RSC Chemical Biology

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Chem. Biol.*, 2024, 5, 1111

Improved synthesis of the unnatural base NaM, and evaluation of its orthogonality in *in vitro* transcription and translation†

Anthony V. Le [ab] and Matthew C. T. Hartman [*ab]

Received 8th June 2024,
Accepted 4th September 2024

DOI: 10.1039/d4cb00121d

rsc.li/rsc-chembio

Unnatural base pairs (UBP) promise to diversify cellular function through expansion of the genetic code. Some of the most successful UBPs are the hydrophobic base pairs 5SICS:NaM and TPT3:NaM developed by Romesberg. Much of the research on these UBPs has emphasized strategies to enable their efficient replication, transcription and translation in living organisms. These experiments have achieved spectacular success in certain cases; however, the complexity of working *in vivo* places strong constraints on the types of experiments that can be done to optimize and improve the system. Testing UBPs *in vitro*, on the other hand, offers advantages including minimization of scale, the ability to precisely control the concentration of reagents, and simpler purification of products. Here we investigate the orthogonality of NaM-containing base pairs in transcription and translation, looking at background readthrough of NaM codons by the native machinery. We also describe an improved synthesis of NaM triphosphate (NaM-TP) and a new assay for testing the purity of UBP containing RNAs.

Introduction

Unnatural base pairs (UBP) have emerged as a promising strategy for expanding the genetic codes of organisms.[1–4] Early UBPs based on altering the donor–acceptor hydrogen bonding patterns of nucleobases[2,3,5–7] met with challenges with incomplete orthogonality due to the formation of tautomeric forms.[8] Benner has solved some of these issues through careful nucleobase modifications, developing two new base pairs (dS-dB, and dZ-dP).[9]

Kool, Hirao, and Romesberg developed new UBPs that rely solely on hydrophobic and pi–pi stacking interactions rather than H-bonding.[10–12] The most advanced of these UBPs contain the NaM-5SICS and NaM-TPT3 pairs. These pairs have been validated in replication, transcription, and translation in living *E. coli* and have been used for industrial preparation of proteins with non-canonical amino acids.[13–20]

Of the 96 possible codons containing a single NaM or TPT3 base, only a small subset has been tested. Certain base pairs are not efficiently retained in the plasmid DNA,[21] although this can be corrected using the CRISPR/Cas9 system.[4] Some rules have emerged for the codons tested, including preference for locating the UBP in the second position and the requirement for at least one G–C base pair in the other codon positions for efficient translation.[21]

Although the downstream integration of UBPs into organisms is an exciting goal for synthetic biology, the requirements of working *in vivo* also present many challenges that prevent fundamental experiments exploring the function of UBPs. *In vivo*, an unnatural base-containing tRNA or mRNA must be efficiently replicated, and transcribed before it is read by the ribosome. The precise concentrations of each system component is also difficult to control and optimize. *In vivo* studies may therefore miss important fundamental insights about the efficiency of UBPs and their true orthogonality.

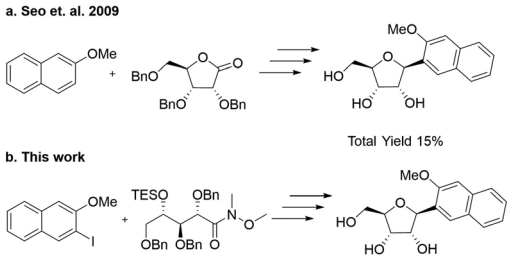
In this paper we focus on the efficiency and orthogonality of the NaM base in transcription and translation *in vitro*, learning new important insights about the functionality of this unnatural base. We also report a new, high-yielding synthetic pathway for the synthesis of NaM-ribose and its triphosphates.

Results and discussion

Romesberg's strategy for the synthesis of the NaM nucleoside involved first alkylation of a protected ribolactone, followed by deoxygenation and protecting group removal (Scheme 1a).[22] The overall reported yield of his approach was low, and in our hands, the initial alkylation step was difficult to control.

[a] *Virginia Commonwealth University, Department of Chemistry, 1001 W Main St. Richmond, VA 23284, USA. E-mail: mchartman@vcu.edu*
[b] *Virginia Commonwealth University, Massey Cancer Center, 401 College St. Richmond, VA 23219, USA*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4cb00121d

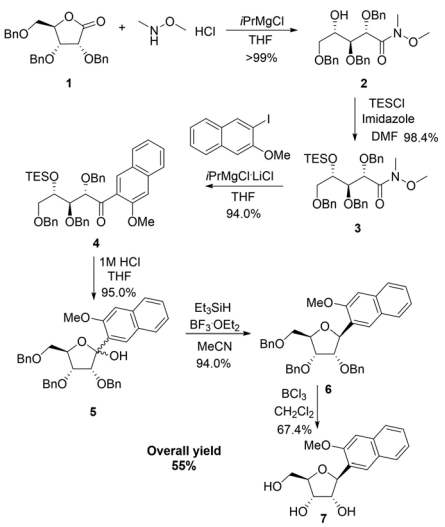


Scheme 1 (a) Original synthetic strategy involving direct alkylation of ribolactone. (b) Our new approach involving a Weinreb amide precursor.

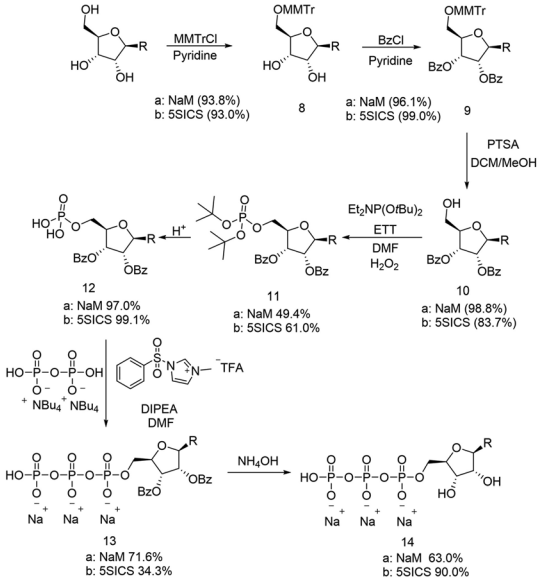
We therefore considered an alternative involving a Weinreb amide (Scheme 1b).[23] Our synthesis started by reacting commercially available ribolactone **1** with dimethylhydroxylamine giving Weinreb amide **2** (Scheme 2). We noticed that **2** is very sensitive to reversion to **1**, so we protected the C-4 OH. After some optimization (Table S1, ESI†), we found that the TES protecting group, as opposed to other silyl groups, gave the right balance of stability, reactivity, and removability for the subsequent alkylation and deprotection steps.

The subsequent aryl-ketone formation with 2-iodo-3-methoxy-napthalene[24] required some optimization (Table S2, ESI†), but the maximum yield of ketone **4** was achieved using turbo-Grignard conditions (Scheme 2). Recyclization was achieved by removal of the TES group under acidic conditions to form **5** as a mixture ratio of the linear and ring closed isomers 1.0 : 0.76 : 0.40. Removal of the hydroxyl group was achieved with Et_3SiH and $BF_3{\cdot}OEt_2$ leading to **6** as a mixture of inseparable isomers 7.1 : 1 (β : α). For the final debenzylation to give the NaM nucleoside **7**, we found that the milder Lewis Acid BCl_3[25] was superior to BBr_3. The overall yield for the 6 steps was 55% in Scheme 2.

In parallel, we also prepared the 5SICS nucleoside using Romesberg's reported method.[22] To form the NaM and 5SICS triphosphates, we found Romesberg's phosphorylation strategy gave multiple products, so we decided to selectively protect the 2′ and 3′ OH groups with benzoyl groups using a tritylation/



Scheme 2 Synthesis of NaM *via* Weinreb amide intermediate.



