Development of an Artificial Genetic System Capable of Darwinian Evolution

by

Hanyang Yu

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2013 by the
Graduate Supervisory Committee:

John C. Chaput, Chair
Julian J.-L. Chen
Hao Yan

ARIZONA STATE UNIVERSITY

May 2013

ABSTRACT

The principle of Darwinian evolution has been applied in the laboratory to nucleic acid molecules since 1990, and led to the emergence of in vitro evolution technique. The methodology of in vitro evolution surveys a large number of different molecules simultaneously for a pre-defined chemical property, and enrich for molecules with the particular property. DNA and RNA sequences with versatile functions have been identified by in vitro selection experiments, but many basic questions remain to be answered about how these molecules achieve their functions. This dissertation first focuses on addressing a fundamental question regarding the molecular recognition properties of in vitro selected DNA sequences, namely whether negatively charged DNA sequences can be evolved to bind alkaline proteins with high specificity. We showed that DNA binders could be made, through carefully designed stringent in vitro selection, to discriminate different alkaline proteins.

The focus of this dissertation is then shifted to in vitro evolution of an artificial genetic polymer called threose nucleic acid (TNA). TNA has been considered a potential RNA progenitor during early evolution of life on Earth. However, further experimental evidence to support TNA as a primordial genetic material is lacking. In this dissertation we demonstrated the capacity of TNA to form stable tertiary structure with specific ligand binding property, which suggests a possible role of TNA as a pre-RNA genetic polymer. Additionally, we discussed the challenges in in vitro evolution for TNA enzymes and developed the necessary methodology for future TNA enzyme evolution.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my wife Zhe for her love, support and patience. Earning a Ph.D. degree in a foreign country would be an unimaginably difficult mission for me to accomplish if she was not by my side encouraging and spurring me on along the way. As a scientific researcher herself, she initiated numerous stimulating discussions with me and I benefited a lot from her sparkling ideas. I am especially indebted to her for her devoting more time to taking good care of our daughter. I am also thankful for my daughter Audrey, who brought innumerable little moments of fun and happiness to my life and also made me become a more independent and responsible person. I am grateful to my parents for their understanding and support. They provided everything I need to finish my study in China and encouraged me to study abroad even though knowing they would only be able to see me very infrequently thereafter. After Audrey was born, they traveled a long way from China to an unfamiliar country to see us and took over the chief responsibility of babysitting Audrey, so that I could focus more on my research. I owe them too much for their dedication and sacrifice. I want to express gratitude to my parents-in-law. They have been supporting Zhe and me in many ways. Their help of looking after Audrey was crucial to release our anxiety and stress as new parents.

I want to acknowledge my mentor, Dr. John C. Chaput. He inspired me to explore a thrilling field of science full of challenges, and provided a privileged opportunity to work on exciting research projects. He taught me how to design experiments seamlessly, how to present results passionately and how to write articles accurately. Under his guidance, I was able to grow from a student to a qualified researcher armed with critical

thinking and solid experimentation skills. He broadened my knowledge by encouraging me to attend seminars, conferences and to converse with world renowned scientists. I am also grateful to the members of my committee, Dr. Julian Chen and Dr. Hao Yan, for their time and advices. They challenged me to think more deeply about the significance of my projects and gave valuable suggestions on how to present research to a broad audience. I also want to thank all the former and present graduate students, postdoctoral fellows, and lab researchers who I have interacted with along the way. Their advice and support have been instrumental to my accomplishments. I want to especially thank my collaborators Dr. Noureddine Fahmi, Dr. Su Zhang, Bing Jiang and Matthew R. Dunn. Without their contributions, my research projects would not be as successful and complete as it is today.

I would like to thank Department of Chemistry and Biochemistry at ASU for accepting me into the Biochemistry PhD program and offering the opportunity to meet excellent faculty, talented students, brilliant researchers and friendly staff. I would also like to thank the Biodesign Institute and the Center for Evolutionary Medicine and Informatics for creating a collaborative and inspirational working environment. I would like to thank the Graduate & Professional Student Association at ASU for sponsoring me to go to an international conference and to share my research findings with the scientific community.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

APPENDIX Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

In Vitro Evolution of Natural and Artificial Nucleic Acids

1.1. Abstract

The principle of Darwinian evolution has been applied to individual biopolymers (nucleic acids and polypeptides) in the laboratory and led to a powerful technique named in vitro evolution. By exerting a pre-defined selection criterion on a random-sequence pool of enormous diversity, in vitro selection can isolate molecules with particular properties such as ligand binding and catalysis as rare as 1 in 10^{14}. For over two decades, laboratories all over the world have employed in vitro evolution to generate high affinity reagents to bind a wide range of targets and novel enzyme motifs to catalyze specific chemical reactions. The best example of in vitro evolution in practical application is the identification of Macugen, a therapeutic RNA approved by the Food and Drug Administration. However, natural nucleic acids suffer from their susceptibility to nuclease degradation. Consequently, chemists and biologists have been attempting to expand the technology of in vitro evolution to modified nucleic acid analogs. This field is referred to as synthetic genetics, which explores the structural and functional properties of synthetic genetic polymers by in vitro evolution.

This chapter is largely focused on one type of synthetic genetic polymer called threose nucleic acid (TNA). TNA is different from DNA and RNA in that it contains a threose in place of (deoxy)ribose in RNA (or DNA). This chapter extensively summarizes the initial discovery of TNA, the following structural analysis and a lot of efforts towards the goal of in vitro evolution of TNA.

1.2. In Vitro Selection of Natural Nucleic Acids

Darwinian evolution describes how a population changes over time in a particular environment condition (Figure 1.1A). Briefly, a common ancestor can generate multiple progenies with different mutations. These mutations can be harmful, beneficial or neutral to the survivability of the progenies under a certain selection pressure. Consequently, some progenies can survive and reproduce while others will become extinct, which ultimately leads to differential reproduction.

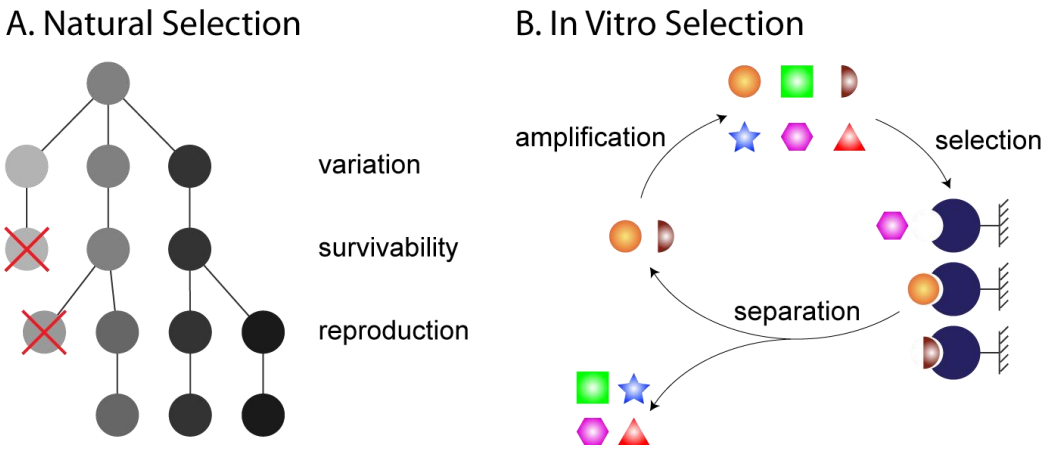


Figure 1.1. Natural selection and in vitro selection. (A) Darwinian evolution in nature drives differential survival based on effects of variations. (B) Darwinian evolution in a test tube generates molecules with pre-defined functions.

Although Darwinian evolution is often discussed on the level of organisms and species, the same principle can also be applied to individual molecules in the laboratory (Figure 1.1B). This technique is called in vitro selection. Typically, a pool of random sequences, which is referred to as a library, is first generated to mimic a diverse population and to initiate the selection. This library is subjected to a pre-defined selection criterion, such as ability to bind a target. Functional molecules are separated from the

nonfunctional ones (e.g. binders vs. non-binders) and subsequently amplified to re-generate a pool enriched with the desired property. This process of selection and amplification is repeated iteratively until the pool is dominated by functional sequences.

1.2.1. In Vitro Selection of Nucleic Acid Aptamers. In vitro selection can be easily applied to nucleic acid molecules because single-stranded RNA can not only fold into stable tertiary structures (phenotype for selection) but also act as template sequences (genotype) in amplification reaction. In the first example of in vitro selection for RNA molecules that bind specific ligands,[1] Ellington and Szostak synthesized a pool of 10^{13} different RNA sequences by in vitro transcription. These RNA sequences all have a random region of 100 nucleotides flanked by defined sequences necessary for primer hybridization. The pool was incubated with different organic dyes attached to affinity columns. These dyes appear to mimic metabolic cofactors and have possible hydrogen-bond donor and acceptor groups as well as planar surfaces for stacking interactions. Seven parallel in vitro selections were performed and yielded RNAs with different discriminating specificity. These individual RNA sequences were termed "aptamers", from the Latin word 'aptus' meaning "to fit". In the same year 1990, Tuerk and Gold synthesized an RNA library with an eight-base randomized region and selected for sequences that could bind T4 DNA polymerase.[2] Starting from over 60000 non-identical sequences, they identified two binders by repeated cycles of selective amplification. This method was called systematic evolution of ligands by exponential enrichment (SELEX). Conceivably, the protocol of SELEX or in vitro selection could be developed to yield high-affinity RNA ligands for any target molecule.

Two years later after the original in vitro selection for RNA aptamers, Ellington and Szostak selected DNA aptamers for the same organic dyes.[3] Starting from 10^{13} different DNA sequences containing 120 bases of random region flanked by defined primer-binding sites, they isolated DNA sequences that could fold into specific ligand-binding structures. Interestingly, these DNA sequences, if constructed in RNA backbone, would not interact with the same ligands. And this phenomenon has been observed in multiple cases ever since. The selection of DNA aptamers suggested that DNA has at least similar functional potential to RNA, and might be developed into better pharmacological reagents than RNA due to its greater stability.

One important characteristic of aptamer binding is that it can be very specific towards its cognate target. In a classic example by Polisky and co-workers (Figure 1.2A),[4] RNA aptamers were selected to bind theophylline specifically. In order to improve the selection stringency, they designed a "counter-SELEX" step in which caffeine, an analog of theophylline, was used to competitively elute RNAs bound to theophylline with low selectivity. Indeed, the selection yielded an RNA aptamer that could bind theophylline with a dissociation constant (K_d) of 0.32 μM and more importantly, the K_d of this aptamer binding caffeine is 10000-fold higher. This discriminative binding is pronounced especially considering caffeine differs from theophylline only by one additional methyl group at N-7 position. More recently, Chaput and co-workers have selected DNA aptamers that could recognize histone peptide with subtle post-translational modifications (Figure 1.2B).[5] Histone H4 peptide with acetylated lysine at position 16 (H4K16Ac) was chosen because of its importance in regulating gene activation and silencing. The in vitro selection included similar negative and positive

selection steps, and yielded an aptamer that could bind H4K16Ac with a K_d of 21 nM and discriminate unmodified H4 peptide by 2400-fold. This level of selectivity outperforms commercial antibody raised to bind H4K16Ac target, which shows only 15-fold specificity against H4 off-target sequence.

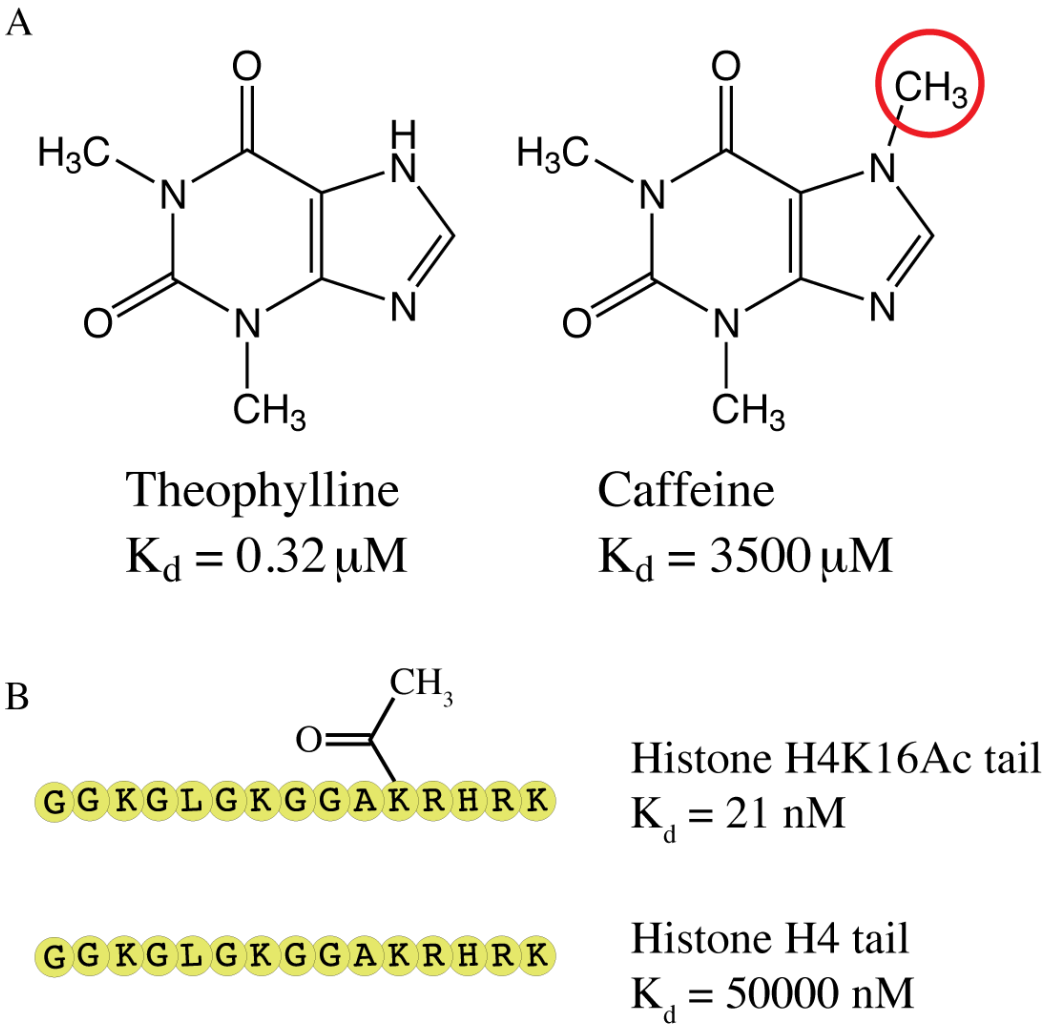


Figure 1.2. Aptamers show high specificity towards cognate target. (A) RNA aptamer distinguishes theophylline from caffeine by 10000-fold.[4] (B) DNA aptamer specifically recognizes acetylated lysine by 2400-fold selectivity.[5]

1.2.2. In Vitro Selection of Nucleic Acid Enzymes. In vitro selection can be also used to isolate nucleic acid structures with catalytic activities. In the first example of selecting RNA enzymes (ribozymes) that could specifically cleave single-stranded DNA,[6] Robertson and Joyce started with the *Tetrahymena* ribozyme, a self-splicing group I intron that catalyzes phosphoester transfer reactions on RNA substrates (Figure 1.3A). This involves attack by the 3'-terminal guanosine of the ribozyme at a phosphodiester bond within an RNA substrate. The product is the ribozyme joined to the substrate sequence downstream of the target phosphodiester bond. This ligation product offers a selectable tag for isolating catalytically active RNA sequences. The selection occurs during reverse transcription (RT) because the RT primer is hybridized across the ligation junction and initiate synthesis of complementary DNA only from catalytic RNAs. By performing an iterative process of mutation, selection and amplification, they identified a variant that could work on a DNA substrate more efficiently than the wildtype ribozyme. This example of in vitro mutagenesis and selective amplification is referred to directed evolution. Directed evolution often starts with a scaffold already showing a certain functional property, introduces random mutations to sites important for its function, and aims to either improve its original activity or alter to create a novel function. Different from in vitro selection, which typically starts with a random-sequence library and surveys an unbiased sequence space, directed evolution is focused on a local function optimum and examines the neighborhood in a sequence-function landscape.

After the identification of ribozymes in the laboratory, Breaker and Joyce reasoned that DNA might also be able to function as an enzyme considering the similarity of DNA and RNA. To test this hypothesis, they carried out in vitro selections to enrich

for DNA enzymes (deoxyribozymes) that could promote divalent metal ion dependent cleavage of an RNA phosphodiester bond embedded within an otherwise all-DNA sequence (Figure 1.3B).[7,8] Starting with 10^{14} different DNA sequences containing 50 random nucleotides, they isolated catalytically active molecules by incubating the library immobilized on affinity column with divalent metal ions and recovering DNAs that could catalyze RNA hydrolysis reaction and thus cleave themselves off the column. RNA-cleaving deoxyribozymes dependent on Pb^{2+}, Zn^{2+}, Mn^{2+}, Mg^{2+}, respectively, were identified and showed about 10^5-fold rate enhancement compared with uncatalyzed reaction. In an effort to identify deoxyribozymes that could cleave any RNA substrate, Santoro and Joyce carried out a similar in vitro selection using an extension primer containing 12, instead of one, embedded ribonucleotides as potential cleavage sites.[9] This selection generated two deoxyribozymes 8-17 and 10-23. They both promote RNA cleavage by forming two substrate binding arms (7 base pairs each) and one small catalytic core (13 and 15 nucleotides, respectively). Enzyme 8-17 cleaves between an A and G, while 10-23 can cleave the linkage between a purine and a pyrimidine. Deoxyribozyme 8-17 has been observed for multiple times in the subsequent independent in vitro selections ever since, suggesting this motif presents the simplest solution for catalyzing RNA cleavage reaction. Additionally, variants of 8-17 have been made to be able to cleave all 16 dinucleotide junctions, making it a general RNA-cleaving enzyme that can target any RNA substrate sequence.[10,11] Enzyme 10-23 cleaves between a purine and a pyrimidine residue, and thus in principle could target any mRNA start codon (A-UG). As a proof of concept, deoxyribozyme 10-23 was prepared to target a 17-mer RNA corresponding to the translation initiation region of HIV-1 gag/pol mRNA sequence. The

catalyzed RNA cleavage reaction proceeded with a k_{cat} of 0.15 min^{-1} and K_m of 0.47 nM under simulated physiological conditions.[9]

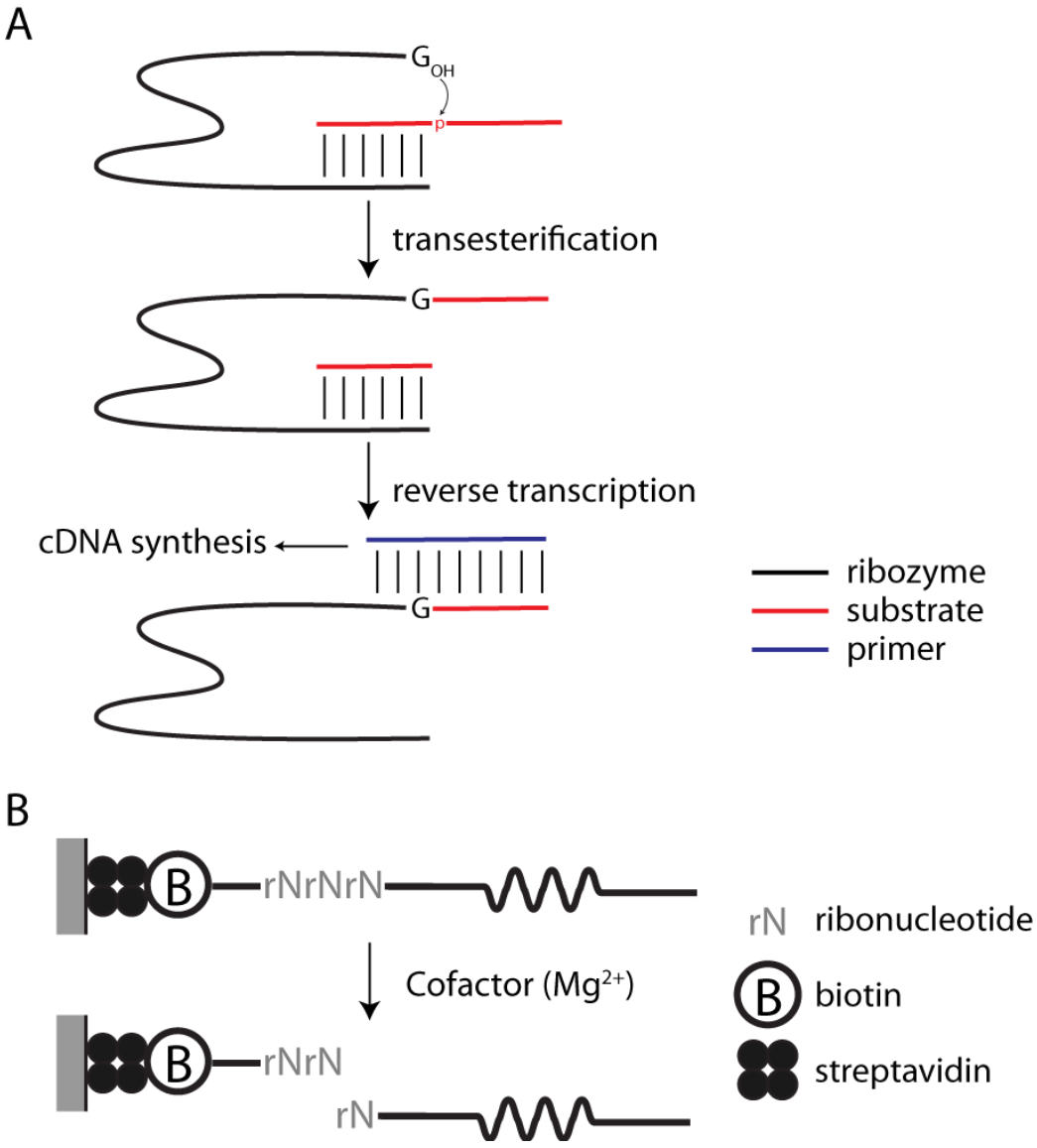


Figure 1.3. Nucleic acid enzyme selection. Selection for catalysis usually involves (A) self-tagging[6] or (B) self-cleaving.[9]

1.3. In Vitro Selection of Modified Nucleic Acids.

Nucleic acid aptamers and enzymes have been selected to perform a variety of functions,[12,13] and thus represent a novel class of diagnostic and therapeutic reagents. Of

particular interest are aptamers that show high-affinity and high-specificity binding to biological target molecules, because such binding events often alter or inhibit the biological functions of the target molecules. However, aptamers composed of natural nucleic acid sequences do not perform well in vivo due to their intrinsic sensitivity to nucleases. Consequently, different chemical modifications have been introduced to RNA and DNA structures to create nuclease-resistant binders. This is not trivial because 1) modified monomer building blocks are often not commercially available and have to be chemically synthesized in the laboratory and 2) natural DNA and RNA polymerases have strict substrate selectivity and do not always recognize unnatural nucleoside triphosphates. Fortunately, recent advances in both organic synthetic chemistry and polymerase engineering and evolution have enabled several successful demonstrations of in vitro selections on artificial genetic polymers. This field is referred to as synthetic genetics, which explores the functional and structural properties of synthetic genetic polymers by in vitro selection.[14] Because these unnatural nucleic acids harbor different chemical modifications, synthetic genetics opens the door to an unexplored sequence space with novel functionalities and structural properties.

Modifications have been introduced to different moieties on DNA and RNA structures: nucleobase, furanose ring and phosphodiester linkage. In a recent example of selecting DNA aptamers with modified bases,[15] Eaton and co-workers synthesized new dUTP derivatives with different substitutions at C-5 position. As expected, these triphosphate derivatives could not be incorporated into DNA by polymerase chain reaction (PCR). Instead, they could be efficiently recognized by two archaeal DNA polymerases in primer extension reactions. Several of these modified dUTPs were used in

an in vitro selection for selecting aptamers to a protein target necrosis factor receptor superfamily member 9. After eight rounds of selection, the survived DNA pool showed an apparent K_d of 4 nM towards the target protein. In contrast, a parallel selection using all natural triphosphate substrates resulted in a round 8 DNA pool that did not show binding for the target protein at up to 100 nM. These data clearly showed the advantage of adding more functional groups in aptamer selection.

Nucleoside triphosphates with modified bases are typically well tolerated by polymerases, but nucleotides with altered backbones are normally poor substrates for natural enzymes. This latter class includes subtle changes, such as 2' modification on sugar ring with amino, fluoro or methoxy groups, to dramatic changes like substituting sugar-phosphate backbone with protein-like peptide bond linkage. Recognition of these analogs often requires rational design and/or directed evolution of polymerase enzymes. A recent review article by Chaput extensively examines the in vitro selection of artificial genetic polymers with modified backbones.[14]

1.4. Threose Nucleic Acid

1.4.1. Initial Synthesis of TNA. In an effort to uncover the criteria by which nature chose RNA, or more precisely β-D-ribofuranosyl (5'->3') nucleic acid, as the molecular basis of life's genetic system, Eschenmoser led a systematic research program to synthesize and evaluate the chemical properties of natural nucleic acid structure alternatives.[16] The project focused on oligonucleotide systems with potentially natural sugars ranging from hexose to tetrose. The rational is that whatever prebiotic chemistry gave rise to RNA should have also produced a series of analogs, and that these analogs in principle might have worked as alternative genetic materials for Earth's early organisms.

Perhaps one of the most important criteria for an unnatural nucleic acid to work as a genetic polymer is the capacity to form stable intrasystem base pairing, which is essential for self replication. Eschenmoser and co-workers have chosen three families of oligonucleotide systems (Figure 1.4) and studied their base pairing properties with themselves and with natural nucleic acids.

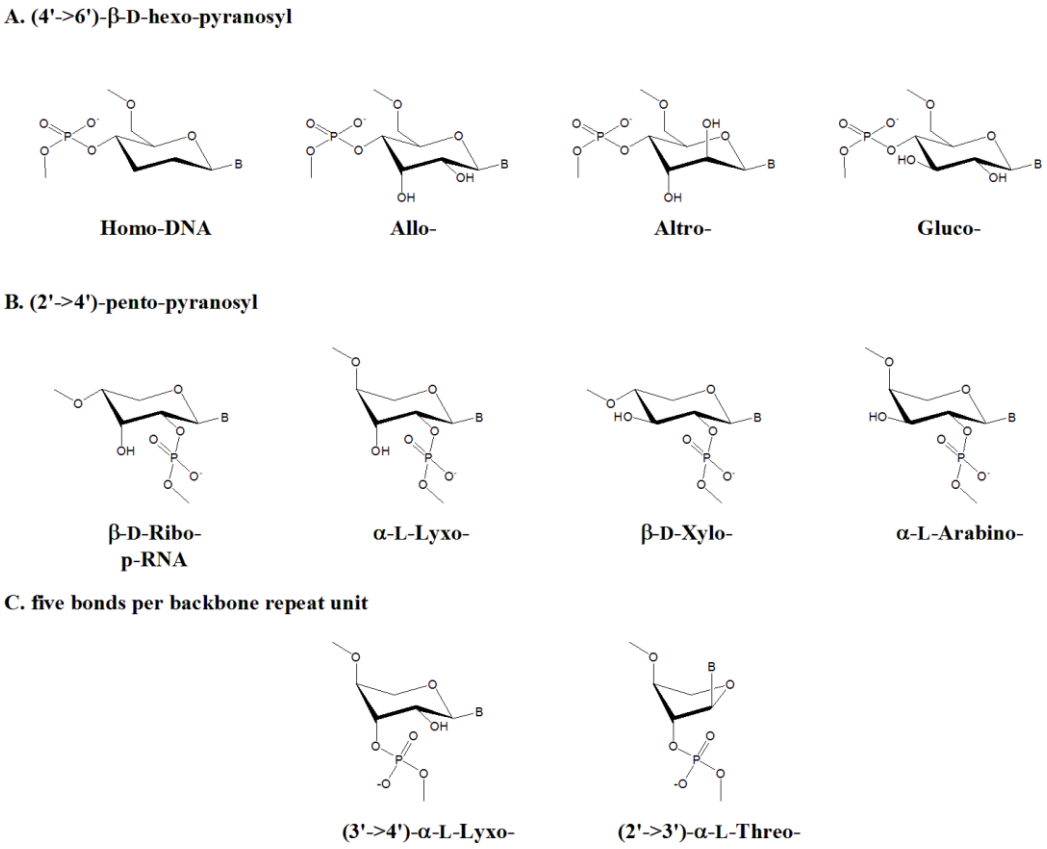


Figure 1.4. Constitution and configuration of the repeating units of selected nucleic acid alternatives studied by Eschenmoser.[16] (A) (4'->6')-β-D-hexo-pyranosyl oligonucleotide systems with homo-DNA as a model. (B) (2'->4')-pento-pyranosyl oligonucleotide systems with pRNA as a prototype. (C) oligonucleotide systems with five, instead of six, bonds in the phosphodiester backbone repeating unit. B: base.

1.4.1.1. β-D-Hexo-pyranosyl Oligonucleotide System. The project of examining alternative nucleic acid structures was initiated with a model study of "homo-DNA", which differs from DNA by expanding the deoxyribose furanose ring with one additional methylene group to form dideoxyallose pyranose ring with 4'->6' linkage (Figure 1.4A). Homo-DNA was the first experimental example of a backbone-modified oligonucleotide system with stronger Watson-Crick base pairing than DNA. Moreover, its base pairing mode is orthogonal to natural nucleic acid systems. However, homo-DNA was never considered a potentially natural nucleic acid alternative but only a model system for the family of β-D-hexopyranosyl oligonucleotides, because its pyranose ring does not have the 2' and 3' hydroxyl groups and is thus not directly derived from a natural hexose.

Three oligonucleotide systems within this family were studied: β-D-allo-, β-D-altro-, and β-D-gluco-pyranosyl oligonucleotides. Different from the observations of homo-DNA, none of these systems exhibit appreciable Watson-Crick base pairing (Figure 1.5).[16] The contrast between homo-DNA and its fully hydroxylated counterparts was interpreted as the result of steric hindrance caused by the hydroxyl groups. Based on the NMR structure of homo-DNA duplex, insertion of a hydroxyl group to the C2' equatorial position would lead to a steric clash of this OH group with the neighboring nucleobase. Indeed, the 2'-deoxy analog of β-D-allo-pyranosyl nucleic acids show similar melting temperature as homo-DNA, while the 3'-deoxy system has thermostability as weak as the β-D-allo-pyranosyl series.

These results led to the conclusion that hexopyranosyl-(6'->4') oligonucleotide system could not act as viable competitors of RNA as nature's genetic

polymer. Because the steric bulk of hexopyranose ring was responsible for its inability to base pair, the research focus was shifted to less bulky pentose-derived nucleic acid alternatives, no longer asking "why pentose and not hexose" but rather "why ribose and not other pentose" and "why ribofuranose and not ribopyranose".

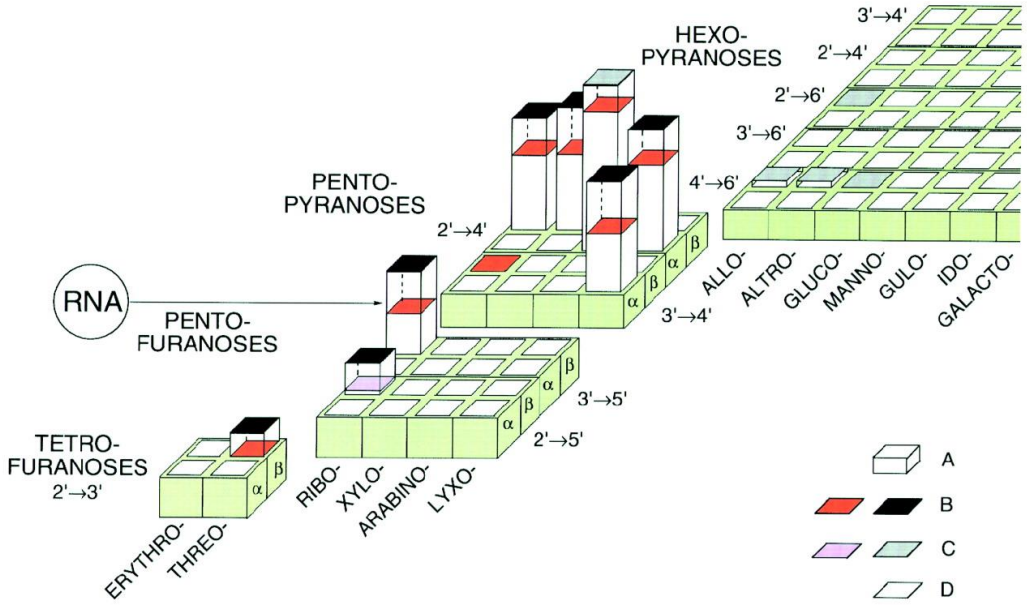


Figure 1.5. Pairing-strength landscape of hexose-, pentose-, and tetrose-derived oligonucleotide systems. (A) Column height unit is 10 ℃ above 0 ℃; (B) observed T_m values, red for A_8:T_8 and black for A_{12}:T_{12}; (C) estimated T_m values, purple for A_8:T_8 and gray for A_{12}:T_{12}; and (D) not investigated. From Eschenmoser, A. *Science* **1999**, *284*, 2118-2124. Reprinted with permission from AAAS.

