Nucleic Acids Research, 2021 **1**
https://doi.org/10.1093/nar/gkab923

Transliteration of synthetic genetic enzymes

Yajun Wang[1], Xiaolin Liu[1], Mouhamad Shehabat[2], Nicholas Chim[2] and John C. Chaput[2,3,4,5,*]

[1]College of Pharmaceutical Sciences, Soochow University, Suzhou 215123, China, [2]Departments of Pharmaceutical Sciences, University of California, Irvine, CA 92697. USA, [3]Department of Chemistry, University of California, Irvine, CA 92697. USA, [4]Department of Molecular Biology and Biochemistry, University of California, CA 92697. USA and [5]Department of Chemical and Biomolecular Engineering, University of California, CA 92697. USA

Received July 21, 2021; Revised September 07, 2021; Editorial Decision September 25, 2021; Accepted September 27, 2021

ABSTRACT

Functional nucleic acids lose activity when their sequence is prepared in the backbone architecture of a different genetic polymer. The only known exception to this rule is a subset of aptamers whose binding mechanism involves G-quadruplex formation. We refer to such examples as transliteration—a synthetic biology concept describing cases in which the phenotype of a nucleic acid molecule is retained when the genotype is written in a different genetic language. Here, we extend the concept of transliteration to include nucleic acid enzymes (XNAzymes) that mediate site-specific cleavage of an RNA substrate. We show that an in vitro selected 2′-fluoroarabino nucleic acid (FANA) enzyme retains catalytic activity when its sequence is prepared as α-L-threofuranosyl nucleic acid (TNA), and vice versa, a TNA enzyme that remains functional when its sequence is prepared as FANA. Structure probing with DMS supports the hypothesis that FANA and TNA enzymes having the same primary sequence can adopt similarly folded tertiary structures. These findings provide new insight into the sequence-structure-function paradigm governing biopolymer folding.

INTRODUCTION

The sequence-structure-function paradigm describes the relationship between the sequence and function of natively folded structures (1). This connection is a central tenet of biology as it explains how changes to the primary sequence of a biopolymer can affect, either positively or negatively, the ability of that molecule to fold into a tertiary structure having a particular function. With the advent of high throughput DNA sequencing (2), it is now possible to study ensembles of sequence changes (i.e. mutations) by constructing fitness landscapes, which are contour plots that depict the distribution of functional molecules in sequence space (3,4). In these plots, peaks are formed by groups of related sequences that share a common functional activity, such as ligand binding or catalysis, while valleys comprise the totality of inactive sequences (5). Although sequence spaces tend to be very large (e.g. a 40-mer RNA has 10^{24} possible sequence combinations), in vitro selection experiments reveal that certain RNA motifs, such as the hammerhead ribozyme and ATP-binding RNA aptamer, are sufficiently common that they can be repeatedly discovered from different pools of random-sequence oligonucleotides (6,7). Such examples suggest that Darwinian evolution may have channeled early life toward the simplest, and therefore the most common solutions to complex biochemical problems.

The extent to which different classes of nucleic acid molecules share overlapping peaks in the fitness landscape is an interesting, but largely unexplored question in molecular evolution. Classical biochemistry studies clearly show that the catalytic properties of ribozymes are not retained when the RNA sequence is prepared as DNA (8,9). This observation, which reflects differences in the sugar pucker and backbone inclination angle between DNA and RNA (10,11), led to the unwritten rule that functional activity cannot be transferred between different classes of nucleic acid molecules. The only known exception is a subset of aptamers whose binding mechanism involves ligand recognition by a G-quadruplex structure. Examples of such interactions include an RNA riboflavin aptamer that continues to bind riboflavin when its primary sequence is synthesized as DNA (12) and a hemin-binding DNA aptamer that binds hemin and continues to exhibit peroxidase activity when prepared as RNA (13). An RNA dopamine aptamer was reported as a non-G-quadruplex example of backbone switching (14), but subsequent analysis revealed that the DNA version functions by non-specific electrostatic interactions and is therefore not a folded aptamer structure (15).

Here, we introduce the term 'transliteration' to describe cases in which the phenotype of a nucleic acid molecule is retained when the genotype is written in a different genetic language. Although such examples are rare among natural genetic polymers, we postulate that the growing number of

*To whom correspondence should be addressed. Tel: +1 949 824 8149; Email: jchaput@uci.edu

synthetic genetic polymers (16,17), also referred to as xeno-nucleic acids or XNAs (18), offer new opportunities for discovering examples whereby sequence and functional information can be transferred between different classes of nucleic acid molecules (19). Our investigation led us to discover cases of transliteration among nucleic acid enzymes (XNAzymes) that mediate site-specific RNA cleavage. We show that an in vitro selected 2′-fluoroarabino nucleic acid (FANA) enzyme (FANAzyme) remains active when prepared as α-L-threofuranosyl nucleic acid (TNA), and vice versa, a TNA enzyme (threozyme) that remains functional when prepared as FANA. Structure probing studies demonstrate that FANA and TNA catalysts constructed with the same primary sequence adopt similarly folded tertiary structures. Together, these findings offer new insight into the sequence-structure-function paradigm by providing the first example of transliteration among nucleic acid enzymes.

MATERIALS AND METHODS

General information

2'F-araNTPs (faATP, faCTP, faGTP, faUTP) were obtained from Metkinen Chemistry (Kuusisto, Finland). TNA triphosphates (tATP, tCTP, tGTP, tTTP) were chemically synthesized following known procedures (20,21). DNA and RNA oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), purified by denaturing polyacrylamide gel electrophoresis, and quantified by UV absorbance via NanoDrop (ThermoFisher, Waltham, MA). YM-3 microcentrifugal concentrators were purchased from EMD Millipore (Billerica, MA). The Monarch DNA gel extraction kit was purchased from New England Biolabs (Ipswich, MA). *Thermococcus gorgonarius* (Tgo) DNA polymerase (exo-), Kod-RS (22) and Kod-RSGA (23) TNA polymerases (exo-), and Bst-LF were expressed and purified from *E. coli* as previously described (24). Sequenase™ Version 2.0 DNA polymerase was purchased from ThermoFisher (Pudong, Shanghai).

In vitro selection of RNA-cleaving threozymes from a three-letter library (A, G, T)

Library preparation. For each round of selection, 1 nmol of the 5′-biotinylated selection primer containing an RNA substrate was annealed to 1 nmol of a DNA library. The DNA library was transcribed into TNA by incubating for 3 h at 55°C in a 1 ml reaction volume. The reaction contained 1 μM selection primer-DNA library duplex, 1× ThermoPol buffer, 1 μM RNA blocker oligonucleotide, 100 μM of each tATP, tGTP and tTTP, and 1 μM Kod-RS TNA polymerase. Following incubation, the reaction was quenched with 0.5 M EDTA (pH 8.0) to give a final concentration of 25 mM.

