ACS Synthetic Biology
Letter
pubs.acs.org/synthbio

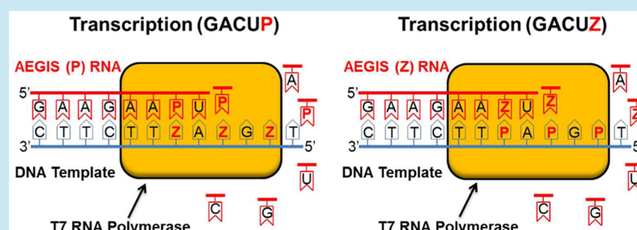
Transcription, Reverse Transcription, and Analysis of RNA Containing Artificial Genetic Components

Nicole A. Leal,[†,‡] Hyo-Joong Kim,[†,‡] Shuichi Hoshika,[†,‡] Myong-Jung Kim,[†,‡] Matthew A. Carrigan,[§] and Steven A. Benner*[,†,‡]

[†]Foundation for Applied Molecular Evolution (FfAME), 720 SW Second Avenue, Suite 201, Gainesville, Florida 32601, United States
[‡]Firebird Biomolecular Sciences LLC, 13709 Progress Boulevard, Box 17, Alachua, Florida 32615, United States
[§]Department of Natural Sciences, Santa Fe College, 3000 NW 83rd Street L209, Gainesville, Florida 32606, United States

Supporting Information

ABSTRACT: Expanding the synthetic biology of artificially expanded genetic information systems (AEGIS) requires tools to make and analyze RNA molecules having added nucleotide "letters". We report here the development of T7 RNA polymerase and reverse transcriptase to catalyze transcription and reverse transcription of xNA (DNA or RNA) having two complementary AEGIS nucleobases, 6-amino-5-nitropyridin-2-one (trivially, **Z**) and 2-aminoimidazo[1,2a]-1,3,5-triazin-4(8H)-one (trivially, **P**). We also report MALDI mass spectrometry and HPLC-based analyses for oligomeric GACU**ZP** six-letter RNA and the use of ribonuclease (RNase) A and T1 RNase as enzymatic tools for the sequence-specific degradation of GACU**ZP** RNA. We then applied these tools to analyze the GACU**ZP** and GACT**ZP** products of polymerases and reverse transcriptases (respectively) made from DNA and RNA templates. In addition to advancing this 6-letter AEGIS toward the biosynthesis of proteins containing additional amino acids, these experiments provided new insights into the biophysics of DNA.



KEYWORDS: *synthetic biology, artificially expanded genetic information systems, transcription, T7 RNA polymerase, reverse transcriptase*

One grand challenge in synthetic biology seeks to produce unnatural genetic systems that do things that natural nucleic acids do, including reproduction, evolution, adaptation, and translation to give encoded biopolymers such as proteins. Pursuing such challenges forces scientists to solve unscripted problems using available theory. If that theory is inadequate, the challenge fails, and fails in a way that cannot be ignored. Thus, synthesis can drive innovation, discovery, and paradigm change in ways that hypothesis-driven research cannot.[1,2]

Efforts to meet this challenge have come from our work to expand the DNA alphabet by shuffling hydrogen bonding groups, adding nucleotides to the standard four that base pair following an expanded set of Watson−Crick pairing rules, referred to as Artificially Expanded Genetic Information Systems (AEGIS).[3−7] Other groups have examined nucleotide analogs having more divergent structures that might perform as replicable units of genetic systems.[8−10] *In vitro*, artificially expanded genetic systems have been shown to indeed be replicable.[4,11−17] Nucleic acids with extra nucleotides have been incorporated into functional molecules that specifically bind target proteins[18] and, as part of a genetic system that fully supports adaptation and evolution *in vitro*, to give DNA molecules that bind to cancer cells.[19] Most recently, the Romesberg group showed that *E. coli* could maintain intracellularly a nonstandard nucleobase pair for up to 9 h, if fed triphosphates and engineered to transport these inside of cells.[20]

Adding nucleotides to the four standard nucleotides in xNA has technological prospects. For example, it offers the potential to increase the genetic "lexicon", the amino acid words that are encoded into translated proteins. While this was first done in these laboratories 20 years ago,[21] recent authors have expressed a renewed enthusiasm for this as a goal of synthetic biology, with some suggesting that some of the more exotic genetic systems might guide translation in living cells.[20]

To make this vision a reality for an alternative genetic system, we must develop (often from scratch) many of the synthetic, manipulative, and analytic tools that are taken for granted by molecular biologists who work with standard nucleic acids. In particular, to use AEGIS components to encode transcribed RNA, polymerases must be found that can biosynthesize RNA containing AEGIS ribonucleosides from DNA templates that contain the encoding AEGIS 2′-deoxyribonucleosides. Further, classical tools to analyze product RNA directly must be developed so that they can be applied to expand genetic alphabets. Further, reverse transcriptases might be developed to analyze the products of that transcription.

Received: June 23, 2014
Published: August 19, 2014

© 2014 American Chemical Society

We recently reported the large-scale synthesis of both the deoxyribo and ribonucleoside forms of AEGIS components carrying, as the replicable heterocycle, 6-amino-5-nitropyridin-2-one, trivially designated d**Z** and r**Z**, and 2-aminoimidazo[1,2-a]-1,3,5-triazin-4(8H)-one, trivially designated as d**P** and r**P**[3,22] (Figure 1). This laid the grounds for a broad effort to develop

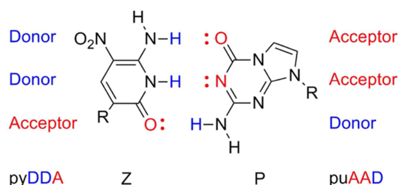


Figure 1. Structures of **Z** and **P** showing their arrangements of hydrogen bonding patterns different from those found on standard nucleobases.

the manipulative and analytical technology needed to support the conversion of six-letter information encoded in DNA to give the corresponding information in encoding RNA molecules. We report these technologies here.