1.4.1.2. Pento-pyranosyl Oligonucleotide System. This family has four diastereoisomeric members: β-D-ribo-, α-L-lyxo-, β-D-xylo-, and α-L-arabino-pyranosyl oligonucleotide systems with 2'->4' linkage (Figure 1.4B). All the members form stronger intrasystem Watson-Crick base pairing than RNA itself, and they exhibit cross pairing with other members within the family (Figure 1.5). However, they do not cross

pair with natural RNA.[16] These results imply that although these systems form Watson-Crick base pairing within the family, there seems to be a different Watson-Crick "language" that natural nucleic acids use, and consequently members in this family can not "communicate" with natural RNA polymers. It suggested that pentopyranosyl nucleic acids might have existed as competitors of RNA, but very unlikely as ancestors, at least not immediate ancestors due to the lack of a direct exchange mechanism of genetic information. It also led to the conclusion that nature did not choose RNA based on the criterion of maximal base pairing strength, because the melting temperature of natural RNA duplex is inferior than that of any member in this family.

The primary object within this family is β-D-ribopyranosyl nucleic acid (pRNA), which is the pyranosyl isomer of natural RNA. An NMR structure of pRNA duplex derived from sequence CGAATTCG revealed a ladder-like duplex structure with a weak left-handed twist. Moreover, the base stacking is overwhelmingly interstand instead of intrastrand as in natural nucleic acid systems such as B-DNA. This quasi-linear structure of pRNA is very different from that of RNA and explains how pRNA could cross pair with other family members but not with natural RNA.

1.4.1.3. α-L-Tetro-furanosyl Oligonucleotide System. The RNA analogs derived from tetrose were initially not considered because an oligonucleotide backbone containing six bonds per repeat unit (as RNA does) can not be constructed from a four-carbon sugar. A tetrose-derived nucleic acid has at most five bonds in its backbone repeating unit, and was thought unable to form stable base pairs with RNA simply because it can not stretch itself long enough to bind RNA properly. However, an observation in the (4'->2') pentopyranosyl system changed the perspective. Switching

from 2'->4' linkage to 3'->4' linkage in α-L-lyxopyranosyl series results in shortening of the phosphodiester bridge from six to five bonds (Figure 1.4C), but did not lead to loss of base pairing as expected (Figure 1.5).[17] It turns out this switch is accompanied by the change of the orientation of one of the C-O bonds in the bridge from equatorial to axial. By adopting a quasi-trans-diaxial position, the five bonds can bridge a wider gap than originally thought and maintains the base pairing property.

Realizing the first impression that six-atom repeat is essential for alternative genetic polymers is incorrect, Eschenomser and co-workers synthesized a tetrose-based nucleic acid alternative called α-L-threofuranosyl nucleic acid (TNA), and studied its intra- and inter-system base pairing properties.[18] They discovered that TNAs do form stable Watson-Crick double helices and, perhaps more importantly, that TNAs form stable base pairing with complementary strands of RNAs and DNAs. This unique cross pairing property of TNA provides a plausible mechanism for the transfer of information between two different genetic systems, for example, TNA and RNA.

Generally speaking, TNA duplexes show comparable thermostability to that of RNA and DNA duplexes. But it does show some sequence dependence: duplexes with strands composed of alternating purine-pyrimidine bases are more stable than the homo-oligomeric strands with the same length. In fact, the homobasic strands display exceptional behavior in that their thermostability is unusually low. For example, the poly(A) and poly(T) strands of TNA show no pairing as octamers A_8 and T_8, and the dodecamer A_{12} exhibited weak pairing with T_{12} (T_m = 14 ℃). The melting temperature was eventually increased to 40 ℃ for 16-mers and 53 ℃ for 20-mers (Table 1.1).[19]

Table 1.1. T_m values (℃) of duplexes from intra- and inter-system base pairing.[19]

Entry	Duplex strands	X = T	X = U
1	$t(A_8) + t(X_8)$	-	-
2	$t(A_{12}) + t(X_{12})$	14.6	-
3	$t(A_{16}) + t(X_{16})$	39.8	12.0
4	$t(A_{20}) + t(X_{20})$	52.6	25.7
5	$t(A_8X_8)$	38.3	18.2
6	$t((A_4X_4)_2)$	45.1	29.5
7	$t((A_2X_2)_4)$	54.9	44.4
8	$t((AX)_8)$	74.5	68.6
		Y = A	Y = D
9	$t(Y_{12}) + t(T_{12})$	14.6	45.7
10	$t(Y_4T_3YTYT_2YT_2Y) + t(TA_2TA_2TATA_3T_4)$	56.2	66.6
11	$t(A_4T_3ATAT_2AT_2A) + t(TY_2TY_2TYTY_3T_4)$	56.2	66.9
12	$t(Y_4T_3YTYT_2YT_2Y) + t(TY_2TY_2TYTY_3T_4)$	56.2	78.8
13	$t(Y_4T_3YTYT_2YT_2Y) + r(UA_2UA_2UAUA_3U_4)$	52.0	65.0
14	$r(A_4U_3AUAU_2AU_2A) + t(TY_2TY_2TYTY_3T_4)$	57.0	74.0

-: no detectable pairing. t: TNA. r: RNA.

Besides its unique base pairing properties, TNA also stands out among natural and artificial nucleic acid systems because of its chemical simplicity. A tetrose-derived structure allows for reactant economy in monomer formation that could be based on $C_2 + C_2 \rightarrow C_4$ chemistry, whereas the corresponding derivation for ribose in RNA

would require two different building blocks (C_2 + C_3 -> C_5) and very likely yield tetrose and hexose byproducts.[20]

The unexpected ability of TNA to mimic RNA in a Watson-Crick double helix, combined with the chemical simplicity of threose synthesis relative to ribose under prebiotic conditions, suggests that TNA might have been a precursor of RNA.[20,21] The criteria for a possible RNA progenitor include: 1) potentially derivable using prebiotic chemistry and simpler than RNA; 2) capacity to exchange information with RNA; 3) capacity to store and replicate genetic information; 4) capacity to express a chemical phenotype (receptor and catalysis) and 5) potential for evolution.[21] TNA meets the first three requirements, at least in theory, as demonstrated in Eschenmoser's investigation. The catalytic potential of TNA is restricted by the absence of a free hydroxyl group compared with RNA, and the diversity of catalytic activities might be an important consideration of why nature arrived at the choice of RNA. For the same reason, the hydrolytic stability of TNA is similar to that of DNA, with no detectable decomposition after months, while RNA decomposes rapidly with a half-life of 12 h under slightly basic condition (pH = 8) at 35 ℃.[19] Therefore, an RNA progenitor might have had limited catalytic activity combined with high chemical stability. Nevertheless, answering the question of whether TNA polymer could be catalytically active is a daunting challenge because TNA polymers are not known to exist naturally and can not be identified by the same process of initial discovery of ribozymes. Evaluating the potential of TNA to evolve requires an experimental re-creation of TNA replication in the laboratory and accessing the ability of TNA to mutate and to inherit beneficial variations under certain selection pressures.

1.4.2. Synthesis of TNA Phosphoramidites and Triphosphates. L-threose was prepared from the cheaply available starting material L-ascorbic acid in six steps, which yielded protected L-threofuranosyl ring as a mixture of α/β anomers.[19] The protected nucleobases were then glycosylated with the protected sugars to afford L-threofuranosyl nucleosides. The 2' and 3' hydroxyl groups were deprotected to remove benzoyl groups and re-protected by a DMT (4,4'-dimethoxytrityl) group. As expected, this would yield a mixture of 2'-O-DMT and 3'-O-DMT regioisomers. The 3'-isomer could be used for phosphoramidite synthesis,[19] while the 2'-isomer could be used for triphosphate synthesis (Figure 1.6).[22]

TNA phosphoramidites with six different bases were made for the synthesis of TNA oligonucleotides: adenine (A), thymine (T), cytosine (C), guanine (G), uracil (U) and 2,6-diaminopurine (D). TNA base pairing strength was studied in terms of T_m values for different sequences. Interestingly, T:A base pair seems to increase the thermostability of TNA duplexes compared with U:A base pair. For example, A_{16}:T_{16} TNA duplex has a T_m of 39.8 ℃ while A_{16}:U_{16} duplex has a T_m of only 12.0 ℃ (Table 1.1).[19] As expected, Replacement of A with D has a dramatic effect on duplex stability because D can form one additional hydrogen bond with T. For example, the T_m increases by over 30 ℃ when t(D_{12}) instead of t(A_{12}) is paired with t(T_{12}) (Table 1.1).[19] It has also been shown that A->D substitution could improve the efficiency of TNA-templated TNA ligation[23] and DNA-templated TNA polymerization.[24,25] TNA triphosphates with four different bases (A, D, C and G) were made and used as substrates in enzyme-mediated TNA polymerization.[22]

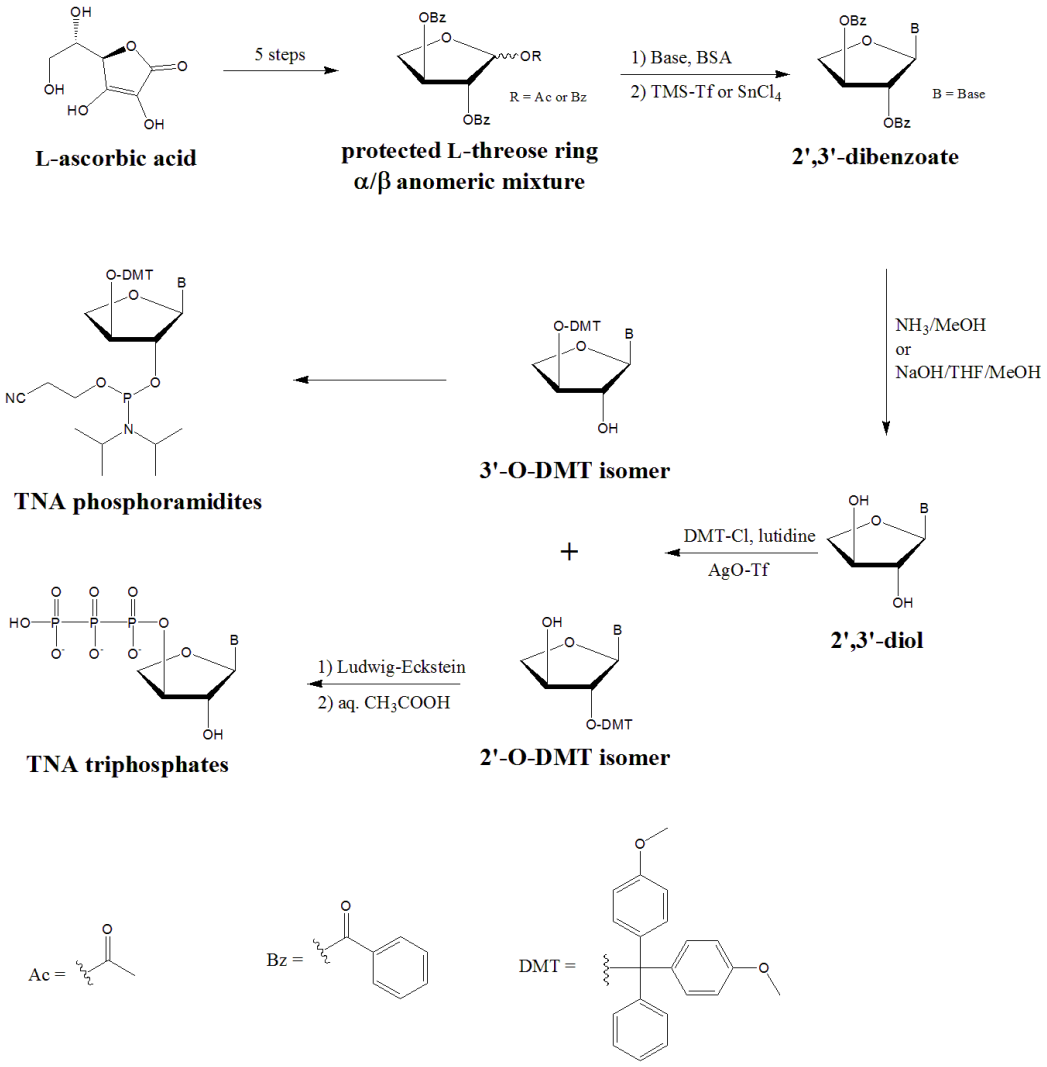


Figure 1.6. Synthesis scheme of TNA monomer building blocks.[19,22]

TNA analogs containing phosphoramidate linkages have also been constructed. This includes (3'-NH)-TNA and (2'-NH)-TNA and they can cross pair with TNA as well as RNA and DNA.[26] Comparison of T_m values of homo-duplex revealed that the intrasystem base pairing strength decreases in the following order: (3'-NH)-TNA > RNA ≈ TNA > DNA > (2'-NH)-TNA. The (3'-NH)-TNA has been shown stable towards hydrolytic cleavage under slightly basic condition, similar to unmodified TNA itself. But

in acidic solution (pH = 0.3) the half-life of (2'-NH)-TNA is only 1 h at room temperature.

TNA polymer has also been show to promote nonenzymatic oligomerization of ribonucleotides in a template-directed manner.[27] In this particular case, an RNA-TNA chimeric strand forms a hairpin structure with an overhanging stretch of seven TNA cytidine residues as template. Incubation of such a hairpin template with activated RNA monomer building blocks guanosine 5'-phosphoro-2-methylimidazole led to full length RNA polymerization product after ten days. And digestion with RNase T1 revealed that the phosphodiester bonds formed are 3', 5'-linked. This demonstration models the "genetic takeover" process from a pre-RNA polymer to RNA.

1.4.3. Structure of TNA Duplex. The unique base pairing properties of TNA has inspired several laboratories to investigate the structural basis of how TNA polymer could arrange its shorter backbone to form stable base pairing with natural nucleic acids (or alternatively, how DNA and RNA could adapt to the geometric constraints imposed by TNA).

The first crystal structure of a B-form DNA duplex containing one single TNA residue in each strand was solved by Egli and co-workers in 2002.[28] The sequence of Dickerson-Drew dodecamer (DDD) CGCGAATTCGCG was used and the 7th residue was replaced by a TNA thymine nucleoside. The structure of such modified B-form duplex [d(CGCGAA)tTd(TCGCG)]$_2$ was determined at 1.2 Å resolution. The overall duplex conformation was nearly identical to the native all-DNA reference DDD duplex. In fact, almost all the residues were superimposable with the corresponding nucleotides in native duplex, except the modified threose sugar rings and the adjacent phosphate

groups. The sugar rings adopted C4'-*exo* pucker and the 3' and 2' C-O bonds occupied a quasi-trans-diaxial orientation. Assuming such an arrangement, the 3' and 2' phosphate groups were separated farthest away from each other, which could partially compensate for its shorter backbone when it formed hetero-duplex with DNA or RNA. Even at this maximized separation, the distance between these two phosphorus atoms was 5.80 Å (6.09 Å in the other strand), significantly shorter than the distance between the adjacent phosphates in 2'-deoxyribonucleotides in B-form DNA (Figure 1.7A-B). Remarkably, the introduction of such a modified residue caused conformational changes only at the local level and without affecting the overall duplex structure. The base stacking interactions were virtually unchanged, consistent with the thermostability of TNA duplex and the capacity of TNA to cross pair with RNA and DNA.

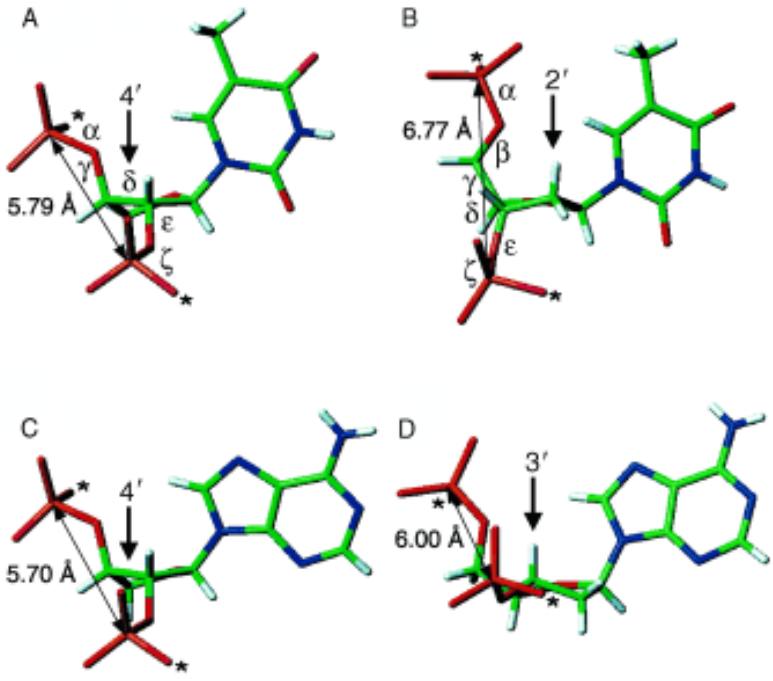


Figure 1.7. Conformations adopted by TNA and DNA nucleotides in B- and A-form duplexes. A) B-TNA; B) B-DNA; C) A-TNA; D) A-DNA. The sugar pucker is 4'-*exo* for

B-TNA and A-TNA, 2'-*endo* for B-DNA and 3'-*endo* for A-DNA. The distances between neighboring phosphate groups are shown. From Pallan, P. S. et al. *Angew. Chem. Int. Ed.* **2003**, *42*, 5893-5895. Reprinted with permission from John Wiley and Sons.

In 2003 a similar crystal structure of DNA duplex containing one TNA residue per strand was solved, aiming to answer the question of why TNA cross pairs more strongly with RNA than with DNA.[29] A decamer with the sequence GCGTATACGC was used where the T6 residue was modified as a 2'-O-methyl analog to lock the duplex in the A-form geometry. And the A7 residue was replaced by a TNA adenine nucleotide to probe its local conformation as well as its effect on the overall structure. It was found that within an A-form duplex the adenine TNA nucleotides also adopted C4'-*exo* sugar pucker and the intranucleotide P_i-P_{i+1} distance was 5.70 Å (Figure 1.7C). These observations were very similar to the behaviors of TNA residues in the B-form DNA context, implying that TNA existed in a rigid conformation independent of its environment. Indeed, the preferred sugar pucker of threose is C4'-*exo* and the vicinal phosphate groups occupy the quasi-trans-diaxial position when TNA is embedded within an A-form or B-form duplex or present as an isolated nucleoside. And the distance of neighboring phosphate groups at 5.80 Å is probably the upper limit of the intranucleotide P_i-P_{i+1} distance in TNA. Then the ability of TNA to cross pair with DNA must be explained by the more flexible structure of DNA that could conformationally adjust to meet the structural requirements of TNA, but not vice versa. In other words, in order to form a hetero-duplex with TNA, the deoxyribose rings in DNA strand have to adopt C3'-*endo* puckering to reduce the neighboring phosphate distance from approximately 7 Å (B-from) to 6 Å (A-form) (Figure 1.7D). But TNA polymer cannot stretch itself long

enough to be able to base pair with B-form DNA. This structural analysis offered a convincing rationalization of 1) why TNA hybridizes more strongly with RNA than with DNA and 2) why RNA constitutes a better template in directing TNA ligation than DNA.

An all-TNA duplex NMR structure was reported in 2008 (Figure 1.8).[30] The TNA octamer CGAATTCG formed a curved right-handed double helix with antiparallel strands pairing in Watson-Crick mode. While some helical parameters of TNA duplex such as base pair stretch were very similar to those of A- and B-DNAs, others such as stagger were different from A- and B-helices (Table 1.2). The average value of base pair step rise was 3.1 Å, smaller than that in canonical A- and B-form DNAs but still within the optimal base-stacking distance of 3.3 Å. The slide value between two adjacent base pairs was between -1.1 and -1.5 Å, which is close to that of A-DNA (-1.53 Å) but different from B-DNA (0.23 Å). The x-displacement value in a helix measures the diameter of the cavity in the center of the helix. It is almost zero in B-DNA (0.05 Å) but distinctly negative in A-DNA (-4.17 Å). The average x-displacement for TNA duplex is -4.6 Å, which again suggests TNA is structurally similar to A-DNA and RNA. Consistent with previous results, the P_i-P_{i+1} distance in TNA duplex is around 5.9 Å, similar to that shown by A-type DNA (6.0 Å) but different from that of B-DNA (6.8 Å). The similarity of TNA and RNA could be further seen in the shallow and wide minor groove in TNA helix, which closely resembled that of RNA helix but different from the narrow and deep minor groove in B-DNA duplex. Due to its short sequence, a comparison for the major groove in TNA duplex is difficult.

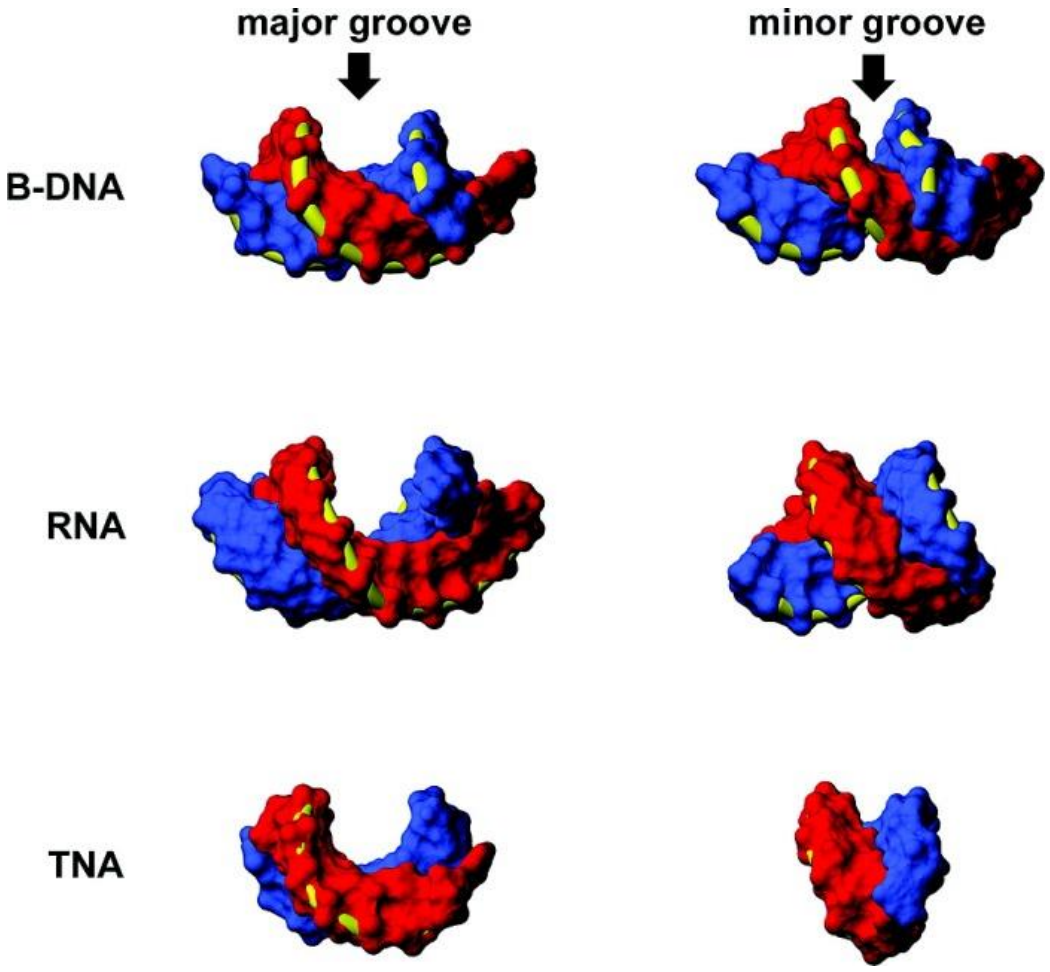


Figure 1.8. Comparison of groove widths in B-DNA, RNA and TNA duplexes. From Ebert, M. O.; Mang, C.; Krishnamurthy, R.; Eschenmoser, A.; Jaun, B. *J. Am. Chem. Soc.* **2008**, *130*, 15105-15115. Reprinted with permission from American Chemical Society.

Table 1.2. Comparison of selected helical parameters of TNA duplex with A- and B-DNA.[30,40]

Parameter	TNA	A-DNA	B-DNA
Complementary base-pair parameters			
Stretch (Å)	-0.2 ± 0.1	-0.18 ± 0.10	-0.15 ± 0.12
Stagger (Å)	0.2 ± 0.1	0.02 ± 0.25	0.09 ± 0.19
Base-pair step parameters			
Slide (Å)	-1.3 ± 0.2	-1.53 ± 0.34	0.23 ± 0.81
Rise (Å)	3.1 ± 0.4	3.32 ± 0.20	3.32 ± 0.19
Local helical parameters			
x-displacement (Å)	-4.6 ± 0.6	-4.17 ± 1.22	0.05 ± 1.28
Helical rise (Å)	2.3 ± 0.4	2.83 ± 0.36	3.29 ± 0.21
Inclination (degree)	24.1 ± 3.0	14.7 ± 7.3	2.1 ± 9.2

All the structural analyses of TNA, including TNA residues in B-DNA, A-DNA, all-TNA helices and as monomer nucleoside derivatives, revealed a quasi-trans-diaxial arrangement for the two C-O bonds at 2’ and 3’ positions. This suggested that the adoption of this conformation by TNA might be partially due to the electrostatic repulsion between the two neighboring phosphate groups. It this was true, one would think an even simpler oligonucleotide system, such as an acyclic analog of TNA lacking the endocyclic oxygen atom as well as the 4’-methylene group, would still be able to

cross pair with the natural systems. This type of acyclic analog is glycerol nucleic acid (GNA).[31] Indeed, GNA can cross pair with RNA, suggesting GNA is also a possible progenitor candidate of RNA. However, it was not clear what the evolutionary relationship between GNA and TNA is. Were they consecutive polymers along the same evolutionary pathway to RNA, or were they just individual competing genetic systems of RNA? To address this question, Chaput and co-workers synthesized two complementary sequences in GNA, TNA and RNA backbone, respectively.[32] They found that although the three polymers all showed stable intrasystem base pairing and that both GNA and TNA could cross pair with complementary strands of RNA, GNA does not cross pair with TNA. This provided experimental evidence that GNA and TNA are not sequential polymers during the early emergence of RNA. But they are still potential ancestors of RNA because there might be an unknown adaptor polymer that could bridge these two systems. This result also suggests that there are more than one Watson-Crick "languages", as the example of pRNA has proven.

1.4.4. Towards Evolution of TNA Polymers. Since its initial discovery in 2000,[18] TNA has generated considerable interests as a potential RNA progenitor due to its unique base pairing properties with itself and RNA and the chemical simplicity of threose relative to ribose. In order to further evaluate the feasibility of TNA as a primitive genetic polymer, people have been attempting to evolve functional TNA molecules such as aptamers and enzymes by in vitro selection. Just like RNA polymerases and reverse transcriptases could support in vitro selection for RNA aptamers and ribozymes, enzymes that could recognize TNA as substrate are required for any in vitro TNA selection. Unfortunately, such "TNA polymerases" do not exist naturally, and therefore screening

of commercial DNA polymerases, and evolution and engineering of natural DNA polymerases to alter or broaden its substrate specificity have been used to identify polymerases capable of accepting TNA as substrate.

The initial study was focused on enzymes that catalyze DNA polymerization on TNA template. This involved challenging different commercial DNA polymerases in a primer extension assay to extend the DNA primer with dNTPs in a TNA template-directed manner.[33] Considering the structural similarity of TNA to RNA, several RNA-directed DNA polymerases were included together with other DNA-directed DNA polymerases in this study. It turned out that the most proficient enzymes were MMLV RT and SuperScript II RT. MMLV RT is a reverse transcriptase isolated from a retrovirus called Moloney murine leukemia virus, and SuperScript II RT is the commercialized version of MMLV RT, with three point mutations to reduce its RNase H activity. They both could extend a DNA primer with at least nine residues on a TNA template.

What is perhaps more challenging is to identify polymerases that could incorporate TNA triphosphates, because DNA polymerases have very stringent specificity for the incoming nucleotide. For example, most DNA polymerases could prevent NTP incorporation and preferred dNTP incorporation by 280-4400000 fold.[34] In an attempt to identify enzymes that could catalyze TNA incorporation, over ten DNA polymerases were screened in a primer extension assay for the ability to extend the DNA primer with TNA residues on a DNA template.[24] Only Bst DNA polymerase I and Deep Vent (exo-) DNA polymerase were able to catalyze up to three consecutive tTTP additions after 24-h incubation. Bst DNA polymerase is the full length polymerase from *Bacillus stearothermophilus*. Deep Vent (exo-) DNA polymerase is an engineered form

of native Deep Vent DNA polymerase from archaeon *Pyrococcus* species GB-D with two point mutations to eliminate its 3'->5' proofreading exonuclease activity. Generally, Deep Vent (exo-) DNA polymerase is more efficient than Bst DNA polymerase because the former could yield quantitative conversion of primer to full length product but the latter generated a major product by incorporation of only one TNA residue. Similarly, Herdewijn and co-workers tested Vent (exo-) DNA polymerase and HIV reverse transcriptase in a primer extension assay.[35] They found Vent (exo-) DNA polymerase were able incorporate two tTTP or tUTP nucleotides with minor products at three or four threosyl building blocks incorporation. However, HIV reverse transcriptase could not extend the primer by more than two successive TNA residues, which is consistent with previous report.

A major discovery came in 2005 when an enzyme called Therminator DNA polymerase was found to be able to catalyze the polymerization of 50 consecutive TNA nucleotides within 24 h.[36,37] Therminator DNA polymerase is an engineered version of 9°N DNA polymerase from the extremely thermophilic marine archaeon *Thermococcus* species 9°N-7 with three point mutations (D141A/E143A/A485L) to eliminate its exonuclease activity and to enhance its ability to incorporate modified nucleotides such as dideoxynucleotides, ribonucleotides and acyclonucleotides.[38] In contrast, other mutants of 9°N DNA polymerase including single mutant Y409V and double mutant Y409V/A485L appeared to be less efficient. But they could still incorporate more TNA residues than other archaeal family B DNA polymerases such as wildtype 9°N DNA polymerase, Deep Vent (exo-) DNA polymerase and Vent (exo-) DNA polymerase.

A detailed kinetic study of enzyme-mediated single nucleotide incorporation revealed dramatic difference between Deep Vent (exo-) DNA polymerase and Therminator DNA polymerase.[37] The steady-state kinetic parameters of Deep Vent (exo-)-mediated tNTP incorporation to a DNA primer were compared with those of dNTP incorporation. It was shown the unnatural incoming nucleotide reduced the enzyme efficiency by 20-100 fold, mostly due to increased K_M which might reflect the loss of binding interaction with tNTP substrate. Similar experiments were performed for Therminator DNA polymerase and revealed that tNTPs are only slightly less efficient (2-10-fold) compared with dNTP substrates, again primarily due to a modest increase in K_M. These results were in agreement with previous primer extension assays using Deep Vent (exo-) and Therminator DNA polymerase. In the presence of Mn^{2+}, which is known to relax substrate specificity of polymerase, the enzymatic efficiency of Therminator DNA polymerase was further improved to closely approximate that of dNTP incorporation in the absence of Mn^{2+}. In order to examine the effects of TNA residues in DNA primer on the continued polymerization of TNA triphosphates, similar kinetics experiments were done using a DNA/TNA chimeric primer ending with five TNA residues on the 3' terminus. It is intuitive to imagine that the polymerase processivity will drop when it moves from natural primer to unnatural primer due to possible loss of critical contacts. Indeed, the catalytic efficiency of Deep Vent (exo-) DNA polymerase incorporating the second and third tTTP dropped 360- and 2600-fold, respectively, compared with the incorporation of the first TNA residue.[24] However, the continued TNA synthesis mediated by Therminator DNA polymerase was not adversely affected by the presence of TNA region within the primer. Instead, a slight increase (2-6-fold) was observed in the

catalytic efficiency of Therminator-mediated TNA polymerization. This implied that Therminator DNA polymerase might be able to form direct and possibly stronger contact with the TNA primer, thus allowing processive TNA polymerization.

Identifying polymerases that recognize TNA either as template or monomer building blocks is one important step towards the goal of in vitro selection for functional TNA molecules. However, a successful in vitro selection requires the polymerization process accurate, because too many mutations will dilute the enriched pool with erroneous sequences and might lead to failure in in vitro selection. Consequently, several studies measured the fidelity of enzyme-mediated TNA-directed DNA polymerization process and DNA-dependent TNA synthesis. In an initial study of DNA synthesis on a 9-nt TNA template, the most efficient TNA-dependent DNA polymerase SuperScript II was found to have a fidelity value of 98.9%, which is significantly higher ($p < 0.001$) than that (99.6%) of SuperScript II-catalyzed DNA synthesis on DNA template.[33] This relatively low fidelity is likely due to the non-proofreading nature of the polymerase and the presence of mutagenic manganese ions. The fidelity of Therminator-catalyzed tNTP incorporation was also estimated in a dropout assay.[36] In this assay, two parallel primer extension reactions were performed with one containing all four tNTPs and the other containing only three types of tNTPs. In theory, if the fidelity was 100%, the three-tNTP reaction would not be able to generate any full-length product because the polymerase could not read through a reference nucleotide when the complementary tNTP was omitted. Consequently, any full-length product observed in this reaction was attributed from misincorporation at the reference position, and fidelity could be calculated as:

$$1-\frac{\text{percentage of full-length product in three-tNTP reaction}}{\text{percentage of full-length product in four-tNTP reaction}}$$

This assay estimated the fidelity of individual tNTP incorporation event to be >99% (tGTP), 98% (tCTP), 95% (tDTP), 94% (tTTP). However, these values were very likely to be underestimated, because in the three-tNTP reactions the correct building block was absent and thus it encouraged misincorporation to occur whereas a true fidelity should reflect the competition between the desired and mismatched nucleotides.

