Single enzyme RT-PCR of full-length ribosomal RNA

Michael J. Hammerling ⓘ [1], Danielle J. Yoesep[1], and Michael C. Jewett[1,2,3,4,5,*]

[1]Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA, [2]Center for Synthetic Biology, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA, [3]Simpson Querrey Institute, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA, [4]Chemistry of Life Processes Institute, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA and [5]Robert H. Lurie Comprehensive Cancer Center, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208, USA

*Corresponding author: E-mail: m-jewett@northwestern.edu

Abstract

The ribosome is a two-subunit, macromolecular machine composed of RNA and proteins that carries out the polymerization of α-amino acids into polypeptides. Efforts to engineer ribosomal RNA (rRNA) deepen our understanding of molecular translation and provide opportunities to expand the chemistry of life by creating ribosomes with altered properties. Toward these efforts, reverse transcription PCR (RT-PCR) of the entire 16S and 23S rRNAs, which make up the 30S small subunit and 50S large subunit, respectively, is important for isolating desired phenotypes. However, reverse transcription of rRNA is challenging due to extensive secondary structure and post-transcriptional modifications. One key challenge is that existing commercial kits for RT-PCR rely on reverse transcriptases that lack the extreme thermostability and processivity found in many commercial DNA polymerases, which can result in subpar performance on challenging templates. Here, we develop methods employing a synthetic thermostable reverse transcriptase (RTX) to enable and optimize RT-PCR of the complete *Escherichia coli* 16S and 23S rRNAs. We also characterize the error rate of RTX when traversing the various post-transcriptional modifications of the 23S rRNA. We anticipate that this work will facilitate efforts to study and characterize many naturally occurring long RNAs and to engineer the translation apparatus for synthetic biology.

Key words: synthetic biology; ribosome construction; *in vitro*; directed evolution; RT-PCR

1. Introduction

Directed evolution of the ribosome and its associated translation factors has emerged as a promising opportunity to create new classes of enzymes, therapeutics and materials with diverse genetically encoded chemistry (1–5). The key idea is that the ribosome can be engineered to make proteins and polymers that selectively incorporate non-canonical monomers (6). To date, such efforts have incorporated a wide range of non-canonical α- (7), β- (8, 9), γ- (10–13), δ-, ε-, ζ- (13), D- (14, 15), aromatic (16–18), aliphatic (16, 19), malonyl (17), N-alkylated (20), cyclic (9, 13, 21), aminobenzoic (22) and oligomeric (10, 23, 24) amino acid analogs, among others. While the incorporation of
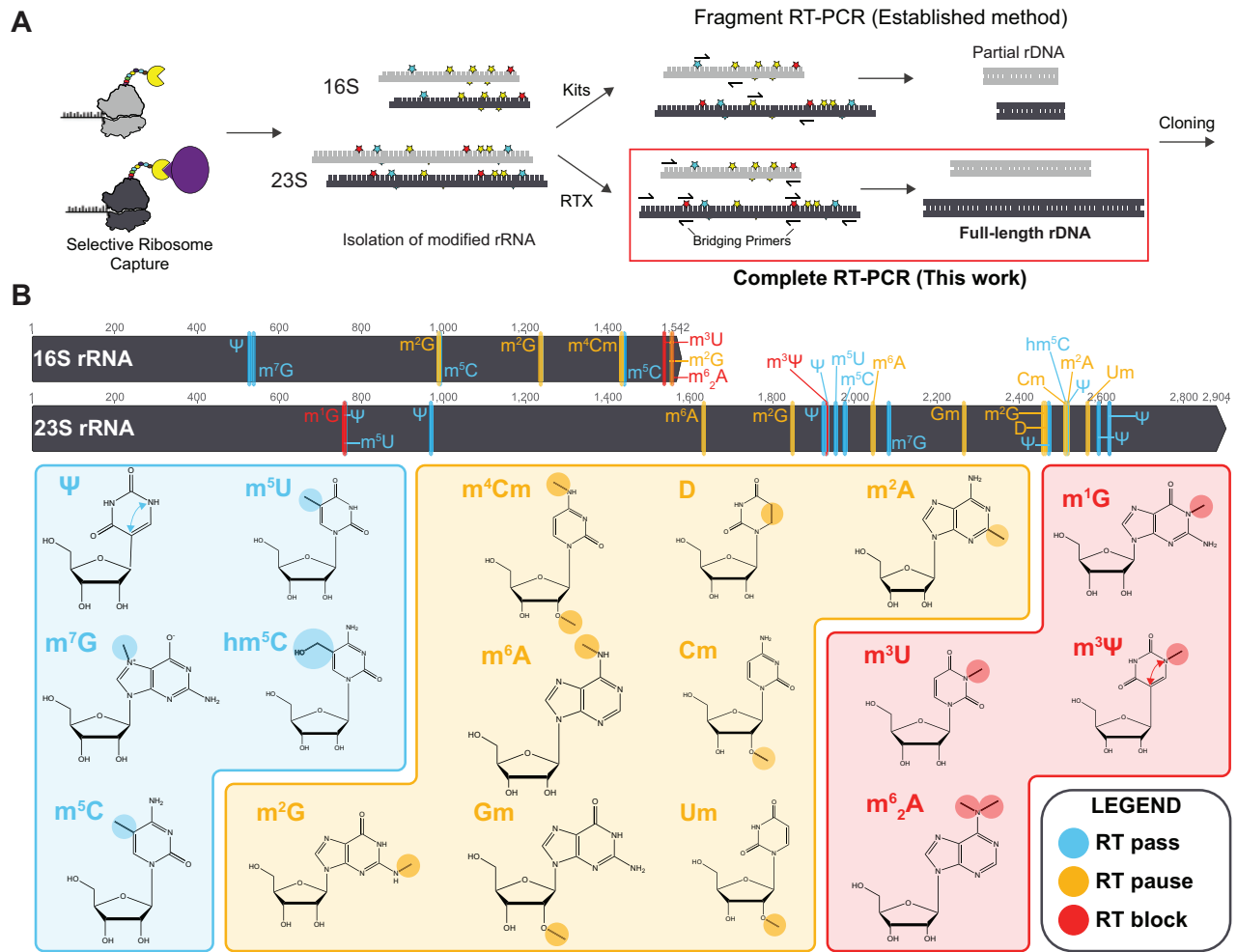
such diverse chemistries into peptides and proteins has facilitated exciting applications (e.g. macrocyclic foldamer–peptide drugs (25, 26)), there is poor compatibility with the natural translation apparatus for numerous classes of non-canonical monomers (e.g. backbone-extended amino acids) leading to incorporation inefficiencies. An especially challenging constraint is the ribosome, which has evolved to polymerize α-amino acids. Because its function is necessary for life, cell viability restricts the mutations that can be made to ribosomes.

