

Selection and evolution of NTP-specific aptamers

Laure Weill, Dominique Louis and Bruno Sargueil*

Centre de Génétique Moléculaire, CNRS, Avenue de la Terrasse, 91190 Gif sur Yvette, France

Received March 29, 2004; Revised April 29, 2004; Accepted September 1, 2004

ABSTRACT

ATP occupies a central position in biology, for it is both an elementary building block of RNA and the most widely used cofactor in all living organisms. For this reason, it has been a recurrent target for *in vitro* molecular evolution techniques. The exploration of ATP-binding motifs constitutes both an important step in investigating the plausibility of the 'RNA world' hypothesis and a central starting point for the development of new enzymes. To date, only two RNA motifs that bind ATP have been characterized. The first one is targeted to the adenosine moiety, while the second one recognizes the 'Hoogsteen' face of the base. To isolate aptamers that bind ATP in different orientations, we selected RNAs on an affinity resin that presents ATP in three different orientations. We obtained five new motifs that were characterized and subsequently submitted to a secondary selection protocol designed to isolate aptamers specific for cordycepin. Interestingly, all the ATP-binding motifs selected specifically recognize the sugar-phosphate backbone region of the nucleotides. Three of the aptamers show some selectivity for adenine derivatives, while the remainder recognize any of the four nucleotides with similar efficiency. The characteristics of these aptamers are discussed along with implications for *in vitro* molecular evolution.

INTRODUCTION

Since the development of SELEX, an *in vitro* molecular evolution procedure, much effort has been expended to isolate DNA and RNA motifs that bind to a given ligand. These ligand moieties range from simple ions to complex targets such as whole cells (1–3). A plethora of experiments in this field have provided potential molecular tools (3–5), insights into nucleic acid structure (6–9), models in molecular evolution and support for the 'RNA world' hypothesis (10). In a few cases, SELEX has also been used as a first step towards the development of new catalysts (11,12). For any of these purposes, it has been of particular interest to select for molecules that bind

nucleobases, nucleosides, nucleotides or their derivatives (13–23). The first such example of this was the selection of an RNA aptamer for ATP, isolated by Sassanfar and Szostak (18) and structurally characterized by NMR (24–26). Because the backbone conformation of this RNA is reminiscent of the greek letter zeta, it has been referred to as the 'zeta-shaped ATP aptamer'. This motif primarily recognizes the adenosine moiety and has been subsequently re-isolated in selection experiments to find aptamers to *S*-adenosylmethionine, *S*-adenosylhomocysteine and NAD⁺ (14,18,22,27). Recently, a motif very similar to the zeta-shaped aptamer has been identified in the pRNA of phi29 bacteriophage (28). The zeta-shaped motif was subsequently used as a starting point for developing self-kinasing RNA, catalysts that use ATP as a substrate. Five ribozymes that have the ability to self-kinase were isolated, but for most of these it is uncertain whether they have retained the zeta-shaped motif, and it is also uncertain whether the incorporation of ATP-binding motifs in the initial pool significantly improved the selection process (11). In fact, other ribozymes have been successfully isolated from totally random pools, in no more rounds of selection than were required to isolate ribozymes from pools that were biased to include a substrate-binding region (29–37). In this respect, it should be noted that, in the above-cited selection, only one substrate-binding domain was used, which was derived from an aptamer that presents the substrate in a single orientation and potentially buries a significant fraction of the ligand surface area. It could be that in order to select for novel ribozymes from aptamers, it is advantageous to present the ligand in a manner that mimics the solution state; i.e. in more than one orientation such that various modes of substrate binding can be explored.

In the process of selecting an RNA enzyme for which one of the substrates is ATP, we isolated new RNA motifs that bind ATP. In order to bypass the potential problems discussed above, our selection was performed under conditions that mimic ATP recognition in solution. This was achieved by using an affinity column that comprised a mixed population of ATP-derivatized agaroses, in which the selection moiety was coupled in three different orientations. The five different RNA aptamers that were thus obtained differ markedly from the previously known motifs. To evaluate the adaptability of these aptamers, the primary selected pool was submitted to secondary selections in order to generate more diversity, better understand the properties of these motifs and obtain aptamers that bind the adenosine analogue, cordycepin

*To whom correspondence should be addressed. Tel: +33 1 69 82 31 54; Fax: +33 1 69 82 43 86; Email: sargueil@cgm.cnrs-gif.fr
Present address:
Dominique Louis, EVOLOGIC, 2, rue Gaston Crémieux, 91000 EVRY, France

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

(3' deoxyadenosine). The specificity of each of these motifs was determined; each of them recognizes the phosphate-ribose moiety of ATP. While three of them show some selectivity for the base, one recognized any nucleotide triphosphate, and the last is targeted to the elementary building block of RNA backbone, a ribose monophosphate.

MATERIALS AND METHODS

Pool construction

A 104mer DNA oligonucleotide containing 60 random nucleotides flanked by two primer-binding sites (GGTTGGCAGC-AGAAGATAGCAN₆₀GTCGGTCAAGGGAGGGATCCTA) was synthesized on an 'Expedite' synthesizer (Perseptive Biosystems) using the following phosphoramidite mixture for the random positions: A/C/G/T:25/30/25/20. The synthesized 104mer was deprotected in 30% NH₄OH, precipitated and purified on a 6% denaturing polyacrylamide gel. An aliquot of 5 nmol of DNA templates were obtained and amplified by large-scale PCR using the following primers (Upper: GAT-AATACGACTCACTATAGGTTGGCAGCAGAAGATAGC-AG, Lower: TAGGATCCCTCCCTTGACCGAC). The upper primer introduces a T7 promoter sequence (underlined). A 100 ml PCR containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl₂, 0.07% NP40, 5% acetamide, 0.2 mM dNTPs, 500 U *Taq* DNA polymerase, 0.8 μM each primers and 0.5 nmol of DNA template was performed under the following conditions: 94°C, 3 min 30 s (5 min for the first cycle); 55°C, 5 min; 72°C, 7 min or 20 min for the last cycle. The reaction was stopped after six cycles. The pool complexity was estimated to be $\sim 3.5 \times 10^{14}$ sequences (38). The DNA was phenol and chloroform extracted, and ethanol precipitated.

The initial RNA pool was prepared by *in vitro* transcription using 100 μg of DNA template in total volume of 2 ml (40 mM Tris-HCl, pH 8, 25 mM MgCl₂, 1 mM spermidine, 4 mM NTPs, 10 mM DTT, 1000 U T7 RNA polymerase, 2 U pyrophosphatase) for 7 h at 37°C. The DNA template was then digested with 100 U of RQ1 DNase for 1 h at 37°C, the transcription mixture was phenol extracted, ethanol precipitated and purified on a 6% denaturing acrylamide gel. For the second round, the PCR and transcription reaction were carried out in 500 μl; for the following rounds, we followed the same procedure in 100 μl. Minimal motifs were transcribed on partial duplexes using two oligonucleotides as template (39).

Selection procedures

Primary selection. To minimize the enrichment of RNA molecules that bind the matrix, the pool was pre-selected on Tris-derivatized cyanogen bromide-activated agarose (0.4 ml) for the first four rounds. The selection column (0.8 ml) contains ~ 2 μmol of ATP linked to activated agarose through the 2'/3'hydroxyl, C8 or N6 position (Sigma). The selection and pre-selection columns were equilibrated and saturated with two volumes of binding buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 5 mM MgCl₂) + 1 μg/ml tRNA and 0.1 μg/ml BSA and subsequently washed with 5 vol of binding buffer. The RNA pool (the final concentration and the volume of RNA solution for each round is indicated in Table 1) was dissolved in a suitable volume of 50 mM Tris-HCl, pH 7.5, incubated at 80°C for 5 min. An aliquot of 150 mM NaCl and 5 mM MgCl₂

Table 1. Course of the ATP aptamer selection

Round	RNA quantity (nmol)	[RNA] (μM)	Pre-selection	Percentage of input RNA eluted with	
				GTP	ATP
1	11	2	Yes	—	ND
2	2.5	1	Yes	—	ND
3	0.35	0.2	Yes	—	0.3
4	0.35	0.2	Yes	—	4.7
5	0.2	0.1	No	—	25.3
6	0.2	0.1	No	28	<0.5
7	0.2	0.1	No	—	23

ND = Not determined.

(final) were added and the mixture was immediately incubated at 37°C for 10 min.

The RNA solution was loaded onto the pre-selection column and eluted with 2 column volumes of binding buffer. The eluted RNA was then applied to the ATP-agarose column. The column was then washed with 10 column volumes of binding buffer, and the elution of aptamers was performed with 4 column volumes of binding buffer + 5 mM ATP.

The eluted aptamers were ethanol precipitated in the presence of 5 μg of glycogen and 0.5 M ammonium acetate. The excess ATP was eliminated by gel filtration (5 ml of sephadex G-50), and the RNA was ethanol precipitated a second time.

RNA was reverse-transcribed in a final volume of 40 μl with Superscript II RT (Invitrogen) and PCR-amplified using DynaZyme II DNA polymerase (Finnzymes) in a total volume of 300 μl.

To enhance the specificity of the aptamers, a counter-selection with GTP was carried out during round 6. After loading the RNA pool, the column was washed with 10 vol of binding buffer, and eluted with binding buffer + 5 mM GTP before the ATP elution step.

Secondary selections. To perform a secondary selection at higher stringency, the concentration of ATP linked to agarose on the selection column (0.8 ml) was reduced to 1 nM (instead of 2.5 mM) and the RNA/ATP ratio on the column was 100/1. Two rounds of selection were carried out under these conditions.

Selection of the cordycepin (3' deoxyadenosine) aptamers was performed as described above for ATP. The RNA was pre-selected on 200 μl of a Tris-derivatized column and selected on a 600 μl cordycepin-derivatized cyanogen bromide-activated Sepharose (Amersham). Cordycepin coupling was monitored using ³²P-cordycepin triphosphate (NEN) on a separate aliquot. The resulting cordycepin agarose contains 0.36 μmol of cordycepin/ml. The course of the selection is summarized in Table 2.

