Accurate and efficient one-pot reverse transcription and amplification of 2’ fluoro-modified nucleic acids by commercial DNA polymerases

*Arianna S. Thompson, Susanna E. Barrett, Aurora G. Weiden, Ananya Venkatesh, Madison K. C. Seto, Simone Z. P. Gottlieb and Aaron M. Leconte**

W.M. Keck Science Department of Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, California 91711, United States of America

Corresponding Author
*Telephone: 909-607-9353. Email: aleconte@kecksci.claremont.edu

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This work was supported by the National Science Foundation (CHE-1752924 CAREER award to A.M.L.).

Notes

The authors declare no competing financial interests.

ABSTRACT: DNA is a foundational tool in biotechnology and synthetic biology but is limited by sensitivity to DNA-modifying enzymes. Recently, researchers have identified DNA polymerases that can enzymatically synthesize long oligonucleotides of modified-DNA (M-DNA) that are resistant to DNA-modifying enzymes. Most applications require M-DNA to be reverse transcribed, typically using a RNA reverse transcriptase, back into natural DNA for sequence analysis or further manipulation. Here, we tested commercially available DNA-dependent DNA polymerases for their ability to reverse transcribe and amplify M-DNA in a one-pot reaction. Three of the six polymerases chosen (Phusion, Q5, and Deep Vent) were able to reverse transcribe and amplify synthetic 2'F M-DNA in a single reaction with $< 5\times 10^{-3}$ errors per base pair. We further used Q5 DNA polymerase to reverse transcribe and amplify M-DNA synthesized by two candidate M-DNA polymerases (SFP1 and SF4-6), allowing for quantification of the frequency, types, and locations of errors made during M-DNA synthesis. From these studies, we identify SFP1 as one of the most accurate M-DNA polymerases identified to date. Collectively, these studies establish a simple, robust method for conversion of 2'F M-DNA to DNA in less than one hour using commercially available materials, significantly improving the ease of use of M-DNA.

ABBREVIATIONS

DNA – deoxyribonucleic acid
nt – nucleotide
IR – infrared
RT – reverse transcription
RTase – reverse transcriptase
PCR – polymerase chain reaction
dNTPs – deoxyribonucleoside triphosphates
NTPs – nucleoside triphosphates

INTRODUCTION

DNA polymerases efficiently amplify DNA without loss of information, enabling DNA to be used as a foundational tool in biotechnology,[1] nanotechnology,[2] clinical diagnostics,[3] and synthetic biology.[4-5] However, DNA is a heavily regulated and controlled natural biomolecule; many native DNA editing enzymes, such as nucleases, can distort potential signals in experiments that rely on DNA. Due to this limitation, researchers have spent considerable time and effort to identify chemical variants of DNA that have the beneficial amplifiable information coding properties of DNA without the susceptibility to nucleases.[6] These modified DNA (M-DNA) have potential uses in point-of-care diagnostics, therapeutics, and a range of applications.[7-9]

Many M-DNA, especially DNA modified at the 2' position (2'M-DNA) or other positions on the sugar moiety, have been shown to be resistant to nucleases.[10-11] While natural DNA and RNA polymerases do not readily synthesize these M-DNAs, recent efforts have identified a number of DNA polymerase mutants that can synthesize a broad range of M-DNA.[12-15] However, considerably less effort has been expended in identifying, developing and quantitatively assessing candidate enzymes that can convert the M-DNA back into DNA (M-DNA reverse transcriptase, M-DNA-RTase). Accurate and efficient M-DNA-RTases are highly valuable and necessary for sequencing, molecular cloning, and most M-DNA applications.

To date, most M-DNA-RTases are either natural RNA reverse transcriptases[16-17] or DNA polymerases developed during M-DNA polymerase engineering and repurposed as M-DNA-RTases.[18-19] However, these M-DNA-RTase enzymes are often highly error-prone[18] and frequently require overnight incubation periods and/or the use of manganese,[19-20] both of which are likely to decrease the fidelity of M-DNA reverse transcription. The low efficiency and inaccuracy of reverse transcription can be limiting for applications of these M-DNA.

Recently, several groups have evaluated commercially available DNA-dependent DNA polymerases as candidate M-DNA-RTases. Surprisingly, some commercially available DNA polymerases have proven adept at reverse transcribing heavily modified forms of M-DNA, and, in several examples, appear superior to RNA RTases.[17] For example, threose nucleic acid (TNA), a type of 2'M-DNA, can be reverse transcribed by Bst with 80% efficiency in two hours with an error rate of 3.8×10^{-3} errors per base pair.[17] While RNA is not nuclease resistant, understanding reverse transcription of RNA is useful due to its similar chemical nature to many 2'M-DNA. Notably, commercially available DNA polymerases have also been demonstrated to be capable of RNA reverse transcription and, in some cases, to outperform natural RNA RTases.[21]

While there have been some efforts to identify more efficient and accurate enzymes, M-DNA-RTases have been incompatible with one-pot reverse transcription and amplification reaction due to either poor efficiency, poor fidelity, or lack of thermal stability of these enzymes. However, a one-pot reaction where M-DNA could be both reverse transcribed and amplified in a single step would significantly ease utility and facilitate application of M-DNA in clinical diagnostics, especially for point-of-care, and may enable new uses in synthetic biology and clinical diagnostics that are not possible with current technologies.

In this study, we assess whether commercial DNA polymerases have the latent ability to both reverse transcribe and amplify 2' modified nucleotides in a single reaction. 2' fluoro modified nucleic acids, which have been referred to as either 2' fluoro-modified DNA[15,19] or 2' fluoro-modified RNA (here, for clarity, we refer to this polymer as 2' fluoro-modified DNA as our synthesis enzyme is a DNA polymerase mutant and there are no 2'OH substitution on any nucleotides of the oligonucleotide polymer) are among the most frequently used M-DNAs, and has been used in miRNA[22] and in aptamers containing partially 2'F substituted DNA or RNA.[23-26]

Further, 2'F M-DNA is widely commercially available as both the nucleoside triphosphate and as a phosphoramidite, making it one of the most broadly accessible M-DNAs currently in use. Most commonly, 2'F M-DNA is used in a mixed polymer where some of the nucleotides are 2'F modified and the other nucleotides are DNA or RNA.[23-26] These partially substituted M-DNAs are then usually reverse transcribed by a natural RTase and then amplified in a separate reaction.[23-26] While partially substituted 2'F M-DNA has been the focus of 2'F M-DNA applications due to a limited ability of RNA polymerase or DNA polymerases to synthesize fully substituted 2'F M-DNA, recent work has shown that mutant M-DNA polymerases can synthesize long (80nt) fully substituted 2'F M-DNA.[15] Importantly, this synthesis can be performed without manganese using substoichiometric enzyme levels with relatively short extension times (1 hour). These are significantly more mild conditions than are typically used for other M-DNAs making 2'F M-DNA better suited to applications in synthetic biology that encode information *ex vivo* and, potentially, eventually *in vivo*. However, the error rate of these M-DNA syntheses have not been quantitatively evaluated to date.

