Evolution of functional nucleic acids in the presence of nonheritable backbone heterogeneity

Simon G. Trevino[a], Na Zhang[a], Mark P. Elenko[a], Andrej Lupták[b], and Jack W. Szostak[a,1]

[a]Howard Hughes Medical Institute, Center for Computational and Integrative Biology, and Department of Molecular Biology, Simches Research Center, Massachusetts General Hospital, Boston, MA 02144; and [b]Departments of Pharmaceutical Sciences, Chemistry and Molecular Biology and Biochemistry, University of California, Irvine, CA 92697

Edited by Dinshaw J. Patel, Memorial Sloan-Kettering Cancer Center, New York, NY, and approved July 11, 2011 (received for review May 4, 2011)

Multiple lines of evidence support the hypothesis that the early evolution of life was dominated by RNA, which can both transfer information from generation to generation through replication directed by base-pairing, and carry out biochemical activities by folding into functional structures. To understand how life emerged from prebiotic chemistry we must therefore explain the steps that led to the emergence of the RNA world, and in particular, the synthesis of RNA. The generation of pools of highly pure ribonucleotides on the early Earth seems unlikely, but the presence of alternative nucleotides would support the assembly of nucleic acid polymers containing nonheritable backbone heterogeneity. We suggest that homogeneous monomers might not have been necessary if populations of heterogeneous nucleic acid molecules could evolve reproducible function. For such evolution to be possible, function would have to be maintained despite the repeated scrambling of backbone chemistry from generation to generation. We have tested this possibility in a simplified model system, by using a T7 RNA polymerase variant capable of transcribing nucleic acids that contain an approximately 1:1 mixture of deoxy- and ribonucleotides. We readily isolated nucleotide-binding aptamers by utilizing an in vitro selection process that shuffles the order of deoxy- and ribonucleotides in each round. We describe two such RNA/DNA mosaic nucleic acid aptamers that specifically bind ATP and GTP, respectively. We conclude that nonheritable variations in nucleic acid backbone structure may not have posed an insurmountable barrier to the emergence of functionality in early nucleic acids.

abiogenesis | genetic takeover | RNA progenitor | origin of life

Given that RNA is likely to have played a key role early in the evolution of life (1), understanding the origins of the RNA-based biosphere is a critical aspect of understanding the origin of life. The difficulties associated with the prebiotic synthesis of a macromolecule as complicated and fragile as RNA have stimulated interest in the hypothesis that RNA was preceded by simpler and/or more stable progenitor nucleic acids (2–4). The systematic exploration of nucleic acids that are structurally related to RNA has revealed that the formation of stable Watson–Crick base-paired, antiparallel duplex structures is compatible with a surprising degree of variation in the sugar-phosphate backbone (5). These studies imply that many distinct nucleic acids are in principle capable of mediating the inheritance of genetic information. On the other hand, recent advances in the prebiotic chemistry leading to the pyrimidine ribonucleotides (6) have revived the prospects of the RNA-first model, with the attendant advantage of avoiding a difficult genetic takeover. Both RNA-first and RNA-late models assume that life began with a single, well-defined genetic polymer. Perhaps the greatest problem implicit in this assumption is the challenge of explaining the origin of pure nucleotide monomers on the early Earth. Indeed, the supposed requirement for a prebiotic pool containing a high concentration of pure ribonucleotides, ready to be polymerized into the primordial genetic material was parodied by Orgel et al. as "the molecular biologist's dream" (7). Here we discuss and begin to test the idea that life may have started in a less well-defined manner, from mixtures of chemically diverse monomer building blocks.

Any genetic polymer capable of Darwinian evolution must be able to replicate and must be able to encode useful functions. Of these two properties, replication is probably more tolerant of structural heterogeneity in the sugar-phosphate backbone of a nucleic acid, because the key to information transfer during replication is the complementarity of the nucleobases, which act as molecular recognition modules that are largely independent of the backbone structure (8, 9). For example, the transfer of information from DNA to RNA and back to DNA is catalyzed by RNA polymerases and reverse transcriptases. Work from our laboratory has extended this concept to the transcription of DNA into a complementary threose nucleic acid (TNA) polymer (10–12). More generally, the formation of Watson–Crick base-paired duplexes between different nucleic acids suggests that information transfer should be possible between many different nucleic acids, including variants such as 3′-amino RNA and DNA (13–15), 2′-amino RNA and DNA (16, 17), 4′-3′ lyxopyranosyl-NA (18), glycerol nucleic acid (GNA) (19), TNA (20), and others. As long as two polymers can base-pair with each other, it seems reasonable that a mixed polymer containing monomers of each type could still engage in base-pair mediated replication. Not all structural variation will be tolerated—indeed it is well-known that certain monomer types act as chain terminators in replication reactions, e.g., the presence of L-ribonucleotides is incompatible with the chemical replication of RNA by standard D-ribonucleotides (21). Thus, the requirement for replication imposes restrictions on the composition of nucleotide mixtures that would allow for the emergence of evolving systems, but these requirements are far from absolute.

If replication can tolerate chemical heterogeneity, what about the requirements for function? Functional nucleic acids, including ribozymes, deoxyribozymes, aptamers, and riboswitches, achieve their catalytic and binding abilities through the formation of specifically folded structures (22). Formation of a stable structure allows functional groups from both nucleobases and the backbone to interact specifically with binding targets. The involvement of backbone functional groups in tertiary interactions as well as in ligand binding or catalysis suggests that attaining a heritable functional state imposes significantly greater requirements on backbone homogeneity than does replication. Because replication will shuffle the order of variable backbone units, the progeny of a nucleic acid strand that exhibits a particular fold and

Author contributions: S.G.T., N.Z., M.P.E., A.L., and J.W.S. designed research; S.G.T., N.Z., M.P.E., and A.L. performed research; S.G.T., N.Z., M.P.E., A.L., and J.W.S. analyzed data; and S.G.T., N.Z., and J.W.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: szostak@molbio.mgh.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1107113108/-/DCSupplemental.

function may not retain that fold and function. However, this conclusion is based entirely on molecules that have evolved, either in nature or in the laboratory, from homogeneous RNA or DNA molecules. The possibility remains that functional molecules might be able to evolve from heterogeneous nucleic acid populations.