Mutagenic PCR

Mutagenic PCR was performed as described by Vartanian *et al.* (40), in a total volume of 100 μl containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2.5 mM MgCl₂, 0.5 mM MnCl₂, 30 μM dCTP and dATP, 1 mM dTTP and dGTP, 1 μM each primers and 20 fmol of DNA template. The PCR was carried out for 40 cycles under the following conditions: 95°C, 30 s (5 min, for the first cycle); 60°C, 30 s; 72°C, 10 min.

Table 2. Course of the cordycepin aptamer selection

Round	RNA quantity (nmol)	[RNA] (μ M)	Pre-selection	Percentage of input RNA eluted with	
				ATP	Cordycepin
1	11	2	Yes	—	<0.5
2	2.5	1	Yes	—	0.5
3	0.35	0.2	Yes	1.4	<0.1
4	0.35	0.2	Yes	—	19.5
5	0.2	0.1	Yes	19	1
6 ^a	0.2	0.1	Yes	—	<0.1

^aOn round 6, cordycepin agarose was diluted with Tris-derivatized sepharose to a final concentration of 5 pmol cordycepin/ml sepharose. An aliquot of 200 μ l of cordycepin sepharose were used.

Sequence analysis

The RNA pool was reverse-transcribed and PCR amplified as described above and cloned into pGEM-T (Promega) for sequencing using ABI PRISM BigDye Terminators Sequencing Kit (Applied Biosystems).

K_D estimation

The K_D was estimated by spin-column chromatography. A 1 ml disposable syringe was filled with Sephadex G-50 equilibrated in TE buffer of pH 8, and centrifuged at 1500 r.p.m. for 5 min. The process was repeated until the volume of packed column was \sim 1 ml of Sephadex. An aliquot of 200 μ l of RNA solution (0.1 μ M) was heat denatured and renatured as described above, then incubated for 15 min with increasing concentration of ATP (from 10 μ M to 1 mM) containing 2 μ l of [γ -³²P]ATP (3000 Ci/mmol). Of note, incubation times up to 4 h did not change the outcome, suggesting that the equilibrium is quickly reached. The mixture was loaded onto a Sephadex G-50 column and immediately centrifuged for 5 min. The radioactivity of the eluted solution was measured by scintillation counting. A 10 μ l aliquot of the mixture before gel filtration was counted to evaluate the input c.p.m. The same procedure was performed without RNA to determine the blank values for each ATP concentration. The proportion of RNA in complex, F_{RNA} , is equal to $F_{ATP} \times [ATP]_{in}/[RNA]_{in}$, where F_{ATP} , the fraction of ATP bound to RNA, is equal to (eluted c.p.m. – eluted c.p.m. blank)/input c.p.m.; $[ATP]_{in}$ and $[RNA]_{in}$ are the input concentrations. F_{RNA} values were plotted against ATP concentration, and the K_D values were determined by fitting the data to the following equation: $F_{RNA} = [ATP]/([ATP] + K_D)$ using Kaleidagraph (Synergy software). Most of the K_D determinations were at least duplicated (see Figure 4). The SD was calculated for each of the duplicated points, and then expressed as a percentage. The weighted mean of the latter was considered as representative of the experimental error, calculated from 53 results and was 24%; this is reported with the K_D values.

Analysis of the secondary structure of aptamers

The secondary structure of RNA molecules was probed using dimethylsulfate (DMS) and carbodiimide metho-*p*-toluene sulfonate (CMCT), as described by Butcher and Burke (41). Five picomoles of RNA were resuspended in 18 μ l of 50 mM HEPES, pH 8 (or 50 mM potassium borate, pH 8 for CMCT), heated for 2 min at 80°C, 2 μ l of 1.5 M NaCl and 50 mM MgCl₂ was then added and the mixture was slowly cooled to

room temperature. DMS (160 mM final) or CMCT (30 mM final) was added and the mixture was incubated for 10 min. The modification reaction was stopped by the addition of 10 μ g of tRNA and immediate ethanol precipitation on dry ice in the presence of 0.5 M ammonium acetate. The RNA was then resuspended in 0.5 M ammonium acetate, ethanol precipitated, rinsed with 70% ethanol and resuspended in 5 μ l of TE (41). Modifications were revealed by reverse transcription using a ³²P-labelled primer according to the manufacturer's instructions (Superscript II, Invitrogen).

Secondary structure models were established using the mfold energy minimization algorithm developed by Zuker (42), and using chemical probing data to provide constraints.

Determination of aptamer specificity

An aliquot of 10 pmol of ³²P-labelled RNA aptamers were heat denatured for 5 min at 80°C in 90 μ l of 50 mM Tris, pH 7.5, and renatured by slow cooling after the addition of 10 μ l of 1.5 M NaCl, 50 mM MgCl₂. RNA was then loaded onto 500 μ l of ATP-agarose (mixture of the three media; 2 μ mol of ATP/ml; saturated and equilibrated). The affinity column was washed with 10 vol of binding buffer, and the elution was then carried out by sequential addition of 4 vol of the analogue under scrutiny at three different concentrations: 2 mM, then 5 mM and finally 10 mM. A last elution with 4 vol of 10 mM ATP was carried out to verify that binding occurred normally if no elution was observed in the previous steps, or to confirm that at least 75% of the bound aptamers were eluted by the analogues. Fractions collected along the process were counted by scintillation, and the value obtained for each fraction was divided by the sum of the c.p.m. values for all fractions. The percentage of binding was then plotted for each fraction.

Probability of finding a particular motif in a random sequence

The probability to find the zeta-shaped aptamer in a 60 nt long random sequence was calculated according to Sabeti *et al.* (43), using the characteristics of the zeta-shaped aptamer defined in Ref. (14). This aptamer ($n = 32$ nt) has 12 determined core nucleotides, 4 arbitrary positions in a loop, a minimum of 7 non-restricted Watson–Crick base pairs, one restricted base pair (G–C) and a cyclic permutation factor of 2. The redundancy of the aptamer is $r = 4^4$ (arbitrary positions) $\times 4^7$ (Watson–Crick base pairs) $\times 2 = 8.38 \times 10^6$. The probability to find a n nucleotide motif in a 1 nt long random sequence is $P = r/4^n \times (1 - n + m)!/m!(1 - n)!$ where m is the cyclic permutation factor. The probability to find the zeta-shaped aptamer in our 60 nt long random sequence is $P = 1.98 \times 10^{-10}$. For a longer version of the aptamer containing 11 non-constrained Watson–Crick base pairs, $r = 2.15 \times 10^9$ and $P = 4 \times 10^{-13}$.

RESULTS

Selection of ATP aptamers

The starting pool contained $\sim 3.5 \times 10^{14}$ different RNA sequences, derived from a 60 nt long random region flanked by two constant sequences. RNA molecules were selected on an affinity column made of three different affinity media, presenting ATP in different orientations. ATP was coupled

to an activated agarose either by its 2'/3' hydroxyl, C8 or N7. RNAs that are bound to the matrix were eliminated during the first four cycles by pre-selection on a Tris-derivatized agarose column. During these cycles, very little RNA was specifically eluted by ATP (Table 1). A sharp increase in eluted RNA was observed after the fifth round, where up to 25% of the molecules were specifically retained on the ATP-agarose column. In the penultimate round, RNAs bound on the column were counter-selected with GTP, in an attempt to increase the overall selectivity for ATP. It is worth noting that almost all the bound RNA was eluted with GTP. The same percentage of binding (25%) was recovered on the following round. After this seventh round of selection, 41 cDNA clones were sequenced. Based on their sequence homology, they can be classified into five different classes (Figure 1).

Sixty percent of the isolated sequences fall into class I, and show few variations if any (an average of 1 nt per sequence). For this reason, a pertinent secondary structure model could not be directly defined from sequence covariation data. However, an artificial phylogeny was subsequently derived (see below). Class II is a five-member family, four of which are identical. The fifth sequence (C7) has been regrouped with the four others on the basis of the presence of a 14 nt long conserved sequence. Part of this conserved region can base pair to a portion of primer one, but the rest of the sequence can be folded into several different conformations, such that a reliable secondary structure could not be established. Four clones with three different sequences were grouped in class III. These sequences can be folded into a secondary structure with two helices separated by an asymmetrical bulge with sequence conservation on both sides. Class IV and class V are unique sequences found in four and two occurrences, respectively. As all classes comprise at least two identical members, we conclude that the most commonly represented aptamers present in the final pool were identified. Surprisingly, we did not

find the ATP aptamer originally identified by Sassanfar and Szostak (18), although it was most probably present in the original pool (see Discussion) and binds to the C8 tethered ATP-agarose on which it was originally selected (data not shown).

Aptamer binding to ATP in solution is inhibited at high RNA concentration

To verify that all the isolated motifs effectively bind ATP, a representative of each family was transcribed and individually loaded on the column containing a mixture of the three differently derivatized ATP-agarose beads. For each of them, a similar fraction of about 25% of the input RNA was retained and specifically eluted by ATP. It was also verified that none of the aptamers binds to a Tris-blocked agarose matrix, and that random-sequence RNA is not retained by the ATP-agarose column. To confirm that these aptamers bind their substrate in solution, they were pre-incubated with ATP before loading on the column, and, as expected, the percentage of retention fell significantly, to <6% (see for example the class I aptamers in Figure 2).