A more thorough fidelity study was carried out by Szostak and co-workers with different templates, reaction temperatures, incubation time, and tNTP concentrations.[25] In this study, the term fidelity was different from the conventional definition as the accuracy of single nucleotide incorporation, because TNA could not be directly sequenced by current sequencing technology. Instead, the fidelity determined in this study reflected the combined effects of Therminator-mediated TNA synthesis and SuperScript II-mediated TNA reverse transcription. To rule out the interference from leftover template contaminant, a marker nucleotide (thymidine) in the primer sequence was intentionally designed to be non-complementary to the template (thymidine). Consequently, after TNA synthesis and reverse transcription, the cDNA strand should contain an adenosine residue in this position while the original DNA template has a thymidine. By examining the nucleotide identity in this position, one could distinguish whether the sequencing result was amplified from cDNA or from leftover template contaminant.

Using two different template sequences, they found the fidelity was over 99% and did not show any sequence dependence. Additionally, adenine and diaminopurine were equally accurately recognized in DNA template, causing no difference in fidelity values,

although diaminopurine-containing templates supported dramatically more efficient TNA polymerization than adenine-containing templates. Two different reaction temperatures were tested and it was found the fidelity was essentially the same at 55 ℃ and 75 ℃, but at lower temperature it yielded much more full-length extension products of TNA synthesis. Although the yield of full-length TNA products always increased with time, the fidelity of TNA transcription dropped with prolonged incubation. One could imagine that some molecules would stop extending due to a misincorporation event, which could have destabilizing effect on the interaction between the primer terminus and template and cause the primer to lose critical contacts with the polymerase active site. This would lead to truncated extension product if the reaction was stopped at early incubation time. However, with longer reaction time such molecules would be eventually extended especially considering Therminator DNA polymerase lacks 3'->5' exonuclease activity. Indeed, the error rate increased drastically from 0.004 from 0.04 when the incubation time increased from 40 min to 3 h. Previously, highly biased tNTP concentration was used with 60-fold more tDTP than tGTP, because it was thought tGTP could form wobble base pair with thymine and compete with the correct incorporation of tDTP. In this study, more equal molar tNTP ratios were used and its effect on fidelity was found to be equivalent to that of biased tNTP mixture, suggesting that heavily biased tNTP concentrations were not necessary to maintain accurate TNA synthesis. Interestingly, in the 3-h TNA synthesis reaction, two most dominant mutations were D->T and G->C transversions, presumably caused by D-D and G-G mispairs, respectively, through Hoogsteen edge during TNA transcription. In a later study, the D-D mispair was avoided

by switching back to the natural nucleobase adenine and the G-G mispair was observed again as a major mutation event.[39]

1.5. Projects

In this dissertation, we report three research projects addressing fundamental questions in the field of synthetic genetics. In the first project, we ask a basic question regarding the specificity of nucleic acid aptamers: can a DNA aptamer can be made to discriminate alkaline proteins with high specificity? Some might think it is impossible because DNA has a negatively charged backbone and alkaline proteins typically have positive charges on the surface. Consequently, the electrostatic attraction between these two types of molecules makes selection for high-specificity aptamers difficult. In the second project, we access the fitness of threose nucleic acid as a primordial genetic polymer. As a potential RNA progenitor, TNA needs to be able to fold and perform functions like ligand binding and catalysis. We address key obstacles in TNA evolution and demonstrate TNA polymer is able to form stable tertiary structure with specific ligand binding pocket. This demonstration provides experimental evidence to support TNA as an RNA ancestor during early evolution of life on Earth. The third project aims to develop applicable methodology towards the goal of isolating a TNA enzyme with catalytic activity, as catalysis is a prerequisite for any primitive genetic material. This requires identifying enzymes that allow genetic information flow between DNA and TNA efficiently and faithfully. The establishment of such an in vitro replication system for TNA enables future in vitro selection for TNA with more sophisticated functions such as catalysis.

1.6. References

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

(2) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505-510.

(3) Ellington, A. D.; Szostak, J. W. *Nature* **1992**, *355*, 850-852.

(4) Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425-1429.

(5) Williams, B. A.; Lin, L.; Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2009**, *131*, 6330-6331.

(6) Robertson, D. L.; Joyce, G. F. *Nature* **1990**, *344*, 467-468.

(7) Breaker, R. R.; Joyce, G. F. *Chem. Biol.* **1994**, *1*, 223-229.

(8) Breaker, R. R.; Joyce, G. F. *Chem. Biol.* **1995**, *2*, 655-660.

(9) Santoro, S. W.; Joyce, G. F. *Proc. Natl. Acad. Sci.* **1997**, *94*, 4262-4266.

(10) Cruz, R. P.; Withers, J. B.; Li, Y. *Chem. Biol.* **2004**, *11*, 57-67.

(11) Schlosser, K.; Gu, J.; Sule, L.; Li, Y. *Nucleic Acids Res.* **2008**, *36*, 1472-1481.

(12) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611-647.

(13) Joyce, G. E. *Angew. Chem. Int. Ed.* **2007**, *46*, 6420-6436.

(14) Chaput, J. C.; Yu, H.; Zhang, S. *Chem. Biol.* **2012**, *19*, 1360-1371.

(15) Vaught, J. D.; Bock, C.; Carter, J.; Fitzwater, T.; Otis, M.; Schneider, D.; Rolando, J.; Waugh, S.; Wilcox, S. K.; Eaton, B. E. *J. Am. Chem. Soc.* **2010**, *132*, 4141-4151.

(16) Eschenmoser, A. *Science* **1999**, *284*, 2118-2124.

(17) Reck, F.; Wippo, H.; Kudick, R.; Bolli, M.; Ceulemans, G.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **1999**, *1*, 1531-1534.

(18) Schoning, K.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347-1351.

(19) Schoning, K. U.; Scholz, P.; Wu, X. L.; Guntha, S.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. *Helv. Chim. Acta* **2002**, *85*, 4111-4153.

(20) Orgel, L. *Science* **2000**, *290*, 1306-1307.

(21) Herdewijn, P. *Angew. Chem. Int. Ed.* **2001**, *40*, 2249-2251.

(22) Zou, K. Y.; Horhota, A.; Yu, B.; Szostak, J. W.; McLaughlin, L. W. *Org. Lett.* **2005**, *7*, 1485-1487.

(23) Wu, X. L.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **2002**, *4*, 1283-1286.

(24) Chaput, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 9274-9275.

(25) Ichida, J. K.; Horhota, A.; Zou, K. Y.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Res.* **2005**, *33*, 5219-5225.

(26) Wu, X.; Guntha, S.; Ferencic, M.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **2002**, *4*, 1279-1282.

(27) Heuberger, B. D.; Switzer, C. *Org. Lett.* **2006**, *8*, 5809-5811.

(28) Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. *J. Am. Chem. Soc.* **2002**, *124*, 13716-13721.

(29) Pallan, P. S.; Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. *Angew. Chem. Int. Ed.* **2003**, *42*, 5893-5895.

(30) Ebert, M. O.; Mang, C.; Krishnamurthy, R.; Eschenmoser, A.; Jaun, B. *J. Am. Chem. Soc.* **2008**, *130*, 15105-15115.

(31) Zhang, L. L.; Peritz, A.; Meggers, E. *J. Am. Chem. Soc.* **2005**, *127*, 4174-4175.

(32) Yang, Y.-W.; Zhang, S.; McCullum, E. O.; Chaput, J. C. *J. Mol. Evol.* **2007**, *65*, 289-295.

(33) Chaput, J. C.; Ichida, J. K.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 856-857.

(34) Brown, J. A.; Suo, Z. *Biochemistry* **2011**, *50*, 1135-1142.

(35) Kempeneers, V.; Vastmans, K.; Rozenski, J.; Herdewijn, P. *Nucleic Acids Res.* **2003**, *31*, 6221-6226.

(36) Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 2802-2803.

(37) Horhota, A.; Zou, K. Y.; Ichida, J. K.; Yu, B.; McLaughlin, L. W.; Szostak, J. W.; Chaput, J. C. *J. Am. Chem. Soc.* **2005**, *127*, 7427-7434.

(38) Gardner, A. F.; Jack, W. E. *Nucleic Acids Res.* **2002**, *30*, 605-613.

(39) Yu, H.; Zhang, S.; Chaput, J. C. *Nat. Chem.* **2012**, *4*, 183-187.

(40) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakked, Z.; Sklenar, H.; Suzuki, M.; Tung, C. S.; Westhof, E.; Wolberger, C.; Berman, H. M. *J. Mol. Biol.* **2001**, *313*, 229-237.

Chapter 2

Aptamers can Discriminate Alkaline Proteins with High Specificity

Reprinted with permission from Yu, H.; Jiang, B.; Chaput, J.C. Aptamers can discriminate alkaline proteins with high specificity, *ChemBioChem* **2011**, *12*, 2659-2666.

2.1. Abstract

Aptamers are single-stranded nucleic acids that fold into stable three-dimensional structures with ligand binding sites that are complementary in shape and charge to a desired target. Aptamers are generated by an iterative process known as in vitro selection, which permits their isolation from pools of random sequences. While aptamers have been selected to bind a wide range of targets, it is generally thought that these molecules are incapable of discriminating strongly alkaline proteins due to the attractive forces that govern oppositely charged polymers (e.g., polyelectrolyte effect). Histones, eukaryotic proteins that make up the core structure of nucleosomes are attractive targets for exploring the binding properties of aptamers because these proteins have positively charged surfaces that bind DNA through noncovalent sequence-independent interactions. Previous selections by our lab and others have yielded DNA aptamers with high affinity but low specificity to individual histone proteins. Whether this is a general limitation of aptamers is an interesting question with important practical implications in the future development of protein affinity reagents. Here we report the in vitro selection of a DNA aptamer that binds to histone H4 with a K_d of 13 nM and distinguishes other core histone proteins with 100 to 480-fold selectivity, which corresponds to a $\Delta\Delta G$ of up to 3.4 kcal

mol^{-1}. This result extends our fundamental understanding of aptamers and their ability to fold into shapes that selectively bind alkaline proteins.

2.2. Introduction

Aptamers are short nucleic acid polymers (DNA or RNA) that fold into well-defined three-dimensional structures whose surfaces include binding sites that are complementary in shape and charge to a desired target. Aptamers were first discovered in 1990 when two labs independently reported the generation of RNA molecules with specific ligand binding properties from pools of random sequences.[1,2] In the original papers, Ellington and Szostak called these RNA molecules "aptamers" from the Latin *aptus*, to fit, while Tuerk and Gold labeled this process "systematic evolution of ligands by exponential enrichment" (SELEX). SELEX is sometimes referred to as in vitro selection or test tube evolution since this laboratory procedure mimics the natural process of Darwinian evolution.[3-7] In these experiments researchers create a survival-of-the-fittest environment in which individual molecules compete against one another to overcome a selective pressure that is predefined, but often requires binding to a desired target. The small fraction of molecules that meet this requirement are collected and amplified to restore the population to its original size and create progeny molecules that can be further challenged in subsequent rounds of in vitro selection and amplification. Progeny molecules have the ability to inherit genetic mutations, either by intentional mutagenesis or through random mistakes made by a polymerase that can improve the fitness of the molecule for its intended function or lead to deleterious effects that cause the sequence to be removed from the pool.

The ability to harness the power of evolution at the molecular level has led to the development of straightforward procedures for creating tailor-made affinity reagents in the laboratory.[8] Since those initial experiments aptamers have been shown to display a wide range of structural plasticity, and it is now clear that aptamers can be selected to bind almost any kind of molecular target from small molecules to whole cells.[3,7,9] One major hallmark of aptamers is their ability to bind discrete targets with high specificity. An aptamer generated to bind theophylline, for example, recognizes its cognate ligand 10 000 times better than it recognizes caffeine, which differs from theophylline by only one methyl group.[10] More recently, our lab developed an aptamer that recognizes an acetyllysine posttranslational modification in a polypeptide sequence with 2400-fold specificity.[11]

The strong recognition properties of aptamers combined with the ease by which they can be produced has fueled strong interest in the use of aptamers as affinity reagents in many areas of biotechnology and molecular medicine.[12,13] Aptamers function efficiently in standard protein-binding assays, including ELISA,[14] Western blot analyses,[15] microarrays,[16] and affinity chromatography.[17] In one example, an L-selectin aptamer was used to purify the human L-selectin receptor from Chinese hamster ovary cells.[17] In this case, pure protein was obtained in a single step with 15 000-fold enrichment and 83% recovery. Aptamers have also been used as recognition elements in a variety of biosensors and analytical devices.[18] For example, an aptamer-based dipstick assay was developed to detect cocaine,[19] and a colorimetric assay now exists to monitor the levels of lead in the environment.[20] Aptamers are also gaining attention as therapeutic agents.[21] The aptamer Macugen is now approved by the FDA for the treatment of patients

affected by neovascular age-related macular degeneration.[22] This VEGF aptamer functions as a drug by inhibiting the binding of VEGF-165 to its receptor. In clinical trials, 80% of the patients treated with this aptamer showed stable or improved vision three months after treatment.[22]

Despite the success that aptamers have achieved in recent years, many basic questions remain about how these molecules fold into shapes with discrete ligand-binding functions.[23-25] The ability for aptamers to target alkaline proteins constitutes an important aspect of this general problem as many proteins have highly basic surfaces. Clearly a greater understanding of the binding properties of aptamers is needed if these molecules are to be used as affinity reagents on a scale as large as the human proteome.[26] Conventional wisdom suggests that aptamers should be incapable of folding into structures that selectively recognize positively charged proteins due to the attractive forces that govern polymers of opposite charge. This problem, commonly referred to as the polyelectrolyte effect, occurs when negatively charged polymers like DNA interact with positively charged polymers like protein to create a ligand binding interaction that releases water molecules and counter ions that previously solvated overlapping regions of both polymers.[27] The magnitude of the polyelectrolyte effect is an important constraint on the ability of aptamers to target alkaline proteins, as aptamers would need to first overcome the barrier that defines the complementary attraction of oppositely charged polymers in order to bind a basic protein with high specificity. While the polyelectrolyte effect has been the subject of previous computational studies,[28] very little experimental consideration has been given to the thermodynamic properties of aptamers and their ability to bind alkaline proteins.

We chose to explore this problem by attempting to evolve DNA aptamers with high specificity to histone H4. Histones are eukaryotic proteins that package DNA into nucleosomes. The core proteins that make up the nucleosome are histones H2A, H2B, H3, and H4.[29] We hypothesized that histone H4 represented an ideal target for this investigation as a previous study by our lab produced a histone-binding aptamer with high affinity but low specificity.[30] Similar results were also achieved by Gonzalez and co-workers in their generation of DNA aptamers to *Leishmania infantum* histone proteins H2A and H3.[31,32] Collectively, these examples led us to wonder whether this was a general problem of aptamer binding or a specific problem related to the previous selection strategies. To explore this question in greater detail, we carried out an in vitro selection using counter selection steps to determine whether aptamers could be generated that distinguished histone H4 from the three other core histone proteins. The best aptamer identified in this selection binds to histone H4 with low nanomolar affinity and discriminates against histone proteins H2A, H2B, and H3 by ∼100-500-fold. By comparison, all previous selections yielded aptamers with only two- to five-fold specificity. This result demonstrates that aptamers have the ability to fold into structures that distinguish highly basic proteins of similar structure and function; however, these sequences are sufficiently rare that they require strong counter selection conditions to isolate.

2.3. Results and Discussion

2.3.1. Selection for single-stranded DNAs that bind to histone H4. DNA sequences that bind to histone H4 were isolated by iterative rounds of in vitro selection and amplification. The initial pool contained 10^{12} unique single-stranded DNA molecules

with a central random region of 50 unbiased nucleotide positions flanked on both sides with distinct primer-binding sites. To isolate molecules with affinity to the N-terminal region of histone H4, peptides reflecting the N-terminal tail of histone proteins H2A, H2B, H3, and H4 were used in place of the whole proteins. This substitution was feasible because this region of the protein remains natively unstructured when DNA threads itself around the histone octamer to form the nucleosome core.[29] The selection strategy (Figure 2.1) included a negative selection step to remove molecules that bound the off-target histone sequences of H2A, H2B, and H3, followed by a positive selection step to isolate molecules with affinity to the desired histone H4 target sequence. For each round of selection, the pool was incubated with the off-target peptides H2A, H2B, and H3, which were modified with a C-terminal biotin residue to enable their capture on a streptavidin-affinity matrix. Molecules that remained in the pool were incubated with the desired H4 peptide, and functional aptamers were separated from the unbound pool by injecting the mixture onto a neutral coated capillary. Five injections were made for each round of selection and 10^{11} molecules were sampled in the first round of the selection.

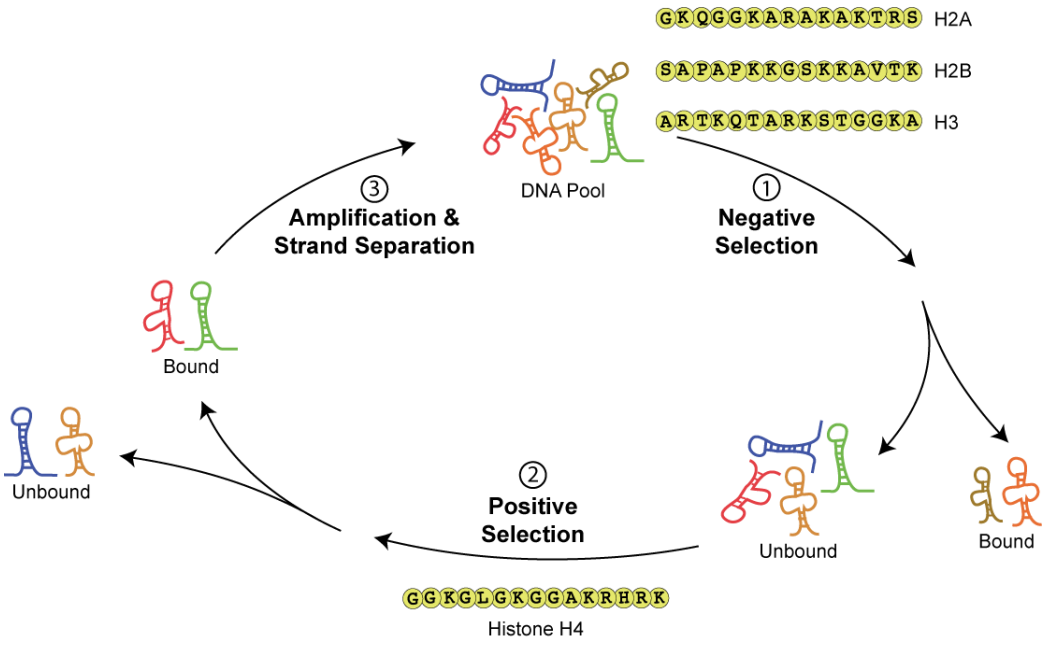


Figure 2.1. In vitro selection strategy used to generate DNA aptamers with high affinity and specificity to histone H4 protein. For each round of selection and amplification, the pool of DNA was pre-cleared by removing DNA sequences with affinity to the N-terminal tails of histone proteins H2 A, H2 B and H3. Molecules collected in the flow-through were incubated with the N-terminal tail of histone H4 and the bound fraction was separated from the unbound pool by capillary electrophoresis. The set of bound molecules was recovered, amplified and used to generate a new pool of sequences for input into the next round of selection.

Capillary electrophoresis (CE) was chosen for the positive selection step because this technique leads to a higher partitioning efficiency than is commonly observed for traditional gravity filtration.[33] This in turn reduces the number of selection cycles required to generate high quality aptamers from ≥10 to just three or four rounds of selection and amplification. In the case of IgE, for example, an aptamer was generated after four rounds of CE-based selection that exhibited similar binding properties to an

aptamer produced after 15 rounds of traditional selection.[34] A second major advantage of CE is that aptamer binding occurs free in solution, which obviates the need for complicated conjugation chemistry that can occlude surface binding sites or alter the native protein structure. Collectively, these advantages make CE an increasingly popular separation technique for the in vitro selection of aptamers that bind peptides and proteins.[35-38] We have previously used this method to generate a DNA aptamer with >2000-fold specificity to an acetylated lysine residue in a short polypeptide sequence.[11]

To favor the selection of aptamers with high specificity to histone H4, the ratio of the DNA pool to the different histone tails was adjusted in the negative and positive selection steps to maintain high selective pressure on the pool of evolving molecules. In rounds one and two, the ratio of the DNA pool to the off-target histones was 100:1, which was stringent enough to remove DNA sequences that bound the off-target peptides, but permissive enough to allow desirable molecules to remain in the pool. The stringency was then increased in rounds three and four by reducing the ratio to 1:1, which favored the removal of molecules with weaker affinity to the H2A, H2B, and H3 peptide sequences. For each round of positive selection, the ratio was reversed such that the H4 peptide was present at limiting amounts relative to the DNA pool (1:1000). By limiting the target peptide, we aimed to increase competition between the pool and desired histone tail. After four rounds of selection, the DNA pool was cloned and sequenced. We obtained 23 clones and analyzed their sequences by calculating their predicted secondary structures using the computer program mFold[39] (See Figure S1 in APPENDIX A). Five of the clones are predicted to fold into structures that are dominated by a simple stem-

loop or internal bulge motif. The remaining clones adopt more complicated structures that contain tandem stem-loop motifs. The presence of many highly structured sequences suggests that sophisticated functions, such as the ability to discriminate subtle differences between peptides of similar sequence and composition, require molecules with significant structural complexity.

2.3.2. Affinity and specificity of the DNA aptamers. Of the 23 sequences, eight representative clones with different secondary structures were chosen for further analysis. The eight sequences were constructed by solid-phase DNA synthesis, purified by gel electrophoresis, and assayed for affinity to histone H4 whole protein by dot blot analysis.[40] Close inspection of the dissociation constants (K_d values) reveal that all eight clones bind to histone H4 with K_d values of 1 to 10 nM, indicating these sequences are all capable of high affinity binding (Table 2.1). To examine the specificity of the selected aptamers, dissociation constants were measured for the four strongest binders to the off-target whole proteins H2A, H2B, and H3. In keeping with the literature,[41] this study defined specificity as the ratio of the off-target K_d to the on-target K_d, and aimed to produce aptamers with at least 100-fold specificity to each of the off-target proteins. Results from our initial specificity study demonstrate that the selected clones are relatively specific against histone proteins H2A and H2B (50- to 150-fold), but fail to discriminate histone H3 by more than tenfold (Table 2.2). Creating aptamers that distinguish the N-terminal tail of histone H4 from the N-terminal tail of histone H3 is a challenging problem as previous selections performed in the absence of counter selection methods yielded aptamers with only two- to fivefold selectivity.[30-32]

Table 2.1. Sequence and dissociation constants of round 4 aptamers to histone H4 protein.[a]

Clone	Sequence	K_d (nM)
4.33	TGGTGGGGTTCCCGGGAGGGCGGCTACGGGT TCCGTAATCAGATTTGTGT	1.3 ± 0.3
4.60	TTGGCCCCGCGTGATCATTGAGGGGAGGAGC CGAGGCGGGTCCAAGATTG	2.6 ± 0.4
4.32	ATCACGATATGCCCGGCTCATCGGGGTTCAG TTGGGCGGTCACATGGAAA	2.8 ± 0.6
4.36	TTCAAGGCGGCGAGATTTAGTGGTTGGGAGG CTGTACGCCCTACGTGAAC	2.9 ± 1.0
4.58	TTTAACGTAACCTGCAAGGCGGGGAGGTGCG AGCCCCGTGTGTGGCTTGC	3.8 ± 1.6
4.51	GGTGCTCAGGGAACTGTCTGAGGGATCAGGC TTAAGCCTGTCGAGCAGTT	5.6 ± 0.7
4.65	TCGCATGGAGGGCAGAGCCGCCTGCCGGGAT CCGGCCCTCTTGGGCGGGC	7.4 ± 0.5
4.57	TAGTCCCAAGGCACATAAGGGCCGGAGGTCT AGCGTCAGGGATACAGAGA	8.8 ± 1.9

[a] binding condition: 100 mM NaCl, 5 mM $MgCl_2$, 10 mM HEPES (pH 7.5).

2.3.3. Salt effects. Because electrostatic attraction between the negatively charged DNA backbone and positively charged histone protein might account for the low selectivity observed for the selected aptamers, we decided to examine the role of metal ions on ligand binding affinity. By increasing the concentrations of monovalent and divalent metal ions in the binding buffer, we aimed to stabilize the tertiary structure of the aptamer fold and simultaneously satisfy competing charges on the protein surface. To test this possibility, we chose clone 4.33 for further analysis as this sequence showed the highest degree of specificity to histone H3. Raising the salt concentration from 100 mM NaCl and 5 mM $MgCl_2$ to 500 mM NaCl and 10 mM $MgCl_2$ increased the binding specificity of clone 4.33 for histone H4 versus histone H3 from tenfold to nearly 30-fold (Table 2.3). A similar increase in specificity was observed against histones H2A and H2B (up to 422- and 86-fold, respectively). We noticed that increasing the salt concentration beyond this level did not translate into further increases in specificity, indicating that all of the metal binding sites on the aptamer and protein were saturated under the higher salt conditions (data not shown). We speculate that the change in specificity is due to the formation of new intramolecular contacts within the aptamer structure. This hypothesis is consistent with the observation that clone 4.33 adopts a third stem-loop motif when its predicted secondary structure is calculated under conditions that simulate the higher salt concentration (See Figure S2 in APPENDIX A).

Table 2.2. Affinity and specificity of representative aptamers.[a]

	H4		H3		H2A		H2B	
Clone	K_d (nM)	Sp.	K_d (nM)	Sp.	K_d (nM)	Sp.	K_d (nM)	Sp.
4.33	1.3 ± 0.3	--	14 ± 4	11	204 ± 22	157	84 ± 5	65
4.60	2.6 ± 0.4	--	17 ± 5	7	332 ± 137	128	270 ± 117	104
4.32	2.8 ± 0.6	--	18 ± 5	6	287 ± 44	103	153 ± 32	55
4.36	2.9 ± 1.0	--	21 ± 3	7	249 ± 46	86	213 ± 70	73

[a] Specificity (Sp.) is defined as the ratio of K_d (off-target) to K_d (on-target).

2.3.4. Directed evolution. In an effort to isolate aptamers with greater specificity for histone H4, we used directed evolution to optimize clone 4.33 for improved ligand binding affinity and specificity. We created a second-generation library based on the parent sequence of clone 4.33 in which each nucleotide position in the aptamer sequence was doped with a 15% mixture of the other three nucleotides. This level of mutagenesis was intended to optimize contacts within the aptamer structure and produce mutations that would lead to greater discrimination between histone H4 and the other three histone proteins. As a precaution new primer binding sites were added to the flanking regions to avoid the unwanted enrichment of aptamers from the original library. The doped library was subjected to three iterative rounds of directed evolution using two different selection strategies. The first strategy was performed in a manner identical to the original in vitro selection with a negative selection step performed on streptavidin-coated beads to remove molecules with affinity to the off-target sequence followed by a CE-based positive selection step to recover molecules that bound the N-terminal tail of histone H4. For each selection round, the ratio of the off-target to pool and on-target to pool was maintained at 1:1 and 1:1000, respectively. In the second selection strategy, both the negative and

positive selection steps were performed using traditional affinity chromatography methods to separate the bound molecules from the unbound pool. After three rounds of directed evolution, both libraries were cloned and sequenced to examine the diversity of molecules that remained in each pool.

Table 2.3. Affinity and specificity of aptamer 4.33 under low and high salt conditions.

	H4 (pI = 11)[a]		H3 (pI = 11)[a]		H2A (pI = 11)[a]		H2B (pI = 10)[a]	
Condition	K_d (nM)	Sp.	K_d (nM)	Sp.	K_d (nM)	Sp.	K_d (nM)	Sp.
Low salt[b]	1.3 ± 0.3	--	14 ± 4	11	204 ± 22	157	84 ± 5	65
High salt[c]	9.6 ± 6.2	--	273 ± 78	28	4050 ± 70	422	825 ± 3	86

[a] pI: isoelectric point. [b] low salt condition: 100 mM NaCl, 5 mM $MgCl_2$, 10 mM HEPES (pH 7.5). [c] high salt condition: 500 mM NaCl, 10 mM $MgCl_2$, 10 mM HEPES (pH 7.5).

Eight clones from the CE-based selection and nine clones from bead-based selection were aligned with clone 4.33 (See Figure S3 in APPENDIX A). The average number of mutations per sequence was 7.6, which closely approximates the number of mutations expected for a library of 50 nucleotides that was doped at a level of 15% per nucleotide position. Close inspection of the aligned sequences reveals several small patches of conserved nucleotides that are distributed among numerous single-point mutations. To examine the extent to which any of the selected sequences showed higher selectivity for histone H4, we randomly chose three sequences from the output of each selection and measured their affinity and specificity for histone H4. Each sequence was synthesized by solid-phase DNA synthesis, purified by gel electrophoresis, and assayed for affinity to histone proteins H2A, H2B, H3, and H4. Surprisingly, only clone 3.13 isolated from the CE-based selection showed high selectivity to histone H4 (Table 2.4). The remaining clones were unable to distinguish histone H4 from histone H3 by more

than ~20-fold, which is less than the parent sequence (clone 4.33). Aptamer CE-3.13, however, binds histone H4 with a K_d of 13 nM and discriminates against histone proteins H2A, H2B, and H3 by 477-, 165-, and 100-fold, respectively (Figure 2.2). This result emphasizes the challenge of isolating aptamers with reasonable selectivity to highly basic proteins, but provides evidence that such sequences are not so rare that they cannot be discovered by in vitro selection.

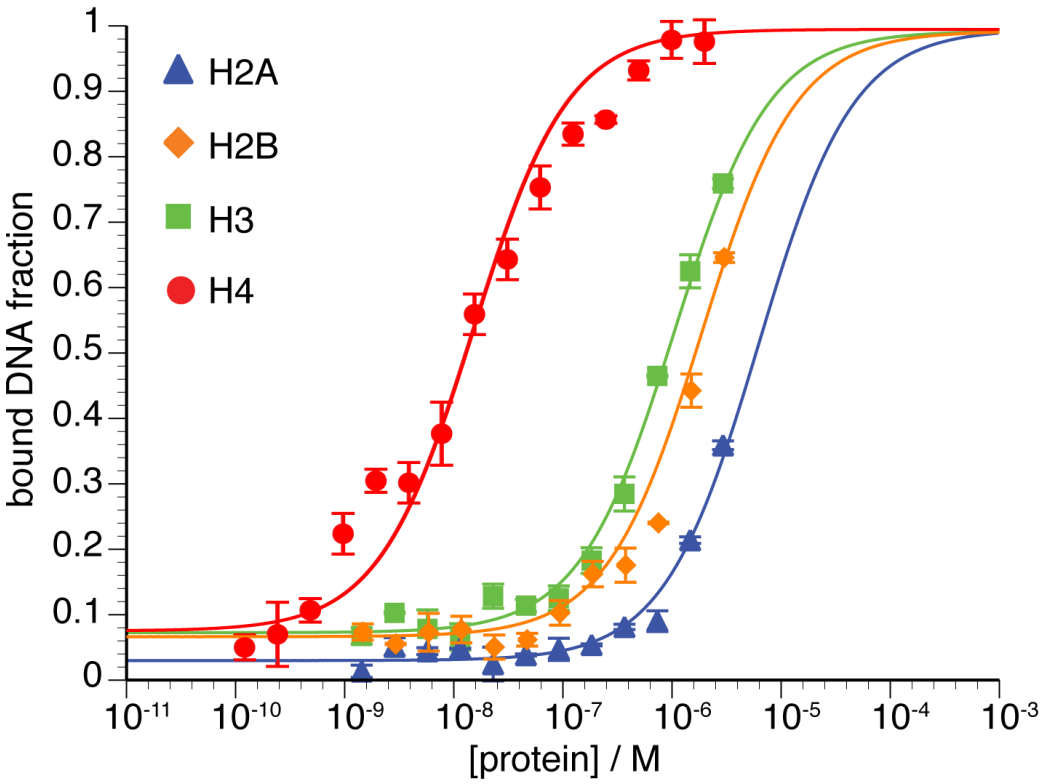


Figure 2.2. Solution binding affinity and specificity of aptamer CE-3.13. Equilibrium binding affinity was measured by dot blot analysis for histone proteins H4 (red circle), H3 (green square), H2 B (orange diamond) and H2 A (blue triangle).

2.3.5. Studies of site-directed mutations. To examine the genetic changes that led to improved specificity, we compared the predicted secondary structure for the parent sequence to the evolutionary optimized variant. This analysis demonstrates that four of

the eight single-point mutations occur in regions of the sequence that define the predicted secondary structure (Figure 2.3A). We therefore reverted each of the four point mutations individually back to their original nucleotide, and measured the solution binding affinity for histone proteins H3 and H4. While the four revertant clones recognized histone H4 with K_d values that are within twofold of aptamer CE-3.13, none of the sequences was able to distinguish histone H3 by more than 20-fold (Figure 2.3B). This loss in selectivity suggests that each of the four point mutations play an important role in the folding and recognition properties of aptamer CE-3.13.

Table 2.4. Affinity and specificity of representative aptamers isolated by directed evolution.[a]

	H4		H3		H2A		H2B	
Clone	K_d (nM)	Sp.	K_d (nM)	Sp.	K_d (nM)	Sp.	K_d (nM)	Sp.
Parental sequence								
4.33	9.6 ± 6.2	--	273 ± 78	28	4050 ± 70	422	825 ± 3	86
CE-based directed evolution								
3.1	55 ± 5	--	318 ± 16	6	5500 ± 800	100	5400 ± 2100	98
3.5	71 ± 8	--	490 ± 95	7	10500 ± 700	148	5500 ± 800	77
3.13	13 ± 1	--	1300 ± 50	100	6200 ± 100	477	2150 ± 70	165
Bead-based directed evolution								
3.12	22 ± 1	--	415 ± 12	19	5950 ± 70	270	3900 ± 400	177
3.13	11 ± 2	--	264 ± 7	23	4800 ± 300	425	3800 ± 500	332
3.16	20 ± 2	--	121 ± 2	11	3750 ± 70	332	1800 ± 40	159

[a] binding condition: 500 mM NaCl, 10 mM $MgCl_2$, 10 mM HEPES (pH 7.5).

