How to cite: Angew. Chem. Int. Ed. 2020, 59, 21870–21874
International Edition: doi.org/10.1002/anie.202003779
German Edition: doi.org/10.1002/ange.202003779

Protein Modification

Suppression of Formylation Provides an Alternative Approach to Vacant Codon Creation in Bacterial In Vitro Translation

*Minglong Liu, Vito Thijssen, and Seino A. K. Jongkees**

Abstract: *Genetic code reprogramming is a powerful approach to controlled protein modification. A remaining challenge, however, is the generation of vacant codons. We targeted the initiation machinery of E. coli, showing that restriction of the formyl donor or inhibition of the formyl transferase during in vitro translation is sufficient to prevent formylation of the acylated initiating tRNA and thereby create a vacant initiation codon that can be reprogrammed by exogenously charged tRNA. Our approach conveniently generates peptides and proteins tagged N-terminally with non-canonical functional groups at up to 99% reprogramming efficiency, in combination with decoding the AUG elongation codons either with native methionine or with further reprogramming with azide- and alkyne-containing cognates. We further show macrocyclization and intermolecular modifications with these click handles, thus emphasizing the applicability of our method to current challenges in peptide and protein chemistry.*

Genetic code reprogramming provides many opportunities in the generation of biomolecules with novel and useful properties. Examples include antibody-drug conjugates, conjugated vaccines, mechanistic studies, and generating new catalysts.[1–3] Reprogramming of the genetic code requires a tRNA charged with an unnatural amino acid and a vacant codon. Multiple approaches have been developed for the first requirement and these are generally robust. For tRNA aminoacylation in cells or in vivo the most common approach is by engineered tRNA and amino-acyl tRNA synthetase pairs,[4] while in vitro more chemoenzymatic approaches are possible such as by ligation of a pdCpA-amino acid with a truncated tRNA,[5] or by acylation of an activated amino acid onto in vitro generated tRNA by a catalytic RNA called flexizyme.[6] Mis-recognition of near cognates by the endogenous amino-acyl tRNA synthetases also provides a convenient approach,[7] where possible. With flexizyme-mediated acylation a broad range of moieties have proven amenable to ribosomal translation, with the *N*-terminus proving particularly permissive and allowing boron clusters,[8] exotic peptide fragments,[9] and even foldamers.[10] In addition, the *N*-terminus of proteins is a common target for chemical modification.[11] Approaches for generating a vacant initiating codon are thus particularly valuable.

One crucial challenge that remains only partially solved is an efficient way to generate a vacant codon to reprogram. The most common current approach is reprogramming of a stop codon,[12] and is also compatible with initiation.[13] However, this is in competition with release factors, which in many cases results in low efficiency and truncated products. Sense codon reprogramming during in vitro translation can efficiently provide a vacant codon, but requires the use of a reconstituted recombinant translation system to allow for omission of canonical amino acids and prevents generation of sequences with all 20 canonical amino acids.[6,14] There clearly remains a demand for alternative approaches.

Recently, during efforts to generate multiply reprogrammed peptides using an in vitro translation system[15] we observed that methionine near-cognates such as homopropargylglycine (Hpg) and azidohomoalanine (Aha) competed with flexizyme-charged initiator aminoacyl-tRNA (Ac-Phe-$tRNA^{fMet}$), giving a mixture of reprogrammed and formyl-Hpg-initiated products (Figure 2). While not unexpected, as initiation with these near-cognates has been reported before,[16] it lead us to investigate ways of improving the ratio of desired (reprogrammed) product and to enable access to a broader range of N-terminal non-canonical amino acids.

