Expanding the structural and functional diversity of RNA: analog uridine triphosphates as candidates for *in vitro* selection of nucleic acids

Narendra K. Vaish[1,2], Andrew W. Fraley[1], Jack W. Szostak[2] and Larry W. McLaughlin[1,*]

[1]Department of Chemistry, Boston College, Chestnut Hill, MA 02467, USA and [2]Howard Hughes Medical Institute, Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA

ABSTRACT

Two analog uridine triphosphates tethering additional functionality, one a primary amino group and the second a mercapto group, were prepared and tested for their compatibility with *in vitro* RNA selection procedures. 5-(3-Aminopropyl)uridine triphosphate (UNH$_2$) as a uridine substitute was a more effective substrate for T7 RNA polymerase than 5-(2-mercapto-ethyl)uridine triphosphate (USH). However, both functioned in transcription assays of 100 nt templates to generate RNA transcripts in quantities sufficient to initiate RNA selection procedures. Transcription of RNA pools with T7 RNA polymerase and UNH$_2$ or USH occurred with efficiencies of 43 and 29%, respectively, of the values obtained for native UTP transcription. In addition, the transcribed RNA containing roughly 25% UNH$_2$ residues exhibited better substrate properties for SuperScript™ II RNase H reverse transcriptase than did RNA transcripts containing ~25% of the USH analog. With either analog, both transcription and reverse transcription proceeded with high fidelity for insertion of the analog residue.

INTRODUCTION

The discovery of catalytic RNA (1,2) has changed the traditional view of nucleic acids from that of only information storage and message transfer agents to one of important catalytic roles in the cell. Since Altman's and Cech's initial discovery and the observation that some RNAs exhibit enhanced cleavage activity (3), several natural RNAs with catalytic motifs have been also been identified (4) and numerous RNA catalytic motifs and aptamers have been selected and/or evolved through *in vitro* selection experiments (5,6). *In vitro* selection or SELEX procedures involve the isolation and amplification of RNA and DNA sequences, which exhibit selected properties, from a large pool of random sequences through repeated cycles. Numerous examples exist where nucleic acids have been selected as aptamers for proteins or small molecules (5–7), as well as for chemical manipulations, including those of

nucleic acid ligases and polymerases (8). Nevertheless, the functional scope of nucleic acids is likely limited since they possess a narrow range of functionality, at least when compared with proteins (9,10). Those functionalities that nucleic acids contain are principally employed in Watson–Crick base pairing, leaving sugar hydroxyls (in the case of RNA) or charged phosphodiesters available to participate in binding or catalysis. A logical step towards the creation of novel aptameric and catalytic nucleic acids would be the incorporation of additional functionality into the common nucleotide triphosphate building blocks and their use for SELEX procedures. Nucleotide analogs could provide sites for metal coordination or new hydrogen bonding interactions or provide electrophiles and nucleophiles, as well as acidic and basic groups to facilitate binding and/or catalysis. Modified nucleic acids could provide novel motifs for ligand-binding pockets with diverse molecular recognition capabilities that would not be present using native RNA/DNA sequences.

Several research groups have focused on the development of modified nucleotides with added functionality as building blocks for *in vitro* selection procedures. The majority of these studies have focused on the modification and selection of DNA, presumably due to its enhanced chemical stability relative to RNA, but RNA plays a more active role in the cell and might be expected to exhibit a greater catalytic potential than DNA. Sakthivel and Barbas synthesized a series of 5-modified dUTP analogs which included both cationic and anionic residues as well as L-histidine (11). Additionally, Helene and co-workers have simultaneously incorporated two modified deoxynucleotides into a DNA transcript (12). Recently an amino-functionalized deoxyuridine was used to select for a DNA-based ATP aptamer (13). Similarly, a DNA aptamer has been selected in which 5-(1-pentynyl)-2′-deoxyuridine was essential for binding (14). RNA cleavage activity has been observed with a DNA sequence containing imidazole-modified uridines (15).

RNA sequences in comparison to DNA sequences may offer a wider variety of folded structural motifs, as is observed in biology. For example, the presence of 2′-hydroxyls may assist in interdomain binding to provide better tertiary structural stability (16). Recently Eaton and co-workers have used a pyridyl-derivatized uridine nucleotide to select a Diels-Alderase ribozyme while an imidazole-modified uridine nucleotide was employed to select an amide synthase ribozyme (10,17–21).

*To whom correspondence should be addressed. Tel: +1 617 552 3622; Fax: +1 617 552 2705; Email: larry.mclaughlin@bc.edu

Both of the selected RNAs were shown to be dependent on the presence of the modified nucleotides for their respective catalytic activities. Other RNA-like building blocks available for selection assays include 2′-deoxy-2′-fluoro- and 2′-deoxy-2′-amino analogs (22,23).

We report here the synthesis of two modified uridine nucleotide triphosphate building blocks: one is 5-(3-aminopropyl)uridine 5′-triphosphate, a UTP derivative tethering a primary amine (UNH$_2$); the second is 5-(2-mercaptoethyl)uridine 5′-triphosphate, in which UTP tethers a mercaptan (USH). These analogs have been examined for their ability to be incorporated into RNA by T7 RNA polymerase, their use as templates for reverse transcriptase and their fidelity with respect to RNA selection protocols.