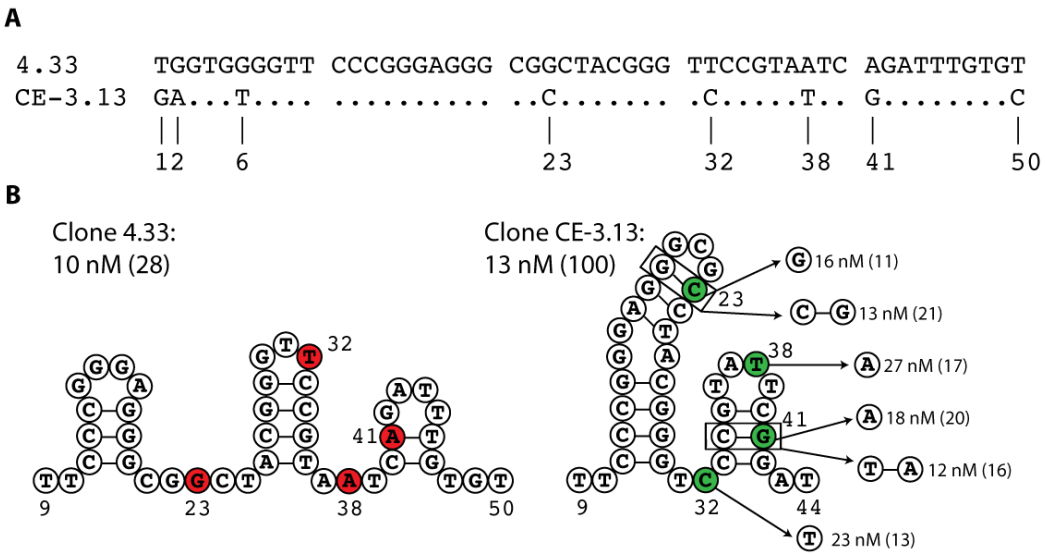


Figure 2.3. Aptamer binding motif and analysis. A) Sequence alignment of aptamers 4.33 and CE-3.13 showing the eight genetic mutations that gave rise to an evolutionary optimized variant with high specificity to histone H4. B) The predicted secondary structures of aptamers 4.33 and CE-3.13 are shown with the mutated positions highlighted in red and green, respectively. Binding affinity and specificity (parentheses) of each reversion and compensatory mutation is shown next to the structure of aptamer CE-3.13.

To further explore the evolved mutations, we generated variants that contained compensatory mutations in stem-loop regions of the predicted secondary structure. Two clones were constructed that restore Watson–Crick base pairs to the C23G and G41A revertants by changing the G:G and C:A mismatches to C:G and T:A base pairs, respectively. These engineered clones bind histone H4 with K_d values equivalent to the evolutionary optimized aptamer, but again fail to restore selectivity to the aptamer sequence (Figure 2.3B). This result demonstrates that the selected mutations, G23C and A41G, which form G:C and C:G base pairs in aptamer CE-3.13, impart additional

functionality beyond simply maintaining a contiguous helix in the stem-loop region of the secondary structure. One possibility is that these mutations form new contacts within the architecture of the aptamer that rigidify its structure and limit the amount of flexibility in and around the ligand-binding pocket.

To gain a better understanding of how the structure of aptamer CE-3.13 contributes to its recognition of the N-terminal tail of histone H4, we performed a hydroxyl radical footprinting analysis in the absence and presence of histone proteins H2A, H2B, H3, and H4. Consistent with the high specificity of aptamer CE-3.13 for histone protein H4, protection of the DNA backbone occurred to a greater extent for histone H4 than for any of the other three histone proteins (See Figure S4 in APPENDIX A). Careful analysis of the resulting gel indicates that histone proteins H2A, H2B, and H3 protect residues 34–38 of the aptamer. This relatively small region is likely due to weak electrostatic interactions between the DNA backbone and the three histone proteins. However, the case is quite different for histone H4, which protects a much larger region of the aptamer from hydroxyl cleavage. Here a clear footprint is observed for residues 31–45, which constitute a strong binding interface with histone H4. Combining this information with the predicted secondary structure suggests that the second stem-loop motif forms a binding pocket that is complementary in shape and charge to the N-terminal tail of histone H4. Indeed, three of the four genetic mutations observed in aptamer CE-3.13 occur in this region of the oligonucleotide, which implies that each of these mutations play an important role in the binding of histone H4. The fourth mutation, which occurs in the first stem-loop motif could help with aptamer stability by improving the packing interactions between the two stem-loop motifs. This result is consistent with

the interpretation that the four selected mutations increase protein binding specificity by rigidifying the aptamer structure.

2.3.6. Implications of aptamer selection. We have applied the strategy of in vitro selection and directed evolution to isolate a single-stranded DNA molecule with high affinity and specificity to histone H4. When we began this study it was not obvious a priori that a random pool of DNA sequences would contain a nucleic acid molecule that folded itself into a shape that recognized an alkaline protein with high affinity and specificity. A previous study by our lab that aimed to produce a set of DNA aptamers to histone H4 yielded a number of high affinity sequences (K_d ~5–10 nM); however, the best sequence could only discriminate histone H3 by a factor of five.[30] Similar results were obtained by Ramos and co-workers on histone proteins H2A and H3, which produced aptamers with only two- to threefold specificity.[31,32] Whether this was a general problem of aptamers (that is, the potential inability of negatively charged polymers to fold into shapes that recognize positively charged polymers with high selectivity) or simply a limitation of the previous selection strategy was unclear. We therefore designed a new selection strategy that included the use of stringent counter selection steps between iterative rounds of selection and amplification, since this approach has been widely used to generate aptamers with specific ligand binding properties.[10,11,42] The goal of this selection was to remove DNA molecules from the pool that exhibited high affinity to the off-target histone proteins H2A, H2B, and H3. In doing so, we aimed to address the broader question of whether aptamers could be used to bind alkaline proteins with high specificity.

Comparison of the binding properties of the aptamers isolated by directed evolution (Table 2.4) reveals a striking difference in the tolerance of each molecule for the off-target histone proteins. For example, aptamer CE-3.13, which was isolated using the capillary electrophoresis method, is significantly more fit in terms of its ability to bind histone H4 than all of the other DNA aptamers. This aptamer binds to histone H4 with a solution binding affinity of 13 nM and distinguishes the three remaining core histone proteins by a factor of 100- to 477-fold, which corresponds to a binding energy of up to 3.4 kcal mol^{-1}. In contrast, the less fit aptamers also bind to histone H4 with low nanomolar affinity, and are able to distinguish histone proteins H2A and H2B with high specificity (≥100-fold), but struggle with their ability to discriminate histone proteins H4 and H3. This problem was observed in our previous selection and likely stems from the fact that both proteins have similar sequence composition in their N-terminal tails, which was the protein region targeted in both selections. While it is exciting to wonder whether the isolation of aptamer CE-3.13 was due to the high partitioning efficiency of the capillary electrophoresis-based separation, many additional aptamers will need to be tested before this question can be answered.

One interesting observation to come from the aggregate set of binding data is that aptamers with high affinity are not automatically more specific for their target ligands. Although it has long been assumed that the easiest way to improve aptamer specificity is to increase its shape and charge complementary for a given target,[43] a recent study by Szostak and co-workers suggests that specificity is a physical property that emerges when biopolymers adopt folded structures that are reinforced with additional intramolecular contacts.[44] This revised aptamer binding theory takes into account the free energy term

provided by intramolecular contacts that contribute to the overall stability of the tertiary structure. According to this model, it is expected that as an aptamer evolves from an initial simple motif to a more complex tertiary structure it will acquire additional structural elements that allow it to form a more rigid ligand-binding pocket that is less able to accept target analogues. This hypothesis is consistent with the binding properties of our aptamers and suggests that aptamer CE-3.13 adopts a folded structure, either as a free molecule or in the bound state that is more rigid than the other aptamers that we tested.

Mutagenesis data supports the prediction that aptamer CE-3.13 represents a complex solution to the chemical problem of how a DNA molecule would fold itself into a tertiary structure with a ligand binding site that is capable of selectively recognizing the alkaline protein histone H4. Single-nucleotide revertants constructed for each of the four genetic mutations that occur in the region of the sequence that defines the predicted secondary structure maintain high affinity binding but abate specificity. Furthermore, specificity is not restored when the C23G and G41A revertants are modified with compensatory mutations that change the G_{19}:G_{23} and C_{34}:A_{41} mismatches to C_{19}:$G_{23}$and T_{34}:A_{41} base pairs, respectively. Since positions 23 and 41 occur in adjoining helices of the predicted secondary structure, successful restoration of specificity would have meant that these mutations were selected to maintain two contiguous helices in the aptamer structure. However, since neither compensatory mutation allowed the aptamer to recover specificity, it is reasonable to assume that both mutations play a greater role in aptamer folding. This prediction is supported by our footprinting analysis (See Figure S4 in APPENDIX A).

A second interesting observation to emerge from our results was that a limited sampling of aptamers (in this case six aptamers were examined after directed evolution) yielded a DNA molecule that was capable of achieving high specificity. One interpretation that is consistent with our results is that the counter selection method used to isolate these aptamers provided access to complex structures that are capable of folding into rigid shapes with well-defined ligand binding sites, but that these structures are still somewhat rare when compared to simpler structures that continue to dominate the pool. This scenario agrees with the long held belief that in vitro selection tends to produce the simplest solutions to a given biochemical problem. This hypothesis is evident from our previous selection for histone-binding aptamers, which selected for protein-binding affinity only and produced molecules with high affinity but low specificity. In contrast, the current selection strategy, which included a direct selection for specificity allowed us to favor the enrichment of aptamers with specific ligand binding properties by removing many of the simpler solutions from the pool. We speculate that our previous selection contained aptamers that were capable of high specificity but these molecules were so rare that random sampling of the selection output could not identify them.

2.4. Conclusion

In summary, we provide evidence that nucleic acid aptamers can be evolved by in vitro selection to fold into shapes that recognize alkaline proteins with high specificity. Because these aptamers are rare relative to simpler molecules that bind with high affinity but low specificity, their isolation requires strong counter selection measures that deplete the pool of low specificity binders. In the broader context of aptamer binding, these results suggest that aptamers could be used as affinity reagents to target a wide range of

human proteins, including structures whose surfaces are dominated by an abundance of positive charge.

2.5. Experimental Design

2.5.1. General. DNA oligonucleotides were purchased from Integrated DNA Technologies and purified by denaturing polyacrylamide gel electrophoresis. Histone peptides (H4, GGKGLGKGGAKRHRK; H3, ARTKQTARKSTGGKA; H2A, GKQGGKARAKAKTRS; H2B, SAPAPKKGSKKAVTK) were purchased from Sigma–Aldrich in >95% purity. Histone peptides with a C-terminal biotin residue (H4, GGKGLGKGGAKRHRK-Biotin; H3, ARTKQTARKSTGGKAGK-Biotin; H2A, GKQGGKARAKAKTRSGK-Biotin; H2B, SAPAPKKGSKKAVTK-Biotin) were purchased from New England Peptide in >95% purity. Histone proteins H2A, H2B, H3, and H4 were purchased from New England BioLabs. The 100-mer DNA library containing a random region of 50 nucleotides flanked on both sides with constant PCR primer binding sites, and a second generation DNA library based on clone 4.33 were purchased from the Keck Facility at Yale University.

2.5.2. In vitro selection. For each round of selection, the DNA library was amplified by PCR using a 6-carboxyfluorescein (6-FAM)-labeled forward primer (5′-FAM-GAG CTA CGT ACG AGG ATC CGG TGA G-3′) and a biotin-labeled reverse primer (5′-Biotin-GGA CCT GGG GCC GAA GCT TAG CAG T-3′). The pool was made single-stranded by immobilizing the dsDNA onto streptavidin-coated agarose beads and eluting the top strand with NaOH (0.15 M). The single-stranded library was neutralized, ethanol precipitated, and folded by heating to 95 ℃ for 5 min and cooling on ice for 10 min in selection buffer (NaCl (100 mM), $MgCl_2$ (5 mM), HEPES (10 mM), pH

7.5). The DNA library was incubated for 1 h at 24 ℃ with histone peptides H2A, H2B and H3 derivatized with a C-terminal biotin residue. After 1 h, the solution was passed through a column of streptavidin-coated agarose beads. The unbound fraction was collected, concentrated by ethanol precipitation, and refolded. The DNA pool was incubated with the histone H4 peptide for 1 h at 24 ℃, and histone H4 aptamers were isolated by separating the bound molecules from the unbound library by capillary electrophoresis. After four rounds of in vitro selection and amplification, the library was cloned and sequenced to examine the diversity of molecules that remained in the pool.

2.5.3. Capillary electrophoresis. Capillary electrophoresis was performed on Beckman ProteomeLab PA 800 Protein Characterization System. Prior to use, the glass capillary (0.1 mm inner diameter, total length=60 cm) was rinsed with water and equilibrated with selection buffer. A small portion (70 nL) of the library/peptide mixture was injected onto the capillary using pressure injection (0.5 psi for 5 s) and electrophoresis was performed under a constant voltage of 15 kV at 20 ℃ for 35 min. Laser-induced fluorescence (LIF) was used to monitor the separation of 6-FAM-labeled DNA (λ_{ex}=488 nm; λ_{em}=520 nm). Five injections were performed for each round of in vitro selection.

2.5.4. Directed evolution. Directed evolution was performed to optimize clone 4.33 (5′-CAC GAC TCT CAC CTC ATA GC tgg tgg ggt tcc cgg gag ggc ggc tac ggg ttc cgt aat cag att tgt gt CTG GTT CTG TAG ACG GCT TG-3′). A degenerate DNA library was constructed by solid-phase DNA synthesis using mixtures of phosphoramidite monomers that allowed for 15% mutagenesis to occur at each nucleotide position in the aptamer sequence. Lower case bases in clone 4.33 denote a region of the sequence that

contains 85% of the wild-type nucleotide and 5% of each of the other three bases. New PCR primers were used to avoid possible contamination with the first-generation library. The DNA library was amplified by PCR and made single-stranded by denaturing on streptavidin-coated agarose beads. The pool of single-stranded DNA was split into two parts, and two separate selections were carried out in parallel. The first selection was performed as described above with the exception that only histone H3 peptide was used in the negative selection step. All other steps were the same, including the solution-phase separation of the bound aptamers by capillary electrophoresis. The second selection was perform in a similar manner with the exception that the positive selection step was performed by capturing the portion of DNA that remained bound to histone H4 peptide on a streptavidin-coated magnetic beads, washing to remove the unbound molecules, and amplifying the bound material by PCR. After three rounds of in vitro selection and amplification, both libraries were cloned and sequenced to examine the diversity of molecules that remained in the pool.

2.5.5. DNA sequencing and analysis. DNA sequences present in the output of each selection were amplified by PCR and cloned into a pJET DNA cloning vector (Fermentas). The vectors were transformed into *E. coli* TOP10 competent cells and grown on ampicillin agar plates at 37 ℃ with an overnight incubation. Individual colonies were randomly picked and checked by colony PCR to ensure that the vector contained the insert. Positive clones were grown in liquid media, mini-prepped, and sequenced at the ASU Sequencing Facility. The predicted secondary structures were determined using the computer program mFold.[39]

2.5.6. Dot blot binding assay. DNA aptamers (150 pmol) were labeled with ^{32}P by incubating with [γ-^{32}P]ATP and T4 polynucleotide kinase for 1 h at 37 ℃. The [^{32}P]-labeled aptamers were desalted on a sephadex G-25 column, diluted with selection buffer and folded by heating at 95 ℃ for 5 min and cooling on ice for 10 min. The purified aptamers were then divided into 12 tubes and incubated for 1 h at 24 ℃ with the histone protein poised at concentrations that span the expected K_d (typically 0.1–100 nM). After 1 h, the solutions were placed into a vacuum minifold dot blot apparatus and the bound aptamers were partitioned away from the free DNA by passing the solution through a layer of nitrocellulose and nylon membranes.[40] To reduce any nonspecific binding and retention of the free DNA, the nitrocellulose membrane was presoaked for 10 min in KOH (0.4 M) and rinsed with water until the pH returned to neutral. Prior to analysis, both membranes were equilibrated in selection buffer for 30 min at 4 ℃ followed by passing selection buffer through the wells with vacuum. Samples were then loaded into the dot blot minifold and vacuum was applied to separate the bound aptamer from the unbound DNA. Aptamers that are bound to histone become captured on the surface of the nitrocellulose membrane (top layer), while unbound DNA passes through the nitrocellulose layer and becomes captured on the nylon membrane (bottom layer). The wells were then washed with selection buffer, dried, and the amount of radioactivity present on both membranes was determined by phosphorimaging. The protein-bound aptamer fraction and protein concentration were used to determine the K_d using the following equation:

$$\frac{I_b}{I_b + I_u} = c_1 + c_2 \times \frac{[protein]}{[protein] + K_d}$$

Here I_b and I_u are the intensity of protein-bound aptamer and free aptamer, respectively, c_1 and c_2 are constants. Dissociation constants were calculated using a nonlinear least-squares regression analysis performed with the DeltaGraph program.

2.5.7. Structural probing by hydroxyl radical footprinting. Hydroxyl radical footprinting reactions were performed similar to previously described.[45] Briefly,[^{32}P]-5′-end labeled aptamer was incubated in high salt conditions in the presence or absence of histone proteins and equilibrated for 1 h at room temperature (total volume 10 μL). The hydroxyl radical cleavage reaction was prepared by carefully spotting a fresh Fe^{II}-EDTA solution ($(NH_4)_2Fe(SO_4)_2$ (3 mM), EDTA (6 mM), 1 μL), sodium ascorbate (30 mM, 1 μL), and 1.8% hydrogen peroxide (1 μL, freshly diluted from a 30% stock) as three separate drops on the wall of the tube. The reaction was initiated by simultaneously mixing the three individual reagent drops together and immediately adding this mixture to the aptamer solution. The reaction was quenched by addition of stop solution (100 mM thiourea, 7 μL) after 1 min of digestion.

2.6. References

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

(2) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505-510.

(3) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611-647.

(4) Gold, L.; Polisky, B.; Uhlenbeck, O.; Yarus, M. *Annu. Rev. Biochem.* **1995**, *64*, 763-797.

(5) Bittker, J. A.; Phillips, K. J.; Liu, D. R. *Curr. Opin. Chem. Biol.* **2002**, *6*, 367-374.

(6) Famulok, M.; Mayer, G.; Blind, M. *Acc. Chem. Res.* **2000**, *33*, 591-599.

(7) Mayer, G. *Angew. Chem. Int. Ed.* **2009**, *48*, 2672-2689.

(8) Bunka, D. H.; Stockley, P. G. *Nat. Rev. Microbiol.* **2006**, *4*, 588-596.

(9) Jayasena, S. D. *Clin. Chem.* **1999**, *45*, 1628-1650.

(10) Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425-1429.

(11) Williams, B. A.; Lin, L.; Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2009**, *131*, 6330-6331.

(12) Mukhopadhyay, R. *Anal. Chem.* **2005**, *77*, 114A-118A.

(13) Famulok, M.; Hartig, J. S.; Mayer, G. *Chem. Rev.* **2007**, *107*, 3715-3743.

(14) Drolet, D. W.; Moon-McDermott, L.; Romig, T. S. *Nat. Biotechnol.* **1996**, *14*, 1021-1025.

(15) Murphy, M. B.; Fuller, S. T.; Richardson, P. M.; Doyle, S. A. *Nucleic Acids Res.* **2003**, *31*, e110.

(16) Bock, C.; Coleman, M.; Collins, B.; Davis, J.; Foulds, G.; Gold, L.; Greef, C.; Heil, J.; Heilig, J. S.; Hicke, B.; Hurst, M. N.; Husar, G. M.; Miller, D.; Ostroff, R.; Petach, H.; Schneider, D.; Vant-Hull, B.; Waugh, S.; Weiss, A.; Wilcox, S. K.; Zichi, D. *Proteomics* **2004**, *4*, 609-618.

(17) Romig, T. S.; Bell, C.; Drolet, D. W. *J. Chromatogr. B. Biomed. Sci. Appl.* **1999**, *731*, 275-284.

(18) Tombelli, S.; Minunni, M.; Mascini, M. *Biosens. Bioelectron.* **2005**, *20*, 2424-2434.

(19) Liu, J.; Mazumdar, D.; Lu, Y. *Angew. Chem. Int. Ed.* **2006**, *45*, 7955-7959.

(20) Wang, Z.; Lee, J. H.; Lu, Y. *Adv. Mater.* **2008**, *20*, 3263-3267.

(21) White, R. R.; Sullenger, B. A.; Rusconi, C. P. *J. Clin. Invest.* **2000**, *106*, 929-934.

(22) Siddiqui, M. A.; Keating, G. M. *Drugs* **2005**, *65*, 1571-1577.

(23) Patel, D. J.; Suri, A. K. *J. Biotechnol.* **2000**, *74*, 39-60.

(24) Macaya, R. F.; Schultze, P.; Smith, F. W.; Roe, J. A.; Feigon, J. *Proc. Natl. Acad. Sci.* **1993**, *90*, 3745-3749.

(25) Long, S. B.; Long, M. B.; White, R. R.; Sullenger, B. A. *RNA* **2008**, *14*, 2504-2512.

(26) Taussig, M. J.; Stoevesandt, O.; Borrebaeck, C. A.; Bradbury, A. R.; Cahill, D.; Cambillau, C.; de Daruvar, A.; Dubel, S.; Eichler, J.; Frank, R.; Gibson, T. J.; Gloriam, D.; Gold, L.; Herberg, F. W.; Hermjakob, H.; Hoheisel, J. D.; Joos, T. O.; Kallioniemi, O.; Koegl, M.; Konthur, Z.; Korn, B.; Kremmer, E.; Krobitsch, S.; Landegren, U.; van der Maarel, S.; McCafferty, J.; Muyldermans, S.; Nygren, P. A.; Palcy, S.; Pluckthun, A.; Polic, B.; Przybylski, M.; Saviranta, P.; Sawyer, A.; Sherman, D. J.; Skerra, A.; Templin, M.; Ueffing, M.; Uhlen, M. *Nat. Methods* **2007**, *4*, 13-17.

(27) Lohman, T. M. *CRC Crit. Rev. Biochem.* **1986**, *19*, 191-245.

(28) Jayaram, B.; DiCapua, F. M.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 5211-5215.

(29) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251-260.

(30) Lin, L.; Hom, D.; Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2007**, *129*, 14568-14569.

(31) Ramos, E.; Pineiro, D.; Soto, M.; Abanades, D. R.; Martin, M. E.; Salinas, M.; Gonzalez, V. M. *Lab. Invest.* **2007**, *87*, 409-416.

(32) Ramos, E.; Moreno, M.; Martin, M. E.; Soto, M.; Gonzalez, V. M. *Oligonucleotides* **2010**, *20*, 207-213.

(33) Berezovski, M. V.; Musheev, M. U.; Drabovich, A. P.; Jitkova, J. V.; Krylov, S. N. *Nat. Protoc.* **2006**, *1*, 1359-1369.

(34) Mendonsa, S. D.; Bowser, M. T. *J. Am. Chem. Soc.* **2004**, *126*, 20-21.

(35) Mendonsa, S. D.; Bowser, M. T. *J. Am. Chem. Soc.* **2005**, *127*, 9382-9383.

(36) Mosing, R. K.; Mendonsa, S. D.; Bowser, M. T. *Anal. Chem.* **2005**, *77*, 6107-6112.

(37) Berezovski, M.; Musheev, M.; Drabovich, A.; Krylov, S. N. *J. Am. Chem. Soc.* **2006**, *128*, 1410-1411.

(38) Tang, J.; Xie, J.; Shao, N.; Yan, Y. *Electrophoresis* **2006**, *27*, 1303-1311.

(39) Zuker, M. *Nucleic Acids Res.* **2003**, *31*, 3406-3415.

(40) Wong, I.; Lohman, T. M. *Proc. Natl. Acad. Sci.* **1993**, *90*, 5428-5432.

(41) Carothers, J. M.; Oestreich, S. C.; Davis, J. H.; Szostak, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 5130-5137.

(42) Sazani, P. L.; Larralde, R.; Szostak, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 8370-8371.

(43) Eaton, B. E.; Gold, L.; Zichi, D. A. *Chem. Biol.* **1995**, *2*, 633-638.

(44) Carothers, J. M.; Oestreich, S. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2006**, *128*, 7929-7937.

(45) Jain, S. S.; Tullius, T. D. *Nat. Protoc.* **2008**, *3*, 1092-1100.

Chapter 3

Darwinian Evolution of an Alternative Genetic System Provides Support for TNA as an RNA Progenitor

Reprinted with permission from Yu, H.; Zhang, S.; Chaput, J.C. Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor, *Nat. Chem.* **2012**, *4*, 183-187.

3.1. Abstract

The pre-RNA world hypothesis postulates that RNA was preceded in the evolution of life by a simpler genetic material, but it is not known if such systems can fold into structures capable of eliciting a desired function. Presumably, whatever chemistry gave rise to RNA would have produced other RNA analogues, some of which may have preceded or competed directly with RNA. Threose nucleic acid (TNA), a potentially natural derivative of RNA, has received considerable interest as a possible RNA progenitor due to its chemical simplicity and ability to exchange genetic information with itself and RNA. Here, we have applied Darwinian evolution methods to evolve, *in vitro*, a TNA receptor that binds to an arbitrary target with high affinity and specificity. This demonstration shows that TNA has the ability to fold into tertiary structures with sophisticated chemical functions, which provides evidence that TNA could have served as an ancestral genetic system during an early stage of life.

3.2. Introduction

The question of why nature chose ribofuranosyl nucleic acids as the molecular basis of life's genetic material has prompted a systematic analysis of nucleic acid structure with regard to the chemical etiology of RNA.[1] Results from these studies reveal that several potentially natural RNA alternatives are capable of Watson–Crick base pairing, demonstrating the capacity for these molecules to store genetic information. However, to establish a primitive metabolism, primordial genetic polymers would have also needed to fold themselves into stable tertiary structures with sophisticated functions such as ligand binding and catalysis.[2] Determining which, if any, of these RNA analogues are capable of folding into shapes that elicit a specific pre-defined function would help constrain models that describe the origin and early evolution of life on Earth. Of the genetic systems studied to date, (3′,2′)-α-L-threose nucleic acid (TNA, Figure 3.1) has generated considerable interest as a possible RNA progenitor because of the chemical simplicity of threose relative to ribose and the ability of TNA to form stable helical structures with complementary strands of itself and RNA.[3-5] This latter feature, which provides a plausible mechanism for the transfer of information between successive genetic systems, is remarkable considering that TNA polymers have a sugar-phosphate backbone that is one atom shorter than that of RNA.

Molecular evolution provides a powerful approach for investigating the functional properties of nucleic acids.[6,7] Until now, this method has been limited to DNA and RNA or close structural analogues thereof (for a review see ref. 8), because these were the only polymers with enzymes that could transcribe, reverse-transcribe and amplify genetic information.[8] Extending this approach to artificial genetic systems like TNA requires (i)

polymerases that can transcribe a library of random DNA sequences into TNA; (ii) a method of selection that allows functional members to be separated from the non-functional pool; and (iii) a strategy to amplify functional molecules that maintains the integrity of the selected sequence. Here, we address this problem by applying Darwinian evolution methods to evolve a TNA receptor with ligand-binding activity. This property of biopolymer folding implies that TNA, like DNA and RNA, can fold into structures with complicated functions—a prerequisite for any RNA progenitor in a hypothetical pre-RNA world.

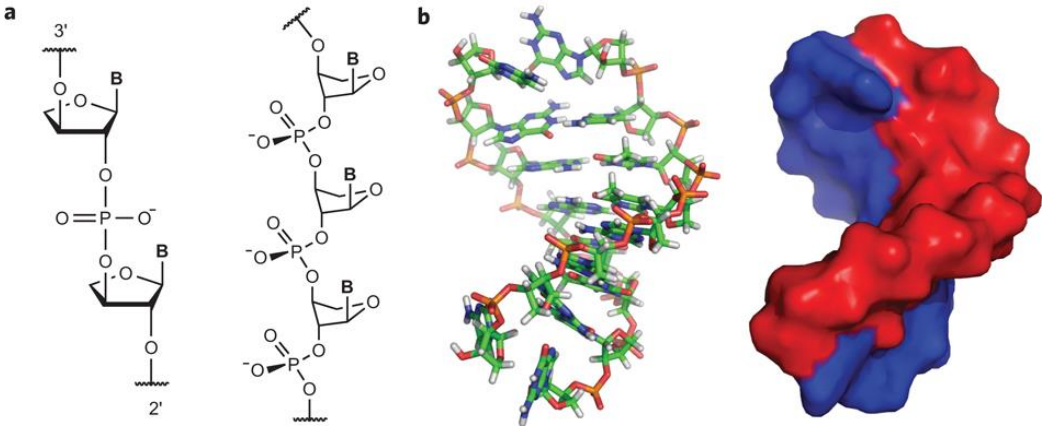


Figure 3.1. Structure of TNA. (A) Constitutional (left) and configurational (right) structures for the linearized backbone of an α-L-threofuranosyl-(3′ → 2′)-oligonucleotide. TNA contains one less atom per backbone repeat unit than natural RNA and DNA. B represents a nucleobase. (B) Solution NMR structure for the duplex formed from the self-pairing complementary sequence 3′-CGAATTCG-2′ (ref. 5). Duplex structures were rendered in PyMol with models showing ball and stick (left) and surface (right) images. TNA adopts a right-handed structure with helical parameters similar to A-form RNA.

3.3. Results and Discussion

The discovery by Eschenmoser and colleagues that TNA can undergo informational base pairing with complementary strands of itself and RNA inspired others to begin developing the methodology necessary to explore the functional properties of TNA by *in vitro* selection. Early work in this area showed that certain DNA polymerases can synthesize short segments of DNA on a TNA template,[9] whereas other polymerases can synthesize limited stretches of TNA on a DNA template.[10,11] Because the activity observed in these assays was insufficient to support *in vitro* selection, additional enzymes were screened for enhanced efficiency. The most active enzyme discovered thus far is therminator DNA polymerase, an engineered variant of 9°N DNA polymerase.[12,13] Under optimal conditions, therminator can transcribe an 80 nucleotide (nt) DNA template into TNA with high efficiency and fidelity. Because natural polymers of this length can fold into shapes with defined ligand binding sites, it was postulated that therminator could be used to generate pools of TNA molecules for *in vitro* selection.[12] Unfortunately, early attempts at *in vitro* selection were hindered by the inability of therminator to copy a DNA library into TNA.[14] This observation yielded an unexpected conundrum: why is it that therminator can transcribe individual DNA templates into TNA, but fails to do so when the substrate is changed from a single sequence to a large pool of random sequences?

We reasoned that individual templates analysed in previous studies might have contained unintentional design elements, such as a minimal secondary structure that favoured TNA synthesis. To address this problem, we developed a primer extension assay to examine TNA synthesis under a variety of conditions and substrate mixtures. A synthetic DNA primer was annealed to a synthetic DNA library (L1) that contained a

central random region of 50 nts followed by a second primer-binding site that could be used for downstream amplification using the polymerase chain reaction (PCR). We synthesized all four TNA triphosphates (tNTPs, Figure 3.2A) and used these molecules as substrates in the primer-extension assays. In these experiments, therminator DNA polymerase was challenged to extend the DNA primer with 70 TNA residues (Figure 3.2B). We used the diaminopurine analogue of adenine in place of adenosine 3′-triphosphate, because this substitution is known to increase the thermodynamic stability of the TNA–DNA heteroduplex.[9,15]

We began by attempting to copy an unbiased random-sequence DNA library (L1) into TNA. Consistent with earlier findings by Szostak and co-workers, primer extension failed under all conditions tested, which included varying the incubation time, temperature and salt concentration (Figure 3.2C).[14] To help identify which nucleotides were inhibiting transcription, we repeated the primer extension assay using defined mixtures of tNTP and dNTP substrates. By systematically substituting tNTP substrates for dNTP substrates, we discovered that the random library could be copied into full-length TNA product whenever the TNA substrate tCTP was replaced with the DNA substrate dCTP (Figure 3.2C). This finding suggested that repeating G nucleotides in the DNA library cause therminator DNA polymerase to pause during TNA transcription. Because unbiased DNA libraries like L1 would have an abundance of short G-repeats, we predicted that the individual templates used in previous studies were deficient in these motifs. This hypothesis was borne out by comparing representative sequences from the L1 library to those DNA templates used previously, which showed that nine out of ten

sequences in the L1 library contained at least one GGG or GGGG repeat, but none was present in the individual templates.

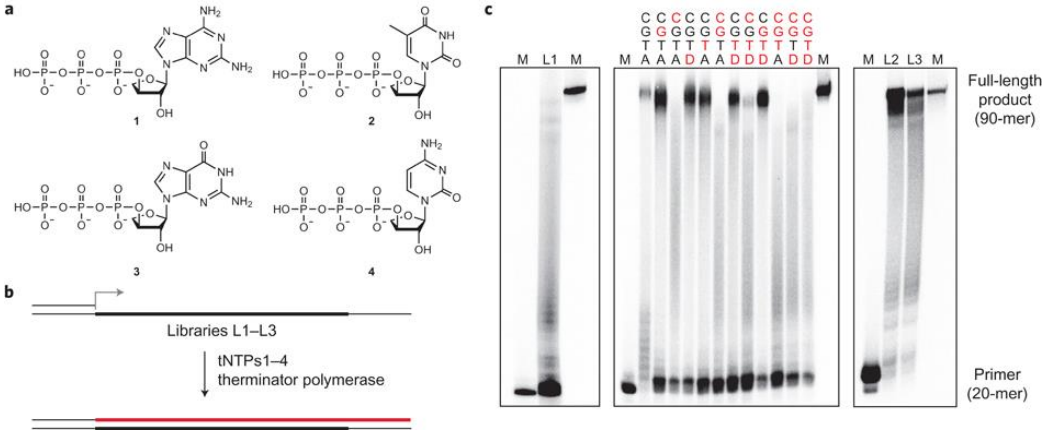


Figure 3.2. Synthesis of TNA libraries by enzyme-mediated primer extension. (A) Chemical structure of TNA triphosphates (tNTPs): tDTP, 1; tTTP, 2; tGTP, 3; tCTP, 4. (B) Schematic diagram of the primer-extension assay used to evaluate DNA libraries L1–L3. Each library contains a central random region of 50 nts flanked on either side by a 20 nt constant region. Library compositions: L1, equal distribution of A, C, G and T; L2, equal distribution of A, C and T; L3, one-half equivalent of G relative to A, C and T. (C) Therminator-mediated TNA transcription assays analysed by denaturing polyacrylamide gel electrophoresis. Primer extension of L1 with tNTPs 1–4 yields only trace amounts of full-length product (left panel). Primer extension of L1 using defined combinations of dNTPs (black) and tNTPs (red) leads to full-length product when tCTP is replaced with dCTP (centre panel). Primer extension across libraries L2 and L3 leads to full-length product in ~60% and ~30% yield, respectively (right panel). M: DNA marker.