Here, we systematically evaluate commercially available DNA-dependent DNA polymerases for the ability to both reverse transcribe and then amplify fully-substituted 2'F M-DNA efficiently and accurately. Development of one-pot reverse transcription and amplification reactions of 2'F M-DNA would further improve the ease of use and broaden the potential applications of this polymer.

MATERIALS / EXPERIMENTAL DETAILS

M-DNA reverse transcription Assay

Reaction mixture contained 5' IRDye700-labeled primer, P24 (40 nM, IDT DNA; see Table S1 for sequence), either T57 or T57(24F) (80 nM, IDT DNA; see Table S1 for sequence), and the enzyme manufacturer's recommended reaction buffer (Table S2). To anneal the primer/template, reaction mixture was heated to 98° C for 2 minutes and cooled to 40° C over 15 minutes. Nucleoside triphosphates (200 μM; New England Bio Labs) and variable enzyme (0.01 U/μL, all enzymes from New England Biolabs) were added, and the mixture was incubated at 50° C on a pre-equilibrated heat block for variable times up to 60 minutes and quenched with two volume equivalents of QBO (Formamide (95%; Acros), EDTA (12.5 mM, Sigma Aldrich), Orange G (< 0.1 mg/mL; Sigma Aldrich)). Quenched samples were incubated at 100° C for 5 minutes with urea (6 M; Research Products International) and analyzed on a 10% TBE-Urea polyacrylamide gels (Bio-Rad). The gels were imaged on an Odyssey CLx gel imager (Li-Cor, Lincoln, NE) and band intensitites were quantified using ImageStudio software (Li-Cor). Each experiment was performed twice and all qualitative trends were reproducibly observed.

Enzymatic M-DNA synthesis

Reactions were performed with 5' IRDye700-labeled primer K017(40 nM; see Table S1 for sequence), template K021 (80nM; see Table S1 for sequence) in Tris buffer (50 mM, pH 8.5; Fisher), $MgCl_2$ (6.5mM; Fisher Scientific), Acetylated BSA (0.05 mg mL^{-1}; Promega), KCl (50 mM; Fisher Scientific). Reaction mixture was annealed by heating to 98°C for 2:05 minutes then slowly cooled as follows: 80°C for 2:00, 70°C for 3:00, 65°C for 2:00, 60°C for 1:00, 55°C for 1:00, 50°C for 2:00, 45°C for 2:00, 40°C for 2:00. Either milli-Q water, dNTPs (200μM, New England Biolabs), or 2'F modified NTPs (200μM; TriLink Biotechnologies) were added and either SFM4-6 or SFP1 enzyme (20 nM). Reactions were incubated at 50°C on a heat block for two

hours. From each reaction, a 3 μL analytical aliquot was removed and quenched with two volume equivalents of QBO ((Formamide (95%; Acros), EDTA (12.5 mM, Sigma Aldrich), Orange G (< 0.1 mg/mL; Sigma Aldrich)) and observed on a 10% TBE-Urea gel (Bio-Rad) to confirm synthesis.

M-DNA purification by DNase treatment

To the remaining volume of reaction mixture described in B i., Turbo DNase was added (final concentration 0.11 U/μL; Invitrogen) and samples were incubated at 37°C on a heat block for 40 minutes. Samples were then purified according to manufacturer's protocol using Qiaquick Nucleotide Removal Kit (Qiagen) and samples were used directly in RT/Amp (described in Methods of the main manuscript).

RT/Amp Reaction for enzymatically synthesized, fully substituted 2'F M-DNA

PCRs were performed using barcoded primers to allow multiplexing when performing next-generation HT-Seq (see Table S4 for sequences). Each 50 μL reaction contained 1 μL purified M-DNA or equivalent (see Supporting Information for details on M-DNA synthesis and purification), 1x Q5 Reaction Buffer (New England Biolabs), 0.5 μM of each primer, 0.4 mM dNTPs (New England Biolabs), 6% DMSO (Fisher Scientific), milliQ purified water, and 0.02 U Q5 DNA polymerase (New England Biolabs). PCR was performed using 1 RT cycling conditions (see Table S5) in a C1000 Touch Thermal Cycler (Bio-Rad). PCR products were purified using DNA Clean and Concentrator-25 kit (Zymo Research), according to manufacturer's protocol for PCR products. Products were visualized on 2% agarose gel containing ethidium bromide and a benchtop UV transilluminator. DNA concentration was quantified using Qubit 3 Fluorometer and Qubit dsDNA

HS Assay Kit (ThermoFisher Scientific). Samples were submitted to Genewiz for Amplicon-EZ sequencing and analyzed using a custom Python script (see Supporting Information).

RESULTS AND DISCUSSION

First, we assessed the ability of different commercially available DNA polymerases to reverse transcribe 2'F modified nucleic acids. We designed and synthesized a 57 nucleotide (nt) template that contains 24 consecutive 2'F modified nucleotides in the center of a DNA oligonucleotide T57(24F) (Figure 1A); the length of 2'F modified portion of the sequence was limited by the capabilities of commercial oligonucleotide synthesis. Using a near-IR fluorophore labeled primer (P24) that anneals to the DNA portion of the template and the first 8nt of the 2'F region allows us to directly observe reverse transcription of the subsequent 16 2'F nt by the polymerases (Figure 1A). As a comparison and control for polymerase activity, we also synthesized a 57nt unmodified DNA template (T57) with an identical sequence to T57(24F); we performed all experiments in parallel with both the T57 and T57(24F) templates.

We evaluated a panel of six commercial DNA polymerases for their ability to extend a DNA primer using the T57(24F) template. For each enzyme, we incubated 1U of the enzyme with either a natural or modified template in the commercial supplier's recommended buffer at 50 °C for varying times up to 60 minutes. Surprisingly, all of the DNA polymerases are able to reverse transcribe the 2'F M-DNA with at least 3% yield of full-length product in 15 minutes, albeit to varying degrees (Figure 1B, Figure S1). Each showed at least 5% conversion to the full-length product in one hour (Figure S1). The most efficient enzyme for reverse transcription was Bst 3.0, which was able to reverse transcribe the primer most completely (approximately 90% conversion of the primer to full-length product) and most efficiently (even at 5 minutes, a majority of the

primer was converted to full-length product, far outpacing the other enzymes). The least efficient enzymes were Deep Vent and Vent which, even after 60 minutes, were only able to convert approximately 10% of the primer to full-length product. Phusion, Q5, and Therminator DNA Polymerase were able to perform the task at intermediate levels. Importantly, these reactions were conducted under simple, manufacturer's recommended buffers and concentrations and did not require addition of manganese or long incubation times.