■ RESULTS AND DISCUSSION

T7 RNA polymerase directs the incorporation of rZTP opposite d**P** and rPTP opposite d**Z**. Figure 2 shows transcription reactions using both rZTP and rPTP AEGIS ribonucleoside triphosphates in experiments with templates containing their AEGIS encoding nucleotides. With a fully standard system, the results are clear. Transcription in the presence of all triphosphates gave full-length product without pausing; omission of a needed triphosphate (here, rATP needed to complement a template T at position 4, 6, and following) caused transcription to terminate.

With rZTP as the substrate and its encoding d**P** present in the template at positions 16, 18, and 20 (with position 1 in the DNA template complementary to the G initiating the transcript), full length product was also seen, but with "stutter" bands characterizing transcription efforts that were aborted early in the synthesis.[23] If rZTP was excluded, similar stutter bands were seen, as well as pausing at position 16. These results show that T7 RNA polymerase incorporates rZTP opposite d**P** in a template. Further, multiple r**Z**s could evidently be incorporated into a single RNA product opposite multiple template d**P**s. Finally, T7 RNA polymerase misincorporates a standard nucleotide (perhaps CTP[11]) opposite template d**P** when rZTP is not available.

With rPTP as the substrate and its encoding d**Z** present in the template, when rPTP was absent, very little pausing was seen. This is consistent with the greater facility shown by DNA polymerases to misincorporate dGTP opposite template d**Z** when dPTP is absent.[11] In these cases, dPTP (if it is present) successfully competes with dGTP as a partner for d**Z**, and the correct full length sequence is formed. However, the ease with which dGTP is misincorporated opposite template d**Z** prevents "minus' experiments from being reliable tools to assess fidelity. As this was evidently true for T7 RNA polymerase as well, further analytical tools were developed to provide convincing evidence for successful synthesis of GACU**ZP** RNA by T7 RNA polymerase.

MALDI mass spectrometry was then developed to analyze GACU**ZP** oligonucleotides synthesized by T7 RNA polymerase. Representative spectra for Z3 RNA, (5′-ppp- GGC AGA

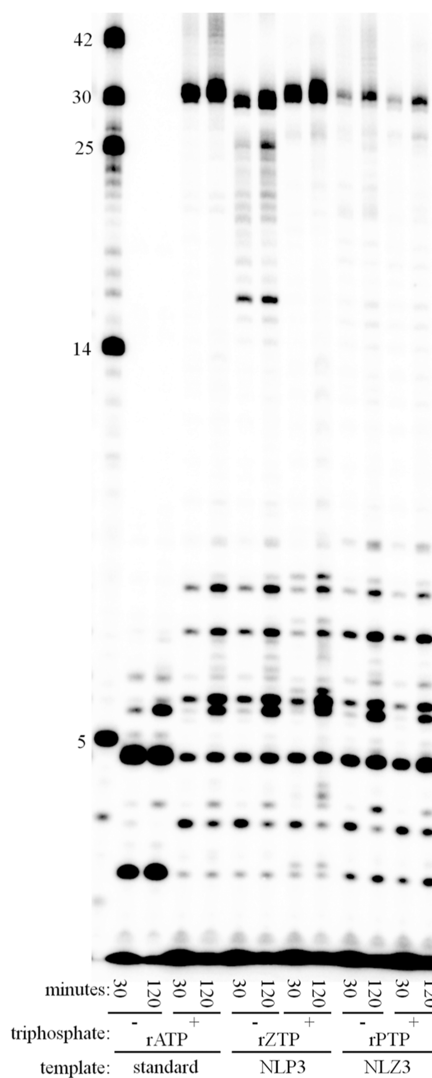


Figure 2. PAGE (20%) showing the RNA products of transcription from the indicated templates containing the AEGIS components at positions 16, 18, and 20 (d**P** in NLP3; d**Z** in NLZ3). Reactions with and without (+ and −) rZTP or rPTP were done in the presence of d**P** or d**Z** containing templates, respectively. The control reaction, labeled "standard" uses only standard nucleotides in a template containing multiple Ts (at positions, 4, 6, 8, 11, 12, and following) with and without (+ and −) complementary rATP. The gel is visualized by phosphorimager detecting P-32 introduced into the RNA products from α-labeled rGTP.

GAG GAA GAA **ZUZ** C**Z**A CAG GCA AGC; calculated mass = 10130.95) and P3 RNA, 5′-ppp-GGC AGA GAG GAA GAA **PUP** C**P**A CAG GCA AGC; calculated mass = 10119.01) are shown in Figure 3. In addition to showing evidence for the expected full length transcript (N) product, the MALDI mass spectrum also indicate "N+1" peaks. T7 RNA polymerase has been shown to add one or more nontemplated nucleotides to the 3′-end of transcripts and is referred to as N+1 activity, which results in heterogeneity of transcription products.[24,25]

Separately, we developed a degradative analytical tool to estimate the total composition of the transcribed product RNA molecules. Here, the oligoribonucleotide was completely converted to 5′-nucleotides by P1 nuclease treatment, followed by dephosphorylation using calf intestinal phosphatase. The products were analyzed by reversed phase HPLC. The number