In an effort to design DNA libraries that were better suited for TNA transcription, we synthesized DNA libraries L2 and L3 that minimized the occurrence of G residues in the random region. L2 contained equal amounts of A, C and T, and L3 limited the

frequency of G nucleotides to 50% the occurrence of A, C and T. Primer-extension assays performed on L2 and L3 (Figure 3.2C) yielded substantial amounts of the fully extended DNA primer (~60% and ~30%, respectively). In both cases, therminator DNA polymerase was able to extend the DNA primer with 70 sequential TNA nucleotides. This result demonstrated, for the first time, that combinatorial TNA libraries, which represent a nucleic acid system beyond the local structural neighbourhood of DNA and RNA, could be generated by enzyme-mediated extension of a DNA primer annealed to a DNA library.

To evaluate the fitness of TNA as a primordial biopolymer, we used an *in vitro* selection strategy to enrich for TNA molecules with receptor activity. The approach taken is analogous to other display technologies like mRNA display and DNA display that provide a covalent link between genotype (DNA) and phenotype (TNA).[16,17] The following strategy was originally devised by Ichida and Szostak, but was never implemented because of the earlier limitations of synthesizing TNA libraries on DNA templates.[12] We chose this selection strategy because it provides a clever solution to the problem of how to evolve an alternative genetic system *in vitro* when a reverse transcriptase is not available to convert surviving molecules back into DNA for amplification by PCR. Accordingly, a single-stranded DNA library was constructed with a stable stem-loop structure that serves as a DNA primer for the extension of TNA across the random region (Figure 3.3A). Extension of the primer leads to the formation of a chimeric TNA–DNA hairpin duplex. The TNA portion of the duplex was displaced by extending a separate DNA primer annealed to the stem-loop region of the hairpin with DNA. The product of the strand displacement is a TNA molecule that is physically linked

to its own double-stranded DNA message. To enrich for TNA molecules with specific ligand-binding activity, the pool of TNA–DNA fusions was incubated with a desired target and bound sequences were recovered and amplified by PCR. The coding strand of the DNA duplex was made single-stranded and folded into self-priming hairpin templates to facilitate another selection cycle.

Successive rounds of *in vitro* selection and amplification were performed using the L2 library, which gave the highest yield of full-length TNA product in our primer-extension assay. Although the L2 library generates TNA polymers that lack cytidine, we reasoned that this was not a significant concern as cytidine may not have been present in the first genetic material due to its tendency to undergo spontaneous deamination ($t_{1/2}$ = 340 yr at pH 7 and 25 ℃).[18] Furthermore, it has been shown that ribozymes missing cytidine can be generated by *in vitro* evolution, demonstrating that a three-letter genetic alphabet can still retain the ability to fold and function.[19] Using the strategy described above, we selected for TNA molecules that could bind to human thrombin and therefore exhibit an arbitrary specific function. We chose capillary electrophoresis as our separation technique because it leads to much higher partitioning efficiency than is commonly observed with traditional gravity filtration methods.[20-22] TNA–DNA fusion molecules carried the fluorescent label 6-carboxy-fluorescein to facilitate detection by laser-induced fluorescence. For each selection cycle, five 70 nl portions of library (~1×10^{11} TNA–DNA fusion molecules) were injected onto a neutral coated capillary. Functional sequences were recovered by allowing the unbound pool to pass into a waste vial and applying pressure to the column to isolate the bound molecules in a separate vial. After three rounds, a slower migrating peak corresponding to TNA–DNA fusion

molecules bound to human thrombin was visible (See Figure S1 in APPENDIX B), indicating that the pool had become enriched in TNA aptamers with affinity to thrombin.

We cloned and sequenced the DNA to determine the diversity of molecules that remained in the pool. Several clones contained isolated G residues, indicating that a low level of random mutagenesis had occurred during either the strand displacement or PCR amplification step of the selection. Although some mutations may have occurred during the TNA transcription step, these genetic changes would not have been carried through the selection, as the TNA molecules were not reverse transcribed back into DNA. Ten representative sequences were chosen for individual analysis. Each of these TNA molecules was synthesized by extending a DNA primer annealed to a synthetic DNA template with TNA and purifying the DNA-tagged TNA molecules by denaturing gel electrophoresis. We noticed that the representative sequences transcribed with much greater efficiency than the starting pool (1 h versus 10 h), indicating that the selection favoured the enrichment of DNA templates that transcribed with high efficiency. The resulting TNA aptamers have equilibrium dissociation constants (K_d) for thrombin that fall within the range 200–900 nM (See Table S1 in APPENDIX B), which is similar to previously evolved DNA and RNA aptamers that bind to thrombin.[23,24]

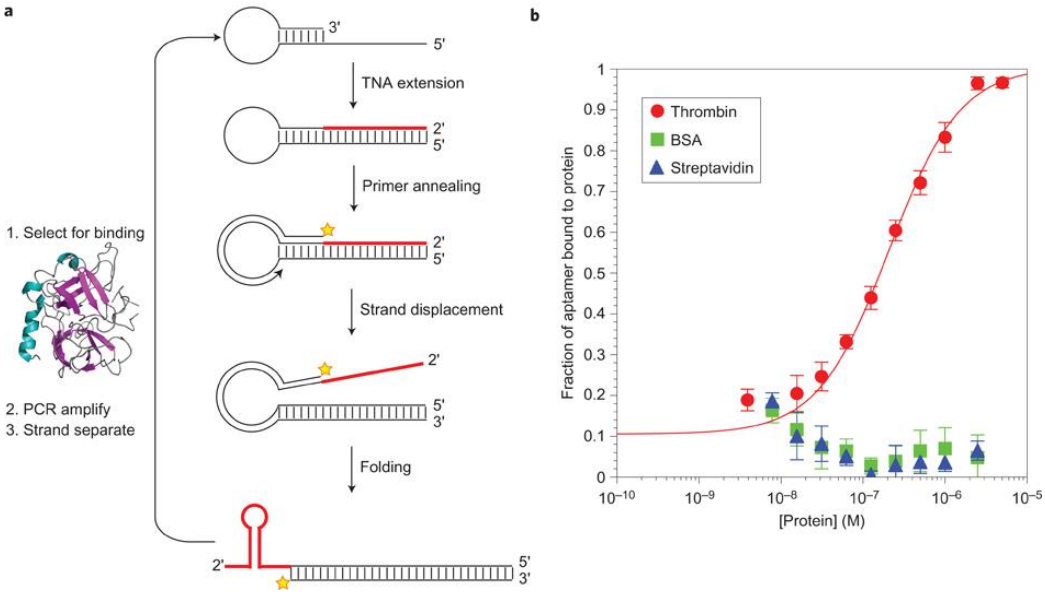


Figure 3.3. Evolution of TNA receptors *in vitro*. (A) *In vitro* selection strategy designed to isolate TNA aptamers with affinity to human thrombin. The DNA library encodes a random region of 50 nt positions flanked on the 3′ end with a stem-loop structure that serves as a DNA primer and a fixed-sequence primer-binding site located at the 5′ end. The DNA primer is extended with tNTPs to produce a chimeric TNA–DNA hairpin. A separate DNA primer modified with 6-carboxy-fluorescein (star) is annealed to the stem-loop region and extended with DNA to displace the TNA strand. The resulting pool of TNA–DNA fusion molecules is incubated with the protein target. Bound molecules are separated from the unbound pool by capillary electrophoresis and amplified by PCR. The dsDNA is made single-stranded and the coding strand is annealed to generate a new pool of DNA templates for the next selection cycle. (B) Equilibrium binding affinity measurement of the core binding domain of TNA aptamer 3.12. The aptamer (3′-TGTTDTDGDDDDDDTDDTGGTGGGGGGTTTDGDTDDDGGG G-2′) binds to

human thrombin with a K_d of 200 nM and shows no detectable affinity for BSA or streptavidin. Error bars, standard deviation of each data point (n = 3).

We determined the minimum binding region for the highest-affinity aptamer (clone 3.12, $K_d \approx 200$ nM at 25 ℃) by deletion analysis. TNA variants were generated by primer extension using DNA templates that contained incremental deletions from both ends of the coding region. We compared the K_d for each deletion construct to the parent sequence, which defined an internal core boundary of 41 nts that retained high-affinity binding (Figure 3.3B, See Figure S2 in APPENDIX B). The affinity of the aptamer did not change when the DNA primer was exchanged for a different sequence and a randomly generated TNA sequence had no affinity to human thrombin (data not shown). Similarly, the DNA primer alone had no affinity to human thrombin (See Figure S2 in APPENDIX B), ruling out the possibility that the DNA primer was involved in ligand binding. Taken together, these results demonstrate that the selected aptamer probably functions by forming specific contacts to the protein surface. To show that the TNA aptamer adopts a folded structure that is specific to human thrombin, we measured its solution binding affinity against two common proteins, bovine serum albumin (BSA) and streptavidin. The TNA molecule had no detectable affinity for either off-target protein (Figure 3.3B), indicating that the evolved aptamer binds to thrombin with high affinity and specificity.

The idea that life on Earth arose during an ancestral period in which RNA stored genetic information and catalysed metabolically relevant chemical reactions was discussed extensively over 40 years ago,[25-27] and was later termed ‘the RNA world hypothesis’.[28] Although the chemical plausibility of the RNA world remains strong,[29] it is

not clear that RNA was the first genetic material.[30,31] Problems associated with the prebiotic synthesis of ribose and non-enzymatic replication of RNA have fueled interest in the possibility that RNA was preceded in the evolution of life by a genetic system that was more accessible on the primitive Earth.[32,33] Determining which genetic systems are both simpler than RNA and capable of Darwinian evolution is a daunting challenge. Chemical synthesis is required to build substrates that are not otherwise available and enzymes must be discovered that can make unnatural polymers with high efficiency and fidelity. In the case of TNA, for example, a decade-long effort was required to establish the methodology necessary to evolve TNA molecules *in vitro*. Although new advances in protein engineering are making it easier to create polymerases that recognize unnatural substrates,[34,35] future examination of structurally diverse genetic systems by *in vitro* evolution will require similar advances in nucleic acid chemistry.

3.4. Conclusion

In summary, we have shown that TNA has the ability to fold into tertiary structures that can elicit a complex function. The best TNA aptamer identified in our selection binds to human thrombin with an activity similar to previously evolved RNA aptamers,[23,24] demonstrating that the chemical problem of ligand binding is not unique to the ribofuranosyl nucleic acid polymers of RNA and DNA. This fact was not clear at the beginning of our study, and indeed one could imagine that purely chemical constraints, such as a shorter backbone repeat unit, might preclude the ability of TNA to fold into structures that function with a desired activity. The fact that TNA does not appear to be limited in this regard suggests that it may be possible to isolate novel TNA enzymes from pools of random sequences using *in vitro* evolution. We suggest that selections of this

type could be used to further examine the fitness of TNA as an RNA progenitor in a hypothetical TNA world.

3.5. Methods

3.5.1. Oligonucleotides and TNA triphosphates. DNA primer and templates were purchased from Integrated DNA Technologies. DNA library L1 was purchased from the Keck Facilities at Yale University. DNA libraries L2 and L3 were synthesized on an automated ABI 3400 DNA synthesizer, deprotected in concentrated NH_4OH for 18 h at 55 ℃, butanol precipitated, and purified by denaturing polyacrylamide gel electrophoresis. TNA triphosphates were synthesized as previously described.[36]

3.5.2. Primer extension assay. The DNA primer (5′-GACACTCGTATGCAGTAGCC) was 5′-end labelled by incubation in the presence of [γ-^{32}P] ATP with T4 polynucleotide kinase for 1 h at 37 ℃. ^{32}P-labelled primer was annealed to the DNA template (5′-TGTCTACACGCAAGCTTACA-N_{50}-GGCTACTGCATACGAGTGTC) 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) by heating at 95 ℃ for 5 min and cooling on ice. Primer extension reactions contained 100 μM tNTPs (or a combination of defined tNTPs and dNTPs mixtures), 500 nM primer–template complex, 1 mM DTT, 100 μg ml^{-1} BSA, 1.25 mM $MnCl_2$ and 0.1 U μl^{-1} Therminator DNA polymerase (New England Biolabs). Reactions were performed by adding the tNTP substrates to a solution containing all other reagents, and heating the mixture for 10 h at 55 ℃. Primer extension products were analysed by 20% denaturing polyacrylamide gel electrophoresis.

3.5.3. *In vitro* selection. For each round of selection, the self-priming DNA library L2 (5′-TGTCTACACGCAAGCTTACA-N_{50}-GGCTACTGCATACGAGTGTCac tacgtacccacaacctcggccgtaccacggtacgtagtgacactcgtatgcagtagcc; lower-case letters denote the self-priming stem-loop region) was extended with tNTPs to generate a pool of chimeric TNA–DNA hairpin structures. The extension product was phenol/chloroformextracted, desalted and subjected to a strand invasion step. A DNA primer (5′-FAM-AAGGCTACTGCATACGAGTGTCACTACGTACCGTGGTACGGC CGAGGTTG-3′) was annealed to the stem-loop region of the chimeric TNA–DNA hairpin. The primer–template complex was extended with therminator DNA polymerase using 2.5 mM dNTPs in 1× ThermoPol buffer for 30 min at 55 ℃, followed by an additional heating step of 90 min at 65 ℃. The resulting pool of double-stranded DNA–TNA fusion molecules was incubated with human α-thrombin (Haematologic Technologies) in selection buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM $MgCl_2$, pH 7.4) for 1 h at 25 ℃. The ratio of human thrombin to the TNA library was decreased over the course of the selection to favour the enrichment of high-affinity aptamers. In round 1, the ratio was 10:1; in round 2 the ratio was 1:1; and in round 3, the ratio was 1:10. Bound molecules were separated from the unbound pool using capillary electrophoresis. Electrophoresis was performed on a Beckman ProteomeLab PA 800 Protein Characterization System. Before use, the glass capillary (inner diameter, 0.1 mm; total length, 60 cm) was rinsed with water and equilibrated with selection buffer. A small portion (70 nl) of the library–thrombin mixture was injected onto the capillary using pressure injection (0.5 psi for 5 s) and electrophoresis was performed under a constant voltage of 10 kV for 30 min at 20 ℃. Laser-induced fluorescence (LIF) was used to

monitor the separation of 6-FAM labelled DNA–TNA fusion molecules (excitation, 488 nm; emission, 520 nm). Five injections were performed for each round of *in vitro* selection. The thrombin-bound TNA–DNA fusion molecules were collected and the DNA portion was amplified by PCR using a top strand biotinylated primer. The bottom strand was isolated by denaturing the double-stranded DNA on streptavidin-coated agarose beads and annealed to generate a new pool of DNA templates as input for the next round of selection. After three rounds of *in vitro* selection and amplification, the library was cloned and sequenced to examine the diversity of molecules that remained in the pool.

3.5.4. Filter binding assay. Representative TNA molecules were synthesized by primer extension and purified by denaturing polyacrylamide gel electrophoresis as described above with the exception that the extension time was reduced to 1 h and the DNA templates contained an A_{10}-overhang at the 3′-end to enable the product to be separated from the template by denaturing polyacrylamide gel electrophoresis. The DNA portion of the TNA molecule was labelled with ^{32}P, desalted, and annealed by heating for 5 min at 95 ℃ and cooling on ice. Folded structures were incubated with thrombin poised at concentrations spanning the expected K_d (typically 1 nM–1 μM) at 25 ℃. After 1 h, the protein-bound TNA molecules were partitioned away from the unbound fractions using vacuum to pass the solution through a layer of nitrocellulose and nylon membranes. Both membranes were quantified by phosphorimaging, and dissociation constants were calculated using nonlinear least-squares regression analysis performed with DeltaGraph program as previously described.[37]

3.6. References

(1) Eschenmoser, A. *Science* **1999**, *284*, 2118-2124.

(2) Benner, S. A.; Ellington, A. D.; Tauer, A. *Proc. Natl. Acad. Sci.* **1989**, *86*, 7054-7058.

(3) Schoning, K.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347-1351.

(4) Orgel, L. *Science* **2000**, *290*, 1306-1307.

(5) Ebert, M. O.; Mang, C.; Krishnamurthy, R.; Eschenmoser, A.; Jaun, B. *J. Am. Chem. Soc.* **2008**, *130*, 15105-15115.

(6) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611-647.

(7) Joyce, G. E. *Angew. Chem. Int. Ed.* **2007**, *46*, 6420-6436.

(8) Keefe, A. D.; Cload, S. T. *Curr. Opin. Chem. Biol.* **2008**, *12*, 448-456.

(9) Chaput, J. C.; Ichida, J. K.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 856-857.

(10) Chaput, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 9274-9275.

(11) Kempeneers, V.; Vastmans, K.; Rozenski, J.; Herdewijn, P. *Nucleic Acids Res.* **2003**, *31*, 6221-6226.

(12) Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 2802-2803.

(13) Horhota, A.; Zou, K. Y.; Ichida, J. K.; Yu, B.; McLaughlin, L. W.; Szostak, J. W.; Chaput, J. C. *J. Am. Chem. Soc.* **2005**, *127*, 7427-7434.

(14) Ichida, J. K.; Horhota, A.; Zou, K. Y.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Res.* **2005**, *33*, 5219-5225.

(15) Wu, X. L.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **2002**, *4*, 1283-1286.

(16) Roberts, R. W.; Szostak, J. W. *Proc. Natl. Acad. Sci.* **1997**, *94*, 12297-12302.

(17) Rosenbaum, D. M.; Liu, D. R. *J. Am. Chem. Soc.* **2003**, *125*, 13924-13925.

(18) Levy, M.; Miller, S. L. *Proc. Natl. Acad. Sci.* **1998**, *95*, 7933-7938.

(19) Rogers, J.; Joyce, G. F. *Nature* **1999**, *402*, 323-325.

(20) Lin, L.; Hom, D.; Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2007**, *129*, 14568-14569.

(21) Williams, B. A.; Lin, L.; Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2009**, *131*, 6330-6331.

(22) Mendonsa, S. D.; Bowser, M. T. *J. Am. Chem. Soc.* **2004**, *126*, 20-21.

(23) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. *Nature* **1992**, *355*, 564-566.

(24) Kubik, M. F.; Stephens, A. W.; Schneider, D.; Marlar, R. A.; Tasset, D. *Nucleic Acids Res.* **1994**, *22*, 2619-2626.

(25) Woese, C. R.; Harper & Row: 1967.

(26) Crick, F. H. *J. Mol. Biol.* **1968**, *38*, 367-379.

(27) Orgel, L. E. *J. Mol. Biol.* **1968**, *38*, 381-393.

(28) Gilbert, W. *Nature* **1986**, *319*, 618-618.

(29) Cech, T. R. *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a006742.

(30) Joyce, G. F. *Nature* **2002**, *418*, 214-221.

(31) Yarus, M. *Cold Spring Harb Perspect Biol* **2011**, *3*, a003590.

(32) Joyce, G. F.; Schwartz, A. W.; Miller, S. L.; Orgel, L. E. *Proc. Natl. Acad. Sci.* **1987**, *84*, 4398-4402.

(33) Joyce, G. F. *New Biologist* **1991**, *3*, 399-407.

(34) Loakes, D.; Holliger, P. *Chem. Commun.* **2009**, 4619.

(35) 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. *Science* **2012**, *336*, 341-344.

(36) Zou, K. Y.; Horhota, A.; Yu, B.; Szostak, J. W.; McLaughlin, L. W. *Org. Lett.* **2005**, *7*, 1485-1487.

(37) Yu, H.; Jiang, B.; Chaput, J. C. *ChemBioChem* **2011**, *12*, 2659-2666.

Chapter 4

Progress towards an in vitro Selection for RNA-cleaving TNA Enzymes

Partially adapted with permission from Yu, H.; Zhang, S.; Dunn, M.R.; Chaput, J.C. An Efficient and Faithful in Vitro Replication System for Threose Nucleic Acid, *J. Am. Chem. Soc.* **2013**, *135*, 3583-3591.

4.1. 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.

4.2. 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 4.1A), 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 $\sim 10^{15}$ 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

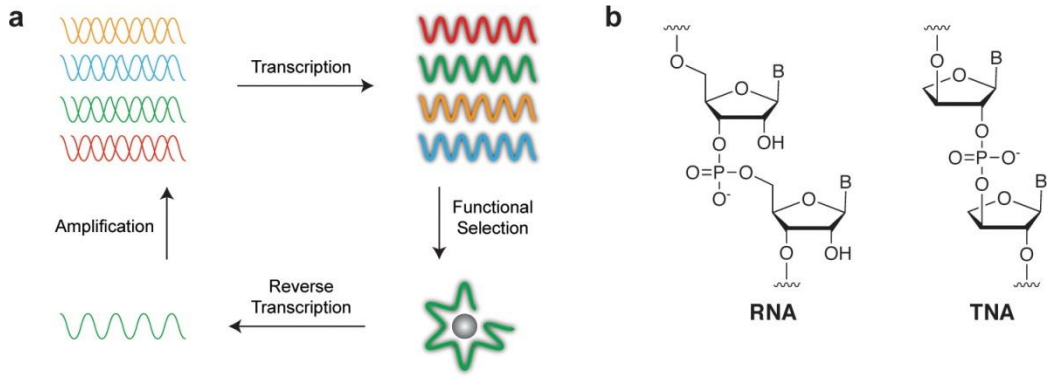


Figure 4.1. Darwinian evolution of natural and artificial nucleic acid polymers. (A) In vitro evolution of nucleic acid polymers requires polymerases that can transcribe, reverse transcribe, and amplify genetic information in the laboratory. For RNA, each cycle of in vitro selection and amplification 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).

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 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 4.1B).[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.

4.3. Results

4.3.1. 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,

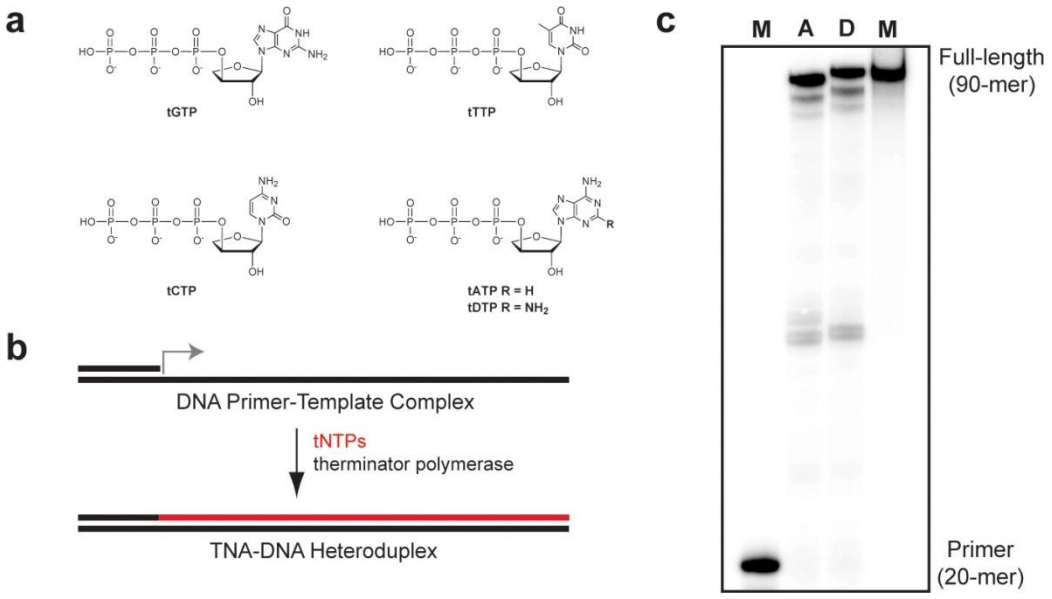


Figure 4.2. Enzyme-mediated transcription of a DNA library into a library of TNA sequences. (A) Chemical structures of TNA triphosphates (tNTPs). Diaminopurine (tDTP) is an analogue of adenine (tATP) that forms three hydrogen bonds with thymine. (B) Schematic representation of DNA primer extension reaction 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 ℃ 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.

tCTP, and tGTP) as well as the diaminopurine analogue (tDTP) of adenine threofuranosyl 3′-triphosphate (Figure 4.2A). 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,[25,28,33] 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 4.2B, 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 ℃ 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. In both cases, the DNA primer was completely extended with TNA residues to make the desired full-length product (Figure 4.2C). 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.

4.3.2. 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 4.3A). 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 ℃. 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 (See Figure S1 in APPENDIX C).

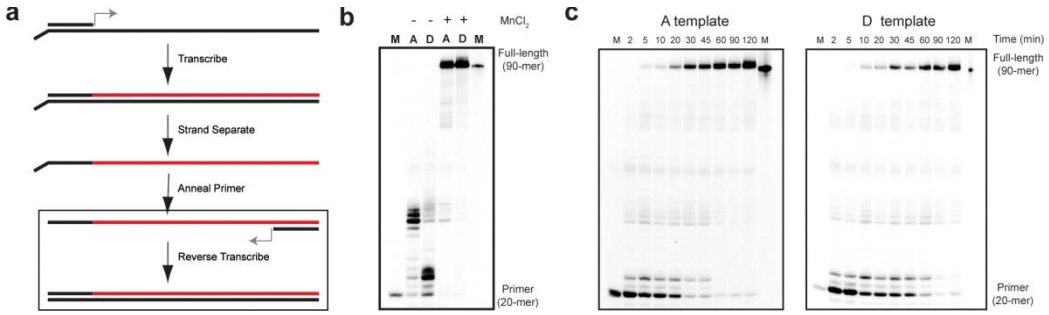


Figure 4.3. Enzyme-mediated reverse transcription of individual TNA molecules into DNA. (A) Schematic representation of TNA synthesis and reverse transcription. Boxed region highlights the reverse transcription step analyzed by gel electrophoresis. The TNA region of each strand is shown in red, while the cDNA region is shown in black. (B) SuperScript II-mediated TNA reverse transcription performed for 24 h at 42 ℃ 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.

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 ℃. 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 4.3B). 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 4.3C). 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.

4.3.3. 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 (See Figure S2 in APPENDIX C). 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 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 (See Figure S3 in APPENDIX C). 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 (See Table S3 in APPENDIX C and Figure 4.4A).

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 in APPENDIX C). 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 4.4B). 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% (See Table S3 in APPENDIX C and Figure 4.4C).

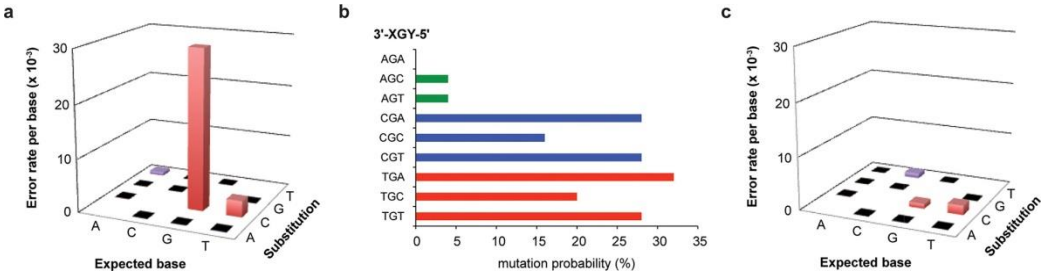


Figure 4.4. Fidelity of TNA replication using a four-letter genetic alphabet. (A) The mutation profile of TNA replication indicates a high frequency of G → C substitutions during TNA transcription. (B) Analysis of the local sequence context upstream and downstream of the misincorporation site indicates 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}.

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 (See Figure S5 in APPENDIX C). 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 4.5A). Consistent with the efficient replication of the L2 library, the template 3NT.ATC exhibits an overall fidelity of replication of 99.6% (Figure 4.5B), 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 4.5C). 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.

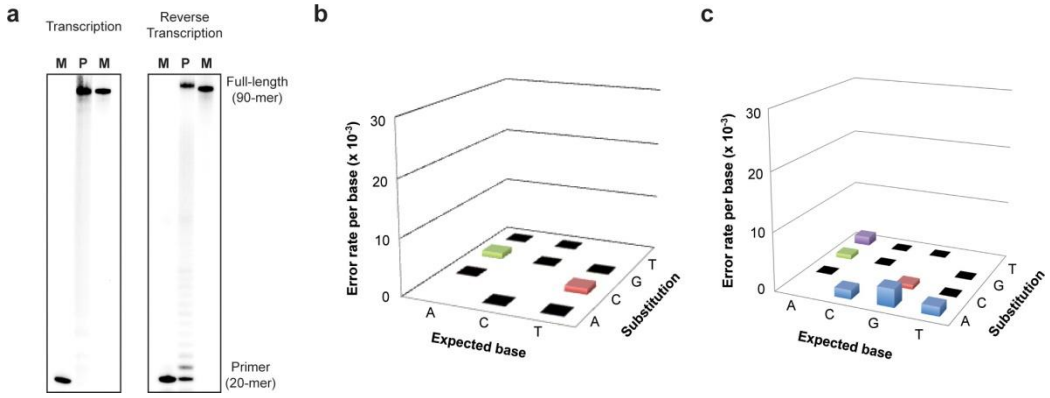


Figure 4.5. An efficient and faithful replication system for TNA. (A) Replication of a three letter TNA library. A DNA library composed of three nucleotides (A, C, and T) transcribes into TNA (left panel) and reverse transcribes back into DNA (right panel) with high primer-extension efficiency. (B) Mutation 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).

4.3.4. 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 4.6). As expected, the L2 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]

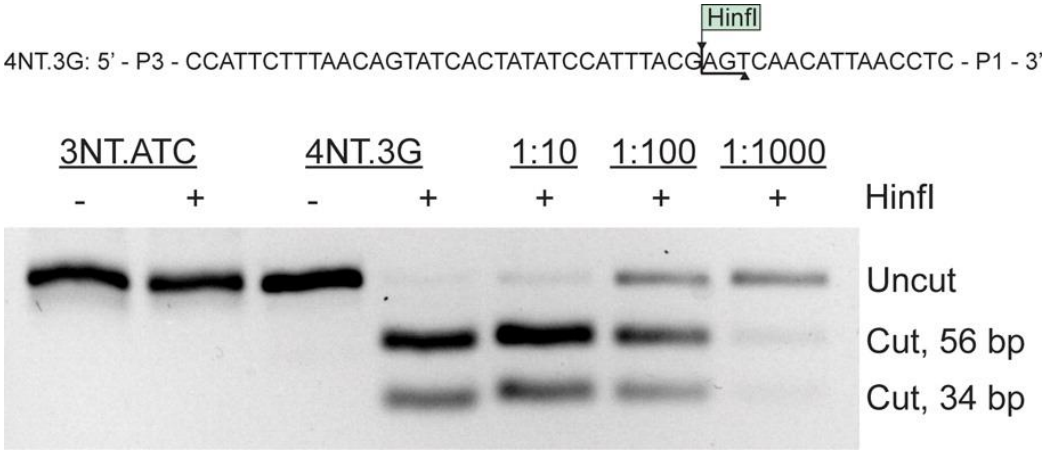


Figure 4.6. Enrichment of biotinylated TNA from a TNA library. 10^{14} molecules of the L2 library were doped with 1/10th, 1/100th, and 1/1000th molar ratio of biotinylated

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.