Given that omitting the methionine cognate is not an option in this case, we turned our attention to the bacterial formylation apparatus. In bacterial initiation, the enzyme methionyl-tRNA formyltransferase (MTF) is responsible for transferring a formyl group from 10-formyltetrahydrofolate (10-CHO-THF) to methionyl-$tRNA^{fMet}$ generated by methionyl-tRNA synthetase (MetRS).[17] This formylated methionyl-tRNA (^{f}Met-$tRNA^{fMet}$) is then recognized by initiation factors (IF1, IF2-GTP, and IF3) and the ribosome to form the pre-30S initiation complex.[18,19] Typically the N-terminal methionine and its formyl group are removed following translation in a series of enyzmatic steps by peptide deformylase (PDF) and methionine amino peptidase (MAP). In some cases, however, the N-terminal processing is attenuated and this can be exploited by changing the identity of the second amino acid in the protein sequence to provide proteins that retain a methionine analogue.[20] It should also be noted that there also exist eubacteria that do not require formylation for initiation of protein synthesis.[21] Nonetheless, for *Escherichia coli* the formyl moiety is an important recognition factor for the initiation process,[22] and so we reasoned that if

[*] M. Liu, V. Thijssen, Dr. S. A. K. Jongkees
Department of Chemical Biology and Drug Discovery
Utrecht Institute for Pharmaceutical Sciences, Utrecht University
Universiteitsweg 99, 3584CG Utrecht (The Netherlands)
E-mail: s.a.k.jongkees@uu.nl

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202003779.

© 2020 The Authors. Published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited, and is not used for commercial purposes.

we could disrupt the activity of MTF we could prevent generation of fMet-tRNA^{fMet} or its ana-logues and thus bias initiation in favor of an exogenously charged aminoacyl-tRNA^{fMet}. To that end, in the current work we explore the possibility of preventing the formylation process in initiation of bacterial in vitro translation, either by substrate restriction or by addition of an inhibitor, to create vacant start codons without removal of any canonical amino acids (Figure 1).

Translation of a model peptide was carried out in the "PURExpress" in vitro translation sytem (a commercialised version of the PURE system[15]) with homopropargylglycine substituted for methionine and supplemented with 25 μM initiator aminoacyl-tRNA charged with *N*-acetyl phenylalanine by means of flexizyme[6] (an aminoacylating ribozyme), Ac-Phe-tRNA^{fMet}. This gave a mixture of reprogrammed and formyl-Hpg-initiated products in a ratio of roughly 2:1 (Figure 2A and S1). To prevent MTF activity, and thus hopefully suppress this undesired formylated product, we removed its source of the formyl group by using a previously reported custom energy solution (SolAFD-)[6] that omits 10-CHO-THF. Using this in an otherwise identical translation to above, still substituting homopropargylglycine for methionine and adding 25 μM of Ac-Phe-tRNA^{fMet}, changed the ratio of desired product to side product to 92:8, while also changing the nature of the side product to the non-formylated form (Figure S2). Increasing the concentration of Ac-Phe-tRNA^{fMet} to 50 μM was found to give only the desired product **P1**, with all initiation side-products being undetectable by UPLC-MS (Figure 2B and S3). Simple removal of 10-CHO-THF from the translation is thus sufficient for reprogramming of the initiation codon even in the presence of a substrate for MetRS, provided that sufficient amounts of exogenous aminoacyl-tRNA^{fMet} are added.

As an alternative to 10-CHO-THF omission, we also investigated whether MTF activity could be suppressed by a competitive inhibitor to achieve the same result. This removes the need for a custom energy solution, allowing reprogramming with only simple commercial reagents. One candidate inhibitor was found in methotrexate, a chemotherapy drug that targets dihydrofolate reductase and which is a close analogue of 10-CHO-THF. This has previously been shown to be a modest inhibitor of algal MTF.[23] Adding

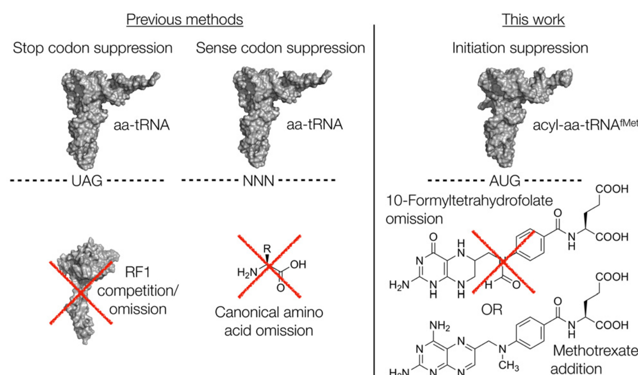

Figure 1. Previous and current approaches to vacant codon creation. aa = amino acid, RF1 = release factor 1.