We suggest several different ways in which functional nucleic acids with heterogeneous backbones [mosaic nucleic acids (MNA)] (23) could evolve through repeated cycles of replication and selection. Some folded structures may simply be unaffected by a particular type of chemical heterogeneity, and thus might form equally well whether made of one polymer, or a second polymer, or a mosaic of the two. Alternatively, a backbone functional group from one polymer might be required at only one or a few specific positions in the folded structure, in which case a significant fraction of backbone-scrambled progeny would retain activity. More interesting is the possibility that functional groups from one polymer would be required at one or more positions, whereas different functional groups from the second polymer would be required at other specific positions. In this case, neither polymer alone could access the functional structure, but a fraction of mosaic transcripts could be functional.

To test the idea that functional structures could evolve from mosaic nucleic acids, despite the presence of nonheritable variation in the sugar-phosphate backbone, we undertook the selection of RNA/DNA mosaic aptamers that recognize nucleotide ligands. We found that ATP- and GTP-binding aptamers emerged from mosaic libraries as easily as from homogeneous RNA or DNA libraries, although the aptamers resulting from the mosaic selections exhibited weaker ligand affinity.

Results

Synthesis of Transcripts Containing both Deoxy- and Ribonucleotides. We refer to nucleic acids containing a mixture of deoxy- and ribonucleotides in random order as MNA (23). The Y639F T7 RNA polymerase utilizes both ribonucleotide triphosphates (rNTPs) and deoxyribonucleotide triphosphates (dNTPs) as substrates for transcription (24). We generated MNA transcripts from template DNA (Fig. 1 *A* and *D*) in optimized Y639F T7 RNA polymerase transcription reactions that included all eight canonical dNTPs and rNTPs. To prepare MNA transcripts with equal proportions of deoxy- and ribonucleotides, we examined a series of dNTP:rNTP substrate ratios and measured the ratio of deoxy- to ribonucleotides (d/r ratio) in the resulting transcripts. We measured the d/r ratio by hydrolyzing each MNA sample in strong base, so as to cleave the strand on the 3′ -side of each ribonucleotide, but not deoxyribonucleotide. Digested products were analyzed by anion-exchange HPLC (Fig. 1 *B* and *C*); the chromatogram shows clear separation of mono-, di-, tri-, tetra- and pentanucleotide products, each of which contains one 3′-ribonucleotide (Fig. 1*E*). Mono-, di-, tri-, tetra- and pentanucleotide peaks derive 100%, 50%, 33%, 25%, and 20% of absorbance from ribonucleotides, respectively. Thus, the approximate sugar content in the mosaic transcript can be calculated based on the relative absorbance of these peaks, neglecting hypochromic shifts, by simply dividing the fraction of absorbance due to ribonucleotides by the total absorbance of the nucleic acid fragments (Table S1). We found that an input d/r ratio of 9:1 results in transcripts with a d/r ratio close to 1 (Fig. 1*E*).

In Vitro Selection Scheme and Progress of the Selection. ATP and GTP aptamers were isolated from a large pool of random sequence MNA transcripts by repeated cycles of affinity chromatography and amplification by RT-PCR followed by transcription of a new, enriched MNA pool. We began the first round of the in vitro selection with a library of 6.2×10^{14} 100-nt MNA transcripts. This library consisted of MNA transcribed from an equimolar mixture of two DNA templates, one containing a 64-nt random region, and one containing a designed stem loop flanked by two 24-nt random regions (25, 26) (Fig. 1*A*). For each round, PAGE-purified, ^{32}P- radiolabeled MNA was incubated with a precolumn of underivatized agarose beads to capture matrix-binding sequences (Fig. 2*A*). Flow-through from this column was incubated with ATP or GTP conjugated via the γ-phosphate to an agarose bead matrix. After extensive washing, aptamers were competitively eluted with buffer containing free ligand (Fig. 2*B*). Eluted MNA was reverse transcribed and amplified by PCR to generate template DNA for the transcription of MNA for the subsequent round of selection. Multiple rounds of selection yielded DNA sequence pools enriched for MNA sequences with ligand binding activity. A significant fraction (20–30%) of the input MNA was bound to the column and then eluted by free ligand at round 8 in both selections (Fig. 2 *B* and *C*). Every sequence cloned (of 69) from rounds 7 and 8 of the ATP selection contained two G-rich motifs that were identical to those previously seen in an ATP aptamer in vitro selection from a DNA library (27) (Fig. 3*A*, Table S2). Similarly, nearly all sequences cloned (146 out of 156) from rounds 7 and 8 of the GTP selection contained a single, novel G-rich motif (Fig. 4*A*, Table S2). To further characterize these MNA aptamers, we examined the binding specificity and affinity of individual sequences containing these motifs.