To determine the affinity constant of the selected RNAs, a constant low concentration of ATP was incubated with an increasing concentration of ATP aptamers (from 0.1 to 30 μ M). The mixture was then submitted to 'equilibrium ultrafiltration' or dialysis to separate the RNA-ATP complex from the free ATP (44). No more than 8% of the complex could be isolated using these methods. This prompted us to test the ability of the aptamers to bind ATP as a function of their concentration. Because class I aptamers were the most abundant, and a secondary structure model with conserved sequences could be readily defined for the class III aptamers (Figure 1), we chose class I and class III aptamers to perform most of the experiments presented below. ATP class I and class III

Starting Pool

gguuggcagcagaagauagcag... (N 60) ...gucggucaagggagggauccua

ATP class I

C32 -UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGGCACUACAAGCGAGUUAUCUCCUUGAU- 25 clones

ATP class II

C20 -gcagaagauagcagCCUGUCGUUCGUUCUUGUCCGACUCCAUUCGCAGUCGAAGUG-GUAACUAUCUUUAAGG- 4 clones

C7 -gcagaagauagcagUCUGUCUGAUCCACUUCUGCGACAGGAUCCGGGGAGUGCUUUCCGUAACUAUCUUUAUGg- 1 clone

ATP class III

C27 -----uagcagAACAAGCAGAUCGGCAAGGGUUUCAUCGGGAUCGAAGGUAAACUUGAUCGUGCGCUGAAUC- 1 clone

C4 -----uagcagAACAAGCAGGAAACAUCAGACCCAUGUC-----AUAGAUUGUUUCUGAGCGCUGAAUCGCAUUC- 2 clones

C22 -23nt--GUGGAACAAGCAGGGGAAAG-----CCUGAGCGCGCACGUU- 1 clone

ATP class IV

C26 -CACGGCUAAAGCCUUAACAUCAGAGCUUACUUCUCGACUCCUUGGCCAGUCAAAUCGCUU- 4 clones

ATP class V

C25 -UUCUCUCAUCGCACGUUUGACCUAAGGGUUAACGACGGUCCAAUGGUUGUCUCCAUUUA- 2 clones

Figure 1. Sequences of the starting pool and of the five classes of ATP aptamers selected. The primer sequences are in lower case letters, and were omitted from the aptamer sequences, if not involved in pairings. Putative paired sequences are underlined, and conserved sequences are shaded. The number of occurrences of each sequence is indicated at the right.

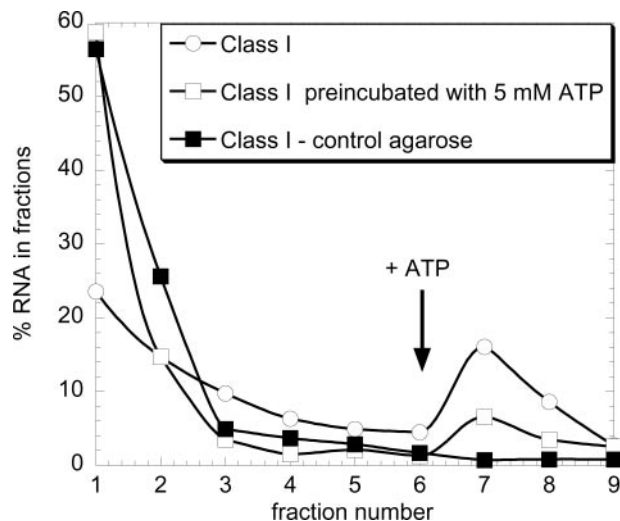


Figure 2. Single sequences specifically bind ATP in solution. Cloned aptamer sequences were individually transcribed in the presence of [32 P]UTP and loaded at 0.1 μ M on a Tris-blocked or ATP-agarose column. Aptamers were pre-incubated or not with 5 mM ATP. The column was then washed with 10 column vol of binding buffer and subsequently eluted with 4 vol of ATP 5 mM. Each fraction was then counted in a scintillation counter. Equivalent results were obtained with one member from each class of ATP aptamer.

aptamers were separately loaded onto ATP-agarose columns at different concentrations. As shown in Figure 3, binding to the column was severely inhibited by increasing concentrations of RNA, with 50% inhibition being reached around 300–400 nM. This inhibition was observed whether or not a denaturing-renaturing step was performed, suggesting that the inhibition could proceed through tertiary, rather than secondary interactions. However, we have no further indication that this is actually the case. This experiment was not performed with the three other aptamer classes. However, we observed that, from round 5, the entire pool did not bind the column at concentrations over 1 μ M. Therefore, we anticipate that the other classes of aptamers would behave similarly.

Binding affinity and salt dependence for complex formation of the ATP aptamers

We determined the binding affinity of the RNA aptamers for ATP by gel filtration on G50 spin columns (see Materials and Methods). To validate this method, we measured the dissociation constant for the zeta-shaped aptamer, and obtained a K_D of 4 ± 1 μ M, which correlates well with the 6–8 μ M value observed with other equilibrium methods by Sassanfar and Szostak (18). To reliably detect and quantify the complexes up to ATP concentrations of 1 mM, we used a constant RNA concentration of 100 nM, even though at this concentration binding is partly inhibited. Thus, the apparent K_D values should be considered as upper limits, and the actual dissociation constants may be significantly lower. Class I, class II and class III ATP aptamers have apparent K_D values ranging from 21 to 39 μ M, while class IV has much lower affinity with a measured dissociation constant of 294 μ M (Figure 4). The K_D for class V molecules, which were a minority, was not determined. At saturating concentrations of ATP, only 66–77% of the RNA bound ATP; this may reflect either a limitation of the

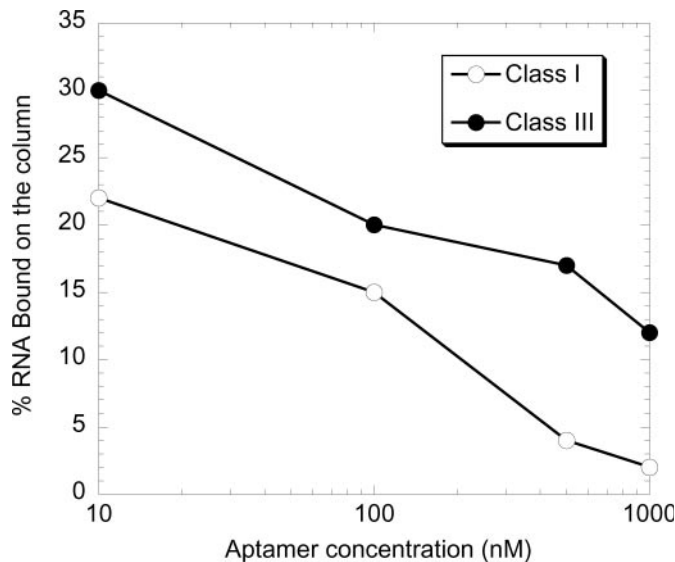


Figure 3. High RNA concentration inhibits class I and class III aptamer-binding. A constant amount of 32 P-labelled RNA aptamer (50 pmol) was loaded onto an ATP-agarose column at increasing concentrations (10 nM to 1 μ M). The column was then washed with 10 column vol of binding buffer and subsequently eluted with 4 vol of ATP 5 mM. Each fraction was then counted in a scintillation counter. The percentage of aptamer eluted by ATP is plotted against RNA concentration.

	K_D (μ M)
Class I	27 ± 7
Class II	39 ± 12
Class III	21 ± 4
Class IV	294 ± 27

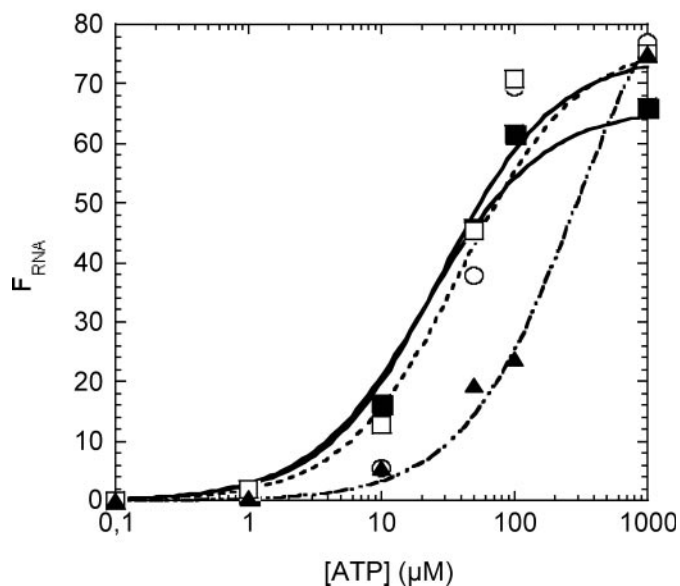


Figure 4. Affinity of the aptamers for ATP. The K_D for each aptamer and F_{RNA} , the fraction of RNA bound to ATP, were determined as described in the Materials and Methods section. Data points correspond to the mean of two (class I, II and IV) to four (class III) independent experiments.

technique or, possibly, that a fraction of the aptamers are misfolded.

The ionic requirements for aptamer–ATP complex formation were investigated by measuring the binding of the individual aptamers on ATP–agarose, in the presence of varying concentrations of NaCl or MgCl₂. The selection procedure had been performed in 150 mM NaCl and 5 mM MgCl₂. The data indicate that monovalent salts are not required and can be omitted from the binding reaction. In contrast, divalent ions are strictly required for all five classes. The ionic dependence of complex formation was studied in greater detail for the class I and class III molecules: class III molecules reach their optimal binding in the presence of 10 mM Mg²⁺, while class I molecules need less magnesium (half of the maximal retention on ATP–agarose is obtained with 2 mM Mg²⁺, data not shown). We note that such a quantitative study may be complicated by the fact that ATP itself binds magnesium. Both class I and class II molecules bind ATP in the presence of 5 mM manganese, with an efficiency comparable to what is observed in the presence of 5 mM magnesium (data not shown).

Secondary structure probing revealed five different motifs

We used two chemical reagents, DMS and CMCT, to probe the accessibility of adenines, cytosines, guanines and uracils

present in the aptamer motifs. These experiments were carried out in the presence or absence of Mg²⁺ and both with and without saturating concentrations of ATP. The residues that were chemically modified in the presence of Mg²⁺ were used as single-stranded constraints for the secondary structure prediction software mfold (42). The resulting secondary structure models are shown in Figure 5. Interestingly, most of the residues protected upon magnesium addition are located in regions predicted to be internal loops. The only noticeable exceptions are the nucleotides located in two loops of class II molecules that have the potential to base pair. This suggests that these sequences could form a pseudoknot structure which, like other pseudoknots, might be stabilized by magnesium (45–47). Addition of saturating concentrations of ATP did not induce any protection or perturbation in the modification pattern of any of the motifs, possibly because only a limited set of functional groups were probed. Alternatively, it is possible that the RNA concentration (250 nM) used for probing experiments inhibits ATP binding, and that some of the protections observed in the presence of magnesium are related to some specific aggregation of the aptamers (see Discussion).