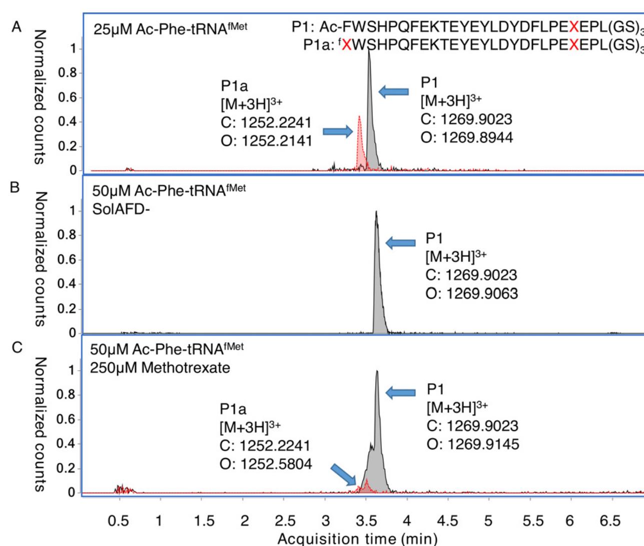

Figure 2. UPLC-MS extracted-ion chromatograms (EIC) from analyses of translation tests with reprogrammed initiation. A) Competition between 25 μM Ac-Phe-tRNA^{fMet} and endogenous initiators (69:31 **P1**/**P1a**). B) As for (A) but using SolAFD- and with 50 μM Ac-Phe-tRNA^{fMet} (100% **P1**, byproduct undetectable). C) As for (A), but with 250 μM methotrexate and 50 μM Ac-Phe-tRNA^{fMet} (93:7 **P1**/**P1a**). X = homopropargylglycine, ^{f} = *N*-formylation.

25 μM Ac-Phe-tRNA^{fMet} and 100 μM methotrexate to a translation reaction using the commercial translation kit components gave an improved fraction of reprogrammed initiation product (83%, Figure S4), with the only detectable side product still being initiation with formylated homopropargylglycine (17%). This reaction contains the substrates 10-CHO-THF and homopropargylglycine as well as the enzymes MetRS and MTF, and so should be formylation-competent. Increasing the methotrexate concentration to 250 μM and the Ac-Phe-tRNA^{fMet} concentration to 50 μM gave a further improvement to 93% fractional yield of the fully reprogrammed product (Figure 2C and S5). This indicates that initiation reprogramming by MTF suppression by addition of an inexpensive commercially available inhibitor is compatible with the presence of all natural translation components.

An additional trace side product that could be detected in some cases was a peptide starting from the second amino acid ("initiator-truncated"), which is also observed for vacant codon creation by methionine omission in cases of less efficient reprogramming.[10,24] A control reaction using the custom SolAFD- but with no flexizyme-mediated reprogramming (still containing the Met analogue Hpg) afforded peptide product with the initiating amino acid present in the formylated form (**P2a**; 25%), to our surprise, as well as the free amine (**P2b**; 63%) and the initiator-truncated peptide (**P2c**; 12%; Figure 3A and S6). To investigate whether this initiator-truncated product can be exploited as a further alternative to initiation with formyl-methionine, we tested whether addition of excess un-acylated in vitro-transcribed tRNA^{fMet} to the translation mix would improve the yield and purity of initiator-truncated product. An additional 50 μM tRNA^{fMet} indeed gave clean production of an abundant amount of this initiator-truncated product (Figure 3B and

 © 2020 The Authors. Published by Wiley-VCH GmbH

Figure 3. EIC from UPLC-MS analyses of translation without reprogramming initiation. A) Control translation in the absence of both formyl donor and reprogrammed initiator tRNA (25:63:12 **P2a/P2b/P2c**). B) As for (A) but with additional uncharged tRNA^{fMet}. X = homopropargylglycine, ^{f} = *N*-formylation.

