SCIENTIFIC REP⚙RTS

OPEN

A specific, promoter-independent activity of T7 RNA polymerase suggests a general model for DNA/RNA editing in single subunit RNA Polymerases

Subha Narayan Sarcar & Dennis L. Miller

Insertional RNA editing has been observed and characterized in mitochondria of myxomycetes. The single subunit mitochondrial RNA polymerase adds nontemplated nucleotides co-transcriptionally to produce functional tRNA, rRNA and mRNAs with full genetic information. Addition of nontemplated nucleotides to the 3′ ends of RNAs have been observed in polymerases related to the mitochondrial RNA polymerase. This activity has been observed with T7 RNA polymerase (T7 RNAP), the well characterized prototype of the single subunit polymerases, as a nonspecific addition of nucleotides to the 3′ end of T7 RNAP transcripts _in vitro_. Here we show that this novel activity is an editing activity that can add specific ribonucleotides to 3′ ends of RNA or DNA when oligonucleotides, able to form intramolecular or intermolecular hairpin loops with recessed 3′ ends, are added to T7 RNA polymerase in the presence of at least one ribonucleotide triphosphate. Specific ribonucleotides are added to the recessed 3′ ends through Watson-Crick base pairing with the non-base paired nucleotide adjacent to the 3′ end. Optimization of this activity is obtained through alteration of the lengths of the 5′-extension, hairpin loop, and hairpin duplex. These properties define a T7 RNAP activity different from either transcriptional elongation or initiation.

RNA editing is an alteration of the genetic information in RNA relative to its DNA template. It is an additional step in gene expression, where specific nucleotides at specific positions are substituted for, added to, or deleted from the RNA in order to produce functional RNAs, especially mRNAs with a functional open reading frame. The term "RNA editing" was introduced in 1986 to describe the insertion of four uridine nucleotides into specific sites of mitochondrial cox2 mRNA in Trypanosome mitochondria[1]. There are two major forms of RNA editing, [A] substitutional, involving nucleotide modification, typically a transamination or deamination, which converts one nucleotide to another, for example, editing of tRNAs in _Acanthamoeba castellanni_[2] or single nucleotide conversions such as A to I[3]), C to U or U to C[4] and [B] insertional in which a nucleotide (or nucleotides) are inserted or deleted to produce a functional RNA[5–7].

Two different types of insertional editing have been observed, post-transcriptional and co-transcriptional, each exemplified by a mitochondrial DNA system in which that type of insertional RNA editing is used extensively to produce functional mitochondrial RNAs. Post-transcriptional RNA editing, as exemplified by RNA editing in the mitochondria of trypanosomes, requires an endonuclease to break the RNA backbone at a specific site, a terminal transferase to add a nucleotide or nucleotides to the 3′ end of the RNA at the break site, and an RNA ligase to restore the RNA backbone at the site of the insertion or deletion[8–12]. Specificity of the nucleotide insertion or deletion site and the number of nucleotides inserted or deleted at a site is determined by an antisense RNA, termed a guide RNA [gRNA], which has the genetic information to produce functional mRNAs significantly different from the template DNA[13].

Co-transcriptional insertional RNA editing as exemplified by RNA editing in the mitochondria of myxomycetes such as _Physarum polycephalum_ produce functional mRNAs, rRNAs, and tRNAs[7,14–17] different from their

Department of Biological Sciences, The University of Texas at Dallas, Richardson, Texas, 75083-0688, USA. Correspondence and requests for materials should be addressed to D.L.M. (email: dmiller@utdallas.edu)

1

mtDNA template through non-templated addition of specific nucleotides at specific sites in nascent RNAs by the mitochondrial RNA polymerase (mtRNAP)[15,18–20]. Two major questions in the mechanism of this type of editing are: [A] How is the correct nucleotide inserted without a DNA template? [B] How is the correct site of insertion determined? Miller & Miller[20] and Visomirski-Robic & Gott[19] have each proposed that an unknown mechanism causes the mitochondrial RNA polymerase to pause during elongation of RNA on the DNA template, and insert a specific, non-templated nucleotide (or occasionally nucleotides), before continuing DNA-templated elongation. Miller & Miller[20] have shown that the Physarum mtRNAP is able to add nucleotides to the 3′ end of RNAs in the absence of DNA *in vitro*. However, with complex RNA populations without DNA templates *in vitro* any of the four ribonucleotides can be added.

Both aspects of the specificity of this co-transcriptional RNA editing involve the mtRNAP. Mitochondrial RNAPs are single subunit polymerases[21] found in essentially all eukaryotic organisms and have conserved structural and functional similarities that indicate that they evolved from a common ancestor present at or near the advent of the eukaryotic line[18]. Interestingly, the conservation of these structural and functional similarities extend to other single subunit polymerases, such as the bacteriophage RNAPs[22–24], DNA polymerases (DNAPs) such as the Klenow fragment of DNAP I in *E. coli*[25–27], and reverse transcriptases (RNA-directed DNA polymerases) such as the HIV-1 reverse transcriptase[28–30], constituting a super-family of single subunit polymerases with structural and functional homology.

The prototype RNA polymerase for this group of polymerases is the T7 RNAP encoded by the T7 *E. coli* bacteriophage DNA and it has been extensively studied[23,24,31,32]. The crystal structure of T7 RNAP has been solved along with the different domains and their functions corresponding to both initiation and elongation processes[33–35]. It is a single subunit ~99 kDa protein (883 amino acid)[36], found in Bacteriophage T7 and is primarily involved in transcription activity without the need of any other accessory transcription factors. The specificity and robustness of T7 RNA polymerase's transcription activity comes from its specificity towards its promoter, a 23 nucleotide dsDNA sequence[37–39]. This specificity towards the promoter, involves the promoter recognition loop[40] as well as the N terminus domain in T7 RNA polymerase[41,42]. T7 RNA polymerase can successfully carry out transcription on a synthetic ssDNA oligonucleotide template containing double stranded T7 promoter sequence[43,44].

In addition to *de novo* initiation which requires a promoter, single stranded DNA template, and the four ribonucleotide triphosphates; and elongation which requires a duplex with one strand having a recessed 3′ end (primer), the other strand 5′ extended (template), and four ribonucleotide triphosphates; there have been reports of a third activity in which addition of ribonucleotides occurs at the 3′ end of RNAs in the absence of a promoter or DNA template[45–48]. This activity was incorrectly attributed to the primer extension activity of elongation. Until recently, this activity of T7 RNAP has not been well characterized and most of the research on it has been focused on ways to prevent the addition of non-templated nucleotides to the 3′ ends of run-off RNAs[46,49,50]. Here we report the characterization of this activity. We show [A] that the T7 RNAP can add nucleotides to the 3′ end of RNAs and DNAs in a process that is neither elongation nor initiation, but is a type of RNA/DNA editing, [B] that the added nucleotide is added specifically as determined by base pairing of either an inter- or intramolecular template with an extended 5′ end relative to the recessed 3′ end, and [C] that the properties of the oligonucleotide that is provided determine the efficiency of nucleotide addition and distinguish this activity from either initiation or elongation activities. We will discuss this activity in terms of a general model of RNA editing by single subunit polymerases that may explain the specificity of nucleotide insertion during RNA editing by the *Physarum* mtRNAP.

Material and Methods

Oligonucleotides.　Synthetic single stranded DNA oligonucleotides (10 nmol scale) were synthesized by Eurofin-Operon. Oligonucleotides were dissolved in water to 200 μM and stored at 4 °C.

Oligonucleotide naming convention.　Each oligonucleotide is named by a five number sequence separated by dashes where the first number is the oligonucleotide length, the second number is the length of the single stranded 5′ extension. The third and fifth number are the length of the double stranded region (stem of hairpin), and the fourth number is the length of the hairpin loop.

T7 RNAP, reaction buffer, ribonucleotide triphosphates.　T7 RNA polymerase (New England Biolabs (NEB) catalog number M0251S) was supplied in: 100 mM NaCl, 50 mM Tris-HCl (pH 7.9), 1 mM EDTA, 20 mM 2-mercaptoethanol, 0.1% Triton X-100 and 50% glycerol and stored in a −20 °C freezer.

NEB 10 × RNA pol reaction buffer was supplied. 1 × RNA Pol Reaction Buffer contains 40 mM Tris-HCl, 6 mM MgCl$_2$, 2 mM spermidine, 10 mM dithiothreitol, pH 7.9 @ 25 °C, 1 mg/ml BSA (Bovine Serum Albumin).