Scheme 3 Unnatural nucleotide triphosphorylation strategy.

benzoylation/detritylation sequence (Scheme 3). Each reaction proceeded in high yield for both the 5SICS and NaM base giving compounds **10a/b**. We reacted **10a/b** with di-*t*-butyl *N*,*N*-diisopropylphosphoramidite followed by oxidation to give the protected monophosphates **11a/b**. Deprotection under acidic conditions gave **12a/b**. We found that 5SICS was unstable in TFA, so we opted to use HCl in 1,4 dioxane for the deprotection instead. **12a/b** was then coupled with pyrophosphate using 1-methyl-3-benzenesulfonylimidazolium triflate to produce a clean unnatural base-TP product **13a/b**.[26] **13a/b** was then debenzoylated with NH_4OH to give the final NaM and 5SICS triphosphates **14a/b** (Scheme 3) as their sodium salts with a total yield of 19/15% starting from **7**/**20**. Although we generated lower yields than the reported single step phosphorylation method, this route avoids difficult separations of water-soluble phosphate isomers.[27]

With the 5SICS and NaM triphosphates in hand, we moved on to tests of their efficiency in *in vitro* transcription. Using phosphoramidite chemistry, we first prepared 16 ssDNAs for transcription *via* T7 RNA polymerase (Fig. 1a). Each contained a single NYN codon (Y = 5SICS) for transcription into mRNAs containing NaM. We included a native template containing the CAG leucine codon as a control. We first initiated our studies by performing *in vitro* transcriptions in the absence of NaM-TP and were surprised to find that a full-length product was formed for all templates, even though the expected truncated products were also observed (Fig. S1, ESI†). We then titrated in NaM-TP and found that the percentage of full-length products increased with increasing NaM-TP concentration and saturated above 1.25 mM of NaM-TP (Fig. 1b and c). We then performed *in vitro* transcription reactions for each of the 16 NYN templates with and without 2.5 mM NaM-TP (Fig. S2, ESI†) and computed the improvements in full-length bands with each template (Fig. 1d). In every template we observed

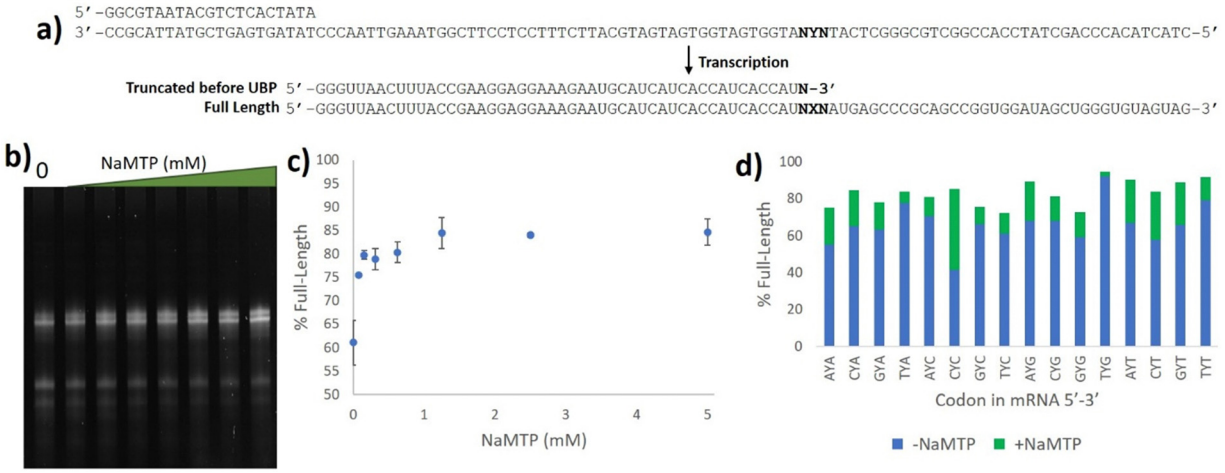


Fig. 1 Optimizing NaM incorporation in T7 transcription with d5SICS. (a) ssDNAs containing all possible NYN codons where Y = d5SICS and X = NaM. The T7 promoter primer and the expected products from the transcription reaction are shown. (b) 10% Urea PAGE gel containing *in vitro* transcription products with increasing concentrations of NaM-TP (0–5 mM) with a ssDNA template containing the CYT codon. Bands were visualized by SYBR-green II (c) Plot showing the % full-length calculated by measuring the relative intensity of the full-length bands *vs.* truncated bands in each lane in (b). (d) % Full-length stacked bar graph for transcriptions with all 16 ssDNA templates with and without NaM-TP computed from gel image Fig. S2 (ESI†). The ssDNA template codons are shown in Table S3 (ESI†).

significant readthrough of the d5SICS base in the absence of NaM-TP, although improvements in the ratio of full-length to truncated bands were observed for every template in the presence of NaM-TP at 2.5 mM thus confirming at least partial NaM-TP incorporation.

Although we were able to optimize our NaM-TP incorporation efficiency, it was unclear how pure our RNA products were since in all cases, we observed a full-length band in the absence of NaM-TP. Several strategies have been employed to assess UBP incorporation efficiency. Hirao created a biotinylated version of his *y*-base UBP and used a gel shift assay to monitor incorporation.[28] Romesberg envisioned a similar approach that utilized an amine modified MMO2 which was biotinylated post-transcriptionally and then analyzed *via* gel.[29] Both approaches require a separate synthesis to deliver the desired product and assume that the biotinylated UBP will incorporate with the same efficiency as the unmodified UBP. Another common strategy to investigate incorporation efficiency involves radiolabeling the NTP preceding the UBP during transcription, followed by 2D-TLC after complete RNase I digestion to the 3′ monophosphates to analyze the monomer distribution.[30,31] While powerful, this assay is typically performed with short model templates, to prevent the UBP signal from being overshadowed by the canonical bases. With longer templates slight misincorporations can easily fall within the measurement error.

We used two approaches to monitor the purity of our RNAs. First, we initiated a series of reverse transcriptions using superscript II (SSII)[19,32,33] (Fig. 2a) on full-length, gel-purified RNAs prepared with increasing concentrations of NaM-TP. We omitted any deoxy UBP triphosphates in these reactions and therefore expected that SSII would terminate at the NaM base if present. RT reactions were performed with a fluorescein-labeled primer for detection (Fig. 2a). Surprisingly, we did not observe any truncation products from reverse transcription. This is a different result compared to that reported by Eggert *et al.* who observed 23% truncation when reverse transcribing NaM-containing RNAs with SSII in the absence of any complementary deoxy unnatural NTP.[32] Notably, reverse transcriptions of RNAs prepared in the presence of NaM-TP produced a higher band than those transcribed in the absence of NaM-TP, and this band increased as more NaM-TP was added (Fig. 2b). Above 2.5 mM we observed almost exclusively this higher band (Fig. 2b and c). The band transcribed in the absence of NaM-TP was the same length as the band produced by a control template containing CUG in place of the NaM (Fig. S3, ESI†), thus this higher band is unique to RNAs containing NaM. This band was present in all 16 templates containing NaM (Fig. S3, ESI†) and its intensity ranged from 65–95% depending on the template (Fig. 2d).

To get a better idea of the identity of this higher band, we performed next-generation sequencing of the reverse transcribed products after PCR amplification (with only the standard dNTPs) of the template containing the GXA codon. Surprisingly, there was no indication that these longer sequences were present in the PCR products, although we could detect A, C, G, and T incorporation or a deletion in the PCR products (Fig. S4, ESI†) in the position where we would expect NAM to be. We therefore suspect that this higher band is a product that cannot be amplified from the cDNA by PCR using our primers.