Selection. 500 μl of streptavidin-coated magnetic beads were washed 3× with 500 μl of wash buffer (50 mM Tris–HCl, pH 7.5, 200 mM NaCl and 1 mM EDTA). The EDTA quenched primer extension product was incubated with the streptavidin beads for 20 min at room temperature to immobilize the TNA–DNA duplex on the beads. Following two more washes with 500 μl of wash buffer, the template strand was removed by five quick wash steps (<30 s per wash) with 500 μl of cold NaOH (0.1 M) containing 1 mM EDTA. The bound TNA strand was immediately neutralized using 1 ml of neutralization buffer (50 mM Tris–HCl, pH 6.0, 1 mM EDTA), followed by a final 1 ml wash with nuclease-free H_2O. The TNA library was then incubated in 100 μl of cleavage buffer (50 mM Tris–HCl, pH 7.5, 200 mM NaCl, 50 mM $MgCl_2$). For the initial round of selection, the library was incubated in cleavage buffer containing 50 mM $MgCl_2$ for 20 h at room temperature (24°C). For the subsequent rounds, the $MgCl_2$ concentration in cleavage buffer was gradually reduced to 10 mM and the incubation time was shortened to 0.5 h for the last round of selection (round 7).

Following incubation with $MgCl_2$, a magnetic field was applied to the beads and the supernatant (90 μl) was recovered, desalted by exchanging into 500 μl of H_2O, and concentrated to the final volume of 11.5 μl using EMD Millipore YM-3 microcentrifuge concentrators. The population of threozymes was reverse transcribed into cDNA by incubating for 3 h at 50°C with Bst-LF. The reaction (20 μl) contained 1x ThermoPol buffer supplemented with 3 mM of $MgCl_2$, 1 μM of reverse transcription primer PBS13, and 500 μM of each dNTP. The cDNA was amplified by PCR using Taq DNA polymerase with the PBS11/PBS13 primer pair: 95°C for 8 min, N cycles of (95°C for 25 s; 58°C for 15 s; 72°C for 30 s). The number of cycles (N) was optimized for each round of selection by sampling PCR reactions every other cycle up to 20 cycles. The amplified DNA was used as template for a second PCR reaction in which PBS11 was replaced by PEGylated PBS11 (PEG-PBS11) following PCR procedures as described above. The second PCR product was purified by denaturing PAGE, and the non-PEGylated strand was used as template for the next round of selection. A total of seven rounds of selection were performed to enrich the fast-to-cleave population (see Supplementary Table S2 for the stringency parameters used for each round of selection).

Illumina next generation sequencing. cDNA amplicons of round 5 and round 7 were generated by PCR using Taq DNA polymerase. Polyclonal cDNA amplicons were purified by 2% agarose using Monarch DNA gel extraction kit. The cDNA amplicons were then made into barcoded Illumina libraries using the Apollo 324 platform and PrepX ILM DNA kit and protocol (Wafergen Bio-Systems, Fremont, CA). The barcoded libraries were spiked into a multiplex of Illumina libraries. The multiplex was denatured and clustered at 12 pM for sequencing on a HiSeq 2500 (Illumina, San Diego, CA) in rapid run mode (8 million reads per amplicon library, single-end 100 cycles) by UCI Genomics High Throughput Facility. Data were analyzed on UCI HPC (https://hpc.oit.uci.edu) using in-house scripts, and sequences were ordered by abundance (Top 5 abundant sequences from R5 and R7 are summarized in Supplementary Table S3).

Truncation analysis. The most abundant threozymes identified by NGS analysis of rounds 5 and 7 (Supplementary Table S3) were constructed as full-length enzymes and as truncated variants by primer-extension. Both full-length

sequences contained internal substrate binding sites, suggesting that the core binding domain was shorter than the 25 nt random region. The truncated variants were synthesized by primer extension using templates that removed a 3′ segment of the sequence that corresponded to the outer substrate binding domain (Supplementary Figure S1).

Template-switching experiments

XNA enzyme preparation. The FANA version of each enzyme was enzymatically transcribed using wild-type Tgo DNA polymerase (exo-) and purified by 10% denaturing PAGE (8 M urea) as previously described (25). The TNA version of each enzyme was enzymatically transcribed in 1x ThermoPol buffer containing 1 μM of primer-template complex, 1 μM of Kod-RSGA and 100 μM of each tNTP. A synthetic DNA template containing $(AAC)_7$ repeats at the 3′-end, originally designed for strand separation, and a 16-nt PBS11 primer were annealed in ThermoPol buffer by heating for 5 min at 90°C and cooling for 10 min at 4°C. TNA transcription was initiated by the addition of tNTPs and Kod-RSGA. Reactions were incubated for 90 min at 55°C and quenched with 0.5 M EDTA (pH 8.0) to a final concentration of 25 mM and lyophilized to dryness. Formamide stop buffer (99% deionized formamide, 25 mM EDTA) that equaled the transcription volume was used to resuspend the dry pellet followed by heat-denaturation for 15 min at 95°C. After cooling for 5 min on ice, the sample was resolved by 10% denaturing purification PAGE (8 M urea), and the gel was visualized by UV-shadowing. Full-length TNA transcript was excised, electroeluted, exchanged into H_2O using YM-3 microcentrifugal device, and UV quantified for concentration by NanoDrop. All of the FANA and TNA catalysts were validated by MALDI-ToF mass spectrometry (Supplementary Table S4).

RNA-cleavage assays. The RNA-cleavage assays were performed under single-turnover conditions with the concentration of substrate and enzyme poised at 0.5 and 2.5 μM, respectively, in reaction buffer at 24°C. The Hammerhead ribozyme and the 10-23 DNAzyme systems were assayed in buffer containing 50 mM Tris–HCl (pH 7.5) supplemented with 5 mM of $MgCl_2$ and 200 mM NaCl. The FANAzyme and threozyme systems were assayed in 50 mM CHES buffer (pH 8.5) supplemented with 50 mM $MgCl_2$ and 200 mM NaCl. Each version of the enzyme and IR-680 labeled RNA substrate were annealed in buffer by heating for 5 min at 90°C and cooling for 5 min on ice. Reactions were initiated by the addition of NaCl and $MgCl_2$ to the reaction. Kinetic reaction progress was monitored by quenching 1.5 μl of reaction using 15 μl (10 equivalents, v/v) of formamide stop buffer (99% deionized formamide, 25 mM EDTA) at designated time points. Samples were denatured for 15 min at 95°C and analyzed by 15% denaturing PAGE (8 M). Gels were visualized on a LI-COR Odyssey CLx.

DMS probing of XNAzymes and secondary structure analysis

Preparation of XNAzyme constructs for DMS probing. Dimethyl sulfate (DMS) modification was adapted from an approach previously reported (26,27), where the sequence of XNAzyme 12-7 was flanked by two terminal SHAPE cassettes. FANA and TNA versions of the constructs were prepared by primer extension using Tgo-WT and Kod-RSGA to copy the same DNA template into FANA and TNA, respectively. The transcription reactions were purified by 8% denaturing PAGE (8 M urea). Full-length transcripts were excised, electroeluted, exchanged into H_2O using YM-3 microcentrifugal devices, and UV quantified for concentration by NanoDrop.

XNAzyme folding and DMS modification. 5 μl of annealing mixture containing 2 μM of FAM labeled XNAzyme, 5 μM of 2′-OMe-guanine RNA substrate, and 400 mM of NaCl was denatured at 95°C for 15 min, and slowly cooled to room temperature. Annealed XNAzyme-substrate complexes were subsequently treated with 4 μl of 2.5× folding buffer (750 mM cacodylate, pH 7.0, and 125 mM $MgCl_2$) for 30 min to allow for catalytic folding. The folded mixtures were treated with 1 μl of DMS (1.7 M in absolute ethanol) by incubating at 17°C for 6 min, followed by quenching with 10 μl of 2-mercaptoethanol. The quenched reactions were then precipitated by 10 volumes of absolute ethanol, resuspended in H_2O, desalted by YM-3 microcentrifugal devices, and lyophilized to dryness. No-reagent control reactions were treated with 1 μl of absolute ethanol.