MATERIALS AND METHODS

Materials

Solvents and reagents were purchased from Aldrich Chemical Co. (Milwaukee, WI) and used without further purification unless otherwise noted. Anion exchange chromatography was performed using a 250 ml glass column packed with Sephadex DEAE-A25 from Pharmacia (Piscataway, NJ). Reverse phase HPLC was performed on a Beckman System Gold with UV detection at 260 nm and a 4.6 × 250 mm stainless steel column packed with Ranin C18 ODS-Hypersil (100 Å, 5 μm). The flow rate of 1.5 ml/min was used with an aqueous solution buffered with 20 mM NaH$_2$PO$_4$ (pH 5.5) and a linear gradient of 0–70% methanol over 1 h. Anion exchange HPLC was performed using a Beckman 421A controller with a Beckman 163 detection unit set to 260 nm and a 4.6 × 100 mm Peek column packed with POROS 20 HQ from Perseptive Biosystems (Framingham, MA). Beckman 114M solvent delivery units were used to obtain a flow rate of 7 ml/min with a gradient of 0–1 M NaCl aqueous solution buffered with 20 mM Tris–HCl (pH 7.0) over 4.5 min. NMR spectra were obtained on 400 and 500 MHz Varian FT-NMR spectrometers using the deuterated solvent as an internal standard. ^{31}P NMR spectra were obtained on a 500 MHz Varian FT-NMR using phosphoric acid as an external standard. Nucleotide triphosphate NMR samples were dissolved in a 50% (v/v) D$_2$O/H$_2$O solution buffered with 10 mmol Tris–HCl (pH 7.0) and 10 mmol EDTA. Mass spectral analysis for nucleoside triphosphates was obtained on a Micromass MALDI TOF SPEC ZE (Boston College Mass Spectrometry Laboratory) and high resolution mass spectrometry was performed by the University of Illinois Mass Spectrometry Laboratory. 2′-Deoxyphosphosphoramidites were purchased from Glen Research (Sterling, VA) and nucleoside-bound controlled pore glass (CPG) supports were purchased from CPG Inc. UV analyses were performed on a Beckman DU 640 spectrophotometer.

[α-^{32}P]CTP (3000 Ci/mmol, 250 μCi) and [γ-^{32}P]ATP (6000 Ci/mmol, 250 μCi) were purchased from ICN (Costa Mesa, CA). T7 RNA polymerase (200 U/μl) was purchased from Ambion (Austin, TX) or prepared from over-producing plasmid pAR1219 according to the method of Zawadzki and Gross (24). SuperScript II RNase H$^-$ Reverse Transcriptase (200 U/μl) was purchased from Life Technologies (Grand Island, NY). *Taq* DNA polymerase, snake venom phosphodiesterase and calf intestinal alkaline phosphatase were purchased from

Boehringer (Mannheim, Germany). The common nucleotide triphosphates were purchased from Pharmacia (Piscataway, NJ) and the TA cloning kit was purchased from Invitrogen (Carlsbad, CA). Polyacrylamide gel electrophoresis was performed in an apparatus from Life Technologies (Grand Island, NY). Gels were imaged in a Phosphorimager 425 from Molecular Dynamics (Sunnyvale, CA).

Methods

5-Propyl trifluoroacetamide uridine (1). 5-Propargyl trifluoroacetamide uridine (0.5 g, 1.36 mmol), prepared according to the reported procedure (25), was dissolved in 200 ml of methanol, and 10% Pd/C (0.20 g) was added. The solution was hydrogenated in a Parr Hydrogenation Apparatus at 85 p.s.i. for 48 h at ambient temperature and the reaction was monitored over time by ^{1}H NMR. Upon completion the reaction mixture was filtered through Celite and the Celite was rinsed with methanol (200 ml). The solvent was removed by rotary evaporation, yielding 0.468 g (92%) of the desired product. No further purification was necessary, however, recrystallization can be accomplished using ethyl alcohol/ethyl ether. UV (ethanol) λ$_{max}$ 260 nm (ε 1.01 × 10^3 l/M cm). ^{1}H NMR (DMSO-d$_6$) δ = 1.63–1.68 (m, 2H), 2.19–2.24 (m, 2H), 3.16–3.21 (m, 2H), 3.54–3.64 (dd, 2H, H5′), 3.83–3.84 (d, 1H, H4′), 3.98 (m, 1H, H2′), 4.05 (d, 1H, H3′), 5.10 (s, 1H, OH), 5.13 (s, 1H, OH), 5.37 (s, 1H, OH), 5.78–5.79 (d, 1H, H1′), 7.76 (s, 1H, H6), 9.45 (s, 1H, NH), 11.34 (s, 1H, NH) p.p.m. ^{13}C NMR (DMSO-d$_6$) δ = 25.2, 28.7, 62.3, 71.3, 74.9, 86.2, 89.2, 114.1, 119.3, 119.4, 138.2, 152.1, 164.9 p.p.m. HRMS: calculated for C$_{14}$H$_{18}$F$_3$N$_3$O$_7$ (M+H$^+$), 398.1175; found 398.1175.

General procedure for nucleotide triphosphate synthesis. The modified nucleoside (0.345 mmol) was dried in a flask under vacuum overnight. Trimethylphosphate (1.9 ml) and Proton Sponge (0.11 g, 1.5 equiv.) were added to the flask and the solution was allowed to stir under N$_2$ cooled by an ice/methanol bath (–5°C). Distilled phosphorus oxychloride (44 μl, 1.5 equiv.) was added and the reaction was allowed to stir for 4 h with cooling. Tributylamine (0.32 ml) and tributylamine pyrophosphate (4.0 ml of a 0.5 M solution in DMF) were added and the reaction was allowed to stir for an additional 45 min with cooling. The reaction was quenched with triethylammonium bicarbonate (0.5 M, 20 ml). The extent of the reaction could be monitored at this point by anion exchange HPLC. The solvents were removed by rotary evaporation and the remaining crude mixture was dissolved in 10 ml of water. The product was purified using a Sephadex DEAE A-25 column with a linear gradient of 0–1 M NaCl buffered with 20 mM Tris–HCl (pH 7.0) (triphosphates eluted at ~0.4 M NaCl) and desalted on a Sephadex G-10 column.

