126/science.aad0344> (2015).
53. Wang, K. H., Schmied, W. H. & Chin, J. W. Reprogramming the Genetic Code: From Triplet to Quadruplet Codes. *Angew. Chem. Int. Edit.* **51**, 2288–2297. <https://doi.org/10.1002/anie.201105016> (2012).
54. Englander, M. T. *et al.* The ribosome can discriminate the chirality of amino acids within its peptidyl-transferase center. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 6038–6043. <https://doi.org/10.1073/pnas.1424712112> (2015).
55. Fleisher, R. C., Cornish, V. W. & Gonzalez, R. L. Jr. d-Amino Acid-Mediated Translation Arrest Is Modulated by the Identity of the Incoming Aminoacyl-tRNA. *Biochemistry* **57**, 4241–4246. <https://doi.org/10.1021/acs.biochem.8b00595> (2018).
56. Wang, J. F., Kwiatkowski, M. & Forster, A. C. Kinetics of tRNA(Pyl)-mediated amber suppression in *Escherichia coli* translation reveals unexpected limiting steps and competing reactions. *Biotechnol. Bioeng.* **113**, 1552–1559. <https://doi.org/10.1002/bit.25917> (2016).
57. Fujino, T., Goto, Y., Suga, H. & Murakami, H. Reevaluation of the D-amino acid compatibility with the elongation event in translation. *J. Am. Chem. Soc.* **135**, 1830–1837. <https://doi.org/10.1021/ja309570x> (2013).
58. Tsiamantas, C. *et al.* Ribosomal incorporation of aromatic oligoamides as peptide sidechain appendages. *Angew. Chem. Int. Ed. Engl.* **59**, 4860–4864. <https://doi.org/10.1002/anie.201914654> (2020).
59. Marks, J. *et al.* Context-specific inhibition of translation by ribosomal antibiotics targeting the peptidyl transferase center. *Proc. Natl. Acad. Sci. USA* **113**, 12150–12155. <https://doi.org/10.1073/pnas.1613055113> (2016).
60. Agirrezabala, X. *et al.* A switch from alpha-helical to beta-strand conformation during co-translational protein folding. *Embo J.* **1**, 1 (2022).
61. Lehmann, J. Induced fit of the peptidyl-transferase center of the ribosome and conformational freedom of the esterified amino acids. *RNA* **23**, 229–239. <https://doi.org/10.1261/rna.057273.116> (2017).
62. Melnikov, S. V. *et al.* Mechanistic insights into the slow peptide bond formation with D-amino acids in the ribosomal active site. *Nucl. Acids Res.* **47**, 2089–2100. <https://doi.org/10.1093/nar/gky1211> (2019).
63. Huang, Y. & Sprinzl, M. Peptide bond formation on the ribosome: the role of the 2'-OH group on the terminal adenosine of peptidyl-tRNA and of the length of nascent peptide chain. *Angew. Chem. Int. Ed. Engl.* **50**, 7287–7289. <https://doi.org/10.1002/anie.201005245> (2011).
64. Katunin, V. I., Muth, G. W., Strobel, S. A., Wintermeyer, W. & Rodnina, M. V. Important contribution to catalysis of peptide bond formation by a single ionizing group within the ribosome. *Mol. Cell* **10**, 339–346. [https://doi.org/10.1016/s1097-2765\(02\)00566-x](https://doi.org/10.1016/s1097-2765(02)00566-x) (2002).
65. Sharma, P. *et al.* 2A peptides provide distinct solutions to driving stop-carry on translational recoding. *Nucl. Acids Res.* **40**, 3143–3151. <https://doi.org/10.1093/nar/gkr1176> (2012).
66. J.D., B. & M.D., R. in *Recoding: Expansion of Decoding Rules Enriches Gene Expression. Nucleic Acids and Molecular Biology* Vol. 24 (eds Atkins J. & Gesteland R.) (Springer, 2010).
67. Kawakami, T., Ishizawa, T. & Murakami, H. Extensive reprogramming of the genetic code for genetically encoded synthesis of highly N-alkylated polycyclic peptidomimetics. *J. Am. Chem. Soc.* **135**, 12297–12304. <https://doi.org/10.1021/ja405044k> (2013).
68. Fujino, T., Goto, Y., Suga, H. & Murakami, H. Ribosomal synthesis of peptides with multiple beta-amino acids. *J. Am. Chem. Soc.* **138**, 1962–1969. <https://doi.org/10.1021/jacs.5b12482> (2016).
69. Cruz-Vera, L. R., Hernandez-Ramon, E., Perez-Zamorano, B. & Guarneros, G. The rate of peptidyl-tRNA dissociation from the ribosome during minigene expression depends on the nature of the last decoding interaction. *J. Biol. Chem.* **278**, 26065–26070. <https://doi.org/10.1074/jbc.M301129200> (2003).