Generation of artificial phylogenies

Diversity was reintroduced in the seventh round pool using mutagenic PCR (40). The efficiency of mutagenesis was

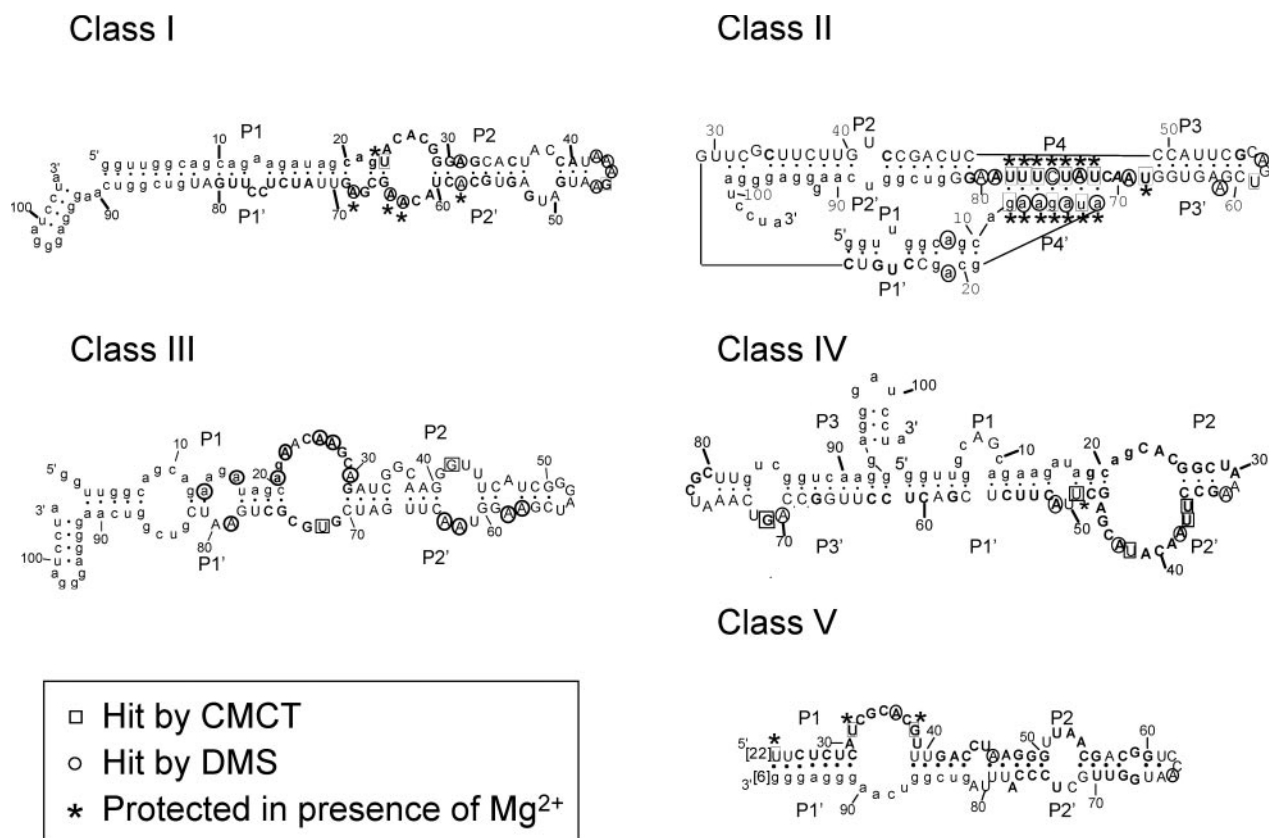


Figure 5. Secondary structure of the five ATP-binding motifs. Secondary structure models of the five aptamer motifs as defined by structure probing, sequence comparisons and energy minimization (42). The sequences shown are those of clone C14 (class I), C20 (class II), C27 (class III), C15 (class IV) and C25 (class V). Squares and circles indicate nucleotides hit by CMCT and DMS, respectively. A–U and G–U base pairs at the ends of helices have been implemented even if one of the residues is hit by a modifying agent. *P_n'*s designate the paired segments. The sequences of conserved regions (see Figure 6) are shown in bold, the variability within these sequences is indicated in Figure 7. Primer sequences were omitted from class V aptamers secondary structure model because they are not involved in any potential pairing. The number of nucleotides omitted is indicated in between brackets.

evaluated by amplifying a single sequence from class I aptamers under mutagenic conditions. The mutation frequency, including insertions and deletions, was $9.1 \pm 1.7\%$ over the 1101 nt sequenced (calculated with a confidence interval of 95%). This is in good agreement with mutagenesis frequencies previously observed using this method (40). However, we observe a strong bias for transition over transversion mutations (5:1), which differs significantly from what was reported by Vartanian *et al.* (2:1) (40). The mutagenized class I clone was re-submitted to the selection process for two rounds, after which the sequences of 13 individual clones were determined. The mutation frequency in the selected aptamers ($7.9 \pm 2.3\%$) was close to what we observed in the non-selected pool, indicating that mutants were not strongly selected against.

The mutagenized seventh round pool was submitted to secondary selection under the original screening conditions. In parallel, this pool was also subjected to two rounds of selection under competitive binding conditions, where RNA was in excess relative to ATP ($[RNA] = 100 \times [ATP]$). The two secondary selections did not give significantly different results. In particular, no striking bias in the relative class abundance was observed. Therefore, the two sets of results were pooled together for analysis. A total of 69 clones were sequenced (40 belong to class I, 2 to class II, 8 to class III, 16 to class IV and 3 to class V). The alignment of sequences obtained from the primary and the secondary selections revealed or confirmed the presence of conserved sequences in class I to class IV aptamers (an insufficient diversity was generated from class V molecules to define reliable conserved sequences) (Figure 6).

A comparison with the secondary structure models showed the following: (i) For all but the class II aptamers, most of the conserved sequences are located in internal bulges, while variable sequences are in terminal loops or in paired regions. Regions predicted to be paired with primer sequences are also essentially conserved. (ii) The secondary structure models defined by the probing experiments described above are largely confirmed by the artificial phylogeny analysis (Figures 5 and 6).

Minimal motifs were designed for class I, class III, class IV and class V aptamers according to the sequence conservation and predicted secondary structures (Figure 7). We did not take into account the few residues conserved in terminal loops of class I and class IV aptamers. We did not attempt to design a short version of class II RNA, since there was insufficient information on the length requirement between the two portions of the postulated pseudoknot. Class I, class III and class IV minimal motifs were synthesized, they bind ATP with a similar efficiency to their corresponding full length aptamers, while a class V shortened motif (spanning nucleotides 5–21 and 60–71 capped by a UUCG loop) failed to bind ATP.

Isolation of a cordycepin aptamer

The versatility of the RNA motifs and their potential to evolve into structures that could recognize an ATP analogue were evaluated. The seventh round pool of the original selection was mutagenized as described above, and subjected to a SELEX procedure to isolate RNA that selectively binds 3' deoxyadenosine (cordycepin). To increase the selectivity for

cordycepin, a counter-selection step for RNA that binds ATP was introduced on the third and fifth rounds. Finally, to retain only the best aptamers, competition between RNA molecules was introduced in the sixth round, by using a 100-fold excess of RNA over target (see Table 2). To follow the evolution of the pool, we cloned and sequenced molecules after the second, third, fourth and sixth rounds. Amongst the nine clones sequenced from round 2, five belong to class IV and four to class I. In the next round, after going through the ATP counter-selection, eight clones belonged to class IV aptamers and only one to class III. Twelve clones were analysed both after round 4 and round 6, and they were all found to belong to class IV. Remarkably, round 2 and round 3 class IV molecules still show some variations in the conserved sequences, but all 24 molecules from rounds 4 and 6 show the same conserved regions, strictly identical to those of the aptamer IV isolated in the primary selection (Figure 1). The sequences outside of the binding motif still vary greatly between clones, indicating that these molecules originated independently.

A quantitative analysis showed that the new aptamers bind cordycepin triphosphate in solution with an apparent K_D of 39 μM , which is an 8-fold tighter binding than for ATP (data not shown). Of note, the difference between the apparent K_D of the aptamers for cordycepin versus cordycepin triphosphate has not been measured. Among the other motifs isolated in the primary selection, only class III molecules bind to cordycepin, although with reduced affinity compared to ATP (see below).

The five classes of aptamers show distinct substrate selectivity

To better understand the selection results described above, we determined which of the three ATP–agarose analogues used in the column matrix were bound by each aptamer. Surprisingly, the five classes of aptamers bind to the three media with comparable efficiency (see Figure 8A for class IV aptamers). In contrast, the zeta-shaped ATP aptamer (18) is retained only in the medium on which it has been selected (C8-linked ATP, data not shown).

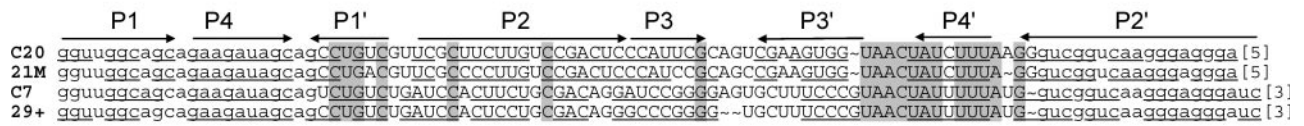
The specificity of the five classes of ATP aptamers was further studied using different ATP analogues. Labelled aptamers were bound on ATP–agarose and subsequently eluted with 4 volumes of 2, 5 or 10 mM ATP analogues (see Figure 8B and Materials and Methods). The results are summarized in Figure 9. To roughly estimate the relative affinity for the different analogues, we measured K_D values for the ligands for which a radioactive isotope was available. The class IV aptamer is eluted by 2 mM ATP and cordycepin. The apparent K_D values for these ligands are 294 and 39 μM , respectively. The dissociation constant of the class III aptamer for dATP is about 5 mM, and this RNA is eluted from the column by 5 mM dATP. The K_D determination method failed to detect any binding between the class I aptamer and dATP, indicating that the dissociation constant is over 20 mM.