To overcome this compatibility challenge, new methods for engineering ribosomes have been developed *in vivo* and *in vitro*. *In vivo*, powerful positive/negative selections have been used to

Figure 1: Schematic of RT-PCR of full-length rRNA and rRNA modifications. (**A**) Diagram of the established methods for converting rRNA into rDNA and the methods developed in this work. rRNA from ribosomes selected by directed evolution is purified and must be reverse transcribed to recover successful genotypes. In prior work, rDNA libraries have focused on small portions of the ribosome which lack post-transcriptional modifications that block reverse transcription to enable library recovery (top arrows) This work documents the development of methods to reverse transcribe the entirety of the 16S and 23S rRNAs to enable construction and selection of libraries covering the whole ribosome (bottom arrows). (**B**) Locations and structures of post-transcriptional modifications of *E. coli* rRNAs. *Escherichia coli* rRNAs are extensively modified, with the 16S and 23S rRNAs containing as many as 11 and 24 modifications, respectively. In addition to their significance in ribosome assembly and function, these modifications may be categorized by the impact on RT-PCR. Permissive modifications (cyan) have little to no impact on polymerase extension, pausing modifications (orange) cause the polymerase to pause briefly and blocking modifications (red) cause the polymerase to stop and fall off the message. Blocking modifications generally sterically disrupt the Watson–Crick interface, while pausing modifications may modestly interfere with Watson–Crick base-pairing or polymerase binding to the RNA backbone.

engineer orthogonal ribosomes (27, 28) and quadruplet decoding ribosomes (29), as well as ribosomes for β-amino acid incorporation (30) and D-amino acid incorporation (31, 32). In addition, the development of orthogonal tethered ribosomes has led to new functions inaccessible to the natural ribosome (33–37). *In vitro*, directed evolution of ribosomes using ribosome display offers the compelling advantages of flexibility and throughput, allowing access to lethal ribosomal genotype and larger rRNA variant libraries (38, 39). Despite these advances, key challenges remain. One such challenge to the *in vitro* approach is the recovery of full-length ribosomal cDNA (rDNA) from successful 16S and 23S rRNA sequences to enable the directed evolution of the whole ribosome (Figure 1A).

Typical methods for rRNA recovery, including 16S rRNA profiling broadly used for measuring microbial diversity in an environmental sample (40–47), rely on PCR of the rDNA (not rRNA) from genomic material in the sample. Additionally, these methods target only a small region of the 16S rRNA, as this is all that is required for microbial profiling. In contrast, RT-PCR methods for directed evolution of the ribosome for novel function would ideally allow recovery of the entire 16S and 23S rRNA, enabling diversification, selection and recovery of disparate regions of the molecule (Figure 1A). Unfortunately, rRNA is a challenging template for RT-PCR due to the presence of extensive secondary structure and post-transcriptional modifications (PTxMs), which can interfere with RT-PCR. These modifications can be categorized as being permissive to reverse-transcription, inducing pausing, or altogether blocking polymerization (48) (Figure 1B). Blocking PTxMs generally disrupt the Watson–Crick interface of a base, while pausing PTxMs may weaken the base-pairing interaction or interfere with binding of the polymerase to ribose backbone. The 16S and 23S rRNAs of the *Escherichia coli* ribosome contain 11 and 24 known modifications, respectively (49) (Supplementary Tables S1 and S2). For these reasons, recent

efforts at *in vitro* directed evolution of the rRNA were limited to a small fragment of the 23S rRNA that did not contain any blocking PTxMs (38, 39).

In this study, we set out to develop a robust method for recovery of rDNA of the entire 16S rRNA and 23S rRNAs from translating ribosomes using a single enzyme. In this method, we leverage RT-PCR with a thermostable reverse transcriptase derived from the DNA polymerase KOD to identify regions of the 16S and 23S rRNAs which are intransient to RT-PCR. Then, we optimized methods to bypass or polymerize through these challenging regions using RTX. Finally, we quantify the error rate when RTX polymerizes across various modified RNA nucleotides. We anticipate that our robust approach for RT-PCR of full-length 16S and 23S rRNA will facilitate efforts to engineer the ribosome. In addition, our work lays the groundwork to use RTX for the characterization of, in principle, any naturally occurring long RNAs such as long messenger RNAs (mRNAs) and long non-coding RNAs (lncRNAs).

2. Results

The goal of this study was to develop a methodology for producing consistent, robust RT-PCR product of full-length 16S and 23S rRNAs. We anticipated that this would be difficult because of the complex rRNA secondary structure resulting from the prevalence of post-transcriptional modifications (PTxMs) (48) and the challenges surrounding the RT processivity. To start, we assessed the ability of a commercial reverse transcriptase kit; namely, SuperScript^{TM} III, to achieve this goal using the manufacturer's recommended protocol. rRNA from 70S ribosomes purified from *E. coli* MRE600 cells was used as the test template for RT-PCR experiments. Primers were designed with a high melting temperature to ensure proper annealing to the complex structure of the rRNA (Supplementary Tables S3 and S4). As expected, this attempt failed to produce a robust single-band product representing full-length rDNA.

We hypothesized that we could enable RT-PCR of full-length 16S and 23S rRNA by (i) employing a highly processive, thermostable reverse transcriptase and (ii) taking measures to enable the polymerase to either bypass or traverse the PTxMs present in rRNA. For the reverse transcriptase, we opted to utilize a recently developed synthetic RT (RTX) derived from the DNA polymerase KOD (50). Capable of recognizing both RNA–DNA and DNA–DNA duplexes and polymerizing DNA, this engineered polymerase performs both the RT step and subsequent PCR steps in a one-pot RT-PCR reaction. As a result, RTX had the potential to increase sensitivity by allowing multiple rounds of reverse-transcription (in contrast to commercial one-step kits, in which most of the RT enzyme is inactivated after one round), while also improving the quantity of the correct final product produced. To bypass PTxMs present in rRNA, we postulated that we could use bridging primers or long extension times.

To assess the potential for RTX to RT-PCR full-length rRNA, we first examined the ability of RTX to successfully polymerize various regions of the rRNA. Initially, RT-PCR was performed on an 1135 bp region spanning positions 784–1800 of the 23S rRNA which is known to lack blocking PTxMs. RTX produced a robust band at 1135 bp (Figure 2A). As a control, we also carried out PCR reactions using three additional DNA polymerases (Pfx, Phusion and Q5), which cannot reverse transcribe RNA. None of these polymerases produced bands of the correct size, indicating that our sample was free of contaminating genomic ribosomal cDNA.

We next assessed the ability of RTX to RT-PCR 9 regions of the 23S rRNA and 6 regions of the 16S rRNA known to have PTxMs (Figure 2B). The exact start point, endpoint and expected length of each product can be found in Supplementary Table S5. We started with the 23S rRNA. RT-PCR products spanning positions 1–770, 1867–2218 and 2535–2904 of the 23S rRNA, as well as nucleotides 1–980 and 945–1497 of the 16S rRNA were amenable to RT-PCR under standard conditions [68°C—30 min, 35× (95°C—30 s, 68°C—30 s/kb)]. Longer 23S rRNA RT-PCR reactions spanning position 745 (1-methylguanosine), 1915 (3-methyl-pseudouridine), 2251 (2′-O-methylguanosine) and the region from 2498 to 2504 containing several PTxMs did not reliably produce full-length product in our initial tests (Figure 2C, products 2, 5, 7–9). However, shorter targets spanning the blocking PTxM at position 1915 (Figure 2C, products 3, 4) did produce products of the correct size, suggesting that RTX may be capable of traversing blocking PTxMs.