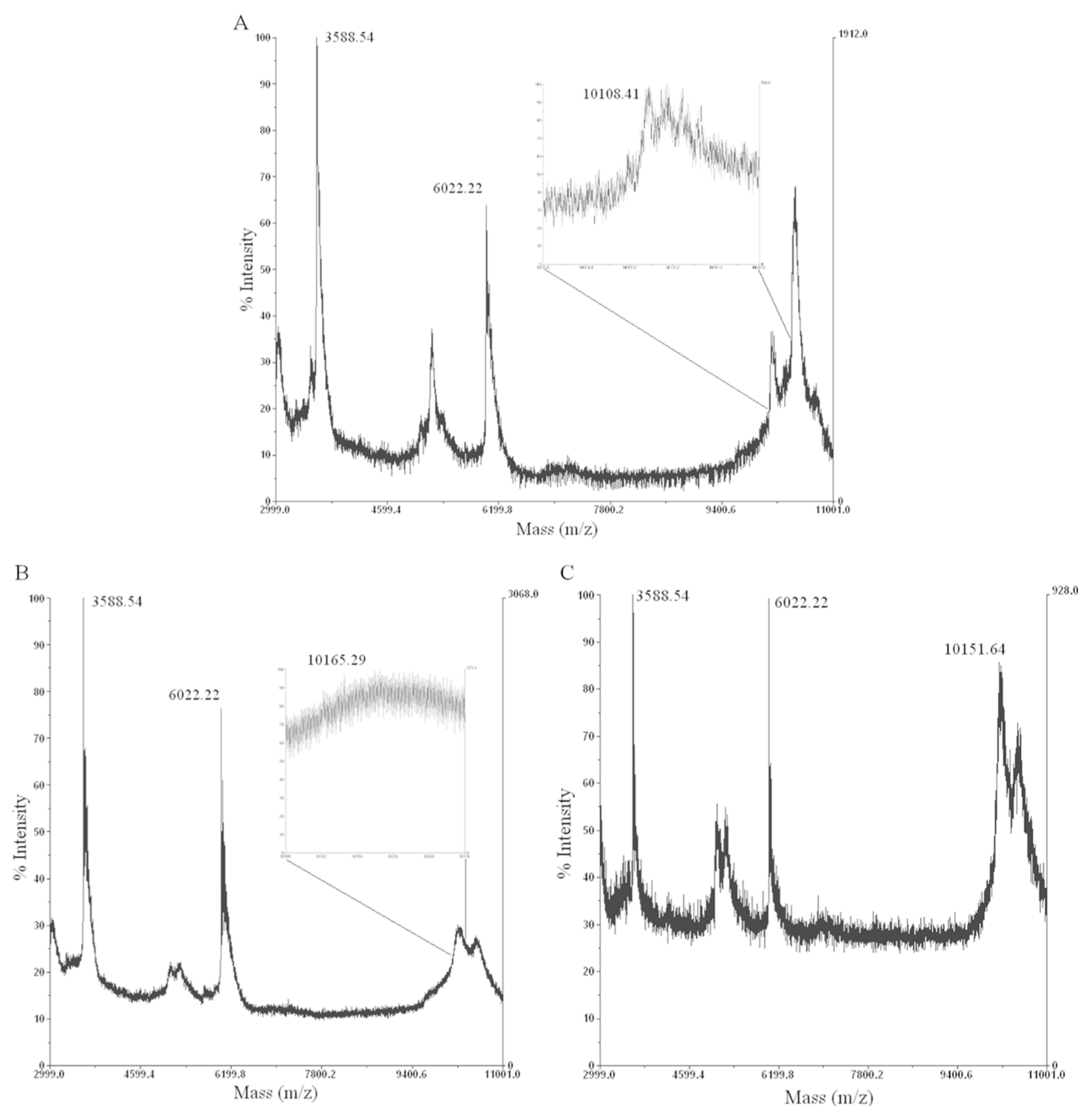


Figure 3. MALDI TOF spectra of oligomeric RNA generated on an Applied Biosystems Voyager system 1102. (A) Standard transcript with a calculated molecular weight (calc. MW) of 10103 and found MW of 10108. (B) P3 transcript with a calc. MW of 10119 and a found MW of 10165. (C) Z3 transcript with a calc. MW of 10131 and a found MW of 10151.

of nucleosides was determined by the comparison of the area of each nucleoside peak to the area of G (Supporting Information, Figures S1−S5). These two analytical methods are complementary. The first (MALDI-MS) establishes the successful formation of oligomeric product but can miss minor products. The second (HPLC) can assess the overall success of the incorporation of nonstandard AEGIS component, even though it does not demonstrate oligomeric product formation. These analyses are, of course, qualitative and cannot make a precise, quantitative statement about the fidelity of incorporation of the non-natural nucleotide when all triphosphates are present.

We then developed an enzymatic tool to confirm that information encoded by a DNA template was indeed captured in a product RNA molecule. This involved the reverse transcription of product RNA molecules back to DNA molecules. Separate work has already shown that GACT**ZP**-containing DNA molecules can be PCR amplified and, subsequently, sequenced by deep sequencing instruments.[11]

SuperScript III (Invitrogen) reverse transcriptase was used to reverse transcribe RNA. Initial results are shown in Figure 4. With the template built from only standard nucleotides, a "minus" experiment lacking dATP to complement rU in the RNA template showed substantial pausing at the site where the template rU was present. However, with prolonged incubation, reverse transcriptase either mismatched or skipped the uncomplemented nucleotide in the template to produce a long product.

