N-Methyl Amino Acids

DOI: 10.1002/anie.201007686

Optimal Codon Choice Can Improve the Efficiency and Fidelity of *N*-Methyl Amino Acid Incorporation into Peptides by In-Vitro Translation**

*Alexander O. Subtelny, Matthew C. T. Hartman, and Jack W. Szostak**

The generation of peptides containing *N*-methyl amino acids (*N*-Me AAs) by in-vitro translation is of interest as a potential route to the directed evolution of pharmaceutically useful modified peptides.[1–11] However, *N*-Me AAs are suboptimal substrates for the translational apparatus,[12] and as a result, premature termination of translation can occur or incorrect amino acids can be misincorporated in place of the desired *N*-Me AA, resulting in reduced incorporation efficiency and fidelity of the *N*-Me AA. We sought to determine the main cause of such misincorporation and truncation events, and then to optimize the incorporation of *N*-Me AAs.

We examined the efficiency and fidelity with which three different *N*-Me AAs (*N*-Me Leu, *N*-Me Thr, and *N*-Me Val) are incorporated at a single position in a short peptide (MH_6MX_mEP, X_m = *N*-Me AA, M = Met, E = Glu, P = Pro, H = His) using each of the codons of the corresponding natural amino acid (Figure 1). We showed previously that these three *N*-Me AAs can be incorporated into peptides by in-vitro translation[11] using the fully reconstituted PURE system for translation (PURE = protein synthesis using recombinant elements, an in vitro system containing only purifies components).[13] We directed the incorporation of these *N*-Me AAs into peptides by supplying in-vitro translation reaction mixtures with total tRNA that had been enzymatically precharged with Leu, Thr, or Val and then chemically *N*-methylated as previously described (see Experimental Section).[7,11] The presence of the other 19 amino acid/aminoacyl-tRNA synthetase (AARS) pairs in the translation reaction allowed us to determine the propensity of these aminoacyl-tRNAs (AA-tRNAs) to be misincorporated. For each *N*-Me AA, we identified the codon that minimized competing misincorporation of natural amino acids carried by tRNAs with near-complementary anticodons. We required that this codon also afford highly efficient incorporation of

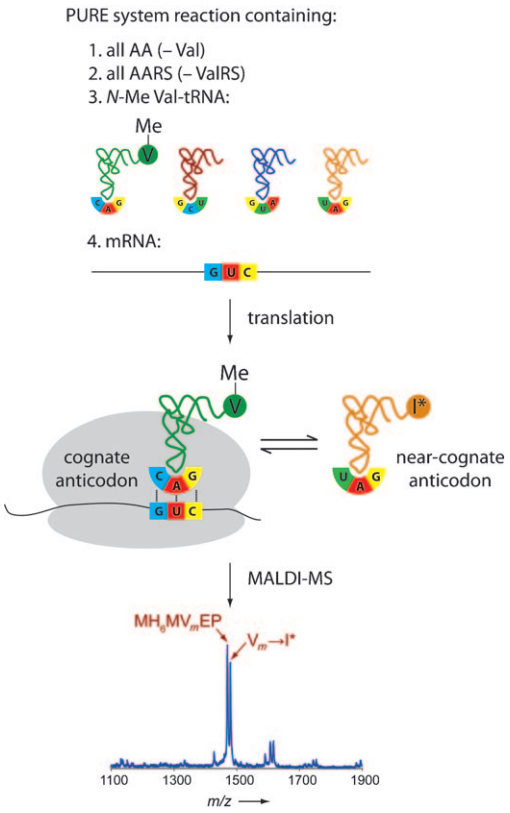

Figure 1. Assay to screen for codons that direct efficient and faithful incorporation of an *N*-Me AA into a short peptide (MH_6MX_mEP, X_m = *N*-Me AA), using *N*-Me Val and the GUC codon as an example. In-vitro translation reactions using the PURE system were carried out with 1) total tRNA that had been enzymatically precharged with Val and chemically *N*-methylated (*N*-Me Val-tRNA), 2) all 19 other natural amino acid/aminoacyl-tRNA synthetase pairs, and 3) mRNA encoding MH_6MV_mEP where *N*-Me Val incorporation was directed by the GUC codon. Translation products were analyzed by MALDI-TOF MS to assess the efficiency of *N*-Me Val incorporation (i.e. absence of truncation products) and the extent of misincorporation of non-cognate amino acids. In the example above, using the GUC codon resulted in significant misincorporation of Ile (I; the asterisk represents an isotope label to break the mass degeneracy with *N*-Me Val). After repeating the assay for each of the remaining Val codons, the codon(s) that yielded the *N*-methyl peptide with no misincorporation or truncation products were judged to be optimal.