4.3.5. 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 ℃ 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 4.7A). 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 4.7B). 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 4.7A-B). This result demonstrates that 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

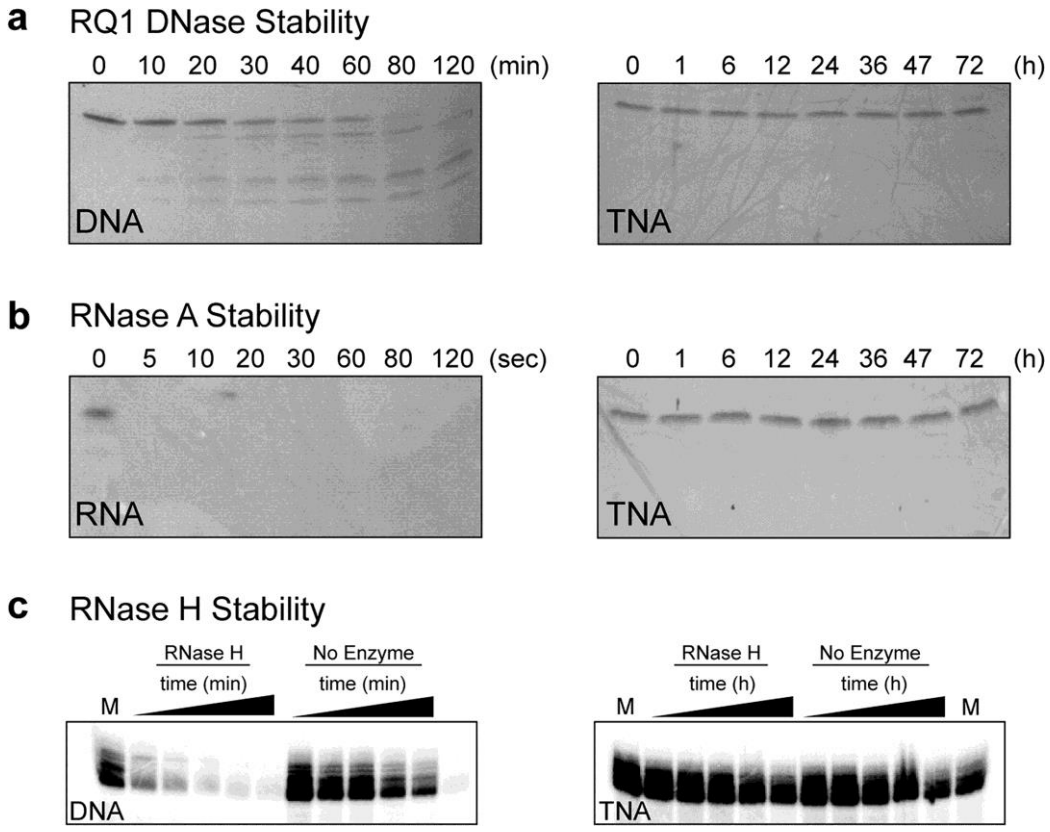


Figure 4.7. TNA sensitivity to nuclease degradation. Nuclease stability of synthetic DNA, RNA, and TNA oligonucleotides were monitored over time by denaturing polyacrylamide gel electrophoresis. (A) In the presence of RQ1 DNase, DNA exhibits a half-life of ~30 min, while TNA remains undigested after 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.

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 ℃ 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 4.7C).

4.4. 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 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.

4.5. Conclusion

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.

4.6. Progress towards an in vitro selection for RNA-cleaving TNA enzymes

Nucleic acid enzymes can be identified by in vitro selection as well. Compared with aptamer selection, which usually separates target-binding library members from the unbound sequences by affinity chromatography, enzyme selection aims to enriching for catalytically active molecules that catalyzes a specific chemical reaction on a substrate. Consequently, in order to separate functional sequences from the pool for amplification, the target chemical reaction typically involves acquisition or loss of a distinct tag from the enzyme itself.

RNA hydrolysis was chosen as the target reaction because 1) it is a model reaction that many naturally occurring ribozymes catalyze, and 2) numerous ribozymes and deoxyribozymes have been identified by in vitro selection to promote specific RNA cleavage reaction. A classic example for selecting RNA-cleaving DNA enzymes (deoxyribozymes) was performed by Santoro and Joyce in 1997 (Figure 4.8).[43] Briefly, a DNA library was used as template in a DNA primer extension reaction. The DNA primer

itself was modified with a biotin moiety on the 5' end and contained 12 embedded ribonucleotides. After primer extension, the double-stranded DNA was immobilized onto a streptavidin affinity column, and the template strand was removed by NaOH denaturation. Then the catalysis reaction buffer containing divalent metal ioins such as Mg^{2+} was added to help single-stranded DNA library members to fold into tertiary structures. Sequences that can fold and promote RNA hydrolysis reaction would cleave themselves off the column and could be amplified for the next round of selection.

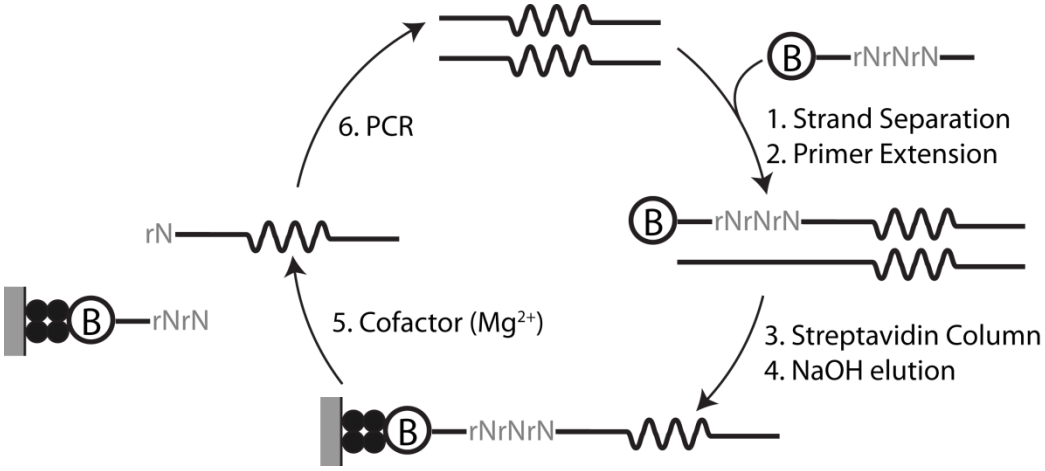


Figure 4.8. In vitro selection strategy for RNA-cleaving DNA enzymes. An embedded RNA substrate is linked to a random DNA library by primer extension. In the presence of divalent metal ion cofactor, DNA sequences that catalyze intramolecular RNA hydrolysis reaction are eluted from affinity column and amplified.

I validated this experimental design and successfully identified a deoxyribozyme that catalyze RNA cleavage at a specific position. I then attempted to adapt this selection scheme to TNA. The primer extension and selection steps are similar but the amplification is a bit different. Because selected TNA molecules are not recognized by Taq DNA polymerase and thus can not be directly amplified by PCR, I first converted

TNA back to complementary DNA (cDNA) using SuperScript II catalyzed reverse transcription, and then amplified cDNA to generate a pool for next cycle of selection.

Close inspection of each step of in vitro selection with control experiments revealed that it suffered heavily from contamination. Namely, a control PCR experiment using TNA as template was compared with normal RT-PCR. Surprisingly, the no-RT control produced very similar bands to regular RT-PCR amplification (Figure 4.9). This result suggested that the recovered TNA molecules were contaminated by leftover DNA templates. In other words, the duplex denaturation by NaOH was incomplete. Using harsher denaturation conditions such as prolonged incubation time and increased NaOH concentration might be helpful but could potentially induce nonspecific RNA hydrolysis as well and thus consume all substrates before the actual selection step.

Now that we realized that NaOH denaturation was not an effective approach to separate the two strands in TNA-DNA heteroduplex, we attempted to achieve the partition by denaturing gel electrophoresis. Since TNA exhibits essentially the same mobility as DNA sequence of the same length, we had to introduce certain linkers to the DNA template in order to induce a mobility shift (Figure 4.10A). We chose polyethylene glycol (PEG) because it is not accepted by polymerases and thus acts as a terminator in PCR and primer extension. We used a PEG-modified primer and a regular primer to amplify the DNA library by PCR. This produced a long PEG-modified strand and a short unmodified strand due to the blocking effect of PEG. The PEG-modified template was separated and purified by denaturing polyacrylamide gel electrophoresis (Figure 4.10B), and used in TNA transcription. Then the TNA strand was separated from DNA template on a denaturing gel. However, the TNA molecules purified using this method still

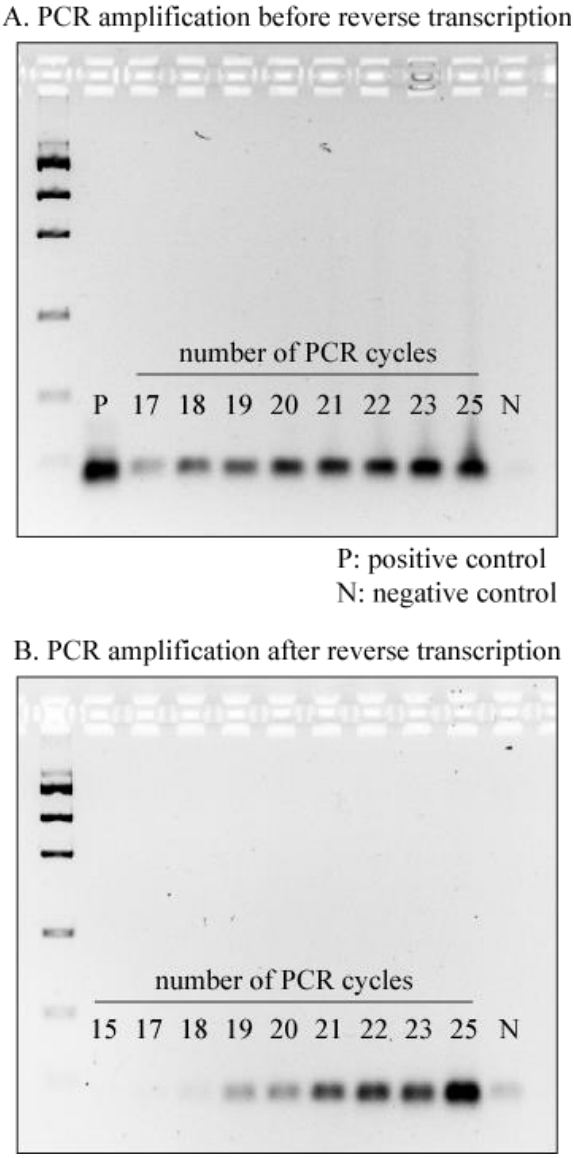


Figure 4.9. PCR amplification in in vitro selection for RNA-cleaving TNA enzymes. (A) Selected TNA molecules could be amplified before reverse transcription, implying the presence of DNA contaminant. (B) RT-PCR yielded similar band intensity to no-RT controls, suggesting a massive contamination by DNA template. P: positive control using DNA as template; N: negative control without template.

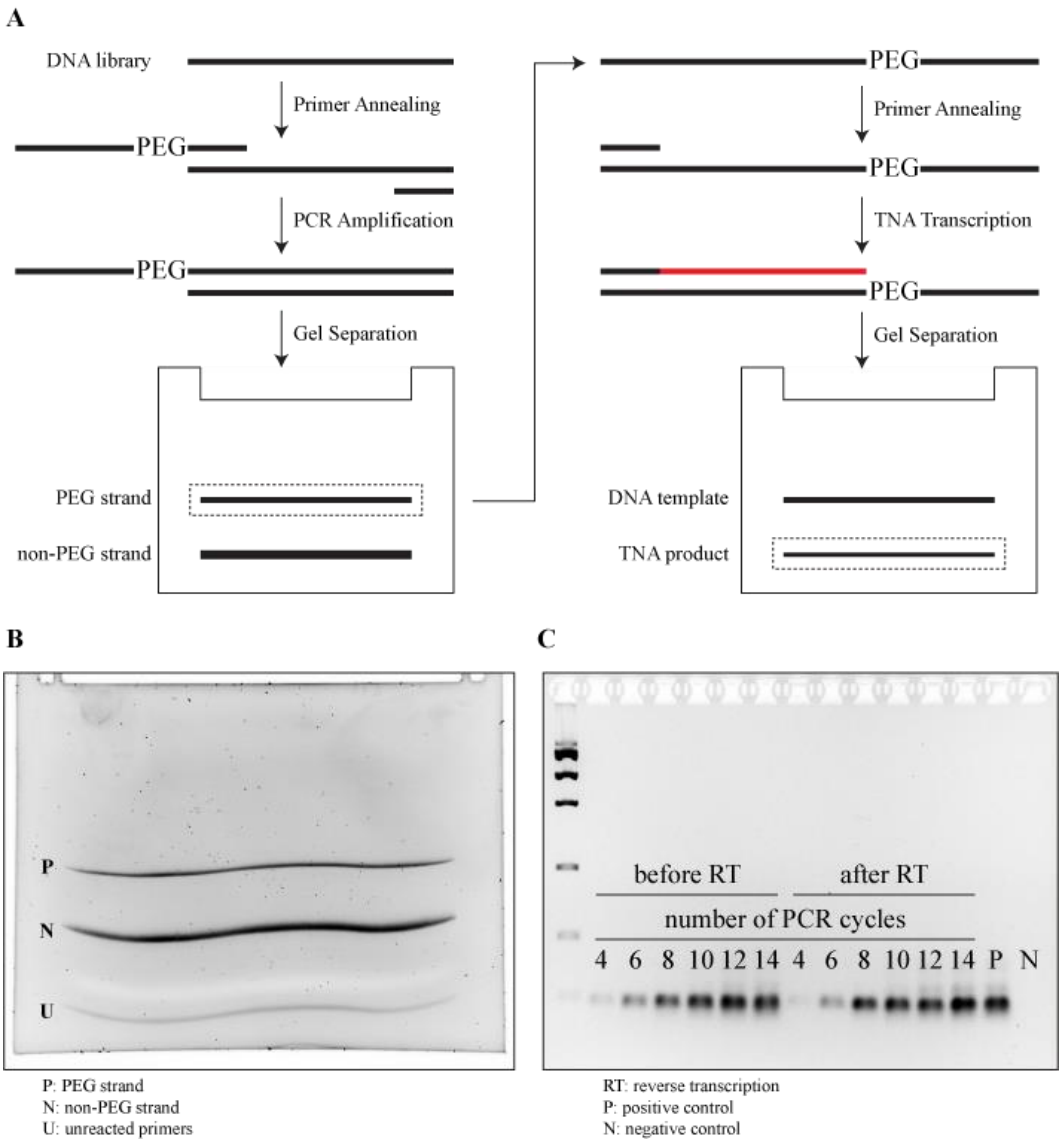


Figure 4.10. Separation method using PEG-modified DNA template. (A) PEG-modified template is first generated by PCR and purified by denaturing PAGE. PEG-template is then used in TNA transcription and facilitates separation of TNA product. (B) Denaturing gel shows separation of PEG strand, non-PEG strand and unreacted primers after PCR. (C) Purified TNAs are amplified before RT, implying the presence of contaminant.

contained DNA contaminant, and could be directly amplified by PCR (Figure 4.10C). It was not entirely clear how the DNA contaminant managed to survive the purification process.

Next, we designed a different approach, no longer trying to separate one strand exclusively from the other but accepting that NaOH denaturation is incomplete and then aiming to selectively amplify cDNA instead of contaminant after reverse transcription. This involves using a new RT primer, which has an overhanging region on its 5' end, during TNA reverse transcription. In the subsequent PCR amplification, one of the primers shares the same sequence as the overhanging region of RT primer. By implementing this design, PCR should preferentially amplify the cDNA sequence over contaminant because the latter lacks one primer binding site (Figure 4.11A). As expected, RT-PCR exhibited exponential amplification, while the no-RT control yielded no detectable band (Figure 4.11B). Although the amplification pattern was encouraging, the generated bands were of incorrect size (expected: 110bp; observed: << 100bp). We hypothesize that the overhang PCR primer binds unintentionally to an internal binding site on cDNA and generate short amplification product. To test this hypothesis, we synthesized 11 additional pairs of RT primers and PCR primers. Unexpectedly, they all generated dominant bands well below the correct length (Figure 4.11C). It seemed unlikely that each of the new primer design would bind internally to a random cDNA library. The mechanism underlying the appearance of these short amplification products remained mysterious.

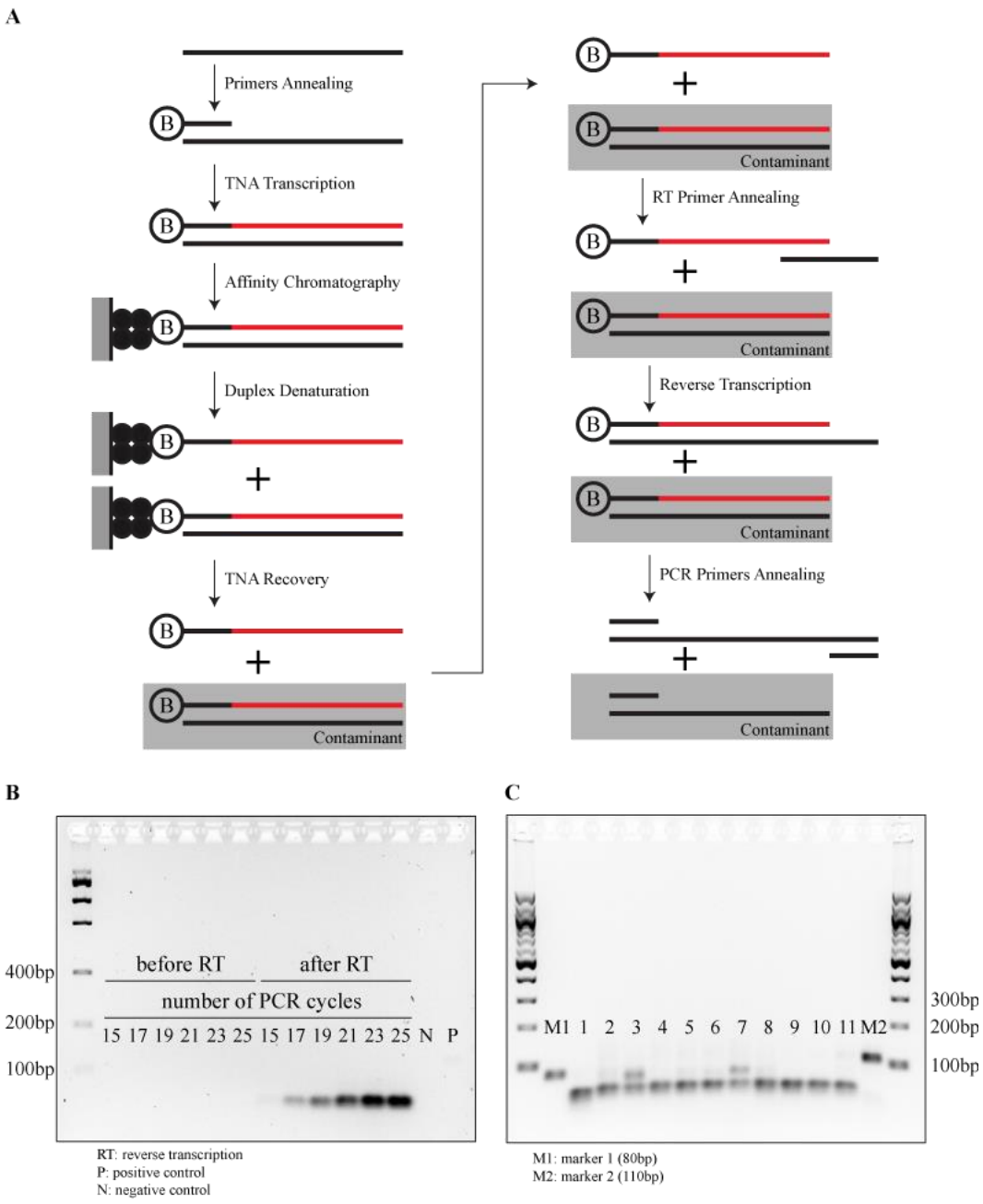


Figure 4.11. RT-PCR using overhanging RT primers. (A) The selective amplification occurs during RT-PCR because the contaminant lacks one primer binding site. (B) PCR amplification of TNAs before and after RT. (C) RT-PCR using different overhang sequences.

Since TNA is resistant to nuclease-mediated degradation, DNase treatment should represent an effective approach to remove any residual DNA contaminant. To implement a DNase-assisted decontamination, the DNA primer used in TNA transcription has to be substituted by an RNA primer so that it remains intact for RT-PCR. The whole selection procedure stays the same except for 1) the substitution of NaOH denaturation by DNase treatment in template strand removal and 2) the substitution of DNA primer by RNA primer in TNA transcription (Figure 4.12A). Indeed, DNase treatment was very effective at eliminating contaminant, and the selected TNA molecules could not be amplified unless being reverse transcribed first (Figure 4.12B). The in vitro selection for RNA-cleaving TNA enzymes is currently in process: another graduate student Matthew R. Dunn has made significant progress in this project and showed appreciable enrichment for catalytic activity after six rounds of selection.

4.7. Methods

4.7.1. 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 ℃. The ^{32}P-labeled primer was annealed to the DNA template (See Table S1 in APPENDIX C) 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 ℃] by heating at 95 ℃ 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

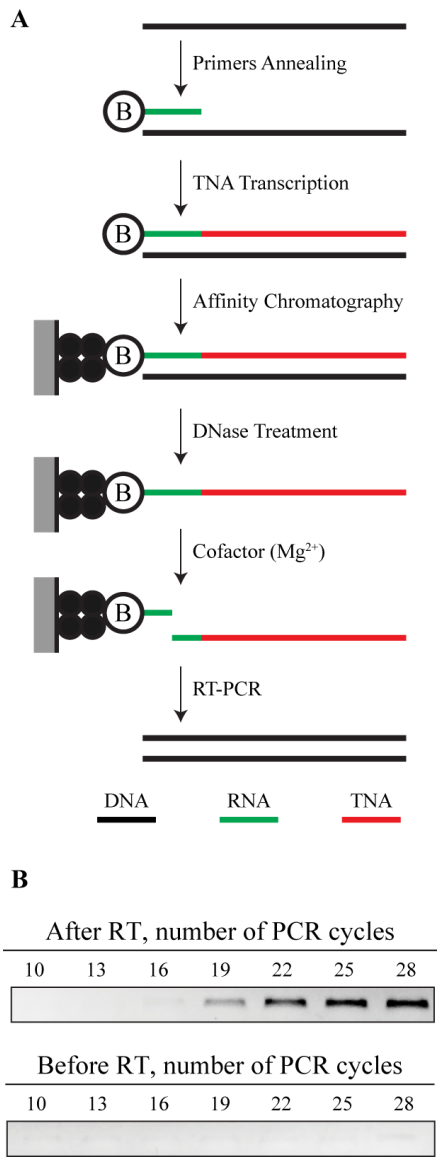


Figure 4.12. In vitro selection for RNA-cleaving TNA enzymes using DNase treatment to remove contaminant. (A) An RNA primer is used in TNA transcription and DNase is used to degrade DNA template. (B) The selected TNA could only be amplified after reverse transcription.

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 ℃. 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).

4.7.2. 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.

4.7.3. 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 ℃)] by heating at 95 ℃ 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 ℃. 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).

4.7.4. 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.

4.7.5. 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 ℃ 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 ℃. 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 ℃, and the digestion products were analyzed by 4% agarose gel electrophoresis.

4.7.6. Nuclease Stability Assay. DNA, RNA, and TNA oligonucleotide substrates (1 nmol) were incubated for up to 72 h at 37 ℃ 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.

4.7.7. 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.

4.8. 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. *Nat. Rev. Drug Discov.* **2010**, *9*, 537-550.

(7) Mayer, G. *Angew. Chem. Int. Ed.* **2009**, *48*, 2672-2689.

(8) Joyce, G. E. *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.; 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. *Science* **2012**, *336*, 341-344.

(17) Loakes, D.; Holliger, P. *Chem. Commun.* **2009**, 4619.

(18) Schoning, K.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347-1351.

(19) Orgel, L. *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. Y.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Res.* **2005**, *33*, 5219-5225.

(28) Horhota, A.; Zou, K. Y.; 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*, Unit 4.51.

(32) Zhang, S.; Yu, H.; Chaput, J. C. *Curr. Protoc. Nucleic Acid Chem.* **2013**, *4*, Unit 4.54.

(33) Wu, X. L.; 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.* **1993**, *90*, 804-808.

(35) Damha, M. J.; Wilds, C. J.; Noronha, A.; brunker, 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.* **1989**, *86*, 4076-4080.

(37) Brudno, Y.; Birnbaum, M. E.; Kleiner, R. E.; Liu, D. R. *Nat. Chem. Biol.* **2010**, *6*, 148-155.

(38) Wrenn, S. J.; Harbury, P. B. *Annu. Rev. Biochem.* **2007**, *76*, 331-349.

(39) Zamaratski, E.; Pradeepkumar, P. I.; Chattopadhyaya, J. *J. Biochem. Bioph. 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.* **1998**, *95*, 7933-7938.

(42) Rogers, J.; Joyce, G. F. *Nature* **1999**, *402*, 323-325.

(43) Santoro, S. W.; Joyce, G. F. *Proc. Natl. Acad. Sci.* **1997**, *94*, 4262-4266.

Chapter 5

Summary and Outlook

5.1. Conclusions

Nucleic acid aptamers have been considered a promising class of pharmaceuticals, and the small size and phosphodiester backbone places aptamers between small molecule drugs and antibody therapeutics. The advantages of aptamers include: 1) relative ease and low cost of identification and synthetic production, 2) versatile repertoire of various target recognitions and chemical modifications, and 3) non-immunogenicity to elicit innate immune responses. However, aptamers can not compete with the rivals in other aspects such as cellular uptake, serum stability and excretion.

Although aptamers have demonstrated their ability to target a broad range of molecules, the advances in molecular recognition mechanisms have been very uneven for different classes of targets. Considering the chemical nature of an aptamer is a polyanion with hydrophobic nucleobases, there should be several modes of interactions aptamers are capable of, such as electrostatic attraction, hydrogen bonding and π-π stacking. Therefore, the targets of aptamers can be roughly categorized as: small molecules, carbohydrates, nucleic acids and peptides/proteins. Perhaps the simplest mode of interactions is that of aptamers binding nucleic acid targets and the rules governing such interactions are relatively well understood. Small molecule binding aptamers have been identified by in vitro selection[1] and also exist naturally as riboswitches.[2] Artificial allosteric riboswitches have also been constructed for theophylline or tetracycline induced RNA splicing.[3,4] Aptamers bind these hydrophobic small molecules by forming three way junctions and projecting the helix-terminal base pairs towards the central cavity

to create a hydrophobic binding pocket.[5] While providing opportunities for hydrogen bonding, carbohydrates represent poor targets for aptamer selection due to their inherent flexibility and absence of aromatic ring structures or positive charges in most cases. Consequently the majority of carbohydrate binding aptamers show affinities in the micrmolar range.[6]

A lot of aptamers have been generated towards various peptide and protein targets, with the best known examples of VEGF-binding RNA aptamer (Macugen)[7] and thrombin-binding DNA aptamers.[8] A common perception is that positively charged proteins elicit high affinity binding of aptamers but undermine selectivity due to nonspecific electrostatic attraction. The first project in this dissertation evaluated the question regarding aptamer specificity towards positively charged proteins and demonstrated that aptamers could readily discriminate related alkaline proteins.[9] An inverse scenario, which is perhaps more challenging, is the selection of aptamers towards negatively charged proteins. This is exemplified in the aptamer selection towards the cancer relevant proteins in epidermal growth factor receptor family (EGFR or ErbB). ErbB2 has resisted multiple attempts to select aptamers in large part due to its abundant negative surface charges. This challenge has recently been overcome by the use of a small extracellular domain of the receptor, which is near neutral and offers suitable bait for the selection of 2'-fluoro-modified RNA aptamers with K_d of 3.5 nM.[10] An alternative approach for selecting aptamers towards negatively charged targets is the use of nucleotides modified with positively charged functiona groups during the selection. Examples include the use of aminohexyl modified dTTP derivative in the selection for

Sialyllactose[11] and the use of arginine-like dUTP analog in the selection for glutamic acid.[12,13]

Given nucleic acid aptamers differ fundamentally from proteins in the inherent chemical complexity of their building blocks, aptamers have been challenged for the ability to rival proteins in functional potentials. A recent example involves the quest for an equivalent nucleic acid counterpart of green fluorescent protein (GFP).[14] In this case, RNA aptamers were selected to bind GFP-like fluorophores and screened for binding induced fluorescence. Remarkably, these selections yielded multiple aptamer-fluorophore pairs and altogether the aptamer-activated fluorescence almost spanned the entire visible spectrum. One RNA-fluorophore complex, which exhibits green fluorescence and therefore is termed "Spinach", are excellent mimics of GFP and can be imaged in living cells.[14] Additionally, Spinach has been engineered to fuse with other aptamers including *S*-adenosyl methionine (SAM) and adenosine diphosphate (ADP) to create allosteric sensors, in which ligand binding stabilizes Spinach and induces increased fluorescence. These RNAs have been used to monitor the SAM and ADP levels in *E. coli*.[15]

While in vitro selection has been used to isolate functional DNA and RNA molecules for over 20 years, expanding this technology to unnatural genetic polymers remains a challenging problem. There are two major obstacles: 1) unnatural monomer building blocks are often not commercially available and 2) polymerases that accept modified nucleic acid analogs do not exist naturally.[16,17] These problems are more severe for polymers with backbone modifications such as TNA. The second project cleared a key hurdle that hindered the efficient enzymatic synthesis of a TNA library from a DNA library and enabled in vitro selection for TNA.[18] Without a proper "reverse transcriptase"

for TNA in place, DNA display offered a brilliant bypass strategy to select TNA receptors for thrombin.[19] The demonstration that TNA could form stable tertiary structures with specific ligand binding capacity provided experimental evidence to support TNA as a primordial genetic polymer. This fact was not entirely clear at the beginning of the study because one can imagine that TNA polymer, which contains a shorter backbone repeat unit, might be too stiff to fold into any secondary or tertiary structures. The fact that TNA does not seem to be limited in this regard, together with a similar demonstration that another synthetic genetic polymer HNA (hexitol nucleic acid) can also form aptamer structures,[20] suggested that the ability to inherit and evolve and the ability to fold and function is not restricted to natural RNA and DNA. Artificial genetic polymers represent promising alternatives to natural nucleic acids in a lot of aspects.

Binding is often just the first step in catalysis, and the latter is a prerequisite for any primitive genetic polymer in early organisms. In order to answer the big question of whether TNA can fold into structures with catalytic activities, we initiated the third project aiming to develop an in vitro replication system for TNA.[21] Since in vitro selection for nucleic acid enzymes usually involves self-tagging or self-cleaving reactions,[22,23] the DNA display strategy used in aptamer selection can not be readily adapted to an enzyme selection. Instead, a complete TNA replication system composed of transcription and reverse transcription processes is required to select any TNA enzymes. In the third project, we established an in vitro replication system for TNA using commercial enzymes, and assessed its feasibility to support TNA evolution in terms of efficiency, fidelity and enrichment factor under certain selection pressure.[21] The

development of this system enables future in vitro selection for TNA molecules with more sophisticated functions such as catalysis.

5.2. Future Directions

An in vitro selection for RNA-cleaving TNA enzymes is currently underway. We chose RNA hydrolysis as the target chemical reaction because many naturally occurring ribozymes catalyze this reaction such as hammerhead ribozymes[24,25] and glmS riboswitch.[26] In fact, the first natural ribozymes group I self-splicing intron and RNase P, discovered by Cech and Altman, are RNA-cleaving enzymes.[27,28] Moreover, many RNA-cleaving ribozymes and deoxyribozymes have been identified by in vitro selection in the laboratory[29,30] and these are excellent references for future TNA enzymes comparison.

An RNA-cleaving TNA enzyme might have been used as a defense mechanism by early TNA-organisms to fight against genetic takeover by RNA, but failed to survive the battle for unknown reasons. A more relevant enzyme in a hypothetical TNA world is self-replicase, a TNA-dependent TNA polymerase that is capable of copying itself into multiple copies. Such enzyme is essential for primitive life forms that use TNA as genetic information carrier and catalyst, and offers exceptional evolutionary advantage to the host organisms. Ideally, the identification of enzymes with such activity would greatly support a TNA world hypothesis. However, in vitro selection for TNA replicases is unusually difficult, if not impossible at all. Nucleic acid enzymes with such complicated activities would be extraordinarily large, at least hundreds of residues. For example, the most productive ribozyme polymerase is almost 200-nucleotide long.[31] The complexity of a randomized 200-mer region is $4^{200} = 10^{120}$. On the other hand, an in vitro selection typically examines 10^{15} different molecules, a tiny sampling from the theoretical

sequence space. Although we do not know the exact frequency of a nucleic acid replicase appearing in an unbiased pool, it is expected to be far more rare than any in vitro selected aptamers or ribozymes given its functional complexity.

A more realistic enzymatic activity for in vitro TNA selection is ligase. A ligase is essentially a polymerase without processivity, meaning that a ligase catalyzes just one step of a polymerization reaction before dissociating from substrates. Besides the replication system mentioned above, in vitro selection for TNA ligases requires synthesis of TNA oligonucleotides, as substrates and template. Compared with RNA ligase ribozyme catalyzed reaction, TNA ligation is simpler in that the attacking substrate has only one hydroxyl group and does not raise regioselectivity issues as in RNA ligation. However, the triphosphate group being attacked in TNA is not as accessible as in RNA due to the absence of exocyclic methylene group. Moreover, TNA is expected to be not as catalytically versatile as RNA due to the lack of a free hydroxyl function. Consequently, in vitro selection for TNA ligases would be more challenging than RNA ligase selection. A seemingly logical starting point would be directed evolution of existing RNA ligase ribozymes[32] and gradual substitution of RNA components of this system with TNA backbones. Although functional RNAs almost inevitably lose their functional properties when constructed in DNA with the same sequence and vice versa, initiating the in vitro selection for TNA ligases with an existing catalytically active scaffold of RNA ligase might be beneficial and provide a faster route to TNA ligase identification than starting from an unbiased random library.

TNA and HNA are the first examples of in vitro evolution of artificial genetic polymers, but there are many other nucleic acid analogs with various modifications.

Collectively, these derivatives are termed XNA (xeno-nucleic acids). Synthetic genetics explores the structural and functional properties of XNA polymers by in vitro evolution, and is now expanding to include a growing list of structurally diverse XNA polymers (Figure 5.1). Very recently, Liu and co-workers demonstrated non-enzymatic translation of DNA into synthetic polymers that are structurally unrelated to nucleic acids.[33] In contrast to XNAs, which are still in the structural neighborhood of DNA and RNA, the system developed by Liu includes structurally diverse synthetic polymers such as α-D-peptides and β-peptides, greatly expanding the sequence space accessible by laboratory evolution.

In vitro genetics of synthetic polymers has implications in several disciplines such as exobiology, biomedicine and synthetic biology. For example, TNA has been considered a potential RNA progenitor during early evolution of life on Earth. However, the chemical simplicity of threose relative to ribose and the cross pairing property of TNA with RNA are not sufficient to support TNA as a pre-RNA genetic polymer. What is also required for a primordial genetic material is its capacity to function as a catalyst to promote metabolic reactions. In vitro evolution provides a powerful tool for the identification of functional XNA molecules. Study of this kind will help constrain models for evaluating a given synthetic genetic polymer for a particular chemical function.

