nature chemistry

ARTICLES

PUBLISHED ONLINE: 10 JANUARY 2012 | DOI: 10.1038/NCHEM.1241

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

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

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.

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, Fig. 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.

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

Center for Evolutionary Medicine and Informatics in the Biodesign Institute and Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-5301, USA; †These authors contributed equally to the project. *e-mail: john.chaput@asu.edu

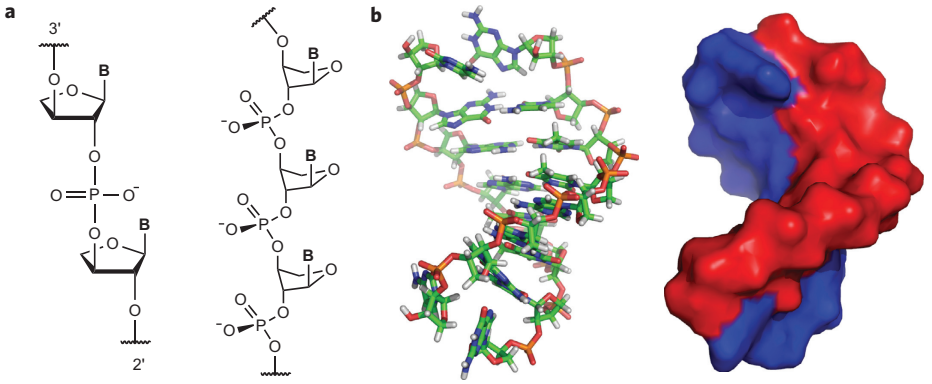

Figure 1 | Structure of TNA. a, Constitutional (left) and configurational (right) structures for the linearized backbone of an α-L-threofuranosyl-(3′ → 2′)-oligonucleotide. TNA contains one less atom per backbone repeat unit than natural RNA and DNA. **B** represents a nucleotase. **b**, Solution NMR 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.

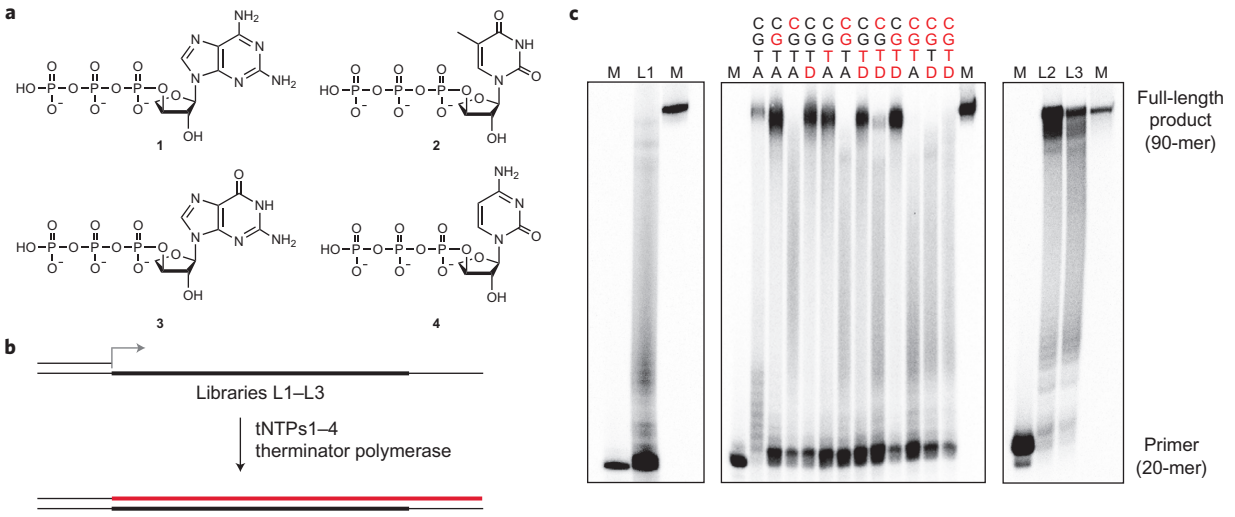

Figure 2 | Synthesis of TNA libraries by enzyme-mediated primer extension. a, Chemical structure of TNA triphosphates (tNTPs): tDTP, **1**; tTTP, **2**; tGTP, **3**; tCTP, **4**. **b**, Schematic diagram of the primer-extension assay used to evaluate DNA libraries L1–L3. Each library contains a central 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.

downstream amplification using the polymerase chain reaction (PCR). We synthesized all four TNA triphosphates (tNTPs, Fig. 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 (Fig. 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 (Fig. 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 (Fig. 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.