Reverse transcription (RT) of DMS modified samples. The desalted samples from the DMS modification step were resuspended in 10 μl of H_2O. Reverse transcriptions were performed using Bst-LF DNA polymerase. Briefly, 5 μl of the recovered XNA construct and 10 pmol of IR-680 labeled RT primer were annealed in 1x ThermoPol buffer by denaturing for 15 min at 95°C and cooling for 5 min on ice. Primer extension reaction was then initiated by the addition of dNTPs to 0.5 mM, $MgCl_2$ to 3 mM, and Bst-LF DNA polymerse to ∼1 μM. All reactions were run for 180 min at 55°C, and then quenched by the addition of EDTA (pH 8.0) to 25 mM. 1 μl of each quenched RT reaction was transferred to 9 ul of formamide stop buffer (99% deionized formamide, 25 mM EDTA) and then denatured for 15 min at 95°C. Then the RT samples were purified by 15% denaturing PAGE (8 M urea). Gels were visualized on a LI-COR Odyssey CLx.

Generation of DNA ladder by Sequenase V2.0 mediated incorporation of ddTTP and ddGTP residues as chain terminators. DMS probing construct consisting of DNA was prepared by following the same procedures as that for XNAzymes. 1.2 μM of the construct and 1 μM of IR-680 labeled RT primer were annealed in 1× Sequenase buffer by denaturing for 15 min at 95°C and cooling for 5 min on ice. Primer extension reaction was then initiated by the addition of sequencing mix (3 mM dNTPs and 300 μM ddGTP/ddTTP) to a concentration of 1.2 mM, 50 mM DTT, and 6.5 U/μl Sequenase™ Version 2.0 DNA polymerase. All reactions were run for 180 min at 60°C, and then quenched by the addition of EDTA (pH 8.0) to 25 mM. 1 μl of each quenched RT reaction was transferred to 9 ul of formamide stop buffer (99% deionized formamide, 25 mM EDTA) and denatured for 15 min at 95°C. Then the RT

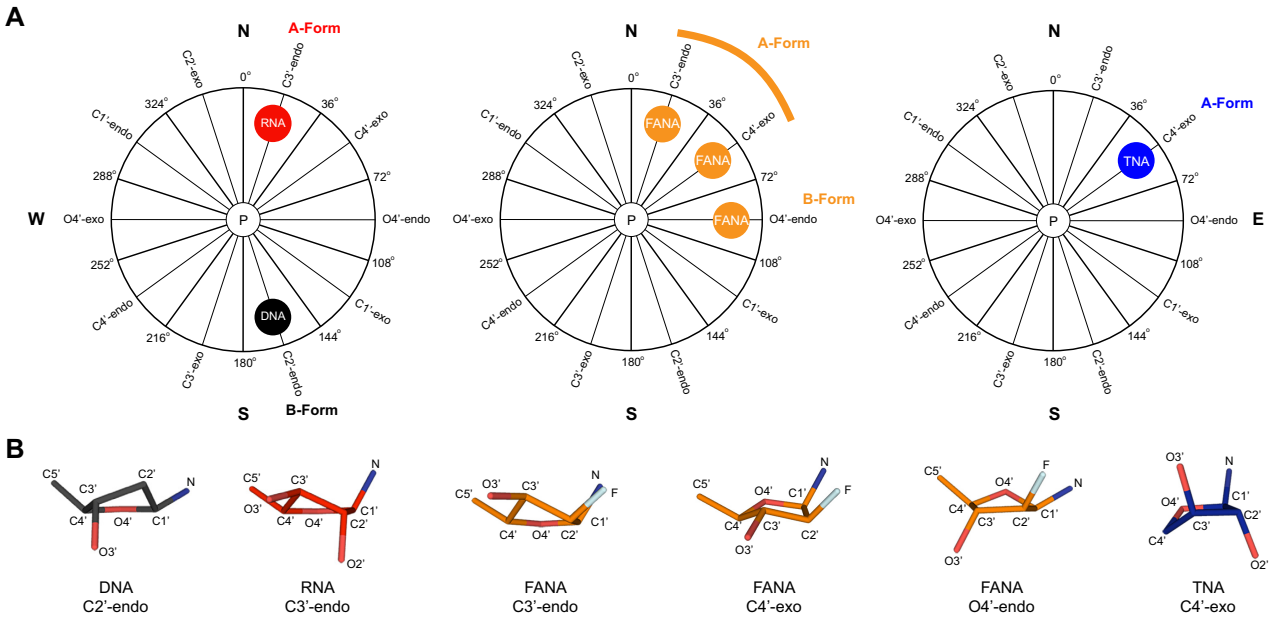

Figure 1. The pseudorotation phase angles observed for the sugar pucker found in DNA, RNA, FANA, and TNA. (**A**) B-form DNA favors a C2′-*endo* sugar pucker, while A-form RNA prefers a C3′-*endo* conformation. FANA is capable of adopting both C3′-*endo* and C4′-*exo* sugar pucker conformations in A-form helices and an O4′-*endo* pucker in B-form helices. TNA is limited to A-form helices with a C4′-*exo* sugar pucker. (**B**) Sugar conformations observed in crystallographic structures of antiparallel duplexes. PDB codes: 3BSE (DNA), 3ND4 (RNA), 2FIL and 388D (FANA), and 1N1O (TNA).

samples were purified by 15% denaturing PAGE (8 M urea). Gels were visualized on a LI-COR Odyssey CLx.

RESULTS

Transliteration demands that two different genetic polymers have structurally similar backbone architectures so that the functional information encoded in one genetic polymer is maintained when the sequence is prepared in a different genetic polymer. In an effort to identify classes of nucleic acid molecules that might be capable of transliteration, we evaluated the helical parameters of antiparallel duplex structures that have been solved by NMR or X-ray crystallography (19). From this analysis, we observed that 2′-deoxy-2′-fluoroarabino nucleic acid (FANA) is able to adopt multiple different sugar conformations (28–30), including the C3′-*endo* and C4′-*exo* sugar puckers commonly found in RNA and TNA duplexes, respectively (31–33). Based on this observation, we hypothesized that FANA was a reasonable candidate for transliteration (Figure 1), as the polymorphic nature of the arabinose sugar overlaps with the ribose and threose conformations found in RNA and TNA, respectively.

To explore this possibility in an experimental setting, four chemically distinct nucleic acid enzymes that share a common biochemical function but derive from different classes of nucleic acid molecules (Figure 2A) were selected for biochemical analysis. Each enzyme catalyzes the site-specific cleavage of an RNA substrate following a conserved reaction mechanism that produces a 2′,3′-cyclic phosphate and a 5′-hydroxyl group at the scissile phosphodiester bond (Figure 2B). The set of nucleic acid enzymes includes the hammerhead ribozyme, a naturally occurring ribozyme originally found in plant viruses (34); DNAzyme 10-23, a completely unnatural product of laboratory evolution that has been extensively evaluated as a gene silencing agent in clinical trials (35); FANAzyme 12-7, the first XNAzyme known to exhibit saturating kinetics resembling classic enzyme behavior (25); and threozyme 7-1s (Supplementary Figure S1), a TNA enzyme evolved for the current study that derives from a genetic language having a backbone repeat unit that is one atom shorter than the backbone repeat unit found in DNA and RNA (36).