Most of the 16S rRNA was amenable to RT-PCR, noting that the longest products used reverse primers that annealed just 5′ of three blocking PTxMs at the 3′ end of the gene (Figure 1B). Single products were produced spanning positions 1–980 and 945–1497. The nearly full-length RT-PCR reaction spanning positions 1–1497 of the 16S rRNA produced some full-length product, but also a more dominant smaller off-product.

From these initial reactions, two strategies to recover full-length cDNA using 16S and 23S rRNAs as templates were attempted. The first RT-PCR strategy for achieving full-length ribosomal cDNA was to utilize 'bridging primers' to anneal across problematic blocking or pausing post-transcriptional modifications. This would allow the production of constituent fragments of the 23S and 16S cDNA, which would be built into full-length products using overlap PCR in subsequent PCR cycles. Given previous results, positions 745 (1-methylguanosine), 1915 (3-methylpseudouridine), 2251 (2′-O-methylguanosine), and the region from 2498-2504 of the 23S rRNA were the prime candidates for bypassing with bridging primers (Figure 3A, Supplementary Table S2). Given the initial results for RT-PCR of the 16S rRNA (Figure 2B and C), we chose a single pair of bridging primers spanning modifications m^{2}G and m^{5}C at positions 966–967 (Figure 3C, Supplementary Table S1).

We next designed bridging primers to anneal across each of these PTxMs (Supplementary Table S3). For the 23S rRNA, all possible combinations of these primers were tested to assess the minimal set of primers necessary to enable RT-PCR of the 23S rRNA. We found that including bridging primer sets 1 and 2 (corresponding to PTxMs at positions 745 and 1915 that disrupt Watson–Crick base pairing) were the minimal set that enabled robust RT-PCR of the 23S rRNA (Figure 3A). Using this primer set for the 23S rRNA and the single set designed for the 16S rRNA, we found that this strategy enabled recovery of full-length rDNA using elongation steps of between 3 and 6 min for the 23S rRNA (Supplementary Figure S1), and between 2 and 4 min with the 16S rRNA (Supplementary Figure S2). Optimal results were achieved from the protocol with 6-min elongation for the 23S rRNA (Figure 3B) and the 3-min elongation for the 16S rRNA (Figure 3C), enabling recovery of full-length rDNA from as low as 40 pg of rRNA in each case.

While the bridging primer strategy allowed the production of the correct product, RT-PCR using bridging primers entails loss of genotype fidelity, due to recombining of the regions broken up by the bridging primers. This may be undesirable for ribosome engineering applications. Additionally, we were only able to recover cDNA from 97% of the 16S rRNA gene using bridging primers, since the three blocking PTxMs of the 16S rRNA are so
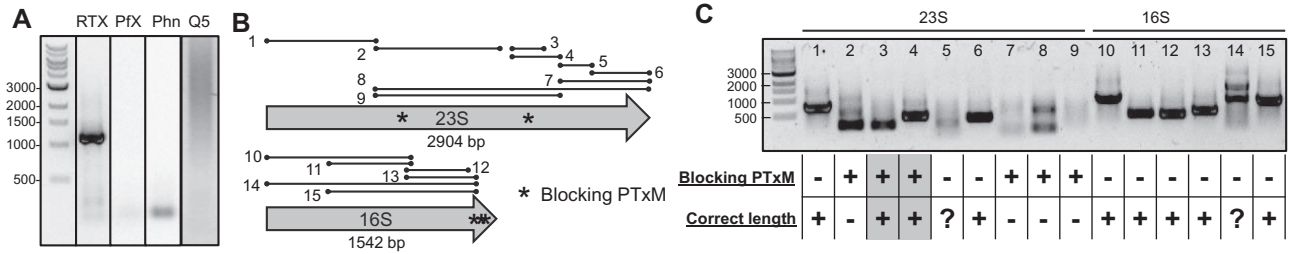
Figure 2: Post-transcriptional modifications inhibit RT-PCR of the rRNA. (**A**) RT-PCR spanning positions 725 to 1858 of MRE600 23S rRNA from cells yields a robust product, while PCR using DNA polymerases PfX, Phusion (Phn) or Q5 do not produce the correct product. (**B**) Diagram of the RT-PCR products attempted using RTX. Products were chosen to yield information about the ability of RTX to traverse PTxMs. (**C**) For the 23S rRNA, products spanning post-transcriptional modifications do not produce product in 4 of 6 cases, (products 2, 7, 8 and 9) suggesting that these modifications inhibit RTX polymerization. However, the two products containing PTxMs that did appear to produce the correct product (3 and 4, shaded cells) suggest that RTX can traverse blocking PTxMs if given enough time. For the 16S rRNA, all products except 14 conclusively form a product of the correct length, while product 14 appears to form a minor product of approximately correct size. Gels are representative of three independent experiments.
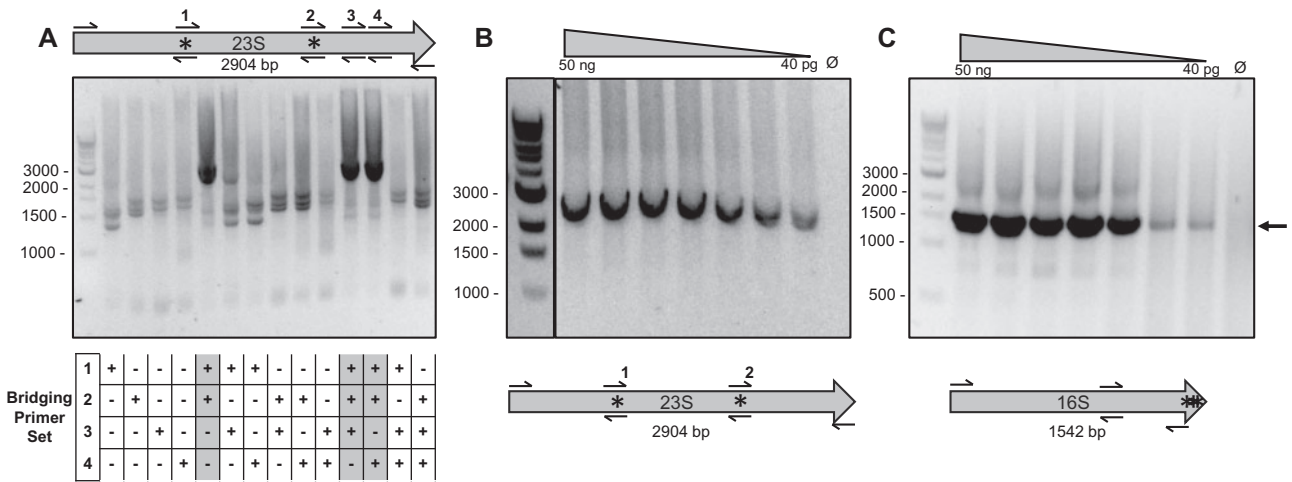

Figure 3: Bridging primers enable RT-PCR of the entire 23S rRNA. (**A**) Diagram of the bridging primer sets tested for the 23S rRNA at top. An agarose gel of RT-PCRs attempted on the 23S rRNA using RTX and combinations of bridging primers is depicted. The sets of bridging primers used in each lane are indicated in each column of the table below the gel, with '+' indicating the presence and '−' indicating the absence of the primer set. Shaded columns in the table indicate primer combinations which produced primarily the correct full-length product. The minimal primers for successful RT-PCR was primer sets 1 and 2 together, which span blocking PTxMs. (**B**) RT-PCR of the full-length 23S rRNA using RTX and primer sets 1 and 2 together from a gradient of rRNA concentrations. (**C**) RT-PCR of >97% of the 16S rRNA (1497 bases) using a single set of bridging primers. In all panels, the asterisk represents blocking PTxMs. Gels are representative of three independent experiments.

proximal to the 3′ end of the RNA. We wondered if there might be alternative strategies for achieving full-length RT-PCR of the 23S and 16S rRNAs that did not require bridging primers.