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 (Fig. 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

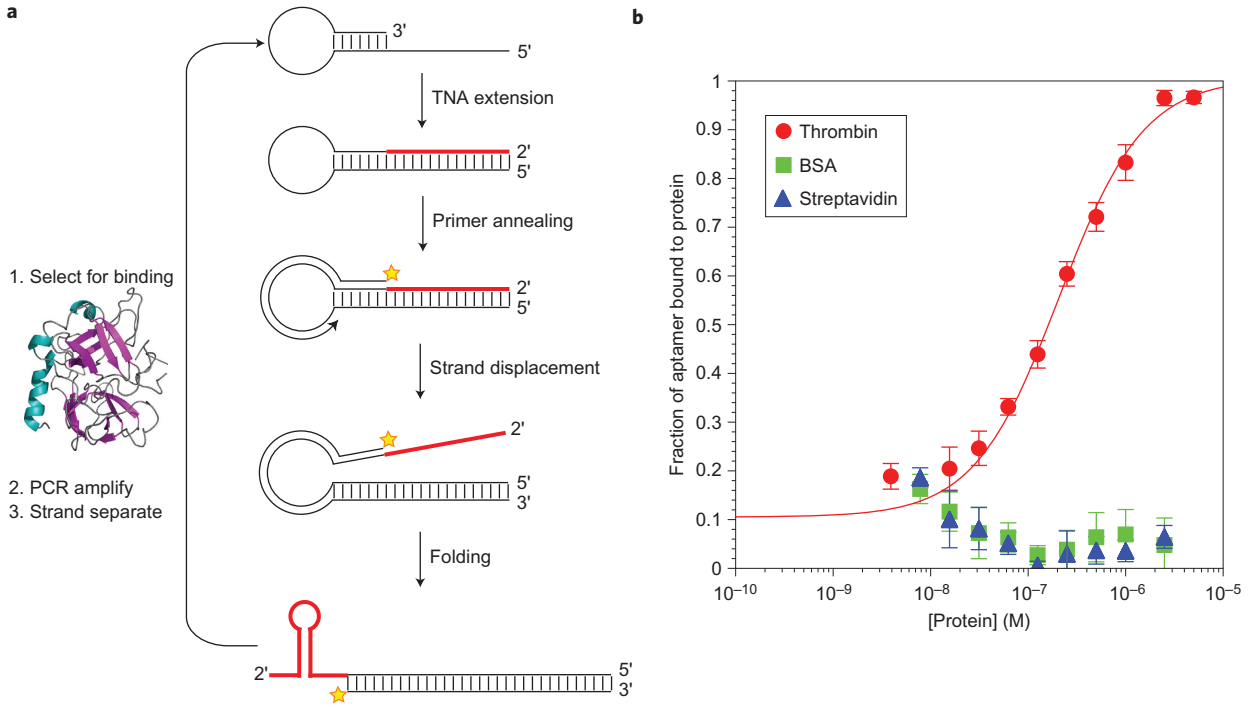

Figure 3 | Evolution of TNA receptors *in vitro*. a, *In vitro* selection strategy designed to isolate TNA aptamers with affinity to human thrombin. The DNA library encodes a random region of 50 nt positions flanked on the 3′ end with a stem-loop structure that serves as a DNA primer and a fixed-sequence 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′-TGTTDTDGDDDDDDTDDT GGTGGGGGGTTTDGDTDDDGGGG-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 ($n = 3$).

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 (Fig. 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 °C)[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 ($\sim 1 \times 10^{11}$ 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 (Supplementary Fig. S1), 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 (Supplementary Table S1), which is similar to previously evolved DNA and RNA aptamers that bind to thrombin[23,24].