With the AEGIS templates, substantial amounts of pausing were observed in "minus" experiments that lacked the complementary AEGIS triphosphate, with full-length product obtained, presumably by mismatching a standard nucleotide opposite the AEGIS complement. As before, the raw data suggest that this mismatching is easier when r**Z** is in the templates, and less easy when r**P** is in the template. Nevertheless, in both cases, the presence of the AEGIS complement produces full product with considerably less pausing, suggesting that the reverse transcriptase is able to synthesize duplexes containing both the **Z·P** and **P·Z** nucleobase pairs. Further, the results when authentic synthetic RNA molecules were used as templates for reverse transcriptase were quite similar to those when the T7 RNA polymerase transcripts were used as the templates for the reverse

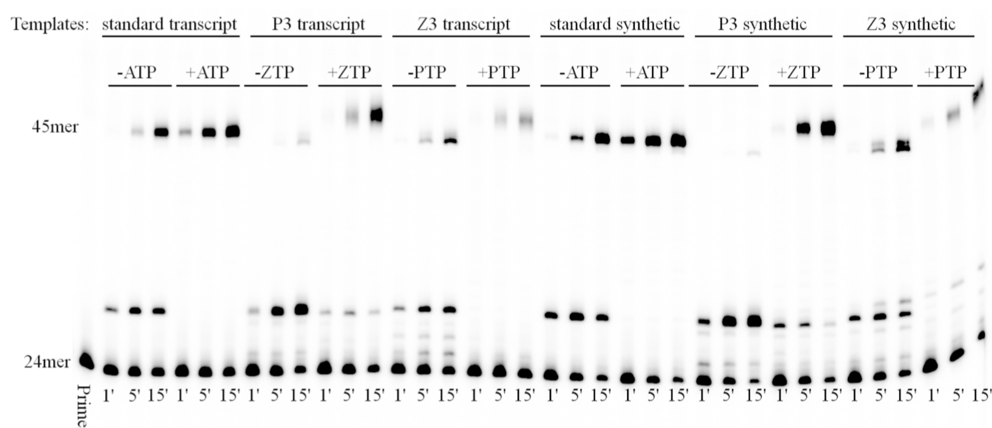


Figure 4. PAGE (12%) showing the DNA products obtained by reverse transcription (pH 8.3, measured at 25 °C, reaction at 37 °C) of transcribed and synthetic RNA in the absence (−) and presence (+) of deoxyribotriphosphates that complement the target AEGIS ribonucleotides at positions 2, 4, and 6. Demonstrating the incorporation of both d**Z** and d**P** at the expected sites with pausing (essentially complete after 1 min incubation, with template r**Z** more easily complemented by a standard nucleotide mismatch than d**P**). The behavior of the synthetic RNA templates paralleled closely the behavior as a template of the RNA made by transcription using T7 RNA polymerase. Control uses a template containing a U at position 5, without and with complementary dATP.

transcription reaction. This comparison suggests that the preponderance of the transcripts have the structure identical to the authentic synthetic RNA.

We then sought to adapt 6-letter RNA to some classical tools originally developed a half-century ago to analyze standard RNA. One classical tool fragments RNA under alkaline conditions. The products of alkaline hydrolysis have a free 5′-OH group, with the linking phosphate ending up at the 2′- or 3′-positions, after opening of a 2′,3′-cyclic phosphate intermediate.

Alkaline hydrolysis was easily applied to GACU**ZP** RNA. Gel analysis of the product mixture gave a "ladder" (Figure 5). Since this ladder could be "counted", it was useful for analyzing the products of site-specific cleavage of GACU**ZP** RNA.

T1 nuclease was also classically used to analyze oligomeric single stranded RNA. This nuclease cleaves specifically at the 3′-ends of unpaired G generating free 5′-OH group and 2′- or 3′-phosphates. Results of digestion by T1 of several synthetic and transcript RNA molecules are shown (respectively) in Figures 5 and 6. T1 nuclease gave the expected products with all standard RNA. Interestingly, despite its general structural similarity to G, **P** in an RNA template was not recognized as a substrate by T1 nuclease. In addition to providing a useful analytical tool, this suggests that T1 uses N1 and N7 of G as recognition sites, a suggestion consistent with crystallographic data suggesting that the epsilon oxygen of Glu46 contacts N1 and the backbone NH of Asn43 contacts N7 on guanosine.[26] Perhaps as expected, r**Z** was not recognized as a cleavage site by T1 nuclease.

Ribonuclease (RNase) A is another enzyme classically used to degrade RNA. RNase A prefers, however, to cleave after (on the 3′-ends of) pyrimidine sites, a specificity attributed to the size of its binding pocket.[27] RNase A digestion gives the expected fragment pattern with standard nucleotides. Interestingly, RNase A also recognizes r**Z** as a site for cleavage. Thus, the standard C and Z3 RNA molecules (where three rCs had been replaced by three r**Z**s) gave very similar digestion patterns.

In all cases, the authentic synthetic RNA and the transcribed RNA behaved the same in digestion reactions with T1 and RNase A. This again suggests that the structures of the principal components of the synthetic and transcribed RNA are the same. This is especially evident after considering that the synthetic RNA contained a few minor impurities that generated few minor bands in the unreacted (UR) lanes. The intensities of these bands must be subtracted from the bands in other lanes to see the extent of the correspondence.

These results suggest some interesting conclusions about the biophysics of DNA built from artificially expanded genetic information systems. For example, with respect to mismatches in DNA, standard G is more easily mismatched against AEGIS **Z** than standard C is mismatched against AEGIS **P**. This was also observed in this work.