70. Karimi, R., Pavlov, M. Y., Heurgue-Hamard, V., Buckingham, R. H. & Ehrenberg, M. Initiation factors IF1 and IF2 synergistically remove peptidyl-tRNAs with short polypeptides from the P-site of translating *Escherichia coli* ribosomes. *J. Mol. Biol.* **281**, 241–252. <https://doi.org/10.1006/jmbi.1998.1953> (1998).
71. Fredrick, K. & Noller, H. F. Accurate translocation of mRNA by the ribosome requires a peptidyl group or its analog on the tRNA moving into the 30S P site. *Mol. Cell* **9**, 1125–1131. [https://doi.org/10.1016/s1097-2765\(02\)00523-3](https://doi.org/10.1016/s1097-2765(02)00523-3) (2002).
72. Semenkov, Y. P., Rodnina, M. V. & Wintermeyer, W. Energetic contribution of tRNA hybrid state formation to translocation catalysis on the ribosome. *Nat. Struct. Biol.* **7**, 1027–1031. <https://doi.org/10.1038/80938> (2000).
73. Pierson, W. E. *et al.* Uniformity of peptide release is maintained by methylation of release factors. *Cell Rep* **17**, 11–18. <https://doi.org/10.1016/j.celrep.2016.08.085> (2016).
74. Dedkova, L. M. & Hecht, S. M. Expanding the scope of protein synthesis using modified ribosomes. *J. Am. Chem. Soc.* **141**, 6430–6447. <https://doi.org/10.1021/jacs.9b02109> (2019).
75. Lee, J., Schwarz, K. J., Kim, D. S., Moore, J. S. & Jewett, M. C. Ribosome-mediated polymerization of long chain carbon and cyclic amino acids into peptides in vitro. *Nat. Commun.* **11**, 4304. <https://doi.org/10.1038/s41467-020-18001-x> (2020).
76. Milon, P. *et al.* Transient kinetics, fluorescence, and FRET in studies of initiation of translation in bacteria. *Methods Enzymol.* **430**, 1–30. [https://doi.org/10.1016/S0076-6879\(07\)30001-3](https://doi.org/10.1016/S0076-6879(07)30001-3) (2007).
77. Rodnina, M. V. *et al.* Thiostrepton inhibits the turnover but not the GTPase of elongation factor G on the ribosome. *Proc. Natl. Acad. Sci. USA* **96**, 9586–9590 (1999).
78. Rodnina, M. V. & Wintermeyer, W. GTP consumption of elongation factor Tu during translation of heteropolymeric mRNAs. *Proc. Natl. Acad. Sci. USA* **92**, 1945–1949 (1995).
79. Gu, S. Q., Peske, F., Wieden, H. J., Rodnina, M. V. & Wintermeyer, W. The signal recognition particle binds to protein L23 at the peptide exit of the *Escherichia coli* ribosome. *RNA* **9**, 566–573 (2003).
80. Wohlgenuth, I., Pohl, C., Mittelstaet, J., Konevega, A. L. & Rodnina, M. V. Evolutionary optimization of speed and accuracy of decoding on the ribosome. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2979–2986. <https://doi.org/10.1098/rstb.2011.0138> (2011).
81. Ziehr, D. R., Ellis, J. P., Culviner, P. H. & Cavagnero, S. Production of ribosome-released nascent proteins with optimal physical properties. *Anal. Chem.* **82**, 4637–4643. <https://doi.org/10.1021/ac902952b> (2010).

82. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675. <https://doi.org/10.1038/nmeth.2089> (2012).
83. Kawashima, S., Ogata, H. & Kanehisa, M. AIndex: Amino Acid Index Database. *Nucl. Acids Res.* **27**, 368–369. <https://doi.org/10.1093/nar/27.1.368> (1999).

### Acknowledgements

We thank M. Mittelstaet and A. L. Konevega for guidance during the initial phase of this study, I. Wohlgemuth for support with data analysis, interpretation and steric modeling, and F. Hummel, O. Geintzer, S. Kappler, C. Kothe, A. Pfeiffer, T. Steiger, and M. Zimmermann for expert technical assistance. This work was supported by the fellowship #17C174 from the Novartis Foundation for Medical-Biological Research to MT, the European Research Council (ERC) Advanced Investigator Grant RIBOFOLD to MVR (proposal number n° 787926) and by the Max Planck Society.

### Author contributions

M.T. prepared materials, designed constructs, and performed initial experiments. A.D. performed and analyzed the experiments. M.T. and M.V.R. interpreted the data and wrote the paper. All authors designed experiments and contributed to writing of the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-16932-7>.

**Correspondence** and requests for materials should be addressed to M.V.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022