S7). Although accurate absolute quantification of the product is not possible with the UPLC-MS method used, this reaction gave no detectable peptide side products and peak area was comparable to the initiation-reprogrammed translations above. Free tRNA^{fMet} is clearly initiation-competent in the PURE system, and addition of an excess is a practical solution to its low efficiency. This is consistent with reported upregulation of expression of tRNA^{fMet} as an adaptation to deletion or inhibition of MTF in *E. coli*, allowing initiation of protein synthesis to resume.[25] Uncharged tRNA addition thus provides an alternative approach for production of peptides with a free amine at the *N*-terminus, as compared to post-translational removal of the initiating formyl methionine by the enzymes peptide deformylase and methionine aminopeptidase (PDF/MAP).[26] Our approach does not require any additional proteins to be purified and also removes restrictions imposed by MAP's substrate scope for the amino acid in the second position.[27]

Encouraged by these results, we next investigated the scope of the 10-CHO-THF omission approach for initiation with amino acids bearing different functional groups as *N*-acyl modifications. In addition to simple capping by an acetyl group, we investigated a chloroacetyl group for thioether head-to-sidechain cyclisation with cysteine,[28] an aminomethylbenzyl group for fluorogenic oxidative cyclization/derivatization reactions with hydroxytryptophan,[29] a fluorescein derivative, and biotin. For decoding of the elongation AUG codon we included two "clickable" near cognates of methionine, homopropargylglycine and azidohomoalanine. In all cases a good ratio of desired product to side-products was obtained, and in comparable yields to a non-reprogrammed translation by peak area in UPLC-MS (Table 1 and S1, Figures S1–S18). Methionine is also able to be included in this approach (Table 1 entry 9 and Figure S17), allowing translation of otherwise completely native sequences with only initiation reprogrammed and showing that generation of the vacant codon is not dependent on the use of a near-cognate. These reactions were not further optimized with additional exogenous initiating tRNA and observed variations in flexizyme acylation yields were not accounted for (Figure S19), so there remains scope to optimize in cases where complete suppression of the side products is needed. Unlike recently reported approaches to *N*-terminal reprogramming that are based on engineered bacteria,[30,31] our approach allows convenient access to a much broader scope of functional groups as it does not need orthogonal tRNA/aminoacyl tRNA synthetase pairs to fill the vacated initiation codon.

Given that we were able to show translation of clickable near-cognates of methionine along with reprogramming of the initiating position, we investigated the application of copper-catalyzed azide-alkyne cycloaddition reactions to macrocyclization of the products.[32] A peptide **P3** initiated with *N*-(4-pentynoyl) phenylalanine ("Pyn-Phe") and containing a downstream azidohomoalanine was treated with Cu^{I} directly in the translation mixture. Clean conversion to the macrocyclized product **P4** was shown by retention shift of the isobaric product (Figure 4) and further confirmed by treatment with tris(2-carboxyethyl)phosphine (TCEP), which gave no detectable azide reduction. This click macrocyclization reaction is particularly convenient as it was carried out without any intermediate purification or removal of oxygen, and is anticipated to be compatible with mRNA display as well as further

Table 1: Products of translation reprogrammed with various combinations of initiating amino acid and methionine analogues for the peptide sequence (M)WSHPQFEKTEYEYLYDYDFLPE(M)EPLGSGSGS.