In DNA, this asymmetry has been interpreted in terms of the pK_a of various protonated and deprotonated forms of the nonstandard nucleobases.[28] As an isolated nucleoside, **Z** deprotonates with a p$K_a \approx 7.8$. In its deprotonated form, **Z** is strictly complementary (in the Watson−Crick sense) to G. Correspondingly, as an isolated nucleoside, the **P** heterocycle protonates with a p$K_a \approx 5.6$; protonated **P** is a strict Watson−Crick complement of C. With the T7 RNA polymerase and SuperScript III reverse transcriptase reactions being run at a nominal pH of 7.8 and 8.3, respectively, we expect that the deprotonated **Z**·G mismatch should be more effective in supporting duplex synthesis than the protonated **P**·C mismatch. These experiments provide for synthetic biologists a tool-box of resources that allows them to analyze RNA molecules that contain two complementary AEGIS components, **Z** and **P**, throughout the sequence of an oligomeric RNA. Further, they provide synthetic biologists the tools needed to make AEGIS RNA from AEGIS DNA, and AEGIS DNA from AEGIS RNA.

This sets the stage for the next step in the development of an AEGIS synthetic biology, including the use of DNA containing extra codons based on the AEGIS expanded alphabet to encode mRNA and tRNA that might increase the number of amino acids within the protein lexicon. Therefore, these results move the ball one step further toward a fully functional synthetic biology, including expanded genetic translation systems operating *in vivo*.

■ METHODS

Transcription Using Standard, Z-Containing, and P-Containing Templates. Double stranded DNA templates

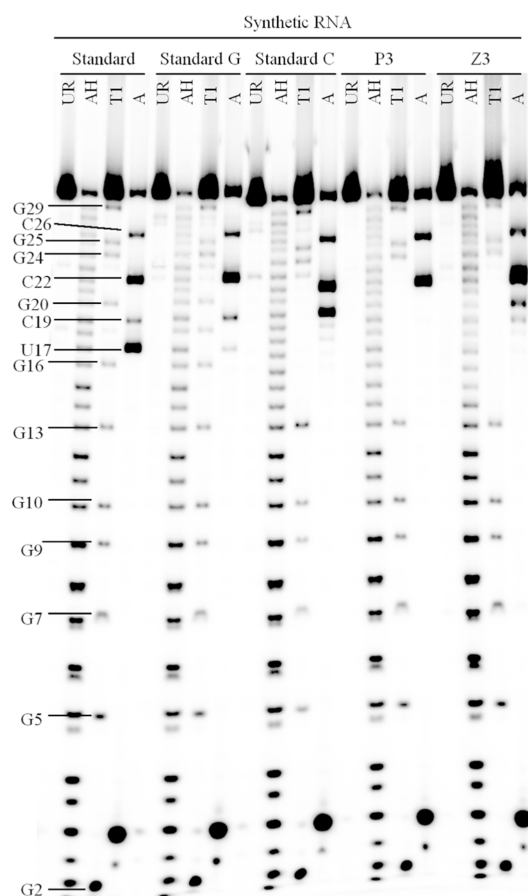


Figure 5. PAGE (20%) showing the RNase degradation profile of synthetic RNA. Lanes are labeled: UR, unreacted; AH, alkaline hydrolysis; T1, denaturing T1digestion; and A, RNase A digestion. RNase T1 is sequence specific for single-stranded RNAs. It cleaves 3′-end of unpaired G residues. RNase A is an endoribonuclease that cleaves single stranded RNA at the 3′end of C and U residues. Standard RNA sequence: 5′-HO-GGC AGA GAG GAA GAA **G**U**A** C**G**A CAG GCA AGC. The other sequences vary at positions 16, 18, and 20 (underlined in the standard RNA sequence) by substituting these positions with a rG, rC, r**P**, or r**Z** for Standard G, Standard C, P3, or Z3, respectively.

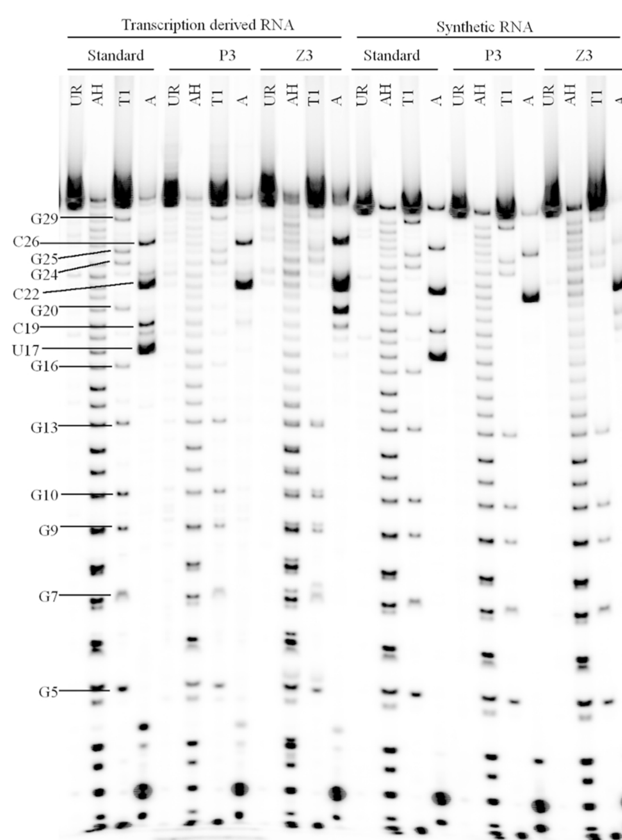


Figure 6. PAGE (20%) showing the degradation profile of transcribed and synthetic RNA. Lanes are labeled: UR, unreacted; AH, alkaline hydrolysis; T1, denaturing T1 digestion; and A, RNase A digestion. Standard RNA sequence: 5′-GGC AGA GAG GAA GAA **G**U**A** C**G**A CAG GCA AGC. The other sequences vary at positions 16, 18, and 20 (underlined in the standard RNA sequence) by substituting these positions with a rG, rC, r**P**, or r**Z** for Standard G, Standard C, P3, or Z3, respectively. Transcripts contain a 5′-triphosphate while synthetic RNA has a 5′−OH. T1 cannot digest r**Z** or r**P** while RNase A is able to digest r**Z** but not r**P**.