As an alternative strategy, we assessed the possibility of RTX read-through of PTxMs by testing for long extension times. To do so, we attempted RT-PCR of the 23S rRNA with products spanning various regions of increasing length with elongation steps of 8, 10, 12 and 14 min (Figure 4A, Supplementary Figure S3 and Table S4). The exact start point, endpoint and expected length of each product can be found in Supplementary Table S6. We observed improved production of desired products with increasing extension duration. In particular, products K and L, comprising ~2000 nucleotides and spanning the two blocking PTxMs we identified earlier (Figure 2), benefited from increasing extension time, with little or no observed product in the 8-minute extension step and a clear observable product in the 14-min

condition (Figure 4B, Supplementary Figure S3). These results encouraged us to try even more extreme extension durations to achieve full-length products of these templates. Given our previous success using a forward primer at the 5′ extremity of the 23S rRNA, we held this primer constant and attempted RT-PCRs on products ranging from 2055 to 2904 bases in length with extension steps of 20, 25 and 30 min (Supplementary Figure S4). We achieved robust products up to 2675 bases in length, with 25- and 30-min extension steps yielding the strongest bands (Figure 4C). Our failure once again to achieve full-length products for the 23S rRNA despite the absence of blocking or pausing PTxMs between positions 2675 and 2904 led us to implicate a problematic reverse primer in this RT-PCR, and thus to attempt a variety of primers at the 3′ extremity of the 23S rRNA. Using a variety of reverse primers and varying elongation times, we were able to achieve full-length product of the 23S rRNA (Figure 4D, Supplementary Figure S5 and Table S4) Notably,
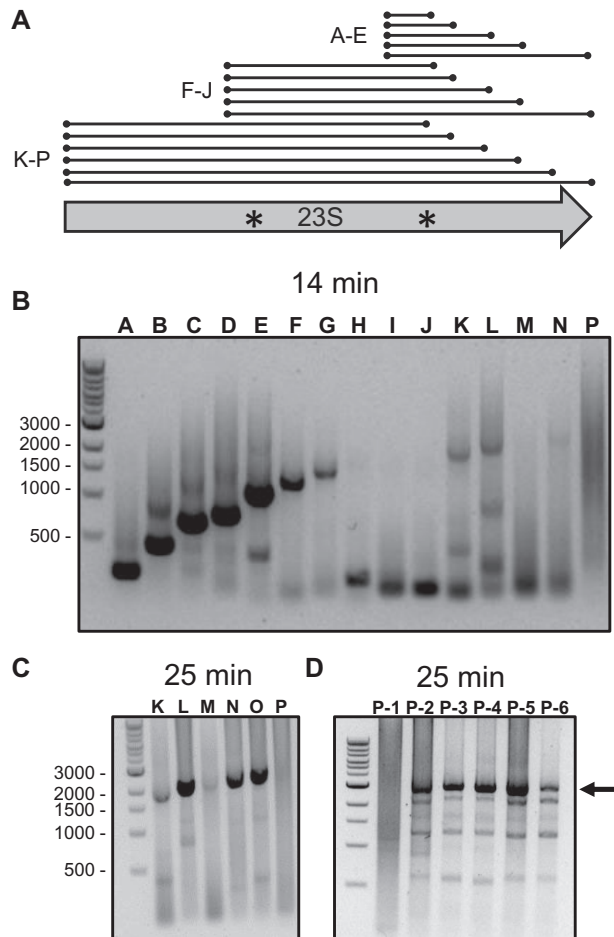
Figure 4: RT-PCR of entire 23S rRNA using long elongation steps. (**A**) Diagram of the RT-PCR reactions of the 23S rRNA attempted. Products are labeled with letters from A–P from the top down. (**B**) RT-PCR elongation steps of 8, 10, 12 and 14 min were attempted (agarose gel of 14-min reactions shown). Longer elongation steps promote formation of correct products that were not obtained using standard reaction conditions. (**C**) RT-PCR elongation steps of 20, 25 and 30 min were performed (agarose gel of 25-min reactions shown). Elongation steps of this length enable production of lengths up to but not including the full-length product. (**D**) The reverse primer was varied in the full-length 23S rRNA RT-PCR reaction to troubleshoot this reaction. Correct full-length product was obtained in all reactions with a new reverse primer, though with some smaller off-target products as well. Arrow indicates correct full-length RT-PCR product of the 23S rRNA. Gels are representative of three independent experiments.

however, these RT-PCR reactions contained both low and high molecular weight aberrant products.

While long extension times enabled production of full-length 23S rDNA product, we hypothesized that extension times of this length would not be necessary for all the cycles of a PCR reaction, since the initial 30-min cycles would produce template cDNA for future cycles which should be more amenable to PCR. To test whether shortening these steps reduced off-product while retaining full-length product, we performed reactions with one 30-min extension step, followed by remaining cycles with, 3-, 4-, 5- or 6-min extension steps for the 23S rRNA (Supplementary Figure S6). We found that 5-min or longer extension steps were still required for the 23S rRNA. This suggests that RTX requires more time to traverse the difficult 23S rRNA template than typical templates, even when it has already been reverse transcribed into cDNA by the first long RT cycles.

Using the information gained from optimization of the 23S rRNA RT-PCR, we attempted RT-PCR of the 16S rRNA using no

bridging primers, an initial 30-min extension step and 2-, 3- or 4-min elongation steps while cycling. While we were able to produce the correct product under all cycling conditions, we found that RT-PCR of the 16S rRNA without bridging primers generally resulted in production of many off-products as well, likely requiring extraction of the correct band for downstream applications (Supplementary Figure S7).

Commercial one-step RT-PCR kits use a combination of a reverse transcriptase and a thermostable DNA polymerase to enable a single step of reverse-transcription followed by PCR in a single reaction vessel. Given the prevalence of low-molecular weight off-products in some of our RT-PCR reactions, we wanted explore methods that might reduce these undesirable side-reactions. To test whether a method comparable to one-step RT-PCR kits might improve our results with RTX, we attempted a two-enzyme strategy that mimicked their methodology. We hypothesized that the added benefit of thermostability of RTX would enable multiple initial RT steps before the proofreading DNA polymerase takes over (Section 5). However, our results were comparable for the RTX alone and RTX + Q5 reactions for amplification of the 23S rRNA, indicating that this dual enzyme strategy did not improve amplification of the 23S rRNA when using RTX (Figure 5A). However, this strategy did result in more reliable production of full-length 16S cDNA (Figure 5B). We anticipate that future work with other dual-enzyme mixes including RTX would be useful to reduce off-products.