5-(3-Trifluoroacetamidopropyl)uridine 5′-triphosphate (2). Prepared from 137 mg (0.345 mmol) of **1** as described above, 0.221 mmol of **2** was obtained (64% yield), as determined by UV analysis. The sample was then lyophilized and stored at –20°C. ^{31}P NMR (D$_2$O) δ = –20.55 (t), –10.25 (d), –9.75 (d) p.p.m. MALDI-TOF: expected *m/z* 636.98; found *m/z* 638.03.

5-(3-Aminopropyl)uridine 5′-triphosphate (3). Deprotection of the amine was accomplished by treating **2** at 55°C overnight in ammonium hydroxide (23% aq.). Ammonium hydroxide was

then removed *in vacuo* and the deprotected nucleoside triphosphate was dissolved in water. Purity was confirmed by anion exchange HPLC and deprotection was confirmed by treating 1 A_{260} unit of **3** with alkaline phosphatase followed by nucleoside analysis by reverse phase HPLC. MALDI-TOF: expected *m/z* 551.01; found *m/z* 550.81.

*5-(2-t-Butyldithioethyl)uridine 5′-triphosphate (**4**).* 5-(2-*t*-Butyldithioethyl) uridine was prepared according to the reported procedure (26) and then converted to the triphosphate and purified as described above. The product was obtained in 32% yield and was stored as the disulfide in a freezer (–20°C) until further use. Purity was confirmed by anion exchange HPLC. While deprotection of the mercapto group occurred immediately before transcription, deprotection (for conditions see RNA transcription reactions) was optimized at this point by treating 1 A_{260} unit with alkaline phosphatase, followed by nucleoside analysis by reversed phase HPLC. ^{31}P NMR (D$_2$O) –20.76 (t), –10.19 (d), –10.09 (d) p.p.m. MALDI-TOF: expected *m/z* 627.98; found *m/z* 651.17 (M+Na$^+$).

Oligonucleotide syntheses. Oligodeoxynucleotides used as primers or templates were synthesized using conventional phosphoramidite chemistry on an Applied Biosystems 381A DNA synthesizer. After the synthesis was complete the 5′ DMT group was removed and the oligos were cleaved from the solid support with 23% aq. NH$_4$OH at 55°C overnight. The resulting solution was filtered through a 0.22 μM filter and the DNA precipitated with 10 vol *n*-butanol. Oligodeoxynucleotide strands were purified by denaturing PAGE and isolated by passive elution from the gel in water and finally precipitated with 300 mM NaCl and 3 vol ethanol.

RNA transcription reactions. Initial transcription studies were performed using a double-stranded T7 RNA polymerase promoter followed by a single-stranded template (27), which coded for a 5mer and contained one site for analog incorporation (transcript, 5′-GGUCG-3′). The transcript also contained a single site for incorporation of radiolabeled [α-^{32}P]CMP for subsequent imaging after denaturing PAGE. Subsequent transcription reactions of 100mer templates were performed using a double-stranded pool DNA template that contained an arbitrary tract of 60 nt with 25% doping of uridine. For cloning and sequencing studies a single sequence was designed where uridine composed 26% of the sequence with six of these uridines present in the reverse transcriptase primer binding site (100mer transcript, 5′-*GCACACCAAUUACAUUCACU*CAC-GGAUCUCUGGUAACUAACUAUACAAGCCCGAUACC-ACCUUCCUGUAGCUGGGUGCC*CAUCCAUCAUCUACU-ACUCA*-3′; italic at the 5′-terminus indicates the PCR primer binding site and at the 3′-end the site for binding by the reverse transcription primer). Additional templates were designed for the incorporation of USH that contained sites for six uridine residues with none of these residues in the reverse transcriptase primer binding site (transcript, 5′-*GGCACUGCACACACAC-CCGCA*GCAUGCAUGCAUGCAUGCAU*GAGAGAGAGCA-GGAGAAGAG*-3′; italics indicate the reverse transcription primer and PCR primer binding sites).

The polymerization reaction to generate the RNA 5mer contained 40 mM Tris–HCl (pH 8.0), 20 mM MgCl$_2$, 1.0 mM spermidine, 10 mM DTT, 0.01% Triton X-100, 2.0 mM each

nucleoside triphosphates, 5–10 μCi [α-^{32}P]CTP, 1 μM DNA template and 10 U/μl T7 RNA polymerase. For transcription reactions containing a uridine triphosphate analog, uridine triphosphate was omitted from the reaction. The polymerization reaction to generate the RNA 100mer contained 40 mM Tris–HCl (pH 8), 35 mM MgCl$_2$, 1 mM spermidine, 10 mM DTT, 0.01% Triton X-100, 5 mM each nucleoside triphosphates, 5–10 μCi [α-^{32}P]CTP, 200 nM dsDNA template and 400 U/μl T7 RNA polymerase. Reactions were stopped after 3 h by the addition of gel loading buffer [7 M urea, 50 mM EDTA (pH 8.0), 0.05% xylene cyanol and 0.05% bromophenol blue] equal in volume to that of the transcription reaction.