Entry	Initiator[a] (R)	Elongator[a] (R′)	Product ratio (*N*-terminus as shown)[b] (reprogrammed)	(formyl-Met or analogue)	(Met or analogue)	(free amine)
1	ClAc-Tyr	Hpg	> 99	0	0	0
2	ClAc-Tyr	Aha	> 99	0	0	0
3	AMB-Phe	Hpg	77	4	14	5
4	AMB-Phe	Aha	65	5	14	16
5	(5/6)FAM-Phe	Hpg	78	4	9	9
6	(5/6)FAM-Phe	Aha	72	7	11	10
7	Biotin-Phe	Hpg	> 99	0	0	0
8	Biotin-Phe	Aha	> 99	0	0	0
9	Ac-Phe	Met	> 99	0	0	0
10	Pyn-Phe	Aha	> 99	0	0	0

[a] Aha: azidohomoalanine, Hpg: homopropargylglycine, AMB: 4-aminomethylbenzoic acid, ClAc: chloroacetyl, FAM: carboxyfluorescein, Pyn: 4-pentynoic acid. [b] Determined by relative peak areas in UPLC-MS EIC traces. R and R′ reflect the initiator and AUG elongator, respectively.

© 2020 The Authors. Published by Wiley-VCH GmbH

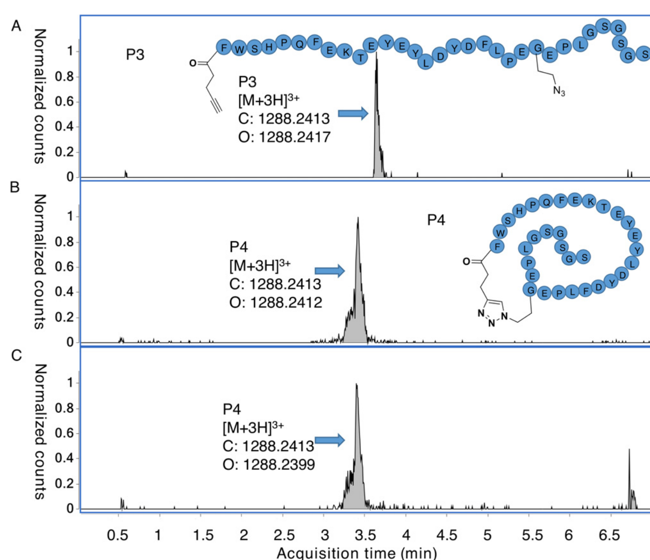

Figure 4. EIC from UPLC-MS analyses of peptide macrocyclisation following reprogrammed translation. A) Peptide translation with an alkynyl initiator and downstream azidohomoalanine (100% **P3**, no detectable side products). B) The same peptide after overnight treatment with 1.2 mM Cu^{I}.THPTA and aminoguanidine hydrochloride scavenger. C) Macrocyclized product from (B) treated with 35 mM TCEP for one hour (mass peak for reduction not detectable).

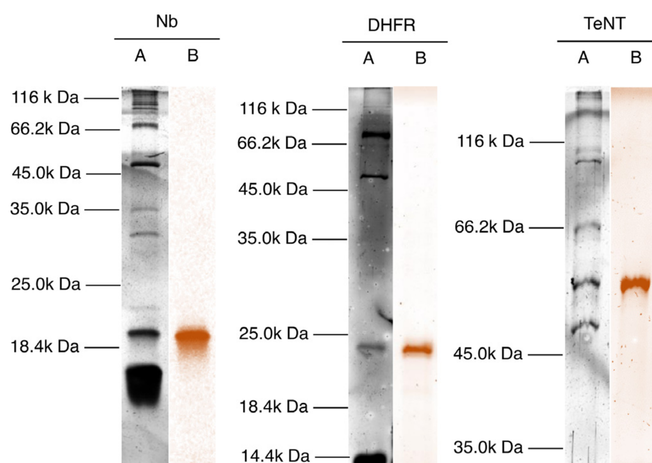

Figure 5. SDS-PAGE showing translation of proteins under a multiply reprogrammed genetic code (Biotin-Phe initation and Aha decoding of AUG elongator codons). A: silver staining, B: Cy3 fluorescence (color applied in processing), Nb: nanobody, DHFR: dihydrofolate reductase, TeNT: tetanospasmin toxoid.

modifications such as chemical *S*-glycosylation at cysteine residues.[33]

