JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

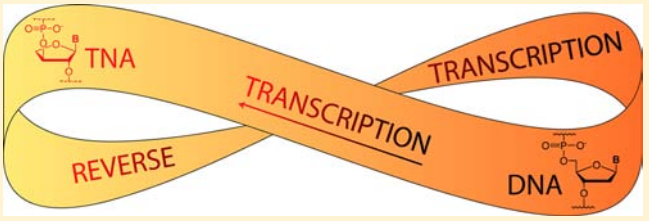
An Efficient and Faithful in Vitro Replication System for Threose Nucleic Acid

Hanyang Yu, Su Zhang, Matthew R. Dunn, and John C. Chaput*

Center for Evolutionary Medicine and Informatics in the Biodesign Institute and Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-5301, United States

Supporting Information

ABSTRACT: The emerging field of synthetic genetics provides an opportunity to explore the structural and functional properties of synthetic genetic polymers by in vitro selection. Limiting this process, however, is the availability of enzymes that allow for the synthesis and propagation of genetic information present in unnatural nucleic acid sequences. Here, we report the development of a transcription and reverse-transcription system that can replicate unnatural genetic polymers composed of threose nucleic acids (TNA). TNA is a potential progenitor of RNA in which the natural ribose sugar found in RNA has been replaced with an unnatural threose sugar. Using commercial polymerases that recognize TNA, we demonstrate that an unbiased three-letter and two different biased four-letter genetic alphabets replicate in vitro with high efficiency and high overall fidelity. We validated the replication system by performing one cycle of transcription, selection, reverse transcription, and amplification on a library of 10^{14} DNA templates and observed ~380-fold enrichment after one round of selection for a biotinylated template. We further show that TNA polymers are stable to enzymes that degrade DNA and RNA. These results provide the methodology needed to evolve biologically stable aptamers and enzymes for exobiology and molecular medicine.



INTRODUCTION

Darwinian evolution allows for the generation of nucleic acid molecules with specific, predefined functions from large pools of random sequences.[1−3] The molecular biology steps needed to isolate these molecules are now well established, and these procedures are routinely used to identify nucleic acid sequences that can fold themselves into shapes that catalyze chemical reactions or bind to specific targets.[4−8] In the case of RNA, for example (Figure 1a), a large population of diverse RNA sequences is constructed by in vitro transcription. The pool of RNA is then assayed en masse for molecules that exhibit a desired phenotype by imposing a selection constraint on the population to remove nonfunctional sequences from the pool. The small number of functional molecules that exhibit the desired phenotype are isolated, reverse-transcribed back into DNA, and amplified using the polymerase chain reaction (PCR) to produce a new population of progeny molecules that has become enriched in a particular trait. Unlike traditional screening methods where molecules are sampled on an individual basis, the process of in vitro selection and amplification makes it possible to search vast regions of sequence space for functional molecules that could be as rare as one in ~10^{15} different sequences.[9,10]

Applying the principles of Darwinian evolution to artificial genetic polymers with unnatural nucleic acid backbones makes it possible to explore the structural and functional properties of xeno-nucleic acids (XNA).[11] This emerging area of science, termed synthetic genetics, is motivated by the desire to create functional synthetic genetic polymers that advance our understanding of life as well as provide molecular tools with practical applications in molecular medicine and synthetic biology.[12,13] Achieving the goal of artificial genetics requires organic chemistry to synthesize unnatural nucleic acid substrates that are not otherwise available and advanced protein engineering methods to create polymerases that can replicate artificial genetic polymers in the laboratory. Because synthetic genetic polymers have the potential to carry out functional roles with greater nuclease stability than natural genetic polymers, XNA replication by engineered polymerases has generated significant interest.[14] The main barrier to the development of effective methods for XNA replication has been the difficulty of identifying polymerases that can copy DNA into XNA and vice versa. Until recently, most examples of polymerase-mediated XNA synthesis resulted in limited XNA incorporation due to the high specificity of natural polymerases.[15] However, this paradigm is now changing as recent advances in protein engineering have produced a new generation of polymerases that can copy genetic information back and forth between DNA and XNA.[16] The practical utility of these enzymes was revealed when an XNA polymer composed of hexose nucleic acid (HNA) was evolved for ligand binding affinity. This achievement demonstrated that

Received: December 7, 2012
Published: February 25, 2013

   dx.doi.org/10.1021/ja3118703 | J. Am. Chem. Soc. 2013, 135, 3583−3591

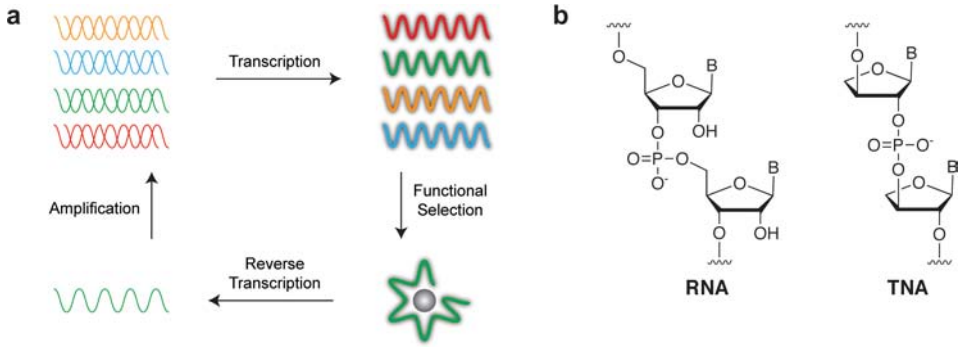


Figure 1. Darwinian evolution of natural and artificial nucleic acid polymers. (a) In vitro evolution of nucleic acid polymers requires polymerases that can transcribe, reverse transcribe, and amplify genetic information in the laboratory. For RNA, each cycle of in vitro selection and amplification involves transcription of a DNA library into RNA, isolation of RNA molecules with a desired function, reverse transcription of functional molecules back into cDNA, and amplification of the resulting cDNA molecules by PCR. Applying the process of Darwinian evolution to XNA, requires engineered polymerases that can copy genetic information back and forth between DNA and XNA. (b) Constitutional structures for the linearized backbones of RNA (left) and α-L-threofuranosyl-(3′→2′) nucleic acid, TNA (right). TNA has a backbone repeat unit that is one atom shorter than the backbone repeat unit found in RNA (and DNA).

heredity and evolution, two important hallmarks of life, are no longer limited to the natural polymers of DNA and RNA.[16]