Figure 5.1. XNA chemical space. The *x*-, *y*- and *z*-axes representing sugar, base and backbone modifications, respectively. Most currently described XNAs (except PNA) lie on the axes, comprising only modifications to one of the three 'variables'. More divergent phenotypes should become accessible through a fuller exploration of the XNA space, that is, the replication and evolution of XNAs comprising a combination of modifications to base, sugar and backbone. Grey spheres represent these possible variants with multiple modifications not yet described in the XNA space. From Pinheiro, V. B.; Holliger, P. *Curr. Opin. Chem. Biol.* **2012**, *16*, 245-252. Reprinted with permission from Elsevier.

In addition to answering fundamental questions related to life's alternative genetic materials, synthetic genetics also provides an opportunity to develop nuclease-resistant

nucleic acid tools for molecular medicine. Natural aptamers and enzymes have proven their versatility in the chemical capacity, but their applications are limited by the susceptibility to nuclease-mediated degradation. Sometimes this problem can be overcome by post-selection modifications, but it is often accompanied by undesired consequences such as destabilization of the tertiary structure and interference of the original function. Synthetic genetics offers an alternative to produce intrinsic bio-stable functional polymers, which are invisible to the nucleases in complex biological environment.

Some artificial genetic polymers could also find potential use in future synthetic biology projects. Although synthetic biologists have been reprogramming cellular processes to create genetically modified organisms that are useful and beneficial to us, this engineering process is not easy because the metabolic pathways are often not independent but involve extensive cross communication and form inter-connected matrix-like network. Consequently, attempts to rewire one pathway are often supplemented by necessary modifications on several related genes, and still may lead to unexpected outcomes due to cellular homeostasis, genetic redundancy, epistasis and pleiotropy. Another concern associated with synthetic biology is safety issues. The society always fear that a pathogenic bacterial or virus strain designed for research purposes might escape into wild and cause disastrous consequences. Artificial genetic polymers offer potential solutions to these concerns because some of they are orthogonal to the natural biological machinery, which means they could be designed as a separate system and minimally interact with the natural system. Even if they fled out of the laboratory, they would not be able to survive and pose any threat to society because XNA

substrates and enzymes do not exist naturally.[34] This molecular orthogonality essentially creates a genetic firewall between nature and synthetic biology.[35]

5.3. References

(1) Sassanfar, M.; Szostak, J. W. *Nature* **1993**, *364*, 550-553.

(2) Mandal, M.; Breaker, R. R. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 451-463.

(3) Kim, D. S.; Gusti, V.; Pillai, S. G.; Gaur, R. K. *RNA* **2005**, *11*, 1667-1677.

(4) Weigand, J. E.; Suess, B. *Nucleic Acids Res.* **2007**, *35*, 4179-4185.

(5) Stojanovic, M. N.; Worgall, T. S. *Curr. Opin. Chem. Biol.* **2010**, *14*, 751-757.

(6) Sun, W.; Du, L.; Li, M. *Curr. Pharm. Des.* **2010**, *16*, 2269-2278.

(7) Siddiqui, M. A.; Keating, G. M. *Drugs* **2005**, *65*, 1571-1577.

(8) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. *Nature* **1992**, *355*, 564-566.

(9) Yu, H.; Jiang, B.; Chaput, J. C. *ChemBioChem* **2011**, *12*, 2659-2666.

(10) Kim, M. Y.; Jeong, S. *Nucleic Acid Ther.* **2011**, *21*, 173-178.

(11) Masud, M. M.; Kuwahara, M.; Ozaki, H.; Sawai, H. *Biorg. Med. Chem.* **2004**, *12*, 1111-1120.

(12) Kuwahara, M.; Ohsawa, K.; Kasamatsu, T.; Shoji, A.; Sawai, H.; Ozaki, H. *Nucleic Acids Symp. Ser.* **2005**, 81-82.

(13) Ohsawa, K.; Kasamatsu, T.; Nagashima, J.; Hanawa, K.; Kuwahara, M.; Ozaki, H.; Sawai, H. *Anal. Sci.* **2008**, *24*, 167-172.

(14) Paige, J. S.; Wu, K. Y.; Jaffrey, S. R. *Science* **2011**, *333*, 642-646.

(15) Paige, J. S.; Nguyen-Duc, T.; Song, W.; Jaffrey, S. R. *Science* **2012**, *335*, 1194.

(16) Chaput, J. C.; Yu, H.; Zhang, S. *Chem. Biol.* **2012**, *19*, 1360-1371.

(17) Pinheiro, V. B.; Holliger, P. *Curr. Opin. Chem. Biol.* **2012**, *16*, 245-252.

(18) Yu, H.; Zhang, S.; Chaput, J. C. *Nat. Chem.* **2012**, *4*, 183-187.

(19) Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 2802-2803.

(20) 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. *Science* **2012**, *336*, 341-344.

(21) Yu, H.; Zhang, S.; Dunn, M. R.; Chaput, J. C. *J. Am. Chem. Soc.* **2013**, *135*, 3583-3591.

(22) Robertson, D. L.; Joyce, G. F. *Nature* **1990**, *344*, 467-468.

(23) Santoro, S. W.; Joyce, G. F. *Proc. Natl. Acad. Sci.* **1997**, *94*, 4262-4266.

(24) Forster, A. C.; Symons, R. H. *Cell* **1987**, *50*, 9-16.

(25) Forster, A. C.; Symons, R. H. *Cell* **1987**, *49*, 211-220.

(26) Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R. *Nature* **2004**, *428*, 281-286.

(27) Cech, T. R.; Zaug, A. J.; Grabowski, P. J. *Cell* **1981**, *27*, 487-496.

(28) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849-857.

(29) Tang, J.; Breaker, R. R. *Proc. Natl. Acad. Sci.* **2000**, *97*, 5784-5789.

(30) Breaker, R. R.; Joyce, G. F. *Chem. Biol.* **1994**, *1*, 223-229.

(31) Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. *Science* **2011**, *332*, 209-212.

(32) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411-1418.

(33) Niu, J.; Hili, R.; Liu, D. R. *Nat. Chem.* **2013**, *5*, 282-292.

(34) Herdewijn, P.; Marliere, P. *Chem. Biodivers.* **2009**, *6*, 791-808.

(35) Schmidt, M. *Bioessays* **2010**, *32*, 322-331.

Bibliography

Chapter 1 References

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

(2) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505-510.

(3) Ellington, A. D.; Szostak, J. W. *Nature* **1992**, *355*, 850-852.

(4) Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425-1429.

(5) Williams, B. A.; Lin, L.; Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2009**, *131*, 6330-6331.

(6) Robertson, D. L.; Joyce, G. F. *Nature* **1990**, *344*, 467-468.

(7) Breaker, R. R.; Joyce, G. F. *Chem. Biol.* **1994**, *1*, 223-229.

(8) Breaker, R. R.; Joyce, G. F. *Chem. Biol.* **1995**, *2*, 655-660.

(9) Santoro, S. W.; Joyce, G. F. *Proc. Natl. Acad. Sci.* **1997**, *94*, 4262-4266.

(10) Cruz, R. P.; Withers, J. B.; Li, Y. *Chem. Biol.* **2004**, *11*, 57-67.

(11) Schlosser, K.; Gu, J.; Sule, L.; Li, Y. *Nucleic Acids Res.* **2008**, *36*, 1472-1481.

(12) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611-647.

(13) Joyce, G. E. *Angew. Chem. Int. Ed.* **2007**, *46*, 6420-6436.

(14) Chaput, J. C.; Yu, H.; Zhang, S. *Chem. Biol.* **2012**, *19*, 1360-1371.

(15) Vaught, J. D.; Bock, C.; Carter, J.; Fitzwater, T.; Otis, M.; Schneider, D.; Rolando, J.; Waugh, S.; Wilcox, S. K.; Eaton, B. E. *J. Am. Chem. Soc.* **2010**, *132*, 4141-4151.

(16) Eschenmoser, A. *Science* **1999**, *284*, 2118-2124.

(17) Reck, F.; Wippo, H.; Kudick, R.; Bolli, M.; Ceulemans, G.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **1999**, *1*, 1531-1534.

(18) Schoning, K.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347-1351.

(19) Schoning, K. U.; Scholz, P.; Wu, X. L.; Guntha, S.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. *Helv. Chim. Acta* **2002**, *85*, 4111-4153.

(20) Orgel, L. *Science* **2000**, *290*, 1306-1307.

(21) Herdewijn, P. *Angew. Chem. Int. Ed.* **2001**, *40*, 2249-2251.

(22) Zou, K. Y.; Horhota, A.; Yu, B.; Szostak, J. W.; McLaughlin, L. W. *Org. Lett.* **2005**, *7*, 1485-1487.

(23) Wu, X. L.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **2002**, *4*, 1283-1286.

(24) Chaput, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 9274-9275.

(25) Ichida, J. K.; Horhota, A.; Zou, K. Y.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Res.* **2005**, *33*, 5219-5225.

(26) Wu, X.; Guntha, S.; Ferencic, M.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **2002**, *4*, 1279-1282.

(27) Heuberger, B. D.; Switzer, C. *Org. Lett.* **2006**, *8*, 5809-5811.

(28) Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. *J. Am. Chem. Soc.* **2002**, *124*, 13716-13721.

(29) Pallan, P. S.; Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. *Angew. Chem. Int. Ed.* **2003**, *42*, 5893-5895.

(30) Ebert, M. O.; Mang, C.; Krishnamurthy, R.; Eschenmoser, A.; Jaun, B. *J. Am. Chem. Soc.* **2008**, *130*, 15105-15115.

(31) Zhang, L. L.; Peritz, A.; Meggers, E. *J. Am. Chem. Soc.* **2005**, *127*, 4174-4175.

(32) Yang, Y.-W.; Zhang, S.; McCullum, E. O.; Chaput, J. C. *J. Mol. Evol.* **2007**, *65*, 289-295.

(33) Chaput, J. C.; Ichida, J. K.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 856-857.

(34) Brown, J. A.; Suo, Z. *Biochemistry* **2011**, *50*, 1135-1142.

(35) Kempeneers, V.; Vastmans, K.; Rozenski, J.; Herdewijn, P. *Nucleic Acids Res.* **2003**, *31*, 6221-6226.

(36) Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 2802-2803.

(37) Horhota, A.; Zou, K. Y.; Ichida, J. K.; Yu, B.; McLaughlin, L. W.; Szostak, J. W.; Chaput, J. C. *J. Am. Chem. Soc.* **2005**, *127*, 7427-7434.

(38) Gardner, A. F.; Jack, W. E. *Nucleic Acids Res.* **2002**, *30*, 605-613.

(39) Yu, H.; Zhang, S.; Chaput, J. C. *Nat. Chem.* **2012**, *4*, 183-187.

(40) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakked, Z.; Sklenar, H.; Suzuki, M.; Tung, C. S.; Westhof, E.; Wolberger, C.; Berman, H. M. *J. Mol. Biol.* **2001**, *313*, 229-237.

Chapter 2 References

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

(2) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505-510.

(3) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611-647.

(4) Gold, L.; Polisky, B.; Uhlenbeck, O.; Yarus, M. *Annu. Rev. Biochem.* **1995**, *64*, 763-797.

(5) Bittker, J. A.; Phillips, K. J.; Liu, D. R. *Curr. Opin. Chem. Biol.* **2002**, *6*, 367-374.

(6) Famulok, M.; Mayer, G.; Blind, M. *Acc. Chem. Res.* **2000**, *33*, 591-599.

(7) Mayer, G. *Angew. Chem. Int. Ed.* **2009**, *48*, 2672-2689.

(8) Bunka, D. H.; Stockley, P. G. *Nat. Rev. Microbiol.* **2006**, *4*, 588-596.

(9) Jayasena, S. D. *Clin. Chem.* **1999**, *45*, 1628-1650.

(10) Jenison, R. D.; Gill, S. C.; Pardi, A.; Polisky, B. *Science* **1994**, *263*, 1425-1429.

(11) Williams, B. A.; Lin, L.; Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2009**, *131*, 6330-6331.

(12) Mukhopadhyay, R. *Anal. Chem.* **2005**, *77*, 114A-118A.

(13) Famulok, M.; Hartig, J. S.; Mayer, G. *Chem. Rev.* **2007**, *107*, 3715-3743.

(14) Drolet, D. W.; Moon-McDermott, L.; Romig, T. S. *Nat. Biotechnol.* **1996**, *14*, 1021-1025.

(15) Murphy, M. B.; Fuller, S. T.; Richardson, P. M.; Doyle, S. A. *Nucleic Acids Res.* **2003**, *31*, e110.

(16) Bock, C.; Coleman, M.; Collins, B.; Davis, J.; Foulds, G.; Gold, L.; Greef, C.; Heil, J.; Heilig, J. S.; Hicke, B.; Hurst, M. N.; Husar, G. M.; Miller, D.; Ostroff, R.; Petach, H.; Schneider, D.; Vant-Hull, B.; Waugh, S.; Weiss, A.; Wilcox, S. K.; Zichi, D. *Proteomics* **2004**, *4*, 609-618.

(17) Romig, T. S.; Bell, C.; Drolet, D. W. *J. Chromatogr. B. Biomed. Sci. Appl.* **1999**, *731*, 275-284.

(18) Tombelli, S.; Minunni, M.; Mascini, M. *Biosens. Bioelectron.* **2005**, *20*, 2424-2434.

(19) Liu, J.; Mazumdar, D.; Lu, Y. *Angew. Chem. Int. Ed.* **2006**, *45*, 7955-7959.

(20) Wang, Z.; Lee, J. H.; Lu, Y. *Adv. Mater.* **2008**, *20*, 3263-3267.

(21) White, R. R.; Sullenger, B. A.; Rusconi, C. P. *J. Clin. Invest.* **2000**, *106*, 929-934.

(22) Siddiqui, M. A.; Keating, G. M. *Drugs* **2005**, *65*, 1571-1577.

(23) Patel, D. J.; Suri, A. K. *J. Biotechnol.* **2000**, *74*, 39-60.

(24) Macaya, R. F.; Schultze, P.; Smith, F. W.; Roe, J. A.; Feigon, J. *Proc. Natl. Acad. Sci.* **1993**, *90*, 3745-3749.

(25) Long, S. B.; Long, M. B.; White, R. R.; Sullenger, B. A. *RNA* **2008**, *14*, 2504-2512.

(26) Taussig, M. J.; Stoevesandt, O.; Borrebaeck, C. A.; Bradbury, A. R.; Cahill, D.; Cambillau, C.; de Daruvar, A.; Dubel, S.; Eichler, J.; Frank, R.; Gibson, T. J.; Gloriam, D.; Gold, L.; Herberg, F. W.; Hermjakob, H.; Hoheisel, J. D.; Joos, T. O.; Kallioniemi, O.; Koegl, M.; Konthur, Z.; Korn, B.; Kremmer, E.; Krobitsch, S.; Landegren, U.; van der Maarel, S.; McCafferty, J.; Muyldermans, S.; Nygren, P. A.; Palcy, S.; Pluckthun, A.; Polic, B.; Przybylski, M.; Saviranta, P.; Sawyer, A.; Sherman, D. J.; Skerra, A.; Templin, M.; Ueffing, M.; Uhlen, M. *Nat. Methods* **2007**, *4*, 13-17.

(27) Lohman, T. M. *CRC Crit. Rev. Biochem.* **1986**, *19*, 191-245.

(28) Jayaram, B.; DiCapua, F. M.; Beveridge, D. L. *J. Am. Chem. Soc.* **1991**, *113*, 5211-5215.

(29) Luger, K.; Mader, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. *Nature* **1997**, *389*, 251-260.

(30) Lin, L.; Hom, D.; Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2007**, *129*, 14568-14569.

(31) Ramos, E.; Pineiro, D.; Soto, M.; Abanades, D. R.; Martin, M. E.; Salinas, M.; Gonzalez, V. M. *Lab. Invest.* **2007**, *87*, 409-416.

(32) Ramos, E.; Moreno, M.; Martin, M. E.; Soto, M.; Gonzalez, V. M. *Oligonucleotides* **2010**, *20*, 207-213.

(33) Berezovski, M. V.; Musheev, M. U.; Drabovich, A. P.; Jitkova, J. V.; Krylov, S. N. *Nat. Protoc.* **2006**, *1*, 1359-1369.

(34) Mendonsa, S. D.; Bowser, M. T. *J. Am. Chem. Soc.* **2004**, *126*, 20-21.

(35) Mendonsa, S. D.; Bowser, M. T. *J. Am. Chem. Soc.* **2005**, *127*, 9382-9383.

(36) Mosing, R. K.; Mendonsa, S. D.; Bowser, M. T. *Anal. Chem.* **2005**, *77*, 6107-6112.

(37) Berezovski, M.; Musheev, M.; Drabovich, A.; Krylov, S. N. *J. Am. Chem. Soc.* **2006**, *128*, 1410-1411.

(38) Tang, J.; Xie, J.; Shao, N.; Yan, Y. *Electrophoresis* **2006**, *27*, 1303-1311.

(39) Zuker, M. *Nucleic Acids Res.* **2003**, *31*, 3406-3415.

(40) Wong, I.; Lohman, T. M. *Proc. Natl. Acad. Sci.* **1993**, *90*, 5428-5432.

(41) Carothers, J. M.; Oestreich, S. C.; Davis, J. H.; Szostak, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 5130-5137.

(42) Sazani, P. L.; Larralde, R.; Szostak, J. W. *J. Am. Chem. Soc.* **2004**, *126*, 8370-8371.

(43) Eaton, B. E.; Gold, L.; Zichi, D. A. *Chem. Biol.* **1995**, *2*, 633-638.

(44) Carothers, J. M.; Oestreich, S. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2006**, *128*, 7929-7937.

(45) Jain, S. S.; Tullius, T. D. *Nat. Protoc.* **2008**, *3*, 1092-1100.

Chapter 3 References

(1) Eschenmoser, A. *Science* **1999**, *284*, 2118-2124.

(2) Benner, S. A.; Ellington, A. D.; Tauer, A. *Proc. Natl. Acad. Sci.* **1989**, *86*, 7054-7058.

(3) Schoning, K.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347-1351.

(4) Orgel, L. *Science* **2000**, *290*, 1306-1307.

(5) Ebert, M. O.; Mang, C.; Krishnamurthy, R.; Eschenmoser, A.; Jaun, B. *J. Am. Chem. Soc.* **2008**, *130*, 15105-15115.

(6) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611-647.

(7) Joyce, G. E. *Angew. Chem. Int. Ed.* **2007**, *46*, 6420-6436.

(8) Keefe, A. D.; Cload, S. T. *Curr. Opin. Chem. Biol.* **2008**, *12*, 448-456.

(9) Chaput, J. C.; Ichida, J. K.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 856-857.

(10) Chaput, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 9274-9275.

(11) Kempeneers, V.; Vastmans, K.; Rozenski, J.; Herdewijn, P. *Nucleic Acids Res.* **2003**, *31*, 6221-6226.

(12) Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 2802-2803.

(13) Horhota, A.; Zou, K. Y.; Ichida, J. K.; Yu, B.; McLaughlin, L. W.; Szostak, J. W.; Chaput, J. C. *J. Am. Chem. Soc.* **2005**, *127*, 7427-7434.

(14) Ichida, J. K.; Horhota, A.; Zou, K. Y.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Res.* **2005**, *33*, 5219-5225.

(15) Wu, X. L.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **2002**, *4*, 1283-1286.

(16) Roberts, R. W.; Szostak, J. W. *Proc. Natl. Acad. Sci.* **1997**, *94*, 12297-12302.

(17) Rosenbaum, D. M.; Liu, D. R. *J. Am. Chem. Soc.* **2003**, *125*, 13924-13925.

(18) Levy, M.; Miller, S. L. *Proc. Natl. Acad. Sci.* **1998**, *95*, 7933-7938.

(19) Rogers, J.; Joyce, G. F. *Nature* **1999**, *402*, 323-325.

(20) Lin, L.; Hom, D.; Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2007**, *129*, 14568-14569.

(21) Williams, B. A.; Lin, L.; Lindsay, S. M.; Chaput, J. C. *J. Am. Chem. Soc.* **2009**, *131*, 6330-6331.

(22) Mendonsa, S. D.; Bowser, M. T. *J. Am. Chem. Soc.* **2004**, *126*, 20-21.

(23) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. *Nature* **1992**, *355*, 564-566.

(24) Kubik, M. F.; Stephens, A. W.; Schneider, D.; Marlar, R. A.; Tasset, D. *Nucleic Acids Res.* **1994**, *22*, 2619-2626.

(25) Woese, C. R.; Harper & Row: 1967.

(26) Crick, F. H. *J. Mol. Biol.* **1968**, *38*, 367-379.

(27) Orgel, L. E. *J. Mol. Biol.* **1968**, *38*, 381-393.

(28) Gilbert, W. *Nature* **1986**, *319*, 618-618.

(29) Cech, T. R. *Cold Spring Harb Perspect Biol* doi: 10.1101/cshperspect.a006742.

(30) Joyce, G. F. *Nature* **2002**, *418*, 214-221.

(31) Yarus, M. *Cold Spring Harb Perspect Biol* **2011**, *3*, a003590.

(32) Joyce, G. F.; Schwartz, A. W.; Miller, S. L.; Orgel, L. E. *Proc. Natl. Acad. Sci.* **1987**, *84*, 4398-4402.

(33) Joyce, G. F. *New Biologist* **1991**, *3*, 399-407.

(34) Loakes, D.; Holliger, P. *Chem. Commun.* **2009**, 4619.

(35) 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. *Science* **2012**, *336*, 341-344.

(36) Zou, K. Y.; Horhota, A.; Yu, B.; Szostak, J. W.; McLaughlin, L. W. *Org. Lett.* **2005**, *7*, 1485-1487.

(37) Yu, H.; Jiang, B.; Chaput, J. C. *ChemBioChem* **2011**, *12*, 2659-2666.

Chapter 4 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. *Nat. Rev. Drug Discov.* **2010**, *9*, 537-550.

(7) Mayer, G. *Angew. Chem. Int. Ed.* **2009**, *48*, 2672-2689.

(8) Joyce, G. E. *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.; 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. *Science* **2012**, *336*, 341-344.

(17) Loakes, D.; Holliger, P. *Chem. Commun.* **2009**, 4619.

(18) Schoning, K.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347-1351.

(19) Orgel, L. *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. Y.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Res.* **2005**, *33*, 5219-5225.

(28) Horhota, A.; Zou, K. Y.; 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*, Unit 4.51.

(32) Zhang, S.; Yu, H.; Chaput, J. C. *Curr. Protoc. Nucleic Acid Chem.* **2013**, *4*, Unit 4.54.

(33) Wu, X. L.; 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.* **1993**, *90*, 804-808.

(35) Damha, M. J.; Wilds, C. J.; Noronha, A.; brunker, 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.* **1989**, *86*, 4076-4080.

(37) Brudno, Y.; Birnbaum, M. E.; Kleiner, R. E.; Liu, D. R. *Nat. Chem. Biol.* **2010**, *6*, 148-155.

(38) Wrenn, S. J.; Harbury, P. B. *Annu. Rev. Biochem.* **2007**, *76*, 331-349.

(39) Zamaratski, E.; Pradeepkumar, P. I.; Chattopadhyaya, J. *J. Biochem. Bioph. 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.* **1998**, *95*, 7933-7938.

(42) Rogers, J.; Joyce, G. F. *Nature* **1999**, *402*, 323-325.

(43) Santoro, S. W.; Joyce, G. F. *Proc. Natl. Acad. Sci.* **1997**, *94*, 4262-4266.

Chapter 5 References

(1) Sassanfar, M.; Szostak, J. W. *Nature* **1993**, *364*, 550-553.

(2) Mandal, M.; Breaker, R. R. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 451-463.

(3) Kim, D. S.; Gusti, V.; Pillai, S. G.; Gaur, R. K. *RNA* **2005**, *11*, 1667-1677.

(4) Weigand, J. E.; Suess, B. *Nucleic Acids Res.* **2007**, *35*, 4179-4185.

(5) Stojanovic, M. N.; Worgall, T. S. *Curr. Opin. Chem. Biol.* **2010**, *14*, 751-757.

(6) Sun, W.; Du, L.; Li, M. *Curr. Pharm. Des.* **2010**, *16*, 2269-2278.

(7) Siddiqui, M. A.; Keating, G. M. *Drugs* **2005**, *65*, 1571-1577.

(8) Bock, L. C.; Griffin, L. C.; Latham, J. A.; Vermaas, E. H.; Toole, J. J. *Nature* **1992**, *355*, 564-566.

(9) Yu, H.; Jiang, B.; Chaput, J. C. *ChemBioChem* **2011**, *12*, 2659-2666.

(10) Kim, M. Y.; Jeong, S. *Nucleic Acid Ther.* **2011**, *21*, 173-178.

(11) Masud, M. M.; Kuwahara, M.; Ozaki, H.; Sawai, H. *Biorg. Med. Chem.* **2004**, *12*, 1111-1120.

(12) Kuwahara, M.; Ohsawa, K.; Kasamatsu, T.; Shoji, A.; Sawai, H.; Ozaki, H. *Nucleic Acids Symp. Ser.* **2005**, 81-82.

(13) Ohsawa, K.; Kasamatsu, T.; Nagashima, J.; Hanawa, K.; Kuwahara, M.; Ozaki, H.; Sawai, H. *Anal. Sci.* **2008**, *24*, 167-172.

(14) Paige, J. S.; Wu, K. Y.; Jaffrey, S. R. *Science* **2011**, *333*, 642-646.

(15) Paige, J. S.; Nguyen-Duc, T.; Song, W.; Jaffrey, S. R. *Science* **2012**, *335*, 1194.

(16) Chaput, J. C.; Yu, H.; Zhang, S. *Chem. Biol.* **2012**, *19*, 1360-1371.

(17) Pinheiro, V. B.; Holliger, P. *Curr. Opin. Chem. Biol.* **2012**, *16*, 245-252.

(18) Yu, H.; Zhang, S.; Chaput, J. C. *Nat. Chem.* **2012**, *4*, 183-187.

(19) Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 2802-2803.

(20) 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. *Science* **2012**, *336*, 341-344.

(21) Yu, H.; Zhang, S.; Dunn, M. R.; Chaput, J. C. *J. Am. Chem. Soc.* **2013**, *135*, 3583-3591.

(22) Robertson, D. L.; Joyce, G. F. *Nature* **1990**, *344*, 467-468.

(23) Santoro, S. W.; Joyce, G. F. *Proc. Natl. Acad. Sci.* **1997**, *94*, 4262-4266.

(24) Forster, A. C.; Symons, R. H. *Cell* **1987**, *50*, 9-16.

(25) Forster, A. C.; Symons, R. H. *Cell* **1987**, *49*, 211-220.

(26) Winkler, W. C.; Nahvi, A.; Roth, A.; Collins, J. A.; Breaker, R. R. *Nature* **2004**, *428*, 281-286.

(27) Cech, T. R.; Zaug, A. J.; Grabowski, P. J. *Cell* **1981**, *27*, 487-496.

(28) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849-857.

(29) Tang, J.; Breaker, R. R. *Proc. Natl. Acad. Sci.* **2000**, *97*, 5784-5789.

(30) Breaker, R. R.; Joyce, G. F. *Chem. Biol.* **1994**, *1*, 223-229.

(31) Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. *Science* **2011**, *332*, 209-212.

(32) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411-1418.

(33) Niu, J.; Hili, R.; Liu, D. R. *Nat. Chem.* **2013**, *5*, 282-292.

(34) Herdewijn, P.; Marliere, P. *Chem. Biodivers.* **2009**, *6*, 791-808.

(35) Schmidt, M. *Bioessays* **2010**, *32*, 322-331.

APPENDIX A

SUPPLEMENTAL INFORMATION FOR CHAPTER 2

Supporting Information

Aptamers can Discriminate Alkaline Proteins with High Specificity

Hanyang Yu, Bing Jiang, and John C. Chaput*

Center for Evolutionary Medicine and Informatics, and Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-5301.

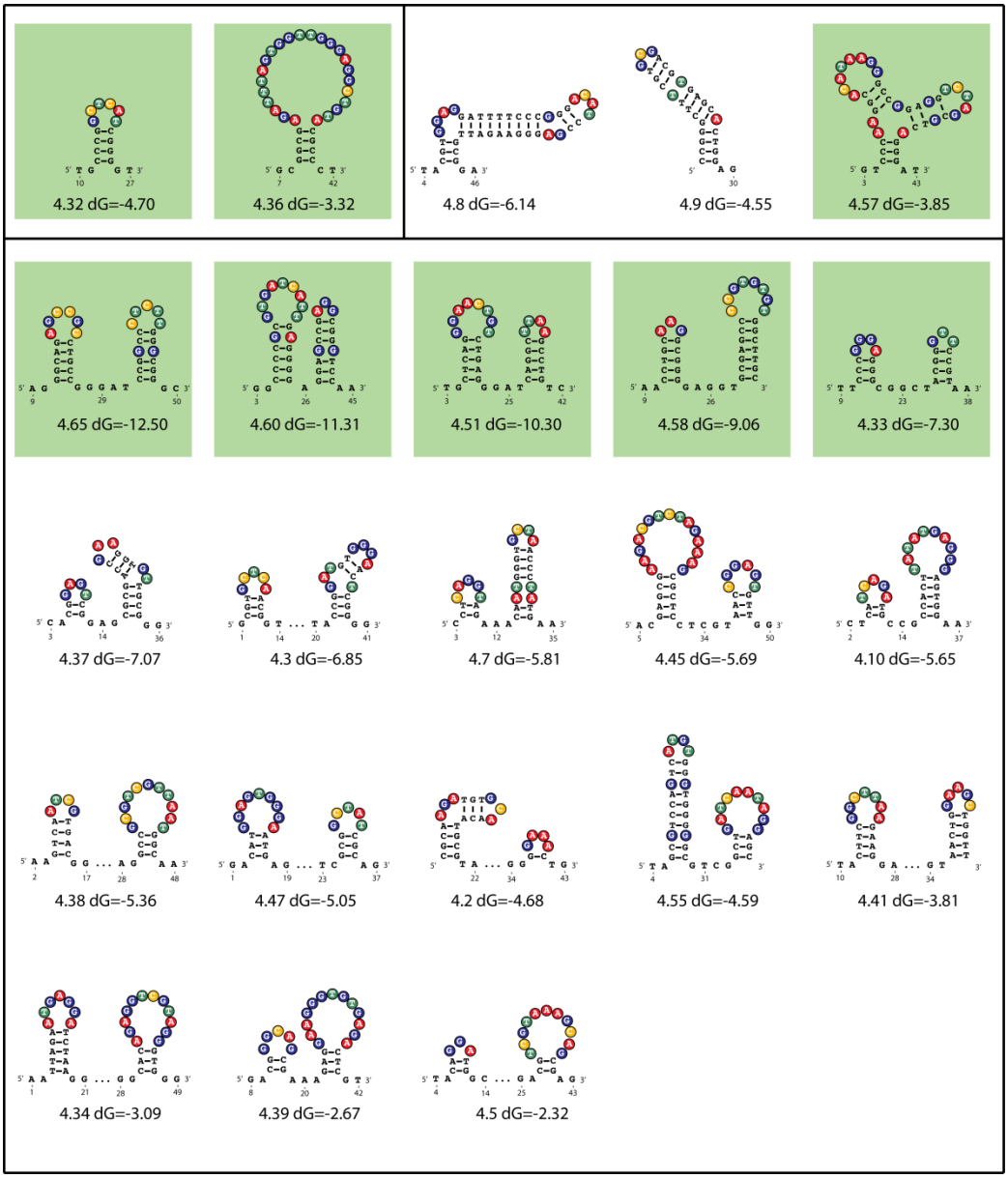


Figure S1. Predicted secondary structures and free energy values (kcal/mol) of 23 clones isolated after four iterative rounds of in vitro selection and amplification. Aptamers were folded under conditions that simulated a salt concentration of 100 mM NaCl and 5 mM $MgCl_2$, and a temperature of 25 ℃. With the exception of clones 4.32 and 4.36, which show simple stem-loop structures and clones 4.8, 4.9 and 4.57, which have stem-loop motifs modified with a small bulge, most of the clones (18 out of 23) adopt more complicated structures with tandem stem-loop motifs in their predicted secondary structure. Eight representative clones chosen for further analysis are highlighted in green.

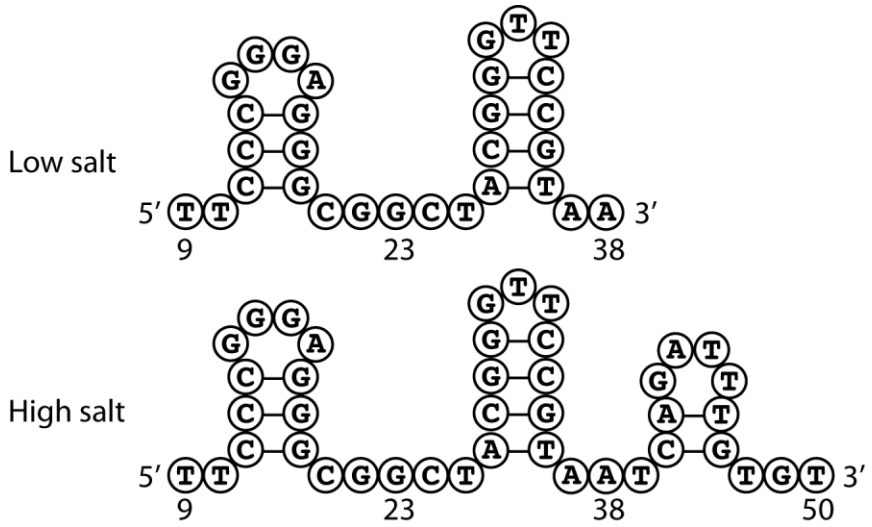


Figure S2. Predicted secondary structures (unstructured region not shown) of clone 4.33 calculated under simulated low salt (100 mM NaCl, 5 mM $MgCl_2$) and high salt (500 mM NaCl, 10 mM $MgCl_2$) conditions. When the concentrations of metal ions increase, clone 4.33 adopts a third stem-loop motif in its secondary structure.

