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

Simon G. Trevinoa, Na Zhangb, Mark P. Elenkoc, Andrej Lupták, and Jack W. Szostakd,e

aHoward Hughes Medical Institute, Center for Computational and Integrative Biology, and Department of Molecular Biology, Simches Research Center, Massachusetts General Hospital, Boston, MA 02144; and bDepartments 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.

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
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 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. 1A 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. 1B and C); the chromatogram shows clear separation of mono-, di-, tri-, tetra- and pentanucleotide products, each of which contains one 3′-ribonucleotide (Fig. 1E). Mono-, di-, tri-, tetra- and pentanucleotide peaks derive 100%, 50%, 33%, 25%, and 20% of absorbance from ribonucleotides, respectively. Thus, the approximated 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 the 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. 1E).

**In Vitro Selection Scheme and Progress of the Selection.** ATP and GTP aptamers were isolated from a large pool of random sequence DNA 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. 1A). For each round, PAGE-purified, 32P-labeled MNA was incubated with a precolumn of underivatized agarose beads to capture matrix-binding sequences (Fig. 2A). 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. 2B). 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. 2B and C). Every sequence cloned (of 69) from round 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. 3A, 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. 3A, Table S2). To further characterize these MNA aptamers, we examined the binding specificity and affinity of individual sequences containing these motifs.

**Characterization of an ATP-Binding MNA Aptamer.** Because MNA pools contain a heterogeneous combination of sugars in the back-
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. 3A), and for each version we measured the fraction bound and specifically eluted from an ATP-agarose column (Fig. 3B). 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. 3B 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,

---

**Fig. 2.** In vitro selection scheme and progress. (A) Transcribed, C3p-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 DNA or MNA (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%.
N6-methyl ATP, 1-methyl ATP, and 7-deaza dADP (Fig. 3C). 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. Indeed, the MNA ATP aptamer 74 displayed $K_d$ 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. 4A), we hypothesized that this motif is essential for binding activity. Indeed, deletion of the G-rich motif (5′ ATTAGGGGGC...GTTGGAT 3′) abolishes activity as determined by affinity chromatography (Fig. S9). However, chemically synthesized DNA, RNA, and DNA/RNA chimeric versions of this sequence displayed $K_d$ that are approximately 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.

**Table 1. Aptamer dissociation constants**

<table>
<thead>
<tr>
<th>Aptamer</th>
<th>Ligand</th>
<th>Sequence</th>
<th>Nucleic acid</th>
<th>Line width $K_d (\mu M)$</th>
<th>$R^2$</th>
<th>waterLOGSY (H2) $K_d (\mu M)$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP 74</td>
<td>ATP</td>
<td>full-length</td>
<td>MNA</td>
<td>407 (H8)</td>
<td>0.9194</td>
<td>384</td>
<td>0.9914</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>302 (H2)</td>
<td>0.9892</td>
<td></td>
<td></td>
</tr>
<tr>
<td>74-1</td>
<td>ATP</td>
<td>CCTGGGGGGAUGTAAUUGCGGAGGAAGG</td>
<td>chNA</td>
<td>—</td>
<td>399</td>
<td>0.9879</td>
<td></td>
</tr>
<tr>
<td>74-2</td>
<td>ATP</td>
<td>CCGGGGGAUGTAAUUGCGGAGGAAGG</td>
<td>chNA</td>
<td>—</td>
<td>744</td>
<td>0.9908</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ATP</td>
<td>CCTGGGGGGAUGTAAUUGCGGAGGAAGG</td>
<td>DNA</td>
<td>—</td>
<td>29.2</td>
<td>0.9404</td>
<td></td>
</tr>
<tr>
<td>GTP 812</td>
<td>GTP</td>
<td>full-length</td>
<td>MNA</td>
<td>346 (H8)</td>
<td>0.9954</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

Underlined positions in synthetic chimeric sequences represent ribonucleotides.