Despite powerful enzyme engineering strategies,[17] developing efficient replication systems to the growing list of XNA polymers remains a challenging problem in synthetic genetics. The demand for chemical synthesis, coupled with complex evolutionary strategies for polymerase development, suggests that considerable effort would be required before enzymes are available that can routinely copy diverse sets of XNA polymers. Recognizing this problem, we have sought to identify conditions that would allow commercial polymerases to transcribe and reverse transcribe XNA polymers with high efficiency and fidelity. We have focused our efforts on α-L-threofuranosyl nucleic acid (TNA), an artificial XNA polymer in which the natural five-carbon ribose sugar found in RNA has been replaced with an unnatural four-carbon threose sugar (Figure 1b).[18] This substitution leads to a nucleic acid polymer with a backbone repeat unit that is one-atom shorter than the backbone repeat unit found in natural DNA and RNA. TNA has generated considerable interest as a possible RNA progenitor due to the chemical simplicity of threose relative to ribose and the ability for TNA to cross-pair opposite complementary strands of RNA.[19−22] This latter feature, which provides a plausible mechanism for the passage of genetic information between successive genetic systems, is likely due to the helical geometry of the TNA duplex, which is structurally similar to natural A-form DNA and RNA.[23]

The discovery of TNA as a possible RNA progenitor inspired several laboratories to begin developing the methodology needed to explore the functional properties of TNA by in vitro selection. Much of the early work in this area focused on identifying natural or engineered polymerases that could recognize TNA either in the template or as a nucleoside triphosphate. From these studies, we identified several DNA polymerases that could synthesize short sequences of DNA on a TNA template and other polymerases that could copy limited stretches of TNA on a DNA template.[24,25] Herdewijn reported similar findings for the transcription of tTTP on a DNA template using thermophilic polymerases.[26] While these results suggested that TNA is not easily recognized by natural enzymes, subsequent screening did lead to the discovery of Therminator DNA polymerase, an engineered archeal family B DNA polymerase, that functions as an efficient DNA-dependent TNA polymerase.[27,28] Ichida and Szostak developed a DNA display strategy to generate functional TNA molecules by in vitro selection.[29] This method establishes a genotype—phenotype link by extending a library of self-priming DNA templates with TNA, which allows each TNA sequence to become physically connected to its own DNA message. Using DNA display, we evolved a TNA aptamer with high affinity and high specificity to human thrombin.[30] This demonstration, along with the evolution of an HNA aptamer,[16] shows that at least two different XNA polymers can fold into tertiary structures that can perform sophisticated types of chemical functions.

In the current article we report a complete replication system for TNA that allows for the storage and propagation of synthetic genetic information using commercial enzymes. This approach was developed to expand the range of evolutionary strategies that could be used to evolve TNA aptamers and catalysts in the laboratory. Several key developments were made to facilitate the in vitro replication of TNA polymers with high efficiency and fidelity. First, adenine threofuranosyl 3′-triphosphate was used in place of diaminopurine to establish a set of TNA substrates that contain only natural nucleobases. Second, putative reverse transcriptases were rigorously tested and optimized to identify the conditions needed to efficiently copy long TNA templates into DNA. Third, fidelity tests were performed on several template designs to define the sequence constraints of in vitro TNA replication. Fourth, a mock in vitro selection was performed to demonstrate the feasibility of our replication system to support the in vitro selection of functional TNA molecules. Last, the nuclease stability of TNA was evaluated by measuring the sensitivity of TNA to natural enzymes that degrade DNA and RNA. Taken together, these results advance the field of synthetic genetics by providing the methodology needed to evolve biologically stable TNA aptamers and enzymes for exobiology and molecular medicine.

RESULTS

In Vitro Transcription of TNA. We began by chemically synthesizing each of the α-L-threofuranosyl nucleoside triphosphates (tNTPs) required for our study.[31,32] This included TNA triphosphates with all four natural bases (tTTP, tATP, tCTP, and tGTP) as well as the diaminopurine analogue (tDTP) of adenine threofuranosyl 3′-triphosphate

dx.doi.org/10.1021/ja3118703 | J. Am. Chem. Soc. 2013, 135, 3583−3591

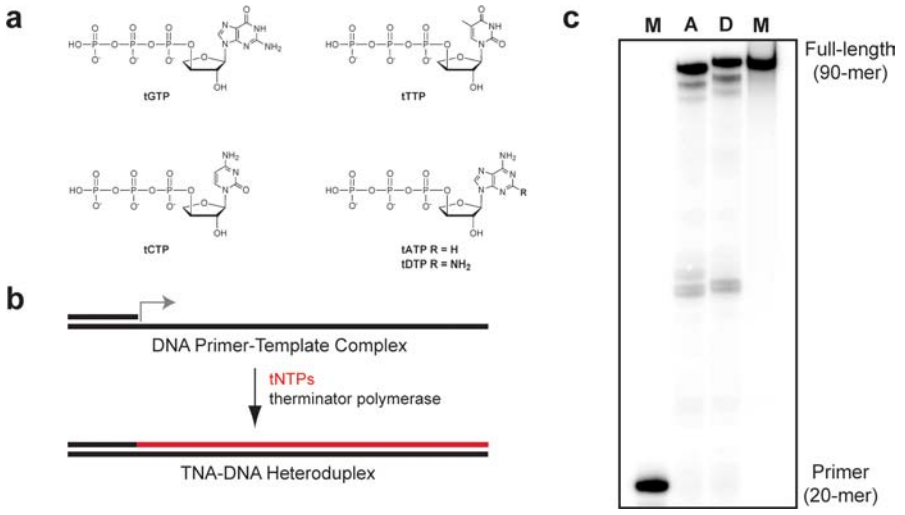


Figure 2. Enzyme-mediated transcription of a DNA library into a library of TNA sequences. (a) Chemical structures of TNA triphosphates (tNTPs). Diaminopurine (tDTP) is an analogue of adenine (tATP) that forms three hydrogen bonds with thymine. (b) Schematic representation of DNA primer extension reaction used to synthesize long TNA strands. The DNA primer–template complex is given in black, while the extended TNA product is shown in red. (c) Therminator-mediated TNA transcription reactions performed for 1 h at 55 °C and analyzed by denaturing polyacrylamide gel electrophoresis. A refers to primer extension reactions performed using tATP, D refers to primer extension reactions that contain tDTP in place of tATP, and M refers to the marker.