The base modifications present in rRNA are known to cause elevated error rates in reverse-transcription reactions, a fact that has been leveraged recently to identify modified nucleotides from transcriptomic data (51–54). Since we used the version of RTX lacking proofreading exonuclease function, we anticipated that using this enzyme to polymerize through modified nucleotides—especially those canonically thought to block polymerization—would result in elevated error rates at these positions. To assess the error rate of RTX across the 23S rRNA, cDNA from a full-length RT-PCR reaction of the 23S rRNA lacking bridging primers was deep sequenced, and divergence of the observed base from wild type at each position was quantified (Figure 6A). As expected, the two blocking PTxMs in the 23S rRNA (m^1G and $m^3\Psi$ at positions 745 and 1915, respectively) had elevated misincorporation rates of 88.2% (s.d. = 0.003, $n = 4$) and 29.2% (s.d. = 0.002, $n = 4$), respectively. These two modifications cause widely differing mutational spectra as well which were consistent across 4 replicates, with 'C' and 'A' as the dominant missense mutation for positions 745 and 1915, respectively (Figure 6B). While most other permissive and pausing modifications resulted in error rates comparable to the surrounding region of the 23s rRNA, modifications m^6A and U_m at positions 1618 and 2552 also had elevated rates of missense mutation (Figure 6A, inset). This is consistent with previous work showing that reverse transcription across m^6A results in an elevated error rate (52). Interestingly, a number of other positions which are not reported to be modified display an elevated mutation rate above the trend for that region of the 23s rRNA. It is unclear whether these positions are the result of sequencing artifacts or reflect real molecular heterogeneity or modification, and further investigation may be necessary to discern the source of the elevated rate of mismatch mutations at these positions (51, 53).

3. Discussion

We developed and optimized methodologies to achieve full-length RT-PCR of the 23S and 16S rRNAs. Specifically, we
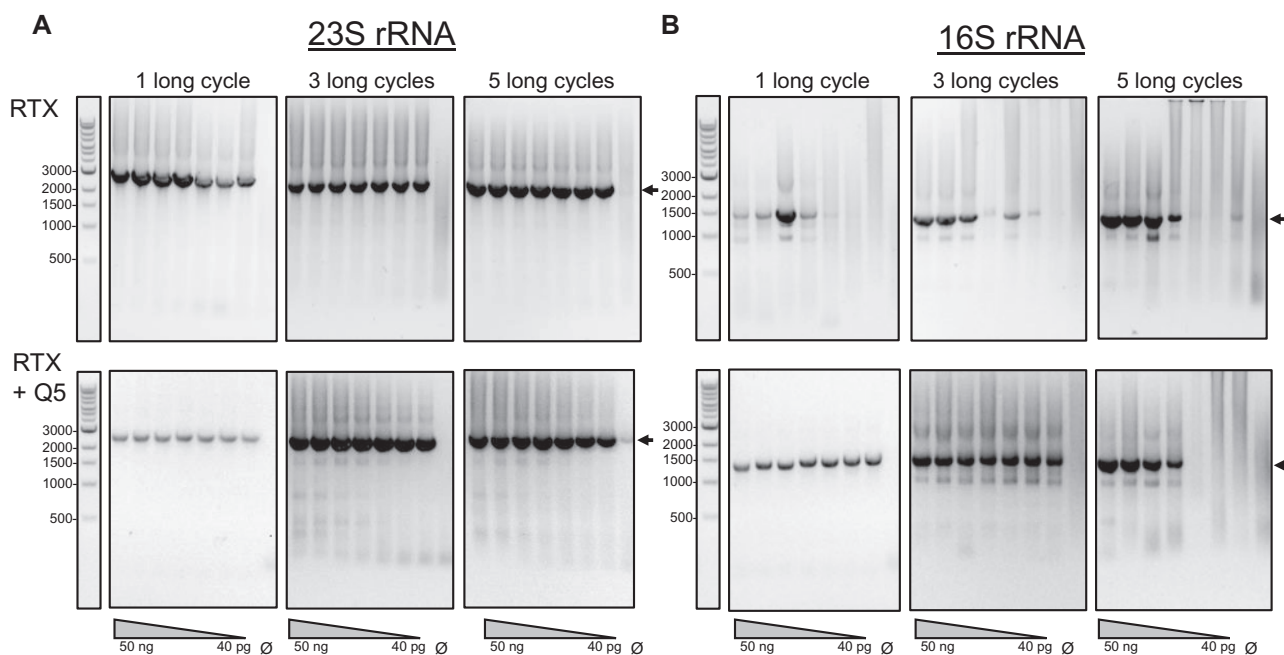
Figure 5: Two-enzyme strategy for RT-PCR of full-length 23S and 16S rRNA. (**A**) Amplification of 23S cDNA from purified rRNA was attempted using RTX alone or RTX reactions spiked into Q5 polymerase PCRs after 1, 3 or 5 long (30 min) RTX extension steps. Reactions were performed on 2-fold serially diluted rRNA ranging from 50 ng to 40 pg total rRNA. Little difference was noted between the RTX alone and RTX + Q5 conditions for the 23S rRNA, indicating that a two-enzyme strategy does not improve recovery of 23S rDNA. (**B**) The same conditions were tested for recovery of 16S cDNA. The two-enzyme strategy does improve reliability of the recovery of 16S cDNA. Gels are representative of three independent experiments.
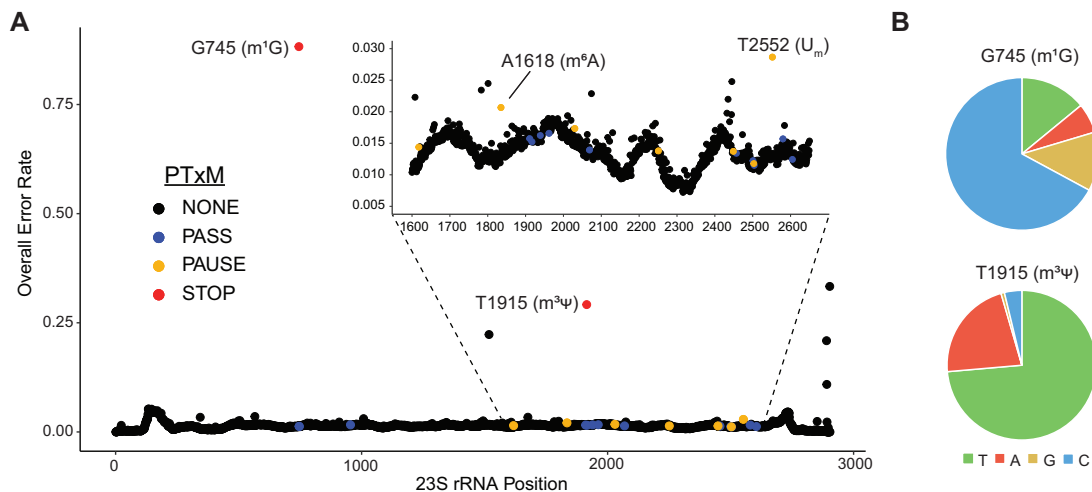
Figure 6: Error rate of RTX traversing post-transcriptional modifications. (**A**) The average error rate ($n = 4$ independent RT-PCR reactions) of each position of the 23 s rRNA is plotted and color-coded for positions with no PTxM (black), or polymerization-permissive (blue), -pausing (yellow) or -blocking (red) PTxM. The inset is a zoomed in plot of positions 1600–2700 which contain the bulk of the permissive and pausing PTxMs to enable greater resolution of plotting. Blocking PTxMs induce a greatly elevated error rate, while select positions with pausing PTxMs (1618, 2552) also show an increase in error rate compared to the trend for that region of the 23S rRNA. For most points, standard deviation error bars are smaller than the diameter of the point and thus not depicted for clarity of plotting. (**B**) Pie charts depicting the RTX mutation spectrum at positions 745 and 1915 with blocking PTxMs m[1]G and m[3]Ψ, respectively.