**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. 4A); all versions displayed similar binding and elution profiles (Fig. 4B). 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. 4C). 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. 4C). 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. 4A), we hypothesized that this motif is essential for binding activity. Indeed, deletion of the G-rich motif (5′ ATTAGGGGGC...GTTGGAT 3′) abolishes activity as determined by affinity chromatography (Fig. S9). However, chemically synthesized DNA, RNA, and DNA/RNA chimeric versions of this sequence displayed $K_d$ that are approximately 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.
some activity when synthesized as RNA (38). Similarly, DNA ver-
mon function. For example, in vitro selection studies have iden-
similar or identical DNA and RNA sequences can share a com-
mentary bases of the reverse PBS (underlined). (\[equation\])
In most cases, after all wash regimes, background binding by the MNA to the
eluted by each analogue is shown; the range of values did not exceed 5%.
The average percentage (of two independent experiments) of aptamer
gue wash and a subsequent GTP wash (internal control) is reported as 100%.
aptamer 812 were determined by the competitive elution of MNA bound to
MNA GTP aptamer 812. Atomic positions (gray) recognized by the MNA GTP
bone may be required to disrupt both opportunities for binding.
sites (28). Thus, multiple alterations to the sugar-phosphate back-
aptamer selection (27) retrieved the same adenosine-binding
We conclude that, far from representing an evolutionary dead-
end, MNA could have provided a valuable source of heritable
biochemical activities (48, 49), thus allowing multiple functions to
be coded by the same genetic information.
May be required to disrupt both opportunities for binding.
In contrast, our selection for GTP aptamers from the MNA pool

taken place on the early Earth remains incomplete, it is difficult to
assess the sugar compositions of early mosaic nucleic acids. Some
table 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 en-
able 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 se-
quences may therefore strike an advantageous balance by provid-
ing 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 32P-radiolabeled 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. 1A) 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 selection complexity of 6.2 x 10^{15}. Both templates yielded similar amount of DNA in each reaction (the four DNA 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), MgCl₂ (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 (6.2 mM) or GTP (5.2 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), MgCl₂ (12 mM), MES (10 mM pH 6.2) and free and 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 (Omniscript RT, Qiagen) and PCR-amplified by standard techniques (forward primer, 5′-CATCGATGCTAGTCGTAACGATCC 3′; reverse primer, 3′-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 (6.2 mM) or GTP (5.2 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 binding 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% D₂O (final volume 300 μl), heated at 55°C for 5–7 min and cooled to room temperature. Aptamer concentration was 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 51Text).

ACKNOWLEDGMENTS. We thank T. MacAllister for the generous gift of a plasmid containing histidine-tagged Y639F T7 RNA polymerase gene; G. Ruvkun, D.P. Bartel, A.M. van Ojen, 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.

Supporting Information

Trevino et al. 10.1073/pnas.1107113108

SI Text

Aptamer Dissociation Constant Determination by NMR. The binding equilibrium of the aptamer and ligand can be described by the simple model: \( A + nL = AL_n \), where \( A \) is the unbound aptamer, \( L \) is the free ligand, and \( n \) is the number of binding sites. The microscopic dissociation constant, \( K_d^{[n]} = [A][L]^n/[AL_n] \), where \( h \) is the Hill coefficient, which approaches the value of \( n \) for fully cooperative binding. Because the number of GTP aptamer binding sites remains unknown, one binding site was assumed. Additionally, the DNA ATP aptamer is known to have two binding sites (1). Thus, for the calculation of free ligand concentration in the ATP aptamer titrations (see below), we assumed two binding sites to conserve proper stoichiometry. However, because our data indicate that the extent of ligand cooperativity is small, a Hill coefficient of 1 was assumed (indicating no negative or positive cooperativity). In this case, \( K_d^{[1]} \) is equal to the apparent dissociation constant, \( K_d \), which can be determined by plotting complex \( AL_n \) formation as a function of free ligand concentration.

Dissociation Constant Determination by Water-Ligand Observed Gradient Spectroscopy (waterLOGSY). The waterLOGSY signal intensities of titrated ligand (ATP or GTP), starting at a concentration of 80–160 \( \mu \)M, were recorded both in the absence or presence of a constant amount of aptamer (56–300 \( \mu \)M). The hydrated free ligand generates a weak, positive NOE, whereas bound ligand within ligand–aptamer complex generates a strong, negative NOE. Because the exchange rate between bound and free state ligand is faster than the NMR timescale, the signal resonances of bound and free state ligand are merged together. Therefore, the directly observed waterLOGSY signal is a combined weighted average resulting from both the free and bound ligand NOE signals at equilibrium.