To further confirm the purity of our mRNAs, we digested a subset of our transcripts using RNase A and looked for the incorporation of NaM in the expected fragments. RNase A cleaves RNAs at the 3′ side of C or U, and so we chose transcripts that were predicted to produce NaM-containing fragments within a unique mass range (the 5 nt region) (Fig. 2e). We compared the RNAs transcribed in the presence and absence of NaM-TP. The transcriptions performed in the

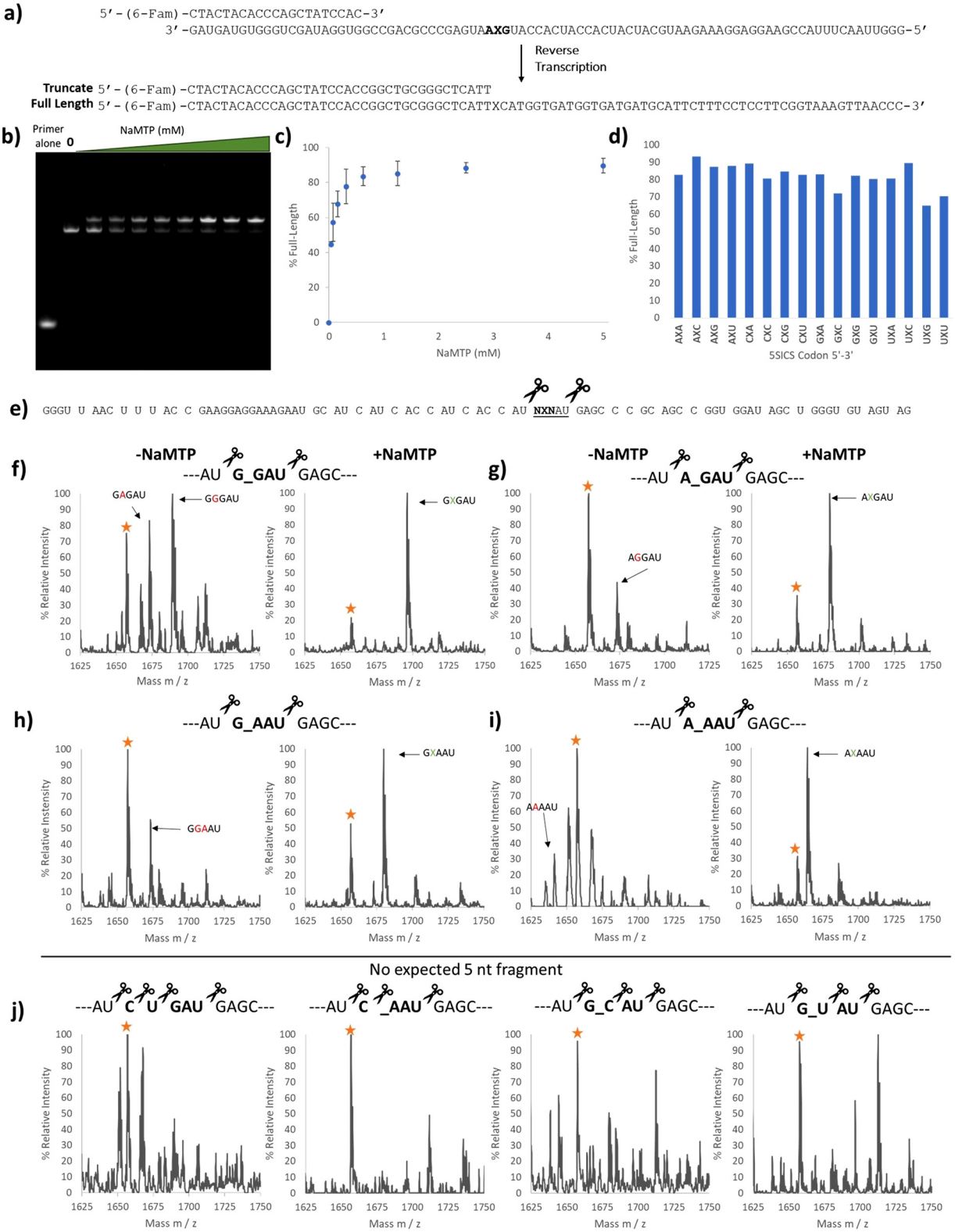


Fig. 2 Purity validation of transcription products containing NaM. (a) Reverse transcription template containing NaM at the second position. The primer was labeled with 6-(FAM) and the expected full-length and truncated products are shown. (b) 10% Urea PAGE image with each lane containing SSII reverse transcription products using UBP mRNA containing the GXA codon prepared with increasing amounts of NaM-TP (0–5 mM) (Fig. 1b). (c) Plot showing the percentage of the higher band in each lane in (b). (d) Bar graph showing the percentage of the higher band for all 16 reverse transcribed templates. The RNA templates for the RT reactions were prepared with 2.5 mM NaM-TP. (e) Predicted RNase A digest products of the NaM-containing RNAs. RNase A cleaves on the 3′ end of C and U, so templates containing neither C nor U in the codon region would give a unique 5 nucleotide (nt) fragment. (f) MALDI-TOF MS of the RNase A digest products in the 5 nt region for the template encoding the GXG RNA transcribed without (left) and with

(right) NaM-TP. Yellow star labeled as a background peak (g) MALDI-TOF MS of the RNase A digest products in the 5 nt region for the template encoding the AXG RNA transcribed with (Right) and without (Left) NaM-TP. Yellow star labeled as a background peak or potential AAGAU fragment. (h) MALDI-TOF MS of the RNase A digest products in the 5 nt region for the template encoding the GXA RNA transcribed with (Right) and without (Left) NaM-TP. Yellow star labeled as a background peak or potential GAAAU fragment. (i) MALDI-TOF MS of the RNase A digest products in the 5 nt region for the template encoding the AXA RNA transcribed with (Right) and without (Left) NaM-TP. Yellow star labeled as a background peak or potential AGAAU fragment. (j) MALDI-TOF MS of the RNase A digest products in the 5 nt region for the template mRNA # 01 CUG (Left), mRNA # 06 CXA (middle-left), mRNA # 11 GXC (middle-right) and mRNA # 13 GXC (right). Yellow star labeled as a background peak. Full mass spectrums are shown in Fig. S5–S16 (ESI†) and the expected and observed masses are summarized in Table S5 (ESI†). mRNA templates are shown in Table S4 (ESI†).

absence of NaM-TP gave several peaks. In line with the sequencing results, we were able to identify G and A misincorporation (Fig. 2f–i left and Fig. S5–S16, ESI†). (We note here that C and U misincorporation would be digested by RNase A into smaller fragments indistinguishable from other fragments). In contrast, those mRNAs transcribed in the presence of NaM-TP (Fig. 2f–i, right, Fig. S5–S16, ESI†) all showed a major peak corresponding to NaM incorporation as well as an additional peak (labeled with a star). While this peak could correspond to a natural base misincorporation (G or A depending on the template), it was also present in four templates that would not be expected to give any fragments in this range: one natural control template and 3 UBP templates (Fig. 2j). We therefore suspect that this peak comes from another source (*e.g.* incomplete digestion). Taken together, these studies show that *in vitro* transcription using 2.5 mM NaM-TP delivers NaM-containing RNAs with minimal misincorporation of natural nucleotides, aligning our work with literature reports.[27,34]

With confidence in our NaM-mRNA purity, we moved to testing our NaM-mRNAs in *in vitro* translation. We first investigated the potential consequences of translating our mRNAs in the absence of any unnatural base-containing tRNAs. We wondered if the ribosome would be able to read through these codons, and if so, if there were any trends. We tested both wild-type (wt) and hyperaccurate (mS12) ribosomes, known to more readily reject near-cognate tRNAs.[35,36] We used the recently developed affinity-clamp fluorescent translation assay for these studies in which each template has a C-terminal affinity-clamp sequence (encoding PQPVDSWV).[37] In this assay peptides containing the C-terminal tag will bind to the cyPET-yPET-affinity clamp protein causing a decrease in the ratio of 527/475 fluorescence emission which correlates to yield (Fig. S17–S20, ESI†). Yields are shown in Fig. 3b alongside a control template with a native leucine codon (CUG). For most of the codons, very low yields were observed, further confirming the purity of our mRNA products, especially when observing the mS12 results. With wt ribosomes, there were exceptions. The CXA and GXC codons both showed >40% misdecoding yield; AXC and GXU showed misreading between 20 and 40%. For mS12 ribosomes the codons with appreciable yield tended to be the same, but misreading yields in all cases were significantly suppressed; no codons gave yields above 20% (Fig. 3b).

We also performed Ni-NTA capture of the translated peptides using the N-terminal His-tag and used MALDI to detect any full-length peptides formed (Fig. S21–S54, ESI†). Based on the amino acid incorporated, we inferred how the ribosome was able to interpret the UBP codon. These results are shown as heatmaps in Fig. 3c. Yields in the heatmaps are relative to the template containing a CUG leucine codon and a control translation reaction lacking mRNA.