Finally, we tested translation of a small panel of proteins with (+)-biotinyl phenylalanine as the initiating amino acid and azidohomoalanine reprogramming the elongation AUG codons, with formylation suppressed by methotrexate inhibition of MTF. Chosen as model proteins were a nanobody, a protein used in folding studies (*E. coli* dihydrofolate reductase), and a carrier protein used in conjugate vaccine development (tetanospasmin toxoid). Protein products were reacted with DBCO-Sulfo-Cy3 in a copper-free click reaction under mild denaturing conditions, then pulled down by magnetic streptavidin beads and analyzed by SDS-PAGE gel detected by both fluorescence and silver staining (Figure 5 and S20). Pull-down by biotin confirms the initiation reprogramming while attachment of fluorophore confirms orthogonal reprogramming of the elongator AUG codons, together demonstrating facile generation of doubly reprogrammed proteins amenable to further covalent modification. Quantification of yield was estimated by densitometry of the silver stain band for a "non-clicked" product referenced to a standard curve of bovine serum albumin to give 40 ngμL^{-1} (Figure S21), consistent with the manufacturer's estimated yield for non-reprogrammed translation. We envisage our method to be suited for applications such as generating immobilized proteins for protein interaction experiments, conjugate vaccines carrying multiple epitopes, and generating proteins with multiple fluorophores for conformational studies.

We have shown here that suppression of formylation is sufficient to liberate the initiation codon in *E. coli*-derived in vitro translation, achievable by removing the formyl donor or by inhibiting the enzyme MTF. We have demonstrated reprogramming of the genetic code on this vacated codon using amino acids acylated with non-canonical functional groups such as reactive handles, a fluorophore, and biotin, and we have shown this to be applicable to protein translation as well as peptides. This initiation reprogramming is compatible with translation of downstream AUG codons by methionine, or exploiting mis-acylation of near cognates by MetRS to incorporate clickable handles to provide convenient access to doubly modified proteins. The method we report here provides a novel and practical approach to creating a vacant initiation codon, and allows rapid access to *N*-terminal reprogramming with a broad range of chemical functionality using simple molecular biology techniques and easily accessible reagents.

Acknowledgements

We are grateful for funding support provided by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 746631 to SAKJ, and for a CSC scholarship to ML. We thank Prof. P van Bergen en Henegouwen and Prof. GJPH Boons (both of Utrecht University) for providing Nb and TeNT plasmids, respectively.

Conflict of interest

The authors declare no conflict of interest.

Keywords: Bioconjugation · cyclic peptides · genetic code reprogramming · protein modification · protein engineering