Oligonucleotide radiolabeling.　The *in vitro* oligonucleotide labeling protocol was derived from the *in vitro* transcription protocol recommended by New England Biolabs (NEB) for their T7 RNA polymerase[51]. For radiolabeling, (α-^{32}P) rGTP, or (α-^{32}P) rATP (3000 Ci/mmol 10 mCi/ml; Perkin Elmer, product numbers BLU006H and BLU003H, respectively) were used. *In vitro* oligonucleotide radiolabeling was assayed in reactions containing 6 μl of 200 μM ssDNA oligonucleotides, 2 μl 10 × RNA polymerase buffer, 2 μl of 5 mM ribonucleoside triphosphates, and 1 μl of α-^{32}P radiolabeled rATP/rGTP to a final volume of 20 μl. The reactions were incubated at 37 °C water bath for 60 minutes. After incubation 2 μl of 6× loading dye (10 mM Tris-HCl (pH 7.6) 0.03% bromophenol blue, 60% glycerol, 60 mM EDTA) was added to individual reaction mixtures and samples were run on 15% TBE (Tris-Borate EDTA) polyacrylamide gels at a constant 200 V and 75 mA, until the bromophenol blue migrated three-quarters of the length of the gel. Experiments were routinely repeated on both denaturing and non-denaturing gels to check their reproducibility.
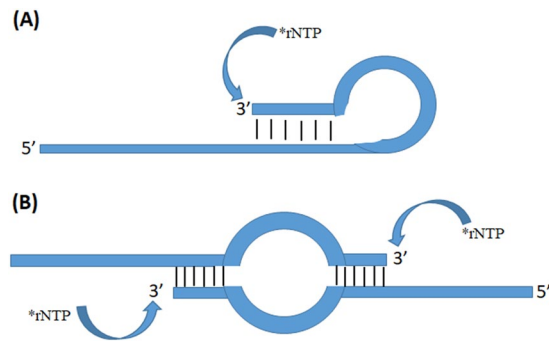
Figure 1. Diagram of the Intramolecular and Intermolecular base pairing forms in the oligonucleotides used in the study. (**A**) intramolecular base pairing form of the oligonucleotide showing the site of nucleotide addition. (**B**) intermolecular base pairing form of the oligonucleotide showing the sites of nucleotide addition.

Detection and Quantification.   Nucleic acids in the gels were stained with $1\times$ TBE, Ethidium Bromide for 10 min and then visualized under Bio-Rad Gel Doc™. Detection of incorporated radioactivity was achieved by exposure to a phosphor screen (Molecular Dynamics) and scanned on a Phosphor Imager (Storm 840, Amersham Biosciences). Quantification of the phosphor image was done using ImageQuaNT software version 5.2 (Molecular Dynamics). Only the most intense band in each lane was quantitated. Graphical representation of major band intensities below each gel are shown to allow easier interpretation.

Results

Several laboratories have reported that single subunit RNAPs could add radiolabeled nucleotide triphosphates to the 3′ end of RNA in the absence of a DNA template[20,48,52] and that T7 RNAP could add ribonucleotides to the 3′ end of DNA and RNA[44,46,47,53–55]. We hypothesized that these observations indicated that these RNAPs had a novel activity similar to the RNA editing observed in mitochondria of *Physarum polycephalum*[7] and other myxomycetes[56]. To study this activity, we have developed an *in vitro* assay for labeling the 3′ ends of RNA and DNA oligonucleotides using T7 RNAP and a single radiolabeled ribonucleotide triphosphate. In contrast to using complex mixtures of RNA or DNA, with these simple oligonucleotide sequences, some oligonucleotides failed to label while others labeled efficiently. We have determined that this labeling specificity depends on the ability of the 3′ end of the oligonucleotide to anneal with an internal complementary sequence which positions the 3′ end next to a nucleotide that is not base paired in the 5′ extension of the oligonucleotide. This can take place within the same oligonucleotide (intramolecular base pairing) or between complementary sequences in two separate oligonucleotides (intermolecular base pairing) as shown in Fig. 1. Most of the oligonucleotides of this type exists in both forms in solution. Form B (Fig. 1B) can also be labeled, since it has two recessed 3′ ends.

An example of these experiments is shown in Fig. 2. Two similar DNA oligonucleotides, 36-6-6-18-6-A and 36-6-6-18-6-G, each 36 nucleotides in length, and each with the potential to form intra or intermolecular duplexes of 6 base pairs with recessed 3′ ends (Fig. 2B), were incubated with T7 RNAP in the presence of either radiolabeled GTP or ATP. These oligonucleotides varied from one another by a single nucleotide in the first position of the 5′ extension template. Oligonucleotide 36-6-6-18-6-A has a T nucleotide in this position which will serve as a template for an A to be added to the 3′ double stranded end. Whereas oligonucleotide 36-6-6-18-6-G has a C nucleotide in the template position. After incubation the oligonucleotides were applied to a 15% polyacrylamide gel to separate the oligonucleotide from the unreacted radiolabeled nucleotides. Figure 2A shows an autoradiograph of the gel. Oligonucleotide 36-6-6-18-6-A was labeled with ATP, but not GTP, while oligonucleotide 36-6-6-18-6-G was labeled with GTP, but not ATP. Oligonucleotides similar to 36-6-6-18-6 but lacking the 6 base pair complementarity were not labeled (data not shown). Reactions with oligonucleotides and labeled nucleotides but in which the T7RNAP was omitted, were similarly negative (data not shown).

This experiment demonstrates that when presented with an oligonucleotide with a 3′ end with the potential to base pair as a recessed 3′ end, either inter- or intra- molecularly, the T7 RNAP is able to add the radiolabeled nucleotide to the 3′ end, but only if it can form a Watson-Crick base pair with the nucleotide in the 5′ extension (template) next to the nucleotide base paired with the 3′ end nucleotide (primer).

To better understand this self-templated labeling at the 3′ end and to optimize this reaction, the single stranded DNA oligonucleotides have been systematically altered to determine the importance and contribution of [1] the overall length, [2] the length and base composition of the potential duplex, [3] the length of the 5′ extension, and [4] loop length (number of nucleotides separating the complementary sequences) with the goal of optimizing the reaction.

Contribution of 5′ extension length (V-V-6-8-6 series).

To determine the contribution of 5′ extension length, oligonucleotides were designed with increasing 5′ extension lengths (0–16 nucleotides), constant base pairing length (6 pairs), constant loop length (8 nucleotides), and increasing overall length (20–36 nucleotides).

Figure 3 shows the results of these experiments. Oligonucleotide labeling was absent in the absence of a single stranded template (no 5′ extension, 20-0-6-8-6). Labeling increased with 1, 2, 3, and 5 nucleotide 5′ extensions. 5′ extensions of 10 and 16 nucleotides were most intensely labeled and were essentially equal in intensity. 5′ extensions from 1–10 nucleotides had essentially linearly increased labeling (a slope of 10 units per nucleotide
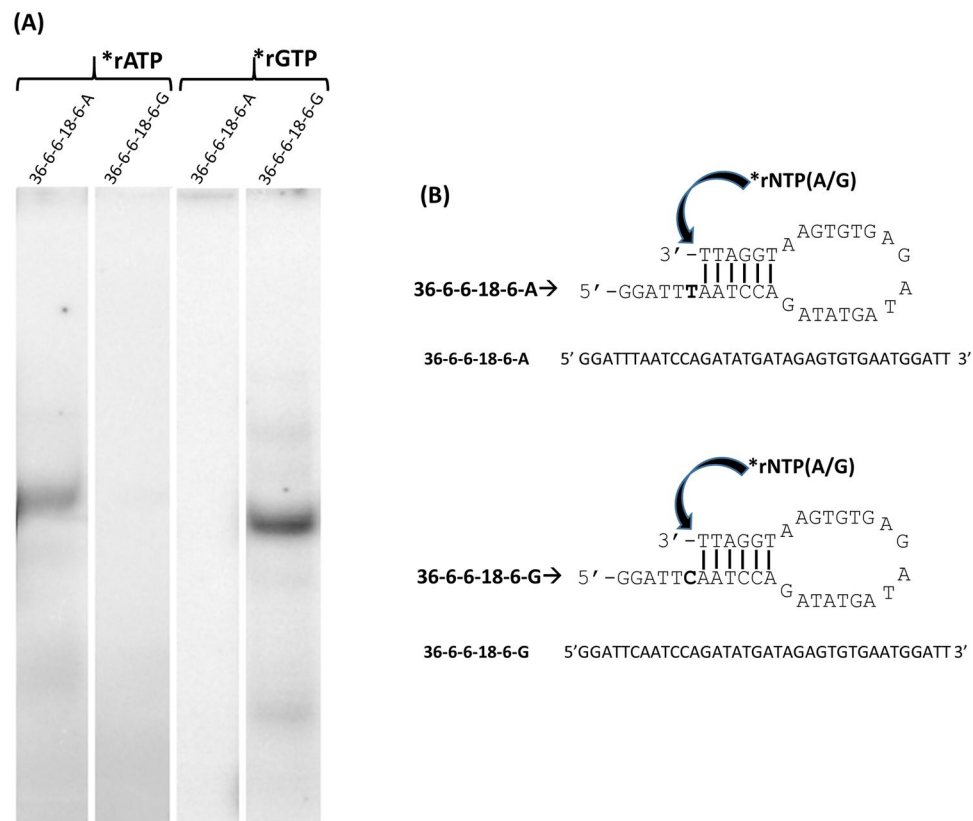
Figure 2. Specificity of oligonucleotide labeling by T7 RNA polymerase. (**A**) Autoradiograph of 15% polyacrylamide gel which shows 3′ end labeling by T7 RNAP on DNA oligonucleotides (36mer) in the presence of the α-^{32}P GTP (right two lanes) or α-^{32}P ATP (left two lanes). (**B**) The sequence and potential looped structure of the DNA oligonucleotides labeled in (**A**).

extension length). Above a length of 10 nucleotides the labeling reached a plateau. Upper (lower mobility) bands in Fig. 3A show labeling of oligonucleotides with stable intermolecular base pairing to be proportional with lower (higher mobility) bands with intramolecular base pairing. In the 6-8-6 series oligonucleotide labeling increased with 5′ extension length up to about 10 nucleotides.