[*] A. O. Subtelny, Prof. Dr. M. C. T. Hartman, Prof. Dr. J. W. Szostak
Howard Hughes Medical Institute, Department of Molecular Biology and Center for Computational and Integrative Biology
Massachusetts General Hospital
185 Cambridge Street, Boston, MA 02114 (USA)
Fax: (+1) 617-643-3328
E-mail: szostak@molbio.mgh.harvard.edu
Prof. Dr. M. C. T. Hartman
Department of Chemistry and Massey Cancer Center, Virginia Commonwealth University (USA)
A. O. Subtelny
Department of Biology, Massachusetts Institute of Technology (USA)

[**] We thank Drs. A. Bell, Y. Guillen, K. Josephson, C.-W. Lin, F. Seebeck, D. Treco, R. Green, and U. RajBhandary for reagents and useful discussions. This work was supported by the Howard Hughes Medical Institute, and A.O.S. was supported in part by the Harvard College Research Program.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201007686.

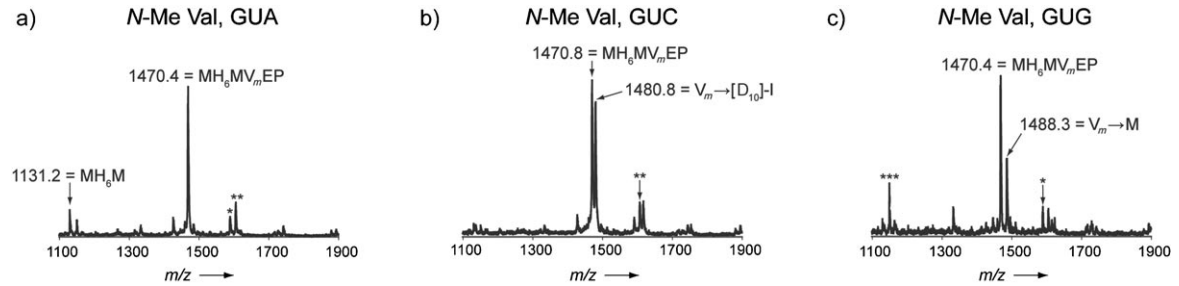

Figure 2. Efficiency and fidelity of *N*-Me Val incorporation when directed by one of three Val codons (GUA, GUC, or GUG). a)–c) MALDI-TOF mass spectra of the products of translation reactions supplied with *N*-Me Val-tRNA and mRNA encoding the peptide MH_6MV_mEP (m_{calcd} = 1470.7). Misincorporation products are indicated by an arrow from the templated to the misincorporated amino acid, whereas truncation products are denoted by the sequence of the shortened peptide. The asterisk (*) and triple asterisk (***) indicate peaks corresponding to unknown species (m_{obs} = 1591 and 1151, respectively), while the double asterisk (**) peak is consistent with the sodium trifluoroacetate adduct of the protonated *N*-methyl peptide (m_{obs} = 1608).

the *N*-Me AA, as defined by the lack of truncation products arising from premature termination of translation.

N-Me Val was most faithfully incorporated when encoded by the GUA codon, with MALDI-TOF mass spectrometry analysis of the translation products showing one major peak consistent with the desired *N*-methyl peptide (Figure 2a; Table S1 in the Supporting Information). In contrast, the three other valine codons (GUC, GUG, and GUU) resulted in more significant misincorporation (Figure 2b,c; and Figure S1f in the Supporting Information). Since Ile, Leu, and Asn have very similar or identical masses to *N*-Me Val such that misincorporation of one of these would yield an indistinguishable MS peak, we verified that using the GUA codon indeed furnished the *N*-methyl peptide by repeating the experiment with *N*-Me Val-tRNA where Val was deuterium ($[D_8]$) labeled. As expected, the major peptide peak shifted by 8 mass units, showing that no other cryptic peptide of the original mass was present (Figure S1c in the Supporting Information). When *N*-Me Val was encoded by the GUG codon, a peak corresponding to Met misincorporation was observed, consistent with binding of $tRNA^{Met}$ owing to G:U base-pairing between the 5′ G of the codon and the 3′ U of the N^4-acetyl-CAU anticodon (Figure 2c).[14] A translation using *N*-Me $[D_8]$Val-tRNA shifted the original mass peak and ruled out the presence of cryptic mistranslated peptides (Figure S1e in the Supporting Information). When the GUC codon was used, a single major peak with the expected peptide mass was observed. Performing the same experiment with deuterated analogues of Ile, Leu and Asn ($[D_{10}]$Ile, $[D_7]$Leu and $[D_3]$Asn) to unambiguously identify any misincorporation products revealed another large peak corresponding to misincorporation of $[D_{10}]$Ile ($[D_{10}]$I; Figure 2b). Since no additional misincorporations were revealed upon repeating the translation with both *N*-Me $[D_8]$Val-tRNA and free $[D_{10}]$Ile (Figure S1d in the Supporting Information), Ile misincorporation is the major process competing with incorporation of *N*-Me Val, when directed by the GUC codon. Similar but less extensive misincorporation of $[D_{10}]$Ile was also seen with the GUU codon (Figure S1f in the Supporting Information). Misrecognition of the GUC and GUU codons by $tRNA^{Ile}$ may result from the predominant Ile isoacceptor, which recognizes the AUC and AUU codons, forming a G:U base-pair at the 5′ position of these Val codons.[14]