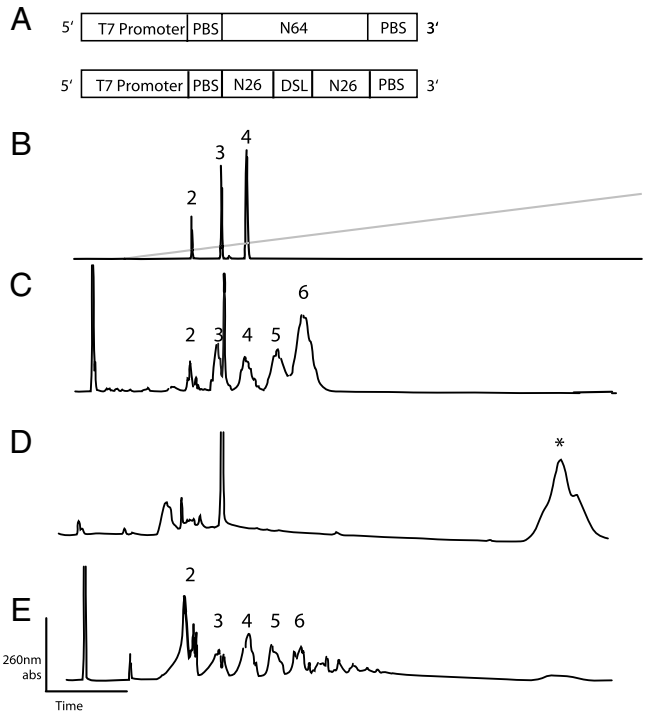

Fig. 1. Sequence and sugar content of the MNA starting pool. (*A*) Schematic of DNA template including a 5′ T7 promoter and random regions flanked by constant primer binding sites. Random regions consist of 64 degenerate bases (*Top*) or a designed stem loop (5′ CTCGCGAACGAG 3′) flanked on both sides by 26 degenerate bases (*Bottom*). (*B–E*) Anion-exchange HPLC chromatograms; absorbance at 260 nm (y axis) vs. time, 0–30 min (x axis). Peak numbers denote negative charge at pH 8.0. (*B*) 2, 5′-AMP; 3, 5′-ADP; 4, 5′-ATP. Gray line depicts 0–2 M KCl gradient. (*C*) Chemically synthesized degenerate DNA oligomer standards with no terminal phosphates. Peak 2, trinucleotide; 3, tetranucleotide; 4, pentanucleotide; 5, hexanucleotide; 6, heptanucleotide. (*D*) Full-length MNA, denoted by asterisk (*). (*E*) MNA base digests generated from a 9:1 (dNTP:rNTP) substrate ratio transcription reaction. Peak 2, ribonucleotide 2′/3′ monophosphate (r); 3, dinucleotide (d-r); 4, trinucleotide (d-d-r); 5, tetranucleotide (d-d-d-r); 6, pentanucleotide (d-d-d-d-r); 7, hexanucleotide (d-d-d-d-d-r), where d denotes deoxyribose and r denotes ribose.

Characterization of an ATP-Binding MNA Aptamer. Because MNA pools contain a heterogeneous combination of sugars in the back-

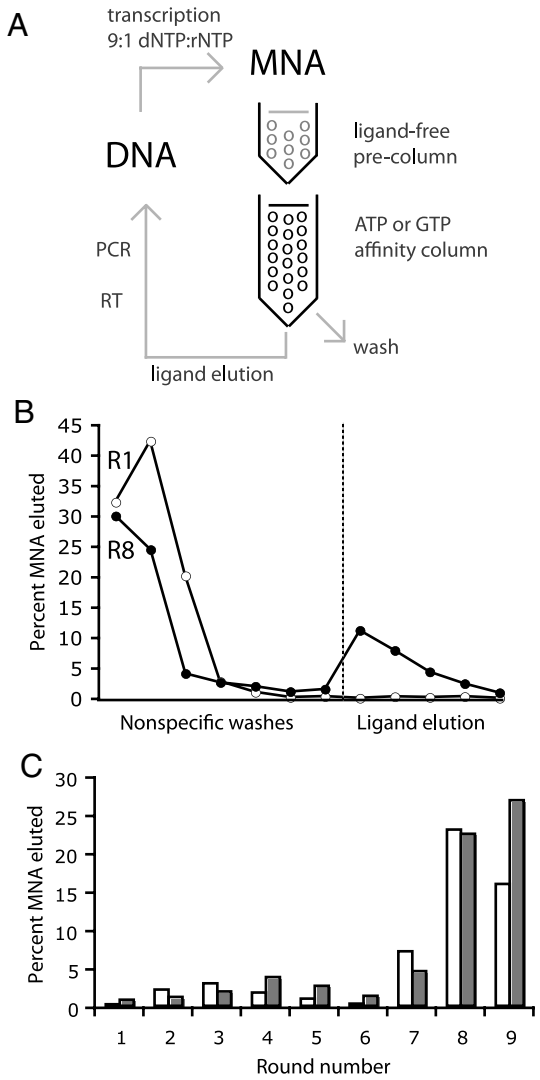

Fig. 2. In vitro selection scheme and progress. (*A*) Transcribed, ^{32}P- labeled MNA consisting of approximately 50% deoxy- and ribonucleotides was incubated with a ligand-free precolumn for 1 h. Flow-through from this column was incubated with either ATP- or GTP-derivatized agarose for 30 min. After a wash regime, aptamers were specifically eluted by four 30-min incubations with free ligand, then reverse-transcribed (RT) and PCR-amplified to generate the next pool of template DNA. (*B*) Elution profile from rounds 1 (○) and 8 (●) of the ATP aptamer selection. (*C*) Percent of input MNA eluted by free ligand washes for each round of the ATP (white) and GTP (gray) selections. In the first round, the precolumn retained approximately 10% of MNA; in subsequent rounds approximately 50% of total counts were retained.