Contribution of loop length (V-2-9-V-9 and V-2-6-V-6 series). To determine the contribution of loop length to labeling efficiency, oligonucleotides were designed with constant 5′ extensions (2 nucleotides), constant double stranded regions (either 9 nucleotides or 6 nucleotides), and variable loop lengths (6–24 nucleotides) and variable overall length (20 nucleotides to 44 nucleotides). Nucleotide labeling efficiencies are shown in Fig. 4A.

Labeling is seen for oligonucleotides with both intramolecular and intermolecular base pairing. Figure 4B shows the relative intensities of the oligonucleotides with intramolecular base pairing. For oligonucleotides with a nine base pair double stranded duplex, the optimum labeling was seen in oligonucleotides with 12–18 nucleotide loops, loops longer or shorter than this were labeled less intensely. For oligonucleotides with a six base pair double stranded duplex, maximum labeling was seen for 18 to 21 nucleotides in the loop. Again, shorter or longer loop lengths labeled less well. Loop sizes for maximum labeling vary with the length of base pairing in the double stranded region. The smaller base pairing region (6 base pairs) gives optimum labeling with larger loop sizes (18–21 nucleotides) relative to the larger base pairing region (9 base pairs) which has optimum labeling with a smaller loop size (12–18 nucleotides). Overall length does not correlate with labeling intensity, for example, the two oligonucleotides that were 38 nucleotides in length gave dramatically different labeling intensities.

Contribution of base paired region (32-2-V-V-V series, V-10-V-8-V series). To determine the contribution of the length of the base pairing to 3′ end labeling efficiency, the base pairing length was systematically altered while keeping the overall length (32 nucleotides) and the 5′ extension length (2 nucleotides) constant (Fig. 5). To keep overall length constant, the loop length was decreased as the length of the base pairing region was increased. Oligonucleotides with 0, 1, 2, or 3 base pairing potential did not label efficiently, while base pairing regions of 6 or 9 base pairs (loop lengths of 18 nucleotides or 9 nucleotides respectively) did label efficiently. Surprisingly, oligonucleotides with the potential for even longer base pairing regions (12 and 15 base pairs) also did not label significantly. Again, overall length did not correlate with labeling efficiency.

To further determine whether the contribution of the base pairing length to 3′ end labeling efficiency, we varied the length of the base pairing region while holding the length of the loop region (at either 8 nucleotides or 12 nucleotides) and the 5′ extension constant (at either 2 nucleotides or 10 nucleotides) constant. The V-10-V-8-V
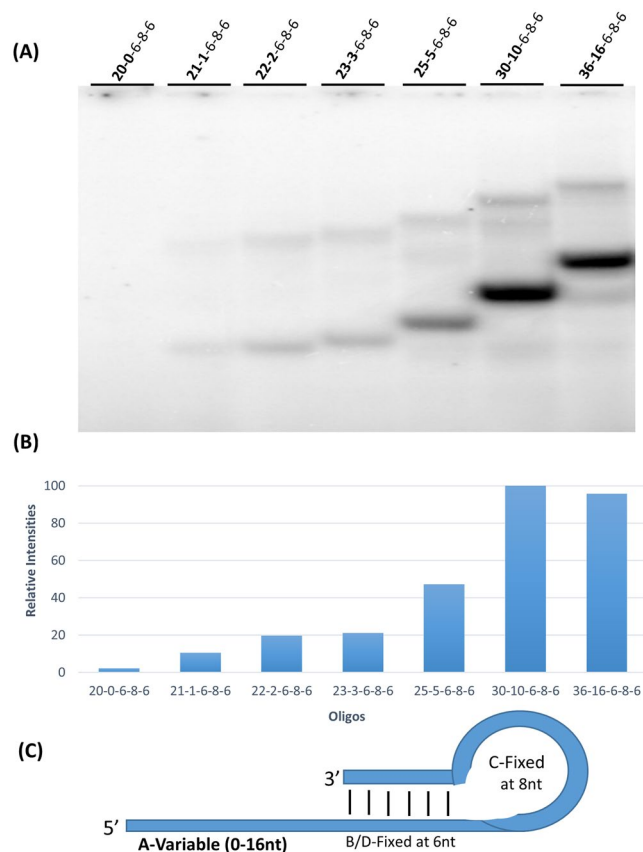
Figure 3. Effect of 5′-extension length variation on oligonucleotide labeling. (**A**) Autoradiograph of 15% polyacrylamide gel which shows 3′ end labeling by T7 RNAP on DNA oligonucleotides 36-0-6-8-6 to 20-16-6-8-6 in the presence of the α-^{32}P ATP where there is a systematic increase of 5′ extension from 0 to 16 nucleotides. (**B**) Graphical representation showing the relative intensities of oligonucleotide labeling in the most intensely labeled (intramolecularly base paired) bands in each lane, normalized against the maximum intensity in the same gel. Minor low mobility (stable intermolecularly base paired) bands were not quantitated, but give consistent results with the intramolecularly base paired form. (**C**) The general intramolecular structure of the oligonucleotides used in (**A**) showing the number of nucleotides present in each feature of the hairpin loop and 5′ extension.

series is shown in Fig. 6A; the V-2-V-12-V series is shown in Fig. 6B. In Fig. 6A the duplex length is varied from 0 to 6 base pairs with constant 5′ extension (10 nucleotides) and constant loop length (8 nucleotides). Labeling efficiency increases with base pairing length. Maximum label efficiency is at 6 base pairs. In Fig. 6B the base paired region length is varied from 3 base pairs to 18 base pairs with constant 5′ extension (2 nucleotides) and loop length (12 nucleotides). Labeling efficiency increased with duplex length up to 9 base pairs but decreased with lengths greater than 9 base pairs to nearly zero at 18 base pairs. The 9 base pair sequence that gave the maximum labeling is 5′- TATAATATT-3′.

Labeling efficiency of oligonucleotides with intermolecular base pairing. To determine if this optimum base pairing length for labeling efficiency also pertained to intermolecular base pairing and to eliminate the contribution of loop length to labeling efficiency, a series of oligonucleotides with varying lengths of sequence complementary to a template oligonucleotide were designed. Figure 7D shows the sequences of these oligonucleotides. Oligonucleotide 30-15-7-1-7-G provides the 5′ extension nucleotide which serves as the template. It has the potential to form an intramolecular hairpin loop with a 7 base pair stem, 5′ extension, and recessed 3′ end with an adjacent non-base paired C nucleotide, or an intermolecular base pairing with itself with two 7 base pair regions each with a 5′ extension and a recessed 3′ end with an adjacent non-base paired C nucleotide. Oligonucleotide 30-15-7-1-7-A is identical to oligonucleotide 30-15-7-1-7-G except the non-base paired nucleotide next to the recessed 3′ end is a T instead of a C nucleotide. A third oligonucleotide 45-15-6-18-6 has the potential to form an intramolecular hairpin with a 6 base pair stem, 18 nucleotide loop, a 5′ extension, a recessed 3′ end with an adjacent non-base paired C nucleotide. This oligonucleotide serves as a positive labeling control when used with radiolabeled GTP.

Below the template strands in Fig. 7D are eight 15 nucleotide primer oligonucleotides. Oligonucleotide 15–15 is complementary to the 15 nucleotides on the 3′ end of both 30 nucleotide template oligonucleotides. The remaining 15 nucleotide primers retained 12 to 0 nucleotides of complementarity to the template at their 3′ end.
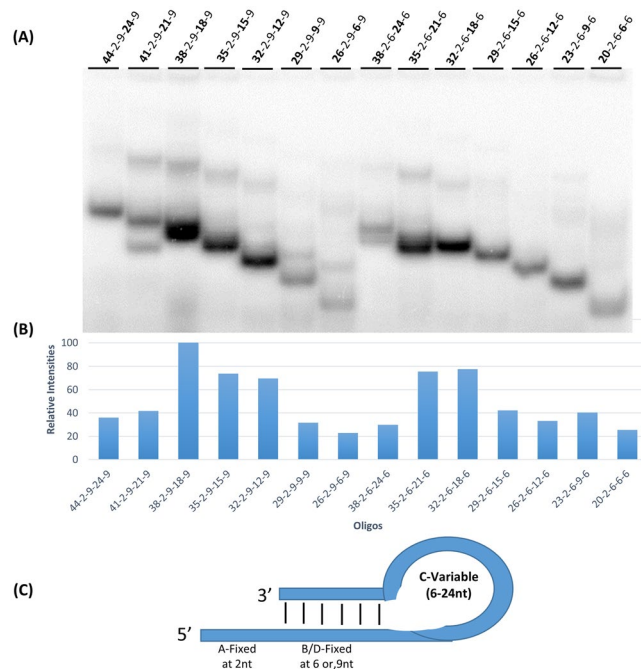
Figure 4. Effect of loop length variation on oligonucleotide labeling. (**A**) Autoradiograph of 15% polyacrylamide gel which shows 3′ end labeling by T7 RNAP on DNA oligonucleotides in the presence of the α-^{32}P ATP where the loop length is systematically varied. (**B**) Graphical representation showing the relative intensities of the most intensely labeled bands in each lane (intramolecularly base paired). Minor bands corresponding to intermolecularly base paired oligonucleotides were not quantitated. (**C**) The general structure of the oligonucleotides used in (**A**) showing the number of nucleotides present in each feature of the hairpin loop and 5′ extension.