There are global trends revealed by this dataset. First, most of the codons that are misread at detectable levels have at least one G or C (Fig. 3b) with AXU as an exception. This is in alignment with Romesberg's finding that successful UBP pairs contain at least 1 G–C pair. Second, there doesn't seem to be a nucleotide bias in the misdecoding. Both the wild-type and mS12 ribosomes were able to misread the NaM base as either U, C, A or G depending on the context. Third, the hyperaccurate mS12 ribosomes were more efficient at discriminating against misdecoding by native AA-tRNAs than wild-type ribosomes. This is evidenced by the lower yields shown in Fig. 3b, many of which were at background levels. The difference in yields between hyperaccurate and wild-type ribosomes is further evidence of the purity of our NaM-containing mRNAs as templates containing standard nucleotides in place of NaM would have been expected to give the same results for both types of ribosomes.[35] Finally, it is interesting to compare our results with Katoh and Suga's recent paper that investigated background readthrough with standard codons.[38] Suga found that NUN and NCN codons are more easily mis-decoded that NAN and NGN codons. Our data shows that NXN codons overall appear to be misread at low frequencies in comparison with the native codons, which speaks to their relatively high level of orthogonality.

The CXA codon is notable for its particularly high background yield, due to misdecoding as glutamine. It is interesting that the $tRNA^{Gln}$ that is the primary decoder of the CAA codon has a unique modified base, 5-carboxymethylaminomethyl-2-thiouridine, ($cmnm^5s^2U$) base in its anticodon.[39,40] This particular base is not found in any other tRNAs in *E. coli* (although there are close analogs).[40] The $cmnm^5s^2U$ base is known to improve the geometry of the pair with A, preventing frameshifting.[40] Perhaps this enhanced pairing also enables the ribosome to overcome the mismatched NaM:U base pair in the second codon position. It is also notable that next most readthrough codons (GXC, GXU, and AXC) in our data set were shown to have little background readthrough by Romesberg.[21] This discrepancy suggests that the surrounding sequence context could influence UBP codon readthrough as it does for stop codon suppression.[41] The fact that these same 3 codons are among the most efficient codons for *in vivo* incorporation with ncAAs is also intriguing and highlights that: (1) the ribosome has an innate ability to recognize these codons and (2) background readthrough can be overcome in the presence of a complementary unnatural base-containing tRNA.

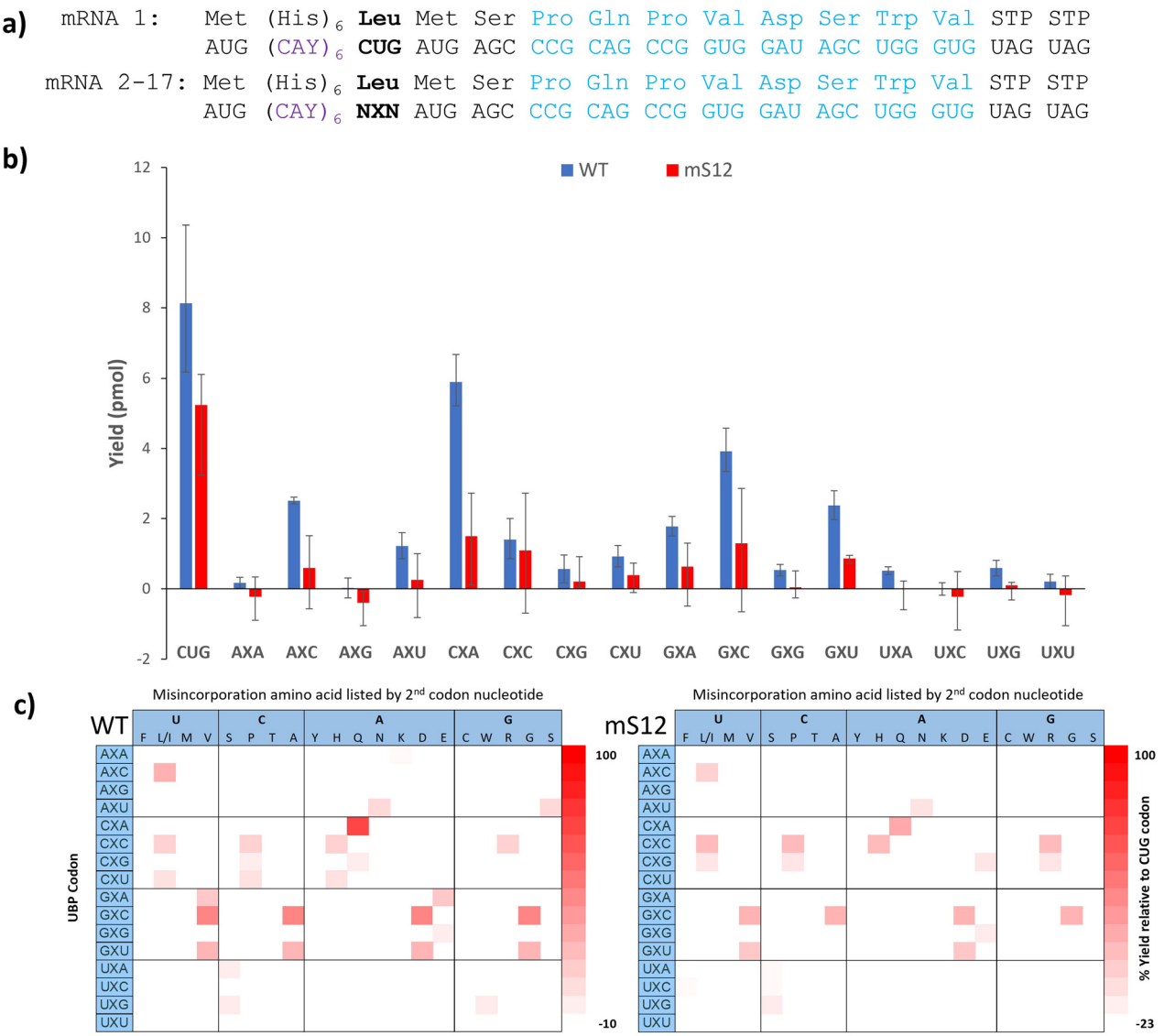


Fig. 3 Investigating the orthogonality of NaM codons. Translations were performed using wt or hyperaccurate (mS12) ribosomes. (a) Templates tested in translation. The C-terminus of each template encodes an affinity clamp peptide used to monitor yield. (b) *In vitro* translation yields in pmol calculated after 90 min at 37 °C. Bar graphs colored in blue are yields generated from using wt ribosomes and bar graphs colored in red are yields generated from using mS12 Ribosomes. The codon in the mRNA is shown below each data point. (c) Misdecoding events represented by a heat map. The yields were measured relative to the positive control mRNA containing the CUG codon and are shown in b. Each colored box in the heat map represents a detectable peak after His-tag purification.

Finally, it is also interesting that some codons were read by multiple tRNAs. The GXC codon is prominent in this regard as we observed 4 different full length peptide products. This means that the ribosome was able to accept at least 4 different tRNAs in response to this codon (albeit at moderate to low yield).