A parallel series of experiments with *N*-Me Thr showed good efficiency and fidelity when incorporation was directed by the ACU codon or, to a lesser extent, the ACA codon (Figure 3a,c; Figure S2c,f and Table S2 in the Supporting Information). In contrast, use of the ACG codon resulted in Met misincorporation and the MH_6M truncation product (Figure 3b; Figure S2e in the Supporting Information), while the ACC codon led to some Ser misincorporation (Figure S2d in the Supporting Information).

In the case of *N*-Me Leu, the CUG and CUU codons yielded the desired *N*-methyl peptide with few misincorporation or truncation products (Figure 4a; Figure S3e,f and Table S3 in the Supporting Information). The CUC codon also afforded efficient incorporation of *N*-Me Leu with no misincorporation and only low levels of truncation (Figure S3d in the Supporting Information). In contrast, the CUA codon did not allow for efficient incorporation of *N*-Me Leu, with the

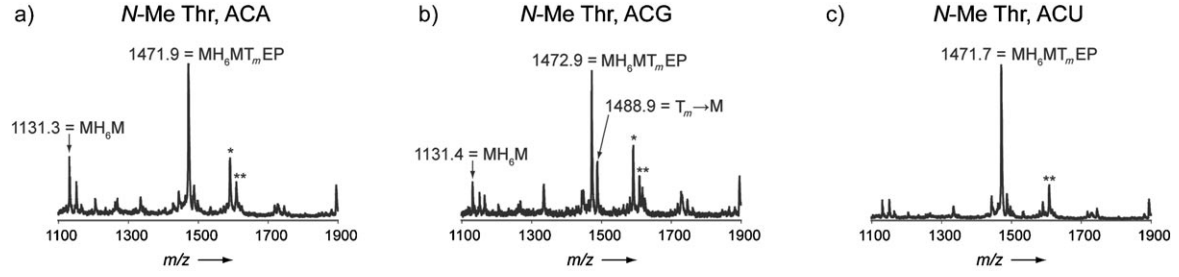

Figure 3. Efficiency and fidelity of *N*-Me Thr (T_m) incorporation when encoded by different Thr codons (ACA, ACG, or ACU). a)–c) MALDI-TOF mass spectra of the products of translation reactions supplied with *N*-Me Thr-tRNA and mRNA encoding MH_6MT_mEP (m_{calcd} = 1472.6). Misincorporation products, truncation peptides and other minor peaks are indicated as in Figure 2.