To facilitate the study, each enzyme was synthesized in all four genetic languages. For example, the nucleotide sequence of the hammerhead ribozyme was prepared as RNA, DNA, FANA and TNA. The DNA and RNA versions of the catalysts, as well as the 5′ labeled RNA substrates used to test for catalytic activity, were all generated by solid-phase oligonucleotide synthesis (Supplementary Tables S1–S3). Because XNA oligonucleotides are more difficult to generate by solid-phase synthesis than DNA and RNA, the FANA and TNA versions of each catalyst were enzymatically prepared using polymerases (Supplementary Figures S2, S3) that were previously established to copy DNA templates into either FANA or TNA (37). All of the nucleic acid enzymes used in this study were purified by denaturing polyacrylamide gel electrophoresis (PAGE) and validated by mass spectrometry (Supplementary Table S4).

Denaturing PAGE was used to assess the catalytic activity of each construct under single-turnover conditions in a *trans*-cleavage assay format. Consistent with published reports (7,25,35), the hammerhead ribozyme, DNAzyme 10-23, and FANAzyme 12-7 each exhibit robust catalytic profiles (Figure 3) with product formation observed upon magnesium chloride addition. By comparison, threozyme 7-1s, which was evolved from a 3 nt library, is a slower enzyme with reaction kinetics that are more consistent

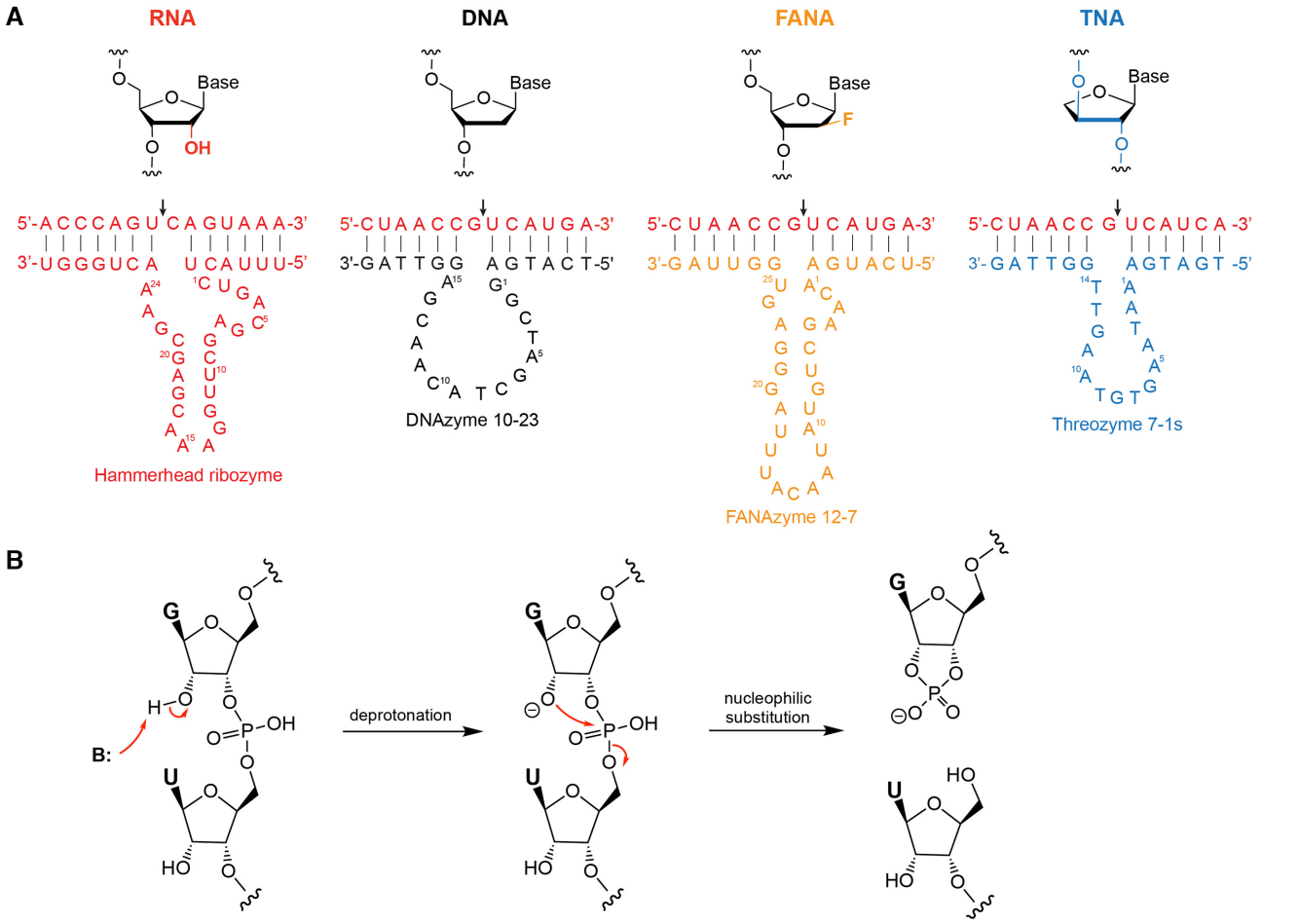

Figure 2. Nucleic acid enzymes. (**A**) The chemical structure and sequence of nucleic acid catalysts that share a common RNA cleavage activity but derive from different genetic systems. Arrow indicates the cleavage site. Color code: RNA (red), DNA (black), FANA (orange) and TNA (blue). (**B**) Mechanism of phosphodiester bond cleavage by RNA-cleaving nucleic acid enzymes.

with nucleic acid enzymes that have been isolated from libraries of reduced nucleotide diversity (38,39). These data confirm that the control catalysts are properly folded and functional.

Our screen for transliteration activity among the RNA cleaving nucleic acid enzymes revealed a number of striking observations. First, the catalytic activity of the hammerhead ribozyme as well as that of DNAzyme 10-23 does not recapitulate when their sequence is prepared in a different genetic language (Figure 3). This conclusion was confirmed by repeating the assay under more permissive reaction conditions (Supplementary Figure S4) that tested for weaker activity over a longer period of time (22 h, 50 mM $MgCl_2$, pH 8.5). This result indicates that DNA and RNA catalysts have strict geometric constraints that limit the transfer of functional information to the genetic systems evaluated in this study. Second, the in vitro selected FANA and TNA enzymes exhibit a remarkable capacity to transfer functional information between one another. As indicated in Figure 3, FANAzyme 12-7 remains functional when its sequence is prepared in a TNA backbone architecture, and likewise, threozyme 7-1s remains active when its sequence is prepared in a FANA backbone. Consistent with the notion that transliteration is due to structural compatibility rather than some unknown feature of the nucleic acid sequence, the DNA and RNA versions of both enzymes are inactive even after 22 h of incubation in buffer containing 50 mM $MgCl_2$ (pH 8.5).