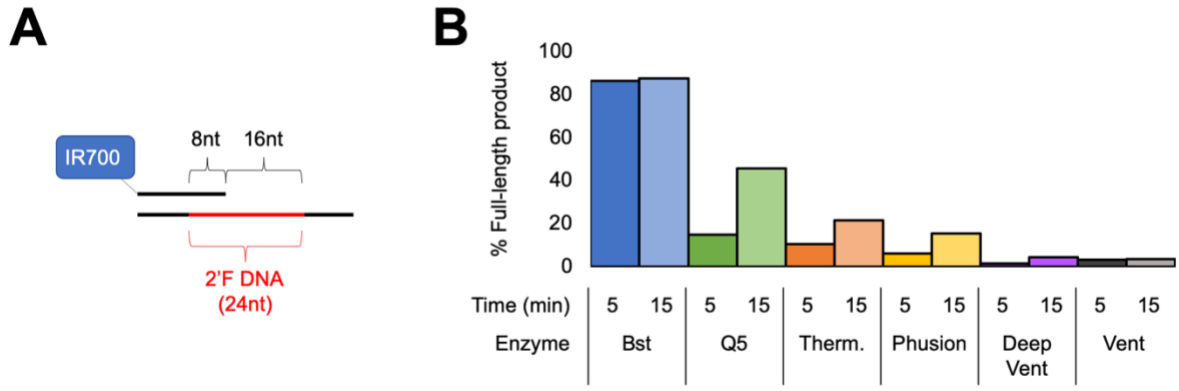


Figure 1. Reverse transcription of 2'F M-DNA by commercial DNA polymerases. (A) Primer-template construct used for characterization of reverse transcription. (B) Full-length product observed in RT reactions by different DNA polymerases after 5 and 15 minutes. Therm. is Therminator.

Interestingly, past studies surveying commercial DNA polymerases for RT ability of RNA[21] and TNA[16-17] both identified Bst as the most active RTase from a broad panel of enzymes. With RNA, Bst was shown to be more efficient under some conditions than commonly used RNA RTases, such as AMV RTase.[21] Further, Bst 3.0 is an efficient RTase for two types of M-DNA (TNA[10] and FANA[27]). Thus, our data add to a rapidly growing body of literature demonstrating Bst may be broadly useful for reverse transcription of a number of M-DNA substrates. Recent structural studies[27] show that the conformation of the enzyme during M-DNA reverse transcription is different when binding different substrates, implying that Bst may possess structural plasticity during DNA synthesis that enables its ability to reverse transcribe a range of M-DNA substrates.

Further studies will help elucidate whether this is unique to Bst or whether this occurs in other DNA polymerases.

While these results indicate that many DNA polymerases can reverse transcribe 2'F M-DNA, it is challenging to interpret the relative efficiency of these enzymes since the temperature used (50° C) is suboptimal for most of the enzymes. It is possible that Bst 3.0 shows better activity due to its lower optimal temperature relative to the other enzymes and, at higher temperatures, other enzymes may be more efficient. Most importantly, from these data, we conclude that all of these enzymes possess the latent ability to perform the reaction, greatly expanding the possible enzymes that may be used with 2'F M-DNA.

Intriguingly, several of the polymerases capable of reverse transcription are commonly used in PCR, suggesting the possibility of one-pot reverse transcription and amplification (RT/Amp) reactions. If feasible, this one-pot method would be operationally simpler, more robust, and more time and resource efficient than traditional reverse transcription methods. To assess the ability of these enzymes to perform RT/Amp, we evaluated each enzyme in a PCR containing either T57 or T57(24F) as a template. To examine amplification of as many 2'F modified nucleotides as possible, primers were designed that overlap only with the DNA portion of T57, and PCRs were performed under the standard manufacturer recommended conditions for each enzyme if available.

Phusion, Q5, Vent, and Deep Vent were able to reverse transcribe and amplify the 2'F M-DNA under standard PCR conditions (Figure 2A, Figure S2). These reactions were all performed under mild reaction conditions typical of normal PCR; the buffer used was from the manufacturer, and the reaction required only 40 minutes for Phusion and Q5. Bst 3.0 and Therminator were not able to RT/Amp T57(24F) but were also unable to amplify natural DNA, indicating that these enzymes are not suitable for PCR, in general, which, for both enzymes, has been well-established.[28] Both

enzymes showed strong reverse transcription abilities and may be useful in a mixture of enzymes in a one-pot reaction,[21] which has previously been done with RNA, or in other applications that do not require amplification.

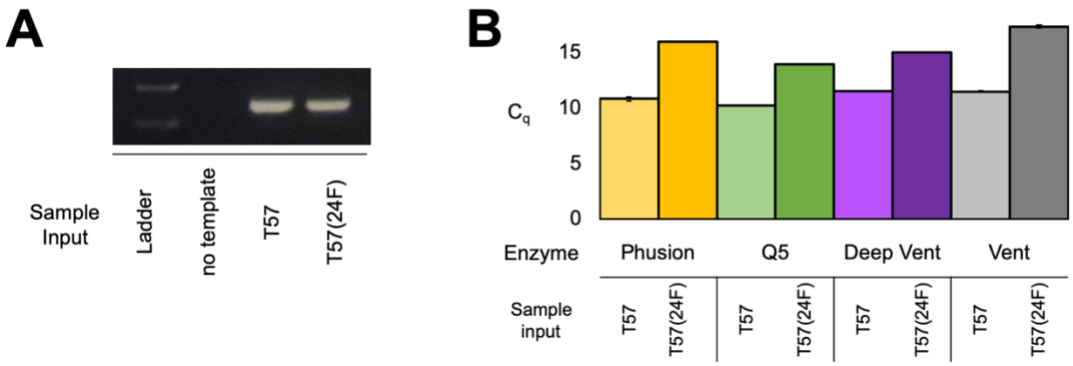


Figure 2. Commercial DNA polymerases can reverse transcribe and amplify 2'F M-DNA in a one-pot reaction. (A) PCR product created by Phusion DNA polymerase. 'no template' indicates a negative control where template is not added, but all other PCR components are added. (B) C_q of PCR amplification of T57 and RT/amplification of T57(24F) under optimized conditions for each enzyme.