When USH was used for transcription, the *t*-butylthio protecting group was removed just prior to transcription by the addition of 10 mM Tris–HCl (pH 8.2) and 10 mM DTT to the protected 5-(2-*t*-butyldithioethyl)uridine 5′-triphosphate (4 mM) for 10 min at room temperature. The transcription reaction was stopped by the addition of gel loading buffer and DTT was increased to 100 mM (28). Samples were preheated at 95°C for 5 min prior to analysis either using 20 or 10% denaturing PAGE depending upon the length of the transcript. Upon completion of electrophoresis, transcription efficiency was determined by imaging in a phosphorimager for [α-^{32}P]CMP incorporation. Transcripts for use in reverse transcription reactions and analysis by snake venom phosphodiesterase/ reverse phase HPLC were isolated from the gels by passive elution in water and precipitated using 300 mM NaCl and 3 vol absolute ethanol.

Nucleoside analysis of RNA transcripts. An aliquot of 1–2 A_{260} units of an RNA transcript was incubated (37°C) in a 10 μl solution with Tris–HCl (pH 7.6) and MgCl$_2$ with alkaline phosphatase (1 U) and snake venom phosphodiesterase (1 U) for 4 h. Aliquots of 1 μl were diluted to 30 μl for reverse phase HPLC analysis and co-injections were performed with the respective nucleoside analogs.

T4 polynucleotide kinase labeling of reverse transcription primers. The T4 polynucleotide kinase reaction mixture contained 70 mM Tris–HCl (pH 7.6), 100 mM KCl, 10 mM MgCl$_2$, 1 mM 2-mercaptoethanol, 20 μCi [γ-^{32}P]ATP, 15 μM oligonucleotide and 1 U/μl T4 polynucleotide kinase in a 20 μl reaction mixture. The reaction mixture was incubated for 30 min at 37°C, the reaction stopped by heating at 95°C for 5 min and the primer used directly for reverse transcription.

Reverse transcription and PCR. The reverse transcription reaction mixture contained 50 mM Tris–HCl (pH 8.3), 40 mM KCl, 6 mM MgCl$_2$, 10 mM DTT, 0.1 mg/ml BSA, 0.25 mM each deoxynucleoside triphosphates, 4 μM 5′-terminally labeled 3′ primer, 1–10 pmol RNA transcript and 10 U/μl SuperScript™ II RNase H$^-$ Reverse Transcriptase in a 50 μl reaction mixture. The reaction mixture was incubated for 30 min at 42°C, stopped by heating at 95°C for 5 min and was directly used for PCR. The PCR reaction mixture contained 50 mM KCl, 10 mM Tris–HCl (pH 8.3), 1.5 mM MgCl$_2$, 0.25 mM each deoxynucleotide triphosphates, 4 μM DNA primers and 0.025 U/μl *Taq*. PCR was performed in a PTC-200 Peltier thermocycler (MJ Research, Waltham, MA), with a typical cycle of 94°C for 2 min, 92°C for 2 min, 55°C for 1 min and

72°C for 1 min and cycled between 92 and 55°C (see also cloning protocol).

Cloning and sequencing. Following the PCR reaction, the dsDNA product was cloned using a TA cloning kit (Invitrogen) according to the manufacturer's instructions and single colonies were chosen for sequencing. TA cloning utilizes the property of *Taq* polymerase, whereby the enzyme adds several 2′-deoxy-adenosines at the 3′-end of the DNA in a non-specific manner. For this, during the last PCR cycle the extension time was increased to 10 min.

RESULTS AND DISCUSSION

In comparing the functionalities contained in proteins to that of nucleic acids, at least two groups, an aliphatic amine and the corresponding thiol, could offer aptameric and catalytic nucleic acids new types of interactions. The aliphatic amine, in contrast to the aromatic amines of nucleobases, has a pK_a between 9 and 10 and would be largely in the protonated state at neutral pH values. The protonated amine, as a lysine-like functionality, could be valuable for its ability to interact with the charged phosphate backbone of nucleic acid aptamers and facilitate folding of the RNA into new types of active complexes. The amine could also function as a general acid in complexes involving catalysis. The corresponding aliphatic thiol, a cysteine-like functionality, could stabilize active complexes through the formation of covalent disulfide linkages. Additionally, the thiol is an effective ligand for metal coordination.

For the introduction of new functionality into NTP building blocks for RNA selection, NTP derivatives must meet a series of requirements: (i) the nucleotide triphosphate must be a substrate for a RNA polymerase and be incorporated into RNA transcripts with high fidelity; (ii) the transcription products must be recognized and copied with high fidelity by reverse transcriptase to generate the complementary cDNA prior to PCR amplification; (iii) the modified nucleotide must completely replace the native nucleotide in the transcript to preserve sequence information; (iv) modified nucleotide triphosphates must be relatively easily obtained.