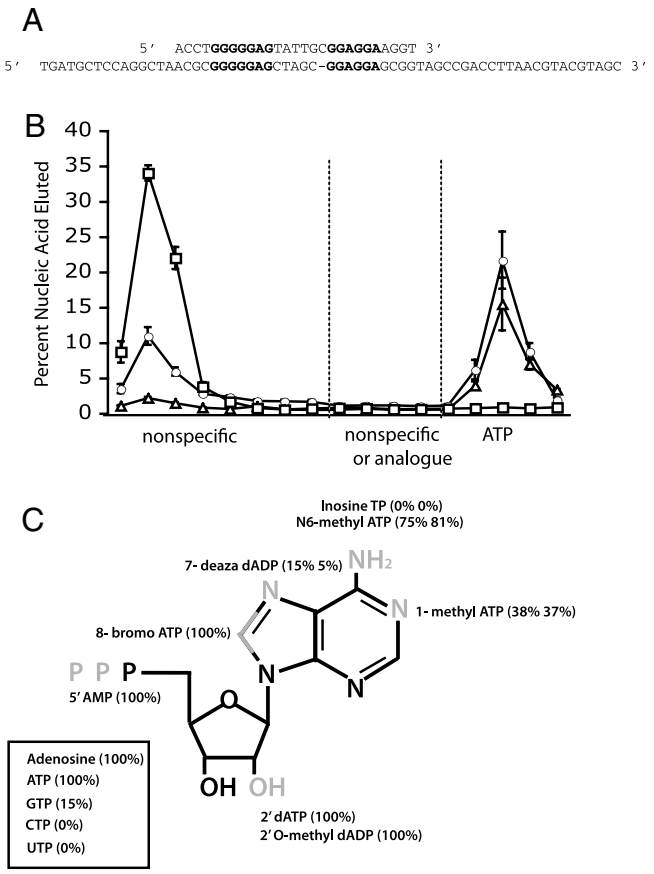

Fig. 3. MNA ATP aptamer 74 column binding assays. (*A*) Sequence of the previously identified DNA ATP aptamer (*Top*) and the MNA ATP aptamer 74 (*Bottom*), (for PBS sequences see Fig. S4, legend). Common aptamer sequence motifs are shown in bold. (*B*) Elution profile. MNA (○, $n = 6$), DNA (△, $n = 3$), or RNA (□, $n = 3$) was incubated with ATP-agarose (2 mM) for approximately 10 min. Fractions of one column volume were collected for a series of nonspecific and ATP washes (2 mM). Error bars represent SEM. (*C*) Specificity of the MNA and DNA ATP aptamer 74. Atomic positions (gray) recognized by the MNA or DNA ATP aptamer 74 were determined by the competitive elution of DNA or MNA bound to ATP-agarose with ATP analogues (2 mM). Total MNA or DNA (second value, when shown) eluted by each analogue wash followed by a subsequent ATP wash (internal control) is reported as 100%. The average percentage (of two independent experiments) of aptamer eluted by each analogue is shown; the range of values did not exceed 7%. In most cases, after all wash regimes, residual binding by the MNA to the derivatized agarose ranged from 15 to 40%.

bone, we first investigated whether deoxy- or ribonucleotides are required at any specific positions for ligand binding activity. We prepared DNA, MNA, and RNA versions of the full-length MNA ATP aptamer 74 sequence (Fig. 3*A*), and for each version we measured the fraction bound and specifically eluted from an ATP-agarose column (Fig. 3*B*). Whereas the MNA and DNA versions of the full-length MNA ATP aptamer 74 sequence had similar binding and elution activity, the RNA version exhibited no observable binding. These results imply that our selection conditions retrieved an aptamer sequence that requires deoxyribonucleotides, but not ribonucleotides, at one or more specific positions for activity.

To determine whether target recognition by the full-length MNA ATP aptamer exhibits typical aptamer-like specificity, we carried out competitive elution assays with several ATP analogues. In these experiments, aptamers bound to ATP-agarose were eluted first with buffer containing a nucleotide analogue, followed by buffer containing ATP to serve as an internal control (Fig. 3 *B* and *C*). In general, the MNA ATP aptamer 74 displays considerable molecular discrimination as it does not bind UTP or CTP, and has only slight affinity for GTP. Because the MNA ATP aptamer 74 binds adenosine, but does not recognize the α or β phosphates, the 2′ OH, or the C8 position of ATP, it is likely that most contacts are made on the nucleobase. These results, and the fact that the MNA ATP aptamer 74 and the previously isolated DNA ATP aptamer contain similar G-rich motifs suggest that they also share similar binding strategies.

Previous work showed that the DNA ATP aptamer recognizes the N1, N6, and N7 of the adenosine moiety (27). A solution structure of the DNA ATP aptamer indicates that these positions are important because of hydrogen bonding between these functional groups of ATP and the minor groove face of an invariant guanine in the G-rich aptamer bulge-loop, which forms a distorted but continuous duplex with a widened minor groove (28). Both DNA and MNA versions of the MNA ATP aptamer 74 sequence discriminate similarly against inosine triphosphate,

N6-methyl ATP, 1-methyl ATP, and 7-deaza dADP (Fig. 3*C*). These results suggest that the presence of ribonucleotides in the MNA ATP aptamer structure does not compromise the specificity of the aptamer.

To compare the ATP-binding activity of the MNA aptamer 74 to that of other known ATP-binding macromolecules, we sought to measure its binding affinity. Traditional methods for measuring binding affinity require large amounts of material and are problematic because of the inefficiency of MNA transcription (23, 29). To characterize the binding affinity of the MNA ATP aptamer 74, we employed two NMR-based ligand titration methods. Both the water-ligand observed gradient spectroscopy (waterLOGSY) NMR technique (30, 31) and the line-width method (31) require only 15–50 nmol of aptamer and are appropriate for detecting low micromolar to low millimolar binding interactions. In these methods, aptamer in the free and bound states can be determined by titration of ligand into a constant amount of aptamer until binding saturation is observed (*SI Text*). We measured the apparent K_d of the heterogeneous full-length MNA ATP 74 aptamer to be approximately 300–400 μM by each method (Table 1, Fig. S1– S3). This affinity is much weaker than that of biological adenine-binding riboswitches (32) and of previously in vitro-selected adenosine nucleic acid aptamers (27, 33), which tend to have nanomolar to low micromolar affinities for their targets. However, the apparent K_d of the ATP MNA aptamer 74 represents the average activity of the entire heterogeneous pool of nucleic acid.

We then identified the minimal sequence motifs that confer binding activity in the MNA ATP aptamer 74 sequence (Table S2). An MNA ATP aptamer 74 deletion mutant that lacks both DNA ATP aptamer sequence motifs has no binding activity observable by affinity chromatography (Fig. S4). We hypothesized that the previously identified 25-base DNA ATP aptamer sequence might be sufficient for MNA ATP-binding activity. To evaluate whether the MNA version of this minimal sequence is functional, we compared the binding affinity of two chemically synthesized DNA/RNA chimeric (chNA) versions of the ATP aptamer sequence. These DNA/RNA chNAs contain similar fractions of deoxy- and ribonucleotides fixed at arbitrary positions and in complementary orders (Table 1). The chNA ATP aptamer 74-1 has a K_d of approximately 400 μM whereas the chNA ATP aptamer 74-2 has a K_d of 744 μM. Because these binding affinities are similar to the full-length MNA ATP aptamer (apparent $K_d \sim 350$ μM), these results suggest that this minimal sequence confers the activity of the full-length MNA sequence (Table 1, Fig. S5). We also measured the binding affinities of DNA and RNA versions of this sequence. The DNA ATP aptamer has a K_d of 29 μM by this method (Table 1, Fig. S5), in general agreement with a previous estimation of $6 +/- 3$ μM (27). In contrast, the RNA version of this sequence has much weaker binding affinity of approximately 1.9 mM (Table 1, Fig. S5), consistent with previous results showing that it does not possess binding activity detectable by affinity chromatography (27) (Fig. 3*B*).