We determined the minimum binding region for the highest-affinity aptamer (clone 3.12, $K_d \approx 200$ nM at 25 °C) 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 (Fig. 3b, Supplementary Fig. S2). 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 (Supplementary Fig. S2), 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 (Fig. 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 (ref. 34; and V.B. Pinheiro *et al.*, manuscript in preparation), future examination of structurally diverse genetic systems by *in vitro* evolution will require similar advances in nucleic acid chemistry.

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.

Methods

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 °C, butanol precipitated, and purified by denaturing polyacrylamide gel electrophoresis. TNA triphosphates were synthesized as previously described[35].

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 °C. ^{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 °C 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 °C. Primer extension products were analysed by 20% denaturing polyacrylamide gel electrophoresis.

***In vitro* selection.** For each round of selection, the self-priming DNA library L2 (5′-TGTCTACACGCAAGCTTACA-N_{50}-GGCTACTGCATACGAGTGTCa ctacgtacccacaacctcggccgtaccacggtacgtagtgacactcgtatgcagtagcc; 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/chloroform extracted, desalted and subjected to a strand invasion step. A DNA primer (5′-FAM-AAGGCTACTGCATACGAGTGTCACTACGTACCG TGGTACGGCCGAGGTTG-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 °C, followed by an additional heating step of 90 min at 65 °C. 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 °C. 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 °C. 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.

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 °C 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 °C. 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[36].

Received 30 September 2011; accepted 29 November 2011; published online 10 January 2012