Having established that FANA and TNA are capable of transliteration, we wished to evaluate the catalytic properties of active and inactive versions of these enzymes to determine if the genotype-phenotype linkage is maintained in a predictable way such that functional catalysts remain active and non-functional catalysts remain inactive. FANAzymes 12-10 and 12-20 are single-point variants of FANAzyme 12-7 known to support and inhibit RNA-cleavage activity, respectively (25). Likewise, threozymes 5-1s bulge and 5-13 are novel enzymes discovered in the current study that also support and inhibit catalytic activity, respectively. When assayed for activity in their native and compatible backbone structures (Figure 4), the functional catalysts remain active and the non-functional catalysts remain inactive. Similar results were obtained for another example of active and inactive versions of the FANAzyme (Supplementary Figure S5) confirming that a programmable relationship is maintained across the two backbone architectures.

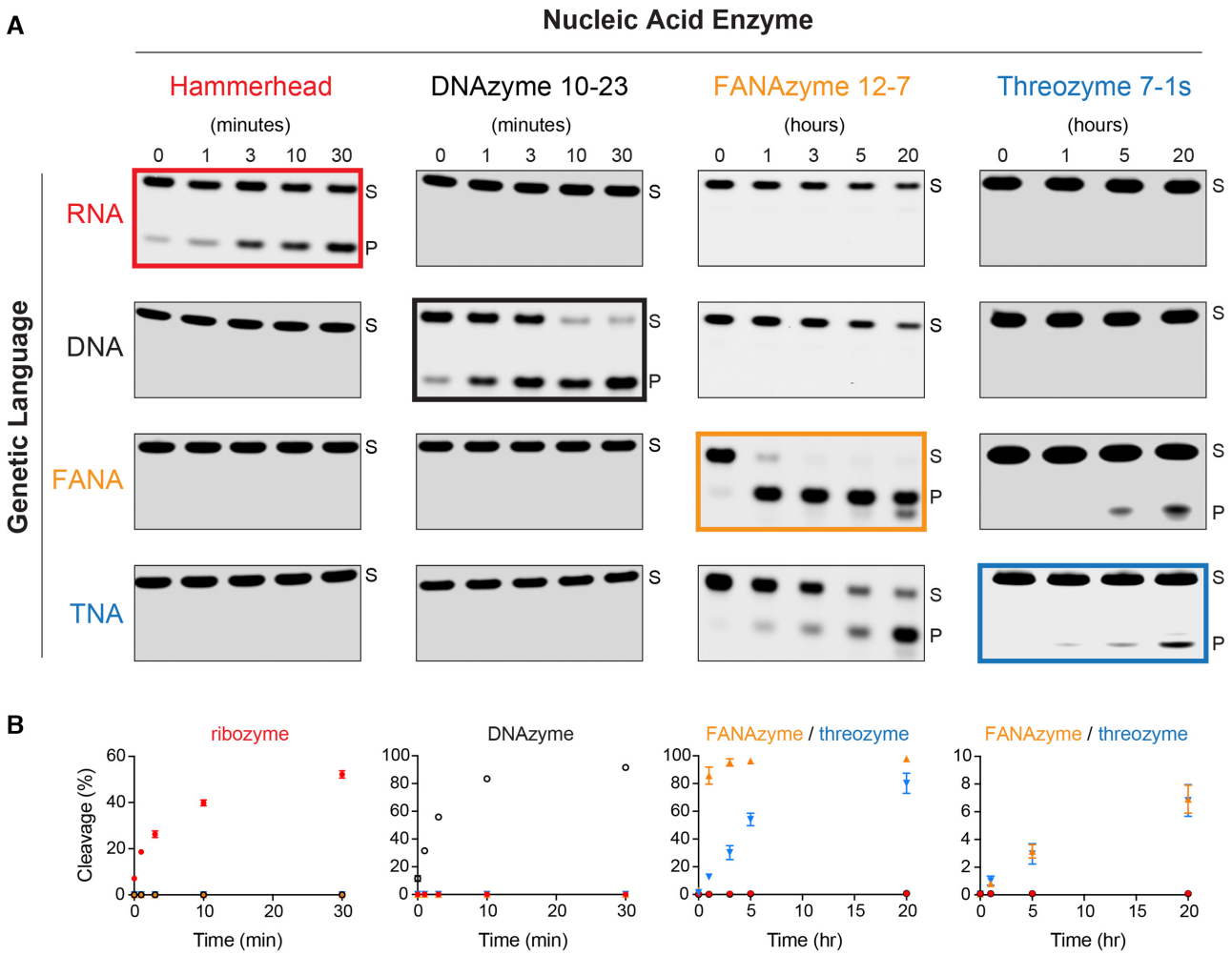

Figure 3. Transliteration. (**A**) Time-dependent RNA cleavage activity observed for known and putative RNA-cleaving enzymes. Ribozymes and DNAzymes were evaluated in buffer containing 5 mM $MgCl_2$ (pH 7.5). FANAzymes and threozymes were evaluated in buffer containing 50 mM $MgCl_2$ (pH 8.5). Bold colored boxes indicate wild-type catalysts. Boxes framed with thin black lines indicate synthetic versions of the wild-type sequence prepared in a different genetic language. Gel images are representative of experimental data from three experimental replicates. (**B**) Kinetic plots showing product formation as a function of time. Abbreviations: S, substrate band and P, band corresponding the to the 5′ cleavage product. $n = 3$; error bars represent standard deviation, s.d. Colors: RNA (red squares), DNA (black circles), FANA (orange triangles), and TNA (blue inverted triangles).

The surprising discovery of a non-G-quadruplex example of transliteration observed between two chemically different classes of nucleic acid enzymes (FANA and TNA) isolated from separate *in vitro* selection experiments that probed different regions of sequence space raises the interesting question of whether two different XNA versions of the same enzyme adopt similarly folded tertiary structures. To explore this question in greater detail, we separately prepared FANAzyme 12-7 in the genetic languages of FANA and TNA and evaluated their folding topology by structure probing with dimethyl sulfate (DMS), a reagent that specifically modifies unpaired adenine and cytosine residues present in the tertiary structure of folded nucleic acid molecules. The assay was performed using an inactive version of the RNA substrate that carried a 2′-*O*-methyl residue at the guanosine cleavage site (Figure 5A). Analysis of the reverse transcription products of the DMS treated XNAzyme constructs by denaturing PAGE (Figure 5B) reveals that the FANA and TNA versions of FANAzyme 12-7 adopt similar global folds with DMS reactive A and C bases located at the same nucleotide positions (Figure 5C). Relative to FANAzyme 12-7, the A and C bases in threozyme 12-7 display slightly stronger DMS reactivity profiles (Figure 5C), which may be due to the reduced reverse transcription efficiency of TNA templates as compared to FANA templates. Nevertheless, the data supports the hypothesis that FANA and TNA enzymes having the same primary sequence can adopt similarly folded tertiary structures, which is consistent with the sequence-structure-function paradigm.