To eliminate the contribution from free ligand and obtain the waterLOGSY signal intensity due exclusively to bound ligand—which is proportional to the formation of complex \( AL_n \)—the waterLOGSY signal intensities recorded in the absence of aptamer were subtracted from those recorded in the presence of aptamer. Nonlinear fitting of the data was executed with Prism 5.0 (GraphPad Software Inc.) according to the Hill equation, Eq. S1:

\[
I_{\text{waterlogsy}} = I_{\text{max}} \left( \frac{[L]_{\text{free}}}{K_d^{[1]} + [L]_{\text{free}}} \right)^h. \tag{S1}
\]

where \( I_{\text{waterlogsy}} \) is the intensity of NOE signal attributed to bound ligand (arbitrary units), \( I_{\text{max}} \) is the saturation level of the signal (corresponding to 100% \( AL_n \) complex formation), \( h \) is the Hill coefficient, \( K_d^{[1]} \) is the microscopic dissociation constant, and \( [L]_{\text{free}} \) is the concentration of ligand in the free state and can be calculated by using Eq. S2:

\[
[L]_{\text{free}} = [L]_{\text{total}} - n[A]I_{\text{waterlogsy}}/I_{\text{max}}. \tag{S2}
\]

where \([L]_{\text{total}}\) is the total concentration of ligand, \([A]\) is the total concentration of aptamer, and the ratio of \( I_{\text{waterlogsy}}/I_{\text{max}} \) represents the fractional saturation level.

Starting with an approximated \( I_{\text{max}} \) value, \([L]_{\text{free}}\) can be estimated from Eq. S2. \([L]_{\text{free}}\) and \( I_{\text{waterlogsy}}\) are used as inputs to non-linearly fit the Hill equation, Eq. S1 thereby resulting in a more accurate \( I_{\text{max}} \) value as well as an estimated \( K_d \) value. This process can be iterated several times until the \( K_d \) and \( I_{\text{max}} \) values are in fair agreement as indicated by the \( R^2 \) value.

Dissociation Constant Determination by \( ^1H \) NMR Line Width. Watergate 1D \(^1\)H watergate NMR spectra were recorded for ligand titration series in the presence of a constant amount of aptamer (corresponding to aptamer concentrations of 60–80 \( \mu \)M). For each ligand concentration, the line width of the NMR signal attributed to the nucleotide H8 or H2 was measured at half peak maximum. The observation of a consistently sharp internal reference line width signal indicated satisfactory field homogeneity for each titrated sample. In this method, broadening of the H2 or H8 ligand signal is indicative of an increased transverse relaxation rate (\( R_2 \)) adopted by the ligand while complexed with the aptamer. Attenuation of line broadening occurs as more ligand remains in the unbound state (2). Because of relatively weak binding (micromolar), the exchange rate between the bound and free states is much faster than the NMR timescale, which causes the ligand signal resonances due to both bound and free states to merge. Fast exchange does not significantly contribute to extra line width broadening and may be neglected.

Thus, the observed line width signal is related to \( R_{2ob} = F_f \times R_{2b} + F_f \times R_{2g} \), where \( F_f \) and \( F_f \) are the fraction of ligand in the free and bound states, respectively, and \( R_{2b} \) and \( R_{2g} \) are the transverse relaxation rates of ligand in the free and bound states, respectively. This equation can be arranged to \( F_f = (R_{2ob} - R_{2g})/(R_{2b} - R_{2g}) \), to reflect the fraction of bound ligand, \( R_{2b} \) was determined by a titration of ligand in the absence of aptamer.

The dissociation constant (\( K_d \)) can be determined by plotting changes in the observed line width of NMR signal as a function of the free concentration of ligand, according to the Hill equation:

\[
\theta = \frac{[L]_{\text{free}}}{(K_d^{[1]} + [L]_{\text{free}})^h},
\]

where \( \theta \) is the fraction of aptamer that is bound to ligand, or:

\[
\theta = F_f[L]_{\text{total}}/[A]_{\text{total}}.
\]

where \([A]_{\text{total}}\) is the total concentration of aptamer, \( n \) is the number of binding sites, and \([L]_{\text{free}}\) is the concentration of ligand in the free state. Because ligand concentration is in excess to aptamer, the total concentration of ligand, \([L]_{\text{total}}\) is \( [L]_{\text{free}}\). Additionally, a Hill coefficient of 1 is assumed such that \( K_d^{[1]} \) is approximately equal to \( K_d \).

Thus, we obtain the equation:

\[
R_{2ob} = (R_{2b} + n[A](R_{2b} - R_{2g})/(K_d + [L]_{\text{free}})). \tag{S3}
\]

This approximation allows for the calculation of \( K_d \) and \( R_2 \) by iteratively refitting the data (Eq. S3) until \( K_d \) and \( R_2 \) values are in fair agreement as indicated by the \( R^2 \) value. Nonlinear fitting of the data was executed with Prism 5.0 (GraphPad Software Inc.).


Fig. S1. NMR experimental data for the titration of the mosaic nucleic acid (MNA) ATP aptamer 74 with ATP. One-dimensional $^1$H NMR waterLOGSY spectra of a series of ATP ligand concentrations (indicated) generated in the presence (A–J) of approximately 80 μM MNA ATP aptamer 74. The expanded spectral region that contains ATP aromatic nucleobase proton signals is shown. The signal intensities are on a uniform arbitrary scale throughout the titration.