Figure 7A shows the labeling of each of the individual oligonucleotides by T7 RNAP in the presence of radio-labeled GTP. Both 45-15-6-18-6 and 30-15-7-1-7-G are labeled since they can form intramolecular or intermolecular base paired regions with a recessed 3′ end next to a C nucleotide that can base pair with GTP. 30-15-7-1-7-A is not labeled even though it can form the same base paired regions with a recessed 3′ end. However, in this case, the base adjacent to the 3′ end is an A nucleotide which cannot base pair with GTP. None of the 15 base pair primer oligonucleotides are labeled since they cannot form intermolecular or intramolecular base pairing regions with recessed 3′ ends.

Figure 7B shows the labeling of the oligonucleotides by T7 RNAP in the presence of radiolabeled GTP when the primer oligonucleotides are mixed with the template oligonucleotides in various combinations. The lanes of the left side of Fig. 7B show each of the eight 15 nucleotide primers mixed with 30-15-7-1-7-G. In each case 30-15-7-1-7-G is labeled indicating intramolecular or intermolecular base pairing with itself. In addition, primer oligonucleotides 15-15, 15-12, 15-9 and 15-6 are labeled indicating that they can form stable base pairing with the template to produce a recessed 3′ end next to the C nucleotide. Primer oligonucleotides 15-3, 15-2, 15-1, and 15-0 do not label significantly indicating that they cannot form stable base pairing regions with the template. Figure 7C is a graphic representation of the relative labeling intensities of the primer oligonucleotides. The 15-9 primer oligonucleotide labels most intensely indicating that an optimum length or base pairing stability has been reached.

On the right side of Fig. 7B primer oligonucleotides 15-15, 15-12, 15-9, and 15-6 were mixed with template oligonucleotide 30-15-7-1-7-A. Although these oligonucleotides can form the same base pairing as the oligonucleotides on the left side of Fig. 7B, these oligonucleotides do not label because there is not a C nucleotide complementary to GTP next to the 3′ end.

Interdependence of optimization parameters. While overall oligonucleotide length does not correlate with labeling efficiency, experiments to determine the optimum length of the 5′ extension, the double stranded region (stem), and the loop for the efficiency of labeling the 3′ end with a specific nucleotide give optimum lengths for one of the parameters when the other two parameters are held constant. However, the optimums for a parameter do not appear be independent, since a different optimum for a parameter is observed when the other two parameters are fixed at different lengths. To determine the interdependence of the three parameter lengths, 27 oligonucleotides were created with 3, 6, or 9 nucleotides in the 5′ extension, double stranded region, or loop in all possible combinations [Fig. 8]. The relative labeling intensity of each oligonucleotide is shown.

Analysis of these results allows several generalizations to be made about the interdependence of the parameter lengths. Labeling efficiency generally increases with 5′ extension length for a given stem length/loop length. However, different stem length/loop length combinations give significantly different efficiencies with 3, 6, or 9 nucleotides in the 5′ extension.
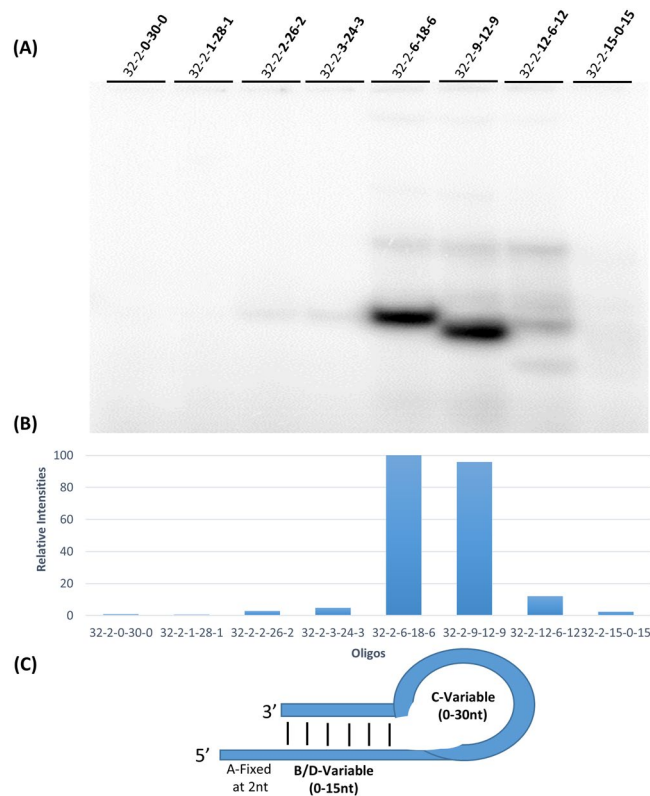
Figure 5. Effect of loop length variation and duplex length variation on oligonucleotide labeling. (**A**) Autoradiograph of 15% polyacrylamide gel which shows 3′ end labeling by T7 RNAP on DNA oligonucleotides in the presence of the α-^{32}P ATP where the loop length and duplex length are systematically varied. (**B**) Graphical representation showing the relative intensities of oligonucleotide labeling with varying loop length after exposure, normalized against the maximum intensity in the same gel. Minor, intermolecularly base paired bands were not quantitated. (**C**) The general structure of the oligonucleotides used in (**A**) showing the number of nucleotides present in each feature of the hairpin loop and 5′ extension.

The labeling efficiencies generally either decrease with the length of the double stranded stem region (3 base pairs >6 base pairs >9 base pairs) for six 5′ extension/loop combinations or peak with 6 base pairs (6 base pairs >3 or 9 base pairs) for the three 5′ extension/loop combinations with loops of 9 nucleotides.

Labeling efficiencies vary with loop length in several different ways. For the three oligonucleotides with 6 base pair stems, 9 nucleotides loops have the highest labeling efficiency (9 nucleotides >3 or 6 nucleotides). For the three oligonucleotides with 3 base pair stems, 9 nucleotide loops have the lowest efficiency (3 or 6 nucleotides >9 nucleotides). For the remaining three oligonucleotides with 9 base pair stems, 6 nucleotide loops have the lowest labeling efficiency (3 and 9 nucleotides >6 nucleotides or 3 nucleotides >6 or 9 nucleotides).

The highest labeling efficiencies are with oligonucleotides 12-3-3-3-3 and 27-6-6-9-6 followed closely by 30-9-6-9-6, 15-6-3-3-3, and 18-9-3-3-3. Oligonucleotides in the X-X-3-3-3 series have high labeling efficiency, essentially independent of 5′ extension length. Oligonucleotides of the X-X-6-9-6 series have significantly higher labeling efficiencies with 6 or 9 nucleotide (longer) 5′ extensions.

Labeling efficiency and stem length. Since the formation of a base paired region (stem) is necessary to put a recessed 3′ primer end on the 5′ extension template, it is surprising that longer stem lengths generally give decreased labeling efficiencies. While at least three nucleotides of base pairing complementarity are needed for significant labeling, labeling efficiencies with increasingly longer base pairing potential generally increase up to a peak of about 5 – 7 base pairs and then decline with increasing length, where loop length and 5′ extension length are held constant. This duplex optimum could be caused by at least two possible constraints, [A] a duplex length constraint (static model) in which only duplexes of intermediate length can be bound productively, and [B] a duplex stability constraint (dynamic model) where a duplex is required but where labeling efficiency decreases with duplex stability, indicating that both single stranded and duplex forms are necessary.

Efficiency probably does not depend directly on duplex length since different loop lengths and 5′ extension lengths, when held constant, give a different optimum duplex length (e.g. 9 nucleotides, see Figs 5B and 6B).