References

1. Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science* **284,** 2118–2124 (1999).
2. Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl Acad. Sci. USA* **86,** 7054–7058 (1989).
3. Schoning, K-U. *et al.* Chemical etiology of nucleic acid structure: the α-threofuranosyl-(3′-2′) oligonucleotide system. *Science* **290,** 1347–1351 (2000).
4. Orgel, L. E. A simpler nucleic acid. *Science* **290,** 1306–1307 (2000).
5. Ebert, M-O., Mang, C., Krishnamurthy, R., Eschenmoser, A. & Jaun, B. The structure of a TNA–TNA complex in solution: NMR study of the octamer duplex derived from α-(L)-threofuranosyl-(3′-2′)-CGAATTCG. *J. Am. Chem. Soc.* **130,** 15105–15115 (2008).
6. Wilson, D. S. & Szostak, J. W. *In vitro* selection of functional nucleic acids. *Annu. Rev. Biochem.* **68,** 611–647 (1999).
7. Joyce, G. F. Forty years of *in vitro* evolution. *Angew. Chem. Int. Ed.* **46,** 6420–6436 (2007).
8. Keefe, A. D. & Cload, S. T. SELEX with modified nucleotides. *Curr. Opin. Chem. Biol.* **12,** 448–456 (2008).
9. Chaput, J. C., Ichida, J. K. & Szostak, J. W. DNA polymerase-mediated DNA synthesis on a TNA template. *J. Am. Chem. Soc.* **125,** 856–857 (2003).
10. Chaput, J. C. & Szostak, J. W. TNA synthesis by DNA polymerases. *J. Am. Chem. Soc.* **125,** 9274–9275 (2003).
11. Kempeneers, V., Vastmans, K., Rozenski, J. & Herdewijn, P. Recognition of threosyl nucleotides by DNA and RNA polymerases. *Nucleic Acids Res.* **31,** 6221–6226 (2003).
12. Ichida, J. K. *et al.* An *in vitro* selection system for TNA. *J. Am. Chem. Soc.* **127,** 2802–2803 (2005).
13. Horhota, A. *et al.* Kinetic analysis of an efficient DNA-dependent TNA polymerase. *J. Am. Chem. Soc.* **127,** 7427–7434 (2005).
14. Ichida, J. K., Horhota, A., Zou, K., McLaughlin, L. W. & Szostak, J. W. High fidelity TNA synthesis by therminator polymerase. *Nucleic Acids Res.* **33,** 5219–5225 (2005).
15. Wu, X., Delgado, G., Krishnamurthy, R. & Eschenmoser, A. 2,6-Diaminopurine in TNA: effect on duplex stabilities and on the efficiency of template-controlled ligations. *Org. Lett.* **4,** 1283–1286 (2002).
16. Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proc. Natl Acad. Sci. USA* **94,** 12297–12302 (1997).
17. Rosenbaum, D. M. & Liu, D. R. Efficient and sequence-specific DNA-templated polymerization of peptide nucleic acid aldehydes. *J. Am. Chem. Soc.* **125,** 13924–13925 (2003).
18. Levy, M. & Miller, S. L. The stability of RNA bases: implications for the origin of life. *Proc. Natl Acad. Sci. USA* **95,** 7933–7938 (1998).
19. Rogers, J. & Joyce, G. F. A ribozyme that lacks cytidine. *Nature* **402,** 323–325 (1999).
20. Lin, L., Hom, D., Lindsay, S. M. & Chaput, J. C. *In vitro* selection of histone H4 aptamers for recognition imaging microscopy. *J. Am. Chem. Soc.* **129,** 14568–14569 (2007).
21. Williams, B. A., Lin, L., Lindsay, S. M. & Chaput, J. C. Evolution of a histone H4–K16 acetyl-specific DNA aptamer. *J. Am. Chem. Soc.* **131,** 6330–6331 (2009).
22. Mendonsa, S. D. & Bowser, M. T. *In vitro* evolution of functional DNA using capillary electrophoresis. *J. Am. Chem. Soc.* **126,** 20–21 (2004).
23. Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. & Toole, J. J. Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355,** 564–566 (1992).
24. Kubik, M. F., Stephens, A. W., Schneider, D., Marlar, R. A. & Tasset, D. High-affinity RNA ligands to human α-thrombin. *Nucleic Acids Res.* **22,** 2619–2626 (1994).
25. Woese, C. R. *The Genetic Code: The Molecular Basis for Genetic Expression* (Harper & Row, 1967).
26. Crick, F. H. C. The origin of the genetic code. *J. Mol. Biol.* **38,** 367–380 (1968).
27. Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **38,** 381–393 (1968).
28. Gilbert, W. The RNA world. *Nature* **319,** 618 (1986).
29. Cech, T. R. The RNA worlds in context. *Cold Spring Harb. Perspect. Biol.* http://dx.doi.org/10.1101/cshperspect.a006742 (2011).
30. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418,** 214–221 (2002).
31. Yarus, M. Getting past the RNA world: the initial Darwinian ancestor. *Cold Spring Harb. Perspect. Biol.* http://dx.doi.org/10.1101/cshperspect.a003590 (2011).
32. Joyce, G. F., Schwartz, A. W., Miller, S. L. & Orgel, L. E. The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proc. Natl Acad. Sci. USA* **84,** 4398–4402 (1987).
33. Joyce, G. F. The rise and fall of the RNA world. *The New Biologist* **3,** 399–407 (1991).
34. Loakes, D. & Holliger, P. Polymerase engineering: towards the encoded synthesis of unnatural polymers. *Chem. Commun.* 4619–4631 (2009).
35. Zou, K., Horhota, A., Yu, B., Szostak, J. W. & McLaughlin, L. W. Synthesis of α-L-threofuranosyl nucleoside triphosphates (tNTPs). *Org. Lett.* **7,** 1485–1487 (2005).
36. Yu, H., Jiang, B. & Chaput, J. C. Aptamers can discriminate alkaline proteins with high specificity. *ChemBioChem* **12,** 2659–2666 (2011).

Acknowledgements

The authors thank members of the Chaput laboratory for helpful comments and suggestions. This work was supported by start-up funds from the Biodesign Institute at Arizona State University.

Author contributions

J.C. conceived the project and wrote the manuscript. H.Y., S.Z. and J.C. designed the experiments. H.Y. and S.Z. performed the experiments and wrote initial drafts of the manuscript. All authors discussed the results and commented on the manuscript.

Additional information

The authors declare no competing financial interests. Supplementary information and chemical compound information accompany this paper at www.nature.com/naturechemistry. Reprints and permission information is available online at http://www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.C.C.