Class I molecules recognize a ribose monophosphate with an efficiency comparable to ATP. The important binding determinants are therefore the phosphate, because it does not bind to ribose alone, and the 2'-OH and 3'-OH groups, because it does not bind cordycepin (3'-deoxyadenosine) and poorly binds to 2' deoxyATP.

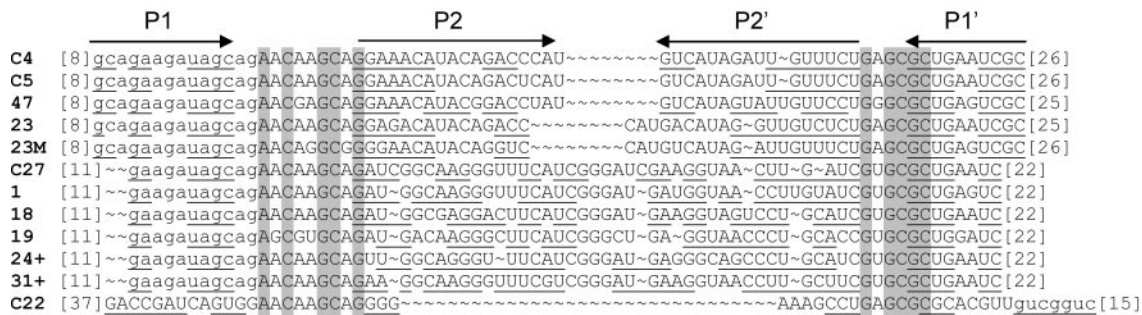
Class I aptamers

		P1	P2	P2'	P1'	
C14 (X12)	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
C3	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>CAUACAAGCGAGUUUCUCCUUGAU</u>	[22]	
1441	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCGCUACCGUAAAGAACGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
1437	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACCACCAUAGAGAUUGAUGGGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
1438	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAGUGAUGAGUGCGC</u>	<u>~CACAGGC~AGUUACCUCCUAGGU</u>	[22]	
1430	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAAAAGAAUGAUGAGCGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGUU</u>	[22]	
1432	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGACGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
10	[7]	<u>agcagaagauagcag</u>	<u>UACAUGGGAGCACUACCAUAAAGAAUGACGGGUGUAC</u>	<u>~UACAA~CGAGUUUCUCCUUGAC</u>	[22]	
12	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUGCCAUAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
25	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUAUU</u>	[22]	
29	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACGAGCG~GUUUCUCCUUGAU</u>	[22]	
37	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUGCCAUAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
38	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUGCUAUAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
43	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACCACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGUU</u>	[22]	
8M	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
16M	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGUCGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
17M	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCGCUACCGUAAAGGACGCGCAGCGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGGU</u>	[22]	
15	[7]	<u>agcagaagauagcag</u>	<u>UAC~GGGGGCACUACCAUAAAGGACGAUGAGUGCGC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
41	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGGGCACUACCAUAGAGGAUGAAGGGCGCAC</u>	<u>~UACAAGCGAGUUACCUUCUUGAU</u>	[22]	
22	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCGCUACCAUAAAGAAUGAUGAGCGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
20M	[7]	<u>agcagaagauagcag</u>	<u>UACACAGGAGUACUACUAAAGAGCAUUGCGCGU</u>	<u>~UGCAAGCGAGUUACCUCCUUGAU</u>	[22]	
22M	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAAUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGUU</u>	[22]	
1440	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGUACUACCAU~AAGGAUUGGGUACAC</u>	<u>~UACAAGCGAGCUACCUCCUUGAU</u>	[22]	
1435	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCGCUACCGUAAACAAUGUUGAACGCAC</u>	<u>~UACAAGCGAGCUAUCUCCU~GAU</u>	[22]	
1442	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCGCCUCCAUAAGGAAUGAUGAGAGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	
1443	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCAUACCAUUAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	
1444	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCGCUACCAAGAGAAUGUUGAGCGCAC</u>	<u>~UACAAGCGAGCUACCUCCUUGGU</u>	[22]	
1434	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	
24	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	
20	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	
27	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACCACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUAUU</u>	[22]	
21	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAAGAGCGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	
45	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGUACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	
12M	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	
4M	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	
2	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACCGCCAUACAGAGUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUUUCUUGAC</u>	[22]	
14	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCCAUUCUCCUUGAC</u>	[22]	
28	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCCAUUCUCCUUGAU</u>	[22]	
11M	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAAGAGUGCAC</u>	<u>~UACAAGCGAGCCAUUCUCCUUGGU</u>	[22]	
17	[7]	<u>agcagaagauagcag</u>	<u>UACAC~GGAGUACUACCAUAAAGAAUGACGAAUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGGU</u>	[22]	
1436	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACCACUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	
C1	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCUGUCUCCUUGAU</u>	[22]	
C6	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCUAGUUUCUCCUUGAU</u>	[22]	
C9	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGCCAUUCUCCUUGAU</u>	[22]	
C11	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAA~GAAUGAUGAGUGCAC</u>	<u>~UAUAAAGCGAGUUUCUCCUUGAU</u>	[22]	
C16	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAA~CAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
C17	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGGUGAGUGCGC</u>	<u>~UACAAGCGAGUCUUCUCCUUGAU</u>	[22]	
C21 (X2)	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGGUGAGUGCAC</u>	<u>~UACAAGCGAGUCUUCUCCUUGAU</u>	[22]	
C40 (X6)	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUCUUCUCCUUGAU</u>	[22]	
3+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGGUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
5+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGGUGAGUGCAC</u>	<u>~UACAAGCGAGUUACCUCCUUGAG</u>	[22]	
7+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGGUGAGUGCAC</u>	<u>~UGCAAGCGAGUUUCUCCUUGGU</u>	[22]	
9+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACCGCCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
12+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
22+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
21+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUGCCAUAAGAAUGGUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAA</u>	[22]	
25+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACCACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUCUUCUCCUUGAU</u>	[22]	
27+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGGUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
28+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUCUUCUCCUUGAU</u>	[22]	
33+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACUACCAUAAAGAAUGAUGAGUGCAC</u>	<u>~UACAAGCGAGUUUCUCCUUGAU</u>	[22]	
34+	[7]	<u>agcagaagauagcag</u>	<u>UACACGGGAGCACAACCAUAAAGGUGAUGGUGAGUGCAC</u>	<u>~UACAAGCGAGCUAUCUCCUUGAU</u>	[22]	

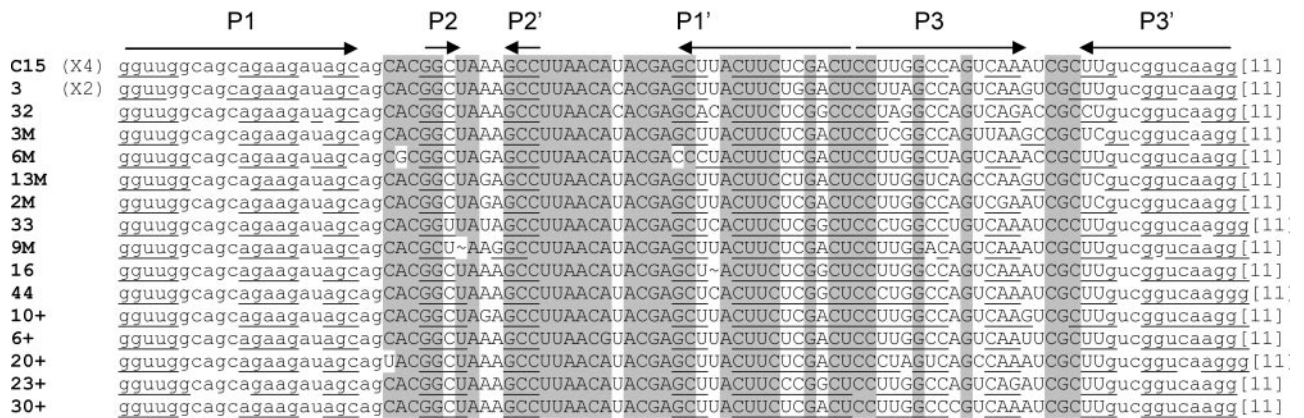
Class II aptamers



Class III aptamers



Class IV aptamers



Class V aptamers

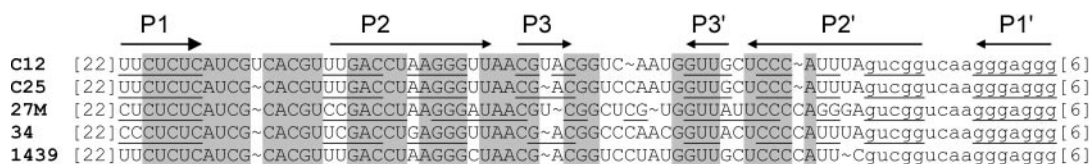


Figure 6. Alignments of the aptamer sequences yielded by the primary and secondary selections. The clone name is on the left, the number of occurrence of each sequence is mentioned when superior to one. The residues that do not vary in 95% of the sequences are shaded. Potentially paired nucleotides are underlined, and the paired regions are indicated by 'Pn' and an arrow at the top of each alignment. Alignments are based on sequence homology, potential pairings are sometimes slightly shifted from one clone to another. Arrows cover the whole region potentially involved in the pairing for all sequences. Primer sequences are in lower case letters and were not shaded even though they are invariant. They were omitted when not involved in potential pairings with a variable region, the number of nucleotides omitted is indicated in between brackets. The clones named 'n+' were selected under stringent conditions.

The binding of the class II ATP aptamers appears to have few determinants in the ribose. Rather, the principal binding determinant appears to be the gamma phosphate of ATP. It is likely to contact one of the non-bridging oxygens through a Mg²⁺ ion, because most of the binding is lost upon replacement of one of the non-bridging oxygen with sulphur and is

partially recovered in the presence of manganese ion (48). Although it was not available for us to test, its minimal substrate is likely to be a ribose triphosphate.