To meet these criteria, and provide added functionality to RNA sequences, two uridine triphosphate building blocks, UNH$_2$ and USH, were prepared. Propargylamine was protected with trifluoroacetic anhydride, coupled to 5-iodouridine using previously described procedures (25) and reduced by hydrogenation to the N-protected 5-(3-aminopropyl)uridine derivative. The protected amino analog was then converted to the triphosphate using POCl$_3$ in trimethyl phosphate (29) and purified by DEAE–Sephadex chromatography. Treatment with concentrated ammonia hydrolyzed the trifluoroacetamide protecting group to generate the UTP derivative tethering the free amine (see Fig. 1). The purity of the final product was confirmed by HPLC. By comparison, the thiol analog, prepared according to Piccirilli and co-workers (26), was also converted to the triphosphate derivative, but stored with the thiol protected as the disulfide. The disulfide could then be cleaved to unmask the free thiol (see Fig. 1) just prior to its use with T7 RNA polymerase.
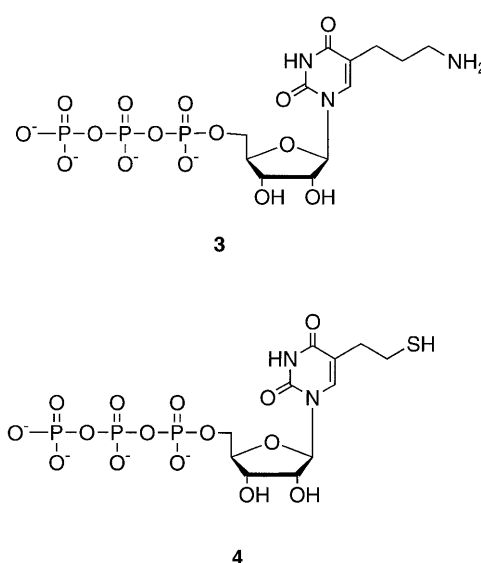

Figure 1. 5-(2-Amino)propyl uridine 5′-triphosphate (UNH$_2$) and 5-(2-mercapto-ethyl)uridine 5′-triphosphate (USH).

Transcription reactions

While a number of protocols for polymerization reactions were examined, the simplest reactions were performed using a double-stranded T7 RNA polymerase promoter sequence followed by a single-stranded template. In this case, the single-stranded template coded for a 5mer that contained one site for analog incorporation. The transcript additionally contained a single site for incorporation of [α-^{32}P]CMP for analysis of the reaction. Transcriptions for 100mers were performed using a double-stranded promoter and double-stranded template and resulted in the incorporation of multiple radiolabeled cytidine residues. Various RNA 100mers were transcribed from a random pool of dsDNA templates in which uridine comprised at least 25% of the transcribed RNA sequence; six of the uridines were present in the reverse transcription primer binding site. Additional templates were designed and used that contained six sites coding for uridine residues with no coded uridine residues in the reverse transcription primer binding site.

For the simplest assays, those resulting in 5mers, T7 RNA polymerase was able to incorporate UNH$_2$ and USH into their respective strands at rates observed to be slower than that obtained with native UTP (Fig. 2). After a 30 min reaction period UNH$_2$ incorporation reached 68% of the native UTP value and USH incorporation was 58% of the native UTP value. When UNH$_2$ was used in the transcription reaction to produce a pool of RNA 100mers, containing a randomized tract of 60 nt with 25% of the transcript sequence as UNH$_2$ (of which six were contained in the reverse transcription primer binding site), transcription efficiency was 43% relative to that with UTP. Transcription for USH in this case was only 4% as efficient as that with native triphosphates. Both transcripts exhibited a gel shift effect during PAGE due to reduced mobility relative to that of the native transcript (Fig. 3).
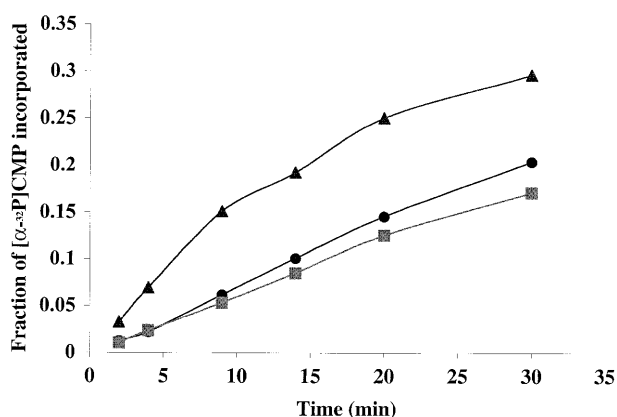
Figure 2. Time course of the incorporation of the single modified nucleotide in the 5mer transcript (5′-GGUCG-3′). Filled triangle, UTP transcription; filled circle, UNH$_2$ transcription; filled square, USH transcription. Fraction of label incorporated reflects the total isotope available in solution.
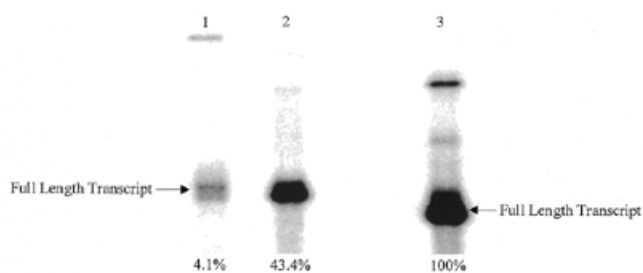

Figure 3. Transcription of dsDNA primer–template after 3 h. A 100mer template was transcribed containing a randomized tract of 60 nt, where the transcript contained 25% uridines in the random region and six uridines in the reverse transcription primer binding site. The transcription reaction contained 5 mM each NTP and 35 mM MgCl$_2$ and was analyzed after a 3 h incubation period. Lane 1, USH transcription; lane 2, UNH$_2$ transcription; lane 3, UTP transcription.

Figure 4. A pool of 60mer transcripts containing six uridines per sequence but no uridines in the reverse transcription primer binding site. Lane 1, 5 mM USH, ATP, CTP and GTP; lane 2, 5 mM NTPs; lane 3, 2.5 mM USH and 5 mM ATP, CTP and GTP.