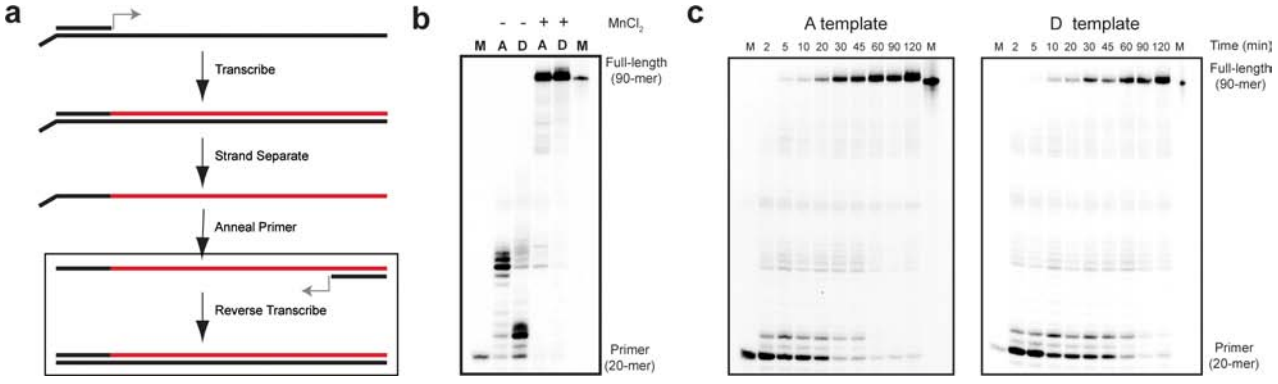


Figure 3. Enzyme-mediated reverse transcription of individual TNA molecules into DNA. (a) Schematic representation of TNA synthesis and reverse transcription. Boxed region highlights the reverse transcription step analyzed by gel electrophoresis. The TNA region of each strand is shown in red, while the cDNA region is shown in black. (b) SuperScript II-mediated TNA reverse transcription performed for 24 h at 42 °C and analyzed by denaturing polyacrylamide gel electrophoresis. Mn^{2+} is required to convert TNA into full-length cDNA. (c) Time course analysis of cDNA synthesis on TNA templates. The A and D templates refer to TNA templates containing either adenosine or diaminopurine in the TNA strand, and M refers to the DNA primer and full-length product markers.

(Figure 2a). Previous studies have established that the diaminopurine modification strongly enhances the thermodynamic stability of TNA/TNA, TNA/RNA, and TNA/DNA duplexes (for example, $\Delta\Delta G$ = 4.7 kcal/mol, tD_{12}/tT_{12} versus tA_{12}/tT_{12}).[33] This modification also accelerates the rate of nonenzymatic template-directed ligation of TNA ligands and improves the efficiency of polymerase-mediated extension of tTTP residues on a DNA template.[25,33] While our earlier work focused exclusively on the use of tDTP as substrate for TNA synthesis,[25,28,30] we have become concerned that the diaminopurine analogue might complicate the analysis of future TNA aptamers and enzymes. One could imagine that the presence of an additional proton-donor group on the adenine base would make secondary structure prediction more difficult due to the enhanced potential for alternative non-Watson–Crick base pairing modes.[34] A further concern is that structural differences between TNA and natural DNA and RNA are no longer limited to the sugar–phosphate backbone, which could obfuscate future comparisons made with previously evolved aptamers and enzymes.

To address these concerns, we examined the efficiency of tATP as a substrate for Therminator DNA polymerase. As illustrated in Figure 2b, a synthetic DNA primer was annealed to a synthetic DNA library that contained a random region of 50-nts flanked on either side with a 20-nt primer-binding site. Therminator DNA polymerase was challenged to extend the DNA primer with up to 70 sequential TNA residues to produce a library of TNA molecules containing either adenine or diaminopurine nucleotides in the product strands. Primer-extension assays were performed by incubating the polymerase for 1 h at 55 °C in reaction buffer supplemented with 1.25 mM $MnCl_2$. We have previously shown that manganese ions dramatically enhance the efficiency of TNA synthesis.[24] Analysis of the extension products by denaturing polyacrylamide gel electrophoresis reveals that tATP and tDTP are equally efficient substrates for Therminator DNA polymerase.

dx.doi.org/10.1021/ja3118703 | J. Am. Chem. Soc. 2013, 135, 3583–3591

In both cases, the DNA primer was completely extended with TNA residues to make the desired full-length product (Figure 2c). While we have constructed TNA libraries with diaminopurine residues, this was the first demonstration where a TNA library was prepared using all four natural nucleobases. Since no difference in the amount of full-length product was observed between the two sets of in vitro transcription reactions, we concluded that tATP is an efficient substrate for Therminator DNA polymerase in the enzyme-mediated polymerization of TNA.

In Vitro Reverse Transcription of TNA into DNA. The in vitro selection of XNA molecules in the laboratory requires enzymes that can transcribe and reverse transcribe XNA polymers with high efficiency and fidelity. In a recent new advance, Pinheiro et al. used a compartmentalized self-tagging strategy to evolve several polymerases with XNA activity.[16] One of these enzymes, RT521, was created from TgoT, a variant of the replicative polymerase from *Thermococcus gorgonarius*, for the ability to reverse transcribe HNA back into DNA. In addition to HNA reverse transcriptase activity, RT521 was also found to reverse transcribe other XNA polymers with varying degrees of efficiency. This included arabinonucleic acids, 2′-fluoro-arabinonucleic acids and TNA.[35] The observation that RT521 could reverse transcribe portions of a TNA template into DNA led us to consider this enzyme as a possible polymerase for the replication TNA polymers in vitro.

To examine the activity of RT521 as a TNA-dependent DNA polymerase, we performed a polymerase activity assay to access the ability for RT521 to reverse transcribe long TNA templates into DNA. Because it is not possible to generate long TNA polymers by solid-phase synthesis, we transcribed a DNA template into TNA using Therminator DNA polymerase (Figure 3a). The resulting TNA polymer was purified by denaturing polyacrylamide gel electrophoresis and used as a template for reverse transcription. A second DNA primer was then annealed to the 2′-end of the TNA strand, and reverse transcription was attempted by incubating the primer−template complex with RT521 for 24 h at 65 °C. Although some variation was observed among the different TNA templates, the best primer-extension reaction produced full-length products that were barely detectable by polyacrylamide gel electrophoresis (Figure S1).

In an attempt to improve the efficiency of TNA-dependent DNA polymerization by RT521, we explored a variety of conditions that have proven helpful in the past. To our surprise, varying the reaction time, salt conditions, and enzyme concentration all proved ineffective. Even the addition of manganese ions, which is known to relax the specificity of many DNA polymerases,[36] inhibited the reaction. The presence of diaminopurine residues in the TNA template also failed to improve the yield of full-length product. The limited DNA synthesis observed in these reactions may reflect an unknown sequence specificity of the enzyme. Alternatively, it is also possible that the sample of RT521 used in our study was less active than the sample used in the original study by Pinheiro et al. However, close examination of the previous reverse transcription reaction reveals a substantial amount of truncated product, suggesting that RT521 may require further optimization before it can function as an efficient TNA-dependent DNA polymerase.[16]

Recognizing the limitations of RT521, we pursued other enzymes as possible candidates for a TNA reverse transcriptase. In this regard, we have previously screened a wide range of natural and mutant DNA and RNA polymerases for the ability to copy a short chimeric DNA−TNA template containing nine TNA residues in the template region.[24] This study identified the reverse transcriptases MMLV and SuperScript II (SSII) as efficient TNA-dependent DNA polymerases that could copy a short TNA template into DNA with ∼30% full-length product conversion observed after an incubation of 1 h at 42 °C. To determine whether these enzymes could be made to function on longer TNA templates, we explored a range of conditions that would allow the enzymes to copy a 90-nt TNA template back into DNA. Since it was possible that diaminopurine would enhance the efficiency of reverse transcription, we performed the polymerase activity assay on in vitro transcribed TNA containing either adenine or diaminopurine nucleotides in the template strand. Preliminary studies indicated that SSII functioned with greater efficiency and reproducibility than MMLV. Subsequent optimization of this reaction led us to discover conditions that enabled SSII to reverse transcribe the entire TNA template into DNA (Figure 3b). Optimal extension was observed using new enzyme and a reaction buffer that contained a freshly prepared solution of 1.5 mM $MnCl_2$. Under these conditions, the adenine- and diaminopurine-containing TNA templates are efficiently reverse transcribed back into DNA. In the absence of $MnCl_2$, the reaction is significantly impeded with SSII terminating reverse transcription early into the primer extension process.