The class III aptamer minimal substrate is adenosine, with binding reliant upon 2'-OH, 3'-OH and the N3 of adenine, since dATP, cordycepin and N3-methyl adenine are poorly

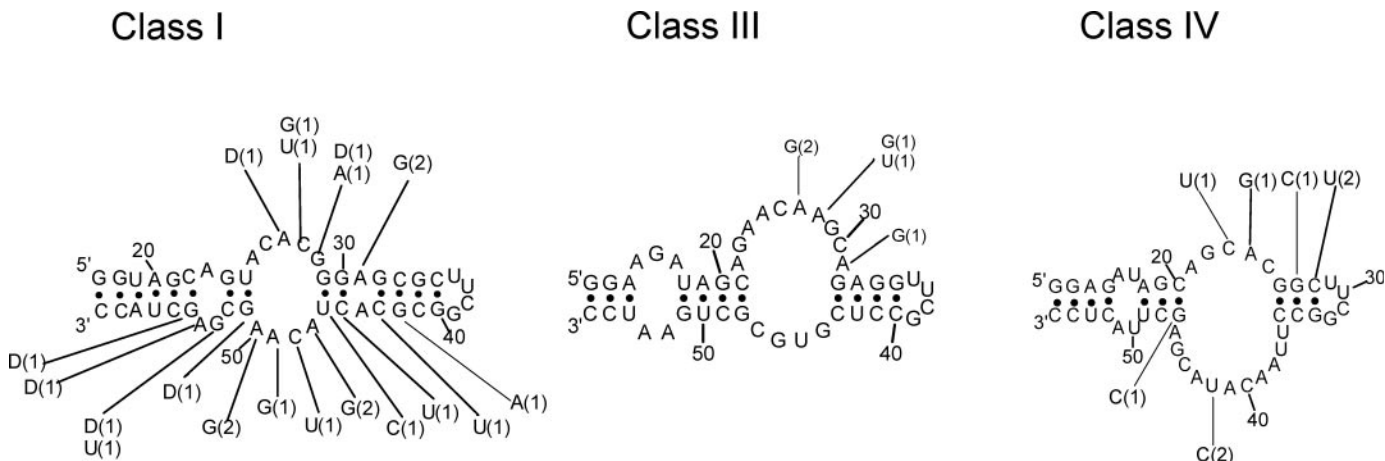


Figure 7. Minimal motifs proposed for the class I, class III and class IV aptamers. For the secondary structure models of the minimal motifs of class I, class III and class IV aptamers, alterations of the conserved sequences are indicated, and the number of occurrences of each change is mentioned in parentheses. D stands for deletion. These motifs show similar binding properties than the aptamers originally selected.

or not recognized. Nevertheless, the binding to ATP can be competed by high concentrations of UTP and GTP, showing that the most important determinants, likely, lie within the ribose and not in the base itself.

Surprisingly, class IV molecules, originally selected as ATP aptamers, appear to be more efficient at binding cordycepin. These aptamers display a low affinity for adenosine, but are efficiently eluted by ATP. Although the phosphate group is not absolutely required for binding, it does play some role, because ribose phosphate elutes the aptamer more efficiently than ribose. We find that high concentrations of adenine are required to out-compete ATP binding.

The optimal substrate for class V molecules is the entire ATP molecule. No isolated moieties or ATP derivatives had a comparable binding affinity. At high concentrations, class V aptamers could be eluted by adenosine, GTP, ribose phosphate, and at even higher concentration by *N*6-methyl adenosine. The binding pocket is therefore specific for both the nature of the base and the sugar-phosphate moiety.

DISCUSSION

The selected motifs bind to the three derivatives of ATP-agarose

To isolate new RNA aptamers other than the well-characterized zeta-shaped motif, we used a selection medium on which the substrate was linked through different positions, *N*6, 2'/3'-OH and C8. We expected the selection to yield three different classes of aptamers, each binding to only one affinity medium. In contrast, the selected molecules were those that bound to all three matrices, i.e. those that can bind to all of the substrate present. The actual concentration of the substrate moiety targeted is likely to be crucial in the very first steps of selection, when aptamers are necessarily in an infinitesimal amount. This is probably the most important selection pressure in the SELEX experiments reported here, and could be the reason why we did not find the zeta-shaped aptamer among the 110 isolated sequences. The probability to find this aptamer in a 60 nt long random sequence is in between 2×10^{-10} and

4×10^{-13} [(14,43) and see Material and Methods]. Since we started with about 3.5×10^{14} different sequences, the zeta-shaped aptamer was more than likely present in the original pool and is known to bind to one of the affinity matrices used for the selection. Nevertheless, the zeta-shaped aptamer must account for less than 1% of the aptamers from the final pool. We believe that it was out-competed by the other aptamers that have less affinity and specificity for ATP, but can bind to all three of the different ATP-agarose derivatives present in the column matrix.

Future SELEX experiments involving affinity media that incorporate ligands randomly bound at multiple sites should take this sort of phenomenon into account. Such would be the case, for example, when proteins are tethered at multiple sites through cyanogen bromide or sulphydryl targeting reagents. Our results suggest that such a protocol could result in the early invasion of molecules targeted to the most available site, even if they are not the best possible aptamers.

In our case, the 'multiple orientations' strategy led to the most striking result of this selection: the five classes of aptamers that we recovered bind ATP through its sugar-phosphate moiety. Thus, even though they show a preference for ATP, most of these aptamers should be considered as 'nucleotide-binding motifs'. The counter-selection step with GTP should have eliminated such molecules, but it was performed at the sixth round, a stage at which those binders had already invaded the pool.

The selected aptamers may aggregate at high concentrations

At concentrations over 1 μ M, the RNA pools from the fifth and later rounds, as well as the class I and class III aptamers, fail to bind ATP. This could be due to misfolding, but, in the light of the selectivity of the aptamers we isolated, we propose that aggregation, caused by a specific intermolecular binding of those aptamers to their own sugars and phosphates, is responsible for this lack of binding. The concentration resulting in half-inhibition of binding for class I and class III molecules is \sim 300–400 nM. Under these conditions, the actual

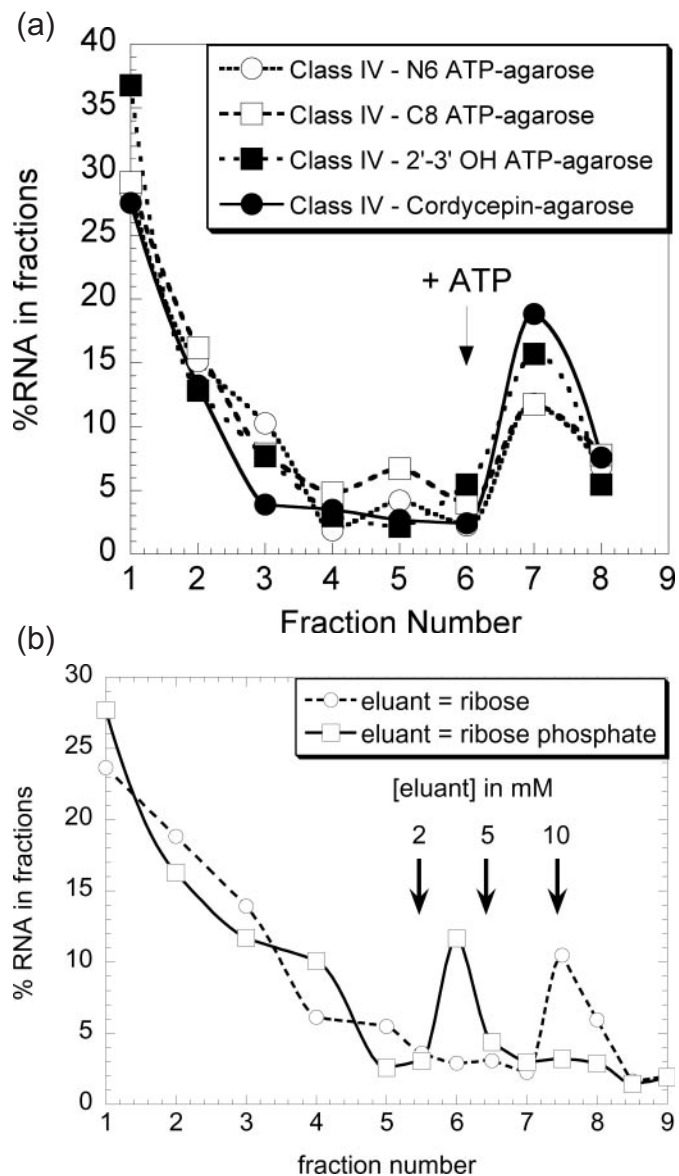


Figure 8. (a) Class IV binds to each of the three differently derivatized ATP-agarose affinity matrices and on cordycepin-agarose: 100 μ l of a 0.1 μ M solution of 32 P-labelled class IV aptamer was loaded onto each of the four affinity matrices. The columns were then washed with 10 column vol of binding buffer, and subsequently eluted with 4 vol of 5 mM ATP. Each fraction was counted and the results were plotted as percentages of the total amount of RNA. Equivalent results were obtained with the four other classes of aptamer with the three different ATP-agarose matrices. (b) Class I aptamer minimal substrate is a ribose phosphate: 100 μ l of a 0.1 μ M solution of 32 P-labelled class I aptamer was loaded onto a mixture of the three ATP-agarose matrices. The column was then washed with 10 column vol of binding buffer, and subsequently eluted with 4 vol of 2, 5 and 10 mM of ribose or ribose phosphate. A last elution with 10 mM ATP was performed to confirm that at least 75% of the bound aptamer was eluted. Each fraction was counted and the results were plotted as percentages of the total amount of RNA. This procedure was used for all the data presented in Figure 8.

concentration of ribose phosphate is \sim 33–44 μ M, which is in good agreement with the K_D observed for the binding of these molecules to ATP. The chemical structure probing of these molecules was performed at 0.25 μ M, and it is possible that

at this concentration the aggregation of aptamers is partially responsible for the protection pattern observed in presence of magnesium. Thus, some of the positions involved in binding may have been shielded from the probe, even though ATP was not present in the solution. This would be particularly relevant for those positions that are located within the internal bulges, and are likely to form binding pockets.