Fig. S2. NMR experimental data for the titration of the mosaic nucleic acid (MNA) ATP aptamer 74 with ATP. One-dimensional $^1$H NMR waterLOGSY spectra of a series of ATP ligand concentrations (indicated) generated in the absence (K–N) of approximately 80 μM MNA ATP aptamer 74. The expanded spectral region that contains ATP aromatic nucleobase proton signals is shown. The signal intensities are on a uniform arbitrary scale throughout the titration.
Fig. S3. Determination of the MNA ATP aptamer 74 dissociation constant for ATP. (A) The calibrated change in relative waterLOGSY intensity for the ATP H2 proton as a function of increasing total ligand concentration (●) calculated by subtracting the observed signal acquired in the absence of aptamer (□) from observed signals acquired in the presence of aptamer (▵). Lowercase letters correspond to experimental data in Fig. S1 and Fig. S2. (B) The change in fractional occupancy as a function of free ligand concentration, calculated by nonlinear fitting of the experimental data of the change in calibrated relative waterLOGSY intensity against ATP concentrations (see Materials and Methods).

Fig. S4. Mosaic nucleic acid (MNA) ATP aptamer 74 column binding assays. (A) DNA sequence of the MNA ATP aptamer 74 flanked by the PBS regions GGGA-GAGGAGAAACG and GGATCGTTACGACTAGCATCGATG (5′ to 3′). The previously identified DNA ATP aptamer motifs are shown in bold. Bases deleted in the MNA ATP aptamer 74 deletion mutant are underlined. (B) The full-length MNA ATP aptamer 74 sequence (black, n = 6), or the MNA ATP aptamer 74 deletion sequence (white, n = 3) was incubated with ATP agarose (2 mM) for 5–15 min, washed with 12 column volumes of binding buffer followed by three column volumes of buffer containing free ATP ligand (2 mM). Percent of MNA eluted from the column by free ATP is shown. After all wash regimes, background binding by the MNA to the derivatized agarose ranged from 15 to 40%.
Fig. S5. ATP aptamer binding curves for ATP. Binding curves for (A) DNA, (B) RNA, (C) RNA/DNA 74-1, and (D) 74-2 versions of the minimal ATP aptamer sequence. Data collected by the waterLOGSY methodology (see Materials and Methods).

Fig. S6. NMR experimental data for the titration of the mosaic nucleic acid (MNA) GTP aptamer 812 with GTP. One-dimensional $^1$H spectra of a series of GTP (A–H) ligand concentrations (indicated) in the presence of approximately 80 µM MNA GTP aptamer 812. The expanded spectral region that contains aromatic nucleobase proton signals is shown. The peak line widths are as labeled. The height of displayed signals is normalized to the same level.
Fig. 57. NMR experimental data for the titration of the mosaic nucleic acid (MNA) GTP aptamer 812 with ATP. One-dimensional $^1$H spectra of a series of ATP (I–O) ligand concentrations (indicated) in the presence of approximately 80 $\mu$M MNA GTP aptamer 812. The expanded spectral region that contains aromatic nucleobase proton signals is shown. The peak line widths are as labeled. The height of displayed signals is normalized to the same level.
Fig. S8. Mosaic nucleic acid (MNA) GTP aptamer 812 dissociation constant determination for GTP. (A) Observed NMR line width upon titration with GTP (●) or ATP (▵). Lowercase letters correspond to experimental data in Fig. S6 and Fig. S7. (C) The fraction of bound ligand as a function of increasing free GTP concentration was determined by nonlinear fitting to the observed data (see Materials and Methods).

Fig. S9. Mosaic nucleic acid (MNA) GTP aptamer 812 column binding assays. (A) DNA sequence of the MNA GTP aptamer 812 including six bases complement to the reverse PBS (italic). The G-rich 812-1 sequence that is deleted in the MNA GTP aptamer 812 deletion is underlined. (B) The full-length MNA GTP aptamer 812 sequence (black, n = 6), or the MNA GTP aptamer 812 deletion sequence (white, n = 3) was incubated with GTP agarose (5 mM) for 5–15 min, washed with 12 column volumes of binding buffer followed by three column volumes of buffer containing free GTP ligand (5 mM). Percent of MNA eluted from the column by free GTP is shown. After all wash regimes, background binding by the MNA to the derivatized agarose ranged from 15 to 40%.
Table S1. Determination of the approximate fraction of ribose and deoxyribose in mosaic nucleic acid (MNA)