We chose three NaM codons that have been shown by Romesberg to be highly efficient in translation, AXC, GXU, and GXC for further analysis.[21] *In vivo* work with these codons has utilized the pyrrolysyl tRNA ($tRNA^{Pyl}$); here we chose to use *E. coli* $tRNA^{Leu1}$ which natively has the CAG anticodon and reads the CUG codon.[37,42,43] *E. coli* LeuRS does not recognize the anticodon of the tRNA,[44] making it amenable to substitution with an unnatural nucleotide. Using templates containing NaM (Table S4, ESI†), we created tRNAs containing the GYU, AYC, and GYC anticodons using 2.5 mM 5SICS triphosphate during transcription. All 3 tRNAs were able to be charged with Leucine and were detectable *via* MALDI-MS after undergoing the reductive amination/nuclease digestion assay (Fig. S55–S57, ESI†).[45] AXC tRNA/GXU only showed misdecoding by native tRNAs. The GYU tRNA/AXC codon pairing gave some of the correct mass peak; however, it was impossible to distinguish this peak from the background isoleucine misdecoding

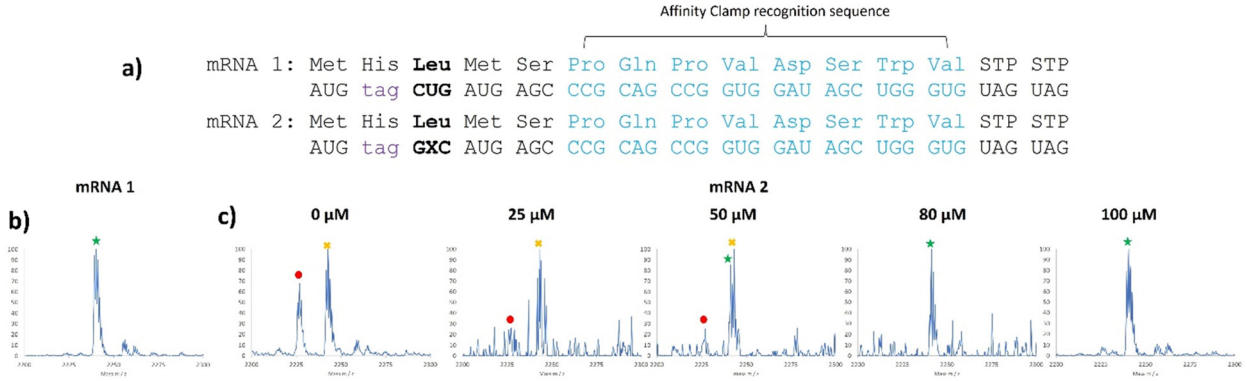


Fig. 4 5SICS-tRNA$_{GYC}$ is able to read the GXC codon *in vitro*. (a) mRNAs used for testing of translation yield and their corresponding expected peptide sequences, where GXC codes for Leucine. (b) MALDI-MS spectrum of the translation products using mRNA **1** showing the expected peak corresponding to leucine incorporation. (c) MALDI-MS spectrum of the translation products using mRNA **2** at increasing concentrations of 5SICS-tRNA$_{GYC}$. Peptide sequences: MH$_6$V̲MSPQPVDSWV labeled as a red circle, MH$_6$D̲MSPQPVDSWV labeled as a yellow "X" and the desired product, MH$_6$L̲MSPQPVDSWV, labeled as a green star. Masses for the labeled peaks are shown in Table S6 (ESI†).

(Fig. S58, compare with Fig. S25, ESI†) and so this codon was not pursued further. The GYC tRNA/GXC codon pairing showed the expected peptide product containing leucine as well as other misdecoding peaks when we tested at a concentration of 40 μM (Fig. S58–S60, ESI†).

We then titrated this tRNA into the translation mixture and observed clean MALDI-MS spectra at 100 μM (Fig. 4c), comparable to the native CUG codon experiment (Fig. 4b), highlighting the good orthogonality with this codon and that background readthrough can be overcome with sufficient complementary tRNA.

Summary/conclusion

Here we report an improved synthesis of NaM-TP; we used the flexibility of *in vitro* transcription and translation systems to test its tolerance with the translation system. To monitor the purity of UBP products from transcription reactions we used a combination of reverse transcription and a newly developed RNase A assay. We then tested the ability of the translation apparatus to readthrough sixteen codons containing NaM (NXN) in the absence of complementary tRNAs. Codons CXA, AXC, GXC, GXU yielded >20% yield of mis-decoding using wt ribosomes; however, mS12 hyperaccurate ribosomes showed diminished misdecoding. Using MALDI-MS we were also able to infer how the ribosomes interpreted the NaM base during translation. Finally, we tested the fidelity and orthogonality of three UBP codon-anticodon in translation and found that mRNA$_{GXC}$-tRNA$_{GYC}$ delivered high orthogonality at 100 μM of 5SICS-tRNA$_{GYC}$. Taken together, our work shows that combining reconstituted *in vitro* translation and UBPs can lead to rapid assessment of UBP fidelity in transcription and translation. Our data shows that investigators performing UBP experiments with certain codons (in particular CXA and GXC) should be proceed with caution, realizing that they will likely be competing with background readthrough. It also points to the possibility of using hyperaccurate ribosomes as a strategy to minimize background readthrough of NaM codons. Down the line, it should also be possible to use this system alongside new UBP DNA synthesis strategies[46] to quickly design and test the effects of engineered components for improved fidelity. A list of rigorously validated UBP codons and strategies maximize their incorporation will enable new advances in synthetic biology and the generation of peptide libraries *via* display technologies.

Materials and methods

General

Salts and solvents. Reagents for synthesis were purchased from Fisher Scientific, Sigma-Aldirch, TCI, VWR, and Acros Organics. Dowex Cation Exchange Resin was purchased from Acros Organics. C_{18} Resin was purchased from Sigma Aldrich. UBP phosphoramidites 5SICS & NaM were purchased from Berry & Associates. UBP ssDNA oligos and primer were synthesized from Keck Lab. Salts and NTPs (ATP, GTP, CTP, and UTP) were purchased from Fisher Scientific, Sigma-Aldrich, and Thermo-Fisher. A mix of buffered phenol/$CHCl_3$/isoamylalcohol was purchased from Acros Organics. Enzymes were purchased from Thermo Fischer or NEB. Affinity clamp protein was prepared using literature reference.[37] Natural amino acids were purchased from Fluka and dissolved in water to 10 mM, pH was adjusted to 7.2–7.6 with KOH and sterile filtered (0.22 μm) for subsequent storage at −20 °C.

Instrumentation. DNA, RNA, Optical density, and protein concentrations were quantified through UV-vis NanoDrop ND-1000 Spectrophotometer. pH was measured with a Mettler Toledo SevenEasy S20 pH Meter. Mass spectrometry was carried out through Voyager DE-Pro MALDI-TOF, and samples were prepared with α-cyano-4-hydroxycinnamic acid (CHCA) purchased through Millipore-Sigma (C2020-10G) or α-cyano-4-chlorocyanocinnamic acid (ClCA) prepared using literature protocol.[47] High-resolution mass spectrometry was done through PerkinElmer AxION 2 TOF in positive mode using an LC column; samples were injected into the column *via* a syringe pump. Affinity clamp assays were run using Agilent BioTek

Synergy H1 Microplate Reader H1M. Benchtop centrifuges are Sorvall legend micro 17R refrigerated and for larger volumes, Sorvall Superspeed RC2-B was used.

Synthesis of NaM and 5SICS triphosphate. Experimental details and compound characterization is shown in the ESI.† The 5SICS base was synthesized using the literature protocol[22] with some optimization shown on ESI† pages 12–13.

RNA UREA PAGE purification and analysis. 10% polyacrylamide gel was cast and polymerized in a 10 × 10 cm plate with 1.0 mm spacing or 20 × 20 cm with 2.0 mm spacing. Urea was added to the *in vitro* transcription reaction to a final concentration of 5 M. The RNA samples were then loaded onto the gel and were eluted for 1–2 h at 150 V for a 10 × 10 cm plate and 25 W for a 20 × 20 cm plate.

The RNAs were eluted by crushing and soaking in 2 mL dH_2O at 80 °C for 30 min. The solution was then filtered through a membrane syringe filter and then extracted with butanol to a final volume of 100–500 μL of dH_2O. The water layer was then separated and the RNA product was precipitated with 0.1 volumes of 3 M NaOAc (pH = 5.5) and 3 volumes of ethanol. The solution was then chilled at −20 °C for 30 min and then centrifuged. The precipitate was then filtered and washed with 70% Ethanol followed by ethanol to obtain the final RNA product.

The 10 × 10 gels were stained using SYBR green 1× (Thermo Fisher Scientific) in 1× TBE solution for 30 min. The gels were then imaged using the Bio-Rad ChemiDoc MP imaging system.