To assess the efficiency of SSII-mediated reverse transcription, we performed a time course analysis to compare the rate of product formation as a function of template composition. Analysis of product formation over time revealed that reverse transcription of the adenine-containing template is complete in 1 h, while the diaminopurine-containing template requires nearly 2 h to copy the TNA template into DNA (Figure 3c). The higher efficiency of the adenine-containing template further supports the use of tATP as a substrate for TNA synthesis. Taken together, the transcription and reverse transcription results demonstrate that commercial enzymes can be made to replicate TNA polymers with high efficiency, which is remarkable considering the significant structural differences between the threofuranosyl and (deoxy)ribofuranosyl backbones of TNA and DNA (or RNA), respectively.

Fidelity of TNA Replication. We measured the fidelity of TNA replication by sequencing the cDNA product of the reverse transcription reaction after amplification by PCR. This fidelity assay measures the aggregate fidelity of a complete replication cycle (DNA → TNA → DNA), which is operationally different than the more restricted view of fidelity as the accuracy of a single-nucleotide incorporation event. The fidelity determined by this assay is the actual accuracy with which full-length TNA is synthesized and reverse transcribed and therefore reflects the combined effects of nucleotide misincorporation, insertions and deletions (indel) and any mutations that occur during PCR amplification and cloning. Several controls were implemented to ensure that the sequencing results represented the true fidelity of TNA replication (Figure S2). First, to eliminate any possibility of contamination by the starting DNA template, the DNA primer−template complex used for TNA transcription was partially unpaired and contained additional nucleotides in the primer strand to facilitate separation of the TNA product by denaturing polyacrylamide gel electrophoresis. Second, all PCR amplification steps were performed using a negative control that contained the purified TNA template prior to reverse

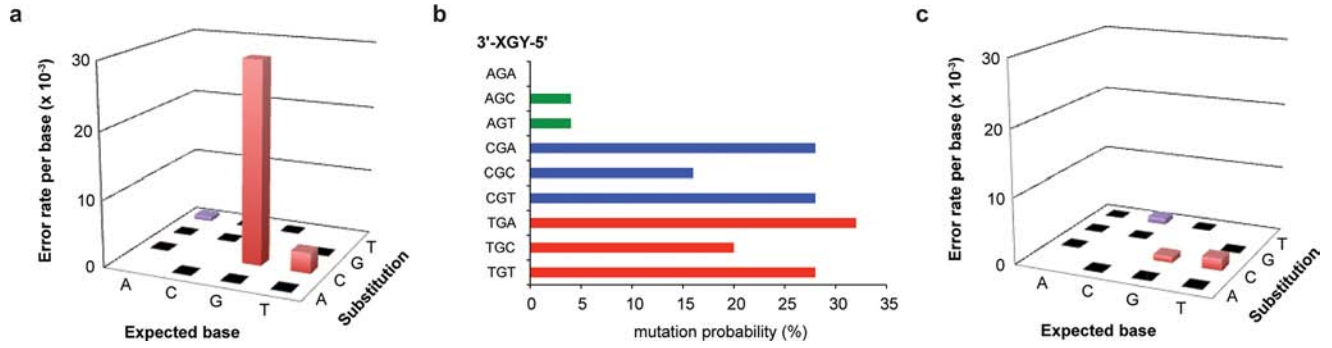


Figure 4. Fidelity of TNA replication using a four-letter genetic alphabet. (a) The mutation profile of TNA replication indicates a high frequency of G → C substitutions during TNA transcription. (b) Analysis of the local sequence context upstream and downstream of the misincorporation site indicates a sequence-specific context that favors mutagenesis when G residues in the DNA template are preceded by pyrimidines (C or T). (c) Substituting tCTP for dCTP suppresses dG:tG mispairing and reduces the error rate from 36×10^{-3} to 3.5×10^{-3}.

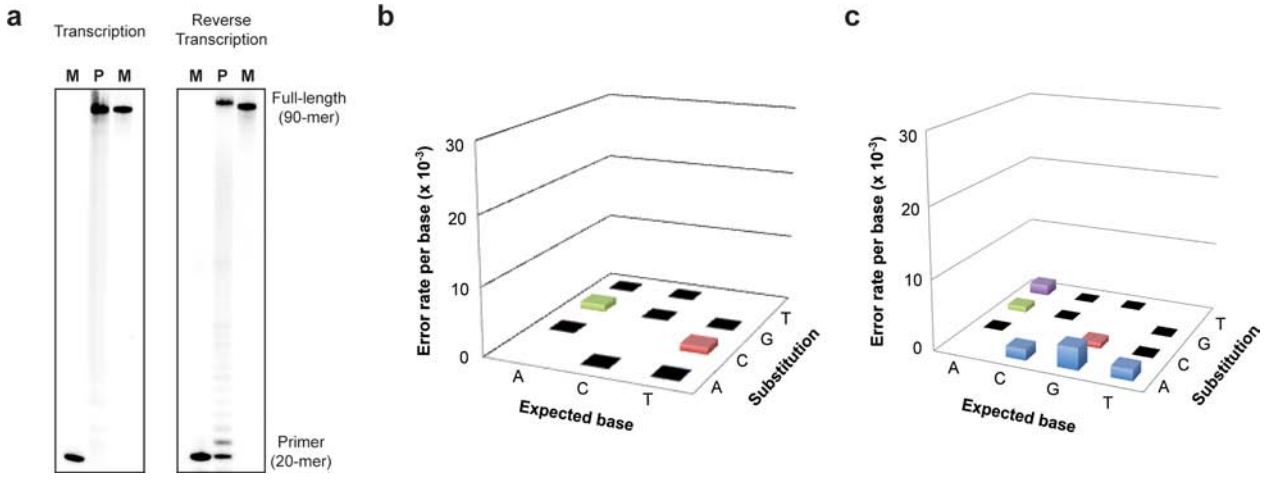


Figure 5. An efficient and faithful replication system for TNA. (a) Replication of a three letter TNA library. A DNA library composed of three nucleotides (A, C, and T) transcribes into TNA (left panel) and reverse transcribes back into DNA (right panel) with high primer-extension efficiency. (b) Mutation profile demonstrates that ACT sequences replicate with an error rate of 3.8×10^{-3} (99.6% fidelity). (c) Replication of a biased four letter TNA library. A DNA library composed of ACT-AG replicates with an error rate of 10.0×10^{-3} (99.0% fidelity).

transcription. In no cases did we observe a DNA band in this lane, demonstrating that the purification step effectively separated the TNA transcript from the DNA template (Figure S3). Third, to unambiguously demonstrate that each DNA sequence derived from a complete cycle of TNA replication, the DNA primer used for TNA transcription was engineered to contain a single-nucleotide mismatch that resulted an A → T transversion in the sequenced product. These controls allowed us to determine the actual fidelity of TNA replication with confidence.