To optimize RT/Amp reactions, we systematically varied [Mg^{2+}] and cycling conditions; we observed that RT/Amp of 2'F M-DNA was more sensitive to changes in either of these parameters than PCR amplification of unmodified DNA (Supporting Results, Figure S3). Under the optimal conditions for each enzyme, we observe C_q values for the RT/Amp that are within three to six cycles of the equivalent amplification of natural DNA (Figure 2B). These data show that the reverse transcription of 2'F M-DNA is relatively facile and occurs at efficiencies that are within one to two orders of magnitude of natural DNA-templated DNA synthesis. Perhaps most importantly, this demonstrates that a one-pot reverse transcription and amplification reaction is possible in less than an hour. Typically, M-DNA or RNA reverse transcription reactions require approximately one hour for the reverse transcription reaction, and then require purification followed by a separate PCR amplification. This one-pot reaction is much more efficient and

significantly operationally simpler, which will be helpful in potential applications in synthetic biology and point-of-care diagnostics.

Since RTases are relatively error prone while most of these DNA polymerases are considered high fidelity, we also examined the accuracy of each enzyme during the RT/Amp reaction. We submitted the products of the RT/Amp reaction for next generation high-throughput DNA sequencing (HT-Seq). For each RT/Amp reaction, at least 8,000 reads including at least 200,000 reverse transcribed nucleotides were used in the analysis. Based on our efficiency experiments, we systematically evaluated each polymerase under three different conditions. Although efficiency of the RT/Amp reaction was sensitive to $[Mg^{2+}]$ and cycling conditions, we did not see substantial differences in the error rate under different buffer conditions or different cycling conditions (Supporting Results, Figure S4).

For Phusion, Q5, and Deep Vent, we observed that at least 90% of the sequence reads originating from T57(24F) matched the original sequence (Figure 3A), corresponding to error rates ranging from 2.7×10^{-3} to 4.2×10^{-3} errors per base pair (epbp); this number reflects the error rate for all steps including chemical synthesis, reverse transcription, and amplification. Q5 was the most accurate of the enzymes although the difference was relatively small. These values were only slightly less accurate than the amplification of T57 that was done in parallel; Phusion, Q5, and Deep Vent all showed >95% matched sequence reads when using T57 as a template, reflecting error rates that ranged from 1.6×10^{-3} to 2.2×10^{-3} epbp.

In contrast, Vent showed a markedly less accurate synthesis for both T57(24F) and T57. With T57(24F), Vent performed RT/Amp with an error rate of 7.5×10^{-3} epbp, leading to only 84% of the sequences matching the correct sequence. Similarly, Vent was fairly error prone when amplifying T57, which lead to only 89% retention of the correct sequence during amplification.

Because of the volume of sequence data generated, we were able to quantify the types of errors that were observed. For Phusion, Q5, and Deep Vent, deletions composed the majority of errors observed (Figure 3B). This is likely to reflect, at least in part, deletions that originate during chemical synthesis, which are the most typical type of errors during chemical synthesis.[29] In contrast, Vent's most frequent errors were substitutions, which largely accounts for the increased error rate of Vent, and indicates that the main source of the errors is likely the polymerase itself rather than chemical synthesis.

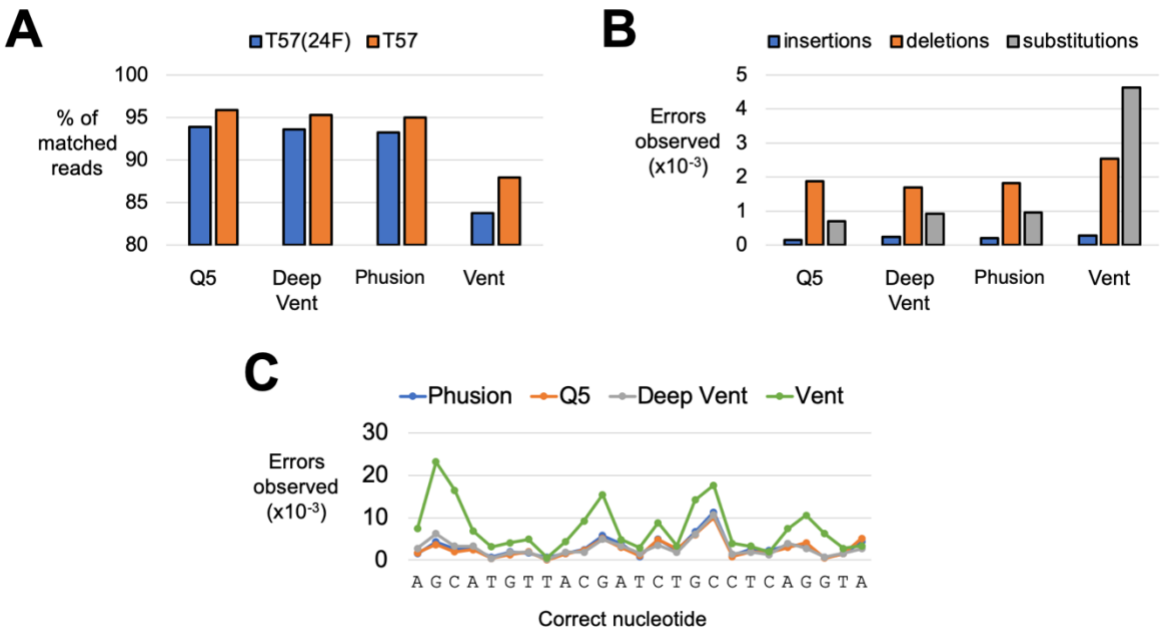


Figure 3. High-throughput DNA sequencing of RT/Amp products. (A) Percentage of reads that match the initial sequence using different DNA polymerases. (B) Types of errors observed with different DNA polymerases. (C) Errors observed at individual template positions in the initial sequence. All reactions were performed using buffer supplemented with Mg^{2+} and 1RT cycling conditions (see SI).

We also evaluated the sequence context in which errors were made by calculating the error rate of each position of the test sequence. Interestingly, the positional dependence of Phusion, Q5, and Deep Vent look nearly identical (Figure 3C). This stark similarity may further indicate that the errors observed are largely originating from the chemical synthesis rather than the RT/Amp

reaction since such similar profiles are unlikely to originate from reverse transcription and amplification alone. For Vent, six positions of the sequence were incorrectly synthesized at least 10×10^{-3} epbp; all of these encoded synthesis of either C or G (Figure 3C). Interestingly, the effect is highly sequence dependent as several positions encoding C or G were synthesized by Vent with error rates $< 5 \times 10^{-3}$ epbp.