were prepared by independently combining equimolar ratios of top strand (NLT1:5′-GCGTAATACGACTCACTATAGG-3′) and bottom strand (NL standard: 3′-CGC ATT ATG CTG AGT GAT ATC CGT CTC TCC TTC TTC ATG CTG TCC GTT CG-5′, NLP3: 3′-CGC ATT ATG CTG AGT GAT AT**C** CGT CTC TCC TTC TT**P** A**P**G **P**TG TCC GTT CG-5′, or NLZ3: 3′-CGC ATT ATG CTG AGT GAT AT**C** CGT CTC TCC TTC TT**Z** A**Z**G **Z**TG TCC GTT CG-5′) in 20 mM NaCl, 40 mM Tris pH 7.8, 6 mM $MgCl_2$, 2 mM spermidine, and 10 mM dithiothreitol, heating to 95 °C for 1 min followed by cooling to room temperature at 0.1 °C/s. This produced a double stranded T7 RNA polymerase promoter region (underlined in the sequences above) followed by a transcriptional start site and three AEGIS bases upstream separated by a standard base.

Transcriptions (20 μL) contained 40 mM Tris (pH 7.8), 20 mM NaCl, 6 mM $MgCl_2$, 2 mM spermidine, 10 mM dithiothreitol, 1 μCi/μL α^{32}P-GTP (PerkinElmer), T7 RNA polymerase (purified in house used at 1 unit per 20 μL reaction), 0.2 μM template DNA, 500 μM of each standard ribotriphosphate (ATP, GTP, CTP, UTP) and 500 μM of either rZTP or rPTP in assays containing either **P**-containing or **Z**-containing templates, respectively. Transcription initiated with the incorporation of rG at position 1 opposite dC in the template. AEGIS bases are at positions 16, 18, and 20. Reactions were incubated at 37 °C for 30 min or 2 h and were quenched with the addition of 3-fold excess 95% formamide. Samples were resolved on a 20% PAGE and visualized using a Personal Molecular Imager System (Biorad).

Purification of RNA. Transcription reactions (400 μL) were prepared as described above in the absence of α^{32}P-GTP using standard, **Z**-containing or **P**-containing double stranded DNA templates. Reactions were incubated at 37 °C for 16 h followed by phenol chloroform isoamyl alcohol (25:24:1) treatment to remove proteins and ethanol precipitation to recover RNA. Transcription reactions were resolved on a 2 mm thick 12% PAGE, UV shadowed to visualize the major product, which was then excised from the gel. RNA was extracted from the matrix using 0.3 M sodium acetate, recovered by ethanol precipitation and dissolved in nuclease free water. This purified RNA was used in subsequent experiments.

RNA Digestion by P1 Nuclease for HPLC Analysis. Purified transcripts and synthetic RNA were digested with P1

nuclease by incubation at 37 °C for 24 h followed by calf intestinal phosphatase treatment (CIP) by incubation at 37 °C for 24 h. Synthetic RNA (with a 5′-hydroxyl) used in these studies have the following sequences;

Standard: 5′-GGC AGA GAG GAA GAA **GUA CGA** CAG GCA AGC-3′;

Standard G: 5′-GGC AGA GAG GAA GAA **GUG** C**GA** CAG GCA AGC-3′;

Standard C: 5′-GGC AGA GAG GAA GAA **CUC** CCA CAG GCA AGC-3′;

P3:5′-GGC AGA GAG GAA GAA **PUP** C**P**A CAG GCA AGC-3′; and Z3:5′-GGC AGA GAG GAA GAA **ZUZ** C**Z**A CAG GCA AGC-3′. Transcribed RNA used in these studies are the same as listed above however have a 5′-triphosphate instead of a 5′-hydroxyl. The digested RNA was analyzed by reversed phase HPLC (Column: Sunfire C18 5 μm, 3.0 × 150 mm. Eluent: A = 25 mM TEAA, B = CH_3CN. Gradient: 100% A for 5 min, then 93% A and 7% B in 20 min. Flow rate: 0.5 mL/min). Each peak was identified by the comparison of authentic nucleoside from its retention time and UV absorption profile: cytidine, 5.2 min (λ_{max}; 270 nm); uridine, 8.3 min (λ_{max}; 262 nm); r**P** nucleoside, 13.7 min (λ_{max}; 259 nm); inosine, 14.7 min (λ_{max}; 248 nm); guanosine, 15.5 min (λ_{max}; 253 nm); r**Z** nucleoside, 16.6 min (λ_{max}; 382 nm); hypothetical product from A, 19.7 min; adenosine, 20.7 min (λ_{max}; 259 nm). Inosine is the deamination product of adenosine, generated during the degradation sequence through a contaminating adenosine deaminase activity present in the commercial phosphatase. Thus, the total A equals the adenosine plus inosine peaks (Supporting Information, Figures S1−S5).