To determine whether duplex stability is a critical factor in labeling efficiency, we assayed 25 different oligonucleotides with constant loop length (8 nucleotides), constant 5′ extension (10 nucleotides), and duplex regions varying from 3 nucleotides to 7 nucleotides and from 7 to 18 hydrogen bonds (Figs 9 and S1). Except for 5 of the 25 nucleotides (with duplex sequences of AGATGT, GAGTG, GTTAAG, AGA, and ATGGA) labeling efficiency increased with hydrogen bond number up to 14 hydrogen bonds and decreased between 14 and 18 hydrogen
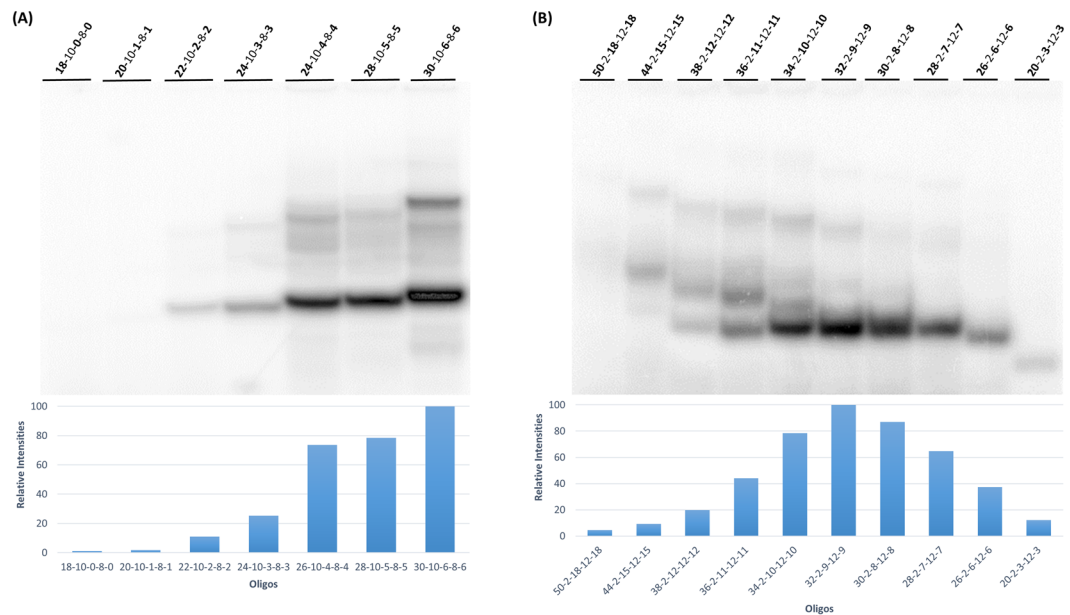
Figure 6. Effect of duplex length variation on oligonucleotide labeling. (**A**) Relative intensities of oligonucleotide labeling where duplex length is varied and where 5′-extension is fixed at 10 nucleotides and loop length is fixed at 8 nucleotides. (**B**) Relative intensities of oligonucleotide labeling where duplex length is varied and where 5′-extension is fixed at 2 nucleotides and loop length is fixed at 12 nucleotides. In both (**A**) and (**B**), the most intensely labeled (intra molecularly base paired) bands were normalized against the maximum intensity in the same gel. Minor low mobility (stable intermolecularly base paired) bands were not quantitated but appeared consistent with the intramolecular base paired form.
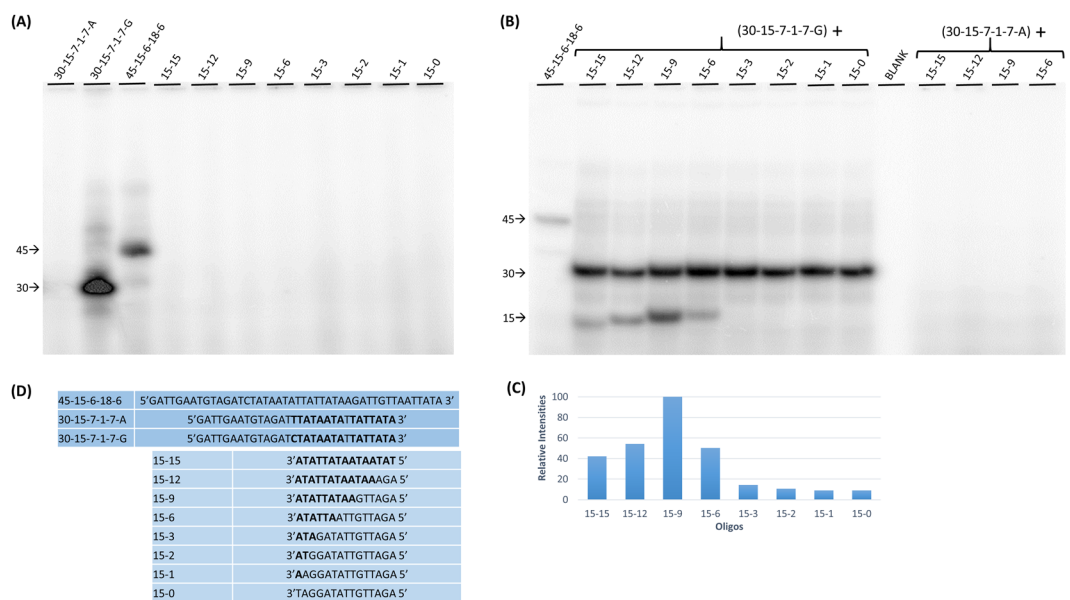

Figure 7. Oligonucleotide labeling with intermolecular duplex formation. (**A**) rGTP labeling of the individual oligonucleotides listed in (**D**). The numbers on the left are the size of oligonucleotides in nucleotides. (**B**) rGTP labeling of oligonucleotides in reactions involving a mixture of primer and template oligonucleotides. The left side of (**B**) shows the labeling of primers with various lengths of base pairing potential with template 30-15-7-1-7-G. The right side of (**B**) shows the labeling of primers with various lengths of base pairing potential with template 30-15-7-1-7-A. (**C**) Graphical representation showing the relative intensities of oligonucleotide primer labeling (with varying template base pairing potential) after exposure, normalized against the maximum intensity of the primer oligonucleotides in the same gel. (**D**) List of the sequences of the oligonucleotides used in this experiment. The upper three oligonucleotides are used as templates, while the lower eight oligonucleotides are used as primers.
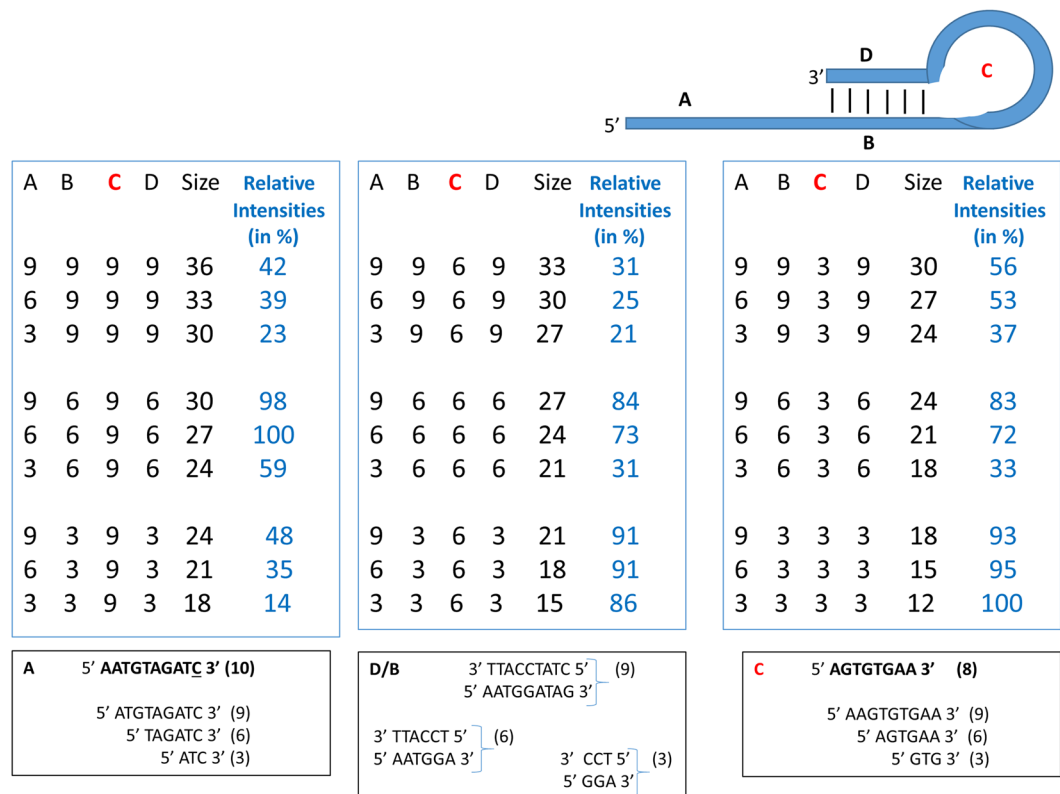
A	B	C	D	Size	Relative Intensities (in %)
9	9	9	9	36	42
6	9	9	9	33	39
3	9	9	9	30	23
9	6	9	6	30	98
6	6	9	6	27	100
3	6	9	6	24	59
9	3	9	3	24	48
6	3	9	3	21	35
3	3	9	3	18	14

A	B	C	D	Size	Relative Intensities (in %)
9	9	6	9	33	31
6	9	6	9	30	25
3	9	6	9	27	21
9	6	6	6	27	84
6	6	6	6	24	73
3	6	6	6	21	31
9	3	6	3	21	91
6	3	6	3	18	91
3	3	6	3	15	86