We began by measuring the fidelity of TNA replication for the adenine-containing template used in the reverse transcription assay with SSII. This template, referred to as 4NT.3G, derives from a single sequence that was present in the L3 library.[30] The L3 library was designed to overcome the problem of polymerase stalling at G-repeats by reducing the occurrence of G residues in the template to 50% the occurrence of A, C, and T. Our earlier work on TNA transcription established the L3 library as an efficient design strategy for generating pools of full-length TNA molecules.[30] While TNA replication on 4NT.3G resulted in an overall fidelity that was comparable with other XNA replication systems (96.4%),[16] detailed analysis of the mutation profile indicated that G → C transversions account for 90% of the genetic changes (Table S3 and Figure 4a).

Since iterative replication cycles of the L3 library would eventually bias TNA replication toward a population of DNA sequences that were overly enriched in cytidine residues, we decided to ascertain the propensity for mutagenesis by examining the role of nearest-neighbor effects in the DNA template. We designed a synthetic DNA template (4NT.9G) containing all of the possible combinations of A, C, and T nucleotides on the 3′ and 5′ side of a central G residue. We avoided the triplets NGG, GGN, and GGG due to their ability to terminate primer extension (for example, see Figure S4). We found that the frequency of a G → C transversion is ∼25% when a pyrimidine (C or T) precedes G in the template, but only ∼3% when G is preceded by A (Figure 4b). No correlation was observed between the identity of the 5′ nucleotide residue and the frequency of transversion, suggesting that mutagenesis occurs during the transcription step of TNA replication. We tested this hypothesis by repeating the triplet fidelity study using a nucleotide mixture in which the tCTP substrate was replaced with dCTP. Under these conditions, mutagenesis is suppressed, and the overall fidelity of TNA replication increases to 99.6% (Table S3 and Figure 4c).

dx.doi.org/10.1021/ja3118703 | J. Am. Chem. Soc. 2013, 135, 3583−3591

While the precise molecular details of the G → C transversion remain unknown, our results suggest that base stacking plays an important role in the misincorporation of tGTP opposite deoxyG in the template. This prediction is supported by the fact that the frequency of dG:tG mispairing increases 10-fold when G-nucleotides in the template are preceded by pyrimidine residues, indicating that purine residues (A or G) on the growing TNA strand stabilize the incoming tGTP substrate via base stacking interactions. However, this observation is also consistent with the strong slide between adjacent base pairs in the crystal and NMR structures of TNA-containing DNA and all-TNA duplexes that result in significant cross-strand stacking (i.e., template dG and incoming tGTP).[21−23] To better understand the problem of dG:tG mispairing, we measured the fidelity of TNA replication using different combinations of template and substrate (Figure S5). Biasing the nucleotide mixture with lower amounts of tGTP and higher amounts of tCTP increased the fidelity to 97.6% and reduced the problem of G → C transversions. Substituting tGTP for dGTP and assaying a template devoid of C residues produced similar results with 97.5% and 98.2% fidelity, respectively. The mutational profiles obtained under these conditions provide evidence that dG:tG mispairing can be overcome by engineering DNA templates to avoid the problem of nucleotide misincorporation.

In an effort to further improve the fidelity of TNA replication, we examined the mutational profile of two different types of DNA templates that were designed for high fidelity replication. The first template, 3NT.ATC, contained a central region of 50-nts that was composed of a random distribution of A, T, and C residues that were flanked by two 20-nt fixed-sequence primer-binding sites. This sequence derived from library L2, which we used previously to evolve a TNA aptamer to human thrombin.[30] We found that the L2 library transcribes and reverse transcribes with very high efficiency as judged by the amount of starting primer that is extended to full-length TNA product and the absence of any significant truncated products (Figure 5a). Consistent with the efficient replication of the L2 library, the template 3NT.ATC exhibits an overall fidelity of replication of 99.6% (Figure 5b), which is similar to the fidelity of in vitro RNA replication. Similar results (99.0% fidelity) were obtained with a four-nucleotide sequence, 4NT.9GA, which is identical to the DNA template 4NT.9G, except that each of the nine G residues in the template was preceded by an adenine nucleotide to minimize dG:tG mispairing in the enzyme active site (Figure 5c). These results demonstrate that commercial enzymes are capable of replicating TNA with high efficiency and fidelity, both of which are essential for future in vitro selection experiments.

Selection of a Biotinylated TNA Strand From a TNA Library. To test the ability of our replication system to support in vitro selection, we performed in vitro transcription, selection, and reverse transcription on a model DNA library. We synthesized a biotinylated DNA template, 4NT.3G, containing a HinfI restriction site in the coding region and primer binding sites that are identical in sequence to the L2 library. The DNA template was combined with 10^{14} L2 library members at ratios of 1:10, 1:100, and 1:1000 (DNA template to L2 library) to create three doped-sequence libraries. The three DNA pools were transcribed into TNA, and the biotinylated TNA−DNA heteroduplexes were selected by incubating the mixtures on streptavidin-coated beads. The beads were thoroughly washed to remove the nonbiotinylated L2 members, and TNA strands that remained on the beads were recovered by denaturing the DNA−TNA heteroduplex with NaOH. The solution was neutralized, reverse-transcribed back into DNA, and amplified by PCR.

To determine the enrichment from this one-round of mock selection, the PCR amplified DNA was digested with HinfI, and the ratio of digested DNA from each of the three pools was measured after agrose gel electrophoresis. Negative (L2 library) and positive (4NT.3G only) controls were used to ensure that HinfI cuts the 4NT.3G biotyinlated template with high efficiency and specificity (Figure 6). As expected, the L2

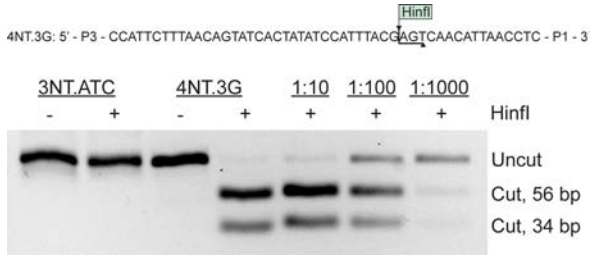


Figure 6. Enrichment of biotinylated TNA from a TNA library. 10^{14} molecules of the L2 library were doped with 1/10th, 1/100th, and 1/1000th molar ratio of biotinylated 4NT.3G DNA template containing a HinfI restriction site. The resulting pools were subjected to one round of in vitro transcription, selection for binding to immobilized streptavidin, reverse transcription, and PCR amplification. HinfI digestion of amplified molecules reveals ∼380-fold enrichment of the biotinylated molecules from the pool.