Characterization of a GTP-Binding MNA Aptamer. We examined the binding activity of DNA, RNA, and MNA versions of the GTP aptamer 812 by affinity chromatography (Fig. 4*A*); all versions displayed similar binding and elution profiles (Fig. 4*B*). Therefore, our selection identified an aptamer sequence that does not require any particular arrangement of deoxy- or ribonucleotides for binding.

To probe the specificity of this aptamer, we carried out competitive elution column binding assays with multiple GTP analogues (Fig. 4*C*). The MNA GTP aptamer 812 displays typical aptamer-like specificity, as it does not bind ATP, CTP, or UTP. Hydrogen bond formation by the Watson–Crick and Hoogsteen faces of guanine are likely critical for binding, as the aptamer is unable to bind analogues with alterations of these positions (Fig. 4*C*). Furthermore, the MNA GTP aptamer 812 binds all tested GTP analogues with modifications that do not disrupt hydrogen bonding of the nucleobase (e.g., deletion of phosphates, or substitutions at the 2′ or 3′ position). These results suggest that the MNA GTP aptamer 812 primarily recognizes the guanine nucleobase, a strategy employed by several families of RNA GTP aptamers (34).

To quantitatively characterize the binding of the full-length MNA GTP aptamer 812, we first attempted to measure its apparent K_d by the NMR waterLOGSY method, but were unable to detect a sufficiently strong waterLOGSY signal. We therefore turned to the line-width method, and measured an apparent K_d of 346 μM (Table 1, Fig. S6–S8). This affinity is much weaker than that of biological guanine-binding riboswitches (32) and many in vitro-selected GTP aptamers, although some aptamers of similar affinity have been characterized (25, 26). Upon titration of the MNA GTP aptamer 812 with increasing amounts of ATP, we observed no change in the line-width signal, indicating an absence of ATP-binding (Fig. S7).

Finally, because nearly every sequence from the GTP MNA selection contains an identical G-rich motif (Fig. 4*A*), we hypothesized that this motif is essential for binding activity. Indeed, deletion of the G-rich motif (5′ ATTAGGGGCGGCGGGGTGGGAT 3′) abolishes activity as determined by affinity chromatography (Fig. S9). However, chemically synthesized DNA, RNA, and DNA/RNA chNA versions of this sequence displayed K_ds that are approximatley 30 times higher than that of the full-length MNA GTP aptamer 812. This result suggests that unidentified motifs present in the full-length sequence contribute to full binding activity.

Discussion

We have identified two MNA nucleotide-binding aptamers by in vitro selection. Each selection round placed selective pressure for functional binding on MNA, which contained a mixed, nearly equivalent fraction of deoxyribose or ribose sugars in the sugar-phosphate backbone. As a consequence, binders that were dependent on fixed sugar arrangements were strongly selected against. Under these conditions, aptamer sequences that retain function despite variations in most or all sugar positions were retrieved.

Table 1. Aptamer dissociation constants

				Line width		waterLOGSY (H2)	
Aptamer	Ligand	Sequence	Nucleic acid	K_d (μM)	R^2	K_d (μM)	R^2
ATP 74	ATP	full-length	MNA	407 (H8) 302 (H2)	0.9194 0.9892	384	0.9914
74-1	ATP	CCTGGGGGAGTAUUGCGGAGGAAGG	chNA	—		399	0.9879
74-2	ATP	CCUGGGGGAGUAUUGCGGAGGAAGG	chNA	—		744	0.9908
	ATP	CCTGGGGGAGTATTGCGGAGGAAGG	DNA	—		29.2	0.9404
	ATP	CCUGGGGGAGUAUUGCGGAGGAAGG	RNA	—		1.95×10^3	0.9927
GTP 812	GTP	full-length	MNA	346 (H8)	0.9954	—	

Underlined positions in synthetic chimeric sequences represent ribonucleotides.