leveraged two strategies: (i) bridging primers that anneal over PTxMs to bypass them and (ii) long extension steps to polymerize through these positions. Taken together, these methods should facilitate efforts to study and engineer mutant ribosomes.

We present two strategies since they each have advantages and disadvantages. The use of bridging primers has the advantage of a robust correct product and fewer off-products for both the 23S and 16S rRNAs. Additionally, for directed evolution studies, recombination of distal regions of the rRNA that arise in this approach would add new genetic diversity of the ribosome pool under selection. The disadvantage of this approach is that the bridging primers result in a loss of genotype fidelity. RT-PCR using a long extension step to polymerize through PTxMs may be the more desirable strategy in situations where the genotype of distal regions of the ribosome must be preserved. However, the RT-PCR itself is less robust to synthesis of a single target product. Moreover, the RTX enzyme used here

has an elevated error rate when polymerizing through blocking PTxMs. Studying the effect of removing PTxMs individually and in combination would help elucidate the role of these modifications in translation function.

Up to now, efforts to engineer ribosomes *in vitro* that rely on RT-PCR have been limited by the inability to recover the entire 16S and 23S rRNA (38, 39). Looking forward, the methods reported here should enable diversification, selection and recovery of disparate regions of the ribosome, setting the stage for new directions at the interface of chemical and synthetic biology.

This work also has unexpected implications for the study of post-transcriptional modifications not only in rRNAs, but also in other naturally occurring long RNAs such as long mRNAs and lncRNAs. Our findings suggest the potential of RTX to not only reverse-transcribe these RNAs in their entirety but, but also to aid in the initial discovery of post-transcriptional modification of these molecules by analyzing the mutation spectrum at each position using deep sequencing. This may serve as the initial step in high-throughput pipelines to discover and characterize these modifications in naturally occurring long RNAs.

4. Materials and methods

4.1 Purification of RTX

pET_RTX_(exo-) (Addgene #102786) was transformed into BL21 (DE3) cells and inoculated in LB overnight. Cells were diluted at a ratio of 1:250 and induced at mid-log phase with 1 mM IPTG. Protein was expressed at 18°C overnight. Cells were pelleted at 5000 × *g* for 10 min and resuspended in Buffer A (10 mM phosphate, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol, pH 7). Cells were lysed in an Avestin Emulsiflex-B15 homogenizer and benzonase (Millipore) was added and incubated for 1 h at 37°C to degrade cellular nucleic acids. Cell lysates were then heated at 85°C for 25 min, cooled on ice and cell debris was pelleted. Cleared cell lysate was incubated with Qiagen nickel-NTA resin for 1 hour at 4°C. Resin was washed 5 times with wash buffer (50 mM NaH$_2$PO$_4$, 300 mM NaCl, 6 mM BME, pH 8) with 10 mM imidazole and 3 times with 20 mM imidazole. Protein was eluted using the same buffer composition plus 200 mM imidazole. Purified protein was dialyzed into storage buffer (50 mM Tris-HCl, 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.1% Tween20, 50% glycerol, pH 8.0), concentrated to 0.4 mg/ml using Centriprep centrifugal filters, 3000 MWCO and stored in aliquots at −80°C.

4.2 Purification of ribosomal RNA

Lysate generated from MRE600 cells harvested at OD600 of 0.5 was centrifuged for 30 min at 30 000 × g to remove cellular debris. Supernatant was layered in a 1:1 volumetric ratio on a high-salt sucrose cushion containing 20 mM Tris–HCl pH 7.2, 500 mM NH$_4$Cl, 10 mM MgCl$_2$, 0.5 mM EDTA, 6 mM BME and 37.7% sucrose in a Ti70 ultracentrifuge tube and centrifuged at 4°C and 90 000 × g overnight. The next morning the supernatant was removed and spun at 4°C and 150 000 × g for an additional 3 h. The supernatant was removed, and the remaining pellet, consisting primarily of 70S ribosomes, was gently washed with Buffer C [10 mM Tris-OAc (pH = 7.5 at 4°C), 60 mM NH$_4$Cl, 7.5 mM Mg(OAc)$_2$, 0.5 mM EDTA, 2 mM DTT] until the pellet was glassy. This pellet was then resuspended in Buffer C for at least 2 h shaking at 4°C. To precipitate RNA, two volumes of glacial acetic acid were added to the sample and incubated for 10 min at 4°C.

The sample was then centrifuged at 16 000 × g for 30 min to pellet rRNA. Pelleted rRNA resuspended in nuclease-free water and purified using a Qiagen RNEasy Mini Kit, eluted using nuclease-free water and stored in aliquots at −80°C.

4.3 Reverse transcription PCR using RTX

RT-PCR using RTX was carried out in 1× RTX buffer [60 mM Tris-HCl (pH 8.4), 25 mM (NH$_4$)$_2$SO$_4$, 10 mM KCl, 200 µM dNTPs, 1 mM MgSO$_4$, 1 M betaine, 0.4 µg RTX exo-]. Initially, the cycling conditions indicated in the original manuscript were utilized: 68°C—30 min, 35× (95°C—30 s, 68°C—30 s/kb). Following optimization, the cycling parameters were adjusted to 2 min/kb for the 23S rRNA. Following RT-PCR, 5 µl of each reaction was run on a 1% agarose gel in TAE running buffer to assess success of the RT-PCR reaction.

4.4. Two enzyme RT-PCR using RTX and Q5 polymerase

One-step mixes featuring either RTX buffer or Q5 buffer were attempted, with poor results from both buffer conditions, likely owing to the buffers substantially different pH and salt composition. To allow each enzyme to perform under more ideal conditions, we assembled 25 µL RT reactions to amplify the 16S or 23S rRNA in RTX buffer, allowing 1, 3, or 5 long (30 min) extension steps. After completion of the RT-PCR cycles, 5 µL of this reaction was added to a 25 µL Q5 polymerase reaction (NEB #M0491) and cycled using standard reaction conditions. The RTX reactions were also returned to the thermocycler and cycled for an addition 35 cycles using the protocol 35x (95°C –30 sec, 68°C 25 min). 5 µL of each of these reactions were run on a 1% TAE agarose gel to assess product formation.