Overall, these data indicate that RT/Amp reactions proceed at high fidelity for Q5, Phusion, and Deep Vent, but not when using Vent. For Q5, Phusion, and Deep Vent, the difference between RT/Amp of T57(24F) and amplification of T57 is approximately 1×10^{-3} to 2×10^{-3} epbp. From these experiments, it is not clear that this very small error rate difference is from the enzymatic reverse transcription or, possibly, due to errors during chemical synthesis of 2'F M-DNA. Considering that a majority of the errors were deletions and that 2'F M-DNA chemical synthesis is more challenging than chemical synthesis of DNA, it seems likely that chemical synthesis may contribute to the small difference between RT/Amp of 2'F M-DNA and amplification of DNA, if not explain the difference entirely.

Gratifyingly, the overall error rates for M-DNA RT/Amp are comparable to those observed in one-step RT/Amp of RNA,[30] including by commercially available kits such as Superscript IV One-Step PCR System. Notably, when they have been directly compared, past studies observing M-DNA reverse transcription have often shown larger differences in error rates between DNA-templated DNA synthesis and M-DNA-templated DNA synthesis.[16, 20] Most importantly, the error rate of the RT/Amp performed by Q5, Phusion, and Deep Vent is quite small, showing unambiguously that these enzymes can, in a single reaction, reverse transcribe and amplify these M-DNAs with accuracy that is equal to or better than reverse transcription of RNA by natural RTases.

In our initial experiments, we evaluated these properties on an oligonucleotide that possessed 24 consecutive 2'F modified nucleotides, but still contained DNA in the primer binding regions. Considering the efficiency and accuracy of the one-pot RT/Amp reaction on a synthetic substrate, we wondered whether a longer, enzymatically synthesized, fully substituted M-DNA, more typical of those used in many M-DNA applications, could also be a substrate for the reaction. Importantly, unlike T57(24F), an enzymatically synthesized 2'F M-DNA is completely substituted with modified nucleotides. Recently, we reported that both SFP1 and SFM4-6 are able to synthesize 2'F M-DNA of 80nt length in one hour. A qualitative dropout fidelity assay indicated that SFP1 may be more accurate than SFM4-6.[15] Thus, we sought to both evaluate the ability of DNA polymerases to perform this RT/Amp reaction on a longer, enzymatically synthesized, fully substituted M-DNA substrate and apply the RT/Amp reaction to calculate a quantitative error rate and error spectrum for these two enzymes, which has not been previously calculated for these 2'F M-DNA syntheses.

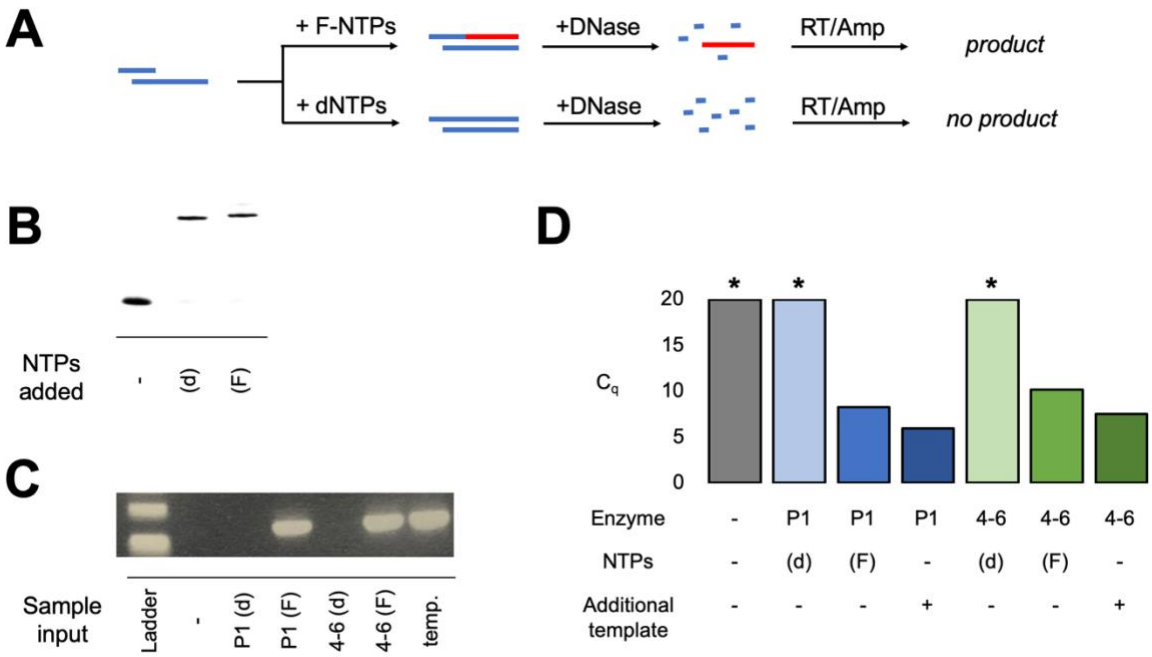


Figure 4. Enzymatic synthesis and RT/Amp of 2'F M-DNA. (A) Scheme depicting workflow with either 2'F NTPs or dNTPs. (B) SFP1 mediated synthesis of DNA or 2'F M-DNA. 20nM enzyme was incubated with 40nM DNA and 200μM NTPs for 2 hr. (C) RT/Amp of DNA and M-DNA generated using different M-DNA synthesis enzymes and NTPs. 'temp.' indicates a control reaction where template is added immediately prior to PCR. (D) Quantitative RT/Amp of DNA and M-DNA generated using different enzymes and NTPs. * indicates the critical threshold was not crossed during the experiment. For all figures, (F) indicates 2'F NTPs were used in the synthesis; (d) indicates dNTPs were used in the synthesis. 'Additional template' or 'temp.' indicates that template was added prior to PCR as a control; in these control experiments, no M-DNA or DNA synthesis was performed. (n=3).