Figure 5. Reverse phase HPLC chromatographs for nuclease digestion products: (**a**) RNA 60mer pool transcript containing 15 UNH$_2$ residues; (**b**) RNA 15mer transcript 5′-GGCGU*ACGU*ACGU*A-3′ in which the U* residues are USH.

Although poor transcription with USH could be attributed to the polymerase aborting elongation prior to strand completion, no aborted sequences were observed on the denaturing gel. One possible explanation for the poor transcription with USH is that termination occurs very early after the initiation of polymerization and prior to the incorporation of a radiolabeled cytidylate. An additional explanation for reduced activity with USH is the relatively high concentration of DTT present in those transcription assays. USH is deprotected immediately before transcription using 10 mM DTT and then used directly without purification. If the DTT is removed from the solution of triphosphate, dimerization of the triphosphates mediated through disulfide bond formation ensues. Some improvement in transcription efficiency was obtained by lowering the concentration of USH in the polymerization mixtures from 5 to 2.5 mM, thus lowering the DTT concentration. Under those conditions, transcription efficiency increased to a value of 10% of that observed with UTP (data not shown).

The dsDNA template was redesigned such that it contained only six uridine residues in a 60mer transcript and no uridines present in t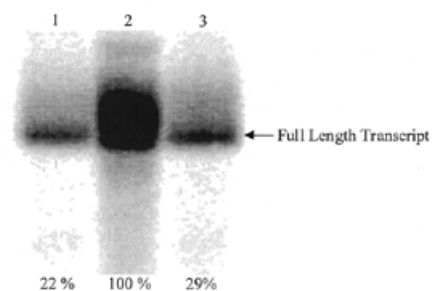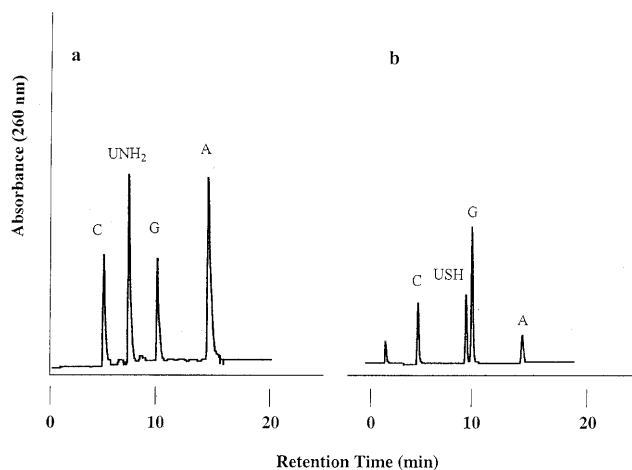he reverse transcription primer binding site (see below). The transcription efficiency with USH using the new construct increased to 29% with 2.5 mM USH and to 22% with 5 mM USH, in both cases with 5 mM of the other three common nucleotide triphosphates (Fig. 4). Transcription was also examined in the presence of reduced concentrations of ATP, CTP and GTP (at 2.5 mM as well as 2.5 mM USH) and transcription efficiency in this case fell to 18% of native UTP (data not shown). USH incorporation was optimal with 5 mM ATP, CTP and GTP and 2.5 mM USH in the transcription reaction.

To confirm that the transcripts produced in these assays contained the analog residues, RNA products (in the absence of [α-^{32}P]CMP incorporation) were isolated and subjected to digestion with snake venom phosphodiesterase and alkaline phosphatase. The digestion mixture was then analyzed by reverse phase HPLC (Fig. 5a and b). Figure 5a illustrates the analysis of the digestion products from a pool of 60mer RNAs that each contained 15 (25%) UNH$_2$ residues and Figure 5b shows the digestion analysis of products from an RNA 15mer containing only three USH residues. In both cases the analyses clearly indicate the presence of the analog residues.

Reverse transcription reactions and fidelity

For cloning and sequencing purposes necessary to measure the fidelity of incorporation of the modified nucleotides, a single 100mer RNA was transcribed r(*GCACACCAAUUACAUUC-ACU*CACGGAUCUCUGGUAACUUAACUAUACAAGCC-CGAUACCACCUUCCUGUAGCUGGGUGCC*CAUCCAU-CAUCUACUACUCA*). Reverse transcription of this 100mer RNA occurred with an efficiency of 25% for the UNH_2 ($U = UNH_2$) transcript relative to the native RNA template, while no reverse transcription was observed for the USH transcript (data not shown). Cloning and sequencing of the amplified dsDNA oligos obtained from the UNH_2-containing RNA revealed that transcription and reverse transcription had proceeded with high fidelity. From five UNH_2 sequences, only one was mutated; this sequence contained two mutations in which two contiguous cytidines at positions 65 and 66 were thymidines (Fig. 6). These mutations could have resulted from misincorporation of UNH_2, but since they occurred in but a single sequence, it seems more likely to represent mutations which occurred within the normal fidelity range for *Taq* polymerase. All of the UNH_2 residues appear to have been faithfully incorporated for template dA residues. By comparison, no mutations were observed for the sequences obtained using UTP.

Reverse transcription of the 100mer RNA containing USH residues (~25%) did not result in significant amounts of cDNA product. We prepared a new construct in which all the uridine residues were removed from the reverse transcription primer binding site to test whether poor primer hybridization and/or poor binding by the reverse transcriptase to the heteroduplex containing USH residues was responsible for the low quantities of reverse transcription product. This construct exhibited 80–90% efficiency for reverse transcription of USH RNAs relative to RNAs containing uridine (Fig. 7).