T7 NaM-containing mRNA Transcription. *In vitro* transcription reactions (500 μL) contained final concentrations of 40 mM Tris (pH 7.8) 0.1% Triton, Spermidine (2.5 mM), DTT (10 mM), NTPs (5 mM each), GMP (4 mM), RiboSafe RNase inhibitor (0.2 U μL^{-1} Bioline Cat#: C755H60), inorganic pyrophosphatase (1 μ mL^{-1}), T7 forward primer (1 μM 5′-GGCGTAATACGACTCACTATA-3′), ssDNA template containing a NYN codon (0.125 mM), NaM-TP (2.5 mM) and T7 RNA Polymerase (0.2 mM). The solution was then incubated at 37 °C overnight. The reaction was quenched by adding urea (0.3 mg μL^{-1}) and the solution was loaded on a 20 cm × 20 cm 10% UREA page gel.

T7 5SICS-containing tRNA transcription. *In vitro* transcription reactions (500 μL) contained final concentrations of 40 mM Tris (pH 7.8) 0.1% Triton, Spermidine (2.5 mM), DTT (10 mM), NTPs (7 mM each), GMP (4 mM), RiboSafe RNase inhibitor (0.2 U μL^{-1} Bioline Cat#: C755H60), inorganic pyrophosphatase (1 μ mL^{-1}), T7 forward primer (1 μM GGCGTAATACGACTCACTATA), ssDNA template containing a NYN codon (0.125 μM), 5SICS-TP (2.5 mM) and T7 RNA Polymerase (0.2 mM). The solution was then incubated at 37 °C overnight. The reaction was quenched by adding urea (0.3 mg μL^{-1}) and the solution was loaded on a 20 cm × 20 cm 10% UREA page gel.

NaM-containing mRNA reverse transcription. Reverse transcription reactions (10 μL) contained final concentrations of First-strand reaction buffer (1×)(Thermo-Fisher Scientific Cat#: 18064022), DTT (10 mM), dNTPs (0.5 mM each), RNAse inhibitor (0.4 U μL^{-1}), mRNA 6-fam-primer (0.1 μM) (6-fam)-CTACTACACCCAGCTATCCAC) and mRNA NXN (1 μM). The solution was then incubated at 95 °C for 1 min and cooled to 25 °C. Then Superscript II 10 U μL^{-1} was added and mixed *via* pipetting. The reaction was then incubated at 37 °C for 4 h and then quenched with heat at 50 °C for 30 min.

Sequencing analysis of RT products. (100 μL) contained final concentrations of 1X Q5 reaction buffer, dNTPs (0.2 mM), Forward Primer (200 nM 5′-GGCGTAATACGACTCACTATAGGG-TTAACTTTACCGAAGGAGGAAAGA-3′), Reverse Primer (200 nM; 5′–CTACTACACCCAGCTATCCACCGGCTGCGGGCTCAT–3′), ssDNA from reverse transcription reaction (1 nM) and Q5 polymerase (1 μL). PCR was then carried out on a DNA Engine thermocycler (BioRad), beginning with an initial 1 minute denature step at 95 °C, followed by twenty-two cycles of the following: 98 °C for 20 seconds, oligo T_m for 40 seconds, 72 °C for 40 seconds. PCR products were submitted to Azenta and sequenced using the Amplicon-EZ format. The sequencing was analyzed using the Aptasuite software,[48] and 130 882 reads were analyzed. 59.3% of the reads consisted of the 5 outcomes described in Fig. S4 (ESI†).

RNase digestion of RNAs. In a 1.5 mL autoclaved Eppendorf tube, the components were added in the following order: 50 mM 3-hydroxypicolinic acid, mRNA (NXN) (30 pmol), and 20 U RNase A in a final volume of 20 μL.[49] The reaction was mixed *via* pipetting and then incubated at 37 °C for 1 h. 2 μL of 1 M HCl was then added with mixing followed by concentration on a Speed-Vac 37.7 °C for 30 min. The solid was then dissolved in 5 μL of a MALDI Matrix solution composed of 3-hydroxy picolinic acid : picolinic acid : citric acid diammonium (9 : 1 : 1) dissolved in MeCN : H_2O (7 : 3).

Charging of leucine onto $tRNA^{Leu}$. A solution mixture of HEPES/KOH (30 mM pH = 7.4), transcribed 5SICS-$tRNA^{Leu}$ NYN (50–400 μM), $MgCl_2$ (15 mM), KCl (25 mM), ATP (6 mM), inorganic pyrophosphatase (0.001 mg mL^{-1}) (Sigma-Aldrich Cat#: I5907-1 mg), Leucine (400 μM) LeuRS (0.3–0.51 μM) and bovine serum albumin (BSA, 0.24 mg mL^{-1}) previously dialyzed into deionized water (dH_2O), in a final volume of 50 μL. The reaction was incubated for 60 min at 37 °C before quenching by adding 0.1 volumes of aq. NaOAc (3 M pH = 5.2). The RNA-containing aqueous phase was extracted by mixing it with a mixture of phenol/$CHCl_3$/isoamyl alcohol (25/24/1) (unbuffered) and subsequently vortexed and centrifuged to separate layers. Then aqueous layer was removed and transferred to a clean vial mixed with an equal volume of $CHCl_3$ and vortexed and centrifuged again. The aqueous layer was removed into a clean vial and mixed with 3 volumes of ice-cold ethanol. The vial was incubated at −20 °C for 20 min and centrifuged for 20 min at 4 °C at 17 000 × *g*. The pellet was washed with 500 μL of 70% Ethanol and then 500 μL of 100% ethanol to remove salts. Finally, the pellet was left to air dry and resuspended in NaOAc (12.5 μL, 100 mM pH = 5.0). Reductive amination was carried out by mixing 6.25 μL of the previous tRNA-aa preparation with 3.75 μL of dH_2O, (4-formylphenoxypropyl) triphenylphosphonium bromide in MeOH (12.5 μL, 63 mM) and fresh $NaBH_3CN$ dissolved in 50 mM NaOAc pH = 5.0 (2.5 μL, 200 mM). The reaction was incubated at 37 °C on a tumbler for 2 hours before quenching with 0.1 volume of NH_4OAc

(4.4 M pH = 5.0). The reaction product was recovered through ethanol precipitation and the resulting pellet was resuspended in NH_4OAc (2.25 μL, 200 mM pH = 5.0). 0.25 μL of Nuclease P1 (1 U μL^{-1} in 200 mM NH_4OAc pH = 5.0) (Wako Cat#: 145-08221) was added and incubated at rt for 20 min. After incubation, the reaction mixture was quenched on ice, and 1 μL was mixed with 9 μL of MALDI matrix α-cyano-4-hydroxycinnamic acid (CHCA) 10 mg mL^{-1} in MeCN:1% TFA (1 : 1) and spotted onto the MALDI plate for further analysis.

***In vitro* translation background readthrough.** Experiments were carried out in the presence of 5SICS-$tRNA_{NYN}$ and the absence of 5SICS-$tRNA_{NYN}$. In a 15 μL microplate well, the components were added in the following order: dH_2O, TS-solution HEPES-KOH pH = 7.6 (50 mM), AcOK (100 mM), $MgOAc_2$ (6 mM), spermidine (10 mM), dithiothreitol (1 mM), creatine phosphate (20 mM), ATP/GTP (1.5 mM each, potassium exchanged), total *E. coli* tRNA (100 mg mL^{-1}, deacylated and dialyzed overnight against 50 mM tris/HCl pH = 9, precipitated and resuspended in dH_2O), factor mix (EF-G (0.52 μM), IF-1 (2.7 μM), IF-2 (0.4 μM), IF-3 (1.5 μM), MTF (0.6 μM), RF-1 (0.3 μM), RF-3 (0.17 μM), RRF (0.5 μM), EF-Tu (10 μM), EF-Ts (4.1 μM), Ribosomes (1.2 μM), affinity Clamp (200 nmol) amino acids required for the template (12.5–100 μM each), 20 aminoacyl-tRNA synthetases (0.03–1.23 μM), inorganic pyrophosphatase (1 μg mL^{-1}), creatine kinase (4 μg mL^{-1}), myokinase from rabbit (3 μg mL^{-1}) (Sigma-Aldrich Cat#: M3003-2.5KU), nucleoside 5′-diphosphate Kinase (1.1 μg mL^{-1}) (Sigma-Aldrich Cat#: N2635-100 UN). To the resulting mix, methionine (10 μM) was added followed by the desired mRNA template (3 μM). For negative controls, the mRNA template was left out. The reaction mixture was incubated for 1.5 hours at 37 °C in Agilent Biotek Synergy H1 microplate reader H1MF and quenched with wash buffer (TBS buffer, 10 mM BME). Then, the reaction was mixed with 10 μL of Ni-NTA resin (MCLabs Cat#: NINTA-500) and left for binding for 1 h with tumbling at rt. After binding, the supernatant was removed by centrifugation (1 min, rt, 7800 × *g*) and washed 3 × 50 μL with wash buffer (TBS buffer, 10 mM BME). Elution was done by incubating the beads for 15 min with 50 μL of trifluoroacetic acid 1% in dH_2O at rt and subsequent centrifugation (1 min, rt, 7800 × *g*). A 200 μL pipette tip packed with a C_{18} resin, was washed with 15 μL MeCN, 15 μL MeCN:0.2% TFA (1 : 1) and 15 μL 0.2% TFA, centrifuging each time for 1 min at 1000 × *g*. The peptide eluted in TFA was loaded onto the tip. Each tip was washed 2× with 15 μL 0.2% TFA. The peptide was eluted with 4-chloro-α-cyanocinnamic acid (6.2 mg mL^{-1} in MeCN:0.2% TFA (7 : 3)), spotted in the MALDI-TOF plate and analyzed.