We developed a simple workflow to synthesize M-DNA, purify the M-DNA using nuclease digestion, and amplify the sequences using RT/Amp (Figure 4A). Using previously described conditions,[15] in parallel reactions, we synthesized M-DNA with the two enzymes using a longer template encoding 80nt of synthesis (Figure 4B). For all experiments, parallel reactions were performed with either no NTPs or with dNTPs as controls. Following synthesis, all samples were incubated with TurboDNase, an endonuclease commonly used to remove DNA from in vitro transcription reactions, to remove the unmodified DNA primer and template strands; observation by TBE-Urea PAGE gel showed that no DNA primer remained (Figure S5). Following purification to remove DNase, we subjected all samples to RT/Amp by Q5 DNA polymerase; we

chose Q5 DNA polymerase as it was most accurate in prior model experiments. As expected, only the 2'F M-DNA containing sample showed PCR product (Figure 4C); the DNA sample did not show product indicating that the DNA had been degraded in the workflow. Additional control experiments varying DNAse digestion confirm that the resulting PCR product in the 2'F sample is not due to contaminating template (Supporting Results, Figure S6). Quantification of this PCR showed a C_q value for 2'F M-DNA ranging from 8-10 cycles; the DNA control shows no product after 20 cycles (Figure 4D). A control in which template was added after DNA digestion and M-DNA purification has C_q values of 5-6 (Figure 4D), which mirrors our ΔC_q with RT/Amp of T57(24F) and T57. Collectively, these results suggest that completely substituted 2'F M-DNA can be reverse transcribed and amplified efficiently.

To quantify the error rates between SFM4-6 and SFP1, we subjected the PCR products to HT-Seq. To estimate the errors introduced during chemical synthesis, amplification, and HT-Seq, we also amplified the template of the synthesis reaction using PCR and submitted it for HT-Seq as a control. Our sequencing analysis contained >14,000 reads for each of the three samples, resulting in >650,000 total nucleotides sequenced in each condition.

SFP1 shows a significantly higher accuracy for the M-DNA synthesis relative to SFM4-6. We observed 19.1×10^{-3} epbp when SFM4-6 synthesizes the M-DNA, 5.3×10^{-3} epbp when SFP1 synthesizes the M-DNA, and 3.0×10^{-3} epbp in a template amplification control where the template was PCR amplified in parallel to the RT/Amp reaction (Figure 5A). Relative to the template control, SFM4-6 shows a 6-fold increase in insertions, a 3-fold increase in deletions, and a 27-fold increase in substitutions while SFP1 shows approximately equal frequency for insertions and deletions, but a 6-fold increase in substitution errors.

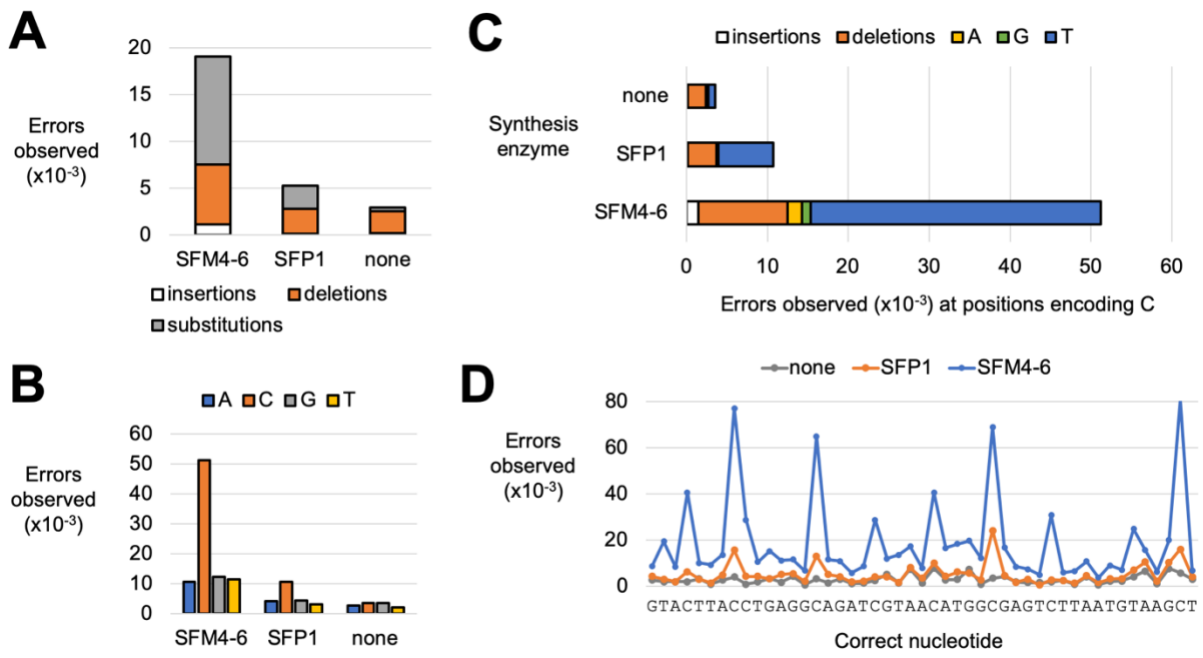


Figure 5. Accuracy of M-DNA synthesis determined by high-throughput sequencing. (A) Total errors observed for each synthesis enzyme and template control. (B) Frequency of errors observed for each encoded nucleotide. (C) Types of errors observed at positions encoded by C. (D) Frequency of errors at each template position of the synthesis reaction. For all reactions, none indicates the template amplification control.

To further examine this, we next asked whether the templating nucleotide influences the error rate for each reaction. While the template amplification control had fairly similar error rates independent of the encoded nucleotide (ranging from 2.2 to 3.6 × 10^{-3} epbp), both SFP1 and SFM4-6 show dramatically increased error rates at positions encoding C incorporation; relative to the template control, SFP1 has a 3-fold increase and SFM4-6 has a 13-fold increase (Figure 5B). Importantly, these differences were not observed in prior studies with T57(24F), indicating that these errors originate in the M-DNA synthesis. Thus, while the template control generates 3.5 × 10^{-3} epbp when C incorporation is encoded by the template, SFP1 generates 10.7 × 10^{-3} epbp and SFM4-6 generates 51.2 × 10^{-3} epbp. At positions encoding T, G, or A, SFP1 possesses an error rate within 2-fold of the template amplification control; in contrast, SFM4-6 shows 3.5- to 5-fold increase at these other positions. Thus, it appears that errors introduced by SFP1 largely relate to

misincorporation against template dG while SFM4-6 possesses an increased error rate at all positions of the template.

We can further examine what types of errors are made when dG is the template nucleotide (Figure 5C). For the template amplification control, a majority of the errors are deletions (2.3×10^{-3} epbp), which likely reflects errors made during chemical synthesis. For SFP1, both increases in misincorporation of 2'F-TTP against dG (6.8×10^{-3} epbp) and deletions at this position (3.6×10^{-3} epbp) are observed. Similarly, SFM4-6 shows the greatest increase in 2'F-TTP misincorporation (36×10^{-3} epbp) and deletions (11×10^{-3} epbp) when the template is dG. Collectively, these data demonstrate that misincorporation of 2'F-TTP against dG is the most prevalent source of errors during M-DNA synthesis by both of these enzymes, and that the error rate of SFM4-6 when dG is the template nucleotide (51.2×10^{-3} epbp) may limit some applications of this enzyme.