A	B	C	D	Size	Relative Intensities (in %)
9	9	3	9	30	56
6	9	3	9	27	53
3	9	3	9	24	37
9	6	3	6	24	83
6	6	3	6	21	72
3	6	3	6	18	33
9	3	3	3	18	93
6	3	3	3	15	95
3	3	3	3	12	100

A   5′ **AATGTAGAT<u>C</u> 3′ (10)**
5′ ATGTAGATC 3′ (9)
5′ TAGATC 3′ (6)
5′ ATC 3′ (3)

D/B   3′ TTACCTATC 5′ (9)
5′ AATGGATAG 3′
3′ TTACCT 5′ (6)
5′ AATGGA 3′   3′ CCT 5′ (3)
5′ GGA 3′

C   5′ **AGTGTGAA 3′ (8)**
5′ AAGTGTGAA 3′ (9)
5′ AGTGAA 3′ (6)
5′ GTG 3′ (3)

Figure 8. A list of 27 oligonucleotides of the type shown at the top, used to test the interdependence of optimization parameters for 3′ oligonucleotide labeling. Columns denote (**A**) the length of their 5′ extensions, (**C**) the length of their hairpin loops, (**B**,**D**) duplex length, total length (size), and the relative labeling intensities. Labeling intensities are shown as the relative intensities of the set of 27 oligonucleotides where the most intense band (here, in two oligonucleotides) is defined as 100%. Boxes A, B/D, and C are the actual sequences of each length of each part of the oligonucleotides.
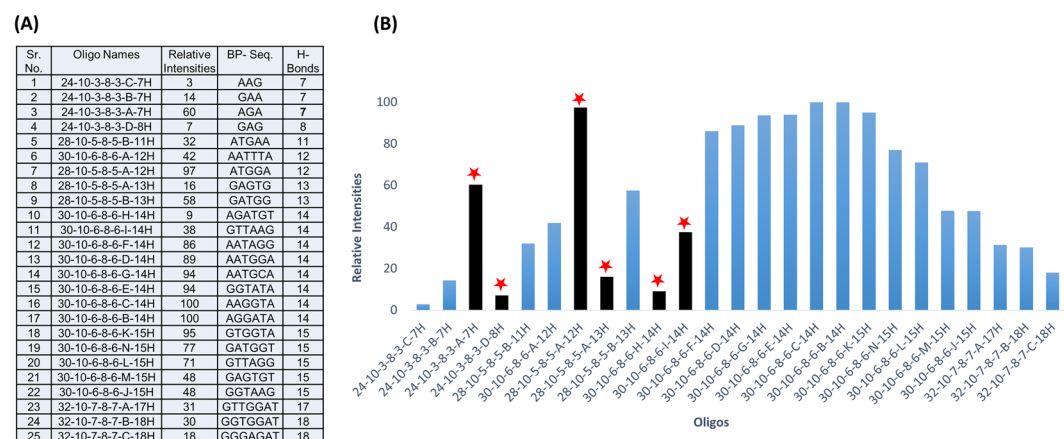
(A)

Sr. No.	Oligo Names	Relative Intensities	BP- Seq.	H-Bonds
1	24-10-3-8-3-C-7H	3	AAG	7
2	24-10-3-8-3-B-7H	14	GAA	7
3	24-10-3-8-3-A-7H	60	AGA	7
4	24-10-3-8-3-D-8H	7	GAG	8
5	28-10-5-8-5-B-11H	32	ATGAA	11
6	30-10-6-8-6-A-12H	42	AATTTA	12
7	28-10-5-8-5-A-12H	97	ATGGA	12
8	28-10-5-8-5-A-13H	16	GAGTG	13
9	28-10-5-8-5-B-13H	58	GATGG	13
10	30-10-6-8-6-H-14H	9	AGATGT	14
11	30-10-6-8-6-I-14H	38	GTTAAG	14
12	30-10-6-8-6-F-14H	86	AATAGG	14
13	30-10-6-8-6-D-14H	89	AATGGA	14
14	30-10-6-8-6-G-14H	94	AATGCA	14
15	30-10-6-8-6-E-14H	94	GGTATA	14
16	30-10-6-8-6-C-14H	100	AAGGTA	14
17	30-10-6-8-6-B-14H	100	AGGATA	14
18	30-10-6-8-6-K-15H	95	GTGGTA	15
19	30-10-6-8-6-N-15H	77	GATGGT	15
20	30-10-6-8-6-L-15H	71	GTTAGG	15
21	30-10-6-8-6-M-15H	48	GAGTGT	15
22	30-10-6-8-6-J-15H	48	GGTAAG	15
23	32-10-7-8-7-A-17H	31	GTTGGAT	17
24	32-10-7-8-7-B-18H	30	GGTGGAT	18
25	32-10-7-8-7-C-18H	18	GGGAGAT	18

(B)

Figure 9. Relative labeling intensities of oligonucleotides with fixed 5′-extension (10 nucleotides), loop length (8 nucleotides) and duplex length (6 nucleotides), but with different duplex sequences. (**A**) A list of the 25 oligonucleotides assayed in Fig. S1 showing [1] the oligonucleotide names, [2] the relative intensities of the 3′ end labeling as determined from the autoradiographs in panels (A) and (B) of Fig. S1, [3] the base pair sequence of the duplex, 5′ to 3′, and [4] the number of hydrogen bonds in the base pairs of the duplex. (B) Graphical representation of the relative labeling intensities of the 25 oligonucleotides from the list in panel (A) plotted against the number of hydrogen bonds in their duplex. The five samples in black with a red star are exceptions to the general trend.

bonds, consistent with duplex stability being the major factor in labeling efficiency and supporting a dynamic model of labeling (Fig. 9). The five exceptions (shown in black with red stars in Fig. 9B) had both relative labeling efficiencies higher (AGA and ATGGA) and lower (AGATGT, GAGTG, and GTTAAG) than expected for the number of hydrogen bonds in the duplex of the oligonucleotide. In addition, the remaining eight oligonucleotides with 14 hydrogen bonds in its duplex varied in efficiency from 86% to 100%. Likewise, the five oligonucleotides with duplexes with 15 hydrogen bonds varied from 48% to 95% relative labeling efficiency. The five exceptional oligonucleotides and the variation of labeling efficiency in groups of oligonucleotides with the same number of hydrogen bonds indicate that while duplex stability (hydrogen bond number) generally correlates with labeling efficiency, variation within this correlation is observed which may be due to the base composition or to a specific nucleotide sequence motif in the duplex.

In order to determine the cause of the variation in relative labeling efficiency within oligonucleotides with duplexes with the same number of hydrogen bonds, we compared relative labeling efficiencies of oligonucleotides with different base compositions but with the same number of hydrogen bonds in the duplex. There are 15 different base compositions of 6 nucleotides with 14 hydrogen bonds, and there are 16 different base compositions of 6 nucleotides with 15 hydrogen bonds. Three similar six base compositions with 14 hydrogen bonds (A_3G_2T, A_3GCT, $A_2G_2T_2$) and two similar base composition with 15 hydrogen bonds (AG_3T and A_2G_3T) were compared. The four oligonucleotides with base composition A_3G_2T or A_3GCT varied in relative labeling efficiency from 86% to 100%, and the four oligonucleotides with base composition $A_2G_2T_2$ varied in relative labeling efficiency from 9% to 100%. Likewise, the four oligonucleotides with base composition AG_3T varied in relative labeling efficiency from 48–95%. These wide variations in relative labeling efficiency with oligonucleotides with the same base composition in their duplex eliminated base composition as a factor in relative labeling efficiency.

To determine the role of specific sequence motifs as a factor in relative labeling efficiency, we scanned the 13 duplex sequences with 6 base pairs and 14 or 15 hydrogen bonds in the duplex for common motifs which were consistent in relative labeling efficiency. In general, sequences with GG motifs in the duplex labeled more efficiently (48% to 100%) than oligonucleotides in which the G nucleotides in the duplex were separated by other nucleotides (9% to 48%).

Discussion

This paper describes and characterizes a novel DNA/RNA editing activity of T7 RNAP with nucleotide specificity. This novel activity is detected when an oligonucleotide is provided that has a duplex with a recessed 3′ end (primer) and a 5′ extension (template) where the first unpaired base specifies the labeling nucleotide through Watson-Crick base pairing. This novel editing activity differs from the two classic transcription activities, initiation and elongation. It differs from initiation in that nucleotide addition occurs in the absence of a promoter, and differs from elongation in that nucleotides are not added processively on the template, but only one or a few nucleotides are added since each nucleotide added to the duplex ultimately decreases the labeling efficiency. Probably the most unique feature of this novel activity is the addition of a DNA templated ribonucleotide to a DNA 3′ end to form a DNA primed, RNA-DNA duplex. Although T7 RNAP will only add ribonucleotides to 3′ ends to create RNA strands, it is able to add ribonucleotides to 3′ ends of either RNA or DNA and will use either RNA or DNA templates to specify ribonucleotide addition. Therefore, either RNA or DNA oligonucleotides able to form intramolecular and intermolecular duplexes with recessed 3′ ends can be specifically labeled with radiolabeled ribonucleotides as well as DNA oligonucleotide templates with RNA primers with recessed 3′ ends or RNA oligonucleotide templates with DNA primers with recessed 3′ ends.