***In vitro* translation with complementary precharged 5SICS-$tRNA^{Leu}$.** *In vitro* translation experiments were performed as above, except for the addition of the 5SICS-$tRNA^{Leu}$ NYN at concentrations of 25–100 μM.

Data availability

The data supporting this article have been included as part of the ESI.†

Conflicts of interest

Authors declare no conflicts of interest.

Acknowledgements

The authors thank the NIH (R03CA259876 and R01GM143396) for funding of this work.

References

1 S. A. Benner and A. M. Sismour, Synthetic Biology, *Nat. Rev. Genet.*, 2005, **6**(7), 533–543.
2 I. Hirao, T. Mitsui, M. Kimoto and S. Yokoyama, An Efficient Unnatural Base Pair for PCR Amplification, *J. Am. Chem. Soc.*, 2007, **129**(50), 15549–15555.
3 J. D. Bain, C. Switzer, R. Chamberlin and S. A. Benner, Ribosome-Mediated Incorporation of a Non-Standard Amino Acid into a Peptide through Expansion of the Genetic Code, *Nature*, 1992, **356**(6369), 537–539.
4 Y. Zhang, B. M. Lamb, A. W. Feldman, A. X. Zhou, T. Lavergne, L. Li and F. E. Romesberg, A Semisynthetic Organism Engineered for the Stable Expansion of the Genetic Alphabet, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**(6), 1317–1322.
5 M. Ziemniak, J. Kowalska, M. Lukaszewicz, J. Zuberek, K. Wnek, E. Darzynkiewicz and J. Jemielity, Phosphate-Modified Analogues of M7GTP and M7Gppppm7G - Synthesis and Biochemical Properties, *Bioorg. Med. Chem.*, 2015, **23**(17), 5369–5381.
6 J. Horlacher, M. Hottiger, V. N. Podust, U. Hübscher and S. A. Benner, Recognition by Viral and Cellular DNA Polymerases of Nucleosides Bearing Bases with Nonstandard Hydrogen Bonding Patterns, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**(14), 6329–6333.
7 S. Moran, R. X.-F. Ren and E. T. Kool, A Thymidine Triphosphate Shape Analog Lacking Watson–Crick Pairing Ability Is Replicated with High Sequence Selectivity, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**(20), 10506–10511.
8 C. R. Geyer, T. R. Battersby and S. A. Benner, Nucleobase Pairing in Expanded Watson-Crick-like Genetic Information Systems, *Structure*, 2003, **11**(12), 1485–1498.
9 S. Hoshika, N. A. Leal, M.-J. Kim, M.-S. Kim, N. B. Karalkar, H.-J. Kim, A. M. Bates, N. E. Watkins, H. A. SantaLucia, A. J. Meyer, S. DasGupta, J. A. Piccirilli, A. D. Ellington, J. SantaLucia, M. M. Georgiadis and S. A. Benner, Hachimoji DNA and RNA: A Genetic System with Eight Building Blocks, *Science*, 2019, **363**(6429), 884–887.
10 A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari and F. E. Romesberg, Discovery, Characterization, and Optimization of an Unnatural Base Pair for Expansion of the Genetic Alphabet, *J. Am. Chem. Soc.*, 2008, **130**(7), 2336–2343.
11 Y. Wu, A. K. Ogawa, M. Berger, D. L. McMinn, P. G. Schultz and F. E. Romesberg, Efforts toward Expansion of the

Genetic Alphabet: Optimization of Interbase Hydrophobic Interactions, *J. Am. Chem. Soc.*, 2000, **122**(32), 7621–7632.

12 K. M. Guckian, T. R. Krugh and E. T. Kool, Solution Structure of a DNA Duplex Containing a Replicable Difluorotoluene–Adenine Pair, *Nat. Struct. Biol.*, 1998, **5**(11), 954–959.

13 J. Oh, J. Shin, I. C. Unarta, W. Wang, A. W. Feldman, R. J. Karadeema, L. Xu, J. Xu, J. Chong, R. Krishnamurthy, X. Huang, F. E. Romesberg and D. Wang, Transcriptional Processing of an Unnatural Base Pair by Eukaryotic RNA Polymerase II, *Nat. Chem. Biol.*, 2021, **17**(8), 906–914.

14 D. A. Malyshev, K. Dhami, T. Lavergne, T. Chen, N. Dai, J. M. Foster, I. R. Corrêa and F. E. Romesberg, A Semi-Synthetic Organism with an Expanded Genetic Alphabet, *Nature*, 2014, **509**(7500), 385–388.

15 Y. J. Seo, D. A. Malyshev, T. Lavergne, P. Ordoukhanian and F. E. Romesberg, Site-Specific Labeling of DNA and RNA Using an Efficiently Replicated and Transcribed Class of Unnatural Base Pairs, *J. Am. Chem. Soc.*, 2011, **133**(49), 19878–19888.

16 E. L. Tae, Y. Wu, G. Xia, P. G. Schultz and F. E. Romesberg, Efforts toward Expansion of the Genetic Alphabet: Replication of DNA with Three Base Pairs, *J. Am. Chem. Soc.*, 2001, **123**(30), 7439–7440.

17 L. Li, M. Degardin, T. Lavergne, D. A. Malyshev, K. Dhami, P. Ordoukhanian and F. E. Romesberg, Natural-like Replication of an Unnatural Base Pair for the Expansion of the Genetic Alphabet and Biotechnology Applications, *J. Am. Chem. Soc.*, 2014, **136**(3), 826–829.

18 J. Oh, M. Kimoto, H. Xu, J. Chong, I. Hirao and D. Wang, Structural Basis of Transcription Recognition of a Hydrophobic Unnatural Base Pair by T7 RNA Polymerase, *Nat. Commun.*, 2023, **14**, 195.

19 E. S. Hoffmann, M. C. De Pascali, L. Neu, C. Domnick, A. Soldà and S. Kath-Schorr, Reverse Transcription as Key Step in RNA *in Vitro* Evolution with Unnatural Base Pairs. RSC, *Chem. Biol.*, 2024, **5**, 556–566.

20 J. L. Ptacin, C. E. Caffaro, L. Ma, K. M. San Jose Gall, H. R. Aerni, N. V. Acuff, R. W. Herman, Y. Pavlova, M. J. Pena, D. B. Chen, L. K. Koriazova, L. K. Shawver, I. B. Joseph and M. E. Milla, An Engineered IL-2 Reprogrammed for Anti-Tumor Therapy Using a Semi-Synthetic Organism, *Nat. Commun.*, 2021, **12**, 4785.

21 E. C. Fischer, K. Hashimoto, Y. Zhang, A. W. Feldman, V. T. Dien, R. J. Karadeema, R. Adhikary, M. P. Ledbetter, R. Krishnamurthy and F. E. Romesberg, New Codons for Efficient Production of Unnatural Proteins in a Semisynthetic Organism, *Nat. Chem. Biol.*, 2020, **16**(5), 570–576.