<table>
<thead>
<tr>
<th>Nucleotide fragment size</th>
<th>Ratio (d/r)</th>
<th>% total fragments</th>
<th>% RNA</th>
<th>RNA in full-length, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Starting pool</td>
<td>1:1</td>
<td>73</td>
<td>3.3</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>13.5</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.7</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Starting pool</td>
<td>9:1</td>
<td>73</td>
<td>3.3</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Aptamer 74</td>
<td>9:1</td>
<td>24</td>
<td>0.8</td>
<td>44.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.3</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.3</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>Aptamer 812</td>
<td>9:1</td>
<td>28</td>
<td>0.8</td>
<td>50.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>21</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

MNA was digested by potassium hydroxide to cleave MNA at ribonucleotide linkages under conditions that fully digest an RNA transcript to mononucleotides but do not digest a DNA primer. The resultant product is a complex pool of nucleotide fragments that each contain one ribonucleotide. Because each fragment is characterized by a different charge, nucleotide monomers, dimers, trimers, and so forth can be resolved by HPLC (260 nm absorbance), where RNA and DNA absorb similarly. Thus, absorbance due to ribonucleotides for the monomer fraction is 100% whereas absorbances due to ribonucleotides for dimer and trimer fractions are 50% and 33%, respectively. Based on the relative absorbance of all observed fractions, it is possible to estimate the amount of RNA in the original MNA transcript. In all cases, peaks that eluted later than pentamers had combined absorbances (<5%) and were discounted from calculations.

Table S2. Sequences of the ATP and GTP mosaic nucleic acid (MNA) aptamers obtained from in vitro selections in this study

<table>
<thead>
<tr>
<th>Selection</th>
<th>Name</th>
<th>Sequence</th>
<th>Copy no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>74</td>
<td>GTGGCAGCGGTAGCGGGGGAGTGTCGCTTCGCGAGGGAGGAGCCACAGGAGCCGCTCC</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCCCTAGAGGAAGAATAGTACGGCAGGGGAGACATGCGGGGGGAGACGAGTACCTGCTCTAG</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGATGCTCCAAGTTAACCGGGGGAGGTACGGAGAGCAGTGGAGCCACCTTACGTAAGC</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATCTCTACGCTCTGCAGGGGAGCAGTCGGCAGGGGGGAGAGTCTACATGTAGT</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGGAGGAGGAGGATAGTTGGGAGAAGATGTCTCCTGGGGGAGCTAATCGAATCTCTGGTGTC</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>812</td>
<td>TACGCTCCAAGGAGAATCCTGCTCCGAGGGGAGGAGGCTTACGGGCTGTCGGGGGAGC</td>
<td>3</td>
</tr>
<tr>
<td>GTP</td>
<td></td>
<td>CCTTACGCTCTACACCTGGAGAACTACCTGGAATACACCGGGAAGGCGAGGGGGGGAT</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TCGAGATTCGCTATCTATTGGAAGTACCCGAAGGAGGAGGAGGAGGAGGAGGAT</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGTACTGGATGCTATCACGATTCCTGTGTTATCTATCTATACCTATGGGAGGGGGGGATTGAT</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGGTGTGCTCAGAACGATTGGGGCGAGATGGGCTAGTACGTCGGCACACCCCTGGAT</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTCTGCTGCAGACGCTGCGCAGGAGACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAT</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCTTACGCTCTCACCCATTGCACACTGCTAGAATACACCGGGAAGGCGAGGGGGGGAT</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CCTTACGCTCTACACCTGGAGAACTACCTGGAATACACCGGGAAGGCGAGGGGGGGAT</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGGAGGAGGAGGATAGTTGGGAGAAGATGTCTCCTGGGGGAGCTAATCGAATCTCTGGTGTC</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ACACATCCAGGAGAAGAATCCTGCTCCGAGGGGAGGAGGCTTACGGGCTGTCGGGGGAGC</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TACGCTCCAAGGAGAATCCTGCTCCGAGGGGAGGAGGCTTACGGGCTGTCGGGGGAGC</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATCTCTACGCTCTGCAGGGGAGCAGTCGGCAGGGGGGAGAGTCTACATGTAGT</td>
<td>3</td>
</tr>
</tbody>
</table>

The region shown corresponds to the initial random region of the library. Only sequences for which three or more isolates were obtained are shown. Common G-rich motifs for the ATP aptamer sequences (GGGGAG and GGAGGAG) and for the GTP selection (GGGG, GG, GGGG, GG) are shown in bold. Partial reverse primer binding sites are underlined.