Unfortunately, aggregation renders these molecules unamenable to many biophysical and biochemical methods (NMR, circular dichroism, ultraviolet melting and gel shift assay). However, X-ray crystallography may be a suitable approach if conditions can be found under which the observed aggregation results in an ordered lattice.

The binding sites of all aptamers are left unchanged by different secondary selections

We conducted three different secondary selections from the primary selected pool. The first selection consisted of re-selecting the molecules under the exact same conditions of the primary selection, while the second one increased the selective pressure by diminishing the ratio of target over RNA. As stated in the results section, no important differences were observed between the two selected pools; the secondary structure and the conserved sequence remained the same. Nevertheless, a close examination of the selected molecules showed that aptamers isolated under non-stringent conditions are characterized by a suboptimal secondary structure featuring short pairings, mismatches and single nucleotide bulges. Therefore, the most productive folding can potentially be competed by alternative structures. The aptamers obtained under more stringent conditions show the same conserved sequences, which probably indicates that the binding sites have not been improved. However, we notice that in most of the re-selected molecules, the secondary structure has been strengthened, the mismatches corrected and the pairings elongated (see Figures 6 and 10 for examples). Thus, the secondary selection yielded molecules that could bind more readily to ATP-agarose, which could be related to an increased folding efficiency.

Interestingly, a close examination of the conserved regions revealed common sequences between different aptamers. Class I and class III aptamers share two stretches of conserved sequences, 5' CAG(U/A)ACA 3' (nucleotides 20–26, 20–22 belonging to the primer sequences) and 5' CAAGC 3' (nucleotides 62–66 and 25–29 in class I and class III respectively). One or the other could be the base for recognition of the ribose, which is the common determinant of the minimal substrate of these aptamers. On the other hand, this may be fortuitous since these sequences do not show the same modification pattern in both aptamers. The class I and class IV aptamer-binding sites both contain 5' CACGG 3' (nucleotides 25–29 and 23–27 respectively). Both of these classes recognize ribose-phosphate, but with different affinity and selectivity.

Isolation of a cordycepin-binding aptamer

The third secondary selection was aimed at deriving a cordycepin-specific aptamer; however, it failed to yield a new RNA motif. We had been expecting molecules whose ATP-binding pocket would have been modified to specifically accommodate cordycepin. Instead, it appears that the RNA molecules

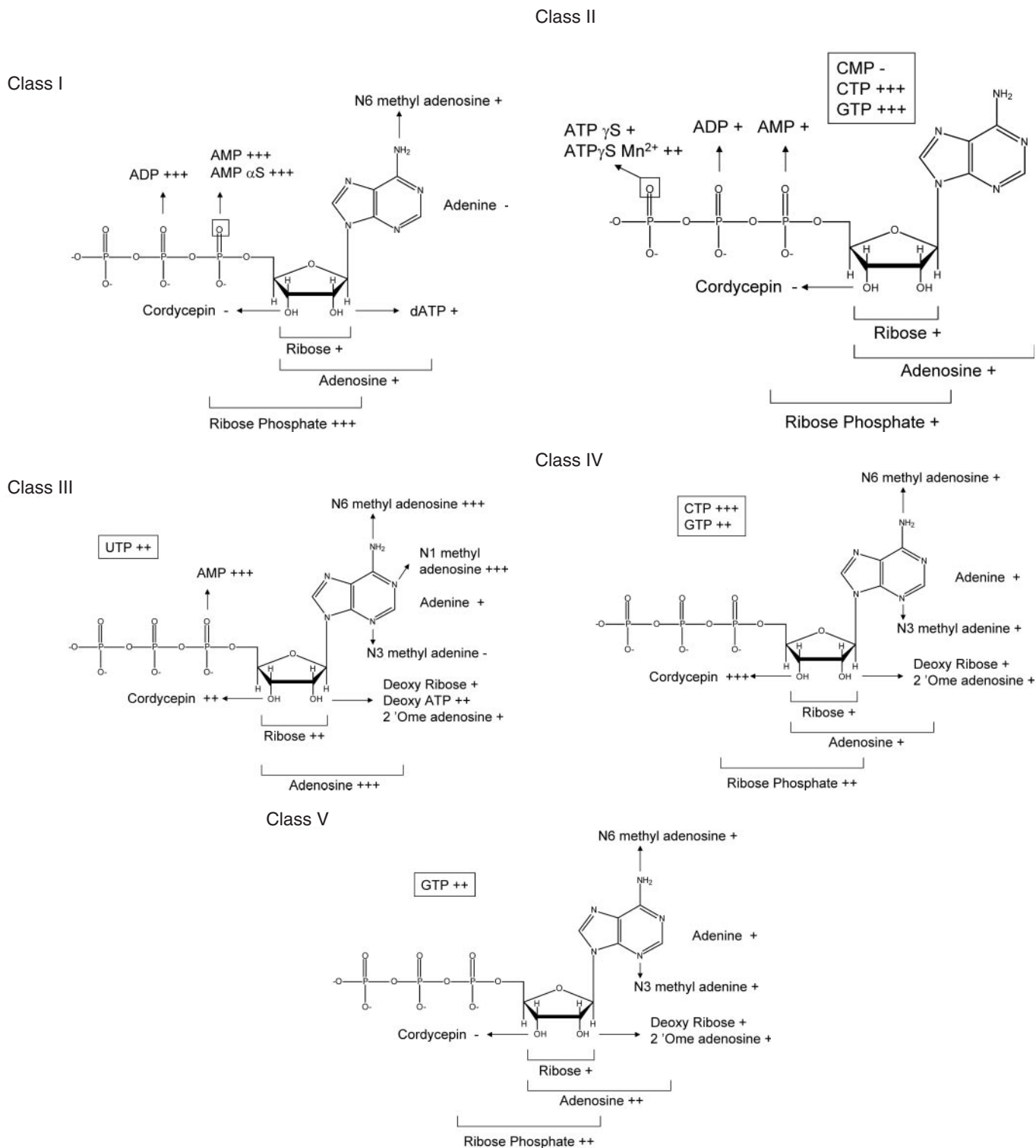
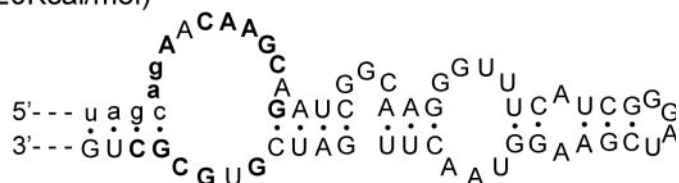


Figure 9. Specificity of the five ATP-binding motifs. Selectivity of the different aptamers was determined as specified on Figure 8B, using a set of analogues. Eluant was used at three different concentrations: (+++), (++) and (+) indicate that aptamers were efficiently eluted at 2, 5 or 10 mM respectively. The (-) sign indicates that aptamers could not be eluted by the analogue considered.

selected were the ones that had both the best affinity for cordycepin and the worst affinity for ATP. This is perfectly exemplified by the course of the selection; before counter-selection, two types of molecules are equally present (see

Results). On the one hand, class I molecules that do not significantly bind cordycepin and were present as a vast majority in the starting pool, and, on the other hand, class IV aptamers that were poorly represented in the starting pool, but have a

A – Class III aptamer from the primary selection
($\Delta G = -20\text{Kcal/mol}$)



B – Class III aptamer selected under stringent conditions
($\Delta G = -26\text{Kcal/mol}$)

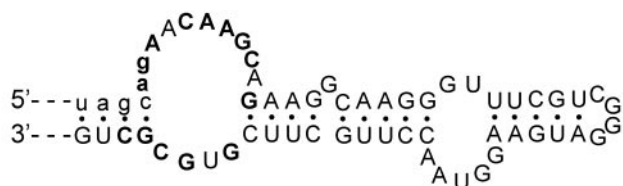


Figure 10. Selection under more stringent condition results in consolidation of aptamer secondary structure. (A) Class III aptamer isolated from the primary selection. (B) Class III aptamer selected under more stringent conditions. The parts not shown are identical in the two secondary structure models. The ΔG values are those reported by the mfold software.

good affinity for cordycepin. Only class IV molecules went through the counter-selection step with ATP and, although the diversity was not lost, only one of the binding sites previously selected by the primary selection was finally retained. What is the most likely reason that the secondary selection did not yield a novel binding pocket? The simplest explanation would be that not enough diversity had been introduced in the starting pool of the secondary selection. Since we wanted to explore the same sequence space rather than generate completely novel aptamers, the starting sequences were rather lightly mutagenized (10%). We anticipated that this should be enough, since it was previously shown that a 44 nt long citrulline aptamer could be converted into an arginine binder by only three mutations (9). Similarly, the specificity of the natural GTP-binding site present in group I introns can be changed by mutating only one base pair (49,50). In contrast, Huang and Szostak (17) reported recently that turning the zeta-shaped ATP aptamer into a GTP aptamer required a 24% mutation rate (on average) from the original motif. Likewise, it may be impossible to easily convert any of the pre-selected ATP-binding sites into a cordycepin-binding pocket. Moreover, a molecule that bound cordycepin more efficiently than ATP was already present in the unmutagenized starting pool.

Potential basis for class IV aptamer selectivity

One puzzling property of the class IV aptamers is their strong affinity for cordycepin (K_D for cordycepin triphosphate = $39\ \mu\text{M}$), with almost no affinity for adenosine, although they do bind to ATP ($K_D = 294\ \mu\text{M}$). This could reflect the fact that the main determinant for recognition lies within the ribose, potentially in the sugar pucker itself and in the influence it has on the *syn* or *anti* conformations. The sugar pucker of nucleotides and nucleosides is in dynamic equilibrium between two

main conformations: *C'2 endo* and *C'3 endo*. The relative distribution in solution of the two main conformers is influenced by the nature of the base, the electro-negativity of 2' and 3' substituents and the phosphorylation state of the 5' oxygen (51). For a given combination of these factors, the sugar pucker is correlated with the *syn* or *anti* position of the base relative to the ribose. The *syn* orientation is correlated with a *C'2 endo*, while a *C'3 endo* favors the *anti* conformation. Thus, ATP and cordycepin are essentially in the *C'3 endo-anti* conformation, but for adenosine, the different conformations are equally represented (51–53). Therefore, it is possible that the class IV aptamer-binding pocket preferentially accommodates a *C'3 endo-anti* conformation. This would also explain why these aptamers bind more efficiently to CTP (mainly in the *anti* conformation) than to GTP (mainly in the *C'2 endo-syn* conformation). Alternatively, the 3'-OH could be disadvantageous by itself (presumably due to a steric clash), but could be partially compensated by the ATP triphosphate, which provides a supplementary site of interaction. In this respect, it should be noted that the K_D , we derived for cordycepin, is likely to be overestimated, since it was actually determined using cordycepin triphosphate.