DISCUSSION

In this study, we extend the sequence-structure-function paradigm by establishing that relatively simple catalytic domains comprised of FANA and TNA oligonucleotides have the capacity to undergo transliteration. This finding represents the first case in which a non-G-quadruplex

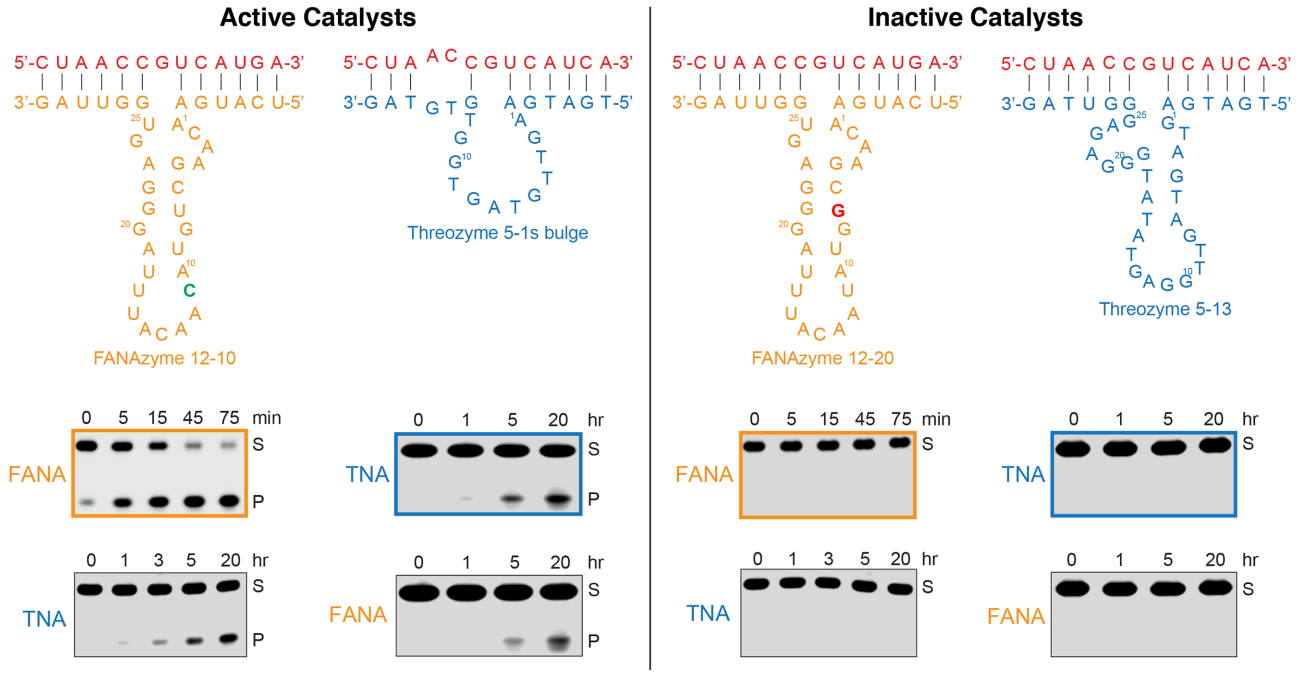

Figure 4. Programmable information transfer. Time-dependent RNA cleavage activity observed for active and inactive FANA and TNA catalysts. Template switching reveals that functional activities are programmable such that active FANA and TNA catalysts remain active and inactive FANA and TNA catalysts remain inactive when prepared in their opposing genetic polymer. Bold colored boxes indicate the wild-type catalyst. The green and red nucleotides in FANAzymes 12-10 and 12-20 denote genetic mutations discovered in the original 12-7 FANAzyme sequence (5). Boxes framed with thin black lines indicate synthetic analogs of wild-type catalysts produced in another genetic language. Abbreviations: S, substrate band and P, band corresponding to the 5′ cleavage product. $n = 3$; gel images are representative of experimental data from at least 3 experimental replicates.

structure remains active when its primary sequence is prepared in a different language as well as the first case in which transliteration has been shown to occur for a biochemical activity other than ligand binding. One particularly interesting aspect of this result is the significant chemical differences that exist between the chemical structures of FANA and TNA (Figure 2). Similar to DNA and RNA, FANA has a six-atom backbone repeat unit with a five-carbon sugar unit that is connected by 3′,5′-phosphodiester linkages. By comparison, TNA has a five-atom backbone repeat unit with a four-carbon sugar that is connected by 2′,3′-phosphodiester linkages. Thus, from a purely chemical perspective, these differences are more substantial than the chemical differences separating RNA from DNA, which are incapable of undergoing transliteration even though they differ by a single 2′-hydroxyl group (8,9).

The critical aspect of transliteration is not chemical similarity but structural similarity. In the current example, we suspect that the ability of FANA and TNA to transfer structural and functional information entirely on the basis of their primary sequence is due to the shared ability of FANA and TNA polymers to adopt a C4′-*exo* sugar conformation. This hypothesis is consistent with the long standing view that the sugar pucker, more than any other structural parameter, is responsible for determining the helical geometry of a nucleic acid duplex (10). To what extent this phenomenon extends to other catalytic activities is unclear, as the field of synthetic genetics has only recently produced its first examples of catalytic XNA molecules (40). In addition to individual selection experiments, it may be possible to explore this question more thoroughly by constructing comprehensive sequence-fitness maps that query defined libraries of FANA and TNA molecules for common solutions to the same biochemical problem (4). For example, if the selection imposes a constraint for catalytic activity, then overlapping peaks in the fitness landscape should lead to the discovery of homologous sequences from different classes of genetic molecules that catalyze the same chemical reaction.

Although more work is needed to understand the structural similarities and distribution of transliteration events in sequence space, the occurrence of these events in catalytic motifs raises the question of whether transliteration is a structural anomaly or possible evolutionary path between different nucleic acid worlds. A growing body of evidence exists to support the RNA world hypothesis in which life evolved from RNA-based organisms that stored genetic information and catalyzed chemical reactions (41). Whether RNA was the first genetic polymer or an evolutionary intermediate in the path to extant life is less clear (42). Arguments against an RNA-later model have largely focused on the loss of functional activity that was predicted to occur as life transitioned from a pre-RNA world to the RNA world (43). However, the demonstration that transliteration is in fact physically possible for small catalytic motifs suggests that this problem may be less severe than previously thought and could be overcome by identifying pre-RNA