4.5 Deep sequencing analysis of reverse transcribed cDNA

RT-qPCR of the 23S and 16S rRNAs was performed using only external primers and the protocol: 68°C—30 min, 35× (95°C—30 s, 68°C—X min), where X = 6 min and 2 min, respectively. The 23S cDNA was processed for sequencing using the NEXTNext Ultra II FS DNA Library Prep Kit and barcoded using NEBNEXT Multiplex Oligos for Illumina Primers Sets 1 and 2. Samples were submitted for sequencing at Genewiz using the Illumina HiSeq 4000 platform with 2 × 150 bp paired end reads. Raw sequencing reads were filtered using *sickle* (Available at https://github.com/najoshi/sickle) and merged using PANDAseq (55). Paired reads were then aligned to the reference sequence using BWA (56). SAM files were converted to BAM files and sorted, and pileup files were generated all using SAMtools (57). A table of nucleotide by nucleotide variation was generated using the pileup2acgt function in Sequenza (58). Statistical analysis of these data tables and figure generation was performed using R. For error rate analysis, positions of the 23S rRNA which are polymorphic in MRE600 *E. coli* were excluded from analysis, as heterogeneity at these positions reflects polymorphism and not RTX error rate.

SUPPLEMENTARY DATA

Supplementary Data are available at SYNBIO Online.

for technical advice and Ashty Karim (ORCID: 0000-0002-5789-7715) for feedback on the manuscript.

References

1. Dedkova,L.M. and Hecht,S.M. (2019) Expanding the scope of protein synthesis using modified ribosomes. *J. Am. Chem. Soc.*, 141, 6430–6447.
2. Hammerling,M.J., Krüger,A. and Jewett,M.C. (2019) Strategies for in vitro engineering of the translation machinery. *Nucleic Acids Res.*, 29, 1–16. doi: 10.1093/nar/gkz1011.
3. Chin,J.W. (2017) Expanding and reprogramming the genetic code. *Nature*, 550, 53–60.
4. Arranz-Gibert,P., Vanderschuren,K. and Isaacs,F.J. (2018) Next-generation genetic code expansion. *Curr. Opin. Chem. Biol.*, 46, 203–211.
5. Tharp,J.M., Krahn,N., Varshney,U. and Söll,D. (2020) Hijacking translation initiation for synthetic biology. *ChemBioChem*, 21, 1387–1396.
6. Liu,C.C. and Schultz,P.G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.*, 79, 413–444.
7. Rogers,J.M. and Suga,H. (2015) Discovering functional, non-proteinogenic amino acid containing, peptides using genetic code reprogramming. *Org. Biomol. Chem.*, 13, 9353–9363.
8. Katoh,T. and Suga,H. (2018) Ribosomal incorporation of consecutive β-amino acids. *J. Am. Chem. Soc.*, 140, 12159–12167.
9. Lee,J., Torres,R., Kim,D.S., Byrom,M., Ellington,A.D. and Jewett,M.C. (2020) Ribosomal incorporation of cyclic β-amino acids into peptides using in vitro translation. *Chem. Commun.*, 56, 5597–5600.
10. Ohshiro,Y., Nakajima,E., Goto,Y., Fuse,S., Takahashi,T., Doi,T. and Suga,H. (2011) Ribosomal synthesis of backbone-macrocyclic peptides containing γ-amino acids. *ChemBioChem*, 12, 1183–1187.
11. Tsiamantas,C., Kwon,S., Rogers,J.M., Douat,C., Huc,I. and Suga,H. (2020) Ribosomal incorporation of aromatic oligoamides as peptide sidechain appendages. *Angew. Chem. Int. Ed.*, 59, 4860–4864.
12. Katoh,T. and Suga,H. (2020) Ribosomal elongation of cyclic γ-amino acids using a reprogrammed genetic code. *J. Am. Chem. Soc.*, 142, 4965–4969.
13. Lee,J., Schwarz,K.J., Kim,D.S., Moore,J.S. and Jewett,M.C. (2020) Ribosome-mediated polymerization of long chain carbon and cyclic amino acids into peptides in vitro. *Nat. Commun.*, 11, 4304.
14. Goto,Y., Murakami,H. and Suga,H. (2008) Initiating translation with D-amino acids. *RNA*, 14, 1390–1398.
15. Katoh,T., Tajima,K. and Suga,H. (2017) Consecutive elongation of D-amino acids in translation. *Cell Chem. Biol.*, 24, 46–54.
16. Lee,J., Schwieter,K.E., Watkins,A.M., Kim,D.S., Yu,H., Schwarz,K.J., Lim,J., Coronado,J., Byrom,M. and Anslyn,E.V. (2019) Expanding the limits of the second genetic code with ribozymes. *Nat. Commun.*, 10, 1–12.
17. Ad,O., Hoffman,K.S., Cairns,A.G., Featherston,A.L., Miller,S.J., Söll,D. and Schepartz,A. (2019) Translation of diverse aramid- and 1,3-dicarbonyl-peptides by wild type ribosomes in vitro. *ACS Cent. Sci.*, 5, 1289–1294.
18. Kawakami,T., Ogawa,K., Hatta,T., Goshima,N. and Natsume,T. (2016) Directed evolution of a cyclized peptoid-peptide chimera against a cell-free expressed protein and proteomic profiling of the interacting proteins to create a protein-protein interaction inhibitor. *ACS Chem. Biol.*, 11, 1569–1577.
19. Torikai,K. and Suga,H. (2014) Ribosomal synthesis of an amphotericin-B inspired macrocycle. *J. Am. Chem. Soc.*, 136, 17359–17361.
20. Kawakami,T., Ishizawa,T. and Murakami,H. (2013) Extensive reprogramming of the genetic code for genetically encoded synthesis of highly N-alkylated polycyclic peptidomimetics. *J. Am. Chem. Soc.*, 135, 12297–12304.
21. Katoh,T., Sengoku,T., Hirata,K., Ogata,K. and Suga,H. (2020) Ribosomal synthesis and de novo discovery of bioactive foldamer peptides containing cyclic β-amino acids. *Nat. Chem.*, 12, 1081–1088.
22. Katoh,T. and Suga,H. (2020) Ribosomal elongation of aminobenzoic acid derivatives. *J. Am. Chem. Soc.*, 142, 16518–16522.
23. Goto,Y., Katoh,T. and Suga,H. (2011) Flexizymes for genetic code reprogramming. *Nat. Protoc.*, 6, 779–790.
24. Goto,Y. and Suga,H. (2009) Translation initiation with initiator tRNA charged with exotic peptides. *J. Am. Chem. Soc.*, 131, 5040–5041.
25. Oba,M. (2019) Cell-penetrating peptide foldamers: drug-delivery tools. *ChemBioChem*, 20, 2041–2045.
26. Rogers,J.M., Kwon,S., Dawson,S.J., Mandal,P.K., Suga,H. and Huc,I. (2018) Ribosomal synthesis and folding of peptide-helical aromatic foldamer hybrids. *Nat. Chem.*, 10, 405–412.
27. Wang,K., Neumann,H., Peak-Chew,S.Y. and Chin,J.W. (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.*, 25, 770–777.
28. Rackham,O. and Chin,J.W. (2005) A network of orthogonal ribosome-mRNA pairs. *Nat. Chem. Biol.*, 1, 159–166.
29. Neumann,H., Wang,K., Davis,L., Garcia-Alai,M. and Chin,J.W. (2010) Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature*, 464, 441–444.
30. Maini,R., Chowdhury,S.R., Dedkova,L.M., Roy,B., Daskalova,S.M., Paul,R., Chen,S. and Hecht,S.M. (2015) Protein synthesis with ribosomes selected for the incorporation of β-amino acids. *Biochemistry*, 54, 3694–3706.
31. Dedkova,L.M., Fahmi,N.E., Golovine,S.Y. and Hecht,S.M. (2003) Enhanced D-amino acid incorporation into protein by modified ribosomes. *J. Am. Chem. Soc.*, 125, 6616–6617.
32. Dedkova,L.M., Fahmi,N.E., Golovine,S.Y. and Hecht,S.M. (2006) Construction of modified ribosomes for incorporation of D-amino acids into proteins. *Biochemistry*, 45, 15541–15551.
33. Orelle,C., Carlson,E.D., Szal,T., Florin,T., Jewett,M.C. and Mankin,A.S. (2015) Protein synthesis by ribosomes with tethered subunits. *Nature*, 524, 119–124.