Aptamers described in this work have been isolated in order to subsequently develop RNA enzymes that use ATP as a cofactor. As they appear to be specific for nucleotides in general, they may be a good starting point for deriving polymerases that need to bind all nucleotides indiscriminately.

ACKNOWLEDGEMENTS

We would like to thank D. Menay for oligonucleotides synthesis, and F. Michel, G. Bassi and S. Butcher for critical reading of the manuscript. This work was supported by the

CNRS and a grant from le ministère de la recherche Française (ACI).

REFERENCES

- Famulok, M. (1999) Oligonucleotide aptamers that recognize small molecules. *Curr. Opin. Struct. Biol.*, **9**, 324–329.
- Gold, L., Polisky, B., Uhlenbeck, O. and Yarus, M. (1995) Diversity of oligonucleotide functions. *Annu. Rev. Biochem.*, **64**, 763–797.
- Jayasena, S.D. (1999) Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin. Chem.*, **45**, 1628–1650.
- Brody, E.N. and Gold, L. (2000) Aptamers as therapeutic and diagnostic agents. *J. Biotechnol.*, **74**, 5–13.
- Burgstaller, P., Jenne, A. and Blind, M. (2002) Aptamers and aptazymes: accelerating small molecule drug discovery. *Curr. Opin. Drug Discov. Devel.*, **5**, 690–700.
- Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820–825.
- Feigon, J., Dieckmann, T. and Smith, F.W. (1996) Aptamer structures from A to zeta. *Chem. Biol.*, **3**, 611–617.
- Patel, D.J. and Suri, A.K. (2000) Structure, recognition and discrimination in RNA aptamer complexes with cofactors, amino acids, drugs and aminoglycoside antibiotics. *J. Biotechnol.*, **74**, 39–60.
- Yang, Y., Kochoyan, M., Burgstaller, P., Westhof, E. and Famulok, M. (1996) Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy. *Science*, **272**, 1343–1347.
- Bartel, D.P. and Unrau, P.J. (1999) Constructing an RNA world. *Trends Cell Biol.*, **9**, M9–M13.
- Lorsch, J.R. and Szostak, J.W. (1994) *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature*, **371**, 31–36.
- Wilson, C. and Szostak, J.W. (1995) *In vitro* evolution of a self-alkylating ribozyme. *Nature*, **374**, 777–782.
- Burke, D.H. and Hoffman, D.C. (1998) A novel acidophilic RNA motif that recognizes coenzyme A. *Biochemistry*, **37**, 4653–4663.
- Burke, D.H. and Gold, L. (1997) RNA aptamers to the adenosine moiety of S-adenosyl methionine: structural inferences from variations on a theme and the reproducibility of SELEX. *Nucleic Acids Res.*, **25**, 2020–2024.
- Huizenga, D.E. and Szostak, J.W. (1995) A DNA aptamer that binds adenosine and ATP. *Biochemistry*, **34**, 656–665.
- Koizumi, M. and Breaker, R.R. (2000) Molecular recognition of cAMP by an RNA aptamer. *Biochemistry*, **39**, 8983–8992.
- Huang, Z. and Szostak, J.W. (2003) Evolution of aptamers with a new specificity and new secondary structures from an ATP aptamer. *RNA*, **9**, 1456–1463.
- Sassanfar, M. and Szostak, J.W. (1993) An RNA motif that binds ATP. *Nature*, **364**, 550–553.
- Meli, M., Vergne, J., Decout, J.L. and Maurel, M.C. (2002) Adenine–aptamer complexes: a bipartite RNA site that binds the adenine nucleic base. *J. Biol. Chem.*, **277**, 2104–2111.
- Davis, J.H. and Szostak, J.W. (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc. Natl Acad. Sci. USA*, **99**, 11616–11621.
- Kiga, D., Futamura, Y., Sakamoto, K. and Yokoyama, S. (1998) An RNA aptamer to the xanthine/guanine base with a distinctive mode of purine recognition. *Nucleic Acids Res.*, **26**, 1755–1760.
- Gebhardt, K., Shokraei, A., Babaie, E. and Lindqvist, B.H. (2000) RNA aptamers to S-adenosylhomocysteine: kinetic properties, divalent cation dependency, and comparison with anti-S-adenosylhomocysteine antibody. *Biochemistry*, **39**, 7255–7265.
- Lauhon, C.T. and Szostak, J.W. (1995) RNA aptamers that bind flavin and nicotinamide redox cofactors. *J. Am. Chem. Soc.*, **117**, 1246–1257.
- Dieckmann, T., Butcher, S.E., Sassanfar, M., Szostak, J.W. and Feigon, J. (1997) Mutant ATP-binding RNA aptamers reveal the structural basis for ligand binding. *J. Mol. Biol.*, **273**, 467–478.
- Dieckmann, T., Suzuki, E., Nakamura, G.K. and Feigon, J. (1996) Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA*, **2**, 628–640.
- Jiang, F., Kumar, R.A., Jones, R.A. and Patel, D.J. (1996) Structural basis of RNA folding and recognition in an AMP–RNA aptamer complex. *Nature*, **382**, 183–186.
- Burgstaller, P. and Famulok, M. (1994) Isolation of RNA–aptamers for biological cofactors by *in vitro* selection. *Angew. Chem.*, **106**, 1163–1166.
- Shu, D. and Guo, P. (2003) A viral RNA that binds ATP and contains a motif similar to an ATP-binding aptamer from SELEX. *J. Biol. Chem.*, **278**, 7119–7125.
- Unrau, P.J. and Bartel, D.P. (1998) RNA-catalysed nucleotide synthesis. *Nature*, **395**, 260–263.
- Huang, F., Yang, Z. and Yarus, M. (1998) RNA enzymes with two small-molecule substrates. *Chem. Biol.*, **5**, 669–678.
- Huang, F. and Yarus, M. (1997) 5'-RNA self-capping from guanosine diphosphate. *Biochemistry*, **36**, 6557–6563.
- Illangasekare, M., Sanchez, G., Nickles, T. and Yarus, M. (1995) Aminoacyl-RNA synthesis catalyzed by an RNA. *Science*, **267**, 643–647.
- Tsukiji, S., Pattnaik, S.B. and Suga, H. (2003) An alcohol dehydrogenase ribozyme. *Nature Struct. Biol.*, **10**, 713–717.
- Chapman, K.B. and Szostak, J.W. (1995) Isolation of a ribozyme with 5'-5' ligase activity. *Chem. Biol.*, **2**, 325–333.
- Bartel, D.P. and Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science*, **261**, 1411–1418.
- Lohse, P.A. and Szostak, J.W. (1996) Ribozyme-catalysed amino-acid transfer reactions. *Nature*, **381**, 442–444.
- Santoro, S.W. and Joyce, G.F. (1997) A general purpose RNA-cleaving DNA enzyme. *Proc. Natl Acad. Sci. USA*, **94**, 4262–4266.
- Geiger, A., Burgstaller, P., von der Eltz, H., Roeder, A. and Famulok, M. (1996) RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity. *Nucleic Acids Res.*, **24**, 1029–1036.
- Milligan, J.F. and Uhlenbeck, O.C. (1989) Synthesis of small RNAs using T7 RNA polymerase. *Meth. Enzymol.*, **180**, 51–62.
- Vartanian, J.P., Henry, M. and Wain-Hobson, S. (1996) Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Res.*, **24**, 2627–2631.
- Butcher, S.E. and Burke, J.M. (1994) Structure-mapping of the hairpin ribozyme. Magnesium-dependent folding and evidence for tertiary interactions within the ribozyme-substrate complex. *J. Mol. Biol.*, **244**, 52–63.
- Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
- Sabeti, P.C., Unrau, P.J. and Bartel, D.P. (1997) Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool. *Chem. Biol.*, **4**, 767–774.
- Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425–1429.
- Egli, M., Minasov, G., Su, L. and Rich, A. (2002) Metal ions and flexibility in a viral RNA pseudoknot at atomic resolution. *Proc. Natl Acad. Sci. USA*, **99**, 4302–4307.
- Gonzalez, R.L., Jr and Tinoco, I., Jr (1999) Solution structure and thermodynamics of a divalent metal ion binding site in an RNA pseudoknot. *J. Mol. Biol.*, **289**, 1267–1282.
- Nixon, P.L. and Giedroc, D.P. (1998) Equilibrium unfolding (folding) pathway of a model H-type pseudoknotted RNA: the role of magnesium ions in stability. *Biochemistry*, **37**, 16116–16129.
- Eckstein, F. (2002) Developments in RNA chemistry, a personal view. *Biochimie*, **84**, 841–848.
- Been, M.D. and Perrotta, A.T. (1991) Group I intron self-splicing with adenosine: evidence for a single nucleoside-binding site. *Science*, **252**, 434–437.
- Michel, F., Hanna, M., Green, R., Bartel, D.P. and Szostak, J.W. (1989) The guanosine binding site of the Tetrahymena ribozyme. *Nature*, **342**, 391–395.
- Saenger, W. (1984) *Principles of Nucleic Acid Structure*. Springer-Verlag, NY.
- Westhof, E., Plach, H., Cuno, I. and Ludemann, H.D. (1977) Proton magnetic resonance studies of 2'-, 3'-, and 5'-deoxyadenosine conformations in solution. *Nucleic Acids Res.*, **4**, 939–953.
- Guschlbauer, W. and Jankowski, K. (1980) Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Res.*, **8**, 1421–1433.