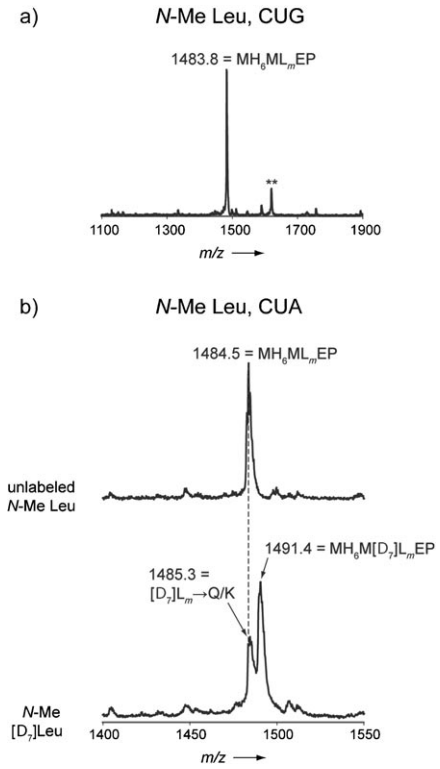

Figure 4. a) Incorporation of *N*-Me Leu when directed by the CUG codon. Misincorporation products, truncation peptides and other minor peaks are indicated as in Figure 2. b) Incorporation of *N*-Me Leu when templated by the CUA codon. Use of this codon resulted in the formation of a Gln (Q) or Lys (K) misincorporation product (m_{calcd} = 1485.6 or 1485.7, respectively), which has a very similar mass to the desired *N*-Me Leu containing peptide (m_{calcd} = 1484.7), resulting in overlapping MS peaks that are indistinguishable (top). The presence of the misincorporation product was revealed by performing an identical translation with isotopically labeled *N*-Me Leu (*N*-Me [D_7]Leu), which separated the two peaks (bottom). For the full mass spectrum for each experiment, see Figure S3c in the Supporting Information.

mass spectrum showing a large peak representing truncation product and another corresponding to Gln or Lys misincorporation whose presence was revealed through isotope labeling of *N*-Me Leu (Figure 4b; Figure S3c in the Supporting Information). The UUA and UUG codons yielded the MH_6M truncation peptide as the major product, with the desired *N*-Me peptide present only as a minor peak (Figure S3g,h in the Supporting Information). The increased amount of truncation product could result from the lower binding energy associated with a larger number of A:U base pairs in the codon–anticodon interaction. It is likely not due to low levels of the corresponding tRNAs, since each of these is at least as abundant as the relatively inabundant cognate tRNA for the CUC codon, which yields efficient incorporation of *N*-Me Leu.[14] However, other *N*-Me AAs templated by A/U-rich codons (such as *N*-Me Asn, encoded by AAC) are incorporated with reasonable efficiency under the same translation conditions.[11] Others have also reported that unnatural amino acid incorporation efficiency does not necessarily correlate with codon–anticodon pairing stability when using another in-vitro translation system.[15]

In summary, experiments in which the codon templating the incorporation of *N*-Me Val, *N*-Me Thr, and *N*-Me Leu into a short peptide was varied identified one or more codons for each analogue (GUA for *N*-Me Val, ACU for *N*-Me Thr, and CUG or CUU for *N*-Me Leu) that afforded optimal incorporation efficiency and fidelity. Since these codons generally correspond to the more abundant tRNA isoacceptors for each amino acid, it is possible that the higher levels of misincorporation seen with other codons resulted from depletion of the corresponding *N*-Me AA-tRNAs during translation to such low levels that AA-tRNAs with near-cognate anticodons could compete successfully for incorporation. In cases where no other AA-tRNA could recognize the codon, premature termination occurred. As a corollary, when choosing between synonymous codons that direct the efficient and faithful incorporation of an *N*-Me AA in our assay for more extended incorporation experiments, involving longer reaction times or multiple incorporations of the *N*-Me AA into the same peptide, the codon with the most abundant cognate isoacceptor (e.g. CUG for *N*-Me Leu)[14] would be expected to be optimal. This "abundance" hypothesis is supported by previous experiments in which the incorporation of multiple *N*-methyl residues was enhanced by periodic supplementation of the translation reaction mixture with additional aliquots of *N*-Me AA-tRNA.[11] In another study which measured the rescue of inactive, single-residue luciferase mutants by misincorporation of the wild type amino acid in *E. coli* cells, rescue was greater when the mutant codon had less-abundant cognate tRNAs, presumably because these could compete less effectively with the near-cognate, wild-type AA-tRNA.[16] An alternative hypothesis is that optimal incorporation efficiency and fidelity of an *N*-Me AA could require a codon-specific sequence context.[15] Finally, the variation in incorporation efficiency and fidelity among synonymous codons could be due to differences in the kinetics of cognate isoacceptor recognition and GTP hydrolysis by the thermally unstable elongation factor EF-Tu, accommodation of the *N*-Me AA-tRNA, and/or peptidyl transfer.[12]

Although effective for reactions of short duration, supplementing translations with more *N*-Me AA-tRNA quickly becomes impractical owing to the limited amount of total tRNA that can be added to a translation reaction. Supplementation with individual purified tRNAs that have been aminoacylated and *N*-methylated has been shown to allow for improved incorporation of *N*-Me AAs.[11] Clearly, the ideal approach would involve the continuous regeneration of *N*-Me AA-tRNAs during in-vitro translation. The "flexizyme" aminoacyl-tRNA synthetase ribozyme might provide one approach to this problem, although it would be limited to a single amino acid until variants with good amino acid specificity are evolved.[17] Moreover, its charging efficiency would have to be optimized for the conditions of translation reactions. Alternatively, the evolution or design of AARS variants with high specificity for the *N*-Me residue (or other modified residue of choice) could in principle allow for continuous regeneration of charged tRNA and thus enhanced efficiency and fidelity of incorporation.