MALDI. A 10 μL solution of matrix-buffer (made from 3-hydroxypicolinic acid (8 mg) and diammonium citrate (1 mg) in water (180 μL)) was mixed with 8 μL of aqueous solution of RNA (concentration of 20 to 40 μM) and 2 μL of internal standard DNA (purchased from Integrated DNA Technologies, T12 oligo, MW = 3588.54 and T20 oligo, MW = 6022.22, each 100 μM). An aliquot (1.5 μL) of the above mixture was then applied to a sample plate and evaporated. The samples were then analyzed using a PerSeptive Voyager DE instrument (20 000 V, negative mode, 700−900 ns delay time).

Reverse Transcription in the Absence and Presence of dATP, dPTP, or dZTP. Reverse transcription primer template complexes were prepared by combining a DNA primer (RT0915A: 5′-^{32}P-$(A)_{15}$GCT TGC CTG (synthesized by IDT)) with 2-fold excess of RNA template transcript (Standard: 3′-CGA ACG GAC AGC AUG AAG AAG GAG AGA CGG-5′; Z3 transcript: 3′-CGA ACG GAC AZC ZUZ AAG AAG GAG AGA CGG-5′ or P3 transcript 3′-CGA ACG GAC APC PUP AAG AAG GAG AGA CGG-5′) heating to 60 °C for 5 min and chilling on ice to eliminate formation of RNA secondary structure. The primer binding site is underlined in sequences listed above. Reverse transcriptase reactions contained a final concentration of 0.025 μM ^{32}P-labeled DNA primer, 0.05 μM RNA templates, 0.25 mM dNTPs, with and without dATP, dPTP, or dZTP, 50 mM Tris-HCl (pH 8.3 at 25 °C), 75 mM KCl, 3 mM $MgCl_2$, 100 mM DTT, and 50 units of SuperScript III. Reactions were incubated at 37 °C for 1, 5, and 15 min and quenched with 4-fold excess quench buffer (10 mM EDTA in 95% formamide). Samples were resolved on a 12% PAGE and visualized using a Personal Molecular Imager System.

RNA Labeling, Alkaline Hydrolysis, Denaturing Ribonuclease T1, and RNase A Digestion. RNA (~20 pmol each) was treated with alkaline phosphatase (10 units, NEB) in 50 mM Tris-HCl (pH 7.9), 100 mM NaCl, 10 mM $MgCl_2$, and 1 mM dithiothreitol and incubated at 37 °C for 1 h. RNA was then purified by phenol chloroform isoamyl alcohol (25:24:1) extraction and ethanol precipitation. Extracted RNA was 5′ radiolabeled with ^{32}P-γATP and polynucleotide kinase (10 units, 37 °C for 30 min). For alkaline hydrolysis, RNA was incubated at 90 °C for 10 min in 50 mM sodium carbonate buffer pH 9.2. Reactions were quenched with buffer (10 M urea, 15 mM EDTA, and 0.025% SDS, xylene cyanol, and bromophenol blue). For ribonuclease T1 (Ambion) and RNase A (Ambion) digestions, RNA was treated with enzyme (1 unit) in 250 μM sodium citrate pH 5.0, 6 M urea, 9 mM EDTA at 55 °C for 15 min. Reactions were quenched with the addition of loading buffer (10 M urea, 15 mM ethylenediamintetetraacetic acid (EDTA), and 0.025% sodium dodecyl sulfate (SDS), xylene cyanol, and bromophenol blue). Reactions were resolved on a 20% PAGE and visualized using a Personal Molecular Imager System.

■ ASSOCIATED CONTENT

Supporting Information

HPLC traces and nucleoside analyses of degraded RNA (Figures S1−S5). This material is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +1 352 271 7005. Fax: +1 352 271 7076. Email: manuscripts@ffame.org.

Author Contributions

N.A.L. designed the experiments and performed the enzymology, H.J.K. performed the analytical chemistry and provided the **Z**, S.H. provided oligonucleotide synthesis, M.J.K. provided the **P**, and M.A.C. provided early polymerase work and helpful discussion. N.A.L. and H.J.K. contributed equally to this work.

Funding

The Defense Threat Reduction Agency (HDTRA1-13-1-0004), DARPA (N66001-12-C-4211), Nucleic Acids Licensing, and Firebird Biomolecular Sciences (under a contract from the Office of the Secretary of Defense, W911NF-12-C-0059) provided support for this work.

Notes

The authors declare the following competing financial interest:Various authors are listed as inventors on various patents and patent applications relating to this work.

■ ACKNOWLEDGMENTS

We are indebted to support from the Defense Threat Reduction Agency (HDTRA1-13-1-0004), DARPA Foundries (N66001-12-C-4211), Nucleic Acids Licensing, and Firebird Biomolecular Sciences (under a contract from the Office of the Secretary of Defense, W911NF-12-C-0059) for support of this work.

■ ABBREVIATIONS

Z, Nucleoside carrying a 6-amino-5-nitropyridin-2-one; **P**, Nucleoside carrying a 2-aminoimidazo[1,2-a]-1,3,5-triazin-4(8H)-one heterocycle; AEGIS, Artificially Expanded Genetic Information System

REFERENCES

(1) Malaterre, C. (2013) Synthetic biology and synthetic knowledge. *Biol. Theory 8*, 346−356.

(2) Benner, S. A. (2013) Synthesis as a route to knowledge. *Biol. Theory 8*, 357−367.

(3) Yang, Z., Hutter, D., Sheng, P., Sismour, A. M., and Benner, S. A. (2006) Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern. *Nucleic Acids Res. 34*, 6095−6101.