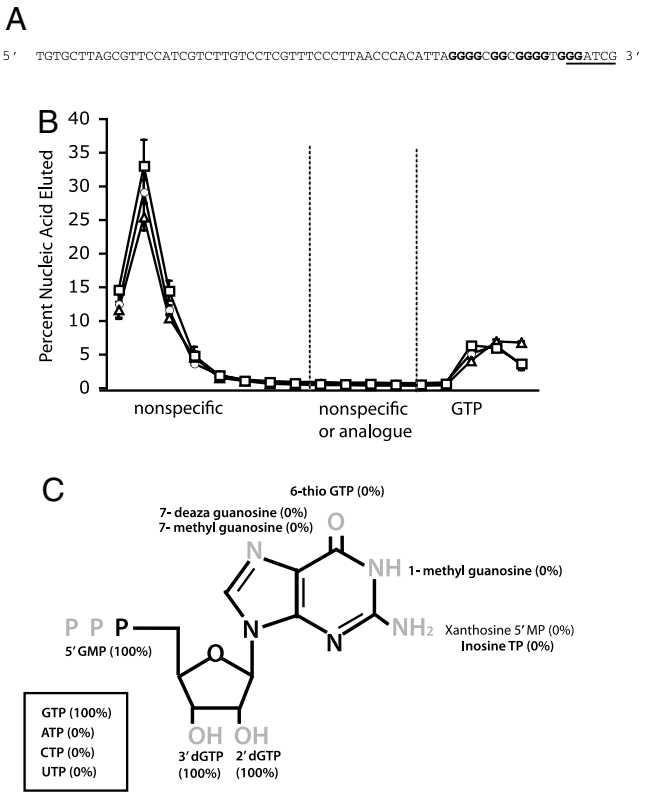

Fig. 4. MNA GTP aptamer 812 column binding assays. (*A*) DNA sequence of the MNA GTP aptamer 812 showing the G-rich motif (bold) and six complementary bases of the reverse PBS (underlined). (*B*) Elution profile. MNA (○, $n = 6$), DNA (△, $n = 3$), or RNA (□, $n = 3$) was incubated with GTP-agarose (5 mM) for approximately 10 min. Fractions of one column volume were collected for a series of nonspecific and GTP washes (5 mM). (*C*) Specificity of the MNA GTP aptamer 812. Atomic positions (gray) recognized by the MNA GTP aptamer 812 were determined by the competitive elution of MNA bound to GTP-agarose with GTP analogues (5 mM). Total MNA eluted by each analogue wash and a subsequent GTP wash (internal control) is reported as 100%. The average percentage (of two independent experiments) of aptamer eluted by each analogue is shown; the range of values did not exceed 5%. In most cases, after all wash regimes, background binding by the MNA to the derivatized agarose ranged from 15 to 40% of MNA.

We conclude that, far from representing an evolutionary dead-end, MNA could have provided a valuable source of heritable functionality for early organisms.

Our MNA ATP aptamer selection and a previous DNA ATP aptamer selection (27) retrieved the same adenosine-binding sequence motifs. This aptamer sequence may have been favored under our selection conditions because it contains two binding sites (28). Thus, multiple alterations to the sugar-phosphate backbone may be required to disrupt both opportunities for binding. In contrast, our conditions did not retrieve any sequences similar to a previously isolated RNA aptamer that binds ATP (33), as expected because this aptamer requires at least two specific 2′-hydroxyls at defined locations for detectable activity (35–37). Similarly, our selection for GTP aptamers from the MNA pool did not yield any sequences related to those identified in previous selections from an RNA pool (25, 26) suggesting that 2′-hydroxyls are required for activity in these aptamers.

We show that MNA aptamer sequences retain function despite the substitution of many, or all positions for deoxy- or ribonucleotides (Fig. 3*B*, Fig. 4*B*, Table 1). Other studies have indicated that similar or identical DNA and RNA sequences can share a common function. For example, in vitro selection studies have identified a DNA aptamer sequence that binds heme and retains some activity when synthesized as RNA (38). Similarly, DNA versions of some in vitro-selected RNA aptamers retain binding activity toward their targets (39). Additionally, a ribozyme ligase sequence was evolved into an active deoxyribozyme by only a few mutations (40). Finally, classical studies of various DNA-RNA chNA versions of the hammerhead ribozyme have demonstrated that some DNA substitutions can increase (41) or decrease activity (42). These results lend experimental support to genetic transfer hypotheses (43, 44). We suggest that MNA might facilitate such genetic transitions, e.g., as RNA-based cells evolved the ability to make DNA, they may have needed to tolerate the incorporation of deoxynucleotides into functional RNA molecules.

As our understanding of the prebiotic chemical processes that took place on the early Earth remains incomplete, it is difficult to assess the sugar compositions of early mosaic nucleic acids. Some testable constraints arise from the requirement for sequence-dependent duplex formation. DNA and RNA, as well as DNA/RNA chNA, readily form heteroduplexes (45), suggesting that DNA and RNA-based MNA can also form heteroduplexes. Not all pairs of nucleic acids can form heteroduplexes, for example GNA and TNA do not form heteroduplexes (20). Thus, MNA composed of similar fractions of GNA and TNA may not form duplexes, in which case GNA- and TNA-based MNA may not represent a viable informational biopolymer.

Other variant nucleotides that might be considered as potential components of a heterogeneous nucleotide pool include the threose nucleotides, arabino-nucleotides, and various amino-sugar nucleotides. However, the laboratory generation and amplification of such mosaic nucleic acid pools would require mutant DNA or RNA polymerase that could incorporate such nucleotides into polymers and transcribe them back into RNA or DNA. Some naturally occurring error-prone or bypass DNA polymerases as well as artificially evolved mutant polymerases show a remarkable ability to incorporate nucleotides with sugar modifications into polymers (46, 47). Such polymerases may enable the identification of additional types of structural variation that are tolerable in the evolution of functional molecules.

The binary MNAs we have described consist of 2^n distinct MNA molecules for each sequence, where n is the length of the sequence. It is likely that some arrangements of sugars disrupt tertiary structures necessary for functional activity. Future work will determine if these relaxed structures exist and if they might provide superior templates for copying reactions. MNA sequences may therefore strike an advantageous balance by providing functional activity as well as serving as enhanced templates for replication. More speculatively, we suggest that alternative sugar orders may even stabilize distinct folded structures with distinct biochemical activities (48, 49), thus allowing multiple functions to be coded by the same genetic information.

Materials and Methods

Y639F T7 RNA Polymerase Expression and Activity. Histidine-tagged Y639F T7 RNA polymerase encoded on a plasmid (courtesy of W.T. McAllister, Newark, NJ), was expressed in BL21 Star DE3 bacterial cells (Invitrogen) and purified as previously described (50) using Ni-nitrilotriacetate column chromatography.

MNA Transcription Reactions. Y639F T7 RNA polymerase MNA transcription buffer included Tris·HCl (40 mM, pH 8.0), $MgCl_2$ (30 mM), $MnCl_2$ (3.75 mM), nucleotides [16 mM total, 9:1 ratio of deoxy:ribo nucleotide triphosphates; 0.4 mM each rATP, rCTP, rGTP and rUTP; 3.6 mM each dATP, dGTP, dCTP and dTTP (Sigma-Aldrich)], DTT (20 mM), spermidine (3.6 mM), and 0.01% Triton X-100. High concentrations of DNA template (1–5 μM) were necessary for acceptable product yields. Reactions were carried out at 37 °C for 4–6 h. For selection and column binding assays, MNA was transcribed in the presence of a trace amount of ^{32}P-radiolabeled α-ATP (Perkin-Elmer).