library, which contains only A, T, and C in the random region remained undigested after incubation with HinfI, while the 4NT.3G control template digested quantitatively. Comparison of the relative intensity of cut versus uncut DNA in the doped pools revealed enrichment of ∼380-fold for the 4NT.3G sequence. This result is consistent with the level of enrichment previously observed for other in vitro selections of biotinylated templates.[37,38]

Nuclease Stability. A major goal of synthetic genetics is to create nuclease resistant aptamers and enzymes that function in complex biological environments. To evaluate the nuclease stability of TNA, we synthesized a synthetic TNA 16-mer having the sequence 3′-AAAATTTATTTATTAA-2′ by solid-phase phosphoramidite chemistry. The TNA oligonucleotide was tested for nuclease stability against the enzymes RQ1 DNase and RNase A, which degrade DNA and RNA, respectively. In both cases, 1 nmol of the TNA sample was incubated at 37 °C in a reaction buffer of 40 mM Tris-HCl, 10 mM $MgSO_4$, and 1 mM $CaCl_2$ (pH 8.0) for the DNase digestion and a reaction buffer of 50 mM NaOAc (pH 5.0) for the RNase digestion. The samples were removed at specified time points, quenched with urea, and analyzed by denaturing polyacrylamide gel electrophoresis. As a control, synthetic DNA and RNA strands with the same sequence were incubated with their respective nuclease and analyzed under time frames that coincided with their degradation. As expected, the DNA sample is rapidly degraded in the presence of RQ1 DNase and exhibited a half-life of ∼30 min (Figure 7a). The case was even more extreme for the RNA sample, which degraded in a matter of seconds and exhibited a half-life of <10 s (Figure 7b). In contrast to the natural DNA and RNA samples, the TNA sample remained undigested even after 72 h in the presence of pure nuclease (Figure 7a,b). This result demonstrates that

dx.doi.org/10.1021/ja3118703 | J. Am. Chem. Soc. 2013, 135, 3583−3591

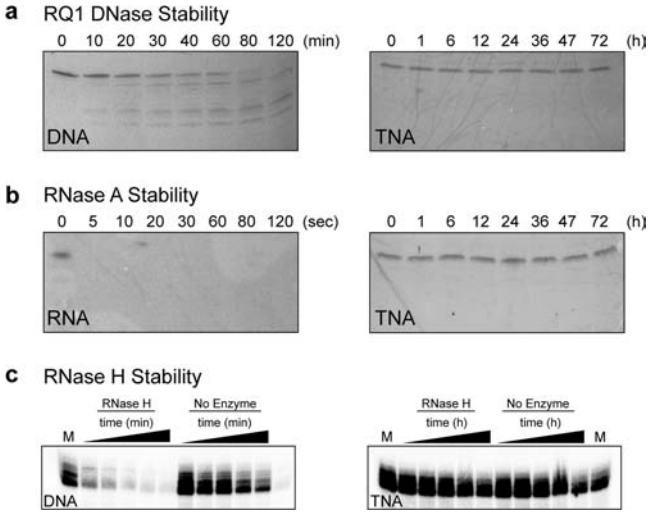


Figure 7. TNA sensitivity to nuclease degradation. Nuclease stability of synthetic DNA, RNA, and TNA oligonucleotides were monitored over time by denaturing polyacrylamide gel electrophoresis. (a) In the presence of RQ1 DNase, DNA exhibits a half-life of ~30 min, while TNA remains undigested after 72 h. (b) In the presence of RNase A, RNA is digested in less than 5 s, while TNA remains intact after 72 h. (c) RNase H digestion using DNA and TNA probes that are complementary to an RNA target indicates that TNA is not a substrate for RNase H. The reaction time for DNA and TNA was 30 min and 16.5 h, respectively.

enzymes that degrade DNA and RNA do not easily recognize the threofuranosyl backbone of TNA.

Antisense oligonucleotides are widely used to alter intracellular gene expression patterns by activating RNase H activity.[39] RNase H is an endoribonuclease that specifically hydrolyzes the phosphodiester bonds of RNA in DNA–RNA duplexes to produce 3′ hydroxyl and 5′ monophosphate products. Given the importance of alternative nucleic acid structures as antisense therapeutics,[40] we felt that it would be interesting to examine the recognition properties of TNA–RNA hybrids by RNase H. We hybridized a 16-mer TNA oligonucleotide to the target site of a 70-mer synthetic RNA strand produced by in vitro transcription. To establish a positive control for RNase H activity, the analogous 16-mer DNA probe was hybridized to the RNA target. The DNA and TNA samples were incubated at 37 °C in the presence and absence of the enzyme in buffer containing 10 mM Tris-HCl, 25 mM KCl, 1 mM NaCl, and 0.5 mM $MgCl_2$ (pH 7.5). Samples were removed at specified time points, quenched with urea, and analyzed by denaturing polyacrylamide gel electrophoresis. As expected, the DNA–RNA hybrid is rapidly degraded (half-life <1 min) in the presence of RNase H, while the TNA–RNA hybrid remained intact even after an incubation of 16.5 h indicating that TNA does not promote RNase H activity in vitro (Figure 7c).

■ DISCUSSION

The ability to replicate TNA using DNA as an intermediate provides a convenient format for assessing the fidelity of TNA replication. We used this assay to define the parameters of TNA replication by examining the fidelity of different DNA library designs. Results from these assays indicate that DNA libraries composed of a three-letter genetic alphabet of A, C, and T or a four-letter genetic alphabet of A, C, T, and G′, where all G residues are preceded by A, can be made to replicate with high efficiency and high overall fidelity (>99%). Transcription of DNA libraries with these nucleotide compositions leads to the synthesis of TNA molecules that either lack C or contain C at reduced frequency. However, this difference may not be a significant concern for studies that aim to examine the origin and evolution of early genetic polymers, as cytidine may have been less prevalent on the early Earth due to its tendency to undergo spontaneous deamination to uridine.[41] Furthermore, previous in vitro selections experiments have shown that functional cytidine-free molecules can be obtained by in vitro evolution, demonstrating that three-nucleotide libraries do not prevent genetic polymers with reduced nucleobase diversity from folding into shapes, and can elicit a specific chemical function.[42]

Although not explicitly examined in the current work, another potential library is one that derives from DNA sequences consisting of tandem repeats of 3′-ANH-5′, where N is any natural DNA nucleotide and H is an unbiased random mixture of A, C, and T. DNA libraries with this design would code for TNA molecules with greater sequence diversity than the L2 library, as they would contain all four TNA nucleotides. By ensuring that adenosine residues precede each G nucleotide in the template, fidelity of the incoming tCTP is maintained during TNA synthesis. Presumably, this library design would outperform the H-only design or an H-library with where all G residues occur at fixed positions.