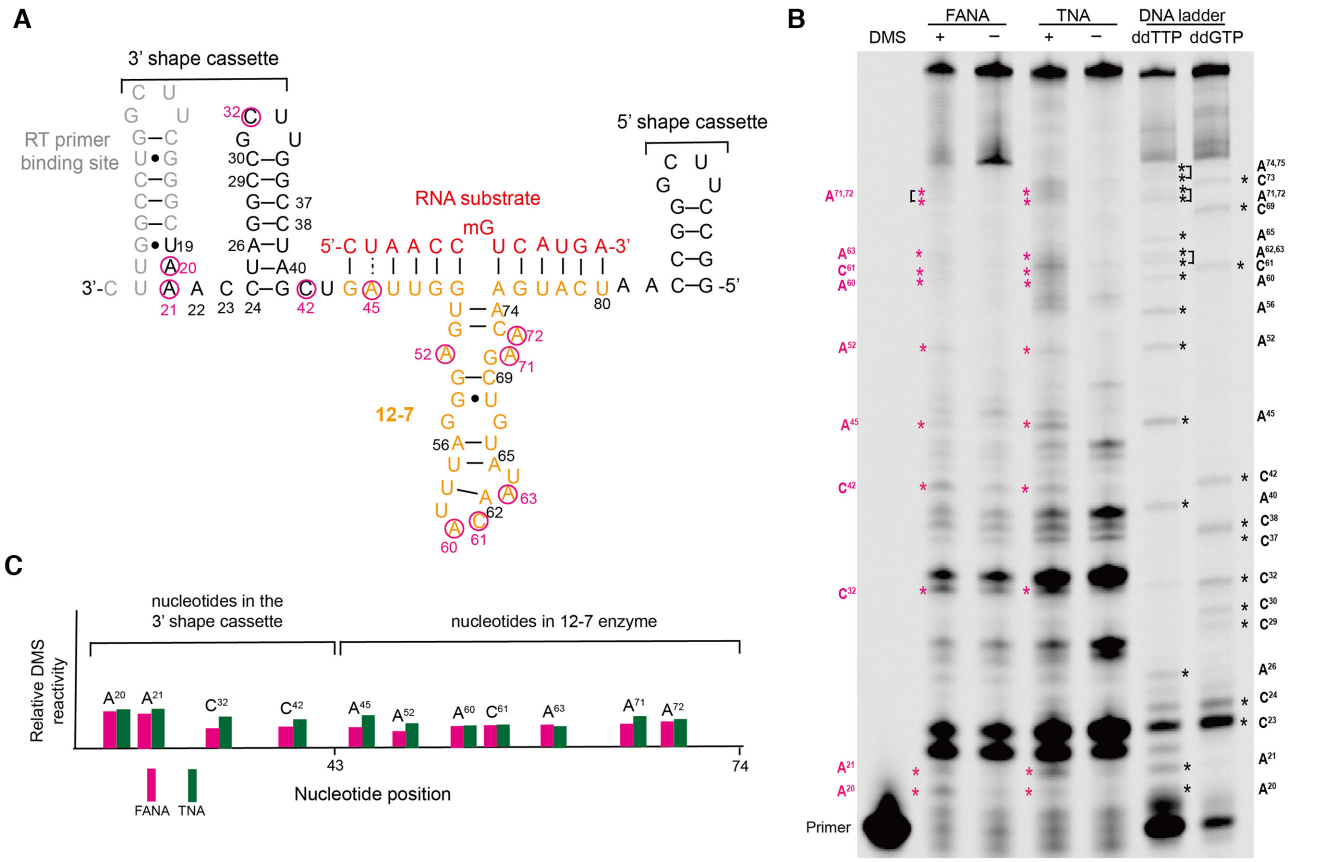

Figure 5. Chemical probing of XNAzyme 12-7 secondary structure by dimethyl sulfate (DMS) footprinting. (**A**) RNA cleaving FANA or TNA 12-7 embedded in structural cassettes, complexed with an inactive RNA substrate containing a 2′-*O*-methyl G residue at the cleavage site. DMS reacts predominantly with unpaired A and C. (**B**) Denaturing PAGE analysis of reverse transcription (RT) products using DMS treated XNAzyme as templates. (**C**) A summary of DMS reactive nucleosides revealed by denaturing PAGE gel in (B). The comparison of the DMS reactivity patterns reveals that FANA and TNA 12-7 adopt similar global conformations. Positions circled in magenta in (A) display DMS reactivities in both FANA and TNA, the latter of which was in general slightly stronger than the former presumably due its inefficiency for RT.

world candidates that are capable of transferring sequence information and functional information when prepared as RNA.

Transliteration could also find practical value in the evolution of oligonucleotide therapeutics (44). As an example, the superior biological stability of TNA relative to FANA could be accessed through a process where functional molecules are first evolved using the FANA system and then transitioned to a more stable backbone structure through TNA synthesis (45). This approach would reduce the cost of generating TNA reagents, as FANA building blocks and polymerases are commercially available (25), whereas their TNA equivalents can only be obtained through chemical synthesis and enzyme engineering (37). In this way, one genetic language could lend to the discovery of biologically stable and therapeutically relevant reagents of another genetic language.

In summary, we offer FANA and TNA as a model for studying a genetics phenomenon, termed transliteration, in which the phenotype of a nucleic acid enzyme is retained when the genotype is written in a different genetic language. We suggest that further investigations into transliteration could have basic and practical value by unlocking hidden elements of nucleic acid structure.

DATA AVAILABILITY

All data are available from the authors upon request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank members of the Chaput laboratory for helpful discussions and critical reading of the manuscript.

Author contributions: Y.W. and J.C. conceived the project and designed the experiments. Y.W., X.L., M.S. and N.C. performed all the experiments. J.C. wrote the manuscript with drafts from Y.W. and N.C. All authors reviewed and commented on the manuscript.

FUNDING

W.M. Keck Foundation; National Science Foundation [CHE: 2001434, MCB: 1946312]. Funding for open access charge: NSF.

Conflict of interest statement. None declared.