The results we have described above have significant implications for the design of random sequence libraries, and for the synthesis of modified peptides by in-vitro translation. Libraries employed for peptide and protein selection experiments are typically generated using random or biased nucleotide compositions; thus, many codons will afford suboptimal incorporation efficiency and fidelity for modified amino acids, making it more difficult to evolve peptides with desired properties. Our results support strategies of library synthesis based on the assembly of triplet building blocks, with each triplet corresponding to the optimal codon for each unnatural amino acid.

Experimental Section

All materials were as described previously.[11]

Synthesis of *N*-methyl aminoacyl-tRNA: We followed the strategy of Merryman and Green,[7] with modifications as described previously.[11] Briefly, *E. coli* total tRNA was charged with a natural amino acid (4 mM) using the corresponding aminoacyl-tRNA synthetase (AARS; 10–20 μM) for 30 min at 37 °C in charging buffer (for composition, see Ref. [11]). The aminoacylation reaction was quenched by adding 0.1 volume 3 M NaOAc pH 5.0, and the AA-tRNA purified by phenol-chloroform extraction and ethanol precipitation. To protect against bis-methylation of the amino group, the AA-tRNA was mono-derivatized with *o*-nitrobenzaldehyde (30 mM) at 37 °C for 1 h in the presence of sodium cyanoborohydride (20 mM) under slightly acidic conditions (55 mM NaOAc, 33% dioxane, pH 5.0). Formaldehyde was then added to 10 mM, and the reaction incubated at room temperature for another 1 h. The methylation reaction was quenched by adding 0.1 volume 4.4 M NH_4OAc pH 5.0, and the resulting bis-alkylated AA-tRNA purified by ethanol precipitation. Removal of the photolabile *o*-nitrobenzyl group was achieved by irradiating the bis-alkylated AA-tRNA with UV light from a 450 W Hanovia lamp for 30 min. The resulting mono-*N*-methyl AA-tRNA was ethanol precipitated and resuspended in 5 mM KOAc pH 5.0.

Preparation of mRNA templates: Synthetic oligonucleotide duplexes encoding the C-terminal four amino acids of the peptide MH_6MX_mEP (X_m = *N*-Me AA) were cloned into BamHI/NdeI-digested pETDual (a pET12b derivative possessing a His_6-encoding sequence upstream of the NdeI site). The resulting constructs were transformed into *E. coli* Top10 cells (Invitrogen), which were selected on ampicillin-containing LB (Luria broth) plates. Single colonies were picked and used to initiate 3 mL overnight LB cultures from which plasmid was purified using a Plasmid Mini purification kit (Qiagen). This plasmid was used as a template for polymerase chain reactions (PCR) using primers that corresponded to the T7 promoter and terminator sequences. After PCR products were purified using a QIAquick PCR Purification Kit (Qiagen), they were added to overnight in-vitro transcription reactions with T7 RNA polymerase.[18] The reactions were extracted with phenol–chloroform and the transcripts precipitated with 2.5 M LiCl (1 h on ice, 30 min at −20 °C) and again with ethanol before being resuspended in water.

In-vitro translations: Translation reactions were performed using the PURE system, whose composition was as described previously,[11] with the following modifications. With the exception of the natural form of the *N*-Me AA and the corresponding AARS, all amino acids (100 μM each, unless otherwise indicated) and AARSs were present during translations. For some experiments, Asn, Ile, and Leu were replaced with the isotope-labeled analogues $[D_3]$Asn, $[D_{10}]$Ile, and $[D_7]$Leu at a final concentration of 50 μM each. Peptides were radiolabeled with 0.1 μM ^{35}S-Met (2600 dpm/fmol) and 10 μM Met.