A third library that we developed involves the synthesis of mixed-backbone chimeric DNA–TNA polymers that contain deoxyC residues interspersed among an otherwise intact TNA backbone. Although less interesting as a primordial genetic polymer, this system could find widespread use in molecular medicine where the goal is to evolve nuclease resistant molecules for diagnostic and therapeutic applications. The advantage of the chimeric library strategy is that the TNA backbone provides the molecular functionality needed to achieve nuclease resistance, while the tCTP to dCTP substitution allows replication to occur on unconstrained pools of purely random nucleic acid sequences.

Our study, in conjunction with seminal work on XNA replication by Pinheiro et al.,[16] provides the molecular tools needed to evolve functional XNA molecules in the laboratory. While these advances mark the emergence of synthetic genetics,[12,13] many challenges still face those interested in exploring the structural and functional properties of unnatural genetic polymers. Perhaps the most significant problem of all is the limited availability of substrates and enzymes that can be used to evolve XNA molecules in the laboratory. The current study addresses this problem in part by developing the conditions needed to replicate synthetic TNA polymers using commercial enzymes. Unfortunately, TNA substrates, like most XNA substrates, are only accessible by chemical synthesis, which requires extensive knowledge of organic chemistry to construct and purify nucleoside triphosphates. Given the tremendous potential for synthetic genetics to impact the fields of exobiology, synthetic biology, and medicine, it seems likely that XNA reagents will become more available with time. A second challenge is to improve existing XNA replication systems so that they function with higher efficiency and fidelity. Our analysis of TNA replication, for example, demonstrates that certain XNA systems are subject to replication biases that may require further optimization of the reaction conditions or possibly even refinement of the polymerases themselves. By

dx.doi.org/10.1021/ja3118703 | J. Am. Chem. Soc. 2013, 135, 3583–3591

fine-tuning each of the different XNA replication systems, it should be possible to create robust tools that help push the field of synthetic genetics into mainstream molecular biology.

In summary, we have established the methodology needed to replicate TNA polymers by an enzyme-mediated process that involves the reverse transcription of TNA into DNA, amplification of the DNA by PCR, and forward transcription back into TNA. We demonstrate that a complete replication cycle occurs with high efficiency and high overall fidelity and yields TNA molecules that are resistant to nuclease degradation. By inserting appropriate selection constraints into the replication cycle, it should be possible to evolve biologically stable aptamers and 'threozymes' for exobiology and molecular medicine.

METHODS

TNA Synthesis by Primer Extension on a DNA Template. The DNA primer P1 was 5′-end labeled by incubation in the presence of [γ-^{32}P] ATP with T4 polynucleotide kinase for 1 h at 37 °C. The ^{32}P-labeled primer was annealed to the DNA template (Table S1) in 1× ThermoPol buffer [20 mM Tris-HCl, 10 mM $(NH_4)_2SO_4$, 10 mM KCl, 2 mM $MgSO_4$, 0.1% Triton X-100, pH 8.8 at 25 °C] by heating at 95 °C for 5 min and cooling on ice. Primer extension reactions were performed in 10 μL volumes containing 100 μM tNTPs (or a combination of defined tNTP and dNTP mixtures), 500 nM primer–template complex, 1 mM DTT, 100 μg/mL BSA, 1.25 mM $MnCl_2$, and 0.1 U/μL Therminator DNA polymerase. Reactions were initiated by adding the tNTP substrates to a solution containing all other reagents and heating the mixture for 1 h at 55 °C. Primer extension products were analyzed by 20% denaturing polyacrylamide gel electrophoresis, imaged with a phosphorimager, and quantified using ImageQuant software (GE Healthcare Biosciences, Pittsburgh, PA).

Generation of TNA Template for Reverse Transcription. TNA synthesis reactions were performed as described above using unlabeled DNA primer P1 in a 400 μL reaction. After incubation for 1 h at 55 °C, the TNA product was separated from the DNA template by 10% denaturing polyacrylamide gel electrophoresis and stained with SYBR Gold. The band corresponding to the TNA product was excised, and the gel slices were electroeluted for 2 h at 200 V. The final solution was ethanol precipitated and quantified by UV absorbance.

TNA Reverse Transcription by Primer Extension Assay. The ^{32}P-labeled DNA primer P3 was annealed to the TNA template in 1× first strand buffer [50 mM Tris-HCl, 75 mM KCl, 3 mM $MgCl_2$ (pH 8.3 at 25 °C)] by heating at 95 °C for 5 min and cooling on ice. Primer extension reactions contained 500 μM dNTPs, 100 nM primer–template complex, 10 mM DTT, 3 mM $MgCl_2$, 1.5 mM $MnCl_2$, and 10 U/μL SuperScript II reverse transcriptase. Reactions were initiated by adding the enzyme to a solution containing all other reagents and heating the reaction mixture for 1 h at 42 °C. Primer extension products were analyzed by 20% denaturing polyacrylamide gel electrophoresis, imaged with a phosphorimager and quantified using ImageQuant software (GE Healthcare Biosciences, Pittsburgh, PA).

Fidelity Assay. DNA sequencing was used to measure the fidelity for the overall process of TNA replication and cloning. DNA templates of a defined sequence were transcribed into TNA as described above using primer P2. Primer P2 has an internal reference nucleotide that is designed to unambiguously distinguish cDNA obtained from TNA replication from the starting DNA template. The DNA–TNA heteropolymer was purified by denaturing polyacrylamide gel electrophoresis and reverse transcribed back into DNA. The resulting cDNA strand was amplified by PCR using primers that matched the outside region of P2 (i.e., P3 and P4). AccuPrime Taq high fidelity DNA polymerase was used to minimize possible mutations caused by PCR. Additionally, separate PCR reactions were performed on purified TNA templates to confirm that the PCR product was amplified from cDNA generated in TNA reverse transcription. PCR products were cloned into pJET1.2 vector, transformed into *E. coli* XL1-Blue competent cells, and grown to log phase, and the vector was isolated using PureYield Plasmid Miniprep System (Promega, Madison, WI). Isolated vectors were sequenced at the ASU DNA Sequencing Facility.

Streptavidin Binding Selection. DNA library L2, which codes for unbiased mixtures of A, T, and G TNA residues, was mixed with a biotinylated 4NT.3G template in a 10:1, 100:1 or 1000:1 ratio favoring the library. The doped pools were transcribed into TNA by primer extension. The mixture was then incubated for 15 min at 24 °C with streptavidin agarose resin pre-equilibrated in binding buffer [10 mM Tris, 50 mM NaCl, 1 mM EDTA, pH 7.4]. Nonbinders were removed with 10 column volumes of binding buffer, and bound TNA molecules were eluted by incubating with 20 mM NaOH for 15 min at 37 °C. Four separate elutions were collected. Recovered TNA strands were ethanol precipitated, reverse transcribed, and amplified by PCR. PCR amplified DNA was digested with HinfI (5 units) for 1 h at 37 °C, and the digestion products were analyzed by 4% agarose gel electrophoresis.