© 2020 The Authors. Published by Wiley-VCH GmbH

[1] S. L. Richardson, K. K. Dods, N. A. Abrigo, E. S. Iqbal, M. C. Hartman, *Curr. Opin. Chem. Biol.* **2018**, *46*, 172–179.
[2] Y. Huang, T. Liu, *Synth. Syst. Biotechnol.* **2018**, *3*, 150–158.
[3] K. Wals, H. Ovaa, *Front. Chem.* **2014**, *2*, 15.
[4] "Directed Evolution of Orthogonal Pyrrolysyl-tRNA Synthetases in Escherichia coli for the Genetic Encoding of Noncanonical Amino Acids": M. J. Schmidt, D. Summerer, in *Noncanonical Amino Acids*, Springer, Berlin, **2018**, pp- 91–111.
[5] M. Duca, S. Chen, S. M. Hecht, *Methods* **2008**, *44*, 87–99.
[6] Y. Goto, T. Katoh, H. Suga, *Nat. Protoc.* **2011**, *6*, 779–790.
[7] M. C. T. Hartman, K. Josephson, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 4356–4361.
[8] Y. Yin, N. Ochi, T. W. Craven, D. Baker, N. Takigawa, H. Suga, *J. Am. Chem. Soc.* **2019**, *141*, 19193–19197.
[9] Y. Goto, H. Suga, *J. Am. Chem. Soc.* **2009**, *131*, 5040–5041.
[10] J. M. Rogers, S. Kwon, S. J. Dawson, P. K. Mandal, H. Suga, I. Huc, *Nat. Chem.* **2018**, *10*, 405–412.
[11] C. B. Rosen, M. B. Francis, *Nat. Chem. Biol.* **2017**, *13*, 697–705.
[12] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182–188.
[13] S. Mamaev, J. Olejnik, E. K. Olejnik, K. J. Rothschild, *Anal. Biochem.* **2004**, *326*, 25–32.
[14] Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami, H. Suga, *ACS Chem. Biol.* **2008**, *3*, 120–129.
[15] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, T. Ueda, *Nat. Biotechnol.* **2001**, *19*, 751–755.
[16] S. Horiya, J. K. Bailey, J. S. Temme, Y. V. Guillen Schlippe, I. J. Krauss, *J. Am. Chem. Soc.* **2014**, *136*, 5407–5415.
[17] D. Kahn, M. Fromant, G. Fayat, P. Dessen, S. Blanquet, *Eur. J. Biochem.* **1980**, *105*, 489–497.
[18] A. Antoun, M. Y. Pavlov, M. Lovmar, M. Ehrenberg, *EMBO J.* **2006**, *25*, 2539–2550.
[19] A. Zorzet, M. Y. Pavlov, A. I. Nilsson, M. Ehrenberg, D. I. Andersson, *Mol. Microbiol.* **2010**, *75*, 1299–1313.
[20] B. Wiltschi, L. Merkel, N. Budisa, *ChemBioChem* **2009**, *10*, 217–220.
[21] D. T. Newton, C. Creuzenet, D. Mangroo, *J. Biol. Chem.* **1999**, *274*, 22143–22146.
[22] U. Varshney, U. L. RajBhandary, *J. Bacteriol.* **1992**, *174*, 7819–7826.
[23] A. Gambini, P. Crosti, R. Bianchetti, *Biochim. Biophys. Acta Enzymol.* **1980**, *613*, 73–78.
[24] J. Lee, K. E. Schwieter, A. M. Watkins, D. S. Kim, H. Yu, K. J. Schwarz, J. Lim, J. Coronado, M. Byrom, E. V. Anslyn, et al., *Nat. Commun.* **2019**, *10*, 5097.
[25] A. I. Nilsson, A. Zorzet, A. Kanth, S. Dahlström, O. G. Berg, D. I. Andersson, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 6976–6981.
[26] T. Kawakami, A. Ohta, M. Ohuchi, H. Ashigai, H. Murakami, H. Suga, *Nat. Chem. Biol.* **2009**, *5*, 888–890.
[27] Q. Xiao, F. Zhang, B. A. Nacev, J. O. Liu, D. Pei, *Biochemistry* **2010**, *49*, 5588–5599.
[28] K. Iwasaki, Y. Goto, T. Katoh, H. Suga, *Org. Biomol. Chem.* **2012**, *10*, 5783–5786.
[29] Y. Yamagishi, H. Ashigai, Y. Goto, H. Murakami, H. Suga, *ChemBioChem* **2009**, *10*, 1469–1472.
[30] R. M. Vincent, B. W. Wright, P. R. Jaschke, *ACS Synth. Biol.* **2019**, *8*, 675–685.
[31] J. M. Tharp, O. Ad, K. Amikura, F. R. Ward, E. M. Garcia, J. H. D. Cate, A. Schepartz, D. Söll, *Angew. Chem. Int. Ed.* **2020**, *59*, 3122–3126; *Angew. Chem.* **2020**, *132*, 3146–3150.
[32] Y. Sako, J. Morimoto, H. Murakami, H. Suga, *J. Am. Chem. Soc.* **2008**, *130*, 7232–7234.
[33] S. A. K. Jongkees, S. Umemoto, H. Suga, *Chem. Sci.* **2017**, *8*, 1474–1481.

Manuscript received: March 13, 2020
Revised manuscript received: August 4, 2020
Accepted manuscript online: August 25, 2020
Version of record online: September 28, 2020

 © 2020 The Authors. Published by Wiley-VCH GmbH