The reverse transcription efficiency was rechecked for UNH_2 with a new construct containing no uridine residues in the reverse transcription primer binding site and with the randomized tract of 60 nt containing 25% uridines. The efficiency of reverse transcription in this case for the UNH_2-containing transcript was observed to be indistinguishable from that of the native sequence (Fig. 8).

Cloning and sequencing of the amplified dsDNA oligos from the new construct for USH-containing sequences revealed that transcription and reverse transcription had both proceeded with high fidelity (Fig. 6). Of the six thymidines in the transcript one is reintroduced into the dsDNA by its presence in the PCR primer. Therefore, only five thymidines were considered for fidelity observations. Of the five clones from transcripts containing uridine, one exhibited two mutations where GC was replaced by TT in the PCR primer binding site and in a second clone one deletion occurred in the T7 promoter region (data not shown). From these five clones for USH transcripts, one exhibited deletion of a single G in the (GCAT × 5) region, while another sequence had an insertion of C in this region. The third and fourth clones resulted in the complete deletion of one GCAT block and the fifth clone had the sequence GCATGCAT deleted. Despite the fact that all the clones from the reverse transcribed USH template were mutations that were either deletions or insertions, none of these occurred at the sites of incorporation of the USH analog residues. We attribute the high rate of deletion and insertion events for USH transcripts as likely being derived from a highly stabilized secondary structure of the RNA template, which could arise as the result of the formation of disulfide crosslinks (even though these reactions were performed in the presence of DTT). Such artifacts have been reported previously in both reverse transcription (30) and PCR (30,31). We believe that these deletions arose during the reverse transcription step where reverse transcriptase simply reads past the stable secondary structures. The error rate of *Taq* polymerase is in the range 0.1×10^{-4}–2×10^{-4} per nucleotide per pass on the template, depending on reaction conditions (32,33). Mutations may have arisen from the ability of *Taq* DNA polymerase to introduce

```
Sequences from UNH2 transcription
No. of clones
  (3)   GCACACCAATTACATTCACTCACGGATCTCTGGTAACTTAACTATACAAGCCCGATACCACCTTCCTGTAGCTGGGT
  (1)   GCACACCAATTACATTCACTCACGGATCTCTGGTAACTTAACTATACAAGCCCGATACCACCTTTTTGTAGCTGGGT


Sequences from UTP transcription
No. of clones
  (5)   GCACACCAATTACATTCACTCACGGATCTCTGGTAACTTAACTATACAAGCCCGATACCACCTTCCTGTAGCTGGGT




Sequences from USH transcription
No. of clones
  (1)   GCTAATACGACTCACTATAGGCACTGCACACACACCCGACGCATGCCATGCATGCATGCATGAGAGAGAGCAGGAGAAGAG
  (1)   GCTAATACGACTCACTATAGGCACTGCACACACACCCGACGCATGC ATGCATGCAT-CATGAGAGAGAGCAGGAGAAGAG
  (2)   GCTAATACGACTCACTATAGGCACTGCACACACACCCGACGCATGC ATGCATGCAT----GAGAGAGAGCAGGAGAAGAG
  (1)   GCTAATACGACTCACTATAGGCACTGCACACACACCCGACGCATGC ATGCAT--------GAGAGAGAGCAGGAGAAGAG


Sequences from UTP transcription
No. of clones
  (4)   GCTAATACGACTCACTATAGGCACTGCACACACACCCGACGCATGCATGCATGCATGCAT GAGAGAGAGCAGGAGAAGAG
  (1)   TTTAATACGACTCACTATAGGCACTGCACACACACCCGACGCATGCATGCATGCATGCAT GAGAGAGAGCAGGAGAAGAG
  (1)   GCTA-TACGACTCACTATAGGCACTGCACACACACCCGACGCATGCATGCATGCATGCAT GAGAGAGAGCAGGAGAAGAG
```

Figure 6. Sequencing results from both USH and UNH_2 transcription reactions. In both cases control reactions with UTP were used. Numbers in parentheses indicate the number of clones obtained.
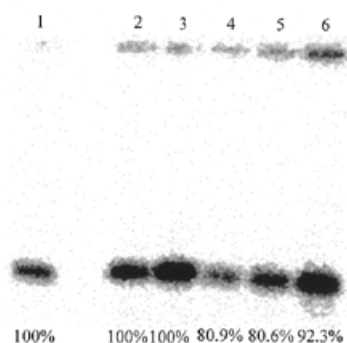
Figure 7. Reverse transcripts from RNA containing USH and native UTP. The 60mer transcript contained six uridines but no uridines in the reverse transcription primer binding site. Lane 1, 0.5 pmol UTP RNA as a control; lane 2, 1 pmol UTP RNA as a control; lane 3, 2 pmol UTP RNA as a control; lane 4, reverse transcript from 0.5 pmol USH-containing RNA; lane 5, reverse transcript from 1 pmol USH-containing RNA; lane 6, reverse transcript from 2 pmol USH-containing RNA.
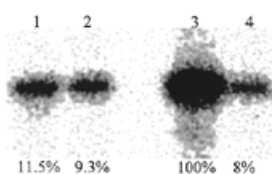

Figure 8. Reverse transcription of RNA containing UNH$_2$ and native UTP. The 100mer transcript contained a randomized tract of 60 nt with 25% uridines but no uridines in the reverse transcription primer binding site. Lane 1, reverse transcript from 1 pmol UNH$_2$-containing RNA; lane 2, reverse transcript from 1 pmol UTP-containing RNA; lane 3, 10 pmol UTP RNA as a control; lane 4, 1 pmol UTP RNA as a control.

mutations and not due to the poor fidelity of the analog, but in such cases similar mutations should have been present in the UTP control reactions.