Nuclease Stability Assay. DNA, RNA, and TNA oligonucleotide substrates (1 nmol) were incubated for up to 72 h at 37 °C in presence of RQ1 DNase or RNase A using the manufacture's recommended conditions. The DNase reaction contained 1× RQ1 DNase reaction buffer [40 mM Tris-HCl, 10 mM $MgSO_4$, 1 mM $CaCl_2$, pH 8.0] and 0.2 U/μL of RQ1 RNase-free DNase in reaction volume of 10 μL. The RNase reaction contained 50 mM NaOAc (pH 5.0) and 0.24 μg/μL RNase A in a reaction volume of 10 μL. Time course reactions were performed by initiating multiple reactions in parallel, removing individual tubes at defined time points, quenching the reaction by the addition of 7 M urea and 20 mM EDTA, and storing the quenched reactions at −20 °C until the time course was complete. Time-dependent oligonucleotide stability against DNase or RNase was analyzed by 20% denaturing polyacrylamide gel electrophoresis and visualized by UV shadowing.

RNase H Assay. RNA template T1 was synthesized by in vitro transcription using T7 RNA polymerase. After purification by denaturing PAGE, the RNA transcript was dephosphorylated using calf intestinal alkaline phosphatase and then 5′-end labeled by incubation in the presence of [γ-^{32}P] ATP with T4 polynucleotide kinase. ^{32}P-labeled RNA template T1 (25 pmol) was incubated with a complementary DNA oligonucleotide probe S2 or TNA oligonucleotide probe S3 (50 pmol) for 15 min at 37 °C. Each reaction contained 44 μL of reaction buffer [10 mM Tris-HCl, 25 mM KCl, 1 mM NaCl, and 0.5 mM $MgCl_2$, pH 7.5] and 6 μL RNase H (5 U/μL). Control tubes received buffer in place of enzyme. Aliquots were removed at the indicated time points, quenched by the addition of 7 M urea and 20 mM EDTA, and analyzed by 20% denaturing polyacrylamide gel electrophoresis.

ASSOCIATED CONTENT

Supporting Information

General information, Tables S1–S3, and Figures S1–S5. This information is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

john.chaput@asu.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully acknowledge V. Pinheiro and P. Holliger for the polymerase RT521. We wish to thank members of the Chaput lab for helpful comments and suggestions. This work was supported by the Biodesign Institute at ASU.

REFERENCES

(1) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818–822.

(2) Robertson, D. L.; Joyce, G. F. *Nature* **1990**, *344*, 467−468.
(3) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505−510.
(4) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611−647.
(5) Famulok, M.; Hartig, J. S.; Mayer, G. *Chem. Rev.* **2007**, *107*, 3715−3743.
(6) Keefe, A. D.; Pai, S.; Ellington, A. D. *Nat. Rev. Drug Discovery* **2010**, *9*, 537−550.
(7) Mayer, G. *Angew. Chem., Int. Ed.* **2009**, *48*, 2672−2689.
(8) Joyce, G. F. *Angew. Chem., Int. Ed.* **2007**, *46*, 6420−6436.
(9) Szostak, J. W. *Trends Biochem. Sci.* **1992**, *17*, 89−93.
(10) Joyce, G. F. *Curr. Opin. Struct. Biol.* **1994**, *4*, 331−336.
(11) Herdewijn, P.; Marliere, P. *Chem. Biodivers.* **2009**, *6*, 791−808.
(12) Chaput, J. C.; Yu, H.; Zhang, S. *Chem. Biol.* **2012**, *19*, 1360−1371.
(13) Pinheiro, V. B.; Holliger, P. *Curr. Opin. Chem. Biol.* **2012**, *16*, 245−252.
(14) Joyce, G. F. *Science* **2012**, *336*, 307−308.
(15) Keefe, A. D.; Cload, S. T. *Curr. Opin. Chem. Biol.* **2008**, *12*, 448−456.
(16) Pinheiro, V. B.; et al. *Science* **2012**, *336*, 341−344.
(17) Loakes, D.; Holliger, P. *Chem. Commun.* **2009**, 4619−4631.
(18) Schoning, K.-U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347−1351.
(19) Orgel, L. E. *Science* **2000**, *290*, 1306−1307.
(20) Yang, Y.-W.; Zhang, S.; McCullum, E. O.; Chaput, J. C. *J. Mol. Evol.* **2007**, *65*, 289−295.
(21) Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. *J. Am. Chem. Soc.* **2002**, *124*, 13716−13721.
(22) Pallan, P. S.; Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. *Angew. Chem., Int. Ed.* **2003**, *42*, 5893−5895.
(23) Ebert, M.-O.; Mang, C.; Krishnamurthy, R.; Eschenmoser, A.; Jaun, B. *J. Am. Chem. Soc.* **2008**, *130*, 15105−15115.
(24) Chaput, J. C.; Ichida, J. K.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 856−857.
(25) Chaput, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 9274−9275.
(26) Kempeneers, V.; Vastmans, K.; Rozenski, J.; Herdewijn, P. *Nucleic Acids Res.* **2003**, *31*, 6221−6226.
(27) Ichida, J. K.; Horhota, A.; Zou, K.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Res.* **2005**, *33*, 5219−5225.
(28) Horhota, A.; Zou, K.; Ichida, J. K.; Yu, B.; McLaughlin, L. W.; Szostak, J. W.; Chaput, J. C. *J. Am. Chem. Soc.* **2005**, *127*, 7427−7434.
(29) Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 2802−2803.
(30) Yu, H.; Zhang, S.; Chaput, J. C. *Nat. Chem.* **2012**, *4*, 183−187.
(31) Zhang, S.; Chaput, J. C. *Curr. Protoc. Nucleic Acid Chem.* **2012**, *4*; DOI: 10.1002/0471142700.nc0451s50.
(32) Zhang, S.; Yu, H.; Chaput, J. C. *Curr. Protoc. Nucleic Acid Chem.* **2013**, in press.
(33) Wu, X.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **2002**, *4*, 1283−1286.
(34) Skelly, J. V.; Edwards, K. J.; Jenkins, T. C.; Neidle, S. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 804−808.
(35) Damha, M. J.; Wilds, C. J.; Noronha, A.; Brukner, I.; Borkow, G.; Arion, D.; Parniak, M. A. *J. Am. Chem. Soc.* **1998**, *120*, 12976−12977.
(36) Tabor, S.; Richardson, C. C. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4076−4080.
(37) Wrenn, S. J.; Weisinger, R. M.; Halpin, D. R.; Harbury, P. B. *J. Am. Chem. Soc.* **2007**, *129*, 13137−13143.
(38) Brudno, Y.; Birnbaum, M. E.; Kleiner, R. E.; Liu, D. R. *Nat. Chem. Biol.* **2010**, *6*, 148−155.
(39) Zamaratski, E.; Pradeepkumar, P. I.; Chattopadhyaya, J. *J. Biochem. Biophys. Methods* **2001**, *48*, 189−208.
(40) Deleavey, G. F.; Damha, M. J. *Chem. Biol.* **2012**, *19*, 937−954.
(41) Levy, M.; Miller, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7933−7938.
(42) Rogers, J.; Joyce, G. F. *Nature* **1999**, *402*, 323−325.