To further evaluate whether this effect is sequence dependent, we also examined the specific positions where mutations were made in the sequence (Figure 5D). Consistent with the aggregate data, SFP1 makes fewer errors than SFM4-6 at every template position. Also consistent with the bulk error rates, for SFM4-6, the nine most incorrectly synthesized positions all encode C (there are nine C's in the sequence). For SFP1, the five most incorrectly synthesized positions are also C; however, some C-encoding positions possess a much lower error rate. This suggests that there is some sequence dependence to the errors created for both enzymes. The template control shows only two of top nine positions are C, indicating that the bias towards errors at positions encoding C are likely for the M-DNA synthesis or reverse transcription rather than chemical synthesis, amplification, or HT-Seq.

Collectively, the use of this HT-Seq assay allows us to obtain detailed and quantitative information about the error frequency and spectrum of these two M-DNA polymerases. SFP1 creates approximately 3-fold fewer errors than SFM4-6, which ensures that nearly 80% of sequences that go through a cycle of chemical synthesis of DNA, enzymatic synthesis of M-DNA, reverse transcription, amplification of M-DNA, and HT-Seq retain the original sequence information. Notably, this compares favorably to a template only control, which has gone through all steps except M-DNA synthesis and reverse transcription, and which results in 90% sequence retention through this process. Additionally, the overall error rate of 5.3×10^{-3} epbp is similar to the error rate of the most accurate M-DNA polymerase previously identified.[31] In contrast, only 44% of sequences synthesized by SFM4-6 possess a retained sequence through the same process.

The use of the HT-Seq assay also allows us to understand the specific error spectra of these enzymes and may inform applications. For example, when performing SELEX with M-DNA, a more error-prone polymerase such as SFM4-6 may be better suited; notably the primary application of SFM4-6 and related polymerases to date has been in aptamer identification.[23, 32-33] However, for applications requiring accurate synthesis, including synthetic biology applications where this information may be synthesized and recovered, SFP1 may be superior.

Importantly, this quantitative assay shows that the distribution of errors made by these M-DNA polymerases is not symmetrical and may lead to bias; fortunately, the simple assay described here should enable the rapid optimization of M-DNA synthesis to provide a more even distribution of errors, possibly by manipulating the [F-NTP] in a manner similar to how error-prone PCR was optimized.[34] Although additional work is needed, the difference in error rates and error spectra described here may also have applications in synthetic biology by differential expression of these M-DNA polymerases.[35]

CONCLUSION

Overall, the work described here identifies a number of commercially available DNA polymerases that can reverse transcribe and amplify either chemically or enzymatically synthesized 2'F M-DNA in a one-pot reaction in less than one hour. This represents a dramatic improvement in the ease of use, time required, and accuracy of the conversion of M-DNA into DNA. We have also shown here that 2'F M-DNA can be synthesized, purified, reverse transcribed and amplified using widely available commercial reagents, with the sole exception of the M-DNA polymerase. Further, application of these simple and robust methods have shown that SFP1 is quantitatively more accurate than SFM4-6, providing guidance in application and further engineering of these different M-DNA polymerase enzymes in the broader synthetic biology community.

Accession ID (UniProt):

Phusion, Therminator, and Q5 DNA polymerases are proprietary enzymes from New England Biolabs.

Deep Vent DNA polymerase (Q51334)

Vent DNA polymerase (P30317)

Bst 3.0 is a proprietary mutant of Bst DNA polymerase I (P52026)

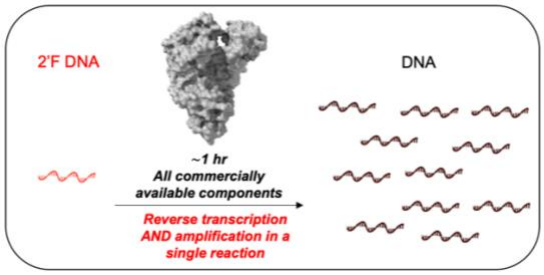
SFP1 and SFM4-6 are both mutants of Taq DNA polymerase I (P19821)

Supporting Information. Supporting Results, Methods, Figures S1-S7, and Tables S1-S5. This material is available free of charge via the Internet at http://pubs.acs.org.

For Table of Contents Use Only

Accurate and efficient one-pot reverse transcription and amplification of 2' fluoro-modified DNA by commercial DNA polymerases

*Arianna S. Thompson, Susanna E. Barrett, Aurora G. Weiden, Ananya Venkatesh, Madison K. C. Seto, Simone Z. P. Gottlieb and Aaron M. Leconte**