Determination of Nucleotide Content in Transcripts. MNA purified by PAGE was heated to 50 °C in 0.5 M KOH for 3.5 h to hydrolyze ribonucleotide linkages. Products were analyzed by anion-exchange (DNAPac PA 100, Dionex) HPLC (Agilent) on a gradient (0–2 M KCl, Tris pH 8.0).

In Vitro Selection Procedure. Equal amounts of two DNA template libraries (25, 26) (Integrated DNA Technology, Fig. 1*A*) that contain either a fully randomized region (64 bases), or a designed stem loop flanked on each side by 26 bases of random sequence, were combined to generate a final sequence complexity of 6.2×10^{14}. Both templates yielded similar amount of MNA in each reaction (about four MNA copies of each sequence in the first round of each selection). PAGE-purified MNA transcribed from this pool was heated at 55 °C for 5–7 min in folding buffer that contained KCl (200 mM), $MgCl_2$ (5 mM), and MES (10 mM, pH 6.2) and allowed to cool. MNA was applied to a precolumn of Sepharose 4B (Sigma) for 1 h to remove matrix-binding sequences. Flow-through was incubated with either γ-phosphate-linked ATP (~2 mM) or GTP (~5 mM) agarose resin (Innova Bioscience) for 30 min. Subsequently, the column was washed with seven column volumes (200 μL each) of binding buffer. Aptamers were specifically eluted by four consecutive incubations (30 min each) with a column volume of elution buffer that contained KCl (200 mM), $MgCl_2$ (12 mM), MES (10 mM pH 6.2) and free ATP (2 mM) or GTP (5 mM). To remove free ATP or GTP, MNA was desalted on an NAP-5 column before ethanol precipitation. MNA was reverse transcribed (Ominiscript RT, Qiagen) and PCR-amplified by standard techniques (forward primer, 5′ TAATACGACTCACTATAGGGAGAGGAGAGAAACG3′; reverse primer, 5′ CATCGATGCTAGTCGTAACGATCC 3′). DNA pools from later rounds were cloned (One Shot Top 10), sequenced (SeqWright), and aligned (JalView) for analyses.

Affinity Column Binding Assays. Radiolabeled, folded nucleic acid (~100 pmol) was incubated with 200 μL of γ-phosphate-linked ATP (~2 mM) or GTP (~5 mM) agarose resin (Innova Bioscience) that had been equilibrated in binding buffer on a column for approximately 10 min. Resin was washed with eight column volumes of folding buffer, three column volumes of folding buffer (or analogue elution buffer), and three columns of ATP (or GTP) elution buffer. One column volume of folding buffer was used to clear the column after nucleotide elution buffer steps.

Aptamer Binding Affinity Studies by NMR. To prepare samples, MNA (PAGE-purified), DNA, RNA, and chNA aptamer (Integrated DNA Technology) samples were dissolved in binding buffer containing 10% D_2O (final volume 300 μL), heated at 55 °C for 5–7 min and cooled to room temperature. Aptamer dissociation constants were determined by titrating ATP or GTP into a constant amount of aptamer. The fraction of ligand-bound aptamer was determined by NMR line width or waterLOGSY methods (51, 30, 31, and *SI Text*).

ACKNOWLEDGMENTS. We thank T. McAllister for the generous gift of a plasmid containing histidine-tagged Y639F T7 RNA polymerase gene; G. Ruvkun, D.P. Bartel, A.M. van Oijen, R. Bruckner, A. Ricardo, I. Budin, and D. Treco for helpful discussions. This work was supported in part by a grant from the National Science Foundation. J.W.S. is an Investigator of the Howard Hughes Medical Institute.