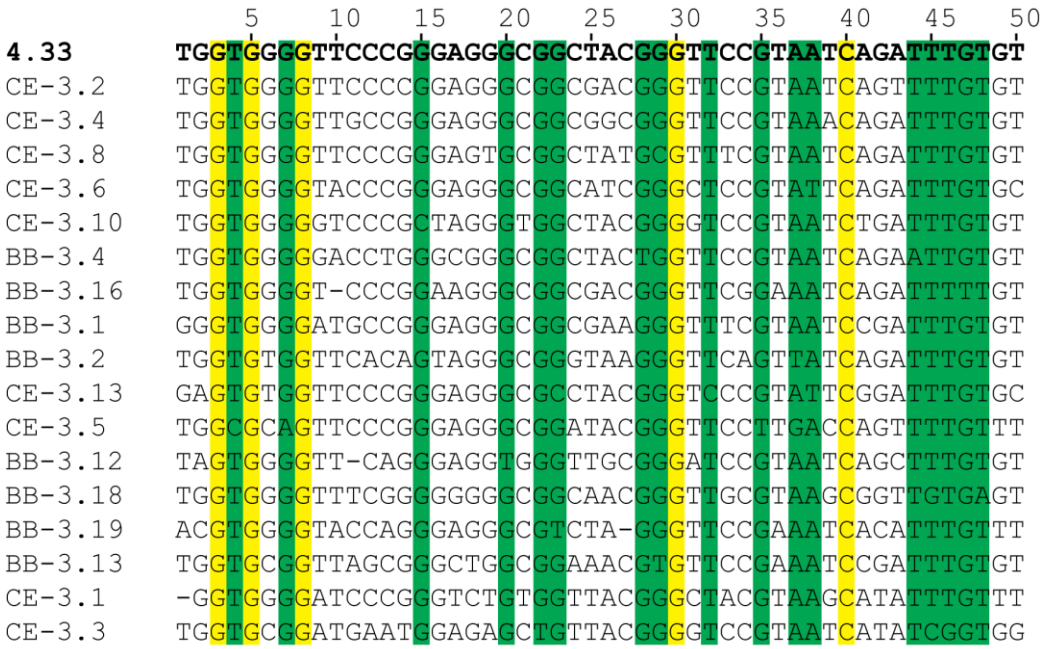


Figure S3. Sequence alignment of the clones generated by directed evolution of clone 4.33. The sequences are aligned in order of increasing mutations from top to bottom. Conserved nucleotides are highlighted in yellow, and semi-conserved bases are highlighted in green.

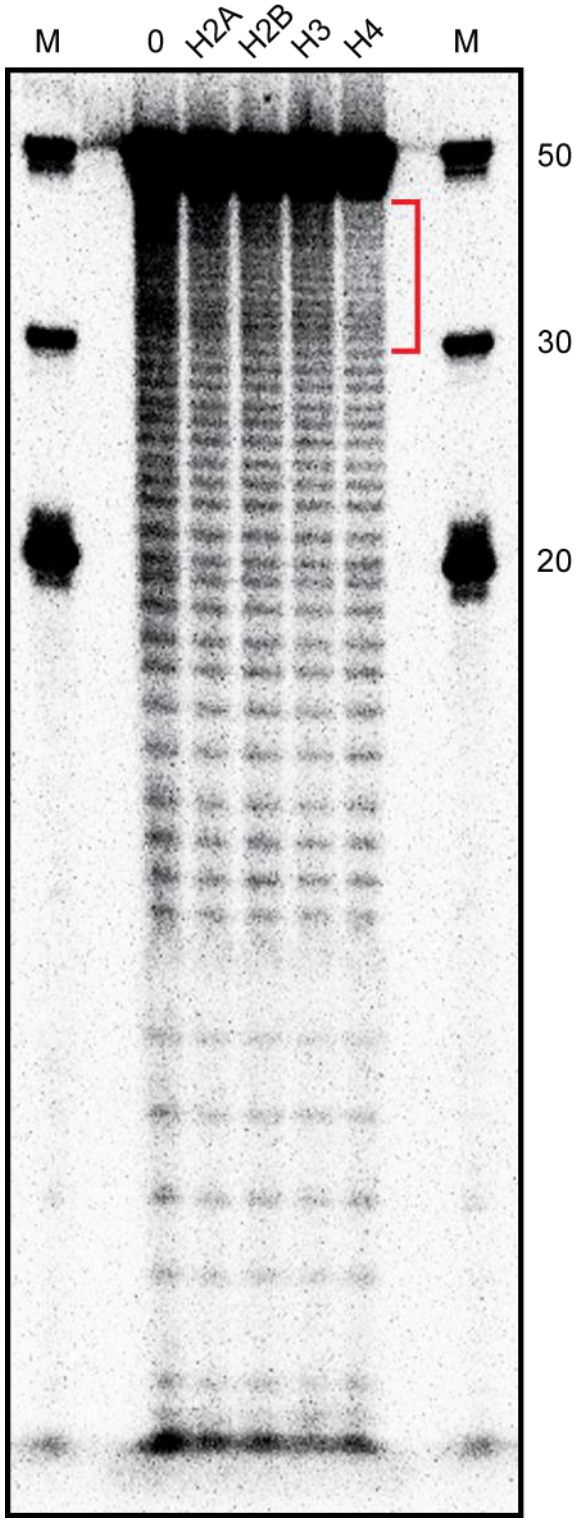


Figure S4. Footprinting of aptamer CE-3.13 complexed with and without histone proteins using hydroxyl radical. The protected region by the H4 protein is denoted by bracket on the right side of the H4 lane.

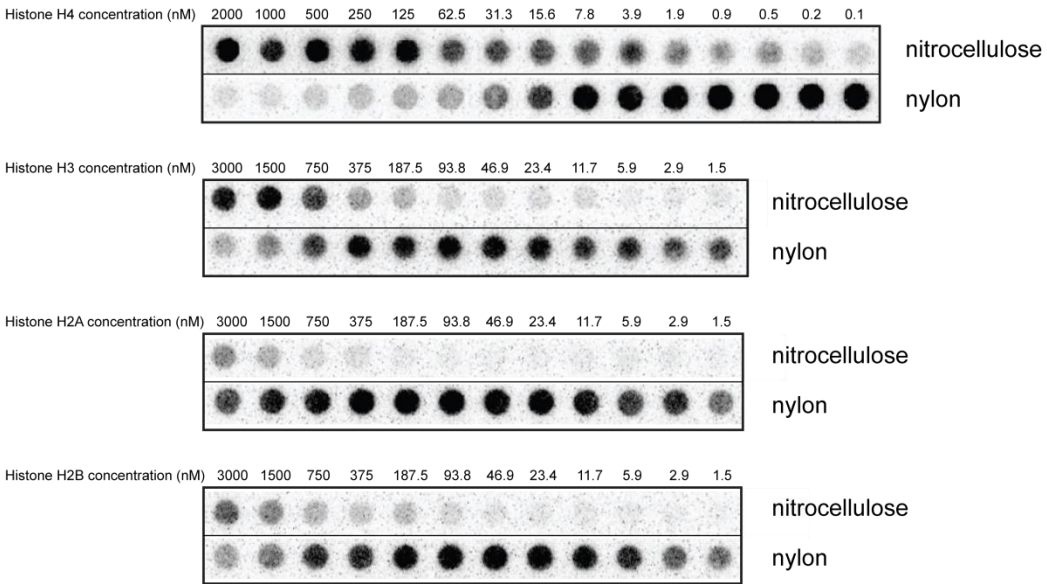


Figure S5. Dot blot filter binding asssay. For each histone protein, the top row is nitrocellulose membrane capturing protein-bound aptamer, and the bottom row is nylon membrane with free aptamer.

APPENDIX B

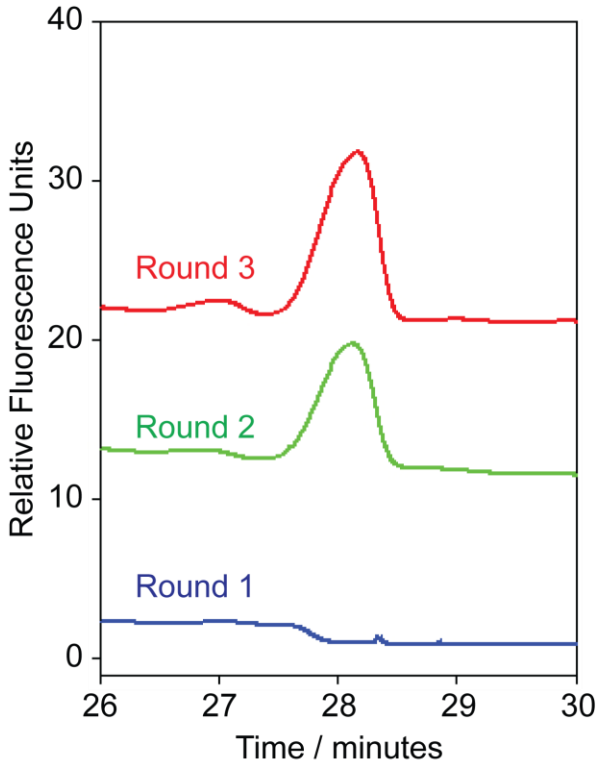
SUPPLEMENTAL INFORMATION FOR CHAPTER 3

Supplementary Information

Darwinian evolution of an alternative genetic system provides support for TNA as an RNA progenitor

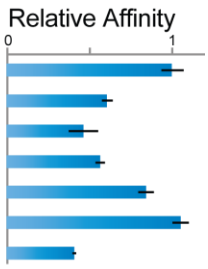
Hanyang Yu[†], Su Zhang[†] and John C. Chaput*

Center for Evolutionary Medicine and Informatics, and Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-5301.



Supplementary Figure 1 Enrichment of functional TNA molecules. A peak corresponding to the thrombin-bound TNA aptamers became visible in the electropherogram after iterative rounds of in vitro selection and amplification.

Truncation	Sequence (3'--2')
3.12	TGTTDTDGDDDDDDTDDTGGTGGGGGGTTTDGDTDDDGGGGTDTGDGDG
3.12-3t4	DTDGDDDDDDTDDTGGTGGGGGGTTTDGDTDDDGGGGTDTGDGDG
3.12-3t8	DDDDDDTDDTGGTGGGGGGTTTDGDTDDDGGGGTDTGDGDG
3.12-3t12	DDTDDTGGTGGGGGGTTTDGDTDDDGGGGTDTGDGDG
3.12-2t4	TGTTDTDGDDDDDDTDDTGGTGGGGGGTTTDGDTDDDGGGGTDTG
3.12-2t8	TGTTDTDGDDDDDDTDDTGGTGGGGGGTTTDGDTDDDGGGG
3.12-2t12	TGTTDTDGDDDDDDTDDTGGTGGGGGGTTTDGDTDDD



Supplementary Figure 2 Truncation analysis of TNA aptamer 3.12. Sequential deletion of four nucleotide segments from the 3' and 2' ends defined a core region of 41 nucleotides (3.12-2t8) that retained high affinity binding to human thrombin.

Supplementary Table 1 Sequence and Binding Affinity for TNA Aptamers.

Clone	Sequence (3'—2')	K_d (nM)
3.12	TGTTDTDGDDDDDDTDDTGGTGGGGGGTTTDGDT DDDGGGGTDTGDGDG	190
3.14	GGGGDGTGTGTCGGDDGTTGTGGGTTDGDTTDTT DGTGDDGTGDTGDDGG	250
3.9	DDGDDTTGGGGGDTTGGTDGGGDGDGDDGDGGTT DDDGTDDGGTDTDTD	320
3.11	TTGTTDDGTDTGDTGGTTTDTGTDDGGTGGDTDT TTTGDGTTGDGTGGGG	420
3.1	TTTTTGDTGTGGDGTCTGTTGGDTTGGDDGTTGT TTTTGTGDDGGGG	510
3.2	DGTGGDGTGTTDGTDDTGGTTTTDDGTGGGGTGD GDTGGGDDTTTGDTTT	590
3.10	GDTTGDGTGGDGDTGDTTTDDTTTGDGGTGTGTT GGDTTGTTGGG	590
3.5	DTGGTGDTTTGTTTDTGTGDDCDTGDTGTDTGDG DTTTTGGGGTGTTTTG	750
3.13	TGDGTDGDTTGGGDDTGTGTGDTTTGGGDTGTGT GGGGGTDDDTTDG	830
3.6	TGGGTDTDTGTTDGGDTDTDTGTTDDTTTDTDGG TTTGTGDGTGGTTTDG	880

APPENDIX C

SUPPLEMENTAL INFORMATION FOR CHAPTER 3

Supplemental Information

An Efficient and Faithful In vitro Replication System for Threose Nucleic Acid

Hanyang Yu, Su Zhang, Matthew 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.

General. DNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), purified by denaturing polyacrylamide gel electrophoresis, desalted on sephadex G-25 size exclusion column, concentrated, and quantified by UV absorbance. TNA triphosphates (tNTPs) and oligonucleotide were synthesized using standard methods.[1,2] RNA oligonucleotide S1 was purchased from Thermo Fisher Scientific (Lafayette, CO). The enzymes T4 polynucleotide kinase, calf intestinal alkaline phosphotase, Therminator™ DNA polymerase, Restriction enzyme HinfI and RNase H were purchased from New England Biolabs (Ipswich, MA). Other enzymes used in this study include, SuperScript™ II reverse transcriptase and AccuPrime™ High Fidelity Taq DNA Polymerase purchased from Invitrogen (Grand Island, NY); RQ 1 RNase-free DNase purchased from Promega (Madison, WI); and RNase A purchased from Sigma-Aldrich (St. Louis, MO). CloneJET™ PCR cloning kit was purchased from Fermentas (Waltham, MA).

Table S1: Primers and templates used in TNA transcription, reverse transcription and fidelity experiments.

TNA transcription primer P1 and its binding site in various templates are underlined. Primer P2, used in generation of TNA template for reverse transcription, is similar on its 3' end to P1 except for the internal mismatch reference shown in bold. P2 also has a 5' overhang, which is designed to be non-complementary to any region within the templates. TNA reverse transcription primer P3 and its counterpart in various templates are double underlined. Primer P3.ATG, used for fidelity determination of template 3NT.ATG, is composed of A, T and G only, and derived from P3 with all the C's changed to G's. PCR primer P4 is identical to the 5' overhang of P2 to avoid possible amplification from template contaminant. All the templates have a 50mer central region used for fidelity measurement, and a poly(dA) tail on the 3' end to facilitate separation on denaturing gel. Within the central 50mer region of interest, all the G residues are shown in red. In template 4NT.9G, every G residue is positioned in one of nine possible sequence environments (HGH, H = A, C, T). Template 4NT.9GA is derived from 4NT.9G with the modification that the position immediately downstream of every G residue is changed to A.

Name	Sequence
Primer P1	5′-GACACTCGTATGCAGTAGCC-3′
Primer P2	5′-CTTTTAAGAACCGGACGAACGACACTCGT**T**TGCAGTAGCC-3′
Primer P3	5′-TGTCTACACGCAAGCTTACA-3′
Primer P3.ATG	5′-TGTGTAGAGGGAAGGTTAGA-3'
Primer P4	5′-CTTTTAAGAACCGGACGAAC-3'
Template 4NT.8G	5′-TGTCTACACGCAAGCTTACAATGATCAACTCCGCAACTGTCATACCTGCG ACAAGACTACACGTTAAGCAGGCTACTGCATACGAGTGTCAAAAAAAAAA-3′
Template 4NT.3G	5′-TGTCTACACGCAAGCTTACACCATTCTTTAACAGTATCACTATATCCATT TACGAGTCAACATTAACCTCGGCTACTGCATACGAGTGTCAAAAAAAAAA-3′
Template 4NT.10G.1	5′-TGTCTACACGCAAGCTTACAAACTCCGTTACCGATTCAGCTTAGGATCCG ATCAAGCTTATGGTCCACGTGGCTACTGCATACGAGTGTCAAAAAAAAAA-3′
Template 4NT.10G.2	5′-TGTCTACACGCAAGCTTACAAACGCCATGACCTAGGCAACGTACAATGCT ATCGGCCTTGTAATCGACTTGGCTACTGCATACGAGTGTCAAAAAAAAAA-3′
Template 4NT.9G	5′-TGTCTACACGCAAGCTTACATTAAGACTCGCCATGTTACGATCTGCCAAG TACAGCCTTGAATCGTCACTGGCTACTGCATACGAGTGTCAAAAAAAAAA-3′
Template 4NT.9GA	5′-TGTCTACACGCAAGCTTACATTAAGACTCGACATGATACGATCTGACAAG AACAGACTTGAATCGACACTGGCTACTGCATACGAGTGTCAAAAAAAAAA-3′
Template 3NT.ATC	5′-TGTCTACACGCAAGCTTACAAACTCCATTACCTATTCAACTTACAATCCT ATCAACCTTATAATCCACTTGGCTACTGCATACGAGTGTCAAAAAAAAAA-3′
Template 3NT.ATG	5′-TGTGTAGAGGGAAGGTTAGATTAAGAATGGGTATGTTAGGATATGGTAAG TAGAGATTTGAATGGTTAGTGGCTACTGCATACGAGTGTCAAAAAAAAAA-3'
Library L2	5′-TGTCTACACGCAAGCTTACAH_{50}GGCTACTGCATACGAGTGTC-3′

Table S2: Oligonucleotide substrates used in nuclease digestion reactions.

Oligonucleotide substrates S1-S3 share the same sequence except that T is replaced by U in RNA substrate S1. Within the RNA template T1, the region complementary to DNA substrate S2 or TNA substrate S2 is underlined.

Name	Sequence
RNA substrate S1	5′-r(AAAAUUUAUUUAUUAA)-3′
DNA substrate S2	5′-d(AAAATTTATTTATTAA)-3′
TNA substrate S3	3′-t(AAAATTTATTTATTAA)-2′
RNA template T1	5′-r(GGGAGGAGGAUUACCCCUCGUUAAUAAAUAAAUUUUCUCUCGUGAUCGG GUAGCUGGACGCGACGGGUCC)-3′

Table S3: Fidelity of TNA replication under various transcription conditions.

Template	tNTP	dNTP	Bases Read	Substitution	Indel*	Total	Fidelity
4NT.3G	ATCG	None	1650	34.5×10^{-3}	1.2×10^{-3}	35.8×10^{-3}	96.4%
4NT.9G	ATCG	None	1250	32.8×10^{-3}	7.2×10^{-3}	40.0×10^{-3}	96.0%
4NT.9G	ATG	C	1150	3.5×10^{-3}	0.0×10^{-3}	3.5×10^{-3}	99.7%
4NT.9G	ATC	G	1400	22.1×10^{-3}	2.9×10^{-3}	25.0×10^{-3}	97.5%
4NT.9GA	ATCG	None	1300	9.2×10^{-3}	0.8×10^{-3}	10.0×10^{-3}	99.0%
3NT.ATG	ATC	None	1150	16.5×10^{-3}	1.7×10^{-3}	18.3×10^{-3}	98.2%
3NT.ATC	ATG	None	1050	1.9×10^{-3}	1.9×10^{-3}	3.8×10^{-3}	99.6%
4NT.9G	ATCG[#]	None	1500	18.0×10^{-3}	6.0×10^{-3}	24.0×10^{-3}	97.6%

* Indel: insertion + deletion.

[#] tATP (100 μM), tTTP (100 μM), tCTP (150 μM), tGTP (50 μM). Regular TNA transcription reaction contains 100 μM of each tNTP.

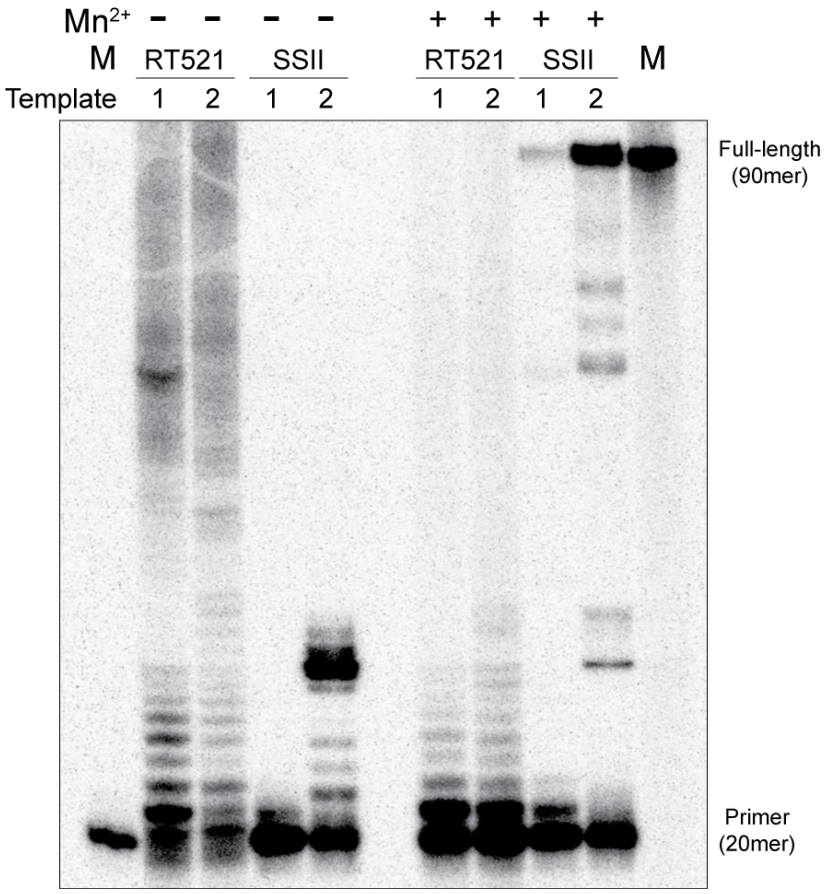


Figure S1: TNA reverse transcription by RT521 and SuperScript II (SSII) reverse transcriptase. TNA reverse transcription reactions were evaluated by challenging different enzymes to extend a DNA primer annealed to TNA template with dNTPs in the absence or presence of Mn^{2+}. RT521 could incorporate several monomers under optimal conditions[3] [100 nM primer-template complex, 1x ThermoPol buffer, 500 μM dNTPs, 1 mM $MgSO_4$ and 0.02 μg/μl RT521 enzyme. Incubation for 24 hours at 65°C] before pausing on the primer-template complex, and no discrete band for full-length product was observed. Also, manganese ions seemed to inhibit RT521's activity. SuperScript II reverse transcriptase could yield substantial amounts of full-length products in the

presence of $MnCl_2$.[4] TNA template 1 and 2 was synthesized on DNA templates 4NT.8G and 4NT.3G, respectively (Table S1). M: marker.

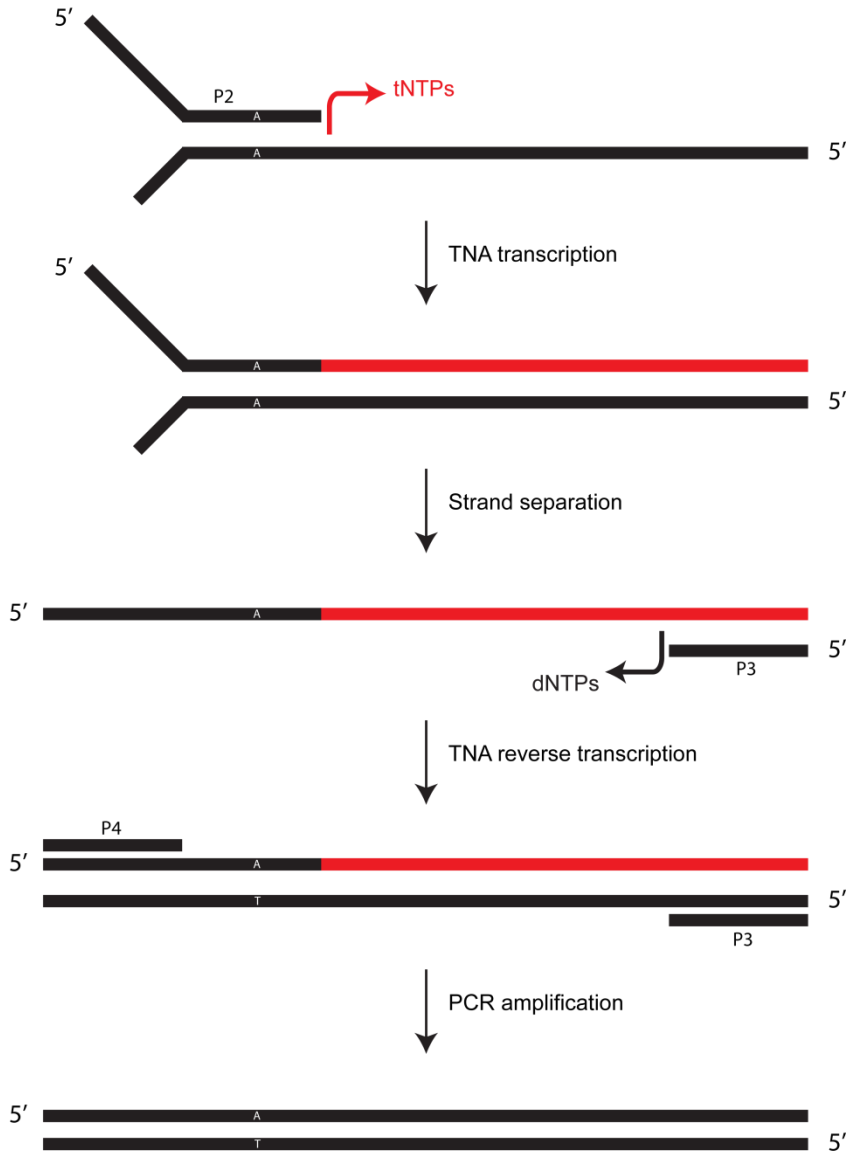


Figure S2: Schematic of fidelity measurement. Primer P2 (Table S1) is designed to have a non-complementary overhang on the 5' end and an internal A:A mismatch reference position[5]. After the primer is extended with tNTPs, the chimeric DNA-TNA strand is separated from DNA template strand by denaturing PAGE. The purified TNA sequence is used as template in reverse transcription reaction, which generates full-length cDNA strand that is amplified by PCR. One of the PCR primer (P4) shares the same sequence as the 5' overhang in P2 so that only full-length cDNA can be amplified even if

some original DNA template contaminant is present. After PCR amplification, the internal reference position has a T residue and can be ambiguously distinguished from the original DNA template sequence.

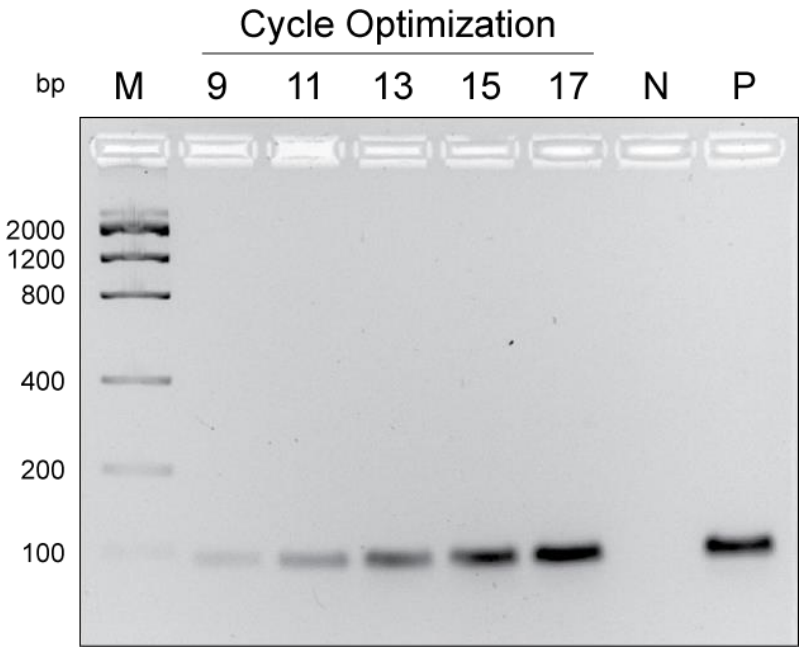


Figure S3: Amplification of cDNA after TNA reverse transcription. PCR amplification of cDNA generated in TNA reverse transcription was analyzed on 2% agarose gel. Cycle optimization showed exponential enrichment of cDNA sequences. N: negative control using purified TNA strand before reverse transcription as template. P: positive control using DNA library 3NT.ATC as template. M: low DNA mass ladder.

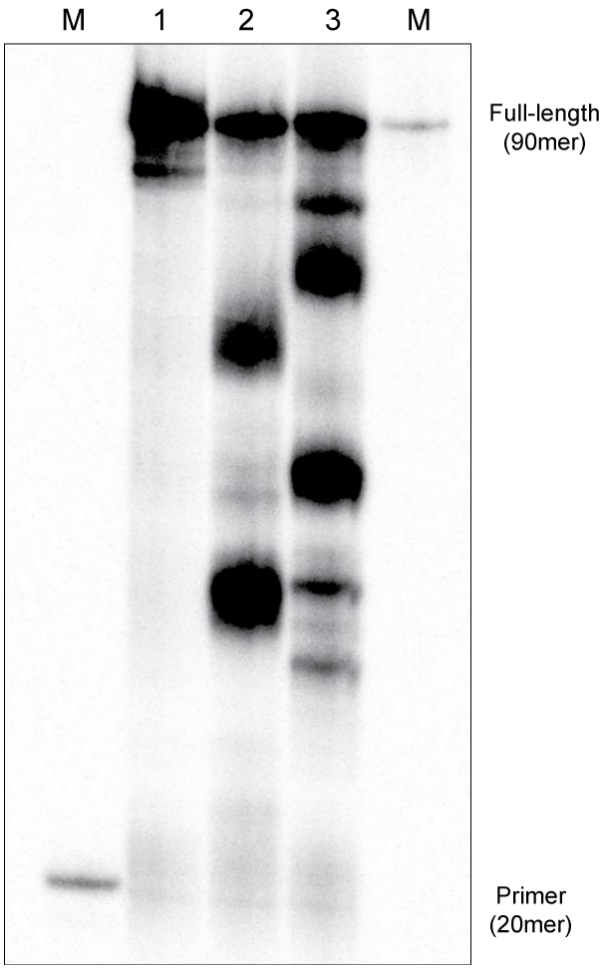


Figure S4: TNA transcription on templates containing continuous GG regions. TNA transcription reactions on different DNA templates were analyzed on 20% denaturing polyacrylamide gel. TNA transcription on DNA templates containing multiple GG repeats (DNA template 4NT.10G.1 and 4NT.10G.2 on lane 2 and 3, respectively) (Table S1) generated truncated sequences with lower yield in full-length products,[6] while DNA template devoid of G residues (3NT.ATC on lane 1) (Table S1) directed almost quantitative conversion of DNA primer to full-length DNA-TNA heteropolymer.

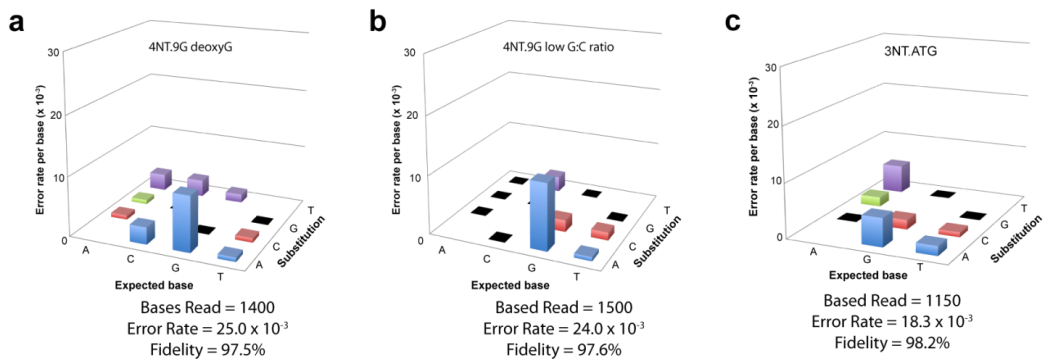


Figure S5: Substitution profile and overall fidelity of TNA replication under different conditions. TNA replication were examined under different conditions where G:G mispair was disfavored, such as **a**, using dGTP in place of tGTP during TNA transcription; **b**, decreasing tGTP:tCTP ratio during TNA transcription; and **c**, excluding tGTP during TNA transcription on a DNA template devoid of C (3NT.ATG in Table S1).

References

(1) Zhang, S.; Chaput, J. C. *Curr. Protoc. Nucleic Acid Chem.* **2012**, *4*, Unit 4.51.

(2) Zhang, S.; Yu, H.; Chaput, J. C. *Curr. Protoc. Nucleic Acid Chem.* **2013**, *4*, Unit 4.54.

(3) 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. *Science* **2012**, *336*, 341-344.

(4) Chaput, J. C.; Ichida, J. K.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 856-857.

(5) Ichida, J. K.; Horhota, A.; Zou, K.; McLaughlin, L. W.; Szostak, J. W. *Nucleic Acids Res.* **2005**, *33*, 5219-5225.

(6) Yu, H.; Zhang, S.; Chaput, J. C. *Nat. Chem.* **2012**, *4*, 183-187.

APPENDIX D

CO-AUTHOR APPROVAL

I verify that the following co-authors have approved of my use of our publications in my dissertation.

John C. Chaput (Arizona State University)

Su Zhang (Arizona State University)

Bing Jiang (Arizona State University)

Matthew R. Dunn (Arizona State University)