(4) Yang, Z., Sismour, A. M., Sheng, P., Puskar, N. L., and Benner, S. A. (2007) Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Res. 35*, 4238−4249.

(5) Yang, Z., Chen, F., Chamberlin, S. G., and Benner, S. A. (2010) Expanded genetic alphabets in the polymerase chain reaction. *Angew. Chem., Int. Ed. Engl. 49*, 177−180.

(6) Switzer, C. Y., Moroney, S. E., and Benner, S. A. (1993) Enzymic recognition of the base pair between isocytidine and isoguanosine. *Biochemistry 32*, 10489−10496.

(7) Benner, S. A. (2004) Understanding nucleic acids using synthetic chemistry. *Acc. Chem. Res. 37*, 784−797.

(8) Hirao, I. (2006) Unnatural base pair systems for DNA/RNA-based biotechnology. *Curr. Opin. Chem. Biol. 10*, 622−627.

(9) Henry, A. A., and Romesberg, F. E. (2003) Beyond A, C, G, and T: Augmenting nature's alphabet. *Curr. Opin. Chem. Biol. 7*, 727−733.

(10) Mitsui, T., Kimoto, M., Sato, A., Yokoyama, S., and Hirao, I. (2003) An unnatural hydrophobic base, 4-propynylpyrrole-2-carbaldehyde, as an efficient pairing partner of 9-methylimidazo[(4,5)-b]pyridine. *Bioorg. Med. Chem. Lett. 13*, 4515−4518.

(11) Yang, Z., Chen, F., Alvarado, J. B., and Benner, S. A. (2011) Amplification, mutation, and sequencing of a six-letter synthetic genetic system. *J. Am. Chem. Soc. 133*, 15105−15112.

(12) Hirao, I., Kimoto, M., Mitsui, T., Fujiwara, T., Kawai, R., Sato, A., Harada, Y., and Yokoyama, S. (2006) An unnatural hydrophobic base pair system: Site-specific incorporation of nucleotide analogs into DNA and RNA. *Nat. Methods 3*, 729−735.

(13) Hirao, I., Mitsui, T., Kimoto, M., and Yokoyama, S. (2007) An efficient unnatural base pair for PCR amplification. *J. Am. Chem. Soc. 129*, 15549−15555.

(14) Malyshev, D. A., Seo, Y. J., Ordoukhanian, P., and Romesberg, F. E. (2009) PCR with an expanded genetic alphabet. *J. Am. Chem. Soc. 131*, 14620−14621.

(15) Malyshev, D. A., Pfaff, D. A., Ippoliti, S. I., Hwang, G. T., Dwyer, T. J., and Romesberg, F. E. (2010) Solution structure, mechanism of replication, and optimization of an unnatural base pair. *Chemistry (Easton) 16*, 12650−12659.

(16) Piccirilli, J. A., Krauch, T., Moroney, S. E., and Benner, S. A. (1990) Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature 343*, 33−37.

(17) Chelliserrykattil, J., Lu, H., Lee, A. H., and Kool, E. T. (2008) Polymerase amplification, cloning, and gene expression of benzo-homologous "yDNA" base pairs. *Chembiochem 9*, 2976−2980.

(18) Kimoto, M., Yamashige, R., Matsunaga, K., Yokoyama, S., and Hirao, I. (2013) Generation of high-affinity DNA aptamers using an expanded genetic alphabet. *Nat. Biotechnol. 31*, 453−457.

(19) Sefah, K., Yang, Z., Bradley, K. M., Hoshika, S., Jimenez, E., Zhang, L., Zhu, G., Shanker, S., Yu, F., Turek, D., Tan, W., and Benner, S. A. (2014) *In vitro* selection with artificial expanded genetic information systems. *Proc. Natl. Acad. Sci. U.S.A. 111*, 1449−1454.

(20) Malyshev, D. A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J. M., Correa, I. R., Jr., and Romesberg, F. E. (2014) A semi-synthetic organism with an expanded genetic alphabet. *Nature 509*, 385−388.

(21) Bain, J. D., Switzer, C., Chamberlin, A. R., and Benner, S. A. (1992) Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. *Nature 356*, 537−539.

(22) Kim, H. J., Leal, N. A., Hoshika, S., and Benner, S. A. (2014) Ribonucleosides for an artificially expanded genetic information system. *J. Org. Chem. 79*, 3194−3199.

(23) Martin, C. T., Muller, D. K., and Coleman, J. E. (1988) Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry 27*, 3966−3974.

(24) Kao, C., Zheng, M., and Rudisser, S. (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA 5*, 1268−1272.

(25) Milligan, J. F., Groebe, D. R., Witherell, G. W., and Uhlenbeck, O. C. (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res. 15*, 8783−8798.

(26) Hubner, B., Haensler, M., and Hahn, U. (1999) Modification of ribonuclease T1 specificity by random mutagenesis of the substrate binding segment. *Biochemistry 38*, 1371−1376.

(27) Findlay, D., Herries, D. G., Mathias, A. P., Rabin, B. R., and Ross, C. A. (1962) The active site and mechanism of action of bovine pancreatic ribonuclease. 7. The catalytic mechanism. *Biochem. J. 85*, 152−153.

(28) Yang, Z., Durante, M., Glushakova, L. G., Sharma, N., Leal, N. A., Bradley, K. M., Chen, F., and Benner, S. A. (2013) Conversion strategy using an expanded genetic alphabet to assay nucleic acids. *Anal. Chem. 85*, 4705−4712.