22 Y. J. Seo, S. Matsuda and F. E. Romesberg, Transcription of an Expanded Genetic Alphabet, *J. Am. Chem. Soc.*, 2009, **131**(14), 5046–5047.

23 Y. Xie, T. Hu, Y. Zhang, D. Wei, W. Zheng, F. Zhu, G. Tian, H. A. Aisa and J. Shen, Weinreb Amide Approach to the Practical Synthesis of a Key Remdesivir Intermediate, *J. Org. Chem.*, 2021, **86**(7), 5065–5072.

24 R. R. Kadiyala, D. Tilly, E. Nagaradja, T. Roisnel, V. E. Matulis, O. A. Ivashkevich, Y. S. Halauko, F. Chevallier, P. C. Gros and F. Mongin, Computed CH Acidity of Biaryl Compounds and Their Deprotonative Metalation by Using a Mixed Lithium/Zinc-TMP Base, *Chem. – Eur. J.*, 2013, **19**(24), 7944–7960.

25 K. Okano, K. I. Okuyama, T. Fukuyama and H. Tokuyama, Mild Debenzylation of Aryl Benzyl Ether with BCl3 in the Presence of Pentamethylbenzene as a Non-Lewis-Basic Cation Scavenger, *Synlett*, 2008, 1977–1980.

26 S. Mohamady, A. Desoky and S. D. Taylor, Sulfonyl Imidazolium Salts as Reagents for the Rapid and Efficient Synthesis of Nucleoside Polyphosphates and Their Conjugates, *Org. Lett.*, 2012, **14**(1), 402–405.

27 A. W. Feldman, V. T. Dien, R. J. Karadeema, E. C. Fischer, Y. You, B. A. Anderson, R. Krishnamurthy, J. S. Chen, L. Li and F. E. Romesberg, Optimization of Replication, Transcription, and Translation in a Semi-Synthetic Organism, *J. Am. Chem. Soc.*, 2019, **141**(27), 10644–10653.

28 K. Moriyama, M. Kimoto, T. Mitsui, S. Yokoyama and I. Hirao, Site-Specific Biotinylation of RNA Molecules by Transcription Using Unnatural Base Pairs, *Nucleic Acids Res.*, 2005, **33**(15), 1–8.

29 Y. J. Seo, D. A. Malyshev, T. Lavergne, P. Ordoukhanian and F. E. Romesberg, Site-Specific Labeling of DNA and RNA Using an Efficiently Replicated and Transcribed Class of Unnatural Base Pairs, *J. Am. Chem. Soc.*, 2011, **133**(49), 19878–19888.

30 I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, K. Kodama, T. Yokogawa, K. Nishikawa and S. Yokoyama, An Unnatural Base Pair for Incorporating Amino Acid Analogs into Proteins, *Nat. Biotechnol.*, 2002, **20**, 177–182.

31 Y. J. Seo, S. Matsuda and F. E. Romesberg, Transcription of an Expanded Genetic Alphabet, *J. Am. Chem. Soc.*, 2009, **131**(14), 5046–5047.

32 F. Eggert, K. Kurscheidt, E. Hoffmann and S. Kath-Schorr, Towards Reverse Transcription with an Expanded Genetic Alphabet, *ChemBioChem*, 2019, **20**(13), 1642–1645.

33 A. X. Z. Zhou, X. Dong and F. E. Romesberg, Transcription and Reverse Transcription of an Expanded Genetic Alphabet in Vitro and in a Semisynthetic Organism, *J. Am. Chem. Soc.*, 2020, **142**(45), 19029–19032.

34 J. Oh, J. Shin, I. C. Unarta, W. Wang, A. W. Feldman, R. J. Karadeema, L. Xu, J. Xu, J. Chong, R. Krishnamurthy, X. Huang, F. E. Romesberg and D. Wang, Transcriptional Processing of an Unnatural Base Pair by Eukaryotic RNA Polymerase II, *Nat. Chem. Biol.*, 2021, **17**(8), 906–914.

35 B. Shakya, O. G. Joyner and M. C. T. Hartman, Hyperaccurate Ribosomes for Improved Genetic Code Reprogramming, *ACS Synth. Biol.*, 2022, **11**(6), 2193–2201.

36 C. A. L. McFeely, B. Shakya, C. A. Makovsky, A. K. Haney, T. Ashton Cropp and M. C. T. Hartman, Extensive Breaking of Genetic Code Degeneracy with Non-Canonical Amino Acids, *Nat. Commun.*, 2023, **14**(1), 5008.

37 G. N. Kerestesy, K. K. Dods, C. A. L. McFeely and M. C. T. Hartman, Continuous Fluorescence Assay for In Vitro Translation Compatible with Noncanonical Amino Acids, *ACS Synth. Biol.*, 2024, **13**(1), 119–128.

38 T. Katoh and H. Suga, A Comprehensive Analysis of Translational Misdecoding Pattern and Its Implication on Genetic Code Evolution, *Nucleic Acids Res.*, 2023, **51**(19), 10642–10652.
39 A. Cappannini, A. Ray, E. Purta, S. Mukherjee, P. Boccaletto, S. N. Moafinejad, A. Lechner, C. Barchet, B. P. Klaholz, F. Stefaniak and J. M. Bujnicki, MODOMICS: A Database of RNA Modifications and Related Information. 2023 Update, *Nucleic Acids Res.*, 2024, **52**(D1), D239–D244.
40 M. E. Armengod, S. Meseguer, M. Villarroya, S. Prado, I. Moukadiri, R. Ruiz-Partida, M. J. Garzón, C. Navarro-González and A. Martínez-Zamora, Modification of the Wobble Uridine in Bacterial and Mitochondrial TRNAs Reading NNA/NNG Triplets of 2-Codon Boxes, *RNA Biol.*, 2014, **11**(12), 1495–1507.
41 W. T. Pedersen and J. F. Curran, Effects of the Nucleotide 3′ to an Amber Codon on Ribosomal Selection Rates of Suppressor TRNA and Release Factor-1, *J. Mol. Biol.*, 1991, **219**(2), 231–241.
42 Z. Cui, V. Stein, Z. Tnimov, S. Mureev and K. Alexandrov, Semisynthetic TRNA Complement Mediates in Vitro Protein Synthesis, *J. Am. Chem. Soc.*, 2015, **137**(13), 4404–4413.
43 C. A. L. McFeely, K. K. Dods, S. S. Patel and M. C. T. Hartman, Expansion of the Genetic Code through Reassignment of Redundant Sense Codons Using Fully Modified TRNA, *Nucleic Acids Res.*, 2022, **50**(19), 11374–11386.
44 R. Giege, M. Sissler and C. Florentz, Universal Rules and Idiosyncratic Features in TRNA Identity, *Nucleic Acids Res.*, 1998, **26**(22), 5017–5035.
45 M. C. T. Hartman, K. Josephson and J. W. Szostak, Enzymatic Aminoacylation of TRNA with Unnatural Amino Acids, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**(12), 4356–4361.
46 Y. Cao, J. Bai, J. Zou, Y. Du and T. Chen, One-Pot Enzymatic Preparation of Oligonucleotides with an Expanded Genetic Alphabet via Controlled Pause and Restart of Primer Extension: Making Unnatural Out of Natural, *ACS Synth. Biol.*, 2023, **12**(9), 2691–2706.
47 S. Carulli, C. D. Calvano, F. Palmisano and M. Pischetsrieder, MALDI-TOF MS Characterization of Glycation Products of Whey Proteins in a Glucose/Galactose Model System and Lactose-Free Milk, *J. Agric. Food Chem.*, 2011, **59**(5), 1793–1803.
48 J. Hoinka, R. Backofen and T. M. Przytycka, AptaSUITE: A Full-Featured Bioinformatics Framework for the Comprehensive Analysis of Aptamers from HT-SELEX Experiments, *Mol. Ther. Nucleic Acid.*, 2018, **11**, 515–517.
49 R. Matthiesen and F. Kirpekar, Identification of RNA Molecules by Specific Enzyme Digestion and Mass Spectrometry: Software for and Implementation of RNA Mass Mapping, *Nucleic Acids Res.*, 2009, **37**(6), e48.