REFERENCES

1. Aschenbrenner, J.; Marx, A., DNA polymerases and biotechnological applications. *Curr Opin Biotechnol* **2017,** *48*, 187-195.
2. Chidchob, P.; Sleiman, H. F., Recent advances in DNA nanotechnology. *Curr Opin Chem Biol* **2018,** *46*, 63-70.
3. Kanwar, J. R.; Roy, K.; Maremanda, N. G.; Subramanian, K.; Veedu, R. N.; Bawa, R.; Kanwar, R. K., Nucleic acid-based aptamers: applications, development and clinical trials. *Curr Med Chem* **2015,** *22* (21), 2539-57.
4. Dien, V. T.; Morris, S. E.; Karadeema, R. J.; Romesberg, F. E., Expansion of the genetic code via expansion of the genetic alphabet. *Curr Opin Chem Biol* **2018,** *46*, 196-202.
5. Cameron, D. E.; Bashor, C. J.; Collins, J. J., A brief history of synthetic biology. *Nat Rev Microbiol* **2014,** *12* (5), 381-90.
6. Keefe, A. D.; Cload, S. T., SELEX with modified nucleotides. *Curr Opin Chem Biol* **2008,** *12* (4), 448-56.
7. Rangel, A. E.; Chen, Z.; Ayele, T. M.; Heemstra, J. M., In vitro selection of an XNA aptamer capable of small-molecule recognition. *Nucleic Acids Res* **2018,** *46* (16), 8057-8068.
8. Taylor, A. I.; Arangundy-Franklin, S.; Holliger, P., Towards applications of synthetic genetic polymers in diagnosis and therapy. *Curr Opin Chem Biol* **2014,** *22*, 79-84.
9. Alves Ferreira-Bravo, I.; Cozens, C.; Holliger, P.; DeStefano, J. J., Selection of 2'-deoxy-2'-fluoroarabinonucleotide (FANA) aptamers that bind HIV-1 reverse transcriptase with picomolar affinity. *Nucleic Acids Res* **2015,** *43* (20), 9587-99.
10. Appella, D. H., Non-natural nucleic acids for synthetic biology. *Curr Opin Chem Biol* **2009,** *13* (5-6), 687-96.
11. Prakash, T. P., An overview of sugar-modified oligonucleotides for antisense therapeutics. *Chemistry & biodiversity* **2011,** *8* (9), 1616-41.
12. Chen, T.; Romesberg, F. E., Directed polymerase evolution. *FEBS letters* **2014,** *588* (2), 219-229.
13. Houlihan, G.; Arangundy-Franklin, S.; Holliger, P., Engineering and application of polymerases for synthetic genetics. *Curr Opin Biotechnol* **2017,** *48*, 168-179.
14. Schultz, H. J.; Gochi, A. M.; Chia, H. E.; Ogonowsky, A. L.; Chiang, S.; Filipovic, N.; Weiden, A. G.; Hadley, E. E.; Gabriel, S. E.; Leconte, A. M., Taq DNA Polymerase Mutants and 2'-Modified Sugar Recognition. *Biochemistry* **2015,** *54* (38), 5999-6008.
15. Rosenblum, S. L.; Weiden, A. G.; Lewis, E. L.; Ogonowsky, A. L.; Chia, H. E.; Barrett, S. E.; Liu, M. D.; Leconte, A. M., Design and discovery of new combinations of mutant DNA polymerases and modified DNA substrates. *Chembiochem* **2017,** *18* (8), 816-823.
16. Chaput, J. C.; Ichida, J. K.; Szostak, J. W., DNA polymerase-mediated DNA synthesis on a TNA template. *J Am Chem Soc* **2003,** *125* (4), 856-7.
17. Dunn, M. R.; Chaput, J. C., Reverse Transcription of Threose Nucleic Acid by a Naturally Occurring DNA Polymerase. *Chembiochem* **2016,** *17* (19), 1804-1808.
18. Pinheiro, V. B.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S. Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P., Synthetic genetic polymers capable of heredity and evolution. *Science* **2012,** *336* (6079), 341-4.

19. Chen, T.; Hongdilokkul, N.; Liu, Z.; Adhikary, R.; Tsuen, S. S.; Romesberg, F. E., Evolution of thermophilic DNA polymerases for the recognition and amplification of C2'-modified DNA. *Nat Chem* **2016,** *8* (6), 556-62.
20. Tsai, C. H.; Chen, J.; Szostak, J. W., Enzymatic synthesis of DNA on glycerol nucleic acid templates without stable duplex formation between product and template. *Proc Natl Acad Sci U S A* **2007,** *104* (37), 14598-603.
21. Shi, C.; Shen, X.; Niu, S.; Ma, C., Innate Reverse Transcriptase Activity of DNA Polymerase for Isothermal RNA Direct Detection. *J Am Chem Soc* **2015,** *137* (43), 13804-6.
22. Wilson, C.; Keefe, A. D., Building oligonucleotide therapeutics using non-natural chemistries. *Curr Opin Chem Biol* **2006,** *10* (6), 607-14.
23. Thirunavukarasu, D.; Chen, T.; Liu, Z.; Hongdilokkul, N.; Romesberg, F. E., Selection of 2'-Fluoro-Modified Aptamers with Optimized Properties. *J Am Chem Soc* **2017,** *139* (8), 2892-2895.
24. Davydova, A.; Vorobyeva, M.; Bashmakova, E.; Vorobjev, P.; Krasheninina, O.; Tupikin, A.; Kabilov, M.; Krasitskaya, V.; Frank, L.; Venyaminova, A., Development and characterization of novel 2'-F-RNA aptamers specific to human total and glycated hemoglobins. *Anal Biochem* **2019,** *570*, 43-50.
25. Ruckman, J.; Green, L. S.; Beeson, J.; Waugh, S.; Gillette, W. L.; Henninger, D. D.; Claesson-Welsh, L.; Janjić, N., 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165). Inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *The Journal of biological chemistry* **1998,** *273* (32), 20556-67.
26. Shubham, S.; Hoinka, J.; Banerjee, S.; Swanson, E.; Dillard, J. A.; Lennemann, N. J.; Przytycka, T. M.; Maury, W.; Nilsen-Hamilton, M., A 2'FY-RNA Motif Defines an Aptamer for Ebolavirus Secreted Protein. *Sci Rep* **2018,** *8* (1), 12373.
27. Jackson, L. N.; Chim, N.; Shi, C.; Chaput, J. C., Crystal structures of a natural DNA polymerase that functions as an XNA reverse transcriptase. *Nucleic Acids Res* **2019,** *47* (13), 6973-6983.
28. Ishino, S.; Ishino, Y., DNA polymerases as useful reagents for biotechnology - the history of developmental research in the field. *Frontiers in microbiology* **2014,** *5*, 465.
29. Ma, S.; Saaem, I.; Tian, J., Error correction in gene synthesis technology. *Trends Biotechnol* **2012,** *30* (3), 147-54.
30. Orton, R. J.; Wright, C. F.; Morelli, M. J.; King, D. J.; Paton, D. J.; King, D. P.; Haydon, D. T., Distinguishing low frequency mutations from RT-PCR and sequence errors in viral deep sequencing data. *BMC Genomics* **2015,** *16* (1), 229.
31. Larsen, A. C.; Dunn, M. R.; Hatch, A.; Sau, S. P.; Youngbull, C.; Chaput, J. C., A general strategy for expanding polymerase function by droplet microfluidics. *Nat Commun* **2016,** *7*, 11235.
32. Shao, Q.; Chen, T.; Sheng, K.; Liu, Z.; Zhang, Z.; Romesberg, F. E., Selection of Aptamers with Large Hydrophobic 2'-Substituents. *J Am Chem Soc* **2020,** *142* (5), 2125-2128.
33. Liu, Z.; Chen, T.; Romesberg, F. E., Evolved polymerases facilitate selection of fully 2'-OMe-modified aptamers. *Chem Sci* **2017,** *8* (12), 8179-8182.
34. Cadwell, R. C.; Joyce, G. F., Randomization of genes by PCR mutagenesis. *PCR methods and applications* **1992,** *2* (1), 28-33.

35. Ravikumar, A.; Arzumanyan, G. A.; Obadi, M. K. A.; Javanpour, A. A.; Liu, C. C., Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **2018,** *175* (7), 1946-1957.e13.