1. Joyce GF (1989) RNA evolution and the origins of life. *Nature* 338:217–224.
2. Piccirilli JA (1995) Origin of life. RNA seeks its maker. *Nature* 376:548–549.
3. Joyce GF, Schwartz AW, Miller SL, Orgel LE (1987) The case for an ancestral genetic system involving simple analogs of the nucleotides. *Proc Natl Acad Sci USA* 84:4398–4402.
4. Engelhart AE, Hud NV (2010) Primitive genetic polymers. *Cold Spring Harb Perspect Biol* 2, doi: 10.1101/cshperspect.a002196.
5. Eschenmoser A (1999) Chemical etiology of nucleic acid structure. *Science* 284:2118–2124.
6. Powner MW, Gerland B, Sutherland JD (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* 459:239–242.
7. Joyce GF, Orgel LE (1993) Prospects for understanding the origin of the RNA world. *The RNA World*, eds RF Gesteland and JF Atkins (Cold Spring Harbor Laboratory Press, Plainview, NY), pp 1–25.
8. Benner SA, Ricardo A, Carrigan MA (2004) Is there a common chemical model for life in the universe? *Curr Opin Chem Biol* 8:672–689.
9. Benner SA, Hutter D (2002) Phosphates DNA, and the search for nonterrean life: A second generation model for genetic molecules. *Bioorg Chem* 30:62–80.
10. Ichida JK, Horhota A, Zou K, McLaughlin LW, Szostak JW (2005) High fidelity TNA synthesis by Therminator polymerase. *Nucleic Acids Res* 33:5219–5225.
11. Horhota A, et al. (2005) Kinetic analysis of an efficient DNA-dependent TNA polymerase. *J Am Chem Soc* 127:7427–7434.
12. Ichida JK, et al. (2005) An in vitro selection system for TNA. *J Am Chem Soc* 127:2802–2803.
13. Rothlingshofer M, et al. (2008) Chemical primer extension in seconds. *Angew Chem Int Ed Engl* 47:6065–6068.
14. Tohidi M, Zielinski WS, Chen CH, Orgel LE (1987) Oligomerization of 3′-amino-3′ deoxyguanosine-5′phosphorimidazolidate on a d(CpCpCpCpC) template. *J Mol Evol* 25:97–99.
15. Zielinski WS, Orgel LE (1985) Oligomerization of activated derivatives of 3′-amino-3′-deoxyguanosine on poly(C) and poly(dC) templates. *Nucleic Acids Res* 13:2469–2484.
16. Lohrmann R, Orgel LE (1976) Polymerization of nucleotide analogs I: Reaction of nucleoside 5′-phosphorimidazolides with 2′-amino-2′-deoxyuridine. *J Mol Evol* 7:253–267.
17. Schrum JP, Ricardo A, Krishnamurthy M, Blain JC, Szostak JW (2009) Efficient and rapid template-directed nucleic acid copying using 2′-amino-2′,3′-dideoxyribonucleoside-5′-phosphorimidazolide monomers. *J Am Chem Soc* 131:14560–14570.
18. Reck F, et al. (1999) L-alpha-lyxopyranosyl (4′ → 3′) oligonucleotides: A base-pairing system containing a shortened backbone. *Org Lett* 1:1531–1534.
19. Zhang L, Meggers E (2005) An extremely stable and orthogonal DNA base pair with a simplified three-carbon backbone. *J Am Chem Soc* 127:74–75.
20. Yang YW, Zhang S, McCullum EO, Chaput JC (2007) Experimental evidence that GNA and TNA were not sequential polymers in the prebiotic evolution of RNA. *J Mol Evol* 65:289–295.
21. Joyce GF, et al. (1984) Chiral selection in poly(C)-directed synthesis of oligo(G). *Nature* 310:602–604.
22. Wilson DS, Szostak JW (1999) In vitro selection of functional nucleic acids. *Annu Rev Biochem* 68:611–647.
23. Conrad F, Hanne A, Gaur RK, Krupp G (1995) Enzymatic synthesis of 2′-modified nucleic acids: Identification of important phosphate and ribose moieties in R Nase P substrates. *Nucleic Acids Res* 23:1845–1853.
24. Sousa R, Padilla R (1995) A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J* 14:4609–4621.
25. Carothers JM, Oestreich SC, Davis JH, Szostak JW (2004) Informational complexity and functional activity of RNA structures. *J Am Chem Soc* 126:5130–5137.
26. Davis JH, Szostak JW (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc Natl Acad Sci USA* 99:11616–11621.
27. Huizenga DE, Szostak JW (1995) A DNA aptamer that binds adenosine and ATP. *Biochemistry* 34:656–665.
28. Lin CH, Patel DJ (1997) Structural basis of DNA folding and recognition in an AMP-DNA aptamer complex: Distinct architectures but common recognition motifs for DNA and RNA aptamers complexed to AMP. *Chem Biol* 4:817–832.
29. Yang JH, Usman N, Chartrand P, Cedergren R (1992) Minimum ribonucleotide requirement for catalysis by the RNA hammerhead domain. *Biochemistry* 31:5005–5009.
30. Dalvit C, Fogliatto G, Stewart A, Veronesi M, Stockman B (2001) WaterLOGSY as a method for primary NMR screening: Practical aspects and range of applicability. *J Biomol NMR* 21:349–359.
31. Fielding L, Rutherford S, Fletcher D (2005) Determination of protein—ligand binding affinity by NMR: Observations from serum albumin model systems. *Magn Reson Chem* 43:463–470.
32. Kim JN, Breaker RR (2008) Purine sensing by riboswitches. *Biol Cell* 100:1–11.
33. Sassanfar M, Szostak JW (1993) An RNA motif that binds ATP. *Nature* 364:550–553.
34. Carothers JM, Oestreich SC, Szostak JW (2006) Aptamers selected for higher-affinity binding are not more specific for the target ligand. *J Am Chem Soc* 128:7929–7937.
35. Dieckmann T, Butcher SE, Sassanfar M, Szostak JW, Feigon J (1997) Mutant ATP-binding RNA aptamers reveal the structural basis for ligand binding. *J Mol Biol* 273:467–478.
36. Dieckmann T, Suzuki E, Nakamura GK, Feigon J (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA* 2:628–640.
37. Jiang F, Kumar RA, Jones RA, Patel DJ (1996) Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature* 382:183–186.
38. Travascio P, Bennet AJ, Wang DY, Sen D (1999) A ribozyme and a catalytic DNA with peroxidase activity: Active sites versus cofactor-binding sites. *Chem Biol* 6:779–787.
39. Lauhon CT, Szostak JW (1995) RNA aptamers that bind flavin and nicotinamide redox cofactors. *J Am Chem Soc* 117:1246–1257.
40. Paul N, Springsteen G, Joyce GF (2006) Conversion of a ribozyme to a deoxyribozyme through in vitro evolution. *Chem Biol* 13:329–338.
41. Taylor NR, Kaplan BE, Swiderski P, Li H, Rossi JJ (1992) Chimeric DNA-RNA hammerhead ribozymes have enhanced in vitro catalytic efficiency and increased stability in vivo. *Nucleic Acids Res* 20:4559–4565.
42. Perreault JP, Wu TF, Cousineau B, Ogilvie KK, Cedergren R (1990) Mixed deoxyribo- and ribo-oligonucleotides with catalytic activity. *Nature* 344:565–567.
43. Dworkin JP, Lazcano A, Miller SL (2003) The roads to and from the RNA world. *J Theor Biol* 222:127–134.
44. Gesteland RF, Cech TR, Atkins JF (2006) *The RNA World, Third Edition* (Cold Spring Harbor Laboratory Press, Plainview, NY), pp 46–47.
45. Shaw NN, Arya DP (2008) Recognition of the unique structure of DNA:RNA hybrids. *Biochimie* 90:1026–1039.
46. Henry AA, Romesberg FE (2005) The evolution of DNA polymerases with novel activities. *Curr Opin Biotechnol* 16:370–377.
47. Brudno Y, Liu DR (2009) Recent progress toward the templated synthesis and directed evolution of sequence-defined synthetic polymers. *Chem Biol* 16:265–276.
48. Landweber LF, Pokrovskaya ID (1999) Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: The spandrels of RNA evolution. *Proc Natl Acad Sci USA* 96:173–178.
49. Schultes EA, Bartel DP (2000) One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science* 289:448–452.
50. He B, et al. (1997) Rapid mutagenesis and purification of phage RNA polymerases. *Protein Expression Purif* 9:142–151.
51. Ludwig C, Guenther UL (2009) Ligand based NMR methods for drug discovery. *Front Biosci* 14:4565–4574.