There have been a few reports of non-templated nucleotide addition to 3′ ends of nucleic acids with T7 RNAP[57] and other single subunit RNAPs[20,44,46,47,53–55]. These reports did not mention the potential for nucleotide specificity presumably because they were using complex templates that positioned 3′ recessed ends at many positions, and therefore, next to essentially any nucleotide. Rather than studying this activity, most of the focus was on blocking non-templated nucleotide addition which produced *in vitro* transcripts with unintended sequences. A few studies using less complex templates demonstrated that T7 RNAP could add nucleotides to RNA in the absence of DNA, presumably through inter- and intramolecular formation of RNA-RNA duplexes[48]. Konarska and Sharp[44] have reported ribonucleotide addition to RNA templated, RNA primed oligonucleotides but did not explore the mechanisms of this activity and did not report activity on DNA templated or DNA primed oligonucleotides. Krupp[45] observed promoter independent RNA synthesis on DNA templates able to form intermolecular or intramolecular hairpin loops that could be used as DNA primers for RNA synthesis forming RNA-DNA duplexes. However, none of these reports discussed the actual mechanism and criteria for T7 RNAP to add these apparently non DNA-templated nucleotides to the 3′ end. These reactions were all done with four ribonucleotide triphosphates present and often produced complex gel patterns that were complicated to interpret. Here we used DNA oligonucleotides with a single radiolabeled ribonucleotide triphosphate present to simplify interpretation.

To determine their contribution to labeling efficiency, we have systematically varied the three parts of the oligonucleotide (duplex, 5′ extension, loop). In general, labeling efficiency increases with 5′ extension length up to about 10 nucleotides. The loop length can affect labeling efficiency depending on duplex length, but is essentially optional since intermolecular base pairing between two separate oligonucleotides can also produce 3′ end labeling. Small loops when present seem to destabilize a duplex and longer than optimal duplexes can be activated by a small loop while optimal length duplexes may be inactivated by a short loop. On the other hand, longer loops may be detrimental since they increase the distance between the component strands of the duplex decreasing the probability of duplex formation. The most critical portion of these oligonucleotides is the duplex which positions the 3′ end recessed relative to the 5′ end. The duplex length for optimal labeling is about six base pairs for duplexes 33% GC rich (14 hydrogen bonds), about 9 nucleotides at 0% GC (18 hydrogen bonds), or about three base pairs at 66% GC (8 hydrogen bonds).

The observation that duplex stability is a major factor in labeling efficiency implies a dynamic, two-step mechanism for labeling. Although the duplex must transiently form in order to prime for labeling, duplex instability implies that an unfolded single strand intermediate is required for polymerase binding. If the polymerase binds the unfolded single strand oligonucleotide in a productive way, so that the duplex can form with the 3′ end base paired and at the active site of the T7 RNAP, a nucleotide complementary to the template base binds, and a phosphodiester bond is formed incorporating the base into the duplex. In this model, phosphodiester bond formation is coupled to the release of the T7 RNAP. This non-processive mechanism is in contrast to elongation which uses a processive, primer extension on template mechanism. Although duplex stability appears to be a major factor in labeling efficiency, additional sequence motifs may enhance RNA polymerase binding and therefore contribute to labeling efficiency. One motif that seems to enhance 3′ end labeling in oligonucleotides is consecutive G nucleotides (GG motif).

One of the major differences between RNA polymerases and DNA polymerases is the ability of RNAP to initiate synthesis *de novo* by binding to a promoter and then transitioning to elongation through promoter clearance. Early RNAPs probably lacked promoter recognition domains and initiated through primer extension like their closely-related DNA polymerases or through some other non-promoter mechanism. Several contemporary examples of such RNA polymerases exist[20,48,52]. It is likely that the RNA/DNA editing activity of T7 RNAP derives from such a primal activity. Mutations in the gene for T7 RNAP produce altered enzymes which support this idea. For example, some mutated enzymes have been found to be catalytically active although they bind promoter weakly, and in some instances mutated enzymes which are transcriptionally inactive have strong binding to the promoter[58]. Also, mutated T7 RNA polymerase[59] with two amino acid substitutions (Y639F and S641A) has altered specificity towards promoter, but gains the ability to utilize dNTPs and catalyze RNA and DNA synthesis from circular supercoiled plasmid DNA[60]. This Y639F mutant retains the ability to use RNA or DNA templates and can display de novo initiated or primed DNA-directed DNA polymerase, reverse transcriptase, RNA-directed RNA polymerase or DNA-directed RNA polymerase activities depending simply on the templates and substrates presented to it in the synthesis reaction[55].

This novel DNA/RNA editing activity in T7 RNAP is probably present in all single subunit polymerases and may be the basis of nucleotide specificity during RNA editing by the mtRNAP in mitochondria of myxomycetes[61]. One of the major questions about Myxomycete RNA editing is how the identity of the inserted nucleotide is determined. Based on the observed *in vitro* mechanism of editing in T7 RNAP, we propose a model of nucleotide specificity in RNA editing for the related *Physarum* mtRNAP. At an RNA editing site, the *Physarum* mtRNAP pauses and then may either stay in place, release, or back up on the template to allow the 3′ end of the RNA to loop back on a complementary sequence either in the RNA or in the non-template strand of DNA, and add a nucleotide at the 3′ end. This positioning of the 3′ end allows specific addition to the 3′ end using the vacant position next to the 3′ end to specify the added nucleotide through Watson-Crick base pairing. The transient base pairing of the RNA is then released resulting in the insertion of a nucleotide at the 3′ end not templated by the template strand of DNA. Elongation is then resumed as a phosphodiester bond is formed between the unpaired 3′ nucleotide and the next templated nucleotide. *In vitro* presentation of the *Physarum* mitochondrial RNAP (mtRNAP) with oligonucleotides, similar to those described above, would be a test of this hypothesis. Confirmation of this novel activity in *Physarum* mtRNAP would be a major step toward understanding the co-transcriptional mechanism of RNA editing by the mtRNAP in mitochondria of myxomycetes.

References

1. Benne, R. *et al*. Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* **46**, 819–826 (1986).
2. Lonergan, K. M. & Gray, M. W. Editing of transfer RNAs in Acanthamoeba castelanii mitochondria. *Science* **259**, 812–815 (1993).
3. Rubio, M. A. T., Ragone, F. L., Gaston, K. W., Ibba, M. & Alfonzo, J. D. C to U editing stimulates A to I editing in the anticodon loop of a cytoplasmic threonyl tRNA in Trypanosoma brucei. *J Biol Chem* **281**, 115–120 (2006).
4. Gaston, K. W. *et al*. C to U editing at position 32 of the anticodon loop precedes tRNA 5′ leader removal in trypanosomatids. *Nucl Acids Res* **35**, 6740–6749 (2007).
5. Feagin, J. E., Abraham, J. M. & Stuart, K. Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei. *Cell* **53**, 413–422 (1988).
6. Gott, J. M., Visomirski, L. M. & Hunter, J. L. Substitutional and insertional RNA editing of the cytochrome c oxidase subunit 1 mRNA of Physarum polycephalum. *J Biol Chem* **268**, 25483–25486 (1993).
7. Mahendran, R., Spottswood, M. R. & Miller, D. L. RNA editing by cytidine insertion in mitochondria of Physarum polycephalum. *Nature* **349**, 434–438, https://doi.org/10.1038/349434a0 (1991).
8. Ringpis, G.-E. *et al*. Mechanism of U insertion RNA editing in trypanosome mitochondria: the bimodal TUTase activity of the core complex. *J Mol Biol* **399**, 680–695 (2010).
9. Gao, G. *et al*. Uridine insertion/deletion RNA editing in Trypanosomatids: specific stimulation *in vitro* of Leishmania tarentolae REL1 RNA ligase activity by the MP63 zinc finger protein. *Protist* **161**, 489–496 (2010).
10. Alatortsev, V. S., Cruz-Reyes, J., Zhelonkina, A. G. & Sollner-Webb, B. Trypanosoma brucei RNA editing: coupled cycles of U deletion reveal processive activity of the editing complex. *Mol Cellular Biol* **28**, 2437–2445 (2008).
11. Sabatini, R. & Hajduk, S. L. RNA Ligase and Its Involvement in Guide RNA/mRNA Chimera Formation. *J Biol Chem* **270**, 7233–7240 (1995).
12. Cruz-Reyes, J., Zhelonkina, A. G., Huang, C. E. & Sollner-Webb, B. Distinct functions of two RNA ligases in active Trypanosoma brucei RNA editing complexes. *Mol Cellular Biol* **22**, 4652–4660 (2002).
13. Kable, M. L., Seiwert, S. D., Heidmann, S. & Stuart, K. RNA editing: a mechanism for gRNA-specified uridylate insertion into precursor mRNA. *Science* **273**, 1189–1195 (1996).
14. Byrne, E. M. & Gott, J. M. Cotranscriptional editing of Physarum mitochondrial RNA requires local features of the native template. *RNA* **8**, 1174–1185 (2002).
15. Cheng, Y. W., Visomirski-Robic, L. M. & Gott, J. M. Non-templated addition of nucleotides to the 3′ end of nascent RNA during RNA editing in Physarum. *EMBO J* **20**, 1405–1414, https://doi.org/10.1093/emboj/20.6.1405 (2001).
16. Antes, T., Costandy, H., Mahendran, R., Spottswood, M. & Miller, D. Insertional editing of mitochondrial tRNAs of Physarum polycephalum and Didymium nigripes. *Mol Cell Biol* **18**, 7521–7527 (1998).
17. Mahendran, R. *et al*. Editing of the mitochondrial small subunit rRNA in Physarum polycephalum. *EMBO J* **13**, 232–240 (1994).