REFERENCES

1. Redfern,O.C., Dessailly,B. and Orengo,C.A. (2008) Exploring the structure and function paradigm. *Curr. Opin. Struct. Biol.*, **18**, 394–402.
2. Goodwin,S., McPherson,J.D. and McCombie,W.R. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.*, **17**, 333–351.
3. de Visser,J.A.G.M. and Krug,J. (2014) Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.*, **15**, 480–490.
4. Blanco,C., Janzen,E., Pressman,A., Saha,R. and Chen,I.A. (2019) Molecular fitness landscapes from high-coverage sequence profiling. *Annu. Rev. Biophys.*, **48**, 1–18.
5. Smith,J.M. (1970) Natural selection and the concept of a protein space. *Nature*, **225**, 563–564.
6. Vu,M.M., Jameson,N.E., Masuda,S.J., Lin,D., Larralde-Ridaura,R. and Luptak,A. (2012) Convergent evolution of adenosine aptamers spanning bacterial, human, and random sequences revealed by structure-based bioinformatics and genomic SELEX. *Chem. Biol.*, **19**, 1247–1254.
7. Salehi-Ashtiani,K. and Szostak,J.W. (2001) In vitro evolution suggests multiple origins for the hammerhead ribozyme. *Nature*, **414**, 82–84.
8. Jeffries,A.C. and Symons,R.H. (1989) A catalytic 13-mer ribozyme. *Nucleic. Acids. Res.*, **17**, 1371–1377.
9. Chowrira,B.M., Berzal-Herranz,A., Keller,C.F. and Burke,J.M. (1993) Four ribose 2′-hydroxyl groups essential for catalytic function of the hairpin ribozyme. *J. Biol. Chem.*, **268**, 19458–19462.
10. Saenger,W. (1984) In: *Principles of Nucleic Acid Structure*. Springer-Verlag, NY.
11. Pallan,P.S., Lubini,P., Bolli,M. and Egli,M. (2007) Backbone-base inclination as a fundamental determinant of nucleic acid self- and cross-pairing. *Nucleic Acids Res.*, **35**, 6611–6624.
12. Lauhon,C.T. and Szostak,J.W. (1995) RNA aptamers that bind flavin and nicotinamide redox cofactors. *J. Am. Chem. Soc.*, **117**, 1246–1257.
13. Travascio,P., Li,Y. and Sen,D. (1998) DNA-enhanced peroxidase activity of a DNA-aptamer-hemin complex. *Chem. Biol.*, **5**, 505–517.
14. Walsh,R. and DeRosa,M.C. (2009) Retention of function in the DNA homolog of the RNA dopamine aptamer. *Biochem. Biophys. Res. Commun.*, **388**, 732–735.
15. Alvarez-Martos,I. and Ferapontova,E.E. (2017) A DNA sequence obtained by replacement of the dopamine RNA aptamer bases is not an aptamer. *Biochem. Biophys. Res. Commun.*, **489**, 381–385.
16. Chaput,J.C., Yu,H. and Zhang,S. (2012) The emerging world of synthetic genetics. *Chem. Biol.*, **19**, 1360–1371.
17. Pinheiro,V.B., Loakes,D. and Holliger,P. (2012) Synthetic polymers and their potential as genetic materials. *Bioessays*, **35**, 113–122.
18. Chaput,J.C. and Herdewijn,P. (2019) What Is XNA? *Angew. Chem. Int. Ed. Engl.*, **58**, 11570–11572.
19. Anosova,I., Kowal,E.A., Dunn,M.R., Chaput,J.C., Van Horn,W.D. and Egli,M. (2016) The structural diversity of artificial genetic polymers. *Nucleic Acids Res.*, **44**, 1007–1021.
20. Sau,S.P., Fahmi,N.E., Liao,J.-Y., Bala,S. and Chaput,J.C. (2016) A scalable synthesis of α-L-threose nucleic acid monomers. *J. Org. Chem.*, **81**, 2302–2307.
21. Liao,J.-Y., Bala,S., Ngor,A.K., Yik,E.J. and Chaput,J.C. (2019) P(V) reagents for the scalable synthesis of natural and modified nucleoside triphosphates. *J. Am. Chem. Soc.*, **141**, 13286–13289.
22. Nikoomanzar,A., Vallejo,D. and Chaput,J.C. (2019) Elucidating the determinants of polymerase specificity by microfluidic-based deep mutational scanning. *ACS Synth. Biol.*, **8**, 1421–1429.
23. Nikoomanzar,A., Vallejo,D., Yik,E.J. and Chaput,J.C. (2020) Programmed allelic mutagenesis of a DNA polymerase with single amino acid resolution. *ACS Synth. Biol.*, **9**, 1873–1881.
24. Nikoomanzar,A., Dunn,M.R. and Chaput,J.C. (2017) Engineered polymerases with altered substrate specificity: expression and purification. *Curr. Protoc. Nucleic Acid Chem.*, **69**, 4.75.
25. Wang,Y., Ngor,A.K., Nikoomanzar,A. and Chaput,J.C. (2018) Evolution of a general RNA-cleaving FANA enzyme. *Nat. Commun.*, **9**, 5067.
26. Homan,P.J., Favorov,O.V., Lavender,C.A., Kursun,O., Ge,X., Busan,S., Dokholyan,N.V. and Weeks,K.M. (2014) Single-molecule correlated chemical probing of RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 13858–13863.
27. Wilkinson,K.A., Merino,E.J. and Weeks,K.M. (2006) Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.*, **1**, 1610–1616.
28. Li,F., Sarkhel,S., Wilds,C.J., Wawrzak,Z., Prakash,T.P., Manoharan,M. and Egli,M. (2006) 2′-Fluoroarabino- and arabinonucleic acid show different conformations, resulting in deviating RNA affinities and processing of their heteroduplexes with RNA by RNase H. *Biochemistry*, **45**, 4141–4152.
29. Berger,I., Tereshko,V., Ikeda,H., Marquez,V.E. and Egli,M. (1998) Crystal structures of B-DNA with incorporated 2′-deoxy-2′-fluoro-arabino-furanosyl thymines: implications of conformational preorganization for duplex stability. *Nucleic Acids Res.*, **26**, 2473–2480.
30. Trempe,J.F., Wilds,C.J., Denisov,A.Y., Pon,R.T., Damha,M.J. and Gehring,K. (2001) NMR solution structure of an oligonucleotide hairpin with a 2′F-ANA/RNA stem: implications for RNase H specificity toward DNA/RNA hyrbrid duplexes. *J. Am. Chem. Soc.*, **123**, 4896–4903.
31. Wing,R., Drew,H., Takano,T., Broka,C., Tanaka,S., Itakura,K. and Dickerson,R.E. (1980) Crystal structure analysis of a complete turn of B-DNA. *Nature*, **287**, 755–758.
32. Ebert,M.-O., Mang,C., Krishnamurthy,R., Eschenmoser,A. and Jaun,B. (2008) The structure of a TNA-TNA complex in solution: NMR study of the octamer duplex derived from a-(L)-threofuranosyl-(3′-2′)-CGAATTCG. *J. Am. Chem. Soc.*, **130**, 15105–15115.
33. Wilds,C.J., Wawrzak,Z., Krishnamurthy,R., Eschenmoser,A. and Egli,M. (2002) Crystal structure of a B-form DNA duplex containing (L)-a-threofuranosyl (3′-2′) nucleosides: a four-carbon sugar is easily accommodated into the backbone of DNA. *J. Am. Chem. Soc.*, **124**, 13716–13721.
34. Pabon-Pena,L.M., Zhang,Y. and Epstein,L.M. (1991) Newt satellite 2 transcripts self-cleave by using an extended hammerhead structure. *Mol. Cell Biol.*, **11**, 6109–6115.
35. Santoro,S.W. and Joyce,G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 4262–4266.
36. Schöning,K.U., Scholz,P., Guntha,S., Wu,X., Krishnamurthy,R. and Eschenmoser,A. (2000) Chemical etiology of nucleic acid structure: the a-threofuranosyl-(3′→2′) oligonucleotide system. *Science*, **290**, 1347–1351.
37. Nikoomanzar,A., Chim,N., Yik,E.J. and Chaput,J.C. (2020) Engineering polymerases for applications in synthetic biology. *Q. Rev. Biophys.*, **53**, e8.
38. Reader,J.S. and Joyce,G.F. (2002) A ribozyme composed of only two different nucleotides. *Nature*, **420**, 841–844.
39. Rogers,J. and Joyce,G.F. (1999) A ribozyme that lacks cytidine. *Nature*, **402**, 323–325.
40. Taylor,A.I., Houlihan,G. and Holliger,P. (2019) Beyond DNA and RNA: the expanding toolbox of synthetic genetics. *Cold Spring Harb. Perspect. Biol.*, **11**, a032490.
41. Gilbert,W. (1986) The RNA world. *Nature*, **319**, 618.
42. Engelhart,A.E. and Hud,N.V. (2011) Primitive genetic polymers. *Cold Spring Harb. Perspect. Biol.*, **2**, a002196.
43. Bohler,C., Nielsen,P.E. and Orgel,L.E. (1995) Template switching between PNA and RNA oligonucleotides. *Nature*, **376**, 578–581.
44. Khvorova,A. and Watts,J.K. (2017) The chemical evolution of oligonucleotide therapies of clinical utility. *Nat. Biotechnol.*, **35**, 238–248.
45. Culbertson,M.C., Temburnikar,K.W., Sau,S.P., Liao,J.-Y., Bala,S. and Chaput,J.C. (2016) Evaluating TNA stability under simulated physiological conditions. *Bioorg. Med. Chem. Lett.*, **26**, 2418–2421.