50 μL translation reactions were supplied with 1.5 A_{260} units of total tRNA that had been enzymatically precharged with Val, Leu, or Thr (or their stable isotope-labeled equivalents) and chemically *N*-methylated. Messenger RNA was added to 1.0 μM, and the reactions incubated at 37 °C for 30 min before being quenched by the addition of 100 μL wash buffer (50 mM Tris-HCl, 300 mM NaCl, pH 8.0) and supplemented with 50 μL of suspended Ni-NTA (NTA = nitrilotriacetic acid) agarose beads (Qiagen). The beads were incubated with the reaction mixture on a rotator at room temperature for 1 h, washed twice with wash buffer and twice with deionized water, and the bound peptides eluted with 50 μL 1% trifluoroacetic acid for 15 min.

After elution from the Ni-NTA agarose beads, the translation products were supplemented with 500 fmol of each of two peptide mass standards (angiotensin II, $m = 1046.5$; $m = 3494$). They were then further purified and concentrated using C18 ZipTip reverse phase microchromatography columns (Millipore, Billerica, MA). The bound peptides were eluted with a saturated α-cyano-4-hydroxycinnamic acid solution containing 50% acetonitrile and 0.1% trifluoroacetic acid, and spotted onto a MALDI target plate. Mass spectra were obtained using an Applied Biosystems Voyager MALDI-TOF.

Received: December 7, 2010
Published online: March 4, 2011

Keywords: amino acids · codon–anticodon recognition · directed evolution · peptidomimetics · translational fidelity

[1] J. D. Bain, D. A. Wacker, E. E. Kuo, A. R. Chamberlin, *Tetrahedron* **1991**, *47*, 2389 – 2400.
[2] J. D. Bain, E. S. Diala, C. G. Glabe, D. A. Wacker, M. H. Lyttle, T. A. Dix, A. R. Chamberlin, *Biochemistry* **1991**, *30*, 5411 – 5421.
[3] J. A. Ellman, D. Mendel, P. G. Schultz, *Science* **1992**, *255*, 197 – 200.
[4] H. H. Chung, D. R. Benson, P. G. Schultz, *Science* **1993**, *259*, 806 – 809.
[5] A. Frankel, S. W. Millward, R. W. Roberts, *Chem. Biol.* **2003**, *10*, 1043 – 1050.
[6] Z. Tan, A. C. Forster, S. C. Blacklow, V. W. Cornish, *J. Am. Chem. Soc.* **2004**, *126*, 12752 – 12753.
[7] C. Merryman, R. Green, *Chem. Biol.* **2004**, *11*, 575 – 582.
[8] B. Zhang, Z. Tan, L. G. Dickson, M. N. L. Nalam, V. W. Cornish, A. C. Forster, *J. Am. Chem. Soc.* **2007**, *129*, 11316 – 11317.
[9] M. C. T. Hartman, K. Josephson, C.-W. Lin, J. W. Szostak, *PLoS One* **2007**, *2*, e972.
[10] T. Kawakami, H. Murakami, H. Suga, *Chem. Biol.* **2008**, *15*, 32 – 42.
[11] A. O. Subtelny, M. C. T. Hartman, J. W. Szostak, *J. Am. Chem. Soc.* **2008**, *130*, 6131 – 6136.
[12] M. Y. Pavlov, R. E. Watts, Z. Tan, V. W. Cornish, M. Ehrenberg, A. C. Forster, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 50 – 54.
[13] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, T. Ueda, *Nat. Biotechnol.* **2001**, *19*, 751 – 755.
[14] a) H. Dong, L. Nilsson, C. G. Kurland, *J. Mol. Biol.* **1996**, *260*, 649 – 663; Anticodon base modifications were obtained from: b) H. Inokuchi, F. Yamao in *tRNA: Structure, Biosynthesis, and Function* (Eds.: D. Söll, U. L. RajBhandary), American Society for Microbiology, Washington, DC, **1995**, pp. 17 – 30. If an anticodon sequence was present in Ref. [14a] but not the Ref. [14b], it was assumed to be unmodified. Conversely, anticodon sequences present in the Ref. [14b] but not in Ref. [14a] were disregarded.
[15] A. C. Forster, *Nucleic Acids Res.* **2009**, *37*, 3747 – 3755.
[16] E. B. Kramer, P. J. Farabaugh, *RNA* **2007**, *13*, 87 – 96.
[17] H. Murakami, A. Ohta, H. Ashigai, H. Suga, *Nat. Methods* **2006**, *3*, 357 – 359.
[18] J. F. Milligan, O. C. Uhlenbeck, *Methods Enzymol.* **1989**, *180*, 51 – 62.