18. Miller, M. L., Antes, T. J., Qian, F. & Miller, D. L. Identification of a putative mitochondrial RNA polymerase from Physarum polycephalum: characterization, expression, purification, and transcription *in vitro*. *Curr Genet* **49**, 259–271, https://doi.org/10.1007/s00294-005-0053-y (2006).

19. Visomirski-Robic, L. M. & Gott, J. M. Insertional editing in isolated Physarum mitochondria is linked to RNA synthesis. *RNA* **3**, 821–837 (1997).

20. Miller, M. L. & Miller, D. L. Non-DNA-templated addition of nucleotides to the 3′ end of RNAs by the mitochondrial RNA polymerase of Physarum polycephalum. *Mol Cell Biol* **28**, 5795–5802, https://doi.org/10.1128/MCB.00356-08 (2008).

21. Tracy, R. L. & Stern, D. B. Mitochondrial transcription initiation: promoter structures and RNA polymerases. *Curr Genet* **28**, 205–216 (1995).

22. Masters, B. S., Stohl, L. L. & Clayton, D. A. Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* **51**, 89–99 (1987).

23. Cermakian, N., Ikeda, T. M., Cedergren, R. & Gray, M. W. Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucl Acids Res* **24**, 648–654 (1996).

24. Cermakian, N. *et al.* On the evolution of the single-subunit RNA polymerases. *J Mol Evol* **45**, 671–681 (1997).

25. Clark, J. M. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucl Acids Res* **16**, 9677–9686 (1988).

26. Yang, S. W., Astatke, M., Potter, J. & Chatterjee, D. K. Mutant Thermotoga neapolitana DNA polymerase I: altered catalytic properties for non-templated nucleotide addition and incorporation of correct nucleotides. *Nucl Acids Res* **30**, 4314–4320 (2002).

27. Gao, G., Orlova, M., Georgiadis, M. M., Hendrickson, W. A. & Goff, S. P. Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc Natl Acad Sci USA* **94**, 407–411 (1997).

28. Ong, J. L., Loakes, D., Jaroslawski, S., Too, K. & Holliger, P. Directed Evolution of DNA Polymerase, RNA Polymerase and Reverse Transcriptase Activity in a Single Polypeptide. *J Mol Biol* **361**, 537–550, https://doi.org/10.1016/j.jmb.2006.06.050 (2006).

29. McAllister, W. T. & Raskin, C. A. The phage RNA polymerases are related to DNA polymerases and reverse transcriptases. *Mol Microb* **10**, 1–6 (1993).

30. Lanciault, C. & Champoux, J. J. Single unpaired nucleotides facilitate HIV-1 reverse transcriptase displacement synthesis through duplex RNA. *J Biol Chem* **279**, 32252–32261 (2004).

31. Delarue, M., Poch, O., Tordo, N., Moras, D. & Argos, P. An attempt to unify the structure of polymerases. *Protein Eng* **3**, 461–467 (1990).

32. Chamberlin, M. & Ryan, T. 4 Bacteriophage DNA-Dependent RNA Polymerases. *The Enzymes* **15**, 87–108 (1982).

33. Sousa, R., Chung, Y. J., Rose, J. P. & Wang, B.-C. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* **364**, 593–599 (1993).

34. Jeruzalmi, D. & Steitz, T. A. Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J* **17**, 4101–4113, https://doi.org/10.1093/emboj/17.14.4101 (1998).

35. Cheetham, G. M., Jeruzalmi, D. & Steitz, T. A. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399**, 80–83, https://doi.org/10.1038/19999 (1999).

36. Moffatt, B. A., Dunn, J. J. & Studier, F. W. Nucleotide sequence of the gene for bacteriophage T7 RNA polymerase. *J Molecular Biol* **173**, 265–269 (1984).

37. Oakley, J. L., Strothkamp, R. E., Sarris, A. H. & Coleman, J. E. T7 RNA polymerase: promoter structure and polymerase binding. *Biochem* **18**, 528–537 (1979).

38. McAllister, W. T. & Carter, A. D. Regulation of promoter selection by the bacteriophage T7 RNA polymerase *in vitro*. *Nucl Acids Res* **8**, 4821–4837 (1980).

39. Brown, J. E., Klement, J. F. & McAllister, W. T. Sequences of three promoters for the bacteriophage SP6 RNA polymerase. *Nucl Acids Res* **14**, 3521–3526 (1986).

40. Sousa, R. Structural and mechanistic relationships between nucleic acid polymerases. *Trends Biochem Sci* **21**, 186–190 (1996).

41. Cheetham, G. M. & Steitz, T. A. Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305–2309 (1999).

42. Rong, M., Castagna, R. & McAllister, W. Cloning and purification of bacteriophage K11 RNA polymerase. *Biotechnique*s **27**, 690, 692, 694-690, 692, 694 (1999).

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

44. Konarska, M. M. & Sharp, P. A. Replication of RNA by the DNA-dependent RNA polymerase of phage T7. *Cell* **57**, 423–431 (1989).

45. Krupp, G. Unusual promoter-independent transcription reactions with bacteriophage RNA polymerases. *Nucl Acids Res* **17**, 3023–3036 (1989).

46. Triana-Alonso, F. J., Dabrowski, M., Wadzack, J. & Nierhaus, K. H. Self-coded 3′-extension of run-off transcripts produces aberrant products during *in vitro* transcription with T7 RNA polymerase. *J Biol Chem* **270**, 6298–6307 (1995).

47. Huang, J., Brieba, L. G. & Sousa, R. Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochem* **39**, 11571–11580 (2000).

48. Cazenave, C. & Uhlenbeck, O. C. RNA template-directed RNA synthesis by T7 RNA polymerase. *Proc Natl Acad Sci* **91**, 6972–6976 (1994).

49. Kao, C., Zheng, M. & Rüdisser, S. 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 (1999).

50. Nacheva, G. A. & Berzal-Herranz, A. Preventing nondesired RNA-primed RNA extension catalyzed by T7 RNA polymerase. *Eur J Biochem* **270**, 1458–1465 (2003).

51. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning: a laboratory manual*. 10.27-10.37 (Cold spring harbor laboratory press, 1989).

52. Biebricher, C. & Luce, R. Template-free generation of RNA species that replicate with bacteriophage T7 RNA polymerase. *EMBO J* **15**, 3458 (1996).

53. Arnaud-Barbe, N., Cheynet-Sauvion, V., Oriol, G., Mandrand, B. & Mallet, F. Transcription of RNA templates by T7 RNA polymerase. *Nucl Acids Res* **26**, 3550–3554 (1998).

54. Konarska, M. M. & Sharp, P. A. Structure of RNAs replicated by the DNA-dependent T7 RNA polymerase. *Cell* **63**, 609–618 (1990).

55. Sousa, R. & Padilla, R. A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J* **14**, 4609–4621 (1995).

56. Krishnan, U., Barsamian, A. & Miller, D. L. Evolution of RNA editing sites in the mitochondrial small subunit rRNA of the Myxomycota. *Methods Enzymol* **424**, 197–220, https://doi.org/10.1016/S0076-6879(07)24009-1 (2007).

57. Krupp, G. RNA synthesis: strategies for the use of bacteriophage RNA polymerases. *Gene* **72**, 75–89 (1988).

58. Gross, L., Chen, W.-J. & McAllister, W. T. Characterization of bacteriophage T7 RNA polymerase by linker insertion mutagenesis. *J Mol Biol* **228**, 488–505 (1992).

59. Raskin, C. A., Diaz, G. A. & McAllister, W. T. T7 RNA polymerase mutants with altered promoter specificities. *Proc Natl Acad Sci USA* **90**, 3147–3151 (1993).

60. Kostyuk, D. *et al*. Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. *FEBS letters* **369**, 165–168 (1995).

61. Miller, D., Padmanabhan, R. & Sarcar, Subha N. In *Myxomycetes: Biology, Systematics, Biogeography and Ecology Biology, Systematics, Biogeography and Ecology* (eds Steven Stephenson, Carlos Alvarado) 107–143 (Academic Press, 2017).

Author Contributions

S.N.S. and D.L.M. conceptualized the project. S.N.S. performed and visualized the experiments. S.N.S. and D.L.M.analyzed the data. S.N.S. and D.L.M. wrote the paper.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-32231-6.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.