CONCLUSIONS

Although not as efficient as native NTP derivatives, both the UNH$_2$ and USH analog triphosphates function as substrates for T7 RNA polymerase in place of UTP. The reverse transcription reactions with template containing these analog residues are also effective, provided that their presence in the primer binding site is minimized. High fidelity incorporation of the analogs is observed, but owing to possible secondary structure effects, some deletions are apparent when using the USH derivative. In general both analogs should function as effective

building blocks for RNA selection procedures and permit the introduction of new functionality into RNA aptamers or catalysts.

REFERENCES

1. Kruger,K., Grabowski,P.J., Zaug,A.J., Sands,J., Gottschling,D.E. and Cech,T.R. (1982) *Cell*, **31**, 147–157.
2. Guerrier-Takada,C., Gardiner,K., Marsh,T., Pace,N. and Altman,S. (1983) *Cell*, **35**, 849–857.
3. Usher,D.A. and McHale,A.H. (1976) *Proc. Natl Acad. Sci. USA*, **73**, 1149–1153.
4. Eckstein,F. and Lilly,D.J.M. (eds) (1996) *Nucleic Acids and Molecular Biology*, Vol. 10, *RNA Catalysis*. Springer-Verlag, Heidelberg, Germany.
5. Breaker,R.R. (1997) *Chem. Rev.*, **97**, 371–390.
6. Osborne,S.E. and Ellington,A.D. (1997) *Chem. Rev.*, **97**, 349–370.
7. Abelson,J.N. (ed.) (1996) *Methods Enzymol.*, **267**.
8. Hager,A.J., Pollard,J.D. and Szostak,J.W. (1996) *Chem. Biol.*, **3**, 717–725.
9. Benner,S.A., Allenmann,R.K., Ellington,A.D., Ge,L., Glasfeld,A., Leanz,G.F., Krauch,T., MacPherson,L.J., Moronet,S., Piccirilli,J.A. and Weinhold,E. (1987) *Cold Spring Harbor Symp. Quant. Biol.*, **52**, 53–63.
10. Tarasow,T.M. and Eaton,B.E. (1998) *Biopolymers*, **48**, 29–37.
11. Sakthivel,K. and Barbas,C.F. (1998) *Angew. Chem. Int. Ed. Engl.*, **37**, 2872–2875.
12. Perrin,D.M., Garestier,T. and Helene,C. (1999) *Nucl. Nucl.*, **18**, 377–391.
13. Battersby,T.R., Ang,D.N., Burgstaller,P., Jurczyk,S.C., Bowser,M.T., Buchanan,D.D., Kennedy,R.T. and Benner,S.A. (1999) *J. Am. Chem. Soc.*, **121**, 9781–9789.
14. Latham,J.A., Johnson,R. and Toole,J.J. (1994) *Nucleic Acids Res.*, **22**, 2817–2822.
15. Santoro,S.W., Joyce,G.F., Sakthivel,K., Gramatikova,S. and Barbas,C.F. (2000) *J. Am. Chem. Soc.*, **122**, 2433–2439.
16. Pyle,A.M. and Cech,T.R. (1991) *Nature*, **350**, 628–631.
17. Eaton,B.E. and Pieken,W.A. (1995) *Annu. Rev. Biochem.*, **64**, 837–863.
18. Dewey,T.M., Mundt,A.A., Crouch,G.J., Zyzniewski,M.C. and Eaton,B.E. (1995) *J. Am. Chem. Soc.*, **117**, 8474–8475.
19. Dewey,T.M., Zyzniewski,M.C. and Eaton,B.E. (1996) *Nucl. Nucl.*, **15**, 1611–1617.
20. Tarasow,T.M., Tarasow,S.L. and Eaton,B.E. (1997) *Nature*, **389**, 54–57.
21. Wiegand,T.W., Janssen,R.C. and Eaton,B.E. (1997) *Chem. Biol.*, **4**, 675–683.
22. Aurup,H., Williams,D.M. and Eckstein,F. (1992) *Biochemistry*, **31**, 9636–9641.
23. Kujau,M.J. and Wolfl,S. (1998) *Nucleic Acids Res.*, **26**, 1851–1853.
24. Zawadzki,V. and Gross,H.J. (1991) *Nucleic Acids Res.*, **19**, 1948.
25. Hobbs,F. (1989) *J. Org. Chem.*, **54**, 3420–3422.
26. Sun,S.G., Tang,X.Q., Marchant,A., Anjaneyulu,P.S.R. and Piccirilli,J.A. (1996) *J. Org. Chem.*, **61**, 5708–5709.
27. Milligan,J.F., Groebe,D.R., Witherell,G.W. and Uhlenbeck,O.C. (1987) *Nucleic Acids Res.*, **15**, 8783.
28. He,B.K., Riggs,D.L. and Hanna,M.M. (1995) *Nucleic Acids Res.*, **23**, 1231–1238.
29. Ludwig,J. (1981) *Acta Biochim. Biophys. Acad. Sci. Hung.*, **16**, 131–133.
30. Buell,G.N., Wickens,M.P., Payvar,F. and Schimke,R.T. (1987) *J. Biol. Chem.*, **253**, 2471–2482.
31. Green,I.R. and Sargan,D.R. (1991) *Gene*, **109**, 203–210.
32. Keohavong,P. and Thilly,W.G. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 9253–9257.
33. Eckert,K.A. and Kunkel,T.A. (1990) *Nucleic Acids Res.*, **18**, 3739–3744.