34. Schmied,W.H., Tnimov,Z., Uttamapinant,C., Rae,C.D., Fried,S.D. and Chin,J.W. (2018) Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature*, 564, 444–448.

35. Fried,S.D., Schmied,W.H., Uttamapinant,C. and Chin,J.W. (2015) Ribosome subunit stapling for orthogonal translation in E. coli. *Angew. Chem. Int. Ed.*, 54, 12791–12794.

36. Carlson,E.D., d'Aquino,A.E., Kim,D.S., Fulk,E.M., Hoang,K., Szal,T., Mankin,A.S. and Jewett,M.C. (2019) Engineered ribosomes with tethered subunits for expanding biological function. *Nat. Commun.*, 10, 3920.

37. Aleksashin,N.A., Leppik,M., Hockenberry,A.J., Klepacki,D., Vázquez-Laslop,N., Jewett,M.C., Remme,J. and Mankin,A.S. (2019) Assembly and functionality of the ribosome with tethered subunits. *Nat. Commun.*, 10, 930.

38. Cochella,L. and Green,R. (2004) Isolation of antibiotic resistance mutations in the rRNA by using an in vitro selection system. *Proc. Natl. Acad. Sci. USA*, 101, 3786–3791.

39. Hammerling,M.J., Fritz,B.R., Yoesep,D.J., Kim,D.S., Carlson,E.D. and Jewett,M.C. (2020) In vitro ribosome synthesis and evolution through ribosome display. *Nat. Commun.*, 11, 1108.

40. Lu,T., Stroot,P.G. and Oerther,D.B. (2009) Reverse transcription of 16S rRNA to monitor ribosome-synthesizing bacterial populations in the environment. *Appl. Environ. Microbiol.*, 75, 4589–4598.

41. Tsuji,H., Matsuda,K. and Nomoto,K. (2018) Counting the countless: bacterial quantification by targeting rRNA molecules to explore the human gut microbiota in health and disease. *Front. Microbiol*, 9, 1417.

42. Matsuda,K., Tsuji,H., Asahara,T., Kado,Y. and Nomoto,K. (2007) Sensitive quantitative detection of commensal bacteria by rRNA-targeted reverse transcription-PCR. *Appl. Environ. Microbiol*, 73, 32–39.

43. Matsuda,K., Tsuji,H., Asahara,T., Takahashi,T., Kubota,H., Nagata,S., Yamashiro,Y. and Nomoto,K. (2012) Sensitive quantification of *Clostridium difficile* cells by reverse transcription-quantitative PCR targeting rRNA molecules. *Appl. Environ. Microbiol.*, 78, 5111–5118.

44. Sakaguchi,S., Saito,M., Tsuji,H., Asahara,T., Takata,O., Fujimura,J., Nagata,S., Nomoto,K. and Shimizu,T. (2010) Bacterial rRNA-targeted reverse transcription-PCR used to identify pathogens responsible for fever with neutropenia. *J. Clin. Microbiol.*, 48, 1624–1628.

45. Cox,C.J., Kempsell,K.E. and Gaston,J.S.H. (2003) Investigation of infectious agents associated with arthritis by reverse transcription PCR of bacterial rRNA. *Arthritis Res. Ther.*, 5, 1–8.

46. Fujimori,M., Hisata,K., Nagata,S., Matsunaga,N., Komatsu,M., Shoji,H., Sato,H., Yamashiro,Y., Asahara,T., Nomoto,K. et al. (2010) Efficacy of bacterial ribosomal RNA-targeted reverse transcription-quantitative PCR for detecting neonatal sepsis: a case control study. *BMC Pediatr.*, 10, 53.

47. Engstrand,L., Nguyen,A.M.H., Graham,D.Y. and El-Zaatari,F.A.K. (1992) Reverse transcription and polymerase chain reaction amplification of rRNA for detection of Helicobacter species. *J. Clin. Microbiol.*, 30, 2295–2301.

48. Motorin,Y., Muller,S., Behm-Ansmant,I. and Branlant,C. (2007) Identification of modified residues in RNAs by reverse transcription-based methods. *Methods Enzymol.*, 425, 21–53.

49. Cantara,W.A., Crain,P.F., Rozenski,J., McCloskey,J.A., Harris,K.A., Zhang,X., Vendeix,F.A.P., Fabris,D. and Agris,P.F. (2011) The RNA modification database, RNAMDB: 2011 update. *Nucleic Acids Res.*, 39, 195–201.

50. Ellefson,J.W., Gollihar,J., Shroff,R., Shivram,H., Iyer,V.R. and Ellington,A.D. (2016) Synthetic evolutionary origin of a proof-reading reverse transcriptase. *Science (80)*, 352, 1590–1593.

51. Sas-Chen,A. and Schwartz,S. (2019) Misincorporation signatures for detecting modifications in mRNA: not as simple as it sounds. *Methods*, 156, 53–59.

52. Potapov,V., Fu,X., Dai,N., Corrêa,I.R., Tanner,N.A. and Ong,J.L. (2018) Base modifications affecting RNA polymerase and reverse transcriptase fidelity. *Nucleic Acids Res.*, 46, 5753–5763.

53. Schwartz,S. and Motorin,Y. (2017) Next-generation sequencing technologies for detection of modified nucleotides in RNAs. *RNA Biol.*, 14, 1124–1137.

54. Lee,J., Kladwang,W., Lee,M., Cantu,D., Azizyan,M., Kim,H., Limpaecher,A., Gaikwad,S., Yoon,S., Treuille,A., EteRNA Participants. et al. (2014) RNA design rules from a massive open laboratory. *Proc. Natl. Acad. Sci. USA*, 111, 2122–2127.

55. Masella,A.P., Bartram,A.K., Truszkowski,J.M., Brown,D.G. and Neufeld,J.D. (2012) PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics*, 13, 31.

56. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25, 1754–1760.

57. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079.

58. Favero,F., Joshi,T., Marquard,A.M., Birkbak,N.J., Krzystanek,M., Li,Q., Szallasi,Z. and Eklund,A.C. (2015) Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.*, 26, 64–70.